International Journal of Current Engineering and Technology
E-ISSN 2277 – 4106, P-ISSN 2347 – 5161

Available at http://inpressco.com/category/ijcet

Research Article

Hybrid Approach for Heart Disease Prediction using Machine Learning Framework

Neha S. Ahire and Shivaji R. Lahane

Department of Computer Engineering Gokhale Education Society's R. H. Sapat College of Engineering Management Studies and Research, Nashik- 5

Received 10 Nov 2020, Accepted 10 Dec 2020, Available online 01 Feb 2021, **Special Issue-8 (Feb 2021)**

Abstract

Data analytics has initiated to play a crucial role within the evolution of health care research and development. it's given tools to collect , oversee, investigate, and absorb huge volumes of unique, organized, and unstructured information produced by current healthcare systems. Big data processing has been newly practiced towards aiding the method of healthcare precaution suggestion and disease risk prediction. Medical data processing has great potential to explore the hidden models in data sets of the medical domain. These models are often used for make a clinical diagnosis these data should be collected during a standardized form. Of the medical profiles six attributes are extracted, like age, sex, vital sign and blood glucose etc. can predict the likelihood of a patient contracting heart condition . These attributes are introduced within the machine learning algorithms, classification of decision tree in heart condition prediction, applying the technique of knowledge mining to cardiopathies treatment; it can provide a reliable performance like that achieved within the diagnosis of heart condition . For these medical industries it could offer a far better diagnosis and treatment of the patient to be achieved good quality of services. the most advantages of this document are: Timely detection of heart condition and its diagnosis in time and supply treatment at a reasonable cost.

Keywords: *Machine learning, heart disease prediction, feature selection, prediction Model, classification algorithms.*

Introduction

Acute myocardial infarct , commonly mentioned as heart issues is that the commonest cause for sudden deaths in city and village areas. Detecting attack on time is of paramount importance as delay in predicting may cause severe damage to cardiac muscle , called myocardium resulting in morbidities and mortalities. Even after having severe and unbearable pain , the person may neglect to travel to a doctor thanks to several reasons including his professional reasons, personal reasons or simply overconfidence that they how they will have heart problem . repeatedly , people don't realize that the pain they're affected by could also be a attack and cause death as they're not educated on the topic . When web application and mobile application is one among the foremost widely used technology nowadays, developing an application to predict the episode of attack will yield productive leads to diagnosing of excluding attack of the pain someone is affected by heart problem . This may cause early prediction of heart attack resulting in early presentation to and evaluation by a doctor and early treatment. pain is that the commonest and significant symptom of a heart issues. although, another features also are susceptible to have attack . during this era, modern life science has been enriched with many modern technology and biological equipment that reduce the general deathrate greatly But disorder, cancer, chronic respiratory illness and diabetes are getting fatal at an alarming rate. Predicting attack on time is of paramount importance as delay in detecting may cause severe damage to cardiac muscle , called myocardium resulting in morbidities and mortalities. Acute myocardial infarct occurs when there's a sudden, complete blockage of a arteria coronaria that supplies blood to a neighborhood of heart also referred to as attack . A blockage can develop thanks to a buildup of plaque, a substance mostly made from fat, cholesterol and cellular waste products. thanks to an insufficient blood supply, a number of the guts muscles begin to die. Without early medical treatment this damage are often permanent. Medical sector is rich with information but the major issues with medical data mining are their volume and complexity, poor mathematical categorization and canonical form. We have used advanced machine learning techniques to discover knowledge from the collected medical datasets. Reducing the delay time between onset of a heart attack and seeking treatment is a major issue. Individuals who are busy in their homes or offices with

their regular works and rural people having no knowledge on the symptoms of heart attack may neglect the chest discomfort. They may not have exact intention to neglect it but they may pass on the time and decided to go to a doctor or hospital after a while. But for heart attack, time matters most. There are numerous Mobile Health apparatuses accessible to the shopper in the anticipation of CVD, for example, self observing portable applications. Current science shows the proof on the utilization of the huge swath of cell phones, for example, utilization of cell phones for correspondence and criticism, Smartphone applications. As medicinal finding of coronary episode is significant yet confused and expensive undertaking, we will proposed a framework for therapeutic determination that would upgrade restorative mind and lessen cost. Our aim is to provide a ubiquitous service that is both feasible, sustainable and which also make people to assess their risk for heart attack at that point of time or later.

