

Kronik Böbrek Hastalığı Tespitinde Farklı Sınıflandırma Yöntemleri ve Farklı Topluluk Algoritmalarının Birlikte Kullanımının Sınıflandırma Performansına Etkisi

The Impact on the Classification Performance of the Combined Use of Different Classification Methods and Different Ensemble Algorithms in Chronic Kidney Disease Detection

Kübra Eroğlu¹, Tuğba Palabaş²

¹Elektrik-Elektronik Mühendisliği Bölümü, ²Bilgisayar Mühendisliği Bölümü

İstanbul Arel Üniversitesi

kubraeroglu@arel.edu.tr, tugbapalabas@arel.edu.tr

Özet

Yapılan çalışmanın amacı kronik böbrek hastalığı tespitinde farklı sınıflandırma yöntemleri ve topluluk algoritmalarının performans değerlendirme sonuçlarının karşılaştırılmasıdır. Çalışmada altı farklı temel sınıflandırıcı (naive bayes, k en yakın komşuluk (KNN), destek vektör makineleri (DVM), J48, rastgele ağaç, karar tablosu) ve üç farklı topluluk algoritması (adaboost, bagging, rastgele alt uzaylar) kullanılmıştır. Sınıflandırma sonuçları üç farklı performans değerlendirme ölçütü (doğruluk, κ , ROC eğrisi altında kalan alan (AUC)) kullanılarak değerlendirilmiştir. Elde edilen performans değerlendirme sonuçlarına göre J48 temel sınıflandırıcısının bagging ve rastgele alt uzaylar topluluk algoritmalarıyla, rastgele ağaç temel sınıflandırıcısının ise bagging topluluk algoritması ile birlikte kullanımı % 100 sınıflandırma başarısı sağlamıştır.

Abstract

The aim of this study is to compare the performance assessment results of the different classification methods and ensemble algorithms for the detection of chronic kidney disease. Six different basic classifier (naive bayes, k nearest neighbor (KNN), support vector machines (SVM), J48, random trees, decision tables) and three different ensemble algorithm (adaboost, bagging, random subspace) are used in the study. Classification results were evaluated using three different performance evaluation criteria (accuracy, κ , the area under the ROC curve (AUC)). According to the performance evaluation results, J48 basis algorithm for use with bagging and random subspace ensemble algorithms and random tree basis algorithm for use with bagging ensemble algorithm has provided 100% classification success.

1. Giriş

Kronik böbrek yetmezliği yaşamı tehdit eden önemli iş gücü kaybına ve çeşitli komplikasyonlara yol açan, hemen her yaş grubunu en çok da genç erişkinleri etkileyen bir hastalıktır [1-2]. Böbrek fonksiyonlarında veya glomerular filtrasyon hızında önemli bir azalış olması durumunda ortaya çıkmakta olup tansiyon yüksekliği, kemik hastalığı, kansızlık, kalp ve kan damarı hastalıkları gibi komplikasyonlara yol açabilmektedir. Kronik böbrek hastalığı çoğunlukla uzun süre belirti vermediği için tanısı, kanda kreatinin ölçümü ve glomerüler filtrasyon hızı ölçümü ve idrar- kan örnekleri ile yapılan testler sonucunda konulabilir. Bununla birlikte radyolojik testler sonucunda da hastalık tespiti yapılabilir. Tespiti yapılamayan kronik böbrek hastalığında ileride karşılaşılabilecek en büyük problem böbrek yetmezliği ve diyaliz veya böbrek nakline sebep olabilecek böbrek işlev kaybıdır. Bununla birlikte kalp-damar hastalıklarına bağlı olarak erken ölümler de görülebilmektedir.

Kronik böbrek hastalığının erken evrelerinde böbrek fonksiyonunun azalması ile kardiyovasküler hastalık ve erken ölümler gibi olumsuz sonuçlar daha yüksek olmaktadır [3]. Erken teşhis sağlanırsa gelecek dönemde görülebilecek komplikasyon riski azaltılabilir. Kronik böbrek yetmezliğinin fiziksel etkileri dışında psikolojik anlamda moralite düşüklüğü, hastalığın evresine bağlı olarak sürekli sağlık merkezine bağımlı yaşama, tedavi sürecinin ekonomik etkileri gibi sosyal, ekonomik ve psikolojik etkileri de bulunmaktadır. Kronik böbrek hastalığı oluşturabileceği hayati tehlike ve tedavi sürecindeki yüksek maliyet sebebiyle önemli bir hastalık olup teşhisi ne kadar erken konulursa organ kaybı ve hayati tehlike riski o ölçüde azalacaktır. Bu sebeptendir ki kronik böbrek hastalığının belirtilerinin erken teşhis edilmesi ve teşhis konulması durumunda tedavi sürecine girilmesi hem hayati hem de ekonomik anlamda büyük bir önem taşımaktadır. Konuyla ilgili literatürde ki bir takım çalışmalardan bahsedilecek olursa; [4] ve [5]' de kronik böbrek hastalığının farklı evrelerini tespit edebilmek için

