

Chronic Kidney Disease Prediction with Reduced Individual Classifiers

Merve Doğruyol Başar¹ , Aydın Akan² 

¹Department of Electrical and Electronics Engineering, Istanbul University Cerrahpaşa School of Engineering, İstanbul, Turkey

²Department of Biomedical Engineering Department, İzmir Katip Celebi University School of Engineering, İzmir, Turkey

Cite this article as: M. Doğruyol Başar, A. Akan, "Chronic Kidney Disease Prediction with Reduced Individual Classifiers", *Electrica*, vol. 18, no: 2, pp. 249-255, 2018.

ABSTRACT

Chronic kidney disease is a rising health problem and involves conditions that decrease the efficiency of renal functions and that damage the kidneys. Chronic kidney disease may be detected with several classification techniques, and these have been classified using various features and classifier combinations. In this study, we applied seven different classifiers (Naïve Bayes, HoeffdingTree, RandomTree, REPTree, Random Subspaces, Adaboost, and IBk) for the diagnosis of chronic kidney disease. The classification performances are evaluated with five different performance metrics, i.e., accuracy, kappa, mean absolute error (MAE), root mean square error (RMSE), and F measures. Considering the classification performance analyses of these methods, six reduced features provide a better and more rapid classification performance. Seven individual classifiers are applied to the six features and the best results are obtained using individual random tree and IBk classifiers.

Keywords: Chronic kidney disease, reduced individual classifiers, data mining, classification

Introduction

Chronic Kidney Disease (CKD), which is a universal public health problem, describes the degeneration of renal function. Kidneys filter fluids from the blood to produce urine [1]. When the glomerular filtration rate (GFR) reaches a lower stage, high residual urine and wastes can accumulate in the body. Primarily, when the deprivation of nephrons and detracting of operational renal mass reaches a certain point, the surviving nephrons initiate a period of irrecoverable sclerosis that causes an onward regression in GFR [2-4].

In Turkey, there is a rising rate of kidney failure, with serious implications including a high amount of fund waste. At the same time, CKD is mainly associated with the elderly population. In our country, despite the lower number of elderly people, the CKD ratio is considerably high. However, for some younger patients and patients over 65 years of age, CKD typically causes the progressive loss of kidney function. Cardiovascular disease and chronic renal failure risk factors have increased and are also associated with CKD. Kidney disease is one of the top causes of death in Turkey [5, 6].

Chronic Kidney Disease can easily be cured in the early stages. However, degenerative changes in kidney tissue progresses very rapidly and incurable inflammation in the kidneys accumulates in a short period of time. It develops into many complications such as malnutrition deficiency, loss of body mass, muscle weakness, edema, hypertension, fatigue, attenuated exercise capacity, disrupted cognitive and immune function, lower quality of life, advancement of cardiovascular disease, heart failure, enhanced cardiovascular mortality, pericarditis, lupus, peripheral neuropathy, gastrointestinal symptoms, loss of appetite, skin manifestations and malnutrition [6, 7].

Patients with CKD need to get comprehensive treatment. Furthermore, some therapies can overcome the symptoms, decrease and slow the development of the disease. CKD may be the

Corresponding Author:

Merve Doğruyol Başar

E-mail:

mervedb@istanbul.edu.tr

Received: 27.02.2018

Accepted: 12.04.2018

© Copyright 2018 by Electrica

Available online at

<http://electrica.istanbul.edu.tr>

DOI: 10.26650/electrica.2018.99255

cause of many related problems in the community, such as economic, social and medical problems. Ultrasound, CT scan, kidney biopsy, and a greater rate of drug consumption can cause economical and medical problems. In addition, the constant use of drugs and the ensuing complications can cause social problems as well [7].

In the last few decades, numerous studies have been conducted in order to understand and examine CKD. In the literature, many scholars have used different classification techniques for the perception of CKD. Data mining is the process of pulling out intended data from the major dataset. These techniques are used in various applications such as medical diagnosis, face recognition and data filtering.

