Afyon Kocatepe Üniversitesi Fen ve Mühendislik Bilimleri Dergisi

Afyon Kocatepe University Journal of Science and Engineering

AKÜ FEMÜBİD **22** (2022) 061302 (1307-1313) AKU J. Sci. Eng. **22** (2022) 061302 (1307-1313)

DOI: 10.35414/ akufemubid.1153610

Araştırma Makalesi / Research Article

An Application of Feature Selection Methods to Compare the Performances of Classification Algorithms

Mustafa DEMİR[1], İbrahim KILIÇ[2]

[1] *Afyon Kocatepe Üniversitesi, Fen Edebiyat Fakültesi, İstatistik Bölümü, Afyonkarahisar.*

[2] *Afyon Kocatepe Üniversitesi, Veterinerlik Fakültesi, Biyostatistik Bölümü, Afyonkarahisar.*

Sorumlu yazar e-posta: mustafademir08@gmail.com ORCID ID: http://orcid.org/0000-0001-5684-7778

ibrahimkilic@aku.edu.tr ORCID ID: http://orcid.org/0000-0003-0595-8771

Geliş Tarihi: 03.08.2022 Kabul Tarihi: 27.10.2022

Abstract

In this study, it is aimed to determine fewer and significant variables with the help of feature selection methods among a large number of variables in the data discussed. Feature selection methods are effective methods that have great importance in statistics in recent years and provide great convenience to researchers. Depending on the technique used in the method, different numbers of variables are included in the model, but the correct classification rates may vary. In this context, being able to express the variables in a data set with a large number of variables of interest with a high classification percentage and fewer new variables makes positive contributions to issues such as time and cost. The variables in the data set discussed in this study were firstly analyzed with different feature selection methods and new data sets were created. Afterwards, these new data sets containing different numbers of variables were analyzed with different machine learning techniques and the best machine learning technique was determined. In this study, chronic kidney disease data were handled and the variables in the data set were classified with different feature selection methods. When the results of the study are examined, the highest classification rate with 99.75% was obtained from the correlation-based feature selection method, which includes the random forest and multilayer perceptron technique, and the filter method, which includes the k-nearest neighbor technique, with the same rate. The results of the study show that the percentage of correct classification obtained from this study is higher than that of other studies, when compared with other studies using the same dataset.

Keywords
Clustering methods; Feature selection methods; Machine learning; Chronic kidney disease.

Sınıflandırma Algoritmalarının Performanslarının Karşılaştırılması için Özellik Seçim Yöntemleri Üzerine Bir Uygulama

Öz

Bu çalışmada ele alınan bir verinde yer alan çok sayıdaki değişken arasından özellik seçim yöntemleri yardımı ile daha az sayıda ve anlamlı değişkenlerin belirlenmesi amaçlanmıştır. Özellik seçim yöntemleri son yıllarda istatistik bilimi içerisinde büyük önem arz eden etkili ve araştırmacılara büyük kolaylıklar sağlayan yöntemlerdir. Yöntem içerisinde kullanılan tekniğe bağlı olarak farklı sayıda değişkenlerin modele alınmasına sebep olmakla beraber doğru sınıflandırma oranları değişebilmektedir. Bu bağlamda ilgilenilen çok dayıda değişkene sahip bir veri seti içerisindeki değişkenlerin yüksek bir sınıflama yüzdesi ile daha az sayıda yeni değişkenle ifade edilebilmesi zaman, maliyet gibi konularda olumlu katkılar sunmaktadır. Bu çalışmada ele alınan veri setinde yer alan değişkenler öncelikle farklı özellik seçim yöntemleri ile analiz edilerek yeni veri setleri oluşturulmuştur. Daha sonra oluşturulan bu yeni ve farklı sayıda değişken içeren ver setleri, farklı makine öğrenme teknikleri ile analiz edilerek en iyi makine öğrenme tekniği belirlenmiştir. Bu çalışma kronik böbrek hastalığı verileri ele alınarak farklı özellik seçimleri yöntemleri ile veri setinde yer alan değişkenler sınıflandırılmıştır. Çalışma sonuçları incelendiğinde en yüksek sınıflandırma oranı %99.75 ile rassal orman ve çok katmanlı algılayıcı tekniğini içeren korelasyon tabanlı özellik seçimi yönteminden ve yine aynı oran ile k en yakın komşu tekniğini içeren filtre yönteminden elde edilmiştir. Çalışma sonuçları daha önceden aynı veri seti kullanılarak yapılan diğer araştırmalarla karşılaştırıldığında, bu çalışmadan elde edilen doğru sınıflama yüzdesinin diğer çalışmalardan daha yüksek olduğunu göstermektedir.

