



RSESLIBKNN MAKİNE ÖĞRENMESİ YÖNTEMİ KULLANILARAK PARKİNSON HASTALIĞININ TANISI

İlknur BÜTÜNER¹ , Burak KAPLAN² , Kemal ADEM^{3,*} 

^{1,2,3} Aksaray Üniversitesi, Yönetim Bilişim Sistemleri Bölümü, 68100, Aksaray, Türkiye

ÖZET

Parkinson hastalığı, insanların yaşam kalitesini etkileyen nörolojik bir hastalıktır. Parkinson hastalığı merkezi sinir sistemini olumsuz etkileyen bir hastalıktır. Hastaların ölümüne yol açabilmektedir. Bu nedenle, Parkinson hastalığının erken tespiti son derece önemlidir. Parkinson hastalığına ait belirtiler, potansiyel olarak gelişmiş makine öğrenme tekniklerine dayanan bilgisayar destekli tanı sistemleri ile tespit edilebilir. Bu çalışmada Parkinson hastalığı tanısı için kNN, RseslibKnn ve A1DE makine öğrenmesi yöntemleri kullanılmıştır. Çalışmanın amacı Parkinson hastalığı veri kümesi üzerinde makine öğrenmesi yöntemlerinin başarı oranlarının karşılaştırılarak en uygun karar destek sisteminin sunulmasıdır. Veri kümesi olarak 'UC Irvine Machine Learning Repository' veri tabanından elde edilen, 252 örnekten ve 753 öznelikten oluşan veri kümesi kullanılmıştır. Literatür üzerinde farklı çalışmalar da incelenip karşılaştırılmıştır. Deneysel çalışmalar farklı çapraz geçerlilikler üzerinden yapılmış olup bunların ortalaması başarı sonucu olarak verilmiştir. Çalışma sonucunda, parkinson hastalığı veri kümesi kNN, RseslibKnn ve A1DE makine öğrenmesi yöntemleri ile sınıflandırılmış ve daha sonra eğitim ve test sonuçları doğruluk, duyarlılık ve özgüllük değerleri temel alınarak değerlendirilmiştir. Farklı çapraz geçerlilik değerleri ile ele alınan tüm yöntemler incelendiğinde en yüksek başarı sonucu %97,61 doğruluk oranı ortalaması ile RseslibKnn yöntemi vermiştir. Değerlendirme sonucunda RseslibKnn makine öğrenmesi yönteminin Parkinson hastalığının tespiti ile ilgili karar destek sistemleri üzerine önerilerde bulunulmuştur.

Anahtar kelimeler: Makine öğrenmesi, Parkinson hastalığı, RseslibKnn, A1DE, kNN

DIAGNOSIS OF PARKINSON'S DISEASE USING RSESLIBKNN MACHINE LEARNING METHOD

ABSTRACT

Parkinson's disease is a neurological disease that affects the quality of life of people. Parkinson's disease is a disease that negatively affects the central nervous system. It can lead to death of patients. Therefore, early detection of Parkinson's disease is extremely important. Symptoms of Parkinson's disease can be detected with computer-assisted diagnostic systems based on potentially advanced machine learning techniques. In this study, kNN, RseslibKnn and A1DE machine learning methods were used for the diagnosis of Parkinson's disease. The aim of the study was to compare the success rates of the machine learning methods on the Parkinson's disease dataset and to present the most appropriate decision support system. The data set was composed of 252 samples and 753 attributes from the 'UC Irvine Machine Learning Repository' database. Different studies on the literature are also examined and compared. Experimental studies were conducted on different cross-validations and the average of these studies was given as a result of success. At the end of the study, the Parkinson's disease data set was classified with kNN, RseslibKnn and A1DE machine learning methods and then the training and test results were evaluated based on accuracy, sensitivity and specificity values. When all the methods examined with different cross validity values were examined, RseslibKnn method gave the highest success result with an average of %97,61 accuracy rate. As a result of the evaluation, the recommendations of the RseslibKnn machine learning method on decision support systems for the detection of Parkinson's disease were made.

