

**DERİN ÖĞRENME TABANLI BAS KONUŞ/İTERKOM
SİSTEMLERİNDE SES DUYGU ANALİZİ**

KÜBRA SİNAN

**YÜKSEK LİSANS TEZİ
BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI**

**DANIŞMAN
PROF. DR. PAKİZE ERDOĞMUŞ**

DÜZCE, 2024

T.C.
DÜZCE ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

DERİN ÖĞRENME TABANLI BAS KONUŞ/İTERKOM
SİSTEMLERİNDE SES DUYGU ANALİZİ

Kübra SİNAN tarafından hazırlanan tez çalışması aşağıdaki jüri tarafından Düzce Üniversitesi Lisansüstü Eğitim Enstitüsü Bilgisayar Mühendisliği Anabilim Dalı'nda **YÜKSEK LİSANS TEZİ** olarak kabul edilmiştir.

Tez Danışmanı

Prof. Dr. Pakize ERDOĞMUŞ

Düzce Üniversitesi

Jüri Üyeleri

Prof. Dr. Pakize ERDOĞMUŞ

Düzce Üniversitesi

Doç. Dr. Abdullah Talha KABAKUŞ

Düzce Üniversitesi

Dr. Öğr. Üyesi Fatma AKALIN

Sakarya Üniversitesi

Tez Savunma Tarihi: 22/01/2024

BEYAN

Bu tez çalışmasının kendi çalışmam olduğunu, tezin planlanmasından yazımına kadar bütün aşamalarda etik dışı davranışımın olmadığını, bu tezdeki bütün bilgileri akademik ve etik kurallar içinde elde ettiğimi, bu tez çalışmasıyla elde edilmeyen bütün bilgi ve yorumlara kaynak gösterdiğimi ve bu kaynakları da kaynaklar listesine aldığımı, yine bu tezin çalışılması ve yazımı sırasında patent ve telif haklarını ihlal edici bir davranışımın olmadığını beyan ederim.

22 Ocak 2024

Kübra SİNAN

TEŞEKKÜR

Yüksek Lisans öğrenimimde ve bu tezin hazırlanmasında gösterdiği her türlü destek ve yardımından dolayı çok değerli hocam Prof. Dr. Pakize ERDOĞMUŞ' a en içten dileklerimle teşekkür ederim.

Bu çalışma boyunca ve tez yazımı sırasında manevi desteklerini hiçbir zaman esirgemeyen ve motivasyonumu yüksek tutmamı sağlayan, hayatımın her aşamasında olduğu gibi, maddi ve manevi desteklerini hiçbir zaman esirgemeyen, varlıkları ile bana güç veren kıymetli aileme, en içten sevgi, saygı ve teşekkürlerimi sunarım.

22 Ocak 2024

Kübra SİNAN

İÇİNDEKİLER

Sayfa No

ŞEKİL LİSTESİ	vi
ÇİZELGE LİSTESİ	vii
KISALTMALAR.....	viii
ÖZET	ix
ABSTRACT	x
1. GİRİŞ.....	1
2. GENEL BİLGİLER	3
2.1. YAPAY ZEKÂ	3
2.2. MAKİNE ÖĞRENMESİ.....	4
2.2.1. Denetimli Öğrenme	5
2.2.2. Denetimsiz Öğrenme	6
2.2.3. Pekiştirmeli Öğrenme	7
2.3. DERİN ÖĞRENME.....	7
2.4. TRANSFER ÖĞRENME	9
2.5. LİTERATÜR ÇALIŞMALARI.....	20
3. MATERYAL VE YÖNTEM	27
3.1. VERİ SETİNİN BELİRLENMESİ	29
3.2. VERİLERİN EĞİTİM, DOĞRULAMA VE TEST OLARAK BÖLÜNMESİ	34
3.3. ÖZELLİK ÇIKARMA	35
3.4. ÖN İŞLEME	39
3.5. MODEL EĞİTİMİ.....	39
3.6. SINIFLANDIRMA	41
3.7. MODELİN PERFORMANS DEĞERLENDİRMESİ.....	48
4. SONUÇ	58
5. KAYNAKLAR	60
ÖZGEÇMİŞ.....	66

ŞEKİL LİSTESİ

Sayfa No

Şekil 2.1 Yapay Zekâ, Makine Öğrenmesi ve Derin Öğrenme (Edureka, 2023)	3
Şekil 2.2 Makine Öğrenmesi (Masraff, 2024)	5
Şekil 2.3 Denetimli Makine Öğrenmesi (Turhost, 2021)	6
Şekil 2.4 Denetimsiz Makine Öğrenmesi (Turhost, 2021)	6
Şekil 2.5 Pekiştirmeli Makine Öğrenmesi (Turhost, 2021)	7
Şekil 2.6 Derin Öğrenme Tanımı (Yılmaz ve Orman, 2021)	8
Şekil 2.7 Derin Öğrenme Mimarisi (Onar, 2020)	9
Şekil 2.8 Transfer Öğrenme Mimarisi (Jain Anchit, 2018)	10
Şekil 2.9 Transfer Öğrenme Uygulaması (Dwivedi Rohit, 2020)	11
Şekil 2.10 BERT Transformatör Modeli (Devlin vd. 2019)	13
Şekil 2.11 GPT Transformatör Modeli (Radford vd. 2018)	14
Şekil 2.12 AST Transformatör Modeli (Gong vd. 2021)	16
Şekil 3.1 Ses Duygu Analizi Geliştirilen Model Süreci	28
Şekil 3.2 EMO-DB Ses Verilerinin Kayıt Şekli (Burkhardt vd. 2005)	29
Şekil 3.3 EMO-DB Mutlu Duygu Sınıfının Dalga Formu	30
Şekil 3.4 EMO-DB Kızgın Duygu Sınıfının Dalga Formu	30
Şekil 3.5 EMO-DB Üzgün Duygu Sınıfının Dalga Formu	30
Şekil 3.6 EMO-DB Korkmuş Duygu Sınıfının Dalga Formu	31
Şekil 3.7 EMO-DB İğrenmiş Duygu Sınıfının Dalga Formu	31
Şekil 3.8 EMO-DB Veri Seti Duygu Dağılımları	34
Şekil 3.9 EMO-DB Veri Setinin Eğitim, Doğrulama ve Test Alt Kümelerine Bölünmesi (Agrawal Samarth, 2021)	34
Şekil 3.10 EMO-DB Mutlu Duygu Sınıfının Spektrogramı	36
Şekil 3.11 EMO-DB Kızgın Duygu Sınıfının Spektrogramı	37
Şekil 3.12 EMO-DB Üzgün Duygu Sınıfının Spektrogramı	37
Şekil 3.13 EMO-DB Korkmuş Duygu Sınıfının Spektrogramı	37
Şekil 3.14 EMO-DB İğrenmiş Duygu Sınıfının Spektrogramı	38
Şekil 3.15 Önerilen Modelin Çalışma Metodolojisi	57

ÇİZELGE LİSTESİ

Sayfa No

Çizelge 2.1 Wav2Vec2 Üzerine Yapılan Çalışmalar.....	18
Çizelge 2.1 (devamı). Wav2Vec2 Üzerine Yapılan Çalışmalar	19
Çizelge 2.1 (devamı). Wav2Vec2 Üzerine Yapılan Çalışmalar	20
Çizelge 3.1 EMO-DB Veri Seti Seslendirme Sanatçı Bilgileri (Emodb Bilderbar,2024)	32
Çizelge 3.2 EMO-DB Veri Seti Duygu Sınıf İsimlendirmeleri.....	32
Çizelge 3.3 EMO-DB Veri Seti Cümleleri	33
Çizelge 3.4 EMO-DB Veri Seti Duygu Sınıfları	33
Çizelge 3.5 Wav2Vec2 Hiperparametreleri ve Açıklamaları	35
Çizelge 3.6 Wav2Vec2 Hiperparametreleri ve Değerleri	39
Çizelge 3.7 Veri Bağdaştırıcıları ve Açıklamaları.....	40
Çizelge 3.8 Veri Bağdaştırıcı Parametreleri ve Değerleri	41
Çizelge 3.9 Eğitim Verilerinin Ayrılmasına Göre Eğitim Doğruluk Oranı Grafikleri ve Açıklamaları	42
Çizelge 3.9 (devamı). Eğitim Verilerinin Ayrılmasına Göre Eğitim Doğruluk Oranı Grafikleri ve Açıklamaları	43
Çizelge 3.10 Eğitim Verilerinin Ayrılmasına Göre Eğitim Karmaşıklık Matrisleri ve Açıklamaları	44
Çizelge 3.10 (devamı). Eğitim Verilerinin Ayrılmasına Göre Eğitim Karmaşıklık Matrisleri ve Açıklamaları	45
Çizelge 3.10 (devamı). Eğitim Verilerinin Ayrılmasına Göre Eğitim Karmaşıklık Matrisleri ve Açıklamaları	46
Çizelge 3.10 (devamı). Eğitim Verilerinin Ayrılmasına Göre Eğitim Karmaşıklık Matrisleri ve Açıklamaları	47
Çizelge 3.11 EMO-DB Veri Setinin Eğitim Doğrulama ve Test Veri Setlerine Ayrılma Oranlarına Göre Değişen Doğruluk Değerleri	49
Çizelge 3.12 Verilerin Ayrılmasına Göre Sınıflandırma Raporları ve Değerlendirmeleri.....	51
Çizelge 3.12 (devamı). Verilerin Ayrılmasına Göre Sınıflandırma Raporları ve Değerlendirmeleri.....	52
Çizelge 3.12 (devamı). Verilerin Ayrılmasına Göre Sınıflandırma Raporları ve Değerlendirmeleri.....	53
Çizelge 3.12 (devamı). Verilerin Ayrılmasına Göre Sınıflandırma Raporları ve Değerlendirmeleri.....	54
Çizelge 3.12 (devamı). Verilerin Ayrılmasına Göre Sınıflandırma Raporları ve Değerlendirmeleri.....	55
Çizelge 3.12 (devamı). Verilerin Ayrılmasına Göre Sınıflandırma Raporları ve Değerlendirmeleri.....	56

KISALTMALAR

ANN	Yapay Sinir Ağları
AST	Ses Spektrogram Transformatörü
BERT	Transformatörlerden Çift Yönlü Kodlayıcı Temsilleri
CNN	Evrişimsel Sinir Ağları
DCNN	Uygulamalı Evrişimsel Sinir Ağları
DL	Derin Öğrenme
EMO-DB	Berlin Duygusal Konuşma Veritabanı
EER	Eşit Hata Oranı
GMM	Gauss Karışım Modeli
GPRS	Genel Paket Radyo Servisi
GPT	Üretken Önceden Eğitilmiş Transformatör
GSM	Mobil İletişim İçin Küresel Sistem
HMM	Gizli Markov Modeli
IEMOCAP	Etkileşimli Duygusal İkili Hareket Yakalama Veritabanı
LSTM	Uzun Kısa Süreli Bellek
MFCC	Mel Frekansı Kepstral Katsayıları
ML	Makine Öğrenmesi
NLP	Doğal Dil İşleme
SER	Otomatik Konuşma Duygu Tanıma
SVM	Destek Vektör Makineleri
UBM	Evrensel Arka Plan Modeli
TurEV-DB	Türkçe Duygu-Ses Veritabanı
TL	Transfer Öğrenme
WA	Ağırlıklı Doğruluk Oranı
WER	Kelime Hata Oranı
WERR	Kelime Hata Oranı Azaltma
YZ	Yapay Zeka

ÖZET

DERİN ÖĞRENME TABANLI BAS KONUŞ/İTERKOM SİSTEMLERİNDE SES DUYGU ANALİZİ

Kübra SİNAN

Düzce Üniversitesi

Lisansüstü Eğitim Enstitüsü, Bilgisayar Mühendisliği Anabilim Dalı

Yüksek Lisans Tezi

Danışman: Prof. Dr. Pakize ERDOĞMUŞ

Ocak 2024, 65 sayfa

Metrolarda bulunan Bas-Konuş sistemleri yolcuların makinistle iletişime geçmelerini sağlayan sistemlerdir. Yolcular bu sistem aracılığı ile acil durumlarda makinistle irtibata geçerek bir sıkıntı olduğunu söyleyebilmektedir. Ancak mevcut Bas Konuş/ Interkom sistemlerinde bir öncelik olayı olmadığı için makinist rastgele butona basan biriyle de görüşebilmektedir. Dolayısıyla önemli /acil bir durum için basan yolcu görüşmek için daha sonraya kalabilmekte ve geç müdahaleye sebep olabilmektedir. Bu çalışmada, bas/konuş sistemlerinde karşılaşılan bu sorunlara çözüm getirmek amacı ile, son yıllarda geliştirilen derin öğrenme mimarilerinden Wav2Vec2 kullanılmıştır. Bas-konuş cihazından butona basan yolcu konuştuğunda, insan sesini algılayan sistem devreye girmektedir. Bas-konuş cihazından butona basan yolcu konuştuğunda, eğitilmiş model ile duygu sınıflandırma yapılarak, yolcunun gerçekten acil bir durumda olup olmadığı algılanmaktadır. Bu çalışmada eğitilen duygu sınıflandırma modelinin, sisteme entegre edilmesi ve tahmin edilen duygu sınıfına göre, sistem otomasyonu geliştirilmesi, acil olmayan durumların otomatik bir sistem cevabı ile geçilmesini ve acil durumların geciktirilmemesi amaçlanmaktadır. Çalışmada EMO-DB veri setinde yer alan, duygu sınıflarına göre önceden eğitilmiş bir Wav2Vec2 modeli kullanılarak transfer öğrenme yoluyla ses verileri 2D görüntülere, yani ses spektrogramlarına dönüştürülmüştür. Elde edilen veri seti ile transfer öğrenme yolu ile eğitilen model de, %92 doğruluk değerine ulaşıldığı görülmüştür.

Anahtar Sözcükler: Derin öğrenme, Duygu sınıflandırma, Transfer öğrenme, Transformatör

ABSTRACT

DEEP LEARNING BASED VOICE EMOTION ANALYSIS IN INTERCOM SYSTEMS

Kübra SİNAN

Düzce University

Graduate School, Department of Computer Engineering.

Master Thesis

Supervisor: Prof. Dr. Pakize ERDOĞMUŞ

January 2024, 65 pages

Push-to-Talk systems in subways are systems that allow passengers to communicate with the engineer. Through this system, passengers can contact the engineer in case of emergency and tell him/her that there is a problem. However, since there is no priority in the current Push to Talk/Intercom systems, the engineer can talk to someone who presses the button randomly. Therefore, the passenger who presses the button for an important/emergency situation may be delayed to be interviewed later and may cause late intervention. In this study, Wav2Vec2, one of the recently developed deep learning architectures, is used to solve these problems in push-to-talk systems. When the passenger presses the button on the push-to-talk device, the human voice recognition system is activated. When the passenger presses the button on the push-to-talk device, the trained model is used to classify the emotion and detect whether the passenger is really in an emergency or not. The aim of this study is to integrate the trained emotion classification model into the system and to develop system automation according to the predicted emotion class, to pass non-emergency situations with an automatic system response and not to delay emergency situations. In the study, audio data in the EMO-DB dataset were converted into 2D images, i.e. audio spectrograms, by transfer learning using a Wav2Vec2 model pre-trained according to emotion classes. It was observed that the model trained by transfer learning with the obtained data set achieved 92% accuracy.

Keywords: Deep learning, Emotion classification, Transfer learning, Transformers

1. GİRİŞ

Bas Konuş Sistemleri GSM/GPRS, Telsiz ve İnternet altyapısını kullanıp sesin ve/veya verinin ilgili altyapı teknolojisi ile iletilmesini sağlayan sistemlerdir. Bas Konuş sistemi, birebir veya grup olarak sesin ve/veya verinin yaklaşık gerçek zamanlı iletilmesini sağlayan haberleşme sistemleridir.

Bas-Konuş-Bırak-Dinle mantığı üzerine geliştirilmiş teknolojiye sahiptir ve genelde iletişimi tek-yönlü olup grup içerisinde bir kişi konuşurken diğerleri dinleme konumundadır, özel durumlarda ise yetkili kişiler aynı anda konuşma yapabilmektedir. (TED Bilişim, 2024)

Bas konuş/İnterkom sistemler metrolarda, veya tramvaylarda yolcu ile iletişim kurulabilmesi adına büyük öneme sahiptir. Metrolarda bulunan Bas konuş/İnterkom sistemleri yolcuların acil durumlarda makinistle irtibata geçerek durumu aktarması amacıyla kullanılmaktadır. Yolcuların bas konuş sistemlerini bilinçli kullanmaması nedeniyle makinist gereksiz yere meşgul edildiğinden böyle bir çalışmanın geliştirilmesine ihtiyaç duyulmuştur.

Bas konuş sistemlerde bu meşguliyeti ortadan kaldırabilmek, bir nebze de olsa azaltabilmek ve bu sistemlerin daha etkin kullanılabilmesini sağlayabilmek adına bas konuş sistemlerden alınacak ses verileriyle duygu analizi yapılması planlanmıştır. Duygu durumu, dış etmenlerden, kişisel etmenlerden ve insanlar arasındaki iletişim ve ilişkilerden etkilenen tespit edilmesi ve ayırt edilmesi zor bir durumdur. Literatürde insanların duygularını analiz etmek için birçok yöntemin çalışıldığı vurgulanmaktadır. Bunların arasında ses, mimikler, bedensel hareketler olduğu bilinmektedir. Çalışmamız sonucunda elde edilen insan sesleri analizi ile ileride yolculara yönelik geniş kapsamlı bir uygulamaya dönüştürülmesinde ön ayak olması hedeflenmiştir. Veri seti olarak EMO-DB kullanılmıştır. Tüm veri seti, yedi duyguyla hareket eden 10 cümleden oluşan 535 örnek içermektedir. Bu veri seti için seçilen duygular nötr, kızgın, korkmuş, mutlu, üzgün, iğrenmiş ve sıkılmıştır. Toplanan veriler önceden eğitilmiş model olan Wav2Vec2 modeli transfer öğrenme ve transformatörler yoluyla modele işlenmiştir ve sınıflandırılmıştır. Sonuçlar kullanıcıların duygularını %92 doğrulukla sınıflandırdığını göstermektedir.

Raylı sistemlerdeki bas konuş/İnterkom sistemleri, toplu taşıma araçlarında sesli iletişimi yönetmek ve acil durumlarda yolculara hızlı ve etkili bilgi aktarımını sağlamak açısından kritik bir rol oynamaktadır. Bu sistemler, tren ve tramvay gibi araçlarda personel ile yolcular arasında iletişimi kolaylaştırarak, güvenlik, konfor ve seyahat deneyimini geliştirmeyi amaçlamaktadır. Ancak, sesli iletişimde kullanılan bu sistemlerin, yolcuların ve personelin duygusal durumlarını anlamak ve bu bilgiyi etkili bir şekilde kullanmak açısından daha gelişmiş bir yaklaşımı benimsemeye ihtiyaç duyduğu görülmektedir.

Bas konuş/İnterkom sistemlerden alınacak ses verileri kullanılarak yapay zekanın birleştirilmesiyle literatür için önemli adımlar atılabileceği düşünülmüştür.



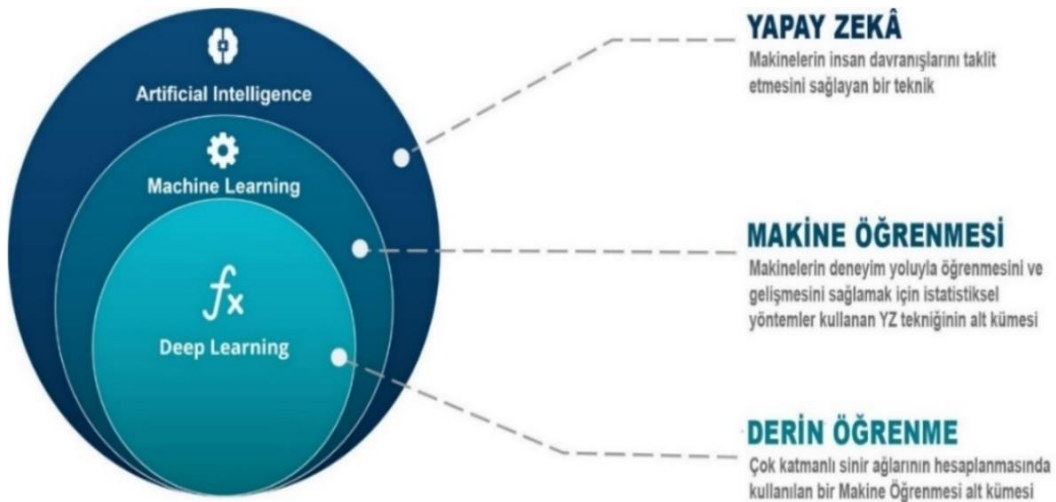
2. GENEL BİLGİLER

2.1. YAPAY ZEKÂ

Yapay zekâ (YZ) , bilgisayar sistemlerine insan benzeri zekâ özellikleri kazandırmak amacıyla tasarlanmış bir disiplindir. Yapay zekâ, bilgisayarların öğrenme, çıkarım, problem çözme, dil anlama ve görsel algılama gibi yeteneklere sahip olmalarını amaçlamaktadır.

Yapay zekânın makine öğrenmesi, doğal dil işleme, görüntü işleme gibi alanlarda sıklıkla kullanıldığı bilinmektedir. Yapay zekâ; sağlık sektöründe hastalık teşhisi, genetik analiz ve tedavi planlaması gibi alanlarda, finans sektöründe risk yönetimi, ticaret stratejileri ve müşteri hizmetleri gibi konularda, otomotiv sektöründe sürücüsüz araç teknolojisi üzerinde, eğitim sektöründe öğrenci performans analizi, öğrenci öneri sistemleri ve öğrenme materyali geliştirmeleri gibi alanlarda aktif rol aldığı görülmüştür (Üstündağ, 2022).

Yapay zekâ, teknolojik gelişmelerle birlikte hızla ilerlemekte ve birçok sektördeki süreçleri dönüştürmektedir.



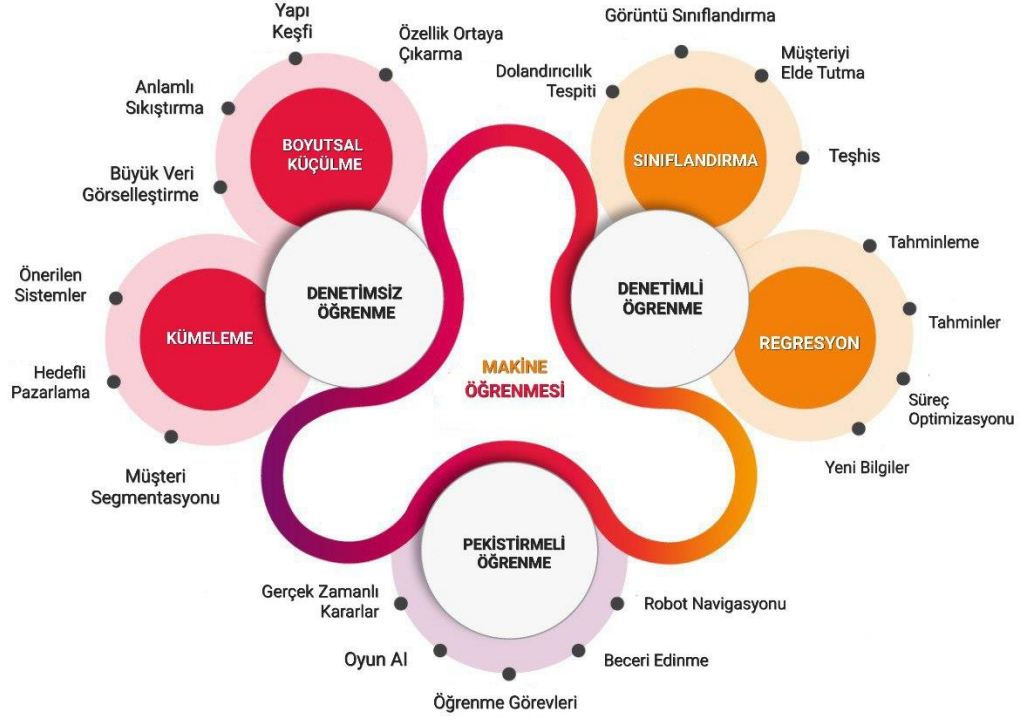
Şekil 2.1 Yapay Zekâ, Makine Öğrenmesi ve Derin Öğrenme (Edureka, 2023)

2.2. MAKİNE ÖĞRENMESİ

Makine öğrenmesi (ML), bilgisayar sistemlerinin verilere dayanarak örüntüler öğrenmesini sağlayan ve bu öğrenmeyi kullanarak belirli görevleri gerçekleştiren bir alanı ifade eden yapay zekânın bir alt dalıdır. Bu öğrenme süreci, programın açık bir şekilde programlanmadan önce veri üzerinden genelleme yapmasına olanak tanır. Makine öğrenmesi kural tabanlı bir teknik olmayıp bilgiyi doğrudan veriden ayrıştırarak öğrenmektedir. Bu yöntemlerin önceden belirlenmiş denklemleri veya çözümü bulan açıkça programlanmış kararları yoktur. Sadece verileri kullanarak çözüme giden yolu oluşturur veya denklemi öğrenir.

