



A Comparative Study for the Prediction of Heart Attack Risk and Associated Factors Using MLP and RBF Neural Networks

¹Rüstem Yılmaz , ²Fatma Hilal Yağın 

¹Samsun Gazi State Hospital, Department of Cardiology, Ilkyadim, Samsun, Turkey. (e-mail: drustemyilmaz@hotmail.com).

²Inonu University Department of Biostatistics and Medical Informatics, Faculty of Medicine, Malatya, Turkey. (e-mail: hilal.yagin@inonu.edu.tr).

ARTICLE INFO

Received: Sep.,23.2021

Revised: Nov.,9.2021

Accepted: Nov.,17.2021

Keywords:

Heart Attack
Machine Learning
Neural Networks
Classification
Variable Importance

Corresponding author: Fatma Hilal Yağın

✉ hilal.yagin@inonu.edu.tr

☎ +90 422 341 0660/1321

ISSN: 2548-0650

DOI: <https://doi.org/10.52876/jcs.1001680>

ABSTRACT

The aim of this study is to develop a predictive classification model that can identify risk factors for heart attack disease. In the study, patients with a low and high probability of having a heart attack were examined. Variable importance was calculated to identify risk factors. The radial basis function and multilayer perception neural networks were used to compare the classification prediction results. MLP model criteria; Accuracy 0.911, F1 score 0.918, Specificity 0.92, Sensitivity 0.903, while RBF model criteria were obtained as accuracy 0.797, F1 score 0.812, Specificity 0.84, Sensitivity 0.765. The first three most important factors that may be associated with having a heart attack were obtained as trestbps, oldpeak, and chol. According to the prediction results of the heart attack, it can be said that the model created with the MLP neural network has more successful predictions than the model created with the RBF neural network. In addition, estimating the importance values of the factors most associated with a heart attack (obtaining the most important biomarkers that may cause a heart attack) is a promising result for the diagnosis, treatment and prognosis of the disease.

1. INTRODUCTION

HEART attacks are usually caused by a blockage in the coronary artery. Despite improvements in the number of people surviving a heart attack, heart disease still remains the biggest killer. Early recognition of the cardinal symptoms of this disease is important for appropriate management, to prevent poor clinical outcomes [1,2].

Therefore, to shorten the time from heart attack onset to the hospital visit, it is important to improve the public's knowledge of stroke warning symptoms and signs. The symptoms of Myocardial Infarction (MI) include chest pain, which travels from left arm to neck, shortness of breath, sweating, nausea, vomiting, abnormal heart beating, anxiety, fatigue, weakness, stress, depression, and other factors [2,3].

Older age, male gender, lower education level, lack of regular exercise, unmarried status, unemployment, poor economic status, poor health behaviours (high-salt diet, no health screening), poor psychological status (self-perceived high stress and self-perceived poor health), and the presence

of hypertension or dyslipidemia independently predicted poorer understanding of Cardiovascular disease (CVD). The heart attack can be prevented by taking earlier action to lower those risks by controlling diet, fat, cholesterol, salt, smoking, nicotine, alcohol, drugs, monitoring of blood pressure every week, doing exercise every day, and losing body weight. As well as stopping smoking, other important ways of reducing risks are eating healthily, staying within safe limits of alcohol consumption, taking regular exercise and reaching and maintaining an ideal weight. Medication may also be prescribed to lower risks, including ACE inhibitors, antiplatelet therapy, beta-blockers and statins [2,4].

The ability to diagnose the cardiac disease quickly, accurately, and accurately plays a critical role in adopting preventative actions to avoid death. The electrocardiogram (ECG) determines the MI by electrical signals in the heart and damage to the blood supply to the heart muscle. The common blood tests are troponin and creatine kinase (CK-MB). ECG testing is used to differentiate between two types of

myocardial infarctions based on the shape of the tracing. An ST section of the tracing higher than the baseline is called an ST-segment elevation myocardial infarction (STEMI) which requires aggressive treatment. The coronary angiography or X-ray of the heart and blood vessels is performed to see the narrowing of coronary arteries. In addition, extensive studies have been made and various machine learning models are used to classify and predict heart disease [1,2].