Keywords: Machine learning, Diagnosis of Parkinson's, RseslibKnn, A1DE, kNN

* Sorumlu yazar / Corresponding author, e-posta / e-mail: kemaladem@aksaray.edu.tr
Geliş / Received: 05.12.2019 Kabul / Accepted: 08.06.2020 doi: 10.28948/ngmuh.655720

1. GİRİŞ

Parkinson Hastalığı (PH), hastaların yaşam kalitelerini etkisi altına alan ve Alzheimer hastalığından sonra en yaygın nörolojik hastalıktan biridir [1]. Anatomik olarak omuriliği beyne bağlayan bir köprü olan “beyin sapı” bölgesinde bulunan, gri cevher çekirdeklerinin hasar görmesi sonrasında, bedensel hareketlerin kontrolünü sağlayan dopamin hormonu salgılayan hücrelerin yapılarının bozulup normal fonksiyonlarını yapamayacak hale gelmesi ile birlikte ortaya çıkar [2]. Yapısı bozulan hücrelerin neden hasara uğradığı henüz belirlenememiştir. Parkinson Hastalığına ait belirtiler dopamin salgılayan bu hücrelerin %60 ile %80’inin kaybolması sonucu ortaya çıkmaktadır [2]. Genellikle Parkinson hastalığı 50 ve 60’lı yaşlarda ortaya çıkmaya başlar. Yaklaşık 10-20 yıllık bir süreçte hastalıkta belirgin ilerlemeler görülmektedir. Daha öncesi bir süreçte hastalığın tanımlanması oldukça zordur [3]. PH tanısındaki temel zorluk, hastaların, dopamin hormonlarının beyindeki bir yerden bir yere taşınmasından sorumlu olan dopaminerjik nöronlarının yarısından fazlası hasara uğrayana kadar tipik motor semptomların görülmemesidir [4]. Parkinson hastalığının en çok görülen belirtilerinden birisi titreme ve sertliktir. Dinlenirken ellerin birden titremesi ve semptom sertlik olarak adlandırılan kaslardaki sertlik bu hastalığın belirtilerindedir. Hastalığın diğer belirtileri ise hareketlerin yavaşlığı, konuşma işlevinin bozukluğu, zayıf ses, değişken konuşma hızı ve kısa konuşma ile ilgili olan problemlerdir. Hastalığın erken teşhisi hastanın yaşam kalitesini artırır. Vücudun dopamin seviyesini arttırmak için takviye olarak kullanılan levodopa ilacı hastalığın ilk evrelerinde kullanılır. Hastalık ilerlediğinde ilacın yan etkileri de artmaktadır. İlaç artık etkisini kaybettiğinde hastalığın tedavisi için cerrahi bir müdahale gerekmektedir [5]. Parkinson hastalığının sayısının artmasıyla hastalığın teşhis edilmesindeki bu zorlu süreç teknolojinin gelişmesiyle ve yeni makine öğrenmesi yöntemlerinin kullanılmasıyla daha kolay hale gelmiştir. Makine öğrenmesi yöntemleri ile Parkinson hastalığının tespiti üzerine yapılan çalışmalar literatürde karşımıza çıkmaktadır.

Literatür taramasında Parkinson hastalığının teşhisi ve sınıflandırılması ile ilgili makine öğrenmesi yöntemleri kullanan 4 farklı çalışma incelenmiştir. Bu çalışmaların ilkinde ‘UC Irvine Machine Learning Repository’ (UCI)’den alınan 195 adet ses kaydından oluşan veri seti üzerinde Bayes Net, Naive Bayes, Multilayer Perceptron, SMO, J48 makine öğrenmesi yöntemleri kullanılarak Weka yazılım platformu aracılığı ile analizler yapılmıştır. Yapılan bu analizler sonucunda Parkinson Hastalığının teşhisinde elde edilen doğruluk, duyarlılık ve özgüllük değerleri incelenerek karşılaştırılmıştır. Doğruluk, duyarlılık ve özgüllük ile en başarılı sonucu elde ettiği görülmektedir [5]. İkinci çalışmada Matlab yazılım platformunda yapılmış olup Cvpartition fonksiyonu kullanılarak verilerin %80 i eğitim ve %20 si test için kullanılmıştır. Naive Bayes, kNN, Diskriminant Analiz, Lojistik Regresyon Analizi ve SVM makine öğrenmesi yöntemleri kullanılarak analizler yapılmıştır. Parkinson Hastalığının teşhisinde doğruluk, duyarlılık ve özgüllük değerleri göz önüne alınarak yapılan analizin sonucunda kNN makine öğrenmesi tekniğinin %96,97 doğruluk ile en başarılı sonucu elde ettiği görülmektedir [6].

