PERFORMANCE EVALUATION OF THE ENSEMBLE LEARNING MODELS IN THE CLASSIFICATION OF CHRONIC KIDNEY FAILURE

Z. Kucukakcali, and I. Balikci Cicek

Abstract— *Aim:* This study aims to classify the CKF by applying the ensemble learning method, which is an important sub-field of machine learning, on the open access CKF data set.

Materials and Methods: In this study, the ensemble learning methods Bagging, Boosting and Stacking methods were applied to the open access data set named "Chronic Kidney Disease". The performance of the models used was evaluated with accuracy, sensitivity, specificity, positive predictive value, and negative predictive value.

Results: Accuracy, , sensitivity, specificity, positive predictive value and negative predictive value obtained from the Bagging model were 96.5, 96.8, 96, 97.5 and 94.7 respectively. Accuracy, , sensitivity, specificity, positive predictive value and negative predictive value obtained from the Boosting model were 98.75, 98, 1, 1 and 96.7 respectively. Accuracy, , sensitivity, specificity, positive predictive value and negative predictive value obtained from the Stacking model were 99.25, 99.6, 98.9, 99.2 and 99.3 respectively.

Conclusion: The findings obtained from this study showed that successful results were obtained with the ensemble learning model for the kidney failure data set.

Keywords— Chronic kidney failure, classification, machine learning, ensemble learning.

1. INTRODUCTION

CHRONIC kidney failure (CKF), which has become an important public health problem in the world and in our country today, is a disease that can occur due to many reasons, results in irreversible loss of kidney functions, negatively affects the quality of life of individuals and requires lifelong treatment and follow-up [1]. Clinically, CKF is defined as a structural and functional disorder that can be demonstrated by blood, urine, and imaging methods resulting in a decrease in nephron count and nephron functions as a result of a decrease in glomerular filtration rate (GFR) that lasts for more than three months. [2]. Nowadays, it is reported that the incidence of CKF is increasing rapidly. Considering the results of the studies, it is seen that the rate of chronic kidney failure varies between 10-16% in the world.

✉ **Zeynep KUCUKAKCALI,** Inonu University Department of Biostatistics and Medical Informatics, Faculty of Medicine, Malatya, Turkey, (zeynep.tunc@inonu.edu.tr)

İpek BALIKCI CICEK, Inonu University Department of Biostatistics and Medical Informatics, Faculty of Medicine, Malatya, Turkey, (ipek.balikci@inonu.edu.tr)

Manuscript received Sep 19, 2020; accepted Oct 21, 2020.
Digital Object Identifier:

It is estimated that more than 500 million people worldwide have CKF. [3]. Chronic kidney failure (CKF) is an increasingly common health problem worldwide. This disease, which is very expensive to treat, can lead to negative consequences when evaluated from a prognostic point of view. The most important consequences are the progression of kidney disease, acute and chronic complications due to renal dysfunction, cardiovascular mortality, and morbidity. Therefore, CKF, which can occur due to many reasons and results in irreversible loss of kidney functions, is a disease that negatively affects the quality of life of individuals and requires lifelong treatment and follow-up [4].

Data mining is the finding of relationships and rules that will enable us to make predictions about the future from a large amount of data with the help of computer programs. [5]. Data mining; includes a combination of techniques from different disciplines such as database technology, statistics, machine learning, pattern recognition, neural networks, data visualization, and spatial data analysis [6]. Machine learning, one of these techniques, is a subfield of artificial intelligence that aims to make predictions about new data when they are exposed to new data by performing data-based learning. Machine learning systems aim to completely eliminate the need for human intuition or to gain the ability to make decisions through human-machine cooperation [7]. The logic of ensemble learning, which is an important subfield of machine learning, is based on the idea that many classifiers can be combined to increase the rate of correct prediction using a single basic classifier. In other words, the method of ensemble learning is based on the idea of combining many basic classifiers to obtain a more accurate and reliable model (meta classifier) compared to the classification success that a basic classifier (model) can achieve [8]. In this way, ensemble learning methods increase the predictive power of weak classifiers [9]. For these reasons, ensemble learning methods are highly preferred recently.

