Classification of Factors Affecting Renal Failure by Machine Learning Methods

Burçin Şeyda ÇORBA ZORLU[1], *Pelin KASAP[1]

*[1]Ondokuz Mayıs University, Faculty of Science and Letters, Department of Statistics, SAMSUN

(Alınış / Received: 28.10.2019, Kabul / Accepted: 24.01.2020, Online Yayınlanma/ Published Online: 01.04.2020)

Keywords
Machine Learning,
Classification,
Decision Trees,
Renal Failure

Abstract: Machine learning methods are widely used for data analysis in health research. The aim of this study is to classify the factors that affect renal failure by using various machine learning methods such as Artificial Neural Networks (Multilayer Perceptron), Support Vector Machines, Naive Bayes, Decision Trees, Random Forests, K-Nearest Neighborhood algorithms. In this study, 237 patients who have been in emergency unit in Hospital of Numune in Ankara and were older than 18 years and have upper gastrointestinal bleeding symptoms have been selected. Here, 34 variables such as age, gender, blood values, other diseases etc. which affect renal failure have been used to make classification with machine learning methods. When machine learning methods are compared according to the accuracy rates, F-measure, sensivity, specifity and Kappa values, it has been found that decision trees algorithm performs well.

Makine Öğrenmesi Yöntemleri ile Böbrek Yetmezliği Hastalığını Etkileyen Faktörlerin Sınıflandırılması

Anahtar Kelimeler
Makine Öğrenmesi,
Sınıflandırma,
Karar Ağaçları,
Böbrek Yetmezliği

Öz: Makine öğrenmesi yöntemleri, sağlık araştırmalarında veri analizi için yaygın olarak kullanılmaktadır. Bu çalışmanın amacı, Yapay Sinir Ağları (Çok Katmanlı Algılayıcı), Destek Vektör Makineleri, Naive Bayes, Karar Ağaçları, Rastgele Orman Algoritması, K-En Yakın Komşu Algoritması gibi çeşitli makine öğrenmesi yöntemlerini kullanarak böbrek yetmezliğini etkileyen faktörleri sınıflandırmaktır. Bu çalışmada, Ankara Numune Hastanesi'nde acil servise gelen, 18 yaşından büyük ve üst gastrointestinal kanama belirtileri bulunan 237 hasta seçilmiştir. Burada makine öğrenmesi yöntemleri ile sınıflandırma yapmak için böbrek yetmezliğini etkileyen yaş, cinsiyet, kan değerleri, diğer hastalıklar vb. gibi 34 değişken kullanılmıştır. Makine öğrenmesi yöntemleri doğruluk oranları, F-ölçütü, duyarlılık, özgüllük ve Kappa değerlerine göre karşılaştırıldığında, karar ağaçları algoritmasının iyi performans gösterdiği bulunmuştur.

1. Introduction

Machine learning methods are widely used in many other fields such as health, industry, agriculture, aviation, quality control, insurance etc [1;2;3;4;5;6].

In addition, these methods have attracted great interest in different branches of health such as psychiatry, gastroentrology, diabetes, renal diseases, cancer prediction etc [7;8; 9;10;11;12;13].

There are many machine learning methods widely used in the literature. Some of these methods are Multilayer Perceptron-Artificial Neural Networks (MLP-ANNs), Support Vector Machines (SVM), Naive Bayes, Decision Trees, Random Forests (RF), K-Nearest Neighborhood (K-NN) algorithms. There are many studies in the field of health using these methods.

Department of Statistics, Ondokuz Mayıs University, Samsun
*Corresponding author, email: pelin.kasap@omu.edu.tr

Greco et al [10] have applied to a transplant population classification trees to build predictive models of graft failure, evaluating the interactions between body mass index and other risk factors.

Lofaro et al [9] have shown two classification trees to predict chronic allograft nephropathy, through an evaluation of routine blood and urine tests.

Tangri et al [14] purpose to compare the factors that predict peritoneal dialysis technique survival using ANNs and logistic and Cox regression methods.

Kumari and Godara [15] study data mining classification methods such as RIPPER classifier, Decision Tree, ANNs and SVM. They compare the performance of these methods on cardiovascular disease dataset.

Gupta et al [16] present an overview of the current research being carried out using the data mining methods to enhance the breast cancer diagnosis and prognosis. They aim to summarize various review and technical articles on diagnosis and prognosis of breast cancer.

Krishnaiah et al [17] examine the use of classification based data mining methods such as Rule based, Decision tree, Naïve Bayes and ANNs on lung cancer data. They aim to propose a model for early detection and correct diagnosis of the disease.

Martínez-Martínez et al [11] aim to predict the hemoglobin value using a database of European hemodialysis patients provided by Fresenius Medical Care for improving the treatment of this kind of patients. They use classification methods such as Linear Models, ANNs, SVM and Regression Trees.

Kunwar et al [18] aim to predict chronic kidney disease using classification methods such as Naive Bayes and ANNs.

Aziz and Rehman [19] introduce the application of decision tree classification method for the detection of heart disease.

Davazdahemami and Delen [20] investigate using data mining approach the general confounding role of the common diabetes medications in developing acute renal failure in a large group of patients with diabetes mellitus.

Mezzatesta et al [12] aim at predicting death and cardiovascular diseases in dialysis patients using machine learning methods.

Renal failure is disease condition in which kidney fails to function properly resulting in elevation of serum Creatinine because of decrease in glomerular filtration rate [21]. Patients with acute renal failure are often more prone to as gastrointestinal complications such as bleeding. Increased bleeding sensitivity in patients may be due to different reasons such as other critical diseases, inflammation, hemodialysis use, etc [22; 23]. A review about renal diseases is given Table 1.

