



# Parkinson Hastalığı Tespitinde Farklı Boyutsallık İndirgeme Yöntemlerinin Karşılaştırılması

Hakan Gündüz<sup>1\*</sup>

<sup>1</sup> Düzce Üniversitesi, Teknoloji Fakültesi, Bilgisayar Mühendisliği Bölümü, Düzce, Türkiye, (ORCID: 0000-0003-2152-5490)

(İlk Geliş Tarihi 5 Aralık 2019 ve Kabul Tarihi 31 Aralık 2019)

(DOI: 10.31590/ejosat.655795)

**ATIF/REFERENCE:** Gündüz, H. (2019). Parkinson Hastalığı Tespitinde Farklı Boyutsallık İndirgeme Yöntemlerinin Karşılaştırılması. *Avrupa Bilim ve Teknoloji Dergisi*, (17), 1164-1172.

## Öz

Parkinson Hastalığı (PH), bireylerin çoklu motor ve motor olmayan özelliklerini doğrudan etkileyen ilerleyici bir sinir hastalığıdır. PH'nin ilk evresinde bireyler genellikle ses bozulmalarıyla karşı karşıya kalır. Bu durumda PH'nin erken tespitinde kişilerin ses kayıtlarından yararlanılır. Ses kayıtlarından sinyal işleme yöntemleriyle çıkarılan öznelikler yapay öğrenme yöntemlerine girdi olarak verilerek bireylerin hastalığa sahip olup olmadığı tespit edilir. Bu çalışmada bireylerin ses kayıtlarından çıkarılan öznelikler iki farklı yapay öğrenme yöntemine girdi olarak verilmiş ve bireyler Parkinson hastası veya sağlıklı olarak sınıflandırılmıştır. Oluşturulan modeller UCI Makine Öğrenmesi deposundan alınan veri kümesi ile eğitilmiştir. Hem eğitilen yapay öğrenme modellerinin karmaşıklığını azaltmak hem de modellerin aşırı öğrenmesini engellemek için öznelikler üzerinde iki farklı boyutsallık indirgeme yöntemi uygulanmıştır. İlk yöntem olan Temel Bileşenler Analizi (TBA)'yle yeni bir öznelik alt uzayı oluşturmak için öznelik kümesi orijinal boyuttan daha az boyuta sahip olan yeni bir alt uzaya yansıtılır. Oluşturulan yeni öznelik uzayında yüksek varyansa sahip bileşenler seçilirken; varyansı düşük bileşenler ihmal edilir. İkinci yöntem olan Özyinelemeli Öznelik Eleme (ÖÖE)'de özneliklere yapay öğrenme yöntemleri kullanılarak ilgililik puanları atanır. İlk aşamada tüm öznelik kümesini kullanan bir model oluşturulur ve her öznelik için bir ilgililik puanı hesaplanır. Sonraki aşamada en az ilgililik puanına sahip öznelik ihmal edilerek model yeniden oluşturulur ve ilgililik puanları tekrar hesaplanır. Bu işlem öznelik kümesinde istenilen sayıda öznelik kalana kadar devam ettirilir. Kullanılan iki Boyutsallık indirgeme yöntemiyle öznelik uzayının boyutları azaltılmış ve indirgenmiş öznelik vektörleriyle Destek Vektör Makineleri (DVM) ve Gradyan Arttırıcı Makineler (GAM) sınıflandırıcıları eğitilmiştir. Elde edilen veri kümesinin örnek sayısı görece az olduğundan sınıflandırıcıların eğitiminde Bireyi Dışarda Bırakan Çapraz Doğrulama (BDBÇD) prosedürü kullanılmıştır. Veri kümesi aynı zamanda dengesiz sınıf dağılımına sahip olduğundan modellerin performans değerlendirmesinde doğruluk oranıyla birlikte F-ölçütü ve Matthews Korelasyon Katsayısı (MKK) ölçütleri kullanılmıştır. Alınan tüm deneysel sonuçlar irdelendiğinde en yüksek sınıflandırma başarısına sadece 13 öznelik kullanılarak erişildiği görülmüştür. ÖÖE yöntemiyle seçilen 13 öznelikle GAM sınıflandırıcısı eğitilerek 0,881 doğruluk oranı elde edilmiştir. Doğruluk oranı öznelik seçimi yapılmadan elde edilen sonuçlara göre yaklaşık %2 oranında artmıştır. Aynı artış sınıfların ayırt edilebilirliğini gösteren MKK oranında da olmuştur. Boyutsallık indirgeme işlemi olmadan elde edilen MKK oranı 0,62 iken ÖÖE yöntemiyle öznelik seçimi yapıldığında oran 0,67'ye yükselmiştir. Kullanılan diğer boyutsallık indirgeme yöntemi olan TBA ise öznelik seçimsiz modellere göre sınıflandırma başarısı arttırmamasına rağmen, aynı başarı oranlarına daha az sayıda öznelikle erişmiştir.

**Anahtar Kelimeler:** Parkinson Hastalığı Tespiti, Boyutsallık İndirgeme, Temel Bileşenler Analizi, Özyinelemeli Öznelik Eleme, Destek Vektör Makineleri, Gradyan Arttırıcı Makineler.

## Comparison of Different Dimensionality Reduction Methods in the Detection Parkinson's Disease

### Abstract

Parkinson's Disease (PD) is a progressive neural disease that directly affects multiple motor and non-motor features of the individuals. PD individuals are often confronted with sound distortion in the first stage of the disease. In this case, the voice recordings of the people

\* Düzce Üniversitesi, Teknoloji Fakültesi, Bilgisayar Mühendisliği Bölümü, Düzce, Türkiye, ORCID: 0000-0003-2152-5490, [hakangunduz@itu.edu.tr](mailto:hakangunduz@itu.edu.tr)

are used for the early detection. The features extracted from the sound recordings by signal processing methods are given as input to machine learning methods for the detection of the PD. In this study, the features extracted from the voice recordings of individuals were given as input to two different machine learning models for the detection of PD. The models were trained with the dataset obtained from the UCI Machine Learning repository. Two different dimensionality reduction methods were applied on the features in order to reduce the complexity of the trained models and to prevent the over-fitting. The first method, Principal Components Analysis (TBA), projects original feature space into a new subspace that has fewer dimensions than the original. In order to reduce feature dimensions, components with high variances in the new feature space are selected. In the second method, Recursive Feature Elimination (RFE), relevance scores are assigned to the features by using machine learning methods. In the first step, a model that uses the entire set of features is created and a relevance score is calculated for each feature. In the next stage, the model is rebuilt by neglecting the feature with the least relevance score and the relevance scores are recalculated. This process is continued until the desired number of features remains in the feature set. After dimensionality reduction process, Support Vector Machines (SVM) and Gradient Boosting Machines (GBM) classifiers were trained with selected features. Since the number of instances in the dataset was small, One Person Out Cross Validation (OPOCV) was used in classifier training. Due to having data imbalance problem, F-Measure and Matthews Correlation Coefficient (MCC) metrics were used along with accuracy in the performance evaluation. When all the experimental results were examined, it was found out that the highest classification success was achieved by using only 13 features. The GBM classifier was trained with 13 features selected by RFE method to obtain an accuracy of 0.881. Accuracy rate increased by about 2% according to the results obtained without feature selection. The same increase was also seen in the rate of MCC that shows the degree of the class distinguishability. While MCC rate obtained without dimensionality reduction was 0.62, the ratio increased to 0.67 when the feature selection was done with the RFE. PCA, which is the other dimensionality reduction method, did not increase the classification success compared to without selection, but achieved the same success rates with fewer features.

