

## Machine Learning-Based Classification of Parkinson's Disease Using Voice Signal Features

Süleyman ALTUNOK<sup>1\*</sup>

Bilgisayar Mühendisliği, İskenderun Teknik Üniversitesi, Hatay, Türkiye.

### ABSTRACT

Parkinson's disease (PD) is a progressive neurodegenerative disorder in which early and objective diagnosis remains clinically challenging. Because vocal impairment appears in the majority of PD patients even at early stages, sustained-vowel speech recordings have been widely investigated as a non-invasive biomarker source. In this study, the UCI Parkinson's Disease Classification dataset, comprising 756 voice recordings collected from 252 subjects (188 PD patients and 64 healthy controls) at Istanbul University, was used. A total of 754 acoustic features spanning baseline, intensity, formant, bandwidth, vocal fold, MFCC, wavelet and tunable Q-factor wavelet transform (TQWT) groups were reduced to the 100 most discriminative attributes using an ANOVA F-test based selection procedure. Six supervised classifiers - Support Vector Machine, Random Forest, k-Nearest Neighbors, Multi-Layer Perceptron, Gradient Boosting and Decision Tree - were trained and evaluated using a subject-grouped five-fold cross-validation scheme, in which recordings belonging to the same individual were kept strictly within a single fold to prevent data leakage. Gradient Boosting achieved the most balanced performance with 81.22% accuracy and 72.37% macro F1-score, followed closely by Random Forest with 81.23% accuracy. The results demonstrate that classical machine learning pipelines, when evaluated under subject-independent validation, provide realistic and clinically meaningful estimates of PD classification performance from voice signals.

### ARTICLE INFO

**Received** 2026-05-28 ,

**Accepted** 2026-06-25,

**Publication Date** 2026-06-30

### Keywords:

Parkinson's Disease, Voice Analysis, Machine Learning, Feature Selection, Classification

**Distributed Under** CC-BY 4.0



\*Corresponding Author: Süleyman ALTUNOK, E-mail: suleymanaltunok.lee21@iste.edu.tr

## Ses Sinyali Özniteliklerini Kullanarak Parkinson Hastalığının Makine Öğrenmesi Tabanlı Sınıflandırılması

### ÖZET

Parkinson hastalığı (PH), erken ve objektif tanısı klinik açıdan hâlâ güçlükler barındıran, ilerleyici bir nörodejeneratif hastalıktır. Hastalığın erken evrelerinde bile hastaların büyük bir kısmında ses bozukluklarının ortaya çıkması nedeniyle, sürdürülen ünlü seslerine ait konuşma kayıtları, invaziv olmayan bir biyobelirteç kaynağı olarak yaygın biçimde araştırılmaktadır. Bu çalışmada, İstanbul Üniversitesi'nde 252 bireyden (188 PH hastası ve 64 sağlıklı birey) toplanan 756 ses kaydından oluşan UCI Parkinson Hastalığı Sınıflandırma veri seti kullanılmıştır. Temel özellikler, şiddet, formant frekansı, bant genişliği, vokal kord, MFCC, dalgacık dönüşümü ve ayarlanabilir Q-faktörlü dalgacık dönüşümü (TQWT) gruplarına ait toplam 754 akustik öznitelik, ANOVA F-testi tabanlı bir seçim yöntemiyle en ayırt edici 100 özneliğe indirgenmiştir. Destek Vektör Makinesi, Rastgele Orman, k-En Yakın Komşu, Çok Katmanlı Algılayıcı, Gradient Boosting ve Karar Ağacı olmak üzere altı denetimli sınıflandırıcı, aynı bireye ait kayıtların tek bir katlamada tutularak veri sızıntısının önleildiği, birey-bazlı beş katlı çapraz doğrulama şemasıyla eğitilmiş ve değerlendirilmiştir. Gradient Boosting modeli %81,22 doğruluk ve %72,37 makro F1-skoru ile en dengeli performansı sergilemiş, bunu %81,23 doğrulukla Rastgele Orman modeli yakından takip etmiştir. Elde edilen sonuçlar, birey-bağımsız doğrulama altında değerlendirildiğinde klasik makine öğrenmesi yöntemlerinin ses sinyallerinden PH sınıflandırmasında gerçekçi ve klinik açıdan anlamlı performans tahminleri sunduğunu göstermektedir.

### MAKALE BİLGİSİ

#### Keywords:

Parkinson Hastalığı, Ses Analizi, Makine Öğrenmesi, Öznitelik Seçimi, Sınıflandırma.

