



**ZİHİNSEL İŞ YÜKÜNÜN EEG VE GÖZ İZLEME VERİLERİ
KULLANILARAK MAKİNE ÖĞRENMESİ TEKNİKLERİYLE
MODELLENMESİ**

Şeniz HARPUTLU AKSU

**DOKTORA TEZİ
ENDÜSTRİ MÜHENDİSLİĞİ ANA BİLİM DALI**

**GAZİ ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

EKİM 2023

ETİK BEYAN

Gazi Üniversitesi Fen Bilimleri Enstitüsü Tez Yazım Kurallarına uygun olarak hazırladığım bu tez çalışmada;

- Tez içinde sunduğum verileri, bilgileri ve dokümanları akademik ve etik kurallar çerçevesinde elde ettiğimi,
- Tüm bilgi, belge, değerlendirme ve sonuçları bilimsel etik ve ahlak kurallarına uygun olarak sunduğumu,
- Tez çalışmada yararlandığım eserlerin tümüne uygun atıfta bulunarak kaynak gösterdiğimi,
- Kullanılan verilerde herhangi bir değişiklik yapmadığımı,
- Bu tezde sunduğum çalışmanın özgün olduğunu,

bildirir, aksi bir durumda aleyhime doğabilecek tüm hak kayıplarını kabullendiğimi beyan ederim.

Şeniz HARPUTLU AKSU

04/10/2023

ZİHİNSEL İŞ YÜKÜNÜN EEG VE GÖZ İZLEME VERİLERİ KULLANILARAK MAKİNE ÖĞRENMESİ TEKNİKLERİYLE MODELLENMESİ

(Doktora Tezi)

Şeniz HARPUTLU AKSU

GAZİ ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

Ekim 2023

ÖZET

Bu tez çalışması kapsamında, zihinsel iş yükünün elektroensefalografi (EEG) ve göz izleme verilerine bağlı olarak analiz edilmesi ve makine öğrenmesi algoritmaları kullanılarak tahmin edilmesi amaçlanmıştır. Öncelikle 4 katılımcı (2 kadın, 2 erkek) üzerinde pilot uygulama gerçekleştirilmiş, EEG ve göz izleme cihazları ile ayrı ayrı ölçümler alınarak sonuçlar analiz edilmiştir. Temel uygulamada ise deneyler 15 üniversite öğrencisi (7 kadın, 8 erkek) üzerinde uygulanmış olup, EEG ve göz izleme cihazlarından eş zamanlı kayıtlar alınmıştır. Deneyler sırasında, farklı seviyelerde zihinsel iş yükünün tahmini için n-geri hafıza görevi ve NASA-Task Load Index (TLX) öznel değerlendirme ölçeği kullanılmıştır. Görev zorluk seviyesi arttıkça sağ ve sol göz bebeği boyutunun, odaklanma sayısının, seğirme sayısı ve süresinin ve ayrıca, göz kırpması sayısı ve süresinin arttığı; odaklanma süresi ile ilgili değişkenlerin ise azaldığı görülmüştür. EEG ile ilgili olarak elde edilen sonuçlar; prefrontal, frontal ve frontal merkez bölgelerindeki yani ön beyindeki teta gücünün görev zorluğu arttıkça arttığını göstermektedir. Görev zorlaştıkça frontal bölgelerdeki alfa gücü artmakta iken temporal, parietal ve oksipital bölgelerdeki alfa gücü ise azalmaktadır. Düşük beta gücünün ise görev zorlaştıkça hemen hemen tüm beyin bölgelerinde önemli ölçüde azaldığı gözlemlenmiştir. Dört sınıflı sınıflandırma problemi kapsamında, seçilen 18 özellik ile zihinsel iş yükü seviyesi %70,26 doğruluk oranı ile tahmin edilebilmiştir. Sonuçlar, göz izleme ve EEG ölçümlerinin zihinsel iş yükünün analizinde önemli bir rol oynadığını göstermektedir. Bu kapsamda, göz izleme ve EEG ölçümlerinin eş zamanlı olarak kayıt altına alındığı deney süreci sonucunda farklı makine öğrenmesi algoritmalarının karşılaştırmalı olarak kullanıldığı dört sınıflı zihinsel iş yükü tahmin modeli ile literatüre önemli katkı sağlanmıştır.

Bilim Kodu : 90603

Anahtar Kelimeler : Elektroensefalografi, göz izleme, makine öğrenmesi, nöroergonomi, zihinsel iş yükü

Sayfa Adedi : 202

Danışman : Prof. Dr. Erman ÇAKIT

MENTAL WORKLOAD MODELING BY USING MACHINE LEARNING
TECHNIQUES BASED ON EEG AND EYE TRACKING DATA

(Ph. D. Thesis)

Şeniz HARPUTLU AKSU

GAZİ UNIVERSITY

GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

October 2023

ABSTRACT

In this thesis, it was aimed to analyze the mental workload based on EEG and eye tracking data and to classify it through machine learning algorithms. First of all, a pilot application was carried out on 4 participants (2 women, 2 men), and after taking separate measurements with EEG and eye tracking devices the results were analyzed. In the basic application, the experiments were carried out on 15 university students (7 women, 8 men), and simultaneous recordings were taken from EEG and eye tracking devices. During the experiments, the n-back memory task and the NASA-Task Load Index (TLX) subjective rating scale were used to estimate different levels of mental workload. It was observed that as the task difficulty increased, the right and left pupil diameters, the number of fixations, the number and duration of saccades, as well as the number and duration of blinks increased, while the variables related to fixation duration decreased. The results obtained in relation to the EEG shows that theta power in the prefrontal, frontal and frontal center regions increases as the task difficulty increases. As the task becomes more difficult, alpha power increases in the frontal regions and decreases in the temporal, parietal and occipital regions. Low beta power was also observed to decrease significantly in almost all brain regions as the task became more difficult. Within the scope of the four-class classification problem, the mental workload level could be predicted with an accuracy rate of 70,26% with 18 selected features. The results show that eye tracking and EEG measurements play an important role in the classification of mental workload. In this context, a significant contribution has been made to the literature with the four-class mental workload estimation model, in which different machine learning algorithms are used comparatively, as a result of the experimental process in which eye tracking and EEG measurements are recorded simultaneously.

Science Code : 90603

Key Words : Electroencephalography, eye tracking, machine learning,
neuroergonomics, mental workload

Page Number : 202

Supervisor : Prof. Dr. Erman ÇAKIT

TEŞEKKÜR

Tez çalışmam boyunca, hem değerli görüşleri ile bu çalışmaya katkı sağlayan hem de akademik gelişimimi destekleyen yönlendirmeleriyle yolumu aydınlatan çok değerli tez danışmanım Sayın Prof. Dr. Erman ÇAKIT'a hiçbir zaman benden esirgemediği cesaret veren ve anlayışlı yaklaşımı sebebiyle teşekkürlerimi ve şükranlarımı sunarım. Tez sürecinin başında bir süre tez danışmanlığımı yürüten Sayın Prof. Dr. Metin DAĞDEVİREN'e sağladığı değerli katkılar ve kıymetli yol göstericiliği sebebiyle ayrıca teşekkür ederim.

Değerli fikirlerini paylaşarak tez çalışmam süresince beni yönlendiren ve teşvik eden çok kıymetli Tez İzleme Komitesi hocalarım, Sayın Prof. Dr. Ergün ERASLAN'a ve Sayın Prof. Dr. Feyzan ARIKAN ÖKTEMER'e çok teşekkür ederim. Değerli tez jürisi üyeleri Sayın Doç. Dr. Aylin ADEM ve Sayın Dr. Öğr. Üyesi İbrahim YILMAZ'a kıymetli değerlendirmeleri ve katkıları için ayrıca teşekkür ederim.

Tez danışmanım ile birlikte sunmuş olduğumuz 6483 numaralı “Zihinsel İş Yükünün Analizine İlişkin Bir Model Önerisi” başlıklı lisansüstü tez projesini finansal olarak destekleyen Gazi Üniversitesi Bilimsel Araştırma Projeleri (BAP) Koordinasyon Birimine, göz izleme cihazını bünyesinde barındıran İnsan-Bilgisayar Etkileşimi Araştırma ve Uygulama Laboratuvarından faydalanmama imkan sağlayan Orta Doğu Teknik Üniversitesine, deney süreci boyunca söz konusu laboratuvarın kullanımı konusunda desteklerini esirgemeyen Sayın Serkan ALKAN ile Sayın Alpay KARAGÖZ'e ve ayrıca deneylere katılan tüm gönüllü katılımcılara bu çalışmanın gerçekleşmesini mümkün kıldıkları için teşekkürü borç bilirim.

Son olarak, bu süreçte varlığıyla bana her zaman güç veren başta sevgili eşim Hamdi Kurtcebe AKSU olmak üzere hayatımın her anında sevgisini ve desteğini hissettiren canım aileme sonsuz teşekkür ederim.

İÇİNDEKİLER

	Sayfa
ÖZET	iv
ABSTRACT.....	v
TEŞEKKÜR.....	vi
İÇİNDEKİLER	vii
ÇİZELGELERİN LİSTESİ.....	x
ŞEKİLLERİN LİSTESİ.....	xii
RESİMLERİN LİSTESİ	xv
SİMGELER VE KISALTMALAR.....	xvi
1. GİRİŞ.....	1
2. TEMEL KAVRAMLAR.....	13
2.1. Bilişsel ve Zihinsel İş Yüğü	13
2.2. İşleyen Bellek ve N-Geri Hafıza Testleri.....	14
2.3. NASA-TLX.....	16
2.4. Göz İzleme	17
2.4.1. Odaklanma	20
2.4.2. Seçirme	21
2.4.3. Göz bebeği tepkisi.....	22
2.4.4. Göz kırpma.....	22
2.5. Beyin Görüntüleme	23
2.5.1. Beynin yapısı ve sinirsel aktivite	23
2.5.2. Beyin aktivitesini izleme yöntemleri	26
2.5.3. EEG	27
3. LİTERATÜR ARAŞTIRMASI.....	35

3.1. Göz İzleme Değişkenleri ile Zihinsel İş Yükü Arasındaki İlişkinin Yönünü Araştıran Çalışmalar.....	35
3.2. EEG Değişkenleri ile Zihinsel İş Yükü Arasındaki İlişkinin Yönünü Araştıran Çalışmalar	41
3.3. Zihinsel İş Yükü Seviyesinin Tahmininde Makine Öğrenmesi Algoritmalarının Kullanıldığı Çalışmalar	44
3.4. Problemin Tanımı ve Literatüre Katkı	53
4. MATERYAL VE METODOLOJİ	57
4.1. Veri Kümesi ve Deney Prosedürü.....	57
4.1.1. Veri toplamada kullanılan cihazlar, yazılımlar, ölçekler ve anketler	58
4.1.2. Katılımcılar	63
4.1.3. Deney prosedürü.....	64
4.1.4. Elde edilen veri dosyaları.....	71
4.1.5. Veri ön işleme ve değişken türetme.....	73
4.1.6. Analizde kullanılan değişkenler.....	78
4.2. Makine Öğrenmesi	80
4.2.1. Makine öğrenmesi ve ilişkili kavramlar.....	80
4.2.2. Makine öğrenmesi algoritmaları ve hiperparametreleri.....	85
4.2.3. Özellik seçimi.....	96
4.3. Model Performans Metrikleri.....	100
5. ARAŞTIRMA BULGULARI VE TARTIŞMA	109
5.1. Pilot Uygulamalarda Elde Edilen Bulgular	110
5.2. Temel Uygulamada Elde Edilen Bulgular	112
5.2.1. Performans ve öznel iş yükü değerlendirme sonuçlarına ilişkin bulgular	113
5.2.2. Göz izleme değişkenlerine ilişkin bulgular	129
5.2.3. EEG değişkenlerine ilişkin bulgular	134

	Sayfa
5.2.4. EEGLAB STUDY çalışması ile elde edilen bulgular	138
5.2.5. Makine öğrenmesi algoritmaları ile elde edilen tahmin modellerine ilişkin bulgular	150
6. SONUÇ	173
KAYNAKLAR	179
EKLER	195
ÖZGEÇMİŞ	201

ÇİZELGELERİN LİSTESİ

Çizelge	Sayfa
Çizelge 2.1. EEG dalgalarının frekans aralıkları.....	30
Çizelge 3.1. Zihinsel iş yükü sınıflandırmada makine öğrenmesi kullanılan çalışmalar	52
Çizelge 4.1. Karmaşıklık matrisinin gösterimi.....	101
Çizelge 4.2. Sınıflandırma modellerinin başarı ölçüleri.....	102
Çizelge 5.1. Normallik testi.....	113
Çizelge 5.2. Tanımlayıcı istatistikler	114
Çizelge 5.3. Çarpıklık ve basıklık değerleri	117
Çizelge 5.4. Ağırlıklı toplam NASA-TLX skoru ile alt boyutları arasındaki ilişki	118
Çizelge 5.5. Görev zorluk seviyesi ile performans değişkenleri arasındaki ilişki	120
Çizelge 5.6. NASA-TLX ile performans değişkenleri arasındaki ilişki	121
Çizelge 5.7. Görev zorluk seviyesine göre K-W testi sonuçları.....	122
Çizelge 5.8. NASA-TLX ve performans değişkenleri için K-W testi sıralamaları	122
Çizelge 5.9. Görev zorluk seviyelerine göre Mann-Whitney U testi sonuçları	124
Çizelge 5.10. Cinsiyete göre Mann-Whitney U testi sonuçları.....	126
Çizelge 5.11. Sınıflara göre Mann-Whitney U testi sonuçları.....	127
Çizelge 5.12. Göz izleme değişkenlerine ilişkin korelasyon testi sonuçları	129
Çizelge 5.13. Göz izleme değişkenlerine ilişkin K-W testi sonuçları	130
Çizelge 5.14. Göz izleme değişkenleri için K-W testi sıralama değerleri	131
Çizelge 5.15. Göze ilişkin değişkenler ile zihinsel iş yükü arasındaki ilişki.....	133
Çizelge 5.16. Görev zorluk seviyesi ile EEG değişkenleri arasındaki ilişki.....	134
Çizelge 5.17. NASA-TLX ile EEG değişkenleri arasındaki ilişki.....	136
Çizelge 5.18. EEG değişkenleri ile zihinsel iş yükü arasındaki ilişki	137
Çizelge 5.19. 111 değişkenli dört sınıflı modellerin performans sonuçları	152

Çizelge	Sayfa
Çizelge 5.20. 111 değişkenli dört sınıflı modele ilişkin yöntemsel karşılaştırma	153
Çizelge 5.21. 111 değişkenli dört sınıflı modele ilişkin hata matrisi.....	154
Çizelge 5.22. 111 değişkenli dört sınıflı modele ilişkin sınıflandırma performansı.....	155
Çizelge 5.23. Dört, üç ve iki sınıflı modellerin sınıflandırma performansları.....	155
Çizelge 5.24. Dört sınıflı modelde ÇDÖÖE yöntemi ile seçilen özellikler	157
Çizelge 5.25. Boyut azaltılmış ve ayarlanmış modellerin performans sonuçları.....	158
Çizelge 5.26. Model başarısının literatür ile karşılaştırılması	165
Çizelge 5.27. 111 değişkenli regresyon modellerinin performans sonuçları.....	168

ŞEKİLLERİN LİSTESİ

Şekil	Sayfa
Şekil 1.1. Tez çalışmasının adımları	10
Şekil 2.1. Beyin lobları	24
Şekil 2.2. Sinyal görüntüleri	28
Şekil 2.3. Prefrontal lobdan alınan EEG sinyaline ilişkin güç spektrumu örneği.....	30
Şekil 2.4. Farklı bantlardaki EEG dalgaları	33
Şekil 4.1. Uygulamada izlenen yol haritası	57
Şekil 4.2. 14 kanallı EEG cihazının elektrot yerleşimi.....	60
Şekil 4.3. Ham EEG ekranı.....	60
Şekil 4.4. Katılımcıların cinsiyet dağılımı	64
Şekil 4.5. Katılımcıların sınıf dağılımı	64
Şekil 4.6. Deney prosedürü.....	66
Şekil 4.7. EEG bağlantı kalitesi ekranı	67
Şekil 4.8. EEG sinyal kalitesi ekranı	67
Şekil 4.9. Göz mesafesi belirleme ekranı.....	68
Şekil 4.10. Kalibrasyon ekranı.....	68
Şekil 4.11. Kalibrasyon onay ekranı	69
Şekil 4.12. N-geri görev kuralları	70
Şekil 4.13. Göz kırpma hareketinin tespitine ilişkin algoritma	75
Şekil 4.14. HFD ve bant gücü ekranı.....	77
Şekil 4.15. Makine öğrenmesi çeşitleri.....	83
Şekil 4.16. Özellik seçiminde filtreleme yöntemleri yaklaşımı	98
Şekil 4.17. Özellik seçiminde sarmal yöntemler yaklaşımı.....	99
Şekil 4.18. Özellik seçiminde gömülü yöntemler yaklaşımı	99

Şekil	Sayfa
Şekil 4.19. AİK eğrisi örneği	103
Şekil 5.1. Ağırlıklı toplam NASA-TLX skoru ve zorluk seviyesi arasındaki ilişki	119
Şekil 5.2. Doğru yanıt oranı ve zorluk seviyesi arasındaki ilişki.....	119
Şekil 5.3. Ortalama tepki süresi ve zorluk seviyesi arasındaki ilişki.....	120
Şekil 5.4. Göz izleme değişkenlerinin zorluk seviyesine göre değişimi.....	132
Şekil 5.5. Filtreleme öncesi sinyal görüntüsü	139
Şekil 5.6. Filtreleme sonrası sinyal görüntüsü	139
Şekil 5.7. STUDY oluşturma ekranı	140
Şekil 5.8. STUDY tasarımı belirleme ekranı	141
Şekil 5.9. STUDY giriş ekranı.....	141
Şekil 5.10. Bileşen çıkarma kriterlerini belirleme ekranı	142
Şekil 5.11. Bileşen çıkarma öncesi etiketleme ekranı.....	142
Şekil 5.12. Ön hesaplama ekranı.....	143
Şekil 5.13. Veri görselleştirme ekranı.....	144
Şekil 5.14. İstatistiksel parametreleri belirleme ekranı.....	145
Şekil 5.15. Beyin haritaları	146
Şekil 5.16. Zorluk seviyesine göre bant gücünde anlamlı farklılık gösteren frekanslar.	147
Şekil 5.17. Farklı zorluk seviyeleri için teta gücü spektrumu.....	148
Şekil 5.18. Farklı zorluk seviyeleri için alfa gücü spektrumu	149
Şekil 5.19. Modellerin doğruluk performanslarının karşılaştırılması.....	153
Şekil 5.20. Dört sınıflı model için HGAM tarafından seçilen 27 özelliğin önem sırası.	157
Şekil 5.21. Boyut azaltma ve hiperparametre ayarlama sonrası performans değişimi ...	159
Şekil 5.22. Dört sınıflı model için GAM tarafından seçilen 18 özelliğin önem sırası....	161
Şekil 5.23. Dört sınıflı model için erken durdurma sonrası performans karşılaştırma ...	162

Şekil	Sayfa
Şekil 5.24. Üç sınıflı model için GAM tarafından seçilen 48 özelliğin önem sırası	162
Şekil 5.25. Üç sınıflı model için erken durdurma sonrası performans karşılaştırma.....	163
Şekil 5.26. İki sınıflı model için GAM tarafından seçilen 37 özelliğin önem sırası.....	164
Şekil 5.27. İki sınıflı model için erken durdurma sonrası performans karşılaştırma.....	164
Şekil 5.28. Erken durdurma sonrası performans karşılaştırma	165
Şekil 5.29. Modellerin HKOK performanslarının karşılaştırılması.....	168
Şekil 5.30. Regresyon modeline katkı sağlayan en önemli değişkenler	169
Şekil 5.31. Regresyon modeli için GAM tarafından seçilen 41 özelliğin önem sırası ...	170
Şekil 5.32. Regresyon modeli için erken durdurma sonrası performans karşılaştırma ..	171
Şekil 5.33. Farklı hiperparametre değerlerine göre performans değişimi	171

RESİMLERİN LİSTESİ

Resim	Sayfa
Resim 2.1. Gözün kameradan görünüşü.....	18
Resim 4.1. EMOTIV EPOC X EEG başlığı ve USB aparatı.....	59
Resim 4.2. Tobii X2-60 göz izleme cihazı.....	62
Resim 4.3. İnsan bilgisayar etkileşimi laboratuvarı.....	65
Resim 4.4. Deneylerin uygulanışı.....	72

SİMGELER VE KISALTMALAR

Bu çalışmada kullanılmış simgeler ve kısaltmalar, açıklamaları ile birlikte aşağıda sunulmuştur.

Simgeler	Açıklamalar
AF	Prefrontal
C	Central
%	Yüzde
dB	Desibel
F	Frontal
Hz	Hertz
mm	Milimetre
ms	Milisaniye
L1	Lasso Regresyon
L2	Ridge Regresyon
O	Oksipital
P	Pariyetal
R²	Determinasyon Katsayısı
T	Temporal

Kısaltmalar	Açıklamalar
AGAM	Aşırı Gradyan Artırma Makineleri (eXtreem Gradient Boosting-XGBoost)
AİK	Alıcı İşletim Karakteristiği (Receiver Operating Characteristic-ROC)
BAP	Bilimsel Araştırma Projeleri
BBA	Bağımsız Bileşenler Analizi (Independent Component Analysis-ICA)
ÇDÖÖE	Çapraz Doğrulama ile Özyinelemeli Özellik Eleme (Recursive Feature Elimination with Cross Validation-RFECV)
çd	çapraz doğrulama (cross validation-cv)
DAA	Doğrusal Ayırma Analizi (Linear Discriminant Analysis-LDA)

Kısaltmalar	Açıklamalar
DN	Doğru Negatifler
DNO	Doğru Negatif Oranı
DP	Doğru Pozitifler
DPO	Doğru Pozitif Oranı
DVM	Destek Vektör Makineleri (Support Vector Machines-SVM)
EAA	Eğri Altındaki Alan (Area Under Curve-AUC)
EEG	Elektroensefalografi (Electroencephalography-EEG)
EEGLAB	Elektroensefalografi Laboratuvarı (Electroencephalography Laboratory-EEGLAB)
EKG	Elektrokardiyografi (Electrocardiography-ECG)
EMG	Elektromiyografi (Electromyography-EMG)
EOG	Elektrookülografi (Electrooculography-EOG)
fMRG	fonksiyonel Manyetik Rezonans Görüntüleme (functional Magnetic Resonance Imaging-fMRI)
fyKS	fonksiyonel Yakın Kızılötesi Spektroskopisi (functional Near Infrared Spectroscopy-fNIRS)
GAM	Gradyan Artırma Makineleri (Gradient Boosting Machines-GBM)
GDY	Galvanik Deri Yanıtı (Galvanic Skin Response-GSR)
GİBP	Görsel İşitsel Bilişsel Psikomotor (Visual Auditory Cognitive Psychomotor-VACP)
GKKY	Göz Kapağı Kapanma Yüzdesi (Percentage of Eyelid Closure-PERCLOS)
GUGT	Görevle Uyarılan Göz Bebeği Tepkisi (Task Evoked Pupil Response-TEPR)
HFD	Hızlı Fourier Dönüşümü (Fast Fourier Transform-FFT)
HGAM	Hafif Gradyan Artırma Makineleri (Light Gradient Boosting Machines-LightGBM)
HKO	Hata Kareler Ortalaması (Mean Squared Error-MSE)
HKOK	Hata Kareler Ortalamasının Karekökü (Root Mean Squared Error-RMSE)
HKT	Hata Kareler Toplamı (Sum of Squared Error-SSE)
IMPRINT	Improved Performance Research Integration Tool

Kısaltmalar	Açıklamalar
İA	İlgi Alanları (Areas of Interest-AOI)
KA	Karar Ağaçları (Decision Trees-DT)
KEYK	K En Yakın Komşu (K Nearest Neighborhood-KNN)
KH	Kalp Hızı (Heart Rate-HR)
KHD	Kalp Hızı Değişkenliği (Heart Rate Variability-HRV)
K-S	Kolmogorov-Smirnov
K-W	Kruskal-Wallis
MATLAB	Matris Laboratuvarı (Matrix Laboratory-MATLAB)
MEG	Manyetoensefalografi (Magnetoencephalography-MEG)
MKK	Matthew Korelasyon Katsayısı (Matthew's Correlation Coefficient-MCC)
NASA-TLX	National Aeronautics and Space Administration-Task Load Index
NTD	Negatif Tahmin Değeri
OİP	Olaya İlişkin Potansiyeller (Event Related Potentials-ERP)
OMH	Ortalama Mutlak Hata (Mean Absolute Error-MAE)
ÖÖE	Özyinelemeli Özellik Eleme (Recursive Feature Elimination-RFE)
pBBE	pasif Beyin Bilgisayar Etkileşimi (passive Brain Computer Interactions-pBCI)
PET	Pozitron Emisyon Tomografisi (Positron Emission Tomography-PET)
PTD	Pozitif Tahmin Değeri
RO	Rastgele Ormanlar (Random Forests-RF)
SPSS	Statistical Package for Social Sciences
SWAT	Subjective Workload Assessment Technique
VBA	Visual Basic for Applications
WP	Workload Profile
YN	Yanlış Negatifler
YNO	Yanlış Negatif Oranı
YP	Yanlış Pozitifler
YPO	Yanlış Pozitif Oranı
YSA	Yapay Sinir Ağları (Artificial Neural Networks-ANN)

1. GİRİŞ

Problemin tanımı

İş yükü fiziksel çabadan ortaya çıksa bile görevleri gerçekleştirirken meydana gelen zihinsel taleplerin etkileşimini yansıtan bilişsel bir yapıya sahiptir (Cain, 2007). Diğer yandan, günümüzün endüstriyel sistemlerinde otomasyonun artmasıyla birlikte; bedensel çalışma yerini gözle görülür bir biçimde zihinsel çalışmaya bırakmıştır. Sürdürülebilir dikkat ve problem çözme becerisi, fiziksel güç ve dayanıklılıktan daha önemli hale gelmiştir. Dolayısıyla, insan-makine sistemlerinde etkinliğin artırılabilmesi amacıyla çalışanların anlama, neden bulma, bilginin kullanımı gibi zihinsel süreçlerini inceleyen ergonomi dalı olan bilişsel ergonomi öne çıkmış; çalışmalar, geleneksel ergonomide olduğu gibi bedensel yorgunluğun ve konforsuzluğun nedenleri ve sonuçlarından daha çok zihinsel iş yükü kaynaklarına ve etkilerine yönelmiştir. Bilişsel ergonomi insanın bilişsel yetenek ve kısıtlılıkları ile makine, görev ve çevre arasında denge bulmaya çalışmaktadır. Bu doğrultuda; dikkatin dağılımı, karar verme, öğrenme yetenekleri, stres, insan hatası, zihinsel yükün bilişsel boyutu gibi durumları incelemektedir. Fiziksel ergonomi çalışma kalitesine odaklanırken, bilişsel ergonomi işin ve çıktının kalitesine odaklanmaktadır (Bommer ve Fendley, 2018). Özellikle zihinsel iş yükünün ölçülebilmesi veya tahmin edilebilmesine yönelik olarak yapılan çalışmalar bilişsel ergonomi içerisinde önemli bir yer teşkil etmektedir.

İş yükü birçok farklı değişkenin bir araya gelmesinden kaynaklanan çok boyutlu bir kavramdır (Cain, 2007). Doğrudan gözlenemez, ancak davranışların izlenmesi ile psikolojik ve fizyolojik süreçlerin ölçülmesinden çıkarılabilir. Cain'e (2007) göre iş yükü, çalışan ve sistem performansını tahmin etmek amacıyla ölçülmektedir. Zihinsel iş yükü seviyesini bilmek, operatöre atanacak görev sayısına ve görevlerin zorluk seviyelerinin ne olacağına karar verebilmek açısından önemlidir. Bu sayede görev performansını yüksek tutabilmek mümkün olacaktır. İş yükünün ölçümündeki nihai hedefin ise, çalışma koşullarının iyileştirilmesi olduğu varsayılmaktadır. Bu nedenle zihinsel iş yükünün değerlendirilmesi süreci; ergonominin temel amacı olan daha konforlu, daha tatmin edici, daha verimli ve daha güvenli işyerleri elde etmeyi hedefleyen çalışmaların önemli bir parçası olmayı sürdürmektedir (Rubio, Diaz, Martin ve Puente, 2004). Özetle, ergonomi

alanıyla uğraşanlar zihinsel iş yükü ölçümünden insan-makine sistemlerinin etkinliğini değerlendirmede, performansı arttırmada ve farklı tasarımlar arasından seçim yapmada yararlanmaktadırlar. Zihinsel iş yükünün insan performansını ve dolayısıyla insan-makine sistemlerinin etkinliğini doğrudan etkilediği düşünüldüğünde, kritik görevler başta olmak üzere askeriye, havacılık, sağlık, bilgisayar ve araç kullanımı gibi pek çok alanda zihinsel iş yükü ölçümünün önemle üzerinde durulan bir çalışma alanı olması kaçınılmazdır.

Görevlerin etkin dağılımı için zihinsel iş yükünü doğru ve etkin olarak ölçen araçlara ihtiyaç vardır. Ancak, çok boyutlu olması yani pek çok faktörden etkilenmesi sebebiyle zihinsel iş yükünün doğrudan ve kolaylıkla ölçülmesi mümkün değildir. Bu doğrultuda, zihinsel iş yükünü en doğru şekilde ölçen yöntemleri bulmak veya en doğru tahmin eden modelleri geliştirmek bu alanla ilgilenen araştırmacıların temel hedefleri olmuştur.

Rusnock, Borghetti ve McQuaid (2015) tarafından, zihinsel iş yükü ölçümünde kullanılan araçlar nesnel-öznel ve deneysel-analitik şeklinde iki boyutta ele alınmıştır. Nesnel ölçümler gerçek yaşamdan elde edilen bilgi ve verilere dayanmaktadır. Verinin toplandığı kişiye bağlı değildir. Öznel yöntemlerde ise verinin toplandığı kişilerin görüşleri esas alınmaktadır. Ölçüm yöntemi deneysel ise deneyimle, analitik ise analizle elde edilmektedir. Deneysel yöntemler, iş yükünü doğrudan ölçmeyi amaçlamakta; analitik ölçekler ise görev, operatör ve çevreden elde edilen bilgiyle çıkarım yapmaya çalışmaktadır.

Nesnel-deneysel ölçütler, birincil/ikincil performans ölçüleri ve fizyolojik yöntemlerdir. Öznel-deneysel araçlara öznel iş yükü değerlendirme ölçekleri örnek gösterilmektedir. Nesnel-analitik araçlar matematiksel modeller, öznel-analitik araçlar ise uzman görüşleridir. Nesnel-analitik yöntemler; görev, çevre ve bireye özgü bilgilerin girdi olarak kullanıldığı tahmini iş yükünü ölçen matematiksel modellerdir. Bu modeller genellikle görev tasarımı veya yeniden tasarım sürecinde kullanılmaktadır. Deneysel ölçümlerin elde edilmesi zor olduğunda ayrıntılı görev analizi ile elde edilmektedir. Avantajları; elde edilen iş yükü değerlerinde tutarlılık, farklı kaynak gereksinimleri için farklılaştırılmış iş yükü değerlerini içerme yeteneği ve herhangi bir zaman ölçeğinde iş yükü değeri hesaplayabilme yeteneği şeklinde sıralanabilir. Öznel-analitik yöntemlerde deneyimli uzman görüşlerine dayanarak zihinsel iş yükünün tahmin edilmesi amaçlandığından bu

yöntemlerden akademik çalışmalarda yararlanılmamaktadır. Akademik çalışmalarda genellikle öznel-deneysel ve nesnel-deneysel yöntemler tercih edilmektedir.

Öznel-deneysel değerlendirme yöntemlerinin literatürde en sık başvurulanan yöntemler arasında yer almasının başlıca sebebi kullanım kolaylığıdır. Maliyetsiz ve kolayca uygulanabilmesi ve basit hesaplamalarla anlamlı ve güvenilir sonuçlar elde etmeye imkan sağlaması sebebiyle başta National Aeronautics and Space Administration-Task Load Index (NASA-TLX) olmak üzere; Subjective Workload Assessment Technique (SWAT), Workload Profile (WP), Cooper-Harper Ölçeği ve Bedford Ölçeği en çok kullanılan öznel-deneysel yöntemlerdir. Diğer yandan; öznel yöntemlerin, avantajlarının yanında bazı dezavantajları da mevcuttur (Choi, Lee, Ha ve Seong, 2018):

- Fiziksel ve zihinsel iş yükünün karışma potansiyeli
- Dış ortamdaki gelen zihinsel zorlayıcıların gerçek iş yükünden ayrıştırılma güçlüğü
- Operatörün öznel olarak derecelendiremediği bilgilerin bilinçsiz bir şekilde işlenmesi
- Öznel derecelendirmelerin ve görev performansının ayrışması
- İyi ifade edilmiş bir sorunun gerekliliği
- Operatörün kısa süreli hafızasına bağlı olma

Öznel-deneysel yöntemlerin dezavantajlarını ortadan kaldırmak ve zihinsel iş yükünü bireylerin algıladığından bağımsız, daha güvenilir bir ölçüm ile elde edebilmek amacıyla nesnel-deneysel tekniklerden yararlanılmaktadır. Başlıca nesnel-deneysel yöntemler performans ölçüleri ve fizyolojik yöntemlerdir. Çalışanların birincil ya da ikincil görevlerdeki performansları ölçülerek zihinsel iş yükünün ölçülmesini hedefleyen çalışmalar mevcuttur. Yaygın olarak kullanılan birincil görev performans ölçüleri hız, doğruluk, tepki veya yanıt süreleri ve hata oranlarıdır. Gözlemlenen hata sayısının olası hata sayısına bölümüne karşılık gelen hata olasılığı da kullanılan performans metrikleri arasındadır (Bommer ve Fendley, 2018). İkincil görev ölçüleri ise, çalışanın birincil görevleri yerine getirirken kalan kapasitesinin bir işaretidir ve nispeten daha teşhis edicidir (Cain, 2007). İkincil görevin kullanıldığı çalışmalarda, bireylerin dikkat kaynaklarını her iki göreve ayırarak dağıtmaları istenir. İstenen aktiviteler dikkat sınırını aştığında performans düşmektedir (DiDomenico ve Nussbaum, 2011). İkincil görevler arasında ritmik dokunma, rastgele sayı üretimi, zaman tahmini ve zaman üretimi yer almaktadır.

İnsan-makine etkileşimi sırasında zihinsel iş yükünü araştırmak için fizyolojiden yararlanan çalışmalar ise çoğunlukla zihinsel göstergeler olarak nörofizyolojik, kardiyovasküler ve göz davranışlarına ilişkin ölçümlerin değerlendirilmesini içermektedir. Kalp hızı (KH) ve değişkenliği (KHD), galvanik deri yanıtı (GDY), elektromiyografi (EMG), elektrokardiyografi (EKG), elektroensefalografi (EEG), fonksiyonel yakın kızılötesi spektroskopisi (fYKS), fonksiyonel manyetik rezonans görüntüleme (fMRG), manyetoensefalografi (MEG), elektrookülografi (EOG) ve olaya ilişkin potansiyeller (OİP) gibi biyolojik belirteçler zihinsel iş yükü değerlendirmesi için kullanılan nörofizyolojik ölçütlerdir. Fizyolojik ölçümler sürekli ve bireye özgüdür. Gerçek zamanlı ve hızlı geri bildirim alınabilen yöntemlerdir. Özellikle nörofizyolojik ölçümlerin bilişsel talep dalgalanmalarını ayırt etmede hem davranışsal hem de öznel yöntemlere üstün geldiği yaygın olarak gösterilmiştir. Zihinsel iş yükünün çevrimiçi nörofizyolojik ölçümleri, sadece izleme teknikleri olarak değil, aynı zamanda işleyen faaliyetler sırasında kullanıcıya destek araçları olarak da önemli rol oynamaktadır. Bilişsel aktivitedeki değişiklikler gerçek zamanlı olarak ölçülebildiğinden, operatörün görev sırasında optimum zihinsel iş yükü seviyesini sürdürmesine yardımcı olmak için görev gereksinimini manipüle etmek de mümkün olabilmektedir. Bu doğrultuda, nörofizyolojik iş yükü değerlendirmesi, gerçek operasyonel ortamlardaki pasif Beyin Bilgisayar Etkileşimi (pBBE) uygulamalarını gerçekleştirmek için kullanılabilen bir yöntem olarak karşımıza çıkmaktadır (Arico ve diğerleri, 2016). Fizyolojik yöntemlerin en önemli avantajları aşağıda sıralanmaktadır (Dirican ve Göktürk, 2011):

- Nesneldir, katılımcının algıladığına bağımlı değildir. Bu özellik ölçüm tekniğinin güvenilirliğini artırır.
- Değişen seviyelerdeki farklı bilişsel süreçlere karşı duyarlılık gösterir.
- İkincil görev performans ölçüleri gibi esas göreve doğrudan müdahale etmemekte ve onu kesintiye uğratmamaktadır.
- Gerekli bilgileri örtük (gizli) olarak sağlar.
- Sürekli sinyallerdir ve bu nedenle gerçek zamanlı olarak kullanılabilir. Araştırmacılara hem kısa vadeli hem de uzun vadeli bedensel inceleme yapma imkânı sağlar.

Boyut, ağırlık ve güç kısıtlamaları, pBBE uygulamalarını gerçekleştirmek için kullanılabilecek nörolojik geri besleme türlerini sınırlamaktadır. Örneğin, fMRG ve MEG tekniklerinin uygulanması bir oda büyüklüğünde ve taşınabilir olmayan teçhizat gerektirmesi sebebiyle zordur. EEG ve fYKS, operasyonel ortamlarda kullanılabilen pBBE uygulamalarını gerçekleştirmek için doğrudan kullanılabilecek en muhtemel adaylardır (Arico ve diğerleri, 2016).

Yukarıda açıklanan pek çok yöntem zihinsel iş yükü ölçümünde gösterge olarak kullanılmakla birlikte, zihinsel iş yükünü tek başına mükemmel şekilde ölçebilen bir metot bulunmamaktadır (Cegarra ve Chevalier, 2008). Örneğin, her ne kadar katılımcının algıladığından bağımsız nesnel bir yöntem olsa da zihinsel iş yükünü yalnızca performans ölçülerine göre değerlendirmek bizi her zaman doğru sonuca ulaştırmayabilir. Çünkü zihinsel iş yükünün en kilit özelliklerinden biri performans çıktılarından bağımsız davranabilmesidir. Çalışanların zorlu görevlerle karşılaşmalar dahi, çaba ve motivasyonlarını artırmak suretiyle talep edilen performans çıktı seviyesini sürdürebilmeleri mümkün olabilmektedir. Benzer şekilde, performanstaki düşüş de iş yüküne özgü olmayabilir. Hem düşük hem de yüksek zihinsel iş yükü seviyelerinde düşük performans gözlenebilmektedir. Yani düşük performansa sebep olan başka faktörler olması mümkündür. Dolayısıyla performans ile zihinsel iş yükü arasındaki ilişki doğrusal değildir. Performansında değişiklik tespit edilmeden önceki zaman periyodunda, operatör uzun süredir yüksek zihinsel iş yüküne maruz kalmış olabilir. Bu durumda performanstaki değişiklik tespit edildikten sonra iş yükünde düzenlemeye gidilmesi performans düşüşünü önlemek için geç kalınmış bir hamle olacaktır (Ayaz ve diğerleri, 2012; Rusnock ve diğerleri, 2015).

Zihinsel iş yükü değerlendirme teknikleri, birincil görev performansını ihlal etmeden görev taleplerindeki bilişsel dalgalanmalara karşı hassas olmalıdır. Bu bakımdan, zihinsel iş yükünü, asıl görevin yerine getirilmesi sırasında öznel yöntemler kullanarak ölçmek, kullanıcının performansını olumsuz etkileyebilmektedir. Bunu önlemek için NASA-TLX gibi öznel yöntemler genellikle görev tamamlandıktan sonra uygulanmakta, bunun sonucu olarak da görevin yapıldığı süre içerisindeki iş yükü dalgalanmalarına duyarsız kalmaktadır. Bu yöntemlerin algılanan ortalama iş yükünü ölçtüğü varsayılmasına rağmen elde edilen sonuçlar en son hissedilen veya hatırlanan iş yüküne yanlı bir eğilim göstermektedir. Yani operatörün kısa süreli hafızasına bağımlı kalmaktadır (Choi ve

diğerleri, 2018). Özellikle uzay ve robotik cerrahi gibi kritik görevlerde kullanılan karmaşık yüksek hassasiyetli insan-makine sistemlerinin etkinliği ve güvenliği; bilişsel hazır olma, iş yükünü yönetme ve operatörlerinin durumsal farkındalığı ile yakından ilgilidir. Öznel operatör raporları ile davranışsal ölçütler zihinsel aşırı yüklenmeyi izlemek için tek başlarına yeterince güvenilir değildir ve olumsuz sonuçlara yol açabilir (Ayaz, M. İzzetoğlu, K. İzzetoğlu ve Onaral, 2016). EOG, EKG ve GDY aktiviteleri gibi bazı fizyolojik tekniklerin bile gerçek zamanlı nesnel veri sağladıkları halde farklı zihinsel durumlarla (stres, yorgunluk, uyuşukluk) korelasyon göstermeleri nedeniyle tek başlarına kullanımlarının zihinsel iş yükü ölçümü açısından yanıltıcı olabileceği, sadece beyne doğrudan bağlı diğer beyin görüntüleme teknikleriyle kombinasyon halinde kullanıldığında faydalı sonuç verecekleri gösterilmiştir. Fizyolojik yöntemlerin birlikte kullanıldığı çalışmaların daha etkili sonuçlar verdiği örneklere literatürde sıklıkla rastlanmaktadır (Heine ve diğerleri, 2017; Liu, Ayaz ve Shewokis, 2017).

Yukarıda belirtilen farklı tekniklerle ölçülmeye çalışılan zihinsel iş yükü kavramının önemi göz önüne alındığında, farklı ölçüm yöntemlerini bir arada kullanarak zihinsel iş yükünü en doğru şekilde tahmin eden modelin geliştirilmesi, literatürde hala tam olarak başarılmış sayılamayacak bir konudur. Bu alanda yapılan araştırmalar son yıllarda oldukça önem kazanmıştır.

Çalışmanın amaçları

Bu tez kapsamında göze ilişkin veriler ve EEG sinyalleri ile zihinsel iş yükü arasında ilişki olup olmadığının araştırılması ve makine öğrenmesi algoritmaları aracılığıyla zihinsel iş yükünün tahmin edilmesine yönelik sınıflandırma ve regresyon modelleri geliştirilmesi hedeflenmiştir. Bu doğrultuda, daha güçlü ve teşhis edici bir zihinsel iş yükü modeli elde etmek için tek bir fizyolojik yöntemin kullanılması sakıncalı bulunduğundan bu tez çalışmasında da hibrit bir model tercih edilmiştir. Göz izleme ve EEG ölçümleri ile elde edilen veriler, performans ölçüleri ve öznel değerlendirme ölçeği NASA-TLX ile desteklenmiştir.

Monod ve Kapitaniak (1999) tarafından, bir içsel bilişsel yük etkeni olarak kabul edilen görev zorluğunun bilişsel yükü doğrudan etkileyeceği öne sürülmüştür. Hatırlanacak öğelerin sayısına göre belirlenen görev zorluğunun bir bellek görevindeki performansı

engellediği gösterilmiş ve yine aynı değişkenin psikofizyolojik ölçümleri, özellikle de otonom sinir sisteminin kontrolü altındaki bileşenleri etkilediği kanıtlanmıştır (Galy, Cariou ve Melan, 2012). Zihinsel iş yükünün en önemli belirleyicilerinden biri olması sebebiyle, tez kapsamında oluşturulan sınıflandırma modelinde tahmin edilen hedef değişken olarak görev zorluk seviyesi seçilmiştir.

Son yıllarda sıklıkla kullanılması ve geçerliliğinin bu tezde yer verilen pek çok çalışmayla gösterilmiş olması sebebiyle, çalışmada fizyolojik yöntemlerden yararlanılmıştır. Göz izleme tekniğinin tercih edilmesinin sebebi diğer fizyolojik yöntemlere göre daha az görev kesintisine sebep olması ve tek bir sensör ile gözle ilgili birden fazla bilgi türüne ulaşılabilir olmasıdır. EEG yöntemi ise, zamansal olarak gerçek bilgiye ulaşmada etkili bir araç olması ve beyin sinyallerinin zihinsel iş yükünün analizindeki etkinliğinin pek çok çalışma ile kanıtlanmış olması sebebiyle tercih edilmiştir. Hem EEG hem de göz izleme tekniklerinin zihinsel iş yüküne ilişkin çalışmalarda kullanımı üçüncü bölümde detaylandırılacaktır.

Sonuçların daha doğru yorumlanabilmesi açısından fizyolojik yöntemlerin başka yöntemlerle birlikte kullanılmasına ihtiyaç duyulmaktadır. Bu doğrultuda zihinsel iş yükü analizine NASA-TLX ve performans ölçütleri de dâhil edilmiştir. Özellikle uygulanan görevlerde kullanılan zorluk seviyelerinin zihinsel iş yükünde beklenen değişime yol açıp açmadığı, bu yöntemler ile test edilmiştir. Diğer bir ifadeyle; görev zorlaştıkça zihinsel iş yükünün artacağı, dolayısıyla görev zorluk seviyelerinin zihinsel iş yükü seviyelerine karşılık geldiği varsayılmakla birlikte, söz konusu varsayımın geçerliliğinin NASA-TLX ve performans verileriyle test edilmesi de bu çalışmanın analiz sürecinin bir parçasını oluşturmaktadır.

Tezin temelde dört amacı bulunmaktadır:

- Tezin “görev zorluk seviyelerinin zihinsel iş yükü seviyelerine karşılık geldiği” temel varsayımını destekleyecek bulguların elde edilmesi (performans, öznel iş yükü değerlendirme ve görev zorluk seviyeleri arasındaki ilişkinin analiz edilmesi)
- EEG ve gözle ilgili değişkenlerin görev zorluk seviyesi, öznel iş yükü değerlendirme ve performans ile aralarındaki ilişkinin analiz edilmesi ve literatürdeki benzer çalışmalar ile karşılaştırılması

- EEG ve gözle ilgili değişkenleri kullanarak toplam NASA-TLX skorunu tahmin eden regresyon modelinin ve zihinsel iş yükü seviyesi olarak değerlendirilen görev zorluk seviyesini tahmin eden sınıflandırma modelinin farklı makine öğrenmesi algoritmaları kullanılarak geliştirilmesi
- Ham EEG verisini işleyen ve beynin farklı bölümlerinde, farklı görev koşulları altında, bant gücündeki değişimi analiz eden Elektroensefalografi Laboratuvarı (EEGLAB) STUDY çalışmasını gerçekleştirmek suretiyle literatüre bir uygulama örneği sunulması

Çalışmanın özgün yönleri

Zihinsel iş yükünün sınıflandırma problemi olarak ele alındığı çalışmalar; deneyde kullanılan görev türü, modelin girdi değişkenlerinin elde edilmesi için kullanılan teknikler, tahmin edilen sınıf sayısı ve problemin çözümünde kullanılan algoritmalar açısından farklılık göstermektedir. Beklenen zihinsel manipülasyonu; aritmetik işlemler (Borys, Plechawska-Wojcik, Wawrzyk ve Wesolowska, 2017a; Plechawska-Wojcik, Tokovarov, Kaczorowska ve Zapala, 2019; Zarjam, Epps ve Lovell, 2015), rakam-sembol eşleştirme testleri (Kaczorowska, Wawrzyk ve Plechawska-Wojcik, 2020) simülasyon görevleri (Yan, Wei ve Tran, 2019) gibi görev türlerinden yararlanarak sağlayan çalışmalar mevcuttur. Bu çalışmada görev türü olarak bir sonraki bölümde açıklanacak olan n-geri hafıza testleri kullanılmıştır. Literatür araştırması bölümünde detaylarına yer verilecek olan çalışmalarda, zihinsel iş yükünün kategorik olarak tahmin edildiği modeldeki sınıf sayısı arttıkça problem karmaşıklığı artmakta ve model performansı düşmektedir. Mümkün olduğunca fazla kategorinin ele alındığı bir problemde, zihinsel iş yükünün içinde bulunduğu seviyeyi doğru tahmin edebilen bir modelin elde edilebilmesi, ele alınan girdi değişkenleri ile zihinsel iş yükü düzeyinin derinlemesine ayrıştırılabildiği anlamına gelir ki bu durum daha detaylı bir analizle karar verme ortamına imkan sağlayacaktır. Bu doğrultuda, çalışmanın temel motivasyonlarından biri literatürde genellikle iki ya da üç sınıflı olarak ele alınan problemi dört sınıflı olarak ele almak ve mümkün olan en iyi doğruluk oranını veren modeli elde etmek olmuştur. Diğer yandan, katılımcıların göz izleme ve EEG verilerinin eş zamanlı olarak toplandığı bir deney prosedürü izlenmesi de bu çalışmanın özgün yönlerinden birini oluşturmaktadır.

İlgili literatür incelendiğinde; özellikle son on yılda, fizyolojik veriler ile makine

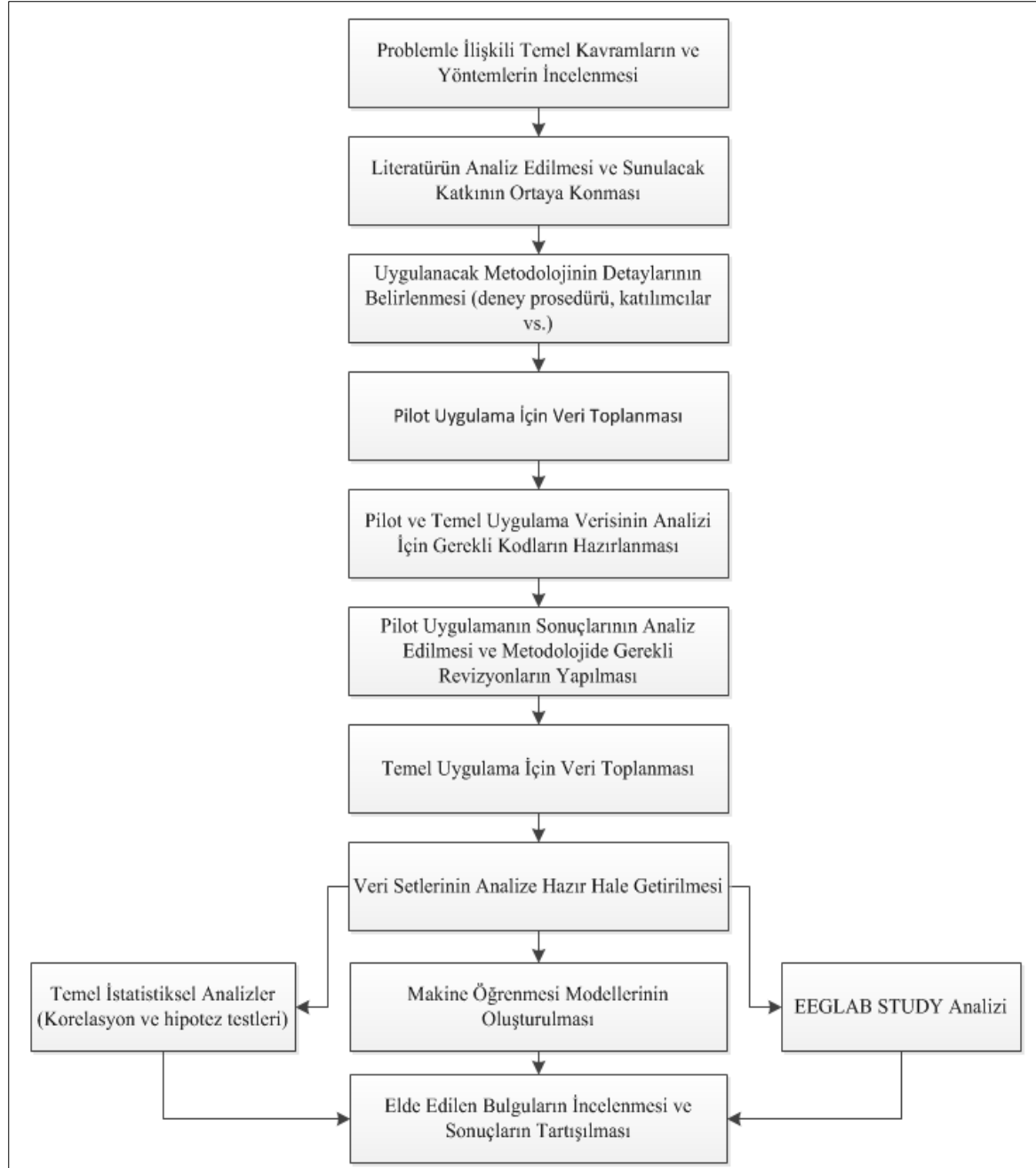
öğrenmesine dayalı metodolojiler kullanılarak bir zihinsel iş yükü tahmin modeli oluşturmaya yönelik çalışmalara ağırlık verildiği görülmektedir. Bu çalışmada, zihinsel iş yükü regresyon ve sınıflandırma problemi şeklinde iki farklı açıdan incelenmiştir. EEG ve göz izleme değişkenlerine ilişkin değerleri bilinen katılımcıların zihinsel iş yüküne karşılık geldiği varsayılan görev zorluk seviyelerinin kategorik olarak tahmin edilebilmesi sınıflandırma problemi, yine aynı fizyolojik değişkenler kullanılarak ağırlıklı toplam NASA-TLX skorunun tahmin edilebilmesi ise regresyon problemi olarak ele alınmıştır. Her iki problem çözümü için de makine öğrenmesi algoritmalarından yararlanılmıştır. Bu kapsamda, son on yılda geliştirilen ve daha önce EEG ve göz izleme teknikleriyle zihinsel iş yükünün tahminini amaçlayan çalışmalarda kullanımına rastlanmayan makine öğrenmesi algoritmaları da analize dâhil edilmiştir. EEG ve göz tepkisine ilişkin değişkenlerin hem öznel iş yükü değerlendirme sonuçları ve zorluk seviyesi ile muhtemel ilişkileri araştırılmış hem de ele alınan hedef değişkenleri tahmin etmek amacıyla kurulan modellerin başarısına katkıları incelenmiştir. Model optimizasyonu sürecinde ayarlanması gereken algoritma hiperparametreleri ve özellik seçim yöntemleri araştırılmış ve uygulanmıştır. Bu aşamada, tez boyunca model optimizasyonu ve model ayarlama ifadelerinin birbirinin yerine kullanıldığını belirtmekte fayda görülmektedir. Benzer şekilde, model başarısı ve model performansı ifadeleri de birbiriyle aynı anlamda kullanılmıştır. Değişken ve özellik ifadelerinin birbiri yerine kullanıldığı durumlar olmakla birlikte, özellik ifadesi daha çok önem taşıyan değişkenler için kullanılan öznitelik anlamında kullanılmıştır. Ayrıca, önemli görülen kavramların literatürdeki İngilizce kullanımları parantez içerisinde verilmiştir.

Belirtilen analizlere ek olarak, ham EEG verisinin Matris Laboratuvarı (MATLAB) yazılımı tabanlı EEGLAB uygulaması ile işlenerek analize dâhil edilmesi hem istatistiksel analiz sonuçlarının görselleştirilerek yorumlanabilmesine katkı sağlamış hem de söz konusu programın kullanımına ilişkin literatüre örnek bir uygulama sunulmuştur.

Yapılan çalışmada bu şekilde bir yol izlemek, hem EEG sinyallerine hem de göz hareketlerine ilişkin veri dağılımlarında, zihinsel iş yükü durumlarına karşılık gelen gizli yapıları ortaya çıkarması açısından önem taşımaktadır. Özellikle gerçek zamanlı veri sağlayan fizyolojik ölçümlere dayanan başarılı tahmin modellerinin varlığı, zihinsel iş yükünün gerçek zamanlı olarak takip edilebilmesi ve bu sayede zihindeki aşırı yüklenmeleri önleyici erken uyarı sistemlerinin oluşturulabilmesi açısından umut vadetmektedir.

Çalışmanın adımları

Belirtilen amaçları gerçekleştirmek üzere yapılan tez çalışmasının bundan sonraki kısmı Şekil 1.1’de belirtildiği şekilde yapılandırılmıştır.



Şekil 1.1. Tez çalışmasının adımları

Bu doğrultuda, ikinci bölümde; tez çalışmasında kullanılan zihinsel iş yükü, işleyen bellek, göz izleme ve beyin görüntüleme ile ilişkili temel kavramlar ve yöntemler hakkında bilgi sunulmuştur.

Tezin üçüncü bölümünde, zihinsel iş yükü ölçümünde EEG ve göz izleme yöntemlerinin kullanıldığı çalışmalar ile makine öğrenmesi algoritmalarının kullanıldığı çalışmalara yer verilmiş ve literatürün mevcut durumu tartışılarak tezin sunacağı katkı vurgulanmıştır.

Dördüncü bölümde; tez kapsamında izlenen yaklaşım, sürecin kolay bir biçimde kavranmasına olanak sağlayan bir akış şeması ile sunulmuştur. Analizin temel verisini elde etmek için gerçekleştirilen uygulama detaylandırılmıştır. Metodolojide kullanılan makine öğrenmesi, sınıflandırma, model başarısı, özellik seçimi gibi kavramlar açıklanmış, bu kavramlar ile ilişkili diğer kavramlar, kullanılan yöntemler ve algoritmalar hakkında bilgi verilmiştir.

Beşinci bölümde pilot çalışmalarda ve temel uygulamada elde edilen istatistiksel bulgular ortaya konmuş, ham EEG verisinin analizinde kullanılan EEGLAB uygulamasının detaylarına yer verilmiş ve elde edilen sonuçlar sunulmuştur. Aynı bölümün devamında; makine öğrenmesi algoritmalarıyla kurulan modellerin sonuçları ile özellik seçimi ve model ayarlama sonrası elde edilen sonuçlara yer verilmiştir. Elde edilen bulgular, önceki çalışmalar ile benzerlikleri ve farklılıkları göz önüne alınarak tartışılmıştır.

Son olarak altıncı bölümde ise, gerçekleştirilen tez çalışması ve ulaşılan temel sonuçlar özetlenmiş, ileriye dönük uygulamalar konusunda öneriler sunulmuştur.

2. TEMEL KAVRAMLAR

Bu bölümde, çalışmanın temelini oluşturan ve tez boyunca pek çok kez tekrarlanacak olan zihinsel iş yükü, n-geri hafıza testleri, NASA-TLX, göz izleme, beyin görüntüleme, EEG, makine öğrenmesi, sınıflandırma, regresyon, model performansı, özellik seçimi gibi kavramlar açıklanacaktır. Literatürde yapılan çalışmaların daha rahat anlaşılabilmesi amacıyla, söz konusu çalışmalar detaylandırılmadan önce yukarıda belirtilen kavramlara ve ilişkili tanımlamalara yer verilmesinde fayda görülmüştür.

2.1. Bilişsel ve Zihinsel İş Yükü

Bilişsel yük (cognitive load) ile zihinsel yük (mental load) kavramlarının birbirleriyle ilişkili kavramlar olduğu aşıkardır. Her ne kadar aralarındaki güçlü ilişki sebebiyle literatürde iç içe kullanılan kavramlar olsa da, bu aşamada aralarındaki farkın ortaya konması önemli görülmektedir. Bilişsel yük üç kategoride ele alınmaktadır. "İçsel bilişsel yük", zihinde işlenmekte olan öğelerin içsel doğası (örneğin görev zorluğu) tarafından sebep olunan yüke karşılık gelmektedir. İş organizasyonu, zaman baskısı ve gürültü gibi dış faktörlerin neden olduğu yük ise "dışsal bilişsel yük" olarak ifade edilmektedir. Üçüncü bilişsel yük kategorisi olan "bağlı bilişsel yük" ise, otomasyon sırasında işleyen belleğe yüklenen yük olarak tanımlanmıştır. Zihinsel yük bilişsel yükün yalnızca görev ve çevresel taleplerden kaynaklanan kısmını göstermektedir. Bu nedenle bir görevin gerektirdiği bilişsel talep olarak tanımlamak mümkündür. Zihinsel yük, zihinsel çaba ve performans, bilişsel yükün ölçülebilir üç boyutunu oluşturmaktadır (Galy ve diğerleri, 2012). Bu doğrultuda zihinsel yükün bilişsel yükün bir alt bileşeni olduğunu söylemek mümkündür.

Zihinsel iş yükü, bir kişinin bir görevi yerine getirirken ortaya koyduğu genel bilişsel çaba olarak tanımlanabilir. Zihinsel iş yükü, halihazırda kullanılabilir zihinsel kaynak kapasitesi ile görevin gerektirdiği zihinsel kaynak miktarı arasında oluşan fark ile ilişkili bir kavramdır. Sweller (1988), ortaya koyduğu bilişsel yük teorisinde, bireyin çalışan bellek kapasitesinin sınırlı olduğunu varsaymakta, çalışanların bilişsel kaynaklarını ellerindeki birincil görevi etkili bir şekilde tamamlamak üzere yönlendirebilmeleri için görev tasarımcılarının dışsal bilişsel yük kaynaklarını azaltmaları gerektiğini öne sürmektedir. Eğer görevin gerektirdiği zihinsel kaynak miktarı, mevcut kapasiteden daha fazla ise aşırı

zihinsel yüklenme ortaya çıkacaktır. (Bommer ve Fendley, 2018). Söz konusu aşırı yüklenme, fazladan zihinsel kaynak ayrılmasını gerektirecektir. Fazladan kaynak talebinin ise işlem verimliliğini azaltabileceği düşünülmektedir (Galy ve diğerleri, 2012). Zihinsel iş yükündeki artış aşırı yüklenme düzeyine ulaştığında, bilgi işlemede gecikme, hatalar ve hatta kazalar ile sonuçlanabileceğinden; görev tasarımı, atama ve performans değerlendirme süreçleri mutlaka zihinsel iş yükü değerlendirmeyi de içermelidir (DiDomenico ve Nussbaum, 2011).

Bu aşamada; bilişsel yükün psikolojik, fizyolojik ve çevresel pek çok faktörden etkilenen bir kavram olması sebebiyle tezin kapsamının bilişsel iş yükünün nispeten ölçülebilir boyutlarını içeren zihinsel iş yükü olarak belirlendiğinin, literatür araştırması sırasında irdelenen çalışmalarda da bilişsel iş yükü olarak ifade edilen kavramla çoğunlukla zihinsel iş yükünün kastedildiğinin belirtilmesinde yarar görülmektedir.

2.2. İşleyen Bellek ve N-Geri Hafıza Testleri

Bellek önceden öğrenilmiş bilgileri tutma, hatırlama ve geri çağırma görev alan mekanizma olarak bilinmektedir. Bellek türleri içerisinde en bilinenleri uzun ve kısa süreli bellektir. İşleyen ya da çalışan bellek (working memory) olarak bilinen bellek türü ise, kısa süreli belleğin işlevi olan bilgiyi belirli bir süre zarfında tutmaya ek olarak o bilgiyi dönüştürmekte, zihinsel sürecin sonundaki hedefe ulaşmak için gerekli bilgi parçacıkları ile bir araya getirerek işlemekte ve yapılandırmaktadır. Beynin görev esnasında fiziksel olarak gözlemlenmesini mümkün kılan gelişmiş teknolojiler sayesinde, işleyen belleğin anatomik olarak konumunun beynin frontal, pariyetal ve temporal loblarında olduğu saptanmıştır. Günlük yaşamda gerçekleştirilen konuşma, anlama, bir metni veya konuşmayı takip etme ve anlamlandırma, dinlerken not alma, karar verme, planlama, hesaplama gibi pek çok becerinin gerçekleştirilmesindeki en temel birim olarak bilinen bu yapının önemi açıktır (Altun ve Çevik, 2012).

Bilişsel yük teorisinin doğrudan işleyen belleğin kapasitesi göz önüne alınarak türetildiği (Norman, 2013) düşünüldüğünde, deneysel çalışmalarda oluşturulmak istenen zihinsel iş yükü manipülasyonu için işleyen belleği hedef alan görev türlerinin seçilmesi şaşırtıcı olmayacaktır. Bu doğrultuda, hem işleyen bellek kapasitesinin güvenilir bir göstergesi olması ve hem de geçerliliğinin pek çok çalışma ile gösterilmiş olması nedeniyle n-geri (n-

back) görevler literatürde sıklıkla kullanılmaktadır (Öztürk, 2018).

N-geri görevler, ilk kez Kirchner (1958) tarafından "hızlı değişen bilgi"nin kısa süreli depolanmasında yaşa bağlı farklılıkların değerlendirilmesi için kullanılmış ve daha sonra Mackworth (1959) tarafından n seviyeleri artırılarak geliştirilmiştir.

N-geri görevlerde bireye farklı depolama birimlerini aktif kılan seri uyaranlar sunulmakta ve bu uyaranlar eğer kendisinden n-adım sonra tekrarlamışsa kullanıcının tepki vermesi beklenmektedir. N-geri görevler görevin tasarımına göre tekli ya da ikili (dual) olarak gerçekleştirilebilmektedir. Literatürde ikili n-geri görevler olarak belirtilen testlerde ikili ifadesi ile kast edilen görevin modalitesidir (ses, renk, konum, şekil, harf vb.). Deney tasarımında uyaranın sadece konum ya da sadece ses özelliklerinin değiştirilmesi söz konusu olduğu gibi, çalışmanın ihtiyacına göre konum ve ses, konum, ses ve renk gibi özelliklerinin değiştirilmesiyle giderek karmaşılaşan deney ortamlarının tasarlanması da söz konusu olabilmektedir. Daha çok çalışan belleğin geliştirilmesi amacıyla ikili n-geri görevlerden yararlanılmaktadır. Çoğunlukla görsel ve işitsel uyaranların birlikte kullanıldığı ikili n-geri görevlere rastlanmaktadır (Altun ve Çevik, 2012; Öztürk, 2018).

N-geri görevler doğası gereği bilginin hem depolama hem de manipülasyonunu (örneğin uyaran bir öncekiyle aynı ama n öncekiyle aynı değil) birleştirmekte, bu yönüyle geleneksel çalışan bellek görevlerinin ötesine geçmektedir. N seviyesinin, söz konusu depolama ve/veya manipülasyon süreçlerinin gerektirdiği zihinsel talebe karşılık gelen “işleyen bellek yükü” ile orantılı olduğu kabul edilmektedir (Öztürk, 2018).

N-geri, zorluk seviyesi ile dorsolateral ve ventrolateral prefrontal korteksi de içeren kortikal aktivasyon arasında güçlü korelasyonlara sahip, iyi karakterize edilmiş bir paradigmadır. N-geri görevler; kodlama, geçici depolama ve tekrar, seri sıra takibi, güncelleme, karşılaştırma ve yanıt süreçleri, işleyen bellek ve dikkat işlevlerini gerektirmektedir (Ayaz ve diğerleri, 2012). Bilginin eş zamanlı depolanması, zihinsel olarak kaydırılması ve manipülasyonunu sağlayan n-geri görevler, Pozitron Emisyon Tomografisi (PET) ve fMRG çalışmalarında kullanılmış ve bu görevlerin dorsolateral prefrontal korteks (alın korteksi) aktivasyonunu gerçekleştirdiği görülmüştür. Bu bölge işleyen belleğin yer aldığı düşünülen alandır. Ayrıca, akıcı zekâyı ölçmede sıklıkla kullanılan nöropsikolojik testlerden biri olan Raven’in matrisler testi ile n-geri görevlerden

elde edilen performans düzeyleri arasında gözlenen korelasyonun, prefrontal nöral aktivasyon düzeylerindeki farklılıklara bağlı olarak oluştuğunu gösteren çalışmalar mevcuttur. Bu sonuçlar da n-geri görevlerin dış geçerliliğini yani genellenebilirliğini sağlamaktadır (Altun ve Çevik, 2012).

Zihinsel iş yükünün sınıflandırılmasına ilişkin çalışmalarda da n-geri görevlerin kullanımına sıklıkla rastlanmaktadır (Grimes, Tan, Hudson, Shenoy ve Rao, 2008; Herff ve diğerleri, 2014; Ke ve diğerleri, 2015; Liu ve diğerleri, 2017; Tjolleng ve diğerleri, 2017). Geçerliliğinin pek çok çalışmayla kanıtlanmış olması sebebiyle, bu tez çalışmasında n-geri hafıza testleri kullanılmıştır. Çalışmada kullanılan n-geri görevler tezin dördüncü kısmı olan “Materyal ve Metodoloji” bölümünde detaylandırılmıştır.

2.3. NASA-TLX

NASA-TLX, algılanan iş yükünü ölçen çok boyutlu öznel bir değerlendirme aracıdır. Bu araç, belirli bir işle ilgili belirli iş yükü kaynaklarının bütünsel bir iş yükü derecesinin hesaplanmasında kullanılabileceği bir yöntem sağlamaktadır (Hart ve Staveland, 1988). NASA-TLX, altı alt gruba karşılık gelen altı sorudan oluşmaktadır (zihinsel talep, fiziksel talep, zamansal talep, performans, çaba, rahatsızlık seviyesi). Zihinsel talep sorusu ile ne ölçüde zihinsel aktiviteye ihtiyaç duyulduğu, fiziksel talep sorusuyla ise bedensel yorgunluk hissedilip hissedilmediği sorgulanmaktadır. Zamansal talep alt boyutu kişinin üzerinde hissettiği zaman baskısını, performans ise görevi gerçekleştirirken duyduğu tatmin ve başarı hissini ölçmeyi amaçlamaktadır. Çaba sorusuyla, görevin gerektirdiği performans seviyesi için harcanan eforun ne düzeyde olduğuna, rahatsızlık seviyesi sorusuyla ise görev esnasında ne ölçüde gergin, rahatsız ve güvensiz hissedildiğine cevap aranmaktadır. Katılımcılardan bu soruların her birine cevap verebilmeleri için sürekli ölçekte 20 aralıktan birini işaretlemeleri istenmektedir. Her aralık 5 puan değerindedir. Ölçekte sayısal bir ifade olmamasına rağmen; hesaplama sürecinde 5 puana karşılık gelen ilk aralık düşük/zayıf anlamındaki yanıtı, 100 puana karşılık gelen son aralık ise yüksek/iyi anlamındaki yanıtı temsil etmektedir. Ölçekte yalnızca performans alt boyutu tersten puanlanmaktadır. NASA-TLX'in küçük iş yükü değişikliklerine karşı yüksek duyarlılığını gösteren çalışmalar mevcuttur (Galy ve diğerleri, 2012). Diğer yandan, son yıllarda NASA-TLX gibi öznel değerlendirme yöntemlerinin tek başına kullanıldığı yöntemlere az rastlanmaktadır. NASA-TLX'in zihinsel iş yükü analizinde güvenilirliğinin kanıtlanmış

olması (Hart, 2006) sebebiyle genellikle yeni önerilen bir yöntemin geçerliliğini ve tercih edilebilirliğini göstermek amacıyla özellikle fizyolojik yöntemlerle bir arada kullanıldığı çalışmalar mevcuttur (Choi ve diğerleri, 2018; Di Stasi, Antoli, Gea, ve Cañas, 2011; Evans ve Fendley, 2017; Heine ve diğerleri, 2017; Jo, Myung, ve Yoon, 2012; Orlandi ve Brooks, 2018; C. Yang, L. Yang, Cheng, Jou ve Chiou, 2012).

Ağırlıklı NASA-TLX yönteminde katılımcılara görevi gerçekleştirirken hangi ölçütü daha yoğun olarak yaşadıkları sorulmakta, iki seçenekten birini seçmeleri beklenmektedir. 15 ikili karşılaştırma neticesinde elde edilen ağırlıklar alt boyut puanlarıyla çarpıldıktan sonra 15'e bölünerek ağırlıklı toplam NASA-TLX skorları elde edilmektedir. Ağırlıklandırılmış TLX'in, öznel iş yükünü daha iyi ayırt etme potansiyeli gösterdiği ortaya konmuştur (Ma, McCrory ve Claudio, 2021).

2.4. Göz İzleme

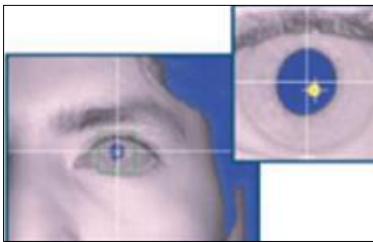
Göz izleme (eye tracking); bir kişinin belli bir zamanda nereye baktığının ve gözünü bir noktadan diğerine nasıl hareket ettirdiğinin araştırmacı tarafından bilinebilmesi amacıyla göz hareketlerinin ölçüldüğü bir tekniktir (Poole ve Ball, 2006). Uygulamalı göz hareketi takibi ve analizi ilk kez 1870'li yıllarda gerçekleştirilmiştir. Başlangıçta kornea ile teması gerektiren müdahaleci (invazif) bir yöntem iken, müdahaleci olmayacak şekilde ilk kez, korneadan yansıyan ışığı kullanan ve gözün yalnızca yatay konumunu ışığın düştüğü bir fotoğraf plakasına kaydeden Dodge ve Cline (1901) tarafından kullanılmıştır. Başlangıçta göz hareketlerinin yalnızca belli bir görsel uyarının hareket, kontrast ve konum özellikleri ile olan ilişkisi incelenmekteydi. Göz hareketlerinin 1930'lu yıllardan itibaren havacılık, otomobil kullanımı, X-ray görüntüleme ve reklamcılık gibi alanlarda gerçekleştirilen performans ve kullanılabilirlik değerlendirmelerinde kullanımı artmakla birlikte, bilişsel süreçler ile olan ilişkisi 1970'lerden itibaren önem kazanmaya başlamıştır. Ancak, verinin toplanmasından öte analizinin çok fazla çaba gerektirmesi sebebiyle çalışmalar belli bir süre veri analizini kolaylaştırma hedeflerine yönelmiştir (Cansız, 2012; Goldberg ve Kotval, 1999; Jacob ve Karn, 2003).

Göz izleme cihazları, bakış koordinatlarını belirlemek ve odaklanma, seğirme gibi göz hareketlerini ölçmek amacıyla kullanılan cihazlardır. Günümüzde, masaüstü bilgisayarlara sabitlenmiş, web kamerası şeklinde bilgisayar monitörüne takılabilen ve gözlük tipinde

kullanılabilen çeşitleri bulunmaktadır.

Müdahaleci olmayan kamera tabanlı göz izleyiciler çoğunlukla gözdeki iki sınır çizgisinden yararlanmaktadır. Biri göz bebeği ile iris arasındaki, diğeri ise iris ve göz akı arasındaki sınırdır. Limbus olarak adlandırılan iris-göz akı sınırı, göz kapaklarından dolayı takibi zorlaştırmaktadır. Göz bebeği-iris sınırı ise limbusa göre daha az kontrastlıdır. Bu nedenle, göz bebeği izleme sistemlerinde gerekli kontrast, görünür ışık yerine kızılötesi veya yakın kızılötesi ışık kaynağı altında sağlanabilmektedir. Görünmeyen yakın kızılötesi ışık kaynakları birçok video kamera tarafından algılanabilmektedir. Işık kaynağı kameraya paralel yerleştirildiğinde, kamera göz bebeğini daha parlak görür. Işık kaynağının kamera ile paralel konumunda izlendiği yöntemle parlak göz bebeği izlemesi, paralel olmayan konumunda izlenmesine ise karanlık göz bebeği izlemesi denilmektedir. Bazı göz izleme sistemleri, daha iyi kontrast sağlamak üzere gözün iki görüntüsünü birleştirmek için kamerayla hem paralel hem de paralel olmayan ışık kaynakları kullanmaktadır (Cansız, 2012).

Günümüzde çoğu ticari göz izleme sistemi kornea-yansıma/göz bebeği-merkez metoduna göre çalışmaktadır. Cihazla gömülü bir kızılötesi kameradan çıkan kızılötesi ışın doğrudan göze gelmekte ve hedef göz metriklerinin kolay takibi için güçlü yansımalarla sebep olmaktadır. Retinaya giren ışığın büyük bir kısmı göz bebeğinin parlak ve iyi görünmesi için geri yansıtılmaktadır. Bu durum literatürde “parlak göz bebeği etkisi” olarak bilinmektedir. Kornea yansıması (corneal reflection) da yine küçük ama keskin görünen kızılötesi ışın tarafından oluşturulmaktadır. Bu sayede göz bebeğinin merkezi ve kornea yansımasının konumu tespit edilmekte, aralarındaki vektör ölçülmekte ve farklı trigonometrik hesaplamalar yoluyla ilgilenilen ilgi noktası bulunabilmektedir (Poole ve Ball, 2006). Resim 2.1’de gözün kameradan görünüşü ve göze yansıyan ışık görülebilmektedir.



Resim 2.1. Gözün kameradan görünüşü (Cansız, 2012)

Video tabanlı göz izleyicilerde bir kalibrasyon süreci ile her bir kişinin göz hareketleri özelliklerine göre ince ayar yapılması gerekmektedir. Kalibrasyon sürecinde ekranda bir nokta görüntülenmekte ve eğer göz belirli bir eşik süresinden daha uzun süre belirlenen alanda sabitleniyorsa kornea-yansıma/göz bebeği-merkez ilişkisi belirli bir (x,y) koordinatına karşılık gelecek şekilde kaydedilmektedir (Poole ve Ball, 2006).