**Keywords:** Detection of Parkinson's Disease, Dimensionality Reduction, Principal Component Analysis, Recursive Feature Elimination, Support Vector Machines, Gradient Boosting Machines.

## 1. Giriş

Sağlık bilişim sistemleri, son yıllarda önemli hastalıkların tespitinde ve gelişiminin izlenmesinde yaygın olarak kullanılmaktadır. Yapay öğrenmeye dayalı bilişim sistemleri, özellikle 60 yaşın üzerindeki kişilerde sıkça görülen Parkinson Hastalığı (PH)'nin izlenmesinde kullanılır [1]. PH, çoklu motor ve motor olmayan özelliklerde kayba neden olan ilerleyici bir sinir hastalığıdır [2]. Hastalığın erken teşhisinin yaşam kalitesini artırması nedeniyle, PH'nin tespiti için yüksek doğruluğa ve güvenilirliğe sahip tıp bilişim sistemleri gerekir. Bu sistemler aynı zamanda hekimlerin iş yükünü azaltmayı da hedefler [3]-[7]. PH saptama sistemleri, genel anlamda hastalığın hangi aşamada olduğunu tanımaya odaklanır. Hastalığın en yaygın belirtilerinden biri de ses problemidir ve hastalığın erken safhalarından itibaren ses bozukluklarına sıklıkla rastlanır. Bu nedenle, ses bozukluklarına dayalı sağlık sistemleri son PH tespit çalışmalarında öncü konumdadır [3]-[7]. Ses sinyallerini kullanan bu sistemler ilintili ses özelliklerini kullanarak Parkinson hastalarının sağlıklı bireylerden otomatik olarak ayırt edebilir. Bu sistemlerde klinik olarak ilintili öznelikler elde etmek için çeşitli konuşma sinyal işleme teknikleri kullanılır. Elde edilen bu öznelikler yapay öğrenme yöntemlerine girdi olarak verilerek PH tespitinde güvenilir kararlar alınır. PH çalışmalarında kullanılan veri kümelerinin öznelik vektörlerinin boyutunun yüksek olması ve örnek sayısının az olması nedeniyle, sinyallerden elde edilen öznelikler üzerinde boyutsallık indirgeme gerçekleştirilir. Boyutsallık indirgeme öznelik uzayının yüksek boyutlu olmasının negatif etkilerini ve veri seyrekliği problemini ortadan kaldırır. Boyut indirgenmesi, öznelik seçimi veya öznelik çıkartımı teknikleriyle yapılmaktadır.

PH tespitinde boyutsallık indirgmeden sonraki aşama uygun yapay öğrenme yönteminin seçilmesidir. Son yıllarda yapılan PH çalışmalarında Yapay Sinir Ağları (YSA), Destek Vektör Makineleri (DVM), Rastgele Orman (RF) ve K-En Yakın Komşular (KEYK) gibi yapay öğrenme modelleri sıklıkla kullanılmıştır. Bahsedilen bu modellerden KYEK sadeliği ve sonuçları anlamlandırma kolaylığı açısından ön plandadır. YSA, DVM ve RO modelleri ise doğrusal yapıya sahip olmayan verileri modellemede başarılıdır [8-11]. PH çalışmalarında kullanılan yapay öğrenme modeli ne olursa olsun; modellerin başarısı verilerden seçilen özneliklerin kalitesi ile doğrudan ilgilidir. Boyutsallık indirgeme yöntemleri veride gürültü olarak yer alan öznelikleri ihmal ederken ilintiye sahip öznelikleri ön plana çıkarmaktadır. Böylece tekrarlı ve aynı bilgiyi içeren öznelikler elenir ve sınıflandırma performansında artış olur [12].

Bu çalışmada İstanbul Üniversitesi Cerrahpaşa Tıp Fakültesi Nöroloji Bölümünde kaydedilmiş 188 Parkinson hastası ve 64 sağlıklı bireye ait ses kayıtlarından farklı sinyal işleme teknikleriyle elde edilen öznelikler kullanılmıştır [13]. Bu özneliklere farklı boyutsallık indirgeme yöntemleri uygulanmıştır. Öznelik çıkarma yöntemi olarak Temel Bileşen Analizi (TBA) kullanılırken; öznelik seçimi için Özyinelemeli öznelik eleme (ÖÖE) yönteminden yararlanılmıştır. Boyut indirgemesi yapılan öznelik vektörleri DVM ve Gradyan Arttırıcı Makineler (GAM) sınıflandırıcılarına girdi olarak verilmiştir. Sınıflandırıcıların eğitimi Bireyi Dışarda Bırakan Çapraz Doğrulama (BDBÇD) prosedürüyle yapılmıştır. Eğitilen sınıflandırıcıların performansı Doğruluk Oranı, F-ölçütü ve Matthews Korelasyon Katsayısı ölçütleriyle değerlendirilmiştir. Son olarak alınan sınıflandırma sonuçları aynı veri kümesinin kullanıldığı diğer literatür çalışmalarıyla kıyaslanmıştır. Çalışmada devam eden bölümleri şu şekildedir: İkinci bölümde literatürde yer alan PH tahmini çalışmalarından bahsedilmiştir. Üçüncü bölümde kullanılan veri kümesinin detayları hakkında bilgi verilmiş ve boyutsallık indirgeme, yapay öğrenme ve performans değerlendirme yöntemleri açıklanmıştır. Son bölüm olan 4. Bölüm'de ise çalışmada elde edilen sonuçlar gösterilmiş ve değerlendirilmiştir.

## 2. İlgili Çalışmalar

Bu bölümde, yapay öğrenme algoritmalarını kullanan PH sınıflandırma çalışmaları özetlenmiştir. PH olan bireylerde arasında yavaş hareket, duruş ve denge eksiklikleri gibi pek çok belirtiler olmasına rağmen, hastalığın en önemli belirtisi konuşma ve boğumlanma değişimi olarak tanımlanan “Disfoni”dir [13]. Disfoninin PH'nin en anlamlı öncüsü olması PH çalışmalarının ses sinyallerine odaklanmasının en önemli sebebidir. Literatürdeki birçok çalışma 31 örneğe (23 PH ve 8 sağlıklı birey) sahip 195 ses kaydından oluşan kamuya açık veri setini kullanmıştır [14]. Çalışmalarda kullanılan diğer bir PH veri seti de çoklu konuşma kaydına sahip 40 örnek (20 sağlıklı ve 20 PH birey) içermektedir [4]. Her iki veri kümesindeki ses örneklerinden ses temel frekansı, temel frekanstaki değişkenlik ölçümleri, genliktaki değişkenlik ölçümleri gibi genel öznelikler elde edilmiştir. PH sınıflandırma çalışmalarının çoğu bu veri kümelerindeki öznelikleri içerdiğinden her iki veri kümesinden elde edilen öznelikler “Temel Öznelikler” olarak adlandırılır. Temel özneliklerin yanı sıra, PH tespitinde sinyal işleme tekniklerine dayanan diğer özneliklerden de yararlanır. Sinyal-gürültü oranı (İng. Signal to noise ratio), Mel-frekans Kepstral Katsayıları (MFKK) (İng. Mel Frequency Cepstral Coefficients) ve Ayarlanabilir Q-faktör Dalgacık Dönüşümü (AQDD) (İng. Tunable Q-Factor Wavelet Transform) gibi araçlarla PH sınıflandırma görevleri için etkin öznelikler elde edilebilir [13].

