



Temel Bileşen Analizi Yöntemleri Kullanarak Parkinson Hastalığının Otomatik Teşhisi

Şule Yücelbaş^{1*}, Cüneyt Yücelbaş²

¹ Hakkari Üniversitesi, Mühendislik Fakültesi, Elektrik-Elektronik Mühendisliği Bölümü, Hakkari, Türkiye (ORCID: 0000-0002-6758-8502)

² Hakkari Üniversitesi, Mühendislik Fakültesi, Elektrik-Elektronik Mühendisliği Bölümü, Hakkari, Türkiye (ORCID: 0000-0002-4005-6557)

(İlk Geliş Tarihi 21 Mayıs 2019 ve Kabul Tarihi 1 Haziran 2019)

(DOI: 10.31590/ejosat.568544)

ATIF/REFERENCE: Yücelbaş, Ş., & Yücelbaş, C. (2019). Temel Bileşen Analizi Yöntemleri Kullanarak Parkinson Hastalığının Otomatik Teşhisi. *Avrupa Bilim ve Teknoloji Dergisi*, (16), 294-300.

Öz

Parkinson rahatsızlığı çok yavaş ilerleyen sinsi bir beyin hastalığıdır. Bu hastalığın teşhis yöntemleri arasında, kişilere ait seslerin analizi de bulunmaktadır. Ses analizi ile Parkinson'un erken teşhisi, kullanılan çeşitli yöntemler sayesinde mümkün olmaktadır. Bu çalışma kapsamında 188 Parkinson hastası ve 64 sağlıklı kişiye ait kaydedilmiş ses sinyallerine Ayarlanabilir Q-faktör Dalgacık Dönüşümü (AQDD) metodu uygulanması sonucunda elde edilen özellikler kullanılmıştır. AQDD özelliklerine, boyut indirgeme yöntemlerinden temel bileşen analizi (TBA) ve bunun çeşitlerinden olan kernel TBA (KTBA) ile olasılıksal TBA (OTBA) uygulanmıştır. Daha sonra boyutları indirgenen yeni veri gruplarına ayrı ayrı k-kat çapraz doğrulama yöntemi uygulanarak eğitim-test verileri elde edilmiştir. Sonraki aşamada ise, boyut indirgeme yöntemlerinin etkinliğinin araştırılması için veriler Rastgele Orman (RO) algoritması ile ayrı ayrı sınıflandırılmış ve elde edilen sonuçlar ayrıca istatistiksel ölçütlerle yorumlanmıştır. Sınıflandırma sonuçları açısından boyut azaltma yöntemleri içerisinde en başarılısı %87.56 doğruluk oranı ile OTBA olmuştur. Ayrıca bu yöntem sonucunda ROC ve PRC alan değerleri yaklaşık 0.95 bandına ulaşarak hasta ve sağlıklı sınıf ayrışımının mükemmel yaklaştığını kanıtlamıştır. Gerçek yaşam uygulamalarına uygun olan bu çalışmanın performans sonuçları, aynı verinin kullanıldığı literatürdeki tek çalışma ile kıyaslanmış ve bu çalışmada diğer çalışmaya nazaran daha yüksek istatistiksel oranların elde edildiği görülmüştür. Ayrıca verilerin kaydedildiği kişi sayısının literatürdeki diğer çalışmalara göre yüksek oluşu, çalışmanın bu alandaki önemini arttırmaktadır.

Anahtar Kelimeler: Parkinson hastalığı, boyut indirgeme yöntemleri, AQDD, TBA.

Automatic Diagnosis of Parkinson's Disease Using Principal Component Analysis Methods

Abstract

Parkinson's disease is a sneaky brain disorder that progresses very slowly. The diagnostic methods of this disease include the analysis of individual voices. The earliest detection of Parkinson's with voice analysis is made possible by various methods. In this study, the results obtained from the application of Tunable Q-factor Wavelet Transformation (TQWT) method to the recorded audio signals of 188 Parkinson's patients and 64 healthy individuals were used. The principal component analysis (PCA) and its types (kernel PCA (KPCA) and Probabilistic PCA (PPCA)) which are dimension reduction methods have been applied to the TQWT features. Afterwards, k-fold cross validation method was applied to the new data groups and the training-test data were obtained. In the next step, the data were separately classified by random forest (RF) algorithm to investigate the effectiveness of the dimension reduction methods and the results were also interpreted by statistical criteria. In terms of classification results, OTBA was the most successful in size reduction methods with 87.56% accuracy rate. In addition, as a result of this method, ROC and PRC area values reached a band

* Sorumlu Yazar: Hakkari Üniversitesi, Mühendislik Fakültesi, Elektrik-Elektronik Mühendisliği Bölümü, Hakkari, Türkiye, ORCID: 0000-0002-6758-8502, suleyucelbas@hakkari.edu.tr

of about 0.95, proving that patient and healthy class decomposition approached perfection. The performance results of this study, which is suitable for real-life applications, were compared with the single study in the literature in which the same data was used, and this study showed that higher statistical ratios were obtained in comparison to the other study. Moreover, the high number of people in whom the data was recorded compared to other studies in the literature increases the importance of the study in this field.

Keywords: Parkinson's disease, dimension reduction methods, TQWT, PCA

1. Giriş

Parkinson hastalığı (PH), beyinde dopamin üretmekle görevli hücrelerin eksikliğinde ortaya çıkan bir rahatsızlıktır [1, 2]. Beynin ilgili bölgesinde bulunan hücrelerin zarar görmesi sonucunda dopamin yeteri kadar üretilmez. Bundan dolayı kişinin hareketleri yavaşlar, dinlenme anında titremeler görülür ve kaslarda sertlik meydana gelir [1, 2]. Bu hastalık, 40 yaşını geçmiş erkeklerde daha sık görülür. Bayanlarda görülme oranı erkeklere göre daha azdır. PH kronik bir hareket bozukluğu olup belirtiler zamanla kendini gösterir. Beyinde üretilen dopamin miktarı azaldıkça bu rahatsızlık da ilerleme kat eder ve hastalık belirtileri daha yoğun olarak görülür. Ancak ortaya çıkan belirtiler kişiye göre değişiklik gösterebilir [1, 2]. PH'nin birincil motor belirtileri vücutta titremeler, hareket duraklanması, vücudun çeşitli organlarında sertlik ve dengesizliktir. Bazı hastalarda ise belirti olarak sadece tek bir elin titremesi görülür. Hastalığın başlangıç aşamasında çoğunlukla vücudun tek tarafında belirtiler başlar ve zamanla diğer semptomlar da kendini göstermeye başlar [1-5]. Ayrıca kişinin yüz mimiklerinde azalma, suratta donuk bir ifade ve uyku bozuklukları da görülebilmektedir. Bu durum kişilerde ses ve telaffuz bozukluklarına da yol açmaktadır. Bu hastalığın kimlerde yavaş veya hızlı ilerleyeceğini önceden tespit edebilmek şu an için imkân dâhilinde değildir [1-5].