In this study, 237 patients older than 18 years with symptoms of upper gastrointestinal bleeding who were admitted to the emergency department of Ankara Numune Hospital were selected. Here, 34 different variables such as age, gender, blood values, other diseases, etc, that affect renal failure, are investigated in order to classify with machine learning methods. The aim of this study is to compare the six machine learning methods such as Support Vector Machines, Naive Bayes, Decision Trees, Random Forests, K-Nearest Neighborhood algorithms used for classification on renal failure data set. Thus, as a result of the comparison of the methods, the factors that affect renal failure are classified using the method that best meets the criteria such as accuracy rate, F-measure, sensivity, specifity, Kappa value.

The organization of the paper is as follows: In Section 2, we briefly introduce machine learning methods such as SVM, Naive Bayes, Decision Trees, RF, K-NN algorithms. In Section 3, we give attributes of renal failure data set and results of classification methods. Finally, we interpret the obtained results in Section 4.

Table 1. A Review about renal diseases

Authors	Publication year	Topic	Algorithms
Vijayarani et al [24]	2015	Kidney disease	SVM, ANNs
Vijayarani and Dhayanand [25]	2015	Kidney disease	SVM, Naive Bayes
Subasi et al [26]	2017	Chronic Kidney Disease	RF, ANNs, SVM, K-NN, C4.5 Decision Tree.
Kunwar et al [18]	2016	Chronic Kidney Disease	Naive Bayes, ANNs
Ani et al [27]	2016	Chronic renal failure	Back propagation ANNs, Naive Bayes, LDA classifier, K-NN, Random subspace classification.
Sinha and Sinha [28]	2015	Chronic Kidney Disease	K-NN, SVM
Al-Hyari et al [29]	2014	Chronic renal failure	ANNs, Decision tree, Naive Bayes, SVM, Logistic regression.
Polat et al [30]	2017	Chronic Kidney Disease	SVM
Khitan et al [31]	2017	Chronic Kidney Disease	SVM Generalized linear model, Decision tree, ANNs, RF.
Barbieri et al [32]	2015	Renal disease	ANNs
Dimitrov et al [33]	2003	Renal disease	Decision tree models, Bayesian approach.

2. Material and Method

There are various machine learning methods used in the field of health. Each of these methods has advantages and disadvantages. Classification is one of the most important tasks in machine learning. There are different machine learning methods used for classification. In this section, these methods are briefly introduced.

2.1. Decision Trees

Decision trees are one of the most widely used data mining classification methods. Decision trees are a tree structure consisting of roots, branches and leaves [16]. Each node in a decision tree represents the property of the sample to be classified and each branch represents the value in the node. The decision trees start with the root and move through the branches to the leaves at the bottom by sorting their attribute values [34].

Many algorithms such as ID3, C4.5, C5, Sliq, Sprint, Classification and Regression Tree (CART), Chi-squared Automatic Interaction Detection (CHAID), Logistic Model Tree can be used in decision trees. The most commonly used mathematical algorithms for splitting in decision tree are Entropy, Chi-squared test and Gini index. In ID3, C4.5 and C5 algorithms are used Entropy based information gain; in CART algorithm is used Gini index and in CHAID algorithm is used Chi-squared test [15]. Entropy formula to be used to calculate information gain is given in Equation (1), when data set consists of several classes such as $C_1, C_2, \dots, C_n$ and T represents class values, the probability of a class is $p_i = C_i/|T|$ and Entropy formula

$$Entropy\ (T) = -\sum_{i=1}^{n}(p_i * log_2(p_i)). \quad (1)$$

Gini index formula is given in Equation (2), when p_i is the probability of an object being classified to a particular class [35]

.

$$Gini = 1 - \sum_{i=1}^{n}({p_i}^2). \quad (2)$$

After the data is referred to one of these decision trees algorithms, the data is processed in the algorithm and the decision tree is created. This created decision tree class is applied on unknown data to determine the classes of this data is provided [36].

For the basic algorithm in decision trees, firstly a learning set is created. The attribute that best distinguishes the samples in this learning set is determined. With the selected attribute, the node of the tree is created and subnodes and leaves are created. The instances in the data set on these child nodes are identified. For each subset of data that is created, the process is terminated if all instances belong to the same class or there are no attributes left to divide the instances, or there are no instances containing the value of the remaining attributes. Otherwise, the best distinguishing attribute is determined and operations are continued [37]. One of important issue in decision trees formation is the pruning of the formed trees structure. Decision trees algorithm divides training data into subsets containing only one class, so a complex tree structure emerges. Therefore, it may be possible to place a leaf in the decision tree instead of discarding a subtree. Pruning removes portions of the decision tree that do not affect or contribute to the classification accuracy [35;42;43].

2.2. Random Forest

Random forest (RF) algorithm, which can be defined as tree type classifiers, is one of the ensemble classification methods. RF separates each node into branches using the best of the randomly selected variables in each node [44]. After each data set is generated interchangeably from the original data set, trees are developed using random selection and has no pruning. This is the reason why random forest algorithm is faster and more accurate than other algorithms. RF is also resistant to excessive compliance and can be used with more trees [45].

RF uses the CART algorithm to produce trees and at each node, the branches are formed according to the criteria of the CART algorithm [46]. In order to create the tree structure in the algorithm, the samples to beused in each node and the number of trees to be created must be determined. The number of trees in the decision forest is determined by users. When a new object is to be classified, it is processed by the determined decision trees and the determination of the class is made by selecting the highest rate obtained from each tree [47].

