

**T.C.**  
**FIRAT ÜNİVERSİTESİ**  
**FEN BİLİMLERİ ENSTİTÜSÜ**



**KONUŞMA OLMAYAN SİNYALLERİN EVRİŞİMSEL SİNİR  
AĞLARI TABANLI YAKLAŞIMLAR İLE SINIFLANDIRILMASI**

**Fatih DEMİR**

Doktora Tezi

ELEKTRİK ELEKTRONİK MÜHENDİSLİĞİ ANABİLİM DALI

Elektrik Elektronik Mühendisliği Teknolojileri Programı

NİSAN 2020

**T.C.**  
**FIRAT ÜNİVERSİTESİ**  
**FEN BİLİMLERİ ENSTİTÜSÜ**

Elektrik Elektronik Mühendisliği Anabilim Dalı  
Elektrik Elektronik Mühendisliği Teknolojileri Programı

Doktora Tezi

**KONUŞMA OLMAYAN SİNYALLERİN EVRİŞİMSEL SİNİR AĞLARI**  
**TABANLI YAKLAŞIMLAR İLE SINIFLANDIRILMASI**

Tez Yazarı  
**Fatih DEMİR**

Danışman  
Prof. Dr. Abdulkadir ŞENGÜR

NİSAN 2020  
ELAZIĞ

**T.C.**  
**FIRAT ÜNİVERSİTESİ**  
**FEN BİLİMLERİ ENSTİTÜSÜ**

Elektrik Elektronik Mühendisliği Anabilim Dalı  
Elektrik Elektronik Mühendisliği Teknolojileri Programı

**Doktora Tezi**

---

Başlığı: Konuşma Olmayan Sinyallerin Evrimsel Sinir Ağları Tabanlı Yaklaşımlar ile Sınıflandırılması

Yazarı: Fatih DEMİR

İlk Teslim Tarihi: 19.3.2020

Savunma Tarihi: 15.4.2020

---

**TEZ ONAYI**

Fırat Üniversitesi Fen Bilimleri Enstitüsü tez yazım kurallarına göre hazırlanan bu tez aşağıda imzaları bulunan jüri üyeleri tarafından değerlendirilmiş ve akademik dinleyicilere açık yapılan savunma sonucunda OYBİRLİĞİ ile kabul edilmiştir.

Danışman:	Prof. Dr. Abdulkadir ŞENGÜR Fırat Üniversitesi, Teknoloji Fakültesi	<i>İmza</i> Onayladım
İkinci Danışman:		Onayladım
Başkan:	Doç. Dr. Melih Cevdet İNCE Fırat Üniversitesi, Teknoloji Fakültesi	Onayladım
Üye:	Prof. Dr. Abdulkadir ŞENGÜR Fırat Üniversitesi, Teknoloji Fakültesi	Onayladım
Üye:	Doç. Dr. Murat KARABATAK Fırat Üniversitesi, Teknoloji Fakültesi	Onayladım
Üye:	Dr. Öğr. Ü. Ömer Faruk ALÇİN Malatya Turgut Özal Üniversitesi, Mühendislik Fakültesi	Onayladım
Üye:	Dr. Öğr. Ü. Muzaffer ASLAN Bingöl Üniversitesi, Mühendislik Fakültesi	Onayladım

Bu tez, Enstitü Yönetim Kurulunun ...../...../20..... tarihli toplantısında tescillenmiştir.

*İmza*

Prof. Dr. Soner ÖZGEN  
Enstitü Müdürü

## BEYAN

Fırat Üniversitesi Fen Bilimleri Enstitüsü tez yazım kurallarına uygun olarak hazırladığım “Konuşma Olmayan Sinyallerin Evrimsel Sinir Ağları Tabanlı Yaklaşımlar ile Sınıflandırılması ” Başlıklı Doktora Tezimin içindeki bütün bilgilerin doğru olduğunu, bilgilerin üretilmesi ve sunulmasında bilimsel etik kurallarına uygun davrandığımı, kullandığım bütün kaynakları atıf yaparak belirttiğimi, maddi ve manevi desteği olan tüm kurum/kuruluş ve kişileri belirttiğimi, burada sunduğum veri ve bilgileri unvan almak amacıyla daha önce hiçbir şekilde kullanmadığımı beyan ederim.

15/4/2020

**Fatih DEMİR**



# ÖNSÖZ

Yapay zeka uygulamaları zaman geçtikçe hayatımıza daha çok girmeye başlamıştır. Özellikle görsel verilerle ilgilenen uygulamalarda önemli derecede ilerleme sağlanmıştır. Bununla birlikte özellikle konuşma olmayan sesleri ve biyomedikal işaretleri içeren uygulamalarda gürültü ve arka plan seslerinden dolayı istenen başarıma her zaman ulaşamamaktadır. Bu ses türlerinde iyi bir performans sağlanırsa kalp seslerindeki anormalliklerin önceden teşhis edilmesi veya bir işitme engelliye etrafındaki acil duyması gereken seslerin bildirilmesi gibi hayatı kolaylaştırabilecek uygulamaların oluşturulması mümkün olacaktır.

Bu tez çalışmasında, konuşma olmayan seslerin ve biyomedikal işaretlerin sınıflandırma performansı, yeni bir makine öğrenme teknikleri ile artırılmaya çalışılmıştır. Yapılan deneysel çalışmalarda önerilen yöntemlerin sınıflandırma performanslarının sonuçları literatürde kullanılan diğer yöntemlerle karşılaştırılmıştır.

Çalışmalarımın her noktasında beni etkin bir şekilde yönlendiren, her türlü zorlukta yanımda olan, hem akademik hem de karakteriyle rol modeli olan danışman hocam Prof. Dr. Abdulkadir ŞENGÜR'e teşekkürlerimi bir borç bilirim.

Akademik kariyer boyunca hep yanımda olan ve beni sürekli çalışmalarım doğrultusunda teşvik eden Prof. Dr. Hasan GÜL ve Prof. Dr. Servet YILDIZ hocalarıma çok teşekkür ederim.

Yazılım konusunda farklı bakış açıları ile bakmamı sağlayan ve yardımını esirgemeyen Öğ. Gör. Dr. Yaman AKBULUT'a teşekkür ederim. Tez konusu ile ilgili ortak çalışma fırsatı bulduğum Doç. Dr. Varun Bajaj'a teşekkür ederim.

Bütün meslek seçimlerimde yanımda olan ve beni destekleyen babama, anneme, kardeşime, eşime ve hayatıma renk katan kızım Selin İpek DEMİR'e teşekkür ederim.

**Fatih DEMİR**  
ELAZIĞ, 2020

# İÇİNDEKİLER

	Sayfa
ÖNSÖZ.....	iv
İÇİNDEKİLER .....	v
ÖZET .....	vii
ABSTRACT .....	viii
ŞEKİLLER LİSTESİ .....	ix
TABLolar LİSTESİ .....	xi
SİMGELER .....	xiii
KISALTMALAR .....	xiv
<b>1. GİRİŞ .....</b>	<b>1</b>
1.1. Motivasyon .....	1
1.2. Çevresel Seslerin Otomatik Sınıflandırılması.....	3
1.3. Biyomedikal Ses ve İşaretlerin Sınıflandırılması .....	6
1.4. Tezin Literatüre Katkıları .....	8
1.5. Tezin Organizasyonu .....	9
<b>2. SPEKTROGRAM GÖRÜNTÜLER .....</b>	<b>10</b>
2.1. Giriş .....	10
2.2. Kısa Zamanlı Fourier Dönüşümü .....	10
2.3. Pencere ve Pencereleme İşlemi .....	13
2.3.1. Hamming Pencereleme İşlemi.....	13
2.3.2. Hann Pencerelemesi .....	13
2.4. Spektrogram .....	14
<b>3. ÖZİNİTELİK VE SINIFLANDIRMA AŞAMASINDA KULLANILAN YÖNTEMLER.....</b>	<b>16</b>
3.1. Evrişimsel Sinir Ağları .....	16
3.2. ESA Kullanım nedenleri.....	16
3.3. Evrişim İşlemi .....	20
3.4. Aktivasyon İşlemi.....	25
3.5. Havuzlama Katmanı .....	28
3.6. Toptan Normalleştirme Katmanı .....	30
3.7. Tam Bağlı Katman .....	31
3.8. Seyreltme Katmanı .....	32
3.9. Eşiksiz En Büyük İşlev Katmanı .....	33
3.10. ESA'da Sık Kullanılan Optimizasyon Teknikleri.....	34
3.10.1. Uyarlanabilir Öğrenme Oranları Kullanan Optimizasyon Teknikleri .....	36
3.11. Önceden Eğitilmiş ESA Modelleri .....	38
3.11.1. AlexNet .....	39
3.11.2. VGGNet .....	40
3.11.3. GoogleNet .....	41
3.11.4. ResNet .....	42
3.12. Sınıflandırma İşlemi .....	43
3.12.1. K-En Yakın Komşuluk.....	44
3.12.2. Destek Vektör Makineleri (DVM) .....	46
3.12.3. Karar Ağaçları .....	53

3.13. Sınıflandırmada Normalleştirme .....	55
3.13.1. Z-skor Normalleştirme Yöntemi .....	56
3.13.2. Min-Max Normalleştirme Yöntemi .....	56
3.14. Sınıflandırmada Eğitim ve Test Verileri .....	56
3.15. Sınıflandırmadan Sonraki Performans Ölçüm Kriterleri .....	58
<b>4. MATERYAL VE METOT .....</b>	<b>59</b>
4.1. Kalp Seslerinin Derin Öznitelikler ile Sınıflandırılması .....	59
4.1.1. Giriş .....	59
4.1.2. Önerilen Yöntem .....	59
4.2. Derin Öznitelikler ve Derin Öğrenme Tabanlı Yaklaşımların EMG ve Akciğer Sinyalleri .....	60
Üzerinden Denenmesi .....	60
4.2.1. Giriş .....	60
4.2.2. Önerilen Yöntem .....	60
4.3. Çevresel Seslerin Derin Öznitelikler ve Öznitelik Seçme Yaklaşımları ile Sınıflandırılması .....	62
4.3.1. Giriş .....	62
4.3.2. Önerilen Yöntem .....	63
4.4. Uçtan Uca Yapılandırılmış Bir ESA Modelinden Derin Öznitelikler Çıkarılması ile .....	63
Sınıflandırma .....	63
4.4.1. Giriş .....	63
4.5. Önerilen Yöntem .....	64
<b>5. BULGULAR .....</b>	<b>66</b>
5.1. Kalp Seslerine Derin Öznitelikler Tabanlı Yaklaşımın Uygulanması .....	66
5.1.1. Veri Seti .....	66
5.1.2. Deneysel Çalışmalar .....	67
5.2. EMG Sinyalleri ve Akciğer Seslerinden Oluşan Veri Setleri için Derin Öznitelikler ve Transfer .....	71
Öğrenme Yaklaşımlarının Uygulanması .....	71
5.2.1. EMG Sinyalleri için Uygulama .....	71
5.2.2. Akciğer Sesleri için Uygulama .....	74
5.3. Çevresel Seslere Derin Öznitelikler Tabanlı Yaklaşımın Önceden Eğitilmiş ESA modelleri ile .....	77
Uygulanması .....	77
5.3.1. Veri Setleri .....	77
5.3.2. Deneysel Çalışmalar .....	78
5.4. Çevresel Seslere Derin Öznitelikler Tabanlı Yaklaşımın Uçtan Uca Yapılandırılmış .....	81
ESA modelleri ile Uygulanması .....	81
<b>6. SONUÇLAR .....</b>	<b>89</b>
6.1. Sonuçların Değerlendirilmesi .....	90
ÖNERİLER .....	92
KAYNAKLAR .....	93
ÖZGEÇMİŞ .....	A

# ÖZET

---

## Konuşma Olmayan Sinyallerin Evrişimsel Sinir Ağları Tabanlı Yaklaşımlar ile Sınıflandırılması

**Fatih DEMİR**

Doktora Tezi

FIRAT ÜNİVERSİTESİ  
Fen Bilimleri Enstitüsü

Elektrik Elektronik Mühendisliği Anabilim Dalı  
Elektrik Elektronik Mühendisliği Teknolojileri Programı  
Nisan 2020, Sayfa: xxiv + 101

---

Makine öğrenmesi algoritmalarının gelişmesi ile birlikte her geçen sene değişik türdeki verilerin otomatik sınıflandırılması ve tanınması üzerine olan ilgi artmıştır. Bu ilgi alanları arasında yer alan konuşma olmayan sesler ve biyomedikal işaretlerin, arka plan sesi ve gürültüden dolayı sınıflandırılması zordur. Literatürde, bu tür sınıflandırma problemleri genellikle, akustik öznitelikler ve geleneksel sınıflandırma algoritmaları kullanan yöntemler ile aşılmaya çalışılmıştır.

Bu çalışmada, konuşma olmayan seslerin ve biyomedikal işaretlerin sınıflandırma problemi, son zamanlarda popüler hale gelen ve derin öğrenmenin alt alanı olan Evrişimsel Sinir Ağları (ESA) temelli yaklaşımlar ile çözülmeye çalışılmıştır. Bu kapsamda akciğer sesleri, kalp sesleri, EMG sinyalleri ve çevresel sesleri içeren kullanıma açık veri setleri üzerinde deneysel çalışmalar yapılmıştır.

Önerilen yöntemlerde, genel olarak derin öznitelikler tabanlı ve transfer öğrenme tabanlı yaklaşımlarla çalışmalar yapılmıştır. Derin öznitelikler tabanlı yaklaşımın, transfer öğrenme tabanlı yaklaşıma ve aynı veri setini kullanan diğer geleneksel yöntemlere göre daha iyi bir sınıflandırma performansı verdiği görülmüştür. Ayrıca tezin son bölümünde, derin öznitelikler çıkarmak için büyük görüntü verileriyle eğitilmiş ESA modelleri kullanmak yerine çevresel sesler ile eğitilmiş olan ve mimarisi kullanıcı tarafından ayarlanabilen bir ESA modeli önerilmiştir. Bu modelin çevresel sesleri sınıflandırmada, hem önceden eğitilmiş ESA modelleriyle derin öznitelik çıkarma yaklaşımına hem de aynı veri setini kullanan diğer yöntemlere göre daha iyi sınıflandırma performansı verdiği görülmüştür.

**Anahtar Kelimeler:** Konuşma olmayan sesler ve biyomedikal işaretler, Evrişimsel Sinir Ağları, Derin öznitelikler, Transfer öğrenme

# ABSTRACT

---

## Classification of Non-speech Signals with Convolutional Neural Network Based Approaches

**Fatih DEMİR**

Ph.D. Thesis

FIRAT UNIVERSITY

Graduate School of Natural and Applied Sciences

Department of Electrical and Electronics Engineering

Electrical and Electronics Engineering Technologies

April 2020, Pages: xxiv + 101

---

The interest on auto-classification and recognition of different types of data, has increased with development of machine-learning algorithms every passing year. The classification of non-speech sounds and biomedical signals taking part at these interest fields are difficult because of background sound and noise. In the literature, such classification problems are generally tried to be overcome by methods using acoustic features and traditional classification algorithms

In this study, the classification problem of non-speech sounds and biomedical signals are tackled with Convolutional Neural Networks (CNN) based approaches, which have become popular recently and are a subfield of deep learning. In this context, the experimental studies are realized on public datasets including lung sounds, heart sounds, EMG signals and environmental sounds.

In the proposed methods, studies are generally conducted with deep features based and transfer learning based approaches. It is shown that deep feature-based approaches give better the classification performance than the transfer learning-based approaches and the other conventional methods using same dataset. Besides, at the last part of the thesis, an CNN model whose architecture can be arranged by user and trained with environmental sounds, is proposed to extract deep features instead of using CNN models, which is pre-trained with big image data. It is concluded that this model presents better classification performance than deep feature extraction approach with pre-trained CNN models and other methods using same dataset for the classification of environmental sounds.

**Keywords:** Non-speech sounds and biomedical signals, Convolutional Neural Networks, Deep features, Transfer learning

## ŞEKİLLER LİSTESİ

	Sayfa
Şekil 2.1. Otobüs sesinin ayrık zamanlı KZFD dönüşümü.....	12
Şekil 2.2. Süpermarket sesinin ayrık zamanlı KZFD dönüşümü.....	12
Şekil 2.3. 64 tane örnek için Hamming pencerelemesi.....	13
Şekil 2.4. 64 tane örnek için Hann pencerelemesi.....	14
Şekil 2.5. Otobüs sesinin spektrogramı .....	15
Şekil 2.6. Süpermarket sesinin spektrogramı .....	15
Şekil 3.1. Klasik bir sinir ağında çıkış biriminin giriş birimleri ile etkileşim durumu .....	17
Şekil 3.2. Basit bir evrişim ağında çıkışı biriminin giriş birimleri ile etkileşim durumu .....	18
Şekil 3.3. Klasik bir sinir ağında giriş biriminin çıkış birimleri ile etkileşim durumu .....	18
Şekil 3.4. Basit bir sinir ağında giriş biriminin çıkış birimleri ile etkileşim durumu .....	18
Şekil 3.5. İki katmanlı bir evrişim ağında çıkışı biriminin giriş birimleri ile etkileşim durumu .....	19
Şekil 3.6. Basit bir sinir parametre paylaşım durumu .....	20
Şekil 3.7. Evrişimsel bir ağda parametre paylaşımı .....	20
Şekil 3.8. Ayrık zamanlı evrişim işleminin örnek gösterimi .....	22
Şekil 3.9. İki boyutlu matris durumu için ayrık zamanlı evrişim işleminin örnek gösterimi.....	24
Şekil 3.10. 224×224×3 boyutundaki orijinal görüntü ve farklı çekirdek matrisleri ile evrişim işlemleri .....	25
Şekil 3.11. Sigmoid Aktivasyon Fonksiyonu ve Türevinin Değişimi .....	26
Şekil 3.12. DDB katmanında kullanılan aktivasyon fonksiyonu.....	27
Şekil 3.13. Evrişim çıktılarına DDB katmanının uygulanması.....	28
Şekil 3.14. Maksimum havuzlama örnek gösterimi .....	29
Şekil 3.15. Maksimum havuzlama ile görüntülerin değişimi .....	30
Şekil 3.16. Tam Bağlı Katman temsili gösterimi .....	32
Şekil 3.17. Seyreltme katmanındaki seyreltme senaryoları .....	33
Şekil 3.18. EEBİ ve ÇD maliyet işlevi örnek gösterimi .....	34
Şekil 3.19. RGİ ve RGİB optimizasyon teknikleri için yitim değerinin değişimi .....	36
Şekil 3.20. Dışbükey olmayan bir fonksiyonun gösterimi.....	37
Şekil 3.21. LeNet ESA modeli .....	39
Şekil 3.22. AlexNet modeli .....	40
Şekil 3.23. VGG16 modeli .....	41
Şekil 3.24. VGG19 modeli .....	41
Şekil 3.25. Başlangıç katmanının içeriği .....	42

Şekil 3.26. GoogleNet Modeli.....	42
Şekil 3.27. ResNet modelinin artık katman yapısı .....	43
Şekil 3.28. K-EYK örnek gösterimi .....	45
Şekil 3.29. İki sınıf için DVM üstündüzlem gösterimi.....	47
Şekil 3.30. DVM’de doğrusal ayrılmama durumu .....	49
Şekil 3.31. Çok terimli çekirdek fonksiyonu ile ayırtaç ve destek vektörlerin durumu.....	52
Şekil 3.32. Gauss çekirdek fonksiyonu ile ayırtaç ve destek vektörlerin durumu .....	53
Şekil 3.33. İki sınıflı karar ağacı örnek gösterimi .....	54
Şekil 3.34. İki sınıflı düzensizlik işlevinin olasılık durumuna göre değişimi.....	55
Şekil 3.35. 10-Kat çapraz geçişleme örnek gösterimi .....	57
Şekil 3.36. Karmaşıklık matrisi .....	58
Şekil 4.1. Kalp Seslerinin Derin Öznitelikler ile Sınıflandırılması .....	60
Şekil 4.2. Derin Öznitelikler ile EMG sinyallerinin sınıflandırılması .....	61
Şekil 4.3. Transfer Öğrenme ile EMG sinyallerinin ve akciğer seslerinin sınıflandırılması .....	62
Şekil 4.4. Çevresel seslerin derin öznitelikler ile sınıflandırılması .....	63
Şekil 4.5. Önerilen yöntemin temsili gösterimi .....	64
Şekil 4.6. Uçtan uca eğitilen ESA modelinin hiyerarşik yapısı.....	65
Şekil 4.7. RAKT algoritmasının temsili gösterimi .....	65
Şekil 5.1. Veri Seti-A içi farklı sınıflardaki kalp sesleri ve spektrogramları.....	68
Şekil 5.2. AlexNet ESA modeli için transfer öğrenme.....	72
Şekil 5.3. En iyi performansı veren model ve yöntem için karmaşıklık matrisi.....	76
Şekil 5.4. ESC-10 veri seti için her sınıftan bir ses dosyasının spektrogram görüntüleri.....	78
Şekil 5.5. mFMU algoritması ile seçilen öznitelik sayısına göre sınıflandırma doğruluğu değişimi .....	79
Şekil 5.6. Kullanılan ESA modelinin katmanlarının boyutsal parametreleri.....	82
Şekil 5.7. UrbanSound8K veri seti için öznitelik setinin dağılımı .....	83
Şekil 5.8. UrbanSound8K veri seti için öznitelik setinin dağılımı .....	83
Şekil 5.9. UrbanSound8K veri seti için karmaşıklık matrisi .....	84
Şekil 5.10. DCASE 2017-ASC veri seti için karmaşıklık matrisi .....	85

## TABLolar LİSTESİ

	Sayfa
<b>Tablo 3.1.</b> Yaygın kullanılan ESA modelleri ile bilgiler .....	39
<b>Tablo 3.2.</b> DVM algoritmasında yaygın kullanılan çekirdek fonksiyonları.....	52
<b>Tablo 5.1.</b> Veri setleri ile ilgili bilgiler .....	67
<b>Tablo 5.2.</b> Veri seti-A için ESA modellerinin kombinasyonlarına göre kesinlik sonuçları .....	69
<b>Tablo 5.3.</b> Veri Seti-B için ESA modellerinin kombinasyonlarına göre kesinlik sonuçları.....	69
<b>Tablo 5.4.</b> Veri Seti-A ve Veri Seti-B kullanan diğer yöntemler .....	69
<b>Tablo 5.5.</b> Veri Seti-A için önerilen yöntemin diğer yöntemlerle karşılaştırılması .....	70
<b>Tablo 5.6.</b> Veri Seti-B için önerilen yöntemin diğer yöntemlerle karşılaştırılması.....	70
<b>Tablo 5.7.</b> Derin özniteliklerin kombinasyonları ile her elektrot için doğruluk sonuçları .....	72
<b>Tablo 5.8.</b> Transfer öğrenme ile elde edilen doğruluk sonuçları.....	73
<b>Tablo 5.9.</b> Transfer öğrenme ile elde edilen diğer performans ölçütleri .....	73
<b>Tablo 5.10.</b> Önerilen yöntemler ile diğer yöntemlerin karşılaştırılması.....	74
<b>Tablo 5.11.</b> ICBHI veri setinde 1 ses dosyası içindeki sınıfların durumu için örnek gösterim .....	74
<b>Tablo 5.12.</b> Sınıflara göre ayrılan ses dosyaların sayısal durumu .....	75
<b>Tablo 5.13.</b> ESA modelleri için derin öznitelikler ve transfer öğrenme ile sınıflandırma sonuçları .....	76
<b>Tablo 5.14.</b> En iyi performansı veren model ve yöntem için diğer performans kriterlerinin sonuçları .....	76
<b>Tablo 5.15.</b> Önerilen yöntemlerle ile diğer yöntemlerin karşılaştırılması.....	77
<b>Tablo 5.16.</b> Kullanılan veri setlerinin temel özellikleri.....	77
<b>Tablo 5.17.</b> AlexNet ile VGG16 modellerinden derin öznitelikler elde edilmesi, öznitelik birleştirme..... ve öznitelik seçme durumlarında sınıflandırma doğrulukları.....	79
<b>Tablo 5.18.</b> ESC-50 ve ESC-10 veri seti için önerilen yöntemin diğer yöntemler ile karşılaştırılması .....	80
<b>Tablo 5.19.</b> UrbanSound8K veri seti için önerilen yöntemin diğer yöntemler ile karşılaştırılması .....	80
<b>Tablo 5.20.</b> DCASE-2013 veri seti için önerilen yöntemin diğer yöntemler ile karşılaştırılması.....	81
<b>Tablo 5.21.</b> TBK boyutlarına göre sınıflandırma doğrulukları .....	82
<b>Tablo 5.22.</b> ESA modelleri ile öznitelik çıkarma süreleri.....	84
<b>Tablo 5.23.</b> UrbanSound8K veri seti için ESA modellerinden derin öznitelik çıkarımı kullanılarak..... elde edilen sınıflandırma sonuçları.....	85
<b>Tablo 5.24.</b> DCASE-2017 ASC veri seti için ESA modellerinden derin öznitelik çıkarımı kullanılarak..... elde edilen sınıflandırma sonuçları.....	86
<b>Tablo 5.25.</b> DCASE-2017 ASC veri setini kullanan diğer yöntemlere göre önerilen yöntemin her sınıf .....	86
..... için karşılaştırılması .....	86

<b>Tablo 5.26.</b> DCASE-2017 ASC veri setini kullanan diğ er yöntemlere göre önerilen yöntemin .....	
her sınıf için karşılaştırılması.....	87
<b>Tablo 5.27.</b> UrbanSound8K veri seti için önerilen yöntemin diğ er performans kriterlerine göre .....	
her sınıf için aldığı değ erler.....	88
<b>Tablo 5.28.</b> UrbanSound8K veri seti için önerilen yöntemin diğ er performans kriterlerine göre .....	
her sınıf için aldığı değ erler.....	88



## SİMGELER

$C$	: Düzgünleştirme parametresi
$D$	: Düzensizlik fonksiyonu
$K$	: Çekirdek
$L$	: Lagrange çarpanı
$M_i$	: Momentum değişkeni
$m_b$	: Mini-toptan ortalama
$o$	: Aktivasyon işlemi
$s$	: Evrişim işlemi
$S_i$	: Eşiksiz en büyük işlev çıkışı
$v_b$	: Mini toptan değişinti
$x$	: Giriş sinyali
$y$	: Çıkış
$\omega_i$	: Ağırlık değeri
$\omega(.)$	: Sürekli zaman pencereleme fonksiyonu
$\omega[.]$	: Ayrık zaman pencereleme fonksiyonu
$\delta_i$	: Sapma değeri
$\emptyset(.)$	: Çekirdek uzayı

## KISALTMALAR

AG	: Ayırtaç gücü
AFD	: Ayrık Fourier dönüşümü
ÇD	: Çapraz düzensizlik
ÇSS	: Çevresel sesleri sınıflandırma
MLP	: Çok katmanlı algılayıcı
DÖ	: Derin Öğrenme
DVM	: Destek Vektör Makinesi
DDB	: Düzleştirilmiş Doğrusal Birim
EEBİ	: Eşiksiz En Büyük İşlev
ET	: Eşleşme takibi
ESA	: Evrimsel Sinir Ağları
FKG	: Fonokardiyogram
FD	: Fourier dönüşümü
GKM	: Gauss karma modelleri
K-EYK	: K-En Yakın Komşu
KZEF	: Kısa zamanlı enerji fonksiyonu
MFKK	: Mel Frekans Kepstral Katsayıları
mFMU	: Minimum fazlalık maksimum uygunluk
RAKT	: Rastgele altuzay K-EYK topluluk
RGİ	: Rastgele Gradyan İnişi
RGİM	: Rastgele Gradyan İnişi Momentum
SMM	: Saklı Markov Modeli
SGS	: Sıfırdan geçiş sayısı
TBK	: Tam Bağlı Katman
TBA	: Temel bileşen analizi
YSA	: Yapay Sinir Ağları
Yİ	: Youden indeksi
ZFG	: Zaman-frekans gösterimi

# 1. GİRİŞ

İnsanlar, çevrelerinde gelişen olayları algılamak ve çevrelerine yanıt vermek için hem görme hem de işitme yeteneğinden etkin bir şekilde yararlanmaktadır. Bugünün robotik uygulamalarında ise bu durum oldukça sınırlıdır ve genellikle görsel temellidir. Robotik uygulamalarda herhangi bir çevresel ortamı algılamak için (örneğin, iç veya dış konumların belirlenmesi) sadece görsel bilgiler kullanırsa, görüntünün herhangi bir nedenle kısmen veya tamamen alınmadığı durumlarda sistemin sağlamlığı veya kullanımı kaybolacaktır [1-4]. Bu nedenle ses bilgisi ile sistem güvenilirliğini arttırmak için işitme sağlama yollarını bulma konusunda yeni ilgi alanları oluşmuştur [5, 6].

## 1.1. Motivasyon

İşitme ile ilgili sesler genel anlamda konuşmacı sesleri ve konuşma olmayan çevresel sesler olmak üzere iki alanda değerlendirilir. Konuşmacı seslerinin sınıflandırılmasında akustik öznitelikler ve klasik makine öğrenme teknikleri ile yüksek başarımlar sağlanmıştır [7-10]. Fakat çevresel seslerin sınıflandırılması sınıf çeşitliliğinin çokluğu, benzer özellikteki sınıfların varlığı, gürültü ve arka plan sesinin asıl sesi bastırmasından dolayı güçleşmektedir. Bu çevresel sesler canlı, canlı olmayan ve yapay varlıklardan üretilmektedir. Çevrede, bu sesler karışık biçimde bulunur ve bu seslerin her birini ayrı ayrı dinleyip ayırt etmek ve tanımlamak insanlar için hem zordur hem de zaman alıcıdır. Bu bakımdan, otomatik çoklu çevresel ses sınıflandırma sistemlerinin geliştirilmesinin önemi büyüktür [11]. Çevresel sesleri sınıflandırma (ÇSS) birçok uygulama alanı içermektedir. Giyilebilirler ve içeriğe duyarlı uygulamalar da ÇSS algoritmaları kullanılmaktadır [12, 13]. Tiyatro, sinema ve sınıf gibi sessizliğin gerekli olduğu bir ortamda kullanıcının ortamını temel alarak bildirim seçeneğini otomatik olarak değiştirebilen bir cep telefonu uygulaması veya işitme engelli ve ağır işitenler için kapı zili, araba kornası, yangın alarmı gibi duyması öncelikli sesleri bildiren akıllı cihaz uygulamaları ÇSS algoritmalarını kullanmaktadır [14, 15]. Ayrıca adli vakalarda suç tespiti ve akıllı ev güvenlik sistemlerinde ÇSS içeren algoritmalar uygulamalar kullanılmaktadır [16].

Tıbbi alanlarda gelişen teknolojiyle birlikte medikal cihazlardan alınan verilerin sayısallaştırılması ve saklanması mümkün hale gelmiştir. Bu sayısallaştırma ile birlikte verilerin işlenmesi ve sınıflandırılması üzerine çalışmalara literatürde çoğunlukla yer verilmiştir [17]. Bu çalışmaların arasında kalp ve akciğer konusunu ihtiva edenlerin önemi büyüktür. Çünkü, Dünya Sağlık Örgütü tarafından açıklanan verilere göre dünyada ölüm oranının en fazla olduğu hastalık kalp hastalığıdır ve ölüm oranındaki üçüncü sırada yer alan hastalık ise akciğer hastalığıdır [17, 18]. Bu nedenle, bu hastalıkların önceden analizinin yapılarak teşhis edilmesi büyük önem kazanmıştır.

Vücuttaki organların sesleri, tıbbi uzmanlar için bir veri kaynağıdır. Tecrübeli bir doktor, organların normal çalışma sesini analiz ederek sorunu bulabilir. Organların doğal sesleri haricindeki anormal bir ses, hastalığın teşhis edilmesine yardımcı olmaktadır [18]. Normal kalp atış ritmindeki ses sinyallerinin frekansı 60 ila 500 Hz arasında değişir. Bu frekanslar dışındaki kalp sesleri anormal olarak nitelendirilmektedir. Akciğerlerden gelen 1400 Hz veya üzerindeki sesler de, bir sorun olduğu anlamına gelmektedir [19].

Kalp ve akciğer hastalıklarında en çok kullanılan basit ve rutin yöntemlerden birisi stetoskoplara bu organların seslerinin dinlenmesidir. Stetoskobun basit veya karmaşık yapıya sahip çeşitleri olsa da genel olarak çalışma prensipleri ses iletimi üzerinedir. Yassı koni şeklindeki diyafram, vücutta dinlenmek istenen bölgeye konulur. Bu diyafram, ortam şartlarından yalıtılmış bir zarın oluşmaktadır. Dokundurulan yüzeyden gelen ses dalgaları zarı titreştirir. Zar, konik parça içindeki havaya baskı uygular ve oluşan hava basıncı elastik boru veya tüpten geçerek kulaklığa ulaşır. Kulaklıktaki uygun parçalar, sesi kulağa ulaştırır. Bazı stetoskoplarda “çan” adı verilen bir kısım daha bulunur. Çan, alçak sesleri yükseltmeye yarar [20]. Yeni geliştirilen dijital stetoskoplara çeşitli örnekleme frekanslarında (4KHz, 10KHz ve 44.1KHz) ses sinyallerinin kaydı yapılabilmektedir. Böylelikle, hastanın oskültasyonu ile ilgili bir veri tabanı oluşturabilmektedir. Elde edilen bu veriler makine öğrenme algoritmaları ile yüksek doğruluk oranında sınıflandırılabilirse uzman bir doktorun teşhisine yardımcı olmakla birlikte hastanın durumu hakkında daha net karar verilir [21, 22].

Sık sık yere düşme, merdiven çıkarken aksama, kalkarken ve otururken zorluk çekme, solunum ve yutkunmadaki sıkıntılar gibi kas rahatsızlıkları insan hayatını olumsuz yönde etkilemektedir. Ayrıca bu kas rahatsızlıkları, ALS ve Parkinson gibi sinir sistemini de ilgilendiren kalıcı hastalıklardan habercisi olabilir. Bu yüzden kas rahatsızlıklarının önceden teşhis etmek hastalığın ileri düzeyde seyretmemesi için çok önemlidir [23, 24]. Kas anormalliklerini teşhis etmek için sinir ve çizgili kasların, kasılma durumlarına göre elektriksel potansiyellerini gösteren Elektromiyografi (EMG) sinyalleri kullanılmaktadır ve EMG sinyallerinin güçlü ve ileri düzey yöntemlerle tespiti ve kaydedilmesi biyomedikal mühendisliğinde çok önemli bir gereklilik haline gelmiştir [25-28]. Kaydedilen birçok EMG sinyalinin işledikten sonra doğru bir şekilde otomatik olarak sınıflandırılabilen uygulamalar sayesinde uzman doktorların iş yoğunluğu önemli miktarda azaltacaktır [29].

Yukarıdaki paragraflarda anlatılan nedenler doğrultusunda bu tez çalışmasında, konuşma olmayan seslerin ve biyomedikal işaretlerin, Evrişimsel Sinir Ağları (ESA) tabanlı mantık çerçevesinde sınıflandırma problemi ele alınmış ve bu problem çözülmeye çalışılmıştır.

## 1.2. Çevresel Seslerin Otomatik Sınıflandırılması

Ses sınıflandırmasıyla ilgili ilk çalışmalar seslerin spektral özniteliklerini çıkardıktan sonra bu öznitelikleri geleneksel sınıflandırıcılara besleme üzerine dayandırılmıştır. Szu ve Tefler [30] konuşmacı ve çevresel seslerin sınıflandırması üzerine çalışmışlardır. Çalışmalarında sabit dalgacık parametreleri kullanmak yerine ortogonal uyarlanabilir dalgacık parametrelerini kullanmışlardır ve bu çalışmaları Yapay Sinir Ağları (YSA) sınıflandırıcısı üzerinde daha iyi sonuç vermiştir. Frstrup ve Watkins [31] deniz hayvanlarının seslerinin sınıflandırılması için çalışmalarda bulunmuşlardır. *Hızlı Fourier* dönüşümüne dayanan 27 tane spektral öznitelik çıkarmışlardır. Bu öznitelikler Morrison [32] tarafından önerilen doğrusal bir sınıflandırıcıya ve, Clark ve Pregibon [33] tarafından önerilen ağaç tabanlı bir sınıflandırıcıya uygulanmıştır. Nooralahiyen vd. [34] trafik yönetim sistemlerine yardım etmek için online olarak trafikteki dört tip aracı seslerine göre sınıflandırmak için çalışmışlardır. Dijital ses kaydedici bir mikrofon ile sesleri kaydettikten sonra otomatik düzeltme analizine dayanan Doğrusal Öngörü Kodlama parametre dönüşümü ile ön işleme yapmışlardır. Zaman Gecikmeli Sinir Ağları ile eğitim ve test süreci gerçekleştirilmiş ve %100 doğruluk ile sınıflandırma yapılmıştır. Zhang ve Kuo [35] konuşma, müzik gibi çeşitli çevresel sesleri içeren ses kayıtlarının gerçek zamanlı sınıflandırılması üzerine çalışmışlardır. Ses özniteliklerini çıkarmak için sıfırdan geçiş sayısı (SGS) , kısa zamanlı enerji fonksiyonu (KZEF) ve sesin hem harmonik hem de harmonik olmayan bileşenlerini kullanılmıştır. Sezgisel kural tabanlı bir sınıflandırıcı ise gerçek zamanlı sınıflandırma için kullanılmıştır. Liu ve Wan [36] çevresel sesleri otomatik olarak sınıflandırmak için içerik tabanlı teknoloji geliştirmiştir. Ardışık ileri yönde seçim yöntemi ile 87 adet öznitelik çıkarılmıştır. K-En Yakın Komşu (K-EYK), Gauss karışım modeli (GKM) ve olasılıksal sinir ağı sınıflandırıcılarının performansı karşılaştırılmıştır. Deneyler kullanıma açık ve yazarlar tarafından oluşturulan daha kapsamlı bir veri tabanında gerçekleştirilmiştir. Öklid uzaklık ölçümü ile birleştirilmiş olasılıksal sinir ağı sınıflandırıcısı ile en iyi sonuç alınmıştır. Lu ve Zhang [37] konuşmacı sesleri ile çevresel sesleri otomatik olarak ayırdıktan sonra çevresel seslerin kendi içerisinde sınıflandırması için çalışmışlardır. Öznitelik çıkarım aşamasında SGS, KZEF ve spektrum akısı kullanılmıştır. Konuşmacı sesleri ile çevresel seslerin ayrılması için K-EYK sınıflandırıcısı kullanılmıştır. Çevresel seslerin sınıflandırılması için ise, eşik yöntemine dayanan bir kural tabanı geliştirilmiştir. Toyoda vd. [38] çevresel sesleri sınıflandırmak için halka açık bir veri tabanı üzerinde çalışmışlardır. Ses sinyallerinin frekans ve güç bileşenlerini çok katmanlı bir algılayıcıda kullanmayı tercih etmişlerdir. Wang vd. [39] çevresel seslerin özniteliklerini oluşturmak için MPEG-7 düşük seviyeli tanımlayıcı, spektral ağırlık merkezi, spektral yayılım ve spektral yassılık kombinasyonunu kullanmıştır. Sınıflandırma safhasında en iyi sonucu almak için, Destek Vektör Makinesi (DVM), ve K-EYK sınıflandırıcılarının tahmin skorlarının belirli ağırlık oranları ile birleştirilmesi tercih edilmiştir. Chu vd. [40] çevresel seslerden öznitelik çıkarmak için Mel Frekans Keprtral Katsayıları (MFKK)

ve eşleşme takibi (ET) yöntemini kullanmışlardır. Sınıflandırma aşamasında K-EYK ve GKM algoritmaları kullanılmıştır. En iyi sınıflandırma sonucu MFKK ile ET'nin birlikte kullanıldığı öznelikler ve GKM sınıflandırıcısı ile elde edilmiştir. Su vd. [41] çevresel seslerden öznelik çıkarmak için düzensizliği en aza indirgeyen yerel ayırtaç düzlem algoritmasını kullanmıştır. Önerilen metodun sınıflandırma performansı Saklı Markov Modeli (SKM) ile değerlendirilmiştir. Uz Kent vd. [42] konuşma olmayan seslerin öznelik çıkarımı için genellikle konuşma ve müzik seslerinde iyi sonuç veren MFKK ile birlikte otomatik düzeltme fonksiyonu kullanmayı tercih etmiştir. Eğitim ve test işlemi için K-EYK, radyal temelli çekirdek kullanan YSA, doğrusal ve Gauss çekirdeğini kullanan DVM kullanılmıştır. En iyi sonucu sınıflandırma doğruluğunu, Gauss çekirdeğini kullanan DVM vermiştir. Wang vd. [43] çevresel seslerin özneliklerini oluşturmak için Gabor fonksiyonunda düzensiz frekans aralıklarını kullanarak ölçek-frekans haritaları çıkarmıştır. Daha sonra ET yöntemi ile ölçek-frekans haritaları daha az boyutlu olacak şekilde ayrıştırılmıştır. Ayrıştırılan her haritanın boyutu doğrusal ayırtaç analizi ve temel bileşen analizi (TBA) ile azaltılmıştır. Sınıflandırma için DVM algoritması kullanılmıştır. Ayrıca düzenli frekans aralıklarına göre düzensiz frekans aralıklarının sınıflandırma doğruluğunda daha iyi sonuç verdiği gözlemlenmiştir. Rakotomamonjy ve Gasso [44] çevresel ses sinyallerinin sabit-Q dönüşüm tabanlı zaman-frekans gösterimini içeren resimler oluşturmuştur. Bu resimlerinden histogramları kullanılarak öznelikler çıkarılmıştır. Sınıflandırma aşamasında doğrusal bir DVM algoritması kullanılmıştır.

Krizhevsky ve Sutskever [45] ImageNet yarışması için kullanmış oldukları kendi ESA mimarisi ile en iyi sınıflandırma sonucunu elde etmiştir. Simonyan ve Zisserman [46] 2 yıl sonra düzenlenen ImageNet yarışmasında da kendi ESA mimarileri ile birinci olmuştur. Bu iyi sonuçlardan sonra ESA tabanlı modeller makine öğrenme alanında sıkça kullanılmaya başlanmıştır.

ÇSS problemi için, diğer ESA modellerine öncülük edecek ilk ESA çalışmaları Pizcak [47] tarafından sağlanmıştır. Pizcak 2 tane evrişim katmanı, 2 tane maksimum havuzlama katmanı ve 2 tane tam bağlı katmandan oluşan bir ESA yapısı ile çalışmıştır. Sınıflandırma doğruluğu, 10-kat çapraz doğrulama ile %74 olmuştur. Salamon vd. [48], Pizcak'taki ESA modeline göre hem daha fazla evrişim katman ekleyerek hem de veri artırma yöntemlerini kullanarak sınıflandırma doğruluğunu %5 kadar geliştirmiştir. Chen vd. [49] klasik bir ESA modelinde maksimum havuzlama katmanları kullanmak yerine genişletilmiş evrişim katmanlarını farklı boyutlarda kullanmayı tercih etmiştir. Kullanılan yöntemde, ayrı ayrı 1, 2 ve 3 tane genişletilmiş evrişim katmanlarıyla test yapıldıktan sonra en iyi sınıflandırma sonucuna 3 tane genişletilmiş evrişim katmanı ile ulaşılmıştır.

DCASE 2017 ASC yarışması ile 15 sınıflı ÇSS probleminin çözümü için 76 farklı çalışma yapılmıştır. Bu çalışmaların hepsinde ESA tabanlı modeller kullanılmıştır. Bu paragrafta yer verilen çalışmalar DCASE 2017 ASC yarışmasında en iyi değerlendirme sonucunu almış ilk 6

çalışmadır. Park vd. [50], her alt bant için güç yoğunluğunu analiz eden bir eşdeğişinti matrisini ve her alt bant için güç değişimini analiz eden bir çift Fourier dönüşüm görüntüsünü öznitelik çıkarımı için kullanmıştır. 3 tane evrişim katmanı, 3 tane maksimum havuzlama katmanı ve 3 tane tam bağlı katman içeren bir Evrişimsel Sinir Ağı (ESA) modeliyle eğitim ve test işlemi gerçekleştirilmiştir. Eğitim süresini azaltmak için her evrişim katmanının çıkışında toptan normalleştirme işlemi uygulanmıştır. Test verileri için en iyi sınıflandırma doğruluğu bu yöntemle %72.6 olmuştur. Lehner vd. [51] topluluklar ile sınıflandırma modelini, i-vector, tam ESA ve ESA alıntı olarak adlandırılan üç farklı model için kullanmayı seçti. I-vektör modelinde, öznitelik çıkarmak için MFKK ve sınıflandırma yapmak için DAA algoritması kullanılmıştır. Tam ESA modelinde, öznitelikler için bir spektrogram işlemi ve evrişim, maksimum havuzlama ve global ortalama havuzlama katmanlarından oluşan 5 katmana sahip bir ESA yapısı oluşturulmuştur. ESA alıntı modelinin, rastgele seçilen 2 saniyelik spektrogram bölümlerinin kullanılması haricinde tam ESA modeli ile aynı olduğu belirtilmiştir. Daha sonra, topluluk algoritması ile sınıflandırma yapılmış ve en iyi sonuca % 73.8 ile ulaşılmıştır. Hyder vd. [52] Mel ve logaritmik ölçekli filtreleri kullanılarak spektrogram görüntüleri oluşturmuştur. Daha sonra, ESA model yapısını temel alınarak farklı filtre sayı ve boyutlardan oluşan alt modeller oluşturulmuştur. Alt modeller, spektrogram görüntüleri kullanılarak eğitilip test edilmiştir. Sınıflandırma performansını arttırmak için, her alt modelden elde edilen en iyi sonuç kullanılmıştır. 10 tane alt model arasında en iyi sınıflandırma doğruluğu %74,1 olarak hesaplanmıştır. Weiping vd. [53], ön işleme aşamasında standart ve Sabit-Q dönüşüm spektrogramları kullanılmıştır. Özniteliklerin çıkarılması için, bu temsiller ayrı ayrı 5 katmanlı derin bir ESA modelini beslemiştir. Eğitilmiş bu ESA modellerinin tam bağlı katmanlarından alınan özniteliklerin birleştirilmesinden sonra en iyi sınıflandırma doğruluğu DVM sınıflandırıcısı ile % 77.7 olarak kaydedilmiştir. Han vd. [54], stereofonik gösterimler, harmonik-vurmali kaynak ayırımı ve arka plan çıkarma işlemlerinden oluşan üç farklı temsil oluşturmuştur. Bu temsiller ayrı ayrı aynı yapıdaki ESA modelleriyle eğitilmiştir. Daha sonra her ESA modelinin çıkışından alınan skorlar, topluluk modelinde sınıflandırma yapmak için kullanılmıştır. En yüksek sınıflandırma doğruluğu, karar verme aşamasında ortalama olasılık yöntemini kullanan topluluk modeliyle %80,4 olarak gerçekleşmiştir. Mun vd. [55] ek eğitim veri seti üretmek için çekişmeli üretici ağlar tabanlı bir yöntem önermiştir. Öznitelikler, eğitilmiş bir ESA modelinin tam bağlı katmanlarından elde edilmiştir. DVM, çok katmanlı algılayıcı (MLP), tekrarlayan sinir ağı ve ESA sınıflandırıcılarından oluşturulan bir topluluk modeli sınıflandırma için tercih edilmiştir. Değerlendirme verilerindeki genel ortalama doğruluğu bu topluluk modeliyle %83,3 olarak hesaplanmıştır.