ultrasonik görüntülerden yararlanılmıştır. [6]' da kronik böbrek hastalarının sürekli kontrolünü gerçekleştirmek için uzaktan gözetim sistemi tasarımı ve uygulaması önerilmiştir. [7]' de kronik böbrek hastalığı sınıflandırma çalışmasında yapay sinir ağları ve naive bayes, [8]' de ise destek vektör makineleri, KNN ve naive bayes yöntemi kullanılmıştır. Yapılan çalışmanın amacı ise kronik böbrek hastalığı tespitinde daha önce literatürde kullanılmış olan ve olmayan toplam altı farklı sınıflandırma yöntemi ile hastalığın tespitinin yapılması ve sınıflandırma başarısının belirlenen bir takım kriterler ile değerlendirilerek (doğruluk (%), kappa, AUC) değerlendirme sonuçlarının karşılaştırılmasıdır. Bununla birlikte üç farklı topluluk algoritması kullanılarak (adaboost, bagging, rastgele alt uzaylar) topluluk algoritmalarının sınıflandırma başarısı üzerindeki etkileri ayrıntılı olarak incelenmiştir.

2. Yöntem

2.1. Materyal

Çizelge 1: Kronik böbrek hastalığı tespitinde kullanılan veri grubu.

Kronik Böbrek Hastalığı Tespiti İçin Kullanılan 24 Veri	Kısaltma	Veri Tipi
Yaş (age)	age	Nicel
Kan Basıncı (blood pressure)	bp	Nicel
Spesifik Gravite (specific gravity) (1.005, 1.010, 1.015, 1.020, 1.025)	sg	Nitel
Albumin (0, 1, 2, 3, 4, 5)	al	Nitel
Şeker (sugar) (0, 1, 2, 3, 4, 5)	su	Nitel
Kırmızı Kan Hücreleri (red blood cells) (normal/ anormal)	rbc	Nitel
İltihap Hücresi (pus cell) (normal/ anormal)	pc	Nitel
İltihap Hücre Kümeleri (pus cell clumps) (Mevcut/Mevcut Değil)	pcc	Nitel
Bakteri (bacteria) (Mevcut/Mevcut Değil)	ba	Nitel
Rastgele Kan Şekeri (blood glucose random)	bgr	Nicel
Kan Üresi (blood urea)	bu	Nicel
Serum Kreatini (serum creatinine)	sc	Nicel
Sodyum (sodium)	sod	Nicel
Potasyum (potassium)	pot	Nicel
Hemoglobin (hemoglobin)	hemo	Nicel
Sıkıştırılmış Eritrosit Hacmi (packed cell volume)	pcv	Nicel
Beyaz Kan Hücreleri Sayımı (white blood cell count)	wc	Nicel
Kırmızı Kan Hücreleri Sayımı (red blood cell count)	rc	Nicel
Hipertansiyon (hypertension) (Evet/Hayır)	htn	Nitel
Diabet (diabetes mellitus) (Evet/Hayır)	dm	Nitel
Koroner Arter Hastalığı (coronary artery disease) (Evet/Hayır)	cad	Nitel
İştah (appetite) (İyi/Kötü)	appet	Nitel
Ayak Ödemi (pedal edema) (Evet/Hayır)	pe	Nitel
Anemi (anemi) (Evet/Hayır)	ane	Nitel