Yumuşak kontakt lensler tipik olarak sınır problemlerine neden olmamaktadır. Ancak sert olanlar, merceğin altında sıkışan kir veya tozdan kaynaklı olarak parlak göz bebeği sistemlerinde sınır sorunlarına neden olabilmektedir. Tipik olarak tek yönlü düzeltici gözlükler (miyop, hipermetrop veya astigmattan sadece biri) yansıma önleyici bir kaplamaya sahip olmadıkça sorun yaratmamaktadırlar. Araştırma yapan kişinin kaliteli göz izleme verileri elde edebilmesi için, göz tembelliği vb. göz rahatsızlığı durumları ile gözlük/kontakt lens kullanımı gibi hususlara dikkat etmesi gerekmektedir (Cansız, 2012).

Göz izlemede anlamlı sonuçların derlenebilmesi için ilgi alanlarının (İA) belirlenmesi kritik önem taşımaktadır. İlgi alanları, bir görüş alanı üzerinde belirlenen, anlamlı bilgi taşıyan ve spesifik bir bilgi kaynağını işaret eden bölgelerdir. Görevi tamamlamak için hangi tür görsel taramanın gerekli olduğuna ve hangi arayüz kullanıldığına bağlı olarak ilgi alanları büyük ölçüde göreve özgüdür (Ellis, 2009). İA, daha çok gözün yöneldiği yerin önem taşıdığı reklamcılık ve pazarlama çalışmalarında kullanılmaktadır.

Bakış, bilişsel ilginin doğrudan ölçüsüdür (Dirican ve Göktürk, 2011). Bu nedenle, zihinsel iş yükü ölçümü için göz tepkilerine ilişkin verilerden yararlanan pek çok çalışma mevcuttur. Söz konusu çalışmaların detaylarına “Literatür Özeti” bölümünde yer verilmiştir.

Zihinsel iş yükünün analizinde göz hareketlerinden yararlanmanın potansiyel bir faydası kısa sürede meydana gelen iş yükü dalgalanmalarını yakalayabilme imkanı sağlamasıdır (Ahlstrom ve Friedman-Berg, 2006). Chen ve Epps (2013), zihinsel iş yükü çalışmalarında göz aktivitelerinden yararlanma nedenlerini 4 maddede toplamaktadır:

- Tek bir sensör ile gözle ilgili farklı bilgi türleri (göz bebeği çapı, göz kırpma, odaklanma ve seğirme) elde edilebilmektedir.
- Göz aktivitesinden her zaman ve her yerde yararlanılabilmektedir.

- Göz bebeği tepkisi ve göz kırpmasının görsel ve işitsel görevlerle ilişkili olduğu gösterilmiştir.
- Göz aktivitesine ilişkin veriler diğer fizyolojik sinyallerin toplanmasına kıyasla daha az görev kesintisine sebep olmaktadır.

Ayrıca göz izleme yöntemi; cihaz kurulumunun hızlı olması, doğru ve yeterli veri elde edilebilmesi ve kalibrasyonu geçersiz kılmadan başın küçük miktarlarda mobilizasyonunu tolere edebilmesi açısından da avantajlıdır (Cansız, 2012).

Göz hareketleri temelde odaklanma ve seğirme hareketlerinden meydana gelmektedir. Odaklanma ve seğirme hareketleri, göz hareketi davranışı mekaniğinin temelini oluşturan ölçülerdir. Bu nedenle göz hareketlerinin incelendiği pek çok çalışmada, bu iki önemli metriğin istatistiksel alt kümeleri hesaplanmakta ve veri setinde toplanmaktadır. Çalışmalarda sıklıkla kullanılan diğer göz izleme metrikleri ise göz bebeği tepkisi ve göz kırpmadır. Zihinsel iş yükü ölçümünde kullanılan temel göz tepkileri aşağıda detaylandırılmıştır.

2.4.1. Odaklanma

Odaklanma (fixation) gözün belli bir noktaya sabitlenmesidir. Bir kişinin görsel odak aralığında mevcut bulunan her türlü bilgiyi toplamak ve yorumlamak için geçirdiği süredir. Bu süre zarfında görsel algı ve bilişsel süreçler gerçekleşmektedir. Dolayısıyla odaklanma yoğun bilişsel süreçler ile alakalı bir göz hareketidir (Ellis, 2009; Pan ve diğerleri, 2004). Odaklanmalardaki değişiklikler sıklıkla belirli bir görevin dinamik talepleriyle ilişkili olmaktadır. Bir odaklanmanın minimum süresinin 80 milisaniye (ms) (Borys ve diğerleri, 2017b; Wu, Liu, Jia, Congchi ve Yan, 2019), 100 ms (Chen ve Epps, 2013; Yan ve diğerleri, 2019) veya 200 ms (Ellis, 2009); maksimum süresinin ise, 900 ms (Wu ve diğerleri, 2019) veya 1200 ms (Yan ve diğerleri, 2019) olabileceğini belirten çalışmalar mevcuttur.

Verimliliğin iş yükü ile negatif ilişkili olduğunun gösterildiği bir çalışmada toplam odaklanma sayısının da verimlilik ile negatif ilişkili olduğu ortaya konmuştur. Toplam odaklanma sayısının görev süresine oranını gösteren odaklanma hızı da tıpkı toplam odaklanma sayısı gibi iş yükü ile pozitif yönlü bir ilişki içindedir. Odaklanma hızı; daha

fazla çaba gerektiren görsel taramanın, düşük performansın ve hafıza görevlerinde daha uzun tarama sürelerinin bir göstergesidir (Ellis, 2009). Odaklanma hızı ölçüsünden en çok okuma görevleri, insan faktörleri çalışmaları ve kullanılabilirlik testlerinde yararlanılmaktadır (Yan ve diğerleri, 2019). Daha uzun süreli odaklanmalar belli bir süre boyunca bilişsel işlem yükündeki artışların göstergesidir. Ortalama ve maksimum değerleri de dâhil olmak üzere odaklanma süresi, uçuş simülatörü görev araştırmasında iş yükünün arttığına işaret etmektedir (Ellis, 2009).

Odaklanma süresinin görsel ilgi alanından bilgi çıkarmanın zorluğuyla, odaklanma hızının ise odaklanılan ögenin önemiyle ilişkili olduğu düşünülmektedir (Chen ve Epps, 2013).

Odaklanma hareketine çok benzeyen sürekli bakış (gaze) tepkisi, tek bir ilgi alanındaki odaklanmaların gruplandırılmasını analiz etmektedir. Bakış metrikleri daha çok ilgi alanına ve neyi temsil ettiğine odaklanmaktadır (Ellis, 2009).

2.4.2. Seğirme

Seğirme (saccade) bir odaklanmadan diğerine gerçekleşen kısa ve hızlı göz hareketidir. Bir seğirme hareketi bir odaklanma hareketini diğerine bağlamaktadır. Diğer bir ifadeyle, odaklanma ve seğirme bir yere bakarken sırayla gerçekleşmektedir. Bir seğirmenin farklı bileşenleri arasında seğirmenin uzunluğu (görsel açı), saniyedeki hızı ve yönü yer almaktadır. Okuma sırasında göz her seferinde birkaç kelimeye odaklanabilmek için bir odaklanmadan diğerine geçen saniyede 4-5 hızlı hareket gerçekleştirebilmektedir. Seğirme sırasında beyne görsel sinyaller iletilmemektedir. Bu nedenle bilgi alınan her bir odaklanmanın devamında seğirme gerçekleşmekte ve yeni bilgilerin gözlemlenebilmesi için başka bir odaklanma hareketi gerekmektedir. Odaklanmalar yalnızca bilgi ile dolu sonlu bir alanı kapsadığından, seğirme hareketleri istenen bilgi alanını takip eder, böylece odaklanmalar beynin genel görüntüyü yorumlaması için gerekli tüm bilgileri toplayabilir. Odaklanmalarla sonuçlanan seğirme hareketleri nesneleri ve sahneleri kodlayan görsel-motor belleğin oluşmasına izin vermektedir (Ellis, 2009). En yaygın kullanılan parametreler seğirme sayısı, seğirme süresi, seğirme hızı (birim zamandaki seğirme sayısı) ve seğirme genliğidir. Seğirme genliği, kesin hedef noktalarını yakalamanın zorluğuna işaret etmektedir (Chen ve Epps, 2013). Daha büyük uzunluktaki seğirme hareketlerinin sıklığının azalan verimlilikle ve algılanan iş yükündeki potansiyel artış ile ilişkili olduğu

gösterilmiştir (Ellis, 2009).

2.4.3. Göz bebeği tepkisi

Göz bebeği, irisin retinaya ışık girmesine izin veren bölümüdür (Benedetto ve diğerleri, 2011). Göz bebeği çapındaki değişimin temel işlevi retinayı korumak (ışık refleksi) ve ayrıca yaklaşan bir nesne için odadaki değişime tepki vermektir (yakın refleksi). Işık refleksi ve yakın refleksi ile karşılaştırıldığında bilişsel aktivitelerdeki varyasyonları nispeten daha az yansıtmasına rağmen, göz bebeği değişimleri bilişsel süreçlerdeki değişikliklerden kaynaklanabilmektedir. Kişinin görsel alanındaki nesneler neredeyse sabitse göz bebeği yanıt spektrumundaki düşük frekans bileşenleri görevle uyarılmış göz bebeği tepkisi (GUGT) olarak düşünülebilir (Chen ve Epps, 2013). Göz bebeği çapındaki küçük ve istemsiz dalgalanmalar bilişsel sürece tepki olarak tanımlanmakta ve GUGT olarak adlandırılmaktadır. Bu tip görevlere örnek olarak; kısa ve uzun süreli hafıza görevleri, zihinsel aritmetik, cümle kurma, tetikte olma, görsel ve işitsel algı görevleri verilebilir. Göz bebeği çapındaki büyük dalgalanmalar ise genellikle ışık kaynaklı değişimler olarak yorumlanmaktadır. Göz bebeğinin farklı aydınlatma koşullarına duyarlı olması sebebiyle pratikte zihinsel iş yükü ölçümü için kullanılması zordur. Bu sebeple daha güvenilir zihinsel iş yükü tahminleri elde etmek için çok fazla GUGT verisinin kayda alınması ve ortalama değerlerin kullanılması gibi yöntemlere başvurulmaktadır (Marquart, Cabrall ve Winter, 2015).

Her ne kadar göz bebeği çapı bir takım taleplere ve duygusal durumlara karşı hassas olması sebebiyle tanısıl açıdan zayıf kalabilse de (Cain, 2007), göz bebeğindeki büyümenin bilgi işleme talebindeki artışın göstergesi olma eğiliminde olduğu söylenebilmektedir (Ellis, 2009). Wilson (2004) göz bebeği çapının iş yükündeki hızlı değişimlere karşı yüksek duyarlılık gösterdiğini ve artan bilişsel süreç seviyelerinde arttığını ortaya koymuştur.

2.4.4. Göz kırpma

Göz kırpma (blink), göz kapağının hızlıca kapanıp açılma hareketidir (Benedetto ve diğerleri, 2011). Göz kırpma fonksiyonel olarak dakikada yalnızca iki-dört defa meydana gelir. İşlevsel olmayan diğer türleri ise: refleksi olarak kırpma (koruyucu bir tepki), bilerek

kırpma (birine yanıt) ve içsel kırpmadır. İçsel yani endojen kırpma bilinçsizce meydana gelir. Göz kırpma davranışlarının çoğunluğu bilinçsizce meydana gelen endojen kırpma şeklindedir. Merkezi olarak kontrol edildiği ve bilişle bağlantısı olduğu için bu tür göz kırpma zihinsel iş yükü ölçümünde kullanılır. En yaygın kullanılan göz kırpma parametreleri göz kırpma hızı (birim zamandaki göz kırpma sayısı), göz kırpma süresi ve görevle ilişkili bilginin gelmesinden göz kırpmaya kadar geçen süre yani göz kırpmadaki gecikmedir. Göz kırpma ile ilişkili bir diğer parametre ise Göz Kapağı Kapanma Yüzdesi (GKKY) olarak adlandırılan ve göz kapağının göz bebeğinin en az %80'ini kapattığı zaman yüzdesini ifade eden ölçüdür (Chen ve Epps, 2013; Marquart ve diğerleri, 2015). Göz kırpma parametreleri, görev talepleri ve yorgunluk seviyesi hakkında anlamlı bilgi vermektedir.

Göz kırpma hızının görev çeşitliliği ne olursa olsun iş yükü ile ilişkili olduğu kanıtlanmıştır. Temel inanç, yüksek iş yükünün, daha fazla odaklanmış dikkat ve görsel yükte genel bir artış gerektirdiğidir. Bu nedenle artan iş yükü ile birlikte daha az göz kırpmasının gerçekleşmesi, bu sayede kritik bilgileri kaçırma olasılığının azalması beklenmektedir. Daha fazla bilgi toplama gereksinimi, gözün açık kaldığı sürenin daha fazla olmasına ve böylece göz kırpma hızının düşmesine neden olmaktadır (Ellis, 2009). Göz kırpma süresi artan iş yükü ile birlikte düşerken, göz kırpmadaki gecikme artan zihinsel taleple birlikte artmaktadır (Castor, 2003).

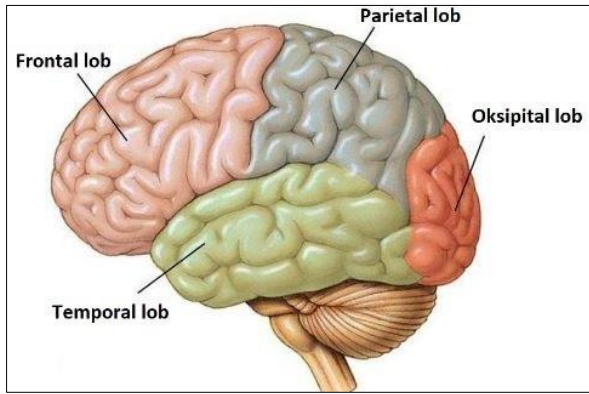
2.5. Beyin Görüntüleme

Bu bölümde; öncelikle beynin yapısı ve sinirsel aktivitenin nasıl gerçekleştiği açıklanacak, daha sonra ise başta EEG olmak üzere beyin görüntüleme tekniklerine değinilecektir.

2.5.1. Beynin yapısı ve sinirsel aktivite

Beyin, insan vücudunun gerçekleştirdiği tüm fonksiyonları kontrol eden, merkezi sinir sisteminin en önemli organıdır. Vücut ağırlığının yaklaşık %2'sini (ortalama 1400 gram) oluşturan beynin yapısında, nöron adı verilen ve beyin ile vücut arasındaki iletişimi sağlayan yaklaşık 100 milyar sinir hücresi bulunmaktadır. Beynin sinirsel aktivitesi doğum öncesi gelişimin 17. ve 23. haftaları arasında başlamaktadır (Bezgin, 2016; Kaya, 2022; Selvi, 2022).

Beyin duyuşal bilgiyi almaktan ve işlemekten, karar vermekten ve eylemleri başlatmaktan sorumludur. Beynin farklı bölgeleri insan vücudunun farklı fonksiyonlarını yönetmektedir. Sağ ve sol yarımküre olmak üzere ikiye ayrılan beyin geliştikçe serebral kortekste (beyin kabuğu) kıvrımlar oluşur ve ceviz görünümlü bir yapıya sahip olur. Bu kıvrımlar “girus” olarak bilinmektedir. Kıvrımlar arasındaki sığ oluklar da “sulkus” adını almaktadır. Sulkusların pozisyonları beyni ana dört loba ayırmaktadır. Beynin ön kısmında frontal lob, arka kısmında oksipital lob, yan kısmında pariyetal lob, şakak kısmında ise temporal loblar bulunmaktadır (Kaya, 2022). Şekil 2.1’de beynin bölümleri gösterilmiştir.



Şekil 2.1. Beyin lobları

Frontal lob

Alnın hemen arkasında yer alan frontal lob beynin en büyük lobudur. Kişiliği ve davranışları belirleyen frontal lobun ön uç bölümünde prefrontal korteks yer almaktadır. Prefrontal kortekste hem davranışların kontrolü hem de konuşma vb. öğrenilmiş motor hareketleri yönetilmektedir. Görevlere sistematik olarak yaklaşma, hedef belirleme, geri bildirimi değerlendirme, davranışları planlama, duygusal tepkiler ve karar verme gibi işlevler bu bölümde gerçekleştirilmektedir. Kısacası frontal lob, düşünme, planlama yapma ve hareket etmeden sorumludur (Bezgin, 2016; Kaya, 2022; Selvi, 2022).

Temporal lob

Temporal lob şakak bölümlerinde bulunmakta, işitme, dil, hafıza ve öğrenme gibi fonksiyonları yönetmektedir. Hem bellek görevi görmek hem de duyguları kontrol etmektedir (Bezgin, 2016; Kaya, 2022; Selvi, 2022).

Pariyetal lob

Pariyetal lob, duyu organlarından gelen bilgileri birleştirmek ve duyu algılama fonksiyonlarını yönetmek ile sorumludur. Görsel ve duyuşsal algılar ile tetiklenen doğal kas ve motor hareketlerinin gerçekleştiği bölümdür (Bezgin, 2016; Kaya, 2022; Selvi, 2022).

Oksipital lob

Beynin arka bölümünde yer alan oksipital lob daha karmaşık bir yapıya sahiptir. Sesin beyinde işlendiği yer olan oksipital lob ayrıca görsel algı ve hafıza ilişkilendirmede de önemli rol oynamaktadır (Bezgin, 2016; Kaya, 2022; Selvi, 2022).

Nöronlar merkezde bir çekirdek içeren bir hücre gövdesinden (soma) oluşmaktadır. Hücre gövdesinin etrafı dendrit adı verilen yapıyla çevrelenmiştir. Bu yapılar nöron hücresine girdi sağlamayı amaçlayan yapılardır. Dendritlerden uzanan akson adı verilen yapılar diğer nöronlara bilgi iletimi için kullanılan terminal butonları (sinapslar) ile son bulmaktadır. Nöronlar birbirlerine bilgi iletimi yaparken hücrelerindeki aksiyon potansiyelinden yararlanırlar. Aksiyon potansiyeli, hücre merkezinden başlayarak akson boyunca akson terminallerine kadar ilerleyen bir voltaj değişimdir (Selvi, 2022).

Nöronlar iyi bir iletkenlik özelliğine sahip olmamakla birlikte iyon değişimi ile elektriksel akımı yaratmakta ve iletmektedirler (Bezgin, 2016). Nöron hücresi içerisindeki iyonlar sadece hücre zarındaki gözenekler vasıtası ile dışarıya geçiş yapabilirler. Nöronların hücre zarında mikron karede bine yakın gözenek bulunmaktadır. Bu gözenekler hücre içi ve dışarısındaki iyon geçiş durumuna göre açılıp kapanabilmektedir. Bu gözenekler hücre içi ve dışı arasında yalnızca sodyum, potasyum, klorür ve kalsiyum iyonlarının geçişine duyarlıdır. Hücre zarı kanalları farklı uyarılma potansiyellerinde farklı iyonların geçişine izin vermektedir (Selvi, 2022).

Sinir hücreleri elektrokimyasal sinyaller üreterek çevreye farklı frekanslarda dalgaların yayılmasını sağlarlar. Bu dalgalar gözle görülemeseler bile bazı elektronik cihazlar tarafından algılanabilmektedirler. Saniyedeki titreşim sayısına (frekans) göre delta, teta, alfa, beta ve gama olarak tanımlanmaktadır (Bezgin, 2016).

2.5.2. Beyin aktivitesini izleme yöntemleri

Beynin fonksiyonel süreçlerini izlemek için kullanılan girişimsel olmayan teknikler MEG, fMRG, fYKS ve EEG'dir. Beyinde meydana gelen fonksiyonel hareketlilikler, büyük çaplı sinirsel ağ etkileşimi sonucu oluşan elektrik, manyetik ve elektromanyetik aktivitelerdir. Fonksiyonel nörogörüntüleme çalışmalarının amacı, verilmiş olan zihinsel çalışmayı kapsayan nöral ağın odak yerinin tespitidir. Fonksiyonel nörogörüntüleme yöntemleri içerisinde yer alan ve zamansal çözünürlüğü en yüksek olan yöntemler MEG ve EEG'dir. Görüntüleme; kafa yüzeyindeki elektrik potansiyellerinin ölçülmesi ile gerçekleştiriliyorsa EEG, manyetik alanların ölçülmesi ile gerçekleşiyorsa MEG olarak isimlendirilmektedir (Bezgin, 2016). Konumsal çözünürlüğü yüksek olan yöntemler ise fMRG ve fYKS'dir (Selvi, 2022).

fMRG

fMRG oksijenli ve oksijensiz kanın manyetik duyarlılığındaki farkın görüntülenmesi ile elde edilmektedir. Beynin aktif olan nöronları faaliyetlerini gerçekleştirmek için oksijene ihtiyaç duymaktadır. Aktif olan nöronlarda oksijen fazlalığı fMRG yardımı ile beyin aktif bölümlerinin görüntülenmesini sağlamaktadır. Büyük boyutlu olması ve yüksek maliyeti sebebiyle tercih edilmemektedir.

MEG

Beyindeki elektriksel nöron aktivitesi manyetik bir alan oluşmasına sebep olmaktadır. Manyetoensefalografi cihazları beyin manyetik aktivitesini izlemeye yönelik geliştirilmiş girişimsel olmayan cihazlardır (Selvi, 2022).

Beyinde doğal olarak oluşan elektrik akımları tarafından üretilen manyetik alanları kaydederek beyin aktivitesini haritalamaktadır. MEG emsalsiz zamansal çözünürlük sunmaktadır. fMRG'nin en iyi ihtimalle birkaç yüz milisaniyelik zamansal çözünürlüğü ile karşılaştırıldığında, MEG olayları milisaniye hassasiyetinde çözebilmektedir (Kaya, 2022). Taşınabilirlik açısından esnek olmaması ve cihaz maliyetleri sebebiyle beyin bilgisayar etkileşimi uygulamalarında kullanımı sınırlıdır.

fYKS

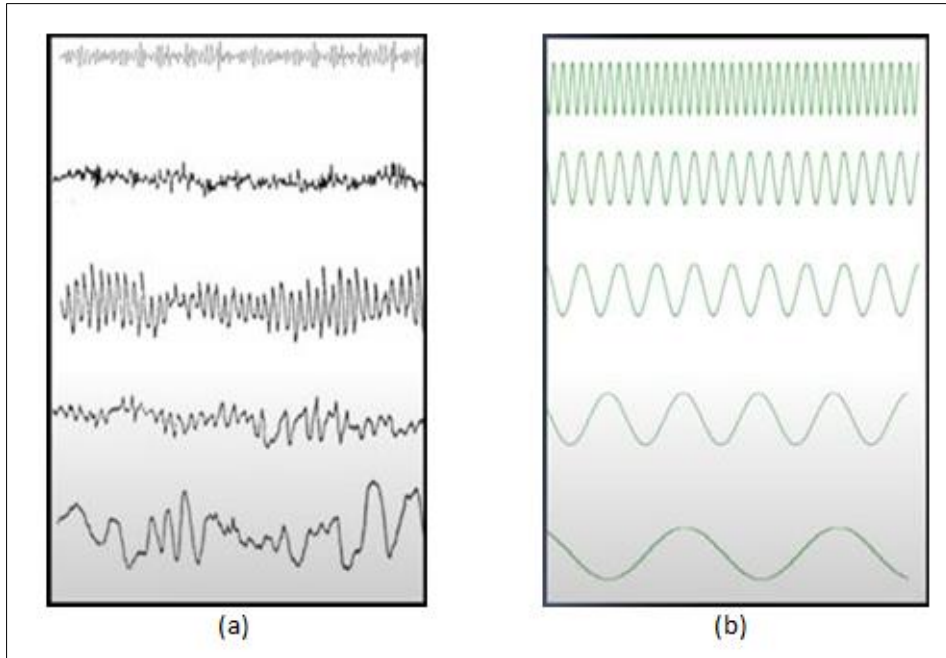
fYKS yöntemi, zihinsel iş yükü ölçümünde kullanılan, prefrontal kortekste kanın oksijenlenmesindeki değişikliklerin izlendiği bir optik beyin izleme teknolojisidir. fYKS yöntemi, beyindeki hemodinamik değişiklikleri fMRG'ye benzer şekilde ölçer, ancak fYKS sessizdir, daha yüksek çözünürlük sağlar, katılımcıları kapalı bir alanla sınırlamaz ve katılımcının uzanmasını gerektirmez. Bu sayede beyin aktivitesine bağlı hemodinamik değişiklikler sadece laboratuvar ortamlarında değil aynı zamanda çalışma koşullarında da izlenebilmektedir. fYKS, fMRG ile karşılaştırıldığında baş hareketlerinin neden olduğu gürültülere karşı daha toleranslıdır. Katılımcıların egzersiz yapmasını gerektiren deneylerde bile kullanılmıştır (Ayaz ve diğerleri, 2012). Son yıllarda zihinsel iş yükü ölçümünde dikkat çeken bir yöntem olmakla birlikte, cihaz maliyetinin yüksek olması sebebiyle yaygın bir kullanıma henüz sahip değildir.

Bir diğer beyin görüntüleme tekniği olan EEG, tezde kullanılan yöntem olması sebebiyle ayrı bir başlık altında incelenmiştir.

2.5.3. EEG

Beyin işlemlerinin tümü çok sayıda hücrenin birlikte çalışması esasına dayanmaktadır. Bu nedenle sinir hücrelerinin toplu etkinliğinin ölçülmesine dayanan yöntemler sinir bilim çalışmalarında sıklıkla kullanılmaktadır. EEG, bilişsel görevler sırasında sıklıkla beyin etkileşiminin bir ölçüsü olarak kabul edilen elektriksel aktivitenin bir temsili olarak kullanılmaktadır. EEG cihazları, kafatasına bağlanan elektrotlar aracılığıyla beynin yaydığı dalgaları ölçümlemeyi sağlamaktadır (Bezgin, 2016). EEG'de her bir elektrot konum olarak denk geldiği kortikal bölgeyi temsil etmektedir. Diğer nörogörüntüleme yöntemlerine kıyasla EEG'nin zamansal çözünürlüğü oldukça yüksektir. EEG; hem gerçek zamanlı nesnel bir değerlendirme sağlar, hem taşınabilir cihazı sayesinde uygulanması kolaydır, hem de müdahaleci olmayan pasif bir ölçüm yöntemi olduğundan uzun süreli ve tekrarlanan kullanım için güvenli bir tekniktir. Yetenekli bir operatör ya da tıbbi bir işlem gerektirmemektedir (Grimes ve diğerleri, 2008). EEG indekslerinin beyin aktivitesi dalgalanmalarına oldukça duyarlı olduğu kanıtlanmıştır. Beyin dalgalarının sonuçları ile detaylı ve çeşitli analizler yapılabilmesi mümkündür (Choi ve diğerleri, 2018).

EEG sinyalleri büyük nöron gruplarının senkronize uyarılmasından gelmektedir. Ritim, nöronların senkron aktivitesinin bir ürünüdür. Hücre uyarımının eş zamanlılığı EEG sinyalinin genliğini ve ritmini belirlemektedir. Sinüs dalgası şeklindeki sinyaller sinüzoid olarak adlandırılmaktadır. Genlik, sinüzoidin y eksenindeki yüksekliğine; faz ise, x eksenindeki kaymasına karşılık gelmektedir. Sinyal ritmik olsa bile sinüzoid olmak zorunda değildir. Ritim, örüntünün düzenli aralıklarla tekrarlandığı anlamına gelmektedir (her saniye davula vurulması gibi). Farklı frekanslarda izlenen idealize edilmiş ve gerçek sinyal görüntüleri karşılaştırmalı olarak Şekil 2.2’de görülmektedir (URL-1). Yukarıdan aşağıya sırasıyla gama, beta, alfa, teta ve delta frekansları verilmiştir.



Şekil 2.2. Sinyal görüntüleri a) gerçek b) idealize edilmiş

EEG sinyalleri çoğunlukla zaman-frekans analizleri veya OİP kullanılarak analiz edilmektedir. OİP’de, trial olarak belirtilen görev denemelerindeki voltaj değişimlerinin zaman eksenli ortalaması alınmaktadır. Ortalama alınması farklı zamanlarda meydana gelen voltaj değişimlerinin yok sayılmasına yani anlamlı sinyalin düzleşerek kaybolmasına ve bilgi kaybına neden olmaktadır. Zaman-frekans analizinde ise bu durumu zaman ekseninde görebilmek mümkündür.

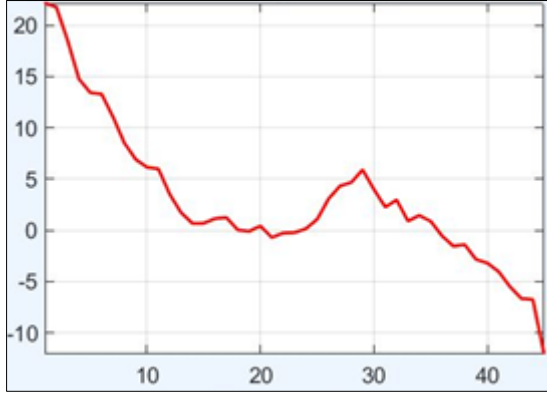
Ritim bazlı analizlerin motivasyonu çoğu EEG özniteliğinin oldukça güçlü bir şekilde ritmik olmasından ileri gelmektedir. Zaman-frekans analizleri ve spektral analizler ritim

bazlı analizlerdir. Zaman-frekans analizinde herhangi bir karmaşık sinyali zaman ekseninden frekans eksenine yaymak için Fourier adı verilen bir dönüşüm işlemi gerçekleştirilmektedir. Fourier, herhangi bir durağan ve sürekli sinyalin farklı genlik ve fazlardaki sonsuz sinüzoid toplamı ile tam olarak tanımlanabileceğini belirtmiştir. Fourier analizinin amacı, toplamı orijinal zaman serisini yeniden oluşturacak şekilde her bir sinüzoid için doğru genliği ve fazı öğrenmektir. Fourier dönüşümünde bir sinyal bulunduğu uzaydaki gösteriminden frekans uzayındaki gösterimine dönüştürülmektedir. Hızlı Fourier Dönüşümü (HFD), Fourier dönüşüm işleminin hızlı hesaplanmasını sağlayan bir algoritmadır (URL-1).

EEG verisi tipik olarak, belirli sinirsel aktivite türlerine karşılık geldiği gözlemlenen frekans bandı setindeki sinyalin spektral gücüne bakılarak analiz edilmektedir. Zihinsel iş yükü analizinde kullanılan EEG değişkenlerinin çoğu da spektral karakteristiklere dayanmaktadır (Borys ve diğerleri, 2017a; Choi ve diğerleri, 2018; Grimes ve diğerleri, 2008). Nitekim spektral EEG ölçülerinin zihinsel yorgunluğun geçerli ve güvenilir göstergeleri olduğu ortaya konmuştur (Lal ve Craig, 2002). En çok kullanılan spektral değişkenler; katılımcının bir beyin bölgesinde belli bir frekans bandındaki salınımların genliği (mikrovolt), güç spektral yoğunluğu ($\text{mikrovolt}^2/\text{hertz}$) ve logaritmik dönüşümü ile elde edilen güç spektral yoğunluğudur (desibel) (Li, Ng, Yu, Yiu ve Lyu, 2023).

EEG spektrumu (tayfı) tüm EEG dalga boylarını içeren bir sınıflandırma biçimidir. Belirlenen farklı frekans aralıklarının genliklerine bakılarak güç spektrumları oluşturulmakta ve analiz edilmektedir. Güç, orijinal sinyal ile bu frekanstaki karmaşık bir sinüzoid arasındaki kovaryansı yansıtmaktadır (URL-1). Diğer bir deyişle, güç spektrumu zaman eksenindeki sinyalin belli bir frekansta sinüs dalgasına ne kadar benzediğinin bir ölçüsüdür. EEG spektrumları, farklı bilişsel aktiviteler sırasında meydana gelen aktivite seviyelerini belirlemek için analiz edilmektedir. Prefrontal bölgeden alınan EEG sinyaline ilişkin gücün, frekans bantlarına göre değişimini gösteren bir güç spektrumu örneği Şekil 2.3'te verilmiştir. X eksen frekansı, y eksen ise gücü göstermektedir.

Kafa derisinin birkaç bölgesinden eş zamanlı kayıtlar alınarak elektriksel aktivite dağılımlarını gösteren topografik haritalar da üretilebilmektedir. Bu haritaların irdelenmesi, beyinde yüksek iş yükü ile ilişkili beyin bölümlerini belirlemeye yarayan elektriksel aktivite örüntülerini açığa çıkarmayı sağlamaktadır.



Şekil 2.3. Prefrontal lobdan alınan EEG sinyaline ilişkin güç spektrumu örneği

Yapılan çalışmalarda EEG ölçümleri beş ayrı frekans bandı doğrultusunda ele alınmaktadır (Borys ve diğerleri 2017a; Choi ve diğerleri, 2018; Kosti ve diğerleri, 2018; Grafton ve Tipper, 2012). Fiziksel ve zihinsel rahatlama halinde yayılan dalgaların saniyede kaç kez salındığını gösteren Hertz (Hz) cinsinden frekans aralıkları (bantlar) Çizelge 2.1’de verilmiştir (URL-2).

Çizelge 2.1. EEG dalgalarının frekans aralıkları

Dalga Tipi	Frekans Bandı
Delta	0-4 Hz
Teta	4-8 Hz
Alfa	8-12 Hz
Düşük Beta	12-16 Hz
Yüksek Beta	16-25 Hz
Gama	25-45 Hz

Delta bandı

0-4 Hz aralığında değişen frekanstaki beyin dalgalarıdır. En büyük genliğe sahip bu dalgalar kendinde olmama ve derin uyku haliyle ilişkilidir (Kosti ve diğerleri, 2018). Boyun hareketleri veya göz kırpması gibi dış gürültülerden (artefakt) kolayca etkilenebilen banttır. Boyun ve çenenin büyük kaslarının neden olduğu artefakt sinyallerini gerçek delta sinyalleriyle karıştırmak çok kolaydır, çünkü kaslar derinin yüzeyine yakın olmakla beraber büyük sinyaller üretmektedir. Oysa ilgili delta sinyalleri beyin derinliklerinde

oluşmakta ve kafatasından geçerken ciddi şekilde zayıflamaktadır. Bebeklerde baskın olan bu ritimlerin uyanık olan yetişkinlerde görülmesi anormaldir. Büyüme hormonunu uyaran bu dalgaların yoğunluğu ilerleyen yaşlarda azalmaktadır (Bezgin, 2016; Kaya, 2022).

Teta bandı

Azalan bilinç, uyuklama hali ve hafıza performansı ile ilişkilidir (Choi ve diğerleri, 2018; Kosti ve diğerleri, 2018). Teta dalgaları hayal kurma durumunu da kapsar. Ayrıca yaratıcılığın geliştirilmesine yardımcı yararları bulunmaktadır. İnsanların kendilerini daha kabul edilebilir hissetmelerini sağlar. Bazı araştırmacılara göre frontal merkez bölgesindeki EEG sinyallerinde gözlenen belirgin teta faaliyeti zihinsel konsantrasyonun yanı sıra kaygıdan uzak meditatif bir durumu yansıtmaktadır (Choi ve diğerleri, 2018). Duygusal gerginlik ve düş kırıklığı durumlarında da ortaya çıkabilmektedir (Bezgin, 2016). Bir teta dalgasına genellikle diğer frekanslar eşlik etmektedir ve uyarılma düzeyiyle ilişkili görünmektedir (Kaya, 2022).

Alfa bandı

Alfa bandı frekansı 8 ile 12 Hz arasında değişmektedir. İlk tanımlanan EEG aktivitesi olup, yetişkin insan beyninin ana ritmidir (Bezgin, 2016). Bilinçli ancak fiziksel ve zihinsel olarak rahat bir duruma karşılık gelir. Bilinçli düşünme ile bilinçaltı arasında bir köprü gibidir (Kosti ve diğerleri, 2018). Bir kişi rahat olduğunda, beyindeki alfa dalgalarının oranı artar. Ayrıca bu dalgalar gerektiğinde insanların sakinleşmelerine yardımcı olmakta ve rahatlamış bir zihni teşvik etmektedirler. Bazı araştırmacılar artan alfa gücünün rahatlamış bir ruh haline ve dışsal uyaranlara karşı gelişmiş otonom tepkilere neden olduğunu öne sürmektedirler (Choi ve diğerleri, 2018). Uyanıklık durumunda baskın olan alfa dalgaları başın arka yarısında ve genellikle beyin oksipital bölgesinde bulunmaktadır. Alfa dalgası, tüm beyin aktivitesi alanındaki en belirgin ritimdir. Çoğu denek gözleri kapalıyken alfa dalgaları üretir, bu yüzden bunun beyin görsel bölgeleri tarafından daha önce üretilen bir örüntünün beklenmesi veya taranmasından başka bir şey olmadığı iddia edilmiştir. Bu durumlarda oluşan alfa dalgaları gözlerin açılması, duygusal uyaranların alınmaya başlanması, yabancı seslerin duyulması, kaygı durumu, zihinsel odaklanma veya dikkat ile azalmakta veya ortadan kalkmaktadır (Kaya, 2022). EEG tabanlı bir iş yükü endeksi önermek için yapılan çalışmada alfa dalgaları rahatlama durumu ile ilgili bir

değişken olarak değerlendirilmiştir (Choi ve diğerleri, 2018).

Beta bandı

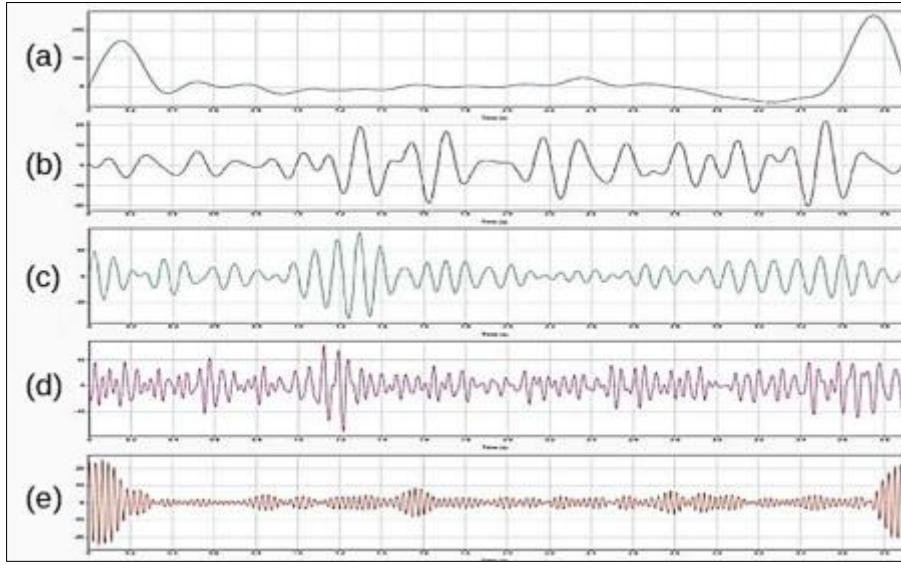
Beta bandı frekansı 12 ile 25 Hz arasında değişmektedir. 12-16 Hz arası düşük beta, 16-25 Hz arası yüksek beta olarak değerlendirilmektedir. Bu bant; aktif dikkat, konuşma, yargılama, karar verme, yüksek konsantrasyon ile ilişkilendirilir. Diğer bir deyişle aktif bir zihin ve endişeli düşüncelerle yakından ilişkilidir. Bilinçli ve aynı zamanda mantıklı düşünmeyi içerir. Diğer yandan, çok fazla beta dalgasına maruz kalmak aşırı strese ve/veya kaygıya yol açabilir. Bu durum karmaşık problemlerin çözümü gibi yoğun zihinsel aktiviteden kaynaklanmış olabilir. Dolayısıyla, beta bandı dalgaları yüksek konsantrasyon ve stres ile ilgili bir değişken olarak kabul edilmektedir (Choi ve diğerleri, 2018). Genellikle frontal ve parietal loblarda görülmektedir (Kaya, 2022). Doğru miktarda beta dalgası odaklanmaya ve okul veya işle ilgili görevleri kolayca tamamlamaya imkan sağlamaktadır (Kosti ve diğerleri, 2018).

Gama bandı

Gama bandı frekansı 25 ile 45 Hz arasında değişmektedir. Yüksek düzeyde bilişsel işlemi kapsar. Gama dalgaları öğrenme, hafıza ve bilgi işleme açısından önemlidir. EEG'deki gama bandı aktivitesi üst düzey bilgi işlem, görsel bilgi işlem, ve dikkate dayalı algısal mekanizmalar ile yakından ilişkilidir (Choi ve diğerleri, 2018; Kosti ve diğerleri, 2018). Dolayısıyla, gama bandı dalgaları yüksek zihinsel iş yükü ve uyarılma ile ilgili bir değişken olarak kabul edilmektedir.

Özetle; en yavaş kortikal ritimlerin aktif olmayan bir beyinle, en hızlı kortikal ritimlerin ise bilgi işlemeyle ilgili olduğu söylenebilmektedir. Farklı bantlardaki EEG dalgaları Şekil 2.4'te verilmiştir (Kaya, 2022).

Özellikle alfa ve teta aktivitesinin zihinsel iş yükü düzeylerini ayırmada etkili olduğu doğrulanmıştır (Zammouri, Moussa ve Mebrouk, 2018). Ayrıca, parietal ve oksipital beta gücünün zihinsel iş yükündeki değişimle ilişkili olduğunu (Yin ve Zhang, 2017) ve değişen görev performansına yanıt olarak arka korteksteki beta ve gama bantlarında güç artışı izlendiğini gösteren çalışmalar mevcuttur (Chuang ve diğerleri, 2012).



Şekil 2.4. Farklı bantlardaki EEG dalgaları a) Delta, b) Teta, c) Alfa, d) Beta e) Gama

Zihinsel iş yükü ölçümünde EEG yönteminin kullanıldığı çalışmalara tezin üçüncü kısmı olan “Literatür Araştırması” bölümünde yer verilmiştir.

3. LİTERATÜR ARAŞTIRMASI

Tez çalışmasının bu bölümünde; öncelikle göz izleme ve EEG sinyalleri aracılığıyla elde edilen değişkenlerin zihinsel iş yükü ile ilişkisi olup olmadığını ve varsa ilişkinin yönünü istatistiksel olarak inceleyen çalışmalara yer verilmiştir. Dolayısıyla 3.1. ve 3.2. bölümlerindeki amaç, yapılan çalışmalarda zihinsel iş yükünün analizinde hangi göz izleme ve EEG değişkenlerinin kullanıldığını ve zihinsel iş yükü ile söz konusu değişkenler arasında istatistiksel olarak anlamlı bulunan ilişkinin yönünü vurgulamaktır. Sonrasında ise tezin bir diğer temel amacı doğrultusunda, makine öğrenmesi algoritmalarından yararlanarak zihinsel iş yükünü tahmin eden modeller ortaya koyan çalışmalar sunulmuştur. Bu çalışmalarda ise; hangi fizyolojik yöntemler ile veri toplandığına, sınıflandırma probleminin kaç sınıflı olarak ele alındığına, kullanılan makine öğrenmesi algoritmalarının neler olduğuna ve model başarı düzeyine odaklanılmıştır. Son bölümde verilen özet tabloda makine öğrenmesine ilişkin çalışmalar ve model başarıları özetlenmiştir. Son kısımda ise; literatürün mevcut durumu analiz edilmiş, problemin tanımı ve tezin sağlayacağı katkı ortaya konmuştur. Göz izleme ve EEG değişkenleri ile zihinsel iş yükü arasındaki ilişkinin yönüne odaklanan literatür araştırmasının sonuçlarına ilişkin özet tablolar, “5. Araştırma Bulguları ve Tartışma” bölümünde bu tezin sonuçları ile karşılaştırmalı olarak verilecektir.

3.1. Göz İzleme Değişkenleri ile Zihinsel İş Yükü Arasındaki İlişkinin Yönünü Araştıran Çalışmalar

1970’lerden 2000’lerin başlarına kadar yapılan çalışmalarda genellikle göz bebeği boyutu ve göz kırpmasının zihinsel iş yükü ile ilişkisine odaklanılmıştır. Juris ve Velden (1977) tarafından yapılan çalışmada dört alternatifli zorunlu seçim görevini gerçekleştiren deneklerin göz bebeklerinde artan iş yükü ile birlikte büyüme gözlenmiştir. Uçuş sırasında iletişim görevinin gerçekleştirildiği bir diğer çalışmada, deneklerin göz bebeklerinin boyutunun artan iş yükü ile birlikte arttığı, göz kırpma hızlarında ise anlamlı bir değişim olmadığı tespit edilmiştir (Casali ve Wierwille, 1983). Hava trafik kontrolörlerinin yüksek ve düşük iş yükü altında incelendiği bir çalışmada ise, iş yükündeki artışın göz kırpma hızıyla negatif ilişkili olduğu gösterilmiştir (Brookings, Wilson ve Swain, 1996). Aynı

çalışmada, göz kırpmadaki gecikme olarak da ifade edilen göz kırpma aralığı arasındaki sürenin iş yükü ile pozitif ilişkili olduğu gösterilmiştir.

2000'li yıllarda da göz bebeği ve göz kırpmada davranışının zihinsel iş yükü ile ilişkisi merak konusu olmaya devam etmiştir. Recarte ve Nunes (2000) sürüş sırasında görsel ve işitsel görevler gerçekleştirilirken meydana gelen zihinsel iş yükündeki artışın göz bebeği boyutu üzerinde önemli etkisi olduğunu ortaya koymuşlardır. Beatty ve Lucero-Wagoner (2000), göz bebeğindeki büyümenin çeşitli bilişsel görevlerin zorluk seviyelerini ayırtmada başarılı olduğunu göstermişlerdir. Göz bebeği boyutunun görev zorluk seviyesine karşı duyarlı olduğu Iqbal, Zheng ve Bailey (2004) tarafından da doğrulanmıştır. Göz izleme yönteminin kullanıldığı bir diğer çalışmada ise, göz bebeğinin artan büyüme sıklığının görev süresince ortaya çıkan zihinsel iş yükü ile ilişkili olduğu ortaya konmuştur (Hampson, Opris ve Deadwyler, 2010). Gao, Wang, Song, Li, ve Dong (2013) tarafından yapılan çalışmada, bir nükleer güç tesisindeki karmaşık acil eylem uygulamaları esnasında göz bebeği boyutunun düşük karmaşıklıkta görevdekine nazaran daha büyük olduğu tespit edilmiştir. Reiner ve Gelfeld (2014) iş yükünü tahmin etmek için üç boyutlu sanal gerçeklik ortamında göz bebeği değişikliklerini izlemişlerdir. Tekrarlayan görevlerin uygulandığı çalışmada sonuçlar zihinsel iş yükünün birkaç görev tekrarıyla azaldığını göstermektedir. Naskrent, Grzywinski, Polowy, Tomczak ve Jelonek (2022) kerestecilik sektöründe biçerdöver operatörlerinin seçirme ve göz bebeği hareketlerini kayıt altına alarak zihinsel iş yükünü inceledikleri çalışmalarında, ağaç kesimi ve ağaç işleme gibi operasyonların gündüz ve gece uygulamaları arasında fark olup olmadığını araştırmışlardır. Aralarında çok fark olmamakla birlikte zihinsel zorlanmayla birlikte göz bebeği çapının arttığı doğrulanmıştır. Radhakrishnan (2023) ve diğerleri, sürüş simülatörü tabanlı çalışmalarında, farklı araba takip senaryolarının ve devralma durumlarının sürücülerin zihinsel iş yükü üzerindeki etkisini incelemişlerdir. Göz bebeği çapının zihinsel iş yükündeki değişime duyarlı olduğu ve öznel değerlendirme sonuçları ile benzer eğilim gösterdiği sonucuna ulaşılmıştır. Ayrıca göz bebeği çapının standart sapmasına bakılarak elde edilen sonuçlar, sürücülerin zihinsel iş yükü seviyelerinin otomatik araç takibi sırasında manuel araç takibine kıyasla önemli ölçüde daha fazla dalgalandığını göstermiştir.

Göz bebeği boyutunun zihinsel iş yükü ile birlikte arttığına yönünde elde edilen benzer bulguların aksine, göz kırpmada hareketinin sayı, hız ve süre gibi bileşenlerinin zihinsel iş

yükü ile ilişkisinin yönü açısından farklı sonuçların elde edildiği çalışmalar dikkati çekmektedir. Kaneko ve Sakamoto (2001) tarafından yapılan çalışmada algılanan görsel yorgunluk ile göz kırpma hızı arasında pozitif bir ilişki olduğu ortaya konmuştur. Söz konusu bulgunun aksine, hava trafik kontrolü ve uçuş görevlerinde artan iş yükü ile birlikte göz kırpma hızı ve gözün kapalı kalma süresinin azaldığını gösteren çalışmalar mevcuttur. Bu farklılığın görev türünden kaynaklanabileceği düşünülmektedir. Görsel uyaran ile tetiklenen ve görsel talep gerektiren iş yükü söz konusu olduğunda göz kırpma hızının azaldığını söylemek mümkündür. Özellikle uçuş görevlerinin görsel olarak daha zorlayıcı bölümlerinde zihinsel iş yükü arttıkça göz kırpma hızının anlamlı ölçüde düştüğü ortaya konmuştur (Wanyan, Zhuang, Lin, Xiao ve Song, 2018; Wilson, 2002). Ancak bazı durumlarda ekstra iş yükü eklenmesi göz kırpma hızında artışa sebep olabilmektedir. Örneğin bir uçuş kontrol görevine eş zamanlı bir hafıza görevi eklendiğinde göz kırpma sıklığının yalnızca uçuş görevinin gerçekleştirildiği koşula kıyasla arttığı görülmüştür (Wilson ve diğerleri, 2004). Castor (2003) tarafından ise, göz kırpma hızı ile iş yükü arasındaki ilişkinin zayıf olduğu ancak göz kırpma sürelerinin artan iş yükü ile azaldığı öne sürülmüştür. Yapılan çalışmalar neticesinde, göz kırpma hızından ziyade göz kırpma sürelerinin daha duyarlı ve güvenilir ölçüler olduğu söylenebilmektedir (Benedetto ve diğerleri, 2011). Benedetto ve diğerleri (2011) araç içi bilgi sistemlerinin göz kırpma davranışı üzerindeki etkilerini incelemek amacıyla tekli ve ikili sürüş simülasyon görevi koşulunda göz kırpma hareketlerini karşılaştırmışlardır. Araç içi bilgi sistemleri ile etkileşimin olduğu ikili görev koşulunda daha kısa süreli göz kırpmalar görülürken, sürüş görevi süresi uzadıkça daha uzun süreli göz kırpmaların olduğu ortaya çıkmıştır. Gao ve diğerleri (2013) tarafından yapılan çalışmada, göz kırpma hızının görev zorluk seviyesindeki değişime duyarlı olduğu ve göz kırpma süresinin de hem düşük hem de yüksek zorluk seviyesinde görev periyodu boyunca arttığı gözlenmiştir. Bu çalışmalar, görev süresi uzadıkça göz kırpma sürelerinin de uzadığını göstermektedir. John ve diğerleri (2022) tarafından yapılan çalışmada, hava trafik kontrolü operatörleri için tasarlanan üç farklı zorluk seviyesindeki izleme ve çarpışma tahmini görevleri sırasında 24 katılımcının EEG ve KHD ile birlikte göz izleme kayıtları toplanmıştır. Göz izleme ve zorluk seviyeleri arasındaki ilişki incelendiğinde, göz bebeği çapının her iki görev türünde de zorluk seviyesi arttıkça arttığı, göz kırpma sayısının ise görev zorlaştıkça azaldığı görülmüştür.

Göz bebeği boyutu ve göz kırpmaya davranışına ek olarak, odaklanma ve seğirme göz hareketlerinin zihinsel iş yükü ile ilişkisinin irdelendiği çalışmalar da mevcuttur. Matematiksel hesaplama görevi verilen bir çalışmada, katılımcıların göz bebeği büyüklüğü ve göz kırpmaya hızı görev zorluk seviyesi ile birlikte artmış, seğirme hızı ve seğirme uzunluğu ise azalmıştır. Aynı çalışmada görsel hedef sıklığı arttıkça göz kırpmaya süresinin azaldığı gözlenmiştir. Çalışmanın sonuçlarına göre, okülomotor yani göz küresi hareketine ilişkin göstergeler hem zorluk seviyesinden hem de göz hareketlerinin sıklığından etkilenmektedir (Nakayama, Takahashi ve Shimizu, 2002). Odaklanma hareketinin incelendiği bir görsel uzamsal bellek görevi sırasında alınan ölçümler doğrultusunda, hedef yoğunluğu arttıkça odaklanma sıklığının arttığı ve odaklanma sıklığı ile zihinsel iş yükü arasında pozitif ilişki olduğu ortaya konmuştur (Van Orden, Limbert, Makeig ve Jung, 2001). Pan ve diğerleri (2004) tarafından yapılan çalışmada, web sayfası arayüzü ile ilgili verilen bir görevde seğirme hızının; görev zorluğu, zihinsel iş yükü ve yorgunluk ile negatif ilişkili olduğu gösterilmiştir. Görev kesintisine sebep olmaması nedeniyle göz davranışlarının sürekli iş yükü göstergeleri olarak kullanılabileceğinin vurgulandığı bir diğer çalışmada, odaklanma sayısı ve gözün kapalı kalma süresinin görev zorluğu arttıkça arttığı görülmüştür (Huang, Xu, Hildebrandt ve Lau, 2019).

Buchholz ve Kopp (2020) tarafından yapılan çalışmada ise, öncelikle zihinsel iş yükü öznel değerlendirme skorlarının görev zorluk seviyesi arttıkça arttığı gösterilmiştir. Bu sonuç yapılan çalışmada zihinsel iş yükünün tetiklendiğini doğrulamıştır. Yapılan ek analizler; artan görev zorluğu ile birlikte, ortalama odaklanma sayısı, ortalama odaklanma süresi ve ortalama seğirme süresinde anlamlı ölçüde düşüş olduğunu göstermiştir. Odaklanma sayı ve sürelerindeki bu düşüş literatürdeki diğer bulguların tersi niteliğindedir. Ayrıca görev zorluğu arttıkça, odaklanma ve seğirme hızları ile ortalama seğirme sayısının anlamlı ölçüde arttığı ortaya konmuştur. Seğirme hızındaki bu artışın da literatürdeki bazı çalışmalar ile çeliştiği söylenebilmektedir. Neticede, görev zorluğu ile odaklanma sayısı negatif ilişkili iken, seğirme sayısı pozitif ilişkili bulunmuştur. Ek analizler ayrıca göz bebeği boyutunun zor görevden orta ve düşük zorluktaki görevlere doğru anlamlı ölçüde azaldığını doğrulamıştır. Düşük ve orta zorluktaki görevler arasında ise göz bebeği boyutu açısından anlamlı bir fark bulunamamıştır.

Göz izleme yönteminin hem performans ölçüleri, hem öznel yöntemler hem de diğer fizyolojik yöntemler ile birlikte kullanıldığı çalışmalara son yıllarda sıklıkla

rastlanmaktadır. Marchitto, Benedetto, Baccino ve Cañas (2016), hava trafik çakışmalarını fark etme görevinin zihinsel iş yüküne etkisini, kayıt altına alınan göz hareketleri (ortalama odaklanma süresi, toplam odaklanma sayısı vs.) aracılığıyla araştırmışlardır. Trafik çakışması olmadığı zamanla kıyaslandığında çakışma olduğu zaman görev karmaşıklığı artmış ve bunun zihinsel iş yüküne yansıdığı gözlenmiştir. Bahsi geçen çalışmada, zihinsel iş yükü ölçümü çok boyutlu ele alınarak hem performans ölçüsü (ortalama hata oranı ile tepki zamanı), hem sözlü öznel yöntem, hem de gözle ilgili fizyolojik verilerden yararlanılmıştır. Orlandi ve Brooks (2018) fizyolojik ölçülerden; göz bebeği çapı, EKG ve EEG'yi, öznel değerlendirme tekniklerinden ise bir Likert ölçeği ile NASA-TLX'i kullandıkları çalışmalarında elde ettikleri fizyolojik verileri öznel verilerle karşılaştırmışlardır. Sonuçta, Likert ölçeği sonuçları ile kalp atış hızı ve göz bebeği çapı arasındaki düşük korelasyon dikkati çekmiştir.

Göz değişkenlerinin NASA-TLX ve alt boyutlarıyla ilişkisine bakılarak zihinsel iş yükünün analiz edildiği çalışmalar olduğu gibi, her iki yöntemin görev zorluk seviyesi ile ilişkisinin araştırıldığı çalışmalar da mevcuttur. Di Stasi ve diğerleri (2011) zihinsel iş yükünü çok boyutlu olarak değerlendirerek öznel ve performansa dayalı teknikler ile ve göz izleme tekniğini birlikte kullandıkları çalışmalarında alışveriş amaçlı iki farklı web tabanlı arama görevinde katılımcıları karşılaştırmışlardır. Sonuçta hedef odaklı grubun, satın alacağı şeyi araştırarak alacak olan araştırmacı gruba oranla; daha kısa sürede, daha kısa süreli odaklanma, daha küçük göz bebeği boyutu ve daha az seğirme davranışıyla görevi tamamladıkları ve araştırmacı gruptan daha düşük NASA-TLX skorlarına sahip oldukları görülmüştür. Rauffet, Chauvin, Nistico, Judas ve Norbert (2016) tarafından yapılan çalışmada, bir denizaltı görevinde Konfigürasyon A ve Konfigürasyon B olarak belirtilen iki farklı zorluk seviyesinde göz bebeği boyutunun görev zorlaştıkça arttığı, NASA-TLX skorunun da zor olan Konfigürasyon B görevinde daha yüksek ölçüldüğü görülmüştür. Yan, Tran, Chen, Tan ve Habiyaemye (2017) nükleer güç tesisindeki kontrol arayüzü ile yeniden tasarlanan arayüz arasındaki farkı araştırmak üzere NASA-TLX ve göz parametrelerini incelemişlerdir. Yeniden tasarlanan arayüz ile birlikte hem algılanan iş yükü, hem hata oranı, hem de görevi tamamlama süresi azalmıştır. Bununla birlikte göz kırpma hızı artmış, toplam odaklanma süresi ve odaklanma hızı ise azalmıştır. Söz konusu çalışmada elde edilen sonuçlar ışığında; görev zorluğu ile göz kırpma hızının pozitif, toplam odaklanma süresi ve odaklanma hızının ise negatif ilişkili olduğunu söylemek mümkündür.

Tran, Yan, Habiyaemye ve Wei (2017), zihinsel iş yüküne dayalı performans tahmini elde etmeyi amaçlayan çalışmalarında dördü gözle alakalı olan altı adet fizyolojik ölçüt ile birlikte NASA-TLX'i kullanmışlardır. NASA-TLX skorunun; göz bebeği boyutu, göz kırpma süresi ve odaklanma süresi ile pozitif, göz kırpma hızı ile ise negatif korelasyonlu olduğu bulunmuştur. Evans ve Fendley (2017) otomasyon seviyeleriyle bilişsel iş yükünün fizyolojik etkileri arasındaki ilişkiyi incelemişlerdir. Çalışma neticesinde NASA-TLX skorları, görevin tamamlanma süresi ve odaklanma hızı ile otomasyon seviyeleri arasında anlamlı bir ilişki olduğu ortaya çıkmıştır.

Das, Maiti ve Krishna (2020) gözle alakalı değişkenlerin NASA-TLX'in alt boyutları ilişkisini araştırdıkları çalışmalarında odaklanma sıklığının zihinsel talep, zaman baskısı ve toplam iş yükü ile anlamlı ölçüde pozitif yönlü ilişkili olduğunu ortaya koymuşlardır. Korelasyon katsayıları sırasıyla 0,24; 0,29 ve 0,30 olarak bulunmuştur. Odaklanma süresi ise, zihinsel talep, zaman baskısı, performans ve toplam iş yükü ile pozitif ilişkilidir (r : 0,31; 0,32; 0,26; 0,30). Odaklanma süresi ayrıca bir performans ölçüsü olan görevi tamamlama süresi ile 0,25 korelasyon katsayısı ile anlamlı ölçüde ilişkili bulunmuştur. Aynı çalışmada, seğirme süresinin zihinsel talep, zaman baskısı ve toplam iş yükü ile negatif yönlü ilişkili olduğu bulunmuştur (r : -0,32; -0,28; -0,35). Odaklanma/seğirme oranının ise, zaman baskısı ve toplam iş yükü arttıkça arttığı gözlenmiştir.

Yapılan çalışmalar doğrultusunda, odaklanma sürelerinin zihinsel iş yükü ile birlikte arttığı kanısı yaygın olmakla birlikte, görev zorlaştıkça odaklanma sürelerinde düşüş olduğunu gösteren çalışmalar da mevcuttur. Borys ve diğerleri (2017b); ortalama odaklanma süresi, maksimum odaklanma süresi ve odaklanma süresinin standart sapması değişkenleri ile görev zorluk seviyesi arasında negatif korelasyon olduğunu tespit etmişlerdir. Korelasyon katsayıları sırasıyla -0,27; -0,24 ve -0,27 olarak bulunmuştur.

Kaczorowska, Plechawska-Wojcik ve Tokovarov (2021) tarafından gerçekleştirilen ve üç zorluk seviyesinden oluşan rakam sembol eşleştirme testinin uygulandığı çalışmada, ortalama seğirme genliği ile maksimum seğirme süresinin zorluk seviyesi arttıkça arttığı, odaklanma süresinin ise görev zorlaştıkça azaldığı gösterilmiştir. Belirlenen zorluk seviyelerinden orta ve yüksek zorluktaki görevler arasında sadece maksimum seğirme genliği açısından anlamlı fark bulunmuştur.

Literatürde seğirme genliği değişkeninin zihinsel iş yükündeki dalgalanmalara duyarlı olduğu konusunda uzlaşmakla birlikte, aralarındaki ilişkinin yönüyle ilgili elde edilen bulguların birbiriyle çeliştiği gözlenmektedir. Ahlstrom ve Friedman-Berg (2006) tarafından, karmaşık hava trafik kontrolü operasyonları sırasında seğirme genliğinin anlamlı ölçüde azaldığı raporlanmıştır. Bir başka çalışmada ise zaman baskısı ve toplam iş yükü arttıkça seğirme genliğinin arttığı sonucu ortaya çıkmıştır (Das ve diğerleri, 2020). Borys ve diğerleri (2017b) tarafından yapılan çalışmada da ortalama ve maksimum seğirme genliği değişkenlerinin görev zorluk derecesi ile pozitif ilişkili olduğu görülmüştür.

Son yıllarda göz hareketlerine ilişkin farklı değişkenlerin de zihinsel iş yüküne duyarlılığı araştırılmaya başlanmıştır. Bu doğrultuda, göz kapağının göz bebeğinin en az %80'ini kapattığı zaman yüzdesini ifade eden bir ölçü olan GK KY değişkeninden zihinsel iş yükü ile ilgili çalışmalarda yararlanıldığı görülmektedir. Han, Zhou, Zong ve Teng (2023) tarafından, GK KY ile yorgunluk seviyesi arasında pozitif yönlü bir ilişki olduğu ortaya konmuştur.

3.2. EEG Değişkenleri ile Zihinsel İş Yükü Arasındaki İlişkinin Yönünü Araştıran Çalışmalar

İnsan beyni işaretlerinin gözlem geçmişi 1900'lü yılların başına dayanmaktadır. Teknolojik gelişmelerle birlikte 1980'li yıllarda kayıt altına alınması mümkün hale gelen EEG sinyalleri, 1990'lı yıllarda dijital ortamlarda gerçek zamanlı izleme ve değerlendirme süreçlerinde kullanılmaya başlanmıştır. Başlangıçta genellikle görsel uyaranlara karşı verilen tepkiler ve söz konusu uyaranlara karşı beynin hangi bölümlerinin harekete geçtiği incelenmiştir. Son yıllarda ise, insanların yaşamlarını kolaylaştırmak üzere bilgisayar ile etkileşimlerini arttırmak büyük ölçüde önem kazanmıştır. Yapılan çalışmalar, insan düşüncesinin bilgisayar ortamına aktarılabilmesi hususunu merak konusu haline getirmiştir. Bu motivasyon, EEG cihazlarının tıbbi klinik araştırmalardan ziyade beyin bilgisayar arayüzü uygulamalarında kullanılmaya başlanmasına sebep olmuştur (Selvi, 2022).

Bu kapsamda yapılan ilk çalışmalarda; EEG sinyallerinin uyanıklık, dikkat ve performansla ilişkisini gösteren çalışmalara odaklanılmıştır. Özellikle teta bandında artan

EEG güç spektrumları ve alfa gücündeki değişiklik ile uyanıklıktaki azalma ve performanstaki bozulmanın ilişkili olduğu gösterilmiştir (Davies, 1965; Gale, Davies ve Smallbone, 1977; Morrel, 1966). 1980'li ve 1990'lı yıllarda da EEG'nin dikkatteki değişkenliğe duyarlı olduğunu ve zihinsel talep gerektiren uzun süreli işlerin ardından performanstaki düşüşü öngörebildiğini gösteren çalışmalar yapılmıştır (Gevins, Bressler, Cuttillo, Illes ve Fowler-White, 1990; Matousek ve Petersen, 1983). Uyanıklık ve dikkatteki azalmanın EEG sinyalinde yine özellikle teta ve alfa dalgalarında kesin işaretler oluşturabildiği ortaya konmuştur (Makeig, 1993; Makeig ve Jung, 1995). Ayrıca Okogbaa, Shell ve Filipusic (1994), beta bandındaki güç artışının, artan uyanıklık ve uyarılma ile ilişkili olduğuna işaret etmiştir. EEG sinyallerinin görev zorluğu ve yorgunluk kavramlarıyla ilişkisine 1990'lı yılların sonlarında başladığı görülmektedir. Kirov, Warsawskaya ve Voynov (1996) tarafından, alfa gücünün dikkat gerektiren bir görevde, teta gücünün ise daha zorlu aritmetik işlemler boyunca arttığı gösterilmiştir. Klimesh (1999) ise, artan yorgunluk ile birlikte beyindeki alfa aktivitesinin oldukça azaldığını öne sürmüştür.

Özellikle 2000'li yılların başından itibaren EEG dalgalarının artan zihinsel iş yükü ile olan ilişkisi daha yaygın bir şekilde incelenmeye başlanmıştır. Yapılan çalışmalarda zihinsel iş yükü ile ilişki genellikle teta, alfa, beta ve gama dalgalarındaki değişim göz önüne alınarak incelenmektedir. Delta dalgalarının kendinde olmama ve derin uyku haliyle ilişkili olması sebebiyle, görevle tetiklenen zihinsel iş yükünün analiz edildiği çalışmalarda düşük frekanslı bu işaretlerin kullanımına pek rastlanmamaktadır. Zihinsel iş yükü ile birlikte beyinde özellikle alfa ve teta aktivitesinde değişkenlik olduğunu gösteren pek çok çalışma bulunmaktadır. Ayrıca, beynin özellikle frontal, pariyetal ve oksipital bölgelerinin zihinsel iş yüküne duyarlı olduğu gösterilmiştir (Lim, Sourina ve Wang, 2018).

Menon, Rivera, White, Glover ve Reiss (2000), yüksek bir zihinsel iş yükü koşulu altında görev talebi arttıkça pariyetal ve oksipital alfa gücünün azaldığını, frontal teta gücünün ise arttığını ortaya koymuşlardır. Aynı koşulda pariyetal kortikal alanlarda artmış bir teta gücü de rapor edilmiştir (Gundel ve Wilson, 1992). Uzun süreli zihinsel eforun ardından meydana gelen birikimli bir yorgunluk seviyesi ile merkez bölgede artan teta gücü arasında anlamlı ilişki olduğunu ortaya koyan çalışmalar mevcuttur (Akin, Kurt, Sezgin ve Bayram, 2008). Holm, Lukander, Korpela, Sallinen ve Müller (2009) tarafından yapılan çalışmada da yüksek zihinsel iş yükü koşulu altında frontal bölgedeki teta gücünde artış olduğu,

pariyetal ve oksipital bölgedeki alfa gücünde ise düşüş olduğu gösterilmiştir. Ayrıca Chuang ve diğerleri (2012) tarafından yapılan çalışmada oksipital korteksten gelen yüksek frekanslı beta ve gama bantlarındaki güç artışının, sürüş görevlerindeki değişen görev performansına yanıt olarak ortaya çıktığı bulunmuştur.

Belirli bir bant gücünün bölgesel olarak ele alındığı çalışmalar ile söz konusu frekans bandına ilişkin tüm elektrotlardan kaydedilen toplam gücün izlendiği çalışmalar, zihinsel iş yükü ve bant gücü arasında farklı yönde ilişki durumları ortaya koyabilmektedir. Bu doğrultuda, zihinsel iş yükünün artmasıyla birlikte pariyetal ve oksipital alfa gücünde meydana gelen azalmaya işaret eden çalışmaların aksine, alfa aktivitesi ile bilişsel performans arasında pozitif bir korelasyon bulunduğunu öne süren çalışmalar da mevcuttur (Wan, Nan, Vai ve Rosa, 2014).

Farklı frekanslardaki EEG bant gücü ile görev zorluk seviyeleri arasındaki anlamlı ilişkiyi ortaya koyan çalışmalardan birinde, bir tetikte olma görevi sırasında görev zorlaştıkça teta ve alfa gücünün zamanla bariz bir şekilde arttığı (Kamzanova, Metthews, Kustubayeva ve Jakupov, 2011), daha zorlayıcı olan Stroop testi boyunca ise yalnızca teta gücünün arttığı (Barwick, Arnett ve Slobounov, 2012) gözlenmiştir. Zammouri ve diğerleri (2018), geliştirdikleri sınıflandırıcıya dayanarak bilişsel görevdeki zorluk seviyesi arttıkça merkez ve arka bölümlerdeki teta ve alfa gücünün azaldığını ortaya koymuşlardır. Ayrıca, görevlerin zorluk derecesindeki artışa karşılık beta gücünün de arttığı gösterilmiştir (Howells, Stein ve Russell, 2010). Lim ve diğerleri (2018) tarafından üç farklı zorluk seviyesinde gerçekleştirilen çalışmada, yüksek düzeyde zihinsel aktivite sırasında beynin ön bölgesindeki teta ve beta gücünde artış, arka bölgesindeki alfa gücünde ise düşüş gözlenmiştir. Kosti ve diğerleri (2018) bilgisayar programcılarının kod anlama ve kod yazım hatalarını bulma şeklinde iki farklı görevi gerçekleştirirken topladıkları EEG verilerinden beyin aktivitelerinde ortaya çıkan farkı analiz etmiş ve kodu anlama görevinin hataları bulma görevine göre daha fazla zihinsel iş yüküne neden olduğunu ortaya çıkarmışlardır.

Özellikle teta ve alfa dalgalarının zihinsel iş yükündeki değişkenliğe duyarlılığının kanıtlanması ile birlikte, son yıllarda yapılan bazı araştırmalarda zihinsel iş yükü seviyesini daha net ayırtılabilecek oransal değişkenler türetilmeye çalışılmaktadır. Örneğin, teta/alfa oranının görev zorluk seviyesi ile birlikte arttığı ortaya konmuştur (Plechawska-Wojcik ve

diğerleri, 2019). Frontal teta/pariyetal alfa oranının görev zorluk seviyesi ile ilişkisinin incelendiğı başka bir çalışmada da anlamlı ölçüde pozitif yönlü ilişki elde edilmiştir (Holm ve diğerleri, 2009).

Choi ve diğerleri (2018) farklı frekanslardaki EEG dalga gücüne karşılık gelen değişkenleri oransal olarak ele aldıkları çalışmalarında EEG tabanlı iş yükü indeksi adlı bir ölçüm yöntemi geliştirmiş ve bunu NASA-TLX ile karşılaştırarak yüksek korelasyon elde etmişlerdir. Söz konusu indeks; oransal beta gücü ile oransal gama gücünün toplamının, oransal teta gücü ile oransal alfa gücünün toplamına bölünmesiyle elde edilen bir ölçüdür. Yapılan çalışmada, önerilen indeks ile hata sayısı arasındaki korelasyon ve NASA-TLX ile hata sayısı arasındaki korelasyon karşılaştırılmış, önerilen indeksin zihinsel iş yükü değerlendirmede ve insan performansının tahmininde bir ölçüm tekniğı olarak tercih edilebilirliği kanıtlanmıştır. EEG tabanlı iş yükü indeksi arttıkça iş yükü artmakta, azaldıkça azalmaktadır.

Guan, Chai, Zhang, Li ve Niu (2021), iş yükündeki artışla birlikte frontal teta gücünün ve pariyetal teta/alfa gücü oranının önce önemli ölçüde arttığını, sonra hafifçe azaldığını gözlemlemişlerdir. Pariyetal alfa gücü ise önce önemli ölçüde azalmış, sonrasında bir miktar artmıştır.

Merkezi, pariyetal ve oksipital beta gücünün zihinsel iş yükündeki değişimle ilişkili olduğunu gösteren çalışmalar yapılmıştır (Yin ve Zhang, 2017; Plechawska-Wojcik ve diğerleri, 2019). Ayrıca, frontal, pariyetal ve oksipital gama gücünün görev zorluk seviyesi ve algılanan zihinsel iş yükü ile pozitif ilişkili olduğu ortaya konmuştur (Harputlu Aksu ve Çakıt, 2022).

3.3. Zihinsel İş Yükü Seviyesinin Tahmininde Makine Öğrenmesi Algoritmalarının Kullanıldığı Çalışmalar

Son yıllarda bir takım yazılımlardan yararlanarak zihinsel iş yükünü sayısal olarak tahmin etmek ya da en doğru şekilde sınıflandırmak için çeşitli iş yükü modelleri geliştiren çalışmalar bulunmaktadır. Bunlar, matematiksel modeller olduğu gibi makine öğrenmesine dayalı tahmin modelleri de olabilmektedir. Zihinsel iş yükünü seviye olarak tahmin etmeye yarayan bir sınıflandırma modeli yerine sayısal olarak tahmin etmeye yönelik nesnel-

analitik modellerin çoğu Görsel İşitsel Bilişsel Psikomotor (GİBP) yaklaşımı ile oluşturulmaktadır. Görsel, işitsel, bilişsel ve psikomotor kaynakların iş yüküne etki ettiği kabul edilen çoklu kaynak teorisinden yararlanan GİBP yaklaşımında belirtilen her bir talep kanalı için oluşturulan 0-7 arasında bir talep ölçeği vardır. Daha açıklayıcı ve detaylı bilinen durumlar için belirli ara değerler verilebilmektedir (Örneğin görsel okuma görevi 5,9 değerine sahiptir). Görev ağındaki görevlerin her birine her kanal için bir talep değeri atanır. Daha sonra talep değerleri toplanır ve genel iş yükü puanı oluşturulur. Böylece zamanın her bir noktası için elde edilen iş yükü puanı, iş yükü profilinin oluşturulmasını sağlar (Rusnock ve diğerleri, 2015). GİBP yöntemine dayanan ve sürekli iş yükü profillerinin elde edilmesini sağlayan bir takım yazılımlar geliştirilmiştir. Bunlardan literatürde en sık rastlanana Improved Performance Research Integration Tool (IMPRINT) yazılımıdır (Borghetti ve diğerleri, 2017; Kandemir, Handley ve Thompson, 2018; Mitchell, D. K., 2000; Rusnock ve Borghetti, 2018; Rusnock ve diğerleri, 2015). Amerika Birleşik Devletleri'nin askeri ihtiyaçları doğrultusunda geliştirilen bir yazılım olması sebebiyle erişimi ve kullanım alanı sınırlıdır.

Tezin kapsamı gereği zihinsel iş yükünü seviye olarak tahmin etmeye yarayan bir sınıflandırma modeli geliştirilmek istendiğinden, bu bölümde belirtilen amaç için makine öğrenmesi algoritmalarından yararlanan çalışmalardan bahsedilecektir.

Büyük boyutlu verinin girdi olarak kullanıldığı sınıflandırma modellerinin çözümü için makine öğrenmesi algoritmalarının kullanımına ihtiyaç duyulmuştur. Örneğin, ekranda verilen kelimelerden düzgün cümleler oluşturarak kalem ile tablet arayüzüne yazma görevi verilen bir çalışmada, belirlenen 21 yazma özelliğine (kalemin maksimum basıncı, kalemi tutma eğikliği, kalemin havada yazmadan bekleme süresi, iki hareket arasında 100 milisaniyeyi aşan bekleme sayısı vb.) ilişkin değerler kaydedilmiştir. Ayrıca algılanan zorluk seviyesi ve iş yükü 1-9 Likert ölçeğine göre sorularak öznel değerlendirmeler alınmıştır ve görev tamamlama süreleri ölçülmüştür. Verilen kelime sayısına göre belirlenen üç zorluk seviyesinin gerçekten de algılanan zorluk ve iş yükü seviyesinde beklenen değişikliğe sebep olup olmadığı tek yönlü varyans analizi ile doğrulanmıştır. Yapay Sinir Ağları (YSA) algoritması ile elde edilen modelin sınıflandırma doğruluğu %76,27 olarak elde edilmiştir (Lin, Xie, Chen ve Tang, 2013).

