Progress in Nutrition 2020; Vol. 22, N. 2: 660-670 DOI: 10.23751/pn.v22i2.9830

Original article

Prediction of heart disease by classifying with feature selection and machine learning methods

Cengiz Gazeloğlu
Faculty of Science Literature, Department of Statistics, Suleyman Demirel Universtiy, Isparta, Turkey

Abstract. *Study Objectives*: Cardiovascular diseases are among the most common diseases experienced by human beings. In addition, these diseases require spending too much money to be treated. According to the World Health Organization report, 56 million death cases occurred in the World in 2012. *Methods:* The aim to determine the method (s) with the most accurate classification rate of cardiovascular diseases by using machine learning and feature selection methods. To fulfill this aim, 18 machine learning methods divided into 6 different categories, and 3 different feature selection was used in this study. These methods were analyzed via WEKA, Python and MATLAB computer program. *Results:* According to the results of the analysis, SVM (PolyKernel) with an 85.148% ratio was found to be the most successful machine learning algorithm without feature selection. After the Correlation-based Feature Selection (CFS) feature selection, the most successful algorithm was Naive Bayes and Fuzzy RoughSet with a ratio of 84.818%. However, after using Chi-Square feature selection, the most successful algorithm was found to be the RBF Network algorithm with 81.188% ratio. *Conclusion:* Consequently, it is recommended that specialist doctors who want to classify heart disease should use the SVM (PolyKernel) algorithm if they are not going to use feature selection whereas they should use should the Naive Bayes algorithm if they are going to use CFS as a feature selection. Additionally, if they are to use Fuzzy Rough Set and Chi-Square as the feature selection, it is recommended that they use the RBFNetwork algorithm.

Key words: Machine Learning, Feature Selection, Heart Disease, Classification, Artificial Intelligence

Introduction

According to the World Health Organization report, 56 million death cases occurred in the World in 2012. 38 million of these deaths were caused by Noncommunicable Diseases (NCD) and especially cardiovascular diseases, cancer, chronic airway diseases. One third (28 million) of these deaths took place in countries with low and middle-income. Of these deaths, 7.4 million were due to heart attacks (ischemic heart disease) and 6.7 million were due to stroke. 46.2 percent (17.5 million) of deaths due to noncommunicable diseases were caused by cardiovascular diseases. Cardiovascular diseases are responsible for 37 percent of deaths under the age of 70 depending on the noncommunicable diseases. It is predicted that deaths due to cardiovascular diseases will be 22.2 million in 2030 (1). As can be seen from the estimation by experts, very important risks await the human species. This risk threatens all countries economically, regardless of whether they are developed or not.

As with all diseases, early diagnosis is very important in people with heart disease. Because when it is too late for the treatment of the disease, both large amounts of money are spent and it causes problems in recovery. The main problems on this issue are that in the first stage of the disease, the diagnosis of the disease is not made with the help of computer-aided systems and the process is very slow since the final decision is made by the doctors. In addition, results are directly

affected due to conditions such as the education taken, working conditions, number of patients per physician. In addition, the low rate of making the right decision at the first stage can be seen as another problem. In order to make a full diagnosis, people who apply to health institutions are repeatedly tested. This state means both time and financial loss. The way to minimize the damage in these situations is to benefit from computer-aided smart systems. The main reason why such smart systems have not been used until now is that artificial intelligence and computer systems were not so common. Besides, the fact that the technology was not so advanced and accessible made the transition to these smart systems difficult. This does not mean that the human factor will be completely eliminated in decision-making processes. On the contrary, the human factor will become even more effective. Because ultimately, it is the people who produce these smart systems. The only goal here is to speed up the process and make decisions with high accuracy by minimizing human errors.

In this study, a solution has been made with the mentioned speed and correct decisions, computer aided smart systems, machine learning algorithms. 18 machine learning methods and 3 feature selections were used in the study so as to find the combination that best predicts heart disease (best method and feature selection). No feature selection was applied to the relevant data set. After the selection of CFS, Fuzzy-RoughSet, and Chi-Square features, the results of 18 different machine learning algorithms, which are collected in 6 different categories, were compared in the related tables. This study will not only help specialists working in the field of heart disease to diagnose but also will accelerate the experts much more in the context of time. As the study is taken into consideration in this respect, it is an important study that will facilitate the work of specialists.

Material and Method

Participants

Data used in the study have been taken from https://www.kaggle.com/ronitf/heart-disease-uci18 website and the related data set belongs to 303 patients and they are formed of 14 variables. These mentioned data were collected at a health center in Cleveland, Ohio, one of the central northern states of the United States. The variables of the data set are given in (Table 1) below.

