

Article

Heart Disease Forecast using Machine Learning Methods

Sapna Bansal¹, Deepika Upadhyay², Jeetendar Saini³

¹M.Tech Scholar, ^{2,3}Assistant Professor, Department of Computer Science Institute of Engineering & Technology, IET, Alwar, Rajasthan, India.

INFO

Corresponding Author:

Sapna Bansal, Department of Computer Science Institute of Engineering & Technology.

E-mail Id:

sapnaagrawal@gmail.com

Orcid Id:

<https://orcid.org/0000-0002-7602-3132>

How to cite this article:

Bansal S, Upadhyay D, Saini J. Heart Disease Forecast using Machine Learning Methods. *J Adv Res Instru Control Engg* 2020; 7(1): 12-14.

Date of Submission: 2020-02-26

Date of Acceptance: 2020-03-15

ABSTRACT

AI is a part of computerized reasoning that permits PC frameworks to gain legitimately from models, information, and experience. The calculations get an info esteem and foresee a yield for this by utilizing certain factual strategies. Arrangement is an amazing AI method that is generally utilized for forecast. The primary point of AI is to make clever machines which can think and work like people. In social insurance, AI is making frameworks that can assist specialists with giving progressively precise or powerful analyses for specific conditions.

From last few decades Heart related diseases are the leading cause of death worldwide and has emerged as the most life-threatening disease. For this, multiple machine learning approaches used to understand the data and predict the HF chances in a medical database. Here in this paper, we have discussed about machine learning and various algorithms used for prediction of heart diseases.

Keywords: Machine Learning, Heart Disease, Decision Tree, SVM, Random Forest, Naive Bayes

Introduction

Machine learning, by its definition, is a field of computer science that evolved from studying pattern recognition and computational learning theory in artificial intelligence.

It is the building of algorithms that can learn from and make predictions on data sets. Our topic is about prediction of heart diseases by processing patient's datasets to whom we need to predict the chance of occurrence of a heart disease. Some major heart diseases are:

Heart is another important organ of the human body after brain it pumps blood to every part of our body. If it fails to function correctly, then the brain and various other organs will stop working, and within few minutes, the person will die. According to WHO (World Health Organization) 17 million people die every year because of heart diseases. One-third of all global deaths are due to heart diseases.

Smoking, family history, cholesterol, poor diet, high blood pressure, physical inactivity, and alcohol intake are careful to be risk factors for heart disease.

Types of Machine Learning

Machine learning methods are categorized into following categories: Supervised Learning, Unsupervised learning, Reinforcement learning. Supervised learning and Unsupervised learning are well established approaches and most commonly used but Reinforcement learnings are newer and complex but it has generate incredible results.

Supervised Learning

The Goal of supervised learning is to learn the mapping between a set of input and output. In this method, input and output is provided to the computer along with feedback during the training. Supervised learning has two cases Classification and Regression.



Table 1.Different types of Heart disease

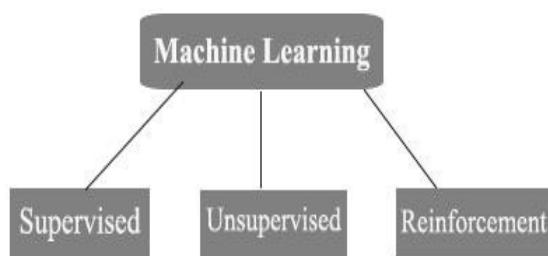
Heart Disease	Reasons
Cardiac arrest	An unexpected loss of heart function, consciousness and breathing occur suddenly.
Congestive heart failure	The heart does not pump blood as well as it should, it is the condition of chronic
Knock	occurs due to strengthening, blocking or lessening of blood vessels
Stroke	Interruption of blood supply occur damage to the brain.
Arrhythmia	When heart beat loses its regular rhythm whether it may too fast and too slow.
Heart Failure	When the Heart does not pump blood around the body efficiently.
Congenital heart disease	Some deformities of the heart that have been present since birth.

Unsupervised Learning

In this case, no such training is provided leaving computers to find the output on its own. Unsupervised learning is mostly applied on transactional data. The most communal form of unsupervised learning is Clustering and Dimensionality reduction.

Reinforcement Learning

The main goal in reinforcement learning is to find the best possible policy. This is less common and much more complex method of machine learning.

