



## A machine learning approach to classify mental workload based on eye tracking data

Şeniz Harputlu Aksu<sup>ID</sup>, Erman Çakıt\*<sup>ID</sup>

Department of Industrial Engineering, Faculty of Engineering, Gazi University, 06570, Maltepe, Ankara, Türkiye

### Highlights:

- Collecting eye tracking data and obtaining variables
- Classification of mental workload at 4 different levels
- Usage of GBM, LightGBM, XGBoost algorithms

### Keywords:

- Mental workload
- Eye tracking
- Machine learning
- Classification

### Article Info:

Research Article

Received: 28.12.2021

Accepted: 01.05.2022

### DOI:

10.17341/gazimmfd.1049979

### Acknowledgement:

This work was funded by Gazi University Scientific Research Projects Unit under Grant No. 6483.

### Correspondence:

Author: Erman Çakıt  
e-mail: ecakit@gazi.edu.tr  
phone: +90 312 582 3820

### Graphical/Tabular Abstract

In this study, we tested whether the change in mental workload during N-Back memory tasks is related to eye movements, compared it with the results obtained previously in the literature, and developed a classification model that predicts the level of mental workload with machine learning algorithms using eye tracking data. The steps of the study are given in Figure A.

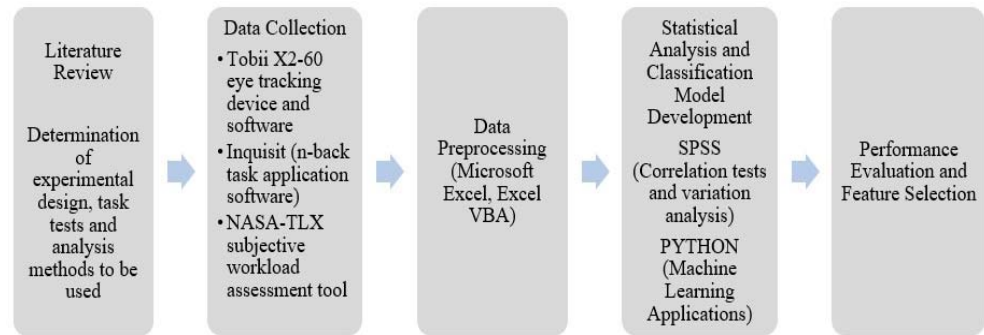


Figure A. Study Steps

### Purpose:

The primary objective of this study was to develop machine learning algorithms for classifying mental workload using eye tracking data.

### Theory and Methods:

This study proposed machine learning algorithms for classifying mental workload using eye tracking data. Four participants performed the N-Back memory task and National Aeronautics and Space Administration-task load index (NASA-TLX) to induce different levels of mental workload. In total, 792 samples were collected from four participants. Twenty-seven eye tracking metrics were selected as independent variables. One output variable reflecting the difficulty level of N-Back memory was classified.

### Results:

As a result of these experiments, it was revealed that almost all eye tracking metrics considered in this study were significantly correlated to both weighted NASA-TLX total score and N-Back memory task difficulty level. As the task difficulty increased, pupil diameter, number of saccades, number of blinks, and blink duration increased, while fixation duration decreased. The results obtained for the two classes of classification problem reached the accuracy of 68% with 14 eye-tracking features due to problem complexity. The results obtained for the two classes of classification problem reached the accuracy of 84% with 27 eye-tracking features as input and the LightGBM algorithm. To determine the degree to which the input variables contribute to the determination of the output variable, a sensitivity analysis was conducted using the gradient boosting machines (GBM) algorithm. The left eye pupil diameter was found to be the most effective metric in the classification of the task difficulty level. The results from the analysis indicate that eye tracking metrics play an important role in the classification of mental workload.

### Conclusion:

The results from the analysis indicate that eye tracking metrics play an important role in the classification of mental workload. Among the classification models developed with machine learning algorithms, the most successful results were obtained with tree-based algorithms. Algorithms such as GBM, LightGBM, XGBoost, which have been developed recently and therefore are not frequently used in the literature on this subject, have achieved better results compared to others.



## Göz izleme verilerine bağlı olarak zihinsel iş yükünü sınıflandırmada makine öğrenmesi algoritmalarının kullanılması

Şeniz Harputlu Aksu<sup>ID</sup>, Erman Çakıt\*<sup>ID</sup>

Gazi Üniversitesi, Mühendislik Fakültesi, Endüstri Mühendisliği Bölümü, 06570 Maltepe Ankara, Türkiye

### Ö N E Ç I K A N L A R

- Göz izleme verilerinin toplanması ve değişkenlerin elde edilmesi
- Zihinsel iş yükünün 4 farklı seviyede sınıflandırılması
- GBM, LightGBM, XGBoost algoritmalarının kullanımı

### Makale Bilgileri

Araştırma Makalesi

Geliş: 28.12.2021

Kabul: 01.05.2022

### DOI:

10.17341/gazimmfd.1049979

### Anahtar Kelimeler:

Zihinsel iş yükü,  
göz izleme,  
makine öğrenmesi,  
sınıflandırma

### ÖZ

Bu çalışmada, zihinsel iş yükünün göz izleme verilerine bağlı olarak makine öğrenmesi algoritmaları aracılığıyla sınıflandırılması amaçlanmıştır. Dört katılımcı (iki kadın ve iki erkek) ile yapılan deneyler sırasında, farklı düzeylerde zihinsel iş yükünün tahmini için n-geri hafıza görevi ve NASA-Task Load Index (TLX) öznel değerlendirme ölçeği kullanılmıştır. Bağımsız değişkenler olarak 27 göz izleme parametresi seçilmiş ve çıktı değişkeni olarak görev zorluk seviyesi sınıflandırılmıştır. Çalışmada ele alınan hemen hemen tüm göz izleme parametrelerinin hem ağırlıklı NASA-TLX toplam skoru hem de görev zorluk seviyesi ile anlamlı ölçüde ilişkili olduğu ortaya çıkmıştır. Görev zorluğu arttıkça göz bebeği boyutu, seçirme sayısı, göz kırpma sayısı ve göz kırpma süresi artarken sabitleme süresi ile ilgili değişkenlerin ise azaldığı gözlenmiştir. Dört sınıflı sınıflandırma problemi kapsamında, problemin karmaşıklığı sebebiyle (seçilen 14 öznelik ile) ancak %68 doğruluğa ulaşılabilmektedir. İki sınıflı olarak ele alınan sınıflandırma problemi için elde edilen sonuçlar incelendiğinde ise, 27 göz izleme özelliği ve LightGBM algoritması ile kurulan modelde %84 doğruluğa ulaşılmıştır. Girdi değişkenlerinin çıktı değişkeninin belirlenmesine katkısını incelemek için gradyan artırma makineleri (GBM) algoritması kullanılarak bir duyarlılık analizi yapılmış ve sol göz bebeği çapı ortalamasının görev zorluk seviyesinin sınıflandırılmasında en etkili parametre olduğu görülmüştür. Sonuçlar, göz izleme ölçümlerinin zihinsel iş yükünün sınıflandırılmasında önemli bir rol oynadığını göstermektedir.

## A machine learning approach to classify mental workload based on eye tracking data

### H I G H L I G H T S

- Collecting eye tracking data and obtaining variables
- Classification of mental workload at 4 different levels
- Usage of GBM, LightGBM, XGBoost algorithms

### Article Info

Research Article

Received: 28.12.2021

Accepted: 01.05.2022

### DOI:

10.17341/gazimmfd.1049979

### Keywords:

Mental workload,  
eye tracking,  
machine learning,  
classification

### ABSTRACT

The primary objective of this study was to develop machine learning algorithms for classifying mental workload using eye tracking data. Four participants (two females and two males) performed the N-Back memory task and National Aeronautics and Space Administration-task load index (NASA-TLX) to induce different levels of mental workload. Twenty-seven eye tracking metrics were selected as independent variables. One output variable reflecting the difficulty level of N-Back memory was classified As a result of these experiments, it was revealed that almost all eye tracking metrics considered in this study were significantly correlated to both weighted NASA-TLX total score and N-Back memory task difficulty level. As the task difficulty increased, pupil diameter, number of saccades, number of blinks, and blink duration increased, while fixation duration decreased. The results obtained for the two classes of classification problem reached the accuracy of 68% with 14 eye-tracking features due to problem complexity. The results obtained for the two classes of classification problem reached the accuracy of 84% with 27 eye-tracking features as input and the LightGBM algorithm. To determine the degree to which the input variables contribute to the determination of the output variable, a sensitivity analysis was conducted using the gradient boosting machines (GBM) algorithm. The left eye pupil diameter was found to be the most effective metric in the classification of the task difficulty level. The results from the analysis indicate that eye tracking metrics play an important role in the classification of mental workload.

## 1. Giriş (Introduction)

Zihinsel iş yükü operatörün kullanılabilir zihinsel kaynak kapasitesi ile işin gerektirdiği zihinsel kaynak arasındaki fark ile ilişkilidir. Şayet görevin gerektirdiği kaynak miktarı kişinin kullanılabilir kapasitesini aşarsa aşırı yüklenme olacaktır [1]. Zihindeki aşırı yüklenme, bireyin fazladan zihinsel kaynak tahsis etmesini gerektirir. Ekstra kaynaklara olan bu talebin işlem verimliliğini azaltabileceği düşünülmektedir [2]. Ayrıca aşırı zihinsel iş yükü düzeyleri, hatalara veya bilgi işlemede gecikmeye neden olabileceğinden, mesleki bir görevin tasarım ve değerlendirmesi mutlaka zihinsel iş yükünün değerlendirilmesini de içermelidir [3]. Bu sebeple, ergonomi alanıyla uğraşanlar insan makine sistemlerinin etkinliğini değerlendirmede, performansı arttırmada ve farklı tasarımlar arasından seçim yapmada zihinsel iş yükü ölçümünden yararlanmaktadır.

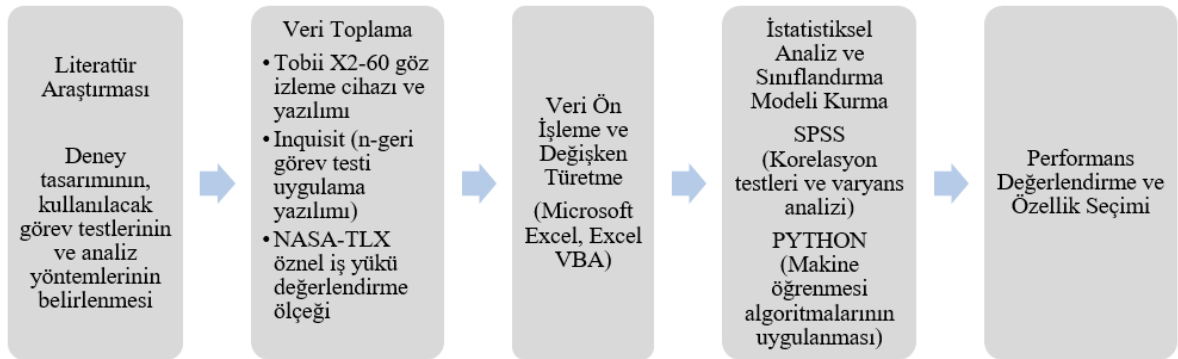
Operatörün zihinsel iş yükünü bilmek, ona atanacak görev sayısına ve zorluk seviyelerine karar vermek ayrıca görev performansını yüksek tutmak açısından önemlidir. Etkin bir görev dağılımı ile uyarlanabilir bir otomasyon için operatör zihinsel iş yükünü doğru ve etkin olarak ölçen araçlara ihtiyaç vardır [4]. Zihinsel iş yükünün insan performansını ve dolayısıyla insan-makine sistemlerinin etkinliğini doğrudan etkilediği düşünüldüğünde, zihinsel iş yükü ölçümünün kritik görevler başta olmak üzere askeriye [4], havacılık [5], bilgisayar [6] ve araç kullanımı [7] gibi pek çok alanda önemle üzerinde durulan bir çalışma alanı olduğu görülmektedir. Diğer yandan 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.