PH hastaları çoğunlukla ses yüksekliğini, kararsızlığı ve sıklık anormalliklerini doğrudan etkileyen ses bozulmalarıyla karşı karşıyadır. Ses kırılmaları ve bozulmuş ses kalitesi de PH bireylerde görülen diğer bozukluklardır. Konuşma işleme teknikleri, bahsedilen bu anormallikleri tespit etmek ve PH ile ilişkili ses özneliklerinin otomatik olarak çıkarmak için sıklıkla tercih edilir. Son on yılda, PH tespitinde ses öznelikleri kullanıldığı birçok yapay öğrenme çalışması yapılmıştır. Örneğin, Tsanas ve diğ. ses özneliklerine sahip yeni bir PH saptama modeli önermiş ve yüksek ilinti puanına sahip ilk 10 özneliği seçmek için farklı öznelik seçim teknikleri uygulamıştır. Minimum Fazlalık Maksimum İlgi (MFMİ) (İng. Minimum Redundancy Maximum Relevance- mRmR), Kabartma (İng. Relief) ve Yerel Öğrenme-Temelli Öznelik Seçimi (İng. Local Learning-Based Feature Selection) yöntemleri öznelik seçiminde kullanılmış ve seçilen özneliklerin performansları Rastgele Orman (İng. Random Forest) ve Destek Vektör Makineleri (DVM) (İng. Support Vector Machines) sınıflandırıcıları değerlendirilmiştir. Bu sınıflandırıcılar parıltı (İng. shimmer) ve ses kıvrım (İng. vocal fold) uyarma özneliklerinden yararlanarak % 98,6'ya varan hassasiyet oranına ulaşmıştır [15].

Vikas ve Sharma, Parkinson hastalarını sağlıklı bireylerden ayırmak için yeni bir model önermiş ve modele girdi olarak bireylerin ses sinyallerinden çıkarılan farklı öznelik kümeleri verilmiştir. Çalışmada gırtlak nabızı, MFKK, perde, titreme ve ısıltı özneliklerinin PH üzerindeki etkileri incelenmiş, MFKK ve gırtlak nabızı özneliklerinin PH ve sağlıklı bireylerde farklı karakteristiğe ve yüksek dalgalanmalara sahip olduğu sonucuna varılmıştır. PH bireylerin sağlıklı bireylerden daha yüksek titreşim ve ısıltı değerlerine sahip oldukları da tespit edilmiştir [8]. Parisi ve diğ. ise PH'nin erken teşhisi için yapay zekâ tabanlı bir sınıflandırma sistemi kurmayı amaçlamıştır. Çalışmada kullanılan veri kümesi California-Irvine Üniversitesi (UCI) Makine Öğrenimi deposundan alınmıştır. Veri kümesinde 68 bireye ait disfoni öznelikleri ve klinik puanlar mevcuttur. Sınıflandırma sistemine giriş olarak verilecek öznelikleri belirlemek için özel maliyet fonksiyonuna sahip Çok Katmanlı Algılayıcı (ÇKA) (İng. Multi Layer Perceptron) eğitilmiş ve eğitim sonucunda en yüksek önem puanına sahip 20 öznelik seçilmiştir. Seçilen öznelikler Lagrange Destek Vektör Makineleri (LDVM) sınıflandırıcısına verilerek sınıflandırma gerçekleştirilmiştir. Öznelik seçimine ve sınıflandırmaya dayalı bu melez sistemin performansı benzer çalışmalarla karşılaştırılmış, önerilen sistemin doğruluk oranının %100'e ulaştığı görülmüştür [16].

Sakar ve diğ. tarafından yapılan yakın dönem çalışmasında, bireylerin PH tanıları için ses sinyallerine AQDD uygulanmıştır. Elde edilen AQDD özneliklerinin başarısı, PH çalışmalarında yaygın olarak kullanılan diğer ses öznelikleriyle karşılaştırılmıştır. Deneyler 252 kişinin çoklu ses örnekleriyle yapılmış ve bu örneklerden farklı öznelik kümeleri çıkarılmıştır. Öznelik altkümeleri girdi verisi olarak farklı türde sınıflandırıcılara verilmiş ve bu sınıflandırıcıların çıktıları çoğunluk oyu (İng. Majority Voting) yöntemiyle birleştirilmiştir. Bu çalışma, AQDD özneliklerinin PH sınıflandırmada sıklıkla kullanılan diğer ses özneliklerinden daha yüksek başarıya sahip olduğunu göstermiştir. Ek olarak, MFKK ve AQDD özneliklerinin birleşimine MFMİ öznelik seçimi uygulanmış ve seçilen özneliklerin sınıflandırma performansını arttırdığı tespit edilmiştir [13].

Yücelbaş ve Yücelbaş, Sakar ve diğ. çalışmasıyla aynı veri kümesini kullanmış ve veri kümesindeki AQDD öznelikleri üzerine farklı Bağımsız Bileşen Analizi (BBA) yöntemleri uygulayarak boyut indirgemesi yapmıştır. Çalışmada indirgenmiş öznelik kümeleri Rastgele Orman (RO) sınıflandırıcısına giriş olarak verilmiş ve sınıflandırıcılar eğitilmiştir. Eğitilen sınıflandırıcıların sonuçları farklı değerlendirme ölçütleriyle sınanmıştır. Çalışmanın sonuçları irdelendiğinde; en yüksek sınıflandırma başarısına %82.01 ile yeniden yapılanma Bağımsız Bileşen Analizi (YBBA) yöntemiyle erişildiği görülmüştür [32].

Bahsedilen ilgili PH çalışmalarında yapay öğrenme yöntemleriyle beraber genellikle ses tabanlı özneliklerin kullandığı açıktır. Ses temelli özneliklerin yanında; elektroensefalogram (EEG), akıllı kalem ve giyilebilir algılayıcılar gibi farklı veri kaynaklarından öznelik çıkaran çalışmalarda bulunmaktadır [17-19].

## 3. Materyal ve Metot

### 3.1. Veri Kümesi

Çalışmada kullanılan veri kümesi UCI Makine Öğrenimi deposundan alınmıştır. Veri kümesi yakın zamanda Sakar ve diğ. tarafından yapılan çalışmada kullanılmıştır [13]. Veri kümesi ile ilgili istatistiki bilgiler Tablo 1'de verilmiştir.

