



**ODYOMETRİ SİSTEM TASARIMI VE EEG
SİNYALLERİ KULLANILARAK MAKİNE
ÖĞRENMESİ YÖNTEMLERİ İLE İŞİTME TESTİ**

Mustafa KÜÇÜKAKARSU

**2021
YÜKSEK LİSANS TEZİ
BİYOMEDİKAL MÜHENDİSLİĞİ**

**Tez Danışmanı
Dr. Öğr. Üyesi Ahmet Reşit KAVSAOĞLU**

**ODYOMETRİ SİSTEM TASARIMI VE EEG
SİNYALLERİ KULLANILARAK MAKİNE
ÖĞRENMESİ YÖNTEMLERİ İLE İŞİTME TESTİ**

Mustafa KÜÇÜKAKAR SU

**T.C.
Karabük Üniversitesi
Lisansüstü Eğitim Enstitüsü
Biyomedikal Mühendisliği Anabilim Dalında
Yüksek Lisans Tezi
Olarak Hazırlanmıştır**

**Tez Danışmanı
Dr. Öğr. Üyesi Ahmet Reşit KAVSAOĞLU**

**KARABÜK
Haziran 2021**

Mustafa KÜÇÜKAKARSU tarafından hazırlanan “ODYOMETRİ SİSTEM TASARIMI VE EEG SİNYALLERİ KULLANILARAK MAKİNE ÖĞRENMESİ YÖNTEMLERİ İLE İŞİTME TESTİ” başlıklı bu tezin Yüksek Lisans Tezi olarak uygun olduğunu onaylarım.

Dr. Öğr. Üyesi Ahmet Reşit KAVSAOĞLU
Tez Danışmanı, Biyomedikal Mühendisliği Anabilim Dalı

Bu çalışma, jürimiz tarafından oy birliği ile Biyomedikal Mühendisliği Anabilim Dalında Yüksek Lisans tezi olarak kabul edilmiştir. 24/06/2021

Unvanı, Adı SOYADI (Kurumu) İmzası

Başkan : Dr. Öğr. Üyesi Sabri ALTUNKAYA (NEÜ)

Üye : Doç. Dr. Mehmet Akif ERDEN (KBÜ)

Üye : Dr. Öğr. Üyesi Ahmet Reşit KAVSAOĞLU (KBÜ)

KBÜ Lisansüstü Eğitim Enstitüsü Yönetim Kurulu, bu tez ile, Yüksek Lisans derecesini onamıştır.

Prof. Dr. Hasan SOLMAZ
Lisansüstü Eğitim Enstitüsü Müdürü



“Bu tezdeki tüm bilgilerin akademik kurallara ve etik ilkelere uygun olarak elde edildiğini ve sunulduğunu; ayrıca bu kuralların ve ilkelerin gerektirdiği şekilde, bu çalışmadan kaynaklanmayan bütün atıfları yaptığımı beyan ederim.”

Mustafa KÜÇÜKAKARSU

ÖZET

Yüksek Lisans Tezi

ODYOMETRİ SİSTEM TASARIMI VE EEG SİNYALLERİ KULLANILARAK MAKİNE ÖĞRENMESİ YÖNTEMLERİ İLE İŞİTME TESTİ

Mustafa KÜÇÜKAKARSU

Karabük Üniversitesi

Lisansüstü Eğitim Enstitüsü

Biyomedikal Mühendisliği Anabilim Dalı

Tez Danışmanı:

Dr. Öğr. Üyesi Ahmet Reşit KAVSAOĞLU

Haziran 2021, 52 sayfa

Bu çalışmanın amacı, işitme testi (odyometri) işlemlerinin EEG sinyalleri ile otonom bir şekilde makine öğrenmesi yöntemleri kullanılarak gerçekleştirilmesidir. Standart işitme testlerinde test edilen kişiye verilen farklı genlik ve dalgaboylarına sahip sesler, rassal bir şekilde MATLAB GUI ile tasarlanan arayüz ile verilmiştir. Kişi, kulaklık ile dinlediği rassal büyüklükteki sesleri duyduğunu belirtmiş, duymamış ise herhangi bir işlem yapmamıştır. Eş zamanlı olarak EEG (elektroensefalografi) sinyalleri takip edilmiş ve kişinin duyduğu ve duymadığı seslerin beyinde oluşturduğu dalgalar kayıt altına alınmıştır. Test bitiminde oluşturulan EEG verileri ön işleme adımlarından geçirilmiş ve sonrasında öznitelik çıkarımı yapılmıştır. MATLAB arayüzünden alınan duydu ve duymadı bilgileri EEG sinyalleri ile birleştirilmiş ve kişinin hangi sesleri duyduğu hangi sesleri duymadığı tespit edilmiştir. Arayüz aracılığı ile verilen sesler arasında ki bekleme süresinde kişiye herhangi bir ses verilmemiştir. Dolayısıyla bu

zamanlar da EEG sinyallerinde duyulmadı olarak işaretlenmiştir. Bu yüksek lisans tez çalışmasında, beyin sinyalleri Brain Products Vamp 16 EEG cihazı ile ölçülmüştür ve sonrasında Brain Vision Recorder program ile MATLAB kullanılarak EEG ham verisi oluşturulmuştur. Duyulan ve duyulmayan seslerin beyinde oluşturduğu sinyal verisinden veri seti oluşturulduktan sonra PYTHON programa dili ile makine öğrenmesi işlemleri gerçekleştirilmiştir. MATLAB ile oluşturulan ham veri, Python programlama dili ile alınmış, üzerinde ön işleme adımları tamamlandıktan sonra sınıflandırma algoritmaları üzerinde makine öğrenmesi yöntemleri uygulanmıştır. Analizlerde Naïve Bayes, LGBM, SVC, DTC, KNC, LR, RFC algoritmaları uygulanmış ve tahmin bakımından en başarılı algoritmanın %84 başarı oranı ile LGBM sınıflandırma algoritması olduğu tespit edilmiştir.

Anahtar Sözcükler : Odyometri, İşitme Testi, EEG, Makine Öğrenmesi,
Sınıflandırma Algoritmaları

Bilim Kodu : 92504

ABSTRACT

M. Sc. Thesis

AUDIOMETRY SYSTEM DESIGN AND HEARING TEST WITH MACHINE LEARNING METHODS USING EEG SIGNALS

Mustafa KÜÇÜKAKARSU

Karabük University

Institute of Graduate Programs

Department of Biomedical Engineering

Thesis Advisor:

Asst. Prof. Dr. Ahmet Reşit KAVSAOĞLU

June 2021, 52 pages

The purpose of this study is the realization of hearing test (audiometry) procedures carried out autonomously with EEG signals using machine learning methods. Sounds, which are provided in different amplitudes and wavelengths to the person tested in standard hearing tests, are provided randomly with the interface designed with MATLAB GUI. The person stated that he heard random-sized sounds he listened to with headphones, and if he did not, he did not take any action. Simultaneously, EEG (electroencephalography) signals were tracked, and the waves created in the brain by the sounds heard and unheard by the person were recorded. At the end of the test, the generated EEG data were subjected to pre-processing steps, and then feature extraction was carried out. Heard and unheard information received from the MATLAB interface combined with EEG signals and heard and unheard sounds by the person was determined. No sound was provided to the person during the waiting period between the sounds provided via the interface. Therefore, EEG signals were marked as unheard

during these times. In this master thesis, the brain signals were measured with the Brain Products Vamp 16 EEG device and then, raw EEG data were created by using Brain Vision Recorder software and MATLAB. Once the data set is generated from the signal data formed by the heard and unheard sounds in the brain, machine learning processes were carried out with the PYTHON programming language. The raw data generated with MATLAB processed with the Python programming language and machine learning methods were applied on classification algorithms once the pre-processing steps are completed. Naïve Bayes, LGBM, SVC, DTC, KNC, LR, RFC algorithms were used in the analysis, and with a success rate of 84%, the LGBM classification algorithm was determined to be the most successful algorithm in terms of prediction.

Key Words : Audiometry, Hearing Test, EEG, Machine Learning, Classification Algorithms.

Science Code : 92504

TEŐEKKÜR

Bu tez alıŐmasının planlanmasında, araŐtırılmasında, yürütülmesinde ve oluşumunda ilgi ve desteęini esirgemeyen, engin bilgi ve tecrübelerinden yararlandıęım, yönlendirme ve bilgilendirmeleriyle alıŐmamı bilimsel temeller ışığında şekillendiren sayın hocam Dr. Öğr. Üyesi Ahmet ReŐit KAVSAOęLU'na sonsuz teşekkürlerimi sunarım.

Sevgili aileme manevi hiçbir yardımı esirgmeden yanımda oldukları için tüm kalbimle teşekkür ederim.

İÇİNDEKİLER

	<u>Sayfa</u>
KABUL.....	ii
ÖZET	iv
ABSTRACT.....	vi
TEŞEKKÜR.....	viii
İÇİNDEKİLER	ix
ŞEKİLLER DİZİNİ.....	xii
ÇİZELGELER DİZİNİ	xiii
SİMGELER VE KISALTMALAR DİZİNİ.....	xiv
BÖLÜM 1	1
GİRİŞ	1
BÖLÜM 2	6
MATERYAL VE YÖNTEM.....	6
2.1. İŞİTME TESTİ ARAYÜZÜ	6
2.1.1 MATLAB GUI İşitme Testi Arayüzü	6
2.1.2 Kulaklık	9
2.2. EEG KAYIT CİHAZI	10
2.3. RECORDER YAZILIMI	13
2.4. DİJİTAL PORT YAPILANDIRMASI	15
2.5. VERİ SETİ OLUŞTURMA	16
2.5.1 Öznitelik Çıkarımı	19
2.5.1.1 Zaman Uzayında Öznitelik Çıkarımı	19
2.5.1.2 Frekans Uzayında Öznitelik Çıkarımı	21
BÖLÜM 3	22
MAKİNE ÖĞRENMESİ ORTAMI.....	22
3.1. VERİ ÖN İŞLEME	23

	<u>Sayfa</u>
3.1.1 Count Vectorizer (Sayma Vektörü).....	23
3.1.2 TFIDF Yöntemi	24
3.2. SINIFLANDIRMA	24
3.2.1. Naïve Bayes Sınıflama Algoritması	25
3.2.2 Hafif Gradyan Güçlendirme Makinesi Sınıflama Algoritması.....	25
3.2.3 Rastgele Orman Sınıflama Algoritması.....	26
3.2.4. K-En Yakın Komşular Sınıflama Algoritması	26
3.2.5. Destek Vektör Makineleri Sınıflama Algoritması.....	27
3.2.6. Karar Ağaçları Sınıflama Algoritması.....	28
3.2.7. Lojistik Regresyon.....	29
3.3. TEST ADIMLARI	30
3.3.1. Confusion Matrix (Karışıklık Matrisi)	30
3.3.1.1. Tahmin Hatası.....	31
3.3.1.2 Accuracy (Doğruluk) Oranı	31
3.3.1.3 Sensitive, Recall (Duyarlılık).....	32
3.3.1.4 Precision (Hassasiyet).....	32
3.3.1.5 Specificity (Özgüllük).....	32
3.3.1.6 F-1 Score (F-1 Skoru)	32
3.3.1.7 ROC (Receiver Operating Characteristics – İşlem Karakteristik Eğrisi).....	33
3.3.1.8 MCC (Matthews Correlation Coefficient).....	33
3.3.1.9 Log Loss (Log Kaybı).....	34
3.3.1.10 K-Fold Cross Validation (K Katlamalı Çapraz Doğrulama)	34
BÖLÜM 4	35
DENEYSEL UYGULAMALAR VE SONUÇLARI.....	35
4.1 MATLAB GUI İLE İŞİTME TESTİ.....	35
4.2. MAKİNE ÖĞRENMESİ ÇALIŞMALARI	36
4.2.1 16 Kanal İçin Makine Öğrenmesi Algoritmaları Sonuçları.....	37
4.2.2 1-13 Arası Kanallar için Makine Öğrenmesi Algoritmaları Sonuçları ...	39
4.2.3 1-8 Arası Kanallar için Makine Öğrenmesi Algoritmaları Sonuçları.....	41

	<u>Sayfa</u>
BÖLÜM 5	43
SONUÇLAR VE ÖNERİLER	43
KAYNAKLAR	46
ÖZGEÇMİŞ	52



ŞEKİLLER DİZİNİ

Sayfa

Şekil 1.1. İşitme testi esnasında oluşan EEG sinyalleri ile makine öğrenmesi çalışması blok şeması.....	2
Şekil 2.1. İşitme testi arayüzü başlangıç ekranı.	8
Şekil 2.2. Test sonunda temsili duyulan ve duyulmayan sesler.	9
Şekil 2.3. İşitme testi için kullanılan kulaklık.	9
Şekil 2.4. Brain Products V-Amp 16 EEG kayıt aleti ve aksesuarları.	10
Şekil 2.5. EEG cihazı önden yakından görünüşü.	11
Şekil 2.6. EEG kayıt cihazının arka görünümü.	11
Şekil 2.7. Brain Product ActiCap Xpress.	12
Şekil 2.8. Elektrotların uygun yerlere konumlanmasını sağlayan kep.	12
Şekil 2.9. EEG donanımının kişiye uygunlanmış hali.....	12
Şekil 2.10. EEG elektrotlarının kafaya bağlantı şeması.....	13
Şekil 2.11. Recorder yazılımı ayarları.....	14
Şekil 2.12. EEG verilerinin Recorder yazılımı üzerinde görüntülenmesi.....	14
Şekil 2.13. Örnek .amrk dosyası.....	16
Şekil 3.1. KNN algoritması ile bitki sınıflandırma örneği.	27
Şekil 3.2. Destek vektörleri ile optimum hiper düzlem tespiti.	28
Şekil 3.3. Matlab GUI işitme testi arayüzü ile gerçekleştirilen işitme testi sonucu.	33
Şekil 4.1. 16 Kanal makine öğrenmesi performans sonuçları birinci bölümü.	36
Şekil 4.2. Tüm deneysel çalışmalardan elde edilen verilerde P300 etkisi.	37
Şekil 4.3. 16 kanal makine öğrenmesi ROC eğrileri sonuçları	38
Şekil 4.4. 13 kanal makine öğrenmesi ROC eğrileri sonuçları	40
Şekil 4.5. 8 kanal makine öğrenmesi ROC eğrileri sonuçları	42
Şekil 5.1. LGBM algoritmasının kanal sayısına göre ROC eğrileri.....	45

ÇİZELGELER DİZİNİ

	<u>Sayfa</u>
Çizelge 2.1. Saf EEG verisi.	17
Çizelge 2.2. S etiketlerinin konumları.	18
Çizelge 2.3. Makine öğrenmesi algoritmaları için uygun veri formatı.	19
Çizelge 3.1. Count Vectorizer çalışması.	23
Çizelge 3.2. Count Vectorizer tamamlanmış hali.	24
Çizelge 3.3. Confision Matrix (Karışıklık Matrisi).	30
Çizelge 3.4. 5-Fold Cross Validation.	34
Çizelge 4.1. 16 Kanal EEG verisi için makine öğrenmesi sonuçları ilk bölümü.....	37
Çizelge 4.2. 16 Kanal EEG verisi için makine öğrenmesi sonuçları ikinci bölümü..	39
Çizelge 4.3. 13 Kanal EEG verisi için makine öğrenmesi sonuçları ilk bölümü.....	39
Çizelge 4.4. 13 Kanal EEG verisi için makine öğrenmesi sonuçları ikinci bölümü..	39
Çizelge 4.5. 8 Kanal EEG verisi için makine öğrenmesi sonuçları ilk bölümü.....	41
Çizelge 4.6. 8 Kanal EEG verisi için makine öğrenmesi sonuçları ikinci bölümü...	41
Çizelge 5.1. LGBM algoritmasının kanal sayısına göre başarı metrikleri.....	45

SİMGELER VE KISALTMALAR DİZİNİ

SİMGELER

dB : desibel

Hz : heartz

kHz : kiloheartz

max : maksimum

ms : milisaniye

t : zaman

KISALTMALAR

AUC	: Area Under Curve (Eđri Altındaki Alan)
AUX	: Auxiliary (Yardımcı-Destek)
DTC	: Decision Tree Classification (Karar Ağaçları Sınıflandırma)
DVD	: Digital Versatile Disc (Çok Amaçlı Sayısal Disk)
EEG	: Electroencephalography (Elektroensefalografi)
EN	: Elman Naural Network (Elman Sinir Ađı)
GUI	: Graphical User Interface (Grafiksel kullanıcı arayüzü)
KNN	: K-Nearest Neighbors (K En Yakın Komşular)
KNN	: K-Nearest Neighbors (K-En Yakın Komşular)
LCD	: Liquid Crystal Display (İnce Tabakalı Transistör)
LED	: Light Emitting Diode (Işık Yayan Diyot)
LGBM	: Light Gradient Boosting Machine (Hafif Gradyan Güçlendirme Makinesi)
LGBMC	: Light Gradient Boosting Machine Classification (Hafif Gradyan Güçlendirme Makinesi Sınıflandırması)
LPT	: Lineprinter (Paralel Port - Paralel Kapı)
LR	: Logistic Regression (Lojistik Regresyon)
MLP	: Multilayer Perceptron Neural Network (Çok Katmanlı Algılayıcı Sinir Ađı)
RFC	: Random Forest Classification (Rastgele Orman Sınıflandırma)
ROC	: Receiver Operating Characteristics (İşlem Karakteristik Eğrisi)
SGDC	: Stochastic Gradient Descent Classification (Stokastik Gradyan İnişi Sınıflandırması)
SVC	: Support Vector Classification (Destek Vektör Sınıflandırması)
SVM	: Support Vektör Machine (Destek Vektör Makinesi)
TFIDF	: Term Frequency-Inverse Document Frequency (Terim Sıklığı-Ters Doküman Sıklığı)
USB	: Universal Serial Bus (Evrensel Seri Veriyolu)

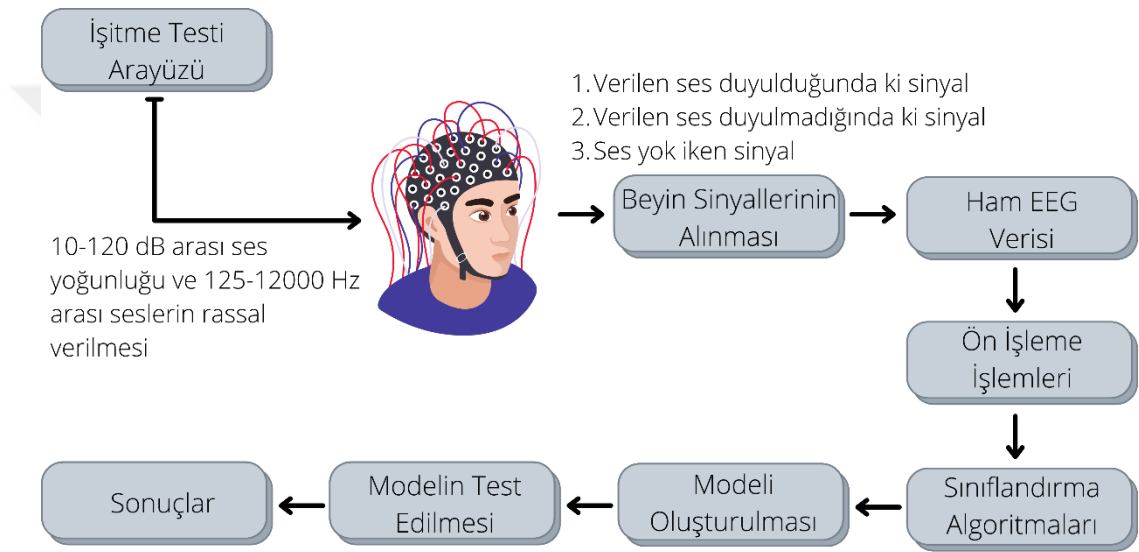
BÖLÜM 1

GİRİŞ

Günümüzde işitme kaybı her yaştan insanda görülen kronik bir hastalık haline gelmiştir. Dünya Sağlık Örgütü'ne göre dünyada 466 milyon insanda işitme kaybı bulunuyor ve bu kişilerin %75'i gelişmekte olan ülkelerde yer alıyor [1]. İşitme bozukluğu ölümcül bir hastalık olmasa da depresyon, iletişim problemleri, günlük hayatta fonksiyonel problemler gibi sorunlara neden olabilmektedir [2]. Yapılan araştırmalara göre, işitme kaybı tespiti yapılan ve sonrasında işitme cihazı kullanan kişilerin, sonraki evrelerde tespit edilip cihaz kullanmaya başlayanlara göre daha fazla fayda sağladığı görülmüştür [3]. Bu nedenle işitme testi ihmal edilmeden yapılmalıdır. Saf ses odyometri testi, işitme değerlendirmesinin temelidir [4].

