T.C.

YILDIZ TEKNİK ÜNİVERSİTESİ

FEN BİLİMLERİ ENSTİTÜSÜ

YAPAY ZEKÂ TEKNİKLERİ KULLANILARAK PARKİNSON HASTALIĞININ TEŞHİSİ

Ebru Gizem YILMAZ

YÜKSEK LİSANS TEZİ

Veri Bilimi ve Büyük Veri Anabilim Dalı

Veri Bilimi ve Büyük Veri Programı

Danışman

Prof. Dr. Gülhayat GÖLBAŞI ŞİMŞEK

Ekim, 2025

T.C.

YILDIZ TEKNİK ÜNİVERSİTESİ

FEN BİLİMLERİ ENSTİTÜSÜ

YAPAY ZEKÂ TEKNİKLERİ KULLANILARAK PARKİNSON HASTALIĞININ TEŞHİSİ

Ebru Gizem YILMAZ tarafından hazırlanan tez çalışması 31.10.2025 tarihinde aşağıdaki jüri tarafından Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü Veri Bilimi ve Büyük Veri Anabilim Dalı, Veri Bilimi ve Büyük Veri Tezli Yüksek Lisans Programı **YÜKSEK LİSANS TEZİ** olarak kabul edilmiştir.

Prof. Dr. Gülhayat GÖLBAŞI ŞİMŞEK
Yıldız Teknik Üniversitesi
Danışman

Jüri Üyeleri

Prof. Dr. Gülhayat GÖLBAŞI ŞİMŞEK, Danışman
Yıldız Teknik Üniversitesi _____

Prof. Dr. Fatma NOYAN TEKELİ, Üye
Yıldız Teknik Üniversitesi _____

Dr. Öğr. Üyesi Zehra CİVAN, Üye
Fenerbahçe Üniversitesi _____

Danışmanım Prof. Dr. Gülhayat GÖLBAŞI ŞİMŞEK sorumluluğunda tarafımca hazırlanan "Yapay Zekâ Teknikleri Kullanılarak Parkinson Hastalığının Teşhisi" başlıklı çalışmada veri toplama ve veri kullanımında gerekli yasal izinleri aldığımı, diğer kaynaklardan aldığım bilgileri ana metin ve referanslarda eksiksiz gösterdiğimi, araştırma verilerine ve sonuçlarına ilişkin çarpıtma ve/veya sahtecilik yapmadığımı, çalışmam süresince bilimsel araştırma ve etik ilkelerine uygun davrandığımı beyan ederim. Beyanımın aksinin ispatı halinde her türlü yasal sonucu kabul ederim.

Ebru Gizem YILMAZ

İmza

Bu çalışma, Yıldız Teknik Üniversitesi Bilimsel Araştırma Proje Koordinatörlüğü'nün FYL – 2024 – 6458 numaralı projesi ile desteklenmiştir.



Aileme

TEŞEKKÜR

Bu tez sürecinde bana destek olan tüm değerli insanlara en içten teşekkürlerimi sunarım.

Öncelikle, akademik yolculuğum boyunca bana rehberlik eden ve bilgi birikimiyle ufkumu genişleten danışmanım Gülhayat GÖLBAŞI ŞİMŞEK'e minnettarım. Yönlendirmeleri, sabrı ve desteği olmadan bu çalışma tamamlanamazdı.

Erasmus+ programı kapsamında bulunduğum Almanya Münih Teknik Üniversitesi'nde, tez sürecime akademik katkı sunan ve danışmanlık desteğiyle çalışmamın gelişimine önemli katkılarda bulunan Prof. Dr. Martin WERNER'e teşekkür ederim.

Ailem ve arkadaşlarım, her an yanımda olarak moral ve motivasyon kaynağım oldular. Varlıkları, bu sürecin zorluklarını aşmamda büyük rol oynadı.

Ayrıca, araştırmam sırasında bana yardımcı olan akademisyenlere ve diğer tüm kaynak sağlayıcılara teşekkür ederim. Paylaştıkları bilgiler ve gösterdikleri kolaylık, çalışmamın daha verimli ilerlemesini sağladı.

Son olarak, yüksek lisans sürecimde karşılaştığım tüm zorluklara rağmen hiç yılmadan devam etme cesaretini kendimde bulmamı sağlayan herkese teşekkür ediyorum.

Ebru Gizem YILMAZ

İÇİNDEKİLER

SİMGE LİSTESİ viii

KISALTMA LİSTESİ ix

ŞEKİL LİSTESİ x

TABLO LİSTESİ xi

ÖZET xii

ABSTRACT xiv

1 GİRİŞ 1

1.1 Parkinson Hastalığına Genel Bakış 1

1.1.1 Parkinson Hastalığının Belirtileri ve Etkileri 1

1.2 Tezin Kapsamı 3

1.3 Tezin Organizasyonu 3

2 LİTERATÜR TARAMASI 4

2.1 Parkinson Hastalığı Teşhisinde Yapay Zekâ Kullanımı 4

3 YÖNTEM 9

3.1 Araştırma Yöntemi 9

3.2 Veri Seti ve Kaynağı 10

3.3 Özellik Çıkarımı ve Uygulama 11

3.3.1 Permütasyon Entropi 12

3.3.2 Negentropi 16

3.3.3 Adım/Dakika Hesaplama 19

3.3.4 Neurokit-Toplam Ayak Kuvvetleri Filtreleme 20

3.3.5 Sağ - Sol Ayak Denge Farkı..……..…...22

3.3.6 Gini Değeri ……..……..…28

3.4 Makine Öğrenmesi Modelleri 31

3.4.1 Destek Vektör Makineleri ……31

3.4.2 Yapay Sinir Ağları…….....34

3.4.3 Lojistik Regresyon ……...37

3.5 Değerlendirme Ölçütleri 40

4 UYGULAMA VE BULGULAR 43

4.1 Model Eğitimi ve Test Süreci 43

4.2 Deneysel Sonuçlar.....44

4.3 Modellerin Karşılaştırılması 50

5 SONUÇ 52

5.1 Genel Sonuçlar.....52

KAYNAKÇA 53

TEZDEN ÜRETİLMİŞ YAYINLAR 56

SİMGE LİSTESİ

w	Ağırlık vektörü
$f(.)$	Aktivasyon fonksiyonu
p	Anlamlılık düzeyi
$N_{Anomali}$	Belirli bir eşik değerini aşan anomalilerin toplam sayısı
b	Bias terimi
y	Çıktı değeri
I	Düzensizlik skoru
G	Gini değeri
x	Girdi özellikleri vektörü
Hz	Hertz – frekans birimi
H&Y	Hoehn & Yahr
D	İki ayak arasındaki denge farkı
X	İlgili veri noktası
F	Kuvvet
log	Logaritma
d	Marj genişliği
min	Minimum
S_n	n-elemanlı tüm olası permütasyonların kümesi
J(y)	Negentropi fonksiyonu
μ	Ortalama
$S_{Anomali}$	Ortalama anomali şiddeti
n	Örneklem büyüklüğü
$p(p)$	Permütasyon gözlenme olasılığı
H_n	Shannon entropi
Z	Standartlaştırılmış skor (Z-puanı)
t	t istatistiği
V	Varyans değeri
N	Veri kümesindeki toplam nokta sayısı

KISALTMA LİSTESİ

BPHDÖ	Birleşik Parkinson Hastalığı Derecelendirme Ölçeği
DVM	Destek Vektör Makineleri
DYTK	Dikey Yer Tepkisi Kuvveti
KG	Kontrol Grubu
KT-YİD	Komşuluk Temelli Yerel İkili Desen
PH	Parkinson Hastalığı
PHG	Parkinson Hastası Grubu
SHAP	SHapley Additive exPlanations
YSA	Yapay Sinir Ağı

ŞEKİL LİSTESİ

Şekil 3.1 Ayak tabanı sensör konumları **11**
Şekil 3.2 Gruplar arası permütasyon entropisi karşılaştırması boxplot grafiği **15**
Şekil 3.3 Sol ve sağ ayak için negentropi dağılımlarının boxplot grafiği **17**
Şekil 3.4 Ham ve filtrelenmiş ayak kuvveti verileri **21**
Şekil 3.5 Sağ – sol ayak denge farkı anomali sayısı **24**
Şekil 3.6 Sağ – sol ayak denge farkı anomali şiddeti **25**
Şekil 3.7 Sağ – sol ayak denge farkı varyansı **26**
Şekil 3.8 Sağ – sol ayak denge farkı düzensizlik skoru **26**
Şekil 3.9 Destek vektör makineleri optimal hiperdüzlem ve marjlar **32**
Şekil 3.10 Biyolojik ve yapay nöron karşılaştırması **34**
Şekil 3.11 Yapay sinir ağları katmanları **35**
Şekil 3.12 Lojistik Fonksiyon Eğrisi **38**
Şekil 4.1 DVM özniteliklerin tahmin üzerindeki etkisi **45**
Şekil 4.2 YSA özniteliklerin tahmin üzerindeki etkisi **47**
Şekil 4.3 Lojistik regresyon özniteliklerin tahmin üzerindeki etkisi **49**

TABLO LİSTESİ

Tablo 2.1 Literatür özeti **8**
Tablo 3.1 Katılımcıların demografik özellikleri **10**
Tablo 3.2 Ayak tabanı sensör konumları **10**
Tablo 3.3 Sol ve sağ ayak permütasyon entropisi istatiksel değerleri **15**
Tablo 3.4 Permütasyon entropisi Welch t-testi **16**
Tablo 3.5 Negentropi istatiksel özeti **18**
Tablo 3.6 Negentropi t-testi değerleri **18**
Tablo 3.7 Adım/dakika karşılaştırması **20**
Tablo 3.8 Adım/dakika t-testi sonuçları **20**
Tablo 3.9 Zirve ve çukur noktalarına ait t-testi sonuçları **22**
Tablo 3.10 Sağ – sol ayak denge farkı metrikleri t-testi sonuçları **27**
Tablo 3.11 Gini değerleri **29**
Tablo 3.12 KG ve PHG gini indeksi değerleri t-testi sonuçları **30**
Tablo 3.13 Karışıklık matrisi **40**
Tablo 4.1 Destek vektör makineleri sınıflandırma sonuçları **44**
Tablo 4.2 DVM modeli öznitelik etkileri **46**
Tablo 4.3 Yapay sinir ağları sınıflandırma sonuçlar **46**
Tablo 4.4 YSA modeli öznitelik etkileri **48**
Tablo 4.5 Lojistik regresyon sınıflandırma sonuçları **48**
Tablo 4.6 Lojistik regresyon modeli öznitelik etkileri **50**

ÖZET

Yapay Zekâ Teknikleri Kullanılarak Parkinson Hastalığının Teşhisi

Ebru Gizem YILMAZ

Veri Bilimi ve Büyük Veri Anabilim Dalı
Yüksek Lisans Tezi

Danışman: Prof. Dr. Gülhayat GÖLBAŞI ŞİMŞEK

Bu tez, Parkinson hastalığı (PH) teşhisi için yapay zekâ tekniklerinin kullanılmasını araştırmaktadır. PH, Alzheimer hastalığından sonra en sık görülen nörodejeneratif hastalıktır.

Teşhis genellikle subjektif klinik değerlendirmelere dayanır ve bu durum, teşhisin doğruluğunu ve zamanlamasını etkileyebilir. Bu tezin amacı, hastalığın erken teşhisini sağlamak için yürüyüş sinyal verilerinden daha objektif bir yöntem sunmaktır. Bu amaçla, yapay zekâ modelleri kullanılacaktır. Bu modeller, hastaların yürüyüş özelliklerini analiz ederek Parkinson hastalığının belirtilerini tespit edecektir.

Bu çalışma, farklı yapay zekâ modellerinin performanslarını karşılaştırarak en uygun modeli seçmeyi hedeflemektedir. Aynı zamanda ayrıntılı ve kapsamlı tek bir veri seti üzerinde yöntemlerin bütüncül bir şekilde incelenmesi ile yapay zekâ tekniklerinin Parkinson hastalığının teşhisindeki etkinliğini artırmayı hedeflemektedir.

Yürüyüş verilerinin yapay zekâ teknikleri kullanılarak analizi, hastalığın erken teşhisindeki hataları azaltarak hastaların yaşam kalitesini arttırma potansiyeline

sahiptir. Ayrıca, bu tez, yapay zekânın tıbbi teşhislerde nasıl kullanılabileceğine dair değerli bilgiler sunabilir.

Anahtar Kelimeler: Parkinson hastalığı, yapay zekâ, teşhis, yürüyüş analizi, makine öğrenmesi

ABSTRACT

Diagnosis of Parkinson Disease Using Artificial Intelligence Techniques

Ebru Gizem YILMAZ

Department of Data Science and Big Data
Master of Science Thesis

Supervisor: Prof. Dr. Gülhayat GÖLBAŞI ŞİMŞEK

This thesis explores the use of artificial intelligence techniques for diagnosing Parkinson's disease. Parkinson’s Disease is the second most common neurodegenerative disease after Alzheimer's.

Diagnosis is often based on subjective clinical evaluations, which can affect the accuracy and timing of diagnosis. The aim of this thesis is to provide a more objective method using gait signal data for early diagnosis of the disease. For this purpose, artificial intelligence models will be used. These models will analyze patients' gait characteristics to detect symptoms of Parkinson's disease.

This study aims to compare the performance of various artificial intelligence models to select the most suitable one. Additionally, by conducting a comprehensive analysis on a single detailed dataset, it seeks to enhance the effectiveness of artificial intelligence techniques in diagnosing Parkinson’s disease.

The analysis of gait data using artificial intelligence techniques has the potential to reduce errors in early diagnosis and improve the quality of life for patients. Furthermore, this thesis can provide valuable insights into how artificial intelligence can be utilized in medical diagnostics.

Keywords: Parkinson's disease, artificial intelligence, diagnosis, gait analysis, machine learning

1
GİRİŞ

1.1 Parkinson Hastalığına Genel Bakış

Hücre kaybı ile sinsi başlayan ve yavaş seyreden hastalıklara nörodejeneratif hastalıklar denilmektedir. PH beyinde dopamin içeren nigrostriatal nöronları içerecek şekilde yaygın nörodejenerasyonun bulunduğu, Alzheimer hastalığından sonra en sık görülen nörodejeneratif hastalıktır. Parkinson hastalığının motor bulguları beyinde hareketlerimizden sorumlu olan hücrelerin ufak bir bölümünün hasara uğraması ve eksilmesi (dejenerasyon) sonucu ortaya çıkar. Bu hücreler bilgileri bir sinir hücresinden diğerine gönderen dopamin adı verilen kimyasal bir madde salgılar. Beyinde yeterli dopamin yapılamazsa hareket ve duruş işlevleri etkilenerek Parkinson hastalığı belirtileri ortaya çıkar. Dopamin eksikliğinin temel belirtileri hareketlerde yavaşlama, hareket yeteneğinin azalması ve titremedir. Hastalık yavaş bir şekilde ilerler. Hastadan hastaya belirtilerin varlığı, şiddeti ve hastalığın ilerleme hızı farklıdır [1].

PH ilk olarak 1817 yılında James Parkinson tarafından detaylı olarak tanımlanmıştır. Yaşlanma, PH'nin önde gelen nedenlerinden biridir ve prevalansı yaşla birlikte artar. Prevalans çalışmaları, 65 yaşın üstündeki nüfusun yaklaşık %1'inin ve 80'li yaşlarda %3-4'ünün bu hastalıktan etkilendiğini göstermektedir [2]. Ayrıca, araştırmacılar, çevresel faktörler ve PH arasındaki ilişkileri araştırmak için çeşitli çalışmalar yapmışlardır, burada PH'nin insidans oranının sigara içme, kafein alımı ve diğer faktörlerle ilişkili olduğu kanıtlanmıştır. Ancak, çevresel faktörlerin PH üzerindeki etkisi, bileşen faktörlerin uzun vadeli etkisi nedeniyle net bir şekilde belirlenememiştir. PH için kesin tanısal biyobelirteçler bulunamadığından, erken belirtiler ve klinik muayeneler ana tanısal ölçütlerdir [3].

PH olan hastaların motor bozukluklarını değerlendirmek için çeşitli ölçekler kullanılır, ancak bu ölçeklerin çoğu geçerlilik ve güvenilirlik açısından tam olarak değerlendirilmemiştir. Hoehn ve Yahr ölçeği, hastaların gruplarını karşılaştırmak ve hastalığın ilerlemesini değerlendirmek için yaygın olarak kullanılır. Bu ölçek, evre 0'dan (hastalığın hiçbir belirtisi yok) evre 5'e (tekerlekli sandalyeye bağlı veya yardım olmadan yatağa bağımlı) kadar değişir. Birleşik Parkinson Hastalığı Derecelendirme Ölçeği (BPHDÖ) ise motor bozulmanın değerlendirilmesi için başka bir ölçektir. BPHDÖ ile yapılan çalışmalar, PH'nin seyrinin doğrusal olmadığını göstermiştir. Ayrıca, bu çalışmalar hastalığın erken evresinde ve postural dengesizlik yürüyüş zorluğu olan PH hastalarında bozulma hızının daha hızlı olduğunu öne sürmektedir [4].

1.1.1 Parkinson Hastalığının Belirtileri ve Etkileri

Parkinson hastalığının en erken belirtileri enterik sinir sistemi, alt beyin sapı ve koku yollarında ortaya çıkmaktadır. Koku duyusu kaybı veya azalması, uyku bozuklukları ve kabızlık, sonraki aşamada ise titreme ve hareketlerde yavaşlama gibi motor belirtiler görülmektedir. Hastalık genellikle motor semptomlarla tanı almaktadır [1]. Başlıca klinik belirtileri; istirahat tremoru, bradikinezi, rijidite ve ilerleyen yıllarda ortaya çıkan postüral refleks bozukluğudur. Erkeklerde görülme sıklığı kadınlara oranla biraz daha fazladır [2].

Yürüyüş ve koşu paternleri, karmaşık beyin ağları tarafından düzenlenebilir. Bu ağlar, motor ve bilişsel fonksiyonlardan sorumlu kortikal bölgeleri içerir. Yürüme bozuklukları ve anormallikleri, PH'nin temel semptomlarındandır. Geçen on yılda, yürüme bozuklukları ve yürüme anormalliklerinin analiz edilmesiyle nicel bir araç haline gelmiştir. Hareket semptomları açısından, yürüme bozukluklarına ve anormalliklerine yol açan üç ana yön vardır. 1) Titreme: titreme genellikle ellerde veya uzuvlarda başlar ve genellikle dinlenirken daha sık görülür. 2) Hareket yavaşlığı: hastalar, sağlıklı bireylerle karşılaştırıldığında azalmış yürüme hızı ve adım uzunluğunu gösterir. 3) Kas sertliği: kasların yüksek gerilimi, hastaların postüründe artan rijitliğe neden olur, bu da insan yürüyüşü sırasında istikrarı etkileyebilir. Ayrıca, hareket dışı semptomlar (örneğin, biliş bozukluğu, depresyon, anksiyete) de anormal yürüme paternlerine katkıda bulunur. Gelişmiş algılama teknolojileri tarafından desteklenen yürüme analizi, klinik laboratuvar çalışmalarından günlük yaşam

ortamlarına kadar gerçekleştirilebilir, bu da yürüme temelli PH tespiti, izlenmesi ve müdahaleleri için fırsatlar sunar. [3].