Parkinson hastalığının erken ortaya çıkan semptomları arasında titreme, sertlik ve vokal bozukluğu (disfoni) bulunur. Sonuç olarak, konuşma göstergeleri, disfonik bulgulara dayanan Parkinson hastalığının tanımlanmasında önemlidir. Disfoni semptomlarına dayalı Parkinson hastalığı tanısı için makine öğrenimi temelli tekniklerin performansının değerlendirildiği Matlab programı üzerinde yapılan “Disfoni önlemlerine göre Parkinson hastalığının teşhisinde makine öğrenme yöntemlerinin performansı” adlı üçüncü çalışma incelenmiş olup, 22 adet ses bozukluğu ölçümü ile gerçekleştirilmiştir. Çalışmada SVM, k-NN, RBFNN, NB, LDA, RT ve MDC makine öğrenmesi algoritmaları kullanılmıştır. En başarılı sonucu %92 doğruluk oranı ile SVM makine öğrenmesi tekniğinin vermiş olduğu görülmektedir [7].

Sağlıklı kişiler ve Parkinson hastalığı hastaları arasındaki el hareketi ve kas koordinasyonundaki farklılıkları üzerinde yapılan “Parkinson hastalığında basit çizim hareketlerinin makine öğrenimine dayalı sınıflandırılması” adlı dördüncü çalışma incelenmiştir. Veri setinde 20 sağlıklı birey 24 Parkinson hastası ele alınmıştır. Çalışma weka üzerinde yapılmış olup, Naive Bayes, AdaBoost (J48), Log. Regression, J48, SVM ve Random Forest makine öğrenmesi yöntemleri kullanılmıştır. Naive Bayes makine öğrenmesi yönteminin %90,90 doğruluk oranı ile en yüksek değeri verdiği görülmektedir [8].

Bu çalışmada Parkinson hastalığı üzerine ‘UC Irvine Machine Learning Repository’ (UCI)’den alınan 252 kayıt ve 753 öznitelige sahip olan veri kümesi üzerinde kNN, RseslibKnn ve A1DE makine öğrenmesi yöntemleri kullanılarak Parkinson hastalığının teşhisinde en doğru sonucun elde edilmesi amaçlanmıştır. Literatür araştırmasında yapılan çalışmalarda göz önüne alındığında çalışmamızda kullanılan RseslibKnn makine öğrenmesi yöntemi %97,61 başarı oranı ile en yüksek sonucu vermektedir. İlgili veri kümesinde RseslibKnn ve A1DE makine öğrenmesi yöntemleri ilk defa uygulanmıştır. En başarılı sonucu veren makine öğrenmesi tekniğinin karar destek sistemi olarak kullanılması önerilmektedir.

2. MATERYAL VE METOT

Bu çalışmada hastalıklı ve sağlıklı bireylerin toplamından oluşan 252 kayıt ve 753 öznitelige sahip ‘UC Irvine Machine Learning Repository’ (UCI) veri tabanından alınan veri kümesi kullanılmıştır [9]. Veri kümesi ile ilgili sayısal bilgiler Tablo 1’de verilmektedir.

RSESLIBKNN MAKİNE ÖĞRENMESİ YÖNTEMİ KULLANILARAK PARKİNSON HASTALIĞININ TANISI

Tablo 1. Çalışmaya katılan bireyler ile ilgili sayısal bilgiler

Hasta Bireyler		Sağlıklı Bireyler		Toplam Kayıt
Kadın	81	Kadın	41	252
Erkek	107	Erkek	23	
Toplam	188	Toplam	64	
Hasta Bireyler Yaş Grubu		Sağlıklı Bireyler Yaş Grubu		
33 – 87		41 - 82		

Tablo 1’de görüldüğü gibi, hasta bireylerin sayısı sağlıklı bireylere göre daha fazladır. Bu durum çalışmada kullanılan yöntemlerin duyarlılık değerlerinin özgüllük değerlerine göre daha yüksek çıkabilmesi sonucunu ortaya koymaktadır. Çalışmada belirtilen 753 özneliğin makine öğrenmesi modellerine giriş olarak verilmesi ile sınıflandırma işlemi gerçekleştirilerek kişinin Parkinson hastası olup olmadığı hakkında karar verilmektedir.

