

TC.  
İSTANBUL KÜLTÜR ÜNİVERSİTESİ  
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

MODLAR ARASI TRANSFER ÖĞRENİMİ İLE SES SİNYALLERİNDEN DUYGU  
TANIMA

YÜKSEK LİSANS TEZİ

Fahreddin Raşit KILIÇ

2200007048

Anabilim Dalı: Bilgisayar Mühendisliği

Programı: Bilgisayar Mühendisliği

Tez Danışmanı: Doç. Dr. Fatma Patlar Akbulut

TEMMUZ 2023

TC.  
İSTANBUL KÜLTÜR ÜNİVERSİTESİ  
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

MODLAR ARASI TRANSFER ÖĞRENİMİ İLE SES SİNYALLERİNDEN DUYGU  
TANIMA

YÜKSEK LİSANS TEZİ

Fahreddin Raşit KILIÇ

2200007048

Anabilim Dalı: Bilgisayar Mühendisliği

Programı: Bilgisayar Mühendisliği

Tez Danışmanı: Doç. Dr. Fatma Patlar Akbulut

Jüri Üyeleri: Dr.Ögr.Üyesi Öznur Şengel

Prof.Dr. Özgür K. Şahingöz

TEMMUZ 2023

# ÖNSÖZ

“Modlar arası transfer öğrenimi ile ses sinyallerinden duygu tanıma” konulu yüksek lisans tez çalışmamdaki her aşamada bilgi ve tecrübesiyle bana yardımcı olan değerli tez danışmanım Doç. Dr. Fatma Patlar AKBULUT’a teşekkürler ediyorum.

Eğitim hayatım boyunca bana desteğini hiçbir zaman esirgemeyen sevgili aileme en içten teşekkürlerimi sunarım.

**Fahreddin Raşit KILIÇ**  
**Temmuz 2023**

# İÇİNDEKİLER

|                                                                |    |
|----------------------------------------------------------------|----|
| ÖNSÖZ .....                                                    | i  |
| İÇİNDEKİLER .....                                              | ii |
| KISALTMALAR .....                                              | v  |
| ŞEKİL LİSTESİ .....                                            | vi |
| TABLO LİSTESİ .....                                            | ix |
| ÖZET .....                                                     | x  |
| ABSTRACT .....                                                 | xi |
| 1.GİRİŞ .....                                                  | 1  |
| 1.1 Tezin Literatüre Katkısı .....                             | 3  |
| 1.2 Tezin Organizasyonu .....                                  | 3  |
| 2. İLGİLİ ÇALIŞMALAR .....                                     | 4  |
| 2.1 Transfer Öğrenmesi .....                                   | 4  |
| 2.2 Görüntü Duygu Sınıflama .....                              | 6  |
| 2.3 Ses Sinyallerinden Duygu Sınıflama .....                   | 7  |
| 3. TEMEL TANIM, KAVRAM VE YÖNTEMLER .....                      | 12 |
| 3.1 Duygu ve ses sinyali sınıflandırma .....                   | 12 |
| 3.2 Ses ve Ses Sınıflandırması İçin Kullanılan Yöntemler ..... | 14 |
| 3.3 Duygu Sınıflama Görevi .....                               | 17 |
| 3.4 Özellik Çıkarım Yöntemleri .....                           | 19 |
| 3.4.1 Mel Frekansı Kepstrum Katsayıları (MFCC – MFKK) .....    | 19 |
| 3.4.1.1 Çerçeveleme .....                                      | 20 |
| 3.4.1.2 Pencereleme .....                                      | 21 |
| 3.4.1.3 FFT .....                                              | 21 |
| 3.4.1.4 Mel Frekansı Dönüşümü .....                            | 21 |
| 3.4.1.5 Kepsturum Katsayıları .....                            | 22 |
| 3.4.2 Log Mel Filter Bank (LMFB) .....                         | 22 |
| 3.5 Sınıflandırıcılar .....                                    | 22 |
| 3.5.1 Yapay sinir ağları .....                                 | 22 |
| 3.5.2 Derin öğrenme .....                                      | 27 |
| 3.5.3 Evrimsel Sinir Ağı .....                                 | 28 |
| 3.5.3.1 Giriş Katmanı .....                                    | 29 |
| 3.5.3.2 Konvolüsyon Katmanı .....                              | 30 |
| 3.5.3.3 Havuzlama katmanı .....                                | 31 |
| 3.5.3.4 Düzleştirme Katmanı .....                              | 32 |
| 3.5.3.5 Tam Bağlantılı Katman .....                            | 32 |

|                                                                                             |    |
|---------------------------------------------------------------------------------------------|----|
| 3.5.3.6 Dropout Katmanı .....                                                               | 33 |
| 3.5.3.7 Sınıflandırma Katmanı .....                                                         | 33 |
| 3.5.4 Yinelemeli Sinir Ağları (RNN) .....                                                   | 33 |
| 3.5.4.1 Uzun-Kısa Süreli Bellek Ağları (LSTM) .....                                         | 34 |
| 3.6 Birleştirme Yöntemleri .....                                                            | 36 |
| 4. MATERYAL .....                                                                           | 38 |
| 5. YÖNTEM .....                                                                             | 41 |
| 5.1 Veri Seti .....                                                                         | 41 |
| 5.2 Ses Öznitelik Çıkartıcıları .....                                                       | 45 |
| 5.3 Ses Sınıflama Yöntemi .....                                                             | 45 |
| 5.4 Görüntü Sınıflama Yöntemi .....                                                         | 45 |
| 5.5 Transfer Öğrenmesi Yöntemi .....                                                        | 46 |
| 5.6 Yöntemde Kullanılan Parametre Deneyleri .....                                           | 46 |
| 6. UYGULAMA .....                                                                           | 48 |
| 6.1 MFCC tabanlı ses sinyali duygu tanıma .....                                             | 49 |
| 6.1.1 MFCC tabanlı konuşma tipi ses sinyali duygu tanıma .....                              | 49 |
| 6.1.1.1 Konuşma tipi ses sinyali, MFCC öznitelik çıkartıcısı, Dense katmanları .....        | 49 |
| 6.1.1.2 Konuşma tipi ses sinyali, MFCC öznitelik çıkartıcısı, LSTM katmanları .....         | 51 |
| 6.1.2 MFCC tabanlı şarkı tipi ses sinyali duygu tanıma .....                                | 53 |
| 6.1.2.1 Şarkı tipi ses sinyali, MFCC öznitelik çıkartıcısı, Dense katmanları .....          | 53 |
| 6.1.2.2 Şarkı tipi ses sinyali, MFCC öznitelik çıkartıcısı, LSTM katmanları .....           | 55 |
| 6.2 LMFB tabanlı ses sinyali duygu tanıma .....                                             | 57 |
| 6.2.1 LMFB tabanlı konuşma tipi ses sinyali duygu tanıma .....                              | 57 |
| 6.2.1.1 Konuşma tipi ses sinyali, LMFB öznitelik çıkartıcısı, Dense katmanları .....        | 58 |
| 6.2.1.2 Konuşma tipi ses sinyali, LMFB öznitelik çıkartıcısı, LSTM katmanları .....         | 60 |
| 6.2.2 LMFB tabanlı şarkı tipi ses sinyali duygu tanıma .....                                | 62 |
| 6.2.2.1 Şarkı tipi ses sinyali, LMFB öznitelik çıkartıcısı, Dense katmanları .....          | 62 |
| 6.2.2.2 Şarkı tipi ses sinyali, LMFB öznitelik çıkartıcısı, LSTM katmanları .....           | 64 |
| 6.3 Görüntü tabanlı duygu tanıma için derin öğrenme modeli .....                            | 66 |
| 6.4 Duygu tanıma performansını arttırmak için transfer öğrenmesi .....                      | 68 |
| 6.5 Duygu tanıma performansını arttırmak için erken birleştirme yöntemi .....               | 70 |
| 7. SONUÇ .....                                                                              | 73 |
| 7.1 Modlar arası transfer öğrenimi yönteminin geleneksel yaklaşımla karşılaştırılması ..... | 73 |
| 7.2 Erken birleştirme yönteminin geleneksel yaklaşımla karşılaştırılması .....              | 73 |
| 7.3 MFCC ve LMFB özellik çıkarıcı yöntemlerin karşılaştırılması .....                       | 74 |

|                                                                     |    |
|---------------------------------------------------------------------|----|
| 7.4 Konuşma ve Şarkı tipi ses sinyallerinin karşılaştırılması ..... | 74 |
| 7.5 Dense ve LSTM tipi katman türlerinin karşılaştırılması .....    | 74 |
| 7.6 Sınırlamalar, zorluklar ve öneriler.....                        | 74 |
| 8. KAYNAKÇA.....                                                    | 75 |



## KISALTMALAR

MFCC : Mel Frequency Cepstral Coefficients (Mel Frekanslı Cepstral Katsayıları)  
DL: Deep Learning (Derin Öğrenme)  
LSTM: Long Short-Term Memory (Uzun Kısa Süreli Hafıza)  
ML: Machine Learning (Makine Öğrenmesi)  
LMFB: Log Mel Filter Bank  
FFT: Fast Fourier Transform (Hızlı Fourier Dönüşümü)  
CNN: Convolutional Neural Network  
ESA: Evrişimli Sinir Ağları  
RAVDESS: The Ryerson Audio-Visual Database of Emotional (Ryerson Görsel-İşitsel Duygusal Veri Seti)  
CREMA-D: Crowd-sourced Emotional Multimodal Actors Dataset (Kalabalık Kaynaklı Duygusal Çok Modlu Aktörler Veri Kümesi)  
SAVEE: Crowd-sourced Emotional Multimodal Actors Dataset ( Kalabalık Kaynaklı Duygusal Çok Modlu Aktörler Veri Kümesi)  
MSP-IMPROV: Multimodal database of spontaneous improvisations (Çoklu Modlu Spontan Performans Doğaçlaması Veri Seti)  
TESS: Toronto Emotional Speech Set (Toronto Duygusal Konuşma Seti)  
BAUM-1: Bahçeşehir University Multimodal Affective Database (Bahçeşehir Üniversitesi Çoklu Modlu Duygu Veri Seti)  
IEMOCAP: Interactive emotional dyadic motion capture database (Etkileşimli duygusal ikili hareket yakalama veri seti)  
EmoDB: Berlin Database of Emotional Speech (Berlin Duygusal Konuşma Veri Seti)  
TL: Transfer Learning (Transfer Öğrenmesi)

## ŞEKİL LİSTESİ

- Şekil 1.1 – Transfer Öğrenmesi Yöntemi
- Şekil 3.1 - Plutchik'in duygular çarkı
- Şekil 3.2 - Russel'in Çember Modeli
- Şekil 3.3: Ses Sinyali Grafiği
- Şekil 3.4: Tipik Bir Duygu Tespit Yaklaşımı
- Şekil 3.5: Duygu Sınıflama Aşamaları
- Şekil 3.6: MFCC Adımları
- Şekil 3.7: Çerçeveleme Yöntemi
- Şekil 3.8: Sinir Hücresi Yapısı
- Şekil 3.9: Yapay Sinir Ağı Yapısı
- Şekil 3.10: Sigmoid Fonksiyonu
- Şekil 3.11: RELU Fonksiyonu
- Şekil 3.12: Swish Fonksiyonu
- Şekil 3.13: Basamak Fonksiyonu
- Şekil 3.14: Sızdıran RELU Fonksiyonu
- Şekil 3.15: Üslü Doğrusal Birim Fonksiyonu
- Şekil 3.16: Softmax Fonksiyonu
- Şekil 3.17: Tanh Fonksiyonu
- Şekil 3.18: Parametrelili RELU Fonksiyonu
- Şekil 3.19: Lineer Fonksiyon
- Şekil 3.20: Yapay Sinir Ağı Katman Yapısı
- Şekil 3.21: Yapay Sinir Ağı ve Derin Sinir Ağı Yapısı
- Şekil 3.22: Evrişimsel Sinir Ağı Yapısı
- Şekil 3.23: Evrişimsel Sinir Ağı Filtre Yapısı
- Şekil 3.24: Evrişimsel Sinir Ağı Havuzlama Yapısı
- Şekil 3.25: Evrişimsel Sinir Ağı Düzleştirme Katmanı Fonksiyonu
- Şekil 3.26: Evrişimsel Sinir Ağı Düzleştirme Yapısı
- Şekil 3.27: Dropout Uygunlandıktan Sonraki Evrişimsel Sinir Ağı
- Şekil 3.28: Yinelemeli Sinir Ağı Yapısı
- Şekil 3.29: LSTM Ağ Yapısı
- Şekil 5.1 – Veri Seti Veri Türleri Dağılımı
- Şekil 5.2 – Nötr Ses Sinyali Örneği

Şekil 5.3 - Sakin Ses Sinyali Örneği  
Şekil 5.4 - Mutlu Ses Sinyali Örneği  
Şekil 5.5 - Üzgün Ses Sinyali Örneği  
Şekil 5.6 - Kızgın Ses Sinyali Örneği  
Şekil 5.7 - Korkulu Ses Sinyali Örneği  
Şekil 5.8 - İğrenme Ses Sinyali Örneği  
Şekil 5.9 - Şaşırmış Ses Sinyali Örneği  
Şekil 6.1: Konuşma, MFCC, Dense Uygulaması Karışıklık Matrisi  
Şekil 6.2: Konuşma, MFCC, Dense Uygulaması Doğruluk Grafiği  
Şekil 6.3: Konuşma, MFCC, Dense Uygulaması Kayıp Grafiği  
Şekil 6.4: Konuşma, MFCC, LSTM Uygulaması Karışıklık Matrisi  
Şekil 6.5: Konuşma, MFCC, LSTM Uygulaması Doğruluk Grafiği  
Şekil 6.6: Konuşma, MFCC, LSTM Uygulaması Kayıp Grafiği  
Şekil 6.7: Şarkı, MFCC, Dense Uygulaması Karışıklık Matrisi  
Şekil 6.8: Şarkı, MFCC, Dense Uygulaması Doğruluk Grafiği  
Şekil 6.9: Şarkı, MFCC, Dense Uygulaması Kayıp Grafiği  
Şekil 6.10: Şarkı, MFCC, LSTM Uygulaması Karışıklık Matrisi  
Şekil 6.11: Şarkı, MFCC, LSTM Uygulaması Doğruluk Grafiği  
Şekil 6.12: Şarkı, MFCC, LSTM Uygulaması Kayıp Grafiği  
Şekil 6.13: Konuşma, LMFB, Dense Uygulaması Karışıklık Matrisi  
Şekil 6.14: Konuşma, LMFB, Dense Uygulaması Doğruluk Grafiği  
Şekil 6.15: Konuşma, LMFB, Dense Uygulaması Kayıp Grafiği  
Şekil 6.16: Konuşma, LMFB, LSTM Uygulaması Karışıklık Matrisi  
Şekil 6.17: Konuşma, LMFB, LSTM Uygulaması Doğruluk Grafiği  
Şekil 6.18: Konuşma, LMFB, LSTM Uygulaması Kayıp Grafiği  
Şekil 6.19: Şarkı, LMFB, Dense Uygulaması Karışıklık Matrisi  
Şekil 6.20: Şarkı, LMFB, Dense Uygulaması Doğruluk Grafiği  
Şekil 6.21: Şarkı, LMFB, Dense Uygulaması Kayıp Grafiği  
Şekil 6.22: Şarkı, LMFB, LSTM Uygulaması Karışıklık Matrisi  
Şekil 6.23: Şarkı, LMFB, LSTM Uygulaması Doğruluk Grafiği  
Şekil 6.24: Şarkı, LMFB, LSTM Uygulaması Kayıp Grafiği  
Şekil 6.25: Görüntü Sınıflandırma Uygulaması Karışıklık Matrisi  
Şekil 6.26: Görüntü Sınıflandırma Uygulaması Doğruluk Grafiği  
Şekil 6.27: Görüntü Sınıflandırma Uygulaması Kayıp Grafiği

Şekil 6.28: Transfer Öğrenimi Uygulaması Karışıklık Matrisi

Şekil 6.29: Transfer Öğrenimi Uygulaması Doğruluk Grafiği

Şekil 6.30: Transfer Öğrenimi Uygulaması Kayıp Grafiği

Şekil 6.31: Erken Birleştirme Uygulaması Karışıklık Matrisi

Şekil 6.32: Erken Birleştirme Uygulaması Doğruluk Grafiği

Şekil 6.33: Erken Birleştirme Uygulaması Kayıp Grafiği



## **TABLO LİSTESİ**

Tablo 3.1: Ekman'ın duygu kategorizasyonu

Tablo 3.2: Aktivasyon Fonksiyonları Listesi ve Formülasyonları

Tablo 4.1: Literatürde kullanılan veritabanları

Tablo 6.1: Konuşma, MFCC, Dense Uygulaması Sınıflama Raporu

Tablo 6.2: Konuşma, MFCC, LSTM Uygulaması Sınıflama Raporu

Tablo 6.3: Şarkı, MFCC, Dense Uygulaması Sınıflama Raporu

Tablo 6.4: Şarkı, MFCC, LSTM Uygulaması Sınıflama Raporu

Tablo 6.5: Konuşma, LMBF, Dense Uygulaması Sınıflama Raporu

Tablo 6.6: Konuşma, LMFB, LSTM Uygulaması Sınıflama Raporu

Tablo 6.7: Şarkı, LMFB, Dense Uygulaması Sınıflama Raporu

Tablo 6.8: Şarkı, LMFB, LSTM Uygulaması Sınıflama Raporu

Tablo 6.9: Görüntü Verilerinden Duygu Tanıma Sınıflama Raporu

Tablo 6.10: Transfer Öğrenmesi ile Duygu Tanıma Sınıflama Raporu

Tablo 6.11: Erken Birleştirme Yöntemi ile Duygu Tanıma Sınıflama Raporu

|                    |   |                                     |
|--------------------|---|-------------------------------------|
| Üniversite         | : | İstanbul Kültür Üniversitesi        |
| Enstitü            | : | Lisansüstü Eğitim Enstitüsü         |
| Anabilim Dalı      | : | Bilgisayar Mühendisliği             |
| Programı           | : | Bilgisayar Mühendisliği             |
| Tez Danışmanı      | : | Dr. Öğr. Üyesi Fatma Patlar AKBULUT |
| Tez Türü ve Tarihi | : | Yüksek Lisans – Mayıs 2023          |

## ÖZET

### MODLAR ARASI TRANSFER ÖĞRENİMİ İLE SES SİNYALLERİNDEN DUYGU TANIMA

**Fahreddin Raşit Kılıç**

İnsanların konuşma sırasında ifade ettikleri duyguları anlamak, uygun şekilde tepki vermek için önemlidir. Ses sinyallerinden anlayabileceğimiz bilgileri maksimize edebilmek için ilgili ses ve görüntünün transfer öğrenimi yöntemi ile analiz edilmesi önemlidir. Duygu tanıma çalışmalarıyla alakalı olarak derin öğrenme ve yapay zekâ algoritmalarıyla araştırmalar hız kazanmıştır. Özellikle yapay zekâ ve robotik sistemlerde, doğal ve empatik bir insan-makine etkileşimi sağlamak için ses sinyallerinden duygu analizi esastır. Bu sistemler sayesinde kullanıcı deneyimini zenginleştirerek daha etkili ve tatmin edici hizmetler sunulabilmektedir.

Duygu analizi sağlık sektöründe de önemli bir rol oynamaktadır. Psikolojik hastalıkların teşhis ve takibinde, hastaların duygu durumlarını doğru bir şekilde tespit etmek, uygun tedavi ve müdahalelerin gerçekleştirilmesi için kritiktir. Eğitim sektöründe ise, öğrencilerin ve öğretmenlerin duygusal durumlarını anlamak, eğitim ve öğretim süreçlerini daha etkili hale getirmektedir. Reklam ve pazarlama alanında, tüketici duygularını analiz etmek, müşteri memnuniyetini ve marka sadakatini artırarak satışları ve karlılığı yükseltmektedir. Ayrıca, duygu analizi, oyun endüstrisinde daha gerçekçi ve etkileyici oyun deneyimleri sunmak için de kullanılmaktadır.

Bu tez çalışmasında, ses sinyallerinden ve ses sinyallerine ait ilgili görüntülerden transfer öğrenme yöntemi ile bu verilerin duygu durumlarını tespit etmeye yönelik gelişmiş sınıflandırma ve analiz yöntemlerini kullanarak doğru duygu tahminlerinde bulunmayı hedeflenmektedir. Bu çalışmada veri seti nötr, sakin, mutlu, üzgün, kızgın, korkulu, tikslenme ve şaşırma olmak üzere 8 farklı duygu durumu kullanılmıştır. Ses verilerini analiz edebilmek için MFCC ve Log Mel Filter Bank olmak üzere iki yöntem, Dense ve LSTM olmak üzere iki derin öğrenme tekniği kullanılmıştır. Video veri setini analiz edebilmek içinse CNN ağ modeli kullanılmıştır. Toplamda 11 farklı uygulama gerçekleştirilen bu uygulamada modellerin başarısı analiz edilmiş ve sonuç olarak görüntü verilerinden sınıflama gerçekleştiren modelden konuşma ses sinyalleri verilerinden sınıflama gerçekleştiren modele transfer öğrenmesi yöntemi ile bilgi aktarımı gerçekleştirilip %6,78'lik başarı artışı sağlanmıştır. Ayrıca MFCC yönteminin LMFB'a göre daha başarılı olduğu, şarkı ses türünün ise konuşma ses türüne göre daha yüksek doğrulukla etiketlendiği görülmüştür.

Anahtar Kelimeler : Ses sinyalleri, duygu analizi, derin öğrenme, transfer öğrenmesi, CNN, LSTM, MFCC, LMFB, Dense

|                      |   |                                 |
|----------------------|---|---------------------------------|
| University           | : | İstanbul Kültür University      |
| Institute            | : | Institute of Graduate Education |
| Branch               | : | Computer Engineering            |
| Program              | : | Computer Engineering            |
| Thesis Advisor       | : | Dr. Fatma Patlar AKBULUT        |
| Thesis Type and Date | : | Master Degree – May 2023        |

## ABSTRACT

### EMOTION RECOGNITION FROM AUDIO SIGNALS WITH CROSS-MODAL TRANSFER LEARNING

Fahreddin Raşit Kılıç

Understanding the emotions people express during conversation is important to responding appropriately. In order to maximize the information we can understand from audio signals, it is important to analyze the relevant audio and video with the transfer learning method. Related to emotion recognition studies, research has gained momentum with deep learning and artificial intelligence algorithms. Especially in artificial intelligence and robotic systems, emotion analysis from sound signals is essential to providing a natural and empathetic human-machine interaction. Thanks to these systems, more effective and satisfying services can be provided by enriching the user experience.

Sentiment analysis also plays an important role in the healthcare industry. In the diagnosis and follow-up of psychological diseases, it is critical to accurately determine the mood of the patients and to carry out appropriate treatment and interventions. In the education sector, understanding the emotional states of students and teachers makes education and training processes more effective. In the field of advertising and marketing, analyzing consumer emotions increases sales and profitability by increasing customer satisfaction and brand loyalty. In addition, sentiment analysis is also used in the gaming industry to provide more realistic and immersive gaming experiences.

The objective of this thesis is to accurately predict emotional states based on audio signals and simultaneous facial images using a cross-domain transfer learning approach. Eight distinct states of emotion were used in this analysis: neutral, calm, happy, sad, angry, fearful, disgusted, and surprised. Traditional signal processing methods, such as MFCC and Log Mel Filter Bank, and deep learning techniques, such as Densely connected network, CNN, and LSTM, were used to analyze the audio data. The videos were evaluated by employing a CNN network model. In this application, which consisted of a total of 11 different applications, the success of the models was analyzed, and as a result, information was transferred from the model performing the classification from the image data to the model performing the classification from the speech audio signal data using the cross-modal transfer learning, resulting in a 6.78 percent improvement in success. In addition, it was discovered that the MFCC method was more effective than the LMFB, and that the song voice type was identified with greater precision than the speech voice type.

Keywords: Audio signals, sentiment analysis, deep learning, transfer learning, CNN, LSTM, MFCC, LMFB

# 1.GİRİŞ

Duygular, her zaman insanlar arasındaki iletişimin temel öğelerinden olmuştur ve insanlar arası etkileşimlerin anlaşılması ve yorumlanması açısından büyük önem taşır. Bu durum araştırmacıları insan duygularının bilgisayar sistemlerinde anlaşılması için motive etmiştir [1]. İnsanlarda duygu tanıma gerçekleştirmek için birçok fizyolojik verilerden faydalınınsa da temel veriler yüz ifadesinde ki mimikler ve ses sinyalleridir. Veri, görüntü veya ses olsa da taşıdığı bilgilerin sınıflanabilirlik kalitesi kişinin cinsiyeti, yaşı, ruh hali gibi birçok çok faktöre bağlıdır [3]. Bu tez içerisinde ayrıca aynı duyguyu taşıyan ses verilerinin söyleyiş tarzındaki farklılıklarının modelin başarısına nasıl etki ettiği de incelenecektir.

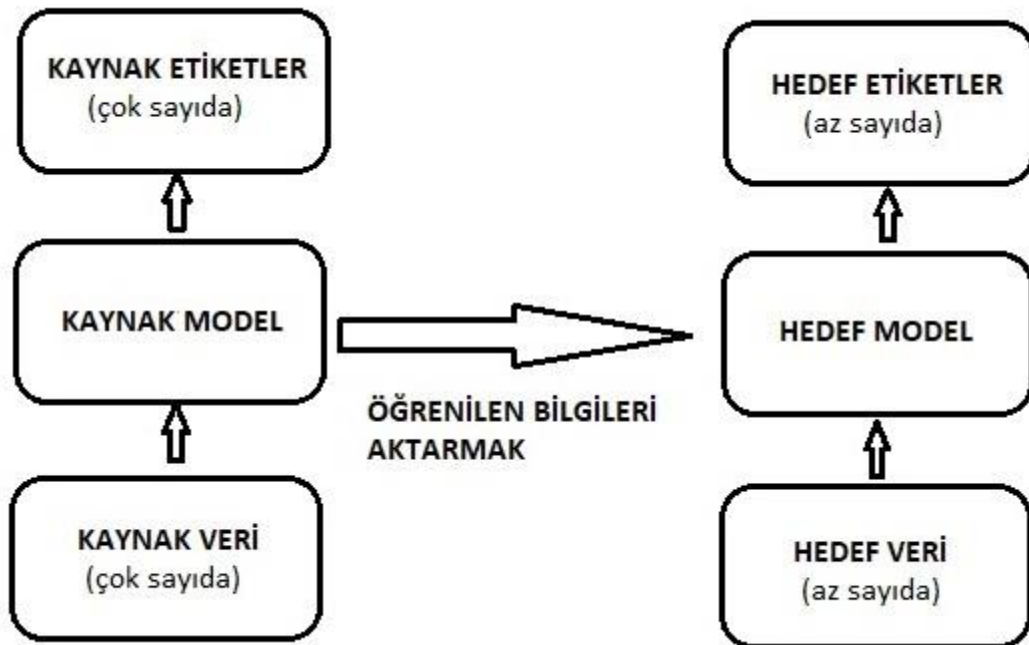
Duyguları ses, metin ve görüntü gibi farklı ortamlardan sınıflamayı amaçlayan birçok model ortaya çıkmıştır. Bu modeller insan-makine etkileşimi, sağlık hizmetleri, reklamcılık ve pazarlama gibi çeşitli alanlarda uygulama potansiyeline sahiptir. Örneğin, sağlık hizmetlerinde özellikle ruh sağlığını ilgilendiren durumlarda duyguların analizi hastaların yüz ifadesindeki mimikler ve sesindeki sinyallerden gerçekleştirilebilir, doktorların doğru teşhis koymasındaki başarıyı arttırabilir [2],[10]. Ayrıca reklamcılık ve pazarlama sektörlerinde mimik ve ses verilerinden duygu durum analizi sayesinde müşterilere özel pazarlama stratejileri geliştirilebilir.

Ses sinyallerini sınıflandırmak, konuşma sinyali boyunca farklı duygu sınıflarını da içerebilmesi, doğru yöntemin belirlenebilmesi, sınıflandırıcının en uygun parametrelerinin bulunabilmesi, duyguyu içerebilecek bölümün atlama boyutu ve pencere boyutu gibi parametrelere doğrudan bağımlı olması bakımından görüntü verilerini sınıflandırmaya göre daha karmaşıktır ve bunun sonucunda da ses verilerinin sınıflanabilirliği daha zayıf olabilir. Bu problemleri minimize etmek için ses sinyali ideal pencere boyutlarında (20- 30 ms) incelenmelidir. [1],[3].

Duygu tanımada yaklaşımlar genel olarak tekil ve çoklu olarak iki başlık altında toplanabilir. Tekil yaklaşımda veri tek başına ses veya görüntü olabilir ve bu verinin işlenmesiyle duygu tanıma yapılır. Tekil yaklaşımın avantajı spesifik bir veri seti üzerinde detaylı araştırmalar yapmaya imkân vermesidir ama bu gerçek hayatta yeterli olmayabilir çünkü hayatta karşılaştığımız durumlar çoğu zaman karmaşık veri türlerinden meydana gelir. İşte bu noktada çoklu yaklaşım kullanılması avantaj sağlar. Çoklu yaklaşımda birçok kaynaktan farklı

veri türleri birlikte değerlendirilebilir, örneğin bu tez çalışmasında ses ve görüntü bilgileri birleştirilerek duygu tanınmanın başarısı artırılmaya çalışılmıştır. [3],[7],[8].

Transfer öğrenmesi, derin öğrenme ve makine öğrenmesi alanlarında popülerlik kazanmaktadır çünkü önemli avantajlar sunar. Transfer öğrenmesinde genel amaç önceden eğitilmiş modellerin bilgisini kullanarak daha az etiketli veriyle yeni modeli eğitirken başarıyı maksimize etmektir. Bu özellikle etiketli veri setine erişimin maliyetli olduğu durumlarda çok büyük bir avantajdır. Transfer öğrenmesi önceden eğitilmiş modellerin bilgisini yeni modelde kullandığı için modelin eğitilirken daha iyi bir başlangıç noktasından başlamasını sağlar. Ayrıca yeni model önceden eğitilmiş modeldeki bilgilerin üzerine inşa edildiğinden daha kısa bir eğitim süresi sağlanır [4]. Bu avantajlarına rağmen transfer öğrenmesinde başarı artışı garanti değildir. Bazen kaynak modelden hedef modele transfer öğrenmesi gerçekleştirmek hedef modeldeki performansı arttırmayabilir ve hatta başarının düşmesine sebep olabilir. Bu durum literatürde olumsuz öğrenme (negative learning) olarak adlandırılmıştır. Olumsuz öğrenmeden kaçınmak için transfer öğrenmesi yapılacak alanların ve öğrenme modellerinin benzer olması tavsiye edilmektedir [5],[91]. Şekil 1.1’de transfer öğrenmesini özetleyen bir görsel gösterilmiştir [6].



Şekil 1.1: Transfer Öğrenmesi Yöntemi

Ses sinyallerinden ve görüntü verilerinden duygu tanıma 3 başlık altında toplanabilir. İlk adımda veri seti seçimi ve organizasyonu gerçekleştirilmelidir. İkinci adımda veri setinden öznitelik çıkarmak için uygun yöntemlerin seçilmesi. Son adımda ise uygun sınıflandırma modeli ve tekniği öznitelik matrisine uygulanmalıdır [1],[9].

## 1.1 Tezin Literatüre Katkısı

Bu tez çalışmasında RAVDESS veri seti içerisindeki 8 farklı duygu içeren konuşma ses sinyalleri ve görüntü verileri ile 6 farklı duygu içeren şarkı tipi ses sinyalleri verileri kullanılmıştır. Toplamda 11 farklı uygulama gerçekleştirilmiştir. Uygulamalarda, özellik çıkarıcı olarak MFCC veya LMFB, katman tipi olarak Dense veya LSTM kullanılmıştır.

Çalışmanın literatüre sağladığı katkılar şunlardır:

Ses sinyallerindeki duygunun sınıflanmasında veri tipinin, özellik çıkarıcı yöntemin ve yapay sinir ağı yapısındaki kullanılan katman tipinin başarıyı nasıl etkilediği gösterilmiştir. Transfer öğrenmesi yöntemi ile görüntülerdeki duyguyu sınıflayan modelden konuşma ses sinyallerindeki duyguyu sınıflayan modele bilgi aktarımı gerçekleştirilmiş, konuşma ses sinyallerinden duygu sınıflayan modelin başarısı arttırılarak literatüre katkı sağlanmıştır. Erken birleştirme yöntemi ile görüntülerdeki duyguyu sınıflayan modelden konuşma ses sinyallerindeki duyguyu sınıflayan modele öznitelik aktarımı gerçekleştirilmiş, konuşma ses sinyallerinden duygu sınıflayan modelin başarısı arttırılarak literatüre katkı sağlanmıştır.

## 1.2 Tezin Organizasyonu

Bölüm 1’de tezin konusu belirtilmiş ve çalışmaya giriş yapılmıştır.

Bölüm 2’de tezin çalışma konusuyla alakalı literatürdeki ilgili çalışmalar gösterilmiştir.

Bölüm 3’de tezde kullanılan temel tanım ve kavramlar açıklanmıştır.

Bölüm 4’de çalışmada kullanılan materyal hakkında bilgiler verilmiştir.

Bölüm 5’de çalışmada gerçekleştirilen uygulamaların yöntemi açıklanmıştır.

Bölüm 6’da çalışmada gerçekleştirilen uygulamalar ve sonuçları gösterilmiştir.

Bölüm 7’de çalışmada gerçekleştirilen uygulamaların sonuçlarının performans değerlendirmeleri yapılmıştır.

Bölüm 8’de çalışmada kullanılan kaynaklar sıralanmıştır.

## 2. İLGİLİ ÇALIŞMALAR

Literatürde transfer öğrenimi, görüntü duygu sınıflama ve ses duygu sınıflama alanlarında çeşitli yöntemler kullanılarak çeşitli konulara çözüm bulunmaya çalışılmıştır. Bu çalışmaların bazıları aşağıda incelenmiştir.

### 2.1 Transfer Öğrenmesi

Fırıldak K. ve Talu M. F. 2019 yılında evrişimsel sinir ağlarında transfer öğrenmesinin başarısını incelemiştir. Özellikle görüntü verilerini içeren veri setlerinde yüksek başarısından dolayı evrişimsel sinir ağları tercih edilir. Ama evrişimsel sinir ağları eğitilmesinin yüksek maliyetli olması gibi bir dezavantajı vardır. Ayrıca hedeflenen göreve yönelik en uygun veri setini her zaman elde edemeyebiliriz. Bu nedenle kapsamlı ve başarısını kanıtlamış veri setleriyle eğitilmiş bir evrişimli sinir ağları modelinin ağırlıklarının transfer öğrenmesinde kullanılması çok yaygındır. Yazarlar bu çalışmada AlexNet'ten transfer edilen ağırlıklarla farklı veri kümeleri için ESA'ların sınıflama başarısını incelemiştir. Sonuç olarak, transfer öğrenme ile ESA'ların sınıflama doğruluğunun arttığı ve kaynak ESA'nın eğitim veri kümesinin hedef veri kümesine olan benzerliğinin de önemli bir etken olduğu gösterilmiştir [11].

Literatürde transfer öğrenimi ile alakalı bir diğer çalışmayı 2019 yılında AKILOTU B. N. ve arkadaşları hazırlamıştır. Yazarlar bu çalışmada sıtma hastalığını transfer öğrenmesi yönteminden faydalananarak tespit etmeye çalışmıştır. Sıtma, dişi anofel türü sivrisinek ısırığı kaynaklı bulaşıcı ve ölümcül olabilen bir hastalıktır bu nedenle erken tespiti hayati önem taşımaktadır. Sıtma virüsünün tespiti birkaç yöntemle yapılmaktadır ve bu yöntemler hem maliyetli hemde tedavinin geç başlamasına sebep olabilir. Belirtileri grip virüsünün etkilerine benzer ve başta tropik bölgeleri etkiler. Yazarlar bu çalışmada veri kümesi olarak Giemsa ile boyanmış kırmızı kan hücreleri içeren sıtma virüsü ile sağlıklı hücrelerin görüntülerini kullanmıştır. Derin öğrenme yapısı olarak AlexNet ve VGGNet modellerinin Fc6 ve Fc7 katmanlarının ağırlıkları kullanılmıştır. Sınıflandırıcı olarak SVM (destek vektör makineleri) kullanılmıştır ve sonuç olarak çalışmada en yüksek başarıyı transfer öğrenme aracılığıyla %96.02 doğruluk ile VCC16 ağı sağlamıştır [12].