Data mining techniques, such as clustering, classification and so on, play a big role in the extraction of unknown knowledge from the major databases. Classification is a supervised learning technique and predefines subgroups. Classification algorithm necessitates the classes to be identified based on the data attribute value. It defines the classes taking into account the characteristics of the data. The training algorithm uses these predefined specimens to determine the set of parameters required for appropriate segregation [8].

Jena and Kamila [9] predicted and analyzed kidney disease using Waikato Environment for Knowledge Analysis (Weka) tool and they used different algorithms such as Support Vector Machine (SVM), J48 classifier, Naïve Bayes classifier, Multilayer perceptron, conjunctive rule and decision table [10].

Chaudhary and Garg [11] developed a prediction system using A-priori and k-means algorithms for prediction and these algorithms were also used to predict patients with kidney failure. The authors analyzed 42 attributes of the data using machine learning tools and evaluated the data using Receiver Operating Characteristic (ROC) plots.

Baby and Vital [12] proposed a diagnosis and prediction system and analyzed the data using the Weka tool. The authors used AD Trees, J48 classifier, K-Star algorithm, Naïve Bayes classifier and Random Forest and evaluated ROC plots. Their study showed that K-Star algorithm and Random Forest classifier were the best methods for their dataset.

Sinha [13] proposed the performance evaluation of two data mining techniques. They used K-Nearest Neighbor (KNN) and SVM to compare accuracy and precision. The authors showed that the KNN classifier provided better results than SVM in terms of accuracy and precision values [13].

Vijayarani and Dhayanand [14] used SVM and Artificial Neural Network (ANN) for the diagnosis of kidney disease. The authors compared the performance of two classifiers in terms of accuracy and execution time and used Weka tool for execution. After the performance a comparison of the two data mining techniques

was undertaken and the authors concluded that ANN is better than the other algorithm for the diagnosis of kidney stone.

Against this background, in this work, we use seven different classification techniques to evaluate the existence of CKD in humans. With this purpose, we focus on Naïve Bayes, Hoëffding Tree, Random Tree, REP Tree, Random Subspaces, Adaboost and IBk classifiers and compare their accuracy, kappa, root mean square error (RMSE), mean absolute error (MAE) and F measure values.

Our analyses show that classifiers can be applied to all features and reduced features. After that, the best results are obtained using individual Random Tree and IBk classifiers using 6 best reduced features. The aim of this work is to reduce the number of classifiers and thus help early treatment of the CKD patients.

Methodology

In this section, we provide the considered methodology with the purpose of analyzing the data.

Material

Chronic Kidney Disease dataset was created for the analysis of kidney disease and obtained from "UC Irvine Machine Learning Repository" database. This dataset contains twenty-four features and four hundred instances are used in this comparative analysis. The features in the CKD dataset are age, blood pressure, specific gravity, albumin, sugar levels, cerum creatinine, hemoglobin, hypertension, red blood cells, pus cell, pus cell clumps, bacteria, blood glucose, blood urea, cerum creatinine, sodium, potassium, hemoglobin, packed cell volume, white blood cell count, red blood cell count, hypertension, diabetes mellitus, coronary artery disease, appetite, pedal edema and anemia.

Our features contain 11 numeric and 13 nominal parts. Table 1 shows the description of the features in CKD. These factors are closely associated with kidney disease [15].

Kidney Disease Factors

Chronic Kidney Disease occurs slowly and gradually over many years. In deficient kidney function, the conditions cannot operate as in their former stages. Five stages of kidney disease can be given. As shown in Table 2, the GFR number is a reference for the kidney function activation. If the GFR number decreases, kidney disease becomes worse [16].

Classification Techniques

In this work, we use seven different classification techniques for the CKD data [17].