2.1. kNN (k En Yakın Komşuluk)

kNN sınıflandırmasında, en yakın komşulara göre sınıflandırma yapılır. En yakın komşular için k değeri değişebilir. k, bilinmeyen bir olayın nasıl sınıflandırılacağına karar vermek için; kaç değer, komşu olarak değerlendirileceğini belirler. Bilinmeyen bir örnek verildiğinde, en yakın bir komşu sınıflandırıcı, bilinmeyen numuneye en yakın olan k eğitim numuneleri için desen uzayını arar. En yakın komşu hesaplanmasında Öklid, Minkowski ve Manhattan uzaklık ölçütleri kullanılmaktadır [10]. Bilinmeyen örnek, en yakın komşuları arasında en yaygın sınıfa atanır. k = 1 olduğunda, bilinmeyen örneğe desen alanında kendisine en yakın olan eğitim örneğinin sınıfı atanır. En yakın komşu sınıflandırıcıya sahip bir test örneğini sınıflandırmak için geçen süre, sınıflandırıcıda tutulan eğitim örneği sayısı ile doğrusal olarak artar. Büyük bir depolama gereksinimi vardır. Ayrıca, farklı özellikler farklı kapsamlar için sonucu etkilediğinde de kötü performans gösterir. kNN sınıflandırma algoritmasının performansını etkileyebilecek parametre kullanılacak en yakın komşu sayısıdır. Varsayılan olarak en yakın bir komşu kullanılmaktadır [11].

2.2. RseslibKnn

RseslibKnn makine öğrenmesi yöntemi, parametrik olmayan tembel sınıflandırma ve tahminleme yöntemidir. Veri büyüklüğü, optimal k değerini öğrenme ve belirleme, yetenekleri kontrol etme, en uygun çekirdek fonksiyonunu belirleme, yoğunluk temelli metrik ve bunun gibi bazı özellikleri tanımlar. Sınıflandırma işlemlerini gerçekleştirirken temel kNN yönteminden farklı olarak en yakın komşu sayısını ve mesafe ölçütlerini otomatik olarak bulabilen bir algoritma yapısına sahiptir. Sınıflandırıcı, iki arama budama kriterine sahip bir metrik ağacı kullanarak en yakın komşu aramasını uygular. Uygulanan arama yöntemi sınıflandırıcının çok büyük veri kümeleri için çalışmasını sağlar [12], [13].

Algoritma önce mesafelerin ağırlıklı toplamına dayalı bir mesafe ölçüsünü hesaplar ve daha sonra bir indeklemeye ağacı oluşturur. Ağırlıklar üç farklı yöntem kullanılarak hesaplanabilir: mesafe tabanlı yöntem, doğruluk tabanlı yöntem ve algılayıcı tabanlı yöntem. Sınıflandırma, sette en yakın komşuları bulmak ve karar için oy vermek suretiyle yapılır. RseslibKnn, görüntü sınıflandırmasında kullanılan son teknoloji tekniklerin bir temsili, ek olarak geliştirilen kNN modelidir [14].

2.3. A1de

Ortalama bir bağımlılık tahmincisi olan A1DE (Averaged N Dependence Estimators) artarak öğrenmeyi destekleyen bir sınıflandırma tekniğidir. Naive Bayes’in bütün küçük alanlarının ortalamasını hesaplayarak Naive Bayes’ten daha başarılı sınıflandırma performansı sağlar. Sabit yapıya sahip basit modelleri tanımlar. Performansların iyileştirilmesi yüksek hesaplama maliyetlerine neden olmaktadır. Bu algoritma öğrenmede hassas bir şekilde sınıflandırma yaparken hesaplama açısından da daha verimlidir [15]. Üst düzey olasılıkları daha fazla genellemek için A1DE kullanmak mümkündür.

$$p(Y = y | X = x) = \sum_{s \in S^n} p(Y = y, x_s) \prod_{i=1}^D p(X_i = x_i | Y_i = y_i, x_s) \binom{D}{n} \quad (1)$$

Eşitlik 1’de S^n kümenin n büyüklüğündeki tüm alt kümelerini gösterir $\{1, \dots, D\}$. A1DE algoritması tüm sınıflandırıcıların tahminlerini bir araya getirerek tahminin elde edildiği bağımlılık sınıflandırıcıları grubunu öğrenerek çalışır. n, bağımlılık tahmin edicisi, bir öznelilik olma olasılığı, sınıf değişkeni ve en çok n diğer öznelilik koşulunu anlatır. A1DE algoritmasında, n

İ. Bütüner, B. Kaplan, K. Adem

özniteliklerinin diğer tüm özniteliklere ebeveyn olarak ayarlandığı n özelliklerinin her birleşimi için bir n-bağımlılık sınıflandırıcısı oluşturulur.

