

PAPER • OPEN ACCESS

## Cardiovascular disease prediction using deep learning techniques

To cite this article: Syed Nawaz Pasha *et al* 2020 *IOP Conf. Ser.: Mater. Sci. Eng.* **981** 022006

View the [article online](#) for updates and enhancements.



The Electrochemical Society  
Advancing solid state & electrochemical science & technology

**240th ECS Meeting** ORLANDO, FL

Orange County Convention Center Oct 10-14, 2021

Abstract submission due: April 9

SUBMIT NOW

# Cardiovascular disease prediction using deep learning techniques

Syed Nawaz Pasha<sup>1</sup>, Dadi Ramesh<sup>2</sup>, Sallauddin Mohammad<sup>3</sup>, A. Harshavardhan<sup>4</sup> and Shabana<sup>5</sup>

<sup>1</sup>Department of CSE & Member of Collaboratory for Social Innovation, S R Engineering College, Warangal, Telangana, India.

<sup>2</sup>Department of CSE, S R Engineering College, Warangal, Telangana, India.

<sup>3</sup>Department of CSE & Member in Center for Embedded System & IOT, S R Engineering College, Warangal, Telangana, India.

<sup>4</sup>Department of CSE, S R Engineering College, Warangal, Telangana, India.

<sup>5</sup>Department of CSE, Sumathi Reddy Institute of Technology for Women, Ananthasagar, Hasanparthy, Warangal, Telangana, India

E-mail: Snp786@gmail.com

**Abstract:** Cardiovascular Disease or coronary illness is one of the significant dangerous infections in India as well as in the entire world. It is estimated that 28.1 % of deaths occur due to heart diseases. It is also the major cause for significant number of deaths which as more than 17.6 million in the year 2016. So proper and timely diagnosis, treatment of such diseases require a system that can predict with precise accuracy and reliability. Intensive research is carried out by various researchers using diverse machine learning algorithms to forecast the heart disease taking different datasets which consists of different attributes that result in heart attack. In this paper we analyzed the dataset collected from kaggle which consists of attributes related to heart disease such as age, gender, blood pressure, cholesterol and so on. We have also investigated the accuracy levels of various machine learning techniques such as Support Vector Machines (SVM), K-Nearest Neighbor (KNN), Decision Trees (DT). The performance and accuracy of above algorithms is not so well when executed using large dataset, so here we tried to improving the prediction accuracy using Artificial Neural Network (ANN), Tensor Flow Keras.

**Keywords:** Cardiovascular Disease, Deep Learning, Support Vector Machines, K-Nearest Neighbor, Decision Tree (DT), Artificial Neural Network (ANN).

## 1. Introduction

The most important part of human body is the heart which pumps blood to every other part. In the event that the heart doesn't work appropriately, at that point the mind and different organs will quit working which prompts the demise of the individual inside couple of moments. So proper functioning of the heart is of vital importance. Heart diseases have developed as one of the most indispensable reason for death all around the globe. As indicated by World Health Organization, heart related diseases are answerable for the taking 17.7 million lives consistently, 31% of every single worldwide demise [1]. In India as well, heart related sicknesses have become the main source of mortality. Thus it

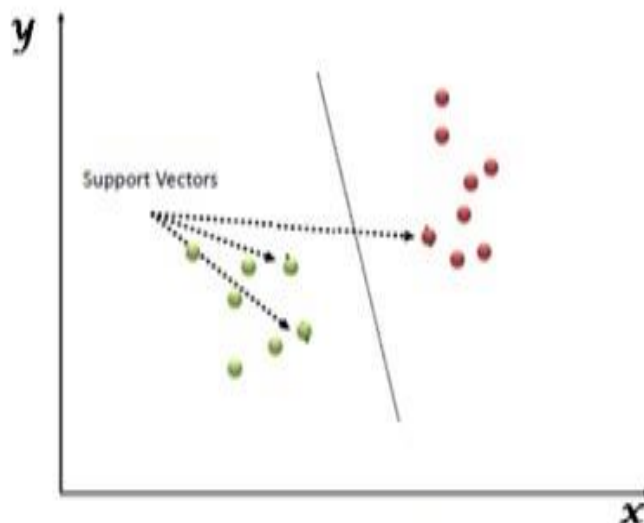


has become important for precise and accurate prediction of heart related diseases. For this many researchers all around the globe started working on predicting heart related diseases by analyzing the massive datasets. Various machine learning techniques have the potential to work on massive datasets and draw useful conclusions. The machine learning models run on various algorithms and hence these algorithms have become important in predicting the existence or non existence of heart diseases accurately.

