RESEARCH ARTICLE

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

Burçin KURT 1, İlknur BUÇAN KIRKBİR 1,2

1Department of Biostatistics and Medical Informatics, Karadeniz Technical University, Trabzon, Turkey
2Department 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 deaths 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: burcin.kurt@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 using the 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)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Sex</td>
</tr>
<tr>
<td>2</td>
<td>Ecg Change</td>
</tr>
<tr>
<td>3</td>
<td>St Segment Change</td>
</tr>
<tr>
<td>4</td>
<td>Chronical Disease</td>
</tr>
<tr>
<td>5</td>
<td>Heart Disease</td>
</tr>
<tr>
<td>6</td>
<td>Patient Pedigree</td>
</tr>
<tr>
<td>7</td>
<td>Ck-Mb (Creatine Kinase)</td>
</tr>
<tr>
<td>8</td>
<td>Hs Troponin (High Sensitivity)</td>
</tr>
</tbody>
</table>

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

![Flow diagram of the study](image)

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,

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

where

$$
\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt
$$
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 β = (β_1, β_2, ..., β_p)^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             | -4.880   | -137.700   | -0.035  | 0.972   | -58.429| -0.492|
| Change.No              | -0.620   | 0.190      | -3.264  | 0.001   | -0.992| -0.252|
| Chronic.Disease.No     | 0.352    | 0.202      | 1.740   | 0.082   | -0.037| 0.745 |
| Heart.Disease.No       | -0.539   | 0.249      | -2.162  | 0.031   | -1.036| -0.058|
| 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

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>Mean rank (SO)</th>
<th>Sum of ranks (ST)</th>
<th>U</th>
<th>z</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heart Attack</td>
<td>192</td>
<td>230.91</td>
<td>44334.5</td>
<td>4529.5</td>
<td>-11.296</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Not Heart Attack</td>
<td>158</td>
<td>108.17</td>
<td>17090.5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>350</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>Mean rank (SO)</th>
<th>Sum of ranks (ST)</th>
<th>U</th>
<th>z</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heart Attack</td>
<td>158</td>
<td>107.48</td>
<td>16982</td>
<td>4421</td>
<td>-11.409</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Not Heart Attack</td>
<td>192</td>
<td>231.47</td>
<td>44443</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>350</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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, chronic 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

<table>
<thead>
<tr>
<th>Variable</th>
<th>Chi-square</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td>21.18</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>ECG change</td>
<td>118.16</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>ST segment change</td>
<td>34.09</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Chronic Disease</td>
<td>23.78</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Patient Pedigree</td>
<td>24.28</td>
<td></td>
</tr>
<tr>
<td>Heart Disease</td>
<td>0.005</td>
<td>0.95</td>
</tr>
</tbody>
</table>

\(^*p<0.05\)

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, chronic disease, and patient pedigree variables have been selected due to the significant difference. However, chronic 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 chronic disease, heart disease, and patient pedigree. The distributions of numerical and categorical variables are given in Figure 2.
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.
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

<table>
<thead>
<tr>
<th>Optimum Parameter Values</th>
<th>SVM (Sigmoid)</th>
<th>SVM (Radial)</th>
<th>SVM (Linear)</th>
<th>SVM (Polynomial)</th>
<th>DT</th>
<th>ANN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cost=1, Gamma=0.2, Coef=0</td>
<td>Cost=1, Gamma=0.17</td>
<td>Cost=1, Gamma=0.2</td>
<td>Cost=1, Gamma=0.2</td>
<td>Tree size=6</td>
<td>Hidden layer number=2, Hidden_layer1_neuron_number=5, Hidden_layer2_neuron_number=2, Epoch=1000, Threshold=0.01, Activation function=logistic</td>
<td></td>
</tr>
</tbody>
</table>

Table 7. The performance results of developed diagnosis models

<table>
<thead>
<tr>
<th>Machine Learning Based Models</th>
<th>With all 8 parameters</th>
<th>SVM (Radial)</th>
<th>DT</th>
<th>ANN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensitivity (%)</td>
<td>83.5</td>
<td>93.8</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>Specificity (%)</td>
<td>85.4</td>
<td>93.7</td>
<td>4.2</td>
<td></td>
</tr>
<tr>
<td>Accuracy (%)</td>
<td>84</td>
<td>93.8</td>
<td>68.3</td>
<td></td>
</tr>
<tr>
<td>With selected 5 parameters</td>
<td>SVM (Radial)</td>
<td>DT</td>
<td>ANN</td>
<td></td>
</tr>
<tr>
<td>Sensitivity (%)</td>
<td>79</td>
<td>91.7</td>
<td>98</td>
<td></td>
</tr>
<tr>
<td>Specificity (%)</td>
<td>83</td>
<td>97.9</td>
<td>93.7</td>
<td></td>
</tr>
<tr>
<td>Accuracy (%)</td>
<td>81</td>
<td>93.8</td>
<td>96.5</td>
<td></td>
</tr>
</tbody>
</table>

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](image)

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 chronic 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.
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