PH'nin teşhisinde belirli bir test yoktur. Alanında uzman bir nörolog hastanın tıbbi geçmişini, klinik bulgularını ve fiziki muayenesini analiz ederek PH'yi teşhis edebilmektedir [5, 6]. Bu hastalığın tedavisinde asıl amaç kişinin aktif, bağımsız ve kendi ihtiyaçlarını kendisinin yapabilir duruma gelmesini sağlamaktır. Günümüzde kesin bir tedavisi olmamasına rağmen, kişilere verilen sınırlı medikal destek ile belirtilerin kontrol altına alınması hedeflenmektedir [5, 6]. Bu kişilere ayrıca akıllı egzersiz uygulamaları, denge hareketleri ve yaşam tarzı değişiklikleri faydalı olabilmektedir. Konuşmasında bozukluklar meydana gelmiş hastaların konuşma ve dil uzmanları eşliğinde konuşma egzersizleri yapmaları faydalı olabilmektedir. Bunun yanında cerrahi tedavi de uygulanabilmekte, fakat genellikle ilk tercih edilen yol olmamaktadır [5-7]. Ancak hastalığın düzeltilmediği ve medikal destek ile yapılan rehabilitasyon işlemlerine rağmen bir sonuç elde edilemediği durumlarda bu yöntemle başvurulabilmektedir [7].

Tıbbi literatürde PH'nin teşhis ve tedavisi için birçok yayın mevcuttur. Mühendislik alanında ise bu hastalığın uzman sistemler tarafından otomatik teşhisine odaklanılmıştır. Bu amaçla, çoğunlukla kişilerden alınan ses veya yürüme kayıtları kullanılmıştır. Daha önceden de belirtildiği gibi Parkinson hastalarının büyük çoğunluğunda (yaklaşık %90) hastalığın erken evrelerinde görülen ve en önemli belirtilerden biri olan telaffuz bozuklukları ortaya çıkmaktadır [8, 9]. Bu nedenle, PH'nin otomatik olarak teşhisinde ses ile ilgili özellikler uzman sistemlerde yaygın olarak kullanılmıştır. Ses sinyallerini kullanarak gerçekleştirilen yayınlardaki asıl amaç, ilgili ses özelliklerini kullanarak hastaların sağlıklı kişilerden otomatik olarak ayırt edilmesidir [9-13].

Mühendislik alanında yapılan otomatik Parkinson hastalığının tespiti için tasarlanan sistemler araştırmacılar tarafından tediagnosis olarak tanımlanmıştır. Tediagnosis sisteminin tasarlandığı çalışmalardan biri Sakar ve Kursun (2010) tarafından sunulmuş ve sistem dâhilinde sınıflandırıcı olarak DVM kullanılarak sonuçlar değerlendirilmiştir [10]. Bu alanda aynı amaç doğrultusunda yapılan bir diğer çalışmada ise, 4 farklı sınıflandırıcının (Sinir ağları, DM sinir ağları, regresyon ve karar ağaçları) sistem üzerindeki etkisi incelenmiştir [14]. Benzer bir çalışmada da paralel ileri beslemeli sinir ağları mimarisi üzerinde odaklanılarak çeşitli ağ yapılarının otomatik PH tespiti üzerindeki etkisi araştırılmıştır [15]. Sakar ve arkadaşları, hasta ve sağlıklı deneklerin ses kayıtlarından bazı özellikleri çıkardıktan sonra sınıflandırıcılar aracılığı ile elde ettikleri sonuçları analiz etmişlerdir [16]. Aynı amaç doğrultusunda sunulan benzer bir çalışmada da sadece sınıflandırıcı olarak bulanık k-en yakın komşu algoritmasından yararlanılmıştır [17]. Bu çalışmaya çok benzer başka bir çalışmada ise farklı olarak sadece parçacık sürü optimizasyon temelli bulanık k-en yakın komşu sistemi kullanılarak ortalama %97.47 doğruluk oranı elde edilmiştir [18]. Son yıllarda yapılan çalışmalara bakıldığında; PH teşhisinde çıkarılan vokal özelliklerin sistem üzerindeki etkinliği makine öğrenme teknikleri ile analiz edilmiş ve en iyi sonuç DVM'de %96.4 olarak elde edilmiştir [19]. Bir başka çalışmada çeşitli sınıflayıcı sistemleri (lineer diskriminant analizi, k-en yakın komşu, naive bayes, regresyon ağaçları, radyal temelli fonksiyon sinir ağları, DVM, Mahalanobis uzaklığı) denenmiş ve en iyi sonuca %92 ile DVM'de ulaşılmıştır [20]. Bu alanda literatürde gerçekleştirilmiş çalışmaların ortak noktaları özetle; veri kaydı, veri işleme, özellik çıkarma, özellik seçme ve sınıflandırma olarak görülmektedir. Amaç olarak ise, genellikle PH'nin yüksek doğrulukla otomatik sınıflandırılmasına odaklanılmıştır. Literatür incelendiğinde kullanılan özellikler üzerinde boyut değiştirme ve indirgeme yöntemlerinin yeteri kadar araştırılmadığı, bunların sistemler üzerindeki etkilerinin incelenmediği görülmüştür.

