



T.C.

TOKAT GAZİOSMANPAŞA ÜNİVERSİTESİ

LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

**ÇEVRESEL SESLERİN EVRİŞİMSEL SİNİR AĞLARI
İLE SINIFLANDIRILMASI**

Yalçın DİNÇER

YÜKSEK LİSANS TEZİ

Mekatronik Mühendisliği Anabilim Dalı

TOKAT

2021

Her Hakkı Saklıdır

ETİK SÖZLEŞME

Tokat Gaziosmanpaşa Üniversitesi Lisansüstü Eğitim Enstitüsü tez yazım kılavuzuna göre, Dr. Öğr. Üyesi Özkan İNİK danışmanlığında hazırlamış olduğum “Çevresel Seslerin Evrişimsel Sinir Ağları ile Sınıflandırılması” adlı Yüksek Lisans tezinin bilimsel etik değerlere ve kurallara uygun, özgün bir çalışma olduğunu, aksinin tespit edilmesi halinde her türlü yasal yaptırımını kabul edeceğimi beyan ederim.

28/12/2021

Yalçın DİNÇER



ÖNSÖZ

Yüksek lisans ders ve tez aşamaları boyunca her türlü yardımı, bilgiyi ve ilgiyi gösteren değerli fikirleriyle tecrübelerimin artmasını sağlayan, güven ve desteğini hiçbir zaman esirgemeyen, saygı değer danışmanım Dr. Öğr. Üyesi Özkan İNİK hocama en içten teşekkürlerimi sunarım.

Yalçın DİNÇER
TOKAT-2021



ÖZET

ÇEVRESEL SESLERİN EVRİŞİMSEL SİNİR AĞLARI İLE SINIFLANDIRILMASI

Dinçer, Yalçın

Yüksek Lisans, Mekatronik Mühendisliği Ana Bilim Dalı

Tez Danışmanı: Dr. Öğr. Üyesi Özkan İnik

Aralık 2021, xv + 80 sayfa

Çevresel faaliyetler canlı veya cansız varlıklar tarafından oluşmaktadır. Bu faaliyetlerin sonuçlarını temsil edebilecek ve aynı zamanda ortam hakkında bilgi verebilecek ses verisi önem kazanmaktadır. Kentlerde meydana gelen gürültü kirliliği, trafik problemleri, güvenlik sistemleri, akıllı takip sistemleri, sağlık hizmetleri, yerel hizmetler gibi faaliyetlerin işleyişini ve temel bilgilerini elde etmek için ses verisinden faydalanılmaktadır. Bu anlamda Çevresel Seslerin Sınıflandırılması (ÇSS) kritik önem kazanmaktadır. Artan veri miktarı ve çözümlemedeki zaman kısıtlamalarından dolayı anlık otomatik olarak seslerin tanımlanmasını sağlayan yeni ve güçlü yapay zekâ yöntemlerine ihtiyaç duyulmaktadır. Bu alanda son yıllarda Derin Öğrenme (DÖ) mimarilerinin birçok farklı alanda elde etmiş olduğu yüksek performanstan dolayı ÇSS işlemi DÖ mimarileri ile yapılması amaçlanmıştır.

Yapılan bu tez kapsamında üç farklı ÇSS veri setinin sınıflandırılması için Derin Öğrenmenin temel mimarisi olan Evrişimsel Sinir Ağları (ESA) modelleri tasarlanmıştır. Her bir veri seti için özgün olarak tasarlanan birden fazla ESA modelleri içerisinde en yüksek doğruluk oranına sahip ESA modelleri elde edilmiştir. Bu veri setleri sırasıyla ESC10, ESC50 ve UrbanSound8K veri setleridir. Bu veri setlerindeki ses kayıtları 32x32x3 ve 224x224x3 boyutuna sahip görüntü formatına çevrilmiştir. Böylelikle toplamda altı farklı görüntü formatında veri seti elde edilmiştir.

Bu veri setlerinin sınıflandırılması için geliştirilen özgün ESA modelleri sırasıyla, ESC10_ESA32, ESC10_ESA224, ESC50_ESA32, ESC50_ESA224, URBANSOUND8K_ESA32 ve URBANSOUND8K_ESA224 olarak isimlendirilmiştir. Bu modeller veri setleri üzerinde 10-Kat Çapraz Doğrulama (10-Fold Cross Validation) yapılarak eğitilmiştir. Elde edilen sonuçlarda, ESC10_ESA32, ESC10_ESA224, ESC50_ESA32, ESC50_ESA224, URBANSOUND8K_ESA32 ve URBANSOUND8K_ESA224 modellerinin ortalama doğruluk oranları sırasıyla %80.75, %82.25, %54.55, %72.15, %88.60 ve %84.33 olarak elde edilmiştir. Elde edilen sonuçlar aynı veri setleri üzerinde literatürde yapılan diğer temel çalışmalarla

karşılaştırıldığında önerilen modellerin ESC10 ve UrbanSound8K veri setinde daha iyi sonuçlar elde ettiği görülmüştür. ESC50 veri setinde ise bir çalışma haricinde diğer çalışmalardan daha iyi olduğu anlaşılmıştır.

Anahtar Kelimeler: Derin Öğrenme (DÖ), Evrişimsel Sinir Ağı (ESA), Çevresel Ses Sınıflandırılması (ÇSS), Görüntü Sınıflandırma, ESC10, ESC50, UrbanSound8K.



ABSTRACT

CLASSIFICATION OF ENVIRONMENTAL SOUNDS WITH CONVOLUTIONAL NEURAL NETWORKS

Dinçer, Yalçın

Master's Thesis, Department Of Mechatronic Engineering

Advisor: Asst. Prof. Dr. Özkan İnik

December 2021, xv + 80 pages

The environmental activities consist of living or non-living things. The sound data that can represent the results of these activities and also provide information about the environment gains importance. Sound data is used to obtain basic information and the operation of activities such as noise pollution, traffic problems, security systems, smart tracking systems, health services, local services in cities. In this sense, Environmental Sounds Classification (ESC) gains critical importance. Due to the increasing amount of data and time constraints in analysis, new and strong artificial intelligence methods that enable automatic recognition of sound are needed. Due to the high performance that Deep Learning (DL) architectures have achieved in many different areas in recent years, it is aimed to perform ESC process with DL architectures.

Within the scope of this thesis, the Convolutional Neural Networks (CNN) models, which are the basic architecture of Deep Learning, have been designed for the classification of three different CNN dataset. CNN models that have the highest accuracy have been obtained among more than one CNN models designed originally for each dataset. These data sets are ESC10, ESC50 and UrbanSound8K data sets, respectively. The sound recordings in these datasets have been converted to image formats that has 32x32x3 and 224x224x3 dimensions. Thus, dataset that a total of six different image format were obtained.

The original CNN models developed to classify these data sets are named as ESC10_CNN32, ESC10_CNN224, ESC50_CNN32, ESC50_CNN224, URBANSOUND8K_CNN32 and URBANSOUND8K_CNN224, respectively. These models have been trained by performing 10-fold Cross Validation on the datasets. In the results obtained, the average accuracy rates of ESC10_CNN32, ESC10_CNN224, ESC50_CNN32, ESC50_CNN224, URBANSOUND8K_CNN32 and URBANSOUND8K_CNN224 models were found to be 80.75%, 82.25%, 54.55%, 72.15%, 88.60% and 84.63%, respectively. When the results obtained are compared with other studies in the literature on the same data sets, it was seen that the proposed

models achieved better results in the ESC10 and UrbanSound8K data sets. In the ESC50 dataset, it was found to be better than other studies, except for one study.

Keywords: Deep Learning (DL), Convolutional Neural Network (CNN), Environmental Sound Classification (ESC), Image Classification, ESC10, ESC50, UrbanSound8K.



