



İZMİR BAKIRÇAY ÜNİVERSİTESİ

LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ
AKILLI SİSTEMLER MÜHENDİSLİĞİ A.B.D.

ARAÇ CAN VERİLERİNİN VE EKG'NİN İŞLENMESİYLE SÜRÜCÜ
YORGUNLUK TESPİT SİSTEMİ GELİŞTİRME

YÜKSEK LİSANS TEZİ

KAAN KAHRAMAN

Tez Danışmanı: Dr. Öğr. Üyesi
Esra EKİNCİ

ŞUBAT 2025



**ARAÇ CAN VERİLERİNİN VE EKG'NİN
İŞLENMESİYLE SÜRÜCÜ
YORGUNLUK TESPİT SİSTEMİ
GELİŞTİRME**

Yüksek Lisans

Kaan KAHRAMAN İzmir 2025

**ARAÇ CAN VERİLERİNİN VE EKG'NİN İŞLENMESİYLE SÜRÜCÜ
YORGUNLUK TESPİT SİSTEMİ GELİŞTİRME**

Kaan KAHRAMAN

YÜKSEK LİSANS

**Akıllı Sistemler Mühendisliği Anabilim Dalı
Danışman: Dr. Öğr. Üyesi Esra EKİNCİ**

**İzmir
İzmir Bakırçay Üniversitesi
Lisansüstü Eğitim Enstitüsü
Şubat 2025**

JÜRİ VE ENSTİTÜ ONAYI

İzmir Bakırçay Üniversitesi Lisansüstü Eğitim Enstitüsü Akıllı Sistemler Mühendisliği Anabilim dalında öğrenim görmekte olan Kaan Kahraman'ın "Araç CAN Verilerinin ve EKG'nin İşlenmesiyle Sürücü Yorgunluk Tespit Sistemi Geliştirme" başlıklı tezi 27/02/2025 tarihinde aşağıdaki jüri tarafından değerlendirilerek "İzmir Bakırçay Üniversitesi Lisansüstü Eğitim-Öğretim ve Sınav Yönetmeliği'nin ilgili maddeleri uyarınca, Anabilim dalında Yüksek Lisans tezi olarak kabul edilmiştir.

Jüri Üyeleri	Unvanı Adı Soyadı	İmza
Üye (Tez Danışmanı)	Dr. Öğr. Üyesi Esra EKİNCİ	
Üye	Dr. Öğr. Üyesi Erdinç ÖNER	
Üye	Doç. Dr. Murat OTURAKÇI	
Üye		
Üye		

Prof. Dr. Özge TÜZÜN ÖZMEN
Lisansüstü Eğitim Enstitüsü Müdürü

ÖZET

Araç Can Verilerinin ve EKG'nin İşlenmesiyle Sürücü
Yorgunluk Tespit Sistemi Geliştirme

Kaan KAHRAMAN

Akıllı Sistemler Mühendisliği Anabilim Dalı

İzmir Bakırçay Üniversitesi, Lisansüstü Eğitim Enstitüsü, Şubat 2025

Danışman: Dr. Öğr. Üyesi Esra EKİNCİ

Günümüzde, yorgunluk kaynaklı kazaların önlenmesi için, sürücü yorgunluğunun tespiti konusunda doğru ve güvenilir sistemlerin geliştirilmesi büyük önem taşımaktadır. Bu çalışma, sürücü yorgunluğunu tespit etmek için hem fizyolojik verileri (HRV) hem de aracın CAN-Bus haberleşmesini kullanan hibrit bir sistem önererek literatüre yenilikçi bir katkı sağlamaktadır. Literatürde yer alan hibrit yaklaşımlar ise araç CAN-Bus verilerini simülasyon ortamında elde ederek geliştirilmiştir. Ancak bu çalışma iki farklı veri kaynağını gerçek dünyada birleştirerek daha yüksek doğrulukta tespit sağlamayı ve oluşturulan makine öğrenmesi modeli sonucunda sadece araç sinyalleri ile yorgunluğu tespit etmeyi amaçlamaktadır. Yöntem olarak, gerçek sürücülerden çeşitli çevresel ve trafik koşulları altında EKG ve CAN-Bus verileri eş zamanlı olarak toplanmış ve bu veriler makine öğrenimi algoritmalarıyla analiz edilmiştir. Üç sınıflı bir yorgunluk tespit modeli geliştirilmiş olup, bu model sürücülerini düşük, orta ve yüksek yorgunluk seviyelerine ayırmıştır. Elde edilen sonuçlar, modelin %85 doğruluk oranıyla yorgunluk tespitinde başarılı olduğunu göstermiştir.

Anahtar Sözcükler: Makine Öğrenmesi; HRV; EKG; CAN-BUS; Sürücü Yorgunluğu

ABSTRACT

Development of a Driver Drowsiness Detection System by Processing Vehicle CAN Data and ECG Signals

Kaan KAHRAMAN of student Department of Intelligent Systems Engineering

İzmir Bakırçay University, Graduate Education Institute, February 2025

Supervisor: Asst. Prof. Dr. Esra EKİNCİ

The development of accurate and reliable systems for detecting driver drowsiness is critically important to prevent fatigue-related accidents. This study introduces an innovative hybrid system that combines physiological data (HRV) and vehicle CAN-Bus communication to detect driver fatigue, offering a novel contribution to the literature. Unlike existing hybrid approaches, which typically use CAN-Bus data generated in simulation environments, this study integrates real-world data from two distinct sources to achieve higher detection accuracy and aims to identify fatigue solely through vehicle signals using a machine learning model. ECG and CAN-Bus data were simultaneously collected from real drivers under various environmental and traffic conditions, and these data sets were analyzed using machine learning algorithms. A three-class fatigue detection model was developed to classify drivers into low, medium, and high drowsiness levels. The results showed that the model achieved an 85% accuracy rate in detecting driver drowsiness.

Keywords: Machine Learning; HRV; ECG; CAN-BUS; Driver Drowsiness

TEŐEKKÜR

Bu alıőmanın hazırlanması sürecinde bilgi ve deneyimleriyle bana rehberlik eden ve alıőmanın her aőamasında desteklerini esirgemeyen, deęerli tez danıőmanım Dr. Öğr. Üyesi Esra EKİNCİ ve Do. Dr. Murat OTURAKÇI'YA sonsuz teőekkürlerimi sunarım.

Tez alıőmam sırasında sağladıęı kaynaklar, teknik destek ve motive edici sözleriyle yanımda olan Furkan KORKMAZ'A, TOFAŐ Otomotiv Yazılımları Őube Müdürü Utku KARAKAYA'YA ve ayrıca tüm ekip arkadaşlarıma teőekkürü bir bor bilirim.

Ayrıca alıőmalarım boyunca sabır ve anlayıőlarıyla beni destekleyen sevgili aileme, dostlarıma ve tüm sevdiklerime de teőekkürlerimi sunarım.

ETİK İLKE VE KURALLARA UYGUNLUK BEYANNAMESİ

Bu tezin bana ait, özgün bir çalışma olduğunu; çalışmamın hazırlık, veri toplama, analiz ve bilgilerin sunumu olmak üzere tüm aşamalarında bilimsel etik ilke ve kurallara uygun davrandığımı; bu çalışma kapsamında elde edilen tüm veri ve bilgiler için kaynak gösterdiğimi ve bu kaynaklara kaynakçada yer verdiğimi; bu çalışmanın İzmir Bakırçay Üniversitesi tarafından kullanılan Turnitin bilimsel intihal tespit programıyla tarandığını ve hiçbir şekilde “intihal içermediğini” beyan ederim. Herhangi bir zamanda, çalışmamla ilgili yaptığım bu beyana aykırı bir durumun saptanması durumunda, ortaya çıkacak tüm ahlaki ve hukuki sonuçları kabul ettiğimi bildiririm.

Kaan KAHRAMAN

İÇİNDEKİLER

Sayfa

ARAÇ CAN VERİLERİNİN VE EKG'NİN İŞLENMESİYLE SÜRÜCÜ YORGUNLUK TESPİT SİSTEMİ GELİŞTİRME	i
JÜRİ VE ENSTİTÜ ONAYI.....	ii
ÖZET	iii
ABSTRACT.....	iv
TEŞEKKÜR	v
ETİK İLKE VE KURALLARA UYGUNLUK BEYANNAMESİ.....	vi
İÇİNDEKİLER.....	vii
TABLolar DİZİNİ.....	ix
ŞEKİLLER DİZİNİ.....	x
KISALTMALAR DİZİNİ	xi
1. GİRİŞ.....	1
2. ÇEŞİTLİ YORGUNLUK ALGILAMA TEKNİKLERİ.....	4
2.1. Davranışsal Parametreler ile Yorgunluk Tespiti.....	5
2.1.1. Göz kırpma frekansı.....	5
2.1.2. Göz kırpma oranı.....	6
2.1.3. Esneme hareketleri ile tespiti.....	7
2.2. Araç Parametreleri ile Yorgunluk Tespiti.....	8
2.2.1. Araç direksiyon hareketleri	8
2.2.2. Şerit değiştirme sıklığı.....	8
2.3. Fizyolojik Veriler ile Yorgunluk Tespiti	9
2.3.1. HRV-EKG.....	9
2.3.2. EEG.....	10
3. ÇALIŞMANIN AMACI	12
3.1. Çeşitli Makine Öğrenmesi Teknikleri	17
3.1.1. XGBoost algoritması.....	17

3.1.2. K-NN (K-NearestNeighbor) Algoritması	19
<i>3.1.2.1. Öklid mesafesi</i>	20
<i>3.1.2.2. Manhattan mesafesi</i>	20
<i>3.1.2.3. Chebyshev mesafesi</i>	21
3.1.3. Karar Ağacı	21
<i>3.1.3.1. Karar ağacı algoritması türleri</i>	22
3.1.4. Rastgele Orman	23
3.1.5. SVM (Destek Vektör Makineleri)	26
<i>3.1.5.1. Lineer kernel</i>	30
<i>3.1.5.2. Polinomsal kernel</i>	30
<i>3.1.5.3. RBF kernel</i>	30
4. METOT	31
5. ARAÇTAN VE ADAYLARDAN VERİ TOPLAMA	33
6. BULGULAR	38
6.1. XGBoost Algoritması Sonuçları	38
6.2. KNN Algoritması Sonuçları	40
6.3. Karar Ağacı Algoritması Sonuçları	43
6.4. Destek Vektör Makineleri (SVM)Algoritması Sonuçları	45
6.5. Rastgele Orman Algoritması Sonuçları	47
7. TARTIŞMA	51
8. SONUÇ	54
KAYNAKÇA	55
ÖZGEÇMİŞ	63

TABLolar DİZİNİ

Sayfa

Tablo 1.1. 2013-2022 Yılları Arasında Türkiye'deki Trafik Kazaları ve Ölüm/Yaralanma Sayıları	1
Tablo 1.2. Ölümlü ve Yaralanmalı Trafik Kazalarındaki Kusur Oranları	2
Tablo 2.1. Şirketlerin Yorgunluk Algılama Sistemleri	5
Tablo 3.1. Sürücü Yorgunluğu Tespiti ile İlgili Literatür Taraması(EKG).....	14
Tablo 3.2. XGBoost Parametre Özellikleri	18
Tablo 4.1. Adayların Özellikleri	31
Tablo 4.2. Araç Sinyallerinin Özellikleri	32
Tablo 5.1. Hesaplanan RR süreleri.....	35
Tablo 5.2. Belirlenen Sınıflar	36
Tablo 6.1. XGBoost Algoritması Parametreleri	38
Tablo 6.2. XGBoost Algoritması Sınıflandırma Raporu	39
Tablo 6.3. KNN Algoritması Parametreler	41
Tablo 6.4. KNN Algoritması Sınıflandırma Raporu	42
Tablo 6.5. Karar Ağacı Algoritması Parametreler	43
Tablo 6.6. Karar Ağacı Algoritması Sınıflandırma Raporu.....	44
Tablo 6.7. SVM Algoritması Parametreler	45
Tablo 6.8. SVM Algoritması Sınıflandırma Raporu	46
Tablo 6.9. Rastgele Orman Algoritması Parametreler.....	47
Tablo 6.10. Rastgele Orman Algoritması Sınıflandırma Raporu	49
Tablo 7.1. Tüm Algoritmaların Karşılaştırılması	51

ŞEKİLLER DİZİNİ

Sayfa

Şekil 1.1. Yıllara Göre Sürücü Yorgunluğu Konusundaki Bilimsel Yayınların Dağılımı	3
Şekil 2.1. Göz Kırpma Oranında Noktalar	6
Şekil 2.2. Ağız Açıklığı Oranında Noktalar.....	7
Şekil 2.3. Aracın Şeritten Ayrılması	8
Şekil 2.4. EEG elektrotları yerleşimi.....	11
Şekil 3.1. Rastgele Orman Algoritması Adımlar	24
Şekil 3.2. Rastgele Orman Üyelik Sınıfı	25
Şekil 3.3. Hiper Düzlem	26
Şekil 3.4. Optimum Hiper Düzlem Belirleme	28
Şekil 3.5. Doğrusal Olarak Ayrışamayan Veriler.....	28
Şekil 3.6. Gevşeme Değişkeni ile Optimum Düzlem Belirleme	29
Şekil 5.1. Blue PiraT Mini ve MaM Sense EKG modülü	33
Şekil 5.2. Sistem Tasarımı	34
Şekil 5.3. Kubios SNS indeks Yorumlaması	36
Şekil 5.4. Kubios Yazılımı Arayüz	37
Şekil 6.1. XGBoost Algoritması Öğrenme Eğrileri.....	40
Şekil 6.2. KNN Algoritması Öğrenme Eğrileri	43
Şekil 6.3. Karar Ağacı Algoritması Öğrenme Eğrileri.....	45
Şekil 6.4. SVM Algoritması Öğrenme Eğrileri	47
Şekil 6.5. Rastgele Orman Algoritması Öğrenme Eğrileri	50

KISALTMALAR DİZİNİ

HRV	: Kalp atım hızı deęişkenlięi
EKG	: Elektrokardiyografi
EEG	: Elektroensefalografi
EOG	: Elektrookülografi
MAR	: Aęız Açıklık Oranı
EAR	: Göz Açıklık Oranı
ADAS	: Gelişmiş Sürücü Destek Sistemleri
ANS	: Otonom Sinir Sistemi
GBM	: Gradyan Artırma Makinesi
SNS	: Sempatik Sinir Sistemi
DT	: Karar Aęacı
SVM	: Destek Vektör Makinesi
PNS	: Parasempatik Sinir Sistemi
SNS	: Sempatik Sinir Sistemi

1. GİRİŞ

Türkiye yol ağı, sık sık ciddi trafik kazalarıyla karşı karşıya kalmakta ve bu kazalar her yıl binlerce ölüm, yaralanma ve maddi hasara yol açmaktadır. Tablo 1.1'e göre toplamda 1.232.957 trafik kazası meydana gelmiş olup, bunların 235.176'sı ölüm veya yaralanmayla sonuçlanmıştır. Bu kazalar sonucunda 288.696 kişi yaralanmış ve 5.229 kişi hayatını kaybetmiştir (ölüm sayısı, olaydan sonraki 30 gün içinde yaralanmaları nedeniyle hayatını kaybedenleri de kapsamaktadır). TÜİK 2022 – Karayolu Trafik Kaza İstatistiklerine dayanmaktadır. [1]

Tablo 1.1. 2013-2022 Yılları Arasında Türkiye'deki Trafik Kazaları ve Ölüm/Yaralanma Sayıları (TÜİK 2022 – Karayolu Trafik Kaza İstatistikleri)

Yıl	Trafik kaza sayısı	Ölümlü/yaralanmalı kaza sayısı	Ölü sayısı	Yaralı sayısı
2013	1,207,354	183,030	3,685	274,829
2014	1,199,010	193,215	3,524	285,059
2015	1,313,359	210,498	7,530	304,421
2016	1,182,491	213,149	7,300	303,812
2017	1,202,716	213,325	7,427	300,383
2018	1,229,364	217,898	6,675	307,071
2019	1,168,144	204,538	5,473	283,234
2020	983,808	177,867	4,866	226,266
2021	1,186,353	224,418	5,362	274,615
2022	1,232,957	235,176	5,229	288,696

Tablo 1.2. Ölümlü ve Yaralanmalı Trafik Kazalarındaki Kusur Oranları
(TUİK 2022 – Karayolu Trafik Kaza İstatistikleri)

Yıl	Sürücü Kusuru	Sürücü Kusuru %	Yolcu Kusuru	Yolcu Kusuru %	Yaya Kusuru	Yaya Kusuru %	Yol Kusuru	Yol Kusuru %	Araç Kusuru	Araç Kusuru %
2013	162,327	88.69	774	0.42	16,458	8.99	1,913	1.05	1,558	0.85
2014	171,236	88.62	901	0.47	18,115	9.38	1,841	0.95	1,122	0.58
2015	187,980	89.3	915	0.43	18,522	8.8	1,916	0.91	1,165	0.55
2016	190,954	89.59	869	0.41	18,612	8.73	1,717	0.81	997	0.47
2017	191,717	89.87	782	0.37	18,095	8.48	1,619	0.76	1,112	0.52
2018	194,928	89.42	1,916	0.88	18,396	8.44	1,300	0.6	1,135	0.52
2019	180,042	88.02	2,572	1.26	16,726	8.17	1,045	0.51	5,153	2.03
2020	157,128	88.02	2,577	1.44	12,520	7.04	897	0.5	4,745	2.65
2021	195,382	87.06	3,941	1.76	15,388	8.28	936	0.42	5,761	2.57
2022	204,233	86.84	2,753	1.17	22,234	9.45	902	0.38	5,054	2.15

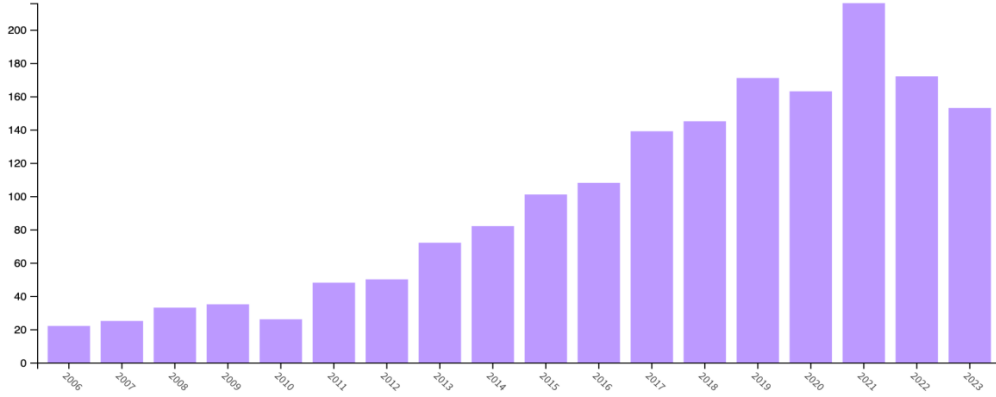
Bu endişe verici istatistikler, trafik güvenliği ile ilgili önemli sorunları bir kez daha gözler önüne sermektedir. (Tablo1.2.) Dünya geneli incelendiğinde ise trafik kazaları nedeniyle her yıl yaklaşık 1.200.000 ölüm ve 50 milyon yaralanma meydana gelmektedir [2]. Özellikle, uykulu halde araç kullanmak, dünya genelinde araç kazalarının başlıca nedenlerinden biridir. Polis raporlarına göre, ABD Ulusal Karayolu Trafik Güvenliği İdaresi (NHTSA), her yıl yaklaşık 100.000 kazanın doğrudan sürücü yorgunluğundan kaynaklandığını tahmin etmektedir. Bu kazalar, yaklaşık 12,5 milyar dolarlık maddi kayıpla sonuçlanmıştır [3].

Ayrıca, ABD'de yorgunluk nedeniyle araç kullanmak, ölümcül trafik kazalarının %16,5'inden ve yaralanmayla sonuçlanan çarpışmaların %12,5'inden sorumludur (Amerikan Otomobil Birliği Trafik Güvenliği Vakfı, 2010). Buna ek olarak, ABD'deki sürücülerin %14,5'i, araç kullanırken yarı uyku halinde olduklarını bildirmiş ve bunların yaklaşık %2'si, yorgunluk veya uykulu halde araç kullanma kaynaklı kazalara karışmıştır Tüm bu istatistiksel veriler doğrultusunda, sürücü yorgunluğundan kaynaklanan kazaların incelenmesi ve bu sorunların çözümüne yönelik geliştirilen çözümlerin tartışılması büyük önem taşımaktadır [4].

Uykulu olma durumu, basitçe uyuma eğilimi olarak tanımlanabilir. Uyku genellikle üç aşamaya ayrılır; uyanıklık, hızlı göz hareketi olmayan (NREM) uyku ve hızlı göz hareketi (REM) uykusu. NREM aşaması, uyanıklıktan uykuya geçiş, hafif uyku ve derin uyku olmak üzere üç evreye daha ayrılabilir.

