# Heart Disease Detection Using Artificial Neural Network


Şilan Karadağ
Department of Electronics and Communication Engineering
Istanbul Technical University
Istanbul, Turkey
karadags15@itu.edu.tr

Merve Duru
Department of Electronics and Communication Engineering
Istanbul Technical University
Istanbul, Turkey
duru15@itu.edu.tr

Şevval Yıldız
Department of Electronics and Communication Engineering
Istanbul Technical University
Istanbul, Turkey
yildizs15@itu.edu.tr



***Abstract*— Cardiovascular disease is one of the most leading causes of death in which is considered an important disease. Many people have lost their lives due to heart diseases throughout history. Numerous software packages and various algorithms are proposed to develop robust medical decision support systems with the researchers' aid. This article is investigating the application of the ANN method for the detection of heart disease. The dataset used in this article has 13 input for training the neural network. The accuracy obtained is 85% and F1-score of 86.73%.**



***Keywords*— heart disease, disease diagnosis, machine learning, artificial neural network, multilayer perceptron, Cleveland database, supervised learning**


## I. Introduction

As human life itself depends on the heart's operate effectively, if it does not operate properly, it will have an effect on various elements of the human body. Cardiovascular disease (CVD) results in severe illness, disability, and death. Heart disease continues the world's primary reason for death for hundreds of years. 1/3 deaths of the planet's deaths are caused by coronary illness; therefore the death rate is on top of cancer mortality rates [1]. Heart disease describes a range of conditions that affect individuals' hearts. The term "cardiovascular disease" includes a wide range of conditions that affect the heart and the blood vessels and the manner in which blood is pumped and circulated through the body. By improving cardiovascular health, quality of life is improved through prevention, detection, and treatment of risk factors for heart attack can be identified earlier, and deaths from cardiovascular disease can be reduced. This is one in each explanation why researchers focus on data mining which is accustomed to diagnosing heart diseases with high accuracy to avoid misdiagnosis. It reduces the cost of treatment compared to traditional ways. In this research, the study of previous works is studied, the most appropriate prediction model is proposed using artificial neural network.

## II. Dataset

### A. Dataset Description

Dataset is one of the factors to increase success for training an artificial neural network. Using a reliable data set should be ensured that training of the artificial neural network we have established is carried out more successfully.

In line with this information, the "Heart Disease Dataset" was chosen to use in solving the problem. This dataset was created with data from the following four institutions:

1. Cleveland Clinic Foundation
2. Hungarian Institute of Cardiology, Budapest
3. V.A. Medical Center, Long Beach, CA
4. University Hospital, Zurich, Switzerland

However, ML researchers have only used the Cleveland database in previous studies. The data collected from these institutions through selected doctors were compiled in 1988 and turned into a data set. This data set has been used in many different studies. These studies are described under the "Previous Study" section.

The data set contains 76 attributes collected from 303 people. While most of these are detailed health data, 14 of them have been specifically selected for their impact on heart disease diagnosis. Like most studies, these 14 attributes were used for our work. The features used are briefly described below.

**Target:** In this dataset, 0 refers to the existence, and 1 refers to the absence of heart disease.

**Age:** The risk of coronary heart disease increases in men over 45 and women over 55. For this reason, age information is an essential factor in the diagnosis of heart disease. The distribution of age vs. sex with the target class is shown in Figure 1.

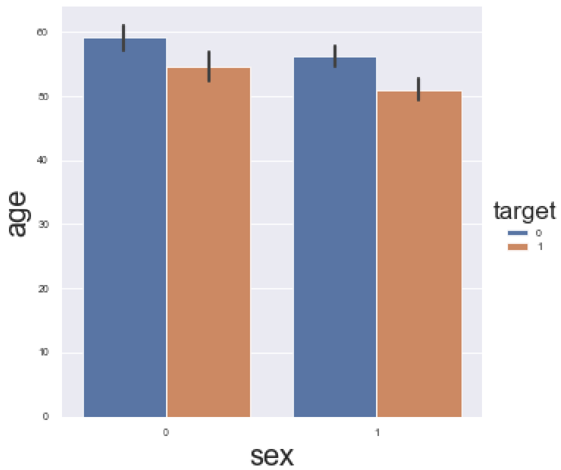


**Figure 1**. Distribution of age vs. sex with the target class.

**Sex:** Since the limit values of health information change according to gender information, gender information is necessary to ensure these values' correct interpretation. The sex value was entered as 1 for male and 0 for female.

**Chest Pain Type:** Chest pain facilitates the diagnosis of heart diseases depending on the type. For this data set, the type of chest pain was divided into four, and the individuals' information was taken as 1-4.

Value 1: typical angina
Value 2: atypical angina
Value 3: non-anginal pain
Value 4: asymptomatic

Chest pain, defined as angina, occurs as a result of insufficient blood and oxygen delivery to the heart. There are three criteria to look for classification of angina as typical or atypical.

