

Heart Disease Prediction Systems and a comparison of methods to improve Public healthcare.

*Sharan Shetty¹

Mukesh Patel School of Technology Management and Engineering NMIMS(Deemed-to-be) University, Maharashtra, India

Abstract: Around 17.9 million people die of cardiovascular diseases every year. 80% of the deaths occur in low and middle-income countries. Most of the population doesn't have sufficient resources to conduct expensive checkups to detect such diseases. As a result, the disease manifest and then ends up taking countless innocent lives. Consequently, there is a need to build up some strategy/way with the goal that it is conceivable to foresee the ailments inexpensively. In this paper, a comparison between the different types of heart disease prediction systems, based on their precision, accuracy, and reliability has been discussed. Methods like MLP, SVM, ANN, Random Forests have been used to early predict heart diseases in individuals.

Keywords: Healthcare, Heart, Disease, Machine learning, prediction algorithms, Cardiovascular, Support Vector Machine, Random Forest, Multi-Layer Perceptron, Artificial Neural Network.

1 Introduction

Lately, the measure of the danger of death because of heart sicknesses has gone up 85%. To fight this a ton of clinical examination is being financed by both private, as well as government organizations. Hopkins medication provided cardiovascular research which ended up being probably the greatest research identified with cardiovascular sicknesses Deaths due to cardiovascular diseases in India increased from 1.4 million in 1991 to 2.9 million in 2017, and more than half the deaths caused by heart sicknesses in 2016 were in persons less than 75 years of age, according to the study. That is 28% of the infected populace cleaned because of coronary illness alone. In the previous decades, the measure of heart cases has been expanding at troubling rates Most of the common reasons behind this incorporate, smoking, drug misuse, stress, terrible eating regimen, and so forth. The point here is to distinguish coronary illness by utilizing AI. For this, a ton of data, running from Age, sex to Body Mass Index would be considered. An illness is a specific unusual condition that contrarily influences the structure of an individual. As such, perceiving certain diseases has been challenging using human insight alone. Computer-based intelligence is assisting this condition.

However, there are two problems when considering this. First is the detection of heart disease. There are a lot of types of heart diseases. As such there are tons of parameters that are needed to be considered when detecting heart diseases. The Age of an individual, Sex, both resting and active Blood Pressure, the Heart Rate, Diabetic or Non- Diabetic sugar levels, Hyper cholesterol, BMI all contribute to these parameters. Because most of these parameters are complex/ use expensive instruments to be determined, the average population refrain from regular heart-checkups. This is the second major issue when it comes to the prevention of heart diseases. In these dire situations, with increasing economic needs, there is a need to find a way or develop a system that can predict/detect heart diseases in a human, by using some of the parameters already mentioned. Such a system would take into account, those parameters which are easier to determine, like age, sex, bpm, etc. This would help the common folk to get their heart check-ups regularly, at a reasonably cheaper price/ or for free, to detect and diagnose heart diseases, and to get treatment as soon as possible.

The remaining portion of the paper is designed as follows. Section 2 provides a summary of the problems in existing methods. Section 3 provides solutions to those problems mentioned in 2. Section 4 talks about how the various machine learning models compare against each other. Section 5 concludes the research with future studies and puts light on the scope of this research.

2 Problems in Existing Methods

In this section, some of the problems existing in the algorithms have been discussed.

2.1 Problem with the Naïve Bayes Model

The author has used the Naive Bayes method, which considers that all of the attributes (parameters) are independent of each other. Such an assumption cannot be used for computing heart disease parameters, as all of them are somehow related. Age affects an individual's heart rate. Such relations cannot be taken into account when using the Naive Bayes system. One of the issues with the Naive Bayes is the mutual independence issue as mentioned above [4]. Also, the author uses parameters that have to be obtained by expensive clinical tests like ECG, CT scans, etc. Such tests are complex, and let alone very expensive for the common people to afford. The idea here is to make a system that is accurate and less complex, but also affordable by people. So that more people make use of this and public health care benefits out of it.[5]

2.2 Problem with exhaustive search methodology

A lot of techniques have been used to make data predictions and have helped to improve the accuracy of machine learning models while reducing training time. One of the methods is to evaluate all possible combinations by training and testing the model with each dataset. E.g.: If there are 15 features then there are $2^{15}=32768$ possible combinations. If a system takes 1 minute to train and test every model, then it'll take 32768 minutes, i.e. 546 hours, i.e. 22 days. Hence it is impractical.