- Aşama 1: Uyanıklıktan uykuya geçiş
- Aşama 2: Hafif uyku
- Aşama 3: Derin uyku

NREM uykusu, genellikle "sürücü yorgunluğu" olarak adlandırılan ilk evredir. Amerika Birleşik Devletleri Ulusal Karayolu Trafik Güvenliği İdaresi'nin (NHTSA) raporlarına göre sürücülerdeki yorgunluk, tepki süresinin yavaşlamasına, dikkat azalmasına ve beyin işlem kapasitesinin düşmesine yol açabilir [6]. Bu nedenle, sürücü yorgunluğundan kaynaklanan kazaları önlemek için sürücülerin yolculukları sırasında izlenmesi önemli hale gelmiştir. Şekil 1.1 incelendiğinde, sürücü yorgunluğu konusu özellikle 2019'dan sonra Web of Science yayınlarında önemli ölçüde ilgi görmüştür. Bu yayınların %60'ı, özellikle 2010'dan sonra sensör teknolojileri, veri madenciliği, derin öğrenme ve yüz tanıma alanlarındaki gelişim trendi ile yakından ilişkili olan bilgisayar bilimiyle ilgili olmuştur [7].



Şekil 1.1. Yıllara Göre Sürücü Yorgunluğu Konusundaki Bilimsel Yayınların Dağılımı (WOS - Web of Science Sürücü Yorgunluğu)

2. ÇEŞİTLİ YORGUNLUK ALGILAMA TEKNİKLERİ

Yorgunluk algılama metotları genellikle üç sınıfa ayrılmaktadır. Bunlar; davranışsal parametrelere dayalı teknikler, araç parametrelerine dayalı teknikler ve son olarak fizyolojik parametrelere dayalı tekniklerdir [8].

Davranışsal parametreler, yorgunluğu tespit etmek için kullanılan müdahalesiz yöntemlerdir. Bu teknikler, sürücünün göz oranı (EAR), göz kırpmaya sıklığı, yüz ifadeleri gibi davranışlarını analiz ederek yorgunluğunu değerlendirir. Esneme tabanlı tespit ise, ağız geometrisindeki değişiklikleri belirleyen bir görüntü işleme yaklaşımı olup, uyuklu bir sürücünün tespit edilmesine yardımcı olur [9].

İkinci kategoride, sürücü yorgunluğu, genellikle CAN-BUS'tan toplanan araç kontrol sistemlerinin analizi yoluyla değerlendirilir. Bu yöntem, direksiyon simidinde, gaz pedalına ve fren pedallarına yerleştirilen sensörler kullanılarak şerit değiştirme sıklığı, direksiyon hareketleri ve araç hızındaki değişiklikler gibi sürüş davranışlarını gözlemleyerek yorgunluğun tespit edilmesini içerir. Direksiyon simidi hareketi ve şerit pozisyonunun standart sapması, en yaygın kullanılan göstergeler arasındadır [10]. Bu çalışmalar, gerçek sürücülerden toplanan verilerle veya sürüş simülatörleri kullanılarak yapılabilir. Ancak simülatör kullanan çalışmalar, gerçek hayat koşullarını tam olarak yansıtamadıkları için yanıltıcı sonuçlar verebilir. Örneğin, kötü hava koşullarında veya bakımsız yollarda sürüş yapıldığında, sürücüler şeritten daha sık sapabilirler. Benzer şekilde engebeli bir arazide kullanılan bir araç, simülatörlere kıyasla daha fazla direksiyon hareketi gösterebilir. Sonuç olarak, simülatör araştırmalarının bulgularının, gerçek hayattaki sürüş durumlarını tam olarak yansıtmayabileceği ve bu nedenle yanlış sonuçlara yol açabileceği söylenebilir.

Son kategori, sürücünün durumunu değerlendirmek için elektronik cihazların cilde bağlanarak ölçümlerin alınmasını içerir. Bu yöntemler arasında Elektroensefalografi (EEG), Elektrokardiyografi (EKG) ve Elektrokülografi (EOG) gibi teknikler bulunmaktadır [11].

Sürücülerin yorgunluk dereceleri fizyolojik ölçümlerle doğru bir şekilde değerlendirilebilir. Ancak bu ölçümler için gereken araçların yüksek maliyeti, günlük kullanımda zorluklar ve dış faktörlerin neden olabileceği hatalı veriler göz önünde bulundurulmalıdır.

Literatürde davranışsal, fizyolojik ve araç tabanlı ölçümleri birleştiren hibrit yaklaşımlar da görülmektedir. Özellikle, çeşitli araştırmalar çevresel faktörler, EEG ve kamera ile sürücü takibi ile toplanan araç tabanlı ölçümlerle yorgunluk tespitini birleştirmiştir [12]. Bununla birlikte, sensör ve büyük veri analiz tekniklerindeki son gelişmeler nedeniyle bu konuda çok daha fazla araştırma yapılmaktadır. Buna rağmen, hibrit çalışmalarla ilgili literatürde hala boşluklar bulunmaktadır. Öte yandan, her bir yöntemde meydana gelebilecek hatalar, birden fazla yaklaşımı birleştiren hibrit yöntemlerle azaltılabilir.

Günümüzde otomotiv şirketleri, yorgunluk tespiti için çoğunlukla araç tabanlı ve davranışsal parametrelere dayalı yöntemler kullanmaktadır. Aşağıda yer alan tablo 2.1’de bazı otomobil üreticilerinin yorgunluk tespiti için kullandığı yöntemler listelenmiştir [13].

Tablo 2.1. Şirketlerin Yorgunluk Algılama Sistemleri (Doudou, M., Bouabdallah, A. ve Berge-Cherfaoui V., 2020)

Şirket	Kullanılan Özellik
Volvo	Şerit Sapması
Ford	Şerit Sapması
Subaru	Şerit Sapması
Volkswagen	SWM (Direksiyon Hareketi) + Şerit Sapması
Mercedes-Benz	SWM (Direksiyon Hareketi)
Lexus	Araç Salınımı

2.1. Davranışsal Parametreler ile Yorgunluk Tespiti

2.1.1. Göz kırpma frekansı

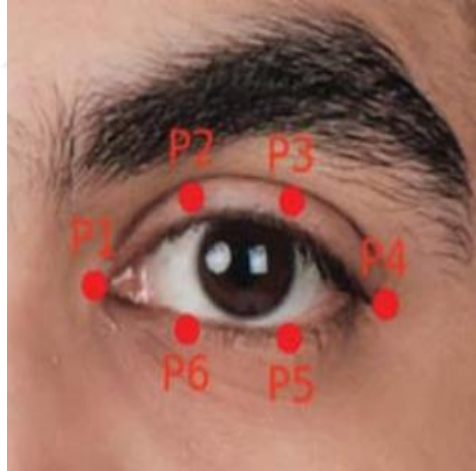
Göz kırpma frekansı, birim zaman başına gerçekleşen göz kırpma sayısını ifade eder. Sürücü yorgun olduğunda, göz kırpma frekansı normal değerine çıkar. Bu yaklaşım, doğruluğunun yüksek olması ve güvenilirliği sayesinde yorgunluk tespitinde günümüzde yaygın olarak kullanılmaktadır [14]. PERCLOS (belirli bir zaman diliminde gözün kapalı kaldığı sürenin yüzdesi), göz kırpma sıklığını ölçmek için en popüler yöntemdir, çünkü yüksek PERCLOS skorları uykululuk ile güçlü bir şekilde ilişkilidir [15]. Ancak sürücü gözlerini uzun süre kapatmasa bile yorgunluktan kaynaklı dikkatsiz davranışlar sergileyebilmektedir [16].

2.1.2. Göz kırpma oranı

Göz Kırpma Oranı (EAR) formülü, göz kırpmalarını skaler bir değer kullanarak tespit eder. Örneğin, bir sürücü daha sık göz kırptığında, bu kişinin uykulu bir durumda olduğunu gösterir. Bu nedenle, göz kırpma sıklığını hesaplamak için göz şeklinin doğru bir şekilde tespit edilmesi gerekmektedir. Yüz görüntüsünden tespit edilen işaret noktaları kullanılarak EAR göz açıklık durumunun bir tahmini olarak kullanılır [17]. Göz oranı, denklem 1’de gösterildiği gibi tanımlanabilir.

$$EAR = \frac{|p2 - p6| + |p3 - p5|}{2|p1 - p4|} \quad (1)$$

Bu denklemde, P1, P2, P3, P4, P5 ve P6 iki boyutlu olarak yüzün belirli noktalarındaki işaretlerin konumlarını temsil eder. Gözün belirli noktaları şekil 2.1’de gösterilmiştir. Denklemin pay kısmı, dikey mesafeyi hesaplar [18].



Şekil 2.1. Göz Kırpma Oranında Noktalar (S. Cheamanunkul ve S. Chawla, 2020)

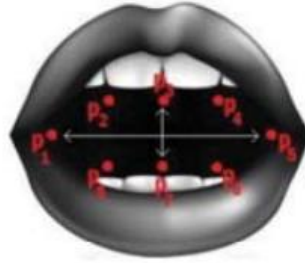
Göz Kırpma Oranı (EAR), EEG ve EOG gibi yoğun sinyal ve bilgisayar işlemeye ihtiyaç duymadan göz kırpmalarını değerlendirmek için kullanılan basit ama yüksek performanslı bir yaklaşımdır [19].

2.1.3. Esneme hareketleri ile tespiti

Bu sistem, cilt tespiti adımıyla başlar. Yüzleri tespit etmek için en etkili yöntem, cilt rengini ve dokusunu belirlemektir. Cilt tespiti için RGB ve YCbCr gibi çeşitli renk uzayları kullanılır ve bu uzaylarda dönüşümler yapılır. Cilt tespit süreci, insan yüzüne sahip olma olasılığı olan bölgeleri bulmak için bir ön işleme adımı olarak kabul edilir [20]. Daha sonra, esneme tespiti, ağız koordinatlarından MAR (Ağız Oranı) ölçülerek denklem 2’de olduğu gibi tahmin edilir.

$$MAR = \frac{||p_2 - p_8|| + ||p_3 - p_7|| + ||p_4 - p_6||}{3||p_1 - p_5||} \quad (2)$$

Şekil 2.2’de gösterildiği gibi, MAR'ın (Ağız Oranı) pay kısmı, dikey ağız koordinatları arasındaki mesafeyi hesaplar, payda ise yatay ağız koordinatları arasındaki mesafeyi değerlendirir (p1, p2 ...p8, sekiz ağız koordinatını temsil eder) [21].



Şekil 2.2. Ağız Açıklığı Oranında Noktalar
(A. Ghourabi, H. Ghazouani, ve W. Barhoumi, 2020)

Şekil 2.2. ve denklem 2 incelendiğinde, ağız açık olduğunda MAR değerinin arttığı gözlemlenmiştir. Ancak bu yöntem tek başına yorgunluğu tespit etmek için kullanılabilecek bir yöntem değildir [22].

2.2. Araç Parametreleri ile Yorgunluk Tespiti

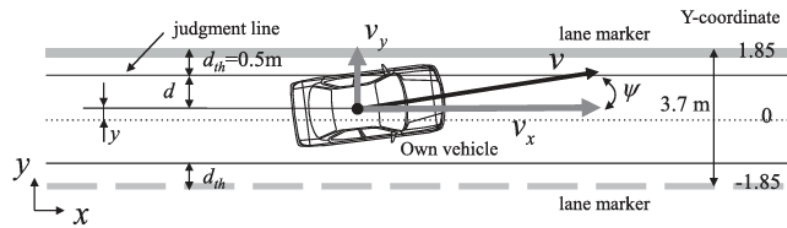
2.2.1. Araç direksiyon hareketleri

Araç direksiyon manevralarını tanımlamak, aracın açısal hızına yöneliktir ve ardından tehlikeli direksiyon manevralarını tanımlamak için farklı direksiyon hareketlerine odaklanılır [23]. Ayrıca literatürde, yorgun bir sürücüde zamanla artan yorgunluk durumu ile direksiyon üzerindeki tutma kuvvetinde bir azalma olduğu kaydedilmektedir. Bu nedenle, bazı araştırmacılar direksiyon üzerindeki tutma kuvvetine dayalı bir yorgunluk tespit sistemi geliştirmiştir. Ancak bu yaklaşım sürücünün sürüş alışkanlıkları, hava durumu, yol ve trafik koşulları gibi dış faktörlerden kolayca etkilenebilir [24].

2.2.2. Şerit değiştirme sıklığı

Dalgınlık ve yorgunluk kaynaklı şeritten istemsiz çıkış, trafikte insanların hayatını riske atan ana sebeplerin başında gelmektedir. Sürücünün dalgın olması nedeniyle bir ölümcül çarpışma aniden gerçekleşebilir. Bu nedenle, bu tür kazaları en aza indirmek için, aracı şeridinden sapıp saptadığını, sürücüye uyarabilecek otomatik bir uyarı sistemine ihtiyaç vardır [25].

Günümüzdeki, yeni üretilen çoğu araçta Gelişmiş Sürücü Destek Sistemleri (ADAS) sıkça kullanılmaktadır. LWD (Şerit İhlal Uyarı) sistemi, (ADAS) sisteminin önemli bir modüllerinden birisidir. Bu sistem, aracın mevcut şeritten istem dışı olarak çıkmasını algılar ve sürücüyü sesli, görsel veya titreşimli uyarılarla bilgilendirir. LWD sistemleri, özellikle uzun yolculuklar veya sürücünün yorgun olabileceği durumlarda, trafik kazalarını önlemeye yardımcı olur [26].



Şekil 2.3. Aracın Şeritten Ayrılması (Y. Saito, M. Itoh, ve T. Inagaki, 2016)

Şeritten ayrılma tahmini için denklem 3 formülü kullanılabilir.

$$t_a(t) = d(t) / v_y(t) = d(t) / \{ v(t) \cdot \sin \psi(t) \} \quad (3)$$

Denklem 3 ve şekil 2.3 incelendiğinde, aracın yanal hareketiyle ilgili olarak belirli bir mesafeye ulaşması için geçen süreyi ifade etmektedir. Aracın yanal hızı $v_y(t) = v(t) \cdot \sin \psi(t)$ olduğu için, belirli bir yanal mesafe $d(t)$ 'yi kat etmesi için gereken süre $d(t) / v_y(t)$ olarak hesaplanır. Burada $v_y(t)$ aracın toplam hız büyüklüğünü $\psi(t)$ ise yön açısını temsil etmektedir. Bu sayede aracın mevcut hızına ve yön açısına bağlı olarak, şerit ihlali veya sürücünün şerit değiştirme süresi hesaplanabilmektedir [27].

2.3. Fizyolojik Veriler ile Yorgunluk Tespiti

2.3.1. HRV-EKG

Kalp hızı değişkenliği (HRV), EKG'de ölçülen ardışık kalp atımları arasındaki sürelerin değişkenliği anlamına gelir. Bir diğer değişle HRV, RR aralıklarının dalgalanması olup, vücudun otonom sinir sisteminin (ANS) sempatik ve parasempatik aktivitelerinin bir göstergesi olarak kabul edilen önemli bir fizyolojik parametredir [28].

Otonom sinir sistemi (ANS) aktivitesi, stres, aşırı yorgunluk ve uyku hali gibi durumlarda değişiklikler gösterir. Kişi çevresel uyarıcılara tepki verebildiği durumlarda genellikle sempatik aktivitede bir artış ve/veya parasempatik aktivitede bir azalma ile karakterize edilirken, Dinlenme gevşeme veya rahatlama durumlarında ise parasempatik aktivitede bir artış ve/veya sempatik aktivitede bir azalma ile tanımlanmaktadır.

Kalp hızı değişkenliği (HRV) analizi, yorgunluk değerlendirmesinde hem zaman domaini hem de frekans domaini yöntemlerini kullanmaktadır. Zaman domaini analizi, kalp atımları arasındaki sürelerin (RR aralıkları) değişkenliğini ölçerek HRV'nin genel seviyesini değerlendirir. Örneğin, SDNN ve RMSSD gibi metrikler, yorgunluk sırasında HRV'nin azalıp azalmadığını anlamak için kullanılabilir. Ancak zaman domaini analizi, sempatik ve parasempatik sinir sistemi aktivitelerini ayrıntılı olarak ayırt edemez. Ancak frekans domaini analizi, yorgunlukla ilişkili otonom sinir sistemi dengesizliklerini daha ayrıntılı bir şekilde ortaya koymaktadır.

Frekans domaininde, düşük frekans (LF) bandı (0.04–0.15 Hz) gücü, sempatik ve parasempatik sinir sistemi aktiviteleri ile diğer mekanizmaların etkisi altında şekillenir. Normalize edilmiş birimler kullanıldığında, genellikle sempatik aktivitenin baskınlığını gösteren bir ölçüt olarak kabul edilir. Yüksek frekans (HF) bandı (0.15–0.4 Hz) gücü ise, solunum frekansının 0.15–0.4 Hz aralığında olduğu varsayıldığında, parasempatik aktivitenin bir göstergesi olarak değerlendirilir. Sempatik ve parasempatik sistemler arasındaki denge LF/HF oranı ile ölçülür ve bu oran, otonom sinir sisteminin dengesini anlamak için önemli bir hesaplama parametresidir. [29], [30] Yüksek bir LF/HF oranına sahip olmak sempatik baskınlığı yansıttığı, bunun da yorgunluk durumu ile ilişkili olabileceği bazı çalışmalar tarafından araştırılmıştır [31].

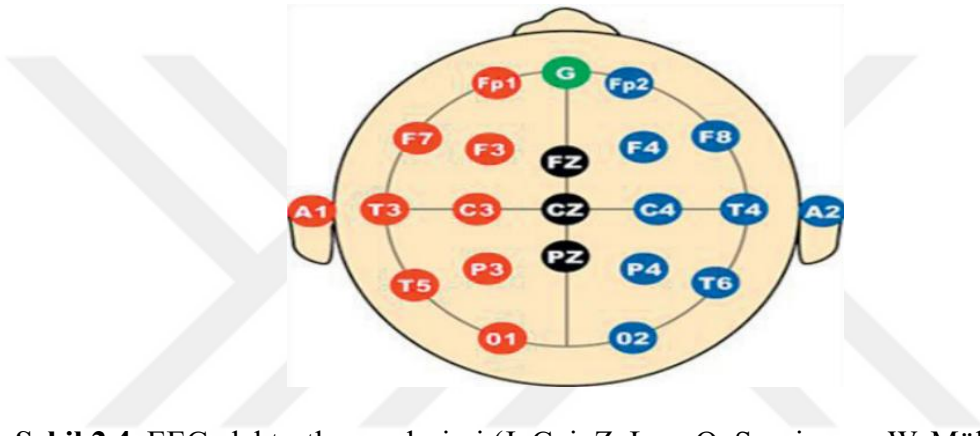
2.3.2. EEG

EEG, kafa derisi üzerinde voltaj farkını ölçer; bu fark, beynin yüzey tabakasındaki binlerce piramidal nöronun sinaptik aktiviteleri sonucunda oluşan iyon akımlarından kaynaklanır. EEG sonucunda elde edilen voltaj son derece küçüktür. Bu değerler genellikle mV cinsinden ölçülür ve gürültülere karşı son derece hassastır. EEG sinyalleri, doğası gereği durağan değildir; bu da farklı zaman aralıklarında farklı frekans değerlerine sahip oldukları anlamına gelir. Tipik bir EEG sinyalindeki beş ana frekans bandı şu şekildedir;

- 1) Delta – (0.5 – 4) Hz
- 2) Teta – (4 – 8) Hz
- 3) Alfa – (8 – 12) Hz
- 4) Beta – (12 – 30) Hz
- 5) Gamma – >30 Hz

Yorgunluk tespiti için özellikle Teta, Alfa ve Beta alt bantları önem taşır. Alfa aktivitesi, bir kişi dinlenirken veya gözlerini kapattığında baskın hale gelir. Uyanık olma durumundan uykuya veya uykulu hale geçiş sırasında, Alfa dalgaları azalırken Teta dalgaları yavaşça artar. Ayrıca, bir kişinin sakinleştirici türevi veya hipnotik ilaçlar kullandığında, Beta dalgalarının önemli ölçüde arttığı kanıtlanmıştır.

EEG sinyalleri, kafa derisine elektrotlar yerleştirilerek ölçülür. Mevcut EEG sistemlerinde 1 ila 256 arasında elektrot bulunur ve elektrotların yerleşimi, 10–20 sistemi veya Uluslararası 10–20 sistemi olarak adlandırılan resmi bir standarda göre yapılır. Şekil2.4’te, nörologlar ve ilgili sağlık profesyonelleri tarafından kullanılan bir elektrot konfigürasyonunu göstermektedir. İngilizce harflerle işaretlenen F, P, T, O ve C bölgeleri sırasıyla frontal lob, parietal lob, temporal lob, oksipital lob ve merkezi lob ile ilişkilidir. A1 ve A2, referans elektrotların yerleştirildiği bölgeleri temsil eder. Fpz, frontal polar bölgeyi, Pz ise nazofaringeal pozisyonu ifade eder. [32], [33]



Şekil 2.4. EEG elektrotları yerleşimi (J. Cui, Z. Lan, O. Sourina, ve W. Müller-Wittig, 2023)

3. ÇALIŞMANIN AMACI

Bu çalışmanın araştırma soruları şunlardır:

Araştırma Sorusu 1: Sürüş esnasında sürücülerden fizyolojik verileri doğru bir şekilde toplamak mümkün müdür?