Experimental design

In this experimental study, for the prediction of heart disease Decision tree, ADTree, k-NN, RoughtSet, logistic regression, randomforest, NBTree, RBFNetwork, FuzzyRoughNN, FuzzyNN, NN, MLP, Naïve Bayes and SVM(Poly Kernel, NormalizedPoly Kernel, Puk and RBF Kernel) classification algorithms have been used and the correct classification rates of these classification algorithms are given in Table 3. Also, the results of the ROC, TP, FP, and Kappa Statistics analyzes are given in Table 5. Besides, (Table 2) presents the variables used as a result of the analysis of 3 different feature selection algorithms used in the study.

Table 1. The description of heart disease dataset

Feature name	Features
var1	age in years
var2	Sex
var3	Cp
var4	Trestbps
var5	Chol
var6	Fbs
var7	Restecg
var8	Thalach
var9	Exang
var10	Oldpeak
var11	Slope
var12	Ca
var13	Thal
var14	Target (1 or 0)

Table 2. Feature selection algorithms and sub features

Cfs Sub Set Eval	V3, V4, V8, V9, V10, V12
Fuzzy Rough Set	V1, V3, V4, V5, V8, V10, V12
Chi-Square	V1, V3, V4, V5, V8, V10, V12

Statistical analysis

Hypotheses to be tested in the study are given above.

Machine learning algorithms differ according to the structure of the data to be used and the solution of the problem. Here is what needs to be done: to know the data structure used in the study and to define the problem well. Depending on these situations, it is necessary to determine the most appropriate method or methods. If the appropriate method is not determined, there will be differences between the accuracy rates of the algorithms used. Depending on this state:

H_1: There is a difference between machine learning algorithms used in computer aided smart systems in terms of accuracy rates in the diagnosis of heart disease.

Parameters are facts that are tried to be estimated using statistics and provide information about the universe. There are many methods used to estimate these parameters. The important thing is to be able to estimate the parameter consistently with the correct method. There are many situations that affect this consistent estimate. The most effective of these situations is the number of parameters used in the solution of the problem. The number of parameters should be neither too few nor too many. It should be at an optimum level. If the appropriate number is not determined, the success rates of the algorithms used will be directly affected. Depending on this situation:

H_2: The number of parameters used in diagnosis has an effect on the accuracy rate of the algorithm.

Feature selection is the process of evaluating which parameters are effective and how effective they are on the result. There are many feature selections in the literature. The important thing here is not to use feature selection in studies. It is necessary to find out whether feature selection has a positive effect on the result of the study. Depending on that:

H_3: The feature selection used affects the results of machine learning algorithms.

The results of the algorithms used in the studies are evaluated only on the correct percentages of success in most of them. However, these evaluations mislead readers. Success rates are necessarily important. Yet, it is also necessary to decide which of the algorithms are better in Type I and which in Type II error types. Because some algorithms give accurate results on the correct classification rate, while others give more successful results on the wrong classification rate. Depending on that:

H_4: It is incorrect to evaluate the performance of machine learning algorithms only on the correct percentages of success.

Classification algorithms

The concept of classification can be defined as distributing data between classes that are defined under certain rules on a data set. There are many classification methods in the literature. The important point here is to determine the correct classification algorithm appropriate to the data set and the success rate of the algorithm used is high.

The Classification Algorithms Used in the Study;

K-nearest neighbors (KNN) is the classification method for classifying unknown examples by searching the closest data in pattern space (2). KNN predicts the class by using the Euclidean distance defined as follows:

$$d(x,y) = \sqrt{\sum_{i=1}^{k}(x_i - y_i)^2}$$

The Euclidean distance is used to measure the distance for finding the closest examples in the pattern space. The class of the unknown example is identified by a majority voting from its neighbors.

In addition, Euclidean distance is among the most used distance measurements (3).

Regression is the description of the relationship between a response variable and one or more explanatory variables. The result variable usually takes two or more values. Recently, the logistic regression model has become a standard model (4).

Naive Bayes; Bayesian network consists of a structural model and a set of conditional probabilities. The structural model is a directed graph in which nodes represent attributes and arcs represent attribute dependencies. Attribute dependencies quantified by conditional probabilities for each node given its parents. Bayesian networks are often used for classification problems (5).

The decision tree is a method which is easy to understand and interpret classification (6). The decision tree method is one of the most popular algorithms in

classification algorithms. This method is based on entropy. The most important point in constructing decision trees is to determine which variable is the first loop, that is, the root loop (7).

Alternating decision tree (ADTree) is one of the machine learning methods used for classification. It is closely related to the decision tree method, which is another method of machine learning. It was originally developed by Freund and Mason in 1999. The method consists essentially of decision nodes and prediction nodes. A classification is made by following all the paths in which the decision nodes are correct and collect all the prediction nodes passed.

Naive Bayes Tree (NBTree) is a hybrid algorithm, which deploys a naive Bayes classifier on each leaf node of the built decision tree and has demonstrated remarkable classification performance (8).