Naïve Bayes: The Naïve Bayes algorithm is based on Bayes theorem and is a probabilistic classifier. It calculates a set of proba-

Table 1. Feature properties of CKD data

Features	Type	Abbr.	Results
Age	num.	age	Avg. 51,5
Blood pressure (mm/Hg)	num.	bp	76.5
Blood glucose (mgs/dl)	num.	bgr	148.04
Blood urea (mgs/dl)	num.	bu	57.43
Serum Creatinine (mgs/dl)	num.	sc	3.07
Sodium (mEq/L)	num.	sod	Avg.137.53
Potassium (mEq/L)	num.	pot	Avg. 4.63
Hemoglobin (gms)	num.	hemo	Avg. 12.53
Packed cell volume	num.	pcv	Avg. 38.88
White blood cell count (cells/cumm)	num.	wbc	Avg. 8406.12
Red blood cell count (millions/ cmm)	num.	rbc	Avg. 4.71
Specific gravity (1.005, 1.010, 1.015, 1.020, 1.025)	nom.	sg	1.005 7 1.010 84 1.015 75 1.020 106 1.025 81
Albumin (0,1,2,3,4,5)	nom.	al	0 1 2 3 4 5 199 44 43 43 24 1
Sugar Degree (0,1,2,3,4,5)	nom.	su	0 1 2 3 4 5 290 1318 14 13 3
Red Blood Cells (normal, abnormal)	nom.	rbc	47 abnormal
Pus cell (normal, abnormal)	nom.	pc	76 abnormal
Pus cell clumps (present/ notpresent)	nom.	pcc	42 present
Bacteria (present/ notpresent)	nom.	ba	22 present
Hypertension (yes/no)	nom.	htn	147 yes
Diabetes Mellitus (yes/no)	nom.	dm	137 yes
Coronary artery disease (yes/no)	nom.	cad	34 yes
Appetite (good/poor)	nom.	appet	82 poor
Pedal edema (yes/no)	nom.	pe	76 yes
Anemia (yes/no)	nom.	ane	60 yes

Table 2. GFR and CKD relation

Stages of CKD	GFR (min/1.73m ²)
Stage 1 Normal kidney function	90 +
Stage 2 Digestible loss of kidney function	89 - 60
Stage 3a Digestible to slight loss of kidney function	59 - 45
Stage 3b Slight to severe loss of kidney function	44 - 30
Stage 4 Severe loss of kidney function	29 - 15
Stage 5 Kidney failure or dialysis	- 15

bilities by calculating the frequency and combinations of values in the data. It is exemplified by a vector $X = (x_1, x_2, \dots, x_n)$ representing n features (independent variables) and appoints to this sample probabilities $p(C_k | x_1, \dots, x_n)$ for each of k possible outcomes or classes.

Using Bayes' theorem, the conditional probability can be decomposed as follows:

$$p(C_k | X) = \frac{p(C_k)p(X|C_k)}{p(X)} \quad (1)$$

The Naive Bayes algorithm performs well and learns rapidly in several supervised classification processes.

Hoeffding Tree

A Hoeffding tree is an algorithm that is competent at learning from a big dataset. Hoeffding trees take advantage of small samples to choose a proper splitting feature. This idea is assisted by the Hoeffding bound, which quantifies the number of observations needed to estimate some statistics within a detected precision. The bound states probability is $1 - d$, the mean of a random variable of range R will not differ from the estimated mean after n independent observations by more than,

$$e = \sqrt{\frac{R^2 \ln(1/d)}{2n}} \quad (2)$$

where R is the base of the number of possible classes, n is the number of observations and e is the bound.

Random Tree (RT)

Random Tree is a supervised classifier and an ensemble learning algorithm. RT can generate many individual learners and cope with both classification and regression problems. Ran-

dom trees are an ensemble of tree estimators called forest. RT takes the input features and classifies all the trees in the forest. The classifier reply is the average of the replies over all the trees in the forest.

Reduce Error Pruning Tree (REPTree)

REPTree makes a regression tree by reducing the data gain as well as pruning the regression tree. It only sorts the data by considering numeric values and copes with deficient values by splitting the data into smaller pieces, as J48 does.

Random SubSpace

Random Subspace classifier is based on the ensemble learning algorithm. It constitutes a decision tree-based classifier that maintains the highest accuracy on training data. The classifier involves multi-trees and constructs them systematically with the feature vector and trees structured in chosen subspaces randomly.