Günümüzde işitme testleri odyolog tarafından bir klinikte ses geçirmez bir kabinde yapılmaktadır. Test sırasında kişiye yoğunluğu 0 dB ile 110 dB arasında olan farklı seviyelerde (250 Hz, 500 Hz, 1000 Hz, 2000 Hz, 4000 Hz, 8000 Hz) saf tonda sesler yalıtımlı kulaklık aracılığı ile verilir [5]. Birey, verilen sesi duyduğunu belirtirse ses yoğunluğu 10 dB azaltılır, aksi durumda 5 dB arttırılır. Bu işitme testi prosedürüne Hughson Westlake Yöntemi adı verilir [6]. İşitme testi esnasında sesler verildiğinde eğer duyulmuş ise beyinde bir etki oluşturacaktır. Ses verilmediğinde ve verilen sesler duyulmadığında beyinde bir etkinin oluşması beklenmeyecektir. Bu çalışmada ise tasarlanan arayüz ile beyinde etki uyandırmak adına sinyaller rassal olarak verilmiştir. Sesler, genlik ve frekans olarak rastgele bir şekilde kişiye verilmiştir. Kişi duydu ise; o frekansa ait daha yüksek dalga boylu sesleri de duyacağı düşünülerek o sesler verilmemiş ama daha düşük yoğunlukta sesler verilmiştir. Kişi duymadı ise; duymadığı sesin frekansında, daha düşük yoğunluktaki sesleri de duymayacağı düşünülerek verilmemiş ama daha yüksek yoğunluktaki sesler verilmiştir.

Bu tez çalışmasında, Şekil 1.1’de blok şeması gösterilen sistem tasarlanmıştır. İşitme testi kullanıcı arayüzü MATLAB GUI ortamında hazırlanmış olup ham EEG verilerinin önışlemesi ve makine öğrenmesi algoritmalarının çalıştırılması PYTHON programlama dili ile gerçekleştirilmiştir. İşitme testi arayüzü ile test yapılan kişiye rassal olarak ses sinyalleri verilmiş ve kişinin duyduğu seslere butona ya da klavyeden bir tuşa basması beklenmiştir. Bu sırada oluşan beyin sinyallerini gerçek zamanlı alabilmek ve işleyebilmek için Brain Products’ın V-Amp 16 model ürünü kullanılmış ve böylece kendi EEG verilerimiz üzerinde çalışılmıştır.



Şekil 1.1. İşitme testi esnasında oluşan EEG sinyalleri ile makine öğrenmesi çalışması blok şeması.

Test esnasında verilen sesler, beyinde bir değişime yol açmış ve bu durum EEG sinyallerine yansımıştır. Ham EEG sinyal verileri, duyulan ve duyulmayan seslere göre ayrıştırılmış ve sonrasında önışlemeye tabi tutulmuştur. Model başarısını daha iyi ölçümleyebilmek adına, işlemler 16 kanal, 13 kanal ve 8 kanal EEG verisi kullanılacak şekilde olarak üç şekilde denenmiştir. Oluşturulan modelin başarı oranlarını test etmek için eğitim verisinin %30’u kullanılmıştır. Modellerin tahmin başarılarının tespiti için, K katlamalı çapraz doğrulama (k-fold cross-validation), karışıklık matrisi (confusion matrix), doğruluk oranı (accuracy), MSE, duyarlılık oranı (recall), hassasiyet oranı (precision), özgüllük oranı (specificity), F1 skor, MCC, log-kayıp (logloss) yöntemleri kullanılmıştır.

Bu tez çalışmasında, Bölüm 1’de gerçekleştirilen sistem blok şeması üzerinden konuya giriş yapıldıktan sonra literatür taraması sonucunda elde edilen bilgilere yer verilmiştir. Bölüm 2’de ise kullanılan materyal ve metotlar hakkında bilgi verilmiştir. Bölüm 2’de ayrıca işitme testi arayüzü, EEG kayıt cihazı, Recorder yazılımı ve işaretleyicilere de değinilmiştir.

Bölüm 3’te veri önışleme, sınıflandırma algoritmaları ve test adımlarını içeren makine öğrenmesi ortamı hakkında bilgi verilmiştir. Bölüm 4’te işitme testi arayüzü ile gerçekleştirilen deneysel uygulamalar ve EEG kayıt cihazı ile kaydedilen beyin sinyallerinin makine öğrenmesi algoritmalarının sonuçları hakkında bilgi verilmiştir. Bölüm 5’te ise tez çalışmasının genel bir değlendirmesi yapılmıştır.

Literatürde bu konuda yapılan çalışmaların bazıları şunlardır:

Marwa Gargouri ve arkadaşları, bireysel olarak işitme testi gerçekleştirmek adına taşınabilir bir sistem geliştirilmiştir. Ses sinyalini üretmek ve test sonuçlarını değlendirmek için Raspberry Pi3 B + kartı kullanmışlardır. Bilgisayar özelliklerini sağlayan bu karta kulaklık ve fare bağlanarak işitme testini gerçekleştirmişlerdir. Sağ kulak test edilirken sol kulağa ses verilmemiş, sol kulak test edilirken sağ kulağa ses verilmemiştir. Başlangıç olarak 40 dB ve 1000 Hz ses sinyali vermişler ve kişi duyduğunu belirtmeyene kadar 20 dB’lik adımlarla desibel değeri azaltılmış ve ardından duyduğunu belirtene kadar 10 dB’lik adımlarla desibel arttırmışlardır. Kişinin duyduğu sinyallere göre işitme testini tamamlamışlardır [7].

İlay Saka yaptığı çalışmasında, işitme testini hem hazır bir mobil uygulama ile hem de işitme engeli ölçeği ile gerçekleştirmiştir. İşitme engeli ölçeği, 25 maddeden oluşup cevap olarak “Evet”, “Hayır”, “Bazen” seçenekleri bulunmaktadır. Her bir “Hayır” cevabı için 0 puan, “Bazen” cevabı için 2 puan, “Evet” cevabı için 4 puan verilmektedir. Tüm puanlar toplandığında 0-17 arası için “engel yok”, 18-42 arası için “hafif-orta derecede engel”, 43’ten büyük değerler için “belirgin engel” değlendirilmesi yapılmıştır. Mobil uygulama ile yaptığı çalışmada, bireylere farklı dB ve frekanslarda sesler verilmiş ve kişinin duyup duymadığını belirtmesi beklenmiştir. İlkay Saka yaptığı çalışmada, işitme cihazının kullanımının yapılan

testlerin sonuçlarına bir etkisi olmadığını tespit etmiş ve kendi kendine yapılabilen işitme testi uygulamalarını işitme cihazı kullanımından bağımsız bir şekilde uygulanabileceğini belirtmiştir [8].

Sadık Özçelik ve arkadaşları yaptıkları çalışmada, müzik eğitiminin müzikal işitmeye ve algılamaya etkisini istatistiksel ölçümler ve sinir ağları analizi ile göstermişlerdir. Çalışmada Eğitim Fakülte'sine giren öğrencilere odyometrik, tek tonlu dikey işitme, çok tonlu yatay işitme, melodi ve ritim işitme testleri yapılmıştır. Öğrenciler iki yıl sonunda aynı testlere tabi tutulmuş ve müzik bölümü öğrencileri ile diğer bölüm öğrencilerinin tahmini sinir ağları yöntemi ile gerçekleştirilmiştir. Sırasıyla müzik bölümü öğrencilerini %92, diğer bölüm öğrencilerini %88 başarı oranıyla sınıflandırdığını göstermişlerdir [9].

Rajkumar S. ve arkadaşlarının yaptığı çalışmada, yapay sinir ağlarını yöntemlerini kullanarak işitme kaybı değerini ve uygun işitme cihazı kazanç değerini tespit etmişlerdir. İşitme engellerin bireylerin %86'sının, yapılan çalışmaların önerdiği kazanç değerine sahip işitme cihazlarından memnun kaldıkları belirtilmiştir [10].

Franz Furbass ve arkadaşları çalışmalarında, derin öğrenme ile geliştirdikleri algoritma ile bireylerin EEG verisi üzerinden epilepsi hastası olup olmadığını %80 başarı ile doğru tahmin etmişlerdir [11].

Su-Lim Tan ve arkadaşları, saf ton odyometri testi ile konuşma tanıma testini birleştiren ve bunu bir mikrodenetleyici ile tasarlayan bir çalışma gerçekleştirmişlerdir. Konuşma tanıma testi, gürültülü bir ortamda kişinin konuşmayı anlama yeteneğini ölçen bir testtir. Bu test odyologlar tarafından mikrofon ile konuşarak ya da bir kayıt ortamından hazır konuşmalar dinletilerek yapılmaktadır. Su-Lim Tan ve arkadaşlarının yaptığı çalışmada bu iki testi odyologlardan bağımsız bir şekilde yapan bir sistem geliştirmişlerdir [12].

Ykhlef Fayçal ve arkadaşları yaptıkları çalışmada, standart bir ses kartına sahip olan bir bilgisayar yardımı ile saf ton ses ve beyaz gürültü maskeleyen sinyali kullanılarak işitme testi sistemi oluşturmuşlardır. Beyaz gürültü, bütün frekanslarda aynı akustik

enerjiye sahip geniş bantlı gürültü türüdür. Beyaz gürültü, dikkat bozacak, rahatsızlık verecek ve odaklanmayı önleyecek sesleri maskeleyen özelliği taşımaktadır. Yapılan çalışmada gerçek bir odyometre cihazı ile üretilenlere benzer sesler üretebilen ve hastaların işitme kaybı eşiklerini tespit eden bir sistem geliştirilmiştir [13].

Muhammad Yeamin Hossain ve arkadaşları yaptıkları çalışmada, insanlarda motor-nöron davranışı tespiti için EEG sinyalleri kullanarak, makine öğrenmesi sınıflandırma yöntemlerinden SVM (Support Vector Machine) algoritması ile yapay zeka modellemesi oluşturulmuştur [14].

N. Sriram yaptığı çalışmada, EEG sinyalleri ile otonom bir şekilde işitme kaybını tespit eden bir çalışma ortaya koymuştur. Çalışmayı yenidoğan bebekler üzerinde test etmiş ve işitsel uyarıların beyinde etkilerini incelemiştir. Toplanan EEG verileri üzerinde ileri besleme ve geri besleme sinir ağı modelleri kullanılmıştır. Çalışmalar, Multilayer Perceptron Neural Network (MLP) ve Elman Neural Network (EN) yapay sinir ağları sınıflandırma algoritmaları ile yürütülmüştür. EN algoritmasının öğrenme süresinin MLP yöntemine göre daha uzun sürse de sınıflandırma tahmin başarısının daha yüksek olduğu görülmüştür [15].

Paulraj M P ve arkadaşlarının yaptığı çalışmada, yenidoğan bebeklerin beyinlerinin farklı desibel ve frekanslardaki seslere verdiği tepkileri EEG ile incelenmiştir. Yapay sinir ağları sınıflandırma algoritmalarından MLP ve EN ile yapılan çalışmalar neticesinde sol kulak için %79,99, sağ kulak için %82,78 en yüksek doğruluk oranı tespit edilmiştir [16]

BÖLÜM 2

MATERYAL VE YÖNTEM

Beyinde uyaran oluşturmak için farklı frekans ve desibelde ses sinyalleri, tasarlanan MATLAB GUI üzerinde test yapılan kişiye kulaklık aracılığı ile verilmektedir. Duyulan sesler için kişi tarafından bilgisayarın faresi ile butona ya da klavyeden bir tuşa basılmıştır. Duyulmayan sesler ve iki farklı ses arasında ki bekleme sürelerinde kişi bir işlem yapmamıştır. Bu işitme testi boyunca beyin dalgalarını takip etmek için EEG cihazı kanalları kafatasına yerleştirilmiş ve beyin dalgaları kayıt altına alınmıştır. EEG sinyalleri duyulan ve duyulmayan sinyaller olarak ayrıştırıldıktan sonra makine öğrenmesi için ön işleme çalışmalarına tabi tutulmuş ve sonrasında sınıflandırma algoritmaları ile eğitim ve test işlemleri gerçekleştirilmiştir. Tez çalışmasının bu bölümünde arayüz tasarımları ve kullanılan EEG veri elde etme donanım ve yazılımları anlatılmaktadır.

2.1. İŞİTME TESTİ ARAYÜZÜ

İşitme testlerinin yapılabilmesi için, sesleri istenen formatta üreten ve uygun biçimde test yapılacak kişiye verilmesi gerekmektedir. İşitme testinde kullanılacak seslerin üretilmesi için yazılıma ve ses kartı donanımına, iletilmesi için kulaklığa ihtiyaç vardır. Bu bölümde işitme testi için gerekli olan yazılım ve donanımlar hakkında bilgiler verilmiştir.

2.1.1 MATLAB GUI İşitme Testi Arayüzü

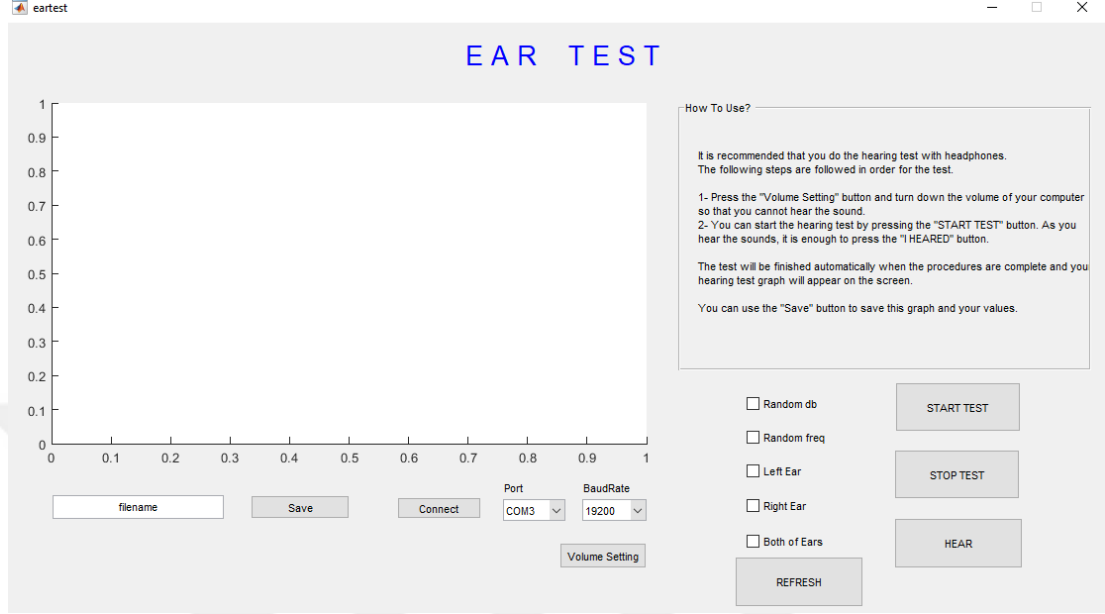
İşitme testleri, işitsel hastalık tespitinde oldukça önemli bir yere sahiptir. Günümüzde işitme testleri, klinik ortamda, sestan izole edilmiş odalarda bir odyolog tarafından yapılmaktadır.

Görevli odyolog, test yapılacak kişiye bir kulaklık ve sesleri duyduğunu belirtebilmesi için bir buton vermektedir. Odyolog kişiye 125 Hz, 250 Hz, 500 Hz, 1 kHz, 2 kHz, 4 kHz ve 8 kHz frekans değerlerine sahip yoğunluğu 0 dB – 110 dB arasında değişen ses sinyalleri vermektedir. Hasta duyduğunu belirtirse, aynı frekansta ki ses için yoğunluğu 10 dB düşürülür ve tekrar cevap beklenir. Hasta duyduğunu belirtmez yani butona basmaz ise aynı frekans değeri için ses yoğunluğu 5 dB arttırılır. Bu işitme testi yöntemi Hughson Westlake Artan Yöntemi olarak bilinmektedir [17]. Bu tez çalışmasında beyinde uyarıları net algılayabilmek adına 125 Hz, 250 Hz, 500 Hz, 1 kHz, 2 kHz, 4 kHz, 8 kHz ve 12 kHz frekanslı ve 10 dB, 20 dB, 30 dB, 40 dB, 50 dB, 60 dB, 70 dB, 80 dB, 90 dB, 100 dB, 110 dB, 120 dB yoğunluğa sahip sesler rassal olarak verilmiştir. Rastgele frekans ve yoğunluk eşleştirmesi yapılmış ve bu test yapılan kişiye verilmiştir.

8 farklı frekans değeri ve 12 farklı desibel değeri olduğu düşünülürse, bir kulak için testin tamamlanabilmesi için 96 sesin kişiye verilmesi gerekmektedir. İki kulak için gereken toplam ses sayısı 192 olacaktır. Sayıca fazla ses vermek hem testlerin uzun sürmesine hem de test yapılan kişinin dikkat dağınıklığına neden olabilmektedir. Bu nedenle kişiye verilen ses, eğer kişi tarafından duyuldu ise; aynı frekans daha yüksek desibelde ki sesleri de duyacağı kabul edilmiş ve o sesler kişiye verilmemiştir. Aynı şekilde, kişinin duymadığı sesler için aynı frekansta ama daha düşük desibele sahip seslerin de duyulamayacağı düşünülerek, bu desibelde ki sesler de kişiye verilmemiştir. İşitme testi arayüzü Şekil 2.1’de görülmektedir.