1.2 Tezin Kapsamı

Bu tez, Parkinson hastalığının erken teşhisinde kullanılabilecek makine öğrenmesi temelli bir sistemin geliştirilmesini kapsamaktadır. Çalışma kapsamında, açık kaynaklı bir veri seti kullanılarak bireylerin yürüme verileri analiz edilmiş ve bu verilerden çeşitli özellikler çıkarılmıştır. Özellikle permütasyon entropisi, negentropi, adım sayısı ve Gini değeri gibi özellikler kullanılarak hastalık belirtilerinin belirlenmesi amaçlanmıştır.

Geliştirilen sistemde destek vektör makineleri, yapay sinir ağları ve lojistik regresyon gibi farklı makine öğrenmesi modelleri karşılaştırmalı olarak kullanılmış; bu modellerin sınıflandırma başarıları değerlendirilmiştir. Tezin kapsamı, yalnızca tanısal süreç ile sınırlıdır. Tedavi önerileri veya hastalığın ilerleyişinin modellenmesi bu çalışmanın kapsamı dışında bırakılmıştır.

Bu kapsamda, çalışmanın amacı Parkinson hastalığının teşhisinde kullanılabilecek yüksek doğruluklu ve yorumlanabilir bir model geliştirmek ve çeşitli algoritmaların bu konudaki performanslarını karşılaştırmaktır.

1.3 Tezin Organizasyonu

Bu tez beş ana bölümden oluşmaktadır. Birinci bölümde, Parkinson hastalığına genel bir bakış sunulmakta ve tez konusu ile ilgili temel bilgiler verilerek araştırmanın kapsamı tanımlanmaktadır. İkinci bölümde, literatürde yer alan yapay zekâ temelli teşhis yaklaşımları incelenmiş ve önceki çalışmalarla karşılaştırmalar yapılmıştır. Üçüncü bölümde, araştırma yöntemi detaylandırılmış; kullanılan veri seti, özellik çıkarım teknikleri ve uygulanan makine öğrenmesi algoritmaları açıklanmıştır. Dördüncü bölümde, deneysel uygulamalara ve elde edilen bulgulara yer verilmiş; farklı modellerin performansları karşılaştırılmıştır. Son bölümde ise genel sonuçlar değerlendirilmiş ve gelecekteki çalışmalara yönelik öneriler sunulmuştur.

2
LİTERATÜR TARAMASI

2.1 Parkinson Hastalığı Teşhisinde Yapay Zekâ Kullanımı

Parkinson hastalığının erken teşhis edilmesi, hastaların yaşam kalitesini artırabilir ve tedavi sürecini iyileştirebilir. Bu bağlamda, yapay zekâ tekniklerinin kullanımı, yürüyüş özelliklerinin analiziyle PH teşhisinde önemli bir potansiyele sahiptir. Son yıllarda, yürüyüş verilerinin analizi yoluyla Parkinson hastalığının tanısına yönelik çalışmalar artmış; bu alanda yapay zekâ tekniklerinin kullanımı dikkat çekici bir ivme kazanmıştır. Yürüyüş örüntüleri, hastalığın motor semptomlarını yansıtan zengin biyomekanik bilgiler içerdiğinden, bu verilerin sınıflandırma algoritmalarıyla işlenmesi, tanı doğruluğunu artıran etkili modellerin geliştirilmesine olanak tanımaktadır. Özellikle sensör teknolojilerinin ilerlemesiyle birlikte, bireylerin adım süresi, salınım ve destek fazları, topuk kuvveti gibi parametreleri yüksek hassasiyetle ölçülebilmekte; bu veriler üzerinden anlamlı öznitelikler çıkarılarak sınıflandırma algoritmalarına beslenmektedir.

Ulusal ve uluslararası literatürde, yürüyüş verileri üzerinden yapay zekâ algoritmalarının kullanıldığı ve Parkinson hastalığının teşhisine yönelik başarılı sonuçlar elde edildiği çeşitli çalışmalar yer almaktadır.

Wu ve Krishnan (2009) [5] tarafından gerçekleştirilen bir çalışmada, Parkinson hastalarının yürüyüş değişkenliği adım süresi verileri üzerinden incelenmiştir. Adım süresi ve alt fazları olan salınım ve destek sürelerinin olasılık yoğunluk fonksiyonları Parzen window yöntemiyle tahmin edilmiş, ardından bu verilere ait istatistiksel öznitelikler elde edilmiştir. Leave-one-out çapraz doğrulama yöntemi kullanılmış ve Destek Vektör Makineleri (DVM) algoritması ile %90,32 doğruluk oranı elde edilmiştir.

Lee ve Lim (2012) [6], idiopatik Parkinson hastalarının yürüyüş verilerinden elde edilen wavelet (dalgacık) tabanlı öznitelikleri kullanarak Ağırlıklı Bulanık Üyelik Fonksiyonlu Sinir Ağı algoritmasıyla sınıflandırma gerçekleştirmiştir. Çalışmada uygulanan üç farklı ön işleme yönteminden elde edilen sinyallerin wavelet dönüşümüne tabi tutulmasıyla, toplamda 40 istatistiksel öznitelik çıkarılmıştır. Bu öznitelikler, Ağırlıklı Bulanık Üyelik Fonksiyonlu Sinir Ağı ile beslenmiş; özellikle üçüncü yöntemde %77,33 doğruluk, %81,10 sensitivite ve %65,48 spesifiklik değerlerine ulaşılmıştır.

Perumal ve Sankar (2016) [7], Parkinson hastalığının erken teşhisinde yürüyüş ve titreme özelliklerinin birlikte değerlendirilmesini amaçlamış ve bu doğrultuda giyilebilir sensörlerden elde edilen veriler üzerinde analizler gerçekleştirmiştir. Çalışmada, adım mesafesi, destek ve salınım fazları, topuk kuvveti gibi yürüyüş parametreleri ile titreme sinyallerine ait frekans-temelli öznitelikler kullanılmıştır. Elde edilen veriler, istatistiksel analizler ve makine öğrenmesi algoritmaları ile değerlendirilmiştir. Lineer Diskriminant Analiz ile %86,9 doğruluk oranı elde etmiştir.

Yurdakul vd. (2020) [8], Parkinson hastalığının tanısında yürüyüş sırasında ayaklara etki eden Dikey Yer Tepkisi Kuvveti (DYTK) sinyallerinden öznitelik çıkarmak amacıyla Komşuluk Temelli Yerel İkili Desen (KT-YİD) algoritmasını kullanmışlardır. Geleneksel YİD algoritmalarına kıyasla sinyalin zamansal komşuluk ilişkilerini daha etkin biçimde kodlayan bu yöntem ile elde edilen öznitelikler üzerinde Student’s t-testi uygulanmış ve anlamlı özellik setleri oluşturulmuştur. Sınıflandırma aşamasında basit yapılı bir yapay sinir ağı kullanılmış; model %98,3 doğruluk ile yüksek performans sergilemiştir.

Wahid vd. (2015) [9], Parkinson hastalığını sınıflandırmak amacıyla uzaysal-zamansal yürüyüş parametrelerini bireylerin yaş, boy, kilo, cinsiyet ve yürüme hızı gibi fiziksel özelliklerinden arındırmak için çoklu regresyon normalleştirme stratejisi uygulamışlardır. Çalışmada normalize edilmiş veriler üzerinde Karar Ağacı algoritmasıyla %92,6 doğruluk elde edilmiş, ham verilerde ise DVM ile %80,4 doğruluk elde edilmiştir.

Alam vd. (2017) [10], yürüyüş sırasında elde edilen dikey yer tepki kuvveti sinyallerinden çıkartılan öznitelikler ile Parkinson hastalığını sınıflandırmak

amacıyla sıralı ileri özellik seçimi (Sequential Forward Selection) yöntemi uygulamıştır. Seçilen öznitelikler arasında stride süresi varyasyonu (adım süresi değişkenliği), swing süresi varyasyonu (salınım süresi değişkenliği) ve Center of Pressure (basınç merkezi) koordinatları yer almıştır. Sınıflandırma işlemi cubic kernel (kübik çekirdek) kullanılan DVM algoritmasıyla gerçekleştirilmiş ve model %93,6 doğruluk oranına ulaşmıştır.

Joshi vd. (2017) [11], sağlıklı kontrol bireyleri ile Parkinson hastaları arasındaki farkı belirlemek amacıyla yürüyüşten elde edilen uzamsal-zamansal parametreleri dalgacık dönüşümü (wavelet transform) ile temsil etmiş ve bu öznitelikleri DVM sınıflandırmıştır. Analiz edilen yürüyüş parametreleri arasında adım süresi (stride interval), salınım süresi (swing interval) ve destek süresi (stance interval) yer almakta olup veriler her iki bacak için ayrı ayrı değerlendirilmiştir. Tek bir parametre kullanıldığında, örneğin sol destek süresi veya sağ salınım süresi ile %90,32 doğruluk elde edilirken; sol bacaktan elde edilen tüm parametreler bir araya getirildiğinde doğruluk oranı %100'e ulaşmıştır.

Bu çalışmaların yanı sıra, Balaji vd. (2020) [12] Parkinson hastalığının erken teşhisi ve evre sınıflandırması amacıyla dikey yer tepki kuvveti zaman serisi verilerinin istatistiksel analizini gerçekleştirmiş ve anlamlı spatiotemporal özellikleri belirlemiştir. Çalışmada, karar ağacı, destek vektör makineleri, topluluk sınıflandırıcı ve Bayes sınıflayıcı olmak üzere dört farklı denetimli makine öğrenimi algoritmasının performansı karşılaştırılmış; Hohen & Yahr (H&Y) ölçeğine göre hastalık evreleri %99,4 doğruluk oranıyla sınıflandırılmıştır.

DL Zhao vd. (2018b) [13], Parkinson hastalığının teşhisi ve evre sınıflandırması amacıyla dikey yer tepki kuvveti verilerine dayalı hibrit bir derin öğrenme modeli önermiştir. Çalışmada, yürüyüş verilerinin uzamsal ve zamansal örüntülerini öğrenmek üzere iki kanallı bir mimari geliştirilmiş; bu mimari, evrişimsel sinir ağı ve uzun kısa süreli bellek yapılarının birleşiminden oluşmaktadır. Evrişimsel sinir ağı uzun kısa süreli bellek tabanlı hibrit yaklaşım hem PH tespiti hem de Hoehn & Yahr (H&Y) ölçeğine göre evre sınıflandırmasında yüksek doğruluk oranları sağlamış ve özellikle Si veri setinde %98,8 doğruluk elde edilmiştir.

Manap, Tahir vd. (2011a) [14], Parkinson hastalığına sahip bireylerin yürüyüş örüntülerini analiz etmek amacıyla yapay sinir ağı (YSA) tabanlı bir sınıflandırma

modeli geliştirmiştir. Çalışmada, 12 Parkinson hastası ve 20 sağlıklı bireyden temel ve kinetik yürüyüş verileri toplanmış; bu veriler dört gruba ayrılarak üçü eğitim, biri test seti olarak kullanılmıştır. Modelin genellenebilirliğini değerlendirmek amacıyla çapraz doğrulama yöntemi uygulanmıştır. YSA sınıflandırıcısı, temel ve kinetik özelliklerle %81,25 doğruluk oranı elde etmiş; kinematik özelliklerin dahil edilmesiyle bu oran %84,38'e yükselmiştir. Farklı özellik türlerinin birleştirilmesiyle (veri füzyonu) doğruluk oranı %87,5'e ulaşmıştır. Bu bulgular, YSA tabanlı modellerin gait verileri üzerinden Parkinson hastalığının teşhisinde etkili bir araç olabileceğini ve çoklu özellik entegrasyonunun sınıflandırma performansını artırabileceğini göstermektedir.

Manap ve arkadaşlarının yürüttüğü diğer bir çalışmada [15], Parkinson hastalığına özgü patolojik yürüyüş özelliklerini belirlemek amacıyla DYTK verileri vücut ağırlığına oranlanarak normalize edilmiştir. Hastalar ve sağlıklı bireyler arasındaki farkları incelemek için bağımsız t-testi ve Pearson korelasyon testi uygulanmıştır. Elde edilen istatistiksel sonuçlara göre, adım uzunluğu, yürüme hızı, diz açısı ve dikey yer tepki kuvveti en ayırt edici dört özellik olarak belirlenmiş ve bu özellikler YSA modeline girdi olarak verildiğinde %95,63 sınıflandırma doğruluğuna ulaşılmıştır.

Tahir & Manap (2012) çalışmasında [16] ise Parkinson hastaları ile sağlıklı bireylerin yürüyüş verileri, temel spatiotemporal, kinematik ve kinetik özellikler olarak üç grup altında incelenmiştir. Veriler DYTK normalizasyonu (vücut ağırlığına göre), intra-grup ve inter-grup normalizasyon yöntemleriyle ön işleme tabi tutulmuştur. Sınıflandırma modelleri olarak YSA ve DVM algoritmaları denenmiştir. Özellikle intra-grup normalizasyon ve yalnızca basic spatiotemporal özellikler kullanıldığında her iki model de %100 doğruluk elde etmiştir. Buna karşın diğer özellik grupları veya kombine veri setlerinde doğruluk oranları biraz daha düşük olmuştur.

Literatürde yer alan çalışmaların yöntemsel yaklaşımları, kullanılan veri türler ve sınıflandırma algoritmaları açısından hem ortak hem de farklı yönler barındırmaktadır. Bu benzerlikleri ve ayrışmaları daha sistematik biçimde değerlendirebilmek amacıyla, incelenen araştırmaların temel özellikleri Tablo 2.1'de karşılaştırmalı olarak özetlenmiştir.

Tablo 2.1 Literatür özeti

Çalışma	Veri	Veri Türü	Algoritma	Doğruluk	Açıklama
Wu & Krishnan (2009)	15 PH 16 KG	Adım süresi, alt fazlar	DVM	%90,3	Leave-one-out çapraz doğrulama
Lee & Lim (2012)	93 PH 73 KG	Wavelet öznitelikler	Ağırlıklı Bulanık Üyelik Fonksiyonlu YSA	%77,3	Sensitivite: %81,1, Spesifikite: %65,5
Perumal & Sankar (2016)	93 PH 73 KG	Yürüyüş, Titreme	Lineer Diskriminant Analiz	%86,9	Erken teşhis için öznitelik kombinasyonu
Yurdakul vd. (2020)	93 PH 73 KG	DYTK, KT-YİD öznitelik	YSA	%98,3	Student t-testi ile anlamlı özellikler
Wahid vd. (2015)	23 PH 26 KG	Uzamsal-zamansal yürüyüş	Karar Ağacı, DVM	%92,6 %80,4	Fiziksel özelliklerden arındırılmış veri
Alam vd. (2017)	29 PH 18 KG	DYTK öznitelikler	DVM	%93,6	Özellik seçimi: stride/swing ve CoP
Joshi vd. (2017)	15 PH 16 KG	Uzamsal-zamansal parametreler	DVM	%90,3 %100	Tüm parametreler birleştirildiğinde %100
Balaji vd. (2020)	93 PH 73 KG	DYTK zaman serisi	Karar Ağacı, DVM, Bayes	%99,40	Evre sınıflandırması H&Y ölçeği
DL Zhao vd. (2018)	93 PH 73 KG	DYTK, spatiotemporal	Hibrit CNN-LSTM	%98,80	Hem PH tespiti hem evre sınıflandırması
Manap vd. (2011-2012)	12 PH 20 KG	DYTK, kinematik/kinetik	YSA, DVM	%81,25 %100 (Intra grup)	Veri füzyonu ve grup içi normalizasyon ile en yüksek doğruluk

3
YÖNTEM

3.1 Araştırma Yöntemi

Bu çalışmada, PH teşhisinde yapay zekâ tekniklerinin etkinliğini değerlendirmek amacıyla açık kaynak bir yürüyüş veri seti kullanılmıştır. Araştırma, tanımlayıcı ve deneysel bir yöntem izlemekte olup, veri analizi ve makine öğrenmesi modellerinin uygulanmasını kapsamaktadır.

Çalışma süreci aşağıdaki aşamalardan oluşmaktadır:

- Veri Seti ve Kaynağı: Açık erişimli bir veri tabanından yürüyüş verileri temin edilmiştir. Bu veriler, Parkinson hastaları ile sağlıklı bireylerin yürüyüş özelliklerini içermektedir
- Özellik Çıkarımı: Verilerden permütasyon entropi, negentropi, adım/dakika hesaplama, Neurokit ile toplam ayak kuvvetlerinin filtrelenmesi, sağ-sol ayak denge farkı ve Gini değeri gibi yürüyüşe özgü çoklu özellikler çıkarılmıştır.
- Makine Öğrenmesi Modelleri: Elde edilen öznitelikler, destek vektör makineleri, yapay sinir ağları ve lojistik regresyon algoritmaları ile sınıflandırılmıştır.
- Model Değerlendirmesi: Sınıflandırma modelleri doğruluk, sensitivite ve spesifiklik gibi ölçütlerle değerlendirilmiş; Parkinson hastalığına sahip bireylerin sağlıklı bireylerden ayrıştırılmasında model performansları karşılaştırılmıştır.

Bu yaklaşım hem literatürdeki benzer çalışmalarla uyumlu hem de tez kapsamında geliştirilen özgün öznitelik çıkarım teknikleriyle desteklenmiştir.

3.2 Veri Seti ve Kaynağı

Bu tez çalışmasında 93 PH ve 72 KG verisi bulunan açık erişimli bir veri seti [17] kullanılmıştır. İdiyopatik Parkinson hastalığı tanısı almış bireyler ile sağlıklı kontrol grubunun yürüyüş ölçümlerini içermektedir. Katılımcılar düz bir zemin üzerinde kendi belirledikleri doğal yürüyüş hızlarında yürürken dikey zemin tepki kuvveti ölçümleri kaydedilmiştir. Katılımcı demografisi Tablo 3.1'de özetlenmiştir.

