Avrupa Bilim ve Teknoloji Dergisi
Sayı 20, S. 448-455, Aralık 2020
© Telif hakkı EJOSAT'a aittir
Araştırma Makalesi



www.ejosat.com ISSN:2148-2683

European Journal of Science and Technology
No. 20, pp. 448-455, December 2020
Copyright © 2020 EJOSAT
Research Article

Kronik Böbrek Hastalığının Erken Tanısı için Yeni Bir Klinik Karar Destek Sistemi

Can Eyüpoğlu [1*]

[1*] Milli Savunma Üniversitesi, Hava Harp Okulu, Bilgisayar Mühendisliği Bölümü, İstanbul, Türkiye (ORCID: 0000-0002-6133-8617), caneyupoglu@gmail.com, ceyupoglu@hho.edu.tr

(İlk Geliş Tarihi 27 Mayıs 2020 ve Kabul Tarihi 25 Ekim 2020)

(**DOI:** 10.31590/ejosat.743652)

ATIF/REFERENCE: Eyüpoğlu, C. (2020). Kronik Böbrek Hastalığının Erken Tanısı için Yeni Bir Klinik Karar Destek Sistemi. *Avrupa Bilim ve Teknoloji Dergisi*, (20), 448-455.

Öz

Kronik böbrek hastalığı dünya çapında bir sağlık sorunudur. Erken tanı ve tedavi sayesinde bu hastalığın ilerlemesini yavaşlatmak veya durdurmak mümkün olmaktadır. Klinik karar destek sistemleri, tıp doktorlarına klinik karar verme görevlerinde yardımcı olmak amacıyla tasarlanan sağlık bilgi teknolojisi sistemleridir. Bu çalışmada kronik böbrek hastalığının erken tanısı için yeni bir klinik karar destek sistemi önerilmiştir. Önerilen sistemin özellik çıkarma ve sınıflandırma aşamalarında sırasıyla temel bileşen analizi (principal component analysis-PCA) ve rastgele ormanlar (random forests-RF) teknikleri kullanılmıştır. Önerilen sistemin performansı, altı farklı performans metriği ile klasik makine öğrenmesi algoritmaları ve literatürde daha önce yapılan çalışmalar ile kıyaslanmıştır. Test sonuçları, önerilen sistemin başarılı olduğunu ve kronik böbrek hastalığının erken tanısı için karar vermede doktorlara destek olabileceğini göstermektedir.

Anahtar Kelimeler: Kronik Böbrek Hastalığı, Erken Tanı, Klinik Karar Destek Sistemi, Temel Bileşen Analizi, Rastgele Ormanlar.

A New Clinical Decision Support System for Early Diagnosis of Chronic Kidney Disease

Abstract

Chronic kidney disease is a worldwide health problem. It is possible to slow or stop the progression of this disease thanks to early diagnosis and treatment. Clinical decision support systems are health information technology systems designed to assist medical doctors in clinical decision making tasks. In this study, a new clinical decision support system is proposed for the early diagnosis of chronic kidney disease. Principal component analysis (PCA) and random forests (RF) techniques are used in the feature extraction and classification phases of the proposed system, respectively. The performance of the proposed system has been compared with classical machine learning algorithms and previous studies in the literature using six different performance metrics. The test results show that the proposed system is successful and can assist doctors in making decisions for early diagnosis of chronic kidney disease.

Keywords: Chronic Kidney Disease, Early Diagnosis, Clinical Decision Support System, Principal Component Analysis, Random Forests.

[*] Sorumlu Yazar: caneyupoglu@gmail.com, ceyupoglu@hho.edu.tr

1. Giriş

Kronik böbrek hastalığı, aylık veya yıllık dönemlerde böbrek fonksiyonlarındaki artan oranlı kayıplar olarak ifade edilebilir (World Kidney Day, 2020). Aynı zamanda kronik böbrek hastalığı, böbreğin yapısını ve işlevini etkileyen heterojen bozukluklar için kullanılan genel bir terimdir (Levey ve Coresh, 2012). Böbreklerin her birinde nefron adı verilen yaklaşık bir milyon küçük filtre vardır ve nefronlar hasar görürse, çalışmayı bırakırlar. Bir süre için sağlıklı nefronlar ekstra işi üstlenebilir. Ancak hasar devam ederse, giderek daha fazla nefron çalışmayı durdurur. Kalan nefronlar belli bir noktadan sonra kanı sağlıklı tutacak kadar iyi filtreleyemezler. Böbrek fonksiyonu belirli bir noktanın altına düştüğünde ise buna böbrek yetmezliği denir. Böbrek yetmezliği tüm vücudu etkiler ve kişinin kendini çok hasta hissetmesini sağlar. Tedavi edilmeyen böbrek yetmezliği hayatı tehdit edici olabilir (World Kidney Day, 2020).

Böbrek hastalığının en yaygın nedenleri yüksek tansiyon ve diyabettir. Yüksek tansiyon, tüm böbrek yetmezliği vakalarının dörtte birinden fazlasına neden olmaktadır. Diyabet ise tüm vakaların yaklaşık üçte birinin nedeni olarak belirlenmiştir ve çoğu gelişmiş ülkede böbrek yetmezliğinin en yaygın sebebidir. Daha az yaygın olan diğer durumlar arasında iltihaplanma veya enfeksiyonlar bulunur. Kronik böbrek hastalığı kalıtsal da olabilir. Bazen de genişlemiş prostat veya böbrek taşları nedeniyle idrar sisteminin uzun süreli tıkanıklığı sonucunda ortaya çıkabilir. Özellikle bazı ağrı kesici ilaçların uzun süre kullanılması da bu hastalığa sebep olabilir. Genellikle doktorlar soruna neyin yol açtığını belirleyemezler (World Kidney Day, 2020).