Büyük boyutlu fizyolojik verilerin otomatik olarak analiz edilmesi ihtiyacı da makine öğrenmesi ilkelerine dayalı yaklaşımların kullanımını yaygınlaştırmıştır. Fizyolojik yöntemlerden hem performans tahmininde hem de doğrudan iş yükü tahmininde yararlanılmaktadır. Van Orden, Jung ve Makeig (2000), yorgunluk kaynaklı performans değişiminin modellenmesini amaçlayan çalışmalarında; tüm katılımcıların farklı iki günde gerçekleşen oturumlarda elde edilen göz izleme verilerini kullanarak genel bir regresyon modeli, her bir katılımcı için her iki oturumdan elde edilen verileri kullanarak ise kişiselleştirilmiş sinir ağları modeli elde etmişlerdir. Sinir ağları modelinin Hata Kareler Ortalamasının Karekökü (HKOK) değeri regresyon modeline kıyasla daha düşük çıkmıştır. Araç sürücülerinin zihinsel iş yükü üzerine yapılan bir çalışmada elde edilen ileri beslemeli yapay sinir ağı sınıflandırıcısı tetikte olma, uyuşukluk ve uyku durumlarını %90'ın üzerinde doğruluk performansı ile tahmin edebilmiştir (Subasi, 2005). Sassaroli ve diğerleri (2008) tarafından yapılan çalışmada; prefrontal kortekste meydana gelen hemodinamik değişiklikler baz alınarak farklı zihinsel iş yükü seviyelerini ayırt etmek üzere, 3 katılımcıya fYKS yöntemi uygulanmıştır. Oksihemoglobin ve deoksihemoglobin genliğine dayanarak zihinsel iş yükü seviyelerini belirlemek için K En Yakın Komşu (KEYK) algoritması kullanılmıştır ($k=3$). Sınıflandırma başarı oranları %44-%72 aralığında bulunmuştur. Bu oran, rastgele verilerin kullanıldığı duruma karşılık gelen şans seviyesinden (%19,1) daha yüksektir. Tek bir parametreye (konsantrasyon değişikliklerinin genliğine) dayanan basit bir sınıflandırma algoritmasıyla bile zihinsel iş yükü seviyeleri ayrıştırılabilmektedir.

Tjolleng ve diğerleri (2017) tarafından yapılan çalışmada ise, sürücülerin zihinsel iş yükünü tahmin etmek için EKG verilerine dayalı ileri beslemeli bir YSA modeli geliştirilmiştir. EKG sinyalleri 15 erkek katılımcı üzerinde ölçülmüştür. Katılımcılar sürüş simülasyon görevini bir ikincil n-geri görevi ile birlikte veya ikincil görev olmadan gerçekleştirmişlerdir. Sürüş sırasında kalp tepkilerindeki bireysel farklılıkları telafi etmek için her bir katılımcının EKG sinyallerine en hassas iki EKG ölçümünün seçimi, üç (düşük, orta ve yüksek) iş yükü seviyesinin tanımlanması ve seçilen EKG ölçülerinin normalleştirilmesi şeklinde üç aşamalı bir veri işleme prosedürü uygulanmıştır. YSA sınıflandırma modelinin doğruluğu, öğrenme verileri için %95, test verileri için ise %82 olarak bulunmuştur.

Bu çalışmada olduğu gibi zihinsel iş yükü sınıflandırma modellerinde EEG ve göz izleme yöntemlerinin girdi olarak kullanıldığı pek çok çalışma bulunmaktadır.

Göz izleme yönteminin kullanıldığı bir tahmin çalışmasında Wu ve diğerleri (2019) tarafından; uzman olmayan 32 kişi ve uzman olan 7 kişiden 2 farklı zorluk seviyesindeki operasyonel görevleri gerçekleştirmeleri istenmiştir. Göz bebeği çapı, göz kırpma hızı, odaklanma hızı ve seğirme hızının girdi olarak kullanıldığı modelde NASA-TLX skoru ise çıktı olarak ele alınmıştır. YSA ile elde edilen modelin doğruluğu test verisi için %97 olarak bulunmuştur. Ancak model uzman olan 7 kişide test edildiğinde orijinal NASA-TLX verisi ile modelin tahmin ettiği arasında anlamlı bir fark oluşmuştur. Tahmin verisinde yaklaşık %10 azaltma yapılarak daha anlamlı sonuçlar elde edilebileceği yorumlanmıştır.

Kaczorowska ve diğerleri (2020) tarafından 26 katılımcının rakam sembol eşleştirme testi esnasındaki göz izleme verilerine dayanarak kurulan ikili sınıflandırma modellerinde Destek Vektör Makineleri (DVM), KEYK ve Rastgele Ormanlar (RO) algoritmalarından yararlanılmış, DVM algoritması ile %97 doğruluk oranı elde edilmiştir. Yine Kaczorowska ve diğerleri (2021) tarafından gerçekleştirilen ve 29 kişiye üç zorluk seviyesinden oluşan rakam sembol eşleştirme testinin uygulandığı çalışmada, göz izleme verileri ile kurulan 20 değişkenli modelin doğruluk performansı lojistik regresyon ve RO algoritmalarıyla %95 olarak elde edilmiştir. Elasticnet regresyonu kullanılarak yapılan boyut azaltma sayesinde belirlenen en önemli 4-5 öz nitelik ile yeniden kurulan modelin performansı %97'ye yükselmiştir.

Zihinsel iş yükünün n-geri görevler kullanılarak sınıflandırıldığı ve iyi bir doğruluk performansının elde edildiği ilk çalışmalardan biri Grimes ve diğerleri (2008) tarafından gerçekleştirilmiştir. 8 kişi üzerinde yapılan çalışmada 32 kanallı EEG cihazı ile alınan ölçümler sonucunda kurulan 2 zorluk seviyeli modelinin doğruluk oranı %99, 4 zorluk seviyeli modelin doğruluk oranı ise %88 olarak bulunmuştur.

Lim ve diğerleri (2018) tarafından üç farklı zorluk seviyesinde çoklu görev testlerinin uygulandığı bir çalışmada EEG güç değişkenlerinin girdi olarak kullanıldığı sınıflandırma modelinin çözümünde DVM algoritması kullanılmış, %69 doğruluk oranı elde edilmiştir.

Yin ve Zhang (2017) 7 kişi üzerinde yaptıkları çalışmada zihinsel iş yüküne ilişkin ikili sınıflandırma temelli bir tahmin modeli geliştirmişlerdir. Kortikal fonksiyonların hiyerarşik ve katmanlı yapısından hareketle YSA şeklinde ele alınan modelde EEG sinyal verileri kullanılarak derin-öğrenme temelli uyarlanabilir bir sınıflandırıcı geliştirilmiştir. Uyarlanabilir olması, gizli tabakadaki nöronların ağırlıklarının, test prosedürü sırasında güncellendiği anlamına gelmektedir. Çalışmanın sonunda modelin tahmin edicisinin %85,7 olarak elde edilen doğruluk performansı diğer klasik sınıflandırıcılar (destek vektör makinaları, en yakın komşu vb.) ile karşılaştırılmış ve bu yöntemin statik sınıflandırıcılardan üstün geldiği gösterilmiştir.

Jusas ve Samuvel (2019) tarafından 9 kişi üzerinde yapılan çalışmada, katılımcıların belirli bir eylemi prova veya simüle ettiği zihinsel süreçleri içeren 4 farklı motor imgeleme görevi kullanılmıştır. Yalnızca EEG fizyolojik verilerinin kullanıldığı çalışmada en iyi sınıflandırma performansı Doğrusal Ayırma Analizi (DAA) kullanılarak %64 şeklinde elde edilmiştir.

Qu ve diğerleri (2020) 10 kişinin EEG verileri üzerinde yaptıkları çalışmada 3 seviyeli zihinsel iş yükünü DVM algoritmasını kullanarak %79,8 doğruluk performansı ile tahmin etmişlerdir. 32 kanallı cihaz ile toplanan EEG sinyallerinin girdi olarak kullanıldığı modeller 4 farklı makine öğrenmesi algoritması aracılığıyla oluşturulan denek bağımlı sınıflandırıcılar ile çözdürülmüş en iyi sonuç DVM algoritması ile elde edilmiştir. Veri ön işleme adımında Bağımsız Bileşenler Analizi (BBA) prosesinin uygulandığı metot ile daha yüksek doğruluk oranı elde edildiği ortaya konmuştur.

Pei, Wang, Bezerianos ve Li (2021) tarafından 7 kişi üzerinde yapılan çalışmada bir simülasyon uçuş görevi sırasında 64 kanallı EEG cihazı ile alınan ölçümler sonucunda farklı EEG öznitelikleri kullanılarak kurulan modelde RO algoritması ile %84,3 doğruluk performansı elde edilmiştir. Sadece bant gücü değişkenleri ile kurulan 3 seviyeli modelin sınıflandırma performansı ise %75,9 olmuştur.

EEG yönteminin kullanıldığı bir başka zihinsel iş yükü sınıflandırma çalışmasında Zhou ve diğerleri (2022) 7 farklı zorluk seviyesinde işleyen bellek görevleri ve matematiksel toplama görevlerinin art arda verildiği bir deney prosedürü oluşturmuşlardır. KEYK,

DVM, DAA ve YSA algoritmaları kullanılarak görevler arası koşulda çalıştırılan modellerde en iyi sonucu %56 ile DVM ve DAA algoritmaları sağlamıştır.

Şaşmaz ve diğerleri (2023) tarafından MS hastalarının ve sağlıklı bireylerin bilgisayar ve sanal gerçeklik ortamında verilen görevler sırasında meydana gelen zihinsel iş yüklerinin düşük, orta ve yüksek şeklinde sınıflandırılabilmesi için önerilen modelde; MS hastalarının zihinsel iş yükü seviyesi bilgisayar ortamında %96, sanal gerçeklik ortamında %94 doğruluk performansı ile ayrıştırılabilmektedir. Sağlıklı bireylerin zihinsel iş yükü seviyesi ise bilgisayar ortamında %95, sanal gerçeklik ortamında %94 oranında doğru tahmin edilmiştir. EEG değişkenlerinden yararlanılan modelde ortalama sınıflandırma performansı %83,4 bulunmuştur.

Liu ve diğerleri (2017) tarafından yapılan çalışmada 21 katılımcının 3 farklı koşul altında iş yükü seviyelerinin sınıflandırılması için katılımcılardan bir n-geri hafıza görevini tamamlamaları istenmiştir. Katılımcıların EEG, fYKS verileri ile kalp atış hızı, solunum sayısı, kalp atış hızındaki değişkenlik gibi diğer fizyolojik verileri eş zamanlı kayıt altına alınmıştır. EEG ve fYKS yöntemlerinin tek başına kullanılmasına kıyasla birlikte kullanımının daha doğru sonuç verdiği ortaya çıkmıştır (doğruluk oranı: %65).

Borys ve diğerleri (2017a) tarafından 13 erkek lisans öğrencisi üzerinde yapılan çalışmada üç farklı zorluk seviyesinde aritmetik görevler gerçekleştirmesi istenen deneklerin hem göz hem de EEG verileri kayıt altına alınmıştır. Göze ilişkin 6 değişkenden oluşan ikili sınıflandırma modelinde %90 doğruluk, göze ilişkin 17 değişkenden oluşan üçlü sınıflandırma modelinde ise %73 doğruluk performansı elde edilebilmiştir. EEG özelliklerinin modellere katkı sağlayan değişkenler arasında yer almaması dikkati çekmiştir.

Bilgisayar tabanlı bir uçuş görevinin 28 kişi tarafından gerçekleştirildiği bir başka çalışmada, 32 kanallı EEG cihazı ile katılımcıların beyin sinyalleri, 256 Hz frekanslı göz izleme cihazı ile de göz verileri toplanmıştır (Li ve diğerleri, 2023). Lineer regresyon, destek vektör makineleri, rastgele ormanlar ve yapay sinir ağları algoritmalarının kullanıldığı çalışmada, 3 kategorili sınıflandırmada en iyi sonuç göz izleme ve EEG değişkenlerinin birlikte kullanıldığı modelde %54,2 doğruluk oranı ile elde edilmiştir. Destek vektör makinelerinde doğrusal, polinomial ve radyal tabanlı fonksiyon

kullanılmış, radyal tabanlı fonksiyon diğerlerinden daha yüksek performans sağlamıştır. 2 kategorili sınıflandırmada ise lineer regresyon ile %82,7 doğruluk oranına ulaşılmıştır.

Birçok araştırmacı, bir makine öğrenmesi modelindeki eğitim ve test verisi setlerinin aynı görev için aynı günde ve aynı katılımcıdan toplandığında zihinsel iş yükü sınıflandırma doğruluğunun özellikle ikili sınıflandırmada %90'dan yüksek olabileceğini göstermiştir. Çünkü nörofizyolojik sinyaller, zihinsel iş yükü değişkenliği açısından durağan olmayan, kişiye ve göreve özgü özelliğe sahiptir. Ancak, örneğin EEG özelliklerinin katılımcılara, görevlere ve günlere göre değişmesi, statik bir zihinsel iş yükü sınıflandırıcısının sınıflandırma doğruluğunu düşürebilmektedir. Nörofizyolojik sinyallerin katılımcıya özgü olma durumunun olası nedenleri, bireysel özelliklerin, fiziksel koşulların ve uzman bilgisinin farklı operatörler için farklı olmasıdır. Göreve özgü olmasının ise; gözle takip, işitsel yanıt ya da zihinsel-aritmetik odaklı görevlerde zihinsel gereksinimlerin farklılık göstermesinden kaynaklandığı söylenebilmektedir. Somut bir uygulama ile belirli bir insan-makina görevi ve belirli bir operatör için bir zihinsel iş yükü sınıflandırıcısı tasarlanabilmektedir. Ancak, farklı görev koşullarında farklı kişilerden toplanan verilere dayanarak eğitilen ve test edilen zihinsel iş yükü sınıflandırıcılarının genelleme yeteneğini artırmak son yıllarda önem kazanmıştır (Yin ve Zhang, 2017).

Operatör iş yükünü doğru bir şekilde tahmin edebilen uyarlanabilir bir sistem sayesinde otomasyondan çok daha verimli bir şekilde yararlanılabilmesi mümkündür. Uyarlanabilir bir öğrenmede sınıflandırma eğitimi ve testi için aynı denekten veri kullanmak uygun değildir. Ancak yapılan çalışmalarda iş yükü sınıflandırıcılarının çoğu bireye özgüdür, yani her bir katılımcı için yeni bir sınıflandırıcı oluşturulur ve eğitilir. Bireye özgü birden fazla sınıflandırıcı kullanmak yerine birden fazla birey için kullanılabilecek tek bir sınıflandırıcı geliştirmek zihinsel iş yükü sınıflandırmasında çok büyük bir problem olan katılımcılar arası farklılıkları açıkça modelleyebilmek açısından önem taşımaktadır. Bu, daha önce modele öğretilmemiş yani modelin tanımadığı yeni bir bireyin iş yükünü ayırt edebilen bir tahmin modeli geliştirmek için önemli bir adımdır. Bu amaçla son yıllarda yapılan çalışmalar mevcuttur. Z. Wang, Hope, Z. Wang, Ji, ve Gray (2012) tarafından yapılan çalışmada hiyerarşik bayes modelin çapraz denek sınıflandırıcısı, 8 kişiden oluşan bir grup katılımcıdan alınan EEG verisi ile eğitilmiş ve test edilmiştir. Çalışmanın sonucunda, çapraz denek sınıflandırıcısının bireye özgü sınıflandırıcılar ile karşılaştırılabilir bir performans sergilediği ortaya konmuştur. Zihinsel iş yükü seviyesini denekten bağımsız

değerlendirmeyi amaçlayan bir diğer çalışmada, klasik makine öğrenmesi teknikleri (destek vektör makineleri, karar ağacı, k en yakın komşu ve rastgele orman) kullanılmıştır. Deneysel prosedürün ilk 5 bölümünde 17'şer aritmetik işlem, son bölümünde ise 16 aritmetik işlem yapan 11 katılımcının EEG verileri toplanmıştır. Hem denek bağımlı hem de denek bağımsız uygulamada en yüksek doğruluk oranı k en yakın komşu algoritması ile elde edilmiştir (%90,5, %91,5) (Plechawska-Wojcik ve diğerleri, 2019). Çapraz denek iş yükü tahmininin araştırıldığı başka bir çalışmada ise, bir katılımcının bireysel iş yükü diğerlerinden gelen verilere dayanarak tahmin edilmiştir. Artan zorluk derecesine sahip matematik problemlerini çözmeleri istenen 10 katılımcının EEG verileri kaydedilmiş ve doğrusal regresyon ile analiz edilmiştir. Gerçek iş yükü değeri ile tahmin edilen değer arasındaki korelasyon katsayısının ortalaması denek bağımlı model için 0,88, denek bağımsız model için ise 0,84 bulunmuştur. HKOK değerleri ise 0,6-7,2 arasında değerler alan bir iş yükü ölçeği için sırasıyla 1,02 ve 1,58 olarak elde edilmiştir (Walter, Wolter, Rosenstiel, Bogdan ve Spüler, 2014).

Smith, Borghetti ve Rusnock (2015) tarafından yapılan çalışmada, iş yükü tahmininde çapraz uygulama hatasını en azlayacak uygun yöntemi bulmak için iki farklı regresyon-ağaç yöntemi karşılaştırılmıştır. Hem çapraz denek hem de çapraz görev uygulaması yapılmıştır. Görevler arası uygulamada elde edilen sonuçlar tatmin edici olmazken, her iki algoritmanın da denekler arası iş yükü tahmin hatasını anlamlı ölçüde azalttığı ortaya konmuştur. Rastgele ormanlar algoritması çapraz denek uygulamasında en iyi sonucu vermiştir.

Dört farklı kanaldan fYKS verilerinin toplandığı bir çalışmada; 5 katılımcıdan, yalnızca sürüş, sürüşle birlikte 1-geri görev, sürüşle birlikte 2-geri görev olmak üzere 3 farklı zorluk seviyesindeki görevleri yerine getirmeleri istenmiştir. 60 deney sonucunda toplanan veriler ile elde edilen modellerin denek bağımlı-denek bağımsız ve kanal bağımlı-kanal bağımsız sınıflandırıcıları performans açısından karşılaştırılmıştır. Denek bağımlı sınıflandırma doğruluğu %81,30-95,40 arasında değişirken, denek bağımsız sınıflandırma doğruluğu %84,90-89,50 arasında bulunmuştur. Hem kanal hem de denek bağımsız sınıflandırma doğruluğu ise %82,90 olarak elde edilmiştir (Le, Aoki, Murase ve Ishida, 2018).

Çizelge 3.1'de literatürde zihinsel iş yükü tahminini sınıflandırma problemi olarak ele alan çalışmalar özetlenmiştir. Çizelgede yer alan çalışmalarda elde edilen model performansları

bu tezde elde edilen modellerin sonuçları ile “5. Araştırma Bulguları ve Tartışma” bölümünde karşılaştırılacaktır.

Çizelge 3.1. Zihinsel iş yükü sınıflandırmada makine öğrenmesi kullanılan çalışmalar

Yazar (Yıl)	Katılımcı Sayısı	Sınıf Sayısı	Ölçüm Yöntemi	Kullanılan Algoritma	Doğruluk Performansı
Subasi (2005)	30	3	EEG	YSA	Tetikte olma: %95+-3 Uyuşukluk: %93+-4 Uyku: %92+-5
Grimes ve diğerleri (2008)	8	2-4	EEG	Naif Bayes	2 sınıflı model: %99 4 sınıflı model: %88
Sassaroli ve diğerleri (2008)	3	3	fYKS	KEYK (k=3)	%44,4–%72,2
Wang ve diğerleri (2012)	8	3	EEG	YSA Naif Bayes	%30-84
Lin ve diğerleri (2013)	49	3	Belirlenen 21 yazma özelliği	YSA	%76,27
Borys ve diğerleri (2017a)	13	2-3	EEG + Göz İzleme	KA DAA Lojistik regresyon DVM KEYK	2 sınıflı model: %90 3 sınıflı model: %73
Liu ve diğerleri (2017)	21	3	EEG + fYKS	DAA Naif Bayes	%65
Tjolleng ve diğerleri (2017)	15	3	EKG	YSA	%82
Yin ve Zhang (2017)	7	2	EEG	Derin öğrenme	%85,7
Le ve diğerleri (2018)	5	3	fYKS	KA DAA Lojistik Regresyon DVM KEYK	Denek bağımlı model: %81,30-95,40 Denek bağımsız model: %84,90-89,50
Lim ve diğerleri (2018)	48	3	EEG	DVM	%69
Jusas ve Samuvel (2019)	9	4	EEG	DAA	%64

Çizelge 3.1. (devam) Zihinsel iş yükü sınıflandırmada makine öğrenmesi kullanılan çalışmalar

Yazar (Yıl)	Katılımcı Sayısı	Sınıf Sayısı	Ölçüm Yöntemi	Kullanılan Algoritma	Doğruluk Performansı
Plechawska-Wojcik ve diğerleri (2019)	11	3	EEG	DVM KA KEYK RO	Denek bağımlı model: %71,1-90,5 Denek bağımsız model: %70,4-91,5
Wu ve diğerleri (2019)	32+7	2	Göz izleme	YSA	%97
Kaczorowska ve diğerleri (2020)	26	2	Göz izleme	DVM KEYK RO	%97
Qu ve diğerleri (2020)	10	3	EEG	DVM	%79,8
Kaczorowska ve diğerleri (2021)	29	3	Göz izleme	Lojistik regresyon RO	%97
Pei ve diğerleri (2021)	7	3	EEG	RO	Farklı EEG öznitelikleri ile: %84,3 Yalnızca bant gücü ile: %75,9
Zhou ve diğerleri (2022)	45	7	EEG	KEYK DVM DAA YSA	%56
Li ve diğerleri (2023)	28	2-3	EEG + Göz İzleme	Doğrusal regresyon DVM RO YSA	2 sınıflı model: %82,7 3 sınıflı model: %54,2
Şaşmaz ve diğerleri (2023)	17 hasta 28 sağlıklı	3	EEG	DVM RO DAA YSA	%83,4

3.4. Problemin Tanımı ve Literatüre Katkı

İlgili literatür incelendiğinde, hem EEG'nin hem de göz izleme yönteminin zihinsel iş yükü analizinde önemli rol oynadığı görülmüş; ancak, her iki yönetime ilişkin ele alınan

değişkenler ile zihinsel iş yükü arasındaki ilişkinin yönü hakkında tutarlı olmayan sonuçlar olduğu göze çarpmıştır. Özellikle göze ilişkin değişkenlerin, bazı çalışmalarda zihinsel iş yükü ile birlikte arttığı bazı çalışmalarda ise azaldığı ortaya çıkmıştır. Bu doğrultuda, söz konusu iki yönetime ilişkin verilerin görev zorluk seviyesi ile ilişkisinin incelenmesi ve benzer çalışmalar ile karşılaştırılarak literatüre katkı sağlanması amaçlanmıştır.

Zihinsel iş yükü seviyesinin tahmin edilmeye çalışıldığı benzer çalışmalarda sınıf sayısı genellikle iki ya da üç sınıf olarak ele alınmakta, iş yükü seviyelerine ilişkin bilginin ayrıntı düzeyini artırabilmek için sınıf sayısını yüksek tutmayı amaçlayan az sayıda çalışma bulunmaktadır. Diğer yandan, sınıf sayısı arttıkça problemin karmaşılaşması sebebiyle tahmin modellerinin performansı düşmektedir. Bu çalışmada sınıflandırma modeli dört sınıflı olarak ele alınmış ve zorluk seviyesinin tahmin edildiği bu modele ek olarak, NASA-TLX skorunun tahmin edildiği bir regresyon modeli de geliştirilmeye çalışılmıştır.

Literatürde, EEG ve göz izleme yöntemlerinin aynı anda kullanıldığı çalışmalar oldukça az sayıdadır. Her iki cihazdan toplanan verinin senkronizasyonunun sağlanması hem deney hem de analiz sürecini zorlaştıran bir unsur olmakla birlikte, zihinsel iş yükü dalgalanmalarından etkilendiği kanıtlanan fizyolojik yöntemlerin birlikte kullanımının model performansını artırdığı yönündeki bazı bulgular bu çalışmada EEG ve göz izleme yöntemlerinin birlikte kullanımını teşvik etmiştir. Diğer yandan, literatürde hem n-geri görevlerin test edildiği hem de EEG ve göz izleme yöntemlerinin birlikte kullanıldığı bir çalışma ile karşılaşılmamıştır. Tez bu yönüyle diğer çalışmalardan farklılık göstermektedir.

Ülkemizde bilişsel yorgunluğun sınıflandırılmasına yönelik olarak daha önce yapılan iki tez çalışmasında yalnızca EEG yönteminden yararlanılmış ve problem ikili sınıflandırma olarak ele alınmıştır (Ekim, 2021; Korkmaz, 2022). Göz izleme tekniği ise genellikle pazarlama ve psikoloji çalışmalarında kullanılmakta olup, ülkemizde zihinsel iş yükü analizinde göz izleme yönteminden yararlanan bir tez çalışması bulunmamaktadır. Bu bağlamda, tez çalışmasının ele alınan konu ve kullanılan yöntemler açısından ulusal anlamda da yenilik taşıdığı söylenebilmektedir.

Hem ulusal hem de uluslararası literatüre bakıldığında benzer çalışmaların özellikle elektrik elektronik mühendisliği ve bilgisayar mühendisliği alanlarında çalışan

araştırmacılar tarafından gerçekleştirildiği dikkati çekmiş, disiplinler arası bu çalışma konusunun ergonominin en önemli meselelerinden biri olarak görülen zihinsel iş yükü analizinde endüstri mühendisliği bakış açısıyla da incelenmesi gerektiği değerlendirilmiştir.

Tez kapsamında hazır bir veri setinin kullanımı yerine, belirlenen hedef kitleden bizzat toplanan verinin kullanılmış olması bu çalışmayı farklı kılan unsurlardan bir diğeridir. Bu nedenle, özellikle veri toplama süreci çalışmanın en az analiz aşaması kadar önemli görülmektedir. Problemin tanımlanmasından başlayarak; yol haritasının oluşturulması, pilot çalışmaların gerçekleştirilmesi, gerçek deneylerin gerçekleştirilmesi yoluyla ham verinin toplanması, verinin işlenmesi ve farklı açılardan analiz edilmesi aşamasına kadar bilimsel araştırma sürecinin tüm adımları bu çalışma kapsamında gerçekleştirilmiştir.

Veri ön işleme sürecinde gerçekleştirilen ve göz kırpmaya hareketine ilişkin değişkenlerin diğer değişkenlerden yararlanılarak türetilmesi için izlenen süreç, literatürden edinilen bilgiler ile desteklenerek oluşturulan ancak literatürde rastlanmayan yeni bir yaklaşımdır. Yaklaşımın algoritması ve ilişkili makro kodları verilerek değişken türetmeye ilişkin bir yol haritası sunulması sağlanmıştır.

Benzer şekilde, ham EEG verisinin EEGLAB STUDY programı aracılığıyla nasıl analiz edilebileceği ekran görüntüleri ile birlikte açıklanmış ve ele alınan gerçek bir problem için elde edilen istatistiksel sonuçların ve grafiklerin nasıl yorumlanacağına dair bir uygulama örneği sunulmuştur.

“3.3. Zihinsel İş Yükü Seviyesinin Tahmininde Makine Öğrenmesi Algoritmalarının Kullanıldığı Çalışmalar” bölümünde bahsedilen araştırmalarda en çok kullanılan makine öğrenmesi yöntemlerinin KEYK, DVM, YSA ve RO olduğu görülmektedir. Yakın zamanda geliştirilen Gradyan Artırma Makineleri (GAM), Hafif Gradyan Artırma Makineleri (HGAM), Aşırı Gradyan Artırma Makineleri (AGAM) gibi algoritmaların bu alanda kullanımı ise oldukça sınırlıdır. Tez, görece yeni bu algoritmalar ile elde edilen sonuçları ve model optimizasyonu amacıyla farklı hiperparametre kombinasyonları denenerek sağlanması hedeflenen iyileştirmeleri de tartışmayı hedeflemiştir. Söz konusu algoritmaların tercih edilmesinin sebebi elbette yalnızca yeni olmaları değil aynı zamanda güçlü sınıflandırıcı ve regresörler elde etmekte sağladıkları başarının son yıllarda

gösterilmiş olmasıdır. Özellikle HGAM, veri bilimciler ve makine öğrenmesi ile ilgilenenleri bir araya getiren ve makine öğrenmesi yarışmaları sunan platformlarda kazananlar arasında açık ara favori gradyan artırma karar ağacı kütüphanesi olmuştur (URL-3). Ayrıca, HGAM algoritmasının daha önce vurgulanan öğrenme hızı da bu çalışma kapsamında tercih edilmesinin en önemli sebeplerindendir.

Neticede, her iki türde fizyolojik verinin eş zamanlı olarak kayıt altına alındığı deney süreci sonucunda farklı makine öğrenmesi algoritmalarının karşılaştırmalı olarak kullanıldığı dört sınıflı zihinsel iş yükü tahmin modeli ile literatüre katkı sağlanması hedeflenmiştir.

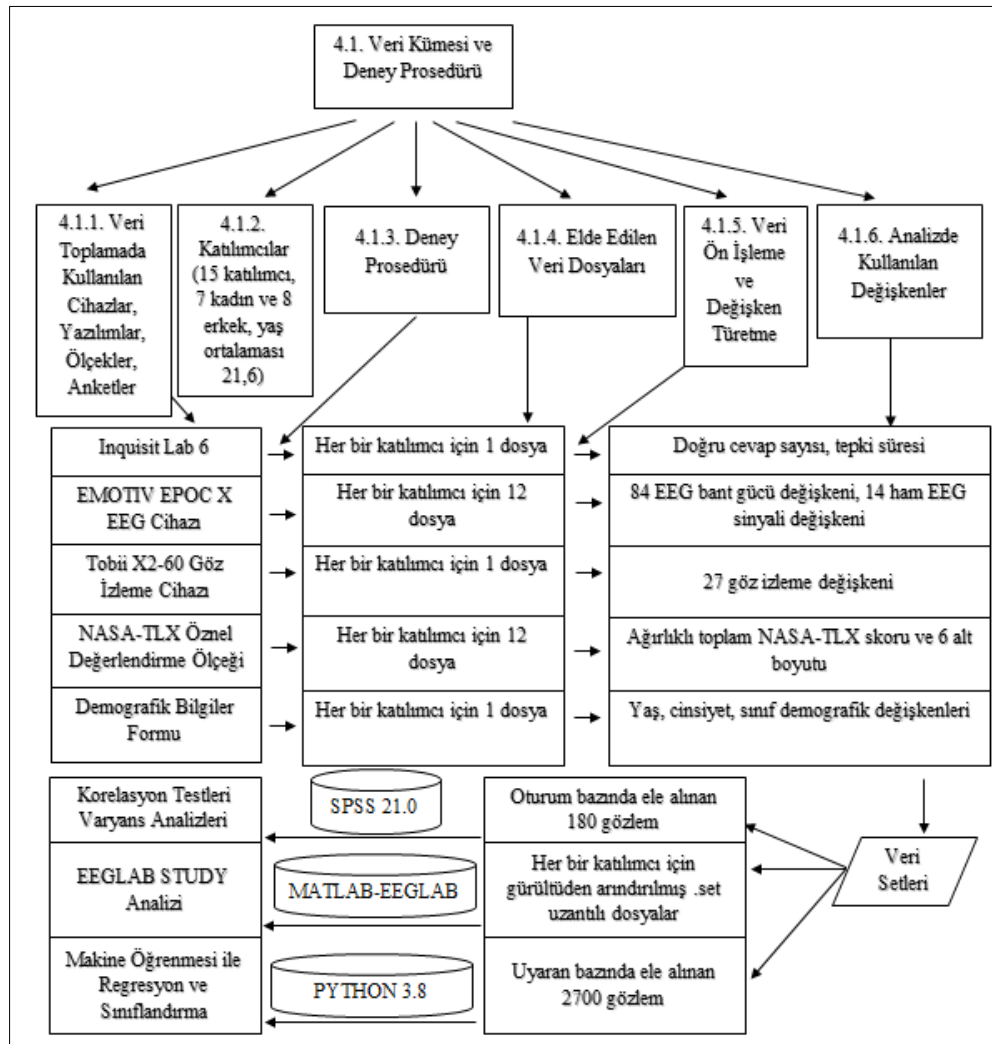
Sonraki bölümde, EEG ve göz izleme değişkenlerinin zihinsel iş yükü ile ilişkisini araştırmak ve söz konusu fizyolojik verileri kullanarak zihinsel iş yükünü tahmin eden sınıflandırma ve regresyon modellerini geliştirmek üzere yapılan çalışmada izlenen süreç sistematik olarak ele alınmış ve kullanılan yöntemler detaylandırılmıştır.

4. MATERYAL VE METODOLOJİ

Bu bölümde metodolojide izlenen yaklaşımın sistematik olarak açıklanması amaçlanmıştır. Bu nedenle; öncelikle deneysel sürecin adımlarına yer verilmiş, daha sonra da analizde kullanılan makine öğrenmesi yöntemleri ve başarı ölçütleri detaylandırılmıştır.

4.1. Veri Kümesi ve Deney Prosedürü

Bu bölümde; kullanılan cihazlar ve yazılımları, deneye katılan katılımcılar, deney prosedürü, elde edilen veri dosyaları, veri ön işleme ile elde edilen ve analizde kullanılan değişkenler açıklanacaktır. Bu doğrultuda; tez çalışması sürecinde izlenen yol haritasını gösteren şema Şekil 4.1’de verilmiştir.



Şekil 4.1. Uygulamada izlenen yol haritası

Alt bölümler Şekil 4.1’de ifade edildiği sırayla detaylandırılacaktır.

4.1.1. Veri toplamada kullanılan cihazlar, yazılımlar, ölçekler ve anketler

Inquisit Lab 6

Çalışmada; bilimsel çalışmalar yapan araştırmacıların faydalanması için hazırlanan ve literatürde kullanılarak geçerliliği kanıtlanmış olan bilişsel, sosyal, nörofizyolojik ve çevrimiçi psikolojik deneyler içeren bir yazılım olan Inquisit programından yararlanılmıştır. Inquisit Lab 6 açık kaynak kodlu bir yazılım olup belirli parametreler, deneyi yapan araştırmacı tarafından değiştirilebilmektedir. Inquisit Lab 6 yazılımında mevcut bulunan ve 4 farklı zorluk seviyesinde uygulanabilen (0-geri, 1-geri, 2-geri, 3-geri) standart görsel bir n-geri görev seti kullanılmıştır. Uyarı olarak harflerin ekranda belirdiği tekli n-geri görevler gerçekleştirilmiştir.

EMOTIV EPOC X EEG cihazı ve yazılımı

EMOTIV EPOC X EEG cihazı ve EMOTIV Pro 3 cihaz yazılımı kullanılmıştır. Söz konusu cihaz Gazi Üniversitesi Bilimsel Araştırma Projeleri (BAP) lisansüstü tez projesi desteği ile elde edilmiştir. Taşınabilir EEG cihaz paketi; giyilebilir EEG başlığı, yazılım ile bağlantıyı sağlayan bir USB aparatı, şarj kablosu, elektrotlara takmak için keçe parçalar ve çok amaçlı lens solüsyonundan oluşmaktadır. Cihazın çözünürlüğü 14 bit, örnekleme hızı ise saniyede 128 Hz’dir. Cihazın giyilebilir başlığı ve USB aparatı Resim 4.1’de gösterilmiştir.

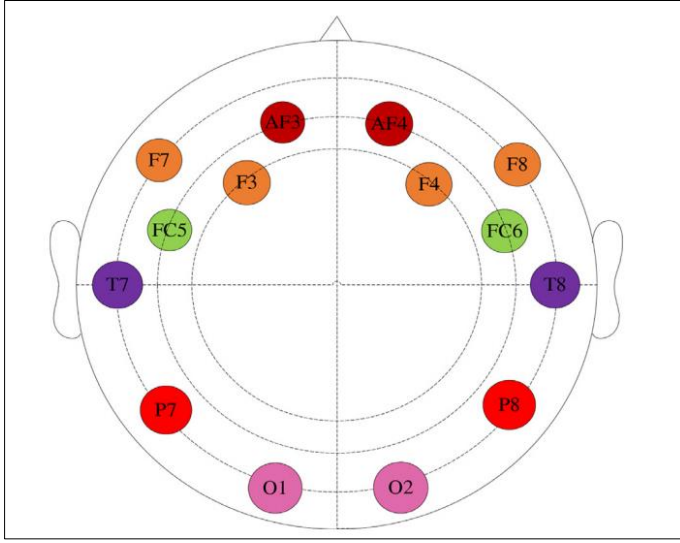
Cihaz kullanımı için öncelikle lens solüsyonunda ısıtılan ve fazla suyu sıkılan keçe bantlar EEG başlığındaki elektrot yuvalarına yerleştirilmektedir. USB aparatı bilgisayara takıldıktan sonra başlıktaki güç düğmesine basılmakta, USB’deki her iki ışık da yandıysa bağlantı sağlanmış anlamına gelmektedir (URL-4). Daha sonra EMOTIV Pro 3 programı açılarak katılımcıya EEG başlığı giydirilmektedir. Bağlantı kalitesi ve EEG kalitesi kontrol edildikten sonra kaydın başlatılması gerekmektedir (URL-2).



Resim 4.1. EMOTIV EPOC X EEG başlığı ve USB aparatı

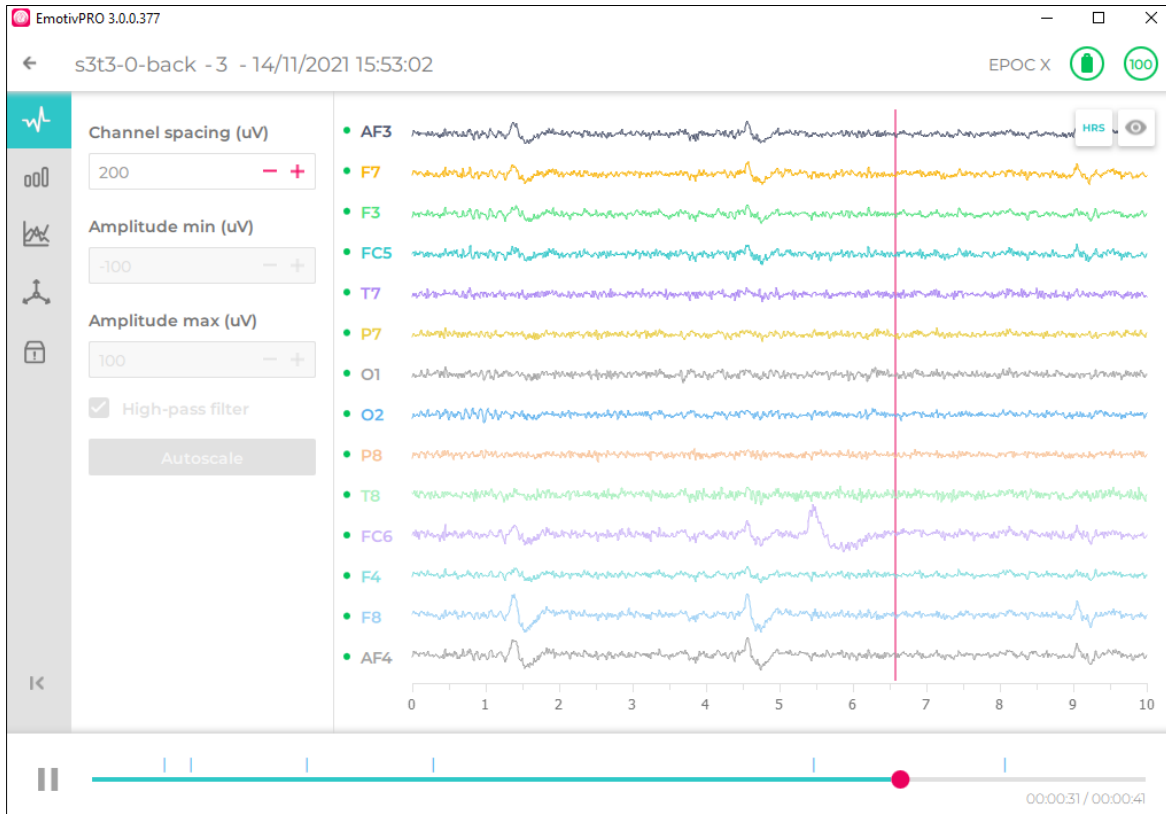
EEG cihazları için kabul edilmiş elektrot yerleşim standartları vardır. Bunlar 10–20, 10–10, 10–5 olarak adlandırılmaktadır. Bu standartlar EEG elektrotların kafatası üzerinde nerede konumlanacağı hakkında bilgi vermektedir. En yaygın kullanılan elektrot yerleşim sistemi 10–20 olarak adlandırılan sistemdir. Kafatası üzerindeki elektrot yerleşim bölgesi burun üzerindeki boşluktan (nasyon) başlayarak kafatasının arkasında ensenin üstündeki şiş bölgeye (inyon) kadar olan alanı kapsar. Bu alandaki 10–20 sistemi elektrotların yerleştiği kafatası üzerindeki bölümler arasında uzunluk oranı %20’dir. 10–20 standardında beynin kafatası yüzeyini kapsayan altı ana bölüm ismi kullanılmaktadır. Bu bölümler AF (prefrontal), F (frontal), C (central), O (oksipital), P (pariyetal) ve T (temporal) olarak adlandırılmıştır. Beynin sol tarafında bulunan elektrotlar numaralandırılırken tek sayılar, sağ tarafında bulunan elektrotlar numaralandırılırken çift sayılar kullanılmaktadır. Her kafatası bölümünün merkezini sembolize eden elektrot z ile gösterilmektedir. Genellikle sol ve sağ kulak memesi altında A1 ve A2 olarak adlandırılan iki adet referans elektrotu bulunmaktadır (Selvi, 2022).

EPOC X cihazı 14 kanallı olup beynin prefrontal (alın bölgesi), frontal (ön bölge), temporal (şakak bölgesi), pariyetal (yanal bölge) ve oksipital (arka bölge) loblarından gelen EEG sinyallerini toplamaya imkan sağlamaktadır. AF3 ve AF4 kanalları prefrontal bölgeden; F3, F4, F7 ve F8 kanalları frontal bölgeden; FC5 ve FC6 kanalları frontal merkez bölgeden; T7 ve T8 kanalları temporal bölgeden, P7 ve P8 kanalları pariyetal bölgeden, O1 ve O2 kanalları ise oksipital bölgeden gelen EEG sinyallerini iletmektedir. 14 kanallı cihazın 10 – 20 elektrot yerleşim standardı Şekil 4.2’de verilmiştir.



Şekil 4.2. 14 kanallı EEG cihazının elektrot yerleşimi (Ji ve diğerleri, 2023)

EEG sinyalleri yazılım aracılığıyla anlık olarak izlenebilmekte ve ham verisi kaydedilerek dış ortama aktarılabilir. Ham EEG veri akışı görüntüsü Şekil 4.3'te yer almaktadır. EEG sinyalleri, tüm kanallar için eş zamanlı olarak görüntülenebilmektedir. Sağ üst köşede 100 olarak görünen değer EEG ölçüm kalitesini göstermektedir.



Şekil 4.3. Ham EEG ekranı

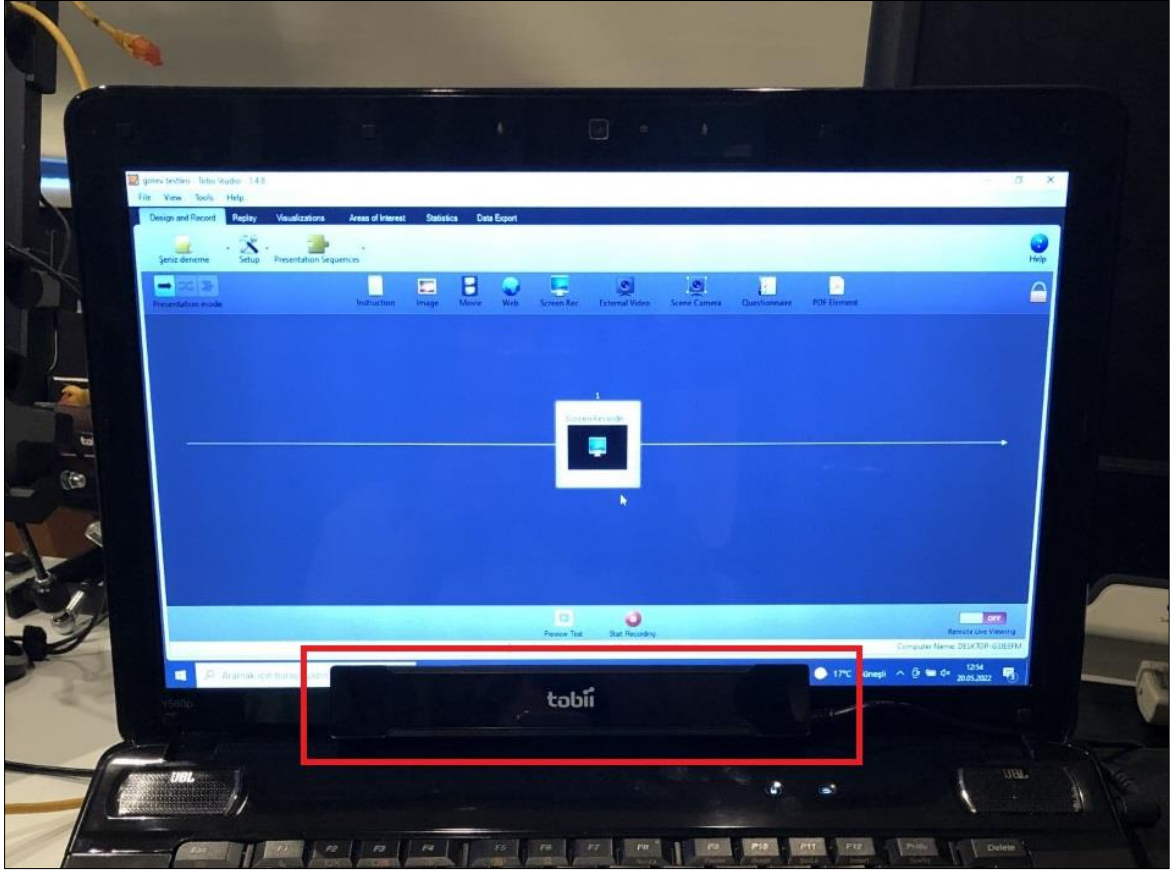
Tobii X2-60 göz izleme cihazı ve yazılımı

Deneylerde kullanılan göz izleme cihazı Orta Doğu Teknik Üniversitesi İnsan-Bilgisayar Etkileşimi Araştırma ve Uygulama Laboratuvarı'ndan temin edilmiştir. Pandemi sonrası tekrar kullanıma açılan söz konusu laboratuvarın belirlenen çalışma takvimi doğrultusunda iki hafta süreyle bu araştırma için kullanımına izin verilmiştir. Söz konusu laboratuvarında mevcut bulunan cihazlar Tobii T120, X2-60 ve Glass-2'dir. T120 model cihaz laboratuvarında yer alan masaüstü bilgisayarda sabit olup 120 Hz frekansta kayıt almayı sağlamaktadır. Glass-2 model cihaz gözlük şeklinde giyilebilir bir cihaz olup daha çok pazarlamaya yönelik çalışmalarda kullanılmaktadır.

Bu çalışmada uygulanacak testlerin bilgisayar başında gerçekleştirilecek olması sebebiyle giyilebilir bir göz izleme cihazı elzem görülmemiştir. Taşınabilir olması ve kişisel dizüstü bilgisayarda kullanım imkanı sağlaması nedeniyle tercih edilen Tobii X2-60 için kullanım izni alınmıştır.

Deney uygulama sürecinin ilk gününde X2-60 cihazında yaşanan bağlantı problemi kaynaklı teknik aksaklık nedeniyle göz izleme cihazı olarak laboratuvarında sabit bulunan T120 cihazı kullanılmıştır. Ancak elde edilen veri dosyasında seğirme hareketine ilişkin değişkenlerin yer almaması sebebiyle söz konusu cihazın çalışmanın devamı için verimli olmayacağına karar verilmiştir. Tobii X2-60 cihazından laboratuvar bünyesinde iki adet bulunmaktadır. Bağlantı problemi yaşanan cihaz yerine kullanılmak üzere diğer X2-60 cihazı üniversitedeki başka bir araştırmacıdan temin edilebilmiştir. Cihaz ile birlikte Tobii Studio 3.4.8 yazılımının kullanıcı sınırlamalı anahtarının deneylerin gerçekleştirileceği iki haftalık süreçte kullanılması için de izin alınmıştır.

Tobii X2-60 cihazı görev testlerinin gerçekleştirildiği dizüstü bilgisayar ekranının alt kısmına mıknatıslı arka yüzeyi sayesinde yerleştirilmekte, 60 Hz frekansta (saniyede 60 kayıt) veri toplamaya imkan sağlamaktadır. Cihazın yerleşimi Resim 4.2'de verilmiştir.



Resim 4.2. Tobii X2-60 göz izleme cihazı

NASA-TLX öznel değerlendirme ölçeği

Tez çalışmasında ağırlıklı NASA-TLX yöntemi kullanılmış, bu doğrultuda katılımcılardan NASA-TLX'in 6 alt boyutunu kendilerine verilen ikili karşılaştırma tabloları aracılığıyla önceliklendirmeleri talep edilmiştir. Çalışmada ağırlıklı NASA-TLX yönteminin kullanılmasının amacı, daha yoğun olarak hissedilen alt boyutun toplam skora etkisini artırmaktır.

Demografik bilgiler formu

Tez çalışmasında insan üzerinde deney yapılıyor olması sebebiyle, deneylere başlanmadan önce Gazi Üniversitesi Etik Komisyonu'ndan çalışmanın gerçekleştirilebilmesinde etik açıdan bir sakınca bulunmadığını gösteren "Etik Kurul İzni" alınmıştır. Bu kapsamda, katılımcılardan deney öncesi "Katılımcılar İçin Bilgilendirilmiş Gönüllü Olur Formu"nu (EK-1) imzalamaları istenmiştir. Ayrıca yine deney öncesinde katılımcılara "Demografik

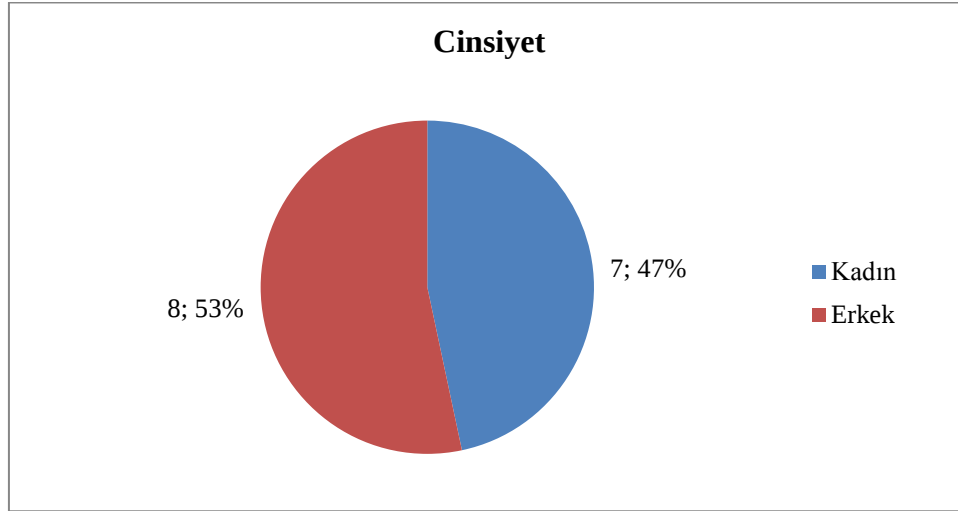
Bilgiler Formu” (EK-2) doldurtularak yaş, cinsiyet, sınıf gibi bilgileri talep edilmiştir. Demografik bilgilerin talep edilmesi ile; hissedilen iş yükünün, EEG sinyallerinin, göz hareketlerinin ve performansın, belirtilen katılımcı özelliklerine göre değişip değişmediğinin analiz edilmesi amaçlanmıştır. Aynı formda ilaç, alkol ve kafein kullanımı, görme problemi, nörolojik sorunlar vb. durumlar da sorgulanmış; böylelikle, deney sonuçlarını etkilemesinden kaçınılan faktörlerin önceden tespit edilerek katılımcı grubunun homojenliğinin sağlanması hedeflenmiştir. Katılımcıların tamamının benzer motivasyonla deneylere katılım sağladığı gözlem yoluyla da değerlendirilmiş, deneklerin tutum ve davranışlarında analizi etkileyebileceği düşünülen herhangi bir sorunla karşılaşılmamıştır.

4.1.2. Katılımcılar

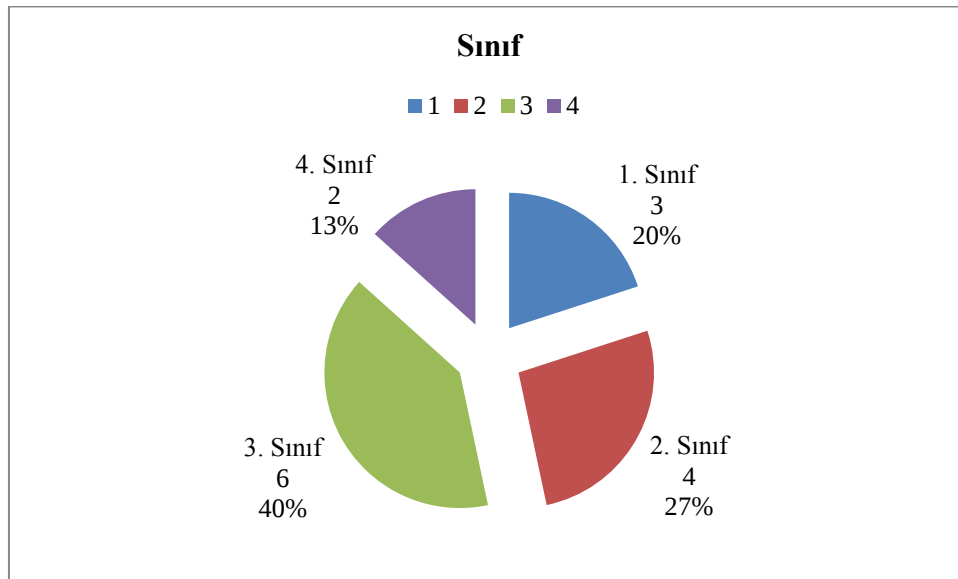
Deneyler başlangıçta herhangi bir nörolojik rahatsızlığı ve görme problemi bulunmayan (görme sorunu yok/gözlük veya lens ile düzeltilmiş) 19-25 yaş aralığında 25 (12 kadın, 13 erkek) üniversite lisans öğrencisi üzerinde gerçekleştirilmiştir. Deneklerin 22’si Gazi Üniversitesi Endüstri Mühendisliği Bölümü öğrencisi, 3’ü ise Orta Doğu Teknik Üniversitesi öğrencisidir. İlk gün deneye katılım sağlayan 3 kişiye ilişkin veriler, göz izleme dosyalarının eksik olması sebebiyle çalışmanın dışında tutulmuştur. Dolayısıyla tüm çalışmaya ilişkin veri hazırlama işlemleri 22 kişiye ait dosyalar üzerinden sürdürülmüştür. Veri ön işleme adımının ardından güvenilir EEG sinyaline ilişkin verinin girdi olarak kullanılmasının sağlanması amacıyla, analizlere yine 19-25 yaş aralığında 15 kişilik (7 kadın, 8 erkek; yaş ortalaması 21,6) veri setiyle devam edilmesine karar verilmiştir. Katılımcıların 3’ü birinci sınıf, 4’ü ikinci sınıf, 6’sı üçüncü sınıf, 2’si ise dördüncü sınıf öğrencisidir. Analizde kullanılan veri setinin ait olduğu katılımcıların cinsiyet ve sınıf dağılımları sırasıyla Şekil 4.4 ve Şekil 4.5’te verilmiştir.

Nielsen ve Landauer (1993) 15 denek kullanıldığında probleme ilişkin resmin hemen hemen tamamının netleşeceğini öne sürmüşlerdir. Örneklem büyüklüğü ile ilgili bir diğer çalışmada gözlem sayısının bağımsız değişken sayısının 15-20 katı olması gerektiği belirtilmiştir (Siddiqui, 2013). Bir diğer yaklaşım ise, olması gereken $n \geq 50 + 8(k)$ formülü ile hesaplanabileceğini öne sürmektedir (Green, 1991). Bu doğrultuda, benzer çalışmalarda kullanılan denek sayıları da göz önüne alınarak 15 katılımcıya ilişkin 2700

adet gözlemin EEG ve göz değişkenlerinden oluşan 111 değişkenli analiz için yeterli olduğu değerlendirilmiştir.



Şekil 4.4. Katılımcıların cinsiyet dağılımı



Şekil 4.5. Katılımcıların sınıf dağılımı

4.1.3. Deney prosedürü

Deneylemin tamamı Orta Doğu Teknik Üniversitesi İnsan Bilgisayar Etkileşimi Laboratuvarında aynı fiziksel koşullar altında gerçekleştirilmiştir. Laboratuvar ortamı yapay ışık (florasan) ile ışıklandırılmıştır ve ses yalıtımlıdır. Laboratuvarın görünümü Resim 4.3'te yer almaktadır.

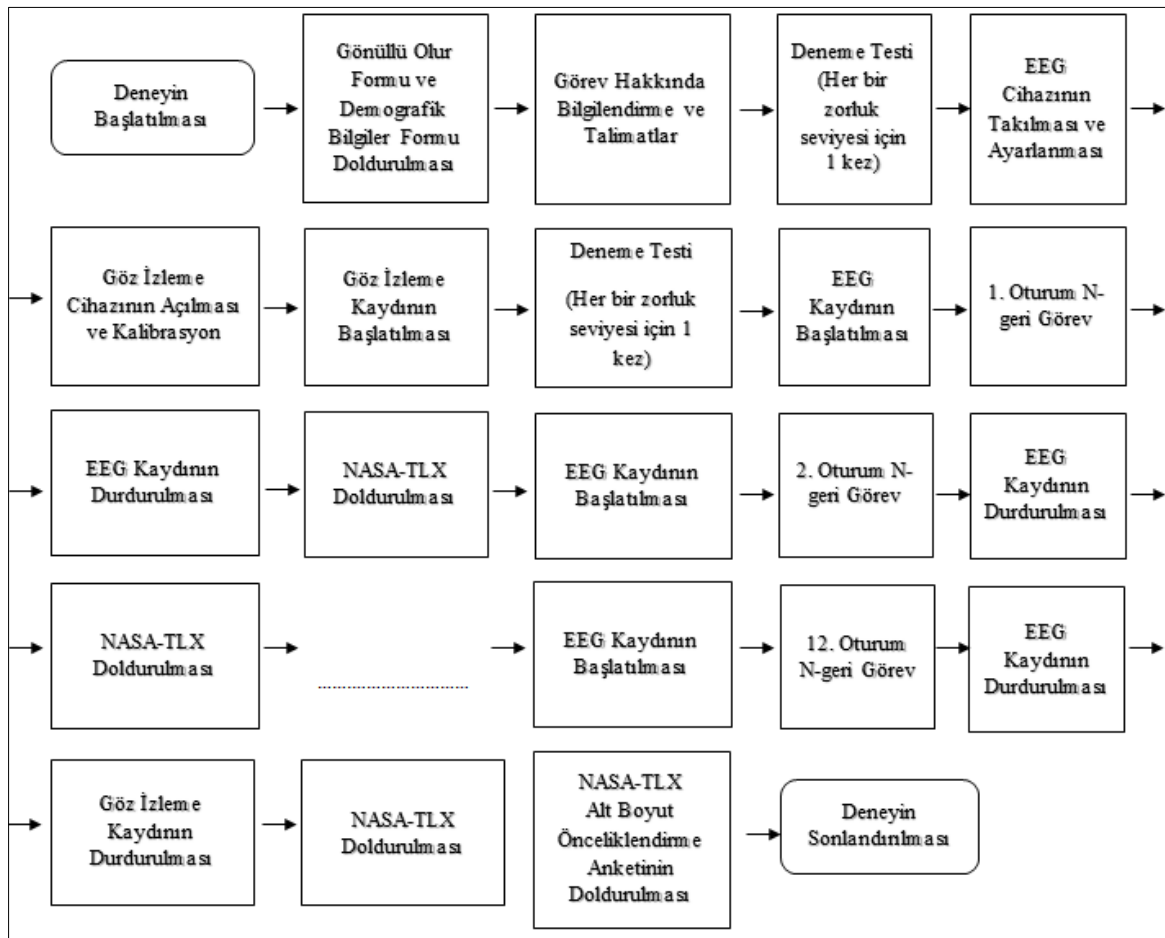
Katılımcılardan deneylere başlamadan önce “Katılımcılar İçin Bilgilendirilmiş Gönüllü Olur Formu” ile “Demografik Bilgiler Formu”nu doldurmaları istenmiştir. Ayrıca katılımcılara sözlü olarak, çalışmanın amacı, gerçekleştirilecek n-geri testler ve yapılması gerekenler hakkında bilgi verilmiştir. Görev başlamadan önce ekranda görevin nasıl gerçekleştirileceği ve zorluk seviyesiyle alakalı yazılı açıklamalar da yer almaktadır. Katılımcılara testlerin anlaşılmasını sağlamak amacıyla bir kez kayıt dışı bir kez de kayıt başladıktan sonra olmak üzere ikişer kez deneme testi uygulanmıştır. Ancak, kayda dâhil edilen deneme testi verisi analizin dışında tutulmuştur.



Resim 4.3. İnsan bilgisayar etkileşimi laboratuvarı

Deneme testlerinde her bir testin sonunda başarı yüzdesi ekrana gelmekte, gerçek testlerde ise herhangi bir geri bildirim yapılmamaktadır. Katılımcıdan görevi yapamadığını düşündüğünde moralini bozmaması istenmiş, önemli olanın yalnızca göreve odaklanıp mümkün olduğunca çaba göstererek tüm oyunları tamamlamaya çalışması olduğu vurgulanmıştır. Şayet belli bir anda, hafızasındaki harf zincirinde kopukluk yaşarsa gördüğü son harften itibaren oyuna devam etmeye çalışması talep edilmiştir. Uygulanan deney prosedürü ise Şekil 4.6’da verilmiştir.

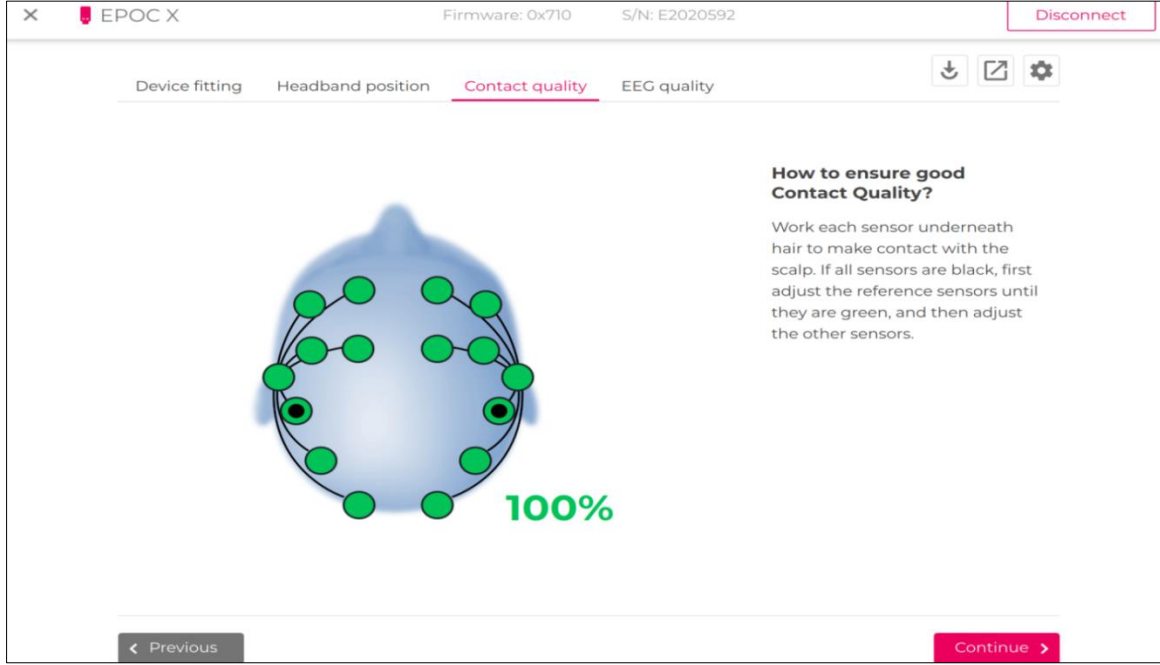
Oyunun anlaşılıp anlaşılmadığının test edilmesi amacıyla her bir zorluk seviyesi için birer kez uygulanan deneme testinin ardından katılımcının klavyeye kolay uzanabileceği şekilde rahatça oturması istenmiştir. Hem EEG hem de göz izleme kalitesinin sürdürülebilirliği için hareket etmeyen bir çalışma koltuğu kullanılmıştır. EEG başlığı katılımcının başına yerleştirilmiş, kendisinden sakın olması, deney boyunca konuşmaması ve kafasını mümkün olduğunca hareket ettirmemesi istenmiştir. Önce bağlantı kalitesi daha sonra EEG kalitesi kontrol edilmiş, bağlantı kalitesi %100 olmadan deneye başlanmamıştır. Bağlantı kalitesini gösteren ekran Şekil 4.7’de, EEG kalitesini gösteren ekran Şekil 4.8’de verilmiştir.



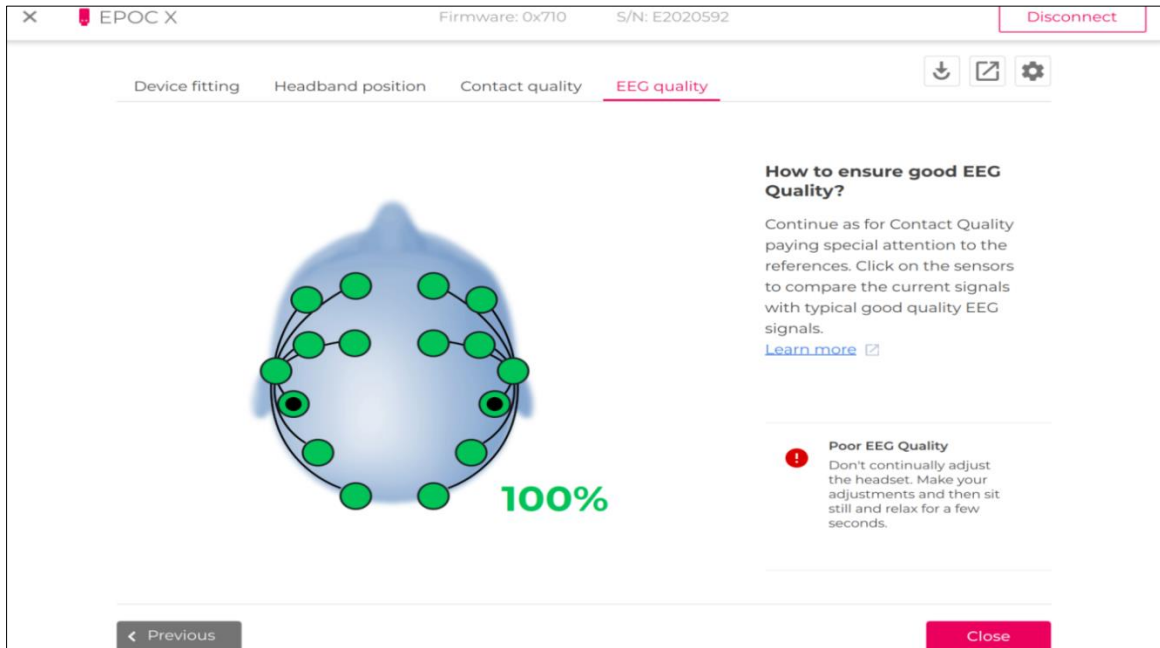
Şekil 4.6. Deney prosedürü

EEG sinyali göz kırpması ve diğer kas hareketleri gibi içsel kaynaklar ile cihazların güç adaptörleri gibi dışsal kaynaklar sebebiyle oluşan gürültülere karşı hassas olduğu için, %100 kaliteli EEG sinyalini deney boyunca sürdürebilmek mümkün değildir; ancak, veri ön işleme adımıyla verinin söz konusu gürültülerden arındırılması ve güvenilir EEG sinyaline ilişkin verinin girdi olarak kullanılması sağlanmıştır. Yine de değişken EEG

kalitesi deney boyunca takip edilmiş bazı durumlarda oturum aralarında kalitenin düştüğü elektrotlara NASA-TLX formunun doldurulması esnasında elle müdahale edilmiştir.



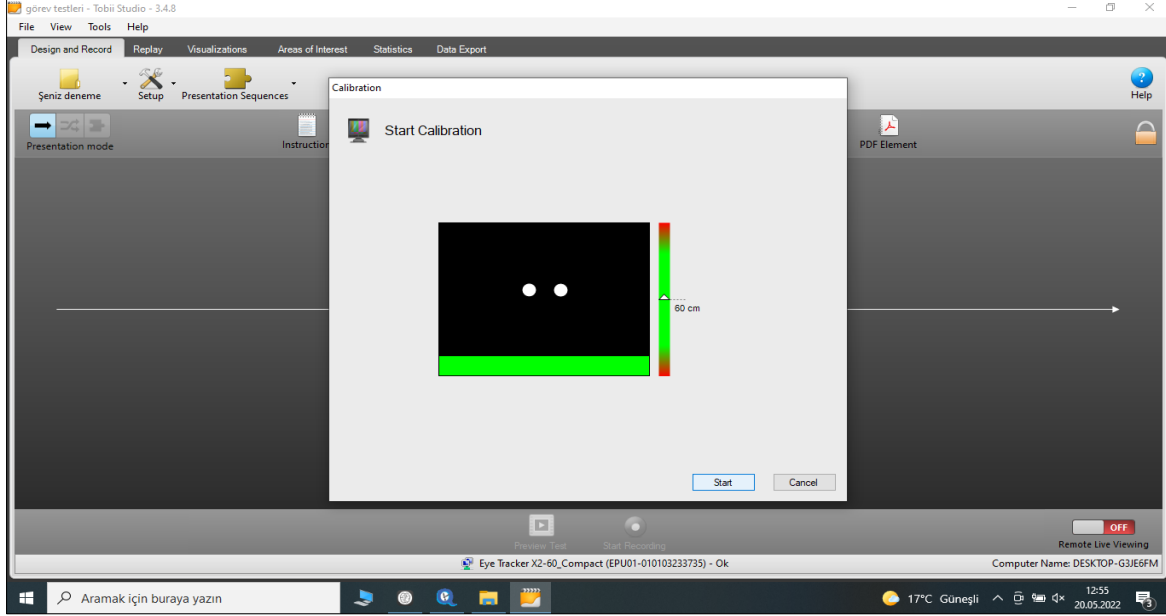
Şekil 4.7. EEG bağlantı kalitesi ekranı



Şekil 4.8. EEG sinyal kalitesi ekranı

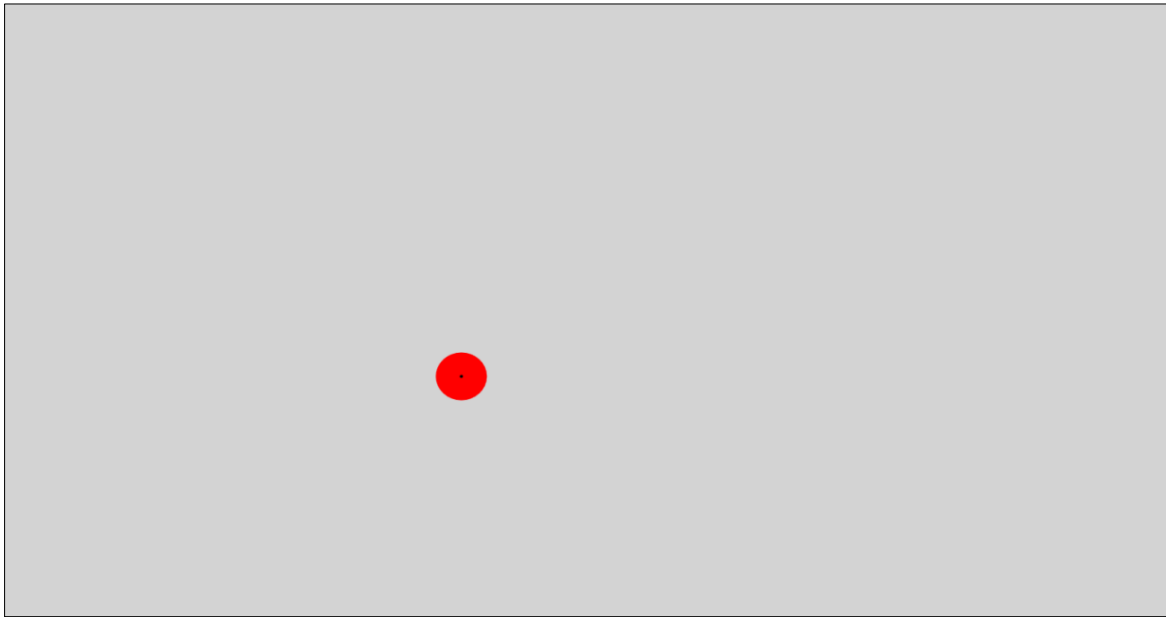
EEG cihazının kayda hazır hale getirilmesinin ardından göz izleme cihazı açılmış ve kalibrasyon öncesi gözün cihaza mesafesi kontrol edilmiştir. Tüm deneklerin gözleri

ekrana, dolayısıyla göz izleme cihazına 50-70 cm uzaklıkta olacak şekilde oturmaları sağlanmıştır. Mesafe, Şekil 4.9’da yeşil ile gösterilen sınırlara göre ayarlanmıştır.



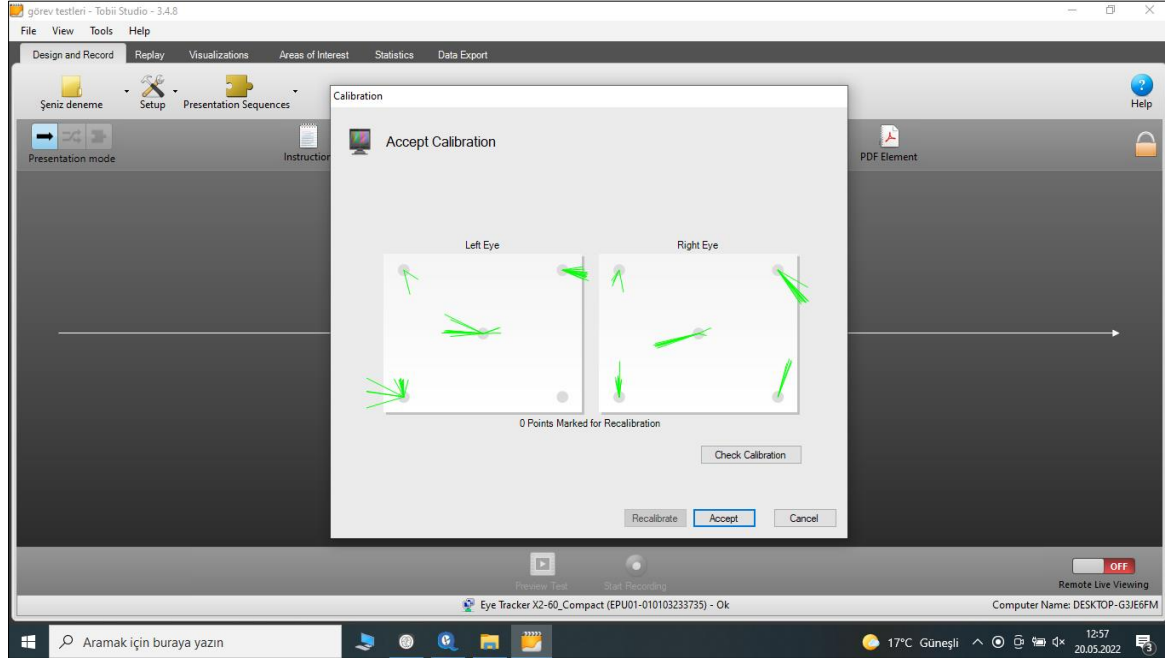
Şekil 4.9. Göz mesafesi belirleme ekranı

Kayda başlanmadan önce kalibrasyon gerçekleştirilmiştir. Kalibrasyon yapılırken katılımcıdan Şekil 4.10’da gösterildiği şekilde ekranda hareket eden kırmızı bir noktayı gözleriyle takip etmesi istenmektedir.



