KSÜ Mühendislik Bilimleri Dergisi, 25(4), 2022
Araştırma Makalesi

KSU J Eng Sci, 25(4), 2022
Research Article

Kahramanmaras Sutcu Imam University
Journal of Engineering Sciences

Geliş Tarihi : 18.07.2022
Kabul Tarihi : 26.09.2022

Received Date : 18.07.2022
Accepted Date : 26.09.2022

EVALUATION OF PERFORMANCE OF CLASSIFICATION ALGORITHMS IN PREDICTION OF HEART FAILURE DISEASE

KALP YETMEZLİĞİ HASTALIĞININ TAHMİN EDİLMESİNDE SINIFLANDIRICI ALGORİTMALARININ PERFORMANSLARININ DEĞERLENDİRİLMESİ

Cevdet COŞKUN[1] (ORCID: 0000-0002-1351-0590)
Fatma KUNCAN[1*] (ORCID: 0000-0003-0712-6426)

[1] Siirt University, Department of Computer Engineering, Siirt, Turkey

*Corresponding Author: Fatma KUNCAN, fatmakuncan@siirt.edu.tr

ABSTRACT

Success rates and performances of Gaussian Naive Bayes, Support Vector Machines, Linear Discriminant Analysis, Decision Tree and Random Forest classifier algorithms from machine learning methods were evaluated using the Heart Failure Prediction dataset. Label encoder method was used primarily in data preprocessing techniques on the data set. Catalog data (5 pieces) in the data set have been converted into numerical data. In addition, it was observed that there were negative values in the data in a field and this situation was converted to values in the range of 0 - 1 with min-max conversion methods. After the pre-processing, analyzes were made with classification algorithms. As a result of these analyzes, a success rate of 90.76% was achieved with the random forest algorithm, which is an ensemble classifier. In the study, 80% of the data was used for training and 20% for testing. Of the 184 data used for the test, 102 of them were patients with heart failure and 72 of them were from those without the disease. The success of the random forest algorithm in estimating those with heart failure disease was 93.1% (95 observations), and the success in predicting those without the disease was 87.8% (72 observations).

Keywords: Data mining, heart failure, classification algorithms, machine learning

ÖZET

Kalp Yetmezliği Tahmin veri seti kullanılarak makine öğrenmesi yöntemlerinden Gaussian Naive Bayes, Support Vector Machines, Linear Discriminant Analysis, Decision Tree ve Random Forest sınıflandırıcı algoritmalarının başarı oranları ve performansları değerlendirilmiştir. Data set üzerinde öncelikle veri ön işleme tekniklerinde label encoder yöntemi kullanılmıştır. Data setteki katalog veriler (5 adet) sayısal verilere dönüştürülmüştür. Ayrıca bir alandaki verilerde negatif değerlerin olduğu gözlemlenmiş ve bu durum min-max dönüşüm yöntemleri ile 0 - 1 aralığındaki değerlere dönüştürülmüştür. Yapılan ön işlemlerden sonra sınıflandırma algoritmaları ile analizler yapılmıştır. Bu analizler neticesinde bir ensemble (topluluk) sınıflandırıcı olan random forest algoritması ile %90,76 oranında bir başarı elde edilmiştir. Yapılan çalışmada verilerin %80'i eğitim, %20'si test için kullanılmıştır. Test için kullanılan 184 tane verinin 102 tanesi kalp yetmezliği hastalığı olanlar, 72 tanesi ise hastalığı olmayanlardan oluşmaktadır. Random forest algoritmasının kalp yetmezliği hastalığı olanları tahminlime başarısı %93,1 (95 gözlem), hastalığı olmayanları tahminlime başarısı ise %87,8 (72 gözlem) olarak gerçekleşmiştir.

Anahtar Kelimeler: Veri madenciliği, kalp yetmezliği, sınıflandırma algoritmaları, makine öğrenmesi

To Cite: COŞKUN, C., & KUNCAN, F., (2022). EVALUATION OF PERFORMANCE OF CLASSIFICATION ALGORITHMS IN PREDICTION OF HEART FAILURE DISEASE. *Kahramanmaraş Sütçü İmam Üniversitesi Mühendislik Bilimleri Dergisi*, 25(4), 1-4.

INTRODUCTION

Heart failure is a type of disease that occurs as a result of the heart's ability to pump enough blood to other organs as the heart rhythm deteriorates. Heart failure is a type of cardiovascular disease. Cardiovascular diseases (CVDs) are caused by situations such as heart attack, hypertension, heart valve disease, stress, diminished life comfort, smoking, and alcohol usage. Heart failure is mostly detected in the later stages of the disease. Other conditions that can cause heart failure include congenital chronic heart diseases, inherited heart muscle diseases, heart muscle infections, heart valve issues, and arrhythmias (Heart Failure: Investigation of an Epidemic, 2013).