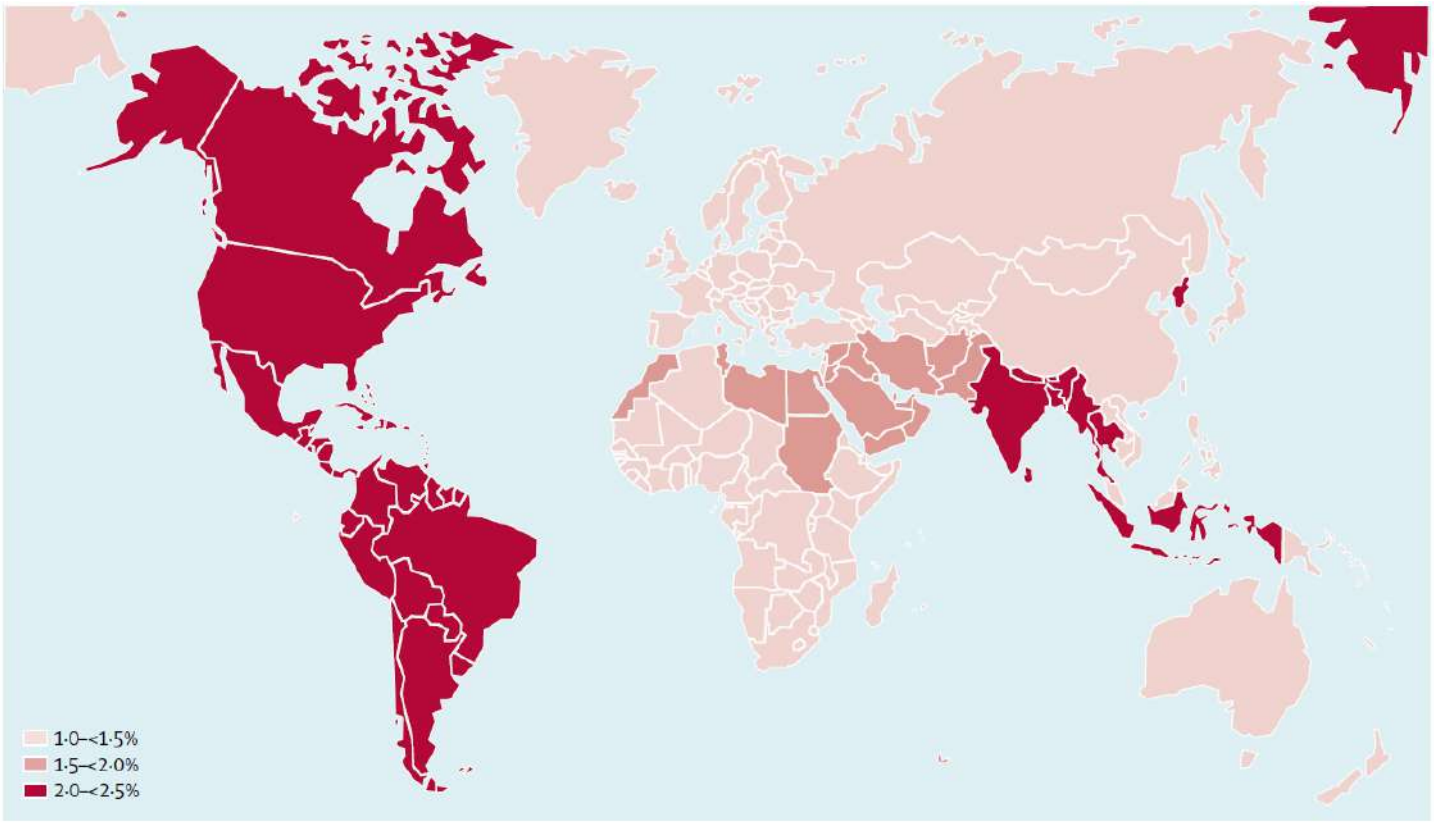
Dünya Sağlık Örgütü'nün küresel tahminlerine göre 2012 yılında kronik böbrek hastalığı sebebiyle 864.226 kişi hayatını kaybetmiştir. Bu sayı dünya çapındaki ölümlerin %1,5'ine karşılık gelmektedir. Önde gelen ölüm nedenleri listesinde 14'üncü sırada yer alan kronik böbrek hastalığı, her 100.000 kişide 12,2 kişinin ölümüne sebebiyet vermiştir. Bu hastalığın ölüm oranının 2030 yılına kadar 100.000'de 14 kişiye ulaşacağı tahmin edilmektedir (Webster vd., 2017; World Health Organization, 2020). Dünya çapında böbrek hastalığına dayanan toplam ölüm oranı ve kronik böbrek hastalığının görülme sıklığı Şekil 1'de gösterilmektedir.