Veri girdileri üzerinde eğitim alarak bir model oluşturmak, oluşturulan modeli analiz etmek ve çıkarımlar yapmak, ardından verilerin sonuçlarını tahminlemek, makine öğrenmesinin temel taşını oluşturur. Bu süreç, örüntülerin veya anlamlı verilerin oluşabilmesi için var olan verilerin detaylı bir şekilde analiz edilmesini gerektirir. Analiz aşamasından sonra, modelin performansını değerlendirmek ve çıkarımlar yapmak önemlidir. Bu adım, modelin ne kadar iyi veya kötü performans gösterdiğini anlamak için gerçek ve tahmin edilen sonuçlar arasındaki uyumu değerlendirir. İstatistiksel analiz yöntemleri kullanılarak eski verilere dayanarak yeni verilere yönelik tahminlerde bulunabilen model oluşturulur (Yaraş, 2020).

Oluşturulan model parametrelerle test edilir, başarı ve kayıp oranları belirlenir. Veri eklenmesiyle modelin başarısının artırılması hedeflenir. Bu sürekli geri besleme ve iyileştirme döngüsü, modelin genelleme yeteneğini artırarak daha güvenilir ve etkili tahminler yapmasına olanak tanır. Makine öğrenmesi, öğrenme türlerine göre üçe ayrılır: Denetimli Öğrenme, Denetimsiz Öğrenme ve Pekiştirmeli Öğrenme (Varghese ve Sharma, 2018).

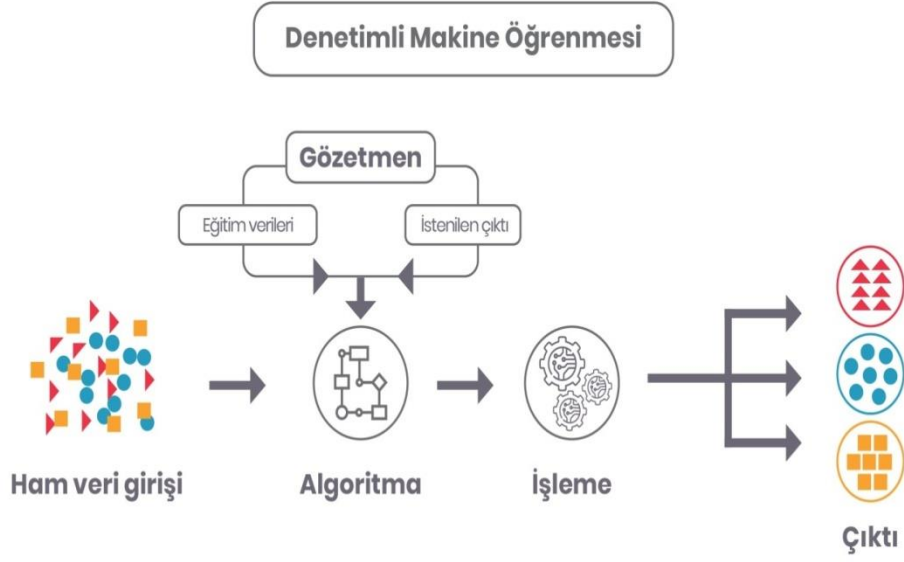


Şekil 2.2 Makine Öğrenmesi (Masraff, 2024)

2.2.1. Denetimli Öğrenme

Denetimli öğrenme, bilinmekte olan ve etiketlenmiş verilerin giriş ve çıkış fonksiyonları arasındaki ilişkinin öğrenildiği bir öğrenme türüdür. Model oluşturulduğunda, eğitim verileri kullanılarak model eğitilir. Bu süreçte, giriş ve çıkış verileri arasında bir ilişki kurulur ve model, bu ilişkiyi genelleme yeteneği kazanarak yeni veriler üzerinde tahminlerde bulunabilir.

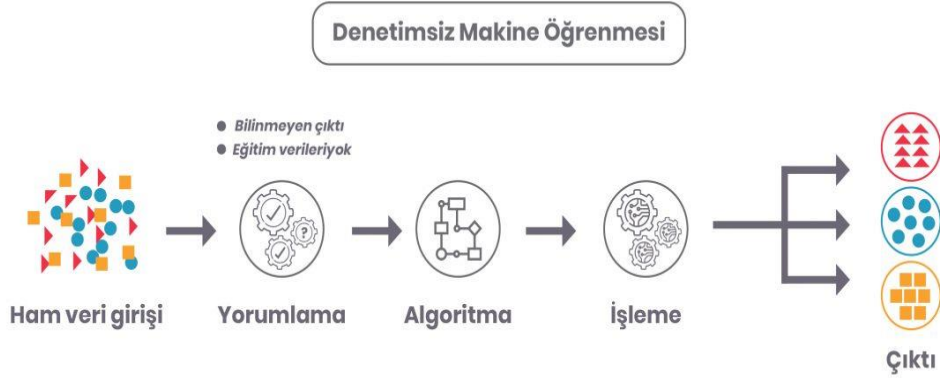
Denetimli öğrenme genellikle regresyon ve sınıflandırma problemleri olarak iki ana kategoriye ayrılır. Regresyon problemlerinde, modelin çıkışı sürekli bir değerdir ve belirli bir niceliği tahmin etmeye yöneliktir. Sınıflandırma problemlerinde ise çıkış, belirli bir sınıfa aittir ve model, verilen girdiye dayanarak bu sınıfı tahmin etmeye çalışır (Phelan vd. 2020).



Şekil 2.3 Denetimli Makine Öğrenmesi (Turhost, 2021)

2.2.2. Denetimsiz Öğrenme

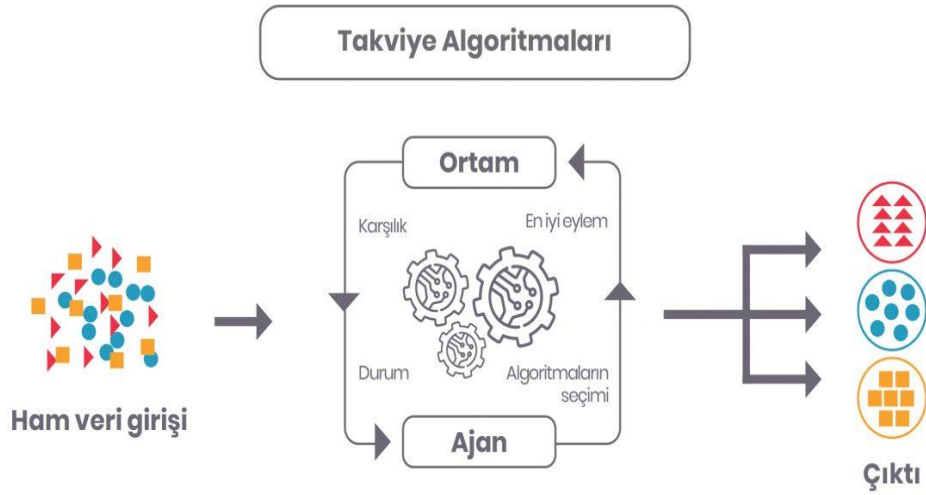
Denetimsiz öğrenme, sadece giriş verilerini kullanır. Geliştirilen model, giriş verilerine göre öğrenme işlemini gerçekleştirir ve çıktı verilerini keşfeder. Bu öğrenme türünün, modelin giriş verileri hakkında bilgi edinmesi ve bu girişlerin yapıya ait doğru dağılımını sağlamak genel amacıdır. Bu süreçte model, öğrenme aşamasında giriş verilerini analiz eder, bu verilerin içerdiği desenleri anlamaya çalışır ve ardından doğru çıktıları üretebilmek için bu desenleri kullanır (Goodfellow, Bengio ve Courville, 2017).



Şekil 2.4 Denetimsiz Makine Öğrenmesi (Turhost, 2021)

2.2.3. Pekiştirmeli Öğrenme

Pekiştirmeli /takviyeli öğrenme, bir sistem veya yapı için kendi başına karar alabilen yapıyı, belirli bir amacı gerçekleştirebilmek için doğru kararları nasıl öğreneceğini gösteren öğrenme türüdür. Pekiştirmeli öğrenme türünde, sistem, çevresiyle etkileşimde bulunarak elde ettiği deneyimlerden ders çıkarır ve bu deneyimlere dayanarak en iyi kararları vermeyi öğrenir. Bu algoritma türü, kuralları belirleyerek deneme-yanılma yöntemiyle olasılıkları ortaya çıkarmaya çalışır (Hacıbeyoğlu, 2006).



Şekil 2.5 Pekiştirmeli Makine Öğrenmesi (Turhost, 2021)

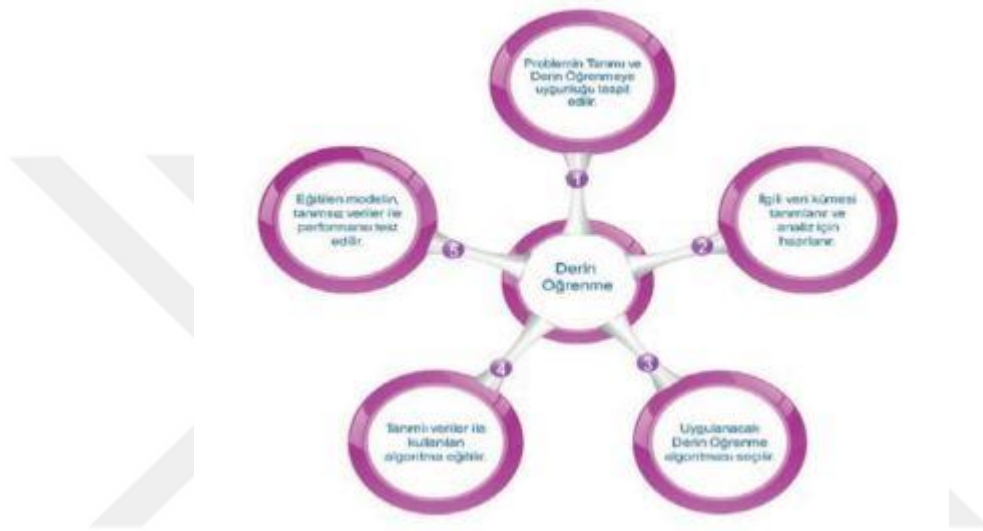
Derin öğrenme modellerinin kullanımı, verilerin soyut özelliklerini daha üst düzeyde temsil ederek ve daha karmaşık problemleri çözerek klasik makine öğrenmesi yöntemlerinden ayrılır (Zhao, Zheng, Xu ve Wu, 2019). Büyük veri setlerinin kullanımı, makine öğrenmesi yöntemlerindeki başarı oranını düşürdüğü için, derin öğrenme modelleriyle çalışmak başarı oranını artırabileceği düşünülmektedir. (Talbi, 2020).

2.3. DERİN ÖĞRENME

Derin öğrenme (DL), makine öğrenmesine kıyasla birçok avantaja sahiptir. Yapay sinir ağları temelli makine öğrenmesi genellikle giriş, gizli ve çıkış katmanları olmak üzere üç katmandan oluşur. Ancak derin öğrenme, gizli katman sayısını artırarak daha fazla veri ve özellik çıkarımını mümkün kılar. Derin öğrenme, her katman geçişinde iç parametreleri güncelleyerek, bilgiyi tutan ağırlıkları nasıl değiştirmesi gerektiğini

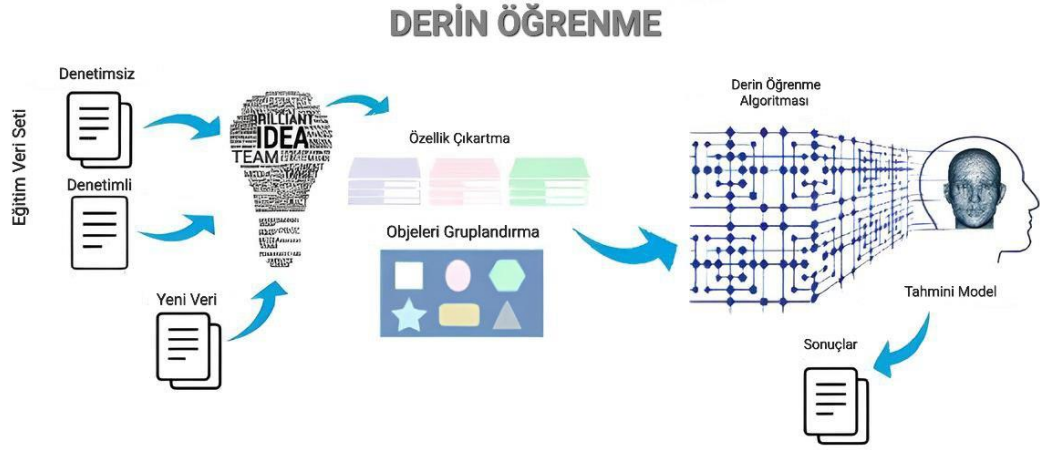
keşfeden geri yayılım algoritması sayesinde çalışır. Derin öğrenme, programcının tüm adımlarını belirlemesi yerine girdi ve çıktı arasında otomatik olarak özellik çıkarımını yapar (Türkdamar, 2023).

Derin öğrenme, verinin temsilini öğrenmeyi amaçlayan bir öğrenme yöntemidir. Bu yöntem, çoklu soyutlama yapısının bir araya getirilmesiyle oluşan çok katmanlı işleme katmanlarından oluşan ve genellikle çok katmanlı sinir ağları olarak adlandırılan bu yapıları kullanan bir yapı olarak bilinmektedir (LeCun, Bengio ve Hinton, 2015).



Şekil 2.6 Derin Öğrenme Tanımı (Yılmaz ve Orman, 2021)

Derin öğrenme algoritmalarının geliştirilmesi sayesinde, görüntüler içindeki objelerin tespit edilmesi ve bu objeler arasındaki ilişkiler de analiz edilebilir hale gelmiştir. Derin öğrenme, görüntülerden özellik çıkarma ve sınıflandırma işlemi üzerine odaklanan makine öğrenmesi ile birleşerek, nesne algılama, tıbbi görüntü sınıflandırma, görüntü ve konuşma tanıma, doğal dil işleme, görüntü iyileştirme, kullanıcılara kişiselleştirilmiş öneriler sunma, siber güvenlik ve sağlık sektöründe hastalıkların otomatik olarak teşhis edilmesi gibi alanlarda geniş bir uygulama alanı bulmuştur (Goodfellow vd. 2017).



Şekil 2.7 Derin Öğrenme Mimarisi (Onar, 2020)

Görüntü sınıflandırma, dil öğrenme, video analizi ve konuşma ses tanısı gibi süreçlerde derin öğrenme uygulamaları, Grafik İşlemci Üniteleri (GPU) kullanılarak yürütülmektedir. GPU teknolojisinin kullanımı, yüksek miktarda etiketli veri setlerinden özellik çıkarma yapan sistemlerin kurulmasını sağlar ve derin öğrenme alt yapılarını daha fazla veri ve daha az güç kullanarak destekler. Temelde derin öğrenme, çok katmanlı işleme yapan algılayıcı bir modelden oluşur (Feldman ve Anderson, 2012).

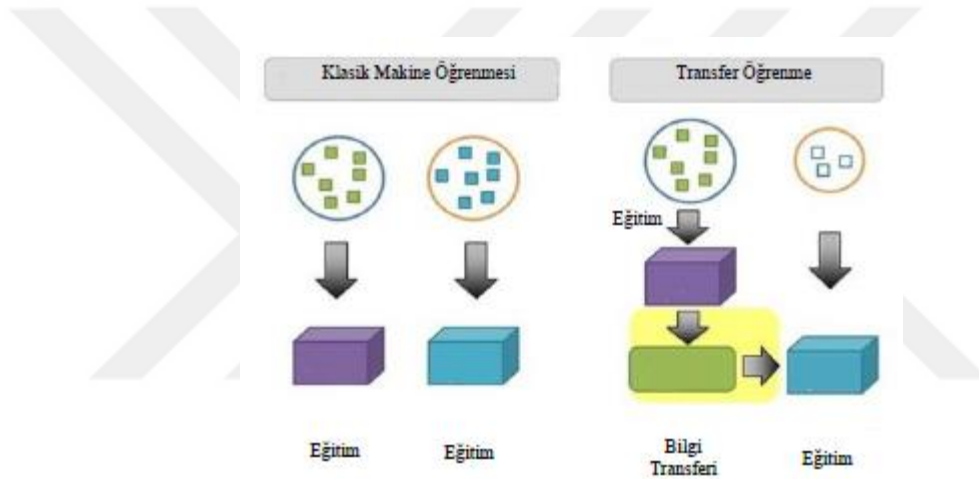
Konuşma tanıma alanındaki Derin Öğrenme kullanımındaki artış, daha eski bir tarihe, 1990'lı yıllara kadar uzanmaktadır. Ancak, özellikle son birkaç yılda, Derin Öğrenme tekniklerinin sesle duygu tanıma konusundaki kullanımında belirgin bir artış görülmüştür.

2.4. TRANSFER ÖĞRENME

Transfer öğrenme (TL), önceden öğrenilmiş bilgilerin yeni bir alana uygulanması amacıyla kullanılan bir makine öğrenmesi tekniğidir. Bu yöntem, yeni etki alanının, daha önce öğrenilenle benzer özelliklere sahip olması gerektiği temeline dayanır. Temel motivasyonu, yeni problemleri daha hızlı, daha etkili ve daha verimli bir şekilde çözmektir (Top, 2018).

Transfer öğrenme, bir makine öğrenimi algoritması türü olmaktan ziyade, öğrenme modellerini eğitirken kullanılan bir teknik veya yaklaşımdır. Önceki eğitimden elde edilen bilgi, yeni bir görevin gerçekleştirilmesine yardımcı olmak için tekrardan kullanılır. Orijinal eğitilmiş model, genellikle yeni ve görünmeyen verilere adapte

olabilmesi için yüksek düzeyde genelleme yeteneği gerektirmektedir. Transfer öğrenme, her yeni görev için sıfırdan eğitim alınması gerekmeyen bir yaklaşım sunmaktadır. Yeni makine öğrenimi modellerini eğitmek için hem kaynaklar hem de zaman açısından maliyetli olabilmektedir. Bu nedenle, transfer öğrenme, kaynaklardan ve zamandan tasarruf sağlar. Büyük veri kümelerini doğru bir şekilde etiketlemek de genellikle çok zaman almaktadır. Kuruluşların karşılaştığı verilerin çoğu, özellikle bir makine öğrenimi algoritmasını eğitmek için gerekli geniş veri kümeleriyle uyumsuz olabilir. Transfer öğrenme, bir modelin mevcut bir etiketli veri kümesinde eğitilebileceği ve daha sonra benzer ancak etiketsiz verilere uygulanabilecek bir görevde kullanılabileceği bir yöntem sunmaktadır (Özaydın, 2023).



Şekil 2.8 Transfer Öğrenme Mimarisi (Jain Anchit, 2018)

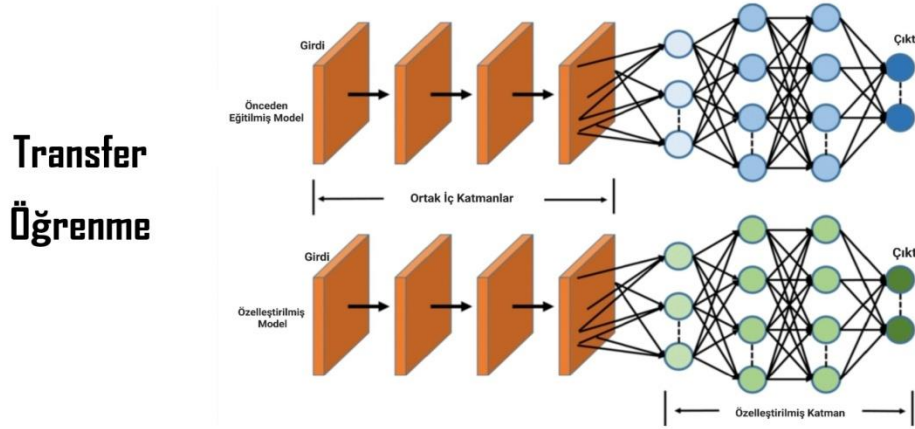
Transfer öğrenme, önceden eğitilmiş bir model kullanarak yeni bir görevde performansı artırmayı amaçlayan bir makine öğrenimi stratejisidir. Bu önceden eğitilmiş model, genellikle büyük ve karmaşık bir veri seti üzerinde önceden eğitilmiş olduğu için öğrendiği özellikleri, ilişkileri ve desenleri temsil eden içsel parametreleri içermektedir.

Önceden eğitilmiş model, genellikle bir temsil öğrenme göreviyle ilişkilidir. Temsil öğrenme, modelin veriyi anlaması, genelleme yapması ve belirli bir konuda bilgi edinmesi için içsel temsiller oluşturmasını ve optimize edilmiş halini içermektedir.

Transfer öğrenme sürecinde, önceden eğitilmiş model, daha önce öğrenilen temsilleri, özellikleri ve desenleri yeni bir görevde kullanmak üzere adapte etmeye çalışır. Bu genellikle yeni görevin gereksinimlerine uyacak şekilde modelin ince ayarını içerir ve

önceden eğitilmiş model, daha önce öğrenilen genel bilgileri, benzer bir görevde veya aynı bağlamda kullanılabilir hale getirmek için özelleştirilir.

Bu yaklaşım, özellikle sınırlı etiketli veri setleri veya yeni görevler için yeterli veri bulunmadığı durumlar gibi senaryolarda, modelin genel bilgilerini kullanarak öğrenmeyi hızlandırabilir ve başarı şansını artırabilmektedir.



Şekil 2.9 Transfer Öğrenme Uygulaması (Dwivedi Rohit, 2020)

Transfer öğrenme ve transformatörler, doğal dil işleme uygulamalarında sıkça kullanılan ve tamamlayıcı tekniklerdir. Geniş veri setleri üzerinde eğitilen önceden eğitilmiş transformatör modelleri, genel dil bilgisi ve özelliklerini öğrenir. Bu modeller daha sonra belirli bir göreve veya veri kümesine özgü hale getirilmek üzere ince ayarlanır.

Transformatörler ilk defa 2017 yılında "İhtiyacınız Olan Tek Şey Dikkat" başlıklı makalede tanıtılmıştır. Özellikle doğal dil işleme (NLP) modellerinde büyük bir etki yaratmış ve büyük bir başarı elde etmiştir. Transformatör modelinin tanıtıldığı bu makale, önceki modellere kıyasla özellikle paralel hesaplamalar ve uzun mesafeli bağımlılıkları daha etkin bir şekilde ele almış bir mimari sunmuştur.

