ESKİŞEHİR TECHNICAL UNIVERSITY JOURNAL OF SCIENCE AND TECHNOLOGY
A- APPLIED SCIENCES AND ENGINEERING

2023, 24(2), pp. 107-120, DOI: 10.18038/estubtda.1025092

RESEARCH ARTICLE

CLINICAL DECISION SUPPORT SYSTEM FOR EARLY DIAGNOSIS OF HEART ATTACK USING MACHINE LEARNING METHODS

Burçin KURT [1], İlknur BUÇAN KIRKBİR [1, 2]

[1] Department of Biostatistics and Medical Informatics, Karadeniz Technical University, Trabzon, Turkey
[2] Department of Public Health Nursing, Karadeniz Technical University, Trabzon, Turkey

ABSTRACT

Heart attack which is the main cause of death for both men and women is the leader among deaths due to heart diseases. Therefore, early diagnosis is very important for patients who are having a heart attack. Therefore, the study aimed to develop a clinical decision support system for the diagnosis of a heart attack to help physicians. In the study, variables were obtained accompanied by physicians by statistical analysis methods, where the optimum variables were selected from these variables considering the patient's unconscious state in some cases. Different decision models were developed using probit regression, decision tree, SVM, and ANN methods. As a result, the developed clinical decision support models for heart attack diagnosis were compared and evaluated. Consequently, the best diagnosis model was obtained using ANN with selected variables. In addition to these, the proposed study is significantly noticed with a sensitivity of 98% and specificity of 93.7% for heart attack diagnosis with optimum variables compared to similar studies in the literature. By using the proposed decision support system, it is possible to determine whether a patient has a heart attack or not and help the physician in the process of diagnosis of a heart attack.

Keywords: Heart attack, Machine learning, Clinical decision support system

1. INTRODUCTION

Cardiovascular diseases are the leading cause of death among all causes of death, especially ischemic heart diseases, and cerebrovascular diseases constitute the first two causes of death [1]. In 2012, 38 million deaths out of 56 million des in worldwide were caused by non-contagious diseases, especially heart and vascular diseases, cancer, and chronic airway diseases. In 2012, 46.2% (17.5 million) of non-contagious diseases worldwide were caused by cardiovascular disease. Of these deaths, 7.4 million depend on heart attacks [1]. Cardiovascular diseases are responsible for 37% of those under 70 deaths due to non-contagious diseases. It seems that cardiovascular diseases will continue to be the number one cause of death globally for a long time.

Myocardial infarction (MI) occurs when the heart muscle cells cannot get enough oxygen because of not getting enough blood, and is also called a heart attack. As a result, damage may occur and even death may result if the heart muscle is left without oxygen for a long time. While 50% of deaths due to heart attacks occur in the first hour, this rate rises to 80% in the first 24 hours [2]. Duration of diagnosis and treatment of patients play a big role in deaths which are from heart attacks. Computer programs or machine learning techniques can be used to reduce the mortality rate, improve the accuracy of disease diagnosis and mainly reduce the diagnosis time. Therefore, it aimed to develop a clinical decision support system to help physicians for prediction of a heart attack, in the study.

Doğan et al. developed a system for the diagnosis of heart attack using the decision tree method with different biochemical variables [3]. LDH, CK, CKMB, AST, and ALT enzymes were used as input to predict MI+ or MI-, and the proposed system has been evaluated on 61 patients. The developed system has performed 100% success on the patient data of 50 heart attacks and 11 non-diagnosed heart attacks

*Corresponding Author: burcinnkurt@gmail.com
Received: 17.11.2021 Published: 21.06.2023

diagnosed by physicians. In another study conducted in the literature [4], the clinical symptoms; myoglobin, mass concentration, CG, creatine kinase MB activity, creatinine kinase, and cardiac troponin T values were compared and used with the decision tree method for heart attack diagnosis. 91% sensitivity and 90% specificity values were obtained on 69 test dataset for the heart attack diagnosis system which was developed the using decision tree method.

In a study conducted by Dangare et al. in 2012, a model was developed that can predict the risk of heart disease by using artificial neural networks and data mining methods. A Heart Disease Prediction system (HDPS) was developed using a neural network classifying as "has heart disease" and "has no heart disease". The HDPS system used 13 variables such as sex, cholesterol, and blood pressure to predict the likelihood of a patient getting heart disease. Nearly 100% success was obtained on 270 test data [5].In a similar study in 2011, two different models were developed that can diagnose heart disease using radial-based (RBF) artificial neural networks (ANN) and support vector machine (SVM) methods. The dataset has 214 records with 19 variables and the outcome values are 0-Myalgia, 1-Myocardial Infarction (MI), 2- Ischemic Heart Disease (IHand D), and 3- Unstable Angina (UA). 84.66% sensitivity and 88.5%specificity values were obtained for the ANN model on the 214 test data, and 82.4% sensitivity and 82.10% were obtained ned for the SVM model [6].

In this study, we have developed clinical decision support models for early diagnosis of heart attack using probit regression and machine learning methods which are decision tree, SVM, and ANN with biochemical, ECG, and demographic variables which are given in the below section. Furthermore, we have compared the performance of models using statistical scales and selected the best model for clinical decision support. Using the selected model with these variables for a patient, it can be diagnosed as a heart attack or not.

2. METHODS AND MATERIALS

2.1. Dataset

In the study, the data of 350 patients who came to Karadeniz Technical University Faculty of Medicine Farabi Hospital Emergency Medicine Service with chest pain between the years September 2013 and April 2016, with or without a heart attack diagnosis were used. The data were obtained retrospectively with the ethical approval dated 09 May 2016 and numbered 2016/45, which was given in Figure A1 in Appendix. The parameters in the data set were obtained by examining documents such as laboratory test results, epicrisis reports, and angiography results under the supervision of a specialist physician. Conditions with false positive results for CK-MB and troponin (polymyositis/dermatomyositis (inflammation of the muscles), muscular dystrophies (muscle disease), chronic renal failure, and chronic hemodialysis patients, patients who have received intramuscular (intramuscular) injections in the last 24 hours, Patients who had trauma or skeletal muscle damage during the day, patients with a hemolytic blood disease, and patients with shock were excluded from the study.