Fuzzy-Rough Nearest Neighbour (FRNN) is the extension of the K-nearestneighbor algorithm by using the fuzzy-rough uncertainty. The fuzzy uncertainty concept is used to measure the distance between the test pattern and the neighbor. It also helps to represent the neighbour to be in many classes. Due to the lack of features some of the neighbors and the test patterns may be indistinguishable hence the concept of rough uncertainty is used. The neighborhood structure is artificial, so the roughness emerges (9).

The Fuzzy Nearest Neighbor (FuzzyNN) classifier is well known for its effectiveness in supervised learning problems. K-NN classifies by comparing new incoming examples with a similarity function using the samples of the training set. The fuzzy version of the kNN accounts for the underlying uncertainty in the class labels, and it is composed of two different stages. The first one is responsible for calculating the fuzzy membership degree for each sample of the problem in order to obtain smoother boundaries between classes. The second stage classifies similarly to the standard kNN algorithm but uses the previously calculated class membership degree (10).

Among the various methods of supervised statistical pattern recognition, the Nearest Neighbour (NN) rule achieves consistently high performance, without a priori assumptions about the distributions from which the training examples are drawn. It involves a training set of both positive and negative cases. A new sample is classified by calculating the distance to the nearest training case; the sign of that point then determines the classification of the sample. The k-NN classifier extends this idea by taking the k nearest points and assigning the sign of the majority (11).

Rough clustering analysis is based on data tables called information systems. The so-called information system is a table that provides information about related objects in terms of some features. In these tables, system conditions and decision variables are generally separated from each other. Such information system tables are called decision tables. The decision table is explanatory of the conditions that must be met. Each decision table contains a series of inter-related rules. Each decision algorithm reveals well-known probability features. The best examples of these are probability and bayes theorems. These features give a new method to draw conclusions from the data (12).

Multilayer Perceptron (MLP) is a multi-layered artificial neural network. The basic structure of MLP consists of at least 3 layers. The first layer is called the input layer, the second latent layer, and the last layer is the output layer. In addition, MLP uses a back-propagated supervised learning technique in the training of the dataset.

In spite of the fact that the capacity control principle (the SRM principle) was discovered in the middle of the 1970s. the development of this principle-which led to new types of algorithms, the so-called Support Vector Machines (SVM) started only in the 1990s. One of the most influential developments in the theory of machine learning in the last few years is Vapnik's work on support vector machines (SVM) (13). The aim of it is to group data according to the support vectors. These groupings are more suitable for linear data sets. However, for nonlinear data sets, data sets can be linearized and implemented with kernel functions. In this study, Poly Kernel, Normalized Poly Kernel, Puk and Radial Basis Function (RBF) Kernel functions have been employed.

The radial basis function network (RBFNetwork) method is generally used estimations in time series, classification problems, system controls, modeling field. It is almost the same as artificial neural networks. The only difference from artificial neural networks is that it uses the radial basis function as the activation function.

Random forests are a combination of tree predictors

such that each tree depends on the values of a random vector sampled independently and with the same distribution for all trees in the forest. The generalization error for forests converges as to a limit as the number of trees in the forest becomes large. The generalization error of a forest of tree classifiers depends on the strength of the individual trees in the forest and the correlation between them (14).

Genetic programming (GP) is an evolutionary technique used for generating computer programs based on a high level description of the problem to be solved. This innovative flexible and interesting technique has been applied to solve numerous interesting problems. Classification is one of the ways to model the problems of face recognition, speech recognition, fraud detection and knowledge extraction from databases. GP has emerged as a powerful tool for classifier evolution. Classification is a common real world activity. It is used to put entities or patterns into predefined classes (15).

Feature selection

Feature selection is an important set of algorithms used to achieve more consistent results by improving the correct classification rates or performances of the methods used in machine learning systems. In this study, CFS, Fuzzy Rough Set and Chi-Square algorithms are used as feature selection algorithm.

Correlation-based Feature Selection (CFS) is a simple correlation-based filtering algorithm. One point to note here is that features with low correlation should be ignored. It should be ensured that the remaining features are highly correlated with each other. CFS's feature subset evaluation function equation below is repeated here for ease of reference:

$$M_s = \frac{k\overline{r_{cf}}}{\sqrt{k + k(k-1)\overline{r_{ff}}}}$$

where is the heuristic "merit" of a feature subset S containing k features, is the mean feature-class correlation (f S), and is the average feature-feature intercorrelation (16).

Fuzzy Rough Sets; A fuzzy-rough set is a generalization of a rough set, derived from the approximation of a fuzzy set in a crisp approximation space. This corresponds to the case where the values of the conditional attribute are crisp and the decision attribute values are fuzzy. The main focus of fuzzy-rough sets is to define lower and upper approximation of the set when the universe of the fuzzy set becomes rough because of quivalence relation or transforming the equivalence relation to similar fuzzy relation (17).