**Tablo 1.** Veri kümesiyle ilgili istatistiki bilgiler.

	PH	Sağlıklı
Birey Sayıları	188	64
Bireylerin Cinsiyet Dağılımı (E/K)	107/81	23/41
Bireylerin Yaş Aralığı	33-87	41-82

Veri toplama aşamasında mikrofonun frekans tepkisi 44.1 KHz'e ayarlanmış ve her birey "a" harfini seslendirmiştir. Her birey için 3 farklı kayıt alınmış böylece elde edilen ses kaydı örneğinin sayısı 756 olmuştur. Bireylere ait ses kaydı örneklerine farklı sinyal işleme teknikleri uygulanarak toplamda 752 adet öznelik elde edilmiştir. Bu özneliklerin türleri ve sayıları Tablo 2'de belirtilmiştir.

**Tablo 2.** Veri kümesindeki öznelik türlerinin sayıları.

Öznelik Türü	Öznelik Sayısı
Temel Öznelikler	21
Akustik Öznelikler	33
Mel-Frekans Kepstral Katsayıları(MFKK)	84
Dalgacık Dönüşümü (DD)	182
Ayarlanabilir Q-faktörü Dalgacık Dönüşümü (AQDD)	432

Literatür çalışmaları PH'nin erken dönemde bile kişinin konuşmasını etkilediği göstermiştir [20]. Bu sebeple konuşma öznelikleri PH'yi değerlendirmek ve hastalığın tedaviden sonraki evrimini izlemek için sıklıkla kullanılır. Titreme ve parıltı tabanlı öznelikler, temel frekans parametreleri, harmoniklik parametreleri, Yinelili Zaman Yoğunluk Entropisi (İng. Recurrence Time Density Entropy), Ardışık Bağımlı Dalgalanma Analizi (İng. Detrended Fluctuation Analysis) ve Perde Periyodu Entropisi (İng. Pitch Period Entropy), PH çalışmalarında sıkça kullanılan konuşma öznelikleridir. Sakar ve diğ. çalışmalarında bahsedilen bu öznelikleri Temel Öznelikler olarak adlandırmıştır [13]. Konuşma yoğunluğu, formant frekansları ve bant genişliği gibi Akustik Öznelikler ise konuşma sinyallerinden elde edilen spektrogramlarla oluşturulmuş olup, PH sınıflandırmasında kullanılan diğer anahtar özelliklerdir.

İnsan kulağının özelliklerini taklit eden MFKK, konuşmacı tanıma, otomatik konuşma tanıma, biyomedikal ses tanıma ve PH teşhisi gibi farklı görevlerde konuşma sinyallerinden gürbüz bir öznelik çıkarma yöntemi olarak bilinir. PH dil ve dudak hareketlerinde hızlı bozulmaları yol açtığından, MFKK'lar bu bozulmaları tespit etme yeteneğine sahiptir [6]. PH bireylerin konuşma örneklerinde uzun süreli sesli harflerin tam periyodikliğinde ani değişiklikler görülmektedir. Dalgacık Dönüşümü (DD) (İng. Wavelet Transform) bölgesel ölçekteki bu dalgalanmaları tespit etmede önemli bir araçtır. DD'nin ham temel konuşma sinyali F0'dan elde ettiği belirli öznelikler, PH tanısında birçok çalışmada kullanılmıştır [13].

Ayarlanabilir Q-faktörü Dalgacık Dönüşümü (AQDD), öznelik çıkarımı için kullanılan başka bir yöntemdir. AQDD, sinyal davranışına göre sinyalleri daha iyi kalitede dönüştürmek için 3 ayarlanabilir parametreden (Q (Q faktörü), r (artıklık) ve J (seviye sayısı)) yararlanır. Q faktörü parametresi, sinyallerdeki salınımların sayısı ile doğrudan ilgilidir ve zaman alanında yüksek salımlı sinyaller için nispeten yüksek Q faktörü değeri seçilir. J ayrıştırma aşamasındaki seviye sayısı olarak kabul edilir. R parametresi, zaman şeklini etkilemeden dalgacık konumunu belirlemek için aşırı çınlamayı kontrol eder. Daha önce belirtildiği gibi, PH bireyler ses katlama titreşimindeki periyodikliği yitirirler. Bu nedenle, kullanılan veri kümesindeki AQDD'nin parametreleri, konuşma sinyallerinin zaman bölgesi özellikleri dikkate alınarak ayarlanır [13].

### 3.2. Boyutsallık İndirgeme Yöntemleri

Boyutsallık indirgeme, bir modelin karmaşıklığını azaltmanın ve aşırı öğrenmeyi engellemenin bir yoludur. Boyutsallık indirgeme öznelik seçimi ve öznelik çıkarma olmak üzere iki ana kategoriye ayrılır. Öznelik seçimi yoluyla, orijinal özneliklerin bir alt kümesini seçilirken, öznelik çıkartmada yeni bir öznelik alt uzayı oluşturmak için öznelik kümesi farklı bir uzaya yansıtılmaktadır.

Boyutsallık indirgeme yapay öğrenme algoritmasının hesaplama verimliliğini arttırmasının yanında, modellerin tahmin performansı iyileştirilebilir [22].

Çalışmada boyutsallık indirgeme işlemi için Temel Bileşenler Analizi (TBA) (İng. Principal Component Analysis) ve Özyinelemeli Öznitelik Eleme (ÖÖE) (İng. Recursive Feature Elimination) yöntemleri kullanılmıştır. TBA yöntemiyle öznitelik uzayı yansıtma matrisiyle farklı bir uzaya yansıtılırken, ÖÖE yöntemiyle orijinal özniteliklerin belirli sayıda alt kümeleri seçilir. Kullanılan yöntemlerin detayları aşağıdaki alt bölümlerde açıklanmıştır.

### 3.2.1. Temel bileşenler analizi (TBA)

Temel Bileşenler Analizi (TBA): Temel Bileşen Analizi (TBA) (İng. Principal Component Analysis), özellikle öznitelik çıkarımı ve boyutsallığı azaltma için farklı alanlarda yaygın olarak kullanılan denetimsiz bir doğrusal dönüşüm tekniğidir. TBA'nın diğer popüler uygulamaları arasında keşifsel veri analizi, zaman serisi sinyallerindeki gürültülerin giderilmesi ve biyoinformatik alanındaki genom verilerinin ve gen ekspresyon seviyelerinin analizi bulunur [23]. TBA, öznitelikler arasındaki korelasyona dayanarak verilerdeki örüntüleri çıkarmamıza yardımcı olur. TBA yüksek boyutlu verilerde maksimum varyansın yönlerini bulmayı amaçlar ve bunu orijinal boyuttan daha az boyutta olan yeni bir alt alana yansıtır. Bunun için, örnek  $x$  vektörünü orijinal  $d$ -boyutlu öznitelik uzayından daha az boyutlara sahip yeni bir  $k$ -boyutlu öznitelik alt uzayına yansıtılmamıza izin veren bir  $d \times k$  boyutlu dönüşüm matrisi olan  $W$  oluşturulur.

$$\begin{aligned} \mathbf{x} &= [x_1, x_2, x_3 \dots x_d], & \mathbf{x} \in \mathbb{R}^d, \mathbf{z} &= \mathbf{x}\mathbf{W}, & \mathbf{W} &\in \mathbb{R}^{d \times k} \\ \mathbf{z} &= [z_1, z_2, z_3 \dots z_k], & \mathbf{z} &\in \mathbb{R}^k \end{aligned} \quad (1)$$

Orijinal  $d$  boyutlu verilerin bu yeni  $k$  boyutlu alt uzaya dönüştürülmesinin bir sonucu olarak, birinci ana bileşen mümkün olan en büyük varyansa sahiptir. Birinci bileşeni izleyen diğer ana bileşenler ise bu bileşenlerin diğer ana bileşenlerle ilişkisiz olduğu (ortogonal) kısıtı göz önüne alındığında yine en büyük varyansa sahip olacaktır. TBA yansıtılmaları veri ölçeklendirmeye karşı oldukça duyarlı olduğundan, TBA öncesi öznitelikler standart hale getirilmeli ve tüm özniteliklere eşit önem verilmelidir [24].