Distributed Under CC-BY 4.0



### GİRİŞ

Yapay zeka (YZ), günümüzde hemen hemen her sektörde devrim niteliğinde bir dönüşüm yaratmaktadır. Sağlıkta hastalık teşhisinden ilaç keşfine, finansta risk analizinden dolandırıcılık tespitine, otomotivde otonom sürüşten tarımda verim artırıcı akıllı sistemlere, eğitimde kişiselleştirilmiş öğrenme platformlarından üretime kadar geniş bir yelpazede kullanılmaktadır. Yapay zeka, özellikle sağlık sektöründe köklü bir dönüşüm yaratmaktadır. Hastalıkların erken teşhisinden tedavi planlamasına, görüntüleme analizlerinden (MR, CT, röntgen) kanser tespiti ve kalp hastalıklarının öngörülmesine kadar pek çok kritik alanda kullanılmaktadır. Doktorlara ikinci bir görüş sunan yapay zeka sistemleri, radyolojik görüntüleri saniyeler içinde analiz ederek insan

gözünden kaçabilecek ince detayları yakalayabilmekte, kişiselleştirilmiş tedavi protokolleri oluşturmakta ve nadir hastalıkların teşhis sürecini hızlandırmaktadır. Ayrıca ilaç keşfi ve geliştirme süreçlerini dramatik şekilde kısaltmakta, ameliyat robotları ile cerrahi hassasiyeti artırmakta, hasta izleme sistemleri ile kronik hastalık yönetimini iyileştirmekte ve salgın tahmin modelleriyle kamu sağlığına önemli katkılar sağlamaktadır. Yapay zeka sayesinde sağlık hizmetleri daha erişilebilir, daha hızlı, daha doğru ve daha kişiselleştirilmiş hale gelerek, hem hasta outcomes'larını iyileştirmekte hem de sağlık sistemlerinin verimliliğini önemli ölçüde artırmaktadır (Camgözlü ve Kutlu, 2021;Tohma, 2025; Daşcı, 2025;Bel ve ark. 2025; Kaderoglu ve Satır, 2025).

Parkinson hastalığı (PH), merkezi sinir sisteminde substantia nigra bölgesindeki dopaminerjik nöronların ilerleyici biçimde kaybı sonucu ortaya çıkan, tremor, bradikinezi, kas rijiditesi ve postural instabilite gibi motor bulguların yanı sıra çok sayıda motor olmayan belirtiyile de seyreden kronik bir nörodejeneratif hastalıktır (Jankovic, 2008). Hastalığın kesin tanısı halen büyük ölçüde nörologların klinik muayenesine dayanmakta olup, özellikle erken evrelerde semptomların diğer parkinsonizm sendromlarıyla örtüşmesi nedeniyle tanı doğruluğu sınırlı kalabilmektedir. Bu durum, objektif, tekrarlanabilir ve düşük maliyetli tarama araçlarına duyulan ihtiyacı ön plana çıkarmaktadır.

Yapılan klinik gözlemler, PH hastalarının yaklaşık %90'ının hastalığın erken döneminde bir tür ses ve konuşma bozukluğu geliştirdiğini göstermektedir; bu bozukluklar arasında ses tellerinin titreşim düzensizliği, azalmış ses şiddeti, monoton tonlama ve nefesli ses kalitesi sayılabilir (Sakar ve ark., 2019). Ses üretiminin büyük ölçüde istemsiz kas kontrolüne dayanması, konuşma sinyallerini hastalığın erken dönem belirtilerini yansıtan doğal ve invaziv olmayan bir biyobelirteç kaynağı hâline getirmektedir. Bu nedenle son on beş yılda, sürdürülen ünlü fonasyonlarından ya da bağlantılı konuşmadan çıkarılan akustik özneliklerin PH tanısında kullanılmasına yönelik çalışmalar hızla artmıştır (Tsanas ve ark., 2012; Little ve ark., 2009).

Konuşma sinyali işleme alanındaki gelişmeler, jitter ve shimmer gibi klasik dizinlerin ötesine geçilerek Mel Frekanslı Kepstral Katsayıları (MFCC), dalgacık dönüşümü tabanlı öznelikler ve ayarlanabilir Q-faktörlü dalgacık dönüşümü (TQWT) gibi daha zengin temsillerin çıkarılmasına olanak sağlamıştır (Sakar ve ark., 2013; Sakar ve ark., 2019). Bu yüksek boyutlu öznelik uzayları, makine öğrenmesi tabanlı sınıflandırıcılarla birleştirildiğinde, sağlıklı bireyler ile PH hastalarını yüksek doğrulukla ayırt edebilen karar destek sistemlerinin geliştirilmesine imkân tanımaktadır. Ancak literatürdeki birçok çalışmada, aynı bireye ait birden fazla ses kaydının eğitim ve test kümeleri arasında rastgele dağıtılması sonucunda veri sızıntısı olduğu ve bildirilen başarı oranlarının gerçek genelleme performansını olduğundan yüksek gösterdiği bilinmektedir (Hossain & Amenta, 2024).

