DOI: 10.37990/medr.1031866 Med Records 2022;4(2):171-8

MEDICAL RECORDS-International Medical Journal

Research Article

Estimation of Risk Factors Related to Heart Diseases With Multilayer Perceptron Model

Kalp Hastalıklarına İlişkin Risk Faktörlerinin Multilayer Perceptron Modeli ile Tahmini

Mehmet Gunata[1], Ahmet Kadir Arslan[2], Cemil Colak[2], Hakan Parlakpinar[1]

[1]Inonu University, Faculty of Medicine, Department of Medical Pharmacology, Malatya, Turkey
[2]Inonu University, Faculty of Medicine, Department of Biostatistics and Medical Informatics, Malatya, Turkey

Abstract

Aim: Heart diseases (HD) refer to many diseases such as coronary heart disease, heart failure, and heart attack. Every year, approximately 647.000 people die in the United States (U.S.) from HD. Genetic and environmental risk factors have been identified due to numerous studies to determine HD risk factors.
Material and Method: In this study, the Multilayer Perceptron (MLP) model was constructed to predict the risk factors related to HD in both genders. The relevant dataset consisted of 270 individuals, 13 predictors, and one response/target variable. Model performance was evaluated using overall accuracy, the area under the ROC (Receiver Operating Characteristics) curve (AUC), sensitivity, and specificity metrics.
Results: The performance metric values for accuracy, AUC, sensitivity and specificity were obtained with 95% CI, 0.876 (0.79-0.937), 0.935 (0.877-0.992), 0.921 (0.786-0.983) and 0.843 (0.714-0.93), respectively. According to the relevant model findings, blood pressure, the number of significant vessels coloured by fluoroscopy, and cholesterol variables were the three most crucial HD classification factors.
Discussion: It can be said that the model used in the present study offers an acceptable estimation performance when all performance metrics are considered. In addition, when compared with the studies in the literature from both data science and statistical point of view, it can be stated that the findings in the current study are more satisfactory.
Conclusion: Due to the predictive performance in this study, the MLP model can be recommended to clinicians as a clinical decision support system. Finally, we propose solutions and future research pathways for the various computational materials science challenges for early HD diagnosis.

Keywords: Heart disease, multilayer perceptron, risk factors, prediction, clinical decision support system

Öz

Amaç: Kalp hastalıkları (HD); koroner kalp hastalığı, kalp yetmezliği ve kalp krizi gibi birçok hastalığı ifade eder. Amerika Birleşik Devletleri'nde (U.S.) her yıl yaklaşık 647.000 kişi HD'den ölmektedir. HD risk faktörlerini belirlemeye yönelik çok sayıda çalışma neticesinde genetik ve çevresel risk faktörleri tanımlanmıştır.
Materyal ve Metot: Bu çalışmada, her iki cinsiyette de kalp hastalığına bağlı risk faktörlerini tahmin etmek için Multilayer Perceptron (MLP) modeli oluşturulmuştur. İlgili veri seti 270 kişiden, 13 tahmin ediciden ve bir yanıt/hedef değişkeninden oluşmaktadır. Model performansı, genel doğruluk, ROC (Alıcı Çalışma Karakteristikleri) eğrisi (AUC) altındaki alan, duyarlılık ve özgüllük metrikleri kullanılarak değerlendirildi.
Bulgular: Doğruluk, AUC, duyarlılık ve özgüllük için performans metrik değerleri sırasıyla 95% CI, 0.876 (0.79-0.937), 0.935 (0.877-0.992), 0.921 (0.786-0.983) ve 0.843 (0.714-0.93) şeklinde elde edildi. İlgili model bulgularına göre, kan basıncı, floroskopi ile renklendirilen önemli damar sayısı ve kolesterol değişkenleri en önemli üç HD sınıflandırma faktörü olarak görüldü.
Tartışma: Bu çalışmada kullanılan modelin tüm performans ölçütleri dikkate alındığında kabul edilebilir bir tahmin performansı sunduğu söylenebilir. Ayrıca hem veri bilimi hem de istatistiksel açıdan literatürdeki çalışmalarla karşılaştırıldığında, mevcut çalışmadaki bulguların daha tatmin edici olduğu ifade edilebilir.
Sonuç: Bu çalışmadaki öngörücü performans nedeniyle, MLP modeli klinik karar destek sistemi olarak klinisyenlere önerilebilir. Son olarak, erken HD teşhisi için çeşitli hesaba dayalı bilim alanında çözümler ve yeni araştırmalar öneriyoruz.