2021 yılında Elmas B., transfer öğrenimi ile ağaç kabuğu görüntülerinden ağaç türlerinin tanımlanması konulu bir çalışma hazırlamıştır. Yazar bu çalışmada ağaçların daha verimli

kullanılabilinmesini amaçlamaktadır. Bu çalışmada veri seti olarak 59 farklı ağaç türünden 24.686 ağaç kabuğu görüntüsü kullanılmıştır. AlexNet, DenseNet201, ResNet18, ResNet50, ResNet101, VGG16, VGG19 olmak üzere önceden eğitilmiş 7 adet evrimsel sinir ağı kullanılmıştır. Ayrıca bu çalışmada ağların derinlik ve parametre sayısı gibi özelliklerin çözüme etkisi de incelenmiştir. Veri boyutunun ağ parametre sayısına oranla düşük kalması nedeniyle parametre sayısı arttıkça başarının düştüğü, derinliğin artırıldıkça başarının yükseldiği görülmüştür. Sonuç olarak diğer ağların başarı oranı %93 ile %95 arasındayken en başarılı iki ağ olan DenseNet201 ve ResNet50'in başarıları sırasıyla %99.51 ve %99.42 olmuştur [13].

2020 yılında Fırıldak K. ve Talu M. F., transfer öğrenimi ve imge üretiminden faydalanarak cilt lezyonlarını sınıflandırma doğruluğunu iyileştirmeyi amaçlayan bir çalışma hazırlamışlardır. Dünyada sıkça rastlanan kanser çeşitlerinden biri olan cilt kanseri yapılan araştırmalara göre ABD'de her yıl 10000 civarı insanın hayatını kaybetmesine sebep olmaktadır.[15] Bilgisayarlı görü sistemlerinde cilt kanserinin yüksek doğrulukla tespit edilmesi hayatidir. Yazarlar bu makalede HAM10000 veri kümesini kullanmışlardır. Bu çalışmada transfer öğrenmesi aşamasında ImageNet veri kümesiyle eğitilmiş AlexNet modelindeki ağırlıkların bir kısmı ya da tamamını yeni ESA'ya aktarılmıştır. İmge üretimi aşamasında ise gerçek kanser görüntülerinden yeni kanser görüntüleri üretmek için klasik üretme ve ÇÜA yöntemlerinden faydalanılarak veri setinin büyütülmesi amaçlanmaktadır. Bu çalışmada klasik üretme ve DEÇÜA yöntemlerinin başarısı ayrıca değerlendirilmiştir. DEÇÜA, klasik veri artırım yöntemine göre %15 ile %20 arasında sınıflama doğruluğunu iyileştirdiği ortaya çıkmıştır. Sonuç olarak bu çalışmada kısmi ağırlıklı transfer öğrenme ve DEÇÜA temelli imge üretim tekniği birlikte kullanıldığında en yüksek doğruluk oranına (%93) ulaşıldığı gösterilmiştir [14].

2016 yılında Karl Weiss ve arkadaşları transfer öğrenmesinin geleneksel makine öğrenmesi yöntemlerine kıyasla daha verimli ve esnek bir çözüm olduğunu vurgulamışlardır. Makine öğrenmesi, verilerden öğrenme ve tahmin yapma yeteneği olan bir bilgisayar bilimi dalıdır. Makine öğrenmesi yöntemleri, farklı alanlarda birçok uygulama için kullanılmaktadır. Ancak, makine öğrenmesi yöntemleri genellikle belirli bir alan veya görev için tasarlanmıştır ve farklı alan veya görevlerde aynı performansı gösteremezler. Bu nedenle, farklı alan veya görevler arasında bilgi aktarımı yapabilen ve öğrenme sürecini hızlandıran ve iyileştiren transfer öğrenme gibi yeni yöntemlere ihtiyaç vardır. Transfer öğrenme, kaynak alan veya görevde öğrenilen bilgiyi hedef alan veya görevde kullanarak öğrenme performansını artırmayı

amaçlayan bir makine öğrenmesi tekniğidir. Transfer öğrenme yöntemleri, homojen ve heterojen transfer öğrenmesi olarak iki ana kategoriye ayrılabilir. Homojen transfer öğrenme, kaynak ve hedef alanların aynı özellik uzayına sahip olduğu, ancak veri dağılımlarının farklı olabileceği transfer öğrenme yöntemidir. Bu yöntemde, kaynak ve hedef alanlar arasındaki dağılım farkını azaltmak için çeşitli teknikler kullanılır. Örneğin, özellik seçimi, özellik dönüştürme, örnek seçimi veya ağırlıklandırma gibi teknikler homojen transfer öğrenmede kullanılabilir. Heterojen transfer öğrenme ise kaynak ve hedef alanların hem farklı özellik uzaylarına hem de farklı veri dağılımlarına sahip olduğu transfer öğrenme yöntemidir. Bu yöntemde, kaynak ve hedef alanlar arasında bir bağlantı kurmak için çeşitli teknikler kullanılır. Örneğin, ortak bir gizli uzaya eşleştirme, ortak bir alt uzay bulma veya ortak bir temsil öğrenme gibi teknikler heterojen transfer öğrenmede kullanılabilir. Transfer öğrenme yöntemleri, bilgisayar görüşü, doğal dil işleme, biyoenformatik ve robotik gibi farklı alanlarda başarıyla uygulanmıştır. Sonuç olarak bu çalışmada, transfer öğrenmenin tanımı, yöntemleri ve uygulamaları hakkında genel bir bilgi verilmiştir. Transfer öğrenmenin geleneksel makine öğrenmesinin sınırlamalarını aştığı ve farklı alanlarda verimli bir şekilde öğrenme sağladığı gösterilmiştir. Ayrıca, transfer öğrenme için kullanılabilen yazılımlar ve gelecekteki araştırma yönleri de sunulmuştur [30].

Chenjing Cai ve arkadaşları Transfer öğrenmenin ilaç keşfinde nasıl kullanılabileceği anlatan bir çalışma hazırlamışlardır. Makalede transfer öğrenmenin avantajlarına ve zorluklarına değinilmektedir. Makalede yazarlar, transfer öğrenmenin ilaç keşfi alanında büyük potansiyele sahip olduğu ancak henüz tam olarak kullanılmadığı vurgulanmaktadır. Yazarlar makalede sonuç olarak, transfer öğrenmenin ilaç keşfi için gelecekte aktif öğrenme ve pekiştirmeli öğrenme gibi yöntemlerle daha yaygın olarak kullanılabileceğini belirtmektedir [31].

## 2.2 Görüntü Duygu Sınıflama

Duygu analizi ilk defa Nasukava ve Yin tarafından 2003 yılındaki bir araştırmalarında kullanılmıştır. Çalışmalarında önceden belirledikleri nesneler için duygu çıkarımı yapmışlardır. Duyguları nesnelere göre olumlu ve olumsuz kutup ile ilişkilendirme çalışmışlardır. Çalışmada duygunun kelimelerle nasıl ifade edildiği anlaşılmaya çalışılmıştır. Diğer bir önemli çıkarım ise nesnelerin duygularla ilişkisinin doğru çıkarım yapabilme kabiliyeti olmuştur. Bu çalışma web

üzerinde çalışan haber kaynaklarını analiz ederek duygu çıkarımı yapılmıştır ve analiz sonucunda(yüzdelere) yüksek başarı sağlanmıştır [16].

Schuller ve takımı (2004), sürücü güvenliğini sağlamak amacıyla bir uyarı durumunda araç içi kontrol mekanizmalarının devreye alındığı bir çalışma Schuller ve ark tarafından 2004 yılında yapılmıştır. Bu çalışmada, nitelikler Doğrusal Ayrımcı Analizi (LDA) vasıtasıyla düzenlenip başlıca 33 nitelik seçilmiştir. Bu nitelikler aracılığıyla, sürücünün hislerini denetleyerek sürücü yanlışlarını düzeltme ve emniyeti temin etme hedeflenmiştir [17].

2021 yılında Elmas B., transfer öğrenimi ile mantar görüntülerinden mantar türlerinin tanımlanması konulu bir çalışma hazırlamıştır. Yazar bu çalışmada kritik öneme sahip mantarların korunması ve sürdürülebilir stratejik planlarının yapılmasına katkı sağlamayı amaçlamaktadır. Bu çalışmada veri seti 18 aileye ait 472 farklı mantar sınıfından veriler içerilmektedir. AlexNet, ResNet18, ResNet50, ResNet101, VGG16, VGG19 olmak üzere önceden eğitilmiş 6 adet evrimsel sinir ağı kullanılmıştır. Sonuç olarak en başarılı ağı doğruluk oranı %97.62 olmuştur [25].

## 2.3 Ses Sinyallerinden Duygu Sınıflama

Duygu tanıma çalışmalarında etkili sonuçlar sunan metotlardan bir tanesi Mel Frekans Kepstral Katsayıları (MFCC)'dir. 2011 yılında Bozkurt ve arkadaşları ses sinyallerinden ağırlıklı MFCC (WMFCC) niteliklerini çıkartarak farklı duygu durumlarını tanıma işlemini başarmışlardır. Çalışma için FAU Aibo veri setini kullanmışlardır. Sonuç olarak , iki-sınıflı duygu tanıma doğruluğunda maksimum %70.29, beş-sınıflı duygu tanıma modelinde maksimum %43.59 başarımlarına ulaşarak standart spektral yöntemlere kıyasla daha yüksek bir başarımla gerçekleştirmişlerdir [2],[18].

2015 yılında Durukal tarafından konuşma sinyallerinin analiz edilip bireyin duygusal durumunun belirlenmesine yönelik gerçekleştirilen çalışmada, MFCC katsayıları nitelik çıkartıcı olarak kullanılarak farklı duygu durumlarının sınıflanması görevinde başarıları incelenmiştir. Bu çalışmada sınıflandırma yöntemleri olarak kNN ve DVM tercih edilmiştir. MFCC özellik çıkartım yönteminin kullanıldığı çalışmalarda LPC, prozodik ve kalite özellikleri kullanan diğer çalışmalarla karşılaştırıldığında daha yüksek başarı oranlarına ulaştığı belirlenmiştir [2],[19].

2016 yılında Karimi ve Sedaaghi, EmoDB veri setinden faydalananak kızgın, sıkıntılı, korkulu ve mutlu duygu durumlarını sınıflamaya odaklanmıştır. Temel frekans, enerji, MFCC, LPC ve PLP özellikleriyle bunların maksimum, minimum, ortalama ve varyans değerleri hesaplanarak elde edilmiştir. DVM ve Bayes sınıflandırma yöntemleri kullanılarak, konuşmacıya bağlı olarak ve 10 çapraz doğrulama kullanılan bu çalışmada, DVM ile %93.10 ve Bayes ile %83.73 duygu tanıma doğruluğu elde edilmiştir [2],[20].

Milton ve Tamil Selvi tarafından 2014 yılında yapılan bir çalışmada MFCC, LPC ve AR öznitelik çıkartıcıları kullanmışlardır. Sınıf-bağımlı çoklu sınıflandırıcı yöntemi kNN, GKM ve YSA yöntemleriyle gerçekleştirmişlerdir. Sınıflandırma süreci için yalnızca konuşmacı bağımsız, yalnızca metin bağımsız ve hem konuşmacı hem metin bağımsız olmak üzere üç grup hazırlanmıştır. Performans değerlendirmesi için tek çıkışlı 10 çapraz doğrulama yöntemi uygulanmış ve özgüllük, duyarlılık ve f-ölçümü bu çalışma için esas ölçütler olarak ele alınmıştır. Analiz edilen sonuçlar SD değeri AR ve MFCC özellikleri birlikte kullanıldığında konuşmacı bağımsız sınıflandırmada %57.8, metin-bağımsız sınıflandırmada %83.7 ve hem konuşmacı hem metin bağımsız sınıflandırmada ise %80.2 olarak hesaplanmıştır [2],[21].

2018 yılında Koolagudi ve arkadaşları, Telugu Filmlerinden oluşturdukları ses sinyalleri veri seti üzerinde duygu tanıma çalışması yapmışlardır ve çalışmada spektral ve prozodik özellikler beraber değerlendirilmiştir. Prozodik özellikler için minimum, maksimum, ortalama ve standart sapma değerleri hesaplanmış ve MFCC öznitelik çıkarıcısıyla birleştirilmiştir. Çalışmada kızgın, korku, mutlu, normal ve üzgün olmak üzere 5 duygu durumu ele alınmıştır. Çalışmada GKM, k-Ortalama kümeleme ve Vektör Nicelmesi (VQ) ve YSA yöntemleri sınıflandırma işlemi için kullanılmıştır. Sonuç olarak GKM yöntemi en iyi sonucu vermiştir [2],[22].

Araştırmacılar tarafından Kısa zamanlı Fourier dönüşümü (STFT) alternatifi olarak ortaya çıkartılan dalgacık dönüşümü (CWT) duygu sınıflama alanında sıkça başvurulmuş özellik çıkartma metotlarındandır. Ray ve Chan tarafından 2001 yılında yapılan bir çalışmada CWT ile elde edilen özellikler GA ile sınıflandırılmıştır. Ray ve Chan (2001) ile Go ve ark. (2003) ise yüz görüntülerinden ve konuşma sinyallerinden oluşturdukları bir veri seti üzerinde ayrık dalgacık tabanlı çoklu çözünürlük analizi kullanan bir duygu sınıflama metodu geliştirmişlerdir [2],[23],[24].

2020 yılında Reddy ve Vijayarajan, EmoDB ve Telugu veri setleri üzerinde duygu tanıma amaçlı bir çalışma hazırlamışlardır. Öznitelik çıkartıcı olarak MFCC, LPC, DWT ve Mel Frekanslı Ayrık Dalgacık Dönüşümü (MFDWT) kullanan yazarlar sınıflama için DVM ve kNN yöntemlerini kullanmışlardır. Sonuç olarak maksimum doğruluğu kızgın duygu durumunda %91.38 başarımla elde etmişlerdir [2],[26].

Ses sinyallerinden elde edilebilecek nitelikler oldukça geniştir ve duygu sınıflamada akustik ve spektral özelliklerle birlikte içerikte önemli bir rol oynamaktadır. 2010 yılında Tawari ve Trivedi çalışmasında pitch ve yoğunluk metrikleri için kepsrum öznitelik çıkartma yöntemini kullanmıştır. İki farklı veri seti üzerinden yürütülen çalışmada DVM sınıflandırıcı ile maksimum %84 doğruluğa ulaşılmıştır [3],[7].

2010 yılında Paleari ve Chellali video ve ses verilerinden duygu sınıflama amacıyla hibrit bir model kullanmışlardır. Ses sinyalini sınıflayabilmek için enerji, perde, titreşim, doğrusal öngörü katsayıları ve Mel-frekans kepsral katsayıları (MFCC) gibi öznitelik çıkartıcılar kullanılmıştır. Sınıflama yöntemi olarak Yapay Sinir Ağları (YSA) kullanılan modelde MFCC özellik çıkartısı ile öfke ve kızgınlık duygu durumlarının %90 civarında başarımla tahmin edilebildiği görülmüştür [3],[28].

Sesin hem spektral hem de prosodik özelliklerini kullanan bir çalışmada Xie ve Guan (2012) duygu sınıflama için öznitelikleri birleştirerek Çekirdek Entropi Bileşen Analizi (ÇEBA) yöntemini önermişlerdir. ÇEBA yöntemi, orijinal öznitelik vektörünün özelliklerini koruyan daha az boyutlu bir alt veri kümesi oluşturarak, verilerin boyutları arttıkça ortaya çıkan boyutluluk problemini çözmüştür. Çalışmada SMM sınıflandırıcı kullanılmıştır ve model, gürültülü veriye rağmen yüksek bir başarımla elde etmiştir [3],[29].

Ses sinyallerinden duygu sınıflama konulu yapılan bir çalışmada 2019 yılında Zhijie Han ve arkadaşları duygunun karmaşık doğası sebebiyle duygu tanımda zorlukların olduğunu belirtmektedir. Yazarlar bu çalışmada transfer öğrenmeyi, yöntemlerini açıklamaktadır ve bazı uygulamalar gerçekleştirmektedir. Yazarlar çalışmada transfer öğrenmenin özellik ve parametre tabanlı olarak iki farklı yönteminin olduğunu söylemektedir. Makalede yazarlar derin öğrenmeyi, çoklu görev öğrenmesi ve ince ayar olmak üzere iki başlık altında açıklamışlardır.

Makalede sonuç olarak derin öğrenme ve transfer öğrenmenin birlikte kullanılmasını başarıyı arttıracığını belirtilmektedir [32].

Ses sinyallerinden duygu sınıflama üzerine 2021 yılında Cristina Luna-Jiménez ve arkadaşları tarafından bir çalışma hazırlanmıştır. Yazarlar duygu tanımının birçok alanda uygulanabilirliğinin olması nedeniyle araştırmacılar tarafından ilgi gördüğünü belirtmektedir. Bu çalışmada, ses sinyalleri ve görüntü verileri içeren RAVDESS veri setini kullanmıştır. Yazarlar çalışmada ses sinyalleri ve görüntü verilerini birlikte işleyen bir hibrit model önerilmektedir. Ses sinyallerinin sınıflandığı bölümde en iyi sonuç PANN'lerin CNN-14'ünün ince ayarlaması neticesinde elde edilmiştir. Görüntü verilerinin sınıflandırıldığı bölümde ise önceden eğitilmiş Uzamsal Dönüştürücü Ağ önerilmektedir. Çalışmada transfer öğrenmesinde hedef ve kaynak model arasında uyumsuzluğun olabileceği ve bu problemlerin giderilmesi için yeni yöntemler ihtiyacı ortaya konulmuştur. Sonuç olarak yazarlar bu çalışmada ses sinyallerinden ve görüntü verilerinden elde edilen bilgileri birleştirerek veri seti üzerindeki 8 farklı duygu durumunu %80.08 başarı oranıyla sınıflamışlardır [33].

Ses sinyallerinden öznitelik çıkartma yöntemleri arasındaki en popüler yöntemlerden birisi olan Mel-Frekans Kepstral Katsayıları (MFCC) yönteminin farklı parametre değerlerinden nasıl etkilenebileceği 2001 yılında Zheng Fang ve arkadaşları tarafından araştırılmıştır. Bu çalışmada yazarlar optimum performans koşullarını bulabilmek için filtrelerin şekli, sayısı ve yerleşimi gibi faktörler üzerinde çeşitli testler gerçekleştirmiştir. Çalışma veri seti olarak “863” veri seti kullanılmıştır. Ayrıca yazarlar çalışmada Frekans Bant Enerjisi Mel-Frekans Kepstral Katsayıları (FBE-MFCC) yöntemi önermiştir. Önerilen yöntemin sınıflamadaki hata oranını %10 oranında düşürdüğü tespit edilmiştir [34].

MFCC'nin DWT (Dinamik Zaman Çarpıtması) ile beraber kullanımını konu alan bir çalışma 2010 yılında Lindasalwa Muda ve arkadaşları tarafından hazırlanmıştır. Bu çalışmada ses sinyallerinin özelliklerinden bahsedilmiştir. Yazarlara göre ses sinyallerinin çok fazla bilgi taşıyabilmesi analiz edilmesini zorlaştıran faktörlerden birisidir. Bu zorluklara çözüm için öznitelik çıkartma ve özellik eşleştirme gibi sinyal analiz metotlarına başvurulmalıdır. MFCC ve DWT'nin başarılı bir şekilde birlikte kullanılabileceği gösterilmiştir [35].

Wei Han ve arkadaşları 2006 yılında hazırladıkları bir çalışmada MFCC özniteliklerini çıkartmak için yüksek verimle hesaplama maliyetinde düşüş sağlayan metot önermişlerdir.

Çalışmada yazarlar önerdikleri metodun hesaplama maliyetinde %53'lük bir düşüş sağlarken model başarısında yalnızca %1.5'lik bir düşüş sağladığı belirtilmektedir [36].

K. V. Krishna Kishore ve arkadaşları 2013 yılında ses sinyallerinden duygu tanıma konulu bir çalışma yayımlamışlardır. Çalışmada ses sinyallerinden özellik çıkartımı için Mel-frekans Kepstral Katsayısı (MFCC) ve Altbant tabanlı Kepstral Parametre (SBC) yöntemleri kullanılmıştır. Makalede ses sinyalleri için SAVEE veri seti, sınıflandırıcı olarak ise Gaussian Karışım Modeli (GMM) kullanılmıştır. Sonuç olarak ses sinyallerini sınıflandırmada Altbant tabanlı Kepstral Parametre (SBC) yöntemi %70, Mel-frekans Kepstral Katsayısı (MFCC) yöntemi %51 başarımlı oranı göstermiştir [37].



### 3. TEMEL TANIM, KAVRAM VE YÖNTEMLER

#### 3.1 Duygu ve ses sinyali sınıflandırma

Duyguların karar vermede ve yaşamımızı sürdürmede temel işlevi vardır. İnsan akıl yürütme ve duygular gibi etmenlerinde tesiri altındadır. Aynı fikirlere sahip iki insan farklı duyguları hissettiği sürece farklı davranacaklardır [48],[49],[50],[51],[52]. Bu durum duyguların insanları aksiyon almak için motive ettiği ve çoğu zaman insanın davranış şekillerini belirlediğini belirtir [48],[53]. Duyguları saat piliyle kıyaslayan araştırmacılar bir saatin pil olmadan çalışmadığı gibi duyguların etkilemediği insanda temel görevlerini yerine getirecek enerjiden mahrumdur [48],[54].

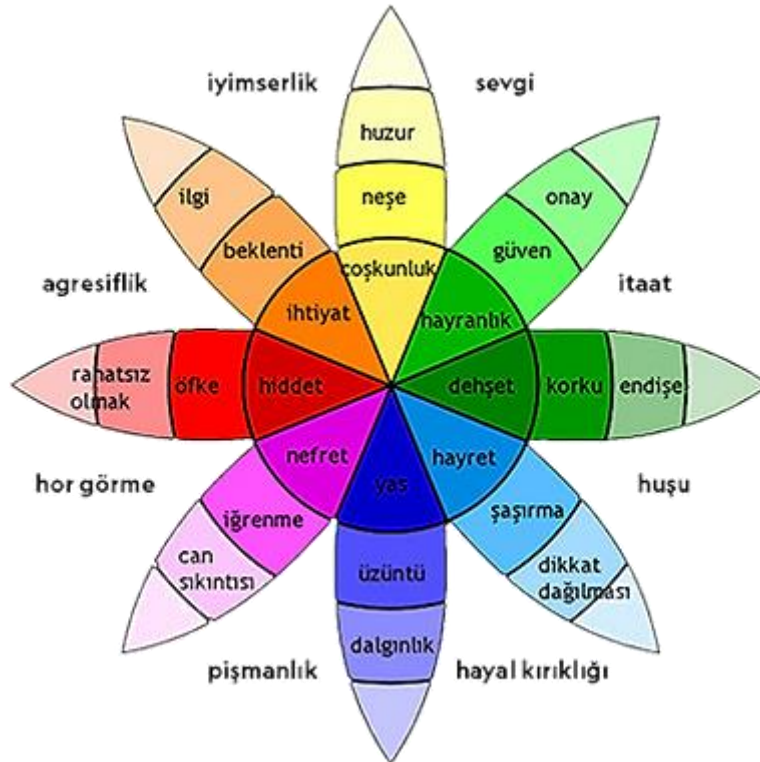
Duygular uyum sağlamaya odaklıdır. Bu durum duyguların organizmaya değerlendirme süreçleri aracılığıyla ilişkili olduğunu gösterir. Duygular insanın koruyucu mekanizmasında rol alır ve yapması veya yapmaması gereken davranışlarda belirleyicidir [48],[55]. Duyguların, davranışlar ile doğal bir bağlantısının olduğu varsayılmaktadır [48],[56].

Charles Darwin (1872) duyguların biyolojik kökenli olduğunu ve tüm insanlık kültürleri arasında evrensel olduğunu öne sürmüştür [3],[57]. Ekman'dan önce bilim camiası duyguların kültürel etkilerle şekillendiğini düşünmekteydi. Ekman 1974 yılında kültürel olarak minimum düzeyde değişime uğramış olan Papua Yeni Gine'de gerçekleştirdiği çalışmalar sonucu tüm insanlık için ortak duygu tezahürlerini 6 farklı kategoriye ayırmıştır [3],[58]. Ekman, yaptığı çalışmalarla yalnızca yüz mimiklerine dayanarak tespit edilemeyen pozitif ve negatif duyguları da ekleyerek duygu durumlarını yenilemiştir [3],[59]. Tablo 3.1'de Ekman'ın duygu kategorizasyonu tablosu gösterilmektedir.

| Ana duygular | Ana duygu önerileri |
|--------------|---------------------|
| Hiddet       | Keyif               |
| Tiksinme     | Küçümseme           |
| Endişe       | Memnuniyet          |
| Neşe         | Hicap               |
| Keder        | Coşku               |
| Hayret       | Pişmanlık           |
|              | Övünç               |
|              | Sükunet             |
|              | Tatmin              |
|              | Haz                 |
|              | Sıkılganlık         |

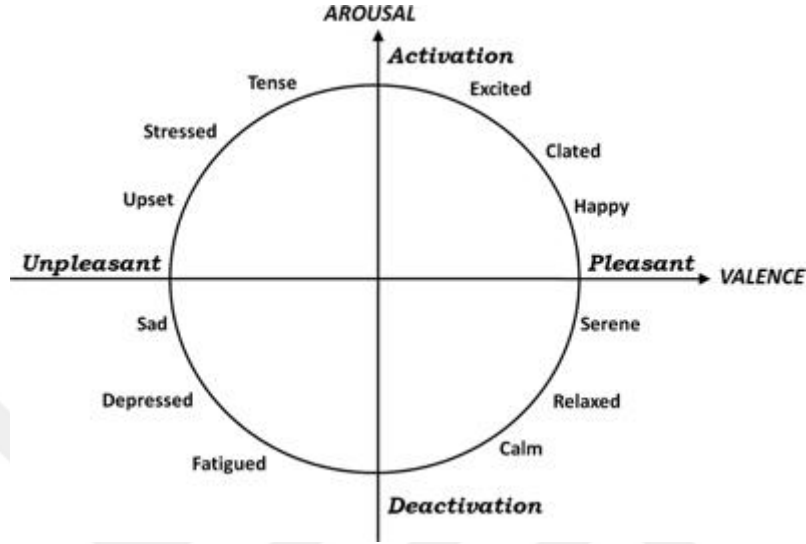
Tablo 3.1: Ekman'ın duygu kategorizasyonu

Plutchik 1980 yılında, duygulara ilişkin yeni bir kavram önerdi [60]. Duygular çarkı adını verdiği bu kavramda farklı duygular birleştirilerek yeni duygular ortaya çıkar. Bu yaklaşımda duygular, yoğunluk bakımından kendi içinde sıralanır; coşku, neşe ve huzur gibi. Yakın konumdaki duygular birbirine benzerken karşıt konumdaki duygular ise birbirlerinin zıttıdır. Şekil 3.1'de Plutchik'in duygular çarkı gösterilmektedir [3].



Şekil 3.1 Plutchik'in duygular çarkı

Duygu tanımlanmasında değerli bir çalışma olan Russel (1990) duygu koordinat sistemi, 3 özelliğe ayrılır. Birincini özellik değer, ikinci özellik uyarılma, üçüncü ve son özellik baskınlıktır [3],[61]. Şekil 3.2’de Russel’in çember modeli gösterilmektedir.



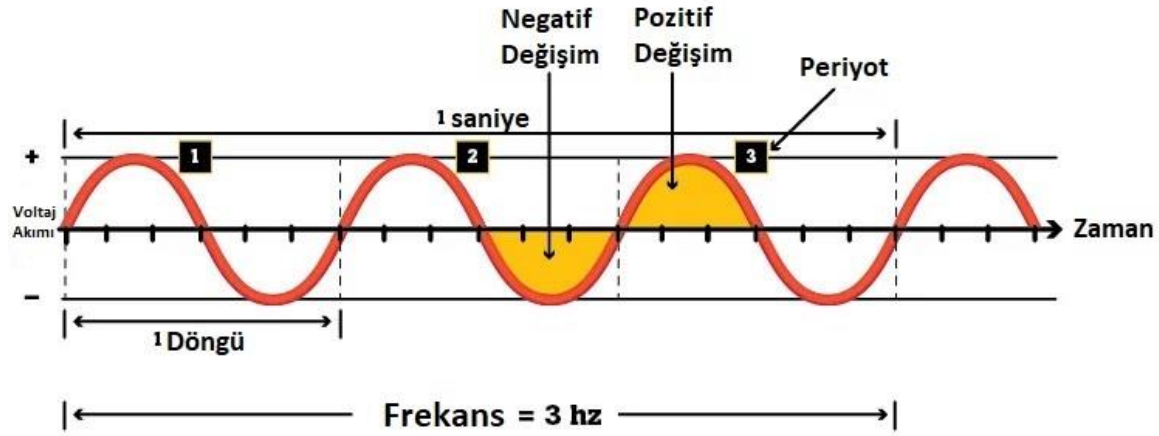
Şekil 3.2 Russel'in Çember Modeli

### 3.2 Ses ve Ses Sınıflandırması İçin Kullanılan Yöntemler

Ses havada titreşerek yayılır ve kulağa ulaştığında kulakta fiziksel bir baskı üretir. Kulaktaki özelleşmiş yapı sayesinde bu baskı elektrik sinyallerine dönüştürülür ve sinir hücreleri aracılığıyla beynin ilgili bölümüne iletilir. Beynin ilgili bölümünde bu ses sinyalleri işlenir ve böylece ses anlam kazanır. Sesin bilgisayar sistemlerinde kullanılabilmesi için kulak ile benzer çalışma prensibine sahip mikrofona benzeri araçlar kullanılır. Ses sinyali, sesin elektromanyetik biçimidir ve iki farklı biçimi vardır: analog ve dijital. Analog ses sinyali, sesin tüm niteliklerini koruyarak kopyalanan formu temsil eder. Öte yandan, dijital ses sinyali, belirli zaman aralıklarında örneklenen analog sinyalin, sadece 1 ve 0'lardan oluşan sayısal sisteme dönüştürülerek elde edilen biçimidir. Bu şekilde, ses verisi daha uygun bir şekilde işlenip saklanabilir. Ses dalgalarının frekans, dalga boyu, periyot ve hız gibi özellikleri bulunur [3], [62].

Ses dalgalarının belirli bir zaman diliminde ne kadar sıklıkla titreştiği, frekans olarak adlandırılır ve bu değer Hertz (Hz) cinsinden ölçülür. İnsan kulağının algılama yeteneği, yaklaşık olarak 20 Hz'den başlayarak 20 kHz'e kadar uzanan bir frekans aralığına sahiptir

[3],[62]. Bir tam dalganın oluşması için gereken zaman dilimine periyot denir ve genellikle saniye cinsinden ölçülür. Şekil 3.3’de ses sinyali grafiği gösterilmektedir [88].



Şekil 3.3: Ses Sinyali Grafiği

1940’lı yıllarda bilim insanları, insanın karar alma sürene dair yaptıkları araştırmalarda topla ve tetikle nöron modelini ortaya koymuşlardır. Bu ilerleme, yapay zekâ araştırmalarının 1950’li yıllarda başlatılmasını sağlamıştır [3].

Öğrenme insan için geçmişten elde edilen bilgileri geleceğe yönelik kullanabilme yeteneğidir. Bu durum makine öğrenmesi içinde benzerdir. Eğer makine deneyimlerinden elde ettiği bilgilerle performansını artırabilirse, başarılı bir makine öğrenimi sürecinden bahsedilebilir. Makine öğrenmesinin ana hedefi, elimizdeki bilgiyi kullanarak yeni bilgiler türetmek, bu bilgileri makinelerin davranışlarına uyarlamak, yeni verilerin işlenmesi ve tahminler yapılması süreçlerine dayanmaktadır. Bu süreçler birbirleriyle ilişkili ve iç içe geçmiş olan veri madenciliği, yapay zekâ, örüntü tanıma ve istatistik gibi çeşitli alanlarla alakalıdır. Bu disiplinlerin her biri, makine öğreniminin temel yapısına katkı sağlar ve onun daha gelişmiş hale gelmesini destekler [3].

Makine öğrenme yöntemleri sınıflama, kümeleme, regresyon, öznitelik çıkarımı, ilişki tespitini içerir. Sınıflama, verileri belirli niteliklere dayanarak önceden tanımlanmış sınıflara ayırma işlemidir. Kümeleme, sınıflarının önceden etiketlenmediği verilerde, benzer niteliklere sahip olanların kümelerini ortaya çıkarma sürecidir. Regresyon, hedef değişken ile bağımlı değişken arasındaki ilişkiyi tahmin etmeye yarayan bir dizi istatistiksel yöntemdir ve iki amaçla

kullanılmaktadır: Bağımlı değişken ile bağımsız değişkenler arasındaki ilişkinin gücünü ölçmek ve bu ilişkiye göre tahminler yapmak [3].

Öznitelik çıkarımı için verinin çok sayıda özelliği olsa da hangi özelliklerin verinin kümesine, sınıfına veya değerine etki ettiği bilinmeyebilir. Bu durumda öznitelik çıkarımı, mevcut özelliklerin bir alt kümesini belirleme ya da mevcut niteliklerin kombinasyonlarından yeni nitelikler elde etme işlemi olarak söylenebilir. İlişki tespiti, veri kümeleri arasındaki ilişkileri bulma sürecidir ve gerçek hayatta özellikle müşterilerin hangi ürünleri beraber satın aldığına yönelik tahminlerde kullanılır [3].

Duygu tanıma çalışmalarında, yaygın olarak literatürde sıkça başvurulanan makine öğrenimi yöntemleri şöyledir [3][7][27][63][64][65]:

- Yapay Sinir Ağları
- Genetik Algoritmalar
- Gizli Markov Modelleri
- Bayes Öğrenme ve Bayes Ağları
- Karar Ağaçları ve Karar Ormanları
- Kümeleme
- Destek Vektör Makineleri

Makine öğrenmesi türleri denetimli, denetimsiz ve pekiştirmeli öğrenme olmak üzere üçe ayrılır. Bu tezde denetimli ve pekiştirmeli öğrenme türleri kullanılmıştır. Denetimli öğrenme, verilerin ve sonuçlarının (sınıflarının) bilindiği durumlarda bu verileri ve sonuçlarını içeren bir model oluşturmayı hedefleyen makine öğrenimi yöntemidir. Etiketli eğitim verileri sisteme sağlanır ve makine, gelecekteki bilinmeyen veriler için doğru tahminlerde bulunabilecek bir model oluşturarak bağımsız bir şekilde öğrenir. Makine, beklenen çıktıya en yakın sonucu sağlayacak en uygun işlevi bulmaya çalışır. Makine en yüksek doğrulukla modeli bulmaya çalışır [3].

Eğitim aşamasından sonra, makinenin geliştirdiği model test verileri ile sınanır. Test sonuçlarından, doğru olarak belirlenmiş sonuçların, toplam test sayısına oranı doğruluk oranıdır ve bu aynı zamanda modelin başarı oranını gösterir gösterir [3]. Sınıflandırma, gözetimli öğrenme yöntemlerinden biridir. Denetimli öğrenmede hata sisteme öğrenme olarak geri

beslenir ve hata bilgisini sistemin öğrenme sürecine dahil eder böylece denetimsiz öğrenmeye kıyasla eğitimi daha az süre sürer [3].

Etiketlenmemiş verilerle çalışan denetimsiz öğrenme, hedef değerleri ve etiketleri bilinmeyen veri kümelerini kullanır. Amaç, veri kümesi içindeki gizli yapıları ve bağlantıları keşfetmektir. Ancak, her zaman veriler arasında belirgin bir ilişki bulunmayabilir ve bu nedenle elde edilen sonuçların tam olarak doğru olması beklenmemelidir [3].