<table>
<thead>
<tr>
<th>Literature Studies</th>
<th>Method</th>
<th>Sens. (%)</th>
<th>Spec. (%)</th>
<th>Acc. (%)</th>
<th>Patients (Test Samples)</th>
<th>Data set Information</th>
<th>Variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>Doğan et al. (2007) [3]</td>
<td>Decision Tree (Heart attack)</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>50Ml(+), 11 Ml(-)</td>
<td>Not given in the paper</td>
<td>CK, CKMB, LDH, AST and ALT</td>
</tr>
<tr>
<td>Mair et al. (1995) [4]</td>
<td>Decision Tree (Heart attack)</td>
<td>91</td>
<td>90</td>
<td>90</td>
<td>45 Ml(+), 49 unstable angina pectoris, 20 chest pain caused by other diseases</td>
<td>Emergency room of a Department of Internal Medicine (University Hospital)</td>
<td>ECG, creatine kinase, creatine kinase MB activity and mass concentration, myoglobin, and cardiac troponin T</td>
</tr>
<tr>
<td>Ghumbre et al. (2011) [6]</td>
<td>RBF network, SVM (Heart disease)</td>
<td>82.4</td>
<td>84.06</td>
<td>82.1</td>
<td>88.5</td>
<td>82.2</td>
<td>4</td>
</tr>
<tr>
<td>Soni et al. (2011) [25]</td>
<td>Decision Tree, Naive Bayes, ANN (Heart disease)</td>
<td>-</td>
<td>-</td>
<td>89</td>
<td>86.5</td>
<td>3</td>
<td>85.5</td>
</tr>
<tr>
<td>Husra et al. (2017) [26]</td>
<td>RBF network (Hannan et al.)</td>
<td>-</td>
<td>-</td>
<td>90-97</td>
<td>75</td>
<td>Sahara Hospital, Aurangabad</td>
<td>Age, P(T(Previous History), P2(Present History), P3(personnel History), P4(Physical Examination), CVX(Cardio Vascular System), RS(Respiratory System), PA(Per Abdomen), CNS(Central Nervous System), ECG (Electrocardiography) and BI (Blood Investigation)</td>
</tr>
<tr>
<td>Mujawar et al. (2015) [27]</td>
<td>Learning Vector Quantization (LVQ) (Chen et al.) (Heart disease)</td>
<td>85</td>
<td>70</td>
<td>80</td>
<td>150</td>
<td>Machine Learning Repository of UCI</td>
<td>Age, sex, chest pain type, trestbps, cholesterol, fasting blood sugar, resting ekg, max heart rate, exercise induced angina, old peak, slope, number of vessels colored, and thal</td>
</tr>
<tr>
<td>Aydin et al. (2016) [28]</td>
<td>Naive Bayes &amp; Modified K-Means (Heart disease)</td>
<td>93</td>
<td>89</td>
<td>91</td>
<td>100 Heart disease</td>
<td>100 Not heart disease</td>
<td>Cleveland Heart Disease database</td>
</tr>
<tr>
<td>Florence et al. (2014) [29]</td>
<td>Bagging AdaBoostM1 Forest Naive Bayes RBF Network IBK NNẻ (Heart disease)</td>
<td>11.1</td>
<td>22.2</td>
<td>33.3</td>
<td>44.4</td>
<td>44.4</td>
<td>33.3</td>
</tr>
<tr>
<td>Proposed Study</td>
<td>Decision Tree (Heart attack)</td>
<td>98</td>
<td>53</td>
<td>-</td>
<td>76</td>
<td>Machine Learning Repository of UCI</td>
<td>Sex, Age, Cardiac Duration, Cholesterol, Signal Level, Possibility of Attack (Yes/No)</td>
</tr>
<tr>
<td></td>
<td>ANN SVM Decision Tree</td>
<td>98</td>
<td>79</td>
<td>93.7</td>
<td>96.5</td>
<td>145</td>
<td>Karadeniz Technical University Farabi Hospital</td>
</tr>
<tr>
<td></td>
<td>Probit Regression</td>
<td>81.3</td>
<td>81.7</td>
<td>81.4</td>
<td>-</td>
<td>Karadeniz Technical University Farabi Hospital</td>
<td>Sex, ECG change, HS Troponin (high sensitivity troponin), Cardiac disease, Patient Pedigree</td>
</tr>
</tbody>
</table>

Table 7. Comparison of the proposed study with similar studies in the literature
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. İlkınur Buçan Kırkbir; Collected the data; Performed the analysis.