Dünya nüfusunun yaklaşık %10'u kronik böbrek hastalığından etkilenmektedir ve her yıl milyonlarca kişi uygun tedaviye erişemedikleri için ölmektedir (National Kidney Foundation, 2020). Dünya çapında 2 milyondan fazla insan hayatta kalmak için diyaliz tedavisi görmektedir ya da böbrek nakli olmaktadır. Fakat bu sayı gerçekten yaşamak için tedaviye ihtiyaç duyan insanların sadece %10'unu temsil etmektedir. Böbrek yetmezliği sebebiyle tedavi gören 2 milyon kişinin çoğunluğu Amerika Birleşik Devletleri, İtalya, Almanya, Brezilya ve Japonya olmak üzere sadece beş ülkededir. Bu beş ülke dünya nüfusunun sadece %12'sini temsil etmektedir. Dünya nüfusunun yarısından fazlasını oluşturan yaklaşık 100 gelişmekte olan ülkede hastaların sadece %20'si tedavi edilmektedir (Couser vd., 2011). Böbrek yetmezliği için tedavi gören tüm hastaların %80'inden fazlası, sağlık hizmetlerine erişebilen ve büyük yaşlı nüfusuna sahip olan zengin ülkerlerde yaşamaktadır. Böbrek yetmezliği vakalarının, yaşlı insanların sayısının arttığı Çin ve Hindistan gibi gelişmekte olan ülkelerde orantısız bir şekilde artacağı tahmin edilmektedir (Jha vd., 2013). Orta gelirli ülkelerde diyaliz veya böbrek nakli tedavisi, ihtiyacı olan insanların çoğunluğu için büyük bir mali yük oluşturmaktadır. Diğer 112 ülkede ise birçok insan hiçbir tedavi görememektedir. Bu da tedavi edilmeyen böbrek yetmezliği sebebiyle yılda 1 milyondan fazla insanın ölümüne neden olmaktadır (Couser vd., 2011; National Kidney Foundation, 2020).

Kronik böbrek hastalığının erken evrelerinde olan bireylerin büyük çoğunluğunda bu hastalık teşhis edilmez. Böbrek hastalığı genellikle sessizce ilerler ve sıklıkla herhangi bir semptom yaratmadan önce böbrek fonksiyonlarının çoğunu tahrip eder. Azalan böbrek fonksiyonunun erken tanısı çok önemlidir. Çünkü erken tanı sayesinde böbrek hasarı veya bozulması diğer komplikasyonlarla kendini göstermeden önce uygun tedaviye başlanabilir (World Kidney Day, 2020). Bilgisayar bilimleri ve mühendisliği alanlarında kronik böbrek hastalığının erkan tanısı için makine öğrenmesi ve veri madenciliği tekniklerine dayanan bazı çalışmalar yapılmıştır. Xun vd. (2010), kronik böbrek hastalığı olan Çinli hastalardaki glomerüler filtrasyon hızını tahmin etmek için radyal tabanlı fonksiyon ağı (radial basis function network-RBFN) kullanmıştır. Çalışmada RBFN, 224'ü erkek ve 103'ü kadın olmak üzere toplam 327 hastaya ait kayıtlardan oluşan veri seti üzerinde test edilmiştir. Deneysel sonuçlar, RBFN'nin glomerüler filtrasyon hızının tahmin edilmesinde kullanılabileceğini göstermektedir.

Al-Hyari vd. (2013) tarafından yapılan çalışmada kronik böbrek yetmezliği olan hastaları teşhis etmek için sınıflandırma tabanlı bir sistem önerilmiştir. Veri seti olarak 11 ile 81 yaş aralığındaki 102 kişiden toplanan ve 15 farklı nitelikten oluşan kayıtlar kullanılmıştır. Sınıflandırı olarak ise karar ağacı (decision tree-DT), Naive Bayes (NB) ve yapay sinir ağı (artificial neural network-ANN) algoritmaları seçilmiştir. Test sonuçlarından DT'nin %92,2'lik sınıflandırma doğruluğu ile diğer iki algoritmadan daha iyi performansa sahip olduğu görülmüştür. Gupta vd. (2016), kronik böbrek hastalığı dahil olmak üzere 11 farklı kronik hastalık için hastalıklar ve ilgili tanı testleri arasındaki ilişkiyi makine öğrenmesi tekniklerini kullanarak analiz etmiştir. Çalışmanın sonucunda kronik böbrek hastalığının teşhisinde AdaBoost tekniğinden yararlanılarak %98,67'lik eğitim ve %88,66'lik test doğruluğu performansı elde edilmiştir.

Salekin ve Stankovic (2016) tarafından yapılan çalışmada kronik böbrek hastalığının tespitinde kullanılmak üzere makine öğrenmesi tekniklerine dayanan bir yaklaşım ileri sürülmüştür. Bu yaklaşımda ANN, *k*-en yakın komşu (*k*-Nearest Neighbours-*k*-NN) ve C4.5 karar ağacı teknikleri, 250'si kronik böbrek hastası ve 150'si sağlıklı olmak üzere toplam 400 kişinin kayıtlarından oluşan bir veri seti üzerinde test edilmiştir. C4.5 tekniği hastalığın tespitinde ANN ve *k*-NN tekniklerinden daha iyi performans göstermiştir. Son yıllarda yapılan bir çalışmada ise Ogunleye ve Wang (2018), Salekin ve Stankovic (2016) ile aynı veri setini kullanmış ve XGBoost (eXtreme Gradient Boosting) tekniğini temel alan bir otomatik teşhis sistemi önermiştir. Bu sistem %97,58'lik sınıflandırma doğruluğu ile kronik böbrek hastalığını teşhis etmiştir. Daha sonra yazarlar önceki çalışmalarında kullandıkları yöntemi geliştirmiş (Ogunleye ve Wang, 2019) ve yine aynı veri seti üzerinde %100'lük sınıflandırma başarısı elde etmiştir.