- A pressure not exceeding 10 minutes, pain in the form of compression, under the breastbone.
- The emergence of pain with exertion or stress
- Relaxation with rest

If all three of these symptoms are shown, chest pain type is defined as typical angina. However, if there are only two, it is defined as atypical angina. If there is a pain in the chest, but the heart does not cause this pain, this type of pain is called non-anginal pain. The reasons may be things like reflux, rib fracture. If there is no pain, people are classified as asymptomatic.

**Resting Blood Pressure:** High blood pressure, also known as hypertension, is defined as a risk factor for heart disease if it is not controlled. The interval for resting blood pressure is between 94 and 200 mm Hg in the dataset.

**Serum Cholesterol:** Calculated using the LDL, HDL and triglyceride values resulting from the measurements. LDL is called bad cholesterol, HDL is good cholesterol. The interval for serum cholesterol is between 126 and 564 mg/dl. The calculation is:

Serum Cholesterol = HDL + LDL + (Triglyceride / 5)

**Fasting Blood Sugar:** The fasting blood sugar value was entered as 1 or 0 according to whether this value is less than or more than 120(in mg/dL). While 1 is entered for values greater than 120, this data takes 0 for small values.

**Resting Electrocardiographic Results:** The results were analyzed by reviewing the electrocardiographic results and divided into three headings.

Value 0: normal
Value 1: having ST-T wave abnormality (T wave inversions or ST elevation or depression of $> 0.05$ mV)
Value 2: showing probable or definite left ventricular hypertrophy by Estes' criteria

If the results are as they should be, the typical result is recorded with 0. As a result, if there are ST elevation or depression, T wave inversion problems, the value is recorded as 1. Value 2 is showing LVH. Left ventricular hypertrophy, or LVH, is a term for a heart's left pumping chamber that has thickened and may not be pumping efficiently.

**Maximum Heart Rate:** The maximum heart rate should not exceed a constant value that is different for all ages. Unexpected values of the maximum heart rate may be a sign of heart disease. The interval for maximum heart rate is between 71 and 202.

**Exercise-Induced Angina:** It is the condition that occurs in the middle of the chest, under the breastbone, by exercise triggering. The presence or absence states are entered as 1 (pain occurs) or 0 (no pain occurs).

**Oldpeak:** ST depression is generally induced by exercise relative to rest. The increase in St depression is discussed under "Resting Electrocardiographic Results". The interval for resting blood pressure is between 0 and 6.2.

**Slope:** The slope of the peak exercise ST segment. There are three values for the slope:

Value 1: upsloping
Value 2: flat
Value 3: downsloping

**Number of major vessels:** In fluoroscopy, using X-ray, structures that not visible in the standard film are stained with medicine and made visible. The value to be entered as information is between 0 and 3 which describes how many vessels are colored by fluoroscopy.

**Thal:** It is input from the result of the nuclear stress test. The camera is used to monitor the tracer by injecting a tracer into the body. If coronary stenosis is detected when a myocardial segment takes up nuclear tracer at rest, it is defined as "reversible defect". Scarred myocardium from prior infarct will not take up tracer at all and is referred to as a "fixed defect". In this dataset, 1 refers to normal; 2 refers to fixed defect; 3 refers to "reversible defect".

The dataset does not include missing values. Since all values are numeric, encoding is not necessary.

TABLE I. LIST OF FEATURES

| ***Name*** | ***Shortening*** | ***Value*** |
|---|---|---|
| Age | Age | [29, 77] |
| Sex | Sex | 1 = male,<br>0 = female |
| Chest Pain Type | Cp | [1, 4] |
| Resting Blood Pressure | Trestbps | [94, 200] |
| Serum Cholesteral | Chol | [126,564] |
| Fasting Blood Sugar | Fbs | 1 = true,<br>0 = false |
| Resting ECG Results | Restecg | [0, 2] |
| Maximum Heart Rate | Thalach | [71,202] |
| Exercise Induced Angina | Exang | 1 = yes,<br>0 = no |
| Oldpeak | Oldpeak | [0, 6.2] |
| Slope | Slope | [1, 3] |
| Number of Major Vessels | Ca | [0, 3] |
| Thal | Thal | [1, 3] |
| Target | Target | 0 = existence<br>1= absence |