"Gini index" is the test criterion of the selected variable for branching. If the training and test data are not predetermined, 2/3 and 1/3 of the entire data set are taken into training data and test data, respectively. For decision trees that will form the decision forest, the samples are created using the number of trees bootstrap technique and the test and training data are separated for each sample. All trees are calculated error rate with the test data and then the test error of the decision forest is calculated by averaging these error rates. A weight is given to all trees according to the test error rate and weight and test error rate are inversely proportional. That means Decision tree with the highest error rate has lowest weight, the lowest error rate has the highest weight. All trees are subjected to a voting process for the classification process according to the determined weights. The tree with the highest rating is determined as the class estimation [48;49].

2.3. Naive Bayes

The Naive Bayes Classifier, which is a simplified version of Bayes' theorem by the independence theorem, also known as the simple Bayes classification algorithm [50]. The Naive Bayes Classifier, one of the Machine Learning algorithms, represents a supervised learning algorithm as well as a statistical method for classification. Test data is classified according to these predefined classes. For each test data, the probability of entering the predefined sample classes is calculated and classified according to the highest probability value [51].

The most important difference between the Naive Bayes classifier and the Bayes Theorem is that the classifier focuses on finding the target class rather than the probability value. The value in the denominator in the Bayes Theorem is neglected in the Naive Bayes Classifier since it is common in the probability calculations of all target states [52]. When $X = \{x_1, x_2, \dots, x_n\}$ be a set of n attributes, X is considered as evidence and H be some hypothesis means, the data of X belongs to specific class C, according to the Bayes theorem probability that the hypothesis H holds given evidence i.e. data sample X, $P(H|X)$, is given as follows:

$$P(H|X) = P(X|H) * P(H)/P(X) \quad (3)$$

2.4. Artificial Neural Networks

Artificial neural networks (ANNs), inspired by the work of the human brain, are one of the research field of artificial intelligence brain [53]. Artificial neural network (ANN) has many superior features compared to other systems. They are capable of modelling not only linear, but also complex nonlinear functions. Also, they have ability to learn from examples, self-organization, ability to quickly calculate, ability to find complex relationships between input variables, no need for assumptions, and generalization [15;54].One of most widely used ANN is called multi-layer perceptron ANN (MLP-ANN) is given in Figure 1.

The MLP-ANN model consists of an input, one or more hidden layers and an output layer, and such networks are referred to as feed-forward neural network models since the information flow is continuously forward and there is no feedback.

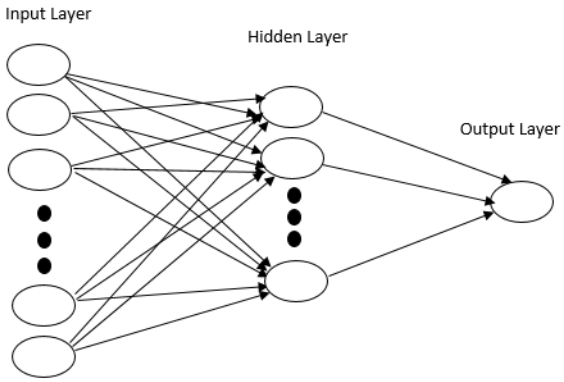

Figure 1. The structure of MLP-ANN

Back propagation algorithm is mostly used as learning algorithm in MLP-ANN. The training of the network is staged as the progression of information from the input layer to the output layer, the calculation of the error in the output units, and the backward propagation, changing the weights relative to the backward propagating error. This algorithm allows to find the weights to produce the most suitable solution for the given training set. Network training is provided by minimization of the error function in Equation (4).

$$E = \frac{1}{2}\sum_{s=1}^{S}(t_s - o_s)^2 \quad (4)$$

Here t_s, o_s are given output and output values computed by the net, respectively, S is the number of experimental data points utilized for the training and the coefficient $1/2$ is to facilitate the derivation of the function. The amount of change in weights is obtained by taking the derivative of this function [54;55].

2.5 Support Vector Machine

Support Vector Machines (SVM) was proposed by Vapnik[56] for solving of classification and regression type problem [56]. Support vector machines, one of the machine learning algorithms, are algorithms based on statistical learning theory. The aim of the support vector machines, which has a supervised learning algorithm, is to obtain the optimal separation to classify [56;57]. In other words, it is to maximize the distance between the support vectors of different classes [16].

The type of SVM used in classification applications is known as Support Vector Classification, and the type used in regression applications is known as Support Vector Regression [58]. SVM was initially used to

solve linear binary classification problems and then developed to solve non-linear or multi-class classification problems, so it is used in the solution of both linear and nonlinear problems.

SVM transforms to a linear high dimensional attribute space using kernel function to solve a nonlinear and multi-dimensional problem [59]. The data in the nonlinear input space is converted to linear high dimensional attribute space with this kernel function. It is known that the main task of the SVM approach based on the selection of these kernel functions. It can be produced different SVMs as selecting different kernel functions such as polynomial, linear, two layer neural networks etc [60].

2.6 K-Nearest Neighborhood (K-NN) algorithms

The closest neighbor algorithm proposed by Cover and Hart [61] is one of the best known, old, simple classification methods and is mostly used among machine learning algorithms [38].
Among the supervised learning methods, the K-NN algorithm uses the sample data shown earlier to determine the class of new data. Distances are calculated by processing the new data and sample data one by one [39]. These distances between the data can be calculated using various distance functions such as Manhattan Distance Function, Minkowski Distance Function, Euclidean Distance Function. k samples at the nearest distance to the new data are selected. In the set of selected samples, the new data belongs to that class if there are the most samples belonging to that class [34].
The use of different types of distances can affect the performance of the algorithm. In the literature, Euclidean distance, the most commonly used to calculate distances between samples, is given for n-dimensional distance between 2 points in Equation (5).

$$d(x,y) = \sqrt{\sum_{i=1}^{n}(X_i - Y_i)^2} \quad (5)$$

Other parameters that are effective and important in the performance of the K-NN algorithm are neighbor number (k) and weighting method. Again, the performance of the algorithm can be compared by changing the number of k and weighting methods [40].