Bu çalışmada İstanbul Üniversitesi Cerrahpaşa Tıp Fakültesi Nöroloji Bölümünde kaydedilmiş 188 Parkinson hastası ve 64 sağlıklı kişiye ait ses sinyallerine Ayarlanabilir Q-faktör Dalgacık Dönüşümü (AQDD) metodu uygulanması sonucunda elde edilen özellikler kullanılmıştır [21]. Bu özelliklere boyut indirgeme yöntemlerinden temel bileşen analizi (TBA) ve bunun çeşitlerinden olan kernel TBA (KTBA) ile olasılıksal TBA (OTBA) uygulanmıştır. Daha sonra bu veriler k-kat çapraz doğrulama yöntemi aracılığıyla eğitim ve test grupları elde edilmiştir. Çalışma kapsamında, boyut indirgeme yöntemlerinin etkisinin araştırılması için elde edilen veriler Rastgele Orman (RO) sistemi ile sınıflandırılmış ve sonuçlar doğru pozitif oranı veya recall (DPO), yanlış pozitif oranı (YPO), kesinlik veya hassasiyet (Precision), F skor (F), ROC alanı, PRC (Precision-Recall) alanı ve ortalama sınıflandırma doğruluğu (OSD) gibi ölçütlerle değerlendirilmiştir. Sınıflandırma sonuçları açısından boyut azaltma yöntemleri içerisinde en başarılısı %88 doğruluk oranı ile OTBA olmuştur. Ayrıca bu yöntem sonucunda ROC ve PRC alan değerleri yaklaşık 0.95 ile hasta ve sağlıklı sınıf ayrışımının mükemmelle yaklaşımını kanıtlamıştır. Gerçek yaşam uygulamalarına uygun olan bu çalışmanın performans sonuçları, aynı verinin kullanıldığı literatürdeki tek çalışma [21] ile kıyaslanmış ve bu çalışmada diğer çalışmaya nazaran daha yüksek istatistiksel oranların elde edildiği görülmüştür.

2. Materyal ve Yöntem

Parkinson hastalığının bilinen sebebi beynin, kasları kontrol eden iletişim mekanizmasının yavaş bir şekilde bozulmasıdır. Konuşmayı sağlayan yüz kasları da bu durumdan etkilendiği için hastaların telaffuzları çoğunlukla daha yumuşak ve tekdüze hale gelir. Buna bağlı olarak kişiler bazı sesleri çıkarmakta ve söylemekte zorlanırlar [22]. Hastalığın bu belirtisinden dolayı çalışma kapsamında İstanbul Üniversitesi Cerrahpaşa Tıp Fakültesi Nöroloji Bölümünde yaşları 33 ile 87 arasından değişen 188' i Parkinson hastası ve 64' ü sağlıklı olmak üzere toplam 252 kişiye /a/ harfi üçer kez tekrarlatılarak kayıt altına alınan veriler kullanılmıştır [21].

Parkinson'un tespitinde kişilerin kayıt altına alınmış ses sinyallerine birçok sinyal işleme ve özellik çıkarma yöntemi uygulanabilir. Bu yöntemlerden biri de ayarlanabilir AQDD dir [23]. Verilerin alındığı referansta [21] en iyi sonuçlar AQDD yöntemi ile elde edildiğinden bu çalışmada da bu yöntemin uygulanması neticesinde elde edilen özellik grubu kullanılmıştır. AQDD yönteminin uygulanması sonucunda her bir veriden 36 alt-bant elde edilmiş ve her banttan da 12 farklı özellik (enerji, shannon entropi, log enerji entropi, Teager-Kaiser enerji operatörü (TKEO) ortalaması, TKEO standart sapması, medyan, ortalama, standart sapma, minimum, maksimum, skewness ve kurtosis) çıkarılmıştır [21, 24-31]. Böylelikle her data kendi içinde $36 \times 12 = 432$ boyuttan oluşmuştur. Ayrıca her deneğe /a/ harfi üçer kez tekrarlatılarak kayıt altına alındığından $252 \times 3 = 756$ adet veri üzerinden çalışma yürütülmüştür. Buna göre toplam veri 756x432 olarak elde edilmiştir. Daha sonra 432 boyutlu bir veri, uzman sistemin çalışmasını ağırlaştıracağından bu veri seti üzerinde TBA, KTBA ve OTBA boyut indirgeme yöntemleri uygulanmıştır. Bunların sonucunda elde edilen yeni boyutlu datalardan eğitim-test grupları oluşturmak amacıyla 10-kat çapraz doğrulama yöntemi kullanılmıştır. Son olarak eğitim-test veri setleri RO algoritmasına girdi olarak verilerek uzman sistem sonuçlandırılmıştır.

2.1. TBA, KTBA ve OTBA Boyut İndirgeme Yöntemleri

Bu çalışma kapsamında kullanılan ilk boyut azaltma yöntemi TBA olmuştur. Bu yöntem, yeni değerler oluşturmak suretiyle verinin boyut sayısını hesaplamak ve azaltmak için başvurulan etkili bir veri analiz metodudur [32, 33]. Standart bir TBA algoritmasında üzerinde çalışılacak verilerin gerekli işlemler sonucunda elde edilmesi ilk adım olarak kabul edilir. Sonraki aşamada, veriye ait ortalama değer her elemandan çıkarılarak merkezleme işlemi gerçekleştirilir. Daha sonraki aşamada kovaryans matrisi, iki veya daha fazla rastgele değişken kümesi arasındaki korelasyonun gücünün bir ölçüsünü sağlamak için hesaplanır. Dördüncü adımda ise, kovaryans matrisinin öz vektörleri ve öz değerleri hesaplanır. Beşinci aşamada bileşenler seçilerek bir özellik vektörü oluşturulur. Burada en yüksek öz değere sahip öz vektör, veri kümesinin temel bileşeni olarak belirlenir. Son aşamada ise, boyutu değiştirilmiş yeni veri seti elde edilmiş olur [33].