Bu çalışmada, AnDE algoritması ailesinden A1DE algoritması kullanılmaktadır. Zaman karmaşıklıklarının yüksek olması nedeniyle, $n > 2$ olduğunda A1DE algoritması nadiren kullanılmaktadır. A1DE için hem eğitim hem de test aşamaları için alan karmaşıklığı $O(|Y| \binom{D}{n+1} v^{n+1})$ değerlerine eşittir; burada n ebeveynlerin sayısıdır (sınıf hariç). Eğitim aşaması için zaman karmaşıklığı $O(N \binom{D}{n+1})$ ve tek bir örneği sınıflandırmak için test aşaması ise $O(|Y|D \binom{D}{n})$ 'dir [16].

3. BULGULAR VE TARTIŞMA

Bu çalışmadaki makine öğrenmesi yöntemleri WEKA yazılım platformu üzerinde çalıştırılmıştır. Analizler için Intel Core i7-4510U işlemciye sahip, 8 GB DDR3 belleğe sahip olan bir bilgisayar kullanılmıştır. Makine öğrenmesi yöntemlerinde kullanılan parametreler deneysel çalışmalar sonucunda belirlenmiş ve RseslibKnn yönteminde $k=1$ ve Öklid uzaklık ölçütü, kNN yönteminde $kNN = 1$, batch_size = 100, A1DE yönteminde Frekans Limiti = 1, Ondalık Basamak Numarası = 2 olarak tercih edilmiştir. Sonuçların objektif olarak değerlendirilmesi için veri kümesinde Karar Ağacı, Rasgele Orman, SVM gibi temel makine öğrenmesi yöntemleri kullanılarak analiz edilmiştir. Tüm analizler sonunda içinden en yüksek başarı oranını veren 15 farklı makine öğrenmesi yöntemi kullanılarak 10 fold değeri üzerinde deneysel çalışmalar yapılmış olup sonuçlar Tablo 2'de verilmiştir. Makine öğrenmesi yöntemlerinin değerlendirilmesinde doğruluk, duyarlılık ve özgüllük değerleri incelenmiştir.

Tablo 2. Kullanılan makine öğrenmesi yöntemlerinin 10 fold'da Duyarlılık, Özgüllük ve Doğruluk analizi değerleri

Sınıflandırma	Çapraz Geçerlilik	Duyarlılık	Özgüllük	Doğruluk
SMO	10	0,93	0,64	%85,44
SIMPLE LOGISTIC	10	0,93	0,64	%85,31
kNN	10	0,97	0,89	%95,23
MLP CLASSIFIER	10	0,88	0,69	%83,46
LDA	10	0,78	0,69	%75,52
RANDOM COMMITTEE	10	0,95	0,69	%88,49
SYS FOR	10	0,95	0,65	%87,69
FT	10	0,89	0,67	%83,59
J48 CONSOLIDATED	10	0,84	0,70	%80,82
SPEGASOS	10	0,86	0,66	%80,82
RSESLIBKNN	10	0,99	0,96	%98,28
A1DE	10	0,82	0,78	%81,08
DI4JMLP CLASSIFIER	10	0,82	0,70	%78,70
CLASSIFICATION VIA REGRESSION	10	0,91	0,64	%83,99
BAYES NET	10	0,78	0,71	%76,19

Tablo 2'de 15 farklı makine öğrenmesi yöntemleri ile elde edilen değerler detaylı bir şekilde verilmiştir. Tablo 2'de görüldüğü gibi en başarılı makine öğrenmesi yöntemleri kNN, RseslibKnn ve A1DE'dir. Sonuçların objektif olarak değerlendirilmesi için bu yöntemler ayrıca 5 ve 7 fold değerleri üzerinde de analiz edilmiştir. 3 farklı çapraz geçerlilik ile elde edilen başarı oranlarının ortalaması Tablo 3'te görüldüğü gibi modellerin başarı oranı olarak sunulmuştur.