## 2. Related work

### 2.1 Support vector machine

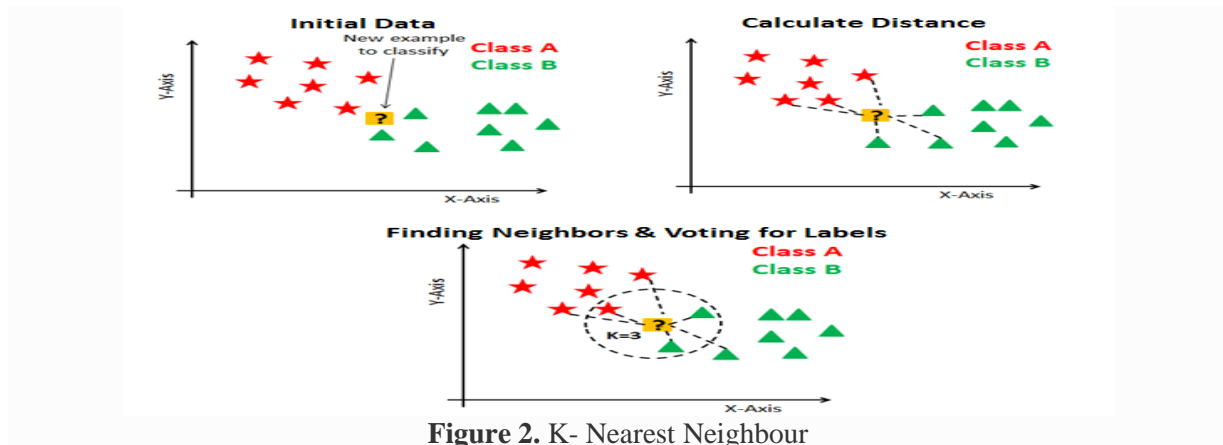
It is a supervised machine learning techniques which attempts to find the optimal hyper-plane by learning through training data of target labeled tuples. The objective is to discover the hyper-plane that see various classes. There can be multiple hyper-planes that can do this task but the objective is to find the hyper-plane with highest margin that maximizes distance between classes. The new data point that need to be classified can be classified easily using the above hyper-plane [2].



**Figure 1.** Support Vector Machine

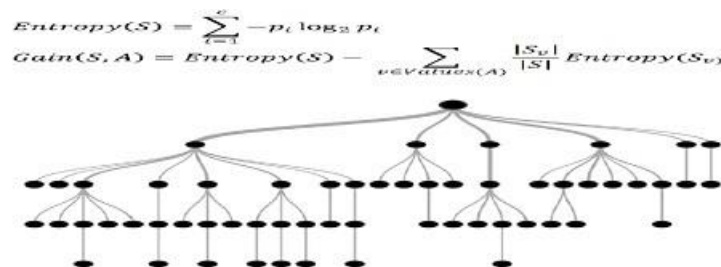
### 2.2 *K* – nearest neighbour

Fix and Hodge proposed the KNN model which is one of the simple and most widely used supervised classification algorithm. It works by calculating distance between the test and training datapoints in order to identify its nearest neighbours. The new sample is then assigned to the class of its closest neighbor. In KNN, *K* represents the number of closest neighbors [3]. The classification is greatly influenced by this value of '*K*'. KNN works with numeric data which uses various measures such as Euclidean, Manhattan, Minkowsky [4].



### 2.3 Decision tree

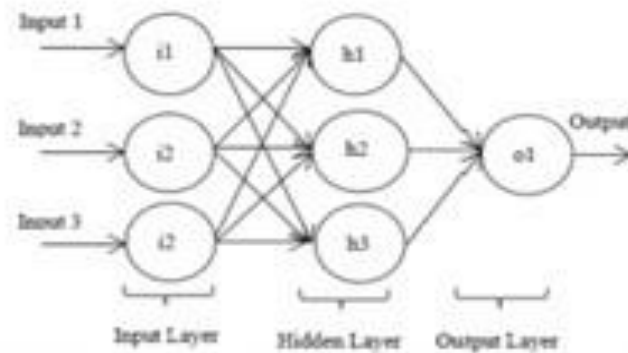
Decision Tree is one of the supervised learning technique which is capable of handling both continuous and categorical attributes. It initiates the classification by dividing the dataset into two or more similar sets according to the most significant predictors [5][6]. It then calculates the entropy of each and every attribute. The attribute that has maximum information gain or minimum entropy is selected as splitting attribute. The same process continues recursively on remaining attributes.