İÇİNDEKİLER

ETİK SÖZLEŞME	II
JÜRİ KABUL VE ONAY.....	III
ÖNSÖZ	IV
ÖZET	V
ABSTRACT.....	VII
İÇİNDEKİLER	IX
SİMGELER VE KISALTMALAR	XI
ŞEKİL LİSTESİ.....	XII
TABLO LİSTESİ	XIV
1. GİRİŞ	1
1.1. Tezin Amacı ve Önemi	1
1.2. Tezin Organizasyonu	2
2. KAYNAK ARAŞTIRMASI	3
3. MATERYAL VE METOTLAR	7
3.1. Derin Öğrenme (DÖ)	7
3.2. Evrimsel Sinir Ağları (ESA)	10
3.2.1. Giriş (Input) katmanı	11
3.2.2. Evrişim (Convolutional) katmanı	12
3.2.3. Aktivasyon (Activation) katmanı.....	14
3.2.4. Ortaklama (Pooling) katmanı.....	14
3.2.5. Tam bağlantılı katman (Fully Connected Layer).....	15
3.2.6. Sınıflandırma (softmax) katmanı	16
3.3. Görüntü Sınıflandırmada Kullanılan Derin Öğrenme Modelleri.....	16
3.3.1. LeNet	16
3.3.2. AlexNet.....	17
3.3.3. VggNet.....	17
3.3.4. GoogLeNet.....	19
3.3.5. Microsoft RestNet.....	20
3.4. Ses ve Özellikleri	20
3.4.1. Sesin özellikleri.....	21
3.4.2. Sesin yayılma hızı	21
3.4.3. Frekans.....	21
3.4.4. Dalga boyu	22
3.4.5. Ses basıncı.....	22
3.4.6. Ses genliği ve şiddeti	22
3.5. Kentsel Sesler	22
3.6. Veri Setleri	24

3.6.1. ESC10 veri seti	24
3.6.2. ESC50 veri seti	25
3.6.3. UrbanSound8K	28
4. ÖNERİLEN YÖNTEM	30
4.1. ESC10_ESA32.....	31
4.2. ESC10_ESA224.....	32
4.3. ESC50_ESA32.....	34
4.4. ESC50_ESA224.....	36
4.5. URBANSOUND8K_ESA32	38
4.6. URBANSOUND8K_ESA224	40
5. DENEYSEL ÇALIŞMALAR.....	42
5.1. Performans Ölçütleri.....	42
5.2. ESC10 Veri Seti Üzerinde Yapılan Çalışmalar	43
5.3. ESC50 Veri Seti Üzerinde Yapılan Çalışmalar	50
5.4. UrbanSound8K Veri Seti Üzerinde Yapılan Çalışmalar	63
5.5. Önceki Çalışmalarla Karşılaştırılması	71
6. SONUÇ VE ÖNERİLER.....	73
KAYNAKLAR	75
ÖZGEÇMİŞ	80

SİMGELER VE KISALTMALAR

DÖ	: Derin Öğrenme
ESA	: Evrişimsel Sinir Ağları (Convolutional Neural Networks-CNN)
ÇSS	: Çevresel Seslerin Sınıflandırılması
CPU	: Merkezi İşlem Birimi (Central Processing Unit)
GPU	: Grafik İşlem Birimi (Graphic Processing Unit)
ILSVRC	: Büyük Ölçekli Görsel Tanıma Yarışması (ImageNet Large Scale Visual Recognition Challenge)
ESC	Environmental Sounds Classification
IoT	Nesnelerin İnterneti
ReLu	: Düzleştirilmiş Doğrusal Birim Katmanı
RGB	: Kırmızı, Yeşil, Mavi Renk uzayı (Red, Green, Blue)
CWT	Sürekli Dalgacık Dönüşümü
TP	: Doğru Pozitif (True Positive)
FP	: Yanlış Pozitif (False Positive)
FN	: Yanlış Negatif (False Negative)
TN	: Doğru Negatif (True Negative)

ŞEKİL LİSTESİ

Şekil 3.1. Derin öğrenme modellerinin yıllara göre ILSVRC yarışmasında elde ettikleri hata değerleri (inik 2019).....	9
Şekil 3.2. Yapay zekâ, makine öğrenmesi ve derin öğrenme arasındaki ilişki.....	9
Şekil 3.3. Derin öğrenmede ESA mimarilerinin temel yapısı	10
Şekil 3.4. ESA mimarisine verilen tek kanallı giriş görüntüsü.....	11
Şekil 3.5. ESA mimarisine verilen renkli bir görüntü ve onun renk kanalları.....	11
Şekil 3.6. ESA mimarisinde konvolüsyon katmanlarındaki farklı temsillerin elde edilmesi ve soldan sağa doğru insan yüzü, araba, fil ve sandalyeye ait farklı temsillerin çıkarılması (Lee ve ark 2009)	12
Şekil 3.7. ESA mimarisinde konvolüsyon işleminin uygulanması.....	13
Şekil 3.8. Düzleştirilmiş doğrusal birim katmanı	14
Şekil 3.9. ESA mimarisinde pooling (havuzlama) uygulama işlemi.....	15
Şekil 3.10. ESA mimarisinde tam bağlantılı katmanın yapısı (Tammina 2019)	16
Şekil 3.11. ESA mimarisinde sınıflandırma katmanı	16
Şekil 3.12. ESA mimarisinde LeNet-5 mimarisi (LeCun ve ark 1998).....	17
Şekil 3.13. ESA mimarisinde AlexNet mimarisi (Krizhevsky ve ark 2012).....	17
Şekil 3.14. ESA mimarisinde VggNet ağ mimarisi (Hassan 2018).....	18
Şekil 3.15. ESA mimarisinde GoogleNet/ Inception ağı (Szegedy ve ark 2015).....	19
Şekil 3.16. ESA mimarisinde Microsoft ResNet ağı (He ve ark 2016).....	20
Şekil 3.17. Hava moleküllerinin sıkışma ve seyrekleşme durumlarına bağlı olarak oluşan ses dalgası (Huang ve ark 2001).....	20
Şekil 3.18. Fiziksel özellikleri bakımından ses dalgası	21
Şekil 3.19. ÇSS'lerin taksonomisi (Salamon ve ark 2014).....	23
Şekil 3.20. ESC-10 veri setindeki sınıflara ait kayıtların grafiksel gösterimi	24
Şekil 3.21. (a) ESC50 veri setindeki her bir sınıfın ses kayıtlarına ait örnek bir grafik gösterimi	26
Şekil 3.21. (b) ESC50 veri setindeki her bir sınıfın ses kayıtlarına ait örnek bir grafik gösterimi	26
Şekil 3.21. (c) ESC50 veri setindeki her bir sınıfın ses kayıtlarına ait örnek bir grafik gösterimi	27
Şekil 3.21. (d) ESC50 veri setindeki her bir sınıfın ses kayıtlarına ait örnek bir grafik gösterimi	27
Şekil 3.21. (e) ESC50 veri setindeki her bir sınıfın ses kayıtlarına ait örnek bir grafik gösterimi	28
Şekil 3.22. UrbanSound8k veri setindeki sınıflardaki her bir kayda ait örnek gösterim	29
Şekil 4.1. Tezde uygulanacak yöntemin mantıksal tasarımı.....	30

Şekil 4.2. ESC10_ESA32 modelinin mimarisi	31
Şekil 4.3. ESC10_ESA32 modelinin birinci evrişim katmanındaki filtreler	31
Şekil 4.4. ESC10_ESA32 modelinin birinci evrişim katmanındaki filtrelerin giriş görüntüsüne uygulandıktan sonra elde edilen özellik haritaları	32
Şekil 4.5. ESC10_ESA224 modelinin mimarisi	33
Şekil 4.6. ESC10_ESA224 modelinin birinci evrişim katmanındaki filtreler	33
Şekil 4.7. ESC10_ESA224 modelinin birinci evrişim katmanındaki filtrelerin giriş görüntüsüne uygulandıktan sonra elde edilen özellik haritaları	33
Şekil 4.8. ESC50_ESA32 modelinin mimarisi	34
Şekil 4.9. ESC50_ESA32 modelinin birinci evrişim katmanındaki filtreler	35
Şekil 4.10. ESC50_ESA32 modelinin birinci evrişim katmanındaki filtrelerin giriş görüntüsüne uygulandıktan sonra elde edilen özellik haritaları	35
Şekil 4.11. ESC50_ESA224 modelinin mimarisi	36
Şekil 4.12. ESC50_ESA224 modelinin birinci evrişim katmanındaki filtreler	37
Şekil 4.13. ESC50_ESA224 modelinin birinci evrişim katmanındaki filtrelerin giriş görüntüsüne uygulandıktan sonra elde edilen özellik haritaları	37
Şekil 4.14. URBANSOUND8K_ESA32 modelinin mimarisi	38
Şekil 4.15. URBANSOUND8K_ESA32 modelinin birinci evrişim katmanındaki filtreler	39
Şekil 4.16. URBANSOUND8K_ESA32 modelinin birinci evrişim katmanındaki filtrelerin giriş görüntüsüne uygulandıktan sonra elde edilen özellik haritaları	39
Şekil 4.17. URBANSOUND8K_ESA224 modelinin mimarisi	40
Şekil 4.18. URBANSOUND8K_ESA224 modelinin birinci evrişim katmanındaki filtreler	40
Şekil 4.19. URBANSOUND8K_ESA224 modelinin birinci evrişim katmanındaki filtrelerin giriş görüntüsüne uygulandıktan sonra elde edilen özellik haritaları	41
Şekil 5.1. ESC10_ESA32 modelinin eğitim aşamasındaki yakınsama grafiği	43
Şekil 5.2. ESC10_ESA224 modelinin eğitim aşamasındaki yakınsama grafiği	47
Şekil 5.3. ESC50_ESA32 modelinin eğitim aşamasındaki yakınsama grafiği	50
Şekil 5.4. ESC50_ESA224 modelinin eğitim aşamasındaki yakınsama grafiği	57
Şekil 5.5. URBANSOUND8K_ESA32 modelinin eğitim aşamasındaki yakınsama grafiği	64
Şekil 5.6. URBANSOUND8K_ESA224 modelinin eğitim aşamasındaki yakınsama grafiği	68