Transformatörler genel olarak çok başlı dikkat mekanizması üzerine kurulmuştur. Dikkat mekanizmasındaki her başlığın girişindeki her elemanın diğer elemanlarına olan önemini öğrenme yeteneği sağladığı ve özellikle dil işleme görevlerinde kelime bağlamını anlama açısından güçlü olmasına katkıda bulunduğu savunulmuştur. Bu model mimarisi gereği tekrar eden katmanlardan oluşur ve her katmanda dikkat mekanizması birden fazla katman içermektedir. Transformatörler modelin sıralı veri setleri üzerinde

etkili olabilmesi ve giriş elemanlarının sırasını tanımlayabilmek için pozisyonel kodlama kullanılmaktadır. Sonuç olarak transformatör modelleri, kodlayıcı-kod çözücü mimarilerinde en yaygın kullanılan tekrarlayan katmanları çok başlı öz dikkat ile değiştirerek, tamamen dikkate dayalı ilk dizi dönüştürme modeli olarak bilinen modellerdir (Vaswani vd. 2017).

Transformatör modelinde temsil edilen öz dikkat mekanizmaları, dil anlama ve bilgisayarla görme uygulamaları dahil olmak üzere çeşitli makine öğrenimi problemlerinin ilerlemesine katkıda bulunmuştur. Evrimsel Sinir Ağları veya Tekrarlayan Sinir Ağları gibi mimarilerden farklı olarak, Transformatörler bir dizideki her bir öge çifti arasında doğrudan etkileşim sağlayıp uzun vadeli bağımlılıkları yakalamada güçlü hale getirmiştir (Zhai vd. 2021).

Transformatör modelleri, özellikle büyük dil modelleri olarak bilinen önceden eğitilmiş modeller, doğal dil işleme alanında önemli bir yer tutmaktadır. Transfer öğrenme, önceden eğitilmiş transformatör modellerini alır ve belirli bir görev için adapte etmeye çalışır. Bu, modelin genel dil bilgisi ve özelliklerini, daha spesifik bir görevde kullanılabilir hale getirmeyi içerir.

Bu yaklaşım, genellikle daha az veri kullanarak daha yüksek performans elde etmeye yardımcı olmaktadır. Önceden eğitilmiş transformatör modelleri, geniş veri setlerindeki dil desenlerini kapsamlı bir şekilde öğrenir ve bu nedenle daha spesifik görevler için daha iyi bir başlangıç noktası sağlar. Transfer öğrenme, bu genel bilgiyi daha özel bir bağlamda kullanmak için bu modelleri ayarlamayı amaçlar, bu da doğal dil işleme uygulamalarında etkili ve verimli çözümler sağlamaya yardımcı olur.

Transformatör mimarisi temelinde birçok farklı model ve model ailesi geliştirilmiştir. Transformatörler, özellikle dil işleme alanında bir dönüm noktası olarak kabul edilen Transformatörlerden Çift Yönlü Kodlayıcı Temsilleri (BERT)(Devlin vd. 2019), Üretken Önceden Eğitilmiş Transformatör (GPT)(Radford vd. 2018), Ses Spektrogram Transformatörü (AST)(Gong vd. 2021) ve Wav2Vec2 gibi modellerde yaygın olarak kullanılmaktadır. Bu modeller, çeşitli dil işleme görevlerinde üstün performans sergileyerek dil modellemesi için yeni bir evrenin başlangıcını temsil etmektedir.

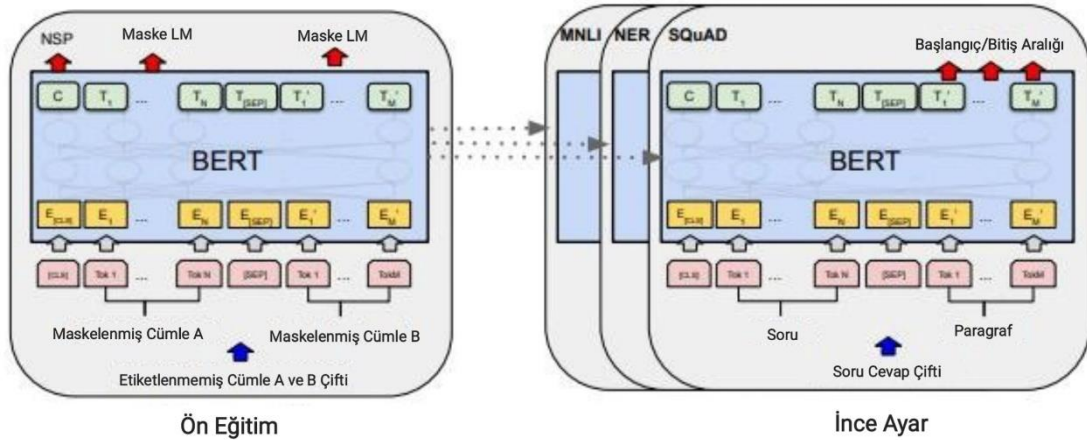
Transformatörlerden Çift Yönlü Kodlayıcı Temsilleri (BERT): Transformatörlerden türetilen BERT, yeni bir dil temsil modelini tanıtmaktadır. Bu model, diğer dil temsil modellerinden farklı olarak modelin kelimenin bağlamını anlama yeteneğini artırmak

adına transformatör kodlayıcı olarak tasarlanmıştır. Tüm katmanlarda sol ve sağ bağlamı aynı anda ele alarak etiketsiz metinlerden çift yönlü temsilleri önceden eğitmektedir.

BERT' in önemli özelliği, öğrenilen dil temsillerini elde etmek için bir kodlayıcı kullanmasıdır. Kodlayıcı önceden eğitilmiş geniş bir veri kümesinden elde edilen temsilleri içerdiği için büyük bir öneme sahiptir. Bu temsiller, ardından görevle ilgili bir çıktı katmanı eklenerek belirli görevlerde kullanılabilir. Bu sayede, önceden eğitilmiş bir BERT modeli, göreve özgü mimari değişiklikler olmaksızın soru yanıtlama veya dil çıkarımı gibi çeşitli görevlerde tercih edilmektedir.

Kodlayıcı bloğunda giriş metinler kelime gömme katmanına geçirilir ve bu katman her kelimeyi vektör uzayında bir noktaya dönüştürür. BERT kodlayıcı bloğunda her kelimenin cümledeki konumunu temsil etmek için pozisyon kodlama yapmaktadır. BERT modeli, birden çok katmanlı transformatör bloğu çift yönlü dikkat mekanizması ve birkaç tam bağlı katman içermektedir. BERT' in kodlayıcı aşamasının çıkış kısmında, her bir kelimenin zengin bir temsili içeren bir dizi vektörü bulunur. Bu temsiller genellikle bir görev için özel çıkış katmanlarına bağlanarak ince ayar yapılır.

BERT, kavramsal olarak basit olmakla birlikte deneysel olarak da güçlüdür. Bu model, dil anlama yeteneklerini geliştirmek ve çeşitli doğal dil işleme görevlerinde yüksek performans elde etmek için yaygın bir şekilde kullanılmaktadır (Devlin vd. 2019).

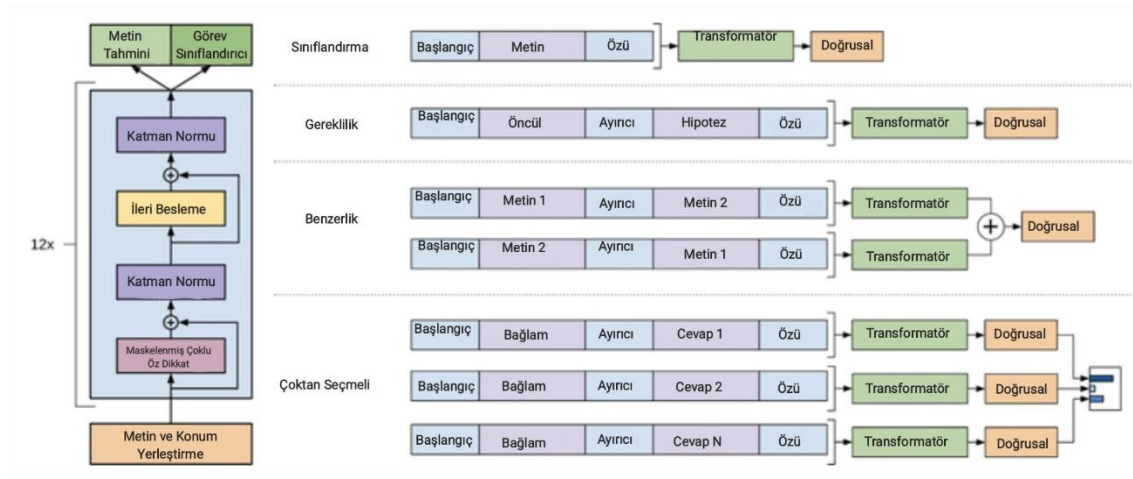


Şekil 2.10 BERT Transformatör Modeli (Devlin vd. 2019)

Üretken Önceden Eğitilmiş Transformatör (GPT): Diğer transformatör modelleri gibi transformatör mimarisine dayanmaktadır. GPT, büyük bir dil modelini denetimsiz bir

şekilde önceden eğitmeyi amaçlamıştır. Model, büyük miktarda metin verisi üzerinde denetimsiz öğrenme yaparak dilin temel yapısını anlamaya çalışır. Bu model, dikkat mekanizmalarını kullanarak uzun menzilli bağlamaları ve dil yapılarını anlamaya yardımcı olur. GPT Modeli, bir dil modeli olarak denetimsiz olarak önceden eğitilir. Dil anlama görevinde hedef görevlere özgü olarak ince ayarlanabilir durumdadır. Bu durum, modelin dilin yapısını öğrenmesine ve genel dil anlama yeteneğini geliştirmesine olanak tanır. GPT Modeli, önceden eğitilmiş denetimsiz temsillerini, daha küçük etiketli veri setleri üzerinde çeşitli görevlere aktararak transfer öğrenme prensiplerinden faydalanmaktadır (Radford vd. 2018).

Doğal dil için denetimsiz temsil öğrenimindeki ilerlemeden esinlenmiş özellikle dil işleme görevlerinde başarılı olduğu bilinen bir mimaridir. Bu transformatör modeli, genellikle büyük ölçekli ve çeşitli internet kaynaklarından toplanan dil verilerinin karışımından oluşan verilerle bu verileri anlamak, dil yapılarını öğrenmek ve genel dil anlama yeteneklerini geliştirmek üzere eğitilmiştir. GPT, modelin kelimenin bağlamını daha iyi anlamasını sağlamak ve dil anlama yeteneklerini artırmak için çift yönlü dil temsili kullanmaktadır. Dolayısıyla da dil anlama, özetleme ve metin oluşturma gibi uygulama alanlarında tercih edilmektedir (Chen vd. 2020).

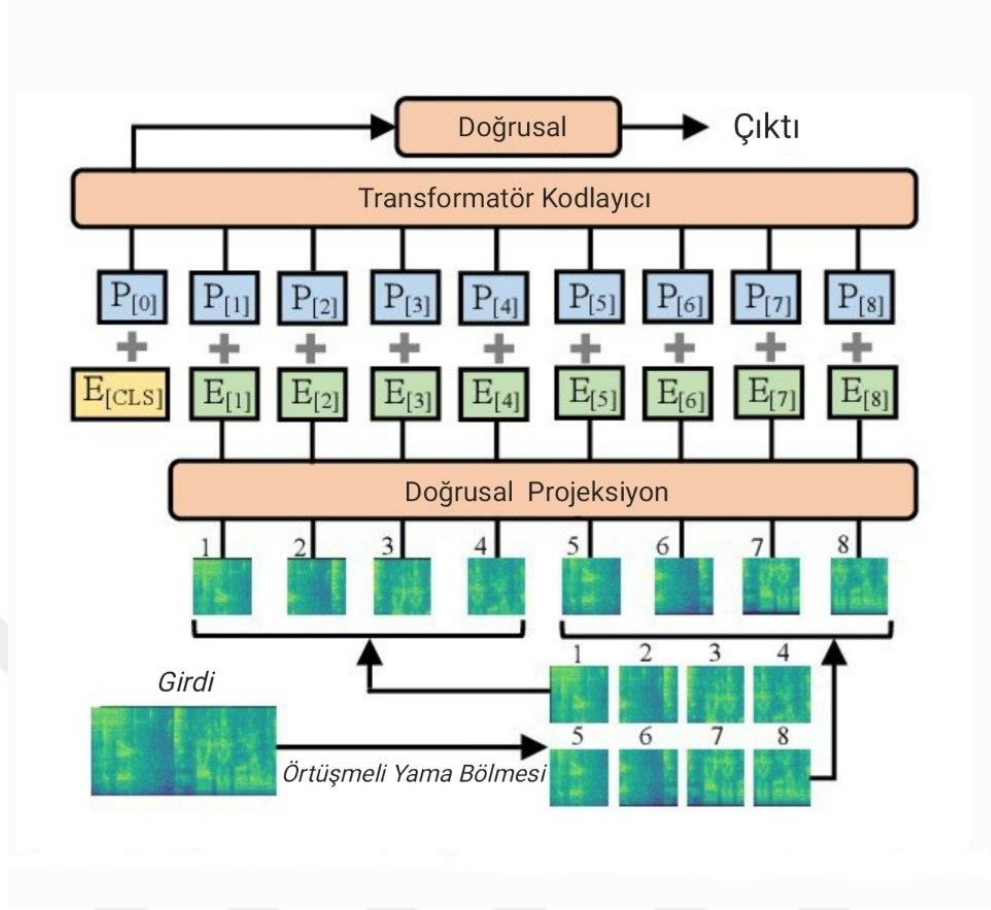


Şekil 2.11 GPT Transformatör Modeli (Radford vd. 2018)

Ses Spektrogram Transformatörü (AST): Ses Spektrogram Transformatörü, doğal dil işleme görevleri için önerilen transformatör mimarisine dayanmaktadır. AST, ses verilerini temsil etmek için spektrogramları girdi olarak almaktadır. Spektrogramlar, ses

dalgalarının frekans ve zaman bilgilerini içeren görsel temsillerdir. AST' nin model mimarisi, dikkat mekanizmalarını kullanarak uzun menzilli bağıntıları öğrenme yeteneğine sahiptir. AST' nin en önemli özelliği dikkat mekanizmasıdır. Dikkat mekanizması, modelin girdinin farklı kısımlarına odaklanmasına ve önemli bilgileri vurgulamasına olanak tanır. Bu durum, özellikle uzun süreli bağıntıları anlama ve temsil etme açısından önemlidir. AST, genellikle büyük bir ses veri seti üzerinde önceden eğitilir. Bu eğitim sürecinde, model, veri setindeki ses spektrogramları üzerinde belirli bir görevi çözmek için öğrenir. Ses spektrogramları, modelin anlaması için sayısal özelliklere dönüştürülür. Bu adım, önceden eğitilmiş bir model veya özel bir özellik çıkartma katmanını içerir. AST modeli, önceden eğitilmiş modellerin özelliklerini kullanarak belirli bir ses işleme görevi için ince ayar yapılmaktadır. Eğitildikten sonra, ses verilerinden anlamlı temsiller çıkartabilir veya belirli bir sınıflandırma görevi için kullanılabilir hale gelmektedir.

Son on yılda, evrişimli sinir ağları, ses sınıflandırması için yaygın bir model bileşeni haline gelmiş olup ses spektrogramlarından ilgili etiketlere doğrudan eşleme öğrenmeyi amaçlamışken AST, evrişimli sinir ağlarını kullanmaması ve tamamen dikkat mekanizmalarına dayanması bakımından bu çalışmalardan farklıdır. Ses spektrogramları, ses sinyallerini zaman-frekans alanında temsil etmek için kullanılan bir görselleştirme yöntemidir. Ses Spektrogram Transformatör modelleri, dil işleme, görüntü işleme ve ses işleme gibi çeşitli görevlerde başarıyla kullanılan derin öğrenme modeli olarak tercih edilen bir transformatör haline gelmiştir (Gong vd. 2021).



Şekil 2.12 AST Transformatör Modeli (Gong vd. 2021)

Tez çalışmasında transformatör modellerinden Wav2Vec2 modeli kullanılmıştır.

Wav2Vec2: Ham ses verilerinden temsillerin kendi kendine denetimli öğrenimi yapan transformatör modelidir. Wav2Vec2 modeli, büyük bir ses veri seti üzerinde önceden eğitilir. Ses veri seti genellikle geniş bir konuşma kayıtlarını içerir ve model, ses dalgalarını anlamlı temsillere çevirmeyi öğrenir. Model, karşılaştırmalı öğrenme prensiplerine dayanarak eğitilir. Karşılaştırmalı öğrenme, bir ses örneği ile ilgili (pozitif örnek), diğer ses örnekleri ile (negatif örnekler) karşılaştırılmasını içermektedir. Wav2Vec2 modeli, aynı konuşmacıya ait ses örneklerini birbirine daha yakın, farklı konuşmacılara ait örnekleri ise birbirine daha uzak konumlandırmayı öğrenir. Wav2Vec2, ses dalgalarından özellik çıkartma yetenekleriyle öne çıkmaktadır. Ses dalgalarını analiz ederek, model sesin yüksek seviyede özelliklerini çıkartır ve bunları anlamlı temsillere dönüştürür. Wav2Vec2, kuantalı çıkartma adı verilen bir teknikte çalışmaktadır. Kuantalı çıkartma, özelliklerin daha az bit kullanarak temsil edilmesini sağlar ve daha hafif modeller elde edilmesine yardımcı olur (Baeovski vd. 2020).

Wav2Vec2, yarı-denetimli öğrenme yaklaşımını benimsemiştir. Bu, modelin sadece etiketlenmiş ses verileriyle değil, aynı zamanda etiketlenmemiş verilerle de eğitilmesini içerir. Bu sayede genel ses özelliklerini daha iyi öğrenmesi ve genelleme yeteneğini artırması amaçlanır. Wav2Vec2 modeli, önceden eğitilmiş bir model olarak sunulmakta ve daha spesifik görevler için ince ayar yapmak üzere kullanılabilir (Baevski vd. 2020).

Wav2Vec2, son zamanlarda ses verileri üzerinde dil işleme görevlerinde popüler hale gelen bir dil modelidir. Özellikle konuşmayı metine dönüştürme ve ses duygu tanıma gibi ses işleme görevlerinde yüksek performans sergiler. Bu model, konuşma tanıma görevleri için özel olarak tasarlanmış olup doğal dil işleme alanında önemli bir konuma sahiptir.

Wav2Vec2 modelinde özellik çıkartma aşamasında Evrişimsel Sinir Ağı kullanılmaktadır. Wav2Vec2 modeli, ses dalgalarını işlemek için özel olarak tasarlanmış bir dizi dönüşüm katmanı içerir. Bu dönüşüm katmanları, giriş ses dalgalarını yüksek seviyeli özelliklere dönüştürmek için evrişim işlemlerini içermektedir. Wav2Vec2 modelindeki evrişim katmanları, özellik çıkartma aşamasında ses dalgalarını analiz etmek ve giriş veriyi daha yüksek seviyeli özellikler haline getirmek için kullanılır.

Wav2Vec2, geniş bir veri kümesinde eğitildiği için çeşitli dil desenlerini kapsamlı bir şekilde öğrenir. Bu model, özellikle konuşmacı değişiklikleri ve gürültü gibi ses verilerindeki zorlukları ele alarak yüksek doğruluk sağlar.

Wav2Vec2, ses verileri üzerinde dil işleme görevlerinde başarılı performansı ile dikkat çeker ve bu alanda birçok uygulama için etkili bir araç olarak kullanılır (Bayraktar, 2023).

Wav2Vec2 transformatör modelinin son dönemlerde yapılan çalışmaları aşağıdaki çizelgede verilmiştir.

Çizelge 2.1 Wav2Vec2 Üzerine Yapılan Çalışmalar

Referans Adı	Yayınlan ma Yılı	Kullanılan Veri seti	Kullanım Amacı	Kullanılan Performan s Metriği	Başarım Oranı(%)
wav2vec 2.0: A Framework for Self- Supervised Learning of Speech Representatio ns (Baevski vd.,2020)	2020	Librispeech	Konuşma Tanıma	Kelime Hata Oranı	4.8
Unsupervised Speech Recognition (Baevski vd.,2021)	2021	Librispeech	Konuşma Tanıma	Fonem Hata Oranı	11.3
Wav2vec-C: A Self- supervised Model for Speech Representatio n Learning (Sadhu vd.,2021)	2021	Gerçek Dünya Uzak Alan Gürültülü Verileri	Konuşma Tanıma	Kelime Hata Oranı	0.7
Emotion Recognition from Speech UsingWav2ve c 2.0 Embeddings (Pepino vd., 2021)	2021	IEMOCAP RAVDESS	Duygu Tanıma	Doğruluk Oranı	IEMOCAP: 64.3 RAVDESS: -

Çizelge 2.2 (devamı). Wav2Vec2 Üzerine Yapılan Çalışmalar

Robust Wav2vec 2.0: Analyzing Domain Shift In Self-Supervised Pre-Training (Hsu vd., 2021)	2021	Sesli Kitaplar	Konuşma Tanıma	Kelime Hata Oranı Azaltma Oranı (WERR)	73
Brazilian Portuguese Speech Recognition Using Wav2vec 2.0 (Gris vd.,2021)	2021	BP (Brezilya Portekizcesi) Verileri	Konuşma Tanıma	Kelime Hata Oranı (WER)	10.5
Fine-Tuning Wav2vec2 For Speaker Recognition (Vaessen vd.,2021)	2021	VoxCeleb Veriseti	Konuşmacı Tanıma	Eşit Hata Oranı (EER)	1.88

Çizelge 2.3 (devamı). Wav2Vec2 Üzerine Yapılan Çalışmalar

Improving Mispronunciation Detection with Wav2vec2-based Momentum Pseudo-Labeling for Accentedness and Intelligibility Assessment (Yang vd., 2022)	2022	L2 Speech	Yanlış Telaffuz Tespiti	Fonem Hata Oranı Azaltma	5.35
Exploring Wav2vec 2.0 FineTuning For Improved Speech Emotion Recognition (Chen vd., 2023)	2023	IEMOCAP	Duygu Tanıma	Ağırlıklandırılmamış Doğruluk Oranı	74.3

2.5. LİTERATÜR ÇALIŞMALARI

Literatürde, ses duygu analizi ve derin öğrenme yöntemleri üzerine yapılan çalışmalar incelenmiş ve bu alanda elde edilen başarılar ile eksiklikler belirlenmiştir. Ayrıca, transfer öğrenme yöntemi ve transformatörler, özellikle ses duygu analizinde, benzer çalışmalarda nasıl başarıyla kullanıldığına dair örnekler incelenmiştir.

2020 yılında yapılan bir yüksek lisans tezinde, koklear implant kullanan bireylerin duygularını sözel olarak ifade etme yeteneklerini değerlendirmek amacıyla derin öğrenme tekniklerini kullanmaktadır. Çalışmada, katılımcıların eğitim öncesi ve eğitim sonrası ses tabanlı duygu analizleri üzerinden derin öğrenme modeli değerlendirilmiştir.

Eğitim öncesi ve eğitim sonrası analizler arasındaki farklar, katılımcıların koklear implant kullanımının duygusal ifade becerilerine etkisini ölçmek amacıyla incelendiği görülmüştür (Uludağ, 2020).