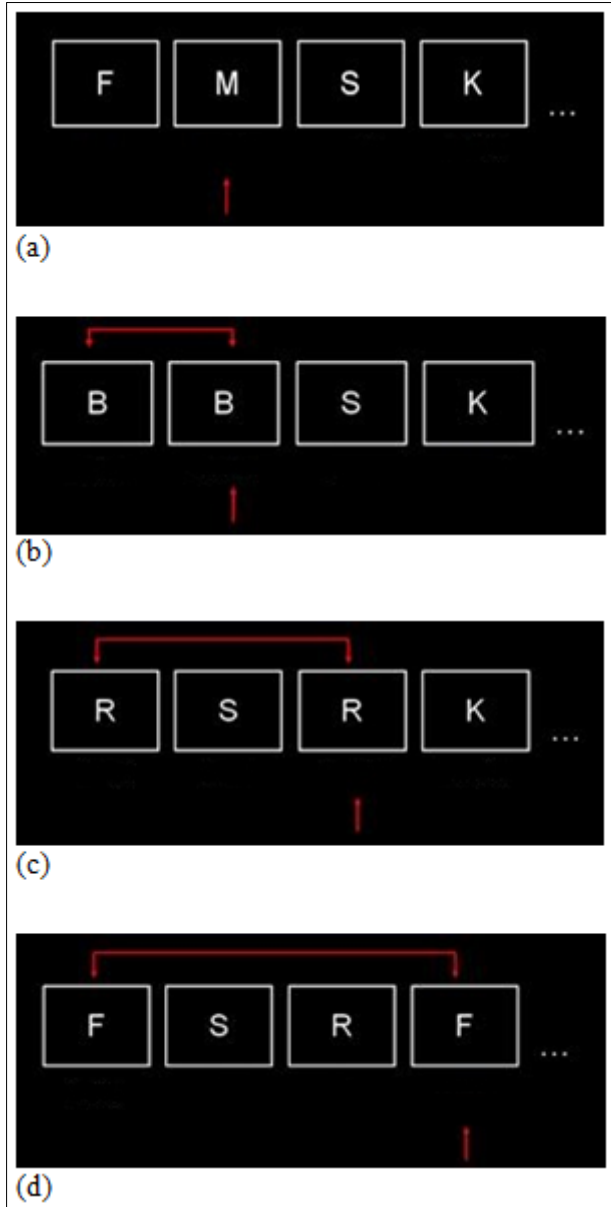
Şekil 4.10. Kalibrasyon ekranı

Kayıtlar, kalibrasyon seviyesi kontrol edildikten sonra başlatılmıştır. Şekil 4.11’de kalibrasyon onay ekranı görülmektedir.



Şekil 4.11. Kalibrasyon onay ekranı

4 seviyeden (0-geri, 1-geri, 2-geri, 3-geri) oluşan n-geri göreve ilişkin gerçek testlerde her bir zorluk seviyesi 3 kere uygulanmış, yani bir katılımcı tarafından söz konusu görev için toplamda 12 deneme gerçekleştirilmiştir. Deneme sırası, zorluk seviyesi rassal olacak şekilde uygulanmıştır. Rassal uygulanmasının sebebi, zihinsel iş yükünde meydana gelmesi beklenen dalgalanmanın zorluk seviyesinin önceden bilinmesi kaynaklı oluşabilecek yanlı bir duruma mahal vermeyecek şekilde gerçekleşmesini sağlamaktır. Katılımcılardan ekranda belli bir süre görünen harfleri hafızalarında tutup, 0-geri görevde M harfini gördüklerinde klavyede “M” tuşuna basmaları eğer harf M değil ise “L” tuşuna basmaları istenmektedir. 1, 2 veya 3 geri görevde ise; ekranda görünen harf sırasıyla 1 önceki, 2 önceki veya 3 önceki harfle aynı ise (hedef/eşleşme durumu) “M” tuşuna basmaları, eğer harf aynı değilse (hedef olmayan/eşleşmeme durumu) “L” tuşuna basmaları istenmektedir. Görev kuralları Şekil 4.12’de verilmiştir. Her bir harf ekranda 0,5 saniye görünmekte ve bir sonraki harf ekrana gelene kadar 2 saniye süre geçmektedir. Yani, katılımcının yanıt verebileceği toplam süre 2,5 saniyedir. 0-geri görevde ekrana toplam 15 harf, 1-geri görevde 16 harf, 2-geri görevde 17 harf, 3-geri görevde ise 18 harf gelmektedir. Her bir zorluk seviyesi için beklenen yanıt sayısı 15’tir.



Şekil 4.12. N-geri görev kuralları a) 0-geri b) 1-geri c) 2-geri d) 3-geri

Pilot çalışmalarda yalnızca eşleşme durumunda klavyede “A” tuşuna basılması diğer durumda herhangi bir şey yapılmaması istenirken, gerçek testlerde hem klavyede basılması istenen tuşlar “M” ve “L” olarak değiştirilmiş hem de daha çok performans verisi elde edilebilmesi amacıyla eşleşmeme durumu için de tepki verilmesi beklenmiştir. Harflerin “M” ve “L” olarak belirlenmesinin sebebi katılımcıların büyük bir kısmının sağ el kullanıyor olmasıdır. İşaret parmakları ile M’ye orta parmakları ile L’ye basmaları istenmiştir. Sol el kullanan 2 katılımcı ise, görevi rahat gerçekleştirebilmeleri için sol el işaret parmakları ile “M”ye (eşleşme durumu), sağ el işaret parmakları ile de “L”ye (eşleşmeme durumu) basmaları doğrultusunda yönlendirilmiştir.

Testlerde X, W gibi İngilizce harfler de içeren birbirinden farklı toplam 20 adet ünsüz harf kullanılmıştır. Eşleşme/eşleşmeme oranı 1/2 olarak belirlenmiştir. Yanıt beklenen 15 soru olduğu düşünüldüğünde, 5'inde eşleşme durumu 10'unda eşleşmeme durumu olacak şekilde harfler rassal şekilde ekrana gelmektedir. Katılımcılardan harf ekrana gelir gelmez mümkün olan en kısa sürede yanıt vermeleri istenmiştir.

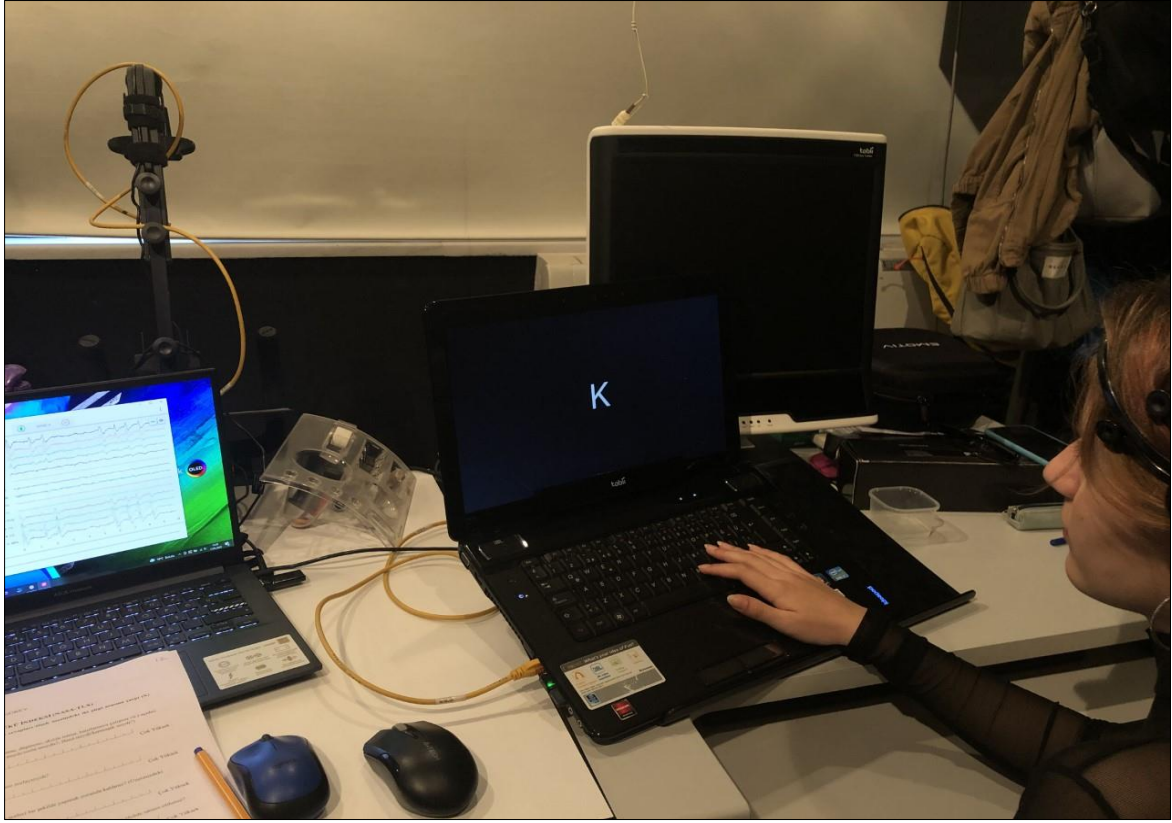
Her bir denemenin ardından NASA-TLX ölçeği (EK-3) aracılığıyla katılımcılardan testleri gerçekleştirirken deneyimledikleri zihinsel iş yükünü öznel olarak değerlendirmeleri istenmiştir. Dolayısıyla bir katılımcı her deneme sonunda 1 kez olmak üzere toplamda 12 kez NASA-TLX ölçeğini doldurmuştur. Bununla birlikte katılımcılardan NASA-TLX'in 6 alt boyutunu kendilerine verilen ikili karşılaştırma tabloları aracılığıyla önceliklendirmeleri talep edilmiştir. 15 ikili karşılaştırma neticesinde elde edilen ağırlıklar sayesinde ağırlıklı toplam NASA-TLX skorları elde edilmiştir. Bir katılımcı ile gerçekleştirilen kayıt süresi yaklaşık 20 dakikadır. Deneylerin uygulanışı esnasında çekilen bir fotoğraf Resim 4.4'te verilmiştir.

4.1.4. Elde edilen veri dosyaları

Her bir katılımcı için birer adet “Demografik Bilgiler Formu” mevcuttur. Kağıt üzerinde doldurulması istenen bu formdaki bilgilerden analizde kullanılacak olan cinsiyet, sınıf gibi değişkenlere ilişkin veriler bilgisayar ortamına girilmiştir.

Oturum aralarında doldurulması istendiği için NASA-TLX ölçeğinden de her bir katılımcı için 12 form bulunmaktadır. Bu formlardaki veriler de her bir alt boyut için bilgisayar ortamına girilmiştir. Formun son sayfasında bulunan alt boyut önceliklendirme formunda yapılan ikili karşılaştırmalar neticesinde ağırlıklandırılan skorların toplamı ile de “Ağırlıklandırılmış Toplam NASA-TLX Skoru” elde edilmiştir.

Inquisit Lab 6 aracılığıyla performans değişkenlerinin kayıt altına alınması sayesinde her bir katılımcı için birer adet performans dosyası elde edilebilmesi mümkün olmuştur.



Resim 4.4. Deneylerin uygulanışı

Tobii göz izleme cihazının, görevin gerçekleştirildiği dizüstü bilgisayara yerleştirilmesi gerektiği için göz izleme kayıtları da görevin gerçekleştirildiği bilgisayarda tutulmuştur. Tam ekran uygulanan ve 12 oturumdan oluşan tüm görev boyunca oyuna müdahale edilmesinin hem rassallık düzenini bozacak olması hem de Inquisit Lab yazılımının performans ölçümünü kesintiye uğratabilecek olması sebebiyle göz izleme kaydı 12 oturum boyunca kesintisiz bir şekilde alınmıştır. Dolayısıyla her bir katılımcı için birer adet göz izleme veri dosyası elde edilmiştir.

EEG kayıtlarının alınması ise cihazın bağlı bulunduğu bilgisayara bakılmasını gerektirmediğinden bağlantı ve EEG kalitesinin takibi amacıyla görevlerin gerçekleştirildiği bilgisayarın hemen yanındaki başka bir bilgisayarda tutulmuştur. Oturum aralarında oluşacak kafa hareketlerinden etkilenen sinyal kalitesi düzene girdikten sonra yeni oturumla beraber her seferinde EEG kaydı yeniden başlatılmıştır. Böylelikle 12 oturum için 12 adet EEG kayıt dosyası elde edilmiştir. Yalnızca test uygulanırken alınan kayda ilişkin dosyaların hem daha az gürültü içermesi ve hem de veri boyutu açısından yalınlığı sayesinde veri temizleme adımıındaki güçlüğü azaltması hedeflenmiştir. İki farklı

bilgisayardan alınan kayıtların eş zamanlılığının sağlanabilmesi için ise klavyede test başlangıç noktaları için işaretçiler belirlenmiştir. Inquisit testinin başlaması için “Space” tuşuna basılması gerektiğinden, göz izleme cihazının bulunduğu bilgisayarda “Space” tuşuna basıldığı anda EEG cihazının bağlı olduğu bilgisayarda da “S”ye basılmıştır. Başlangıç noktalarından itibaren cihazların örnekleme oranları (60 Hz ve 128 Hz) dikkate alınarak bitiş noktalarının belirlenmesi ve veri temizleme işlemi yapılması amaçlanmıştır.

4.1.5. Veri ön işleme ve değişken türetme

Göz izleme verisinin temizlenmesi

Tobii X2-60 cihazı 60 Hz frekansta (saniyede 60 kayıt) veri toplamaya imkan sağlamaktadır. Göz izleme cihazıyla katılımcılardan elde edilen ve başlangıçta her bir katılımcı için yaklaşık 70000 satır kayıttan oluşan excel dosyası üzerinde veri temizleme işlemi gerçekleştirilmiştir. Veri temizlemenin ilk aşamasında kayıt sırasında klavyede basılan tuşlardan yararlanılmıştır. Inquisit testinin başlaması için “Space” tuşuna basılması gerektiğinden, yazılımdan elde edilen raporda klavyeye basma olayı anlamına gelen “KeyPressEvent” değişkeninin “Space” olduğu kayıtlar her bir oturum için başlangıç kaydı olarak belirlenmiştir. Başlangıç noktalarından itibaren cihazın örnekleme oranı (60 Hz) dikkate alınarak bitiş noktaları belirlenmiş ve bu sayede oturum aralarındaki gereksiz kayıtlar silinmiştir. 12 ana oturumun gerçekleştirilmesinden önce her bir zorluk seviyesi için birer kez uygulanan test (practice) oturumlarına ilişkin kayıtlar da analizin dışında tutulmuştur. Satır sayıları azaltılarak elde edilen dosyalar her bir katılımcı için 29700 kayıttan (198 uyarı x 2,5 saniye x 60 kayıt) oluşmaktadır.

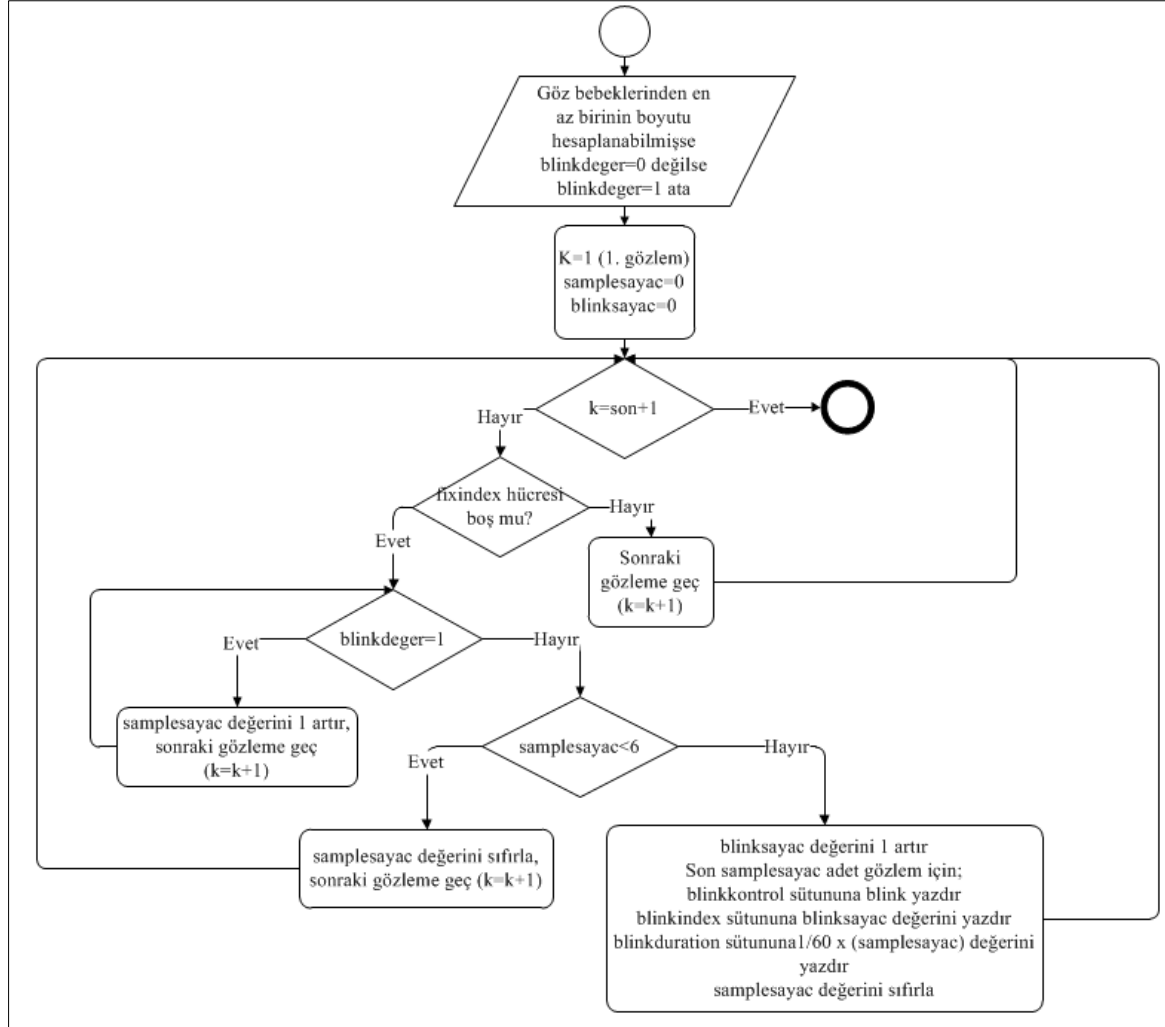
Tobii Studio yazılımı aracılığıyla “fixation”, “saccade” ve “unclassified” olmak üzere 3 türde göz olayı kayıt altına alınmaktadır. Dolayısıyla yazılım tarafından odaklanma ve seğirme hareketleri tespit edilebilmekte, göz kırpması verisi ise doğrudan elde edilememektedir. Bunun için, “unclassified” yani sınıflandırılmamış olarak tanımlanan kayıtlar için ayrı bir analiz yapılması gerekliliği ortaya çıkmıştır. Sağ ve sol göz bebeği boyutlarına ilişkin veriler de yazılım aracılığıyla doğrudan elde edilebilmektedir. Göz bebeklerine ilişkin ayrı ayrı ele alınan değişkenlere ek olarak her iki gözün ortalaması ve maksimumu alınarak elde edilen değişkenler de analize eklenmiştir.

Odaklanma ve seğirme ile ilgili değişkenlerin elde edilmesinde odaklanma indeksi (“Fixation Index”), seğirme indeksi (“Saccade Index”) ve göz hareketinin süresi (“Gaze Event Duration”) değişkenlerinden yararlanılmıştır. “Fixation Index” ve “Saccade Index” değişkenleri bir bakış olayının (odaklanma/seğirme) kaçınıcı odaklanma/seğirme olduğunu gösteren değişkenlerdir. Bir bakış olayının başlangıcından bitişine kadar olan sürede göz izleme cihazının pek çok kayıt alması sebebiyle aynı bakış olayını ifade ettiği halde tekrarlayan kayıtlar mevcuttur. Bu sebeple her bir harf uyarını süresince meydana gelen birbirinden farklı yani eşsiz kaç adet odaklanma/seğirme olduğunu belirlemek için “Fixation Index” ve “Saccade Index” değerinin yinelendiği satırlar kaldırılmıştır. Ayrıca odaklanma değişkenlerine ilişkin veriler analize dâhil edilmeden önce odaklanma süresi 80 ms altında ve 1200 ms üstünde olan kayıtlar analiz dışı bırakılmıştır (Borys ve diğerleri, 2017a; Borys ve diğerleri 2017b; Wu ve diğerleri, 2019; Yan ve diğerleri, 2019).

Göz kırpma, gözün tamamen veya kısmen kapanması durumudur (Ahlstrom ve Friedman-Berg, 2006). Göz kırpma hareketinin tespitine ilişkin bir çalışmada göz bebeğinin merkezi esas alınmıştır. Göz açıksa göz bebeği merkezi hesaplanabilmiş ve ilgili değişkene 1 değeri atanmış, göz kapalıysa merkez nokta hesaplanamamış ve değişken 0 değerini almıştır (Naveed, Sikander ve Khiyal, 2012). Bu yaklaşımdan hareketle, Tobii Studio yazılımının sağladığı sağ göz bebeği (“PupilRight”) ve sol göz bebeği (“PupilLeft”) değişkenlerinden yararlanılarak, söz konusu değişkenlere ilişkin hesaplama yapılabildiyse gözün açık olduğu, her iki değişkende de herhangi bir veri bulunmuyorsa gözün kapalı olduğu kabul edilmiştir. “blinkdeger” isimli bir değişken tanımlanarak bu değişkene gözün açık ve kapalı durumları için sırasıyla 0 ve 1 değerleri atanmıştır.

Bir göz kırpma toplamda 100-300 ms sürmekte, 500 ms üzerindeki göz kırpmalar ise uyuklama hali olarak kabul edilmektedir (Johns, 2003). Bu çalışmada kullanılan cihazın frekansının 60 Hz olması sebebiyle bir kayıt satırının süresi 1/60 saniyedir. Bu nedenle, 100 ms ve üzerindeki durumların göz kırpma olarak değerlendirilebilmesi için gözün kapalı kaldığı yani “blinkdeger” değişkeninin 1’e eşit olduğu kayıt sayısının minimum 6 olması gerekmektedir. Ek olarak, bir göz kırpma olayı iki odaklanma hareketi arasında gerçekleşeceği için (Ahlstrom ve Friedman-Berg, 2006) Tobii Studio yazılımının odaklanma yani “fixation” olarak belirlediği kayıtlar, göz kırpma taramasının dışında tutulmalıdır. Excel Visual Basic for Applications (VBA) aracılığıyla, belirtilen şartları sağlayan kayıtlar belirlenmiş ve her biri için kaçınıcı göz kırpma olduğunu gösteren “Blink

Index” değeri atanmıştır. Göz kırpma kayıtlarına karşılık gelen süreler ise, bir kayıt süresinin (1/60 saniye) her bir göz kırpma olayındaki kayıt sayısı ile çarpımı alınarak hesaplatılmış ve “blinkduration” değişkenine yazdırılmıştır. Göz kırpma hareketini tespit etmek için kullanılan makro kodunun akış şeması Şekil 4.13’te verilmiştir. Göz kırpma süresi 500 ms üzerinde tespit edilen kayıtlar analiz dışı bırakılmıştır (Johns, 2003).



Şekil 4.13. Göz kırpma hareketinin tespitine ilişkin algoritma

Bu şekilde göz kırpma yani “blink” olarak tespit edilen kayıtlar, bir sonraki aşamada farklı bir kod bloğuyla her bir uyarın bazında eşsiz şekilde saydırılmış ve yine uyarın bazında karşılık gelen toplam süreler belirlenmiştir. Bu aşamada Şekil 4.10’daki algoritma ile belirlenen “blinkindex” değişkeninden yararlanılmıştır.

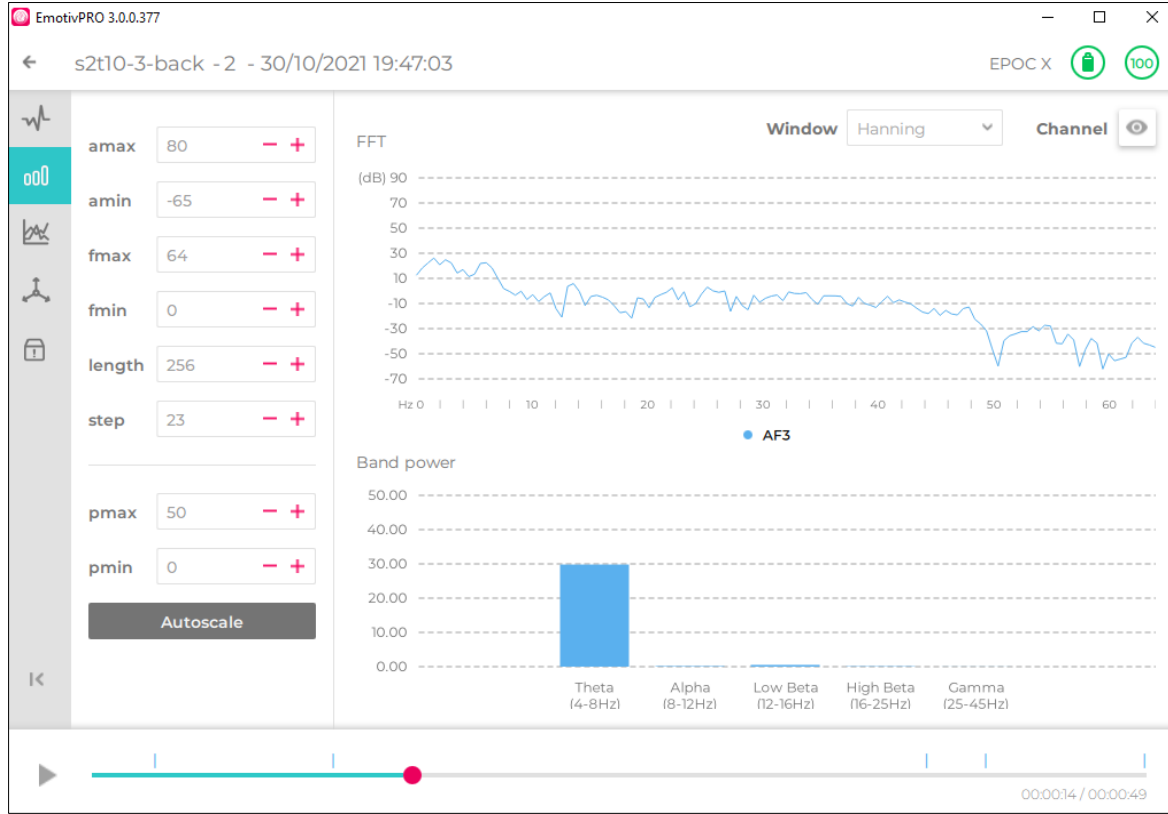
Pivot tablolar aracılığıyla her bir uyarın bazında odaklanma, seğirme, seğirme genliği, göz bebeği boyutu ve göz kırpma hareketleri ile ilişkili sayı değişkenleri ile sürelerine ilişkin

toplam, ortalama, maksimum, minimum ve standart sapma deęişkenleri elde edilmiştir. Odaklanma, seęirme ve göz kırpma deęişkenlerine ilişkin deęerlerin 0 olduęu gözlemlerde, söz konusu deęerler gözlemin içinde bulunduęu ve aynı zorluk seviyesine karşılık gelen oturumdaki ortalama deęerler ile deęiştirilmiştir.

EEG verisinin temizlenmesi

EEG sinyali göz kırpma, kas hareketleri, cihazların güç adaptörleri gibi gürültülere karşı hassas olduęu için, %100 kaliteli EEG sinyalini deney boyunca sürdürebilmek mümkün değildir. Yine de deęişken EEG kalitesi deney boyunca takip edilmiş bazı durumlarda oturum aralarında kalitenin düştüęü elektrotlara elle müdahale edilmiştir.

EMOTIV Pro yazılımı (URL-2) her bir kanal bazında gerçek zamanlı olarak veya kaydedilen veri üzerinde frekans analizi yapmakta ve grafiksel olarak seçilen her bir kanal için frekans bantlarındaki gücü göstermektedir. “HDF/Bant Gücü” ekranında gerçek zamanlı veya kayıt üzerinden izlenebilen görüntü Şekil 4.14’te verilmiştir. Dikkat edilecek olursa, HFD grafiğinde daha önce zaman ekseninde izlenen ham EEG verisi (Bkz. Şekil 4.3) frekans eksenine taşınmıştır. Her bir frekansın güç deęişimi dB (desibel) cinsinden görüntülenebilmektedir. HFD ile alt bantlarına ayrıştırılan EEG sinyallerinin her bir frekans bandı için elde edilen güç yoğunlukları da ekranın alt kısmında yer alan “Bant Gücü” grafiğinden izlenebilmektedir. Yazılım tarafından; teta bandı 4-8 Hz, alfa bandı 8-12 Hz, gama bandı 25-45 Hz frekans aralığı olarak ele alınmıştır. Beta bandı ise düşük beta (12-16 Hz) ve yüksek beta (16-25 Hz) olmak üzere iki alt banda ayrılmıştır. Çalışmanın devamında ele alınan deęişkenler bu bant aralıkları doğrultusunda analiz edilmiştir. “HDF/Bant Gücü” verisinin dış ortama aktarılması ile birlikte her bir kanal için beş farklı frekans bandına ilişkin güç verisine sayısal olarak erişebilmek mümkündür. Bant gücü verisi dış ortama 128 Hz frekanslı ham EEG verisinden farklı olarak 8 Hz frekans ile aktarılmaktadır.



Şekil 4.14. HFD ve bant gücü ekranı

Her bir katılımcı için oturum bazında elde edilen csv uzantılı 12’şer EEG dosyası öncelikle excel formatına dönüştürülmüştür. Bu dosyalarda işaretçi (marker) olarak belirlenen “S” klavye tuşuna basıldığı andaki kayıttan önceki satırlar silinmiştir. Cihazın saniyede 128 kayıt topladığı göz önüne alınarak oturumun bitiş noktası belirlenmiş ve bu kayıttan sonraki satırlar da silinerek; 0-geri görevler için 4800 (128 kayıt x 2,5 saniye x 15 uyarı), 1-geri görevler için 5120 (128 kayıt x 2,5 saniye x 16 uyarı), 2-geri görevler için 5440 (128 kayıt x 2,5 saniye x 17 uyarı) ve 3-geri görevler için 5760 (128 kayıt x 2,5 saniye x 18 uyarı) satırlık excel dosyaları elde edilmiştir. Her bir kanal için EEG kalitesini yüzde değer olarak gösteren değişkenler yazılım tarafından sağlanmaktadır. Bu değişkenlerin filtre olarak kullanıldığı pivot tablolar oluşturularak EEG kalitesinin %75’in üzerinde ölçüldüğü kayıtlar için yazılım tarafından sağlanan bant gücü değişkenlerinin ortalama değerleri uyarı bazında elde edilmiştir. “AF3.Teta” değişkeni AF3 kanalından elde edilen ortalama teta gücünü, “AF3.YüksekBeta” değişkeni ise AF3 kanalından elde edilen ortalama yüksek beta gücünü göstermektedir. Düşük EEG kalitesi (%75’in altı, genellikle %25) sebebiyle ortalama güç değişkeninin yanıt süresi boyunca elde edilemediği ve bu nedenle gözlemlere ilişkin eksik değerlerin fazla olduğu kişilere ilişkin EEG verisinin analizin dışında tutulmasına karar verilmiştir.

4.1.6. Analizde kullanılan değişkenler

Demografik değişkenler

Katılımcıya ilişkin cinsiyet ve sınıf bilgilerini içeren değişkenlerdir.

Göreve ilişkin bilgileri içeren değişkenler

Gözlemin hangi katılımcı tarafından, kaçınıcı oturumda, kaçınıcı uyaran esnasında gerçekleştiğini ve hangi zorluk seviyesine karşılık geldiğini gösteren değişkenlerdir.

EEG değişkenleri

14 kanal için ayrı ayrı 5 frekans bandındaki (teta, alfa, düşük beta, yüksek beta, gama) gücü gösteren 70 adet EEG değişkeni kullanılmıştır. Ek olarak literatürde teta/alfa gücü oranının zorluk seviyesi ile pozitif ilişkili olduğunu gösteren çalışmalar ışığında, her bir kanal bazında söz konusu orana karşılık gelen 14 değişken daha analize eklenmiştir (Guan ve diğerleri, 2021; Lim ve diğerleri, 2018). Örneğin, AF3 kanalındaki teta/alfa gücü oranını gösteren değişken “AF3.TetaAlfa” olarak belirtilecektir. EEG değişkenlerinin birimi dB olarak verilmiştir.

Göz izleme değişkenleri

Odaklanma, seğirme, göz bebeği boyutu ve göz kırpma ile ilişkili 27 adet göz izleme değişkeninden yararlanılmıştır. Bunlar:

- Odaklanma sayısı, odaklanma toplam süresi, odaklanma süresi ortalaması, standart sapması, maksimum ve minimum değerleri,
- Seğirme sayısı, seğirme toplam süresi, seğirme süresi ortalaması, standart sapması, maksimum ve minimum değerleri,
- Seğirme genliği ortalaması, standart sapması ve maksimum değeri,

- Sol göz bebeği çapı ortalaması, standart sapması ve maksimum değeri, sağ göz bebeği çapı ortalaması, standart sapması ve maksimum değeri, her iki göz için ortalama göz bebeği çapı ortalaması, standart sapması ve maksimum değeri,
- Göz kırpma sayısı, toplam ve ortalama süreleridir.

Göz izleme değişkenlerinden sayı olarak belirtilenlerin birimi adet, süre olarak belirtilenlerin birimi ms, çap olarak belirtilenlerin birimi ise mm olarak kullanılmıştır.

Özel değerlendirmeye ilişkin değişkenler

NASA-TLX'in 6 alt boyutuna ilişkin skorlar ve ikili karşılaştırmalar neticesinde elde edilen ağırlıklı toplam skor kullanılmıştır. Alt boyutlar tezin devamında aşağıdaki şekilde belirtilecektir:

- Zihinsel talep
- Fiziksel talep
- Zaman baskısı
- Performans
- Çaba
- Gerginlik

Performansa ilişkin değişkenler

Oturum bazında ele alınan ilk veri setinde Inquisit Lab 6 yazılımı aracılığıyla elde edilen aşağıdaki performans değişkenleri kullanılmıştır:

- Doğru Yanıt Sayısı
- Hedefi Bulma Sayısı (n-gerideki harf ile aynı harf olduğunu tespit etme)
- Hedef Olmayan Bulma Sayısı (n-gerideki harf ile aynı harf olmadığını tespit etme)
- Toplam Tepki Süresi (Bir oturum süresince beklenen 15 yanıtın toplamda kaç ms gecikme ile verildiği)
- Ortalama Tepki süresi (Bir oturum süresince beklenen 15 yanıtın ortalama kaç ms gecikme ile verildiği)

- Doğru Yanıt Oranı (Doğru yanıt sayısının beklenen yanıt sayısına (15) oranı)

Yukarıda belirtilen değişkenler oturum bazında hesaplanabildiği için uyaran bazında ele alınan ikinci veri setinde sadece “tepki süresi” ve “doğruluk” değişkenleri performans ölçüsü olarak kullanılmıştır. Tepki süresi; iki uyarının ekrana gelmesi arasında geçen 2500 milisaniyenin kaçınıcı milisaniyesinde yanıt verildiğine ilişkin değişkendir. Hiç yanıt verilemediyse tepki süresi 2500 milisaniye olarak alınmaktadır. Doğruluk ise, verilen yanıtın doğru olup olmama durumunu gösteren 0-1 değişkendir.

Sonuç olarak tezde; 2 adet demografik değişken, 84 adet EEG değişkeni, 27 adet göz değişkeni, 6 adet NASA-TLX değişkeni ve 6 adet performans değişkeni kullanılmıştır.

Analizler tezin amaçları doğrultusunda üç farklı veri seti üzerinde gerçekleştirilmiştir. Veri setlerine ilişkin detaylı açıklama tezin beşinci kısmı olan “Araştırma Bulguları ve Tartışma” bölümünde verilmiştir.

4.2. Makine Öğrenmesi

Bu bölümde; tezde kullanılan makine öğrenmesi yöntemine ilişkin kavramlar açıklanacak, makine öğrenmesi algoritmaları ve özellik seçim yöntemleri detaylandırılacaktır.

4.2.1. Makine öğrenmesi ve ilişkili kavramlar

Veri madenciliği (data mining); veri tabanı teknolojisi, istatistik, yapay zeka (artificial intelligence), makine öğrenmesi (machine learning), örüntü tanımlama (pattern recognition) ve veri görselleştirmesi (data visualization) gibi alanlar arasında köprü görevi gören çok disiplinli bir alandır (Özekes, 2003). Veri madenciliğinin yapay zekanın bir alt alanı olduğunu öne süren çalışmalar da mevcuttur. Yapay zeka, rasyonel ve akıllıca düşünen hesaplama araçları geliştirmeyi amaçlamaktadır. Yapay zekanın bilimsel hedefi, bilgi temsilinin akıllı davranmaya imkan sağlayacak ilkelerini anlamaktır. Yapay zekanın mühendislik hedefi ise, gerçek dünya sorunlarını insanlardan daha etkili ve verimli çözebilen hesaplama araçları yaratmaktır. Bu kapsamda yapay zeka, alt alanları olan bir araç kutusu olarak görülebilir (Schuh ve diğerleri, 2019).

Veri madenciliğinde kullanılan yöntemlerden bir kısmı klasik teknikler diyebileceğimiz istatistiksel yöntemlerdir. Diğer yöntemler ise genellikle istatistiği temel alan ama daha çok makine öğrenmesi ve yapay zeka destekli yeni nesil yöntemlerdir (Savaş, Topaloğlu ve Yılmaz, 2012).

Bir bilgisayar bilimi alanı olan makine öğrenmesi de yapay zekanın bir dalı olarak kabul edilmektedir (Voyant ve diğerleri, 2017). Yapay zekanın alt alanları olarak görülen makine öğrenmesi ve veri madenciliği arasındaki fark ise, makine öğrenmesi yöntemlerinin her ikisinde farklı amaçlar için kullanılmasıdır. Makine öğrenmesinde bilgi örtük olarak saklanmakta ve kullanılan yöntemler hesaplama araçlarının optimizasyonu amacına hizmet etmek için kullanılmaktadır. Veri madenciliğinde ise, bilginin verilerden elde edilmesi, açık bir şekilde depolanması, görselleştirilmesi, erişilebilir hale getirilmesi ve kullanıcı tarafından yorumlanabilmesi amacıyla makine öğrenmesi yöntemleri kullanılmaktadır (Schuh ve diğerleri, 2019). Dolayısıyla makine öğrenmesinin, değerli fakat örtük bilginin ortaya çıkarılmasını amaçlayan veri madenciliğinde kullanılan en etkin araçlardan biri olduğunu söylemek mümkündür.

Büyük boyutlu verilerin analiz ihtiyacı ile birlikte makine öğrenmesi teknikleri teknoloji, bilim ve ticarete en sık uygulanan yöntemlerden biri haline gelmiştir (Belkin, Hsu, Ma ve Mandal, 2019). Makine öğrenmesi yöntemleri, girdi değişkenleri ve çıktı değişkenleri olarak da bilinen bağımsız ve bağımlı değişkenler arasındaki ilişkiyi kullanarak doğrudan verilerden öğrenen uyumlu ve esnek modeller elde edilebilmeyi sağlamaktadır. Modelin öğrenmesi ifadesi ile bir veri setinde yer alan ve önceden bilinmeyen bağımlılıkların çıkarımı kastedilmekte, bu sayede yeni çıktı tahminlerinin yapılması mümkün olabilmektedir.

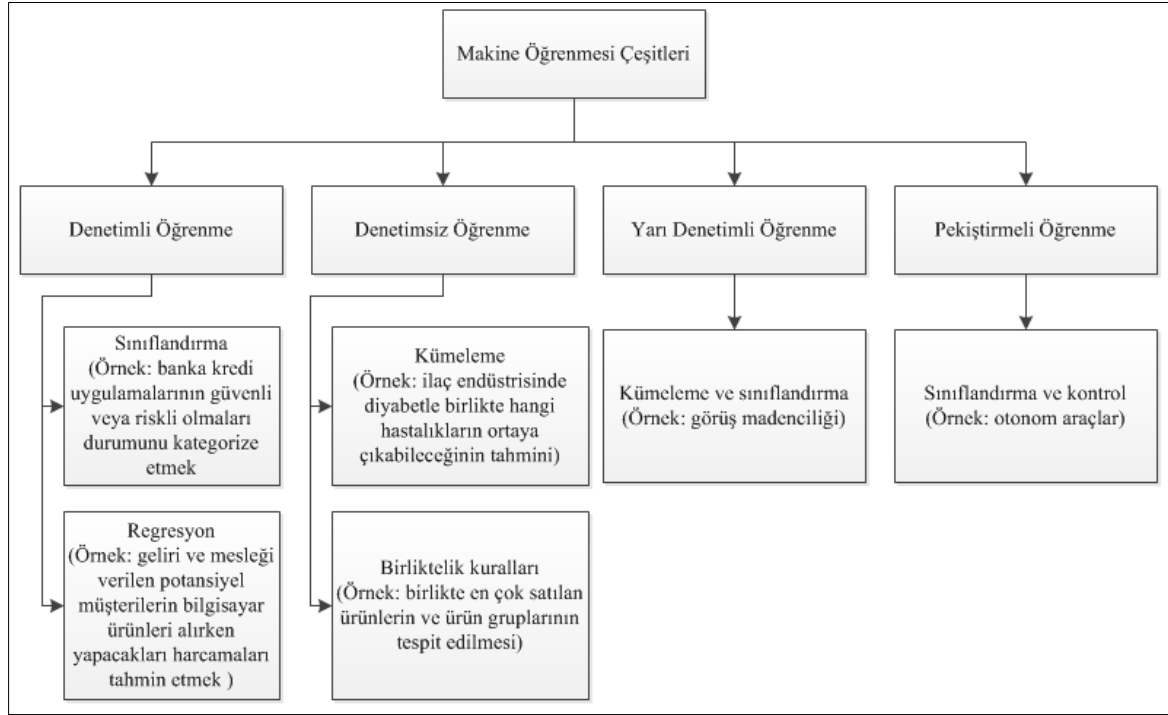
Veri madenciliğinde modeller temelde tahmin edici ve tanımlayıcı işlevlerine göre iki farklı grupta ele alınmaktadır. Tahmin edici modeller sınıflandırma ve regresyon modelleri, tanımlayıcı modeller ise birliktelik kuralları ve kümeleme modelleridir (Özekes, 2003). Tahmin edici modellerin denetimli (supervised) makine öğrenmesi yaklaşımı, tanımlayıcı modellerin ise denetimsiz (unsupervised) makine öğrenmesi yaklaşımı ile elde edildiklerini söylemek mümkündür. Denetimli öğrenmede; öngörülemeyen veriler için çıktı tahmini yapabilen genellenebilir bir algoritma elde etmek amacıyla, etiketlenmiş bir bağımlı (hedef) değişkenin olduğu veri setiyle eğitilen ve bu sayede bağımsız değişkenler

ile bağımlı değişken arasındaki ilişkileri bularak öğrenen modelin kurulması amaçlanmaktadır. Denetimsiz öğrenmede ise; etiketlenmemiş bir veri setinde bağımsız değişkenlere bakılarak anlamlı kümeler oluşturulması ve bu kümelerin analiz edilmesi amacıyla model kurulmaktadır. Bu sayede veri grupları ve gizli örüntüler keşfedilmektedir. Örneğin; ilaç endüstrisinde, diyabetle birlikte hangi hastalıkların ortaya çıkabileceğini tahmin etmek, birlikte en çok satılan ürün ve ürün gruplarını tespit etmek gibi problemlerin çözümü için denetimsiz öğrenmeden yararlanılabilmesi mümkündür (Çolakoğlu, 2020; Özçakır ve Çamurcu, 2007).

Temelde iki gruba ayrılan denetimli ve denetimsiz makine öğrenmesi yaklaşımlarına ek olarak yarı denetimli makine öğrenmesi ve pekiştirmeli makine öğrenmesi yaklaşımları da kullanılmaktadır. Yarı denetimli makine öğrenmesi yönteminden, denetimli öğrenme yöntemlerinin kısıtlarını ortadan kaldırmak amacıyla özellikle görüş madenciliği ve duygu analizi çalışmalarında yararlanılmaktadır. Yarı denetimli öğrenmede, büyük miktarda etiketsiz veri ile az sayıda etiketli veri eğitim amacıyla kullanılmaktadır (Onan ve Korukoğlu, 2016). Pekiştirmeli öğrenmede ise çevre ile etkileşime girerek öğrenebilecek bir sistem oluşturmak amaçlanmaktadır. Algoritmanın davranışı, tanımlanmış bir hedefe yönelik kararların doğru veya yanlış olmasına bağlı olan bir dizi ödül ve ceza yoluyla şekillenmektedir. Algoritmanın verilen örnekler ile eğitildiği denetimli öğrenmeden farklı olarak pekiştirmeli öğrenmede, algoritmanın hangi eylemlerinde ödülü en üst düzeye çıkarırken cezayı en aza indirdiğini keşfetmesi için deneme yanılma temelinde özgürce davranmasına izin verilmektedir (Vieira, Pinaya ve Mechelli, 2020). Pekiştirmeli makine öğrenmesi yaklaşımına otonom araçlar örnek verilebilir. Kendi kendine sürüş gerçekleştiren araçlarda kullanılan modeller, hangi rotada hangi hızla gidileceğini bu yaklaşım aracılığıyla öğrenmekte ve karar vermektedir (Çolakoğlu, 2020).

Şekil 4.15'te insan denetimi olup olmamasına göre kategorize edilen makine öğrenmesi çeşitleri şema halinde verilmiştir.

Tez çalışmasında tahmin edici modellerin elde edilmesi hedeflendiğinden, bu bölümde yalnızca regresyon ve sınıflandırma kavramları üzerinde durulmuş ve bu tür problemlerin çözümünde kullanılan makine öğrenmesi algoritmalarına yer verilmiştir.



Şekil 4.15. Makine öğrenmesi çeşitleri

Sınıflandırma, bilimsel araştırmanın ana konularından biridir. Bilimsel araştırmanın konusu olan her bir bilgi alanı, içeriğini sistematik bir şekilde yapılandırmak için sınıflandırma sistemlerine ihtiyaç duyar. Sınıflandırma kategorileri, öğelerin/nesnelerin belirli bir bağlamdaki benzerliklerine veya farklılıklarına göre tanımlanmaktadır. Sınıflandırma yaklaşımı, bilgiyi sıralama ve organize etmede, kapsamı tanımlamada ve çalışmaları basitleştirmede, bilgiye erişimi kolaylaştırmada ve yinelenen nesneleri tespit etmede kullanılmaktadır (Botchkarev, 2018). Bu doğrultuda, istatistiksel anlamda sınıflandırma problemlerinin çözümü gözlemlerin benzer/farklı özelliklerine göre kategorilerine ayrılmasına dayanmakta, şayet bir tahmin modeli kuruluyorsa modelin başarısı için gözlemlerin doğru kategoride tahmin edilebilmeleri önem taşımaktadır. Sınıflandırma problemleri, ikili (binary) sınıflandırma, çoklu sınıflandırma, çoklu etiketleme problemleri şeklinde üç gruba ayrılmaktadır. İkili sınıflandırmada bir gözlem iki sınıftan birine aittir. Çoklu sınıflandırmada bir gözlem ikiden fazla sınıf arasından sadece birine ait iken, çoklu etiketlemede birden fazla sınıfa ait olabilmektedir (Erickson ve Kitamura, 2021).

Regresyon ise iki ya da daha fazla değişken arasındaki ilişkinin fonksiyonunu belirlemekte kullanılan bir yaklaşım olup; tahmin modellerinde, açıklanmak istenen bağımlı değişken

bu fonksiyondan hareketle tahmin edilmektedir.

Sınıflandırmada kategorik değerler tahmin edilirken, regresyon süreklilik gösteren değerlerin tahmin edilmesinde kullanılmaktadır. Örneğin, bir sınıflandırma modeli banka kredi uygulamalarının güvenli veya riskli olmalarını kategorize etmek amacıyla kurulurken, regresyon modeli geliri ve mesleği verilen potansiyel müşterilerin bilgisayar ürünleri alırken yapacakları harcamaları tahmin etmek için kurulabilmektedir (Özekes, 2003). Regresyon modelleri; e-ticaret fiyatlandırma sistemleri, hava tahmini, borsa tahmini, süper çözünürlüklü görüntüleme ve görüntü sıkıştırma gibi geniş bir uygulama yelpazesine sahiptir (URL-5).

Bu aşamada, makine öğrenmesi çalışmalarında yaygın olarak kullanılan öğrenici ve sınıflandırıcı kavramlarını açıklamakta fayda görülmektedir. Öğrenici (learner) terimiyle genellikle modelin eğitim aşamasındaki durumuna vurgu yapılmaktadır. Sınıflandırıcı (classifier) ise tahmin yapabilen, yalnızca girdi verilerini alıp her bir örnek için çıktı verilerini üretebilen bir algoritmadır. Bir modeli uyumlama öğrenicinin işleviyken, değerleri tahmin etme işlevi sınıflandırıcıya aittir. Diğer bir deyişle, verilerden karar ağacını oluşturan program öğrenicidir, karar ağacı ise sınıflandırıcıdır. Regresör (regressor) ile sınıflandırıcının (classifier) aynı kavrama karşılık geldiğini, sadece model çıktısının doğası gereği farklılaştığını söylemek mümkündür.

Genellenebilirlik, bir modelin öğrendiği bilgileri daha önce hiç görmediği bir veri seti üzerinde kullanabilme kabiliyetidir. Kısacası, modelin eğitildiği veri seti dışında yeni ve farklı bir veri seti ile karşılaştığında ne ölçüde doğru tahmin yapabileceği ile ilişkili bir kavramdır. İyi bir model genellenebilir olmalıdır; ancak, aşırı uyum (overfitting) göstermemelidir. Aşırı uyum veya aşırı öğrenme, modelin ezberlediği anlamına gelmekte ve farklı bir veri seti ile karşılaştığında performansının düşmesine neden olmaktadır. Bu durum, modelin genellenebilirliğini azaltan bir faktördür. Aşırı öğrenen bir model, eğitim verilerine çok fazla uymakta ve bu nedenle eğitim verilerinde yüksek performans gösterirken test verilerinde düşük performans göstermektedir. Aşırı öğrenmenin önüne geçmek için, gözlem sayısını artırma, hiperparametre ayarlaması yapma, özellik seçim yöntemleri ile boyut azaltma ve erken durdurma gibi yöntemlere başvurulmaktadır. Modelin yeterince öğrenememesi (underfitting) de genellenebilirlik açısından önemli bir sorundur. Yetersiz öğrenmede hem eğitim hem de test performansı modelin temel

örüntüleri tam olarak keşfedememesi sebebiyle düşük çıkmaktadır. Bu sorunun çözümünde; modelin karmaşıklığını artırmak, veri sayısını artırmak ve hiperparametre ayarı yapmak önemlidir.

Makine öğrenmesi modelleri kurulurken veri seti rassal olarak eğitim ve test verisi şeklinde ayrılmaktadır. Eğitim ve test verileri yaygın olarak %70-%30, %75-%25 ya da %80-%20 oranlarında bölünmektedir. Model; eğitim verisi ile eğitilmekte, test verisi ile de modelin performansı test edilmektedir. Bu sayede modelin daha önce hiç görmediği bir veri seti ile denenmesi mümkün olmaktadır. Ancak bu rassal bölme işlemi yüksek varyansa eğilimli olması sebebiyle farklı test kümeleri için farklı sonuçlar verebilmektedir. Çünkü bazı bölmelerde kolay sınıflandırılabilen örnekler test kümesine girerken bazı bölmelerde zor sınıflandırılabilen örnekler test kümesine dâhil olmaktadır. Bu durumla başa çıkabilmek ve sonuçların güvenilirliğini artırmak için çapraz doğrulama (çd) yöntemine başvurulmaktadır. Çapraz doğrulama, modelin genellenebilirliğini ve güvenilirliğini sağlamada kullanılan yaygın bir yaklaşımdır. Literatürde k katlı çapraz doğrulama olarak kullanılan yöntemde veri seti k adet eşit parçaya bölünmektedir. Her bir parça birer kez test edilecek şekilde k farklı eğitim-test seti türetilmekte, modelin performansı her bir test seti üzerinde ayrı ayrı değerlendirilmektedir. Bu sayede, modelin performansı farklı veri parçaları üzerinde test edilerek daha güvenilir bir değerlendirme yapılabilir. En sık kullanılanlar, 5-katlı ve 10-katlı çapraz doğrulama yöntemleridir (Lim ve diğerleri, 2018; Jusas ve Samuvel, 2019; Plechawska-Wojcik ve diğerleri, 2019; Pei ve diğerleri, 2021).

4.2.2. Makine öğrenmesi algoritmaları ve hiperparametreleri

Doğrusal ve doğrusal olmayan regresyon ve sınıflandırma modellerinin çözümü için farklı algoritmalar kullanılmaktadır. Doğrusal modellerin çözümü için basit doğrusal regresyon, çoklu doğrusal regresyon, ridge regresyon, lasso regresyon, elasticnet regresyon ve lojistik regresyon algoritmaları kullanılmaktadır. Doğrusal olmayan modellerin çözümü için ise k en yakın komşu, destek vektör makineleri, yapay sinir ağları, karar ağaçları, rastgele ormanlar, gradyan artırma makineleri, hafif gradyan artırma makineleri, aşırı gradyan artırma ve kategori artırma algoritmalarından yararlanılmaktadır.

Alması gereken değer, modeli tasarlayan kişiye bırakılan, probleme ve veri setine göre değişiklik gösteren parametreler hiperparametre olarak adlandırılmaktadır (URL-6). Model

tahmin performansını arttırmak, modelin öğrenme süresini optimize etmek ve aşırı öğrenmenin önüne geçmek için hiperparametrelerin çalışmaya özgü şekilde ayarlanması gerekmektedir. Aşağıda çalışma prensipleri açıklanan makine öğrenmesi algoritmalarında model başarımı için optimize edilmesi gereken hiperparametrelere değinilmiştir.

Basit doğrusal regresyon

Amaç, bir bağımsız değişken ile bir bağımlı değişken arasındaki ilişkiyi modellemek ve bağımlı değişkenin model tarafından tahmin edilen değeri ile gerçek değer arasındaki farkı yani hatayı minimize etmektir. Basit doğrusal regresyon tahmin modeli Eş. (4.1) ile gösterilmektedir (URL-7; Kalaycı, 2008).

$$\hat{y}_i = b_0 + b_1 x_i \quad (4.1)$$

$\hat{y}_i$, i. gözlem için tahmin edilen bağımlı değişkenin değerini, x_i ise, i. gözleme ilişkin bağımsız değişkenin değerini göstermektedir. Farklı x_i değerleri için $\hat{y}_i$ değerini en az hatayla tahmin edebilen b_0 ve b_1 değerlerini elde etmek amaçlanmaktadır.

$$y_i = b_0 + b_1 x_i + e_i \quad (4.2)$$

$$e_i = y_i - \hat{y}_i \quad (4.3)$$

Eş. (4.2) ile gösterilen örneklem modelinde y_i , bağımlı değişkenin gerçek değerlerini vermektedir. Model sonucunda en iyi b_0 ve b_1 değerlerinin bulunması ile birlikte, Eş. (4.3) ile belirtilen hata değişkeni (e_i) en azlanmış olacaktır.

Hata kareler toplamını (HKT) en azlayan değişken katsayısının (b_1) ve regresyon sabitinin (b_0) belirlenmesi, modelin başarısını gösteren bir ölçüdür. Eş. (4.4) HKT performans ölçüsüne ilişkin denklemi göstermektedir. Denklemden toplam gözlem sayısı n ile gösterilmiştir.

$$HKT = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (4.4)$$

Çoklu doğrusal regresyon

Amaç, iki veya daha fazla bağımsız değişken ile bağımlı değişken arasındaki ilişkiyi modellemek ve bağımlı değişkenin model tarafından tahmin edilen değeri ile gerçek değer arasındaki farkı yani hatayı minimize etmektir. Çoklu doğrusal regresyon tahmin modeli Eş. (4.5) ile gösterilmiştir. Gözlem sayısı (i) ve değişken sayısı (p) arttıkça b katsayılarını bulmak zorlaşmaktadır.

$$y_i = b_0 + b_1 x_{i1} + b_2 x_{i2} + \dots + b_j x_{ij} + \dots + b_p x_{ip} \quad (4.5)$$

Varsayımlarının olması ve aykırı gözlemlere duyarlı olması bu yöntemin dezavantajlarını oluşturmaktadır. Çoklu doğrusal regresyon yöntemi, aşağıdaki varsayımlara sahiptir (URL-7; Kalaycı, 2008):

- Hatalar normal dağılıma sahiptir.
- Hatalar birbirinden bağımsızdır ve aralarında otokorelasyon yoktur.
- Her bir gözlem için hata terimleri varyansları sabittir.
- Değişkenler ile hata terimleri arasında ilişki yoktur.
- Bağımsız değişkenler arasında çoklu doğrusal ilişki problemi yoktur.

Ridge regresyon (L2)

Amaç, HKT'yi minimize etmektedir. Ancak ridge regresyonda HKT formülüne, değişken katsayılarının kareleri toplamının bir ayar parametresi sabitiyle (α) çarpımı eklenerek bir düzeltme işlemi yapılmıştır. Ayar parametresi kullanıcı tarafından belirlenmektedir. Eş. (4.6) ile ridge regresyon modelindeki HKT denklemi gösterilmektedir.

$$HKT = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 + \alpha \underbrace{\sum_{j=1}^p \beta_j^2}_{\text{ceza terimi}} \quad (4.6)$$

Bu şekilde hata varyansı azaltılmaktadır. Çoklu doğrusal bağlantı problemi olduğunda yani bağımsız değişkenler arasında doğrusal bir ilişki olduğunda etkili bir yöntemdir. Tüm değişkenler ile modeli kurmakta ancak ilgisiz değişkenlerin katsayılarını sıfıra yaklaştırmaktadır. Bu sayede değişken önem seviyelerini karşılaştırmaya ve değişken

seçimi yapmaya yarar (Hoerl ve Kennard, 2000; URL-7).

Lasso regresyon (L1)

1996 yılında öne sürülen bir regresyon analizi ve değişken seçimi yöntemidir. HKT formülüne, değişken katsayılarının mutlak değerleri toplamının bir ayar parametresi sabitiyle çarpımı eklenerek bir düzeltme işlemi yapılmıştır. Eş. (4.7) ile lasso regresyon modelindeki HKT denklemi gösterilmektedir.

$$HKT = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 + \alpha \underbrace{\sum_{j=1}^p |\beta_j|}_{\text{ceza terimi}} \quad (4.7)$$

Ceza terimi eklenerek düzeltilmiş bu formülün minimize edilmeye çalışılması nedeniyle doğası gereği değişken katsayıları küçülecek ve bazıları sıfıra eşitlenecektir. Böylelikle yorumlanabilir bir final modeli elde edilirken tahmin varyansı da azaltılmış olacaktır. L1 olarak da adlandırılan lasso regresyon yöntemi sayesinde ilgisiz değişkenlerin modelde yer almasının önüne geçilmiş olur (URL-7; Tibshirani, 1997).

ElasticNet regresyon

Bir başka düzenleme ve değişken seçim yöntemidir. L1 ve L2 yöntemlerinde kullanılan ceza terimlerinin her ikisi de HKT formülüne eklenerek düzeltme işlemi yapılır. Eş. (4.8) ile elasticnet regresyon modelindeki HKT denklemi gösterilmektedir.

$$HKT = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 + \alpha_1 \underbrace{\sum_{j=1}^p \beta_j^2}_{\text{ceza terimi}} + \alpha_2 \underbrace{\sum_{j=1}^p |\beta_j|}_{\text{ceza terimi}} \quad (4.8)$$

Bu denklemde iki tane ayar parametresi (α_1 ve α_2) yer almakta ve bu durum hangi ceza teriminin etkisinin daha fazla olacağını belirlemeye fırsat vermektedir. Özellikle değişken sayısı gözlem sayısından büyük olduğunda etkili bir yöntemdir (URL-7; Zou ve Hastie, 2005).

Lojistik regresyon

Cox (1958) tarafından geliştirilmiştir. Lojistik regresyonda bağımlı değişken kategoriktir. Yöntem, doğrusal modelin katsayılarının belirlenebilmesi için maksimum olabilirlik tahminini kullanmaktadır (Benerradi, Maior, Marinescu, Clos ve Wilson, 2019). Her bir gözleme ilişkin olarak bağımlı değişkenin alabileceği sınıf değerlerinin olasılıkları hesaplanır ve test edilen gözlemlerin bilinmeyen sınıfları belirlenir. Çoklu doğrusal regresyonun sınıflandırma problemindeki karşılığıdır. Amaç, bağımlı ve bağımsız değişkenler arasındaki ilişkiyi tanımlayan doğrusal bir model kurmaktır. Doğrusal regresyonda aranan varsayımlar burada aranmadığı için daha esnek kullanım sağlar. Bağımlı değişkenin 1 olarak tanımlanan değerinin gerçekleşme olasılığı hesaplanmaktadır. Dolayısıyla bağımlı değişkenin alacağı değerler ile ilgilenilmez. Lojistik fonksiyon sayesinde üretilen değerler 0 ile 1 arasında olmaktadır (URL-7).

İş yükü seviyelerinin belirlenmeye çalışıldığı bir sınıflandırma probleminde lojistik regresyon yaklaşımı kullanılarak destek vektör makineleri ile elde edilene yakın bir sonuç elde edilmiş ve basit bir yaklaşım olmasına rağmen lojistik regresyonun hız avantajı sayesinde diğer algoritmaların yanında yer alabildiği ortaya konmuştur (Benerradi ve diğerleri, 2019).

K en yakın komşu (KEYK)

Gözlemlerin birbirlerine olan benzerlikleri üzerinden tahmin yapmak üzerine kurulu bir yöntemdir. K en yakın komşu algoritmasında k sayısı belirlenmekte ve her bir gözlemin kendisine en yakın k komşusuyla olan uzaklıkları (örneğin öklit uzaklığı) hesaplanmaktadır. Sınıf etiketi bilinmeyen gözlemin sınıfı, uzaklığının en az olduğu komşuların sınıf etiketlerine göre belirlenmektedir. Sınıflandırmada en sık sınıfa, regresyonda ise ortalama değere göre tahmin yapılmaktadır. Bu algoritmada, kaç adet komşu gözlemin göz önünde bulundurulacağını gösteren “n_neighbors” yani k parametresinin optimizasyonu model performansını artırmak için önem taşımaktadır. Bu algoritma, veri seti çok büyük olduğunda yetersiz kalabilmektedir (URL-7).

EEG yöntemi kullanılarak zihinsel iş yükünün sınıflandırıldığı bir çalışmada diğer üç algoritmaya (destek vektör makineleri, rastgele ormanlar, karar ağaçları) kıyasla en iyi

doğruluk oranı KEYK algoritmasıyla elde edilmiştir (Plechawska-Wojcik ve diğerleri, 2019).

Destek vektör makineleri (DVM)

Vapnik ve Chervonenkis (1991) tarafından geliştirilen bir algoritmadır. Amaç; bir marjın aralığına, en fazla sayıda gözlemi en küçük hata ile dâhil edebilecek doğruyu ya da eğriyi belirlemektir. Veriyi en doğru şekilde ayrıştırabilen bir hiper düzlem bulunması hedeflenmektedir. Algoritma bunu yaparken eş zamanlı olarak, hiper düzlemin kendisine en yakın olan her bir veri noktasından uzaklığını maksimize etmeye çalışmaktadır (Benerradi ve diğerleri, 2019). Destek vektör makineleri modelinin gösterimi Eş. (4.9), Eş. (4.10), Eş. (4.11) ve Eş. (4.12) ile verilmiştir.

Amaç fonksiyonu:

$$\text{Min } \frac{1}{2} \|W\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*) \quad (4.9)$$

Kısıtlar:

$$y_i - (w_i \times x_i) - b \leq \varepsilon + \xi_i \quad (4.10)$$

$$(w_i \times x_i) + b - y_i \leq \varepsilon + \xi_i^* \quad (4.11)$$

$$\varepsilon, \xi_i \geq 0 \quad (4.12)$$

Denklemlerde w , i . gözlem için x bağımsız değişkenlerinin katsayılarını; b , regresyon sabitini; ε , regresyon doğrusu/eğrisinden uzaklığı gösteren marjın aralığını; C , ceza parametresini, ξ_i ise i . gözlemin marjın aralığından uzaklığını göstermektedir.

DVM algoritması küçük veya orta büyüklükteki veri setleri için uygundur. Aykırı gözlemlere karşı dayanıklıdır. Uygulanmadan önce farklı ölçü birimlerine sahip değişkenlerin standartlaştırılması gerekmektedir (URL-7; URL-8). Model optimizasyonu sürecinde, ceza parametresi olan “ C ” ile “kernel” parametresi için ayarlama yapılmalıdır. Kernel, sınıfları ayırmaya yarayan hiper düzleme ilişkin fonksiyondur. İki boyutta açıklanamayan değişimleri boyut arttırarak çözüyormuş gibi yapılan hilelere “Kernel Trick” denilmektedir. İki boyutta açıklayamadığımız veri setimizi daha fazla boyutta açıklamak için kullanılan kernel hilesi metoduna “polinomial kernel” denir. Her bir noktanın belirli bir noktaya ne kadar benzediğini normal dağılım ile hesaplayan, ona göre

sınıflandıran kernel trick metoduna ise “radyal tabanlı fonksiyon” denilmektedir. Hiperparametreleri belirlerken doğrusal kernel için “linear”, polinomiyal kernel için “poly”, doğrusal olmayan radial basis function için ise “rbf” yazılmalıdır. Hiç bir gözlemin marjin içinde yer almadığı marjine “hard margin”, bazı gözlemlerin marjin içerisinde yer aldığı duruma ise “soft margin” denilmektedir. İkisi arasındaki denge “C” parametresi ile sağlanır. C arttıkça marjin daralır ancak model aşırı öğrenmeye meyilli olur. Model aşırı öğrenirse C’nin azaltılması gerekir (URL-8).

Literatürde göz izleme veya EEG verilerinin kullanıldığı ve en iyi zihinsel iş yükü sınıflandırma sonuçlarının DVM ile elde edildiği çalışmalar mevcuttur (Borys ve diğerleri, 2017a; Kaczorowska ve diğerleri, 2020; Shahid ve Rasool, 2020).

Yapay sinir ağları (YSA)

Yapay sinir ağlarının temel mantığı biyolojik sinir sistemi ve insan beyninin yapısına dayanmaktadır. Yapay sinir ağlarındaki nöronlar hesaplama bileşenleri ya da işlem birimlerini temsil etmektedir. Beyin ise, nöronların bağlantı yollarını ayarlayarak bilgiyi depolamakta ve işlemektedir. Bir yapay sinir ağı modeli girdi katmanı, gizli katman ve çıktı katmanından oluşmaktadır. Katmanlardaki her bir nöron bir sonraki katmandaki tüm nöronlar ile bağlantılıdır. Yapay sinir ağlarında tahmin değerinin hedeflenen değerle eşleşebilmesi amacıyla, birçok girdi verisi ve çıktı hedefi kombinasyonu kullanılmaktadır. Ağ çıktısı ile gerçek çıktı arasındaki farkın minimize edilebilmesi için girdilerin ağırlıkları değiştirilmekte ve bu işlem kabul edilebilir bir hata miktarına ulaşıncaya kadar devam etmektedir. Bu şekilde, bir sinir ağının nöronlar arası bağlantı değerlerini (ağırlıkları) ayarlayan belli bir fonksiyon uygulanarak eğitilebilmesi mümkündür. Çıktı katmanındaki çıktılar bir toplama ve aktivasyon fonksiyonu aracılığıyla üretilmektedir. Girdi katmanında ise herhangi bir bilgi işleme operasyonu gerçekleşmemektedir (Wu ve diğerleri, 2019; Yan ve diğerleri, 2019). Sınıflandırma problemlerinde aktivasyon fonksiyonu belirli bir sınıfa ait olma olasılık değerlerini elde etmek için kullanılmaktadır. En bilineni sigmoid fonksiyonudur ve 0-1 arasında değer almaktadır (URL-7). Yapay sinir ağları pek çok zihinsel iş yükü tahmin ve sınıflandırma çalışmasında kullanılmıştır (Baldwin ve Penaranda, 2012; Tjolleng ve diğerleri, 2017; Wu ve diğerleri, 2019; Yan ve diğerleri, 2019).

En önemli hiperparametreleri ceza parametresi anlamına gelen “Alfa” ve gizli katman boyutlarını gösteren “hidden_layer_sizes” parametreleridir. Ayrıca ağırlıkları optimize etmek için kullanılacak olan yöntemi belirlemeye yarayan “solver” ve aktivasyon fonksiyonunu gösteren “activation” parametresinin ayarlanması da model performansını artıracaktır. Solver olarak “adam” büyük boyutlu veriler için uygundur. Şayet veri seti küçükse “lbfgs” seçilmesi daha hızlı ve iyi performans verecektir. Aktivasyon fonksiyonu olarak regresyonda “relu” kullanılmaktadır, sınıflandırmada ise “logistic” kullanılması uygun olacaktır (URL-7).

Karar ağaçları (KA)

Hedef değişkenin tahmin değerine ulaşmak amacıyla belirlenen kuralların karar noktaları olarak düğümlerde toplandığı akış şemasına benzeyen ağaç yapılarına karar ağaçları denilmektedir. Bu yapılarda, dâhili düğümler bir teste karşılık gelir. Bir gözleme ilişkin girdi verisi karar ağacına girdiğinde düğümlere karşılık gelen kural setlerine göre dallanma gerçekleşmekte ve ağacın sonlandığı noktada tahmin sınıfı belirlenmektedir (Duru, 2019). Karar ağaçları yapılarında amaç, belli karar kuralları oluşturarak gözlemleri homojen alt gruplara ayırmaktır. Bu sayede veri seti içindeki karmaşık yapılar basit karar yapılarına dönüştürülmektedir. Genellenebilirlik kaygısı yoktur. Aşırı öğrenmeye eğilimlidir (URL-7). Yaygın olarak kullanılan birçok özel karar ağacı algoritması bulunmaktadır (Kaya ve Pehlivan, 2014).

Karar ağaçlarında, ağacın maksimum derinlik seviyesinin ayarlanması için kullanılan “max_depth” parametresi ile bir düğümden dallanılabilmesi için gerekli minimum gözlem sayısını gösteren “min_samples_split” parametresi optimize edilmelidir. Çok dallanma aşırı öğrenmeye, az dallanma eksik öğrenmeye sebep olacağı için “max_depth” parametresinin ayarlanması gerekmektedir (URL-7).

Rastgele ormanlar (RO)

RO algoritmasının temeli bagging (bootstrap aggregating) olarak adlandırılan ve bootstrap (yerine koymalı rastgele örnekleme) yaklaşımından türetilen yönteme dayanmaktadır. Bagging yönteminde birden fazla karar ağacının ürettiği tahminler bir araya getirilerek değerlendirilmektedir. Öncelikle, t adet ağaç için n gözlemden m ’er adet gözlem bootstrap

yöntemiyle seçilmekte, böylece t adet tahmin çıktısı elde edilmiş olmaktadır. RO algoritmasında hem gözlem seçimi hem de değişken seçimi rassal olarak yapılmaktadır. Tüm değişkenler arasından seçilen daha az sayıda değişken arasından seçim yapılarak dallanma işlemi gerçekleştirilmektedir. Regresyon problemlerinde değişken sayısının üçte biri, sınıflandırma problemlerinde ise değişken sayısının karekökü kadar değişken seçilerek dallanılmaktadır (URL-7). Rassallıkla birlikte genellenebilirliğin iyileştirilmesi amaçlanmaktadır. Brieman (2001)'a göre, ormandaki ağaç sayısı arttıkça, ormanın genelleme hatası bir sınıra yakınsamaktadır. Bir orman sınıflandırıcısının hatası, ormandaki her bir ağacın gücüne ve aralarındaki korelasyona bağlıdır. Dâhili tahminler aracılığıyla değişken önem seviyeleri de ölçülmektedir.

RO algoritmasında özellikle kullanılan ağaç sayısı ($n_estimators$) ve her dallanmada rastgele seçim yapılacak değişken sayısı ($max_features$) parametrelerinin ayarlanması önemlidir (Borghetti ve diğerleri, 2017).

Göz hareketine ilişkin fizyolojik veriler kullanılarak zihinsel iş yükünün tahmin edilmeye çalışıldığı bir çalışmada en iyi performans RO algoritması ile elde edilmiştir (Smith ve diğerleri, 2015). Sadece iki parametrenin ayarlandığı bir yöntem olması sebebiyle uygulanması kolaydır. Aşırı öğrenmeye dirençli olması ve doğrusal olmayan verileri modelleme yeteneği ile de öne çıkmaktadır (Borghetti ve diğerleri, 2017).

Gradyan artırma makineleri (GAM)

Zayıf öğrencileri bir araya getirerek güçlü bir öğrenci ortaya çıkarma fikri “boosting” terimi ile ifade edilmektedir. AdaBoost (adaptive boosting), boosting fikrini hayata geçiren algoritmaya verilen isimdir. GAM, AdaBoost’un sınıflandırma ve regresyon problemlerine uyarlanabilen geliştirilmiş versiyonudur. Diferansiyellenebilen herhangi bir kayıp fonksiyonunu optimize edebilen “gradient descent” algoritmasını kullanmaktadır (URL-7). Söz konusu algoritmada, bir grup zayıf karar ağacı sınıflandırıcısı iteratif olarak eğitilmekte ve kayıp fonksiyonu optimize edilerek güçlü bir sınıflandırıcı elde edilmesi hedeflenmektedir (Friedman, 2001). Kayıp fonksiyonu (loss function), her bir gözlem için modelin yaptığı tahmindeki hatayı yani artık değeri vermektedir. Bir ağaçtaki her bir yaprak için, içinde bulunan gözlemlerin artıklarının ortalaması alınarak o ağaca ilişkin sonuçlar elde edilir. Tahmin değeri ise, ağaçtan elde edilen sonuçların öğrenme oranı

(learning rate) denilen katsayı ile çarpılarak önceki ağacın sonuçlarına eklenmesi ile hesaplanmaktadır. Yeni artık değer her bir iterasyonda düşecektir. Öğrenme oranı, aşırı öğrenme sorununu aşmak için kullanılan bir hiperparametredir (URL-9). RO algoritmasından farkı ağaçların birbirine bağımlı olmasıdır. Bu algoritmada üretilen ağaçların oluşturduğu sonuçlar üzerinden yeni ağaçlar oluşturulmaktadır (URL-7).

Öğrenme oranını yani iteratif olarak her bir ağacın katkısını gösteren “learning_rate”, ağacın derinliğini gösteren “max_depth”, kullanılacak ağaç sayısını gösteren “n_estimators”, her ağaç için rastgele örneklenerek gözlemlerin oranını belirten “subsample” ve kayıp fonksiyonunu gösteren “loss” parametresi en önemli hiperparametreleridir. Kayıp fonksiyonu için “huber”, “quantile” ve “lad” güçlü yöntemlerdir. En küçük kareler fonksiyonu için “ls” kullanılmalıdır (URL-7).

Hafif gradyan artırma makineleri (HGAM)

GAM algoritmasında birçok optimizasyon tekniği benimsenmiş olsa da, fazla sayıda değişkenin yer aldığı büyük boyutlu ve büyük hacimli veri setleri söz konusu olduğunda verimlilik ve ölçeklenebilirlik hala tatmin edici değildir. Bunun önemli bir nedeni, her bir değişken için, bilgi kazancını tahmin etmek amacıyla tüm gözlemlerin taranması gerekliliğidir (Ke ve diğerleri, 2017). Çok zaman alan bu sorunu çözmek için Ke ve diğerleri (2017) tarafından iki yeni teknik önerilmiştir: “Gradyan Tabanlı Tek Taraflı Örnekleme” ve “Özel Özellik Paketleme”. “Gradyan Tabanlı Tek Taraflı Örnekleme” ile, küçük gradyanlara sahip gözlemlerin önemli bir kısmı hariç tutulur ve bilgi kazancı tahmini için yalnızca geri kalan önemli kısım kullanılır. Böylece veri hacmi azaltılmış olur. “Özel Özellik Paketleme” ile ise, değişken boyutunun azaltılması amacıyla özel değişkenler bir pakette bir araya getirilir. Bu sayede HGAM ile, GAM’dan 20 kat daha hızlı öğrenme süresinde hemen hemen aynı doğruluk performansı elde edilebilmektedir.

Karar ağaçlarında seviye odaklı (level-wise or depth-wise) veya yaprak odaklı (leaf-wise) şeklinde iki strateji kullanılabilmektedir. Seviye odaklı stratejide ağaç büyürken ağacın dengesi korunmaktadır. Yaprak odaklı stratejide ise kaybı azaltan yapraklardan bölünme işlemi devam etmektedir. HGAM bu özelliği sayesinde diğer boosting algoritmalarından ayrılmaktadır. Model, yaprak odaklı strateji ile daha az hata oranına sahip olmakta ve daha hızlı öğrenmektedir. Çünkü karar ağaçlarının eğitim süresi yapılan hesaplama ve

dolayısıyla bölünme sayısı ile doğru orantılıdır (URL-10). Zeng ve diğerleri (2019) tarafından yapılan çalışmada, EEG yönteminin kullanıldığı bir zihinsel iş yükü sınıflandırma problemi için geliştirilen HGAM tabanlı sınıflandırıcı diğer sınıflandırıcılardan üstün gelmiştir.