TABLO LİSTESİ

Tablo 3.1. ESA mimarisinde VggNet modeline ait farklı konfigürasyonlar (Simonyan ve Zisserman 2014).....	18
Tablo 3.2. ESA mimarisinde GoogLeNet parametreleri (Szegedy ve ark 2015)	19
Tablo 3.3. ESC50 veri setindeki kategoriler ve her bir kategorideki sınıfların isimleri .	25
Tablo 4.1. ESC10_ESA32 modelindeki her bir katmana ait parametre değerleri	32
Tablo 4.2. ESC10_ESA224 modelindeki her bir katmana ait parametre değerleri	34
Tablo 4.3. ESC50_ESA32 modelindeki her bir katmana ait parametre değerleri.....	36
Tablo 4.4. ESC50_ESA224 modelindeki her bir katmana ait parametre değerleri	38
Tablo 4.5. URBANSOUND8K_ESA32 modelindeki her bir katmana ait parametre değerleri	39
Tablo 4.6. URBANSOUND8K_ESA224 modelindeki her bir katmana ait parametre değerleri	41
Tablo 5.1. Karışıklık matrisi	42
Tablo 5.2. (a) ESC10_ESA32 modelinin test verisi üzerinde elde ettiği karışıklık matrisi (Fold 1-6)	44
Tablo 5.2. (b) ESC10_ESA32 modelinin test verisi üzerinde elde ettiği karışıklık matrisi (Fold 7-10)	45
Tablo 5.3. ESC10_ESA32 modelinin 10-kat çaprazlama ile ESC10 veri seti üzerinde %80.75 ortalama doğruluğun elde edildiği karışıklık matrisi.....	46
Tablo 5.4. (a). ESC10_ESA224 modelinin test verisi üzerinde elde ettiği karışıklık matrisi (Fold 1-2)	47
Tablo 5.4. (b). ESC10_ESA224 modelinin test verisi üzerinde elde ettiği karışıklık matrisi (Fold 3-8)	48
Tablo 5.4. (c). ESC10_ESA224 modelinin test verisi üzerinde elde ettiği karışıklık matrisi (Fold 9-10)	49
Tablo 5.5. ESC10_ESA224 modelinin 10-kat çaprazlama ile ESC10 veri seti üzerinde %82.25 ortalama doğruluğun elde edildiği karışıklık matrisi.....	49
Tablo 5.6. (a) ESC50_ESA32 modelinin test verisi üzerinde elde ettiği karışıklık matrisi (Fold 1-2)	51
Tablo 5.6. (b) ESC50_ESA32 modelinin test verisi üzerinde elde ettiği karışıklık matrisi (Fold 3-4)	52
Tablo 5.6. (c) ESC50_ESA32 modelinin test verisi üzerinde elde ettiği karışıklık matrisi (Fold 5-6)	53
Tablo 5.6. (d) ESC50_ESA32 modelinin test verisi üzerinde elde ettiği karışıklık matrisi (Fold 7-8)	54
Tablo 5.6. (e) ESC50_ESA32 modelinin test verisi üzerinde elde ettiği karışıklık matrisi (Fold 9-10)	55

Tablo 5.7. ESC50_ESA32 modelinin 10-kat çaprazlama ile ESC50 veri seti üzerinde %54.55 ortalama doğruluğun elde edildiği karışıklık matrisi.....	56
Tablo 5.8. (a) ESC50_ESA224 modelinin test verisi üzerinde elde ettiği karışıklık matrisi (Fold 1-2)	58
Tablo 5.8. (b) ESC50_ESA224 modelinin test verisi üzerinde elde ettiği karışıklık matrisi (Fold 3-4)	59
Tablo 5.8. (c) ESC50_ESA224 modelinin test verisi üzerinde elde ettiği karışıklık matrisi (Fold 5-6)	60
Tablo 5.8. (d) ESC50_ESA224 modelinin test verisi üzerinde elde ettiği karışıklık matrisi (Fold 7-8)	61
Tablo 5.8. (e) ESC50_ESA224 modelinin test verisi üzerinde elde ettiği karışıklık matrisi (Fold 9-10)	62
Tablo 5.9. ESC50_ESA224 modelinin 10-kat çaprazlama ile ESC50 veri seti üzerinde %72.15 ortalama doğruluğun elde edildiği karışıklık matrisi.....	63
Tablo 5.10. (a) URBANSOUND8K_ESA32 modelinin test verisi üzerinde elde ettiği karışıklık matrisi (Fold 1-6)	65
Tablo 5.10. (b) URBANSOUND8K_ESA32 modelinin test verisi üzerinde elde ettiği karışıklık matrisi (Fold 7-10)	66
Tablo 5.11. URBANSOUND8K_ESA32 modelinin 10-kat çaprazlama ile UrbanSound8K veri seti üzerinde %88.60 ortalama doğruluğun elde edildiği karışıklık matrisi	67
Tablo 5.12. (a) URBANSOUND8K_ESA224 modelinin test verisi üzerinde elde ettiği karışıklık matrisi (Fold 1-6)	69
Tablo 5.12. (b) URBANSOUND8K_ESA224 modelinin test verisi üzerinde elde ettiği karışıklık matrisi (Fold 7-10)	70
Tablo 5.13. URBANSOUND8K_ESA224 modelinin 10-kat çaprazlama ile UrbanSound8K veri seti üzerinde %84.33 ortalama doğruluğun elde edildiği karışıklık matrisi	71
Tablo 5.14. ESC10 veri seti üzerinde yapılan çalışmaların karşılaştırılması	72
Tablo 5.15. ESC50 veri seti üzerinde yapılan çalışmaların karşılaştırılması	72
Tablo 5.16. UrbanSound8K veri seti üzerinde yapılan çalışmaların karşılaştırılması....	72

1. GİRİŞ

1.1. Tezin Amacı ve Önemi

İnsanların yerleşik hayata geçmesinden itibaren günümüze kadar hızla büyüyen kentlerde artan nüfus ile beraber çevresel faaliyetlerde artmaktadır. Hızla gelişen bilgi ve iletişim teknolojileri ile beraber kentlerin sosyal yapıları da etkilenmektedir. Özellikle artan nüfus ile birlikte kentlerdeki faaliyetlerin çeşitlenmesi ve artması kaçınılmazdır. Meydana gelen her faaliyetin kendine özgü oluşumu ve beraberinde bir sonuç ile sonlanmaktadır. Bu gelişim ile birlikte kentlerde meydana gelen her faaliyetin sonuçları toplanıp işlenebilir veri haline getirilmesi gerekmektedir. Dolayısıyla kentlerin eğitim, sağlık, sosyal, güvenlik, yerel hizmetler gibi farklı alanlarda meydana gelen gelişmelerle artan faaliyetlerin sürdürülebilir etkin çözümlere ve kentlerde verilen hizmetlerin hızlı ve anında yerine getirilmesi için bu faaliyetlerin sonuçlarından çıkan verilerin işlenip güncel teknolojik modeller oluşturulmalıdır. Bu modellerin ihtiyaç duyulan duruma hızlı ve doğru bir şekilde karar alıp müdahale edebilmesi için modellere kaynak sağlayan verilerin faaliyeti tam temsil etmesi gerekmektedir.

Çevremizde faaliyetler canlı veya cansız varlıklar tarafından oluşmaktadır. Bu faaliyetlerin sonuçlarını temsil edebilecek ve aynı zamanda ortam hakkında bilgi verebilecek ses verisi önem kazanmaktadır. Kentlerde meydana gelen gürültü kirliliği, trafik problemleri, güvenlik sistemleri, akıllı takip sistemleri, sağlık hizmetleri, yerel hizmetler gibi faaliyetlerin işleyişini ve temel bilgilerini elde etmek için ses verisinden faydalanılmaktadır. Bu anlamda Çevresel Seslerin Sınıflandırması (ÇSS) kritik önem kazanmaktadır. Bu seslerin işlenmesi için klasik makine öğrenme yöntemleri ve ses işleme yöntemleri kullanılarak birçok çözüm üretilmeye çalışılmıştır. Fakat artan veri miktarı ve çözümlemedeki zaman kısıtlamalarından dolayı anlık otomatik olarak seslerin tanımlanmasını sağlayan yeni güçlü yapay zekâ yöntemlerine ihtiyaç duyulmaktadır. Bu alanda son yıllarda derin öğrenme mimarilerinin birçok farklı alanda elde etmiş olduğu yüksek performanstan dolayı ÇSS'nin işlenmesi derin öğrenme mimarileri ile sınıflandırılması amaçlanmıştır.

Yapılan bu çalışmada üç farklı ÇSS veri setinin sınıflandırılması için özgün derin öğrenme mimarileri geliştirilmiştir. Bu veri setleri sırasıyla 10, 50 ve 8 sınıftan oluşmaktadır. Her bir veri seti için özgün Evrişimsel Sinir Ağları (ESA) modelleri

tasarlanmıştır. Model tasarım aşamasında en az parametreye sahip modeller elde edilmeye çalışılmıştır.