Uygulaması gerçekleştirilen diğer TBA çeşidi ise KTBA yöntemidir. Bu metod, standart TBA'nın bir uzantısı olup doğrusal olmayan veri dağılımları için kullanılmaktadır [34-36]. Kısaca KTBA, kernel tanımlı bir özellik uzayında standart TBA'nın bir uygulama çeşididir. Bu TBA yönteminde; U_k özellik uzayında ilk öz vektörler tarafından kapsanan alt-uzayı ifade etsin. Bu alt-uzay içerisinde yeni verinin k-boyutlu vektör yansıması aşağıdaki gibi hesaplanabilir [37];

$$P_{U_k}(\phi(X)) = \left(\sum_{i=1}^l \alpha_i^j k(x_i, x) \right)_{j=1}^k \quad (1)$$

burada;

$$\alpha^j = \lambda_j^{-1/2} v_j \quad (2)$$

denklemi ile kernel matrisinin ilgili öz vektör ve öz değerleri cinsinden verilmiştir. Kısaca denklem (1) KTBA'nın temelini oluşturmaktadır [37].

Çalışma kapsamında kullanılan yöntemlerin sonucusu ise OTBA'dır. Uzman sistemlerin başarısı çıkarılan özelliklerin seçimine ve hatalı verilerin temizlenmesine bağlıdır. Ancak veri setleri içerisinde sistem performansını etkileyen hatalı-eksik verilerin tespit edilmesi zor ve zahmetli bir işlemdir. Bu tür problemlerde uygulanabilecek TBA yöntemi, OTBA olmuştur. Bu metod sayesinde en yüksek kovaryans değerindeki yansıtılmış vektörler çıkarılır. Bu vektörler daha sonra özellik matrisinin boyutunu azaltmak için kullanılır [38]. OTBA yoluyla oluşturulabilecek giriş verileri şu şekilde formüle edilebilir [39];

$$y_s = \tau x_s' + \mu + \xi(s) : s \in \{1, 2, \dots, M\} \quad (3)$$

burada y_s , $R \times 1$ boyutuna sahip gizli rastgele değişkenlerin bir vektörünü temsil eder. y_s 'nin, bağımsız olduğu ve aynı şekilde sıfır ortalama ve 1 varyansına sahip olduğu varsayılır. Ayrıca, τ Giriş gözlemleri ve gizli uzay arasındaki eşleşmeyi gösteren bir $C \times R$ parametre matrisidir. Kısaca, doğrusal bir haritalama türüdür. μ , her değişkenin ortalamasına sahip olan bir $C \times 1$ parametre vektörüdür. $\xi(s)$ $C \times 1$ boyutuna sahip bir vektördür ve rastgele hataları temsil eder. Varyansı ve ortalaması sıfır olduğu, bağımsız ve aynı şekilde normal dağılmış olduğu varsayılmıştır [38].

2.2. k-Kat Çapraz Doğrulama Yöntemi

Bu yöntem eğitim ve test veri ayırma işlemi için kullanılan genel bir tekniktir. Bu teknikte ilk olarak k sayısı belirlenir. Belirlenen k sayısına göre veri seti k-tane alt kümeye ayrılır. Alt kümelerden bir tanesi test, diğer k-1 tanesi ise eğitim için ayrılarak sistem çalıştırılır. Daha sonra bir başka alt küme test için ayrılırken diğerleri eğitim için ayrılır ve sistem bir kez daha çalıştırılır. Bu şekilde her alt küme test ve eğitim verisi olarak kullanılmış olur. Bu yöntemde amaç hazırlanan veri setindeki her verinin test ve eğitim

işlemine tabi tutulmasıdır. Tüm aşamalar bittikten sonra her adımda elde edilen başarı ve hata oranlarının ortalaması hesaplanır ve genel bir başarı oranı elde edilir. Bu çalışmada 10-kat çapraz doğrulama yöntemi kullanılmıştır.

2.3. Sınıflandırma İşlemi ve Performans Ölçütleri

Uzman sistemler günümüzde birçok alanda gerçek uzman gibi davranan, düşünen, tepki verip sonuca ulaşan yapay zekâ temelli algoritmalar bütünüdür. Bu sistemlerden beklenen, hızlı ve yüksek doğrulukta sonuç üretmeleridir. Çünkü uzman sistemlerin oluşturulması ve kullanılmasının en büyük sebebi, kullanıldığı alandaki bir eksikliği ya da kararsızlığı gidermektir. Sağlık alanında başvuru uzman sistemler, hasta ve sağlıklı deneklere ait ilgili verilerin analizinin yapılması sonucunda bunları birbirinden hızlı ve yüksek doğrulukta ayırt edebilmektedir. Bu işlemler uzman sistem bünyesinde bulunan sınıflandırıcı algoritmalar aracılığıyla gerçekleştirilmektedir.

Bu amaç ile gerçekleştirilen çalışma içerisinde Parkinson hastalığının otomatik teşhisi için birçok sınıflandırıcı (yapay sinir ağları, sıralı minimal optimizasyon temelli destek vektör makinaları, basit lojistik, J48 ve RO) denenmiş ancak en iyi performans sonuçları RO [40] algoritması ile elde edilmiştir. Kullanılan bu sınıflayıcı sistemin çalıştırılması ve ulaşılan performans sonuçlarının değerlendirilmesi Weka 3.8 arayüz yazılımı [41] bünyesinde gerçekleştirilmiştir. Oluşturulan uzman sistem kullanılmadan önce, 10-kat çapraz doğrulama yöntemi ile eğitim-test verileri oluşturulmuştur. Ayrıca veri setleri içerisinde hasta olanlar 0, sağlıklı olanlar ise 1 olarak etiketlenerek sistem çalıştırılmıştır. En iyi sonuçların elde edildiği RO algoritmasının sistem parametreleri ise varsayılan olarak (BagSizePercent:100, BatchSize:100, MaxDepth:0, NumDecimalPlaces:2, NumExecutionSlots:1, NumFeatures:0, NumIterations:100 ve Seed:1) seçilmiştir. Kullanılan performans ölçütleri oluşturulan uzman sistem tarafından doğru ve yanlış olarak tespit edilen veri sayılarına göre oluşturulmuş konfüzyon (karışıklık) matrisine göre yapılmıştır. Konfüzyon matrisi temelde hedef değerler ile sistem tarafından elde edilen değerlere bakarak bir sınıflayıcı algoritmanın performans sonuçlarını değerlendirmek için bu alanda en çok başvuru bir araçtır. Bu matristeki veriler kullanılarak DPO, YPO, Precision, F, ROC, PRC ve OSD ölçütleri elde edilmiştir [42-46].