Data mining is the process of extracting essential information from large datasets in a variety of disciplines, including medicine, business, and education. One of the most rapidly expanding areas of artificial intelligence is machine learning. These algorithms can analyze large amounts of data from a variety of sectors, including the medical field. It is a computer-assisted alternative to traditional prediction modelling for gaining knowledge of complicated and non-linear interactions among various components by reducing errors in projected and actual results. Data mining is the process of analyzing large datasets in order to uncover hidden critical decision-making information for future analysis. Patients' data is abundant in the medical area. These data must be analyzed using a variety of machine learning algorithms. These data are analyzed by healthcare professionals for them to make efficient diagnostic decisions. Through analysis, medical data mining using classification algorithms gives clinical assistance. It puts classification algorithms to the test to predict cardiac disease in patients [5-9].

In this study, Radial Basis Function and Multilayer Perceptron models were established and their results were compared to predict the disease effectively and accurately by identifying heart attack risk factors. In the continuation of the article, Section 2 consists of the materials and methodology used, Section 3 consists of analysis results and findings, and Section 4 consists of discussion.

2. MATERIAL AND METHODS

2.1. Dataset

An open-access data set titled "Health care: Data set on Heart Attack Possibility" is accessible at <https://www.kaggle.com/nareshbhat/health-care-data-set-on-heart-attack-possibility>.

TABLE I

THE VARIABLES IN THE DATASET AND THEIR DEFINITIONS	
Variable	Variable Description
age	age
sex	sex
cp	chest pain type (4 values)
trestbps	resting blood pressure
chol	serum cholestorol in mg/dl
fbs	fasting blood sugar > 120 mg/dl
restecg	resting electrocardiographic results (values 0,1,2)
thalach	maximum heart rate achieved
exang	exercise induced angina
oldpeak	ST depression induced by exercise relative to rest
slope	the slope of the peak exercise ST segment
ca	number of major vessels (0-3) colored by flourosopy
thal	0 = normal; 1 = fixed defect; 2 = reversable defect
target	0= less chance of heart attack 1= more chance of heart attack

The information was collected from heart-attack-possibility. In the data set used, there are 303 patients. While 138 (45.5%) of the patients had a low risk of heart attack, 165 (54.5%) of the patients had a high risk of a heart attack. The variables in the data set and their definitions are given in Table 1.

2.2. Neural networks

Artificial neural networks, or neural networks, are a subset of artificial intelligence. As shown in the taxonomy in Fig. 1, multilayer perceptron or radial basis function is one type of neural network [10].

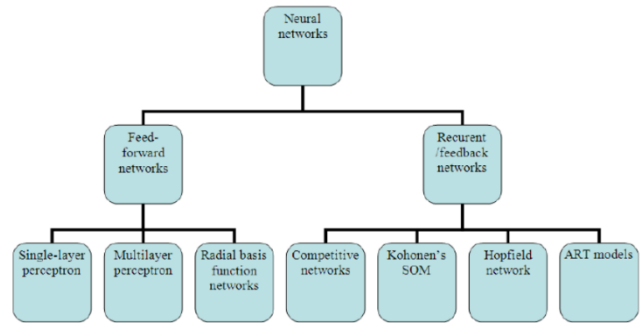


Fig. 1. A taxonomy of neural networks

2.2.1. Radial Basis Function (RBF)

In data analysis and pattern recognition applications, FUNCTIONAL estimation is a critical issue. Radial basis functions (RBFs) have features that make them suited for usage as universal approximators. A weighted sum of kernels can be used to define a continuous function. A kernel decomposition can be used in a two-layer neural network structure, with each kernel implemented by a hidden unit.

The network is given input-output pairs of samples selected from an observation set in supervised learning, and the learning algorithm identifies the rules that describe the given mapping. Because it can approximate any regular function and trains quicker than a multilayer perceptron, the radial basis function neural network (RBFN) can be used for a wide range of applications. The fact that RBFN has only two levels of weights and each layer may be determined sequentially contributes to the faster learning speed. An RBFN has three layers: an input, a middle layer, and an output layer. The pattern classes are represented by the input layer, which corresponds to the input vector space. As a result, determining the middle layer and the weights between the middle and output layers is all that is required to complete the architecture. When the middle layer is determined, the weights between the input and the middle layer are fixed [5,11].