Two different classes of risk factors can be defined for cardiovascular diseases: modifiable risk and unmodifiable risk. Age, gender, race, and genetic predisposition are non-modifiable risk factors, but obesity, hypertension, diabetes, hyperlipidemia, and use of tobacco and alcohol are modifiable risk factors. Modifiable risk factors can be eliminated or reduced by healthy eating, effective stress management, regular physical activity, not using (or reducing) smoking alcohol, hygienic measures, comfortable living environments, regular and quality sleep, doing sports, and taking regular walks. In addition, continuous and regular physician controls can prevent cardiovascular diseases from threatening our lives with early diagnosis in case of a possible disease.

The most common cause of death worldwide is cardiovascular disease particularly in developing nations like Turkey where a decline in infant mortality increases the risk of cardiovascular disease in adulthood. According to estimates, 42% of all fatalities in Turkey occurred as a result of coronary disease in 2001 (Onat, 2001). According to the data of the World Health Organization (WHO), cardiovascular disease caused 17.9 million deaths and 388.3 million healthy life years lost worldwide in 2019. Cardiovascular disease causes 32.2 percent of all fatalities and 15.3 percent of all years with a disability (World Health Organization, 2022b, 2022a).

Machine learning has become a valuable tool in the medical area, as well as in other fields. Studies that help doctors identify diseases have become more common these days. Especially the changes in the health sector bring the need for auxiliary resources. It is necessary to employ artificial intelligence more widely as a result. With machine learning, some diseases are easier to diagnose and treat.

In data mining, very successful results are obtained with biomedical signals, and the data is obtained accordingly. With digitalization, more data is collected in hospitals and medical centers, and these are categorically classified and brought to the scientific world for scientific research. Such developments not only enable the scientific world to progress but also make the treatment of people's diseases faster, easier, and more successful. It is an undeniable fact that the average human lifespan is getting longer with technological developments. The prolongation of human life is undoubtedly due to the developments in the health sector. At this point, positive developments are observed in the diagnosis and treatment of some diseases day by day. Artificial intelligence, data mining and machine learning methods, which are in good relations with the medical world, are being processed more and more by researchers and scientists day by day and offered to doctors in the health sector. By this means, doctors get more accurate results at the point of diagnosis and decision making.

In our study, heart failure estimation was made using classification algorithms on a dataset obtained from Kaggle. The prediction successes of patients with heart failure were interpreted through the classification algorithms used. The dataset from Kaggle is the Heart Failure Prediction Dataset (Fedesoriano, 2021). On this data set, Gaussian Naive Bayes, Support Vector Machines (SVM), Linear Separation Analysis (LDA), Decision Trees, and Random Forest classifier algorithms analyzes were evaluated using the python programming language.

The Heart Failure Prediction Dataset is a newly introduced dataset to the research world. There are not many studies in the literature with this dataset yet. However, this situation will change over time and many researchers will do different studies with this data set.

LITERATURE SURVEY

In the literature search, some studies were found and the results were examined. Srinivas and Katarya (2022), in their study, created a new classifier model using the hyperparameters of XGBoost classifiers and performed performance evaluation on this dataset and a few more similar datasets. The researchers named their newly designed method hyOPTXg. In their study, they tested their methods on three different data sets and compared their success rates with previous classification algorithms. In the analysis made with the hyOPTXg classification

algorithm, 94.7% success rate in the Cleveland dataset obtained from Kaggle, 89.3% success rate in the heart failure dataset obtained from Kaggle (the dataset we used in our study) and finally, In the cardiovascular disease data set they got from the UCI, they had an 88.5 percent success rate.

In their study, Reddy et al. (2022) applied two different classification methods on the heart failure dataset (the dataset we used in our study) obtained from Kaggle and evaluated their performance. Data was preprocessed on the data set and two of the attributes of the data set, "age and gender", were not included in the evaluation. This situation has been evaluated as the quality that can be ignored in detecting the disease. The study assessed 299 individuals' data from the data set. 200 were employed in educational settings, while 99 were used for testing. The following success rates were acquired by researchers using the Gaussian Naive Bayes and Decision Tree classification methods. They had an 86.0 percent success rate with Gaussian Naive Bayes and an 82.0 percent success rate with Decision Tree.

Coşar and Deniz (2021), in their study, used feature selection methods on the heart failure data set they obtained from Kaggle (the data set we used in our study), determining the features that will detect heart failure disease and running the classification algorithms in this way. Researchers used three different classification algorithms and evaluated their performance. The study's classification algorithms and success rates were, correspondingly, as follows. Estimates were made using the Random Forest algorithm, which had an accuracy rate of 88%, the Logistic Regression technique, which had an accuracy rate of 85%, and the kNN (K-Nearest Neighbors- nearest neighbor algorithm), which had an accuracy rate of 70%.