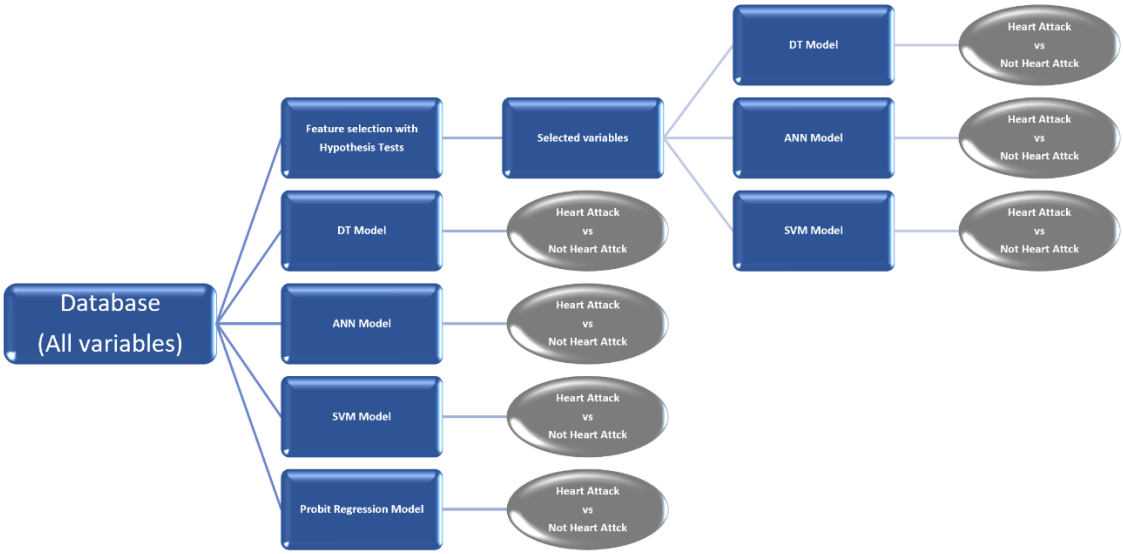
The patients were divided into two groups those with a heart attack (experimental group) and without a heart attack (control group). The diagnosis of heart attack was made according to the World Health Organization (WHO) criteria. 192 and 158 of them were diagnosed with heart attack and not heart attack respectively. The data was collected during a heart attack episode. The variables proposed by the physicians to diagnose a heart attack are given in Table 1.

The descriptive statistics of the categorical and numerical variables for patients diagnosed with heart attack and not heart attack are given in Appendix Table A1 and A2 respectively (using R software). The analyzes were implemented using open-source R 3.3. and 3.6 versions programming language ("C50", "neuralnetwork", "e1071", "caret", "pROC", "ggplot2", and "stats" packages).

Table 1. The variables for heart attack diagnosis (proposed by the physicians)

	Variable	Description
1	Sex	Male/Female
2	Ecg Change	Yes/No
3	St Segment Change	Yes/No
4	Chronical Disease	Yes/No
5	Heart Disease	Yes/No
6	Patient Pedigree	Yes/No
7	Ck-Mb (Creatine Kinase)	Iu/L
8	Hs Troponin (High Sensitivity Troponin)	Ng/Ml

The flow diagram of the study is given in Figure 1.



DT: Decision Tree, SVM: Support Vector Machine, ANN: Artificial Neural Networks

Figure 1. The flow diagram of the study

2.2. Probit regression (Probit model)

Probit regression is used to model dichotomous or binary dependent variables whose distribution is assumed to be a proxy for a true underlying continuous normal distribution [7]. It is a binary classification model which classifies samples according to their predicted probabilities for each class.

The probit model uses a similar approach to logistic regression and is also a popular method for an ordinal or a binary response model. It has a probit link function which uses the inverse of the cumulative distribution function of the standard normal distribution to transform probabilities to the standard normal variable and is most often estimated using the standard maximum likelihood procedure, such an estimation being called a probit regression [8]. Thus,

$$\Phi^{-1}(\pi_i) = x_i\beta + \varepsilon_i \qquad (1)$$

where

$$\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} dt \quad , \qquad (2)$$

Φ, x_i and β are the cumulative distribution function of the standard normal distribution, the ith row of the X matrix (n × p data matrix), that i, the ith record in the dataset and $\beta = \left(\beta_1, \beta_2, \ldots, \beta_p\right)^T$ respectively.

We implemented probit regression for our dataset using R open-source software and obtained the output as given in Table 2. The coefficients, z-statistic (sometimes called a Wald z-statistic), standard errors, and the associated p-values are presented as an output. As seen in Table 2; Sex(Male), Hs.Troponin, ECG.Change(No), Chronical.Disease(No) and Patient.Pedigree(No) is statistically significant. The probit regression coefficients show the change in the z-score or probit index for a one-unit change in the predictor. For example, for a one-unit increase in Hs.Troponin, the z-score increases by 4.742e-04. Furthermore, we obtained confidence intervals for the coefficient estimates, created by profiling the likelihood function [9, 10].