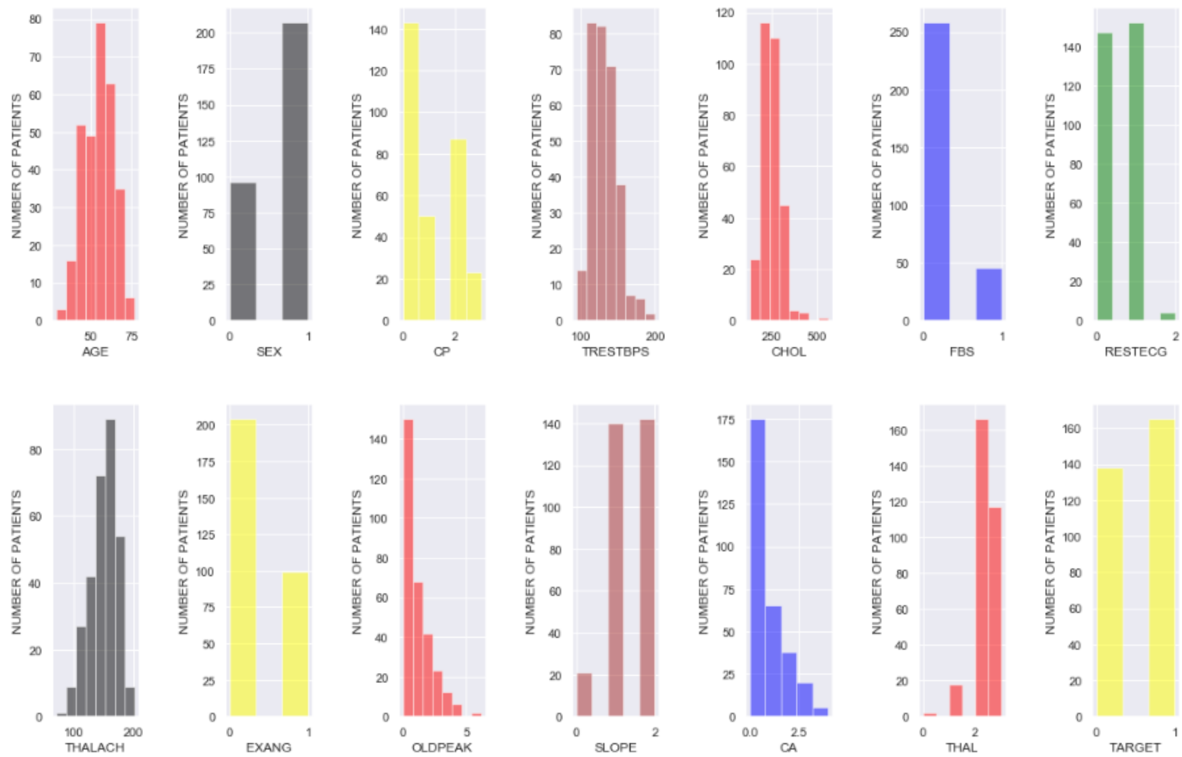


**Figure 2**. Distribution of features vs. the number of patients.

For better understanding, the list of features is shown in Table 1. The distribution of features according to the number of patients can be seen in Figure 2.

As the dataset does not contain too many features, feature selection is not necessary. However, understanding the dataset helps us to read results. Since understanding the relation between features is not easy without having a medical background, it can be obtained using a correlation matrix given in Figure 3.

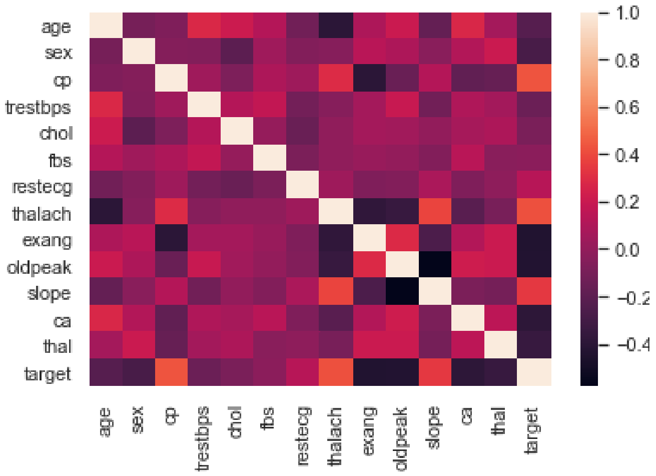


**Figure 3.** Correlation matrix

## B. *Data Preprocessing*

*a) Train-Test Split:* Before creating the model, the data should be divided into two parts as training and testing. The purpose of this application is to measure the success of the model trained with the training data set with the test data set. When using functions that separate data in the Scikit learn library, the point to be considered is that the distinction is made randomly. Since the problem we are dealing with is not a time series problem, it is aimed to have better classification with the randomly separated dataset. In the dataset, data and target values are specified, and target values are reshaped to 1D array. Then, data is splitted up to form the train and test sets. The test size and random state values are set to 0.33 and 0.42, respectively.

*b) Standardization:* After these steps, the "StandardScaler" method of the "sklearn.preprocessing" package was used to normalize the data distribution, to avoid the results, to make accurate predictions and eliminate the scale differences between the features as well. StandartScaler is a method in which the mean value takes 0 and the standard deviation takes 1, and the distribution is close to normal. This method performs the operations of subtracting the average value from each data and dividing the result by the variance value. As a result of the scaling process, it is aimed to increase the model performance.

Standardization formula is (1):

$$z = \frac{x-\mu}{\sigma} \tag{1}$$

The equation for the mean is (2):

$$\mu = \frac{1}{N}\sum_{i=1}^{N}(x_i) \tag{2}$$

Equation (3) refers to the standard deviation:

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i-\mu)^2} \tag{3}$$

## III. Previous Study

Tremendous works in the literature related to heart disease diagnosis using data mining techniques have motivated our work. Since its creation, this database has been used by many researchers investigating different classification problems with various classification algorithms. The researchers in the medical field diagnose and predict the diseases in addition to providing effective care for patients by employing data mining techniques. The data mining techniques have been employed by numerous works in the literature to diagnose diverse diseases, for instance: diabetes, hepatitis, cancer, heart diseases and more [2].