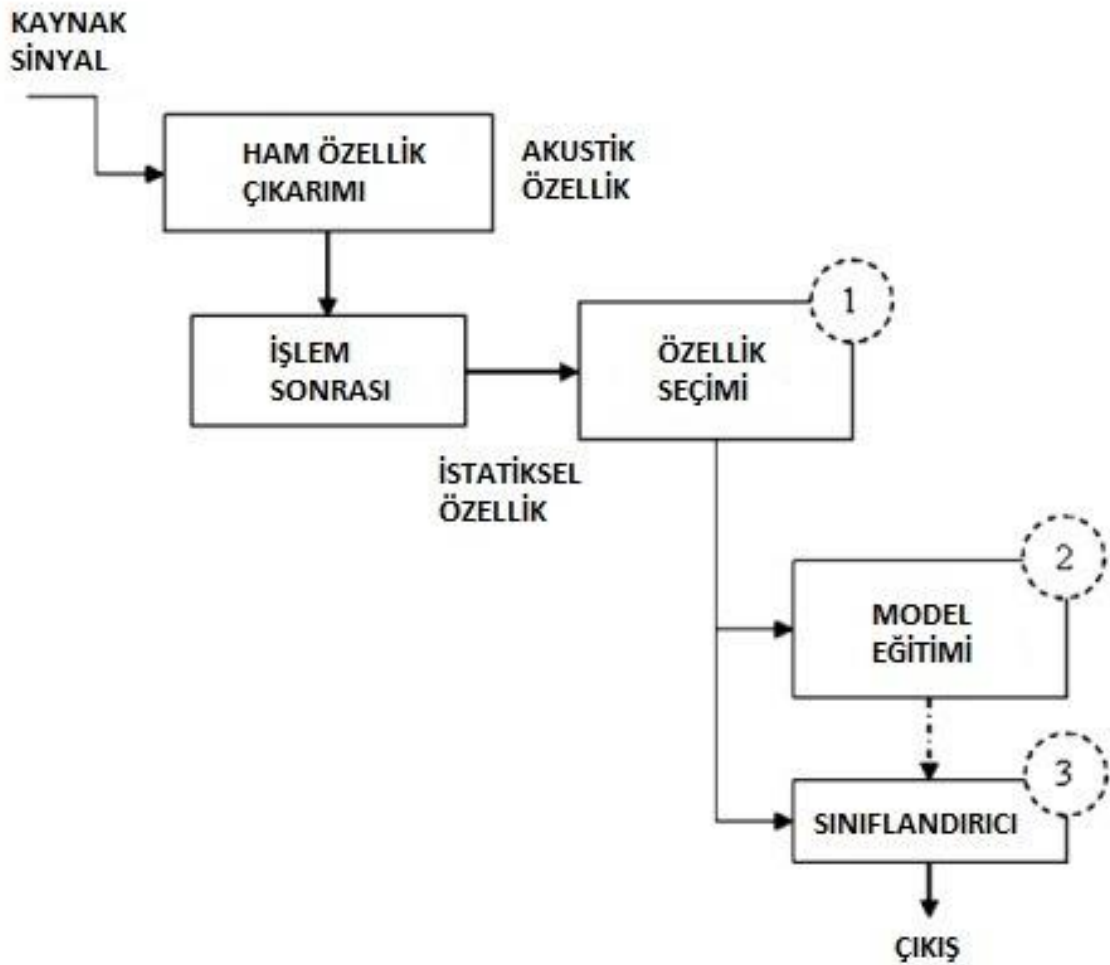
Pekiştirmeli öğrenme, bir yapay zekâ modelinin, deneyimlerinden öğrenerek ve eylemleri sonucu elde ettiği geri bildirimlerle performansını artırma sürecidir. Bu öğrenme türünde, model belirli bir görevi yerine getirirken ödül ve ceza mekanizması aracılığıyla doğru ve yanlış eylemleri ayırt etmeye çalışır. Model, zaman içinde pozitif geri bildirimlerle ödüllendirilen eylemleri tekrarlamayı öğrenirken, ceza ile sonuçlanan eylemlerden kaçınma eğiliminde olur. Pekiştirmeli öğrenme, özellikle karar verme ve etkileşimli ortamlarda başarılı bir şekilde kullanılmaktadır [89].

### **3.3 Duygu Sınıflama Görevi**

Duygu tanıma alanında ses sinyallerini analiz etmek için çeşitli yöntemler bulunsa da literatürde sıklıkla makine öğrenme problemlerine dayalı sınıflandırma yöntemleri tercih edilmektedir. Bu tür bir problemde genel yaklaşım, duyguları bilinen ses sinyallerini eğitim ve test verisi olarak ikiye ayırmaktır. Eğitim verisinden konuşmacıya özgü bilgiler, farklı öznitelik çıkarım yöntemleri kullanarak elde edilir ve duygular etiketlenir. Sistem, öğrenme algoritması ile eğitim verileri üzerinde çalıştırılır ve örnek bir model üretilir. Bu model, test verisindeki etiketli duygular üzerinde denetlenerek başarı ölçülür [3].

Başarılı bulunmayan modellerde, öznitelikler veya öğrenme algoritmaları değiştirilebilir. Başarılı bir model oluşturulduğunda, duygusu bilinmeyen ses sinyalleri sisteme verilir ve doğru duygu tanımlaması elde edilmesi beklenir. Tanımlamanın doğruluğunu etkileyen birçok faktör bulunmaktadır; özellikle, eğitim ve test verisinin benzer özelliklere sahip olması ve problem uzayını temsil edebilmesi önemlidir. Aksi halde, model daha önce karşılaşmadığı ses sinyallerini doğru bir şekilde tanımlayamaz. Bu durumda, eğitim verisine yeni örnekler eklenerek model geliştirilebilir [3].

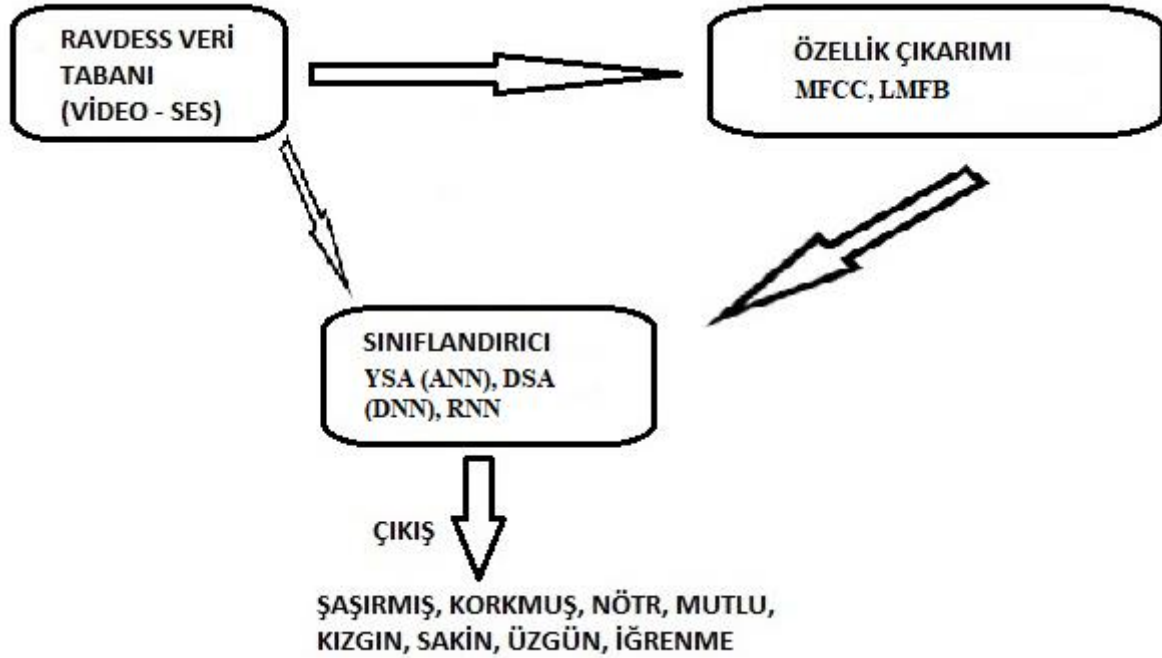
Eğitim verisinin yetersiz olduğu durumlar da benzer sorunlara yol açabilir. Başka bir potansiyel sorun, boyutluluk sorunu olarak bilinir; öznitelik vektörlerinin sayısı ve veri boyutlarının artması, veri uzayının hacminin üstel olarak büyümesine neden olur. Boyutluluk sorununun yaşandığı durumlarda, orijinal uzayı dönüştürülmüş bir uzaya taşıyarak, orijinal özniteliklerin özelliklerini koruyan daha küçük bir veri kümesi oluşturulabilir. Bu yaklaşımla, hem alakasız verilerin (gürültü) ayıklanması hem de öğrenme algoritmasının işlem süresinin kısaltılması mümkündür. Duygu tanıma sistemlerinin örnek bir blok şeması, şekil 3.4’de sunulmuştur [3].



Şekil 3.4: Tipik Bir Duygu Tespit Yaklaşımı

Bu tez çalışmasında veri seti olarak RAVDESS veri seti kullanılmıştır. Duygu tanıma için ses sinyallerini kullanırken izlenecek temel süreçler özellik çıkarımı ve sınıflandırma aşamalarını içerir. Şekil 3.5’de, ses sinyallerinden duygu durumunu belirlemek için gerekli

adımlar sunulmaktadır. Ses sinyalinin özellikler elde edildikten sonra sınıflandırıcı algoritma ile işleme tabi tutulur. Sonuç olarak, sınıflandırma yöntemiyle verinin hangi duygu kategorisine dahil olduğu saptanır.

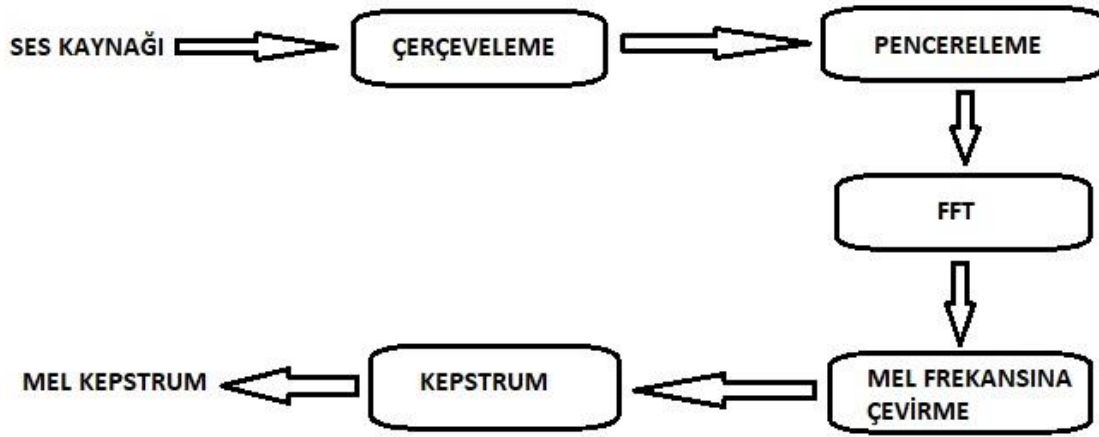


Şekil 3.5: Duygu Sınıflama Aşamaları

### 3.4 Özellik Çıkarım Yöntemleri

#### 3.4.1 Mel Frekansı Kepstrum Katsayıları (MFCC – MFKK)

Mel Frekansı Kepstrum Katsayıları (MFCC), ses sinyalinin duygu tanıma çalışmalarında özellik çıkarımı için yaygın olarak kullanılan yöntemlerdendir. MFCC'nin başarısı, insanın işitme frekans aralığını benzetme yeteneğinden kaynaklanır, bu sayede daha doğru analizler sağlar. Ayrıca MFCC, ses sinyalleri yapısındaki değişikliklere ve dalgalanmalara karşı oldukça dirençlidir. Şekil 3.6'de Reynolds tarafından önerilen MFCC çıkarma yöntemi, bir dizi adımı takip ederek ses sinyallerinden bu özellikleri elde etmeyi amaçlar. Bu süreç, ses analizi ve duygu tanıma için sağlam bir temel sağlar [1],[67].

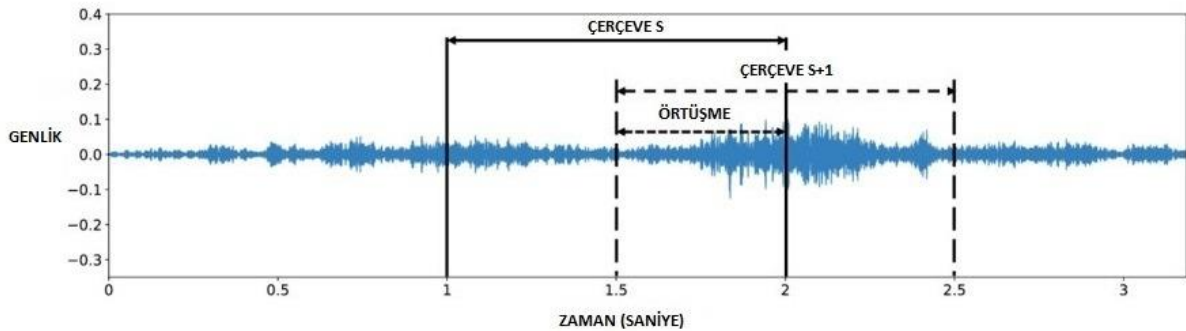


Şekil 3.6: MFCC Adımları

### 3.4.1.1 Çerçeveleme

Ses sinyallerinin özellikleri küçük zaman dilimlerinde değişmediğinden analiz bu küçük zaman dilimlerinde gerçekleştirilmelidir ve ses sinyali, parametrelerin değişmediği küçük zaman parçalarına bölünmelidir [1][68]. Bu parçalara çerçeve denir. Araştırmacıların deneyimleri neticesinde en optimum çerçeve uzunluğunun 20 ile 30 milisaniye arasında olduğu ortaya çıkmıştır [1][69]. Ses sinyalinin tamamını kapsayacak şekilde FFT hesaplamak veride spektral bilgisinin zayıflamasına neden olur. Bu yüzden ses sinyalinin tamamını kapsayacak şekilde FFT hesaplanması yerine her çerçevenin ayrı ayrı FFT'si hesaplanmalıdır [1][70].

Çerçeveler, bir önceki çerçevenin bir kısmını içerisinde bulundurur. Çerçeveleme süresince birbirlerinin bir bölümünü kapsayan çerçevelerin kullanılmasının amacı çerçeveler arası geçişlerde meydana gelebilecek kesinti etkisini minimize etmek ve çerçeve geçiş sürecini pürüzsüz hale getirmektir [71][72][73][74]. Şekil 3.7'de çerçeveleme yöntemine ait bir şekil gösterilmektedir [75].



Şekil 3.7: Çerçeveleme Yöntemi

### 3.4.1.2 Pencereleme

MFCC özelliklerini çıkartmak için ikinci adımda çerçeveler pencereleme işlemine tabi tutulur. Bu adımın amacı çerçevelerin başındaki ve sonunda ki kesintili bölümü iyileştirerek spektral bozulmayı minimize etmektir [1][76]. Sıklıkla tercih edilen pencereleme fonksiyonları şunlardır [77]:

- Hanning
- Dikdörtgen
- Hamming
- Welch
- Bartlett
- Blackman-Harris
- Nuttall Pencereleeri

Bu çalışmada da sıkça tercih edilen Hanning pencereleme fonksiyonu kullanılmıştır ve fonksiyona ait denklem (1) de gösterilmiştir [77].

$$w(j) = 1/2 * (1 - \cos(2 * \pi * j/(N))) \quad (1)$$

### 3.4.1.3 FFT

Bu adımda zaman boyutunda bulunan N adet örnekten oluşan her bir çerçeveye Hızlı Fourier Dönüşümü (Fast Fourier Transform – FFT) uygulanır. Burada amaç veriyi zaman boyutundan frekans boyutuna taşımak ve bir spektrogram elde etmektir [78]. Denklemi (2)’deki gibidir [78].

$$x(k) = \sum_{j=0}^{N-1} X(j)e^{i2\pi jk/N} \quad (2)$$

### 3.4.1.4 Mel Frekansı Dönüşümü

İnsan kulağının ses sinyallerini lineer bir şekilde algılamadığı bilinmektedir. Mel ölçeğine göre 1 kHz’ye kadar lineer 1 kHz üzerindeki değerler için logaritmiktir. Frekans ve mel ölçeği arasındaki denklem x’de verilmiştir [1].

$$mel(f) = 295 \log(1 + f/700) \quad (3)$$

### 3.4.1.5 Kepsturum Katsayıları

Sinyalin dalga boyu spektrumu, mel ölçeğine göre eşit aralıklarla dağıtılıp %50 kesişen N adet üçgen bant geçiren filtre ile çarpmaya tabi tutulur. Bu işlem, her filtredeki enerjinin logaritmasını elde etmek için yapılır. Mel spektrum katsayılarının logaritması, gerçel sayılar olduklarından, (4) ayrık kosinüs dönüşümü uygulanarak zaman bölgesine geri dönülür. Bu süreç sonucunda elde edilen katsayılar, mel frekanslı kepsral katsayılar (MFCC) olarak adlandırılır [3].

$$C_n = \sum_{k=1}^K (\log S_k) \cos \left[ n \left( k - \frac{1}{2} \right) \frac{\pi}{K} \right] \quad (4)$$

Denklemden  $S_k$ ,  $k = 1, 2, 3, 4, \dots, K$  mel spektrum katsayılarıdır.

### 3.4.2 Log Mel Filter Bank (LMFB)

Log Mel Filter Bank aslında MFCC'nin sadece ayrık kosinüs değişimi uygulanmamış halidir [90].

## 3.5 Sınıflandırıcılar

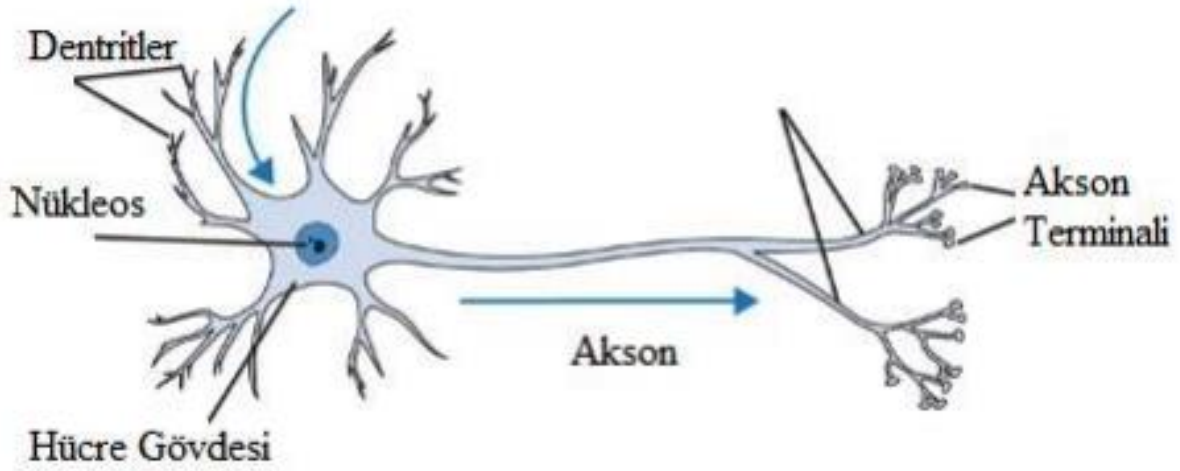
Bu tez çalışmasında sınıflandırıcı olarak Yapay Sinir Ağları, Derin Sinir Ağları ve Evrişimli Sinir Ağları kullanılmıştır.

### 3.5.1 Yapay sinir ağları

Yapay sinir ağları, insan beyninin çalışma prensiplerinden ilham alarak oluşturulmuş, birbirine bağlantılar vasıtasıyla entegre edilen ve her bir bileşenin kendi hafızasına sahip olduğu işlem birimlerinden meydana gelen eşzamanlı ve dağıtık bilgi işleme sistemleridir [2]. Başka bir ifadeyle, aksonlar aracılığıyla bilgi iletimini gerçekleştiren sinir hücreleri topluluğunun oluşturduğu sinir sistemiyle benzerlik gösteren ve genellikle uyum sağlayabilen (adaptif) basit unsurların yoğun paralel bir şekilde bir araya gelerek meydana getirdiği ağ yapıları olarak nitelendirilmiştir [2].

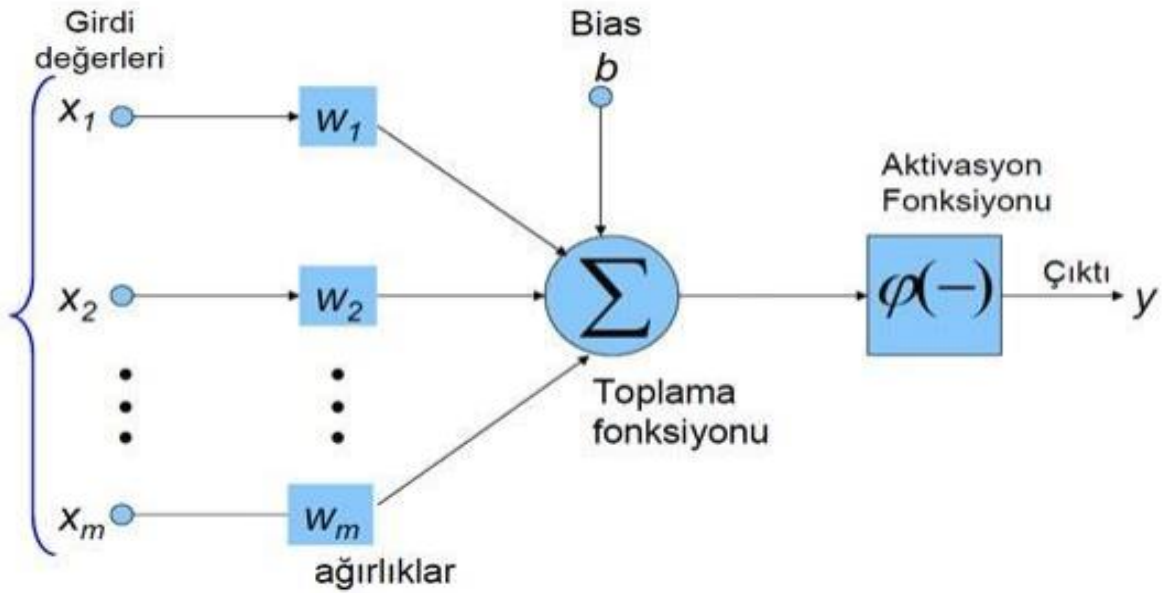
Yapay sinir ağlarının (YSA) temel bileşeni, işlem elemanı ya da düğüm olarak isimlendirilen yapay sinir hücreleridir. Biyolojik sinir hücrelerine kıyasla daha basit yapıda olan

bu yapay sinir hücreleri, biyolojik sinir hücrelerinin dört ana işlevini taklit etme yeteneğine sahiptir. Yapay bir sinir ağı düğümünün görsel temsili Şekil 3.8'de bulunmaktadır [2].



Şekil 3.8 Sinir Hücresi Yapısı

Şekil 3.9'da, girişler  $x_i$  sembolüyle temsil edilir. Her bir giriş, ağırlık  $w$  ile çarpılır ve bu çarpımların toplamı, eşik değeri olarak adlandırılan bias ile birleştirilir. Ardından, bu değer aktivasyon fonksiyonundan geçirilerek sonuç elde edilir ve  $y$  çıkışı oluşturulur [2].



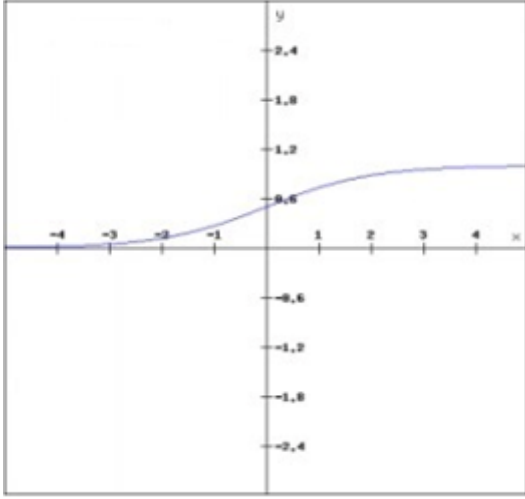
Şekil 3.9: Yapay Sinir Ağı Yapısı

Nöronun işlevini belirleyen ana faktörlerden biri aktivasyon fonksiyonudur. Biyolojik nöronlarda, gelen sinyallerin toplamı belirli bir eşik değerini geçtiğinde tepki oluşur. Yapay

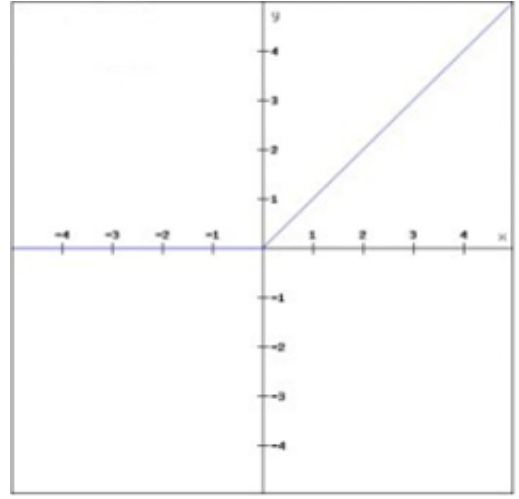
nöronlarda da benzer bir süreç, aktivasyon fonksiyonu sayesinde gerçekleştirilir. Bu fonksiyon, literatürde sıkıştırma veya eşik fonksiyonu olarak da anılır. Tablo 3.2, yaygın olarak kullanılan aktivasyon fonksiyonlarını sergilemektedir. Çözülmeye çalışılan problem türüne ve ağ yapısına göre çeşitli aktivasyon fonksiyonları tercih edilebilir [2]. Şekil 3.10, şekil 3.11, şekil 3.12, şekil 3.13, şekil 3.14, şekil 3.15, şekil 3.16, şekil 3.17, şekil 3.18 ve şekil 3.19, tablo 3.2’de ki aktivasyon fonksiyonlarının görselleştirilmiş halini göstermektedir.

| Aktivasyon Fonksiyonu    | Formülasyon                                                   |
|--------------------------|---------------------------------------------------------------|
| Sigmoid                  | $f(x) = 1/(1 + e^{-x})$                                       |
| Relu                     | $f(x) = \max(0, x)$                                           |
| Swish                    | $f(x) = x/(1 + e^{-x})$                                       |
| İkili Basamak Fonksiyonu | $f(x) = 1 \quad x \geq 0$<br>$f(x) = 0 \quad x < 0$           |
| Sızan Relu               | $f(x) = 0.01x \quad x < 0$<br>$f(x) = x \quad x \geq 0$       |
| Üstel Doğrusal Birim     | $f(x) = x \quad x \geq 0$<br>$f(x) = a(e^x - 1) \quad x < 0$  |
| Softmax                  | $\sigma(\mathbf{z})_j = \frac{e^{z_j}}{\sum_{k=1}^K e^{z_k}}$ |
| Tanh                     | $f(x) = 2\text{sigmoid}(2x) - 1$                              |
| Parametrelili Relu       | $f(x) = x \quad x \geq 0$<br>$f(x) = ax \quad x < 0$          |
| Lineer                   | $F(x) = ax$                                                   |

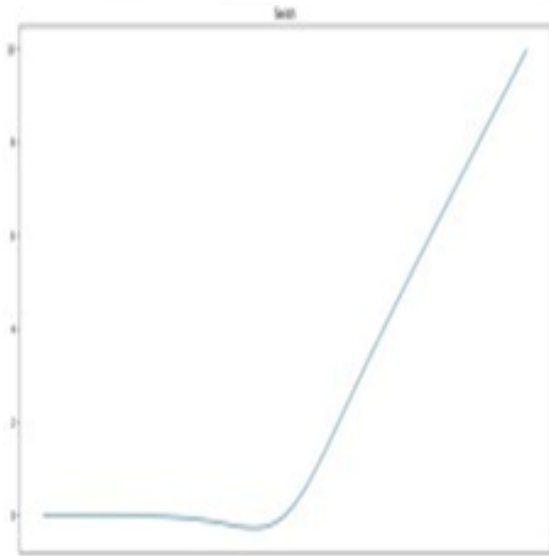
Tablo 3.2: Aktivasyon Fonksiyonları Listesi ve Formülasyonları



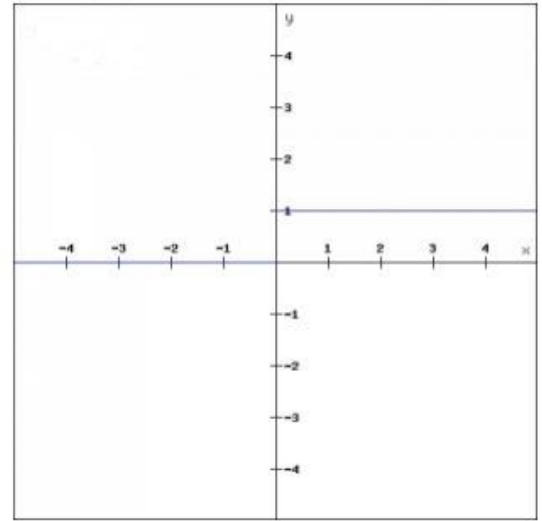
Şekil 3.10: Sigmoid Fonksiyonu



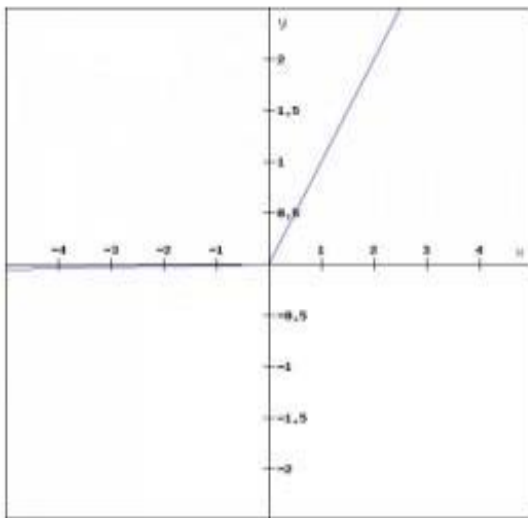
Şekil 3.11: RELU Fonksiyonu



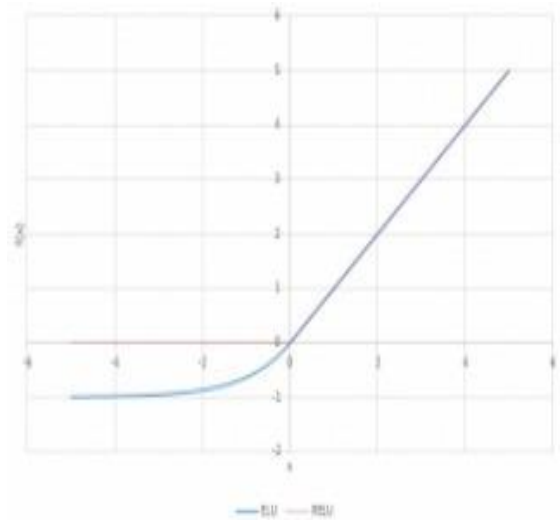
Şekil 3.12: Swish Fonksiyonu Şekil



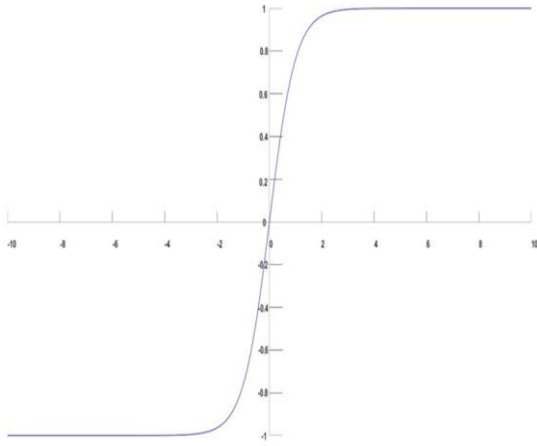
3.13: İkili Basamak Fonksiyonu



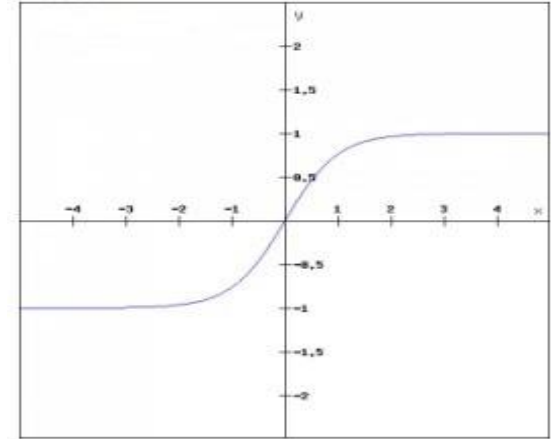
Şekil 3.14: Sızdıran RELU fonksiyonu



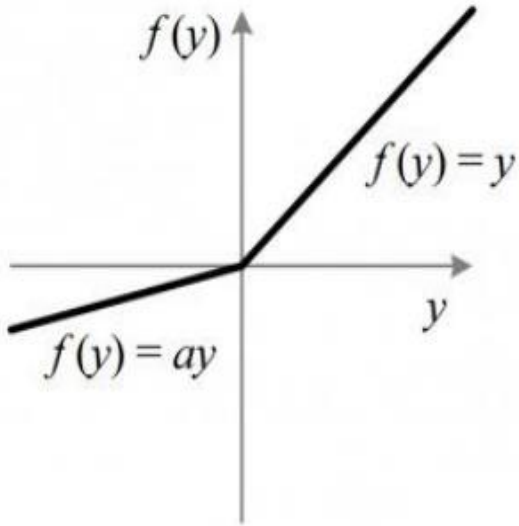
Şekil 3.15: Üslü Doğrusal Birim Fonksiyonu



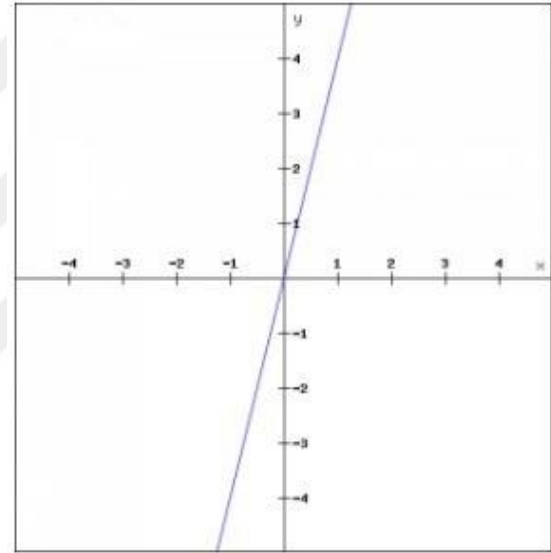
Şekil 3.16: Softmax Fonksiyonu



Şekil 3.17 Tanh Fonksiyonu



Şekil 3.18: Parametreli RELU Fonksiyonu



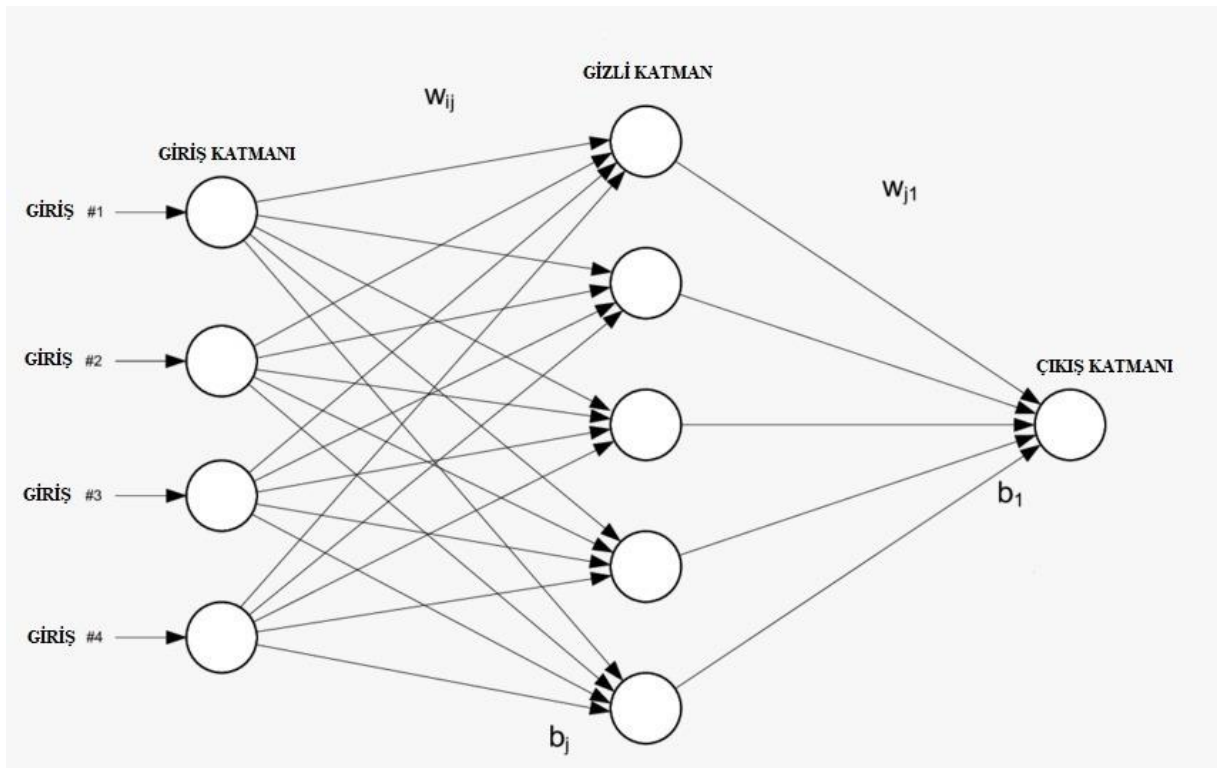
Şekil 3.19: Lineer Fonksiyon [79]

Yapay sinir ağları (YSA), klasik bilgi işleme yöntemlerinden ayrılan ve bu bağlamda daha etkili sonuçlar sağlayabilen özgün özelliklere sahiptir. Bu benzersiz nitelikler arasında eşzamanlı işlem kapasitesi, hatalara karşı tolerans, öğrenme yeteneği ve uygulama kolaylığı bulunmaktadır. Yapay sinir ağları, öğrenme ve ezberleme gibi becerilere sahip olup, veriler arasında bağlantılar kurabilen, geleneksel programlama yöntemlerine ihtiyaç duymayan öz öğrenme sistemleridir [2].

Bu tez çalışmasında, Çok Katmanlı Yapay Sinir Ağı (YSA) modeli temel alınmıştır. Lineer olmayan giriş ve çıkış verileri arasındaki ilişkiyi öğrenmek amacıyla, giriş ve çıkış katmanlarının arasına birden fazla gizli katman eklenen modellere çok katmanlı YSA denir.