**Figure 3. Decision Tree**

### 2.4 Artificial neural network

An Artificial neural Network(ANN), also referred to as "neural Network" (NN).It is a multi-level, that performs computations and numerical models by memic of neurons found in human anatomy. It functions as a single neuron of the human brain [7]. It has various levels and each level has various perceptrons.The output depends on the weight attached to each perceptron. It also has various input layers that performs computations and reassigning of weights and an output layer that generates the output [8].



**Figure 4.** ANN with Input Layer

### 3. Our approach

#### Algorithm Steps:

##### Step1: Splitting the given dataset into training and test

Training set size and test set size is configured as 0.8 and 0.2.

##### Step 2:Model Creation:

In the model input layer consists of 13 nodes and activation function as 'relu'

Hidden layer with 4 nodes and with activation function as 'relu'.

Output layer with 1 node with activation function as 'sigmoid'.

Learning Rate=0.001

Loss=binary\_crossentropy and optimizer='adam'.

##### Step 3: Training

Training is done for 100 epochs and batch size=10

##### Step 4: Result

The result for binary model is obtained is below.

### 4. Results for Binary Model

**Accuracy: 0.8524590163934426**

**Table 1.** Predicted Parameters For ANN Model

	<b>Precision</b>	<b>Recall</b>	<b>F1-Score</b>	<b>Support</b>
<b>0</b>	0.84	0.91	0.88	35
<b>1</b>	0.87	0.77	0.82	26
<b>Accuracy</b>			0.85	61
<b>Micro Avg</b>	0.86	0.84	0.85	61
<b>Weighted Avg</b>	0.85	0.85	0.85	61

**Precision** – Precision is the ratio of the positive observations correctly predicted to the total positive observations predicted [9][12]. Using the above algorithm we have got a precision rate of 0.85 which is quite good.

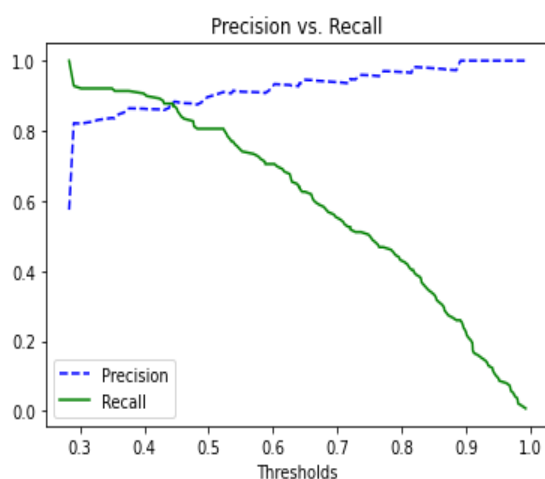
$$\text{Precision} = \text{TP} / (\text{TP} + \text{FP})$$

**Recall** (Sensitivity) - is the proportion of effectively anticipated positive perceptions to all observations in actual class - yes. Using the above algorithm the recall value is predicted as 0.85 which is quite good and is par beyond average [10].

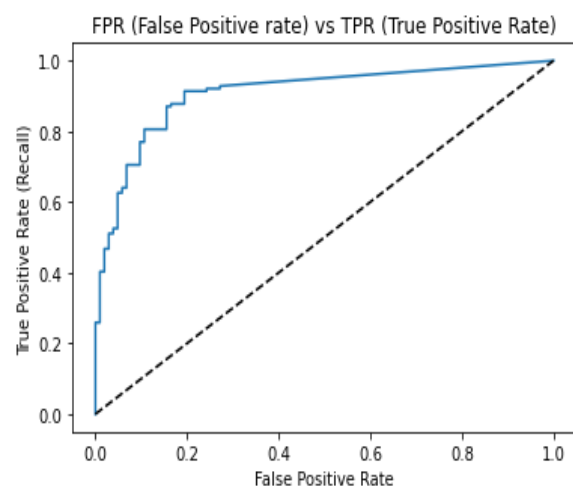
$$\text{Recall} = \text{TP} / (\text{TP} + \text{FN})$$

**F1 score** – It is also called F score or F Measure which is the balance between precision and recall. It is particularly useful when falsepositive and falsenegative are crucial [11]. Using the above algorithm the predicted F1 score is 0.85.