Table 2. The statistical output of the probit regression model for variables

	Estimate	**Std. Error**	**z Value**	**Pr(<|z|)**	**2.5%**	**97.5%**
(Intercept)	0.006	137.700	0.045	0.964	1.011	55.400
Sex.Male	-0.761	0.197	-3.874	***<0.001***	-1.161	-0.376
CK-MB	<-0.001	<0.001	-0.242	0.809	<-0.001	<0.001
Hs.Troponin	<0.001	<0.001	2.423	***0.015***	<0.001	<0.001
ECG Change.No	-1.271	0.178	-7.127	***<0.001***	-1.624	-0.925
ST.Segment Change.No	-4.880	-137.700	-0.035	0.972	-58.429	-0.492
Chronical.Disease.No	-0.620	0.190	-3.264	***0.001***	-0.992	-0.252
Heart.Disease.No	0.352	0.202	1.740	0.082	-0.037	0.745
Patient.Pedigree.No	-0.539	0.249	-2.162	***0.031***	-1.036	-0.058

* z value is the ratio of the *Estimate* to the *Std. Error [11]*

Goodness-of-fit (GOF) measure indicates the fitness of the data to the regression model and there are also many alternative metrics such as measures based on the variance decomposition of the predicted probabilities, measures based on the predicted probabilities, and log-likelihood-based measures. These pseudo-R2 metrics are used as the GOF measures for binary regression models [12]. Furthermore, the pseudo-R2 of McFadden measure uses the two log-likelihood values suggested by Aldrich–Nelson [13] and takes a value between 0 and 1. The obtained Pseudo-R2 of McFadden value is 0.411 for the developed probit model, which can be evaluated as a good model fit [14, 15] and the prediction results are given in the third section. Furthermore, the predicted probabilities of heart attack for statistically significant variables in the probit model are given in Appendix Figure A2.

2.3. Feature Selection and Classification with Machine Learning Methods

In the first step, the feature selection process was implemented using statistical analysis tests and considering some cases. In the statistical analyses for the study, the α of 0.05 was used as the cut-off for significance. If the P value is less than 0.05, we reject the null hypothesis, which means that there is a difference between the means, and decide that a significant difference does exist. Then, in the second step, diagnosis models were developed using three machine learning methods, for the selected variables.

2.3.1. Feature selection using statistical analysis

For the feature selection process, the normal distribution of quantitative data was tested using the Kolmogorov–Smirnov test, and the Mann-Whitney U test was performed to determine whether the values of the variables HS Troponin and CK-MB were significantly different from the patients was who diagnosed with heart attack and not heart attack groups. The results of the Mann-Whitney U test are given in Table 3.

Table 3. The Mann Whitney U test analysis results for HS Troponin and CK-MB variables of patients who were diagnosed with heart attack and not heart attack

HS-Troponin	N	Mean rank (SO)	Sum of ranks (ST)	U	z	p
Heart Attack	192	230.91	44334.5	4529.5	-11.296	**<0.001**
Not Heart Attack	158	108.17	17090.5			
Total	350					
CK-MB	**N**	**Mean rank (SO)**	**Sum of ranks (ST)**	**U**	**z**	**p**
Heart Attack	158	107.48	16982	4421	-11.409	**<0.001**
Not Heart Attack	192	231.47	44443			
Total	350					

As seen in Table 3, p values for HS Troponin ($p<0.001$) and CK-MB ($p<0.001$) were smaller than 0.05, which means that these values were significantly different between heart attack and not heart attack groups. Furthermore, categorical variables were analyzed using the Chi-square test for the feature selection process. According to the results, the categorical variables, which are sex, ECG change, ST segment change, chronical disease, and patient pedigree were obtained significantly different from the patients who were diagnosed with heart attack and not heart attack except heart disease. The Chi-square test results are given in Table 4.

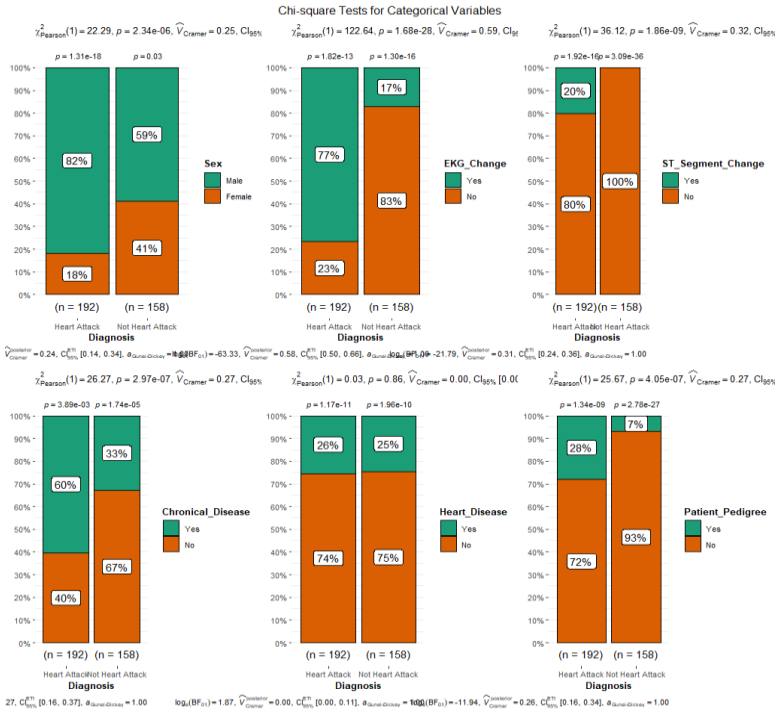
Table 4. The Chi-square test analysis results for categorical variables of patients who were diagnosed with heart attack and not heart attack

Variable	Chi-square	p
Sex	21,18	**<0.001**
ECG change	118,16	**<0.001**
ST segment change	34,09	**<0.001**
Chronic Disease	23,78	**<0.001**
Patient Pedigree	24,28	**<0.001**
Heart Disease	0,005	0.95

*p <0,05

The test gives a Chi-squared statistic, which is a prediction of the goodness of fit of one category relative to the other and can be observed from the frequency of a variable with the expected frequency. The Chi-square test can estimate the role of the random effects in the results and gives a P value which is the probability that the samples have come from the same population. According to the statistical analysis results for the feature selection process; HS Troponin, CK-MB, sex, ECG change, ST segment change, chronical disease, and patient pedigree variables have been selected due to the significant difference. However, chronical disease, heart disease, and patient pedigree information cannot be obtained considering the patient's unconscious state in some cases. Therefore, two different decision system models have been developed and compared. One of them used all 8 variables while the other one used 5 variables except chronical disease, heart disease, and patient pedigree. The distributions of numerical and categorical variables are given in Figure 2.

