

# Heart disease classification comparison among patients and normal subjects using machine learning and artificial neural network techniques

## Abstract

Machine Learning (ML) and Artificial Neural Networks (ANN) have been successfully used for classifications in many of the prediction models. These algorithms provide good accuracy results in many of the applications like fall detection.<sup>1</sup> In this work, we compared the performance of ML and ANN on the Heart Disease data available from Cleveland database.<sup>2</sup> The data has 76 attributes, but we have considered 13 best features by doing correlation and selecting the features that has best correlation index to train our ML and ANN. There are 303 persons data, and we trained our algorithms with 80% of training data and 30% for testing. SVM showed 84% accuracy with sensitivity of 78.5% and specificity of 87.8% whereas ANN gives an accuracy of 87% with sensitivity of 85% and specificity of 88.2%. Overall, both ML and ANN give good accuracy results to distinguish people from with and without heart disease.

**Keywords:** machine learning, artificial neural network, heart disease and classification

Volume 7 Issue 3 - 2021

Pavan Kota,<sup>1</sup> Aishwarya Madenahalli,<sup>2</sup> Rachana Guturi,<sup>3</sup> BT Nukala,<sup>4</sup> Sunil Nagaraj,<sup>4</sup> Santosh Kota,<sup>4</sup> Purna Chandra Neeli<sup>5</sup>

<sup>1</sup>University of North Carolina Greensboro, USA

<sup>2</sup>VF Corporation, USA

<sup>3</sup>IEEE Member, USA

<sup>4</sup>Qorvo Inc, USA

<sup>5</sup>TessolveDTS Inc, INDIA

**Correspondence:** BT Nukala, Qorvo Inc, USA, Tel +18064700275, Email bhargavanukala@qorvo.com

**Received:** May 04, 2021 | **Published:** May 12, 2021

## Introduction

Cardiovascular disease (HD) is among the most complex and life-deadliest diseases that affect human beings. Due to this, the heart can no longer pump enough blood to fulfill the entire body's functions without heart failure.<sup>3</sup> According to the American Heart Association, the United States has a very high heart disease rate.<sup>4</sup> A heart disease's symptoms are shortness of breath, weakness of body parts, swelling of feet, fatigue, and associated signs, caused by functional cardiac or non-cardiological anatomical abnormalities; for example, elevated jugular venous pressure and peripheral edema.<sup>5</sup> Identifying heart disease early on was challenging for investigators, and the results of complex testing impacted life-standards.<sup>6</sup> In developing countries, the diagnosis and treatment of heart disease are often complex, especially since diagnostic apparatus is often unavailable, as are doctors and other resources, resulting in less proper prediction and treatment of heart patients.<sup>7</sup> It is essential to reduce the potential risks associated with heart disease and improve heart security by accurately and properly diagnosing heart disease risk in patients.<sup>8</sup> In order to resolve these complexities in invasive-based diagnosing of heart disease, a noninvasive medical decision support system based on machine learning predictive models such as support vector machine (SVM) and artificial neural network (ANN)<sup>9,10</sup> has been developed by various researchers and widely used for heart disease diagnosis, and due to this machine-learning-based expert medical decision system, the ratio of heart disease death decreased.<sup>11</sup>

Many algorithms have been used for prediction models and out of which, ML and ANNs have the best training models to successfully distinguish between the actual and predicted class. As said before, the work in this paper mainly focuses on the performance comparison of ML and ANN on the available Heart disease dataset from Cleveland database. This data set has 303 persons data and has different attribute information such as age, sex, chest pain type, blood pressure, cholesterol in mg/dl, blood sugar, maximum heart rate etc.

## Materials and methods

The aim of the proposed classification solution is to design a machine-learning-based medical intelligent decision support system

for the diagnosis of heart disease. In the present study, various machines learning predictive models such as ANN, SVM, have been used for classification of people with heart disease and healthy people. We have used 13 attributes which serve as the inputs to our training algorithms out of which the 13<sup>th</sup> attribute is a target value of either 1 (presence of heart disease) or 0 (absence of heart disease). We have used a custom feature which is Cholesterol/age as an additional feature in our training set. Data correlation has been performed among all the 13 attributes to see how strong each attribute has with each other attribute. The methodology of the proposed system structured into five stages including (1) feature selection, (2) cross-validation method, (3) machine learning classifiers, and (5) classifiers' performance evaluation methods. The data correlation among the training attributes has been shown in the Figure 1. Both SVM and ANN were performed in Python 7.1.