$$\text{F1 Score} = 2 * (\text{Recall} * \text{Precision}) / (\text{Recall} + \text{Precision})$$



**Figure 5.** Precision Vs. Recall for our Model



**Figure 6.** FPR vs TPR for our Model

## 5. Comparison with other algorithms

**Table 2.** Accuracy Comparison

Algorithm	Accuracy
SVM	81.97
KNN	67.2
Decision Tree	81.97
ANN(binary model)	85.24

## 6. Conclusion

In this paper we have analyzed different algorithms like Support Vector Machines (SVM), K-Nearest Neighbor (KNN) [13][14], Decision Trees (DT), Artificial Neural Network(ANN),TensorFlow Keras by comparing their accuracy levels on heart attack dataset collected from Kaggle.We also summarized the various prediction parameters obtained using each of the above mentioned algorithms[15].

## References

- [1] Shan Xu ,Tiangang Zhu, Zhen Zang, Daoxian Wang, Junfeng Hu and Xiaohui Duan, 2017 Cardiovascular Risk Prediction Method Based on CFS Subset Evaluation and Random Forest Classification Framework Beijing China IEEE *2nd International Conference on Big Data Analysis* ,228-232.
- [2] Seyedamin Pouriyeh, Sara Vahid, Giovanna Sannino, Giuseppe De Pietro, Hamid Arabnia and Juan Gutierrez 2017 “ A Comprehensive Investigation and Comparison of Machine Learning Techniques in the Domain of Heart Disease”, Heraklion, Greece ,*IEEE 2nd IEEE Symposium on Computers and Communication*.
- [3] Swain, Debabrata, Santosh Pani, and Debabala Swain. 2019, April. An efficient system for the prediction of Coronary artery disease using dense neural network with hyper parameter tuning. *International Journal of Innovative Technology and Exploring Engineering (IJITEE)* **8 (6S)**. 689-695.
- [4] Ghumbre, S.U. and A.A. Ghatol. 2012. Heart disease diagnosis using machine learning algorithm. In *Proceedings of the International Conference on Information Systems Design and Intelligent Applications 2012 (INDIA 2012)* Visakhapatnam India, Springer, **132**,217-225.
- [5] Fahd Saleh Alotaibi 2019. Implementation of Machine Learning Model to Predict Heart Failure Disease, *International Journal of Advanced Computer Science and Applications*, **10(6)**.
- [6] Chen, H., S Y Huang, P S Hong, C H Cheng and E.J. Lin. 2011. HDPS-Heart disease prediction system. *Computing in Cardiology*, Hangzhou, 557–560.
- [7] Miranda E, Irwansyah E, Amelga AY, Kom S, Maribondang MM, Kom S, Salim M and Kom S.2016. Detection of cardiovascular disease risk's level for adults using Naive Bayes classifier. *Healthc Inf Res.* **22(3)** 196–205.
- [8] Bharti, S. 2015. Analytical study of heart disease prediction comparing with different algorithms. *International Conference on Computing, Communication and Automation (ICCA2015)*.Noida,78-82.
- [9] Mohammed Ali Shaik 2019 A Survey on Text Classification methods through Machine Learning Methods. *International Journal of Control and Automation* **12(6)** 390-396.
- [10] Ravi Kumar R, Babu Reddy M and Praveen, 2019 An evaluation of feature selection algorithms in machine learning *international journal of scientific & technology research* **8(12)** 2071-2074.
- [11] M Sheshikala, D Kothandaraman , R Vijaya Prakash and G Roopa,2019. Natural language processing and machine learning classifier used for detecting the author of the sentence *International Journal of Recent Technology and Engineering* **8(3)**,936-939.
- [12] R Ravi Kumar,M Babu Reddy and P Praveen, 2019 Text Classification Performance Analysis on Machine Learning *International Journal of Advanced Science and Technology*, **28(20)**,691-697.
- [13] S Naresh Kumar, P Pramod Kumar, Ch Sandeep, V Thirupthi and S Swetha, 2019. A study on deep q-learning and single stream q-network architecture *International Journal of Advanced Science and Technology*, **28(20)**,586-592.
- [14] Naresh kumar sripada,Shwetha Sirikonda,Nampally Vijay Kumar and Vahini Siruvoru 2019 Support Vector Machines To Identify Information Towards Fixed-dimensional Vector Space *International Journal Of Innovative Technology And Exploring Engineering(IJITEE)* **8(10)**,4452-4455.