Tablo 3.1 Katılımcıların demografik özellikleri

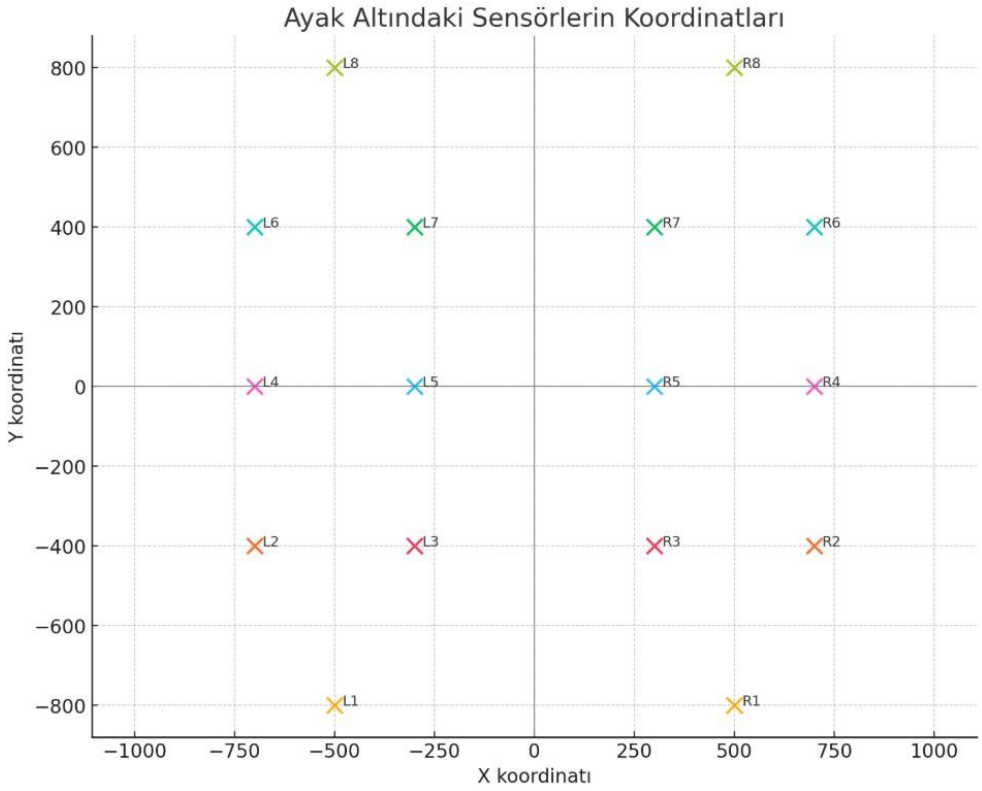
Grup	Katılımcı Sayısı	Ortalama Yaş	Erkek (%)	Kadın (%)
PHG	93	66,3	62%	38%
KG	72	63,6	55%	45%

Her ayağın altında, kuvveti (Newton cinsinden) ölçen 8 sensör (Ultraflex Computer Dyno Graphy, Infotronic Inc.) vardır. Toplamda 16 sensörün her birinden elde edilen veriler sayısallaştırılmış ve saniyede 100 örnekleme frekansıyla kaydedilmiştir. Kayıtlar ayrıca her ayak için 8 sensör çıkışının toplamını yansıtan iki sinyal içermektedir.

Ayak tabanındaki sensörlerin koordinatları Tablo 3.2'de verilmiş ve Şekil 3.1'de gösterilmiştir.

Tablo 3.2 Ayak tabanı sensör konumları

Sol Ayak Sensörleri				Sağ Ayak Sensörleri		
Sensör	**X**	**Y**		**Sensör**	**X**	**Y**
L1	-500	-800		R1	500	-800
L2	-700	-400		R2	700	-400
L3	-300	-400		R3	300	-400
L4	-700	0		R4	700	0
L5	-300	0		R5	300	0
L6	-700	400		R6	700	400
L7	-300	400		R7	300	400
L8	-500	800		R8	500	800



Şekil 3.1 Ayak tabanı sensör konumları

Bu veritabanı ayrıca hastalık şiddeti ölçümlerini (yani, Hoehn & Yahr evrelemesi ve / veya Birleşik Parkinson Hastalığı Derecelendirme Ölçeği kullanılarak) içermektedir [17].

3.3 Özellik Çıkarımı ve Uygulama

Veri setinde yer alan dikey zemin tepki kuvveti sinyalleri, Parkinson hastalığına özgü yürüyüş bozukluklarını ayırt edebilmek amacıyla çeşitli öznitelik çıkarım teknikleriyle analiz edilmiştir. Bu bölümde, sinyal işleme yöntemleri kullanılarak elde edilen özgün öznitelikler açıklanmakta ve bunların sınıflandırma algoritmalarına nasıl entegre edildiği aktarılmaktadır.

Özellikle zaman serisi analizi, entropi tabanlı ölçümler, adım frekansı hesaplamaları ve ayak kuvveti dengesi gibi parametreler üzerinden yürüyüş dinamikleri değerlendirilmiştir. Geliştirilen öznitelikler, Parkinson hastalığına özgü motor semptomların daha hassas biçimde modellenmesini ve sınıflandırılmasını desteklemektedir.

3.3.1 Permütasyon Entropi

Zaman serilerinin karmaşıklığını ölçmek, serinin altında yatan dinamik sistemin doğasını anlamak için kritik bir adımdır. Permütasyon entropisinin temel fikri, bir zaman serisindeki ardışık değerlerin sıralama düzenlerine (permütasyonlarına) odaklanmaktır. Yöntem, verinin mutlak değerlerinden ziyade komşu değerler arasındaki göreli ilişkiyi analiz eder. Bu yaklaşım, yöntemi hem hesaplama açısından verimli kılar hem de gürültüye ve monoton dönüşümlere karşı oldukça dayanıklı hale getirir. Bandt ve Pompe'nin (2002) de belirttiği gibi, bu yöntemin avantajları "basitliği, son derece hızlı hesaplanması, sağlamlığı ve doğrusal olmayan monoton dönüşümlere göre değişmezliğidir." [18].

Permütasyon entropisi, bir zaman serisindeki ardışık değerlerin kendi içindeki sıralanma düzenine (ordinal pattern) dayanarak hesaplanan bir entropi türüdür. Bu yaklaşımda zaman serisi, belirli bir gömme boyutu kullanılarak kayan pencerelere bölünür ve her pencere içindeki değerlerin sıralama yapısı (permütasyonu) belirlenir. Tüm bu permütasyonların olasılık dağılımı elde edildikten sonra klasik Shannon entropisi formülü uygulanır:

$$H_n = -\sum_{p \in S_n} p(p) \cdot log\big(p(p)\big) \tag{3.1}$$

Burada:

- S_n : n-elemanlı tüm olası permütasyonların kümesi
- $p(p)$: her permütasyonun gözlenme olasılığı
- Logaritma genellikle 2 tabanında alınarak bilgi miktarı bit cinsinden hesaplanır.

Permütasyon entropisinin değeri:

- Düzenli dizilerde sıfıra yaklaşır,
- Tamamen rastgele dizilerde maksimum olan log(n!) yaklaşır.

Yöntemin daha iyi anlaşılması için basit bir örnek vermek gerekirse, zaman serisi olarak x= [4,7,9,10,6,11,3] dizisi alındığında ve gömme boyutu n=3 seçildiğinde, diziden toplam 5 adet üçlü alt dizi elde edilir: (4,7,9), (7,9,10), (9,10,6), (10,6,11) ve (6,11,3). Her alt dizinin içindeki sayılar küçükten büyüğe sıralanır ve buna göre indeks sırası bulunur. Örneğin;

(9, 10, 6)

Sayılar: 9, 10, 6

- Küçükten büyüğe sıralanırsa: 6 <9 <10
- Ama orijinal alt dizideki indeksleri:
- 0 → 9
- 1 → 10
- 2 → 6
- Şimdi, en küçük sayı 6, orijinal indeksi 2
- İkinci en küçük sayı 9, orijinal indeksi 0
- En büyük sayı 10, orijinal indeksi 1
- Yani sıralama: (2, 0, 1)

Bu üçlülerin her biri sıralandığında şu permütasyonlara ve indeks sıralarına karşılık gelir:

- (4, 7, 9) → (0, 1, 2)
- (7, 9, 10) → (0, 1, 2)
- (9, 10, 6) → (2, 0, 1)
- (10, 6, 11) → (1, 0, 2)
- (6, 11, 3) → (2, 0, 1)

Bu durumda elde edilen permütasyon frekansları:

- (0,1,2): 2 kez
- (2,0,1): 2 kez
- (1,0,2): 1 kez

Buna göre olasılık dağılımı:

- $p1$=2/5
- $p2$=2/5
- $p3$=1/5

Bu dağılıma Shannon entropisi uygulandığında:

$$H_3 = -\left(\left(\frac{2}{5}\right)\log_2\left(\frac{2}{5}\right) + \left(\frac{2}{5}\right)\log_2\left(\frac{2}{5}\right) + \left(\frac{1}{5}\right)\log_2\left(\frac{1}{5}\right)\right) \approx 1.522\ bit \quad (3.2)$$

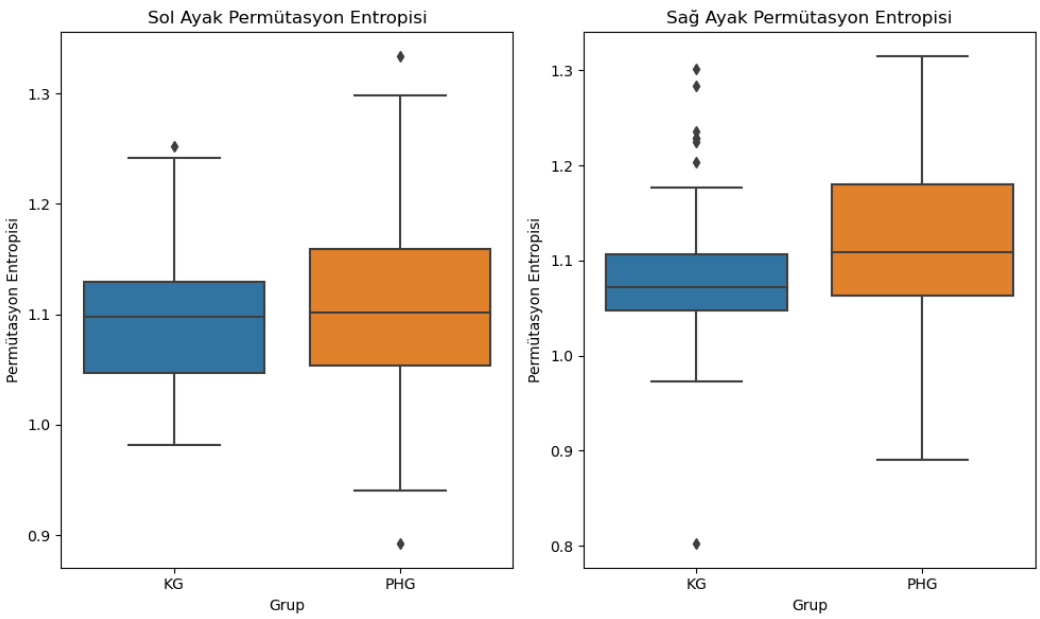
Bu değer, serinin ne tamamen düzenli ne de tamamen rastgele olduğunu, yani orta düzeyde bir karmaşıklık barındırdığını gösterir.

Permütasyon entropisi yalnızca sayısal bir ölçüt sunmakla kalmaz, aynı zamanda farklı sistemlerin zaman serilerini karşılaştırmak veya bir sistemin zaman içindeki durum değişimlerini analiz etmek için de güçlü bir araçtır. Düşük entropi değerleri, sistemin öngörülebilir veya periyodik olduğunu; yüksek entropi değerleri ise karmaşık ya da kaotik bir davranış sergilediğini gösterebilir. Bu nedenle, yöntem hem teorik hem de uygulamalı çalışmalarda yaygın şekilde kullanılmaktadır.

Bu tez kapsamında, yürüyüş verilerinden elde edilen dikey yer reaksiyon kuvveti zaman serilerine permütasyon entropisi yöntemi uygulanarak, Parkinson hastalarının ve kontrol grubunun duruş fazlarındaki karmaşıklık düzeyleri karşılaştırılmıştır. Duruş fazı, yerle temasın olduğu, kuvvet üretiminin aktif olduğu dönemi temsil eder ve motor kontrol açısından hastalıkların etkilerini yansıtma potansiyeli taşır.

Analiz sürecinde her bireyin sol ve sağ ayak toplam kuvvet verileri arasından, belirli bir eşik değerinin üzerindeki bölümler seçilmiştir. Bu seriler üzerinde gömme boyutu n=3 seçilerek permütasyon entropisi hesaplanmıştır. Bu seçim, hem literatürde önerilen aralıklara uygunluk sağlamış hem de kullanılan örnekleme uzunluklarıyla yeterli veri segmenti sunmuştur.

Elde edilen entropi değerleri, her birey için sol ve sağ ayak olmak üzere iki ayrı değer olarak kaydedilmiş ve ilgili gruba göre sınıflandırılmıştır. Bu sayede, her iki grubun duruş fazındaki motor karmaşıklık düzeyi nicel olarak karşılaştırılmıştır. Permütasyon entropisi değerlerinin istatistiksel analizinde normallik sınamaları (örneğin Shapiro-Wilk testi) yapılmıştır. Şekil 3.2'de, sol ve sağ ayak için permütasyon entropisi değerlerinin kontrol ve hasta grupları arasında karşılaştırılması boxplot grafikleri ile gösterilmektedir.



Şekil 3.2 Gruplar arası permütasyon entropisi karşılaştırması boxplot grafiği

Sol ve sağ ayak permütasyon entropisi değerlerine ait istatistiksel karşılaştırmalar, Tablo 3.3'te gösterilmiştir.

Tablo 3.3 Sol ve sağ ayak permütasyon entropisi istatiksel değerleri

Ayak	Grup	Ortalama	Medyan	Çarpıklık
Sağ Ayak	KG	1.0856	1.0718	0.1892
	PHG	1.1191	1.1085	0.1207
Sol Ayak	KG	1.0948	1.0981	0.3133
	PHG	1.1079	1.1016	0.2680

Şekil 3.3'te verilen boxplot ve Tablo 3.3'teki değerlere göre:

- Sol ayak için her iki grupta da ortanca değerler birbirine oldukça yakındır. KG dağılımı biraz daha sıkışık görünmekteyken, PHG daha geniş bir dağılım gözlemlenmektedir.
- Sağ ayak için ise PHG entropi değerleri daha geniş bir aralıkta dağılmış ve ortanca değeri KG'ye kıyasla daha yüksek görünmektedir. Aynı zamanda PHG maksimum değer daha yüksektir.

- Her iki ayak için de PHG uç değerler (outlier) sayısı KG'ye göre daha fazladır; bu durum, hasta grubunda bireysel farklılıkların daha belirgin olabileceğini düşündürebilir.

Gruplar arası fark Welch t-testi ile değerlendirilmiştir. Sonuçları Tablo 3.4'te paylaşılmıştır.

Tablo 3.4 Permütasyon entropisi Welch t-testi

Ayak	**Welch's t-Test T**	**Welch's t-Test p**	**Anlamlılık**
Sağ Ayak	–2.7974	0.0058	Anlamlı ($p < 0.05$)
Sol Ayak	–1.1498	0.2519	Anlamlı değil ($p \geq 0.05$)

Bu bulgular, permütasyon entropisi açısından Parkinson hasta grubu (PHG) ile kontrol grubu (KG) arasında özellikle sağ ayakta anlamlı farklılıklar olduğunu göstermektedir. Welch t-testi sonuçlarına göre sağ ayak entropi değerleri istatistiksel olarak anlamlı şekilde farklılık göstermektedir ($p = 0.0058$). Bu durum, Parkinson hastalarında motor kontrol mekanizmalarının sağ ayakta daha fazla etkilenmiş olabileceğini düşündürmektedir. Ayrıca PHG grubunda gözlenen daha geniş dağılım ve artan uç değer sayısı, bireysel motor semptomların çeşitliliğini yansıtabilir.

Sonuçlar, Parkinson hastalarının duruş fazındaki permütasyon entropisinin kontrol grubuna kıyasla anlamlı derecede farklılık gösterebildiğini ortaya koymuştur. Bu durum, hastalığın motor kontrol sistemine olan etkisinin zaman serisi düzeyinde tespit edilebildiğini göstermektedir. Ayrıca, yöntemin düşük boyutlu, hızlı hesaplanabilir ve gürültüye dayanıklı yapısı sayesinde klinik veri setlerinde etkin şekilde kullanılabileceği de görülmüştür.

3.3.2 Negentropi

Bilgi kuramında entropi, bir sistemin düzensizlik (ya da belirsizlik) düzeyini ifade eden bir ölçüttür. Negentropi (negentropy) ise, sistemin düzensizlikten uzaklaşma miktarını, yani bilgi içeriğini temsil eder. Bu kavram bir sinyalin normal dağılımdan sapma miktarını nicel olarak değerlendirmek için kullanılır [19]. Bu çalışmada

negentropi, Parkinson hastalığına sahip bireylerin yürüyüş verilerindeki motor kontrol düzenliliğini değerlendirmek amacıyla kullanılmıştır.

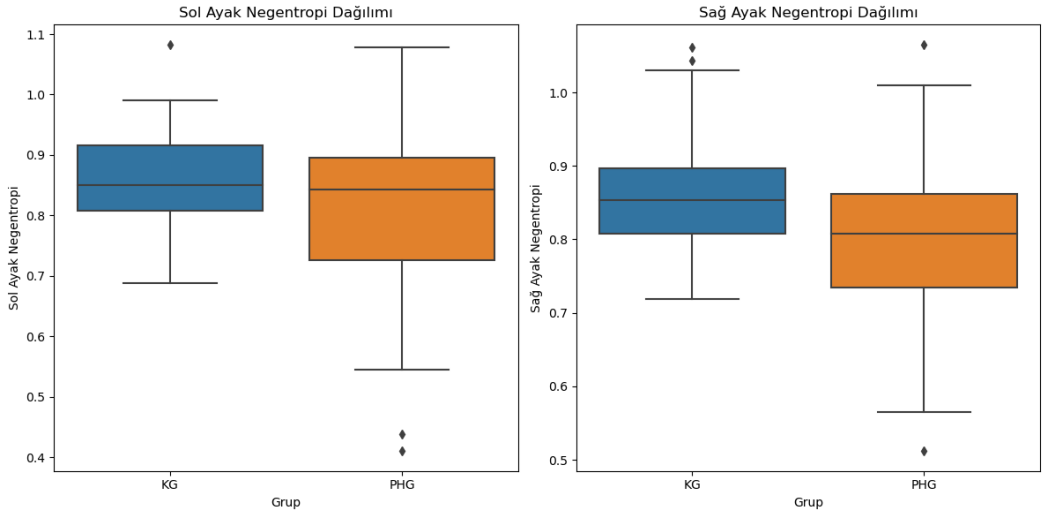
Bu çalışmada kullanılan negentropi formülü, histogram tabanlı negentropi ölçümüne dayalı uyarlamalı bir yaklaşımdır. Verinin normalize edilmiş histogramı üzerinden entropi hesaplanmış, ardından dolu bin sayısının logaritması ile fark alınarak negentropi değeri elde edilmiştir. Bu yöntem Watts&Crow (2022) [20] tarafından önerilen histogramın analizleriyle uyumlu olup, verinin bilgi yoğunluğunu pratik biçimde değerlendirmeyi amaçlamaktadır. Kullanılan negentropi formülü aşağıda yer almaktadır.

$$J(y) = \log(n) - H(y) \tag{3.3}$$

Burada:

- n, histogramdaki dolu bin (aralık) sayısıdır,
- H(y), normalleştirilmiş verinin histogramına dayalı Shannon entropisidir.