Araştırma Sorusu 2: Hem fizyolojik verileri hem de sürücülere ait araç tabanlı ölçümleri aynı anda göz önünde bulunduran bir yorgunluk tespit sistemi önerilebilir mi?

Araştırma Sorusu 3: Araç tabanlı ölçümler üzerine yapılmış önceki çalışmalarda belirlenen parametrelerin yanı sıra, sürücü yorgunluğunu etkileyen başka parametreler var mı?

Araştırma Sorusu 1 ile ilgili olarak, fizyolojik veriler (EKG) sürücülerden yolculukları boyunca toplanıp, incelenecektir. EKG, kalbin biyoelektrik aktivitesini ölçer. EKG sinyali, kalp atışları sırasında kalp kaslarının depolarizasyonu nedeniyle vücudun yüzeyinde oluşan çok düşük seviyeli elektriksel voltajların sonucudur. EKG çıktısında gözlemlenen dalgaların ve parametrelerin belirli adları vardır; Q, R, S, T, QT, ST, RR ve QRS gibi [34]. Ancak yorgunluk tespiti için EKG verilerinden türetilen HRV (Kalp Atış Hızı Değişkenliği) dikkate alınmaktadır. HRV, kalp atışları arasındaki zaman aralıklarındaki değişimi ölçer ve sempatik ve parasempatik sinir sistemlerinin (PSNS) düzenlenmesini yansıtır ve yorgunluğu tespit etmek için önceki çalışmalarda kullanılmıştır [35]. HRV özellikleri, R dalgasının tepe noktalarındaki seriye ve RR dalgaları arasındaki aralıklara bakılarak incelenir. Bu araştırma sorusuna yanıt vermek için fizyolojik ölçümler, toplanan EKG verilerinden HRV değerleri çıkarılarak gerçekleştirilecektir.

Sürücü yorgunluğunun derecesini doğru bir şekilde belirlemek için hem sürücülerden fizyolojik hem de kullanılan test araçlarından çeşitli sinyalleri toplamak, çalışmanın ana hedefidir. Araştırma Sorusu 2'yi yanıtlamak için bu hedeften yola çıkılmıştır. Çalışma, iki farklı kaynaktan alınan verileri bir araya getirerek sürücü yorgunluğunu daha doğru bir şekilde modelleyen ve tahmin eden bir sistem oluşturmak amacıyla çeşitli makine öğrenmesi algoritmalarını kullanmayı amaçlamaktadır. EKG verilerinin mevcut ve güvenilir olması nedeniyle, fizyolojik ölçümler için kullanılacaktır. Aynı zamanda, araçla ilgili bilgiler aracın CAN-Bus (Controller Area Network-Bus) haberleşmesinden elde edilecektir. Ayrıca bu çalışma, veri toplama ve

analiz için hibrit bir yöntem kullanarak literatürdeki boşluğu kapatmayı hedeflemektedir.

Araştırma Sorusu 3 ile ilgili olarak ise, bu çalışma TOFAŞ Türk Otomobil Fabrikası A.Ş. ile iş birliği yaparak gerçek araçlardan CAN-Bus verileri toplanmıştır ve bu verilerin yeni yorgunluk tespiti parametreleri oluşturmak amacıyla kullanılacaktır. Ayrıca makine öğrenimi algoritmaları, bu verileri incelemek ve sürücü yorgunluğuyla anlamlı bir ilişki gösteren parametreleri bulmak amacıyla kullanılacaktır.

Çalışmanın odak noktası, gerçek sürücülerden gerçek zamanlı veri toplamak olup, bu durum literatürde sıklıkla sürüş simülatörlerinden elde edilen verilerle yapılan önceki çalışmalardan ayrılmaktadır. Farklı yaş gruplarından sürücüler dahil edilerek ve çeşitli çevresel koşullarda veri toplanarak, çalışmanın daha anlamlı ve doğru sonuçlar vermesi amaçlanmaktadır. Dolayısıyla, değerlendirme sonuçlarının gerçek dünyada uygulamaya yönelik daha pratik ve etkili çözümler sunması beklenmektedir. Çalışmanın bulguları, iş birliği yapılan TOFAŞ Ar-GE merkezi tarafından sürücü güvenliği teknolojilerini geliştirmek amacıyla doğrudan uygulanacaktır.

EKG sinyalleri kullanılarak sürücü yorgunluğu tespiti konusunda literatürdeki boşluğu belirlemek için Web of Science veri tabanındaki 2020 ve 2024 yılları arasında yayımlanan en güncel makaleler Tablo 3.1'de incelenmiştir. İncelenen çalışmaların sonuçları, bu çalışmadan elde edilen bulgularla birlikte tartışma bölümünde değerlendirilmektedir.

Tablo 3.1. Sürücü Yorgunluğu Tespiti ile İlgili Literatür Taraması (EKG)

Yazar	Yıl	Veri	Ölçüm Tipi			Metot	Performans	Amaç
			Araç	Davranışsal	Fizyolojik			
Chang, et al.	2021	21 kişiyle sürüş simülasyonu		Gözlerin durumları (göz kırpması, kapanma, açılma) Baş hareketleri (dikey, yatay ve dönme)	EKG	Analistik Hiyerarşi Süreci Makine Öğrenmesi	97.97%	EKG kullanarak Sürücü Yorgunluk Tespiti
Tjolleng& Jung	2024	67 kişiyle sürüş simülasyonu			EKG	Yapay Sinir Ağı	Eğitim Doğruluğu : $99 \pm 0.8\%$ ve Test Doğruluğu $95 \pm 3.7\%$	EKG sinyali kullanılarak kullanıcının bilişsel aşırı yüklenme veya uyuklu durumda olup olmadığının belirlenmesi
Sukumar, et al.	2024	13 kişiyle sürüş simülasyonu		Direksiyon simidi sensörleri Baş hareketleri için ivmeölçer	EKG EMG GSR	Synchrosqueezing Sürekli Dalgacık Dönüşümü	97.30%	Sürücü Yorgunluğu ve Stresinin Belirlenmesi
Arefnezhad, et al.	2022	92 kişiyle sürüş simülasyonu		Esnemeyi, uzun göz kırpmaları ve baş sallamayı analiz etmek için video kayıtları	EKG	Evrişimli sinir ağı	Manuel ve otomatik modlarda sırasıyla %77 ve %79	EKG kullanarak Sürücü Yorgunluk Tespiti
Esteves, et al.	2021	13 kişiyle sürüş simülasyonu	Hızlanma ve frenleme gibi dinamik sinyaller doğrudan simülatörden alınmış	Yüz videosu	EKG			Sürücü yorgunluk izleme alanında son teknolojiyi iletirmek
Ebrahimian, et al.	2022	30 kişiyle sürüş simülasyonu			EKG, Solunum	Evrişimli sinir ağları (CNN'ler) uzun kısa süreli bellek (LSTM)	Uykululuğun üç seviyeli ve beş seviyeli sınıflandırmaları sırasıyla %91 ve %67 doğrulukla gerçekleştirilebilir	Üç ve beş seviyeli uykululuk sınıflandırması yaparak endüstriyel kullanıma uygun bir yöntem önerme
Murugan, et al.	2020	10 kişiyle sürüş simülasyonu			EKG	Makine Öğrenimi algoritmaları: Destek Vektör Makinesi (SVM), K-en yakın komşu (KNN)	N. uyuklu, n. görsel dikkatsizlik, n. yorgunluk ve n. bilişsel dikkatsizlik sınıflandırmaları sırasıyla %100,	EKG'den elde edilen fizyolojik sinyaller kullanılarak sürücüde dikkat düşüklüğünün (uykululuk, yorgunluk, görsel dikkatsizlik ve

							%93.1, %96.6 ve %96.6 doğruluğa sahiptir. Beş sınıflı için doğruluk ise %58.3'tür	bilışsel dikkatsizlik tespiti
Babusiak, et al.	2021				EKG Oksimetre fotopletismografi			Direksiyon üzerinde belirli kardiyovasküler hastalıkların (özellikle atriyal fibrilasyon) ve uykululuk durumunun tespiti için bir platform geliştirilmesi
Perkins, et al.	2023	14 kişiye ait verileri içeren DROZY veritabanı		Sürücülerin yüzlerinin ve davranışlarının video kayıtları	EKG EEG EMG EOG	Makine öğrenmesi algoritmaları: k-En Yakın Komşular (kNN), Çok Katmanlı Algılayıcı (MLP) Davranışsal Ölçüm: ResNet101 (evrişimli sinir ağı mimarisi), uzun kısa süreli bellek (LSTM)	93.10%	Sürücü yorgunluğunu tespit etmek için hibrit bir model geliştirilmesi
Fujiwara, et al.	2023	20 kişiyle sürüş simülasyonu			EKG EEG EOG	Sinir ağı modeli (otomatik kodlayıcı)	89%	Sürücü Yorgunluğunu EKG ile tespit etme (RR aralıkları)
Kundinger, et al.	2020	30 kişiyle sürüş simülasyonu		Yüz videosu	Bileklik ile kalp hızı değişkenliği ölçümü EKG	Makine öğrenmesi algoritmaları: 9 farklı yöntem	Kullanıcıya bağımlı testlerde %97.37, kullanıcıdan bağımsız testlerde %78.94 doğruluk	Bilekte takılan bir giyilebilir cihaz kullanarak sürücü yorgunluğunun tespiti.
Wolkow, et al.	2020	59 ağır vasıta sürücüsü ile	Araç izleme sisteminde: sert frenleme ve sert hızlanma olayları toplanır		Bileklik ile kalp hızı değişkenliği ölçümü			Bilekte takılan bir giyilebilir cihaz kullanarak sürücü yorgunluğunu tespit etmek ve sert fren ve ivmelenmeleri önlemek için sürücülerini uyarmak
Muramatsu&Sasaki	2021				EKG: Direksiyon üzerindeki iletken elektrotlar ve koltuk üzerindeki kapasitif elektrotlar kullanılarak			Sürücü yorgunluğunu iletken/kapasitif hibrit EKG R pik ölçüm sistemi kullanarak tespit etme
Linschmann, et al.	2023	20 kişiyle sürüş simülasyonu			Kapasitif EKG ile taşınabilir yastık			Sürücünün hayati belirtilerini izlemek için taşınabilir bir yastık

					kullanılarak kalp ve solunum hızı ölçümü			geliştirme
Seok, et al	2020	11 kişiyle deney			EKG EEG	Derin Q-ağı (DQN) algoritması, geleneksel öznelik mühendisliği ve derin evrişimli sinir ağı (CNN) yöntemleri kullanılarak tasarlanmıştır	>90%	EKG ve EEG kullanarak düşük dikkat seviyesini tespit etme
Du, et al.	2022	50 kişiyle sürüş simülasyonu			Avuç içi ve göğüs kullanılarak EKG	Bulanık evrişimli sinir ağı (FCNN)	85.32% - 96.49%	Akıllı direksiyon üzerinden EKG ölçümü yaparak sürüş yorgunluğunu tespit etme
Sedik, et al.	2023	14 kişiye ait verileri içeren DROZY veri tabanı		Sürücülerin yüzlerinin ve davranışlarının video kayıtları	EKG EEG EMG EOG	Makine öğrenimi algoritmaları: SVM, RF, DT, KNN, QDA, MLP, LR, CNN, ve C-CNN	96%	EKG, EEG, EMG ve EOG verilerine dayalı olarak sürücü yorgunluğunu tespit etmek için bir model geliştirilmesi
Yang, et al.	2023	797 veri(tek konu)			BCG sinyali (EKG) EEG	evrişimli sinir ağları (CNN'ler)uzun kısa süreli bellek (LSTM)	Kalp hızı doğruluğu: %88.49, solunum hızı doğruluğu: %88.37, BCG kullanılarak	Sürücülerin kalp hızlarını EKG'ye benzer şekilde izlemek için temassız bir biyosinyal ölçüm yöntemi geliştirilmesi
Devarajan, et al.	2024		Araç parametrelerini algılamak için sensörler (direksiyon simidi açısı sensörü, kavrama kuvveti sensörü, radar sensörü, IR (sensörü)	Yüz videosu	EKG EEG	Blockchain özellikli Derin Evrişimli Sinir Ağı (BDCNN)		Sürücü, araç ve yol verilerini izlemeyi mümkün kılan hibrit bir sistem geliştirilmesi

3.1. Çeşitli Makine Öğrenmesi Teknikleri

3.1.1. XGBoost algoritması

XGBoost (eXtremeGradientBoosting), Gradient Boosting Machines (GBM) için en etkili ve yaygın olarak kullanılan boosting ağacı algoritmalarından biri olarak kabul edilir. Boosting, regresyon ve sınıflandırma problemlerinde kullanılabilen bir makine öğrenimi tekniğidir. Güçlü problem çözme yetenekleri sayesinde, birçok endüstride geniş çapta benimsenmiştir [36], [37].

XGBoost algoritması, geleneksel GBDT algoritmasından daha etkilidir. Geleneksel GBDT yalnızca birinci dereceden türev bilgilerini kullanırken, XGBoost kayıp fonksiyonu üzerinde ikinci dereceden Taylor açılımını kullanarak daha verimli bir çözüm sunar. Ayrıca, XGBoost bir düzenleme terimi ekleyerek modelin varyansını azaltır, modeli daha basit hale getirir ve aşırı öğrenmenin (overfitting) önlenmesine yardımcı olur [38].

Ana hedef fonksiyonu denklem 4'te verilmiştir:

$$O = \sum_{i=1}^n L(y_i, f(x_i)) + \sum_{k=1}^t R(f_k) + C \quad (4)$$

$R(f_k)$, k. iterasyondaki düzenleme terimini temsil eder ve C sabiti isteğe bağlı olarak dışarıda bırakılabilir. Denklem 5, düzenleme terimi $R(f_k)$ gösterir:

$$R(f_k) = \alpha H + \frac{1}{2} \eta = \sum_{j=1}^H \omega_j^2 \quad (5)$$

Burada, α yaprakların karmaşıklığını, H yaprak sayısını, η ceza değişkenini ve ω her yaprak düğümündeki çıktıyı temsil eder. Ayrıca, GBDT'den farklı olarak, XGBoost ana fonksiyonların yalnızca birinci dereceden türevini kullanmak yerine, ikinci dereceden Taylor açılımını kullanır. Kayıp fonksiyonu ortalama kare hata (MSE) olduğunda, amaç fonksiyonu şu şekilde ifade edilebilir:

$$O = \sum_{i=1}^n [p_i \omega_{q(x_i)} + \frac{1}{2} (q_i \omega_{q(x_i)}^2)] + \alpha H + \frac{1}{2} \eta \sum_{j=1}^H \omega_j^2 \quad (6)$$

Amaç fonksiyonu, kayıp fonksiyonu ve düzenleme terimi olmak üzere iki ana bileşenden oluşmaktadır. Kayıp fonksiyonunu denklem 6'da ilk kısımdır; $\sum_{i=1}^n [p_i \omega_{q(x_i)} + \frac{1}{2} (q_i \omega_{q(x_i)}^2)]$ ve kayıp fonksiyonu da iki bileşenden oluşmaktadır. İlk terim olan $p_i \omega_{q(x_i)}$, her bir örneğin hatasını (kaybını) temsil eder. Burada p_i , hatanın ağırlığıdır. $\omega_{q(x_i)}$ ifadesi ise modelin tahmin ettiği değeri temsil etmektedir. Denklem ikinci kısmı olan $\frac{1}{2} (q_i \omega_{q(x_i)}^2)$ ifadesi ise bu hataların karesel olarak düzenlenmiş halidir. Bu hataları kare olarak büyük hataların etkisi daha fazla dikkate alınır.

Amaç fonksiyonunun yani denklem 6'nın ikinci kısmı olan $\alpha H + \frac{1}{2} \eta \sum_{j=1}^H \omega_j^2$ ise modelin karmaşıklığını kontrol eder. Denklemdeki αH ifadesi, " α " katsayısı ile yaprak düğümlerin sayısını çarparak daha karmaşık ağaçlar için ceza uygular. Yani, modelin fazlasıyla yaprak oluşturması durumunda ceza sayısı artmış olur. $\frac{1}{2} \eta \sum_{j=1}^H \omega_j^2$ ifadesi, Yaprak düğümlerdeki ağırlıkların karelerinin toplamını gösterir. Bu terim, her bir yaprak düğümünün ağırlıklarını sınırlandırarak modelin karmaşık hale gelmesini önlemektedir.

Özetle, ana fonksiyonu optimize etmek, bir ikinci dereceden fonksiyonun minimumunu bulma meselesi haline gelir. Düzenlemenin eklenmesi sayesinde, XGBoost aşırı öğrenmeyi (overfitting) önlemede daha etkilidir [39].

XGBoost'un belirlenmesi gereken çok sayıda hiper parametresi olduğundan, optimal model parametrelerini bulmak, XGBoost modelinin en iyi etkiyi elde edip edemeyeceğinde belirleyici bir rol oynar.

Tablo 3.2. XGBoost Parametre Özellikleri (Y. Jiang, G. Tong, H. Yin, ve N. Xiong, 2019)

Parametre Adları	Aralıklar	Varsayılan Değerler
n_Estimators	-	-
learningRate	[0, 1]	0.3
maxDepth	[0, ∞)	6
minChildWeight	[0, ∞)	1

gammaValue	[0, ∞)	0
subSample	(0, 1]	1
colSampleByTree	(0, 1]	1

Tablo 3.2'ye göre, verilen parametreler incelendiğine $n_estimators$, son verilen denklemin kaç kez tekrarlanacağını, $learning_rate$ her adımda yaprak düğümleri güncellenirken ağırlık ETA yani öğrenme parametresi ile çarpılır. Daha düşük bir öğrenme oranı, her bir ağacın modeli güncellerken yaptığı değişikliklerin etkisini azaltır ve bu da modelin daha yavaş ama potansiyel olarak daha dikkatli öğrenmesini sağlar. Subsample, parametresi, her iterasyonda eğitilecek ağacın ne kadar veri üzerinde çalışacağını belirler. ColSamplebytree her bir ağacın oluşturulması sırasında kullanılan sütunların (özelliklerin) ne kadarının rastgele seçileceğini belirler. Başka bir deyişle, bu parametre, her ağacın eğitiminde kullanılan özelliklerin oranını kontrol eder. Subsample parametresine benzemektedir ancak colsamplebytree veri noktaları yerine sütunları örnekleme [40]. Max Depth parametresi ise, her bir ağacın derinliğini sınırlandırarak, ağaçların ne kadar karmaşık olabileceğini belirler. Bu, özellikle overfitting (aşırı uyum) riskini kontrol altında tutmak için önemlidir. Eğer max_depth değeri artırılırsa, model daha derin ağaçlar oluşturularak daha karmaşık hale gelir ve bu da aşırı uyum riskini artırabilir. Yüksek derinlik değeri modelin belirli bir örneğe özgü ilişkileri öğrenmesine izin verir ve bu da aşırı uyuma yol açabilir [41].

3.1.2. K-NN (K-Nearest Neighbor) Algoritması

K-en yakın komşu (KNN), bir sınıflandırma tekniği olarak kullanılır. Yorumlanmasının kolay olması ve düşük hesaplama süresi nedeniyle yaygın olarak tercih edilir. Bu algoritmada k parametresi kritik bir öneme sahiptir [42]. KNN'nin temel konsepti, verilen bir x örneği ile eğitim veri setindeki örnekler arasındaki mesafeyi hesaplayarak en yakın k komşuyu belirlemektir. KNN daha sonra bu komşular arasında en yaygın olan sınıf etiketini x örneğine atar [43].

Özellikle, iki veri noktası arasındaki mesafe, mesafe matrisi hesaplanarak belirlenir. K-en Yakın Komşu algoritmasında Minkowski, Chebyshev ve Öklid gibi

çeşitli mesafe hesaplama teknikleri vardır, ancak en yaygın kullanılan mesafe matrisi Öklid Mesafesi'dir [44].

3.1.2.1. Öklid mesafesi

Öklid mesafesi, noktalar arasındaki mutlak mesafeyi veya çok boyutlu uzayda vektörlerin doğal uzunluğunu ölçer. Aşağıda verilen denklem genellikle Minkowski mesafe denklemi olarak adlandırılır. Genel mesafe denklemi, denklem 7'de verilmiştir.

$$D = \sqrt[p]{\sum |x_{train_norm} - x_{test_norm}|^p} \quad (7)$$

p=2 olduğunda, denklem 8, Minkowski mesafesi Öklid mesafesine eşit olma durumunu gösterir.

$$D = \sqrt{\sum |x_{train_norm} - x_{test_norm}|^2} \quad (8)$$

Öklid mesafesinin hesaplanması, her boyuttaki özelliklerin mutlak farkına dayanır. Bu nedenle, Öklid mesafesini bir parametre olarak kullanırken, iki vektörün karşılık gelen boyutlarının aynı ölçü birimine sahip olduğundan emin olmak çok önemlidir. Örneğin, farklı birimlerdeki hız (m/s) ve kuvvet (N) farkı hesaplanacaksa, ölçekleme teknikleri uygulanmalıdır. Öklid mesafesi ayrıca L2 mesafesi olarak da adlandırılır [45], [46].