Gizli katmanlardaki nöron sayısını modeli tasarlayan kişi belirler. Örnek olarak Şekil 3.20'de, tek gizli katmana sahip birçok katmanlı YSA modeli sunulmaktadır [80]. Buradaki hedef, ağa verilen girişe karşılık gelen çıkışı üretebilecek en iyi ağırlıkları bulmaktır. Bu süreç, ağın eğitimi olarak adlandırılır [2].

Ağırlıkların, çıkış hatalarına göre sürekli olarak güncellenip en uygun değerlerinin öğrenilmesine yardımcı olan eğitim algoritmalarından biri, Hata Geriye Yayılım Algoritmasıdır. Bu algoritma, modelin eğitim sürecinde önemli bir rol oynar ve çok katmanlı YSA'ların başarılı şekilde öğrenmesini sağlar [2].



Şekil 3.20: Yapay Sinir Ağı Katman Yapısı

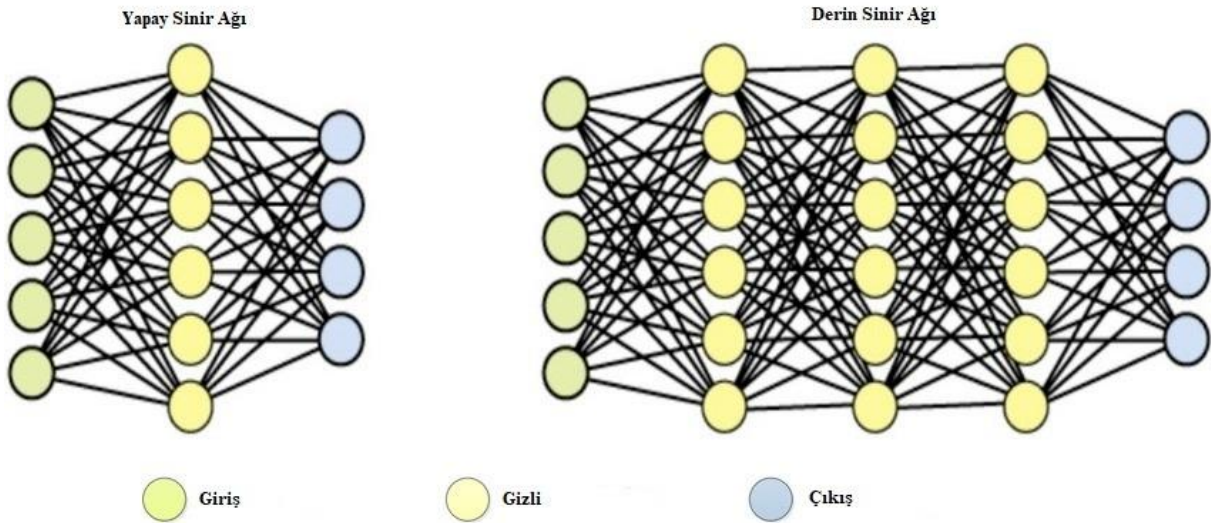
### 3.5.2 Derin öğrenme

Derin Öğrenme, yapay sinir ağları (YSA) üzerine kurulu daha karmaşık bir yapıya sahip algoritmaları ifade eder. Derin öğrenme yaklaşımı, veri özelliklerinin ve temsillerinin çoklu düzeylerini öğrenmeye dayalıdır. Hiyerarşik bir yapı içinde, üst düzey özellikler alt düzey

özelliklere dayanarak oluşturulur. Bu yapı, çeşitli soyutlama derecelerine uygun olarak farklı temsil düzeylerinin öğrenilmesini sağlar [82].

Derin Öğrenme algoritmaları, geleneksel makine öğrenimi yöntemlerinden büyük miktarda veri ve karmaşık yapıları işleyebilme kapasitesiyle ayrılır. Bu, özellikle son yıllarda görüntü işleme alanında milyonlarca etiketli verinin kullanılabilir hale gelmesiyle önem kazanmıştır. Ayrıca, ekran kartı tabanlı hesaplama gücündeki büyük gelişmeler sayesinde derin öğrenme algoritmaları popüler hale gelmiştir [81].

Derin öğrenme teknikleri, pek çok alanda başarı düzeylerini maksimize etmiştir. Tüm dünyada birçok büyük şirket bu algoritmaları kullanarak ürünlerini geliştirmekte ve aynı zamanda kendi derin öğrenme yazılım kütüphanelerini geliştirici topluluklara sunarak bu alanın hızlı ilerlemesine katkıda bulunmaktadır [81]. YSA ve DSA mimarileri örnek görsel benzetimi şekil 3.21’de gösterilmiştir [83].



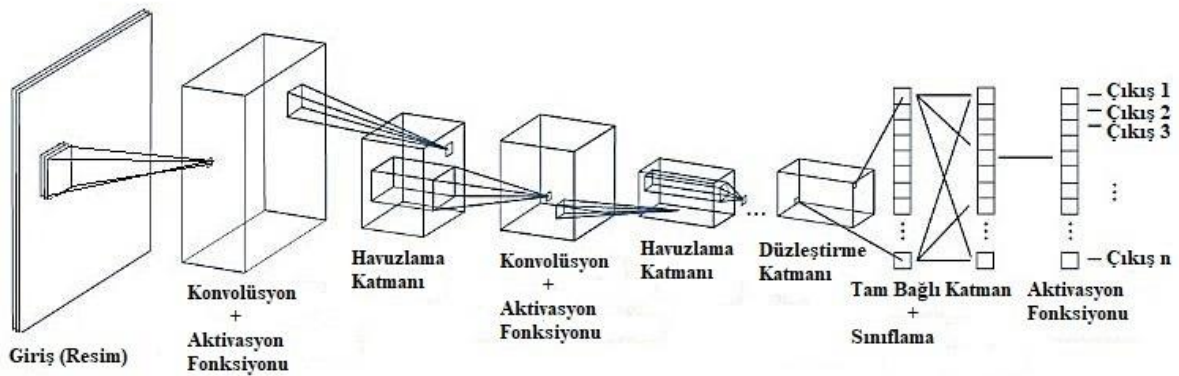
Şekil 3.21: Yapay Sinir Ağı ve Derin Sinir Ağı Yapısı

### 3.5.3 Evrimsel Sinir Ağı

Derin öğrenmenin temel yapısı olarak kabul edilen Evrimsel Sinir Ağları (ESA), Şekil x'te gösterilmektedir. Bu yapıda, başlangıçtaki konvolüsyon (evrimsel), havuzlama katmanlarından, düzleme ve son olarak tam bağlı katmandan oluşmaktadır. Kısacası, ESA'lar

sıralı olarak yerleştirilmiş eğitilebilir bölümleri içerir ve bu bölümlerin ardından bir sınıflandırıcı yer alır. ESA'larda, girdi verileri alındıktan sonra katmanlar üzerinden işlemler gerçekleştirilir ve eğitim süreci ilerler. En sonunda, üretilen sonuç ile hedeflenen sonuç karşılaştırılır ve bir hata değeri oluşur. Bu hatanın tüm ağırlıklara dağıtılabilmesi için geri yayılım algoritması kullanılır. İterasyonlarla birlikte ağırlıkların güncellenmesi gerçekleştirilir ve hata değerinin azaltılması amaçlanır [85].

ESA'lar giriş verisi olarak görüntü, ses ve video gibi farklı sinyalleri kabul edebilir. Başlangıçta görüntü sınıflandırma üzerine odaklanan ESA'lar, son zamanlarda diğer alanlarda da yaygın olarak kullanılmaktadır. Görüntü sınıflandırma süreçlerinde ESA'lar pikselleri, kenar kombinasyonlarından oluşan motiflere, bu motiflerin birleşiminden nesne parçalarına ve nesne parçalarının birleşiminden nesnelere dönüştürür [85]. Şekil 3.22'de evrişimsel sinir ağı yapısı gösterilmektedir [84].



Şekil 3.22 Evrişimsel Sinir Ağı Yapısı

### 3.5.3.1 Giriş Katmanı

Evrişimli sinir ağlarının (CNN) ilk katmanı olan giriş katmanı, belirli bir boyutta verilir ve CNN'lerin başarısı açısından görüntü boyutu büyük önem taşır. Eğer giriş görüntüsünün boyutu artırılırsa hem eğitim süreci hem de sistem performansı olumlu yönde etkilenebilir. Öte yandan, görüntü boyutunun küçük seçilmesi durumunda eğitim süreci hızlanırken, sistemin başarısı azalabilir. Büyük görüntü boyutları seçilirse, eğitim süreci uzar, fakat tasarlanan sistemin performansı yükselebilir. Bu nedenle, giriş görüntüsünün boyutunun, oluşturulan sisteme uygun şekilde seçilmesi önemlidir [84].

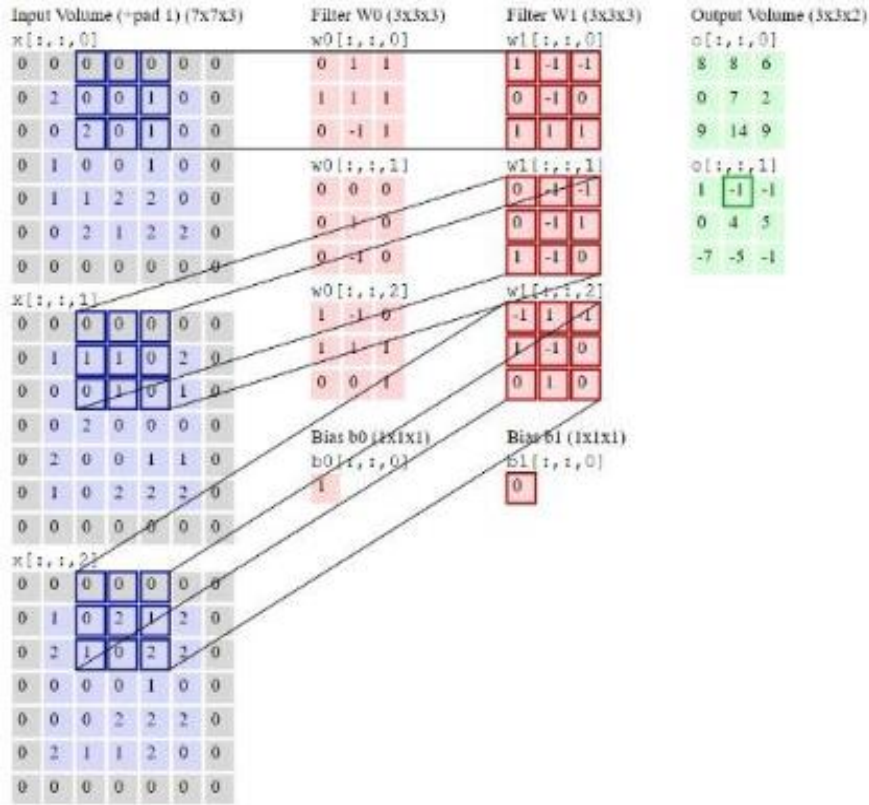
### 3.5.3.2 Konvolüsyon Katmanı

Derin öğrenme mimarilerinde önemli bir yapı olan Evrişimli Sinir Ağları (ESA) temelinde dönüşüm katmanı yer alır. Bu katman, belirli boyutlardaki filtrelerin görüntünün üzerinde gezdirilerek konvolüsyon işlemi gerçekleştirir. Bu işlem sayesinde aktivasyon haritaları veya özellik haritaları oluşturulur. ESA eğitimi sırasında filtre katsayıları, öğrenme sürecinde sürekli olarak güncellenir ve böylece ağ, veride hangi bölgelerin önemli olduğunu belirler [85].

Örnek olarak, 7x7 boyutunda bir renkli (RGB) giriş görüntüsü düşünüldüğünde, giriş veri boyutu 7x7x3 olacaktır. Bir 3x3 boyutunda filtre ile konvolüsyon işlemi gerçekleştirilirken, filtre giriş görüntüsü üzerinde belirli bir adım (stride) ile kaydırılarak dolaşır. Bu işlem, görüntü matrisinin tamamında uygulanır ve filtre katsayıları her renk kanalındaki değerlerle çarpılıp toplamı alınır. Bu süreç her üç renk kanalında gerçekleştirildikten sonra, aktivasyon haritası oluşturulur [85].

Filtre katsayıları farklı olabilir ve tasarımcılar tarafından modele uygun hale getirilir. Aktivasyon haritasındaki değerler, giriş ve çıkış arasında aynı yoğunluk aralığını sağlamak için normalize edilir. Bu işlem, her bir renk kanalı için hesaplanan değerlerin filtre katsayılarının toplamına bölünmesiyle gerçekleştirilir [85].

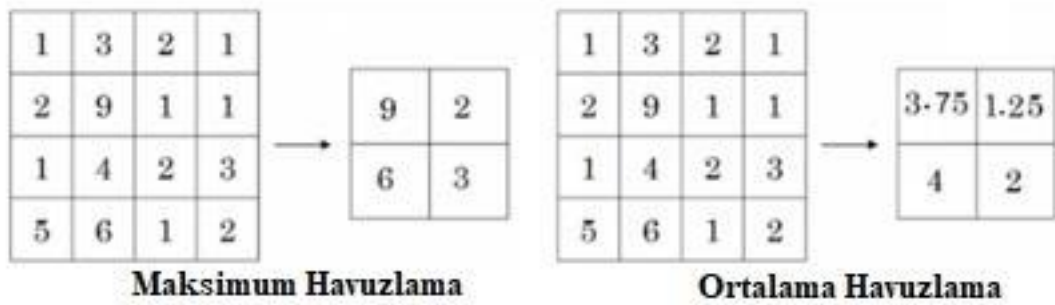
Özellikle, birinci konvolüsyon katmanında 3x3 boyutunda ve toplamda 64 filtreden oluşan bir ESA tasarımı düşünüldüğünde, her bir filtrenin giriş görüntüsüne uygulanması sonucunda elde edilen görüntülerde, her filtrenin giriş görüntüsü üzerindeki etkisi farklıdır. Bu sayede ağ, farklı özellikleri yakalayarak daha doğru ve başarılı sonuçlar elde eder [85]. Şekil 3.23'de evrişimsel sinir ağı filtre yapısı gösterilmektedir [84].



Şekil 3.23: Evrişimsel Sinir Ağı Filtre Yapısı

### 3.5.3.3 Havuzlama katmanı

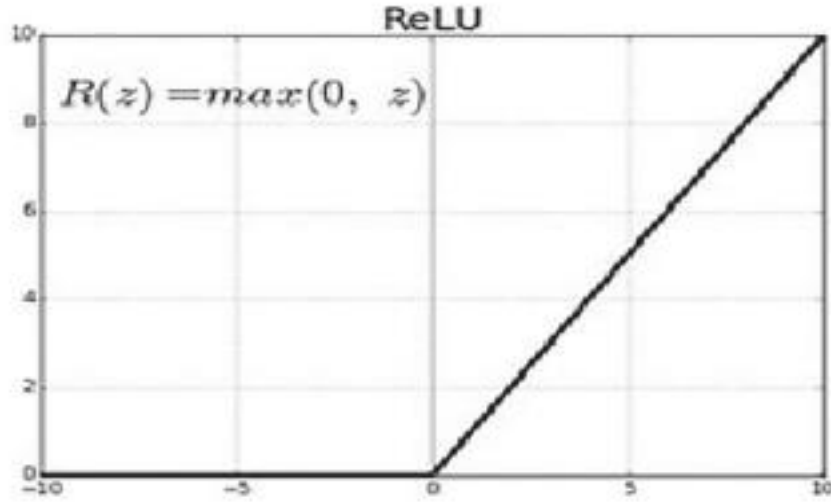
Bu katmanda hedef konvolüsyon katmanından gelen giriş verisinin boyutunu küçültmektir. Bu katmanda da filtreler görüntü üzerinde hareket eder ve yeni değerler hesaplanır. Bu değerler iki farklı türde hesaplanabilir: maksimum havuzlama ve ortalama havuzlama. Ortalama havuzlamada, piksellerin ortalaması hesaplanırken maksimum havuzlamada, görüntüdeki piksellerin en büyük değerleri alınır. Bu katmanında iki önemli parametre bulunmaktadır: filtre boyutu ve adım parametresi. Ortalama havuzlama ve maksimum havuzlama yöntemlerine ait görüntüler 3.24'deki gibidir [84].



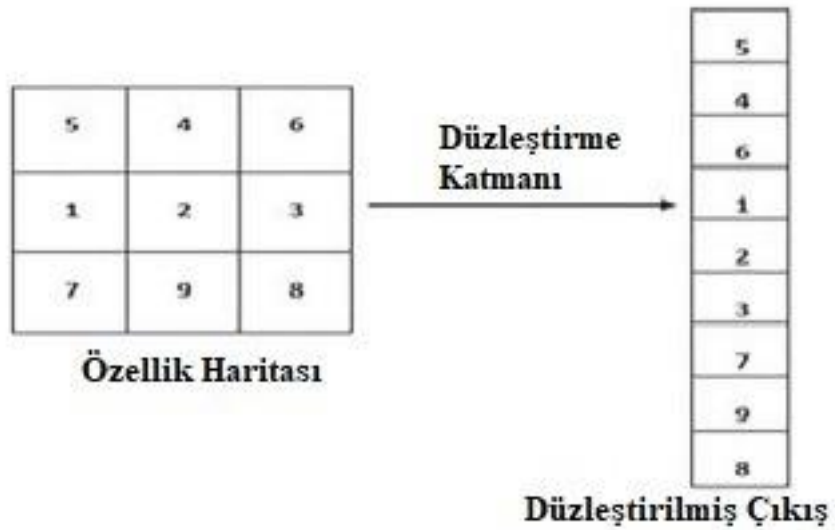
Şekil 3.24: Evrişimsel Sinir Ağı Havuzlama Yapısı

### 3.5.3.4 Düzleştirme Katmanı

Düzleştirme katmanı, konvolüsyon katmanlarının ardından yer alır ve doğrultma katmanı, aktivasyon katmanı olarak da ifade edilebilir. Doğrusal bir yapıda olan bu katmanda amaç negatif değerleri sıfıra dönüştürmektir ve böylece öğrenme süreci hızlanır [85]. Düzleştirme fonksiyonu ve düzleştirme işlemi şekil 3.25 ve 3.26’da gösterilmiştir [86].



Şekil 3.25: Evrişimsel Sinir Ağı Düzleştirme Katmanı Fonksiyonu



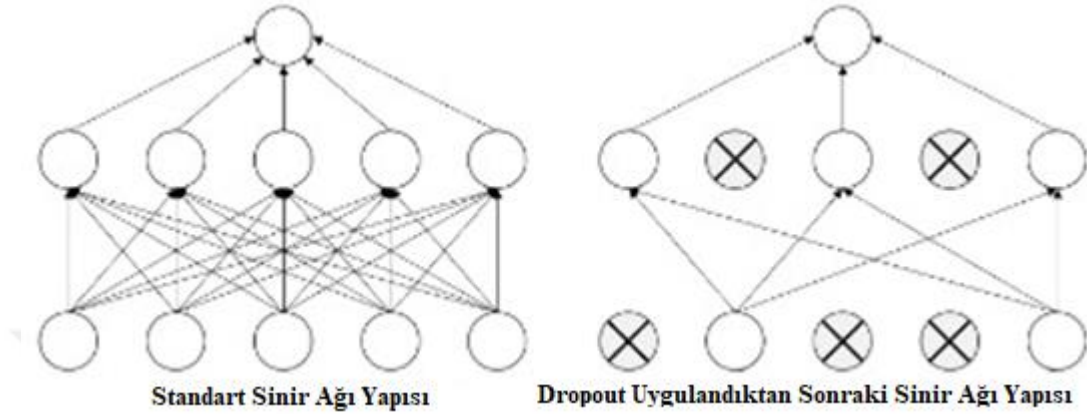
Şekil 3.26: Evrişimsel Sinir Ağı Düzleştirme Yapısı

### 3.5.3.5 Tam Bağlantılı Katman

Evrişimli Sinir Ağları (CNN) yapısında, tam bağlantılı katman evrişim, havuzlama ve düzleştirme katmanlarının ardından konumlanır. Bu katmandaki her düğüm evrişim, havuzlama ve düzleştirme katmanlarındaki tüm düğümlerle bağlantılıdır. Katman sayısı, CNN'nin yapısına bağlı olarak değişiklik gösterir [84].

### 3.5.3.6 Dropout Katmanı

Eğitim sürecinde büyük veri setleri kullanıldığında, Evrişimli Sinir Ağları (ESA) bazen ezberleme eğiliminde olabilir. Ezberlemeyi önlemek amacıyla DropOut katmanı kullanılır . Bu katmanın temel prensibi, ağıdaki belirli düğümleri kaldırmaktır [85]. Şekil 3.27’de ağın orijinal yapısı ve Dropout katmanının uygulanmasından sonra elde edilen yapı gösterilmektedir [84].



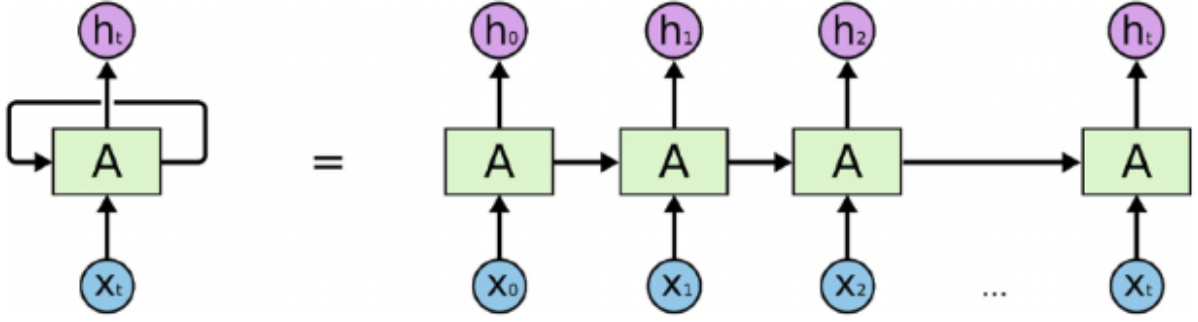
Şekil 3.27: Dropout Uygulandıktan Sonraki Evrişimsel Sinir Ağı

### 3.5.3.7 Sınıflandırma Katmanı

Bu katman tam bağlantılı katmanın ardından yer alır ve CNN'nin sınıflama süreci bu noktada gerçekleşir. Bu katman, girişlerini tam bağlantılı katmandan alırken, çıktılar ise sınıflandırılması gereken nesne sayısına denk gelir [84].

## 3.5.4 Yinelemeli Sinir Ağları (RNN)

1980'lerde ortaya çıkan tekrarlayan sinir ağları (RNN), girdi, bir ya da birden fazla gizli ve çıktı katmanı içeren yapılarıyla dikkat çeker. RNN'lerin temel fikri, tekrar eden modülleri bir bellek olarak kullanarak önceki işlem adımlarındaki kritik bilgileri saklamaktır. İleri beslemeli sinir ağlarından farklı olarak, RNN'ler geri bildirim döngüsüne sahiptir; bu sayede ağ, birden fazla girişi işleyebilir. Bu geri besleme mekanizması, t-1 adımındaki çıktının t adımındaki sonucu ve ardışık adımları etkileyecek şekilde ağa geri döndürülmesini sağlar. Bu nedenle, RNN'ler sıralı veri öğrenmede başarı göstermiştir. Şekil 3.28, RNN'lerin sıralı işleme sürecini gözler önüne sermektedir [87].



Şekil 3.28: Yinelemeli Sinir Ağı Yapısı

Şekil 3.25, temel bir RNN yapısını gösterir; bu yapıda bir girdi birimi, bir çıktı birimi ve tekrarlayan bir gizli birim bulunur. Burada  $x_t$ , zaman adımındaki girdiyi ve  $h_t$ , aynı zaman adımındaki çıktıyı temsil eder. Eğitim sırasında, RNN, ağırlık matrislerini ayarlamak için geri yayılım algoritmasına başvurur. Bu, YSA'daki gradyanların hesaplanması ve ağırlıkların güncellenmesine katkıda bulunan yaygın bir yöntemdir [87].

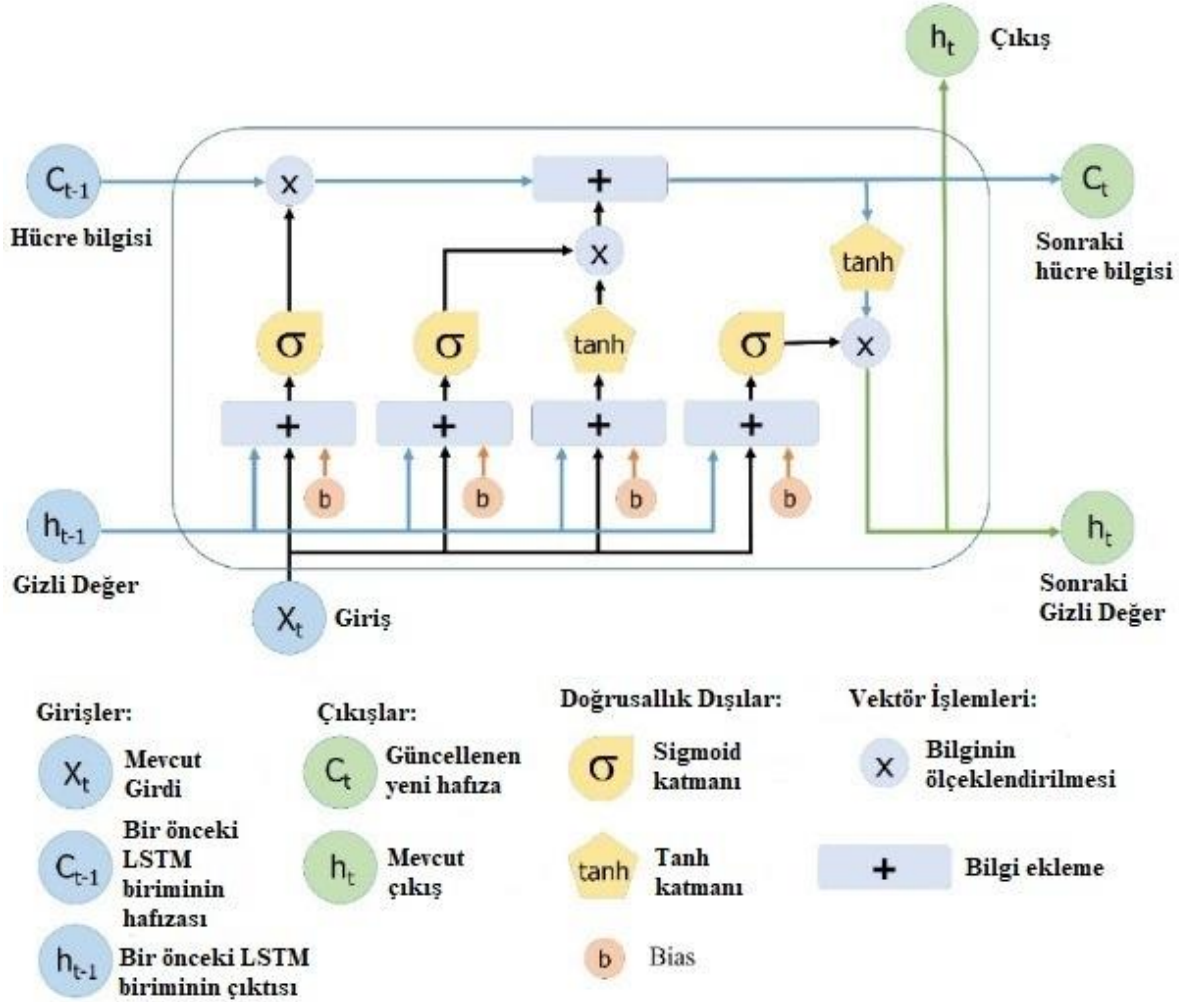
Bununla birlikte, RNN'de uygulanan geri yayılım, zaman içinde geri yayılım (BPTT) olarak bilinir ve alışılmışın dışında bir geri bildirim süreci kullanır. BPTT, son çıktıdan başlayarak katmanları geriye doğru işleyen bir yöntemle her birimin ağırlıklarını, hesaplanan toplam çıktı hatasına göre günceller. Ancak, bilgi döngüleri nedeniyle, sinir ağı modelinde büyük ağırlık güncellemeleri meydana gelir ve gradyanların birikimi, güncelleme süreci sırasında istikrarsız bir ağı ortaya çıkarır [87].

BPTT'nin uzun vadeli bağımlılıkları öğrenmedeki yetersizliği, yok olan ve patlayan gradyan problemleri nedeniyle. Bu durum, tekrarlayan sinir ağlarının eğitiminde önemli zorluklara yol açar [87].

#### 3.5.4.1 Uzun-Kısa Süreli Bellek Ağları (LSTM)

Uzun Kısa Süreli Bellek (LSTM), RNN'nin gelişmiş bir versiyonudur ve araştırmacılar tarafından RNN'nin bazı dezavantajlarını gidermek amacıyla geliştirilmiştir. LSTM'ler, uzun vadeli bağımlılıkları öğrenme ve bilgiyi uzun süreliğine saklama kabiliyetine sahip özel bir RNN çeşididir [87].

LSTM modeli, bir dizi modül şeklinde düzenlenir. Fakat bu tekrarlanan modüller, standart bir RNN'den farklı bir yapıya sahiptir. Tek bir sinir ağı katmanını yerine, bu modüller dört katmandan oluşur ve benzersiz bir iletişim yöntemiyle birbirleriyle etkileşim içindedirler. LSTM sinir ağının yapısal özellikleri Şekil 3.29'da sunulmuştur [87].



Şekil 3.29: LSTM Ağ Yapısı

LSTM ağları, bellek blokları olarak adlandırılan hücrelerden oluşur. İki durum, hücre durumu ve gizli durum, bir hücreden diğerine aktarılır. Hücre durumu, veri akışının temel hattını oluşturur ve esas olarak verilerin ileri yönde değişmeden akmasına izin verir, yalnızca bazı doğrusal dönüşümler gerçekleşebilir [87].

Hücre durumundaki veriler, sigmoid kapılar aracılığıyla eklenip çıkarılabilir. Bir kapı, farklı ağırlıklar içeren bir katmana veya matris işlem dizisine benzer bir yapıdır. LSTM'ler, bu

kapıları kullanarak bellek süreçlerini düzenler ve böylece uzun vadeli bağımlılık problemlerini önlemeye yönelik olarak tasarlanmıştır [87].

Bir LSTM kurarken ilk aşama, gereksiz bilgileri tespit edip hücreden çıkarılmasını sağlamaktır. Bu bilgi ayıklama sürecinde, sigmoid fonksiyonu önceki LSTM biriminin çıktısını ve mevcut zaman noktasındaki girdiyi kullanarak hangi bilgilerin çıkarılması gerektiğine karar verir. Sigmoid fonksiyonu aynı zamanda eski çıktının hangi kısmının atılması gerektiğini belirler. Bu mekanizmaya unutma kapısı denir ve bu mekanizma hücre durumundaki her bir değere karşılık gelen 0 ile 1 arasında değerler alan bir vektördür [87].

### 3.6 Birleştirme Yöntemleri

Cross-Modal transfer öğrenmesi gerçekleştirmeden önce üç farklı modalite birleştirme türünden bahsedilmesi gerekir: erken birleştirme, ara birleştirme ve geç birleştirme. Erken birleştirme, veri kaynaklarının işleme sürecinin başlangıcında birleştirilmesidir. Ham verilerin doğrudan birleştirilebileceği bu yöntemde, genelde öncelikli olarak belirli özelliklerin çıkarılması tercih edilir. Bu aşamada çıkarılan özellikler, daha sonra bireysel verilerin tek bir temsile birleştirilmesinde kullanılır. Birleştirilmiş temsilde verilerin doğru bir şekilde birleştirildiğinden emin olunmalıdır, bu sayede veriler sonraki aşamalar için daha uygun bir şekilde işlenebilir. Eğer veriler uygun şekilde birleştirilmişse, çapraz korelasyonlar arasında var olan bağlantılar ortaya çıkarılabilir ve bu da sistem performansının artmasına yardımcı olabilir. Bununla birlikte, birleştirilen bu düşük seviyeli özelliklerin bazen işlem için gereksiz olabileceği unutulmamalıdır. Dolayısıyla, bu durum erken füzyonun gücünü zayıflatabilir, çünkü alakasız özelliklerin birleştirilmesi sistemin genel verimliliğini düşürebilir. Bu nedenle, hangi özelliklerin çıkarılacağı ve birleştirileceği konusunda dikkatli olunmalıdır [92].

Geç birleştirme, birleştirme teknikleri arasında en basit ve en yaygın kullanılan yöntemdir. Bu yaklaşım, farklı veri türlerini ayrı ayrı tamamlanan işlem akışlarından sonra birleştirir. Her bir veri türü, belirli özelliklerine uygun şekilde özelleştirilmiş yöntemlerle işlenir. Her veri türü için tamamlanan işlem zinciri sonrasında, genellikle bir sınıflandırma görevinde olduğu gibi, tahmini etiketler alınır ve bu sonuçlar en basit haliyle toplanır veya ortalama alınır. Böylece, farklı veri türlerinden elde edilen sonuçlar birleştirilmiş olur. Geç füzyonun dezavantajı, farklı veri türleri arasındaki çapraz korelasyonların kullanımı konusunda bir kısıtlılık sunabilir. Ara birleştirme ise verilerin birleştirilmesi erken ve geç birleştirme

arasında bir yerde, yani işleme sürecinin ortasında gerçekleşir. Bu yöntem diğerlerine göre daha esnektir fakat bu yaklaşımın en büyük zorluğu ne zaman ve nasıl birleştirme yapılacağını belirlemektir [92].



## 4. MATERYAL

Bu çalışmada modlar arası transfer öğrenimi, RAVDESS veri setindeki görüntüler ile duygu sınıflayan modelden öğrenilen bilgileri RAVDESS veri setindeki konuşma ses sinyallerini sınıflayan modele aktararak ses sinyallerini sınıflayan modelin başarısını arttırmayı hedeflemektedir. Bu hedef literatüre transfer öğrenmesi alanında modlar arası transfer öğrenmenin ne kadar başarılı olabileceğini göstermesi açısından katkı sağlamaktadır.

Veri tabanları duygu sınıflama için büyük öneme sahiptir. Çalışmanın performansını doğrudan etkileyebilir. Veri tabanları açık erişimli olup olmadığına içerdiği duygu durumlarının sayısına, içerdiği dile, söyleyiş tarzına, profesyonel bir şekilde oluşturulup oluşturulmadığına göre sıralanabilir [2][38].

Veri tabanlarında karşılaşılabilecek zorlukların başında veri setinde ki duygu durumlarının yeteri kadar yoğunlukta modellenmesinin yapılamamasıdır. Ayrıca kayıtnın frekansı ve bit hızı gibi teknik özelliklerinin kısıtlı olması da çalışmanın kalitesini düşürebilir. [1][39].

Literatürde ses sinyallerinden duygu tanımak için kullanılan başlıca veri setleri EmoDB (Berlin Duygusal Konuşma Veri Seti) [40], Etkileşimli duygusal ikili hareket yakalama veri seti (IEMOCAP) [41], Surrey Sesli-Görsel İfade Edilen Duygu veri seti (SAVEE) [42] ve RAVDESS (Ryerson Sesli-Görsel Duygusal Konuşma ve Şarkı Veri Seti) [43]'dir.

EmoDB veri seti açık erişimli ve almanca dilli bir veri setidir. İçerisinde Nötr, mutluluk, kaygı, üzüntü, iğrenme, sıklık ve öfke olmak üzere 7 farklı duygu durumunu içerir. 5'i erkek 5'i kadın 10 farklı insanın seslendirdiği bu veri setine ait 535 adet söylem bulunmaktadır [40].

SAVEE veri seti ingilizce dilli bir veri setidir. Veri seti öfkeli, nötr, şaşırılmış, mutlu, iğrenme, üzüntülü ve korkulu olmak üzere 7 farklı duygu durumunu içerir. Toplamda 480 İngiltere aksanlı ifade olmak üzere, 4 erkeğin seslendirdiği bu veri setinde her konuşmacıya ait 120 söylem vardır. Veri seti yüksek kalitede görsel ve akustik verileri içerir. Veri setinde ses verileri için örnekleme hızı 44.1 kHz, video verileri için örnekleme hızı 60 fps olarak belirlenmiştir [42].