Çelik (2022) determined the success rates of heart failure disease by using different classification algorithms and deep neural networks (DNN) methods over the data set obtained from Kaggle. In his study, Decision Tree, Random Forest, XGBoost Algorithm, Support Vector Machines (Svm), Gradient Boosting, Proposed DNN classification algorithms were used. In the study, "age and gender" fields were extracted from the attributes of the data set and classification algorithms were tested on the remaining attributes. It was stated that the classifiers achieved higher success, especially after the age and gender fields were removed in the study. The study's findings in conjunction with the age and gender fields are presented in a comparative table with the findings without these areas. In this case, the results obtained were as follows. Decision Tree 79.34%, Random Forest 86.41%, XGBoost Algorithm 84.23%, Support Vector Machines (Svm) 89.13%, Gradient Boosting 84.78%, Proposed DNN 90.22%, respectively (including age and gender fields); (excluding age and gender fields) Decision Tree 80.97%, Random Forest 86.65%, XGBoost Algorithm 89.13%, Support Vector Machines (Svm) 88.58%, Gradient Boosting 88.58%, Proposed DNN 90.76%. With the study, it is observed that deep neural networks (DNN) achieve the highest success both with and without age and gender domains.

Ahmad and Shah (2021) carried out a study on predicting cardiovascular diseases using machine learning techniques in health centers. In this study, various classification algorithms were tested with the data set obtained in kaggle. In this study, it is mentioned that a successful result will be obtained especially with decision trees. However, a fully explanatory result could not be observed regarding this. Logistic Regression, KNeighbors Classifier, Decision Tree, Gradient Boosting Classifier and Random Forest machine learning classification models were used. 99.62% of the decision trees used in the study produced the optimal result. However, the study's details made no mention of the effectiveness of alternative classifier methods. It is unknown if any preprocessing was performed to get this result from the discriminative decision tree.

Rustam et al. (2022) studied several different heart failure datasets in a study called Combining CNN (convolutional neural network) Features to Optimize the Performance of the Ensemble Classifier for Cardiovascular Disease Prediction. One of these datasets is the Heart Failure Prediction Dataset they got from kaggle. New uses of feature extraction with a CNN are proposed in the study. An ensemble model is designed that extends the feature set of a CNN model to train machine learning models such as stochastic gradient descent classifier, logistic regression, and support vector machine. In this context, it has been observed that the proposed model works with an accuracy rate of 0.93. In the study, some transformations were made in the data preprocessing, and “F” and “M” information in the gender category and other categorical data were converted into numerical data with label coding. Data were defined for 0.20 tests and 0.80 training, and the recommended method was run. After the preprocessing with CNN, the performance of the classification algorithms was observed to be higher than before the preprocessing with CNN. Experiment results before CNN; Decision Tree 77%, ADA 83%, SVM 86%, Random Forest 85%, SGLV (recommended linear ensemble Model) 87%. After the use of CNN, the

results were observed as follows; Decision Tree 0.79, ADA 88%, DVM 88%, Random Forest 89%, SGLV (recommended Model) 92%. It was emphasized that the high success rate of this model was effective in the increase in the number of features after feature extraction with CNN.

De Silva and Kumarawadu (2022) conducted a study on the Heart Failure Prediction Dataset obtained from Kaggle with seven different machine learning methods in their study called Performance Analysis of Machine Learning Classification Algorithms in the Case of Heart Failure Prediction. The classification techniques utilized in the study were Logistic Regression, K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Naive Bayes, Decision Trees, Random Forests, and Stochastic Gradient Descent. In the study, no data preprocessing was carried out. All attributes on the data set are used. In the analyzes made in the study, it was emphasized that the Support Vector Machine (SVM) Algorithm, Naive Bayes, and Random Forests classifiers gave more successful results than other classification methods. The success rates obtained in the study were as follows. Logistic Regression 82.60%, Decision Tree 75.55%, KNN 82.61%, SVM 85.86%, Random Forest 85.32%, Stochastic Gradient Descent 79.98, Naive Bayes 58.32. As can be seen from the results, the highest success rate of 85.86% in this study belongs to the provincial SVM classification algorithm.

MATERIAL AND METHOD

In this study, the classification process of whether there is heart disease or not was performed. The 'Heart Failure Prediction Dataset' provided in the Kaggle database is given as input and output to the classifier algorithms. 193 female and 725 male participants make up the 918 data in this data collection. 508 of them have heart failure. The remaining 410 are normal people. There are 12 different attributes as input information of classifier algorithms. These include the following: age, gender, type of angina, resting blood pressure, serum cholesterol, fasting glucose levels, findings of an electrocardiogram (ECG) performed in the resting state, maximum heart rate, exercise-induced angina, depression score, and peak exercise ST segment slope. Bi-state heart disease data is defined as output in classifier algorithms (with or without heart disease). In Table 1, the characteristics and status of the data set are given in detail. 11 independent variables and 1 dependent variable can be seen when the data in the table is analyzed. The dependent variable (HeartDisease) refers to those with and without heart failure disease. 1s in this field represent heart failure disease, 0s represent no heart failure disease. The data in the dataset contains 12 attributes (columns), 918 samples (rows). Of these 12 features, 5 are categorical and 7 are numerical. The classification methods and data preprocessing methods used for data set analysis were made using the pyhton programming language.