### 3.2.2. Özyinelemeli öznitelik eleme (ÖÖE)

Özyinelemeli öznitelik eleme (ÖÖE) özniteliklerin ilgililik puanlarını hesaplarken DVM ve RO gibi yapay öğrenme yöntemlerinden yararlanır [25,26]. Bundan dolayı ÖÖE temel olarak sarıcı (İng. wrapper) bir öznitelik seçim yöntemidir. ÖÖE yöntemi ilk aşamada tüm öznitelik kümesini kullanarak bir model oluşturur ve her öznitelik için bir önem puanı hesaplar. Sonraki aşamada en az önem puanına sahip öznitelik ihmal edilerek model yeniden oluşturulur ve önem puanları tekrar hesaplanır. Bu işlem öznitelik kümesinde istenilen sayıda öznitelik kalana kadar devam ettirilir. Bu nedenle, seçim sonunda istenilen öznitelik alt kümesi boyutu, ÖÖE için bir ayar parametresidir. ÖÖE yönteminde diğer belirlenmesi gereken parametre ise özniteliklerin önem puanlarının belirleneceği yapay öğrenme yöntemidir. DVM, yüksek doğruluğu ve iyi genelleme kabiliyeti nedeniyle ÖÖE için popüler bir algoritmadır. ÖÖE'nin her yinelenmesinde doğrusal DVM modeli eğitilir ve her özniteliğe bir ağırlık değeri atanır. En düşük ağırlığa sahip özniteliklerin sınıflandırma üzerinde etkisi en az olduğundan bir sonraki yenilemede bu öznitelik ihmal edilir. Öznitelik boyutu yüksek olduğunda ise, özniteliklerin birer birer çıkarılması zaman alır. Bu gibi durumlarda, her yinelenmede birden fazla öznitelik ihmal edilebilir [27].

### 3.3. Sınıflandırma Yöntemleri

Çalışmamızda PH sınıflandırması için Destek Vektör Makineleri ve Gradyan Arttırıcı Makineler (İng. Gradient Boosting Machines) sınıflandırıcıları kullanılmıştır. Farklı boyutsallık indirgeme yöntemlerinden elde edilen öznitelik alt kümeleri bahsedilen 2 sınıflandırıcıya girdi olarak verilerek sınıflandırma gerçekleştirilmiştir. Sınıflandırıcıların detaylarına aşağıdaki alt başlıklarda değinilmiştir.

#### 3.3.1. Destek vektör makineleri (DVM)

Destek Vektör Makineleri (DVM) denetimli bir öğrenme modeli olup hem sınıflandırma hem de regresyon problemlerinde kullanılır. İkili sınıflandırma problemlerinde, veriler doğrusal olarak ayrılabilirse, bu ayırım sonsuz sayıda hiper düzlemlerle yapılabilir. DVM'nin sınıfları birbirinden ayırmak için en büyük kenar boşluğuna sahip doğrusal fonksiyonu bulması amaçlanmaktadır. DVM, çekirdek (İng. Kernel) işlemleri yoluyla doğrusal olmayan verileri de başarıyla sınıflandırma yeteneğine sahiptir. Doğrusal olmayan verilerde doğrusal ayrılabilirliği sağlamak için  $n$  boyutlu örnekler çekirdek fonksiyonları kullanılarak yeni bir  $m$  boyutlu uzaya ( $m > n$ ) yansıtılır. Yeni uzaydaki örnekler düzlemler kullanılarak iki sınıfa ayrılır. DVM'deki parametreler kullanılan çekirdek fonksiyonunun türüne bağlı olarak değişir. Örneğin,  $C$  eğitilen modelin karmaşıklığını tanımlayan düzenleme parametresidir. Düşük  $C$  değerleri, yanlış sınıflandırılmış örneklerle sahip olabilecek daha basit bir model sağlarken, daha yüksek  $C$  değerleri, modelin varyansını arttırır ve aşırı öğrenmeye neden olur [28].

#### 3.3.2. Gradyan arttırıcı makineler (GAM)

Arttırma (İng. Boosting), tek bir güçlü öğrenici oluşturmak için bir dizi zayıf öğreniciyi bir araya getiren bir tahmin yöntemidir. Arttırma yöntemi, başarılı performanslarından dolayı birçok probleme başarıyla uygulanmıştır [29,30]. Gradyan Arttırma (GA) [31] kullanıldığında, zayıf öğrenenlerin eğitimi ile modeller eklemeli bir şekilde oluşturulur. Zayıf bir öğrenci olarak doğrusal regresyon, karar ağacı ve regresyon ağacı seçilebilir. GA modelinin formülü aşağıda eşitlikte gösterilmiştir:

$$\hat{y}_i = \sum_{k=1}^K f_k(x_i), \quad f_k \in F \quad (2)$$

Eşitlikte  $K$  öğrenenlerin sayısını,  $f_k$  ise fonksiyon uzayında tanımlı bir fonksiyonu (örneğin karar ağacı veya doğrusal fonksiyon) ifade eder.  $x_i$  veri kümesi örneği ve  $y_i$  ise  $x_i$  örneği için tahmin edilen sınıf etiketidir. GA modelleri yinelemeli olarak eğitilir. Her aşamada, mevcut modelin yanlış sınıflandırmalarını geri kazanmak için bir öğrenici (yeni bir karar ağacı gibi) eğitilir. GA sınıflandırıcısının eğitimi, gradyan düşmesi algoritması ile amaç fonksiyonunun en aza indirilmesiyle gerçekleştirilir [31].

### 3.4. Performans Değerlendirme

Sınıflandırıcıların tahmin performanslarını değerlendirmek için değerlendirme ölçütlerinden faydalanılır. Doğruluk oranı en sık kullanılan ölçütlerden biri olsa da, verilerde sınıf dağılımının dengesiz olması durumunda yanıltıcı sonuçlar verebilir [22]. F-ölçütü ve Matthews Korelasyon Katsayısı gibi değerlendirme ölçütleri, bir sınıflandırıcının sınıf dengesizliği durumunda bile farklı sınıflar arasında ne kadar iyi ayırım yapabileceğini ölçebilir. Bu ölçütler karmaşıklık matrisini temel alarak hesaplanır. Tablo 3'teki gibi karışıklık matrisi, ikili bir sınıflandırma için sınıf başına doğru ve yanlış sınıflandırılmış örneklerin sayısını ifade eder.