Bu ödev kapsamında, UCI Makine Öğrenmesi Deposu'nda yer alan ve İstanbul Üniversitesi Cerrahpaşa Tıp Fakültesi Nöroloji Anabilim Dalı'nda toplanan Parkinson's Disease Classification veri seti kullanılmıştır. Veri seti, 188 PH hastası ve 64 sağlıklı bireyden elde edilen toplam 756 ses

kaydını ve her kayıt için 754 akustik özniteliği içermektedir. Çalışmanın amacı, bu yüksek boyutlu öznitelik uzayı üzerinde istatistiksel bir öznitelik seçimi uygulayarak, altı farklı klasik makine öğrenmesi sınıflandırıcısının (Destek Vektör Makinesi, Rastgele Orman, k-En Yakın Komşu, Çok Katmanlı Algılayıcı, Gradient Boosting ve Karar Ağacı) PH sınıflandırma performansını karşılaştırmalı olarak değerlendirmektir.

Çalışmanın literatürden ayrılan temel noktası, değerlendirme protokolünde bireysel düzeyde grup ayrımı (subject-grouped cross-validation) uygulanmasıdır. Aynı hastaya ait üç tekrarlı kayıt, eğitim ve test kümelerinde asla birlikte bulunmayacak şekilde ayrıştırılarak, modelin bireysel ses özelliklerini ezberlemesinin önüne geçilmiş ve yalnızca gerçekten görülmemiş hastalar üzerindeki genelleme başarısı ölçülmüştür. Bu yaklaşımla elde edilen sonuçların, klinik uygulanabilirlik açısından daha güvenilir bir performans göstergesi sunması hedeflenmiştir.

## MATERYAL VE METOD

### Veri Seti

Bu çalışmada kullanılan veri seti, UCI Makine Öğrenmesi Deposu'nda "Parkinson's Disease Classification" adıyla yayımlanmış olup İstanbul Üniversitesi Cerrahpaşa Tıp Fakültesi Nöroloji Anabilim Dalı'nda toplanmıştır (Sakar ve ark., 2018). Veri seti, yaşları 33 ile 87 arasında değişen (ortalama  $65,1 \pm 10,9$ ) 188 PH hastasından (107 erkek, 81 kadın) ve yaşları 41 ile 82 arasında değişen (ortalama  $61,1 \pm 8,9$ ) 64 sağlıklı bireyden (23 erkek, 41 kadın) oluşan toplam 252 katılımcıya aittir. Ses kayıtları, 44,1 kHz örnekleme frekansına sahip bir mikrofon aracılığıyla, her katılımcıdan sürdürülen /a/ ünlüsünün üç tekrarlı fonasyonu şeklinde toplanmış ve bu sayede toplam 756 kayıt ( $252 \times 3$ ) elde edilmiştir.

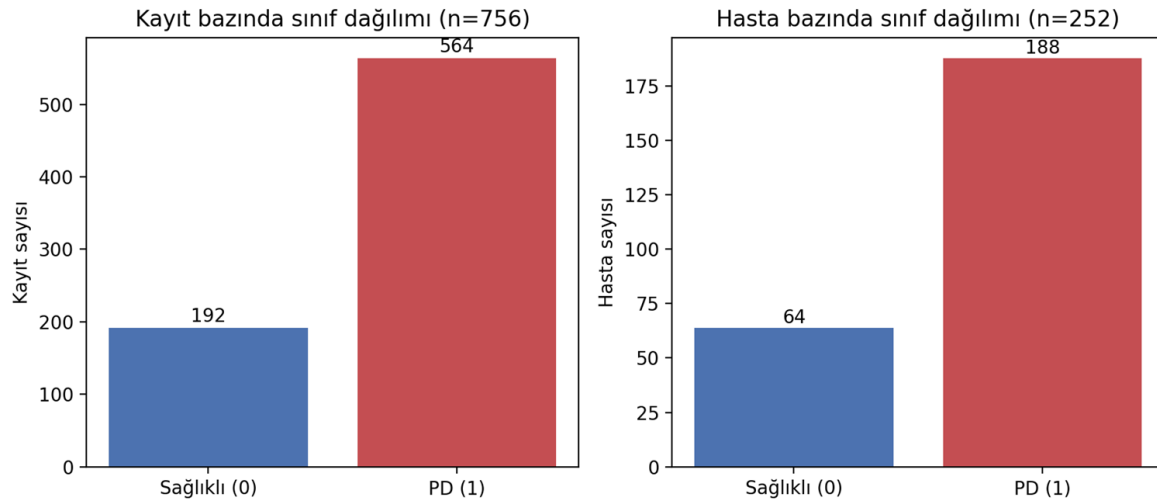
Her ses kaydından, klasik dysphonia ölçütlerinden gelişmiş zaman-frekans dönüşümlerine kadar uzanan toplam 754 akustik öznitelik çıkarılmıştır. Bu öznitelikler; temel (baseline) öznitelikler, şiddet parametreleri, formant frekansları, bant genişliği parametreleri, vokal kord öznitelikleri, MFCC, dalgacık dönüşümü ve TQWT olmak üzere sekiz ana grupta toplanmaktadır. Tablo 1'de bu grupların veri setindeki sütun aralıkları ve öznitelik sayıları özetlenmiştir.