(a)



(b)

Şekil 1. Küresel Olarak Böbrek Hastalığı: (a) Böbrek Hastalığına Dayanan Toplam Ölüm Oranı, (b) Kronik Böbrek Hastalığının Görülme Sıklığı (Webster vd., 2017; World Health Organization, 2020)

Bu çalışmada ise daha önce yapılan çalışmalardan farklı olarak kronik böbrek hastalığının erken tanısında kullanılmak üzere temel bileşen analizi (principal component analysis-PCA) ve rastgele ormanlar (random forests-RF) tekniklerini temel alan yeni bir klinik karar destek sistemi önerilmiştir. Performans metriği olarak doğruluk, kesinlik, duyarlılık, F-ölçütü, MCC (Matthews correlation coefficient-Matthews korelasyon katsayısı) ve AUC (area under curve-eğri altında kalan alan) seçilmiştir. Önerilen sistemin performansı bu metrikler kullanılarak klasik makine öğrenmesi algoritmaları ve daha önce yapılan çalışmalar ile kıyaslanmıştır. Deneysel sonuçlar önerilen sistemin kronik böbrek hastalığını başarılı bir şekilde teşhis ettiğini göstermektedir.

Çalışmanın geri kalanı şu şekilde düzenlenmiştir: 2. bölümde önerilen sistemin üzerinde test edildiği veri seti ve önerilen sistem ayrıntılı olarak anlatılmaktadır. 3. bölümde önerilen sistemin performansının değerlendirilmesi için kullanılan metriklere ve bu metrikler kullanılarak yapılan performans karşılaştırmalarına yer verilmektedir. Son olarak 4. bölümde ise çalışmanın sonuçlarından bahsedilmektedir.

2. Materyal ve Metot

2.1. Veri Seti

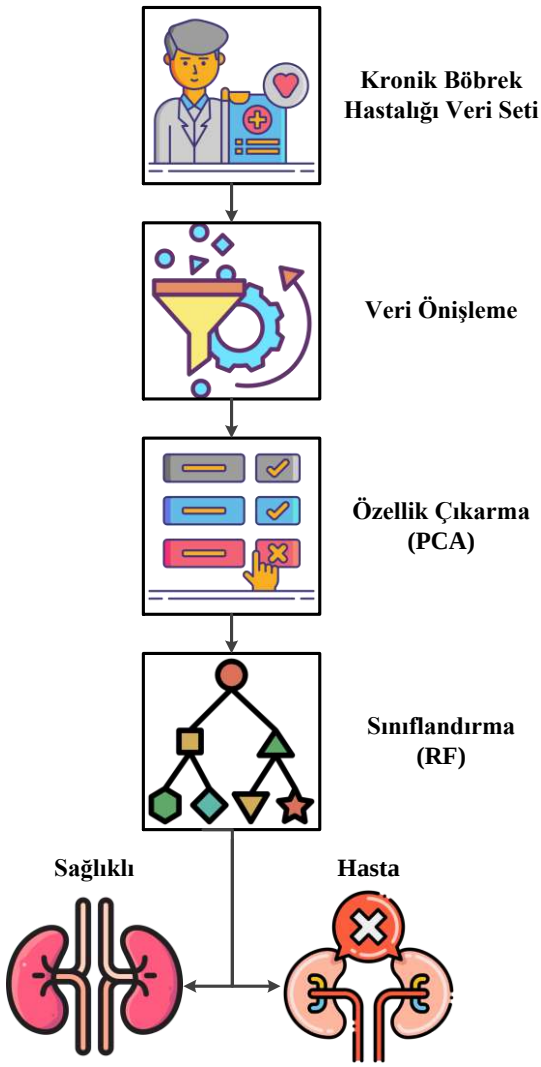
Bu çalışmada kronik böbrek hastalığı tanısı için Kaliforniya Üniversitesi – Irvine Makine Öğrenmesi Deposunda (University of California – Irvine Machine Learning Repository) bulunan Kronik Böbrek Hastalığı Veri Seti (Chronic Kidney Disease Data Set) (Soundarapandian vd., 2015) kullanılmıştır. Veri setinde toplam 400 kişiye ait kayıtlar vardır. Bu kişilerin 250'si kronik böbrek hastası, 150'si ise sağlıklıdır. Veri setinde biri sınıf niteliği (hasta veya sağlıklı) olmak üzere toplam 25 nitelik vardır. Tablo 1, veri setinde bulunan nitelikleri ve bu niteliklerin özelliklerini özetlemektedir.