A model Intelligent Heart Disease Prediction System (IHDPS) built with the aid of data mining techniques (Logistic Regression (LR), Support Vector Machine (SVM), Deep Neural Network (DNN), Decision Tree (DT), Naïve Bayes (NB), Random Forest (RF), and K-Nearest Neighbor (K-NN), Extreme Learnıng Ann (ELANN), Multilayer Perceptron Ann (MLP), Recurrent Fuzzy Neural Networks (RFNN), Backpropagation Neural Network (BPNN), Vector Quantization Algorithm (VQA), Genetic Algorithm (GA)) was proposed by a number of researchers.

Some of the works that are related to this area are described as follows briefly. The summary of the previous studies can be seen in Table 2.

Robert Detrano et al. proposed a probability algorithm for the diagnosis of coronary artery disease. Robert Detrano, who constructed the Cleveland heart disease database, used logistic regression algorithm and obtained 77.0% classification accuracy [3].

Gudadhe et al. realized an architecture base with both the MLP network and the SVM approach. This architecture achieved an accuracy of 80.41% in terms of the classification between two classes (the presence or absence of heart disease, respectively). They trained the multilayer perceptron neural network by the backpropagation algorithm, which is a computationally efficient method, and obtained that MLP with backpropagation can be successfully used for diagnosing heart disease than support vector machine [4].

SY Huang, AH Chen, CH Cheng, PS Hong and EJ Lin trained the classification and prediction via learning Vector Quantization Algorithm. There were three steps in their methodology. The first one was to select 13 clinical features which are important compared to others. The second one was using the Artificial Neural Network algorithm for classification. Lastly, the heart disease prediction system was developed. The accuracy of the prediction rate obtained from the study is almost 80% [5].

Syed Umar Amin, Dr. Rizwan Beg and Kavita Agarwal have proposed a hybrid system using Genetic Algorithms and Artificial Neural Networks. The neural network was trained with backpropagation, which points out two major disadvantages. The first problem is that finding out initial weights that are globally optimized is almost impossible. The second problem is the slowness of the backpropagation algorithm in convergence. The problems were solved by using the Genetic Algorithm for optimizing connection weights. The neural network used in the study had 12 input, 10 hidden and 2 output nodes. The results obtained validation accuracy is 89% [6].

Saba Bashir, M.Younus Javed and Usman Qamar have proposed a hybrid method using Decision Tree, Support Vector Machine and Naive Bayes. The majority voting scheme was obtained by these three classifiers. There were two steps in the proposed approach. The first one was producing every three classifiers' decision. The second one was combining the decisions in order to acquire a new model based on the majority voting scheme. 82% accuracy was obtained from the study to predict heart disease [7].

Jayshril S. Sonawane and D. R. Patil studied Vector Quantization Algorithm to train network by using random order incremental training. There were three layers in the network, including input, hidden and output layers. There were 13 neurons in the input layer and only one single neuron in the output layer that denotes if heart disease present or not. The system performance was improved by training with varying the number of neurons and variable epochs. The result shows that they obtained the highest accuracy of 85.55% [8].

Jayshril S. Sonawane and D. R. Patil used multilayer perceptron neural network in the system. The proposed system had two steps. The first one was the process of accepting 13 clinical data as input and as a last step training the network by the backpropagation algorithm. There were three layers in the network, including input, hidden and output layers. There were 13 neurons, which is equal to the number of clinical data of the heart disease database in the input layer. In the output layer, there was only one single neuron denoting if heart disease present or not. The accuracy rate obtained from the study is 98% [9].

Tulay and Ozkan proposed the prediction of heart disease using multilayer perceptron back propagation neural network. To optimize the network, performance pruning, which defines a set of techniques for trimming the size of the network by nodes, was used. Dimensionality reduction with Principal Component Analysis (PCA) was made by reducing the number of neurons of the input layer from 13 neurons to 8 neurons to improve the performance. The proposed system gives 95% accuracy rate, which is a very good rate according to related studies on this field [10].

Kaan and Ahmet proposed diagnosis of heart disease using genetic algorithm based trained recurrent fuzzy neural networks. The RFNN used in this study had 13 inputs, 7 hidden neurons and 1 output neuron. The approach has an excellent rate of 100% for patients without heart disease that were found to have no heart disease in the testing set. The overall (297 instances) results had a sensitivity of 97.74%, F-score of 0.9626 and accuracy rate is 96.63% [11].

S. Islam Ayon, Md. Milon Islam and Md. Rahat Hossain have compared a number of computational intelligence techniques for the prediction of coronary artery heart disease named as Logistic Regression (LR), Support Vector Machine (SVM), Deep Neural Network (DNN), Decision Tree (DT), Naïve Bayes (NB), Random Forest (RF), and K-Nearest Neighbor (K-NN). They evaluated each technique and found the highest accuracy of 98.15% obtained by DNN [12].