IEMOCAP veri seti ingilizce dilli bir veri setidir. Veri seti 5'i kadın 5'i erkek olmak üzere 10 insan tarafından hazırlanmıştır. Yaklaşık 12 saatlik görsel ve işitsel veri içermektedir. Veri seti heyecanlı, mutlu, nötr, bıkkın, hayal kırıklığına uğramış, kızgın, korkmuş, üzgün, şaşırılmış olmak üzere 9 farklı duygu durumunu içerir [41]. Literatürde kullanılan daha fazla veri setine ait bilgiler tablo 4.1'de derlenmiştir.



| Veri seti                                                                 | Dil           | Boyut                                          | Duygular                                                                                      |
|---------------------------------------------------------------------------|---------------|------------------------------------------------|-----------------------------------------------------------------------------------------------|
| EmoDB (Berlin Duygusal Konuşma Veri Seti) [40]                            | Almanca       | 10 Konuşmacı<br>535 Söylev                     | Nötr, mutluluk, kaygı, üzüntü, iğrenme, sıklık ve öfke                                        |
| Etkileşimli duygusal ikili hareket yakalama veri seti (IEMOCAP) [41]      | İngilizce     | 10 Konuşmacı<br>302 video                      | Heyecanlı, mutlu, nötr, bıkkın, hayal kırıklığına uğramış, kızgın, korkmuş, üzgün, şaşırılmış |
| Surrey Sesli-Görsel İfade Edilen Duygu veri seti (SAVEE) [42]             | İngilizce     | 4 Konuşmacı<br>480 Söylev                      | 7 farklı duygu durumu                                                                         |
| Ryerson Sesli-Görsel Duygusal Konuşma ve Şarkı Veri Seti (RAVDESS) [43]   | İngilizce     | 24 Konuşmacı<br>7356 adet ses ve görsel verisi | Sakin, üzgün, mutlu, şaşırılmış, korkmuş, nötr, kızgın, iğrenme                               |
| Kalabalık Kaynaklı Duygusal Çok Modlu Aktörler Veri Kümesi (CREMA-D) [44] | İngilizce     | 91 Konuşmacı<br>7442 adet söylev               | Mutlu, Üzgün, Öfke, Nötr, Tiksime, Korku                                                      |
| Çoklu Modlu Spontan Performans Doğaçlaması Veri Seti (MSP-IMPROV) [45]    | İngilizce     | 12 Konuşmacı<br>8438 Söylev                    | Kızgın, Üzgün, Mutlu, Nötr                                                                    |
| Enterface Veri Seti [46]                                                  | 14 farklı dil | 42 Oyuncu<br>1166 video                        | Şaşkınlık, korku, iğrenme, mutluluk, üzüntü, öfke                                             |
| Bahçeşehir University Multimodal Affective Database - 1 (BAUM-1) [47]     | Türkçe        | 31 Aktör<br>1502 video                         | Şaşkınlık, Üzüntü, İlgi, Öfke, Mutluluk, Emin olmayan, İğrenme, Korku, Can sıkıntısı          |

Tablo 4.1: Literatürde kullanılan veritabanları

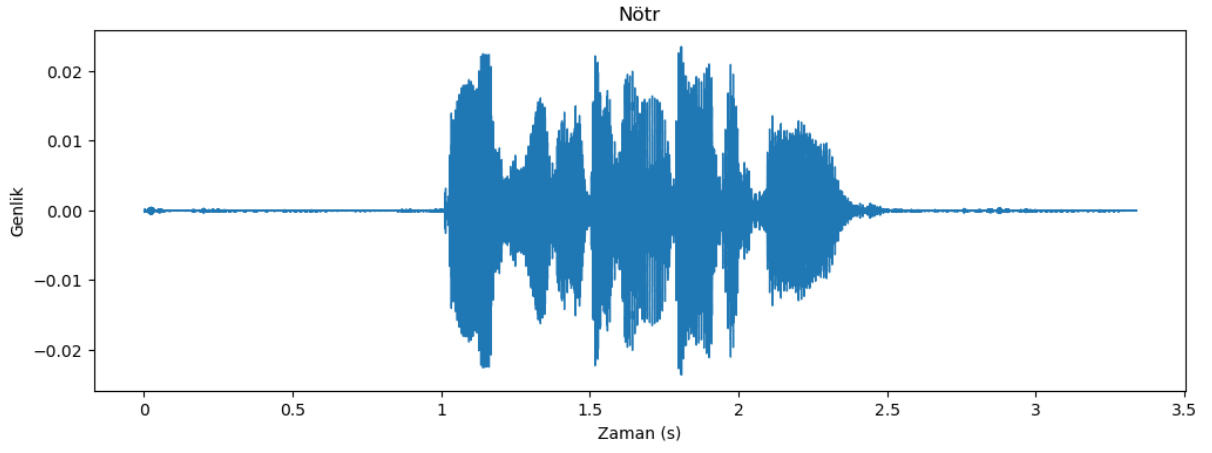
## 5. YÖNTEM

### 5.1 Veri Seti

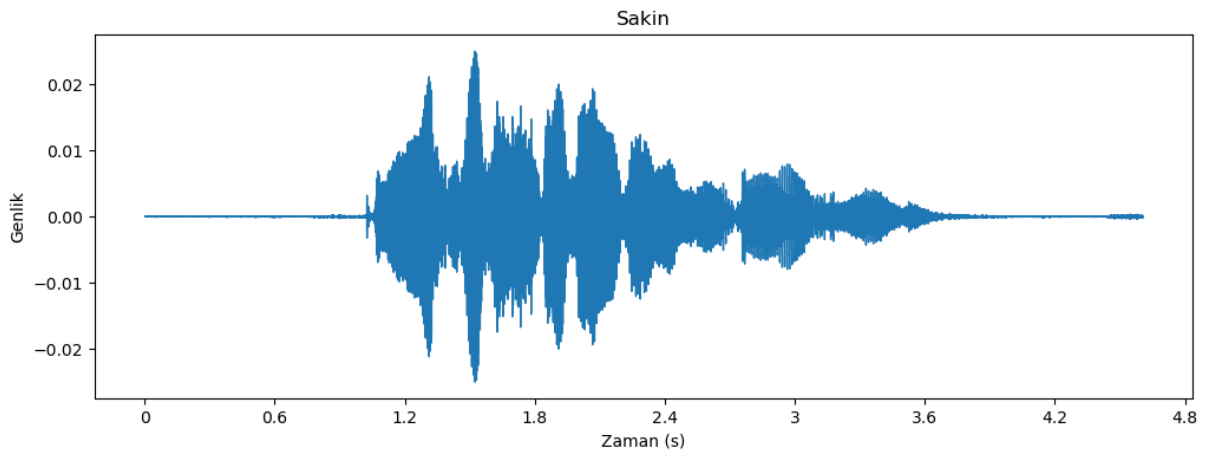
Bu tez çalışmasında kullanılan veri seti Ryerson Sesli-Görsel Duygusal Konuşma ve Şarkı Veri Seti (RAVDESS)'dir [43]. Açık erişimli bir veri seti olan RAVDESS 24 farklı aktör tarafından görsel ve işitsel verilerle oluşturulmuştur. 24 aktörün 12'si erkek 12'si kadındır. Veri seti şaşırmış, sakin, kızgın, mutlu, üzgün, iğrenmiş, korkmuş ve nötr olmak üzere 8 farklı duygu durumunu içerir. Ayrıca veri setinde her ifadenin normal ve güçlü olmak üzere 2 farklı duygu seviyesini ifade eden versiyonu vardır. Veri setinde konuşma ve şarkı olmak üzere 2 farklı söyleyiş tarzı mevcuttur. Veri setindeki söyleyiş tarzı ve özelliklerine göre dağılımı şekil 2.1'de gösterilmiştir. Şekil 5.2'de nötr, Şekil 5.3' de sakin, Şekil 5.4'de mutlu, Şekil 5.5'de üzgün, Şekil 5.6'da kızgın, Şekil 5.7'de korkulu, Şekil 5.8'de iğrenme ve Şekil 5.9'da şaşırmış duygu durumu olmak üzere veri setindeki her duygu durumuna ait rastgele bir ses kaydının sinyal gösterimi gösterilmiştir.



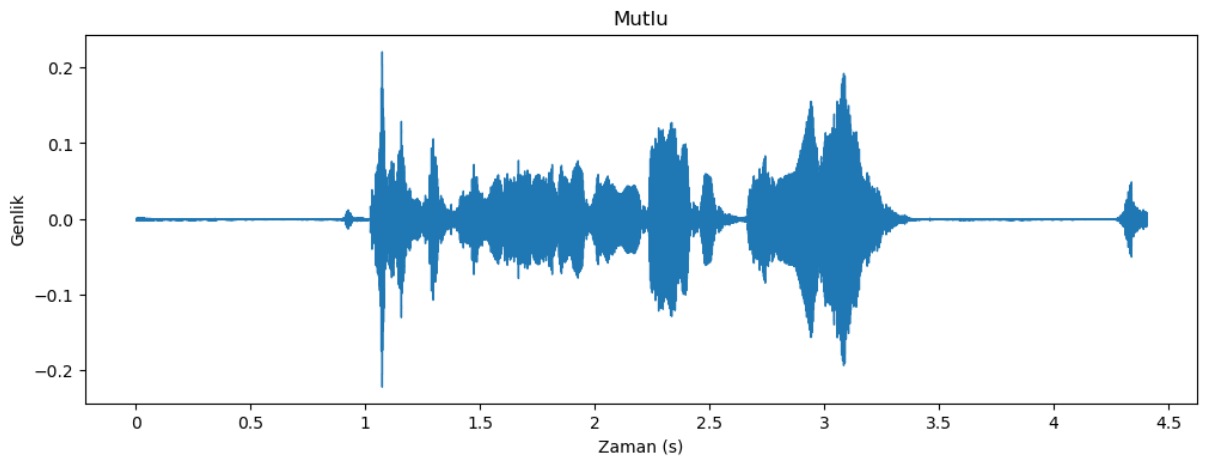
Şekil 5.1: Veri seti veri türleri dağılımı



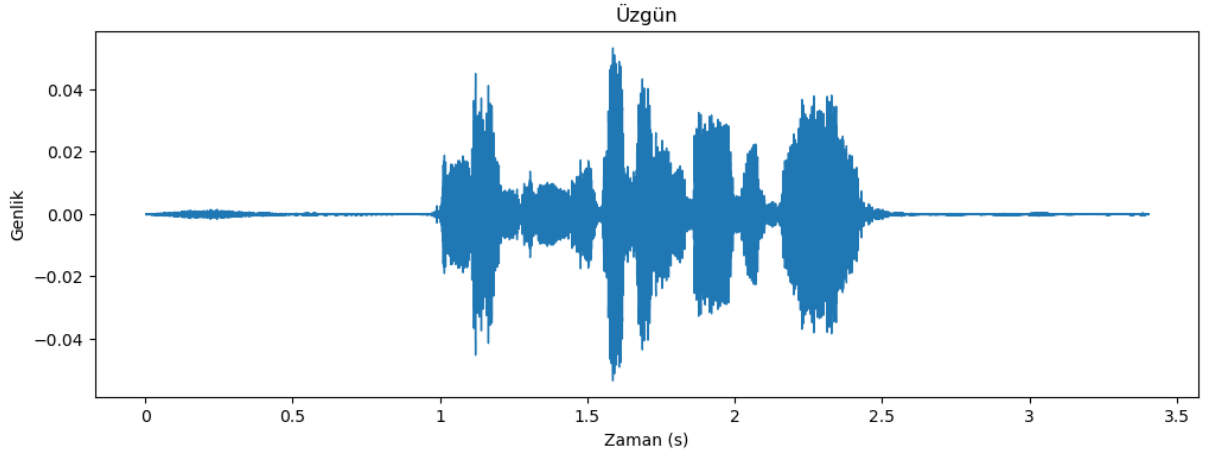
Şekil 5.2: Nötr Ses Sinyali Örneği



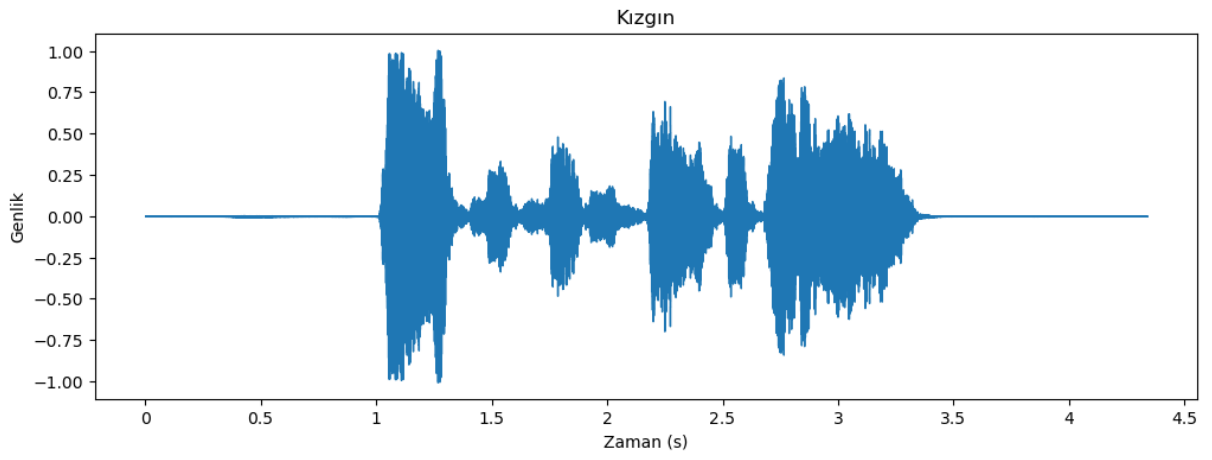
Şekil 5.3: Sakin Ses Sinyali Örneği



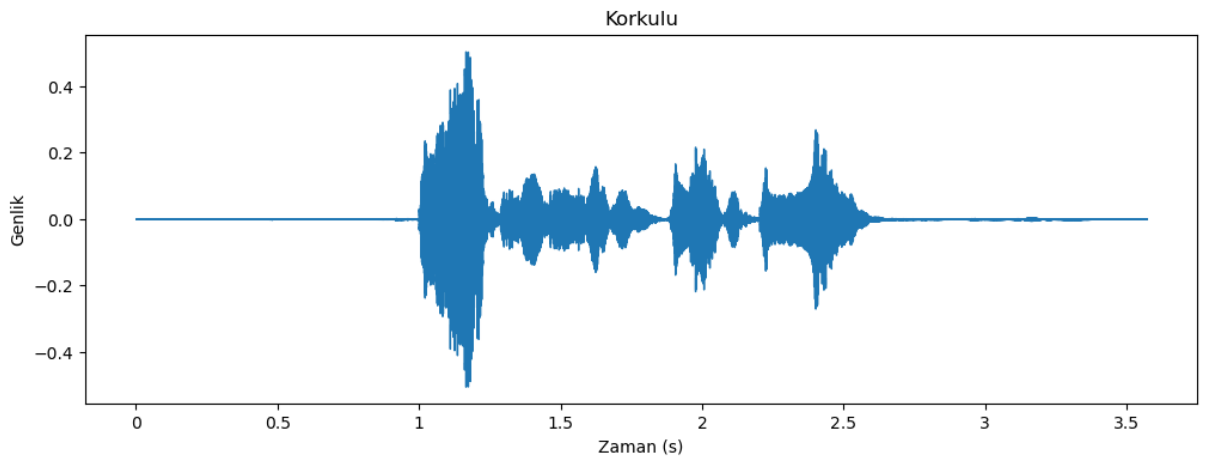
Şekil 5.4: Mutlu Ses Sinyali Örneği



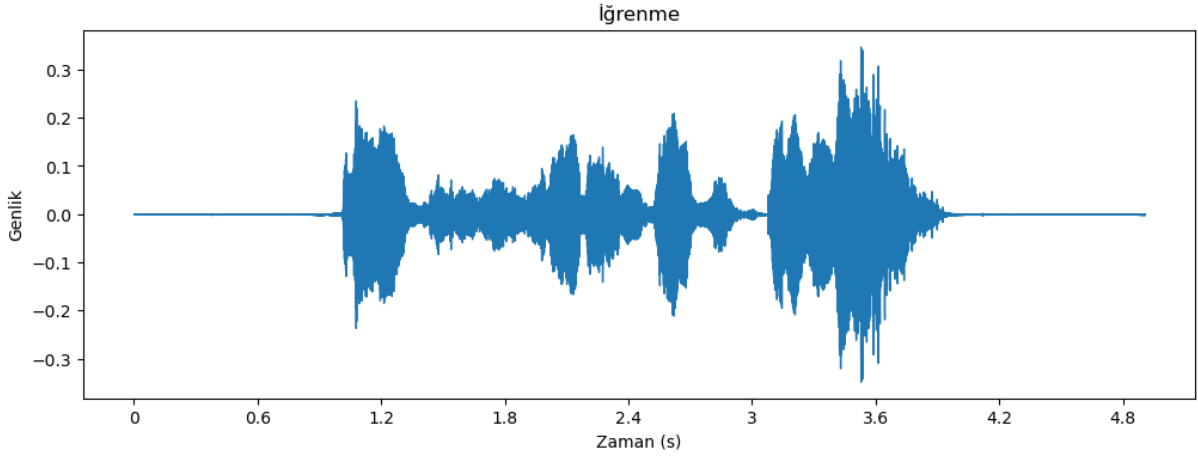
Şekil 5.5: Üzgün Ses Sinyali Örneği



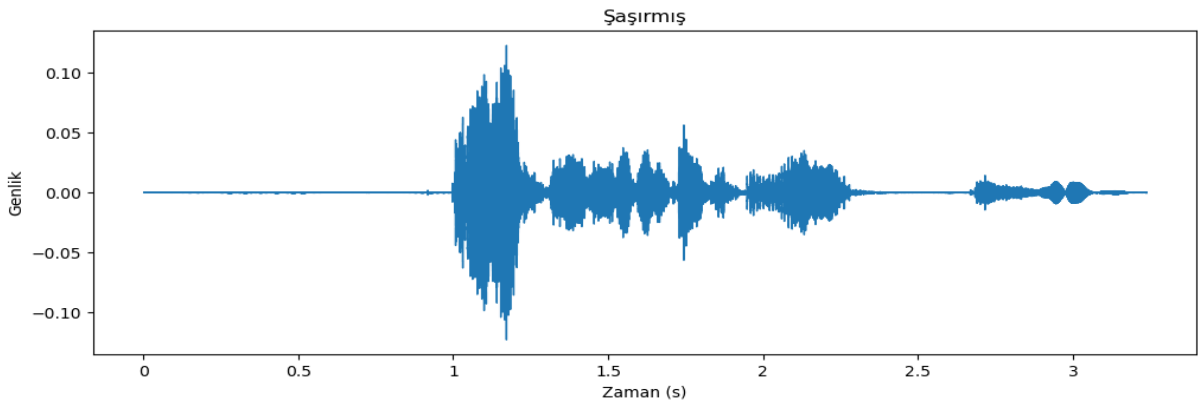
Şekil 5.6: Kızgın Ses Sinyali Örneği



Şekil 5.7: Korkulu Ses Sinyali Örneği



Şekil 5.8: İğrenme Ses Sinyali Örneği



Şekil 5.9: Şaşırılmış Ses Sinyali Örneği

Veri setinde sadece ses içeren veriler 48kHz örnekleme frekansına, 16 bit hıza sahiptir. Sadece ses verilerinin uzantısı Waveform Audio File Format (wav)'dır. Veri setindeki video veriler ise 1280x720 piksel HD çözünürlüğündedir. Ayrıca video verileri H.264 AAC video sıkıştırma standardına 48kHz örnekleme frekansına sahiptir, uzantısı ise mp4 formatındadır [43].

Veri setinde iki farklı söylev vardır. "Kids are talking by the door" (Çocuklar kapıda konuşuyor), ve "Dogs are sitting by the door" (Köpekler kapının yanında oturuyor) olmak üzere veri setinde iki farklı söylev vardır. Veri setinde 1440 sadece konuşma sesi, 1012 sadece şarkı sesi, 2880 konuşma videosu, 2024 şarkı videosu olmak üzere toplam 7356 adet görsel-işitsel kayıt vardır. Veri setinin toplam büyüklüğü 24.8 GB'dir [43].

Bu çalışmada kullanılan görüntüler, veri setindeki her bir videonun tam ortasından, tam ortasından 0.5 saniye öncesinden ve sonrasından, tam ortasından 1 saniye öncesinden ve

sonrasında olmak üzere 5 adet görüntü çıkartılarak oluşturulmuştur. 14400 adet konuşma videosuna ait görüntü, 10120 adet şarkı videosuna ait görüntü olmak üzere 24520 adet görüntü elde edilmiştir.

## 5.2 Ses Öznitelik Çıkartıcıları

Bu tez çalışmasında MFCC ve LMFB öznitelik çıkartıcıları kullanılmıştır ve librosa kütüphanesinden faydalanılmıştır. Her bir ses dosyası librosa.load fonksiyonuyla sisteme yüklendikten sonra her ses dosyasına ait ses zaman serisi verisi örnekleme hızı elde edilir. Daha sonra bu ses dosyasının librosa.MFCC fonksiyonuyla MFCC öznitelikleri çıkartılır. Bu tez çalışmasında MFCC ve LMFB uygulanan tüm çalışmalarda karmaşıklık katsayısı 60 olarak seçilmiştir. LMFB öznitelik çıkartıcısında Ayırık kosinüs değişimi uygulanmaz. MFCC öznitelik çıkartıcısında varsayılan olarak tip 2 ayırık kosinüs değişimi uygulanır, tip 2 yöntemi librosa tarafından önerilir ve daha iyi sıkıştırma yapar. Norm parametresi varsayılan olarak 'ortho' ayarlıdır, bu parametre ayırık kosinüs değişimi için kullanılacak normalizasyon tipidir ve tip 2 ayırık kosinüs değişimi ile beraber librosa tarafından önerilir. n\_fft parametresi 2048, hop\_length parametresi 512 ve pencere fonksiyonu olarak hann fonksiyonu seçilmiştir.

## 5.3 Ses Sınıflama Yöntemi

Ses sınıflama modelimiz için sırasıyla 125, 250 ve 125 nöronlu olacak şekilde 3 adet Dense katmanı kullanılmıştır. Dropout katmanı için 0.5 değeri seçilmiştir. Çıkmış katman nöronunda softmax diğer katmanlarda relu aktivasyon fonksiyonu kullanılmıştır. MFCC ve LMFB işlemlerinden elde edilen öznitelikler öğrenme modeline giriş nöronundaki Dense katmanına input\_shape parametresi aracılığıyla dahil edilmektedir.

## 5.4 Görüntü Sınıflama Yöntemi

80x90 boyutundaki görüntüler ImageDataGenerator.flow\_from\_directory fonksiyonu ile hem sisteme yüklenir hem de bazı ayarlamalar yapılır. Rescale parametresi 1./255 seçilmiştir bu görüntüdeki piksel değerlerinin 0-1 arasına ölçeklenmesini sağlar. Zoom\_range parametresi 0.3 seçilmiştir bu ise görüntülerde %30'a kadar rastgele yakınlaştırma yapar. Horizontal\_flip parametresi ise true olarak belirlenmiştir bu ise görüntüyü yatay ekseninde rastgele bir oranda çevirir. Shuffle parametresi true olarak belirlenmiştir bu veri setindeki verileri rastgele

kariřtırır. Bu ayarlamaların tümü veri setini güçlendirmek ve başarıyı maksimize etmek için gerekleřtirilmiřtir.

Sınıflandırma için yapay sinir ağı modelinde ıkıř nron katmanı hari tüm katmanlarda aktivasyon fonksiyonu olarak relu kullanılmıřtır, ıkıř katmanında ise softmax fonksiyonu kullanılmıřtır. Dropout katmanlarında oran 0.25 seilmiřtir. Modelde 4 adet sırasıyla 32, 64, 128 ve 256 nronlu olacak řekilde Conv2D ağı kullanılmıřtır. Modelde 2 adet sırasıyla 60 ve 125 nronlu olacak řekilde 2 adet Dense katmanı kullanılmıřtır. Normalizasyon için BatchNormalization, MaxPooling2D ve Flatten katmanları kullanılmıřtır.

## 5.5 Transfer ğrenmesi Yöntemi

Bu tez alıřmasında transfer ğrenmesi amacıyla görüntü verilerini sınıflandıran modeli ve konuřma ses sinyallerini sınıflandıran modeli eđittikten sonra görüntü verilerini sınıflandıran modelden ses sinyallerini sınıflandıran modele ağırlık aktarımı gerekleřtirilmiřtir. Transfer ğrenmesi için öncelikle ağırlık aktarımı yapılacak katmanlar belirlenir. Bu tez alıřmasında görüntüleri sınıflandırmak için kullanılan modelin ıkıř katmanının bir önceki katmandan konuřma ses sinyallerini sınıflandıran modelin ilk katmanına ağırlık aktarımı gerekleřtirilmiřtir. Bu iřlem Keras kütüphanesindeki `get_weights()` ve `set_weights()` fonksiyonları aracılıđıyla gerekleřtirilmiřtir.

## 5.6 Yöntemde Kullanılan Parametre Deneyleri

Bu tez alıřmasında kullanılan önemli parametrelerden biri olan mel filtre bankası katsayısı üzerinde deneyler gerekleřtirilmiřtir. Bu deneyler Konuřma tipi ses sinyalleri, MFCC öznitelik ıkartıcısı ve Dense katmanları uygulaması üzerinde gerekleřtirilmiřtir. Mel filtre bankası katsayısı 20 iken modelin sınıflama başarıısı %54 olmuřtur, modelin toplam parametre sayısı ise 66508 olmuřtur. Mel filtre bankası katsayısı 40 iken modelin sınıflama başarıısı %56 olmuřtur, modelin toplam parametre sayısı ise 69008 olmuřtur. Mel filtre bankası katsayısı 60 iken modelin sınıflama başarıısı %59 olmuřtur, modelin toplam parametre sayısı ise 71508 olmuřtur. Mel filtre bankası katsayısı 80 iken modelin sınıflama başarıısı %59 olmuřtur, modelin toplam parametre sayısı ise 74008 olmuřtur. Mel filtre bankası katsayısı 100 iken modelin sınıflama başarıısı %59 olmuřtur, modelin toplam parametre sayısı ise 76508 olmuřtur. Bu durumda mel filtre bankası katsayısı olarak tercih edilmesi en optimum deđerin 60 olduđu gözükmemektedir.

Bu tez çalışmasında kullanılan katman sayısının başarımı nasıl etkileyeceği konusunda da deneyler gerçekleştirilmiştir. Bu deneyler Konuşma tipi ses sinyalleri, MFCC öznitelik çıkartıcısı ve Dense katmanları uygulaması üzerinde gerçekleştirilmiştir. 2 katmanlı Dense yapısı kullanılan uygulamada başarımları %57’de kalırken modelin toplam parametre sayısı 41133 olmuştur. 3 katmanlı Dense yapısı kullanılan uygulamada başarımları %59 iken modelin toplam parametre sayısı 71508 olmuştur. 4 katmanlı Dense yapısı kullanılan uygulamada başarımları %53’de kalırken modelin toplam parametre sayısı 104008 olmuştur. 5 katmanlı Dense yapısı kullanılan uygulamada başarımları %45’de kalırken modelin toplam parametre sayısı 134383 olmuştur. Gerçekleştirilen bu deneyler sonucunda optimum Dense ağı katman sayısının 3 olduğu gözükmemektedir.



## 6. UYGULAMA

Bu tez çalışmasında 11 farklı uygulama gerçekleştirilmiştir ve veri seti olarak tamamıyla RAVDESS veri setindeki ses sinyalleri ve görüntü verileri kullanılmıştır. Ayrıca MFCC ve LMFB öznitelikleri çıkartılan tüm uygulamalarda filtre karmaşıklığı 60 kullanılmıştır ve tüm modellerin amacı duygu sınıflamaktır. Tüm uygulamalarda sınıflama raporu, karışıklık matrisi, doğruluk ve kayıp grafiği gösterilmiştir. Birinci uygulamada konuşma ses sinyallerinden MFCC öznitelikleri çıkartılmıştır ve YSA yapısında Dense katmanları kullanılarak duygu sınıflama çalışması yapılmıştır. İkinci uygulamada konuşma ses sinyalinden MFCC öznitelikleri çıkartılmıştır ve RNN yapısında LSTM katmanları kullanılarak duygu sınıflama çalışması yapılmıştır. Üçüncü uygulamada şarkı tipi ses sinyallerinden MFCC öznitelikleri çıkartılmıştır YSA yapısında Dense katmanları kullanılmıştır. Dördüncü uygulamada şarkı tipi ses sinyallerinden MFCC öznitelikleri çıkartılmıştır RNN yapısında LSTM katmanları kullanılmıştır. Beşinci uygulamada konuşma ses sinyallerinden LMFB öznitelikleri çıkartılmış ve YSA yapısında Dense katmanları kullanılmıştır. Altıncı uygulamada konuşma ses sinyallerinden LMFB öznitelikleri çıkartılmış ve RNN yapısında LSTM katmanları kullanılmıştır. Yedinci uygulamada şarkı tipi ses sinyallerinden LMFB öznitelikleri çıkartılmış ve YSA yapısında Dense katmanları kullanılmıştır. Sekizinci uygulamada şarkı tipi ses sinyallerinden LMFB öznitelikleri çıkartılmış ve RNN yapısında LSTM katmanları kullanılmıştır. Dokuzuncu uygulamada görüntü verilerinden duygu tanıma çalışması yapılmış ve Evrişimli Sinir Ağları yapısı kullanılmıştır. Onuncu uygulamada tezinde temel çalışma konusu olan transfer öğrenmesi uygulaması gerçekleştirilmiştir, bu uygulamada görüntülerden duygu sınıflayan modelden ses verilerinden duygu sınıflayan modele transfer öğrenmesi yöntemiyle bilgi transferi gerçekleştirilmiş ve ses verilerinden duygu sınıflayan modelin başarısının arttığı görülmüştür. On birinci ve son uygulamada erken birleştirme yöntemiyle görüntü verilerini sınıflayan modelden ses verilerinden duygu sınıflayan modele model eğitilmeden önce öznitelik aktarımı yapılmıştır ve ses verilerinden duygu sınıflayan modelin başarısının arttığı görülmüştür. Sınıflama raporunda ki metrikler “Precision”, “Recall” ve “F1-Score” olarak belirlenmiştir. Precision, pozitif olarak tahmin edilen örneklerin kaçının gerçekten de pozitif olduğunu gösterir. Recall, gerçekten pozitif olan örneklerinin ne kadarının doğrulukla tahmin edildiğini gösterir. F1-Score ise Precision ve Recall değerlerinin harmonik ortalamasıdır, precision ve recall değerleri arasında bir denge sağlar.

Tez çalışmasında gerçekleştirilen uygulama sıralaması:

1. Konuşma tipi ses sinyali, MFCC öznitelik çıkartıcısı, Dense katmanları

2. Konuşma tipi ses sinyali, MFCC öznitelik çıkartıcısı, LSTM katmanları
3. Şarkı tipi ses sinyali, MFCC öznitelik çıkartıcısı, Dense katmanları
4. Şarkı tipi ses sinyali, MFCC öznitelik çıkartıcısı, LSTM katmanları
5. Konuşma tipi ses sinyali, LMFB öznitelik çıkartıcısı, Dense katmanları
6. Konuşma tipi ses sinyali, LMFB öznitelik çıkartıcısı, LSTM katmanları
7. Şarkı tipi ses sinyali, LMFB öznitelik çıkartıcısı, Dense katmanları
8. Şarkı tipi ses sinyali, LMFB öznitelik çıkartıcısı, LSTM katmanları
9. Görüntü tabanlı duygu tanıma için derin öğrenme modeli
10. Duygu tanıma performansını arttırmak için transfer öğrenmesi
11. Duygu tanıma performansını arttırmak için erken birleştirme yöntemi

## **6.1 MFCC tabanlı ses sinyali duygu tanıma**

MFCC tabanlı ses sinyali duygu tanıma başlığı altında 4 farklı uygulama gerçekleştirilmiştir. Birinci uygulamada konuşma tipi ses sinyali, MFCC öznitelik çıkartıcısı ve Dense katmanları kullanılmıştır. İkinci uygulamada konuşma tipi ses sinyalleri, MFCC öznitelik çıkartıcısı ve LSTM katmanları kullanılmıştır. Üçüncü uygulamada şarkı tipi ses sinyalleri, MFCC öznitelik çıkartıcısı ve Dense katmanları kullanılmıştır. Dördüncü uygulamada ise şarkı tipi ses sinyalleri, MFCC öznitelik çıkartıcısı ve LSTM katmanları kullanılmıştır. MFCC tabanlı ses sinyali duygu tanıma başlığı altında öncelikle konuşma tipi ses sinyalleri sonrada şarkı tipi ses sinyallerini içeren uygulamalar gerçekleştirilecektir

### **6.1.1 MFCC tabanlı konuşma tipi ses sinyali duygu tanıma**

MFCC tabanlı konuşma tipi ses sinyali duygu tanıma başlığı altında 2 farklı uygulama gerçekleştirilmiştir. Birinci uygulama konuşma tipi ses sinyali, MFCC öznitelik çıkartıcısı ve Dense katmanları kullanılmıştır. İkinci uygulamada ise konuşma tipi ses sinyalleri, MFCC öznitelik çıkartıcısı ve LSTM katmanları kullanılmıştır.

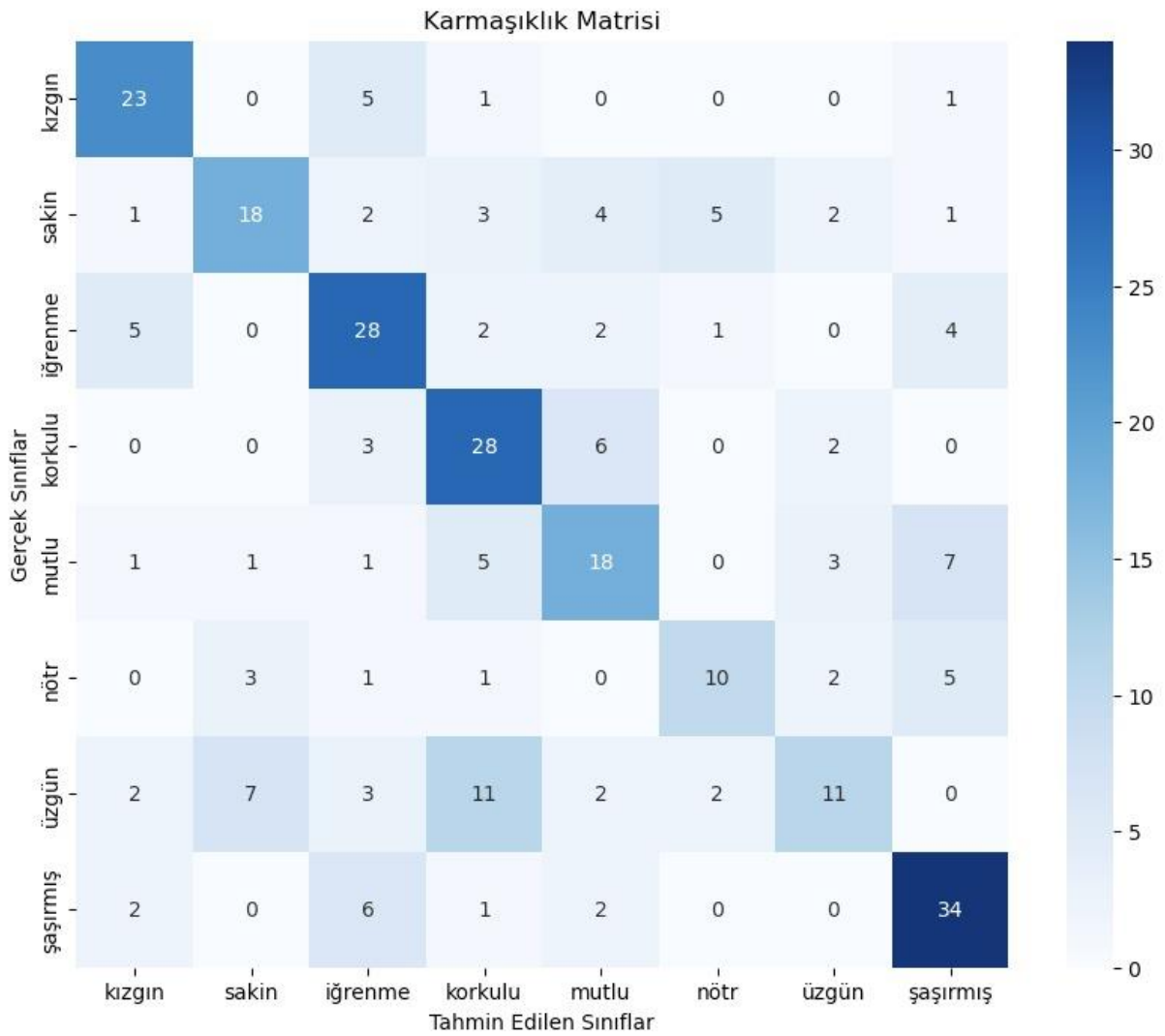
#### **6.1.1.1 Konuşma tipi ses sinyali, MFCC öznitelik çıkartıcısı, Dense katmanları**

MFCC tabanlı konuşma tipi ses sinyali duygu tanıma başlığı altındaki birinci uygulamada konuşma tipi ses sinyallerinden MFCC öznitelikleri çıkartılmış olup yapay sinir ağı modelinde dense katmanları kullanılmıştır. Tablo 6.1, bu uygulamada ki sınıflama raporunu göstermektedir. Şekil 6.1, bu uygulamada ki karışıklık matrisini göstermektedir.

|              | Precision | Recall | F1-score | support |
|--------------|-----------|--------|----------|---------|
| kızgın       | 0.68      | 0.77   | 0.72     | 30      |
| sakin        | 0.62      | 0.50   | 0.55     | 36      |
| iğrenme      | 0.57      | 0.67   | 0.62     | 42      |
| korkulu      | 0.54      | 0.72   | 0.62     | 39      |
| mutlu        | 0.53      | 0.50   | 0.51     | 36      |
| nötr         | 0.56      | 0.45   | 0.50     | 22      |
| üzgün        | 0.55      | 0.29   | 0.38     | 38      |
| şaşırmış     | 0.65      | 0.76   | 0.70     | 45      |
| accuracy     |           |        | 0.59     | 288     |
| Macro avg    | 0.59      | 0.58   | 0.57     | 288     |
| Weighted avg | 0.59      | 0.59   | 0.58     | 288     |

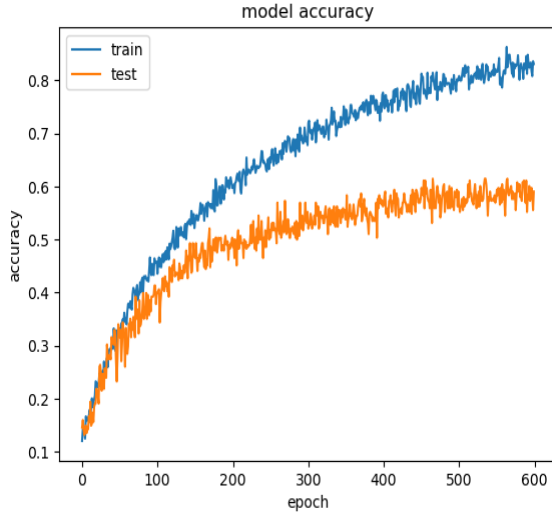
Tablo 6.1: Konuşma, MFCC, Dense Uygulaması Sınıflama Raporu

Tablo 6.1'e göre modelin sınıflama başarı oranı 0.59'dur. En iyi sınıflanan duygu 0.72 ile kızgın, en düşük oran ile sınıflanan duygu 0.38 ile üzgün olmuştur.