Anahtar Kelimeler: Kalp hastalığı, çok katmanlı algılayıcı, risk faktörleri, tahmin, klinik karar destek sistemi

Received: 03.12.2021 **Accepted:** 27.01.2022
Corresponding Author: Mehmet Gunata, Inonu University, Faculty of Medicine, Department of Medical Pharmacology, Malatya, Turkey, **E-mail:** mehmet.gunata@inonu.edu.tr

INTRODUCTION

In the world full of advanced computer technologies, artificial intelligence (machine learning and deep learning), one of the essential computer science fields, is used in many areas, especially in medicine. It is used to predict the presence of diseases such as heart diseases (HD) (1). These estimates can be made by obtaining important data obtained from patients' medical databases using various algorithms (2).

Medical organizations around the world collect a lot of essential health data. Increasing data in today's world can be useful in early diagnosis or triage in medicine using machine learning techniques (3). However, the data collected is extensive, complex, difficult to analyze, and takes much time. These datasets, which are too overwhelming for the human mind to understand, can be easily analyzed using various machine learning techniques. Therefore, these algorithms have recently become very useful for accurately predicting heart-related diseases' presence or absence (4).

Heart Diseases (HD)

Cardiovascular disease (CVD) is a term used for conditions that affect the heart or blood vessels (5). Although there are different variations, CVD is divided into four main headings:

1. Coronary heart disease: Myocardial infarction (MI), angina pectoris, and heart failure (HF)
2. Cerebrovascular disease: Stroke and transient ischemic attack
3. Peripheral artery disease
4. Aortic atherosclerosis and thoracic or abdominal aortic aneurysm

In coronary heart disease, the heart tissue cannot be fed sufficiently due to the cholesterol accumulated in the coronary vessels. The heart cannot receive enough blood due to the cholesterol and fat accumulated in the arteries' wall that supplies the blood to the heart. In MI, also known as a heart attack, it is a fatal condition caused by blockage of blood in the coronary artery wall. HF is a progressive clinical picture that does not fill/discharge due to a problem due to the heart tissue's function or structure, resulting in the peripheral tissues' inability to send enough blood to meet their metabolic needs (5). Angina is a chest pain caused by insufficient blood flow to the heart. Due to the low return, the heart cannot work physiologically. Cerebrovascular disease refers to diseases that affect blood flow to the brain due to congestion or malformation that affect blood vessels. Peripheral artery disease is a condition where the peripheral tissues are not fed enough blood due to narrowing the legs' arteries. Other forms of HD include valvular heart disease, stroke, hypertension, etc.

Epidemiology

CVD is a group of diseases that affect most adults older than 60 years of age and have severe morbidity and mortality. It is common in the general population worldwide, and estimated that 17.3 million deaths from CVD occurred annually in the early 2010s. CVD caused more than 12 million (25.8%) deaths in 1990 and approximately 18 million (32.1%) deaths in 2015 (6). Also, HD and stroke constitute 80% of deaths due to CVD in men and 75% in women (7).

HD more common in older ages, and age is a significant risk factor for CVD. Recently, it was reported that CVD occurs in only 11% of people aged 20-40, while 37% of people aged 40-60, 71% of people aged 60-80, and 85% of people over 80 in the U.S. Approximately 50% of people over the age of 20 have CVD in the U.S., according to the 2019 Heart Disease and Stroke Statistics update of the American Heart Association (8).

HD are the leading cause of death for men, women, and people in most racial and ethnic groups in the U.S. One person with CVD dies every 37 seconds in the U.S. About 647.000 Americans die each year from HD (8). Also, HD cost is about $219 billion annually from 2014 to 2015 in the U.S. (9), including healthcare, medicines expenditure, and loss of productivity from death.

The average age of death from coronary artery disease is around 80 in developed countries; however, it is about 68 in developing countries. The disease's symptoms and diagnosis typically occur in men 7-10 years earlier than in women (7).

HD accounts for approximately one-third and half of total CVD patients. Also, ischemic HD are the most common cause of death in adults in all low, middle, and high-income countries (10). About 9 million people died in 2016 due to ischemic heart disease (5). The lifetime risk of HD has been demonstrated in the Framingham Heart Study in 7.733 individuals aged 40-94 years. The lifetime risk of 40 years old is 49% for men, while women have a lower risk and 32%. Similar results were achieved in a meta-analysis study in which more than 250.000 women and men were included, and data on 18 studies were analyzed (11).