### 1.3. Biyomedikal Ses ve İşaretlerin Sınıflandırılması

Birçok alanda uygulanan yapay zeka algoritmaları 1990'lı yılların başından itibaren sağlık alanında da yoğun olarak kullanılmaya başlanmıştır. İlk başta geleneksel yöntemlerle biyomedikal ses ve işaretler üzerinde çalışılmıştır. 2012 yılından sonra özellikle ESA modeli içeren algoritmalar kullanılmaya başlanmıştır.

Barschdorf vd. [56] stetoskop ile kaydedilen fonokardiyogram (FKG) sinyallerini kullanarak kalp sesindeki anormallikler üzerinde çalışmıştır. FKG sinyallerinin histogramı ve 35Hz, 70Hz, 140Hz ile 250Hz frekans bandındaki spektral yoğunlukları öznitelik çıkarmak için kullanılmıştır. Bu özniteliklere normalleştirme işlemi uygulandıktan sonra geri yayılım algoritması kullanan YSA ile sınıflandırma işlemi gerçekleştirilmiştir. Zhang vd. [57] FKG sinyallerinden öznitelik çıkarmak için kısmi en küçük kareler regresyon tabanlı bir yöntem önermiştir. Daha sonra DVM algoritması ile sınıflandırma yapılmıştır. M.Hamidi vd. [58] iki tane öznitelik çıkartım yöntemi kullanmıştır. İlk yöntemde kalp ses sinyali diziliminden elde edilen bilgiler için eğri uydurma yöntemi kullanılmıştır. İkinci yöntemde, MFKK kullanılmıştır ve her iki öznitelik seti birleştirilmiştir. Sınıflandırma tarafında ise K-EYK algoritması kullanılmıştır. S. Randhawa vd. [59] kalp seslerinin frekans dönüşümünü kullanarak sıfır-geçiş oranı ve efektif değerlerden çıkartılan öznitelikleri Fisher's Diskriminat Oranı yöntemiyle azaltmıştır. Sınıflandırma kısmında ise hem K-EYK hem de YSA kullanılmıştır. J. Low ve K.Choo [60] ön işleme sürecinde, yüksek geçiren filtre ile süzölmüş kalp seslerini ayırık dalgacık dönüşümü kullanarak zaman-frekans bölgesine göre yapılandırdıktan sonra tek boyutlu verileri 48×48'lik yoğunluk haritasına çevirmiştir. Bu yoğunluk haritaları ESA ile eğitilip sınıflandırılmıştır. J. Rubin vd. [61] kalp seslerinden öznitelik çıkartımı aşamasında MFKK'ların spektrogramları alınarak iki boyutlu veri haritaları oluşturulmuştur. Sınıflandırma işleminde ESA algoritmasından faydalanılmıştır. Juan P. Dominguez-Morales vd. [62] işitsel bir algılayıcı ile ses verilerini bilgisayar ortamına aldıktan sonra verileri sonogram görüntülere dönüştürmeyi seçmiştir. Bu görüntüleri de farklı ESA mimarileriyle sınıflandırmıştır. T. Chen [63] MFKK ile öznitelik çıkartımı yaptıktan sonra verileri ESA ile sınıflandırmıştır. Ayrıca, lojistik ayırım, Gauss karma modelleri (GKM), K-EYK ve DVM ile karşılaştırdığında en iyi sınıflandırma sonucunu ESA modelinin verdiği vurgulanmıştır. C. Potes vd. [64] MFKK öznitelik çıkarımından sonra AdaBoost isimli bir model ve ESA ile eğitim yaptıktan sonra en son hangi sınıflandırıcı çıkışının kullanılacağını belirten karar verici bir algoritma kullanmıştır. M. Tschannen vd. [65] ESA modelleri, durum istatistikleri ve güç spektral yoğunluklarını kullanarak karma bir özellik çıkartım vektörü oluşturmuştur. Bu çıkarım vektörü ise DVM ile eğitilip sınıflandırılmıştır.

Serbes vd. [66] , akciğer ses sınıflandırması için çatlak, nefes darlığı ve normal sınıf etiketlerden oluşan ve toplam 6.000 ses dosyasından oluşan bir veri seti kullanılmıştır. Solunum anormalliklerinin tanınması için zaman-frekansı ve zaman-ölçeği analizini kullanan iki tane

öznitelik çıkarım yöntemi tercih edilmiştir. Sınıflandırma aşamasında K-EYK, DVM ve MLP kullanılmıştır ve DVM sınıflandırıcısı ile en iyi doğruluk %97,5 olarak elde edilmiştir. Feng vd. [66] akciğer ses sınıflandırması için 4 sınıf etiketli anormal akciğer seslerini içeren veri seti üzerinde çalışılmıştır. Öznitelik çıkarımı için ayırım ve anlık basıklık fonksiyonları kullanılmıştır. Radyal temelli çekirdek kullanılan DVM sınıflandırıcısı ile %97,7-98,8 arasında değişen doğruluk performansları elde edilmiştir. Mohammed [68] normal ve nefes darlığı içeren ses dosyalarından öznitelik çıkarımı için MFKK kullanmıştır. Daha sonra, önerilen yöntemin GKM ile eğitilip test edilmiştir ve en iyi sınıflandırma doğruluğu %94,2 ile sonuçlanmıştır. Reyes vd. [69] polifonik hırıltılardan oluşan solunum seslerinden öznitelik çıkarmak için yüksek dereceli istatistikler kullanmıştır ve özniteliklerin boyutunu azaltmak için genetik algoritmaları ve fisher ayırtaç oranını kullanmayı tercih etmiştir. Doğruluk performansı K-EYK ve saf bayes algoritmaları kullanan sınıflandırıcılarla %94,6 olmuştur. Orjuela-Cañón vd. [70] anormal akciğer seslerini tespit etmek için basıklık, eşdeğişinti, eğrilik ve Lacunarity fonksiyonları ile öznitelik vektörünü oluşturmuştur. Aşırı öğrenme makinesi ve DVM ile sınıflandırma performans testi yapılmıştır ve AÖM ile daha iyi sonuç alınmıştır. Maruf vd. [71] nefes darlığı ve normal sesleri içeren sınıf etiketlerinden oluşan veri setinin ön işleme aşamasında bant geçiren bir filtreyi gürültüleri ortadan kaldırmak için kullanmıştır. Yükseklik değişimi, enerji ve spektrogram fonksiyonlarından oluşan öznitelikler arasından seçim yapan algoritmayı test etmek için GKM sınıflandırıcısı kullanılmıştır ve sınıflandırma sonucu % 97,56 olmuştur. Bordua vd. [72] spektrogram, MFKK ve yerel ikili desen (YİD) kullanarak akciğer seslerinin zaman-frekans temsilini oluşturmuştur. Bu temsiller farklı yinleme sayılarına sahip 4 tane ESA modelini eğitmek için giriş olarak kullanılmıştır. 4 tane ESA modelinin tahmin sonuçları birleştirilerek topluluk sınıflandırıcısı oluşturulmuştur. Bu yöntemle elde edilen en iyi sonuç %95,56 olmuştur. Chambres vd. [73] mel-bandı, enerji bandı ve MFKK gibi spektral öznitelikleri kullanmıştır. Sınıflandırma aşamasında karar ağacı algoritması ve MFKK öznitelik çıkarımı ile en iyi doğruluk performansı % 49,62 olarak elde edilmiştir. Jakovljević vd. [74] örnekleme frekansları farklı ve gürültü oranı yüksek zorlu bir veri seti üzerinde çalışmıştır. Öznitelik çıkarma aşamasında MFKK kullanılmıştır. Sınıflandırma aşamasında, GMM ve HMM sınıflandırıcılarını birlikte kullanan bir yöntem geliştirilmiştir ve en iyi sınıflandırma doğruluğu %39,56 olarak bulunmuştur. Serbes vd. [75] zorlu bir veri seti üzerinde çalışmıştır. Öznitelikleri çıkarmak için KZFD ile birleştirilmiş dalgacık dönüşümü kullanılmıştır. DVM sınıflandırıcısı ile sınıflandırma performansı test edilirken işlem yükünü azaltmak TBA kullanılmıştır. KZFD+dalgacık dönüşümü ve % 90 TBA varyasyonları kullanılarak en iyi doğruluk sonucu %57,88 olarak kaydedilmiştir.

Bekey vd. [76] EMG sinyallerinin sınıflandırılması için öncülük etmiştir. 6 tane farklı kas grubundan alınan sinyaller ikili kodlama ile işlenerek öznitelik vektörü oluşturulmuş ve doğrusal ayırtaç algoritması ile sınıflandırma yapılmıştır. Graupe vd. [77] özbağlanımlı bir model ile EMG

sinyallerden katsayılar üretmiştir. Bu katsayılar kullanılarak doğrusal olmayan ayırtaç algoritması ile EMG sinyalleri sınıflandırılmıştır. Lee ve Saridis [78] EMG sinyallerinden öznitelik çıkarmak için SGS ve varyans algoritmalarından yararlanmıştır. Sınıflandırma aşamasında ise bayes sınıflandırıcı tercih edilmiştir. Uchida [79] EMG sinyallerinin karakteristik özniteliklerini belirginleştirmek için hızlı Fourier dönüşümlerini almıştır. Bu dönüşümlerden elde ettiği değerleri kullanarak bulanık mantık algoritmasında sınıflandırma işlemi gerçekleştirilmiştir. Hudgins vd. [80] ortalama mutlak değer, ortalama mutlak değer eğimi, SGS, eğim işaret değişiklikleri ve dalga şekli uzunlukları ile çok fonksiyonlu bir algoritmayı öznitelik çıkarımı için tercih etmiştir. Sınıflandırma işleminde YSA algoritması kullanılmıştır. Chan vd. [81] EMG sinyallerini sınıflandırmak için bulanık mantık algoritmasını yeni bir kural tabanı ile uygulamıştır. Lucas vd. [82] EMG sinyallerinin ayrık dalgacık dönüşümünü alarak öznitelik vektörünü oluşturmuştur. Güçlü bir sınıflandırıcı olan DVM ile sınıflandırma işlemi gerçekleştirilmiştir. Phinyomark vd. [83] frekans ve zaman bölgesi 37 tane öznitelik çıkarma fonksiyonundan hangilerinin EMG sinyallerinin sınıflandırılması için daha uygun olduğunu araştırmıştır. Yapılan deneysel çalışmalarda zaman bölgesi fonksiyonlar, frekans bölgesi fonksiyonlara göre daha iyi sınıflandırma doğruluğu vermiştir. Subasi [84] EMG sinyallerinden daha iyi sınıflandırma performansı almak için DVM sınıflandırıcısını eğitirken çekirdek parametrelerini parçacık sürü optimizasyonu ile belirlemiştir. Atzori vd.[85] EMG sinyallerini sınıflandırmak için ESA algoritmasını kullanmıştır. ESA modelinin performansını ölçmek için DVM, K-EYK ve rastgele orman algoritmaları ile karşılaştırma yapılmıştır. Rastgele orman algoritmasından sonra en iyi sonucu ESA vermiştir. Bu durumun eğitim verilerinin yetersizliğinden kaynaklandığı belirtilmiş ve büyük verilerle daha iyi sonuçlar alınacağı vurgulanmıştır. Xia vd. [86] EMG sinyallerinin sınıflandırma performansını artırmak için özyinelemeli sinir ağı ile ESA'yı birleştirmiştir. Phinyomark ve Scheme [87] ESA'dan optimum performans almak için büyük EMG verilerine sahip bir veri seti kullanmıştır. Geleneksel sınıflandırıcılara göre daha iyi ortalama sınıflandırma doğruluğu elde edilmiştir.

#### 1.4. Tezin Literatüre Katkıları

Bu tez çalışması süresince bir tane yurtiçi uluslararası konferans bildirisi sunulmuş ve bir tane yurtdışı uluslararası konferans bildirisi sunulması için gönderilmiş; iki tane uluslararası *Emerging Sources Citation Index* indeksli makale ve bir tane uluslararası *Science Citation Index Expanded* indeksli makale yayınlanmıştır. Ayrıca bir tane uluslararası *Science Citation Index Expanded* indeksli makale yayınlaması için gönderilmiştir. Çalışmalar ile ilgili bilgiler aşağıda belirtilmiştir:

- Demir, F., Şengür, A., Çavaş, M. (2018, September). Heart Sounds Classification with Deep Features and Support Vector Machines. In 2018 International Conference on Artificial Intelligence and Data Processing (IDAP), ss. 1-5, Malatya, Turkey

- Demir, F., Şengür, A., Bajaj, V., Polat, K. (2019). Towards the classification of heart sounds based on convolutional deep neural network, Health information science and systems, cilt 7(1), ss. 1-16
- Demir, F., Şengür, A., Bajaj, V., Polat, K. (2019). Convolutional neural networks based efficient approach for classification of lung diseases, Health information science and systems, cilt 8(1), ss. 1-4
- Demir, F., Bajaj, V., Ince, M. C., Taran, S., Şengür, A. (2019). Surface EMG signals and deep transfer learning-based physical action classification, Neural Computing and Applications, cilt 31(12), ss. 8455-8462

### 1.5. Tezin Organizasyonu

Tez çalışması altı bölümden oluşmaktadır. Birinci bölümde, tez konusu ile ilgili literatürde yapılmış olan çalışmalar ile bilgi verilmiştir. Ayrıca tezin, literatüre katkıları ve organizasyonu ilgili bilgilere de bu bölümde yer verilmiştir.

İkinci ve üçüncü bölümde, önerilen yöntemler hakkında teorik bilgilere değinilmiştir. Bu kapsamda ikinci bölümde, ön işleme aşamasında kullanılan spektrogramlar ile ilgili bilgi verilmiştir. Üçüncü bölümde ise ESA yapısı, ESA modelleri, kullanılan sınıflandırıcılar, verilerin normalleştirilmesi ve performans ölçüm teknikleri hakkında bilgi verilmiştir.

Dördüncü bölümde tez ile ilgili yapılan dört çalışmanın uygulanış metotları hakkında verilmiştir. Önerilen ilk çalışmada, kalp seslerinin derin öznitelikler tabanlı yaklaşımlarla sınıflandırılması üzerinde çalışılmıştır. Bu bölümdeki çalışmalar neticesinde 1 tane yurtiçi uluslararası bildirim sunulmuş ve 1 tane makale yayınlamıştır [157, 158]. Önerilen ikinci çalışmada, derin öznitelikler ve transfer öğrenme tabanlı yaklaşımların sınıflandırma performansları karşılaştırılmıştır. Bu bölümdeki çalışmalar neticesinde 2 tane makale yayınlanmıştır [159, 160]. Önerilen üçüncü çalışmada, derin öznitelikler tabanlı yaklaşımın sınıflandırma performansını daha da artırmak için öznitelik seçme algoritması kullanılmıştır. Bu bölümdeki çalışmalar neticesinde 1 tane yurtdışı uluslararası bildiri sunulması için gönderilmiştir. Önerilen dördüncü çalışmada, bütün katmanları uçtan uca yapılandırılmış bir ESA modeli çevresel seslerden oluşan veri setleri ile eğitilmiştir. Eğitilen bu ESA modelinden derin öznitelikler çıkarılmıştır. Önerilen Yöntem 2 ayrı veri seti üzerinde denenmiştir. Bu bölümdeki çalışmalar neticesinde 1 tane makale yayınlanması için gönderilmiştir.

Beşinci bölümde, önerilen 4 yöntem 10 ayrı veri setinde üzerinde denenmiştir. Önerilen yöntemlerin sonuçları, literatürde aynı veri setini kullanan diğer yöntemlerle karşılaştırılmıştır.

Altıncı bölümde, deneysel uygulama ile elde edilen sonuçlar ayrıntılı olarak irdelenmiştir. Deneysel sonuçların aynı veri setlerini kullanan diğer yöntemler üzerinden değerlendirilmesi yapılmıştır.

## 2. SPEKTROGRAM GÖRÜNTÜLER

Sinyal işleme, genellikle tıbbi hastalık teşhisinde [88-90], endüstriyel işlem kontrolünde [91], hata tespitinde [92] ve diğer birçok alanda öznitelik çıkarma ve sınıflandırma için kullanılmaktadır. Sinyal işlemenin temel amacı, belirli problemler hakkında temel bilgiler vermektir [93].

### 2.1. Giriş

Bu bilgiler genel olarak zaman, frekans veya zaman-frekans bölgesi tabanlı algoritmalar ile sağlanabilir. Zamanın fonksiyonu olarak bir sinyal zaman bölgesinde yer alırken, frekansın fonksiyonu olarak bir sinyal frekans bölgesindedir. Geleneksel sinyal işleme yaklaşımlarında, sadece ya frekans ya da zaman bölgesinde bilgiler sunulmaktadır ve bu yaklaşımlar çoğu zaman sinyallerin karakteristiklerini tam anlamıyla karşılanmayan durağanlığı kullanır. Bununla birlikte, kalp ve akciğer sesleri gibi sinyaller durağan olmayan sinyallerdir [94]. Bu seslerinin sadece zaman veya frekans bölgesini kullanan algoritmalarla öznitelik çıkarımını yapmak ses verisinin karakteristiğini tam anlamıyla yansıtmayacaktır. Dolayısıyla, zaman veya frekans bölgesinin tek başına kullanıldığı yaklaşımlar bu tür sinyaller hakkında kapsamlı bilgi sağlamak için yetersizdir [95]. Bu yüzden zaman-frekans gösterimi (ZFG) sağlayan yaklaşımlar, durağan olmayan sinyaller için daha uygundur. Dahası, son zamanlarda makine öğrenme alanında sıkça kullanılan ESA modellerinde resim haricindeki ses ve biyomedikal işaretlerin ZFG'leri giriş olarak kullanılmaktadır. ZFG'ler zaman ve frekans arasında bir köprü sağlarlar, ayrıca zamansal bilgiler ile spektral bilgileri eşzamanlı olarak sağlarlar. Bu nedenle, ZFG'ler, zamanla değişen frekanslar içeren sinyallerin temsili ve analizi için faydalıdır [96].

### 2.2. Kısa Zamanlı Fourier Dönüşümü

Bir fonksiyonun sabit bir zaman aralığındaki spektral içerikleri ile ilgili analiz yapılabilir. Giriş sinyali zamana bağlı olmadan örneklendiğinde, analiz genellikle bir pencere fonksiyonu ve daha sonra bir Ayrık Fourier Dönüşümü (AFD) uygulanarak yapılmaktadır [97].

Dijital sinyal işlemede işaret, belirli bir zaman aralığı boyunca (çoğunlukla bir pencere fonksiyonuyla tanımlanan) bir ses dalgasının basıncı, bir radyo sinyali veya günlük sıcaklık okumaları gibi zamana göre değişen herhangi bir büyüklük veya sinyaldir. Görüntü işleme sürecinde AFD için örnekler, bir görüntünün satırında veya sütunundaki piksel değerleri olabilir. Bir veri dizisinin AFD'si Denklem 2.1 ile hesaplanır.

$$X(k) = \sum_{n=1}^N x(n) e^{-j2\pi(k-1)(n-1)/N} \quad (2.1)$$

Fakat bir sinyalin AFD'si, sinyalin spektral içeriğinin zamanla nasıl değiştiğini göstermemektedir. Bu durum, pratikte birçok durağan olmayan sinyal için kritik öneme sahiptir. Bu nedenle bir sinyalin ZFG'si için, zamana bağlı olarak spektral değişimlerin uygun bir şekilde tanımlanmasını sağlamak için Fourier tabanlı analizler oluşturulmuştur. Bu nedenle, ZFG'nin temel amacı, belirli bir zamanda frekans eksenini boyunca enerji konsantrasyonunu belirlemektir. İdeal bir durumda, zaman-frekans dönüşümü, anlık bir frekans spektrumunun yerel bilgisini sinyalin zamansal davranışının global bilgisi ile birleştirerek, herhangi bir zamanda meydana gelen frekans bileşenleri hakkında doğrudan bilgi sağlamaktadır [98, 99]

Kısa süreli Fourier dönüşümü (KZFD), zaman içinde değişen bir sinyalin yerel bölümlerinin sinüzoidal frekansını ve faz içeriğini belirlemek için kullanılan Fourier ile ilişkili bir dönüşümdür. KZFD'nin hesaplanması için zaman eksenindeki uzun süreli bir sinyal eşit uzunlukta daha kısa bölümlere bölünür ve ardından her kısaltılmış bölümün Fourier dönüşümü (FD) pencere fonksiyonundan geçirilerek ayrı ayrı hesaplanır. Böylelikle anlık değişen spektrumlar zamanın bir fonksiyonu olarak temsil edilmiş olunur [100].

Sürekli-zaman durumunda, dönüştürülecek fonksiyon sadece belirli bir kısa zaman aralığı için sıfır olmayan bir pencere fonksiyonu ile çarpılır. Elde edilen sinyalin FD'si (tek boyutlu bir fonksiyon), pencere fonksiyonunun zaman eksenini boyunca kaydırılması ile alınır ve böylece sinyalin iki boyutlu (frekans-zaman) bir gösterimi elde edilir. Matematiksel olarak durum Denklem 2.2'deki gibi ifade edilir.

$$X(\tau, \omega) = \int_{-\infty}^{\infty} x(t) \omega(t - \tau) e^{-j\omega t} dt \quad (2.2)$$

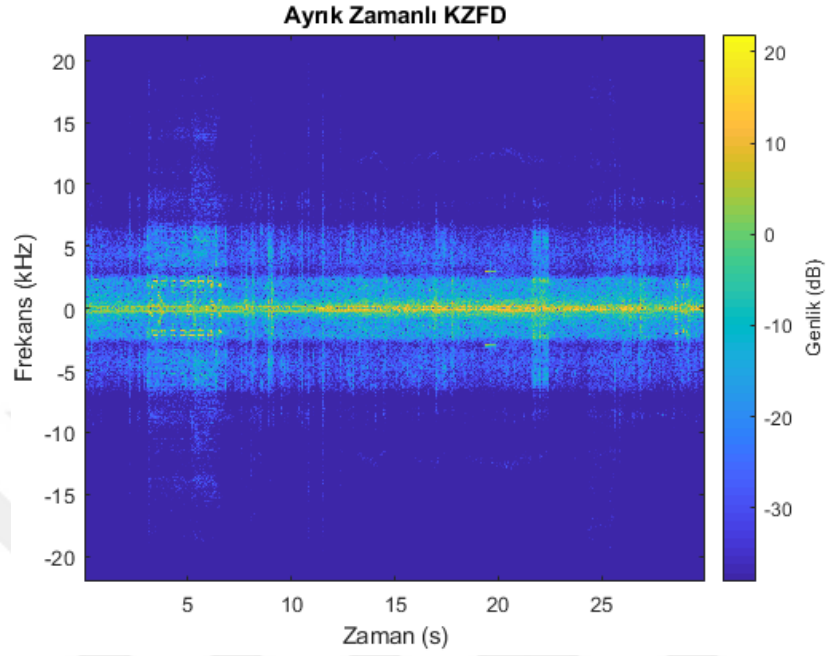
Burada  $\omega$ , Hann, Hamming penceresi veya sıfıra yakın merkezlenmiş Gaussian penceresi gibi pencereleme işlevidir ve  $x(t)$  dönüştürülecek sinyaldir.  $X(\tau, \omega)$ , esasen  $x(t)\omega(t-\tau)$ 'nin FD'sidir ve giriş sinyalinin zaman ve frekans üzerindeki fazını ve büyüklüğünü temsil eden karmaşık bir fonksiyondur.

Ayrık zamanlı durumda ise dönüştürülecek veri daha küçük çerçevelere bölünür. Bu durumda KZFD Denklem 2.3'teki gibi ifade edilir.

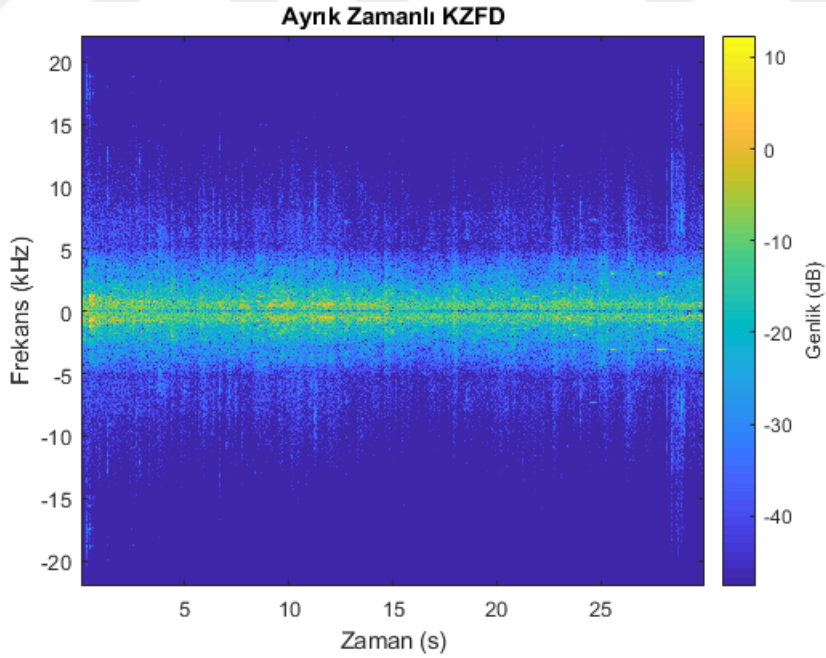
$$X(m, \omega) = \sum_{n=-\infty}^{\infty} x[n] \omega[n - m] e^{-j\omega n} \quad (2.3)$$

Denklem 2.2'deki ile benzer şekilde,  $x[n]$  giriş sinyalidir ve  $\omega[n]$  pencere fonksiyonudur. Bu durumda,  $m$  ayrıktır ve  $\omega$  sürekli, ancak çoğu tipik uygulamada KZFD, hızlı Fourier dönüşümü kullanan bilgisayarlardaki işlemlere uygulanır, bu nedenle her iki değişken de ayrık ve niceldir. Ayrıca sızıntının etkisini en aza indirmek için ve kullanılan pencereleme fonksiyonunun sınır bölgelerindeki sıfır olma durumlarını ortadan kaldırmak için *çakışma* kullanılır. Şekil 2.1 ve 2.2'de sırasıyla otobüs ve süpermarket seslerinin ayrık zamanlı KZFD dönüşümlerinin frekans-zaman gösterimleri verilmiştir. Her iki ses sinyalinin kayıt uzunluğu 30 saniyedir ve kayıt işlemi

44.1 KHz örnekleme frekansı ile gerçekleştirilmiştir. Ayrık zamanlı KZFD'nin pencere boyutu 1000 ve çakışma boyutu 100 olarak ayarlanmıştır.



Şekil 2.1. Otobüs sesinin ayrık zamanlı KZFD dönüşümü



Şekil 2.2. Süpermarket sesinin ayrık zamanlı KZFD dönüşümü

## 2.3. Pencere ve Pencereleme İşlemi

Boyutları büyük olan sinyaller genellikle daha kolay Fourier dönüşümlerinin oluşturulabilmesi için daha küçük boyutlu pencerelere (çerçevelere) ayrılırlar. Ayrılan bu çerçevelerin FD'si incelendiğinde ana frekans dışındaki frekanslarda da genlikler görülür. Bu genliklere spektral sızıntı denir. Spektral sızıntıları yok etmek için pencerenin belli kısımlarını sönmüleyen pencereleme fonksiyonları kullanılır [98].

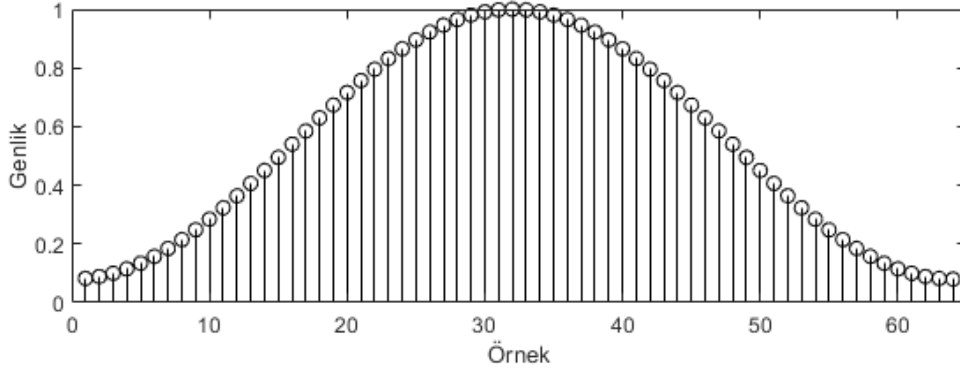
Durağan olmayan sinyaller için Hamming ve Hann daha uyumlu pencerelerdir. Bu pencereler, yaygın olarak, telefon kanalı spektrumu gibi dar bant uygulamalarında kullanılır.

### 2.3.1. Hamming Pencereleme İşlemi

Belirli katsayılarla sahip bu pencere Richard W. Hamming tarafından önerilmiştir [98]. Hamming pencereleme giriş değerlerine göre nasıl hesaplanacağı Denklem 2.4'de verilmiştir. 1'den 64'e kadar artan giriş örneğine göre Hamming pencereleme işleminin çıkışı Şekil 2.3'te gösterilmiştir.

$$\omega(n) = \alpha - \beta \cos\left(\frac{2\pi n}{N-1}\right) \quad (2.4)$$

Burada  $\alpha$ , 0,54 olarak,  $\beta = 1 - \alpha$  olarak kabul edilmektedir.

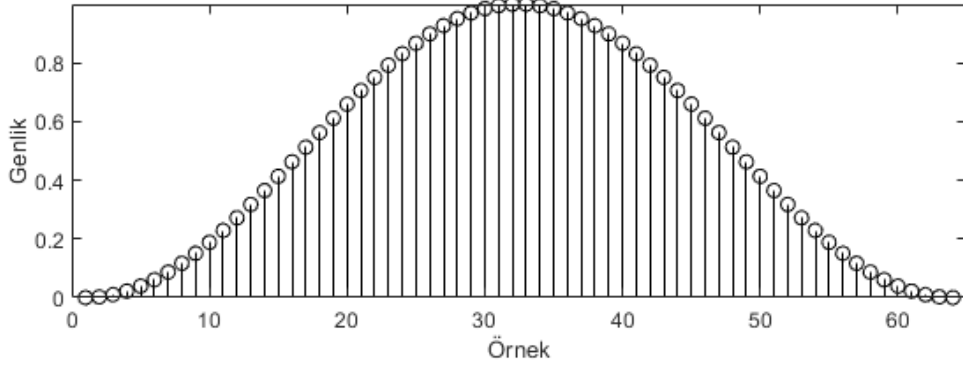


Şekil 2.3. 64 tane örnek için Hamming pencerelemesi

### 2.3.2. Hann Pencerelemesi

Kosinüs ve Hanning'in oluşturduğu takma adlarla geçen *Hann* penceresi, adını Julius von Hann'den alıyor [98]. Hann pencere fonksiyonunun giriş değerlerine göre nasıl hesaplanacağı Denklem 2.5'te verilmiştir. 1'den 64'e kadar artan giriş örneğine göre Hann pencereleme işleminin çıkışı Şekil 2.4'de gösterilmiştir.

$$\omega(n) = 0.5 \left( 1 - \cos\left(\frac{2\pi n}{N-1}\right) \right) = \sin^2\left(\frac{\pi n}{N-1}\right) \quad (2.5)$$



Şekil 2.4. 64 tane örnek için Hann pencerelemesi

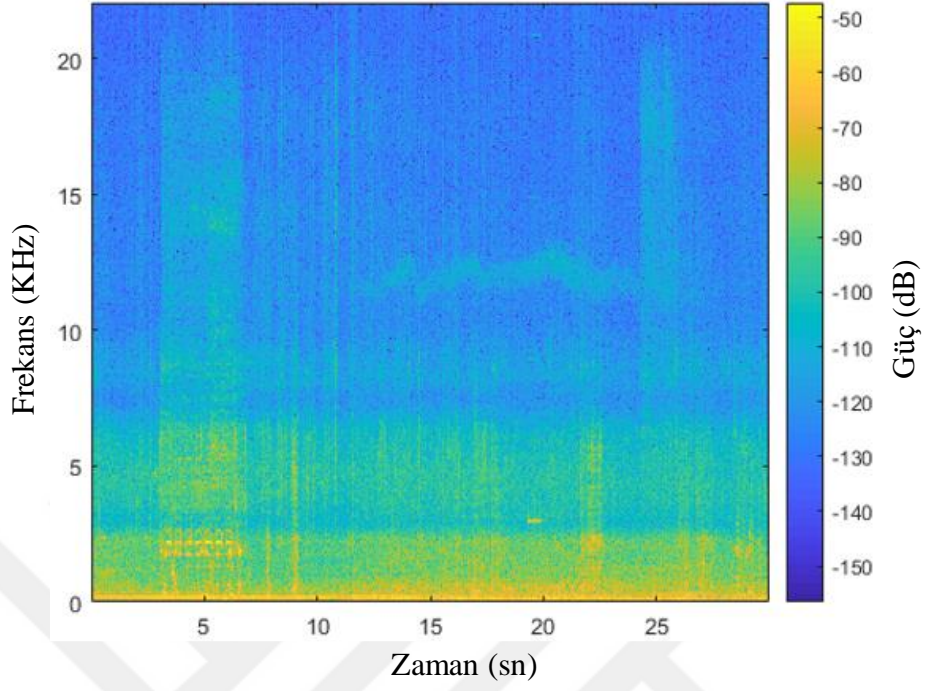
## 2.4. Spektrogram

Spektrogram, zaman değişkenine göre bir sinyalin frekans spektrumunun görsel bir gösterimidir. Spektrogramın gösterim formatında bir eksen zamanı, diğer eksen frekansı temsil etmektedir. Belirli bir zamanda belirli bir frekansın genliğini gösteren üçüncü bir boyut ise, görüntüdeki her bir noktanın yoğunluğu veya rengi ile temsil edilmektedir. Spektrogramın gösterim formatları değişkenlik gösterebilmektedir. Genelde yatay ekseninde zaman değişkeni tercih edilmektedir. Frekans ve genlik eksenleri, grafiğin ne için kullanıldığına bağlı olarak doğrusal veya logaritmik olabilir. Spektrogramda genlik eksenini desibel (dB) cinsinden ifade edilir ve logaritmiktir [102].

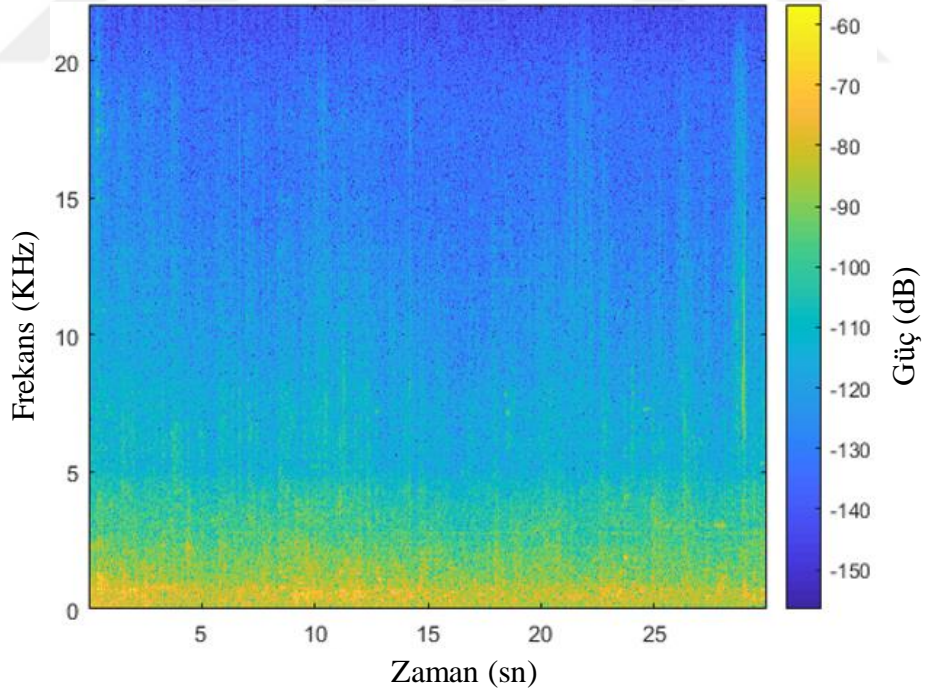
KZFD kullanarak spektrogram oluşturmak sayısal bir işlemdir. Zaman alanındaki sayısal olarak örneklenmiş veriler, genellikle çakışan çerçevelere bölünür ve her çerçevenin frekans spektrumunun büyüklüğünü bulmak için AFD tabanlı hesaplamalar yapılır. Spektrogram işlemi esasen bir ses sinyalinin  $\omega$  pencereli KZFD'sinin kare büyüklüğünün hesaplanmasına yani güç spektral yoğunluğu karşılık gelir [103]. Spektrogram, ayrık zamanlı KZFD ( $X(m, \omega)$ ) kullanımı ile Denklem 2.6'daki gibi tanımlanır.

$$\text{spektrogram}(m, \omega) = |X(m, \omega)|^2 \quad (2.6)$$

Şekil 2.5 ve Şekil 2.6'da sırasıyla Şekil 2.1 ve 2.2'deki ayrık zamanlı KZFD gösterimi yapılan sinyallerin, aynı çerçeve ve çakışma boyutları ile spektrogram gösterimleri verilmiştir. Spektrogramları verilen sinyallerin baskın genliklerinin 0 ile 2.5 KHz arasında olduğu görülmektedir. Ayrıca otobüs ve süpermarket spektrogramları arasındaki desen farkı belirgin şekilde gözükmemektedir.



Şekil 2.5. Otobüs sesinin spektrogramı



Şekil 2.6. Süpermarket sesinin spektrogramı

### 3. ÖZNİTELİK VE SINIFLANDIRMA AŞAMASINDA KULLANILAN YÖNTEMLER

Tezde, öznetelik ve sınıflandırma aşamasında Evrişimsel Sinir Ağları tabanlı yaklaşımlar kullanılmıştır. Sınıflandırma aşamasında literatürde en çok kullanılan algoritmalar seçilmiştir.

#### 3.1. Evrişimsel Sinir Ağları

Görüntü ile ilgili temsil algoritmalarında, pikseller kenarlara, kenarlar motiflere, motifler parçalara, parçalar nesnelere ve sahnelere monte edilmektedir. Bu durum iki yeni soru ortaya çıkarmaktadır: Her aşamada ne tür bir yöntem kullanılmalıdır? Ayrıca bu kadar derin ve çok aşamalı mimariler nasıl eğitilir? Günümüzde, Evrişimsel Ağlar, ilk sorunun cevabı olarak ortaya çıkmıştır. Yakın zamana kadar, ikinci sorunun cevabı gradyan temelli öğrenmeyi kullanmaktı, ancak son zamanlarda Derin Öğrenme (DÖ) konusundaki araştırmalarla birlikte yeni eğitim metotları da geliştirilmiştir [104]. Evrişimsel ağların, tamamen bağlı ağlara göre hesaplama maliyeti açısından daha fazla verimli olmasından dolayı, ESA modelleri ile birden fazla deneysel uygulama gerçekleştirmek ve üst değişkenleri ayarlamak daha kolaylaşmıştır.

ESA, makine öğrenmenin bir alt alanı olan DÖ tarihinde de önemli bir rol oynamıştır. ESA modelleri önceleri, önemli ticari uygulamaları çözen birçok alanda kullanılmıştır. Örneğin, 1990'larda, AT&T' şirketinin sinir ağı araştırma grubu çekleri okumak için ESA modeli geliştirmiştir [105]. 1990'ların sonunda, NEC şirketi tarafından da uygulanan bu sistem ABD'deki bütün çeklerin % 10'unu okumuştur. Daha sonra, evrişimsel ağlara dayanan birçok optik karakter ve el yazısı tanıma sistemi Microsoft tarafından kullanılmıştır [106]. DÖ'ye ticari ilginin mevcut yoğunluğu Krizhevski ve arkadaşlarının [107], ImageNet nesne tanıma yarışmasını kazanmalarından sonra daha da artmıştır.

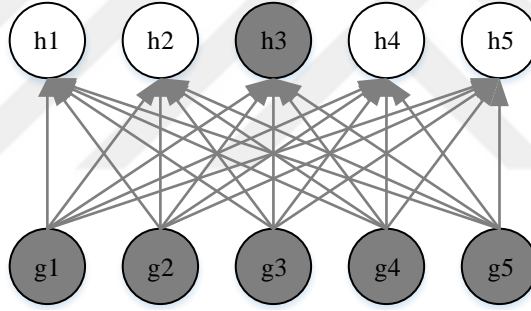
#### 3.2. ESA Kullanım nedenleri

ESA, klasik bir makine öğrenme sisteminin geliştirilmesi için oluşturulmuştur ve bu sistemlere göre avantajları bulunmaktadır. Bu avantajların en önemlileri, seyrek etkileşimler ve parametre paylaşımıdır. Dahası ESA, değişen büyüklükteki girişlerle çalışmak için kolay bir yol sağlamaktadır.

Geleneksel sinir ağı katmanlarında, her bir giriş birimi ve her bir çıkış birimi arasındaki etkileşimi açıklayan ayrı bir parametreye sahip bir ağırlık matrisi ile matris çarpım yapılmaktadır. Bu, her çıkış biriminin her giriş birimiyle etkileşime girdiği anlamına gelir. Bununla birlikte ESA klasik bir YSA'ya göre daha seyrek etkileşimlere sahiptir. Ayrıca bu duruma seyrek bağlantı veya seyrek ağırlıklar da denilmektedir. Bu durum, çekirdeğin girişten daha küçük boyuttaki hale

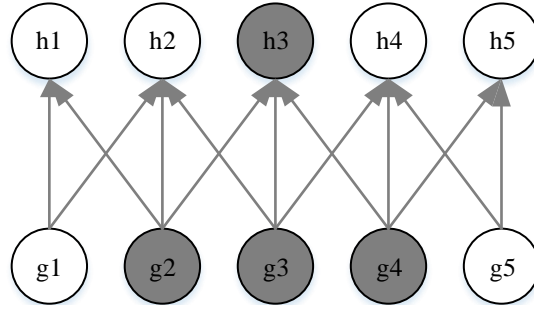
getirilmesiyle gerçekleşir. Örneğin, bir görüntüyü işlerken, giriş görüntüsünde binlerce veya milyonlarca piksel bulunabilir, ancak bu milyonlarca piksel içinden yalnızca onlarca veya yüzlerce pikseli kaplayan çekirdeklerden küçük ve anlamlı öznelilikler saptana bilmektedir. Bu, hem modelin bellek gereksinimlerini azaltmakta hem de istatistiksel verimliliğini artırmaktadır. Bu da, daha az parametre depolanması gerektiği anlamına gelmektedir. Ayrıca, çıkışa kadar olan hesaplarda daha az işlem gerektiği anlamına da gelmektedir. Genellikle verimlilikteki bu gelişmeler ESA'da oldukça büyük oranda olmaktadır. Örneğin, m tane girişi ve n tane çıkışı olan klasik bir YSA katmanında matris çarpımı  $m \times n$  parametre gerektirmektedir ve pratikte kullanılan algoritmalar için bu durum, belli bir çalışma zamanına mal olmaktadır. ESA'da ise k (m'den daha küçük) gibi bir giriş ile bağlanması gereken bağlantı sayısı sınırlanırsa, o zaman seyrek bağlı yaklaşımda sadece  $k \times n$  parametre gerekli olmaktadır ve bu durumda YSA katmanından daha az bir çalışma zamanına gerek duyulmaktadır [108].

Şekil 3.1'de 5 girişli (g1, g2, g3, g4, g5) ve 5 gizli katmanlı bir klasik sinir ağı için çıkış biriminin giriş birimleri ile etkileşim durumu gösterilmiştir.



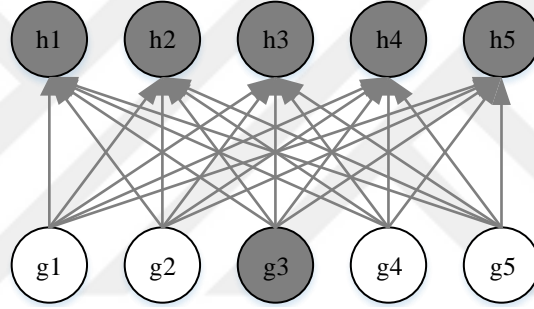
**Şekil 3.1.** Klasik bir sinir ağında çıkış biriminin giriş birimleri ile etkileşim durumu

Şekil 3.1'de gösterildiği gibi klasik sinir ağında h3 çıkış biriminin tüm giriş birimleri ile etkileşimi söz konusudur. Çıkış birimleri ile etkileşimi olan giriş birimlerine alıcı alan da denilmektedir. Şekil 3.2'de ise yine 5 giriş ve 5 çıkış birimi bir evrişim ağında bir çıkış biriminin hangi çıkışlar ile etkileşimde olduğu gösterilmektedir. Bu evrişim ağında filtre boyutu 3 olarak seçildiği için h3 çıkış biriminin g1, g2 ve g3 giriş birimleri ile etkileşimi söz konusudur.



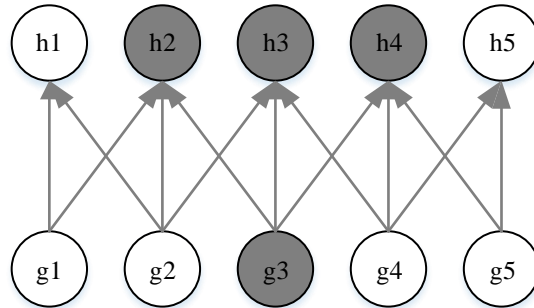
**Şekil 3.2.** Basit bir evrişim ağında çıkışı biriminin giriş birimleri ile etkileşim durumu

Şekil 3.3'te ise klasik bir sinir ağında bir giriş biriminin hangi çıkışlar ile etkileşimde olduğu gösterilmektedir. Şekilden de görüleceği üzere g3 giriş biriminin tüm çıkış birimleri ile etkileşimi vardır.



**Şekil 3.3.** Klasik bir sinir ağında giriş biriminin çıkış birimleri ile etkileşim durumu

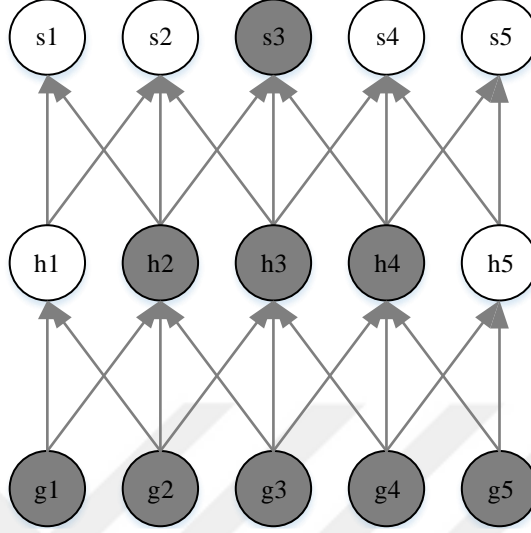
Şekil 3.4'te basit bir evrişim ağında bir giriş biriminin hangi çıkışlar ile etkileşimde olduğu gösterilmektedir. Şekilden de görüleceği üzere filtre boyutu 3 olduğu için g3 giriş biriminin h2, h3 ve h4 çıkış birimleri ile etkileşimi vardır.



**Şekil 3.4.** Basit bir sinir ağında giriş biriminin çıkış birimleri ile etkileşim durumu

Şekil 3.2'de evrişimsel ağda h3 çıkış birimi sadece filtre boyutu kadar giriş ile etkileşimde olsa da evrişimsel ağ daha da derinleştikçe daha da derinde ki çıkış birimleri giriş birimlerinin hepsi

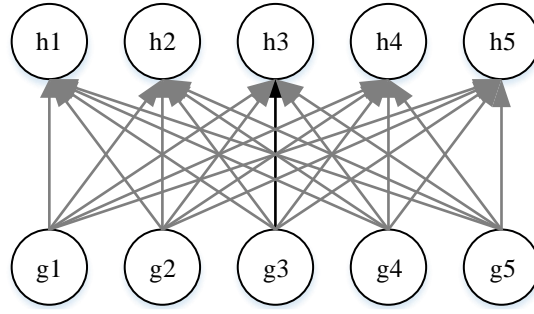
veya çoğu ile etkileşimde bulunur. Bu durum şekil 3.5’de vurgulanan etkileşim yapısına bir katman daha eklenerek gösterilmiştir.