Table 1. Data Used in the Study

Features	Values	Variable	
Age (years)	Age(numerical)	Independent	f0
Gender (M:Male, F:Female)	Sex (0,1)	Independent	f1
Type of chest pain (TA: Typical Angina, ATA: Atypical Angina, NAP: Non-Anginal Pain, ASY: Asymptomatic)	ChestPainType (0,1,2,3)	Independent	f2
Resting blood pressure [mm Hg]	RestingBP(numerical)	Independent	f3
Serum cholesterol (mg/dl)	Cholesterol l(numerical)	Independent	f4
Fasting blood glucose > 120 mg/dl (0 = false; 1 = true)	FastingBS (0,1)	Independent	f5
Resting ECG results (Normal: Normal, ST: Has ST-T wave abnormality (T wave inversions and/or ST elevation or > 0.05 mV depression), LVH: probable or indicates definitive left ventricular hypertrophy)	FastingBS (0,1,2)	Independent	f6
Maximum heart rate numerical value between 60 and 202	MaxHR (numerical)	Independent	f7
Exercise-induced angina (Y: Yes, N: No)	ExerciseAngina (0,1)	Independent	f8
Oldpeak ST Numerical value measured in depression	Oldpeak(numerical)	Independent	f9
Slope of the peak exercise ST segment (Up: curved, Straight: straight, Down: curved)	ST_Slope (0,1,2)	Independent	f10
HeartDisease: 0 = Normal 1 = heart disease	HeartDisease (0,1)	Dependent	

Classification Methods

Gaussian Naive Bayes

One of the top 10 techniques most frequently used in classification is the Naive Bayes method. It is a classifier type that can perform effective classification in various usage areas (Wu et al., 2008). This classifier, which can learn and test quickly, is a generative model-based classification algorithm (Ng and Jordan, 2001). The working principle of Bayesian classification methods is according to Bayes rule and probability theorem. The easiest version of Bayesian methods is the Naive Bayes method. The Gaussian Naive Bayes method, on the other hand, is the Naive Bayes method that works by assuming a Gaussian distribution on the input data (Ali Bagheri et al., 2012).

The mathematical formula "P(Cj|x)" of the Naive Bayes Algorithm is given below in equation 1.

p(x|Cj) : The probability that a sample from class j is x

P(Cj) : First probability of class j

p(x): probability of any sample being x

P(Cj|x) : The probability that a sample with x is of class j

$$P(Cj|x) = \frac{p(x|C_j)P(C_j)}{p(x)} = \frac{p(x|C_j)P(C_j)}{\sum_k p(x|C_k)P(C_k)} \quad (1)$$

Support Vector Machines (SVM)

A straightforward linear SVM classifier functions by drawing a straight line between the two classes. As a result, all of the data points on one side of the line will be put into one category, while the data points on the other side of the line will be put into a separate category. This implies that there are an endless number of rows from which to choose. The fact that the linear SVM method chooses the best row to categorize your data points sets it apart from other techniques like k-nearest neighbors. Choose the line that best divides the data while remaining as far away from the sample points as you can.

The complete machine learning lingo is made clear with the aid of a two-dimensional example. In essence, you have a grid of data points. To avoid placing any data in the incorrect category, you are attempting to split these data points based on the category they should belong in. The line that separates the other data points from the two nearest points, which is what this signifies, must be found.

According to the SVM algorithm, classes are usually represented as {-1,+1}. The discrimination function is created according to the education data of these two classes. This discrimination function may not always correctly classify test data. Because data do not always show a linear distribution during training. In such cases, an incorrect classification is made when the system is tested (Babur, Turhal and Akbaş, 2012).

The main purpose of SVM is to find the plane that maximizes the distance between the points closest to it. As seen in Figure 2, the hyper-leveling that makes the most appropriate distinction is called the optimum hyper-plane, and the points that limit the boundary width are called support vectors (Kavzoğlu and Çölkesen, 2010).

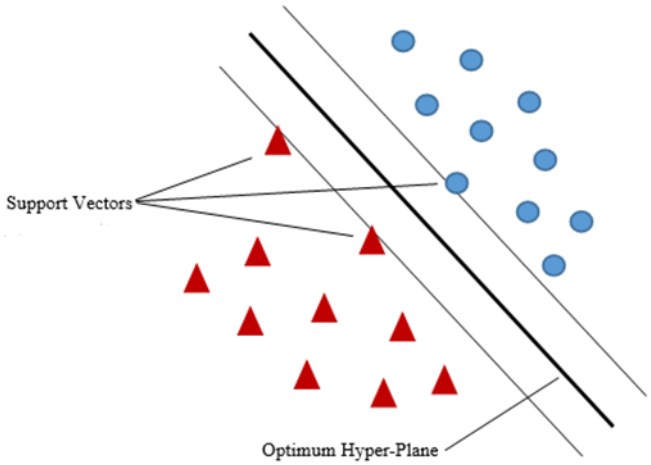

Figure 1. Optimum Hyper-Plane and Support Vectors

Decision Tree

The tree-based learning algorithm Decision Tree is regarded as one of the most popular supervised learning techniques. Methods based on trees are highly accurate, stable, and simple to interpret. They are quite good at mapping nonlinear interactions, unlike linear models. Regression or classification can be modified to address any specific issue. In many kinds of data science tasks, techniques including decision trees, random forests, and gradient reinforcement have been frequently used.