Nesnel olması, katılımcının algısına bağlı olmaması ve sürekli sinyaller sayesinde gerçek zamanlı olarak kullanılabilmesi sebebiyle insan-makine etkileşimi sırasında zihinsel iş yükünü araştırmak için sıklıkla fizyolojik verilerden yararlanılmaktadır [8]. Özellikle göz aktivitesinden her zaman ve her yerde yararlanılabilmesi, tek bir sensör ile gözle ilgili farklı bilgi türleri (göz bebeği çapı, göz kırpması, sabitleme ve seçirme) elde edilebilmesi, göz aktivitesine ilişkin verilerin diğer fizyolojik sinyallerin toplanmasına kıyasla daha az görev kesintisine sebep olması, zihinsel iş yükünün tahmininde göz izleme tekniklerinin kullanılmasını cazip kılmaktadır [9]. Fizyolojik yöntemler ile elde edilen sürekli ve büyük boyutlu verinin analizine ihtiyaç duyulması ile birlikte, son yıllarda zihinsel iş yükünün tahmin edilmesine yönelik çalışmalarda makine öğrenmesi algoritmalarının kullanımı yaygınlaşmıştır. Literatürde bu alanda yapılan çalışmalarda en sık kullanılan algoritmalar k en yakın komşu, destek vektör makineleri, yapay sinir ağları, karar ağaçları ve rastgele ormanlardır.

Bu çalışma kapsamında, n-geri hafıza görevleri sırasında meydana gelen zihinsel iş yükündeki değişimin göz hareketleri ile ilişkili olup olmadığının test edilmesi, daha önce literatürde elde edilen sonuçlar ile karşılaştırılması ve göz izleme verilerinden yararlanarak makine öğrenmesi algoritmaları ile zihinsel iş yükü seviyesini tahmin eden bir sınıflandırma modeli geliştirilmesi hedeflenmiştir. Makine öğrenmesinde sınıflandırma, yeni bir gözlemin bir kategori kümesinden hangisine ait olduğunun, kategorisi belli gözlemlerden oluşan eğitim seti kullanılarak belirlenmesi problemidir. Literatürde zihinsel iş yükü düzeyini belirlemenin sınıflandırma problemi olarak ele alındığı diğer çalışmalarda sınıf sayısı arttıkça problemin zorlaştığı, dolayısıyla model performansının düştüğü görülmektedir [10]. Bu çalışmada, 4 zorluk seviyesine göre zihinsel iş yükünü mümkün olduğunca yüksek doğruluk oranıyla tahmin eden model elde edilmeye çalışılmıştır. Öncelikle uygulanan testlerdeki görev zorluk seviyelerinin beklenen zihinsel iş yükü değişimine yol açıp açmadığı, öznel değerlendirme verileri ile karşılaştırılarak test edilmiştir. Göz izleme cihazı ile elde edilen değişkenlerin haricinde göz kırpması hareketine ilişkin değişkenler ayrı bir analiz ile üretilmiş, elde edilen ilave göz izleme değişkenleri ile modeldeki bağımlı değişken arasındaki beklenen ilişki test edildikten sonra değişkenler model kurma aşamasına dâhil edilmiştir. Kullanılan her bir makine öğrenmesi algoritması ile kurulan modellerin performansları doğruluk (accuracy) değerleri hesaplanarak değerlendirilmiştir. Diğer çalışmalardan farklı olarak; literatürde özellikle göz izleme yöntemi ile zihinsel iş yükünün tahmininde kullanımına rastlanmayan gradyan artırma makineleri (GBM), hafif gradyan artırma makineleri (LightGBM) ve aşırı gradyan artırma (XGBoost) algoritmaları çalışmaya dâhil edilmiş, model performansları incelenerek diğer algoritmalar ile karşılaştırılmıştır. En iyi hiper-parametre değerleri belirlenerek kurulan modeller düzeltilmiş, boyut azaltma yapılarak performanstaki değişim gözlemlenmiş ve sonuçlar tüm algoritmalar için karşılaştırılmıştır. Pilot uygulama olarak elde edilen bulguların ileride yapılacak çalışmalara temel oluşturması hedeflenmiştir. Çalışma adımları Şekil 1'de verilmiştir.

## 2. Literatür Araştırması (Literature Research)

Literatüre bakıldığında farklı ölçüm yöntemlerinin, doğruluk açısından karşılaştırma amaçlı ya da birbirini destekleyerek elde edilen sonuçları pekiştirme amaçlı olarak bir arada kullanıldığı çalışmalar olduğu gibi [11] yeni bir ölçüm yönteminin kullanılabilirliğinin daha önceden etkinliği kanıtlanan yöntemlerle birlikte kullanılarak test edildiği çalışmalara da rastlanmaktadır [12]. Son yıllarda NASA-TLX [13] gibi öz değerlendirme yöntemlerinin tek başına kullanıldığı yöntemlere az rastlanmaktadır. NASA-TLX'in güvenilirliğinin kanıtlanmış olması sebebiyle genellikle yeni önerilen bir yöntemin geçerliliğini ve tercih edilebilirliğini göstermek



Şekil 1. Çalışma adımları (Study Steps)

amacıyla özellikle fizyolojik yöntemlerle bir arada kullanıldığı çalışmalar mevcuttur.

### 2.1. Göz İzleme ve Zihinsel İş Yükü (Eye Tracking and Mental Workload)

Zihinsel iş yükünün analizine ilişkin çalışmalarda göz izleme tekniklerine sıklıkla başvurulduğu görülmektedir. Yapılan çalışmalarda genellikle göz bebeği boyutu, sabitleme, seçirme ve göz kırpmaya hareketlerinden faydalanılmaktadır. Sabitleme gözün belli bir noktaya odaklanmasınıdır. Seçirme ise bir pozisyonun diğerine gerçekleşen hızlı göz hareketidir. Sabitleme ve seçirme bir yere bakarken sırayla gerçekleşmektedir. Göz bebeği boyutundaki değişim, ışık refleksi ve yakın refleksi ile karşılaştırıldığında zihinsel aktivitedeki varyasyonları nispeten daha az yansıtmasına rağmen, kişinin görsel alanındaki nesnelere neredeyse sabit bir şekilde uyarılmış göz bebeği yanıtı olarak düşünülebilir. Fonksiyonel olmayan ve bilinçsizce yapılan göz kırpmaya ise, merkezi olarak kontrol edilmesi ve bilişle bağlantısı olması sebebiyle zihinsel iş yükü ölçümünde kullanılmaktadır [9].

Matematiksel hesaplama görevi verilen bir çalışmada, katılımcıların göz bebeği büyüklüğü ve göz kırpmaya oranı görev zorluk seviyesi ile birlikte artmış, seçirme oranı ve seçirme uzunluğu ise azalmıştır. Aynı çalışmada görsel hedef sıklığı arttıkça göz kırpmaya süresinin azaldığı gözlemlenmiştir. Çalışmanın sonuçlarına göre, okülomotor yani göz küresi hareketine ilişkin göstergeler hem zorluk seviyesinden hem de göz hareketlerinin sıklığından etkilenebilir [14]. Benedetto vd. araç içi bilgi sistemlerinin göz kırpmaya davranışı üzerindeki etkilerini incelemek amacıyla tekli ve ikili sürüş simülasyonu görevi koşulunda göz kırpmaya hareketlerini karşılaştırmışlardır. Sonuçlar görsel iş yükü ölçümünde göz kırpmaya süresinin duyarlı ve güvenilir bir ölçüt olduğunu göstermiştir. Diğer yandan; 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 [15]. Gao vd. tarafından yapılan çalışmada ise, göz kırpmaya oranının görev zorluk seviyesindeki değişime duyarlı olduğu ve göz kırpmaya süresinin de hem düşük hem de yüksek zorluk seviyesinde görev periyodu boyunca arttığı gözlemlenmiştir [16]. Tran vd., zihinsel iş yüküne dayalı performans tahmini elde etmeyi amaçlayan çalışmada sonuçları zihinsel iş yükünün birkaç görevin uygulandığı çalışmada sonuçları zihinsel iş yükünün birkaç görevin tekrarıyla azaldığını göstermektedir [20]. Sabitleme sayısı ve kapalı kalma süresi gibi göz metriklerinin de görev zorluğu arttıkça arttığı gösterilmiştir. Görev kesintisizliği sebep olmaması nedeniyle bu göz davranışlarının sürekli iş yükü göstergeleri olarak kullanılabilirliği vurgulanmıştır [21].

Di Stasi vd. zihinsel iş yükünü çok boyutlu olarak değerlendirerek öznel, performansa dayalı ve nöroergonomik (göz izleme) teknikleri birlikte kullandıkları çalışmalarında alışveriş amaçlı iki farklı web tabanlı arama görevinde katılımcıları karşılaştırmışlardır. Sonuçta hedef odaklı grubun, satın alacağı şeyi araştırarak alacak olan araştırıcı gruba oranla; daha kısa sürede, daha kısa süreli göz sabitleme, daha küçük göz bebeği ve daha az göz seçirmesi davranışlarıyla görevi tamamladığı ve araştırıcı gruptan daha düşük NASA-TLX skorlarına sahip olduğu görülmüştür [6]. Marchitto vd.,

hava trafik çakışmalarını fark etme görevinin bilişsel iş yüküne etkisini kayıt altına alınan göz hareketleri (ortalama göz sabitleme süresi, toplam göz sabitleme sayısı vs.) aracılığıyla araştırmışlardır. Trafik çakışması olmadığı zamanla kıyaslandığında çakışma olduğu zaman görev karmaşıklığı artmış ve bunun zihinsel iş yüküne yansıdığı gözlemlenmiştir. Bahsi geçen çalışmada, zihinsel iş yükü ölçümü çok boyutlu ele alınarak hem performans ölçüsü (ortalama hata oranı ile tepki zamanı) hem sözlü öznel yöntem (verbal online subjective opinion - VOSO) hem de gözle ilgili fizyolojik verilerden yararlanılmıştır [22]. Evans ve Fendley otomasyon seviyeleriyle bilişsel iş yükünün fizyolojik etkileri arasındaki ilişkiyi incelemişlerdir. Çalışma neticesinde; NASA-TLX skorları, görevin tamamlanma süresi ve göz sabitleme oranı ile otomasyon seviyeleri arasında anlamlı bir ilişki olduğu ortaya çıkmıştır [23].