Anahtar kelimeler
Sınıflandırma yöntemleri; Özellik seçim yöntemi; Makine öğrenmesi; Kronik böbrek hastalığı.

1. Introduction

In real-life problems, there is often not enough information about interrelated features. Therefore, many candidate features are identified to better represent the problem of interest. However, this also results in the selection of unnecessary features. Unnecessary features are features that are not directly related to the dependent variable, but affect the learning process and do not contribute to the purpose. Since the data used in many classification problems today is large, it is difficult to obtain a good classifier without removing unwanted features. Reducing the number of redundant or irrelevant features both shortens the working time of the learning algorithm and ensures higher generalization success. Thus, a better approach and perspective to the real-life classification problem is developed (Ay 2019).

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 (Gazeloğlu 2020).

FS is also known as subset selection in the literature. In the FS process, the feature subset obtained from the dataset is selected for the learning algorithm. For the solution space, the set consisting of the smallest size dataset with the highest accuracy rate is considered the best subset. The remaining unimportant features in the dataset are ignored. This stage is an important data preprocessing stage. The main goal of the FS is to provide the highest level of data integrity without using all of the original features. It is however to find the minimum subset of features. In many real-world problems, LS is considered a necessity due to the abundance of redundant, misleading or noisy data. In order to find the optimal solution in the LS results, all feature subsets should be tested (Koç 2016).

According to Forman (2003), the main purpose of LS is defined as the process of choosing the best subset that can represent the original dataset without affecting performance. FS (feature selection or variable selection) is defined as the process of selecting the best k number among n features in the data set by evaluating the features suitable for the algorithm to be used (Karakaş 2020).

The change in technology has enabled the integrated use of information systems with high processing capacity in data mining processes, providing the opportunity to handle data mining processes in the field of machine learning (Beyazıt 2019).

Classification is the most well-known job of data mining. It is the process of assigning inputs to classes by a classifier (model) according to various properties. It is the determination of whether the objects at hand are assigned to a class or to which of the classes. In other words, it is the estimation of the appropriate class for objects or situations. Classification inputs are a training set of observations or examples, each of which will be labeled with a class label. The output is the class label assigned by the model based on each observed feature (Emel and Taşkın 2005).

Chronic Kidney Disease (CKD) is defined as chronic and progressive deterioration in the fluid-soluble balance and metabolic-endocrine functions of the kidney as a result of decreased glomerular filtration value. Renal failure is termed as CKD when the glomerular filtration value decreases to 5-10 ml/min and patients need kidney replacement therapies such as dialysis and kidney transplantation (Akpolat and Utaş 2008).

In this study, it is aimed to compare the performance of classification algorithms based on the features obtained by FS methods. By applying FS methods on the relevant data set, the most effective method(s) was determined and the classification algorithms popularly used in the literature were classified with the help of cross validation. During

the classification process, true positive, false positive, kappa statistics, correct classification and AUC (Area Under the Curve) values below the ROC (Receiver Operating Characteristics) curve were calculated and compared over the relevant values.

2. Material and Method

The data used in the study were collected from Apollo Hospitals in India on 2015-07-03. It contains information about 400 people in total. While 250 of these 400 people have CKD, the remaining 150 people do not have CKD. These data were obtained as a result of blood and urine analysis with 24 variable values. There is a classification variant for whether a person is suffering from chronic kidney disease. In other words, there are 25 variables in total, of which 11 are numerical and 14 are nominal. Data were obtained from (Web.Ref.1). There are also some missing values in the data downloaded from the relevant link. Details of the dataset are shown in Table 1. In addition, 10-fold cross validation was used during the entire classification process. According to the program to be used, only the file extensions are set to the existing data in Excel. In this process, no change was made that would harm the structure of the data, and no variables were removed from the relevant data set.