(a)



(b)

Figure 2. The distribution of (a) numerical and (b) categorical variables of patients who were diagnosed with heart attack and not a heart attack

As seen in Figure 2, the mean, standard deviation, and range values of Hs-Troponin and CK-MB variables for the "heart attack group" seems higher than the "not heart attack group". Furthermore, for categorical variables, only the "Heart Disease" variable is not significantly different between the "heart attack and "not heart attack groups due to the close distribution of the "Heart disease/yes" and "Heart disease/no" frequencies according to the dependent variable.

2.3.2. Classification using machine learning methods

For the second step, developing diagnosis models train and test sets were created. The distribution of train and test sets is given in Table 5.

Table 5. The distribution of train and test sets

	Train Set	Test Set
Heart Attack	95	97
Not Heart Attack	95	63
Total	190	160

192 and 158 patients of the dataset were diagnosed with heart attack and not heart attack respectively. To balance the patient number of each class in the training set, 60% of patients with no heart attack (as it is a minor class), which was 95 patients, and 95 patients with a heart attack (as it is equal to the number of patients with no heart attack), have been used for training. Therefore, in the development of the models, 190 of the 350 data were used as a training set and the remaining 160 were used as the test set for validation. With these train and test sets, three prediction models were developed using SVM [14,16-19], decision tree [20-22], and ANN [23, 24] methods which give successful results for clinical decision support systems in the literature.

3. RESULTS

In this study, heart attack, which is one of the serious diseases of today was handled. For this purpose, decision models were developed using probit regression, SVM, decision tree, and ANN. Furthermore, a feature selection process has been implemented for machine learning methods, whereas the probit regression method has done that by itself. Therefore, two approaches were used for each SVM, ANN, and decision tree model using selected and all features. The optimum parameter values and comparison of performance results of all these approaches are given in Table 6 and Table 7, respectively.

Table 6. The optimum parameter values for the developed machine learning models

	SVM (Sigmoid)	SVM (Radial)	SVM (Linear)	SVM (Polynomial)	DT	ANN
Optimum Parameter Values	Cost=1 Gamma=0.2 Coef.0=0	Cost=1 Gamma=0.17	Cost=1 Gamma=0.2	Cost=1 Gamma=0.2 Degree=3 Coef.0=0	Tree size=6	Hidden layer number=2 Hidden_layer1_neuronnumber=5 Hidden_layer2_neuronnumber=2 Epoch=1000 Threshold=0.01 Activation function=logistic

Table 7. The performance results of developed diagnosis models

Machine Learning Based Models			
With all 8 parameters	**SVM (Radial)**	**DT**	**ANN**
Sensitivity (%)	83.5	93.8	100
Specificity (%)	85.4	93.7	4.2
Accuracy (%)	84	93.8	68.3
With selected 5 parameters	**SVM (Radial)**	**DT**	**ANN**
Sensitivity (%)	79	91.7	**98**
Specificity (%)	83	97.9	**93.7**
Accuracy (%)	81	93.8	**96.5**
Probit Regression Model			
Sensitivity (%)	81.3		
Specificity (%)	81.7		
Accuracy (%)	81.4		

For the developed SVM models, four different kernel functions were applied. Although the performance results of these kernels were almost equal, the best SVM model was obtained with a radial-based kernel

function. As seen in Table 7, the best results were obtained with ANN (MLP) using selected 5 variables considering the patient's unconscious state in some cases. Therefore, it is possible to determine whether automatically diagnosed with a heart attack or not, by using the optimum 5 variables of a patient which are ECG change, ST segment change, gender, HS-troponin, and CK-MB. Furthermore, the ROC curve performance comparison of the machine learning-based heart attack decision models using selected variables is given in Figure 3.

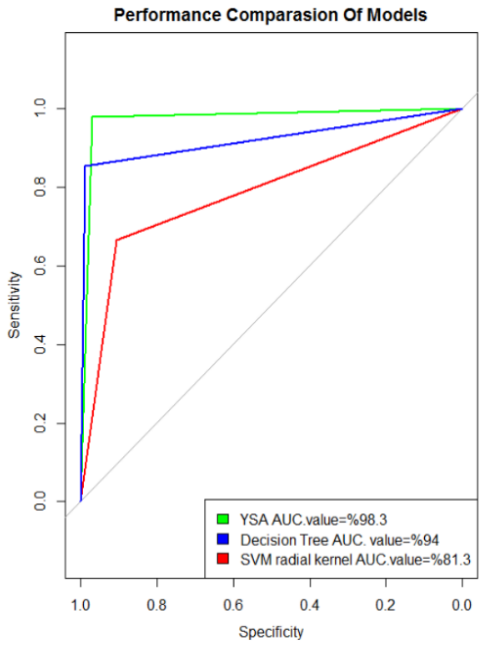


Figure 3. ROC curves of developed diagnosis models

4. DISCUSSION

According to the statistical analysis results for the feature selection process; HS Troponin, CK-MB, sex, ECG change, ST segment change, chronical disease, and patient pedigree variables have been selected due to the significant difference. However, chronical disease, heart disease, and patient pedigree information cannot be obtained considering the patient's unconscious state in some cases. Therefore, two different decision system models were developed and compared one of which used all 8 variables, while the other one used 5 variables except chronical disease, heart disease, and patient pedigree. As a result, a satisfactory successful decision support model has been developed for heart attack diagnosis using the optimum 5 variables.

For the machine learning-based models, the selected variables were obtained using statistical analysis methods and optimum variables were selected from these variables considering the patient's unconscious state in some cases. On the other hand, for the probit model, variables have been determined by the regression approach through model development. When compared to similar studies in the literature, this study is stand out by using a different approach based on probit regression and comparison with machine learning methods. In this study, classification models were implemented for heart attack decisions using probit regression, SVM, ANN, and decision tree methods. Consequently, the best decision support model was obtained using ANN with selected variables. In addition to these, the proposed study is significantly noticed with the high number of test data for heart attack classification with a sensitivity of 98% and specificity of 93.7% compared to similar studies in the literature, which can be seen in Table 8.