2.2.2. Multilayer Perceptrons (MLP)

The multilayer perceptron, unlike other statistical techniques, makes no assumptions about the data distribution. It can represent extremely non-linear functions and be trained to generalize accurately when given new, previously unknown data. These characteristics make the multilayer perceptron an appealing option for constructing numerical models as well as choosing amongst statistical approaches. The multilayer perceptron has numerous uses in the atmospheric sciences, as will be seen. The multilayer perceptron is a model that represents a nonlinear mapping between an input vector and an output vector. It is made up of a system of basic interconnected neurons, or nodes, as shown in Fig. 2. Weights and output signals connect the nodes, and the output signals

are a function of the sum of the node's inputs, modified by a simple nonlinear transfer, or activation, function [5,12].

A multilayer perceptron's construction varies, but it usually consists of many layers of neurons. The input layer serves only to provide the input vector to the network and does not perform any computations. The phrases input and output vectors relate to the multilayer perceptron's inputs and outputs, which can be represented as single vectors in Fig. 2. One or more hidden layers, followed by an output layer, make up a multilayer perceptron. Multilayer perceptrons are fully connected, with each node coupled to every node in the layer above and below it [5].

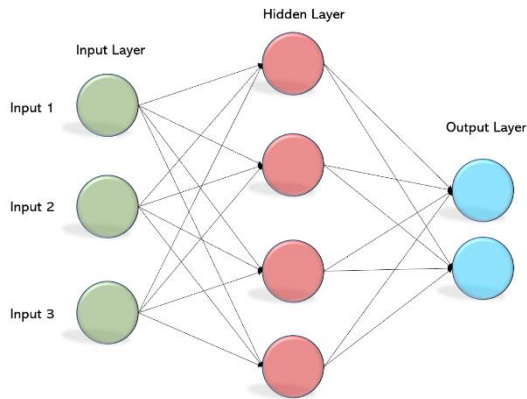


Fig. 2. A multilayer perceptron with two hidden layers.

3. EXPERIMENTAL RESULTS

The classification matrix of RBF and MLP models created in this study to classify heart attacks are given in Table II and Table III, respectively.

TABLE II
CLASSIFICATION MATRIX OF RBF MODEL

Prediction	Reference	
	more chance of heart attack	less chance of heart attack
more chance of heart attack	26	4
less chance of heart attack	8	21

TABLE III
CLASSIFICATION MATRIX OF MLP MODEL

Prediction	Reference	
	more chance of heart attack	less chance of heart attack
more chance of heart attack	28	2
less chance of heart attack	3	23

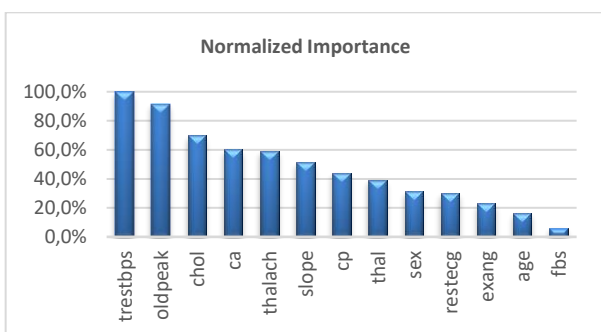


Fig. 3. Variable Importance

TABLE IV
IMPORTANCE SCORES OF FACTORS RELATED TO HEART ATTACK

Factors	Importance
trestbps	0,161
oldpeak	0,147
chol	0,112
ca	0,097
thalach	0,095
slope	0,082
cp	0,071
thal	0,063
sex	0,050
restecg	0,048
exang	0,037
age	0,027
fbs	0,010

When Table IV and Figure 3 are examined; the factors that most effect the risk of having a heart attack were found to be trestbps, oldpeak, and chol.

TABLE V
VALUES FOR THE METRICS OF THE CLASSIFICATIONS PERFORMANCE OF MLP AND RBF MODELS

Metric	MLP	RBF
	Value	
Accuracy	0.911	0.797
F1-Score	0.918	0.812
Specificity	0.92	0.84
Sensitivity	0.903	0.765
Negative Predictive Value	0.885	0.724
Positive Predictive Value	0.933	0.867

The values for the classification performance metrics of the generated RBF and MLP models are given in Table V. For the MLP model is calculated as accuracy 0.911, F1 score 0.918, Specificity 0.92, Sensitivity 0.903, Negative Predictive Value 0.885 and Positive Predictive Value 0.933. The RBF model is calculated as accuracy 0.797, F1 score 0.812, Specificity 0.84, Sensitivity 0.765, Negative Predictive Value 0.86724. In both models, the ability to distinguish positive cases (i.e., patients with high probability of heart attacks) was better (compared to negative cases). In addition, the MLP model was more successful in classifying heart attacks than the RBF model.