At this point, it is aimed to employ a method to calculate reducts for fuzzy rough sets, and only the minimal elements positioned in the discernibility matrix are taken into consideration. Initially, the definition of the relative discernibility relations of the conditional attribute is carried out, then to qualify the minimal elements in the discernibility matrix relative discernibility relations are employed. Followingly, to calculate the minimal elements an algorithm is created. At the end, the designation of new algorithms so as to figure out correct reducts with the minimal elements is achieved (18).

Chi-Square; is one of the most popular feature selection algorithms known in the literature. What lies on the basis of the algorithm is the calculation of the chi-square value between each feature and the target feature. With this calculation, the best chi-square score is determined and the desired number of properties is selected. The formula of the algorithm is given in the equation below.

$$\chi^2 = \frac{(Observed\ frequency - Expected\ frequency)^2}{Expected\ frequency}$$

Results

Table 3 shows the correct classification rates according to the feature selections of 18 different machine learning algorithms. It is useful to specify an important point in feature selection. In FuzzyRoughSet and Chi-Square feature selections, the results are the same. In other words, the same variables were used in the selection of 2 properties.

According to Table 3, the most accurate classification rate has been the RBFNetwork algorithm with approximately 85% as a result of 18 classification algorithms without feature selection. On the other hand, the lowest classification rate was found to be about 65%, which belongs to the FuzzyNN algorithm. After

Table 3. Accurate classification ratios of classification algorithms (as per %)

Model	No Feature Selection	CFS	FuzzyRoughSet and Chi-Square
Logistic Regression	83,828	83,498	80,198
J48	75,247	81,188	78,877
NaiveBayes	83,498	84,818	80,198
KNN	77,227	78,547	73,267
RouhtSet	82,178	78,547	78,217
ADTree	79,538	84,488	75,577
MLP	79,538	79,207	75,577
SVM (PolyKernel)	85,148	84,158	79,868
SVM- (Normalized PolyKernel)	83,828	81,848	78,877
SVM(Puk)	83,828	82,508	78,547
SVM (RBFKernel)	83,828	83,168	77,557
RandomForest	82,178	80,581	77,557
NBTree	79,868	83,828	78,217
RBFNetwork	84,488	83,168	81,188
Fuzzy RoughNN	79,868	77,887	72,937
FuzzyNN	64,686	59,405	64,686
NN	83,168	81,848	78,217
Geneticpro-graming	81,188	82,838	78,877

applying the CFS feature selection method, the best accuracy belongs to the NaiveBayes classification algorithm with approximately 85% while the lowest accuracy was found to be the FuzzyNN algorithm with 59%. After the implementation of FuzzyRoughSet and Chi-Square feature selections, the best algorithm became RBFNetwork with 81% whereas the lowest accuracy rate was about 65% FuzzyNN.

After making CFS feature selection, it has been observed that there is an improvement in J48, NaiveBayes, KNN, ADTree, SVM(PolyKernel), NBTree and Geneticprograming in the accurate classification rates of classification algorithms. On the other hand, a decrease has taken place in the remaining algorithms.

After FuzzyRoughSet and Chi-Square feature selection, there has been improvement in the correct classification rate only in the Decision tree (J48) classification algorithm.

As it is evaluated in general, it is seen that the correct grading rate of the J48 algorithm has always improved in all three feature selection algorithms.

Performance evaluation criteria of classification algorithms

The ROC analysis is used in the determination of the ability to distinguishing power of the test, comparison of various test techniques and in the determination of the appropriate positive threshold.

The area calculated by ROC analysis is one of the most important analysis methods used to evaluate the performance of classification algorithms.

$$AUC = \frac{S_p - n_p(n_n + 1)/2}{n_p n_n}$$

Here indicates the sum of all positive samples while and give the number of positive and negative ones respectively.

TP rate can be defined as the classification of a true condition as true in the test result. It is also known in the literature as Sensitivity.

Sensitivity is the proportion of true positives that are correctly identified by the test (19).

$$Sensitivity = TP/(TP + FN)$$

FP Rate is deciding that it is correct after testing a condition that is incorrect in reality. It is also known in the literature as specificity.

Specificity is the proportion of true negatives that are correctly identified by the test (19).

$$Sensitivity = TN/(FP + TN)$$

Kappa coefficient is a statistic that measures inter-rater agreement for categorical items. It is generally thought to be a more robust measure than simple percent agreement calculation since κ takes into account the agreement occurring by chance. Cohen's kappa measures agreement between two raters only but Fleiss' kappa is used when there are more than two raters. κ may have a value between -1 and +1. A value of kappa equal to +1 implies perfect agreement between the

two raters, while that of -1 implies perfect disagreement. If kappa assumes the value 0, then this implies that there is no relationship between the ratings of the two observers, and any agreement or disagreement is due to chance alone (20) (see Table 4).

