

All tools x

- Export a PDF
- Edit a PDF
- Create a PDF
- Combine files
- Organize pages
- AI Assistant
- Generative summary
- Request e-signatures
- Scan & OCR
- Protect a PDF
- Redact a PDF
- Compress a PDF
- Prepare a form

[View more](#)

Convert, edit and e-sign PDF forms & agreements

[Free trial](#)

19 A comparative analysis of machine learning techniques for effective heart disease prediction

Ab Qayoom Sofi^{1,a} and Ram Kumar^{2,b}

¹Research Scholar, Department of Computer Applications, Lovely Professional University, Punjab, India
²Faculty, School of Computer Applications, Lovely Professional University, Punjab, India

Abstract

Cardiovascular diseases (CVD) are the biggest challenges and problems faced worldwide by the people in general and doctors/clinicians in particular. Heart disease is the leading cause of death in the world and the mortality rate has increased after COVID-19, affecting the individual productivity and national economies badly. Owing to variety of risk factors and types of the disease, it becomes very difficult for the clinicians to diagnose the type of heart disease quickly and accurately. So, there is a dire need that the mortality rate is brought down and the precious lives be saved by effectively predicting the disease early and on time. In this paper, we have compared the performance of different machine learning/deep learning techniques. The experimental results indicated that the K-nearest neighbor with K = 5 achieved the best performance especially in Bi-class mode followed by ensemble algorithm random forest.

Keywords: artificial neural networks, cardio-vascular disease, classification, deep learning, ensemble algorithms, heart disease, K-nearest neighbor, machine learning, random forest, support vector machine

Introduction

As per the WHO, cardiovascular disease (CVD) is spreading like an epidemic. Any condition affecting the heart indicated by the chest pain, abnormal pulse rate and other symptoms is a heart disease (HD). A huge loss is caused to the economy by the HD. Further, the individual suffering from the disease leads a poor-quality life and the overall productivity decreases significantly. So, the heart diseases need to be prevented and checked. If the disease is diagnosed early on time accurately, it can be controlled, and the patients can be given good treatment to prevent the huge losses and the precious lives can be saved. Hence, it is very important to predict the HD early, efficiently, and accurately. Since there are several different types of HD (Swathy and Saruladha, 2021), there is a need to employ the latest machine learning (ML) techniques for earliest identification of the disease with its type to increase the survival rate.

Systems already designed for the prediction of the HD using ML techniques have achieved very good accuracies. But a lot of research can be still done to find the improved techniques for use with a particular data type. More research can be done to select the most significant feature subset after comparison and analysis of several feature selection (FS) algorithms. This paper is an effort to compare some of the popular ML techniques employed for prediction of the heart disease.

Literature Review

Swain et al. (2018) analyzed various ML techniques and obtained best accuracy of 97.5% in artificial neural networks (ANN) technique with 13 features. Panicker (2020) reviewed varied ML algorithms successfully and noted support vector machine (SVM) as popular algorithm used for classification followed by neural networks and ensembles with high accuracy. Convolutional neural networks (CNNs) and deep learning (DL) techniques also have high accuracies of over 95%. Krittanawong et al. (2020) found that SVM and boosting algorithms have been commonly used but comparison of various ML algorithms with conventional classifiers is required for selection of the best algorithm. Mehanović et al. (2020) applied ANN, K-nearest neighbor (KNN), and SVM. They used evaluation methods - percentage split (66:34) and 10-fold cross validation in two ways- multiclass (5 outputs) and bi-class (2 outputs) with majority voting having highest accuracy (87.37%) in bi-class method and 61.16% in multi-class method with percentage split found to be better than 10-fold cross validation. In Bi-class algorithm using 10 fold cross validation, ANN performed better than the