One of the most popular and useful practical approaches to inductive inference is decision tree learning. Discrete-valued target functions can be approximated using the decision tree learning technique, which uses a decision tree to represent the learned function. The majority of the time, classification problems are solved using decision trees, which are supervised learning algorithms with predefined target variables. Both categorical and continuous input and output variables can be used with it. To make learned trees more readable by humans, they can be expressed as if-then rule sets. In the shape of a tree structure, the decision tree generates classification or regression models. An associated decision tree is gradually built when a dataset is divided into smaller categories. The outcome is a tree containing leaf nodes and decision nodes. There are two or more branches on a decision node. A determination or classification is represented by the leaf node. The root node, which is the highest decision node in a tree, corresponds to the best determinant. Both category and numerical data can be processed using decision trees.

A sample data set is A_1, A_2, A_3, …. Assuming that it consists of classes in the form of A_n and n represents the class value; probability of any class is $P_i = (A_i/T|)$. The Entropy of the classes will be as in the equation 2.

$$\text{Entropy(T)} = -\sum_{i=0}^{n} P_i \log_2(P_i) \tag{2}$$

Assuming that the data set is subdivided as T class values T_1, T_2, T_3, … , T_n according to the B attribute, the gain to be obtained will be as in equation 3.

$$\text{Yield(B, T)} = \text{Entropy (T)} - \sum_{i=0}^{n} \frac{|T_i|}{|T|} \text{ Entropi}(T_i) \tag{3}$$

In Equation 4, there is the value of B attribute for T set.

$$\text{Data (B)} = -\sum_{i=0}^{n} \frac{|T_i|}{|T|} \log_2\left(\frac{|T_i|}{|T|}\right) \tag{4}$$

The rate of gain to be obtained can be obtained in equation 5.

$$\text{Yield Ratio} = \frac{Yield(B,T)}{Data(B)} \tag{5}$$

According to this formula, the branch with the highest profit is selected (Öztürk and Semra, 2021).

Random Forest

Machine learning methods like random forest are utilized to address classification and regression issues. It employs community learning, a method that integrates numerous classifications to offer answers to challenging issues. Many decision trees are used in the random forest algorithm. This algorithm trains the "forest" it creates via bagging or bootstrapping. A community meta-algorithm called bagging increases the precision of machine learning techniques. The decision trees' predictions serve as the basis for the random forest algorithm's determination of the outcome. It makes estimations by averaging the results of different trees. It becomes more precise as the number of trees grows. The decision tree algorithm's restrictions are eliminated by random forest. It improves precision and minimizes dataset overfitting.

The root nodes are formed and the nodes are parsed randomly in the random forest algorithm, which is the fundamental distinction between it and the decision tree algorithm. This method is to construct multiple trees in randomly selected subspaces of the feature space. Trees in different subspaces generalize their classifications in complementary ways, and their combined classification can be developed monotonically (Ho, 1998). The bagging method is used by random forest to produce the necessary forecast. Using multiple data samples (training data) as opposed to a single sample is known as bagging. The observations and attributes that are utilized to construct predictions are contained in a training dataset. Depending on the training data provided to the random forest algorithm, decision trees produce varied results. These results are sorted, and the best one is picked as the final result.

Random forest generates multiple decision trees during the classification process. As a result of this process, it increases the classification rate. It gives very successful results in data sets containing categorical variables with a large number of variables and class labels, and in data sets with missing data or showing an unbalanced distribution (Aydın, 2018).

Evaluation Criteria

The application data set is divided into two parts: 80 percent for training and the remaining percent for testing. This is true for all of the classifiers mentioned above. To analyze the performance of the models, a confusion matrix was utilized. The Confusion Matrix is a classification accuracy measuring tool. The matrix's primary components are indicated as true positive (TP), true negative (TN), false positive (FP), and false negative (FN) (FN). Table 2 shows the matrix representation of this circumstance.

Table 2. Confusion Matrix

		Predicted Class	
		Positive	**Negative**
True Class	**Positive**	TP	FN
	Negative	FP	TN

True Positive (TP): Situations where data with positive true class are correctly classified.
False Positive (FP): Situations where data with a negative true class is classified as false.
False Negative (FN): Situations where data with a negative true class is incorrectly classified.
True negative (TN): Cases where data with a negative true class are correctly classified.

By using the complexity matrix data, analysis such as percentage of correct classification, precision, sensitivity and f-score can be made. The percentage of correct classification is given in equation 6 below.

$$\text{Correct Classification Percentage} = \frac{TP+TN}{TP+FN+FP+TN} \quad (6)$$

Precision is obtained by dividing the predicted class positive and correctly predicted data by all positively predicted data. The calculation of the precision ratio is given in equation 7.

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \tag{7}$$

Sensitivity refers to the ratio of data with positive true class and positive predictive class to the sum of all data with positive true class. The sensitivity ratio calculation is given in equation 8.