The purpose of this study is to classify the CKF by applying the ensemble learning method, which is an important sub-field of machine learning, on the open-access CKF data set.

2. MATERIAL AND METHODS

2.1. Dataset

In the study, the ensemble learning method, which is an important sub-field of machine learning, was applied to an open-access data set called "Chronic Kidney Disease". Open access data set named "Chronic Kidney Disease" was obtained from https://www.kaggle.com/abhia1999/chronic-kidney-disease. There are 400 patients in the data set used. 250 (62.5%) of these patients have chronic kidney failure. Explanations about the variables and their properties in the data set are given in Table 1.

TABLE I
EXPLANATIONS ABOUT THE VARIABLES IN THE DATASET AND THEIR PROPERTIES

Variable	Variable Description	Variable Type	Variable Role
Bp	Blood Pressure	Quantitative	Predictor
Sg	Specific Gravity	Quantitative	Predictor
Al	Albumin	Qualitative	Predictor
Su	Sugar	Qualitative	Predictor
Rbc	Red Blood Cell	Qualitative	Predictor
Bu	Blood Urea	Quantitative	Predictor
Sc	Serum Creatinine	Quantitative	Predictor
Sod	Sodium	Quantitative	Predictor
Pot	Pottasium	Quantitative	Predictor
Hemo	Hemoglobin	Quantitative	Predictor
Wbcc	White Blood Cell Count	Quantitative	Predictor
Rbcc	Red Blood Cell Count	Quantitative	Predictor
Htn	Hypertension	Qualitative	Predictor
Class	Predicted Class	Qualitative	Output

3. ENSEMBLE LEARNING METHOD

Machine learning methods classify and infer by learning the pattern in the data stack. Machine learning has developed rapidly in recent years. The rapid development of machine learning has been dependent on the development of statistical algorithms that can extract information from these data with the rapid increase in data stacks in the computer and internet environment. For this reason, many machine learning methods have been developed, some of which are k-nearest neighbor algorithm, simple (naive) Bayes classifier, decision trees, logistic regression analysis, k-means algorithm, support vector machines, and artificial neural networks. Some of these approaches have the ability to predict, some to cluster and some to classify [10]. An important subfield of machine learning is ensemble learning methods. Ensemble learning methods provide a common classification result from the estimates of each classifier by classifying the data of more than one machine learning algorithm separately, rather than classifying the data set of a machine learning algorithm. Thus, according to the prediction results of a machine learning method, the common prediction results obtained from more than one machine learning method provide more accurate, more reliable, and higher performance [7]. Ensemble learning methods are based on the principle that more than one classifier can perform classification with higher accuracy than a single classifier predicts. Ensemble learning methods have found a wide range of applications in recent years with their successful results. Commonly used ensemble learning methods have been successfully applied in many diagnostic and diagnostic studies [11, 12].

An important issue affecting the classification performance in the ensemble learning method is the selection of the appropriate joining method. In determining the joining technique, attention should be paid to the selection of the appropriate joining technique for classifiers. There are different ensemble learning methods according to the joining techniques, the sample selection for the training data set, and the process steps. These methods are the bagging ensemble learning method, the boosting ensemble learning method, and the stacking ensemble learning method [13, 14].

3.1. Bagging

Bagging method is a method that aims to retrain the basic learner by creating new training sets by random selection by substituting from a known training set [15]. In summary, the main purpose of the Bagging method is to obtain new data sets randomly using training data and to increase the success of classification by creating differences. In the bagging method, first, the data set is divided into training and test data. One or more new training sets consisting of n samples are obtained by random selection method by replacing the training set containing N samples. Each basic classifier in the community obtained by the bagging method is trained with training sets containing different examples obtained in this way. Finally, the result of each major classifier is combined with the majority vote [16].