İşitme testi yapılacak kişiye kulaklık takılmıştır ve duyduğu sesleri belirtebilmesi için klavye ya da fare verilmiştir. Arayüz üzerinde “Random db” ve “Random freq” bölümleri seçilirse, sesler rassal frekans ve yoğunlukta verilecektir. Seçilmez ise standart işitme testi biçiminde sıralı frekans ve desibel değerlerinde sesler verilecektir. Bu tez çalışmasında rassal frekans ve desibel değerine sahip seslerin beyinde daha belirgin bir uyarı oluşturacağı düşünülmüştür. “Left Ear” seçeneği ile sadece sol kulak üzerinde, “Right Ear” seçeneği ile sadece sağ kulak üzerinde, “Both of Ears” seçeneği ile iki kulak sıralı önce sol kulak sonrasında sağ kulak olacak şekilde işitme testi gerçekleştirilmektedir. “START TEST” butonuna tıklanması ile seçilen ayarlar ile test başlatılmaktadır. Bu tez çalışmasında, “Both of Ears”, “Random db” ve “Random

freq” seçenekleri seçilerek, sol kulak ve sağ kulak sıralı bir şekilde rassal frekans ve desibel değerleri ile test edilmiştir.

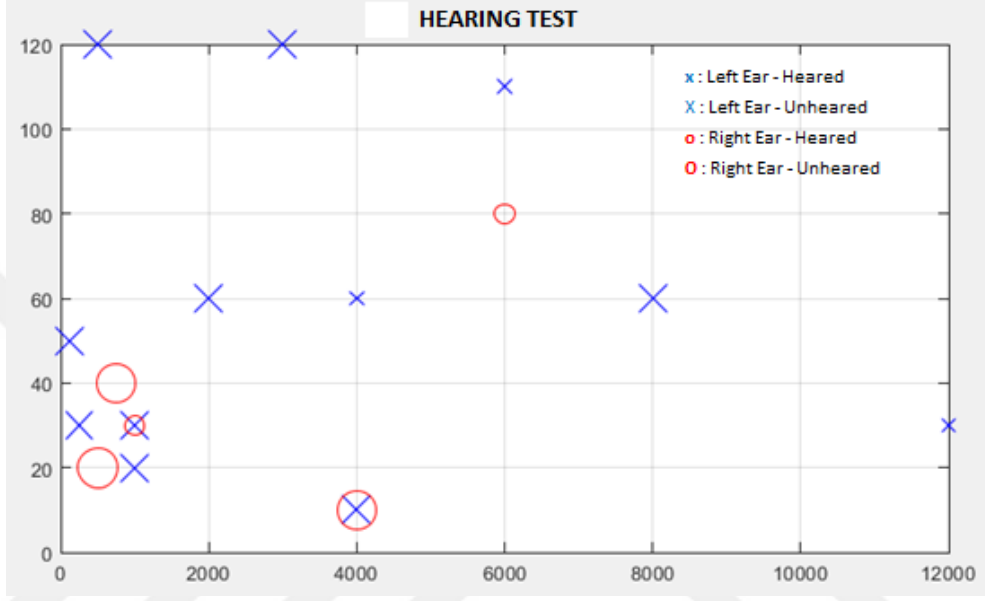


Şekil 2.1. İşitme testi arayüzü başlangıç ekranı.

Test ilk önce sol kulak ile başlamış olup kişi duyduğu seslerden sonra “HEAR” butonuna tıklamış ya da klavyeden bir tuşa basmıştır. Duymadığı sesler ve verilen seslerin aralarındaki bekleme zamanlarında herhangi bir aksiyon almamıştır. Kişiye ilk ses verilmeye başladıktan sonra EEG sinyalleri Recorder yazılımı ile kayıt altına alınmaktadır. Sol kulak için bir ses verildiğinde EEG sinyali üzerinde S1 etiketi işaretlenmiştir. Kişi o sesi duyduğunu belirttiği anda EEG sinyali üzerinde S2 etiketi işaretlenmiştir. Sağ kulak için ise ses verildiğinde S3 etiketi, duyulduğunda ise S4 etiketi işaretlenmiştir. İki ses arasında bekleme süresinin ortasında S5 etiketi işaretlenmiştir. Bu etiketlemelerin işlevleri hakkında ayrıntılı bilgi Bölüm 2.3’te Recorder Yazılımı bölümünde verilmektedir.

Şekil 2.2’de iki kulak için gerçekleştirilen test görülmektedir. Şekil 2.1’de görülen “Both of Ears” seçeneği seçilirse, test iki kulak için yapılır ve sırayla önce sol kulak için test yapılır, sonra sağ kulak için test yapılır. Mavi renkte “X” işareti duyulmayan sesleri, “x” işareti duyulan sesleri ifade etmektedir. Sol kulak testi tamamlandıktan sonra ya da manuel olarak müdahale edildiğinde sağ kulak testine geçilmektedir. Sağ

kulak için gerçekleştirilen testte duyulan sesler için “o” işareti, duyulmayan sesler için “O” işareti kırmızı renk ile gösterilmektedir. Testin tekrarlanması istenmesi ya da yeni bir test yapılması durumunda Şekil 2.1’de görülen ekranda “REFRESH” butonuna tıklanabilmektedir. Bu işitme testi Şekil 2.1’de görülen “Left Ear” ya da “Right Ear” seçenekleri ile tek kulak için de yapılabilmektedir.



Şekil 2.2. Test sırasında temsili duyulan ve duyulmayan sesler.

2.1.2 Kulaklık

Matlab programlama dili ile frekans ve desibel ayarı yapılan ve ses kartı donanımı sayesinde üretilen seslerin, test yapılacak kişiye verilmesi gerekmektedir. Bu projede sesleri kişiye vermek için Trust GXT-350 7.1 Surround USB kulaklık kullanılmıştır. Ancak yazılımın çalışacağı bilgisayarda kablolu-kablosuz olarak tanımlı herhangi bir kulaklığın olması yeterlidir.



Şekil 2.3. İşitme testi için kullanılan kulaklık

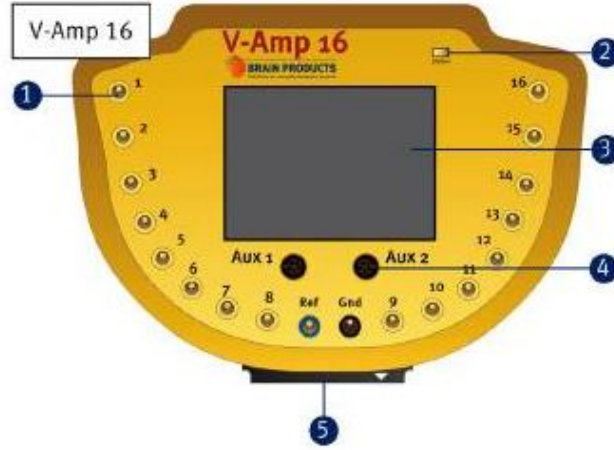
2.2. EEG KAYIT CİHAZI

İşitme testi arayüzü ile gerçekleştirilen işitme testi esnasında EEG kayıtları Brain Products V-Amp 16 EEG kayıt aleti ve Brain Products ActiCap Xpress model kuru elektrot kullanılarak alınmıştır. Şekil 2.4'te EEG kayıtları için kullanılan cihaz ve aksesuarları görülmektedir. Şekil 2.4'te 1 ile gösterilen Brain Products V-Amp 16 EEG kayıt aleti, 2 numara ile gösterilen tetikleme kablosu, 3 numara ile gösterilen dijital veri aktarımı için gereksinim duyulan USB kablosu, 4 numara ile gösterilen kullanım DVD'si görülmektedir [18].



Şekil 2.4. Brain Products V-Amp 16 EEG kayıt aleti ve aksesuarları [18].

Şekil 2.5'te EEG kayıt cihazının ön kısmının daha yakından görünüşü görülmektedir. Şekil 2.5'te 1 numaralı kısım pasif elektrot girişini, 2 numara ile gösterilen kısım cihazın durumunu farklı renklerle gösteren LED bölümünü, 3 numara ile gösterilen kısım LCD ekranı, 4 numara ile gösterilen AUX girişini, 5 numara ile gösterilen kısım ise 40 pinli aktif elektrot girişini ifade etmektedir. [18].



Şekil 2.5. EEG cihazı önden yakından görünüşü [18].

Şekil 2.6’da ise Brain Products V-Amp 16 EEG kayıt cihazının arkadan görüşünü görülmektedir. Şekil 2.6’da 1 ile gösterilen kısım USB kablo girişini, 2 ile gösterilen kısım Trigger (2) girişini, 3 ile gösterilen kısım ise Trigger (1) girişini belirtmektedir.



Şekil 2.6. EEG kayıt cihazının arka görünümü [18].

Bu tez çalışmasında işitme testi esnasında EEG kayıtlarını almak için Şekil 2.7’de 16 elektrotlu Brain Product ActiCap Xpress, Şekil 2.8’de elektrotların kafada uygun yerlerde konumlanmasını sağlayan kep kullanılmıştır. Şekil 2.7’de yer alan Brain Product ActiCap Xpress’te 16 elektrota ek olarak bir tane referans elektrotu ve bir tane toprak elektrotu bulunmaktadır [19]. Şekil 2.8’de görülen elektrotlar Şekil 2.8’de görülen kepe bağlanır ve kep ise elektrotlar saç derisine temas edecek şekilde kafaya, referans ve toprak elektrotları ise kulaklara takılmaktadır. Şekil 2.9’da elektrotların kafaya doğru konumlanmış hali görülmektedir.



Şekil 2.7. Brain Product ActiCap Xpress [19].

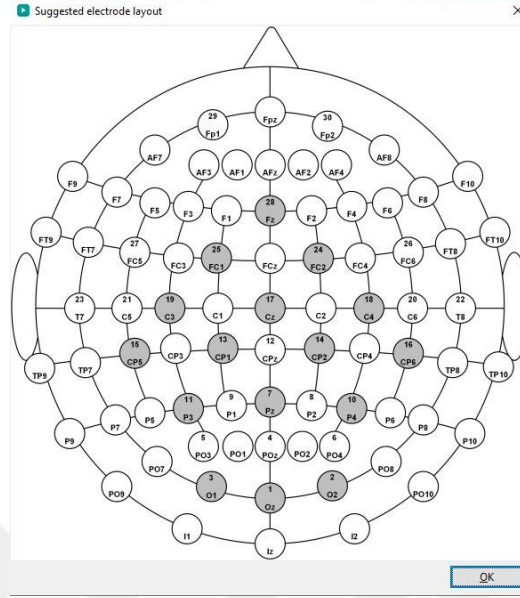


Şekil 2.8. Elektrotların uygun yerlere konumlanmasını sağlayan kep [19].



Şekil 2.9. EEG donanımının kişiye uygunlanmış hali [19].

Şekil 2.10’da bu tez çalışmasında kullanılan 16 kanallı EEG kayıt cihazının elektrotlarının kafada nasıl konumlandırıldığı görülmektedir.



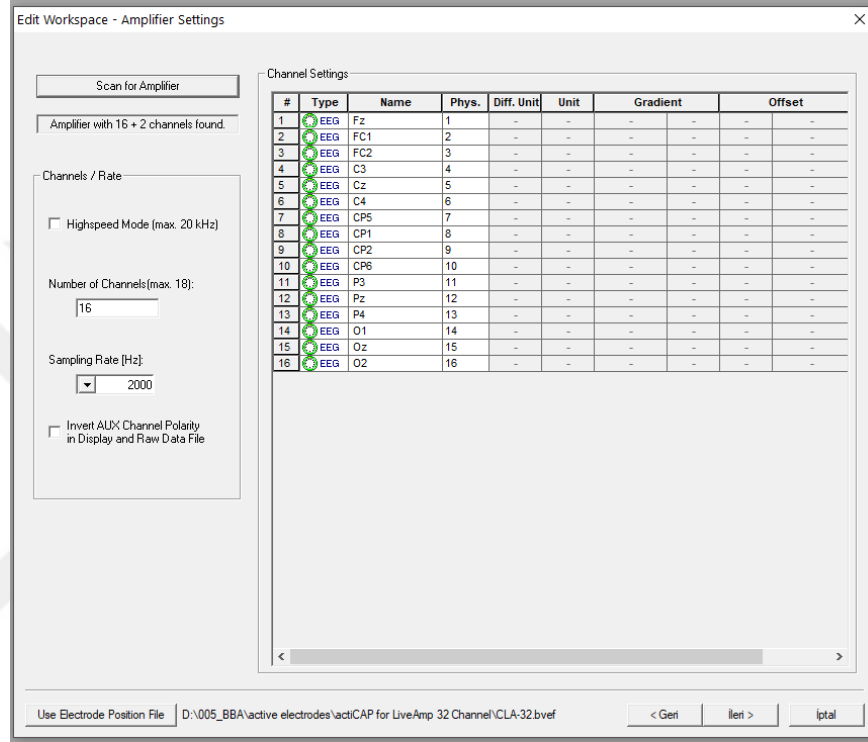
Şekil 2.10. EEG elektrotlarının kafaya bağlantı şeması.

2.3. RECORDER YAZILIMI

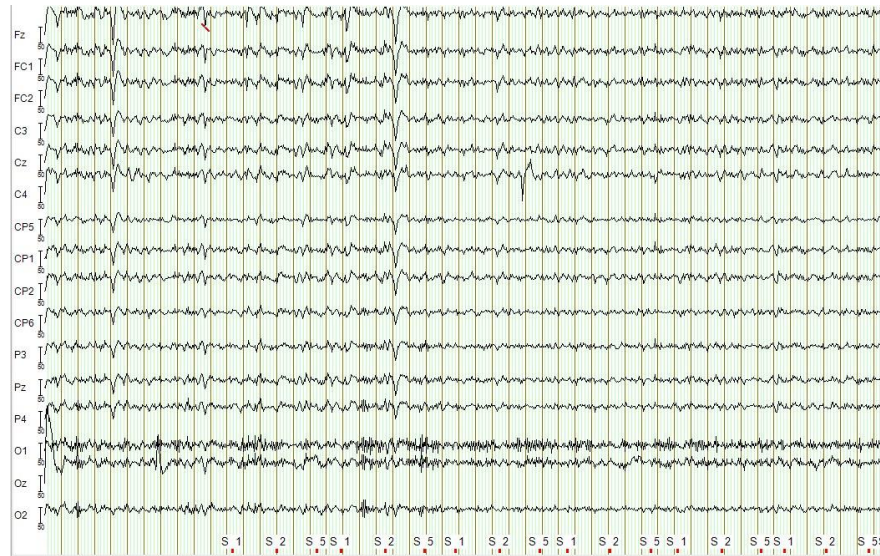
Bölüm 2.2’de anlatılan ve Şekil 2.9’da gösterildiği gibi donanım kurulum süreci tamamlandıktan sonra kaydedilen EEG sinyallerinin bilgisayar ortamında gösterilmesi ve kaydedilmesi için bir yazılıma ihtiyaç vardır. Bu tez çalışmasında kullanılan donanım ile uyumlu olan Recorder adı verilen bir yazılım kullanılmıştır. EEG kayıt cihazı donanımları kişinin kafasına takıldıktan sonra Recorder yazılımı üzerinde Şekil 2.11’deki ayarlamalar yapılır. Bu ayarlara göre 16 kanal saniyede 2000 Hz örnekleme frekansı ile EEG verileri kaydedilecektir.

Recorder yazılımı ile kaydedilen EEG verilerine filtrelemeler yapılabilmektedir. Bu tez çalışmasında 50 Hz Notch (çentik) filtre ve 1-12 Hz band geçiren filtreleme yapılmıştır. Donanım test yapılacak kişiye başarılı bir şekilde bağlanmış ise Şekil 2.12’de ki gibi bir görüntünün Recorder yazılımı üzerinden görüntülenebilmesi gerekmektedir.

Recorder yazılımı ile EEG kayıt cihazı aracılığı ile beyinden alınan sinyaller kaydedilirken, seslerin verildiği ve duyulup-duyulmadığı bilgileri EEG verileri üzerinde işaretlenmelidir. Şekil 2.12’de görülen S1, S2, S3, S4, S5 işaretleri, işitme testi arayüzünden gönderilmiştir. İşaretleyiciler hakkında ayrıntılı bilgiler Bölüm 2.4’te yer almaktadır.



Şekil 2.11. Recorder yazılımı ayarları.



Şekil 2.12. EEG verilerinin Recorder yazılımı üzerinde görüntülenmesi.

2.4. DİJİTAL PORT YAPILANDIRMASI

Matlab işitme testi arayüzü ile yapılan işitme testi esnasında EEG sinyalleri kayıt altına alınmaktadır. Bu amaçla duyulan seslerin belirtilmesi ile eş zamanlı olarak EEG kayıt aletine LPT port aracılığı ile işaretleyici gönderilir. Bu işaretleyiciler sayesinde, sinyalin hangi kulak için olduğu ve duyulup duyulmadığı çıkartılmıştır. Sol kulak için yapılan işitme testinde, ses verildiği anda S1 etiketi, ses duyulduğunda da S2 etiketi gönderilmektedir. Sağ kulak için ise ses verildiği anda S3 etiketi ses duyulduğunda ise S4 etiketi gönderilmektedir. Her iki kulak için gönderilen sesler arasında bekleme süresinde ise S5 etiketi gönderilmiştir. Şekil 2.12’de görüldüğü üzere, etiketler EEG sinyalleri üzerinde işaretlenmiştir. Bu işaretler sayesinde hangi EEG sinyallerinin duyulan sinyal, hangilerinin duyulmayan sinyal olduğu tespit edilmiştir. S1 etiketi ile S5 etiketi arasında S2 etiketi var ise, sol kulak için verilen bu ses duyulmuş, yok ise bu ses kişi tarafından duyulmamıştır. S3 etiketi ile S5 etiketi arasında S4 etiketi varsa sağ kulak için verilen ses duyulmuş, yoksa ise bu ses kişi tarafından duyulmamıştır. S5 etiketi ile bir sonraki sesin verildiğini belirten S1 işareti, sol kulak testi esnasında iki ses arasında ki bekleme zamanını göstermektedir. S5 etiketi ile S3 etiketi arasında ki EEG ise, sağ kulak testi esnasında iki ses arasındaki bekleme zamanını göstermektedir. Bu bekleme zamanında kişinin herhangi bir ses duymadığı düşünülmüştür. İşitme testi bittikten sonra EEG sinyali durdurulmuştur. İşaretleyiciler kayıt durdurulduktan sonra “.vmrk” uzantılı bir dosya formatında kaydedilmektedir. Recorder yazılımı ekranında Configuration bölümünde yer alan Amplifier selmesinden Digital Port Settings penceresi açılmaktadır. Bu pencerede sinyaller yüksek aktif olarak seçilirse ve LPT portunda veriyi oluşturan bitlerden herhangi birisi 0 iken 1 olursa, EEG kayıt cihazı tarafından tetikleme oluştuğu anlamına gelmektedir. Aynı şekilde eğer düşük aktif seçilirse ve LPT portunda veriyi oluşturan bitlerden herhangi birisi 1 iken 0’a geçerse, EEG kayıt aletinden tetikleme oluştuğu anlamına gelmektedir. Algılanan bu bit değerleri “.amrk” dosyasına “Mk10=Stimulus,S 5,61738,0,0” biçiminde kaydolmaktadır. Bu veride Mk10 ifadesi 10. etiket olduğunu, S 5 ifadesi etiket numarasını, 61738 ifadesi EEG sinyalinde kaçınıcı örnek veri olduğunu göstermektedir. Her bir S işareti için aynı biçimde kaydedilmektedir. Şekil 2.13’te görülen “.amrk” örneğinde sağ kulak ve sol kulak için yapılan işitme testi esnasında toplamda 234 işaretleyici bulunmaktadır.