Although it is more common in older ages, it has been shown that the early stages of atherosclerosis begin to occur starting from the second and third years of life according to autopsy data (12). The life span has been long since 1975, CVD's prevalence and related complications are still very high, and treatment costs are relatively high (13).

In a cohort study of over 1.9 million people older than 30 years old who did not have CVD before, patients were followed up for six years. According to this; angina, HF, peripheral artery disease, transient ischemic attack, and abdominal aortic aneurysm constitute 66% of CVD (14).

More than 4 million people die from HD annually among 49 countries in Europe and North Asia. About 1,5 million people experience a heart attack or stroke annually, resulting in over 250.000 deaths in the U.S. (15). Despite advances in medicine and science, CVD's prevalence is increasing rapidly in developing countries. It is estimated that the global HD burden increased by 29% between 1990 and 2010 (16).

Risk factors

There are many known risk factors for HD including genetic predisposition, family history, age, gender, tobacco use, physical inactivity, excessive alcohol consumption, unhealthy diet, obesity, hypertension, diabetes mellitus, hyperlipidemia, psychosocial factors, poverty, low educational status, and air pollution, etc. (5).

Nine potentially changeable risk factors were placed in the INTERHEART study, which included 52 countries. These are smoking, dyslipidemia, hypertension, diabetes, abdominal obesity, psychosocial factors, daily fruit and vegetable consumption, alcohol consumption, and regular physical activity. It has been shown that improving these risk factors can reduce CVD by approximately 90% (17).

Genetic and family history

Genetic factors affect the development of CVD, especially in men under 55 and women under 65. Having a CVD in one of the parents increases CVD's risk in the person three times. The risk of developing HD varies between 15-100% in the presence of positive family history (18).

According to information obtained from several extensive cohort studies following more than 163.000 patients, it has been observed that a positive family member for CVD is associated with the risk of developing HD (19). Also, 3.9 million people who were born between 1950-2008 were followed in another study. It has been found that children who have had two or more early cardiovascular deaths among their first-degree relatives had three times the risk of developing CVD before age 50 (20). More than one single nucleotide polymorphism is associated with CVD in genetic association studies; however, it generally has little individual effects, and genetic contributions to CVD are not fully understood (21). Some mutations associated with leukemia in blood cells are also thought to cause an increased risk of CVD. According to the data obtained from many studies examining CVD's genetic background, it has been observed that the presence of these mutations can cause events and mortality associated with CVD (22).

Age

Age is the most critical risk factor in CVD development and carries about three times more risk every ten years of life (23). Coronary fat lines may begin to form during adolescence (24). Nearly 82% of people aged 65 and over are thought to have died due to HD (25).

Although multiple reasons increase CVD incidence with aging; the most common are pathological changes in blood vessels and increased serum cholesterol levels. The serum total cholesterol level rises with aging in most populations. This increase is around 45-50 years old in males, while in females, it is evident in 60-65 years old (26). Depending on aging, arterial elasticity loss occurs, which can reduce the artery's ability to adapt and lead to coronary artery disease (26).

Gender

Men are more likely to suffer from HD than premenstrual women (23). It is known that coronary HD is seen 2-5 times more in middle-aged men than in women (26). A study conducted by the World Health Organization (WHO) showed that gender had a 40% impact on HD mortality (27). Similar results have been reported showing that gender differences explain approximately half of CVD's risk in another study (26). Also, the predominant hormone in women is an estrogenic effect. Estrogen is known to have significant impacts not only on carbohydrate metabolism but also on the cardiovascular system. It is also known to have significant protective effects on hemodynamics and directly affect improving endothelial cell function. Estrogen production decreases, which negatively affects lipid levels in women after menopause (26).

There are also differences between men and women regarding body weight, height, body fat distribution, and hemodynamic parameters. This situation may cause CVD to be more common in men due to these changes. It is known that male sex is a significant and independent risk for HD. The reason for this is not fully known (28).

According to the data obtained from ONTARGET and TRANSCEND studies (9.378 women, 22.168 men) performed on 31.000 patients followed for an average of 56 months, women were at a 20% lower risk of all major cardiovascular events, including cardiovascular death, than men (29).

Hypertension

It is known that hypertension is the most investigated and important risk factor for many cardiovascular system diseases including HD. In a cohort study of over 1.25 million patients aged 30 years and older without CVD, including 20% with initially treated hypertension, the risk of developing lifetime CVD in patients with hypertension was 63.3% (30).

Smoking

Health risks arising from smoking arise from direct tobacco consumption and passive exposure. Approximately 10% of CVD occurs due to smoking (7). The incidence of MI increases six times in women and three times in men who smoke at least 20 cigarettes daily than non-smokers (31).