Yapılan çalışmada kronik böbrek hastalığı tespiti için 400 kişiye ait 24 farklı değerden (11 nicel ve 13 nitel) oluşan veri kümesi kullanılmıştır. Bahsedilen veri kümesi 'UC Irvine Machine Learning Repository' veri tabanından elde edilmiştir [9]. Veri kümesi 250 kronik böbrek hastası ve 150 sağlıklı olmak üzere toplam 400 bireye ait 24 farklı değerden meydana gelmektedir. Bahsedilen 24 farklı değerlerin bir kısmının varlığı bir kısmının da miktarı hastalığın oluşumunda etkili olabileceği düşünülen değerlerdir. Kronik böbrek hastalığı tespiti için kullanılan veri kümesindeki tüm bireylerden elde edilen 24 farklı değer isimleri sırasıyla Çizelge 1' de ayrıntılı olarak ifade edilmiştir. Çizelgede, veri isimlerinin kısaltılmış karşılıkları ve veri tipine ait bilgiler bulunmaktadır. Çalışmada gerçekleştirilen tüm işlemler 'Waikato Environment for Knowledge Analysis (WEKA)' yazılımı ile gerçekleştirilmiştir.

2.2. Öz Nitelik Çıkarma

Yapılan çalışmada her birey için Çizelge 1 ile ifade edilen 24 farklı kayıt verisi öz nitelik olarak kullanılmıştır. Materyal kısmında bahsedildiği gibi veri tabanında toplam 400 kişiye ait veri kayıt bilgileri bulunmakta olup öz nitelik kümesinin boyutları da (400 x 24)' tür. Her bir kayıt (1 x 24) vektör uzunluğunda olup vektörün her bir sütunu sırasıyla Çizelge 1' de belirtilen nicel ve nitel veri değerlerinden oluşmaktadır. Öz nitelik kümesine ait ayrıntılı bilgiye [9]' dan ulaşılabilir. Veri kümesinde 250 kronik böbrek hastası ve 150 sağlıklı (kronik böbrek hastası olmayan) olmak üzere toplam 400 bireye ait kayıt bilgisi bulunmaktadır. Kronik böbrek hastası olan bireylere ait kayıtlar 'ckd', sağlıklı bireylere ait kayıtlar ise 'notckd' sınıf etiketi ile etiketlenilmiştir (Çizelge 2).

Çizelge 2: Öz nitelik veri kümesi boyutları.

Kronik böbrek hastası bireylere ait kayıtlar (ckd)	Kronik böbrek hastası olmayan (sağlıklı) bireylere ait kayıtlar (notckd)
(250 x 24)	(150 x 24)
Toplam= (400 x 24)	

2.3. Sınıflandırma

Çalışmada amaçlanan, (1 x 24) vektör uzunluğundaki kayıt bilgileri kullanılarak kronik böbrek hastalığı tespitinin gerçekleştirilebilmesidir. Bu amaçla 400 kişiye ait kayıt bilgileri temel sınıflandırıcılar kullanılarak sınıflandırılmıştır. Çalışmada performans karşılaştırması yapabilmek amacıyla altı farklı temel sınıflandırıcı (naive bayes, KNN, DVM, J48, rastgele ağaç, karar tablosu) ve temel sınıflandırıcılara ait sonuçların iyileştirilip daha kararlı hale getirilebilmesi için de üç farklı topluluk algoritması (adaboost, bagging, rastgele alt uzaylar) kullanılmıştır.

Naive bayes, istatistiksel yöntemler kullanarak sınıflandırma yapan ve sınıflardaki niteliklerin birbirinden bağımsız olduğunu varsayan basit ve etkili bir sınıflandırıcı tekniğidir. DVM bir makine öğrenme yöntemi olup iki sınıfa ait verileri birbirinden ayırt edebilen en uygun hiper düzlem veya düzlemi belirlemektedir. KNN yöntemi ise sınıflandırma problemlerinde kullanılan denetimli öğrenme algoritmalarından birisi olup önceden belirlenen bir uzaklık kriterine göre test örneğinin eğitim kümesinde bulunan örneklerle olan yakınlığını hesaplar. Bu işlemin ardından en yakın olan k tane örneği tespit edip bu örnekler en çok hangi

sınıfa ait ise test örneğini o sınıfa dahil eder.