Table 7. Comparison of the proposed study with similar studies in the literature

Literature Studies	Method	Sens. (%)	Spec. (%)	Acc. (%)	Patients (Test Samples)	Data set Information	Variables
Doğan et al. (2007) [3]	Decision Tree (Heart attack)	100	100	100	50MI(+), 11 MI(-)	Not given in the paper	CK, CKMB, LDH, AST and ALT
Mair et al. (1995) [4]	Decision Tree (Heart attack)	91	90	90	45 MI(+), 49 unstable angina pectoris, 20 chest pain caused by other diseases	Emergency room of a Department of Internal Medicine (University Hospital)	ECG, creatine kinase, creatine kinase MB activity and mass concentration, myoglobin, and cardiac troponin T
Ghumbre et al. (2011) [6]	RBF network, SVM (Heart disease)	82.4 84.06	82.1 88.5	82.24 85.05	78 Myalgia (Normal), 139 Heart disease	Not given in the paper	Chest pain types (Left or Right side), Arm pain, Backache, Sweating, Breathlessness, Addiction, Diabetic, MAP, Pulse rate, ECG: ST elevation, ST depression, T elevation, T depression, Q waves, BSL, CKMB test
Soni et al. (2011) [25]	Decision Tree, Naive Bayes, ANN (Heart disease)	-	-	89 86.53 85.53	454	Machine Learning Repository of UCI	Sex, Chest Pain Type, Fasting Blood Sugar, Restecg, Exang, Slope, CA, Thal, Trest Blood Pressure, Serum Cholesterol, Thalach, Oldpeak, Age, Heart disease
Hazra et al. (2017) [26] Review study	RBF network **(Hannan et al.)** (Heart disease)	-	-	90-97	75	Sahara Hospital, Aurangabad	Age, P1(Previous History), P2(Present History), P3(personnel History), P4(Physical Examination), CVS(Cardio Vascular System), RS(Respiratory System), PA(Per Abdomen), CNS (Central Nervous System), ECG (Electrocardiography) and BI (Blood Investigation)
	Learning Vector Quantization (LVQ) **(Chen et al.)** (Heart disease)	85	70	80	150	Machine Learning Repository of UCI	Age, sex, chest pain type, trestbps, cholesterol, fasting blood sugar, resting ecg, max heart rate, exercise induced angina, old peak, slope, number of vessels colored, and thal
Mujawar et al. (2015) [27]	Naive Bayes& Modified K-Means (Heart disease)	93	89	91	100 Heart disease 100 Not heart disease	Cleveland Heart Disease database	Sex, Chest Pain Type, Fasting Blood Sugar, Restecg, Exang, Slope, CA , Thal, Trest Blood Pressure, Serum Cholesterol, Thalach, Oldpeak, Age, Heart disease
Aydın et al. (2016) [28]	Bagging AdaBoostM1 Random Forest Naive Bayes RBF Network IBK NNge (Heart disease)	11.1 22.2 33.3 44.4 33.3 44.4 33.3	-	67.5 67.5 72.5 77.5 82.5 72.5 70	40 40	Long Beach VA Hospital Long Beach VA Hospital	Age, Sex, Chest Pain Type, Resting Blood Pressure, Serum Cholesterol, Fasting Blood Sugar, Resting Electrocardiographic Results, Maximum Heart Rate, Exercise induced angina, ST depression induced by exercise relative to rest,The Slope of The Peak Exercise St Segment, Number of Major Vessels, Defect Type, Heart disease
Florence et al. (2014) [29]	Decision Tree Neural Network (Heart attack)	98	53	-	76	Machine Learning Repository of UCI	Sex, Age, Cardiac Duration, Cholesterol, Signal Leval, Posibility of Attack (Yes/No)
Proposed Study	ANN SVM Decision Tree	**98** 79 91.7	**93.7** 83 97.9	**96.5** 81 93.8	145 97 MI(+), 48 MI(-)	Karadeniz Technical University Farabi Hospital	Sex, ECG change, ST Segment change, CK-MB (creatine kinase), HS Troponin (high sensitivity troponin)
	Probit Regression	81.3	81.7	81.4	-	Karadeniz Technical University Farabi Hospital	Sex, ECG change, HS Troponin (high sensitivity troponin), Chronical disease, Patient Pedigree

By using the proposed heart attack decision support model, it is aimed to reduce the number of repeated laboratory tests and ECG measurements to assist the physician in the process of deciding the probability of having a heart attack in patients who apply to the emergency department with chest pain and to make

a definitive diagnosis. In addition to these, the proposed model can be used as a pilot decision support system in clinics after taking the required permissions. So, it can be developed and widened based on the evaluation results.

5. CONCLUSION

The proposed heart attack decision support system can be used as software by entering required variables or it can be integrated with the patient's tracking system in the hospital. By using the developed heart attack decision support system, it is possible to help the physician in the process of diagnosis of heart attack for patients who apply to the emergency service with chest pain compliment. Furthermore, it is aimed to reduce the number of repeated laboratory tests and ECG measurements so that a definite diagnosis can be made. Therefore, the proposed prediction model could assist in the diagnosis of MI quickly. In addition to these, if the proposed variables can be obtained with the help of portable devices in the future, the patient will be pre-diagnosed with a heart attack while being ambulance, and able to intervene first without wasting time. This will help to reduce the mortality due to heart attacks.

ACKNOWLEDGEMENTS

All necessary permissions such as ethics committee, hospital permit were obtained for the conducted study. No funding to declare.