Table 1. Description of the chronic kidney disease dataset

Feature1	Age	Feature14	Potassium
Feature2	Blood pressure	Feature15	Hemoglobin
Feature3	Specific weight	Feature16	Packed cell volume
Feature4	Albumin	Feature17	White blood cell count
Feature5	Sugar	Feature18	Red blood cell count
Feature6	Red blood cells	Feature19	Hypertension
Feature7	Iris cell	Feature20	Diabetes
Feature8	Iris cell clusters	Feature21	Coronary artery disease
Feature9	Bacteria	Feature22	Appetite
Feature10	Blood sugar	Feature23	Foot edema
Feature11	The amount of urea in the blood	Feature24	Anemia
Feature12	Serum creatinine	Feature25	Class
Feature13	Sodium		

In the application part of the study, all of the FS methods in the Weka package program were applied to the relevant data set, and as a result, correlation, filter and consistency methods were able to explain the data set with fewer features (variables). A total of 25 variables used in the study were explained with 16 variables when the correlation-based FS method was applied, 11 variables when the filter FS method was applied, and 4 variables when the consistency FS method was applied. From this point of view, the performances of the classification methods used in the study are calculated for the case where there is no FS (all the variables are used) and for these three FS methods.

Finally, all classification algorithms 16, 11, 4 and all 25 variables were evaluated. Correct classification rates were made by these variable numbers.

When the machine learning algorithms used in this study are compared in general terms, the basic logic in these algorithms is to predict with the maximum number in line with the class. The main reason why such simple algorithms are included in the study is that they are designed to find out what the result of the simplest algorithm is and to determine whether success is achieved in other algorithms. In fact, other algorithms are a bit more comprehensive and will function after a few processes.

In this experimental study, Decision tree, k-NN (k=2), Multilayer Perceptron (MP), Naive Bayes, RTF network and SVM (Poly Kernel, Normalize Poly Kernel, Puk and RTF Kernel classification algorithms) were used for the CKD data set.

ROC analysis is used to determine the ability to discriminate the strength of the test, to compare various test techniques, and to determine the appropriate positive threshold. ROC analysis is a method used to evaluate the results of classification algorithms (Takıcı 2018). Area Under the Curve (AUC) refers to the area under the ROC curve. The closer this field is to 1, the higher the diagnosis rate.

The accuracy of a classification problem is one of the highest universal evaluation measures and is given in equation 1 (TP - True Positive, TN - True Negative, FP - False Positive, FN - False Negative). The benefit of this measure is that it can find the number of suitably classified test cases from the absolute number of test cases (Rahman et al. 2020).

$$Accuracy\ (\%) = \frac{\left| TP + TN \right|}{\left| TP + TN + FP + FN \right|} \quad (1)$$

The Kappa test is a test that measures the reliability of the agreement of two or more observers. (Congalton and Green 1998, Aydın 2018). If the observed values are greater than or equal to the fit due to chance, κ≥0; If the observed fit is smaller than the fit due to chance, κ<0. If κ=1, perfect fit is achieved. The interpretable range of the kappa coefficient is between 0≤ κ ≤ 1 and is not significant for reliability in the case of negative (κ<0). Kappa value above 0.4 is the desired state. Kappa value is calculated as in Equation 2 (Aydın 2018):

$$\kappa = \frac{(P_0 - P_c)}{(1 - P_c)} \quad (2)$$

Here; P_0 is the accepted rate, P_c is the expected rate.

According to Landis and Koch (1977), the interpretation of the obtained κ values is presented in Table 2 (Web.Ref.1).

Table 2. Interpretation of Kappa values

κ	Explanation
≤ 0	No fit (Worse fit than fit, which may be due to chance)
0.1-0.20	Trivial fit
0.21-0.40	Low degree of compliance
0.41-0.60	Moderate compatibility
0.61-0.80	Good compatibility
0.81-1.00	Perfect fit

On the other hand, it should not be forgotten that the results regarding the Kappa value are also affected by the categorical number. The smaller the number of categories, the larger the calculated kappa value. Another point to note is that if the situation to be examined is a very rare situation, the Kappa value related to the indicator of compliance is also small (Viera and Garrett 2005, Kılıç 2015).

CV is a powerful method for evaluating how well a prediction model can perform on an independent dataset. Cross-Validation (CV) allows the predictive potential of baseline training data to be tested internally without predictive bias. The basic process is simple: randomly divide the data into several equal subsets, then iteratively construct and test predictive models such that each subset is retained once and used once for model testing, while the remaining subsets are used to train the model (Web.Ref.3).