3.2.Boosting

Boosting yöntemi, Schapire tarafından 1990 yılında tanıtılan ve 2000'li yıllara kadar geliştirilen bir toplu öğrenme yöntemi [17-19]. The term "boost" refers to a family of algorithms that transform poor learning methods into powerful learning techniques. Boosting is an ensemble method to improve the model predictions of any learning algorithm, and unlike the Bagging method, the predictors in the Boosting method are created sequentially, although they are not independent of each other. The aim of this method is to combine weak estimators to obtain strong estimator (s). Models are created by assigning weight to observations. In the Boosting method, as in the bagging method, N training sets are created. In this method, models with low variance and bias are obtained by both the presence of the bagging method and the assignment of weight to the observations [19].

3.3.Stacking

The stacking method is a simple ensemble learning technique that creates a meta classifier by combining two or more basic multiple classification models. It is an ensemble model that is trained by combining the estimates of the classification models used. Predictions made from models created by the basic classifier are used as input for each ordered layer and are combined to create a new set of predictions. In the stacking method, basic classification models are trained on the original training data set and then created based on the outputs (estimates) of the basic classification models in the meta-classifier community. The meta-classifier performs the classification process by training on the predicted class labels [20].

3.4. Performance evaluation criteria

The classification matrix for the calculation of performance metrics is given in Table II.

TABLE II
THE METRICS OF THE MODEL'S CLASSIFICATION PERFORMANCE

		Real		
		Positive	**Negative**	**Total**
Predicted	Positive	True positive (TP)	False negative (FN)	TP+FN
	Negative	False positive (FP)	True negative (TN)	FP+TN
	Total	TP+FP	FN+TN	TP+TN+FP+FN

Accuracy = (TP+TN)/(TP+TN+FP+FN)

Sensitivity = TP/(TP+FP)

Specificity = TN/(TN+FN)

Positive predictive value = TP/(TP+FN)

Negative predictive value =TN/(TN+FP)

4. DATA ANALYSIS

Quantitative data are summarized by median (minimum-maximum) and qualitative variables are given by number and percentage. Normal distribution was evaluated with the Kolmogorov-Smirnov test. In terms of input variables, the existence of a statistically significant difference and relationship between the categories of output variable, "ckd" and "notckd" groups, was examined using Mann-Whitney U, Pearson Chi-square test and Yates's correction chi-square test. p<0.05 values were considered statistically significant. In all analyzes, IBM SPSS Statistics 26.0 for the Windows package program was used. Random Forest, Neural Network, Support Vector Machine were used as classifiers in the stacking method of the ensemble learning models. Rapidminer Studio Free version 9.4 is used for ensemble learning methods analysis.

5. RESULTS

Descriptive statistics related to the target variable examined in this study are presented in Table 3 and Table 4. There is a statistically significant difference between the dependent variable classes in terms of other variables other than the “Pot” variable.

TABLE III
DESCRIPTIVE STATISTICS FOR QUANTITATIVE INPUT VARIABLES

	Predicted Class		
Variables	Not-ckd **Median (min-max)**	ckd **Median (min-max)**	**P* value**
Bp	70(60-80)	80(50-180)	<0,001*****
Sg	1,02(1,02-1,03)	1,02(1,01-1,03)	<0,001*****
Bu	33,5(10-57)	55(1,5-391)	<0,001*****
Sc	0,9(0,4-3,07)	2,45(0,5-76)	<0,001*****
Sod	141(135-150)	137,53(4,5-163)	<0,001*****
Pot	4,5(3,3-5)	4,63(2,5-47)	0,515
Hemo	15(12,53-17,8)	11,3(3,1-16,1)	<0,001*****
Wbcc	7750(4300-11000)	8406(2200-26400)	<0,001*****
Rbcc	5,25(4,4-6,5)	4,71(2,1-8)	<0,001*****