3. Araştırma Sonuçları

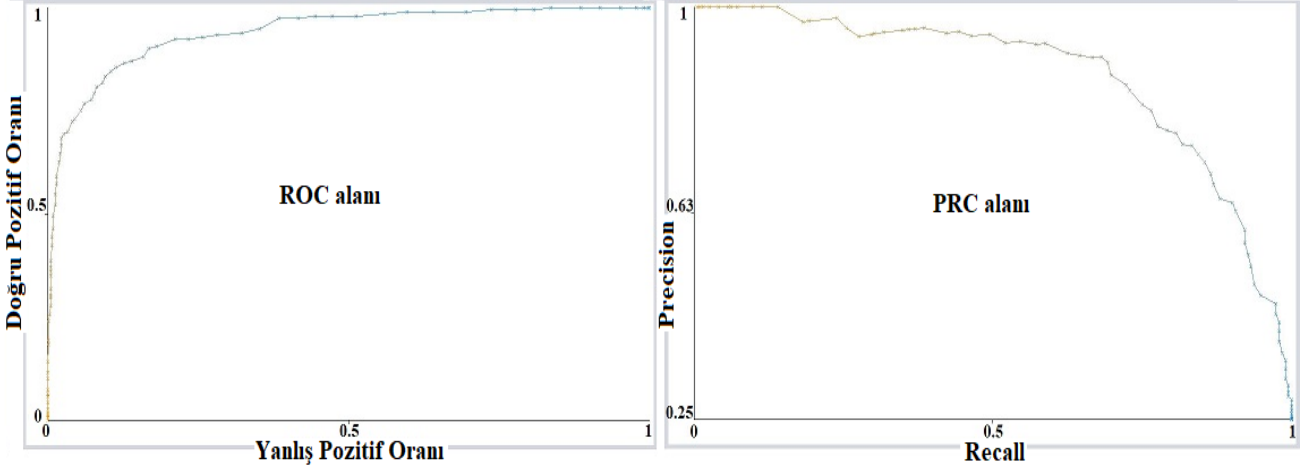
Bu çalışmada bir uzman eşliğinde önceden veri kayıtları yapılmış 188 Parkinson hastası ve 64 sağlıklı olmak üzere toplamda 252 kişi denek olarak ele alınmıştır [21]. Bu deneklere ait kaydedilmiş ses verilerinden çıkarılmış özellik grupları arasından en iyi sonuçların elde edildiği AQDD yöntemine ait özellik grubu [21] çalışma kapsamında kullanılmıştır. Bu özellik grubu üzerinde gerekli düzenlemeler yapılarak 756x432 boyutunda yeni bir veri seti oluşturulmuştur. Sonraki aşamada ise, TBA ve bunun 2 farklı yöntemi 10, 15, 20, 25, 30, 35 ve 40 boyutları için çalıştırılmış ve her yöntem için farklı boyutlarda yeni veri grupları elde edilmiştir. Daha sonra bu yeni boyutlu veri gruplarına tek tek 10-kat çapraz doğrulama yöntemi uygulanarak eğitim-test verileri oluşturulmuştur. Sistemin son aşamasında ise bu veriler RO sınıflandırıcısına girdi olarak verilerek konfüzyon matrisi elde edilmiştir. Bu matristeki sonuç verilerinin değerlendirilmesi neticesinde ulaşılan performans değerlendirme kıstaslarına ait sistem sonuçları ayrıntılı olarak aşağıdaki Tablo 1’de verilmiştir.

Tablo 1. Sistem tarafından TBA ve çeşitlerine göre elde edilen performans sonuçları

(*:Toplam Örnek Sayısı:TÖS - Doğru Sınıflandırılan Örnek Sayısı:DSÖS - Doğru Pozitif Oranı:DPO - Yanlış Pozitif Oranı:YPO- Kesinlik veya hassasiyet:Precision - F-skor:F - Receiver Operating Characteristic alanı:ROC - Precision-Recall alanı:PRC - Ortalama Sınıflandırma Doğruluğu:OSD)

Sonuçlar*	TÖS	DSÖS	DPO	YPO	Precision	F	ROC	PRC	OSD (%)
TBA 10	756	638	0.844	0.366	0.838	0.832	0.875	0.899	84.391
TBA 15	756	634	0.839	0.395	0.834	0.824	0.862	0.889	83.862
TBA 20	756	651	0.861	0.329	0.858	0.852	0.892	0.909	86.111
TBA 25	756	650	0.860	0.353	0.860	0.848	0.895	0.913	85.978
TBA 30	756	641	0.848	0.385	0.847	0.833	0.900	0.915	84.788
TBA 35	756	647	0.856	0.379	0.859	0.841	0.894	0.911	85.582
TBA 40	756	645	0.853	0.380	0.855	0.838	0.887	0.905	85.317
KTBA 10	756	562	0.743	0.678	0.687	0.676	0.566	0.655	74.338
KTBA 15	756	564	0.746	0.691	0.692	0.671	0.563	0.651	74.603
KTBA 20	756	563	0.745	0.719	0.680	0.655	0.528	0.644	74.470
KTBA 25	756	561	0.742	0.720	0.667	0.653	0.500	0.627	74.206
KTBA 30	756	566	0.749	0.718	0.706	0.657	0.550	0.649	74.867
KTBA 35	756	562	0.743	0.719	0.673	0.654	0.501	0.631	74.338
KTBA 40	756	561	0.742	0.720	0.667	0.653	0.537	0.650	74.206
OTBA 10	756	627	0.829	0.422	0.724	0.812	0.856	0.882	82.936
OTBA 15	756	646	0.854	0.369	0.854	0.841	0.889	0.903	85.449
OTBA 20	756	650	0.860	0.374	0.865	0.845	0.930	0.938	85.978
OTBA 25	756	644	0.852	0.366	0.850	0.839	0.910	0.923	85.185
OTBA 30	756	643	0.851	0.425	0.867	0.829	0.933	0.944	85.052
OTBA 35	756	662	0.876	0.341	0.883	0.863	0.940	0.948	87.566
OTBA 40	756	635	0.840	0.450	0.854	0.815	0.933	0.939	83.994