Adaboost Algorithm

This is an ensemble learning algorithm to determine data classification and degradation. At first, it starts with an equal data distribution, then it finds the best classifier which has a weight below the threshold. Then, the algorithm updates the weights and focuses on erroneously classified samples. Therefore, after a certain number of iterations, the most powerful classifiers are combined to improve the overall performance of the classification.

IBk

IBk is a KNN classifier that uses a number of nearest neighbors. Furthermore, it can be specified using leave-one-out cross-validation focalize to an upper limit given by the determined value. Different kinds of search algorithms can be used to accelerate the search of nearest neighbors. For instance, KD-trees, ball trees and cover trees are further options than linear search. In this algorithm, the distance function is used a parameter of the search method.

Performance Measures

In this work, we used five performance measures such as accuracy, kappa, MAE, RMSE and F measures.

Accuracy: Accuracy is described as the percentage of correctly classified instances. Additionally, accuracy is one of the most popular classification metrics in classification data and given by,

$$Accuracy = \frac{TN + TP}{TP + FP + FN + TN} \quad (3)$$

where TN ; true negative, TP ; true positive, FN ; false negative and FP ; false positive.

Kappa statistic

This is a classifier performance measure between two sets of classified data. Kappa result values are between 0 to 1. The results become meaningful with increasing values of kappa, which is expressed as,

$$Kappa = \frac{P(A) - P(E)}{1 - P(E)} \quad (4)$$

Mean absolute error (MAE)

MAE is a quantity used to measure the predictions which deviate from the true probability. $P(i, j)$ is the estimated probability of i coefficient to be of class j . MAE takes values of between 0 to 1, and is given by,

$$MAE = \frac{\sum_{j=1}^c \sum_{i=1}^m |\hat{A}(i, j) - P(i, j)|}{m \times c} \quad (5)$$

Precision

This is given as the correlation number between the correctly classified modules to entire classified fault-prone modules. Precision is the number of correctly prescribed as faulty:

$$precision = \frac{TP}{TP + FP} \quad (6)$$

Recall

Recall is given as the average probability of complete retrieval, and given by,

$$recall = \frac{TP}{TP + FN} \quad (7)$$

Root Mean Square Error (RMSE)

RMSE is the square root of the mean distance between predicted and observed data. Denoting the estimated probability of the suitable data i and the target value for the suitable data j by $P(i, j)$ and T_j , respectively, RMSE is given as follows:

$$E_i = \sqrt{\frac{1}{n} \sum_{j=1}^n (P(i, j) - T_j)^2} \quad (8)$$

Table 3. Classification with Twenty-four Features

		Performance Measures				
		Accuracy	Kappa	MAE	RMSE	F Measure
Classifiers	Naive Bayes	94.17	0.879	0.057	0.230	avg: 0.942
	HoeffdingTree	95	0.896	0.051	0.223	avg: 0.951
	Random Tree	96.67	0.929	0.047	0.161	avg: 0.967
	REPTree	99.17	0.982	0.047	0.095	avg: 0.992
	Random SubSpace	99.17	0.982	0.071	0.117	avg: 0.992
	Adaboost	98.33	0.964	0.021	0.084	avg: 0.983
	IBk	95.83	0.912	0.045	0.204	avg: 0.959

Table 4. Classification with Twelve Features

		Performance Measures				
		Accuracy	Kappa	MAE	RMSE	F Measure
Classifiers	Naive Bayes	96.67	0.929	0.029	0.141	avg: 0.967
	HoeffdingTree	96.67	0.930	0.030	0.143	avg: 0.967
	Random Tree	98.33	0.967	0.027	0.114	avg: 0.983
	REPTree	99.17	0.982	0.047	0.095	avg: 0.992
	Random SubSpace	99.17	0.982	0.087	0.135	avg: 0.992
	Adaboost	98.33	0.964	0.017	0.095	avg: 0.987
	IBk	97.5	0.947	0.028	0.158	avg: 0.975

F-measure

F-measure is the one the combination of both precision and recall. *F-measure* uses the field of information retrieval in order to estimate classification performance, and is calculated by,

$$F - measure = \frac{2 \times precision \times recall}{precision + recall} \quad (9)$$

where *precision* and *recall* are as defined in (6) and (7).