**Şekil 3.5.** İki katmanlı bir evrişim ağındaki çıkış biriminin giriş birimleri ile etkileşim durumu

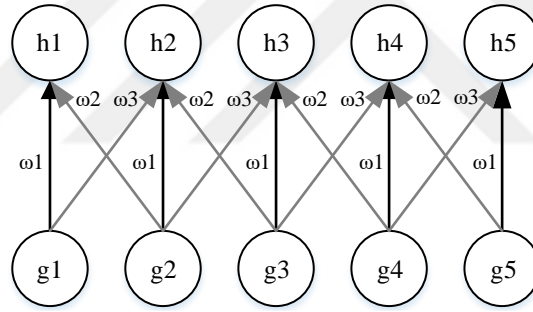
Şekil 3.5’te gösterildiği üzere s3 çıkış biriminin h2, h3 ve h4 çıkış birimleri ile etkileşimi bulunmaktadır. Bu çıkış birimlerinin ise tüm giriş birimleri ile etkileşimi söz konusudur.

Parametre paylaşımı, bir modelde birden fazla işlem için aynı parametrenin kullanılması anlamına gelmektedir. Geleneksel bir sinir ağındaki, ağırlık matrisinin her elemanı bir katmanın çıktısını hesaplamak için sadece bir kez kullanılmaktadır. ESA’da, çekirdeğin her üyesi girişin her konumunda kullanılır. Evrişim işlemi tarafından kullanılan parametre paylaşımı sayesinde her bir yer için ayrı bir parametre seti öğrenmek yerine, yalnızca bir ağırlık matrisinin öğrenilmesi yeterli olmaktadır. Bu durum, ileri yayılımın çalışma zamanını etkilemez. Ancak, evrişim işlemleri ile bellek gereksinimleri, istatistiksel verimlilik ve geri yayılımdaki hesaplama açısından geleneksel bir sinir ağına göre daha fazla verimlilik sağlanmaktadır. Şekil 3.6’da basit bir sinir ağı için parametre paylaşım durumuna bakılacak olursa koyu renkle gösterilen ok (ağırlık) sadece g3 giriş birimi ile h3 çıkış birimi arasında kullanılır. Bu durum diğer ağırlıklar içinde söz konusudur. Bu yüzden parametre paylaşımı yoktur. 5 giriş ve 5 çıkış birimi olan bu sinir ağındaki  $5 \times 5 = 25$  adet farklı ağırlık kullanılmaktadır [108].



Şekil 3.6. Basit bir sinir parametre paylaşım durumu

Şekil 3.7’de gösterilen evrişimsel bir ağda ise parametre paylaşımı ile çok daha az ağırlık kullanılmıştır. Aynı indislere sahip giriş biriminden gelen ağırlıkta  $\omega_1$ , kendi indisinden bir sonraki indisteki giriş biriminden gelen ağırlıkta  $\omega_2$  ve kendi indisinden bir önceki indisteki giriş biriminden gelen ağırlıkta  $\omega_3$  olmak üzere sadece 3 parametre kullanılmıştır. Önemli derecede azalan bu ağırlık parametreleri özellikle geri yayılım algoritmalarında hesaplama maliyetini ciddi oranda düşürmektedir.



Şekil 3.7. Evrişimsel bir ağda parametre paylaşımı

### 3.3. Evrişim İşlemi

Matematikte (özellikle, fonksiyonel analiz) evrişim, bir işaretin başka bir işaret tarafından nasıl değiştirildiğini ifade eden üçüncü bir işlev üreten matematiksel bir işlemdir. Evrişim işlemi genel olarak sürekli ve ayrık zamanlı olarak gerçekleştirilmektedir. Sürekli zamanda evrişim işlevi Denklem 3.1’de gösterilmiştir.

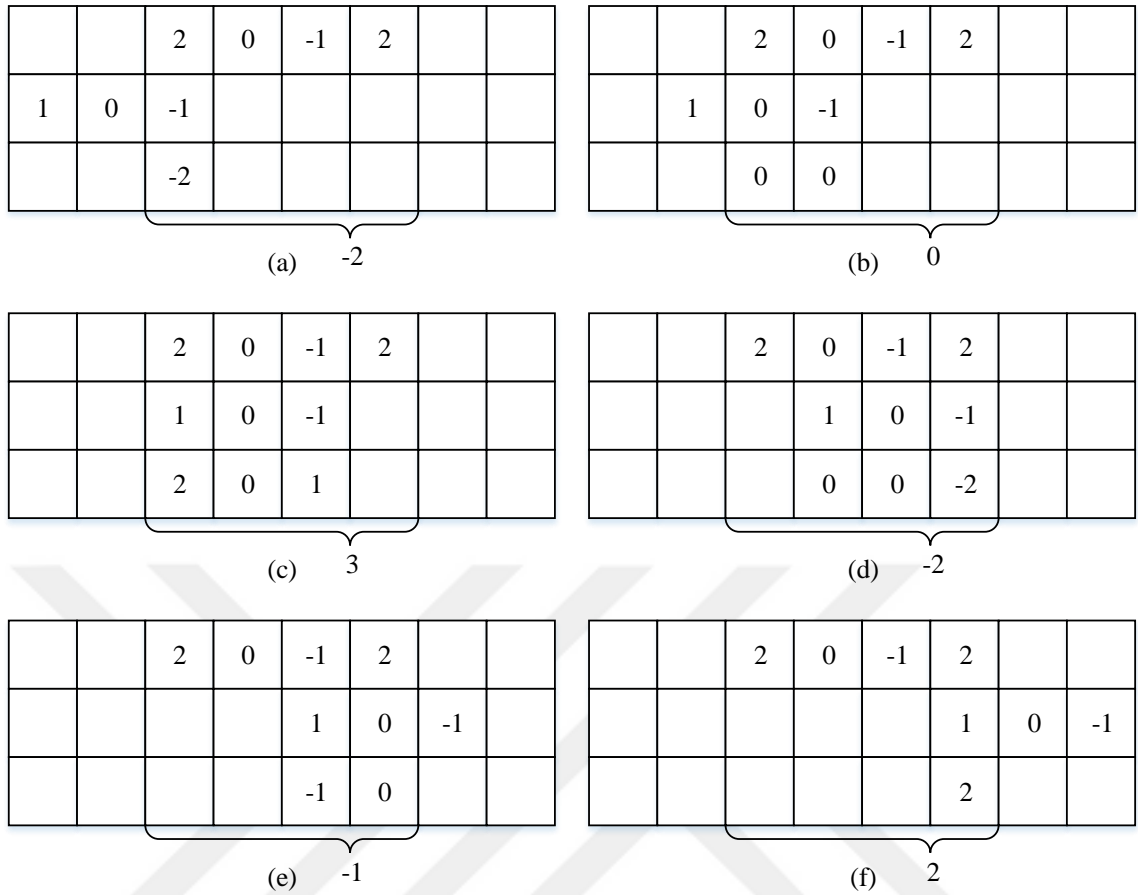
$$s(t) = (x * w)(t) = \int x(a)\omega(t - a)da \quad (3.1)$$

Burada evrişim işleminin simgesel gösterimi yıldız imidir.  $x$ , giriş işareti ve  $\omega$ , ise kullanılmak istenilen filtre işareti olarak farz edilebilir. Buna göre Denklem 3.1’de verilen formüle göre evrişim işlemi  $\omega$  filtre işaretinin  $a$  ekseninde  $t$  kadar ötelenerek ters simetriği alınır. Daha sonra,  $a$  ekseninde  $\omega$  filtre işareti  $x$  giriş işaretine doğru kaydırılır. İki işaretin kesişimde olduğu

tüm durumlara göre integral sınırları ve integrali alınan fonksiyon değişiklik gösterir. Ayrıca, giriş işaretinin ve filtre işaretinin türüne göre evrişim işlemin zorluk derecesi artmaktadır. Bununla birlikte görüntü, ses, video işleme ve makine öğrenme algoritmalarında kullanılan veriler belirli bir örnekleme frekansı ile örneklenecek sayısallaştırıldığı için ayrık zamanlı evrişim işlemi daha çok kullanım alanı bulmaktadır. Ayrık zamanlı evrişim işlemi Denklem 3.2’de gösterildiği hesaplanmaktadır [108].

$$s[n] = (x * w)[n] = \sum_{-\infty}^{\infty} x[k]\omega[n - k] \quad (3.2)$$

Burada Denklem 3.1’den farklı farklı olarak integral yerine toplam sembolü kullanılmıştır. Uygulanan evrişim işleminin yöntemi ise sürekli zamandaki durumla benzerdir. Ayrık  $\omega$  filtre işaretinin  $k$  ekseninde  $n$  kadar ötelenerek ters simetriği alınır. Bu  $k$  ekseninde  $\omega$  filtre işareti  $x$  giriş işaretine doğru kaydırılır. İşaretlerin kesişim bölgeleri toplanır ve ayrık işaret dizisi elde edilir. Ayrık zamanlı evrişim işlemine Şekil 3.8’de örnek verilmiştir. Bu evrişim işleminde  $x[n] = [2, 0, -1, 2]$  ve  $\omega[n] = [1, 0, -1]$  olarak seçilmiştir. Ayrık  $\omega$  filtre işaretinin ters simetriği alınarak ayrık  $x$  işaretine doğru kaydırılmıştır. Şekil 3.8’de gösterilen 6 adıma ((a)-(f)) göre ayrık evrişim işlemi gerçekleştirilmiştir. Bu altı adımda her kesişim bölgesinde değerler toplanmıştır. Böylelikle evrişim işleminin sonucu  $s[n] = [-2, 0, 3, -2, -1, 2]$  olarak elde edilmiştir.



Şekil 3.8. Ayırık zamanlı evrişim işleminin örnek gösterimi

ESA da kullanılan modeller çoğunlukla 2-boyutlu olduğu için evrişim işlemi de 2-boyutlu olarak yapılır. 2- boyutlu bir evrişim işlemi denklem 3.3'te gösterildiği gibi hesaplanır.

$$s[i, j] = \sum_m \sum_n I(m, n)K(i - m, j - n) \quad (3.3)$$

Burada  $I$ , 2-boyutlu giriş işareti  $K$  ise 2-boyutlu uygulanan çekirdektir. Evrişim işlemin sonucu oluşan çıkış işaretinin satır indisi,  $i$  ve sütun indisi ise  $j$ 'dir. Çekirdek ile evrişim işlemine maruz kalan giriş işaretinin çekirdek sınırları içerisindeki indisleri satır ve sütun için sırasıyla  $m$  ve  $n$  olmaktadır. Denklem 3.3'teki evrişim hesaplanması yapılırken dikkat edilmesi gereken önemli bir nokta  $K$  çekirdek değerleri  $I$  giriş değerleri ile çarpılırken  $K$  matrisinin ters simetrisinin kullanılması gerektiğidir.

Denklem 3.3'teki evrişim 2-boyutlu hesaplamasının daha iyi anlaşılması için örnek verilecek

olursa  $I = \begin{bmatrix} 1 & 1 & 0 & 1 \\ 2 & 0 & 1 & 2 \\ 1 & 1 & 2 & 0 \end{bmatrix}$  ve  $K = \begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix}$  için ve çekirdek matrisini ( $K$ ) 1 adım kaydırarak evrişim

işlemini aşağıdaki gibi hesaplanır. Bu hesaplamalarda  $s$  ve  $I$  matrislerinin indisleri 0'dan  $K$

matrisinin indisleri -1'den başlamaktadır ve aşağıdaki eşitlikler kullanılırken  $K$  matrisinin ters simetriği olan  $\begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}$  matrisi kullanılır.

$$s[0,0] = \sum_m \sum_n I(m,n)K(0-m, 0-n) \quad (3.4)$$

$$s[0,0] = I(0,0)K(0,0) + I(0,1)K(0,-1) + I(1,0)K(-1,0) + I(1,1)K(-1,-1) \quad (3.5)$$

$$s[0,0] = 1 \times 1 + 1 \times 0 + 2 \times 2 + 0 \times 1 = 5 \quad (3.6)$$

$$s[0,1] = \sum_m \sum_n I(m,n)K(0-m, 1-n) \quad (3.7)$$

$$s[0,1] = I(0,1)K(0,0) + I(0,2)K(0,-1) + I(1,1)K(-1,0) + I(1,2)K(-1,-1) \quad (3.8)$$

$$s[0,1] = 1 \times 1 + 0 \times 0 + 0 \times 2 + 1 \times 1 = 2 \quad (3.9)$$

$$s[0,2] = \sum_m \sum_n I(m,n)K(0-m, 2-n) \quad (3.10)$$

$$s[0,2] = I(0,2)K(0,0) + I(0,3)K(0,-1) + I(1,2)K(-1,0) + I(1,3)K(-1,-1) \quad (3.11)$$

$$s[0,2] = 0 \times 1 + 1 \times 0 + 1 \times 2 + 2 \times 1 = 4 \quad (3.12)$$

$$s[1,0] = \sum_m \sum_n I(m,n)K(1-m, 0-n) \quad (3.13)$$

$$s[1,0] = I(1,0)K(0,0) + I(1,1)K(0,-1) + I(2,0)K(-1,0) + I(2,1)K(-1,-1) \quad (3.14)$$

$$s[1,0] = 2 \times 1 + 0 \times 0 + 1 \times 2 + 1 \times 1 = 5 \quad (3.15)$$

$$s[1,1] = \sum_m \sum_n I(m,n)K(1-m, 1-n) \quad (3.16)$$

$$s[1,1] = I(1,1)K(0,0) + I(1,2)K(0,-1) + I(2,1)K(-1,0) + I(2,2)K(-1,-1) \quad (3.17)$$

$$s[1,1] = 0 \times 1 + 1 \times 0 + 1 \times 2 + 2 \times 1 = 4 \quad (3.18)$$

$$s[1,2] = \sum_m \sum_n I(m,n)K(1-m, 2-n) \quad (3.19)$$

$$s[1,2] = I(1,2)K(0,0) + I(1,3)K(0,-1) + I(2,2)K(-1,0) + I(2,3)K(-1,-1) \quad (3.20)$$

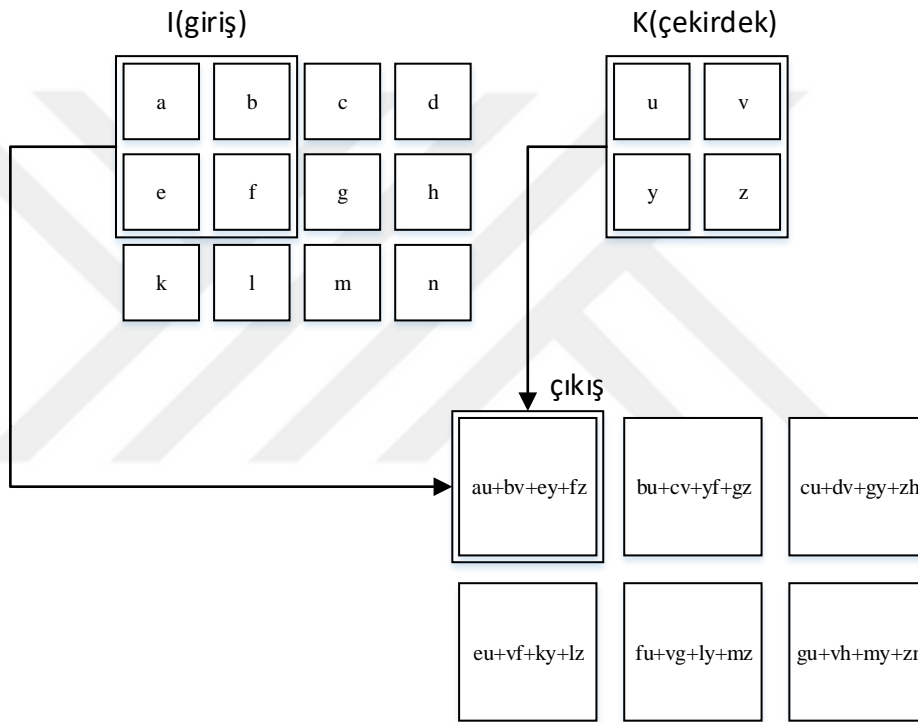
$$s[1,2] = 1 \times 1 + 2 \times 0 + 2 \times 2 + 0 \times 1 = 5 \quad (3.21)$$

Eşitlik 3.6, 3.9, 3.12, 3.15, 3.18 ve 3.21 indisler göz önüne alınarak bir matriste yerine koyulursa,  $S = \begin{bmatrix} 5 & 2 & 4 \\ 5 & 4 & 5 \end{bmatrix}$  çıkışı elde edilir. Ayrıca Denklem 3.3'e göre hesaplanan tüm bu evrişim sonuçlarının, çekirdek matrisinin ters simetriğinin alınmamış durumu ile giriş matrisinin örtüşen kısımları kullanılarak yapılan birebir çarpım ile aynı sonucu verdiğine dikkat edilmelidir. Bu hesaplamaların görsel gösterimi Şekil 3.9'da verilmiştir. Bununla birlikte giriş  $I$  matrisinin boyutu

3×4 ve kullanılan çekirdek matrisinin boyutu 2×2 iken evrişim çıkışında elde edilen matrisin boyutu 1 adım kaydırmayla 2×3 olmuştur. Giriş boyutu ve kullanılan filtre (ağırlık veya çekirdek) boyutuna göre evrişim işleminde elde edilen çıkış matrisinin boyutunun ne olacağı Denklem 3.22 ve 3.23'te verilmiştir.

$$\text{satur boyutu} = \frac{G-F_G+2E}{A_S} + 1 \quad (3.22)$$

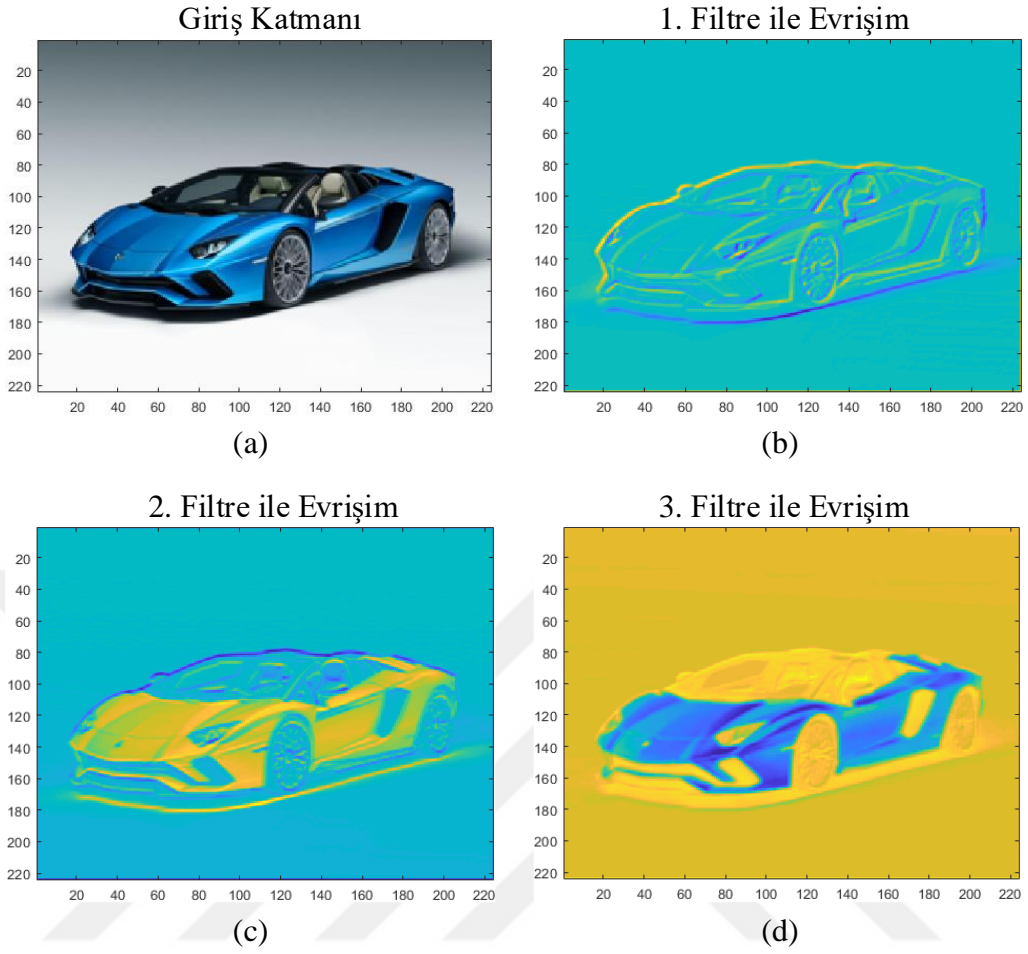
$$\text{sütün boyutu} = \frac{Y-F_Y+2E}{A_S} + 1 \quad (3.23)$$



Şekil 3.9. İki boyutlu matris durumu için ayrık zamanlı evrişim işleminin örnek gösterimi

Burada  $G$  ve  $Y$  sırasıyla giriş matrisinin satır ve sütun uzunluğu,  $F_G$  ve  $F_Y$  çekirdek (filtre) matrisinin satır ve sütun uzunluğu,  $E$  eklenen dolgu uzunluğu ve  $A_S$  ise adım sayısıdır.

Şekil 3.10'da 224×224×3 boyutundaki bir görüntünün 3×3 boyutundaki çekirdek ile evrişim işlemi sonucu oluşan görüntülere yer verilmiştir.



**Şekil 3.10.** 224×224×3 boyutundaki orijinal görüntü ve farklı çekirdek matrisleri ile evrişim işlemleri

Şekil 3.10'nun (a) bölümünde gösterilen orijinal görüntünün boyutu (224×224) değişmeyecek şekilde 3 ayrı evrişim işlemi uygulanmıştır. Boyutun değişmemesi için adım sayısı 1 ve dolgu ekleme ise giriş görüntüsünün satır ve sütun boyutunu 1 adım artıracak şekilde "1" olarak seçilmiştir. Farklı ağırlıklar kullanan çekirdek matrisleri (filtre) ile giriş görüntüsü arasındaki evrişim işleminin çıktıları Şekil 3.10'un (b), (c) ve (d) bölümünde gösterilmiştir. Bu kısımlardan da görüleceği üzere ağırlık değerlerine göre her bölümde giriş görüntüsünün farklı kısımları ön plana çıkmıştır.

### 3.4. Aktivasyon İşlemi

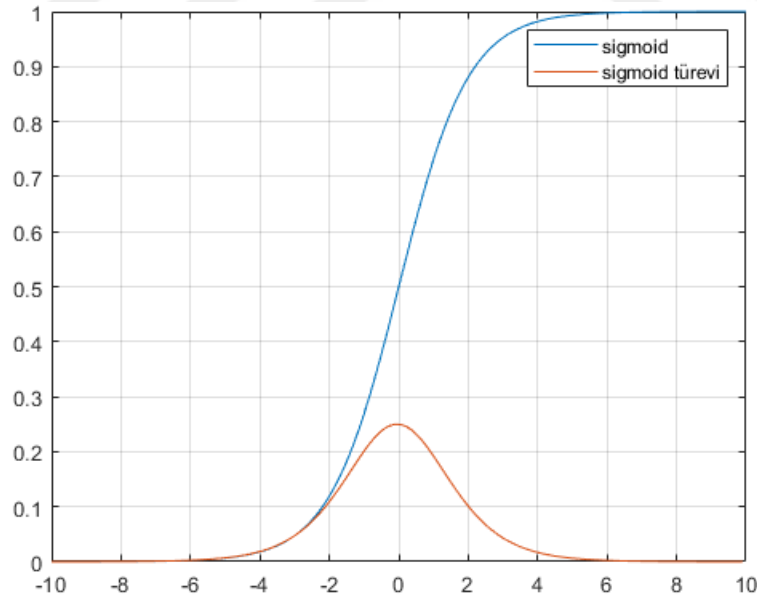
Aktivasyon fonksiyonlarının sinir ağlarındaki rolü önemlidir. Yapay sinir ağlarında aktivasyon fonksiyonlarına, gerçek dünyada var olan doğrusal olmayan bilgileri öğrenmek için ihtiyaç duyulmaktadır. Geleneksel bir sinir ağındaki bir nöronun girişi ağırlıklar ( $\omega_i$ ) ile girdilerin ( $x_i$ ) çarpımı ve bu çarpıma bias değerinin eklenmesinden oluşur ( $\omega_i x_i + b$ ). Bu doğrusal

fonksiyonun doğrusal olmayan bir fonksiyonla etkileşiminden nöron çıkışı oluşturulur. Bu işleme aktivasyon işlemi denir.

Hedeflenen öğrenmeyi gerçekleştirmek için sinir ağlarında kullanılan ağırlıklar çeşitli optimizasyon teknikleri ile sürekli güncellenmektedir. Bu güncelleme yapılırken çıkışta elde edilen hatadan başlayarak geri yayılım algoritması ile ağırlıklar hatayı azaltacak şekilde sürekli değiştirilir. Geri yayılım algoritması kullanımıyla ileri yönde kullanılan tüm denklemlerin zincir kuralı ile kısmi türevleri alınarak parametre değişim değerleri hesaplanır. Bu yüzden uygun aktivasyon fonksiyonunu seçmek bir sinir ağı modelinde çok önemlidir. Geleneksel YSA modellerinde en çok kullanılan aktivasyon fonksiyonları arasında en çok bilineni sigmoid fonksiyonudur [109]. Girdilere göre bir sigmoid fonksiyonun değişimi Denklem 3.24 ile elde edilir.

$$sigmoid = \frac{1}{1+e^{-t}} \quad (3.24)$$

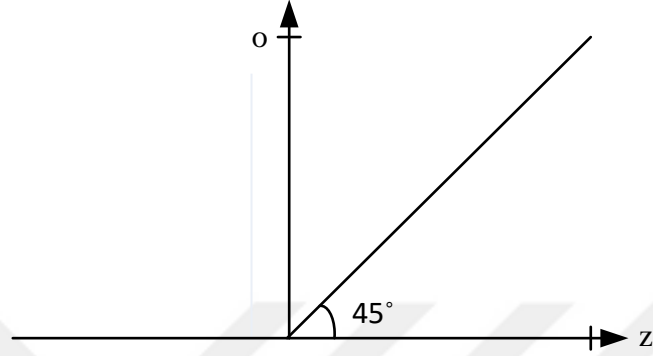
YSA'da popüler olan sigmoid aktivasyon fonksiyonu ESA gibi daha çok katmanlı ağlar kullanılmaya başlandıktan sonra *gradyanların kaybolması* adı verilen sorunlar oluşmuştur. Bu sorun hataları minimize etmek için sigmoid fonksiyonun kısmi türevi alınırken oluşmuştur. Şekil 3.11'de bir sigmoid fonksiyonu ve türevinin nasıl değiştiği gösterilmiştir. Girdi değeri olarak -10 ve 10 arasında düzenli aralıklarla değişen sayılar kullanılmıştır.



Şekil 3.11. Sigmoid Aktivasyon Fonksiyonu ve Türevinin Değişimi

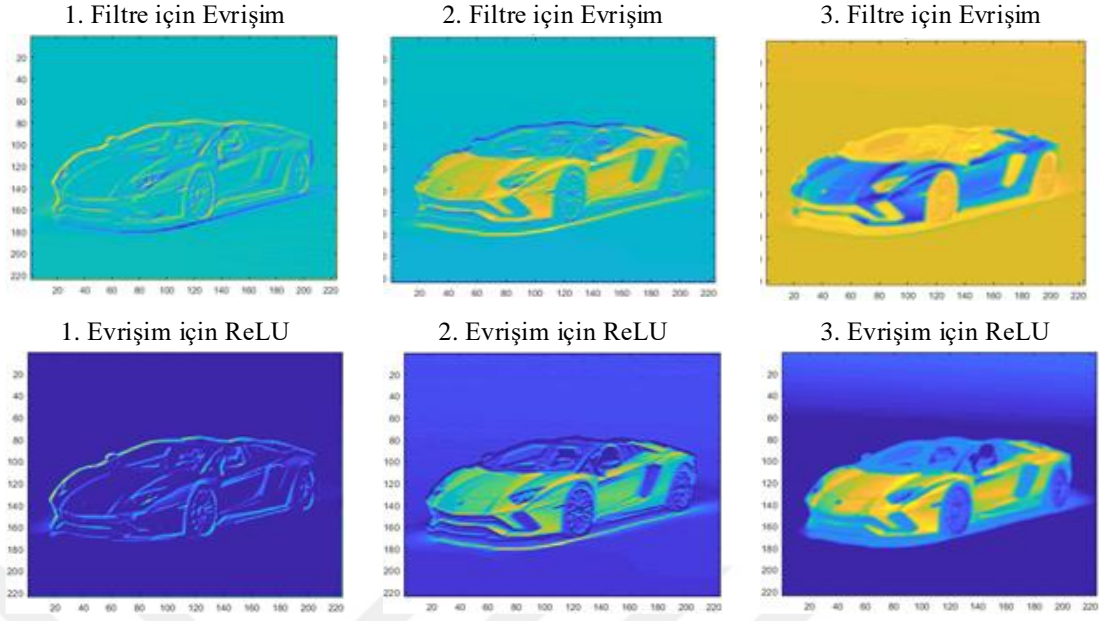
Şekil 3.11'de görüleceği üzere sigmoid fonksiyonun türevinin maksimum olduğu değer 0 girdisinde 0.25'tir. Eğer ESA gibi çok katmanlı bir ağda zincir kuralı ile birlikte birçok sigmoid fonksiyonu kısmi türevinin çarpıldığı düşünülürse bu değerler daha da küçülecektir. Sonuç olarak

eski ağırlık değerleri ile yeni ağırlık değeri hemen hemen aynı olacaktır ve öğrenme işlemi çok zorlaşacaktır [110]. ESA modellerinde bu sorundan kurtulmak için Düzleştirilmiş Doğrusal Birim (DDB) katmanı kullanılmıştır [111]. Bu katmanda kullanılan aktivasyon fonksiyonun girdilere göre değişimi Şekil 3.12’de verilmiştir.



Şekil 3.12. DDB katmanında kullanılan aktivasyon fonksiyonu

Şekil 3.12’de “z” aktivasyon fonksiyonunun girişi, “o” ise aktivasyon fonksiyonunun çıkışıdır ve aktivasyon fonksiyonu  $o = f(z) = \max(0, z)$  ile ifade edilir. Bu aktivasyon fonksiyonu pozitif bölgede doğrusal gibi gözükse de negatif bölgedeki tüm girdi değerlerine karşılık çıkışları 0 yaptığı için doğrusal değildir. Ayrıca DDB katmanında negatif girdilerin sifıra çekilmesi ile öğrenme hızı artmaktadır. Şekil 3.10’da evrişim işlemi uygulanmış görüntülerin DDB katmanında aktivasyon işlemine uygulandıktan sonraki durumu Şekil 3.13’de gösterilmiştir. Şekil 3.13’ün üst kısmında görüntüler 3 farklı ağırlığa sahip filtre ile yapılan evrişim işleminin sonucu görülmektedir. Şekil 3.13’ün alt kısmında ise bu evrişim işlemlerini DDB katmanından geçirilmiş hali gösterilmiştir.



Şekil 3.13. Evrişim çıktılarına DDB katmanının uygulanması

### 3.5. Havuzlama Katmanı

Havuzlama işlemi, genellikle DDB katmanından sonra gerçekleştirilmektedir. Havuzlama işleminin temel amacı bir sonraki evrişim katmanı için giriş matrisinin satır ve sütun boyutunun azaltılmasıdır. Havuzlama katmanında çıkışta matrisin derinlik boyutu değişmez. Ayrıca havuzlama katmanında yapılan bu işleme aşağı örnekleme de denilmektedir. Havuzlama işlemi sonucu veride bilgi kaybı oluşmaktadır. Bu bilgi kaybının iki yönden faydası bulunmaktadır. İlk olarak, kendisinden sonra bulunan ESA katmanları için daha az hesaplama işlemi yapılır. İkinci olarak ise ağır ezberlenmesi önlenir. Evrişim katmanında olduğu gibi belli filtreler giriş matrisine uygulanır. Bu filtreler giriş matrisi üzerinde belli bir adım kaydırma değerine göre gezdirilerek giriş matrisinin filtre sınırları içindeki maksimum, minimum ve ortalama değerlerinden bir tanesi kullanılarak havuzlama işlemi yapılır [111]. Yapıldığı işlemlere göre bu katman maksimum, minimum ve ortalama havuzlama katmanı adını almaktadır. Daha iyi sonuç verdiği için çoğunlukla maksimum havuzlama kullanılır. Havuzlama katmanından önceki evrişim katmanında kullanılan filtre sayısı kadar matriste havuzlama işlemi gerçekleştirilir. Bununla birlikte ESA'larda DDB katmanından sonra havuzlama katmanının kullanılması şart değildir. Havuzlama katmanında giriş ve filtre boyutuna göre çıkış boyutunun ne olacağı Denklem 3.25-3.28'e göre hesaplanır.

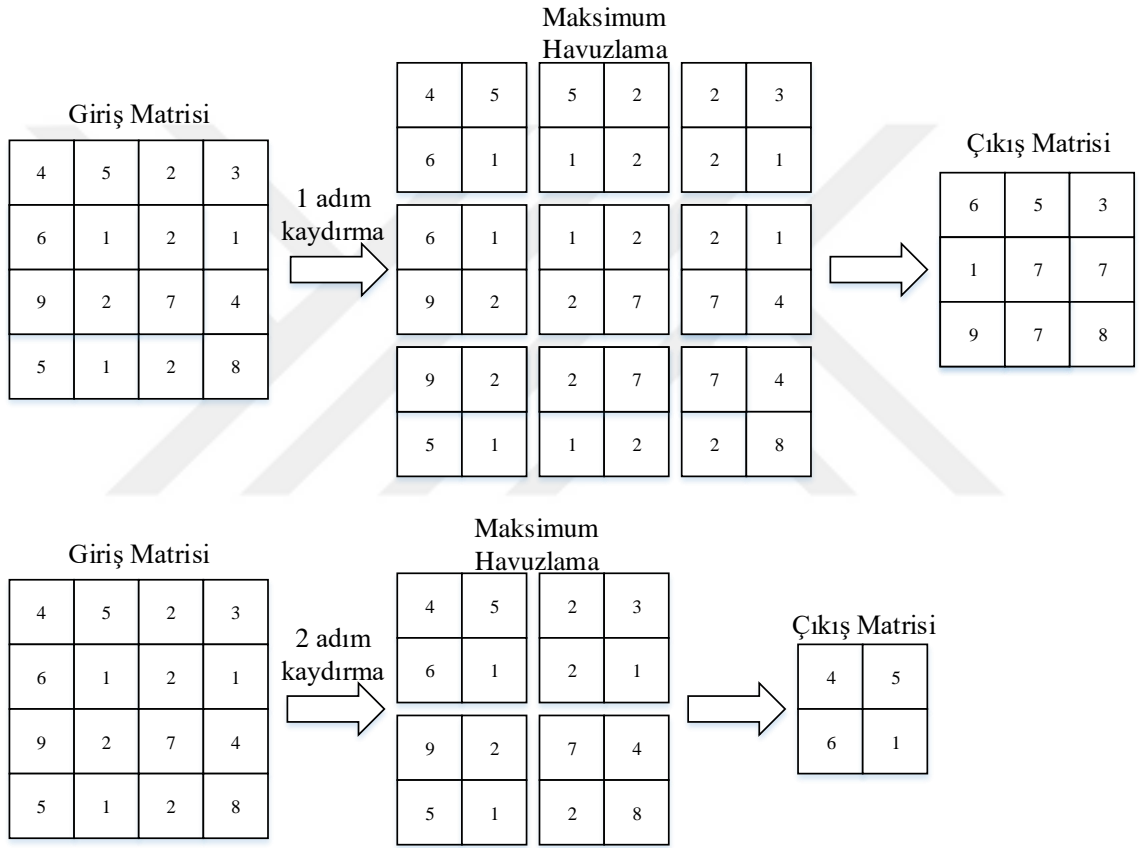
$$\text{Çıkış boyutu} = A_2 \times B_2 \times C_2 \quad (3.25)$$

$$A_2 = \frac{(A_1 - F)}{S} + 1 \quad (3.26)$$

$$B_2 = \frac{(B_1 - F)}{S} + 1 \quad (3.27)$$

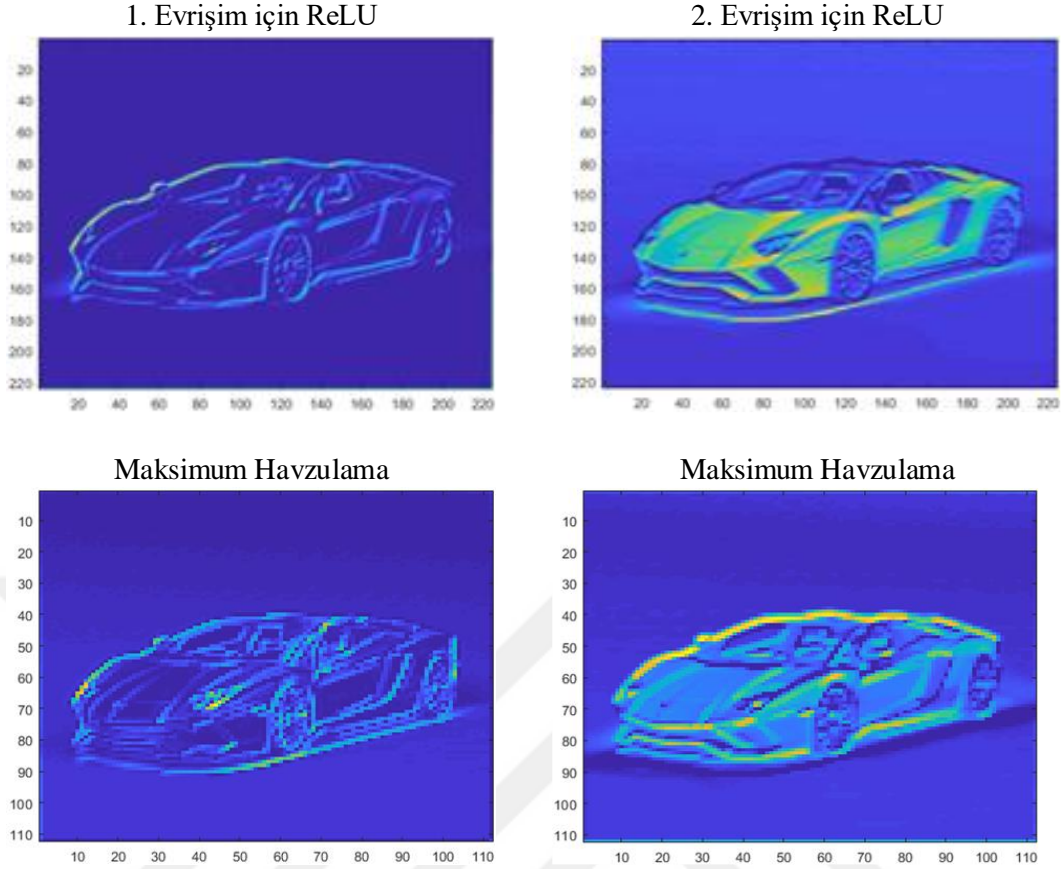
$$C_2 = C_1 \quad (3.28)$$

Burada,  $A_2$ ,  $B_2$  ve  $C_2$  sırasıyla giriş matrisindeki genişlik, yükseklik ve derinlik boyutu olup  $A_2$ ,  $B_2$  ve  $C_2$  ise sırasıyla havuzlama işlemi sonrası çıkış matrisindeki genişlik, yükseklik ve derinlik boyutudur.  $F$ , filtre boyutu ve  $S$ , adım kaydırma boyutudur. Şekil 3.14'te  $4 \times 4$  boyutlu giriş verisine,  $2 \times 2$  boyutlu maksimum havuzlama işlemi yapan filtre hem 1 adım kayma hem de 2 adım kayma ile uygulanmıştır. 1 adım kayma ile çıkış matrisinin boyutu  $3 \times 3$  ve 2 adım kayma ile çıkış matrisinin boyutu  $2 \times 2$  olmuştur.



Şekil 3.14. Maksimum havuzlama örnek gösterimi

Şekil 3.13'te verilen DDB katmanı çıkış görüntülerinin maksimum havuzlama katmanından sonraki değişimi Şekil 3.15'te verilmiştir. Bu katmanda  $2 \times 2$ 'lik filtre boyutu ile 2 adımlık kaydırma yapıldığı için giriş görüntüsünün boyutu  $224 \times 224$ 'ten  $112 \times 112$ 'ye düşmüştür. Ayrıca bu boyut düşümünden dolayı görüntülerin çözünürlüğünün dolayısıyla netliğinin de düştüğü görülmektedir.



Şekil 3.15. Maksimum havzulama ile görüntülerin değişimi

### 3.6. Toptan Normalleştirme Katmanı

Çok derin modeller, çeşitli fonksiyonların veya katmanların birleşimini içermektedir. Gradyan inme, diğer katmanların değişmediği varsayımı altında her bir parametrenin nasıl güncelleneceğini anlatmaktadır. Uygulamada, tüm katmanlar aynı anda güncellenmektedir. Fakat güncelleme yapılırken diğer işlevlerin (diğer katmanlar ve girişleri) sabit kaldığı varsayımı altında hesaplanan güncellemeler kullanılır ve oluşturulan birçok işlev aynı anda değiştirilir. Bu da ESA gibi çok katmanlı ağ yapılarında ağırlıkların istenilen düzeyden kaymasına neden olur. Bu durumda “iç değişken değişme” problemine neden olur. Bu problemi düzeltmek için Ioffe ve Szegedy Toptan Normalleştirme (TN) algoritmasını önermişlerdir [112]. Toptan normalleştirme katmanı hem ağın eğitim süresini hem de ağın ilk başlatılmasında kullanılan ağırlık değerlerine karşı hassasiyeti azaltmaktadır. Bu algoritma Denklem 3.29-3.32’de gösterildiği gibi oluşturulur. Denklem 3.29’da normalleştirilmiş etkilenim ( $\hat{x}_i$ ) değeri, giriş ( $x_i$ ), mini-toptan ortalama ( $m_b$ ) ve mini toptan değışınti ( $v_b$ ) değerleri ile hesaplanır.

$$\hat{x}_i = \frac{x_i - m_b}{\sqrt{v_b^2 + \epsilon}} \quad (3.29)$$

Burada  $\epsilon$ , bir sabittir ve  $v_b$ 'nin küçük olması durumunda sayısal kararlılığın korunmasına yardımcı olur.

$$m_b = \frac{1}{n} \sum_{i=1}^n x_i \quad (3.30)$$

$$v_b = \frac{1}{n} \sum_{i=1}^n (x_i - m_b)^2 \quad (3.31)$$

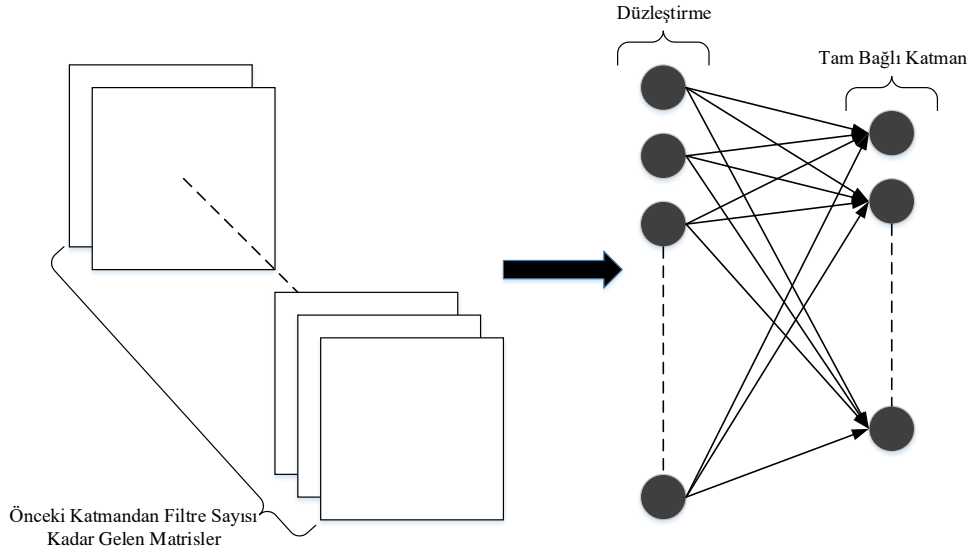
Son olarak TN katmanındaki etkilenim ( $y_i$ ) kayma ve ölçek parametreleri birlikte hesaplanır.

$$y_i = ax_i + b \quad (3.32)$$

Burada  $a$  ve  $b$  sırasıyla kayma ve ölçek faktörleridir. Ayrıca bu faktörler eğitim işlemi süresince öğrenilebilir parametrelerdir.

### 3.7. Tam Bağlı Katman

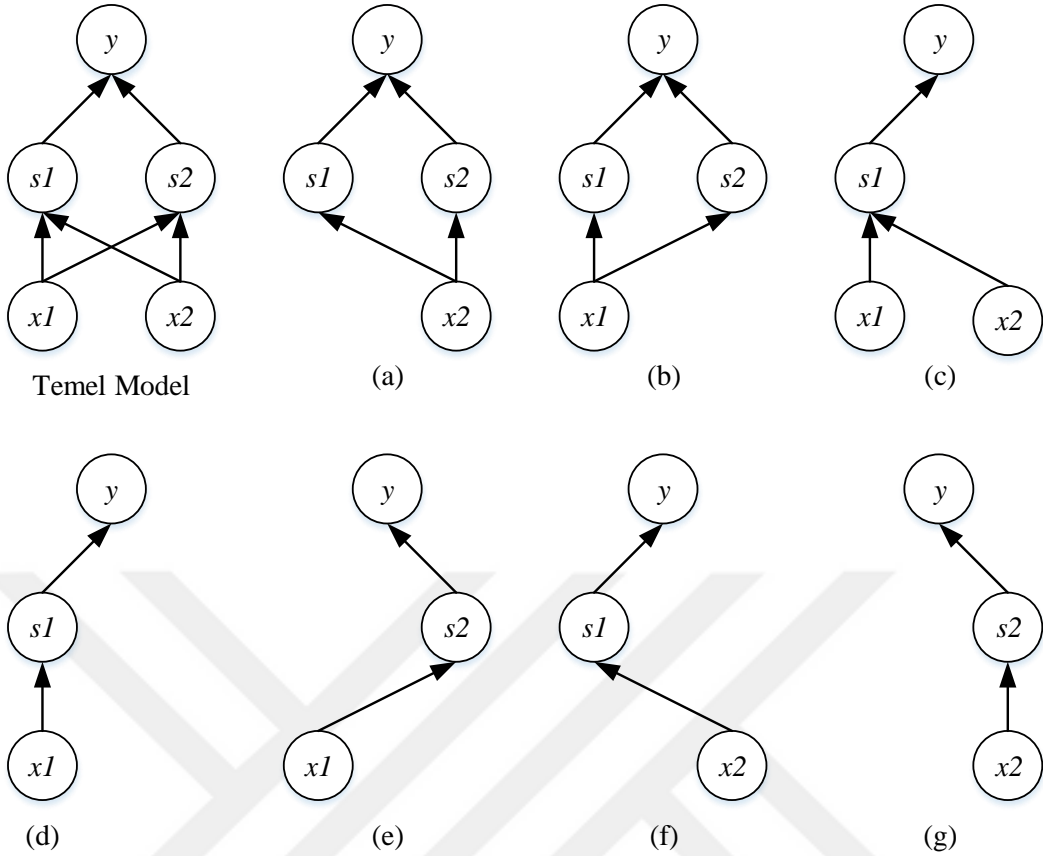
Tam Bağlı Katmanda bir önceki katmandan gelen ve yükseklik, genişlik ile derinlik değerlerine sahip olan matris verileri ilk önce sadece vektör yapısında olacak şekilde düzleştirilir. Bundan sonraki ağ yapısının bağlantı durumu klasik bir MLP ile benzerdir. Kullanılacak olan mimaride isteğe bağlı olarak ara Tam Bağlı Katmanlar oluşturulur. Örnek verilecek olunursa AlexNet ve VGGNet gibi önceden eğitilmiş ESA modellerinin yapısında ara Tam Bağlı Katman (TBK) bulunmaktadır. Ağ mimarisinde en son olarak kullanılacak olan TBK'da her giriş verisi için toplam sınıf sayısı kadar sınıf olasılık skorları bulunmaktadır. Örneğin 4 sınıflı 500 eğitim verisinden oluşan bir makine öğrenme probleminde TBK çıkışında her veri için 4 tane olmak üzere  $500 \times 4$  boyutunda bir yüksek seviyeli öznitelikleri temsil eden matris oluşturulur. TBK'nın temsili gösterimi Şekil 3.16'da verilmiştir. Bu gösterimde önceki katmanlardan gelen filtre sayısı kadar matris verisi düzleştirilmekte ve sonra her eğitim verisi için sınıf sayısı kadar nöron kullanılarak TBK oluşturulmaktadır.