Algoritma eğitimi için hazırlanmış olan veri seti, 18-40 yaş aralığındaki toplam 40 bireyden oluşmaktadır. Bu bireylerden 20'si normal işitme kapasitesine sahipken, diğer 20 birey koklear implant kullanmaktadır. Veri seti nötr, mutluluk, hüznün, öfke, korku ve şaşkınlık olmak üzere altı farklı duygu durumunu içermektedir. Evrişimsel Sinir Ağı (CNN) kullanılarak gerçekleştirilen ses duygu analizi için 40 bireyin seslendirdiği toplam 4800 cümle kaydı kullanılmıştır. Bu cümleler, spektral analiz ile işlenmiş ve gürültüden arındırılmıştır. Ayrıca, daha kaliteli ses kayıtları elde etmek amacıyla her bir cümle 2 kez seslendirilmiştir. Spektral analiz sonrasında 4800 olan ses kaydı yarısına düşmüştür. Derin öğrenme eğitim aşamasında, 2400 ses kaydının %75'i kullanılmıştır. Test aşamasında ise geriye kalan %25'i kullanılmıştır. Eğitim tamamlandıktan sonra, derin öğrenme modelinin test aşamasında elde ettiği doğruluk oranı %93 olarak belirlenmiştir.

Elde edilen sonuçlara göre, derin öğrenme ile ses duygu analizi doğruluğu normal işiten bireylerde %94.3 olarak saptanırken, koklear implant kullanan bireylerde eğitim öncesi %68.25 ve eğitim sonrasında %91.3 olarak ölçülmüştür. Bu sonuçlar, koklear implant kullanan bireylerde derin öğrenme ile gerçekleştirilen ses duygu analizinin eğitim sonrasında önemli bir iyileşme gösterdiğini göstermektedir.

2021 yılında yapılan bir başka tezde ise koklear implantlı, otizm spektrum bozukluğu tanılı, normal gelişim gösteren bireyler ve drama liderlerinin ses-duygu analizlerinin yapılması, sonuçların karşılaştırılması ve değerlendirilmesi amaçlanmıştır (Gök, 2021).

21 hedef cümlelerin bulunduğu veri seti duygu sınıfı cümleleri 7 farklı duyguda (nötr, mutluluk, üzüntü, öfke, korku, şaşkınlık, tiksinti) 20 drama lideri ve 20 normal işiten olmak üzere toplamda 40 birey çalışmaya dahil edilmiştir. Bu cümlelerin içinden 10 tanesi rastgele seçilerek 7 farklı duyguda 30 koklear implant kullanıcısı ve 30 otizm tanılı birey tarafından seslendirilmiş, veri seti oluşturulmuştur. Koklear implant ve otizm grubu aynı cümleleri, oluşturulan eğitim materyalleri uygulanmadan önce ve sonra olmak üzere 2 kez seslendirmiştir. Drama liderleri grubundan 2940 (147 cümle kaydı*20 kişi=2940) ve normal işiten bireyler grubundan 2940 (147 cümle kaydı*20 kişi=2940), işitme kayıplı bireylerden eğitim öncesi 2100 (70 cümle kaydı*30 kişi=2100), eğitim sonrası 2100 (70 cümle kaydı*30 kişi=2100), OSB tanılı bireylerden, eğitim öncesi 2100 (70 cümle

kaydı*30 kişi=2100), eğitim sonrası 2100 (70 cümle kaydı*30 kişi=2100) etiketlenmiş veri seti ses kaydı olmak üzere toplam 14820 kayıt elde edilmiştir. Derin öğrenme algoritmalarından CNN kullanılarak ses duygu analizleri karşılaştırılmıştır. Derin öğrenme modeli FreshNet ile eğitim aşamasında %72.8, test aşamasında %51.6 doğruluk oranı elde edildiği ve diğer test grupları için doğruluk oranları; normal işiten bireyler için %26.6, işitme engelli bireyler için alıştırma materyali öncesi %19.7, alıştırma materyali sonrası %22.3, otizmlili bireyler için alıştırma materyali öncesi %14.71, alıştırma materyali sonrası %16.37 olarak elde edildiği görülmüştür.

Derin öğrenme modelinin uygulandığı başka bir çalışmada veri seti olarak EMO-DB kullanılmış olup 3 farklı duygu sınıfı (öfke, nötr, üzüntü) üzerine çalışılmıştır. Derin Öğrenme algoritmalarından DCNN/10 kullanılarak ses duygu analizleri karşılaştırılmıştır. Veri setinin eğitimi sonucunda %96.97 doğruluğa ulaşıldığı görülmüştür (Harar, Burget & Dutta, 2017).

Zhao, Mao & Chen tarafından yapılan bir çalışmada ise veri seti olarak EMO-DB kullanılmış olup 6 farklı duygu sınıfı (öfke, nötr, üzüntü, heyecanlı, hayal kırıklığı, mutluluk) üzerine çalışılmıştır. DCNN ve LSTM derin öğrenme modelleri kullanılarak duygu analizinde %95.33 'a kadar doğruluğa ulaştıkları görülmüştür (Zhao, Mao & Chen, 2019a).

Shaw, Vardhan & Saxena tarafından yapılan bir çalışmada ise kendi oluşturdukları ses kayıtlarından oluşan veri seti kullanmışlardır. 4 farklı duygu sınıfı (öfke, mutluluk, üzüntü, nötr) üzerine çalışmışlardır. Derin öğrenme algoritmalarından ANN modeliyle ses duygu analizinde %86.87 'a kadar doğruluğa ulaştıkları görülmüştür. (Shaw, Vardhan & Saxena, 2016).

Bir başka literatür çalışmasında, önceden kullanılmamış ses verileri üzerinde duygu analizi yapabilen, kullanıcı etkileşimine uygun ve derin öğrenme temelli bir kompakt sistem geliştirilmesi hedeflenmiştir. Bu sistem, yazılım araçları aracılığıyla kullanıcıların etkileşimli bir şekilde kullanabileceği bir altyapı üzerine kurulmuştur. Bu çalışma, ses verisi üzerinde duygu analizini daha etkili ve kullanıcı dostu hale getirmeyi amaçlamaktadır (Kıvrak, Karasulu, Sozbir&Turkay, 2021).

Derin sinir ağı modelleri, RAVDESS, CREMA-D, Berlin (EMO-DB), ve SAVEE veri kümelerinin birleştirilmesiyle oluşturulan geniş bir veri kümesi üzerinde eğitilmiştir. Bu modeller, ses verilerini girdi olarak alır ve sınıflandırma sonucunda duygu durumunu

belirler. Toplamda 8 farklı duygu durumunu içeren bu sınıflar; doğal, mutlu, üzgün, kızgın, korkmuş, iğrenmiş, şaşırılmış, sıkılmıştır. Eğitim öncesi, veri kümesi beyaz gürültü, kaydırma, esnetme ve perde değiştirme gibi veri artırma yöntemleri kullanılarak zenginleştirilmiştir. Bu veri artırma işlemleri sonucunda toplamda 19636 örnek elde edilmiştir. Bu örneklerin %20'si test kümesinde, %16'sı doğrulama kümesinde, geriye kalan %64'lük kısım ise modelin eğitimi için kullanılmıştır.

Veri setlerini eğitirken derin öğrenme modellerinden LSTM ve CNN kullanılmıştır. Bu yazılımsal araç ile yapılan çalışmada her duygu modelinde farklı doğruluk değerlerine ulaşıldıkları görülmüştür. Doğal duygu sınıfında %92.23, mutlu duygu sınıfında %91.01, üzgün duygu sınıfında %91.34, kızgın duygu sınıfında %93.55, korkmuş duygu sınıfında %88.92, iğrenmiş duygu sınıfında %90.70, şaşırılmış duygu sınıfında %99.54, sıkılmış duygu sınıfında %99.87 doğruluk değerlerine ulaşıldıkları görülmüştür.

2020 yılında yayınlanan bir konferans özetinde (LREC 2020) ; temel veri seti olarak Türkçe Duygu-Ses Veritabanı (TurEV-DB) kullanılmıştır. Bu veri seti 6 amatör aktörün seslendirmelerinden olup 82 kelimenin geçtiği 1738 cümleden oluşmaktadır. Bu veri seti 4 duygudan (kızgın, sakın, mutlu ve üzgün) oluşmuş olup SVM makine öğrenmesi modeli ve CNN derin öğrenme modeliyle eğitilip duygu sınıflandırması yapılmıştır. Eğitim sonunda maksimum % 79 doğruluk oranına ulaşıldıkları görülmüştür (Canpolat, Ormanoğlu & Zeyrek, 2020).

Bir başka çalışmada; bir duygusal durumun tanınması genellikle konuşmacının cinsiyetinin tanınması için düşünülen aynı özelliklerin analizine dayandığı, ancak erkek ve kadın konuşmacılar söz konusu olduğunda duygu tanıma sonuçlarının eşit olmadığı görülmüş olup cinsiyet temelli ses duygu analizi geliştirmesi amaçlanmıştır (Mura, Lamberti, 2020).

Veri seti nötr, mutlu, üzgün, korkmuş, sinirli olmak üzere 5 farklı duygu sınıfından oluşmaktadır. Cinsiyetten bağımsız duygu tanıma sistemlerinin, referansa göre %2 ile %4 arasında değişen bir farkla, cinsiyetten bağımsız duygu tanıma sistemlerinin, ön cinsiyet ayrımını uygulayan sistemlerden daha düşük başarı oranlarına sahip olduğu gösterildiğinden, duygu tespitine yönelik bu cinsiyet temelli yaklaşımın başarılı olduğu kanıtlanmıştır. HMM derin öğrenme modeliyle eğitimi ve SVM sınıflandırıcılarıyla, cinsiyet sınıflandırmasından sonra duygu tanıma başarı oranının %85'ten %92.5'e yükseldiği görülmüştür.

2020 yılında yapılan bir başka çalışmada; müşteri hizmetleri, eğitim ve sağlık hizmetlerinde kullanılmak üzere duygu tanıma uygulamaları sunmak amaçlanmıştır (Anithadevi vd. 2020).

Yapılan bu çalışma MFCC ve GMM modellemesine dayanmaktadır. Bu on yıl içinde MFCC, konuşmacının sese dayalı sınıflandırması için yaygın olarak kullanıldığı. MFCC'nin baskınlığı, çeşitli sınıflandırıcılar üzerindeki yüksek performansı ve doğruluğundan kaynaklandığı düşünülmektedir. GMM modelinin, sınıflandırma için oldukça iyi olduğu ve nispeten daha yüksek doğruluk sağladığı. UBM modeli ile birlikte daha iyi doğruluk sonuçları verdiği görülmüştür. Bu çalışma kapsamında önerilen bu sistem, amazon alexa tarafından bireyin duygularını belirlemek için kullanılabilecek bir bireyin duygularını bulmak için bir uygulama olarak kullanılmaktadır. Veri setinde sinirli, iğrenmiş, korkmuş, mutlu, nötr, şaşırılmış, üzgün olmak üzere 7 duygu sınıfından oluşmaktadır. Bu çalışmada MFCC ile özellikler çıkarılmış olup GMM ve GMM-UBM ile ayrı ayrı eğitimler gerçekleştirilmiş olup GMM modelinde %80.6 GMM-UBM modelinde ise 92.8 doğruluk oranına ulaştığı görülmüştür.

İncelenen bir başka literatür çalışmasında; üretken gürültü modeline sahip çok koşullu Otomatik Konuşma Duygu Tanıma (SER) sisteminin, geleneksel kaydedilen gürültülerden daha güvenilir ve gürültü sağlamlığına karşı kararlı olduğunu göstermek amaçlanmıştır (Tiwari vd. 2022).

Veri seti olarak oldukça temiz bir ortamda kaydedilmiş, 7 duygu kategorisini ortaya çıkaran 535 eylemli sözceden oluşan Berlin Duygusal Veri Tabanı (Emo-DB) ve Etkileşimli Duygusal İkili Hareket Yakalama Veritabanı (IEMOCAP) kullanılmıştır. Emo-DB veri setinde, katılımcılar ezberledikleri senaryoyu prova ederken, IEMOCAP veri setinde katılımcılar belirli duyguları ortaya çıkarmak için tasarlanmış bazı varsayımsal durumları doğaçlama yapmışlardır. Bu çalışmada 4 duygu kategorisinden (Mutlu, Öfke, Nötr, Üzgün) senaryo ve doğaçlama örnekleri üzerinde deneyler yapılmıştır. Veri setinin %80'ini eğitim için ve %20'sini test için kullanmışlardır. Derin öğrenme modellerinden DNN/3 ve Generative gürültü modeli kullanılmıştır. Veri setinin eğitilmesi sonucunda Emo-DB veri setinde %82.73, IEMOCAP veri setinde %62.74 doğruluk oranına ulaştıkları görülmüştür.

2020 yılında yayınlanan bir konferans özetinde (ECTI DAMT & NCON) ; derin öğrenme tekniğinin uygulanmasıyla Tay dili için olumsuz duygu tanıma için bir çerçeve önerilmesi

amaçlanmıştır (Mekruksavanich, 2020).

On farklı konuşmacının seslerinin kaydedilmesiyle duygusal konuşma veri setinin toplanmasını takiben, çalışma, hangi derin öğrenme sınıflandırıcısının makul bir performans sağladığını belirlemek amacıyla SAVEE, RAVDESS, TESS ve CREMA-D olmak üzere dört farklı duygusal konuşma veri seti kullanılmıştır. SAVEE veri seti 400 cümleden, RAVDESS veri seti 7356 cümleden, TESS veri seti 2800 cümleden oluşmakta olup 7 farklı duygu sınıfını (nötr, mutlu, üzgün, kızgın, korkmuş, iğrenmiş, şaşırılmış) kapsamaktadır. CREMA-D veri seti ise 7442 cümleden oluşmakta olup 6 farklı duygu sınıfını (nötr, mutlu, üzgün, kızgın, korkmuş, iğrenmiş) kapsamaktadır. Veri setinin özellik çıkarımında MFCC, eğitiminde derin öğrenme modeli olarak DCNN/6 kullanılmıştır. Veri setlerinin eğitilmesi sonucunda SAVEE veri setinde % 65.83, RAVDESS veri setinde %75.83, TESS veri setinde %55.71, CREMA-D veri setinde %65.77 doğruluk oranlarına ulaşıldığı görülmüştür.

İncelenen bir çalışmada, ses dosyaları üzerinde MFCC özellik çıkarımı yapılmış, ardından bunlar derin öğrenme modeller için girdi olarak kullanıldığı gözlemlenmiştir. Derin öğrenme modelleri arasından VGG, MobileNet, Xception ve ResNet kullanılarak akciğer sesi veri kümesini sınıflandırmak için önceden eğitilmiş derin öğrenme tabanlı modeller kullanılmıştır. Önceden eğitilmiş modellerin akciğer seslerini sınıflandırmada büyük veri kümeleri üzerinde eğitilmiş olması nedeniyle, düşük eğitim süresiyle çok daha iyi sınıflandırma doğruluğuna ulaştığı görülmüştür (Idrisse, 2023).

Bir diğer çalışmada, EMO-DB, RAVDESS, TESS, CREMA-D, SAVEE gibi ses veri setleri kullanılarak, duygusal ifadelerin doğru bir şekilde tanınması için eğitim gerçekleştirilmiştir. Duygusal ses analizinde MFCC, MelSpektrogramları ve MFCC + MelSpektrogramları gibi özellik çıkarım yöntemlerinin derin öğrenme yöntemleri ile makine öğrenmesi sınıflandırma algoritmalarından daha etkili olduğunu ve başarılı olduğu sonucuna varılmıştır (Bayraktar, 2023).

2022 yılında yapılan bir çalışmada ise 375 deneğe ait COVID-19 sonucu pozitif ve 375 deneğe ait COVID-19 verisi negatif olmak üzere 750 deneğe ait test sonucunun öksürük sesi veri kümesi kullanılarak oluşturulan veri kümesi kullanılarak COVID-19 tahmini için sınıflandırma yapılmıştır. Sınıflandırma sonucunda öznelik çıkarma yöntemlerinden sık kullanılan MFCC gibi yöntemlerin yanında Spektrogram, Sıfır Geçiş Oranı, Spektral Azalma, Spektral Ağırlık Merkezi, Ton, Karekök Ortalama gibi yöntemlerin kullanılması

doğruluk oranını arttığı görülmüştür. Sınıflandırma yöntemlerinden derin öğrenme yöntemi olan Derin Sinir Ağları (DNN) modelinin, klasik yöntem algoritmalarından daha yüksek bir doğruluk verdiği görülmüştür (Keleş, 2022).

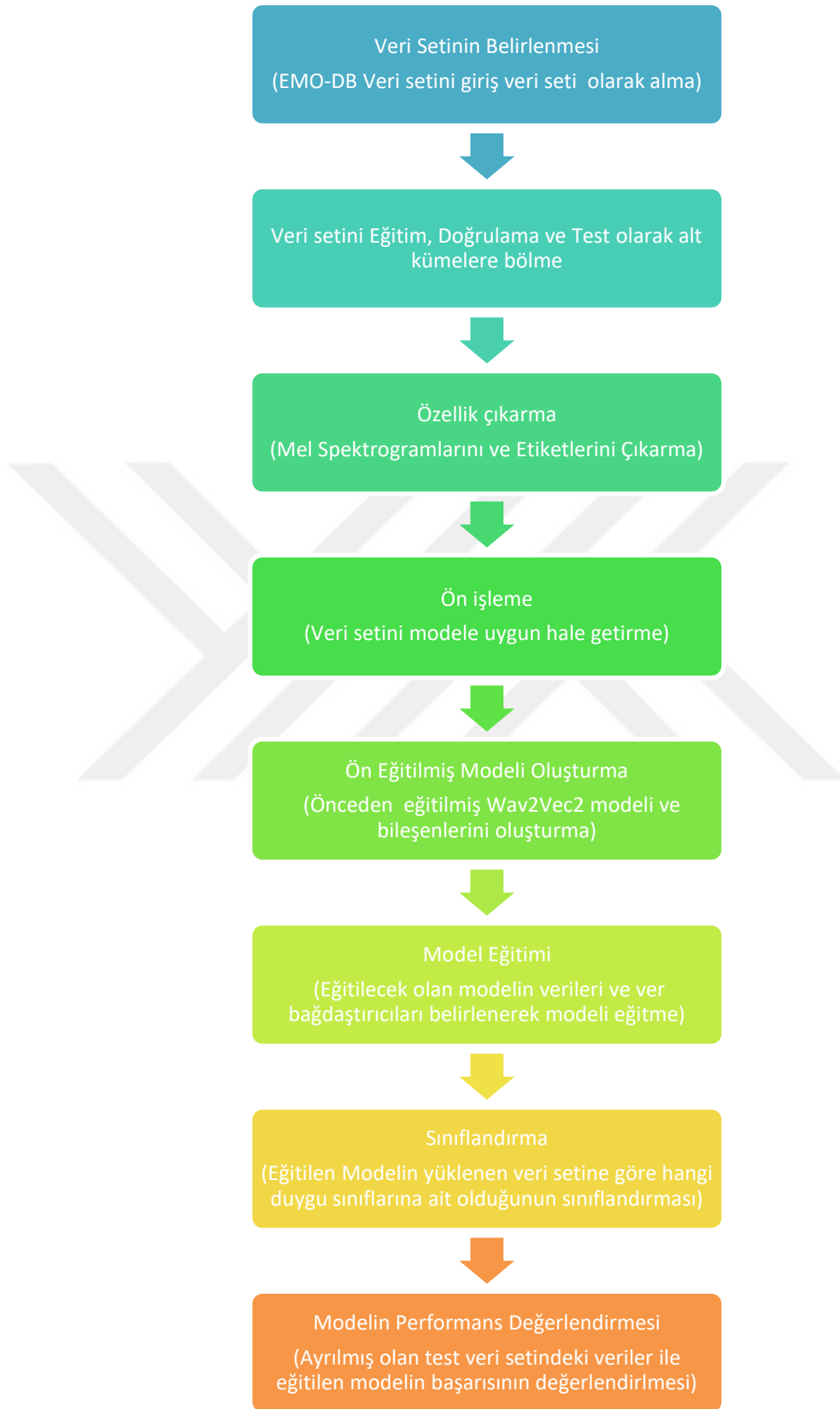


3. MATERYAL VE YÖNTEM

İnsan fiziksel eylemleri, sözleri ve düşünceleri sözlü iletişim halindeyken hem üstü kapalı hem de açık bilgiler sunmaktadır. Konuşma sırasında duygular önemli bir şekilde aktarılır. Duygusal durumların net olarak ölçülemediği durumlara yönelik literatürde bir çok oluşturulmuş veritabanı, bu alandaki araştırmaları doğru yönlendirmek için önemli bir role sahip olmuştur. Konuşma duygusunun tespiti için seçilen veritabanında konuşmacının duygusu, cinsiyeti ve yaşı, ifade tonları ve sesler önemli parametrelerdir.

Ses duygu analizi, bir insanın ses dosyalarının konuşma sinyallerindeki vurgu, tonlama gibi özelliklerinin alınıp işlenerek duygu etiketlerine göre analiz edilip sınıflandırılmasıdır. Ses duygu analizinin günümüzde çağrı merkezlerinde müşteri deneyimlerinin iyileştirilmesi amacıyla, pazar araştırmalarında reklamın etkisinin anlayabilme adına, sağlık alanında sesli asistan uygulamaları gibi bir çok alanda aktif olarak kullanıldığı bilinmektedir

Bu kapsamda çalışma kapsamında tercih edilen veri seti çalışmanın yönünü ve sınıflandırmada yönlendirici olacaktır. Çalışma kapsamında izlediğimiz yol aşağıda detaylı olarak sırasıyla sunulmuştur.



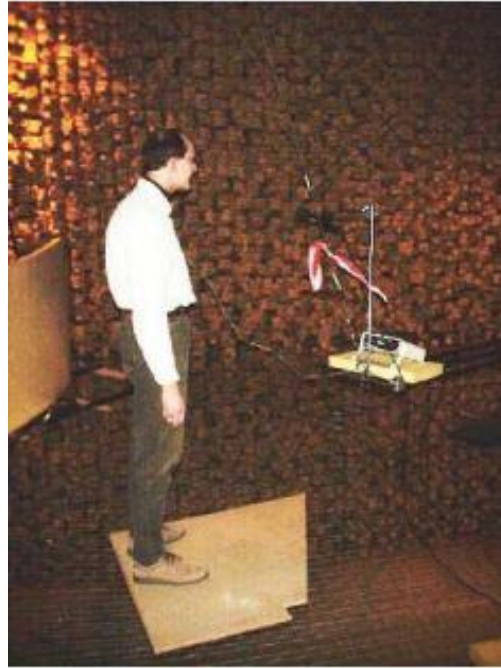
Şekil 3.1 Ses Duygu Analizi Geliştirilen Model Süreci

3.1. VERİ SETİNİN BELİRLENMESİ

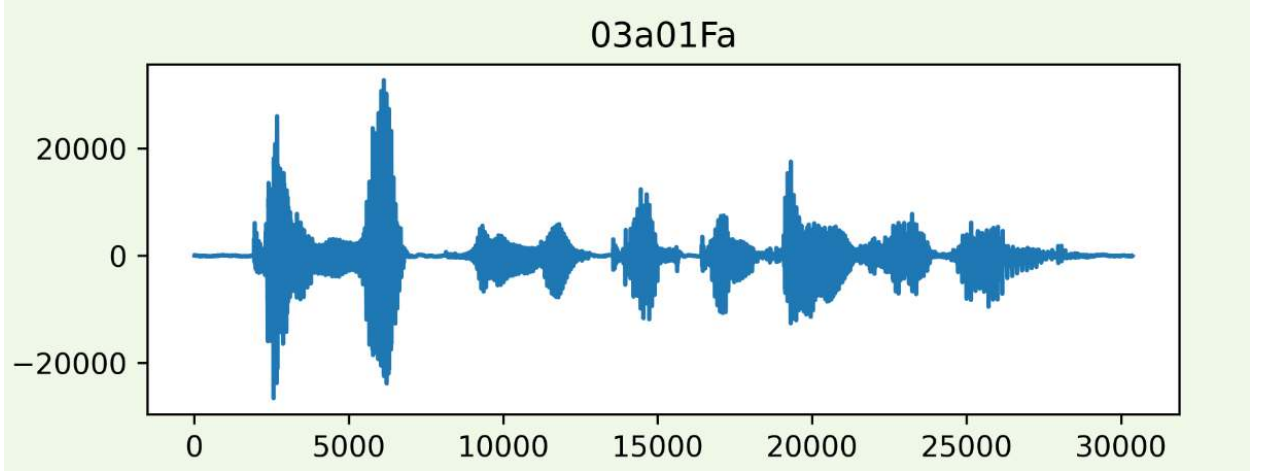
Bu çalışmada Bas Konuş/İnterkom sistemlerinden alınacak verileri temsil etmesi adına bu çalışmada ses duygu analizinde giriş verisi olarak hazır veri seti olan Berlin Duygusal Konuşma Veritabanı (EMO-DB) kullanılmıştır.