Physical inactivity

Insufficient physical activity (less than 5x30 minutes per week) is now ranked 4th among the causes of mortality worldwide (7). According to the 2008 study, 31.3% (28.2% men and 34.4% women) of adults aged 15 and over are not physically active (7). The risk of HD and diabetes mellitus decreases by approximately one-third in those who regularly participate in a moderate sports activity for about 2.5 hours each week. Also, weight control is provided depending on physical activity, and blood glucose and lipid parameters also improve. Physical activity also improves blood pressure and insulin sensitivity, leading to reduced CVD (7).

Diet

It is known that high amounts of saturated fat, trans-fatty acids, salt, low consumption of fruits, vegetables, and fish increase the risk of CVD. WHO reported that 1.7 million people died worldwide due to not consuming enough fruits and vegetables (7). It is known that consuming foods with high fat and sugar content increases important risk factors for CVD, such as obesity (7). It has also been shown that the amount of salt consumed daily affects blood pressure and is an important risk factor for CVD (7). Also, it has been shown that reducing the consumption of saturated fat for about two years reduces the risk of CVD. Also, consuming high sugar foods negatively affects blood lipid and glucose parameters and increases diabetes mellitus and HD incidence. Also, excessive consumption of processed meats increases the risk of CVD (32).

Socioeconomic level

Although CVD affects all countries, it affects low and middle-income countries more than other countries. Although there is not much information, low income and low education levels have a higher risk of CVD (32).

Air pollution

According to the related-studies, long-term exposure to particles in the air under 2.5 micrometers diameter increases the rate of atherosclerosis and inflammation. Short-term (2 hours) exposure to air pollution causes an approximately 50% increase in CVD's mortality rate. It has also been shown that, after just five days of exposure to air pollution, both systolic and diastolic blood pressure increases by 2.8 mmHg and 2.7 mmHg, respectively (33). According to other studies, air pollution causes both arrhythmia and HF. Also, it is associated with carotid artery thickening and acute MI (34).

Machine Learning Algorithms

Today, machine learning algorithms assist clinicians in decision support systems in various tasks in medicine (35). Thanks to data science developments, computer technologies, and neural network-based systems, machine learning technologies are strengthening their position every day in the medical world where large and different data types (image, video, and soon) are abundant (36). As a result of these developments, health services in some areas are aimed to be faster, less costly, and reliable (37).

It is a complicated process that is decided by using clinical data and clinical experience in diagnosing a disease. Making this decision-making process less costly, more accessible, faster, more accurate, and efficient will improve patients' quality of life. As mentioned above, too many people suffer from HD in today's world and die or fight their complications due to these diseases. Also, patients cannot benefit from health services sufficiently because of time limitations and the excessive number of patients. For all these reasons, early and accurate detection of HD will save many patients' lives. Unfortunately, the cardiovascular system's complicated process, late onset of many symptoms, and genetic differences delay inappropriate treatment.

Therefore, there is a need to develop HD estimation systems to assist medical professionals in the early and accurate HD diagnosis. In this study, Multilayer Perceptron (MLP), one of the well-known neural network-based machine learning models, was utilized to diagnose of HD and determine the essential variables/factors as a risk factor related to HD (38).

MATERIAL AND METHOD

The dataset used in this study consisted of 270 individuals, 13 predictors, and one response/output variable. The dataset used in this study was taken from the www.kaggle.com/ronitf/heart-disease-uci. A detailed information table regarding the variables is shown in Table 1, and descriptive statistics of all variables in the dataset are shown in Table 2-3.

The MLP is a type of artificial neural network that is performed in many disciplines. The MLP architecture has one input layer, one or more hidden layers, and one output layer. Each layer comprises of several nodes. The inputs in the first layer nodes are weighted and sent simultaneously to a second or hidden layer. Each of the nodes takes the weighted sum of the outputs of the previous layer as input and implements an activation function to determine its output. The activation function is usually a sigmoid function that converts the output to a number between 0 and 1 (39).

The MLP model was applied to predict HD using these 13 risk factors. The dataset is randomly divided into training and testing at 70% and 30% ratios in order to obtain unbiased model performance metrics at 70% and 30% ratios, respectively. While model training was done in 70% of the dataset, the model's learning performance was tested in the remaining 30%. The tuning hyperparameters for the MLP model are shown in Table 2.