**Şekil 3.16.** Tam Bağlı Katman temsili gösterimi

### 3.8. Seyreltme Katmanı

Seyreltme, çok katmanlı bir YSA modelini düzenli hale getirmek için kolay ve güçlü bir yöntem sunmaktadır [113]. Seyreltme, birçok büyük sinir ağının toplulukları için torbalamanın pratik hale getirilmesi ve aşırı öğrenmenin engellemesi için bir yöntem olarak düşünülmüştür. Torbalamada, birden fazla modelin eğitilmekte ve her test örneğinde birden fazla modelin değerlendirilmektedir. Her modelin büyük bir sinir ağı olduğu durumlarda, ağların eğitilmesi ve edilmesi çalışma süresi ve bellek açısından maliyetli olmaktadır. ESA modellerinde de ağlar eğitilirken eğitim verileri toptan eğitilmez. Önce belli bir boyutta küçültülmüş toptan gruplara (modellere) ayrılır. Daha sonra eğitim işlemi gerçekleştirilir ve her bir grubun çıkışından alınan değerlerin ortalaması alınarak bir çıkış belirlenir. Genelde 5-10 arasında sinir ağından oluşan toplulukların kullanılması yaygındır. Szegedy vd. [114] 2014 yılındaki ILSVRC görüntü sınıflandırma yarışmasını kazanmak için altı tane sinir ağı modeli kullanmıştır, ancak seyreltme katmanını kullanmadığından eğitim işlemi daha hızlı bir şekilde hantallaşmıştır. Bu durumdan sonra büyük ESA modelleri için seyreltme katmanı kullanımı yaygınlaşmıştır. Seyreltme katmanında belli eşik değerine göre her sinir ağı modelinde rastgele bazı düğümler ve bağlantılar işleme sokulmaz. Klasik bir YSA'da seyreltme işleminin temsili gösterimi Şekil 3.17'de verilmiştir. Şekil 3.17'nin (a)-(g) gösteriminde seyreltme işlemi yapılmamış 2 katmanlı basit bir sinir ağındaki farklı rastgele seyreltme senaryoları gösterilmiştir.



Şekil 3.17. Seyreltme katmanındaki seyreltme senaryoları

### 3.9. Eşiksiz En Büyük İşlev Katmanı

ESA'da TBK çıkışında bulunan her giriş verisi için sınıf sayısı kadar vektör değerleri bulunmaktadır. ESA'da bu değerleri uygun sınıf değerine atamak Eşiksiz En Büyük İşlev (EEBİ) kullanılır [115]. EEBİ her veri için olasılık skoru ürettiğinden çok sınıflı sınıflandırma problemlerinde tercih edilir. EEBİ Denklem 3.33 ile hesaplanır.

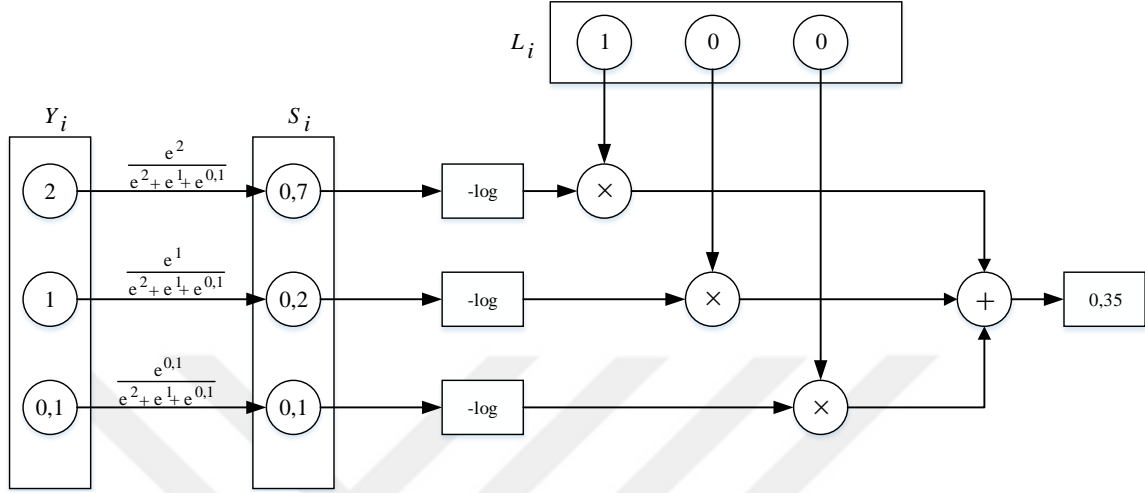
$$S_i = \frac{e^{y_i}}{\sum_j e^{y_j}} \quad (3.33)$$

Burada  $S_i$ , her değer için EEBİ çıkışıdır ve her "j" değeri sınıflara ait olma olasılığı için hesaplanır. TBK'dan gelen değerleri ise " $y_i$ " temsil eder. EEBİ çıkışından alınan tahmin sınıf etiketleri ile gerçek sınıf etiketleri arasındaki hata, Çapraz Düzensizlik (ÇD) maliyet fonksiyonu ile hesaplanır. ÇD maliyet fonksiyonu Denklem 3.34 ile hesaplanır.

$$D(S, L) = - \sum_i L_i \log(S_{(i)}) \quad (3.34)$$

Burada  $L_i$  gerçek sınıf etiketi değerleridir. ÇD maliyet fonksiyonu çıkışında alınan değer ne kadar az ise tahmin hata oranı o kadar az olur. Bu hata oranı istenilen değerlere inene kadar farklı

optimizasyon teknikleri ile ağıdaki bütün öğrenilebilir parametreler güncellenir. Bir önceki katmandan gelen verileri ( $Y_i$ ) kullanarak EEBİ ve sonrasında yer alan ÇD maliyet işlevi ile ilgili basit bir çözüm Şekil 3.18'deki yapıya göre hesaplanmıştır.



Şekil 3.18. EEBİ ve ÇD maliyet işlevi örnek gösterimi

Şekil 3.18'de ilk önce Denklem 3.34 kullanılarak EEBİ olasılık değerleri ve sonrasında Denklem 3.34 kullanılarak ÇD maliyet işlevi çıkışındaki yitim elde edilmiştir. Bu yitim değeri istenilen değere ulaşıncaya kadar eğitim işlemi sürdürülür.

### 3.10. ESA'da Sık Kullanılan Optimizasyon Teknikleri

Her sinir ağında istenilen doğruluk skorlarına ulaşmak için ağılar arası katmanlarda kullanılan ağırlıklar istenilen hata değerine ulaşıncaya kadar sürekli güncellenmesi gerekir. Bu güncelleme işlemi belli kurallar ile yapılır. Ağlarda kullanılan ağırlıkların birbiri ile etkileşimi göz önüne alınarak yeni ağırlık değerleri hesaplanır. Ağırlıkların etkileşimi bir sinir ağı mimarisinde girişten çıkışa kadar devam ettiği için ağda ileri ve geri yönde sürekli matematiksel işlemler yapılmaktadır. Geri yönde hesaplar yapılırken Zincir Kuralı uygulanır. Zincir kuralı, bir çıktının kendisini dolaylı olarak etkileyen bir girişe göre değişimi bulunmak istendiği zaman uygulanır. Basit bir örnek vermek gerekirse  $z = 2y + 5$  ve  $y = 3x + 6$  olsun. Değişken  $z$ 'nin değişken  $x$ 'e göre değişimi,  $z$ 'nin  $y$ 'ye göre kısmi türevi ile  $y$ 'nin  $z$ 'ye göre kısmi türevinin çarpımı ( $\frac{\Delta z}{\Delta x} = \frac{\Delta z}{\Delta y} \frac{\Delta y}{\Delta x}$ ) ile bulunur. Sinir ağlarında ise istenilen hedefe ulaşmak için ağıdaki öğrenilebilir parametreler, yitim işlevi ( $L$ ) temel alınarak güncellenir.

Makine öğrenme algoritmalarında en çok kullanılan optimizasyon tekniklerinin başında Rastgele Gradyan İnişi (RGİ) optimizasyonu gelmektedir. Bir sinir ağında yitim fonksiyonu ve ağırlıklar kullanılarak RGİ optimizasyonu Denklem 3.35'te gösterildiği gibi hesaplanır.

$$\omega_{i+1} = \omega_i - \eta \frac{\partial L}{\partial \omega_i} \quad (3.35)$$

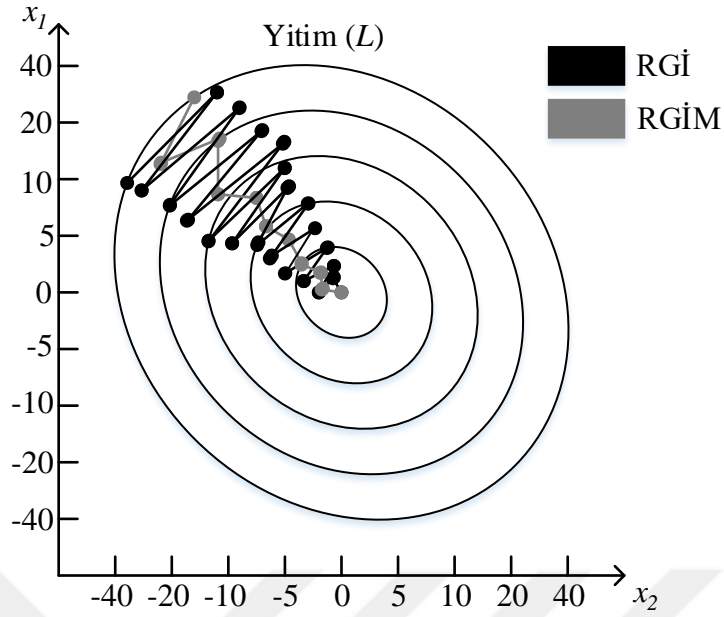
Burada  $\eta$ , öğrenme oranıdır ve keyfi olarak seçilebilir. Fakat tüm eğitim süresince tek bir öğrenme oranı kullanmak makul seçim değildir. Çünkü öğrenme oranı büyük seçilirse yitim işlevini en az hata değerini sağlayan global minimum noktası atlanabilir. Bununla birlikte öğrenme oranı küçük seçilirse eğitim süresi uzayabilir. Bunun için ağ parametreleri güncellenirken öğrenme oranında genelde iki yol izlenir. Bunlardan ilki eğitim başında yüksek öğrenme oranı, ortasında daha küçük bir öğrenme oranı ve sonunda daha da küçük bir öğrenme oranı kullanmaktır. İkinci ve daha akıllı bir yol ise öğrenme oranını, bir önceki öğrenme oranına ve içinde bulunulan andaki döngü sayısına bağlı olarak üstel olarak azaltmaktır. Bu şekilde öğrenme oranının değişimi Denklem 3.36'da verilmiştir.

$$\eta_i = \eta_{i-1} e^{-\frac{i}{20}} \quad (3.36)$$

Fakat öğrenme oranındaki bu ayarlamalara rağmen öğrenilecek verilerde gürültü varsa ve ilk ağ başlatma parametreleri hedef değerden çok uzak seçilmişse eğitim süresi yine de uzamaktadır. Bu problemin üstesinden gelmek için Sutskever vd. [116] RGİ optimizasyonuna bir momentum algoritmasını adapte etmiştir. Bu optimizasyon tekniğine Rastgele Gradyan İnişi Momentum (RGİM) optimizasyonu denilmektedir. RGİM optimizasyonu DÖ algoritmalarında sıkça kullanılmaktadır. RGİM optimizasyon tekniği ile ağırlıklar Denklem 3.37'deki gibi hesaplanır.

$$\omega_{i+1} = \omega_i - \eta \frac{\partial L}{\partial \omega_i}(t) - \alpha \frac{\partial L}{\partial \omega_i}(t-1) \quad (3.37)$$

Burada  $\alpha$ , 0-1 arasında değişen herhangi bir katsayıdır ve genelde 0,9 olarak seçilmektedir. “ $t$ ”, içinde bulunan döngüdeki değerleri, “ $t-1$ ” ise bir önceki döngüde kullanılan değerleri temsil etmektedir. Denklem 3.36'ya dikkat edilecek olunursa, bir önceki döngüde kullanılan değişim değerleri içinde bulunulan döngüdeki değişim değerlerini belli bir oranda etkilemektedir. Böylece öğrenilebilir parametreler hızlı değişirken birden yavaşlamaz veya yavaş değişirken birden hızlanmaz. RGİ ve RGİM optimizasyonları kullanılarak yitim değerlerinin 2 boyutlu eksenindeki değişiminin örnek bir gösterimi Şekil 3.19'da verilmiştir. Bu gösterime göre RGİ optimizasyon tekniği istenilen yitim değerine, RGİM optimizasyon tekniğinden daha fazla adımda ulaşmıştır. Başka bir deyişle RGİM optimizasyon tekniğinin eğitim işlemi daha az zaman almıştır.



Şekil 3.19. RGİ ve RGİM optimizasyon teknikleri için yitim değerinin değişimi

### 3.10.1. Uyarlanabilir Öğrenme Oranları Kullanan Optimizasyon Teknikleri

YSA araştırmacıları öğrenme hızının, güvenilir bir şekilde ayarlanması en zor olan üst değişkenlerden biri olduğunu ve model performansı üzerinde önemli bir etkiye sahip olduğunu fark etmiştir [116]. Yitim işlevi genellikle parametre uzayındaki bazı yönlerde oldukça duyarlıdır. Momentum algoritması ile bu sorunlar bir şekilde hafifletilmiştir, ancak bu yapılırken başka bir üst değişken eklenmiştir. Bu yüzden araştırmacılar öğrenme oranları üzerine yoğunlaşmıştır. Duyarlılık yönlerinin verilerdeki eksen hizaları bağlantılı olduğu kabul edilmesi durumunda, her parametre için ayrı bir öğrenme oranı kullanan algoritmalar geliştirilmiştir.

Delta-bar-delta algoritması, bireysel öğrenme oranlarını eğitim sırasında model parametrelerine uyarlamaya yönelik bir sezgisel yaklaşımdır [117]. Yaklaşım, basit bir fikir üzerine oluşturulmuştur. Bu yaklaşımda verilen bir model parametresine göre yitimin kısmi türevi aynı işaret olarak kalırsa, öğrenme oranı artmalıdır. Bu parametreye göre kısmi türev işareti değişirse, öğrenme oranı düşürülmelidir. Bu algoritma sadece Tam Toptan Optimizasyon teknikleri için uygulanabilmiştir. Bu algoritmanın dezavantajı büyük veriler eğitim süresinin çok uzamasıdır.

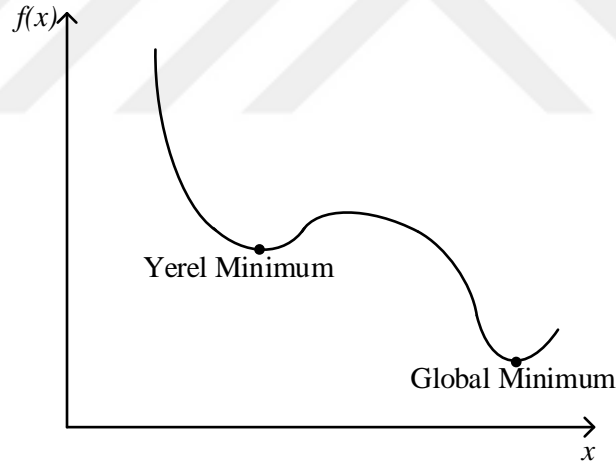
Son zamanlarda ise eğitim verilerini belli gruplar halinde eğitip, parametrelerini uyarlayan optimizasyon teknikleri geliştirilmiştir. 2011 yılında öne sürülen AdaGrad algoritması, tüm model parametrelerinin öğrenme oranlarını, tüm önceki kare değerlerinin toplamının karekökü ile ters orantılı olarak ölçekleyerek uyarlama yapmaktadır [118]. Yitimin en büyük kısmi türevine sahip olan parametreler, öğrenme oranlarında buna bağlı olarak hızlı bir düşüşe sahipken, küçük kısmi türevli parametreler, öğrenme hızlarında nispeten küçük bir düşüşe sahiptir. AdaGrad algoritmasında ağırlıkların değişimi Denklem 3.38 ve 3.39'a göre hesaplanır.

$$\Delta\omega_i(t) = -\frac{\eta}{\sqrt{G_i(t)+\epsilon}} \frac{\partial L}{\partial \omega_i}(t) \quad (3.38)$$

$$G_i(t) = G_i(t-1) + \left(\frac{\partial L}{\partial \omega_i}(t)\right)^2, G_i(0) = 0 \quad (3.39)$$

Burada “ $\epsilon$ ” küçük bir değerdir ve  $G_i(t)$ 'nin sıfır olma durumuna karşı denkleme eklenmiştir. Dışbükey optimizasyonlarda, AdaGrad algoritması istenen bazı teorik özelliklere sahiptir. Fakat bununla birlikte, deneysel uygulamalarda derin sinir ağı modellerinin eğitimi için eğitimin başlangıcından itibaren gradyan karelerin birikiminin etkili öğrenme hızında erken ve aşırı bir düşüşe yol açabileceği bulunmuştur. Bu nedenle AdaGrad optimizasyon tekniği bazı DÖ modellerinde iyi sonuç verirken bazı DÖ modellerinde iyi sonuç vermez. Bu nedenle, dışbükey olmayan ortamda daha iyi performans elde etmek için Adagrad algoritmasına ağırlıklı ortalama eklenerek RMSProp algoritması geliştirilmiştir [119].

Bir sinir ağını eğitilirken Şekil 3.20’de gösterildiği gibi dışbükey olmayan bir işlevle karşılaşıldığında, öğrenme yörüngesi birçok farklı yapıdan geçebilmekte ve sonunda yerel bir minimum bölgeye ulaşabilmektedir.



Şekil 3.20. Dışbükey olmayan bir fonksiyonun gösterimi

Böyle bir durumda AdaGrad algoritması öğrenme oranını, gradyanların karesinin bütün geçmişine göre küçültürken global minimum noktasına varmadan çok küçültülmüş olabilir. Bu durumu engellemek için RMSProp algoritması, geçmişi aşırı geçmişten silmek için ağırlıklı olarak azalan bir ortalama kullanmaktadır. RMSProp algoritmasında Denklem 3.38 değişmezken Denklem 3.39, Denklem 3.40’da gösterildiği gibi değiştirilir.

$$G_i(t) = \beta G_i(t-1) + (1-\beta) \left(\frac{\partial L}{\partial \omega_i}(t)\right)^2, G_i(0) = 0 \quad (3.40)$$

Burada  $\beta$ , 0-1 arasında değişen hareketli ortalamanın skaler değişimini belirleyen bir sayıdır ve genellikle 0,9 olarak seçilir.

RMSProp optimizasyon tekniğinde momentum değerine ayarlama yapan bir algoritma bulunmamaktadır. 2015 yılında RMSProp optimizasyon tekniğine momentum algoritması uyarlanarak güçlü bir optimizasyon tekniği geliştirilmiştir [120]. Bu optimizasyon tekniğine Adam algoritması denilmiştir. Adam algoritması RMSProp algoritmasının Denklem 3.38’de kullanılan  $\frac{\partial L}{\partial \omega_i}(t)$  gradyanın “ $M_i(t)$ ” momentum değişkeni ile değiştirilmesiyle oluşturulmuştur. Bu değişim Denklem 3.41 ve 3.42’de gösterilmiştir.

$$\Delta \omega_i(t) = -\frac{\eta}{\sqrt{G_i(t)+\epsilon}} M_i(t) \quad (3.41)$$

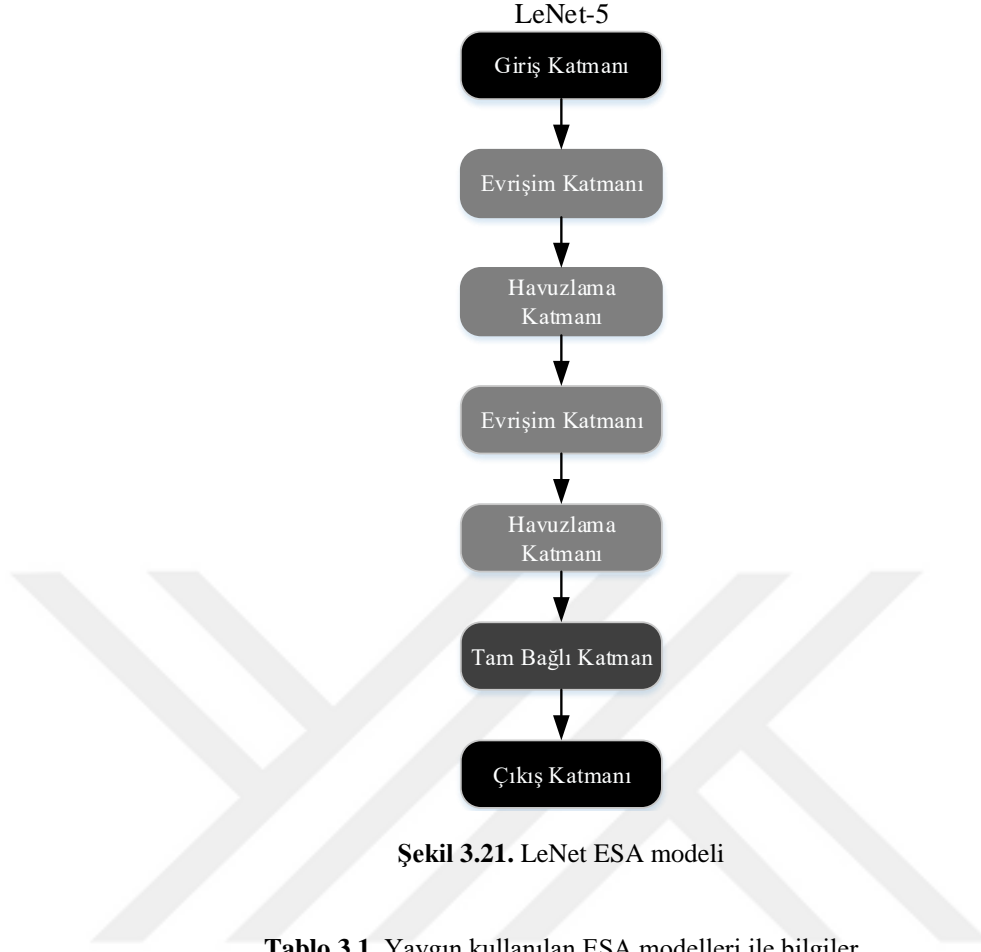
$$M_i(t) = \alpha M_i(t-1) + (1-\alpha) \frac{\partial L}{\partial \omega_i}(t), M_i(0) = 0 \quad (3.42)$$

Böylelikle Adam optimizasyon algoritmasında hem momentum hem de öğrenme oranı uyarlanabilir şekilde ayarlanabilmektedir.

### 3.11. Önceden Eğitilmiş ESA Modelleri

LeChun vd. [121] 1998 yılında Şekil 3.21’de gösterilen ESA mimarisi ESA modellerine öncülük etmiştir. Bu ESA modelinin 7 seviyeden oluşan ve çok derin olmayan bir yapısı bulunmaktadır ve giriş katmanının boyutu  $32 \times 32$  olacak şekilde oluşturulmuştur. Bu ESA modeli yüksek çözünürlüklü ve çok sayıda görüntülerin tanınması için yetersiz kalmıştır.

LeNet-5 modelinden sonraki ESA modelleri, büyük ölçekli resimleri tanımaya yönelik sınıflandırma problemleri için oluşturulmuştur. Bu ESA modelleri arasında ILSVRC yarışmaları için hazırlanan modeller yarışmada derece yapmış modellerdir. Bu modellerin hepsi 1000 sınıflı olmak üzere 1 milyondan fazla görüntü ile eğitilmiştir. Bu modellere genel olarak önceden eğitilmiş ESA modelleri denilmektedir. Bu modellerle ilgili bilgiler Tablo 3.1’de verilmiştir.



Şekil 3.21. LeNet ESA modeli

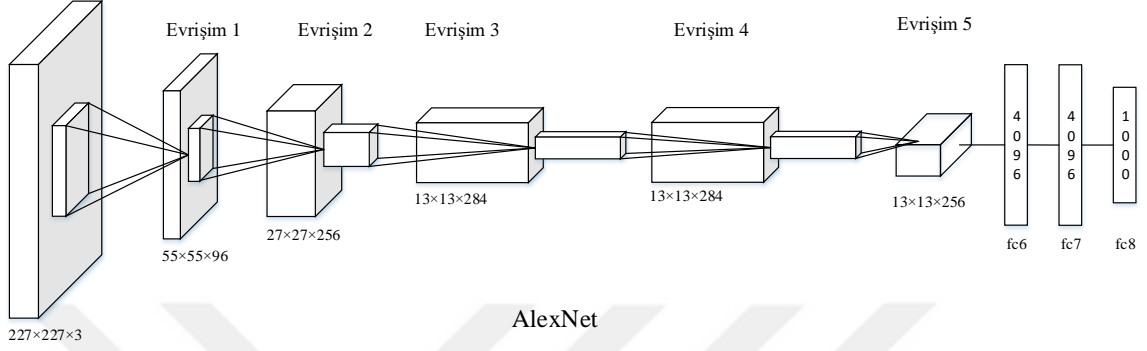
Tablo 3.1. Yaygın kullanılan ESA modelleri ile bilgiler

ESA modelleri	Katman sayısı	Parametre Sayısı(milyon)	Giriş Boyutu	Yarışma Adı
AlexNet	8	61	227×227	ILSVRC 12
VGG16	16	138	224×224	ILSVRC 14
VGG19	19	144	224×224	ILSVRC 14
GoogleNet	22	7	224×224	ILSVRC 14
ResNet18	72	11,7	224×224	ILSVRC 15
ResNet50	177	25,6	224×224	ILSVRC 15
ResNet101	342	44,6	224×224	ILSVRC 15

### 3.11.1. AlexNet

Alex vd. [45] diğerleri tarafından tasarlanan ESA modeli 2012 yılında düzenlenen ILSVRC görüntü tanıma yarışmasında birinci olmuştur. Ağ yapısı LeNet-5 modeline benzemektedir. Aynı seviyede 8 blok katman olmak üzere toplam 25 katmandan oluşan bir mimarisi mevcuttur. Evrişim filtreleri 11×11, 5×5 ve 3×3 boyutunda olup maksimum havuzlama katmanı(filtre boyutu 3×3 ve 2 adımlık kayma ), DDB katmanı, ikisi 4096 diğeri 1000 boyutlu olmak üzere 3 tane TBK ve seyreltme katmanı kullanılmıştır. Ayrıca evrişim katmanlarının çıkışında model performansını hızlandırmak için normalleştirme işlemi yapılmıştır. AlexNet modeli ile yarışmada alınan en iyi hata sonucu %16,4 olmuştur. AlexNet modelinin giriş katmanından son TBK'ya (fc8) kadar olan mimarisi Şekil 3.22'de verilmiştir. TBK katmanından sonra EEBİ katmanı ile sınıflandırma işlemi

yapılmaktadır. Her evrişim işlemimin ardından DDB katmanı ve maksimum havuzlama katmanı bulunmaktadır. Ayrıca fc6-fc7 ve fc7-fc8 TBK katmanları arasında seyreltme katmanı kullanılmıştır. Ayrıca AlexNet modelinde parametreleri optimize etmek için RGİM optimizasyon tekniği kullanılmıştır.

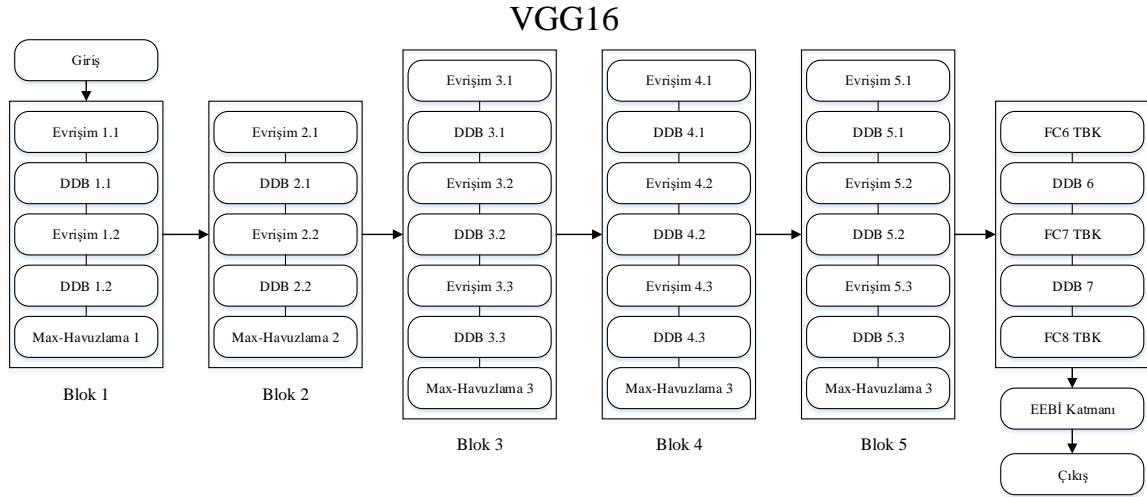


Şekil 3.22. AlexNet modeli

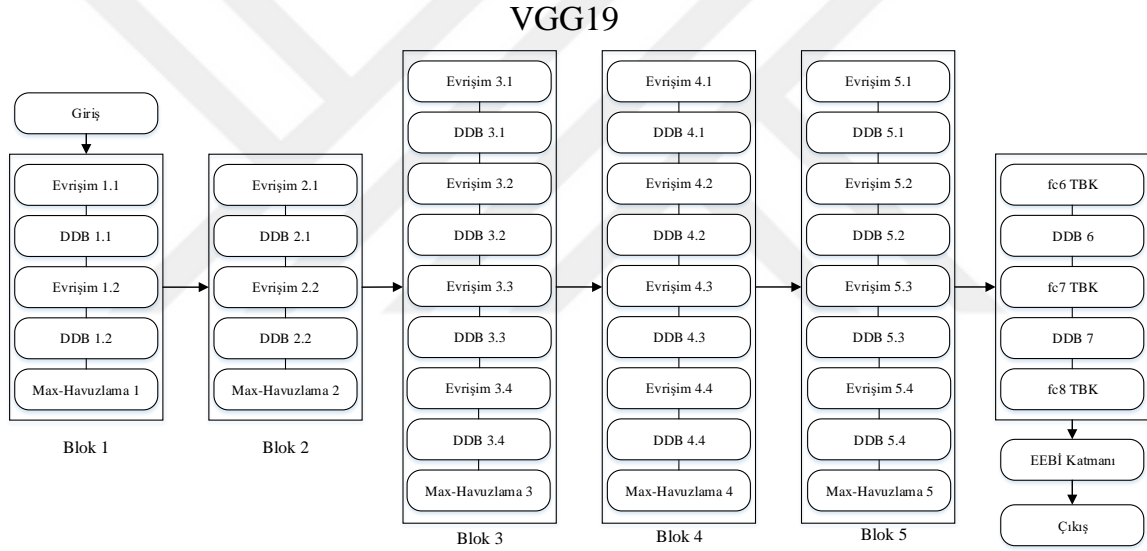
### 3.11.2. VGGNet

Simonyan vd. [46] diğerleri tarafından tasarlanan ESA modeli 2014 yılında düzenlenen ILSVRC görüntü tanıma yarışmasında GoogleNet modelinin ardından ikinci olmuştur. Blok yapısı itibariyle AlexNet modeline benzemekle beraber daha fazla  $3 \times 3$  boyutunda evrişim filtreleri kullanılmıştır. VGGNet modelinin eğitimi, yüksek kapasiteli 4 tane grafik işlemci birimi ile 3 hafta süresinde gerçekleştirilmiştir. VGGNet modelinin VGG16 ve VGG19 olmak üzere 2 modeli bulunmaktadır. VGG16 modelinin mimarisi Şekil 3.23'te verilmiştir. Şekilden görüleceği üzere Blok 1, 2 ve Blok 3, 4, 5'in içinde bulunan katmanların hiyerarşisi birbirine benzerdir. Ayrıca TBK katmanlarının yapısı AlexNet modeli ile birebir aynıdır ve %50 ile seyreltme katmanı kullanılmıştır. VGG16 modelinde kullanılan bütün evrişim filtrelerinin boyutu  $3 \times 3$ 'tür ve 1 adımlık kayma ve dolgu ile filtreleme yapılmaktadır. Blok numaralarına göre kullanılan filtre sayıları değişmektedir. Blok 1, 2, 3, 4 ve 5 için kullanılan filtre sayısı sırasıyla 64, 128, 256, 512 ve 512'dir. Tüm mimari boyunca kullanılan maksimum havuzlama filtresinin boyutu  $2 \times 2$ 'dir ve 2 adımlık kayma ile filtreleme yapılır.

VGG19 modelinin yapısı Şekil 3.24'te verilmiştir. Şekilden görüleceği üzere VGG19 modeli VGG16 modelinin Blok 3, 4 ve 5 nolu kısımlarına birer adet evrişim ve DDB katmanı eklenmesiyle oluşturulmuştur.



**Şekil 3.23. VGG16 modeli**



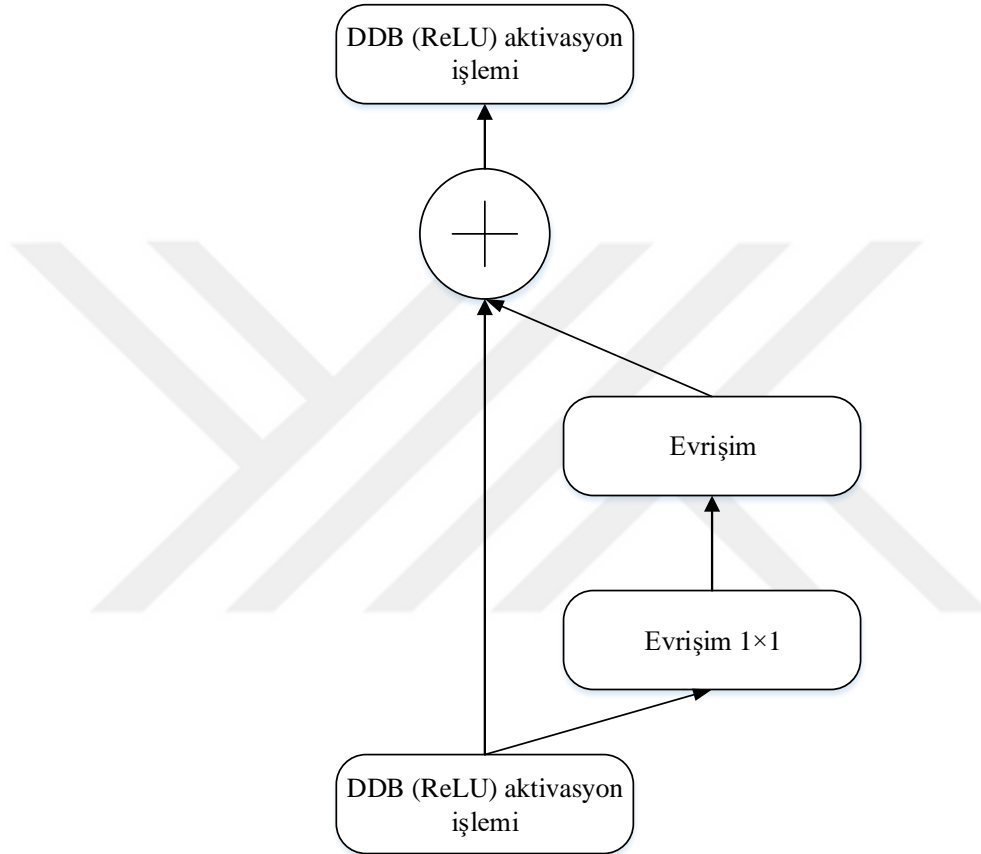
**Şekil 3.24. VGG19 modeli**

### 3.11.3. GoogleNet

Şezgedy vd. [114] diğlerleri tarafından tasarlanan ESA modeli, 2014 yılında düzenlenen ILSVRC görüntü tanıma yarışmasında birinci olmuştur. Ayrıca bu model başarımı yaklaşık 7 milyon parametre ile sağlamıştır. Modelin geneli, içerisinde Şekil 3.25'te gösterilen ağ içinde ağ yapısını içeren başlangıç katmanından oluşmaktadır. Başlangıç katmanının ilk seviyesine bir önceki başlangıç katmanının çıkışları bağlanmıştır. Bir önceki başlangıç katmanın çıkışı ise direkt olarak ikinci seviyedeki evrişim katmanına bağlanmıştır. Ayrıca başlangıç katmanın çıkışındaki tüm bağlantılar girişlerle paralel olarak bağlanmıştır. GoogleNet modelinde Şekil 3.26'da görüldüğü gibi katmanlar arası 2 ayrı yerden de çıkış almak mümkündür. Modelde  $1 \times 1$ ,  $3 \times 3$ ,  $5 \times 5$  ve  $7 \times 7$  boyutlu evrişim filtreleri kullanılırken  $3 \times 3$  maksimum ve  $5 \times 5$  ile  $7 \times 7$  ortalama havuzlama



mimarisinde artık katmanın ilk seviyesindeki evrişim filtrelerinin boyutu  $1 \times 1$  seçilerek artık katmandaki tüm verilerin optimizasyonu sağlanmaktadır. Artık katmanın, DDB aktivasyon fonksiyonu çıkışından direkt bir toplam alarak evrişim çıkışlarında öğrenmenin durduğu koşullarda bile çıkışa bir önceki aktivasyon değerinin aktarılması üzerine kurulmuş bir yapısı söz konusudur. Artık katman yapısı Şekil 3.27’de gösterilmiştir.



Şekil 3.27. ResNet modelinin artık katman yapısı

### 3.12. Sınıflandırma İşlemi

Veri bilimcileri, işlem yapılabilir büyük verilerdeki örüntüyü keşfetmek için birçok farklı türde makine öğrenme algoritması kullanmıştır. Yani sınıflandırma performansının en üst düzeye taşımaya çalışılmıştır. Günümüze kadar kullanılan algoritmalar genel olarak, yüksek düzeyde tahminlerde bulunma konusunda veriler hakkında öğrendiklerini temel alarak gözetimli ve gözetimsiz öğrenme olarak iki gruba ayrılabilir.

Gözetimsiz öğrenme, önceden belirlenmiş etiketleri olmayan veri setinde önceden bilinmeyen kalıpları bulunmasına yardımcı olan ve kendi kendine organize bir öğrenme türüdür.

Aynı zamanda “öz-örgütlenme” olarak da bilinir ve verilen girdilerin olasılık yoğunluklarının modellenmesine izin verir.

Gözetimsiz öğrenmede sadece girdiler bulunmaktadır ve bu girdilerin karakteristiklerine göre belli algoritmalar yardımıyla grupsal ayırım yapılmaya çalışılır. Gözetimsiz öğrenmede kullanılan ana yöntemlerden ikisi, Temel Bileşen Analizi (TBA) ve Kümelemedir. Kümeleme, etiketlenmemiş, sınıflandırılmamış veya kategorize edilmemiş verileri gruplayan bir makine öğrenim dalıdır. Kümeleme, gözetimli öğrenmede algoritmik ilişkileri tahmin etmek için veri kümelerini paylaşılan özelliklere göre gruplara ayırmak için kullanılmaktadır. Kümelemede geri bildirim cevap vermek yerine, küme analizi verilerdeki ortaklıklar arasındaki uzaklıklara göre hareket edilir ve her yeni veri parçasındaki bu ortak noktaların varlığına veya yokluğuna dayanarak tepki verilir. Bu yaklaşım, her iki gruba da uymayan anormal veri noktalarını tespit etmeye yardımcı olur. Kümelemede kullanılan algoritmalara K-Merkezli ve GKM algoritmaları örnek olarak gösterilebilir. TBA algoritmasında ise verilerdeki değişim büyüklüğüne göre yeni eksenler seçilip verilerin doğrusal olarak birbirinden ayrılması sağlanır.

Makine öğrenme algoritmalarının çoğunluğunda gözetimli öğrenmeyi kullanılır. Gözetimli öğrenme de, giriş değişkenlerinin “ $x$ ” ve bir çıkış değişkenlerinin “ $x$ ” olduğu durumda girişten  $y = f(x)$  çıkışına kadar eşleme (mapping) fonksiyonunu öğrenmek için bir algoritma kullanılmaktadır. Gözetimli denetimde, eşleme fonksiyonu o kadar iyi tahmin edilmelidir ki, yeni giriş verileri sisteme eklendiğinde çıkış değişkenleri doğru tahmin edilmelidir. Gözetimli öğrenme, algoritmayı eğitmek için kullanılan verilerin doğru cevaplarla etiketlenmesini gerektirir. Örneğin, bir sınıflandırma algoritması, hayvanın türleriyle ve bazı tanımlayıcı özelliklerle uygun şekilde etiketlenmiş bir görüntü veri kümesi üzerinde eğitildikten sonra hayvanları tanımayı öğrenecektir. Gözetimli öğrenmeye yapısal bağıntı, doğrusal bağıntı, DVM ve karar ağaçları algoritmaları örnek olarak verilebilir. Gözetimli öğrenme problemleri, bağlanım ve sınıflandırma problemleri olmak üzere iki gruba ayrılabilir. Her iki problemde öznitelik değişkenlerinden bağımlı özniteliğin değerini tahmin edebilen özlü bir modelin oluşturulması amaçlanmıştır. İki görev arasındaki fark, bağımlı özniteliğin bağlanım için sayısal ve sınıflandırma için kategorik olmasıdır. Örneğin yaş, ücret tahmini gibi problemler sayısal değerler olduğu için bağlanım problemi, bir kişinin hasta olup olmaması, bir e-mailin spam içerip içermemesi veya bir hayvanın türünün belirlenmesi gibi problemler kategorik olduğu için sınıflandırma problemidir. Ele alınan tez konusunda gözetimli denetim ile birlikte kategorik verilerle çalışıldığı için sınıflandırma problemi üzerine durulmuştur [123].

### 3.12.1. K-En Yakın Komşuluk

K-En Yakın Komşuluk (K-EYK) algoritması, gözetimli öğrenme kategorisinde olup hem sınıflandırma hem de bağlanım problemleri için kullanılmakla beraber genelde pratiğe dayalı

uygulamalarda sınıflandırma probleminin çözümü için tercih edilmektedir. Cover ve Hart tarafından [124] 1967 yılında öne sürülen K-EYK algoritmasında önceden belirlenmiş etiketli olan bir veri seti mevcuttur. Sınıflandırılacak olan yeni veriler bu veri seti üzerinden sınıflandırılır. Yeni verinin hangi kategoriye ait olduğunu belirlemek için etiketli veri setindeki verilerle uzaklık durumuna bakılır. Bu uzaklıklar Denklem 3.43-45 arasında verilen Öklid, Minkowski ve Manhattan gibi uzaklık denklemleri ile hesaplanır.

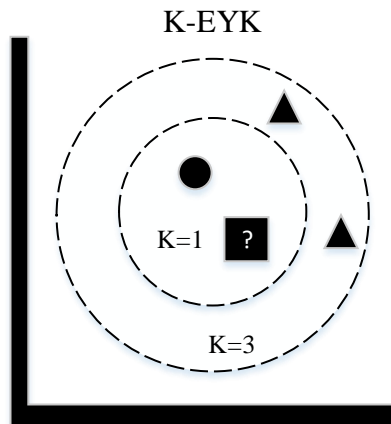
$$\text{Öklid} = \sqrt{\sum_{i=1}^N (x_i - y_i)^2} \quad (3.43)$$

$$\text{Manhattan} = \sum_{i=1}^N |x_i - y_i| \quad (3.44)$$

$$\text{Minkowski} = (\sum_{i=1}^N (|x_i - y_i|)^k)^{1/k} \quad (3.45)$$

Minkowski denkleminde  $k=1$  verilirse Manhattan,  $k=2$  verilirse Öklid uzaklık denklemleri elde edilir.

Kaç tane veriye göre yeni verinin kategorisinin belirleneceği ise K parametresi ile belirlenir. Yeni verinin etiketi yeni veriye en yakın K kadar verinin içerisinde çoğunluk oylaması yapılarak seçilir. Şekil 3.28’de K-EYK sınıflandırma algoritması ile ilgili bir örnek gösterim verilmiştir. Şekil 3.28’de  $K=1$  olarak seçilirse yeni verinin etiketi “daire”,  $K=3$  olarak seçilirse yeni verinin etiketi “üçgen” olur. Dolayısıyla yeni verinin etiketinde K parametresinin etkisi fazladır. Bunların dışında K-EYK algoritmasında eğitime ihtiyaç duyulmaması, gürültülere karşı dirençli olması algoritmanın avantajı iken, büyük veri setlerinde hesaplama maliyetinin artması, performansın sadece kontrol edilmeyen uzaklığa göre değişmesi ve K parametresine göre sonucun değişkenlik göstermesi dezavantajdır.



Şekil 3.28. K-EYK örnek gösterimi

### 3.12.2. Destek Vektör Makineleri (DVM)

Hem bağlantım hem de sınıflandırma işleminde kullanılan DVM algoritması 1995 yılında Vapnik tarafından önerilmiştir [125]. DVM algoritmasında ayırtaç tabanlı yöntem uygulanır. İki sınıf arasındaki en geniş kenar payı ve bununla birlikte en uygun üstündüzlem optimizasyonla tahmin edilmeye çalışılır. DVM algoritması doğrusal ve doğrusal olmayan sınıflandırma problemlerinde güçlü bir sınıflandırıcı olarak kullanılmaktadır.