Literature Survey

M.Akhil jabbar B.L Deekshatulua Priti Chandra [1] demonstrates that K-Nearest neighbor (KNN) is very simple, most popular, highly efficient and effective technique for pattern recognition. KNN is a straight forward classifier, where parts are classified based on the class of their nearest neighbor. Medical data bases are big volume in nature. If the data set contains excessive and irrelevant attributes, classification may create less accurate result. Heart disease is the best cause of death in INDIA. In Andhra Pradesh coronary illness was the best reason for mortality representing 32%of all passings, a rate as high as Canada (35%) and USA. Subsequently there is a need to characterize a choice emotionally supportive network that encourages clinicians to make prudent strides. Right now another strategy which consolidates KNN with hereditary method for compelling arrangement. Hereditary procedure perform worldwide hunt in complex huge and multimodal scenes and give ideal arrangement.

MA.Jabbar, B.L Deekshatulu, Priti Chandra [2] centers another methodology for applying affiliation manages in the Medical Domain to find Heart Disease Prediction. The human services industry gathers gigantic measure of social insurance information which, lamentably are not mined to find concealed data for powerful dynamic. Revelation of concealed examples and connections frequently goes unexploited. Information mining procedures can help cure this circumstance. Data mining have found numerous applications in Business and

Scientific domains. Association rules, classification, clustering are major areas of interest in data mining.

Chaitrali S Dangare [3] has investigated expectation frameworks for Heart infection utilizing increasingly number of information properties. The work utilizes medicinal terms, for example, sex, pulse, cholesterol like 13 ascribes to foresee the probability of patient getting a Heart ailment. Up to this point, 13 characteristics are utilized for expectation. This examination work included two additional characteristics for example stoutness and smoking. The machine learning classification algorithms, such as Decision Trees, Naive Bayes, and Neural Networks are analyzed on Heart disease database.

Amma, N.G.B [4] has proposed medical Diagnosis Systems that plays important role in medical practice and are used by medical practitioners for diagnosis and treatment. In this work, a medical diagnosis system is defined for predicting the risk of cardiovascular disease. This system is built by combining the relative advantages of genetic technique and neural network. Multilayered feed forward neural networks are particularly adapted to complex classification problems. The weights of the neural network are determined using genetic technique because it finds acceptably good set of weights in less number of iterations.

Sayantan Mukhopadhyay , Shouvik Biswas , Anamitra Bardhan Roy , Nilanjan Dey [5] demonstrated wide scope of heart condition is characterized by intensive assessment of the highlights of the ECG report. Programmed extraction of time plane highlights is important for distinguishing proof of imperative cardiovascular maladies. This work introduces a multi-goals wavelet change based framework for recognition 'P', 'Q', 'R', 'S', 'T' tops complex from unique ECG signal. 'R-R' time slip by is a significant minutia of the ECG signal that compares to the heartbeat of the related individual. Unexpected increment in stature of the 'R' wave or changes in the estimation of the 'R-R' indicate different peculiarities of human heart. Correspondingly 'P-P', 'Q-Q', 'S-S', 'T-T' likewise compares to different peculiarities of heart and their pinnacle adequacy additionally conceives other cardiovascular illnesses. Right now the 'PQRST' tops are stamped and put away over the whole sign and the time interim between two back to back 'R' tops and different pinnacles interim are estimated to discover abnormalities in conduct of heart, assuming any.

Sahar H. El-Khafifand Mohamed A. El-Brawany[6] presented that ECG signal is notable for its nonlinear changing conduct and a key trademark that is used right now; nonlinear segment of its elements changes more naturally among ordinary and anomalous conditions than does the straight one. As the higher-request insights (HOS) keep up stage data, this work utilizes one-dimensional cuts from the higher-request ghostly district of typical and ischemic subjects. A feed forward multilayer neural system (NN) with mistake back engendering (BP) learning method was utilized as a robotized ECG classifier to discover the chance of perceiving ischemic coronary illness from typical ECG signals.