1.2. Tezin Organizasyonu

Sunulan bu tez çalışmasında Bölüm 1’de tezin amacı, önemi ve organizasyonu açıklanmıştır. Bölüm 2’deki kaynak araştırması kısmında ÇSS ile ilgili Derin Öğrenme (DÖ) yöntemleri hakkında literatürde yapılan çalışmalar belirtilmiştir. Bölüm 3’te tezde kullanılan materyal hakkında bilgilendirmeler yapılmıştır. Bu bölümde Yapay Zekâ (AI), Makine Öğrenmesi (ML), DÖ, ESA, nesne sınıflandırma ve tanımlamada kullanılan DÖ modelleri hakkında bilgiler ve birbiri ile olan ilişkisi anlatılmıştır. Aynı zamanda bu tez kapsamında kullanılan ses veri setleri detaylıca anlatılmıştır. Bölüm 4’te önerilen yöntemin mimarisi, yöntemin eğitim ve test aşaması, yöntemin parametre optimizasyonu gibi yöntem hakkında bilgilendirmeler yapılmıştır. Bu bölümde ayrıca tez kapsamında elde edilen veri setleri hakkında bilgiler sunulmuştur. Bölüm 5’te deneysel çalışmalar anlatılmıştır. Bu bölümde öncelikle performans ölçütleri hakkında gerekli açıklamalar yapılmıştır. Ayrıca segmentasyon ve sınıflandırma için oluşturulan modellerin mimarileri, parametre değerleri ve elde ettikleri performans değerleri verilmiştir. Son olarak önerilen yöntemin diğer yöntemlerle test sonuçları karşılaştırılıp gerekli tartışmalar yapılmıştır. Bölüm 6’da, tez kapsamında elde edilen sonuçlar ve öneriler anlatılmıştır.

2. KAYNAK ARAŞTIRMASI

Ses verisi, görsel verinin sağlayabileceği bilginin yanı sıra anlamsal olarak daha zengin bilgi içerir (Chu ve ark 2009). Özellikle bir ortam hakkında bilgi elde edebilmek için ses verisi daha önemli hale gelmektedir. Günlük hayatta bazı uygulamaların gerçekleştirilmesi için konuşma ve müzik seslerinin aksine çevresel seslerin kullanılması gerekmektedir. Bu sebeple son yıllarda çevresel seslerin sınıflandırılması üzerine çalışmalar yoğunlaşmıştır. ÇSS, konuşma dışı ses sınıflandırma görevinin en önemli konularından biri olarak bilinir (Demir ve ark 2020). Birçok problemde çevresel seslerin sınıflandırılması kritik bir önem taşımaktadır. Gürültü kirliliği analizi (Aumond ve ark 2017, Cao ve ark 2019), gözetleme sistemleri (Radhakrishnan ve ark 2005, Crocco ve ark 2016, Laffitte ve ark 2019), makinenin duyması (Huang 2002, Chu ve ark 2006, Li ve ark 2007, Lyon 2010), çevresel gözetleme (Green ve Murphy 2020), cinayetlerin önceden kestirilmesi (Intani ve Orachon 2013) ve akıllı şehirler (Ntalampiras 2014, Agha ve ark 2017) gibi uygulamalar mevcuttur.

Literatürde ÇSS için farklı istatistiksel ve klasik makine öğrenmesi yöntemleri kullanılmıştır (Dhanalakshmi ve ark 2011, Geiger ve Helwani 2015, Salamon ve Bello 2015, Stowell ve ark 2015, Ludena-Choez ve Gallardo-Antolin 2016, Bisot ve ark 2017, Mulimani ve Koolagudi 2019, Xie ve Zhu 2019). Derin öğrenmenin (Krizhevsky ve ark 2012) 2012 yılından imageNet (Deng ve ark 2012) yarışması ile yüksek bir başarı değeri elde ettiğinden dolayı birçok farklı problemlerin çözülmesinde kullanılmaya başlanılmıştır. Elde edilen bu başarıdan dolayı farklı alanlarda kullanılmaya başlandığı gibi son yıllarda ÇSS için derin öğrenme modelleri sıklıkla kullanılmaya başlamıştır (Piczak 2015, Takahashi ve ark 2016, Boddapati ve ark 2017, Salamon ve Bello 2017, Tokozume ve ark 2017, Zhang ve ark 2017, Li ve ark 2018, Lim ve ark 2018, Abdoli ve ark 2019, Chen ve ark 2019, Su ve ark 2019, Akbal 2020, Medhat ve ark 2020, Mushtaq ve Su 2020, Mushtaq ve ark 2021). Piczak tarafından yapılan çalışmada üç farklı veri setinin sınıflandırılmasında ESA modeli kullanılmıştır. Önerilen ESA modeli iki konvolüsyon katmanı bir pooling katmanı ve iki tamamen bağlı katmandan oluşmaktadır. Elde edilen sonuçlarda ESA modelinin mevcut diğer yöntemlere göre daha iyi performans sergilediği gösterilmiştir (Piczak 2015). Salamon ve ark. tarafından yapılan çalışmada iki temel amaç hedeflenmiştir. Birincisi ÇSS'nin sınıflandırılması için ESA modelinin kullanılması amaçlanmıştır. İkinci olarak ESA modellerinin eğitilmesi için gereken yüksek miktarda verilerin elde edilmesi için veri artırma

işlemleri gerçekleştirilmiştir. Veri artırma ile ESA modelinin daha iyi bir performans sergilediği gösterilmiştir (Salamon ve Bello 2017). Takahashi ve ark. tarafından akustik olayların tespiti için ESA modeli önerilmiştir. Önerilen ESA modeli VGGNET'ten ilham alınarak geliştirilmiştir. Ayrıca etigim işlemlerinde verilen artırılması için yeni bir yöntem önerilmiştir (Takahashi ve ark 2016). Derin öğrenme modellerinin eğitim aşamasında verilerin modeli beslemesi için farklı bir strateji Tokozume ve ark. tarafından geliştirilmiştir. Sınıflar arası öğrenme (Between-Class learning) adındaki bu yöntemde farklı sınıflara ait iki sesi rastgele oranlarda karıştırarak sınıflar arası sesler üretilmektedir. Yapılan çalışmada sınıflar arası öğrenme yönteminin ses tanıma ağlarında ve veri artırmada iyi performans sergilediği belirtilmiştir. Çalışmada ayrıca ÇSS'nin sınıflandırılması için bir ağ tanımlanmış olup önerilen yöntem ile eğitilmiştir. Elde edilen sonuçların insan ses tanıma hata değerinden daha düşük değerler olduğu ifade edilmiştir (Tokozume ve ark 2017). Boddapatia ve ark. tarafından AlexNet ve GoogLeNet derin öğrenme modelleri kullanılarak ESC10, ESC50 ve UrbanSound8K veri setlerinde sınıflandırma işlemi gerçekleştirilmiştir. Veri setlerindeki ses sinyalleri spectrogram, MFCC ve CRP yöntemleri kullanılarak görüntü haline çevrilmiştir. Daha sonra elde edilen görüntülerde AlexNet ve GoogLeNet kullanılarak sınıflandırma işlemi gerçekleştirilmiştir (Boddapati ve ark 2017). Li ve ark. tarafından yapılan çalışmada özgün bir yığılmış ESA modeli önerilmiştir. Önerilen ağda azaltılmış filtre sayıları ile çoklu konvolüsyon katmanları kullanılmıştır. Yöntemde log-mel özellikleri ile ham ses sinyal verileri kullanılarak iki farklı ESA ağı eğitilmektedir. Bu iki model daha sonra Dempster-Shafer (DS) kanıt teorisi kullanılarak birleştirilip ÇSS için DS-CNN adında yeni bir model önerilmiştir. Önerilen modelin ESC10, ESC50 ve UrbanSound8K veri setlerinde ESA tabanlı diğer modellere göre daha iyi performans sergilediği belirtilmiştir (Li ve ark 2018). Su ve ark. tarafından yapılan çalışmada ÇSS'nin daha kapsamlı temsil edilmesi için iki farklı özelliğinin birleştirilmesi önerilmiştir. Sınıflandırma işlemi için dört katmanlı bir ESA ağı tasarlanmıştır. Farklı özelliklerle eğitilen ESA ağları Dempster-Shafer (DS) kanıt teorisi kullanılarak birleştirilmiştir. Önerilen TSCNN-DS modeli UrbanSound8K veri seti üzerinde %97.2 doğruluk oranı elde etmiştir (Su ve ark 2019). Mushtaq ve ark. tarafından yapılan çalışmada akustik kaynak ile mikrofon arasındaki mesafeye çevredeki birçok ses kaynağının dâhil olmasıyla seslerin üst üste binmesi karmaşıklığı arttırdığı vurgulanmıştır. Bu karmaşıklıktan dolayı ESC'nin doğrulanmış ses özellikleriyle birlikte derin konvolüsyonlu sinir ağı (DCNN) kullanılmıştır. Çalışmada ESC-10, ESC-50 ve Kentsel ses (US8K) veri kümeleri için üç ses özneteliği çıkarma tekniği, Mel spektrogramı (Mel), Mel Frekans Cepstral Katsayısı (MFCC)