3. Results

The data set consists of 237 records belongs to patients who have been in emergency unit in Hospital of Numune in Ankara and are larger than 18 years and have upper gastrointestinal bleeding symptoms. The dataset has 34 variables which are 6 continuous and 28 categorical in Table 2 [62].

Table 2. Attributes of renal failure data set

Number	Name	Description and values
1	Age	Age in years
2	Sex	0=male, 1=female
3	Steroid	Whether the patient uses steroid or not (0= not use, 1=use)
4	Diuretic	Whether the patient uses diuretic or not (0= not use, 1=use)
5	DM	Diabetes Mellitus (0= has not DM, 1= has DM)
6	HT	Hypertension (0= has not HT, 1= has HT)
7	CORD	Chronic obstructive respiratory disease (0= has not CORD, 1= has CORD)
8	HF	Heart failure disease (0= has not CORD, 1= has CORD)
9	CIRR	CIRRHOSIS (0= has not CIRR, 1= has CIRR)
10	MLN	Malignancy (0= has not MLN, 1= has MLN)
11	CAD	Coronary Artery Disease (0= has not CAS, 1= has CAD)
12	CVD	Cerebrovascular diseases (0= has not CVD, 1= has CVD)
13	CKD	Chronic Kidney Disease (0= has not CKD, 1= has CKD)
14	HMT	Hematemesis (0= has not HMT, 1= has HMT)
15	MEL	MELENA (0= has not MEL, 1= has MEL)
16	HMZ	Hematochezia (0= has not HMZ, 1= has HMZ)
17	GI	Gastrointestinal bloods (0= has not GI, 1= has GI)
18	Gastritis	(0= has not Gastritis, 1= has Gastritis)
19	Duodenit	(0= has not Duodenit, 1= has Duodenit)
20	VISSVESSEL	Visible Vessel (0= has not VISSVESSEL, 1= has VISSVESSEL)
21	HC	The hematin clot (0= has not HC, 1= has HC)
22	Gulcer	Gastric ulcer (0= has not Gulcer, 1= has Gulcer)
23	Dulcer	Duodenitis ulcer (0= has not Dulcer, 1= has Dulcer)
24	Pulcer	Peptic ulcer (0= has not Pulcer, 1= has Pulcer)
25	ESOP	Esophagitis (0= has not ESOP, 1= has ESOP)
26	EVAR	Esophageal varicose (0= has not EVAR, 1= has EVAR)
27	EGAST	Erosive gastritis (0= has not EGAST, 1= has EGAST)
28	MAL-A	Malignancy-A (0= has not MAL-A, 1= has MAL-A)
29	NA	Sodium (gr/dl)
30	ALB	Albumin (gr/dl)
31	HB	Hemoglobin (gr/dl)
32	WBC	White Blood Cells (k/ul)
33	CR	Creatine (mg/dl)
34	RF	Renal Failure (0= has not RF, 1= has RF)

The total records are divided into two datasets, one is used for training (60%, 70%, 75%, 80% of total records), another is used for testing (40%, 30%, 25%, 20% of total records). The 34th variable RF in Table 1 is predictable attribute and other 33 variables are input variables. Before classifying, 1th variable (age) and between numbers of 29-33 numerical variables have been standardized as [0,1], because the type of variables is an integer. Machine learning methods such as MLP-ANNs, Support Vector Machines (SVM), Naive Bayes, Decision Trees, Random Forests (RF), K-Nearest Neighborhood (K-NN) algorithms have been applied for classifying RF depending on input variables by using programs of Rapidminer Studio and Matlab 2015b.

A confusion matrix has been obtained to calculate accuracy rate, precision, F-measure, sensivity, specifity and Kappa values. This matrix that a matrix representation of the classification results, is given Table 3.

Table 3. Confusion matrix

		Actual Values	
		Positive (1= has RF)	Negative (0=has not RF)
Predicted Values	Positive (1= has RF)	TP	FP
	Negative (0=has not RF)	FN	TN

TP (True Positive): Observation is actually positive, and is predicted to be positive.
FN (False Negative): Observation is actually positive, but is predicted negative.
FP (False Positive): Observation is actually negative, but is predicted positive.
TN (True Negative): Observation is actually negative, and is predicted to be negative.

The calculation formulas of the criteria used to measure the performance of the classification methods are given in Equation (6)-(10):

$$Accuracy\ rate = (TP + TN)/Total \quad (6)$$

$$Precision = TP/(TP + FP) \quad (7)$$

$$Sensivity = Recall = TP/(TP + FN) \quad (8)$$

$$Specifity = TN/(TN + FP) \quad (9)$$

$$F - measure = (2 * Recall * Precision)/(Recall + Precision) \quad (10)$$

Kappa score is also used as performance criteria. It is a measure of how well the classification method actually performs. In other words, if it has a high Kappa score, it can be said that there is a big difference between the accuracy rate and the null error rate [41].

The application details of different classification methods such as RF, MLP-ANN, SVM, K-NN, Naïve Bayes and Decision Trees used in the analysis are summarized as below:

Random Forest: In RF algorithm, Gini index has been used as criterion. Optimal values of number of trees and optimal values of maximal depth have been obtained 99, 83, 34, 35 and 4, 3, 6, 3 for training set 60%, 70%, 75%, 80%, respectively.