***: Mann Whitney U test**

TABLE IV
DESCRIPTIVE STATISTICS FOR QUALITATIVE INPUT VARIABLES

Variables		Predicted Class		**P** value**
		Not-ckd	ckd	
Al	0	145(96,7%)	54(21,6%)	<0,001*****
	1	5(3,3%)	85(34,0%)	
	2	0(0%)	43(17,2%)	
	3	0(0%)	43(17,2%)	
	4	0(0%)	24(9,6%)	
	5	0(0%)	1(0,4%)	
Su	0	150(100%)	189(75,6%)	<0,001*****
	1	0(0%)	13(5,2%)	
	2	0(0%)	18(7,2%)	
	3	0(0%)	14(5,6%)	
	4	0(0%)	13(5,2%)	
	5	0(0%)	3(1,2%)	
rbc	0	0(0%)	47(18,8%)	<0,001******
	1	150(100%)	203(81,2%)	
htn	0	150(100%)	103(41,2%)	<0,001******
	1	0(0%)	147(58,8%)	

***: Pearson's chi-square test;** Yates's correction chi-square test**

In this study, the classification matrices of Bagging, Boosting, and Stacking models, which are among the ensemble learning methods used to classify the CKF dataset, are given in Table V.

TABLE V
CLASSIFICATION MATRICES OF BAGGING, BOOSTING AND STACKING MODELS

Classification Matrix of the Bagging Model			
Prediction	**Reference**		
	CKD	not CKD	**Total**
CKD	242	6	248
not CKD	8	144	152
Total	250	150	400

Classification Matrix of the Boosting Model			
Prediction	**Reference**		
	CKD	not CKD	**Total**
CKD	245	0	245
not CKD	5	150	155
Total	250	150	400

Classification Matrix of the Stacking Model			
Prediction	**Reference**		
	CKD	not CKD	**Total**
CKD	249	2	251
not CKD	1	148	149
Total	250	150	400

The values for the metrics of the classification performance of Bagging, Boosting and Stacking models are given in Table VI.

TABLE VI
VALUES FOR THE METRICS OF THE CLASSIFICATION PERFORMANCE OF BAGGING, BOOSTING AND STACKING MODELS

Models	**Metric**	**Value**
Bagging	Accuracy	96.5
	Sensitivity	96.8
	Specitivity	96
	Positive predictive value	97.5
	Negative predictive value	94.7
Boosting	Accuracy	98.75
	Sensitivity	98
	Specitivity	1
	Positive predictive value	1
	Negative predictive value	96.7
Stacking	Accuracy	99.25
	Sensitivity	99.6
	Specitivity	98.9
	Positive predictive value	99.2
	Negative predictive value	99.3

Accuracy, sensitivity, specificity, positive predictive value and negative predictive value obtained from the Bagging model were 96.5, 96.8, 96, 97.5 and 94.7 respectively. Accuracy, sensitivity, specificity, positive predictive value and negative predictive value obtained from the Boosting model were 98.75, 98, 1, 1 and 96.7 respectively. Accuracy, sensitivity, specificity, positive predictive value and negative predictive value obtained from the Stacking model were 99.25, 99.6, 98.9, 99.2 and 99.3 respectively.

6. DISCUSSION

Chronic kidney failure (CKF) is an important public health problem with increasing frequency in the world and in our country. KRG; it is an important health problem that is commonly seen with objective kidney damage lasting at least three months and/or GFR below 60ml / min, chronic and progressive impairment in the fluid-electrolyte balance, endocrine and metabolic functions of the kidney, increased mortality and decreased quality of life. In population-based studies investigating the prevalence of CKF, similar results have been obtained in the world and in our country. The prevalence of CKF in the world was found to be 11.7-15.1% (average 13.4%) according to the results of the meta-analysis study conducted in 2016. CKF is considered to be a serious public health problem in the world due to its high morbidity rate and increased health expenditures. Therefore, it is an open area for research and new developments. [21-23].