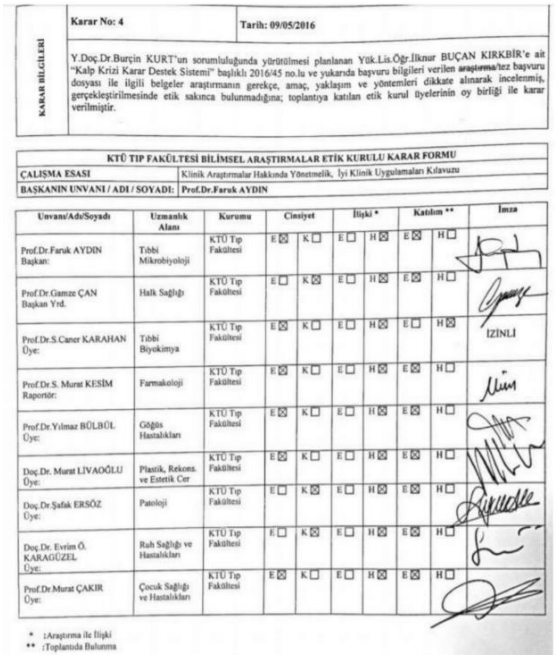
CONFLICT OF INTEREST

The authors confirmed that there are no conflicts of interest regarding the publication of this article.

AUTHORSHIP CONTRIBUTIONS

Burçin Kurt; Contributed data or analysis tools; Performed the analysis; Wrote the paper.
İlknur Buçan Kırkbir; Collected the data; Performed the analysis.

APPENDIX



KARAR BİLGİLERİ

Karar No: 4 — Tarih: 09/05/2016

Y.Doç.Dr.Burçin KURT'un sorumluluğunda yürütülmesi planlanan Yük.Lis.Öğr.İlknur BUÇAN KIRKBİR'e ait "Kalp Krizi Karar Destek Sistemi" başlıklı 2016/45 no.lu ve yukarıda başvuru bilgileri verilen araştırma/tez başvuru dosyası ile ilgili belgeler araştırmanın gerekçe, amaç, yaklaşım ve yöntemleri dikkate alınarak incelenmiş, gerçekleştirilmesinde etik sakınca bulunmadığına; toplantıya katılan etik kurul üyelerinin oy birliği ile karar verilmiştir.

KTÜ TIP FAKÜLTESİ BİLİMSEL ARAŞTIRMALAR ETİK KURULU KARAR FORMU

ÇALIŞMA ESASI	Klinik Araştırmalar Hakkında Yönetmelik, İyi Klinik Uygulamaları Kılavuzu
BAŞKANIN UNVANI / ADI / SOYADI:	Prof.Dr.Faruk AYDIN

Unvanı/Adı/Soyadı	Uzmanlık Alanı	Kurumu	Cinsiyet		İlişki *		Katılım **		İmza
Prof.Dr.Faruk AYDIN Başkan:	Tıbbi Mikrobiyoloji	KTÜ Tıp Fakültesi	E☒	K☐	E☐	H☒	E☒	H☐	
Prof.Dr.Gamze ÇAN Başkan Yrd.	Halk Sağlığı	KTÜ Tıp Fakültesi	E☐	K☒	E☐	H☒	E☒	H☐	
Prof.Dr.S.Caner KARAHAN Üye:	Tıbbi Biyokimya	KTÜ Tıp Fakültesi	E☒	K☐	E☐	H☒	E☐	H☒	İZİNLİ
Prof.Dr.S. Murat KESİM Raportör:	Farmakoloji	KTÜ Tıp Fakültesi	E☒	K☐	E☐	H☒	E☒	H☐	
Prof.Dr.Yılmaz BÜLBÜL Üye:	Göğüs Hastalıkları	KTÜ Tıp Fakültesi	E☒	K☐	E☐	H☒	E☒	H☐	
Doç.Dr. Murat LİVAOĞLU Üye:	Plastik, Rekons. ve Estetik Cer	KTÜ Tıp Fakültesi	E☒	K☐	E☐	H☒	E☒	H☐	
Doç.Dr.Şafak ERSÖZ Üye:	Patoloji	KTÜ Tıp Fakültesi	E☐	K☒	E☐	H☒	E☒	H☐	
Doç.Dr. Evrim Ö. KARAGÜZEL Üye:	Ruh Sağlığı ve Hastalıkları	KTÜ Tıp Fakültesi	E☐	K☒	E☐	H☒	E☒	H☐	
Prof.Dr.Murat ÇAKIR Üye:	Çocuk Sağlığı ve Hastalıkları	KTÜ Tıp Fakültesi	E☒	K☐	E☐	H☒	E☒	H☐	

* :Araştırma ile İlişki
** :Toplantıda Bulunma

Figure A1. The ethical approval

Table A1. The descriptive statistics of categorical variables for each group

Variable	Valid	Frequency	Percent	CumPercent
Sex	Female	35	18.23	18.23
Sex	Male	157	81.77	100.00
Sex	TOTAL	192	NA	NA
EKG_Change	No	45	23.44	23.44
EKG_Change	Yes	147	76.56	100.00
EKG_Change	TOTAL	192	NA	NA
ST_Segment_Change	No	153	79.69	79.69
ST_Segment_Change	Yes	39	20.31	100.00
ST_Segment_Change	TOTAL	192	NA	NA
Chronical_Disease	No	76	39.58	39.58
Chronical_Disease	Yes	116	60.42	100.00
Chronical_Disease	TOTAL	192	NA	NA
Heart_Disease	No	143	74.48	74.48
Heart_Disease	Yes	49	25.52	100.00
Heart_Disease	TOTAL	192	NA	NA
Patient_Pedigree	No	138	71.88	71.88
Patient_Pedigree	Yes	54	28.12	100.00
Patient_Pedigree	TOTAL	192	NA	NA