Table 1. Types and roles of the variables

Variable	Type	Role
Age	Numerical	Predictor
Sex	Categorical	Predictor
Chest pain type	Categorical	Predictor
Blood pressure	Numerical	Predictor
Cholesterol	Numerical	Predictor
FBS over 120	Categorical	Predictor
ECG results	Categorical	Predictor
Max heart rate	Numerical	Predictor
Exercise angina	Categorical	Predictor
ST depression	Numerical	Predictor
Slope of ST	Categorical	Predictor
Number of vessels fluro	Categorical	Predictor
Thallium	Categorical	Predictor
Heart disease	Categorical	Response/output

FBS, fasting blood sugar; ECG, electrocardiogram

Table 2. Tuning hyperparameters of the MLP model	
Parameter	Value
Number of the hidden layer(s)	1
Number of node(s)	2
Activation function	Softmax
Error function	Cross-entropy
Optimization function	Gradient descent
Learning rate	0.4
Momentum	0.9

Model performance was evaluated using overall accuracy ratio, area under the ROC (Receiver Operating Characteristics) curve (AUC), sensitivity, and specificity metrics. IBM SPSS Statistics 25 package program was used for the analyses.

RESULTS

The descriptive statistics of the categorical and numerical variables are summarized in Tables 3 and 4. Also, the ROC curve graph for the performance metrics of the MLP model by the classification matrix is presented in Table 5 and shown in Figure 1. The contributions of the predictive variables to the classification performance of the MLP model are shown in Table 6.

The performance metric values for accuracy, AUC, sensitivity and specificity were obtained with 95% CI, 0.876 (0.79-0.937), 0.935 (0.877-0.992), 0.921 (0.786-0.983) and 0.843 (0.714-0.93), respectively. The general structure and performance metric values of MLP for HD is shown in Figure 2.

Table 3. The descriptive statistics of the categorical variables

Variable	Category	n	%
Heart disease	Negative	150	55.56
	Positive	120	44.44
	Total	270	100.00
Sex	0	87	32.22
	1	183	67.78
	Total	270	100.00
Chest pain type	1	20	7.41
	2	42	15.56
	3	79	29.26
	4	129	47.78
	Total	270	100.00
FBS over 120	0	230	85.19
	1	40	14.81
	Total	270	100.00
ECG results	0	131	48.52
	1	2	0.74
	2	137	50.74
	Total	270	100.00
Exercise angina	0	181	67.04
	1	89	32.96
	Total	270	100.00
Slope of ST	1	130	48.15
	2	122	45.19
	3	18	6.67
	Total	270	100.00
Number of vessels fluro	0	160	59.26
	1	58	21.48
	2	33	12.22
	3	19	7.04
	Total	270	100.00
Thallium	3	152	56.30
	6	14	5.19
	7	104	38.52
	Total	270	100.00

FBS, fasting blood sugar; ECG, electrocardiogram

Table 4. The descriptive statistics of numerical variables

Statistics	Variables				
	Age	Blood Pressure	Cholesterol	Max Heart Rate	ST depression
N	270	270	270	270	270
Mean	54.43	131.34	249.66	149.678	1.05
Median	55	130	245	153.5	0.8
Std. Deviation	9.109	17.86	51.686	23.166	1.145
Minimum	29	94	126	71	0
Maximum	77	200	564	202	6.2

Table 5. The confusion matrix

Heart Disease (Model prediction)	Heart Disease (Real)		
	Positive	Negative	Total
Positive	35	8	43
Negative	3	43	46
Total	38	51	89