J48, sınıflandırma yapmak için entropi kavram bilgisini kullanan bir karar ağacıdır. Budanmış bir C4.5 ağacı üretmek için ve Quinlan'ın C4.5 algoritmasını [10] uygular. Karar verme her bir öznelik veri kümesini entropi farklılıklarını incelemek amacıyla alt kümelerle bölerek yapılır. En yüksek normalize bilgi kazançlı öz nitelikler seçilir [11]. *Rastgele ağaç* sınıflandırma yönteminde her düğümde belirli bir sayıda rastgele seçilen özelliklerden ağaç oluşturulur ve bu yöntemde budama işlemi yoktur [12-13]. *Karar tabloları* sınıflandırma yönteminde bilgi satır ve sütunlar kullanılarak tablo şeklinde saklanılır. Bir karar tablosunun dört bileşeni vardır; durum taslakları, durum kayıtları, eylem taslakları ve eylem girişleri. Sol üst bölge koşulları içerir, sağ üst bölge durum kuralları veya alternatifleri içerir, sol alt bölge alınacak önlemleri, sağ alt bölge ise eylem kurallarını içermektedir. Doğrulama ve doğrulama karar tablosunu çelişki, tutarsızlık, eksiklik ve fazlalık gibi durumlarda kontrol etmek kolaydır [14].

Topluluk algoritmalarında birçok eğitici eğitildikten sonra vermiş oldukları kararlar birleştirilmektedir. *Adaboost* topluluk algoritmasında yanlış sınıflandırılan örnekler daha fazla odaklanılarak başarımın artırılması amaçlanmaktadır. Bu amaçla eğitim kümesindeki tüm örnekler bir ağırlık değeri atanır. Algoritmanın kendini her tekrar etmesinde yanlış sınıflandırılan örnekler ilişkin ağırlık değerleri artırılırken, doğru sınıflandırılmış örneklerin ağırlık değerleri azaltılır. *Bagging* topluluk algoritmasında N boyutlu eğitim kümesinden, aynı boyuttaki (N boyutlu) eğitim kümesi rastgele ve tekrarlı olarak üretilir. Üretilen her bir eğitim kümesi eğitilip sonuçları çoğunluk oylaması ile birleştirilir. *Rastgele alt uzaylar* topluluk algoritmasında ise öğrencilerin her biri eğitim kümesindeki özelliklerin rastgele seçilmiş alt kümesi ile eğitilmektedir.

3. Bulgular

Yapılan çalışmada Çizelge 2' de belirtilen iki farklı etiket ile etiketlendirilmiş (*ckd*, *notckd*) öz nitelik veri grupları 10 katlı çapraz doğrulama yöntemi ile eğitim ve test aşamasından geçirilmiştir. Bu işlemdeki amaç eğitim ve test aşamalarında daha güvenilir sonuçlar elde edilebilmesidir. Eğitim işleminin ardından test verileri altı farklı temel sınıflandırıcı (naive bayes, KNN, DVM, J48, rastgele ağaç, karar tablosu) ile sınıflandırılmıştır. Sınıflandırma sonuçlarının değerlendirilmesi aşamasında üç farklı performans değerlendirme kriteri kullanılmıştır; '*Doğruluk*', '*Kappa*', '*AUC*'. '*Doğruluk*' denklem (1) ile ifade edilmiştir. Denklemden dn , doğru sınıflandırılan negatif örnekleri, dp , doğru sınıflandırılan pozitif örnekleri, yn , yanlış sınıflandırılan negatif örnekleri, yp ise yanlış sınıflandırılan pozitif örnekleri ifade etmektedir.

$$\text{doğruluk} = \frac{dn+dp}{dn+dp+yp+yn} \quad (1)$$

$$\text{duyarlılık} = \frac{dp}{dp+yn} \quad (2)$$

$$\text{seçicilik} = \frac{dn}{dn+yp} \quad (3)$$

Kappa hesabı tanı uyumlarının ölçüldüğü istatistiktir. *Kappa* katsayısı 0-1 arasında değerler alır. Buna göre, '1' mükemmel uyumu, '0' ise zayıf uyumu temsil etmektedir. Diğer performans değerlendirme kriteri olan *AUC* katsayısı ise denklem (2) ve denklem (3) kullanılarak elde edilen alıcı işletim karakteristiği (receiver operating characteristic- ROC) eğrileri altında kalan alan hesabı ile elde edilmektedir. 0.5 üzerindeki *AUC* değerleri iyi performans anlamına gelmektedir. Altı farklı temel sınıflandırıcı kullanılarak elde edilen sınıflandırma performans değerlendirme sonuçları Çizelge 3 ile verilmiştir. Çizelge 3 incelendiğinde J48 ve Karar Tablosu temel sınıflandırıcıları kullanılarak elde edilen performans değerlendirme sonuçlarının diğer temel sınıflandırıcı performans değerlendirme sonuçlarından daha yüksek olduğu görülmektedir.