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP+FN}} \tag{8}$$

F-score expresses the ratio found by taking the harmonic average of the results obtained in equations 7 and 8. F-score ratio calculation is given in equation 9.

$$\text{F}-\text{score} = \frac{2 \text{ x Sensitivity x Precision}}{\text{Precision+ Sensitivity}} \tag{9}$$

Data Preprocessing

Heart Failure Prediction Dataset consists of 5 categorical and 6 numerical content attributes. In the study, categorical content fields (non-numerical fields) were converted into numeric contents using the pyhton programming language. Machine learning methods give more successful results in data with numerical content. Considering this situation, fields with categorical content were converted into numerical content. The categorical fields Sex, ChestPainType, RestingECG, ExerciseAngina and ST_Slope were converted to numerical values on the data set. For this conversion, the "Label Encoder" method was used. This method is used to transform the existing categorical values into representative numerical data. Instead of data in categorical fields, it is provided to assign representative values such as 0,1,2,3,.. This method is made using the "LabelEncoder" class in the "preprocessing" library in the "sklearn package" in the pyhton programming language. Another data preprocessing operation on the data set is to make positive transformations of negative values in some of the existing data in the “old peak” area. This situation causes unsuccessful results in reading and evaluating some data in machine learning methods. To change this situation, the values in the "0-1" range of the data were transformed by making the min-max transformation in the python programming language. Thus, machine learning methods and classification algorithms on the data set provide more successful results.

CONCLUSION AND DISCUSSION

In this study, Gaussian Naive Bayes, SVM, LDA, Decision Tree, and Random Forest classifier algorithms were used using the Heart Failure Prediction Dataset. 80 percent of the data set was designated as training data, and 20 percent as test data, and analyses were conducted in accordance with these allocations. The success rates, precision, sensitivity and f-score rates obtained with the classification algorithms are given in detail in Table 3.

Table 3. Classification Algorithms Success Rate vs. Complexity Matrix Estimation Data

Serial Number	**Classification Algorithm**	**Success Rate (%)**	**Positive Class** (102 samples)		**Negative Class** (82 samples)		**Sensitivity**	**Precision**	**F- Skor**
			TP	**FN**	**FP**	**TN**			
1	Gaussian Naive Bayes	84.78	87	15	69	13	0.85	0.87	0.86
2	Support Vector Machines (SVM)	88.04	93	9	13	69	0.91	0.88	0.89
3	Linear Discriminant Analysis (LDA)	83.15	89	13	18	64	0.87	0.83	0.85
4	Decision Tree	80.97	82	20	15	67	0.80	0.83	0.82
5	Random Forest	90.76	95	7	10	72	0.93	0.91	0.92

In the study with the Python programming language, the performances of data classification algorithms were measured. In the study, 184 out of 918 data in the data set were used as test data. Of these data, 102 were selected from individuals identified as sick, and 82 from non-patients. With the study, the success rates, precision, sensitivity and f-score rates of the classification algorithms were compared. In the results obtained, the highest success rate according to the performances in the classification algorithms belongs to the Random Forest algorithm

with 90.76%. This is seen in the confusion matrix. According to the confusion matrix, the Random Forest algorithm correctly predicts 93.1% (95 data) of the real patients. It also correctly predicts 87.8% (72 data) of people who are not sick. Confusion matrices data of classification algorithms are given in Figure 2, respectively.

A more successful result was obtained in the analysis compared to previous studies. The main reason for the emergence of this situation is considered to be the transformation operations performed in the data preprocessing. Machine learning methods give more successful results when numerical data are used instead of categorical data. Random forest algorithms achieved higher results than other methods. The high success rates also show themselves in the data in the confusion matrix. Likewise, in f-score and sensitivity, the situation gave better results than other classification techniques.

The highest score obtained in previous studies belongs to the study by Çelik (2022), who developed a method based on deep neural networks. In the study, when all the features in the data set were included, the success rate was observed as 90.22% with the Proposed DNN classification algorithm. When the age and gender fields of some data belonging to the data set were excluded from the evaluation, the success rate was measured as 90.76% with the Proposed DNN classification algorithm.

In our study, our success rate was measured as 90.76%, which is the highest score, as in the study by Çelik, which we mentioned above. Success rates were measured the same in the studies, but there are some fundamental differences in both studies. In the study conducted by Çelik, the success rate is measured as 90.22% when all fields in the data set are included. It achieved a success rate of 90.76% by excluding age and gender fields in the data set. In our study, all the qualities in the data set were evaluated and the success rate was measured as 90.76% in this way. Undoubtedly, data preprocessing steps were effective in achieving this success. Another factor is that the random forest method, which is an ensemble learning method, divides the dataset into subsets and repeats the process repeatedly to validate and upgrade the predictions with a sequence decision tree classifier.