### 2.2. Makine Öğrenmesi ve Zihinsel İş Yükü (Machine Learning and Mental Workload)

Büyük boyutlu fizyolojik verilerin otomatik olarak analiz edilmesi ihtiyacının ortaya çıkmasıyla birlikte yapılan çalışmalarda, makine öğrenmesi ilkelerine dayalı yaklaşımlar bu konuda etkili olduğu gösterilmiştir. Aşağıda hem literatürde kullanımına en sık rastlanan hem de bu çalışmada kullanılan makine öğrenmesi algoritmalarının temel prensiplerine ve zihinsel iş yükü çalışmalarında kullanımına değinilmiştir.

#### 2.2.1. K en yakın komşu algoritması (K nearest neighborhood algorithm - kNN)

Temel mantığı gözlemlerin birbirlerine olan benzerlikleri üzerinden tahmin yapmak üzerine kurulu bir algoritmadır. K en yakın komşu algoritmasında k sayısı belirlenmekte ve her bir gözlemin kendisine en yakın k komşusuyla olan uzaklıkları (örn. öklit uzaklığı) hesaplanmaktadır. Sınıfı bilinmeyen gözlemin sınıf etiketi, uzaklığının en az olduğu komşuların sınıf etiketlerine göre belirlenmektedir. Burada k parametresinin optimize edilmesi model performansını artırmak için önem taşımaktadır [24]. Fizyolojik verilere göre zihinsel iş yükünün sınıflandırıldığı bir çalışmada diğer üç algoritmaya (destek vektör makineleri, karar ağaçları, rastgele ormanlar) kıyasla en iyi doğruluk oranı kNN algoritmasıyla elde edilmiştir [25].

#### 2.2.2. Destek vektör makineleri (Support vector machines - SVM)

İlk olarak Vapnik ve Chervonenkis tarafından 1991 yılında geliştirilmiştir [26]. Destek vektör makinelerinde amaç; bir marjın aralığına, en fazla sayıda gözlemi en küçük hata ile dâhil edebilecek doğruyu ya da eğriyi belirlemektir. Veriyi en doğru şekilde ayırabilen bir hiper düzlem bulmaya çalışan algoritmadır [27]. Bunu yaparken eş zamanlı olarak, hiperdüzlemin kendisine en yakın olan her bir veri noktasından uzaklığını maksimize etmeye çalışmaktadır [28]. Literatürde fizyolojik verilerinin kullanıldığı ve en iyi zihinsel iş yükü sınıflandırma sonuçlarının SVM ile elde edildiği çalışmalar mevcuttur [10, 29].

#### 2.2.3. Yapay sinir ağları (Artificial neural networks - ANN)

Yapay sinir ağları, biyolojik sinir sistemi ve insan beyninin yapısına dayanan bir makine öğrenmesi algoritmasıdır. Temel olarak bir yapay sinir ağı modeli girdi katmanı, gizli katman ve çıktı katmanı olmak üzere üç katmandan oluşmaktadır [30]. Bir katmandaki her bir nöron bir sonraki katmandaki tüm nöronlar ile bağlantılıdır. Yapay zekâ öğrenmesinde ağ çıktısının hedeflenen değerle eşleşmesini sağlamak amacıyla, bir ağı eğitmek için birçok girdi verisi ve çıktı hedefi kombinasyonu kullanılmaktadır. Kabul edilebilir bir hata miktarı elde edilinceye kadar ağ çıktısı ile gerçek çıktı arasındaki farklar minimize

edilmekte, bunun için de girdilerin ağırlıkları değiştirilmektedir. Girdi katmanında herhangi bir bilgi işleme operasyonu gerçekleşmezken, çıktı katmanında elde edilen çıktılar bir toplama ve aktivasyon fonksiyonu aracılığıyla üretilmektedir [31-33].

Wu vd. tarafından yapılan çalışmada; uzman olmayan 32 kişi ve uzman olan 7 kişiden 2 farklı zorluk seviyesindeki operasyonel görevleri gerçekleştirmeleri istenmiştir. Göz bebeği çapı, göz kırpması oranı, sabitleme oranı ve seçirme oranını girdi olarak kullandığı modelde NASA-TLX skoru ise çıktı olarak ele alınmıştır. Yapay sinir ağları ile elde edilen modelin doğruluğu test verisi için %97 olarak bulunmuştur [34]. Operatörlerin zihinsel iş yüklerini göz tepkisi verilerine dayanarak tahmin etmek için yapay sinir ağları modeli geliştirilen bir diğer çalışmada da göz tepkisi göstergeleri olarak göz bebeği çapı, göz kırpması oranı, sabitleme oranı ve seçirme oranı kullanılmıştır. Yan vd. tarafından gerçekleştirilen bu çalışmada iki öznel derecelendirme yöntemine ilişkin (NASA-TLX ve SWAT) skorlar ile işlem süresi ve hata sayısı gibi performans göstergeleri de kayıt altına alınmıştır. Geliştirilen yapay sinir ağı modelinin operatörlerin zihinsel iş yüklerini eğitim, doğrulama ve test verileri için sırasıyla 0,971, 0,912 ve 0,918 determinasyon katsayılarıyla ( $R^2$ ) tahmin ettiği ortaya konmuştur. Bu sonuçlar, operatörlerin göz tepkisi verilerine dayanarak zihinsel iş yükü tahminleri için yapay sinir ağları yaklaşımının oldukça doğru bir seçim olduğunu göstermiştir [35].

#### 2.2.4. Karar ağaçları (Decision trees - DT)

Bir sınıf etiketine ulaşmak amacıyla belirlenen kuralların düğümlerde depolandığı akış şemasına benzeyen ağaç yapılarına karar ağaçları denilmektedir. Bir veri ağaca girdiğinde değişkenlere ilişkin belirlenen kural setlerine göre sınıf etiketi belirlenmektedir [36]. Karar ağaçları yapılarında amaç, belli karar kuralları oluşturarak gözlemleri homojen alt gruplara ayırmaktır. Bu sayede veri seti içindeki karmaşık yapılar basit karar yapılarına dönüştürülmektedir [24].

#### 2.2.5. Rastgele ormanlar (Random forests - RF)

RF algoritması bagging adı verilen temeli bootstrap (yerine koymalı rastgele örnekleme) yöntemi ile oluşturulan birden fazla karar ağacının ürettiği tahminlerin bir araya getirilerek değerlendirilmesine dayanmaktadır. Bagging yönteminde t adet ağaç için m gözlemden n'er adet gözlem bootstrap yöntemiyle seçilmekte ve böylece t adet tahmin değeri elde edilmektedir. RF algoritması gözlem seçiminin rassal olarak yapıldığı bagging yöntemine ek olarak değişken seçimini de rassal olarak yapmaktadır. Tüm değişkenler arasından seçilen daha az sayıda değişken arasından seçilerek dallanma işlemi gerçekleştirilmekte ve rassallıkla birlikte genellenebilirliğin iyileştirilmesi amaçlanmaktadır [24].

Göz bebeği hareketine ilişkin fizyolojik değişkenlerin de yer aldığı zihinsel iş yükü tahmin çalışmasında kurulan regresyon modeli en iyi performansı rastgele ormanlar algoritmasıyla göstermiştir [37]. Aşırı öğrenmeye dirençli olması, kolay uygulanabilirliği (sadece iki parametre ayarlama) ve doğrusal olmayan verileri modelleme yeteneği ile öne çıkan bir algoritmadır [38].

#### 2.2.6. Gradyan artırma makineleri (Gradient boosting machines - GBM)

GBM bir grup zayıf karar ağacı sınıflandırıcısını iteratif olarak eğiterek kayıp fonksiyonunu optimize etmeyi ve güçlü bir sınıflandırıcı ortaya çıkarmayı amaçlayan bir algoritmadır [39]. Kayıp fonksiyonu (loss function), her bir gözlem için modelin yaptığı tahmindeki artık değerleri verir. Bir ağaçtaki her bir yaprak için, içinde bulunan gözlemlerin artıklarının ortalaması alınarak o ağaca ilişkin sonuçlar elde edilir. Tahmin, ağaçtan elde edilen sonuçların

öğrenme oranı (learning rate) denilen katsayı ile çarpılarak önceki ağacın sonuçlarına eklenmesi ile hesaplanır. Tahmin edilen değer ile gerçek değer arasındaki farkın yani yeni artık değerlerin her bir iterasyonda düştüğü gözlemlenecektir. Öğrenme oranı, aşırı öğrenme (overfitting) sorununu aşmak için kullanılmaktadır [40].

#### 2.2.7. Aşırı gradyan artırma (XGBoost)

Aşırı Gradyan Artırma (XGBoost), gradyanı artırılmış (gradient boosted) karar ağaçları temeline dayalı çalışan ölçeklenebilir bir makine öğrenme algoritmasıdır [41, 42]. Yeni nesil topluluk öğrenme algoritmalarından olan XGBoost, algoritmanın eğitilmesi sürecinde aşırı öğrenme sorununun önüne geçerek modelin genel doğruluğunu (performansını) artırmaktadır. Bu yöntemin başarısının altında yatan ana neden, öğrenme sürecinde kullandığı amaç fonksiyonudur. Amaç fonksiyonu, kayıp fonksiyonu ve düzenleme (regularization) teriminden oluşmaktadır. Kayıp fonksiyonu, modelin yaptığı her bir tahmin değerinin (predicted class) gerçek değerinden olan farkını hesaplamaktadır. Düzenleme terimi ise, modelin karmaşıklığını kontrol etmekte ve bu da modeldeki aşırı uyumlama sorununu ortadan kaldırmaktadır [43]. Wang ve Guo tarafından 2019 yılında yapılan çalışmada EEG-PPG fizyolojik verileri kullanılarak bilişsel yük sınıflandırılmaya çalışılmış ve SVM, KNN, DT (Decision Trees), RF, XGBoost algoritmaları arasında en yüksek sınıflandırma performansı XGBoost ile elde edilmiştir [44].

#### 2.2.8. Hafif gradyan artırma makineleri (LightGBM)

GBM gibi uygulamalarda birçok mühendislik optimizasyonu benimsenmiş olsa da, özellik boyutu ve veri hacmi büyük olduğunda verimlilik ve ölçeklenebilirlik hala tatmin edici bulunmamaktadır. Bunun önemli bir nedeni, her bir özellik için, bilgi kazancını tahmin etmek amacıyla tüm örneklerin taranması gerekliliğidir [45]. Çok zaman alan bu sorunu çözmek için Ke ve arkadaşları tarafından iki yeni teknik önerilmiştir [45]: Gradyan Tabanlı Tek Taraflı Örnekleme (Gradient-based One-Side Sampling-GOSS) ve Özel Özellik Paketleme (Exclusive Feature Bundling-EFB). GOSS ile küçük gradyanlara sahip veri örneklerinin önemli bir kısmı hariç tutulmakta ve bilgi kazancı tahmini için yalnızca geri kalan önemli kısım kullanılmaktadır. EFB ile ise, değişken boyutunun azaltılması amacıyla özel özellikler bir pakette bir araya getirilmektedir. Bu iki özellik sayesinde LightGBM ile geleneksel GBM'den 20 kat daha hızlı öğrenme süresinde hemen hemen aynı doğruluk performansı elde edilebilmektedir. Zeng vd. 2019 yılında yaptıkları çalışmada, EEG tabanlı bir zihinsel iş yükü sınıflandırma problemi için LightGBM tabanlı bir sınıflandırıcı geliştirmişler ve diğer sınıflandırıcılardan üstünlüğünü ortaya koymuşlardır [46].