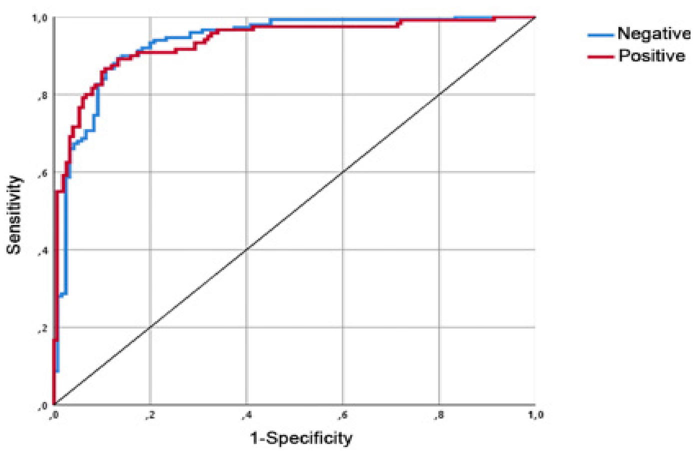

Figure 1. The ROC curve

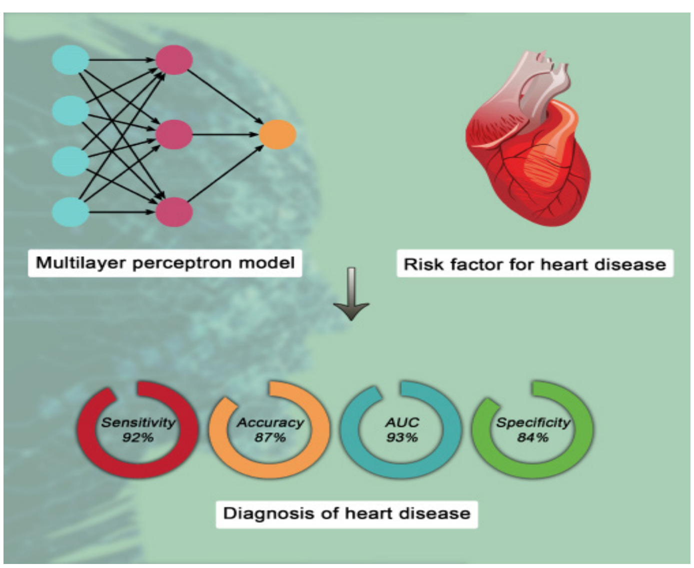

Figure 2. The general structure and performance metric values of MLP for HD

Table 6. The variable importance statistics (In descending order)

Variables	Importance	Importance (%)
Blood pressure	0.155	100.00
Number of vessels fluro	0.141	90.95
Cholesterol	0.109	70.22
Thallium	0.108	69.79
Max heart rate	0.101	65.27
Chest pain type	0.094	60.77
Sex	0.076	48.76
Age	0.052	33.40
ST depression	0.042	27.16
Slope of ST	0.040	25.65
ECG results	0.030	19.03
FBS over 120	0.027	17.49
Exercise angina	0.025	16.37

FBS, fasting blood sugar; ECG, electrocardiogram

DISCUSSION

In this study, HD prediction was made by modeling various risk factors with an artificial neural network model, MLP. Considering similar studies in the literature, HD classification was made using various machine learning models in a study conducted in 2021 (40). Considering the overall accuracy ratios, it was seen that the highest value was obtained with the Random Forest model (84%). This value is considerably lower than the classification performance obtained in the current study.

In another study (41), the individual and ensemble performances of various data mining models were discussed. In related study, in which extensive data preprocessing was performed, the ensemble model (majority vote) formed by combining the support vector machine and logistic regression model gave the highest classification accuracy (91.23%). This ensemble learning approach does not appear to have significantly higher classification performance than the model in the current study. In addition, this ensemble learning model created lacks the knowledge of how predictive variables contribute to the classification performance of the model.

In another study on the prediction of HD (42), the classification performances of Logistic Regression, Decision Forest, Random Forest, Artificial Neural Networks, K-Nearest Neighbors and Support Vector Machines models were compared. According to the results of this study, the Random Forest model correctly classified all cases with 100% accuracy. However, in this study, the classification performance of the model was evaluated only with its overall accuracy. Considering that the Random Forest model is highly prone to overfitting, it is inconvenient to evaluate the classification performance over only one metric. In our study, the performance of the model was evaluated by giving 4 different metrics and their 95% confidence intervals. Thus, the findings were evaluated from both data science and statistics perspectives.

Performance measurement is a significant issue in machine learning. One of the most critical evaluation criteria for checking any classification model's performance is the area under the curve and explains how well the model is at its prediction. The larger the area covered, the better the machine learning models at differentiating the classes given. In addition to the high accuracy of the MLP test used in our study, the AUC value is more excellent than 90%, indicating that the model's predictive value is high. Also, the sensitivity value appears to be greater than the specificity value. This is valuable in the classification of HD in patients, and its high level is of great clinical importance.

In brief, it was found that the three most essential variables in the classification of HDs in the MLP model are blood pressure, the number of major vessels colored by fluoroscopy, and cholesterol. It was also shown to have high accuracy, AUC, sensitivity, and specificity values.

CONCLUSION

The MLP model, which was trained using 13 predictive variables, was found to be successful in the classification of HD. According to the relevant model findings; blood pressure, the number of major vessels colored by fluoroscopy, and cholesterol variables were the three most important HD classification variables. The performance metric values for accuracy, AUC, sensitivity, and specificity were obtained with 87%, 93%, 92%, and 84%, respectively.