2.3 Multiple factors for detecting Heart Diseases.

Heart diseases can be controlled, or rather cured if and only if they are analyzed at an earlier stage. As such detecting heart diseases at an earlier stage is difficult. Age, smoking, sex, family history, cholesterol, poor diet, high blood pressure, physical inactivity, obesity, and alcohol intake are considered to be risk factors for heart disease. Also, hereditary factors like diabetes, high blood pressure can lead to heart diseases. Some of these risk factors are controllable. However, the main problem here is the early detection of heart disease. To predict a heart disease a patient has to go through several scans and tests. These tests are quite complex, let alone expensive. Then the patient needs a cardiologist to look into the data and then draw conclusions. Doing this regularly to predict heart diseases is not feasible, and very expensive. Hence, there is a need to develop a system that can efficiently determine heart diseases in an individual, by only using a limited amount of patient's data with high accuracy and efficiency. The patient can then get all of the parameters checked by a medical professional if the system shows the risk of having a heart disease.[6]

3 Solutions/Countermeasures

3.1 MLP (Multi-Layer Perceptron)

Multi-Layer Perceptron comes under Artificial Neural Networks (ANN). It is a class of feed-forward networks with multiple layers of perceptron having threshold activation. [1]

The author uses MLP (supervised neural network algorithm) with 3 layers. Input, hidden, and output layer. Input is connected to the output with the help of hidden layers. A bias and activation function are given to the inputs to predict outputs.

The MLP is an unsupervised learning algorithm. It uses a method called backpropagation for training the data. The program is implemented using python on PyCharm IDE. It uses SCI-kit learn module. Every MLP network has an activation function. This activation function is a linear function that maps inputs to the output at each node with a weight that is to be defined by the programmer. To build this, the MLP Classifier uses 'ReLU' activation function with 100 hidden layer nodes. There are a lot of activation functions. ReLu, however, is more frequently used nowadays as it easily overcomes the numerical problems related to sigmoid. Such a model can take into account all the different interrelated parameters. One of the fundamental focal points of the MLP is that it can recognize information that isn't straight detachable. Circulatory strain can be influenced by age, and this can be appeared by utilizing the MLP calculation. The more individuals, utilize this model, the more will know about their heart condition, and the number of individuals passing on because of heart illnesses will decrease inevitably.

3.2 Random Forest Model

The author has proposed random forest model to fight the issue of traditional ML models. The Random Search method is far faster and more controllable than these traditional models.[8] That is, if a given result is not desirable, then it is easy to change and repeat the process and many times as the user wants. The Random Forest algorithm, which is a large number of individual decision trees that work as an ensemble. In random forest, each decision tree is obtained from the training sets. Each tree in the model is different from the other. During the classification process, each tree takes part. Here two parameters play a decisive role in the classification. They are the number of trees and the depth of each tree forming the random forest. In general, the more trees in an ensemble, the more robust is the prediction and higher is the accuracy. Further, a confusion matrix is plotted, which helps to find the performance of the classifier and to evaluate the accuracy and precision of the model.

3.3 SVM and ANN

SVM (Support Vector Machine)

Support Vector Machine works by comparing and contrasting 2 data groups and finding suitable hyperplane between the two.[3] The classification is usually better if there is a large difference between the data groups. The result of the prediction, i.e. Yes or No will depend on which side the output lies of the hyperplane. The author starts with pre-processing the data, i.e. removing all the null fields in the data, removing outliers, and replacing the null values with average values. Linear yielded the maximum score for the given parameters.

The kernel in SVM uses two functions: Linear and Non-Linear. A hyperplane is then found by comparing and contrasting the data points of the first and second class. A function is mapped on a graph and the hyperplane is drawn that separates the classes into two outputs: “Yes” or “No”.

SVM yields great results. It gives 84.7% accuracy and 85.6% precision. In terms of sensitivity, it yields 84.12%

ANN (Scaled Conjugate gradient method)

An Artificial Neural Network reproduces the working of a human mind by utilizing neurons (Artificial Intelligence).[4] Information is given to the ANN, which at that point it uses to learn and create yields. Complex issues that are difficult to be unraveled by people or by factual methods have been comprehended by utilizing the ANN model. The system has hundreds or thousands of neurons which convey data. To place it into point of view, the human mind houses several billions of neurons, though a neural system just has hundreds or thousands. Consequently, there is a ton of possibilities in this field.