4. DISCUSSION

Ischemic heart disease (IHD) is one of the leading causes of death and morbidity in the world. MI, which is defined as myocardial cell damage induced by persistent ischemia, is one of these disorders. A heart attack is a physiological condition characterized by extreme chest pain and the likelihood of mortality as a result of heart failure caused by a problem in the coronary arteries. A heart attack occurs when the heart's oxygen supply is cut off due to a rapid decrease or interruption in blood flow in the veins that supply it. It can harm or kill the cardiac muscle fed by the blocked vessel to varying degrees.

Heart attack is the most common health concern in affluent countries, as well as a major health issue that is becoming more prevalent in underdeveloped countries. MI is a severe public health issue that affects the productive age group of the population, generates serious complications from post-acute phase issues, and can lead to death. The World Health Organization (WHO) estimates that 16.7 million people die each year as a result of a heart attack. This equates to one-third of all deaths worldwide (13, 14).

The goal of machine learning and deep learning is to filter out undiscovered patterns and relations in data. These patterns are also utilized in the development of various prediction models. Technology advancements have aided in the automation of numerous functional units across multiple disciplines [15].

Health care is one area where many electronic devices provide a large amount of complex connected data about hospitals, patients, and diseases. This unprocessed data can be a valuable resource, but it must be properly processed. These data can be processed to extract important information [16].

The use of machine learning and data mining techniques in the field of health care have ushered in a new era of computing. Various data mining approaches have been extensively used to efficiently diagnose cardiac disease. The fundamental issue with machine learning models is that they frequently require feature engineering for proper implementation, which can be a time-consuming process. Deep learning and neural networks have been employed extensively for various classification tasks in the health area, particularly in cardiovascular disease, to address the aforesaid challenge [17,18].

Using existing medical records for model training and testing, this article presents a proposed model built using current machine learning approaches to detect and predict heart diseases and heart attacks. In this article, MLP and RBF methods were used for classification. In addition, new patterns such as variable importance were discovered from the analysis. After application, 91.1% and 79.7% accuracy of MLP and RBF were obtained, respectively. The model developed in this study can be used to assist medical professionals/practitioners in detecting and predicting heart disease/crisis to minimize deaths in the healthcare industry, given the annual death rate it causes.

In conclusion, the findings from this study showed that the model (MLP) created in the classification of the severity of the probability of heart attack gives successful predictions. In addition, it is thought that by estimating the importance values of the factors associated with heart attack and determining the most important factors, these results will provide benefits to healthcare professionals for diagnosing and treating a heart attack.