3.1.2.2. Manhattan mesafesi

Manhattan mesafesi, dik açı yapan eksenler boyunca ölçülen iki nokta arasındaki mesafedir. Ayrıca L1 mesafesi veya taksi mesafesi olarak da bilinir. Diğer mesafe metriklerine göre daha az hesaplama gerektirir. p=1 olarak ayarlandığında, Manhattan mesafesi elde edilir, denklem 9'da gösterildiği gibi.

$$D = \sum |x_{train_norm} - x_{test_norm}| \quad (9)$$

3.1.2.3. Chebyshev mesafesi

Chebyshev mesafesi, $p=\infty$ olduğunda Minkowski mesafesinin bir varyantıdır. Bu mesafe, maksimum değer mesafesi ve satranç tahtası mesafesi olarak da bilinir ve denklem 10'daki gibi formüle edilebilir:

$$D = \lim_{p \rightarrow \infty} (\sum |x_{train_norm} - x_{test_norm}|^p)^{\frac{1}{p}} = \max(|x_{train_norm} - x_{test_norm}|) \quad (10)$$

Chebyshev mesafesi, yüksek boyutlu bir uzayda kare hiper küp yapısındaki verileri ölçebilir. Kare hiper küp, tüm kenarları eşit uzunlukta olan çok boyutlu bir uzaydaki bir küptür. Chebyshev mesafesi ise, kare hiper küp üzerindeki mesafeleri ölçmek için uygundur. Çünkü maksimum eksen farkını dikkate alarak kare hiper küpün kenarları boyunca en kısa mesafeyi hesaplar. Öte yandan, Manhattan mesafesi, yüksek boyutlu bir uzayda eşkenar dörtgen hiper küp yapısındaki verileri ölçebilir. Eşkenar dörtgen hiper küp, yüksek boyutlu uzayda köşegenler boyunca şekillenmiş bir küptür. Bu yapı, eşkenar dörtgen şeklinin çok boyutlu bir uzantısı olarak düşünülebilir. Manhattan mesafesi, bu tür yapılar üzerindeki mesafeleri ölçmede etkilidir. Çünkü eşkenar dörtgen hiper küpün köşeleri arasında hareket ederken karşılaşılan mesafeleri doğru bir şekilde yansıtır [47], [48], [49].

3.1.3. Karar Ağacı

Karar ağacı (DT), karar verme ve sınıflandırma problemleri için çok güçlü ve popüler bir veri madenciliği algoritmasıdır. Tıbbi teşhis, radar sinyali sınıflandırması, hava durumu tahmini, kredi onayı ve sahtecilik tespiti gibi birçok gerçek hayattaki uygulamada kullanılmaktadır. Karar ağacı, büyük hacimli ve birçok özneliğe sahip veri setlerinden oluşturulabilir çünkü ağacın boyutu veri setinin boyutuna bağlı değildir. Bir karar ağacının üç ana bileşeni vardır: düğümler, yapraklar ve kenarlar. Her düğüm, verilerin bölümleneceği bir öznelik ile etiketlenmiştir. Her düğüm, özneliğin olası değerlerine göre etiketlenmiş bir dizi kenara sahiptir. Bir kenar ya iki düğümü ya da bir düğüm ile bir yaprağı bağlar. Yapraklar, verilerin sınıflandırılması için bir karar değeri ile etiketlenir. Bir karar ağacını kullanarak karar vermek kök düğümden başlar ve bir

yaprak düğümüne ulaşana kadar dallar boyunca aşağı doğru ilerlerler. Her karar ağacı, veri setinin özniteliklerine göre verileri sınıflandıran bir kural setini temsil eder [50].

3.1.3.1. Karar ağacı algoritması türleri

ID3 algoritması

ID3 algoritması, Ross Quinlan tarafından geliştirilmiş ve bir veri kümesinden karar ağacı oluşturmak için kullanılan bir algoritmadır. Her düğümde, veri sınıflandırması için en yüksek bilgi kazancına sahip olan özellik, test özelliği olarak seçilir. ID3, Occam'ın usturası ilkesine dayalıdır ve daha küçük karar ağaçlarını tercih eder. Ancak her zaman en küçük ağacı oluşturmaz; bu nedenle, sezgisel bir algoritma olarak kabul edilir. ID3 algoritması yalnızca kategorik özellikleri kabul eder ve gürültülü veriler üzerinde çalışırken doğru sonuçlar vermeyebilir. Bu yüzden, karar ağacı oluşturmadan önce veri ön işleme yapılır. Her bir özellik için bilgi kazancı hesaplanır ve en yüksek bilgi kazancına sahip özellik kök düğüm olarak seçilir. Daha sonra, olası tüm çıktılar incelenir ve aynı sınıfa ait olan veriler bir sınıf adı altında toplanırken, farklı sınıflar için bölme işlemi uygulanır.

ID3 algoritmasının avantajları arasında, eğitilen veriyle kolay anlaşılır tahmin kuralları oluşturması, hızlı ve kısa bir karar ağacı inşa etmesi, veri setinin tamamını tarayarak tüm ağacı oluşturması ve test verilerinin düğüm sayısını azaltarak pratiklik sağlaması yer alır. Ancak küçük örneklem verilerinde aşırı sınıflandırma veya aşırı uyum gibi sorunlara yol açabilir. Ayrıca, sürekli verilerin sınıflandırılması pahalı olabilir ve çok sayıda özelliği olan veri kümelerinde aşırı duyarlı olma eğilimindedir. ID3, basitlik sunmasına rağmen, zaman ve hesaplama açısından bazı dezavantajlara sahiptir [51].

C4.5 algoritması

C4.5 algoritması, Quinlan'ın ID3 algoritmasının bir uzantısı olan C4.5, sınıflandırma işlemleri için karar ağaçları oluşturur ve bu nedenle genellikle istatistiksel sınıflandırıcı olarak anılır. ID3 algoritmasına kıyasla daha gelişmiş olan C4.5, hem sürekli hem de kategorik özelliklerle çalışabilir, eksik değerlerle başa çıkabilir ve ağaç

oluřturma sonrasında budama iřlemi yapar. C4.5, varsayılan olarak bir aęaę budama iřlemi geręekleřtirir, bu da daha kk aęaęlar, daha basit kurallar ve daha sezgisel yorumlamalar saęlar.

C4.5 algoritması  temel adımı takip eder: Kategorik zelliklerin ayırımı iin ID3 ile benzer bir yaklařım kullanılır, srekli zellikler iinse ikili ayrımlar oluřturulur. Daha sonra, en yksek kazan oranına sahip zellik seilir. Bu iřlemler, her yeni dal iin tekrarlanır ve durdurma kriteri kontrol edildikten sonra aęaę bymesi sona erer.

C4.5 algoritmasının avantajları arasında, kolay uygulanabilirlięi, kolayca yorumlanabilir modeller oluřturması hem kategorik hem de srekli deęerlere uyum saęlaması, grltyle bařa ıkabilmesi ve eksik deęerlere karřı toleranslı olması yer alır [52].

CART

CART (Classification And Regression Trees), 1984 yılında Breiman tarafından tanıtılan ve hem sınıflandırma hem de regresyon aęaęları oluřturabilen bir algoritmadır. CART algoritması, sınıflandırma aęacını zelliklerin ikili ayrımlarıyla oluřturur ve bu iřlemdede Gini İndeksi, blme zellięini semek iin kullanılır. Aynı zamanda regresyon analizi yapmak iin de regresyon aęacını kullanır. CART'ın regresyon zellięi, belirli bir sre boyunca bir baęımlı deęiřkenin bir dizi baęımsız deęiřken kullanılarak tahmin edilmesi amacıyla kullanılabilir.

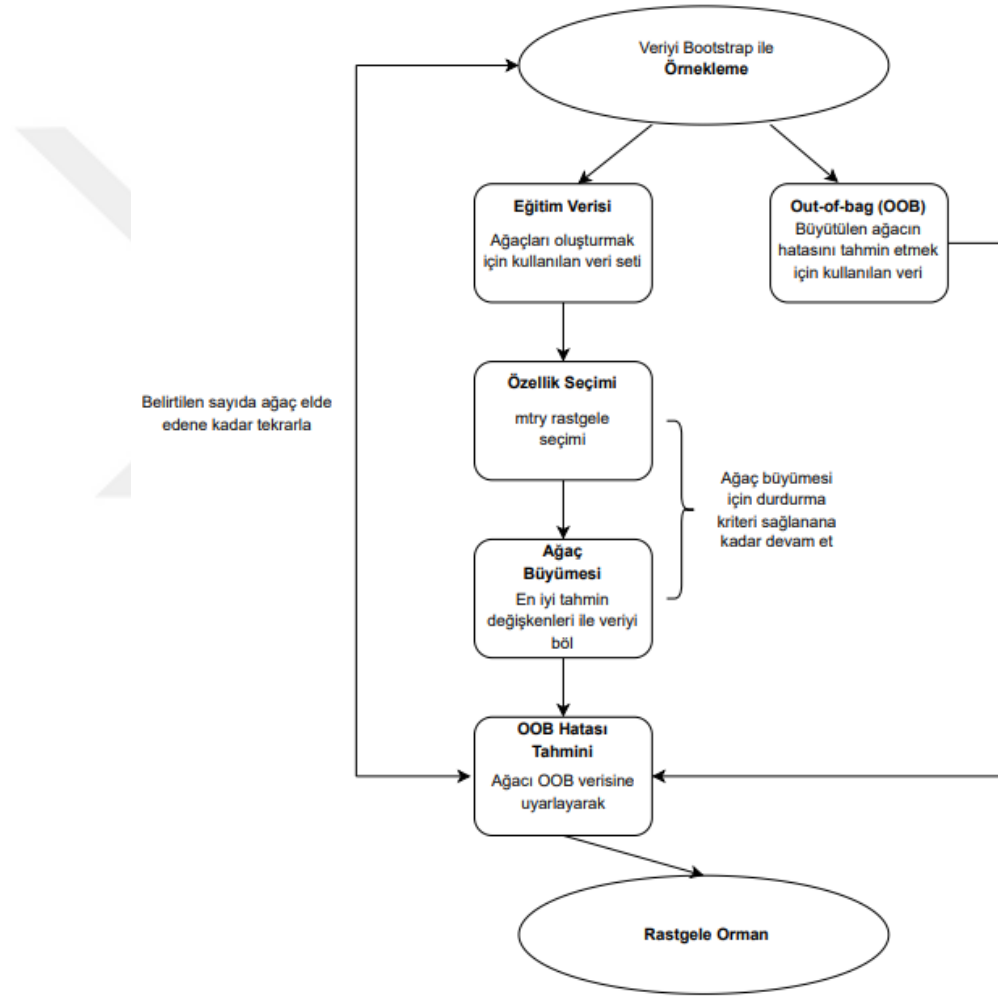
CART algoritmasının avantajları arasında eksik deęerlerle otomatik olarak bařa ıkabilme yeteneęi srekli ve kategorik deęiřkenlerin herhangi bir kombinasyonunu iřleyebilmesi, otomatik deęiřken seimi yapabilmesi ve deęiřkenler arasındaki etkileřimleri belirleyebilmesi bulunur [53].

3.1.4. Rastgele Orman

Random Forest, 2001 yılında Breiman tarafından ilk kez tanıtılan ve tahmin modelleri geliřtirmek iin kullanılabilen popler bir makine ęrenimi algoritmasıdır [54]. Breiman'ın yaklařımında, her bir karar aęacı, her dęmde rastgele seilen bir zellik kmesi zerinden blnerek oluřturulur. Bu zellikler zerinden eęitim verisi kullanılarak en iyi blnme noktası hesaplanır. CART algoritması kullanarak aęaęlar

budama yapılmadan maksimum derinliğe kadar büyütülür. Bu alt uzay rastgeleleştirme yöntemi, torbalama adı verilen bir teknikle birleştirilerek her bir ağacın farklı bir eğitim veri örneği üzerinde oluşturulmasını sağlar. Torbalama yönteminde, eğitim verisinden tekrarlı olarak örnekler çekilir ve her örnekte farklı bir ağaç yetiştirilir [55]. Bu, aynı örneğin birden fazla kez seçilebileceği, diğerlerinin ise hiç seçilmeyebileceği anlamına gelir.

Algoritmanın her bir ağacı şekil 3.1’de olduğu gibi inşa edilir:



Şekil 3.1. Rastgele Orman Algoritması Adımlar (G. Biau, 2012)

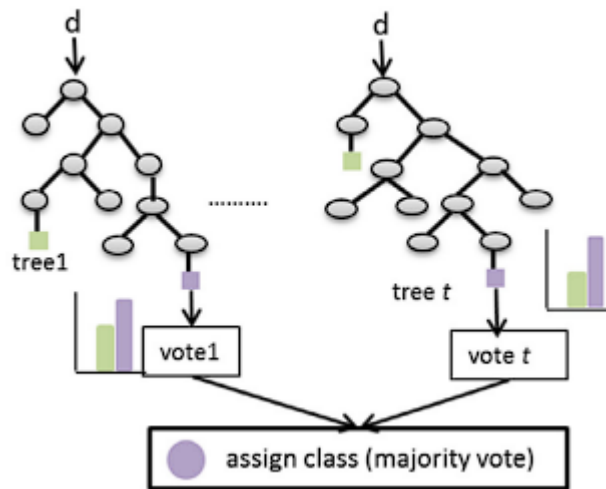
1) Bootstrap Örnekleme: Eğitim setinden örnekler, yer değiştirmeli (bootstrap) bir yöntemle seçilerek bir alt küme oluşturulur. Bu alt küme, orijinal setle aynı büyüklüktedir; ancak bazı örnekler birden fazla kez seçilirken, bazıları hiç seçilmez.

Örnekleme dışında kalan veriler, **Out-Of-Bag (OOB)** olarak adlandırılır ve bu veriler, modelin hata oranını tahmin etmek için kullanılır.

2) Ağacın İnşası: Bootstrap ile oluşturulan örnekler, tür bakımından daha homojen alt gruplar oluşturacak şekilde, özelliklere göre bölünerek bir karar ağacı oluşturulur. Her bir düğümde, mtry sayıda rastgele seçilen özellik değerlendirilir ve en iyi bölünme noktası seçilerek ağaç büyütülür. Bu işlem, her alt düğümde farklı bir mtry özellik seti ile tekrarlanır. Rastgele Orman algoritmasının performansı, mtry parametresinin değerine genellikle duyarlıdır. Bu nedenle, varsayılan olarak, mtry değeri özellik sayısının karekökü olarak alınır. Ancak özel durumlarda bu değer kullanıcı tarafından ayarlanabilir.

3) Maksimum Ağaç Oluşturma: Her bir terminal düğümde yalnızca tek bir tür veriyi (veya uç durumlarda yalnızca bir örneği) içeren bir "maksimum ağaç" inşa edilir[56].

Algoritma, ormanı kullanıcı tarafından tanımlanan ağaç sayısına (Ntree) kadar büyütür ve yüksek varyans ve düşük bias'a sahip ağaçlar oluşturur. Son sınıflandırma kararı, tüm üretilen ağaçlar tarafından hesaplanan sınıf atama olasılıklarının ortalamasını (aritmetik ortalama kullanılarak) alarak alınır. Yeni bir etiketlenmemiş veri girişi, böylece toplulukta oluşturulan tüm karar ağaçlarına karşı değerlendirilir ve her ağaç bir sınıf üyeliği için oy verir. En fazla oyu alan üyelik sınıfı, nihayetinde seçilen sınıf olacaktır [57]. Şekil 3.2'de örnek bir oylama işlemi gösterilmiştir.

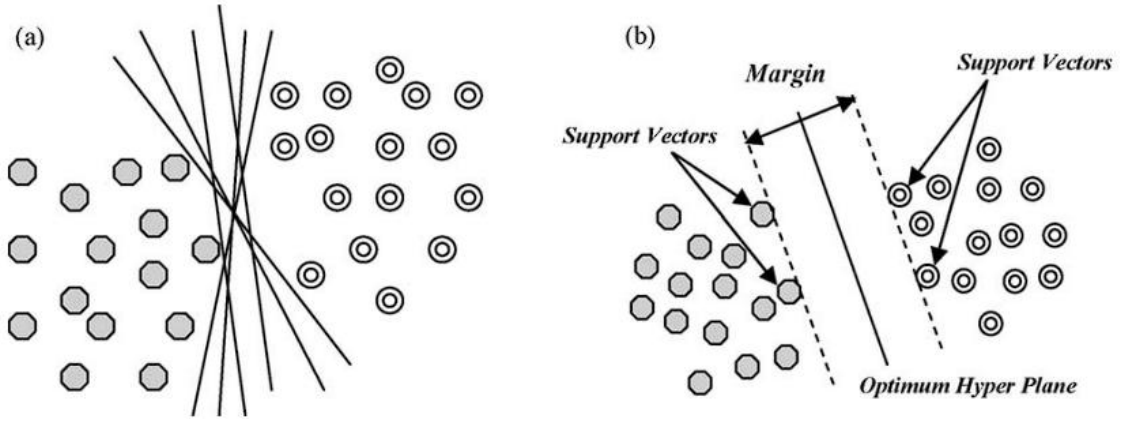


Şekil 3.2. Rastgele Orman Üyelik Sınıfı (M. Belgiu ve L. Drăguț, 2016)

3.1.5. SVM (Destek Vektör Makineleri)

Destek Vektör Makineleri (SVM), istatistiksel öğrenme teorisine dayanarak geliştirilmiştir ve yapısal risk minimizasyonu hipotezinden türetilmiştir. SVM, iyi bir genelleme yeteneği elde etmek için hem ampirik riski hem de modelin genelleme yeteneğini artırmak için yapısal risk minimizasyonu prensibini uygular. Ampirik risk, bir modelin eğitim verisi üzerindeki hata oranını ifade eden bir ölçüttür [58]. Ayrıca SVM sınıflandırma ve regresyon için kullanılan bir dizi denetimli öğrenme yönetimini ifade eder ve geliştirilmiş doğrusal sınıflandırıcılar kategorisinde değerlendirilir. Destek Vektör Makineleri, yüksek boyutlu özellik uzayında doğrusal fonksiyonların hipotez alanını kullanan ve istatistiksel öğrenme teorisinden türetilen bir öğrenme yanlılığı uygulayan optimizasyon teorisi ile eğitilen sistemler olarak tanımlanabilir [59].

SVM'lerin sınıflandırma için amacı, iki sınıfı en iyi şekilde ayıran bir hiper düzlem belirlemektir. Optimum hiper düzlem, eğitim veri setleri kullanılarak belirlenir ve genelleme yeteneği test veri setleri ile doğrulanır.



Şekil 3.3. Hiper Düzlem (T. Kavzoglu ve I. Colkesen, 2009)

K boyutlu veri setlerinin sınıflandırılması için, k-1 boyutlu bir hiper düzlem üretilir. Şekil 3.3 (a), iki sınıfı ayıran birden fazla hiper düzlem bulunmasının bir örneğidir. Ancak iki sınıf arasında maksimum marjini sağlayan yalnızca bir hiper düzlem vardır (Şekil 3.3(b)) ve bu hiper düzlem "optimum hiper düzlem" olarak adlandırılır. Marjin genişliğini sınırlayan noktalara ise "destek vektörleri" denir. Destek Vektörü denklem 11'de gösterilmiştir.