İki sınıflı bir sınıflandırma probleminde sınıf etiketlerinin birisi “+1” diğeri “-1” olsun.  $X = \{x_i, y_i\}$  öğrenme örneklemindeki giriş verileri “ $x_i \in S_1$  ise  $y_i = 1$ ” “ $x_i \in S_2$  ise  $y_i = -1$ ” değerlerini alsın. Bu durumda “ $\omega$ ” ağırlık parametresi ile “ $\omega_0$ ” yanlılık parametrelerinin sağlaması gereken koşullar Denklem 3.46 ve 3.47’de verilmiştir.

$$\omega x_i + \omega_0 \geq +1 \text{ eğer } y_i = +1 \quad (3.46)$$

$$\omega x_i + \omega_0 \leq -1 \text{ eğer } y_i = -1 \quad (3.47)$$

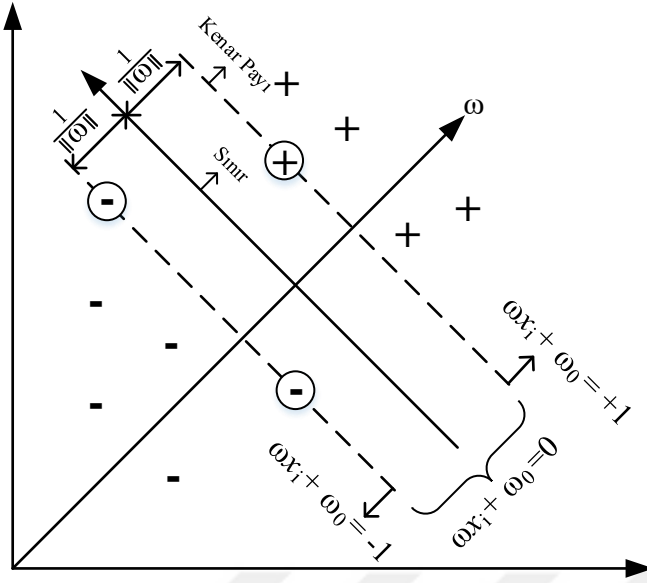
Denklem 3.46 ve 3.47 tek satırda düzenlenerek Denklem 3.48 elde edilir.

$$y_i(\omega^T x_i + \omega_0) \geq +1 \quad (3.48)$$

Bu koşulda  $y_i(\omega x_i + \omega_0) \geq 0$  olmaması istenir. Çünkü giriş verilerinin konumunun, doğruya yakın bölgelerden ziyade daha iyi bir genelleme için doğruya belli bir mesafede olması beklenmektedir. Bu mesafelerin her iki tarafında bulunan en yakın giriş verilerinin sınıra uzaklığına kenar payı denmektedir ve genellemenin iyi olması için kenar payının olabildiğince büyük olması gerekmektedir.

Şekil 3.28’de gösterilen iki sınıflı giriş verilerinin ( $x_i$ ) ayırtaca uzaklığı Denklem 3.49’daki gibi hesaplanır.

$$d_1 = \frac{|\omega x_i + \omega_0|}{\|\omega\|} \quad (3.49)$$



Şekil 3.29. İki sınıf için DVM üstündüzlem gösterimi

$x_i \in \{-1, +1\}$  olması durumunda Denklem 3.49, Denklem 3.50'deki gibi düzenlenir.

$$d_2 = \frac{x_i(\omega x_i + \omega_0)}{\|\omega\|} \quad (3.50)$$

Denklem (3.50)'deki elde edilen değerin en az belli bir eşik değerinde ( $q$ ) olması istenirse Denklem 3.51 elde edilir.

$$\frac{x_i(\omega x_i + \omega_0)}{\|\omega\|} \geq q \quad (3.51)$$

Burada  $\omega$  ağırlıklarını ölçeklemeden  $q$  değerinin olabildiğince büyük olması istenir. Bu nedenle, sadece bir sonuç almak için  $q\|\omega\| = 1$  yaptıktan sonra kenar payını mümkün olduğunca artırmak için  $\|\omega\|$  değerini mümkün olduğunca azaltmak gerekir [125]. Bunun için Denklem 3.48'de denklem 3.52'de gösterildiği gibi  $\|\omega\|$  en küçük değerini bulmak gerekir.

$$\min \frac{1}{2} \|\omega\|^2 \text{ kısıt: } y_i(\omega x_i + \omega_0) \geq +1, \forall i \quad (3.52)$$

Bu “ $m$ ” değerine bağlı ikinci dereceden bir optimizasyon problemidir ve çözüldüğünde  $\omega$  ve  $\omega_0$  için en iyi performansı veren değerler elde edilir. Yani üstündüzlemin en yakınındaki her iki sınıfa ait veri ile sınır arasında en az  $\frac{1}{\|\omega\|}$  kadar ve iki kenar payı arasında en az  $\frac{2}{\|\omega\|}$  kadar uzaklık bulunmaktadır.

En iyi üstündüzleme ulaşılmaya çalışılırken optimizasyon tekniğinde  $m$  yerine giriş sayısı olan “ $N$ ” kullanılır. Böyle bir kullanım doğrusal boyutlara geçmek için kullanılan çekirdek işlevlerinde kolaylık sağlar.

Denklem 3.52'de belirtilen kısıtlamayı Denklem 3.53'te gösterildiği gibi Lagrange çarpanları eklenerek kısıtlama olmadan Denklem 3.54'teki gibi yazılabilir.

$$L_p = \frac{1}{2} \|\omega\|^2 - \sum_{i=1}^N \alpha_i [y_i(\omega x_i + \omega_0)] + \sum_{i=1}^N \alpha_i \quad (3.54)$$

Burada Denklem 3.54'ün  $\omega$  ile  $\omega_0$  değerlerine göre minimum değerde olmasını sağlayan ve  $\alpha_i \geq 0$  çarpanları için maksimum değerde olmasını sağlayan nokta optimizasyon probleminin nihai hedefidir. Bunun için Denklem 3.54'ün  $\omega$  ile  $\omega_0$  değişkenlerine göre türevi alınıp sıfıra eşitlenir.

$$\frac{\partial L_p}{\partial \omega} = 0 \implies \omega = \sum_i \alpha_i y_i x_i \quad (3.55)$$

$$\frac{\partial L_p}{\partial \omega_0} = 0 \implies \sum_i \alpha_i y_i = 0 \quad (3.56)$$

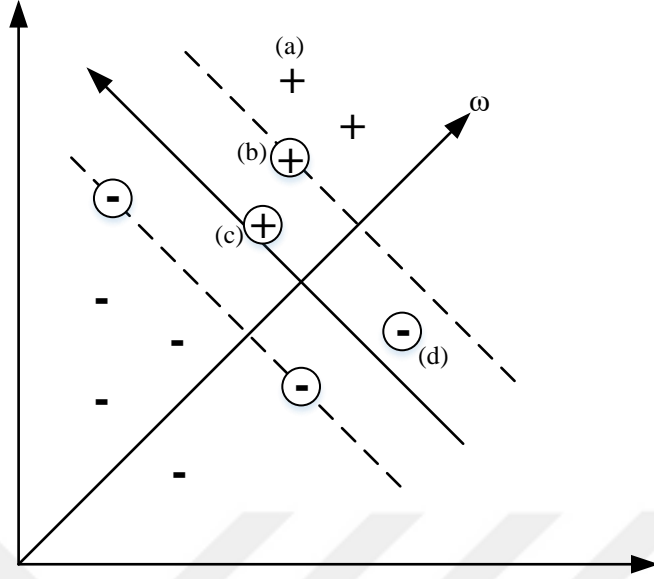
Denklem 3.55 ve 3.56, Denklem 3.54'te yerine koyularak Denklem 3.57 elde edilir.

$$L_d = -\frac{1}{2} \sum_i \sum_j \alpha_i \alpha_j y_i y_j x_i^T x_j + \sum_i \alpha_i \quad (3.57)$$

Denklem 3.57'i  $\alpha_i$  değişkenlerine göre büyütürken kısıtlar,  $\sum_i \alpha_i y_i = 0$  ve her  $i$  değeri için  $\alpha_i \geq 0$  olmasıdır. Bu problem ikinci dereceden optimizasyon problemi olup işlem karmaşıklığı eğitim örneği sayısı olan  $N$  değerine bağlıdır.

Optimizasyon problemi çözülüp  $\alpha_i$  değerleri hesaplandığında birçoğunun sonucunun sıfır olduğu görülür. Bu sonuçlar dışında  $\alpha_i > 0$  koşulunu az bir yüzde ile sağlayan  $x_i$  değerlerine destek vektörleri denir. Destek vektörleri, Şekil 3.28'de kenar payı üzerindeki artı ve eksi sınıfları yuvarlak içine alarak belirtilmiştir. Bu destek vektörleri  $y_i(\omega x_i + \omega_0) = 1$  koşulunu sağlar. Bu koşula göre herhangi bir destek vektör kullanılıp  $\omega_0$  değeri hesaplanabilir. Daha doğru bir tahmin için bu işlem tüm destek vektörlerde uygulanıp ortalaması alınır. Bu şekilde oluşturulan ayırtaca Destek Vektör Makinesi (DVM) denir.  $\alpha_i = 0$  olduğu durumda tüm giriş örnekleri  $y_i(\omega x_i + \omega_0) > 1$  koşulunu sağlamaktadır. Bu giriş örnekleri ayırtaçtan yeterince uzaktır ve herhangi bir bilgi taşımazlar. Bu giriş örneklerinin herhangi bir alt kümesi, eğitim kümesine çıkarılsa veya eklense bile sonuç değişmez. Çünkü DVM, ayırtaç tabanlı olduğundan sadece sınıra yakın giriş örneklerini kullanır ve diğer giriş örnekleri ile ilgilenmez. Bu yüzden, daha az karmaşık olan bir sınıflandırıcı ile destek vektörleri bulunup eğitim kümesinden atılabilir böylece DVM karmaşıklığı indirgenebilir.

Bazı durumlarda da DVM algoritmasında sınıfları doğrusal olarak % 100 doğrulukla ayırmak mümkün olmaz. Bu yüzden en hata ile doğru konumlandırılacak bir üstündüzlem aranır. Böyle bir durumda destek vektörlerin kenar payından sapmalarını gösteren artık değişkenler ( $\delta_i \geq 0$ ) eklenir. Destek vektörlerindeki sapmalarda, ya giriş örneği sınırın yanlış tarafındadır ya da sınırın doğru tarafında fakat yeterince uzak değildir [109]. Bu iki durum Şekil 3.30'un (c) ve (d) olarak etiketlenen destek vektörlerinde olduğu gibidir.



Şekil 3.30. DVM'de doğrusal ayrılmama durumu

Artık değişkene göre Denklem 3.48 yeniden düzenlenip Denklem 3.58 oluşturulur.

$$y_i(\omega^T x_i + \omega_0) \geq 1 - \delta_i \quad (3.58)$$

Burada  $\delta_i = 0$  ise destek vektörleri ile ilgili bir sorun yoktur.  $0 < \delta_i < 1$  ise sınıflandırma doğru yapılmakla beraber destek vektörleri kenar payı ile sınır arasındadır.  $\delta_i \geq 1$  ise destek vektör yanlış taraftadır ve hatalı sınıflandırma yapılmıştır.  $\delta_i \geq 1$ , hatalı sınıflandırılan eğitim örneği sayısı ve  $\sum_i \delta_i$  doğrusal ayrılamayan eğitim örneği sayısını vermektedir. Doğrusal ayrılamayan eğitim örneği sayısını en aza indirmek için bir düzgünleştirme parametresi ( $C$ ), Lagrange optimizasyon problemine Denklem 3.59'daki gibi eklenir.

$$L_p = \frac{1}{2} \|\omega\|^2 + C \sum_i \delta_i \quad (3.59)$$

Düzgünleştirme parametresi ile yapılan hesaplamalarda sadece hatalı yapılan sınıflandırmalar değil aynı zamanda kenar payı ile sınır içerisinde kalan doğru sınıflandırma örnekleri de ele alınmaktadır. Denklem 3.54, düzgünleştirme parametresine göre yeniden düzenlenip Denklem 3.60 elde edilmiştir.

$$L_p = \frac{1}{2} \|\omega\|^2 + C \sum_i \delta_i - \sum_{i=1}^N \alpha_i [y_i(\omega x_i + \omega_0) - 1 + \delta_i] - \sum_{i=1}^N \sigma_i \delta_i \quad (3.60)$$

Burada  $\sigma_i$ ,  $\delta_i$  değerlerinin negatif olmamasını sağlayan güncel Lagrange parametrelerdir. Yine  $L_p$ 'nin  $\omega$  ile  $\omega_0$ 'a göre türevi alınıp sifıra eşitlenirse Denklem 3.55 ve 3.56'nın aynısı elde edilir.  $L_p$ 'nin  $\delta_i$ 'ye göre türevi alınıp sifıra eşitlenirse Denklem 3.61 elde edilir.

$$\frac{\partial L_p}{\partial \delta_i} = C - \alpha_i - \sigma_i = 0 \quad (3.61)$$

$\sigma_i \geq 1$  olduğunda,  $0 \leq \alpha_i \leq C$  koşulunu sağlaması gerekir. Denklem 3.55, 3.56 ve 3.61 Denklem 3.60'da yerine koyulursa Denklem 3.62 elde edilir.

$$L_d = -\frac{1}{2} \sum_i \sum_j \alpha_i \alpha_j y_i y_j x_i^T x_j + \sum_i \alpha_i \quad (3.62)$$

Denklem 3.62'i  $\alpha_i$  değişkenlerine göre büyütürken kısıtlar,  $\sum_i \alpha_i y_i = 0$  ve her  $i$  değeri için  $0 \leq \alpha_i \leq C$  olmasıdır. Denklem 3.62 çözüldüğünde doğru sınıflandırılan ve kenar payından yeterince uzak olan giriş örnekleri için  $0 < \alpha_i$  durumu söz konusudur. Bu giriş örnekleri içindeki destek vektörlerinde,  $\alpha_i \leq C$  koşulunu sağlayanlar tam kenar payının üstündedir ve  $\omega_0$  değerini hesaplamak için kullanılırlar. Ayrıca bu durumda  $\delta_i=0$  ve  $y_i(\omega x_i + \omega_0) = 1$  eşitlikleri de sağlanır.  $\omega_0$  değişkeninin değerini hesaplarken  $y_i(\omega x_i + \omega_0) = 1$  eşitliğindeki destek vektör girişlerinin bir çoğu ile birçok  $\omega_0$  değeri bulup ortalamasını almak daha doğrudur. Kenar payının içinde kalan veya hatalı sınıflandırılan eğitim örneklerinde ise  $\alpha_i = C$  durumu söz konusudur.

DVM'lerdeki optimizasyon beklenen hata oranına ulaşıncaya kadar devam eder. DVM'deki beklenen hata oranı, doğru tahmin edilen destek vektörlerinin sayısının eğitim örneği sayısına bölümünden elde edilen sonuçtan büyük olamaz. Eğer destek vektörleri yanlış sınıflandırılmışsa veya kenar payı içerisinde ise hatalıdır. Bu kayba "menteşe" kaybı denmektedir. Mentşe kaybı işlevinin değeri Denklem 3.63'de verilen koşullu duruma göre değişir.

$$L_h(y_i^t, y_i) = \begin{cases} 0 & \text{eğer } y_i^t y_i \geq 1 \\ 1 - y_i^t & \text{değilse} \end{cases} \quad (3.63)$$

Burada  $y_i^t$ , tahmin edilen çıkışlar,  $y_i$  ise gerçek çıkışlardır. Eğer sınıflandırılacak veriler doğrusal değilse doğrusal olmayan bir model oluşturmak yerine doğrusal olmayan uzaydan başka bir uzaya geçilip doğrusal bir model oluşturulması mümkündür. Yeni uzaydaki model önceki doğrusal olmayan modele denk gelir. Bu işleme çekirdekleme denir. Bu tarz bir modelleme bağlanım içinde kullanılabilir. Yeni uzaydaki boyutların taban işlevlerine göre Denklem 3.63'teki gibi belirtilmektedir.

$$s = \phi(x) \text{ öyle ki } s_m = \phi_j(x), j = 1, 2, \dots, m \quad (3.64)$$

Böylelikle  $d$  boyutuna sahip  $x$  uzayından,  $m$  boyutuna sahip  $s$  uzayına geçilmiş olunur. Yeni uzaydaki ayırtaç Denklem 3.65'te gösterildiği gibi olur.

$$g(x) = \omega \phi(x) = \sum_{j=1}^m \omega_j \phi_j(x) \quad (3.65)$$

Burada  $\omega_0$  değeri Denklem 3.65'te ayrı olarak belirtilmemiştir ve  $\omega_0 = \phi_1(x) = 1$  olarak belirlenmiştir.

Yeni uzayda tam anlamıyla doğrusal ayrılma durumu söz konusu olmadığından daha genel bir tanımla yapılan düzgünleştirme parametresi ile elde edilen denklemleri kullanmak daha doğrudur.

Denklem 3.59 yeni uzayda aynı kalır. Fakat yeni uzaydaki kısıt denklem 3.66'daki gibi bir değişir.

$$y_i \omega^T \phi(x_i) + \geq 1 - \delta_i \quad (3.66)$$

Denklem 3.66'ya Lagrange terimleri eklenirse Denklem 3.67 elde edilir.

$$L_p = \frac{1}{2} \|\omega\|^2 + C \sum_i \delta_i - \sum_{i=1}^N \alpha_i [y_i (\omega \phi(x_i)) - 1 + \delta_i] - \sum_{i=1}^N \sigma_i \delta_i \quad (3.67)$$

$\omega$  ve  $\delta_i$  değerlerine göre türev alınıp sıfıra eşitlenirse Denklem 3.68 ve 3.69 elde edilir ve elde edilen bu denklemler Denklem 3.67'de yerine koyulursa Denklem 3.70 elde edilir.

$$\frac{\partial L_p}{\partial \omega} = 0 \implies \omega = \sum_i \alpha_i y_i \phi(x_i) \quad (3.68)$$

$$\frac{\partial L_p}{\partial \delta_i} = C - \alpha_i - \sigma_i = 0 \quad (3.69)$$

$$L_d = -\frac{1}{2} \sum_i \sum_j \alpha_i \alpha_j y_i y_j \phi(x_i^T) \phi(x_j) + \sum_i \alpha_i \quad (3.70)$$

Denklem 3.70'i  $\alpha_i$  değişkenlerine göre büyütürken kısıtlar,  $\sum_i \alpha_i y_i = 0$  ve her  $i$  değeri için  $0 \leq \alpha_i \leq C$  olmasıdır.

Çekirdeklemdeki ana mantık, taban işlevleri olan  $\phi(x_i^T) \phi(x_j)$  terimlerini iç çarpım olarak eklemek yerine  $K(x_j, x_i)$  çekirdek işlevi olarak eklemektir. Buna göre Denklem 3.70, Denklem 3.71'deki gibi yazılır.

$$L_d = -\frac{1}{2} \sum_i \sum_j \alpha_i \alpha_j y_i y_j K(x_j, x_i) + \sum_i \alpha_i \quad (3.71)$$

Çekirdek işlevi ayrıtaç da ise Denklem 3.72'deki gibi yazılır.

$$g(x) = \sum_i \alpha_i y_i k(x, x_i) \quad (3.72)$$

Denklem 3.72'yi karar fonksiyonu haline getirmek için signum fonksiyonu eklenir ve Denklem 3.73 elde edilir.

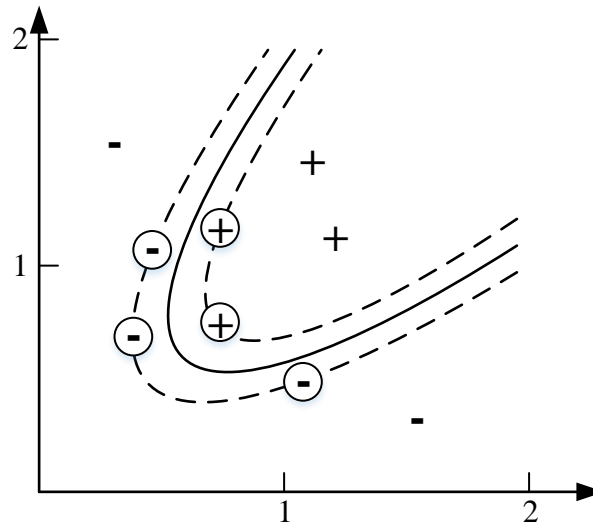
$$f(x) = \text{sign}(\sum_i \alpha_i y_i k(x, x_i) + b) \quad (3.73)$$

Çekirdek fonksiyonu olarak birçok fonksiyon kullanılabilir. En çok kullanılan çekirdek fonksiyonları Tablo 3.2’de verilmiştir.

**Tablo 3.2.** DVM algoritmasında yaygın kullanılan çekirdek fonksiyonları

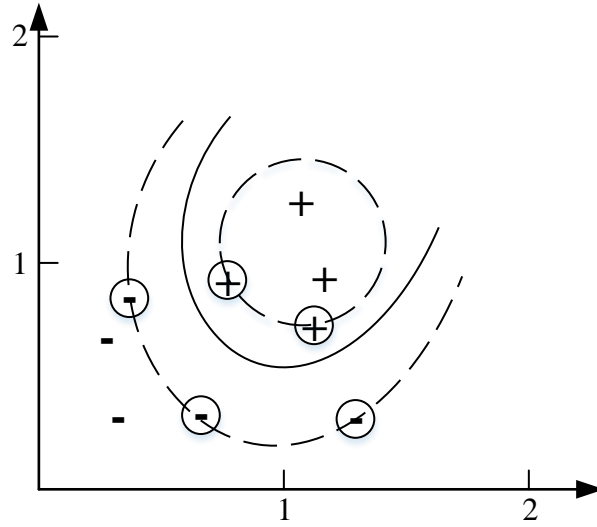
Çekirdek Fonksiyonları	Matematiksel Gösterimleri
Doğrusal	$K(x, x_i) = x^T x_i + c$
Çok Terimli	$K(x, x_i) = (x^T x_i + c)^n$
S biçimli → (Tanh)	$K(x, x_i) = \tanh(2x^T x_i + c)^n$
→ (Sigmoid)	$K(x, x_i) = \text{sigmoid}(2x^T x_i + c)^n$
Dairesel Tabanlı (Gauss) → Öklid uzaklığı	$K(x, x_i) = \exp\left[-\frac{\ x_i - x\ ^2}{2s^2}\right]$
Laplace Tabanlı	$K(x, x_i) = \exp\left[-\frac{\ x_i - x\ }{\sigma}\right]$

Çok terimli çekirdek fonksiyonu için  $c = 1$  ve  $n = 2$  olduğu durumda giriş boyutu iki seçilirse çekirdek fonksiyonu,  $K(x, y) = (x^T y + 1)^2 = (x_1 y_1 + x_2 y_2 + 1)^2$  olarak elde edilir. Bu çekirdek fonksiyonluyla iki sınıflı doğrusal olmayan bir problemde ayırtaç ve destek vektörlerin durumu için örnek gösterim Şekil 3.31’de verilmiştir.



**Şekil 3.31.** Çok terimli çekirdek fonksiyonu ile ayırtaç ve destek vektörlerin durumu

Gauss çekirdek fonksiyonu ile doğrusal olmayan iki sınıflı bir problemde ayırtaç ve destek vektörlerin durumu için örnek gösterim Şekil 3.32’de verilmiştir.



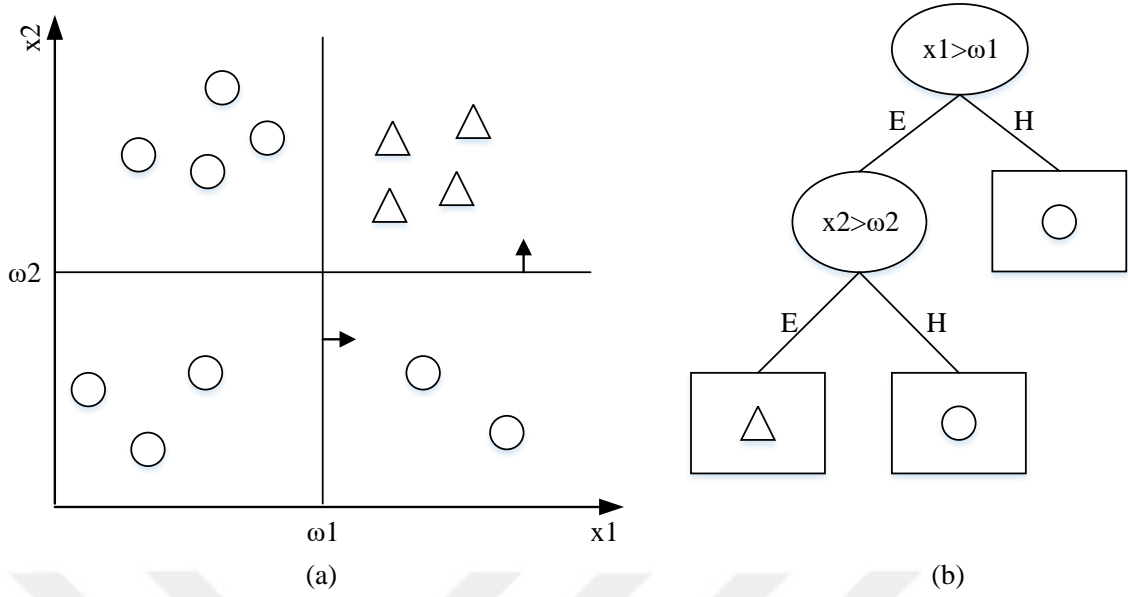
**Şekil 3.32.** Gauss çekirdek fonksiyonu ile ayırtaç ve destek vektörlerin durumu

Gauss çekirdek fonksiyonunda kullanılan “ $s$ ” yarıçapı sınırlarında yumuşatılmasında etkilidir ve yarıçap büyüdükçe sınırlar yumuşar.

İkiden fazla DVM algoritmasıyla sınıflandırma problemleri için “bire karşı bir” ve “bire karşı hepsi” sınıflandırma metotları kullanılır. Bu metotların ortak yanı ikili sınıf varsayımları yaparak oluşturulmasıdır. Bire karşı bir metodunda her sınıf diğer taraftaki bir sınıfa karşı birer birer eğitilir. Bire karşı hepsi metodunda ayırt edilecek sınıf ikili sınıflandırma işleminin bir tarafını, geriye kalan sınıfların hepsi de bir sınıfmış gibi kabul edilerek sınıflandırma işleminin diğer tarafını oluşturur.

### 3.12.3. Karar Ağaçları

Böl ve yönet yaklaşımı ile hareket eden karar ağaçları algoritması ağaç yapısından esinlenmiştir. Karar ağaçları algoritması gözetimli öğrenme sınıfında olup dağılımdan bağımsız kestirim yapar. Diğer dağılımdan bağımsız yöntemlerin (en yakın komşu kestirimi algoritmaları) aksine uzaklığı kullanılmadığı için büyük eğitim örneklerindeki hesap karmaşıklığı oluşmaz. Her karar ağacı iç karar düğüm adı verilen düğümler ile uç yaprak denilen yapılardan oluşur yaprakların çıktılarında tahmin edilen sınıflar olur [109]. Basit bir iki sınıflı bir sınıflandırma probleminde karar ağaçlarının basitçe uygulanış gösterimi Şekil 3.33’te verilmiştir.



Şekil 3.33. İki sınıflı karar ağacı örnek gösterimi

Şekil 3.33'ün (a) kısmında iki sınıflı giriş örneklerinin  $x_1$  ve  $x_2$  değişkenlerine göre dağılımı ve karar düğümünde uygulanacak olan şartın kesim bölgeleri verilmiştir. Bu kesim bölgelerine göre Şekil 3.33'ün (b) kısmında gösterilen karar ağacı akış diyagramı oluşturulmuştur. Bu akış diyagramında elipsler karar düğümlerini, dikdörtgenler ise yaprakları gösterir. Verilen akış diyagramına göre giriş örnekleri  $x_1$  ekseninde  $\omega_1$  kesim bölgesinden küçükse bütün sınıflar daire olduğu için işlem bitirilir. Fakat giriş örnekleri  $x_1$  ekseninde  $\omega_1$  kesim bölgesinden büyükse yeni bir karar düğümüne geçilir ve bu düğümde  $\omega_2$  kesim bölgesinde büyük-küçük olma durumuna bakılır. Bu düğümde hem  $x_1$  hem de  $x_2$  kesim bölgelerinin durumlarına bakılmış olunur. Bu düğümde giriş örnekleri  $\omega_2$  kesim bölgesinden küçükse daire, büyükse üçgen olur ve algoritma sonlandırılır.

Genel anlamda karar ağacı algoritmalarının çalışma mantığı şöyledir:

**Adım 1:** N eğitim kümesi oluştur.

**Adım 2:** N eğitim kümesinde en iyi ayrılmayı sağlayan niteliği bilgi kazancına göre belirle.

**Adım 3:** Belirlen nitelik ile ağacın ana karar düğümünü oluştur.

**Adım 4:** Ana karar düğümünden başlayarak alt karar düğümleri veya yapraklar oluştur.

**Adım 5:** Alt karar düğümlere ait alt veri kümeleri oluştur.

**Adım 6:** 5. adımda oluşturulan her alt veri kümesinde giriş örnekleri için;

- Bütün giriş örnekleri aynı sınıfta ise
- Giriş örneklerini ayıracak nitelik yoksa
- Kalan giriş örneklerinin niteliğini temsil eden örnek bulunmuyorsa işlemi sonlandır. Aksi takdirde Adım 4'e geri dön ve yeni veri kümeleri ile karar düğümleri ve yapraklar oluştur.

Ana karar düğümlerinden başlayarak alt karar düğümlerinin oluşturulmasındaki en önemli faktör bilgi kazancıdır. Bilgi kazancının hesaplanması içinse düzensizlik işlevine başvurulur. İki sınıflı bir problem için düzensizlik Denklem 3.74'te gösterildiği gibi hesaplanır.

$$D = -p \log_2 p - (1 - p) \log_2 (1 - p) \quad (3.74)$$

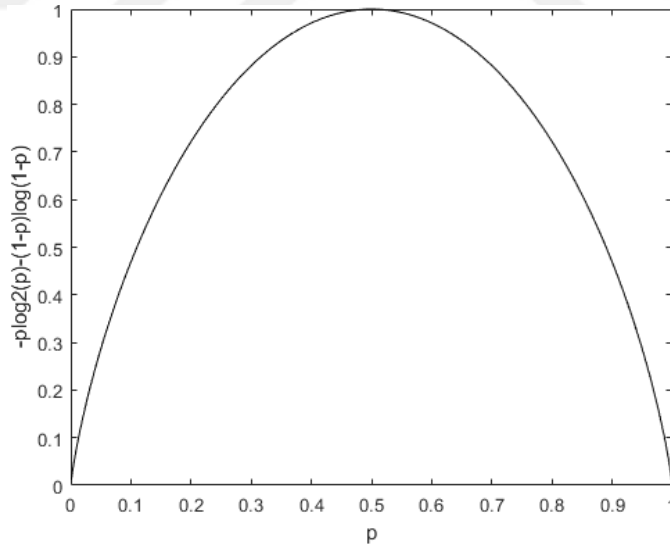
Burada  $p$ , veri kümesinin belirlenen karar düğümünde olma olasılığıdır. Denklem 3.74'ün  $p$  olasılığına göre değişimi Şekil 3.34'te verilmiştir. Sınıf sayısına bağımlı kalmadan düzensizlik işlevi Denklem 3.75'te gösterildiği gibidir.

$$D_n = -\sum_{i=1}^K p_n^i \log_2 p_n^i \quad (3.75)$$

Bu düzensizlik işleviyle  $M$  veri kümesindeki niteliklerin bilgi kazancı Denklem 3.76'ya göre hesaplanır ve hangi niteliğin bilgi kazancı büyükse ana karar düğümü için aranan nitelik o olur.

$$\text{Kazanç}(M, A) = D(M) - \sum P_r D(M_r) \quad (3.76)$$

Burada  $D(M)$ , toplam veri kümesinde her bir giriş örneğine karşılık gelen sınıflar için toplam düzensizlik değeridir.  $P_r$ , aranan niteliğin alt sınıflarındaki örneklerin tüm veri kümesine göre olasılığıdır ve  $D(M_r)$  ise alt sınıflardaki örnekler için düzensizlik değeridir.



Şekil 3.34. İki sınıflı düzensizlik işlevinin olasılık durumuna göre değişimi

### 3.13. Sınıflandırmada Normalleştirme

Normalleştirmenin sağladığı en önemli iki fayda, sınıflandırılacak verinin tutarlılığını artırması ve verilerin tekrar etme durumlarının ortadan kaldırılmasıdır. Normalleştirme etkin bir şekilde uygulandığında verilerin işleme hızı artar ve veri depolama boyutu azalır. Makine

öğrenme algoritmalarında öznitelik çıkarımı yapılırken veriler arasında çok büyük değerler olabilir. Böyle bir durum hem ağın öğrenme hızını azaltır hem de başarımını etkiler [126].

Uygulanan deneysel çalışmalarda öznitelik veri kümesinin normalleştirilmesinde adımları “z-skor” ve “min-max” yöntemleri kullanılmıştır.

### 3.13.1. Z-skor Normalleştirme Yöntemi

Bütün girişlerin belirli bir aralıkta düzenlenmesi için gerçekleştirilen z-skor normalleştirme yöntemi, giriş verilerinin ortalamasını ve standart sapma değerlerini kullanır. Z-skor normalleştirilmesi Denklem 3.77 kullanılarak elde edilir.

$$Z = \frac{x-\rho}{\sigma} \quad (3.77)$$

Burada  $x$  giriş değerlerini,  $\rho$  giriş değerlerinin ortalamasını,  $\sigma$  ise sütun değerleri arasındaki standart sapmayı göstermektedir. Z-skor normalleştirme yöntemi genel olarak sınıflandırma performansını %1 -%10 arasında artırabilmektedir [127].

### 3.13.2. Min-Max Normalleştirme Yöntemi

Öznitelik veri kümesindeki değerlerin doğrusal olarak normalleştirilmesi isteniyorsa min-max normalleştirme yöntemi kullanılır [127]. Bu yöntemle öznitelik veri kümesindeki bütün değerler 0-1 arasına doğrusal bir biçimde yerleştirilir. Min-max normalleştirilmesi Denklem 3.78 ile gerçekleştirilir.

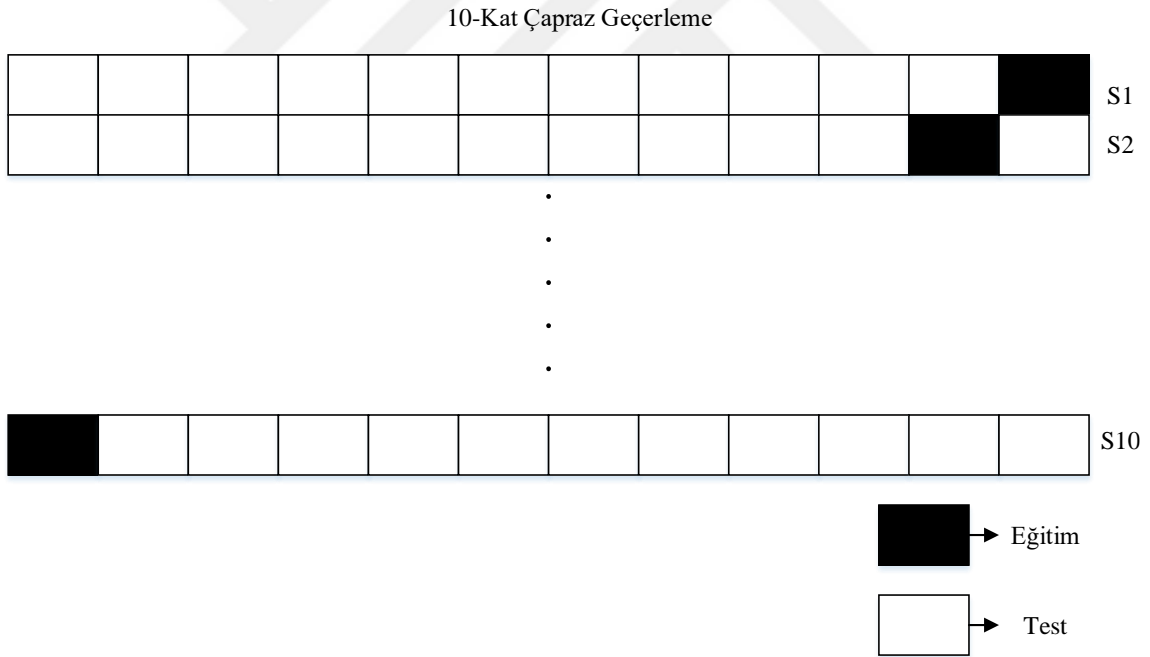
$$x_n = \frac{x-x_{min}}{x_{max}-x_{min}} \quad (3.78)$$

## 3.14. Sınıflandırmada Eğitim ve Test Verileri

Makine öğrenmesinde geliştirilen algoritmaların sınanması için eğitim ve test verisi olarak ayrılması gerekmektedir. Eğitim verisi ile en iyi ayrımı yapacak öğrenilebilir parametreler hesaplanmaktadır. Test verisi ile de öğrenilen bu parametrelerle sınıflandırma doğruluğu ölçülmektedir.

Eğitim ve test verilerinin nasıl ayrılacağı ile ilgili birkaç metot bulunmaktadır. Bunlardan birisi eğitim ve test verilerini belli bir oranda ayırmaktır. Örneğin %80 eğitim ve %20 test olarak ayırım yapılabilir. Bu ayırma işleminin rastgele yapılması genel geçerliliği artırır. Ayrıca bu ayırımın rastgele birkaç kez tekrarlanması da geliştirilen algoritmanın performansı hakkında daha doğru bilgi verir. Diğer bir yöntemde ise veriler veri setini oluşturan organizasyon tarafından ayrılır. Burada ayrılan veriler bir plan dahilinde seçildiği için ayırımın genel geçerliliği söz konusudur. Örneğin DCASE 2017 yarışması için hazırlanan veri setinde geliştirilen algoritmalar eğitim ve test

verisi ayrı ayrı belirtilmiştir. Başka bir diğer yöntemde oluşturulan algoritmaların performansı K-Kat çapraz geçерleme yöntemi ile ölçülmektedir. Bu yöntemde veri kümesi K tane rasgele eşit parçaya bölünür. K tane veri kümesinden bir tanesi her katta test işlemleri için ayrılır. Daha sonra geriye kalan K-1 tane veri kümesi birleştirilerek o katın eğitim veri seti oluşturulmuş olur. Her katta alınan sınıflandırma doğruluğunun ortalaması alınarak başarımları hesaplanır. K-Kat çapraz geçерlemede K sayısı önemlidir. K çok büyük seçilirse eğitim kümesinin yüzdesi artar ve buna karşın test kümesinin yüzdesi çok küçülür. Bu yüzden sağlıklı bir test işlemi gerçekleştirilmiş olmaz. Ayrıca öğrenilecek parametre sayısı eğitim kümesinin sayısı ile doğru orantılı olduğundan öğrenme süresi uzar. K olması gerekenden çok küçük seçilirse bu seferde sınıflandırma performansı düşer. K-Kat çapraz geçерlemede kat sayısı genel olarak 10 alınır. Son yıllarda performansı yüksek bilgisayarların çıkmasıyla K-Kat çapraz geçерleme birden çok kez yapılabilir duruma gelmiştir [109]. Örneğin 10-Kat çapraz geçерlemenin 10 defa yapılması daha doğru bir hata kestirimi yapılmasını sağlayacaktır. 10-Kat çapraz geçерlemenin örnek gösterimi Şekil 3.35'te verilmiştir.



**Şekil 3.35.** 10-Kat çapraz geçerleme örnek gösterimi

Şekil 3.35'te 10 eşit parçaya bölünen veri setinin her katından S1'den S10'a kadar 10 tane test verisi sınıflandırma sonucunun ortalaması alınarak performans ölçümü yapılır.

### 3.15. Sınıflandırmadan Sonraki Performans Ölçüm Kriterleri

Sınıflandırma sonrasında kullanılan yöntemlerin performans ölçümleri için karmaşıklık matrisi temel alınmıştır. Karmaşıklık matrisinin temsili gösterimi Şekil 3.36'da verilmiştir.

	Tahmine Edilen Sınıflar	
Gerçek Sınıflar	Doğru Pozitif (DP)	Yanlış Negatif (YN)
	Yanlış Pozitif (YP)	Doğru Negatif (DN)

Şekil 3.36. Karmaşıklık matrisi

Karmaşıklık matrisinde DP doğru sınıflandırılan değerleri, YP başka sınıfta olması gerekirken ilgili sınıfta olan değerleri, YN ilgili sınıfta olması gerekirken başka sınıfta olan değerleri, DN doğru sınıflandırılan diğer sınıftaki değerleri göstermektedir. Bu değerlere göre sınıflandırma için en sık kullanılan performans ölçüm kriterleri doğruluk (en çok kullanılan), kesinlik, duyarlılık, özgüllük ve F-skor değerleridir. Karmaşıklık matrisindeki değerlere göre bu ölçüm kriterlerinin hesabı Denklem 3.79-3.83'e göre yapılmaktadır.

$$\text{Doğruluk}(D) = \frac{DP+DN}{DP+DN+YN+YP} \quad (3.79)$$

$$\text{Kesinlik}(K) = \frac{DP}{DP+YP} \quad (3.80)$$

$$\text{Duyarlılık}(Dy) = \frac{DP}{DP+YN} \quad (3.81)$$

$$\text{Özgüllük}(\ddot{O}) = \frac{DN}{DN+YP} \quad (3.82)$$

$$F - \text{skor} = 2 \times \frac{\text{Kesinlik} \times \text{Duyarlılık}}{\text{Kesinlik} + \text{Duyarlılık}} \quad (3.83)$$

## 4. MATERYAL VE METOT

Bu bölümde, dört farklı çalışma için uygulanan yaklaşımların nasıl uygulanacağı ayrıntılı bir şekilde verilmiştir.

### 4.1. Kalp Seslerinin Derin Öznitelikler ile Sınıflandırılması

Bu uygulamada kalp seslerinin sınıflandırılması konusu üzerine değinilmiştir. Derin öznitelikleri çıkarmak için popüler ESA modelleri kullanılmıştır.

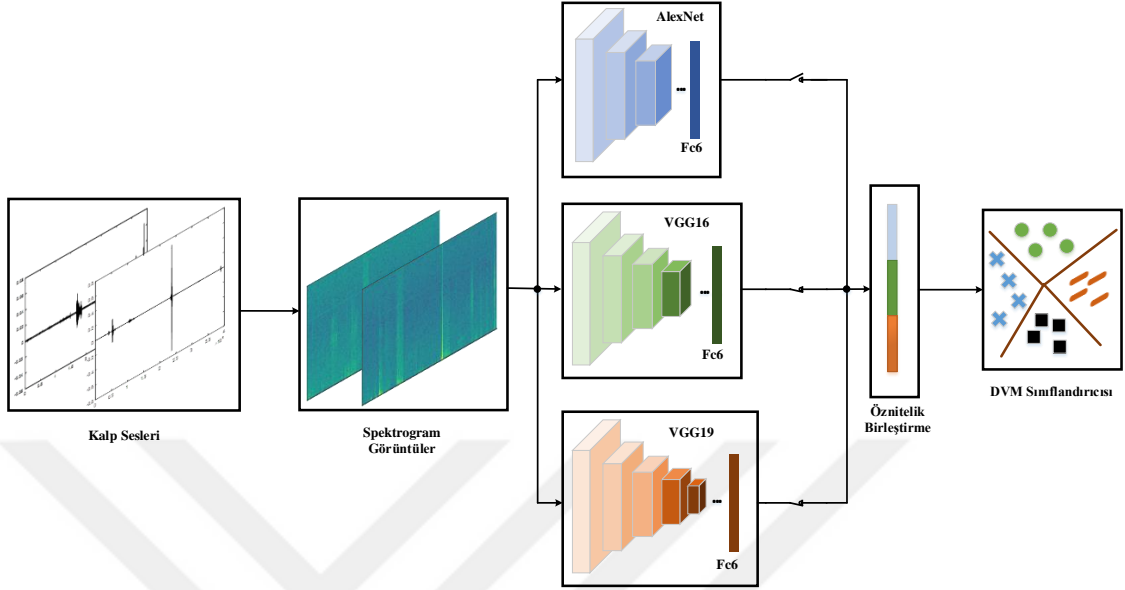
#### 4.1.1. Giriş

Kalp sesleri, kalp hastalıklarının erken teşhisine yardımcı olan akustik karakteristiklere sahiptir. Kalp hastalıklarının tespiti için kalp seslerinde çeşitli sinyal işleme tekniklerinin kullanıldığı çok sayıda yöntem önerilmiştir. Bu bölümde, kalp seslerini temel alan kalp hastalığı tespiti için son yıllarda popüler olan derin öğrenme tabanlı bir metodoloji kullanılmıştır. Önerilen yöntem, spektrogram resimler oluşturma, derin öznitelik çıkarımı ve sınıflandırma gibi ardışık üç aşamadan oluşmaktadır. Spektrogram oluşturma aşamasında, kalp sesleri zaman-frekans dönüşümü kullanılarak spektrogram görüntüleri oluşturulmuştur. Bu spektrogram görüntüleri kullanılan ESA modellerinin girişini beslemiştir. Derin öznitelikler AlexNet, VGG16 ve VGG19 ESA modellerinin TBK'larından elde edilmiştir. Son aşama olan sınıflandırma aşamasında ise güçlü bir sınıflandırıcı olan DVM sınıflandırıcısı kullanılmıştır. Önerilen yöntem, kullanıma açık olan ve bir kalp sesi sınıflandırması için oluşturulan iki ayrı veri seti ile değerlendirilmiştir. Önerilen yöntem, literatürde aynı veri setlerini kullanan diğer yöntemlerle karşılaştırılmıştır. Karşılaştırmalar önerilen yöntemin daha iyi performans verdiğini göstermiştir.

#### 4.1.2. Önerilen Yöntem

Önerilen yöntemin işleyişinin açıklayıcı bir gösterim Şekil 4.1'de gösterilmiştir. Şekil 4.1'e göre kaydedilen kalp sesleri belirli bir örnekleme frekansı ile ayrık verilerden oluşan tek boyutlu vektör uzayına dönüştürülmüştür. Bu vektör uzaylarının spektrogramları alınarak iki boyutlu bir matris yapısı elde edilmiştir. Spektrogramların görüntülerini elde etmek için *Matlab ColorMap'in Viridis* fonksiyonu kullanılmıştır. Öznitelik çıkarımı aşamasında önceden eğitilmiş VGG16, VGG19 ve AlexNet modelleri kullanılmıştır. Bu modellerin ağırlıkları, 1000 sınıflı bir görüntü sınıflandırma problemini çözmek için kullanılan 1 milyondan fazla resim üzerinde öğrenilmiştir. Öznitelikler bu modellerin TBK katmanlarından çıkış alınarak oluşturulmuştur. Daha sonra her bir modelden elde edilen öznitelikler, seti bir vektör uzayını temsil edecek şekilde tekrardan birleştirilmiştir. Sınıflandırma performansını artırmak için bu modellerin üçünden çıkarılan

özniteliklerin birleşim kombinasyonları kullanılmıştır. En son olarak sınıflandırma işlemi için doğrusal çekirdek kullanan DVM sınıflandırıcısı tercih edilmiştir.



Şekil 4.1. Kalp Seslerinin Derin Öznitelikler ile Sınıflandırılması

## 4.2. Derin Öznitelikler ve Derin Öğrenme Tabanlı Yaklaşımların EMG ve Akciğer Sinyalleri Üzerinden Denenmesi

Bu bölümde, Bölüm 4.1’de yer verilen derin öznitelikler tabanlı yaklaşımla beraber transfer öğrenme tabanlı yaklaşıma yer verilmiştir. Aynı metotlar farklı türdeki veriler üzerinde denenmiştir. Böylelikle hangi türdeki yaklaşımın sınıflandırma problemi üzerinde daha iyi sonuç verdiği incelenmiştir.