Her bir kişinin yürüme sırasında elde edilen toplam ayak basınç kuvveti sinyalleri (sol ve sağ ayak için ayrı ayrı) normalize edilmiş ve yukarıdaki formül ile negentropi değerleri hesaplanmıştır. Şekil 3.3'te yer alan boxplot grafikleri, her iki grubun sol ve sağ ayak negentropi dağılımlarını görsel olarak karşılaştırmaktadır.



Şekil 3.3 Sol ve sağ ayak için negentropi dağılımlarının boxplot grafiği

Elde edilen negentropi değerleri gruplar bazında karşılaştırılmıştır ve Tablo 3.5'te gösterilmiştir. İstatistiksel olarak anlamlı bir fark olup olmadığı bağımsız t-testi ile değerlendirilmiştir, Tablo 3.6'da t testi değerleri yer almaktadır.

Tablo 3.5 Negentropi istatiksel özeti

Grup	Ayak	Ortalama	Medyan	Std. Sapma	Varyans	Çarpıklık
KG	Sol	0.8589	0.8499	0.0741	0.0055	0.1239
	Sağ	0.8551	0.8528	0.0752	0.0057	0.6855
PHG	Sol	0.8121	0.8424	0.1202	0.0144	-0.6961
	Sağ	0.8018	0.8082	0.1048	0.0110	-0.2800

- KG hem sol hem de sağ ayak için daha yüksek ortalama negentropi değerlerine sahiptir (≈ 0.86).
- PHG, özellikle sol ayakta daha yüksek varyans göstermektedir. Bu durum, Parkinson hastalarının sinyallerinin daha düzensiz ve dağılımının geniş olduğunu düşündürür. Parkinson grubu içinde bireyler arası fark daha fazladır.
- PHG negatif çarpıklık, bazı bireylerin sinyallerinin çok daha düşük negentropi değerlerine sahip olduğunu (yani oldukça entropik/düzensiz sinyaller taşıdığını) gösterir.

Tablo 3.6 Negentropi t-testi değerleri

Ayak	t-istatistik	p-değeri	Anlamlılık (p <0.05)
Sol Ayak	3.0723	0.0025	Anlamlı
Sağ Ayak	3.7977	0.0002	Anlamlı

Yapılan bağımsız t-testi sonuçlarına göre hem sol hem de sağ ayak negentropi değerleri bakımından KG ile PHG arasında istatistiksel olarak anlamlı farklar tespit edilmiştir. Özellikle sağ ayakta bu farkın daha belirgin olduğu görülmektedir ($p = 0.0002$). Bu bulgular, Parkinson hastalığının yürüyüş sırasında üretilen ayak kuvvet

sinyallerinin yapısını önemli ölçüde etkilediğini ve bu değişimlerin bilgi kuramsal ölçütler ile (negentropi) tespit edilebileceğini ortaya koymaktadır.

KG negentropi değerleri daha istikrarlı, bilgi içeriği yüksek sinyaller üretmektedir. PHG ise daha fazla varyans, daha düşük ortalama, negatif çarpıklık sergilemektedir. Bu da sinyallerin daha az bilgi içerdiğini, daha dağınık olduğunu ve Parkinson hastalığının motor düzensizliklerinin sinyal yapısına doğrudan yansıdığını göstermektedir. Bu istatistiksel bulgular, negentropi değerlerinin Parkinson hastalığının tanısal ayrımında kullanılabilirliğini desteklemektedir.

Sonuç olarak, negentropi Parkinson hastalığının tanısal değerlendirilmesinde potansiyel bir biyobelirteç olarak kullanılabilir. Bu yaklaşım, yürüyüş verilerinden elde edilen kuvvet sinyallerinin bilgi içeriğini ölçerek hastalıkla ilişkili motor bozulmaları nicel olarak ortaya koymaktadır.

3.3.3 Adım/Dakika Hesaplama

Bu çalışmada, bireylerin yürüyüş performanslarını değerlendirmek amacıyla adım/dakika hesaplamaları gerçekleştirilmiştir.

Analizler, sol ve sağ ayaktaki her bir sensor için ayrı yürütülmüştür. Adım sayısı, belirli bir sensörden elde edilen kuvvet değerinin belirli bir eşik değerin üzerine çıkmasıyla bir adım olarak tanımlanmıştır. Ardışık veri noktalarında eşik değeri geçildikten sonra tekrar altına inmesiyle yeni bir adım algılanabilmektedir. Böylece her birey için, sol ve sağ ayak adım sayıları ayrı ayrı belirlenmiştir.

Topuk temasının adım başlangıcına işaret etmesi, klinik yürüyüş analizlerinde geniş kabul gören bir yaklaşımdır. Literatürde “first strike” özellikle topuğun yere temas ettiği anı belirtmektedir [21].

Her iki ayakta sekizer sensör bulunmakta olup, özellikle “DYTK Sol/Sağ Ayak Sensör 1” verileri, topuk bölgesine yerleştirildiğinden adım başlangıcını tespit etmede tercih edilmiştir.

Adım sıklığı (adım/dakika), aşağıdaki formül ile hesaplanmıştır:

$$\frac{Adım}{Dakika} = \frac{Toplam\ Adım\ Sayısı}{Toplam\ Süre\ (saniye)} \times 60 \qquad (3.4)$$

Veriler daha sonra PHG ve KG olarak sınıflandırılmıştır. Her iki grup için adım/dakika hesaplamaları karşılaştırılmış ve istatistiksel anlamlılık analizinde bağımsız t-testi (independent t-test) uygulanmıştır.

Tablo 3.7 Adım/dakika karşılaştırması

Grup	Sol Ayak Ortalama (adım/dk)	Sağ Ayak Ortalama (adım/dk)
KG	55.27	54.95
PHG	51.68	52.19

Tablo 3.8 Adım/dakika t-testi sonuçları

Ayak	t-değeri	p-değeri	Anlamlılık Durumu
Sol Ayak	2.74	0.0068	Anlamlı (**p <0.01**)
Sağ Ayak	1.98	0.0495	Anlamlı (**p <0.05**)

Bu sonuçlar hem sol hem de sağ ayakta adım/dakika açısından KG ve PHG arasında istatistiksel olarak anlamlı farklar olduğunu göstermektedir.

3.3.4 Neurokit – Toplam Ayak Kuvvetleri Filtreleme

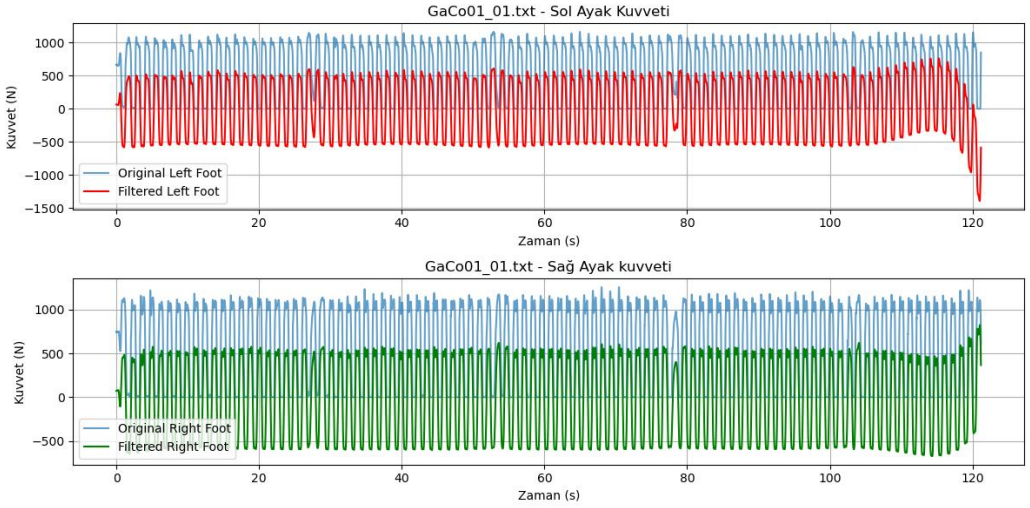
Bu çalışmada, Parkinson hastalığına sahip bireylerin yürüyüş verilerinden elde edilen toplam ayak kuvveti sinyalleri, sinyal işleme teknikleri kullanılarak filtrelenmiştir. Filtreleme işlemi, sinyal işleme alanında, istenmeyen frekans bileşenlerini ortadan kaldırmak veya azaltmak için kullanılır [22]. Bu işlem, sinyalin belirli bir frekans aralığında daha temiz ve anlaşılır hale gelmesini sağlar.

Veri setinde her birey için sol ve sağ ayaklara ait toplam kuvvet değerleri (Toplam Kuvvet Sol Ayak (N) ve Toplam Kuvvet Sağ Ayak (N)) zaman serisi biçiminde yer almaktadır. Bu sinyaller, yürüyüş sırasında ayakların yere uyguladığı dikey kuvvetlerin toplamını temsil etmektedir ve motor fonksiyon bozukluklarının analizi açısından kritik öneme sahiptir.

Filtreleme işlemi için Python ortamında çalışan NeuroKit2 kütüphanesi tercih edilmiştir. Bu kütüphane, biyofizyolojik sinyallerin işlenmesinde yaygın olarak kullanılan, güvenilir ve açık kaynaklı bir araçtır. Filtreleme sürecinde Butterworth

tipi düşük geçişli filtre uygulanmış; bu sayede sinyallerdeki yüksek frekanslı gürültü bileşenleri elimine edilerek daha düzgün ve analiz edilebilir bir sinyal elde edilmiştir. Filtre parametreleri olarak 0.5 Hz alt sınır ve 30 Hz üst sınır belirlenmiştir. Bu frekans aralığı, yürüyüş gibi düşük frekanslı biyomekanik hareketlerin karakteristiğine uygun olarak seçilmiştir.

Filtreleme işleminin etkisini daha somut biçimde gösterebilmek amacıyla, örnek bir bireye ait toplam ayak kuvveti sinyali Şekil 3.4'te görselleştirilmiştir. Grafik, filtreleme öncesi ve sonrası sinyalleri aynı zaman ekseninde karşılaştırmalı olarak sunmaktadır. Bu sayede, Butterworth filtresinin yüksek frekanslı gürültüleri nasıl elimine ettiği ve sinyali nasıl daha düzgün hale getirdiği açıkça gözlemlenebilmektedir.



Şekil 3.4 Ham ve filtrelenmiş ayak kuvveti verileri

Filtrelenmiş sinyaller üzerinden, yürüyüş döngüsünü karakterize eden zirve (maximum) ve çukur (minimum) noktalar belirlenmiştir. Bu noktalar, ayak tabanına uygulanan kuvvetin zaman içerisindeki değişimini temsil etmektedir.

Gruplar arası farkın istatistiksel olarak anlamlı olup olmadığını değerlendirmek amacıyla bağımsız örneklemler için t-testi uygulanmıştır. Her bir özelliğin (zirve ve çukur noktalar için ortalama, maksimum, minimum) KG ve PHG arasında anlamlı fark taşıyıp taşımadığı değerlendirilmiştir.

Bu analizlerin sonuçları Tablo 3.9'da sunulmuştur. Sol ve sağ ayaklara ait ortalama, maksimum ve minimum zirve (peak) ve çukur (trough) değerleri için uygulanan t-

testi sonucunda, gruplar arasında bazı özelliklerde anlamlı farklılıklar gözlemlenmiştir.

Tablo 3.9 Zirve ve çukur noktalarına ait t-testi sonuçları

Özellik	t-istatistiği	p-değeri
Sol Ayak Ortalama Zirve Değeri	-1.595	0.112
Sol Ayak Maksimum Zirve Değeri	-1.795	0.074
Sol Ayak Minimum Zirve Değeri	**-3.034**	**0.003**
Sol Ayak Ortalama Çukur Değeri	-1.584	0.115
Sol Ayak Maksimum Çukur Değeri	-1.803	0.073
Sol Ayak Minimum Çukur Değeri	-1.140	0.256
Sağ Ayak Ortalama Zirve Değeri	-1.613	0.109
Sağ Ayak Maksimum Zirve Değeri	-1.792	0.075
Sağ Ayak Minimum Zirve Değeri	-0.864	0.389
Sağ Ayak Ortalama Çukur Değeri	-1.621	0.107
Sağ Ayak Maksimum Çukur Değeri	-1.781	0.077
Sağ Ayak Minimum Çukur Değeri	**-2.731**	**0.007**

Özellikle Sol Ayak Minimum Zirve Değeri ($p = 0.003$) ve Sağ Ayak Minimum Çukur Değeri ($p = 0.007$) için elde edilen p-değerleri 0.05'in altında bulunmuştur. Bu durum, Kontrol Grubu (KG) ve Parkinson Hastası Grubu (PHG) arasında bu iki özelliğin istatistiksel olarak anlamlı şekilde farklılık gösterdiğini ortaya koymaktadır.

Elde edilen bu sonuçlar, Parkinson hastalarında bazı yürüyüş fazlarının daha düşük kuvvetle gerçekleştiğini ve ayak yere basım örüntüsünde farklılıklar olduğunu düşündürmektedir. Bu farklılıklar, potansiyel olarak hastalığın tanı veya takibinde kullanılabilecek biyobelirteçler olarak değerlendirilebilir.

3.3.5 Sağ – Sol Ayak Denge Farkı

Parkinson hastalığında yürüyüş bozuklukları, ayaklar arasındaki kuvvet dengesizliğiyle kendini gösterebilir. Bu çalışmada, dikey yer reaksiyon kuvveti sensörlerinden elde edilen veriler kullanılarak sağ ve sol ayak arasındaki kuvvet farkı analiz edilmiştir. Bu fark aşağıdaki formül ile hesaplanmıştır.

$$D = \frac{\left|F_{\text{Sağ Ayak}} - F_{\text{Sol Ayak}}\right|}{F_{\text{Sağ Ayak}} + F_{\text{Sol Ayak}}} \tag{3.5}$$

Burada:

- $F_{\text{Sağ Ayak}}$: Sağ ayağın toplam dikey kuvveti (Newton)
- $F_{\text{Sol Ayak}}$: Sol ayağın toplam dikey kuvveti (Newton)
- D: İki ayak arasındaki denge farkı

Formülde pay kısmında, sağ ve sol ayak kuvvetleri arasındaki mutlak fark alınarak denge sapmasının büyüklüğü bulunur. Payda kısmında ise iki ayak toplam kuvvetlerinin toplamı yer alır. Böylece denge farkı, kuvvetlerin toplamına oranlandığı için normalize edilmiş bir ölçü sağlar.

Bu çalışmada, sağ ve sol ayak arasındaki dikey kuvvet farkı denge ölçütü olan D değerleri üzerinde yürüyüş sırasında ortaya çıkan dengesizlikleri tespit etmek için anomali analizi yapılmıştır. Anomaliler, beklenen davranıştan anlamlı sapmalar olarak tanımlanır ve bu sapmaların sayısı, büyüklüğü ve varyasyonu çeşitli metriklerle ölçülmüştür. Veri setinde anormal noktaları belirlemek için aşağıda verilen Z-puanı formülü kullanılmıştır.

$$Z = \frac{X - \mu}{\sigma} \tag{3.6}$$

Burada:

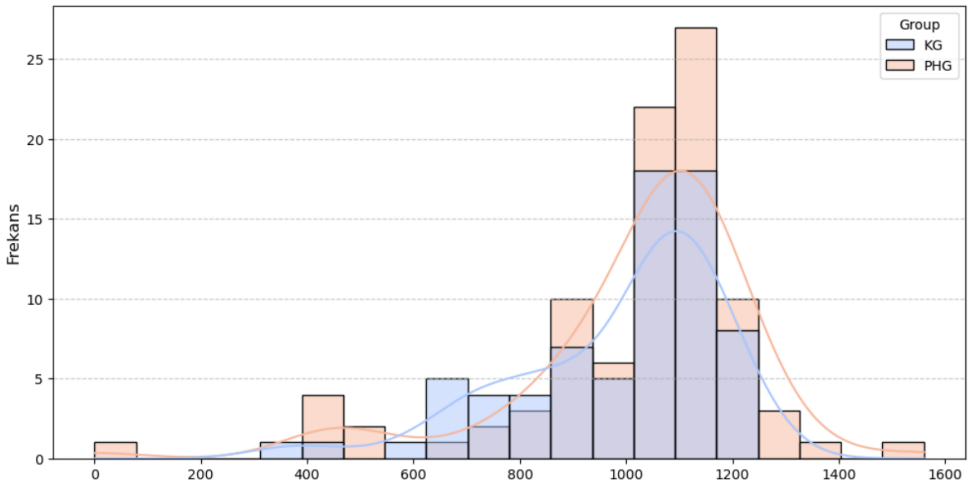
- X= İlgili veri noktası
- μ = Verinin ortalaması
- σ = Verinin standart sapması
- Z= Standartlaştırılmış skor (Z-puanı)

Anomali sayısı, D değerleri arasında standart sapmadan belirlenen eşik değeri aşan veri noktalarının sayısıdır. Aşağıda verilen formül ile hesaplanmıştır.

$$N_{Anomali} = \sum_{i} 1(|Z_i| > 2) \tag{3.7}$$

Burada:

- Z_i: İlgili veri noktası için hesaplanan Z-skoru
- 1: Belirtilen koşul sağlandığında 1, aksi takdirde 0 döndüren gösterge fonksiyonu
- $N_{Anomali}$: Belirli bir eşik değerini aşan anomalilerin toplam sayısı



Şekil 3.5 Sağ – sol ayak denge farkı anomali sayısı

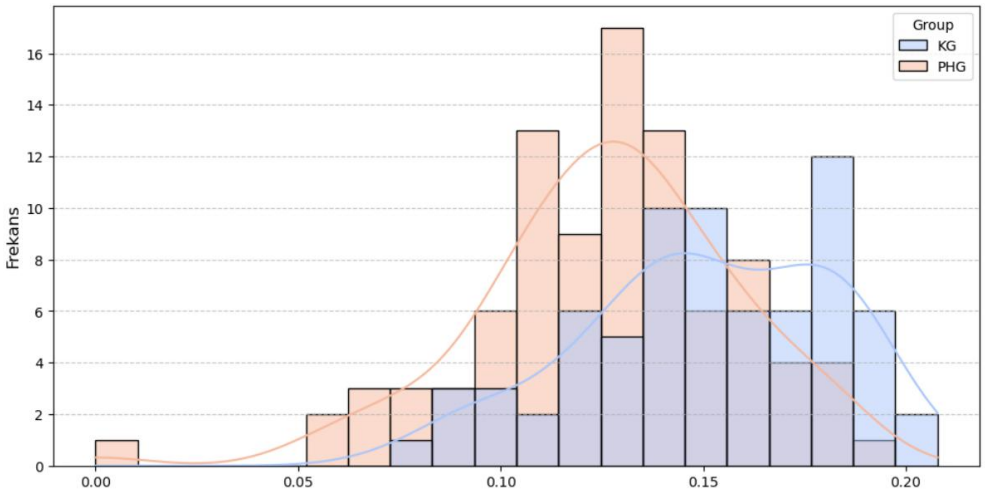
Şekil 3.5'te görüldüğü üzere, PHG bireylerinde sağ–sol ayak denge farkı açısından belirlenen anomali sayısı KG bireylerine kıyasla anlamlı şekilde daha yüksektir. Bu durum, Parkinson hastalığının yürüyüş dinamiklerinde yarattığı bozulmanın nicel bir göstergesi olarak değerlendirilebilir. Histogram dağılımları, PHG grubunda düzensizliklerin daha yoğun ve sistematik olduğunu ortaya koymakta; bu da ilgili metriklerin tanı ve takip süreçlerinde kullanılabilirliğini desteklemektedir.