Tablo 1 incelendiğinde, TBA ve buna ait iki farklı çeşidi arasından en iyi performans sonuçlarının, OTBA'nın 35 boyutlu veri grubunda elde edildiği görülmektedir. Bu veri grubunda toplam 756 adet farklı girdinin 662 tanesi doğru tespit edilerek % 87.566 OSD oranı elde edilmiştir. Ayrıca YPO değerinin sıfıra yaklaşması ve DPO, Precision ile F performans sonuçlarının 1'e yakın çıkması bu OSD sonucunu istatistiksel olarak desteklemiştir. OTBA yönteminin 35-boyutlu veri setini, %86.111 OSD oranı ile 20 boyutlu TBA metodu takip etmiştir. Bu yöntemle ait diğer değerlendirme kıstasları, OTBA-35'e göre daha düşük olarak elde edilmiştir. Her iki yöntem ROC ve PRC ölçütleri açısından değerlendirildiğinde ise, OTBA-35'in üstünlüğü bir kez daha istatistiksel olarak kanıtlanmıştır. OTBA-35'e ait ROC ve PRC değerleri yaklaşık olarak 0.95 olarak elde edilmiş ve Şekil 1'de bu durum gösterilmiştir. Çalışma kapsamındaki verilere ait sınıfların bu iki değere göre birbirinden çok iyi bir şekilde ayrıştığı kanıtlanmıştır.

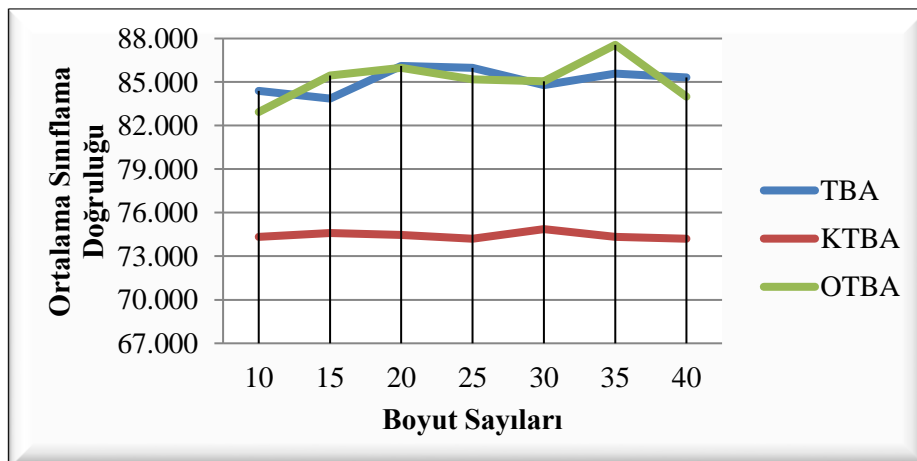


Şekil 1. OTBA-35 yöntemi sonucunda elde edilen ROC ve PRC grafikleri

OTBA ve TBA yöntemlerine göre en kötü sonuçlar KTBA'da elde edilmiştir. Bu TBA çeşidinde boyutlar arası performans kıstasları açısından önemli ölçüde bir fark olmadığı da görülmektedir.

4. Sonuç ve Tartışma

Günümüzde Parkinson gibi bazı kalıcı ve hasta hayatını zorlaştıran rahatsızlıkların önceden tahmin edilmesi ve bunlarla ilgili gerekli tedbirlerin alınması son derece önemlidir. Bu amaçla bir uzmana gerek kalmadan ilgili hastalıkların otomatik olarak tespit edilmesine yönelik araştırmalar önem kazanmaktadır. PH'nin teşhis edilmesinde deneklere ait ses kayıtlarının analizi alanında uzman kişiler tarafından kullanılan yöntemlerden biridir [47]. Eğer bu ses kayıtları çeşitli mühendislik uygulamaları ile işlenerek yüksek doğrulukta otomatik bir PH teşhis sistemleri geliştirilebilirse, bu sistemler hekimlerin kararını destekleyebilir, hatta medikal ve uzman yetersizliği bulunan bölgelerde ön araştırma amacıyla da kullanılabilir. Bu uzman sistemlerin çevrimiçi sürümleri de geliştirilebilirse, kişiler kendi evlerinde sisteme ses kayıtlarını yükleyerek hastalığın kendilerinde bulunup bulunmadığına dair ön tahmini herhangi bir tıbbi muayeneye gerek kalmadan gerçekleştirebilirler. Belirlenen sonuçlar neticesinde kişiler alanında uzman hekimlere yönlendirilebilir. Bu amaçla sunulan bu çalışmada kişilerin ses kayıtlarının girdi olarak kullanıldığı yüksek doğrulukta tepki verebilecek otomatik bir PH uzman sistemi önerilmiştir. Çalışma kapsamında, 188'i Parkinson hastası olmak üzere toplam 252 deneye aynı ifadenin üçer kez tekrarlanmasıyla alınan ses kayıtları kullanılmıştır [21]. Bu ses kayıtlarından çıkarılan AQDD özellikleri üzerinde TBA ve buna ait iki farklı boyut azaltma yönteminin kullanıldığı uzman bir sistem tasarımı gerçekleştirilmiştir. Sistem bünyesinde RO algoritması karar mekanizması olarak kullanılmış ve elde edilen sonuçlar bazı değerlendirme kıstaslarıyla yorumlanmıştır. Analizleri yapılan TBA ve çeşitlerine ait ortalama sınıflandırma doğruluk oranları grafiksel olarak Şekil 2'de gösterilmiştir.



Şekil 2. TBA, KTBA ve OTBA yöntemlerine ait %OSD değerleri

Şekil 2 incelendiğinde en iyi performansın OTBA'nın 35 boyutu için yaklaşık %88 olarak elde edildiği görülmektedir. OTBA-35 yönteminden sonra en iyi sonuç %86 OSD değerini geçen performansı ile TBA yönteminin 20 boyutunda elde edilmiştir. Grafikten de anlaşılacağı üzere en kötü performansı KTBA yöntemi sergilemiştir. Grafikte verilmemiş ancak sonuçlarda elde edilmiş yaklaşık 0.95 ile ROC ve PRC değerleri de OTBA-35'e ait performans çıktılarını desteklemiştir.