Destek vektör makineleri (SVM), ikili sınıflandırma problemlerinde, iki sınıf arasındaki marjı maksimize eden bir hiper düzlem bulmayı hedefler. Şekil 3.4'te eğitim veri seti $\{x_i, y_i\}$ olarak temsil edilen k örnekten oluştuğu düşünülürse. $X_i \in \mathbb{R}^N$, N boyutlu uzayda bir veri noktasını, $Y_i \in \{-1, +1\}$ ise bu veri noktasının sınıf etiketini yani hangi sınıfa ait olduğunu temsil eder. Sınıflar algoritmada, denklem 11'de olduğu gibi ifade edilir.

$$w \cdot x_i + b \geq +1 \quad (11)$$

$$w \cdot x_i + b \leq -1$$

Yukarıda verilen eşitsizliklerin tek bir denklemde yazılışı denklem 12'de verilmiştir.

$$y_i(w \cdot x_i + b) - 1 \geq 0 \quad (12)$$

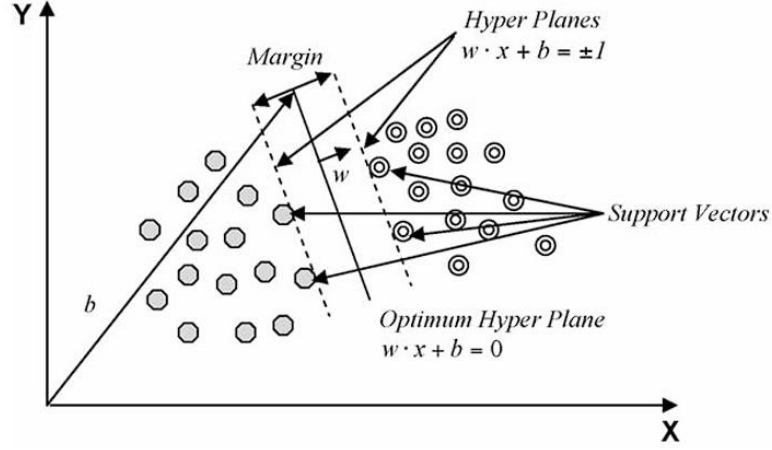
Optimum hiper düzlemin amacı, her iki sınıftaki en yakın veri noktalarının (destek vektörlerinin) hiper düzleme olan uzaklıklarının maksimum olmasını sağlayarak bu iki sınıfı ayırmaktır. Hiper düzlemin denklemi denklem 13'te gösterildiği gibidir. Denklemde w , hiper düzlemin N boyutlu uzayda yönünü belirler. Denklemdeki b ise hiper düzlemin orijinden olan uzaklığını ifade eden önyargıdır (bias). Eğer denklem 12'yi sağlayan bir hiper düzlem varsa, sınıflar doğrusal olarak ayrılabilir. Bu durumda, bu düzlemler arasındaki marj denklem 14'e eşit olur.

$$w \cdot x_i + b = \pm 1 \quad (13)$$

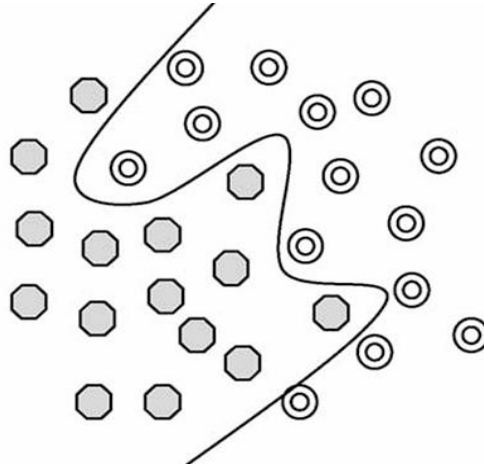
$$2/\|w\| \quad (14)$$

Hiper düzlemi ayırma işlemi optimum olarak, Denklem 12 koşulu altında $\|w\|^2$ değerini en aza indirerek bulunabilir. Bu nedenle, optimum hiper düzlemin belirlenmesi, denklem 15'in optimizasyon problemini çözmeyi gerektirir:

$$\min\left[\frac{1}{2} \|w\|^2\right] \quad (15)$$



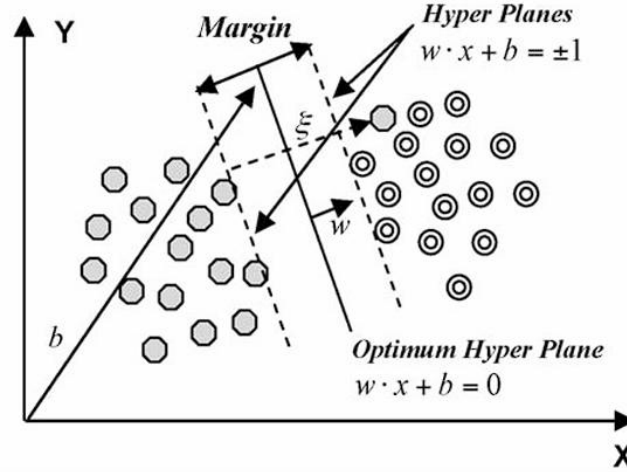
Şekil 3.4. Optimum Hiper Düzlem Belirleme
(T. Kavzoglu ve I. Colkesen, 2009)



Şekil 3.5. Doğrusal Olarak Ayrışamayan Veriler
(T. Kavzoglu ve I. Colkesen, 2009)

Şekil 3.5'te gösterildiği gibi, doğrusal olarak ayrıştırılamayan veriler, çeşitli sınıflandırma problemlerinde karşılaşılan bir durumdur. Bu gibi durumlarda, veri kümeleri giriş uzayında doğrusal bir fonksiyonla iki sınıfa ayrıştırılamaz. Eğitim verileri üzerinde doğrusal denklemlerle tanımlanan bir hiper düzlem oluşturmanın mümkün

olmadığı durumlarda, doğrusal olmayan karar yüzeylerine izin verecek şekilde genişletilebilir. Bu nedenle, optimizasyon problemi ξ gevşeme değişkeni kullanılarak şekil 3.6’da gösterildiği gibi yeniden düzenlenir.



Şekil 3.6. Gevşeme Değişkeni ile Optimum Düzlem Belirleme
(T. Kavzoglu ve I. Colkesen, 2009)

Burada C ifadesi, bir düzenleme sabiti veya ceza parametresidir. Bu parametre, marjı maksimize etme ve hatayı minimize etme gibi iki rekabet eden kriter arasında bir denge kurulmasını sağlar. Gevşeme değişkenleri ξ_i , yanlış sınıflandırılmış noktaların optimum hiper düzleme olan uzaklığını ifade eder. C değeri ne kadar büyük olursa, yanlış sınıflandırılmış örneklere uygulanan ceza o kadar yüksek olur. Yeni optimizasyon denklemi ceza parametresi uygulanarak denklem 16’daki hale dönüşür. Bu denklemde r ifadesi, veri setindeki toplam örnek sayısıdır ve optimizasyonun tüm veri noktalarını kapsamasını sağlar [60].

$$\left[\frac{\|w\|^2}{2} + C \sum_{i=1}^r \xi_i \right] \quad (16)$$

SVM’de çeşitli kernel fonksiyonları vardır. Kernelin işlevi, veriler doğrusal bir hiper düzlem ile ayrıştırılamıyorsa, kernel fonksiyonları bu verileri daha yüksek boyutlu bir uzaya dönüştürür. Bu dönüşüm sayesinde, doğrusal olarak ayrışamayan veriler daha yüksek boyutlu bir uzayda doğrusal bir hiper düzlem ile ayrıştırılabilir hale gelir. Bazı kerneller şu şekildedir:

3.1.5.1. Lineer kernel

Doğrusal kernel, aynı zamanda soft margin olarak da bilinir ve bir hiper düzlemi (düz bir çizgi) bulmaya çalışır ancak bir veya daha fazla veri yanlış sınıflandırmasını tolere edebilir. Verilen yanlış sınıflandırma toleransı miktarı, hiper düzlemin doğruluğunu büyük ölçüde etkiler. Sklearn'de bu tolerans C olarak adlandırılır. C değerinin daha büyük olması, daha az yanlış sınıflandırma toleransı ve daha dar bir marj anlamına gelir. Lineer Kernel denklemi, denklem 17'de gösterilmiştir.

$$K(x, z) = x \cdot z + C \quad (17)$$

3.1.5.2. Polinomsal Kernel

Polinomsal kernel, mevcut özelliklerin polinom kombinasyonlarını uygulayarak yeni özellikler oluşturur. Polinomsal kernelde, hiper düzlemin eğriliğini belirleyecek olan polinom derecesinin değeri olan d parametresi bulunur. Polinomsal Kernel, denklem 18'de gösterildiği gibidir.

$$K(x, z) = (x \cdot z + C)^d, d > 0 \quad (18)$$

3.1.5.3. RBF Kernel

RBF (Radial Basis Function) kernelinde, verilerin haritalanmasını etkileyen bir gamma parametresi bulunur. Gamma, karar sınırı içinde ne kadar eğriliğe izin verildiğinin bir ölçüsüdür; gamma değeri ne kadar büyük olursa, hiper düzlemde o kadar fazla eğrilik oluşur [61]. RBF Kernel denklemi, denklem 19'da gösterildiği gibidir.

$$K(x, z) = \exp(-\gamma \|x - z\|^2), \gamma > 0 \quad (19)$$

4. METOT

Bu çalışmada, sürücü yorgunluk seviyelerini sınıflandırmak ve değerlendirmek amacıyla hem sürücülerden hem de araçlarından toplam 10 saatlik bir veri seti oluşturulmuştur. Veri seti, sürücülerden alınan EKG sinyalleri ve araçlardan alınan CAN-Bus sinyallerini içermektedir. Veriler, farklı yaşa sahip bireylerden ve cinsiyetlerden toplanmıştır. 18 ile 45 yaşları arasındaki, farklı cinsiyetlerden sürücülerden veri toplama sürecinde araçlarını istedikleri süre boyunca kullanmaları istenmiştir. Araç kullanım süresiyle ilgili herhangi bir kısıtlama getirilmemiştir. Sürücülerin özellikleri Tablo 4.1’de detaylandırılmıştır.

Tablo 4.1. Adayların Özellikleri

Sürücü #	Cinsiyet	Sürüş Süresi (dakika)	Yaş
Sürücü_1	Erkek	120	38
Sürücü_2	Erkek	130	28
Sürücü_3	Erkek	130	29
Sürücü_4	Erkek	25	25
Sürücü_5	Kadın	25	43
Sürücü_6	Kadın	25	44
Sürücü_7	Kadın	40	41
Sürücü_8	Kadın	45	35
Sürücü_9	Kadın	60	34

Çeşitli makine öğrenme algoritmaları, K-En Yakın Komşu (KNN), Rastgele Orman, XGBoost, Karar Ağacı ve Destek Vektör Makineleri (SVM) kullanılmış ve karşılaştırılmıştır. Makine öğrenme algoritmalarının bağımlı ve bağımsız değişkenleri, sırasıyla sürücülerden eşzamanlı olarak toplanan EKG verileri ve test araçlarından alınan sinyallerdir. Bu sinyaller, direksiyon açısı, direksiyon hızı, direksiyon torku, gaz pedalı eğimi, motor torku, araç hızı, gaz pedalı konumu ve ani frenleme tepkileri gibi birtakım sinyalleri içermektedir. Toplanan sinyaller CANalyzer programı kullanılarak işlenmiştir. Tablo 4.2, aracın CANBus hattından okunan sinyallerin adları ve açıklamaları hakkında bizlere ayrıntılı bilgiler sunmaktadır.

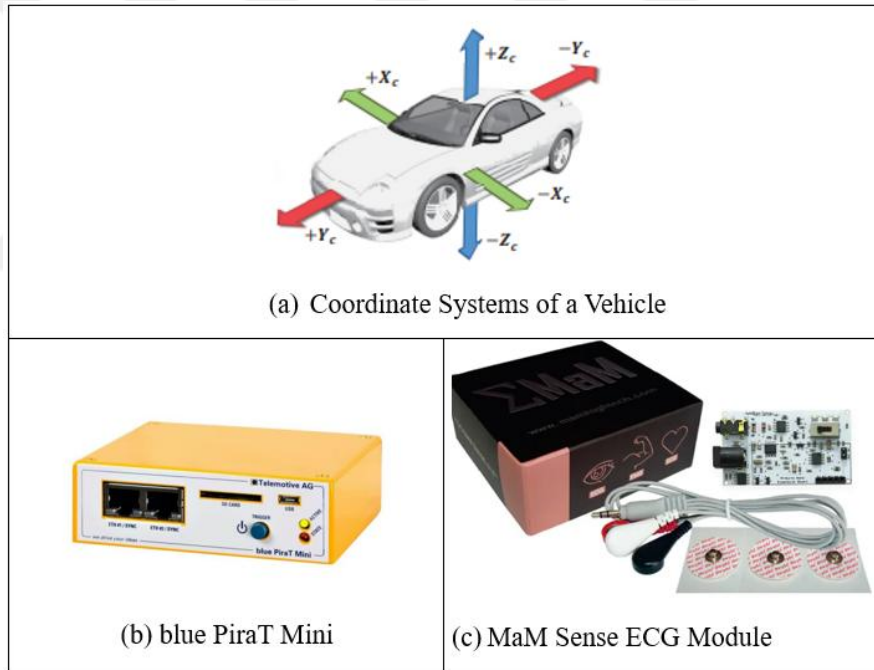
Tablo 4.2. Araç Sinyallerinin Özellikleri

Sinyal İsmi	Açıklama
Direksiyon Açısı	Direksiyon açısını belirtir. (Saat yönünün tersine pozitif)
Direksiyon Hızı	Direksiyon simidi hızını belirtir (Saat yönünün tersi pozitif)
Direksiyon Kolonu Torku	EPS modülünden sürücünün direksiyon simidine uyguladığı tork kuvvetini algılar.
Motor Torku	Motor torkunu belirtir.
Gaz Pedalı Durumu	Sürücünün gaz pedalına basıp basmadığını belirtir. 0 veya 1 değerini alır (0: Basılmamış, 1: Basılmış).
Gaz Pedalı Gradyant	Gaz pedalının pozisyonundaki değişim hızını zaman içinde belirtir, saniye başına yüzde olarak ölçülür (%/s).
Motor Hızı	Motor devrini belirtir (RPM).
Gaz Pedalı Pozisyonu	Sürücünün gaz pedalına bastığı andaki pedal pozisyonunu belirtir.
Şanzıman Durumu	Şanzımanın boşa (nötr) olup olmadığını belirtir.
Fren Pedalı Durumu	Fren pedalının basılı olup olmadığını belirtir. Sinyal, pedal basılı değilse 1, basılıysa 0 değerini alır.
ABS Aktif Durumu	ABS sisteminin etkinleştirilip etkinleştirilmediğini kontrol eder.
Fren Müdahale Durumu	ACC (Adaptif Hız Sabitleyici), ESC (Elektronik Denge Kontrolü) ve ASR (Hızlanma Kayma Düzeni) gibi sürücü dışındaki fren müdahalelerini belirtir.
Acil Frenleme Sistemi Durumu	Acil frenleme sisteminin etkinleştirilip etkinleştirilmediğini belirtir.
ESC Aktif Durumu	ESC (Elektronik Denge Kontrolü) sisteminin etkinleştirilip etkinleştirilmediğini kontrol eder.
Araç Hız Bilgisi	Araç hızını belirtir.
Yanal İvme	Aracın yanal ivmesini temsil eder. (X eksen, Şekil-12(a)).
İleri-Geri İvme	Aracın ileri-geri yönündeki ivmesini gösterir. (Y eksen, Şekil 12(a)).
Z Eksen İvme	Aracın Z eksen boyunca ivmesini temsil eder. Devrilme sensörü, aracın devrilme durumunu ve diğer ilgili koşulları kontrol eder. (Şekil 5.1(a)).

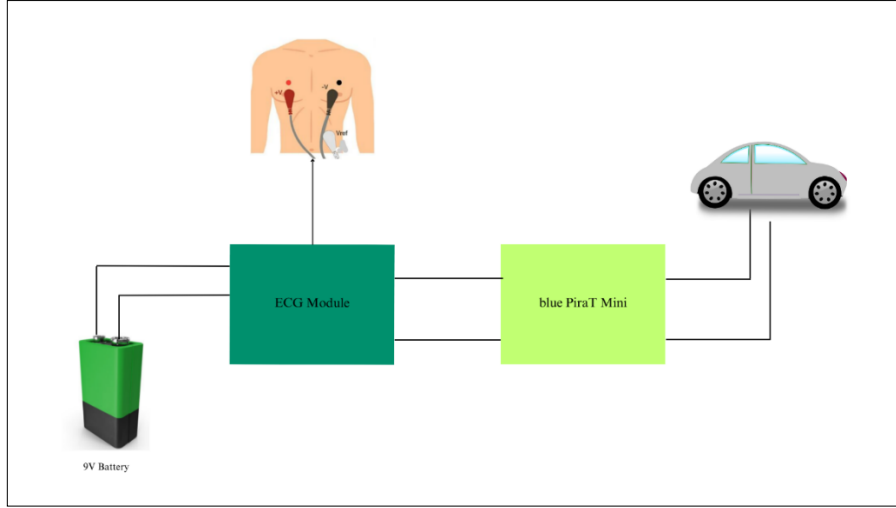
5. ARAÇTAN ve ADAYLARDAN VERİ TOPLAMA

EKG verilerinin ve CAN-Bus sinyallerinin eş zamanlı olarak toplanması, Şekil 5.1(b)'de gösterilen Blue PiraT Mini cihazı kullanılarak gerçekleştirilmiştir. Blue PiraTMini, hem analog verilerin hem de CAN sinyallerinin eşzamanlı elde edilmesini sağlar. Sürücülerden alınan EKG verileri ise, Şekil 5.1 (c)'de gösterilen MaM Sense EKG modülü kullanılarak toplanmıştır.

Şekil 5.2, bir sürücünün fizyolojik durumunu izlemek için kullanılan bir elektrokardiyogram (EKG) modülüne dayalı sistemi göstermektedir. Göğüse yerleştirilen elektrotlar, 9V pil ile çalışan EKG modülünün kalbin elektriksel aktivitesini izlenmesine olanak tanır. EKG modülü tarafından toplanan veriler, Blue PiraT Mini cihazında okunup hafızasında depolanmıştır.



Şekil 5.1. Blue PiraT Mini ve MaM Sense EKG modülü



Şekil 5.2. Sistem Tasarımı

Makine öğrenmesinde kullanılacak hedef değişken yani bağımlı değişken "sürücü yorgunluk seviyesi" için sınıf sayısının belirlenmesi gereklidir. Bu çalışmanın ana amacı, oluşturulan makine öğrenmesi modelinin sadece CAN-Bus girdileri ile sürücü yorgunluğunu sınıflandırmaktır. Ancak denetimli öğrenme (supervised learning) algoritmalarının tahmin edeceği sınıf sayısını objektif olarak belirlemek ve EKG verilerini analiz etmek Kubios yazılımı ve MATLAB programları kullanılmıştır.

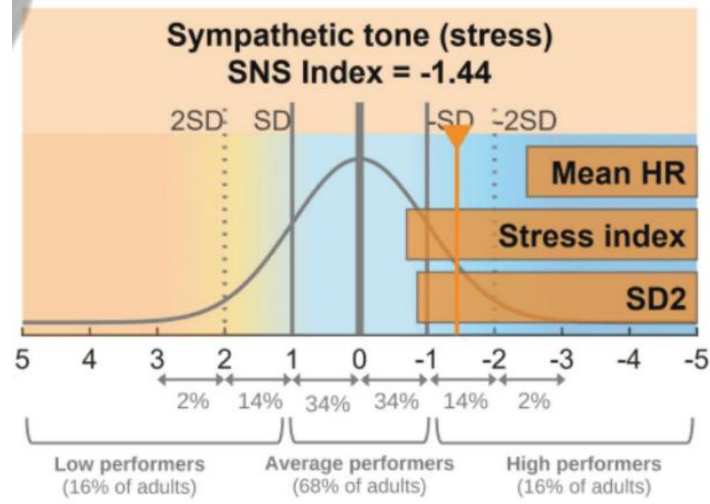
Elektrokardiyogram (EKG) sinyalinin R dalgalarını tespit etmek için Amplitude-Based Technique (ABT) kullanılmıştır. R dalgaları, QRS kompleksinin bir parçasını oluşturdukları için önemlidir ve bu kompleks, kalp ritimlerinin anlaşılması açısından kritik öneme sahiptir. ABT yöntemi, ECG sinyalinde genliğin belirlenen eşik değerinin üzerinde olduğu aralıkları arar. Bu aralıklar, her bir kalp atışı sırasında ventriküllerin depolarizasyonu ile ilişkili elektriksel aktiviteyi temsil eden muhtemel QRS kompleksleri olarak varsayılır. QRS kompleksini oluşturan üç ana dalga Q, R ve S dalgalarıdır. Bunlar arasında en belirgin olanı R dalgasıdır ve bu, kalp atış hızının doğru bir şekilde değerlendirilmesi için gereklidir. Yöntem, QRS komplekslerini tanımladıktan sonra her bir kompleksin zirvesini bulmayı amaçlar. R dalgası, QRS kompleksinin maksimum genlik noktası olup, daha fazla inceleme için kritik bir sinyaldir. Doğru R dalgası tespiti, kalp atış hızı değişkenliğinin değerlendirilmesi için oldukça önemlidir [62]. Tablo 5.1, örnek bir veri kümesi kullanılarak hesaplanan R-R aralığı mesafelerini göstermektedir.

Tablo 5.1. Hesaplanan RR süreleri

0.600000; 0.880000	10.670000; 0.530000
1.480000; 0.580000	11.750000; 0.580000
2.660000; 0.740000	12.570000; 0.600000
0.280000; 0.520000	13.740000; 0.560000
3.320000; 0.550000	14.850000; 0.510000
3.850000; 0.820000	15.960000; 0.580000
4.670000; 0.650000	16.470000; 0.520000
5.320000; 0.520000	17.640000; 0.560000
5.840000; 1.500000	18.240000; 0.590000
7.340000; 0.660000	19.430000; 0.520000
8.560000; 0.560000	21.470000; 0.560000
9.110000; 0.510000	22.680000; 0.510000
9.620000; 0.510000	23.190000; 0.580000
10.130000; 0.540000	

R-R aralığı mesafeleri belirlendikten sonra, veriler HRV (Kalp Hızı Değişkenliği) analizi için Kubios yazılımı kullanılarak işlenmiştir. Bu analiz, çalışmanın makine öğrenme aşamasında kullanılan sınıfların yani hedef değişkenin objektif olarak belirlenebilmesi için ele alınmıştır. Sınıflar, R-R aralıklarından elde edilen frekans alanındaki PNS (Parasempatik Sinir Sistemi) ve SNS (Sempatik Sinir Sistemi) indeksi hesaplamalarına dayanarak Kubios programında 0, 1 ve 2 olarak sınıflandırılmıştır.