Bu veride anomali sayısına ek olarak anomali şiddeti de hesaplanmıştır. Anomali şiddeti, yürüyüş verisinde tespit edilen bir anomali noktasının "ne kadar anormal" olduğunu, yani ortalamadan ne kadar saptığını nicel olarak ifade eder. Anomaliyi sadece saymak yeterli değildir; bazen sayısı az ama etkisi büyük (yüksek şiddetli) anomaliler olabilir. İşte bu yüzden, anomalilerin şiddeti, dengesizliklerin sadece varlığını değil, aynı zamanda ciddiyetini de değerlendirmeye yarar. Anomali şiddeti formülü aşağıda verilmiştir.

$$S_{Anomali} = \frac{1}{N_{Anomali}} \sum_{i \in Anomali} D_i \qquad (3.8)$$

Burada:

- D= İlgili veri noktasının D Difference değeri
- $N_{Anomali}$= Anormal noktaların sayısı
- $S_{Anomali}$= Ortalama anomali şiddeti



Şekil 3.6 Sağ – sol ayak denge farkı anomali şiddeti

Şekil 3.6'da görülen anomali şiddeti dağılımları, PHG bireylerinde düzensizliklerin daha sık ancak daha düşük şiddette gerçekleştiğini göstermektedir. KG bireylerinde ise anomali şiddeti daha geniş bir aralıkta dağılmış ve ortalama değerler daha yüksek gözlenmiştir. Bu durum, Parkinson hastalığının yürüyüş dinamiklerinde yarattığı sürekli fakat hafif sapmalarla karakterize olduğunu ve bu sapmaların tanı süreçlerinde dikkate alınması gerektiğini ortaya koymaktadır.

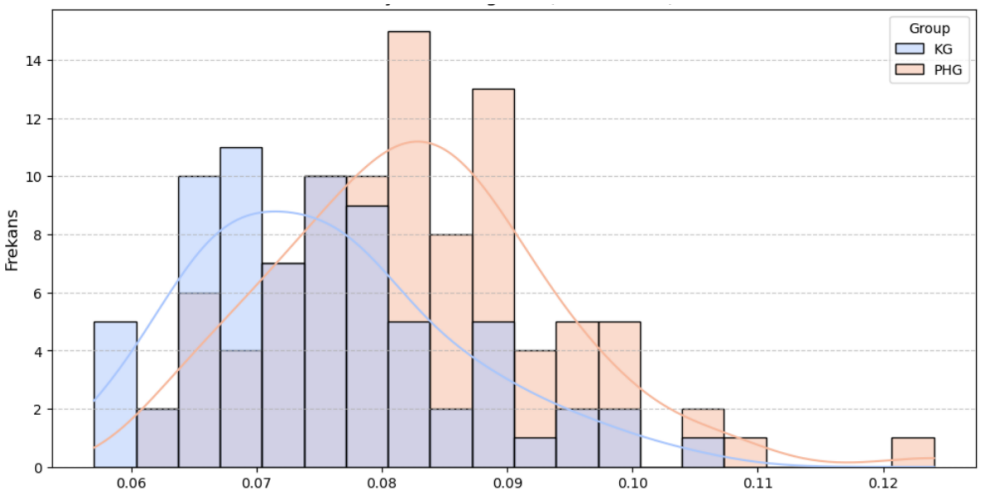
Varyans, ölçülen değerlerin kendi ortalamalarından ne kadar saptığını gösteren istatistiksel bir ölçüttür. Bu bağlamda düşük varyans, kişinin yürüyüşünün tutarlı ve düzenli olduğunu, yüksek varyans ise yürüyüş deseninde değişkenlik ve potansiyel motor kontrol bozukluğu olduğunu gösterebilir.

Varyans, aşağıdaki formül ile hesaplanmıştır.

$$V = \frac{1}{N}\sum_{i=1}^{N}(D_i - \mu)^2 \tag{3.9}$$

Burada:

- D = İlgili veri noktası için D denge fark değeri
- μ = D denge fark değerlerinin ortalaması
- N = Veri kümesindeki toplam nokta sayısı
- V = Varyans değeri

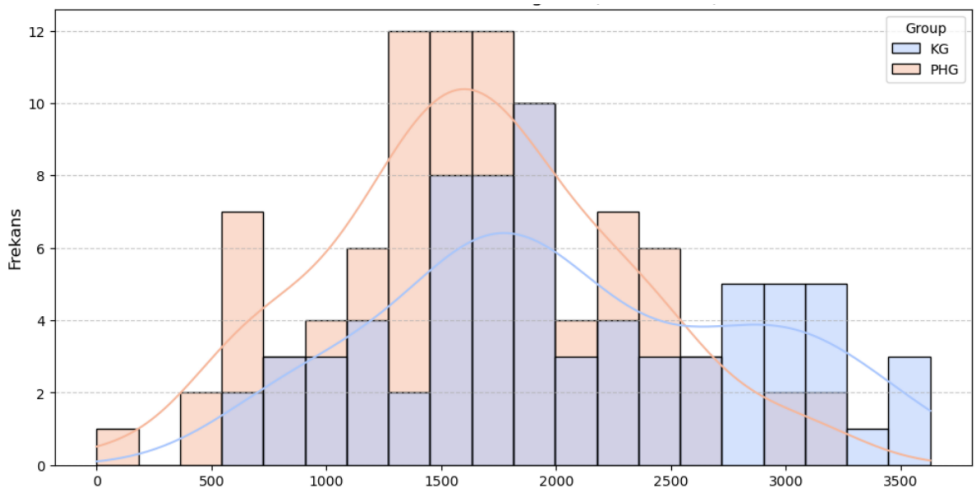


Şekil 3.7 Sağ – sol ayak denge farkı varyansı

Şekil 3.7'de görülen varyans dağılımları, PHG bireylerinde yürüyüş sırasında sağ–sol ayak basınç farklarının daha değişken olduğunu ortaya koymaktadır. Bu durum, Parkinson hastalığının motor semptomlarının bireyler arasında farklı düzeylerde etkili olduğunu ve yürüyüş dinamiklerinde tutarsızlık yarattığını göstermektedir. KG bireylerinde ise varyans değerleri daha düşük ve homojen bir dağılım sergilemiş, bu da sağlıklı yürüyüşün daha istikrarlı olduğunu desteklemektedir.

Anomali sayısı, anomali şiddeti ve varyans ile aşağıdaki formüle göre düzensizlik skoru hesaplanmıştır.

$$I = \frac{N_{\text{Anomali}} \times S_{\text{Anomali}}}{V + 10^{-6}} \tag{3.10}$$



Şekil 3.8 Sağ – sol ayak denge farkı düzensizlik skoru

Şekil 3.8'de görülen düzensizlik skoru dağılımları, KG bireylerinde bazı durumlarda PHG bireylerinden daha yüksek skorlar elde edildiğini göstermektedir. Bu durum, skor formülünde varyansın paydada yer almasıyla açıklanabilir. PHG bireylerinde yürüyüş düzensizlikleri daha değişken olduğu için varyans değerleri yükselmekte ve bu da skorun düşmesine neden olmaktadır. KG bireylerinde ise daha az ama tutarlı düzensizlikler, düşük varyans ile birleştiğinde daha yüksek skorlar üretmektedir. Bu sonuç, düzensizlik skorunun sadece bozukluk miktarını değil, aynı zamanda yürüyüş deseninin tutarlılığını da değerlendirdiğini göstermektedir.

Parkinson hastalarında yürüyüş dinamiklerini değerlendirmek amacıyla, sağ–sol ayak denge farkı üzerinden hesaplanan anomali sayısı, anomali şiddeti, varyans ve düzensizlik skoru gibi metrikler kullanılmıştır. Bu metriklerin, Parkinson Hasta Grubu (PHG) ile Kontrol Grubu (KG) bireyleri arasında anlamlı farklılıklar içerip içermediği bağımsız örneklem t testi ile analiz edilmiştir.

Tablo 3.10'da sunulan t testi sonuçları, iki grup arasında yürüyüş verilerinden elde edilen düzensizlik metrikleri açısından yapılan karşılaştırmayı göstermektedir.

Tablo 3.10 Sağ – sol ayak denge farkı metrikleri t-testi sonuçları

Metrik	**t İstatistiği**	**p Değeri**	**İstatistiksel Anlamlılık**
Anomali Sayısı	-0,993	0,322	Anlamlı değil ($p \geq 0,05$)
Anomali Şiddeti	5,019	$1,42 \times 10^{-6}$	Anlamlı ($p < 0,05$)
Varyans	-4,514	$1,24 \times 10^{-5}$	Anlamlı ($p < 0,05$)
Düzensizlik Skoru	3,773	0,00024	Anlamlı ($p < 0,05$)

Anomali Şiddeti, Varyans ve Düzensizlik Skorlarında istatistiksel olarak anlamlı farklar gözlemlenmiştir.

3.3.6 Gini Değeri

Gini indeksi, dağılımın eşitlik derecesini ölçen istatistiksel bir göstergedir. Bu çalışmada sensörlerden elde edilen kuvvet verilerinin dağılımını incelemek için uyarlanmıştır. Gini değeri, özellikle her bir ayaktaki basıncın farklı sensörler arasında ne ölçüde dengeli dağıldığını değerlendirmek amacıyla kullanılmıştır

Bu çalışmada her ayak tabanında sekiz adet dikey yer tepki kuvveti sensörü bulunmaktadır. Her bir zaman adımında, ilgili ayağın toplam kuvveti F{toplam} ile her sensörün ölçtüğü kuvvet F_I arasındaki oranlar pi hesaplanmıştır. Oranlar şu formülle bulunur:

$$pi(t) = \frac{F_{i(t)}}{F_{\{toplam\}(t)}} \tag{3.11}$$

Burada:

- $pi(t)$: i. sensörün t anındaki kuvvet oranı
- $F_{i(t)}$: i. sensörün t anındaki ölçtüğü kuvvet (N)
- $F_{\{toplam\}(t)}$: ilgili ayağın t anındaki toplam kuvveti (N)

Elde edilen bu oranlar ile Gini indeksi aşağıdaki formülle hesaplanmıştır:

$$G = 1 - \sum_{i=1}^{n} p_i^2 \tag{3.12}$$

Burada:

- G: Gini indeksini temsil eder.
- p_i: her sensörün toplam kuvvet içindeki oranıdır.
- n: sensörlerin toplam sayısıdır.

Formülde $\sum p_I^2$ ifadesi, yük dağılımındaki yoğunlaşmayı gösterir. Tüm sensörlerde eşit yük varsa karelerin toplamı yüksek olur ve gini değeri düşük çıkar. Yük belirli sensörlerde yoğunlaştığında ise kareler toplamı azalır ve gini değeri artar.

Gini değeri, [0, 1] aralığında değişir:

- 0 → Yük tüm sensörlere tamamen eşit dağılmıştır.
- 1 → Yük tamamen tek bir sensöre binmiştir.

Bu çalışmada Gini indeksi, hem sol hem de sağ ayak için ayrı ayrı hesaplanmıştır. Böylece adım sırasında yükün taban yüzeyine nasıl dağıldığı, dengeli mi yoksa belirli bölgelerde mi yoğunlaştığı incelenmiştir. Yüksek Gini değerleri, belirli sensör bölgelerinde kuvvet birikimi olduğunu; düşük değerler ise kuvvetin daha homojen dağıldığını göstermektedir.

Hesaplanan gini değerleri istatistiksel olarak analiz edilmiştir. Analiz sonuçları Tablo 3.11'de yer almaktadır.

Tablo 3.11 Gini değerleri

	Ortalama KG	**Ortalama PHG**	**StdSapma KG**	**StdSapma PHG**
Medyanlar				
GiniSol_Medyan	0.785733	0.799076	0.028634	0.024954
GiniSağ_Medyan	0.788413	0.797800	0.031525	0.023619
Ortalamalar				
GiniSol_Ortalama	0.745605	0.759449	0.044023	0.032026
GiniSağ_Ortalama	0.747112	0.759958	0.047721	0.033328
Std Sapma				
GiniSağ_StdSapma	0.251465	0.237053	0.029790	0.029991
GiniSol_StdSapma	0.253071	0.241579	0.031785	0.027722

PHG grubunda hem sol hem sağ ayakta Gini değerleri daha yüksektir. Bu, Parkinson hastalarında adım sırasında yükün belirli bölgelerde yoğunlaştığını ve denge problemlerinin olabileceğini düşündürür. KG grubunda ise yük daha homojen dağılmış ve bireyler arası varyasyon daha fazladır. Bu durum sağlıklı bireylerin daha çeşitli ve dengeli adım stratejileri kullandığını gösterebilir.

Gini indekslerinin medyan, ortalama ve standart sapma değerleri açısından, sağlıklı grup ile Parkinson hastası grup arasında istatistiksel olarak anlamlı farklar olup olmadığı bağımsız t-testi ile değerlendirilmiştir.

Tablo 3.12 KG ve PHG gini indeksi değerleri t-testi sonuçları

	t-istatistiği	**p-değeri**
GiniSol_Medyan	-3.138	0.0021
GiniSağ_Medyan	-2.110	0.0368
GiniSol_Mean	-2.247	0.0264
GiniSağ_Mean	-1.946	0.0540
GiniSağ_Std	3.073	0.0025
GiniSol_Std	2.434	0.0162

Tablo 3.12'de yer alan bağımsız t-testi sonuçları, sağlıklı grup (KG) ile Parkinson hastası grup (PHG) arasında Gini indeksinin çeşitli ölçümlerinde anlamlı farklılıklar olduğunu göstermektedir. Özellikle, GiniSol_Medyan (p=0.0021), GiniSağ_Medyan (p=0.0368), GiniSol_Mean (p=0.0264), GiniSağ_Std (p=0.0025) ve GiniSol_Std (p=0.0162) ölçümleri açısından gruplar arasında istatistiksel olarak anlamlı farklar bulunmuştur. Bu sonuçlar, Parkinson hastalarının hem sol hem sağ ayakta yük dağılımında farklılıklar olduğunu ve bu dağılımın daha düzensiz veya odaklanmış olduğunu göstermektedir.

Öte yandan, GiniSağ_Mean için bulunan p değeri (0.0540) sınırda bir anlamlılık göstererek, bu ölçümde de gruplar arası farkın olabileceğini ima etmektedir ancak geleneksel anlamda %5 anlamlılık sınırının biraz üzerindedir.

Bu farklılıklar, Parkinson hastalarında adım sırasında yükün daha belirli bölgelerde yoğunlaştığını ve dengenin sağlıklı bireylere kıyasla daha az dengeli olduğunu desteklemektedir.

3.4 Makine Öğrenmesi

Makine öğrenmesi, bilgisayarların verilerden öğrenerek karar verebilmesini sağlayan yapay zekâ alanlarından biridir. Bu yöntemler, özellikle karmaşık veri kümelerinde örüntülerin ortaya çıkarılması, sınıflandırma ve tahminleme problemlerinin çözülmesinde yaygın olarak kullanılmaktadır.

Bu çalışmada, elde edilen verilerin analizi ve sınıflandırılması için farklı makine öğrenmesi algoritmalarından yararlanılmıştır. Tercih edilen yöntemler hem doğruluk oranlarının yüksek olması hem de farklı veri tipleri üzerinde etkin çalışabilmeleri nedeniyle seçilmiştir. Makine öğrenmesi algoritmaları, öğrenme sürecinde verilerden bir model oluşturur ve bu modeli kullanarak yeni gelen veriler üzerinde tahmin veya sınıflandırma işlemleri gerçekleştirir.

Bu bölümde, öncelikle kullanılan makine öğrenmesi yöntemlerinin teorik temelleri ve çalışma prensipleri açıklanacaktır. Ardından, her bir modelin avantajları, dezavantajları ve tez kapsamında hangi amaçla kullanıldığı ayrıntılı olarak sunulacaktır. Çalışmada kullanılan yöntemler arasında Destek Vektör Makineleri, Yapay Sinir Ağları ve Lojistik Regresyon bulunmaktadır.

3.4.1 Destek Vektör Makineleri

Destek Vektör Makineleri, sınıflandırma, regresyon ve aykırı değer tespiti için kullanılan denetimli öğrenme yöntemlerinden oluşan bir algoritma grubudur [23]. DVM, verileri ayıran bir hiperdüzlem belirleyerek sınıflandırma yapar ve bu hiperdüzlemin sınıflar arasındaki marjını maksimize etmeyi hedefler. Böylece modelin genelleme kabiliyeti artar.