Berlin Duygusal Konuşma Veritabanı (EMO-DB), konuşma duygu tanıma için en yaygın kullanılan veri kümelerinden biridir. 10 Almanca cümle, beş kısa cümle ve beş uzun cümleden oluşan simüle edilmiş bir veri setidir. Veri setini oluşturmak için on konuşmacı, beş kadın ve beş erkek kullanılmıştır. Konuşmacıların her biri, beşi uzun, beşi kısa olmak üzere on cümleyi farklı duygularla ifade etmiştir. Ses sinyalinin kaydının yanı sıra prozodik ve ses kalitesinin daha hassas bir şekilde çıkarılabilmesi için seslerle birlikte elektroglossogramlar da kaydedilmiştir. Tüm veri seti, yedi duyguyla hareket eden 10 cümleden oluşan 535 örnek içermektedir. Bu veri seti için seçilen duygular nötr, kızgın, korkmuş, mutlu, üzgün, iğrenmiş ve sıkılmıştır (Abbaschian vd. 2021).

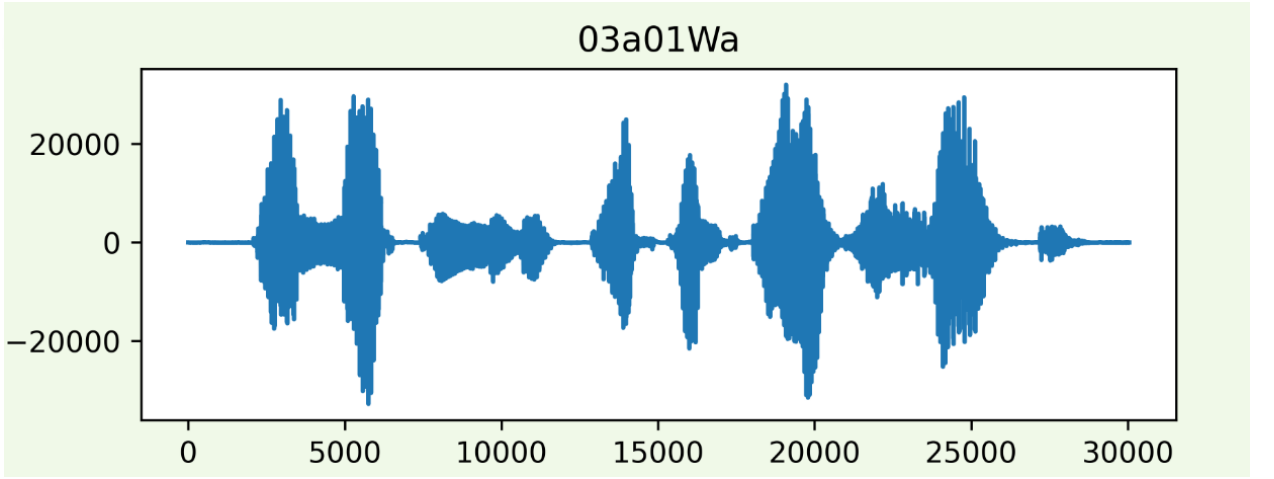
Bu çalışma kapsamında; bas konuş sistemlerde acil durumlarda herhangi bir önceliklendirmeye gerek duyulmayacağı için nötr ve sıkılmış duygu sınıfı örnekleri çıkarılmıştır.



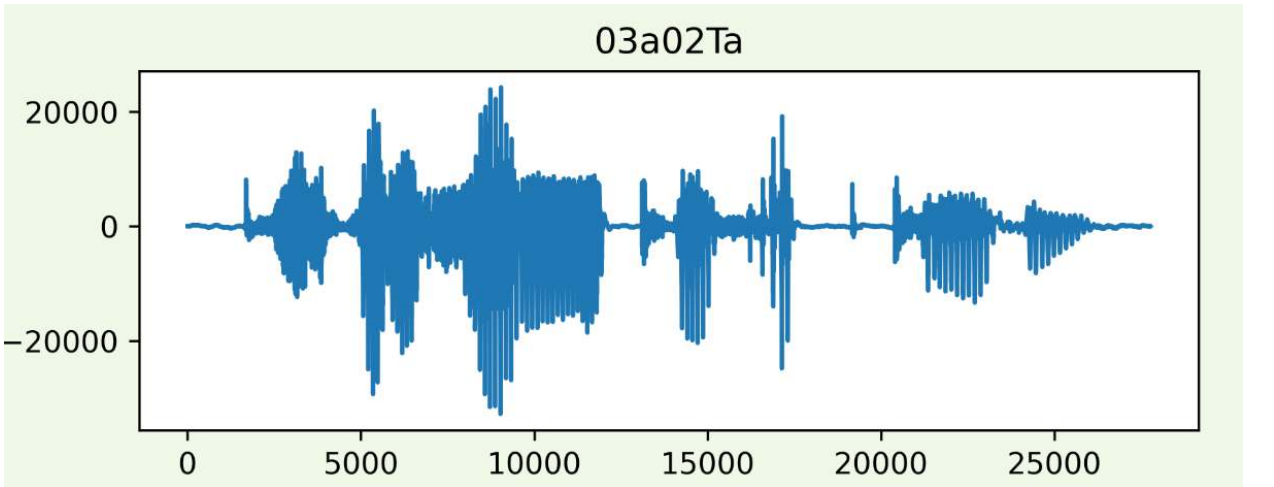
Şekil 3.2 EMO-DB Ses Verilerinin Kayıt Şekli (Burkhardt vd. 2005)



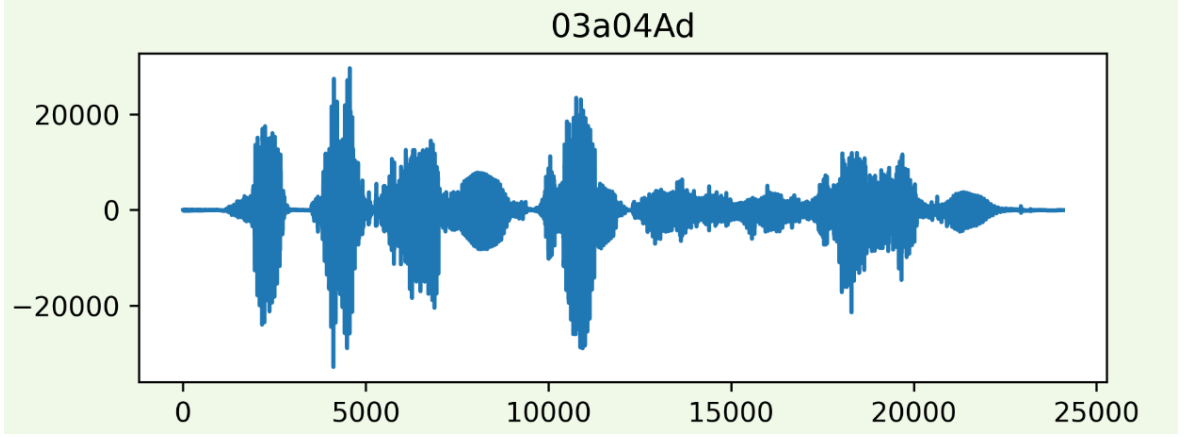
Şekil 3.3 EMO-DB Mutlu Duygu Sınıfının Dalga Formu



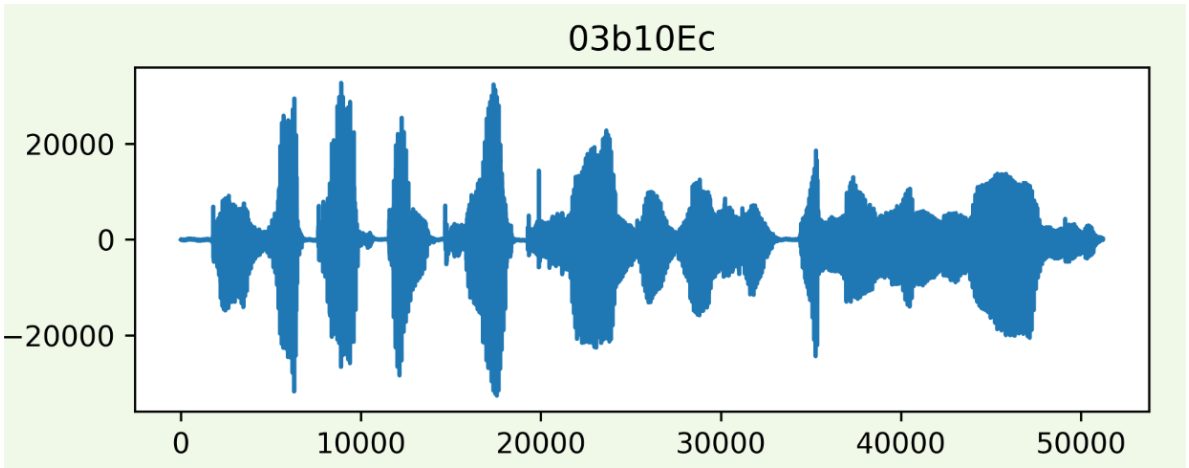
Şekil 3.4 EMO-DB Kızgın Duygu Sınıfının Dalga Formu



Şekil 3.5 EMO-DB Üzgün Duygu Sınıfının Dalga Formu



Şekil 3.6 EMO-DB Korkmuş Duygu Sınıfının Dalga Formu



Şekil 3.7 EMO-DB İğrenmiş Duygu Sınıfının Dalga Formu

Ses dosyalarının her birinin benzersiz bir dosya adı vardır. Bütün kısımların kendine ait farklı anlamları vardır. Her dosya adı toplamda 4 kısımdan (ör. 03a01Fa.wav) oluşur. Bu kısımları soldan sağa biçimde sırasıyla tanımlamak gerekirse;

1. Seslendirme Sanatçı(03'den 16'ya kadar)
2. Metin Kodu (a01, b10 gibi)
3. Duygu (F: Mutlu)
4. İkiden fazla versiyon varsa bunlar a,b,c olarak adlandırılır.

Örneğin; 03a01Fa.wav, Mutluluk duygusuyla Konuşmacı 03, konuşma metni a01'in ses dosyasıdır.

Seslendirme sanatçıları hakkında bilgi vermek gerekirse;

Çizelge 3.1 EMO-DB Veri Seti Seslendirme Sanatçı Bilgileri (Emodb Bilderbar,2024)

Sanatçı Kodu	Sanatçı Cinsiyeti	Sanatçı Yaşı
03	Erkek	31 yaşında
08	Kadın	34 yaşında
09	Kadın	21 yaşında
10	Erkek	32 yaşında
11	Erkek	26 yaşında
12	Erkek	30 yaşında
13	Kadın	32 yaşında
14	Kadın	35 yaşında
15	Erkek	25 yaşında
16	Kadın	31 yaşında

Çizelge 3.2 EMO-DB Veri Seti Duygu Sınıf İsimlendirmeleri

Harf	Duygu
W	Sinirli
F	Mutlu
A	Korkmuş
T	Üzgün
E	İğrenmiş
L	Sıkılmış
N	Nötr

Çizelge 3.3 EMO-DB Veri Seti Cümleleri

Kelime Kodu	Cümle
a01	Bez buzdolabının üzerinde duruyor.
a02	Çarşamba günü teslim edecek.
a04	Bu gece ona söyleyebilirim.
a05	Siyah kâğıt parçası orada, tahta parçasının yanında.
a07	Yedi saat içinde gerçekleşecek.
b01	Masanın altında duran çantalar ne?
b02	Az önce yukarı taşıdılar ve şimdi tekrar aşağı iniyorlar.
b03	Hafta sonları eve gittim ve Agnes 'i gördüm.
b09	Bunu alıp sonra Karl'la bir şeyler içmeye gitmek istiyorum.
b10	Her zaman koyduğumuz yerde olacak.

Çizelge 3.4 EMO-DB Veri Seti Duygu Sınıfları

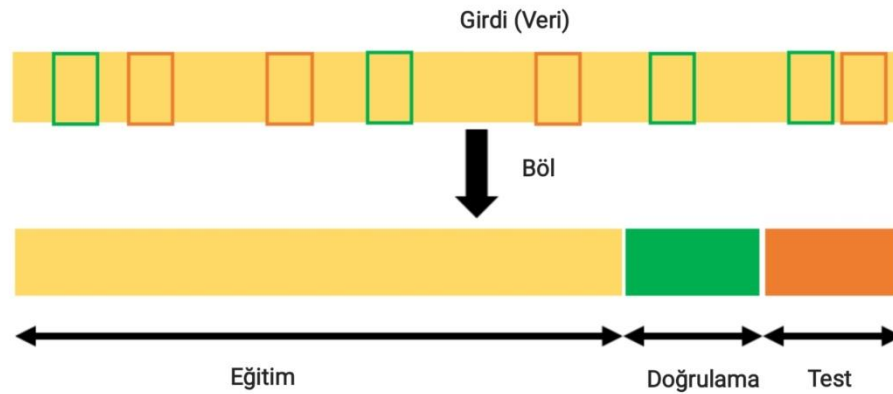
	Kızgın	Mutlu	Üzgün	Korkmuş	Nötr	Sıkılmış	İğrenmiş
Kadın	67	44	37	33	40	46	35
Erkek	60	27	25	36	39	35	11
Toplam=535	127	71	62	69	79	81	46



Şekil 3.8 EMO-DB Veri Seti Duygu Dağılımları

3.2. VERİLERİN EĞİTİM, DOĞRULAMA VE TEST OLARAK BÖLÜNMESİ

Ses duygu analizi için öncelikle EMO-DB veri setinin tüm ses dosyaları okunarak yüklenmiştir. Yüklenen veri seti, ayrılması istenen değerlere göre test, eğitim ve doğrulama olarak alt kümelere ayrılmıştır. Veri seti eğitim sırasında doğruluk değerlerinin artması amacıyla eğitim, doğrulama ve test veri setlerine ayrılması birden fazla oranda değişiklikler yapılarak model eğitimleri gerçekleştirilmiştir.



Şekil 3.9 EMO-DB Veri Setinin Eğitim, Doğrulama ve Test Alt Kümelerine Bölünmesi (Agrawal Samarth, 2021)

3.3. ÖZELLİK ÇIKARMA

Önceden eğitilmiş bir ses modeli olan Wav2Vec2 modeli ve diğer bileşenleri oluşturulmuştur. Önceden eğitilmiş modelin hiperparametrelerini ve özelliklerini içeren yapılandırma dosyaları yüklenmiştir.

Çizelge 3.5 Wav2Vec2 Hiperparametreleri ve Açıklamaları

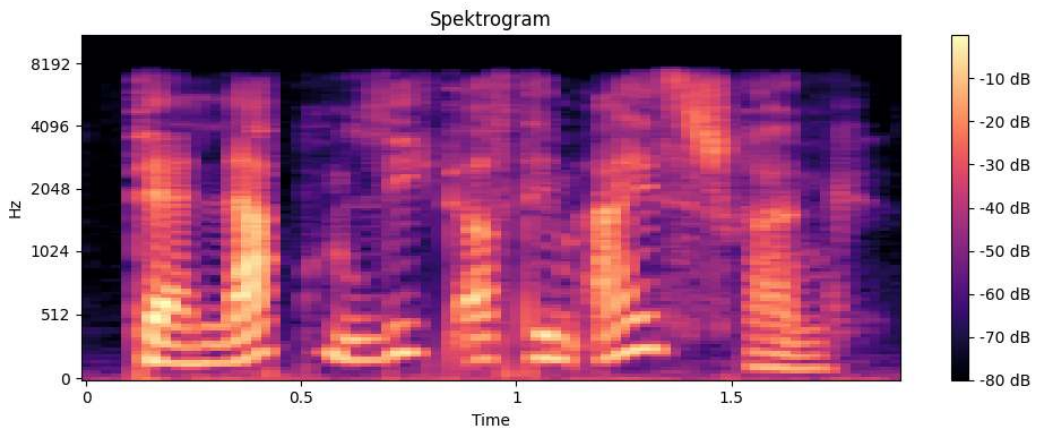
Wav2Vec2 Hiperparametreleri	Açıklaması
sampling_rate	Ses verisinin örnekleme hızıdır.
feature_extractor	Ses dosyalarından özellik çıkartma işlemini gerçekleştiren özelliktir.
num_hidden_layers	Modelin sahip olduğu gizli katman sayısıdır.
num_attention_heads	Gizli katmanlardaki dikkat başlıklarının sayısıdır.
num_labels	Modelin kullandığı etiket sayısıdır.
hidden_size	Gizli katmanlardaki nöron sayısıdır.
classifier_proj_size	Sınıflandırıcıya ait olan tam bağlı katmanın çıkış boyutudur.
intermediate_size	Gizli katmanlardaki ara katman boyutudur.
gradient_checkpointing	Eğitim sırasında bellek kullanımını azaltmak için gradient checkpointing 'in kullanılıp kullanılmayacağını belirten değerdir.
hidden_dropout	Gizli katmanlardaki devre dışı kalma olasılığıdır.
attention_dropout	Dikkat mekanizmasında kullanılan devre dışı kalma olasılığıdır.
layer_norm_eps	Katman normalleştirme sürecindeki epsilon değeridir.

Ses verilerinin önceden işlenmiş modelle uyumlu bir şekilde işlenmesini sağlayabilmek ve ses dalgalarını modelin anlayabileceği sayısal değerlere dönüştürmek için **Wav2Vec2FeatureExtractor** bileşeni ile önceden işlenmiş modelin ses özellik çıkarma modeli yüklenmiştir. Ses özellik çıkarma bileşeninin örnekleme hızı belirlenerek **Wav2Vec2ForSequenceClassification** sınıflandırma model bileşeniyle ile önceden eğitilmiş Wav2Vec2 modeli yüklenmiştir ve bileşenler hazırlanmıştır. Ses verileri işlenerek önceden eğitilmiş modelde sınıflandırma işlemi gerçekleştirilmiştir.

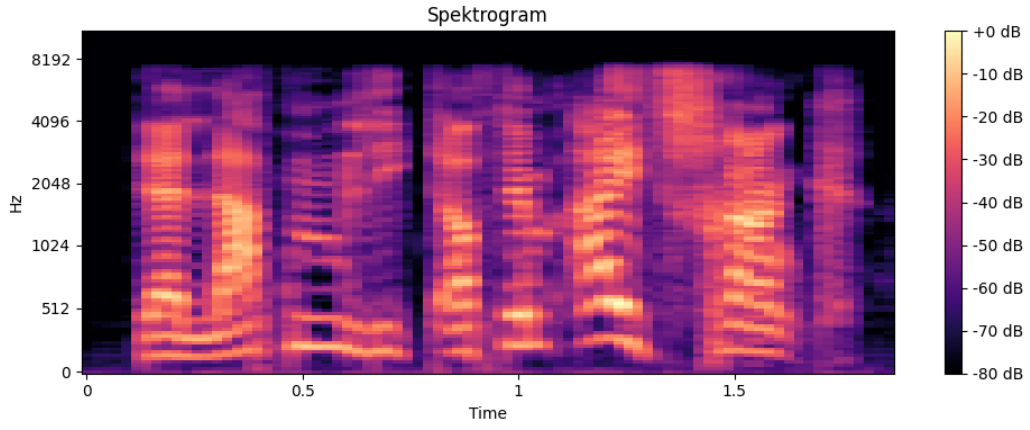
Eğitim, doğrulama ve test veri setleri modele uygun olacak şekilde tasarlanarak verilerin spektrogramlarını ve etiketlerini içerecek formata dönüştürülmüştür.

Spektrogram, bir ses sinyalinin zaman içindeki frekans bileşenlerini görsel bir şekilde temsil eden bir grafikdir. Genellikle ses sinyallerinin analizi ve ses sinyallerini anlamak için kullanılmaktadır. Spektrogram görüntüsünde, dikey eksenle frekanslar, yatay eksenle zamanın pozitif yönde artışı, renk yoğunluğu ise frekans bileşenlerinin genliğini göstermektedir. Yüksek genlikler genellikle kırmızı tonlarıyla temsil edilirken, düşük genlikler mavi tonlarına doğru koyu renklerle gösterilmektedir(Top, 2018).

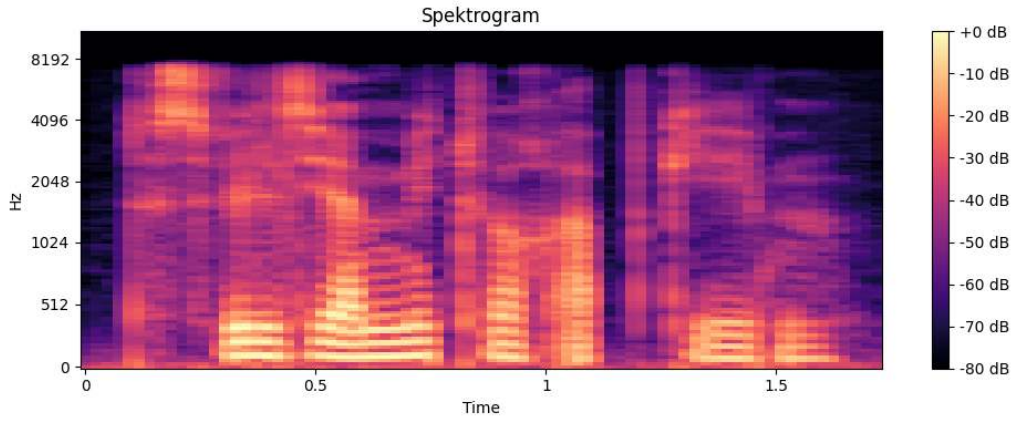
Etiket dediğimiz şey ise; ses dosyalarının hangi duygu sınıflarına ait olduğunu göstermektedir.