APPENDIX

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

<table>
<thead>
<tr>
<th>Variable</th>
<th>Valid Frequency</th>
<th>Percent</th>
<th>CumPercent</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Heart Attack group</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sex Female</td>
<td>35</td>
<td>18.23</td>
<td>18.23</td>
</tr>
<tr>
<td>Sex Male</td>
<td>157</td>
<td>81.77</td>
<td>100.00</td>
</tr>
<tr>
<td>EKG Change No</td>
<td>45</td>
<td>23.44</td>
<td>23.44</td>
</tr>
<tr>
<td>EKG Change Yes</td>
<td>147</td>
<td>76.56</td>
<td>100.00</td>
</tr>
<tr>
<td>ST Segment Change No</td>
<td>113</td>
<td>79.69</td>
<td>79.69</td>
</tr>
<tr>
<td>ST Segment Change Yes</td>
<td>39</td>
<td>20.31</td>
<td>100.00</td>
</tr>
<tr>
<td>Chronic Disease No</td>
<td>70</td>
<td>39.58</td>
<td>39.58</td>
</tr>
<tr>
<td>Chronic Disease Yes</td>
<td>116</td>
<td>60.42</td>
<td>100.00</td>
</tr>
<tr>
<td>Heart Disease No</td>
<td>143</td>
<td>74.48</td>
<td>74.48</td>
</tr>
<tr>
<td>Heart Disease Yes</td>
<td>49</td>
<td>25.52</td>
<td>100.00</td>
</tr>
<tr>
<td>Patient Pedigree No</td>
<td>118</td>
<td>71.88</td>
<td>71.88</td>
</tr>
<tr>
<td>Patient Pedigree Yes</td>
<td>54</td>
<td>28.12</td>
<td>100.00</td>
</tr>
<tr>
<td><strong>Not Heart Attack group</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sex Female</td>
<td>93</td>
<td>58.86</td>
<td>58.86</td>
</tr>
<tr>
<td>Sex Male</td>
<td>93</td>
<td>41.14</td>
<td>100.00</td>
</tr>
<tr>
<td>EKG Change No</td>
<td>131</td>
<td>82.91</td>
<td>82.91</td>
</tr>
<tr>
<td>EKG Change Yes</td>
<td>27</td>
<td>17.09</td>
<td>100.00</td>
</tr>
<tr>
<td>Chronic Disease No</td>
<td>106</td>
<td>67.09</td>
<td>67.09</td>
</tr>
<tr>
<td>Chronic Disease Yes</td>
<td>52</td>
<td>32.91</td>
<td>100.00</td>
</tr>
<tr>
<td>Heart Disease No</td>
<td>119</td>
<td>75.32</td>
<td>75.32</td>
</tr>
<tr>
<td>Heart Disease Yes</td>
<td>39</td>
<td>24.68</td>
<td>100.00</td>
</tr>
<tr>
<td>Patient Pedigree No</td>
<td>147</td>
<td>93.04</td>
<td>93.04</td>
</tr>
<tr>
<td>Patient Pedigree Yes</td>
<td>11</td>
<td>6.96</td>
<td>100.00</td>
</tr>
</tbody>
</table>

(a) Heart Attack group  
(b) Not Heart Attack group

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

<table>
<thead>
<tr>
<th>Heart Attack</th>
<th>CK-MB</th>
<th>Heart Attack</th>
<th>H.S. Troponin</th>
<th>Not Heart Attack</th>
<th>CK-MB</th>
<th>Not Heart Attack</th>
<th>H.S. Troponin</th>
</tr>
</thead>
<tbody>
<tr>
<td>nbr.val</td>
<td>192.00</td>
<td>192.00</td>
<td>158.00</td>
<td>158.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>nbr.null</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>nbr.miss</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>min</td>
<td>1.03</td>
<td>0.01</td>
<td>0.30</td>
<td>0.30</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>max</td>
<td>1579.00</td>
<td>8315.00</td>
<td>4675.00</td>
<td>3985.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>range</td>
<td>1568.97</td>
<td>8314.99</td>
<td>4674.70</td>
<td>3984.99</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>sum</td>
<td>14762.69</td>
<td>13442.97</td>
<td>5487.77</td>
<td>11933.29</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>median</td>
<td>14.80</td>
<td>151.80</td>
<td>2.16</td>
<td>4.86</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mean</td>
<td>76.89</td>
<td>700.14</td>
<td>34.73</td>
<td>75.51</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SE.mean</td>
<td>12.28</td>
<td>97.95</td>
<td>29.57</td>
<td>27.72</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CL.mean.0.95</td>
<td>26.20</td>
<td>193.20</td>
<td>58.40</td>
<td>54.74</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>var</td>
<td>33877.62</td>
<td>1842101.00</td>
<td>138110.20</td>
<td>121347.51</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>std.dev</td>
<td>184.06</td>
<td>1357.24</td>
<td>371.63</td>
<td>348.35</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>coef.var</td>
<td>2.39</td>
<td>1.94</td>
<td>10.70</td>
<td>4.61</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>skewness</td>
<td>5.88</td>
<td>3.09</td>
<td>12.32</td>
<td>4.26</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>kurtosis</td>
<td>38.41</td>
<td>10.15</td>
<td>150.84</td>
<td>98.08</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>kurt.75%</td>
<td>55.01</td>
<td>14.53</td>
<td>186.51</td>
<td>127.78</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>normtest.w</td>
<td>0.40</td>
<td>0.55</td>
<td>0.06</td>
<td>0.20</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>normtest.p</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*nbr.val: Number of instances, nbr.null: Number of missing values, nbr.miss: 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
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