ve Log-Mel düşünülmüştür. Aşırı uyum riskini kaldırmak içinde maksimum havuzlama işlemi uygulanmıştır. Elde edilen doğruluk değerleri sırasıyla ESC-10, ESC-50 ve US8K için %94,94, %89,28 ve %95.37'dir (Mushtaq ve Su 2020). Mushtaq ve ark. tarafından yapılan çalışmada verilerin anlamlı bir şekilde artırılarak ESA ile sınıflandırılmada spektral görüntüler üzerine bir yaklaşım ortaya konulmuştur. Sunulan yaklaşımda Mel spektrogram özelliği kullanılmıştır. Rastgele seçilen 7 veya 9 katmanlı ESA modelleri ile ESC-10, ESC-50 ve Us8k verileri kullanılmıştır. Doğrudan ses üzerinden anlamlı veri artırma yöntemi kullanılarak etkin ve yüksek doğruluk elde edildiği vurgulanmıştır. Kullanılan modeller arasında ResNet-52, ESC-10 için% 99.04, Us8k veri kümeleri için% 99.49 ve DenseNet-161, ESC-50 için% 97.57 doğruluk oranlarına ulaşılmıştır (Mushtaq ve ark 2021). Chen ve ark. tarafından yapılan çalışmada ÇSS problemi genişletilmiş ESA ile ele alınmıştır. Bu yapının kullanılması ile konvolüsyon katmanından sonra kullanılan max-pooling işleminden daha yüksek sınıflama doğruluğu sonucu elde edildiği belirtilmiştir. Aynı zamanda farklı genişleme oranları ile evrişim katmanı sayısının sonuçlara olan etkisi araştırılmıştır. ÇSS probleminde genişletilmiş ESA, maksimum ortaklamalı ESA'dan daha iyi sonuçlar elde ettiği; fakat filtrelerin sayısı ve oranının artırılmasının sınıflama doğruluğunu olumsuz yönde etkilediği ifade edilmiştir (Chen ve ark 2019). Abdoli ve ark. tarafından yapılan çalışmada ÇSS sınıflandırılması için bir boyutlu (1D) ESA ağı kullanılmıştır. Giriş verisi için ses sinyali üzerinden çerçeveler alınmıştır. İnsan işitsel yapısını modelleyen bir Gammatone filtre bankası ile ESA'nın giriş katmanının başlatılması ve farklı girdi filtre boyutlarını oluşturan farklı mimariler araştırılmıştır. Yapılan deneysel çalışmalar sonucunda UrbanSound8K veri seti üzerinde %89 ortalama doğruluk değeri elde edilmiştir. Ham giriş verisinin kullanılması ile uçtan uca yapılan çalışmalar içerisinde en iyi performans sergilediği ifade edilmiştir. Ayrıca önerilen yöntemin literatürdeki diğer modellerden daha az parametreye sahip olduğu belirtilmiştir (Abdoli ve ark 2019). Medhat ve ark. tarafından Koşullu Sinir Ağı (CLNN) ve onun uzantısı olan Maskeli Koşullu Sinir Ağı (MCLNN) önerilmiştir. Yapılan çalışmada sesin doğal halinden faydalanılarak sesin zaman-frekans gösterimi modellenmiştir. Önerilen MCLNN ile farklı müzik ve ÇSS'lerin sınıflandırma işlemi gerçekleştirilmiştir. MCLNN sınıflama doğruluğu ESA tabanlı modellerden daha iyi performans sergilediği belirtilmiştir (Medhat ve ark 2020). Zhang ve ark. tarafından yapılan çalışmada ESA'lardaki evrişim katmanındaki filtrelerin boyutları ve aktivasyon fonksiyonlarının ÇSS'ye etkisi araştırılmıştır. Bu amaçla genişletilmiş ESA tabanlı bir model (D-CNN-ESC) önerilmiştir. Önerilen sistem üç farklı ses veri setine uygulanmıştır. UrbanSound8K veri setinde diğer yöntemlere göre

%10 daha az hata değeri elde edilmiştir (Zhang ve ark 2017). Lim ve ark. tarafından yapılan çalışmada ses olaylarının sınıflandırılması için ESA tabanlı bir yöntem önerilmiştir. Önerilen yöntemin UrbanSound8K (Salamon ve ark 2014), BBC Sound FX (Lim ve ark 2018), DCASE2016 (Mesaros ve ark 2016), ve FREESOUND veri setleri üzerinde 30 farklı ses olayının %81.5 doğruluk oranında sınıflandırdığı görülmüştür (Lim ve ark 2018). Akbal'ın yapmış olduğu çalışmada çevresel seslerden meydana gelen faaliyetin konumunu belirlemenin, dijital adli tıp için önemli olduğu vurgulanmıştır. ESC ile konum belirlenmiş ve bunun için kararlı özellik çıkarma yöntemi sunulmuştur. Çalışılan bu yöntemin 3 temel aşamadan meydana geldiği, bunlar özellik oluşturma, seçme ve sınıflandırma işlemlerini kapsadığını belirtilmektedir. Özellik çıkarımı için tek boyutlu yerel ikili desen (1D-LBP), tek boyutlu üçlü desen (1D-TP) ve istatistiksel özellik oluşturma yöntemleri, sınıflandırma için de kübik (3. Polinom Dereceli Çekirdek) destek vektör makinesi kullanılmıştır. Bu çalışmada önerilen yöntem ESC-10 veri setine uygulanıp veri setindeki seslerin sınıflandırılması sağlanmıştır. Bu önerilen yöntem ile % 90.25 doğruluk oranı elde edilmiştir (Akbal 2020). DÖ modelleri ile elde edilen başarı oranları makine öğrenmesi veya diğer yöntemler ile elde edilen sonuçlardan daha iyi olduğu görülmüştür. Bunun temel sebebi, derin öğrenme modellerindeki özellik keşfinin otomatik yapılması olarak özetlenebilir. Literatürde yapılan çalışmalar incelendiğinde ÇSS'nin sınıflandırılmasına yönelik başarı oranlarının arttırılabilir olduğu görülmüştür.

3. MATERYAL VE METOTLAR

Bu bölümde, öncelikle önerilen yöntemin temelini oluşturan DÖ ve DÖ'nün temel mimarisi olan ESA hakkında detaylı bilgilendirme yapılmıştır. ESA mimarisindeki her bir katman hakkında gerekli açıklamalar yapılmıştır. DÖ'de nesne sınıflandırma ve tanımlama alanında kullanılan modeller hakkında bilgiler verilmiştir. Ayrıca ses ve özellikleri hakkında bilgiler verilmiştir. Son olarak bu tez kapsamında kullanılan veri setleri hakkında bilgiler verilmiştir.

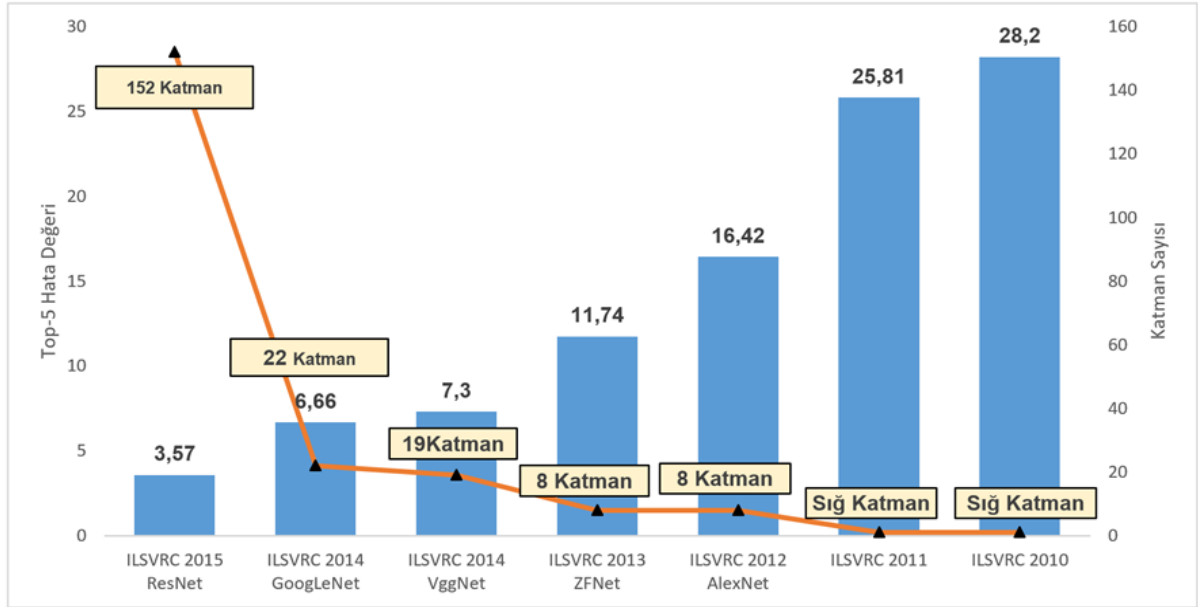
3.1. Derin Öğrenme (DÖ)

Yapay zekâ, makinelerin zeki canlı varlıkların kabiliyetlerini yerine getirme çabasıdır. Bir diğer deyişle zeki davranışların modellenmesi için geliştirilen algoritmaların tümüne verilen bir tanımdır. Bu kapsamda yapay zekâ, uzman sistemler, makine öğrenmesi, doğal dil işleme, ses işleme, görüntü işleme gibi temel alanları kapsamaktadır ve bu alanlarla ilgili algoritmalar geliştirilmektedir. Yapay zekâ terimi ilk defa 1940'lı yıllarda konuşulmaya başlanmıştır. Bu yıllarda insan sinir sisteminin incelenmesi ile aynı işlevi yerine getirecek yapay sinir ağları tasarlanmıştır (McCulloch ve Pitts 1943). İlerleyen yıllarda yapay sinir ağı temelli doğrusal modeller geliştirilmiştir (Rosenblatt 1958, Widrow ve Hoff 1960, Rosenblatt 1962). Bu modellerin doğrusal olmayan problemleri çözemediği ispatlanınca (Minsky 1969) yapay sinir ağı temelli mimarilerinin geliştirilmesi durağan bir sürece girmiştir. Bu süreçlerde makine öğrenmesi algoritmaları geliştirilmeye başlanmıştır.