(a) Heart Attack group

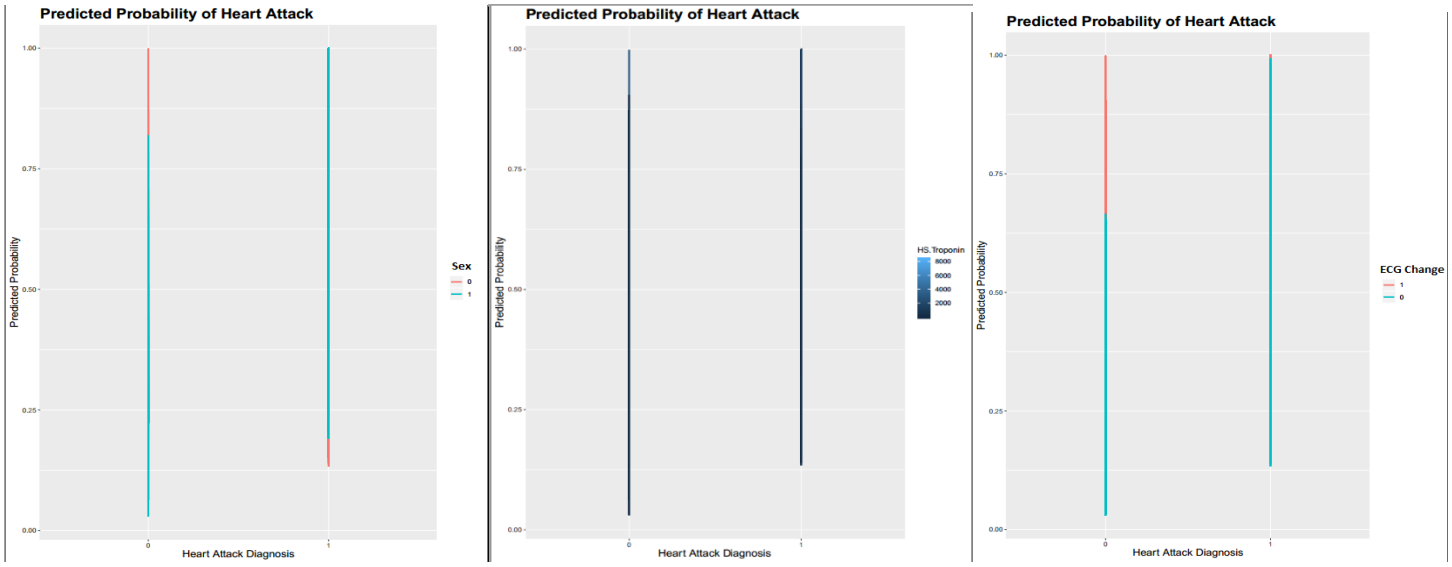
Variable	Valid	Frequency	Percent	CumPercent
Sex	Female	65	41.14	41.14
Sex	Male	93	58.86	100.00
Sex	TOTAL	158	NA	NA
EKG_Change	No	131	82.91	82.91
EKG_Change	Yes	27	17.09	100.00
EKG_Change	TOTAL	158	NA	NA
Chronical_Disease	No	106	67.09	67.09
Chronical_Disease	Yes	52	32.91	100.00
Chronical_Disease	TOTAL	158	NA	NA
Heart_Disease	No	119	75.32	75.32
Heart_Disease	Yes	39	24.68	100.00
Heart_Disease	TOTAL	158	NA	NA
Patient_Pedigree	No	147	93.04	93.04
Patient_Pedigree	Yes	11	6.96	100.00
Patient_Pedigree	TOTAL	158	NA	NA

b) Not Heart Attack group

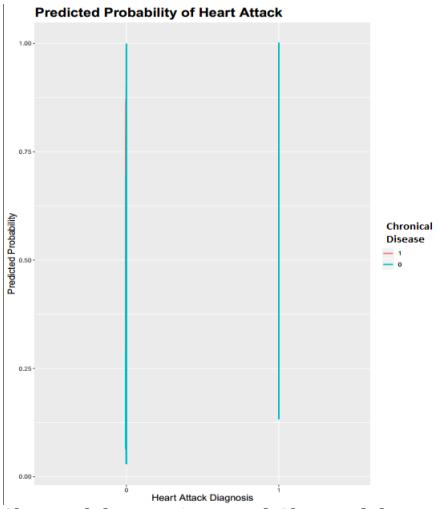
Table A2. The descriptive statistics of numerical variables for patients for each group

	Heart Attack.CK.MB	Heart Attack.HS.Troponin	Not Heart Attack.CK.MB	Not Heart Attack.HS.Troponin
nbr.val	192.00	192.00	158.00	158.00
nbr.null	0.00	0.00	0.00	0.00
nbr.na	0.00	0.00	0.00	0.00
min	1.03	0.01	0.30	0.01
max	1570.00	8315.00	4675.00	3985.00
range	1568.97	8314.99	4674.70	3984.99
sum	14762.69	134426.97	5487.77	11933.29
median	14.80	153.80	2.16	6.46
mean	76.89	700.14	34.73	75.53
SE.mean	13.28	97.95	29.57	27.71
CI.mean.0.95	26.20	193.20	58.40	54.74
var	33877.62	1842101.00	138110.20	121347.51
std.dev	184.06	1357.24	371.63	348.35
coef.var	2.39	1.94	10.70	4.61
skewness	5.68	3.09	12.32	9.26
skew.2SE	16.20	8.81	31.91	23.99
kurtosis	38.41	10.15	150.84	98.08
kurt.2SE	55.01	14.53	196.51	127.78
normtest.W	0.40	0.55	0.06	0.20
normtest.p	0.00	0.00	0.00	0.00

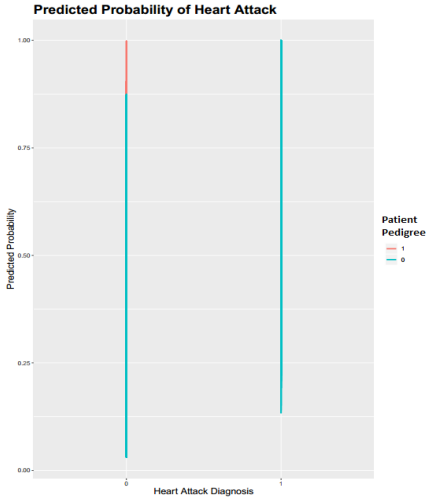
*nbr.val : Number of instances, nbr.null: Number of missing values, nbr.na: Number of NA value



(a)Sex=0, Male and Sex=1, Female (b) Hs. Troponin (c) ECG.change=0,No and ECG.change=1, Yes



(d) Chronical.disease=0, No and Chronical.disease =1, Yes



(e) Patient.pedigree=0, No and Patient.pedigree =1, Yes

Figure A2. The predicted probabilities of heart attack for statistically significant variables in the developed probit model (Heart attack diagnosis=0, No and Heart attack diagnosis =1, Yes)

REFERENCES

[1] Türkiye cardiovascular disease prevention and control program action plan. Ministry of Health, Public Health Institution of Türkiye, 988, 2015.