Due to the predictive performance in this study, the MLP model can be recommended to clinicians as a clinical decision support system. As future research, for boosting classification performance and obtaining more reliable results, implementing ensemble learning-based models to real dataset(s) having more observations and risk factors should be planned.

Congresses: *This study was awarded with the best poster presentation at the BIO Türkiye International Biotechnology Congress at Istanbul, Turkey, March 5-7, 2020.*

Financial disclosures: *The authors declared that this study hasn't received no financial support.*

Conflict of Interest: *The authors declare that they have no competing interest.*

Ethical approval: Since the public dataset was used in this study, ethics committee approval is not required.

REFERENCES

1. Hose DR, Lawford PV, Huberts W, et a. Cardiovascular models for personalised medicine: Where now and where next? Med Eng Phys. 2019;72:38-48.
2. Rasheed J, Hameed AA, Djeddi C, et al. A machine learning-based framework for diagnosis of COVID-19 from chest X-ray images. Interdiscip Sci. 2021;13:103-17.
3. Zhang R, Guo Z, Sun Y, et al. COVID19 XrayNet: A two-step transfer learning model for the covid-19 detecting problem based on a limited number of chest x-ray images. Interdiscip Sci. 2020;12:555-65.
4. Yu AZ, Ramsey SA. A Computational Systems biology approach for identifying candidate drugs for repositioning for cardiovascular disease. Interdiscip Sci. 2018;10:449-54.
5. Gunata M, Parlakpinar H. A review of myocardial ischaemia/reperfusion injury: Pathophysiology, experimental models, biomarkers, genetics and pharmacological treatment. Cell Biochem Funct. 2021;39:190-217.
6. Wang HD, Naghavi M, Allen C, et al. Global, regional, and national life expectancy, all-cause mortality, and cause-specific mortality for 249 causes of death, 1980-2015: a systematic analysis for the Global Burden of Disease Study 2015. Lancet. 2016;388:1459-544.
7. Mendis S, Puska P, Norrving B, Organization WH. Global atlas on cardiovascular disease prevention and control: World Health Organization; 2011.
8. Benjamin EJ, Muntner P, Alonso A, et al. Heart disease and stroke statistics-2019 update: A report from the American heart association. Circulation. 2019;139:e56-e528.
9. Fryar CD, Chen T-C, Li X. Prevalence of uncontrolled risk factors for cardiovascular disease: United States, 1999-2010: US Department of Health and Human Services, Centers for Disease Control and; 2012.
10. Lopez AD, Mathers CD, Ezzati M, et al. Global and regional burden of disease and risk factors, 2001: systematic analysis of population health data. Lancet. 2006;367:1747-57.
11. Berry JD, Dyer A, Cai X, et al. Lifetime risks of cardiovascular disease. N Engl J Med. 2012;366:321-9.