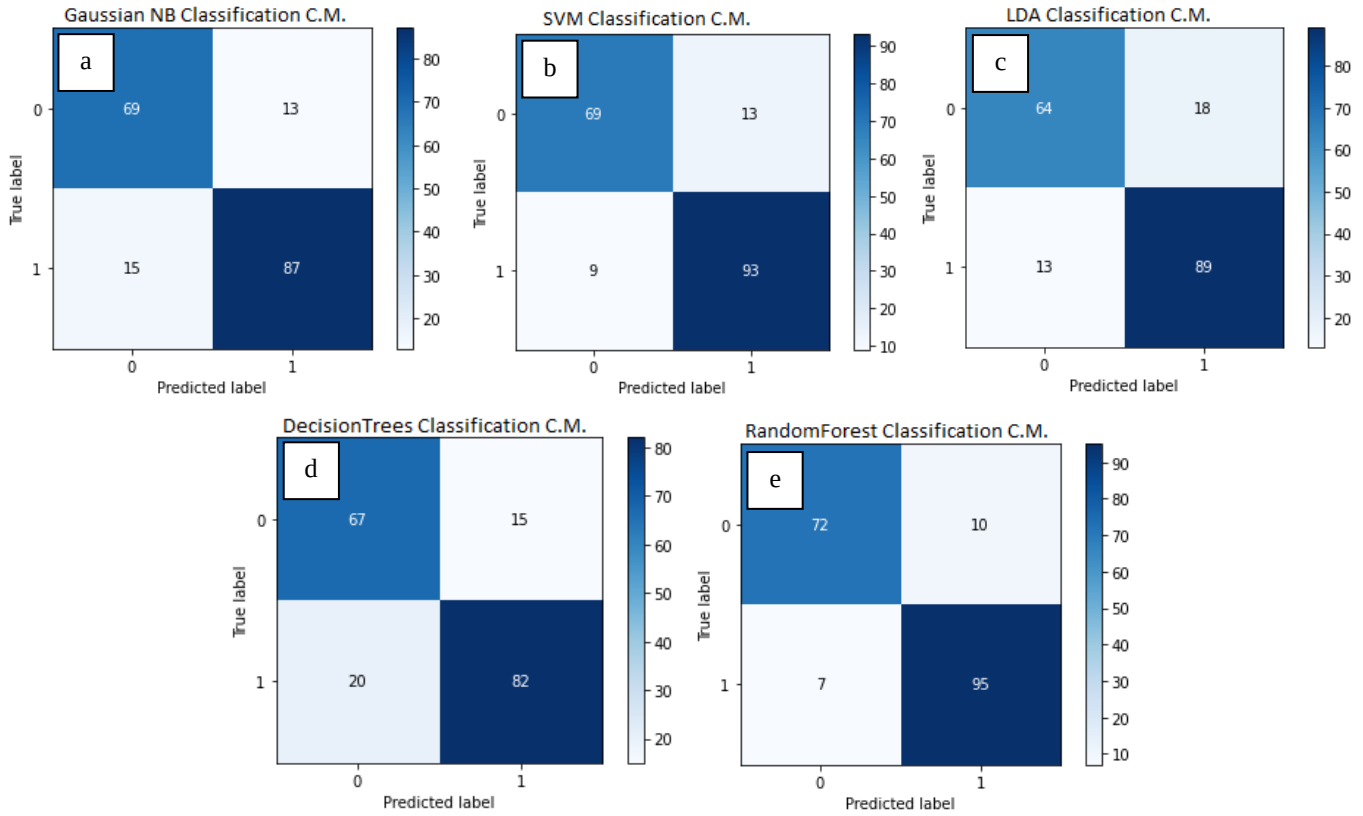

Figure 2. a. Gaussian Naive Bayes Classification Method Confusion Matrix **b.** SVM Classification Method Confusion Matrix **c.** LDA Classification Method Confusion Matrix **d.** Decision Trees Classification Method Confusion Matrix **e.** Random Forest Classification Method Confusion Matrix

In this study, it was aimed to help doctors in the field of health in the prediction of heart failure. Clinical findings such as age, gender, chest pain type, resting blood pressure, cholesterol, fasting blood glucose, resting

electrocardiogram, maximum heart rate, exercise-induced angina, and peak exercise St segment slope were used in the heart failure prediction dataset. In the estimation of heart failure, 184 out of 918 samples in the data set were randomly selected and the most successful ratio was obtained from these samples. When compared to other ensemble learning approaches, the random forest method, which is one of them, produced more accurate predictions. The programming language Python was used to conduct this study. 5 different classification algorithms were run on the same programming language and their performances were measured. The highest success rate in the measurements was obtained with the Random Forest algorithm with 90.76%. The f-score value of the random forest algorithm was also 0.92. When the results were analyzed, it was found that machine learning techniques produced better outcomes when applied to numerical data. The results were better than those of the previous investigations after the data in the catalog were converted into numerical data using the data preprocessing on the data set. In light of this, we can conclude that machine learning techniques provide outcomes with more accuracy when dealing with data with numerical content.