[2] Storrow AB, Gibler WB. Chest pain centers: diagnosis of acute coronary syndromes. Ann Emerg Med 2000; 35: 449-461.

[3] Doğan Ş, Türkoğlu İ, Yavuzkır M. Heart attack detection from cardiac by using decision trees. eJournal of New World Sciences Academy 2007; 2: 39-50.

[4] Mair J, Smidt J, Lechleitner P, Dienstl F, Puschendorf B. A decision tree for the early diagnosis of acute myocardial infarction in nontraumatic chest pain patients at hospital admission. The Journal of Emergency Medicine 1995; 14: 1502-1509.

[5] Dangare CS, Apte SS. A data mining approach for prediction of heart disease using neural networks. International Journal of Computer Engineering and Technology 2012; 3: 30-40.

[6] Ghumbre S, Patil C, Ghatol A. Heart disease diagnosis using support vector machine. International Conference on Computer Science and Information Technology (ICCSIT'2011) 17-18 December 2011; Pattaya, Thailand, 84-88.

[7] Garson GD. Probit Regression and Response Models. Statistical Associates Publishers, 2013.

[8] Razzaghi M. The probit link function in generalized linear models for data mining applications. Journal of Modern Applied Statistical Methods 2013; 12(1): 164-169.

[9] Hosmer D, Lemeshow S. Applied Logistic Regression (Second Edition). New York: John Wiley & Sons, Inc, 2000.

[10] Long JS. Regression Models for Categorical and Limited Dependent Variables. Thousand Oaks, CA: Sage Publications, 1997.

[11] Smithson M, Merkle EC. Generalized Linear Models for Categorical and Continuous Limited Dependent Variables. New York, USA: CRC Press, 2013.

[12] Yazıcı B, Alpu Ö, Yang Y. Comparison of goodness-of-fit measures in probit regression model. Communications in Statistics-Simulation and Computation 2007; 36: 1061–1073.

[13] Aldrich J.H., Nelson F.D.. Probability, logit, and probit models (Quantitative Applications in the Social Sciences). California, US: Sage Publications, 1984.

[14] McFadden D. Quantitative methods for analyzing travel behavior of individuals: some recent developments. London, England: D. Hensher and P. Stopher (eds.), Behavioural Travel Modeling, 1978.

[15] Veall MR., Zimmerman KF. Evaluating pseudo-R2's for binary probit models. Quality and Quantity 1994; 28: 151–164.

[16] Güner N, Çomak E. Predicting the success of engineering students in mathematics lessons using support vector machines. Pamukkale University Journal of Engineering Sciences 2011; 2: 87-96.

[17] Kavzoğlu T, Çölkesen İ. Investigation of the effects of kernel functions on the classification of satellite images with support vector machines. Map Journal 2010; 144: 73-82.

[18] Yılmaz Akşehirli Ö, Ankaralı H, Aydın D, Saraçlı Ö. An alternative approach to medical prediction: Support vector machines. Türkiye Clinics Journal of Biostatistics 2013; 1: 19-28.

[19] Ma Y, Guo G. Support Vector Machines Applications. MN, USA: Springer, 2014.

[20] Berry MJA, Linoff GS. Mastering Data Mining: The Art and Science of Customer Relationship. 1st ed. New York, US: Wiley Computer Publishing, 2000.

[21] Balaban ME, Kartal E. Data mining and machine learning basic algorithms and applications with R language. Istanbul, Türkiye: Caglayan Bookstore, 2015.

[22] Çalış A, Kayapınar S, Çetinyokuş T. An application on computer and internet security with decision tree algorithms in data mining. Journal of Industrial Engineering 2014; 2-19.

[23] Öztemel E. Artificial neural networks. Third edition. İstanbul, Türkiye: Papatya Publishing Education Inc. 2012.

[24] Gönül Y, Ulu Ş, Bucak A. Artificial neural networks and their use in clinical research. Journal of General Medicine 2015; 25: 104-111.

[25] Soni J, Ansari U, Sharma D, Soni S. Predictive data mining for medical diagnosis: An overview of heart disease prediction. International Journal of Computer Applications 2011; 17: 43-48.

[26] Hazra A, Mandal SK, Gupta A, Mukherjee A and Mukherjee A. Heart disease diagnosis and prediction using machine learning and data mining techniques: a review. Advances in Computational Sciences and Technology 2017; 10 (7): 2137-2159.

[27] Mujawar SH, Devale PR. Prediction of heart disease using modified k-means and by using naive bayes. International Journal of Innovative Research in Computer and Communication Engineering 2015; 3(10): 10265-10273.

[28] Aydın S, Ahanpanjeh M, Mohabbatiyan S. Comparison and evaluation of data mining techniques in the diagnosis of heart disease. International Journal on Computational Science & Applications (IJCSA) 2016; 6(1): 1-15.

[29] Florence S, Bhuvaneswari Amma NG, Annapoorani G and Malathi K, "Predicting The Risk of Heart Attacks using Neural Network and Decision Tree", International Journal of Innovative Research In Computer and Communication Engineering, ISSN (Online): 2320-9801, November 2014; Vol. 2, Issue 11, pp. 7025-7028.