M.Vijayavanan, V.Rathikarani, Dr. P. Dhanalakshmi [7] uses automatic ECG classification is a showing tool for the cardiologists in medical diagnosis for effective

treatments. In this work, propose efficient techniques to automatically classify the ECG signals into normal and arrhythmia affected (abnormal) parts. For these categories morphological features are extracted to illustrate the ECG signal. Probabilistic neural network (PNN) is the modeling technique added to capture the distribution of the feature vectors for classification and the performance is calculated. ECG time series signals in this work are bind from MIT-BIH arrhythmia database.

I. S. Siva Rao, T. Srinivasa Rao [8] predicted that heart diseases are the most extensive induce for human dying.

Every year, 7.4 million deaths are attributed to heart diseases (cardiac arrhythmia) including 52% of deaths due to strokes and 47% deaths due to coronary heart diseases. Hence identification of different heart diseases in the primary stages becomes very important for the protection of cardiac related deaths. The existing conventional ECG analysis methods like, RR interval, Wavelet transform with classification algorithms, such as, Support Vector machine K-Nearest Neighbor and Levenberg Marquardt Neural Network are used for detection of cardiac arrhythmia Using these techniques large number of features are extracted but it will not identify exactly the problem.

Saba Bashir, Zain Sikander Khan, Farhan Hassan Khan, Aitzaz Anjum, Khurram Bashir [9] describes the prediction of heart disease in the medical field through the use of data science. Because a lot of research carries out research related to that problem, the accuracy of the forecast has yet to be improved. Therefore, this research focuses on feature selection techniques and algorithms in which multiple data sets on heart disease are used for experimental analysis and to show greater accuracy.

Senthilkumar Mohan, Chandrasegar Thirumalai, And Gautam Srivastava[10] In this paper, they propose a novel method that aims at finding significant features by applying machine learning techniques resulting in improving the accuracy in the prediction of cardiovascular disease. The prediction model is introduced with different combinations of features and several known classification techniques.

Ms.M.C.S.Geetha, Dr.I.Elizabeth Shanthi, Ms.N. Sanfia Sehnaz [11] In this paper, they analyze the commonly used classification algorithms in the medical data set that helps predict heart diseases that are the main ones Cause of death throughout the world. It is complex for doctors Professionals to anticipate the heart attack as required experience and knowledge The healthcare sector today contains hidden but meaningful information to create decisions The experiments carried out reveal this algorithm As expected J48, SIMPLE CART and REPTREE Greater predictive precision than other algorithms.

Zeinab Arabasadi , Roohallah Alizadehsani , Mohamad Roshanzamir , Hossein Moosaei , Ali Asghar Yarifard[12] proposed a highly accurate hybrid method for the diagnosis of coronary artery disease. As a matter of fact, the proposed method is able to increase the performance of neural network by approximately 10 through enhancing its initial weights using genetic algorithm.

Jagdeep Singh, Amit Kamra, Harbhag Singh[13] has proposed the development of a framework based on associative classification techniques on heart dataset for diagnosis of heart based diseases. The implementation of work is done on Cleveland heart diseases dataset from the UCI machine learning repository to test on different data mining techniques. The various attributes related to cause of heart diseases are gender, age, chest pain type, blood pressure, blood sugar etc. that can predict early symptoms heart disease.

Proposed Methodology

We will propose a novel Heart attack prediction mechanism is proposed which first learns deep features and then trains these learned features. Experimental results show the classifier outperforms all other classifiers when trained with all attributes and same training samples. It is also demonstrated that the performance improvement is statistically significant. Prediction of heart attack using a low population, high dimensional dataset is challenging due to insufficient samples to learn an accurate mapping among features and class labels. Current literature usually handles this task through handcrafted feature creation and selection. Naive Bayes and Random Forest is found to be able to identify the underlying structure of data compare to other techniques.

A. Architecture

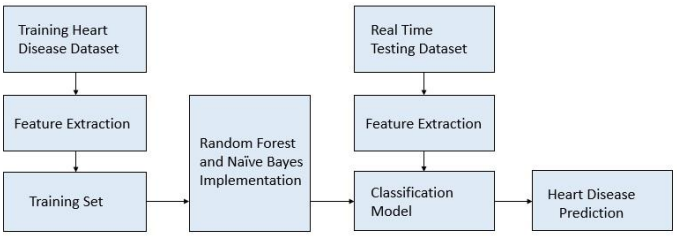

Fig. 1. Proposed System Architecture

B. Algorithms

1. Naive Bayes Algorithm:

Naive Bayes algorithm is the algorithm that learns the probability of an object with certain features belonging to a particular group/class. In short, it is a probabilistic classifier. The Naive Bayes algorithm is named ”naive” because it makes the idea that the occurrence of a particular feature is independent of the occurrence of other features. Here we classify the guts disease supported heart check up attributes. Naive Bayes or Bayes’ Rule is that the basis for several machine learning and data processing methods. The rule (algorithm) is employed to make models with predictive capabilities. It provides new ways of exploring and understanding data.