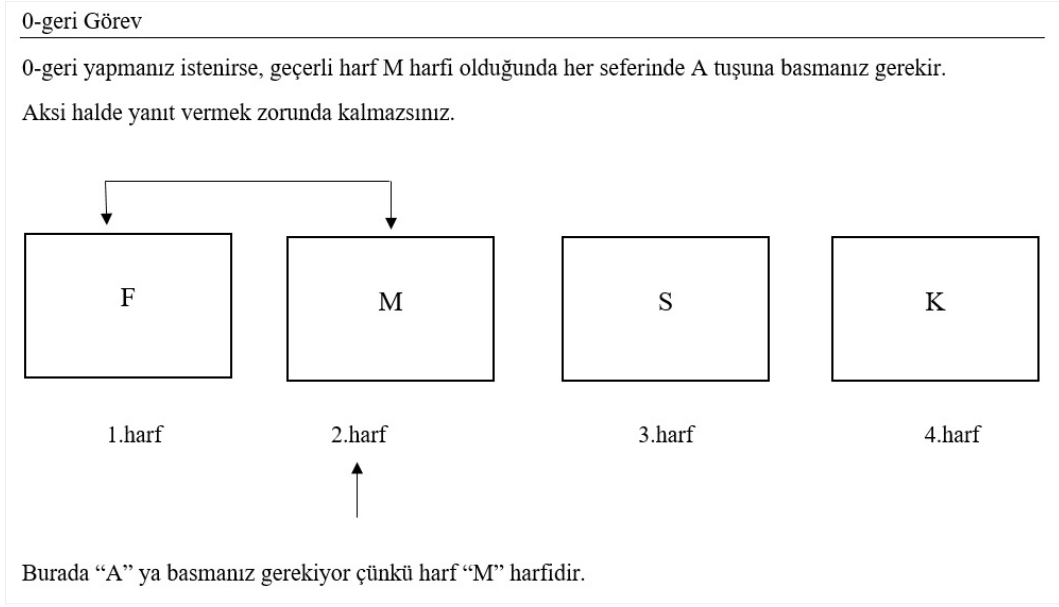
### 3. Yöntem (Methodology)

#### 3.1. Katılımcılar (Participants)

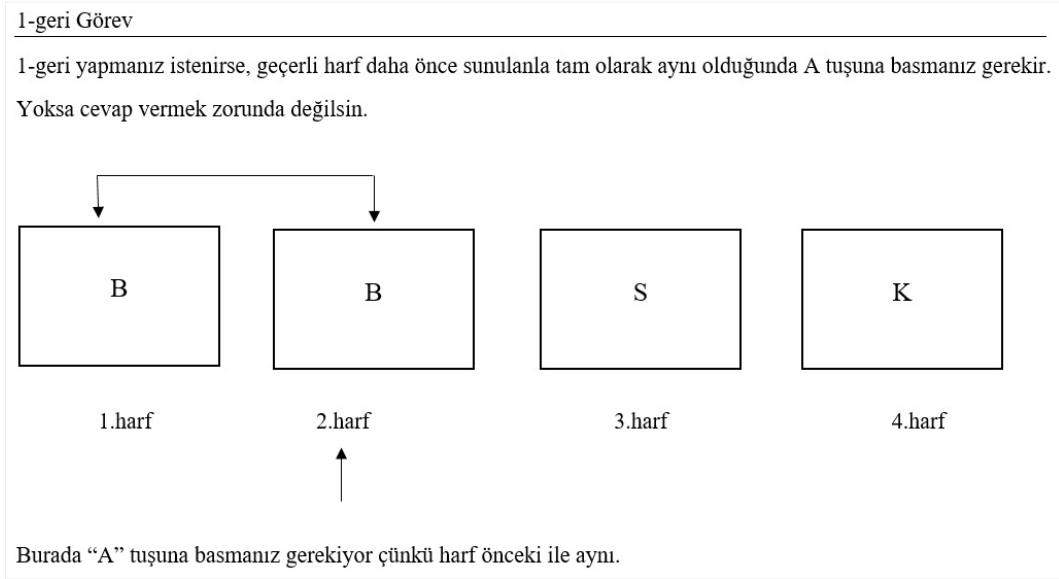
Çalışma, 30-40 yaş aralığında, görme problemi olmayan 4 katılımcı (2 kadın, 2 erkek) üzerinde gerçekleştirilmiştir. Katılımcıların yaş ortalaması 36 olarak bulunmuştur. Çalışmamıza dahil edilen tüm katılımcılardan bilgilendirilmiş onam formu alınmış ve çalışma için yapılan etik kurul başvurusu üniversite etik komisyonu tarafından onaylanmıştır. Katılımcılara sözlü olarak, çalışmanın amacı, gerçekleştirilecek testler ve yapılması gerekenler hakkında bilgi verilmiştir.

#### 3.2. Veri Kümesi (Dataset)

Tobii Studio yazılımı ile "sabitleme (fixation)", "seğirme (saccade)" ve "bilinmeyen (unclassified)" olmak üzere 3 türde göz olayı kayıt altına alınmaktadır. Bu doğrultuda öncelikle, literatürde en sık



Şekil 2. 0-Geri Görev (0-back task)



Şekil 3. 1-Geri Görev (1-back task)

rastlanılardan yola çıkılarak belirlenen 24 adet göz izleme özelliği analize dâhil edilmiştir [28, 34, 35, 47]. Bunlar; sabitleme sayısı, sabitleme toplam süresi, sabitleme süresi ortalaması, standart sapması, maksimum ve minimum değerleri, seçirme sayısı, seçirme toplam süresi, seçirme süresi ortalaması, standart sapması, maksimum ve minimum değerleri, seçirme 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, ortalama göz bebeği çapı ortalaması, standart sapması ve maksimum değeridir. Sabitleme değişkenlerine ilişkin veriler analize dâhil edilmeden önce sabitleme süresi 80 milisaniyenin altında ve 1200 milisaniyenin üstünde olan kayıtlar analiz dışı bırakılmıştır [10, 34, 35, 48]. Göz kırpması yazılım aracılığıyla doğrudan elde edilememiş, göz kırpması hareketlerine ilişkin değişkenlerin elde edilebilmesi amacıyla ayrı bir analiz gerçekleştirilmiştir. Her bir oturumda ekrana gelen harf uyarılarıyla

tetiklenen hafızaya dayalı zihinsel süreçler uyarıcı bazında gerçekleştiği için veri setini oluşturan satırlar uyarıcı bazında ele alınmıştır. 0-geri görevde 15, 1-geri görevde 16, 2-geri görevde 17, 3-geri görevde 18 uyarıcı (harf) ile karşılaştırılması sebebiyle bir katılımcı için toplamda 198 gözlem verisi, toplamda da 792 satırlık veri elde edilmiştir.

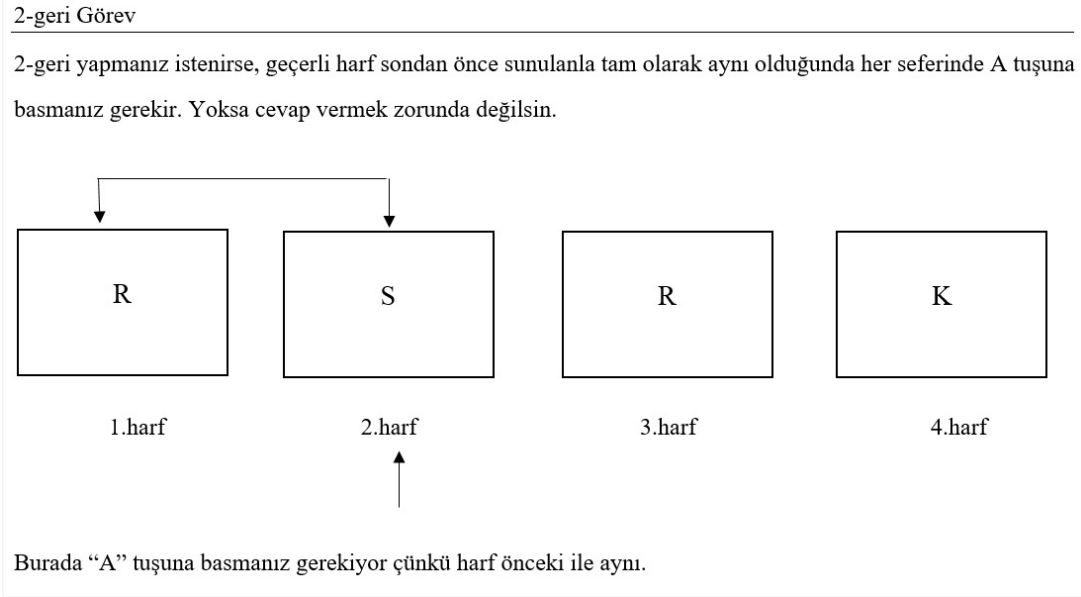
### 3.3. Deney Prosedürü (Experiment Procedure)

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 [49]. Deneylerde, literatürde sıkça rastlanan n-geri görev testleri kullanılmıştır [7, 11, 50]. Çalışmada 4 farklı zorluk seviyesinde n-geri hafıza testleri uygulanmıştır. N sayısı arttıkça görev zorluğu

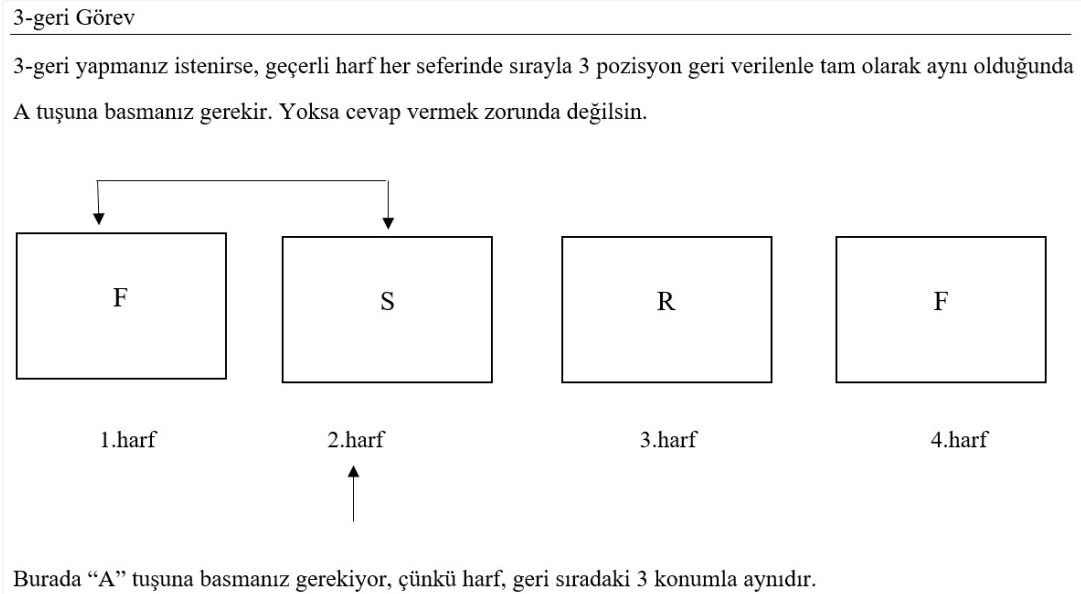
artmaktadır. Katılımcılardan ekranda belli bir süre görünen harfleri hafızalarında tutup, 0-geri görevde M harfini gördüklerinde “A” tuşuna basmaları, değilse hiçbir tuşa basmamaları istenmiştir (Şekil 2). 1, 2 veya 3 geri görevde ise; ekranda görünen harf sırasıyla 1 önceki, 2 önceki veya 3 önceki harfle aynı ise “A” tuşuna basmaları, değilse hiçbir tuşa basmamaları istenmiştir (Şekil 3, Şekil 4, ve Şekil 5).