Çizelge 3: Altı farklı temel sınıflandırıcıya ait performans değerlendirme sonuçları.

Performans Değerlendirme Kriterleri	Naive Bayes	DVM	KNN	J48	Rastgele Ağaç	Karar Tablosu
Doğruluk (%)	95	97.75	95.75	99	95.50	99
Kappa	0.89	0.95	0.91	0.97	0.90	0.97
AUC	1	0.98	0.96	0.99	0.99	0.99

Yapılan çalışmadaki bir diğer amacımız ise topluluk algoritmaları kullanımının kronik böbrek hastalığı tespitindeki başarısını gözlemlemektir. Bu amaçla üç farklı topluluk algoritması (adaboost, bagging, rastgele alt uzaylar) ayrı olarak altı farklı temel sınıflandırıcı ile birlikte kullanıldı ve aynı performans değerlendirme testlerinden geçirildi. Elde edilen sonuçlar Çizelge 4 ile sunulmuştur.

Çizelge 4: Farklı topluluk algoritmaları kullanılarak yapılan sınıflandırmalara ait performans değerlendirme sonuçları.

Topluluk Algoritmaları	Performans Değerlendirme Kriterleri	Naive Bayes	DVM	KNN	J48	Rastgele Ağaç	Karar Tablosu
Adaboost	Doğruluk (%)	98	98.75	95.75	99.50	98	99.50
	Kappa	0.95	0.97	0.91	0.98	0.95	0.98
	AUC	0.99	0.99	0.96	1	0.99	0.99
Bagging	Doğruluk (%)	95.50	97.75	95.75	100	100	98.75
	Kappa	0.90	0.95	0.91	1	1	0.97
	AUC	1	0.99	0.96	1	1	0.99
Rastgele Alt Uzaylar	Doğruluk (%)	95.75	98.25	98.75	100	99.50	99.50
	Kappa	0.91	0.96	0.96	1	0.98	0.98
	AUC	1	1	0.99	1	1	0.99

Çizelge 4 incelendiğinde adaboost topluluk algoritmasının KNN temel sınıflandırıcısı hariç tüm sınıflandırıcılarda, bagging topluluk algoritmasının naive bayes, J48 ve rastgele ağaç sınıflandırıcılarında, rastgele alt uzaylar topluluk algoritmasının ise tüm temel sınıflandırıcılarda performans değerlendirme sonuçlarını artırdığı görülmektedir. Bu artış özellikle Bagging topluluk algoritması ile J48 ve rastgele ağaç temel sınıflandırıcılarının birlikte kullanıldığı durumlarda en yüksek seviyeye (doğruluk=% 100, Kappa= 1, AUC= 1) ulaşmıştır. Çalışmanın bir diğer aşamasında ise her bireye ait 24 öz nitelik arasından 10 tanesi, ilişki tabanlı öz nitelik seçme (correlation based feature selection- CFS) yöntemiyle seçilerek kullanılan öz nitelik sayısı indirgenmiş olup bu durumun kronik böbrek hastalığı tespitindeki performansa etkisi incelenmiştir. CFS yönteminde öz nitelik kümesi alt kümelerle bölünür ve bu alt kümelerin sınıflandırma ile olan

ilişkisi araştırılır. CFS yöntemi kullanılarak seçilen 10 öz nitelik sırasıyla; kan basıncı, spesifik gravite, albümin, kırmızı kan hücreleri, kan şekeri, kreatinin, sodyum, potasyum, hemoglobin, hemotokrit' dir. Öz nitelik seçme sonucunda elde edilen yeni öz nitelik kümesinin boyutları (400 x 10) olmuştur. Yeni öz nitelik kümesinin altı farklı temel sınıflandırıcı ile sınıflandırılması sonucu elde edilen performans değerlendirme sonuçları Çizelge 5 ile verilmiştir.

Çizelge 5: İndirgenmiş öz nitelikler için temel sınıflandırıcılara ait performans değerlendirme sonuçları.