In today's world, as in many sectors, machine learning methods, data mining and artificial intelligence are becoming very important for the health sector. It becomes an ancillary factor to which doctors might commonly refer while making a diagnosis. The primary goal of this study is to assist clinicians in making shorter and more precise judgments of heart failure disease, which has a high prevalence in cardiovascular disorders. Without a doubt, scientific research and studies will advance with time and produce more reliable findings. As the success rate in the studies increases, the solutions needed will increase accordingly. Each new academic study will lead to a different study in the scientific world. Future research can employ artificial intelligence to detect certain diseases early. Automatic identification of heart problems and many other diseases is possible with greater and more insightful data. This makes it possible to identify diseases early and combat them with appropriate means.

ACKNOWLEDGEMENT

This study was carried out in Siirt University Engineering Faculty Human-Computer Interaction Laboratory. The authors of this article thank the Human-Computer Interaction Laboratory staff for their support.

REFERENCES

Ahmad, H. G., & Shah, M. J. (2021). Prediction of Cardiovascular Diseases (CVDs) Using Machine Learning Techniques in Health Care Centers, *Azerbaijan Journal of High Performance Computing,* 4(2), 267-279

Ali Bagheri, M., Montazer, G. A., & Escalera, S. (2012). Error correcting output codes for multiclass classification: application to two image vision problems. The 16th CSI International Symposium on Artificial Intelligence and Signal Processing (AISP 2012), 508–513.

Aydın, C. (2018). Makine öğrenmesi algoritmaları kullanılarak itfaiye istasyonu ihtiyacının sınıflandırılması. *Avrupa Bilim ve Teknoloji Dergisi*, (14), 169-175.

Babur, S., Turhal, U., & Akbaş, A. (2012). DVM Tabanlı Kalın Bağırsak Kanseri Tanısı İçin Performans Geliştirme Performance Improvement for SVM Based Colon Cancer Recognition. ELECO '2012 Elektrik - Elektronik ve Bilgisayar Mühendisliği Sempozyumu 425-428.

Çelik, G. (2022). Prediction of Heart Failure Disease with a Proposed Method Based on Deep Neural Networks. 2nd International Congress of Engineering and Natural Sciences (ICENSS 2022), Ankara/Turkey, 279.

Coşar, M., & Deniz, E. (2021). Makine Öğrenimi Algoritmalar Kullanarak Kalp Hastalklarnn Tespit Edilmesi. *Avrupa Bilim ve Teknoloji Dergisi*, (28), 1112–1116.

De Silva, C., & Kumarawadu, P. (2022). Performance Analysis of Machine Learning Classification Algorithms in the Case of Heart Failure Prediction. 2022 International Wireless Communications and Mobile Computing (IWCMC), 1160–1165.

Fedesoriano. (2021). Heart Failure Prediction Dataset. [Data file and code book]. Retrieved from https://www.kaggle.com/datasets/fedesoriano/heart-failure-prediction.

Heart Failure: Investigation of an Epidemic. (2013). https://doi.org/10.1161/CIRCRESAHA.113.300268

Ho, T. K. (1998, August). C4. 5 decision forests. In Proceedings. Fourteenth International Conference on Pattern Recognition (Cat. No. 98EX170) (Vol. 1, pp. 545-549). IEEE.

Kavzoğlu, T., & Çölkesen, \.Ismail. (2010). Destek vektör makineleri ile uydu görüntülerinin snflandrlmasnda kernel fonksiyonlarnn etkilerinin incelenmesi. *Harita Dergisi*, *144*(7), 73–82.

Ng, A., & Jordan, M. (2001). On discriminative vs. generative classifiers: A comparison of logistic regression and naive bayes. *Advances in Neural Information Processing Systems*, *14*.

Onat, A. (2001). Risk factors and cardiovascular disease in Turkey. Atherosclerosis, 156(1), 1-10.

Öztürk, E., & İçer, S., (2021). Classification of Dermoscopy Images with Feed Forward Neural Network, Decision Trees and Random Forest. *International Journal of Multidisciplinary Studies and Innovative Technologies*, 5(2), 129-135.

Reddy, V. S. K., Meghana, P., Reddy, N. V. S., & Rao, B. A. (2022). Prediction on Cardiovascular disease using Decision tree and Na\"\ive Bayes classifiers. *Journal of Physics: Conference Series*, *2161*(1), 12015.

Rustam, F., Ishaq, A., Munir, K., Almutairi, M., Aslam, N., & Ashraf, I. (2022). Incorporating CNN Features for Optimizing Performance of Ensemble Classifier for Cardiovascular Disease Prediction. *Diagnostics*, *12*(6), 1474.

Srinivas, P., & Katarya, R. (2022). hyOPTXg: OPTUNA hyper-parameter optimization framework for predicting cardiovascular disease using XGBoost. *Biomedical Signal Processing and Control*, *73*, 103456.

World Health Organization. (2022a, July 14). Global health estimates: Leading causes of DALYs.

World Health Organization. (2022b, July 14). Global health estimates: Leading causes of death.

Wu, X., Kumar, V., Ross Quinlan, J., Ghosh, J., Yang, Q., Motoda, H., … others. (2008). Top 10 algorithms in data mining. *Knowledge and Information Systems*, *14*(1), 1–37.