Tahmin edilen sinyalin güç frekans dağılımının (düşük frekans (LF), yüksek frekans (HF) ve (LF/HF) frekans alanı özelliklerinin tahmininde etkili olduğu bilinmektedir. Stres altında, kalp hızı (HR) parasempatik (vagus) sinir sistemi tarafından düşerken, sempatik sinir sistemi tarafından yükseltilir [63]. Bu durumda, artan parasempatik aktivite, daha yüksek PNS indeksi değerleri ile ilişkilidir, bu da daha iyi otonom fonksiyonu ve daha rahat bir durumu gösterir [64]. Bununla birlikte, diğer çalışmalar yüksek LF/HF oranının sempatik baskınlığı gösterdiğini ve bunun yorgunlukla ilişkili olabileceğini öne sürmüştür [65]. Şekil 5.3, Kubios'un sempatik indeks değeri çalışmasının yorumlarını göstermektedir.



Şekil 5.3. Kubios SNS indeksi Yorumlaması

Bu yöntem kullanılarak verilerin LF/HF oranları her üç dakikada bir Kubios programında hesaplanmıştır. Sınıflandırma sürecini hızlandırmak ve aşırı sayıda sınıf oluşumundan kaçınmak için, LF/HF oranlarından türetilen SNS İndeks değerleri en yakın tam sayıya yuvarlanmıştır. Örneğin, 2.55 olan değer, 3'e ve 4.37'nin ise 4'e yuvarlanması işlemi yapılmıştır. Bu yuvarlama işlemi sonucunda oluşan sınıflar, Tablo 5.2'de belirtilmiştir. Ayrıca, araçtan 500 milisaniyede CAN-Bus verisi işlenmiştir. Üç dakikalık aralıklarla yapılan LF/HF oranı hesaplamaları, sınıflandırma algoritmalarında kullanılmak üzere bu 500 milisaniyelik aralıklarla kaydedilen CAN-Bus sinyallerinin çıktısı (Drowsiness_Level) bileşenleri ile birleştirilmiştir.

Tablo 5.2. Belirlenen Sınıflar

RoundedStress Index Value	Class
0 and 1	0 (Not drowsy)
2	1 (Slightly drowsy)
3,4,5	2 (Drowsy)

Kubios yazılımı kullanılarak örnek bir R-R aralık veri setinin segmentasyon süreci, belirlenen sempatik indeks değeri ile Şekil 5.4'te görsel olarak gösterilmiştir.



Şekil 5.4. Kubios Yazılımı Arayüz

Önceki bölümlerde tanımlanan sınıflara göre, bu çalışmada sürücü yorgunluğunu tespit etmek amacıyla XGBoost, K-En Yakın Komşu (K-NN), Karar Ağacı, Destek Vektör Makineleri (SVM) ve Rastgele Orman gibi geleneksel makine öğrenmesi algoritmaları kullanılmıştır.

6. BULGULAR

Bu çalışmada, sürücü yorgunluğu tespit sistemi geliştirmek için, toplanan bir veri seti üzerinde K-NN, SVM, Karar Ağacı, Rastgele Orman ve XGBoost gibi geleneksel bazı makine öğrenmesi modellerinin performansları elde edilmiştir. Bu algoritmaların seçilmesinin en önemli sebepleri, güçlü genelleme yetenekleri, farklı veri yapılarıyla uyumlu olmaları ve hesaplama açısından verimli çözümler sunmalarıdır. Geleneksel makine öğrenmesi yöntemleri, doğrusal ve doğrusal olmayan ilişkileri modelleyebilme kapasiteleri sayesinde geniş bir problem yelpazesinde etkili sonuçlar verebilir. Ayrıca, kolay yorumlanabilir olmaları sayesinde modelin nasıl karar verdiğini anlamak ve gerekli iyileştirmeleri yapmak daha kolaydır. Bu tez çalışmasında, her modelin performansı, en iyi parametre kombinasyonlarına göre değerlendirilmiş ve diğer modellerle karşılaştırılmıştır.

6.1. XGBoost Algoritması Sonuçları

Tablo 6.1, XGBoost algoritmasının çeşitli parametre kombinasyonları ile elde edilen sonuçlarını göstermektedir. Bu sonuçlardan, `n_estimators` değeri 450, `learning_rate` değeri 0.01 ve `max_depth` değeri 8 olarak ayarlanan model, yüksek doğrulama doğruluğu ve eğitim doğruluğu değerlerinin birbirine yakın olması nedeniyle seçilmiştir.

Tablo 6.1.XGBoost Algoritması Parametreleri

# of Estimators	Learning Rate	Maxdepth	Subsample	Colsample bytree	Training Accuracy	Validation Accuracy	Cross Validation Mean Score
500	0.01	6	0.8	0.8	0.7681	0.7583	0.7591
100	0.1	10	0.8	0.8	0.9169	0.8201	0.8276
200	0.05	6	0.8	0.8	0.7924	0.7739	0.7754
100	0.2	4	0.8	0.8	0.774	0.7638	0.7603
100	0.1	6	0.7	0.7	0.7925	0.7758	0.7742

1000	0.01	6	0.8	0.8	0.7935	0.7754	0.7758
300	0.2	6	0.8	0.8	0.924	0.82	0.8265
200	0.1	10	0.8	0.8	0.9771	0.8365	0.8435
500	0.1	6	0.6	0.6	0.8968	0.8112	0.8169
400	0.05	4	0.8	0.8	0.771	0.7618	0.7608
100	0.1	6	0.8	0.8	0.7947	0.7765	0.7748
450	0.01	8	0.9	0.8	0.8077	0.7836	0.7846
250	0.01	8	0.9	0.9	0.7915	0.7738	0.7749

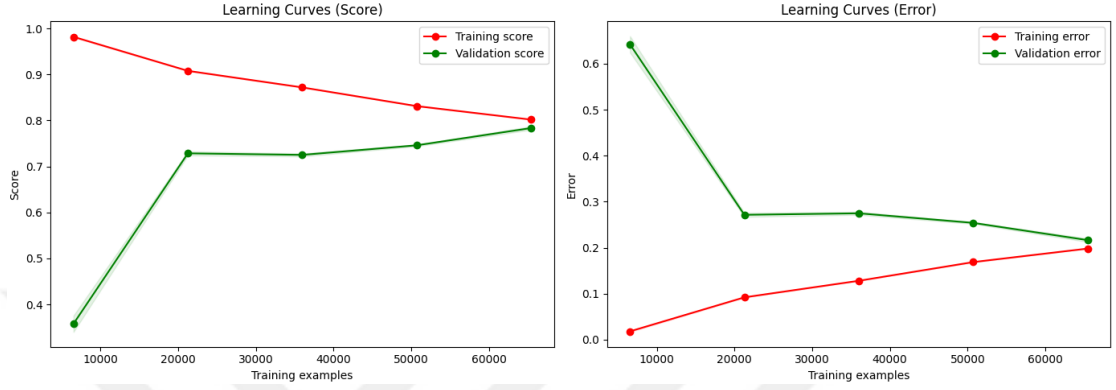
Tablo 6.2'ye dayanarak, modelin performansını değerlendirmek ve diğer sınıflarla karşılaştırıldığında Sınıf 0'ın en iyi performansı sergilediğini görmek mümkündür. Doğru sınıflandırılan 13928 örnekten yalnızca 390 Sınıf 0 örneği yanlış sınıflandırılmıştır. Yüksek geri çağırma (97%) ve doğruluk (79%) değerleri bunu desteklemekte olup, %87'lik bir F1 skoru elde edilmiştir. Bu, modelin Sınıf 0'ı tanımlama ve doğru bir şekilde kategorize etme konusunda oldukça yetkin olduğunu göstermektedir.

Tablo 6.2.XGBoost Algoritması Sınıflandırma Raporu

ConfusionMatrix				Class	precision	recall	f1-score	support	
Actual	0	13928	129	261	0	0.79	0.97	0.87	14318
	1	2721	1411	452	1	0.82	0.31	0.45	4584
	2	1026	179	1704	2	0.71	0.59	0.64	2909
	0	1	2	Macro Avg	0.77	0.62	0.65	21811	
	Predicted			WeightedAvg	0.78	0.78	0.75	21811	

Modelin Sınıf 1 için performansı, oldukça zayıftı. 1.411 doğru sınıflandırmaya kıyasla 3.173 örnek yanlış sınıflandırılmıştır. Ana sorun, 2.721 Sınıf 0 ve 452 Sınıf 2 örneği ile karışıklık yaşanmasıydı, bu da %31 geri çağırma, %82 doğruluk ve zayıf bir %45 F1 skoru ile sonuçlanmıştır. Sınıf 2 'de ise orta düzeyde bir performans sergilenmiştir; 1.704 doğru ve 1.205 yanlış sınıflandırma yapılmıştır. Sınıf 2 için F1 skoru %64, geri çağırma %59 ve doğruluk %71 olarak elde edilmiştir. 1.026 Sınıf 0 ve 179 Sınıf 1 örneği ile yaşanan karışıklık bu sonuçları etkilediği gözlemlenmiştir. Genel

olarak, modelin makro ortalama doğruluk değeri %77, geri çağırma %62 ve F1 skoru %65'ti, bu da özellikle Sınıf 1 ile ilgili önemli sınırlamalar olduğunu göstermektedir. Genel doğruluk %78 olmasına rağmen, Sınıf 1 ve Sınıf 2'deki sınıflandırma hataları nedeniyle ağırlıklı ortalama F1 skoru %75'e düşmüştür.



Şekil 6.1. XGBoost Algoritması Öğrenme Eğrileri

XGBoost modeline ait öğrenme eğrileri Şekil 6.1'de gösterilmektedir. "Öğrenme Eğrileri (Skor)" grafiğinde, başlangıçtaki yüksek performansa rağmen eğitim performansında zamanla hafif bir düşüş görülmektedir. Modelin eğitim verilerine başarılı bir şekilde uyum sağladığını, eğitim skorunun yüksek kalmasına rağmen kademeli olarak düştüğünü gösterir. Diğer taraftan, daha fazla veri ile doğrulama skoru, yaklaşık %35'ten %80'ye yükselmiştir. Bu durum da modelin genelleme yeteneklerinin arttığını göstermektedir. "Öğrenme Eğrileri (Hata)" grafiği ise daha fazla veri eklendikçe doğrulama hatasının hızla düştüğünü, model performansının ve genelleme yeteneğinin iyileştiğini gösterirken, eğitim hatası düşük kalmaya devam etmiştir, ancak hafif bir artış da gözlemlenmiştir.

6.2. KNN Algoritması Sonuçları

Farklı sonuçlar elde etmek için K-En Yakın Komşu (K-NN) algoritmasında kullanılan farklı parametre kombinasyonları Tablo 6.3'te gösterilmiştir. Eğitim doğruluğu %80.49 ve doğrulama doğruluğu %75.44 birbirine oldukça yakın olan N_Neighbors değeri 9, p değeri 1, metric değeri Manhattan ve weights değerinin uniform kombinasyonu bu sonuçlar arasında öne çıkmaktadır.

Tablo 6.3. KNN Algoritması Parametreler

N_Neighbors	Metric	P value	Weights	Accuracy	Validation Accuracy	Training Accuracy	Cross Validation Accuracy
3	Minkowski	2	distance	0.73	0.728	1	0.7336
5	Minkowski	2	distance	0.73	0.7338	1	0.7435
7	Minkowski	2	distance	0.74	0.7375	1	0.7482
9	Minkowski	2	distance	0.74	0.7417	1	0.7511
3	Minkowski	2	uniform	0.73	0.7292	0.8537	0.7391
5	Minkowski	2	uniform	0.73	0.7304	0.8167	0.7454
7	Minkowski	2	uniform	0.73	0.7342	0.8011	0.7457
9	Minkowski	2	uniform	0.74	0.7362	0.7885	0.7467
10	Minkowski	2	uniform	0.74	0.7369	0.7808	0.748
15	Minkowski	2	uniform	0.74	0.7391	0.7693	0.7445
20	Minkowski	2	uniform	0.74	0.7375	0.7607	0.7439
25	Minkowski	2	uniform	0.73	0.7341	0.7538	0.7428
30	Minkowski	2	uniform	0.73	0.7324	0.7506	0.7406
3	Manhattan	1	uniform	0.75	0.7521	0.8703	0.7636
5	Manhattan	1	uniform	0.76	0.7556	0.8343	0.7661
7	Manhattan	1	uniform	0.76	0.7556	0.8173	0.7661
9	Manhattan	1	uniform	0.75	0.7544	0.8049	0.7653
10	Manhattan	1	uniform	0.76	0.7559	0.7985	0.7648
15	Manhattan	1	uniform	0.75	0.7524	0.7853	0.7603
20	Manhattan	1	uniform	0.75	0.7501	0.7748	0.7576
25	Manhattan	1	uniform	0.75	0.748	0.7693	0.7558
30	Manhattan	1	uniform	0.75	0.7456	0.7637	0.753
3	Chebyshev	∞	uniform	0.71	0.705	0.8377	0.7258
5	Chebyshev	∞	uniform	0.71	0.7145	0.8004	0.7305
7	Chebyshev	∞	uniform	0.72	0.7169	0.7833	0.7333
9	Chebyshev	∞	uniform	0.72	0.7203	0.7735	0.736
10	Chebyshev	∞	uniform	0.72	0.7239	0.7673	0.7382
15	Chebyshev	∞	uniform	0.72	0.7224	0.7554	0.7355

20	Chebyshev	∞	uniform	0.72	0.7217	0.7469	0.7352
25	Chebyshev	∞	uniform	0.72	0.7223	0.7428	0.7337
30	Chebyshev	∞	uniform	0.72	0.7206	0.738	0.7321
3	Chebyshev	∞	distance	0.7	0.7033	1	0.7185

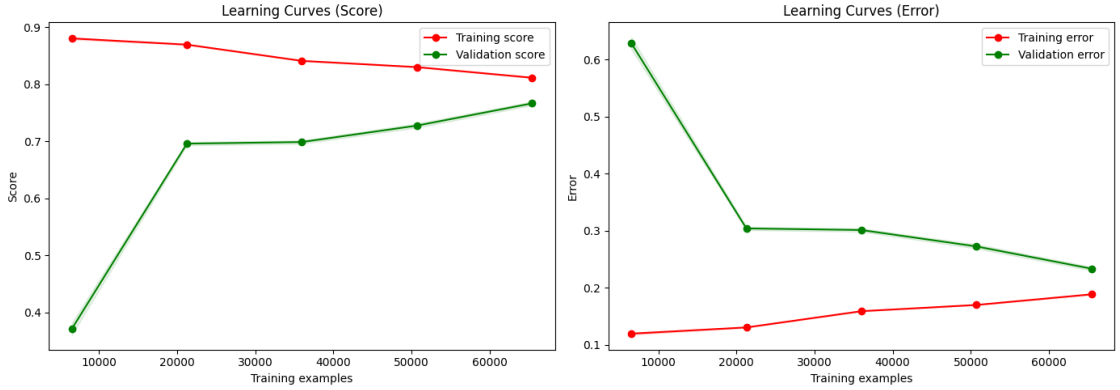
Tablo 6.4'teki Karışıklık Matrisi (Confusion Matrix), modelin Sınıf 0'ı sınıflandırmada doğruluk (0.79), geri çağırma (0.92) ve F1 skoru (0.85) değerleri ile iyi bir performans sergilediğini göstermektedir. Ancak özellikle modelin, Sınıf 0 ve Sınıf 1'i birbirine karıştırması nedeniyle Sınıf 1 için performanslar, geri çağırma (0.36), F1 skoru (0.46) ve doğruluk (0.63) değerleri ile daha kötü bir şekilde sonuçlanmıştır. Sınıf 2 için ise model, ortalama bir performans sergilemiştir. Sınıf 2 için F1 skoru 0.60, geri çağırma 0.57 ve doğruluk 0.64 olarak hesaplanmıştır. Genel olarak, doğruluk (0.69), geri çağırma (0.62) ve F1 skoru (0.64) ile makro-ortalama metrikler, Sınıf 0 için tatmin edici sonuçlar gösterirken, Sınıf 1 ve Sınıf 2 ile ilgili zorluklar saptanmıştır.

Tablo 6.4. KNN Algoritması Sınıflandırma Raporu

ConfusionMatrix				Class	Precision	Recall	F1-score	Support	
Actual	0	13212	675	431	0	0.79	0.92	0.85	14318
	1	2452	1639	493	1	0.63	0.36	0.46	4584
	2	971	291	1647	2	0.64	0.57	0.6	2909
	0	1	2	Macro Avg	0.69	0.62	0.64	21811	
	Predicted			WeightedAvg	0.74	0.76	0.74	21811	

Modelin öğrenme eğrileri Şekil 6.2'de gösterilmektedir. "Öğrenme Eğrileri (Skor)" grafiği, eğitim performansının başlangıçta yüksek olduğunu (yaklaşık %90), ancak veri seti büyüdükçe azaldığını, muhtemelen aşırı öğrenme (over fitting) nedeniyle olduğunu göstermektedir. Öte yandan, doğrulama skoru genelleme yeteneğinin iyileştiğini göstermektedir; başlangıçta düşük (yaklaşık %40) olan doğrulama skoru, daha fazla veri ile iyileşmekte ve yaklaşık %80 seviyesinde sabitlenmiştir. "Öğrenme Eğrileri (Hata)" grafiği, doğrulama hatasının başlangıçta yüksek olduğunu (yaklaşık %65), ancak veri seti büyüdükçe azalarak %20'ye düştüğünü, eğitim hatasının ise düşük kaldığını ve sadece hafifçe arttığını göstermektedir. Bu modeller, modelin başlangıçta

eğitim verileri üzerinde en iyi performansı gösterdiğini ve veri miktarı arttıkça doğrulama verilerine genelleme konusunda daha iyi hale geldiğini ortaya koymaktadır.



Şekil 6.2. KNN Algoritması Öğrenme Eğrileri

6.3. Karar Ağacı Algoritması Sonuçları

Sürücü yorgunluğu sınıflandırma problemi için Karar Ağacı algoritması, Tablo 6.5'te gösterildiği gibi farklı parametre kombinasyonları kullanarak çeşitli sonuçlar üretmiştir. Bu sonuçlar arasında dikkat çeken bir kombinasyon, max_depth: 12, min_samples_split: 6, min_samples_leaf: 2, criterion: 'entropy' ve splitter: 'best' parametreleri olarak gözlemlenmiştir.

Tablo 6.5. Karar Ağacı Algoritması Parametreler

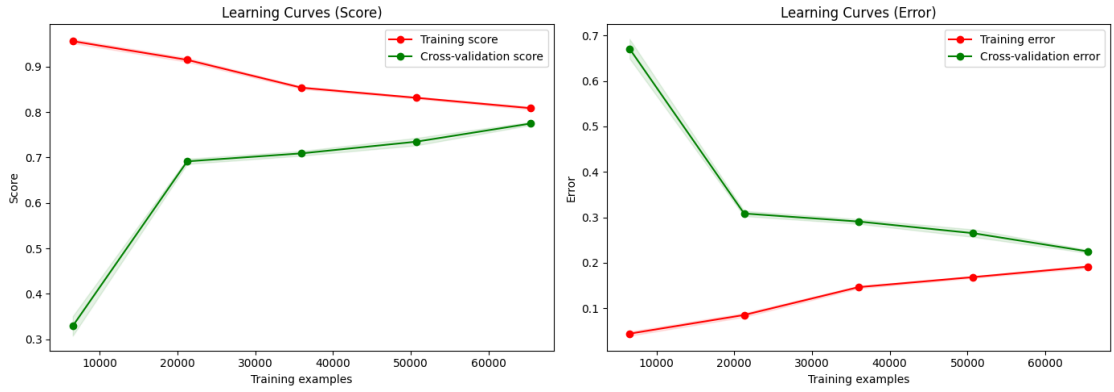
Criterion	Splitter	Max depth	Min samples split	Min samples leaf	Max features	Training Accuracy	Validation Accuracy	Cross Validation Mean Score
gini	best	None	2	1	None	0.9999	0.7664	0.7819
entropy	random	10	5	2	sqrt	0.7253	0.7156	0.7241
gini	best	20	10	5	log2	0.846	0.7308	0.7414
entropy	random	15	2	1	None	0.8015	0.7434	0.7534
gini	random	None	10	2	sqrt	0.7491	0.7046	0.717
entropy	best	5	5	1	log2	0.6631	0.6609	0.6619
gini	best	10	2	3	None	0.7865	0.7598	0.7621
entropy	best	12	6	2	None	0.8143	0.7687	0.7748
entropy	best	7	3	1	sqrt	0.6671	0.6637	0.6862
gini	best	13	6	3	log2	0.7223	0.716	0.7125
entropy	random	9	4	2	None	0.7255	0.719	0.7195
gini	random	11	7	1	sqrt	0.685	0.6771	0.6647
entropy	best	6	5	4	log2	0.6779	0.6749	0.6752
gini	best	16	8	2	None	0.8824	0.7742	0.7837
entropy	random	12	9	3	sqrt	0.6904	0.6869	0.6748
gini	random	15	6	4	log2	0.6832	0.677	0.6783

Karar Ağacı algoritmasının karışıklık matrisi Tablo 6.6'da sunulmuştur. Sınıf 0'da 13050 doğru sınıflandırma ve sadece 1.268 yanlış sınıflandırma ile model, yüksek doğruluk (0.80), geri çağırma (0.93) ve F1 skoru (0.86) ile iyi bir performans sergilemektedir. Sınıf 1 ise daha kötü performans göstermiştir; 2084 doğru sınıflandırmaya karşılık 2500'den fazla yanlış sınıflandırma yapılmış, bunlar arasında 2060 örnek Sınıf 0 olarak yanlış sınıflandırılmıştır. Bu durum, düşük geri çağırma (0.40), doğruluk (0.67) ve F1 skoru (0.50) ile sonuçlanmış ve Sınıf 1'in doğru tanımlanmasında zorluklar olduğunu göstermiştir. Sınıf 2 için ise 1744 doğru ve 1165 yanlış sınıflandırma durumu ile modelin performansı ortalamada kalmıştır. Sınıf 2 için doğruluk (0.65), geri çağırma (0.58) ve F1 skoru (0.61) değerleri elde edilmiştir. Bu, Sınıf 2'nin Sınıf 1'e göre daha güçlü tanındığını, ancak Sınıf 0 kadar güçlü tanınmadığını göstermektedir.