Performans Değerlendirme Kriterleri	Naive Bayes	DVM	KNN	J48	Rastgele Ağaç	Karar Tablosu
Doğruluk (%)	96	96.25	98.75	97	97.75	99
Kappa	0.91	0.92	0.97	0.93	0.94	0.97
AUC	1	0.96	0.99	0.99	0.99	0.99

Çizelge 5, Çizelge 3 ile karşılaştırıldığında naive bayes, KNN, rastgele ağaç ve karar tablosu temel sınıflandırıcıları kullanılarak yapılan sınıflandırma performans değerlendirme sonuçlarında artış olduğu gözlemlenmektedir. İndirgenmiş öz niteliklerden oluşan yeni veri kümesinin sınıflandırılmasında topluluk algoritmalarının başarısı araştırıldığında ise Çizelge 6 elde edilmiştir.

Çizelge 6: İndirgenmiş öz nitelikler için farklı topluluk algoritmaları kullanılarak yapılan sınıflandırmalara ait performans değerlendirme sonuçları.

Topluluk Algoritmaları	Performans Değerlendirme Kriterleri	Naive Bayes	DVM	KNN	J48	Rastgele Ağaç	Karar Tablosu
Adaboost	Doğruluk (%)	98.75	96.75	98.75	99.25	98.25	99.50
	Kappa	0.97	0.93	0.97	0.98	0.96	0.98
	AUC	0.99	0.99	0.99	1	0.99	1
Bagging	Doğruluk (%)	95.75	96.75	99	99.25	98.75	99.75
	Kappa	0.91	0.93	0.97	0.98	0.97	0.99
	AUC	1	0.99	0.99	1	1	1
Rastgele Alt Uzaylar	Doğruluk (%)	97	96.75	97.50	99.75	99.50	98.50
	Kappa	0.93	0.93	0.94	0.99	0.98	0.96
	AUC	1	0.99	0.99	1	1	0.99

Çizelge 6 incelendiğinde adaboost topluluk algoritması kullanımının tüm temel sınıflandırıcı performans değerlendirme sonuçlarını, bagging topluluk algoritması kullanımının naive bayes hariç diğer tüm temel sınıflandırıcı performans değerlendirme sonuçlarını, rastgele alt uzaylar topluluk algoritması kullanımının ise KNN ve karar tablosu hariç tüm temel sınıflandırıcı performans değerlendirme sonuçlarını arttırdığı gözlemlenmektedir.

Elde edilen tüm sonuçlar incelendiğinde kronik böbrek hastalığı tespitinde en iyi performans gösteren sınıflandırıcı ve topluluk algoritmaları Çizelge 7 ile özetlenmiştir. Çizelgeye göre J48 temel sınıflandırıcısının bagging ve rastgele alt uzaylar topluluk algoritmasıyla, rastgele ağaç temel sınıflandırıcısının ise bagging topluluk algoritması ile birlikte kullanımı (öz nitelik sayısı=24 için) % 100 sınıflandırma başarısı sağlamıştır

Çizelge 7: Altı farklı temel sınıflandırıcı için iyi performans değerlendirme sonuçları.

(Sınıflandırıcı Tipi)	Öz Nitelik Sayısı	Topluluk Algoritması	Performans Değerlendirme Kriterleri
Naive Bayes	10	Adaboost	Doğruluk: % 98.75 Kappa: 0.97 AUC: 0.99
DVM	24	Adaboost	Doğruluk: % 98.75 Kappa: 0.97 AUC: 0.99
KNN	10	Bagging	Doğruluk: % 99 Kappa: 0.97 AUC: 0.99
J48	24	Bagging, Rastgele Alt Uzaylar	Doğruluk: % 100 Kappa: 1 AUC: 1
Rastgele Ağaç	24	Bagging	Doğruluk: % 100 Kappa: 1 AUC: 1
Karar Tablosu	10	Bagging	Doğruluk: % 99.75 Kappa: 0.99 AUC: 1