Machine learning methods perform classification and estimation by learning the pattern in the data stack. Machine learning has developed rapidly in recent years. Machine learning methods are one of the technologies used in the diagnosis of diseases and clinical decision support systems in recent years. [24]. Ensemble learning methods, one of the machine learning methods, provide a common classification result from the predictions of each classifier by classifying the data of more than one machine learning algorithm rather than classifying the data set of a machine learning algorithm. Thus, according to the prediction results of a machine learning method, common prediction results obtained from more than one machine learning method provide more accurate, more reliable, and higher performance. Ensemble learning methods have found a wide range of applications in recent years with their successful results. [11, 25].

In this study, ensemble learning models, one of the machine learning methods, were applied to the data set named “Chronic Kidney Disease” which is an open-source data set. According to the results of 3 different models used, the method for the best classification performance is the Stacking method. Accuracy, sensitivity, specificity, positive predictive value and negative predictive value obtained from this model were 99.25, 99.6, 98.9, 99.2 and 99.3 respectively.

In a study conducted with the same data set, the effects of variable selection on the support vector machine method were investigated. According to the results of this study, the highest accuracy was obtained as 98.5 from the support vector machine

 ISSN: 2548-0650

results created with different variable selection methods [26]. In this study, the disease was classified by obtaining an accuracy of 99.25 with the Stacking method.

As a result, the ensemble learning methods used have produced very successful results in the study conducted with the chronic kidney failure data set. Ensemble learning methods offer very high classification performance when correct classifiers and joining techniques are used.

REFERENCES

[1] N. Erol, "Diyaliz tedavisine başlanmayan kronik böbrek yetmezliği hastaları ile hemodiyaliz tedavisi olan hastaların yaşam kalitelerinin karşılaştırılması," Sağlık Bilimleri Enstitüsü, 2010.

[2] E. Crowe, D. Halpin, and P. Stevens, "Early identification and management of chronic kidney disease: summary of NICE guidance," Bmj, vol. 337, pp. a1530, 2008.

[3] M. Eskicioglu, Ü. Eda, and A. Özdemir, "Böbrek hastalarının klinikte yattığı sürede öğrenim gereksinimlerinin tespiti," Uludağ Üniversitesi Tıp Fakültesi Dergisi, vol. 45, pp. 205-210.

[4] Ü. Yıldırım, "Hİpervolemİk kronİk böbrek yetmezlİğİ hastalarinda troponin düzeyleri ve medİkal dİürez tedavisinin troponin düzeyi üzerine etkileri," Uzmanlık, Van yyü üniversitesi, 2020.

[5] H. Akpınar, "Veri tabanlarında bilgi keşfi ve veri madenciliği," İÜ İşletme Fakültesi Dergisi, vol. 29, pp. 1-22, 2000.

[6] D. J. Hand, "Principles of data mining," Drug safety, vol. 30, pp. 621-622, 2007.

[7] R. Polikar, "Ensemble learning," in Ensemble machine learning, ed: Springer, 2012, pp. 1-34.

[8] L. Rokach and O. Maimon, "Clustering methods," in Data mining and knowledge discovery handbook, ed: Springer, 2005, pp. 321-352.

[9] T. G. Dietterich, "Ensemble methods in machine learning," in International workshop on multiple classifier systems, 2000, pp. 1-15.

[10] M. Atalay and E. Çelik, "Büyük Veri Analizinde Yapay Zekâ Ve Makine Öğrenmesi Uygulamalari-Artificial Intelligence and Machine Learning Applications in Big Data Analysis," Mehmet Akif Ersoy Üniversitesi Sosyal Bilimler Enstitüsü Dergisi, vol. 9, pp. 155-172, 2017.

[11] S. Sun, "Ensemble learning methods for classifying EEG signals," in International Workshop on Multiple Classifier Systems, 2007, pp. 113-120.