### 4.2.1. Giriş

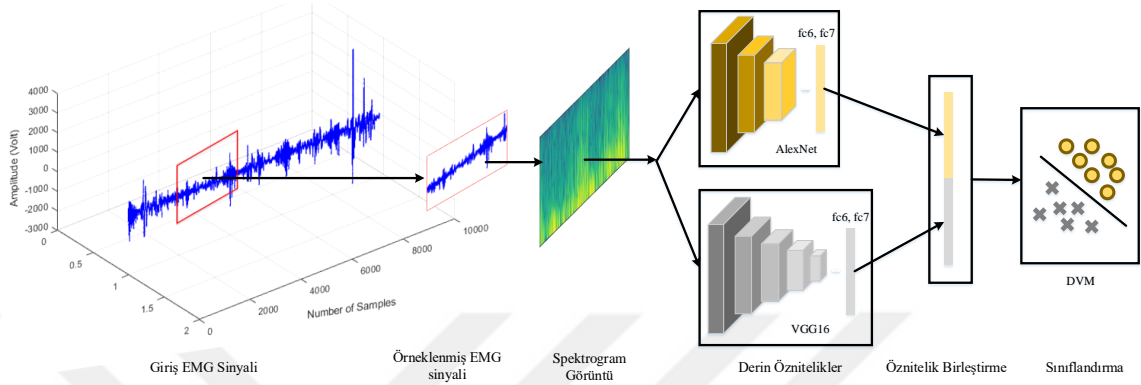
Transfer öğrenmede, öğrenme işlemine bu bilinen ESA modellerinin (AlexNet, VGG16, ResNet) ağırlıkları ile başlanır ve eğitim süresince parametreler değiştirilir.

Önerilen yöntemde her iki yaklaşımın performansları aynı veri setleri üzerinde karşılaştırılmıştır. Ayrıca yaklaşımların genel geçerliliğinin arttırmak için önerilen yöntem, hem EMG sinyallerinden oluşan veri setinde hem de akciğer seslerinden oluşan veri setinde denenmiştir.

### 4.2.2. Önerilen Yöntem

EMG veri seti için kullanılan her iki yöntemin temsili gösterimi Şekil 4.2 ve Şekil 4.3’te verilmiştir. Bu yöntemde kullanılan orijinal EMG sinyallerinin vektör uzayındaki boyutu 10000’dir ve bu EMG sinyallerinden rastgele örnekler alınarak her EMG sinyalinin vektör uzayındaki boyutu 2000 olacak şekilde yeniden boyutlandırılmıştır. Bu EMG sinyallerinden

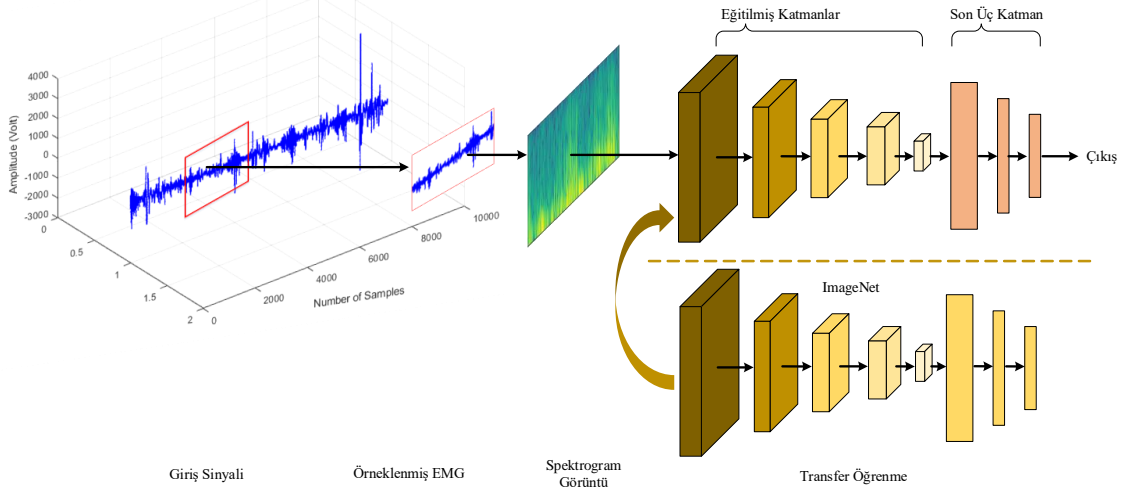
spektrogram görüntüler elde etmek için Bölüm 4.1'deki ön işleme safhalarının aynısı bu bölümde yeniden kullanılmıştır. EMG sinyallerinin spektrogramları alınırken çerçeve boyutu 64 ve çakışma boyutu 8 olarak seçilmiştir. Ayrıca AFD uzunluğu 512 olarak belirlenmiştir.



**Şekil 4.2.** Derin Öznitelikler ile EMG sinyallerinin sınıflandırılması

Şekil 4.2'de gösterildiği gibi derin öznitelikler çıkarılırken VGG 16 ve AlexNet modelleri kullanılmıştır. Bu modellerin fc6 ve fc7 TBK'larından çıkışlar alındıktan sonra öznitelik setinde değişik kombinasyonlarda birleştirilmek üzere kullanılmıştır. Sınıflandırma işleminde ise güçlü bir sınıflandırıcı olan doğrusal DVM algoritması tercih edilmiştir.

Şekil 4.3'te ise transfer öğrenme tabanlı model ile önerilen yöntemin temsili gösterimi verilmiştir. Bu yöntemde transfer öğrenme için AlexNet ESA modeli kullanılmıştır. Önceden eğitilmiş AlexNet ESA modelinin son 3 katmanı çıkarılmıştır. Daha sonra modele yeni TBK, sınıflandırma ve çıkış katmanları eklenerek eğitim ve test işlemi gerçekleştirilmiştir. Burada Şekil 4.3'te belirtilen eğitilmiş katmanlarda AlexNet modelindeki öğrenme parametreleri kullanılmıştır. Transfer öğrenmeye bu parametreler ile başlanmış ve EMG veri setinde en iyi sonucu alacak şekilde optimizasyon gerçekleştirilmiştir.



**Şekil 4.3.** Transfer Öğrenme ile EMG sinyallerinin ve akciğer seslerinin sınıflandırılması

EMG veri seti için yapılan çalışma da derin öznitelikler yaklaşımı transfer öğrenme yaklaşımına göre daha iyi doğruluk sonucu vermiştir. Bu sonucun genel geçerliliğini artırmak için akciğer seslerinden oluşan bir veri seti üzerinden her iki yaklaşımın performansı tekrar denlenmiştir. Ayrıca VGG16 ve AlexNet modeline ek olarak ResNet-50 modeline de her iki yaklaşımda yer verilmiştir.

### 4.3. Çevresel Seslerin Derin Öznitelikler ve Öznitelik Seçme Yaklaşımları ile Sınıflandırılması

Bu bölümde çevresel sesler için derin öznitelik çıkarmak için ESA modellerinin kombinasyonları ile birlikte sınıflandırma performansını artırmak için öznitelik seçme algoritması kullanılmıştır.

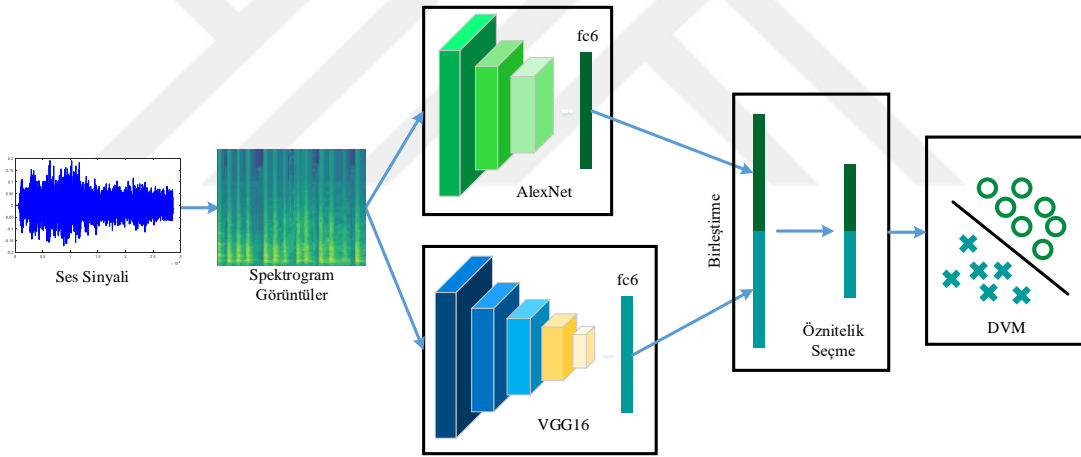
#### 4.3.1. Giriş

Bu bölümde, daha önceki bölümlerde uygulanan derin öznitelikler tabanlı yaklaşım çevresel sesler için uygulanmaktadır. Çevresel sesleri kısaca konuşma sesi olmayan ve çevresel faktörlerden kaynaklanan sesler olarak tanımlamak mümkündür.

Önerilen yöntemde derin öznitelikler tabanlı yaklaşım, Bölüm 4.2’de hem EMG sinyalleri hem de akciğer sesleri için transfer öğrenme tabanlı yaklaşımdan daha iyi performans verdiği için tercih edilmiştir. Önerilen yöntemde ek olarak hem sınıflandırma performansını daha da artırmak hem de test etme süresini azaltmak için minimum fazlalık maksimum uygunluk (mFMU) öznitelik seçme algoritması kullanılmıştır. Önerilen yöntemin genel geçerliliğini artırmak için literatürde en çok kullanılan dört ayrı veri seti tercih edilmiştir ve önerilen yöntem bu veri setlerini kullanan diğer yazarların yöntemleri ile karşılaştırılmıştır.

### 4.3.2. Önerilen Yöntem

Önerilen yöntemin, uygulanış biçiminin temsili gösterimi Şekil 4.4'te verilmiştir. Önerilen yöntemde ilk adım olarak çevresel seslerin spektrogram görüntüleri oluşturulmuştur ve bu görüntüler önceden eğitilmiş ESA modellerinin giriş boyutlarına uygun olacak şekilde yeniden boyutlandırılmıştır. Kullanılan ESA modelleri VGG16 ve AlexNet olmakla beraber derin öznetelikler bu modellerin fc6 TBK'sından elde edilmiştir. Bu öznetelik setinin boyutunu azaltmak için mFMU algoritması kullanılmıştır. Öznetelik seçimi, uğraşılan görevle ilgili olmayan geniş bir öznetelik kümesi içinden bir özellik alt kümesinin ortadan kaldırılması olarak tanımlanabilir. mFMU öznetelik seçim algoritmasında sınıf etiketleri ile yüksek ilintiye sahip olan ve kendi aralarında düşük ilinti içeren bir alt öznetelik seti seçilmektedir. mFMU yaklaşımında, öznetelik uygunluk düzeyini hesaplamak için  $F$  istatistiği ve öznetelik fazlalığını hesaplamak için Pearson ilinti katsayısı kullanılmaktadır. Bu seçimler doğrultusunda boyutu azaltılan öznetelik setinin sınıflandırma işleminde ise doğrusal DVM algoritması kullanılmıştır.



Şekil 4.4. Çevresel seslerin derin öznetelikler ile sınıflandırılması

## 4.4. Uçtan Uca Yapılandırılmış Bir ESA Modelinden Derin Öznetelikler Çıkarılması ile Sınıflandırma

Bu bölümde popüler derin öznetelikler çıkarmak için popüler ESA modelleri kullanmak yerine yeni bir ESA modeli oluşturulmuştur. Bu modelle öznetelikler çıkarılmıştır.

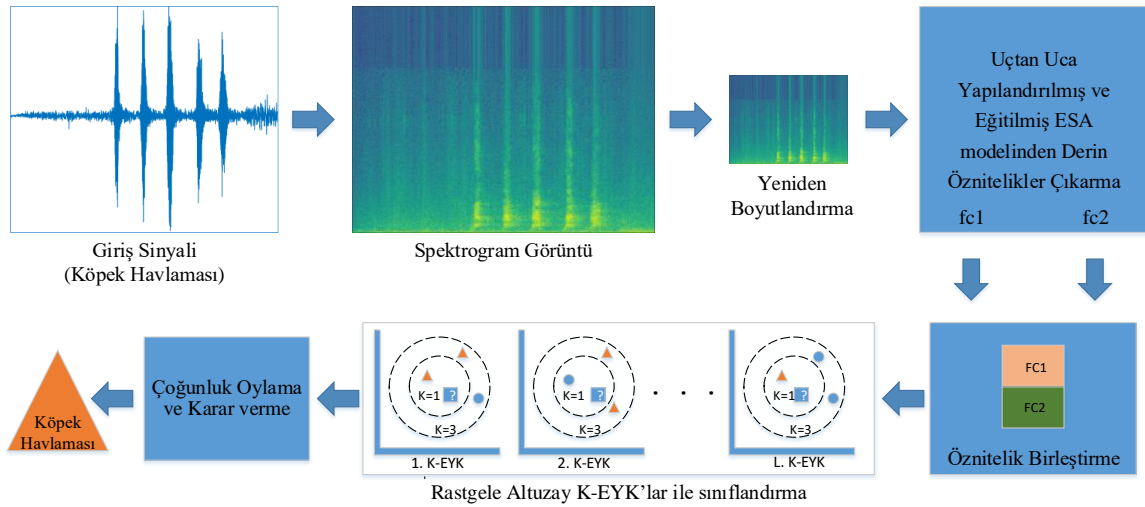
### 4.4.1. Giriş

Bölüm 4.1, 4.2 ve 4.3'te yer alan önerilen yöntemler ile derin öznetelikleri çıkarmak için VGG16 ve AlexNet gibi önceden eğitilmiş ve ağırlıkları paylaşılmış ESA modelleri kullanılmıştır.

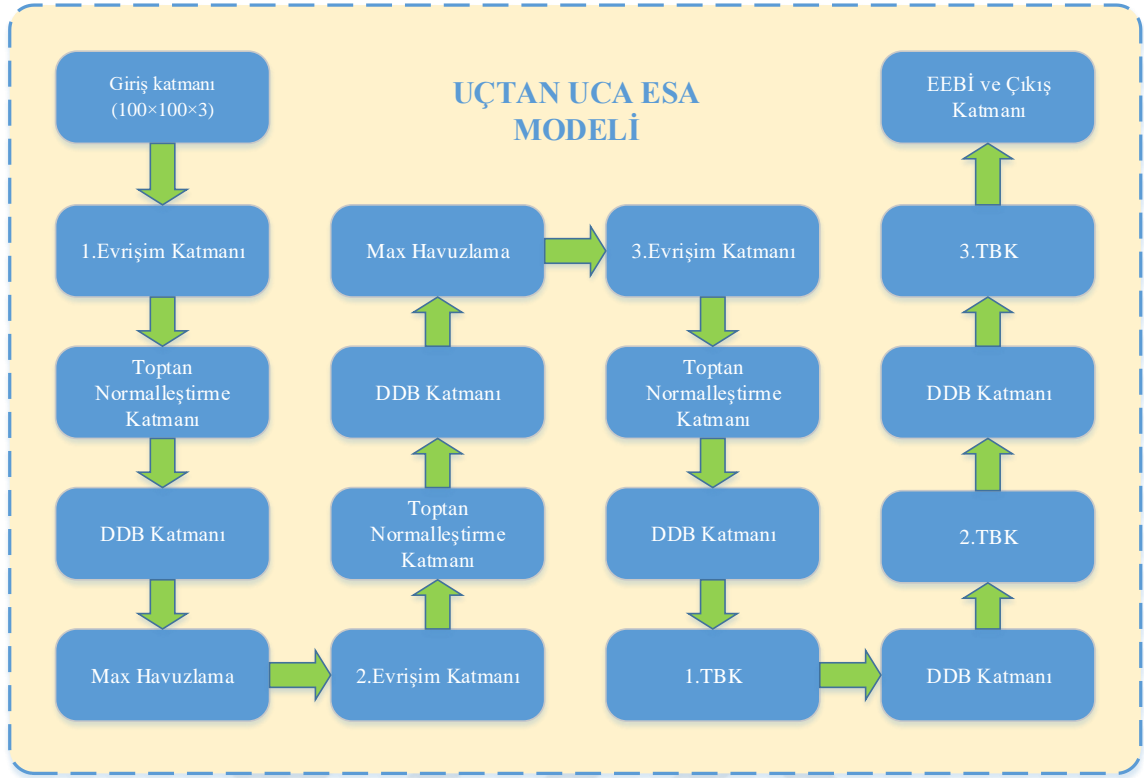
Bu modeller 1 milyondan fazla görüntü ile eğitildiğinden genel olarak herhangi bir veri setinde tatmin edici sonuçlar vermektedir. Ancak ses verilerinin sınıflandırılması için görüntü ile eğitilmiş ESA modelleri kullanmak bir tartışma konusudur. Bu bölümde, çevresel seslerden derin öznitelikler çıkarmak için VGG16 ve AlexNet gibi modeller kullanmak yerine, ses karakteristiklerini daha iyi temsil eden spektrogram görüntüleri ile eğitilen uçtan uca yapılandırılmış bir ESA modeli kullanılmıştır. Eğitim ve test işlemleri için sınıf başına örnek sayısı fazla olan DCASE-2017 ASC ve UrbanSound8K veri setleri kullanılmıştır. Kullanılan ESA modelinde bilinen ESA modellerine göre daha az giriş boyutu, katman sayısı ve TBK çıkış boyutu içeren bir yapı tercih edilmiştir. Böylelikle hem eğitim hem de test süresi önemli ölçüde azalmıştır. Bununla birlikte sınıflandırma performansı ölçmek için rastgele altuzay K-EYK topluluk (RAKT) sınıflandırıcısı kullanılmıştır. Önerilen bu yöntemle hem aynı veri setini kullanan diğer yöntemlere göre hem de bilinen ESA modellerinden çıkarılan derin özniteliklere göre daha iyi bir performans elde edilmiştir.

#### 4.5. Önerilen Yöntem

Şekil 4.5'te önerilen yöntemin temsili gösterimi verilmiştir. Diğer bölümlerde olduğu gibi ilk aşama olarak çevresel seslerin sinyallerinin spektrogram görüntüleri çıkarılmıştır. Bu spektrogram görüntüleri, ESA modelinin giriş katmanının boyutu ( $100 \times 100$ ) ile aynı olacak şekilde yeniden boyutlandırılmıştır. Bu spektrogram görüntülerinin her sınıf için ayrılmış bir kısmı ile uçtan uca yapılandırılmış olan ESA modeli eğitilmiştir. Bu ESA modelinin hiyerarşik yapısı Şekil 4.6'da gösterilmiştir. ESA modeli eğitildikten sonra giriş katmanından ilk ve ikinci TBK'lara (fc1 ve fc2) kadar olan yapı derin öznitelikler çıkarmak için kullanılmıştır. Bu ilk iki TBK'nın birleştirilmesi ile öznitelik seti oluşturulmuştur.

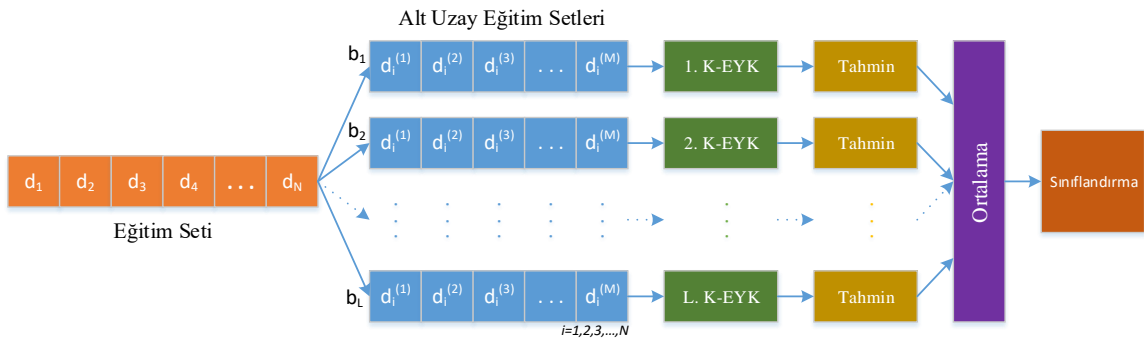


Şekil 4.5. Önerilen yöntemin temsili gösterimi



Şekil 4.6. Uçtan uca eğitilen ESA modelinin hiyerarşik yapısı

Oluşturulan derin öznitelik setini kullanarak sınıflandırma yapmak için RAKT algoritması kullanılmıştır. RAKT algoritmasında öznitelik seti, kullanıcı tarafından belirlenen sayıda ve boyutta alt uzaylara ayrılmaktadır. Her alt uzayda olması gereken öznitelik değerleri öznitelik setinden rastgele seçilmektedir. Daha küçük boyutlu olarak oluşturulan bu alt uzay öznitelik setlerinin her biri K-EYK algoritması ile eğitilir. Her alt uzayda yer alan K-EYK algoritmasının çıkışından alınan tahmin sonuçlarının ortalaması alınır ve sınıf etiketi hangi sınıfa daha uyguna o sınıfa atanır. RAKT algoritmasının temsili gösterimi Şekil 4.7’de verilmiştir.



Şekil 4.7. RAKT algoritmasının temsili gösterimi

## 5. BULGULAR

Bu bölümde, dört farklı çalışma için uygulanan yaklaşımların nasıl uygulandığı açıklanmıştır ve kullanılan veri setleri hakkında bilgi verilmiştir.

### 5.1. Kalp Seslerine Derin Öznitelikler Tabanlı Yaklaşımın Uygulanması

Bu uygulamada kalp seslerinin sınıflandırılması problemi için deneysel çalışmalara yer verilmiştir. Elde edilen sonuçlar ayrıntılı olarak irdelenmiştir.

#### 5.1.1. Veri Seti

Kalp sesleri ritimlerine göre genellikle iki farklı kategoride değerlendirilir. İlk kategoride yer alan sesler uzun süreli ve düşük aralıklı olan (S1), diğeri ise kısa süreli uzun aralıklı olan seslerdir. Kalp sesleri normal düzeni S1 ve S2'nin ardışık olarak meydana gelmesi ile sağlanmaktadır ve S1'den S2'ye kadar olan süre, S2'den S1'e kadar olan süreden daha fazladır. Kalp ritmindeki düzensizlikler S1-S2 hiyerarşisinin bozulmasından ya da S1-S2 arasındaki anormal seslerden kaynaklanmaktadır. Önerilen çalışma bu düzensizlikleri içeren iki farklı veri seti üzerinde gerçekleştirilmiştir [128].

İlk veri setinde (Veri Seti-A) kalp sesleri iStethoscope Pro iPhone adı verilen bir akıllı telefon uygulamasıyla elde edilmiştir. Bu uygulamada sesleri kaydetmek için akıllı telefonun mikrofon girişi kullanılmıştır. Akıllı telefon kalp bölgesine yakın bir yere yerleştirilmiştir ve uygulama içindeki yazılımlarla ses daha da kuvvetlendirilerek 44.1 KHz'lik örnekleme frekansıyla kaydedilmiştir. Kaydedilen sesler 1 tane normal ve 3 tane anormal olmak üzere 4 farklı kategoriye ayrılmıştır. Eğer kalp sesinden eğer hırıltılı ve çatalı bir ses alınıyorsa "hırıltılı" ses olarak, S1-S2 düzeni bozulup sürekli olarak S1-S1-S2 veya S2-S2-S1 oluyorsa "ekstra kalp sesi" veya kalp sesinden ortam gürültüsü alınıyorsa "gürültülü" olarak etiketlenmiştir.

Diğeri veri setinde (Veri Seti-B) kalp sesleri hastane ortamında dijital bir stetoskolla 4 KHz örnekleme frekansıyla kaydedilmiştir. Kaydedilen sesler 1 tane normal ve 2 tane anormal olmak üzere 3 farklı kategoriye ayrılmıştır. Yine bu veri setinde hırıltılı sınıf etiketi olup, diğeri sınıf etiketindeki seslerde (ekstra sistol), Veri Seti-A'daki ekstra kalp sesindeki düzensizliklerin sürekli olmayan durumları söz konusudur.

Her iki veri seti ile bilgiler Tablo 5.1'de verilmiştir. Bu veri setlerinden kaydedilen seslerin kayıt süreleri 1-30 saniye arasında değişmektedir.

**Tablo 5.1.** Veri setleri ile ilgili bilgiler

	Sınıf Etiketleri	Eğitim	Test	Dosya Kayıt Bilgileri
<b>Veri Seti-A</b>	Normal	31	14	44.1 KHz
	Hırıltılı	34	14	(1-30 sn)
	Extra Kalp Sesi	19	8	iStethoscope
	Gürültülü	40	16	Pro iPhone
<b>Veri Seti-B</b>	Normal	31	136	4KHz
	Hırıltılı	34	39	(1-30 sn)
	Ekstra Sistol	19	20	Digital Stetoskop

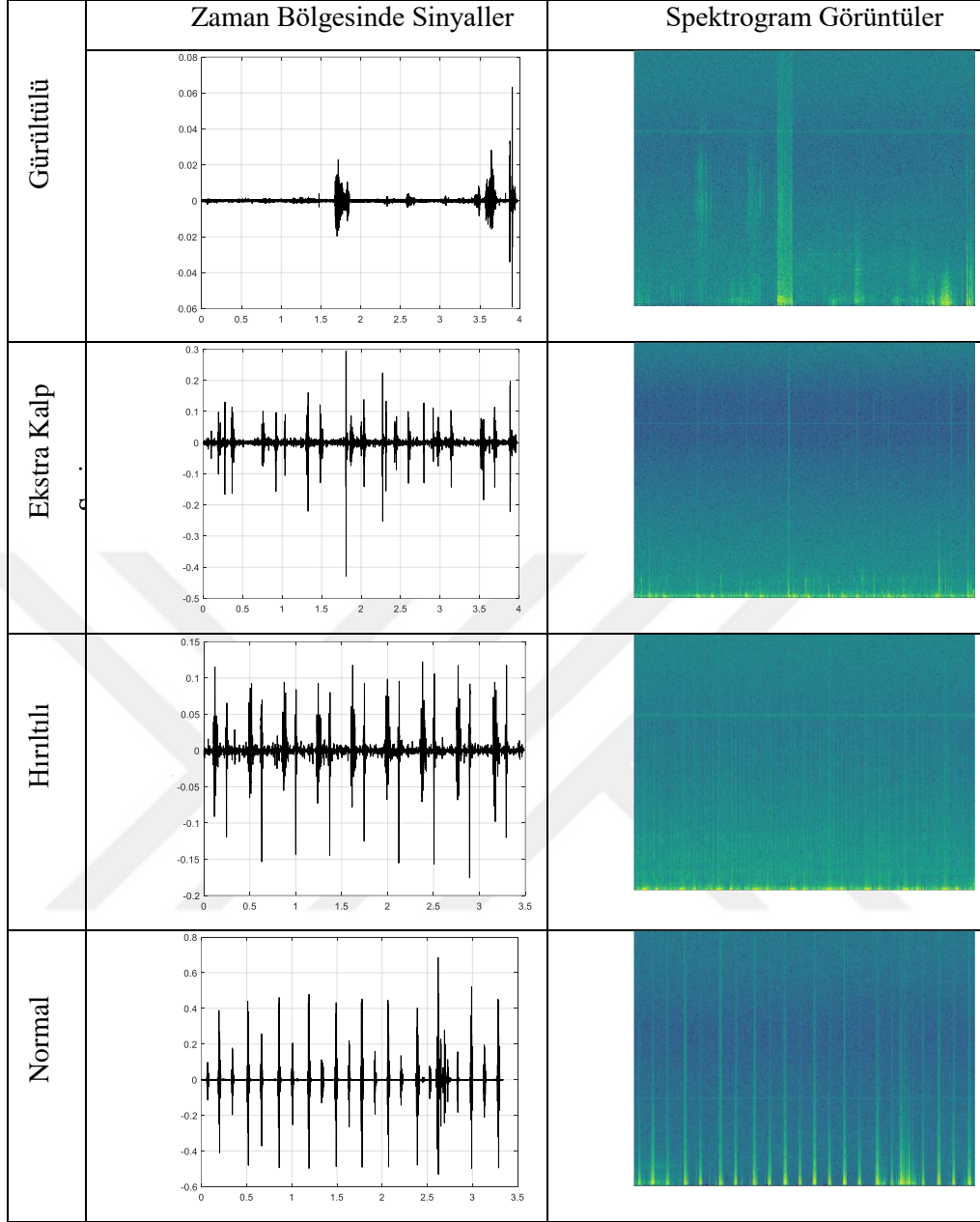
### 5.1.2. Deneysel Çalışmalar

Önerilen yöntem, Tablo 5.1’de verilen eğitim ve test ses dosyası sayılarına göre denenmiştir. Veri setinin paylaşıldığı kaynakta Bölüm 3.15’te verilen performans ölçütlerine ek olarak Youden indeksi (Yİ) ve ayırtaç gücü (AG) gibi kriterlerinde ölçülmesi istenmiştir. Bu performans ölçütlerinin elde edilmesi sırasıyla Denklem 4.1 ve 4.2’de verilmiştir.

$$Yİ = DY - (1 - ÖZ) \quad (4.1)$$

$$AG = \frac{\sqrt{3}}{\pi} \left( \log \frac{DY}{1-DY} + \log \frac{ÖZ}{1-ÖZ} \right) \quad (4.2)$$

Önerilen yöntemde ilk olarak ses sinyallerinin spektrogramları alınmıştır. Kullanılan her iki veri setinde ses sinyallerinin örnekleme frekansları farklı olduğu için tercih edilen pencere ve çakışma boyutları da farklı olmuştur. Veri seti-A için pencere boyutu 1024 çakışma boyutu 256 olarak seçilirken, Veri Seti-B için pencere boyutu 32 çakışma boyutu 8 olarak seçilmiştir. Her iki veri seti için Hamming pencereleme yöntemi kullanılmıştır. Spektrogram işleminden sonra Color Map Viridis fonksiyonu ile spektrogram görüntüleri oluşturulmuştur. Veri Seti-A için her sınıftan bir ses sinyali ve bu ses sinyallerinin spektrogram görüntüleri Şekil 4.2’de gösterilmiştir. Genel olarak yüksek örnekleme frekansında daha büyük pencere ve çakışma boyutları ile daha iyiler sonuçlar alındığı görülmüştür. İdeal pencere boyutu ve çakışma boyutu elde edilen görüntünün çözünürlüğüne göre seçilmiştir. Yüksek örnekleme frekansında küçük pencere boyutu seçmek çözünürlüğü düşük görüntülerin oluşmasına neden olmuştur. Pencere boyutunun çok büyük seçilmesi durumunda ise Fourier katsayılarından dolayı hesaplama maliyeti artmış ve dolayısıyla ön işleme süresi uzamıştır.



Şekil 5.1. Veri Seti-A içi farklı sınıflardaki kalp sesleri ve spektrogramları

Elde edilen spektrogram görüntüleri, önceden eğitilmiş VGG16 ve VGG19 ve AlexNet ESA modellerinin girişine uygun olacak şekilde yeniden boyutlandırılmıştır. VGG16 ve VGG19 ESA modeli için giriş boyutu  $224 \times 224$ , AlexNet ESA modeli için  $227 \times 227$  olacak şekilde boyutlandırılmıştır. Ağırlıkları belli olan bu modellerden derim öznelilikler elde etmek için “fc6” isimli ilk TBK’den çıkış alınmıştır. Bu katmanların çıkışında 4096 boyutlu vektör uzayı elde edilmiştir. Daha sonra bu üç ESA modelinin fc6 çıkışındaki vektör uzaylarının ayrı ayrı ve birleştirilmiş bütün kombinasyonları, düzgünleştirme parametresi  $10^{-4}$ -  $10^3$  arasında değişen DVM sınıflandırıcısı ile denenmiştir. Tablo 5.2’ye göre Veri Seti-A için en iyi sonuç sadece VGG16 ESA modeli için alınmıştır. Toplam 5 tane kesinlik sonucundan 4 tanesinde VGG16 ESA modeli en iyi

sonucu vermiştir. Tablo 5.3'e göre Veri Seti-B için en iyi sonuç VGG16 ve VGG19 ESA modellerinin derin özniteliklerinin birleştirilmesi ile elde edilmiştir. Bu iki modelin birleştirilmesiyle tüm kesinlik sonuçlarında en iyi performansa ulaşılmıştır.

**Tablo 5.2.** Veri seti-A için ESA modellerinin kombinasyonlarına göre kesinlik sonuçları

Veri Seti-A							
Değerlendirme Kriterleri	AlexNet	VGG16	VGG19	AlexNet	AlexNet	VGG16	AlexNet
				VGG16	VGG19	VGG19	VGG19
Normal Kesinliği	0,52	0,59	0,57	<b>0,63</b>	0,47	0,53	0,53
Hırıltı Kesinliği	<b>0,77</b>	<b>0,77</b>	0,71	0,71	0,71	<b>0,77</b>	0,71
Ekstra Kalp Sesi Kesinliği	0,75	<b>0,83</b>	0,6	<b>0,83</b>	0,6	0,67	0,71
Gürültü kesinliği	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>
Toplam Kesinlik	3,04	<b>3,19</b>	2,88	3,17	2,78	2,97	2,96

**Tablo 5.3.** Veri Seti-B için ESA modellerinin kombinasyonlarına göre kesinlik sonuçları

Veri Seti-B							
Değerlendirme Kriterleri	AlexNet	VGG16	VGG19	AlexNet	AlexNet	VGG16	AlexNet
				VGG16	VGG19	VGG19	VGG19
Normal Kesinliği	0,77	0,78	0,78	0,8	0,8	<b>0,82</b>	0,81
Hırıltı Kesinliği	0,55	0,67	0,61	0,64	0,66	<b>0,79</b>	0,69
Ekstra Sistol Kesinliği	0,08	0,2	0,24	0,4	0,08	<b>0,55</b>	0,2
Toplam Kesinlik	1,4	1,65	1,63	1,84	1,54	<b>2,15</b>	1,7

Önerilen yöntemin performansı daha önce literatürde aynı veri setlerini kullanmış metotlarla ve aynı ölçüm kriterlerine göre karşılaştırılmıştır. Daha önce kullanılan metotlarda tercih edilen öznitelik çıkarma ve sınıflandırma yöntemleri hakkında genel bilgi Tablo 5.4'e verilmiştir. Bu yöntemlerin hiç birinde DÖ tabanlı yöntemler kullanılmamıştır.

**Tablo 5.4.** Veri Seti-A ve Veri Seti-B kullanan diğer yöntemler

Metotlar	Öznitelikler	Sınıflandırıcılar
J48 [129]	max, min ve değişinti tabanlı öznitelikler	Karar Ağaçları
MLP [129]	max, min ve değişinti tabanlı öznitelikler	Çok Katmanlı Algılayıcı
CS-UCL [130]	Dalgacık Ayırıştırma+Spektrogram	Karar Ağaçları
SS [133]	Shannon enerji zarfi kullanan otoilinti Shannon enerji zarfi kullanan otoilinti+tekdüze	DVM
SS-PLSR [134]	öznitelikler	DVM
2D-PCA [131]	TBA	KNN
SVM-DM [135]	Ölçeklendirilmiş spektrogram	DVM
SS-TD [132]	Ham sinyal	Motif tabanlı yaklaşım

Önerilen yöntem ile Tablo 5.4'teki metotlar arasında aynı ölçüm kriterlerine göre yapılan karşılaştırma Veri Seti-A için Tablo 5.5'te ve Veri Seti-B için Tablo 5.6'da verilmiştir.

**Tablo 5.5.** Veri Seti-A için önerilen yöntemin diğer yöntemlerle karşılaştırılması

Veri Seti-A									
Değerlendirme Kriterleri	J48	MLP	CS-UCL	SS	SS-PLSR	2D-PCA	SVM-DM	SS-TD	Önerilen Yöntem
Normal Kesinliği	0,25	0,35	0,46	<b>0,67</b>	0,6	0,56	0,62	<b>0,67</b>	0,59
Hırıltı Kesinliği	0,47	0,67	0,31	0,91	0,91	<b>1</b>	0,91	<b>1</b>	0,77
Ekstra Kalp Sesi Kesinliği	0,27	0,18	0,11	0,37	0,44	0,3	<b>1</b>	0,43	0,83
Gürültü Kesinliği	0,71	0,92	0,58	0,76	0,94	0,94	0,64	0,8	<b>1</b>
Gürültü Duyarlılığı	0,63	0,69	0,44	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>
Gürültü Özgüllüğü	0,39	0,44	0,44	0,58	0,64	0,58	0,58	0,64	<b>0,69</b>
Gürültü Yİ	0,01	0,13	0,09	0,58	0,64	0,58	0,58	0,64	<b>0,69</b>
F-skor	0,2	0,2	0,14	0,28	0,3	0,26	0,31	0,3	<b>0,33</b>
Toplam Kesinlik	1,71	2,12	1,47	2,71	2,89	2,8	3,17	2,9	<b>3,19</b>
Normalleştirilmiş Kesinlik	0,45	0,59	0,4	0,72	0,76	0,75	0,76	0,76	<b>0,8</b>

**Tablo 5.6.** Veri Seti-B için önerilen yöntemin diğer yöntemlerle karşılaştırılması

Veri Seti-B									
Değerlendirme Kriterleri	J48	MLP	CS-UCL	SS	SS-PLSR	2D-PCA	SVM-DM	SS-TD	Önerilen Yöntem
Normal Kesinliği	0,72	0,7	0,77	0,74	0,76	0,78	0,77	<b>0,83</b>	0,81
Hırıltı Kesinliği	0,32	0,3	0,37	0,66	0,65	0,57	<b>0,76</b>	0,7	<b>0,76</b>
Ekstra Sistol Kesinliği	0,33	<b>0,67</b>	0,17	0,24	0,33	0,23	0,5	0,15	0,56
Normal hariç duyarlılık	0,22	0,19	0,51	0,24	0,34	0,41	0,34	0,49	<b>0,54</b>
Normal hariç özgüllük	0,82	0,84	0,59	0,84	0,9	0,84	<b>0,95</b>	0,84	0,91
Normal hariç Yİ	0,04	0,02	0,01	0,13	0,24	0,24	0,29	0,33	<b>0,45</b>
AG	0,05	0,04	0,09	0,24	0,36	0,3	0,54	0,39	<b>0,6</b>
Toplam Kesinlik	1,37	1,67	1,31	1,57	1,75	1,58	2,03	1,68	<b>2,15</b>
Normalleştirilmiş Kesinlik	0,6	0,62	0,63	0,67	0,69	0,68	0,74	0,74	<b>0,79</b>

Veri Seti-A için önerilen yöntemle, 10 tane performans ölçütünden 7 tanesinde en iyi sonuç elde edilmiştir. Önemli bir kriter olan toplam kesinlikte ise en iyi sonuca ulaşılmıştır.

Veri Seti-B için önerilen yöntemler 9 tane performans ölçütünden 6 tanesinde en sonuç elde edilmiştir. Toplam kesinlikte ise diğer yöntemlerde elde edilen en iyi sonuca göre 0.12'lik bir fark sağlanmıştır.

## **5.2. EMG Sinyalleri ve Akciğer Seslerinden Oluşan Veri Setleri için Derin Öznitelikler ve Transfer Öğrenme Yaklaşımlarının Uygulanması**

Uygulamalar için kullanılan her iki veri setinde derin öznitelikler ve transfer öğrenmesi yaklaşımlarının sonuçları irdelenmiştir.

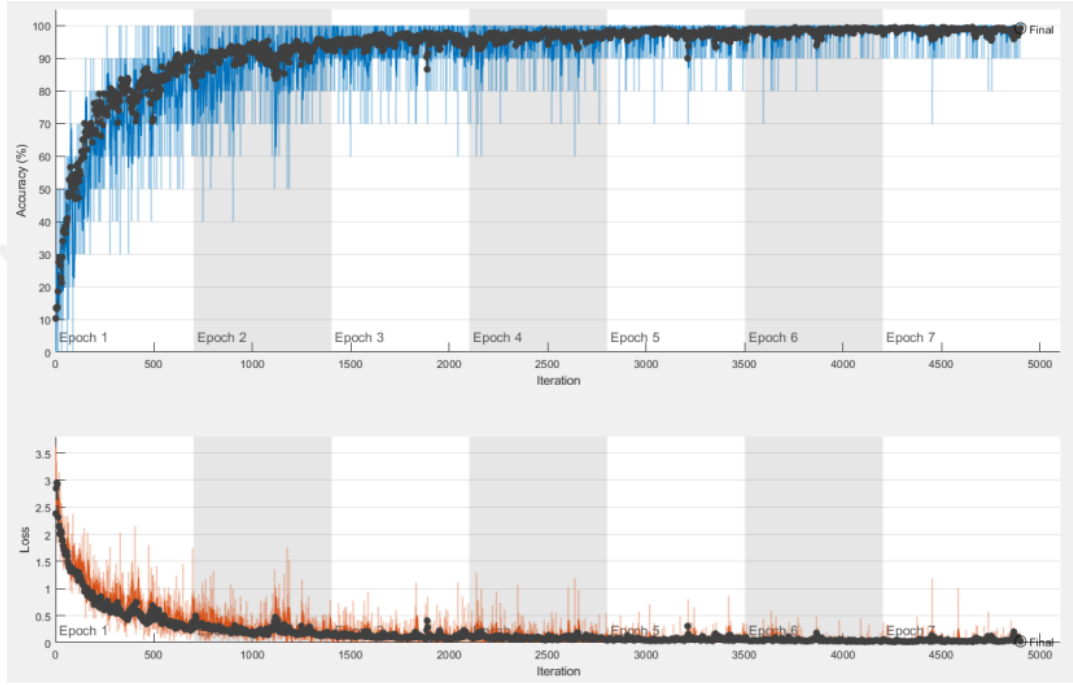
### **5.2.1. EMG Sinyalleri için Uygulama**

Önerilen yöntemlerde, yüzeysel EMG sinyalleri ile fiziksel eylemlerin sınıflandırılması hedeflediğinden, deneysel çalışmalar sırasında 10 farklı fiziksel eylem içeren bir veri seti göz önünde bulundurulmuştur [135]. Veri setinin her bir sınıfında üç erkek denek ve bir bayan denek ile veri seti oluşturulmuştur. Delsys isimli EMG cihazı, yüzeysel EMG sinyallerini kaydetmek için kullanılmıştır. Sinyaller, üst kol, pazı ve üst bacaklara yerleştirilen sekiz deri yüzeysel elektrot ile toplanmıştır. EMG yüzeysel sinyalleri eğilme, alkış, el sıkışma, sarılma, zıplama, koşma, ayakta durma, oturma, yürüme ve sallama fiziksel hareketlerine göre toplanmıştır.

Her elektrot için 10-kat çapraz geçirme testi ile deneyler yapılmıştır. Sonuçlarda, ortalama doğruluk ve elde edilen doğrulukların standart sapması kullanılmıştır. Her sınıf için 1000 spektrogram görüntüsü oluşturulmuştur. Bu nedenle, deneysel çalışmada toplam 10000 spektrogram görüntü kullanılmıştır. Doğrusal DVM sınıflandırıcısı, az miktarda eğitim verileri ile iyi bir sınıflandırma performansı verdiği için tercih edilmiştir. Düzgünleştirme parametresi C,  $10^{-4}$ - $10^3$  aralığında seçilmiştir. Transfer öğrenme modelinin ince ayarı için, mini-toptan büyüklüğü 10 ve başlangıç öğrenme oranı 0,0001 olarak atanmıştır. Maksimum dönem (epoch) sayısı 70 olarak ayarlanmıştır ve önerilen yöntemin optimizasyonu için RGİM metodu kullanılmıştır. Şekil 5.2, ince ayarlı AlexNet modelinin eğitim sürecini göstermektedir. İlk satır, yinelemelere karşı doğruluğun değişimini gösterirken, ikinci satır, yinelemelere karşı yitim değişimini gösterir. AlexNet modelinin ince ayarı için % 70 oranında rastgele veri kullanılmış ve geri kalan veri ile (% 30) ince ayarlı AlexNet ESA modeli test edilmiştir.

Derin öznitelik çıkarımı sonuçları Tablo 5.7'de verilmiştir. Tablo 5.7'de görüldüğü gibi, AlexNet fc6, VGG16 fc6, AlexNet fc6 + fc7, VGG16 fc6 + fc7, AlexNet fc6 + VGG16 fc6, AlexNet fc7 + VGG16 fc7 ve AlexNet fc6 + AlexNet fc7 + VGG16 fc6 + VGG16 fc7 öznitelik kombinasyonları derin öznitelik çıkarımlarında kullanılmıştır. Tablo 5.7'de kullanılan 8 elektrot içinde doğruluk sonuçları verilmiştir. Elektrot-1'deki AlexNet fc6 + VGG16 fc6 derin öznitelik birleştirme işlemi ile önerilen yöntemdeki en iyi sınıflandırma doğruluğuna ulaşılmıştır. İkinci en

iyi doğruluk, ortalama doğruluğun  $99,03 \pm 0,03$  olduğu AlexNet fc6 + AlexNet fc7 + VGG16 fc6 + VGG16 fc7 öznetelik birleştirmesi ile elde edilmiştir. Elektrot-2 için, AlexNet fc6 + AlexNet fc7 + VGG16 fc6 + VGG16 fc7 derin öznetelik birleştirme işlemi ile  $99,01 \pm 0,43$  en iyi üçüncü doğruluk skoruna ulaşılmıştır.  $98,89 \pm 0,37$  doğruluk skoru ise, AlexNet fc6 + VGG16 fc6 derin öznetelik birleştirme tarafından üretilen en iyi ikinci puandır.



Şekil 5.2. AlexNet ESA modeli için transfer öğrenme

Tablo 5.7. Derin özneteliklerin kombinasyonları ile her elektrot için doğruluk sonuçları

Derin Öznetelikler	Elektrotlar							
	1	2	3	4	5	6	7	8
AlexNet fc6	97,89±0,29	97,00±0,40	93,76±0,80	92,89±0,62	96,62±0,55	97,12±0,57	94,78±0,79	96,32±0,38
AlexNet fc6+fc7	98,00±0,35	98,33±0,25	95,04±0,83	94,54±0,75	97,31±0,51	97,66±0,40	95,27±0,53	97,34±0,56
VGG16 fc6	97,83±0,43	95,66±0,72	92,62±0,71	90,60±0,89	95,26±0,58	96,45±0,74	94,80±0,54	95,26±0,67
VGG16 fc6+fc7	97,93±0,35	96,15±0,49	93,68±0,60	92,01±0,88	95,99±0,56	96,97±0,46	95,38±0,91	97,54±0,42
Alexnet+VGG16 fc6	<b>99,04±0,43</b>	98,89±0,37	97,19±0,48	96,62±0,47	98,41±0,23	<b>98,83±0,45</b>	97,51±0,58	98,59±0,35
Alexnet+VGG16 fc7	98,27±0,53	98,00±0,47	95,02±0,60	94,51±0,79	97,08±0,61	97,71±0,54	95,90±0,58	97,32±0,41
Alexnet+VGG16 fc6, fc7	99,03±0,22	<b>99,01±0,43</b>	<b>97,40±0,51</b>	<b>97,32±0,36</b>	<b>98,75±0,25</b>	98,78±0,40	<b>97,86±0,50</b>	<b>98,83±0,34</b>

Elektrot-3 için elde edilen sonuçlar, elektrot 1 ve 2'den biraz daha düşüktür. Elektrot-3 için en iyi başarı, AlexNet fc6 + AlexNet fc7 + VGG16 fc6 + VGG16 fc7 derin öznetelik birleştirmesi ile elde edilen ortalama doğruluk skoru  $97,40 \pm 0,51$  olmuştur. Yakın bir skor  $97,19 \pm 0,48$  doğruluk skoru ile AlexNet fc6 + VGG16 fc6 derin öznetelik birleştirmesi ile elde edilmiştir. En

kötü sonuçlara elektrot 4 ile ulaşılmıştır. En düşük doğruluk skoru,  $90,60 \pm 0,89$ , VGG16 fc6 derin öznitelikleri ile elde edilmiştir. Bu puan aynı zamanda tüm deneysel başarılar arasında en düşük seviyededir. Ayrıca, elektrot 4 için en yüksek doğruluk skoru,  $97,32 \pm 0,36$  ile elde edilmiştir. Bu doğruluğa, AlexNet fc6 + AlexNet fc7 + VGG16 fc6 + VGG16 fc7 derin öznitelik birleştirmesi ile ulaşılmıştır. Elektrotlar 5, 7 ve 8 için, AlexNet fc6 + AlexNet fc7 + VGG16 fc6 + VGG16 fc7 derin öznitelik birleştirmesi uygulanarak, kaydedilen en yüksek doğruluk puanları sırasıyla  $98,75 \pm 0,25$ ,  $97,86 \pm 0,50$  ve  $98,83 \pm 0,34$  olmuştur. Elektrot-6 için, AlexNet fc6 + VGG16 fc6 derin öznitelik birleştirmesi ile,  $98,83 \pm 0,45$  en iyi doğruluk skoruna ulaşılmıştır. Genel olarak sonuçlar değerlendirilirse elektrot-1 ve 2 ile en iyi, elektrot-4 ile en kötü ve diğer elektrotlar ile makul sonuçlar alınmıştır. Elektrot-1 ve 6 dışındaki diğer elektrotlar ile en iyi sonuçlar AlexNet fc6 + AlexNet fc7 + VGG16 fc6 + VGG16 fc7 öznitelik birleştirilmesi ile elde edilmiştir. Elektrot-1 ve 6 ile en iyi sonuçlar ise AlexNet fc6 + VGG16 fc6 öznitelik birleştirilmesi ile elde edilmiştir.