Deneyler sırasında göz hareketleri Tobii X2-60 cihazı ile kayıt altına alınmıştır [51] (Şekil 6). Deneyler, yapay ışık (florasen) ile ışıklandırılmış sessiz bir ortamda yapılmıştır. Her bir katılımcı için kayda başlamadan önce kalibrasyon (5 nokta yöntemi) prosedürü gerçekleştirilmiştir. Kalibrasyon yapılırken katılımcıdan ekranda hareket eden kırmızı bir noktayı gözleriyle takip etmesi istenmektedir. Kayıtlar, kalibrasyon seviyesi kontrol edildikten sonra başlatılmıştır.

Görev başlamadan önce ekranda katılımcı için yazılı açıklamalar yer almaktadır. Katılımcılara testlerin anlaşılmasını sağlamak amacıyla bir kez kayıt dışı bir kez de kayıt başladıktan sonra olmak üzere 2’şer kez deneme testi uygulanmıştır. Ancak, kayda dâhil edilen deneme testi verisi analizin dışında tutulmuştur. 4 seviyeden 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. 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. Her bir denemenin ardından NASA-TLX ölçeği aracılığıyla katılımcılardan testleri gerçekleştirirken deneyimledikleri zihinsel iş yükünü öznel olarak değerlendirmeleri istenmiştir. Dolayısıyla bir katılımcı her deneme sonunda 1 kez olmak üzere

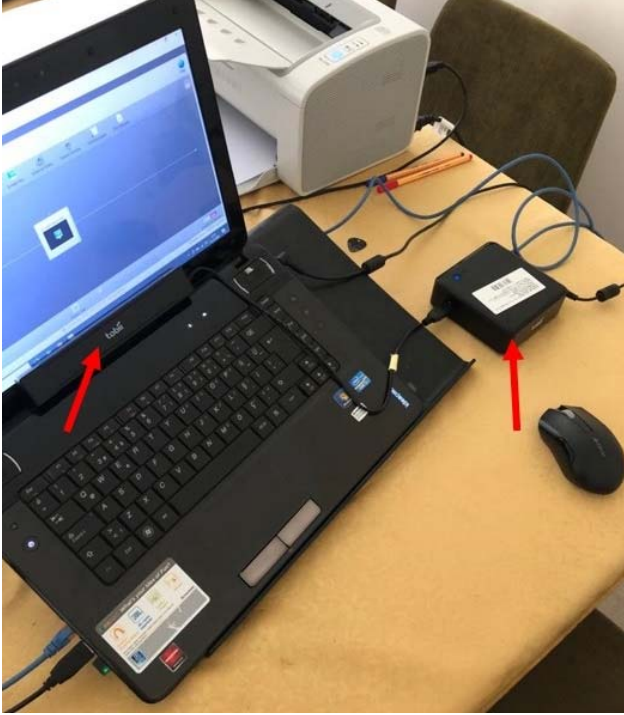


Şekil 4. 2-Geri Görev (2-back task)



Şekil 5. 3-Geri Görev (3-back task)

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 ağırlıklandırılmış toplam NASA-TLX skorları elde edilmiştir. Bir katılımcı ile gerçekleştirilen toplam deney süresi yaklaşık 20 dakikadır.



**Şekil 6.** Göz izleme cihazı ve bağlantı ekipmanları (Eye tracking device and connection equipment)

### 3.4. Veri Ön İşleme (Data Preprocessing)

Deneyde kullanılan Tobii X2-60 cihazı 60 Hz frekansta (saniyede 60 kayıt) veri toplamaya imkân sağlamaktadır. Göz izleme cihazıyla katılımcılardan elde edilen ve başlangıçta her bir katılımcı için yaklaşık 60.000 satırdan oluşan dosya üzerinde veri temizleme işlemi gerçekleştirilmiştir. Tobii Studio yazılımı aracılığıyla alınan raporda yer alan "RecordingTimeStamp" değeri kullanılarak 2500 milisaniye süren uyarıların başlangıç ve bitiş sınırları belirlenmiştir. Bu işlem sonrasında her bir katılımcı için yaklaşık 30.000'er satırlık dosyalar elde edilmiştir.

### 3.5. Göz Kırpma Değişkenlerinin Elde Edilmesi (Obtaining Blink Variables)

Literatürde göz kırpmayla ilişkili parametrelerin zihinsel iş yükü ile ilişkili olduğunu gösteren çalışmalar olması ve dolayısıyla göz kırpmaya hareketine ilişkin elde edilecek değişkenlerin sınıflandırma performansını artıracığı düşüncesiyle, varolan değişkenler kullanılarak değişken türetme işlemi yapılmıştır.

Göz kırpmaya, gözün tamamen veya kısmen kapanması durumudur [52]. Göz kırpmaya hareketinin tespitine ilişkin bir çalışmada göz bebeğinin merkezi esas alınmıştır [53]. Göz açıksa göz bebeği merkezi hesaplanabilmiş ve ilgili değişkene 1 değeri atanmış, göz kapalıysa merkez nokta hesaplanamamış ve değişken 0 değerini almıştır. Bu yaklaşımdan hareketle, Tobii Studio yazılımının sağladığı "PupilRight" ve "PupilLeft" değişkenlerinden yararlanılarak, söz

konusu değişkenlere ilişkin hesaplama yapılabildiyse gözün açık olduğu, her iki değişkenden de herhangi bir veri bulunmuyorsa gözün kapalı olduğu kabul edilmiştir. "Blink Değer" isimli bir değişken tanımlanarak bu değişkene gözün açık ve kapalı durumları için sırasıyla 0 ve 1 değerleri atanmıştır.

Bir göz kırpmaya toplamda 100-300 milisaniye sürmekte, 500 milisaniyenin üzerindeki göz kırpmalar ise uyuklama hali olarak kabul edilmektedir [54]. Bu çalışmada kullanılan cihazın frekansının 60 Hz olması sebebiyle bir kayıt satırının süresi 1/60 saniyedir. Bu nedenle, 100 milisaniye ve üzerindeki durumların göz kırpmaya olarak değerlendirilebilmesi için gözün kapalı kaldığı yani "Blink Değer" değişkeninin 1'e eşit olduğu kayıt sayısının minimum 6 olması gerekmektedir. Ek olarak, bir göz kırpmaya olayı iki sabitleme hareketi arasında gerçekleşeceği için [6], Tobii Studio yazılımının sabitleme yani "fixation" olarak belirlediği kayıtlar, göz kırpmaya taramasının dışında tutulmalıdır. Microsoft Office Excel Visual Basic for Applications (VBA) aracılığıyla, belirtilen şartları sağlayan kayıtlar belirlenmiş ve her biri için kaçınıcı göz kırpmaya olduğunu gösteren "Blink Index" değeri atanmıştır. Göz kırpmaya kayıtlarına karşılık gelen süreler ise, bir kayıt süresinin (1/60 saniye) her bir göz kırpmaya olayındaki kayıt sayısı ile çarpımı alınarak hesaplatılmış ve "Blink Duration" değişkenine yazdırılmıştır. Göz kırpmaya süresi 500 milisaniyenin üzerinde tespit edilen kayıtlar analiz dışı bırakılmıştır [52].

Bu şekilde "göz kırpmaya" olarak tespit edilen kayıtlar, bir sonraki aşamada farklı bir kod bloğuyla her bir uyarı bazında eşsiz şekilde saydırılmış ve yine uyarı bazında karşılık gelen toplam süreler belirlenmiştir. Böylece göz kırpmaya hareketlerine ilişkin olarak göz kırpmaya sayısı, toplam göz kırpmaya süresi, ortalama göz kırpmaya süresi değişkenleri analize dâhil edilmiştir.

Neticede gözle ilgili ele alınan toplam 27 değişkenin zihinsel iş yükü ile ilişkisi araştırılmış ve söz konusu değişkenlerin zihinsel iş yükü sınıfını tahmin etmedeki başarısı makine öğrenmesi algoritmaları kullanılarak test edilmiştir. Sayılara ilişkin değişkenlerin birimi adet, sürelerle ilişkin değişkenlerin birimi milisaniye, seçirme genliğine ilişkin verilerin birimi derece, gözbebeklerinin boyutlarına ilişkin değerlerin birimi ise milimetre olarak alınmıştır.

### 3.6. Değişken Önem Seviyelerinin Belirlenmesi (Determination of Variable Importance Levels)

Veri madenciliği sürecinin önemli aşamalarından biri veri boyutunun azaltılması işlemidir. Veri boyutunun azaltılması, 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. Özellik seçimi, orijinal veri setini temsil edebilecek en iyi altkümenin 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 [55].

Bu çalışmada, değişken seçimi için, ağaç tabanlı makine öğrenmesi algoritmalarının kendi içerisinde barındırdığı gömülü yöntemler incelenmiş ve değişken önem seviyeleri belirlenmiştir. Ek olarak, literatürde "Özyinelemeli Özellik Seçimi" olarak bilinen yöntem uygulanmıştır. Özyinelemeli özellik seçme yöntemi, 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 [56].



#### 4. Araştırma ve Bulgular (Research and Findings)

İstatistiksel analizler SPSS 21.0 yazılımı kullanılarak yapılmıştır. Öncelikle kullanılacak testlerin belirlenebilmesi amacıyla analiz edilecek değişkenlerin normal dağılımdan gelip gelmediği Kolmogorov-Smirnov testleriyle kontrol edilmiştir. Anlamlılık düzeyinin 0,05'in altında çıkması  $H_0$  hipotezinin reddedilmesi yani değişkenlere ilişkin verinin normal dağılımdan gelmediği anlamına gelmektedir. Değişkenlerin büyük çoğunluğuna ilişkin verilerin normal dağılımdan gelmemesi sebebiyle analizlerde parametrik olmayan testlerin kullanılmasına karar verilmiştir. Spearman korelasyon testi kullanılarak göz izleme değişkenleri ile NASA-TLX ve görev zorluk seviyeleri arasında ilişki olup olmadığı araştırılmıştır. Hem ağırlıklı 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 (Tablo 1).