Destek vektör makinelerinin avantajları şunlardır:

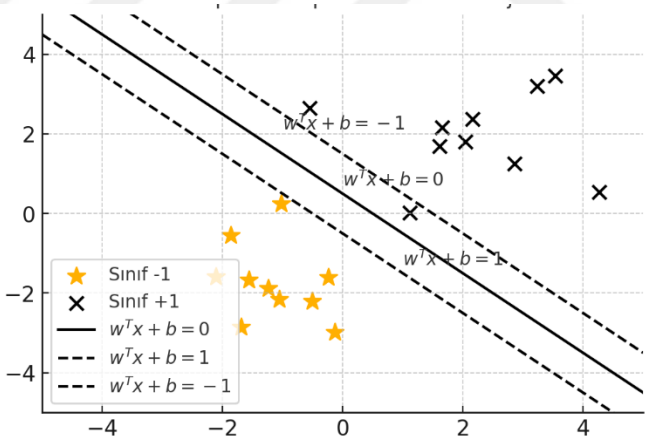
- Yüksek boyutlu uzaylarda etkilidir.
- Boyut sayısının örnek sayısından fazla olduğu durumlarda bile başarılı sonuçlar verebilir.
- Karar fonksiyonunda yalnızca eğitim noktalarının bir alt kümesini (destek vektörleri) kullanır; bu nedenle bellek açısından verimlidir.
- Esnektir: Karar fonksiyonu için farklı çekirdek (kernel) fonksiyonları belirtilebilir. Yaygın çekirdekler sunulmakla birlikte, özel çekirdekler tanımlamak da mümkündür.

Destek vektör makinelerinin dezavantajları şunlardır:

- Özellik (feature) sayısının örnek sayısına göre çok fazla olduğu durumlarda, çekirdek fonksiyonu ve düzenleme (regularization) teriminin seçiminde aşırı uyum (overfitting) riskini önlemek kritik öneme sahiptir.

DVM, Yapısal Risk Minimizasyonu ilkesini izleyerek genelleme yeteneğini artırır ve hem doğrusal hem de doğrusal olmayan, hatta kısmen ayrılabilir veri kümelerinde yüksek doğruluk sağlar.

Doğrusal Olarak Ayrılabilen Durum: Eğitim verileri iki sınıfa ait noktalar olarak verildiğinde, DVM sınıfları en geniş marj ile ayıran optimal hiperdüzlemi bulur [24]. Şekil 3.9'da gösterildiği üzere, hiper düzlem ile en yakın örnekler arasındaki fonksiyonel marj değeri 1 olarak varsayılmaktadır. Bu varsayım, destek vektörlerinin konumunu belirleyerek karar sınırının maksimum marj ile optimize edilmesini sağlar. Böylece DVM, sınıflar arasındaki ayrımı en iyi şekilde gerçekleştirecek hiper düzlemi oluşturur.



Şekil 3.9 Destek vektör makineleri optimal hiperdüzlem ve marjlar

Bu hiper düzlem şu şekilde ifade edilir:

$$w^T x + b = 0 \tag{3.13}$$

Burada:

- w ağırlık vektörü
- x girdi özellik vektörü
- b ise bias değeridir.

Fonksiyonel marj:

$$f(x) = w^T x + b \tag{3.14}$$

Eğer $f(x) > 0$, örnek pozitif sınıfa; $f(x) < 0$, negatif sınıfa ait olabilir.

Marj genişliği, iki sınıf arasındaki boşluğu ifade eder. DVM, bu genişliği maksimize etmeye çalışır. Yani sınıflar arasındaki ayrımı olabildiğince netleştirir.

Marj genişliği:

$$d = \frac{2}{||w||} \tag{3.15}$$

- $||w||$Ağırlık vektörünün normudur.

Optimizasyon problemi:

$$\min \frac{1}{2} ||w||^2 \quad s.t. \; y_i(w^T x_i + b) \geq 1, \quad i = 1, \ldots, n \tag{3.16}$$

Amaç fonksiyonu: $\frac{1}{2} ||w||^2$— marj genişliğini maksimize etmek için normu minimize eder. DVM, her sınıfın örneklerinin karar sınırından en az 1 birim uzaklıkta olmasını ister. Bu yüzden, her örnek için $y_i(w^T x_i + b) \geq 1$ kısıtı eklenir. Bu eşitsizlik, pozitif ve negatif sınıfların karar düzleminin iki yanında en az 1 birim mesafede olmasını garanti eder.

Doğrusal Olmayan Durum: Gerçek dünyada birçok problem doğrusal olarak ayrılamaz. Bu durumda DVM, veriyi çekirdek (kernel) fonksiyonları kullanarak yüksek boyutlu uzaya yansıtır:

$$f(x) = \sum_{i=1}^{n} \alpha_i y_i K(x_i, x) + b \tag{3.17}$$

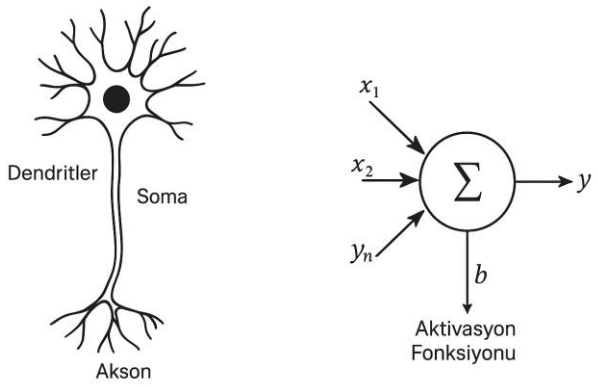
Yaygın kernel fonksiyonları:

- Polinomsal: $K(x, x_i) = (x \,.\, x_i)^q \qquad q = 1, 2, \ldots$
- Gauss : $K(x, x_i) = \exp\left(-\frac{|x - x_i|^2}{2\sigma^2}\right)$
- Çok Katmanlı Algılayıcı: $K(x, x_i) = \tanh(b(x \cdot x_i) - c)$

[24]

3.4.2 Yapay Sinir Ağları

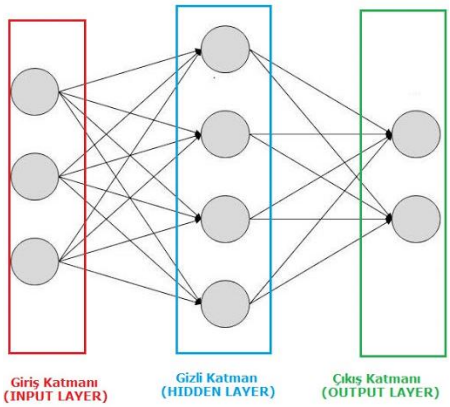
Yapay sinir ağları (YSA), insan beyninin bilgi işleme mekanizmasından esinlenerek geliştirilmiş, öğrenme ve genelleme yeteneklerine sahip hesaplama modelleridir. Biyolojik nöronların sinaptik bağlantılar kurarak bilgi işlediği prensip, YSA'larda matematiksel olarak modellenmiştir. Her bir yapay nöron, kendisine gelen sinyalleri birleştirir, aktivasyon fonksiyonu ile işler ve çıktıyı bir sonraki katmana iletir. Bu yapı sayesinde ağ, karmaşık örüntüleri öğrenebilir ve genelleyebilir [25]. Biyolojik nöron; dendritler (girdi alıcılar), soma (hücre gövdesi) ve akson (çıktı iletici) olmak üzere üç ana bölümden oluşur. Yapay nöronlar ise bu yapının matematiksel bir modellemesidir.



Şekil 3.10 Biyolojik ve yapay nöron karşılaştırması

Biyolojik nöron ve yapay nöron arasındaki bu modelleme, özellikle öğrenme ve adaptasyon süreçlerinin bilgisayar ortamında simüle edilmesine olanak sağlar. Yapay nöronlar, ağırlıklar ve bias değerleri aracılığıyla öğrenme kapasitesine sahiptir ve bu değerler eğitim süreci boyunca optimize edilir.

Yapay sinir ağları, nöronların katmanlar halinde düzenlenmesi ile oluşur. Bu katmanlar arasındaki bağlantı yoğunluğu ve katman sayısı, ağın kapasitesini ve karmaşık örüntüleri öğrenme yeteneğini doğrudan etkiler. Şekil 3.11'de Yapar sinir ağları katmanları yapısı gösterilmiştir.



Şekil 3.11 Yapay sinir ağları katmanları

1. Giriş Katmanı (Input Layer): Ağın dış dünyadan aldığı veri.
2. Gizli Katmanlar (Hidden Layers): Veriyi işleyen katmanlar, doğrusal olmayan özellik çıkarımı sağlar.
3. Çıkış Katmanı (Output Layer): Sonuçları üretir; sınıflandırma veya regresyon sonucudur.

Bir yapay nöronun çalışma prensibi aşağıdaki formül ile gösterilebilir:

$$y = f\left(\sum_{i=1}^{n} w_i x_i + b\right) \tag{3.18}$$

Burada:

- x_i: i. girdi
- w_i: i. girdiye ait ağırlık
- b: bias (eşik değeri)
- f (·): aktivasyon fonksiyonu (ör. sigmoid, ReLU, tanh)
- y: çıktı değeri

Bu formül, her bir yapay nöronun girdileri ağırlıklandırarak toplaması, buna bir bias eklemesi ve ardından aktivasyon fonksiyonundan geçirerek çıktıyı üretmesi sürecini ifade eder.

Aktivasyon fonksiyonları, yapay nöronun doğrusal olmayan davranış göstermesini sağlar [26]. En yaygın kullanılanlar:

Sigmoid Fonksiyonu:

$$\sigma(x) = \frac{1}{1 + e^{-x}} \tag{3.19}$$

Çıkışı [0,1] aralığında sınırlar ve özellikle sınıflandırma problemlerinde kullanılır.

Tanh Fonksiyonu:

$$\tanh(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}} \tag{3.20}$$

Çıkışı [-1,1] aralığındadır, orta değeri sıfır olduğu için bazı durumlarda daha hızlı öğrenme sağlar.

ReLU (Rectified Linear Unit):

$$f(x) = \max(0, x) \tag{3.21}$$

Modern derin ağlarda yaygın olarak kullanılır ve hesaplama açısından verimlidir. Bunun yanında, bazı nöronlar "ölü" hale gelebilir (dead neurons) ve bu nedenle ReLU türevleri (Leaky ReLU, Parametric ReLU) tercih edilebilir.

Eğitim sırasında, w_i ve b değerleri, hata geri yayılımı (backpropagation) yöntemi ile güncellenir.

Önemli bir hiperparametre olarak, learning rate (öğrenme hızı), bu güncellemenin ne kadar hızlı veya yavaş olacağını belirler. Ağırlık güncelleme formülü genellikle şöyle ifade edilir:

$$w_i \leftarrow w_i - \eta \frac{\partial L}{\partial w_i} \tag{3.22}$$

Burada η learning rate'i temsil eder. Çok yüksek bir learning rate, ağırlıkların aşırı değişmesine ve öğrenmenin kararsız hale gelmesine yol açabilir; çok düşük bir learning rate ise öğrenmenin yavaş ve verimsiz olmasına neden olabilir. Bu nedenle, learning rate seçimi ağın performansını doğrudan etkileyen kritik bir hiperparametredir.

Ağırlık güncelleme formülde L kayıp fonksiyonunu temsil eder. Ağ, çıktısı ile gerçek değer arasındaki farkı minimize etmeye çalışır. Bu fark, kayıp fonksiyonu (loss function) ile ölçülür. Örnek olarak MSE (Mean Squared Error) veya Cross-Entropy kullanılabilir.

Ağırlıklar ve bias değerleri, veri seti üzerinden birkaç epoch boyunca güncellenir. Her adımda ağırlıklar, belirli bir batch size kadar veri kullanılarak optimize edilir.

Regulasyon yöntemleri (dropout, L2/L1 cezaları) ağın aşırı öğrenmesini (overfitting) önlemeye yardımcı olur.

Yapay sinir ağları, günümüzde görüntü işleme, doğal dil işleme, ses tanıma ve tahmin modelleri gibi birçok alanda başarılı bir şekilde uygulanmaktadır. Ağın öğrenme kapasitesi, veri miktarı, katman derinliği ve kullanılan aktivasyon fonksiyonları gibi faktörlere bağlıdır. Bu nedenle, ağ tasarımı ve hiperparametre seçimi, performans üzerinde doğrudan etkilidir.

3.4.3 Lojistik Regresyon

Lojistik regresyon, bağımlı değişkenin ikili (binary) veya çok sınıflı kategorik olduğu durumlarda kullanılan, olasılık temelli bir sınıflandırma yöntemidir. Adına "regresyon" denmesine karşın, temel amacı sürekli değer tahmin etmek değil, bir olayın gerçekleşme olasılığını modellemektir [27].

Lojistik regresyonda amaç, girdilerin doğrusal kombinasyonunu logit fonksiyonu aracılığıyla 0 ile 1 arasında bir olasılık değerine dönüştürmektir. Bu dönüşüm, sigmoid (lojistik) fonksiyon kullanılarak yapılır:

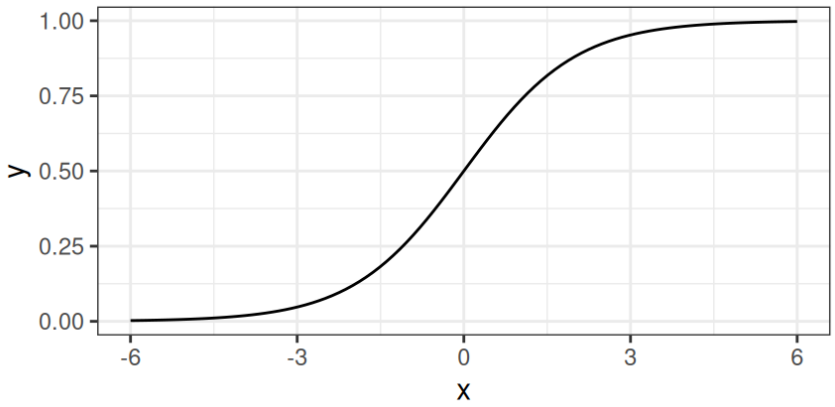
$$p(y = 1 \mid x) = \sigma(z) = \frac{1}{1 + e^{-z}} \tag{3.23}$$

$$z = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n \tag{3.24}$$

Burada:

- $p(y = 1 \mid x)$: Girdinin sınıf 1'e ait olma olasılığı
- $\sigma(z)$: Sigmoid fonksiyonu
- β_0: Sabit terim (bias)
- β_i: Model katsayıları
- x_i: Bağımsız değişkenler

Sigmoid fonksiyonu, Şekil 3.12'de gösterildiği gibi S-şeklinde bir eğriye sahiptir ve bu sayede negatif sonsuzdan pozitif sonsuza kadar olan z değerlerini (0,1) aralığına sıkıştırır.



Şekil 3.12 Lojistik Fonksiyon Eğrisi

Lojistik Fonksiyon Eğrisi giriş değerleri negatiften pozitife doğru arttıkça olasılığı 0'dan 1'e yumuşak bir şekilde dönüştürür ve böylece sınıflandırma kararlarını olasılık temelli olarak verir.

Lojistik regresyon, olasılıkların logaritmik oranını (odds ratio) modellemek için logit fonksiyonunu kullanır:

$$\text{logit}(p) = \ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_1 + \cdots + \beta_n x_n \tag{3.25}$$

Burada:

- $\text{logit}(p)$: Olasılığın logit dönüşümü
- $\ln\left(\frac{p}{1-p}\right)$: Odds oranının doğal logaritması
- $\beta_0 + \beta_1 x_1 + \cdots + \beta_n x_n$: Özelliklerin doğrusal kombinasyonu

Bu dönüşüm, doğrusal regresyonun tahmin gücünü olasılık temelli yorumlamaya imkân verir.

Lojistik regresyon katsayıları, En Yüksek Olabilirlik Tahmini (Maximum Likelihood Estimation, MLE) yöntemiyle hesaplanır. Amaç, gözlemlenen verilerin model tarafından üretilme olasılığını maksimize eden β katsayılarını bulmaktır [28].

Olasılık fonksiyonu:

$$L(\beta) = \prod_{i=1}^{m} p_i^{y_i}(1-p_i)^{1-y_i} \tag{3.26}$$

Burada:

- $L(\boldsymbol{\beta})$: Parametre vektörü β için likelihood
- p_i: Girdi için tahmin edilen olasılık
- y_i: Gerçek çıktı (0 veya 1)
- m: Gözlem sayısı

Logaritmik formu (log-likelihood):

$$\ell(\beta) = \sum_{i=1}^{m} [y_i \ln(p_i) + (1 - y_i) \ln(1 - p_i)] \quad (3.27)$$

Bu, lojistik regresyonun log-likelihood fonksiyonudur. Açıklaması:

- $\ell(\boldsymbol{\beta})$: Parametre vektörü β\boldsymbol{\beta} için log-likelihood
- y_i: Gerçek etiket (0 veya 1)
- p_i: Tahmin edilen olasılık
- m: Gözlem sayısı

Bu fonksiyon, sayısal optimizasyon yöntemleri kullanılarak maksimize edilir.

Lojistik Regresyonun Avantajları:

- Yorumlanabilirlik: Katsayılar, değişkenlerin log-odds üzerindeki etkisini doğrudan gösterir.
- Olasılık çıktısı üretir, bu da karar eşiklerinin esnek belirlenmesini sağlar.
- Hesaplama açısından verimlidir.

Lojistik Regresyonun Dezavantajları:

- Doğrusal ayrılabilirlik varsayımı: Özellikler ile logit arasındaki ilişki doğrusal olmalıdır.
- Çok karmaşık, doğrusal olmayan ilişkileri modelleyemez (bu durumda polinom özellikler veya kernel yöntemleri gerekir).

Lojistik regresyonun temel avantajı, katsayıların hem matematiksel hem de pratik olarak yorumlanabilir olmasıdır; ayrıca olasılık çıktısı üretmesi nedeniyle karar eşiklerini esnek belirleme imkânı sunar. Dezavantajları ise doğrusal varsayımın güçlü olması ve karmaşık ilişkileri tek başına modelleyememesidir.

Lojistik regresyonun yaygın kullanım alanları arasında tıp (hastalık teşhisi), pazarlama (müşteri davranışı tahmini), kredi riski değerlendirmesi ve sosyal bilimler (anket sonuçları analizi) yer alır. Örneğin, bir hastanın belirli bir hastalığa sahip olma olasılığı lojistik regresyon modeli ile 0–1 aralığında tahmin edilebilir.