Why to prefer naive bayes implementation:

1) When the data is high.
2) When the attributes are independent of each other. 3) When we expect more efficient output, as compared to other methods output.

Based on all these information and steps we classify to predict the heart disease depending on heart check up attributes.

2. Random Forest Algorithm

Random forests is an ensemble learning algorithm. the essential premise of the algorithm is that building alittle decision-tree with few features may be a computationally cheap process. If we will build many small, weak decision trees in parallel, we will then combine the trees to make one , strong learner by averaging or taking the bulk vote. In practice, random forests are often found to be the foremost accurate learning algorithms to date.

The algorithm works as follows: for each tree in the forest, we select a bootstrap sample from S where S (i) denotes the ith bootstrap. We then learn a decision-tree using a modified decision-tree learning algorithm. The algorithm is modified as follows: at each node of the tree, instead of examining all possible feature-splits, we randomly select some subset of the features f and F. where F is the set of features. The node then splits on the best feature in f rather than F. In practice f is much, much smaller than F. Deciding on which feature to split is oftentimes the most computationally expensive aspect of decision tree learning. By narrowing the set of features, we drastically speed up the learning of the tree. Step 1: Let the amount of training set be N, and therefore the number of variables within the classifier be M.

Step 2: The number m of input variables to be wont to determine the choice at a node of the tree; m should be much less than M.

Step 3: Choose a training set for this tree by choosing n times with replacement from all N available training cases (i.e. take a bootstrap sample). Use the rest of the cases to estimate the error of the tree, by predicting their classes.

Step 4: For each node of the tree, randomly choose m variables on which to base the choice at that node. Calculate the simplest split supported these m variables within the training set.

Step 5: Each tree is adult and not pruned (as could also be wiped out constructing a traditional tree classifier).

For prediction a replacement sample is pushed down the tree. it's assigned the label of the training sample within the terminal node it finishes up in. This procedure is iterated over all trees within the ensemble, and therefore the average vote of all trees is reported as random forest prediction.

System Analysis

A. Mathematical Model

The Mathematical Model of Heart Disease Prediction is S = {s, e, X, Y, Z} s = Start of the Program e = End of the Program

X = Input

Input should be heart check up dataset.

Y = Output

Dataset will be uploading. Then the further processing will be done and finally resultant prediction should be provided.

Z = Set of System

Z= {Client, D, P, F, C}

Where,

Client, D, F, C are the elements of the set.

Client= User

D= Dataset

P= Pre-processing

F= Feature Extraction

C= Classification

B. Implementation Details

Hardware Requirement:

1) Processor : Intel I3/I5/I7
2) Speed : 1.1 GB(min)
3) RAM :2 GB Min
4) Hard Disk : 40 GB
5) Key Board : Standard Windows Keyboard
6) Mouse
7) Monitor : SVGA 8) LAN Connection

Software Requirement:

1) Operating System : Windows XP/2007/2008
2) Platform : JDK 1.8
3) Application Server : Tomcat 5.0
4) Front End : HTML, Java, JSP
5) Scripts : JavaScript
6) Server side Script : Java Server Pages
7) Editor : Eclipse IDE
8) Database : Mysql
9) Database Connectivity : JDBC

Result and Discussions

The overall accuracy of Na¨ıve bayes and Random Forest classification technique . So this works gives better heart disease prediction compare to existing method.We compared the proposed heart disease prediction accuracy on number of samples and show the result graphically. Let see the following graph and table shows the heart disease prediction accuracy result based on Na¨ıve bayes and Random Forest classification technique.