Experimental Results

In this work, we perform seven different classification metrics to show the performance of classification systems. The result of the classification metrics is based on five performance measures, such as accuracy, kappa, MAE, RMSE and F measures, and these are shown in Table 3-6. All data is produced by the WEKA tool during the classification process. For the classification of healthy and sick people, "notckd" and "ckd" labels are used, respectively.

In our previous work, we used all features as well as the reduced features for the evaluation of two performance measures (accuracy and kappa values). Furthermore, we analyzed the performance measure differences of individual and ensemble classifiers. We showed that both ensemble algorithms and proposed feature selection methods are efficient tools to classify CKD. However, our performance measurement metrics changed differently [18, 19].

In this paper, a 70% training test is first completed for 24 features using 10-fold cross validation. Naïve Bayes, Hoeffding-Tree, RandomTree, REPTree, Random Subspaces, Adaboost and IBk classifiers are used to compare the CKD data. Performance measure metrics of the obtained data are shown in Tables 3-6. Then, we reduce 24 features to 12, 6 and 3 features and analyze the performance measurement metrics.

In Table 3, Random Tree and REPTree classifiers have the same quantity of accuracy, kappa and F measure values. At the same time, these values are the highest scores in all classifier methods. We also show that MAE and RMSE have the smallest values in other classifier methods.

Table 5. Classification with Six Features

		Performance Measures				
		Accuracy	Kappa	MAE	RMSE	F Measure
Classifiers	Naive Bayes	97.5	0.947	0.026	0.114	avg: 0.975
	HoeffdingTree	97.5	0.947	0.026	0.114	avg: 0.975
	Random Tree	100	1	0.028	0.095	avg: 1
	REPTree	99.17	0.982	0.047	0.095	avg: 0.992
	Random SubSpace	99.17	0.982	0.090	0.133	avg: 0.992
	Adaboost	99.17	0.982	0.017	0.085	avg: 0.992
	IBk	100	1	0.003	0.003	avg: 1

Table 6. Classification with Three Features

		Performance Measures				
		Accuracy	Kappa	MAE	RMSE	F Measure
Classifiers	Naive Bayes	94.17	0.879	0.093	0.223	avg: 0.942
	HoeffdingTree	93.33	0.859	0.151	0.270	avg: 0.934
	Random Tree	94.17	0.876	0.090	0.222	avg: 0.942
	REPTree	93.33	0.859	0.111	0.236	avg: 0.934
	Random SubSpace	94.17	0.879	0.162	0.240	avg: 0.942
	Adaboost	93.3	0.859	0.165	0.260	avg: 0.934
	IBk	93.3	0.862	0.088	0.221	avg: 0.934

In Table 4, we reduce 24 features to 12 by using *Information Gain Attribute Evaluator Feature Selection* method. Then, we analyze the data and use the following reduced features: specific gravity, diabetes mellitus, hypertension, red blood cells, albumin, and hemoglobin, packed cell volume, pus cell, pedal edema, appetite, anemia and red blood cell count features. As seen from Table 4, Random Tree and REPTree classifiers have the same quantity of accuracy. Kappa and F measure values as shown in Table 3. In addition, these values are the highest scores in all reduced classifier methods

In Table 5, we reduce 12 features to 6, which are specific gravity, diabetes mellitus, hypertension, red blood cells, albumin, and hemoglobin. Then, we analyze the performance metrics of the data with reduced features. Random Tree and IBk classifiers have the highest accuracy, kappa and F measure values than other classifiers. Random Tree, REPTree and Adaboost classifiers have the same quantity of accuracy, and kappa and F measure values as in the previous comparisons. We also observe that MAE and RMSE values of Adaboost and IBk have smaller values than other techniques.