MLP-ANN: When number of iteration is 100 and activation function is sigmoid, optimal values of learning rate and momentum coefficient have been obtained 0.02, 0.04, 0.04, 0.01 and 0.6, 0.6, 0.5, 0.5 for training set 60%, 70%, 75%, 80%, respectively. The structure of MLP-ANN has been obtained (33-19-2) as numbers of input-hidden-output layers.

SVM: Kernel types linear polynomial and radial are applied all of cases, optimal kernel type is obtained as linear. Maximum iteration number and kernel cache are selected 100000 and 200, respectively.

K-NN: When measure type is selected mixed measure and distance measure is selected Euclidean, the optimal values of k values are obtained as 9, 9, 4, 15 for training set 60%, 70%, 75%, 80%, respectively.

Naïve Bayes: Laplace correction is used for all cases. The ratio of renal failure has not 0,595 and the ratio of renal failure has 0,405 with 33 variables.

Decision Trees: CART and C4.5 algorithm have been applied. For CART algorithm, criteria, maximal depth, minimal leaf size, minimal gain, minimal size for split are selected Gini index, 10, 2, 0.01, 4, respectively. For C4.5 algorithm, when criteria is selected information gain, maximal dept (3, 6, 4, 3), significance level (0.1, 0.5, 0.5, 0.4), minimal gain (0.1, 0.1, 0.01, 0.05), minimal leaf size (1, 1, 1, 2) are obtained for training set 60%, 70%, 75%, 80%, respectively.

In order to evaluate the test set's performance of these different classification methods, the values of performance criteria, are shown in Table 4. According to Table 4, the results show that Decision Trees algorithm mostly outperforms than all other methods in all performance criteria such as sensitivity, specificity, F-measure, accuracy rate and Kappa. Additionally, RF method gives the same result as the decision trees method in terms of 80% training. Therefore, it can be said that Decision tree and Random Forest methods are effective in the classification of renal failure data. However, since Decision Trees algorithm gives better results than the RF algorithm in all training percentages, the renal failure data has been classified using the Decision trees algorithm. Therefore, final comments have been made using Decision Trees algorithm.

Table 4. The test set's performance of classification methods

Training%	Criteria	Random Forest	MLP-ANN	SVM	K-NN	Naive Bayes	Decision Trees
	Accuracy Rate (%)	95,74	85,11	75,53	63,83	92,55	**96,81**
	F-measure (%)	96,43	87,93	81,89	75,36	93,69	**97,30**
60%	**Sensivity (%)**	**96,43**	91,07	92,86	92,86	92,86	**96,43**
	Specifity (%)	94,74	76,32	50,00	21,05	92,11	**97,37**
	Kappa	0,912	0,685	0,458	0,156	0,846	**0,934**
	Accuracy Rate (%)	94,37	88,73	80,28	64,79	90,14	**97,18**
	F-measure (%)	95,35	90,70	85,71	75,73	91,95	**97,62**
70%	**Sensivity (%)**	97,62	92,86	**100,00**	92,86	95,24	97,62
	Specifity (%)	89,66	82,76	51,72	24,14	82,76	**96,55**
	Kappa	0,882	0,764	0,559	0,189	0,793	**0,942**
	Accuracy Rate (%)	94,92	89,83	81,36	59,32	89,83	**96,61**
	F-measure (%)	95,77	91,89	86,42	70,73	91,43	**97,14**
75%	**Sensivity (%)**	97,14	97,14	**100,00**	82,86	91,43	97,14
	Specifity (%)	91,67	79,17	54,17	25,00	87,50	**95,83**
	Kappa	0,894	0,784	0,584	0,085	0,789	**0,930**
	Accuracy Rate (%)	**95,74**	91,49	78,72	72,34	93,62	**95,74**
	F-measure (%)	**96,43**	93,33	84,85	81,16	94,74	**96,43**
80%	**Sensivity (%)**	96,43	**100,00**	**100,00**	**100,00**	96,43	96,43
	Specifity (%)	**94,74**	78,95	47,37	31,58	89,47	**94,74**
	Kappa	**0,912**	0,817	0,517	0,355	0,866	**0,912**

Firstly, the standardized values are converted to the original values for the evaluation of model. The structure of decision tree is given in Figure 2 and the cases showing the results of the classification are explained according to Figure 2.

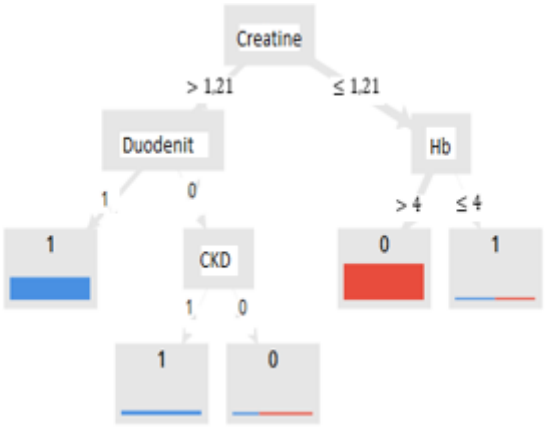

Figure 2. The structure of decision trees

Case 1: Patients with creatinine level greater than 1.21 mg/dl and having duodenitis are included in the class of patients with renal failure.

Case 2: Patients with creatinine level greater than 1.21 mg/dl and having not duodenitis, but having Chronic Kidney disease, they are included in the class of patients with renal failure.

Case 3: Patients with creatinine level greater than 1.21 mg/dl, having not duodenitis and having not Chronic Kidney disease, they are included in the class of patients having not renal failure with more than ratio of 50%.