RSESLIBKNN MAKİNE ÖĞRENMESİ YÖNTEMİ KULLANILARAK PARKİNSON HASTALIĞININ TANISI

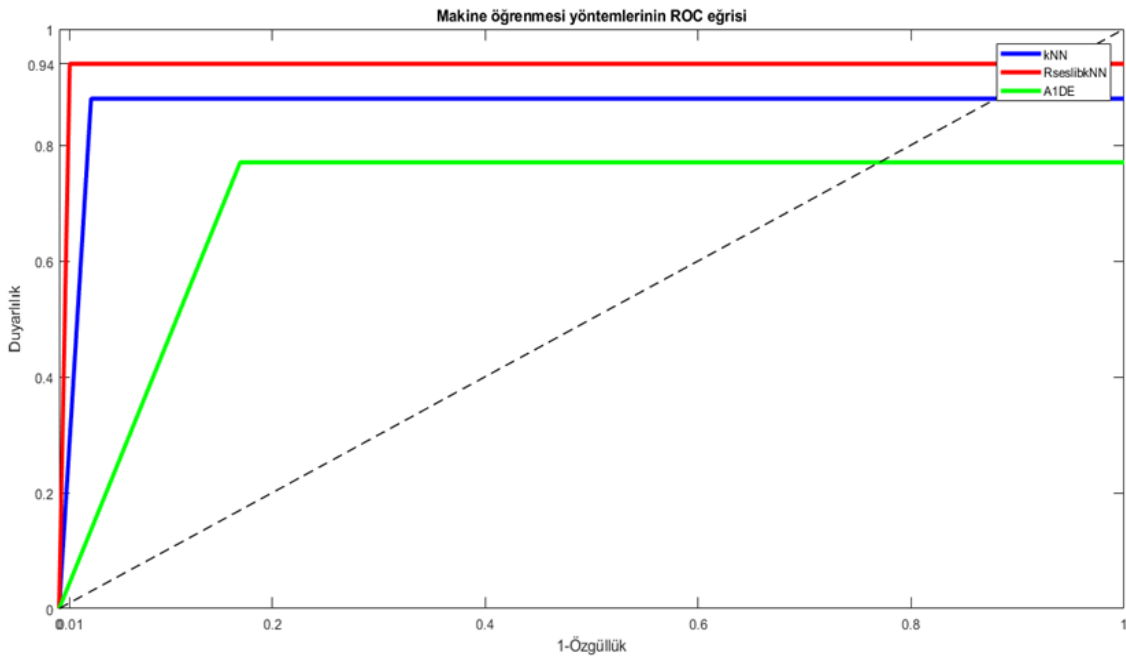
Tablo 3. En başarılı makine öğrenmesi yöntemlerinin farklı foldlarda Duyarlılık, Özgüllük ve Doğruluk analizi değerleri

Sınıflandırma Yöntemi	Çapraz Geçerlilik	Duyarlılık	Özgüllük	Doğruluk	Kappa	F-Ölçüm
kNN	5	0,97	0,89	%94,57	0,8561	0,946
	7	0,97	0,87	%94,31	0,8481	0,943
	10	0,97	0,89	%95,23	0,873	0,952
	Ortalama	0,97	0,88	%94,70	0,8590	0,947
RseslibKnn	5	0,98	0,95	%97,35	0,9302	0,974
	7	0,99	0,92	%97,22	0,9255	0,972
	10	0,99	0,96	%98,28	0,9544	0,983
	Ortalama	0,99	0,94	%97,61	0,9367	0,976
A1DE	5	0,83	0,74	%80,55	0,5267	0,812
	7	0,83	0,78	%81,74	0,5578	0,824
	10	0,82	0,78	%81,08	0,5453	0,818
	Ortalama	0,83	0,77	%81,12	0,5432	0,818

Tablo 3'te üç farklı yöntem ve üç ayrı fold değerinden ortaya çıkan veriler aşağıdaki gibidir;

- 5 fold çapraz geçerlilik sonuçlarına göre kNN, RseslibKnn ve A1DE yöntemlerini değerlendirdiğimizde, en yüksek doğruluk oranını veren yöntem %97,35 ile RseslibKnn, ardından %94,57 ile kNN ve son sırada ise %80,55 ile A1DE'dir.
- 7 fold çapraz geçerlilik sonuçlarına göre kNN, RseslibKnn ve A1DE yöntemlerini değerlendirdiğimizde, en yüksek doğruluk oranını veren yöntem %97,22 ile RseslibKnn, ardından %94,31 ile kNN ve son sırada ise %81,74 ile A1DE'dir.
- 10 fold çapraz geçerlilik sonuçlarına göre kNN, RseslibKnn ve A1DE yöntemlerini değerlendirdiğimizde, en yüksek doğruluk oranını veren yöntem %98,28 ile RseslibKnn, ardından %95,23 ile kNN ve son sırada ise %81,08 ile A1DE'dir. Tüm bu veriler doğrultusunda 10 fold tüm yöntemler için en yüksek doğruluk değerlerini sağlamaktadır.

En başarılı üç makine öğrenmesi yönteminin uygulanması sonucu elde edilen ROC eğrisi Şekil 1'deki gibidir.