Finally, in Table 6, in order to analyze the data, 6 features are reduced to 3, which are specific gravity, diabetes mellitus, and hypertension. Naïve Bayes, Random Tree and Random SubSpaces classifiers have the same and the highest quantity of accuracy, and kappa and F measure values. On the other hand, IBk classifier has smaller MAE and RMSE values than the other parameters.

Conclusion

In this paper, chronic kidney diseases were classified using various features and classifier combinations. Initially, seven individual classifiers were applied to twenty-four features and the best results were obtained using individual REPTree and Random SubSpaces classifiers as 99.17%. Then, the effect of *Information Gain Attribute Evaluator Feature Selection* method was observed and analyzed on the CKD data. Consequently, twelve, six and three best features with the best information values were selected and then seven classifiers were applied to these reduced features.

The best accuracy and kappa values were calculated with Random Tree and IBk classifiers as 100% by using the six best features. More-

over, it was shown that Random SubSpace technique has the highest accuracy and kappa values in every reduced type of features.

The early detection stages of CKD can help in the treatment period of the patients, and at the same time, this may help to prevent the disease from getting worse.

Classification techniques considered in this paper can be used and evaluated to find rapid solutions for the patient. The main aim of this study is to reduce the number of classifiers used so that CKD can be diagnosed efficiently and rapidly.

Peer-review: Externally peer-reviewed

Conflict of Interest: The authors declared that this study has received no financial support.

Financial Disclosure: The authors declared that this study has received no financial support.