Case 4: Patients with creatinine level equal and less than 1.21, having Hb values greater than 4 gr/dl, they are included in the class of patients having not renal failure, else they have renal failure with almost %50 ratio.

Since the best classification results is Decision Trees algoritm with 70% training set, Roc Curve for this case is shown in Figure 3. It can be seen that the correct positive ratio is high and the area under the curve is high (AUC=0,993). Therefore, it can be seen whether the positives have successfully separated from the negatives.

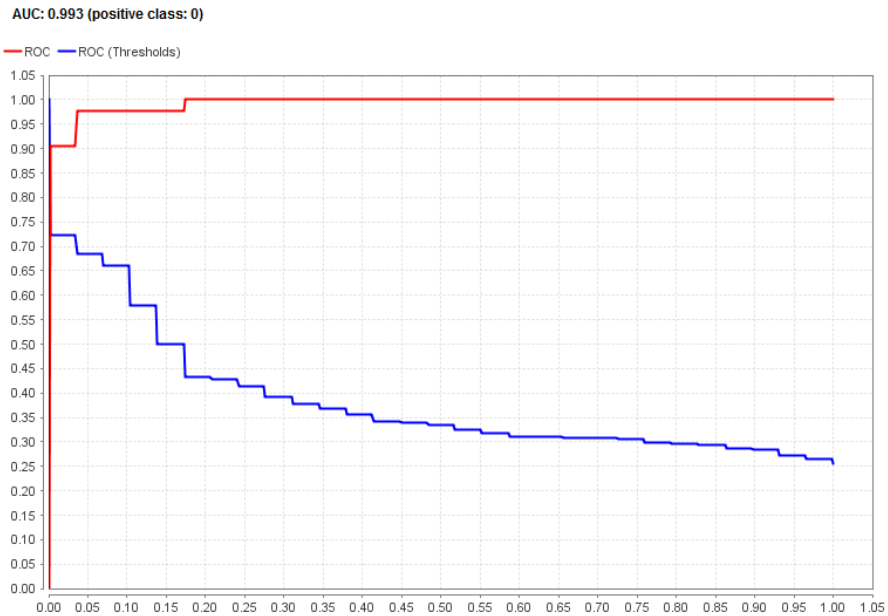

Figure 3. Roc curve for decision trees algoritm with %70 training

4. Discussion and Conclusion

In this study, the data set consists of 237 records of patients older than 18 years who have been in emergency department in Hospital of Numune in Ankara and have upper gastrointestinal bleeding symptoms. The data set has 34 variables which are 6 continuous and 28 categorical. By using these real dataset, it is aimed to compare six machine learning methods commonly used in the literature such as MLP-ANNs, SVM, Naive Bayes, Decision Trees, Random Forests, K-NN algorithms. Firstly, the best performing method is determined in terms of accuracy rate, F-measure, sensivity, specifity and Kappa criteria for renal failure dataset. For this data set, the best method according to these criteria is found to be the decision trees method. Then, renal failure data is classified using decision trees method. Therefore, it can be said that patients with creatinine level above 1.21 mg/dl and with duodenitis are included in the class of patients with renal failure. Patients with creatinine level above 1.21 mg/dl and with non-duodenitis and having chronic renal failure disease are included in the class of patients with renal failure. Also, patients with creatinine level greater than 1.21 mg/dl, with non-duodenitis and having not Chronic Kidney disease, they are included in the class of patients having not renal failure with more than ratio of 50%. Patients with creatinine level equal and less than 1.21, having Hb values greater than 4 gr/dl, they are included in the class of patients having not renal failure, else they have renal failure with almost %50 ratio.

References

[1] Schultz, M., Reitmann, S. 2019. Machine learning approach to predict aircraft boarding. Transportation Research Part C: Emerging Technologies, 98, 391-408.

[2] Maheshwari, A., Davendralingam, N., DeLaurentis, D. A. 2018. A Comparative Study of Machine Learning Techniques for Aviation Applications. In 2018 Aviation Technology, Integration, and Operations Conference p. 3980.

[3] Gümüşçü, A., Tenekeci, M. E., Bilgili, A. V. 2019. Estimation of wheat planting date using machine learning algorithms based on available climate data. Sustainable Computing: Informatics and Systems.

[4] Rehman, T. U., Mahmud, M. S., Chang, Y. K., Jin, J., Shin, J. 2019. Current and future applications of statistical machine learning algorithms for agricultural machine vision systems. Computers and Electronics in Agriculture, 156, 585-605.

[5] Burri, R. D., Burri, R., Bojja, R. R., Buruga, S. 2019. Insurance Claim Analysis using Machine Learning Algorithms. International Journal of Advanced Science and Technology, 127(1), 147-155.

[6] Ferreiro, S., Sierra, B., Irigoien, I., Gorritxategi, E. 2011. Data mining for quality control: Burr detection in the drilling process. Computers & Industrial Engineering, 60(4), 801-810.

[7] Adadi, A., Adadi, S., Berrada, M. 2019. Gastroenterology Meets Machine Learning: Status Quo and Quo Vadis. Advances in Bioinformatics.

[8] Librenza-Garcia, D., Kotzian, B. J., Yang, J., Mwangi, B., Cao, B., Lima, L. N. P. Passos, I. C. 2017. The impact of machine learning techniques in the study of bipolar disorder: a systematic review. Neuroscience &Biobehavioral Reviews, 80, 538-554.