Aynı veri setinin kullanıldığı ve aynı zamanda bu verinin kaynağı konumunda olan araştırma [21] ile gerçekleştirilmiş bu çalışma neticesinde elde edilen sonuçlar kısaca karşılaştırılmıştır. İlgili çalışmada [21] veriler üzerinde birçok yöntem ile özellikler çıkarılmıştır. Ancak her özellik grubu için en iyi sonucun alındığı sınıflayıcı sistem değişikliğe uğramıştır. Bu durum uzman sistemler için hem zaman hem de işlem karmaşıklığı açısından kısıtlayıcı niteliktedir. Gerçekleştirilen bu çalışmada ise birçok sınıflayıcı denenerek bu seçenek en başından bire indirgenmiş, hem en hızlı hem de en yüksek performansı veren RO algoritması kullanılmıştır. Ayrıca gerçekleştirilen bu çalışmada, ilgili araştırmada [21] belirtilen en iyi sonucun elde edildiği AQDD özellik grubu üzerinde çeşitli boyut indirgeme yöntemleri ile uzman sistemin veri yoğunluğu azaltılmaya ve aynı zamanda daha yüksek performans sonuçları elde edilmeye çalışılmıştır. Bahsedilen araştırmada maksimum sınıflandırma performansı 50 özellik ile %86 olarak [21] elde edilmesine karşın bu çalışmada OTBA'nın 35 boyutu için yaklaşık %88 sınıflandırma sonucuna ulaşılmıştır. Bunun yanında ilgili çalışmadan [21] farklı olarak Precision, ROC ve PRC gibi önemli değerlendirme kıstasları ile bu çalışmaya ait performans sonuçlarının etkinliği desteklenmiştir. Çalışma kapsamında kullanılan veri sayısının literatüre göre fazla olması sonuçların güvenilirliği açısından önemlidir. Bu alanda çalışan veya çalışmayı düşünen araştırmacılar daha fazla sayıdaki verilere farklı yöntemleri uygulayarak yüksek performans sonuçlarına ulaşmayı amaçlayabilirler. Ayrıca bu alandaki çevrimiçi sistemlerin geliştirilmesi de literatüre katkı olması ve çeşitli veri tabanlarının oluşturulması açısından önem arz etmektedir.

Kaynakça

- [1] K. Rana. (2014). *Parkinson Hastalığı* [Online]. Available: <http://www.noroloji.org.tr/TNDDData/Uploads/files/parkinson%20hastal%C4%B1%C4%9F%C4%B1.pdf>.
- [2] S. Özekmekçi, H. Apaydın, S. Oğuz, & İ. Zileli. (2013). *Parkinson Hastalığı Hasta ve Yakınları İçin El Kitabı*. İstanbul, Turkey: Baycınar Tıbbi Yayıncılık ve Reklam Hiz. Tic. Ltd. Şti, p. 98.
- [3] J. W. Langston. (2002). Parkinson's disease: current and future challenges. *Neurotoxicology*, vol. 23, no. 4-5, pp. 443-450.
- [4] J. Parkinso. (1817). An essay on the shaking palsy (Printed by Whittingham and Rowland for Sherwood, Neely, and Jones). ed: London.
- [5] J. Jankovic. (2008). Parkinson's disease: clinical features and diagnosis. *Journal of neurology, neurosurgery & psychiatry*, vol. 79, no. 4, pp. 368-376.
- [6] H. Gümüş, Z. Akpınar, & O. Demir. (2013). Erken evre Parkinson hastalığında motor olmayan semptomların değerlendirilmesi. *Türk Nöroloji Dergisi*, vol. 19, no. 3, pp. 97-103.
- [7] Y. Akgün & S. Peker. (2010). Tremor tedavisinde cerrahi girişimler. *Acıbadem Üniversitesi Sağlık Bilimleri Dergisi*, vol. 1 (3), no. 3, pp. 123-127.
- [8] B. Harel, M. Cannizzaro, & P. J. Snyder. (2004). Variability in fundamental frequency during speech in prodromal and incipient Parkinson's disease: A longitudinal case study. *Brain and cognition*, vol. 56, no. 1, pp. 24-29.
- [9] A. Tsanas, M. A. Little, P. E. McSharry, & L. O. Ramig. (2010). Accurate telemonitoring of Parkinson's disease progression by noninvasive speech tests. *IEEE transactions on Biomedical Engineering*, vol. 57, no. 4, pp. 884-893.
- [10] C. O. Sakar & O. Kursun. (2010). Teliagnosis of Parkinson's disease using measurements of dysphonia. *Journal of medical systems*, vol. 34, no. 4, pp. 591-599.
- [11] H. Gürüler. (2017). A novel diagnosis system for Parkinson's disease using complex-valued artificial neural network with k-means clustering feature weighting method. *Neural Computing and Applications*, vol. 28, no. 7, pp. 1657-1666.
- [12] M. A. Little, P. E. McSharry, E. J. Hunter, J. Spielman, & L. O. Ramig. (2009). Suitability of dysphonia measurements for telemonitoring of Parkinson's disease. *IEEE transactions on biomedical engineering*, vol. 56, no. 4, pp. 1015-1022.
- [13] M. Peker, B. Sen, & D. Delen. (2015). Computer-aided diagnosis of Parkinson's disease using complex-valued neural networks and mRMR feature selection algorithm. *Journal of healthcare engineering*, vol. 6, no. 3, pp. 281-302.
- [14] R. Das. (2010). A comparison of multiple classification methods for diagnosis of Parkinson disease. *Expert Systems with Applications*, vol. 37, no. 2, pp. 1568-1572.
- [15] F. Åström & R. Koker. (2011). A parallel neural network approach to prediction of Parkinson's Disease. *Expert systems with applications*, vol. 38, no. 10, pp. 12470-12474.
- [16] B. E. Sakar ve ark. (2013). 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.
- [17] H.-L. Chen ve ark. (2013). An efficient diagnosis system for detection of Parkinson's disease using fuzzy k-nearest neighbor approach. *Expert systems with applications*, vol. 40, no. 1, pp. 263-271.
- [18] W.-L. Zuo, Z.-Y. Wang, T. Liu, & H.-L. Chen. (2013). Effective detection of Parkinson's disease using an adaptive fuzzy k-nearest neighbor approach. *Biomedical Signal Processing and Control*, vol. 8, no. 4, pp. 364-373.
- [19] B. E. Sakar, G. Serbes, & C. O. Sakar. (2017). Analyzing the effectiveness of vocal features in early teliagnosis of Parkinson's disease. *PloS one*, vol. 12, no. 8, p. e0182428.
- [20] S. Lahmiri, D. A. Dawson, & A. Shmuel. (2018). Performance of machine learning methods in diagnosing Parkinson's disease based on dysphonia measures. *Biomedical Engineering Letters*, vol. 8, no. 1, pp. 29-39.
- [21] C. O. Sakar ve ark. (2019). 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. 255-263.