HGAM algoritmasında aşırı öğrenmeyi engellemek için “learning_rate”, “max_depth”, “num_leaves”, “min_data_in_leaf” hiperparametreleri, öğrenme süresini hızlandırmak için “feature_fraction”, “bagging_fraction” ve “num_iteration” hiperparametreleri optimize edilebilir. “Num_leaves”, ağaçta bulunacak yaprak sayısıdır. Ağacın karmaşıklığını kontrol etmede kullanılan en önemli parametredir. Aşırı öğrenmeden kaçınmak için $2^{(\text{max_depth})}$ değerinden küçük olması gerekmektedir. “Max_depth”, kurulacak ağacın derinliğini limitlemek için kullanılır. Aşırı öğrenmeden kaçınmak için optimizasyonu önemlidir. “Min_child_samples”, aşırı öğrenmeyi engellemek için kullanılacak önemli parametrelerden biridir. Bir yaprakta bulunabilecek minimum gözlem sayısını ifade etmektedir. Varsayılan değeri 20’dir ancak optimum değeri veri hacmine ve “num_leaves”e bağlıdır. Büyük bir değer olarak ayarlamak ağacın büyümesini engelleyebilir ve eksik öğrenmeye neden olabilir. “Learning_rate”, yani öğrenme oranı, kurulan ağaçları ölçeklendirmek için 0-1 arasında verilen bir değerdir. Bu değer küçük olması daha iyi tahmin gücüne yardımcı olmakla birlikte öğrenim süresini ve aşırı öğrenme ihtimalini arttıracaktır. “Feature_fraction”, her bir iterasyonda kullanılacak değişken; “bagging_fraction”, her iterasyonda kullanılacak veri sayısının ayarlanabileceği parametrelerdir. “num_iteration”, öğrenme sürecinde yapılacak iterasyon sayısıdır. “Feature_fraction”, “bagging_fraction” ve “num_iteration” sayıları öğrenim süresi ile doğrudan ilgilidir. Bu sayılar ne kadar az olursa öğrenim süresi o kadar az olacaktır. Ancak eksik öğrenmeye dikkat edilmesi de çok önemlidir. Birçok deneme yapılarak optimum sayıya ulaşılması önerilmektedir (URL-10).

Aşırı gradyan artırma makineleri (AGAM)

2014 yılında piyasaya sürülen aşırı gradyan artırma, gradyanı artırılmış karar ağaçları temeline dayalı çalışan ölçeklenebilir bir makine öğrenme algoritmasıdır (Chen ve Guestrin, 2016). Yeni nesil topluluk öğrenme algoritmalarından olan AGAM, algoritmanın eğitilmesi sürecinde aşırı öğrenme sorununun önüne geçerek modelin performansını artırmaktadır. Öğrenme sürecinde kullandığı amaç fonksiyonu ile diğer algoritmalar

farklılaşmaktadır. Amaç fonksiyonu, kayıp fonksiyonu ve düzenlileştirme (regularization) teriminden oluşmaktadır. Kayıp fonksiyonu, daha önce de belirtildiği gibi modelin yaptığı her bir tahmin değerinin gerçek değerinden olan farkını yani artık değeri hesaplamaktadır. Düzenlileştirme terimi ise, modelin karmaşıklığını kontrol etmekte ve bu da modeldeki aşırı öğrenme sorununu ortadan kaldırmaktadır (Üstüner, Abdikan, Bilgin ve Şanlı, 2020). En önemli hiperparametreleri; “n_estimators”, “subsample” ve “max_depth”tir. “n_estimators” modelde kurulacak ağaç sayısını; “subsample”, her bir ağacı oluşturmak için alınan satır oranını; “max_depth” ise ağacın derinliğini ifade etmektedir (URL-7).

Fizyolojik verilerin kullanıldığı bir bilişsel yük sınıflandırma probleminin çözümünde DVM, KEYK, KA, RO, AGAM algoritmaları arasında en yüksek sınıflandırma performansı AGAM ile elde edilmiştir (Wang ve Guo, 2019).

Kategori artırma (Category boosting-CatBoost)

2017 yılında Yandex tarafından geliştirilen ve farklı algoritmik avantajlara sahip olan bir GAM türevidir. Algoritmik üstünlüklerinden birisi permütasyona dayalı bir sıralı artırma özelliğine sahip olması, diğeri ise kategorik değişkenleri işlemeye yönelik yenilikçi yapısıdır (Prokhorenkova, Gusev, Vorobev, Dorogush ve Gulin, 2018). Girdi değişkenleri kategorik olduğunda kullanılmaktadır. En önemli hiperparametreleri; “iterations”, “learning_rate” ve “depth” hiperparametreleridir. Modelde kurulacak ağaç sayısı “iterations” ile belirlenir. Diğer algoritmalarındaki “n_estimators” ile aynı anlama gelmektedir. Derinlik “depth”, öğrenme oranı ise “learning_rate” ile ayarlanabilmektedir (URL-7).

GAM, AGAM, HGAM ve Kategori Artırma algoritmaları “Gradyanı Artırılmış Karar Ağaçları” olarak adlandırılmaktadır. 2022 yılında en çok kullanılan ilk üç kütüphanenin sırasıyla HGAM, Kategori Artırma ve AGAM olduğu açıklanmıştır. Veri bilimciler ve makine öğrenmesi ile ilgilenenleri bir araya getiren ve makine öğrenmesi yarışmaları sunan platformlarda HGAM, kazananlar arasında açık ara favori gradyan artırma karar ağacı kütüphanesi olmuştur. Kazananların çözüm yazılarında ve anketlerde HGAM’dan bahsetme sayısı, Kategori Artırma ve AGAM’ın toplamıyla aynı bulunmuştur (URL-3).

4.2.3. Özellik seçimi

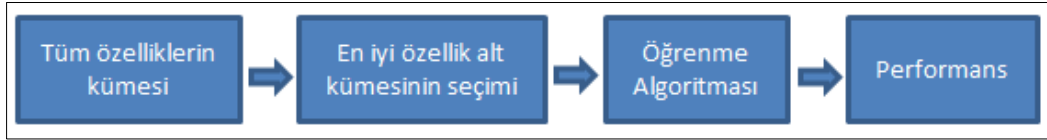
Veri madenciliği sürecinin önemli aşamalarından biri veri boyutunun azaltılması işlemidir. Veri boyutunun azaltılması kısaca, büyük veri kümelerinin depolanması ve analiz edilmesinde karşılaşılan sorunları aşmak için veri kümesinden ilgisiz veya gereksiz değişkenlerin çıkartılması olarak tanımlanmaktadır. Veri boyutunun azaltılması için kullanılan yöntemlerin başında özellik seçimi gelmektedir. Öznitelik seçimi olarak da ifade edilen özellik seçimi, orijinal veri setini temsil edebilecek en iyi alt kümenin seçimi olarak tanımlanmaktadır. Bu işlem, ilgilenilen problem için en faydalı ve en önemli özellikleri seçerek veri kümesindeki özellik sayısını azaltılmayı yani veri boyutunu düşürmeyi amaçlamaktadır (Budak, 2018). Fazla sayıda değişken olması modelin performansını düşüren, anlaşılabilirliğini azaltan ve yavaş çalışmasına neden olan bir faktördür. Bu nedenle özellik seçimi, makine öğrenmesi çalışmalarının önemli bir adımını oluşturmaktadır.

Denetimli öğrenmede özellik seçim yöntemleri üç temel gruba ayrılmaktadır (URL-11):

- Filtreleme Yöntemleri
- Sarmal Yöntemler
- Gömülü Yöntemler

Filtreleme yöntemleri

Bu yöntemler özellik alt kümelerinin filtrelenmesi yoluyla özellik seçimi gerçekleştirdiklerinden dolayı filtreleme yöntemleri adını almaktadırlar. Bu yöntemlerde özelliklerin seçimi, herhangi bir makine öğrenimi algoritmasından bağımsız gerçekleşmektedir. Özellikler çeşitli istatistiksel testlerden elde edilen çıktı değişkeni ile ilişki puanlarına göre seçilmekte ve model bu özellikler seçildikten sonra oluşturulmaktadır. Bilgi kazancı, fischer puanı, varyans analizi ve ısı haritası ile korelasyon matrisi özellik seçimi için kullanılan filtreleme yöntemlerindendir. Filtreleme yöntemleri genellikle bir ön işleme adımı olarak kullanılmaktadır (URL-11; URL-12). Filtreleme yöntemlerinin kullandığı yaklaşım Şekil 4.16'da verilen şemada özetlenmiştir (URL-12).

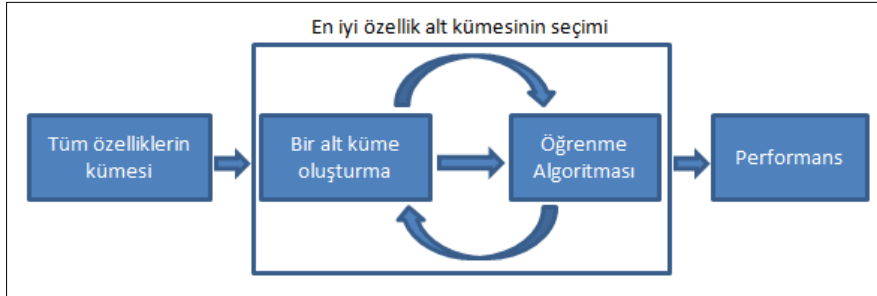


Şekil 4.16. Özellik seçiminde filtreleme yöntemleri yaklaşımı

Sarmal Yöntemler

Sarmal yöntemlerde bir makine öğrenmesi algoritmasına ihtiyaç duyulmakta, belirlenen makine öğrenmesi algoritmasına uyarlanmış en uygun özellik alt kümesi aranmaktadır (Kumar ve Minz, 2014). İteratif ve hesaplama yoğun bir yöntem olmakla birlikte, filtreleme yöntemlerinden daha doğru sonuç vermektedir. Sarmal yöntemlerin kullandığı yaklaşım Şekil 4.17’de verilmiştir (URL-12). Geriye Doğru Eleme, İleriye Doğru Eleme ve Özyinelemeli Özellik Eleme gibi farklı sarmal yöntemler bulunmaktadır. Geriye doğru ve ileriye doğru eleme yönteminde tek yönlü bir işlem söz konusu olduğundan başta çıkarılan özellik daha sonra yeni değişken kombinasyonu ile önemli hale geldiğinde veya eklenen bir özellik daha fazla özellik eklendikten sonra artık yararsız hale geldiğinde bu durumu geri alma şansı yoktur (URL-13). Özyinelemeli Özellik Eleme (ÖÖE) yöntemi ise, en az önemli olan değişkenleri eleyerek en doğru sonuca gitmeye çalışmakta, bu eleme işlemini makine öğrenmesi algoritmasına uyararak ve onun sonucunu değerlendirerek uygulamaktadır. Yöntemdeki işleyiş, verilen model sonucuna göre tahmin edicilerin önem katsayılarını almak ve onları sıralamaktır. Bu sayede en iyi x özelliği seçebilmek mümkün olabilmektedir (URL-14). Yöntem, özellikleri önem derecelerine göre sıralamak için “doğruluk” performans ölçüsünü kullanmaktadır. ÖÖE yöntemi, kullanılacak modeli ve gerekli özellik sayısını girdi olarak almakta, ardından 1 en önemli olmak üzere tüm değişkenlerin sıralamasını vermektedir. Ayrıca “True” (ilgili olma) özelliği ve “False” (alakasız olma) özelliği desteğini de vermektedir (URL-11). Gerekli özellik sayısının modele girdi olarak kullanıcı tarafından verilmesi, optimum değer girilmediği sürece en ideal çözüme ulaşmanın önünde bir engeldir. Bu noktada, ÖÖE yönteminin Çapraz Doğrulama ile Özyinelemeli Özellik Eleme (ÇDÖÖE) versiyonu devreye girmektedir. ÇDÖÖE, kullanıcı tarafından girilen parametrik değer yerine en uygun durma noktasını bulmak için çapraz doğrulama yöntemini kullanmaktadır (URL-14). Böylece, modele en çok katkı sağlayan optimum sayıda özellik belirlenebilmektedir. Belirli bir görev üzerinde eğitilmiş sınıflandırıcının genelleştirilebilirlik kabiliyetinin düşük olacağından hareketle bu zorluğun üstesinden gelmeyi amaçlayan bir çalışmada, ÖÖE yöntemi kullanılarak seçilen

en sağlam özellik alt kümesiyle eğitilip test edildiğinde görevler arası regresyon modelinin performansının önemli ölçüde iyileştiği gözlenmiştir. Çalışmada, özellik seçiminde ÖÖE yöntemini kullanmanın önemi vurgulanmıştır (Ke ve diğerleri, 2015).

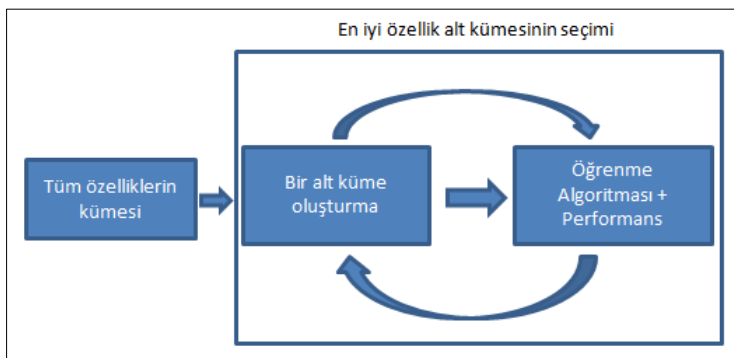


Şekil 4.17. Özellik seçiminde sarmal yöntemler yaklaşımı

Gömülü Yöntemler

Gömülü yöntemler, makine öğrenmesi algoritmasının her iterasyonunda eğitim sürecine en çok katkıda bulunan özellikleri çıkaran yöntemlerdir. Düzenleştirme yöntemleri, katsayı eşiği verilerek özellikleri cezalandıran ve en yaygın kullanılan gömülü yöntemlerdir. Birçok özelliğin ağırlığının sıfır olduğu ve modelden atıldığı, sıfır olmayan ağırlığa sahip geri kalanların ise modele dâhil edildiği Lasso Regresyon gömülü yöntemlere bir örnektir. Bu yöntemlerin dezavantajı, kullanışlı çok fazla algoritma barındırmamasıdır. Gömülü yöntemlerinin kullandığı yaklaşım Şekil 4.18’de verilen şemada özetlenmiştir (URL-12).

Filtreleme yöntemleri, sarmal ve gömülü yöntemlerden daha az doğru sonuç vermektedir. Sarmal ve gömülü yöntemlerin sonuçları daha güvenilir olmakla birlikte işlem süreleri uzundur. Az sayıda değişken olduğunda daha hızlı sonuç vermektedirler (URL-11).



Şekil 4.18. Özellik seçiminde gömülü yöntemler yaklaşımı

Tezde elde edilmek istenen modelde, özellik seçimi için daha güvenilir olması ve en uygun durma noktasını belirlemesi sebebiyle ÇDÖÖE yöntemi kullanılmıştır.

4.3. Model Performans Metrikleri

Bu bölümde, sınıflandırma ve regresyon modellerinin başarısının değerlendirmesinde kullanılan performans metrikleri ayrı ayrı ele alınacaktır.

Sınıflandırma modellerinde başarı değerlendirme

Makine öğrenmesi sınıflandırma performansı, belirli bir makine öğrenmesi algoritmasını uygulayan bir sınıflandırıcının kategoriler arasındaki ayrımı ne ölçüde iyi gerçekleştirebildiği ile ilgili önemli bir konudur. Sınıflandırma modeli değerlendirme ölçüleri, modelin performansını değerlendirmekte ve sınıflandırmanın ne kadar iyi veya kötü olduğunu söylemektedir. Ancak her bir değerlendirme ölçüsü değerlendirmeyi farklı şekilde yapmaktadır. Bu nedenle ele alınan probleme uygun ölçüyü seçmek önem taşımaktadır. Bazı uygulamalarda tek bir değerlendirme ölçüsüne bakmak çözülmek istenen sorunun tüm resmini vermemekte, modelin somut bir değerlendirmesini yapmak için bu bölümde detaylandırılacak olan sınıflandırma ölçülerinin bir alt kümesini kullanmak daha faydalı olabilmektedir.

Kolay anlaşılır olması sebebiyle doğruluk (accuracy) başta olmak üzere en çok kullanılan ölçüler karmaşıklık matrisi tabanlı ölçülerdir (Gong, 2021). Karmaşıklık matrisi (confusion matrix), gözlemlerin gerçek sınıf etiketlerine karşı model tahminlerinin tablo şeklinde görselleştirildiği bir matristir. Karmaşıklık matrisinde her bir satır, ilgili sınıftaki gerçek gözlem sayısını; her bir sütun ise, ilgili sınıfta olduğu tahmin edilen gözlem sayısını temsil etmektedir. Karışıklık matrisi tam olarak bir performans ölçüsü değildir; ancak, diğer ölçülerin hesaplanabilmesi için bir temel oluşturmaktadır. Karmaşıklık matrisindeki değerler; doğru pozitifler (DP), yanlış pozitifler (YP), doğru negatifler (DN) ve yanlış negatifler (YN) olarak adlandırılmaktadır. Bir gözlemin; gerçek sınıfı pozitif iken model tarafından da pozitif olarak tahmin ediliyorsa DP, gerçek sınıfı negatif iken model tarafından da negatif olarak tahmin ediliyorsa DN olarak sayılmaktadır. Şayet gerçek sınıfı negatif iken model tarafından pozitif olarak tahmin ediliyorsa YP, gerçek sınıfı pozitif iken model tarafından negatif olarak tahmin ediliyorsa YN olarak işlenmektedir. İki sınıflı bir

sınıflandırma modeli için karmaşıklık matrisinin gösterimi Çizelge 4.1 ile verilmiştir:

Çizelge 4.1. Karmaşıklık matrisinin gösterimi (Gong, 2021)

		Tahmin edilen sınıf	
		Pozitif	Negatif
Gerçek sınıf	Pozitif	DP	YN
	Negatif	YP	DN

Karmaşıklık matrisindeki değerler ile aşağıda verilen performans ölçütleri hesaplanabilmektedir:

- (a) Duyarlılık (Sensitivity / Recall),
- (b) Özgüllük (Specificity),
- (c) Yanlış pozitif oranı (YPO)
- (d) Yanlış negatif oranı (YNO)
- (e) Pozitif tahmin değeri (PTD) / Kesinlik (Precision),
- (f) Negatif tahmin değeri (NTD)
- (g) Doğruluk (accuracy)
- (h) F1 skoru

Söz konusu ölçülerin açıklamaları ve iki sınıflı bir sınıflandırma modeli için hesaplama formüllerini içeren özet tablo Çizelge 4.2 ile verilmiştir. Bu ölçülerden sıklıkla kullanılanlar; doğruluk, duyarlılık ve kesinliktir. Doğruluk, model tarafından doğru tahmin edilen gözlemlerin tüm gözlem sayısına oranını ifade etmektedir. Pek çok sınıflandırıcı doğruluk oranını optimize etmeye çalışmaktadır (Gong, 2021). Duyarlılık belli bir sınıfa ait gözlemlerin ne ölçüde doğru tahmin edildiğine; kesinlik ise, belli bir sınıfa ilişkin tahminlerin gerçekte ne ölçüde doğru olduğuna dayanmaktadır. Formülasyonlarından da anlaşılacağı üzere, duyarlılık ile kesinlik ters orantılıdır ve aralarında bir denge tutturmak gerekmektedir. Bu noktada hem YP hem de YN gözlem sayılarını eşit derecede dikkate alan F1 skoru devreye girmektedir. F1 skoru duyarlılık ve kesinlik ölçülerinin harmonik ortalaması alınarak hesaplanmakta ve 0 ile 1 arasında değer almaktadır. Özellikle sınıf dengesizliği durumlarında performans değerlendirmede F1 skorunun tercih edilmesi

önerilmektedir. Çünkü yüksek bir F1 skoru için hem YP hem de YN sayıları düşük olmalıdır.

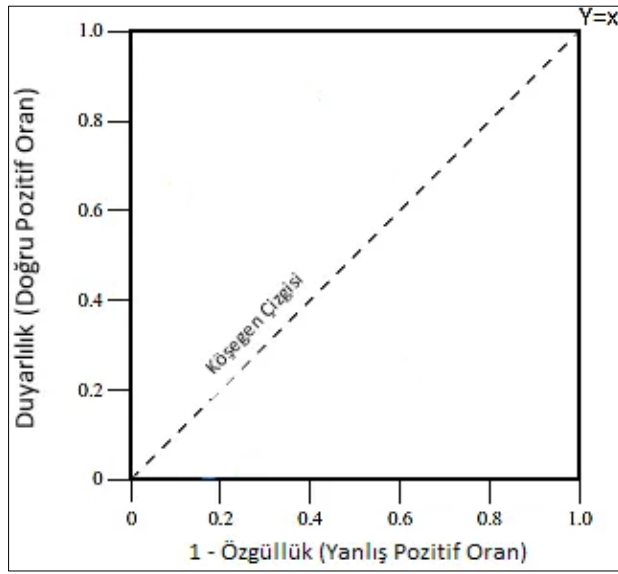
Çizelge 4.2. Sınıflandırma modellerinin başarı ölçüleri (Erickson ve Kitamura, 2021)

Ölçü	Anlamı	Eş Anlamlısı	Matematiksel formülü
Duyarlılık	Pozitif olarak tahmin edilen pozitif gözlemlerin tüm pozitif gözlemlere oranı	Doğru pozitif oranı (DPO)	$\text{Duyarlılık} = \frac{DP}{DP + YN}$
Özgüllük	Negatif olarak tahmin edilen negatif gözlemlerin tüm negatif gözlemlere oranı	Doğru negatif oranı (DNO)	$\text{Özgüllük} = \frac{DN}{DN + YP}$
Yanlış pozitif oranı	Pozitif olarak tahmin edilen negatif gözlemlerin tüm negatif gözlemlere oranı	Yanlış alarm olasılığı	$YPO = \frac{YP}{DN + YP}$
Yanlış negatif oranı	Negatif olarak tahmin edilen pozitif gözlemlerin tüm pozitif gözlemlere oranı	Kaçırma oranı	$YNO = \frac{YN}{DP + YN}$
Kesinlik	Pozitif olarak tahmin edilen pozitif gözlemlerin pozitif tahmin edilen gözlemlere oranı	PTD	$\text{Kesinlik} = \frac{DP}{DP + YP}$
Negatif tahmin değeri	Negatif olarak tahmin edilen negatif gözlemlerin negatif tahmin edilen gözlemlere oranı	-	$NTD = \frac{DN}{DN + YN}$
Doğruluk	Model tarafından doğru tahmin edilen gözlemlerin oranı	-	$\text{Doğruluk} = \frac{DP + DN}{DP + DN + YP + YN}$
F1 skoru	Pozitif tahmin değeri ile duyarlılığın harmonik ortalaması	F skoru	$F1 = \frac{2DP}{2DP + YP + YN}$

Orijinal adı “Receiver Operating Characteristic-ROC” olan Alıcı İşletim Karakteristiği (AİK) eğrisi, model performansını ölçmek için kullanılan bir başka yöntemdir. AİK eğrisi ilk kez 2. Dünya Savaşı sırasında bir radar alıcısının düşman uçaklarını kuş sürülerinden ayırt ederek doğru tespit edebilme başarısını test etmek için kullanılmıştır. AİK eğrisi; x eksenini YPO (1 - özgüllük), y eksenini ise DPO (duyarlılık) olan grafiksel gösterimdir (Erickson ve Kitamura, 2021). Modelin DPO değeri ile YPO değeri cinsinden ne kadar iyi

ayırım yapabildiğini açıklamaktadır. Örnek bir AİK eğrisi Şekil 4.19’da verilmiştir.

“Eğri Altındaki Alan” (EAA), AİK eğrisinin altında kalan alanı vermektedir. 1970’lerde sinyal tespiti teorisinde performansı ifade etmek için kullanılabilecek en iyi metriğin EAA olduğu kabul edilmiştir (Canbek, Sağiroğlu, Temizel ve Baykal, 2017). Bu alanın ölçüsü 0 ile 1 arasında olup, 0 ise bütün tahminlerin yanlış olduğu 1 ise bütün tahminlerin doğru olduğu anlamına gelmektedir.



Şekil 4.19. AİK eğrisi örneği

DPO, gerçekte durum pozitifse bunların kaçını pozitif tahmin ettiğimizi; YPO ise, gerçekte durum negatifken bunların kaçını pozitif olarak tahmin ettiğimizi göstermektedir. AİK grafiğindeki köşegen çizgisi olan $x=y$ doğrusu, bir sınıfı rastgele tahmin etme stratejisini temsil etmektedir, yani köşegende sınıflandırma becerisi bulunmayan modeller yer almaktadır. Gerçeğe uygun doğru sonucu veren eğriler, AİK düzleminin sol üst üçgen kısmında bulunacaktır. Diğer bir ifadeyle, sol üste çıktıkça doğru tahmin sayısı artmaktadır. Sınıflandırıcı sağ alt üçgende bulunuyorsa, rastgele tahminden daha kötü performans gösterdiği şeklinde yorumlanmaktadır (URL-15; Kohl, 2012). Sınıf dengesizliği (sınıflara ait gözlem sayılarının farklı olması) durumlarında EAA iyi bir ölçü değildir (Erickson ve Kitamura, 2021; URL-16). Eğer negatif gözlemler pozitif gözlemlere baskın geliyorsa fazla iyimser bir değerlendirmeye sebep olacaktır. Böyle bir durumda EAA yerine F1 skorunu kullanmak daha mantıklı olacaktır.

Çoklu sınıflandırma problemlerinde karmaşıklık matrisini yorumlamak zor olsa da, öncelikle pozitif ve negatif sınıfları belirlemek gerekmektedir. Sınıflandırma dengeli ise, yani her sınıfın önem derecesi aynı ise, olumlu ya da olumsuz herhangi bir sınıf olmayacaktır. Bu durumda, pozitif ve negatif sınıflar her bir sınıf için ayrı ayrı belirlenmekte, yukarıda açıklanan kesinlik ve duyarlılık gibi performans ölçüleri de sınıf bazında hesaplanmaktadır (URL-17).

Bu tezde ele alındığı şekilde 0, 1, 2, 3 şeklinde etiketlenmiş dört sınıflı bir sınıflandırma modelinde; 0 etiketli olma veya olmama, 1 etiketli olma veya olmama, 2 etiketli olma veya olmama, 3 etiketli olma veya olmama durumları için ayrı ayrı ikili performans ölçüleri hesaplanmaktadır. Örneğin; 0 etiketli öngörülen gözlemlerin kaçısı gerçekten 0 etiketli, 1 etiketli öngörülen gözlemlerin kaçısı gerçekten 1 etiketli, 2 etiketli öngörülen gözlemlerin kaçısı gerçekten 2 etiketli ve 3 etiketli öngörülen gözlemlerin kaçısı gerçekten 3 etiketli şeklinde dört farklı sorunun cevabı ile birlikte dört farklı kesinlik ölçüsü elde edilecektir.

Çoklu sınıflandırma performansının değerlendirilebilmesi için ise bu sınıf bazlı ölçülerin ortalamasını almaya yönelik üç farklı hesaplama yaklaşımı bulunmaktadır. Bunlar; makro, mikro ve ağırlıklı ortalama yaklaşımlarıdır. Makro yaklaşımda; sınıf bazında belirlenen performans ölçülerinin aritmetik ortalaması alınmaktadır. Bu yaklaşım her bir sınıfa eşit ağırlık verdiği için dengeli sınıflandırma problemlerinde tercih edilmektedir. Ağırlıklı ortalama yöntemi; sınıf bazında belirlenen performans ölçülerinin, her bir sınıfta yer alan gözlemlerin toplam gözlem sayısına oranına göre belirlenen ağırlıklar dikkate alınarak ortalamalarının alınması yaklaşımına dayanmaktadır. Dolayısıyla dengeli olmayan sınıflandırma problemleri için uygulanabilir bir yöntemdir. Mikro yaklaşımda ise; karmaşıklık matrisinin diyagonal hücrelerinin toplamının tüm hücrelerin toplamına oranı, yani doğruluk ölçüsü esas alınmaktadır (Erickson ve Kitamura, 2021; URL-17).

Çoklu sınıflandırma problemlerinde kullanılan diğer iki ölçü ise Kappa skoru (Cohen's Kappa) ve Matthew korelasyon katsayısıdır (MKK).

Pozitif-negatif oranının 10:100 olduğu bir veri setinde, sınıflandırıcı tüm negatif örnekleri doğru bir şekilde tahmin ederse %90'ın üzerinde doğruluk elde edilebilecektir. Sınıf dengesizliğini hesaba katmaması sebebiyle bu tip durumlarda doğruluk yanıltıcı bir ölçü olabilmektedir. Ayrıca, makine öğrenimi algoritmaları verilerin olasılıksal varsayımlarına

dayandığından, tahminler üretmenin getirdiği doğal belirsizliği ölçebilen bir puana ihtiyaç duyulmaktadır. Hem şansı hem de sınıf dengesizliğini dikkate alarak genişletilmiş bir doğruluk ölçüsü sunan Kappa skoru adını Jacob Cohen'den almıştır. Hesaplama formülü Eş. (4.13) ile verilmiştir (URL-17):

$$k = \frac{p_0 - p_e}{1 - p_e} = 1 - \frac{1 - p_0}{1 - p_e} \quad (4.13)$$

Eş. 2.13'te p_0 , herhangi bir karmaşıklık matrisinin köşegen hücrelerinin toplamının toplam gözlem sayısına bölümüdür (gözlenen doğruluk). p_e (beklenen doğruluk) ise, gerçek değerler ile yanlış değerlerin tesadüfen uyuşma olasılığını ifade etmektedir. Dolayısıyla p_e , rastgele bir gözlemin o sınıftan olma olasılığı ile rastgele bir tahminin yanlışlıkla o sınıf olarak tahmin edilme olasılığının çarpımlarının her bir sınıf için ayrı ayrı hesaplanarak toplanması ile bulunmaktadır. 0,4'ün üzerindeki kappa skorları orta ölçüde iyi, 0,6'nın üzerindeki önemli ölçüde iyi, 0,8'in üzerindeki ise mükemmele yakın olarak nitelendirilmektedir (Gong, 2021; URL-17).

Matthew korelasyon katsayısı, gerçek değerler ile tahmin edilenler arasındaki ilişkiyi temsil etmektedir. Tıpkı Pearson ve Spearman korelasyon katsayılarında olduğu gibi -1 ile 1 arasında değer almaktadır. 1 katsayısı mükemmel bir sınıflandırıcı anlamına gelirken 0'a yakın bir değer, sınıflandırıcının rastgele tahminden daha başarılı olmadığı anlamına gelmektedir. Matthew korelasyon katsayısı, karmaşıklık matrisindeki dört değeri de içeren tek ölçüdür (Gong, 2021). İkili sınıf için hesaplama formülü Eş. (4.14) ile verilmiştir:

$$MKK = \frac{(DP \times DN) - (YP \times YN)}{\sqrt{(DP + YP) \times (DP + YN) \times (DN + YP) \times (DN + YN)}} \quad (4.14)$$

MKK, sınıf dengesizliği olsa bile her bir sınıfın iyi tahmin edilip edilmediğine odaklanmaktadır. Karmaşıklık matrisi bağlamında bir sınıflandırıcının performansını değerlendirmek için en iyi ölçü olduğu söylenmektedir. 0,7'nin üzerindeki değerler güçlü bir sınıflandırıcıya işaret etmektedir (URL-17).

Sonuç olarak, çoklu sınıflandırma problemlerinde:

- Bir sınıflandırıcının genel performansını tek bir ölçü ile değerlendirebilmek ve diğer sınıflandırıcılar ile kıyaslayabilmek için MKK ve kappa skoruna,
- Dengeli bir sınıflandırmada bir sınıflandırıcının sınıflar arasında ayırım yapma yeteneğini ölçmek için AİK EAA puanına,
- Dengeli olmayan bir sınıflandırmada yanlış pozitifleri ve yanlış negatifleri en aza indiren bir yaklaşım için F1 skoruna,
- Tek bir sınıfın yanlış pozitiflerini azaltmaya odaklanmak için o sınıfın kesinlik ölçüsüne,
- Tek bir sınıfın yanlış negatiflerini azaltmaya odaklanmak için ise o sınıfın duyarlılık ölçüsüne odaklanmak faydalı olacaktır (URL-17).

Regresyon modellerinde başarı değerlendirme

Regresyon modellerinde performans kriteri olarak Ortalama Mutlak Hata (OMH), Hata Kareler Ortalaması (HKO), sıklıkla da Hata Kareler Ortalamasının Karekökü (HKOK) kullanılmaktadır. Ayrıca bağımlı değişkendeki değişimin bağımsız değişkenlerce açıklanma yüzdesi anlamına gelen R^2 değeri de oluşturulan regresyon denklemlerinin değerlendirilmesinde kullanılan önemli bir performans kriteridir.

HKO, regresyon problemlerinde kullanılan en yaygın ölçülerden biridir. Temel olarak tahmin edilen ve gerçek değerler arasındaki hataların karelerinin ortalamasını vermektedir. Hesaplama formülü Eş. (4.15) ile verilmiştir:

$$HKO = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (4.15)$$

Daha önce regresyon modellerinde açıklandığı gibi, n sayıda gözlemden i. gözlem için bağımlı değişkenin gerçek değeri y_i ile, tahmin edilen değeri ise $\hat{y}_i$ ile gösterilmektedir (URL-5). Bazen HKO değeri kolayca karşılaştırılamayacak kadar büyük olabileceğinden, HKOK HKO'dan daha sık kullanılmaktadır (URL-18). HKOK, HKO'nun karekökü alınarak hesaplanmaktadır. HKOK hesaplama formülü Eş. (4.16) ile gösterilmektedir:

$$HKOK = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (4.16)$$

OMH tahmin edilen ve hedef değerler arasındaki ortalama mutlak farkı bulan bir ölçüdür. OMH'nin aykırı değerlere (outliers) karşı HKO'dan daha sağlam olduğu bilinmektedir. Bunun ana nedeni, HKO'da hataların karesi alındığında, diğer gözlemlerden daha yüksek hata değerlerine sahip olan aykırı değerlerin son hata değerinde ve model parametrelerini etkilemede daha fazla baskınlık kazanmasıdır (URL-5). OMH, Eş. (4.17)'de verilen formül ile hesaplanmaktadır:

$$OMH = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (4.17)$$

Determinasyon katsayısı olarak da bilinen R^2 , modeldeki bağımsız değişkenlere göre bağımlı değişkenin varyasyon oranını yani bağımlı değişkendeki değişkenliğin ne kadarının model tarafından açıklanabileceğini ölçmektedir. Örneğin bir modelin R^2 değeri 0,50 ise bu sonuç, gözlemlenen değişkenliğin yaklaşık yarısının modelin girdileriyle açıklanabildiği anlamına gelmektedir. R^2 aşırı öğrenme sorununu dikkate almamaktadır. Bu durumda düzeltilmiş R^2 kullanılmaktadır. Düzeltilmiş R^2 modele fazladan eklenen bağımsız değişkenleri cezalandırmakta ve aşırı öğrenme sorununun önüne geçmektedir. Daima R^2 'den küçük ya da ona eşittir (URL-18).

5. ARAŞTIRMA BULGULARI VE TARTIŞMA

Aynı dört katılımcı üzerinde fakat EEG ve göz izleme için ayrı zamanlarda gerçekleştirilen pilot uygulama aşamasında iki farklı veri seti analiz edilmiştir. Temel uygulama verilerinin analizinde ise üç farklı veri seti üzerinde çalışılmıştır. İlk veri seti oturum bazında ele alınmış öncelikle performans değişkenlerinin ve öznel iş yükü değerlendirme skorlarının farklı zorluk derecesindeki oturumlarda nasıl değiştiği incelenmiştir. Böylece, zihinsel iş yükünün zorluk seviyesine göre beklenen şekilde değişip değişmediği literatürde kanıtlanmış yöntemlerle test edilmiştir. Söz konusu değişkenlerin cinsiyet ve sınıf demografik değişkenleri açısından anlamlı fark gösterip göstermediği de araştırılmıştır. Katılımcı grubunun yaş aralığının düşük olması sebebiyle yaşla ilgili bir analizin gerçekleştirilmesi gerekli görülmemiştir. İlk veri setinin oturum bazında ele alınmasının nedeni, her bir oturum için elde edilebilen hedefi bulma sayısı, hedef olmayanı bulma sayısı, doğru yanıt oranı gibi değişkenlerden yararlanma motivasyonudur. Bu değişkenler kullanılarak performansın zorluk derecesine göre nasıl değişkenlik gösterdiği ve öznel iş yükü değerlendirme sonuçları ile ilişkisi incelenebilmiştir. Bu sayede öncelikle tezin “görev zorluk seviyelerinin zihinsel iş yükü seviyelerine karşılık geldiği” temel varsayımını destekleyecek bulguların elde edilebilmesi amaçlanmıştır. Oturum bazında ele alınan veri seti 180 gözlemden (15 katılımcı x 12 oturum) oluşmaktadır. Analizde kullanılan ikinci veri seti ise uyaran bazında ele alınmıştır. Özellikle EEG ve göz ile ilgili değişkenlerin farklı zorluk seviyesindeki testlerde gösterdiği değişim üzerinde durulmuştur. Uyaran bazında ele alınan veri seti ayrıca, makine öğrenmesi algoritmaları aracılığıyla zihinsel iş yükü seviyesini tahmin eden sınıflandırma ve regresyon modellerini geliştirmek üzere kullanılmıştır. Cihazların sağladığı anlık bilgi akışı avantajının kullanılabilmesi amacıyla ve büyük veri setleri için sınıflandırma ve regresyon problemlerinin üstesinden gelebilecek pek çok güvenilir makine öğrenmesi algoritması bulunması sebebiyle analizde mümkün olduğunca fazla veriden yararlanılmak istenmiştir. Veri setinin uyaran bazında ele alınmasının nedeni zihinsel sürecin görsel algılama, çağırma, eşleştirme ve tepki verme adımlarının sırasıyla her bir uyarının ekrana gelmesinden itibaren 2500 milisaniyelik zaman diliminde gerçekleşmesidir. Dolayısıyla uyaran bazında ele alınan veri seti 2700 gözlemden (15 katılımcı x 12 oturum x 15 uyaran) oluşmaktadır. Analiz edilen üçüncü veri seti ise, EEG cihaz yazılımı aracılığıyla elde edilen ham EEG verisinin, katılımcıların gerçekleştirdiği her bir oturum için MATLAB

aracılığıyla set uzantılı dosyalara dönüştürülen ve EEGLAB ile veri ön işleme sürecinden geçirilen 180 adet set uzantılı dosyadır. 0-geri göreve ilişkin dosyalar 4800 x 14, 1-geri göreve ilişkin dosyalar 5120 x 14, 2-geri göreve ilişkin dosyalar 5440 x 14 ve 3-geri göreve ilişkin dosyalar 5760 x 14 boyutundadır. Gözlem sayıları görev bazında uyarın sayılarına göre farklılık göstermektedir. EEG cihazının 14 kanallı olması sebebiyle veri setindeki sütunlar elektrotlara karşılık gelmektedir. Her bir dosya EEGLAB STUDY ile analiz edilmiş; bununla, beynin farklı bölümlerinde farklı görev koşulları altında bant gücündeki değişimin incelenmesi ve istatistiksel bulguların görselleştirilmesi amaçlanmıştır. İlk veri seti Statistical Package for Social Sciences (SPSS) 21.0 programı ile, ikinci veri seti SPSS 21.0 ve Python 3.8 programları ile analiz edilmiştir. Üçüncü veri setinin analizinde ise, EEGLAB programının 2021 sürümünden yararlanılmıştır.

5.1. Pilot Uygulamalarda Elde Edilen Bulgular

Pilot çalışmalar; pandemi koşulları altında, 30-40 yaş aralığında 4 (2 kadın, 2 erkek) katılımcı üzerinde gerçekleştirilmiştir. Yaş ortalaması 36'dır. Pilot çalışmalarda cihazların aynı zamanlarda temin edilebilmeleri nedeniyle, göz izleme ve EEG ölçümleri aynı katılımcılardan ayrı zamanlarda toplanmış ve ayrı ayrı analiz edilmiştir. Her iki pilot çalışmada da aynı deney süreci uygulanmış, n-geri testler Inquisit Web Player demo sürümü kullanılarak gerçekleştirilmiştir. Görevler sırasında, katılımcılardan ekranda belli bir süre görünen harfleri hafızalarında tutup, 0-geri görevde şayet ekranda görünen harf M ise klavyede "A" tuşuna basmaları değilse hiç bir tuşa basmamaları istenmiştir. 1, 2 veya 3 geri görevde ise; ekranda görünen harf sırasıyla 1 önceki, 2 önceki veya 3 önceki harfle aynı ise "A" tuşuna basmaları değilse hiç bir tuşa basmamaları istenmiştir. Her bir zorluk seviyesi 3 kere uygulanmış, yani bir katılımcı tarafından söz konusu görev için toplamda 12 deneme gerçekleştirilmiştir. Çalışmanın bu aşamasında, pilot test verisi analiz edilerek göz kırpmaya tepkisine ilişkin değişkenler elde edilmeye çalışılmış, geliştirilmesi planlanan tahmin ve sınıflandırma modellerinin elde edilmesi için kullanılacak Python kodları hazırlanmıştır. Görev sırasında ekrana gelen her bir harf uyarını bazında geçen süre bir gözlem satırına karşılık gelecek şekilde veri seti oluşturulmuştur. Problem öncelikle bir regresyon problemi olarak ele alınarak, "Ağırlıklı Toplam NASA-TLX Skoru" bağımlı değişkeni tahmin edilmeye çalışılmıştır. Çalışmanın ikinci kısmında ise yine makine öğrenmesi algoritmaları kullanılarak görev zorluğunun neden olduğu zihinsel iş yükünün

sınıflandırmasına yönelik bir analiz yapılmıştır. Kullanılan her bir algoritmanın performans kriterleri hesaplanarak karşılaştırmalar yapılmıştır.

Göz izleme pilot çalışmasında toplanan verilerin analizi sonucunda, hem ağırlıklı toplam NASA-TLX skoru ile zorluk seviyesinin büyük oranda pozitif ilişkili olduğu hem de her iki değişkenin hemen hemen tüm göz izleme değişkenleriyle ilişkili olduğu ortaya çıkmıştır. Her bir zorluk seviyesi arasında NASA-TLX skorları açısından anlamlı fark olduğu bulunmuştur ($p < 0,001$). Hipotez testleri ile ayrıca, maksimum seğirme süresi, odaklanma sayısı ve seğirme süresinin standart sapması dışında kalan 24 göz izleme değişkeninin de zorluk seviyesine göre anlamlı fark gösterdiği görülmüştür. Seğirme sayısının görev zorluğu arttıkça arttığı, ortalama odaklanma süresine ilişkin değişkenlerin ise zorluk seviyesi ile birlikte azaldığı gözlenmiştir. Dolayısıyla daha zor bir görevi gerçekleştiren katılımcıların ekrana daha kısa sürelerle odaklandığı ve gözlerinin daha çok sayıda seğirme hareketi gerçekleştirdiği sonucu ortaya çıkmıştır. Diğer yandan, ortalama göz bebeği çapının görev zorluğu arttıkça anlamlı ölçüde arttığı görülmüştür. Göz kırpma değişkenlerinin ise algılanan zihinsel iş yükü ve zorluk seviyeleri ile pozitif yönde ilişkili olduğu ortaya çıkmıştır. Görev zorlaştıkça katılımcıların hem daha fazla sayıda göz kırpma hareketi gerçekleştirdikleri hem de göz kırpma sürelerinin uzadığı gözlenmiştir. Göz izleme pilot çalışmasının makine öğrenmesi bölümünde en başarılı sonuçlar ağaç tabanlı makine öğrenmesi algoritmaları ile elde edilmiştir. Sınıf sayısının azaltılması yoluna gidildiğinde dört sınıflı modelin performansına (%68) kıyasla çok daha iyi bir sınıflandırma performansı (iki sınıflı modelde %84) ile karşılaşılabileceği görülmüştür.

EEG pilot çalışmasında ise; BAP desteği ile satın alınan EEG cihazı ve yazılımı incelenmiş, göz izleme pilot çalışmasının gerçekleştirildiği aynı dört katılımcıdan EEG sinyalleri toplanmıştır. Aynı deney prosedürü uygulanarak elde edilen EEG verileri istatistiksel olarak analiz edilmiştir. Artan görev zorluğuyla birlikte beynin ön bölgelerindeki teta dalga gücünün arttığı doğrulanmıştır. NASA-TLX ölçeği aracılığıyla öznel olarak beyan edilen iş yükünün; özellikle ön bölgedeki teta gücü ve ayrıca yan ve arka bölgelerdeki gama gücü ile pozitif ilişkili olduğu, prefrontal yüksek beta gücü ile ise negatif ilişkili olduğu ortaya konmuştur. Görev zorluk seviyeleri arasında 31 EEG değişkeni açısından %1 anlamlılık düzeyinde farklılık olduğu görülmüştür. Yapılan pilot çalışmanın makine öğrenmesi bölümünde, en başarılı sonuçların göz izleme pilot çalışmasında da olduğu gibi ağaç tabanlı algoritmalar ile elde edildiği görülmüştür. Dört

sınıflı problemin en yüksek doğruluk oranı (%68) ile tahmin eden model rastgele ormanlar algoritması ile elde edilmiştir. Model performansına katkı sağlayan en önemli değişkenler ise; prefrontal ve frontal teta gücü, prefrontal yüksek beta gücü, oksipital, parietal ve temporal gama gücü ile oksipital alfa gücüne karşılık gelen değişkenler olarak bulunmuştur. Sınıf sayısının azaltılması yoluna gidildiğinde çok daha iyi bir sınıflandırma performansı (%87) ile karşılaşılabileceği görülmüştür ve iki sınıflı modelde en iyi sonuç göz izleme çalışmasına benzer şekilde GAM algoritması ile elde edilmiştir.

Pilot çalışmalar; cihazların tanınmasını, ölçümler sırasında karşılaşılabilecek zorlukların öngörülmesini ve deney sürecinin test edilmesini sağlamıştır. Bu aşamada, veri ön işleme ve göz kırpma değişkenlerinin türetilmesi amacıyla hazırlanan Excel VBA makroları ve ayrıca, makine öğrenmesi algoritmalarının uygulanması, model değerlendirme ve özellik seçimi için hazırlanan Python kodları temel uygulamadaki analiz sürecine büyük ölçüde katkı sağlamıştır. Belirlenen görev zorluk seviyelerinin zihinsel iş yükündeki değişimi tetikleyip tetiklemediği ve zihinsel iş yükündeki olası değişimin EEG ve göz verilerinde anlamlı bir değişime neden olup olmadığı kontrol edilmiştir. Elde edilen sonuçlar ışığında, n-geri görevlerin önceden belirlenen zorluk seviyelerinin zihinsel iş yükünde dalgalanmaya sebep olduğu, öznel iş yükü değerlendirme sonuçları aracılığıyla doğrulanmıştır. Hem göz izleme hem de EEG'nin zihinsel iş yükü analizinde önemli rol oynadığı görülmüş ve çalışmanın devamında her iki yöntemin birlikte kullanımının faydalı olacağı değerlendirilmiştir.

5.2. Temel Uygulamada Elde Edilen Bulgular

Temel uygulamada pilot çalışmalardan farklı olarak her iki cihazdan eş zamanlı ölçüm alınmıştır. Bu doğrultuda, kayıtlarda eş zamanlılığın sağlanabilmesi için cihaz yazılımlarının tespit edebileceği şekilde klavyede işaretçiler (marker) belirlenmiştir. Diğer yandan; görev testlerinin uygulandığı yazılımın gelişmiş bir versiyonu olan Inquisit Lab 6 (URL-19) kullanılmış ve performansa ilişkin veriler elde edilmiştir. Böylece literatürde zihinsel iş yükü ölçümünde kullanılan nesnel-deneysel yöntemlerden biri daha analize dâhil edilebilmiştir. Testlerin içeriği ve uygulanış şeklini içeren deney süreci literatürdeki benzer çalışmalar örnek alınarak belirlenmiştir (Liu ve diğerleri, 2017). Pilot çalışmalardan farklı olarak, zorluk seviyelerine bağlı olarak meydana gelmesi istenen zihinsel iş yükü değişiminin doğruluk performans kriteri ile ayırt edilebilirliğinin artırılması için görev

sırasında eşleşmeme durumu yani hedef olmayan uyarının tespiti için de yanıt verilmesi beklenmiştir. Bu sayede; eşleşmeme durumu için verilen doğru yanıtın bilinçli verilip verilmediğinin kontrolü performans verilerine bakılarak sağlanmış olmakta, oturum bazında performans göstergesi olarak kullanılan doğru yanıt oranı değişkeninin güvenilirliği temin edilmektedir.

Bu bölümde; performans ve öznel iş yükü değerlendirme sonuçları, EEG ve göz izleme değişkenlerine ilişkin temel istatistiksel bulgular, EEGLAB STUDY analizi ve makine öğrenmesi modellerine ilişkin sonuçlar ayrı başlıklar halinde verilecektir.

5.2.1. Performans ve öznel iş yükü değerlendirme sonuçlarına ilişkin bulgular

Daha önce de belirtildiği gibi, temel uygulamada üç farklı veri seti üzerinde çalışılmıştır. Bu bölümde ilk iki veri seti kullanılarak elde edilen temel istatistiksel bulgular üzerinde durulacaktır. İlk olarak oturum bazında ele alınan veri seti kullanılarak öznel iş yükü skorları, performans ölçüleri ve görev zorluk seviyesi arasındaki ilişki araştırılmıştır. Bu kapsamda uygulanacak korelasyon ve hipotez testlerine karar verilmeden önce, ele alınan değişkenlerin normal dağılımdan gelip gelmediği Kolmogorov-Smirnov (K-S) testi ile çarpıklık ve basıklık değerlerine bakılarak kontrol edilmiştir. Çizelge 5.1’de K-S testi sonuçları Çizelge 5.2’de değişkenlerin zorluk seviyesi bazında ortalama ve standart sapma değerleri, Çizelge 5.3’te ise çarpıklık ve basıklık değerleri verilmiştir.

Çizelge 5.1. Normallik testi

	İstatistik	Anlamlılık
Zihinsel Talep	0,151	0,000
Fiziksel Talep	0,380	0,000
Zaman Baskısı	0,144	0,000
Performans	0,198	0,000
Çaba	0,136	0,000
Gerginlik	0,230	0,000
Ağırlıklı Toplam NASA-TLX Skoru	0,130	0,000
Doğru Sayısı	0,260	0,000
Yanlış Sayısı	0,260	0,000
Hedefi Bulma Sayısı	0,337	0,000
Hedef Olmayan Bulma Sayısı	0,335	0,000
Hedefi Bulamama Sayısı	0,337	0,000
Hedef Olmayan Bulamama Sayısı	0,335	0,000
Ortalama Tepki Süresi	0,157	0,000
Toplam Tepki Süresi	0,157	0,000
Doğru Yanıt Oranı	0,261	0,000

K-S testine göre anlamlılık değerlerinin 0,05'in altında olması, değişkenlerin hiçbirinin normal dağılımdan gelmediğini göstermektedir (Kalaycı, 2008). Uysal ve Kılıç (2022) normallik değerlendirmesinde özellikle çarpıklık katsayısı/standart hata değerinin koşulların çoğunda diğer yöntemlere göre daha düşük birinci tip hata ve daha yüksek güç değerine sahip yöntemlerden biri olduğunu ortaya koymuşlardır.

Çizelge 5.2. Tanımlayıcı istatistikler

	N		İstatistik	Standart Hata
Zihinsel Talep	0	Ortalama	10,22	1,10
		Standart Sapma	7,38	
	1	Ortalama	22,33	2,08
		Standart Sapma	13,96	
	2	Ortalama	47,89	3,00
		Standart Sapma	20,16	
	3	Ortalama	73,11	3,22
		Standart Sapma	21,59	
Fiziksel Talep	0	Ortalama	5,89	0,68
		Standart Sapma	4,56	
	1	Ortalama	6,67	0,62
		Standart Sapma	4,13	
	2	Ortalama	8,89	1,04
		Standart Sapma	6,98	
	3	Ortalama	17,22	4,06
		Standart Sapma	27,21	
Zaman Baskısı	0	Ortalama	12,00	1,44
		Standart Sapma	9,68	
	1	Ortalama	22,89	2,85
		Standart Sapma	19,11	
	2	Ortalama	41,11	2,95
		Standart Sapma	19,80	
	3	Ortalama	60,89	3,00
		Standart Sapma	20,15	
Performans	0	Ortalama	11,22	1,41
		Standart Sapma	9,48	
	1	Ortalama	18,67	2,39
		Standart Sapma	16,04	
	2	Ortalama	34,89	2,97
		Standart Sapma	19,90	
	3	Ortalama	55,00	3,31
		Standart Sapma	22,18	
Çaba	0	Ortalama	13,89	1,67
		Standart Sapma	11,17	
	1	Ortalama	27,67	2,85
		Standart Sapma	19,15	
	2	Ortalama	49,89	3,21
		Standart Sapma	21,55	
	3	Ortalama	69,33	2,70
		Standart Sapma	18,11	

Çizelge 5.2. (devam) Tanımlayıcı istatistikler

	N		İstatistik	Standart Hata
Gerginlik	0	Ortalama	10,67	1,38
		Standart Sapma	9,27	
	1	Ortalama	20,22	2,60
		Standart Sapma	17,45	
	2	Ortalama	34,22	3,49
		Standart Sapma	23,40	
	3	Ortalama	48,67	4,66
		Standart Sapma	31,25	
Ağırlıklı Toplam NASA-TLX Skoru	0	Ortalama	11,56	1,23
		Standart Sapma	8,27	
	1	Ortalama	21,63	2,05
		Standart Sapma	13,75	
	2	Ortalama	43,72	2,75
		Standart Sapma	18,42	
	3	Ortalama	62,86	2,77
		Standart Sapma	18,56	
Doğru Yanıt Sayısı	0	Ortalama	14,62	0,11
		Standart Sapma	0,72	
	1	Ortalama	14,27	0,16
		Standart Sapma	1,05	
	2	Ortalama	13,44	0,35
		Standart Sapma	2,37	
	3	Ortalama	12,29	0,40
		Standart Sapma	2,69	
Yanlış Yanıt Sayısı	0	Ortalama	0,38	0,11
		Standart Sapma	0,72	
	1	Ortalama	0,73	0,16
		Standart Sapma	1,05	
	2	Ortalama	1,56	0,35
		Standart Sapma	2,37	
	3	Ortalama	2,71	0,40
		Standart Sapma	2,69	
Hedefi Bulma Sayısı	0	Ortalama	4,80	0,08
		Standart Sapma	0,50	
	1	Ortalama	4,69	0,09
		Standart Sapma	0,60	
	2	Ortalama	4,24	0,15
		Standart Sapma	1,03	
	3	Ortalama	3,76	0,17
		Standart Sapma	1,11	

Çizelge 5.2. (devam) Tanımlayıcı istatistikler

	N		İstatistik	Standart Hata
Hedef Olmayanı Bulma Sayısı	0	Ortalama	9,82	0,07
		Standart Sapma	0,49	
	1	Ortalama	9,58	0,12
		Standart Sapma	0,81	
	2	Ortalama	9,20	0,25
		Standart Sapma	1,69	
	3	Ortalama	8,53	0,29
		Standart Sapma	1,94	
Hedefi Bulamama Sayısı	0	Ortalama	0,20	0,08
		Standart Sapma	0,50	
	1	Ortalama	0,31	0,09
		Standart Sapma	0,60	
	2	Ortalama	0,76	0,15
		Standart Sapma	1,03	
	3	Ortalama	1,24	0,17
		Standart Sapma	1,11	
Hedef Olmayanı Bulamama Sayısı	0	Ortalama	0,18	0,07
		Standart Sapma	0,49	
	1	Ortalama	0,42	0,12
		Standart Sapma	0,81	
	2	Ortalama	0,80	0,25
		Standart Sapma	1,69	
	3	Ortalama	1,47	0,29
		Standart Sapma	1,94	
Ortalama Tepki Süresi	0	Ortalama	540,91	10,39
		Standart Sapma	69,68	
	1	Ortalama	666,23	19,93
		Standart Sapma	133,71	
	2	Ortalama	838,37	43,54
		Standart Sapma	292,11	
	3	Ortalama	978,24	53,68
		Standart Sapma	360,07	
Toplam Tepki Süresi	0	Ortalama	8113,69	155,81
		Standart Sapma	1045,18	
	1	Ortalama	9993,40	298,98
		Standart Sapma	2005,62	
	2	Ortalama	12575,51	653,17
		Standart Sapma	4381,60	
	3	Ortalama	14673,60	805,15
		Standart Sapma	5401,13	
Doğru Yanıt Oranı	0	Ortalama	0,98	0,01
		Standart Sapma	0,04	
	1	Ortalama	0,95	0,01
		Standart Sapma	0,07	
	2	Ortalama	0,89	0,02
		Standart Sapma	0,16	
	3	Ortalama	0,81	0,03
		Standart Sapma	0,18	

Çarpıklık ve basıklık istatistik değerleri incelendiğinde ise söz konusu değerlerin standart hata değerine bölümünden elde edilen sonucun “Çaba” değişkeni dışındaki diğer tüm değişkenlerde -1,96 - 1,96 aralığının dışında yer alması neticesinde değişkenlerin normal dağılımdan gelmediği sonucuna varılmıştır.

Çizelge 5.3. Çarpıklık ve basıklık değerleri

		İstatistik	Standart Hata
Zihinsel Talep	Çarpıklık	0,548	0,181
	Basıklık	-0,943	0,360
Fiziksel Talep	Çarpıklık	4,763	0,181
	Basıklık	23,816	0,360
Zaman Baskısı	Çarpıklık	0,603	0,181
	Basıklık	-0,758	0,360
Performans	Çarpıklık	0,842	0,181
	Basıklık	-0,376	0,360
Çaba	Çarpıklık	0,343	0,181
	Basıklık	-1,199	0,360
Gerginlik	Çarpıklık	1,000	0,181
	Basıklık	-0,256	0,360
Ağırlıklı Toplam NASA-TLX Skoru	Çarpıklık	0,548	0,181
	Basıklık	-0,854	0,360
Doğru Sayısı	Çarpıklık	-3,027	0,181
	Basıklık	11,824	0,360
Yanlış Sayısı	Çarpıklık	3,027	0,181
	Basıklık	11,824	0,360
Hedefi Bulma Sayısı	Çarpıklık	-1,953	0,181
	Basıklık	4,915	0,360
Hedef Olmayan Bulma Sayısı	Çarpıklık	-3,206	0,181
	Basıklık	11,690	0,360
Hedefi Bulamama Sayısı	Çarpıklık	1,953	0,181
	Basıklık	4,915	0,360
Hedef Olmayan Bulamama Sayısı	Çarpıklık	3,206	0,181
	Basıklık	11,690	0,360
Ortalama Tepki Süresi	Çarpıklık	2,032	0,181
	Basıklık	4,821	0,360
Toplam Tepki Süresi	Çarpıklık	2,032	0,181
	Basıklık	4,821	0,360
Doğru Yanıt Oranı	Çarpıklık	-2,943	0,181
	Basıklık	10,840	0,360

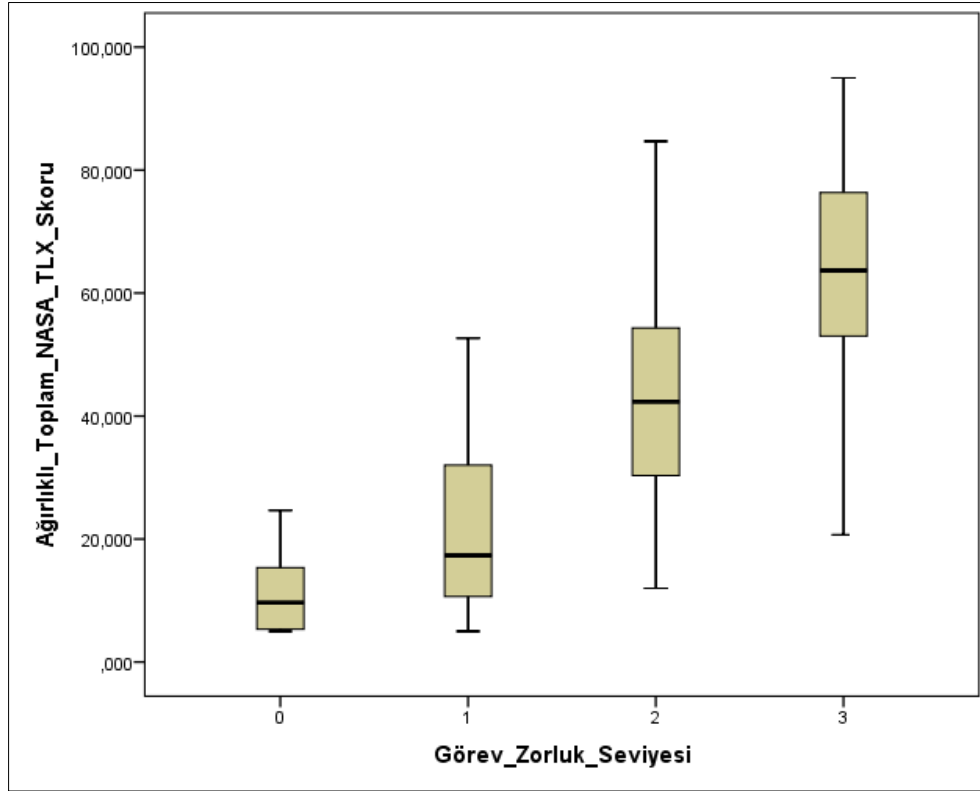
Bu sonuçlar doğrultusunda analizlere parametrik olmayan testlerle devam edilmesine karar verilmiştir. Korelasyon testlerinde parametrik olmayan Spearman korelasyon katsayısından yararlanılmıştır. Görev zorluk seviyesi ile ağırlıklı toplam NASA-TLX skoru arasında %99 güven düzeyinde anlamlı 0,8 korelasyon katsayılı pozitif yönlü ilişki bulunmuştur. Dolayısıyla görev zorluk seviyesi arttıkça katılımcıların algıladıkları iş yükünün arttığı söylenebilmektedir. Bu sonucu, Şekil 5.1’de verilen histogram grafiğinde görebilmek mümkündür. 0-geri görevlerde beyan edilen toplam iş yükü %10 civarında iken, 3-geri görevlerde %60’ın üzerine çıkmıştır. Zihinsel talep, fiziksel talep, zaman baskısı, performans, çaba, gerginlik alt boyutları ile görev zorluk seviyesi arasındaki korelasyon katsayıları ise sırasıyla 0,825; 0,303; 0,744; 0,709; 0,770 ve 0,522 olarak bulunmuştur. Katılımcılara göre görev zorlaştıkça en çok artış gösteren alt boyut zihinsel talep olmuştur.

Korelasyon testlerinin bir sonucu olarak, “Ağırlıklı Toplam NASA-TLX Skoru” ile en çok ilişkili alt boyutların sırasıyla “Zihinsel Talep”, “Çaba”, “Zaman Baskısı” ve “Performans” olduğu görülmüştür. Korelasyon testi sonuçları Çizelge 5.4’te verilmiştir. Korelasyon testi sonuçlarının verildiği çizelgelerde korelasyon katsayısının yanındaki tek yıldız işareti (*) %95 güven düzeyinde anlamlı sonuca çift yıldız işareti (**) ise %99 güven düzeyinde anlamlı sonuca işaret etmektedir.

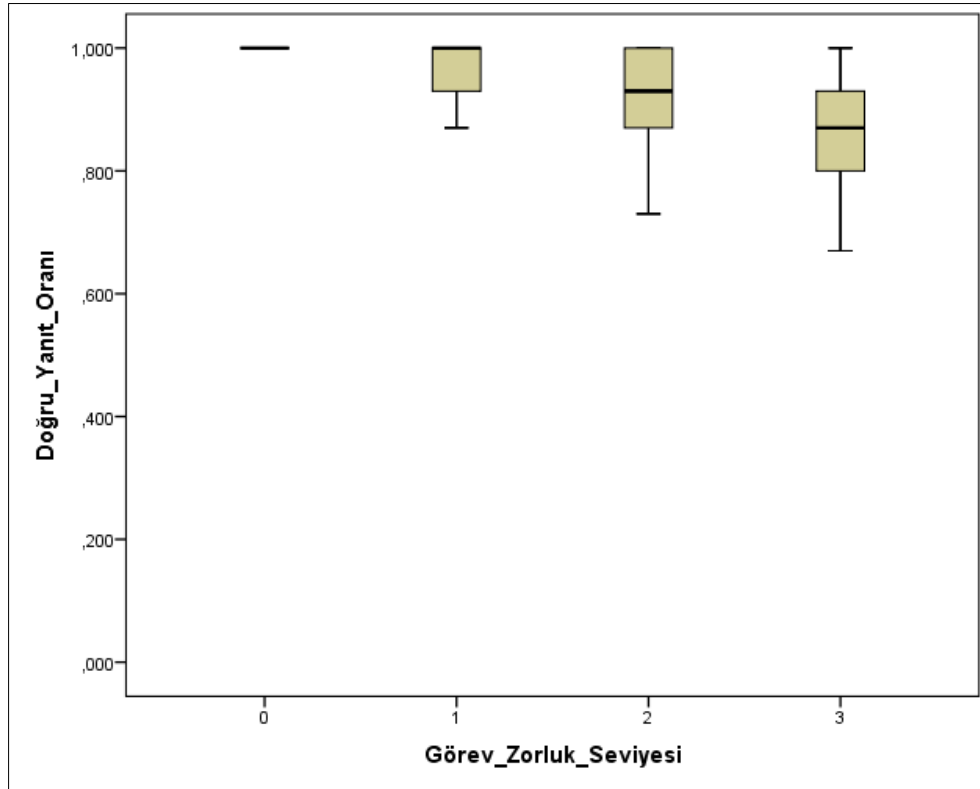
Çizelge 5.4. Ağırlıklı toplam NASA-TLX skoru ile alt boyutları arasındaki ilişki

	Korelasyon Katsayısı	Anlamlılık (2-tarflı)	N
Zihinsel Talep	0,963**	0,000	180
Fiziksel Talep	0,283**	0,000	180
Zaman Baskısı	0,944**	0,000	180
Performans	0,845**	0,000	180
Çaba	0,950**	0,000	180
Gerginlik	0,762**	0,000	180

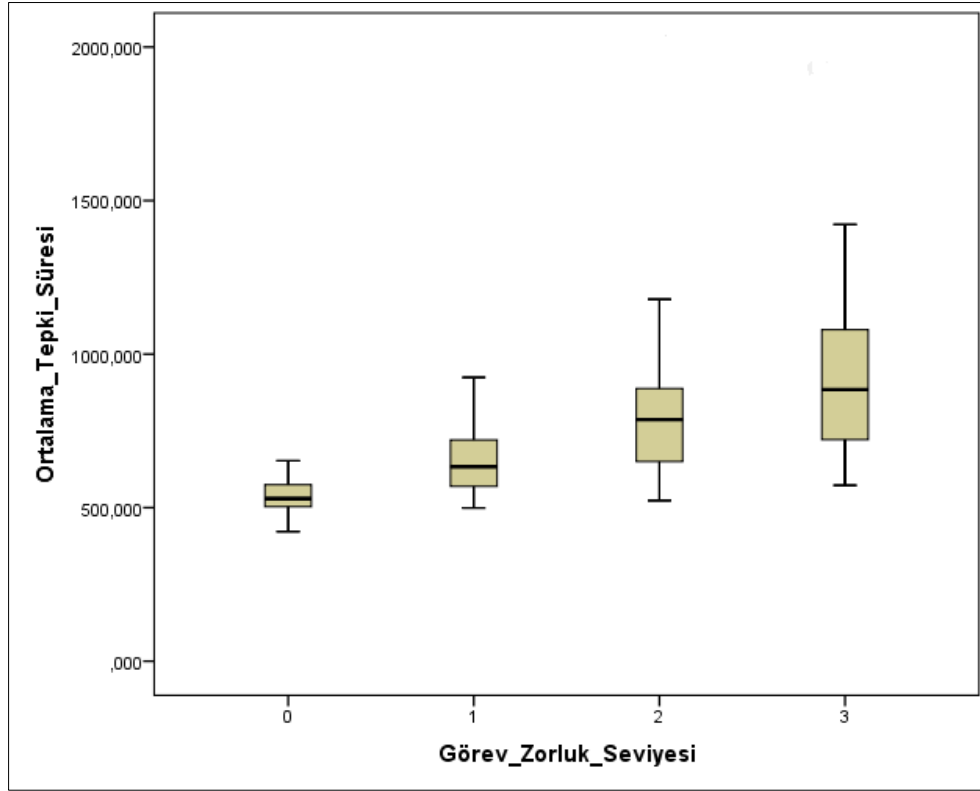
Görev zorlaştıkça ortalama tepki süresinin arttığını, doğru yanıt oranının ise azaldığını gösteren histogram grafikleri Şekil 5.2 ve Şekil 5.3’te görülmektedir. Ortalama değerler ile birlikte incelediğimizde, 0-geri görevlerde doğru yanıt oranı ortalaması %98’i bulurken 3-geri görevlerde %81’e kadar düşmektedir. Benzer şekilde 0-geri görevlerde ortalama tepki süresi yaklaşık 541 ms iken, 3-geri görevlerde ortalama tepki süresi yaklaşık 978 ms olarak ölçülmüştür.



Şekil 5.1. Ağırlıklı toplam NASA-TLX skoru ve zorluk seviyesi arasındaki ilişki



Şekil 5.2. Doğru yanıt oranı ve zorluk seviyesi arasındaki ilişki



Şekil 5.3. Ortalama tepki süresi ve zorluk seviyesi arasındaki ilişki

Çizelge 5.5'te görev zorluk seviyesi ile performans sonuçları arasında da anlamlı ölçüde ilişki olduğu görülmektedir. Özellikle; yanıt için verilen tepki süresindeki gecikme, önceden belirlenen zorluk seviyesi arttıkça büyük ölçüde artmaktadır. Doğru cevap verme oranı ise beklendiği üzere görev zorlaştıkça azalmaktadır.

Çizelge 5.5. Görev zorluk seviyesi ile performans değişkenleri arasındaki ilişki

Ortalama Tepki Süresi	Korelasyon Katsayısı	0,691**
	Anlamlılık (2-terafli)	0,000
	N	180
Doğru Yanıt Oranı	Korelasyon Katsayısı	-0,549**
	Anlamlılık (2-terafli)	0,000
	N	180

Tüm NASA-TLX skorları ile tüm performans değişkenleri arasında anlamlı ölçüde korelasyon bulunmaktadır. Ağırlıklı toplam TLX skoru arttıkça testlerdeki doğru yanıt oranı düşmekte, tepki süresindeki gecikme ise artmaktadır. Korelasyon katsayıları Çizelge

5.6’da verilmiştir. NASA-TLX’in “Performans” alt boyutuna katılımcıların verdikleri puanların gerçek performans sonuçlarıyla tutarlı olduğu göze çarpmaktadır.

NASA-TLX’in “Fiziksel Talep” alt boyutunun hem ağırlıklı toplam TLX skoru hem de performans değişkenleri ile düşük korelasyonlu olduğu görülmektedir. Bu alt boyutun, yapılan ikili karşılaştırmaların tamamında sıfır puan alması ve bu nedenle ağırlık katsayısının sıfır olarak alınması nedeniyle elde edilen ağırlıklı toplam TLX skorunun hesaplanmasında bir etkisi olmamıştır. Bu sebeple, analizin devamında “Fiziksel Talep” alt boyutuna ilişkin sonuçlar dikkate alınmayacaktır.

Çizelge 5.6. NASA-TLX ile performans değişkenleri arasındaki ilişki

NASA-TLX Değişkenleri	İstatistik	Ortalama Tepki Süresi ile Korelasyon Katsayısı	Doğru Yanıt Oranı ile Korelasyon Katsayısı
Zihinsel Talep	Korelasyon Katsayısı	0,561**	-0,536**
	Anlamlılık (2-terafı)	0,000	0,000
	N	180	180
Fiziksel Talep	Korelasyon Katsayısı	0,207**	-0,182*
	Anlamlılık (2-terafı)	0,005	0,015
	N	180	180
Zaman Baskısı	Korelasyon Katsayısı	0,543**	-0,523**
	Anlamlılık (2-terafı)	0,000	0,000
	N	180	180
Performans	Korelasyon Katsayısı	0,475**	-0,596**
	Anlamlılık (2-terafı)	0,000	0,000
	N	180	180
Çaba	Korelasyon Katsayısı	0,556**	-0,521**
	Anlamlılık (2-terafı)	0,000	0,000
	N	180	180
Gerginlik	Korelasyon Katsayısı	0,378**	-0,358**
	Anlamlılık (2-terafı)	0,000	0,000
	N	180	180
Ağırlıklı Toplam NASA-TLX Skoru	Korelasyon Katsayısı	0,555**	-0,569**
	Anlamlılık (2-terafı)	0,000	0,000
	N	180	180

Kruskal-Wallis (K-W) testi ile NASA-TLX ve performans değişkenleri açısından zorluk seviyeleri arasında anlamlı bir fark olup olmadığı incelenmiştir. Ele alınan tüm değişkenlerin en az iki farklı zorluk seviyesi için %99,9 anlamlılık düzeyinde farklılık

gösterdiği ortaya çıkmıştır. K-W testi sonuçları Çizelge 5.7’de verilmiştir. Çizelge 5.8’de değişkenlerin zorluk seviyelerine göre ortalama sıralama (rank) değerlerinin görev zorluk kategorilerine göre düzgün artış veya azalış gösterdiği görülmektedir.

Çizelge 5.7. Görev zorluk seviyesine göre K-W testi sonuçları

NASA-TLX ve Performans Değişkenleri	Ki-Kare	Serbestlik Derecesi	Anlamlılık
Zihinsel Talep	122,442	3	0,000
Zaman Baskısı	99,580	3	0,000
Performans	90,951	3	0,000
Çaba	106,587	3	0,000
Gerginlik	49,675	3	0,000
Ağırlıklı Toplam NASA-TLX Skoru	115,415	3	0,000
Doğru Yanıt Sayısı	46,353	3	0,000
Hedefi Bulma Sayısı	40,807	3	0,000
Hedef Olmayan Bulma Sayısı	27,619	3	0,000
Ortalama Tepki Süresi	87,947	3	0,000
Doğru Yanıt Oranı	55,288	3	0,000

Çizelge 5.8. NASA-TLX ve performans değişkenleri için K-W testi sıralamaları

	Görev Zorluk Seviyesi	N	Ortalama Sıralama
Zihinsel Talep	0	45	35,42
	1	45	66,68
	2	45	112,43
	3	45	147,47
	Toplam	180	
Zaman Baskısı	0	45	41,16
	1	45	68,62
	2	45	110,21
	3	45	142,01
	Toplam	180	
Performans	0	45	45,36
	1	45	67,47
	2	45	108,43
	3	45	140,74
	Toplam	180	

Çizelge 5.8. (devam) NASA-TLX ve performans değişkenleri için K-W testi sıralamaları

	Görev Zorluk Seviyesi	N	Ortalama Sıralama
Çaba	0	45	38,10
	1	45	69,40
	2	45	111,36
	3	45	143,14
	Toplam	180	
Gerginlik	0	45	53,09
	1	45	77,69
	2	45	108,31
	3	45	122,91
	Toplam	180	
Ağırlıklı Toplam NASA-TLX Skoru	0	45	37,06
	1	45	66,31
	2	45	113,44
	3	45	145,19
	Toplam	180	
Doğru Yanıt Sayısı	0	45	119,70
	1	45	103,92
	2	45	85,74
	3	45	52,63
	Toplam	180	
Hedefi Bulma Sayısı	0	45	114,11
	1	45	106,09
	2	45	84,00
	3	45	57,80
	Toplam	180	
Hedef Olmayanı Bulma Sayısı	0	45	111,14
	1	45	97,81
	2	45	89,46
	3	45	63,59
	Toplam	180	
Ortalama Tepki Süresi	0	45	36,20
	1	45	80,69
	2	45	112,40
	3	45	132,71
	Toplam	180	
Doğru Yanıt Oranı	0	45	122,88
	1	45	105,14
	2	45	84,30
	3	45	49,68
	Toplam	180	

Tüm zorluk seviyeleri ikili olarak da karşılaştırılmış ve ele alınan değişkenler açısından aralarında anlamlı bir fark olup olmadığı incelenmiştir. Çizelge 5.9’da görüleceği üzere; ikili olarak karşılaştırılan tüm zorluk seviyeleri arasında, tüm NASA-TLX değişkenleri açısından anlamlı fark olduğu ortaya çıkmıştır. 0-geri görev ile 1-geri görev arasında doğru sayısı ile hedefi ve hedef olmayanı bulma sayıları açısından anlamlı bir fark bulunamamıştır. 1-geri görev ile 2-geri görev arasında ise doğru sayısı ve hedef olmayanı bulma sayısı açısından anlamlı fark bulunamamıştır.