In k-fold cross-validation, the original sample is randomly divided into k equally sized sub-samples. Of the k subsamples, a single subsample is kept as validation data to test the model and the remaining k-1 sub-samples are used as training data. The CV process is then repeated k times (multiples), each k subsamples being used exactly once as validation data. The k results from the folds can then be averaged (or otherwise combined) to produce a single estimate. The advantage of this method is that all observations are used for both training and testing, and each observation is used exactly once for validation. For classification problems, stratified k-fold cross validation is typically used, where floors are selected such that, each floor contains roughly the same proportions of grade labels (Web.Ref.4).

3. Findings

Many studies have aimed to predict CKD with the highest accuracy through various algorithms. All these studies were made to determine which algorithm gave the most accurate results using many algorithms and datasets. As a result of the FS methods used in this study, our classification algorithm results and the algorithms used in similar studies and their results are compared in table 3.5. In addition, in order to make comparisons of these studies, of course, it is necessary to analyze them under the same conditions. The datasets and variables used in these 5 studies are the same. However, some of the classification algorithms gave the same results, while others gave different results, as other authors did not give detailed information about the variables of the algorithms they used. The conditions that made this study superior to the other 4 studies were classified with the help of cross validation on the data reduced by FS methods on the related data set. In addition, this study includes some classification algorithms that other authors do not use. In this way, more algorithms were used and

the chance to determine the algorithm with the highest classification rate was obtained.

Table 3. TP, FP, ROC and Kappa statistics for different Feature selections and classification algorithms

	No FS				Filter				Correlation				Consistency			
	TP	FP	ROC	Kappa	TP	FP	ROC	Kappa	TP	FP	ROC	Kappa	TP	FP	ROC	Kappa
Naive Bayes	0,92	0	1	0,89	0,96	0	1	0,94	0,95	0	1	0,93	0,93	0	0,99	0,91
k-NN (k=2)	0,94	0	0,97	0,92	0,99	0	0,99	0,99	0,97	0	0,98	0,96	0,96	0,04	0,97	0,96
Decision Tree	0,99	0,02	0,99	0,97	0,99	0,03	0,99	0,96	0,99	0,02	0,99	0,97	0,98	0,04	0,99	0,95
Random Forrest	1	1	1	1	0,99	0	1	0,98	0,99	0	1	0,99	0,98	0	0,99	0,97
RTF Network	0,97	0	0,99	0,96	0,99	0,07	0,99	0,98	0,97	0	0,98	0,96	0,98	0,01	0,99	0,96
MP	0,99	0	1	0,99	0,99	0	1	0,98	0,99	0	1	0,99	0,96	0,02	0,99	0,92
DVM (polykernel)	0,96	0	0,98	0,95	0,97	0	0,98	0,96	0,97	0	0,98	0,96	0,92	0	0,96	0,89
SVM (Normal polykernel)	0,96	0	0,98	0,94	0,96	0	0,98	0,95	0,97	0	0,98	0,96	0,89	0	0,94	0,86
SVM (Puk)	0,98	0	0,99	0,97	0,97	0	0,98	0,96	0,98	0	0,99	0,97	0,99	0,04	0,97	0,95
SVM (RTFKernel)	0,92	0	0,96	0,89	0,92	0	0,96	0,90	0,92	0	0,96	0,89	0,82	0	0,91	0,77

Table 3 shows the results of True-Positive (DP), False-Positive (YP), Kappa and ROC analysis of the classification algorithm. DP refers to the proportion of people who are actually sick and not sick as a result of the classification algorithm. YP refers to the proportion of people who are found to be actually sick but not sick as a result of the classification algorithm. This means that the XP ratio should be close to 1 and the FC ratio close to 0.

When Table 3 is examined, it is seen that the Roc analysis of 3 methods is calculated as 1. These methods are CKA, Naive Bayes and Random Forest. That is, all 3 methods have the highest diagnosis rate. However, in the FS methods of interest, it is seen that the DP ratio of MP, Decision Tree, SVM (Puk) and Random Forest is much higher. When examined in terms of Kappa values, it is seen that the random forest shows a perfect agreement between the expected and observed values as a result of the classification.