The remaining part of the paper is organized as follows. Section 4 describes the overall methodology. Section 5 demonstrates the hyperparameter optimization at each step in terms of accuracy and F1-Score. Section 6 illustrates the prediction for understanding the graph. The experimental outcomes analysis is investigated in Section 7. Finally, Section 8 concludes the paper.

TABLE II. SUMMARY OF PREVIOUS STUDY

| ***Publication*** | ***Year*** | ***Title*** | ***Method*** | ***Result Accuracy*** |
|---|---|---|---|---|
| ELSEVIER | 1989 | International Application Of A New Probability Algorithm For The Diagnosis Of Coronary Artery Disease | LR | 77% |
| IEEE | 2010 | Decision Support System For Heart Disease Based On Support Vector Machine And Artificial Neural Network | ANN+SVM+MLP | 80.41% |
| IEEE | 2011 | Hdps: Heart Disease Prediction System | ANN+ VQA | 80% |
| IEEE | 2013 | Genetic Neural Network Based Data Mining In Prediction Of Heart Disease Using Risk Factors | ANN+GA+BPNN | 89% |
| IEEE | 2014 | An Ensemble Based Decision Support Framework For Intelligent Heart Disease Diagnosis | DT+SVO+NB | 82% |
| IEEE | 2014 | Prediction Of Heart Disease Using Learning Vector Quantization Algorithm | ANN+ VQA | 85.55% |
| IEEE | 2014 | Prediction Of Heart Disease Using Multilayer Perceptron Neural Network | ANN+MLP+BPNN | 98%. |
| IEEE | 2017 | Prediction Of Heart Disease Using Neural Network | ANN+BPNN | 95% |
| ELSEVIER | 2017 | Diagnosis Of Heart Disease Using Genetic Algorithm Based Trained Recurrent Fuzzy Neural Networks | RFNN | 96.63% |
| TANDF | 2020 | Coronary Artery Heart Disease Prediction: A Comparative Study of Computational Intelligence Techniques | DNN | 98.15% |

## IV. MODEL

### *A. Layers of Neural Network*

A multilayer perceptron neural network consists of multiple neurons that are organized in layers. While perceptron can only be able to classify linear separated data, multilayer perceptrons can also handle nonlinear separated data. In this binary classification problem, the dataset is not linearly separable, therefore, multilayer perceptron neural network performs well. Starting from here, a 2-layers neural network, one hidden layer and one output layer in addition to input layer have built. The number of hidden layers can be any number, but as it increases, the network gets deeper. Moreover, the complexity and run time of the model increase as well. The number of layers was kept low since there were not too much samples in the data set. As the data contains 13 features, the input layer has 13 nodes and output layer has one node for being binary classification problem. The hidden layer has 5 nodes, but it can accept any number of nodes according to performance of model, it is a hyper parameter and will be discussed under "Hyperparameter Optimization" section. The architecture of the model is shown in Figure 4.

### *B. Weights and Biases*

In order for a neural network to be fully connected, all nodes on one layer must be linked to the nodes on the next layer. For this purpose, all nodes in the input layer are multiplied by a weight and summed with bias. Here, a weight array of dimensions [13,5] and a bias array of dimensions [5,1] are used. Likewise, all nodes in the hidden layer are multiplied by a different weight array and summed with bias. At this stage, a weight array of dimensions [5,1] and a bias array of dimensions [1,1] are used. The initial weights and bias when creating the model are selected randomly. These values are saved and updated in every epoch.