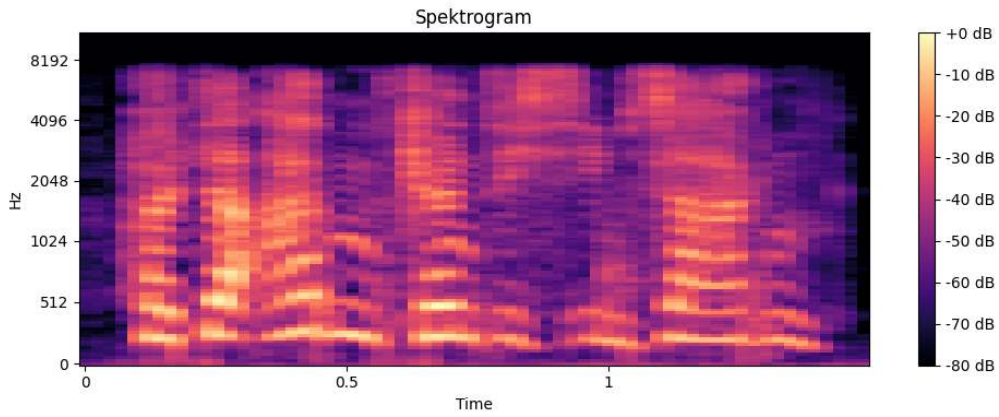
Şekil 3.10 EMO-DB Mutlu Duygu Sınıfının Spektrogramı



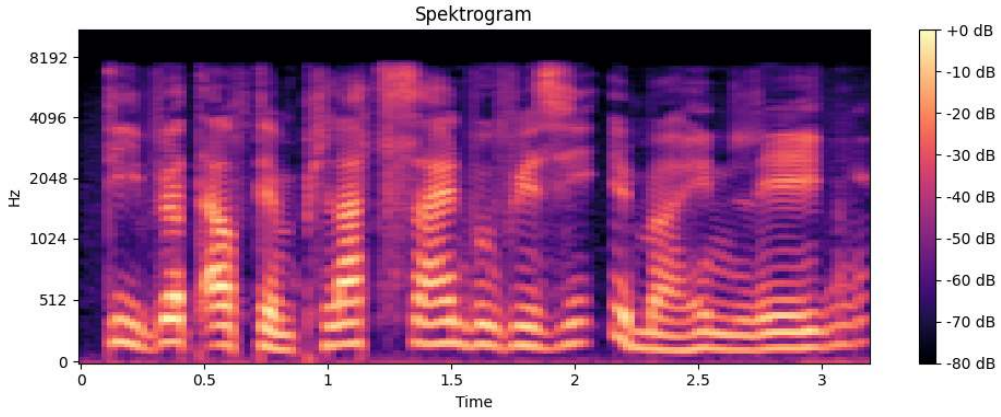
Şekil 3.11 EMO-DB Kızgın Duygu Sınıfının Spektrogramı



Şekil 3.12 EMO-DB Üzgün Duygu Sınıfının Spektrogramı



Şekil 3.13 EMO-DB Korkmuş Duygu Sınıfının Spektrogramı



Şekil 3.14 EMO-DB İğrenmiş Duygu Sınıfının Spektrogramı

Ayrılan veri setleri dönüştürülen bu format ile önceden işlenmiş model olan Wav2Vec2 modeline uygun hale getirilmiştir. Wav2Vec2 modeli ile önceden eğitilmiş bir dil modeline beslenerek, duygu durumunun sınıflandırılması gerçekleştirilmiştir.

İnce ayarlı önceden eğitilmiş Wav2Vec2 modeli için en iyi sonuçları veren değerler aşağıdaki çizelgede verilmiştir.

Çizelge 3.6 Wav2Vec2 Hiperparametreleri ve Değerleri

Wav2Vec2 Hiperparametreleri	Değerleri
sampling_rate	16000
feature_extractor	Wav2Vec2FeatureExtractor
num_hidden_layers	12
num_attention_heads	12
num_labels	5
hidden_size	768
classifier_proj_size	256
intermediate_size	3072
gradient_checkpointing	True
hidden_dropout	0.1
attention_dropout	0.1
layer_norm_eps	1e-05

3.4. ÖN İŞLEME

Eğitim gerçekleştirilmeden önce, modelin farklı boyutlardaki girişleri işleyebilmesini sağlamak amacıyla veri örneklerini modele uygun boyutlara getirebilmek için **DataCollatorWithPadding** kullanılmıştır. Maksimum uzunlukta olan ses dosyalarını alıp modele uygun hale getirebilmek için dolgu işlemi yapılmıştır.

3.5. MODEL EĞİTİMİ

Eğitim verilerini işleyebilmek için veri bağdaştırıcıları ve değerleri belirlenmiştir. Veri bağdaştırıcıları olarak kullanılan parametreler ve ne için kullanıldıkları aşağıdaki gibidir.

Çizelge 3.7 Veri Bağdaştırıcıları ve Açıklamaları

Veri Bağdaştırıcı Parametreleri	Açıklaması
output_dir	Eğitim sırasında oluşturulan modellerin, diğer çıktıların ve eğitim tamamlandığında en iyi modelin kaydedileceği dizindir.
num_train_epochs	Toplam eğitim epok sayısıdır. (Bir epok, modelin eğitim veri setini tamamen geçtiği süreyi ifade etmektedir.)
per_device_train_batch_size	Her bir eğitim cihazındaki (GPU veya TPU) veri yığını boyutudur.
per_device_eval_batch_size	Her bir değerlendirme cihazındaki veri yığını boyutudur. Eğitimde kullanılan veri yığını boyutu ile aynı ya da farklı olarak kullanılabilir.
save_steps	Modelin belirli aralıklarla kaydedilmesi için kullanılan adım sayısıdır.
save_total_limit	Toplamda kaç adet modelin kaydedileceğini belirleyen değerdir. Bu değer aşıldığında eski modeller silinmektedir.
evaluation_strategy	Değerlendirme stratejisidir. (Bu çalışmada steps olarak belirlenmiştir.)
eval_steps	Modelin performansını düzenli aralıklarla kontrol etmek için değerlendirme adım sayısıdır.
logging_steps	Günlüğe kaydedilecek bilgilerin adım sayısıdır.
learning_rate	Modelin ağırlıklarını güncellerken kullanılan öğrenme oranıdır/hızıdır.
load_best_model_at_end	Eğitim tamamlandığında en iyi modelin yüklenip yüklenmeyeceğini belirlemek için tutulur.

Eğitimi gerçekleştirecek olan model için modelin eğitim ve doğrulama veri setleri ve veri bağdaştırıcıları tanımlanarak model eğitilmiştir.

Eğitilen modelin veri bağdaştırıcı parametreleri ve eğitim sırasında verilen değerleri aşağıdaki çizelgede verilmiştir.

Çizelge 3.8 Veri Bağdaştırıcı Parametreleri ve Değerleri

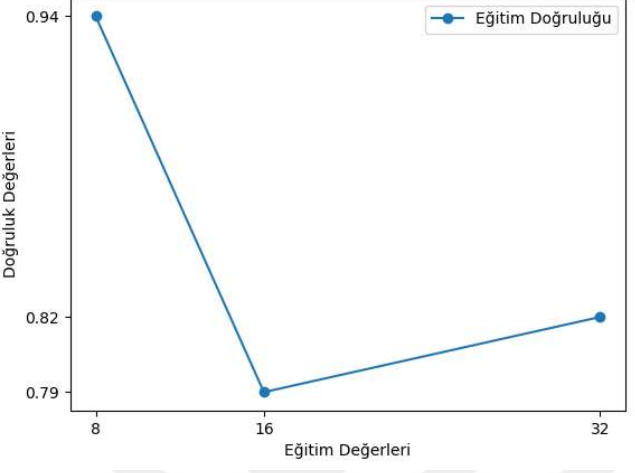
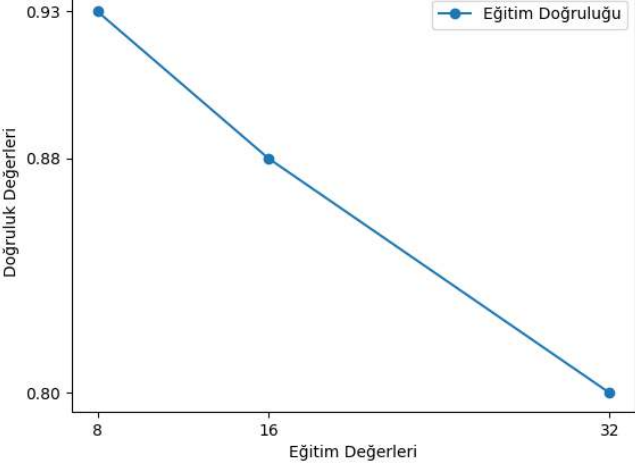
Veri Bağdaştırıcı Parametreleri	Değerleri
output_dir	./results
num_train_epochs	3
per_device_train_batch_size	8,16,32
per_device_eval_batch_size	8,16,32
save_steps	200
save_total_limit	2
evaluation_strategy	steps
eval_steps	100
logging_steps	100
learning_rate	1e-4
load_best_model_at_end	True

3.6. SINIFLANDIRMA

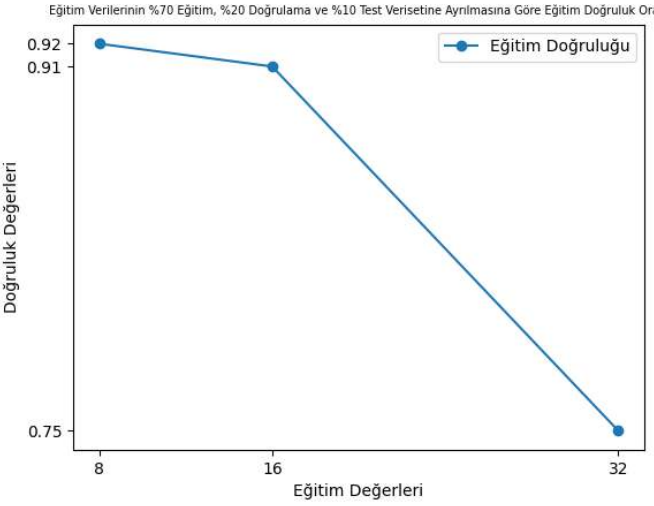
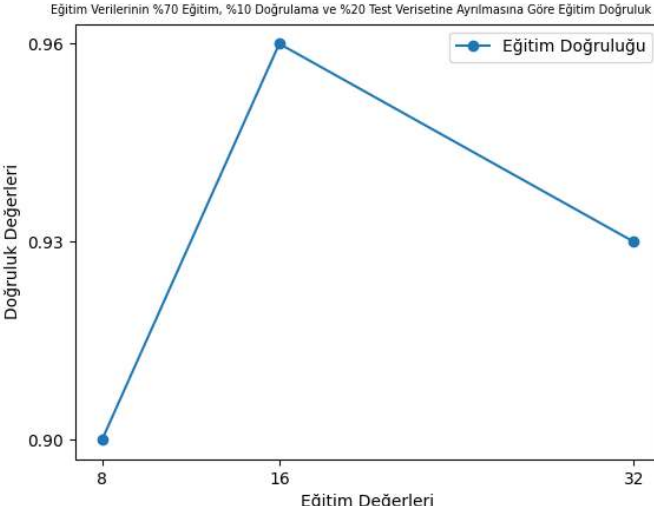
Eğitim gerçekleştirildikten sonra yüklenen eğitim ve doğrulama veri setlerine göre sınıflandırılan ses veri setinde kızgın, üzgün, mutlu, korkmuş ve iğrenmiş olarak duygu sınıflarına göre tahminlemeler yapılmıştır. Eğitilen model için en iyi model kaydedilmiştir.

Eğitim sonrası doğruluk oranı grafiği çıkarılmıştır. Çıkarılan doğruluk oranı grafikleri ve açıklamaları aşağıdaki çizelgede gösterilmiştir.

Çizelge 3.9 Eğitim Verilerinin Ayrılmasına Göre Eğitim Doğruluk Oranı Grafikleri ve Açıklamaları

Doğruluk Oranı Grafiği	Açıklaması								
Eğitim Verilerinin %80 Eğitim, %10 Doğrulama ve %10 Test Verisine Ayrılmasına Göre Eğitim Doğruluk Oranı  <table border="1"> <thead> <tr> <th>Eğitim Değerleri</th> <th>Eğitim Doğruluğu</th> </tr> </thead> <tbody> <tr> <td>8</td> <td>0.94</td> </tr> <tr> <td>16</td> <td>0.79</td> </tr> <tr> <td>32</td> <td>0.82</td> </tr> </tbody> </table>	Eğitim Değerleri	Eğitim Doğruluğu	8	0.94	16	0.79	32	0.82	Eğitim verileri %80 eğitim, %10 doğrulama ve %10 test veri setine ayrıldığında bu ayırım için veri yığını boyutu 16 olduğunda en düşük doğruluk oranını verdiği, veri yığını boyutu 8 iken en yüksek doğruluk oranını verdiği görülmüştür.
Eğitim Değerleri	Eğitim Doğruluğu								
8	0.94								
16	0.79								
32	0.82								
Eğitim Verilerinin %60 Eğitim, %20 Doğrulama ve %20 Test Verisine Ayrılmasına Göre Eğitim Doğruluk Oranı  <table border="1"> <thead> <tr> <th>Eğitim Değerleri</th> <th>Eğitim Doğruluğu</th> </tr> </thead> <tbody> <tr> <td>8</td> <td>0.93</td> </tr> <tr> <td>16</td> <td>0.88</td> </tr> <tr> <td>32</td> <td>0.80</td> </tr> </tbody> </table>	Eğitim Değerleri	Eğitim Doğruluğu	8	0.93	16	0.88	32	0.80	Eğitim verileri %60 eğitim, %20 doğrulama ve %20 test veri setine ayrıldığında bu ayırım için veri yığını boyutu 32 olduğunda en düşük doğruluk oranını verdiği, veri yığını boyutu 8 iken en yüksek doğruluk oranını verdiği görülmüştür.
Eğitim Değerleri	Eğitim Doğruluğu								
8	0.93								
16	0.88								
32	0.80								

Çizelge 3.10 (devamı). Eğitim Verilerinin Ayrılmasına Göre Eğitim Doğruluk Oranı Grafikleri ve Açıklamaları

 Eğitim Verilerinin %70 Eğitim, %20 Doğrulama ve %10 Test Verisetine Ayrılmasına Göre Eğitim Doğruluk Oranı	Eğitim verileri %70 eğitim, %20 doğrulama ve %10 test veri setine ayrıldığında bu ayırım için veri yığını boyutu 32 olduğunda en düşük doğruluk oranını verdiği, veri yığını boyutu 8 iken en yüksek doğruluk oranını verdiği görülmüştür.
 Eğitim Verilerinin %70 Eğitim, %10 Doğrulama ve %20 Test Verisetine Ayrılmasına Göre Eğitim Doğruluk Oranı	Eğitim verileri %70 eğitim, %10 doğrulama ve %20 test veri setine ayrıldığında bu ayırım için veri yığını boyutu 8 olduğunda en düşük doğruluk oranını verdiği, veri yığını boyutu 16 iken en yüksek doğruluk oranını verdiği görülmüştür.

Eğitim gerçekleştirildikten sonra EMO-DB veri setinin eğitim, doğrulama ve test veri setlerine ayrılma oranlarına ve veri yığını boyutlarına (8,16,32) göre değişen karmaşıklık matrisleri kaydedilmiştir.

Kaydedilen karmaşıklık matrisleri grafikleri ve açıklamaları aşağıdaki çizelgede gösterilmiştir.

Çizelge 3.11 Eğitim Verilerinin Ayrılmasına Göre Eğitim Karmaşıklık Matrisleri ve Açıklamaları

Karmaşıklık Matrisi	Açıklaması																																				
Eğitim Verilerinin %80 Eğitim, %10 Doğrulama ve %10 Test Verisetine Ayrılmasına Göre Karmaşıklık Matrisi <table><tr><th>Gerçek Etiketler \ Tahmin Edilen Etiketler</th><th>öğrenmiş</th><th>korkmuş</th><th>mutlu</th><th>sinirli</th><th>üzgün</th></tr><tr><th>öğrenmiş</th><td>5</td><td>1</td><td>1</td><td>0</td><td>0</td></tr><tr><th>korkmuş</th><td>0</td><td>2</td><td>0</td><td>0</td><td>0</td></tr><tr><th>mutlu</th><td>0</td><td>0</td><td>5</td><td>1</td><td>0</td></tr><tr><th>sinirli</th><td>1</td><td>0</td><td>0</td><td>10</td><td>0</td></tr><tr><th>üzgün</th><td>0</td><td>0</td><td>0</td><td>0</td><td>8</td></tr></table>	Gerçek Etiketler \ Tahmin Edilen Etiketler	öğrenmiş	korkmuş	mutlu	sinirli	üzgün	öğrenmiş	5	1	1	0	0	korkmuş	0	2	0	0	0	mutlu	0	0	5	1	0	sinirli	1	0	0	10	0	üzgün	0	0	0	0	8	Veri yığını boyutu 8 ve eğitim verileri %80 eğitim, %10 doğrulama ve %10 test veri setine ayrıldığında üzgün ve korkmuş duygu sınıflarındaki tahminlemelerinin tamamını doğru olarak yaptığı görülmüştür.
Gerçek Etiketler \ Tahmin Edilen Etiketler	öğrenmiş	korkmuş	mutlu	sinirli	üzgün																																
öğrenmiş	5	1	1	0	0																																
korkmuş	0	2	0	0	0																																
mutlu	0	0	5	1	0																																
sinirli	1	0	0	10	0																																
üzgün	0	0	0	0	8																																
Eğitim Verilerinin %80 Eğitim, %10 Doğrulama ve %10 Test Verisetine Ayrılmasına Göre Karmaşıklık Matrisi <table><tr><th>Gerçek Etiketler \ Tahmin Edilen Etiketler</th><th>öğrenmiş</th><th>korkmuş</th><th>mutlu</th><th>sinirli</th><th>üzgün</th></tr><tr><th>öğrenmiş</th><td>2</td><td>1</td><td>3</td><td>0</td><td>1</td></tr><tr><th>korkmuş</th><td>0</td><td>2</td><td>0</td><td>0</td><td>0</td></tr><tr><th>mutlu</th><td>0</td><td>0</td><td>6</td><td>0</td><td>0</td></tr><tr><th>sinirli</th><td>1</td><td>0</td><td>1</td><td>9</td><td>0</td></tr><tr><th>üzgün</th><td>0</td><td>0</td><td>0</td><td>0</td><td>8</td></tr></table>	Gerçek Etiketler \ Tahmin Edilen Etiketler	öğrenmiş	korkmuş	mutlu	sinirli	üzgün	öğrenmiş	2	1	3	0	1	korkmuş	0	2	0	0	0	mutlu	0	0	6	0	0	sinirli	1	0	1	9	0	üzgün	0	0	0	0	8	Veri yığını boyutu 16 ve eğitim verileri %80 eğitim, %10 doğrulama ve %10 test veri setine ayrıldığında üzgün, korkmuş ve mutlu duygu sınıflarındaki tahminlemelerinin tamamını doğru olarak yaptığı görülmüştür.
Gerçek Etiketler \ Tahmin Edilen Etiketler	öğrenmiş	korkmuş	mutlu	sinirli	üzgün																																
öğrenmiş	2	1	3	0	1																																
korkmuş	0	2	0	0	0																																
mutlu	0	0	6	0	0																																
sinirli	1	0	1	9	0																																
üzgün	0	0	0	0	8																																
Eğitim Verilerinin %80 Eğitim, %10 Doğrulama ve %10 Test Verisetine Ayrılmasına Göre Karmaşıklık Matrisi <table><tr><th>Gerçek Etiketler \ Tahmin Edilen Etiketler</th><th>öğrenmiş</th><th>korkmuş</th><th>mutlu</th><th>sinirli</th><th>üzgün</th></tr><tr><th>öğrenmiş</th><td>3</td><td>3</td><td>0</td><td>0</td><td>1</td></tr><tr><th>korkmuş</th><td>0</td><td>2</td><td>0</td><td>0</td><td>0</td></tr><tr><th>mutlu</th><td>0</td><td>0</td><td>6</td><td>0</td><td>0</td></tr><tr><th>sinirli</th><td>0</td><td>1</td><td>1</td><td>9</td><td>0</td></tr><tr><th>üzgün</th><td>0</td><td>0</td><td>0</td><td>0</td><td>8</td></tr></table>	Gerçek Etiketler \ Tahmin Edilen Etiketler	öğrenmiş	korkmuş	mutlu	sinirli	üzgün	öğrenmiş	3	3	0	0	1	korkmuş	0	2	0	0	0	mutlu	0	0	6	0	0	sinirli	0	1	1	9	0	üzgün	0	0	0	0	8	Veri yığını boyutu 32 ve eğitim verileri %80 eğitim, %10 doğrulama ve %10 test veri setine ayrıldığında üzgün, korkmuş ve mutlu duygu sınıflarındaki tahminlemelerinin tamamını doğru olarak yaptığı görülmüştür.
Gerçek Etiketler \ Tahmin Edilen Etiketler	öğrenmiş	korkmuş	mutlu	sinirli	üzgün																																
öğrenmiş	3	3	0	0	1																																
korkmuş	0	2	0	0	0																																
mutlu	0	0	6	0	0																																
sinirli	0	1	1	9	0																																
üzgün	0	0	0	0	8																																

Çizelge 3.12 (devamı). Eğitim Verilerinin Ayrılmasına Göre Eğitim Karmaşıklık Matrisleri ve Açıklamaları

Eğitim Verilerinin %60 Eğitim, %20 Doğrulama ve %20 Test Verisetine Ayrılmasına Göre Karmaşıklık Matrisi

Gerçek Etiketler \ Tahmin Edilen Etiketler	iğrenmiş	korkmuş	mutlu	sinirli	üzgün
iğrenmiş	6	0	0	0	0
korkmuş	0	9	1	0	0
mutlu	0	1	11	2	0
sinirli	0	0	0	19	0
üzgün	0	0	0	0	11

Veri yığını boyutu 8 ve eğitim verileri %60 eğitim, %20 doğrulama ve %20 test veri setine ayrıldığında üzgün, iğrenmiş ve sinirli duygu sınıflarındaki tahminlemelerinin tamamını doğru olarak yaptığı görülmüştür.

Eğitim Verilerinin %60 Eğitim, %20 Doğrulama ve %20 Test Verisetine Ayrılmasına Göre Karmaşıklık Matrisi

Gerçek Etiketler \ Tahmin Edilen Etiketler	iğrenmiş	korkmuş	mutlu	sinirli	üzgün
iğrenmiş	4	0	0	2	0
korkmuş	0	9	0	1	0
mutlu	0	0	10	4	0
sinirli	0	0	0	19	0
üzgün	0	0	0	0	11

Veri yığını boyutu 16 ve eğitim verileri %60 eğitim, %20 doğrulama ve %20 test veri setine ayrıldığında üzgün ve sinirli duygu sınıflarındaki tahminlemelerinin tamamını doğru olarak yaptığı görülmüştür.

Eğitim Verilerinin %60 Eğitim, %20 Doğrulama ve %20 Test Verisetine Ayrılmasına Göre Karmaşıklık Matrisi

Gerçek Etiketler \ Tahmin Edilen Etiketler	iğrenmiş	korkmuş	mutlu	sinirli	üzgün
iğrenmiş	2	1	1	1	1
korkmuş	0	8	2	0	0
mutlu	2	1	8	3	0
sinirli	0	0	0	19	0
üzgün	0	0	0	0	11

Veri yığını boyutu 32 ve eğitim verileri %60 eğitim, %20 doğrulama ve %20 test veri setine ayrıldığında üzgün ve sinirli duygu sınıflarındaki tahminlemelerinin tamamını doğru olarak yaptığı görülmüştür.

Çizelge 3.13 (devamı). Eğitim Verilerinin Ayrılmasına Göre Eğitim Karmaşıklık Matrisleri ve Açıklamaları

Eğitim Verilerinin %70 Eğitim, %20 Doğrulama ve %10 Test Verisetine Ayrılmasına Göre Karmaşıklık Matrisi Gerçek Etiketler Tahmin Edilen Etiketler	Veri yığını boyutu 8 ve eğitim verileri %70 eğitim, %20 doğrulama ve %10 test veri setine ayrıldığında üzgün ve korkmuş duygu sınıflarındaki tahminlemelerinin tamamını doğru olarak yaptığı görülmüştür.
Eğitim Verilerinin %70 Eğitim, %20 Doğrulama ve %10 Test Verisetine Ayrılmasına Göre Karmaşıklık Matrisi Gerçek Etiketler Tahmin Edilen Etiketler	Veri yığını boyutu 16 ve eğitim verileri %70 eğitim, %20 doğrulama ve %10 test veri setine ayrıldığında üzgün, korkmuş ve sinirli duygu sınıflarındaki tahminlemelerinin tamamını doğru olarak yaptığı görülmüştür.
Eğitim Verilerinin %70 Eğitim, %20 Doğrulama ve %10 Test Verisetine Ayrılmasına Göre Karmaşıklık Matrisi Gerçek Etiketler Tahmin Edilen Etiketler	Veri yığını boyutu 32 ve eğitim verileri %70 eğitim, %20 doğrulama ve %10 test veri setine ayrıldığında üzgün duygu sınıfındaki tahminlemelerinin tamamını doğru olarak yaptığı görülmüştür.