Table 4. Performance (%) results for different feature selections and classification algorithms

	No FS	Filter	Correlation	Consistency
Naive Bayes	95.00	97.5	97.00	95.75
k-NN(k=2)	96.25	99.75	98.50	98.25
Decision Tree	99.00	98.25	99.00	97.75
Random Forrest	100	99.5	99.75	98.75
RTF Network	98.50	99.25	98.50	98.25
MP	99.75	99.5	99.75	96.50
DVM (polykernel)	97.75	98.25	98.25	95.00
SVM (Normal polykernel)	97.50	97.75	98.25	93.50
SVM (Puk)	98.75	98.5	98.75	98.00
SVM (RTFKernel)	95.00	95.50	95.00	89.00

Table 4 indicates the performances of the classification methods included in the study for different FS methods. When Table 4. is examined, the highest correct classification rate (99.75%) is obtained with the 3-state correlation-based feature method, the random forest and MLP classification method, and the filter FS method by k- It was obtained from the NN classification method. When Table 4 is examined, it is a remarkable result that the performances obtained from the classification methods made by applying the Consistency FS method are lower than the performances of the classification methods made by applying other FS methods. In addition, the lowest performance

consistency FS with an accurate classification rate of 89% was applied and obtained from the DVM (RTFKernel) classification method.

4. Discussion and Conclusion

When past studies are examined in the diagnosis of CKD, medical methods have generally been successful. In addition to these methods, depending on the technological developments in recent years, developing a new system that will help doctors with the help of various computer-assisted algorithms will both help in the rapid diagnosis of the disorder and reduce the workload of doctors, thus enabling doctors to work more efficiently.

In addition, after the diagnosis of the disease by a doctor, the confirmation of the diagnosis with computerized systems will also eliminate man-made errors. In this study, 10 different classification algorithms were used. In addition, 4 different core functions of DVM are used. In Table 5, previous studies using the same data set of 400 individuals and performance statistics regarding the classification methods they used are given.

Table 5. Accuracy rates obtained from previous studies for the chronic kidney disease dataset

Study	Program	Classification Algorithm	Truth
Celik *et al*.	WEKA	Decision Tree	%91.66
		DVM	%96.11
	WEKA and	k-NN	%98.10
Charleonnan *et al*.	MATLAB	DVM	%98.30
		Logistic Regression	%96.55
		Decision Tree	%94.80
Chetty *et al*.	WEKA	Naive Bayes	%95.00
		DVM	%97.75
		k-NN	%95.75
		Decision Tree	%91.00
Gunarathne *et al*.	---	Decision Forests	%99.10
		Logistic Regression	%95.00
		ANN	%97.50

When Table 5 is examined, Çelik et al. (2016) obtained the correct classification rates of 91.66% in Decision Tree and 96.11% in SVM, among classification algorithms, through the Weka program.

In their study, Charleonnan et al. (2016) obtained the correct classification rates of 98.10% with k-NN, 98.30% with SVM, 96.55% with Logistic Regression and 94.80% with Decision Tree from classification algorithms through Weka and MATLAB software.

Chetty et al. (2015), using the Weka program, they obtained 95.00% correct classification rates with naive bayes, 97.75% with SVM, 95.75% with k-Nearest Neighborhood and 91.00% with Decision Tree.

Gunarathne et al. (2017), among the classification algorithms, they obtained 99.10 % correct classification rates with Decision Forests, 95.00 % with Logistic Regression and 97.50 % with ANN.

When Table 5 is examined, the highest performance among the previous studies was determined by Gunarathne et al. (2017) and the Decision Tree classification algorithm was obtained with 99.10%. In all previous studies, all 25 variables in the CKD data set were included in the process, and these results were obtained without cross-validation, while all the results obtained from this study were obtained as a result of CV.

In this context, it is one of the original results of the study that this study achieved better results than other studies. While compared with the other studies, another important result of this study is, it reduced the number of variables as a result of different FSs, and more efficient results were obtained with fewer variables than many previous studies.

5. References

Akpolat, T. and Utaş, C., 2008. Hemodiyaliz Hekimi El Kitabı, Türk Nefroloji Derneği Yayınları, Ceylan Ofset Samsun, 56-72.

Ay, Ö., 2019. Özellik Seçimi Problemleri İçin Polihedral Konik Fonksiyonlar Temelli Çözüm Yaklaşımı. Yüksek Lisans Tezi, Eskişehir Teknik Üniversitesi Endüstri Mühendisliği Anabilim Dalı, Eskişehir, 43.