Table 4. Evaluation of Kappa Statistical Coefficient (21).

Kappa Statistic	Strength of Agreement
< 0.00	Poor
0.00-0.20	Slight
0.21-0.40	Fair
0.41-0.60	Moderate
0.61-0.80	Substantial
0.81-1.00	Almost Perfect

Table 5 shows the TP, FP, ROC and Kappa Statistic results of the 18 classification algorithms. Besides, the table presents the results of the mentioned classification algorithms without feature selection, CFS feature selection and FuzzyRoughSet feature selection and the results after chi-Square feature selection whose results are the same.

According to Table 5, the best TP ratio among the 18 classification algorithms without any feature selection was calculated with 0.891 in SVM (PolyKernel). However, the lowest TP ratio belongs to the FuzzyNN algorithm with 0.709. Based on these results, it can be said that SVM (PolyKernel) has the ability to call a person as "patient" with a ratio of approximately 90%. However, when CFS, FuzzyRoughSet and Chi-Square feature selections are made in this algorithm,

Table 5. TP, FP, ROC and Kappa Statistic analysis results of classification algorithms (No Feature Selection /Cfs/FuzzyRoughtSet/ Chi-Square)

	TP Rate			**FP Rate**			**ROC Area**			**Kappa Statistic**		
Model	**No Feature Selection**	**CfsSubSetEval**	**FuzzyRoughSet and Chi-Square**	**No Feature Selection**	**CfsSubSetEval**	**FuzzyRoughSet and Chi-Square**	**No Feature Selection**	**CfsSubSetEval**	**FuzzyRoughSet and Chi-Square**	**No Feature Selection**	**CfsSubSetEval**	**FuzzyRoughSet and Chi-Square**
Logistic Regression	0,867	0,879	0,824	0,196	0,217	0,225	0,885	0,896	0,875	0,673	0,665	0,600
J48	0,782	0,855	0,855	0,283	0,239	0,29	0,758	0,837	0,784	0,500	0,618	0,570
NaiveBayes	0,861	0,897	0,848	0,196	0,210	0,254	0,909	0,905	0,874	0,666	0,691	0,598
KNN	0,782	0,824	0,764	0,239	0,261	0,304	0,776	0,783	0,721	0,541	0,565	0,460
RouhtSet	0,873	0,836	0,855	0,239	0,275	0,304	0,817	0,781	0,775	0,638	0,564	0,556
ADTree	0,824	0,885	0,782	0,239	0,203	0,275	0,888	0,895	0,841	0,586	0,685	0,507
MLP	0,812	0,812	0,776	0,225	0,232	0,268	0,879	0,866	0,814	0,587	0,580	0,507
SVM(PolyKernel)	0,891	0,891	0,836	0,196	0,217	0,246	0,848	0,837	0,795	0,699	0,678	0,592
SVM(Normalized PolyKernel)	0,879	0,897	0,903	0,210	0,275	0,348	0,834	0,811	0,778	0,672	0,629	0,565
SVM(Puk)	0,873	0,861	0,885	0,203	0,217	0,333	0,835	0,822	0,776	0,672	0,645	0,560
SVM(RBFKernel)	0,873	0,879	0,806	0,203	0,225	0,261	0,835	0,827	0,773	0,672	0,658	0,546
RandomForest	0,861	0,848	0,836	0,225	0,239	0,297	0,893	0,887	0,854	0,639	0,612	0,543
NBTree	0,812	0,873	0,867	0,217	0,203	0,319	0,877	0,905	0,858	0,594	0,672	0,555
RBFNetwork	0,885	0,879	0,873	0,203	0,225	0,261	0,900	0,888	0,861	0,685	0,658	0,617
FuzzyRoughNN	0,824	0,824	0,776	0,232	0,275	0,326	0,861	0,844	0,811	0,593	0,551	0,451
FuzzyNN	0,709	0,636	0,709	0,428	0,457	0,428	0,641	0,590	0,641	0,283	0,180	0,283
NN	0,873	0,903	0,885	0,217	0,283	0,341	0,902	0,897	0,846	0,658	0,629	0,553
Geneticprograming	0,861	0,891	0,830	0,246	0,246	0,261	0,807	0,822	0,785	0,618	0,650	0,572

the result does not change in CFS whereas this rate reduces to 83% in other feature selections. SVM (Normalized PolyKernel) and NN algorithms give the best results before and after the relevant feature selections are made about the TP rate. This rate is 90%. That is, these algorithms can diagnose a patient with a 90% rate as a result of classification.

