

Chapter 18

Heart Disease Prediction Using Machine Learning

Shiva Shanta Mani B.

Indian Institute of Information Technology, Kottayam, India

Manikandan V. M.

SRM University AP, Andhra Pradesh, India

ABSTRACT

Heart disease is one of the most common and serious health issues in all the age groups. The food habits, mental stress, smoking, etc. are a few reasons for heart diseases. Diagnosing heart issues at an early stage is very much important to take proper treatment. The treatment of heart disease at the later stage is very expensive and risky. In this chapter, the authors discuss machine learning approaches to predict heart disease from a set of health parameters collected from a person. The heart disease dataset from the UCI machine learning repository is used for the study. This chapter discusses the heart disease prediction capability of four well-known machine learning approaches: naive Bayes classifier, KNN classifier, decision tree classifier, random forest classifier.

INTRODUCTION

Nowadays heart disease is one of the leading causes of death in middle-aged people. Heart diseases such as coronary heart disease, heart attack, congestive heart failure, and congenital heart disease are very common in both men and women. As per the study conducted by centers for disease control (CDC), heart disease is a major reason for death in countries like United States, United Kingdom, Australia, and Canada. It is surprising that one in every four deaths in the United States as a result of some kind of serious heart disease. There are different reasons for heart diseases such as smoking habits, eating foods with high-fat content, lack of exercise, etc. The precaution for heart disease includes quitting smoking, controlling the blood pressure, maintaining proper body mass index, and regular exercising.

DOI: 10.4018/978-1-7998-2742-9.ch018

There are two approaches to treat heart diseases: medication and surgery. The heart diseases in its initial stage can be cured using the medication, but the same disease in later stages may need surgical treatment. The heart disease treatment at the later stage is expensive and there is a huge risk even though the advanced surgical technologies are available for treatment. So it will be useful if there is an automated system to predict heart disease. In this paper, we discuss a set of machine learning approaches for heart disease prediction. The experimental study has been carried out using the UCI machine learning repository for heart diseases.

The rest of this chapter is discussing the related work, the proposed scheme, the future research directions, and the conclusion.

BACKGROUND

This paper aims at analyzing the various machine learning approaches for heart disease prediction. A few relevant works in this area are discussed in this section. Heart disease is a term that assigns to a large number of medical conditions related to the heart. These medical conditions describe the abnormal health conditions that directly influence the heart and all its parts. Heart disease is a major health problem. A recent statistics about heart diseases are reported in (Go, 2014). Some preventive mechanisms for heart diseases are discussed in (Shepherd, et al., 1995). A neural network-based heart disease prediction system is discussed in (Singh, Singh, & Pandi-Jain, 2018) and the authors used 15 medical parameters of the patient. A multilayer perceptron neural network with back propagation is used to develop a trained model. A KNN based approach for remote patient monitoring is discussed in (Enriko, Suryanegara, & Gunawan, 2018). A firefly algorithm based heart prediction technique is discussed in (Long, Meesad, & Unger, 2015).

In this manuscript, we used a set of machine learning approaches for heart disease prediction. Four well-known machine learning approaches are used in this study and those are discussed in this section.

1. **Naive Bayes Classifier:** Naive Bayes classifier is based on Bayes theorem. It makes strong independent assumptions between features, which means every pair of features is independent compared to every other feature (Rish, 2019). In this study, the Gaussian Naive Bayes classifier is used. If the training dataset contains null values for some specific parameter then it is better to use naive Bayes classifier. The major advantages of naive Bayes classifier are listed below:
 - Simple and fast
 - Possible to make probabilistic prediction
 - Strong independent attribute assumptions
2. **KNN Classifier:** The KNN stands for K-nearest neighbor and it can be used for both in classification and regression problems (Goldberger & Hinton, 2005). It is widely disposable in real-life scenarios since it is non-parametric, which means it does not make any underlying assumptions about the distribution of data. A given test data is classified by the majority vote of its neighbors. We can choose the value of K where K is the number of nearest neighbors of an object for which we wish to vote.
3. **Decision Tree Classifier:** The decision tree is in the form of a tree structure. It is a simple and widely used classification technique. It breaks the data into subparts, makes decisions, and at the same time decision tree is incrementally developed (Du & Zhan, 2002).

7 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

www.igi-global.com/chapter/heart-disease-prediction-using-machine-learning/263328

Related Content

Data Warehouse With OLAP Technology for the Tourism Industry

Preetvanti Singhand Vijai Dev (2023). *Encyclopedia of Data Science and Machine Learning* (pp. 191-211).

www.irma-international.org/chapter/data-warehouse-with-olap-technology-for-the-tourism-industry/317447

Power Consumption Prediction of IoT Application Protocols Based on Linear Regression

Sidna Jeddou, Amine Baina, Najid Abdallahand Hassan El Alami (2021). *International Journal of Artificial Intelligence and Machine Learning* (pp. 1-16).

www.irma-international.org/article/power-consumption-prediction-of-iot-application-protocols-based-on-linear-regression/287585

Industrial Revolution 4.0 With a Focus on Food-Energy-Water Sectors

Abichal Ghosh, Reddi Kamesh, Siddhartha Moulikand Anirban Roy (2023). *Encyclopedia of Data Science and Machine Learning* (pp. 2199-2210).

www.irma-international.org/chapter/industrial-revolution-40-with-a-focus-on-food-energy-water-sectors/317616

Autoencoder Based Anomaly Detection for SCADA Networks

Sajid Nazir, Shushma Patel and Dilip Patel (2021). *International Journal of Artificial Intelligence and Machine Learning* (pp. 83-99).

www.irma-international.org/article/autoencoder-based-anomaly-detection-for-scada-networks/277436

Machine Learning in 5G Multimedia Communications: Open Research Challenges and Applications

Dragorad A. Milovanovic, Zoran S. Bojkovic and Dragan D. Kukolj (2020). *Handbook of Research on Emerging Trends and Applications of Machine Learning* (pp. 361-382).

www.irma-international.org/chapter/machine-learning-in-5g-multimedia-communications/247572