Şekil 2.13'te görülen “.amrk” dosyasında bulunan tüm S işaretleri, Matlab dilinde yazılan bir program ile ayıklanmıştır. Böylece EEG verileri üzerinde hangi aralıklarda ses verildiği, hangi sesleri duyulduğu ve hangi seslerin duyulmadığı tespit edilmiştir.

```
Brain Vision Data Exchange Marker File, Version 1.0
```

```
[Common Infos]
```

```
Codepage=UTF-8
```

```
DataFile=eartest_halil.eeg
```

```
[Marker Infos]
```

```
; Each entry: Mk<Marker number>=<Type>,<Description>,<Position in data points>,  
; <Size in data points>, <Channel number (0 = marker is related to all channels)>  
; Fields are delimited by commas, some fields might be omitted (empty).  
; Commas in type or description text are coded as "\1".
```

```
Mk1=New Segment,,1,1,0,20210407133319323977
```

```
Mk2=Stimulus,S 1,24705,0,0
```

```
Mk3=Stimulus,S 2,29987,0,0
```

```
Mk4=Stimulus,S 5,34851,0,0
```

```
Mk5=Stimulus,S 1,37749,0,0
```

```
Mk6=Stimulus,S 2,43096,0,0
```

```
Mk7=Stimulus,S 5,47868,0,0
```

```
Mk8=Stimulus,S 1,51619,0,0
```

```
Mk9=Stimulus,S 2,56917,0,0
```

```
Mk10=Stimulus,S 5,61738,0,0
```

```
Mk11=Stimulus,S 1,65067,0,0
```

```
Mk12=Stimulus,S 2,70195,0,0
```

```
Mk13=Stimulus,S 5,75157,0,0
```

```
Mk14=Stimulus,S 1,78373,0,0
```

```
Mk15=Stimulus,S 2,83763,0,0
```

```
Mk16=Stimulus,S 5,88460,0,0
```

```
Mk17=Stimulus,S 1,91299,0,0
```

```
Mk18=Stimulus,S 2,96344,0,0
```

Şekil 2.13. Örnek .amrk dosyası.

2.5 VERİ SETİ OLUŞTURMA

Bu tez çalışmasında hazır veriler ile çalışılmamış olup EEG verileri, Matlab GUI ile tasarlanan işitme testi arayüzü programının verdiği seslerin beyindeki etkileri, Brain Products V-Amp 16 EEG kayıt cihazı ile alınarak oluşturulmuştur. Birbirleri ile akrabalık ilişkisi bulunmayan farklı yaş ve yaşam biçimlerine sahip olan 2'si kadın toplam 9 kişi üzerinde çalışmalar gerçekleştirilmiştir.

Her bir kişi, sessiz bir odada ve aynı koşullarda (sabit sıcaklık, oturma düzeni, test cihazları vb.) işitme testine tabi tutulmuştur. İşitme testi esnasında, kişilere rassal genlikte ve frekansta ses sinyalleri verilmiş ve kişilerin duyduğu sesler olduğunda

klavyeden bir tuşa basmış ya da Matlab GUI işitme testi arayüzü ekranında bulunan “HEAR” butona tıklamıştır. İşitme testi esnasında, işitme testi arayüzü ile verilen seslerin beyinde ki etkileri EEG kayıt cihazı ile kaydedilmektedir. Duyulan sesler için EEG sinyallerine işaretleyiciler konumlandırılmıştır. Böylece EEG sinyallerinin hangi kısımlarının duyulan seslere ait olduğu ve hangi kısımların duyulmayan ya da ses verilmeyen kısımları tayin edilmektedir. Duyulan ve duyulmayan sesler ile veri seti oluşturulmuştur. Şekil 2.14’te bir kişi için EEG kayıt cihazından alınan saf veri görülmektedir. Bu saf veriye önce normalizasyon işlemi uygulanmıştır. Her bir değer için ondalık basamak sayısı fazla olduğu için bu durum sınıflandırma algoritmalarında problem yaratabileceği için yuvarlama işlemi gerçekleştirilmiştir. Sonuca etkisini görmek için ondalık basamak 2 ve 3 olacak şekilde yuvarlama işlemi yapılmıştır. Çizelge 2.1’de normalizasyon işlemi yapıp yuvarlama işlemi yapıldıktan sonraki hali görülmektedir. Çizelge 2.1’de görülen veri seti incelendiğinde 16 satır ve 1126400 sütundan oluşmaktadır. Satırlar EEG kayıt cihazının kanalları temsil ederken sütunlar ise zamana bağlı olarak alınan verileri göstermektedir. EEG kayıt cihazı 16 kanallı olduğu için satır sayısı 16 iken örnekleme frekansı 2000 Hz ile alınan sinyaller sütun olarak eklenmektedir.

Çizelge 2.1. Saf EEG verisi.

Kanal	0	1	...	1126399
0	0,0592088617850095	0,10357900755479932	...	0,28128179255872965
1	1,7305413261055946	1,821337454020977	...	2,170032076537609
2	-0,3118655178695917	-0,21945019252598286	...	0,41984720155596733
3	12,815050780773163	12,923024594783783	...	13,246606290340424
4	1,5578551217913628	1,6642935574054718	...	2,350771427154541
5	-4,742591455578804	-4,663854092359543	...	-3,9622310549020767
6	6,641504913568497	6,724922358989716	...	6,923510134220123
7	5,446863919496536	5,538883060216904	...	6,008538231253624
8	1,4677845872938633	1,553320698440075	...	2,2809989750385284
9	-6,5354034304618835	-6,502767652273178	...	-6,046061590313911
10	3,4234941005706787	3,5690195858478546	...	4,46961373090744
11	1,499308180063963	1,6295140609145164	...	2,4672290310263634
12	-6,896305084228516	-6,8995654582977295	...	-6,919075548648834
13	4,940368607640266	5,120294168591499	...	5,78116662800312
14	0,7688967045396566	0,9239344857633114	...	1,6797062009572983
15	-6,605958938598633	-6,561203300952911	...	-6,21660016477108

Çizelge 2.1’de görülen veri setinde, hangi kısımların duyulan ses hangi kısımların duyulmayan ya da ses verilmeyen zamanlara ait olduğu bilinmemektedir. Bunun için Çizelge 2.2’de görülen işaretleyicilerin konumları kullanılmaktadır. Çizelge 2.2’de, 0 satır S1 etiketini, 1. satır S2 etiketini, 2. satır S3 etiketini, 3. satır S4 etiketini, 4. satır S5 etiketini temsil etmektedir. Sütunlar ise etiketlerin konumlarını temsil etmektedir. 0. satır ve 0. sütun değeri olan 15204 sayısı S1 etiketinin EEG verisi üzerindeki konumunu göstermektedir. 0. sütun ve 4. satır ise S5 etiketinin konumunu göstermektedir. S1 etiketi ile S5 etiketi arasında S2 etiketi varsa, kişi verilen sesi duymuş demektir. S2 etiketinin konumu ise 0. sütun 1. satırda görülen 20524 sayısıdır. Dolayısıyla kişi verilen sesi duymuş demektir. S etiketlerinin çalışmada ki kullanım biçimleri Bölüm 2.1’de ayrıntılı bir şekilde anlatılmıştır.

Çizelge 2.2. S etiketlerinin konumları.

	0	1	2	3	...	33
S1	24705	37749	51619	65067	...	458820
S2	29987	43096	56917	70195	...	463890
S3	540440	554090	566670	579390	...	974860
S4	545780	560720	571830	584480	...	993420
S5	34851	47868	61738	75157	...	468920

Duyulan ve duyulmayan ya da ses verilmemiş zamanlar için EEG sinyalleri S etiketlerine göre ayrıştırıldıktan sonra her bir veri için tüm kanallar birleştirilir ve her bir duyulan ve duyulmayan sinyallere kendi içinde normalizasyon işlemi uygulanır. Sonrasında ondalık basamak sayısı 3 olacak şekilde yuvarlama işlemi yapılmıştır. “Veri” sütunu EEG sinyal değerlerini gösterirken “Durum” sütunu ise o kişinin duyup duymadığını belirtmektedir. 1 değeri sesin duyulduğunu, 0 değeri ise sesin olmadığı ya da sesin duyulmadığını belirtmektedir. Çizelge 2.3’te görülen veri, makine öğrenmesi aşamaları için uygun formattır. Makine öğrenmesi başarılarını arttırmak için, verisine shuffle işlemi(karıştırma) uygulanmaktadır.

Çizelge 2.3. Makine öğrenmesi algoritmaları için uygun veri formatı.

<i>İndeks</i>	<i>Durum</i>	<i>Veri</i>
0	0	'-3.384' '-3.379' '-3.372' '-3.363' '-3.354' '-3.341' '-3.325' '-3.307' '-3.287' '-3.266' '-3.243' '-3.217' '-3.189' '-3.162' ...
1	0	'-0.209' '-0.191' '-0.173' '-0.157' '-0.141' '-0.127' '-0.113' '-0.100' '-0.086' '-0.072' '-0.058' ...
2	0	'-1.126' '-1.138' '-1.150' '-1.160' '-1.168' '-1.177' '-1.185' '-1.193' '-1.198' '-1.202' '-1.205' '-1.206' '-1.207' '-1.207' ...
3	0	'1.401' '1.387' '1.375' '1.363' '1.355' '1.346' '1.339' '1.331' '1.324' '1.318' '1.313' '1.309' '1.305' '1.302' '1.300' ...
4	1	'-0.024' '-0.021' '-0.016' '-0.011' '-0.004' '0.002' '0.008' '0.014' '0.018' '0.023' '0.028' '0.035' '0.043' '0.054' '0.065' '0.077' ...
...
133	1	'1.253' '1.279' '1.306' '1.333' '1.361' '1.391' '1.421' '1.452' '1.483' '1.514' '1.545' '1.573' '1.601' '1.628' '1.656' '1.684' ...

2.5.1 Öznitelik Çıkarımı

Makine öğrenmesi için hazırlanan Çizelge 2.3'te görülen formatta veriler üzerinden öznitelik çıkarımı yapılmaktadır.

2.5.1.1 Zaman Uzayında Öznitelik Çıkarımı

Matlab işitme testi ile verilen seslerin beyinde oluşturduğu etkiyi görmek için yaklaşık 800 ms'lik veri üzerinde filtreleme işlemi gerçekleştirilmiştir. Bu verilerden Eşitlik 2.1 ile Eşitlik 2.9 arasında yer alan formüller ile öznitelikler çıkarılmaktadır.

Genlik: İşitme testi arayüzünden ses verildiği anda başlayıp 700 ms'lik zamanı kapsayan dilimdeki maksimum genlik.

$$s_{\max} = \max(s(t)) \quad (2.1)$$

Gecikme: P300 dalgasının maksimum genliğinin görüldüğü zaman ile sesin verildiği zaman arasındaki farktır.

$$s_{\max} = \max(s(t)) \quad (2.2)$$

Ortalama Sinyal Deęeri:

$$\text{mean} = \left(\sum_1^N x(n) \right) / N \quad (2.3)$$

Standart Sapma:

$$\text{std} = \sqrt{\frac{\sum_1^N (x(n) - \text{mean})^2}{N-1}} \quad (2.4)$$

Karakök Ortalamaları:

$$E1 = \sqrt{\frac{\sum_1^N (x(n))^2}{N}} \quad (2.5)$$

Kök Genlik:

$$E2 = \left(\frac{\sum_1^N \sqrt{|x(n)|}}{N} \right)^2 \quad (2.6)$$

Çarpıklık:

$$E3 = \left(\frac{\sum_1^N (x(n) - \text{mean})^3}{(N-1)(E1)^3} \right) \quad (2.7)$$

Basıklık:

$$E4 = \left(\frac{\sum_1^N (x(n) - \text{mean})^4}{(N-1)(E1)^4} \right) \quad (2.8)$$

Tepe Faktörü:

$$E5 = \frac{\max(x(n))}{E4} \quad (2.9)$$

2.5.1.2 Frekans Uzayında Öznitelik Çıkarımı

EEG sinyallerinin salınım özelliklerinden yararlanılarak frekans uzayında öznitelikler çıkarılmıştır. Eşitlik 2.10, Eşitlik 2.11 ve Eşitlik 2.12 formülleri frekans uzayında öznitelik çıkarımı için kullanılmıştır.

Ortalama Frekans:

$$E1 = \frac{\sum_{k=1}^N \delta(k)}{N} \quad (2.10)$$

Standart Sapma Frekans:

$$E2 = \sqrt{\frac{\sum_{k=1}^N f k s(k)}{\sum_{k=1}^N s(k)}} \quad (2.11)$$

Frekans Merkezi:

$$E3 = (E2)^2 \quad (2.12)$$

BÖLÜM 3

MAKİNE ÖĞRENMESİ ORTAMI

Yapı itibari ile dünyada bilinen en kompleks ve zeki varlık insandır. Buna rağmen, insanlığın karşılaştığı problemlerde, insan yetersiz kalabilmektedir. Bu nedenle insanlık, tarih boyunca karşılaştığı problemlere çözüm bulmak, kendi insani sınırlarını aşabilmek için çalışmalar yapmıştır. İnsan zekasını yansıtan ve insanlar tarafından geliştirilen yapay zeka da bu çalışmalardan birisidir. Alan M. Turing, 1950 yılında yayınladığı “Hesaplayan Makineler ve Zeka” makalesi ile yapay zekanın temellerini atmıştır. Turing, makinelerin zeka yeteneklerini ölçen ve kendi isminin verileceği “Turing Testi” ni oluşturmuştur [20].

Yapay zeka, karşılaşılan sorun bilinir ancak çözüm metodu bilinmiyor iken, sonucu bilinen problemlerden öğrenen ve buna göre çözüm yöntemi üretip, çözümü bilinmeyen problemlerde kullanan, insanlık tarafından üretilmiş sistemler bütünüdür. Başka bir tanımıyla yapay zeka, neden-sonuç ilişkisi çıkarımı, veri üzerinden öğrenerek genelleme yapma, belirlenen unsurları algılama ve tespit etme ve karar verme gibi insani yeteneklerin, bilgisayar tarafından gerçekleştirilmesini sağlayan sistemlerdir [21].

Makine öğrenmesi, insan etkisi olmadan bilgisayarların önceki zamanlara ait veriler üzerinden öğrenip tahmin yapabildiği yapay zeka alanıdır. Makine öğrenmesi (machine learning) kavramı, IBM şirketinde çalışan Arthur Samuel tarafından 1959 yılında ilk kez dile getirilmiştir. 1990 yılından itibaren istatistik bilimi ile bilgisayar bilimlerinin birleşmesi neticesinde makine öğrenmesinin sağlık, eğitim, iş dünyası, teknoloji gibi birçok alanda kullanımı artmıştır [22].

Yapay zeka yöntemleri, sınıflandırma, kümeleme, regresyon, özellik belirleme ve ilişki çıkarımı olmak üzere beş gruba ayrılmaktadır [23]. Bu tez çalışmasında sınıflandırma çalışmaları yapay zekanın makine öğrenmesi dalı kullanılmıştır.

3.1. VERİ ÖN İŞLEME

Birçok makine öğrenmesi algoritmalarında istatistiksel modeller daima vektörler ve matrisler ile çalışmaktadır. Gerçek dünya verilerini makine öğrenmesi algoritmalarına girdi olarak kullanabilmek için uygun formata dönüştürmek gerekmektedir. Bu çalışmada Count Vectorizer (Sayma Vektörü) ve TFIDF (Term Frequency Inverse Document Frequency – Terim Frekans Ters Belge Frekans) yöntemleri kullanılmıştır.

3.1.1 Count Vectorizer (Sayma Vektörü)

Bu çalışmada EEG sinyalleri normalize edildikten sonra text (yazı) formatına dönüştürülmüştür. Text formatına dönüştürülen verilerin makine öğrenmesi algoritmalarında çalıştırılabilmesi için vektörleştirmek gerekmektedir. Count Vectorizer, metnin tamamında geçen her bir kelimenin sıklığına bağlı olarak bir metni bir vektöre dönüştürmek amacıyla kullanılır [24].

Çizelge 3.1.'de çizelgenin sütunları olarak benzersiz 8 kelime, satır olarak 3 metin örneği görülmektedir. Çizelgede her hücrede, söz konusu metinde yer alan kelimenin sayısı yer almaktadır. Büyük-küçük harf ayrımı olacağı için çalışma öncesinde tüm kelimeler küçük harfe dönüştürülmüştür. Sütunda yer alan kelimeler ise alfabetik olarak sıralanmaktadır.

Çizelge 3.1. Count Vectorizer çalışması.

	<i>bir</i>	<i>bu</i>	<i>çalışma</i>	<i>kaynak</i>	<i>metin</i>	<i>örnek</i>	<i>vektör</i>	<i>zaman</i>
1. Metin	0	1	1	0	1	1	0	0
2. Metin	1	0	0	1	1	2	1	0
3. Metin	1	1	1	1	2	0	0	1

Count Vectorizer esnasında veriler kelime olarak tutulmamaktadır. Bu nedenle Çizelge 3.1.'in sütunlarında görülen kelimelere sayısal indeksler verilmektedir. Çizelge 3.2.'de görüldüğü hali ile metinlerin vektörize edilme süreci tamamlanmıştır.

Çizelge 3.2. Count Vectorizer tamamlanmış hali.

0	1	2	3	4	5	6	7
0	1	1	0	1	1	0	0
1	0	0	1	1	2	1	0
1	1	1	1	2	0	0	1

3.1.2 TFIDF Yöntemi

TFIDF, bir terimin döküman içerisindeki önemini gösteren ağırlık faktörüdür. Aynı zamanda TFIDF kavramı, bir metin dizisindeki bir kelimenin metin ile ne kadar alakalı olduğunun hesaplanması olarak da tanımlanabilmektedir. Eşitlik 3.1 ve Eşitlik 3.2'de görülen formüller ile hesaplanmaktadır [25].

$$TFIDF(t,d)=TF(t,d)*IDF(t) \quad (3.1)$$

$$IDF(t) = \log\left(\frac{n+1}{DF(t)+1}\right) + 1 \quad (3.2)$$

3.2. SINIFLANDIRMA

Sınıflandırma, geçmiş verilerin sınıflarının bilindiği durumlarda, bilgisayar tarafından öğrenilerek makine öğrenmesi modeli oluşturulması, yeni gelen ve sınıfı bilinmeyen verinin hangi sınıfta bulunacağını bu model aracılığı ile tespit etme işlemidir.

Makine öğrenmesi çalışmaları, eğitim (train) ve test (test) ismi verilen iki adımdan oluşur. Eğitim aşamasında, bulunması hedeflenen sonuçların bilindiği veri setleri üzerinden öğrenme modeli oluşturulur. Test aşamasında ise, sonucu bilinen veriler modelden geçirilerek sonuçları tahmin edilir. Tahmin edilen sonuçlar ile gerçek sonuçlar kontrol edilerek başarı oranı tespit edilir [26]. Modelin başarısı kabul

edilebilir düzeylerde ise, bu model kullanılarak sonucu bilinmeyen veriler üzerinde tahminleme çalışması gerçekleştirilir.