**Tablo 1. Veri setindeki öznitelik gruplarının dağılımı**

| Öznitelik Grubu             | Sütun Aralığı   | Öznitelik Sayısı |
|-----------------------------|-----------------|------------------|
| Temel Özellikler (Baseline) | Col_3 – Col_23  | 21               |
| Şiddet Parametreleri        | Col_24 – Col_26 | 3                |

|                              |                   |     |
|------------------------------|-------------------|-----|
| Formant Frekansları          | Col_27 – Col_30   | 4   |
| Bant Genişliği Parametreleri | Col_31 – Col_34   | 4   |
| Vokal Kord Özellikleri       | Col_35 – Col_56   | 22  |
| MFCC                         | Col_57 – Col_140  | 84  |
| Dalgacık Dönüşümü (Wavelet)  | Col_141 – Col_322 | 182 |
| TQWT Özellikleri             | Col_323 – Col_754 | 432 |

Şekil 1'de veri setinin sınıf dağılımı hem kayıt bazında (756 kayıt) hem de hasta bazında (252 birey) gösterilmiştir. Görüldüğü üzere veri seti, PH sınıfı lehine belirgin bir dengesizlik içermektedir; bu durum, model değerlendirmesinde yalnızca doğruluk metriğine değil makro ortalama precision, recall ve F1-skoruna da yer verilmesini gerekli kılmaktadır.



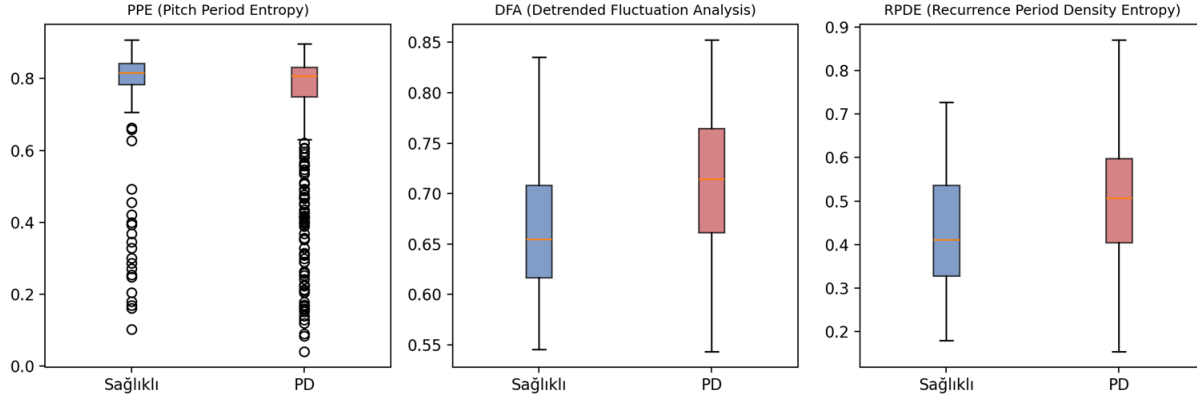
Şekil 1. Veri setindeki sınıf dağılımının kayıt ve hasta bazında gösterimi

### Veri Ön İşleme ve Öznitelik Seçimi

Öznitelik ölçeklerinin sınıflandırıcı performansı üzerindeki olumsuz etkisini gidermek amacıyla, tüm öznitelikler standart ölçeklendirme (StandardScaler) yöntemiyle sıfır ortalama ve birim varyansa sahip olacak şekilde dönüştürülmüştür. Ölçekleme parametreleri yalnızca eğitim kümesi üzerinden hesaplanmış, test kümesine sızıntı oluşmaması için aynı dönüşüm test verilerine ayrı olarak uygulanmıştır.