Tablo 6.6. Karar Ağacı Algoritması Sınıflandırma Raporu

ConfusionMatrix				Class	precision	recall	f1-score	support	
Actual	0	13050	811	457	0	0.8	0.93	0.86	14318
	1	2060	2084	440	1	0.67	0.4	0.5	4584
	2	849	316	1744	2	0.65	0.58	0.61	2909
	0	1	2	Macro Avg	0.71	0.63	0.66	21811	
	Predicted			WeightedAvg	0.75	0.77	0.75	21811	

Modelin öğrenme eğrileri Şekil 6.3'te gösterilmiştir ve bu, modelin çapraz doğrulama ve eğitim sırasında tutarlı performans sergilediğini de göstermektedir. Daha fazla örnek eklendikçe, modelin eğitim performansı sadece hafifçe azalarak yaklaşık %80'e düşmüş, bu da istikrarlı bir eğitim başarısını göstermiştir. Başlangıçta %30 gibi düşük bir seviyede olan doğrulama skoru hızla %75'in üzerine çıkmış, bu da güçlü bir genelleme yeteneği ve az aşırı öğrenme olduğunu göstermektedir. Bu sonuçlar, başlangıçta yüksek olan doğrulama hatasının daha fazla veri ile %20'ye düştüğünü ve eğitim hatasının düşük kalıp sadece hafif bir artış gösterdiğini gösteren "Hata" grafiği ile de desteklenmektedir. Modelin güçlü genelleme yeteneği, eğitim ve doğrulama hataları arasındaki küçük fark ile öne çıkmaktadır.



Şekil 6.3. Karar Ağacı Algoritması Öğrenme Eğrileri

6.4. Destek Vektör Makineleri (SVM) Algoritması Sonuçları

Tablo 6.7, SVM algoritmasının uygulanmasıyla elde edilen sonuçları göstermektedir. Bu sonuçlar arasından kernel: rbf, C değeri: 0.3 ve gamma: scale kombinasyonu seçilmiştir çünkü bu kombinasyon eğitim doğruluğu, doğrulama doğruluğu ve çapraz doğrulama doğruluğu açısından en yakın ve en yüksek değerlere sahiptir.

Tablo 6.7. SVM Algoritması Parametreler

Kernel	C value	Gamma	Training Accuracy	Validation Accuracy	Cross Validation Accuracy	Accuracy
Rbf	1	scale	0.7066	0.704	0.658	0.7
Rbf	0.3	scale	0.6894	0.6867	0.658	0.69
Rbf	0.5	Scale	0.6965	0.693	0.658	0.69
Rbf	1.5	Scale	0.7154	0.7116	0.658	0.71
Rbf	2	Scale	0.7189	0.7153	0.6581	0.72
poly	1	scale	0.6764	0.6742	0.658	0.67
Sigmoid	1	scale	0.5374	0.5403	0.532	0.54

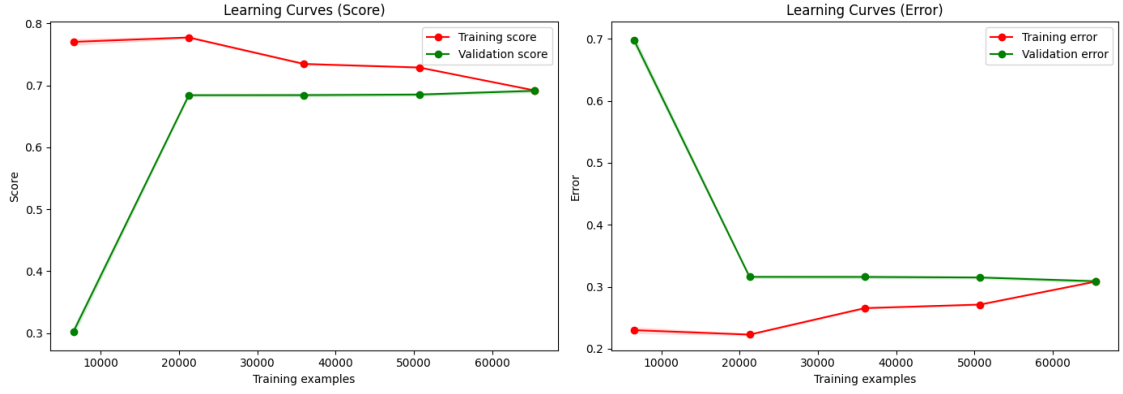
Modelin karışıklık matrisi Tablo 6.8'de gösterilmiştir ve Sınıf 0'ın sınıflandırılmasında önemli bir performans sergilediği görülmüştür. Sınıf 0, 14306 doğru sınıflandırma ve sadece 12 yanlış sınıflandırma ile F1 skoru 0.82, geri çağırma mükemmel şekilde 1.00'dir; ancak Sınıf 0 için doğruluk sadece 0.69'dur. Öte yandan,

model Sınıf 1 ve Sınıf 2 için iyi performans göstermediği gözlemlenmiştir. Sınıf 1 ise sadece 403 doğru sınıflandırmaya sahiptir, geri çağırma 0.09, F1 skoru 0.15 ve doğruluk 0.56'dır. Sınıf 2'de benzer sorunlar sunmaktadır; 271 doğru sınıflandırma, 0.73 doğruluk, 0.09 geri çağırma ve 0.17 F1 skoru ile sınıf ayırımında zorlukları göstermektedir. Genel olarak, makro ortalama doğruluk, geri çağırma ve F1 skorunun sırasıyla 0.66, 0.39 ve 0.38 olduğu görülmektedir. Ağırlıklı ortalama değerler genel doğruluğun 0.69 olduğunu, ancak Sınıf 1 ve Sınıf 2'deki zayıf performans nedeniyle F1 skorunun 0.59'a düştüğünü göstermektedir.

Tablo 6.8. SVM Algoritması Sınıflandırma Raporu

Confusion Matrix					Class	precision	recall	f1-score	support
Actual	0	14306	1	11	0	0.69	1	0.82	14318
	1	4093	403	88	1	0.56	0.09	0.15	4584
	2	2328	310	271	2	0.73	0.09	0.17	2909
		0	1	2	Macro Avg	0.66	0.39	0.38	21811
		Predicted			Weighted Avg	0.67	0.69	0.59	21811

Şekil 6.4'ün sol tarafındaki "Skor" grafiğine bakıldığında, modelin eğitim setinde daha iyi performans gösterdiği, ancak doğrulama setindeki performansının eğitim skorundan hala daha düşük olduğu açıktır. Doğrulama skoru başlangıçta hızla yükseldikten sonra %70 seviyesinde kalırken, eğitim skoru yaklaşık %70 seviyesinde sabit kalır. Model, eğitim verilerini iyi bir şekilde öğrenmektedir, ancak doğrulama verileri üzerinde o kadar iyi performans göstermemektedir, bu da eğitim ve doğrulama skorları arasındaki farkla gösterilmektedir. Bu, modelin doğrulama verilerinin ötesine genelleme yapma yeteneğinin sınırlı olabileceğini bizlere gösterir.



Şekil 6.4. SVM Algoritması Öğrenme Eğrileri

6.5. Rastgele Orman Algoritması Sonuçları

Tablo 6.9, Random Forest algoritmasının bulgularını göstermektedir. Karışıklık matrisi, performans metrikleri ve öğrenme eğrileri, $n_estimators$: 200, max_depth : 15, $min_samples_split$: 8, $min_samples_leaf$: 8 ve $max_features$: 0.5 parametre kombinasyonu için oluşturulmuştur.

Tablo 6.9. Rastgele Orman Algoritması Parametreler

# of Estimators	Max depth	Min samples split	Min samples leaf	Max features	Training Accuracy	Validation Accuracy	Cross-Validation Mean Score
300	20	5	2	sqrt	0.9416	0.8132	0.8195
200	30	10	4	log2	0.9121	0.8105	0.8177
400	10	2	2	0.5	0.7941	0.7738	0.7758
300	20	5	2	sqrt	0.9416	0.8132	0.8195
500	30	10	8	0.5	0.8893	0.8163	0.8255
350	15	3	3	sqrt	0.8449	0.7893	0.7925
250	15	3	3	sqrt	0.8446	0.7898	0.7925
350	20	4	2	0.5	0.9719	0.8346	0.843
300	10	6	3	sqrt	0.7646	0.7517	0.7514
350	20	4	2	sqrt	0.9471	0.8152	0.8209
400	5	2	2	0.5	0.7196	0.7189	0.71.79
500	25	8	8	sqrt	0.8554	0.7958	0.8033
250	20	8	4	sqrt	0.8973	0.8057	0.8113

400	25	8	3	0.3	0.9464	0.8211	0.8296
250	35	8	3	Log2	0.9411	0.8151	0.824
300	20	4	2	0.5	0.9725	0.8346	0.8427
250	15	3	3	Log2	0.8446	0.7898	0.7925
250	20	8	8	0.5	0.882	0.8142	0.8227
200	20	6	6	0.5	0.9017	0.8196	0.8285
500	20	10	8	sqrt	0.8469	0.7936	0.7993
400	20	10	8	sqrt	0.8462	0.7938	0.7994
350	15	8	8	Log2	0.8174	0.7809	0.7844
350	15	8	8	0.5	0.8515	0.8034	0.8093
200	15	8	8	0.5	0.8508	0.8042	0.8094
200	15	10	6	0.5	0.8619	0.8068	0.8123
300	15	5	5	sqrt	0.8316	0.7865	0.7889
500	30	10	8	sqrt	0.8653	0.7963	0.8038
450	18	6	4	0.4	0.9077	0.8188	0.8255
350	12	5	5	0.3	0.8017	0.7758	0.7785
400	15	4	3	Log2	0.8557	0.7931	0.7954
500	20	7	6	sqrt	0.8663	0.7982	0.8054
300	10	3	2	0.45	0.7903	0.7716	0.7733
300	18	6	4	0.4	0.9071	0.8184	0.8254
300	18	6	4	Log2	0.8767	0.8	0.8052
350	10	8	8	sqrt	0.7609	0.7505	0.7497

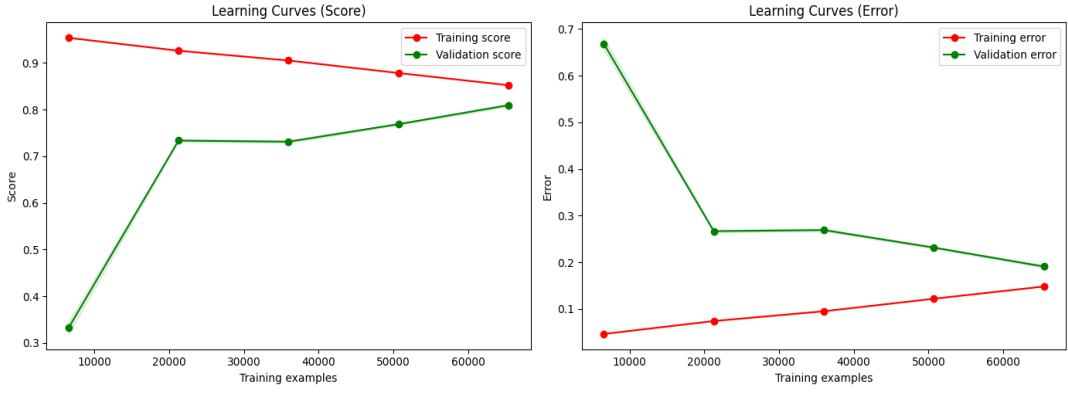
Tablo 6.10, Random Forest modelinin Sınıf 0'ı sınıflandırmada son derece iyi performans sergilediğini göstermektedir. 13896 doğru sınıflandırma ve sadece 422 yanlış sınıflandırma ile model, güçlü bir F1 skoru olan 0.88, yüksek geri çağırma (0.97) ve yüksek doğruluk (0.81) sağlamaktadır. Sınıf 1'in F1 skoru 0.53 olup, doğruluğu yüksek olmasına rağmen (0.86), geri çağırma değeri oldukça düşüktür (0.39), bu da yanlış sınıflandırmaların sıkça meydana geldiğini göstermektedir. Sınıf 2'de ise model, 0.68 F1 skoru, 0.65 geri çağırma ve 0.73 doğruluk ile dengeli bir tanıma yeteneği sergilemektedir. F1 skoru, geri çağırma ve makro-ortalama doğruluk sırasıyla 0.80, 0.67 ve 0.70 olarak hesaplanmıştır, bu da özellikle Sınıf 1 için iyileştirme gerektiren alanlar

olduğunu göstermektedir. Ağırlıklı ortalama doğruluk (0.81), geri çağırma (0.80) ve F1 skoru (0.78) genel kabul edilebilir bir performansı işaret etmektedir. Bu, modelin genel olarak yüksek doğruluk sağladığını, ancak Sınıf 1 için zayıf performans gösterdiğini göstermektedir.

Tablo 6.10. Rastgele Orman Algoritması Sınıflandırma Raporu

Confusion Matrix					Class	precision	recall	f1-score	support			
Actual	0	13896	165	257	0	0.81	0.97	0.88	14318			
	1	2374	1767	443	1	0.86	0.39	0.53	4584			
	2	906	124	271	2	0.73	0.65	0.68	2909			
					0	1	2	Macro Avg	0.8	0.67	0.7	21811
					Predicted			Weighted Avg	0.81	0.8	0.78	21811

Random Forest modeli, eğitim setinde son derece iyi performans göstermektedir, Şekil 6.5'te görüldüğü gibi. Daha fazla veri eklendikçe model, yaklaşık %85 seviyesinde stabilize olarak eğitim verilerine etkili şekilde uyum sağladığını göstermektedir. Model, yaklaşık %95'lik bir eğitim skoru ile başlamaktadır. Başlangıçta yaklaşık %30 olan doğrulama skoru ise hızla %80'e çıkarak güçlü bir genelleme yeteneği sergilemektedir. Şekil 6.5'teki "Öğrenme Eğrileri (Hata)" grafiği, doğrulama hatasının %60'lık bir zirveden %20–25 aralığına düştüğünü, eğitim hatasının ise %10 seviyesinde düşük kaldığını göstermektedir. Bu grafikler, aşırı öğrenmenin çok az olduğunu, modelin genel olarak iyi performans gösterdiğini, eğitim verilerini iyi bir şekilde öğrendiğini ve doğrulama verilerine de başarılı bir şekilde genelleme yaptığını göstermektedir.



Şekil 6.5. Rastgele Orman Algoritması Öğrenme Eğrileri



7. TARTIŞMA

Tablo 7.1’de, beş makine öğrenimi algoritması XGBoost, K-En Yakın Komşular (KNN), Karar Ağacı, Destek Vektör Makinesi (SVM) ve Rastgele Orman Recall, Training Accuracy, Validation Accuracy ve F1 skoru gibi çeşitli performans metriklerine dayanarak karşılaştırılmıştır. Veriler üzerindeki performans açısından, Rastgele Orman en yüksek eğitim doğruluğunu (0.8508) ve doğrulama doğruluğunu (0.8042) sergilerken, yüksek doğruluk (0.81), geri çağırma (0.80) ve F1 skoru (0.78) ile öne çıkmaktadır. Karar Ağacı Algoritması, Rastgele Orman ile karşılaştırıldığında doğruluk (0.75), geri çağırma (0.77) ve doğrulama doğruluğu (0.8143) açısından daha düşük performans değerleri göstermektedir. Rastgele Orman ile karşılaştırıldığında, KNN algoritmasının tüm metriklerde biraz daha düşük sonuçlar gözlemlenmiştir. XGBoost, iyi performans gösteren modellere göre daha düşük bir eğitim doğruluğu sergilemiştir, ancak yine de dengeli doğruluk, geri çağırma ve doğrulama doğruluğu (0.7836) göstermiştir. Diğer modellerle karşılaştırıldığında, SVM modeli en düşük eğitim (0.6894) ve doğrulama doğruluğunu (0.6867) sergiler ve aynı zamanda en düşük doğruluk (0.67), geri çağırma (0.69) ve F1 skoru (0.59) ile genelleme yapmakta zorlandığını gösterir. Bu karşılaştırma, Rastgele Orman’ın KNN, Karar Ağacı ve XGBoost’a göre eğitim performansı açısından daha iyi performans gösterdiğini, SVM’nin ise daha iyi sonuçlar elde etmek için düzenlenmeye veya daha fazla veriye ihtiyaç duyabileceğini göstermektedir. Ek olarak, tüm modellerin sınıflandırma raporu sınıfsal bazlı incelendiğinde, sınıf 1 (Slightly drowsy) sınıfının precision, recall ve f1-score değerleri, diğer sınıflarla karşılaştırıldığında düşük skorlara sahip olduğu gözlemlenmiştir.

Tablo 7.1. Tüm Algoritmaların Karşılaştırılması

Algorithm	Training Accuracy	Validation Accuracy	Precision	Recall	F1-Score
XGBoost	0.8077	0.7836	0.78	0.78	0.75
KNN	0.8049	0.7544	0.74	0.76	0.74
Decisiontree	0.8143	0.7687	0.75	0.77	0.75
SVM	0.6894	0.6867	0.67	0.69	0.59
RandomForest	0.8508	0.8042	0.81	0.8	0.78

Tablo 3.1’de görüldüğü gibi, Web of Science (WoS) veri tabanındaki EKG ile ilgili çalışmalar ayrıntılı olarak incelenmiştir. Literatürdeki çalışmaların çoğunda, EKG genellikle araç ölçüm verileri ile hibrit bir şekilde kullanılmamış, daha çok davranışsal ölçümlerle entegre edilmiştir. Bu çalışmalarda genellikle, yüz tanıma ve EKG verilerinin birleştirilip yorgunluk tespiti için kullanıldığı görülmektedir. EKG ve yüz tanıma yöntemlerinin bir arada kullanılması, yorgunluk tespitinde dikkat çekici derecede yüksek başarı oranlarına ulaşmıştır. Örneğin, [66] Sukumar ve ark. (2024) ise %97,30 doğruluk oranına ulaşmıştır. Ancak yüz tanıma yönteminin bazı dezavantajları olduğu bilinmektedir; örneğin, yüz yapısındaki ırksal özellikler, gözlük veya maske kullanımı gibi faktörler yüz tanıma teknolojisinin her zaman olumlu sonuçlar vermesini engelleyebilir. Ayrıca, literatürdeki bu çalışmalar genellikle yorgunluğu "yorgun" veya "yorgun değil" olarak sınıflandırmakta, bu çalışmada olduğu gibi farklı sınıflara ayırmamaktadır. Bu nedenle, bu çalışmaların yüksek doğruluk değerlerine ulaşması olasıdır.

Ek olarak, literatürdeki birçok çalışma, aslında son teknoloji sürücü yorgunluk izleme sistemlerinin geliştirilmesine odaklanmıştır. Örneğin, [67] Linschmann ve ark. (2023), EKG ölçümü için taşınabilir bir minder tasarlamıştır. [68] Wolkow ve ark. (2020), kalp atış hızını ölçmek için bir bileklik kullanmayı denemiştir. [69] Muramatsu ve Sasaki (2021), direksiyon simidine yerleştirilen iletken elektrotlar ve koltuğa yerleştirilen kapasitif elektrotlarla EKG ölçümü yapmıştır. Bu çalışmaların çoğu, yorgunluk performansını ölçmekten çok, EKG ölçüm teknolojilerini araç içerisinde kullanımını geliştirmeye yönelik araştırmalar olup, EKG ölçümlerinin doğruluğunu artırmayı amaçlamıştır.