**Tablo 3.** İkili sınıflandırma için karışıklık matrisi.

Gerçek/Tahmin Edilen	Pozitif	Negatif
Pozitif	dp	yn
Negatif	yp	dn

Karışıklık matrisinde, dp, yp, yn ve dn sırasıyla doğru pozitif (dp), yanlış pozitif (yp), yanlış negatif (yn) ve doğru negatif (dn) sayımları belirtir. Bu sayımlara dayanarak, doğruluk ve F-Ölçütü şu şekilde hesaplanır [33]:

$$\text{Doğruluk} = \frac{dp+dn}{dp+yp+dn+yn} \quad (3)$$

$$\text{Kesinlik}(K) = \frac{dp}{dp+yp} \quad (4)$$

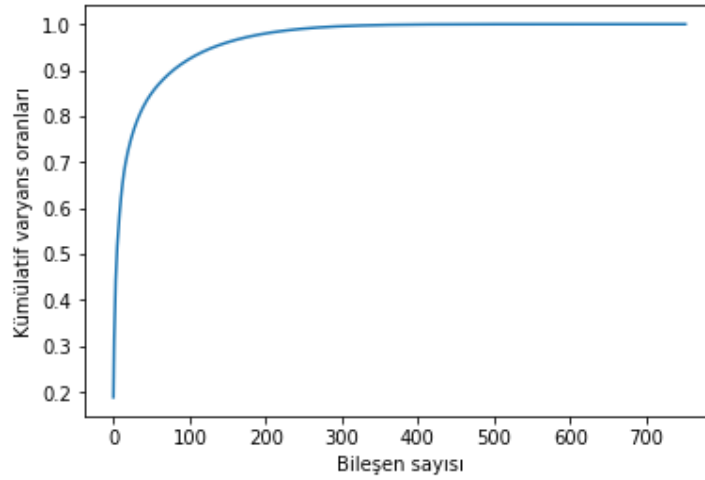
$$\text{Anma}(A) = \frac{dp}{dp+yn} \quad (5)$$

$$\text{F-Ölçütü} = \frac{2 * K * A}{K + A} \quad (6)$$

Matthews Korelasyon Katsayısı (MKK) ise ikili sınıflandırmaların kalitesini ölçmek için kullanılan başka bir ölçüttür. MKK, dp, yp, yn ve dn sayılarını dikkate alır ve genellikle sınıf dağılımı dengesiz olsa bile kullanılabilen dengeli bir ölçüt olarak kabul edilir. MKK esasen gerçek ve öngörülen durumlar arasında bir korelasyon katsayısıdır ve “-1” ile “+1” arasında bir değer alır. “+1” değeri mükemmel bir tahmin gösterirken, “-1” değeri tahmin ile gerçek etiketler arasındaki uyumsuzluğu belirtir.

## 4. Araştırma Sonuçları ve Tartışma

Bu bölümde, farklı boyutsallık indirgeme ve sınıflandırma yöntemleriyle elde edilen deneysel sonuçların ayrıntılarını açıklanmıştır. Elde edilen veri kümesinde az sayıda örnek bulunması nedeniyle, sınıflandırıcıların performans değerlendirilmesi Bireyi Dışarda Bırakan Çapraz Doğrulama (BDBÇD) prosedürüyle yapılmıştır. BDBÇD'nin her yinelemesinde, bir bireye ait olan örnekler test kümesi olarak ayrılırken, kalan diğer bireylerin örnekleri eğitim kümesi olarak kullanılmıştır. Veri kümesinde her bireye ait 3 kayıt olduğundan, bu kayıtların sınıf etiketleri ayrı tahmin edilmiş ve tahminlerin çoğunluğuna bakılarak o kişi için nihai sınıf etiketi belirlenmiştir.



Şekil 1. Bileşenlerin sahip olduğu kümülatif varyans oranları.

Veri kümesiyle yapılan ilk deneylerde herhangi bir boyutsallık indirgeme yöntemi kullanılmamıştır. Veri kümesi örnekleri direkt olarak DVM ve GAM sınıflandırıcılarına verilmiş ve sınıflandırıcıların performansları değerlendirilmiştir. Sınıflandırıcıların parametrelerini en iyilemek için farklı parametre kümeleri denenmiştir. DVM sınıflandırıcısında çekirdek türü ve C parametreleri için değerler bulunurken, GAM sınıflandırıcısı için öğrenici sayısı ve öğrenme oranları en iyilenmiştir. Boyutsallık indirgeme işlemi yapılmadan elde edilmiş sonuçlar Tablo 4'te gösterilmiştir.

Sonuçlar GAM sınıflandırıcısının 0,861 doğruluk ve 0,631 MKK oranlarıyla DVM sınıflandırıcısından daha yüksek performansa sahip olduğunu göstermiştir. F-Ölçütü ve MKK metrikleri açısından bakıldığında ise yine GAM sınıflandırıcısının DVM sınıflandırıcısına üstünlüğü bulunmaktadır. Bu deneylerin ardından boyutsallık indirme yöntemleriyle yeni deneyler gerçekleştirilmiştir. Veri kümesine ilk olarak TBA yöntemi uygulanmış ve her bir temel bileşenin sahip olduğu varyans oranları bulunmuştur. Bileşenlerin varyans oranları kümülatif olarak toplanmış ve Şekil 1'de çizdirilmiştir. Böylece bileşen sayısı ve bileşenlerin sahip olduğu toplam varyans oranları arasındaki ilişki bulunmuştur. İlk deneylerde toplam varyansın %70'ine sahip olan ilk 18 bileşen seçilmiştir. Böylece öznitelik uzayının boyutu 732'den 18'e indirgenmiştir. Seçilen bu 18 bileşen giriş verisi olarak DVM ve GAM sınıflandırıcılarına verilmiştir.

Tablo 4. DVM ve GAM sınıflandırıcılarıyla alınan sınıflandırma sonuçları.

Yöntemler	DVM			GAM		
	Doğruluk Oranı	F-Ölçütü	MKK	Doğruluk Oranı	F-Ölçütü	MKK
Öznitelik Seçimsiz	0,841	0,901	0,543	0,869	0,917	0,632
TBA (18 Öznitelik)	0,837	0,899	0,530	0,845	0,902	0,559
ÖÖE (18 Öznitelik)	<b>0,873</b>	<b>0,918</b>	<b>0,647</b>	<b>0,881</b>	<b>0,923</b>	<b>0,671</b>
TBA (35 Öznitelik)	0,841	0,898	0,553	0,841	0,900	0,545
ÖÖE (35 Öznitelik)	0,861	0,912	0,607	0,873	0,919	0,645
TBA (83 Öznitelik)	0,849	0,907	0,572	0,845	0,901	0,561
ÖÖE (83 Öznitelik)	0,865	0,914	0,620	0,865	0,913	0,622