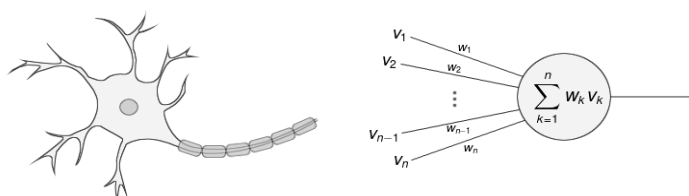


Fig 2

Artificial Neural Network is implemented using the Scaled conjugate gradient method (SCG) on MATLAB. SCG determines and adjusts the step size at each iteration. In SCG a search is performed along the conjugate direction. The data are divided on a random basis. The network can be trained and tested on random values. The author further moves on with the algorithm of the

classification. This is a seven-step process. To begin with, the dataset is imported, any invalid qualities are supplanted by the mean and the dataset is initialized. At that point, he sets the necessary number of concealed layers (input and hidden). Outputs are received from the added weight of input parameters to the hidden layer. Then an Epoch is initialized. An epoch is like a counter. It alludes to the number of times the training/ input vectors are initialized to train the dataset. Then the results are drawn. ANN yields great results. It gives 81.8% accuracy and 83.3% precision. In terms of sensitivity, it yields 68.7%

4 Analysis and Review

Comparison Chart

Solution Name	Concept	Advantages	Disadvantages
SVM	Support Vector Machine works by comparing and contrasting 2 data groups and finding suitable hyperplane between the two.	SVM works relatively well when there is a clear margin of separation between classes.	SVM is not suitable for large datasets. The computation time increases exponentially for large datasets.
ANN	An Artificial Neural Network reproduces the working of a human mind by utilizing neurons	It performs better and faster than the other algorithms mentioned in the literature.	ANN requires a better processor for more computation.
MLP	Multi-Layer Perceptron comes under is a class of feed-forward networks with multiple layers of perceptron having threshold activation.	MLP gave better/accurate predictions than the ones before it. ANN has the ability to learn complex relationships.	MLP Each node is connected to another in a very dense web resulting in redundancy and inefficiency.
Random Forests	The author has proposed random forests model to fight the issue of traditional ML models. The Random Search method is far faster and more controllable than these traditional models.	It has strategies for adjusting blunder in class population unbalanced data sets.	Random forests have been seen to overfit for some datasets with noisy classification/regression tasks.

5 Conclusion and future work

Although the rise in number and seriousness of heart diseases keeps increasing as the population continues to increase, the methods to prevent these have also come up more and more in recent years. The proposed solutions in this paper are prevention, detection, and analysis-based to cover the various kinds of ways in which heart disease can be prevented. In the following papers, it was found out that SVM yielded the best results with the highest accuracy and precision. Moreover, the number of parameters used for SVM data prediction is far less than any other model which makes it the cheapest one. This makes it feasible to be used by people from all types of communities.

This paper aims to highlight the common issue of heart disease, and its prediction and to obtain a method which can accurately determine its presence in an individual. As ML develops, there is hope for much better developments in the field which will provide better accuracy/precision. Better hardware can also be developed in the future which will help in more complex computations, and yield better results in Artificial Neural Networks.

References

- [1] A. Gavhane, G. Kokkula, I. Pandya, and K. Devadkar: Prediction of Heart Disease Using Machine Learning, 2018 Second International Conference on Electronics, Communication, and Aerospace Technology (ICECA), Coimbatore, 2018.
- [2] S. Radhimeenakshi: Classification and prediction of heart disease risk using data mining techniques of Support Vector Machine and Artificial Neural Network, 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, 2016.
- [3] A. Ed-Daoudy and K. Maalmi: Real-time machine learning for early detection of heart disease using big data approach, 2019 International Conference on Wireless Technologies, Embedded, and Intelligent Systems (WITS), Fez, Morocco, 2019.
- [4] Shadab Adam Pattekari and Asma Parveen: Prediction System for heart disease using Naïve Bayes, Department of Computer Science and Engineering Khaja Banda Nawaz College of Engineering.
- [5] A. N. Repaka, S. D. Ravikanti, and R. G. Franklin: Design and Implementing Heart Disease Prediction Using Naïves Bayesian, 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI), Tirunelveli, India, 2019.
- [6] Yanwei Xing, Jie wang, and Zhihong Zhao Yonghong Gao: Combination data mining methods with new medical data to predicting outcome of coronary Heart Disease, Convergence Information Technology international conference.
- [7] M. M. Rathore, A. Paul, A. Ahmad, M. Anisetti and G. Jeon: Hadoop-based intelligent care system (hics): Analytical approach for big data in iot, ACM Transactions on Internet Technology (TOIT)

[8] A. Javeed, S. Zhou, L. Yongjian, I. Qasim, A. Noor, and R. Nour: An Intelligent Learning System Based on Random Search Algorithm and Optimized Random Forest Model for Improved Heart Disease Detection, in IEEE Access.