Bu tez çalışmasında sınıflandırma algoritmalarından, Naïve Bayes, LGBM (Light Gradient Boosting Machine – Hafif Gradyan Güçlendirme Makinesi), SVC (Support Vector Classification – Destek Vektör Makinesi), DTC (Decision Tree Classification – Karar Ağaçları Sınıflandırma), KNN (K-Nearest Neighbor – K-En Yakın Komşular), LR (Logistic Regression- Lojistik Regresyon), RFC (Random Forest Classification- Rastgele Orman Sınıflandırma) kullanılmıştır.

3.2.1. Naïve Bayes Sınıflama Algoritması

Naive Bayes algoritması, Thomas Bayes tarafından 18. yüzyılda geliştirilmiştir [27]. Eşitlik 3.3'te Naïve Bayes algoritmasının matematiksel gösterimi verilmektedir.

$$P(A|B) = \frac{P(B|A)P(A)}{p(B)} \quad (3.3)$$

Burada $P(A)$ ifadesi A olayının gerçekleşme olasılığı, $P(B)$ ifadesi B olayının gerçekleşme olasılığı, $P(B|A)$ ifadesi A olayı gerçekleştiğinde B olayının gerçekleşme olasılığı anlamına gelmektedir [28].

3.2.2 Hafif Gradyan Güçlendirme Makinesi Sınıflama Algoritması

LGBM, zayıf öğrenmenin oluşacağı veri setlerinde güçlü modeller oluşturan bir algoritmadır. 2017 yılından itibaren Microsoft tarafından açık kaynak olarak yayınlanmıştır [29]. LGBM, tahmin başarısını korurken çalışma süresi ve bellek kullanımı açısından karar ağaçlarının özelliklerini geliştiren bir yapıya sahiptir. Algoritma, histogram tabanlı çalışmalar yürütmesi sayesinde büyük veri işleme konusunda başarılıdır [30]. LightGBM, yaprak tabanlı bir yapıda çalıştığı için, ağaç öncelikli olarak yatayda büyümektedir ve ağaç derinliği fazla artmamaktadır. Bu sayede aşırı öğrenme engellenebilmektedir [31].

3.2.3 Rastgele Orman Sınıflama Algoritması

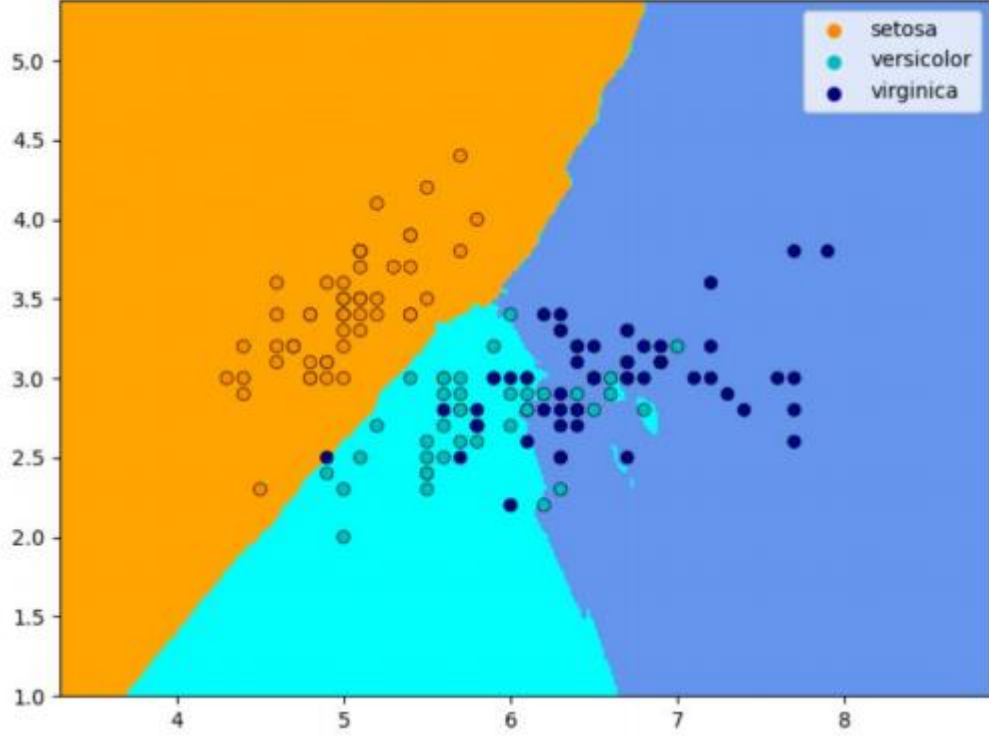
Leo Breiman tarafından 2001 yılında keşfedilen bir oluşumdur. Random Forest (Rastgele Orman) algoritması, tahmin başarı oranının arttırmak için karar ağaçları algoritmasının n defa uygulanması ile kullanılır [32]. Sınıflandırma, regresyon ve öznelik çıkarımı işlemlerinde kullanılmaktadır. RF algoritmasında, ağaç sayısı N ve düğümlerdeki değişken adedini belirleyen M değeri olmak üzere iki tane parametre vardır. Dallara bölmek için tercih edilen değişkenin uygun cut-off değeri olarak bilinen test şartı “gini katsayısı” ile belirlenir. Gini Katsayısı Eşitlik 3.4’deki gibi hesaplanmaktadır [33].

$$\text{GINI}(T) = 1 - \sum_{k=1}^n (S_k)^2 \quad (3.4)$$

GINI indeksi her düğümde hesaplanır ve sıfır olana kadar devam eder. Sıfır olduğunda dallanma sona erer ve hata oranı en düşük olan ağaç en yüksek ağırlığa, hata oranı en yüksek olan ağaç en düşük ağırlığı alır. Sonrasında ağaçların ağırlıklarına göre sınıflar oy kullanır ve bu oylar toplanmaktadır. En yüksek oya sahip ağaç yapısı tespit edilir ve o ağaç tercih edilir [33].

3.2.4. K-En Yakın Komşular Sınıflama Algoritması

KNN (K-Nearest Neighbors) olarak bilinen K En Yakın Komşular algoritması 1966 yılında T.M. COVER ve P.E. HART tarafından oluşturulan “En Yakın Komşu Karar Kuralı” kullanılarak ortaya çıkartılmıştır. KNN, sınıflandırılmaya girmeyen bir verinin, sınıflandırmaya giren veriler arasından hangi sınıflandırma kümesine en yakın olduğunu tespit etmektedir [34]. Şekil 3.1’de KNN algoritması kullanılarak 3 farklı bitki türünün sınıflandırılması ve gerçek türleri görülmektedir.



Şekil 3.1. KNN algoritması ile bitki sınıflandırma örneği [35].

KNN algoritması, Eşitlik 3.5'te görüldüğü üzere eklenecek veri ile sınıfları bilinen veri setinin bileşenleri arasında ki uzaklığın tespit edilip hesaplanması ile çalışmaktadır

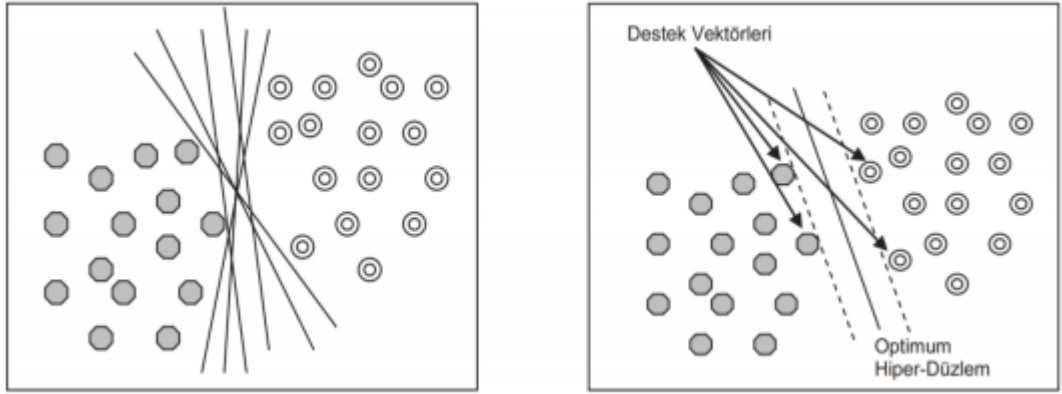
$$d(i,j) = \sqrt{|X_{i1}-X_{j1}|^q + |X_{i2}-X_{j2}|^q + \dots + |X_{ip}-X_{jp}|^q} \quad (3.5)$$

3.2.5. Destek Vektör Makineleri Sınıflama Algoritması

SVC (Support Vector Classification) ya da SVM (Support Vector Machine) olarak isimlendirilen DVM (Destek Vektör Makineleri) sınıflandırma algoritması, hiper düzlemleri kullanan makine öğrenmesi yöntemidir. Diğer doğrusal metotlardan ayrı olarak hiperdüzlemler ayırımı sağlayan aralığın en büyük olacağı biçimde tespit edilmesidir. Yani ayırım, bir tek doğrusal denklemden değil, çok sayıda denklemle ifade edilebilen bir aralıktır [36].

SVC algoritması, 1963 yılında Vladimir Vapnik ve Alexey Chervonenkis tarafından oluşturulmuştur [37]. Bu yöntem kimya [38], fizik [39], biyoloji, teknoloji gibi birçok

alandaki kullanımıdır. SVM, sınıflandırma işlemini gerçekleştirebilmek için sınıfları birbirinden ayıran optimum hiper düzlemi yani karar düzlemi tespit etmektedir [40]. Test aşamasında hangi sınıfa dahil olacağı tahmin edilecek olan veri noktalarının, düzleme olan konumlarına bakılmaktadır. Şekil 3.2’de, soldaki görselde sınıfları birbirinden ayıran çok sayıda hiper düzlem çizilmiştir. Ancak optimum hiper düzlem için sınıflara en yakın noktaların kendisine maksimum uzaklıkta bulunduğu düzlem, Şekil 3.2’de sağ taraftaki görselde gösterilmektedir [41].



Şekil 3.2. Destek vektörleri ile optimum hiper düzlem tespiti [39].

3.2.6. Karar Ağaçları Sınıflama Algoritması

Decision Tree (Karar Ağaçları) algoritması, makine öğrenmesinde sık kullanılan yöntemlerden birisidir. Karar Ağaçları algoritması, regresyon ve sınıflandırma problemleri için kullanılmaktadır. Oluşturma maliyetleri ucuz, yorumlaması kolay ve güvenilirliği yüksek bir algoritma olması sayesinde yaygın olarak kullanılmaktadır. DTC, grubun bütün değerleri aynı sınıf etiketini alana kadar bir kümeleme algoritması ile iterasyon biçimiyle gruplara ayrıştırılması yöntemidir [42].

Karar ağacı sınıflandırma metodu, ağaç yapısını kullanarak tahmin işlemi gerçekleştirir. Bu ağaç yapısında; ağacın düğümlerinde karar değişkenleri, yapraklarında tahmin edilecek hedef değerler bulunmaktadır [43]. Veri setinde ki öznitelikler, karar ağaçlarında düğümleri oluştururlar ve bu düğümler doğru-yanlış ya da evet-hayır şeklinde düğüm sorularını cevaplar ve ikiye bölünür. Veriler bölündükten sonra özniteliklere etki eden etki eden özellik vektörleri incelenir ve

başarı bilgisi yüksek olan düğümler dallanma yapmak için algoritmaya girmektedir [44]. Karar ağaçlarının kararsız yapıda olmaları, verideki en ufak değişiklikte sonucun farklılaşması, aşırı öğrenme gibi problemler olumsuz yönleri bulunmaktadır [45].

Karar ağaçları, arkasında istatistik ve olasılık yapılarının bulunduğu bir makine öğrenmesi yöntemidir. İstatistikte karmaşıklık değeri olarak isimlendirilen entropi değeri çok önemlidir. Entropi, kısaca beklenmeyen durumların yaşanma olasılığına denir. Farklı bir ifade ile entropi, algoritmada yaşanabilecek sapmaları tahmin etmek için gereklidir. Entropinin formülü Eşitlik 3.6'da görülmektedir.

$$H(x) = - \sum S(x) \log S(x) \quad (3.6)$$

Eşitlik 3.6'da bulunan $S(x)$ ifadesi belli bir sınıfa ait grubun oranını göstermektedir ve logaritmik çarpımının ters integrali alınıp hesaplanarak grubun entropisi olan $H(x)$ değeri bulunur. Eşitlik 3.7'de ise karar ağaçları için önemli olan bilgi kazancının formülü bulunmaktadır. Bilgi kazancı, oluşturulan alt grubun karmaşıklık değeri ile tüm veri setinin karmaşıklık değeri arasındaki farktır. Eşitlik 3.7'de S ifadesi bütün veri setini, D ifadesi ise S 'nin yani tüm veri setinin bir alt bölümüdür. V ifadesi ise D 'nin altında bulunan bir karar mekanizmasıdır [44].

$$\text{Kazanç}(S,D) = H(S) - \sum_{V \in D} \frac{|V|}{|D|} H(V) \quad (3.7)$$

3.2.7. Lojistik Regresyon

Logistic Regression (Lojistik Regresyon), Raymond Pearl ve Lowell Reed tarafından 1940 yılında ortaya çıkarılmıştır [46]. Lojistik regresyon, kategorik bir sonucun ihtimalini incelemek için bağımsız değişkenler ile sınıflandırma yapmak için kullanılır [45]. Eşitlik 3.8'de lojistik regresyon eşitliği görülmektedir.

$$W(t) = \Omega \frac{\exp(\alpha + \beta t)}{1 + \exp(\alpha + \beta t)} \quad (3.8)$$

Eşitlik 3.8’de Ω ifadesi, W ’nin doyma seviyesinin üst sınırını, α ifadesi eğrinin x eksenindeki değerini, β ifadesi ise eğrinin eğimini göstermektedir. Bu metot ile olasılık bakımından bağımsız olan değişkenler ile regresyon sonucunda oluşan regresyonlar arasındaki ilişki tespit edilir ve hesaplanır [48]. Lojistik regresyon algoritması, sağlık [49], sosyal bilimler [50], siyasi bilimler [51] gibi alanlarda sıklıkla kullanılmaktadır.

3.3. TEST ADIMLARI

Makine öğrenmesi algoritmaları, eğitim aşamasını tamamladıktan sonra sonucu bilinmeyen veriler ile tahminleme işlemi yapılmaktadır. Sonucu bilinmeyen verilerin tahmin edilen sonuçları tespit edildikten sonra doğruluğu kontrol etmek zordur. Bu tip durumlarda modellerin başarısı test etmek için, eğitim verisinin üçte ikilik bölümü ile eğitim yapılır, kalan üçte birlik bölümü ile test işlemi yapılır. Bu tez çalışmasında da eğitim verileri bölünerek eğitim ve test işlemi yapılmaktadır. Böylece model, eğitime dahil olmayan veriler ile test edilmekte ve başarısı ortaya çıkartılmaktadır.

3.3.1. Confusion Matrix (Karışıklık Matrisi)

Sınıflama modellerinin tahminlerinin doğruluğu test etmek amacıyla kullanılan bir kare matristir. Bu tez çalışmasında birçok makine öğrenmesi algoritması kullanılmıştır ve bu algoritmalar teste tabii tutulur. Test sonuçları kullanıcı tarafından kontrol edilerek çalışma için en uygun algoritma seçilir. Karışıklık matrisinde, modellerin tahminlerinin doğru ve yanlış sayıları gösterilir [52] [53]. Çizelge 3.3’te karışıklık matrisinin genel yapısı görülmektedir.

Çizelge 3.3. Confision Matrix (Karışıklık Matrisi).

		Gerçekte		Toplam
		Pozitif (1)	Negatif (0)	
Tahmin Edilen	Pozitif (1)	TP	FP	TP+FP
	Negatif (0)	FN	TN	FN+TN
Toplam		TP+FN	FP+TN	TP+FN+FP+TN = N

Çizelge 3.3’de yer alan TP, FP, FN ve TN gerçek durum ve yapılan tahminler arasındaki ilişkiyi gösterir [44].

TP: Modelin pozitif tahmin ettiği pozitif durumları ifade eder.

FP: Modelin negatif tahmin ettiği pozitif durumları ifade eder.

FN: Modelin pozitif tahmin ettiği negatif durumları ifade eder.

TN: Modelin negatif tahmin ettiği negatif durumları ifade eder.

Bu çalışmada işitme testi esnasında seslerin duyulması “gerçekte pozitif durum”, duyulmaması “gerçekte negatif durum”, modelin duyulma olarak tahmin etmesi “tahmin edilen pozitif durum”, duyulmama olarak tahmin etmesi “tahmin edilen negatif durum” olarak ifade edilmiştir.

3.3.1.1. Tahmin Hatası

Modelin hatalı olarak tahmin ettiği verilerdir. Eşitlik 3.9’deki formül ile hesaplanmaktadır [54].

$$\frac{FP+FN}{TP+TN+FP+FN} \quad (3.9)$$

3.3.1.2 Accuracy (Doğruluk) Oranı

Toplam olasılık değeri olan 1 sayısından tahmin hatası oranının çıkarılmasıyla bulunan değerdir. Eşitlik 3.10’deki formül ile de hesaplanabilir. Doğruluk oranı, makine öğrenmesi modelinin doğru tahmin ettiği değerlerin oranıdır [55].

$$\frac{TP+TN}{TP+TN+FP+FN} \quad (3.10)$$

3.3.1.3 Sensitive, Recall (Duyarlılık)

Gerçekte pozitif olan durumların hangi oranda model tarafından pozitif olarak tahmin edilmesidir. Eşitlik 3.11'deki gibi hesaplanır [56].

$$\frac{TP}{TP+FN} \quad (3.11)$$

3.3.1.4 Precision (Hassasiyet)

Modelin, çalışmada kullanılan verilerden toplam kaç tanesinin doğru tahmin ettiğini gösterir ve Eşitlik 3.12'deki gibi hesaplanır [57].

$$\frac{TP}{TP+FP} \quad (3.12)$$

3.3.1.5 Specificity (Özgüllük)

Doğru tahmin edilen negatif gerçek değerlerin, tüm negatif gerçek değerlere oranıdır. Eşitlik 3.13'deki formül ile hesaplanır [58].

$$\frac{TN}{TN+FP} \quad (3.13)$$

3.3.1.6 F-1 Score (F-1 Skoru)

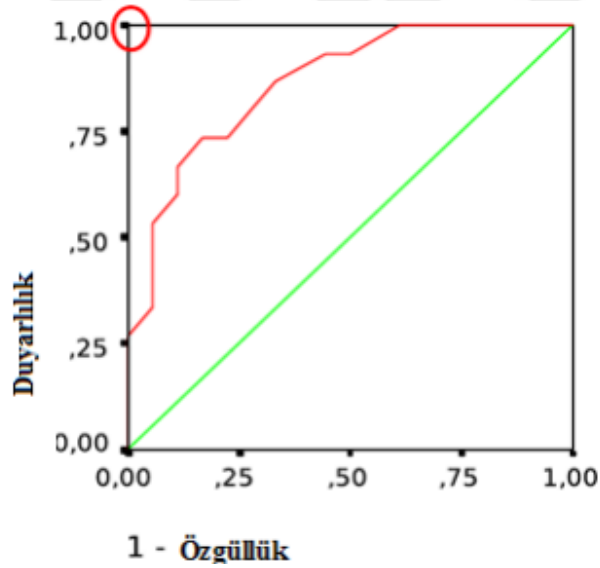
F-1 Score (F-1 Skoru) ya da F-Measure (F-Ölçütü) olarak da isimlendirilen makine öğrenmesi sınıflandırma algoritmalarının başarısını gösteren ve hassasiyet ve geri çağırma değerleri ile hesaplanan bir ölçüt birimidir. Eşitlik 3.14'deki gibi hesaplanır [59].

$$\frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (3.14)$$

3.3.1.7 ROC (Receiver Operating Characteristics – İşlem Karakteristik Eğrisi)

ROC eğrisi, bir veri testine ait duyarlılık ve özgüllük ifadeleri arasındaki ilişkiyi grafiksel olarak gösterir. ROC eğrisi, duyarlılık değerine karşılık yanlış pozitif oranlarının (1-özgüllük) değerlerinin çizilmesi ile elde edilir. Farklı bir ifade ile, TP değeri ile FP değeri arasındaki ilişkinin görsel olarak grafikte gösterimidir. ROC eğrisinin altında kalan alan AUC (Area Under Curve – Eğri Altındaki Alan) ile ifade edilir ve 1'e yaklaştıkça modelin sınıflandırma başarısının arttığı anlaşılmaktadır [60].