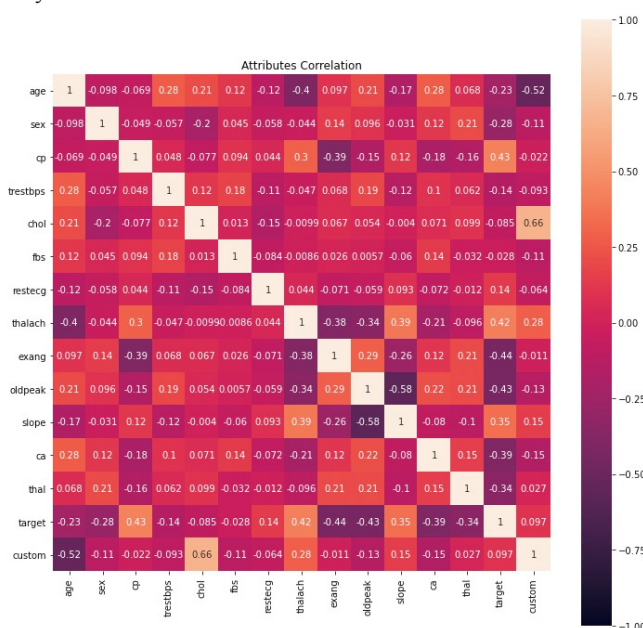


Figure 1 Correlation matrix of the input attributes set.

### Support vector machine

In our SVM, we used a linear kernel function for classification and used Training-Testing (80%-20%) on the dataset. Support Vector Machine (SVM) is a method for patterns recognition/classification on two categories with supervised learning. SVM-light, one of the implementations of SVM proposed by Thorsten Joachims<sup>12,13</sup> is applied to the classification of patients of heart disease with normal persons. In Linear SVM, a hyper-plane or a set of hyper-planes can be used as the separate lines in classification. The higher the margin of separation for the classes that can be created, the better the classification result that can generally be achieved for the Linear SVM.

### Artificial neural network

For training the feed-forward ANN classifier, back propagation was applied according to Duda et al.<sup>14</sup> and a 3-layer system was picked as the standard BP ANN.<sup>15</sup> The input layer of the network has twelve neurons which correspond to the twelve input feature values. There is one hidden layer holding 8 hidden neurons, which number was optimized by adjusting the size of hidden neurons, and two output neurons corresponding to the two target classes the network needs to differentiate.

### Results and discussion

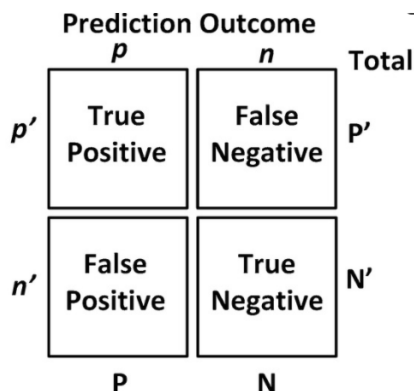
After performing the training of both ML and ANNs, the algorithms have been compared with performance index such as accuracy, sensitivity, and specificity. The results are shown in Table 1. ANN shows better results compared to SVM. In both the classifiers, the training and testing is 80-20% and therefore there are a total of 61 test samples for algorithms to perform classification. The indexes have been calculated using confusion matrix shown in Figure 2. For each SVM and ANN algorithms. The confusion matrix for SVM and ANN are shown in Figure 3 & Figure 4. The equations for Accuracy, Sensitivity and specificity have been calculated using the below equations.

$$Accuracy = \frac{TP+TN}{Total} \text{ Equation 1}$$

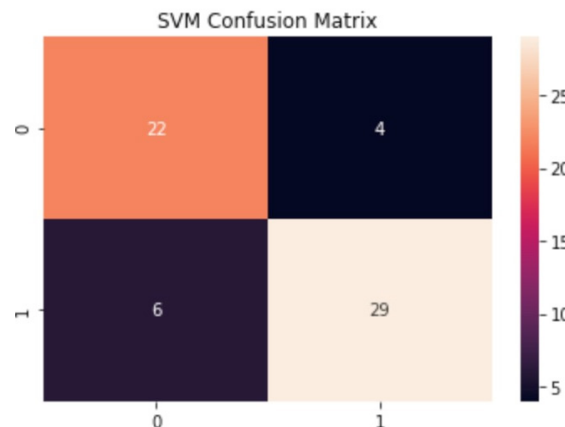
$$Sensitivity = \frac{TP}{TP+FN} \text{ Equation 2}$$

**Table 1** Performance comparison of SVM and ANN

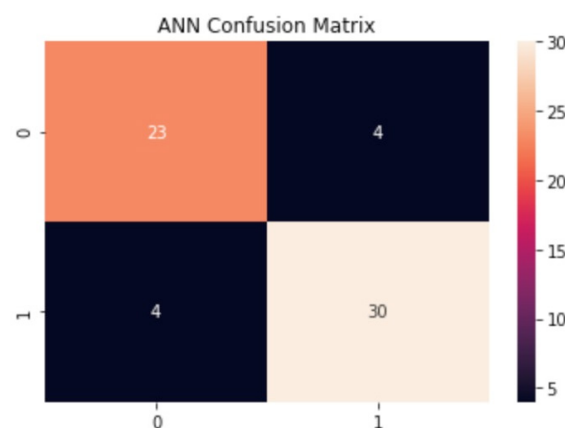
Performance index	SVM	ANN
Accuracy	83.60655738	86.8852459
Sensitivity	78.57142857	85.18518519
Specificity	87.87878788	88.23529412