3.5 Değerlendirme Ölçütleri

Makine öğrenmesi modellerinin performansını değerlendirmek, modelin doğruluğunu ve genelleme yeteneğini anlamak için kritik bir adımdır. Makine öğrenmesi modellerinin başarısını ölçmek için çeşitli değerlendirme ölçütleri kullanılmaktadır. Bu ölçütler, modelin doğru ve yanlış tahminlerini analiz ederek performansını objektif bir şekilde ortaya koyar. Bu amaçla, doğruluk (accuracy), kesinlik (precision), duyarlılık (recall), F1 skoru gibi ölçütler yaygın olarak kullanılmaktadır.

Karışıklık matrisi, sınıflandırma modelinin tahmin performansını görselleştiren temel bir araçtır. Matriste dört temel değer yer alır [29].

Tablo 3.13 Karışıklık matrisi

Tahmin \ Gerçek	**Pozitif**	**Negatif**
Pozitif	TP (Gerçek Pozitif)	FP (Yanlış Pozitif)
Negatif	FN (Yanlış Negatif)	TN (Gerçek Negatif)

- True Positive (TP): Modelin pozitif sınıfı doğru şekilde tahmin ettiği örnek sayısı.
- True Negative (TN): Modelin negatif sınıfı doğru şekilde tahmin ettiği örnek sayısı.
- False Positive (FP): Modelin negatif örneği yanlışlıkla pozitif olarak sınıflandırdığı örnek sayısı.
- False Negative (FN): Modelin pozitif örneği yanlışlıkla negatif olarak sınıflandırdığı örnek sayısı.

Bu dört değer, doğruluk, kesinlik, duyarlılık ve F1 skoru gibi metriklerin hesaplanmasında temel alınır.

Doğruluk (Accuracy), modelin tüm tahminler içinde doğru olanların oranını verir:

$$\text{Doğruluk (Accuracy)} = \frac{TP + TN}{TP + TN + FP + FN} \tag{3.28}$$

Doğruluk, dengesiz veri setlerinde yanıltıcı olabilir. Örneğin, %95 negatif ve %5 pozitif örnek içeren bir veri setinde, tüm örnekleri negatif tahmin eden bir model %95 doğruluk elde eder ama pozitifleri hiç yakalayamaz. Bu nedenle, özellikle sağlık gibi kritik alanlarda Precision, Recall ve F1 Score gibi metriklerle birlikte değerlendirilmesi gerekir.

Kesinlik (precision), modelin pozitif olarak tahmin ettiği örneklerden gerçekte doğru olanların oranını ifade eder:

$$\text{Kesinlik (Precision)} = \frac{TP}{TP + FP} \tag{3.29}$$

Kesinlik, özellikle yanlış pozitiflerin (FP) maliyetinin yüksek olduğu uygulamalarda kritik öneme sahiptir. Örneğin, sağlık alanında bir hastalık teşhis modelinin bir bireyi hasta olarak sınıflandırması, birey üzerinde psikolojik stres yaratabilir ve gereksiz tıbbi müdahalelere yol açabilir. Bu bağlamda, yüksek kesinlik değerleri, modelin pozitif tahminlerinin güvenilirliğini artırır.

Kesinlik metriği, genellikle duyarlılık (recall) ile değerlendirilir. Çünkü bir modelin yalnızca doğru pozitifleri tahmin etmesi yeterli değildir; aynı zamanda mümkün olduğunca çok sayıda gerçek pozitifi de yakalayabilmesi beklenir.

Duyarlılık (recall), modelin gerçek pozitif örnekleri ne kadar iyi yakaladığını gösterir. Yani, tüm pozitif örneklerden modelin kaçını doğru şekilde sınıflandırdığıdır:

$$\text{Duyarlılık (Recall)} = \frac{TP}{TP + FN} \tag{3.30}$$

Recall, özellikle yanlış negatiflerin (FN) kritik sonuçlar doğurabileceği uygulamalarda büyük önem taşır. Sağlık alanında geliştirilen bir teşhis modelinde, hastalıklı bireylerin sağlıklı olarak sınıflandırılması (FN) tedavi sürecinin gecikmesine, hastalığın ilerlemesine ve bireysel ya da toplumsal risklerin artmasına neden olabilir. Bu nedenle, yüksek duyarlılık değeri, modelin gerçek pozitifleri gözden kaçırmadan yakalayabildiğini gösterir.

F1 Skoru, Precision ve Recall değerlerinin harmonik ortalamasıdır ve modelin dengeli performansını gösterir. Harmonik ortalama, iki değerin birbirine yakın olmasını teşvik eder; dolayısıyla F1 Skoru, yalnızca biri yüksekken değil, her ikisi de yüksek olduğunda yüksek değer alır.

$$F1Score = 2\frac{Precision \cdot Recall}{Precision + Recall} \tag{3.31}$$

Makine öğrenmesi modellerinin değerlendirilmesinde tek bir metrik yeterli olmayabilir. Özellikle sağlık gibi yüksek riskli alanlarda, birden fazla ölçütün birlikte kullanılması, modelin genelleme yeteneğini ve güvenilirliğini daha doğru şekilde ortaya koyar.

4
UYGULAMA VE BULGULAR

4.1 Model Eğitimi ve Test Süreci

Bu çalışmada denge verilerinin sınıflandırılması amacıyla üç farklı makine öğrenmesi modeli kullanılmıştır: Destek Vektör Makineleri (DVM), Yapay Sinir Ağları (YSA) ve Lojistik Regresyon.

Veri analizi ve makine öğrenmesi modelleri Python programlama dili ve özellikle Pandas, Scikit-learn ve Matplotlib kütüphaneleri kullanılarak gerçekleştirilmiştir [30]. Modelleme sürecinde kullanılan öznitelikler şunlardır: Yaş, Sol Ayak Toplam Kuvvet Permütasyon Entropisi, Sağ Ayak Toplam Kuvvet Permütasyon Entropisi, Sol Ayak Toplam Kuvvet Negentropisi, Sağ Ayak Toplam Kuvvet Negentropisi, Hız, Sol Ayak Minimum Tepe, Sağ Ayak Minimum Çukur, Anomali Şiddeti, Varyans, Düzensizlik Skoru, GiniSol_Ortalama, GiniSağ_Ortalama, GiniSol_Medyan, GiniSağ_Medyan, GiniSol_Std, GiniSağ_Std,. Veri setindeki hedef değişken olan grup sütunu, “KG” ve “PHG” sınıflarını temsil etmekte olup, sırasıyla 0 ve 1 değerleri ile sayısallaştırılmıştır. Tüm modeller için veri seti, %80 eğitim ve %20 test olacak şekilde ayrılmıştır.

DVM modeli, scikit-learn kütüphanesindeki SVC sınıfı kullanılarak varsayılan RBF (Radial Basis Function) çekirdeği ile eğitilmiştir. Model, eğitim seti üzerinde fit edilmiş ve test setinde sınıflandırma performansı doğruluk, kesinlik, duyarlılık, f1 skor metrikleri ile değerlendirilmiştir.

YSA modeli, TensorFlow Keras kütüphanesi kullanılarak oluşturulmuştur. Giriş katmanında 64 nöron, ara katmanlarda sırasıyla 32, 16 ve 4 nöron bulunmakta, her

katmanda ReLU aktivasyon fonksiyonu ve L2 düzenlileştirme uygulanmaktadır. Aşırı öğrenmeyi (overfitting) önlemek amacıyla her ara katmanda Dropout oranı %50 olarak belirlenmiştir. Çıkış katmanında ise ikili sınıflandırma için sigmoid aktivasyon fonksiyonu kullanılmıştır. Model, Adam optimizasyon algoritması (öğrenme oranı 0.001) ile, binary cross-entropy kayıp fonksiyonu üzerinden 75 epoch boyunca, 55 batch boyutunda eğitilmiştir.

Lojistik regresyon modeli, scikit-learn kütüphanesindeki LogisticRegression sınıfı ile oluşturulmuş, maksimum iterasyon sayısı 1000 olarak ayarlanmıştır. Modelin eğitiminde tüm özellikler standart ölçeklendirilmiş veriler üzerinden kullanılmıştır.

Her üç model de aynı eğitim-test ayrımı ile çalıştırılmış, sonuçlar doğruluk (accuracy), kesinlik (precision), duyarlılık (recall), F1 skoru ve karışıklık matrisi üzerinden değerlendirilmiştir. Bu sayede farklı algoritmaların aynı veri seti üzerindeki performansları objektif olarak karşılaştırılmıştır.

4.2 Deneysel Sonuçlar

Bu bölümde, çalışmada kullanılan üç farklı makine öğrenmesi modelinin eğitim ve test süreçleri sonucunda elde edilen performans çıktıları sunulmaktadır. Elde edilen sonuçlar, her modelin sınıflandırma görevindeki başarımını ayrıntılı biçimde ortaya koymakta ve modellerin veri seti üzerindeki genel davranışlarını gözler önüne sermektedir. Sunulan tablolar, her modelin farklı sınıflardaki performansını özetlemekte ve sonuçların sayısal bir bakışla değerlendirilmesini sağlamaktadır.

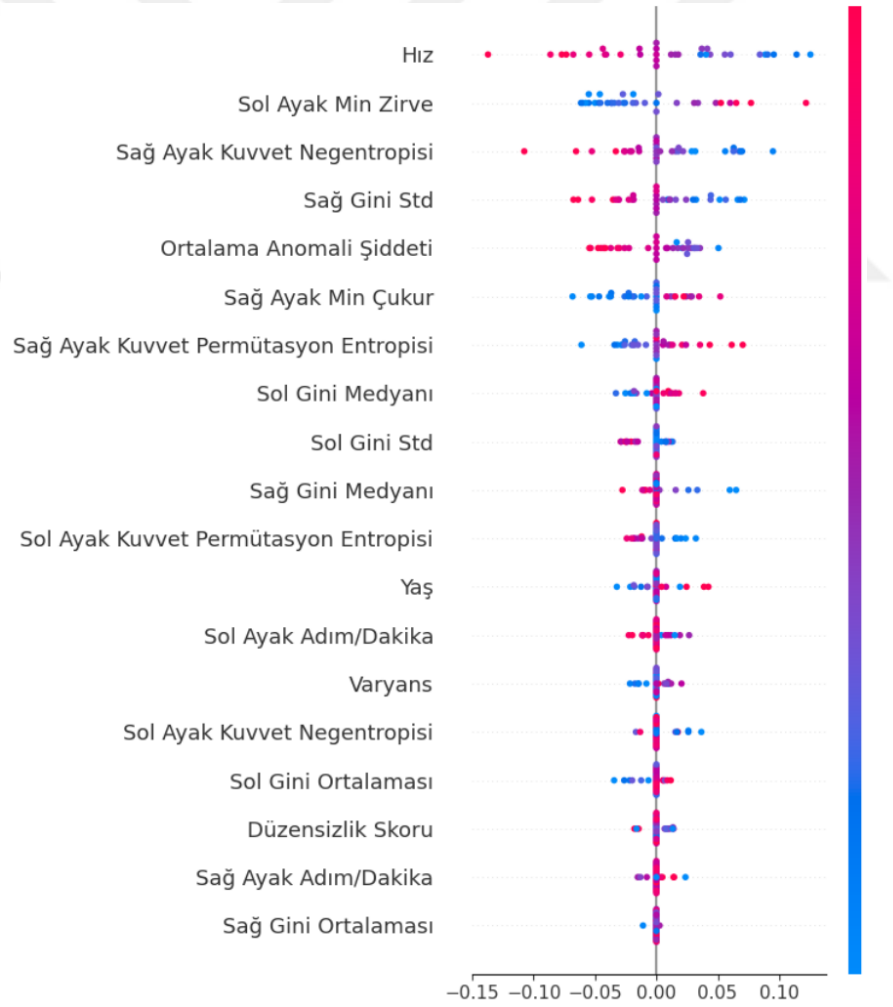
Tablo 4.1 Destek vektör makineleri sınıflandırma sonuçları

	KG	**PHG**
Kesinlik (Precision)	0.75	0.90
Duyarlılık (Recall)	0.82	0.86
F1-Skor	0.78	0.88
Örnek Sayısı	11	22
Doğruluk (Accuracy)	0.85	

Destek vektör makineleri modeli için sınıflandırma raporu Tablo 4.1'de verilmiştir. KG sınıfı için modelin kesinlik değeri 0,75 olarak belirlenmiştir; bu, modelin KG sınıfı için tahmin ettiği pozitif örneklerin %75'inin doğru olduğunu göstermektedir.

Duyarlılık değeri 0,82 olup, modelin gerçek KG örneklerinin %82'sini doğru şekilde tanıdığı anlamına gelmektedir. KG sınıfı için F1 skoru ise 0,78 olarak hesaplanmıştır. PHG sınıfında ise kesinlik 0,90, duyarlılık 0,86 ve F1 skoru 0,88 olarak gözlemlenmiştir. Toplam doğruluk oranı 0,85 olarak bulunmuştur.

Şekil 4.1 SHAP (SHapley Additive exPlanations) yöntemiyle elde edilen öznitelik katkılarını göstermektedir. Yatay eksen, her bir öznitelik için SHAP değerini (model çıktısına etkisini) temsil ederken; dikey eksen, modelde kullanılan öznitelikleri sıralamaktadır. Her bir nokta, bir örneğe ait SHAP değerini ve öznitelik değerini (renk skalası: düşük → mavi, yüksek → kırmızı) göstermektedir. Şekil 4.1'e göre "Hız", "Sol Ayak Min Zirve" ve "Sağ Ayak Kuvvet Negentropisi" gibi özniteliklerin DVM model kararında daha belirleyici olduğu görülmektedir.



Şekil 4.1 DVM özniteliklerin tahmin üzerindeki etkisi

DVM modelinde özniteliklerin model çıktısına olan ortalama etkileri Tablo 4.2'de sunulmuştur.

Tablo 4.2 DVM modeli öznitelik etkileri

Öznitelik	**Ortalama SHAP Değeri**
Hız	0.049634
Sol Ayak Min Zirve	0.039632
Sağ Ayak Kuvvet Negentropisi	0.030688
Sağ Gini Std	0.028323
Ortalama Anomali Şiddeti	0.025222
Sağ Ayak Min Çukur	0.021297
Sağ Ayak Kuvvet Permütasyon Entropisi	0.019195
Sol Gini Medyanı	0.009135
Sol Gini Std	0.008958
Sağ Gini Medyanı	0.008053
Sol Ayak Kuvvet Permütasyon Entropisi	0.007993
Yaş	0.007698
Sol Ayak Adım/Dakika	0.005323
Varyans	0.005180
Sol Ayak Kuvvet Negentropisi	0.005105
Sol Gini Ortalaması	0.004845
Düzensizlik Skoru	0.004041
Sağ Ayak Adım/Dakika	0.002915
Sağ Gini Ortalaması	0.000409

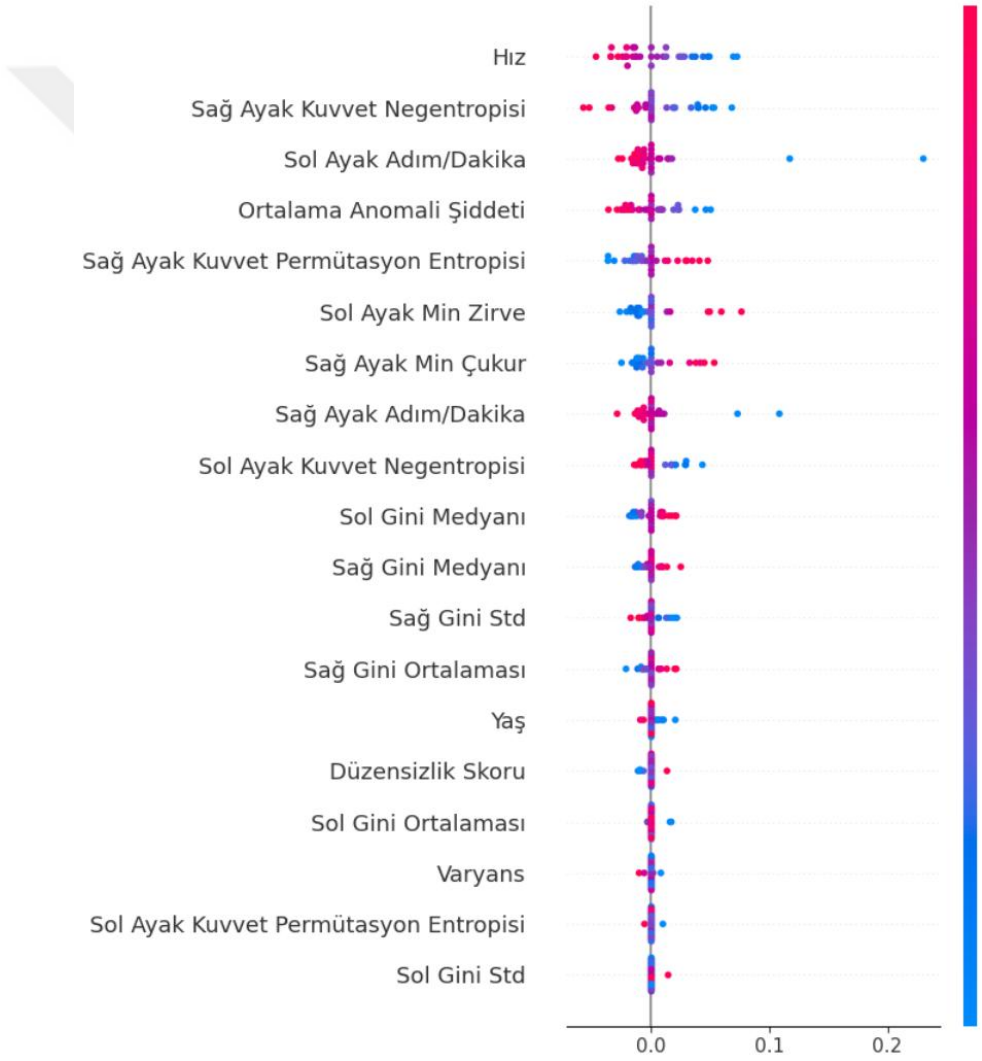
Tablo 4.3 Yapay sinir ağları sınıflandırma sonuçları

	KG	**PHG**
Kesinlik (Precision)	0.82	0.91
Duyarlılık (Recall)	0.82	0.91
F1-Skor	0.82	0.91
Örnek Sayısı	11	22
Doğruluk (Accuracy)	0.88	

Yapay sinir ağları modeli için elde edilen sınıflandırma sonuçları Tablo 4.2'de özetlenmiştir. KG sınıfı için hem kesinlik hem de duyarlılık 0,82 olarak

ölçülmüştür; bu da modelin KG örneklerini doğru tanımada dengeli bir performans sergilediğini göstermektedir. KG sınıfının F1 skoru da 0,82'dir. PHG sınıfında kesinlik, duyarlılık ve F1 skoru 0,91 olarak gözlemlenmiştir. Bu modelin toplam doğruluk değeri 0,88 olup, eğitim ve test verileri üzerinde başarılı bir sınıflandırma performansı sergilediği görülmektedir.