Aydın, C., 2018. Makine Öğrenmesi Algoritmaları Kullanılarak İtfaiye İstasyonu İhtiyacının

Sınıflandırılması. *Avrupa Bilim ve Teknoloji Dergisi*, **14**, 169–175.

Beyazıt, B.E., 2019. Büyük Veri Problemlerinde Performans Arttırmaya Yönelik Özellik Seçimi ve Boyut İndirgeme Optimizasyonu. Yüksek Lisans Tezi, Gazi Üniversitesi Bilişim Enstitüsü, Ankara, 63.

Charleonnan, P., Fufaung, T., Niyomwong, T., Chokchueypattanakit, W., Suwannawach, S., *et al*., 2016. Predictive analytics for chronic kidney disease using machine learning techniques, The 2016 Management and Innovation Technology International Conference, Bangkok, Thailand, 81–83.

Chetty, N., Vaisla, K.S. and Sudarsan, S.D., 2015. Role of attributes selection in classification of chronic kidney disease patients, International Conference on Computing, Communication and Security (ICCCS), Le Meridien, Mauritius, 1–6.

Congalton, R.G. and Green, K., 1998, Assessing the Accuracy of Remotely Sensed Data: Principles and Practices, first edn. Lewis Publications, Boca Raton, 137.

Çelik, E., Atalay, M. and Kondiloğlu, A., 2016. The diagnosis and estimate chronic kidney disease using the machine learning methods. *International Journal of Intelligent Systems and Applications in Engineering*. **4**, 27–31.

Emel, G.G. and Taşkın, Ç., 2005. Veri Madenciliğinde Karar Ağaçları ve Bir Satış Analizi Uygulaması. *Eskişehir Osmangazi Üniversitesi, Sosyal Bilimler Dergisi*, **6**, 221–229.

Forman, G., 2003. An Extensive Empirical Study of Feature Selection Metrics for Text Classification. *Journal of Machine Learning Research*, **3**, 1289–1305.

Gazeloğlu, C., 2020. Prediction of heart disease by classifying with feature selection and machine learning methods. Progress in Nutrition **22**, 660-670.

Gunarathne, W.H.S.D., Perera, K.D.M. and Kahandawaarachchi, K.A.D.C.P., 2017. Performance evaluation on machine learning classification techniques for disease classification and forecasting through data analytics for chronic kidney disease (CKD). IEEE 17th International Conference on Bioinformatics and Bioengineering, 291–296p.

Karakaş, M., 2020. Sınıflandırma Problemlerinde Özellik Seçimi için Karşıtlık Tabanlı Gri Kurt Optimizasyon Algoritması. Yüksek Lisans Tezi, Bilecik Şeyh Edebali Üniversitesi Fen Bilimleri Enstitüsü, Bilecik, 66.

Kılıç, S., 2015. Kappa Test. *Journal of Mood Disorders*, **5**, 142–144.

Koç, İ., 2016. Sınıflandırma Problemlerinde Meta-Sezgisel Optimizasyon Yöntemlerinin Özellik Seçimi ve Ayrıklaştırma Amacıyla Kullanımı. Yüksek Lisans Tezi, Selçuk Üniversitesi Fen Bilimleri Enstitüsü, Konya, 181.

Landis, J.R. and Koch, G.G., 1977. The measurement of observer agreement for categorical data. *Biometrics*, **33**, 159–174.

Rahman, M.M., Usman, O.L., Muniyandi, R.C., Sahran, S., Mohomed, S. and Razak, R.A., 2020. Otizm Spektrum Bozukluğu için Özellik Seçim ve Sınıflandırmasına Yönelik Makine Öğrenim Yöntemlerinin Gözden Geçirilmesi. *Brain Sciences Journal*, **10**, 1–26.

Takıcı, H., 2018. Improvement of heart attack prediction by the feature selection methods. *Turkish Journal of Electrical Engineering Computer Sciences*, **26**, 1–10.

Viera, A.J. and Garrett, J.M., 2005. Understanding interobserver agreement: the kappa statistic. *Family Medicine*, **3**, 360–370.

Web References

1- https://tr.wikipedia.org/wiki/Fleiss%27in_kappa_katsayısı/, (22.06.2021).

2- https://www.kisa.link/KU44/, (12.02.2021)

3- https://blog.goldenhelix.com/cross-validation-for-genomic-prediction-in-svs/, (11.03.2021).

4- https://www.openml.org/a/estimation-procedures/7, (11.03.2021).