4. Sonuçlar ve Yorum

Yapılan çalışmada 150 sağlıklı ve 250 kronik böbrek hastası olmak üzere toplam 400 kişiye ait 24 veri kayıt bilgisi öz nitelik olarak kullanılmış olup bu öz nitelikler ile kronik böbrek hastalığı tespiti gerçekleştirilmiştir. Çalışmada amaçlanan kronik böbrek hastalığı tespitinde altı farklı temel sınıflandırıcıdan en iyi performans gösteren sınıflandırıcıyı bulmak ve bu işlemin ardından üç farklı topluluk algoritması kullanarak bu performansı daha iyi ve kararlı hale getirebilen topluluk algoritmasını belirleyebilmektir. Ayrıca kronik böbrek hastalığı tespitinde öz nitelik olarak kullanılan veri kayıt sayısının indirgenmesinin sınıflandırma performansını nasıl etkilediği de incelenmiştir. Çalışma sonuçları Çizelge 7 ile özetlenmiş olup J48 ve rastgele ağaç temel sınıflandırıcıları ile bagging ve rastgele alt uzaylar topluluk algoritmalarının kronik böbrek hastalığı tespitinde en yüksek performans (doğruluk= %100, kappa=1, AUC=1, öz nitelik sayısı=24) göstermiş olduğu ortaya konulmuştur. Ayrıca öz nitelik sayısı indirgenmesinin (10 öz nitelik) de performansla olumlu anlamda etki ettiği, bu durumun özellikle karar tablosu temel sınıflandırıcısının bagging topluluk algoritması ile birlikte kullanımında yüksek performans (doğruluk= %99.75, kappa= 0.99, AUC= 1) gösterdiği gözlemlenmiştir. Çalışma sonuçlarının konuyla ilgili yapılacak olan diğer çalışmalara ışık tutacağı düşünülmektedir.

5. Kaynaklar

- [1] Şentürk, A., Levent, B. A., ve Tamam, " Hemodiyalize Giren Kronik Böbrek Yetmezliği Olan Hastalarda Psikopatoloji ", *O.M.Ü Tıp Dergisi* , 17(3), 2000, 163-172.
- [2] Çevik, A. İzmir, M. ve diğerleri, " Psikosomatik Bozukluklar ", *Medikomat* , 1996, 55-63.
- [3] Levey, A.,S., Eckardt, K. ve diğerleri, "Definition and classification of Chronic Kidney Disease: A Position Statement from Kidney Disease: Improving Global

- Outcomes (KDIGO)", *Kidney International*, Vol. 67, 2089-2100, 2005.
- [4] Ho, C. Y., Pai, T. W. ve Peng, Y. C., " Ultrasonography Image Analysis for Detection and Classification of Chronic Kidney Disease", *2012 Sixth International Conference on Complex, Intelligent and Software Intensive Systems*, 2012, 624-629.
 - [5] Ho, C. Y., Pai, T. W. ve Peng, Y. C., " Analysis of Ultrasound Images for Identification of Chronic Kidney Disease Stages", *2014 First International Conference on Networks & Soft Computing*, 2014, 380-383.
 - [6] Cuevas, J. R., Dominguez, E. L. ve Velazquez, Y. H., " Telemonitoring Systems for Patients with Chronic Kidney Disease Undergoing Peritoneal Dialysis", *IEEE Latin America Transactions*, 2016, 2000-2006.
 - [7] Kunvar, V., Chander, K. ve diğerleri, " Chronic Kidney Disease Analysis Using Data Mining Classification Techniques", *2016 6th International Conference - Cloud System and Big Data Engineering*, 2016, 300-305.
 - [8] Chetty, N., Vaisla, K. S. ve Sudarsan, S. D., " Role of Attributes Selection in Classification of Chronic Kidney Disease Patients", *Computing, Communication and Security (ICCCS)*, 2015, 1-6.
 - [9] https://archive.ics.uci.edu/ml/datasets/Chronic_Kidney_Disease#
(Erişim Tarihi: Ağustos 2016)
 - [10] Quinlan, R., 'C4.5: Programs for Machine Learning', Morgan Kaufmann Publishers, San Mateo, CA, 1993.
 - [11] Salama, G. I., Abdelhalim, M. B. ve Zeid, M. A., " Experimental Comparision of Classifiers for Breast Cancer Diagnosis", *2012 7th International Conference on Computer Engineering& Systems (ICCES)*, 2012, 180-185.
 - [12] Akçetin, E. ve Çelik, U., "İstenmeyen Elektronik Posta (Spam) Tespitinde Karar Ağacı Algoritmalarının Performans Kıyaslaması", *İnternet Uygulamaları ve Yönetimi*, Cilt No. 5, 45-56, 2014.
 - [13] Breiman, L., 'Random Forests', *Machine Learning*, Vol. 45, 5-32, 2001.
 - [14] Shamim, A. I., Hussain, H. ve Shaikh, M. U., " A Framework for Generation of Rules from Decision Tree and Decision Table", *2010 International Conference on Information and Emerging Technologies (ICIET)*, 2010, 1-6.