2.2. Önerilen Klinik Karar Destek Sistemi

Bu çalışmada kronik böbrek hastalığının erken tanısı için önerilen klinik karar destek sistemi Şekil 2'de gösterilmektedir. Önerilen sistem; veri önişleme, özellik çıkarma ve sınıflandırma adımlarından oluşmaktadır. Tablo 1'de görüldüğü üzere önerilen sistemin test edileceği veri setinde eksik değer barındıran kayıtlar vardır. Veri setindeki 400 kayıt incelendiğinde 242 kayıtta eksik değer olduğu tespit edilmiştir. Eğer eksik değerlerin olduğu kayıtlar silinirse 43'ü hasta ve 115'i sağlıklı olmak üzere toplam 158 kayıt kalacaktır. Bu da kayıtların çoğunluğunun kaybedileceği ve sınıf etiketlerinin oranının dengesiz olacağı anlamına gelmektedir. Bu nedenlerden dolayı önerilen sistemin veri önişleme adımında kayıtları silmek yerine veri setindeki kayıtlarda yer alan eksik değerler yeni değerlerle değiştirilmiştir. Bu değiştirme işlemi, ilgili nümerik veya nominal nitelikteki değerlerin ortalaması alınarak yapılmıştır. Böylece veri seti, bütünlüğü bozulmadan daha iyi hale getirilmiştir.

Tablo 1. Kronik Böbrek Hastalığı Veri Setinin Özeti

Nitelik #	Nitelik Adı	Veri Türü	Değer Aralığı	Eksik Değer Sayısı
1	Yaş	Nümerik	2 – 90	9
2	Kan basıncı	Nümerik	50 – 180	12
3	Özgül ağırlık	Nominal	1,005, 1,010, 1,015, 1,020, 1,025	47
4	Albümin	Nominal	0, 1, 2, 3, 4, 5	46
5	Şeker	Nominal	0, 1, 2, 3, 4, 5	49
6	Kırmızı kan hücreleri	Nominal	Normal, Anormal	152
7	İltihap hücresi	Nominal	Normal, Anormal	65
8	İltihap hücre kümeleri	Nominal	Var, Yok	4
9	Bakteri	Nominal	Var, Yok	4
10	Rastgele kan şekeri	Nümerik	22 – 490	44
11	Kan üre	Nümerik	1,5 – 391	19
12	Serum kreatinin	Nümerik	0,4 – 76	17
13	Sodyum	Nümerik	4,5 – 163	87
14	Potasyum	Nümerik	2,5 – 47	88
15	Hemoglobin	Nümerik	3,1 – 17,8	52
16	Sıkıştırılmış hücre hacmi	Nümerik	9 – 54	71
17	Beyaz kan hücresi sayımı	Nümerik	2200 – 26400	106
18	Kırmızı kan hücresi sayımı	Nümerik	2,1 – 8	131
19	Hipertansiyon	Nominal	Evet, Hayır	2
20	Diyabet (şeker) hastalığı	Nominal	Evet, Hayır	2
21	Koroner arter (kalp) hastalığı	Nominal	Evet, Hayır	2
22	İştah	Nominal	İyi, Kötü	1
23	Ayak ödemi	Nominal	Evet, Hayır	1
24	Anemi	Nominal	Evet, Hayır	1



Şekil 2. Önerilen Klinik Karar Destek Sistemi

Önerilen sistemin özellik çıkarma adımında veri önişleme sonucunda elde edilen veriler üzerinde PCA ile özellik çıkarma işlemi yapılmıştır. PCA, özellik çıkarma ve boyut azaltma için yaygın olarak kullanılan bir tekniktir (Cottrell ve Munro, 1988; Jolliffe, 1986; Wang ve Paliwal, 2003). Bu teknik, sınıflar hakkındaki bilgilerin çoğunun, varyasyonların en büyük olduğu yönlerde yer aldığı varsayımına dayanmaktadır. PCA'nın en yaygın türevi, öngörülen uzaydaki varyansı en üst düzeye çıkaran standartlaştırılmış doğrusal izdüşüm anlamındadır (Hotelling, 1933; Wang ve Paliwal, 2003). Diğer bir ifadeyle PCA, verilerdeki örüntüleri tanımlamanın ve verileri benzerliklerini ve farklılıklarını vurgulayacak şekilde ifade etmenin bir yoludur. Grafiksel olarak temsil edilemeyen yüksek boyutlu verilerde örüntülerin bulunması zor olabilir. PCA tekniği bu gibi durumlarda verileri analiz etmek için güçlü bir araçtır (Smith, 2002). PCA ile ilgili daha ayrıntılı bilgi için ilgili kaynaklardan (Cottrell ve Munro, 1988; Hotelling, 1933; Jolliffe, 1986; Smith, 2002; Wang ve Paliwal, 2003) yararlanılabilir.

Önerilen sistemin sınıflandırma adımında ise RF kullanılmıştır. RF, Breiman (2001) tarafından ileri sürülen sınıflandırma ve regresyon problemleri için model bir araya getirme fikirlerine dayanan popüler ve çok etkili bir algoritmadır. RF, sınıflandırma hatasını en aza indiren eşleme olan Bayes sınıflandırıcısının veya regresyon fonksiyonunun tahmin edicilerini sağlayan bir model oluşturma stratejisidir. RF'nin prensibi, L öğrenme örneğinden gelen ve her bir düğümde X açıklayıcı değişkenlerin bir alt kümesini rastgele seçen çeşitli bootstrap örnekleri kullanılarak inşa edilen birçok ikili karar ağacını birleştirmektir (Genuer vd., 2010). RF tekniği, her bir ağaç bağımsız olarak örneklenen ve ormandaki tüm ağaçlar için aynı dağılıma sahip rastgele bir vektörün değerlerine bağlı olacak şekilde ağaç tahmin edicilerinin bir kombinasyonudur. RF tekniği tahminlemede etkili bir araçtır ve büyük sayılar yasası nedeniyle aşırı uyum göstermez. Doğru türde rastgelelik sağlaması sayesinde iyi bir sınıflandırıcı ve regresördür. Bireysel tahmin edicilerin gücü ve korelasyonları açısından çerçeve, rastgele ormanın tahmin etme yeteneği hakkında bir fikir vermektedir (Breiman, 2001). RF'nin temelleri için (Breiman, 2001) kaynağına başvurulabilir.