Ağırlıklı NASA-TLX skoru ile görev zorluk seviyeleri arasındaki yüksek korelasyon, görev testlerinin önceden belirlenen 4 zorluk seviyesinin deneye katılan katılımcılar tarafından da beklendiği şekilde algılandığını göstermektedir. Hem ağırlıklı NASA-TLX toplam skoru hem de görev zorluk seviyesi değişkeni ile en yüksek korelasyon gösteren göz metrikleri göz bebeği boyutları olmuştur.

Literatürdeki çalışmalara benzer şekilde, görev zorluğu arttıkça göz bebeği boyutunun arttığı görülmektedir. Seçirme sayısının görev zorluğu arttıkça arttığı, sabitleme süresi ile ilgili değişkenlerin ise zorluk seviyesi ile birlikte azaldığı gözlemlenmiştir. Algılanan iş yükündeki artışın sabitleme ve seçirme sayısı ile pozitif, sabitleme süresi ile negatif ilişkili olduğu görülmüştür. Dolayısıyla daha zor bir görevi gerçekleştiren katılımcıların ekrana daha kısa sürelerle odaklandığı ve gözlerinin daha çok sayıda seçirme hareketi gerçekleştirdiği sonucu ortaya çıkmıştır. 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 gerçekleştiren ise, daha çok harfin hafızada tutulmaya çalışılması sonucu ortaya çıkan daha yoğun zihinsel aktivitenin gözün daha çok kez seçirmesine yol açtığı şeklinde yorumlanmaktadır. Göz kırpa değişkenlerinin ise algılanan zihinsel iş yükü ve zorluk seviyeleri ile pozitif yönde ilişkili olduğu ortaya çıkmıştır. Görev zorlaştıkça katılımcıların hem daha fazla sayıda göz kırpa hareketi gerçekleştirdikleri hem de göz kırpa sürelerinin uzadığı görülmektedir. Bu sonuç, literatürde göz kırpa hareketinin görev zorluğu arttıkça azaldığı bazı çalışmalardan farklılık göstermektedir. Görev zorluk seviyeleri arasında NASA-TLX skorları ve göz hareketleri açısından anlamlı ölçüde farklılık olup olmadığının test edilmesi için Kruskal Wallis testi uygulanmıştır. Kruskal Wallis testine göre her bir zorluk seviyesi arasında NASA-TLX skorları açısından anlamlı fark bulunmuştur ( $p < 0,001$ ). Ayrıca, maksimum seçirme süresi, sabitleme sayısı ve seçirme süresinin standart sapması dışında kalan 24 göz izleme değişkeninin de zorluk seviyesine göre anlamlı fark gösterdiği ortaya çıkmıştır (Tablo 2).

Görev zorluk seviyeleri ile NASA-TLX sonuçları arasındaki anlamlı ilişkinin doğrulanmasının ardından, görev zorluk seviyelerinin zihinsel iş yükü seviyelerine karşılık geldiği varsayımı altında, göz verilerinden yararlanarak zihinsel iş yükü seviyesini tahmin eden modelin kurulması aşamasına geçilmiştir. Tahmin modellerinin kurulması aşamasında Python yazılımının 3.8 sürümünden yararlanılmıştır [57]. Ele alınan sınıflandırma problemi için göz izleme değişkenlerine göre en doğru zorluk seviyesini belirleyen sınıflandırma modelini kurmak üzere makine öğrenmesi algoritmaları uygulanmıştır. 27 göz izleme özelliğinin bağımsız değişken, 4 sınıftan oluşan "Görev Zorluk Seviyesi" değişkeninin ise bağımlı değişken olarak ele alındığı sınıflandırma modellerinin performansını gösteren doğruluk (accuracy) değerleri hesaplanmıştır (Tablo 3).

**Tablo 1.** Korelasyon testi sonuçları (Correlation test results)

Değişkenler	Ağırlıklı Toplam TLX Skoru	Görev Zorluk Seviyesi
Görev Zorluk Seviyesi	0,839**	1,000
Sabitleme Sayısı	0,115**	,008
Toplam Sabitleme Süresi	-0,075*	-,131**
Ortalama Sabitleme Süresi	-0,188**	-,187**
Maksimum Sabitleme Süresi	-0,152**	-,168**
Minimum Sabitleme Süresi	-0,203**	-,139**
Sabitleme Süresi Standart Sapması	-0,158**	-,164**
Seçirme Sayısı	0,238**	,223**
Toplam Seçirme Süresi	0,123**	,062
Ortalama Seçirme Süresi	0,036	-,036
Maksimum Seçirme Süresi	0,008	,048
Minimum Seçirme Süresi	-0,160**	-,132**
Seçirme Süresi Standart Sapması	-0,038	,013
Ortalama Seçirme Genliği	0,242**	,148**
Maksimum Seçirme Genliği	0,251**	,131**
Seçirme Genliği Standart Sapması	0,188**	,102**
Sol Göz Bebeği Çapı Ortalaması	0,294**	,301**
Sağ Göz Bebeği Çapı Ortalaması	0,410**	,295**
Ortalama Göz Bebeği Çapı	0,369**	,312**
Sol Göz Bebeği Çapı Maksimum Değeri	0,254**	,287**
Sağ Göz Bebeği Çapı Maksimum Değeri	0,339**	,244**
Maksimum Göz Bebeği Çapı	0,302**	,276**
Sol Göz Bebeği Çapı Standart Sapması	0,049	,100**
Sağ Göz Bebeği Çapı Standart Sapması	0,205**	,117**
Göz Bebeği Çapı Standart Sapması	0,134**	,130**
Göz Kırpa Sayısı	0,165**	,210**
Toplam Göz Kırpa Süresi	0,226**	,200**
Ortalama Göz Kırpa Süresi	0,213**	,147**

**Tablo 2.** Görev zorluk seviyeleri arasındaki farklılıklar (Differences between task difficulty levels)

Değişkenler	Chi-Square	df	Asymp. Sig.
Ortalama Göz Bebeği Çapı	94,626	3	0,000
Sol Göz Bebeği Çapı Ortalaması	86,737	3	0,000
Sağ Göz Bebeği Çapı Ortalaması	86,733	3	0,000
Sol Göz Bebeği Çapı Maksimum Değeri	82,184	3	0,000
Maksimum Göz Bebeği Çapı	82,127	3	0,000
Sağ Göz Bebeği Çapı Maksimum Değeri	65,719	3	0,000
Ortalama Sabitleme Süresi	47,475	3	0,000
Göz Kırpma Sayısı	42,461	3	0,000
Seğirme Sayısı	41,253	3	0,000
Ortalama Seğirme Süresi	41,173	3	0,000
Toplam Göz Kırpma Süresi	37,209	3	0,000
Toplam Seğirme Süresi	34,102	3	0,000
Minimum Sabitleme Süresi	33,379	3	0,000
Maksimum Sabitleme Süresi	31,167	3	0,000
Sabitleme Süresi Standart Sapması	29,895	3	0,000
Göz Bebeği Çapı Standart Sapması	29,078	3	0,000
Minimum Seğirme Süresi	26,114	3	0,000
Ortalama Göz Kırpma Süresi	23,153	3	0,000
Sol Göz Bebeği Çapı Standart Sapması	20,571	3	0,000
Sağ Göz Bebeği Çapı Standart Sapması	19,585	3	0,000
Ortalama Seğirme Genliği	18,694	3	0,000
Toplam Sabitleme Süresi	18,391	3	0,000
Maksimum Seğirme Genliği	15,126	3	0,002
Seğirme Genliği Standart Sapması	11,079	3	0,011
Maksimum Seğirme Süresi	2,775	3	0,428
Sabitleme Sayısı	2,596	3	0,458
Seğirme Süresi Standart Sapması	0,858	3	0,835

En iyi doğruluk oranını elde edebilmek için gerçekleştirilen model ayarlama (tuning) adımı öğrenme oranı, maksimum derinlik ve ağaç sayısı hiper-parametreleri optimize edilmeye çalışılmıştır. Bunun için öncelikle 10-katlı çapraz doğrulama ile veri seti 10 parçaya bölünmüştür. Daha sonra öğrenme oranı, maksimum derinlik ve ağaç sayısı parametreleri için 12 farklı parametre kombinasyonu ile 120 kez (12x10) model fit edilmiştir. Parametre deneme setlerinden elde edilen en iyi parametre değerlerine yakın değerler ile de denemeler yapılmış ve en iyi parametre değerleri GBM algoritması için öğrenme oranı 0,1, maksimum derinlik 4 ve ağaç sayısı 200 olarak bulunmuştur. En iyi doğruluk değeri %65 olarak GBM algoritması kullanılarak elde edilmiştir. Girdi değişkenlerinin çıktı değişkeninin belirlenmesine ne derece katkıda bulunduğunu belirlemek için GBM algoritması kullanılarak bir duyarlılık analizi yapılmış ve sol göz bebeği çapı ortalamasının n-geri hafıza zorluk seviyesinin sınıflandırılmasında en etkili parametre olduğu görülmüştür (Şekil 7).

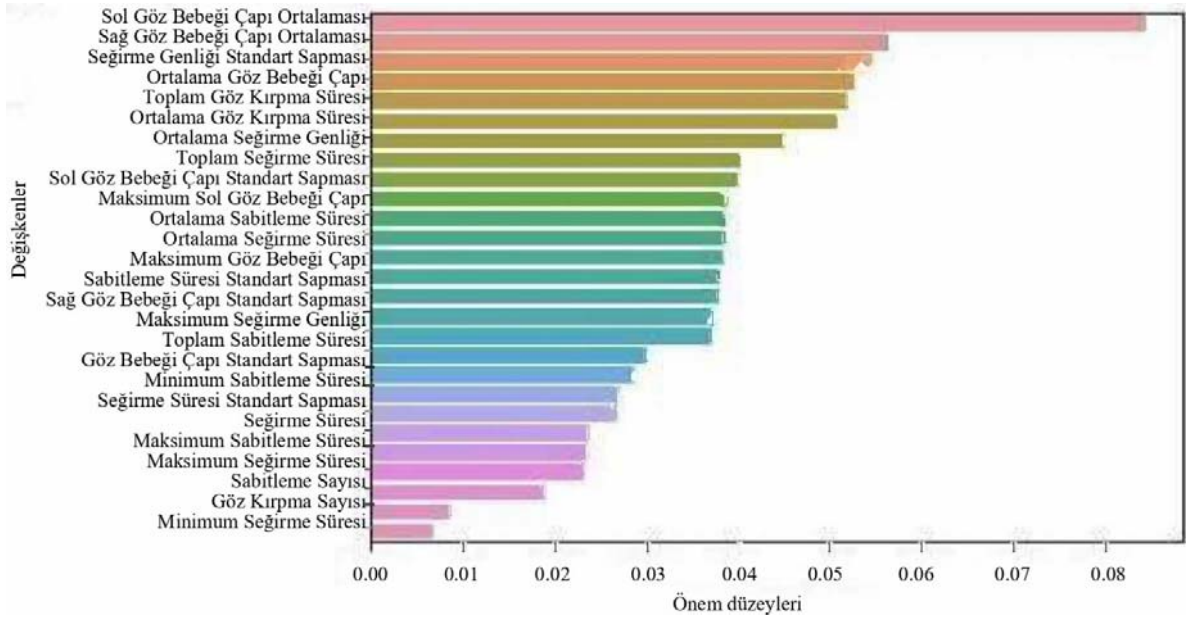
**Tablo 3.** Sınıflandırma Performansları Karşılaştırma Tablosu (Classification Performance Comparison Chart)