Transfer öğrenme için elde edilen doğruluk sonuçları Tablo 5.8'de verilmiştir ve en iyi doğruluğa  $98,65$  ile elektrot-1 üzerinden elde edilmiştir. Elektrot-4 için en düşük doğruluk  $94,20$  olarak üretilmiştir. Hassasiyet, Özgüllük ve AUC Sonuçları Tablo 5.9'da verilmiştir. AUC değeri, hassasiyetin özgüllüğe göre değişimi ile elde edilen eğrinin altında kalan alanın değeridir. Tablo 5.9'da görüldüğü gibi, bütün özgüllük sonuçları aynı ve 1'e eşittir. En iyi hassasiyet ve AUC, elektrot-6 ile sırasıyla  $0,99$  ve  $0,9987$  olarak elde edilmiştir.

**Tablo 5.8.** Transfer öğrenme ile elde edilen doğruluk sonuçları

	Elektrotlar							
	1	2	3	4	5	6	7	8
Transfer öğrenme	<b>98,65%</b>	98,15%	95,10%	94,20%	98,40%	98,23%	96,89%	97,83%

**Tablo 5.9.** Transfer öğrenme ile elde edilen diğer performans ölçütleri

	Elektrotlar							
	1	2	3	4	5	6	7	8
Hassaslık	0,96	0,95	0,90	0,81	0,97	0,99	0,94	0,965
Özgüllük	1	1	1	1	1	1	1	1
AUC	0,9985	0,9986	0,9947	0,9940	0,9967	0,9987	0,9986	0,9986

Tablo 5.10'te ise önerilen yöntemlerle aynı veri setini kullanan diğer yöntemler arasındaki karşılaştırma doğruluk ölçütü üzerinden karşılaştırılmıştır. Önerilen yöntem hem transfer öğrenme hem de derin öznitelikler ile diğer yöntemlerden daha iyi sonuçlara ulaşmıştır.

**Tablo 5.10.** Önerilen yöntemler ile diğer yöntemlerin karşılaştırılması

Yazarlar	Metotlar	Doğruluk (%)
Abdullah vd. 2017 [136]	Matris tabanlı katsayılar, Rastgele Değişken sınıflandırıcısı	92,10
Sukumar vd. 2018 [137]	Değişimsel biçim tabanlı öznitelikler, DVM sınıflandırıcısı	98,17
Sahinbegovi vd. 2017 [138]	YSA	86,25
Önerilen yöntem	Derin özniteliklerin birleştirilmesi, DVM sınıflandırıcısı	<b>99,04</b>
Önerilen yöntem	Transfer Öğrenme	98,65

### 5.2.2. Akciğer Sesleri için Uygulama

EMG veri seti için yapılan çalışma da derin öznitelikler yaklaşımı transfer öğrenme yaklaşımına göre daha iyi doğruluk sonucu vermiştir. Bu sonucun genel geçerliliğini artırmak için akciğer seslerinden oluşan bir veri seti üzerinden her iki yaklaşımın performansı tekrar denenmiştir. Ayrıca VGG16 ve AlexNet modeline ek olarak ResNet-50 modeline de her iki yaklaşımda yer verilmiştir.

Yapılan deneysel çalışmalarda akciğer seslerinin sınıflandırılması için ICBHI 2017 yarışma veri seti kullanılmıştır [139]. Bu veri setinden 920 tane kaydedilmiş akciğer ses kayıtları bulunmaktadır. Ayrıca akciğer sesleri 4 KHz, 10 KHz ve 44,1 KHz olmak üzere üç farklı örnekleme frekansı ile kaydedilmiştir. Kaydedilen 1 tane ses dosyası içinde normal, çatlak ses, nefes darlığı ve çatlak ses + nefes darlığından oluşan 4 farklı sınıfta yer alabilir. Tablo 5.11’de örnek bir ses dosyası içinde bulunan farklı sınıfların nasıl etiketlendiği gösterilmiştir. Etiketler çatlak ses ve nefes darlığı sınıflarının “0” ve “1” atanma durumlarına göre belirlenmiştir. Çatlak ses, “0” ve nefes darlığı, “0” ise sınıf normal, çatlak ses, “1” ve nefes darlığı, “0” ise sınıf çatlak, çatlak ses, “0” ve nefes darlığı, “1” ise sınıf nefes darlığı ve çatlak ses, “1” ve nefes darlığı, “1” ise sınıf çatlak ses + nefes darlığı olmaktadır. Tablo 5.11’te ki “0” ve “1” atanma durumlarına göre 2 tane normal, 2 tane çatlak ses, 2 tane nefes darlığı ve 3 tane çatlak ses + nefes darlığı sınıfından akciğer sesleri bulunmaktadır. Bu sınıflara göre ayırma işlemi 920 tane ses dosyası için yapıldığında her sınıfta bulunan ses dosyasının sayısı Tablo 5.12’de verilmiştir.

**Tablo 5.11.** ICBHI veri setinde 1 ses dosyası içindeki sınıfların durumu için örnek gösterim

Süre İndeksleri	Başlangıç Zamanı (sn)	Bitiş Zamanı (sn)	Çatlak Ses	Nefes Darlığı
1	0,804	3,256	0	0
2	3,256	5,566	0	0
3	5,566	7,851	0	1
4	7,851	10,054	0	1
5	10,054	12,066	1	0
6	12,066	14,47	1	0
7	14,47	16,696	1	1
8	16,696	19,792	1	1

**Tablo 5.12.** Sınıflara göre ayrılan ses dosyaların sayısal durumu

<b>ICBHI Veri Seti</b>	<b>Toplam</b>
Çatlak ses sayısı	1864
Nefes darlığı sayısı	886
Çatlak ses + nefes darlığı sayısı	506
Normal ses sayısı	3642
Toplam ses sayısı	6898

EMG sinyallerinde olduğu gibi akciğer seslerinde de hem derin öznitelikler hem de transfer öğrenme ile sınıflandırma performansı ölçülmüştür. İlk adım olarak, diğer metotlarda olduğu gibi ön işleme aşamasında spektrogram görüntüleri oluşturulmuştur. ICBHI 2017 veri setinde 3 farklı örnekleme frekansı bulunduğu için çerçeve ve çakışma boyutları bu örnekleme frekanslarına göre uygun olarak seçilmiştir. Çerçeve ve çakışma boyutları, 4 KHz'lik örnekleme frekansı için 64 ile 8, 10 KHz'lik örnekleme frekansı için 128 ile 16 ve 44,1 KHz'lik örnekleme frekansı için 512 ve 64 olarak belirlenmiştir. Hamming pencerelemesi kullanılmış ve AFD uzunluğu 3000 olarak seçilmiştir. Spektrogram görüntüleri, AlexNet, VGG16 ve ResNet-50 için sırasıyla 227×227, 224×224 ve 224×224 olacak şekilde yeniden boyutlandırılmıştır.

Derin öznitelikler, AlexNet, VGG16 ve ResNet-50 modellerinin fc6 TBK'larından elde edilmiştir. AlexNet ve VGG16 modellerinin derin öznitelik boyutu 4096, ResNet-50 modelinin derin öznitelik boyutu ise 1000'dir. Bu modellerin sınıflandırma aşamasında Matlab programının Classification Learner menüsü kullanılmıştır. Bu menü DVM, karar ağaçları, KNN gibi sınıflandırıcılarının doğruluk performansını göstermektedir. Ayrıca bu menüde kullanılacak olan sınıflandırıcıların parametreleri otomatik olarak ayarlanmaktadır. Çıkarılan derin öznitelikler ile kullanılan bu menüde en iyi sınıflandırma performansını çok terimli DVM sınıflandırıcısı yakalamıştır.

Transfer öğrenme tabanlı yaklaşımda ise AlexNet, VGG16 ve ResNet-50 modelleri için ayrı ayrı sınıflandırma performansı ölçülmüştür. EMG veri seti için kullanılan eğitim parametrelerinin aynısı ICBHI 2017 veri seti içinde kullanılmıştır.

AlexNet, VGG16 ve ResNet-50 modelleri ile sınıflandırma doğruluğu sonuçları Tablo 5.13'te verilmiştir. Hem derin öznitelikler hem de transfer öğrenme tabanlı yaklaşımda en iyi sınıflandırma performansı VGG16 ESA modeli ile yakalanmıştır.

**Tablo 5.13.** ESA modelleri için derin öznitelikler ve transfer öğrenme ile sınıflandırma sonuçları

ESA Modelleri	Derin Öznitelikler+DVM	Transfer Öğrenme
AlexNet	%60,50	%61,23
ResNet-50	%59,10	%60,05
VGG16	<b>%65,50</b>	%63,09

Tablo 5.13'ten görüleceği üzere en iyi sınıflandırma sonucuna (%65,50) VGG16 modelini kullanan derin öznitelikler metodu ile ulaşılmıştır. AlexNet ve ResNet-50 modelleri, VGG16 modelinin tersine transfer öğrenme yaklaşımında derin öznitelikler yaklaşımına göre daha iyi sonuç vermiştir. VGG16 modeli ile kullanılan derin öznitelikler ve DVM sınıflandırıcısı ile elde edilen en iyi sınıflandırma performansı için karmaşıklık matrisi Şekil 5.3'te verilmiştir. Her sınıf için duyarlılık, özgüllük, kesinlik ve F-skor hesaplamaları bu karmaşıklık matrisi kullanılarak yapılmıştır ve sonuçlar Tablo 5.14'te verilmiştir. Bu sonuçlara göre en iyi duyarlılık, kesinlik ve F-skor değerleri normal ses sınıfında, en iyi özgüllük değeri ise çatlak ses + nefes darlığı sınıfında yakalanmıştır.

**Derin Öznitelikler (VGG16)+DVM**

	1	2	3	4
1	1116	58	634	56
2	134	156	126	90
3	556	59	2856	171
4	104	80	321	381

Tahmin Edilen Sınıflar

1. Çatlak Ses
2. Çatlak Ses+Nefes Darlığı
3. Normal Ses
4. Nefes Darlığı

**Şekil 5.3.** En iyi performansı veren model ve yöntem için karmaşıklık matrisi

**Tablo 5.14.** En iyi performansı veren model ve yöntem için diğer performans kriterlerinin sonuçları

Sınıflar	Duyarlılık	Özgüllük	Kesinlik	F-skor
Çatlak ses	0,60	0,81	0,58	0,59
Çatlak ses+Nefes darlığı	0,31	<b>0,96</b>	0,44	0,36
Normal ses	<b>0,78</b>	0,60	<b>0,73</b>	<b>0,75</b>
Nefes darlığı	0,43	0,93	0,55	0,48
<b>Ortalama değerler</b>	0,53	0,83	0,57	0,55

Önerilen her iki yöntemin ICBHI veri setini kullanan diğer yöntemler göre karşılaştırılması Tablo 5.15'te verilmiştir. Bu sonuçlara göre, transfer öğrenme yaklaşımı ile en iyi sınıflandırma doğruluğunu veren diğer yöntemle göre %5,18 oranında ve derin öznitelikler yaklaşımı ile de %7,62 oranında bir artış sağlanmıştır.

**Tablo 5.15.** Önerilen yöntemlerle ile diğer yöntemlerin karşılaştırılması

Yazarlar	Metot	Doğruluk (%)
N. Jakovljević vd., 2017 [139]	MFKK, SMM	39,56
G.Chambres vd., 2018 [140]	Düşük seviyeli öznitelikler, Karar Ağaçları	49,62
G.Serbes vd., 2018 [141]	KZFD+Dalgacık, DVM	57,88
<b>Önerilen Yöntem</b>	Derin Öznitelikler, DVM	<b>65,50</b>
Önerilen Yöntem	Transfer Öğrenme	63,09

### 5.3. Çevresel Seslere Derin Öznitelikler Tabanlı Yaklaşımın Önceden Eğitilmiş ESA modelleri ile Uygulanması

Önerilen yöntemi değerlendirmek için deneysel çalışmalarda kullanıma açık dört veri seti kullanılmıştır. Bu veri setlerinin isimleri ESC-50, ESC-10, DCASE-2013 ve UrbanSound8K'dır [47, 143, 144]. Her veri seti için sonuçlar ayrıntılı olarak irdelenmiştir.

#### 5.3.1. Veri Setleri

Kullanılan veri setlerinin temel özellikleri Tablo 5.16'da verilmiştir. Bu veri setleri literatürde en çok kullanılan veri setleridir.

**Tablo 5.16.** Kullanılan veri setlerinin temel özellikleri

Veri Setleri	Toplam Ses Sinyal Sayısı	Sınıf Sayısı	Ses Kayıt Sürelerinin Ortalama Uzunluğu
ESC10	400	10	5.01±0.17 sn
ESC50	2000	50	5.01±0.07 sn
DCASE-2013	100	10	30.00±0.00 sn
UrbanSound8K	8732	10	3.61±0.97 sn

ESC-50 veri seti, 50 sınıfın her bir sınıfında 40 ses dosyası içermektedir. ESC-10 veri kümesi ise ESC-50 veri setinin bir alt kümesi olup her sınıfta 40 ses dosyası olmak üzere 10 ses sınıfı ve toplam 400 ses dosyası içermektedir. ESC-50 veri setinden köpek havlaması, yağmur, deniz dalgaları, bebek ağlaması, saat sesi, hapşırma, helikopter, testere, horoz ve yangın çattırması gibi ses sınıfları seçilerek oluşturulmuştur. ESC-50 ve ESC-10 veri kümelerindeki tüm ses dosyaları, 44,1 kHz örnekleme frekansı ile 5 saniye süresince kaydedilmiştir.

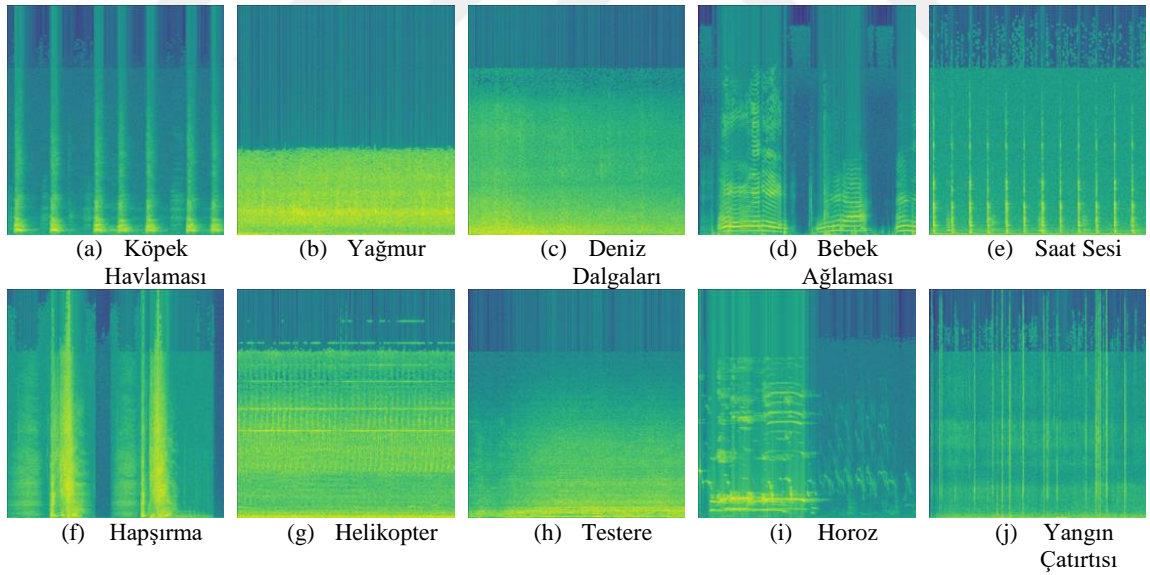
DCASE 2013 veri seti, her sınıfta 10 ses dosyası olmak üzere 10 sınıftan ve 100 tane ses dosyasından oluşmaktadır. Ses dosyaları 44,1 KHz örnekleme frekansıyla 30 saniye boyunca

kaydedilmiştir. DCASE 2013 veri setinin sınıfları otobüs, işlek cadde, ofis, açık hava pazarı, park, ana cadde, restoran, süpermarket, metro ve metro istasyonundan oluşmaktadır.

UrbanSound8K veri setinin sınıf sayısı ESC-10 ve DCASE-2013 veri setleri ile aynı olmakla birlikte çok daha fazla ses dosyasına sahiptir. Toplam 8732 ses dosyası içermekle birlikte bir sınıftaki ses dosya sayısı birbirine yakındır. UrbanSound8K veri setinden her bir ses dosyasının süresi eşit olmamakla birlikte 4 saniyeden kısadır ve 22,05 KHz örnekleme frekansıyla kaydedilmiştir. UrbanSound8k veri seti, klima, araba kornası, oynayan çocuklar, köpek havlaması, matkap, motor rölantisi, silah atış, kaya matkabı, siren ve sokak müziği sınıf etiketlerini içermektedir.

### 5.3.2. Deneysel Çalışmalar

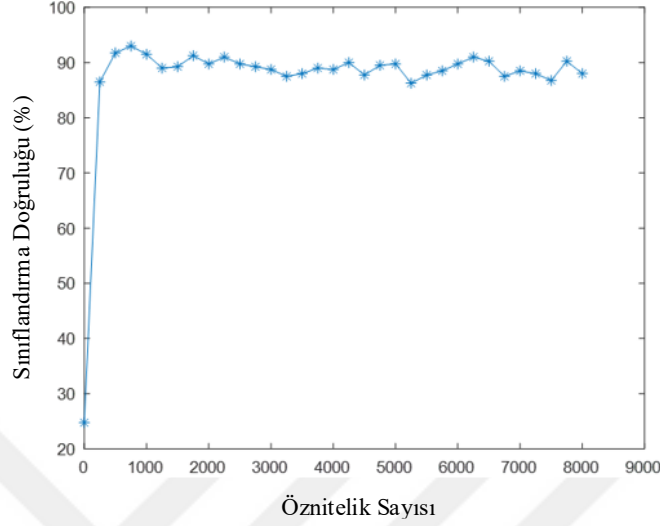
Önerilen yöntemde, ön işleme aşamasında spektrogram görüntüleri oluşturulmuştur. Kaydedilen seslerin örnekleme frekansları birbirine yakın olduğu için bütün veri setlerinde 1024 çerçeve boyutu ve 256 çakışma boyutu kullanılmıştır. Ayrıca AFD boyutu 3000 olarak belirlenmiş ve Hamming pencereleme tipi seçilmiştir. ESC-10 veri seti için her bir sınıftan örnek bir ses dosyasının spektrogram görüntüleri Şekil 5.4'te verilmiştir.



Şekil 5.4. ESC-10 veri seti için her sınıftan bir ses dosyasının spektrogram görüntüleri

Derin öznitelikler AlexNet ve VGG16 ESA modellerinin fc6 isimli TBK'larından elde edilmiştir. Her bir modelden bir ses dosyası için 4096 tane öznitelik çıkarılmıştır. Bu öznitelikler birleştirilerek bir ses dosyası için 8192 tane öznitelik seti oluşturulmuştur. Bu öznitelik setinin boyutu mFMU algoritması ile en iyi sınıflandırma performansını verecek şekilde seçilerek azaltılmıştır. ESC-10 veri seti için seçilen öznitelik sayılarına göre elde edilen sınıflandırma

doğrulukları Şekil 5.5'te verilmiştir. Şekil 5.5'e göre en iyi sınıflandırma doğruluğu 750 tane öznitelik seçimi ile elde edilmiştir.



**Şekil 5.5.** mFMU algoritması ile seçilen öznitelik sayısına göre sınıflandırma doğruluğu değişimi

Seçilen öznitelik seti doğrusal DVM algoritması ile sınıflandırmak üzere kullanılmıştır. DVM sınıflandırıcısının en iyi performansını yakalamak için düzgünleştirme parametresi  $10^{-3}$ - $10^2$  arasında değiştirilmiştir. Veri setleri, eğitim ve test için sırasıyla %80 ve %20 oranında rastgele seçim yapılarak ayrılmıştır.

Alexnet ve VGG16 modellerinin ayrı ayrı derin öznitelikleri, özniteliklerin birleştirilmesi ve birleştirilmiş özniteliklerden (8192 tane öznitelik) 750 tane seçim yapılmış durumlarındaki sınıflandırma doğrulukları Tablo 5.17'de verilmiştir.

**Tablo 5.17.** AlexNet ile VGG16 modellerinden derin öznitelikler elde edilmesi, öznitelik birleştirme ve öznitelik seçme durumlarında sınıflandırma doğrulukları

Öznitelik Durumları	Ortalama Doğruluk (%)			
	ESC-50	ESC-10	DCASE-2013	UrbanSound8K
AlexNet (4096)	69.05	84.75	55.00	66.59
VGG16 (4096)	70.05	86.25	60.00	67.47
Öznitelik Birleştirme (8192)	74.00	88.25	75.00	73.65
Öznitelik Seçme (750)	76.50	93.00	81.00	76.10

Tablo 5.17'de görüldüğü gibi VGG16 ile elde edilen derin öznitelikler bütün veri setlerinde AlexNet modelinden elde edilen derin özniteliklere göre daha iyi performans sağlamıştır. AlexNet ve VGG16 modellerinin özniteliklerinin birleştirilmesi ile sınıflandırma doğruluğu bütün veri setlerinde yeniden artırmıştır. En fazla artış %15 ile DCASE2013 veri setinde olurken, en az artış %2 ile ESC-10 veri setinde olmuştur. mFMU öznitelik seçme algoritmasının kullanılmasıyla

öznitelik boyutu 750 olduktan sınıflandırma doğruluğu bütün veri setlerinde daha da artmıştır. En fazla artış %6 ile DCASE 2013 veri setinde olurken, en az artış %2,45 ile UrbanSound8K veri setinde olmuştur.

ESC-50 ve ESC-10 veri setlerini literatürde daha önce kullanan diğer yöntemler ile önerilen yöntemin sınıflandırma doğruluğuna göre karşılaştırılması Tablo 5.18’de verilmiştir.

**Tablo 5.18.** ESC-50 ve ESC-10 veri seti için önerilen yöntemin diğer yöntemler ile karşılaştırılması

Yazarlar	Metot	Ortalama Doğruluk(%)	
		ESC-50	ESC-10
Piczak [47]	MFKK, DVM	39,60	67,50
Piczak [47]	MFKK, Rastgele Değişken [29]	44,30	72,70
Piczak [145]	ESA	64,50	81,00
Aytar vd. [146]	1 boyutlu ESA ve transfer öğrenme	74,20	92,20
Önerilen Yöntem	Derin Öznitelik birleştirme ve seçme, DVM	<b>76,50</b>	<b>93,00</b>

Önerilen yöntem, literatürde kullanılmış diğer en iyi yönteme göre sınıflandırma doğruluğunu ESC-50 veri seti için %2,5 ve ESC-10 veri seti için %0,8 kadar geliştirmiştir.

UrbanSound8K veri seti için literatürde daha önce kullanan diğer yöntemler ile önerilen yöntemin sınıflandırma doğruluğuna göre karşılaştırılması Tablo 5.19’da verilmiştir.

**Tablo 5.19.** UrbanSound8K veri seti için önerilen yöntemin diğer yöntemler ile karşılaştırılması

Yazarlar	Method	Ortalama Doğruluk (%)
Salamon vd. [48]	ESA	73.00
Piczak [145]	ESA	73.60
Salamon vd. [147]	Küresel K-Ortalamlar ile öznitelik öğrenme, Rastgele Değişken Sınıflandırıcısı	74.00
Önerilen Yöntem	Derin Öznitelik birleştirme ve seçme, DVM	<b>76.10</b>

Önerilen yöntem, literatürde kullanılmış diğer en iyi yönteme göre sınıflandırma doğruluğunu UrbanSound8K veri seti için %2,1 kadar geliştirmiştir.

DCASE-2013 veri seti için literatürde daha önce kullanan diğer yöntemler ile önerilen yöntemin sınıflandırma doğruluğuna göre karşılaştırılması Tablo 5.20’de verilmiştir.

**Tablo 5.20.** DCASE-2013 veri seti için önerilen yöntemin diğer yöntemler ile karşılaştırılması

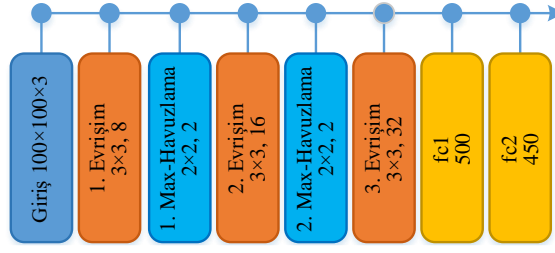
Yazarlar	Method	Ortalama Doğruluk (%)
Rakotomamonjy vd. [148]	Histogram of Oriented Gradient (HOG), DVM	69.00
Roma [149]	MFKK, DVM	77.00
Mesaros vd. [150]	Mel Spektrum, Topluluk Sınıflandırıcısı	78.00
Önerilen Yöntem	Derin Öznitelik birleştirme ve seçme, DVM	81.00
Aytar vd. [146]	1 boyutlu ESA ve transfer öğrenme	<b>88.00</b>

Önerilen yöntem, literatürde kullanılmış diğer en iyi yönteme göre sınıflandırma doğruluğunda en iyi ikinci performansı yakalamıştır.

#### **5.4. Çevresel Seslere Derin Öznitelikler Tabanlı Yaklaşımın Uçtan Uca Yapılandırılmış ESA modelleri ile Uygulanması**

Bu bölümde, önerilen yöntemin uygulanması için UrbanSound8K ve DCASE-2017 ASC [151] veri setleri kullanılmıştır. UrbanSound8K veri setinin yapısı Bölüm 5.3'te belirtilmiştir. DCASE-2017 ASC veri seti çevresel ses yarışması için organize edilmiştir ve bu kapsamda eğitim ve değerlendirme olarak iki bölüme ayrılmıştır. Eğitim veri setinde her birinin uzunluğu 10 saniye olan 4680 tane ses dosyası bulunmaktadır. Değerlendirme veri setinde ise her birinin uzunluğu 10 saniye olan 1620 tane ses dosyası bulunmaktadır. Bu ses dosyaları 44,1 KHz örnekleme frekansıyla kaydedilmiştir. DCASE-2017 ASC veri seti, plaj, otobüs, kafe / restoran, araba, şehir merkezi, orman yolu, bakkal, ev, kütüphane, metro istasyonu, ofis, park, yerleşim bölgesi, tren, tramvay sınıf etiketlerinden oluşmaktadır.

Önerilen yöntemde ses sinyallerinin spektrogram görüntüleri oluşturulurken çerçeve boyutu 1024, çakışma boyutu 256, AFD uzunluğu 3000 seçilmiş ve Hamming pencerelemesi kullanılmıştır. Daha sonra spektrogram görüntülerinin genişliği ve yüksekliği 100×100 olacak şekilde yeniden boyutlandırılmıştır. Uçtan uca yapılandırılan ESA modelinin evrişim, maksimum havuzlama, fc1 ve fc2 katmanlarında kullanılan boyutsal parametreler Şekil 5.6'da gösterilmiştir.



Şekil 5.6. Kullanılan ESA modelinin katmanlarının boyutsal parametreleri

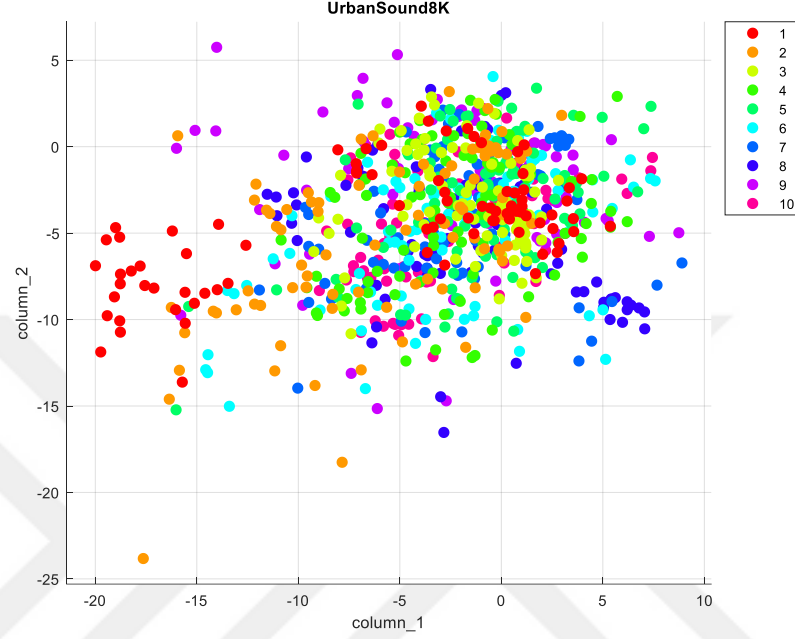
Şekil 5.6’da giriş katmanının boyutu  $100 \times 100 \times 3$  olarak belirlenmiştir. Evrişim katmanlarında, 1. evrişim katmanının filtre boyutu  $3 \times 3$ , filtre sayısı 8, 2. evrişim katmanının filtre boyutu  $3 \times 3$ , filtre sayısı 16, 3. evrişim katmanının filtre boyutu  $3 \times 3$ , filtre sayısı 32 olarak belirlenmiştir. Maksimum havuzlama katmanlarında ise 2 adımlık kayma ile  $2 \times 2$  boyutlu filtreler kullanılmıştır. fc1 ve fc2 katmanlarının boyutları ise sırasıyla 500 ve 450 olarak seçilmiştir. fc1 ve fc2 boyutları önerilen yöntemin her iki veri seti üzerindeki sınıflandırma sonuçlarına göre belirlenmiştir. Her iki veri seti için fc1 ve fc2 boyutlarının değişimi ile alınan sınıflandırma doğrulukları Tablo 5.21’de verilmiştir. ESA modeli eğitilirken başlangıç öğrenme oranı 0,005 ve mini-toptan büyüklüğü 100 olacak şekilde ayarlanmıştır.

Sınıflandırma aşamasında ise kullanılan RAKT algoritmasında kullanılan K-EYK sınıflandırıcılarının en yakın komşu değeri (K) 1 olarak, alt uzay sayısı 64 olarak seçildiğinde [152]’e göre en iyi performansın elde edildiği belirtilmiştir. Her iki veri seti test etmek için 10-Kat çapraz geçiş yöntemi kullanılmıştır.

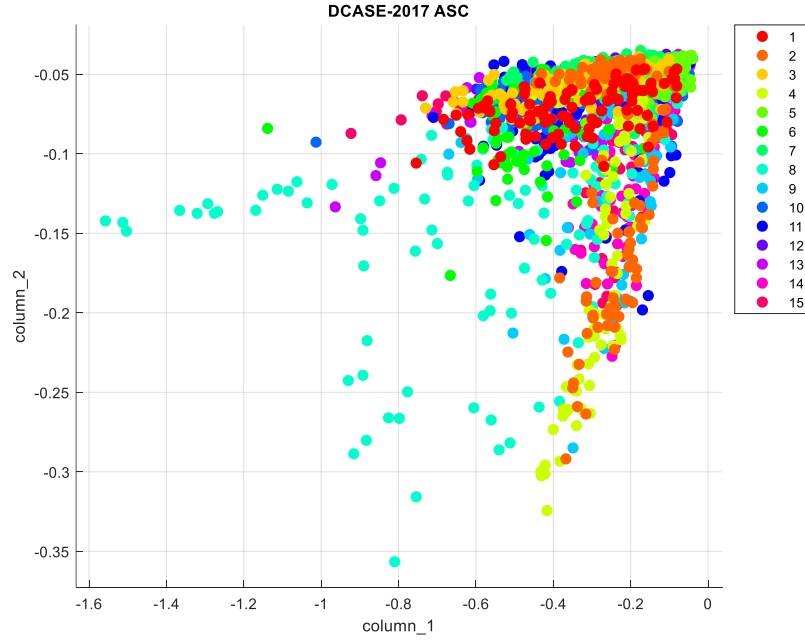
Tablo 5.21. TBK boyutlarına göre sınıflandırma doğrulukları

TBK boyutları		Sınıflandırma Doğrulukları (%)			
		UrbanSound8K		DCASE-2017 ASC	
fc1	fc2	fc1	fc2	fc1	fc2
100	50	78.6	77.4	84.5	84.1
150	100	79.2	78.3	85.7	85.2
200	150	80.3	79.4	86.9	85.6
250	200	81.1	80.4	88.7	87.5
300	250	82.6	81.3	90.2	89.4
350	300	83.1	82.9	91.3	90.8
400	350	83.4	83.3	92.8	91.9
450	400	84.1	83.7	94.7	93.8
<b>500</b>	<b>450</b>	<b>85.5</b>	<b>84.6</b>	<b>95.3</b>	<b>94.2</b>
550	500	84.2	83.9	93.7	92.5
600	550	83.7	82.5	91.4	90.6
650	600	82.8	81.5	86.5	84.2

UrbanSound8K ve DCASE-2017 ASC veri seti için uçtan uca ESA modelinin fc1 ve fc2 katmanlarının birleştirilmesiyle elde edilen öznitelik setlerinin dağılımı Şekil 5.7 ve 5.8’de gösterilmiştir. Tablo 5.22’de, uçtan uca ESA modelinden ve diğer önceden eğitilmiş ESA modellerinden öznitelik çıkarma süreleri verilmiştir.



Şekil 5.7. UrbanSound8K veri seti için öznitelik setinin dağılımı

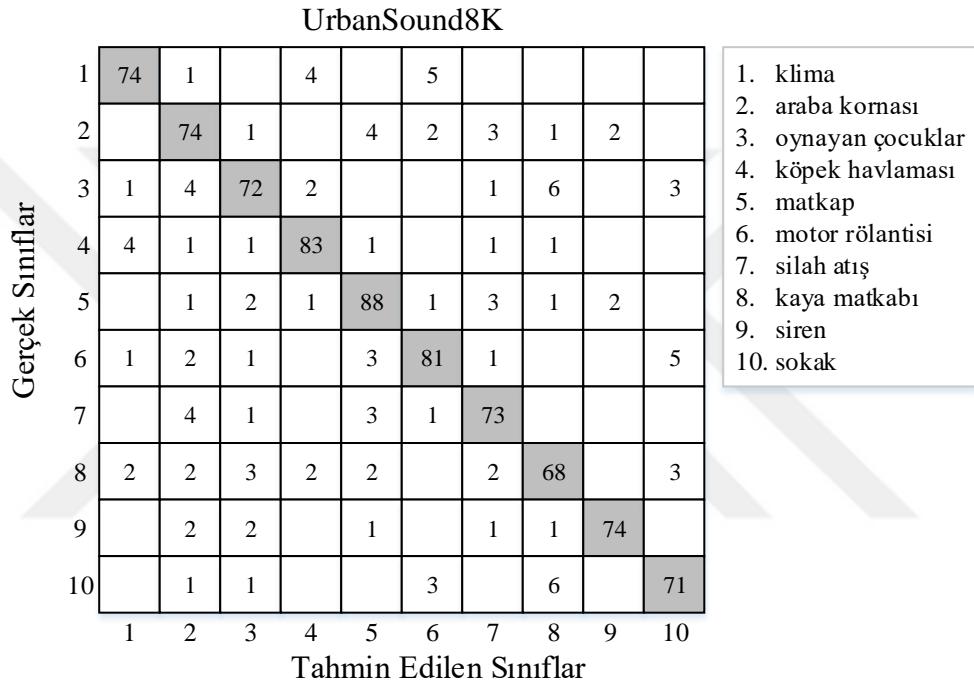


Şekil 5.8. DCASE-2017 ASC veri seti için öznitelik setinin dağılımı

**Tablo 5.22.** ESA modellerinden öznelik çıkarma süreleri

Veri Setleri	Alexnet	VGG16	VGG19	ResNet-50	Önerilen Yöntem
DCASE-2017 ASC	2432 sn	2835 sn	3145 sn	4670 sn	<b>576 sn</b>
UrbanSound8K	2657 sn	3015 sn	3467 sn	4876 sn	<b>650 sn</b>

Her iki veri seti için önerilen yöntemle en iyi sınıflandırma doğruluğunun sağlandığı koşulda karmaşıklık matrislerinin durumu Şekil 5.9 ve 5.10'da verilmiştir.



**Şekil 5.9.** UrbanSound8K veri seti için karmaşıklık matrisi

DCASE-2017 ASC

Gerçek Sınıflar	1	105		2										1		
	2		105		1											2
	3			104				4								
	4		1		104											3
	5	1	1			104								2		
	6						104		1					3		
	7			3				105								
	8			1				1	100	4		2				
	9			1			1		3	101		2				
	10										108					
	11							1	3			104				
	12			1		1							105	1		
	13	2				5	4							97		
	14				1	1									105	1
	15															108
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
	Tahmin Edilen Sınıflar															

1. plaj
2. otobüs
3. kafe / restoran
4. araba
5. şehir merkezi
6. orman yolu
7. bakkal
8. ev
9. kütüphane
10. metro istasyonu
11. ofis
12. park
13. yerleşim bölgesi
14. tren
15. tramvay

Şekil 5.10. DCASE 2017-ASC veri seti için karmaşıklık matrisi

Önerilen metotta kullanılan uçtan uca ESA modelleri ile öznelik çıkarımı kullanılarak alınan sınıflandırma doğruluğu sonuçlarının, bilinen ESA modelleri ile öznelik çıkarımı kullanılarak alınan sınıflandırma doğruluğu sonuçlarına göre karşılaştırılması her iki veri seti için Tablo 5.23 ve 5.24'te verilmiştir.

**Tablo 5.23.** UrbanSound8K veri seti için ESA modellerinden derin öznelik çıkarımı kullanılarak elde edilen sınıflandırma sonuçları

Sınıflandırıcılar	ESA modellerine göre sınıflandırma doğruluğu (%)					
	GoogleNet	AlexNet	VGG16	VGG19	ResNet-50	Uçtan Uca ESA
Karar ağaçları	27,7	35,7	38,1	34,5	35,2	44,1
KNN	59,8	70,35	71,35	69,8	70,8	84,2
SVM	65,4	76,2	77,1	75	75,95	78
Altuzay Ayırtaç Topluluk	36,7	60,5	61,6	59,2	60,65	65,4
RAKT	52,3	71,35	72,35	70,65	70,9	86,7

**Tablo 5.24.** DCASE-2017 ASC veri seti için ESA modellerinden derin öznitelik çıkarımı kullanılarak elde edilen sınıflandırma sonuçları

Sınıflandırıcılar	ESA modellerine göre sınıflandırma doğruluğu (%)					
	GoogleNet	AlexNet	VGG16	VGG19	ResNet-50	Uçtan Uca ESA
Karar ağaçları	40,3	61,56	62,32	61,44	60,1	67,2
KNN	50,65	69,18	70,78	67,16	67,54	79,1
SVM	67,56	<b>80,35</b>	<b>81,2</b>	<b>79,34</b>	<b>79,1</b>	85,4
Altuzay Ayırtaç Topluluk	55,3	68,2	69,26	67,96	66,12	88,2
RAKT	63,2	76,15	79,34,	75,82	74,53	<b>96,23</b>

Önerilen yöntemde kullanılan uçtan uca ESA modelleri ile derin öznitelikler çıkararak RAKT algoritmasıyla sınıflandırma yapmak her iki veri seti için de en sonucu vermiştir. UrbanSound8K ve DCASE-2017 ASC veri seti için sınıflandırma doğruluğu, diğer ESA modellerinden en iyi sonucu veren modele göre (VGG16) sırasıyla %9,6 ve %15,13 olarak geliştirilmiştir.

Önerilen yöntemle aynı veri setlerini kullanan diğer yöntemlerin sınıflandırma doğruluğuna göre her sınıf için karşılaştırılması Tablo 5.25 ve 5.26'da verilmiştir.

**Tablo 5.25.** DCASE-2017 ASC veri setini kullanan diğer yöntemlere göre önerilen yöntemin her sınıf için karşılaştırılması

Sınıf İndeksleri	Sınıflandırma Doğrulukları (%)					Önerilen Metot
	BUETBOSCH1 [156]	DCNN_SVM [155]	F1EnsemSel [154]	GAN_SKMUN [153]		
1	87,0	71,3	78,7	83,3	<b>97,2</b>	
2	59,3	84,3	71,3	74,1	<b>97,2</b>	
3	91,7	79,6	83,3	88,0	<b>96,3</b>	
4	92,6	85,2	93,5	93,5	<b>96,3</b>	
5	94,4	82,4	88,9	94,4	<b>96,3</b>	
6	91,7	78,7	<b>98,1</b>	95,4	96,3	
7	81,5	80,6	79,6	82,4	<b>97,2</b>	
8	<b>97,2</b>	73,1	94,4	88,0	92,6	
9	47,2	59,3	53,7	75,9	<b>93,5</b>	
10	76,9	97,2	100	88,0	<b>100</b>	
11	49,1	81,5	86,1	92,6	<b>96,3</b>	
12	38,0	57,4	44,4	75,9	<b>97,2</b>	
13	58,3	85,2	75,9	86,1	<b>90</b>	
14	81,5	92,6	90,7	67,6	<b>97,2</b>	
15	65,7	57,4	66,7	63,9	<b>100</b>	
Ortalama Doğruluk	74,1	77,7	80,4	83,3	<b>96,2</b>	

**Tablo 5.26.** URBANSOUND8K veri setini kullanan diğer yöntemlere göre önerilen yöntemin her sınıf için karşılaştırılması

Sınıf İndeksleri	Sınıflandırma Doğrulukları (%)				Önerilen Metot
	Baseline System [143]	Piczak CNN [145]	SKM [147]	SB-CNN [48]	
1	50,40	55,70	51,30	48,90	<b>88,1</b>
2	46,97	78,60	63,25	<b>88,13</b>	85,1
3	71,60	82,00	76,60	<b>83,00</b>	80,9
4	74,90	84,00	79,50	90,00	<b>90,2</b>
5	76,60	66,30	79,90	80,20	<b>88,9</b>
6	64,10	67,90	77,20	79,80	<b>86,2</b>
7	90,64	92,51	91,71	<b>94,11</b>	89
8	63,00	62,70	70,20	67,30	<b>81</b>
9	74,00	81,05	75,67	85,79	<b>91,4</b>
10	75,00	76,00	77,00	84,40	<b>86,6</b>
Ortalama Doğruluk	68,57	73,09	73,69	78,65	<b>86,7</b>

DCASE-2017 ASC veri setinde kullanılan diğer yöntemlerin hepsi ESA tabanlı yaklaşımları içermektedir. Bu veri seti bir yarışma setidir ve toplam 76 metot bu yarışma kapsamında değerlendirilmiştir. Önerilen yöntem bu yarışmada en iyi sınıflandırma doğruluğuna sahip 4 metotla karşılaştırılmıştır ve önerilen yöntem, toplam 15 sınıfın 13'ünde en iyi ortalama doğruluk değerine ulaşmıştır. Ortalama doğrulukta ise [153]'e göre %12,9 oranında bir artış sağlanmıştır.

UrbanSound8K veri setinde ise literatürde bu veri setini kullanılan metotlarla karşılaştırma yapılmıştır. Bu metotların ikisi ESA (CNN isimli olanlar) tabanlı yaklaşım kullanırken diğer ikisi geleneksel yöntemler kullanmıştır. Önerilen yöntem, toplam 10 sınıfın 7'sinde en iyi ortalama doğruluk değerine ulaşmıştır. Ortalama doğrulukta ise [48]'e göre %12,9 oranında bir artış sağlanmıştır.

Her iki veri seti üzerinde önerilen yöntemin duyarlılık, özgüllük, kesinlik ve F-skor performans kriterlerine göre her sınıf için aldığı değerler Tablo 5.27 ve 5.28'de verilmiştir. UrbanSound8K veri setinde en iyi performans kriterlerini elde eden sınıf siren olurken, DCASE2017-ASC veri setinde en iyi performans kriterlerini elde eden sınıf metro istasyonu olmuştur.

**Tablo 5.27.** UrbanSound8K veri seti için önerilen yöntemin diğer performans kriterlerine göre her sınıf için aldığı değerler

Sınıflar	Duyarlılık	Özgüllük	Kesinlik	F-skör
Klima	0,8810	0,9899	0,9024	0,8916
Araba kornası	0,8506	0,9771	0,8043	0,8268
Oynayan çocuklar	0,8090	0,9847	0,8571	0,8324
Köpek Havlaması	0,9022	0,9885	0,9022	0,9022
Matkap	0,8889	0,9819	0,8627	0,8756
Motor rölantisi	0,8617	0,9846	0,8710	0,8663
Silah atışı	0,8902	0,9848	0,8588	0,8743
Kaya matkabı	0,8095	0,9797	0,8095	0,8095
Siren	<b>0,9136</b>	<b>0,9950</b>	<b>0,9487</b>	<b>0,9308</b>
Sokak	0,8659	0,9861	0,8659	0,8659

**Tablo 5.28.** UrbanSound8K veri seti için önerilen yöntemin diğer performans kriterlerine göre her sınıf için aldığı değerler

Sınıflar	Duyarlılık	Özgüllük	Kesinlik	F-skör
Plaj	0.9722	0.9980	0.9722	0.9722
Otobüs	0.9722	0.9987	0.9813	0.9767
Kafe/Restoran	0.9630	0.9947	0.9286	0.9455
Araba	0.9630	0.9987	0.9811	0.9720
Şehir merkezi	0.9630	0.9954	0.9369	0.9498
Orman yolu	0.9630	0.9967	0.9541	0.9585
Bakkal	0.9722	0.9967	0.9545	0.9633
Ev	0.9259	0.9974	0.9615	0.9434
Kütüphane	0.9352	0.9947	0.9266	0.9309
Metro istasyonu	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>
Ofis	0.9630	0.9974	0.9630	0.9630
Park	0.9722	<b>1.000</b>	<b>1.000</b>	0.9859
Yerleşim bölgesi	0.8981	0.9954	0.9327	0.9151
Tren	0.9722	<b>1.000</b>	<b>1.000</b>	0.9859
Tramvay	1.000	0.9960	0.9474	0.9730

## 6. SONUÇLAR

Bu tez çalışmasında, konuşma olmayan sesler ve biyomedikal işaretlerin derin öğrenmede popüler olan ESA tabanlı yaklaşımlarla sınıflandırılması üzerine durulmuştur. Bütün çalışmaların uygulanıp deneysel sonuçların elde edilmesinde, Windows 10 işletim sistemine ve Intel Core i7-5500 CPU@ 2.4GHZ, 8 GB bellekli donanımsal özelliklere sahip bir bilgisayara yüklü MATLAB programı kullanılmıştır

Bu doğrultuda ilk yapılan çalışmada, kalp seslerinin derin öznitelikleri çıkartıldıktan sonra DVM algoritması ile sınıflandırması yapılmıştır. Derin öznitelikler çıkartılırken AlexNet, VGG16 ve VGG19 ESA modelleri kullanılmıştır. Bu modellerin ortak tarafı 1 milyondan fazla görüntü içeren aynı veri seti ile eğitilmiş olmalarıdır. Ayrıca her modelin öğrenilebilir parametreleri kullanıma açılmıştır. Derin öznitelikler bu modellerin tam bağlı katmanlarından elde edilmiştir. Ayrıca modellerin özniteliklerin birleştirilmesiyle elde edilen sınıflandırma performansları da ölçülmüştür.