**Şekil 1.** Makine öğrenmesi yöntemlerinin uygulanması sonucu elde edilen ROC eğrisi

Şekil 1’deki ROC eğrisi incelendiğinde her bir yöntem için uygulanan 5, 7 ve 10 fold ile elde edilen değerlendirme ölçütlerinin ortalaması temel alınarak karşılaştırma yapılmıştır. Karşılaştırma sonucunda en yüksek başarı oranı %97,61 doğruluk değeri ile RseslibKnn yöntemi, ardından %94,70 doğruluk değeri ile kNN yöntemi ve son olarak %81,12 doğruluk değeri ile A1DE yöntemi vermektedir. Önerdiğimiz çalışmanın literatürdeki çalışmalarla karşılaştırması Tablo 4’te verilmiştir.

Tablo 4. Literatürde aynı konulardaki farklı çalışmalara ait veri setlerinin karşılaştırılması

Yazarlar	Örnek Sayısı	Öznitelik Sayısı	Kullanılan Yöntemler	Sayısal Parametreler			
				Duyarlılık	Özgüllük	Doğruluk	F-Ölçüm
Badem, 2019 [17]	252	753	kNN	0,91	0,92	%91,83	84,24
Karacı, 2019 [18]	240	44	Topluluk Sınıflandırıcı	-	0,84	%82,00	82
Sakar ve ark., 2013 [19]	1040	26	SVM	0,85	0,90	%85,00	-
Bizal, 2014 [6]	195	22	kNN	-	-	%96,97	97,06
Bu çalışma	252	753	kNN	0,97	0,88	%94,70	94,7
			RseslibKnn	0,99	0,94	%97,61	97,6
			A1DE	0,83	0,77	%81,12	81,8

Tablo 4’te görüldüğü gibi, çalışmada kullanılan RseslibKnn yönteminin literatürde yapılan çalışmalara göre daha başarılı sonuçlar verdiği görülmektedir. Ayrıca, çalışmada kullanılan üç farklı yöntemin eğitim aşamasının zaman performansları ele alındığında, en hızlı sonucu veren yöntem 0.3 saniye ile kNN, ardından 1.71 saniye ile A1DE ve son olarak da 34.84 saniye ile RseslibKnn’dır.

4. SONUÇ

Parkinson hastalığı, dopamin hormonu salgılayan hücrelerin yapılarının bozulup normal fonksiyonlarını yapamayacak hale gelmesi ile birlikte ortaya çıkar. Hastaların yaşam kalitelerini etkisi altına alır ve Alzheimer hastalığından sonra ikinci en çok yaygın olarak görülen nörolojik bir hastalıktır. Yaygın olarak görülmesi nedeniyle bu çalışmada Parkinson hastalığı veri kümesi üzerinde makine öğrenmesi yöntemleri uygulanarak başarı oranları karşılaştırılmış ve en başarılı yöntem karar destek sistemi olarak sunulmuştur. Çalışmada Parkinson hastalığı veri kümesine makine öğrenmesi yöntemi olarak kNN, RseslibKnn ve A1DE yöntemleri uygulanmıştır. RseslibKnn makine öğrenmesi yönteminin ilgili veri kümesi üzerinde ilk defa uygulanması ile %97,61 doğruluk oranı elde edilerek literatürün üzerinde bir başarı oranı yakalanmıştır. Zaman performansı bakımından ise kNN ve A1DE yönteminin RseslibKnn’e göre daha iyi olduğu görülmüştür. RseslibKnn algoritmasının zaman performansının kötü olması eğitim aşamasında çok fazla örnek üzerinde karar vermesinden kaynaklanmaktadır. Algoritmanın test aşamasında sadece bir örnek üzerinde karar vermesi nedeni ile bu durum göz ardı edilebilir. Bu nedenle çalışmada önerilen model sağlık karar destek sistemlerinde alternatif bir yöntem olarak kullanılabilir. Bu çalışmada öznitelik seçme algoritmaları kullanılmamış, yalnızca sınıflandırma algoritmaları denenmiştir. Gelecek çalışmalarda veri kümesine özellik seçme algoritmasının uygulanması ile birlikte farklı makine öğrenmesi yöntemlerinin kullanılması ile başarı oranının artırılması düşünülmektedir. Bu işlem ancak veri setinin güncellenmesi ve hastalık için önemli özniteliklerin seçilmesi ile yapılabilir. Daha önemli öznitelikler seçilerek farklı algoritmaların denenmesi ile başarı oranı yükseltilebilir.