[9] Lofaro, D., Maestripieri, S., Greco, R., Papalia, T., Mancuso, D., Conforti, D., Bonofiglio, R. 2010. Prediction of chronic allograft nephropathy using classification trees. In Transplantation proceedings, Vol. 42, No. 4, pp. 1130-1133, Elsevier.

[10] Greco, R., Papalia, T., Lofaro, D., Maestripieri, S., Mancuso, D., Bonofiglio, R. 2010. Decisional trees in renal transplant follow-up. In Transplantation proceedings, Vol. 42, No. 4, pp. 1134-1136, Elsevier.

[11] Martínez-Martínez, J. M., Escandell-Montero, P., Barbieri, C., Soria-Olivas, E., Mari, F., Martínez-Sober, M. Stopper, A. 2014. Prediction of the hemoglobin level in hemodialysis patients using machine learningtechniques. Computer methods and programs in biomedicine, 117(2), 208-217.

[12] Mezzatesta, S., Torino, C., De Meo, P., Fiumara, G., Vilasi, A. 2019. A machine learning-based approach for predicting the outbreak of cardiovascular diseases in patients on dialysis. Computer Methods and Programs in Biomedicine, 177, 9-15.

[13] Cruz, J. A., Wishart, D. S. 2006. Applications of machine learning in cancer prediction and prognosis. Cancer informatics, 2, 117693510600200030.

[14] Tangri, N., Ansell, D., Naimark, D. 2011. Determining factors that predict technique survival on peritoneal dialysis: application of regression and artificial neural network methods. Nephron Clinical Practice, 118(2), c93-c100.

[15] Kumari, M., Godara, S. 2011. Comparative study of data mining classification methods in cardiovascular disease prediction 1, International Journal of Computer Science and Technology, Vol 2, Issue 2, 304-308.

[16] Gupta, S., Kumar, D., Sharma, A. 2011. Data Mining Classification Techniques Applied for Breast Cancer Diagnosis and Prognosis. Indian Journal of Computer Science and Engineering, 2 (2), 188-195.

[17] Krishnaiah, V., Narsimha, D. G., Chandra, D.N.S. 2013. Diagnosis of lung cancer prediction system using data mining classification techniques. International Journal of Computer Science and Information Technologies, 4(1), 39-45.

[18] Kunwar, V.,Chandel, K., Sabitha, A. S., Bansal, A. 2016. Chronic Kidney Disease analysis using data mining classification techniques. In 2016 6th International Conference-Cloud System and Big Data Engineering (Confluence), pp. 300-305. IEEE.

[19] Aziz, A., Rehman, A. U. 2017. Detection of Cardiac Disease using Data Mining Classification Techniques. IJACSA) International Journal of Advanced Computer Science and Applications, 8(7).

[20] Davazdahemami, B., Delen, D. 2019. The confounding role of common diabetes medications in developing acute renal failure: A data mining approach with emphasis on drug-drug interactions. Expert Systems with Applications, 123, 168-177.

[21] Kumar, A., Kumar, A., Kumar, P., Kumar, P. 2014. U.S. Patent No. 8,668,938. Washington, DC: U.S. Patent and Trademark Office.

[22] Podestà, M. A., Galbusera, M., Remuzzi, G. 2019. Bleeding and Hemostasis in Acute Renal Failure. In Critical Care Nephrology (pp. 630-635). Content Repository Only!.

[23] Lew, S. Q., Ing, T. S. 2019. Gastrointestinal Problems in Acute Kidney Injury. In Critical Care Nephrology pp. 635-640, Content Repository Only!.

[24] Vijayarani, S., Dhayanand, S., & Phil, M. 2015. Kidney disease prediction using SVM and ANN algorithms. International Journal of Computing and Business Research (IJCBR), 6(2).

[25] Vijayarani, S., & Dhayanand, S. 2015. Data mining classification algorithms for kidney disease prediction. International Journal on Cybernetics & Informatics (IJCI), 4(4), 13-25.

[26] Subasi, A., Alickovic, E., & Kevric, J. 2017. Diagnosis of chronic kidney disease by using random forest. In CMBEBIH 2017 (pp. 589-594). Springer, Singapore.

[27] Ani, R., Sasi, G., Sankar, U. R., & Deepa, O. S. 2016. Decision support system for diagnosis and prediction of chronic renal failure using random subspace classification. In 2016 International Conference on Advances in Computing, Communications and Informatics (ICACCI) (pp. 1287-1292). IEEE.

[28] Sinha, P., & Sinha, P. 2015. Comparative study of chronic kidney disease prediction using KNN and SVM. International Journal of Engineering Research and Technology, 4(12), 608-12.

[29] Al-Hyari, A. Y., Al-Taee, A. M., & Al-Taee, M. A. 2014. Diagnosis and classification of chronic renal failure utilising intelligent data mining classifiers. International Journal of Information Technology and Web Engineering (IJITWE), 9(4), 1-12.

[30] Polat, H., Mehr, H. D., & Cetin, A. 2017. Diagnosis of chronic kidney disease based on support vector machine by feature selection methods. Journal of medical systems, 41(4), 55.

[31] Khitan, Z., Shapiro, A. P., Shah, P. T., Sanabria, J. R., Santhanam, P., Sodhi, K., ... & Shapiro, J. I. 2017. Predicting adverse outcomes in chronic kidney disease using machine learning methods: data from the modification of diet in renal disease. Marshall Journal of Medicine, 3(4), 67.

[32] Barbieri, C., Mari, F., Stopper, A., Gatti, E., Escandell-Montero, P., Martínez-Martínez, J. M., & Martín-Guerrero, J. D. 2015. A new machine learning approach for predicting the response to anemia

treatment in a large cohort of end stage renal disease patients undergoing dialysis. Computers in biology and medicine, 61, 56-61.