Şekil 4.2'ye göre, özellikle "Hız" ve "Sağ Ayak Kuvvet Negentropisi" öznitelikleri, modelin tahmin sürecine en yüksek katkıyı sağlayan öznitelikler arasında yer almaktadır



Şekil 4.2 YSA özniteliklerin tahmin üzerindeki etkisi

YSA modelinde özniteliklerin model çıktısına olan ortalama etkileri Tablo 4.4'de sunulmuştur.

Tablo 4.4 YSA modeli öznitelik etkileri

Öznitelik	Ortalama SHAP Değeri
Hız	0.025411
Sağ Ayak Kuvvet Negentropisi	0.021016
Sol Ayak Adım/Dakika	0.019242
Ortalama Anomali Şiddeti	0.016409
Sağ Ayak Kuvvet Permütasyon Entropisi	0.015707
Sol Ayak Min Zirve	0.014616
Sağ Ayak Min Çukur	0.013714
Sağ Ayak Adım/Dakika	0.011130
Sol Ayak Kuvvet Negentropisi	0.010815
Sol Gini Medyanı	0.008886
Sağ Gini Medyanı	0.005638
Sağ Gini Std	0.004982
Sağ Gini Ortalaması	0.004829
Yaş	0.002154
Düzensizlik Skoru	0.001521
Sol Gini Ortalaması	0.001182
Varyans	0.000784
Sol Ayak Kuvvet Permütasyon Entropisi	0.000471
Sol Gini Std	0.000430

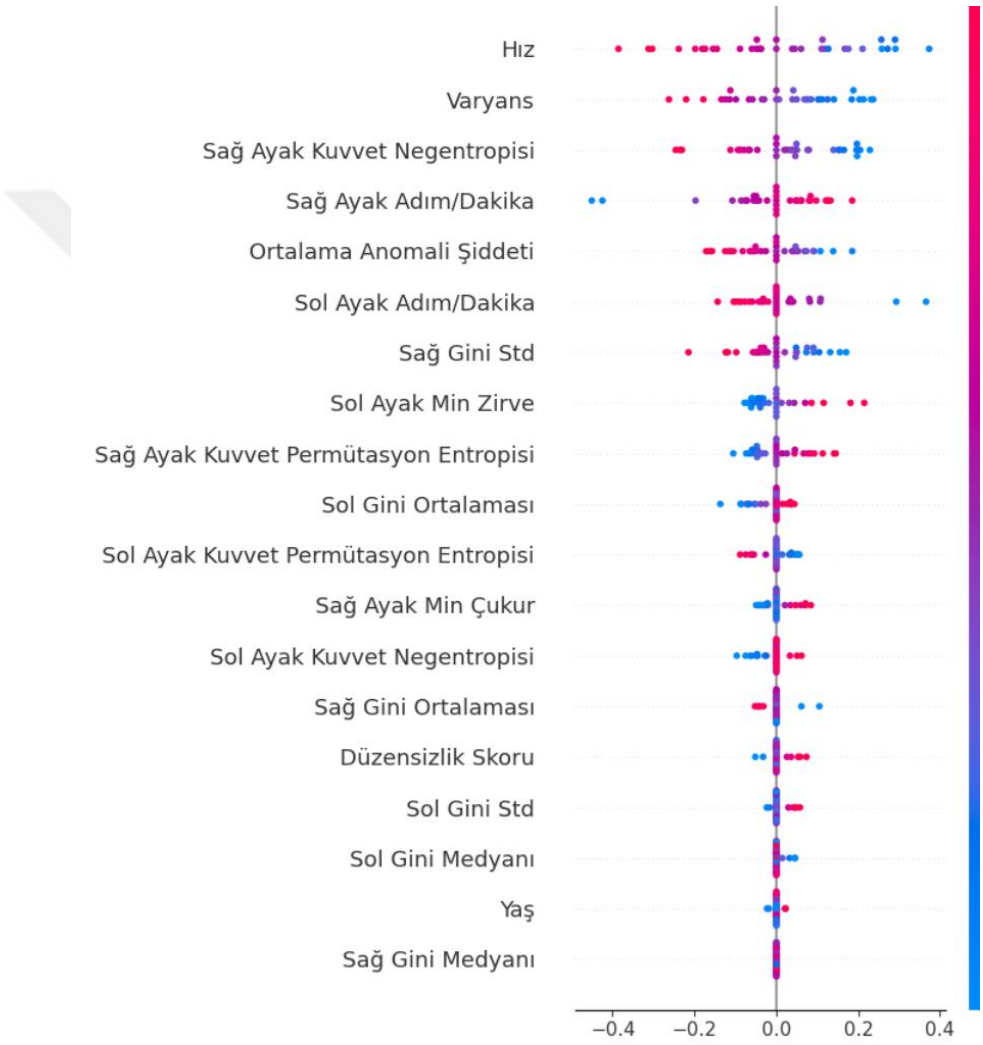
Tablo 4.5 Lojistik regresyon sınıflandırma sonuçları

	KG	PHG
Kesinlik (Precision)	0.71	0.95
Duyarlılık (Recall)	0.91	0.82
F1-Skor	0.80	0.88
Örnek Sayısı	11	22
Doğruluk (Accuracy)	0.85	

Lojistik regresyon modeli için sınıflandırma raporu Tablo 4.3'te gösterilmiştir. KG sınıfı için kesinlik 0,71, duyarlılık 0,91 ve F1 skoru 0,80 olarak belirlenmiştir. Bu

sonuçlar, modelin KG örneklerini tanımada yüksek bir duyarlılık sağladığını, ancak yanlış pozitif tahminlerde bazı hatalar yaptığını göstermektedir. PHG sınıfında ise kesinlik 0,95, duyarlılık 0,82 ve F1 skoru 0,88'dir. Toplam doğruluk oranı 0,85 olarak bulunmuştur.

Şekil 4.3'e göre, özellikle "Hız" ve "Varyans" öznitelikleri, modelin tahmin sürecine en yüksek katkıyı sağlayan öznitelikler arasında yer almaktadır.



Şekil 4.3 Lojistik regresyon özniteliklerin tahmin üzerindeki etkisi

Lojistik regresyon modelinde özniteliklerin model çıktısına olan ortalama etkileri Tablo 4.6'da sunulmuştur.

Tablo 4.6 Lojistik regresyon modeli öznitelik etkileri

Öznitelik	Ortalama SHAP Değeri
Hız	0.161327
Varyans	0.112753
Sağ Ayak Kuvvet Negentropisi	0.102121
Sağ Ayak Adım/Dakika	0.086325
Ortalama Anomali Şiddeti	0.070434
Sol Ayak Adım/Dakika	0.062636
Sağ Gini Std	0.061067
Sol Ayak Min Zirve	0.048652
Sağ Ayak Kuvvet Permütasyon Entropisi	0.048385
Sol Gini Ortalaması	0.026540
Sol Ayak Kuvvet Permütasyon Entropisi	0.022556
Sağ Ayak Min Çukur	0.019927
Sol Ayak Kuvvet Negentropisi	0.019234
Sağ Gini Ortalaması	0.010150
Düzensizlik Skoru	0.010032
Sol Gini Std	0.007992
Sol Gini Medyanı	0.004164
Yaş	0.002565
Sağ Gini Medyanı	0.000000

4.3 Modellerin Karşılaştırılması

Bu bölümde, önceki bölümde sunulan üç farklı makine öğrenmesi modelinin performans sonuçları karşılaştırmalı olarak ele alınacaktır. Modellerin doğruluk, kesinlik, duyarlılık ve F1 skoru gibi metrikler üzerinden değerlendirilmesi, hangi modelin veri seti üzerinde daha dengeli ve etkili çalıştığını ortaya koymaktadır. Karşılaştırma, her modelin farklı sınıflardaki başarımı ve genel sınıflandırma performansı dikkate alınarak yapılacaktır. Bu sayede, modellerin güçlü ve sınırlı yönleri daha net bir şekilde değerlendirilebilecektir.

- **KG Sınıfı:** YSA modeli, KG sınıfında %82 kesinlik ve %82 duyarlılık ile dengeli bir performans sunmaktadır. DVM modeli de iyi bir performans sergilemiş olsa da (%75 kesinlik, %82 duyarlılık), YSA'ya göre pozitif tahminlerde biraz daha hatalıdır. Lojistik regresyon modeli yüksek duyarlılık (0,91) ile KG örneklerini çoğunlukla doğru tanısa da daha düşük kesinlik (0,71) nedeniyle yanlış pozitif tahminler daha fazladır.
- **PHG Sınıfı:** Lojistik regresyon modeli PHG sınıfında %95 kesinlik ile en iyi performansı göstermektedir. YSA modeli %91 kesinlik ve duyarlılık ile güçlü bir performans sunarken, DVM modeli %90 kesinlik ve %86 duyarlılık ile biraz daha düşük bir başarı sergilemiştir.
- **Genel Doğruluk:** YSA modeli 0,88 ile en yüksek genel doğruluk oranına sahiptir. DVM ve lojistik regresyon modelleri ise 0,85 doğruluk ile benzer bir genel performans göstermektedir. Bu durum, YSA'nın veri setindeki karmaşık ilişkileri daha iyi öğrenebildiğini ve hem sınıf bazında hem de genel doğruluk açısından daha dengeli bir performans sağladığını ortaya koymaktadır.

Bu karşılaştırma, modellerin yalnızca genel doğruluk değil, sınıf bazında performans metrikleri üzerinden de değerlendirilmesinin önemini vurgulamaktadır. Sınıf dengesizlikleri ve farklı performans metrikleri, model seçiminde kritik rol oynamaktadır. Özetle, YSA modeli genel olarak daha dengeli ve yüksek performanslı bir sınıflandırma sağlarken, lojistik regresyon ve DVM belirli sınıflarda güçlü, bazı sınıflarda sınırlı performans göstermektedir.

5
SONUÇ

5.1 Genel Sonuçlar

Bu tez çalışmasında, Parkinson hastalığının tanısına yönelik olarak yürüyüş verileri üzerinden sınıflandırma yapan üç farklı makine öğrenmesi algoritması (Destek Vektör Makineleri, Yapay Sinir Ağları ve Lojistik Regresyon) karşılaştırmalı olarak değerlendirilmiştir. Her bir model, özenle seçilmiş öznitelikler kullanılarak eğitilmiş ve test edilmiştir. Elde edilen sonuçlar, modellerin hem genel doğruluk hem de sınıf bazlı metrikler açısından farklı performans özellikleri sergilediğini ortaya koymuştur.

Yapay Sinir Ağı modeli hem KG hem de PHG sınıflarında dengeli ve yüksek performans göstererek en yüksek doğruluk oranına (%88) ulaşmıştır. Bu sonuç, modelin hem kontrol grubu hem de Parkinson hastası bireyleri ayırt etmede tutarlı ve güvenilir bir yapı sunduğunu göstermektedir. DVM ve Lojistik Regresyon modelleri ise bazı sınıflarda daha yüksek duyarlılık veya kesinlik değerleri sunmuş olsa da genel doğruluk açısından YSA modelinin gerisinde kalmıştır.

Çalışmada kullanılan öznitelikler—özellikle entropi, negentropi, adım sayısı, varyans gibi yürüyüş dinamikleri—sınıflar arasında anlamlı ayrımlar oluşturmuş ve modellerin öğrenme başarısını artırmıştır. Bu durum, yürüyüş verilerinin Parkinson hastalığına ilişkin tanısal potansiyelini bir kez daha ortaya koymaktadır.

Sonuç olarak, bu tez çalışması hem veri ön işleme hem de modelleme açısından bütüncül bir yaklaşım sunmakta; Parkinson hastalığının erken tanısına yönelik dijital sağlık çözümleri geliştirilmesinde önemli bir katkı sağlamaktadır.

KAYNAKÇA

[1] Dokuz Eylül Üniversitesi Araştırma Uygulama Hastanesi, "Prof. Dr. Raif Çakmur: Parkinson hastalığında en önemli sorun dopamin eksikliği," [Çevrimiçi]. Erişim adresi: https://hastane.deu.edu.tr/index.php/kurumsal/46-hastanemizden-haberler/2021-prof-dr-raif-cakmur-parkinson-hastaliginda-en-onemli-sorun-dopamin-eksikligi.html. [Erişim tarihi: 25 Nisan 2024].

[2] Türk Nöroloji Derneği ve Türkiye Parkinson Hastalığı Derneği, "Hareket Bozuklukları Tanı ve Tedavi Rehberi," [Çevrimiçi]. Erişim adresi: https://noroloji.org.tr/TNDData/Uploads/files/parkinson_iki%20renk%20ayarl%C4%B1%20(1).pdf . [Erişim tarihi: 25 Nisan 2024].

[3] Y. Guo, J. Yang, Y. Liu, X. Chen ve G. Yang, "Detection and assessment of Parkinson's disease based on gait analysis: A survey", Frontiers in Aging Neuroscience, vol. 14, 2022

[4] J. Jankovic, "Parkinson's disease: clinical features and diagnosis," *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 79, no. 4, pp. 368–376, 2008.

[5] Y. Wu, S. Krishnan, "Statistical Analysis of Gait Rhythm in Patients With Parkinson's Disease", IEEE Transactions On Neural Systems And Rehabilitatıon Engineering, vol. 18, no. 2, 2010

[6] S. Lee, J. S. Lim, "Parkinson's disease classification using gait characteristics and wavelet-based feature extraction", Expert Systems with Applications, vol. 39, pp. 7338-7344, 2012.

[7] S. Perumal, R. Sankar, "Gait And Tremor Assessment For Patients With Parkinson's Disease Usingwearable Sensors", ICT Express 2(4), pp. 168-174, 2016.

[8] O.C. Yurdakul, M.S.P. Subathras, S.T. Georges, "Detection of Parkinson's Disease from Gait Using Neighborhood Representation Local Binary Patterns", Biomedical Signal Processing and Control, vol. 62, 2020.

[9] F. Wahid, R.K. Begg, C.J. Hass, S. Halgamuge, D.C. Ackland, "Classification of Parkinson's Disease Gait Using Spatial-Temporal Gait Features", IEEE Journal of Biomedical and Health Informatics, vol. 19, no: 6, 2015

[10] M.N. Alam, A. Garg, T.T.K. Munia, R. Fazel-Rezai, K. Tavakolian, "Vertical Ground Reaction Force Marker For Parkinson's Disease", PLoS ONE 12(5): e0175951, 2017.

[11] D. Joshi, A. Khajuria, P. Joshi, "An automatic non-invasive method for Parkinson's disease classification", Computer Methods and Programs in Biomedicine ,vol. 145, Pages 135-145, 2017.

[12] B. Elumalai, B.Dass, B. Ramasamy "Supervised machine learning based gait classification system for early detection and stage classification of Parkinson's disease", Applied Soft Computing, vol. 94, 2020.

[13] A. Zhao, L. Qi, J. Li, J. Dong, H.Yu, "A hybrid spatio-temporal model for detection and severity rating of Parkinson's disease from gait data", Neurocomputing, vol. 315, pp. 1-8, 2018.

[14] H.H. Manap, N.M. Tahir, A.İ.M. Yassin, R. Abdullah, "Anomaly gait classification of Parkinson disease based on ANN", System Engineering and Technology (ICSET) ,2011 IEEE International Conference on, 2011

[15] H.H. Manap, N.M. Tahir, A.İ.M. Yassin, "Statistical Analysis Of Parkinson Disease Gait Classification Using Artificial Neural Network", 2011 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT), 2011

[16] H.H. Manap, N.M. Tahir, "Parkinson Disease Gait Classification based on Machine Learning Approach", Journal of Applied Sciences 12(2):180-185, 2012

[17] Goldberger, A., Amaral, L., Glass, L., Hausdorff, J., Ivanov, P. C., Mark, R., ... & Stanley, H. E. (2000). PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. Circulation [Çevrimiçi]. 101 (23), pp. e215–e220. RRID:SCR_007345. (https://physionet.org/content/gaitpdb/1.0.0/)

[18] Bandt, C., & Pompe, B. (2002). Permutation entropy: A natural complexity measure for time series. Physical Review Letters, 88(17), 174102.

[19] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*. New York, NY, USA: John Wiley & Sons, 2001.

[20] S. Watts and L. Crow, "The Shannon entropy of a histogram," *arXiv preprint arXiv:2210.02848*, Oct. 2022.

[21] "Normal gait," *Musculoskeletal Key*, [Çevrimiçi]. Erişim adresi: https://musculoskeletalkey.com/normal-gait/?utm_source=chatgpt.com. [Erişim tarihi: 15 Ağustos 2025].

[22] D. Makowski, T. Pham, Z. J. Lau, J. C. Brammer, F. Lespinasse, H. Pham, C. Schölzel, ve S. A. Chen, "NeuroKit2: A Python toolbox for neurophysiological signal processing," Behavior Research Methods, cilt 53, no. 4, pp. 1689–1696, 2021. doi: 10.3758/s13428-020-01516-y.

[23] Scikit-learn, "Support Vector Machines," 2025.

[24] H. Wang, J. Xiong, Z. Yao, M. Lin ve J. Ren, "Research Survey on Support Vector Machine," *EAI International Conference on Mobile Multimedia Communications*, 2017.

[25] Vikipedi, "Yapay sinir ağları," [Çevrimiçi]. Erişim adresi: https://tr.wikipedia.org/wiki/Yapay_sinir_a%C4%9Flar%C4%B1. [Erişim tarihi: 20 Ağustos 2025].

[26] M. Nielsen, *Neural Networks and Deep Learning*, 2015.

[27] D. W. Hosmer, S. Lemeshow ve R. X. Sturdivant, *Applied Logistic Regression*. Wiley, 2013.

[28] G. James, D. Witten, T. Hastie ve R. Tibshirani, *An Introduction to Statistical Learning with Applications in R*. Springer, 2021.

[29] Google Developers, “Accuracy, Precision, Recall,” 2025.

[30] Python Software Foundation, *Python Language Reference, version 3.x*, 2025.

TEZDEN ÜRETİLMİŞ YAYINLAR

Konferans Bildirileri

1. E. G. Yılmaz ve G. Gölbaşı Şimşek, "Yapay Zekâ Teknikleri Kullanılarak Parkinson Hastalığının Teşhisi," Grad Symposium 2025 – Mühendislikte Sürdürülebilir Çözümler Sempozyumu, İstanbul, Türkiye, Mayıs 2025.