Makine Öğrenmesi, yapay zekânın sayısal öğrenme, model oluşturan, mevcut verilerden öğrenebilen ve yeni veriler hakkında tahmin yapabilen sistemlerdir. Makine öğrenmesi günümüzde hayatımızın her alanında en yaygın bir şekilde kullanılmaktadır. Günümüzde işlenmesi gereken verilerin çok olmasıyla birlikte klasik makine öğrenme modellerinde veri ön işleme aşamasında veri temizleme, veri birleştirme, veri dönüşümü ve veri silme devamında öznitelik belirleme gibi süreçlerde uzmana ihtiyaç duyulmaktadır. 1980' yıllarında yapay sinir ağlarının geriye yayımlı ağların (Rumelhart ve ark 1986, LeCun 1987) geliştirilmesi ile tekrar popüler hale gelip doğrusal olmayan problemlerin çözümünde kullanılmaya başlanmıştır. Fakat bu ağların daha derin tasarlaması için gerekli donanımsal kaynakların olmayışı ve bu ağların eğitime yetecek kadar verilerin bulunmaması bu ağların kullanımını sınırlamıştır. İlerleyen yıllarda teknolojinin gelişmesi, sensörlerin hayatımızın her alanında kullanılması ve Nesnelerin İnterneti (IoT) ile dijitalleşen toplumumuzun her alanındaki faaliyetlerden veriler elde edilmeye başlanmıştır. Ayrıca sabit disk kapasiteleri ve

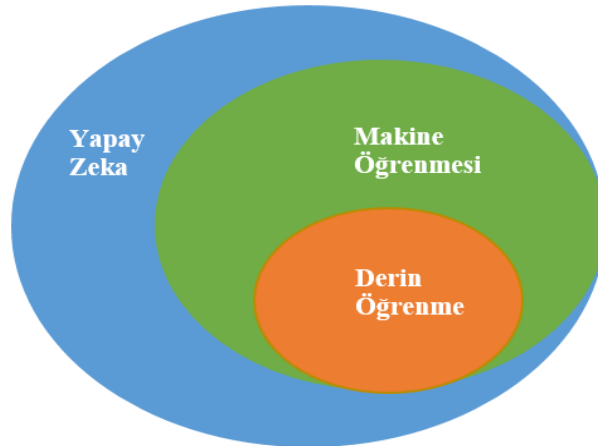
bulut teknolojileri ile artan depolama alanları ile toplumumuzda meydana gelen her faaliyetin sonuçları rahatlıkla veriler haline getirilip depolanabilmektedir. Depolanan bu veri yığınlarının işlenebilmesi ve anlamlı çözümlerin üretilmesi gerekmektedir. Bu amaçla paralel işlemci teknolojileri gelişerek grafik işlemci birimleri (GPU) üretilmeye başlanmıştır. Yeteri kadar verilerin depolanabilmesi ve bu verileri işleyecek donanımsal kaynaklarının geliştirilmesi ile DÖ ağları tasarlanmaya başlanmıştır.

DÖ, birden çok katmanlı öğrenme prensibine dayanan makine öğrenmenin bir alanıdır. İçerisinde birden çok gizli katmanları barındırmasından dolayı birçok parametre ile esneklik kazandırmaktadır. Derin öğrenmeyi yaygınlaştıran etkenlerden biri de her alandaki problemlere uygulanabilir olması ve yüksek doğruluk oranlarının elde edilmesidir (Özkan ve Ülker 2017). DÖ ağları diğer yöntemlerin aksine öğrenme işlemini ham veri üzerinde yaparak; sınıflandırma, tanımlama, segmentasyon gibi problemlerde yüksek performans sergilemektedir. Bu performansın temel sebebi derin ağların ham veri üzerinde kendine özgü özellikleri keşfedebilmesidir. Bu işlemde çözümü yapılan problem için yüksek boyutlarda veriler kullanılmaktadır. Bunun neticesi olarak derin ağların yüksek boyutlu verilerdeki karmaşık yapıları keşfetmede çok iyi olduğu ortaya çıkmıştır ve bu sebeple mühendislik, tıp, ekonomi, savunma sanayisi gibi alanlarda çok yoğun bir şekilde kullanılmaya başlanmıştır (Bengio ve ark 2013, LeCun ve ark 2015). Derin öğrenme ilk olarak 2012 yılında popüler hale gelmiştir. 2012 yılında yapılan ILSVRC (ImageNet Büyük Ölçekli Görsel Tanıma) (Competition 2012) yarışması ile görüntü sınıflandırma probleminde yüksek başarı oranı elde etmiştir Şekil 3.1’de gösterildiği gibi 2012 yılından itibaren bu yarışmadaki bütün modeller derin öğrenme tabanlı geliştirilmiş olup sınıflandırma hata oranı insan eşik hata değerinin altına düşmüştür.



Şekil 3.1. Derin öğrenme modellerinin yıllara göre ILSVRC yarışmasında elde ettikleri hata değerleri (inik 2019)

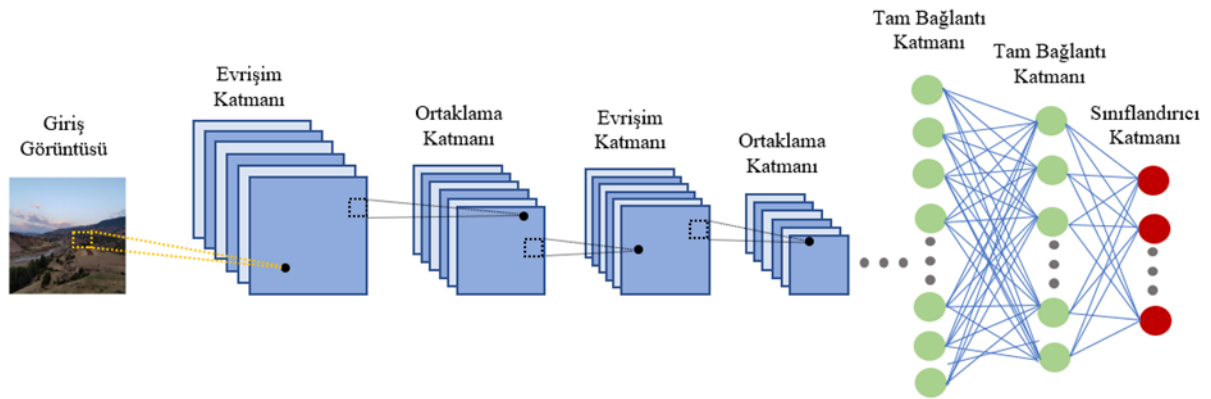
Yapay zekâ, makine öğrenmesi ve derin öğrenmenin şekilsel olarak birbiriyle olan ilişkisi Şekil 3.2’de verilmiştir. Her ne kadar DÖ’nün farklı problemlerde elde etmiş olduğu performans değeri diğer yapay zekâ yöntemlerinden daha yüksek olsa da yapay zekânın bir alt dalı olan makine öğrenmesi disiplini içerisinde bir alan olarak kabul edilmektedir. Farklı araştırmacılar tarafından makine öğrenmesi disiplini dışında farklı bir alan olarak kabul edilmektedir. Çünkü makine öğrenmesindeki algoritmalardan farklı olarak özellik keşfi, derin öğrenmede tamamen farklı bir şekilde yapılmaktadır. İleriki yıllarda DÖ mimarilerinin yapay zekânın bir alt dalı olarak anılması kaçınılmazdır.