REFERENCES

- [1] Brice, J. H., Griswell, J. K., Delbridge, T. R., & Key, C. B. (2002). Stroke: From Recognition by The Public To Management by Emergency Medical Services. *Prehospital Emergency Care*, 6(1), 99-106.
- [2] Park, M. H., Jo, S. A., Jo, I., Kim, E., Eun, S. Y., Han, C., & Park, M. K. (2006). No difference in stroke knowledge between Korean adherents to traditional and western medicine—the AGE study: an epidemiological study. *BMC Public Health*, 6(1), 1-9.
- [3] Küçükakçali Tunç, Z., Çiçek Balıkcı, İ., GÜldoğan, E., & Çolak, C. (2020). Assessment of Associative Classification Approach for Predicting Mortality by Heart Failure. *The Journal of Cognitive Systems*, 5(2), 41-45.
- [4] Küçükakçali Tunç, Z., Çiçek Balıkcı, İ., & GÜldoğan, E. Performance Evaluation of The Deep Learning Models in The Classification of Heart Attack And Determination of Related Factors. *The Journal of Cognitive Systems*, 5(2), 99-103.
- [5] Çiçek Balıkcı, İ., & Küçükakçali, Z. (2020). Classification of Prostate Cancer and Determination of Related Factors with Different Artificial Neural Network. *Middle Black Sea Journal of Health Science*, 6(3), 325-332.
- [6] Çiçek Balıkcı, İ., & Küçükakçali, Z. & Çolak, C. Associative Classification Approach Can Predict Prostate Cancer Based On The Extracted Association Rules. *The Journal of Cognitive Systems*, 5(2), 51-54.
- [7] Küçükakçali Tunç, Z., & Çiçek Balıkcı, İ. Performance Evaluation Of The Ensemble Learning Models In The Classification Of Chronic Kidney Failure. *The Journal of Cognitive Systems*, 5(2), 55-59.
- [8] Perçin, İ., Yağın, F. H., Arslan, A. K., & Çolak, C. (2019, October). An Interactive Web Tool for Classification Problems Based on Machine Learning Algorithms Using Java Programming Language: Data Classification Software. In 2019 3rd International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT) (pp. 1-7). IEEE.
- [9] Leist, A., & Muniz-Terrera, G. (2021). Symposium: Contextual and life-course determinants of later-life cognitive functioning and dementia.
- [10] Yağın, F. H., GÜldoğan, E., Ucuza, H., & Çolak, C. A Computer-Assisted Diagnosis Tool for Classifying COVID-19 based on Chest X-Ray Images. *Konuralp Medical Journal*, 13(S1), 438-445.
- [11] Yegnanarayana, B. (2009). Artificial neural networks. PHI Learning Pvt. Ltd.
- [12] Orr, M. J. (1996). Introduction to radial basis function networks.
- [13] Karaman, U., & Çiçek Balıkcı, İ., Determination Of Cryptosporidium Spp. Risk Factors Using Multilayer Perceptron Neural Network And Radial Based Functional Artificial Neural Network Method. *The Journal of Cognitive Systems*, 5(2), 83-87.
- [14] Després, J. P., Lamarche, B., Mauriège, P., Cantin, B., Dagenais, G. R., Moorjani, S., & Lupien, P. J. (1996). Hyperinsulinemia as an independent risk factor for ischemic heart disease. *New England Journal of Medicine*, 334(15), 952-958.
- [15] McNeer, J. F., Margolis, J. R., Lee, K. L., Kisslo, J. A., Peter, R. H., Kong, Y. I. H. O. N. G., ... & Rosati, R. A. (1978). The role of the exercise test in the evaluation of patients for ischemic heart disease. *Circulation*, 57(1), 64-70.
- [16] Deo, R. C. Machine learning in medicine. *Circulation [Internet]* 2015; 132 (20). 1920–30.
- [17] Sidey-Gibbons, J. A., & Sidey-Gibbons, C. J. (2019). Machine learning in medicine: a practical introduction. *BMC medical research methodology*, 19(1), 1-18.
- [18] Perçin, İ., Yağın, F. H., GÜldoğan, E., & Yoloğlu, S. (2019, September). ARM: An Interactive Web Software for Association Rules Mining and an Application in Medicine. In 2019 International Artificial Intelligence and Data Processing Symposium (IDAP) (pp. 1-5). IEEE.
- [19] Al'Aref, S. J., Anchouche, K., Singh, G., Slomka, P. J., Kolli, K. K., Kumar, A., ... & Min, J. K. (2019). Clinical applications of machine learning in cardiovascular disease and its relevance to cardiac imaging. *European heart journal*, 40(24), 1975-1986.

BIOGRAPHIES

Rüstem Yılmaz obtained his M.B. degree in Medicine from Uludağ University in 2003. He received M.D. degree in cardiology from the İzmir Atatürk Training and Research Hospital in 2009. In 2011, he joined the Department of Cardiology at Gazi State Hospital in Samsun as a specialist. His research interests are coronary artery disease, arrhythmias, and pacemakers.

Fatma Hilal Yağın obtained her BSc. degree in Statistics from Gazi University in 2017. She received MSc. degree in biostatistics and medical informatics from the Inonu University in 2020. She currently continues Ph.D. education in biostatistics and medical informatics from the Inonu University. In 2019, she joined the Department of Biostatistics and Medical Informatics at Inonu University as a researcher assistant. Her research interests are cognitive systems, data mining, machine learning, deep learning, and image processing.