- [22] Ö. Eskidere. (2012). A Comparison Of Feature Selection Methods For Diagnosis Of Parkinson's Disease From Vocal Measurements. *Sigma*, vol. 30, pp. 402-414.
- [23] I. W. Selesnick. (2011). Wavelet transform with tunable Q-factor. *IEEE transactions on signal processing*, vol. 59, no. 8, pp. 3560-3575.
- [24] F. Bulut. (2017). Different Mathematical Models for Entropy in Information Theory. *Bilge International Journal of Science and Technology Research*, vol. 1 (2), pp. 167-174.
- [25] R. M. Gray (1990). Entropy and information. *Entropy and information theory*: Springer, pp. 21-55.
- [26] S. Aydın, H. M. Saraoğlu, & S. Kara. (2009). Log energy entropy-based EEG classification with multilayer neural networks in seizure. *Annals of biomedical engineering*, vol. 37, no. 12, p. 2626.
- [27] J. F. Kaiser. (1990). On a simple algorithm to calculate the 'energy' of a signal. in *International conference on acoustics, speech, and signal processing*: IEEE, pp. 381-384.
- [28] J. F. Kaiser. (1993). Some useful properties of Teager's energy operators. in *1993 IEEE international conference on acoustics, speech, and signal processing*, vol. 3: IEEE, pp. 149-152.
- [29] P. Maragos, J. F. Kaiser, & T. F. Quatieri. (1993). On amplitude and frequency demodulation using energy operators. *IEEE Transactions on signal processing*, vol. 41, no. 4, pp. 1532-1550.
- [30] S. Solnik, P. Rider, K. Steinweg, P. DeVita, & T. Hortobágyi. (2010). Teager–Kaiser energy operator signal conditioning improves EMG onset detection. *European journal of applied physiology*, vol. 110, no. 3, pp. 489-498.
- [31] R. B. Randall & W. A. Smith. (2017). Application of the Teager Kaiser energy operator to machine diagnostics. in *Tenth Dst Group International Conference on Health and Usage Monitoring Systems*.
- [32] J. Luo & G. Oubong. (2009). A Comparison of SIFT, PCA-SIFT and SURF *International Journal of Image Processing*, vol. 3(4), pp. 143-152.
- [33] L. I. Smith. (2002). A tutorial on principal components analysis. *Technical Report OUCS-2002-12*, pp. 1-26.
- [34] H. Hoffmann. (2007). Kernel PCA for novelty detection. *Pattern recognition*, vol. 40, pp. 863-874.
- [35] V. Tsatsishvili, I. Burunat, F. Cong, P. Toivianen, V. Alluri, & T. Ristaniemi. (2018). On application of kernel PCA for generating stimulus features for fMRI during continuous music listening. *Journal of neuroscience methods*, vol. 303, pp. 1-6.
- [36] K. R. Müller, S. Mika, G. Rätsch, K. Tsuda, & B. Schölkopf. (2001). An introduction to kernel-based learning algorithms. *IEEE transactions on neural networks*, vol. 12(2).
- [37] R. Zhang, W. Wang, & Y. Ma. (2010). Approximations of the standard principal components analysis and kernel PCA. *Expert Systems with Applications*, vol. 37(9), pp. 6531-6537.
- [38] S. M. S. Shah, S. Batool, I. Khan, M. U. Ashraf, S. H. Abbas, & S. A. Hussain. (2017). Feature extraction through parallel Probabilistic Principal Component Analysis for heart disease diagnosis. *Physica A: Statistical Mechanics and its Applications*, vol. 482, pp. 796-807.
- [39] M. E. Tipping & C. M. Bishop. (1999). Probabilistic principal component analysis. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 61(3), pp. 611-622.
- [40] L. Breiman. (2001). Random forests. *Machine learning*, vol. 45, no. 1, pp. 5-32.
- [41] I. H. Witten, E. Frank, L. E. Trigg, M. A. Hall, G. Holmes, & S. J. Cunningham. (1999). Weka: Practical machine learning tools and techniques with Java implementations.
- [42] Y. Ma, L. Guo, & B. Cukic. (2006). A statistical framework for the prediction of fault-proneness. *Advances in Machine Learning Application in Software Engineering*, Idea Group Inc, pp. 237-265.
- [43] D. M. Powers. (2011). Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation.
- [44] N. Nicolov. (2012). *Machine Learning with Applications in Categorization, Popularity and Sequence Labeling: 57th and 58nd slides*. [Online]. Available: http://www.slideshare.net/Nicolas_Nicolov/machine-learning-14528792.
- [45] Ş. Yücelbaş, C. Yücelbaş, G. Tezel, S. Özşen, & Ş. Yosunkaya. (2018). Automatic sleep staging based on SVD, VMD, HHT and morphological features of single-lead ECG signal. *Expert Systems with Applications*, vol. 102, pp. 193-206.
- [46] T. Saito & M. Rehmsmeier. (2015). The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLoS one*, vol. 10, no. 3, p. e0118432.
- [47] T. J. Wroge, Y. Özkanca, C. Demiroglu, D. Si, D. C. Atkins, & R. H. Ghomi. (2018). Parkinson's Disease Diagnosis Using Machine Learning and Voice. in *2018 IEEE Signal Processing in Medicine and Biology Symposium (SPMB)*: IEEE, pp. 1-7.