FP Rate is known as the diagnosis of a sick person as "not sick" as a result of the test. It is also called as the alpha error in statistical science. This low rate is very important in terms of making the right decision. That is to say, So the closer this ratio is to zero, the better it is. According to Table 5, there were two algorithms with the lowest FP rate without any feature selection. These are Logistic regression and SVM (PolyKernel) classification algorithms with a 0.196 FP ratio. After applying CFS feature selection, ADTree is the best method among the 18 different classification algorithms with a ratio of 0.203.

ROC analysis is a frequently used method for comparing various tests. In this study, the algorithm with the highest ROC area was found to be the Naive-Bayes without feature selection and after CFS, Fuzzy-Rough and Chi-Square feature selections. Additionally, the NBTree classification algorithm was also the method with the highest ROC area after CFS feature selection. In these mentioned four different cases, the lowest ROC area was calculated in the FuzzyNN classification algorithm.

Kappa statistic is known as an indicator of the harmony between observers in two-class data. In this study, the best fit has been the RBFNetwork classification algorithm with approximately 69% as a result of 18 classification algorithm without feature selection and after FuzzyRoughSet / Chi-Square feature selection. This result is considered to be Substantial according to Landis and Koch. However, the best algorithm as a result of CFS feature selection was found to be the NaiveBayes classification algorithm with about 70%. In all 4 cases, ie without feature selection and after the other 3 feature selections, the lowest rates were calculated in the FuzzyNN algorithm.

Table 6 shows whether the 4 hypotheses used in the study were validated according to the results obtained, and if it is confirmed, there is an explanation of what the result is.

Comparison of the other studies

Table 7 presents the studies on heart disease obtained as a result of the literature review. In this table, you can find the machine learning algorithms used in these studies and the correct classification rates of these algorithms. When these case studies are examined, it is seen that except for this study, only one study (22) appears to classify by using feature selection. What makes our study different from other studies is that a classification is carried out without using feature selection and using 3 different feature selection. In this way, the effect of feature selection on the classification of the disease has been revealed.

Table 6. Hypotheses used in the study

Hypotheses	Whether Confirmed by Study	Result at the End of the Study
H_1	Confirmed	A difference between machine learning algorithms was found. The best algorithm was found to be SVM (PolyKernel) with an 85.14% accuracy rate.
H_2	Confirmed	Initially, 13 variables were used. The good algorithm was SVM (PolyKernel) with 85.14%. 6 variables with CFS the best algorithm with 84.81% ratio is Naive Bayes Fuzzy Rought Set and Chi-Square have the same variables and RBF Network was the best algorithm with a rate of 81.18%.
H_3	Confirmed	Since different variables are used in the CFS, FuzzRough-Set and Chi-Square feature selections used, the success percentages of machine learning algorithms also change.
H_4	Confirmed	Looking at the TP ratios, the best result was the NN algorithm. The best result in FP was Naive Bayes and SVM (polyKernel) while in ROC analysis it is the Naive Bayes and Naive Bayes gave the best results in Kappa Statistics.

Table 7. Some related studies as a result of the literature review and their results

Study	Feature Selection	Best Algorithm and Result
Vembandasamy et al. (23)	No	NB (86.41%)
Das et al. (24)	No	ANN Ensemble (89.01%)
Chen et al. (25)	No	ANN (80%)
Dangre and Apte (26)	No	ANN (Nearly 100%)
Sabarinathan and Sugumaran (27)	No	J48 (85%)
Patel et al. (28)	No	J48 (85%)
Shouman et al. (29)	No	KNN (97.4%)
Wiharto et al. (30)	No	SVM (90%)
Khateeb and Usman (31)		KNN (80%)
Pouriyeh et al. (32)	No	NB (83.49%), DT (77.55%), MLP (82.83), KNN (83.16%), SCRL (69.96), RBF (83.82), SVM (84.15%)
Waghulde and Patil (33)	No	Genetic Neural Approach (6 hidden) (98%) and (10 hidden 84%)
Venkatalakshmi and Shivsankar (34)	No	NB (85.03) and DT (84.01%)
Palaniappan and Awang (35)	No	DT (85.53), NB (86.53) and ANN (85.53%)
Liu et al. (36)	RFRS	RFRS (92.59%)
Ghumbre et al. (37)	No	SVM (86.42) and RBF (80.81%)
Masethe and Masethe (38)	No	J48 (99.07), NB (97.22), REPTREE (99.07), Simple Cart (99.07), and Bayes Net (99.07 %)
Dangare and Apte (39)	No	ANN (Nearly 100%), DT(99.62%) and NB (90.74%)
	No	SVM (Poly Kernel) (85.148 %)
Our Study	Cfs	Naive Bayes (84.818 %)
	Fuzzy Rough Set and Chi-Square	RBF Network (81.188 %)

Discussion and Conclusion

The aim of this study is to determine the machine learning algorithm with the highest accuracy rate of machine learning algorithms, which is one of the computer-aided smart systems in the diagnosis of heart disease. And also, it aims to determine how feature selection affects the performance of these machine learning algorithms and to make determinations about the diagnosis of disease by using TP, FP, Kappa Statistics and ROC analysis as performance criteria.