[33] Dimitrov, B. D., Perna, A., Ruggenenti, P., & Remuzzi, G. 2003. Predicting end-stage renal disease: Bayesian perspective of information transfer in the clinical decision-making process at the individual level. Kidney international, 63(5), 1924-1933.

[34] Alpaydın, E. 2004. Introduction to Machine Learning, MIT Press.

[35] Kavzoğlu, T., Çölkesen, İ., 2010. Karar Ağaçları ile Uydu Görüntülerinin Sınıflandırılması: Kocaeli Örneği. Electronic Journal of Map Technologies, 2 (1), 36-45.

[36] Aksu, M. Ç., Karaman, E. 2017. KararAğaçları ile Bir Web Sitesinde Link Analizi ve Tespiti. Acta Infologica, 1 (2), 84-91.

[37] Phyu, N. T. 2009. Survey of Classification Techniques in Data Mining. IMECS 2009, March 18-20, Hong Kong.

[38] Qiu, X.Y., Kang, K., Zhang, H.X. 2008. Selection of kernel parameters for K-NN, IEEE International Joint Conference on Neural Networks (IJCNN), 61-65.

[39] Peterson, L. E. 2009. K-nearest neighbor. Scholarpedia, 4(2), 1883.

[40] Islam, M. J., Wu, Q. J., Ahmadi, M., Sid-Ahmed, M. A. 2007. Investigating the performance of naive-bayes classifiers and k-nearest neighbor classifiers. November, 2007. 2007 International Conference on Convergence Information Technology (ICCIT 2007), 1541-1546. IEEE.

[41] Dangare, C. S., Apte, S. S. 2012. Improved study of heart disease prediction system using data mining classification techniques. International Journal of Computer Applications, 47(10), 44-48.

[42] Quinlan, J.,R., 1987. Simplifying decision trees, International Journal of Man-Machine Studies, 27, 221-234

[43] Özkan, Y., 2008. Veri Madenciliği Yöntemleri. Papatya Yayıncılık Eğitim, İstanbul, 216.

[44] Breiman, L. 2001. Random Forests. Machine Learning, 45(1), 5- 32. doi: 10.1023/A:1010933404324

[45] Breiman, L., & Cutler, A. 2005. Random Forests. Berkeley.

[46] Han, J., Pei, J., Kamber, M. 2011. Data mining: concepts and techniques. Elsevier.

[47] Pal, M. 2005. Random forest classifier for remote sensing classification. International Journal of Remote Sensing, 26 (1), 217–222.

[48] Akman, M., Genç, Y., Ankaralı, H. 2011. Random Forests Methods and an Application in Health Science. TurkiyeKlinikleri Journal of Biostatistics, 3(1), 36-48.

[49] Karakoyun, M., Hacıbeyoğlu, M. 2014. Biyomedikal Veri Kümeleri ile Makine Öğrenmesi Sınıflandirma Algoritmalarının İstatistiksel Olarak Karşılaştırılması. Dokuz Eylül Üniversitesi Mühendislik Fakültesi Fen ve Mühendislik Dergisi, 16 (48), 30-42.

[50] Lewis, D. D. 1992. An evaluation of phrasal and clustered representations on a text categorization task. In Proceedings of the 15th annual international ACM SIGIR conference on Research and development in information retrieval, June, ACM, 37-50.

[51] Dimitoğlu, G., Adams, J. A., Jim, C. M. 2012. Comparison of the C4.5 and a Naïve Bayes classifier for the prediction of lung cancer survivability. arXiv preprint arXiv, 1206.1121.

[52] Orhan, U., Adem, K. 2012. Naive Bayes Yönteminde Olasılık Çarpanlarının Etkileri. Elektrik Elektronik ve Bilgisayar Mühendisliği Sempozyumu, 723.

[53] Haykin, S. 1999. Neural networks: A comprehensive foundation, Prentice-Hall, New Jersey.

[54] Öztemel, E. 2012. Yapay Sinir Ağları,Üçüncü Basım, Papatya Yayıncılık, İstanbul.

[55] Kartolopoulos, S. V. 1996. Understanding neural network and fuzzy logic, IEEE Press., New York.

[56] Vapnik, V.N. 1995. The Nature of Statistical Learning Theory. Springer-Verlag, New York.

[57] Han, J.,Kamber, M. 2006. Data Mining Concepts and Techniques. 2nd Edition, Morgan Kaufmann, San Francisco.

[58] Güner, N., Çomak, E. 2011. Mühendislik Öğrencilerinin Matematik I Derslerindeki Başarısının Destek Vektör Makineleri Kullanılarak Tahmin Edilmesi. Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi, 17 (2), 87-96.

[59] Huang, T. M., Kecman, V., Kopriva, I. 2006. Kernel Based Algorithms For Mining Huge Data Sets: Supervised, Semi-Supervised, and Unsupervised Learning. Studies in Computational Intelligence, Secaucus, NJ, Springer, USA.

[60] Tan, Y., Wang, J. 2004. A support vector machine with a hybrid kernel and minimal Vapnik-Chervonenkis dimension. IEEE Transactions on knowledge and data engineering, 16(4), 385-395.

[61] Cover, T., Hart, P., 1967. Nearest neighbor pattern classification. IEEE transactions on information theory, 13(1), 21-27.

[62] Cakmak, U., Merhametsiz, O., GokOguz, E., Ercan, Z., Haspulat, A., Ozkan, S. K., ... & Deniz Aylı, M. 2016. Effects of acute kidney injury on clinical outcomes in patients with upper gastrointestinal bleeding. Renalfailure, 38(2), 176-184.