Çizelge 5.9. Görev zorluk seviyelerine göre Mann-Whitney U testi sonuçları

0 geri-1 geri	Mann-Whitney U	Z	Anlamlılık (2-taraflı)
Zihinsel Talep	459,000	-4,582	0,000
Zaman Baskısı	624,000	-3,217	0,001
Performans	693,500	-2,668	0,008
Çaba	527,000	-3,965	0,000
Gerginlik	694,000	-2,665	0,008
Ağırlıklı Toplam NASA-TLX Skoru	519,500	-3,988	0,000
Doğru Yanıt Sayısı	807,000	-1,940	0,052
Hedefi Bulma Sayısı	907,500	-1,194	0,232
Hedef Olmayanı Bulma Sayısı	855,000	-1,787	0,074
Ortalama Tepki Süresi	383,000	-5,080	0,000
Doğru Yanıt Oranı	788,500	-2,155	0,031
0 geri-2 geri	Mann-Whitney U	Z	Anlamlılık (2-taraflı)
Zihinsel Talep	83,000	-7,585	0,000
Zaman Baskısı	155,000	-6,987	0,000
Performans	242,500	-6,301	0,000
Çaba	128,500	-7,172	0,000
Gerginlik	330,500	-5,594	0,000
Ağırlıklı Toplam NASA-TLX Skoru	92,500	-7,432	0,000
Doğru Yanıt Sayısı	636,000	-3,400	0,001
Hedefi Bulma Sayısı	682,500	-3,264	0,001
Hedef Olmayanı Bulma Sayısı	763,000	-2,640	0,008
Ortalama Tepki Süresi	159,000	-6,887	0,000
Doğru Yanıt Oranı	577,000	-3,938	0,000
0 geri-3 geri	Mann-Whitney U	Z	Anlamlılık (2-taraflı)
Zihinsel Talep	17,000	-8,116	0,000
Zaman Baskısı	38,000	-7,934	0,000
Performans	70,000	-7,680	0,000
Çaba	24,000	-8,011	0,000
Gerginlik	329,500	-5,607	0,000
Ağırlıklı Toplam NASA-TLX Skoru	20,500	-8,013	0,000

Çizelge 5.9. (devam) Görev zorluk seviyelerine göre Mann-Whitney U testi sonuçları

0 geri-3 geri	Mann-Whitney U	Z	Anlamlılık (2-taraflı)
Doğru Yanıt Sayısı	280,500	-6,205	0,000
Hedefi Bulma Sayısı	385,000	-5,608	0,000
Hedef Olmayan Bulma Sayısı	490,500	-4,859	0,000
Ortalama Tepki Süresi	52,000	-7,751	0,000
Doğru Yanıt Oranı	215,000	-6,762	0,000
1 geri-2 geri	Mann-Whitney U	Z	Anlamlılık (2-taraflı)
Zihinsel Talep	313,500	-5,659	0,000
Zaman Baskısı	464,000	-4,446	0,000
Performans	480,500	-4,325	0,000
Çaba	443,500	-4,605	0,000
Gerginlik	633,000	-3,086	0,002
Ağırlıklı Toplam NASA-TLX Skoru	351,000	-5,339	0,000
Doğru Yanıt Sayısı	804,000	-1,802	0,071
Hedefi Bulma Sayısı	764,500	-2,342	0,019
Hedef Olmayan Bulma Sayısı	918,000	-,917	0,359
Ortalama Tepki Süresi	609,000	-3,256	0,001
Doğru Yanıt Oranı	771,000	-2,088	0,037
1 geri-3 geri	Mann-Whitney U	Z	Anlamlılık (2-taraflı)
Zihinsel Talep	86,000	-7,492	0,000
Zaman Baskısı	188,000	-6,670	0,000
Performans	189,000	-6,667	0,000
Çaba	146,500	-7,001	0,000
Gerginlik	497,000	-4,193	0,000
Ağırlıklı Toplam NASA-TLX Skoru	92,500	-7,425	0,000
Doğru Yanıt Sayısı	411,500	-5,007	0,000
Hedefi Bulma Sayısı	454,000	-4,914	0,000
Hedef Olmayan Bulma Sayısı	620,500	-3,486	0,000
Ortalama Tepki Süresi	345,000	-5,387	0,000
Doğru Yanıt Oranı	371,000	-5,338	0,000
2 geri-3 geri	Mann-Whitney U	Z	Anlamlılık (2-taraflı)
Zihinsel Talep	371,000	-5,192	0,000
Zaman Baskısı	493,500	-4,202	0,000
Performans	517,500	-4,007	0,000
Çaba	498,000	-4,164	0,000
Gerginlik	752,500	-2,105	0,035
Ağırlıklı Toplam NASA-TLX Skoru	463,500	-4,431	0,000
Doğru Yanıt Sayısı	641,500	-3,060	0,002
Hedefi Bulma Sayısı	727,000	-2,439	0,015
Hedef Olmayan Bulma Sayısı	715,500	-2,593	0,010
Ortalama Tepki Süresi	741,000	-2,191	0,028
Doğru Yanıt Oranı	614,500	-3,277	0,001

Kadın ve erkek katılımcılar arasında NASA-TLX ve performans değişkenleri açısından anlamlı bir fark olup olmadığı incelenmiş ve NASA-TLX'in zaman baskısı alt boyutu açısından aralarında anlamlı fark olduğu görülmüştür. Performans değişkenlerinden ise ortalama tepki süresi açısından anlamlı fark bulunmuştur. Sonuçlar Çizelge 5.10'da verilmiştir. Erkekler tarafından hissedilen zaman baskısı kadınlara göre daha fazladır. Ortalama tepki süresinin de yine erkeklerde daha uzun olduğu görülmüştür.

Çizelge 5.10. Cinsiyete göre Mann-Whitney U testi sonuçları

NASA-TLX ve Performans Değişkenleri	Ki-Kare	Serbestlik Derecesi	Anlamlılık
Zihinsel Talep	3783,500	-0,715	0,474
Zaman Baskısı	3291,000	-2,134	0,033
Performans	3784,000	-0,716	0,474
Çaba	3956,500	-0,217	0,828
Gerginlik	3566,000	-1,351	0,177
Ağırlıklı Toplam NASA-TLX Skoru	3815,000	-0,622	0,534
Doğru Yanıt Sayısı	3772,500	-0,789	0,430
Hedefi Bulma Sayısı	3804,500	-0,740	0,459
Hedef Olmayanı Bulma Sayısı	3745,000	-0,968	0,333
Ortalama Tepki Süresi	3295,000	-2,113	0,035
Doğru Yanıt Oranı	3770,000	-0,796	0,426

K-W testi ile NASA-TLX ve performans değişkenleri açısından katılımcıların bulunduğu sınıflar arasında anlamlı bir fark olup olmadığı incelenmiştir. Hedefi ve hedef olmayanı bulma sayıları dışındaki tüm değişkenlerin en az iki farklı sınıf için %95 anlamlılık düzeyinde farklılık gösterdiği ortaya çıkmıştır. Sınıflar ikili olarak da karşılaştırılmış, hangi sınıflar arasında hangi değişkenler açısından anlamlı fark olduğu incelenmiştir. Sonuçlar Çizelge 5.11'de görülmektedir.

Dört farklı sınıfın ikili karşılaştırmaları neticesinde, özellikle 1. sınıftaki katılımcıların diğer sınıflardaki katılımcılara göre daha düşük NASA-TLX skorları bildirdikleri ve ayrıca, tepki süresi veya doğruluk yüzdesi açısından üst sınıflardan daha iyi performans elde ettikleri dikkati çekmiştir. Ek olarak, 3. sınıftaki katılımcılar, 2. ve 4. sınıftakilere göre tepki süresi açısından daha iyi bir performans göstermişlerdir.

Çizelge 5.11. Sınıflara göre Mann-Whitney U testi sonuçları

1. sınıf-2. sınıf	Mann-Whitney U	Z	Anlamlılık (2-taraflı)
Zihinsel Talep	434,500	-3,904	0,000
Zaman Baskısı	424,000	-3,998	0,000
Performans	628,000	-2,153	0,031
Çaba	435,000	-3,894	0,000
Gerginlik	551,000	-2,906	0,004
Ağırlıklı Toplam NASA-TLX Skoru	421,500	-4,002	0,000
Doğru Yanıt Sayısı	736,000	-1,264	0,206
Hedefi Bulma Sayısı	753,000	-1,208	0,227
Hedef Olmayan Bulma Sayısı	815,000	-0,567	0,571
Ortalama Tepki Süresi	580,000	-2,567	0,010
Doğru Yanıt Oranı	742,500	-1,201	0,230
1. sınıf-3. sınıf	Mann-Whitney U	Z	Anlamlılık (2-taraflı)
Zihinsel Talep	773,000	-3,432	0,001
Zaman Baskısı	776,500	-3,407	0,001
Performans	785,500	-3,349	0,001
Çaba	722,500	-3,755	0,000
Gerginlik	626,000	-4,419	0,000
Ağırlıklı Toplam NASA-TLX Skoru	711,000	-3,818	0,000
Doğru Yanıt Sayısı	885,000	-2,827	0,005
Hedefi Bulma Sayısı	948,500	-2,551	0,011
Hedef Olmayan Bulma Sayısı	993,000	-2,278	0,023
Ortalama Tepki Süresi	1206,000	-0,587	0,558
Doğru Yanıt Oranı	885,000	-2,827	0,005
1. sınıf-4. sınıf	Mann-Whitney U	Z	Anlamlılık (2-taraflı)
Zihinsel Talep	327,000	-1,611	0,107
Zaman Baskısı	380,000	-0,799	0,424
Performans	411,000	-0,323	0,747
Çaba	261,000	-2,605	0,009
Gerginlik	253,000	-2,781	0,005
Ağırlıklı Toplam NASA-TLX Skoru	326,500	-1,594	0,111
Doğru Yanıt Sayısı	348,500	-1,381	0,167
Hedefi Bulma Sayısı	357,000	-1,365	0,172
Hedef Olmayan Bulma Sayısı	372,500	-1,116	0,264
Ortalama Tepki Süresi	304,000	-1,931	0,053
Doğru Yanıt Oranı	348,500	-1,381	0,167

Çizelge 5.11. (devam) Sınıflara göre Mann-Whitney U testi sonuçları

2. sınıf-3. sınıf	Mann-Whitney U	Z	Anlamlılık (2-taraflı)
Zihinsel Talep	1616,500	-0,599	0,549
Zaman Baskısı	1496,500	-1,243	0,214
Performans	1437,000	-1,565	0,117
Çaba	1709,000	-0,102	0,919
Gerginlik	1505,000	-1,202	0,229
Ağırlıklı Toplam NASA-TLX Skoru	1692,500	-0,190	0,849
Doğru Yanıt Sayısı	1410,500	-1,780	0,075
Hedefi Bulma Sayısı	1497,500	-1,370	0,171
Hedef Olmayan Bulma Sayısı	1415,000	-1,927	0,054
Ortalama Tepki Süresi	1192,000	-2,871	0,004
Doğru Yanıt Oranı	1410,000	-1,784	0,074
2. sınıf-4. sınıf	Mann-Whitney U	Z	Anlamlılık (2-taraflı)
Zihinsel Talep	410,500	-1,983	0,047
Zaman Baskısı	323,000	-3,032	0,002
Performans	448,500	-1,534	0,125
Çaba	495,000	-0,971	0,332
Gerginlik	565,000	-0,133	0,894
Ağırlıklı Toplam NASA-TLX Skoru	385,500	-2,276	0,023
Doğru Yanıt Sayısı	550,000	-0,332	0,740
Hedefi Bulma Sayısı	555,000	-0,288	0,773
Hedef Olmayan Bulma Sayısı	531,000	-0,654	0,513
Ortalama Tepki Süresi	568,000	-0,096	0,924
Doğru Yanıt Oranı	548,000	-0,357	0,721
3. sınıf-4. sınıf	Mann-Whitney U	Z	Anlamlılık (2-taraflı)
Zihinsel Talep	696,500	-1,423	0,155
Zaman Baskısı	584,500	-2,375	0,018
Performans	559,000	-2,595	0,009
Çaba	742,500	-1,030	0,303
Gerginlik	716,500	-1,254	0,210
Ağırlıklı Toplam NASA-TLX Skoru	615,500	-2,103	0,035
Doğru Yanıt Sayısı	744,500	-1,052	0,293
Hedefi Bulma Sayısı	779,000	-,788	0,431
Hedef Olmayan Bulma Sayısı	772,000	-,871	0,384
Ortalama Tepki Süresi	585,000	-2,361	0,018
Doğru Yanıt Oranı	744,500	-1,052	0,293

Oturum bazında ele alınan ilk veri setinden elde edilen bulgular neticesinde, önceden belirlenen görev zorluk seviyeleri ile zihinsel iş yükünün tetiklendiği doğrulanmıştır. Bu sonuç doğrultusunda; analize uyaran bazında ele alınan veri seti ile devam edilmiş, EEG ve göz izleme değişkenleri ile görev zorluk seviyesi ve NASA-TLX sonuçları arasındaki ilişki

irdelenmiştir. Hem göz izleme hem de EEG değişkenlerinin normal dağılıma uymaması sebebiyle analizler parametrik olmayan testlerle gerçekleştirilmiştir.

5.2.2. Göz izleme değişkenlerine ilişkin bulgular

Spearman korelasyon testi aracılığıyla göz izleme değişkenleri ile NASA-TLX ve görev zorluk seviyeleri arasında ilişki olup olmadığı araştırılmıştır. Çizelge 5.12'ye göre her iki değişkenin de hemen hemen tüm göz izleme değişkenleriyle ilişkili olduğu ortaya çıkmıştır.

Çizelge 5.12. Göz izleme değişkenlerine ilişkin korelasyon testi sonuçları

Göz İzleme Değişkenleri	Görev Zorluk Seviyesi ile Korelasyon Katsayısı	Ağırlıklı Toplam NASA-TLX Skoru ile Korelasyon Katsayısı
Odaklanma Sayısı	0,062**	0,073**
Odaklanma Toplam Süresi	-0,033	-0,025
Odaklanma Süresi Ortalaması	-0,099**	-0,119**
Odaklanma Süresi Maksimum	-0,084**	-0,106**
Odaklanma Süresi Minimum	-0,091**	-0,094**
Odaklanma Süresi Standart Sapması	-0,050**	-0,061**
Seğirme Genliği Ortalaması	0,079**	0,160**
Seğirme Genliği Maksimum	0,079**	0,152**
Seğirme Genliği Standart Sapması	0,085**	0,137**
Seğirme Sayısı	0,139**	0,121**
Seğirme Toplam Süresi	0,117**	0,111**
Seğirme Süresi Ortalaması	0,057	0,075**
Seğirme Süresi Maksimum	0,090**	0,101**
Seğirme Süresi Minimum	-0,084**	-0,050**
Seğirme Süresi Standart Sapması	0,077**	0,085**
Sol Göz Bebeği Ortalaması	0,109**	0,042*
Sağ Göz Bebeği Ortalaması	0,107**	0,072**
Göz Bebeği Ortalaması	0,102**	0,036
Sol Göz Bebeği Maksimum	0,106**	0,039*
Sağ Göz Bebeği Maksimum	0,033	0,023
Göz Bebeği Maksimum	0,060**	0,041*
Sol Göz Bebeği Standart Sapması	0,080**	0,009
Sağ Göz Bebeği Standart Sapması	0,068**	0,044*
Göz Bebeği Standart Sapması	0,072**	-0,006
Göz Kırpma Sayısı	0,195**	0,187**
Göz Kırpma Toplam Süresi	0,168**	0,144**
Göz Kırpma Süresi Ortalaması	0,058**	-0,001

Görev zorluk seviyesi ile en yüksek korelasyonlu değişkenler; göz kırpma sayısı ve toplam süresi, seğirme sayısı ve toplam süresi, sol ve sağ göz bebeği çapı ortalaması olarak bulunmuştur. Ağırlıklı NASA-TLX skoru ile en çok ilişkili değişkenler ise; göz kırpma sayısı ve toplam süresi, seğirme genliği ortalaması, maksimum değeri ve standart sapması ile seğirme sayısı ve toplam süresidir. K-W testi ile de göz izleme değişkenlerinin tamamının görev zorluk seviyelerine göre anlamlı fark gösterdiği görülmüştür. Sonuçlar Çizelge 5.13'te verilmiştir.

Çizelge 5.13. Göz izleme değişkenlerine ilişkin K-W testi sonuçları

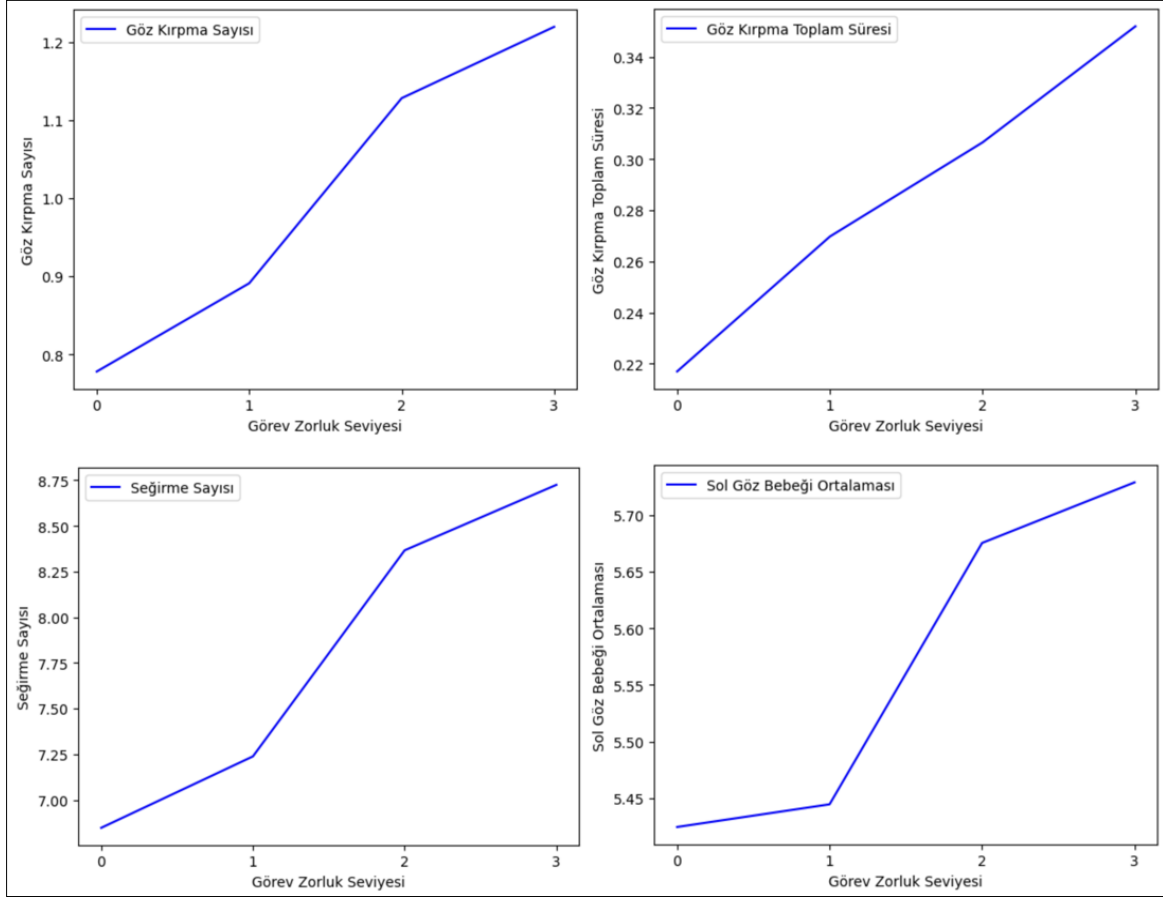
Göz İzleme Değişkenleri	Ki-Kare	Anlamlılık
Odaklanma Sayısı	16,037	0,001
Odaklanma Toplam Süresi	13,327	0,004
Odaklanma Süresi Ortalaması	34,816	0,000
Odaklanma Süresi Maksimum	27,013	0,000
Odaklanma Süresi Minimum	25,367	0,000
Odaklanma Süresi Standart Sapması	17,509	0,001
Seğirme Genliği Ortalaması	40,781	0,000
Seğirme Genliği Maksimum	35,882	0,000
Seğirme Genliği Standart Sapması	31,492	0,000
Seğirme Sayısı	55,448	0,000
Seğirme Toplam Süresi	42,422	0,000
Seğirme Süresi Ortalaması	15,791	0,001
Seğirme Süresi Maksimum	25,668	0,000
Seğirme Süresi Minimum	28,996	0,000
Seğirme Süresi Standart Sapması	19,904	0,000
Sol Göz Bebeği Ortalaması	35,707	0,000
Sağ Göz Bebeği Ortalaması	34,092	0,000
Göz Bebeği Ortalaması	32,543	0,000
Sol Göz Bebeği Maksimum	34,234	0,000
Sağ Göz Bebeği Maksimum	10,977	0,012
Göz Bebeği Maksimum	16,964	0,001
Sol Göz Bebeği Standart Sapması	20,154	0,000
Sağ Göz Bebeği Standart Sapması	39,021	0,000
Göz Bebeği Standart Sapması	18,937	0,000
Göz Kırpma Sayısı	104,888	0,000
Göz Kırpma Toplam Süresi	78,158	0,000
Göz Kırpma Süresi Ortalaması	19,999	0,000

Çizelge 5.14’te verilen değişkenler görev zorluk seviyesine göre düzgün artış ya da azalış gösteren değişkenlerdir. Bu sonuçlar ışığında, çalışmanın devamında kurulacak olan sınıflandırma modelinde dört sınıfı (0-1-2-3) ayırtırmada ve modele katkı sağlamada en etkin değişkenlerin seçirme, göz kırpma ve göz bebeği ile ilişkili değişkenler olacağı tahmin edilmektedir.

Çizelge 5.14. Göz izleme değişkenleri için K-W testi sıralama değerleri

Göz İzleme Değişkenleri	Görev Zorluk Seviyesi	Ortalama Sıralama
Seğirme Sayısı	0	1215,48
	1	1267,59
	2	1439,62
	3	1479,31
Sol Göz Bebeği Ortalaması	0	1245,14
	1	1278,79
	2	1428,32
	3	1449,75
Sağ Göz Bebeği Ortalaması	0	1250,93
	1	1276,79
	2	1424,22
	3	1450,06
Göz Bebeği Ortalaması	0	1253,1
	1	1277,74
	2	1431,02
	3	1440,14
Sol Göz Bebeği Maksimum	0	1248,96
	1	1277,97
	2	1430,38
	3	1444,68
Sol Göz Bebeği Standart Sapması	0	1274,18
	1	1292,81
	2	1415,08
	3	1419,93
Göz Kırpma Sayısı	0	1146,19
	1	1275,20
	2	1454,79
	3	1525,81
Göz Kırpma Toplam Süresi	0	1151,10
	1	1324,09
	2	1420,83
	3	1505,98

Görev zorluk seviyesine göre istatistiksel olarak anlamlı farklılık gösterdiği tespit edilen göz kırpma, seğirme ve göz bebeğine ilişkin değişkenlerin ortalama değerlerinin zorluk seviyesine bağlı değişimini gösteren grafikler Şekil 5.4'te verilmiştir.



Şekil 5.4. Göz izleme değişkenlerinin zorluk seviyesine göre değişimi

Uyaran bazında ele alınan veri setinde verilen yanıtın doğru olup olmadığını gösteren değişken olan doğruluk değişkeninin aldığı 0 ya da 1 değerinin göz izleme değişkenleri açısından anlamlı bir fark gösterip göstermediğine bakıldığında ise, göz kırpma toplam sayı ve süresinin yanlış cevap verilen uyaranlarda daha yüksek olduğu ve göz bebeği çapının da yanlış cevap verilen uyaranlar boyunca daha yüksek ölçüldüğü sonucuna varılmıştır ($p < 0,01$). Kadınlar ve erkekler arasında farklı görev koşullarında göz izleme değişkenleri açısından anlamlı fark görülen değişkenler incelendiğinde, erkeklerin kadınlardan daha çok ve daha uzun süreli seğirme hareketi gerçekleştirdikleri, daha kısa sürelerle daha fazla sayıda odaklandıkları görülmüştür. Özellikle zor görev koşullarında (2-geri ve 3-geri) kadınların sağ ve sol göz bebeği çapı ortalamalarının erkeklerden daha

büyük ölçüldüğü tespit edilmiştir. Diğer yandan, kadınların ortalama göz kırpma süreleri de daha uzundur ($p < 0,01$).

Daha önce “3. Literatür Araştırması” bölümünde tartışıldığı üzere, göz izleme değişkenlerinin zihinsel iş yükü ile ilişkisinin yönü açısından benzer sonuçlar gösteren çalışmalar olduğu gibi farklı sonuçlar ortaya koyan çalışmalar da mevcuttur. Çizelge 5.15’te, tezin sonuçları ile literatürdeki araştırmaların sonuçları karşılaştırmalı olarak verilmiştir.

Çizelge 5.15. Göze ilişkin değişkenler ile zihinsel iş yükü arasındaki ilişki

Göz Değişkeni Türü	Ölçü	Literatürde Zihinsel İş Yükü ile İlişkisinin Yönü (+ / -)	Mevcut Çalışmada Zihinsel İş Yükü ile İlişkisinin Yönü (+ / -)
Odaklanma	Sayısı	+ / -	+
	Süresi	+ / -	-
Seğirme	Sayısı	+	+
	Süresi	+ / -	+
	Genliği	+ / -	+
Göz bebeği	Çapı	+	+
Göz kırpma	Sayısı	+ / -	+
	Süresi	+ / -	+

Literatürde tutarlı sonuçların elde edildiği göz bebeği çapı ile zihinsel iş yükü arasındaki ilişkinin yönü bu çalışmada da benzer bulunmuştur. Literatürde odaklanma sürelerinin zihinsel iş yükü ile birlikte arttığı kanısı yaygın olmakla birlikte, görev zorlaştıkça odaklanma sürelerinde düşüş olduğunu gösteren çalışmalar da mevcuttur (Borys ve diğerleri, 2017b; Kaczorowska, 2021). Bu çalışmada da odaklanma sürelerinin görev zorluk seviyesi ve algılanan zihinsel iş yükü ile birlikte azaldığı görülmüştür. Daha düşük zorluk seviyelerinde, özellikle de 0-geri görevde, uzun odaklanma süresinin ekranda belirecek olan harfi atlamadan görebilmek ve hızlıca klavyedeki tuşa basabilmek için tetikte olma eğiliminden kaynaklanabileceği düşünülmektedir. 3-geri görevi gerçekleştirirken ise, daha çok harfin hafızada tutulmaya çalışılması sonucu ortaya çıkan daha yoğun zihinsel aktivitenin gözün daha çok kez seğirmesine yol açtığı şeklinde yorumlanmaktadır. Diğer değişkenler açısından benzer çalışmalarla tutarsız bir sonuç elde

edilmemiştir. Literatürde birbirinden farklılık gösteren sonuçların ise görev türünden kaynaklanabileceği düşünülmektedir. Örneğin, sadece görsel uyaran yoğun uçuş görevlerinde göz kırpmaya hızının azaldığını, uçuş görevine bir hafıza görevi eklendiğinde ise göz kırpmaya hızının arttığını gösteren çalışmalar mevcuttur (Wanyan ve diğerleri, 2018; Wilson, 2002; Wilson ve diğerleri, 2004). Bu tezde de göz kırpmaya sayı ve süresinin zihinsel iş yükü ile birlikte artmasının görsel uyaranlı bir hafıza görevinin uygulanmış olmasından kaynaklanabileceği değerlendirilmektedir.

5.2.3. EEG değişkenlerine ilişkin bulgular

Spearman korelasyon testi sonucunda görev zorluk seviyesi ile en yüksek korelasyonu gösteren EEG değişkenlerinin AF3, AF4, F7, F8 ve FC5 kanallarından alınan teta dalgalarına ilişkin değişkenler olduğu tespit edilmiştir (korelasyon katsayısı $> 0,3$, $p < 0,01$). Çizelge 5.16’da verilen sonuçlar; prefrontal, frontal ve frontal merkez bölgelerindeki yani ön beyindeki teta gücünün görev zorluğu arttıkça arttığını göstermektedir. Aynı beyin bölgelerinde teta/alfa gücü oranlarına ilişkin değişkenlerin korelasyon katsayıları biraz daha yüksek bulunmuştur.

Çizelge 5.16. Görev zorluk seviyesi ile EEG değişkenleri arasındaki ilişki

EEG Değişkenleri	Korelasyon Katsayısı
AF3.TetaAlfa	0,366**
AF4.TetaAlfa	0,352**
AF4.Teta	0,350**
FC5.TetaAlfa	0,339**
F7.TetaAlfa	0,335**
F7.Teta	0,330**
F8.TetaAlfa	0,322**
AF3.Teta	0,320**
F8.Teta	0,319**
FC5.Teta	0,303**
FC6.TetaAlfa	0,295**
FC6.Teta	0,291**
F3.TetaAlfa	0,290**
F4.TetaAlfa	0,269**
F3.Teta	0,221**
T7.TetaAlfa	0,174**
O1.TetaAlfa	0,173**
F4.Teta	0,164**

Çizelge 5.16. (devam) Görev zorluk seviyesi ile EEG değişkenleri arasındaki ilişki

EEG Değişkenleri	Korelasyon Katsayısı
O1.Alfa	-0,160**
P7.Alfa	-0,154**
T7.Alfa	-0,136**
F7.Alfa	0,135**
O2.TetaAlfa	0,129**
O2.Alfa	-0,126**
P7.TetaAlfa	0,126**
F8.Alfa	0,118**
F3.DüşükBeta	-0,105**
P7.DüşükBeta	-0,104**
T8.TetaAlfa	0,103**
F3.YüksekBeta	-0,096**
F4.Alfa	-0,094**
P8.Alfa	-0,092**
T7.DüşükBeta	-0,089**
O1.DüşükBeta	-0,088**
T8.Alfa	-0,086**
F3.Alfa	-0,081**
AF3.YüksekBeta	-0,080**
P8.TetaAlfa	0,079**
F4.DüşükBeta	-0,073**
FC5.DüşükBeta	-0,071**
AF4.Alfa	0,071**
AF3.DüşükBeta	-0,067**
T7.Teta	0,064**
FC5.YüksekBeta	-0,060**
O1.Gama	0,060**
T8.DüşükBeta	-0,058**
P8.DüşükBeta	-0,056**
P8.Gama	0,056**
F8.Gama	0,052**
AF4.Gama	0,051**
FC5.Alfa	-0,050**

Korelasyon katsayısı daha düşük olmasına rağmen, görev zorluk seviyesinin O1, O2, T7 ve P7 kanallarında ölçülen alfa gücü ile negatif, F7 ve F8 kanallarında ölçülen alfa gücü ile ise pozitif korelasyonlu olduğu ortaya çıkmıştır. Artan görev zorluk derecesi ile frontal bölgelerdeki alfa gücü artarken; temporal, parietal ve oksipital bölgelerdeki alfa gücü azalmaktadır. Görev zorlaştıkça düşük beta gücünün hemen hemen tüm beyin bölgelerinde önemli ölçüde azaldığı gözlemlenmiştir.

Ağırlıklı NASA-TLX toplam skorunun 75 EEG değişkeni ile %99 güven düzeyinde anlamlı korelasyona sahip oluğu görülmüştür. Çizelge 5.17’de en yüksek korelasyonu gösteren EEG değişkenlerine yer verilmiştir. Ağırlıklı NASA-TLX toplam puanı ile en yüksek korelasyon gösteren EEG değişkenlerinin AF4 ve F8'den alınan teta dalgaları ve AF3, AF4 ve F8'den gelen teta/alfa oranları olduğu görülmüştür (korelasyon katsayısı $> 0,3$, $p < 0,01$). Diğer bir deyişle, algılanan zihinsel iş yükü arttıkça prefrontal ve frontal teta gücü artmaktadır. Diğer yandan, algılanan iş yükünün artmasıyla prefrontal, frontal, parietal ve oksipital beyin bölgelerindeki düşük beta gücünün azaldığı görülmektedir (korelasyon katsayısı $> 0,2$, $p < 0,01$).

Çizelge 5.17. NASA-TLX ile EEG değişkenleri arasındaki ilişki

EEG Değişkenleri	Korelasyon Katsayısı
AF4.Teta	0,334**
AF4.TetaAlfa	0,330**
F8.TetaAlfa	0,327**
AF3.TetaAlfa	0,313**
F8.Teta	0,300**
FC6.TetaAlfa	0,295**
F3.TetaAlfa	0,279**
O2.DüşükBeta	-0,271**
P8.DüşükBeta	-0,263**
F7.TetaAlfa	0,260**
FC5.TetaAlfa	0,260**
F4.TetaAlfa	0,260**
F3.DüşükBeta	-0,259**
AF3.DüşükBeta	-0,252**
O2.YüksekBeta	-0,251**
FC6.Teta	0,242**
AF3.YüksekBeta	-0,235**
O1.DüşükBeta	-0,228**
F4.DüşükBeta	-0,227**
O2.Alfa	-0,216**
P7.Alfa	-0,215**
F3.YüksekBeta	-0,209**
P8.Alfa	-0,209**
T7.Gama	0,209**
F7.Teta	0,208**
P7.DüşükBeta	-0,203**
FC6.DüşükBeta	-0,200**

K-W testi ile zorluk seviyeleri arasında EEG değişkenleri açısından anlamlı bir fark olup olmadığı incelenmiştir. Zorluk seviyelerinin 49 adet EEG değişkeni açısından %99 güven düzeyinde farklılık gösterdiği ortaya çıkmıştır. Değişkenlerin sıra değerleri incelendiğinde 49 değişkenden 34'ünün görev zorluk değişimine göre düzgün bir artış veya azalış gösterdiği dikkat çekmiştir. Bu durum söz konusu değişkenlerin zihinsel iş yükünü dört seviyede ayırtırmada katkı sağlayabileceğine işaret etmektedir. Daha zorlu görev koşullarında (2-geri ve 3-geri), teta gücü prefrontal, frontal ve merkezi bölgelerde daha yüksek, alfa gücü temporal, pariyetal ve oksipital bölgelerde daha düşük ve beta gücü hemen hemen tüm beyin bölgelerinde daha düşük bulunmuştur.

Sonuçlar oldukça zayıf korelasyon katsayılarına rağmen %99 güven düzeyinde anlamlı olması sebebiyle dikkate değer görülmüştür. Beta gücünün zorluk seviyesi artarken azalması literatürdekinden farklı bir bulgu olarak karşımıza çıkmaktadır. Çizelge 5.18'de tezin sonuçları ile literatürdeki araştırmaların sonuçları karşılaştırmalı olarak verilmiştir.

Çizelge 5.18. EEG değişkenleri ile zihinsel iş yükü arasındaki ilişki

Beyin bölümleri	Dalga Tipi	Literatürde Zihinsel İş Yükü ile İlişkisinin Yönü (+ / -)	Mevcut Çalışmada Zihinsel İş Yükü ile İlişkisinin Yönü (+ / -)
Frontal	Teta	+	+
	Alfa	+	+
	Beta	+	-
	Gama	+	+
Pariyetal	Teta	+	-
	Alfa	-	-
	Beta	+	-
	Gama	+	+
Oksipital	Teta	-	-
	Alfa	-	-
	Beta	+	-
	Gama	+	+

Tüm korelasyon testleri değişkenler arasında doğrusal bir ilişki olduğu varsayımı altında yapılmıştır. Burada amaç, olası ilişkinin yönü hakkında fikir sahibi olmaktır. Elde edilen düşük korelasyon katsayıları güçlü doğrusal ilişkiye işaret etmemekle birlikte, istatistiksel

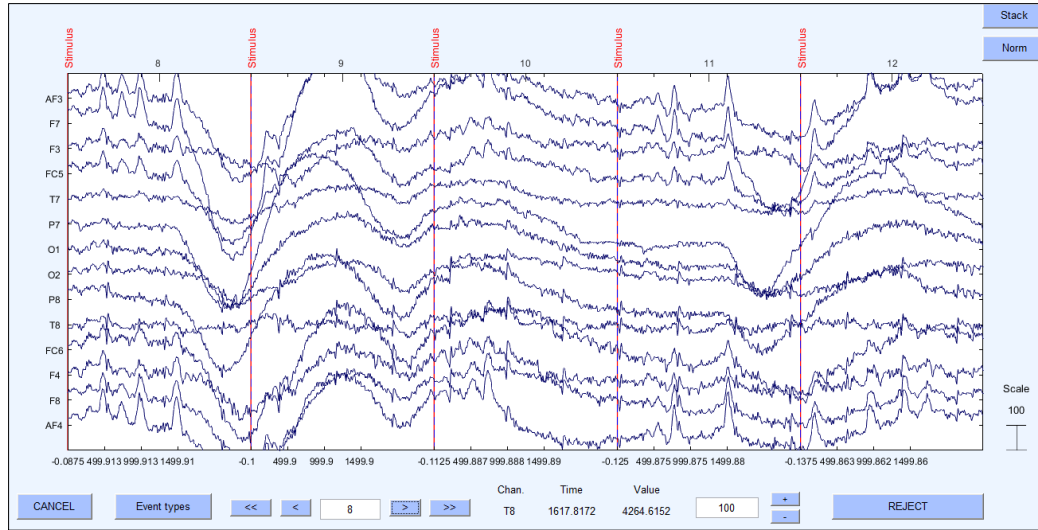
olarak anlamlı bulunan sonuçlar benzer çalışmalarda gerçekleştirildiği şekilde fizyolojik değişkenlerin zihinsel iş yükündeki değişimle birlikte ne yönde değiştiği ve hangi değişkenlerin hedef değişken ile daha fazla ilişkili olduğu hakkında fikir vermektedir. Bu nedenle gerçekleştirilen korelasyon ve hipotez testleri zihinsel iş yükünün analizinde önemli sayılabilecek değişkenlerin tespiti açısından önemli görülmektedir.

5.2.4. EEGLAB STUDY çalışması ile elde edilen bulgular

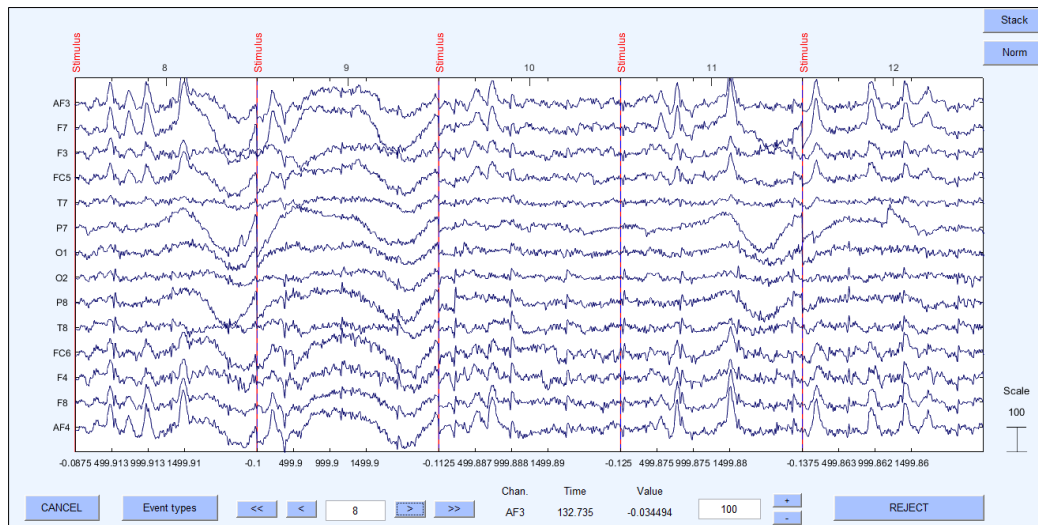
Hem SPSS ile elde edilen istatistiksel bulguların desteklenmesi hem de ham EEG verisinin görselleştirilmiş bir grup analizinin sunulabilmesi amacıyla MATLAB tabanlı EEGLAB programının 2021 sürümünden yararlanılmıştır (URL-20). EMOTIV Pro 3 yazılımı aracılığıyla alınan ham EEG verisi .set dosyasına çevrilmiş, 14 elektrot kanalına ilişkin koordinat verilerini içeren .ced dosyası ile birlikte programa yüklenmiştir. Ham veri üzerindeki gürültüleri ortadan kaldırmak için filtreleme ve veri ön işleme adımları uygulanmıştır. EEGLAB geliştiricileri tarafından önerilen (URL-21) ve literatürdeki çalışmalarda da kullanılan aşağıda açıklanan temel prosedür izlenmiştir (Borys ve diğerleri, 2017a; Borys ve diğerleri, 2017b; Lim ve diğerleri, 2018; Sareen, Singh, Varkey, Achary ve Gupta, 2020):

Gürültü, EEG sinyalinde meydana gelen ve sonuçların güvenilirliğini azaltacak olması sebebiyle istenmeyen bozulmalara denilmektedir. Çalışmada, ön işleme aşamasında, gürültüleri ortadan kaldırmak amacıyla filtreleme işlemi yapılmıştır. EEG sinyallerine kesim frekansı 1 Hz olan bir yüksek geçiren filtre ve bir 50 Hz bant durduran filtre uygulanmıştır. Bu filtreleme işlemi sınırlardaki gürültülerin ortadan kaldırılmasına yardımcı olmaktadır. 30 Hz üzerindeki EEG sinyallerinin genellikle anlamlı ve faydalı frekans bileşenlerine sahip olmadığı düşünülmektedir (Zarjam ve diğerleri, 2015). Durduran filtre olarak 40 Hz (Borghetti ve diğerleri, 2017) ve 50 Hz (Borys ve diğerleri, 2017a; Borys ve diğerleri, 2017b) kullanan çalışmalar mevcuttur. Ana güç hattı frekansı EEG sinyalinde 50 Hz olarak görünebilecek bir gürültüye neden olabilmektedir. Bu hat gürültüsü aynı frekans bandında meydana gelmesi sebebiyle ham EEG sinyalinde karışıklığa neden olmaktadır (Rashid ve Qureshi, 2015). Bu sebeple EEGLAB'ın CleanLine eklentisi kullanılarak bir filtreleme işlemi daha gerçekleştirilmiştir. Son olarak gürültüleri tespit etmek ve ortadan kaldırmak amacıyla veriyi bağımsız bileşenlerine ayırmak için “Bağımsız Bileşenler Analizi – BBA (Independent Component Analysis –

ICA)” uygulanmıştır. Bu yöntem, “runica” işlevi kullanılarak EEGLAB’ta uygulanmıştır. Bağımsız Bileşenler Analizi gibi teknikler dinamik karakteristikleri davranışsal değişimlerle ilişkili olan EEG proseslerini açığa çıkarmaktadır. Eğitim verisi boyunca hem yapay hem de sinirsel olarak üretilen ve EEG katkıları maksimum düzeyde birbirinden bağımsız olan EEG kaynaklarını ayırmak konusunda BBA algoritmaları faydalı olmaktadır. Özellikle göz ve kas hareketleri kaynaklı gürültüleri tespit edip ortadan kaldırmak için bu yöntemle sıklıkla başvurulmaktadır (Delorme ve Makeig, 2004). Filtreleme işlemi öncesi ve sonrasındaki sinyal görüntüsü Şekil 5.5 ve Şekil 5.6’da verilmiştir.



Şekil 5.5. Filtreleme öncesi sinyal görüntüsü



Şekil 5.6. Filtreleme sonrası sinyal görüntüsü

EEGLAB-STUDY, birden fazla denekten, oturumdan ve/veya deneysel bir çalışmanın farklı koşullarından toplanan verileri yönetmek ve işlemek için kullanılmaktadır (URL-21). Bu çalışmada da, 15 kişiden 4 farklı görev koşulu (0-geri, 1-geri, 2-geri, 3-geri) altında 12'şer oturumda (her bir koşul için 3 oturum) toplanan EEG sinyalinin grup analizi amaçlanmıştır. Bu sayede farklı koşullar altında EEG sinyallerinin farklı frekanslardaki gücü açısından istatistiksel olarak anlamlı ölçüde farklılık gösteren beyin bölümlerine ilişkin bulgular elde edilmesi hedeflenmektedir.

Bu doğrultuda öncelikle daha önce filtrelenen 180 adet .set uzantılı dosya STUDY oluşturma ekranına yüklenmiştir. Söz konusu ekranın görüntüsü Şekil 5.7'de verilmiştir. Denek, oturum ve görev zorluk seviyesi bilgileri de sırasıyla subject, session ve condition alanlarına girilmiştir.

Create a new STUDY set -- pop_study()

Edit STUDY set information - remember to save changes

STUDY set name: STUDY15SUB

STUDY set task name: N-BACK

STUDY set notes:

	dataset filename	browse	subject	session	run	condition	group	Select by r.v.	
1	C:\Program Files\Polyspace\R2	...	1	1		2		All comp.	Clear
2	C:\Program Files\Polyspace\R2	...	1	2		1		All comp.	Clear
3	C:\Program Files\Polyspace\R2	...	1	3		0		All comp.	Clear
4	C:\Program Files\Polyspace\R2	...	1	4		3		All comp.	Clear
5	C:\Program Files\Polyspace\R2	...	1	5		3		All comp.	Clear
6	C:\Program Files\Polyspace\R2	...	1	6		0		All comp.	Clear
7	C:\Program Files\Polyspace\R2	...	1	7		2		All comp.	Clear
8	C:\Program Files\Polyspace\R2	...	1	8		3		All comp.	Clear
9	C:\Program Files\Polyspace\R2	...	1	9		1		All comp.	Clear
10	C:\Program Files\Polyspace\R2	...	1	10		2		All comp.	Clear

Important note: Removed datasets will not be saved before being deleted from EEGLAB memory

< Page 1 >

☒ Update dataset info - datasets stored on disk will be overwritten (unset = Keep study info separate).

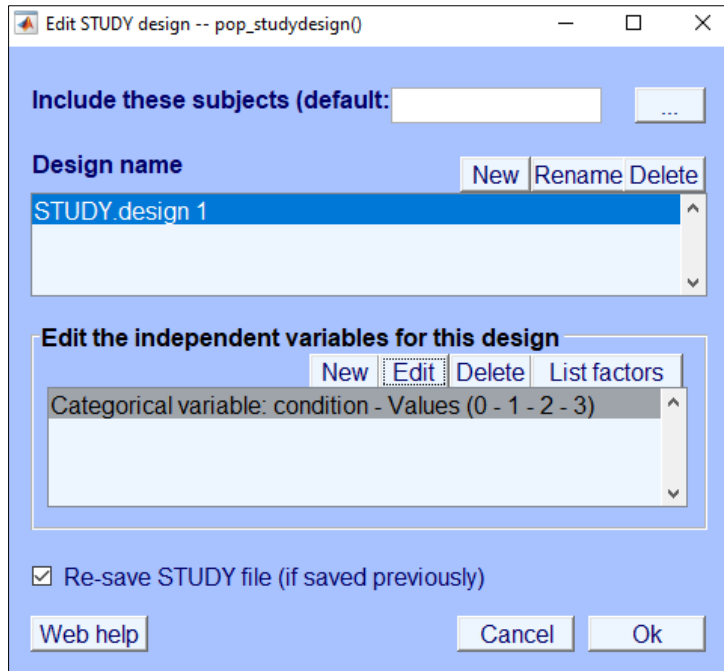
☐ Delete cluster information (to allow loading new datasets, set new components for clustering, etc.)

Help Cancel Ok

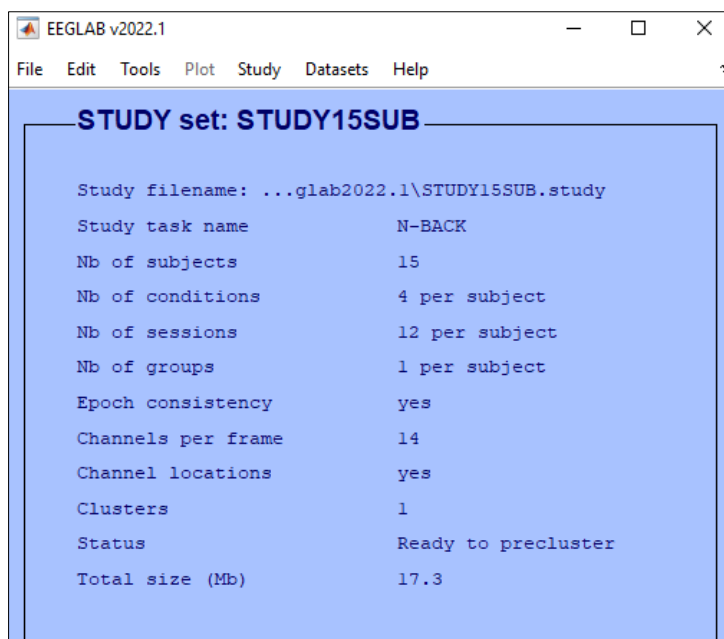
Şekil 5.7. STUDY oluşturma ekranı

STUDY çalışmasında kullanılacak bağımsız değişkenin görev zorluk seviyesi yani “condition” olacağı bilgisi de Şekil 5.8’de verilen “STUDY design” ekranında belirlenmiştir. Daha önce oluşturulan ve var olan bir STUDY dosyası yüklendikten sonra söz konusu STUDY veri setinin ana ekranda görülen bilgileri ise Şekil 5.9’da verilmiştir.

STUDY sonuçlarının analizine başlanmadan önce, bağımsız bileşenler analizi sonucunda elde edilen bileşenlerin kaynağı EEGLAB ICLabel fonksiyonu aracılığıyla sınıflandırılmıştır. %90'ın üzerinde olasılıkla göz, kas veya beyin dışı başka bir kaynaktan oluştuğu tahmin edilen bileşenler çıkarılmıştır. Olasılık değerleri Şekil 5.10'da gösterilen ekrana girilmektedir.



Şekil 5.8. STUDY tasarımı belirleme ekranı



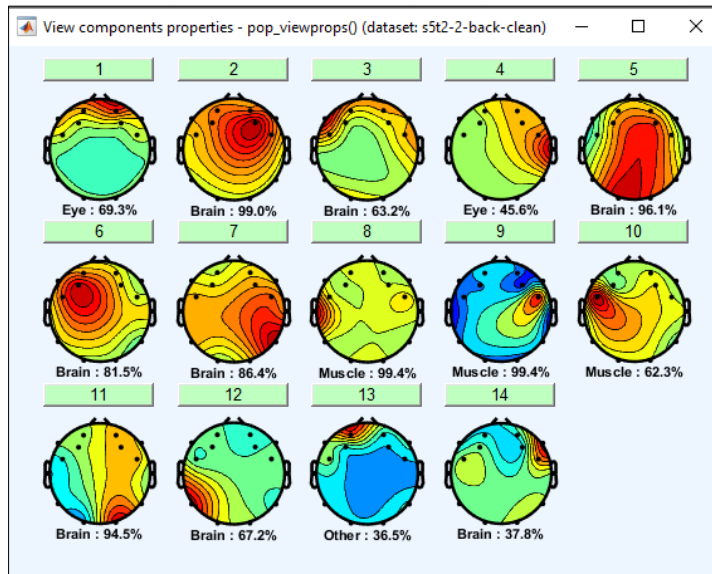
Şekil 5.9. STUDY giriş ekranı

Gereksiz veri kaybının önüne geçmek için program geliştiriciler tarafından bileşen çıkarma işleminin filtreleme ile temizlenmiş .set dosyaları oluşturulurken değil tüm dosyaların yüklendiği grup analizi aşamasında STUDY dosyası üzerinde yapılması önerilmektedir (URL-21).

	Min	Max
Probability range for "Brain"		
Probability range for "Muscle"	0.9	1
Probability range for "Eye"	0.9	1
Probability range for "Heart"		
Probability range for "Line Noise"		
Probability range for "Channel Noise"		
Probability range for "Other"	0.9	1

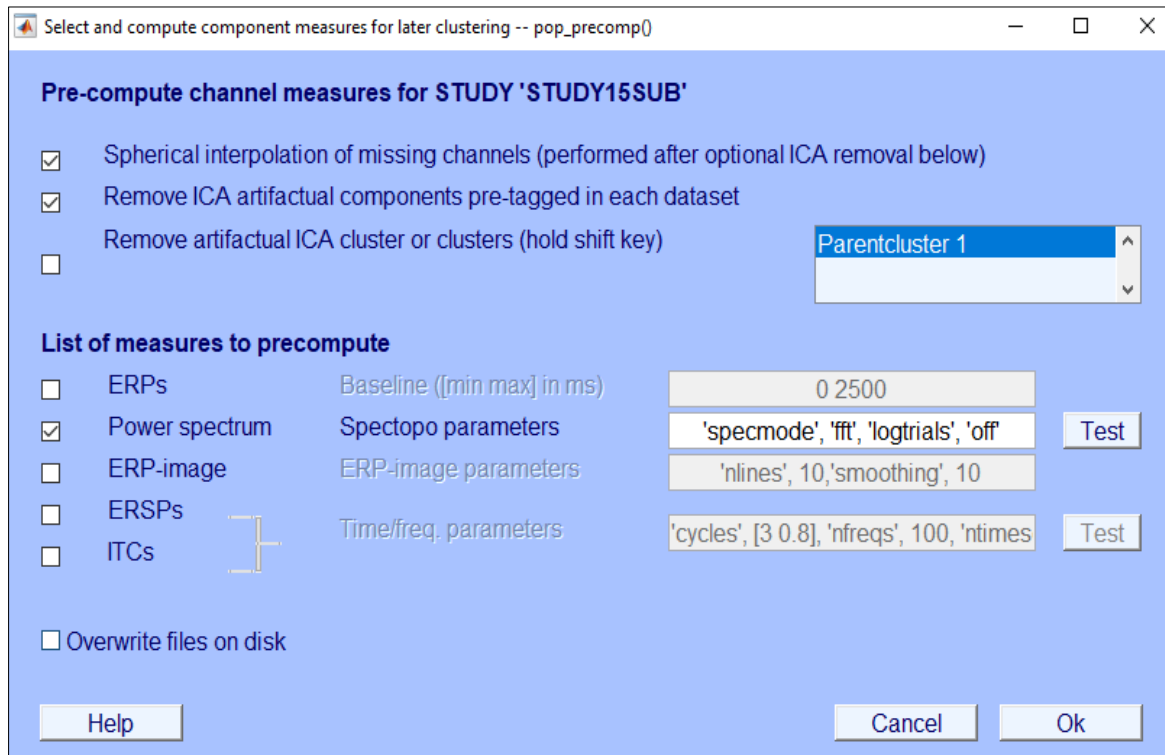
Şekil 5.10. Bileşen çıkarma kriterlerini belirleme ekranı

Örneğin Şekil 5.11’de %90’ın üzerinde olasılıkla kas hareketleri kaynaklı olduğu tespit edilen 2 bileşen analizden çıkarılmıştır.



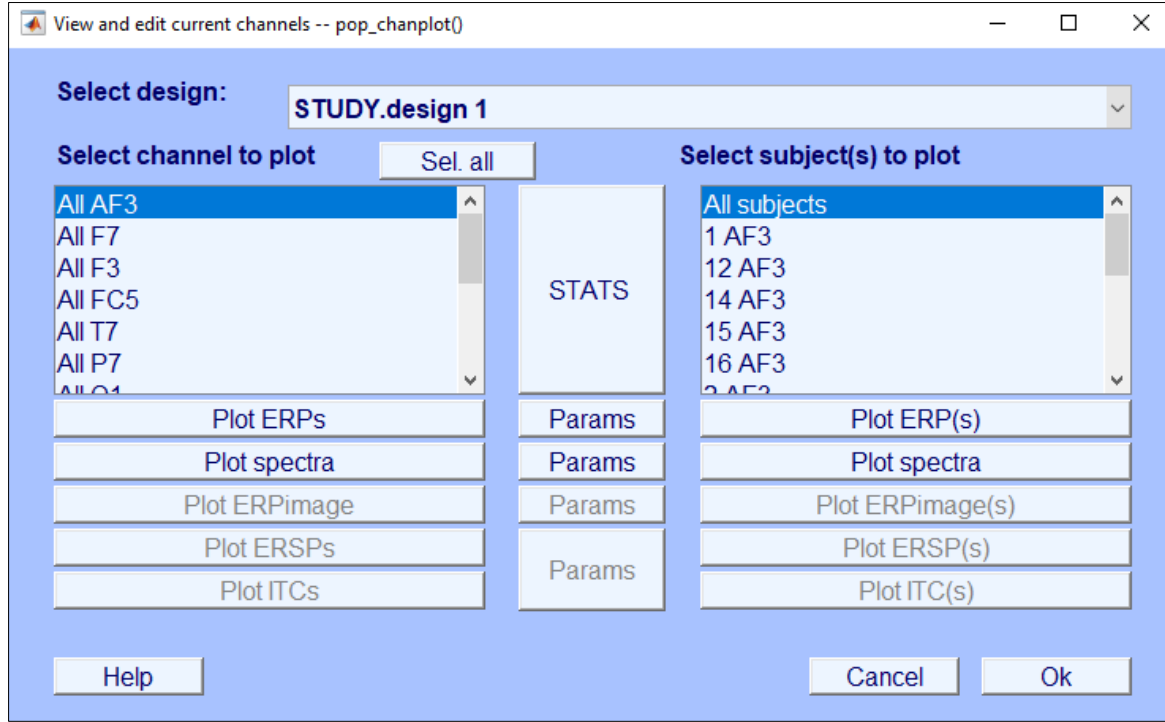
Şekil 5.11. Bileşen çıkarma öncesi etiketleme ekranı

Zihinsel iş yükü analizinde kullanılan EEG ölçütlerinin çoğu spektral karakteristiklere dayanmaktadır (Borys ve diğerleri, 2017b). EEG spektrumu (tayfı) tüm EEG dalga boylarını içeren bir sınıflandırma biçimidir. Belirlenen farklı frekans aralıklarının genliklerine bakılarak güç spektrumları oluşturulmakta ve analiz edilmektedir. Bu çalışmada da analiz güç spektrumları üzerinden yapılacağı için ön hesaplama ekranında “güç spektrumu” seçilmektedir. Ön hesaplama ekranı Şekil 5.12’de görülmektedir:



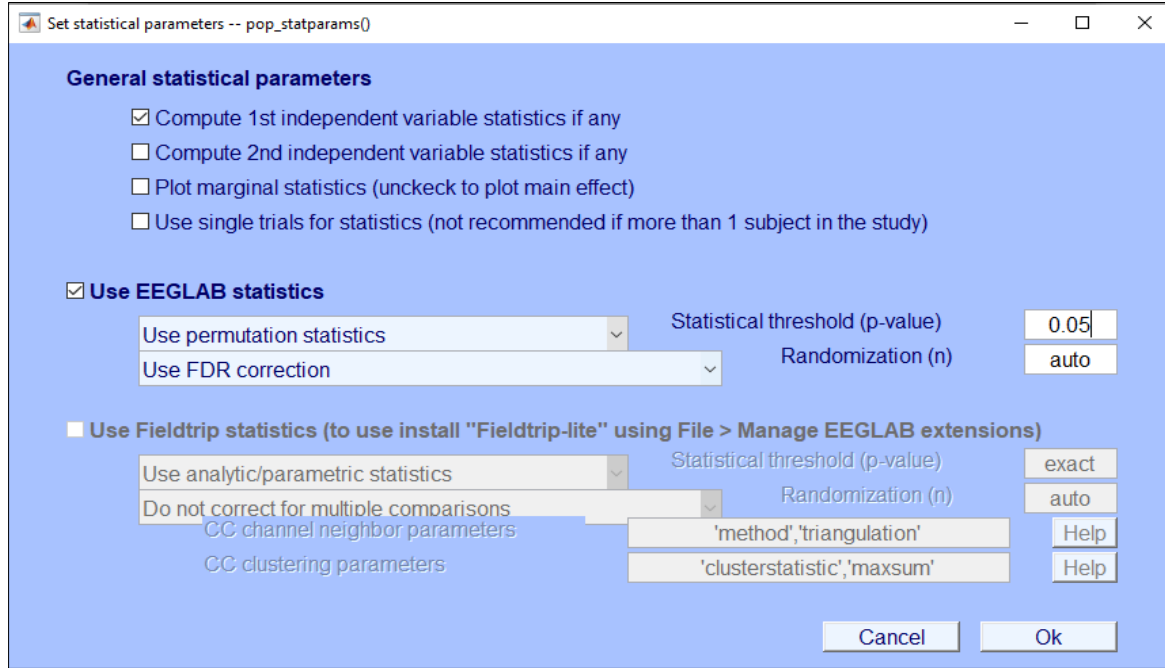
Şekil 5.12. Ön hesaplama ekranı

Hesaplama işleminden sonra spektrumların çizdirilmesi ve beyin topografilerinin (haritalarının) görüntülenebilmesi Şekil 5.13’te verilen veri görselleştirme ekranı aracılığıyla mümkün olabilmektedir. Kullanılacak istatistiksel parametreler belirlenerek kanal ve denek bazında görsel grafikler oluşturulabildiği gibi tüm denekleri ve/veya kanalları içerecek analizler de yapılabilmektedir.



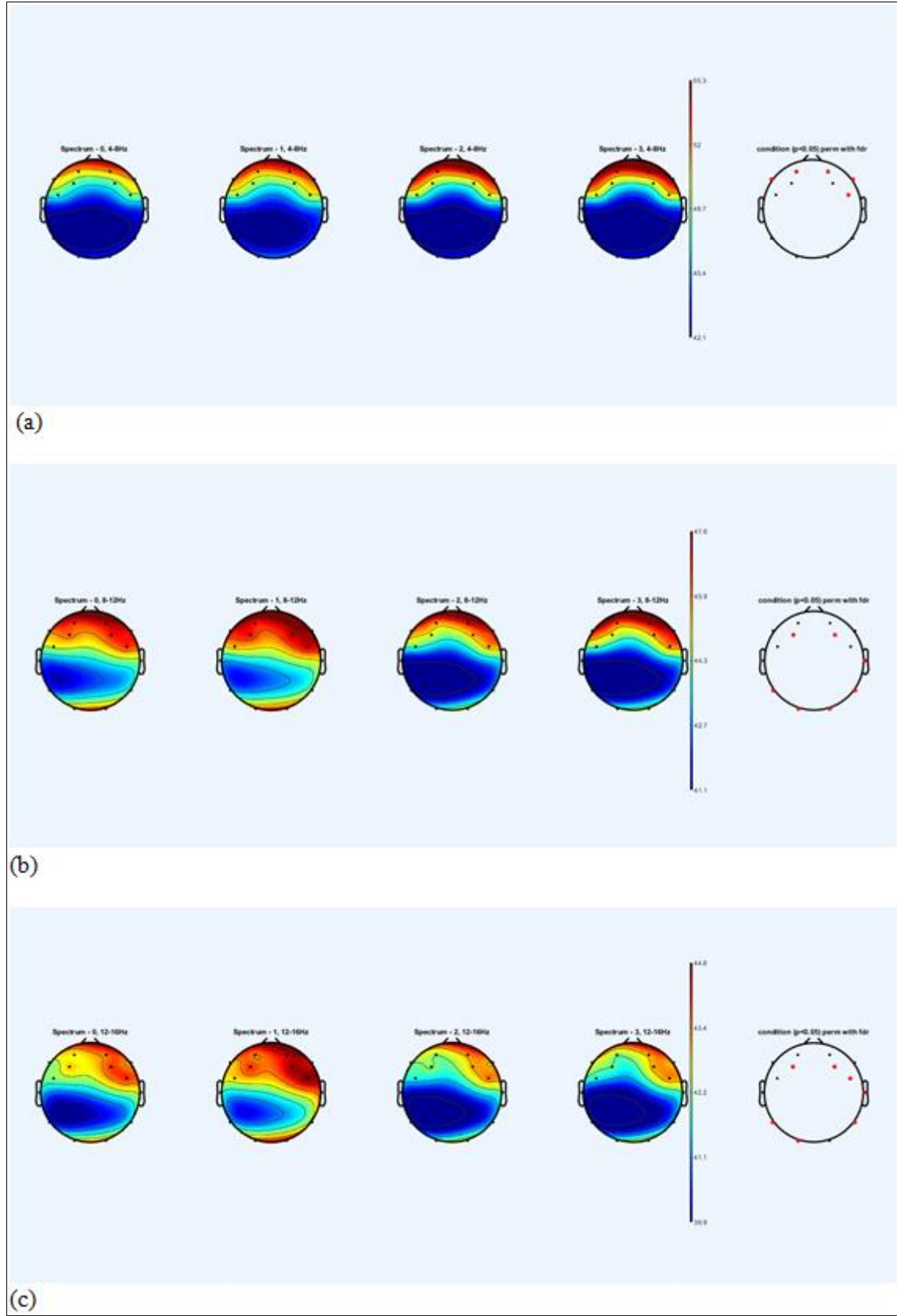
Şekil 5.13. Veri görselleştirme ekranı

Şekil 5.14’te verilen istatistiksel parametreleri belirleme ekranında 1. bağımsız değişken istatistiklerini hesaplama onay kutusuna tıklamamızın sebebi çalışma tasarımında seçilen ikinci bir bağımsız değişken olmamasıdır. Tek bağımsız değişken görev zorluk seviyesi olarak belirlenmiştir. İstatistik türü için parametrik istatistikler yerine açılan listede permütasyon istatistikleri seçilmiştir. Verileri keşfetmek için parametrik istatistikler yeterli olsa da, nihai sonuçları çizmek için permütasyona dayalı istatistikleri kullanmanın daha iyi olacağı program geliştiriciler tarafından önerilmektedir (URL-21). Birden çok karşılaştırmayı düzeltmek için, açılan listeden FDR yöntemi seçilmiş ve p eşik değeri 0,05 olarak girilmiştir. Sonuçlar 0,05 anlamlılık düzeyinde analiz edilecektir. Yanlış Bulgu Oranı (False Discovery Rate, FDR) olarak isimlendirilen bu yaklaşım, yanlışlıkla reddedilen hipotezlerin beklenen oranı olarak ifade edilmektedir. Çoklu test işlemlerinde klasik yöntemlerin gerek deneysel ortak hata oranı ile ilgili sınırlılıklarından dolayı yokluk hipotezlerinin reddedilmesi olasılığını algılamakta güçsüz olmaları, gerekse sabit bir p değerinin dikkate alınmasından kaynaklanan zafiyeti gidermesi açısından FDR yaklaşımı önem arz etmektedir (N. Doğan, İ. Doğan ve Cangür, 2014).



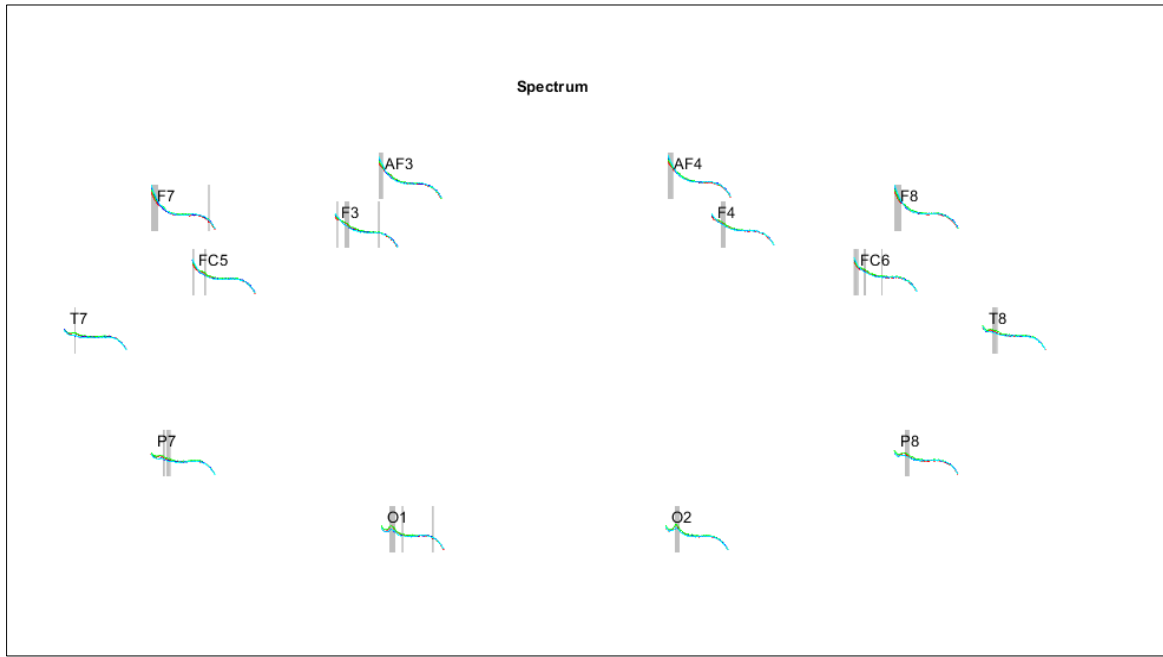
Şekil 5.14. İstatistiksel parametreleri belirleme ekranı

Analiz aşamasında, öncelikle 5 farklı frekans bandında tüm deneklerdeki ortalama gücü gösteren beyin topografileri (haritaları) incelenmiştir. Şekil 5.15'te sadece anlamlı sonuç elde edilen teta (4-8 Hz), alfa (8-12 Hz) ve düşük beta (12-16 Hz) frekanslarına ilişkin topografiler yer almaktadır. Yüksek beta ve gama bantları ise zorluk seviyesine bağlı güç değişiminin anlamlı bulunmamasından dolayı verilmemiştir. Haritalarda yüksek güce sahip bölgeler kırmızı renkle, düşük güce sahip bölgeler ise mavi renkle gösterilmekte, kırmızıdan maviye renk değişimi, gücün azaldığı anlamına gelmektedir. Soldan sağa sırasıyla dört zorluk seviyesi verilmiştir (en soldaki 0-geri, en sağdaki 3-geri). Şekil 5.15a'daki teta frekansı topografyası incelendiğinde, zor görevlere doğru gidildikçe prefrontal ve frontal bölgelerin koyu kırmızıya döndüğü görülmektedir. Alfa frekansı için ise zor görevlere doğru topografyada maviye dönen alanlar, özellikle temporal, parietal ve oksipital bölgelerde alfa frekansındaki gücün azaldığını göstermektedir. Düşük beta frekansına karşılık gelen Şekil 5.15c'deki topografyada ise, görev zorluk seviyesi arttıkça harita genelinde kırmızı rengin açılması ve mavi rengin yoğunluğunun artması dikkati çekmektedir. Bu değişiklik, görev zorlaştıkça beynin farklı bölgelerinden alınan EEG sinyallerinde düşük beta gücünde azalma olduğunu doğrulamaktadır. En sağda kırmızı nokta ile gösterilen elektrotlar, bağlı olduğu beyin bölgesinde, belirtilen banttaki görev koşulları arasında meydana gelen güç değişiminin %95 güvenilirlik düzeyinde anlamlı olduğunu göstermektedir.



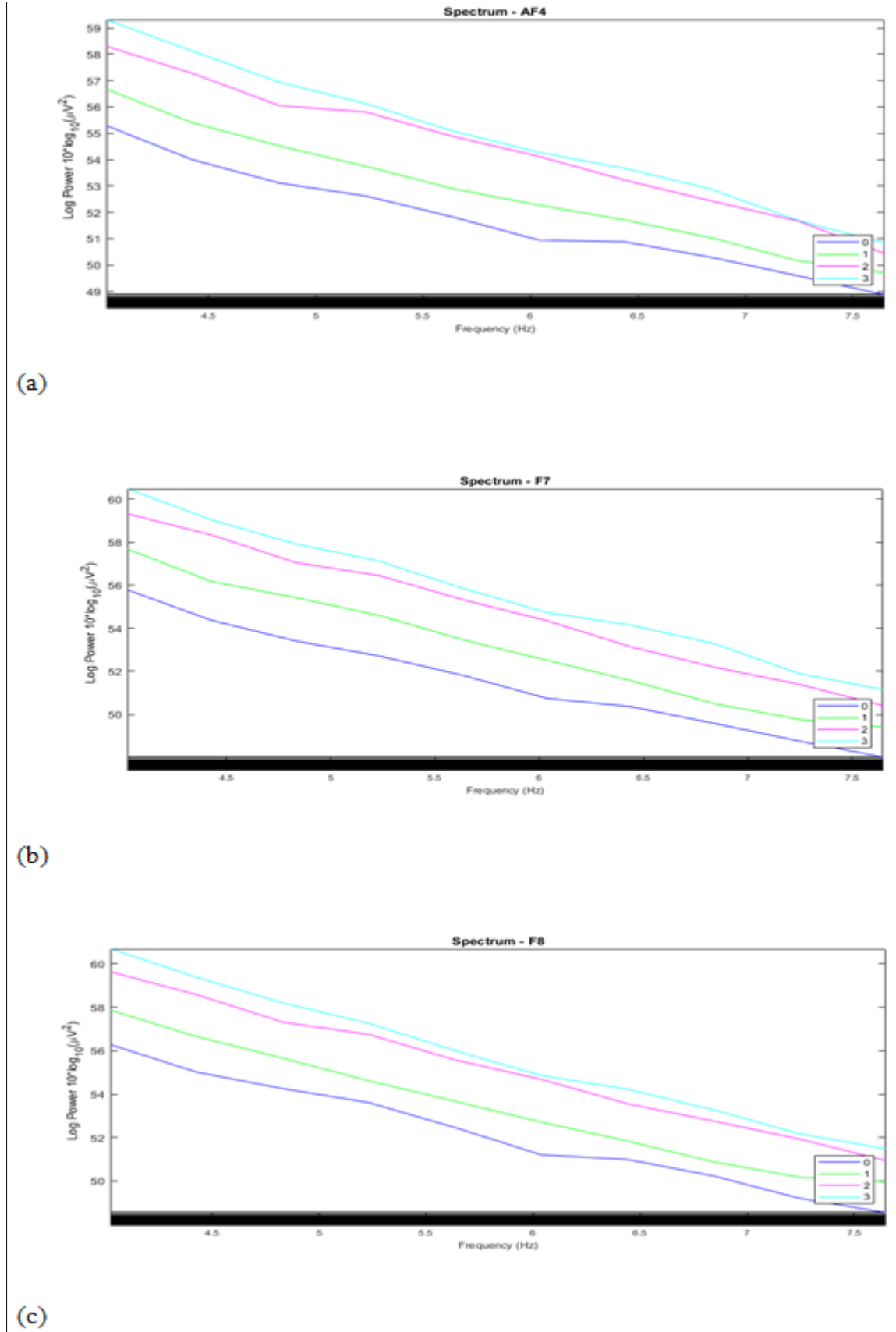
Şekil 5.15. Beyin haritaları a) teta bandı için b) alfa bandı için c) düşük beta bandı için

Görev koşullarına göre bant gücünde anlamlı farklılık gösteren frekans aralıkları kanal bazında Şekil 5.16'da verilmiştir. Grafikler 4-45 Hz frekans aralığındaki güç spektrumlarını göstermektedir. Bu şekilde de prefrontal, frontal ve merkez bölgeleri gösteren AF3, AF4, F7, F8, FC5 ve FC6 kanallarında özellikle düşük frekanslı teta bandında; temporal, pariyetal ve oksipital bölgeleri gösteren O1, O2, P7, P8, T7, T8 kanallarında ise özellikle alfa bandında anlamlı güç değişimi izlendiği görülebilmektedir.

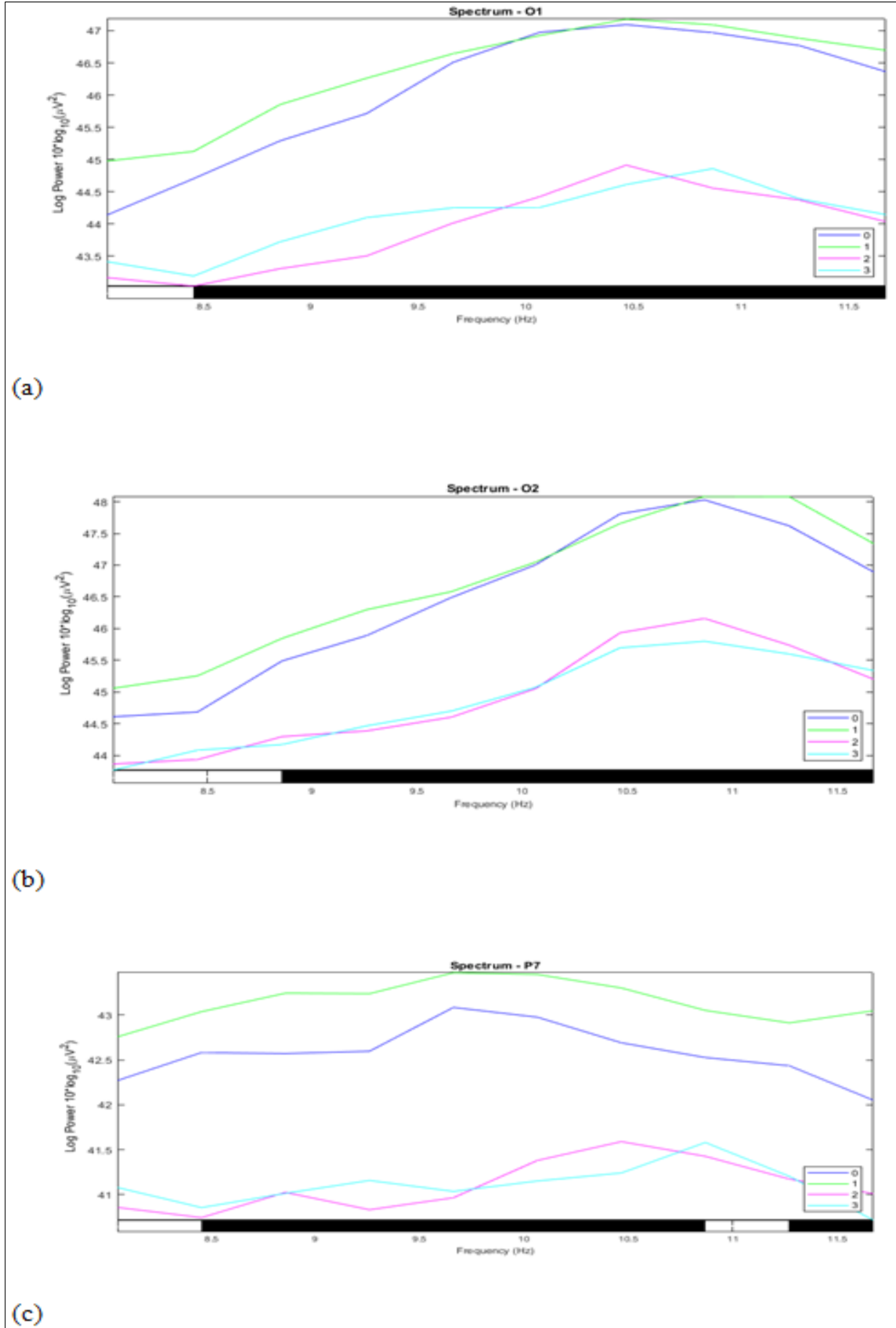


Şekil 5.16. Zorluk seviyesine göre bant gücünde anlamlı farklılık gösteren frekanslar

Şekil 5.17'de özellikle AF4, F7 ve F8 kanallarındaki teta gücünün frekans aralığı (4-8 Hz) boyunca görev zorlaştıkça nasıl arttığı ve zorluk seviyelerinin bu bölgelerdeki teta gücü özelliği sayesinde nasıl ayrıştırılabileceği görülmektedir. 0-geri görev koşulu lacivert, 1-geri görev koşulu yeşil, 2-geri görev koşulu pembe, 3-geri görev koşulu ise açık mavi çizgi ile gösterilmiştir. Benzer şekilde, Şekil 5.18'deki spektrumlardan; kolay olan 0-geri ve 1-geri görevler gerçekleştirilirken O1, O2 ve P7 kanallarındaki alfa (8-12 Hz) gücünün daha zor olan 2-geri ve 3-geri görevler gerçekleştirilirken ölçülenden anlamlı ölçüde farklı olduğu görülebilmektedir. Beyin topografyalarında belirtilen bölgelerde görev zorlaştıkça azaldığını yorumladığımız alfa gücünün bazı frekanslar için 1-geri görevde 0-geri görevdekenden, 3-geri görevde de 2-geri görevdekenden daha yüksek ölçülmesi, beklenenden farklı bir bulgu olarak dikkati çekmiştir. Yine de bu EEG özellikleri sayesinde zor görevleri kolay olanlardan ayrıştırılabileceğimizi söylemek mümkündür.