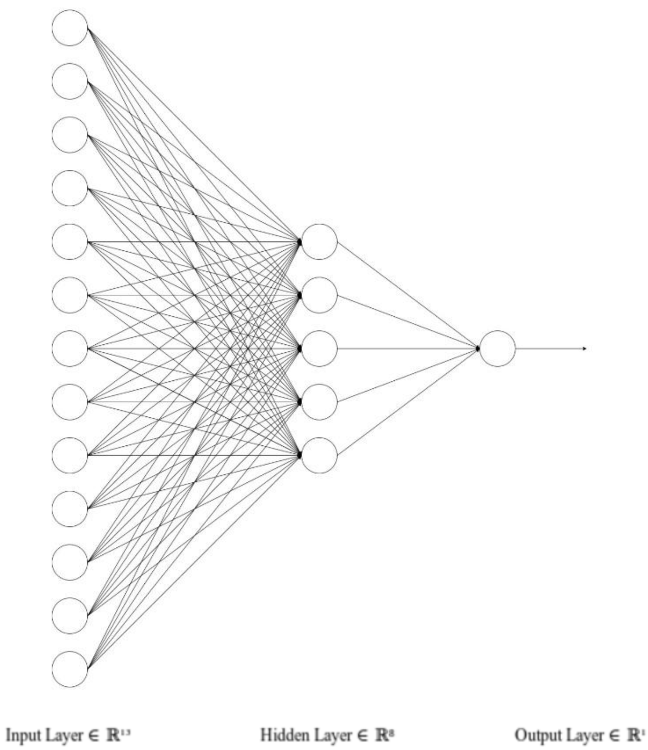


**Figure 4.** The architecture of neural network

### C. Activation Function

After multiplying by weights and summing with biases, the result is passed through an activation function which is used to turn the neuron output on or off depending on a rule or threshold, or function as a transformation that maps the input signals to the required output signals. Non-linear activation functions are useful for this problem. ReLu activation function is used at the end of the input layer, and Sigmoid activation function is used at the end of the hidden layer for this model. But, different activation functions are tested. The activation function is a hyperparameter and will be discussed under "Hyperparameter Optimization" section.

### D. Loss Function and Optimizer

The loss function is a measurement way of model prediction performance. It compares the predicted and true outputs. After each epoch, by updating the weights and biases, loss function decreases and predicted outputs approach to true ones. In this model, the Cross-Entropy loss function is used because it performs well in classification problems.

C : Number of classes

y : True output values

y' : Predicted output values

$$CE = -\sum_i^C y_i \log(\hat{y}_i) \tag{4}$$

After applying C =2, (5) stands for loss function:

$$CE = -\sum_{i=1}^{2} y_1 \log(\hat{y})$$
$$= -y_1 \log(\hat{y}_1) - (1 - y_1)\log(1 - \hat{y}_1) \tag{5}$$

Gradient Descent Algorithm is used to update weights and bias values by minimizing the loss function. Mathematical expression is shown in (6).

$$w* = w - a(\frac{\partial loss}{\partial w}) \tag{6}$$

### E. Learning Rate

The learning rate is the measure of how much the model will change according to the error when weights and bias are updated in each epoch. For this classification problem, 0.01 learning rate was found to be the optimum. The learning rate is a hyperparameter and will be discussed under "Hyperparameter Optimization" section.

### F. Stopping Criteria

The stopping criteria determines the number of epochs needed for training. When stopping criteria is satisfied, the model is not further trained. In this model, epoch number is set to 690 to have the best model performance. Stopping criteria is a hyperparameter and will be discussed under "Hyperparameter Optimization" section.

### G. Forward Propagation

Predicting as a result of operations using input values and weights is called feedforward. The nodes in the input layer are multiplied by weights and summed with biases and then passed through the ReLu activation function to connect to the hidden layer. Afterward, the output nodes of the input layer are multiplied by weights and summed with biases, then passed through the Sigmoid activation function to connect to the output layer. In this way, the predicted output value is obtained. The algorithm of these operations are listed below in (7), (8), (9), (10) and (11).

$$Z1 = X_i.W_i + b_i \tag{7}$$
$$A1 = ReLu(Z1) \tag{8}$$
$$Z2 = Z_i.W_i + b_i \tag{9}$$
$$A2 = Sigmoid(Z2) \tag{10}$$
$$Y_{pred} = A2 \tag{11}$$

The visualization of the described algorithm is shown in Figure 5.

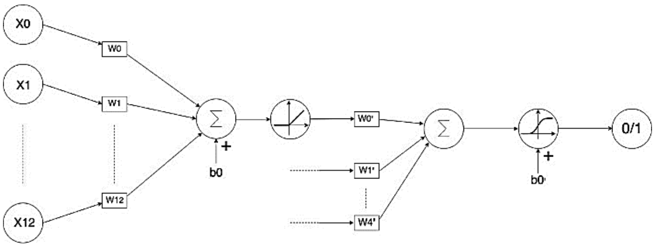


**Figure 5.** Visualization of the algorithm

### H. Back Propagation

First, the network predicts with random weights and bias. Then in the backpropagation phase, weights and bias are updated in each epoch according to the loss value obtained in the feedforward phase so that the feedforward and back forward work in a loop, ensuring that the network is well trained and the predicted output approaches the true output. The first step is calculating the loss of the predictions by finding the difference between the predicted and true output values. The second step is, by changing the weight and bias values, minimizing the loss. For this purpose, the derivative of loss with respect to predicted output value is calculated. After these steps, previous weights and biases are updated to minimize the loss.

## V. Hyperparameter Optimization

Hyperparameter values should be set before the learning process begins. They affect the performance, speed and quality of the model. The number of nodes in hidden layers, activation function, learning rate and stopping criteria are examples of important hyperparameters. By optimizing them, the performance, speed and quality of the model can be improved. For this purpose, different values of these parameters are tried.

The optimization of epoch number takes part in Table 3. The performance of the model increases as the epoch number increases until 690. Then, performance stays stable for a while and decreases after the 1000th epoch. Therefore, the epoch number is set to 690.

Optimization of hidden layer nodes takes part in Table 4. When the number of hidden layers nodes is set to 5, the model performs best. Therefore, the number of nodes in the hidden layer is set to 5. The optimization of activation function takes part in Table 5.