KAYNAKLAR

- [1] Ö. Eskidere, “A Comparison of Feature Selection Methods for Diagnosis of Parkinson's Disease From Vocal Measurements”, *Journal of Engineering and Natural Sciences*, vol. 30, pp. 402-403, 2012.
- [2] E. Akbayır, M. Şen, U. Ay, S. Şenyer, E. Tüzün ve C. İ. Küçükakal, “Parkinson Hastalığının Etiyopatogenezi”, *Deneyisel Tıp Dergisi*, vol. 7, no. 13, pp. 3, 2017.
- [3] F. Özyiğit, S. C. Kabay ve Ö. Arık, “Rational Drug Use And Parkinson Disease”, *Journal Of Contemporary Medicine*, vol. 6, no. 2, pp. 105, 2016.
- [4] E. Taşdemir, “Parkinson Hastalığında Nöroinflamasyonun Rolü”, *SDÜ Tıp Fakültesi Dergisi*, vol. 26, no. 3, pp. 348, 2019.

- [5] D. Narmadha, D. Marudhadevi ve B. Santhi, “Parkinson’s Disease Detection Using Soft Computing Techniques”, *World Applied Sciences Journal*, vol. 29, no. 16, pp. 29-89, 2014.
- [6] Ö. Bizal, “Makine Öğrenmesi Teknikleriyle Parkinson Hastalığının Belirlenmesi”, 2014.
- [7] S. Lahmiri, D. A. Dawson ve A. Shmuel, “Performance of machine learning methods in diagnosing Parkinson’s disease based on dysphonia measures”, *Biomedical Engineering Letters*, vol. 8, no. 1, pp. 29-39, 2018.
- [8] C. Kotsavasiloglou, N. Kostikis, D. Hristu-Varsakelis ve M. Arnaoutoglouc, “Machine learning-based classification of simple drawing movements in Parkinson’s disease”, *Biomedical Signal Processing and Control*, vol. 31, pp. 174-180, 2017.
- [9] C. O. Sakar, G. Serbes, A. Gunduz, H. C. Tunc, H. Nizam, B. E. Sakar, M. Tutuncu, T. Aydin, M. E. Isenkul ve H. Apaydin “A comparative analysis of speech signal processing algorithms for Parkinson’s disease classification and the use of the tunable Q-factor wavelet transform”, *Applied Soft Computing*, vol. 74, pp. 74, 2019.
- [10] S. A. Dudani, “The distance-weighted k-nearest-neighbor rule”, *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 6, no. 4, pp. 325-327, 1976.
- [11] V. Chaurasia ve S. Pal, “A Novel Approach for Breast Cancer Detection using Data Mining Techniques”, *International Journal of Innovative Research in Computer and Communication Engineering*, vol. 2, no. 1, 2017.
- [12] A. Wojna, R. Latkowski ve L. Kowalski, *RSESLIB User Guide*. pp. 45-58, 2019.
- [13] A. Wojna, *Analogy-Based Reasoning In Classifier Construction*, Warsaw: Banacha, 2005.
- [14] R. Jankovi, “Machine Learning Models for Cultural Heritage Image Classification: Comparison Based on Attribute Selection”, *Information*, vol. 11, no. 12, pp. 5-6 2019.
- [15] S. O. M. Kamel, H. M. Harb, N. Hegazi, A. S. Tag El Dein ve H. M. Abd El Kader, “A Novel Ensemble Learner Algorithm for Anomaly Based Detection in Intrusion Prevention System”, *International Journal of Engineering and Technical Research (IJETR)*, vol. 3, no. 12, pp. 24, 2015.
- [16] S. Picek, A. Heuser ve S. Guilley, “Template attack versus Bayes classifier”, *Journal of Cryptographic Engineering*, vol. 7, no. 4, pp. 345-346, 2017.
- [17] H. Badem, “Parkinson Hastalığının Ses Sinyalleri Üzerinden Makine Öğrenmesi Teknikleri İle Tanımlanması”, *Ömer Halisdemir Üniversitesi Mühendislik Bilimleri Dergisi*, vol. 8, no. 2, pp.630-637, 2019.
- [18] A. Karacı, “Detection Of Parkinson's Disease With Hybrid Ensemble Classifier”, 2nd International Turkish World Engineering and Science Congress, Türkiye, 2019.
- [19] B. E. Sakar, M.E. İsenkul, C.O. Sakar, A. Sertbaş, F. Gürgen, S. Delil, H. Apaydin ve O. Kurşun, “Collection and Analysis of a Parkinson Speech Dataset With Multiple Types Of Sound Recordings”, *IEEE Journal Of Biomedical and Health Informatics*, vol. 17, no. 4, pp. 828-834, 2013.