3. Araştırma Sonuçları ve Tartışma

3.1. Performans Metrikleri

Bu çalışmada performans metriği olarak doğruluk, kesinlik, duyarlılık, F-ölçütü, MCC ve AUC kullanılmıştır. Bu metriklerin formülleri aşağıdaki gibidir (Eyupoglu vd., 2018; Yavuz vd., 2017; Yavuz ve Eyüpoğlu, 2019):

$$Doğruluk = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$Kesinlik = \frac{TP}{TP + FP} \quad (2)$$

$$Duyarlılık = \frac{TP}{TP + FN} \quad (3)$$

$$F\text{–}ölçütü = \frac{2 \times Kesinlik \times Duyarlılık}{Kesinlik + Duyarlılık} \quad (4)$$

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP) \times (TP + FN) \times (TN + FP) \times (TN + FN)}} \quad (5)$$

$$AUC = \frac{1}{2}\left(\frac{TP}{TP + FN} + \frac{TN}{TN + FP}\right) \quad (6)$$

Bu formüllerde yer alan gerçek pozitif (true positive-*TP*), yanlış pozitif (false positive-*FP*), gerçek negatif (true negative-*TN*) ve yanlış negatif (false negative-*FN*) sayıları karışıklık matrisindeki sınıflandırma sonuçlarıdır (Sokolova ve Lapalme, 2009; Yavuz ve Eyupoglu, 2019, 2020).

3.2. Performans Sonuçları

Bu çalışmada önerilen klinik karar destek sistemi; Windows 10 Pro 64-bit işletim sistemi, Intel Core i7 8565U işlemci (1.80 GHz) ve 8 GB RAM'e sahip kişisel bir bilgisayar üzerinde çalışan Weka 3.8.4'te gerçeklenmiştir. Önerilen sistemin 10-kat çapraz geçerlemede doğruluk, kesinlik, duyarlılık, F-ölçütü, MCC ve AUC metrikleri kullanılarak elde edilen performans sonuçları ilk olarak klasik makine öğrenmesi algoritmaları ile karşılaştırılmıştır. Bu algoritmalar şunlardır: *k*-NN (Aha vd., 1991), destek vektör makinesi (support vector machine-SVM) (Keerthi vd., 2001), voted perceptron (VP) (Freund ve Schapire, 1999), RBFN (Frank, 2014), NB (John ve Langley, 1995), lojistik regresyon (logistic regression-LR) (Le Cessie ve Van Houwelingen, 1992), Bayesian lojistik regresyon (Bayesian logistic regression-BLR) (Genkin vd., 2007), OneR (Holte,

1993), K* (Cleary ve Trigg, 1995), lojistik model ağaçları (logistic model trees-LMT) (Landwehr vd., 2005), stokastik dereceli azalma (stochastic gradient descent-SGD) (Robbins ve Monro, 1951), C4.5 (Quinlan, 1993) ve AdaBoostM1 (Freund ve Schapire, 1996). Önerilen sistemin performans sonuçlarının klasik makine öğrenmesi algoritmaları ile karşılaştırılması Tablo 2'de gösterilmektedir. Önerilen sistemin doğruluk, kesinlik, duyarlılık, F-ölçütü, MCC ve AUC metrikleri için olan performans sonuçları sırasıyla %99,75, 1, 0,996, 0,998, 0,9947 ve 0,998'dir. Tablodan görüldüğü üzere önerilen sistemin performansı; doğruluk, duyarlılık, F-ölçütü, MCC ve AUC metriklerinde klasik makine öğrenmesi algoritmalarından daha iyidir. Kesinlik metriğindeki performans sonucu ise *k*-NN, SVM, NB, BLR, K*, LMT, SGD ve AdaBoostM1 algoritmaları ile aynıdır ve en iyi performans değeri olan 1'dir.

Önerilen sistemin performansı klasik makine öğrenmesi algoritmaları ile karşılaştırıldıktan sonra literatürde daha önce yapılan ve aynı veri setini kullanan çalışmalarla (Salekin ve Stankovic, 2016; Ogunleye ve Wang, 2018, 2019) kıyaslanmıştır. Tablo 3'te önerilen sistemin duyarlılık, F-ölçütü ve sınıflandırma doğruluğu metrikleri için olan performans sonuçları literatürdeki çalışmalar ile karşılaştırılmaktadır. Tablodan görüldüğü üzere önerilen sistemin 0,9980 olan F-ölçütü performansı Salekin ve Stankovic (2016)'in çalışmasından daha iyidir. Önerilen sistem, duyarlılık açısından Ogunleye ve Wang tarafından 2018 ve 2019 yıllarında yapılan çalışmalarla neredeyse aynı performansı göstermiştir. Sınıflandırma doğruluğu metriğindeki performansı ise Ogunleye ve Wang (2018) çalışmasından daha iyidir ve Ogunleye ve Wang (2018) çalışmasına çok yakındır.