With the data set used in this study, many different methods have been used to do analysis before. Unlike other studies, in this study, the results of the related algorithms were obtained without feature selection. Additionally, it is observed whether there is an improvement in the performance of these algorithms by using 3 different feature selections. When evaluated in this sense, it provides an opportunity to compare how algorithms respond during feature selection. In addition to these results, the accuracy rates of the algorithms used in the diagnosis of heart disease investigated separately in this study in a way that it diagnoses a non-patient as not a patient or diagnoses a patient as a patient. When this study is taken from that point, this study is valuable in terms of covering 18 different machine learning, 6 feature selections and 4 performance evaluation criteria divided into 6 categories, in this respect, a significant deficiency in the literature will be eliminated with this study.

It was obtained that after CFS feature selection, the highest accuracy rate belongs to the NaiveBayes algorithm with approximately 85%. However, in the FuzzyRoughset and Chi-square feature selections, this

ratio belongs to the RBFNetwork classification algorithm with approximately 82%. If experts working in this field want to work with fewer variables, these individuals have the opportunity to work with 6 variables as a result of CFS feature selection, and 7 variables as a result of Fuzzy and chi-square feature selection. In this study there is a total of 13 variables except the variable used to determine the disease class. It is believed that it is a good rate that the number of variables has decreased to 6 and 7 variables in order to classify the disease, which is a reduction of approximately 47%. As a result of the analyzes, it is seen in Table 3 that the highest classification rate is in SVM (PolyKernel) algorithm without any feature selection. Here the readers may come up with a question like this: If the highest accuracy is achieved without a feature selection, why should we then carry out a feature selection? The answer actually lies in the purpose of the study. Thanks to this study the relevant disease is estimated with the highest accuracy rate using fewer variables. Since fewer variables are used, both time and money will be saved. In addition, there will be no loss from the correct classification rate. When evaluated in this sense, it is thought that the study will contribute to the literature greatly.

There is no information about the races of individuals in the data set used. It is considered to explore whether the parameters used in the diagnosis of the disease are effective on the races in the future. Also, if it turns out that the parameters change according to the races, then it is planned to conduct studies on which feature selection algorithm and machine learning algorithm will yield more successful results. Besides, a mobile phone application that will include 18 machine learning, 3 feature selections and 4 performance evaluation criteria will be implemented with the help taken from the relevant departments. It is planned to provide a preliminary assessment opportunity for the diagnosis of the disease by requesting from these people to enter some parameter values.

References

1. Global status report on noncommunicable diseases WHO. https://www.kisa.link/NtkY, Accessed Date: 10. August. 2019
2. Galit S, Nitin RP, Peter CB. Data Mining for Business Intelligence: Concepts, Techniques, and Applications in Microsoft Office Excel with Xlminer, Wiley Publishing, 2010.
3. Saracli S. Performance of rand's C statistics in clustering analysis: an application to clustering the regions of Turkey, Journal of Inequalities and Applications 2013; 1–142.
4. Dawid W, Hosmer SL. Applied Logistic Regression, 2th ed. A Wiley Interscience Publication; 2000.
5. Jiang L, Zhang H, Cai Z. A novel bayes model: hidden naive bayes. IEEE T Knowl Data En 2009; 21(10): 1361–1371.
6. Xing Y, Wang J, Zhao Z. Combination Data Mining Methods with New Medical Data to Predicting Outcome of Coronary Heart Disease, ICCI Presented 2007.
7. Atılgan E, Karayollarında Meydana Gelen Trafik Kazalarının Karar Ağaçları ve Birliktelik Analizi ile Incelenmesi, Hacettepe Üniversitesi, Fen Bilimleri Enstitüsü Bilim Uzmanlığı Tezi, Ankara, 2011.
8. Wang S, Jiang L, Li C. Adapting naive Bayes tree for text classification. Knowledge and Information Systems 2015; 44(1): 77–89.
9. Meenachi L, Ramakrishnan S, Arunithi M, Karthiga R, Karthika S, Nandhini P. Diagnosis of cancer using fuzzy rough set theory. International Research Journal of Engineering and Technology (IRJET) 2015; 3(1): 1203–1208.
10. Maillo J, Luengo J, Garc´ ıa S, Herrera F. A preliminary study on Hybrid Spill-Tree Fuzzy k-Nearest Neighbors for big data classification, IEEE International Conference on Fuzzy Systems (FUZZ-IEEE) 2018.
11. Nearest Neigbour Classifier https://www.kisa.link/Ntl5, Accessed Date: 25. July. 2019
12. Pawlak Z. Rough sets and intelligent data analysis. Information Sciences 2002; 147(1):1–12.
13. Vapnik V N. Estimation of Dependences Based on Empirical Data. Springer Verlag.; 1982.
14. Breiman L. Random Forests. Kluwer Academic Publishers. Manufactured in The Netherlands 2001; 45:5–32.
15. Purohit A, Choudhari NS, Tiwari A. A new mutation operator in genetic programming, ictact journal on soft computing 2013; 3(2): 467–471.
16. Hall AM. Correlation-based feature selection for machine learning Tech. Rep., Doctoral Disertation, University of Waikato, Department of Computer Science 1999.
17. Kumar M, Yadav N. Fuzzy rough sets and its application in data mining field, Advances in Computer Science and Information Technology (ACSIT) 2015; 2(3): 237–240.
18. Heart Disease UCI, https://www.kaggle.com/ronitf/heart-disease-uci, Accessed Date: 20. April. 2019
19. Altman DG, Bland JM. Diagnostic tests. 1: Sensitivity and specificity. BMJ. 1994; 308 (6943): 1552.
20. Kılıç S. Kappa Testi. Journal of Mood Disorders 2015; 5(3):142–144.
21. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977; 33(1):159–74.
22. Liu X. et al., A hybrid classification system for heart disease diagnosis. Computational and Mathematical Methods in Medicine 2017; 2017: 1–11.
23. Vembandasamy K, Sasipriya R, Deepa E. Heart diseases de-