[12] S.-L. Hsieh, S.-H. Hsieh, P.-H. Cheng, C.-H. Chen, K.-P. Hsu, I.-S. Lee, et al., "Design ensemble machine learning model for breast cancer diagnosis," Journal of medical systems, vol. 36, pp. 2841-2847, 2012.

[13] C. Zhang and Y. Ma, Ensemble machine learning: methods and applications: Springer, 2012.

[14] L. Rokach, Pattern classification using ensemble methods vol. 75: World Scientific, 2010.

[15] L. Breiman, "Bagging predictors," Machine learning, vol. 24, pp. 123-140, 1996.

[16] A. J. Ferreira and M. A. Figueiredo, "Boosting algorithms: A review of methods, theory, and applications," in Ensemble machine learning, ed: Springer, 2012, pp. 35-85.

[17] J. H. Friedman, "Greedy function approximation: a gradient boosting machine," Annals of statistics, pp. 1189-1232, 2001.

[18] J. Friedman, T. Hastie, and R. Tibshirani, "Additive logistic regression: a statistical view of boosting (with discussion and a rejoinder by the authors)," The annals of statistics, vol. 28, pp. 337-407, 2000.

[19] Y. Freund and R. E. Schapire, "Experiments with a new boosting algorithm," in icml, 1996, pp. 148-156.

[20] F. Divina, A. Gilson, F. Goméz-Vela, M. García Torres, and J. F. Torres, "Stacking ensemble learning for short-term electricity consumption forecasting," Energies, vol. 11, pp. 949, 2018.

[21] N. R. Hill, S. T. Fatoba, J. L. Oke, J. A. Hirst, C. A. O'Callaghan, D. S. Lasserson, et al., "Global prevalence of chronic kidney disease–a systematic review and meta-analysis," PloS one, vol. 11, pp. e0158765, 2016.

[22] S. F. Yalin, E. Parmaksiz, M. Mese, Z. Dogu, N. D. Çeçen, and Z. B. Bahçebasi, "Evaluating the causes for rejection of potential live-renal donors: Single center experience/potansiyel canli bobrek vericilerinin reddedilme nedenlerinin degerlendirilmesi: Tek merkez deneyimi," Journal of Istanbul Faculty of Medicine, vol. 82, pp. 127-131, 2019.

[23] A. C. Webster, E. V. Nagler, R. L. Morton, and P. Masson, "Chronic kidney disease," The lancet, vol. 389, pp. 1238-1252, 2017.

[24] A. Doganer, "Topluluk öğrenme yöntemleri ile renal hücreli karsnom'un tahmin edilmesi," Doktora, İnönü Üniversitesi, 2020.

[25] Y. Qi, "Random forest for bioinformatics," in Ensemble machine learning, ed: Springer, 2012, pp. 307-323.

[26] H. Polat, H. D. Mehr, and A. Cetin, "Diagnosis of chronic kidney disease based on support vector machine by feature selection methods," Journal of medical systems, vol. 41, pp. 55, 2017.

BIOGRAPHIES

Zeynep KÜÇÜKAKÇALI obtained her BSc. degree in mathematics from Çukurova University in 2010. She received MSc. degree in biostatistics and medical informatics from the Inonu University in 2018. She currently continues Ph.D. degrees in biostatistics and medical informatics from the Inonu University. In 2014, she joined the Department of Biostatistics and Medical Informatics at Inonu University as a researcher assistant. Her research interests are cognitive systems, data mining, machine learning, deep learning.

İpek BALIKÇI ÇİÇEK obtained her BSc. degree in mathematics from Çukurova University in 2010. She received MSc. degree in biostatistics and medical informatics from the Inonu University in 2018. She currently continues Ph.D. degrees in biostatistics and medical informatics from the Inonu University. In 2014, she joined the Department of Biostatistics and Medical Informatics at Inonu University as a researcher assistant. Her research interests are cognitive systems, data mining, machine learning, deep learning.