Makine öğrenmesi modellerinin değerlendirilmesi için tercih edilen önemli değerlerden birisi olan ROC eğrisi (0,0) ve (1,1) noktaları arasında çizilen bir köşegen şeklindedir. AUC ifadesi 0 ile 1 arasında bir değer alır. AUC, 0.5 değerinden küçük ise sistemin hesaplamalarının değersiz olduğunu göstermektedir [61]. Şekil 3.3'te örnek ROC eğrisi görülmektedir.



Şekil 3.3. Örnek ROC eğrisi [62].

3.3.1.8 MCC (Matthews Correlation Coefficient)

MCC, gerçek değerler ile tahmin edilen değerler arasında bir korelasyon katsayısıdır. MCC değeri, -1 ile +1 arasında bir değer alır. Eşitlik 3.15'te yer alan formül ile

hesaplanmaktadır. MCC değeri, 1'e yaklaştıkça modelin sınıflandırma başarısının arttığı anlaşılmaktadır [63].

$$\frac{TP \times TN - FP \times FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \quad (3.15)$$

3.3.1.9 Log Loss (Log Kaybı)

Lojistik Regresyon Kaybı ya da Çapraz Entropi Kaybı olarak da adlandırılan Log Loss, model tahminlerinin, gerçek değerinden ne kadar farklılık gösterdiğine ek olarak belirsizliği hesaba katmaktadır. Tahmin edilen tahminler, gerçek değerlerden ne kadar saparsa Log Loss değeri o kadar yüksek olmaktadır [64].

3.3.1.10 K-Fold Cross Validation (K Katlamalı Çapraz Doğrulama)

K-Fold Cross Validation (K Katlamalı Çapraz Doğrulama), makine öğrenmesi algoritmalarının doğrulunu test etmek için kullanılan bir yöntemdir. Bu yöntemde eğitim verileri rastgele olarak Çizelge 3.4'de görüldüğü gibi k sayısına bölünmüştür. Her bir iterasyonda bölünen parçalardan bir tanesi test için kullanılır, kalan kısımlar eğitim için kullanılmaktadır [65] [66]. K-Fold Cross Validation yönteminin ana avantajı, verinin her bir bölümünün hem eğitimde hem de testte kullanılmasıdır. Bu da model doğruluğu için daha adil test edilmesini sağlamaktadır [67][68]. Çizelge 3.4'te 5-Katlamalı çapraz doğrulamanın genel yapısı görülmektedir.

Çizelge 3.4. 5-Kat Çapraz Doğrulama

İterasyon-1	Test	Eğitim	Eğitim	Eğitim	Eğitim
İterasyon-2	Eğitim	Test	Eğitim	Eğitim	Eğitim
İterasyon-3	Eğitim	Eğitim	Test	Eğitim	Eğitim
İterasyon-4	Eğitim	Eğitim	Eğitim	Test	Eğitim
İterasyon-5	Eğitim	Eğitim	Eğitim	Eğitim	Test

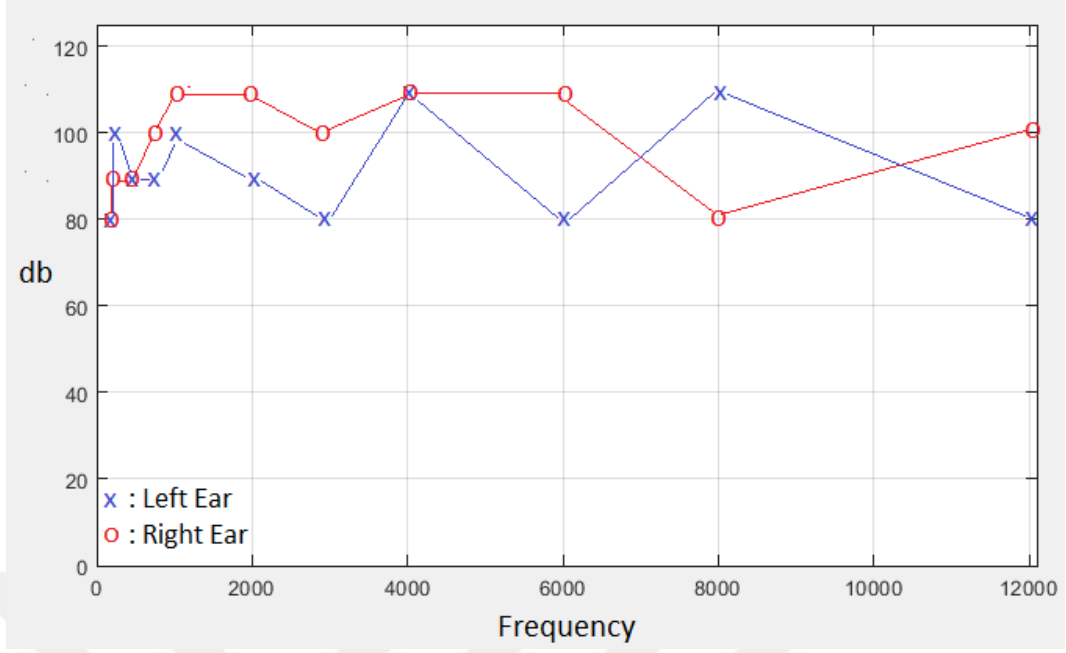
BÖLÜM 4

DENEYSEL UYGULAMALAR VE SONUÇLARI

Bu tez çalışmasında, işitme testi ve eş zamanlı olarak EEG kaydının alınması test ortamında gerçekleştirilmiştir. Sessiz ve sakin bir odada, sabit ve aynı koşullarda (sıcaklık, test aşamaları, test teçhizatı vb.) 9 farklı kişiye Matlab işitme testi arayüzü ile işitme testi yapılmış, her bir sesin ve sessizliğin beyinde ki etkileri EEG kayıt cihazı ile kaydedilmiştir. EEG kayıtları üzerinde veri ön işleme çalışmaları uygulanmış ve sonrasında makine öğrenmesi sınıflandırma algoritmaları denenmiştir.

4.1 MATLAB GUI İLE İŞİTME TESTİ

Tasarlanan Matlab GUI işitme testi arayüzünün çalışması Bölüm 2.1.1’de anlatılmıştır. Klasik odyometri cihazlarında dalgaboyu ve frekans değerleri sıralı bir şekilde verilir. Tasarlanan işitme testi arayüzü ile seslerin dalgaboyu ve frekans değerleri sıralı bir biçimde verilebildiği gibi rassal bir şekilde de verilebilmektedir. Ayrıca sol kulak ve sağ kulak için ayrı ayrı işitme testleri yapılabildiği gibi sıralı bir şekilde de iki kulağın testi yapılabilmektedir. Bu tez çalışmasında işitme testi, sesin dalgaboyu ve frekans ayarları rassal olacak şekilde iki kulak için sıralı bir biçimde gerçekleştirilmiştir. Şekil 4.1’de işitme testi arayüzü ile gerçekleştirilen bir işitme testinin sonucu görülmektedir. Şekil 4.1’de ki işitme testi raporuna göre, her bir frekansta duyulan en düşük genliğe sahip ses sinyali sol ve sağ kulak için ayrı ayrı işaretlenmiştir. Sol kulak mavi renk ve “x” sembolü ile, sağ kulak kırmızı renk ve “o” sembolü ile gösterilmiştir.

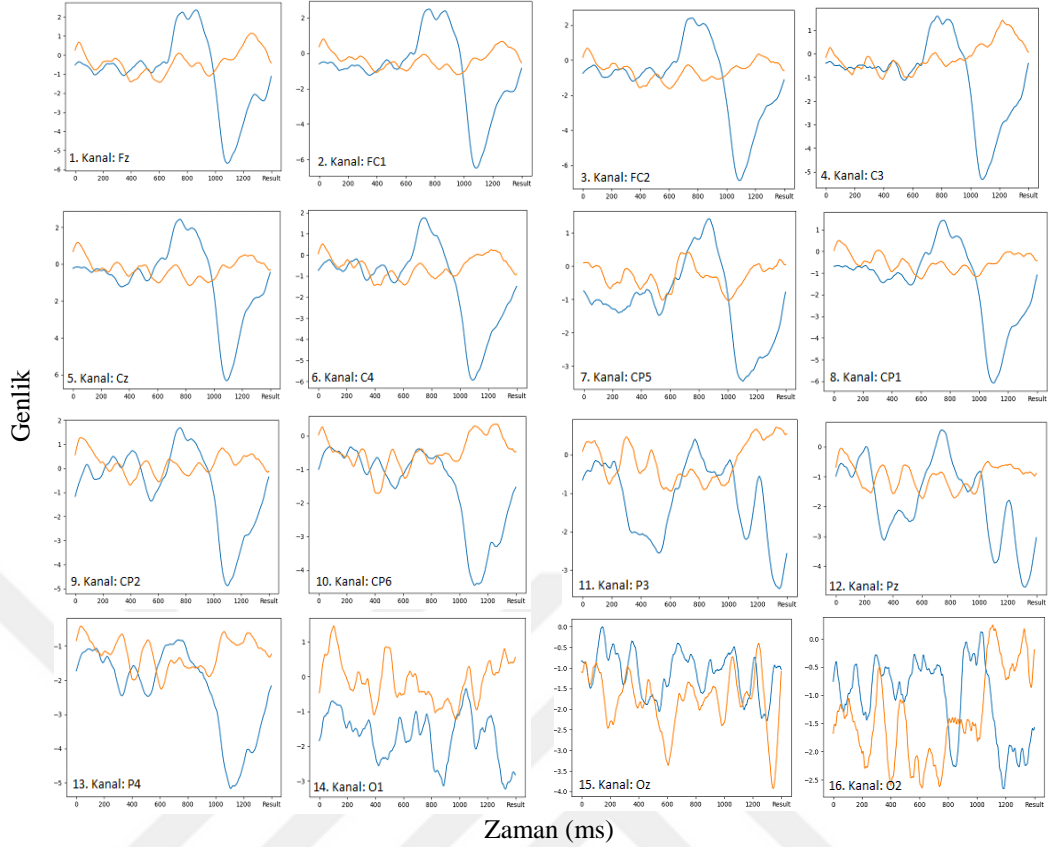


Şekil 4.1. Matlab GUI işitme testi arayüzü ile gerçekleştirilen işitme testi sonucu.

4.2. MAKİNE ÖĞRENMESİ ÇALIŞMALARI

Bölüm 2.1’de ayrıntılı bir şekilde açıklanan MATLAB işitme testi arayüzü ile gönderilen seslerin beyinde oluşturduğu etkisi Bölüm 2.2’de ayrıntılı bir şekilde açıklanan EEG kayıt cihazı ile kaydedilmiştir. Elde edilen saf EEG verileri, işaretleyicilere göre duyulan ve duyulmayan sinyaller ayrıştırılır ve Bölüm 2.5’te açıklandığı üzere makine öğrenmesi için düzenlenmiştir.

Hazırlanan veriseti görselleştirildiğinde, verilen ses uyarılarının EEG sinyalleri üzerinde oluşturmuş olduğu P300 sinyali mavi renk, ses uyarısının olmadığı zaman aralıklarında elde edilen EEG sinyalleri turuncu renk olarak tüm deneysel çalışmalardan elde edilen verilerde P300 etkisi Şekil 4.2’de gösterilmektedir. Şekil 4.2’de görüldüğü üzere 1-13 arasındaki kanallarda duyulan sinyaller ile duyulmayan sinyaller arasında ayrışma varken 1-8 arasındaki kanallarda bu ayrışma daha net olarak görülmektedir. Yapay zeka çalışmaları, 16 kanal için, 1-13 arası kanallar için ve 1-8 arası kanalları için uygulanmış ve değerlendirilmiştir. Hazırlanan veriseti, Bölüm 3.1’de açıklanan veri ön işleme adımlarına tabi tutulduktan sonra Bölüm 3.2’de açıklanan sınıflandırma algoritmaları uygulanmıştır.



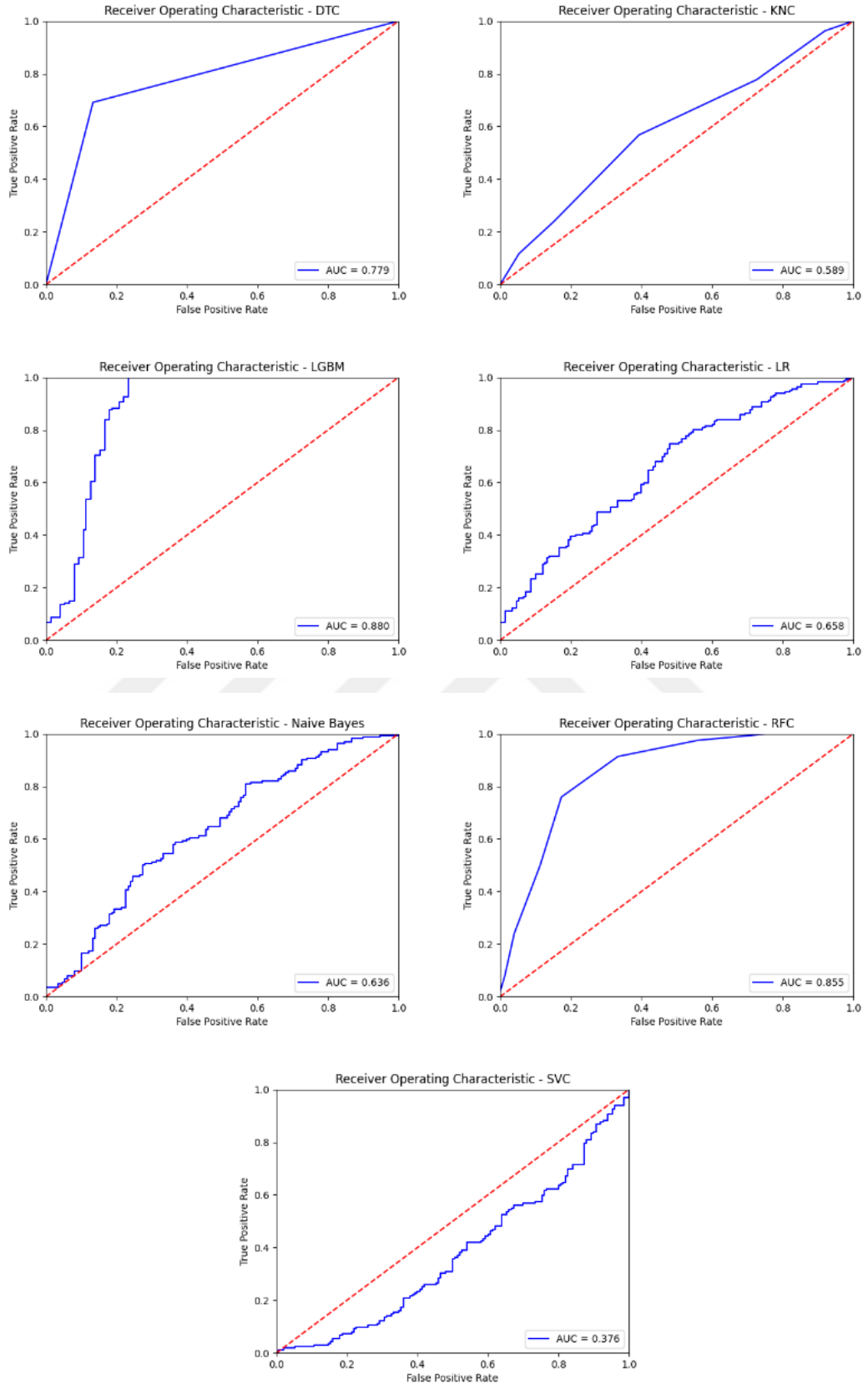
Şekil 4.2. Tüm deneysel çalışmalardan elde edilen verilerde P300 etkisi.

4.2.1 16 Kanal İçin Makine Öğrenmesi Algoritmaları Sonuçları

Matlab işitme testi arayüzünden verilen seslerin beyinde oluşturduğu etkiler, 16 kanallı EEG kayıt cihazı ile kaydedilmektedir. Tüm kanallardan alınan EEG verilerine Bölüm 3.1 ve Bölüm 3.2’de anlatılan makine öğrenmesi yöntemleri uygulanmıştır. Çizelge 4.1 ve Çizelge 4.2’de makine öğrenmesi sınıflandırma algoritmalarının sonuçları görülmektedir.

Çizelge 4.1. 16 Kanal makine öğrenmesi performans sonuçları birinci bölümü.

<i>Algoritma</i>	<i>TN</i>	<i>FP</i>	<i>FN</i>	<i>TP</i>	<i>Accuracy (%)</i>	<i>MSE (%)</i>	<i>Recall (%)</i>	<i>Precision (%)</i>
<i>N.Bayes</i>	142	8	152	10	48,7	51,3	6,2	55,6
<i>LGBM</i>	125	25	28	134	83,0	17,0	82,7	84,3
<i>SVC</i>	150	0	162	0	48,1	51,9	0	-
<i>DTC</i>	130	20	50	112	77,6	22,4	69,1	84,8
<i>KNC</i>	91	59	70	92	58,7	41,3	56,8	60,9
<i>LR</i>	125	25	109	53	57,1	42,9	32,7	67,9
<i>RFC</i>	133	17	81	81	68,6	31,4	50,0	82,7



Şekil 4.3. 16 kanal makine öğrenmesi ROC eğrileri sonuçları.

Çizelge 4.2. 16 Kanal makine öğrenmesi performans sonuçları ikinci bölümü.