Şekil 5.17. Farklı zorluk seviyeleri için teta gücü spektrumu a) AF4 b) F7 c) F8



Şekil 5.18. Farklı zorluk seviyeleri için alfa gücü spektrumu a) O1 b) O2 c) P7

Neticede, EEG yazılımı tarafından sağlanan bant gücü verisi ile elde edilen bulgular, ham EEG verisinin işlenerek görsel olarak analiz edilmesi suretiyle desteklenmiştir. MATLAB tabanlı EEGLAB STUDY aracı, farklı görev koşulları altında EEG sinyallerindeki değişimi grafiksel açıdan inceleme fırsatı vermiştir.

5.2.5. Makine öğrenmesi algoritmaları ile elde edilen tahmin modellerine ilişkin bulgular

Tez çalışmasında, EEG ve göz izleme değişkenlerine ilişkin değerleri bilinen katılımcıların zihinsel iş yüküne karşılık geldiği varsayılan görev zorluk seviyelerinin (4-sınıf) kategorik olarak tahmin edilebilmesi sınıflandırma problemi, yine aynı fizyolojik değişkenler kullanılarak ağırlıklı toplam NASA-TLX skorunun tahmin edilebilmesi ise regresyon problemi olarak ele alınmıştır. Sınıflandırma ve regresyon problemlerinin çözümünde doğrusal olmayan modellerin çözümünde kullanılan makine öğrenmesi algoritmalarından yararlanılmıştır. Çoklu doğrusal regresyonun daha önce bahsedilen pek çok varsayımı gerektirmesi, fazla sayıda değişken bulunması sebebiyle ilişkinin doğrusallığını sağlamanın zor olması, grafiksel olarak incelendiğinde bağımsız değişkenlerin ele alınan bağımlı değişkenler ile ilişkisinin bazı değişkenler için doğrusal olmayan bir eğilim gösterdiğinin tespit edilmesi ve elde edilen korelasyon katsayılarının da güçlü doğrusal bir ilişkiye işaret etmemesi sebebiyle doğrusal olmayan algoritmalar tercih edilmiştir. Modeller Python kullanılarak sırasıyla KEYK, DVM, YSA, RO, GAM, AGAM ve HGAM algoritmaları ile çözdürülmüştür. Kategori artırma algoritması ise bağımsız değişkenler kategorik olduğunda iyi performans gösterdiği için analizin dışında tutulmuştur. 2700 gözlem ve 111 bağımsız değişkenden oluşan veri setinin eğitim ve test için farklı oranlarda ayrıldığı modellerin sonuçları karşılaştırılmış, çalıştırılan tüm modellerde 5 katlı çapraz doğrulama ile modelin genellenebilirliği artırılmaya çalışılmıştır (Lim ve diğerleri, 2018; Plechawska-Wojcik ve diğerleri, 2019; Pei ve diğerleri, 2021). Hiperparametre optimizasyonu için çapraz doğrulamalı ızgara arama (GridSearch) yapılarak çeşitli hiperparametre kombinasyonları arasından en iyi çözümü veren hiperparametre seti araştırılmıştır. Izgara arama yönteminde modelde denenmesi istenen hiperparametreler ve kullanıcı tarafından girilen alternatif değerleri için bütün kombinasyonlar ile ayrı ayrı model kurulmakta ve söz konusu yöntem sayesinde en başarılı hiperparametre kombinasyonu belirlenmektedir. Bu yöntem; kombinasyon sayısı arttığında ve çapraz doğrulama devreye girdiğinde zaman açısından maliyetli olmasına karşın, en iyi

performans gösteren hiperparametre setini belirlemeyi garanti etmesi sebebiyle tercih edilmiştir (URL-22). Çapraz doğrulama yapılan versiyonu GridSearchCV şeklinde belirtilmektedir. Izgara arama yöntemi çapraz doğrulama ile birlikte, belirtilen hiperparametre kombinasyonları arasında en iyi kombinasyonu bulmak için tüm olası kombinasyonları denemekte ve böylece, modeli daha iyi bir şekilde ayarlayarak daha iyi performans elde etmeye yardımcı olmaktadır.

Tüm değişkenlerin kullanıldığı modeller öncelikle tüm algoritmalar ile çözdürülmüş, beş katlı çapraz doğrulama sonucu elde edilen ortalama performans değerleri karşılaştırılmıştır. Analizin sonraki adımlarına en iyi sonucu veren algoritma ile devam edilmiştir. En iyi sonucu veren algoritma kullanılarak öncelikle hem yalnızca EEG değişkenleri hem de yalnızca göz izleme değişkenleri kullanıldığında modellerin nasıl sonuç verdiği araştırılmış, her iki yöntemin birlikte kullanıldığı çözüm ile kıyaslanmıştır. Her iki yöntemin birlikte kullanıldığı 111 değişkenli modelin çözümünde iyileştirmeye gidilebilmesi amacıyla ÇDÖÖE yöntemiyle özellik seçimi yapılarak boyut azaltılmıştır. Boyut azaltılmış modelde daha önce en iyi sonucu veren algoritma için ızgara arama yöntemi kullanılarak en uygun hiperparametre seti araştırılmıştır. Model başarısını artırmak üzere gerçekleştirilen bu hiperparametre ayarlama sürecinin ardından ayarlanmış model çapraz doğrulama ile tekrar test edilmiş ve sonuçlar yorumlanmıştır. Model performans değerlendirme ölçütlerinden; doğruluk (accuracy), duyarlılık (sensitivity), kesinlik (precision), F1, kappa skoru, MKK ve EAA incelenmiştir.

Sınıflandırma probleminin çözümüne ilişkin bulgular

EEG ve göz izleme değişkenlerinin girdi, görev zorluk seviyesi değişkeninin çıktı olarak ele alındığı tahmin modelleri KEYK, DVM, YSA, RO, GAM, AGAM ve HGAM algoritmaları ile çalıştırılmıştır. Sınıflandırma modelinin çıktı değişkeni olarak; 0-geri göreve ilişkin gözlemler 0, 1-geri göreve ilişkin gözlemler 1, 2-geri göreve ilişkin gözlemler 2, 3-geri göreve ilişkin gözlemler ise 3 kodu ile kategorize edilmiştir.

Bazı algoritmalar uygulanmadan önce farklı ölçü birimlerine sahip değişkenlerin standart hale getirilmesi gerekmektedir. Standartlaştırma işlemi değişken sütunlarının ortalama değeri 0 ve standart sapması 1 olacak şekilde standart normal dağılım oluşturmaktır. Özellikle KEYK, DVM ve YSA algoritmaları standartlaştırma veri ön işleme işlemi

gerçekleştirildiğinde daha iyi sonuç vermektedir (URL-7, URL-8). Bu doğrultuda KEYK, DVM ve YSA algoritmaları ile çalıştırılan modellerde “StandardScaler” fonksiyonu kullanılarak veri dönüştürme işlemi gerçekleştirilmiştir.

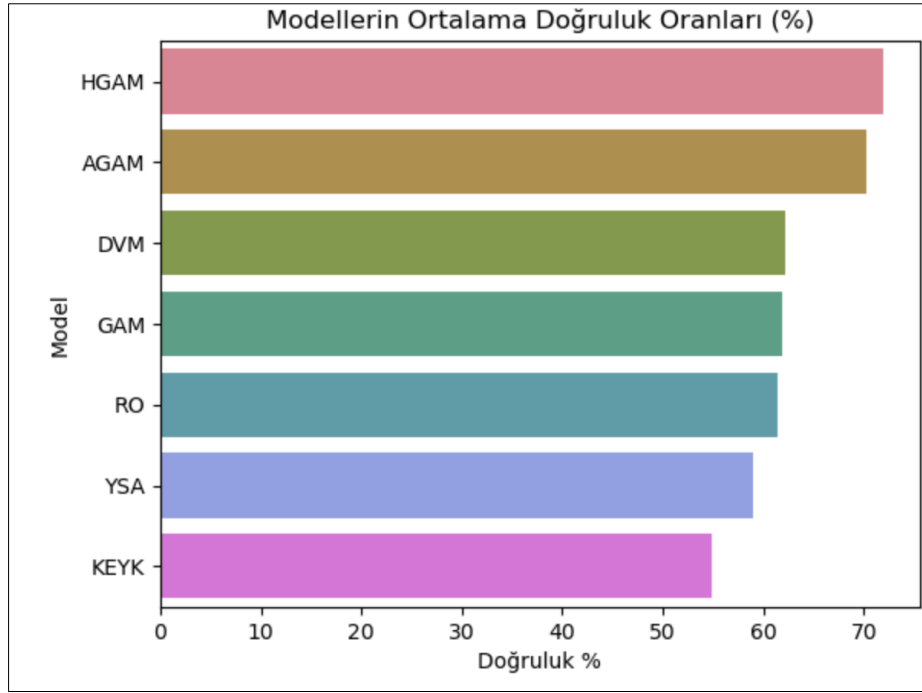
İlk etapta, kullanılan algoritmalarda herhangi bir hiperparametre ayarlama işlemi gerçekleştirilmemiş; ancak KEYK, DVM ve YSA için kernel, solver, aktivasyon gibi temel hiperparametrelerin farklı değerleri birkaç deneme ile test edilerek en uygun değerlere karar verilmiştir. KEYK algoritmasında k değeri 9; DVM algoritmasında kernel radyal, YSA algoritmasında, solver hiperparametresi “adam”, aktivasyon hiperparametresi ise sınıflandırma problemi olması sebebiyle “logistic” seçilmiştir. Ayrıca, YSA modeli çalıştırılırken, iterasyon sayısının iyi sonuca yakınsamada yetersiz kaldığına ilişkin uyarı sebebiyle maksimum iterasyon sayısı 1000 olarak güncellenmiştir.

Çizelge 5.19’da farklı algoritmalarla kurulan 111 değişkenli 4 sınıflı sınıflandırma modellerinin performans değerleri verilmiştir. Verilen değerler 5 katlı çapraz doğrulama ile elde edilen beş ayrı performans sonucunun ortalama değerleridir. Dengeli, yani sınıflar bazında gözlem sayılarının eşit olduğu bir sınıflandırma olması sebebiyle çoklu sınıflandırma modellerinin performans değerlendirmesinde kullanılan mikro yaklaşımdan hareketle temel olarak doğruluk ölçüsü baz alınmış ve modellerin doğruluk performansları birbiriyle kıyaslanmıştır. Sınıflandırıcılar karşılaştırılırken ayrıca çok sınıflı modellerin başarı ölçüsü olarak kullanılan kappa skoru, MKK ve EAA değerleri de incelenmiştir.

Çizelge 5.19. 111 değişkenli dört sınıflı modellerin performans sonuçları

Model	Ortalama Doğruluk (%)	Ortalama Kappa Skoru	Ortalama MKK Skoru	Ortalama EAA Skoru
KEYK	54,85	0,40	0,40	0,80
DVM	62,15	0,49	0,50	0,85
YSA	59,00	0,45	0,45	0,82
RO	61,44	0,49	0,49	0,85
GAM	61,93	0,49	0,49	0,85
AGAM	70,22	0,60	0,60	0,91
HGAM	71,96	0,63	0,63	0,92

Beş katlı çapraz doğrulama ile kurulan modellerin, kullanılan makine öğrenmesi algoritması bazında ortalama doğruluk performanslarına ilişkin karşılaştırma Şekil 5.19’da görülmektedir. Grafik incelendiğinde en yüksek doğruluk performansını gösteren modelin HGAM algoritması ile elde edildiği (%71,96), diğer en iyi sonucu veren algoritmaların ise AGAM, DVM ve GAM olduğu görülmektedir.



Şekil 5.19. Modellerin doğruluk performanslarının karşılaştırılması

En iyi sonucu veren HGAM algoritması kullanılarak yalnızca EEG ya da yalnızca göz değişkenlerinin kullanıldığı modellerin nasıl performans gösterdiği de araştırılmıştır. Çizelge 5.20’de karşılaştırma sonuçları verilmiştir.

Çizelge 5.20. 111 değişkenli dört sınıflı modele ilişkin yöntemsel karşılaştırma

	Doğruluk (%)	Kappa skoru	MKK	EAA
EEG	56,15	0,42	0,42	0,81
GÖZ	65,67	0,54	0,54	0,89
EEG + GÖZ	71,96	0,63	0,63	0,92

Elde edilen performans sonuçları, EEG ve göz değişkenlerinin birlikte kullanıldığı modelin beklendiği şekilde ayrı ayrı kullanıldıkları modellere kıyasla daha iyi sonuç verdiğini ortaya koymuştur. Sadece göz değişkenleri ile çalıştırılan model, sadece EEG değişkenleri ile çalıştırılan modelden daha iyi performans göstermiştir. Bu doğrultuda, hibrit yani her iki yöntemin birlikte kullanıldığı modelde gözle ilgili değişkenlerin modele katkı düzeylerinin daha yüksek olması beklenmektedir.

EEG ve göz değişkenleri ile HGAM kullanılarak kurulan modelde beş katlı çapraz doğrulama ile elde edilen beş ayrı performans sonucundan (72,78; 71,11; 72,04; 72,96; 70,93) birine ilişkin hata matrisi Çizelge 5.21’de görülebilmektedir. Buna göre; gerçekte 0 kodlu sınıfa dâhil olan 132 gözlemden 95’i model tarafından doğru tahmin edilmiştir. Gerçekte 1 sınıfına dâhil olan 132 gözlemden 91’i, 2 sınıfına dâhil olan 137 gözlemden 100’ü, 3 sınıfına dâhil olan 139 gözlemden ise 103’ü doğru tahmin edilmiştir.

Çizelge 5.21. 111 değişkenli dört sınıflı modele ilişkin hata matrisi

		Tahmin Edilen Sınıf			
		0	1	2	3
Gerçek Sınıf	0	95	22	10	5
	1	17	91	13	11
	2	4	12	100	21
	3	7	6	23	103

Daha önce “4.3. Model Performans Metrikleri” bölümünde değinildiği üzere; doğruluk doğru tahmin edilen örneklerin oranını, kesinlik belirli bir sınıfa ilişkin tahminlerin gerçekteki doğruluk oranını, duyarlılık ise belirli bir sınıftaki örneklerin doğru olarak sınıflandırılma oranını vermektedir. 0 olarak tahmin edilen 123 gözlemin 95’i doğru tahmin edilmiştir. %77 olarak elde edilen bu oran kesinlik değerini vermektedir. Gerçekte 0 sınıfına dâhil olan 132 gözlemin 95’inin doğru tahmin edilmiş olması nedeniyle de duyarlılık değeri %72’dir. Doğruluk değeri de doğru tahmin edilen gözlem sayısı olan 389’un test verisindeki toplam gözlem sayısı olan 540’a oranını gösteren %72 olarak elde edilmektedir. Tüm sınıflar için doğruluk, hassasiyet ve kesinlik değerlerini gösteren özet sınıflandırma performansı tablosu Çizelge 5.22’de verilmiştir. Dengeli sınıflandırma

modellerinde tercih edilen makro yaklaşım kullanılarak sınıf bazında belirlenen performans ölçülerinin aritmetik ortalamaları incelenmiştir.

Çizelge 5.22. 111 değişkenli dört sınıflı modele ilişkin sınıflandırma performansı

Sınıf	Kesinlik	Duyarlılık	F1-Skoru	Toplam
0	0,77	0,72	0,75	132
1	0,69	0,69	0,69	132
2	0,68	0,73	0,71	137
3	0,74	0,74	0,74	139
Doğruluk (mikro yaklaşım)	0,72			540
Makro yaklaşım	0,72	0,72	0,72	540

Çapraz doğrulama ile elde edilen beş modelin sınıflandırma performansları incelendiğinde daha çok 1 kodlu sınıfa ait gözlemlerin model tarafından yanlış tahmin edildiği göze çarpmıştır. Çizelge 5.22’de ele alınan modelde yanlış sınıflandırma tahmininin en çok 0-1 ve 2-3 sınıfları arasında gerçekleştiği görülmektedir. Model tarafından 0 ile 1 sınıfları arasında toplamda 39, 2 ile 3 sınıfları arasında ise toplamda 44 hata yapılmıştır. Sınıf sayısının düşürülmesi yoluyla problem karmaşıklığının azaltılmasının sonuçlara nasıl etki edeceğinin araştırılması amacıyla, problem üç sınıflı ve iki sınıflı olarak da ele alınmıştır. İlk olarak duyarlılık değeri en düşük olan 1 kodlu sınıfa ilişkin gözlemler veri setinden çıkarılmış ve yalnızca 0-geri, 2-geri ve 3-geri görevlere ilişkin gözlemler dikkate alınarak zihinsel iş yükünün düşük-orta-yüksek şeklinde sınıflandırılabilmesi hedeflenmiştir. Tüm değişkenleri içeren üç sınıflı modelin doğruluk değeri %80,49 olarak elde edilmiştir. Yalnızca 0-geri ve 3-geri görevlere ilişkin gözlemlerin ele alındığı iki sınıflı modelde ise %89,63 oranında doğruluk performansına ulaşılmıştır. Dört sınıflı, üç sınıflı ve iki sınıflı modellerin sınıflandırma performansları Çizelge 5.23’te verilmiştir.

Çizelge 5.23. Dört, üç ve iki sınıflı modellerin sınıflandırma performansları

	Doğruluk (%)	Kappa skoru	MKK	EAA
4 sınıflı model	71,96	0,63	0,63	0,92
3 sınıflı model	80,49	0,71	0,71	0,93
2 sınıflı model	89,63	0,79	0,79	0,96

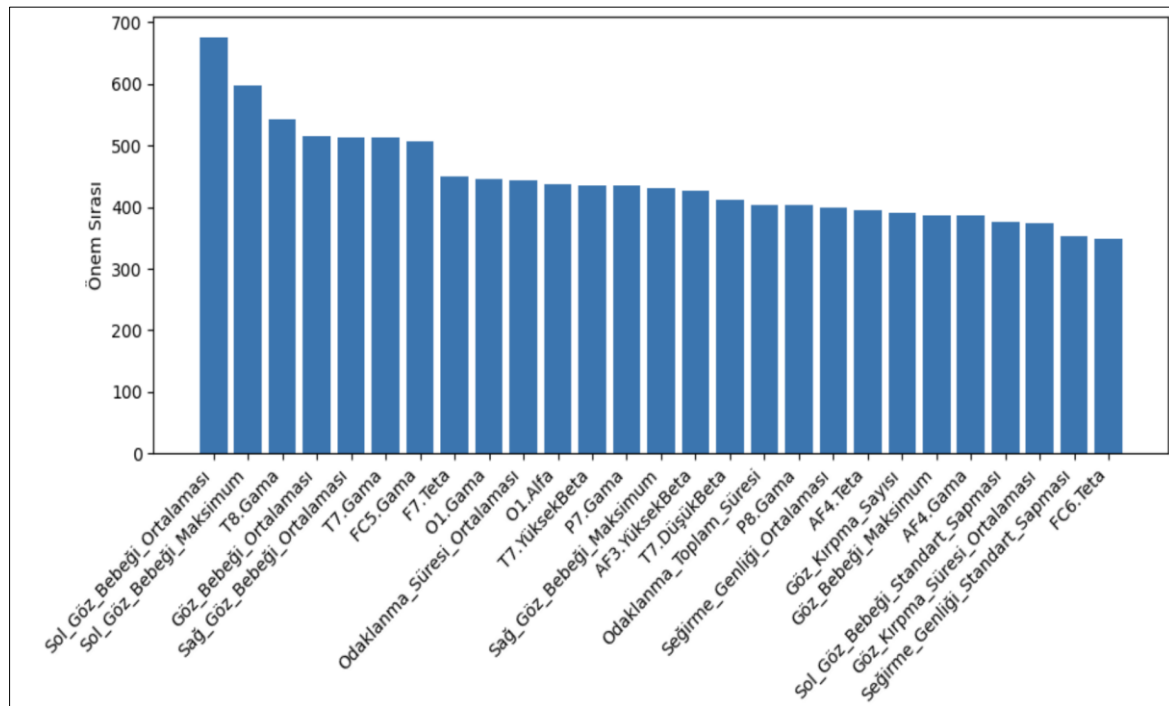
Hem üç sınıflı hem de iki sınıflı modelde güçlü bir sınıflandırıcıya işaret eden 0,70’in üzerindeki kappa ve MKK değerleri dikkati çekmektedir.

Bu aşamada hem sınıflandırma performansını yükseltebilmek hem de modelin anlaşılabilirliğini artırabilmek amacıyla boyut azaltma yapılması gerektiği değerlendirilmiştir. Bu bölüme kadar yapılan analizlerde modeller arası karşılaştırmalar yapılması hedeflenmiş, hiperparametrelerin ayarlanması üzerinde çalışılmamıştır. Veri boyutunun fazla olması sebebiyle bu işlemin hem zaman açısından maliyetli olması hem de özellik seçiminin ayarlanmamış model üzerinden yapılmasının daha anlamlı olması sebebiyle model ayarlama işlemi boyut azaltma işleminden sonraya bırakılmıştır. Ayarlanmamış model üzerinden özellik seçimi ile en etkin değişkenlerin belirlenmesinin ardından boyutu azaltılmış modellerin hiperparametreleri düzeltilerek performans artırılmaya çalışılmıştır. Kullanılan değişkenlerin sayısını azaltmak modelin daha hızlı çalışmasına da katkı sağlamıştır.

Daha önce “4.2.3. Özellik seçimi” bölümünde belirtildiği üzere; veri boyutunun azaltılması için kullanılan yöntemlerin başında özellik seçimi gelmektedir. Özellik seçimi için daha güvenilir olması ve en uygun durma noktasını belirlemesi sebebiyle tezde ÇDÖÖE yöntemi kullanılmıştır. HGAM sınıflandırıcısı kullanılarak beş katlı çapraz doğrulama ile modele en çok katkı sağlayan x sayıda değişkenden oluşan en iyi özellik seti belirlenmeye çalışılmıştır. Optimum durma noktası ÇDÖÖE yöntemi tarafından belirlenmiş, dört sınıflı model için neticede 27 özellik seçilmiştir. Aynı yöntem ile, üç sınıflı modelde 91 değişken, iki sınıflı modelde ise 110 değişken seçilmiştir. Yöntemin en uygun durma noktasını vadetmesi sebebiyle ayrıca bir değişken çıkarma işlemi yapılmamış, yöntem tarafından seçilen değişkenlere sadık kalınmıştır. Dört sınıflı model için seçilen 27 değişken önem sırasından bağımsız bir şekilde Çizelge 5.24’te verilmiştir. Değişkenlerin 14’ünün EEG kaynaklı, 13’ünün ise göz izleme yöntemi ile elde edilen değişkenler olduğu görülmektedir. EEG değişkenlerinden frontal teta ve oksipital alfa gücünün modele en fazla katkı sağlayan değişkenler arasında olması, daha önce yapılan istatistiksel testlerin sonuçları ile tutarlı olması sebebiyle beklenen bir sonuçtur. Hemen hemen tüm beyin bölümlerinde ölçülen gama gücünün dört sınıflı ayrıştırmadaki etkisi ise dikkati çeken bir bulgu olmuştur. Seçilen 27 özelliğin modele sağladığı katkı yönünden önem sırası ise Şekil 5.20’de verilmiştir. En önemli değişkenlerin göz bebeği boyutuna ilişkin değişkenler olduğu dikkati çekmektedir. Odaklanma ve göz kırpma süreleri ile seğirme genliğine ilişkin değişkenler de model tarafından önemli bulunmuştur.

Çizelge 5.24. Dört sınıflı modelde ÇDÖÖE yöntemi ile seçilen özellikler

Değişken Adı	Değişken Kaynağı
AF3.YüksekBeta	EEG
F7.Teta	EEG
FC5.Gama	EEG
T7.DüşükBeta	EEG
T7.YüksekBeta	EEG
T7.Gama	EEG
P7.Gama	EEG
O1.Alfa	EEG
O1.Gama	EEG
P8.Gama	EEG
T8.Gama	EEG
FC6.Teta	EEG
AF4.Teta	EEG
AF4.Gama	EEG
Odaklanma Toplam Süresi	Göz
Odaklanma Süresi Ortalaması	Göz
Seğirme Genliği Ortalaması	Göz
Seğirme Genliği Standart Sapması	Göz
Sol Göz Bebeği Ortalaması	Göz
Sağ Göz Bebeği Ortalaması	Göz
Göz Bebeği Ortalaması	Göz
Sol Göz Bebeği Maksimum	Göz
Sağ Göz Bebeği Maksimum	Göz
Göz Bebeği Maksimum	Göz
Sol Göz Bebeği Standart Sapması	Göz
Göz Kırpma Sayısı	Göz
Göz Kırpma Süresi Ortalaması	Göz



Şekil 5.20. Dört sınıflı model için HGAM tarafından seçilen 27 özelliğin önem sırası

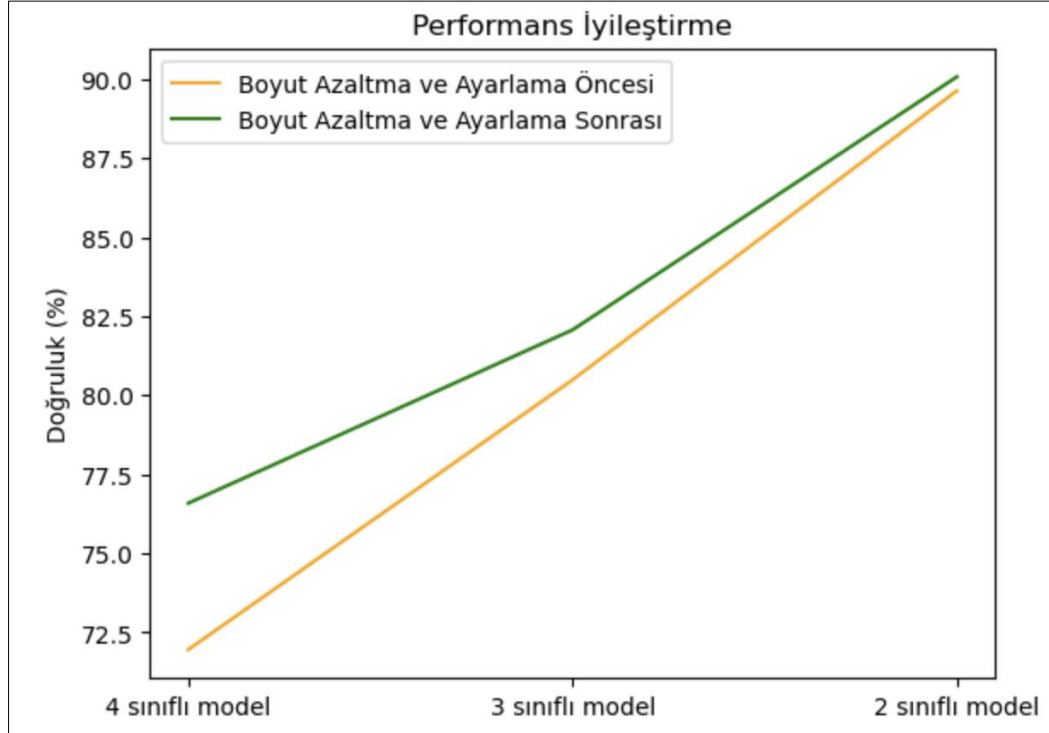
Sadece seçilmiş özelliklerin kullanıldığı modelde çapraz doğrulamalı ızgara arama yöntemi kullanılarak çeşitli hiperparametre kombinasyonları arasından en iyi çözümü veren hiperparametre seti araştırılmıştır. Hiperparametreler için başlangıç değerleri girildikten sonra en iyi hiperparametre seti elde edilmiş, bulunan değerlere yakın alternatif değerler ile tekrar ızgara arama yapılarak performansta iyileşme olup olmadığı kontrol edilmiştir. Yeni hiperparametre seti ile iyileşme gerçekleştiyse belirtilen işlem alternatif değerler kullanılarak tekrarlanmış, gerçekleşmediyse mevcut değerler ile model test edilmiştir. Boyut azaltılmış modellerde, ayarlanmış hiperparametreler ile elde edilen performans sonuçları dört, üç ve iki sınıflı modeller için Çizelge 5.25'te verilmiştir.

Çizelge 5.25. Boyut azaltılmış ve ayarlanmış modellerin performans sonuçları

	Hiperparametreler	Doğruluk (%)	Kappa skoru	MKK	EAA
4 sınıflı model (27 değişken)	öğrenme oranı=0,2 maksimum derinlik=13 ağaç sayısı=750 ağaçtaki yaprak düğümlerinin maksimum sayısı=16	76,59	0,69	0,69	0,93
3 sınıflı model (91 değişken)	öğrenme oranı=0,1 maksimum derinlik=7 ağaç sayısı=750 ağaçtaki yaprak düğümlerinin maksimum sayısı=16	82,07	0,73	0,73	0,94
2 sınıflı model (110 değişken)	öğrenme oranı=0,1 maksimum derinlik=9 ağaç sayısı=300 ağaçtaki yaprak düğümlerinin maksimum sayısı=32	90,07	0,80	0,80	0,97

Boyut azaltma ve hiperparametre ayarlama sonrası sınıflandırma performansındaki değişim Şekil 5.21'de verildiği şekilde olmuştur. Özellikle dört sınıflı modelde daha fazla değişken eliminasyonu gerçekleştirildiği için doğruluk performansındaki değişimin de

daha fazla olduğu gözlenmektedir. Beş katlı çapraz doğrulama ile elde edilen modellerin ortalama doğruluk performansı 27 değişkenli modelde %76,59'a ulaşmıştır.



Şekil 5.21. Boyut azaltma ve hiperparametre ayarlama sonrası performans değişimi

Dört sınıfli modelde %76,59 oranında doğruluk performansına erişilmesine rağmen eğitim setinin doğruluk performansı incelendiğinde %100 olduğu dikkati çekmiştir. Bu durum modelin aşırı öğrenmiş olabileceği riskini ortaya çıkaran ve genellenebilirliği üzerine gölge düşüren bir durumdur. HGAM algoritması her ne kadar öğrenme hızı sebebiyle tercih edilen bir algoritma olsa da aşırı öğrenmeye eğilim göstermektedir. Bu doğrultuda, aşırı öğrenme riskini azaltmak için erken durdurma yöntemine başvurulmuştur.

Erken durdurma (early stopping), makine öğrenmesi modelinin eğitimi sırasında kullanılan bir düzenleme yani modelin genellenebilirliğini artırma tekniğidir. Amacı, modelin aşırı öğrenmesini engellemek veya azaltmaktır. Doğrulama kaybı, bir eğitim iterasyonunun (epoch) sonunda, modelin eğitim verileri üzerindeki performansına bakılmadan, sadece ayrılmış doğrulama veri kümesi üzerindeki performansına dayalı olarak hesaplanan kayıp değeridir. Eğer eğitim kaybı hala düşüyorken, doğrulama kaybı yükselmeye başlarsa, bu aşırı öğrenmenin başladığı bir işarettir. Bu nedenle erken durdurma genellikle doğrulama kaybının artmaya başlaması durumunda eğitimi durdurarak aşırı öğrenmeyi engellemeye

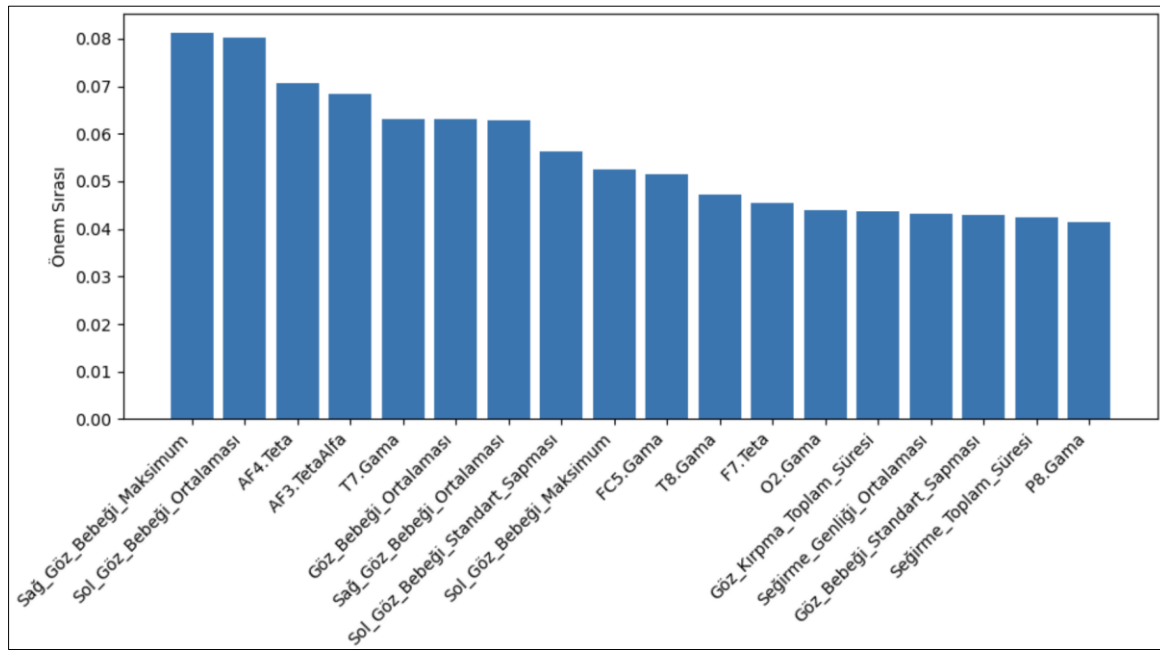
çalışmaktadır. Özetle, erken durdurma, eğitim sürecini, modelin doğrulama verisi üzerindeki performansı düzelmeyi bıraktığında sonlandıran bir stratejidir. Erken durdurma için ayarlanması gereken değişkenler ve parametreler her bir algorithmada farklılık göstermektedir (Prechelt, 2012).

Hem gradyanı artırılmış karar ağaçları algoritmalarının tablo verileri içeren problemlerde günümüzde en gelişmiş teknoloji olmaları (URL-9, URL-23) hem de mevcut problemimizde şu ana kadar en yüksek performans sonuçlarını sağlamaları sebebiyle, erken durdurma yönteminin GAM, AGAM ve HGAM algoritmaları üzerinde uygulanması yoluna gidilmiş, diğer algoritmalara geri dönüş yapılmamıştır. Erken durdurma için ayarlanması gereken hiperparametreler söz konusu algoritmalar için incelenmiştir. Erken durdurma yöntemi öncelikle HGAM algoritmasında denenmiş, ilgili kodda “early_stopping” değişkeni “True” durumuna getirilmiştir. Durdurma kriteri olarak, erken durdurma fonksiyonunun “stopping_rounds” parametresi kullanılmıştır. Eğer söz konusu parametrede belirtilen sayıda ardışık tur boyunca doğrulama kaybı düzelmezse, eğitim erken durdurulmakta ve model kaydedilmektedir. Ancak HGAM algoritması üzerinde uygulanan erken durdurma yöntemi sonucunda da eğitim performansının %100 olarak kaldığı görülmüştür. HGAM ve AGAM algoritmalarının salt erken durdurma yöntemi ile aşırı öğrenme sorunun çözümüne direnç göstermesi sebebiyle GAM algoritması ile erken durdurma yöntemi kullanılmıştır.

“n_iter_no_change” ve “tol” hiperparametreleri GAM modellerinde erken durdurma kriterlerini belirlemek için kullanılmaktadır (URL-24). “n_iter_no_change” parametresi, modelin eğitiminde ardışık olarak kaç iterasyon boyunca doğrulama kaybının veya performansın düzelmediğini belirtmektedir. Eğer belirtilen sayıda iterasyon boyunca herhangi bir düzelme olmazsa, eğitim durdurulmaktadır. HGAM algoritmasındaki “stopping_rounds” hiperparametresi ile benzerdir. “tol” parametresi ise, “n_iter_no_change” süresince doğrulama kaybındaki minimum değişimi belirtmektedir. Eğer son “n_iter_no_change” adet iterasyon boyunca doğrulama kaybı bu değerden daha az bir düşüş gösterirse, bu düşüş ihmal edilmekte ve eğitim durdurulmaktadır. Bu durum küçük değişikliklerin eğitim sürecini gereksiz yere uzatmasını önlemeye yardımcı olduğundan eğitim sürelerinin kısalmasına da katkı sağlamaktadır. Erken durdurma kriteri olan “n_iter_no_change” parametresi için bu çalışmada 50, 100, 200 ve 300 değerleri

denenmiş, elde edilen eğitim ve test performansları doğrultusunda söz konusu parametre dört sınıflı modelde 300, üç sınıflı ve iki sınıflı modelde ise 200 olarak alınmıştır.

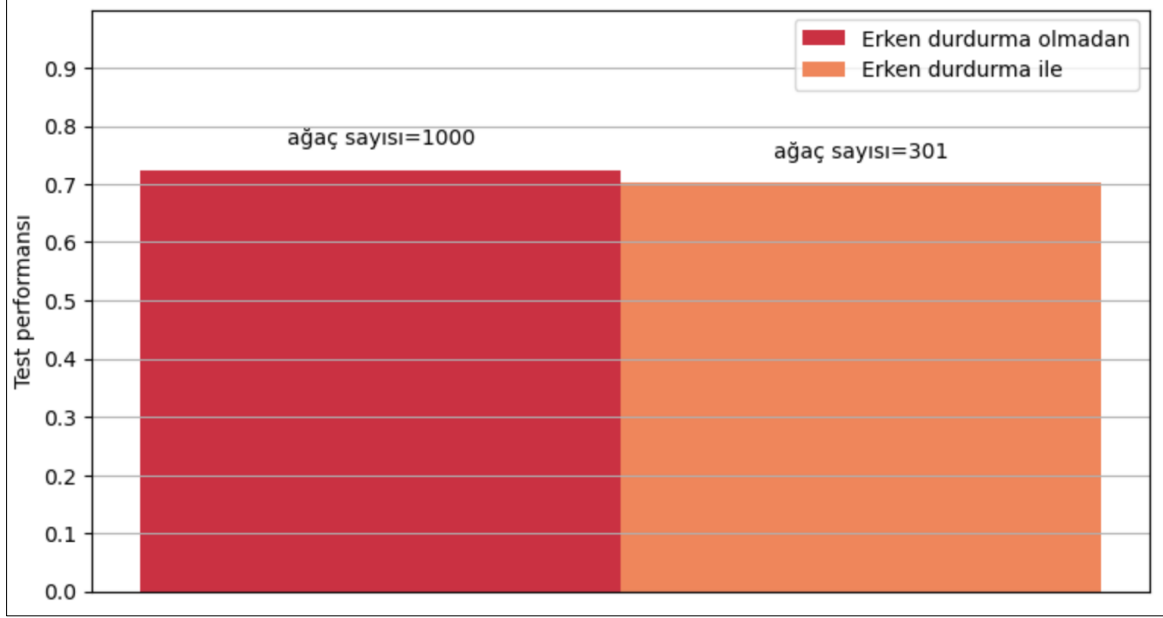
GAM algoritması kullanılarak erken durdurma öncesi ve sonrası test performanslarının karşılaştırılması amaçlandığından öncelikle dört sınıflı model için GAM algoritmasına ÇDÖÖE kullanılarak özellik seçimi yaptırılmıştır. GAM algoritması tarafından belirlenen ve Şekil 5.22’de verilen 18 özellik ile erken durdurma analizine devam edilmiştir.



Şekil 5.22. Dört sınıflı model için GAM tarafından seçilen 18 özelliğin önem sırası

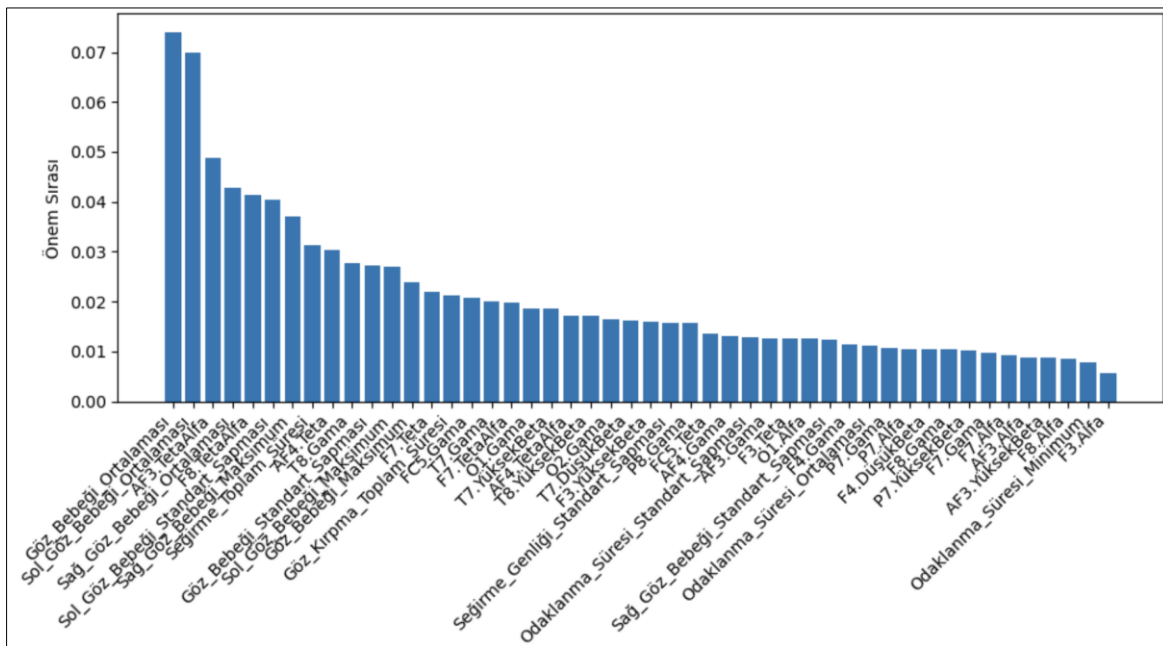
Beş katlı ızgara arama yöntemi ile öğrenme oranı ve maksimum derinlik hiperparametreleri ayarlanmış (öğrenme oranı=0,1, maksimum derinlik=9, ağaç sayısı=1000), sonrasında yine beş katlı çapraz doğrulama ile erken durdurma gerçekleştirilen ve gerçekleştirilmeyen modellerin performansları karşılaştırılmıştır. Erken durdurma kriterleri olarak “n_iter_no_change” parametresi 300, “tol” parametresi ise 0,01 olarak belirlenmiştir. Bir diğer ifadeyle, 300 iterasyon boyunca doğrulama kaybında 0,01’den daha az düşüş gerçekleşirse yani düzelme meydana gelmezse eğitimin durdurulması istenmektedir. Şekil 5.23’te görülebileceği üzere, erken durdurma sonucunda 301 ağaç ile oluşturulan modellerin test doğruluk performansları ortalaması %70,26 olarak bulunmuştur. Erken durdurma yapılmaksızın elde edilen performans ise %72,37’dir. GAM algoritmasının 111 değişkenli ayarlanmamış modelde gösterdiği %61,93 oranındaki doğruluk performansının, boyut azaltma ve hiperparametre ayarlama sonrası yükselişi

dikkati çekmektedir. Erken durdurma sonrası katlı modellerin ortalama eğitim doğruluk performansı %97,09, kappa ve MKK skorları ise 0,60 olarak elde edilmiştir.



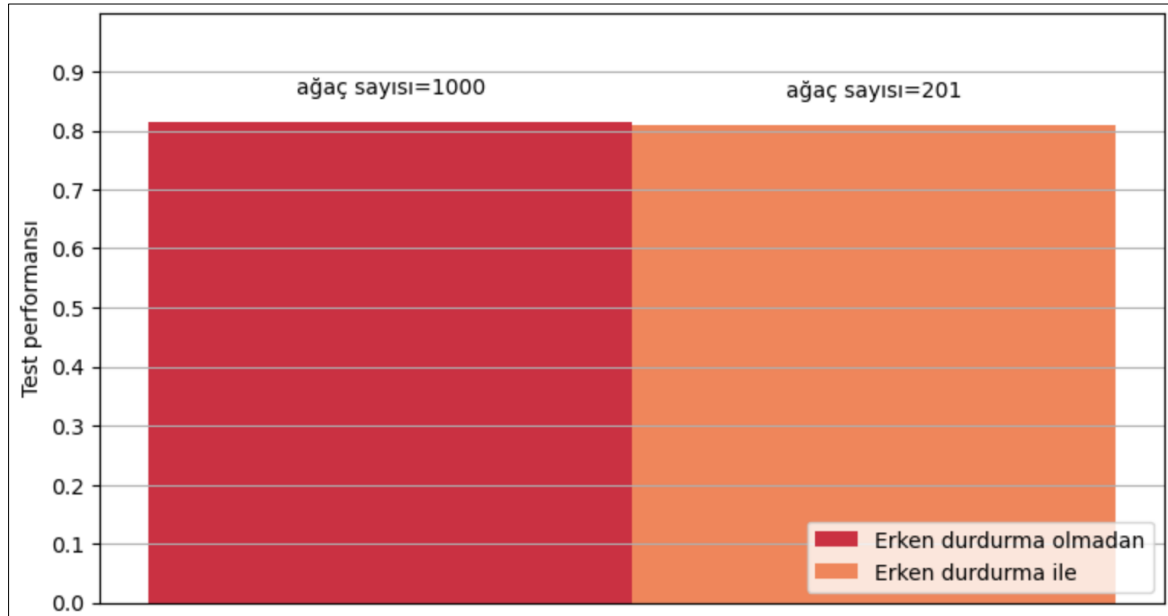
Şekil 5.23. Dört sınıflı model için erken durdurma sonrası performans karşılaştırma

Benzer prosedür üç sınıflı ve iki sınıflı modeller için de gerçekleştirilmiştir. Üç sınıflı model için GAM algoritması tarafından 48 özellik seçilmiştir. Şekil 5.24'te 48 özellik önem sırasına göre sıralanmış halde verilmektedir.



Şekil 5.24. Üç sınıflı model için GAM tarafından seçilen 48 özelliğin önem sırası

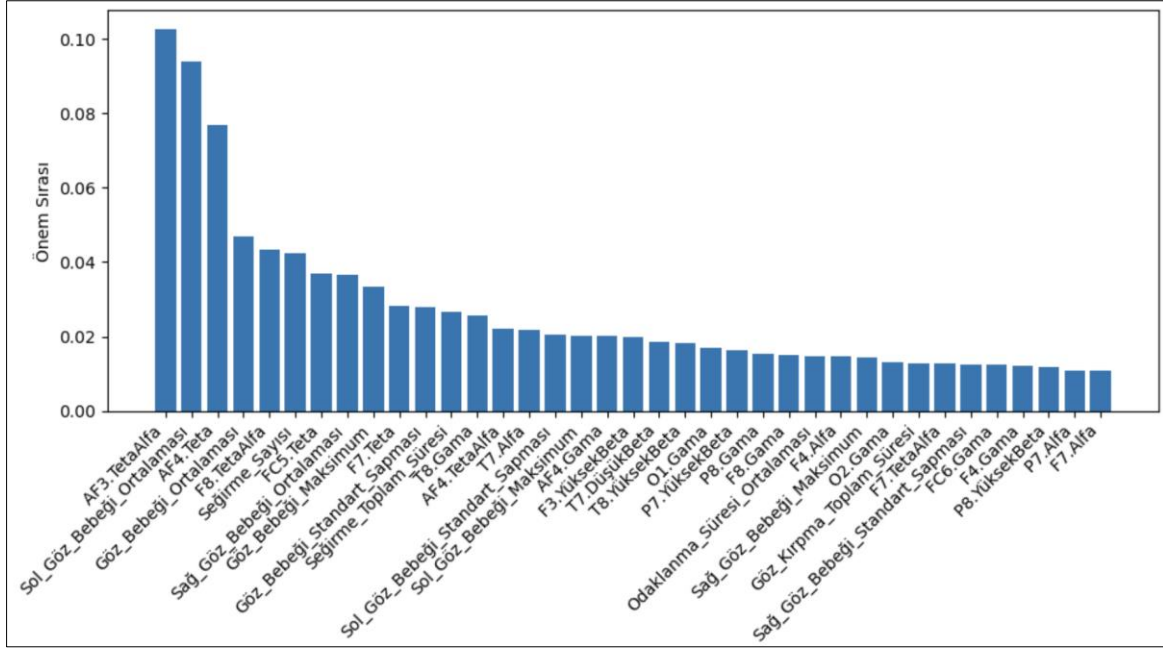
Yalnızca belirlenen 48 değişken kullanılarak uygulanan ızgara arama yöntemiyle en iyi sonucu veren hiperparametreler belirlenmiştir (öğrenme oranı=0,2, maksimum derinlik=9, ağaç sayısı=1000). Ayarlanmış model beş katlı çapraz doğrulama ile önce erken durdurma olmadan daha sonra da “n_iter_no_change” parametresinin 200, “tol” parametresinin ise 0,01 olarak belirlendiği erken durdurma kriterli olarak çalıştırılmış ve sonuçlar karşılaştırılmıştır. Şekil 5.25’te erken durdurma sonrası modelin 201 ağaçtan oluştuğu görülmektedir. Test doğruluk performansı ortalaması %80,79, eğitim doğruluk performansı ortalaması ise %97,85 olarak elde edilmiştir. Erken durdurma olmadığında bu oranlar sırasıyla %81,38 ve %100’dür. Erken durdurma sonrası aşırı öğrenme riskinin azaltıldığı ve modelin daha genellenebilir hale geldiği durumda kappa ve MKK skorları 0,71 olarak elde edilmiştir.



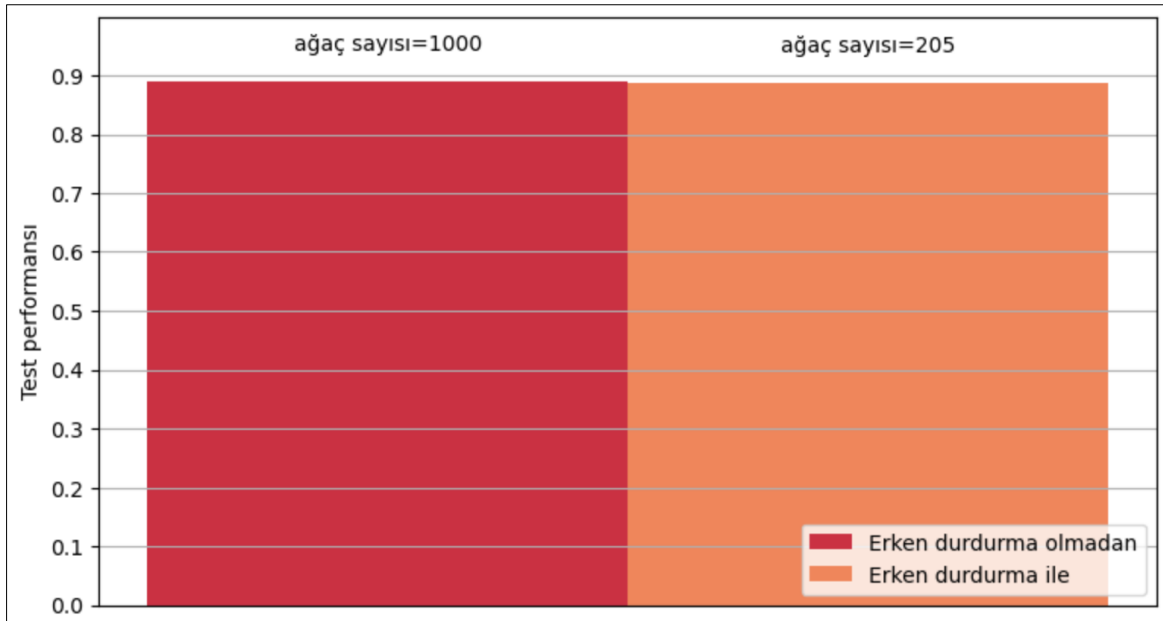
Şekil 5.25. Üç sınıflı model için erken durdurma sonrası performans karşılaştırma

İki sınıflı model için GAM algoritması tarafından 37 değişken seçilmiştir. Seçilen bu değişkenlerin modele sağladığı katkı açısından önem sırası Şekil 5.26’da verilmiştir. 37 değişkenli model üzerinde uygulanan beş katlı ızgara arama yöntemi neticesinde öğrenme oranının 0,2, maksimum derinlik değerinin 9 olarak belirlenmesine karar verilmiştir. Ağaç sayısı 1000’dir. Ayarlanmış model önce erken durdurma olmaksızın, daha sonra da “n_iter_no_change” parametresine 200, “tol” parametresine ise 0,01 değerlerinin verildiği erken durdurma kriterli şekilde beş farklı eğitim ve test seti üzerinde çalıştırılmıştır. Şekil 5.26’da karşılaştırma sonuçları verilmiştir. Erken durdurma sonrası 205 ağaç ile kurulan iki

sınıflı modelin test doğruluk performansı ortalaması %88,74, eğitim doğruluk performansı ortalaması ise %98,63 olarak elde edilmiştir. Erken durdurma uygulanmayan modelde bu oranlar sırasıyla %89,92 ve %100 olarak elde edilmiştir. %100 eğitim doğruluk performansının azaltılması ile aşırı öğrenme riskinin de azaltılması hedeflenmiştir. Erken durdurma sonrası kappa ve MKK skorları 0,77 olarak bulunmuştur.

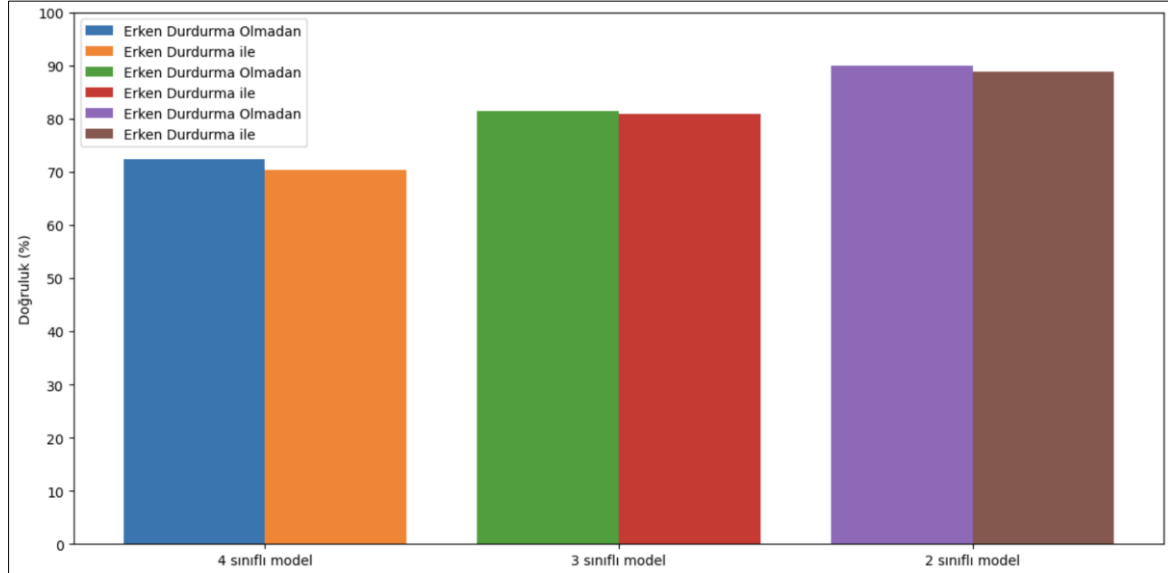


Şekil 5.26. İki sınıflı model için GAM tarafından seçilen 37 özelliğin önem sırası



Şekil 5.27. İki sınıflı model için erken durdurma sonrası performans karşılaştırma

Dört, üç ve iki sınıflı sınıflandırma modellerinin erken durdurma öncesi ve sonrası test doğruluk performanslarındaki değişimi özetleyen grafik Şekil 5.28’de görülebilmektedir.



Şekil 5.28. Erken durdurma sonrası performans karşılaştırma

Aşırı öğrenme riskini azaltmak ve daha genellenebilir bir model elde etmek üzere GAM algoritması kullanılarak yapılan ek analiz neticesinde dört sınıflı modelde %70’e, üç sınıflı modelde %80’e, iki sınıflı modelde de %90’a yakın test performansına ulaşılabilmektedir.

Elde edilen modellerin doğruluk performansları literatürdeki benzer çalışmalar ile Çizelge 5.26’da karşılaştırılmıştır. Çizelgede sadece EEG ve göz izleme değişkenlerinin kullanıldığı çalışmalar dikkate alınmıştır.

Çizelge 5.26. Model başarısının literatür ile karşılaştırılması

Yazar (Yıl)	Katılımcı Sayısı	Sınıf Sayısı	Ölçüm Yöntemi	Doğruluk Performansı
Grimes ve diğerleri (2008)	8	2-4	EEG	2 sınıflı model: %99 4 sınıflı model: %88
Wang ve diğerleri (2012)	8	3	EEG	%30-84
Borys ve diğerleri (2017a)	13	2-3	EEG + Göz İzleme	2 sınıflı model: %90 3 sınıflı model: %73
Yin ve Zhang (2017)	7	2	EEG	%85,7

Çizelge 5.26. (devam) Model başarısının literatür ile karşılaştırılması

Yazar (Yıl)	Katılımcı Sayısı	Sınıf Sayısı	Ölçüm Yöntemi	Doğruluk Performansı
Lim ve diğerleri (2018)	48	3	EEG	%69
Jusas ve Samuvel (2019)	9	4	EEG	%64
Plechawska-Wojcik ve diğerleri (2019)	11	3	EEG	%71,1-90,5
Wu ve diğerleri (2019)	32 uzman + 7 uzman olmayan	2	Göz izleme	%97
Kaczorowska ve diğerleri (2020)	26	2	Göz izleme	%97
Qu ve diğerleri (2020)	10	3	EEG	%79,8
Kaczorowska ve diğerleri (2021)	29	3	Göz izleme	%97
Pei ve diğerleri (2021)	7	3	EEG	Farklı EEG öznitelikleri ile: %84,3 Yalnızca bant gücü ile: %75,9
Zhou ve diğerleri (2022)	45	7	EEG	%56
Li ve diğerleri (2023)	28	2-3	EEG + Göz İzleme	2 sınıflı model: %82,7 3 sınıflı model: %54,2
Şaşmaz ve diğerleri (2023)	17 hasta 28 sağlıklı	3	EEG	%83,4
Mevcut çalışma	15	2-3-4	EEG + Göz İzleme	2 sınıflı model: %89,92 3 sınıflı model: %80,79 4 sınıflı model: %70,26

Çizelge 5.26’da görülebileceği üzere, mevcut çalışmanın sonuçlarının daha iyi olduğu çalışmalar olduğu gibi (Jusas ve Samuvel, 2019; Li ve diğerleri, 2023; Lim ve diğerleri, 2018; Qu ve diğerleri, 2020; Yin ve Zhang, 2017), mevcut çalışmaya kıyasla daha iyi

sonuçların elde edildiği çalışmalar da (Grimes ve diğerleri, 2008; Kaczorowska ve diğerleri, 2021) mevcuttur. Daha iyi sonuç veren EEG çalışmaları incelendiğinde, mevcut çalışmadan farklı olarak cihazdaki kanal sayılarının daha yüksek olduğu dikkati çekmektedir. EEG’de elektrot noktalarının haricindeki kısımlara ilişkin doğrudan ölçüm yapılamamaktadır. Voltaj değerindeki değişimin bir noktadan diğerine sorunsuz, yumuşak bir şekilde gerçekleştiği varsayılmakta, interpolasyon yöntemiyle elektrotlar arası aktivitenin nasıl görünebileceği bir nevi tahmin edilmektedir. EEG aktivitesini lokal olarak kesin bir şekilde ölçmemek EEG yönteminin dezavantajlarından birisidir. Bu nedenle kanal sayısının fazla olmasının lokal olarak daha hassas bir ölçüm yapabilme ve daha iyi tahmin değişkenleri elde edebilme açısından çalışmalarda bir avantaj olabileceği değerlendirilmektedir. Daha iyi sonuç veren göz izleme çalışmaları incelendiğinde ise, daha yüksek katılımcı sayıları dikkati çekmektedir. Bu çalışmada kullanılan katılımcı sayısı analizlerin güvenilirliği açısından yeterli olmakla birlikte, katılımcı sayısının artırılmasının analize dâhil edilebilecek daha fazla sağlıklı veri elde edebilmek açısından önemli olduğu düşünülmektedir.

Regresyon probleminin çözümüne ilişkin bulgular

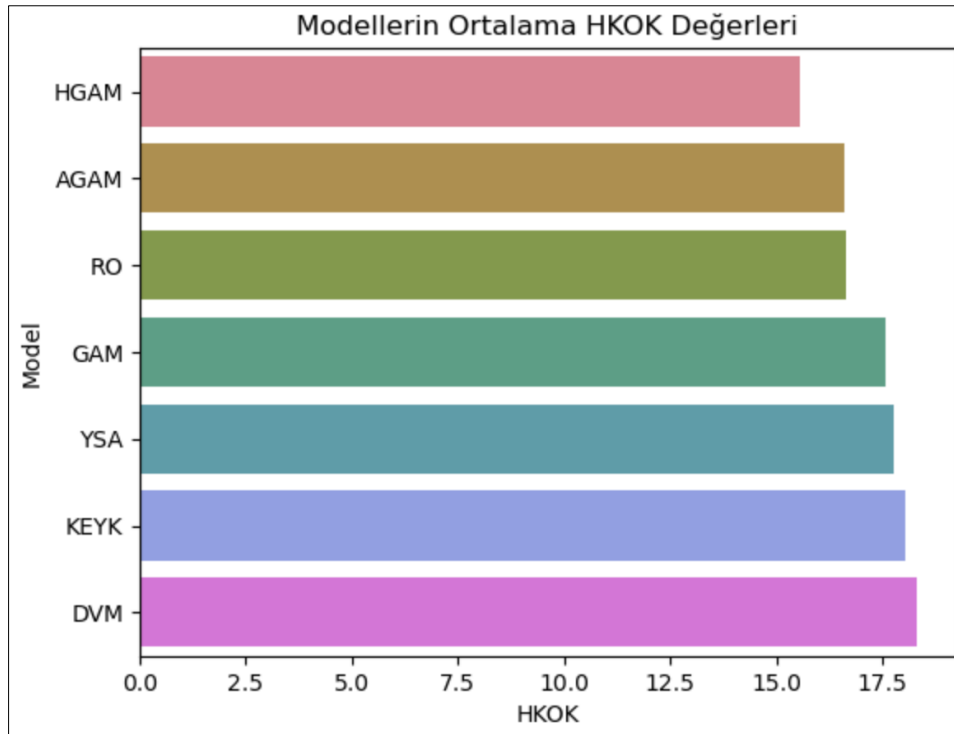
EEG ve göz izleme değişkenlerinin girdi değişkeni, “Ağırlıklı Toplam NASA-TLX Skoru” değişkeninin ise çıktı değişkeni olarak ele alındığı tahmin modelleri KEYK, DVM, YSA, RO, GAM, AGAM ve HGAM algoritmaları ile çalıştırılmıştır. Sınıflandırma probleminin çözümüne benzer şekilde KEYK, DVM ve YSA algoritmaları ile çalıştırılan modellerde “StandardScaler” fonksiyonu kullanılarak veri dönüştürme işlemi gerçekleştirilmiştir. Boyut azaltma öncesi hiperparametre ayarlama yapılmamış, öncelikle algoritmaların sonuçlarının herhangi bir iyileştirme yapılmaksızın karşılaştırılması hedeflenmiştir.

Çizelge 5.27’de farklı algoritmalarla kurulan 111 değişkenli regresyon modellerinin performans değerleri verilmiştir. Verilen değerler beş katlı çapraz doğrulama ile elde edilen beş ayrı performans sonucunun ortalama değerleridir. Performans ölçüsü olarak regresyon problemlerinde kullanılan ve “4.3. Model Performans Metrikleri” bölümünde açıklanan HKOK ve R^2 değerleri incelenmiştir.

Çizelge 5.27. 111 değişkenli regresyon modellerinin performans sonuçları

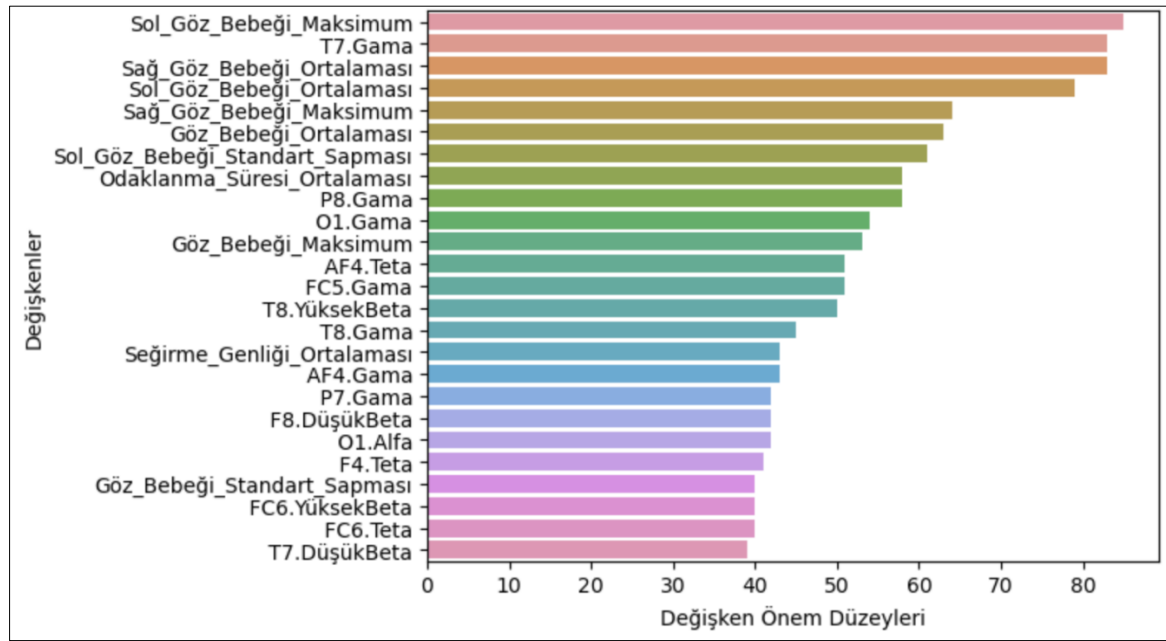
Model	HKOK	R ² (%)
KEYK	18,03	47,90
DVM	18,32	46,21
YSA	17,76	49,42
RO	16,63	55,63
GAM	17,56	50,56
AGAM	16,60	55,73
HGAM	15,54	61,27

Beş katlı çapraz doğrulama ile kurulan modellerin, kullanılan makine öğrenmesi algoritması bazında ortalama HKOK performanslarına ilişkin karşılaştırma Şekil 5.29’da görülmektedir. Grafik incelendiğinde en az hata yapan ve hedef değişkeni en iyi açıklayan modelin HGAM algoritması ile elde edildiği (HKOK=15,54, R²=%61,27), diğer en iyi sonucu veren algoritmaların ise AGAM, RO ve GAM olduğu görülmektedir.



Şekil 5.29. Modellerin HKOK performanslarının karşılaştırılması

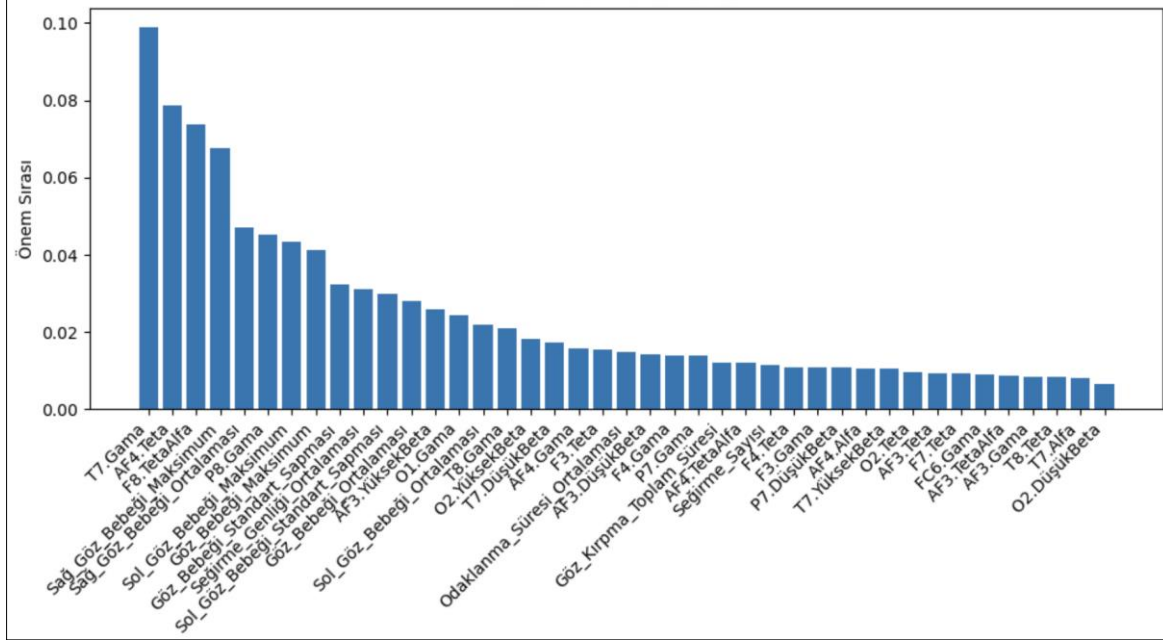
Sınıflandırma probleminin çözümündekine benzer şekilde boyut azaltma yoluyla performansta iyileştirme gerçekleştirilmesi hedeflenmiştir. ÇDÖÖE kullanılarak gerçekleştirilen özellik seçiminde HGAM algoritması tarafından 77 değişken önemli seçilmiştir. En önemli bulunan ilk 25 değişken Şekil 5.30’da verilmiştir. Yine göz bebeği çapı, odaklanma süresi ve seğirme genliğine ilişkin değişkenler ile ön bölge teta, parietal, temporal ve oksipital gama ve oksipital alfa değişkenlerinin modele en fazla katkı sağlayan değişkenlerden olduğu tespit edilmiştir.



Şekil 5.30. Regresyon modeline katkı sağlayan en önemli değişkenler

Izgara arama yöntemi ile belirlenen hiperparametreler kullanılarak (öğrenme oranı=0,1, maksimum derinlik=13, ağaç sayısı=1000, yaprak sayısı=16) beş katlı doğrulama ile 77 değişkenli modeller eğitilmiş ve HKOK değeri 15,09, R^2 değeri ise %63,49 olarak elde edilmiştir. Düzeltilmiş R^2 değeri ise 62,14’tür. Bu sonuçlar; kurulan regresyon modelinin, 100 üzerinden ölçülen ağırlıklı toplam NASA-TLX skorunu 15,09 ortalama hata ile tahmin ettiği ve kullanılan bağımsız değişkenlerin (EEG ve göz değişkenleri) bağımlı değişken varyansının %62,14’ünü açıkladığı şeklinde yorumlanmaktadır. Ancak, eğitilen modelin eğitim hatası incelendiğinde test hatasından oldukça düşük bir hata değeri ile karşılaşılmıştır (HKOK=0,17). Bu durum yine modelin aşırı öğrenmiş olabileceği riskini değerlendirmeyi gerektirmiş, sınıflandırma probleminin çözümüne benzer şekilde GAM algoritmasıyla erken durdurma yöntemi uygulandığında nasıl bir sonuç elde edileceği araştırılmıştır. Bunun için öncelikle GAM algoritmasının belirlediği en önemli değişkenler

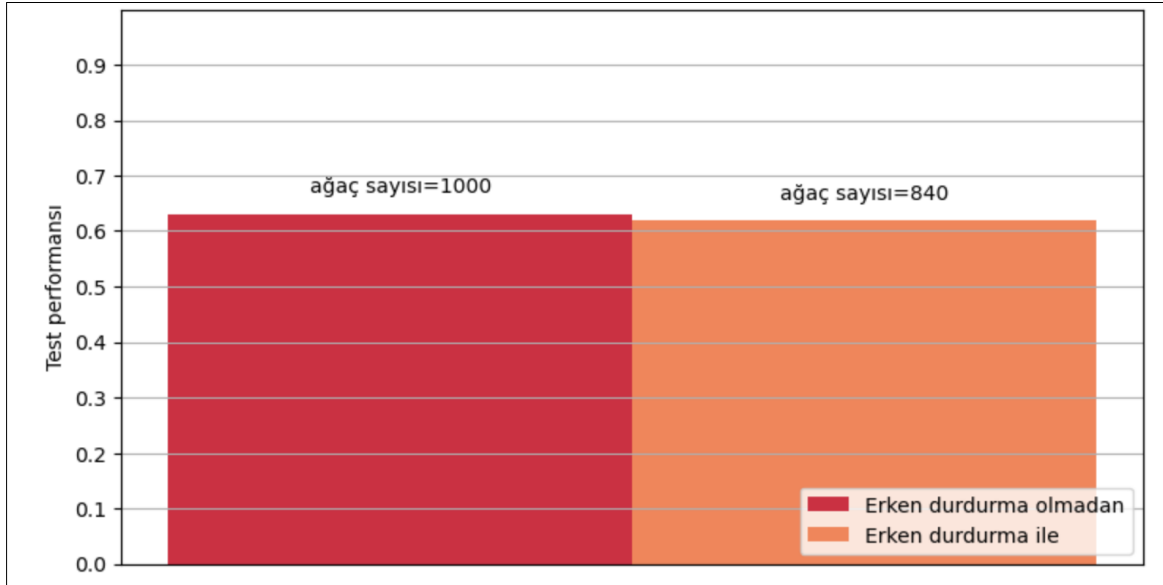
incelenmiştir. Şekil 5.31’de regresyon modeli için GAM algoritmasınca belirlenen en önemli 41 değişken önem sırasına göre verilmiştir.