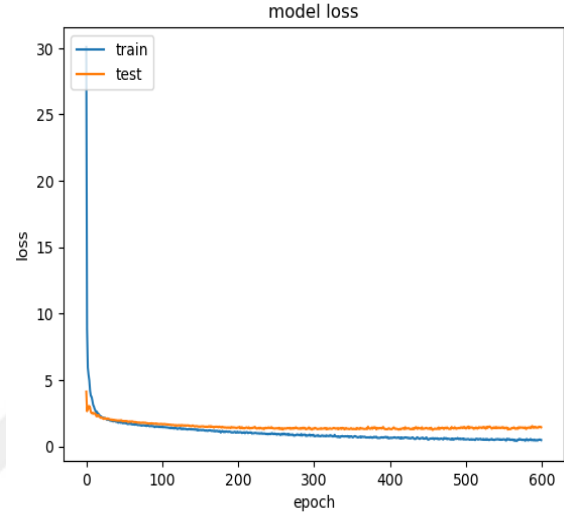


Şekil 6.1: Konuşma, MFCC, Dense Uygulaması Karışıklık Matrisi

Karışıklık matrisine göre model en çok hatayı üzgün sınıfını korkulu olarak tahmin ederek yapmıştır. Şekil 6.2’de bu uygulamanın model öğrenme eğrisi, şekil 6.3’de bu uygulamanın kayıp eğrisi gösterilmektedir.



Şekil 6.2: Konuşma, MFCC, Dense Uygulaması Doğruluk Grafiği



Şekil 6.3: Konuşma, MFCC, Dense Uygulaması Kayıp Grafiği

Şekil 6.2’deki ve şekil 6.3’deki öğrenme ve kayıp eğrilerine bakarak eğitim ve test setindeki kayıpların oldukça kabul edilebilir bir seviyede olduğu, konuşma ses sinyalleri, mfcc ve dense uygulamasının başarısının ideal düzeylerde olduğunu görülmektedir.

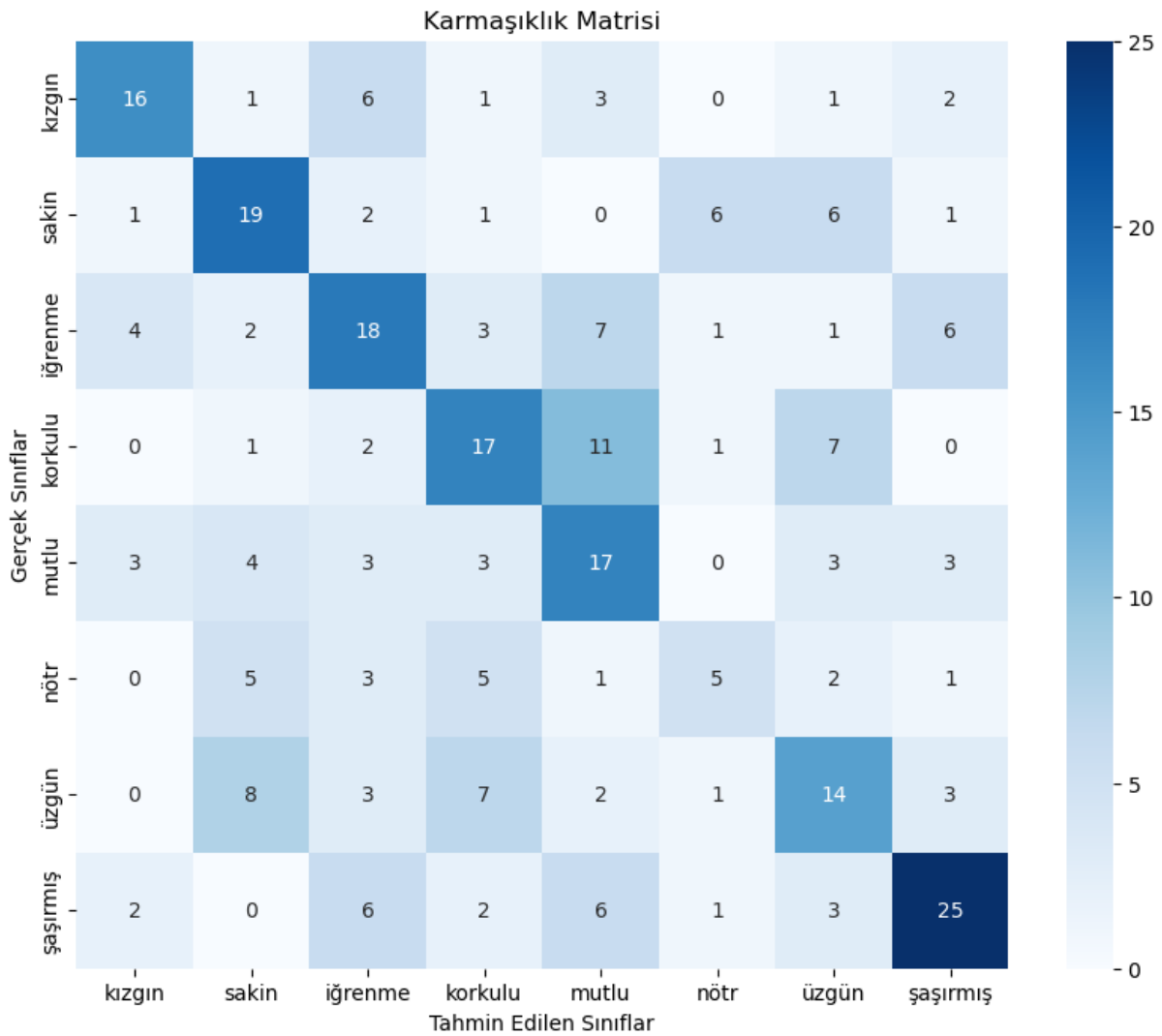
#### 6.1.1.2 Konuşma tipi ses sinyali, MFCC öznitelik çıkartıcısı, LSTM katmanları

MFCC tabanlı konuşma tipi ses sinyali duygu tanıma başlığı altındaki ikinci uygulamada konuşma ses sinyallerinden MFCC öznitelikleri çıkartılmış olup yapay sinir ağı modelinde LSTM katmanları kullanılmıştır. Tablo 6.2, bu uygulamada ki sınıflama raporunu göstermektedir. Şekil 6.4, bu uygulamada ki karışıklık matrisini göstermektedir.

|              | Precision | Recall | F1-score | support |
|--------------|-----------|--------|----------|---------|
| kızgın       | 0.62      | 0.53   | 0.57     | 30      |
| sakin        | 0.47      | 0.53   | 0.50     | 36      |
| igrenme      | 0.42      | 0.43   | 0.42     | 42      |
| korkulu      | 0.44      | 0.44   | 0.44     | 39      |
| mutlu        | 0.36      | 0.47   | 0.41     | 36      |
| nötr         | 0.33      | 0.23   | 0.27     | 22      |
| üzgün        | 0.38      | 0.37   | 0.37     | 38      |
| şaşırmış     | 0.61      | 0.56   | 0.58     | 45      |
| accuracy     |           |        | 0.45     | 288     |
| Macro avg    | 0.45      | 0.44   | 0.45     | 288     |
| Weighted avg | 0.46      | 0.45   | 0.45     | 288     |

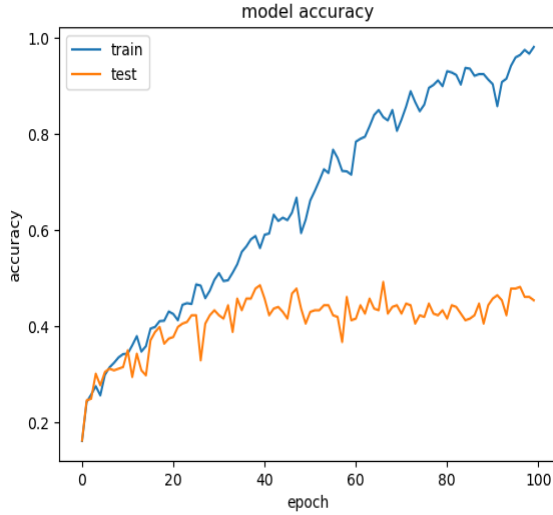
Tablo 6.2: Konuşma, MFCC, LSTM Uygulaması Sınıflama Raporu

Tablo 6.2'ye göre modelin sınıflama başarı oranı 0.45'dir. En iyi sınıflanan duygu 0.58 ile şaşırmış, en düşük oran ile sınıflanan duygu 0.27 ile nötr olmuştur.

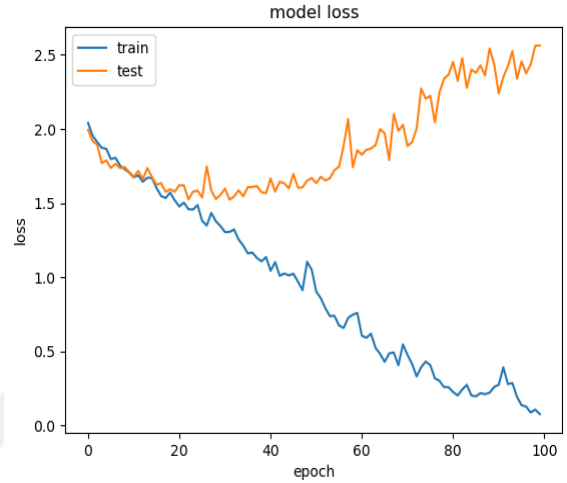


Şekil 6.4: Konuşma, MFCC, LSTM Uygulaması Karmaşıklık Matrisi

Karışıklık matrisine göre model en çok hatayı korkulu sınıfını mutlu olarak tahmin ederek yapmıştır. Şekil 6.5’de bu uygulamanın model öğrenme eğrisi, şekil 6.6’da bu uygulamanın kayıp eğrisi gösterilmektedir.



Şekil 6.5: Konuşma, MFCC, LSTM Uygulaması Doğruluk Grafiği



Şekil 6.6: Konuşma, MFCC, LSTM Uygulaması Kayıp Grafiği

Şekil 6.5’deki ve şekil 6.6’daki öğrenme ve kayıp eğrilerine bakarak 30. epoch’dan sonra eğitim kayıp oranı düşmeye devam etse de test setindeki kayıpların artışı modelin aşırı uyuma eğilimli olduğunu göstermektedir.

## 6.1.2 MFCC tabanlı şarkı tipi ses sinyali duygu tanıma

MFCC tabanlı şarkı tipi ses sinyali duygu tanıma başlığı altında 2 farklı uygulama gerçekleştirilmiştir. Birinci uygulama şarkı tipi ses sinyali, MFCC öznitelik çıkartıcısı ve Dense katmanları kullanılmıştır. İkinci uygulamada ise şarkı tipi ses sinyalleri, MFCC öznitelik çıkartıcısı ve LSTM katmanları kullanılmıştır.

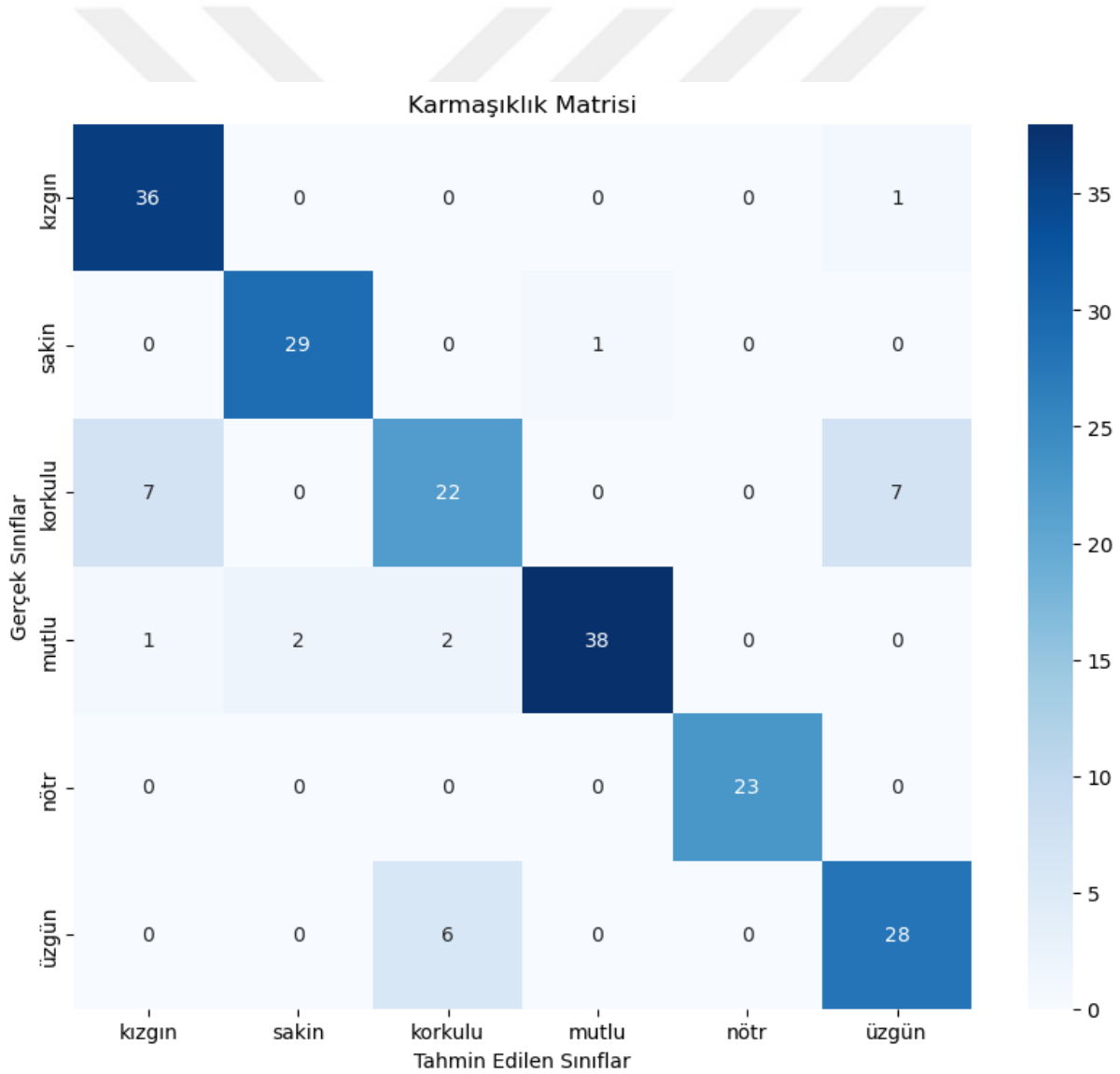
### 6.1.2.1 Şarkı tipi ses sinyali, MFCC öznitelik çıkartıcısı, Dense katmanları

MFCC tabanlı şarkı tipi ses sinyali duygu tanıma başlığı altındaki birinci uygulamada şarkı tipi ses sinyallerinden MFCC öznitelikleri çıkartılmış olup yapay sinir ağı modelinde dense katmanları kullanılmıştır. Tablo 6.3, bu uygulamada ki sınıflama raporunu göstermektedir. Şekil 6.7, bu uygulamada ki karışıklık matrisini göstermektedir.

|              | Precision | Recall | F1-score | support |
|--------------|-----------|--------|----------|---------|
| kızgın       | 0.82      | 0.97   | 0.89     | 37      |
| sakin        | 0.94      | 0.97   | 0.95     | 30      |
| korkulu      | 0.73      | 0.61   | 0.67     | 36      |
| mutlu        | 0.97      | 0.88   | 0.93     | 43      |
| nötr         | 1.00      | 1.00   | 1.00     | 23      |
| üzgün        | 0.78      | 0.82   | 0.80     | 34      |
| accuracy     |           |        | 0.87     | 203     |
| Macro avg    | 0.87      | 0.88   | 0.87     | 203     |
| Weighted avg | 0.87      | 0.87   | 0.86     | 203     |

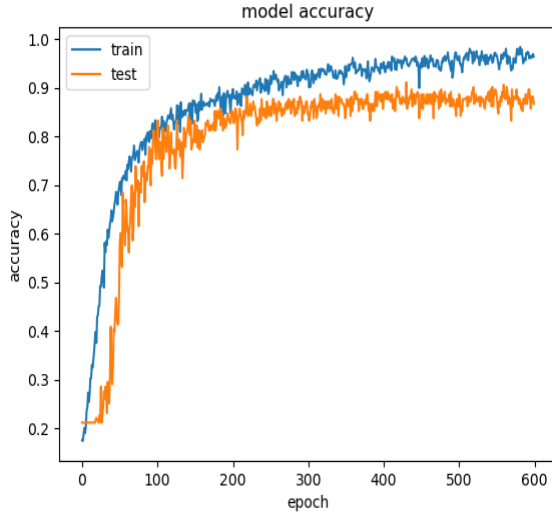
Tablo 6.3: Şarkı, MFCC, Dense Uygulaması Sınıflama Raporu

Tablo 6.3'e göre modelin sınıflama başarı oranı 0.87'dir. En iyi sınıflanan duygu 1.00 ile nötr, en düşük başarı ile sınıflanan duygu 0.67 ile korkulu olmuştur.

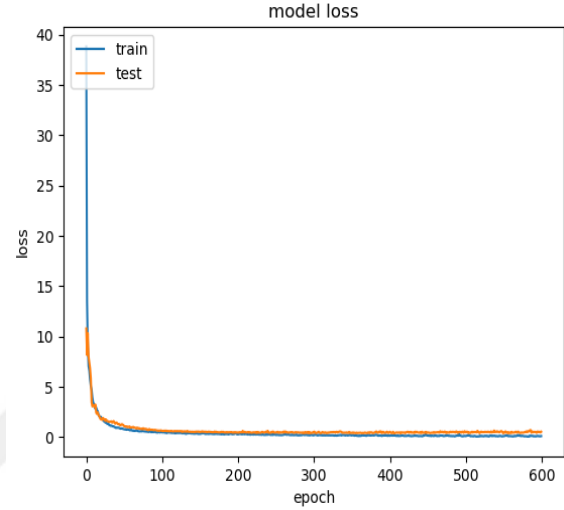


Şekil 6.7: Şarkı, MFCC, Dense Uygulaması Karmaşıklık Matrisi

Karışıklık matrisine göre model en çok hatayı korkulu sınıfını kızgın ve üzgün olarak tahmin ederek yapmıştır. Şekil 6.8’de bu uygulamanın model öğrenme eğrisi, şekil 6.9’da bu uygulamanın kayıp eğrisi gösterilmektedir.



Şekil 6.8: Şarkı, MFCC, Dense Uygulaması Doğruluk Grafiği



Şekil 6.9: Şarkı, MFCC, Dense Uygulaması Kayıp Grafiği

Şekil 6.8’deki ve şekil 6.9’daki öğrenme ve kayıp eğrilerine bakarak eğitim ve test setindeki kayıpların oldukça iyi bir seviyede olduğu, şarkı tipi ses sinyalleri, mfcc ve dense uygulamasının başarısının çok iyi düzeylerde olduğunu görülmektedir.

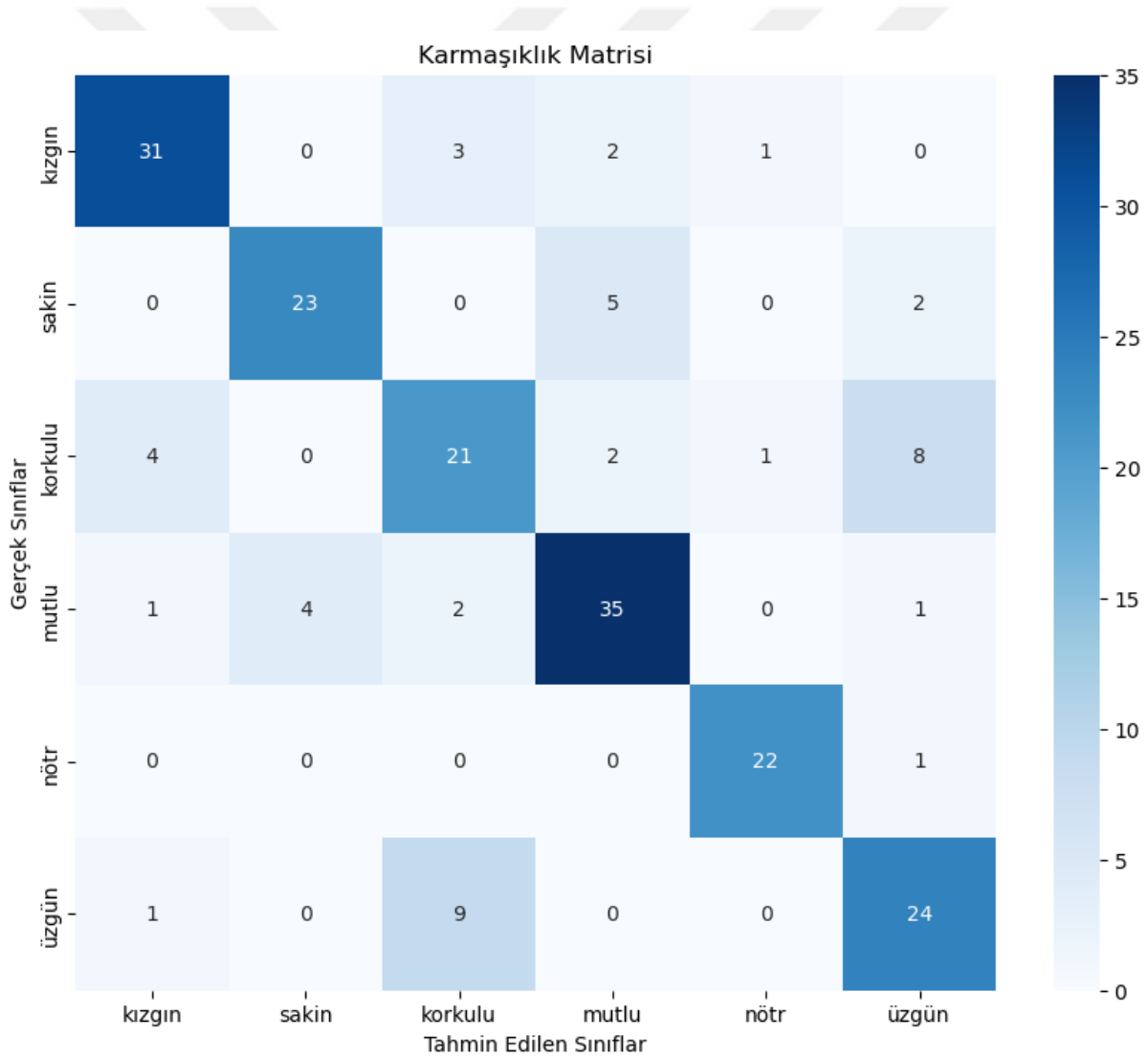
#### 6.1.2.2 Şarkı tipi ses sinyali, MFCC öznitelik çıkartıcısı, LSTM katmanları

MFCC tabanlı şarkı tipi ses sinyali duygu tanıma başlığı altındaki ikinci uygulamada şarkı tipi ses sinyallerinden MFCC öznitelikleri çıkartılmış olup yapay sinir ağı modelinde LSTM katmanları kullanılmıştır. Tablo 6.4, bu uygulamada ki sınıflama raporunu göstermektedir. Şekil 6.10, bu uygulamada ki karışıklık matrisini göstermektedir.

|              | Precision | Recall | F1-score | support |
|--------------|-----------|--------|----------|---------|
| kızgın       | 0.84      | 0.84   | 0.84     | 37      |
| sakin        | 0.85      | 0.77   | 0.81     | 30      |
| korkulu      | 0.60      | 0.58   | 0.59     | 36      |
| mutlu        | 0.80      | 0.81   | 0.80     | 43      |
| nötr         | 0.92      | 0.96   | 0.94     | 23      |
| üzgün        | 0.67      | 0.71   | 0.69     | 34      |
| accuracy     |           |        | 0.77     | 203     |
| Macro avg    | 0.78      | 0.88   | 0.78     | 203     |
| Weighted avg | 0.77      | 0.87   | 0.77     | 203     |

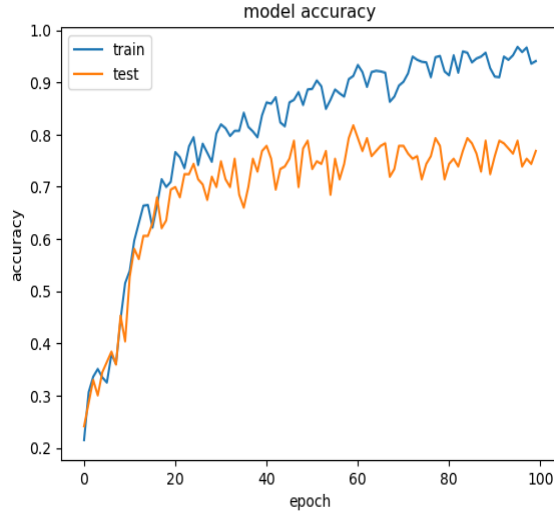
Tablo 6.4: Şarkı, MFCC, LSTM Uygulaması Sınıflama Raporu

Tablo 6.4'e göre modelin sınıflama başarı oranı 0.77'dir. En iyi sınıflanan duygu 0.94 ile nötr, en düşük oran ile sınıflanan duygu 0.59 ile korkulu olmuştur.

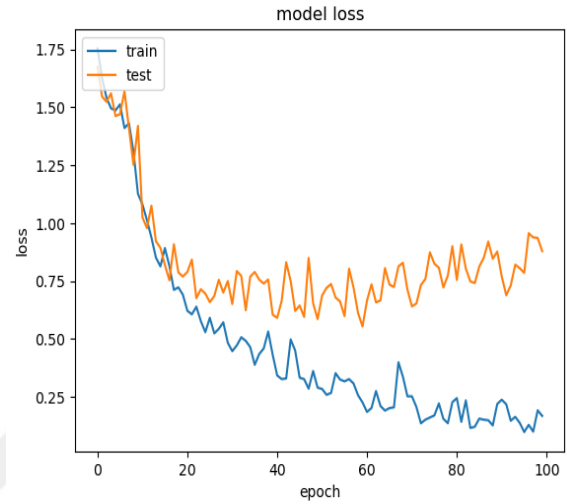


Şekil 6.10: Şarkı, MFCC, LSTM Uygulaması Karmaşıklık Matrisi

Karışıklık matrisine göre model en çok hatayı üzgün sınıfını korkulu olarak tahmin ederek yapmıştır. Şekil 6.11’de bu uygulamanın model öğrenme eğrisi, şekil 6.12’de bu uygulamanın kayıp eğrisi gösterilmektedir.



Şekil 6.11: Şarkı, MFCC, LSTM Uygulaması Doğruluk Grafiği



Şekil 6.12: Şarkı, MFCC, LSTM Uygulaması Kayıp Grafiği

Şekil 6.11’deki ve şekil 6.12’deki öğrenme ve kayıp eğrilerine bakarak 30. epoch’dan sonra eğitim kayıp oranı düşmeye devam etse de test setindeki kayıpların artışta olması modelin aşırı uyuma eğilimli olduğunu göstermektedir yine de 6.1.1.2 uygulamasına kıyasla başarının daha iyi bir seviyede olması uygulamada kullanılan veri setinin önemini göstermektedir.

## 6.2 LMFB tabanlı ses sinyali duygu tanıma

LMFB tabanlı ses sinyali duygu tanıma başlığı altında 4 farklı uygulama gerçekleştirilmiştir. Birinci uygulamada konuşma tipi ses sinyali, LMFB öznitelik çıkartıcısı ve Dense katmanları kullanılmıştır. İkinci uygulamada konuşma tipi ses sinyalleri, LMFB öznitelik çıkartıcısı ve LSTM katmanları kullanılmıştır. Üçüncü uygulamada şarkı tipi ses sinyalleri, LMFB öznitelik çıkartıcısı ve Dense katmanları kullanılmıştır. Dördüncü uygulamada ise şarkı tipi ses sinyalleri, LMFB öznitelik çıkartıcısı ve LSTM katmanları kullanılmıştır. LMFB tabanlı ses sinyali duygu tanıma başlığı altında öncelikle konuşma tipi ses sinyalleri sonrada şarkı tipi ses sinyallerini içeren uygulamalar gerçekleştirilecektir

### 6.2.1 LMFB tabanlı konuşma tipi ses sinyali duygu tanıma

LMFB tabanlı konuşma tipi ses sinyali duygu tanıma başlığı altında 2 farklı uygulama gerçekleştirilmiştir. Birinci uygulama konuşma tipi ses sinyali, LMFB öznitelik çıkartıcısı ve

Dense katmanları kullanılmıştır. İkinci uygulamada ise konuşma tipi ses sinyalleri, LMFB öznitelik çıkartıcısı ve LSTM katmanları kullanılmıştır.

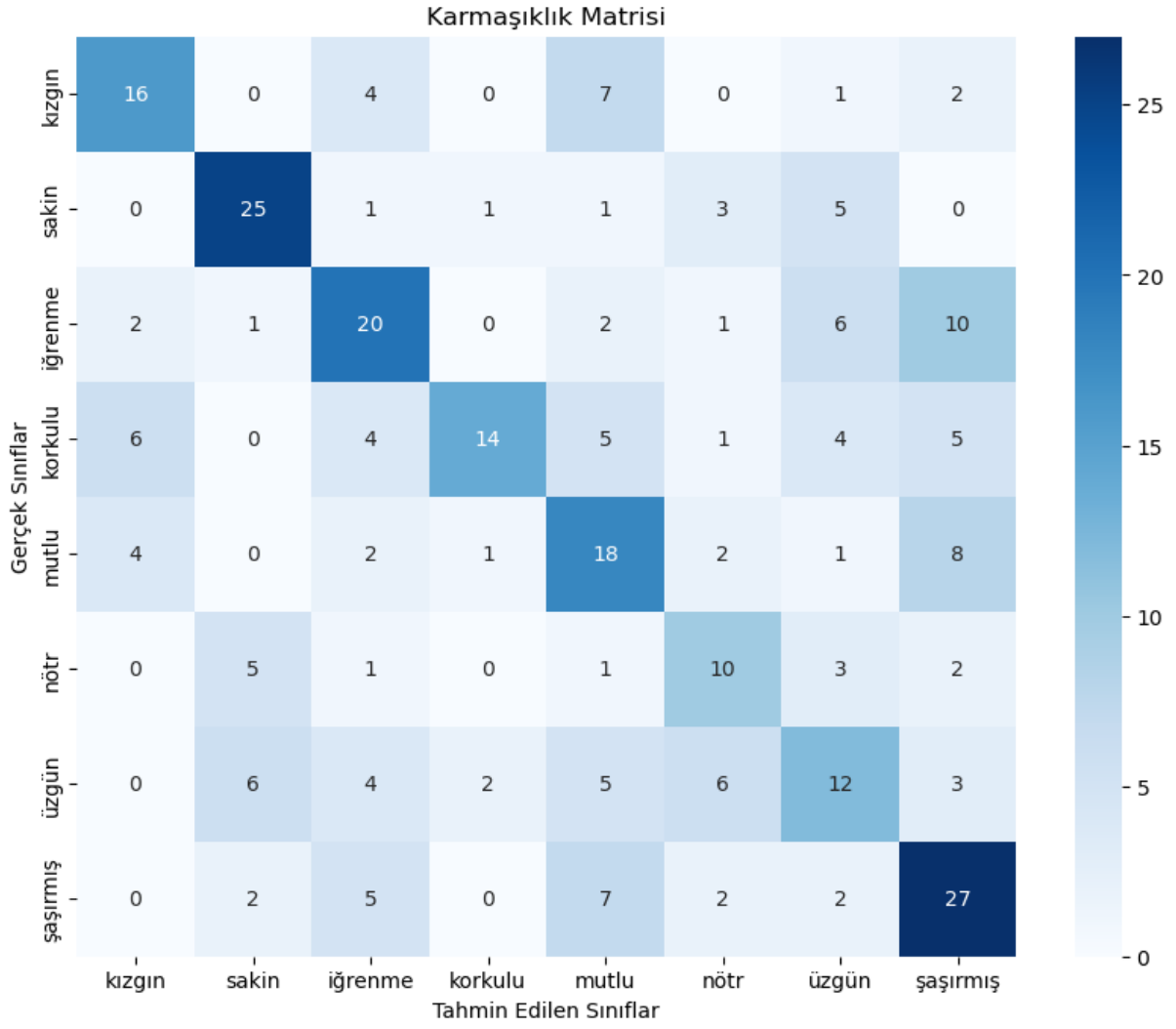
#### 6.2.1.1 Konuşma tipi ses sinyali, LMFB öznitelik çıkartıcısı, Dense katmanları

LMFB tabanlı konuşma tipi ses sinyali duygu tanıma başlığı altındaki birinci uygulamada konuşma tipi ses sinyallerinden LMFB öznitelikleri çıkartılmış olup yapay sinir ağı modelinde Dense katmanları kullanılmıştır. Tablo 6.5, bu uygulamada ki sınıflama raporunu göstermektedir. Şekil 6.13, bu uygulamada ki karışıklık matrisini göstermektedir.

|              | Precision | Recall | F1-score | support |
|--------------|-----------|--------|----------|---------|
| kızgın       | 0.57      | 0.53   | 0.55     | 30      |
| sakin        | 0.64      | 0.69   | 0.67     | 36      |
| iğrenme      | 0.49      | 0.48   | 0.48     | 42      |
| korkulu      | 0.78      | 0.36   | 0.49     | 39      |
| mutlu        | 0.39      | 0.50   | 0.44     | 36      |
| nötr         | 0.40      | 0.45   | 0.43     | 22      |
| üzgün        | 0.35      | 0.32   | 0.33     | 38      |
| şaşırmış     | 0.47      | 0.60   | 0.53     | 45      |
| accuracy     |           |        | 0.49     | 288     |
| Macro avg    | 0.51      | 0.49   | 0.49     | 288     |
| Weighted avg | 0.52      | 0.49   | 0.49     | 288     |

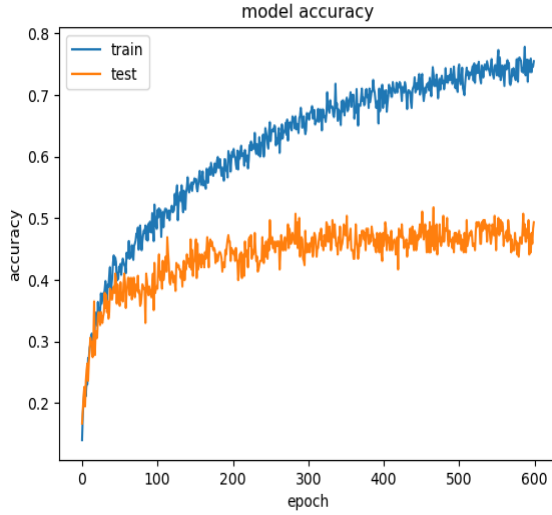
Tablo 6.5: Konuşma, LMFB, Dense Uygulaması Sınıflama Raporu

Tablo 6.5'e göre modelin sınıflama başarı oranı 0.49'dur En iyi sınıflanan duygu 0.67 ile sakın, en düşük oran ile sınıflanan duygu 0.33 ile üzgün olmuştur.