<i>Algoritma</i>	<i>F1 (%)</i>	<i>Specificity (%)</i>	<i>MCC</i>	<i>CrossValidation</i>	<i>Logloss</i>
<i>N.Bayes</i>	11,1	94,7	0,018	[50,7 57,2 51,2]	5,965
<i>LGBM</i>	83,5	83,3	0,66	[76,2 77,5 74,3]	0,397
<i>SVC</i>	-	100	0	[50,2 50,1 50,7]	0,692
<i>DTC</i>	76,2	86,7	0,56	[68,0 71,8 64,9]	7,749
<i>KNC</i>	58,8	60,7	0,175	[58,4 55,0 52,6]	2,191
<i>LR</i>	44,2	83,3	0,185	[57,6 57,4 55,0]	0,688
<i>RFC</i>	62,3	88,7	0,416	[68,0 72,5 66,9]	0,519

4.2.2 1-13 Arası Kanallar için Makine Öğrenmesi Algoritmaları Sonuçları

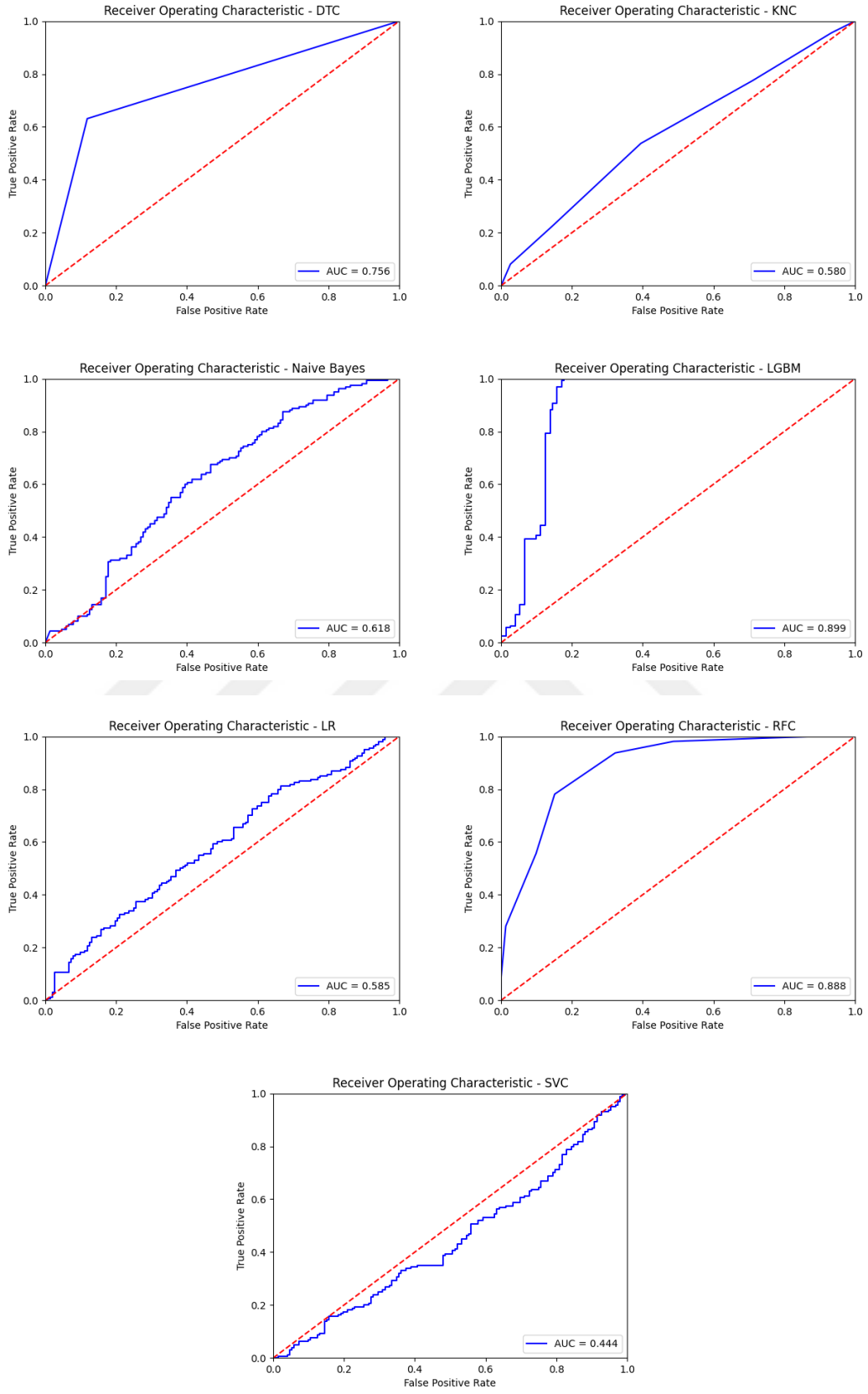
16 kanal üzerinden alınan EEG verilerinde P300 etkisi 1-13 arası kanallarda görülmektedir. Makine öğrenmesi algoritmalarının ilk 13 kanal için çalıştırılması ile aşağıdaki sonuçlar elde edilmiştir.

Çizelge 4.3. 13 Kanal makine öğrenmesi performans sonuçları birinci bölümü.

<i>Algoritma</i>	<i>TN</i>	<i>FP</i>	<i>FN</i>	<i>TP</i>	<i>Accuracy (%)</i>	<i>MSE (%)</i>	<i>Recall (%)</i>	<i>Precision (%)</i>
<i>N.Bayes</i>	132	20	139	21	49,0	51,0	13,1	51,2
<i>LGBM</i>	131	21	29	131	84,0	16,0	81,9	86,2
<i>SVC</i>	152	0	160	0	48,7	51,3	0	-
<i>DTC</i>	134	18	59	101	75,3	24,7	63,1	84,9
<i>KNC</i>	92	60	74	86	57,1	42,9	53,8	58,9
<i>LR</i>	109	43	100	60	54,2	45,8	37,5	58,3
<i>RFC</i>	137	15	71	89	72,4	27,6	55,6	85,6

Çizelge 4.4. 13 Kanal makine öğrenmesi performans sonuçları ikinci bölümü.

<i>Algoritma</i>	<i>F1 (%)</i>	<i>Specificity (%)</i>	<i>MCC</i>	<i>CrossValidation</i>	<i>Logloss</i>
<i>N.Bayes</i>	20,9	86,8	0	[54,8 52,4 49,0]	7,093
<i>LGBM</i>	84,0	86,2	0,681	[76,2 75,4 78,7]	0,383
<i>SVC</i>	-	100	0	[49,5 51,2 48,7]	0,693
<i>DTC</i>	72,4	88,2	0,528	[67,3 68,7 70,7]	8,524
<i>KNC</i>	56,2	60,5	0,143	[51,4 57,6 56,7]	1,885
<i>LR</i>	45,6	71,7	0,098	[51,4 62,2 53,8]	0,690
<i>RFC</i>	67,4	90,1	0,485	[65,1 67,7 67,8]	0,484



Şekil 4.4. 13 kanal makine öğrenmesi ROC eğrileri sonuçları.

4.2.3 1-8 Arası Kanallar için Makine Öğrenmesi Algoritmaları Sonuçları

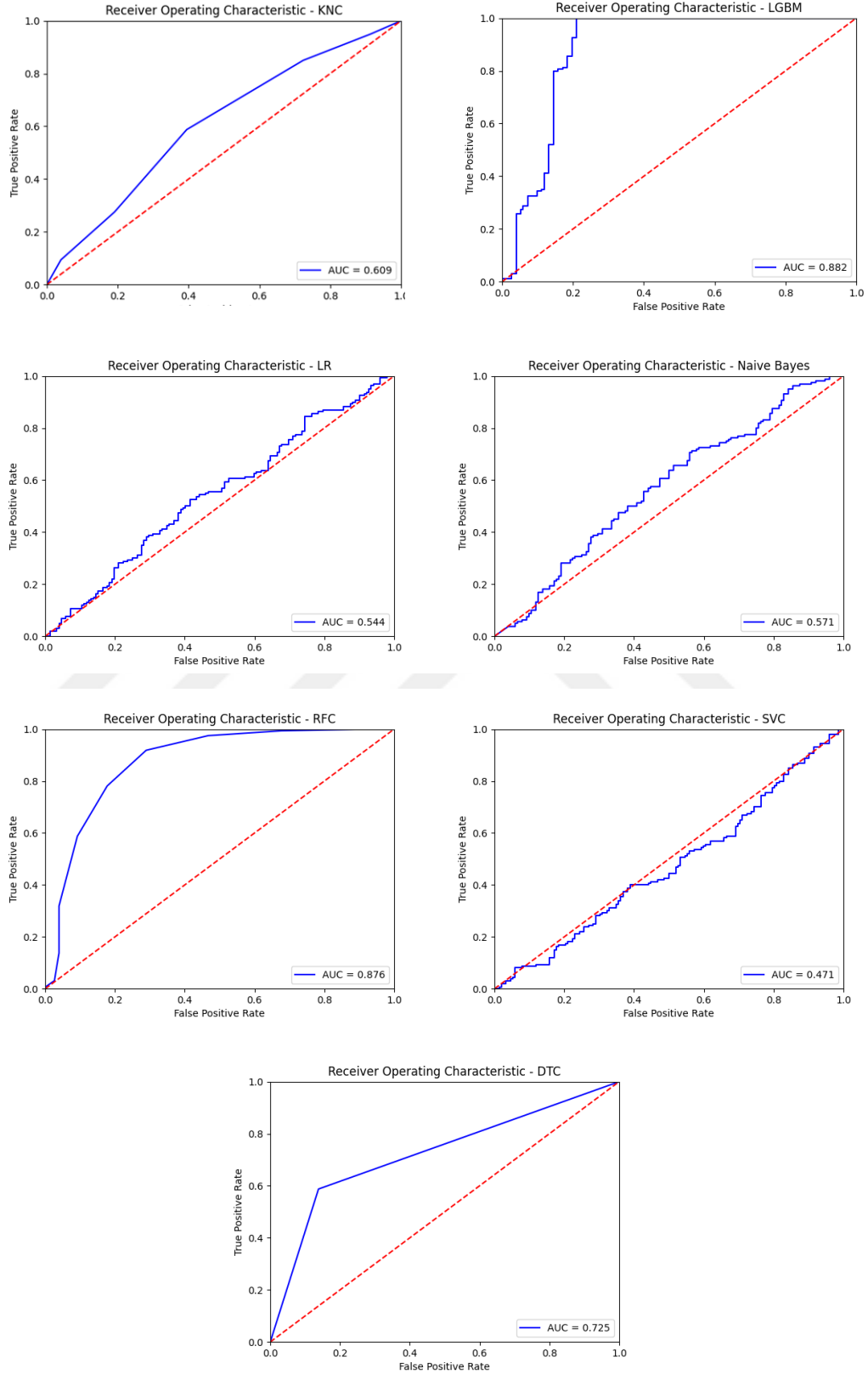
16 kanal üzerinden alınan EEG verilerinde P300 etkisi 1-13. kanallarda görülmesine rağmen 1-8. kanallarda duyulan ve duyulmayan seslerin ayrışması daha belirgin durumdadır. Makine öğrenmesi algoritmalarının ilk 8 kanal için çalıştırılması ile aşağıdaki sonuçlar elde edilmiştir.

Çizelge 4.5. 8 Kanal makine öğrenmesi performans sonuçları birinci bölümü.

<i>Algoritma</i>	<i>TN</i>	<i>FP</i>	<i>FN</i>	<i>TP</i>	<i>Accuracy (%)</i>	<i>MSE (%)</i>	<i>Recall (%)</i>	<i>Precision (%)</i>
<i>N.Bayes</i>	131	21	133	27	50,6	49,4	16,9	56,3
<i>LGBM</i>	124	28	28	132	82,1	17,9	82,5	82,5
<i>SVC</i>	152	0	160	0	48,7	51,3	0	-
<i>DTC</i>	131	21	66	94	72,1	27,9	58,8	81,7
<i>KNC</i>	92	60	66	94	59,6	40,4	58,8	61,0
<i>LR</i>	104	48	97	63	53,5	46,5	39,4	56,8
<i>RFC</i>	138	14	66	94	74,4	25,6	58,8	87,0

Çizelge 4.6. 8 Kanal makine öğrenmesi performans sonuçları ikinci bölümü.

<i>Algoritma</i>	<i>F1 (%)</i>	<i>Specificity (%)</i>	<i>MCC</i>	<i>CrossValidation</i>	<i>Logloss</i>
<i>N.Bayes</i>	26,0	86,2	0,042	[60 52,1 51,69]	7,787
<i>LGBM</i>	82,5	81,6	0,641	[77,4 74,7 76,5]	0,422
<i>SVC</i>	-	100	0	[50,7 50,0 50,9]	0,693
<i>DTC</i>	68,4	86,2	0,466	[70,1 64,9 66,9]	9,631
<i>KNC</i>	59,9	60,5	0,193	[59,8 57,9 54,5]	2,170
<i>LR</i>	46,5	68,4	0,081	[59,6 53,1 56,0]	0,693
<i>RFC</i>	70,1	90,8	0,520	[67,5 68,0 64,7]	0,495



Şekil 4.5. 8 kanal makine öğrenmesi ROC eğrileri sonuçları.

BÖLÜM 5

SONUÇLAR VE ÖNERİLER

Bu çalışmada Matlab işitme testi arayüzü ile gerçekleştirilen işitme testi sırasında 16 kanallı EEG sinyallerinden alınan veriler, 8 kanal, 13 kanal ve 16 kanal olmak üzere üç yöntem ile makine öğrenmesi sınıflandırma algoritmaları kullanılarak incelenmiştir. 16 kanal için yapılan çalışmada en başarılı algoritma %77,5 maksimum Cross Validation değeri ile LGBM olmuştur. 13 kanal için yapılan çalışmada ise %78,7 maksimum Cross Validation değeri ile LGBM olmuştur. 8 kanal için yapılan çalışmada ise %77,4 maksimum Cross Validation değeri ile LGBM olmuştur. P300 etkisinin net olarak görüldüğü kanallar ile tüm kanallar için yapılan çalışmalarda LGBM algoritması en başarılı algoritma olup başarı oranları arasında fazla fark görülmemiştir. Bir EEG sinyalinin duydu-duymadı olarak rassal tahmin oranı %50 iken LGBM algoritması ile %78,7 tahmin oranı tespit edilmiştir.

Doğruluk oranı (accuracy) metriğini incelendiğinde; 16 kanal için yapılan çalışmada %83 ile en başarılı algoritma LGBM, 13 kanal için yapılan çalışmada %84 ile en başarılı algoritma LGBM, 8 kanal için yapılan çalışmada %82,1 ile en başarılı algoritma LGBM olmuştur.

Hassasiyet (precision) metriği incelendiğinde; 16 kanal için yapılan çalışmada %84,3 ile en başarılı algoritma LGBM, 13 kanal için yapılan çalışmada %86,2 ile en başarılı algoritma LGBM, 8 kanal için yapılan çalışmada %87 ile en başarılı algoritma RFC olmuştur.

Duyarlılık (recall) metriği incelendiğinde; 16 kanal için yapılan çalışmada %82,7 ile en başarılı algoritma LGBM, 13 kanal için yapılan çalışmada %81,9 ile en başarılı algoritma LGBM, 8 kanal için yapılan çalışmada %82,5 ile en başarılı algoritma LGBM olmuştur.

F1 skoru metriği incelendiğinde; 16 kanal için yapılan çalışmada %83,5 ile en başarılı algoritma LGBM, 13 kanal için yapılan çalışmada %84 ile en başarılı algoritma LGBM, 8 kanal için yapılan çalışmada %82,5 ile en başarılı algoritma LGBM olmuştur.

MCC metriği incelendiğinde; 16 kanal için yapılan çalışmada 0,66 değeri ile en başarılı algoritma LGBM, 13 kanal için yapılan çalışmada 0,681 değeri ile en başarılı algoritma LGBM, 8 kanal için yapılan çalışmada 0,641 değeri ile en başarılı algoritma LGBM olmuştur.

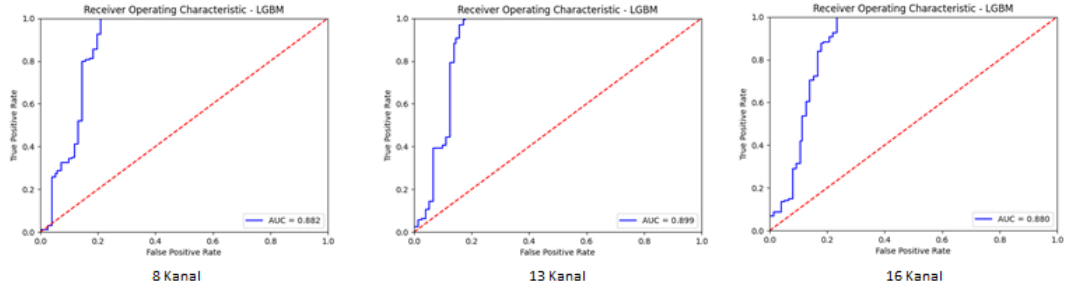
Log-kayıp (Log-loss) metriği incelendiğinde; 16 kanal için yapılan çalışmada 0,397 ile en başarılı algoritma LGBM, 13 kanal için yapılan çalışmada 0,383 değeri ile en başarılı algoritma LGBM, 8 kanal için yapılan çalışmada 0,422 değeri ile en başarılı algoritma LGBM olmuştur.

ROC eğrisinde AUC oranı metriği incelendiğinde; 16 kanal için yapılan çalışmada %88 ile en başarılı algoritma LGBM, 13 kanal için yapılan çalışmada %89,9 ile en başarılı algoritma LGBM, 8 kanal için yapılan çalışmada %88,2 ile en başarılı algoritma LGBM olmuştur.

Makine öğrenmesi model başarı metrikleri incelendiğinde LGBM algoritmasının diğer algoritmalara göre daha başarılı tahminlerde bulunduğu anlaşılmıştır. 16 kanal EEG verisi yerine P300 etkisinin tespit edildiği ilk 13 kanalın ve ilk 8 kanalın EEG verilerinin eğitim için alınıp makine öğrenmesi algoritmaları ile test edildiğinde başarı metriklerinin değiştiği tespit edilmiştir. Çizelge 5.1’de EEG kayıt cihazından alınan 13 kanal EEG verisi ile yapılan eğitimlerin, 16 kanal ve 8 kanal EEG verisi ile yapılan eğitimlere kıyasla daha başarılı olduğu görülmektedir. Şekil 5.1’de ise LGBM algoritmasının 13 kanal EEG verisi ile yapılan eğitimlerin, 16 kanal ve 8 kanal EEG verisi ile yapılan makine öğrenmesi eğitimlerine göre daha başarılı olduğu gösteren ROC eğrileri görülmektedir.

Çizelge 5.1. LGBM algoritmasının kanal sayısına göre başarı metrikleri.

	16 Kanal	13 Kanal	8 Kanal
Accuracy (%)	83,0	84,0	82,1
MSE	17,0	16,0	17,9
Recall (%)	82,7	81,9	82,5
Precision (%)	84,3	86,2	82,5
F1 (%)	83,5	84,0	82,5
Specificity (%)	83,3	86,2	81,6
MCC	0,66	0,681	0,641
CrossValidation (%)	[76,2 77,5 74,3]	[76,2 75,4 78,7]	[77,4 74,7 76,5]
Logloss	0,397	0,383	0,422



Şekil 5.1. LGBM algoritmasının kanal sayısına göre ROC eğrileri.