References

1. A. S. Levey, J. Coresh, E. Balk, A. T. Kausz, A. Levin, M. W. Steffes, R. J. Hogg, R. D. Perrone, J. Lau, G. Eknoyan, "National Kidney Foundation Practice Guidelines for Chronic Kidney Disease: Evaluation, Classification, and Stratification", *Ann Intern Med*, vol. 139, no. 2, pp. 137-147, 2003.
2. A. S. Levey, J. P. Bosch, J. B. Lewis, T. Greene, N. Rogers, D. Roth, "A More Accurate Method To Estimate Glomerular Filtration Rate from Serum Creatinine: A New Prediction Equation", *Ann Intern Med*, vol. 130, no. 6, 461-470, 1999.
3. J. B. Arlet, J. A. Ribeil, G. Chatellier, D. Eladari, S. de Seigneux, J. C. Souberbielle, G. Friedlander, M. de Montalembert, J. Pouchot, D. Prie, M. Courbebaisse, "Determination of the best method to estimate glomerular filtration rate from serum creatinine in adult patients with sickle cell disease: a prospective observational cohort study", *BMC Nephrology*, vol. 13, no. 1, 2012.
4. K. Sumida, M. Z. Molnar, P. K. Potukuchi, F. Thomas, J. L. Lu, J. Jing, V. A. Ravel, M. Soohoo, C. M. Rhee, E. Streja, K. Kalantar Zadeh, C. P. Kovesdy, "Association of Slopes of Estimated Glomerular Filtration Rate With Post-End-Stage Renal Disease Mortality in Patients With Advanced Chronic Kidney Disease Transitioning to Dialysis", *Mayo Clin Proc*, vol. 91, no. 2, pp. 196-207, 2016.
5. A. S. Go, G. M. Chertow, D. Fan, C. E. McCulloch, C. Y. Hsu, "Chronic kidney disease and the risks of death, cardiovascular events, and hospitalization", *N Engl J Med*, vol. 351, no. 13, pp. 1296-1305, 2004.
6. Ş. Şengul, Y. Erdem, V. Batuman, Ş. Erturk, "Hypertension and Chronic Kidney Disease in Turkey", *Kidney Int Suppl* (2011), vol. 3, no. 4, pp. 308-311, 2013.
7. T. Liyanage, T. Ninomiya, V. Jha, B. Neal, H. M. Patrice, I. Okpechi, M. Zhao, J. Lv, A. X Garg, J. Knight, A. Rodgers, M. Gallagher, S. Kotwal, A. Cass, V. Perkovic, "Worldwide access to treatment for end-stage kidney disease: a systematic review", *Lancet*, vol. 385, no. 9981, pp. 1975-1982, 2015.
8. N. R. Hill, S. T. Fatoba, J. L. Oke, J. A. Hirst, C. A. O'Callaghan, D. S. Lasserson, F. D. Hobbs, "Global Prevalence of Chronic Kidney Disease-A Systematic Review and Meta-Analysis", *PLoS One*, vol. 11, no. 7, 2016.
9. L. Jena, N. K. Kamila, "Distributed Data Mining Classification Algorithms for Prediction of Chronic- Kidney-Disease", *International Journal of Emerging Research in Man & Tech*, vol. 4, no. 11, pp. 110-118, 2015.
10. Weka 3: Data Mining Software in Java. (n.d.). Retrieved February 24, 2018, from <http://www.cs.waikato.ac.nz/ml/weka/>.
11. A. Chaudhary, P. Garg, "Detecting and Diagnosing a Disease by Patient Monitoring System", *International Journal of Mechanical Engineering and Information Technology*, vol. 2, no. 6, pp. 493-499, 2014.
12. P. Baby, T. P. Vital, "Statistical Analysis and Predicting Kidney Diseases using Machine Learning Algorithms", *International Journal of Engineering and Technical Research*, vol. 7, no. 7, 2015.
13. P. Sinha, P. Sinha, "Comparative Study of Chronic Kidney Disease Prediction using KNN and SVM", *International Journal of Engineering and Technical Research*, vol. 4, no. 12, 2015.
14. S. Vijayarani, S. Dhayanand, "Data Mining Classification Algorithms for Kidney Disease Prediction", *International Journal on Cybernetics & Informatics*, vol. 4, no. 4, pp. 13-25, 2015.
15. M. Lichman, UC Irvine Machine Learning Repository, University of California "<http://archive.ics.uci.edu/ml/>", Irvine, School of Information and Computer Sciences, 2013.
16. Chronic Kidney Disease Overview. Retrieved February 24, 2018, from <https://www.webmd.com/a-to-z-guides/tc/chronickidney-disease-topic-overview>.
17. S. Mitra, Introduction to machine learning and bioinformatics. Boca Raton: CRC Press, 2008.
18. M. Doğruyol Başar, P. Sarı, N. Kılıç, A. Akan, "Detection of Chronic Kidney Disease by Using Adaboost Ensemble Learning Approach", *24th Signal Processing and Communication Application Conference (SIU)*, 2016.
19. M. Doğruyol Başar, A. Akan, "Detection of Chronic Kidney Disease by Using Ensemble Classifiers", *10th International Conference on Electrical and Electronics Engineering (ELECO)*, pp. 544-547, 2017.



Merve Doğruyol Başar was born in Istanbul, Turkey. She received the B.S. Degree from Marmara University, Istanbul, Turkey, in 2012, and the M.S. Degree from the Istanbul Technical University, Istanbul, Turkey, in 2014. She is currently a research assistant at Istanbul University - Cerrahpasa while pursuing her Ph.D. Degree at Istanbul University. Her primary research interests include biomedical signal processing and machine learning. She is a student member of the IEEE.



Aydın Akan received the B.Sc. degree from the University of Uludağ, Bursa, in 1988, the M.Sc. degree from the Technical University of Istanbul, Istanbul, Turkey in 1991, and the Ph.D. degree from the University of Pittsburgh, Pittsburgh, PA, USA, in 1996, all in electrical engineering. He was with the Department of Electrical and Electronics Engineering, University of Istanbul, between 1996 and 2017. He is currently a Professor at the Department of Biomedical Engineering, Izmir Katip Celebi University, Izmir, Turkey. His current research interests include nonstationary signal processing, time-frequency signal analysis methods and their applications to wireless communications and biomedical engineering. He is a senior member of the IEEE Signal Processing Society and an Associate Editor of the Digital Signal Processing Journal.