Şekil 3.2. Yapay zekâ, makine öğrenmesi ve derin öğrenme arasındaki ilişki

3.2. Evrişimsel Sinir Ağları (ESA)

Derin Öğrenmenin ana mimarisi olarak kabul edilen Evrişimsel Sinir Ağları (ESA), öğrenme işlemini ham veri üzerinde yaparak kendine özgü özellikleri keşfetmesi sonucunda; sınıflandırma, tanımlama, segmentasyon gibi problemlerde yüksek performans sergilemektedir (İnik ve ark 2019, İnik ve ark 2021). Bu sebeple mühendislik, tıp, savunma sanayisi gibi birçok alanda kullanılması sonucunda yaygın hale gelmiştir. Özellikle otomatik özellik keşfi sayesinde devasa verilerde kullanılması ile Amazon, Google, Facebook gibi firmaların algoritmalarında kullanmaya başlamasıyla daha popüler hale gelmiştir (Dev 2017). ESA'lar nesnelerin tanımlanmasında art arda gelen birden çok katmandan oluşan bir katmanın çıktısı diğer katmanın girdisi olan ileri beslemeli çalışma prensibine dayanan ve genel olarak nesneleri parçalayarak özellik haritalarını (feature map) çıkararak ağı eğitmesini sağlayan otomatik sistemlerdir. ESA'ların özellik haritalarını kendi çıkarmasının sağlamış olduğu avantajın yanısıra araştırmacı kullanacağı modelde katman sayısı, katman dizilişi ve bu katmanlardaki parametreleri belirlemede özgür bırakılmıştır. Bu sayede araştırmacıya oluşturacağı modelde en yüksek performans oranına ulaşmasında sayısız deneme alanı bırakılmıştır. Derin Öğrenmede ESA temel yapısı Şekil 3.3'te verilmiştir.



Şekil 3.3. Derin öğrenmede ESA mimarilerinin temel yapısı

ESA mimarisi temelde giriş katmanı, evrişim katmanı, ortaklama katmanı, tam bağlantılı katman ve sınıflandırma katmanı olarak ele alınabilir. ESA'ların çalışma prensibine bakıldığında, giriş katmanında alınan uygun veri evrişim katmanına aktarılır. Art arda gelen katmanlarda işlemler yapılarak ağı eğitilir ve en sonunda bir çıkış verisi oluşturulur. Oluşan bu çıkış verisi istenilen sonuç ile karşılaştırılır ve aradaki fark hata payı olarak kabul edilir ve bu hata payını en aza indirmek için geri yayılım algoritması ile hata payı ağırlıklara aktararak her iterasyonda bu ağırlıklar güncellenip istenilen sonuca yaklaşılr (Özkan ve Ülker 2017,

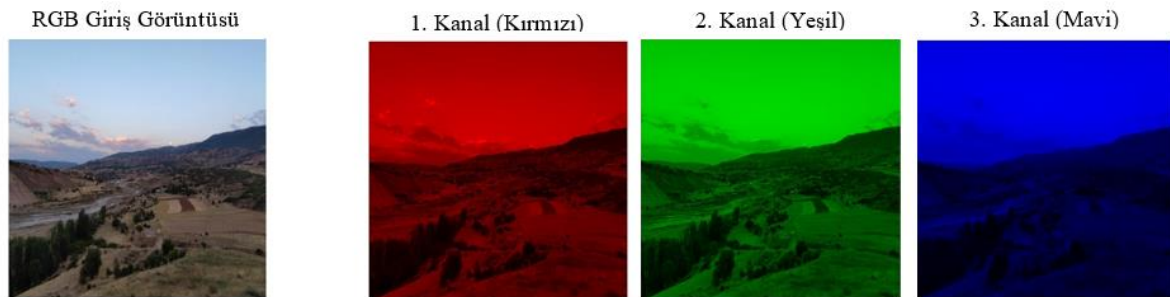
Zhang ve ark 2020). Elde edilecek en iyi performans değerine ulaşmak için ESA'nın parametrelerinden ve katmanlarından faydalanılarak bu işlem devam eder.

3.2.1. Giriş (Input) katmanı

Giriş katmanı ağa beslenecek verinin verildiği ilk katmandır. Bu katmanda dikkat edilmesi gereken, katmana verilecek verinin boyutudur. Tasarlanacak model için giriş verisinin boyutunun büyük olması ağın performans oranını arttırmasında olumlu bir etki yaparken, buna karşın yüksek hesaplama kapasitesi ve daha fazla bellek ihtiyacına gereksinim duyulması olumsuz bir etkisi söz konusudur. Giriş verisinin boyutunun küçük seçilmesi hesaplama karmaşıklığını azaltırken ağın performans oranını düşürmesi sonucunda çok fazla tercih edilmemektedir. Bu sebeple giriş katmanındaki veri boyutu tasarlanacak modelin performans sonucunu doğrudan etkilediği için uzman tarafından tasarlayacak ağ için en uygun giriş verisi seçilmesi gerekmektedir (Özkan ve Ülker 2017, inik 2019). Giriş katmanına verilecek görüntü tek kanallı veya birden fazla kanallı olabilmektedir. Örneğin Şekil 3.4'te verildiği gibi bir görüntü tek kanallı olarak ESA modeline verilebildiği gibi Şekil 3.5'te gösterildiği gibi bu görüntünün renkli (R=Red, G=Green, B=Blue) hali de ESA modeline verilebilmektedir. Renkli görüntü ESA modeline verildiğinde Şekil 3.5'te gösterildiği gibi her bir kanal ayrıca ESA modeline verilir.



Şekil 3.4. ESA mimarisine verilen tek kanallı giriş görüntüsü



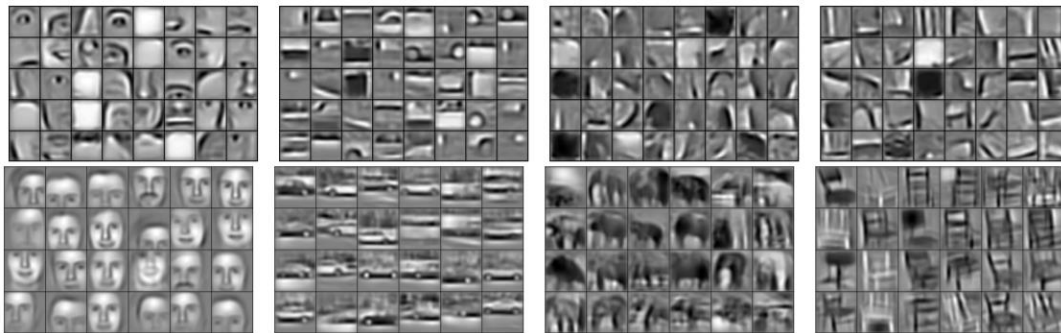
Şekil 3.5. ESA mimarisine verilen renkli bir görüntü ve onun renk kanalları

3.2.2. Evrişim (Convolutional) katmanı

Evrişim katmanı, bir önceki katmanda almış olduğu veri üzerinde belirlenen boyuttaki filtrelerin dolaştırılarak özellik haritalarının (feature map) çıkarıldığı katmandır. Bu sebeple dönüşüm katmanı olarak da bilenen bu katman ESA mimarisinin temelini oluşturur. Tasarımcının seçmiş olduğu filtre giriş görüntüsünün en üst sol tarafından başlayarak yatay ve düşey olarak üzerinden gezinir ve filtredeki değerler ile gezindiği görüntünün ilgili piksel değeri ile çarpılıp toplama alınarak matematiksel işlemler uygulanır. Bu işlem adım (stride) sayısına göre sağa ve aşağıya doğru kaydırılarak gerçekleştirilir. Konvolüsyon işlemi için uygulanan matematiksel formül Eşitlik 1’de verilmiştir.

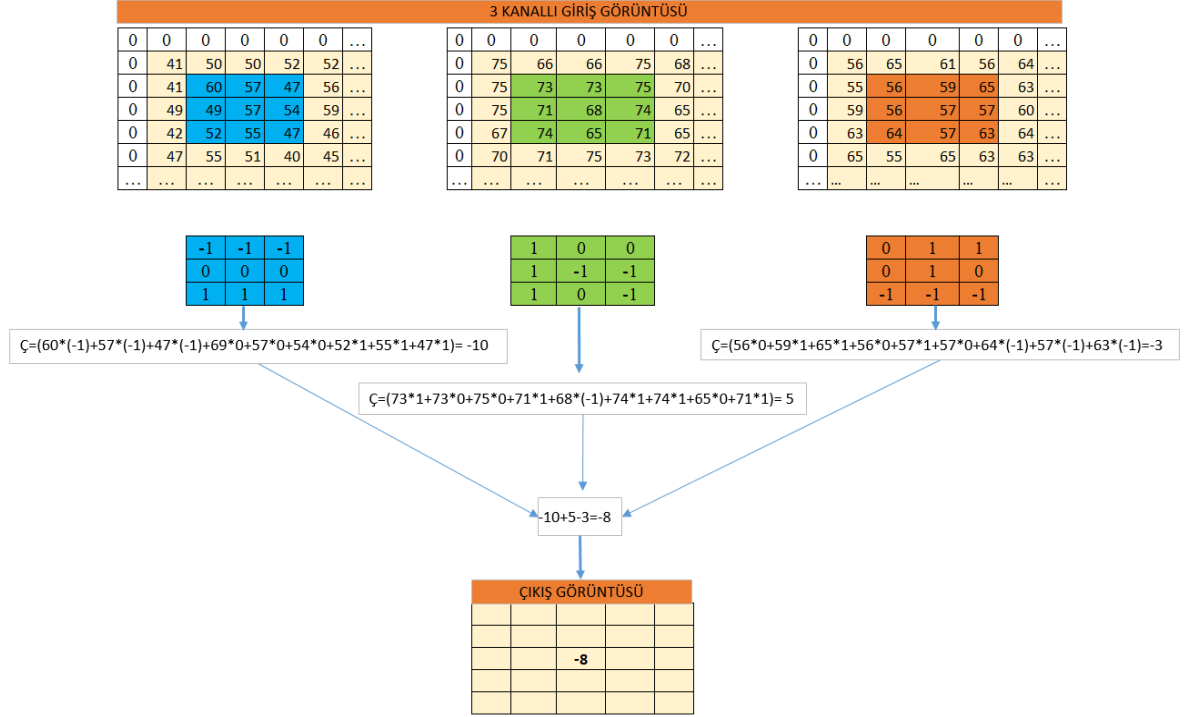
$$y[1,1] = \sum_i \sum_j x[i,j]h[1-i, 1-j] \quad (3.1)$$