Model	27 Değişkenli Doğruluk (Accuracy)	14 Değişkenli Doğruluk (Accuracy)
kNN	0,34	0,38
SVM	0,45	0,54
DT	0,46	0,51
ANN	0,54	0,57
RF	0,59	0,61
XGBoost	0,62	0,63
LightGBM	0,63	0,66
GBM	0,65	0,68

Ayrıca, özyinelemeli özellik seçimi yöntemi uygulanarak en iyi performans gösteren ağaç algoritmaları ile belirlenen en önemli değişkenler incelenmiştir. Tüm sonuçlar göz önünde bulundurulduğunda, öncelikle Kruskal Wallis testi sonucunda hiçbir sınıf arasında ayırt edici bir fark göstermeyen 3 değişken (maksimum

seğirme süresi, sabitleme sayısı, seğirme süresi standart sapması) analizden çıkarılmıştır. Daha sonra, ağaç algoritmalarının önerdiği en önemli değişken kümelerinin dışında kalan 10 değişken daha çıkarılarak boyut sayısı 14'e indirilmiştir. Yeniden kurulacak modelde kullanılacak değişkenler; toplam sabitleme süresi, ortalama sabitleme süresi, toplam seğirme süresi, ortalama seğirme süresi, ortalama seğirme genliği, seğirme genliği standart sapması, sağ göz bebeği çapı ortalaması, sol göz bebeği çapı ortalaması, ortalama göz bebeği çapı, maksimum sağ göz bebeği çapı, maksimum sol göz bebeği çapı, maksimum göz bebeği çapı, toplam göz kırpma süresi ve ortalama göz kırpma süresidir. Daha az bağımsız değişkenli yeni veri seti ile algoritmalar yeniden çalıştırıldığında her bir modelin performansında bir miktar iyileşme olduğu görülmektedir (Tablo 3). En iyi sonucu veren algoritma %68 doğruluk oranıyla GBM algoritması olmuştur. Yani test verisinde bulunan tüm gözlemlerin %68'inin sınıflandırma tahmin edilmiştir. GBM modelinin ayarlama (tuning) adımı farklı hiper-parametre değerleri arasında yapılan denemeler sonucunda; öğrenme oranı 0,1, maksimum derinlik 4 ve ağaç sayısı 200 alındığında %68 başarı elde edilmiştir. %60'ın üzerinde performans gösteren diğer en iyi modeller sırasıyla LightGBM, XGBoost ve RF modelleridir. Borys vd. [10] tarafından 2017 yılında yapılan ve 3 sınıflı tahmin modelleri arasında en yüksek doğruluk oranının kNN algoritmasıyla (17 değişkenli model) %73 olarak elde edildiği çalışma ile kıyaslandığında, bu çalışmada 4 sınıflı model için elde edilen doğruluk oranının tatmin edici olduğu değerlendirilmektedir.

En iyi sınıflandırma performansı olarak elde edilen GBM modeline ilişkin hata matrisi (confusion matrix) Tablo 4'te verilmiştir. 198 tahminin 135'i doğru tahmin edildiği için doğruluk oranı %68'dir (792 satırlık verinin %75'i eğitim verisi, %25'i test verisi olarak kullanılmıştır). Tabloya göre gerçekte 0 kodlu (en düşük zorluk seviyesini gösteren kategori) sınıfa dâhil olan 43 gözlemden 27'si model tarafından doğru tahmin edilmiştir. Gerçekte 1 sınıfına dâhil olan 54 gözlemden 33'ü, 2 sınıfına dâhil olan 49 gözlemden 36'sı, 3 sınıfına dâhil olan 52 gözlemden ise 39'u doğru tahmin edilmiştir.



Şekil 7. GBM Algoritması ile Belirlenen Değişken Önem Seviyeleri (Variable importance levels determined by GBM)

0 olarak tahmin edilen 35 gözlemin 27'si doğru tahmin edilmiştir. %77 olarak elde edilen bu orana kesinlik (precision) değeri denilmektedir. Gerçekte 0 sınıfına dâhil olan 43 gözlemin 27'sinin doğru tahmin edilmiş olması nedeniyle duyarlılık (recall) değeri %63'tür. Tüm sınıflar için kesinlik ve duyarlılık değerleri özet sınıflandırma performansı tablosunda verilmiştir (Tablo 5). Yanlış sınıflandırma tahmininin en çok 0-1 ve 2-3 sınıfları arasında gerçekleştiği görülmektedir. Model tarafından yapılan 0 ile 1 tahminlerinde toplamda 14, 2 ile 3 tahminlerinde ise toplamda 18 hata yapılmıştır. Sınıf sayısının düşürülmesi yoluyla problem karmaşıklığının azaltılmasının sonuçlara nasıl etki edeceğinin araştırılması amacıyla, 0-1 sınıfındaki gözlemler ve 2-3 sınıfındaki gözlemler birleştirilmiş ve modeller tekrar test edilmiştir. Sonuçlar Tablo 6'da verilmiştir. Zihinsel iş yükünün düşük ve yüksek şekilde değerlendirilebileceği 2 sınıflı tahmin modelinde boyut azaltma yapılmadan dahi %84 başarı elde edilmiştir. LightGBM modelinin ayarlama (tuning) adımında farklı hiper-parametre değerleri arasından yapılan denemeler sonucunda; öğrenme oranı 0,2, maksimum derinlik 7 ve ağaç sayısı 300 alındığında %84 başarı elde edilmiştir. Model, test veri seti olarak rassal seçilen 198 adet gözlemin 166'sının sınıfını doğru tahmin etmiştir.

Tablo 4. Hata Matrisi (Confusion Matrix)

		Tahmin Edilen Sınıf			
		0	1	2	3
Doğru Sınıf	0	27	8	5	3
	1	6	33	9	6
	2	0	4	36	9
	3	2	2	9	39

Tablo 5. Özet Sınıflandırma Performansı (Summary of Classification Performance)

	Kesinlik (precision)	Duyarlılık (recall)	f1-score	support
0	0,77	0,63	0,69	43
1	0,70	0,61	0,65	54
2	0,61	0,73	0,67	49
3	0,68	0,75	0,72	52

Doğruluk (accuracy)	0,68	198
---------------------	------	-----

Tablo 6. İki Sınıflı Sınıflandırma Modellerinin Performans Değerleri (Performance Values of 2-Class Classification Models)

Model	27 Değişkenli Doğruluk (Accuracy)
kNN	0,59
SVM	0,70
DT	0,77
ANN	0,77
RF	0,79
XGBoost	0,83
GBM	0,83
LightGBM	0,84

## 5. Sonuçlar (Conclusions)

Bu çalışmada zihinsel iş yükünün tahmini, bir sınıflandırma problemi olarak ele alınmış ve "görev zorluk seviyesi" bağımlı değişkeni Python'da uygulanan makine öğrenmesi teknikleri ile tahmin edilmeye çalışılmıştır. Göz izleme değişkenleri ile zihinsel iş yükünü temsil eden değişkenler arasındaki beklenen ilişki test edilmiş ve özellikle sabitleme, seğirme ve göz bebeği değişkenlerine ilişkin sonuçların literatürdeki çalışmalar ile benzerlik gösterdiği görülmüştür. Görev zorluğu arttıkça; göz bebeği boyutu, seğirme sayısı, göz kırpma sayısı ve süresi artmakta, sabitleme süresi ise azalmaktadır. Göz kırpma sayısı ve süresinin zihinsel iş yükü ile birlikte artması literatürdeki bazı çalışmaların tersi yönünde bir sonuç olarak göze çarpmaktadır. Göz izleme cihaz yazılımının verdiği raporda yer almayan ve tamamen diğer değişkenlerden yararlanılarak türetilen göz kırpma değişkenlerinden özellikle toplam göz kırpma süresi ve ortalama göz kırpma süresinin, kullanılan algoritmalar tarafından önemli değişkenler arasında sayılması, söz konusu değişkenlerin modele katkı sağladığını açıkça göstermiştir.

Makine öğrenmesi algoritmaları ile oluşturulan sınıflandırma modelleri arasında, en başarılı sonuçların ağaç tabanlı algoritmalar ile elde edildiği görülmüştür. Yakın geçmişte geliştirilen ve bu nedenle literatürde bu konuda yapılan çalışmalarda kullanımına sık rastlanmayan GBM, LightGBM, XGBoost gibi algoritmalar ile diğerlerine kıyasla daha iyi sonuçlar elde edilmiştir. Ancak uygulanan

algoritmaların sınıflandırma performansları hem değişken sayısının hem de sınıf sayısının fazla olmasından dolayı ilk analizde düşük çıkmıştır. Katılımcı sayısının az olması da model başarısının istenen düzeyde olmamasının sebeplerinden biri olarak görülmektedir. Boyut azaltmanın performans iyileştirmedeki etkisi bu çalışmada beklenen düzeyde olmamakla birlikte, özellik seçiminin modellerde iyileşme sağladığı açıktır. Ayrıca, sınıf sayısının azaltılması yoluna gidildiğinde çok daha iyi bir sınıflandırma performansı (%84) ile karşılaşılacağı görülmüştür. Hem katılımcı sayısının artması, hem de farklı özellik ve hiper-parametre alt kümelerinin denemesiyle yapılacak ayrıntılı analizlerle birlikte daha fazla iyileşme sağlanabileceği değerlendirilmektedir. Ayrıca sonraki çalışmalarda, model girdisinin EEG vb. farklı fizyolojik yöntemlerle çeşitlendirilmesi ile çalışmanın geliştirilebilmesi mümkündür.

#### Teşekkür (Acknowledgement)

Bu Çalışma Gazi Üniversitesi Bilimsel Araştırma Projeleri Birimi Tarafından Desteklenmiştir. Proje No: 6483. Çalışmada kullanılan göz izleme cihazı Orta Doğu Teknik Üniversitesi İnsan-Bilgisayar Etkileşimi Araştırma ve Uygulama Laboratuvarı'ndan temin edilmiştir. Cihazın kullanımına ilişkin desteklerinden ötürü söz konusu kuruma teşekkür ederiz.