Çizelge 3.14 (devamı). Eğitim Verilerinin Ayrılmasına Göre Eğitim Karmaşıklık Matrisleri ve Açıklamaları

Eğitim Verilerinin %70 Eğitim, %10 Doğrulama ve %20 Test Verisetine Ayrılmasına Göre Karmaşıklık Matrisi

Gerçek Etiketler \ Tahmin Edilen Etiketler	öğrenmiş	korkmuş	mutlu	sinirli	üzgün
öğrenmiş	4	0	0	0	0
korkmuş	0	4	0	0	0
mutlu	0	3	2	0	0
sinirli	0	0	0	13	0
üzgün	0	0	0	0	4

Veri yığını boyutu 8 ve eğitim verileri %70 eğitim, %10 doğrulama ve %20 test veri setine ayrıldığında mutlu duygu sınıfı haricinde tüm duygu sınıflarındaki tahminlemelerinin tamamını doğru olarak yaptığı görülmüştür.

Eğitim Verilerinin %70 Eğitim, %10 Doğrulama ve %20 Test Verisetine Ayrılmasına Göre Karmaşıklık Matrisi

Gerçek Etiketler \ Tahmin Edilen Etiketler	öğrenmiş	korkmuş	mutlu	sinirli	üzgün
öğrenmiş	4	0	0	0	0
korkmuş	0	4	0	0	0
mutlu	0	0	4	1	0
sinirli	0	0	0	13	0
üzgün	0	0	0	0	4

Veri yığını boyutu 16 ve eğitim verileri %70 eğitim, %10 doğrulama ve %20 test veri setine ayrıldığında mutlu duygu sınıfı haricinde tüm duygu sınıflarındaki tahminlemelerinin tamamını doğru olarak yaptığı görülmüştür.

Eğitim Verilerinin %70 Eğitim, %10 Doğrulama ve %20 Test Verisetine Ayrılmasına Göre Karmaşıklık Matrisi

Gerçek Etiketler \ Tahmin Edilen Etiketler	öğrenmiş	korkmuş	mutlu	sinirli	üzgün
öğrenmiş	4	0	0	0	0
korkmuş	0	4	0	0	0
mutlu	0	1	4	0	0
sinirli	0	0	1	12	0
üzgün	0	0	0	0	4

Veri yığını boyutu 32 ve eğitim verileri %70 eğitim, %10 doğrulama ve %20 test veri setine ayrıldığında üzgün, korkmuş ve öğrenmiş duygu sınıflarındaki tahminlemelerinin tamamını doğru olarak yaptığı görülmüştür.

Yukarıdaki karmaşıklık matrisleri incelendiğinde; üzgün duygu sınıfında her zaman doğru tahminlemelerin yapıldığı, doğru olmayan tahminlemelerin en çok olduğu duygu sınıfının mutlu duygu sınıfı olduğu görülmüştür.

3.7. MODELİN PERFORMANS DEĞERLENDİRMESİ

Eğitilen modelin performansı ayrılmış olan test veri setinde değerlendirmeler yapılarak sonuçlandırılmıştır. Sınıflandırmaların eğitim, doğrulama ve test veri setlerinin oranları değiştirilerek değişen doğruluk değerleri kaydedilmiştir.



Çizelge 3.15 EMO-DB Veri Setinin Eğitim Doğrulama ve Test Veri Setlerine Ayrılma Oranlarına Göre Değişen Doğruluk Değerleri

Veri Yığını Boyutu	Eğitim Veri seti(%)	Doğrulama Veri seti(%)	Test Veri seti(%)	İterasyon Sayısı	Gerçekleşen Eğitim Süresi (dk)	Eğitim Doğruluk Değeri(%)	Test Doğruluk Değeri(%)
8	80	10	10	114	10.5	94	87
	60	20	20	90	9	93	85
	70	20	10	102	10	92	92
	70	10	20	102	10	90	91
16	80	10	10	57	9.5	79	89
	60	20	20	45	8	88	87
	70	20	10	51	9	91	87
	70	10	20	51	9	96	89
32	80	10	10	30	9.5	82	89
	60	20	20	24	7.5	80	84
	70	20	10	27	8	75	79
	70	10	20	27	8	93	91

Eğitilen modelin eğitim doğruluk değerinin en yüksek olduğu değere veri yığınının boyutunun 16 ve eğitim verileri %70 eğitim, %10 doğrulama ve %20 test veri setine ayrıldığında eğitim doğruluk değerinin %96 ya ulaştığı, aynı şekilde test doğruluk değerinin en yüksek olduğu değere veri yığınının boyutunun 8 ve eğitim verileri %70 eğitim, %20 doğrulama ve %10 test veri setine ayrıldığında test doğruluk değerinin %92 ye ulaştığı görülmüştür.

Model eğitimi 11 dakika kadar uzun sürede gerçekleşmesine rağmen bu modelin test veri setinde tahminlemelerini tamamlaması 10 saniye kadar kısa sürede gerçekleştiği elde edilmiştir.

Değerlendirilen modelin her bir veri yığını boyutuna ve veri setinin eğitim, doğrulama ve test veri setlerine ayrılma oranlarına göre sınıflandırma raporu çıkarılmıştır.

Sınıflandırma Raporu; sınıflandırma modelinin performansını değerlendirmek için kullanılmaktadır. Bu rapor, modelin her bir sınıf için ne kadar doğru tahmin yaptığını, hangi sınıfları doğru tahmin ettiğini, hangi sınıfları yanlış tahmin ettiğini ve bu sınıflar için hangi performans ölçütlerini elde ettiğini göstermektedir. Sınıflandırma Raporu; performans metriği olarak Kesinlik (Precision), Duyarlılık (Recall), F1 Skoru (F1-Score), Destek (Support) ve Doğruluk (Accuracy) puanlarını görüntülemektedir (Şahin, 2022).

Kesinlik (Precision): Pozitif olarak tahmin edilen örneklerin gerçekten pozitif olma oranını ifade eder.

$$Kesinlik = DP / (DP + YP)$$

Duyarlılık (Recall): Gerçek pozitif örneklerin model tarafından kaçının doğru bir şekilde tahmin edildiğini ifade eder.

$$Duyarlılık = DP / (DP + YN)$$

F1 Skoru: Kesinlik ve duyarlılığın harmonik ortalamasıdır. F1 skoru, kesinlik ve duyarlılık arasındaki dengeyi sağlar.

$$F1 = 2 * (Duyarlılık * Kesinlik) / (Duyarlılık + Kesinlik)$$

Destek (Support): Her bir sınıf için gerçekten var olan örnek sayısını ifade eder.

Doğruluk (Accuracy): Modelin doğru sınıflandırma oranını ifade eder.

$$Doğruluk = (DP + DN) / (DP + YP + DN + YN)$$

Sınıflandırma rapor ve değerlendirmeleri aşağıdaki çizelgede verilmiştir.

Çizelge 3.16 Verilerin Ayrılmasına Göre Sınıflandırma Raporları ve Değerlendirmeleri

Sınıflandırma Raporu	Açıklaması
<pre> precision recall f1-score support iğrenmiş 1.00 0.67 0.80 3 korkmuş 0.90 0.82 0.86 11 mutlu 0.67 0.75 0.71 8 sinirli 0.90 1.00 0.95 9 üzgün 1.00 1.00 1.00 7 accuracy 0.87 38 macro avg 0.89 0.85 0.86 38 weighted avg 0.88 0.87 0.87 38 </pre>	Test veri setinde yapılan sınıflandırmalarda veri yığını boyutunun 8 ve verilerin %80 eğitim, %10 doğrulama ve %10 test veri seti olarak ayrılması
Değerlendirmesi: Bu sınıflandırma raporuna göre; üzgün duygu sınıfında tahminlenenlerin ve tahminlenmesi gerekenlerin hepsinin doğru tahminlendiği incelenmiştir. Diğer duygu sınıfında tahminlenenlerin çoğunlukla doğru olduğu, iğrenmiş duygu sınıfında iğrenmiş olarak tahminlerin doğru olduğu fakat modelin gerçekten iğrenmiş duygu sınıfında yaptığı tahminlemelerinde eksiklikler olduğu görülmüştür. Mutlu duygu sınıfı tahminlemelerinde hatalar olabildiği ve doğruluk değerinin de %87'ye ulaştığı görülmüştür.	
<pre> precision recall f1-score support iğrenmiş 1.00 0.67 0.80 3 korkmuş 1.00 0.82 0.90 11 mutlu 0.67 1.00 0.80 8 sinirli 1.00 0.89 0.94 9 üzgün 1.00 1.00 1.00 7 accuracy 0.89 38 macro avg 0.93 0.87 0.89 38 weighted avg 0.93 0.89 0.90 38 </pre>	Test veri setinde yapılan sınıflandırmalarda veri yığını boyutunun 16 ve verilerin %80 eğitim, %10 doğrulama ve %10 test veri seti olarak ayrılması
Değerlendirmesi: Bu sınıflandırma raporuna göre; üzgün duygu sınıfında tahminlenenlerin ve tahminlenmesi gerekenlerin hepsinin doğru tahminlendiği incelenmiştir. Diğer duygu sınıfında tahminlenenlerin çoğunlukla doğru olduğu, iğrenmiş duygu sınıfında yapılan tahminlerin doğru olduğu fakat modelin gerçekten iğrenmiş duygu sınıfında yaptığı tahminlemelerinde eksiklikler olduğu görülmüştür. Mutlu duygu sınıfında tahminlenenlerde eksiklik olduğu fakat mutlu olarak tahminlenenlerin doğru olduğu ve doğruluk değerinin de %89' a ulaştığı görülmüştür.	

Çizelge 3.17 (devamı). Verilerin Ayrılmasına Göre Sınıflandırma Raporları ve Değerlendirmeleri

	precision	recall	f1-score	support	Test veri setinde yapılan sınıflandırmalarda veri yığını boyutunun 32 ve verilerin %80 eğitim, %10 doğrulama ve %10 test veri seti olarak ayrılması
iğrenmiş	1.00	0.33	0.50	3	
korkmuş	0.85	1.00	0.92	11	
mutlu	0.86	0.75	0.80	8	
sinirli	0.90	1.00	0.95	9	
üzgün	1.00	1.00	1.00	7	
accuracy			0.89	38	
macro avg	0.92	0.82	0.83	38	
weighted avg	0.90	0.89	0.88	38	

Değerlendirmesi: Bu sınıflandırma raporuna göre; üzgün duygu sınıfında tahminlenenlerin ve tahminlenmesi gerekenlerin hepsinin doğru tahminlendiği incelenmiştir. Diğer duygu sınıfında tahminlenenlerin çoğunlukla doğru olduğu, iğrenmiş duygu sınıfında iğrenmiş olarak tahminlerin doğru olduğu fakat modelin gerçekten iğrenmiş duygu sınıfında yaptığı tahminlemelerinde eksiklikler olduğu ve doğruluk değerinin de %89' a ulaştığı görülmüştür.

	precision	recall	f1-score	support	Test veri setinde yapılan sınıflandırmalarda veri yığını boyutunun 8 ve verilerin %60 eğitim, %20 doğrulama ve %20 test veri seti olarak ayrılması
iğrenmiş	1.00	0.78	0.88	9	
korkmuş	0.87	0.81	0.84	16	
mutlu	0.71	0.71	0.71	17	
sinirli	0.83	0.95	0.89	21	
üzgün	1.00	1.00	1.00	12	
accuracy			0.85	75	
macro avg	0.88	0.85	0.86	75	
weighted avg	0.86	0.85	0.85	75	

Değerlendirmesi: Bu sınıflandırma raporuna göre; üzgün duygu sınıfında tahminlenenlerin ve tahminlenmesi gerekenlerin hepsinin doğru tahminlendiği incelenmiştir. Diğer duygu sınıfında tahminlenenlerin çoğunlukla doğru olduğu ve doğruluk değerinin de %85' e ulaştığı görülmüştür.

Çizelge 3.18 (devamı). Verilerin Ayrılmasına Göre Sınıflandırma Raporları ve Değerlendirmeleri

	precision	recall	f1-score	support	Test veri setinde yapılan sınıflandırmalarda veri yığını boyutunun 16 ve verilerin %60 eğitim, %20 doğrulama ve %20 test veri seti olarak ayrılması
iğrenmiş	1.00	0.56	0.71	9	
korkmuş	0.88	0.94	0.91	16	
mutlu	0.86	0.71	0.77	17	
sinirli	0.78	1.00	0.88	21	
üzgün	1.00	1.00	1.00	12	
accuracy			0.87	75	
macro avg	0.90	0.84	0.85	75	
weighted avg	0.88	0.87	0.86	75	

Değerlendirmesi: Bu sınıflandırma raporuna göre; üzgün duygu sınıfında tahminlenenlerin ve tahminlenmesi gerekenlerin hepsinin doğru tahminlendiği incelenmiştir. Diğer duygu sınıfında tahminlenenlerin çoğunlukla doğru olduğu, iğrenmiş duygu sınıfında iğrenmiş olarak tahminlerin doğru olduğu fakat modelin gerçekten iğrenmiş duygu sınıfında yaptığı tahminlemelerinde eksiklikler olduğu ve doğruluk değerinin de %87' ye ulaştığı görülmüştür.

	precision	recall	f1-score	support	Test veri setinde yapılan sınıflandırmalarda veri yığını boyutunun 32 ve verilerin %60 eğitim, %20 doğrulama ve %20 test veri seti olarak ayrılması
iğrenmiş	1.00	0.78	0.88	9	
korkmuş	0.75	0.94	0.83	16	
mutlu	0.82	0.53	0.64	17	
sinirli	0.80	0.95	0.87	21	
üzgün	1.00	1.00	1.00	12	
accuracy			0.84	75	
macro avg	0.87	0.84	0.84	75	
weighted avg	0.85	0.84	0.83	75	

Değerlendirmesi: Bu sınıflandırma raporuna göre; üzgün duygu sınıfında tahminlenenlerin ve tahminlenmesi gerekenlerin hepsinin doğru tahminlendiği incelenmiştir. Diğer duygu sınıfında tahminlenenlerin çoğunlukla doğru olduğu ve doğruluk değerinin de %84' e ulaştığı görülmüştür.

Çizelge 3.19 (devamı). Verilerin Ayrılmasına Göre Sınıflandırma Raporları ve Değerlendirmeleri

	precision	recall	f1-score	support	Test veri setinde yapılan sınıflandırmalarda veri yığını boyutunun 8 ve verilerin %70 eğitim, %20 doğrulama ve %10 test veri seti olarak ayrılması
iğrenmiş	1.00	0.67	0.80	3	
korkmuş	0.91	0.91	0.91	11	
mutlu	0.78	0.88	0.82	8	
sinirli	1.00	1.00	1.00	9	
üzgün	1.00	1.00	1.00	7	
accuracy			0.92	38	
macro avg	0.94	0.89	0.91	38	
weighted avg	0.93	0.92	0.92	38	

Değerlendirmesi: Bu sınıflandırma raporuna göre; üzgün duygu sınıfında tahminlenenlerin ve tahminlenmesi gerekenlerin hepsinin doğru tahminlendiği incelenmiştir. Diğer duygu sınıfında tahminlenenlerin çoğunlukla doğru olduğu, iğrenmiş duygu sınıfında iğrenmiş olarak tahminlerin doğru olduğu fakat modelin gerçekten iğrenmiş duygu sınıfında yaptığı tahminlemelerinde eksiklikler olduğu ve doğruluk değerinin de %92' ye ulaştığı görülmüştür.

	precision	recall	f1-score	support	Test veri setinde yapılan sınıflandırmalarda veri yığını boyutunun 16 ve verilerin %70 eğitim, %20 doğrulama ve %10 test veri seti olarak ayrılması
iğrenmiş	1.00	1.00	1.00	3	
korkmuş	0.77	0.91	0.83	11	
mutlu	0.80	0.50	0.62	8	
sinirli	0.90	1.00	0.95	9	
üzgün	1.00	1.00	1.00	7	
accuracy			0.87	38	
macro avg	0.89	0.88	0.88	38	
weighted avg	0.87	0.87	0.86	38	

Değerlendirmesi: Bu sınıflandırma raporuna göre; üzgün duygu sınıfında tahminlenenlerin ve tahminlenmesi gerekenlerin hepsinin doğru tahminlendiği incelenmiştir. Diğer duygu sınıfında tahminlenenlerin çoğunlukla doğru olduğu, mutlu duygu sınıfında mutlu olarak tahminlerin çoğunun doğru olduğu fakat modelin gerçekten mutlu duygu sınıfında yaptığı tahminlemelerinde eksiklikler olduğu ve doğruluk değerinin de %87' ye ulaştığı görülmüştür.

Çizelge 3.20 (devamı). Verilerin Ayrılmasına Göre Sınıflandırma Raporları ve Değerlendirmeleri

	precision	recall	f1-score	support
igrenmiş	0.00	0.00	0.00	3
korkmuş	0.61	1.00	0.76	11
mutlu	1.00	0.38	0.55	8
sinirli	1.00	1.00	1.00	9
üzgün	0.88	1.00	0.93	7
accuracy			0.79	38
macro avg	0.70	0.68	0.65	38
weighted avg	0.79	0.79	0.74	38

Test veri setinde yapılan sınıflandırmalarda veri yığını boyutunun 32 ve verilerin %70 eğitim, %20 doğrulama ve %10 test veri seti olarak ayrılması

Değerlendirmesi: Bu sınıflandırma raporuna göre; sinirli duygu sınıfında tahminlenenlerin ve tahminlenmesi gerekenlerin hepsinin doğru tahminlendiği incelenmiştir. Diğer duygu sınıfında tahminlenenlerin çoğunlukla doğru olduğu, mutlu duygu sınıfında mutlu olarak tahminlerin doğru olduğu fakat modelin gerçekten mutlu duygu sınıfında yaptığı tahminlemelerinde eksiklikler olduğu, igrenmiş duygu sınıfında başarısız olduğu ve doğruluk değerinin de %79' a ulaştığı görülmüştür.

	precision	recall	f1-score	support
igrenmiş	1.00	0.78	0.88	9
korkmuş	0.76	1.00	0.86	16
mutlu	0.93	0.76	0.84	17
sinirli	0.95	0.95	0.95	21
üzgün	1.00	1.00	1.00	12
accuracy			0.91	75
macro avg	0.93	0.90	0.91	75
weighted avg	0.92	0.91	0.91	75

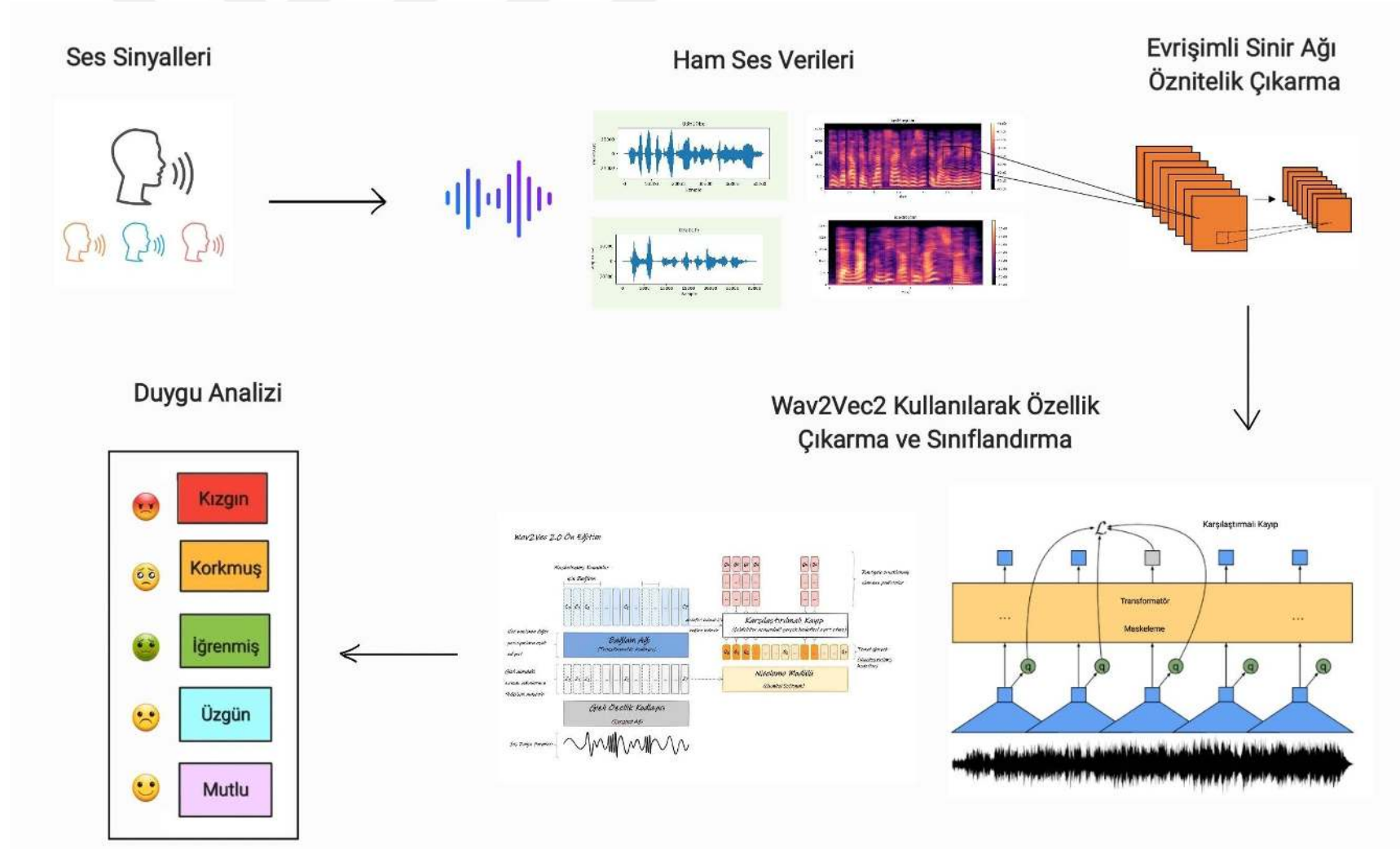
Test veri setinde yapılan sınıflandırmalarda veri yığını boyutunun 8 ve verilerin %70 eğitim, %10 doğrulama ve %20 test veri seti olarak ayrılması

Değerlendirmesi: Bu sınıflandırma raporuna göre; üzgün duygu sınıfında tahminlenenlerin ve tahminlenmesi gerekenlerin hepsinin doğru tahminlendiği incelenmiştir. Diğer duygu sınıfında tahminlenenlerin çoğunlukla doğru olduğu ve doğruluk değerinin de %91' e ulaştığı görülmüştür.