Matlab işitme testi arayüzü ile gerçekleştirilen ve eş zamanlı olarak EEG verilerinin kayıt altına alındığı bu çalışmada, farklı demografik (yaş, cinsiyet vb.) özelliklerde ve farklı meslek gruplarında çalışan daha fazla kişiden verilerin toplanıp makine öğrenmesi algoritmaları ile tahmin edilmesi halinde başarı oranlarının artabileceği tahmin edilmektedir. Ayrıca işitme testi esnasında, testin yapıldığı odanın ses izolasyonuna sahip olması ve kullanılan kulaklığın dış sesleri filtreleme özelliğinin olması, test esnasında verilen seslerin daha doğru bir şekilde duyulmasını sağlayacağı için EEG verilerinin daha doğru kaydedilmesini ve dolayısıyla makine öğrenmesi algoritmalarının başarı oranlarının daha yüksek olmasını sağlayacaktır.

KAYNAKLAR

1. “Deafness and hearing loss,” World Health Organization, [Online] . Available: <http://www.who.int/en/news-room/factsheets/detail/deafness-and-hearing-loss> (2019)
2. Hawthorne G, Hogan A, Giles E, et al. “Evaluating the health-related quality of life effects of cochlear implants: A prospective study of an adult cochlear implant program”. *Int J Audiol* 43: 183 – 192 (2004).
3. Davis A, Smith P, Ferguson M, Stephens D, Gianopoulos I. “Acceptability, benefit and costs of early screening for hearing disability: a study of potential screening tests and models”. *Health Technol Assess*;11: 1–294 (2007)
4. Bess FH, Humes L. Audiologic Measurement. “In: Audiology: the fundamentals. Fourth Edition”, *Lippincott Williams & Wilkins*, Philadelphia, p. 101–58 (2008).
5. International Electrotechnical Commission, “Pure-tone Audiometers”, *CEI-IEC-60645-1 2nd edition* (2001).
6. VENCOSKÝ V. et RUND, “F. Pure tone audiometer”, *20th Annual Conference Proceeding's Technical Computing*, p. 1- 5. (2012)
7. M. Gargouri, M. Chaoui, and P. Wira, “Development of hearing self-assessment pure tone audiometer,” *DTS 2020 - IEEE Int. Conf. Des. Test Integr. Micro Nano-Systems*, pp. 8–12, doi: 10.1109/DTS48731.2020.9196165 (2020).
8. Saka, İlkay. “Altmışbeş Yaş Üzeri Erişkinlerde, Saf Ses Odyometri, Mobil Uygulama İle Kendi Kendine Yapılabilen İşitme Testi Ve Türkçe İşitme Engeli Ölçeği - Yaşlı İle Elde Edilen Sonuçların Karşılaştırmalı Analizi”, Yüksek Lisans Tezi, *Başkent Üniversitesi Sağlık Bilimleri Enstitüsü*, Ankara (2021)
9. S. Özelik and N. Hardala, “The statistical measurements and neural network analysis of the effect of musical education to musical hearing and sensing,” *Expert Syst. Appl.*, vol. 38, no. 8, pp. 9517–9521, doi: 10.1016/j.eswa.2011.01.149 (2011).
10. S. Rajkumar, S. Muttan, V. Saphagirivasan, V. Jaya, and S. S. Vignesh, “Software intelligent system for effective solutions for hearing impaired subjects,” *Int. J. Med. Inform.*, vol. 97, pp. 152–162, doi: 10.1016/j.ijmedinf.2016.10.009 (2017).

11. F. Fürbass, M. A. Kural, G. Gritsch, M. Hartmann, T. Kluge, and S. Beniczky, “An artificial intelligence-based EEG algorithm for detection of epileptiform EEG discharges: Validation against the diagnostic gold standard,” *Clin. Neurophysiol.*, vol. 131, no. 6, pp. 1174–1179, doi: 10.1016/j.clinph.2020.02.032 (2020).
12. S. L. Tan, S. K. Loh, and W. C. Chee, “Speech-enabled pure tone audiometer,” *2007 Int. Symp. Intell. Signal Process. Commun. Syst. ISPACS 2007*, pp. 361–364, 2007, doi: 10.1109/ISPACS.2007.4445898 (2007).
13. Y. Fayçal, B. Wahiba, B. Lotfi, B. Ratiba, and A. Benia, “Computer audiometer for hearing testing,” *Int. Conf. Adv. Electron. Micro-electronics, ENICS 2008*, pp. 111–114, 2008, doi: 10.1109/ENICS.2008.7. (2008)
14. M. Y. Hossain and A. B. M. S. U. Doulah, “Detection of Motor Imagery (MI) Event in Electroencephalogram (EEG) Signals using Artificial Intelligence Technique,” *2020 IEEE East-West Des. Test Symp. EWDTS - Proc.*, no. Mi, pp. 0–5, 2020, doi: 10.1109/EWDTS50664.2020.9224634 (2020).
15. N. Sriraam, “EEG based automated detection of auditory loss: A pilot study,” *Expert Syst. Appl.*, vol. 39, no. 1, pp. 723–731, doi: 10.1016/j.eswa.2011.07.064 (2012).
16. M. P. Paulraj, K. Subramaniam, S. Bin Yaccob, A. H. Bin Adom, and C. R. Hema, “A machine learning approach for distinguishing hearing perception level using auditory evoked potentials,” *IECBES 2014, Conf. Proc. - 2014 IEEE Conf. Biomed. Eng. Sci.* “Miri, Where Eng. Med. Biol. Humanit. Meet,” no. December, pp. 991–996, doi: 10.1109/IECBES.2014.7047661 (2014).
17. FRANKS, John R. “Hearing measurement. Occupational Exposure to Noise: Evaluation, Prevention and Control”. *Geneva: World Health Organisation*, p. 183-231 (2001)
18. Gmbh, B. P., “V-Amp & ImpBox | Operating Instructions”, *Brain Products GmbH*, 1-50 (2016).
19. Instructions, O., “ActiCAP Xpress”, *Brain Products GmbH*, 1–49 (2014).
20. Turing A.M., “Computing Machinery and Intelligence”. *Mind*. 59(236): 433-460 (1950).
21. Erdoğan, Levent., “Coğrafi Bilgi Teknolojileri Ve Akıllı Uygulama Teknolojileri İle Kırsal Kalkınmanın Değerlendirilmesi”, Yüksek Lisans Tezi, *Erciyes Üniversitesi Fen Bilimleri Enstitüsü*, Kayseri (2021)
22. Candan, Hüseyin., “Ses İletim Hızıyla Akciğer Hastalıklarının Teşhisinde Makine Öğrenimi Yöntemlerinin Performanslarının Karşılaştırılması”, Yüksek Lisans Tezi, *Ege Üniversitesi Sağlık Bilimleri Enstitüsü*, İzmir (2019)
23. Alpaydin, E., “Introduction to machine learning”. *MIT press* (2020).

24. M. Özdemir, “Predicting Stock Movements With Machine Learning Using Textual Data”, Yüksek Lisans Tezi, *Boğaziçi Üniversitesi Sosyal Bilimler Enstitüsü*, İstanbul (2020).
25. P. S. Parmar, P. K. Biju, M. Shankar, and N. Kadiresan, “Multiclass Text Classification and Analytics for Improving Customer Support Response through different Classifiers,” *2018 Int. Conf. Adv. Comput. Commun. Informatics*, ICACCI 2018, pp. 538–542, doi: 10.1109/ICACCI.2018.8554881 (2015).
26. Dhage, Sandhya N., C. K. R., “A review on Machine Learning Techniques”, *International Journal on Recent and Innovation Trends in Computing and Communication* 4(3), 395–399 (2017).
27. David Dolan Lewis. “Representation and learning in information retrieval”, Doktora Tezi, *University of Massachusetts, USA*. Order Number: UMI Order No. GAX92-19460 . <https://dl.acm.org/doi/book/10.5555/143044> (1992)
28. İnternet: Ismail, H. A., “Learning Data Science: Day 19 - Naive Bayes - Haydar Ali Ismail”. Medium. <https://haydar-ai.medium.com/learning-data-science-day-19-naive-bayes-3d5d8b93f8e9> (2020)
29. Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, Ye Q, Liu T-Y. “Lightgbm: A Highly Efficient Gradient Boosting Decision Tree”, *Advances in neural information processing systems*, p. 3146e54 (2017)
30. G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, T.-Y. Liu, “Lightgbm: A highly efficient gradient boosting decision tree,” in *Advances in Neural Information Processing Systems*, 2017, pp. 3146–3154.
31. Gümüştaş E., “Kayıp Gözlem İçeren Dengesiz Veri Setlerinin Topluluk Öğrenme Algoritmaları İle Sınıflandırılması”, *Mimar Sinan Güzel Sanatlar Üniversitesi*, İstanbul (2019).
32. Breiman, L., “RANDOM FORESTS”, *Statistics Department University of California Berkeley*, CA 94720, (2001)
33. AKMAN M, GENÇ Y, ANKARALI H (2011). “Random Forests Yöntemi ve Sağlık Alanında Bir Uygulama.”, *Türkiye Klinikleri J Biostat.*, (2011)
34. Cover, T.M ve Hart, P.E., “Nearest Neighbor Pattern Classification”, *IEEE Transactions On Information Theory*, 13(1): 21-27 (1967).
35. İnternet: Scikit-learn, “Supervised Learning”, https://scikit-learn.org/stable/tutorial/statistical_inference/supervised_learning.html, (2021).
36. Hsu, C.-W., Chang, C.-C., & Lin, C.-J., “A Practical Guide to Support Vector Classification”, *National Taiwan University Department of Computer Science*, 1(1), 1–12 (2007).

37. Vapnik V., L. A., “Pattern recognition using generalized portrait method”, *Automation and Remote Control* (24), 774:780 (1963).
38. Ivanciuc, O, “Applications of Support Vector Machines in Chemistry”, *John Wiley & Sons*, 291–400. <https://doi.org/10.1002/9780470116449.ch6> (2007)
39. Sahin, M. Ö., Krücker, D., & Melzer-Pellmann, I.-A., “Performance and optimization of support vector machines in high-energy physics classification problems”, *Deutsches Elektronen-Synchrotron* (2016)
40. Ayhan, S. ve Dođmuş, Ş., “Destek Vektör Makineleriyle Sınıflandırma Problemlerinin Çözümü İçin Çekirdek Fonksiyonu Seçimi”, *Eskişehir Osmangazi Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*. 9(1): 175-201 (2014).
41. Kavzaođlu, T. ve Çölkesen, İ., “Destek Vektör Makineleri ile Uydu Görüntülerinin Sınıflandırılmasında Kernel Fonksiyonlarının Etkilerinin İncelenmesi”, *Harita Dergisi*, (144): 73-82 (2010).
42. Akçay A., “Bilgi ve Belge Yönetiminde Veri Madenciliđi”, Yüksek Lisans Tezi, *İstanbul Üniversitesi Sosyal Bilimler Enstitüsü*, İstanbul (2014).
43. Gülaçar H., “Nesnelerin İnterneti Platformları İçin Makine Öğrenmesi Tabanlı Bir Tahmin Modülü / dü. The 5th International Symposium on Emerging Inter-networks Communication and Mobility”. Yüksek Lisans Tezi, *İstanbul Teknik Üniversitesi Fen Bilimleri Enstitüsü*, İstanbul (2018).
44. Görgün, Mertcan., “Makine Öğrenmesi Yöntemleri İle Kalp Hastalığının Teşhis Edilmesi”, Yüksek Lisans Tezi, *İstanbul Aydın Üniversitesi Lisansüstü Eğitim Enstitüsü*, İstanbul (2020)
45. Yangın G., “Xgboost ve Karar Ağacı Tabanlı Algoritmaların Diyabet Veri Setleri Üzerine Uygulaması”, Yüksek Lisans Tezi, *Mimar Sinan Güzel Sanatlar Üniversitesi Fen Bilimleri Enstitüsü*, İstanbul (2019).
46. Pearl R., Reed L.J., K. J. F., “The Logistic Curve and the Consensus Count of 1940”, *Science*, 14, 895:901 (1940).
47. Peng, C.-Y. J., Lee, K. L., & Ingersoll, G. M., “An Introduction to Logistic Regression Analysis and Reporting”, *The Journal of Educational Research*, 96(1), 3–14 , <https://doi.org/10.1080/00220670209598786> (2002)
48. Baş, S. & Uzun, A., “Tedarik Zincirinde Müşteri Siparişlerinin Lojistik Regresyon Analizi İle Deđerlendirilmesi”, *Ömer Halis Demir Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 11(3), 67-81 (2018)
49. Hewett, T. E., Webster, K. E., & Hurd, W. J., “Systematic Selection of Key Logistic Regression Variables for Risk Prediction Analyses: A Five-Factor Maximum Model. *Clinical Journal of Sport Medicine*”, **Official Journal of the**

Canadian Academy of Sport Medicine 29(1), 78–85,
<https://doi.org/10.1097/JSM.0000000000000486> (2019).

50. Felitti, V. J., Anda, R. F., Nordenberg, D., Williamson, D. F., Spitz, A. M., Edwards, V., ... Marks, J. S., “Relationship of childhood abuse and household dysfunction to many of the leading causes of death in adults. The Adverse Childhood Experiences (ACE) Study”, *American Journal of Preventive Medicine*, 14(4), 245–258 (1998)
51. Cranmer, S. J., Menninga, E. J., & Mucha, P. J., Kantian fractionalization predicts the conflict propensity of the international system”, *Proceedings of the National Academy of Sciences*, 112(38), 11812–11816, <https://doi.org/10.1073/pnas.1509423112> (2015).
52. Candan, Hüseyin. “Ses İletim Hızıyla Akciğer Hastalıklarının Teşhisinde Makine Öğrenimi Yöntemlerinin Performanslarının Karşılaştırılması”, Yüksek Lisans Tezi, *Ege Üniversitesi Sağlık Bilimleri Enstitüsü*, İzmir (2019)
53. Uzosike, Uzoma Beneth, “Kestirimsel Analizde Boyutluluk İndirgenmesi ve Makine Öğrenmesi”, Yüksek Lisans Tezi, *Mersin Üniversitesi Sosyal Bilimler Enstitüsü*, Mersin (2020)
54. Tozlu, İbrahim., “Simplifying Balance Sheet Adjustment Process In Commercial Loan Applications Using Machine Learning Methods”, Yüksek Lisans Tezi, *İstanbul Teknik Üniversitesi Fen Bilimleri Enstitüsü*, İstanbul (2019)
55. Karşlı, Özden Burcu., “Makine Öğrenme Yöntemleri İle Karaciğer Hastalığı Teşhisi”, Yüksek Lisans Tezi, *İstanbul Teknik Üniversitesi Fen Bilimleri Enstitüsü*, İstanbul (2019)
56. O. A. Moruff, A. O. Maruf, and A. Toshio, “Performance Analysis of Selected Machine Learning Algorithms for the Classification of Phishing URLs”, *Faculty of Natural and Applied Sciences Al-Hikmah University*, vol. 13, no. November, pp. 16–20, 20, (2020)
57. E. E. Hassler, D. P. Hale, and J. E. Hale, “A comparison of automated training-by-example selection algorithms for Evidence Based Software Engineering,” *Inf. Softw. Technol.*, vol. 98, no. February, pp. 59–73, doi: 10.1016/j.infsof.2018.02.001 (2018)
58. B. Çil, H. Ayyıldız, and T. Tuncer, “Discrimination of β -thalassemia and iron deficiency anemia through extreme learning machine and regularized extreme learning machine based decision support system,” *Med. Hypotheses*, vol. 138, no. December 2019, 2020, doi: 10.1016/j.mehy.2020.109611 (2019).
59. Z. Ye, S. Guo, D. Chen, H. Wang, and S. Li, “Drilling formation perception by supervised learning: Model evaluation and parameter analysis,” *J. Nat. Gas Sci. Eng.*, vol. 90, no. November 2020, p. 103923, doi: 10.1016/j.jngse.2021.103923 (2021).

60. S. U. Hassan, M. S. M. Zahid, and K. Husain, "Performance comparison of CNN and LSTM algorithms for arrhythmia classification," *Int. Conf. Comput. Intell. ICCI 2020*, no. October, pp. 223–228, doi: 10.1109/ICCI51257.2020.9247636 (2020).
61. Küçükbay, S.R., "Metaloproteinlerin Biyoenformarik Analizi", Yüksek Lisans Tezi, *Başkent Üniversitesi Fen Bilimleri Enstitüsü*, Ankara (2015)
62. Keçeoğlu Ç.R, "Kesme Puanının Belirlenmesinde Lojistik Regresyon Analizi, Diskriminant Analizi ve ROC Eğrisi Yöntemlerinin Karşılaştırılması", Yüksek Lisans Tezi, *Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü*, Ankara (2012)
63. F. Bulut, "Sınıflandırıcı Topluluklarının Dengesiz Veri Kümeleri Üzerindeki Performans Analizleri," *Bilişim Teknol. Derg.*, vol. 9, no. 2, pp. 153–0, doi: 10.17671/btd.81137 (2016)
64. Y. Manzali, M. Chahhou, and M. El Mohajir, "Impure Decision Trees for Auc and Log loss optimization", *Int. Conf. Wirel. Technol. Embed. Intell. Syst. WITS 2017*, 2017, doi: 10.1109/WITS.2017.7934675 (2017).
65. E. Maxwell, T. A. Warner, F. Fang, "Implementation of machine-learning classification in remote sensing: An applied review", *International Journal of Remote Sensing*, vol. 39, no. 9, pp. 2784–2817 (2018).
66. A Ramezan, T. A Warner, A. E Maxwell, "Evaluation of sampling and cross-validation tuning strategies for regional-scale machine learning classification," *Remote Sensing*, vol. 11, no. 2, p. 185, (2019).
67. M. Kuhn K. Johnson," Applied predictive modeling", *Springer*, vol. 26 (2013).
68. J. Zhang, G. S. Okin, B. Zhou, "Assimilating optical satellite remote sensing images and field data to predict surface indicators in the western us: Assessing error in satellite predictions based on large geographical datasets with the use of machine learning", *Remote Sensing of Environment*, vol. 233, p. 111 382 (2019).

ÖZGEÇMİŞ

Mustafa KÜÇÜKAKARSU, ilk ve orta öğrenimini Karabük'te tamamladı. Safranbolu Anadolu Öğretmen Lisesi'nden mezun oldu. 2012 yılında Karabük Üniversitesi Mühendislik Fakültesi Biyomedikal Mühendisliği Bölümü'nde öğrenime başlayıp 2014 yılında Karabük Üniversitesi Mühendislik Fakültesi Bilgisayar Mühendisliği Bölümü'nde Çift Anadal Programı (ÇAP)'na başladı. 2015 yılında Siemens Türkiye'nin düzenlemiş olduğu Siemens İnovasyon Yarışması'nda finalist oldu ve proje fikrini gerçekleştirme süresince Siemens Türkiye'de Uzman Yardımcısı olarak çalıştı. 2017 yılında Biyomedikal Mühendisliği bölümünden üçüncülük ile mezun oldu. Aynı sene çift anadal program bölümü olan Bilgisayar Mühendisliği Bölümü'nden de mezun oldu. 2017 yılında Karabük Üniversitesi Fen Bilimleri Enstitüsü Biyomedikal Mühendisliği Anabilim Dalında yüksek lisans eğitimine başladı. 2020 yılından itibaren bir firmada yapay zeka yazılım geliştirici olarak görevini sürdürmektedir.