Alınan sınıflandırma performansları Tablo 4'te sergilenmiştir. Sonuçlar her iki sınıflandırıcıyla alınan performansın (yaklaşık 0,84 doğruluk oranı) birbirine yakın olduğunu göstermiştir. TBA yönteminin performansını ÖÖE yöntemiyle karşılaştırmak için, ÖÖE ile de en yüksek öznitelik önem puanına sahip 18 öznitelik seçilmiştir ve bu öznitelikler DVM ve GAM sınıflandırıcılarına girdi olarak verilmiştir. Alınan sınıflandırma sonuçları Tablo 4'te listelenmiştir. ÖÖE yöntemiyle seçilen 18 öznitelikle DVM sınıflandırıcısı 0,873 doğruluk ve 0,647 MKK oranlarına erişmiştir. GAM sınıflandırıcısı ise 0,881 doğruluk ve 0,671 MKK oranlarına sahiptir. Toplam varyansın 0,7'sine sahip temel bileşenlerle deneyler yapıldıktan sonra sırasıyla 0,8 ve 0,9 varyans oranlarına sahip bileşenler seçilmiştir. Toplam varyans oranı 0,8 olduğunda seçilen bileşen sayısı 35 olurken; varyans oranı 0,9'a yükseldiğinde ise 83 bileşen seçilmiştir. Bir önceki deneyde olduğu gibi ÖÖE yöntemiyle seçilen öznitelik sayısı TBA ile aynı sayıdadır. Seçilmiş bu özniteliklerle elde edilen sonuçlar Tablo 4'te yer almaktadır. 35 öznitelikle TBA yöntemi hem DVM hem de GAM sınıflandırıcısıyla yaklaşık 0,84'lük doğruluk oranına erişmiştir. ÖÖE yönteminde ise GAM sınıflandırıcısı 0,873 doğruluk ve 0,644 MKK; DVM sınıflandırıcısı ise 0,861 doğruluk ve 0,607 MKK oranlarına sahiptir. 83 öznitelikle alınan sonuçlarda ise DVM ve GAM sınıflandırıcıları TBA yöntemiyle 0,85 civarında

doğruluk oranına ulaşmıştır. Doğruluk oranı ÖÖE yöntemiyle 0,86'ya kadar yükselirken, MKK oranı da 0,62 seviyesine çıkmıştır. Alınan tüm sonuçlar irdelendiğinde en yüksek sınıflandırma başarısına sadece 13 öznelikle kullanılarak erişildiği görülmüştür. ÖÖE yöntemiyle seçilen 13 öznelikle GAM sınıflandırıcısı eğitilerek 0,881 doğruluk oranı elde edilmiştir. Doğruluk oranı öznelikle seçimi yapılmadan elde edilen sonuçlara göre yaklaşık %2 oranında artmıştır. Aynı artış sınıfların ayırt edilebilirliğini gösteren MKK oranında da olmuştur. Boyutsallık indirgeme işlemi olmadan elde edilen MKK oranı 0,62 iken ÖÖE yöntemiyle öznelikle seçimi yapıldığında oran 0,67'ye yükselmiştir.

## 5. Sonuç

İlerleyici bir sinir hastalığı olan Parkinson bireylerin yaşam kalitesini doğrudan etkilemektedir. PH'nin en yaygın belirtilerinden biri ses problemdir ve hastalığın erken evrelerde bireylerin ses özelliklerinde değişikliklere neden olmaktadır. Bu nedenle PH tespiti sistemlerinde bireylerin ses kayıtları sıklıkla kullanılır. Tespit sistemlerinde ham ses kayıtlarından sinyal işleme teknikleriyle ayırt edici öznelikle çıkarılır ve bu öznelikle yapay öğrenme modelleri eğitilir. Sistemlerin yüksek doğruluğa sahip olması için bilgi içeren özneliklerin çıkarılması ve etkin yapay öğrenme yönteminin seçilmesi gereklidir. Yüksek doğruluklu ve güvenilirli sistemler hekimlerin iş yükünü azaltmayı ve kararlarını desteklemeyi hedefler. Bu amaçla yapılan çalışmada bireylerin ses kayıtlarından çıkarılan çeşitli öznelikle kümelerinin yapay öğrenme yöntemlerine girdi olarak verildiği bir PH tanıma sistemi önerilmiştir. Çalışmada UCI veri deposundan elde edilen veri kümesi kullanılmıştır. Veri kümesinde 188'i Parkinson hastası olmak üzere toplam 252 bireye ait çoklu ses kayıtları vardır [21]. Bireylerin ses kayıtlarından farklı türde 754 öznelikle çıkarılmıştır. Genişleyen öznelikle uzayının boyutunu indirgemek için öznelikle üzerine TBA ve ÖÖE boyutsallık indirgeme yöntemleri uygulanmıştır. Kullanılan bu yöntemlerle farklı sayıda ilintili öznelikle seçilmiştir. İndirgenmiş öznelikle DVM ve GAM yapay öğrenme yöntemlerine girdi olarak verilmiş ve modeller BDBÇD prosedürüyle eğitilmiştir. Yapay öğrenme modellerinin performansları farklı ölçütlerle değerlendirilmiştir. Alınan tüm sonuçlar içerisinde en yüksek sınıflandırma başarısına ÖÖE yöntemi ve GAM sınıflandırıcısı çiftiyle erişilmiştir. ÖÖE-GAM çiftiyle yaklaşık olarak 0.881 doğruluk ve 0.67 MKK oranları yakalanmıştır. Bu başarı oranına ÖÖE indirgeme yöntemiyle seçilen sadece 13 öznelikle erişilmiştir. ÖÖE-GAM çiftinin sınıflandırma başarısı öznelikle seçimi olmadan eğitilmiş modellerle de karşılaştırılmıştır. ÖÖE-GAM çifti doğruluk oranını %2, MKK oranını ise %5 geliştirmiştir. Kullanılan diğer boyutsallık indirgeme yöntemi olan TBA ise öznelikle seçimsiz modellere göre sınıflandırma başarısı arttırmamasına rağmen, aynı başarı oranlarına daha az sayıda öznelikle erişmiştir. Bu deneylerde kullanılan öznelikle sayısı 752'den 13 kadar inmiştir. Ek olarak, deneysel sonuçlarımız Sakar ve diğ. çalışmasıyla da kıyaslanmıştır [21]. Bu çalışmada oluşturulan modellerin eğitim metodolojisi (veri seti, eğitim prosedürü ve değerlendirme ölçütlerini) çalışmamızla birebir benzerdir. Böylece çalışmamızda alınan deneysel sonuçlar doğrudan bu çalışmanın sonuçlarıyla karşılaştırılmıştır. Sakar ve diğ. [16], çalışmalarının ilk aşamasında farklı ses öznelikle kümelerini birleştirmiş ve bu özneliklerden MFMİ seçimiyle ilintili olanlarını seçmiştir. Seçilen bu özneliklelerle farklı yapay öğrenme modelleri eğitilmiş ve DVM modeliyle 0.860'lık doğruluk ve 0.590'lık MKK oranları yakalanmıştır. Çalışmamızda önerilen ÖÖE-GAM modeli Sakar ve diğ. tarafından önerilen modele göre doğruluk ve MKK oranlarını sırasıyla %2 ve %5 arttırmıştır (0.881 doğruluk ve 0.67 MKK oranları).