Çizelge 3.21 (devamı). Verilerin Ayrılmasına Göre Sınıflandırma Raporları ve Değerlendirmeleri

<table><tr><td></td><td>precision</td><td>recall</td><td>f1-score</td><td>support</td></tr><tr><td>ığrenmiş</td><td>1.00</td><td>1.00</td><td>1.00</td><td>9</td></tr><tr><td>korkmuş</td><td>0.89</td><td>1.00</td><td>0.94</td><td>16</td></tr><tr><td>mutlu</td><td>0.91</td><td>0.59</td><td>0.71</td><td>17</td></tr><tr><td>sinirli</td><td>0.80</td><td>0.95</td><td>0.87</td><td>21</td></tr><tr><td>üzgün</td><td>1.00</td><td>1.00</td><td>1.00</td><td>12</td></tr><tr><td>accuracy</td><td></td><td></td><td>0.89</td><td>75</td></tr><tr><td>macro avg</td><td>0.92</td><td>0.91</td><td>0.91</td><td>75</td></tr><tr><td>weighted avg</td><td>0.90</td><td>0.89</td><td>0.89</td><td>75</td></tr></table>		precision	recall	f1-score	support	ığrenmiş	1.00	1.00	1.00	9	korkmuş	0.89	1.00	0.94	16	mutlu	0.91	0.59	0.71	17	sinirli	0.80	0.95	0.87	21	üzgün	1.00	1.00	1.00	12	accuracy			0.89	75	macro avg	0.92	0.91	0.91	75	weighted avg	0.90	0.89	0.89	75	Test veri setinde yapılan sınıflandırmalarda veri yığını boyutunun 16 ve verilerin %70 eğitim, %10 doğrulama ve %20 test veri seti olarak ayrılması
	precision	recall	f1-score	support																																										
ığrenmiş	1.00	1.00	1.00	9																																										
korkmuş	0.89	1.00	0.94	16																																										
mutlu	0.91	0.59	0.71	17																																										
sinirli	0.80	0.95	0.87	21																																										
üzgün	1.00	1.00	1.00	12																																										
accuracy			0.89	75																																										
macro avg	0.92	0.91	0.91	75																																										
weighted avg	0.90	0.89	0.89	75																																										
Değerlendirmesi: Bu sınıflandırma raporuna göre; üzgün ve ığrenmiş duygu sınıfında tahminlenenlerin ve tahminlenmesi gerekenlerin hepsinin doğru tahminlendiği incelenmiştir. Diğer duygu sınıfında tahminlenenlerin çoğunlukla doğru olduğu ve doğruluk değerinin de %89' a ulaştığı görülmüştür.																																														
<table><tr><td></td><td>precision</td><td>recall</td><td>f1-score</td><td>support</td></tr><tr><td>ığrenmiş</td><td>0.89</td><td>0.89</td><td>0.89</td><td>9</td></tr><tr><td>korkmuş</td><td>0.80</td><td>1.00</td><td>0.89</td><td>16</td></tr><tr><td>mutlu</td><td>0.92</td><td>0.71</td><td>0.80</td><td>17</td></tr><tr><td>sinirli</td><td>0.95</td><td>0.95</td><td>0.95</td><td>21</td></tr><tr><td>üzgün</td><td>1.00</td><td>1.00</td><td>1.00</td><td>12</td></tr><tr><td>accuracy</td><td></td><td></td><td>0.91</td><td>75</td></tr><tr><td>macro avg</td><td>0.91</td><td>0.91</td><td>0.91</td><td>75</td></tr><tr><td>weighted avg</td><td>0.91</td><td>0.91</td><td>0.90</td><td>75</td></tr></table>		precision	recall	f1-score	support	ığrenmiş	0.89	0.89	0.89	9	korkmuş	0.80	1.00	0.89	16	mutlu	0.92	0.71	0.80	17	sinirli	0.95	0.95	0.95	21	üzgün	1.00	1.00	1.00	12	accuracy			0.91	75	macro avg	0.91	0.91	0.91	75	weighted avg	0.91	0.91	0.90	75	Test veri setinde yapılan sınıflandırmalarda veri yığını boyutunun 32 ve verilerin %70 eğitim, %10 doğrulama ve %20 test veri seti olarak ayrılması
	precision	recall	f1-score	support																																										
ığrenmiş	0.89	0.89	0.89	9																																										
korkmuş	0.80	1.00	0.89	16																																										
mutlu	0.92	0.71	0.80	17																																										
sinirli	0.95	0.95	0.95	21																																										
üzgün	1.00	1.00	1.00	12																																										
accuracy			0.91	75																																										
macro avg	0.91	0.91	0.91	75																																										
weighted avg	0.91	0.91	0.90	75																																										
Değerlendirmesi: Bu sınıflandırma raporuna göre; üzgün duygu sınıfında tahminlenenlerin ve tahminlenmesi gerekenlerin hepsinin doğru tahminlendiği incelenmiştir. Diğer duygu sınıfında tahminlenenlerin çoğunlukla doğru olduğu ve doğruluk değerinin de %91' e ulaştığı görülmüştür.																																														

Çıkarılan sınıflandırma raporuna göre modelin tahminlemelerde en başarılı olduğu duygu sınıfının üzgün duygu sınıfı olduğu, ığrenmiş ve mutlu duygu sınıflarının tahminlemelerinde eksiklikler olduğu, diğer duygu sınıflarında çoğunlukla doğru tahminlemelerde bulunduğu görülmüştür.



Şekil 3.15 Önerilen Modelin Çalışma Metodolojisi

4. SONUÇ

Bu tez çalışması kapsamında daha önce bir bas konuş sistemlerinde ses duygu analizi yapılmadığı için gerçekleştirilen projeyle makinistlerin/kumanda merkezlerinin yolcularla acil durumlarda daha hızlı bir şekilde iletişime geçilerek acil müdahale yapılması hedeflenmektedir.

Tez çalışması kapsamında, EMO-DB veri seti üzerinde uygulanan çalışma ele alınmaktadır. Bu veri setinde bulunan duygu sınıflarının doğru şekilde sınıflandırılabilmesi için ve yüksek doğruluk değerlerine ulaşılması hedeflenmiştir. Tez çalışmasında, derin öğrenme, transfer öğrenme ve transformatör yöntemleri kullanılarak ses duygu analizi alanındaki çalışmalara yeni bir referans oluşturması planlanmıştır.

Yapılan çalışmada EMO-DB veri setindeki ses dosyaları giriş verisi olarak alınarak ses dosyaları eğitim, doğrulama ve test veri setleri olacak şekilde alt kümelerle bölünmüştür. Eğitim, doğrulama ve test alt kümelerine ayrılan veri seti ses spektrogramları ve duygu etiketleriyle özellik çıkarma işlemi yapılmıştır. Bu derin öğrenme yöntemlerinden geçirilerek model, Wav2Vec2 önceden eğitilmiş modeline uygun hale getirilmiştir.

Bu çalışmada önerilen model ile transfer öğrenme yoluyla transformatörler kullanılmıştır. Önceden eğitilmiş bir Wav2Vec2 modeli ile EMO-DB veri seti üzerinde denenmiş %92 doğruluk değerlerine ulaşılmıştır. Modelin en başarılı olduğu duygu sınıfının üzgün duygu sınıfının olduğu, öğrenmiş ve mutlu duygu sınıflarının tahminlemelerinde eksiklikler olduğu geliştirilebileceği görülmüştür. Ulaşılan bu değerler ile ses duygu analizi görevinde Wav2Vec2 modelinin etkili olduğu gösterilmiştir.

Transfer öğrenme ve transformatörler kullanılarak, bu çalışmayla birlikte derin öğrenme ile birlikte kullanıldıklarında ulaşılan yüksek doğruluk oranları görülerek bundan sonraki yapılacak çalışmalara kaynak oluşturmaktadır. Elde edilen doğruluk değerleri ses işleme, konuşma tanıma, duygu analizi gibi alanlarda etkili ve önemli bir referans sağlayacağı görülmektedir.

Eğitim sonucunda Wav2Vec2 modelinin ses duygu analizi ve derin öğrenme modellerinin kent içi raylı sistemlerdeki iletişimi ve ulaşılabilirliği iyileştirmede önemli bir rol oynayabileceği görülmüştür.

Bu tez çalışmasıyla yola çıkılma nedeni olan metrolarda bulunan bas konuş sistemlerinde makinistlerin/kumanda merkezlerinin yolcularla acil durumlarda daha hızlı bir şekilde iletişime geçebilecek acil duruma hızlı bir şekilde dönüş yapılması sağlanacaktır. Sahada uygulanabilir bir çalışma olması nedeniyle sahada uygulamaya koyulduğu takdirde ulaşım hizmetlerinde acil durum müdahaleleri optimize edilerek yolcuya en iyi hizmetin verilmesi sağlanacaktır. En yüksek önceliğin verilmesi gereken duygu sınıfı korkmuş duygu sınıfı olup sıranın kızgın, iğrenmiş, üzgün ve mutlu duygu sınıflarının önceliklendirmesi olarak devam etmesinin en optimize çözüm olacağı düşünülmektedir. Yapılan bu önceliklendime ile duygu sınıflarına göre yolcuların sorunlarının acil durumlarda daha hızlı çözümlere ulaşılması sağlanacaktır.

Bu bulgular ile kent içi raylı sistemlerdeki bas konuş/interkom sistemlerinde acil durumların daha etkili bir şekilde yönetilebileceği gösterilmiştir. Bu çalışma, gelecekte kent içi raylı sistemlerde ses duygu analizi ve iletişim optimizasyonu üzerine daha kapsamlı çalışmaların yapılabilmesi için bir temel oluşturmaktadır.

5. KAYNAKLAR

- Abbaschian, B.J.,Sierra-Sosa D.and Elmaghraby A.(2021). Deep Learning Techniques for Speech Emotion Recognition,from Databases to Models. Computer Science and Engineering Department, University of Louisville, Louisville, KY 40292, USA. *Sensors* 2021. 21(4), 1249.
- Agrawal Samarth. (2021). How to split data into three sets (train, validation, and test) And why? <https://towardsdatascience.com/how-to-split-data-into-three-sets-train-validation-and-test-and-why-e50d22d3e54c>
- Akgözlüoğlu, K. (2021). ‘Transfer Öğrenmesi Tekniği Tabanlı Derin Öğrenme Yöntemiyle Ürün Tanıma’, Yüksek Lisans Tezi, Manisa Celal Bayar Üniversitesi Fen Bilimleri Enstitüsü, Manisa, Türkiye.
- Anithadevi, N.,Gokul, P., Muhil Nandan, S., Magesh, R., Shiddharth, S. (2020). Automated Speech Recognition System For Speaker Emotion Classification. *2020 5th International Conference on Computing, Communication and Security (ICCCS)*.
- Baevski, A., Zhou, H., Mohamed, A., Auli, M. (2020). wav2vec 2.0: A Framework for Self-Supervised Learning of Speech Representations, *34th Conference on Neural Information Processing Systems (NeurIPS 2020)*, Vancouver, Canada.
- Baevski, A., Hsu, W.N., Conneau, A., Auli, M. (2022). Unsupervised Speech Recognition, *35th Conference on Neural Information Processing Systems (NeurIPS 2021)*, San Francisco, California.
- Bardou D., Zhang K., and Ahmad S. M. (2018). Lung sounds classification using convolutional neural networks. *Artificial intelligence in medicine*, vol. 88, 58–69.
- Bayraktar, Ü. (2023). ‘Makine Öğrenimi Tekniklerini Kullanarak Konuşma Duygularını Tanıma’, Yüksek Lisans Tezi, Kocaeli Üniversitesi Fen Bilimleri Enstitüsü, Kocaeli, Türkiye.
- Bgn Jonathan. (2021). An Illustrated Tour of Wav2vec 2.0. <https://jonathanbgn.com/2021/09/30/illustrated-wav2vec-2.html>

- Burkhardt, F., Paeschke, A., Rolfes, M., Sendlmeier, W. F., & Weiss, B. (2005). A database of German emotional speech. *Interspeech 2005*, 1517-1520.
- Canpolat, S., Ormanoglu, Z., Zeyrek, D. (2020). Turkish Emotion-Voice Database (TurEV-DB). *Language Resources and Evaluation Conference (LREC 2020)*.(ss. 368-375).
- Chen, L.W., Rudnicky, A. (2023). Exploring Wav2vec 2.0 Fine Tuning For Improved Speech Emotion Recognition. *ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Rhodes Island, Greece.
- Chen, M., Radford, A., Child, R., Wu, J., Jun, H., Dhariwal, P., Luan, D., Sutskever, I. (2020). Generative Pretraining from Pixels. *ICML'20: Proceedings of the 37th International Conference on Machine Learning*. (ss. 1691–1703).
- Devlin, J., Chang, M.W., Lee, K., Toutanova, K.(2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, *In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, (ss 4171–4186), Minneapolis, Minnesota.
- Dwivedi Rohit. (2020). How is Transfer Learning done in Neural Networks and Convolutional Neural Networks? <https://www.analyticssteps.com/blogs/how-transfer-learning-done-neural-networks-and-convolutional-neural-networks>
- Edureka Web Sayfası. (2023). AI vs Machine Learning vs Deep Learning.<https://www.edureka.co/blog/ai-vs-machine-learning-vs-deep-learning/>
- Emodb Bilderbar Web Sayfası. Berlin Database of Emotional Speech. Geliş tarihi 07 Ocak 2024, gönderen <http://emodb.bilderbar.info/docu/>.
- Feldman, C. ve Anderson, R. (2012). Antibiotic resistance of pathogens causing community-acquired pneumonia. *Seminars in Respiratory & Critical Care Medicine*, 33(3),(ss. 232–243).
- Foto, Ö. (2022). ‘Transfer Öğrenme Modelleri İle Akciğer X-Ray Görüntülerinden Covid-19 Tespiti’, Yüksek Lisans Tezi, Aksaray Üniversitesi Sosyal Bilimler Enstitüsü, Aksaray, Türkiye.
- Goodfellow, I., Bengio, Y. ve Courville, A. (2017). Deep Learning. *MIT Press*, 173-175.

- Gong, Y., Chung, Y.A., Glass J. (2021). AST: Audio Spectrogram Transformer. *MIT Computer Science and Artificial Intelligence Laboratory*, Cambridge, MA 02139, USA.
- Gök, M. (2021). ‘Drama Liderlerinde, İşitme Engelli Bireylerde, Normal İşiten Bireylerde Ve Otizm Spektrum Bozukluğunda Derin Öğrenme İle Ses Duygu Analizi’, Yüksek Lisans Tezi, İstanbul Aydın Üniversitesi Lisansüstü Eğitim Enstitüsü, İstanbul, Türkiye.
- Gris, L.R.S., Casanova, E., Oliveira, F.S., Soares, A.S., Junior, A.C. (2021). Brazilian Portuguese Speech Recognition Using Wav2vec 2.0, *Computational Processing of the Portuguese Language*, (ss. 333-343).
- Guler H. C., Yildiz O., Baysal U., Cinyol F. B., Koksall D., Babaoglu E., and Ulasli S. S. (2020). Classification of abnormal respiratory sounds using machine learning techniques. *Medical Technologies Congress (TIPTEKNO)*, Antalya, Turkey, 1–4.
- Hacıbeyoğlu, M. (2006). ‘Çoklu Etmen Mimarisi ve Takviyeli Öğrenme’, Yüksek Lisans Tezi, Selçuk Üniversitesi Fen Bilimleri Enstitüsü, Konya, Türkiye.
- Harar, P., Burget, R., & Dutta, M. K. (2017). Speech emotion recognition with deep learning. In *2017 4th International Conference on Signal Processing and Integrated Networks (SPIN)* (ss. 137-140). IEEE.
- Hsu, W.N., Sriram, A., Baevski, A., Likhomanenko, T., Xu, Q., Pratap, V., Kahn, J., Lee, A., Collobert, R., Synnaeve, G., Auli, M. (2021). Robust Wav2vec 2.0: Analyzing Domain Shift In Self-Supervised Pre-Training, *Interspeech 2021*, 721-725.
- Idrisse, A.A. (2023). ‘Classification Of Abnormal Respiratory Sounds Using Deep Learning Techniques’, Yüksek Lisans Tezi, Gazi Üniversitesi Bilişim Enstitüsü, Ankara, Türkiye.
- Jain Anchit. (2018). Improve your model accuracy by Transfer Learning. <https://medium.com/data-science-101/transfer-learning-57ce3b98650>
- Keleş, N. (2022). ‘Öksürük Ses Kayıtları Kullanılarak Covid-19 Tahmini’, Yüksek Lisans Tezi, Atatürk Üniversitesi Fen Bilimleri Enstitüsü, Erzurum, Türkiye.
- Kıvrak, E., Karasulu, B., Sozbır, C., Turkay, A., (2021). Ses Özneliklerini Kullanan Ses Duygu Durum Sınıflandırma için Derin Öğrenme Tabanlı Bir Yazılımsal Araç.

LeCun, Y., Bengio, Y. ve Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436-444.

Masraff Web Sayfası. Makine Öğrenmesi. Geliş tarihi 07 Ocak 2024, gönderen <https://www.masraff.co/makine-ogrenmesi/>

Mekruksavanich, S., Jitpattanakul, A., Hnoohom, N. (2020). Negative Emotion Recognition using Deep Learning for Thai Language. *2020 Joint International Conference on Digital Arts, Media and Technology with ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering (ECTI DAMT & NCON)*.

Mura, M., & Lamberti, P. (2020). Human-Machine Interaction Personalization: a Review on Gender and Emotion Recognition Through Speech Analysis. *2020 IEEE International Workshop on Metrology for Industry 4.0 & IoT*.

Onar Ali. (2020, Eylül 21). Derin Öğrenme Yöntemleri ve Uygulamaları. <https://alitunacanar.medium.com/derin-%C3%B6%C4%9Frenme-y%C3%B6ntemleri-ve-uygulamalar%C4%B1-1ce215de40e8>

Özaydın B. (2023). ‘X-Ray Görüntülerinden Alt Solunum Yolu Enfeksiyonlarının Teşhisinde Transfer Öğrenme Yöntemleri Ve Topluluk Öğrenmeye Dayalı Yeni Bir Yaklaşım’, Yüksek Lisans Tezi, Batman Üniversitesi Lisansüstü Eğitim Enstitüsü, Batman, Türkiye.

Parlak, C. (2022). ‘Geliştirilen Yeni Filtrelerin Ve Temel Frekans Tespit Yönteminin Derin Öğrenme İle Konuşma Duygu Analizinde Uygulanması’, Doktora Tezi, Düzce Üniversitesi Lisansüstü Eğitim Enstitüsü, Düzce, Türkiye.

Pepino, L., Riera, P., Ferrer, L. (2021). Emotion Recognition from Speech Using Wav2vec 2.0 Embeddings, *Interspeech 2021*, 3400-3404.

Phelan, A. L., Katz, R. ve Gostin, L. O. (2020). The novel coronavirus originating in Wuhan, China: Challenges for global health governance. *JAMA*, 323(8), 709–710.

Radford, A., Narasimhan, K., Salimans, T., Sutskever, I. (2018). Improving Language Understanding by Generative Pre-Training, *Preprint 2018*.

Sadhu S., He, D., Huang, C.W., Mallidi, S.H., Minhua, W., Rastrow, A., Stolcke, A., Droppo, J., Maas, R. (2021). Wav2vec-C: A Self-supervised Model for Speech

Representation Learning, *Interspeech 2021*, 711-715.

Shaw, A., Vardhan, R. K., & Saxena, S. (2016). Emotion recognition and classification in speech using Artificial neural networks. *International Journal of Computer Applications*, 145(8), 5-9.

Smith, J., Johnson, A. ve Brown, K. (2020). Transfer Learning and Transformers: Enhancing Performance in Natural Language Processing Tasks. *Journal of Artificial Intelligence*, 25(3), 45-62.

Şahin, T. (2022). ‘Dizel Makinanın Makina Öğrenmesi Yöntemi Kullanılarak Modellenmesi Ve Karar-Destek Mekanizması Oluşturulması’, Doktora Tezi, İstanbul Teknik Üniversitesi Lisansüstü Eğitim Enstitüsü, İstanbul, Türkiye.

Talbi, E. G. (2020). Machine learning into metaheuristics: A survey and taxonomy of data-driven metaheuristics. *HAL-Inria*, Hal-02745295.

TED Bilişim. Geliş tarihi 12 Ocak 2024, gönderen <https://www.baskonus.com.tr/>

Tiwari, U., Soni, M., Chakraborty, R., Panda, A., Kumar Kopparapu, S. (2022). Multi-Conditioning And Data Augmentation Using Generative Noise Model For Speech Emotion Recognition In Noisy Conditions. *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*.

Top, A. E. (2018). ‘Classification Of Eeg Signals Using Transfer Learning On Convolutional Neural Networks Via Spectrogram’, Yüksek Lisans Tezi, Ankara Yıldırım Beyazıt Üniversitesi Fen Bilimleri Enstitüsü, Ankara, Türkiye.

Turhost Web Sayfası. (2021, Kasım 26). Makine Öğrenmesi (Machine Learning) Nedir? <https://www.turhost.com/blog/makine-ogrenmesi-machine-learning-nedir/>

Türkdamar, M.U. (2023). ‘Derin Öğrenme İle Nesne Algılamada Transfer Öğrenme Ve İnce Ayar İşlemlerinin Etkinliklerinin Araştırılması’, Yüksek Lisans Tezi, Erciyes Üniversitesi Fen Bilimleri Enstitüsü, Kayseri, Türkiye.

Uludağ, B. (2020). ‘İşitme Kayıplı Bireylerin Derin Öğrenme Tabanlı Ses Duygu Analizi’, Yüksek Lisans Tezi, İstanbul Aydın Üniversitesi Lisansüstü Eğitim Enstitüsü, İstanbul, Türkiye.

Üstündağ, E. (2022). ‘Evrişimli Sinir Ağları, Transfer Öğrenme Ve Destek Vektör Makineleri Kullanarak Akciğer Röntgen Görüntülerinden Covid-19 Sınıflandırması’, Yüksek Lisans Tezi, Afyon Kocatepe Üniversitesi Fen Bilimleri

Enstitüsü, Afyon, Türkiye.

- Vaessen, N., Leeuwen, D.A. (2021). Fine-Tuning Wav2vec2 For Speaker Recognition, *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Singapore.
- Varghese, R. ve Sharma, S. (2018, Haziran). Affordable smart farming using IoT and machine learning. In *2018 Second International Conference on Intelligent Computing&Control Systems (ICICCS)*, Madurai, India.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I. (2017). Attention Is All You Need, *31st Conference on Neural Information Processing Systems (NIPS 2017)*, Long Beach, CA, USA
- Wang, Y., Boumadane, A., Heba, A. (2022). A Fine-Tuned Wav2vec 2.0/Hubert Benchmark For Speech Emotion Recognition, Speaker Verification And Spoken Language Understanding, *arXiv*. doi: 10.48550/ARXIV.2111.02735.
- Yang, M., Hirschi, K., Looney, S.D., Kang, O., Hansen, J.H.L. (2022). Improving Mispronunciation Detection with Wav2vec2-based Momentum Pseudo-Labeling for Accentedness and Intelligibility Assessment, *Interspeech 2022*, Incheon, Korea.
- Yaraş, N. (2020). ‘Vehicle Type Classification with Deep Learning’, Yüksek Lisans Tezi, İzmir Yüksek Teknoloji Üniversitesi Lisansüstü Eğitim Enstitüsü, İzmir, Türkiye.
- Yılmaz, M., & Orman, Z. (2021). LSTM Derin Öğrenme Yaklaşımı ile Covid-19 Pandemi Sürecinde Twitter Verilerinden Duygu Analizi. *Acta Infologica*, 0(0), 0-0.
- Zhai, S., Talbott, W., Srivastava, N., Huang, C., Goh, H., Zhang, R., Susskind, J. (2021). An Attention Free Transformer. *ICLR 2021 Conference Blind Submission*.
- Zhao, J., Mao, X., & Chen, L. (2019a). Speech emotion recognition using deep 1D & 2D CNN LSTM networks. *Biomedical Signal Processing and Control*, 47, 312-323.
- Zhao, Z. Q., Zheng, P., Xu, S. T. & Wu, X. (2019). Object detection with deep learning: a review. *IEEE Transactions on Neural Networks & Learning Systems*, 30(11), 3212–3232.

ÖZGEÇMİŞ

KİŞİSEL BİLGİLER

Adı Soyadı : Kübra SİNAN

Yabancı Dili : İngilizce

ÖĞRENİM DURUMU

Derece	Alan	Okul/Üniversite	Mezuniyet Yılı
Yüksek Lisans	Bilgisayar Müh.	Düzce Üniversitesi	2024
Lisans	Bilgisayar Müh.	Düzce Üniversitesi	2017
Lise		Özel Maltepe Gökyüzü Koleji	2012

YAYINLAR

Sinan, K., Erdoğan, P. (2023). Kent İçi Raylı Sistemler Bas Konuş Interkom Sistemlerinde Derin Öğrenme Ve Transfer Öğrenme İle Ses Duygu Analizi.2. Uluslararası Mühendislik Ve Fen Bilimleri Kongresi, 16-17 Aralık 2023.