12. Webber BJ, Seguin PG, Burnett DG, et al. Prevalence of and risk factors for autopsy-determined atherosclerosis among US service members, 2001-2011. Jama-J Am Med Assoc. 2012;308:2577-83.
13. Koton S, Schneider AL, Rosamond WD, et al. Stroke incidence and mortality trends in US communities, 1987 to 2011. Jama-J Am Med Assoc. 2014;312:259-68.
14. George J, Rapsomaniki E, Pujades-Rodriguez M, et al. How does cardiovascular disease first present in women and men? Incidence of 12 cardiovascular diseases in a contemporary cohort of 1 937 360 people. Circulation. 2015;132:1320-8.
15. Ritchey MD, Wall HK, Gillespie C, et al. Million hearts: prevalence of leading cardiovascular disease risk factors--United States, 2005-2012. Mmwr-Morbid Mortal W. 2014;63:462-7.
16. Moran AE, Forouzanfar MH, Roth GA, et al. The global burden of ischemic heart disease in 1990 and 2010: the Global Burden of Disease 2010 study. Circulation. 2014;129:1493-501.
17. Yusuf S, Hawken S, Ounpuu S, et al. Effect of potentially modifiable risk factors associated with myocardial infarction in 52 countries (the INTERHEART study): case-control study. Lancet. 2004;364:937-52.
18. Patel J, Al Rifai M, Scheuner MT, Set a. Basic vs more complex definitions of family history in the prediction of coronary heart disease: the Multi-Ethnic Study of Atherosclerosis. Mayo Clinic Proceedings; 2018: Elsevier.
19. Paixao AR, Berry JD, Neeland IJ, et al. Coronary artery calcification and family history of myocardial infarction in the Dallas heart study. JACC Cardiovasc Imaging. 2014;7:679-86.
20. Ranthe MF, Carstensen L, Oyen N, et al. Family history of premature death and risk of early onset cardiovascular disease. J Am Coll Cardiol. 2012;60:814-21.
21. MacRae CA, Vasan RS. The Future of Genetics and Genomics: Closing the Phenotype Gap in Precision Medicine. Circulation. 2016;133:2634-9.
22. Jan M, Ebert BL, Jaiswal S. Clonal hematopoiesis. Seminars in hematology; 2017: Elsevier.
23. Finegold JA, Asaria P, Francis DP. Mortality from ischaemic heart disease by country, region, and age: statistics from World Health Organisation and United Nations. Int J Cardiol. 2013;168934-45.
24. D'Adamo E, Guardamagna O, Chiarelli F, et al. Atherogenic dyslipidemia and cardiovascular risk factors in obese children. Int J Endocrinol. 2015;2015:912047.
25. Association AH. Understand your risk of heart attack. American Heart Association. 2012.
26. Sullivan MJ, Higginbotham MB, Cobb FR. Exercise training in patients with chronic heart failure delays ventilatory anaerobic threshold and improves submaximal exercise performance. Circulation. 1989;79:324-9.
27. Jackson R, Chamblеss I, Higgins M, et al. Gender differences in ischaemic heart disease mortality and risk factors in 46 communities: an ecologic analysis. Cardiovascular Risk Factors. 1997:43-54.
28. D'agostino RB, Vasan RS, Pencina MJ, et al. General cardiovascular risk profile for use in primary care. Circulation. 2008;117:743-53.
29. Kappert K, Bohm M, Schmieder R, et al. Impact of sex on cardiovascular outcome in patients at high cardiovascular risk: analysis of the Telmisartan Randomized Assessment Study in ACE-Intolerant Subjects With Cardiovascular Disease (TRANSCEND) and the Ongoing Telmisartan Alone and in Combination With Ramipril Global End Point Trial (ONTARGET). Circulation. 2012;126:934-41.
30. Rapsomaniki E, Timmis A, George J, et al. Blood pressure and incidence of twelve cardiovascular diseases: lifetime risks, healthy life-years lost, and age-specific associations in 1.25 million people. Lancet. 2014;383:1899-911.
31. Prescott E, Hippe M, Schnohr P, et al. Smoking and risk of myocardial infarction in women and men: longitudinal population study. BMJ. 1998;316:1043-7.
32. Micha R, Michas G, Mozaffarian D. Unprocessed red and processed meats and risk of coronary artery disease and type 2 diabetes--an updated review of the evidence. Curr Atheroscler Rep. 2012;14:515-24.
33. Franchini M, Mannucci PM. Air pollution and cardiovascular disease. Thromb Res. 2012;129:230-4.
34. Sun Q, Hong X, Wold LE. Cardiovascular effects of ambient particulate air pollution exposure. Circulation. 2010;121:2755-65.
35. Kalantary S, Pourbabaki R, Jahani A, et al. Development of a decision support system tool to predict the pulmonary function using artificial neural network approach. Concurr Comp-Pract E. 2021;33:e6258.
36. Fajardo JE, Lotto FP, Vericat F, et al. Microwave tomography with phaseless data on the calcaneus by means of artificial neural networks. Med Biol Eng Comput. 2020;58:433-42.
37. Vivekanandan T, Sriman Narayana Iyengar NC. Optimal feature selection using a modified differential evolution algorithm and its effectiveness for prediction of heart disease. Comput Biol Med. 2017;90:125-36.
38. Cherian RP, Thomas N, Venkitachalam S. Weight optimized neural network for heart disease prediction using hybrid lion plus particle swarm algorithm. J Biomed Inform. 2020;110:103543.
39. Ledolter J. Data mining and business analytics with R: John Wiley Sons; 2013.
40. Li Z, Tao D, Zheng J, Zhu C. Determining which physical parameters are significant for heart disease. Journal of Physics: Conference Series. 2021:012062.
41. Jaya Sathwika G, Bhattacharya A. Prediction of CardioVascular Disease (CVD) using Ensemble Learning Algorithms. 5th Joint International Conference on Data Science Management of Data (9th ACM IKDD CODS and 27th COMAD); 2022.
42. Mishra S, Neurkar SV, Patil R, Petkar S. Heart Disease Prediction System. International Journal of Engineering and Applied Physics. 2021;1:179-85.