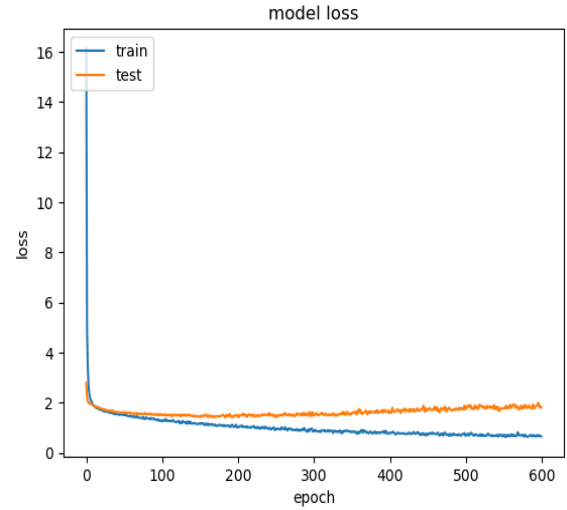


Şekil 6.5: Konuşma, LMFB, Dense Uygulaması Karmaşıklık Matrisi

Karmaşıklık matrisine göre model en çok hatayı iğrenme sınıfını şaşırmış olarak tahmin ederek yapmıştır. Şekil 6.14’de bu uygulamanın model öğrenme eğrisi, şekil 6.15’de bu uygulamanın kayıp eğrisi gösterilmektedir.



Şekil 6.14: Konuşma, LMFB, Dense Uygulaması Doğruluk Grafiği



Şekil 6.15: Konuşma, LMFB, Dense Uygulaması Kayıp Grafiği

Şekil 6.14'deki ve şekil 6.15'deki öğrenme ve kayıp eğrilerine bakarak eğitim ve test setindeki kayıpların oldukça kabul edilebilir bir seviyede olduğu, “konuşma ses sinyalleri, lmfb ve dense” uygulamasının başarısının ideal düzeylerde olduğunu görülmektedir.

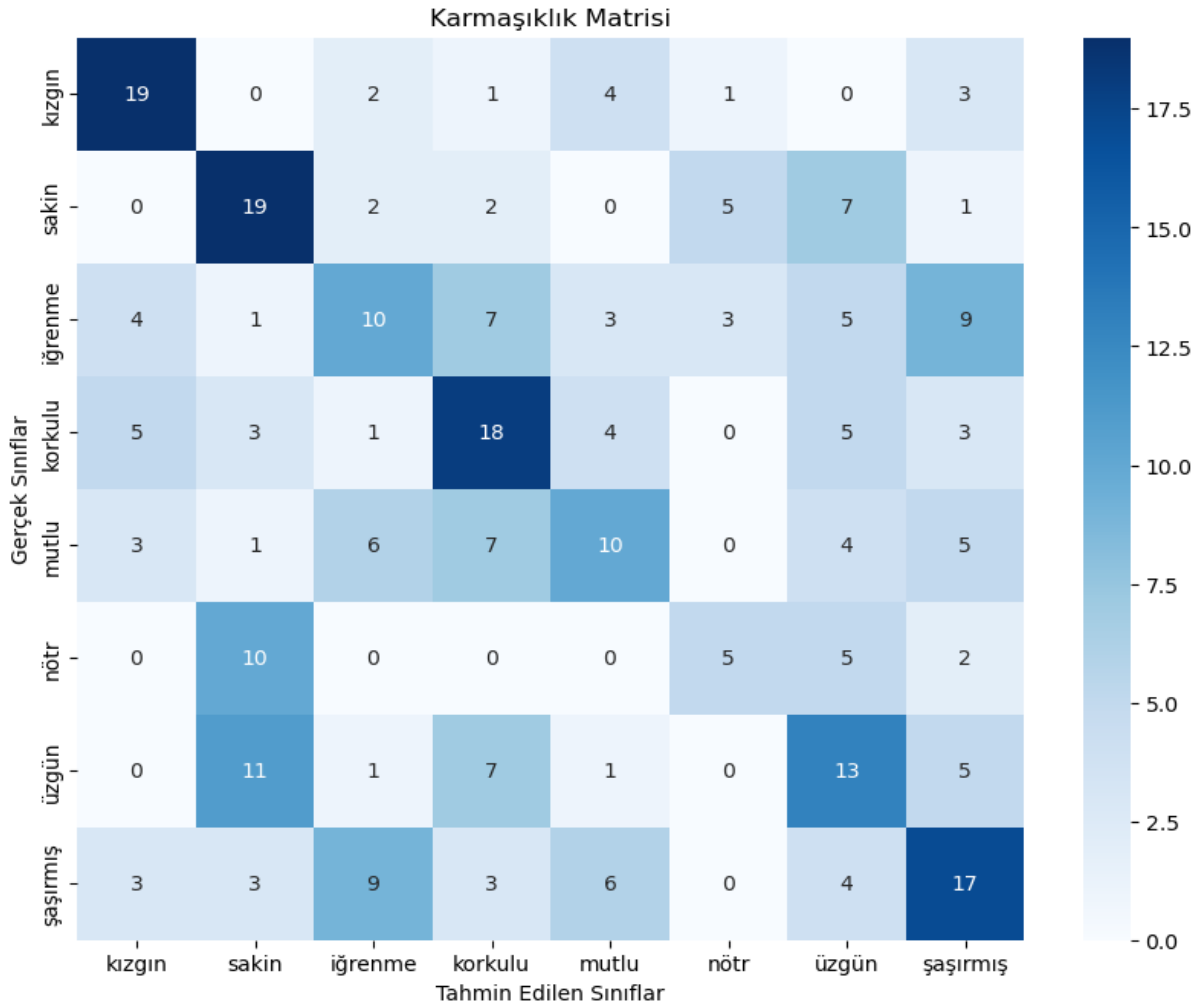
#### 6.2.1.2 Konuşma tipi ses sinyali, LMFB öznitelik çıkartıcısı, LSTM katmanları

LMFB tabanlı konuşma tipi ses sinyali duygu tanıma başlığı altındaki ikinci uygulamada konuşma tipi ses sinyallerinden LMFB öznitelikleri çıkartılmış olup yapay sinir ağı modelinde LSTM katmanları kullanılmıştır. Tablo 6.6, bu uygulamada ki sınıflama raporunu göstermektedir. Şekil 6.16, bu uygulamada ki karışıklık matrisini göstermektedir.

|              | Precision | Recall | F1-score | support |
|--------------|-----------|--------|----------|---------|
| kızgın       | 0.56      | 0.63   | 0.59     | 30      |
| sakin        | 0.40      | 0.53   | 0.45     | 36      |
| iğrenme      | 0.32      | 0.24   | 0.27     | 42      |
| korkulu      | 0.40      | 0.46   | 0.43     | 39      |
| mutlu        | 0.36      | 0.28   | 0.31     | 36      |
| nötr         | 0.36      | 0.23   | 0.28     | 22      |
| üzgün        | 0.30      | 0.34   | 0.32     | 38      |
| şaşırmış     | 0.38      | 0.38   | 0.38     | 45      |
| accuracy     |           |        | 0.39     | 288     |
| Macro avg    | 0.38      | 0.39   | 0.38     | 288     |
| Weighted avg | 0.38      | 0.39   | 0.38     | 288     |

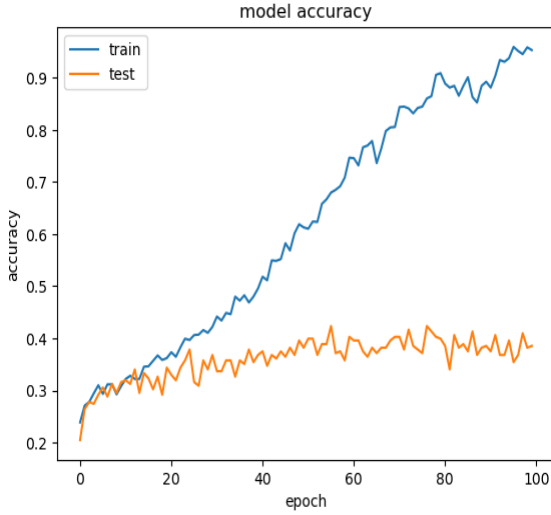
Tablo 6.6: Konuşma, LMFB, LSTM Uygulaması Sınıflama Raporu

Tablo 6.6'e göre modelin sınıflama başarısı 0.39, en iyi sınıflanan duygu 0.59 ile kızgın, en düşük oran ile sınıflanan duygu 0.27 ile öğrenme olmuştur.



Şekil 6.16: Konuşma, LMFB, LSTM Uygulaması Karmaşıklık Matrisi

Karmaşıklık matrisine göre model en çok hatayı üzgün sınıfını sakin olarak tahmin ederek yapmıştır. Şekil 6.17'de bu uygulamanın model öğrenme eğrisi, şekil 6.18'de bu uygulamanın kayıp eğrisi gösterilmektedir.



Şekil 6.17: Konuşma, LMFB, LSTM Uygulaması Doğruluk Grafiği



Şekil 6.18: Konuşma, LMFB, LSTM Uygulaması Kayıp Grafiği

Şekil 6.17'deki ve şekil 6.18'deki öğrenme ve kayıp eğrilerine bakarak 30. epoch'dan sonra eğitim kayıp oranı düşmeye devam etse de test setindeki kayıpların artışı modelin aşırı uyuma eğilimli olduğunu göstermektedir. Modelin başarısının uygulama 6.1.1.2'den düşük olması duygu sınıflamada öznelilik çıkartıcı seçiminin önemini göstermektedir.

## 6.2.2 LMFB tabanlı şarkı tipi ses sinyali duygu tanıma

LMFB tabanlı şarkı tipi ses sinyali duygu tanıma başlığı altında 2 farklı uygulama gerçekleştirilmiştir. Birinci uygulama şarkı tipi ses sinyali, LMFB öznelilik çıkartıcısı ve Dense katmanları kullanılmıştır. İkinci uygulamada ise şarkı tipi ses sinyalleri, LMFB öznelilik çıkartıcısı ve LSTM katmanları kullanılmıştır.

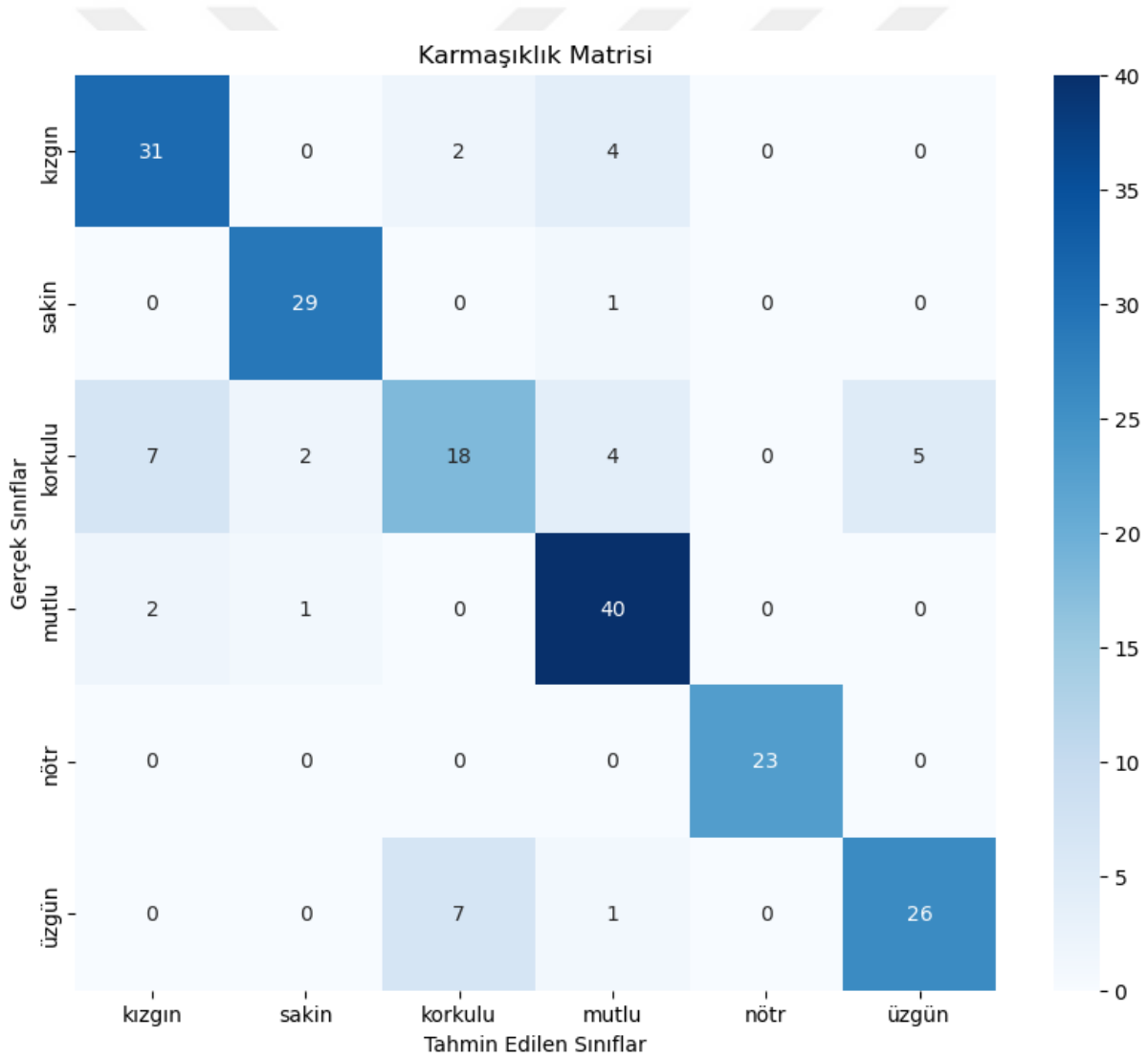
### 6.2.2.1 Şarkı tipi ses sinyali, LMFB öznelilik çıkartıcısı, Dense katmanları

LMFB tabanlı şarkı tipi ses sinyali duygu tanıma başlığı altındaki birinci uygulamada şarkı tipi ses sinyallerinden LMFB öznelilikleri çıkartılmış olup yapay sinir ağı modelinde dense katmanları kullanılmıştır. Tablo 6.7, bu uygulamada ki sınıflama raporunu göstermektedir. Şekil 6.19, bu uygulamada ki karışıklık matrisini göstermektedir.

|              | Precision | Recall | F1-score | support |
|--------------|-----------|--------|----------|---------|
| kızgın       | 0.78      | 0.84   | 0.81     | 37      |
| sakin        | 0.91      | 0.97   | 0.94     | 30      |
| korkulu      | 0.67      | 0.50   | 0.57     | 36      |
| mutlu        | 0.80      | 0.93   | 0.86     | 43      |
| nötr         | 1.00      | 1.00   | 1.00     | 23      |
| üzgün        | 0.84      | 0.76   | 0.80     | 34      |
| accuracy     |           |        | 0.82     | 203     |
| Macro avg    | 0.83      | 0.83   | 0.83     | 203     |
| Weighted avg | 0.82      | 0.82   | 0.82     | 203     |

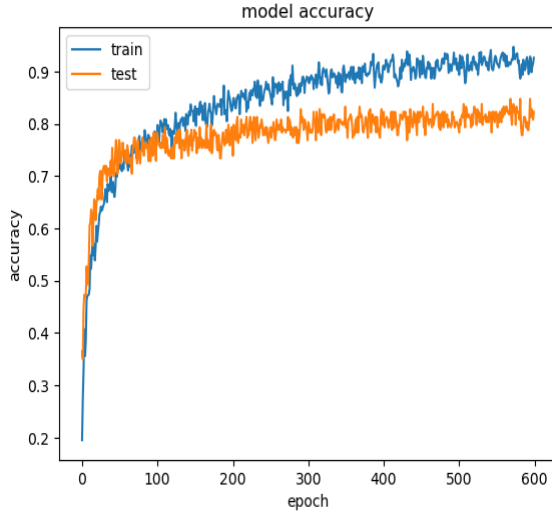
Tablo 6.7: Şarkı, LMFB, Dense Uygulaması Sınıflama Raporu

Tablo 6.7'ye göre modelin sınıflama başarı oranı 0.82'dir En iyi sınıflanan duygu 1.00 ile nötr, en düşük oran ile sınıflanan duygu 0.57 ile korkulu olmuştur.

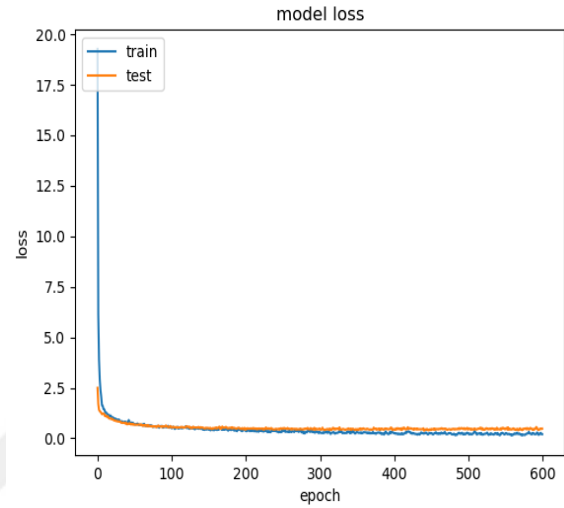


Şekil 6.19: Şarkı, LMFB, Dense Uygulaması Karmaşıklık Matrisi

Karışıklık matrisine göre model en çok hatayı korkulu sınıfını kızgın ve üzgün sınıfını korkulu olarak tahmin ederek yapmıştır. Şekil 6.20’de bu uygulamanın model öğrenme eğrisi, şekil 6.21’de bu uygulamanın kayıp eğrisi gösterilmektedir.



Şekil 6.20: Şarkı, LMFB, Dense Uygulaması Doğruluk Grafiği



Şekil 6.21: Şarkı, LMFB, Dense Uygulaması Kayıp Grafiği

Şekil 6.20’deki ve şekil 6.21’deki öğrenme ve kayıp eğrilerine bakarak eğitim ve test setindeki kayıpların oldukça kabul edilebilir bir seviyede olduğu, şarkı tipi ses sinyalleri, mfcc ve Dense uygulamasının başarısının ideal düzeylerde olduğunu görülmektedir.

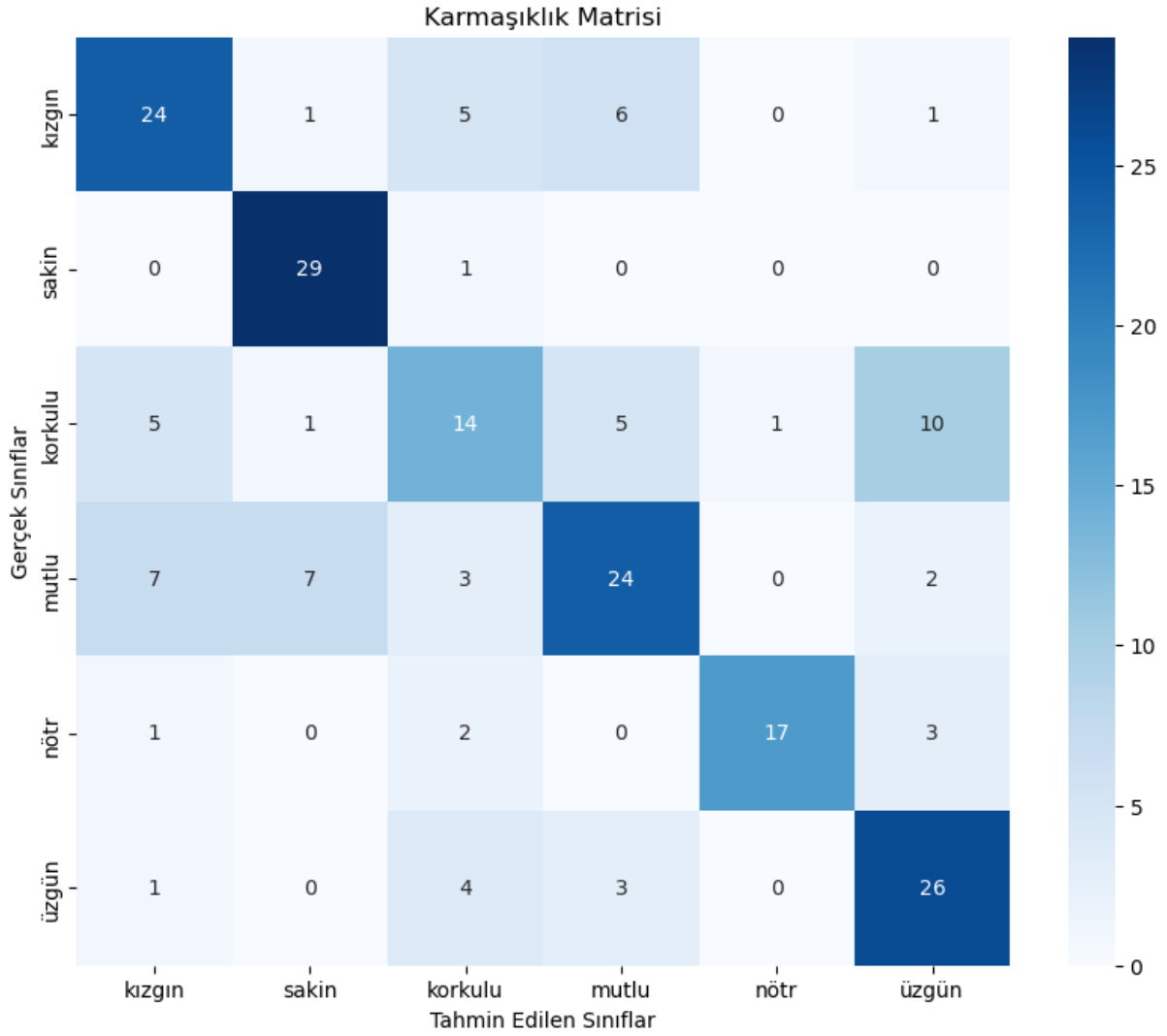
#### 6.2.2.2 Şarkı tipi ses sinyali, LMFB öznitelik çıkartıcısı, LSTM katmanları

LMFB tabanlı şarkı tipi ses sinyali duygu tanıma başlığı altındaki ikinci uygulamada şarkı tipi ses sinyallerinden LMFB öznitelikleri çıkartılmış olup yapay sinir ağı modelinde LSTM katmanları kullanılmıştır. Tablo 6.8, bu uygulamada ki sınıflama raporunu göstermektedir. Şekil 6.22, bu uygulamada ki karışıklık matrisini göstermektedir.

|              | Precision | Recall | F1-score | support |
|--------------|-----------|--------|----------|---------|
| kızgın       | 0.63      | 0.65   | 0.64     | 37      |
| sakin        | 0.76      | 0.97   | 0.85     | 30      |
| korkulu      | 0.48      | 0.39   | 0.43     | 36      |
| mutlu        | 0.63      | 0.56   | 0.59     | 43      |
| nötr         | 0.94      | 0.74   | 0.83     | 23      |
| üzgün        | 0.62      | 0.76   | 0.68     | 34      |
| accuracy     |           |        | 0.66     | 203     |
| Macro avg    | 0.68      | 0.68   | 0.67     | 203     |
| Weighted avg | 0.66      | 0.66   | 0.65     | 203     |

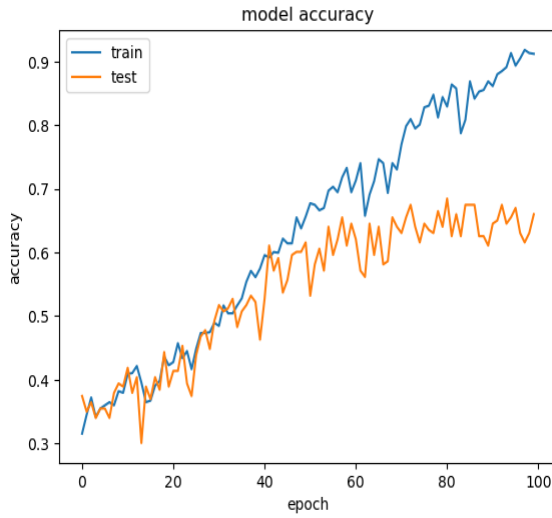
Tablo 6.8: Şarkı, LMFB, LSTM Uygulaması Sınıflama Raporu

Tablo 6.8'e göre modelin sınıflama başarı oranı 0.66'dır En iyi sınıflanan duygu 0.85 ile sakın, en düşük oran ile sınıflanan duygu 0.43 ile korkulu olmuştur.

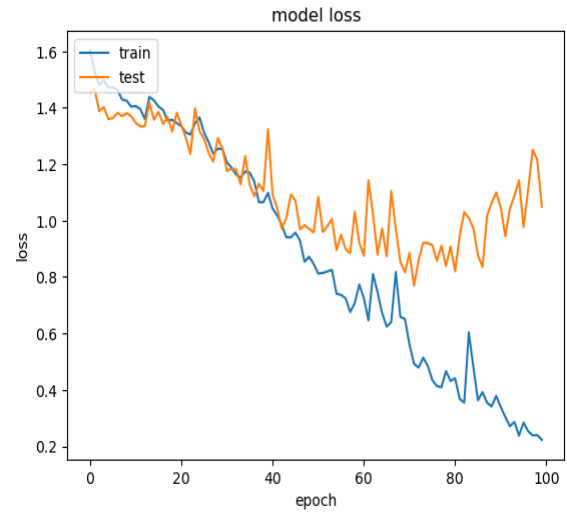


Şekil 6.22: Şarkı, LMFB, LSTM Uygulaması Karmaşıklık Matrisi

Karmaşıklık matrisine göre model en çok hatayı korkulu sınıfını üzgün olarak tahmin ederek yapmıştır. Şekil 6.23'de bu uygulamanın model öğrenme eğrisi, şekil 6.24'de bu uygulamanın kayıp eğrisi gösterilmektedir.



Şekil 6.23: Şarkı, LMFB, LSTM Uygulaması Doğruluk Grafiği



Şekil 6.24: Şarkı, LMFB, LSTM Uygulaması Kayıp Grafiği

Şekil 6.23'deki ve şekil 6.24'deki öğrenme ve kayıp eğrilerine bakarak 70. epoch'dan sonra eğitim kayıp oranı düşmeye devam etse de test setindeki kayıpların artışı modelin aşırı uyuma eğilimli olduğunu göstermektedir.

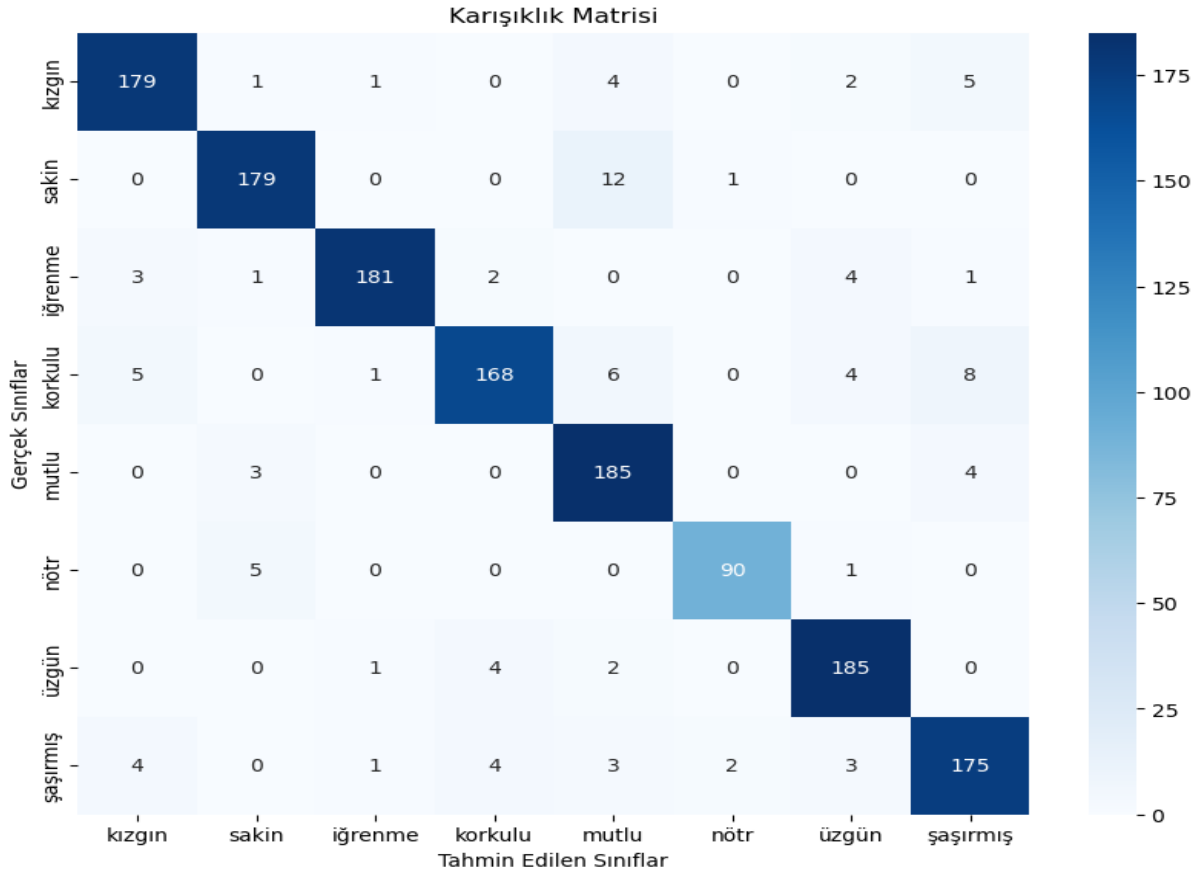
### 6.3 Görüntü tabanlı duygu tanıma için derin öğrenme modeli

Görüntü verilerinden duygu sınıflama gerçekleştirmek amacıyla CNN yapay sinir ağı modeli kullanılmıştır. Tablo 6.9, bu uygulamada ki sınıflama raporunu göstermektedir. Şekil 6.25, bu uygulamada ki karışıklık matrisini göstermektedir.

|              | Precision | Recall | F1-score | support |
|--------------|-----------|--------|----------|---------|
| kızgın       | 0.94      | 0.93   | 0.93     | 192     |
| sakin        | 0.95      | 0.93   | 0.94     | 192     |
| iğrenme      | 0.98      | 0.94   | 0.96     | 192     |
| korkulu      | 0.94      | 0.88   | 0.91     | 192     |
| mutlu        | 0.87      | 0.96   | 0.92     | 192     |
| nötr         | 0.97      | 0.94   | 0.95     | 96      |
| üzgün        | 0.93      | 0.96   | 0.95     | 192     |
| şaşırmış     | 0.91      | 0.91   | 0.91     | 192     |
| accuracy     |           |        | 0.93     | 1440    |
| Macro avg    | 0.94      | 0.93   | 0.93     | 1440    |
| Weighted avg | 0.93      | 0.93   | 0.93     | 1440    |

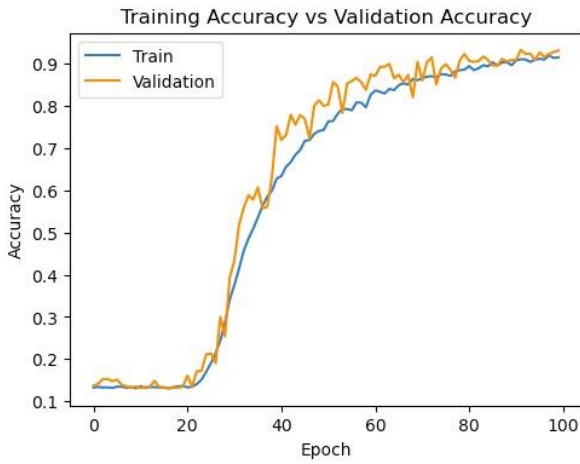
Tablo 6.9: Görüntü Verilerinden Duygu Tanıma Sınıflama Raporu

Tablo 6.9'a göre modelin sınıflama başarı oranı 0.93'dür. En iyi sınıflanan duygu 0.96 ile iğrenme, en düşük oran ile sınıflanan duygular 0.91 ile korkulu ve şaşırılmış olmuştur.

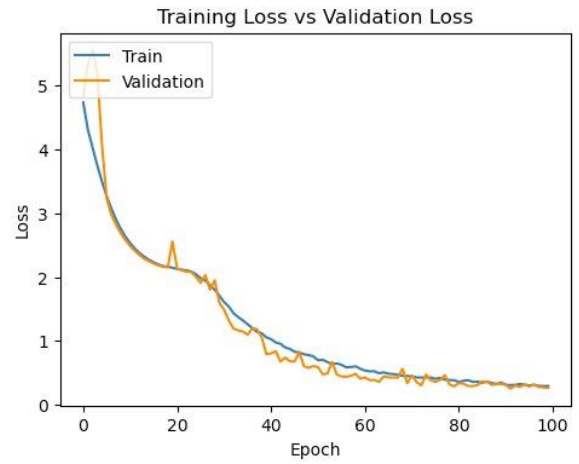


Şekil 6.25: Görüntü Sınıflandırma Uygulaması Karışıklık Matrisi

Karışıklık matrisine göre model en çok hatayı sakın sınıfını mutlu olarak tahmin ederek yapmıştır. Şekil 6.26'da bu uygulamanın model öğrenme eğrisi, şekil 6.27'de bu uygulamanın kayıp eğrisi gösterilmektedir.



Şekil 6.26: Görüntü Sınıflandırma Uygulaması Doğruluk Grafiği



Şekil 6.27: Görüntü Sınıflandırma Uygulaması Kayıp Grafiği

Şekil 6.26'daki ve şekil 6.27'deki öğrenme ve kayıp eğrilerine bakarak eğitim ve test setindeki kayıpların oldukça iyi bir seviyede olduğu, görüntü verilerinden duygu sınıflama uygulamasının başarısının çok iyi düzeylerde olduğunu görülmektedir.

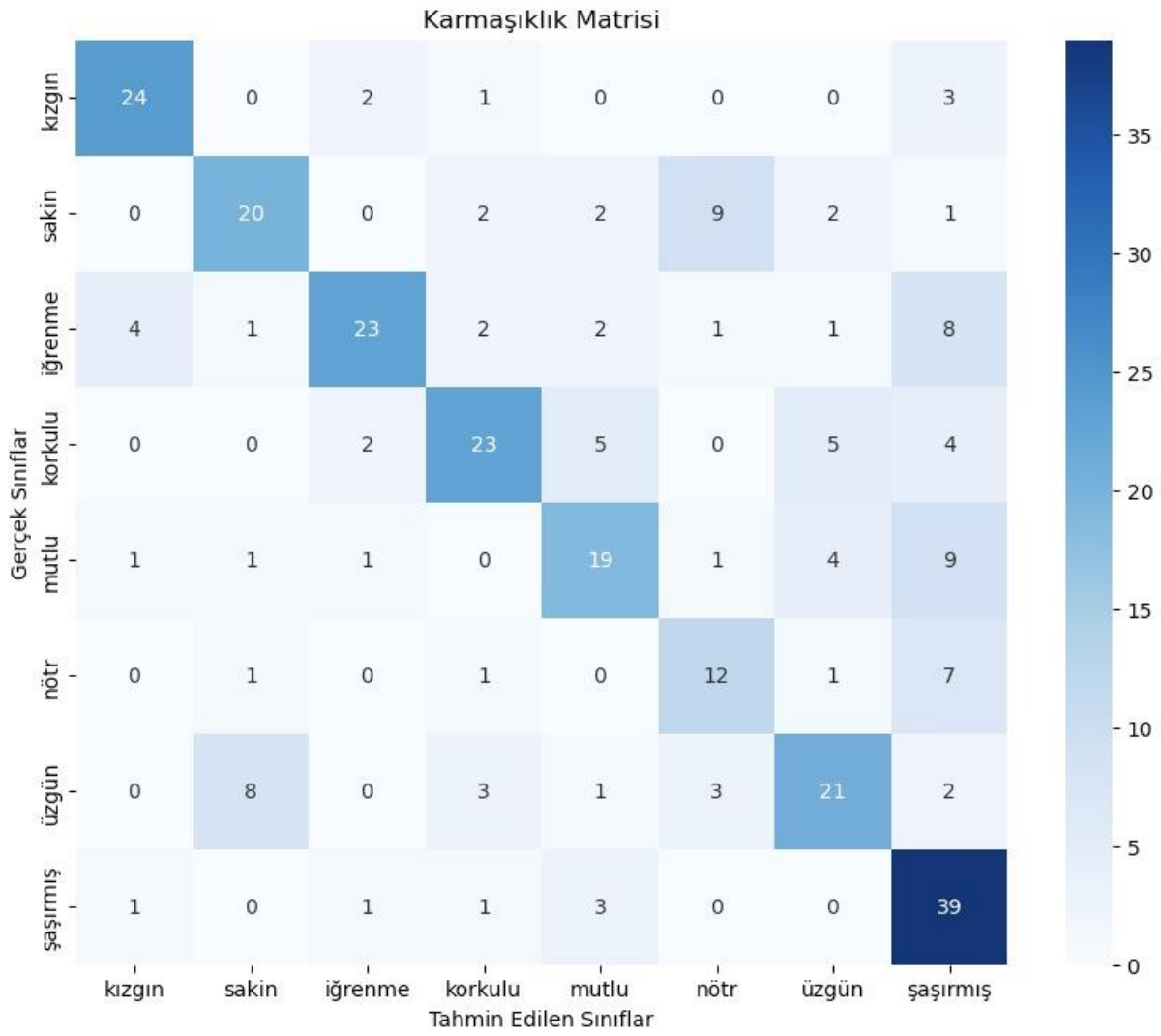
## 6.4 Duygu tanıma performansını arttırmak için transfer öğrenmesi

Görüntü verilerinden duygu sınıflama modelinden ses sinyallerinden duygu sınıflama modeline transfer öğrenmesi gerçekleştirilmiştir. Bu transfer öğrenmesi uygulaması sayesinde 6.1.1.1 uygulamasındaki %59 olan başarı oranı yüzde 6.78'lik artışla %63'e çıkmıştır. Tablo 6.10, bu uygulamada ki sınıflama raporunu göstermektedir. Şekil 6.28, bu uygulamada ki karışıklık matrisini göstermektedir.

|              | Precision | Recall | F1-score | support |
|--------------|-----------|--------|----------|---------|
| kızgın       | 0.80      | 0.80   | 0.80     | 30      |
| sakin        | 0.65      | 0.56   | 0.60     | 36      |
| igrenme      | 0.79      | 0.55   | 0.65     | 42      |
| korkulu      | 0.70      | 0.59   | 0.64     | 39      |
| mutlu        | 0.59      | 0.53   | 0.56     | 36      |
| nötr         | 0.46      | 0.55   | 0.50     | 22      |
| üzgün        | 0.62      | 0.55   | 0.58     | 38      |
| şaşırmış     | 0.53      | 0.87   | 0.66     | 45      |
| accuracy     |           |        | 0.63     | 288     |
| Macro avg    | 0.64      | 0.58   | 0.62     | 288     |
| Weighted avg | 0.65      | 0.59   | 0.63     | 288     |

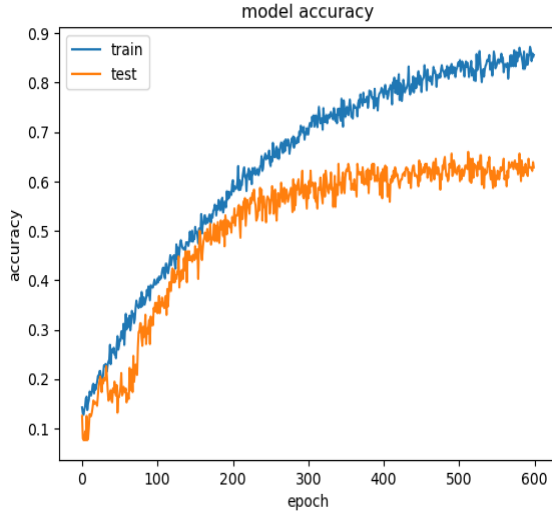
Tablo 6.10: Transfer Öğrenmesi ile Duygu Tanıma Sınıflama Raporu

Tablo 6.10'a göre modelin sınıflama başarı oranı 0.63'dür En iyi sınıflanan duygu 0.80 ile kızgın, en düşük oran ile sınıflanan duygu 0.50 ile nötr olmuştur.