Şekil 5.31. Regresyon modeli için GAM tarafından seçilen 41 özelliğin önem sırası

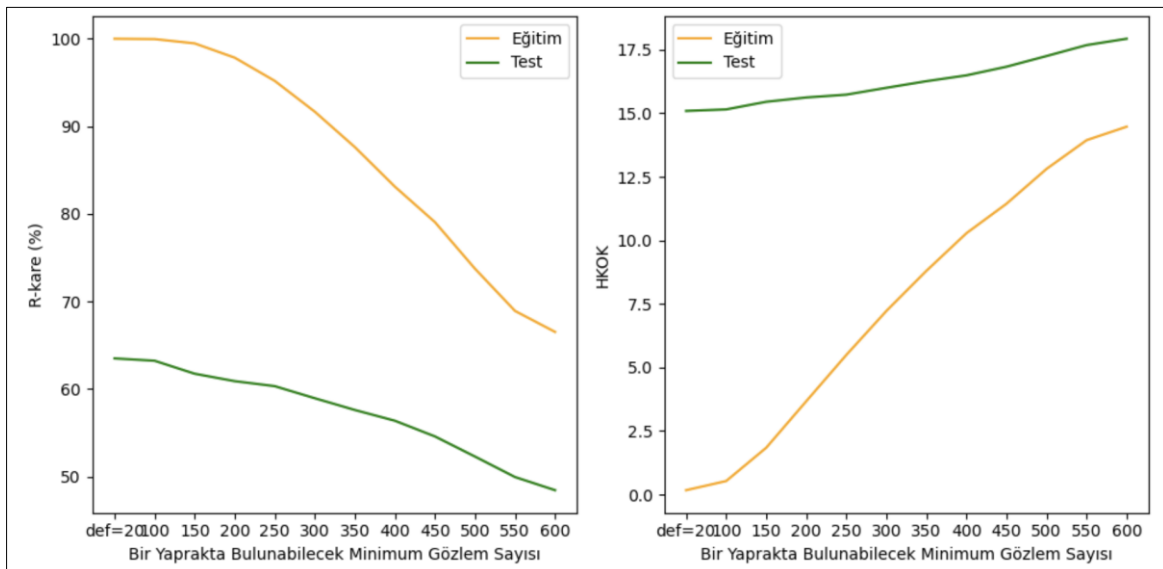
41 değişkenli model üzerinde uygulanan beş katlı ızgara arama yöntemi neticesinde öğrenme oranının 0,1, maksimum derinlik değerinin 7 olarak belirlenmesine karar verilmiştir. Ağaç sayısı 1000’dir. Ayarlanmış model önce erken durdurma olmaksızın, daha sonra da “n_iter_no_change” parametresine 200, “tol” parametresine ise 0,01 değerlerinin verildiği erken durdurma kriterli şekilde beş farklı eğitim ve test seti üzerinde çalıştırılmıştır. Şekil 5.32’de erken durdurma uygulanmayan ve erken durdurma uygulanan modellerin ortalama R^2 değerinin karşılaştırma sonuçları verilmiştir. Erken durdurma sonrası 840 ağaç ile kurulan regresyon modelinin ortalama test HKOK değeri 15,41, test R^2 değeri ise %61,89 olarak elde edilmiştir. Erken durdurma uygulanmayan modelde bu oranlar sırasıyla 15,17 ve %63,09 olarak elde edilmiştir. Erken durdurma uygulanan modelin eğitim performansı incelendiğinde ise; HKOK değerinin 4,97, R^2 değerinin %96,03 olduğu görülmüştür.

Eğitim ve test R^2 değerleri arasındaki farkın daha fazla azaltılabilmesi amacıyla HGAM algoritmasının bir yaprakta bulunabilecek minimum gözlem sayısı (min_child_samples) hiperparametresi için de farklı denemeler gerçekleştirilmiş, bu şekilde modelin aşırı öğrenmesinin önüne geçilmeye çalışılmıştır.



Şekil 5.32. Regresyon modeli için erken durdurma sonrası performans karşılaştırma

Şekil 5.33’te bir yaprakta bulunabilecek minimum gözlem sayısı hiperparametresinin farklı değerleri için eğitim ve test performanslarındaki değişim görülmektedir. Eğitim ve test performansları arasında çok büyük farklar olması modelin genellenebilirliği önünde önemli bir engeldir. Bu nedenle, eğitim ve test R^2 değerleri arasındaki farkın 20’nin altına düştüğü noktalardaki performansın dikkate alınmasının genellenebilir bir model için daha güvenilir olduğu değerlendirilmiştir. Bu noktada, eğitim ve test R^2 değerleri sırasıyla %68,91 ve %49,94 olarak elde edilmiştir. Eğitim ve test HKOK değerleri ise sırasıyla 13,94 ve 17,67’dir.



Şekil 5.33. Farklı hiperparametre değerlerine göre performans değişimi

Gama gücüne ilişkin değişkenlerin bağımlı değişkenler ile yüksek korelasyon katsayılı bir ilişki göstermemelerine rağmen makine öğrenmesi modellerinde tahmine en çok katkı sağlayan değişkenler arasında yer almaları dikkati çekmektedir. Nitekim, özellik seçiminde düşük korelasyonlu değişkenlerin otomatik olarak dışarıda bırakılması önerilmemektedir. Çünkü tek başına sınıf ayırt etme yeteneği düşük olan bir değişken başka değişkenlerle bir araya geldiğinde modele katkısı artabilmektedir. İki rassal değişkenin entropi açısından paylaştığı bilginin ölçüsü anlamına gelen ortak bilgi, değişkenler arasındaki doğrusal olmayan bağlantıları açığa çıkarabilmektedir (URL-25).

Tezin makine öğrenmesi bölümünde mümkün olan en yüksek doğruluk oranı ile zihinsel iş yükü seviyesini tahmin eden sınıflandırma modelinin geliştirilmesi ve daha sonra literatürdeki çalışmalar ile karşılaştırılması öncelikli hedef olmuştur. Regresyon modeli ise hedef değişken olarak ağırlıklı toplam NASA-TLX skorunun alındığı durumda ne ölçüde bir hata yapıldığını ve bağımsız değişkenlerin hedef değişkenin varyansının ne kadarını açıkladığını araştırmak üzere oluşturulmuştur. Problemin bir regresyon problemi olarak ele alınması durumunun incelenmesi amaçlanmıştır. Regresyon modelindeki hata miktarının yüksek çıkmasının nedeninin hedef değişkenin tam olarak sürekli bir yapı göstermemesi (NASA-TLX ölçeğinin kesikli yapısı) ile ilişkili olduğu değerlendirilmektedir. Yine de, ağırlıklı toplam NASA-TLX skorunun algılanan zihinsel iş yükünün farklı seviyelerine karşılık geldiği düşünüldüğünde, 100 üzerinden 15'e yakın bir hatayla çalışan bir modelin en azından katılımcı tarafından hissedilen düşük, orta ve yüksek zihinsel iş yükünü ayırt etmede etkili olacağı değerlendirilmektedir.

6. SONUÇ

Bu tez çalışmasında, zihinsel iş yükünün, n-geri görevler sırasında toplanan EEG ve göz izleme verilerine bağlı olarak analiz edilmesi ve makine öğrenmesi algoritmaları aracılığıyla tahmin edilmesi amaçlanmıştır. Tezin temel amaçları doğrultusunda, kullanılacak cihazların tanınması, ölçümler sırasında karşılaşılabilecek zorlukların öngörülebilmesi ve deney sürecinin test edilmesi amacıyla öncelikle pilot çalışmalar gerçekleştirilmiştir. Cihazların farklı zamanlarda temin edilebilmesi nedeniyle önce göz izleme cihazı ile daha sonra da EEG cihazı ile çalışılmıştır. Pandemi koşulları altında aynı dört kişi üzerinde farklı zamanlarda uygulanan deneylerin sonuçları analiz edilmiştir. Bu aşamada, veri ön işleme ve göz kırpma değişkenlerinin türetilmesi amacıyla hazırlanan Excel VBA makroları ile makine öğrenmesi analizi için hazırlanan Python kodları temel uygulamadaki analiz sürecine büyük ölçüde katkı sağlamıştır. Veri ön işleme ve değişken türetme adımları hem gerçek testlerde toplanacak verinin hazırlanma sürecinde uygulanacak prosedürün netleştirilmesine hem de kullanılacak ilave göz izleme değişkenleri ile modeldeki bağımlı değişkenler arasındaki beklenen ilişkiyi test etmeye imkan sağlamıştır. Pilot çalışmalar sayesinde, hem göz izleme hem de EEG'nin tek başlarına dahi zihinsel iş yükü analizinde önemli rol oynadığı görülmüş ve çalışmanın devamında her iki yöntemin birlikte kullanımının faydalı olacağına karar verilmiştir.

Çalışmanın devamında, daha geniş bir katılımcı grubuyla tezin ana verisini sağlayacak deneyler gerçekleştirilmiştir. Temel uygulamada pilot çalışmalardan farklı olarak ilk kez iki cihaz eş zamanlı kullanılarak ölçüm alınmıştır. Gerçek deneyler sırasında n-geri testlerin uygulandığı yazılımın gelişmiş versiyonunun kullanılması, tepki süresi ve doğru yanıt oranı gibi performans ölçülerinin de kayıt altına alınabilmesini mümkün kılmıştır. Böylece literatürde zihinsel iş yükü ölçümünde kullanılan nesnel-deneysel yöntemlerden biri daha analize dâhil edilebilmiştir. Testlerin içeriği ve uygulanış şeklini içeren deney prosedürü literatürdeki benzer çalışmalar örnek alınarak belirlenmiştir (Liu ve diğerleri, 2017). Pilot çalışmalardan farklı olarak, zorluk seviyelerine göre yapılmak istenen zihinsel manipülasyonun doğruluk performans kriteri ile ayırt edilebilirliğinin artırılması için görev sırasında eşleşmeme durumu için de yanıt verilmesi beklenmiştir. Bu sayede; hedef olmayanı tespit etme durumu için verilen doğru yanıtın bilinçli verilip verilmediğinin

kontrolü sağlanmış, doğru yanıt oranı performans kriterinin güvenilirliği ve ayırt ediciliği artırılmıştır.

Deneyleerin ilk gününde göz izleme cihazında yaşanan bağlantı problemi nedeniyle üç katılımcıdan toplanan göz izleme verisinin eksik olması ve EEG verisinin bazı katılımcılar için sinyal kalitesinin düşüklüğü sebebiyle uyaran bazında eksik veri içermesi, söz konusu katılımcıların analizin dışında bırakılması zorunluluğunu ortaya çıkarmıştır. Laboratuvarın ve göz izleme cihazının sınırlı kullanım süresi nedeniyle deneylerin tekrarlanabilmesi mümkün olmamıştır. Bu durum, çalışmanın kısıtlılıklarından birini oluşturmuştur.

Çalışmanın sonuçları tezin ilk amacı doğrultusunda incelendiğinde; beklendiği şekilde, hem NASA-TLX hem de performans değişkenlerinin görev zorluk seviyesiyle önemli ölçüde ilişkili olduğu ortaya çıkmıştır. Görev zorluk seviyesi arttıkça katılımcıların algıladıkları iş yükünün arttığı ve performanslarının düştüğü söylenebilmektedir. Elde edilen bulgu, uygulanmasına karar verilen testlerin zorluk seviyelerinin zihinsel iş yükünde istenen manipölasyona yol açtığını doğrulamaktadır. Bu sonuç, zorluk seviyelerinin zihinsel iş yükü seviyelerine karşılık geldiği varsayımı altında kurulacak olan sınıflandırma modeline dayanak oluşturması açısından önemlidir. Nitekim, zorluk seviyesinin doğrudan sınıflandırma hedef değişkeni olarak alındığı pek çok çalışma olduğu gibi zihinsel iş yükünün verilen görevlerle tetiklendiği doğrulandıktan sonra analize devam edilen benzer çalışmalar da mevcuttur (Buchholz ve Kopp, 2020). Zorluk seviyesi ile en yüksek korelasyonu gösteren NASA-TLX değişkeni “Zihinsel Talep” olmuştur. NASA-TLX toplam skoruna en fazla katkı sağlayan alt boyutun da aynı değişken olduğu görülmektedir. NASA-TLX’in iş yükü ölçümünde güvenilirliği kanıtlanmış bir yöntem olmasının yanında, bu çalışmada “Fiziksel Talep” alt boyutunun ağırlığının sıfır çıkması da söz konusu öznel iş yükü değerlendirme aracının tezin amacına uygun şekilde zihinsel iş yükü analizine yönelik uygulanmasını ve NASA-TLX ağırlıklı toplam skorunun regresyon probleminde çıktı değişkeni olarak kullanılmasını destekleyici bir bulgu olarak görülmektedir. Yanıt için verilen tepki süresindeki gecikme, önceden belirlenen zorluk seviyesi arttıkça büyük ölçüde artmaktadır. Doğru yanıt verme oranı ise görev zorlaştıkça azalmaktadır. Eşleşmeyi tespit edip doğru cevaplama sayısının eşleşmeme durumunu tespit edip doğru cevaplama kıyasla görev zorluk seviyesi ile daha yüksek korelasyonlu olduğu görülmüştür. Bu sonuç, görev zorlaştıkça hedefi bulmanın daha da zorlaştığını göstermektedir. Katılımcıların algıladıkları iş yükü ve alt boyutları da performans değişkenleri ile ilişkilidir. Tepki

süreleri daha uzun ölçülen testlerde katılımcılar daha çok iş yükü hissettiklerini beyan etmişlerdir. Benzer şekilde doğru yanıt sayıları da azalmıştır. Dört farklı sınıfın ikili karşılaştırmaları neticesinde, birinci sınıftaki katılımcıların diğer sınıflardaki katılımcılara göre daha düşük NASA-TLX skorları bildirdikleri ve ayrıca, tepki süresi ve doğru yanıt oranı açısından üst sınıflardan daha iyi performans elde ettikleri dikkati çekmiştir. Bu bulgular, birinci sınıf öğrencilerinin, testleri daha kolay ve az yorucu buldukları şeklinde yorumlanmaktadır. Cinsiyetler arasında algılanan iş yükü ve performans açısından anlamlı fark olup olmadığı incelendiğinde ise, erkeklerin görevler sırasında kadınlardan daha çok zaman baskısı hissettikleri ve ortalama tepki sürelerinin kadınlardan daha uzun olduğu tespit edilmiştir. Oturum bazında ele alınan ilk veri seti ile gerçekleştirilen analizin en kayda değer sonucu ise, görev zorluk seviyelerinin ikili olarak karşılaştırmaları neticesinde hem tüm NASA-TLX değişkenleri hem de tepki süresi ve doğru yanıt oranı açısından anlamlı fark göstermeleri olmuştur. Bu sonuçların, geliştirilmesi planlanan sınıflandırma modelinde kullanılacak sınıf sayısını destekleyici nitelikte olduğu değerlendirilmiştir.

Görev zorluk seviyesi ile en yüksek korelasyonlu göz değişkenleri; göz kırpma sayısı ve toplam süresi, seğirme sayısı ve toplam süresi, sol ve sağ göz bebeği çapı ortalaması olarak bulunmuştur. Belirtilen değişkenlerin tamamı görev zorluk seviyesi arttıkça artmaktadır. Ağırlıklı toplam NASA-TLX skoru ile en çok ilişkili göz değişkenleri ise; göz kırpma sayısı ve toplam süresi, seğirme genliği ortalaması, maksimum değeri ve standart sapması ile seğirme sayısı ve toplam süresidir. Göz bebeği çapı ortalamaları görev zorluk seviyesi ile olduğu kadar ilişkili bulunmamıştır. Görev zorluk seviyesinden daha fazla ilişkili olduğu tespit edilen değişkenler ise seğirme genliğine ilişkin değişkenlerdir. Göz izleme cihazının verdiği raporda yer almayan ve tamamen diğer değişkenlerden yararlanılarak türetilen göz kırpma değişkenlerinin hem görev zorluk seviyesi ile hem de ağırlıklı toplam NASA-TLX skoru ile en yüksek korelasyona sahip değişkenler olduğu gözlenmiştir. K-W testi ile göz izleme değişkenlerinin tamamının zorluk seviyesine göre anlamlı fark gösterdiği görülmüştür. Görev zorluk seviyesine göre düzgün artış ya da azalış gösteren değişkenler; seğirme, göz kırpma ve göz bebeği ile ilişkili değişkenlerdir. Bu değişkenler tahmin edildiği gibi sınıflandırma modeline de en çok katkı sağlayan değişkenlerden olmuştur. Göz kırpma değişkenlerinin hem istatistiksel testlerde bağımlı değişkenlerle en çok ilişkili göz değişkenleri arasında yer almaları hem de sınıflandırma modellerinde uygulanan özellik seçiminde makine öğrenmesi algoritmaları tarafından model

performansına en çok katkı sağlayan değişkenler arasında seçilmeleri, tez kapsamında gerçekleştirilen değişken türetme yaklaşımının güvenilirliğini destekleyen bulgular olmuştur.

EEG değişkenlerinin analizinde, literatürdeki benzer çalışmalar ile aynı doğrultuda sonuçlar elde edilmiştir. Özellikle prefrontal, frontal ve frontal merkez bölgelerindeki yani ön beyindeki teta gücünün görev zorluğu arttıkça arttığı doğrulanmıştır. Aynı beyin bölgelerinde teta/alfa gücü oranlarına ilişkin değişkenlerin korelasyon katsayıları biraz daha yüksek bulunmuştur. Teta ve alfa değişkenlerinin oran şeklinde kullanımları sayesinde tahmin gücü daha yüksek değişkenler elde edilebildiğini söylemek mümkündür. Bu değişkenler makine öğrenmesi bölümünde de tahmin modellerine en çok katkı sağlayan değişkenler arasına girmiştir. Korelasyon katsayıları daha düşük olmasına rağmen, artan görev zorluk derecesi ile frontal bölgelerdeki alfa gücü artarken; temporal, parietal ve oksipital bölgelerdeki alfa gücü azalmaktadır. Görev zorlaştıkça düşük beta gücünün hemen hemen tüm beyin bölgelerinde önemli ölçüde azaldığı, gama gücünün ise arttığı gözlemlenmiştir.

Çalışmanın makine öğrenmesi bölümünde, en başarılı sonuçların pilot çalışmalarda olduğu gibi yine ağaç tabanlı algoritmalar ile elde edildiği görülmüştür. Yakın geçmişte geliştirilen ve bu nedenle literatürde bu konuda yapılan çalışmalarda kullanımına sık rastlanmayan HGAM, AGAM ve GAM algoritmaları ile yüksek başarı oranları elde edilmiştir. Boyut azaltma ve hiperparametre ayarlama sayesinde performansta ne ölçüde iyileşme sağlanabileceği gösterilmiştir. Ayrıca sınıf sayısının azaltılması yoluna gidildiğinde çok daha iyi sınıflandırma performansları ile karşılaşılacağı görülmüştür. Üç sınıflı ve iki sınıflı modeller beklendiği şekilde daha iyi performans göstermiştir. Sadece odak model üzerinde özellik seçimi, hiperparametre ayarlama ve erken durdurma yöntemlerine başvurulmuş, diğer tüm algoritmalar başlangıçta karşılaştırma amaçlı olarak kullanılmıştır. Hem EEG hem de göz değişkenlerinin girdi olarak kullanıldığı dört sınıflı sınıflandırma modelinde boyut azaltma, hiperparametre ayarlama ve erken durdurma adımlarının ardından yaklaşık %70 oranında doğruluk performansı elde edilmiştir. Modeller tarafından en çok hatanın 1-geri görev sınıfına ilişkin gözlemlerin sınıflandırılmasında yapıldığı tespit edilmiş, bu nedenle 1-geri görev sırasında elde edilen veri sınıfının dikkate alınmadığı üç sınıflı (düşük-orta-yüksek) modeller üzerinde çalışılmıştır. Üç sınıflı sınıflandırmada erken durdurma sonrası yaklaşık %80 doğruluk

oranı elde edilmiştir. Düşük ve yüksek zihinsel iş yükünün tahminini amaçlayan iki sınıflı sınıflandırma modeli ise yaklaşık %90 oranında doğru tahmin yapabilmektedir.

Yalnızca EEG değişkenlerinin ya da yalnızca göz değişkenlerinin kullanıldığı modellere kıyasla her iki fizyolojik yöntemin bir arada kullanıldığı model daha doğru sınıflandırma tahmini sunmuştur. Bu sonuç, farklı kaynaklardan eş zamanlı toplanan verilerin tahmin modeli elde etmede bize daha güvenilir sonuçlar sağladığını ortaya koymakta ve literatürde fizyolojik yöntemlerin bir arada kullanımının daha doğru sonuçlar verdiğini gösteren diğer çalışmaları (Liu ve diğerleri, 2017) desteklemektedir.

Kurulan makine öğrenmesi modellerinde, aşırı öğrenme riskini azaltmak ve modellerin genellenebilirliğini artırmak için çapraz doğrulama yapılmış, topluluk öğrenmesi yöntemleriyle birden fazla model kurulmuş ve erken durdurma yöntemine başvurulmuştur. Aşırı öğrenmeden kaçınmak üzere HGAM algoritmasının diğer hiperparametreleri üzerinde de çalışılmış, örneğin dallanma için gerekli minimum gözlem sayısına karşılık gelen ve “min_child samples” olarak belirtilen hiperparametrenin değeri artırılarak eğitim performansında düşüş gerçekleştirilmiş ancak, belirlenen kriterler doğrultusunda en uygun durma noktasının belirlenebilmesi için GAM ile erken durdurma yöntemi tercih edilmiştir.

Modeller çalıştırılırken GAM algoritmasının öğrenme hızının HGAM algoritmasına kıyasla oldukça yavaş olduğu gözlenmiştir. HGAM algoritmasıyla yüksek doğruluk performansları elde edilebilmiş ancak analizin devamında modelin aşırı öğrenmiş olabileceği riski değerlendirilmiştir. Eğitim ve test performansları arasındaki fark azaltılmaya çalışılmış, bu doğrultuda erken durdurma yöntemi uygulanmıştır. HGAM algoritmasının karmaşık yapısı sebebiyle aşırı öğrenmeye eğilimi doğrulanmış, erken durdurma ile eğitim performansının %100’ün altına indirilebilmesi mümkün olamamıştır. Gradyanı artırılmış karar ağaçları birçok zayıf tahmin modelini, özellikle de karar ağaçlarını birleştirerek yinelemeli olarak güçlü tahmin modelleri oluşturmaya yönelik bir yöntem sınıfı olduğundan sınıflandırma modellerinin analizi GAM algoritması ile sürdürülmüş, erken durdurma yöntemi bu algoritma üzerinden uygulanmıştır.

GAM algoritması tarafından dört sınıflı model performansına katkı sağlayan en önemli değişkenlerin; sağ ve sol göz bebeği çapı, göz kırpma ve seğirme genliğine ilişkin değişkenler ile ön bölge (prefrontal ve frontal) teta gücü, temporal, parietal ve oksipital

gama gücü değişkenleri olduğu tespit edilmiştir. HGAM algoritması tarafından, odaklanma süreleri ile oksipital alfa ve temporal beta değişkenleri de önemli bulunmuştur.

Özetle bu tez, EEG ve göz izleme yöntemlerinin hem NASA-TLX ölçeği ve performansa dayalı yöntemlerle birlikte aynı deney prosedürü içerisinde kullanıldığı hem de verilerin n-geri görevler sırasında toplandığı daha önce gerçekleştirilmemiş özgün bir çalışma olarak literatüre sunulmuştur. EEG ve göz verilerinin eş zamanlı olarak kayıt altına alındığı deney süreci sonucunda makine öğrenmesi algoritmaları aracılığıyla zihinsel iş yükü tahmin modelleri geliştirilmiş, model performansları açısından tatmin edici sonuçlar elde edilmiştir. Özellikle dört sınıflı modelde literatüre kıyasla oldukça iyi bir doğruluk performansına ulaşılmıştır. EEG ve göz izleme değişkenlerinin zihinsel iş yükü ile aralarındaki ilişkiyi gösteren bulgular literatürdeki çalışmaları destekler niteliktedir. Tez çalışması, göz kırpmaya değişkeninin türetilmesine ilişkin farklı bir yaklaşım sunmuştur. Ayrıca tezin EEGLAB STUDY programının kullanımına ve sonuçlarının yorumlanmasına ilişkin bölümü ile literatüre bir uygulama örneği sağlanmıştır. Bu tez doğrultusunda yürütülen BAP projesi sayesinde satın alınan taşınabilir EEG cihazı, tezin belirtilen kapsamda gerçekleştirilebilmesini mümkün kılmakla birlikte, bundan sonra yapılacak bilimsel çalışmalara katkı sağlamak üzere üniversitenin altyapı envanterine eklenmiştir.

Tezin kapsamı gereği, deneyler tek bir görev türünün farklı zorluk seviyeleri uygulanarak gerçekleştirilmiştir. İlerleyen çalışmalarda, farklı özelliklerdeki bir katılımcı grubuyla ya da farklı görev koşulları altında toplanan veriler ile modeller test edilebilir. Bu sayede modellerin genellenebilirliği artırılabilir. Sanal gerçeklik kullanılarak daha gerçekçi görev koşulları altında ve kritik türdeki görevler sırasında elde edilen veriler ile kurulan modeller test edilebilir. Birbiriyle yüksek korelasyon gösteren değişkenlerin incelenmesiyle modellerde daha fazla boyut azaltma gerçekleştirilerek daha yalın modeller elde edilebilir. Doğrudan zihinsel iş yükü değerini tahmin etmeye yönelik çalışmalar hedeflenerek, sürekli iş yükü profillerinin elde edilmesini sağlayan farklı yazılımlar geliştirilebilir ve makine öğrenmesi modellerindeki zihinsel iş yükü çıktı değişkeninin de sürekli bir yapıda elde edilmesi sağlanabilir. Bu sayede, sürekli bir yapıda elde edilen fizyolojik değişkenlere karşılık gelen sürekli yapıdaki zihinsel iş yükü skoru kullanılarak zihinsel iş yükünü seviye olarak değil sayısal olarak tahmin eden modellerin geliştirilebilmesi mümkün olabilecektir.

KAYNAKLAR

- Ahlstrom, U. ve Friedman-Berg, F. (2006). Using eye movement activity as a correlate of cognitive workload, *International Journal of Industrial Ergonomics*, 36, 623-636.
- Akin, M., Kurt, M., Sezgin, N. ve Bayram, M. (2008). Estimating vigilance level by using EEG and EMG signals. *Neural Computing and Applications*, 17(3), 227-236.
- Altun, A. ve Çevik, V. (2012). Measuring working memory by using a multimedia based task. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, Özel Sayı:1, 32-40.
- Arico, P., Borghini, G., DiFlumeri, G., Colosimo, A., Pozzi, S. ve Babiloni, F. (2016). A passive brain-computer interface application for the mental workload assessment on Professional air traffic controllers during realistic air traffic control tasks. *Progress in Brain Research*, 228, 295-328.
- Ayaz, H., Shewokis, P. A., Bunce, S., İzzetoğlu, K., Willems, B. ve Onaral, B. (2012). Optical brain monitoring for operator training and mental workload assessment. *Neuroimage*, 59(1), 36-47.
- Ayaz, H., İzzetoğlu, M., İzzetoğlu, K. ve Onaral, B. (2016). *The use of functional near-infrared spectroscopy in neuroergonomics*. 1. International Neuroergonomics Conference, Paris, 17-25.
- Baldwin C.L. ve Penaranda, B.N. (2012). Adaptive training using an artificial neural network and EEG metrics for within- and cross-task workload classification. *Neuroimage*, 59(1), 48-56.
- Barwick, F., Arnett, P. ve Slobounov, S. (2012). EEG correlates of fatigue during administration of a neuropsychological test battery. *Clinical Neurophysiology*, 123(2), 278-284.
- Beatty, J., ve Lucero-Wagoner, B. (2000). *Handbook of psychophysiology*. (2. Baskı). Cambridge: Cambridge University Press, 142-162.
- Belkin, M., Hsu, D., Ma, S., ve Mandal, S. (2019). Reconciling modern machine-learning practice and the classical bias-variance trade-off. *Proceedings of the National Academy of Sciences*, 116(32), 15849-15854.
- Benedetto, S., Pedrotti, M., Minin, L., Baccino, T., Re, A. ve Montanari, R. (2011). Driver workload and eye blink duration. *Transportation Research Part F: Traffic Psychology and Behaviour*, 14, 199-208.
- Benerradi, J., Maior, H., Marinescu, A., Clos, J. ve Wilson, M. (2019). *Exploring machine learning approaches for classifying mental workload using fNIRS data from HCI tasks*. Halfway to the Future Symposium 2019, Nottingham, 1-11.
- Bezgin, L. (2016). *X, Y, Z kuşağı tüketicilerinin yeniden satın alma kararı üzerinde algılanan marka denkliği öğelerinin etki düzeyi farklılıklarının nörogörüntüleme*

tekniklerinden elektroensefalografi (EEG) ve göz izleme yöntemleriyle belirlenmesine dair deneysel bir çalışma. Yüksek Lisans Tezi, Hitit Üniversitesi Sosyal Bilimler Enstitüsü, Çorum, 65-82.

Bommer, S. ve Fendley, M. (2018). A theoretical framework for evaluating mental workload resources in human systems design for manufacturing operations. *International Journal of Industrial Ergonomics*, 63, 7-17.

Borghetti, B.J., Giametta, J.J. ve Rusnock, C.F. (2017). Assessing continuous operator workload with a hybrid scaffolded neuroergonomic modeling approach. *Human Factors*, 59(1), 134-146.

Borys, M., Plechawska-Wojcik, M., Wawrzyk, M. ve Wesolowska, K. (2017a). *Classifying cognitive workload using eye activity and eeg features in arithmetic tasks*. International Conference on Information and Software Technologies, Druskininkai, 90-105.

Borys, M., Tokovarov, M., Wawrzyk, M., Wesolowska, K., Plechawska-Wojcik, M., Dmytruk, R. ve Kaczorowska, M. (2017b). *An analysis of eye-tracking and electroencephalography data for cognitive load measurement during arithmetic tasks*. 10th International Symposium on Advanced Topics in Electrical Engineering, Bükreş, 287-292.

Botchkarev, A. (2018). Performance metrics (error measures) in machine learning regression, forecasting and prognostics: properties and typology. *ArXiv*, Ön basım, 1-37.

Breiman, L. (2001). Random forests. *Machine Learning*, 45, 5-32.

Brookings, J.B., Wilson, G.F. ve Swain, C.R. (1996). Psychophysiological responses to changes in workload during simulated air traffic control. *Biological Psychology*, 42, 361-377.

Buchholz, V. ve Kopp, S. (2020). *Towards an adaptive assistance system for monitoring tasks: assessing mental workload using eye-tracking and performance measures*. IEEE International Conference on Human-Machine Systems, Roma, 1-6.

Budak, H. (2018). Özellik seçim yöntemleri ve yeni bir yaklaşım. *Süleyman Demirel Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, 22, 21-31.

Cain, B. (2007). A review of the mental workload literature. *DRDO*, Toronto, 1-34.

Canbek, G., Sağıroğlu, S., Temizel, T. ve Baykal, N. (2017). *Binary classification performance measures/metrics: a comprehensive visualized roadmap to gain new insights*. International Conference on Computer Science and Engineering, Antalya, 821-826.

Cansız, Y. (2012). *Effects of way finding affordances on usability of virtual world environments in terms of users' satisfaction, performance, and mental workload: examination by eye-tracking and fnir device*. Yüksek Lisans Tezi, Orta Doğu Teknik

Üniversitesi Fen Bilimleri Enstitüsü, Ankara, 23-41.

- Casali, J.G. ve Wierwille, W.W. (1983). A comparison of rating scale, secondary-task, physiological, and primary-task workload estimation techniques in a simulated flight task emphasizing communications load. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 25(6), 623-641.
- Castor, M.C. (2003). Handbook of mental workload measurement. *GARTEUR*, Stockholm, 1-148.
- Cegarra, J. ve Chevalier, A., (2008). The use of Tholos software for combining measures of mental workload: Toward theoretical and methodological improvements. *Behavior research methods*, 40, 988-1000.
- Chen, S. ve Epps, J. (2013). Automatic classification of eye activity for cognitive load measurement with emotion interference. *Computer Methods and Programs in Biomedicine*, 110(2), 111-124.
- Chen, T. ve Guestrin, C. (2016). *XGBoost: A scalable tree boosting system*. 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, 785-794.
- Choi, M. K., Lee, S. M., Ha, J. S. ve Seong, P. H. (2018). Development of an EEG-based workload measurement method in nuclear power plants. *Annals of Nuclear Energy*, 111, 595-607.
- Chuang, S.W., Ko, L.W., Lin, Y.P., Huang, R.S., Jung, T.P., Lin, C.T. (2012). Co-modulatory spectral changes in independent brain processes are correlated with task performance. *Neuroimage*, 62(3), 1469-1477.
- Cox, D.R. (1958). The regression analysis of binary sequences, *Journal of the Royal Statistical Society Series B (Methodological)*, 20(2), 215-242.
- Çolakoğlu, A.A. (2020). *Makine öğrenmesi algoritmaları ile avrupa havalimanları analizi*. Yüksek Lisans Tezi, Hitit Üniversitesi Sosyal Bilimler Enstitüsü, Çorum, 10-28.
- Das, S., Maiti, J. ve Krishna, O.B. (2020). Assessing mental workload in virtual reality based EOT crane operations: A multi-measure approach. *International Journal of Industrial Ergonomics*, 80(7), 1-14.
- Davies, D.R. (1965). Skin conductance, alfa activity and vigilance. *American Journal of Psychology*, 78(2), 304-306.
- Delorme, A. ve Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics. *Journal of Neuroscience Methods*, 134, 9-12.
- DiDomenico, A. ve Nussbaum, M. (2011). Effects of different physical workload parameters on mental workload and performance. *International Journal of Industrial Ergonomics*, 41(3), 255-260.

- Dirican, A.C. ve Göktürk, M. (2011). Psychophysiological measures of human cognitive states applied in human computer interaction, *Procedia Computer Science*, 3, 1361-1367.
- Di Stasi, L. L., Antoli, A., Gea, M. ve Cañas, J. J. (2011). A neuroergonomic approach to evaluating mental workload in hypermedia interactions. *International Journal of Industrial Ergonomics*, 41(3), 298-304.
- Dodge, R. ve Cline, T. S. (1901). The angle velocity of eye movements. *Psychological Review*, 8(2), 145-157.
- Doğan, N. , Doğan, İ. ve Cangür, Ş. (2014). Birinci tür hatanın kontrolünde yanlış bulgu oranı (false discovery rate) yaklaşımı. *Düzce Üniversitesi Sağlık Bilimleri Enstitüsü Dergisi*, 3(3), 14-19.
- Duru, A . (2019). Determination of increased mental workload condition from eeg by the use of classification techniques. *International Journal of Advances in Engineering and Pure Sciences*, 31(1), 47-52.
- Ekim, A. (2021). *Bilişsel yorgunluğun EEG işaretleri ile sınıflandırılması*. Yüksek Lisans Tezi, Karadeniz Teknik Üniversitesi Fen Bilimleri Enstitüsü, Trabzon, 1-24.
- Ellis, K.K.E. (2009). *Eye tracking metrics for workload estimation in flight deck operations*. Yüksek Lisans Tezi, Iowa Üniversitesi, Iowa, 4-25.
- Erickson, B. ve Kitamura, F. (2021). Magician's corner: 9. performance metrics for machine learning models. *Radiology: Artificial Intelligence*, 3(3), 1-7.
- Evans, D.C. ve Fendley, M. (2017). A multi-measure approach for connecting cognitive workload and automation. *International Journal of Human-Computer Studies*, 97, 182-189.
- Friedman, J. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, 29, 1189-1232.
- Gale, A., Davies, R. ve Smallbone, A. (1977). EEG correlates of signal rate time in task and individual differences in reaction time during a five-stage sustained attention task. *Ergonomics*, 20, 363-376.
- Galy, E., Cariou, M. ve Melan, C. (2012). What is the relationship between mental workload factors and cognitive load types?. *International Journal of Psychophysiology*, 83(3), 269-275.
- Gao, Q., Wang, Y., Song, F., Li, Z. ve Dong, X. (2013). Mental workload measurement for emergency operating procedures in digital nuclear power plants. *Ergonomics*, 56, 1070-1085.
- Gevens, A.S., Bressler, S.L., Cutillo, B.A., Illes, J. ve Fowler-White, R.M. (1990). Effects of prolonged mental work on functional brain topography. *Electroencephalography and Clinical Neurophysiology*, 76, 339-350.

- Goldberg, J.H. ve Kotval, X.P. (1999). Computer interface evaluation using eye movements: methods and constructs. *International Journal of Industrial Ergonomics*, 24, 631-645.
- Gong, M. (2021). A novel performance measure for machine learning classification. *International Journal of Managing Information Technology*, 13(1), 11-19.
- Grafton, S.T. ve Tipper, C.M. (2012). Decoding intention: A neuroergonomic perspective. *NeuroImage*, 59(1), 14-24.
- Green S.B. (1991). How many subjects does it take do to a regression analysis. *Multivariate Behavioral Research*, 26(3), 499-510.
- Grimes D., Tan D. S., Hudson S. E., Shenoy P. ve Rao R. P. (2008). *Feasibility and pragmatics of classifying working memory load with an electroencephalograph*. 26th Annual SIGCHI Conference on Human Factors in Computing Systems, Floransa, 835-844.
- Guan, K., Chai, X., Zhang, Z., Li, Q. ve Niu, H. (2021). *Evaluation of mental workload in working memory tasks with different information types based on EEG*. 43rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Guadalajara, 5682-5685.
- Gundel, A. ve Wilson, G. (1992). Topographical changes in the ongoing EEG related to the difficulty of mental tasks. *Brain Topography*, 5(1), 17-25.
- Hampson, R., Opris, I. ve Deadwyler, S. (2010). Neural correlates of fast pupil dilation in nonhuman primates: relation to behavioral performance and cognitive workload. *Behavioural Brain Research*, 212(1), 1-11.
- Han, J., Zhou, X., Zong, C. ve Teng, F. (2023). *The feasibility of fatigue state monitoring based on eye parameters in virtual reality environment*. 6th International Conference on Intelligent Human Systems Integration: Integrating People and Intelligent Systems, Venedik, 246-254.
- Harputlu Aksu, Ş. ve Çakıt, E., (2022). *Classifying mental workload using EEG data: A machine learning approach*. 13th International Conference on Applied Human Factors and Ergonomics and the Affiliated Conferences, New York, 65-72.
- Hart, S. G. ve Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): results of empirical and theoretical research. *Advances in Psychology*, 52, 139-183.
- Hart, S. G. (2006). *Nasa-task load index (NASA-TLX); 20 years later*. 50th Annual Meeting of Human Factors and Ergonomics Society, San Fransisco, 904-908.
- Heine, T., Lenis, G., Reichensperger, P., Beran, T., Doessel, O. ve Deml, B. (2017). Electrocardiographic features for the measurement of drivers' mental workload. *Applied Ergonomics*, 61, 31-43.
- Herff, C., Heger, D., Fortmann, O., Hennrich, J., Putze, F. ve Schultz, T. (2014). Mental

- workload during n-back task-quantified in the prefrontal cortex using fNIRS. *Frontiers in Human Neuroscience*, 7, 935.
- Hoerl, A. ve Kennard, R. (2000). Ridge regression: biased estimation for nonorthogonal problems. *Technometrics*, 42(1), 80-86.
- Holm, A., Lukander, K., Korpela, J., Sallinen, M. ve Müller, K.M.I. (2009). Estimating brain load from the EEG. *The Scientific World Journal*, 9, 639-651.
- Howells, F. M., Stein, D. J. ve Russell, V. A. (2010). Perceived mental effort correlates with changes in tonic arousal during attentional tasks. *Behavioral and Brain Functions*, 6, 6-39.
- Huang, W., Xu, Y., Hildebrandt, M. ve Lau, N. (2019). *Comparing eye-gaze metrics of mental workload in monitoring process plants*. 16th Engineering International Conference, Orlando, 55-65.
- Iqbal, S.T., Zheng, X.S. ve Bailey, B.P. (2004). *Task-evoked pupillary response to mental workload in human-computer interaction*. Conference on Human Factors in Computing Systems, Viyana, 1477-1480.
- İnternet: URL-1, Time-Frequency Analysis of EEG Time Series Part 1: Fourier Analysis of EEG Signal, Web: <https://www.youtube.com/watch?v=wB417SAbdak&t=18s>, Son Erişim Tarihi: 07.06.2023.
- İnternet: URL-2, EMOTIV Pro 3 Kullanma Kılavuzu, Web: <https://emotiv.gitbook.io/emotivpro-v3/>, Son Erişim Tarihi: 07.06.2023.
- İnternet: URL-3, Winning Toolkit for Competitive ML, Web: <https://mlcontests.com/winning-toolkit/>, Son Erişim Tarihi: 22.08.2023.
- İnternet: URL-4, EMOTIV EPOC X Kullanma Kılavuzu, Web: <https://drive.google.com/file/d/1sxSmnG7Z8OjKnJ55llahBoWuq20UvC1X/view>, Son Erişim Tarihi: 12.06.2023.
- İnternet: URL-5, 20 Popular Machine Learning Metrics. Part 1: Classification & Regression Evaluation Metrics, Web: <https://towardsdatascience.com/20-popular-machine-learning-metrics-part-1-classification-regression-evaluation-metrics-1ca3e282a2ce>, Son Erişim Tarihi: 07.06.2023.
- İnternet: URL-6, Derin Öğrenme Uygulamalarında En Sık Kullanılan Hiperparametreler, Web: <https://medium.com/deep-learning-turkiye/derin-ogrenme-uygulamalarinda-en-sik-kullanilan-hiperparametreler-ece8e9125c4>, Son Erişim Tarihi: 08.06.2023.
- İnternet: URL-7, (2020) Python ile Makine Öğrenmesi (Machine Learning), Çevrimiçi Eğitim, Web: <https://www.udemy.com/course/python-ile-makine-ogrenmesi/>, Son Erişim Tarihi: 08.06.2023.
- İnternet: URL-8, Nedir Bu Destek Vektör Makineleri? (Makine Öğrenmesi Serisi-2), Web: <https://medium.com/deep-learning-turkiye/nedir-bu-destek-vekt%C3%B6r>

makinelere-makine-%C3%B6%C4%9Frenmesi-serisi-2-94e576e4223e, Son Erişim Tarihi: 08.06.2023.

İnternet: URL-9, Gradient Boosted Regresyon Ağaçları, Web: <https://www.veribilimiokulu.com/gradient-boosted-regresyon-agaclari/>, Son Erişim Tarihi: 08.06.2023.

İnternet: URL-10, LightGBM, Web: <https://www.veribilimiokulu.com/LightGBM/>, Son Erişim Tarihi: 08.06.2023.

İnternet: URL-11, Feature Selection with Sklearn and Pandas, Web: <https://towardsdatascience.com/feature-selection-with-pandas-e3690ad8504b>, Son Erişim Tarihi: 14.06.2023.

İnternet: URL-12, Comprehensive Guide on Feature Selection, Web: <https://www.kaggle.com/code/prashant111/comprehensive-guide-on-feature-selection/notebook>, Son Erişim Tarihi: 14.06.2023.

İnternet: URL-13, Özellik Oluşumu ve Özellik Seçimi (Feature Selection)-2, Web: <https://www.veribilimiokulu.com/ozellik-olusumu-ve-ozellik-secimifeature-selection-2/>, Son Erişim Tarihi: 14.06.2023.

İnternet: URL-14, Özellik Oluşumu ve Özellik Seçimi (Feature Selection)-3, Web: <https://www.veribilimiokulu.com/ozellik-olusumu-ve-ozellik-secimifeature-selection-3/>, Son Erişim Tarihi: 14.06.2023.

İnternet: URL-15, Receiver Operating Characteristic (ROC) Eğrisi, Web: <https://mehmetsoyer.medium.com/receiver-operating-characteristic-roc-e%C4%9Frisi-ee5c8b666533>, Son Erişim Tarihi: 08.06.2023.

İnternet: URL-16, What Does Your Classification Metric Tell About Your Data?, Web: <https://towardsdatascience.com/what-does-your-classification-metric-tell-about-your-data-4a8f35408a8b>, Son Erişim Tarihi: 08.06.2023.

İnternet: URL-17, Comprehensive Guide to Multiclass Classification Metrics, Web: <https://towardsdatascience.com/comprehensive-guide-on-multiclass-classification-metrics-af94cfb83fbd>, Son Erişim Tarihi: 08.06.2023.

İnternet: URL-18, Model Performansını Değerlendirmek: Regresyon, Web: <https://medium.com/yaz%C4%B1%C4%B1m-ve-bili%C5%9Fim-kul%C3%BCb%C3%BC/model-performans%C4%B1n%C4%B1-de%C4%9Ferlendirmek-regresyon-48b4afec8664>, Son Erişim Tarihi: 08.06.2023.

İnternet: URL-19, Inquisit Lab 6 Yazılımı, Web: <https://www.millisecond.com/download>, Son Erişim Tarihi: 12.06.2023.

İnternet: URL-20, EEGLAB 2021 Yazılımı, Web: <https://sccn.ucsd.edu/eeglab/downloadtoolbox.php>, Son Erişim Tarihi: 25.07.2023.

İnternet: URL-21, EEGLAB Öğreticisi, Web: <https://eeglab.org/tutorials/>, Son Erişim

Tarihi: 25.07.2023.

İnternet: URL-22, Hiperparametre Optimizasyonu, Web: <https://medium.com/bili%C5%9Fim-hareketi/hiperparametre-optimizasyonu-9ba0e7f32e6f>, Son Erişim Tarihi: 31.07.2023.

İnternet: URL-23, Gradient-Boosted Trees: To Early Stop or Not to Early Stop?, Web: <https://towardsdatascience.com/gradient-boosting-to-early-stop-or-not-to-early-stop-5ea67ac09d83>, Son Erişim Tarihi: 27.08.2023.

İnternet: URL-24, Early Stopping of Gradient Boosting, Web: https://scikit-learn.org/stable/auto_examples/ensemble/plot_gradient_boosting_early_stopping.html, Son Erişim Tarihi: 04.09.2023.

İnternet: URL-25, Feature Selection in Machine Learning: Variable Ranking and Feature Subset Selection Methods, Web: <https://medium.com/@mehulved1503/feature-selection-in-machine-learning-variable-ranking-and-feature-subset-selection-methods-89b2896f2220>, Son Erişim Tarihi: 08.09.2023.

Jacob, R. ve Karn, K. (2003). *The Mind's Eye: Cognitive and Applied Aspects of Eye Movement Research*. (1. Baskı). Amsterdam: Elsevier Science, 573–605.

Ji, Z., Tang, J., Wang, Q., Xie, X., Liu, J. ve Yin, Z. (2023). Cross-task cognitive workload recognition using a dynamic residual network with attention mechanism based on neurophysiological signals. *Computer Methods and Programs in Biomedicine*, 230, 1-16.

Jo, S., Myung, R. ve Yoon, D. (2012). Quantitative prediction of mental workload with the ACT-R cognitive architecture. *International Journal of Industrial Ergonomics*, 42(4), 359-370.

John, A., Singh, A., Do, T., Eidels, A., Nalivaiko, E., Mazloumi Gavgani, A., Brown, S., Bennett, M., Lal, S., Simpson, A., Gustin, S., Double, K., Walker, F., Kleitman, S., Morley, J. ve Lin, C. (2022). Unraveling the physiological correlates of mental workload variations in tracking and collision prediction tasks. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 30, 770-781.

Johns, M. (2003). *The amplitude velocity ratio of blinks: A new method for monitoring drowsiness*. 17th Annual Meeting of the Associated Professional Sleep Societies, Melbourne, 1-6.

Juris, M. ve Velden, M. (1977). The pupillary response to mental overload. *Physiological Psychology*, 5(4), 421-424.

Jusas, V. ve Samuvel, S.G. (2019). Classification of motor imagery using a combination of user-specific band and subject-specific band for brain-computer interface. *Applied Sciences*, 9, 4990.

Kaczorowska, M., Wawrzyk, M. ve Plechawska-Wojcik, M. (2020). *Binary classification of cognitive workload levels with oculography features*. 19th International

- Conference on Computer Information Systems and Industrial Management, Bialystok, 243-254.
- Kaczorowska M., Plechawska-Wojcik, M. ve Tokovarov, M. (2021). Interpretable machine learning models for three-way classification of cognitive workload levels for eye-tracking features. *Brain Sciences*, 11(2), 210.
- Kalaycı, Ş. (2008). *SPSS uygulamalı çok değişkenli istatistik teknikleri*. (3. Baskı). Ankara: Asil Yayın Dağıtım, 199-266.
- Kamzanova, A. T., Matthews, G., Kustubayeva, A. M. ve Jakupov, S.M. (2011). EEG indices to time-on-task effects and to a workload manipulation (cueing). *World Academy of Science, Engineering and Technology*, 80, 19-22.
- Kandemir, C., Handley, H. ve Thompson, D. (2018). A workload model to evaluate distracters and driver's aids. *International Journal of Industrial Ergonomics*, 63, 18-36.
- Kaneko, K. ve Sakamoto, K. (2001). Spontaneous blinks as a criterion of visual fatigue during prolonged work on visual display terminals. *Perceptual and Motor Skills*, 92(1), 234-250.
- Kaya, E. (2022). *Beyin bilgisayar arayüzü uygulamalarında motor görüntüleme EEG sinyallerinin analizi için yeni yaklaşımlar*. Doktora Tezi, Selçuk Üniversitesi Fen Bilimleri Enstitüsü, Konya, 13-29.
- Kaya, Y. ve Pehlivan, H. (2014). *KNN, NN, BAYES, DT ve SVM kullanılarak EKG vurularının sınıflandırılması*. 2nd International Symposium on Innovative Technologies In Engineering and Science, Karabük, 1413-1423.
- Ke, Y., Qi, H., Zhang, L., Chen, S., Jiao, X., Zhou, P., Zhao, X., Wan, B. ve Ming, D. (2015). Towards an effective cross-task mental workload recognition model using electroencephalography based on feature selection and support vector machine regression. *International Journal of Psychophysiology*, 98(2), 157-166.
- Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q. ve Liu, T. (2017). *LightGBM: a highly efficient gradient boosting decision tree*. 31st International Conference on Neural Information Processing Systems, New York, 3149-3157.
- Kirchner, W. K. (1958). Age differences in short-term retention of rapidly changing information. *Journal of Experimental Psychology*, 55(4), 352-358.
- Kirov, V. N., Warsawskaya, L. V. ve Voynov, V. B. (1996). EEG after prolonged mental activity. *International Journal of Neuroscience*, 85, 31-43.
- Klimesch, W. (1999). EEG Alfa and teta oscillations reflect cognitive and memory performance: A review and analysis. *Brain Research Reviews*, 29(2-3), 169-195.
- Kohl, M. (2012). Performance measures in binary classification. *International Journal of Statistics in Medical Research*, 1, 79-81.

- Korkmaz, S.G. (2022). *Bilişsel yorgunluğun EEG sinyallerinin gama bandından tespiti*. Yüksek Lisans Tezi, Karadeniz Teknik Üniversitesi Fen Bilimleri Enstitüsü, Trabzon, 1-85.
- Kosti, M.V., Georgiadis, K., Adamos, D.A., Laskaris, N., Spinellis, D. ve Angelis, L. (2018). Towards an affordable brain computer interface for the assessment of programmers' mental workload. *International Journal of Human-Computer Studies*, 115, 52-66.
- Kumar, V., Minz, S. (2014). Feature selection: a literature review. *Smart Computing Review*, 4(3), 211-229.
- Le, A.S., Aoki, H., Murase, F. ve Ishida, K. (2018). A novel method for classifying driver mental workload under naturalistic conditions with information from near-infrared spectroscopy. *Frontiers in Human Neuroscience*, 12, 1-12.
- Lal S. K. ve Craig A. (2002). Driver fatigue: electroencephalography and psychological assessment. *Psychophysiology*, 39(3), 313-321.
- Li, Q., Ng, K.K.H, Yu, S.C.M., Yiu, C.Y. ve Lyu, M. (2023). Recognising situation awareness associated with different workloads using EEG and eye-tracking features in air traffic control tasks. *Knowledge-Based Systems*, 260, 1-16.
- Lim, W., Sourina, O. ve Wang, L. (2018). STEW: simultaneous task EEG workload dataset. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 99, 1-9.
- Lin, T., Xie, T., Chen, Y. ve Tang, N. (2013). Automatic cognitive load evaluation using writing features: An exploratory study. *International Journal of Industrial Ergonomics*, 43, 210-217.
- Liu, Y., Ayaz, H. ve Shewokis, P. A. (2017). Multisubject "learning" for mental workload classification using concurrent EEG, fNIRS, and physiological measures. *Frontiers in Human Neuroscience*, 11(389), 1-14.
- Ma, J., McCrory, B. ve Claudio, D. (2021). *Comparison of weighted and unweighted task load indices for single-site surgery tasks*. International Symposium on Human Factors and Ergonomics in Health Care, Sanal Konferans, 142-146.
- Mackworth, J. F. (1959). Paced memorizing in a continuous task. *Journal of Experimental Psychology*, 58(3), 206-211.
- Makeig, S. (1993). In low lapses in alertness: coherence of fluctuations in performance and EEG spectrum. *Electroencephalography and Clinical Neurophysiology*, 86, 23-35.
- Makeig, S. ve Jung, T. (1995). Changes in alertness are a principal component of variance in the EEG spectrum. *Neuroreport*, 7, 213-216.
- Marchitto, M., Benedetto, S., Baccino, T. ve Canas, J. (2016). Air traffic control: ocular

- metrics reflect cognitive complexity. *International Journal of Industrial Ergonomics*, 54, 120-130.
- Marquart, G., Cabrall, C.D. ve Winter, J.D. (2015). Review of eye-related measures of drivers' mental workload. *Procedia Manufacturing*, 3, 2854-2861.
- Matousek, M.I. ve Petersen, A. (1983). A method for assessing alertness fluctuations from EEG spectra. *Electroencephalography and Clinical Neurophysiology*, 55, 108-113.
- Menon, V., Rivera, S.M., White, C.D., Glover, G.H. ve Reiss, A.L. (2000). Dissociating pre-frontal and parietal cortex activation during arithmetic processing. *Neuroimage*, 12(4), 357-365.
- Mitchell, D.K. (2000). Mental workload and ARL workload modeling tools. *ARL Report*. Aberdeen, 1-20.
- Monod, H. ve Kapitaniak, B. (1999). *Ergonomie*. Paris: Masson Yayınevi.
- Morrel, L.K. (1966). EEG frequency and reaction time - a sequential analysis. *Neuropsychologia*, 4, 41-48.
- Nakayama, M., Takahashi, K. ve Shimizu, Y. (2002). *The act of task difficulty and eye-movement frequency for the oculo-motor indices*. Symposium on Eye Tracking Research & Applications, New Orleans, 37-42.
- Naskrent, B., Grzywinski, W., Polowy, K., Tomczak, A. ve Jelonek, T. (2022). Eye-tracking in assessment of the mental workload of harvester operators. *International Journal of Environmental Research and Public Health*, 19, 1-11.
- Naveed, S., Sikander, B. ve Khiyal, M. (2012). Eye tracking system with blink detection. *Journal of Computing*, 4, 50-60.
- Nielsen, J. ve Landauer, T. K. (1993). *A mathematical model of the finding of usability problems*. Conference on Human Factors in Computing Systems, Amsterdam, 206-213.
- Norman, G. (2013). Working memory and mental workload. *Advances in Health Sciences Education*, 18, 163-165.
- Okogbaa, O.G., Shell, R.L., Filipusic, D. (1994). On the investigation of the neuro-physiological correlates of knowledge worker mental fatigue using the EEG signal. *Applied Ergonomics*, 25(6), 355-365.
- Onan, A. ve Korukoğlu, S. (2016). Makine öğrenmesi yöntemlerinin görüş madenciliğinde kullanılması üzerine bir literatür araştırması. *Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi*, 22(2), 111-122.
- Orlandi, L. ve Brooks, B. (2018). Measuring mental workload and physiological reactions in marine pilots: Building bridges towards redlines of performance. *Applied Ergonomics*, 69, 74-92.

- Özçakır, F.C. ve Çamurcu, A.Y. (2007). Birlikte kuralı yöntemi için bir veri madenciliği yazılımı tasarımı ve uygulaması. *İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi*, 6(12), 21-37.
- Özekes, S. (2003). Veri madenciliği modelleri ve uygulama alanları. *İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi*, 2(3), 65-82.
- Öztürk, A. (2018). *Transfer and maintenance effects of n-back working memory training in interpreting students: A behavioural and optical brain imaging study*. Doktora Tezi, Orta Doğu Teknik Üniversitesi Enformatik Enstitüsü, Ankara, 1-31.
- Pan, B., Hembrooke, H., Gay, G., Granka, L.A., Feusner, M.K. ve Newman, J.K. (2004). The determinants of web page viewing behavior: an eye-tracking study. *Eye Tracking Research & Application*, 1(212), 147-154.
- Pei, Z., Wang, H., Bezerianos, A. ve Li, J. (2021). EEG-based multiclass workload identification using feature fusion and selection. *IEEE Transactions on Instrumentation and Measurement*, 70, 1-8.
- Plechawska-Wojcik, M., Tokovarov, M., Kaczorowska, M. ve Zapala, D. (2019). A three-class classification of cognitive workload based on EEG spectral data. *Applied Sciences*, 9(24), 1-15.
- Poole, A. ve Ball, L. (2006). *Encyclopedia of human-computer interaction*. (2. Baskı). Michigan: Idea Group Reference, 45-53.
- Prechelt, L. (2012). *Neural networks: tricks of the trade: Early stopping - but when?* (2. Baskı). New York: Springer, 53-67.
- Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A.V. ve Gulin, A. (2018). *CatBoost: unbiased boosting with categorical features*. 32nd International Conference on Neural Information Processing Systems, New York, 6639-6649.
- Qu, H., Shan, Y., Liu, Y., Pang, L., Fan, Z., Zhang, J. ve Wanyan, X. (2020). Mental workload classification method based on EEG independent component features. *Applied Sciences*, 10(9), 1-15.
- Radhakrishnan, V., Louw, T., Gonçalves, R., Torrao, G., Lenné, M. ve Merat, N. (2023). Using pupillometry and gaze-based metrics for understanding drivers' mental workload during automated driving. *Transportation Research Part F Traffic Psychology and Behaviour*, 94, 254-267.
- Rashid, A. ve Qureshi, I. (2015). Eliminating electroencephalogram artefacts using independent component analysis. *International Journal of Applied Mathematics Electronics and Computers*, 3(1), 48-52.
- Rauffet, P., Chauvin, C., Nistico, C., Judas, S. ve Norbert, T. (2016). Analysis of submarine steering: effects of cognitive and perceptual-motor requirements on the mental workload and performance of helmsmen. *Cognition, Technology and Work*, 18, 657-672.

- Recarte, M. A. ve Nunes, L. M. (2000). Effects of verbal and spatial-imagery task on eye fixations while driving. *Journal of Experimental Psychology: Applied*, 6, 31-43.
- Reiner, M. ve Gelfeld, T.M. (2014). Estimating mental workload through event-related fluctuations of pupil area during a task in a virtual world. *International Journal of Psychophysiology*, 93(1), 38-44.
- Rubio, S., Diaz, E., Martin, J. ve Puente, J. M. (2004). Evaluation of subjective mental workload : a comparison of SWAT , NASA-TLX , and workload profile methods. *Applied Psychology: An International Review*, 53(1), 61-86.
- Rusnock, C., Borghetti, B. ve McQuaid, I. (2015). Objective-analytical measures of workload - the third pillar of workload triangulation?. *Foundations Augmented Cognition*, 9183, 124-135.
- Rusnock, C. ve Borghetti, B. (2018). Workload profiles: A continuous measure of mental workload. *International Journal of Industrial Ergonomics*, 63, 49-64.
- Sareen, E., Singh, L., Varkey, B., Achary, K. ve Gupta, A. (2020). EEG dataset of individuals with intellectual and developmental disorder and healthy controls under rest and music stimuli. *Data in Brief*, 30(1), 1-9.
- Sassaroli, A., Zheng, F., Hirshfield, L., Girouard, A., Solovey, E., Jacob, R. ve Fantini, S. (2008). Discrimination of mental workload levels in human subjects with functional near-infrared spectroscopy. *Journal of Innovative Optical Health Sciences*, 31, 227-237.
- Savaş, S., Topaloğlu, N. ve Yılmaz, M. (2012). Veri madenciliği ve Türkiye'deki uygulama örnekleri. *İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi*, 11(21), 1-23.
- Schuh, G., Reinhart, G., Prote, J.P., Sauermann, F., Horsthofer, J., Oppolzer, F. ve Knoll, D. (2019). Data mining definitions and applications for the management of production complexity, *Procedia CIRP*, 81, 874-879.
- Selvi, A. O. (2022). *Makine öğrenmesi algoritmaları ile engellilere yönelik eeg işaretlerinin sınıflandırılması*. Doktora Tezi, Sakarya Uygulamalı Bilimler Üniversitesi Lisansüstü Eğitim Enstitüsü, Sakarya, 6-53.
- Shahid, U. ve Rasool, S. (2020). *EEG based mental workload assessment using machine learning*. Yüksek Lisans Tezi, National University of Sciences and Technology, Islamabad, 1-10.
- Siddiqui K. (2013). Heuristics for sample size determination in multivariate statistical techniques. *World Applied Sciences Journal*, 27(2), 285-287.
- Smith, A., Borghetti, B. ve Rusnock, C. (2015). *Improving model cross-applicability for operator workload estimation*. 59th Human Factors and Ergonomics Society Annual Meeting, Los Angeles, 681-685.

- Subasi A. (2005). Automatic recognition of alertness level from EEG by using neural network and wavelet coefficients. *Expert Systems with Applications*, 28(4), 701-711.
- Sweller, J. (1988). Cognitive load during problem solving: effects on learning. *Cognitive Science*, 12(2), 257-285.
- Şaşmaz, S., Saraoğlu, H., Kabay, S., Akdağ, G., Keskinçiliç, C. ve Tosun, M. (2023). EEG-based mental workload estimation of multiple sclerosis patients. *Signal, Image and Video Processing*, 17(7), 1-9.
- Tibshirani, R. (1997). The lasso method for variable selection in the cox model. *Statistics in medicine*, 16(4), 385-395.
- Tjolleng, A., Jung, K., Hong, W., Lee, W., Lee, B., You, H., Son, H. ve Park, S. (2017). Classification of a Driver's cognitive workload levels using artificial neural network on ECG signals. *Applied Ergonomics*, 59, 326-332.
- Tran, C., Yan, S., Habiyaemye, J. ve Wei, Y. (2017). *Predicting driver's work performance in driving simulator based on physiological indices*. Intelligent Human Computer Interaction International Conference, Evry, 150-162.
- Uysal, İ. ve Kılıç, A. F. (2022). Normal distribution dilemma. *Anadolu Journal of Educational Sciences International*, 12(1), 220-248.
- Üstüner, M., Abdikan, S., Bilgin, G. ve Şanlı, F. (2020). Hafif gradyan artırma makineleri ile tarımsal ürünlerin sınıflandırılması. *Türk Uzaktan Algılama ve CBS Dergisi*, 1(2), 97-105.
- Van Orden, K.F., Jung, T.P. ve Makeig, S. (2000). Combined eye activity measures accurately estimate changes in sustained visual task performance. *Biological Psychology*, 52, 221-240.
- Van Orden, K.F., Limbert, W., Makeig, S. ve Jung, T.P. (2001). Eye activity correlates of workload during a visuospatial memory task. *Human Factors*, 43(1), 111-121.
- Vapnik, N.V. ve Chervonenkis, A.Y. (1991). The necessary and sufficient conditions for consistency in the empirical risk minimization method. *Pattern Recognition and Image Analysis*, 1(3), 283-305.
- Vieira, S., Pinaya, W.H.L., Mechelli, A. (2020). *Machine learning*. Cambridge: Academic Press, 1-20.
- Voyant, C., Notton, G., Kalogirou, S., Nivet, M. L., Paoli, C., Motte, F. ve Fouilloy, A. (2017). Machine learning methods for solar radiation forecasting: A review. *Renewable Energy*, 105, 569-582.
- Walter, C., Wolter, P., Rosenstiel, W., Bogdan, M. ve Spüler, M. (2014). *Towards cross-subject workload prediction*. 6th International Brain-Computer Interface Conference, Graz, 1-4.

- Wan, F., Nan, W., Vai, M.I., Rosa, A. (2014). Resting alfa activity predicts learning ability in alfa neurofeedback. *Frontiers in Human Neuroscience*, 8(500), 1-7.
- Wang, C. ve Guo, J. (2019). A data-driven framework for learners' cognitive load detection using ECG-PPG physiological feature fusion and XGBoost classification. *Procedia Computer Science*, 147, 338-348.
- Wang, Z., Hope, R., Wang, Z., Ji, Q. ve Gray, W. (2012). Cross-subject workload classification with a hierarchical bayes model, *NeuroImage*, 59(1), 64-69.
- Wanyan, X., Zhuang, D., Lin, Y., Xiao, X. ve Song, J.W. (2018). Influence of mental workload on detecting information varieties revealed by mismatch negativity during flight simulation. *International Journal of Industrial Ergonomics*, 64, 1-7.
- Wilson, G.F. (2002). An analysis of mental workload in pilots duringflight using multiple psychophysiological measures. *The International Journal of Aviation Psychology*, 12(1), 3-18.
- Wilson, G., Schlegel, R. ve Veltman, J. (2004). Operator functional state assessment. *NATO RTO*, Brüksel, 1-220.
- Wu, Y., Liu, Z., Jia, M., Congchi, T. ve Yan, S. (2019). Using artificial neural networks for predicting mental workload in nuclear power plants based on eye tracking. *Nuclear Technology*, 206, 1-13.
- Yan, S., Tran, C., Chen, Y., Tan, K., Habiyaemye, J.L. (2017). Effect of user interface layout on the operators' mental workload in emergency operating procedures in nuclear power plants. *Nuclear Engineering and Design*, 322, 266-276.
- Yan, S., Wei, Y. ve Tran, C.C. (2019). Evaluation and prediction mental workload in user interface of maritime operations using eye response. *International Journal of Industrial Ergonomics*, 71, 117-127.
- Yang, C., Yang, L., Cheng, T., Jou, Y. ve Chiou, S. (2012). Assessing mental workload and situation awareness in the evaluation of computerized procedures in the main control room. *Nuclear Engineering and Design*, 250, 713-719.
- Yin, Z. ve Zhang, J. (2017). Cross-session classification of mental workload levels using EEG and an adaptive deep learning model. *Biomedical Signal Processing and Control*, 33, 30-47.
- Zammouri, A., Moussa, A.A. ve Mebrouk, Y. (2018). Brain-computer interface for workload estimation: Assessment of mental efforts in learning processes. *Expert Systems with Applications*, 112, 138-147.
- Zarjam, P., Epps, J. ve Lovell, N. (2015). Beyond subjective self-rating: EEG signal classification of cognitive workload. *IEEE Transactions on Autonomous Mental Development*, 7(4), 1-11.
- Zeng, H., Yang, C., Zhang, H., Wu, Z., Zhang, J., Dai, G., Babiloni, F. ve Kong, W.

- (2019). A LightGBM-based EEG analysis method for driver mental states classification. *Computational Intelligence and Neuroscience*, 2019(4), 1-11.
- Zou, H. ve Hastie, T. (2005). Regularization and variable selection via the elasticnet. *Journal of the Royal Statistical Society Series B Statistical Methodology*, 67, 301–320.
- Zhou, Y., Xu, Z., Niu, Y., Wang, P., Wen, X., Wu, X. ve Zhang, D. (2022). Cross-task cognitive workload recognition based on EEG and domain adaptation. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 30, 50-60.

EKLER

EK-1. Katılımcılar için bilgilendirilmiş gönüllü olur formu



T.C.
GAZİ ÜNİVERSİTESİ

KATILIMCILAR İÇİN BİLGİLENDİRİLMİŞ GÖNÜLLÜ OLUR FORMU

Sizi, Gazi Üniversitesi Etik Komisyonu'ndan 08/09/2020 tarih ve 8 sayılı ile izin alınan ve Doç. Dr. Erman ÇAKIT (Tez danışmanı) ile Şeniz HARPUTLU AKSU (Doktora öğrencisi) tarafından yürütülen “Zihinsel İş Yükünün Analizine İlişkin Bir Model Önerisi” başlıklı araştırmaya davet ediyoruz. Bu çalışmaya katılmak tamamen gönüllülük esasına dayanmaktadır. Çalışmaya katılmama veya katıldıktan sonra herhangi bir anda çalışmadan çıkma hakkına sahipsiniz. Bu çalışmaya katılmanız için sizden herhangi bir ücret istenmeyecektir. Çalışmaya katıldığınız için size bir ödeme yapılmayacaktır. Çalışmadan elde edilecek bilgiler tamamen araştırma amacı ile kullanılacak olup kişisel bilgileriniz gizli tutulacaktır.

Araştırmanın Amacı	Bu proje kapsamında bir zihinsel iş yükü tahmin modeli geliştirilmesi amaçlanmaktadır. Katılımcılardan alınan göz izleme ve EEG ölçüm verileri modelin girdi parametrelerini oluşturacaktır. Çıktı olarak ise zihinsel iş yükü seviyeleri tahmin edilecektir. Makine öğrenme algoritmaları kullanılarak elde edilecek tahmin modelinin doğruluğu ve geçerliliği istatistiksel yöntemler ile test edilecektir. Bu sayede göz izleme ve EEG verilerinin zihinsel iş yükünü belirlemedeki tercih edilebilirliği ve tahmin modelinin en doğru şekilde hangi algoritma ile elde edilebileceği ortaya konacaktır. EEG yöntemi ile birlikte göz izleme teknikleri de kullanılarak elde edilen veri ile farklı görev koşulları altında genelleştirilebilir bir model geliştirilmesi hedeflenmektedir.
Araştırmanın Yöntemi	Deneylerde laboratuvar ortamında katılımcılara zihinsel iş yüküne sebep olması beklenen farklı zorluk derecelerine sahip, n-geri görev, tetikte olma görevi, rakam ve sembol eşleştirme testi ve işitsel hafıza testi gibi testler uygulanacaktır. Testler bilgisayar ortamında uygulanacak olup, eş zamanlı olarak taşınabilir EEG cihazı ve göz izleme cihazı sayesinde beyin görüntüleme verileri ve göz takip verileri kayıt altına alınacaktır. Diğer yandan doğru cevap sayısı, tepki süresi gibi performans göstergeleri de kayıt altına alınacak ve her testin sonunda öznel zihinsel iş yükü skorlarının hesaplanması amacıyla katılımcılara NASA-TLX anketi uygulanacaktır. Elde edilen veriler istatistiksel yazılımlardan (SPSS vb.) yararlanılarak analiz edilecektir. Makine öğrenmesi algoritmalarının ise PYTHON gibi yazılımlar aracılığıyla uygulanması ve zihinsel iş yükü seviyesi tahmin modelinin bu şekilde elde edilmesi planlanmaktadır.
Araştırmanın Öngörülen Süresi (Başlama ve Bitiş Tarihi)	Başlama: Ekim 2020 Bitiş: Haziran 2023
Araştırmaya Katılması Beklenen Katılımcı/Gönüllü Sayısı	15 (Gerekli görüldüğü takdirde artırılacaktır.)
Araştırmanın Yapılacağı Yerler	Orta Doğu Teknik Üniversitesi İnsan-Bilgisayar Etkileşimi Araştırma ve Uygulama Laboratuvarı Gazi Üniversitesi Ergonomi Laboratuvarı
Görüntü ve/veya ses kaydı alınacak mı?	Evet ☐ Hayır ☒

Tablo katılımcıların anlayabileceği biçimde, akademik dil kullanılmadan yazılacaktır.

EK-1. (devam) Katılımcılar için bilgilendirilmiş gönüllü olur formu

KATILIMCI BEYANI

Yukarıda amacı ve içeriği belirtilen bu araştırmaya ile ilgili bilgiler tarafıma aktarıldı. Bu bilgilerden sonra araştırmaya katılımcı olarak davet edildim. Bu çalışmaya katılmayı kabul ettiğim takdirde gerek araştırmaya yürütülürken gerekse yayımlandığında kimliğimin gizli tutulacağı konusunda güvence aldım. Bana ait verilerin kullanımına izin veriyorum. Araştırma sonuçlarının eğitim ve bilimsel amaçlarla kullanımı sırasında kişisel bilgilerimin dikkatle korunacağı konusunda bana yeterli güven verildi. Araştırmanın yürütülmesi sırasında herhangi bir sebep göstermeden çekilebilirim. Araştırma için yapılacak harcamalarla ilgili herhangi bir parasal sorumluluk altına girmiyorum. Bana herhangi bir ödeme yapılamayacaktır. Araştırma ile ilgili bana yapılan tüm açıklamaları ayrıntılarıyla anlamış bulunmaktayım. Bu çalışmaya hiçbir baskı altında kalmadan kendi bireysel onayım ile katılıyorum. İmzalı bu form kağıdının bir kopyası bana verilecektir.

Araştırma yürütücüsü (Tez çalışmalarında Danışman tarafından imzalanacaktır.)

Adı ve Soyadı	Prof. Dr. Erman ÇAKIT	Tarih ve İmza
Adres ve telefonu	Gazi Üniversitesi Mühendislik Fakültesi Endüstri Mühendisliği Bölümü 0 (312) 582 38 20	

Katılımcı

Adı ve Soyadı		Tarih ve İmza
Adres ve telefonu		

Velayet veya Vesayet Altındaki Katılımcılar için Veli/Vasi

Adı ve Soyadı		Tarih ve İmza
Adres ve telefonu		

EK-2. Demografik bilgiler formu

Cinsiyet	() Kadın () Erkek
Yaş	Belirtiniz.....
Boy	Belirtiniz.....
Kilo	Belirtiniz.....
Kaçıncı sınıftasınız?	Belirtiniz.....
Daha önce hafıza testleri ile ilgili herhangi bir uygulama kullandınız mı?	() Evet () Hayır
Hangi elinizi kullanıyorsunuz?	() Sağ () Sol
Günde kaç saat uyuyorsunuz?	Belirtiniz.....
Sigara kullanıyor musunuz?	() Evet () Hayır
Son 12 saat içinde alkol aldınız mı?	() Evet () Hayır
Son 12 saat içinde kafein aldınız mı?	() Evet () Hayır
Son 12 saat içinde herhangi bir ilaç kullandınız mı?	() Evet () Hayır Evet ise belirtiniz.....
Herhangi bir kronik rahatsızlığınız var mı?	() Evet () Hayır Evet ise belirtiniz.....
Herhangi bir görme probleminiz var mı?	() Evet () Hayır Evet ise belirtiniz.....
Gözlük/kontakt lens kullanıyor musunuz?	() Evet () Hayır Evet ise belirtiniz.....
Herhangi bir işitme probleminiz var mı?	() Evet () Hayır Evet ise belirtiniz.....
Herhangi bir psikolojik/psikiyatrik rahatsızlığınız var mı?	() Evet () Hayır Evet ise belirtiniz.....
Herhangi bir nörolojik rahatsızlığınız var mı?	() Evet () Hayır Evet ise belirtiniz.....

EK-3. NASA-TLX ölçeği ve ikili karşılaştırma anketi

Aşağıdaki sorulara verdiğiniz cevapları ölçek üzerindeki iki çizgi arasına çarpı (X) koyarak işaretleyiniz.

Bu test zihinsel (bakma, takip etme, düşünme, akılda tutma, hatırlamaya çalışma vb.) açıdan ne derece zorlayıcıydı? (Kolay mıydı/zorlu muydu?, Basit miydi/karmaşık mıydı?)

Çok Düşük  Çok Yüksek

Bu test fiziksel olarak ne derece zorlayıcıydı?

Çok Düşük  Çok Yüksek

Bu testi ne ölçüde telaşlı ve aceleci bir şekilde yapmak zorunda kaldınız? (Üzerinizdeki zaman baskısını değerlendiriniz.)

Çok Düşük  Çok Yüksek

Bu testi ne ölçüde başarıyla yerine getirdiniz? Performansınızdan ne ölçüde tatmin oldunuz?

Çok Düşük  Çok Yüksek

Bu testi başarıyla yerine getirebilmek için ne ölçüde çaba sarf etmeniz gerekti?

Çok Düşük  Çok Yüksek

Bu testi yaparken ne ölçüde stresli, gergin, isteksiz ve keyifsiz hissettiniz?

Çok Düşük  Çok Yüksek

EK-3. (devam) NASA-TLX ölçeği ve ikili karşılaştırma anketi

Bu testi gerçekleştirirken sizin için daha çok ön plana çıkan faktörü işaretleyiniz. (İkili karşılaştırma yapmanız beklenmektedir. Yan yana yazılan iki faktörden öncelikli olanı seçiniz.)

0-geri

çaba		performans	
zaman baskısı		gerginlik	
zaman baskısı		çaba	
fiziksel talep		gerginlik	
performans		gerginlik	
fiziksel talep		zaman baskısı	
fiziksel talep		performans	
zaman baskısı		zihinsel talep	
gerginlik		çaba	
performans		zihinsel talep	
performans		zaman baskısı	
zihinsel talep		çaba	
zihinsel talep		fiziksel talep	
çaba		fiziksel talep	
gerginlik		zihinsel talep	

1-geri

çaba		performans	
zaman baskısı		gerginlik	
zaman baskısı		çaba	
fiziksel talep		gerginlik	
performans		gerginlik	
fiziksel talep		zaman baskısı	
fiziksel talep		performans	
zaman baskısı		zihinsel talep	
gerginlik		çaba	
performans		zihinsel talep	
performans		zaman baskısı	
zihinsel talep		çaba	
zihinsel talep		fiziksel talep	
çaba		fiziksel talep	
gerginlik		zihinsel talep	

2-geri

çaba		performans	
zaman baskısı		gerginlik	
zaman baskısı		çaba	
fiziksel talep		gerginlik	
performans		gerginlik	
fiziksel talep		zaman baskısı	
fiziksel talep		performans	
zaman baskısı		zihinsel talep	
gerginlik		çaba	
performans		zihinsel talep	
performans		zaman baskısı	
zihinsel talep		çaba	
zihinsel talep		fiziksel talep	
çaba		fiziksel talep	
gerginlik		zihinsel talep	

3-geri

çaba		performans	
zaman baskısı		gerginlik	
zaman baskısı		çaba	
fiziksel talep		gerginlik	
performans		gerginlik	
fiziksel talep		zaman baskısı	
fiziksel talep		performans	
zaman baskısı		zihinsel talep	
gerginlik		çaba	
performans		zihinsel talep	
performans		zaman baskısı	
zihinsel talep		çaba	
zihinsel talep		fiziksel talep	
çaba		fiziksel talep	
gerginlik		zihinsel talep	



Gazili olmak ayrıcalıktır