754 boyutlu öznitelik uzayının doğrudan kullanılması, örneklem sayısına kıyasla aşırı öğrenme riskini artırdığından, ANOVA F-testi tabanlı SelectKBest yöntemiyle sınıflar arasında istatistiksel

olarak en ayırt edici 100 öznelik seçilmiştir. Şekil 2'de, seçilen öznelikler arasında literatürde sıklıkla vurgulanan PPE (Pitch Period Entropy), DFA (Detrended Fluctuation Analysis) ve RPDE (Recurrence Period Density Entropy) ölçütlerinin PH ve sağlıklı sınıfları arasındaki dağılım farkı görselleştirilmiştir (Little ve ark., 2009).



Şekil 2. Seçilen bazı öznelikler için sınıflar arası dağılım karşılaştırması

Model değerlendirmesinde, aynı bireye ait üç kaydın eğitim ve test kümelerine dağılmasını önlemek amacıyla hasta kimliği (id) temel alınarak beş katlı grup çapraz doğrulama (GroupKFold) uygulanmıştır. Her katlamada, bir gruptaki tüm hastaların kayıtları ya tamamen eğitimde ya da tamamen testte yer almış, ölçekleme ve öznelik seçimi işlemleri her katlamada yalnızca o katlamanın eğitim verisi üzerinden yeniden hesaplanmıştır. Bu protokol, literatürde sıkça rastlanan hasta bazlı veri sızıntısı sorununu ortadan kaldırmakta ve modelin gerçekten görülmemiş bireyler üzerindeki başarımını yansıtmaktadır.

### Sınıflandırma Algoritmaları

Öznelik seçimi sonrası elde edilen 100 boyutlu temsil üzerinde, farklı öğrenme mantıklarına sahip altı denetimli sınıflandırma algoritması eğitilmiş ve karşılaştırılmıştır. Her algoritmanın performansı, beş katlı hasta bazlı çapraz doğrulamanın tüm katlamalarından elde edilen doğruluk, precision, recall ve F1-skoru (makro ortalama) değerlerinin ortalaması ve standart sapması olarak Tablo 2'de raporlanmıştır.

### Destek Vektör Makinesi (SVM)

RBF (radyal tabanlı fonksiyon) çekirdeği kullanılan SVM sınıflandırıcısında, doğrusal olarak ayrılamayan öznelik uzayı daha yüksek boyutlu bir uzaya dönüştürülerek sınıflar arasında maksimum kenar boşluğuna sahip bir hiper düzlem aranmıştır. Düzenleştirme parametresi  $C=10$  ve çekirdek katsayısı için varsayılan ölçekli (scale) ayar kullanılmıştır. Model, beş katlamanın ortalamasında %78,58 doğruluk ve %69,70 makro F1-skoru elde etmiştir.

### Rastgele Orman (Random Forest)

300 karar ağacından oluşan Rastgele Orman sınıflandırıcısı, her ağacın öznitelik ve örneklem bazında rastgele alt kümeler üzerinde eğitilmesine dayanan bir topluluk (ensemble) öğrenme yöntemidir. Ağaçlar arası varyansı azaltarak aşırı öğrenmeye karşı nispeten dirençli bir yapı sunan bu model, çalışmada denenen algoritmalar arasında en yüksek ortalama doğruluğa (%81,23) ulaşmış, ancak katlamalar arasındaki standart sapmasının ( $\pm 4,62$ ) diğer bazı modellere kıyasla daha yüksek olduğu gözlenmiştir.

### k-En Yakın Komşu (KNN)

Örnek tabanlı ve parametrik olmayan bir yöntem olan KNN sınıflandırıcısında, bir test örneği, öznitelik uzayındaki en yakın  $k=7$  komşusunun çoğunluk sınıfına atanmıştır. Yüksek boyutlu uzaylarda mesafe ölçütlerinin ayırt ediciliğini yitirmesi (boyutsallık laneti) nedeniyle, KNN modeli %77,24 doğruluk ve %65,83 F1-skoru ile diğer topluluk yöntemlerine kıyasla sınırlı bir performans göstermiştir.

### Çok Katmanlı Algılayıcı (MLP)

İki gizli katmandan (128 ve 64 nöron) oluşan ileri beslemeli bir yapay sinir ağı olan MLP sınıflandırıcısı, erken durdurma (early stopping) mekanizmasıyla aşırı öğrenmeye karşı düzenleniltilerle en fazla 1000 iterasyon boyunca eğitilmiştir. Model, %78,96 ortalama doğruluk elde etmiş; ancak katlamalar arası standart sapmasının ( $\pm 6,70$  doğruluk,  $\pm 10,56$  F1-skoru) en yüksek değerlere sahip model olması, sınırlı örneklem büyüklüğünde sinir ağı tabanlı yaklaşımların kararlılığının daha düşük olabileceğine işaret etmektedir.