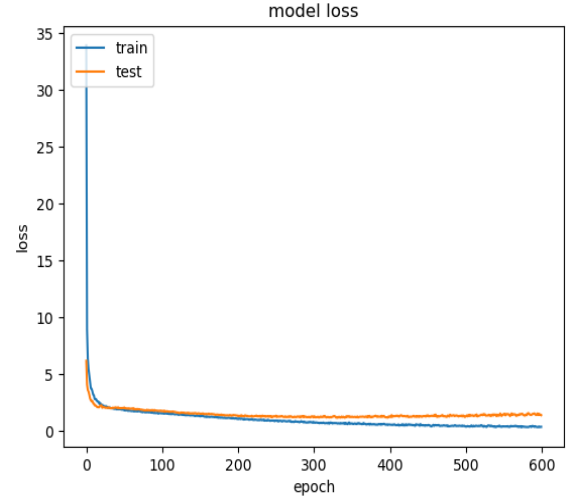


Şekil 6.28: Transfer Öğrenimi Uygulaması Karışıklık Matrisi

Karışıklık matrisine göre model en çok hatayı sakın sınıfını nötr ve mutlu sınıfını şaşırılmış olarak tahmin ederek yapmıştır. Şekil 6.29’da bu uygulamanın model öğrenme eğrisi, şekil 6.30’da bu uygulamanın kayıp eğrisi gösterilmektedir.



Şekil 6.29: Transfer Öğrenmesi Uygulaması Doğruluk Grafiği



Şekil 6.30: Transfer Öğrenmesi Uygulaması Kayıp Grafiği

Şekil 6.29'daki ve şekil 6.30'daki öğrenme ve kayıp eğrilerine bakarak eğitim ve test setindeki kayıpların oldukça kabul edilebilir bir seviyede olduğu transfer öğrenmesi uygulamasının başarısının ideal düzeylerde olduğunu görülmektedir.

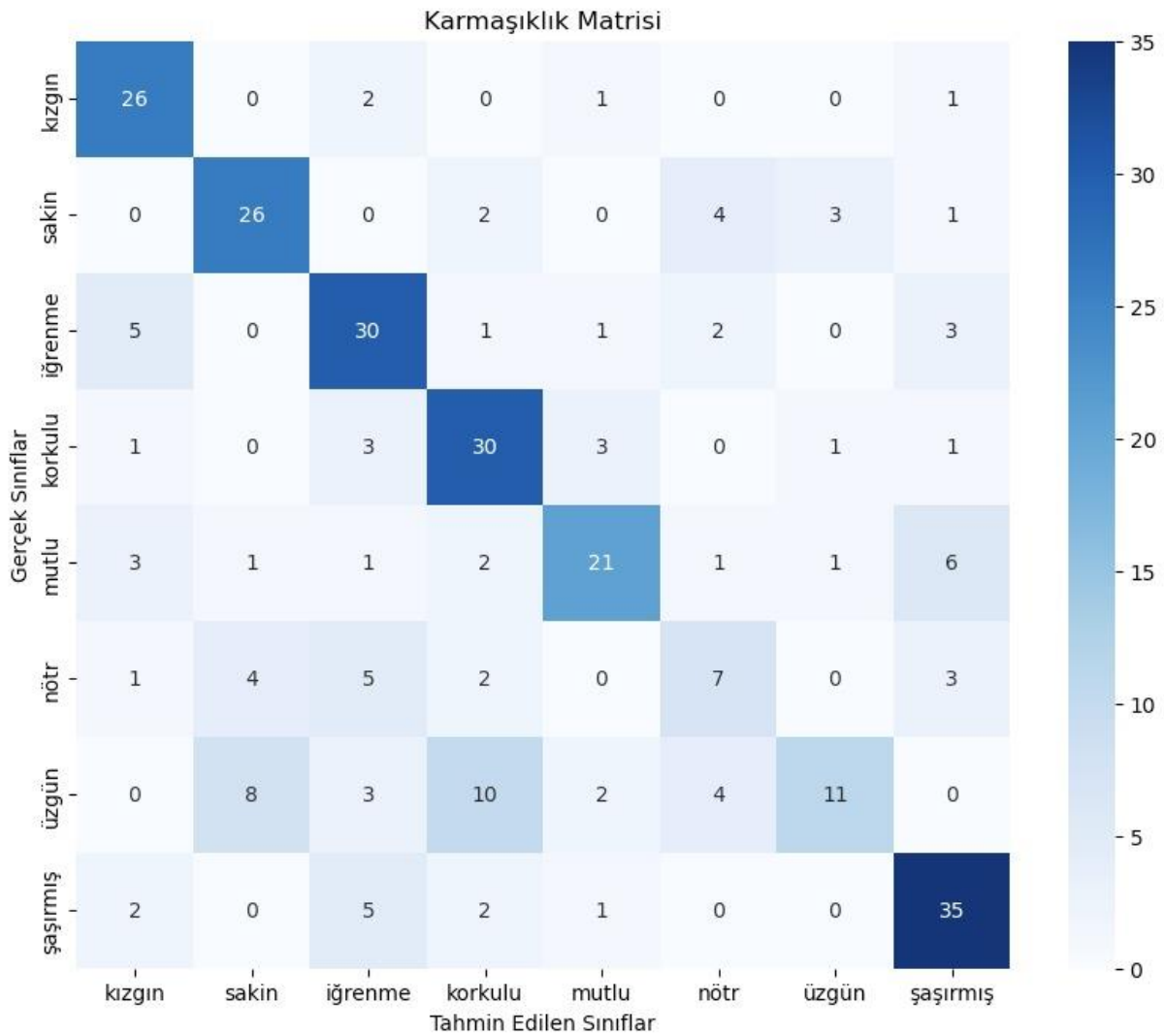
## 6.5 Duygu tanıma performansını arttırmak için erken birleştirme yöntemi

Görüntü verilerinden duygu sınıflama yapan modelden cnn öznitelikleri çıkartılıp, ses sinyallerinden duygu sınıflama yapan modelden daha model eğitilmeden önce mfcc özniteliklerini çıkarttıktan sonra öznitelik aktarımı yapılmıştır. Bu öznitelik aktarımı işleminden sonra ses sinyallerinden duygu sınıflayan model eğitilmiştir ve sonuç olarak modelin başarısının 6.1.1.1 uygulamasındaki %59 olan başarı oranı yüzde 10.1'lik artışla 0.65'e çıkmıştır. Tablo 6.11, bu uygulamada ki sınıflama raporunu göstermektedir. Şekil 6.31, bu uygulamada ki karışıklık matrisini göstermektedir.

|              | Precision | Recall | F1-score | support |
|--------------|-----------|--------|----------|---------|
| kızgın       | 0.68      | 0.87   | 0.76     | 30      |
| sakin        | 0.67      | 0.72   | 0.69     | 36      |
| iğrenme      | 0.61      | 0.71   | 0.66     | 42      |
| korkulu      | 0.61      | 0.77   | 0.68     | 39      |
| mutlu        | 0.72      | 0.58   | 0.65     | 36      |
| nötr         | 0.39      | 0.32   | 0.35     | 22      |
| üzgün        | 0.69      | 0.29   | 0.41     | 38      |
| şaşırmış     | 0.70      | 0.78   | 0.74     | 45      |
| accuracy     |           |        | 0.65     | 288     |
| Macro avg    | 0.63      | 0.63   | 0.62     | 288     |
| Weighted avg | 0.65      | 0.65   | 0.63     | 288     |

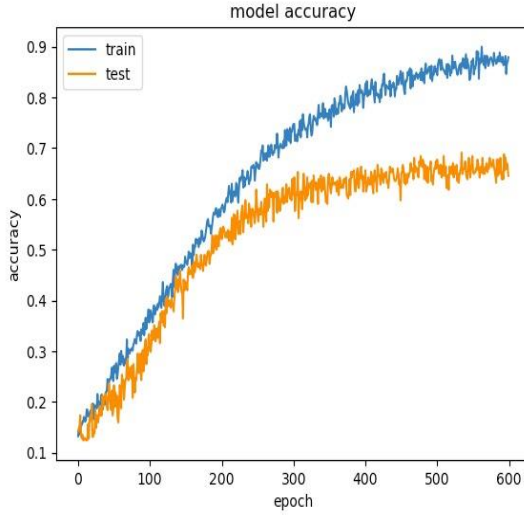
Tablo 6.11: Erken Birleştirme Yöntemi ile Duygu Tanıma Sınıflama Raporu

Tablo 6.11'e göre modelin sınıflama başarı oranı 0.65'dir En iyi sınıflanan duygu 0.76 ile kızgın, en düşük oran ile sınıflanan duygu 0.35 ile nötr olmuştur.

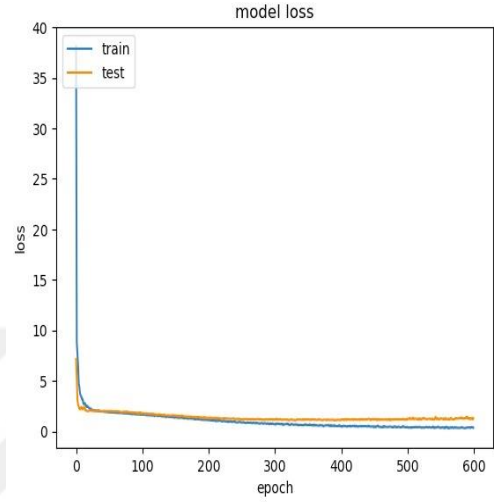


Şekil 6.31: Erken Birleştirme Uygulaması Karmaşıklık Matrisi

Karışıklık matrisine göre model en çok hatayı üzgün sınıfını korkulu olarak tahmin ederek yapmıştır. Şekil 6.32’de bu uygulamanın model öğrenme eğrisi, şekil 6.33’de bu uygulamanın kayıp eğrisi gösterilmektedir.



Şekil 6.32: Erken Birleştirme Uygulaması Doğruluk Grafiği



Şekil 6.33: Erken Birleştirme Uygulaması Kayıp Grafiği

Şekil 6.32’deki ve şekil 6.33’deki öğrenme ve kayıp eğrilerine bakarak eğitim ve test setindeki kayıpların oldukça kabul edilebilir bir seviyede olduğu erken birleştirme yöntemi uygulamasının başarısının ideal düzeylerde olduğunu görülmektedir.

## 7. SONUÇ

Bu tez çalışması kapsamında 11 farklı uygulama gerçekleştirilmiştir. Bu uygulama sonuçlarının yorumlanması 5 bölüme ayrılmıştır. Birinci bölümde modlar arası transfer öğrenimi yönteminin geleneksel yaklaşımla karşılaştırılmaktadır. İkinci bölümde erken birleştirme yöntemi geleneksel yaklaşımla karşılaştırılmaktadır. Üçüncü bölümde özellik çıkartıcı seçiminin karşılaştırılması yapılmaktadır. Dördüncü bölümde veri seti seçiminin karşılaştırılması yapılmaktadır. Beşinci bölümde ise model-katman karşılaştırılması yapılmaktadır.

### 7.1 Modlar arası transfer öğrenimi yönteminin geleneksel yaklaşımla karşılaştırılması

Bu tez çalışmasında transfer öğrenimi için CNN ağ mimarisi ile görüntü verilerindeki duyguyu sınıflayan modelden, konuşma tipi ses sinyali verilerinden MFCC öznitelikleri çıkartılmış ve YSA yapısında Dense katmanları kullanılmış olan modele transfer öğrenimi gerçekleştirilmiştir. %59 sınıflandırma başarıyla konuşma tipi ses sinyallerinden MFCC öznitelikleri çıkartılmış ve YSA yapısında Dense katmanları kullanılmış olan modelin başarıları %6,78'lik yükselişle %63'e çıkmıştır.

### 7.2 Erken birleştirme yönteminin geleneksel yaklaşımla karşılaştırılması

Bu tez çalışmasında erken birleştirme yöntemi ile CNN ağ mimarisi ile görüntü verilerindeki duyguyu sınıflayan modeldeki cnn özniteliklerini, konuşma tipi ses sinyali verilerinden MFCC öznitelikleri çıkartılmış ve YSA yapısında Dense katmanları kullanılmış olan modele model daha eğitilmeden mfcc öznitelikleri çıkartıldıktan sonra öznitelik aktarımı yapılmıştır ve bu birleşik öznitelik matrisi daha sonra ses verilerinden duygu sınıflayan modele verilmiştir. %59 sınıflandırma başarıyla konuşma tipi ses sinyallerinden MFCC öznitelikleri çıkartılmış ve YSA yapısında Dense katmanları kullanılmış olan modelin başarıları %10,17'lik yükselişle %65'e çıkmıştır.

### **7.3 MFCC ve LMFB özellik çıkarıcı yöntemlerin karşılaştırılması**

MFCC özellik çıkartıcı yöntem ile özellik çıkarımı yapılan uygulamaların başarı ortalaması %67 iken LMFB öznelik çıkarım yöntemi özellik çıkarımı gerçekleştirilen uygulamaların ortalaması %59'da kalmıştır. Sonuç olarak ses sinyallerinden duygu tanımda MFCC'nin LMFB'a göre daha başarılı bir yöntem olduğu ortadadır.

### **7.4 Konuşma ve Şarkı tipi ses sinyallerinin karşılaştırılması**

Konuşma tipi ses sinyallerini içeren tüm uygulamaların başarı ortalaması %48 iken şarkı tipi ses sinyallerini içeren tüm uygulamalarda başarı ortalaması %78 olmuştur. Sonuç olarak gerçekleştirilen uygulamalarda veri setinin çok önemli olduğu şarkı tipi ses sinyallerinin konuşma tipi ses sinyallerine göre çok daha yüksek bir başarı ile sınıflandırıldığı görülmüştür.

### **7.5 Dense ve LSTM tipi katman türlerinin karşılaştırılması**

Dense YSA katmanı kullanılan tüm uygulamaların başarı ortalaması %69,25 iken LSTM tipi YSA katmanı kullanılan uygulamaların başarı ortalaması 56,75 olmuştur. Uygulama sonuçlarına göre ses sinyallerinden duygu tanımda Dense katmanlarının LSTM katmanlarına göre daha iyi sonuçlar verdiği gözükmemektedir.

### **7.6 Sınırlamalar, zorluklar ve öneriler**

Modlar arası transfer öğrenimi ile ses sinyallerinden duygu tanıma gerçekleştirirken karşılaşılan problemlerin en başında profesyonel bir şekilde kayıt alınmış video-ses tabanlı bir veri setine erişmektir. Veri setine eriştikten sonra bir başka önemli mesele veri setindeki veri sayısının oldukça fazla ve çeşitli olması gerektiridir. Ses sinyalinden duygu tanıma duygu çeşitlerinin fazla olabilmesi ve özellikle bazı duyguların birbirine çok benzeyebilmesi sebebiyle oldukça zordur. Bu çalışmada karşılaşılan en büyük zorluk ise transfer öğrenimi gerçekleştirirken farklı veri türlerine ait farklı modeller arası bilgi aktarımında anlamlı bilginin kaybolmamasını sağlamaktır. Bu problemlerin çözümü için literatürdeki mevcut veri setlerinin çeşitliliğinin ve boyutunun artması gerekmektedir. Ayrıca transfer öğrenmesi gerçekleştirirken benzer veri türleri ve benzer modeller arası çalışılmalıdır.

## 8. KAYNAKÇA

- [1] Korkmaz, O. E. (2016). Ses sinyalinde duyu tanıma.
- [2] Demircan, S. (2020). Ses sinyallerinden duyu tanıma için farklı yaklaşımlar.
- [3] Erdem, E. S. (2014). Ses sinyallerinde duyu tanıma ve geri erişimi (Master's thesis, Başkent Üniversitesi Fen Bilimleri Enstitüsü).
- [4] Elmas, B. (2021). Evrişimli Sinir Ağları ile Mantar Görüntülerinden Mantar Türlerinin Transfer Öğrenme Yöntemiyle Tanımlanması. Süleyman Demirel Üniversitesi Fen Bilimleri Enstitüsü Dergisi, 25(1), 74-88.
- [5] Aydın, B. (2016). Ters yüz sınıf modelinin akademik başarı, ödev/görev stres düzeyi ve öğrenme transferi üzerindeki etkisi (Master's thesis, Eğitim Bilimleri Enstitüsü).
- [6] Derrick M. (2023). Transfer Learning Guide: A Practical Tutorial With Examples for Images and Text in Keras. Neptune.ai. <https://i0.wp.com/neptune.ai/wp-content/uploads/2022/10/Transfer-learning-idea.jpg?resize=768%2C432&ssl=1>
- [7] Tawari A., Trivedi Mohan., Speech Emotion Analysis: Exploring the Role of Context, IEEE Transactions On Multimedia, 12(16):502-509, 2010.
- [8] Zeng Z., Pantic M., Roisman G., Huang T.S., A Survey of Affect Recognition Methods: Audio, Visual, and Spontaneous Expressions, IEEE Transactions On Pattern Analysis And Machine Intelligence, 31(1):39-58, 2009.
- [9] El Ayadi, M., S. Kamel, M. ve Karray, F., Survey on speech emotion recognition: Features, classification schemes, and databases, Pattern Recognition, 44, 3 (2011) 572-587.
- [10] France, D. J., Shiavi, R. G., Silverman, S., Silverman, M. ve Wilkes, D. M., 2000, Acoustical properties of speech as indicators of depression and suicidal risk, Ieee Transactions on Biomedical Engineering, 47 (7), 829-837.
- [11] FIRILDAK, K., & Talu, M. F. (2019). Evrişimsel sinir ağlarında kullanılan transfer öğrenme yaklaşımlarının incelenmesi. Computer Science, 4(2), 88-95.
- [12] Akılotu, B. N., Kadiroğlu, Z., Şengür, A., & Kayaoğlu, M. (2019). Evrişimsel Sinir Ağları ve Transfer Öğrenme Yöntemi Kullanılarak Sıtma Tespiti. In International Engineering and Science Symposium, Siirt.
- [13] Elmas, B. (2021). Evrişimli sinir ağları ile ağaç kabuğu görüntülerinden ağaç türlerinin transfer öğrenme yöntemiyle tanımlanması. Gazi Üniversitesi Mühendislik Mimarlık Fakültesi Dergisi, 36(3), 1253-1270.

- [14] FIRILDAK, K., TALU, M. F., & ÇELİK, G. (2020). Transfer Öğrenme ve Çekişmeli Üretici Ağ Yaklaşımlarını Kullanarak Cilt Lezyonu Sınıflandırma Doğruluğunu İyileştirilmesi. *Fırat Üniversitesi Mühendislik Bilimleri Dergisi*, 32(2), 491-498.
- [15] Siegel RL, Miller KD, and Jemal A. Cancer statistics, CA. *Cancer J. Clin* 2019; 69: 7–34.
- [16] Nasukawa, T. ve Yi, J., 2003, Sentiment analysis: capturing favorability using natural language processing. *Proceedings of the 2nd international conference on Knowledge capture*. Sanibel Island, FL, USA, ACM: 70-77.
- [17] Schuller, B., Rigoll, G. ve Lang, M., 2004, Speech emotion recognition combining acoustic features and linguistic information in a hybrid support vector machine - Belief Network architecture, 2004 *Ieee International Conference on Acoustics, Speech, and Signal Processing*, Vol I, *Proceedings*, 577-580.
- [18] Bozkurt, E., Erzin, E., Erdem, Ç. E. ve Erdem, A. T., 2011, Formant position based weighted spectral features for emotion recognition, *Speech Communication*, 53 (9-10), 1186-1197
- [19] Durukal, M., 2015, Performance Analysis of MFCC Features On Emotion Recognition from Speech, *International Journal of Scientific and Technological Research*, Vol 1 (No.5).
- [20] Karimi, S. ve Sedaaghi, M. H., 2016, How to categorize emotional speech signals with respect to the speaker's degree of emotional intensity, *Turkish Journal of Electrical Engineering & Computer Sciences*, 24, 1306-1324.
- [21] Milton, A. ve Tamil Selvi, S., 2014, Class-specific multiple classifiers scheme to recognize emotions from speech signals, *Computer Speech & Language*, 28 (3), 727-742.
- [22] Koolagudi, S. G., Murthy, Y. V. S. ve Bhaskar, S. P., 2018, Choice of a classifier, based on properties of a dataset: case study-speech emotion recognition, *International Journal of Speech Technology*, 21 (1), 167-183.
- [23] Ray, S. ve Chan, A., 2001, Automatic feature extraction from wavelet coefficients using genetic algorithms, *Neural Networks for Signal Processing Xi*, 233-241.
- [24] Go, H. J., Kwak, K. C., Lee, D. J. ve Chun, M. G., 2003, Emotion recognition from the facial image and speech signal, *Sice 2003 Annual Conference*, Vols 1-3, 2890- 2895.
- [25] Elmas, B. (2021). Evrişimli Sinir Ağları ile Mantar Görüntülerinden Mantar Türlerinin Transfer Öğrenme Yöntemiyle Tanımlanması. *Süleyman Demirel Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, 25(1), 74-88.

- [26] Reddy, A. P. ve Vijayarajan, V., 2020, Audio compression with multi-algorithm fusion and its impact in speech emotion recognition, *International Journal of Speech Technology*.
- [27] Muhammad G., Alotaibi Y., Alsulaiman, M., Environment Recognition Using Selected MPEG-7 Audio Features and Mel-Frequency Cepstral Coefficients, 2010 Fifth International Conference on Digital Telecommunications, 11-16, 2010.
- [28] Paleari, M., Chellali, R., Huet, B., Features for multimodal emotion recognition: An extensive study, *Cybernetics and Intelligent Systems (CIS)*, 2010 IEEE Conference, pp.90,95, 28-30, 2010.
- [29] Xie Z., Guan L., Multimodal Information Fusion of Audio Emotion Recognition Based on Kernel Entropy Component Analysis, *Multimedia (ISM)*, pp.1,8,2012.
- [30] Weiss, K., Khoshgoftaar, T. M., & Wang, D. (2016). A survey of transfer learning. *Journal of Big data*, 3(1), 1-40.
- [31] Cai, C., Wang, S., Xu, Y., Zhang, W., Tang, K., Ouyang, Q., ... & Pei, J. (2020). Transfer learning for drug discovery. *Journal of Medicinal Chemistry*, 63(16), 8683-8694.
- [32] Han, Z., Zhao, H., & Wang, R. (2019, May). Transfer learning for speech emotion recognition. In 2019 IEEE 5th Intl Conference on Big Data Security on Cloud (BigDataSecurity), IEEE Intl Conference on High Performance and Smart Computing,(HPSC) and IEEE Intl Conference on Intelligent Data and Security (IDS) (pp. 96-99). IEEE.
- [33] Luna-Jiménez, C., Griol, D., Callejas, Z., Kleinlein, R., Montero, J. M., & Fernández-Martínez, F. (2021). Multimodal emotion recognition on ravdess dataset using transfer learning. *Sensors*, 21(22), 7665.
- [34] Zheng, F., Zhang, G., & Song, Z. (2001). Comparison of different implementations of MFCC. *Journal of Computer science and Technology*, 16, 582-589.
- [35] Muda, L., Begam, M., & Elamvazuthi, I. (2010). Voice recognition algorithms using mel frequency cepstral coefficient (MFCC) and dynamic time warping (DTW) techniques. *arXiv preprint arXiv:1003.4083*.
- [36] Han, W., Chan, C. F., Choy, C. S., & Pun, K. P. (2006, May). An efficient MFCC extraction method in speech recognition. In 2006 IEEE International Symposium on Circuits and Systems (ISCAS) (pp. 4-pp). IEEE.
- [37] Kishore, K. K., & Satish, P. K. (2013, February). Emotion recognition in speech using MFCC and wavelet features. In 2013 3rd IEEE International Advance Computing Conference (IACC) (pp. 842-847). IEEE.

- [38] El Ayadi, M., Kamel, M. S. ve Karray, F., 2011, Survey on speech emotion recognition: Features, classification schemes, and databases, *Pattern Recognition*, 44 (3), 572-587.
- [39] Nwe, T. L., Foo, S. W., & De Silva, L. C. (2003). Speech emotion recognition using hidden Markov models. *Speech communication*, 41(4), 603-623.
- [40] Burkhardt, F., Paeschke, A., Rolfes, M., Sendlmeier, W. F., & Weiss, B. (2005, September). A database of German emotional speech. In *Interspeech* (Vol. 5, pp. 1517-1520).
- [41] Busso, C., Bulut, M., Lee, C. C., Kazemzadeh, A., Mower, E., Kim, S., ... & Narayanan, S. S. (2008). IEMOCAP: Interactive emotional dyadic motion capture database. *Language resources and evaluation*, 42, 335-359.
- [42] Jackson, P., & Haq, S. (2014). Surrey audio-visual expressed emotion (savee) database. University of Surrey: Guildford, UK.
- [43] Livingstone, S. R., & Russo, F. A. (2018). The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS): A dynamic, multimodal set of facial and vocal expressions in North American English. *PloS one*, 13(5), e0196391.
- [44] Cao, H., Cooper, D. G., Keutmann, M. K., Gur, R. C., Nenkova, A., & Verma, R. (2014). Crema-d: Crowd-sourced emotional multimodal actors dataset. *IEEE transactions on affective computing*, 5(4), 377-390.
- [45] Busso, C., Parthasarathy, S., Burmanian, A., AbdelWahab, M., Sadoughi, N., & Provost, E. M. (2016). MSP-IMPROV: An acted corpus of dyadic interactions to study emotion perception. *IEEE Transactions on Affective Computing*, 8(1), 67-80.
- [46] Martin, O., Kotsia, I., Macq, B., & Pitas, I. (2006, April). The eNTERFACE'05 audio-visual emotion database. In *22nd international conference on data engineering workshops (ICDEW'06)* (pp. 8-8). IEEE.
- [47] Zhalehpour, S., Onder, O., Akhtar, Z., & Erdem, C. E. (2016). BAUM-1: A spontaneous audio-visual face database of affective and mental states. *IEEE Transactions on Affective Computing*, 8(3), 300-313.
- [48] rezan ÇEÇEN, A. Duygular İnsan Yaşamında Neden Vazgeçilmez ve Önemlidir?. *Çukurova Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 9(9).
- [49] Leventhal, H., & Scherer, K. (1987). The relationship of emotion to cognition: A functional approach to a semantic controversy. *Cognition and emotion*, 1(1), 3-28.
- [50] Lazarus, R. S. (1984). On the primacy of cognition.
- [51] Frijda, N. H. (1988). The laws of emotion. *American psychologist*, 43(5), 349.

- [52] Izard, C. E. (1992). Basic emotions, relations among emotions, and emotion-cognition relations.
- [53] Ekman, P. E., & Davidson, R. J. (1994). The nature of emotion: Fundamental questions. Oxford University Press.
- [54] McLaren, C. (1998). Channeling your emotions.
- [55] Safran, J. D., & Greenberg, L. S. (1991). Emotion in human functioning: Theory and therapeutic implications. *Emotion, psychotherapy, and change*, 3-15.
- [56] Frijda, N. H. (1994). Emotions require cognitions, even if simple ones.
- [57] Darwin C., *The Expression of the Emotions in Man and Animals*, 1872.
- [58] Ekman P. Friesen, W.V., Constants across cultures in the face and emotion, *Journal of Personality and Social Psychology* 17: 124–129, 1971.
- [59] Ekman P., *Basic Emotions*, *Handbook of Cognition and Emotion*, Sussex, UK: John Wiley & Sons, 1999.
- [60] Plutchik R., *Emotion: Theory, research, and experience: Vol. 1. Theories of emotion*, New York: Academic, 1980.
- [61] Bozoklu Ç., Alkibay S., Nöropazarlama kapsamında tv reklamlarına yönelik duygulanım tepkilerinin değerlendirilmesi: yüz kasları analizi ile anket yönteminin karşılaştırılması, 19.Ulusal Pazarlama Kongresi Bildiri Kitabı, S 143-144, 2014.
- [62] Kyriakakis C., *Fundamental and Technological Limitations of Immersive Audio Systems*, *Proceedings Of The IEEE*, Vol. 86, No. 5, 1998.
- [63] Mower E., Mataric, M., Naryanan S., *A Framework for Automatic Human Emotion Classification Using Emotion Profiles*, *IEEE Transactions On Audio, Speech, And Language Processing*, 19(5):1057-1070, 2011.
- [64] Jia J., Zhang S., Meng F., Chai L. , *Emotional Audio-Visual Speech Synthesis Based on PAD* *IEEE Transactions On Audio, Speech, And Language Processing*, 19(3):570-582, 2011.
- [65] Soleymani M., Lichtenauer J., Pun,, T., Pantic, M., *A Multimodal Database for Affect Recognition and Implicit Tagging*, *IEEE Transactions On Affective Computing*, Vol. 3, No. 1:42-55, 2012.
- [66] Hassan, A., 2012, *On Automatic Emotion Classification Using Acoustic Features*, UNIVERSITY OF SOUTHAMPTON, 2012, 169.
- [67] Reynolds, D. A., *A Gaussian Mixture Modeling Approach to Text Independent Speaker Idendification*, Ph.D. thesis, Georgia Ins. of Technology, 1992.

- [68] Schafer, R. W. ve Rabiner, L. R., Digital Representations of Speech Signals, Proceedings of the Ieee, 63 (1975) 662-677.
- [69] Atal, B. S., Automatic Recognition of Speakers from Their Voices, Proceedings of the Ieee, 64 (1976) 460-475.
- [70] Eskidere, Ö. ve Ertaş, F., Mel Frekanslı Kepstrum Katsayılarındaki Değişimlerin Konuşmacı Tanımaya Etkisi, Uludağ Üniversitesi Mühendislik-Mimarlık Fakültesi Dergisi, 14 (2009) 92-110.
- [71] Benba, A., Jilbab, A., Hammouch, A., & Sandabad, S. (2015, March). Voiceprints analysis using MFCC and SVM for detecting patients with Parkinson's disease. In 2015 International conference on electrical and information technologies (ICEIT) (pp. 300-304). IEEE.
- [72] Ch. S. Kumar, P. R. Mallikarjuna, "Design of an automatic speaker recognition system using MFCC, Vector Quantization and LBG algorithm," International Journal on Computer Science and Engineering, Vol. 3, no. 8, 2011.
- [73] J. Martinez, H. Perez, E. Escamilla, M. M. Suzuki, "Speaker recognition using mel frequency cepstral coefficients (MFCC) and Vactor Quantization (VQ) techniques," IEEE Electrical Communications and Computers, Cholula, Puebla, Feb 2012, pp. 248-251.
- [74] Achraf BENBA, Abdelilah JILBAB and Ahmed HAMMOUCH. "Voice analysis for detecting persons with Parkinson's disease using MFCC and VQ." The 2014 International Conference on Circuits, Systems and Signal Processing, Saint Petersburg State Polytechnic University, Saint Petersburg, Russia, September 23-25, 2014.
- [75] Abdoli, S., Cardinal, P., & Koerich, A. L. (2019). End-to-end environmental sound classification using a 1D convolutional neural network. Expert Systems with Applications, 136, 252-263.
- [76] Rabiner, L. R. ve Juang, B. H., Fundamentals of speech recognition, Prentice Hall, Englewood Cliffs, (1993).
- [77] Heinzl, G., Rüdiger, A., & Schilling, R. (2002). Spectrum and spectral density estimation by the Discrete Fourier transform (DFT), including a comprehensive list of window functions and some new at-top windows.
- [78] Bergland, G. D. (1969). A guided tour of the fast Fourier transform. IEEE spectrum, 6(7), 41-52.
- [79] Sharma, S., Sharma, S., & Athaiya, A. (2017). Activation functions in neural networks. Towards Data Sci, 6(12), 310-316.

- [80] Alam, M. (2016). Codes in MATLAB for training artificial neural network using particle swarm optimization. Research Gate, 1-16.
- [81] Rende, F. Ş., Bütün, G., Karahan, Ş., & TÜBİTAK BİLGEM, G. (2016). Derin Öğrenme Algoritmalarında Model Testleri: Derin Testler.
- [82] Şeker, A., Diri, B., & Balık, H. H. (2017). Derin öğrenme yöntemleri ve uygulamaları hakkında bir inceleme. Gazi Mühendislik Bilimleri Dergisi, 3(3), 47-64.
- [83] Aslam, S., Herodotou, H., Ayub, N., & Mohsin, S. M. (2019, December). Deep learning based techniques to enhance the performance of microgrids: a review. In 2019 International Conference on Frontiers of Information Technology (FIT) (pp. 116-1165). IEEE.
- [84] Hacibeyoglu, M., & Ibrahim, M. H. (2018). Human gender prediction on facial mobil images using convolutional neural networks. International journal of intelligent systems and applications in engineering, 6(3), 203-208.
- [85] Özkan, İ. N. İ. K., & Ülker, E. (2017). Derin öğrenme ve görüntü analizinde kullanılan derin öğrenme modelleri. Gaziosmanpaşa Bilimsel Araştırma Dergisi, 6(3), 85-104.
- [86] Suriya, M., Chandran, V., & Sumithra, M. G. (2022). Enhanced deep convolutional neural network for malarial parasite classification. International Journal of Computers and Applications, 44(12), 1113-1122.
- [87] Le, X. H., Ho, H. V., Lee, G., & Jung, S. (2019). Application of long short-term memory (LSTM) neural network for flood forecasting. Water, 11(7), 1387.
- [88] Netes. (n.d.). Frekans Nedir ?. Netes. <https://www.netes.com.tr/netes/dosyalar/kcfinder/images/6004181-dmm-what-is-frequency-715x360-2%281%29.jpg>
- [89] Kaelbling, L. P., Littman, M. L., & Moore, A. W. (1996). Reinforcement learning: A survey. Journal of artificial intelligence research, 4, 237-285.
- [90] Fayek, H. (2016, April 21). Speech Processing for Machine Learning. [Haytham Fayek]. <https://haythamfayek.com/2016/04/21/speech-processing-for-machine-learning.html>
- [91] Camgoz, N. C., Kindiroglu, A. A., Akarun, L., & Aran, O. (2014, April). Domain adaptation for gesture recognition using hidden Markov models. In 2014 22nd Signal Processing and Communications Applications Conference (SIU) (pp. 2050-2053). IEEE.
- [92] Gadzicki, K., Khamsehashari, R., & Zetsche, C. (2020, July). Early vs late fusion in multimodal convolutional neural networks. In 2020 IEEE 23rd international conference on information fusion (FUSION) (pp. 1-6). IEEE.