The optimization of activation function takes part in Table 5. As the network ends with a single unit that takes 0 or 1 values, since the sigmoid activation function takes values between 0 and 1, sigmoid is the most proper activation function for the output layer. For the hidden layer, various activation function types as ReLu, Tanh, Softmax and Sigmoid have been tried. After all, ReLu has given the best result, and it is used for training the model.

TABLE III. EPOCH NUMBER OPTIMIZATION

| *Epoch* | *Accuracy (Train)* | *Accuracy (Test)* | *F1-Score* |
|---|---|---|---|
| 100 | 88 | 78 | 81.034 |
| 150 | 89 | 77 | 80 |
| 200 | 89 | 80 | 82.758 |
| 225 | 90 | 80 | 82.758 |
| 250 | 90 | 80 | 82.758 |
| 300 | 90 | 80 | 82.758 |
| 350 | 90 | 80 | 82.758 |
| 400 | 90 | 78 | 80.701 |
| 500 | 88 | 83 | 85.217 |
| 600 | 89 | 83 | 85.217 |
| 680 | 89 | 83 | 85.217 |
| 690 | 85 | 84 | 85.964 |
| 800 | 89 | 84 | 85.964 |
| 1000 | 89 | 84 | 85.964 |
| 2000 | 93 | 78 | 79.629 |

TABLE IV. HIDDEN LAYER NODES OPTIMIZATION

| *Nodes* | *Accuracy (Train)* | *Accuracy (Test)* | *F1-Score* |
|---|---|---|---|
| 13,2,1 | 74 | 70 | 68.75 |
| 13,3,1 | 88 | 79 | 81.415 |
| 13,4,1 | 88 | 78 | 79.628 |
| 13,5,1 | 87 | 82 | 84.21 |
| 13,6,1 | 86 | 78 | 80.0 |
| 13,7,1 | 85 | 76 | 81.081 |
| 13,8,1 | 89 | 80 | 82.758 |
| 13,9,1 | 90 | 81 | 82.882 |
| 13,10,1 | 75 | 87 | 76.635 |

TABLE V. ACTIVATION FUNCTION OPTIMIZATION

| *Activation Function* | *Accuracy (Train)* | *Accuracy (Test)* | *F1-Score* |
|---|---|---|---|
| ReLu-Sigmoid | 89 | 84 | 85.964 |
| Tanh-Sigmoid | 83 | 71 | 71.280 |
| Softmax-Sigmoid | 87 | 81 | 85.478 |
| Sigmoid-Sigmoid | 89 | 81 | 83.478 |

The optimization of the learning rate takes part in Table 6. The model gives the best performance for 0.01 learning rate. Therefore, learning rate is set to 0.01.

TABLE VI. LEARNING RATE OPTIMIZATION

| *Learning Rate* | *Accuracy (Train)* | *Accuracy (Test)* | *F1-Score* |
|---|---|---|---|
| 0.0001 | 84 | 80 | 81.818 |
| 0.0005 | 88 | 83 | 85.217 |
| 0.001 | 89 | 84 | 85.964 |
| 0.005 | 96 | 78 | 80.0 |
| 0.01 | 98 | 85 | 86.725 |

## VI. PREDICTION

The training data is trained in the model created, and predictions are made on the model using test data.The relationship between epoch number and loss function is shown in Figure 5.

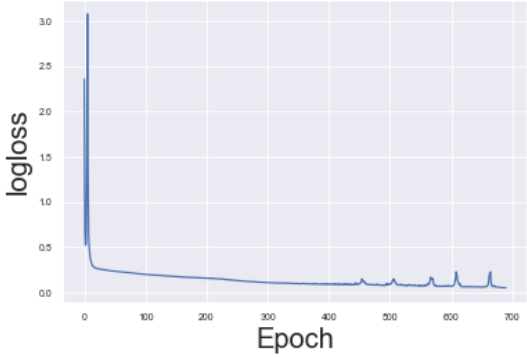


**Figure 5.** Epoch number vs. loss.

It is seen that the loss was very high at the beginning, but then declined rapidly. This shows how fast the network has learned. Confusion Matrix was used to measure the performance of the created algorithm. The Confusion Matrix is a table that visualizes the comparison of the test data with the known values and the prediction algorithm results. The number of correct and incorrect estimates are included in the table. The presentation of the correlation matrix is given in Figure 6.


**Figure 6.** Confusion matrix for the model

The sum of the values on the diagonal gives the correct prediction number. Here, 85 of the 100 test data is correctly predicted. According to the confusion matrix, Accuracy Score can be calculated as in (12):

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (12)$$

The model has 0.85 accuracy for test data.

According to the confusion matrix, F1-Score can be calculated as in (13):

$$F1 - Score = \frac{2*(TP / TP + FN)*(TP / TP + FP)}{(TP / TP + FN) + (TP / TP + FP)} \quad (13)$$

The model has 0.867 F1-Score for test data.

## VII. DISCUSSION

In order to understand the success of the model better, the results were evaluated by comparing it with other machine learning algorithms. The artificial neural network model is created using keras library. The accuracy values of the compared models are shown in Figure 7.