4. Sonuç

Dünya çapında bir sağlık sorunu olan kronik böbrek hastalığının ilerlemesini yavaşlatmak ya da durdurmak için erken tanı ve tedavi çok önemlidir. Doktorlar, hastalıkların teşhisi için karar verirken klinik karar destek sistemlerinden yararlanırlar. Bu çalışmada kronik böbrek hastalığının erken tanısı için PCA ve RF tekniklerine dayanan yeni bir klinik karar destek sistemi ileri sürülmüştür. Önerilen sistemin performansını test ederken doğruluk, kesinlik, duyarlılık, F-ölçütü, MCC ve AUC metrikleri kullanılmıştır. Test sonuçları klasik makine öğrenmesi algoritmaları ve literatürde daha önce yapılan çalışmalar ile karşılaştırılmıştır. Çalışmanın sonuçlarından önerilen sistemin etkili olduğu ve doktorlar tarafından bu hastalığa erken tanı koyarken yardımcı bir araç olarak kullanılabileceği görülmektedir.

Tablo 2. Önerilen Sistemin Performansının Klasik Makine Öğrenmesi Algoritmaları ile Karşılaştırılması

Algoritma	**Doğruluk (%)**	**Kesinlik**	**Duyarlılık**	**F-ölçütü**	**MCC**	**AUC**
k-NN	98,75	1	0,9800	0,9899	0,9739	0,9900
SVM	97,00	1	0,9520	0,9754	0,9389	0,9760
VP	95,75	0,9957	0,9360	0,9649	0,9139	0,9647
RBFN	98,75	0,9960	0,9840	0,9899	0,9736	0,9887
NB	98,75	1	0,9800	0,9899	0,9739	0,9900
LR	96,25	0,9796	0,9600	0,9697	0,9209	0,9633
BLR	95,75	1	0,9320	0,9648	0,9149	0,9660
OneR	98,75	0,9880	0,9920	0,9900	0,9733	0,9860
K*	72,75	1	0,5640	0,7212	0,5715	0,7820
LMT	99,00	1	0,9840	0,9919	0,9790	0,9920
SGD	98,25	1	0,9720	0,9858	0,9637	0,9860
C4.5	99,00	0,9920	0,9920	0,9920	0,9787	0,9893
AdaBoostM1	99,25	1	0,9880	0,9940	0,9842	0,9940
Önerilen Sistem	99,75	1	0,9960	0,9980	0,9947	0,9980

Tablo 3. Önerilen Sistemin Performansının Literatürde Daha Önce Yapılan Çalışmalar ile Karşılaştırılması

Çalışmanın Yazarları	**Çalışmanın Yılı**	**Kullanılan Teknikler**	**Duyarlılık**	**F-ölçütü**	**Doğruluk (%)**
Salekin ve Stankovic	2016	ANN – *k*-NN – C4.5	-	0,9930	-
Ogunleye ve Wang	2018	XGBoost	1	-	97,58
Ogunleye ve Wang	2019	XGBoost	1	-	100
Önerilen Sistem	-	PCA – RF	0,9960	0,9980	99,75

Kaynakça

Aha, D. W., Kibler, D., & Albert, M. K. (1991). Instance-based learning algorithms. *Machine Learning*, *6*(1), 37-66.

Al-Hyari, A. Y., Al-Taee, A. M., & Al-Taee, M. A. (2013, December). Clinical decision support system for diagnosis and management of chronic renal failure. In *2013 IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies (AEECT)* (pp. 1-6). IEEE.

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5-32.

Cleary, J. G., & Trigg, L. E. (1995, July). K*: An instance-based learner using an entropic distance measure. In *12th International Conference on Machine Learning* (pp. 108-114).

Cottrell, G. W., & Munro, P. (1988, October). Principal components analysis of images via back propagation. In *Visual Communications and Image Processing'88: Third in a Series* (pp. 1070-1077). International Society for Optics and Photonics.

Couser, W. G., Remuzzi, G., Mendis, S., & Tonelli, M. (2011). The contribution of chronic kidney disease to the global burden of major noncommunicable diseases. *Kidney International*, *80*(12), 1258-1270.

Eyupoglu, C., Aydin, M. A., Zaim, A. H., & Sertbas, A. (2018). An efficient big data anonymization algorithm based on chaos and perturbation techniques. *Entropy*, *20*(5), 373.

Frank, E. (2014). Fully supervised training of Gaussian radial basis function networks in WEKA. Department of Computer Science, University of Waikato, Hamilton, New Zealand.

Freund, Y., & Schapire, R. E. (1996, July). Experiments with a new boosting algorithm. In *13th International Conference on Machine Learning* (pp. 148-156).

Freund, Y., & Schapire, R. E. (1999). Large margin classification using the perceptron algorithm. *Machine Learning*, *37*(3), 277-296.

Genkin, A., Lewis, D. D., & Madigan, D. (2007). Large-scale Bayesian logistic regression for text categorization. *Technometrics*, *49*(3), 291-304.