### Gradient Boosting

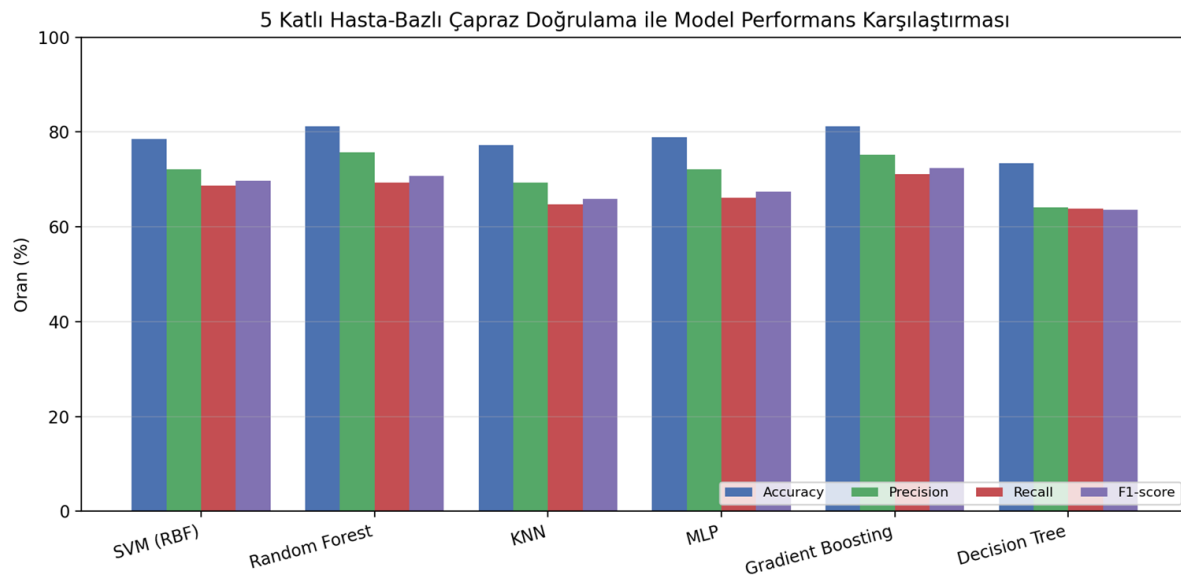
200 zayıf ağaç tahmincisinin ardışık biçimde önceki modellerin hatalarını azaltacak şekilde eklendiği Gradient Boosting algoritması, çalışmada en dengeli sonucu üretmiştir. Model, %81,22 ortalama doğruluk ve tüm sınıflandırıcılar arasında en yüksek makro F1-skoru olan %72,37 değerine ulaşmıştır. Şekil 3'te, altı sınıflandırıcının doğruluk, precision, recall ve F1-skoru bakımından karşılaştırması sunulmuştur; Şekil 4'te ise Gradient Boosting modeline ait beş katlamanın toplamından elde edilen karışıklık matrisi gösterilmektedir.

**Tablo 2. Sınıflandırıcıların 5 katlı hasta-bazlı çapraz doğrulama performans karşılaştırması (ortalama  $\pm$  standart sapma, %)**

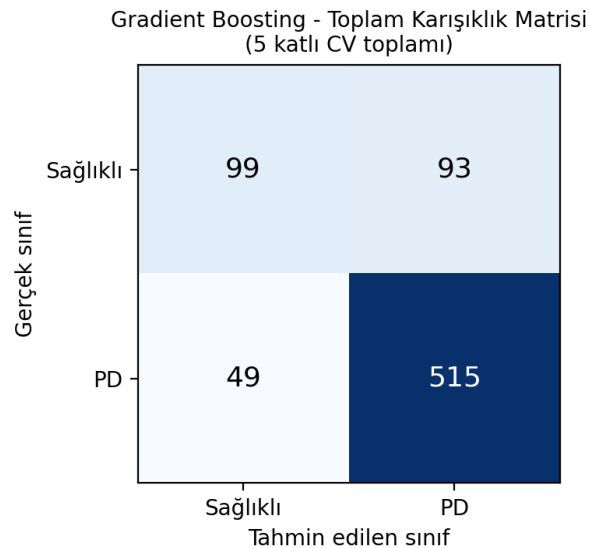
| Sınıflandırıcı     | Doğruluk (%)     | Precision (%)    | Recall (%)       | F1-score (%)     |
|--------------------|------------------|------------------|------------------|------------------|
| SVM (RBF çekirdek) | 78,58 $\pm$ 2,71 | 72,12 $\pm$ 6,01 | 68,68 $\pm$ 3,33 | 69,70 $\pm$ 4,05 |