## Kaynakça

- [1]Launer, L. J., Berger, K., Breteler, M. M., Dartigues, J. F., Baldereschi, M., Fratiglioni, L., ... & Hofman, A. (2000). Prevalence of Parkinson's disease in Europe: A collaborative study of population-based cohorts. *Neurologic Diseases in the Elderly Research Group. Neurology*, 54(11 Suppl 5), S21-3.
- [2]Jankovic, J. (2008). Parkinson's disease: clinical features and diagnosis. *Journal of neurology, neurosurgery & psychiatry*, 79(4), 368-376.
- [3]Tsanas, A., Little, M. A., McSharry, P. E., & Ramig, L. O. (2009). Accurate telemonitoring of Parkinson's disease progression by noninvasive speech tests. *IEEE transactions on Biomedical Engineering*, 57(4), 884-893.
- [4]Sakar, B. E., Isenkul, M. E., Sakar, C. O., Sertbas, A., Gurgun, F., Delil, S., ... & Kursun, O. (2013). Collection and analysis of a Parkinson speech dataset with multiple types of sound recordings. *IEEE Journal of Biomedical and Health Informatics*, 17(4), 828-834.
- [5]Gürüler, H. (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*, 28(7), 1657-1666.
- [6]Peker, M. (2016). A decision support system to improve medical diagnosis using a combination of k-medoids clustering based attribute weighting and SVM. *Journal of medical systems*, 40(5), 116.
- [7]Sakar, B. E., Serbes, G., & Sakar, C. O. (2017). Analyzing the effectiveness of vocal features in early telediagnosis of Parkinson's disease. *PloS one*, 12(8), e0182428.
- [8]Sharma, A., & Giri, R. N. (2014). Automatic recognition of Parkinson's Disease via artificial neural network and support vector machine. *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, 4(3), 2278-3075.
- [9]Shahbakhi, M., Far, D. T., & Tahami, E. (2014). Speech analysis for diagnosis of parkinson's disease using genetic algorithm and support vector machine. *Journal of Biomedical Science and Engineering*, 7(4), 147-156.
- [10]Kubota, K. J., Chen, J. A., & Little, M. A. (2016). Machine learning for large-scale wearable sensor data in Parkinson's disease: Concepts, promises, pitfalls, and futures. *Movement disorders*, 31(9), 1314- 1326.
- [11]Alemami, Y., & Almazaydeh, L. (2014). Detection of Parkinson disease through voice signal features. *The Journal of American Science*, 10(10), 44-47.
- [12]GÜNDÜZ, H., Cataltepe, Z., & Yaslan, Y. (2017). Stock daily return prediction using expanded features and feature selection. *Turkish Journal of Electrical Engineering & Computer Sciences*, 25(6), 4829-4840.



- [13]Sakar, C. O., Serbes, G., Gunduz, A., Tunc, H. C., Nizam, H., Sakar, B. E., ... & Apaydin, H. (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*, 74, 255-263.
- [14]Little, M. A., McSharry, P. E., Hunter, E. J., Spielman, J., & Ramig, L. O. (2008). Suitability of dysphonia measurements for telemonitoring of Parkinson's disease. *IEEE transactions on biomedical engineering*, 56(4), 1015-1022.
- [15]Tsanas, A., Little, M. A., McSharry, P. E., Spielman, J., & Ramig, L. O. (2012). Novel speech signal processing algorithms for high-accuracy classification of Parkinson's disease. *IEEE transactions on biomedical engineering*, 59(5), 1264-1271.
- [16]Parisi, L., RaviChandran, N., & Manaog, M. L. (2018). Feature-driven machine learning to improve early diagnosis of Parkinson's disease. *Expert Systems with Applications*, 110, 182-190.
- [17]Oh, S. L., Hagiwara, Y., Raghavendra, U., Yuvaraj, R., Arunkumar, N., Murugappan, M., & Acharya, U. R. (2018). A deep learning approach for Parkinson's disease diagnosis from EEG signals. *Neural Computing and Applications*, 1-7.
- [18]Pereira, C. R., Weber, S. A., Hook, C., Rosa, G. H., & Papa, J. P. (2016, October). Deep learning-aided Parkinson's disease diagnosis from handwritten dynamics. In *2016 29th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)* (pp. 340-346). Ieee.
- [19]Eskofier, B. M., Lee, S. I., Daneault, J. F., Golabchi, F. N., Ferreira-Carvalho, G., Vergara-Diaz, G., ... & Bonato, P. (2016, August). Recent machine learning advancements in sensor-based mobility analysis: Deep learning for Parkinson's disease assessment. In *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*(pp. 655-658). IEEE.
- [20]Yunusova, Y., Weismer, G., Westbury, J. R., & Lindstrom, M. J. (2008). Articulatory movements during vowels in speakers with dysarthria and healthy controls. *Journal of Speech, Language, and Hearing Research*.
- [21]Selesnick, I. W. (2011). Wavelet transform with tunable Q-factor. *IEEE transactions on signal processing*, 59(8), 3560-3575.
- [22]Gunduz, H., & Cataltepe, Z. (2015). Borsa Istanbul (BIST) daily prediction using financial news and balanced feature selection. *Expert Systems with Applications*, 42(22), 9001-9011.
- [23]Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer.
- [24]Yücelbaş, Ş., & Yücelbaş, C. (2019). Temel Bileşen Analizi Yöntemleri Kullanarak Parkinson Hastalığının Otomatik Teşhisi. *Avrupa Bilim ve Teknoloji Dergisi*, (16), 294-300.
- [25]Granitto, P. M., Furlanello, C., Biasioli, F., & Gasperi, F. (2006). Recursive feature elimination with random forest for PTR-MS analysis of agroindustrial products. *Chemometrics and Intelligent Laboratory Systems*, 83(2), 83-90.
- [26]Yan, K., & Zhang, D. (2015). Feature selection and analysis on correlated gas sensor data with recursive feature elimination. *Sensors and Actuators B: Chemical*, 212, 353-363.
- [27]Weston, J., & Guyon, I. (2012). U.S. Patent No. 8,095,483. Washington, DC: U.S. Patent and Trademark Office.
- [28]Suykens, J. A., & Vandewalle, J. (1999). Least squares support vector machine classifiers. *Neural processing letters*, 9(3), 293-300.
- [29]Freund, Y., & Schapire, R. E. (1996, July). Experiments with a new boosting algorithm. In *icml* (Vol. 96, pp. 148-156).
- [30]Schapire, R. E., & Freund, Y. (2013). *Boosting: Foundations and algorithms*. Kybernetes.
- [31]Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of statistics*, 1189-1232.
- [32] Yücelbaş, C., & Yücelbaş, Ş. (2019). AQDD Özelliklerine BBA Yöntemleri Uygulanarak Parkinson Hastalığının Otomatik Teşhisi. *Bilecik Şeyh Edebali Üniversitesi Fen Bilimleri Dergisi*, 6, 50-58.
- [33] Gündüz, H., Yaslan, Y., & Çataltepe, Z. (2017). Intraday prediction of Borsa Istanbul using convolutional neural networks and feature correlations. *Knowledge-Based Systems*, 137, 138-148.