tection using naive bayes algorithm. International Journal of Innovative Science. Engineering & Technology 2015; 2(9): 441–444.
24. Das R, Turkoglu I, Sengur A. Effective diagnosis of heart disease through neural networks ensembles. Expert systems with applications 2009; 36(4): 7675–7680.
25. Chen A. et al., HDPS: Heart disease prediction system. In Computing in Cardiology, Hangzhou, China: IEEE 2011; 38: 557–560.
26. Dangare C, Apte S. A data mining approach for prediction of heart disease using neural networks. International Journal of Computer Engineering & Technology 2012; 3(3): 30–40.
27. Sabarinathan V, Sugumaran V. Diagnosis of heart disease using decision tree. International Journal of Research in Computer Applications & Information Technology 2014; 2:74–79.
28. Patel J. et al., Heart disease prediction using machine learning and data mining technique. Heart Disease 2015; 7(1):129–137.
29. Shouman M, Turner T, Stocker R. Applying k-nearest neighbour in diagnosing heart disease patients. International Journal of Information and Education Technology 2012; 2(3):220.
30. Wiharto W, Kusnanto H, Herianto H. Performance analysis of multiclass support vector machine classification for diagnosis of coronary heart diseases. International Journal on Computational Science & Applications 2015; 5(5): 27–37.
31. Khateeb N, Usman M. Efficient heart disease prediction system using k-nearest neighbor classification technique. In Proceedings of the International Conference on Big Data and Internet of Thing (BDIOT), New York, NY, USA: ACM, 2017; 21–26.
32. Pouriyeh S. et al., A comprehensive investigation and comparison of machine learning techniques in the domain of heart disease. In Proceedings of IEEE Symposium on Computers and Communications (ISCC). Heraklion, Greece: IEEE, 2017; 204–207.
33. Waghulde N, Patil N. Genetic neural approach for heart disease prediction. International Journal of Advanced Computer Research 2014; 4(3): 778.
34. Venkatalakshmi B, Shivsankar M. Heart disease diagnosis using predictive data mining. International Journal of Innovative Research in Science, Engineering and Technology, 2014; 3(3): 1873–1877.
35. Palaniappan S, Awang R. Intelligent heart disease prediction system using data mining techniques. In IEEE/ACS International Conference on Computer Systems and Applications. Doha, Qatar 2008; 8(8):108–115.
36. Liu X. et al., A hybrid classification system for heart disease diagnosis. Computational and Mathematical Methods in Medicine 2017; 2017: 1–11.
37. Ghumbre S, Patil C, Ghatol A. Heart disease diagnosis using support vector machine. In International conference on computer science and information technology. Pattaya, Thailand: Planetary Scientific Research Centre 2011; 84–88.
38. Masethe H, Masethe M. Prediction of heart disease using classification algorithms. In Proceedings of the world congress on Engineering and Computer Science, San Francisco, USA: International Association of Engineers (IAENG) 2014; 2: 22–24.
39. Dangare C, Apte S. Improved study of heart disease prediction system using data mining classification techniques. International Journal of Computer Applications 2012; 47(10): 44–48.

Correspondence:
Cengiz Gazeloglu
Faculty of Science Literature, Deparment of Statistics,
Suleyman Demirel Universtiy, Isparta, Turkey.
E-mail: cengizgazeloglu@sdu.edu.tr