|                         |              |               |              |               |
|-------------------------|--------------|---------------|--------------|---------------|
| Rastgele Orman          | 81,23 ± 4,62 | 75,71 ± 7,33  | 69,40 ± 8,79 | 70,75 ± 9,22  |
| k-En Yakın Komşu        | 77,24 ± 3,59 | 69,37 ± 6,66  | 64,76 ± 5,91 | 65,83 ± 6,51  |
| Çok Katmanlı Algılayıcı | 78,96 ± 6,70 | 72,15 ± 10,89 | 66,16 ± 9,51 | 67,39 ± 10,56 |
| Gradient Boosting       | 81,22 ± 3,85 | 75,21 ± 6,73  | 71,17 ± 7,47 | 72,37 ± 7,49  |
| Karar Ağacı             | 73,42 ± 4,57 | 64,10 ± 6,92  | 63,85 ± 8,04 | 63,64 ± 7,54  |

Precision, Recall ve F1-score değerleri "macro average" olarak hesaplanmıştır.



Şekil 3. Altı sınıflandırıcının performans metrikleri bakımından karşılaştırması



**Şekil 4. Gradient Boosting modeli için beş katlamanın toplamından elde edilen karışıklık matrisi****Karar Ağacı**

Maksimum derinliği 10 ile sınırlandırılan tek bir Karar Ağacı sınıflandırıcısı, yorumlanabilirliği yüksek olmakla birlikte, topluluk yöntemlerinin sağladığı varyans azaltıcı etkiden yoksun olması nedeniyle çalışmadaki en düşük performansı (%73,42 doğruluk, %63,64 F1-skoru) sergilemiştir. Bu sonuç, tek bir ağaca dayalı modellerin yüksek boyutlu ve göreceli olarak dengesiz veri setlerinde topluluk tabanlı alternatiflere kıyasla daha kırılgan kaldığını doğrulamaktadır.

**Değerlendirme Metrikleri**

Model performansları; doğruluk (accuracy), kesinlik (precision), duyarlılık (recall) ve F1-skoru olmak üzere dört temel metrik ile değerlendirilmiştir. Veri setindeki sınıf dengesizliği göz önünde bulundurularak precision, recall ve F1-skoru, iki sınıfın eşit ağırlıklandırıldığı "macro average" yöntemiyle hesaplanmıştır. Tüm metrikler, beş katlı hasta-bazlı çapraz doğrulamanın her bir katlamasında ayrı ayrı hesaplanmış ve nihai olarak ortalama ile standart sapma değerleri raporlanmıştır.

**SONUÇLAR ve TARTIŞMA**

Bu çalışmada, UCI Parkinson's Disease Classification veri seti üzerinde altı klasik makine öğrenmesi sınıflandırıcısının performansı, hasta bazlı beş katlı çapraz doğrulama protokolü altında karşılaştırmalı olarak incelenmiştir. Elde edilen sonuçlar, topluluk (ensemble) tabanlı yöntemlerin -özellikle Gradient Boosting ve Rastgele Orman'ın- tekil sınıflandırıcılara (Karar Ağacı) ve örnek tabanlı yöntem (KNN) kıyasla daha yüksek ve daha kararlı performans sergilediğini göstermiştir. Gradient Boosting, %81,22 doğruluk ve %72,37 makro F1-skoru ile en dengeli sonucu üretirken, Rastgele Orman %81,23 ile en yüksek ortalama doğruluğa ulaşmıştır.

Bu çalışmada elde edilen doğruluk değerleri (~%77-81), aynı veri setini kullanan ve TQWT tabanlı öznelik mühendisliği ile topluluk öğrenmesini birleştiren Sakar ve ark.'nın (2019) çalışmasında bildirilen yaklaşık %85 doğruluk oranının ve Hossain & Amenta'nın (2024) 10 katlı rastgele çapraz doğrulamayla bildirdiği %84-85 aralığındaki sonuçların bir miktar altında kalmıştır. Bu farkın temel nedeninin, mevcut çalışmada uygulanan hasta-bazlı grup ayrımı olduğu değerlendirilmektedir; zira aynı bireye ait üç kaydın rastgele bölünerek eğitim ve test kümelerine dağıtıldığı yaklaşımlarda, model dolaylı olarak bireyin kendine özgü ses karakteristiğini öğrenebilmekte ve bu durum test başarımını yapay biçimde yükseltebilmektedir. Sunulan çalışmada bu sızıntı türü metodolojik olarak engellenmiş olup, raporlanan performans değerlerinin klinik açıdan daha temkinli ve gerçekçi bir tahmin sunduğu düşünülmektedir.