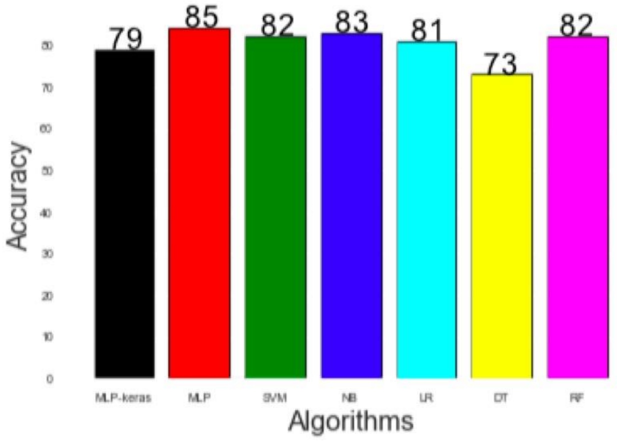


**Figure 7.** Accuracy of Different Methods

As seen in the figure, the model showing the most successful prediction performance among seven machine learning methods was the model proposed in this study. Among the reasons for this success can be the update of the weight of all features by the machine according to the loss function and the implementation of the hyperparameter optimization process.

## VIII. CONCLUSION

In addition to the traditional methods, some alternative ways are needed in heart disease detection, where the rate of death is high if the diagnosis and treatment are delayed. There are many studies using different techniques to produce alternative methods. In this article, a study on heart disease prediction is presented using multilayer perceptron artificial neural networks. In the network structure, there are 13 input neurons whose inputs are feature values of "Cleveland Dataset", a hidden layer containing 5 neurons, each of them is connected to all neurons in the previous layer, and an output neuron indicating whether the patient has heart disease or not. While the network structure is not large and complicated, it provides an advantage in terms of speed and memory; it is seen that the success rate is higher, compared to different machine learning methods. It is hoped that the studies carried out in this way will facilitate the work of healthcare professionals to ease of detection of heart disease. In future studies, an increment of data sample quantity helps to improve the achievement of the developed model.

## REFERENCES

[1] G. Eason, B. Noble, and I. N. Sneddon, “On certain integrals of Lipschitz-Hankel type involving products of Bessel functions,” Phil. Trans. Roy. Soc. London, vol. A247, pp. 529–551, April 1955. *(references)*

[2] A. Bellaachia and Erhan Guven, “Predicting Breast Cancer Survivability using Data Mining Techniques”, the Sixth SIAM International Conference on Data Mining (SDM 2006), Saturday, April 22, 2006.

[3] Detrano R, Janosi A, Steinbrunn W, et al. International application of a new probability algorithm for the diagnosis of coronary artery disease. Am J Cardiol. 1989;64(5):304-310. doi:10.1016/0002-9149(89)90524-9

[4] M. Gudadhe, K. Wankhade and S. Dongre, "Decision support system for heart disease based on support vector machine and Artificial Neural Network," 2010 International Conference on Computer and Communication Technology (ICCCT), Allahabad, Uttar Pradesh, 2010, pp. 741-745, doi: 10.1109/ICCCT.2010.5640377.

[5] A.H. Chen, S.Y. Huang, P.S. Hong, C.H. Cheng and E.J. Lin, HDPS: Heart Disease Prediction System, in: proceeding of International Conference Computing in Cardiology, 2011, pp. 557-5

[6] S. U. Amin, K. Agarwal and R. Beg, "Genetic neural network based data mining in prediction of heart disease using risk factors," 2013 IEEE Conference on Information & Communication Technologies, JeJu Island, 2013, pp. 1227-1231.doi: 10.1109/CICT.2013.6558288

[7] S. Bashir, U. Qamar and M. Younus Javed, "An ensemble based decision support framework for intelligent heart disease diagnosis," International Conference on Information Society (i-Society 2014), London, 2014, pp. 259- 264.doi: 10.1109/i-Society.2014.7009056

[8] J. S. Sonawane and D. R. Patil, "Prediction of heart disease using learning vector quantization algorithm," 2014 Conference on IT in Business, Industry and Government (CSIBIG), Indore, 2014, pp. 1-5.doi: 10.1109/CSIBIG.2014.7056973

[9] S. Sonawane and D. R. Patil, "Prediction of heart disease using multilayer perceptron neural network," International Conference on Information Communication and Embedded Systems (ICICES2014), Chennai, 2014, pp. 1-6.doi: 10.1109/ICICES.2014.7033860

[10] Tulay and Ozkan 2017.Prediction of Heart Disease Using Neural Network.IEEE:719-723

[11] Kaan Uyar and Ahmed 2017.Diagnosis of heart disease using genetic algorithm based trained recurrent fuzzy neural networks. Procedia Computer Science 120 (2017) 588–593

[12] Ayon, Safial & Islam, Md & Hossain, Md. (2020). Coronary Artery Heart Disease Prediction: A Comparative Study of Computational Intelligence Techniques. IETE Journal of Research. 1-20. 10.1080/03772063.2020.1713916.