**Figure 1.**Types of Machine Learning

Machine Learning Techniques For Heart Disease Prediction

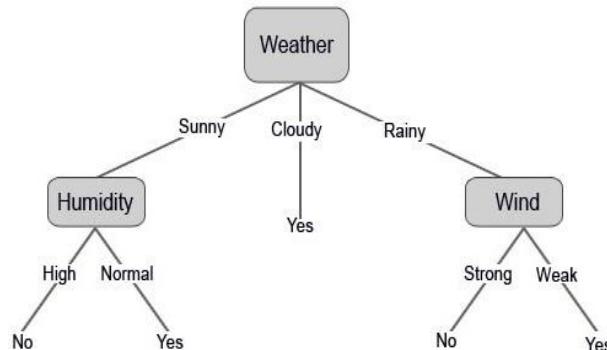
This research used four different models to predict the heart disease using collected dataset i.e. Naïve Bayes, Decision Tree, SVM, Random Forest.

Decision Tree

Decision tree is of supervised learning algorithm. It's a tree like classification model, which built a structure consisting

of branches and nodes according to the number of entities described in the dataset. It is the association amongst info gain and classified information. Decision tree classify instances by sorting them down the tree from the root to some leaf node. It gives 86.3% accuracy for testing phase and 87.3% accuracy for training phase.

In the fig this decision classifies the day according to the weather that it is suitable for playing tennis or not.

**Figure 2.**Example of Decision Tree

Naïve Bayes

The next classifier used in this study is known as Naïve Bayes. Naïve Bayes is a simple but an effective classification technique which is based on the Bayes Theorem. Naïve Bayes is a common approach used to predict classes for different types of dataset such as educational data mining and medical data mining. Naïve Bayes uses the following equation for measuring the probability: -

$$P(c|x) = \frac{P(x|c)P(c)}{P(x)}$$

Likelihood Class Prior Probability
 ↓ ↑
 Posterior Probability Predictor Prior Probability

Random Forest

Random forest is the next model selected and implemented in this research. This is also used for classification technique. As the name suggests, Random Forest technique considers multiple random trees called forest before giving an output. Random forest is used to predict coronary heart disease and it obtains an accuracy of 97.7%. The purpose of merging the random trees through voting in a forest is to opt out the highest forecasted tree, which can enhance the prediction accuracy for future data.

Support Vector Machine (SVM)

Support Vector Machine is an extremely popular supervised machine learning technique which can be used as a classifier

as well as a predictor. In this we perform classification by finding the Hyper-plane that differentiates between two classes very well. SVM is a representation of the learning data as points in space. SVM has implemented on medical data application to predict the accurate class for the heart disease patient. SVM is used with boosting technique to give an accuracy of 84.81%.

Applications of Machine Learning

Machine learning has several benefits it makes decision making faster, it provides the ability to adapt to new changing environment rapidly. Machine learning helps in sympathetic unique data patterns and the quality of outcome will be improved.

Following are some of the applications of machine learning:

- Cognitive Services
- Medical Services
- Language Processing
- Business Management
- Image Recognition
- Face Detection
- Video Games
- Computer Vision
- Pattern Recognition

Conclusion

In the above paper we have studied various classification algorithms that can be used for classification of heart disease databases it can be concluded that there is a huge scope for machine learning algorithms in predicting heart related diseases. The ratio of heart failure patients has been increasing every day. To overcome this dangerous situation and deteriorate the chances of heart failure disease, there is a need of a system that can generate rules or classify the data using machine learning approaches. Therefore, this research discussed, proposed and implemented a machine learning model by combining four different algorithms.

Decision trees performed very poor in some cases. Random Forest has performed very well because they solve the problem of over fitting by employing multiple random trees. Models based on Naive Bayes classifier were computationally very fast and have also performed well. SVM performed extremely well for most of the cases.

References

1. Bouali H, Akaichi J et al. Comparative study of Different classification techniques, heart Diseases use Case. 2014 13th International Conference on Machine Learning and Applications. 2014.
2. Pouriyeh S, Vahid S, Sannino G et al. A Comprehensive Investigation and Comparison of Machine Learning Techniques in the Domain of Heart Disease.
3. Krishnaiah V, Narasimha G, Chandra NS. Heart Disease Prediction System using Data Mining Techniques and Intelligent Fuzzy Approach.
4. Sudhakar K, Manimekalai M. Study of Heart Disease Prediction using Data Mining.
5. Dwivedi AK. Evaluate the performance of different machine learning techniques for prediction of heart disease using ten-fold cross-validation. 2016.
6. Shahi M, Gurm RK. Heart Disease Prediction System using Data Mining Techniques. Orient J. Computer Science Technology.
7. Bouchaffra D, Ykhef F. Mathematical model for machine learning and pattern recognition.
8. An Introduction to Machine Learning Theory and Its Applications: A Visual Tutorial with Examples by Nick Mccrea.