	Naive Bayes	Random Forest
Precision	60.6	52.70
Recall	85.1	87.64
F-Measure	78.8	74.31
Accuracy	90.29	92.26

Conclusion

In this work we have presented a novel approach for classifying heart disease. As a way to validate the proposed method, we will add the patient heart testing result details to predict the type of heart disease using machine learning. Train data sets taken from UCI repository.

Fig. 2. Performance Analysis Graph

Our approach use na¨ıve bayes and random forest technique which is a competitive method for classification. This prediction model helps the doctors in efficient heart disease diagnosis process with fewer attributes. Heart disease is the most common contributor of mortality in India and in Andhra Pradesh. Identification of major risk factors and developing decision support system, and effective control measures and health education programs will decline in the heart disease mortality.

Acknowledgment

I have a tremendous pleasure in presenting the project "Hybrid approach for Heart Disease Prediction using Machine Learning Framework" under the guidance of Prof. S. R. Lahane and PG coordinator Prof. A. S. Vaidya. I am obligated and appreciative to Head of Department Dr. D.V. Patil for their significant direction and consolation. I might likewise want to thank the Gokhale Education Society's R. H. Sapat College of Engineering, Management Studies Research, and Nashik5 for giving the required offices. At last, I must express my sincere heartfelt gratitude to all the Teaching Non-teaching Staff members of Computer Department of GESRHSCOE who helped me for their important time, support, remarks, and thoughts.

References

[1]. M.Akhil jabbar B.L Deekshatulua Priti Chandra International "Classification of Heart Disease Using K-Nearest Neighbor and Genetic Algorithm" Conference on Computational Intelligence: Modeling Techniques and Applications (CIMTA) 2013.

[2]. MA.Jabbar, B.L Deekshatulu, Priti Chandra, "An evolutionary algorithm for heart disease prediction"CCIS,PP 378-389 , Springer(2012).

[3]. Chaitrali S Dangare "Improved Study Of Heart Disease Prediction System Using Data Mining Classification Techniques", International Journal Of Computer Applications, Vol.47, No.10 (June 2012).

[4]. Amma, N.G.B "Cardio Vascular Disease Prediction System using Genetic Algorithm", IEEE International Conference on Computing, Communication and Applications, 2012.

[5]. Sayantan Mukhopadhyay1 , Shouvik Biswas2 , Anamitra Bardhan Roy3 , Nilanjan Dey4' Wavelet Based QRS Complex Detection of ECG Signal'

[6]. International Journal of Engineering Research and Applications (IJERA) Vol. 2, Issue 3, May-Jun 2012, pp.2361-2365

[7]. Sahar H. El-Khafifand Mohamed A. El-Brawany, "Artificial Neural Network-Based Automated ECG Signal Classifier", 29 May 2013.

[8]. M.Vijayavanan, V.Rathikarani, Dr. P. Dhanalakshmi, "Automatic Classification of ECG Signal for Heart Disease Diagnosis using morphological features". ISSN: 2229-3345 Vol. 5 No. 04 Apr 2014.

[9]. I. S. Siva Rao, T. Srinivasa Rao, "Performance Identification of Different Heart Diseases Based On Neural Network Classification". ISSN 09734562 Volume 11, Number 6 (2016) pp 3859-3864.

[10]. Saba Bashir, Zain Sikander Khan, Farhan Hassan Khan, Aitzaz Anjum, Khurram Bashir," Improving Heart Disease Prediction Using Feature Selection Approaches", International Bhurban Conference on Applied Sciences Technology 2019.

[11]. SENTHILKUMAR MOHAN, CHANDRASEGAR THIRUMALAI,

[12]. AND GAUTAM SRIVASTAVA" Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques"IEEE ACCESS 2019

[13]. Ms.M.C.S.Geetha, Dr.I.Elizabeth Shanthi, Ms.N. Sanfia Sehnaz," Analyzing the Suitability Of Relevant Classification Techniques On Medical Data Set For Better Prediction" International conference on I-SMAC 2017.

[14]. Zeinab Arabasadi , Roohallah Alizadehsani , Mohamad Roshanzamir , Hossein Moosaei , Ali Asghar Yarifard " Computer aided decision making for heart disease detection using hybrid neural network-Genetic algorithm" Science Direct 2017.

[15]. Jagdeep Singh, Amit Kamra, Harbhag Singh" Prediction of Heart Diseases Using Associative Classification"2016.