Çalışmamızda, sadece EKG'ye dayalı bir sistemin aksine hem EKG hem de CAN-Bus verilerini inceleyerek entegre bir ölçüm sınıflandırması yapılmıştır. Bu şekilde, sadece CAN-Bus verilerine dayalı olarak EKG verisine ihtiyaç duyulmadan yorgunluk tespiti yapılmaya çalışılmış ve sınıflandırma performansları analiz edilmiştir. Bazı çalışmalarda, bu çalışmada olduğu gibi, yorgunluk farklı sınıflara ayrılarak ölçülmüştür. Örneğin, [70] Ebrahimian ve ark. (2021), EKG ve solunum verilerini kullanarak üç sınıf yorgunluk tespitinde %91, beş sınıf tespitinde ise %67 performans elde etmiştir. Ancak bu çalışmada hem EKG hem de CAN-Bus verileri kullanılarak daha fazla parametre ile üç sınıf yorgunluk tespitinde %85 performans elde edilmiştir. Ayrıca, bu sistem

endüstriyel kullanım için daha uygun bir yapıya sahiptir. Çünkü tasarlanan bu sistemin temel amacı, yorgun sürücüyü EKG verisiyle eğittikten sonra araç CAN sinyallerini kullanarak harici bir EKG cihazına gerek kalmadan sürücü yorgunluğunu, gerçek dünyada ölçülebilir kılmaktır. Murugan ve ark. (2020) ise sadece EKG verilerini kullanarak beş sınıf yorgunluk tespitinde %58,3 performans elde etmiştir. Bu bulgular, entegre bir strateji ve ek özellikler kullanılarak, çalışmamızın daha kapsamlı sonuçlar verdiğini göstermektedir.



8. SONUÇ

Mevcut yaklaşımların ya fizyolojik ölçümleri ya da araç verilerini kullandığı durumların aksine, bu çalışma hem EKG verilerini hem de CAN-Bus sinyallerini birleştirerek sürücü yorgunluğunu tespit etmek için daha kapsamlı ve hassas bir yöntem sunmaktadır. Makine öğrenimi yöntemlerini, özellikle sınıflandırma algoritmalarını kullanarak, çalışma sürücü yorgunluğunu üç farklı kategoriye ayırarak sınıflandırmayı başarmıştır. Sistemin %85 doğruluk oranı, yalnızca EKG verilerine dayanan önceki araştırmalara göre önemli bir ilerlemeyi temsil etmektedir; bu tür araştırmalarda, özellikle çok sınıflı yorgunluk tespitinde doğruluk oranları genellikle %70'in altında kalmaktadır. Bu, hibrit tekniğin yorgunluk tespiti performansını ne kadar iyi artırdığını ve gerçek sürüş koşullarında daha kullanışlı hale getirdiğini göstermektedir.

Yöntem, birkaç sürücüden gerçek zamanlı olarak çeşitli trafik ve çevresel koşullar altında fizyolojik ve araçla ilgili bileşenlerin toplanmasını içermektedir. Yorgunluğun önemli bir göstergesi olan kalp atış hızı değişkenliği (HRV), EKG ölçümleriyle elde edilirken, hız, direksiyon açısı ve sürücü fren tepkimeleri gibi kritik araç metrikleri aracın CAN-Bus hattından elde edilmiştir. Çalışma, fizyolojik ve araç bazlı faktörlerin birleştirilmesinin, bu veri kaynaklarını makine öğrenimi modellerine entegre ederek, hızlı ve güvenilir bir yorgunluk tespiti sağlayabilen daha sağlam bir sistem ile sonuçlandığını ortaya koymuştur. Bu yaklaşım, tespit doğruluğunu artırarak ve araç güvenlik sistemlerinde sadece CAN-Bus hattının okunmasıyla, akıllı sürücü izleme teknolojisinin gelişimine önemli bir katkı sağlamaktadır.

Bu çalışmada toplanan veriler, katılımcılardan farklı zamanlarda ve trafik yoğunluğu nedeniyle eşit koşullar altında toplanmamıştır. Literatürde 10 saatlik veri toplama süresi yeterli görünse de çalışmanın performansını artırmak ve genellemek amacıyla bu sürenin farklı katılımcılar ile uzatılması gerektiği düşünülmektedir. Özellikle, algoritmaların sınıf 1 (Slightly drowsy) sınıfının nispeten düşük precision, recall ve f-1 score değerleri bunun en önemli göstergesidir. Ayrıca, bu çalışmada kullanılan geleneksel makine öğrenimi uygulamalarından elde edilen sonuçlar literatürle karşılaştırıldığında iyi sonuçlar olsa da daha iyi sonuçlar elde etmek mümkündür. Bu, sezgisel/meta-sezgisel hibrit yöntemler denenerek ve daha fazla verinin toplanmasıyla başarılabılır. Gelecekteki çalışmalarda daha fazla veri toplanması ve bu verilerin hibrit yöntemlerle yeniden değerlendirilmesi planlanmaktadır.

KAYNAKÇA

- [1] Türkiye İstatistik Kurumu (TÜİK), “TSI 2022 – Road Traffic Accident Statistics”, 2022. Erişim:28 Eylül 2024. [Çevrimiçi]. Erişim adresi: <https://www.tuik.gov.tr>
- [2] Moradi, A., Nazari, S. S. H., & Rahmani, K. (2019). Sleepiness and the risk of road traffic accidents: A systematic review and meta-analysis of previous studies. *Transportation research part F: traffic psychology and behaviour*, 65, 620-629.
- [3] Sahayadhas, A., Sundaraj, K., & Murugappan, M. (2012). Detecting driver drowsiness based on sensors: a review. *Sensors*, 12(12), 16937-16953.
- [4] Zhang, G., Yau, K. K., Zhang, X., & Li, Y. (2016). Traffic accidents involving fatigue driving and their extent of casualties. *Accident Analysis & Prevention*, 87, 34-42.
- [5] Le Bon, O. (2020). Relationships between REM and NREM in the NREM-REM sleep cycle: a review on competing concepts. *Sleepmedicine*, 70, 6-16.
- [6] National Highway Traffic Safety Administration (NHTSA), “Drowsy Driving and Automobile Crashes”, National Center on Sleep Disorder Research and the National Highway Traffic Safety Administration, Howe, TX, USA, 1998.
- [7] Web of Science (WOS), “Web of Science ‘driver drowsiness’ Search Results”. [Çevrimiçi]. Erişim adresi: www.webofscience.com
- [8] Ramzan, M., Khan, H. U., Awan, S. M., Ismail, A., Ilyas, M., & Mahmood, A. (2019). A survey on state-of-the-art drowsiness detection techniques. *IEEE Access*, 7, 61904-61919.
- [9] Kumar, M. U. S. K. A. N. (2020). Driver drowsiness detection techniques: a review. *Advances in Mathematics: Scientific Journal*, 9(6), 3933-3938.
- [10] Yılmaz, E. K., & Akcayol, M. A. (2022). SUST-DDD: A real-drive data set for driver drowsiness detection. In *Conference of Open Innovations Association (FRUCT) (Vol. 6)*.
- [11] Ngxande, M., Tapamo, J. R., & Burke, M. (2017). Driver drowsiness detection using behavioral measures and machine learning techniques: A review of state-of-art

techniques. 2017 pattern recognition Association of South Africa and Robotics and mechatronics (PRASA-RobMech), 156-161.

[12] Barua, S., Ahmed, M. U., Ahlström, C., & Begum, S. (2019). Automatic driver sleepiness detection using EEG, EOG and contextual information. *Expert systems with applications*, 115, 121-135.

[13] Doudou, M., Bouabdallah, A., & Berge-Cherfaoui, V. (2020). Driver drowsiness measurement technologies: Current research, market solutions, and challenges. *International Journal of Intelligent Transportation Systems Research*, 18, 297-319.

[14] Xiao, Y., & bin Abas, A. (2021). A review on fatigue driving detection. *ASP Transactions on Internet of Things*, 1(3), 1-14.

[15] Park, I., Ahn, J. H., & Byun, H. (2006, August). Efficient measurement of eyeblinking under various illumination conditions fordrowsiness detection systems. In *18th International Conference on Pattern Recognition (ICPR'06)* (Vol. 1, pp. 383-386). IEEE.

[16] Islam, A., Rahaman, N., & Ahad, M. A. R. (2019). A study on tiredness assessment by using eye blink detection. *Jurnal Kejuruteraan*, 31(2), 209-214.

[17] Kamarudin, N., Jumadi, N. A., Mun, N. L., Keat, N. C., Ching, A. H. K., Mahmud, W. M. H. W., ... & Mahmud, F. (2019). Implementation of haarcascade classifier and eye aspect ratio for driver drowsiness detection using raspberry Pi. *Universal Journal of Electricaland Electronic Engineering*, 6(5), 67-75.

[18] Cheamanunkul, S., & Chawla, S. (2020, December). Drowsiness detection using facial emotions and eye aspect ratios. In *2020 24th International Computer Science and Engineering Conference (ICSEC)* (pp. 1-4). IEEE.

[19] Maior, C. B. S., dasChagasMoura, M. J., Santana, J. M. M., & Lins, I. D. (2020). Real-time classification for autonomous drowsiness detection using eye aspect ratio. *Expert Systems with Applications*, 158, 113505.

[20] Bhandari, G. M., Durge, A., Bidwai, A., & Aware, U. (2014). Yawning analysis for driver drowsiness detection. *Int. J. Res. Eng. Technol.*, 3(2), 502-505.

- [21] Ghourabi, A., Ghazouani, H., & Barhoumi, W. (2020, September). Driver drowsiness detection based on joint monitoring of yawning, blinking and nodding. In 2020 IEEE 16th International Conference on Intelligent Computer Communication and Processing (ICCP) (pp. 407-414). IEEE.
- [22] Abtahi, S., Hariri, B., & Shirmohammadi, S. (2011, May). Driver drowsiness monitoring based on yawning detection. In 2011 IEEE International Instrumentation and Measurement Technology Conference (pp. 1-4). IEEE.
- [23] Liu, X., Mei, H., Lu, H., Kuang, H., & Ma, X. (2017). A vehicle steering recognition system based on low-cost smartphone sensors. *Sensors*, 17(3), 633.
- [24] Lu, J., Zheng, X., Tang, L., Zhang, T., Sheng, Q. Z., Wang, C., ... & Zhou, W. (2021). Can steering Wheel detect your driving fatigue?. *IEEE Transactions on Vehicular Technology*, 70(6), 5537-5550.
- [25] Kamble, A., & Potadar, S. (2018, June). Lane departure warning system for advanced drivers assistance. In 2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS) (pp. 1775-1778). IEEE.
- [26] Haloi, M., & Jayagopi, D. B. (2015, June). A robust lane detection and departure warning system. In 2015 IEEE Intelligent Vehicles Symposium (IV) (pp. 126-131). IEEE.
- [27] Saito, Y., Itoh, M., & Inagaki, T. (2016). Driver assistance system with a dual controlscheme: Effectiveness of identifying driver drowsiness and preventing lane departure accidents. *IEEE Transactions on Human-Machine Systems*, 46(5), 660-671.
- [28] Fujiwara, K., Abe, E., Kamata, K., Nakayama, C., Suzuki, Y., Yamakawa, T., ... & Kadotani, H. (2018). Heart rate variability-based driver drowsiness detection and its validation with EEG. *IEEE transactions on biomedical engineering*, 66(6), 1769-1778.
- [29] Shaffer, F., & Ginsberg, J. P. (2017). An overview of heart rate variability metrics and norms. *Frontiers in public health*, 5, 258.
- [30] Vicente, J., Laguna, P., Bartra, A., & Bailón, R. (2016). Drowsiness detection using heart rate variability. *Medical & biological engineering & computing*, 54, 927-937.

- [31] Freitas, A., Almeida, R., Gonçalves, H., Conceição, G., & Freitas, A. (2024). Monitoring fatigue and drowsiness in motor vehicle occupants using electrocardiogram and heart rate – a systematic review. *Transportation research part F: traffic psychology and behaviour*, 103, 586-607.
- [32] Cui, J., Lan, Z., Sourina, O., & Müller-Wittig, W. (2022). EEG-based cross-subject driver drowsiness recognition with an interpretable convolutional neural network. *IEEE Transactions on Neural Networks and Learning Systems*, 34(10), 7921-7933.
- [33] Gurudath, N., & Riley, H. B. (2014). Drowsy driving detection by EEG analysis using wavelet transform and K-means clustering. *Procedia Computer Science*, 34, 400-409.
- [34] Butkevičiūtė, E., Michalkovič, A., & Bikulčienė, L. (2022). Ecg signal features classification for the mental fatigue recognition. *Mathematics*, 10(18), 3395.
- [35] Huang, S., Li, J., Zhang, P., & Zhang, W. (2018). Detection of mental fatigue state with wearable ECG devices. *International journal of medical informatics*, 119, 39-46.
- [36] Dong, W., Huang, Y., Lehane, B., & Ma, G. (2020). XGBoost algorithm-based prediction of concrete electrical resistivity for structural health monitoring. *Automation in Construction*, 114, 103155.
- [37] Pan, B. (2018, February). Application of XGBoost algorithm in hourly PM2.5 concentration prediction. In *IOP conference series: earth and environmental science* (Vol. 113, p. 012127). IOP publishing.
- [38] Chen, J., Zhao, F., Sun, Y., & Yin, Y. (2020). Improved XGBoost model based on genetic algorithm. *International Journal of Computer Applications in Technology*, 62(3), 240-245.
- [39] Amjad, M., Ahmad, I., Ahmad, M., Wróblewski, P., Kamiński, P., & Amjad, U. (2022). Prediction of pile bearing capacity using XGBoost algorithm: model in gnd performance evaluation. *Applied Sciences*, 12(4), 2126.

- [40] Jiang, Y., Tong, G., Yin, H., & Xiong, N. (2019). A pedestrian detection method based on genetic algorithm for optimize XGBoost training parameters. *IEEE Access*, 7, 118310-118321.
- [41] Mitchell, R., & Frank, E. (2017). Accelerating the XGBoost algorithm using GPU computing. *Peer J Computer Science*, 3, e127.
- [42] Moldagulova, A., & Sulaiman, R. B. (2017, May). Using KNN algorithm for classification of textual documents. In 2017 8th international conference on information technology (ICIT) (pp. 665-671). IEEE.
- [43] Yu, L., & Gui-Sheng, C. (2010, March). KNN algorithm improving based on cloud model. In 2010 2nd International Conference on Advanced Computer Control (Vol. 2, pp. 63-66). IEEE.
- [44] Pulungan, A. F., Zarlis, M., & Suwilo, S. (2019). Analysis of braycurtis, Canberra and euclidean distance in knn algorithm. *Sinkron: jurnal dan penelitian teknik informatika*, 4(1), 74-77.
- [45] Malkauthekar, M. D. (2013, October). Analysis of euclidean distance and manhattan distance measure in face recognition. In Third International Conference on Computational Intelligence and Information Technology (CIIT 2013) (pp. 503-507). IET.
- [46] Xiao, Q., Zhong, X., & Zhong, C. (2020). Application research of KNN algorithm based on clustering in big data talent demand information classification. *International journal of pattern recognition and artificial intelligence*, 34(06), 2050015.
- [47] Ehsani, R., & Drabløs, F. (2020). Robust distance measures for k NN classification of cancer data. *Cancer informatics*, 19, 1176935120965542.
- [48] Ooi, H. L., Ng, S. C., & Lim, E. (2013). Ano detectionwith k-nearest neighborusing minkowski distance. *International Journal of Signal Processing Systems*, 1(2), 208-211.
- [49] Patel, H. H., & Prajapati, P. (2018). Study and analysis of decision tree based classification algorithms. *International Journal of Computer Sciences and Engineering*, 6(10), 74-78.

- [50] Charbuty, B., & Abdulazeez, A. (2021). Classification based on decision tree algorithm for machine learning. *Journal of Applied Science and Technology Trends*, 2(01), 20-28.
- [51] Gupta, B., Rawat, A., Jain, A., Arora, A., & Dhami, N. (2017). Analysis of various decision tree algorithms for classification in data mining. *International Journal of Computer Applications*, 163(8), 15-19.
- [52] Navada, A., Ansari, A. N., Patil, S., & Sonkamble, B. A. (2011, June). Overview of use of decision tree algorithms in machine learning. In *2011 IEEE control and system graduate research colloquium* (pp. 37-42). IEEE.
- [53] Priyam, A., Abhijeeta, G. R., Rathee, A., & Srivastava, S. (2013). Comparative analysis of decision tree classification algorithms. *International Journal of current engineering and technology*, 3(2), 334-337.
- [54] Speiser, J. L., Miller, M. E., Tooze, J., & Ip, E. (2019). A comparison of random forest variable selection methods for classification prediction modeling. *Expert systems with applications*, 134, 93-101.
- [55] Biau, G. (2012). Analysis of a random forests model. *The Journal of Machine Learning Research*, 13(1), 1063-1095.
- [56] Dubath, P., Rimoldini, L., Süveges, M., Blomme, J., López, M., Sarro, L. M., ... & Eyer, L. (2011). Random forest automated supervised classification of Hipparcos periodic variable stars. *Monthly Notices of the Royal Astronomical Society*, 414(3), 2602-2617.
- [57] Belgiu, M., & Drăguț, L. (2016). Random forest in remote sensing: A review of applications and future directions. *ISPRS journal of photogrammetry and remote sensing*, 114, 24-31.
- [58] Deka, P. C. (2014). Support vector machine applications in the field of hydrology: a review. *Applied soft computing*, 19, 372-386.
- [59] Moguerza, J. M., & Muñoz, A. (2006). Support vector machines with applications.

- [60] Kavzoglu, T., & Colkesen, I. (2009). A kernel functions analysis for support vector machines for land cover classification. *International Journal of Applied Earth Observation and Geo information*, 11(5), 352-359.
- [61] Rochim, A. F., Widyaningrum, K., & Eridani, D. (2021, December). Performance Comparison of Support Vector Machine Kernel Functions in Classifying COVID-19 Sentiment. In *2021 4th International Seminar on Research of Information Technology and Intelligent Systems (ISRITI)* (pp. 224-228). IEEE.
- [62] Sufi, F., Fang, Q., & Cosic, I. (2007, August). ECG RR peak detection on mobile phones. In *2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (pp. 3697-3700). IEEE.
- [63] Chen, S., Xu, K., Zheng, X., Li, J., Fan, B., Yao, X., & Li, Z. (2020). Linear and nonlinear analyses of normal and fatigue heart rate variability signals for miners in high-altitude and cold areas. *Computer Methods and Programs in Biomedicine*, 196, 105667.
- [64] Berntson, G. G., Thomas Bigger Jr, J., Eckberg, D. L., Grossman, P., Kaufmann, P. G., Malik, M., ... & Van Der Molen, M. W. (1997). Heart rate variability: origins, methods, and interpretive caveats. *Psychophysiology*, 34(6), 623-648.
- [65] Freitas, A., Almeida, R., Gonçalves, H., Conceição, G., & Freitas, A. (2024). Monitoring fatigue and drowsiness in motor vehicle occupants using electrocardiogram and heart rate—a systematic review. *Transportation research part F: traffic psychology and behaviour*, 103, 586-607.
- [66] Sukumar, N., Sivashankar, B., Sumathi, P., & Maurya, O. P. (2024). Physiological and Physical Sensors for Stress Level, Drowsiness Detection, and Behaviour Analysis. *IEEE Transactions on Consumer Electronics*.
- [67] Linschmann, O., Uguz, D. U., Romanski, B., Baarlink, I., Gunaratne, P., Leonhardt, S., & Lueken, M. (2023). A portable multi-modal cushion for continuous monitoring of a driver's vital signs. *Sensors*, 23(8), 4002.
- [68] Wolkow, A. P., Rajaratnam, S. M., Wilkinson, V., Shee, D., Baker, A., Lillington, T., & Howard, M. E. (2020). The impact of heart rate-based drowsiness

monitoring on adverse driving events in heavy vehicle drivers under naturalistic conditions. *Sleephealth*, 6(3), 366-373.

[69] Muramatsu, D., & Sasaki, K. (2021). Noise Reduction Using a Triple-layer Electrode in Conductive/ Capacitive Hybrid Electrocardiogram Measurement. *Sensors and Materials*, 33(12), 4105-4111.

[70] Ebrahimian, S., Nahvi, A., Tashakori, M., Salmanzadeh, H., Mohseni, O., & Leppänen, T. (2022). Multi-level classification of driver drowsiness by simultaneous analysis of ECG and respiration signals using deep neural networks. *International journal of environmental research and public health*, 19(17), 10736.



ÖZGEÇMİŞ

Kişisel Bilgiler

Adı Soyadı	Kaan KAHRAMAN
E-Posta	

Eğitim

Lisans	2022, İzmir Bakırçay Üniversitesi, Mühendislik ve Mimarlık Fakültesi, Elektrik-Elektronik Mühendisliği
--------	--

Yayın Listesi

Bildiri	Kahraman, K, (2025), 2nd International Conference on Modern and Advanced Research, Konya/Turkey
---------	---