İkinci yapılan çalışmada, başka bir derin öğrenme yaklaşımı olan transfer öğrenme yaklaşımı kullanılmıştır. Ayrıca transfer öğrenme tabanlı ve derin öznitelikler tabanlı yaklaşımların performansları karşılaştırılmıştır. Karşılaştırmaların genel geçerliliğini artırmak için önerilen yöntemler hem EMG sinyallerini hem de akciğer seslerini içeren veri setleri üzerinde denenmiştir. Her iki veri seti için de derin özniteliklerin kullanıldığı yöntemin daha iyi sınıflandırma performansı sağlamıştır.

Üçüncü yapılan çalışmada, ikinci çalışmada da derin özniteliklerin kullanıldığı yaklaşımın daha iyi sonuç vermesi üzerine derin öznitelikler tabanlı yaklaşım tercih edilmiştir. Çıkarılan derin öznitelikler üzerinde mFMU öznitelik seçme algoritması kullanılarak sınıflandırma performansı daha da artırılmaya çalışılmıştır. Bu önerilen yöntem, çevresel sesleri içeren dört ayrı veri seti üzerinde denenmiştir. Bütün veri setlerinde öznitelik seçme algoritmasıyla daha iyi sınıflandırma performansına ulaşılmıştır.

Dördüncü yapılan çalışmada, derin öznitelikleri çıkarmak için görüntü ile eğitilmiş AlexNet ve VGG16 gibi önceden eğitilmiş ve ağırlıkları paylaşılmış ESA modelleri kullanmak yerine, çevresel sesler ile eğitilen uçtan uca yapılandırılmış bir ESA modeli kullanılmıştır. Bu ESA modelinin eğitilmesi için veri setlerinin belli bir kısmı kullanılmıştır. Veri setlerinin kalan kısmı ile de derin öznitelikler çıkartılmış ve sınıflandırma yapılmıştır. Önerilen bu yöntem sınıf başına örnek sayısı fazla olan iki veri seti üzerinde uygulanmıştır. Sonuçta, uçtan uca yapılandırılmış ESA modeli ile çıkartılan öznitelikler, görüntü ile önceden eğitilmiş ESA modellerinden çıkartılan özniteliklere göre daha iyi sınıflandırma performansı sağlamıştır. Ayrıca bu sonuçlar daha küçük giriş boyutu, daha az katman ve daha az öznitelik sayısı ile elde edilmiştir.

## 6.1. Sonuçların Değerlendirilmesi

İlk çalışmada, derin öznitelikler çıkarmak için AlexNet, VGG16 ve VGG19 modelleri kullanılmıştır. Derin öznitelikler bu modellerin fc6 isimli tam bağlı katmanlarından çıkarılmıştır. Öznitelik seti oluşturulurken bu ESA modellerinin birleştirilmesi ile oluşturulan kombinasyonlar da kullanılmıştır. Önerilen yöntemler, farklı yöntemlerle (telefon ve stetoskop) ve farklı örnekleme frekanslarıyla kaydedilen kalp seslerini içeren iki ayrı seti üzerinde uygulanmıştır. DVM algoritması ile sınıflandırma işlemi yapıldıktan sonra Veri Seti-A için en iyi toplam kesinlik sonucu (3,19), VGG16 modelinden çıkartılan derin özniteliklerden elde edilirken, Veri Seti-B için ise en iyi toplam kesinlik sonucu (2,15), VGG16 ve VGG19 modelinden çıkartılan derin özniteliklerin birleştirilmesi elde edilmiştir. Ayrıca Veri Seti-A için kullanan diğer yöntemlere göre toplam kesinlik 0,02 kadar ve Veri Seti-B için kullanan diğer yöntemlere göre toplam kesinlik 0,12 kadar artırılmıştır.

İkinci çalışmada, transfer öğrenme ile derin öznitelikler kullanan yaklaşımlar, EMG sinyalleri ve akciğer seslerini içeren iki veri seti üzerinden karşılaştırılmıştır. EMG veri setinde toplam 8 elektrot içinde sınıflandırma sonuçları verilmiştir. Transfer öğrenme tabanlı yaklaşımda en iyi sınıflandırma doğruluğu %98,65, AlexNet modelinin kullanımı ile birinci elektrottan alınmıştır. Derin öznitelikler ile DVM sınıflandırma algoritması kullanan yaklaşımda ise en iyi sonuç %99,04, Alexnet ve VGG16 modellerinin tam bağlı katmanlarından çıkarılan özniteliklerin birleştirilmesi ile birinci elektrottan alınmıştır. Aynı veri setinin kullanan diğer yöntemlere göre sınıflandırma doğruluğu %0,87 kadar geliştirilmiştir. Akciğer seslerini içeren veri setinde ise sınıflandırma doğruluğu, VGG16 modelini kullanan transfer öğrenme yaklaşımında %63,09, VGG16 modelinden elde edilen derin özniteliklerle DVM sınıflandırma algoritması kullanan yaklaşımda %65,5 olmuştur. Aynı veri setinin kullanan diğer yöntemlere göre sınıflandırma doğruluğu %7,62 kadar geliştirilmiştir.

Üçüncü çalışmada ise, ikinci çalışmada da derin öznitelikler tabanlı yaklaşımdan transfer öğrenme tabanlı yaklaşımına göre daha iyi sonuç alınması üzerine, derin öznitelikler tabanlı yaklaşım ile birlikte mFMU öznitelik seçme algoritması kullanılarak sınıflandırma doğruluğunun daha da artırılmıştır. 8192 tane derin öznitelik VGG16 ve AlexNet modelinin fc6 isimli tam bağlı katmanlarının çıkışlarının birleştirilmesi ile elde edilmiştir. mFMU algoritması ile 8192 tane derin öznitelik arasından en iyi sınıflandırma sonucu veren 750 tane derin öznitelik seçilmiştir. Öznitelik seçme işlemi ile sınıflandırma doğruluğu dört veri seti için de artırılmıştır. DCASE-2013 veri seti hariç aynı veri setlerini kullanan diğer yöntemlere göre ise en iyi sınıflandırma doğruluğu elde edilmiştir. DCASE-2013 veri setinde ise en iyi ikinci performansa ulaşılmıştır. En iyi sonucun alınmamasında DCASE-2013 veri setinin ses dosya sayısının azlığı (100) etkili olmuştur. Çünkü büyük veri setleri ESA modelleri ile çalışmaya daha uygundur.

Dördüncü çalışmada, derin öznitelikler çıkarmak için büyük bir görüntü veri seti ile eğitilen AlexNet ve VGG19 gibi modeller kullanılmamıştır. Bunun yerine  $100 \times 100 \times 3$  giriş katmanına, 3 tane evrişim katmanına, 2 tane maksimum havuzlama katmanına, 3 tane DDB katmanına ve 3 tane TBK'ya sahip daha az katmanlı bir ESA modeli kullanılmıştır. Bu ESA modeli çevresel sesleri içeren veri setlerinin belli bir kısmı kullanılarak eğitilmiştir. Derin öznitelikler, uçtan uca yapılandırılmış bu ESA modelinin fc1 (500 tane) ve fc2 (450 tane) isimli tam bağlı katmanlarının birleştirilmesi ile oluşturulmuştur. Daha az boyutlu ve katmanlı bir yapıya sahip olması nedeniyle bu özniteliklerin oluşturulması, önceden eğitilmiş ve ağırlıkları paylaşılmış ESA modellerinden özniteliklerin oluşturulmasına göre daha az zaman almıştır. Önerilen yöntemle her iki veri setini kullanan diğer yöntemlere göre sınıflandırma doğruluğu önemli bir oranda geliştirilmiştir. UrbanSound8K veri seti için sınıflandırma doğruluğu, en iyi sonucu veren diğer yöntemlere göre %8,05 kadar, DCASE-2017 ASC veri seti için sınıflandırma doğruluğu en iyi sonucu veren diğer yöntemlere göre %12,9 kadar geliştirilmiştir. Bu yöntem uygulanırken dikkat edilmesi gereken bir husus ise sınıf başına düşen örnek sayısının çok olması gerekliliğidir. Çünkü örneklerin bir kısmı uçtan uca yapılandırılmış ESA modelini eğitmek için kullanılmaktadır.

## ÖNERİLER

Tezde önerilen yöntemlerde, Derin Öğrenmenin bir alt konusu olan Evrişim Sinir Ağlarını temel alan yaklaşımlar kullanılmıştır. Önerilen yöntemlerin hepsinde geleneksel makine öğrenme algoritmaları kullanan yaklaşımlara göre daha iyi bir performans elde edilmiştir. Bununla birlikte, çevresel seslerin ve biyomedikal işaretlerin sınıflandırılması gelişime açık bir alandır. Bu alanda sınıflandırma performansını artırmak için farklı tarzlarda yaklaşımlar kullanılabilir. İlk olarak, ön işleme aşamasında ses karakteristiklerini daha iyi temsil eden algoritmalar geliştirilip kullanılabilir. İkinci olarak, AlexNet, VGGNet, ve ResNet gibi görüntü ile önceden eğitilmiş Evrişim Sinir Ağları modellerinde olduğu gibi ses ile eğitilmiş Evrişim Sinir Ağları modelleri oluşturulabilir. Bu modelleri oluşturmak için hem büyük miktarda ses verilerine hem de bu verileri eğitebilecek yüksek kapasiteli bilgisayarlara ihtiyaç vardır. Önerilen 4. yöntem, bu uygulamanın bir alt yapısı gibi düşünülebilir. Bu yöntem kısıtlı veri ve kısıtlı donanım kaynakları ile denenmiştir. Yeterli destek bulunması halinde bu uygulama gerçekleştirilecektir. Üçüncü olarak, daha güçlü sınıflandırıcı algoritmaları ile başarımlar artırılabilir. Bu yaklaşımlar denenirken başka önemli bir nokta ise sistem hızıdır. Aynı performansı veya aynı performansa yakın sonuçlar veren daha hızlı algoritmalar genellikle daha çok tercih edilmektedir.

## KAYNAKLAR

- [1] Yanco, H. A. (1998). Wheelchair: A robotic wheelchair system: Indoor navigation and user interface, *In Assistive technology and artificial intelligence*, ss. 256-268, Berlin
- [2] Fod, A., Howard, A. ve Mataric, M. A. J. (2002). A laser-based people tracker, *In Proceedings 2002 IEEE International Conference on Robotics and Automation*, ss. 3024-3029, Washington, DC, USA
- [3] Pineau, J., Montemerlo, M., Pollack, M., Roy, N. ve Thrun, S. (2003). Towards robotic assistants in nursing homes: Challenges and results, *Robotics and autonomous systems*, cilt 42(3/4), ss. 271-281
- [4] Thrun, S., Bennewitz, M., Burgard, W., Cremers, A. B., Dellaert, F., Fox, D. and Schulz, D. (1999). MINERVA: A second-generation museum tour-guide robot, *In Proceedings 1999 IEEE International Conference on Robotics and Automation (Cat. No. 99CH36288C)*, Detroit, MI, USA
- [5] Chu, S., Narayanan, S., Kuo, C. C. J., ve Mataric, M. J. (2006). Where am I? Scene recognition for mobile robots using audio features, *In ICME*, ss. 885-888, Toronto
- [6] Huang, J. (2002). Spatial auditory processing for a hearing robot, *In ICME*, Lausanne, Switzerland
- [7] Itakura, F. (1975). Minimum prediction residual principle applied to speech recognition, *IEEE Transactions on Acoustics, Speech, and Signal Processing*, cilt 23(1), ss. 67-72
- [8] Furui, S. (1986). Speaker-independent isolated word recognition using dynamic features of speech spectrum, *IEEE Transactions on Acoustics, Speech, and Signal Processing*, cilt 34(1), ss. 52-59
- [9] Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J. ve Ekelid, M. (1995). Speech recognition with primarily temporal cues, *Science*, cilt 270(5234), ss. 303-304
- [10] Graves, A., Mohamed, A. R., ve Hinton, G. (2013). Speech recognition with deep recurrent neural networks, *In 2013 ICASSP*, ss. 6645-6649, Vancouver, Canada
- [11] Souli, S. ve Lachiri, Z. (2011). Environmental sounds classification based on visual features, *In Iberoamerican Congress on Pattern Recognition*, ss. 459-466, Berlin, Heidelberg
- [12] Mantyjarvi, J., Huuskonen, P. ve Himberg, J. (2002). Collaborative context determination to support mobile terminal applications, *IEEE Wireless Communications*, cilt 9(5), ss. 39-45
- [13] Waibel, A., Steusloff, H. ve Stiefelhagen, R. (2004). CHIL-Computers in the human interaction loop, 5th Intern, *In Workshop on Image Analysis for Multimedia Interactive Services*, Lisbon, Portugal
- [14] Wang, J. C., Lee, Y. S., Lin, C. H., Siahaan, E. ve Yang, C. H. (2015). Robust environmental sound recognition with fast noise suppression for home automation, *IEEE Transactions on Automation Science and Engineering*, cilt 12(4), ss. 1235-1242
- [15] Ellis, D. P. ve Lee, K. (2004). Minimal-impact audio-based personal archives, *In Proceedings of the 1st ACM workshop on Continuous archival and retrieval of personal experiences*, ss. 39-47, ACM
- [16] Sehili, M. A., Lecouteux, B., Vacher, M., Portet, F., Istrate, D., Dorizzi, B. ve Boudy, J. (2012, November). Sound environment analysis in smart home, *In International Joint Conference on Ambient Intelligence*, ss. 208-223, Berlin, Heidelberg
- [17] Mozaffarian, D., Benjamin, E. J., Go, A. S., Arnett, D. K., Blaha, M. J., Cushman, M. ve Howard, V. J. (2016). Heart disease and stroke statistics-2016 update a report from the American Heart Association, *Circulation*, cilt 133(4), ss. 38-48
- [18] Scambler, T., Holbrook, J., Savic, S., McDermott, M. F. ve Peckham, D. (2018). Autoinflammatory disease in the lung, *Immunology*, cilt 154(4), ss. 563-573

- [19] Jiang, Z. ve Choi, S. (2006). A cardiac sound characteristic waveform method for in-home heart disorder monitoring with electric stethoscope, *Expert Systems with Applications*, cilt 31(2), ss. 286-298
- [20] Schmidt, S. E., Toft, E., Holst-Hansen, C., Graff, C. ve Struijk, J. J. (2008). Segmentation of heart sound recordings from an electronic stethoscope by a duration dependent Hidden-Markov model, *In 2008 Computers in Cardiology*, ss. 345-348, Bologna, Italy
- [21] Wang, H., Chen, J., Hu, Y., Jiang, Z. ve Samjin, C. (2009). Heart sound measurement and analysis system with digital stethoscope, *In 2009 2nd International Conference on Biomedical Engineering and Informatics*, ss. 1-5, Tianjin, China
- [22] Chowdhury, M. E., Khandakar, A., Alzoubi, K., Mansoor, S., M Tahir, A., Reaz, M. B. I. ve Al-Emadi, N. (2019). Real-Time Smart-Digital Stethoscope System for Heart Diseases Monitoring, *Sensors*, cilt 19(12), ss. 2780-2019
- [23] Reaz, M. B. I., Hussain, M. S. ve Mohd-Yasin, F. (2006). Techniques of EMG signal analysis: detection, processing, classification and applications, *Biological procedures online*, cilt 8(1), ss. 11-35
- [24] Linszen, W. H., Stegeman, D. F., Joosten, E. M., Notermans, S. L., van't Hof, M. A. ve Binkhorst, R. A. (1993). Variability and interrelationships of surface EMG parameters during local muscle fatigue, *Muscle & Nerve: Official Journal of the American Association of Electrodiagnostic Medicine*, cilt 16(8), ss. 849-856
- [25] Mambrito, B. ve De Luca, C. J. (1984). A technique for the detection, decomposition and analysis of the EMG signal, *Electroencephalography and clinical neurophysiology*, cilt 58(2), ss. 175-188
- [26] Allison, G. T., Godfrey, P. ve Robinson, G. (1998). EMG signal amplitude assessment during abdominal bracing and hollowing, *Journal of Electromyography and Kinesiology*, cilt 8(1), ss. 51-57
- [27] Hogan, N. (1976). A review of the methods of processing EMG for use as a proportional control signal, *Biomedical engineering*, cilt 11(3), ss. 81-86
- [28] Lowery, M. M., Stoykov, N. S., Taflove, A. ve Kuiken, T. A. (2002). A multiple-layer finite-element model of the surface EMG signal, *IEEE Transactions on Biomedical Engineering*, cilt 49(5), ss. 446-454
- [29] Gut, R. ve Moschytz, G. S. (2000). High-precision EMG signal decomposition using communication techniques, *IEEE transactions on signal processing*, cilt 48(9), ss. 2487-2494
- [30] Szu, H. H., Telfer, B. A. ve Kadambe, S. L. (1992). Neural network adaptive wavelets for signal representation and classification, *Optical Engineering*, cilt 31(9), ss. 1907-1917
- [31] Fristrup, K. M. ve Watkins, W. A. (1993). Marine Animal Sound Classification (No. WHOI-94-13), *Woods Hole Oceanographic Institution Ma*
- [32] Morrison, D. F. (1976). *Multivariate statistical methods*
- [33] Clark, L. A., Pregibon, D., Chambers, J. ve Hastie, T. (1992). Statistical models in S, *chapter Tree-Based Models*, ss. 377-419
- [34] Ashiya, T., Hagiwara, M. ve Nakagawa, M. (1996). IOSES: An indoor observation system based on environmental sounds recognition using a neural network, *IEEE Transactions on Electronics, Information and Systems*, cilt 116(3), ss. 341-349
- [35] Nooralahiyan, A. Y., Dougherty, M., McKeown, D. ve Kirby, H. R. (1997). A field trial of acoustic signature analysis for vehicle classification, *Transportation Research Part C: Emerging Technologies*, cilt 5(3/4), ss. 165-177

- [36] Zhang, T. ve Kuo, C. C. J. (1998). Content-based classification and retrieval of audio, *In Advanced Signal Processing Algorithms, Architectures, and Implementations VIII*, cilt 3461, ss. 432-443.
- [37] Liu, M., ve Wan, C. (2001). A study on content-based classification and retrieval of audio database, *In Proceedings 2001 International Database Engineering and Applications Symposium*, ss. 339-345, Grenoble, France
- [38] Lu, L., Zhang, H. J. ve Jiang, H. (2002). Content analysis for audio classification and segmentation, *IEEE Transactions on speech and audio processing*, cilt 10(7), ss. 504-516
- [39] Toyoda, Y., Huang, J., Ding, S. ve Liu, Y. (2004). Environmental sound recognition by multilayered neural networks, *In The Fourth International Conference on Computer and Information Technology*, ss. 123-127, Wuhan, China
- [40] Wang, J. C., Wang, J. F., He, K. W. ve Hsu, C. S. (2006). Environmental sound classification using hybrid SVM/KNN classifier and MPEG-7 audio low-level descriptor, *In The 2006 IEEE International Joint Conference on Neural Network Proceedings*, ss. 1731-1735, Vancouver, Canada
- [41] Chu, S., Narayanan, S. ve Kuo, C. C. J. (2009). Environmental sound recognition with time-frequency audio features, *IEEE Transactions on Audio, Speech, and Language Processing*, cilt 17(6), ss. 1142-1158
- [42] Su, F., Yang, L., Lu, T., ve Wang, G. (2011). Environmental sound classification for scene recognition using local discriminant bases and HMM, *In Proceedings of the 19th ACM international conference on Multimedia*, ss. 1389-1392, Scottsdale Arizona, USA
- [43] Uzkent, B., Barkana, B. D. ve Cevikalp, H. (2012). Non-speech environmental sound classification using SVMs with a new set of features. *International Journal of Innovative Computing, Information and Control*, cilt 8(5), ss. 3511-3524
- [44] Wang, J. C., Lin, C. H., Chen, B. W. ve Tsai, M. K. (2013). Gabor-based nonuniform scale-frequency map for environmental sound classification in home automation, *IEEE Transactions on Automation Science and Engineering*, cilt 11(2), ss. 607-613
- [45] Krizhevsky, A., Sutskever, I. ve Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks, *In Advances in neural information processing systems*, ss. 1097-1105, Nevada
- [46] Simonyan, K. ve Zisserman, A. (2014). *Very deep convolutional networks for large-scale image recognition*, arXiv preprint arXiv:1409.1556.
- [47] Piczak, K. J. (2015). ESC: Dataset for environmental sound classification, *In Proceedings of the 23rd ACM international conference on Multimedia*, ss. 1015-1018, ACM.
- [48] Salamon, J. ve Bello, J. P. (2017). Deep convolutional neural networks and data augmentation for environmental sound classification, *IEEE Signal Processing Letters*, cilt 24(3), ss. 279-283.
- [49] Chen, Y., Guo, Q., Liang, X., Wang, J. ve Qian, Y. (2019). Environmental sound classification with dilated convolutions, *Applied Acoustics*, cilt 148, ss. 123-132.
- [50] Park, S., Mun, S., Lee, Y. ve Ko, H. (2017). Acoustic scene classification based on convolutional neural network using double image features, *In Proceedings of Detection and Classification of Acoustic Scenes and Events Workshop*, ss. 1-5, Munich, Germany
- [51] Lehner, B., Eghbal-Zadeh, H., Dorfer, M., Korzeniowski, F., Koutini, K. ve Widmer, G. (2017). Classifying short acoustic scenes with I-vectors and CNNs: Challenges and optimizations for the 2017 DCASE ASC task. *IEEE AASP Challenge on Detection and Classification of Acoustic Scene and Events (DCASE)*, Munich, Germany

- [52] Hyder, R., Ghaffarzadegan, S., Feng, Z. ve Hasan, T. (2017). Buet bosch consortium (b2c) acoustic scene classification systems for dcase 2017 challenge, *Detection and Classification of Acoustic Scenes and Events (DCASE)*, Munich, Germany
- [53] Weiping, Z., Jiantao, Y., Xiaotao, X., Xiangtao, L. ve Shaohu, P. (2017). Acoustic scene classification using deep convolutional neural network and multiple spectrograms fusion, *In Detection and Classification of Acoustic Scenes and Events 2017 Workshop (DCASE2017)*, Munich, Germany
- [54] Han, Y., Park, J. ve Lee, K. (2017). Convolutional neural networks with binaural representations and background subtraction for acoustic scene classification, *Detection and Classification of Acoustic Scenes and Events (DCASE)*, ss. 1-5, Munich, Germany
- [55] Mun, S., Park, S., Han, D. K., ve Ko, H. (2017). Generative adversarial network based acoustic scene training set augmentation and selection using SVM hyper-plane, *Proc. DCASE*, ss. 93-97, Munich, Germany
- [56] Barschdorff, D., Ester, S., Dorsel, T. ve Most, E. (1990). Neural network based multi sensor heart sound analysis, *Proceedings Computers in Cardiology*, ss. 303-306
- [57] Zhang, W., Han, J. ve Deng, S. (2017). Heart sound classification based on scaled spectrogram and partial least squares regression, *Biomedical Signal Processing and Control*, cilt 32, ss. 20-28
- [58] Hamidi, M., Ghassemian, H. ve Imani, M. (2018). Classification of heart sound signal using curve fitting and fractal dimension, *Biomedical Signal Processing and Control*, cilt 39, ss. 351-359
- [59] Randhawa, S. K. ve Singh, M. (2015). Classification of heart sound signals using multi-modal features, *Procedia Computer Science*, cilt 58, ss. 165-171
- [60] Varghees, V. N. ve Ramachandran, K. I. (2017). Effective heart sound segmentation and murmur classification using empirical wavelet transform and instantaneous phase for electronic stethoscope, *IEEE Sensors Journal*, cilt 17(12), ss. 3861-3872
- [61] Rubin, J., Abreu, R., Ganguli, A., Nelaturi, S., Matei, I. ve Sricharan, K. (2017). *Recognizing abnormal heart sounds using deep learning*, arXiv preprint arXiv:1707.04642
- [62] Dominguez-Morales, J. P., Jimenez-Fernandez, A. F., Dominguez-Morales, M. J. ve Jimenez-Moreno, G. (2017). Deep neural networks for the recognition and classification of heart murmurs using neuromorphic auditory sensors, *IEEE transactions on biomedical circuits and systems*, cilt 12(1), ss. 24-34
- [63] Chen, T. E., Yang, S. I., Ho, L. T., Tsai, K. H., Chen, Y. H., Chang, Y. F. ve Wu, C. C. (2016). S1 and S2 heart sound recognition using deep neural networks, *IEEE Transactions on Biomedical Engineering*, cilt 64(2), ss. 372-380
- [64] Potes, C., Parvaneh, S., Rahman, A. ve Conroy, B. (2016). Ensemble of feature-based and deep learning-based classifiers for detection of abnormal heart sounds, *In 2016 Computing in Cardiology Conference (CinC)*, ss. 621-624, Canada
- [65] Tschannen, M., Kramer, T., Marti, G., Heinzmann, M. ve Wiatowski, T. (2016). Heart sound classification using deep structured features, *In 2016 Computing in Cardiology Conference (CinC)*, ss. 565-568, Canada
- [66] Serbes, G., Sakar, C. O., Kahya, Y. P. ve Aydin, N. (2013). Pulmonary crackle detection using time–frequency and time–scale analysis, *Digital Signal Processing*, cilt 23(3), ss. 1012-1021
- [67] Jin, F., Sattar, F. ve Goh, D. Y. (2014). New approaches for spectro-temporal feature extraction with applications to respiratory sound classification, *Neurocomputing*, cilt 123, ss. 362-371

- [68] Bahoura, M. (2009). Pattern recognition methods applied to respiratory sounds classification into normal and wheeze classes, *Computers in biology and medicine*, cilt 39(9), ss. 824-843
- [69] Reyes, B. A., Charleston-Villalobos, S., González-Camarena, R. ve Aljama-Corrales, T. (2014). Assessment of time–frequency representation techniques for thoracic sounds analysis, *Computer methods and programs in biomedicine*, cilt 114(3), ss. 276-290
- [70] Orjuela-Cañón, A. D., Gómez-Cajas, D. F. ve Jiménez-Moreno, R. (2014, November). Artificial neural networks for acoustic lung signals classification, *In Iberoamerican Congress on Pattern Recognition*, ss. 214-221, Cham
- [71] Maruf, S. O., Azhar, M. U., Khawaja, S. G. ve Akram, M. U. (2015). Crackle separation and classification from normal Respiratory sounds using Gaussian Mixture Model, *In 2015 IEEE 10th International Conference on Industrial and Information Systems (ICIIS)*, ss. 267-271, Peradeniya
- [72] Bardou, D., Zhang, K. ve Ahmad, S. M. (2018). Lung sounds classification using convolutional neural networks, *Artificial intelligence in medicine*, cilt 88, ss. 58-69
- [73] Chambres, G., Hanna, P., ve Desainte-Catherine, M. (2018). Automatic Detection of Patient with Respiratory Diseases Using Lung Sound Analysis, *In 2018 International Conference on Content-Based Multimedia Indexing (CBMI)*, ss. 1-6, La Rochelle, France
- [74] Jakovljević, N. ve Lončar-Turukalo, T. (2018). Hidden Markov model based respiratory sound classification, *In Precision Medicine Powered by pHealth and Connected Health*, ss. 39-43, Singapore
- [75] Serbes, G., Ulukaya, S. ve Kahya, Y. P. (2018). An automated lung sound preprocessing and classification system based on spectral analysis methods, *In Precision Medicine Powered by pHealth and Connected Health*, ss. 45-49, Singapore
- [76] Bekey, G. A., Chang, C. W., Perry, J. ve Hoffer, M. M. (1977). Pattern recognition of multiple EMG signals applied to the description of human gait, *Proceedings of the IEEE*, cilt 65(5), ss. 674-681
- [77] Graupe, D., Salahi, J. ve Kohn, K. H. (1982). Multifunctional prosthesis and orthosis control via microcomputer identification of temporal pattern differences in single-site myoelectric signals. *Journal of Biomedical Engineering*, cilt 4(1), ss. 17-22
- [78] Lee, S. ve Saridis, G. (1984). The control of a prosthetic arm by EMG pattern recognition, *IEEE Transactions on automatic control*, cilt 29(4), ss. 290-302
- [79] Uchida, M. (1992). The Control Method for the Robot Hand Based on the Fussy Theory, *J. Robotics & Mechatronics*, cilt 4(4), s. 262
- [80] Hudgins, B., Parker, P. ve Scott, R. N. (1993). A new strategy for multifunction myoelectric control, *IEEE Transactions on Biomedical Engineering*, cilt 40(1), ss. 82-94
- [81] Chan, F. H., Yang, Y. S., Lam, F. K., Zhang, Y. T. ve Parker, P. A. (2000). Fuzzy EMG classification for prosthesis control, *IEEE transactions on rehabilitation engineering*, 8(3), ss. 305-311
- [82] Lucas, M. F., Gaufriau, A., Pascual, S., Doncarli, C. ve Farina, D. (2008). Multi-channel surface EMG classification using support vector machines and signal-based wavelet optimization, *Biomedical Signal Processing and Control*, cilt 3(2), ss. 169-174
- [83] Phinyomark, A., Phukpattaranont, P. ve Limsakul, C. (2012). Feature reduction and selection for EMG signal classification, *Expert systems with applications*, cilt 39(8), ss. 7420-7431
- [84] Subasi, A. (2013). Classification of EMG signals using PSO optimized SVM for diagnosis of neuromuscular disorders, *Computers in biology and medicine*, cilt 43(5), ss. 576-586

- [85] Atzori, M., Cognolato, M. ve Müller, H. (2016). Deep learning with convolutional neural networks applied to electromyography data: A resource for the classification of movements for prosthetic hands, *Frontiers in neurorobotics*, cilt 10(9), ss. 102-107
- [86] Xia, P., Hu, J. ve Peng, Y. (2018). EMG-based estimation of limb movement using deep learning with recurrent convolutional neural networks, *Artificial organs*, cilt 42(5), ss. 67-77
- [87] Phinyomark, A. ve Scheme, E. (2018). EMG pattern recognition in the era of big data and deep learning, *Big Data and Cognitive Computing*, 2(3), s. 21
- [88] Challis, R. E. ve Kitney, R. I. (1990). Biomedical signal processing (in four parts), *Medical and Biological Engineering and Computing*, cilt 28(6), ss. 509-524
- [89] Challis, R. E. ve Kitney, R. I. (1991). Biomedical signal processing (in four parts). Part 2. The frequency transforms and their inter-relationships, *Medical and biological engineering and computing*, cilt 29(1), ss. 1-17
- [90] Challis, R. E. ve Kitney, R. I. (1991). Biomedical signal processing (in four parts), *Medical and Biological Engineering and Computing*, cilt 29(1), ss. 1-17
- [91] Stephanopoulos, G. ve Han, C. (1996). Intelligent systems in process engineering: A review. *Computers & Chemical Engineering*, cilt 20(6/7), ss. 743-791
- [92] Mahmoud, M., Jiang, J. ve Zhang, Y. (2003). Active fault tolerant control systems: stochastic analysis and synthesis, *Springer Science & Business Media*, cilt 21(3), ss. 673-701
- [93] Webb, A. R. (2003). *Statistical pattern recognition*, John Wiley & Sons
- [94] Manolakis, D. G., Ingle, V. K. ve Kogon, S. M. (2000). *Statistical and adaptive signal processing: spectral estimation, signal modeling, adaptive filtering, and array processing*, 656, Boston
- [95] Mukhopadhyay, S. ve Sircar, P. (1997). Parametric modelling of non-stationary signals: A unified approach, *Signal Processing*, cilt 60(2), ss. 135-152
- [96] Cohen, L. (1995). *Time-frequency analysis*, 778, Prentice hall
- [97] Frigo, M. ve Johnson, S. G. (1998). FFTW: An adaptive software architecture for the FFT. In *Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP'98 (Cat. No. 98CH36181)*, ss. 1381-1384, Seattle, WA, USA
- [98] Stankovic, L. (1994). An analysis of some time-frequency and time-scale distributions. *In Annales des telecommunications*, cilt 49(9), s.505
- [99] Gröchenig, K. (2013). *Foundations of time-frequency analysis*, Springer Science & Business Media
- [100] Mitra, S. K. ve Kuo, Y. (2006). *Digital signal processing: a computer-based approach (Vol. 2)*. New York: McGraw-Hill
- [101] Schafer, R. W. ve Oppenheim, A. V. (1989). *Discrete-time signal processing*, Englewood Cliffs
- [102] Sejdic, E. ve Djurovic, I. (2008). Quantitative performance analysis of scalogram as instantaneous frequency estimator, *IEEE Transactions on Signal Processing*, cilt 56(8), ss. 3837-3845
- [103] Ojeda-Aguirre, N. A., Garcia-Perez, A., Romero-Troncoso, R. D. J., Morinigo-Sotelo, D., Duque-Perez, O. ve Camarena-Martinez, D. (2019). Reassigned Short Time Fourier Transform and K-means Method for Diagnosis of Broken Rotor Bar Detection in VSD-fed Induction Motors, *Advances in Electrical and Computer Engineering*, cilt 19(2), ss. 61-69
- [104] LeCun, Y., Kavukcuoglu, K., ve Farabet, C. (2010). Convolutional networks and applications in vision, *In ISCAS*, ss. 253-256, Paris, France
- [105] LeCun, Y., Bottou, L., Bengio, Y. ve Haffner, P. (1998). Gradient-based learning applied to document recognition, *Proceedings of the IEEE*, cilt 86(11), ss. 2278-2324

- [106] Simard, P. Y., Steinkraus, D. ve Platt, J. C. (2003). Best practices for convolutional neural networks applied to visual document analysis, In *Icdar*.
- [107] Krizhevsky, A., Sutskever, I. ve Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks, *In Advances in neural information processing systems*, ss. 1097-1105
- [108] Goodfellow, I., Bengio, Y. ve Courville, A. (2016). *Deep learning*, MIT press
- [109] Alpaydm, E. (2013). *Yapay öğrenme*, Boğaziçi Üniversitesi Yayinevi
- [110] Hochreiter, S. (1998). The vanishing gradient problem during learning recurrent neural nets and problem solutions, *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, cilt 6(2), ss. 107-116
- [111] Dahl, G. E., Sainath, T. N. ve Hinton, G. E. (2013). Improving deep neural networks for LVCSR using rectified linear units and dropout, *In 2013 IEEE international conference on acoustics, speech and signal processing*, ss. 8609-8613, Vancouver, Canada
- [112] Ioffe, S., ve Szegedy, C. (2015). *Batch normalization: Accelerating deep network training by reducing internal covariate shift*, arXiv preprint arXiv:1502.03167
- [113] Srivastava, N. (2013), Improving neural networks with dropout, PhD Thesis, University of Toronto
- [114] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D. ve Rabinovich, A. (2015). Going deeper with convolutions, *In Proceedings of the IEEE conference on computer vision and pattern recognition*, 1-9, San Juan, Puerto Rico, USA
- [115] Zang, F. ve Zhang, J. S. (2011). Softmax discriminant classifier, *In 2011 Third International Conference on Multimedia Information Networking and Security*, ss. 16-19, Shanghai, China
- [116] Sutskever, I., Martens, J., Dahl, G. ve Hinton, G. (2013). On the importance of initialization and momentum in deep learning, *In ICML*, ss. 1139-1147, Atlanta, USA
- [117] Jacobs, R. A. (1988). Increased rates of convergence through learning rate adaptation, *Neural networks*, cilt 1(4), ss. 295-307
- [118] Duchi, J., Hazan, E. ve Singer, Y. (2011). Adaptive subgradient methods for online learning and stochastic optimization, *Journal of Machine Learning Research*, cilt 12, ss. 2121-2159
- [119] Tieleman, T. ve Hinton, G. (2012). Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude, *COURSERA: Neural networks for machine learning*, cilt 4(2), ss. 26-31
- [120] Kingma, D. P. ve Ba, J. (2014). *Adam: A method for stochastic optimization*, arXiv preprint arXiv:1412.6980
- [121] LeCun, Y., Bottou, L., Bengio, Y. ve Haffner, P. (1998). Gradient-based learning applied to document recognition, *Proceedings of the IEEE*, cilt 86(11), ss. 2278-2324
- [122] He, K., Zhang, X., Ren, S., ve Sun, J. (2016). Deep residual learning for image recognition. *In Proceedings of the IEEE conference on computer vision and pattern recognition*, San Juan, USA
- [123] Hinton, G. E., Sejnowski, T. J. ve Poggio, T. A. (1999). *Unsupervised learning: foundations of neural computation*, MIT press
- [124] Cover, T. ve Hart, P. (1967). Nearest neighbor pattern classification, *IEEE transactions on information theory*, cilt 13(1), ss. 21-27
- [125] Cortes, C. ve Vapnik, V. (1995). Support-vector networks, *Machine learning*, cilt 20(3), ss. 273-297
- [126] Ba, J. L., Kiros, J. R. ve Hinton, G. E. (2016). *Layer normalization*. arXiv preprint arXiv:1607.06450
- [127] Jain, A., Nandakumar, K. ve Ross, A. (2005). Score normalization in multimodal biometric systems, *Pattern recognition*, cilt 38(12), ss. 2270-2285

- [128] Bentley, P., Nordehn, G., Coimbra, M., Mannor, S. ve Getz, R. (2011). *The pascal classifying heart sounds challenge 2011 (chsc2011) results*
- [129] Gomes, E. F. ve Pereira, E. (2012). Classifying heart sounds using peak location for segmentation and feature construction, *In Workshop Classifying Heart Sounds*, La Palma, Canary Islands, 10.
- [130] Deng, Y., ve Bentley, P. J. (2012). A robust heart sound segmentation and classification algorithm using wavelet decomposition and spectrogram, *In Workshop Classifying Heart Sounds*, La Palma, Canary Island
- [131] Avendano-Valencia, L. D.; Godino-Llorente, J. I.; Blanco-Velasco, M.; and Castellanos-Dominguez, G. (2010). Feature extraction from parametric time–frequency representations for heart murmur detection. *Annals of Biomedical Engineering*, cilt 38(8), ss. 2716-2732.
- [132] Oliveira, S. C., Gomes, E. F. ve Jorge, A. M. (2014.). Heart sounds classification using motif based segmentation, *In Proceedings of the 18th International Database Engineering & Applications Symposium*, ss. 370-371, Porto, Portugal
- [133] Deng, S. W. ve Han, J. Q. (2016). Towards heart sound classification without segmentation via autocorrelation feature and diffusion maps, *Future Generation Computer Systems*, cilt 60, ss. 13-21
- [134] Zhang, W., Han, J., ve Deng, S. (2017). Heart sound classification based on scaled spectrogram and tensor decomposition, *Expert Systems with Applications*, cilt 84, ss. 220-231
- [135] Dua, D. ve Graff, C. (2019). *UCI Machine Learning Repository*, Irvine, CA: University of California, School of Information and Computer Science.
- [136] Abdullah, A. A., Subasi, A. ve Qaisar, S. M. (2017). Surface EMG signal classification by using WPD and ensemble tree classifiers, *In CMBEBIH 2017*, ss. 475-481, Singapore
- [137] Sukumar, N., Taran, S. ve Bajaj, V. (2018). Physical actions classification of surface EMG signals using VMD, *In 2018 ICCSP*, ss. 705-709, Chennai, India
- [138] Šahinbegović, H., Mušić, L., Alić, B. (2017). Distinguishing physical actions using an artificial neural network, *In 2017 XXVI ICAT*, ss. 1-5, Sarajevo, Bosnia and Herzegovina
- [139] Rocha, B. M., Filos, D., Mendes, L., Serbes, G., Ulukaya, S., Kahya, Y. P. ve Kaimakamis, E. (2019). An open access database for the evaluation of respiratory sound classification algorithms, *Physiological measurement*, Cilt 40(3), ss. 35001
- [140] Chambres, G., Hanna, P. ve Desainte-Catherine, M. (2018). Automatic Detection of Patient with Respiratory Diseases Using Lung Sound Analysis, *In 2018 International Conference on Content-Based Multimedia Indexing (CBMI)*, ss.1-6, La Rochelle, France
- [141] Jakovljević, N. ve Lončar-Turukalo, T. (2018). Hidden Markov model based respiratory sound classification, *In Precision Medicine Powered by pHealth and Connected Health*, ss. 39-43, Singapore
- [142] Serbes, G., Ulukaya, S., ve Kahya, Y. P. (2018). An automated lung sound preprocessing and classification system based on spectral analysis methods, *In Precision Medicine Powered by pHealth and Connected Health*, ss. 45-49, Singapore
- [143] Salamon, J., Jacoby, C. ve Bello, J. P. (2014). A dataset and taxonomy for urban sound research, *In Proceedings of the 22nd ACM international conference on Multimedia*, ss. 1041-1044, ACM
- [144] Stowell, D., Giannoulis, D., Benetos, E., Lagrange, M. ve Plumbley, M. D. (2015). Detection and classification of acoustic scenes and events, *IEEE Transactions on Multimedia*, cilt 17(10), ss. 1733-1746.
- [145] Piczak, K. J. (2015). Environmental sound classification with convolutional neural networks, *In 2015 IEEE 25th International Workshop on MLSP*, ss. 1-6, Boston, USA

- [146] Aytar, Y., Vondrick, C. ve Torralba, A. (2016). Soundnet: Learning sound representations from unlabeled video, *In Advances in neural information processing systems*, Barcelona, Spain
- [147] Salamon, J. ve Bello, J. P. (2015). Unsupervised feature learning for urban sound classification, In 2015 IEEE International Conference on Acoustics, *In 2015 ICASSP*, Brisbane, Australia
- [148] Rakotomamonjy, A ve Gasso, G. (2014). Histogram of gradients of time–frequency representations for audio scene classification, *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 23, sayı 1, ss. 142-153
- [149] Roma, G., Nogueira, W., Herrera, P. (2013). Recurrence quantification analysis features for environmental sound recognition. *In 2013 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, ss. 1-4
- [150] Mesaros, A., Heittola, T., Dikmen, O., Virtanen, T. (2015). Sound event detection in real life recordings using coupled matrix factorization of spectral representations and class activity annotations, *In 2015 ICASSP*, Brisbane, Australia
- [151] Mesaros, A., Diment, A., Elizalde, B., Heittola, T., Vincent, E. ve Raj, B., Virtanen, T. (2019). Sound event detection in the DCASE 2017 challenge, *IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP)*, cilt 27(6), ss. 992-1006
- [152] Ho, T. K. (1998). Nearest neighbors in random subspaces, *In Joint IAPR International Workshops on Statistical Techniques in Pattern Recognition (SPR) and Structural and Syntactic Pattern Recognition (SSPR)*, ss. 640-648, Berlin
- [153] Mun, S., Park, S., Han, D. K., ve Ko, H. (2017). Generative adversarial network based acoustic scene training set augmentation and selection using SVM hyper-plane, *DCASE*, Munich
- [154] Han, Y., Park, J., ve Lee, K. (2017). Convolutional neural networks with binaural representations and background subtraction for acoustic scene classification, *DCASE*, Munich
- [154] Hyder, R., Ghaffarzadegan, S., Feng, Z., Hasan, T. (2017). Buet bosch consortium (b2c) acoustic scene classification systems for dcase 2017 challenge, *DCASE*, Munich
- [155] Lehner, B., Eghbal-Zadeh, H., Dorfer, M., Korzeniowski, F., Koutini, K., Widmer, G. (2017). Classifying short acoustic scenes with I-vectors and CNNs: Challenges and optimisations for the 2017 DCASE ASC task, *DCASE*, Munich
- [156] Demir, F., Şengür, A., Çavaş, M. (2018). Heart Sounds Classification with Deep Features and Support Vector Machines, *In 2018 International Conference on Artificial Intelligence and Data Processing (IDAP)*, Malatya, Turkey
- [157] Demir, F., Şengür, A., Bajaj, V., Polat, K. (2019). Towards the classification of heart sounds based on convolutional deep neural network, *Health information science and systems*, cilt 7(1), ss. 1-16.
- [158] Demir, F., Şengür, A., Bajaj, V., Polat, K. (2019). Convolutional neural networks based efficient approach for classification of lung diseases, *Health information science and systems*, cilt 8(1), ss. 1-4.
- [159] Demir, F., Bajaj, V., Ince, M. C., Taran, S., Şengür, A. (2019). Surface EMG signals and deep transfer learning-based physical action classification, *Neural Computing and Applications*, cilt 31(12), ss. 8455-8462

# ÖZGEÇMİŞ

Fatih DEMİR

## KİŞİSEL BİLGİLER

**Doğum Yeri** : Elazığ  
**Doğum Yılı** : 1985  
**Uyruğu** : T.C.  
**Adres** : Fırat Üniversitesi Teknik Bilimler MYO  
**E-posta** : fatidemir@firat.edu.tr  
**Yabancı Diller** : İngilizce (Düzey: Orta)

## EĞİTİM BİLGİLERİ

**Yüksek Lisans** : Üç Fazlı Evriciyle Beslenen Doğrusal Olmayan Yüklerde Harmonik Eliminasyonu  
Fırat Üniversitesi, Fen Bilimleri Enstitüsü, Elektrik-Elektronik Müh. Anabilim Dalı, 2010  
Danışman: Doç. Dr. Mehmet Özdemir  
**Lisans** : Fırat Üniversitesi, Mühendislik Fakültesi, Elektrik-Elektronik Mühendisliği Bölümü, 2007.  
**Lise** : Balakgazi Lisesi, Elazığ, 2003.

## ARAŞTIRMA DENEYİMİ

✓ C, C++, MATLAB  
✓

## AKADEMİK FAALİYETLER

### Makaleler:

- Demir, F., Bajaj, V., Ince, M. C., Taran, S., Şengür, A. (2019). Surface EMG signals and deep transfer learning-based physical action classification, *Neural Computing and Applications*, cilt 31(12), ss. 8455-8462
- Ahmad, S., Agrawal, S., Joshi, S., Taran, S., Bajaj, V., Demir, F. ve Sengur, A. (2020). Environmental sound classification using optimum allocation sampling based empirical mode decomposition, *Physica A: Statistical Mechanics and its Applications*, cilt 537(122613)
- Demir, F., Şengür, A., Bajaj, V., Polat, K. (2019). Convolutional neural networks based efficient approach for classification of lung diseases, *Health information science and systems*, cilt 8(1), ss. 1-4
- Demir, F., Şengür, A., Bajaj, V. ve Polat, K. (2019). Towards the classification of heart sounds based on convolutional deep neural network, *Health information science and systems*, cilt 7(1), ss. 1-9

### Bildiriler:

- Demir, F., Şengür, A., Çavaş, M. (2018). Heart Sounds Classification with Deep Features and Support Vector Machines, *In 2018 International Conference on Artificial Intelligence and Data Processing (IDAP)*, 1-5, Malatya, Turkey