Genuer, R., Poggi, J. M., & Tuleau-Malot, C. (2010). Variable selection using random forests. *Pattern Recognition Letters*, *31*(14), 2225-2236.

Gupta, D., Khare, S., & Aggarwal, A. (2016, April). A method to predict diagnostic codes for chronic diseases using machine learning techniques. In *2016 International Conference on Computing, Communication and Automation (ICCCA)* (pp. 281-287). IEEE.

Holte, R. C. (1993). Very simple classification rules perform well on most commonly used datasets. *Machine Learning*, *11*(1), 63-90.

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, *24*(6), 417.

Jha, V., Garcia-Garcia, G., Iseki, K., Li, Z., Naicker, S., Plattner, B., Saran, R., Wang, A. Y. M., & Yang, C. W. (2013). Chronic kidney disease: global dimension and perspectives. *The Lancet*, *382*(9888), 260-272.

John, G. H., & Langley, P. (1995, August). Estimating continuous distributions in Bayesian classifiers. In *10th Conference on Uncertainty in Artificial Intelligence (UAI'95)* (pp. 338-345).

Jolliffe, I. T. (1986). Principal components in regression analysis. In *Principal component analysis* (pp. 129-155). Springer, New York, NY.

Keerthi, S. S., Shevade, S. K., Bhattacharyya, C., & Murthy, K. R. K. (2001). Improvements to Platt's SMO algorithm for SVM classifier design. *Neural Computation*, *13*(3), 637-649.

Landwehr, N., Hall, M., & Frank, E. (2005). Logistic model trees. *Machine Learning*, *59*(1-2), 161-205.

Le Cessie, S., & Van Houwelingen, J. C. (1992). Ridge estimators in logistic regression. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *41*(1), 191-201.

Levey, A. S., & Coresh, J. (2012). Chronic kidney disease. *The Lancet*, *379*(9811), 165-180.

National Kidney Foundation. (2020). *Global Facts: About Kidney Disease*. Retrieved from https://www.kidney.org/kidneydisease/global-facts-about-kidney-disease#

Ogunleye, A., & Wang, Q. G. (2018, June). Enhanced XGBoost-based automatic diagnosis system for chronic kidney disease. In *2018 IEEE 14th International Conference on Control and Automation (ICCA)* (pp. 805-810). IEEE.

Ogunleye, A., & Wang, Q. G. (2019). XGBoost model for chronic kidney disease diagnosis. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*.

Quinlan, J. R. (1993). *C4.5: Programs for Machine Learning*. Morgan Kaufmann Publishers, San Mateo, CA.

Robbins, H., & Monro, S. (1951). A stochastic approximation method. *The Annals of Mathematical Statistics*, *22*(3), 400-407.

Salekin, A., & Stankovic, J. (2016, October). Detection of chronic kidney disease and selecting important predictive attributes. In *2016 IEEE International Conference on Healthcare Informatics (ICHI)* (pp. 262-270). IEEE.

Smith, L. I. (2002). *A tutorial on principal components analysis*. Technical Report OUCS-2002-12, Department of Computer Science, University of Otago, New Zealand.

Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing & Management*, *45*(4), 427-437.

Soundarapandian, P., Jerlin Rubini, L. & Eswaran, P. (2015). *Chronic Kidney Disease Data Set* [Data file]. Available from https://archive.ics.uci.edu/ml/datasets/chronic_kidney_disease

Wang, X., & Paliwal, K. K. (2003). Feature extraction and dimensionality reduction algorithms and their applications in vowel recognition. *Pattern Recognition*, *36*(10), 2429-2439.

Webster, A. C., Nagler, E. V., Morton, R. L., & Masson, P. (2017). Chronic kidney disease. *The Lancet*, *389*(10075), 1238-1252.

World Health Organization. (2020). *Mortality and global health estimates: Causes of death; Projections for 2015–2030; Projection of death rates*. Retrieved from https://apps.who.int/gho/data/node.main

World Kidney Day. (2020). *Chronic Kidney Disease*. Retrieved from https://www.worldkidneyday.org/facts/chronic-kidney-disease/

Xun, L., Xiaoming, W., Ningshan, L., & Tanqi, L. (2010, October). Application of radial basis function neural network to estimate glomerular filtration rate in Chinese patients with chronic kidney disease. In *2010 International Conference on Computer Application and System Modeling (ICCASM 2010)* (pp. 332-335). IEEE.

Yavuz, E., & Eyupoglu, C. (2019). A cepstrum analysis-based classification method for hand movement surface EMG signals. *Medical & Biological Engineering & Computing*, *57*(10), 2179-2201.

Yavuz, E., & Eyupoglu, C. (2020). An effective approach for breast cancer diagnosis based on routine blood analysis features. *Medical & Biological Engineering & Computing*.

Yavuz, E., Eyupoglu, C., Sanver, U., & Yazici, R. (2017). An ensemble of neural networks for breast cancer diagnosis. In *2017 International Conference on Computer Science and Engineering (UBMK)* (pp. 538-543). IEEE.

Yavuz, E., & Eyüpoğlu, C. (2019). Meme Kanseri Teşhisi İçin Yeni Bir Skor Füzyon Yaklaşımı. *Düzce Üniversitesi Bilim ve Teknoloji Dergisi*, *7*(3), 1045-1060.