**Figure 2** Definition of confusion matrix.



**Figure 3** Confusion matrix of SVM classifier.



**Figure 4** Confusion matrix of ANN classifier.

**Sensitivity (recall):** This measures the actual members of the class that are correctly identified as such. It is also referred to as the true positive rate (TPR). It is defined as the fraction of positive examples predicted correctly by the classification model. Classifiers with large sensitivity have very few positive examples misclassified as the negative class.

**Specificity:** This is also known as the true negative rate. It is defined as the fraction of total negative examples that are predicted correctly by the model/classifier.

$$Specificity = \frac{TN}{TN+FP} \text{ Equation 3}$$

### Conclusion

In this study, a hybrid intelligent machine-learning-based predictive system was proposed for the diagnosis of heart disease. The system was tested on Cleveland heart disease dataset. Two well-known classifiers SVM and ANN were studied with feature selection. Based on our analysis, using performance indices ANN shows accuracy of 88% compared to SVM accuracy of 84%. Feed forward Neural Network with three-layer network has better sensitivity and specificity than linear SVM. Both sensitivity and specificity of ANN has more than 85% accuracy rate. Research work on this topic focuses on the development of a diagnosis system. It is based on two classifiers, one cross-validation method, and performance measurement metrics. The system was tested on Cleveland heart disease dataset to classify HD and healthy subjects. Designing a decision support system through machine-learning-based method will be more suitable for

diagnosis of heart disease. Some irrelevant features slowed down the diagnosis system and made computation time longer. In the future, we will perform more experiments to increase the performance of these predictive classifiers for heart disease diagnosis by using others feature selection algorithms and optimization techniques.

## Acknowledgments

None.

## Conflicts of interest

Authors declare that there is no conflict of interest.

## References

1. Bhargava Teja Nukala, Naohiro Shibuya, Amanda Rodriguez, et al. An Efficient and Robust Fall Detection System using Wireless Gait Analysis Sensor with Artificial Neural Network (ANN) and Support Vector Machine (SVM) Algorithms. *Open Journal of Applied Biosensor*. 2014;3:29–39.
2. <https://archive.ics.uci.edu/ml/datasets/Heart+Disease>
3. Bui AL, Horwich TB, Fonarow GC. Epidemiology and risk profile of heart failure. *Nature Reviews Cardiology*. 2011;8(1):30–41.
4. Heidenreich PA, Trogdon JG, Khavjou OA, et al. Forecasting the future of cardiovascular disease in the United States: a policy statement from the American Heart Association. *Circulation*. 2011;123(8):933–944.
5. Durairaj M, Ramasamy N. A comparison of the perceptive approaches for preprocessing the data set for predicting fertility success rate. *International Journal of Control Theory and Applications*. 2016;9:256–260.
6. Mourão-Miranda J, Bokde ALW, Born C, et al. Classifying brain states and determining the discriminating activation patterns: support vector machine on functional MRI data. *NeuroImage*. 2005;28(4):980–995.
7. Ghwanmeh S, Mohammad A, Al-Ibrahim A. Innovative artificial neural networks-based decision support system for heart diseases diagnosis. *Journal of Intelligent Learning Systems and Applications*. 2013;5(3):176–183.
8. Al-Shayea QK. Artificial neural networks in medical diagnosis. *International Journal of Computer Science*. 2011;8(2):150–154.
9. Nazir S, Shahzad S, Mahfooz, et al. Fuzzy logic based decision support system for component security evaluation. *International Arab Journal of Information Technology*. 2015;15:1–9.
10. Nazir S, Shahzad S, Septem Riza L. Birthmark-based software classification using rough sets. *Arabian Journal for Science and Engineering*. 2017;42(2):859–871.
11. Methaila A, Kansal P, Arya H, et al. Early heart disease prediction using data mining techniques. In *Proceedings of Computer Science & Information Technology (CCSIT-2014)*. 2014;24:53–59.
12. Joachims T. SVM Light Support Vector Machine. 2008.
13. Joachims T. Making Large-Scale Support Vector Machine Learning Practical. In: Schölkopf B, Burges CJC, editors. *Advances in Kernel Methods—Support Vector Learning*. Cambridge. MIT Press; 1999:169–184.
14. Duda RO, Hart PE, Stork DG. *Pattern Classification*. 2nd edn. New York. Wiley-Interscience; 2001.
15. Dreiseitl S, Ohno-Machado L. Logistic Regression and Artificial Neural Network Classification Models: A Methodology Review. *Journal of Biomedical Informatics*. 2002;35:352–359.