Sınıflar arası dengesizlik (564 PH kaydına karşılık 192 sağlıklı kayıt) nedeniyle, tüm modellerde recall ve F1-skoru değerlerinin doğruluk değerinin bir miktar altında kaldığı görülmüştür; bu

durum, modellerin özellikle azınlık sınıfı olan sağlıklı bireyleri PH olarak yanlış sınıflandırma eğiliminde olduğuna işaret etmektedir. Klinik bir tarama aracı için bu tür yanlış pozitiflerin, yanlış negatiflere kıyasla görece olarak daha az riskli kabul edilebileceği düşünülse de, ileride SMOTE gibi aşırı örnekleme tekniklerinin ya da sınıf ağırlıklandırma kayıp fonksiyonlarının bu dengesizliği azaltmak amacıyla değerlendirilmesi önerilmektedir.

Çalışmanın bazı sınırlılıkları da bulunmaktadır. Öznitelik seçiminde kullanılan ANOVA F-testi, öznitelikler arasındaki doğrusal olmayan etkileşimleri dikkate almayan tek değişkenli bir yöntemdir; bu nedenle ileride ilişkili öznitelik gruplarını (örn. TQWT alt bantları) birlikte değerlendiren çok değişkenli seçim yöntemlerinin ya da derin öğrenme tabanlı uçtan uca temsil öğreniminin denenmesi, performansı artırabilir. Ayrıca yalnızca sürdürülen /a/ ünlüsünden elde edilen kayıtların kullanılması, bulguların bağlantılı konuşma veya farklı ünlü seslerine genellenebilirliğini sınırlamaktadır.

Sonuç olarak, bu çalışma, ses sinyali özniteliklerinin Parkinson hastalığı sınıflandırmasında klinik açıdan anlamlı bilgi taşıdığını ve topluluk tabanlı makine öğrenmesi algoritmalarının bu görevde tekil modellere kıyasla daha üstün ve kararlı sonuçlar verdiğini doğrulamaktadır. Aynı zamanda, hasta-bazlı doğrulama protokolünün, literatürdeki iyimser performans tahminlerine karşı daha gerçekçi bir alternatif sunduğu ve gelecekteki çalışmalarda standart bir değerlendirme uygulaması olarak benimsenmesi gerektiği vurgulanmaktadır.

## KAYNAKÇA

- Bel, M., Kutlu, Y., Gürsoy-Demir, H. (2025). Comparative Analysis of Wavelet Types for Feature Extraction in Arrhythmia Classification. *Journal of Artificial Intelligence with Applications*, 6(1), 8-13, doi : 10.5281/zenodo.18105152
- Camgözlü Y, Kutlu Y.(2021). Detection of coronavirus disease (COVID-19) from X-ray images using deep convolutional neural networks. *Natural and Engineering Sciences*, 6(3), 204-211., Doi: 10.28978/nesciences.868087 (Yayın No: 7515288)
- Daşcı, R. (2025). Mental Health Care Seeking Prediction with Machine Learning: An Analysis Based on Lifestyle, Stress, and Demographic Factors. *Health Science and Green Technology*, 1(1), 31-43, doi : 10.5281/zenodo.18649982
- Hossain, M. A., & Amenta, F. (2024). Machine learning-based classification of Parkinson's disease patients using speech biomarkers. *Journal of Parkinson's Disease*, 14(1), 95–109.
- Jankovic, J. (2008). Parkinson's disease: clinical features and diagnosis. *Journal of Neurology, Neurosurgery & Psychiatry*, 79(4), 368–376.
- Kaderoglu, M., Satır, E. (2025). Histopathological Breast Cancer Classification Using a Hybrid Swin Transformer and ConvNeXt Architecture. *Journal of Artificial Intelligence with Applications*, 6(1), 18-25, doi : 10.5281/zenodo.18138734
- Little, M. A., McSharry, P. E., Hunter, E. J., Spielman, J., & Ramig, L. O. (2009). Suitability of dysphonia measurements for telemonitoring of Parkinson's disease. *IEEE Transactions on Biomedical Engineering*, 56(4), 1015–1022.

- Tohma, S. (2025). Classification of Arrhythmia and Cardiac Anomalies from ECG Signals: A Machine Learning-Based Comparison on MIT-BIH and PTB-DB. *Journal of Natural and Engineering Research*, 1(1), 19-29, doi : 10.5281/zenodo.18664503
- Sakar, B. E., Isenkul, M. E., Sakar, C. O., Sertbas, A., Gorgen, F., Delil, S., Apaydin, H., & 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.
- Sakar, C. O., Serbes, G., Gunduz, A., Nizam, H., & Sakar, B. (2018). Parkinson's Disease Classification [Dataset]. UCI Machine Learning Repository. <https://doi.org/10.24432/C5MS4X>
- Sakar, C. O., Serbes, G., Gunduz, A., Tunc, H. C., Nizam, H., Sakar, B. E., Tutuncu, M., Aydin, T., Isenkul, M. 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.
- 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.