#### Kaynaklar (References)

- Bommer S., Fendley M., A theoretical framework for evaluating mental workload resources in human systems design for manufacturing operations, *International Journal of Industrial Ergonomics*, 63, 7-17, 2016.
- Galy E., Cariou M., Mélan C., What is the relationship between mental workload factors and cognitive load types?, *International Journal of Psychophysiology*, 83 (3), 269-275, 2012.
- DiDomenico A., Nussbaum M., Effects of different physical workload parameters on mental workload and performance, *International Journal of Industrial Ergonomics*, 41 (3), 255-260, 2011.
- Rusnock C., Borghetti B., McQuaid I., Objective-Analytical Measures of Workload – the Third Pillar of Workload Triangulation?, *International Conference on Augmented Cognition*, Los Angeles – USA, 124-135, August 2-7, 2015.
- Puma S., Matton N., Paubel P.V., Raufaste E., Yagoubi R., Using theta and alpha band power to assess cognitive workload in multitasking environments, *International Journal of Psychophysiology*, 123, 111-120, 2017.
- Di Stasi L. L., Antolí A., Gea M., Cañas J. J., A neuroergonomic approach to evaluating mental workload in hypermedia interactions, *International Journal of Industrial Ergonomics*, 41 (3), 298-304, 2011.
- Tjolleng A., Jung K., Hong W., Lee W., Lee B., You H., Son H., Park S., Classification of a Driver's cognitive workload levels using artificial neural network on ECG signals, *Applied Ergonomics*, 59, 326-332, 2017.
- Dirican A.C., Göktürk M., Psychophysiological measures of human cognitive states applied in human computer interaction, *Procedia Computer Science*, 3, 1361-1367, 2011.
- Chen S., Epps J., Automatic classification of eye activity for cognitive load measurement with emotion interference, *Computer Methods and Programs in Biomedicine*, 110 (2), 111-124, 2013.
- Borys M., Plechawska-Wojcik M., Wawrzyk M., Wesołowska K., Classifying Cognitive Workload Using Eye Activity and EEG Features in Arithmetic Tasks, *International Conference on Information and Software Technologies*, Druskininkai - Lithuania, 90-105, October 12-14, 2017.
- Liu Y., Ayaz H., Shewokis P.A., Multisubject "Learning" for Mental Workload Classification Using Concurrent EEG, fNIRS, and Physiological Measures, *Frontiers in Human Neuroscience*, 11 (389), 2017.
- Choi M.K., Lee S.M., Ha J.S., Seong P.H., Development of an EEG-based workload measurement method in nuclear power plants, *Annals of Nuclear Energy*, 111, 595-607, 2018.
- Hart S.G., Staveland L.E., Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research, *Advances in psychology*, 52,139-183, 1988.
- Nakayama M., Takahashi K., Shimizu Y., The act of task difficulty and eye-movement frequency for the 'Oculo-motor indices', *Proceedings of the 2002 symposium on Eye tracking research & applications*, 37-42, March, 2002.
- Benedetto S., Pedrotti M., Minin L., Baccino T., Re A., Montanari R., Driver workload and eye blink duration, *Transportation Research Part F: Traffic Psychology and Behaviour*, 14 (3), 199-208, 2011.
- Gao Q., Wang Y., Song F., Li Z., Dong X., Mental workload measurement for emergency operating procedures in digital nuclear power plants, *Ergonomics*, 56, 1070-1085, 2013.
- Tran C.C., Yan S., Habiaryemye J.L., Wei Y., Predicting Driver's Work Performance in Driving Simulator Based on Physiological Indices, *International Conference on Intelligent Human Computer Interaction*, Evry-France, 150-162, December 11-13, 2017.
- Hampson X., Zhuang D., Lin Y., Xiao X., Song J.W., Influence of mental workload on detecting information varieties revealed by mismatch negativity during flight simulation, *International Journal of Industrial Ergonomics*, 64, 1-7, 2018.
- Hampson R., Opris I., Deadwyler S., Neural Correlates of Fast Pupil Dilation in Nonhuman Primates: Relation to Behavioral Performance and Cognitive Workload, *Behavioural brain research*, 212 (1), 1-11, 2010.
- Reiner M., Gelfeld T.M., 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, 2014.
- Huang W., Xu Y., Hildebrandt M., Lau N., Comparing Eye-Gaze Metrics of Mental Workload in Monitoring Process Plants, *International Conference on Human-Computer Interaction*, Orlando-USA, 55-65, July 26-31, 2019.
- Marchitto M., Benedetto S., Baccino T., Cañas J., Air traffic control: Ocular metrics reflect cognitive complexity, *International Journal of Industrial Ergonomics*, 54, 120-130, 2016.
- Evans D.C., Fendley M., A multi-measure approach for connecting cognitive workload and automation, *International Journal of Human-Computer Studies*, 97, 182-189, 2017.
- Keskin M.V., Python ile Makine Öğrenmesi (Machine Learning). [Çevrimiçi Eğitim] <https://www.udemy.com/course/python-ile-makine-ogrenmesi/>. Yayın tarihi 2020. Erişim tarihi Nisan 15, 2021.
- Plechawska-Wojcik M., Tokovarov M., Kaczorowska M., Zapala D., A Three-Class Classification of Cognitive Workload Based on EEG Spectral Data, *Applied Sciences*, 9 (24), 5340, 2019.
- Vapnik N.V., Chervonenkis A.Y., The necessary and sufficient conditions for consistency in the empirical risk minimization method, *Pattern Recognition and Image Analysis*, 1 (3), 283-305, 1991.
- Gökdemir A., Çalhan A., Deep learning and machine learning based anomaly detection in internet of things environments, *Journal of the Faculty of Engineering and Architecture of Gazi University*, 37 (4), 1945-1956, 2022.
- Benerradi J., Maior H., Marinescu A., Clos J., Wilson M., Exploring Machine Learning Approaches for Classifying Mental Workload using fNIRS Data from HCI Tasks, *Proceedings of the Halfway to the Future Symposium*, 1-11, 2019.
- Kaczorowska M., Wawrzyk M., Plechawska-Wojcik M., Binary Classification of Cognitive Workload Levels with Oculography Features, *International Conference on Computer Information Systems and Industrial Management*, Bialystok-Poland, 243-254, October 16-18, 2020.
- Datlca M.T., Çakıt E., Estimation of clustering parameters and anomaly detection in tracking devices with changeable position time, *Journal of the Faculty of Engineering and Architecture of Gazi University*, 36 (1), 373-394, 2021.
- Çakıt E., Durgun B., Cetik O., A neural network approach for assessing the relationship between grip strength and hand anthropometry, *Neural Network World*, 25 (6), 603, 2015.
- Çakıt E., Karwowski W., Understanding the social and economic factors affecting adverse events in an active theater of war: a neural network approach, *International Conference on Applied Human Factors and Ergonomics*, Los Angeles-USA, 215-223, July 17-21, 2017.
- Çakıt E., Karwowski W., Estimating electromyography responses using an adaptive neuro-fuzzy inference system with subtractive clustering,

- Human Factors and Ergonomics in Manufacturing & Service Industries, 27 (4), 177-186, 2017.
34. Wu Y., Liu Z., Jia M., Congchi T., Yan S., Using Artificial Neural Networks for Predicting Mental Workload in Nuclear Power Plants Based on Eye Tracking, *Nuclear Technology*, 206 (1), 94-106, 2020.
  35. Yan S., Wei Y., Tran C.C., Evaluation and prediction mental workload in user interface of maritime operations using eye response, *International Journal of Industrial Ergonomics*, 71, 117-127, 2019.
  36. Duru A.D., 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, 2019.
  37. Smith A.M., Borghetti B.J., Rusnock C.F., Improving Model Cross-Applicability for Operator Workload Estimation, *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, Los Angeles-USA*, 681-685, September, 2015.
  38. Borghetti B.J., Giametta J.J., Rusnock C.F., Assessing Continuous Operator Workload with a Hybrid Scaffolded Neuroergonomic Modeling Approach, *Human Factors*, 59 (1), 134-146, 2017.
  39. Friedman J., Greedy function approximation: A gradient boosting machine, *Annals of Statistics*, 29, 1189-1232, 2001.
  40. Muratlar E.R., Gradient Boosted Regresyon Ağaçları. <https://www.veribilimiokulu.com/gradient-boosted-regresyon-agaclari/>. Yayın tarihi Ocak 24, 2020. Erişim tarihi Mayıs 10, 2021.
  41. Chen T., Guestrin C., XGBoost: A Scalable Tree Boosting System, *International Conference on Knowledge Discovery and Data Mining, San Francisco-USA*, 785-794, August 13-17, 2016.
  42. Çakıt E., Dağdeviren M., Predicting the percentage of student placement: A comparative study of machine learning algorithms, *Education and Information Technologies*, 27 (1), 997-1022, 2022.
  43. Üstüner M., Abdikan S., Bilgin G., Şanlı F.B., Hafif Gradyan Artırma Makineleri ile Tarımsal Ürünlerin Sınıflandırılması, *Turkish Journal of Remote Sensing and GIS*, 1 (2), 97-105, 2020.
  44. Wang C., Guo J., A data-driven framework for learners' cognitive load detection using ECG-PPG physiological feature fusion and XGBoost classification, *Procedia Computer Science*, 147, 338-348, 2019.
  45. Ke G., Meng Q., Finley T., Wang T., Chen W., Ma W., Ye Q., Liu T., LightGBM: a highly efficient gradient boosting decision tree, *Proceedings of Advances in Neural Information Processing Systems (NIP 17)*, Curran Associates Inc., Red Hook, NY, USA, 3146-3154, 2017.
  46. Zeng H., Yang C., Zhang H., Wu Z., Zhang J., Dai G., Babiloni F., Kong W., A LightGBM-Based EEG Analysis Method for Driver Mental States Classification, *Computational Intelligence and Neuroscience*, 1-11, 2019.
  47. Mark J., Curtin A., Kraft A., Sands T., Casebeer W., Ziegler M., Ayaz H., Eye Tracking-Based Workload and Performance Assessment for Skill Acquisition, *International Conference on Applied Human Factors and Ergonomics, Washington D.C. - USA*, 129-141, July 24-28, 2020.
  48. Borys M., Tokovarov M., Wawrzyk M., Wesolowska K., Plechawska-Wojcik M., Dmytruk R., Kaczorowska M., An analysis of eye-tracking and electroencephalography data for cognitive load measurement during arithmetic tasks, *Bucharest - Romania*, 287-292, March 23-25, 2017.
  49. Millisecond Software. Inquisit by Millisecond. <https://www.millisecond.com/>. Erişim tarihi Eylül 15, 2021.
  50. Ke Y., Qi H., Zhang L., Chen S., Jiao X., Zhou P., Zhao X., Wan B., Ming D., 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, 2015.
  51. Tobii X2-60 Eye Tracker. <https://www.tobii.com/>. Stockholm, Sweden: Tobii Pro. Erişim tarihi Ağustos 21, 2021.
  52. Ahlstrom U., Friedman-Berg F.J., Using Eye Movement Activity as A Correlate of Cognitive Workload, *International Journal of Industrial Ergonomics*, 36 (7), 623-636, 2006.
  53. Naveed S., Sikander B., Khiyal M.S.H., Eye Tracking System with Blink Detection, *Journal of Computing*, 4 (3), 51-60, 2012.
  54. Johns M.W., The amplitude velocity ratio of blinks: A new method for monitoring drowsiness, *Sleep*, 26, p. A51, 2003.
  55. Budak H., Özellik Seçim Yöntemleri ve Yeni Bir Yaklaşım, *Süleyman Demirel Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, 22, 21-31, 2018.
  56. Kocakafa T., Özellik Oluşumu ve Özellik Seçimi (Feature Selection)-3. <https://www.veribilimiokulu.com/ozellik-olusumu-ve-ozellik-secimifeature-selection-3/> Yayın tarihi Ocak 10, 2021. Erişim tarihi Mayıs 24, 2021.
  57. Python (Sürüm 3.8). <https://www.anaconda.com/>. Erişim tarihi Ekim 20, 2021.