Konvolüsyon işlemi belirlenen adım kayma sayısına göre gerçekleştirilir ve katmanda belirlenen filtre sayısı kadar çıkış görüntüsü elde edilir. Giriş verisi üç kanallı RGB ise, belirlenen filtre üç kanal üzerinden gezdirilir ve filtre her kanal için çıkarmış olduğu değerler toplanıp özellik haritasını oluşturur. Evrişim katmanının hedefi giriş verisinden kendine özelliklerin çıkarılmasıdır. Evrişim katmanında düşük düzeyden yüksek düzeye doğru farklı veri temsilleri oluşturulur. Genellikle ilk etapta nesnelere ait kenar bilgileri daha sonra bu kenarların birleşiminde oluşan basit şekil parçaları ve sonucunda bu parçaların birleştirilmesi ile nesneler oluşturulur (LeCun ve ark 2015). Şekil 3.6’da sırasıyla insan yüzü, araba, fil ve sandalyeye ait özelliklerin çıkartılıp bir araya getirilerek nesneleri oluşturmuştur (Lee ve ark 2009).



Şekil 3.6. ESA mimarisinde konvolüsyon katmanlarındaki farklı temsillerin elde edilmesi ve soldan sağa doğru insan yüzü, araba, fil ve sandalyeye ait farklı temsillerin çıkarılması (Lee ve ark 2009)

Şekil 3.7’de 5x5x3 boyutundaki RGB giriş görüntüsünün üzerinde adım sayısı 1 olan 3x3 boyutundaki filtrenin konvolüsyon işlemi uygulanması gösterilmiştir.



Şekil 3.7. ESA mimarisinde konvolüsyon işleminin uygulanması

Evrişim katmanında görüntü üzerine uygulanan filtre sonucunda oluşan görüntü boyutu hesabı Eşitlik 3.2’de verilmiştir. ζ filtre uygulandıktan sonra çıkan görüntü boyutu temsil ederken G giriş boyutu, A adım sayısı, F filtre boyutu ve P de eklenen satır sütun sayısını ifade eder.

$$\zeta = \frac{G-F+2xP}{A} + 1 \quad (3.2)$$

Evrişim katmanında Eşitlik 3.2’de görüldüğü gibi çıkış katmanındaki veri boyutunun düştüğü görülmektedir. Örneğin görüntü boyutu 30x30, filtre boyutu 3x3, adım kayma sayısı 1, eklenen satır sayısı 0 ise çıkış görüntü boyutu olan $\zeta = (30-3+2x0)/1 = 28$ boyutunda çıkış görüntüsü oluşur. Burada görüntü boyutunun değişmesinden dolayı bazı özelliklerin kaybolması söz konusudur. Bu özelliklerin kaybolmaması için evrişim katmanında kenar doldurma işlemi gerçekleştirilir. Böylelikle evrişim katmanında görüntü boyutunda herhangi bir değişme olmadan veriler bir sonraki katmana aktarılır yukarıdaki verilen örnekte giriş görüntü boyutu 30x30 iken kenar doldurma işlemi gerçekleştirilmediği için görüntü boyutu 28x28’e indirgenmiş oldu. Eğer kenar doldurma değeri 1 seçilmiş olursa hem yatayda hem de dikeyde görüntü kenarlarına 0 değeri eklenir. Böylelikle görüntünün çıkış değeri $\zeta = (30-3+2x1)/1 = 30$ elde edilir. Dolayısıyla evrişim katmanında herhangi bir boyut azaltma işlemi

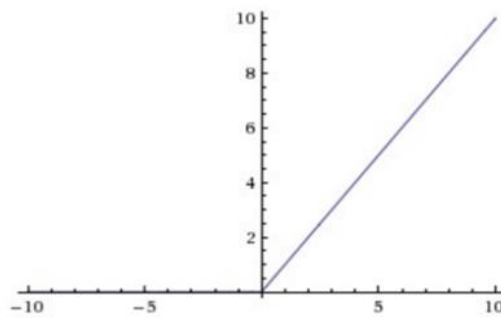
gerçekleştirilmeden gerekli evrişim işlemi gerçekleşir. Bu işlem sıfır doldurma (zero-padding) olarak adlandırılır.

3.2.3. Aktivasyon (Activation) katmanı

ESA mimarilerinde doğrusal olmayan veri özelliklerinin öğrenilebilmesi için aktivasyon fonksiyonu kullanılır. Gerçek dünyada hatalar lineer olmayan özelliklere sahiptir. Bu nedenle doğrusal olmayan aktivasyon fonksiyonları lineer aktivasyona göre daha çok tercih edilmektedir. Bu katmanda Sigmoid, Hiperbolik Tanjant (Hyperbolic Tangent), Basamak (Step), Doğrusal (Linear), Doğrultulmuş Doğrusal Birim (Rectified Linear Unit - ReLU) gibi fonksiyonlar kullanılmaktadır(Sharma ve ark 2017).

Doğrultulmuş Doğrusal Birim (Rectified Linear Unit - ReLU) Fonksiyonu: Doğrusal olmayan ve hızlı çalışma yapısından dolayı ReLU daha fazla popüler olmuştur. ReLU fonksiyonu kullanmanın avantajı, tüm nöronları aynı anda etkinleştirmemesidir (Sharma ve ark 2017). Bu katman Şekil 3.8’de gösterildiği gibi kendisinden önce gelen negatif verileri 0’a çeker diğer verileri ise aynı şekilde bırakır. Bu fonksiyonun formülü Eşitlik 3.3’te verilmiştir. ESA modellerinde yoğun olarak bu fonksiyon kullanılır.

$$f(x) = \begin{cases} 0 & \text{eğer } x < 0 \\ x & \text{eğer } x \geq 0 \end{cases} \quad (3.3)$$



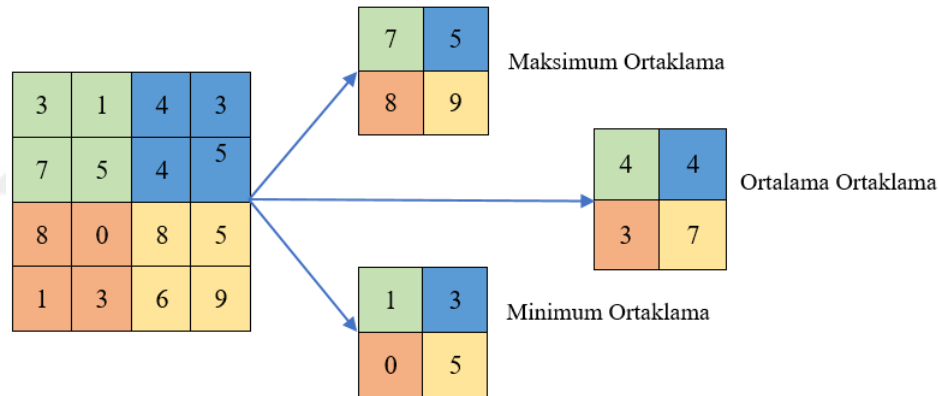
Şekil 3.8. Düzleştirilmiş doğrusal birim katmanı

3.2.4 Ortaklama (Pooling) katmanı

Ortaklama katmanı evrişim işleminden sonra veri boyutunu ve parametre sayısını azaltmak için kullanılmaktadır (Castelluccio ve ark 2015, Kaya ve ark 2020, Zhang ve ark 2021). Veri boyutunda azalma, kullanılacak bellek miktarını ve ihtiyaç duyulan işlem kapasitesini düşürdüğünden dolayı tasarlanacak ağı daha hızlı çalışmasını sağlar. Evrişim katmanındaki

gibi bu filtreler görüntü üzerinden belirli bir adım kayma sayısına göre gezdirilir ve görüntü üzerindeki piksellerin, maksimum havuzlama (maximum pooling), ortalama havuzlama (average pooling) ve minimum havuzlama (minimum pooling) değerleri alınarak havuzlama işlemi yapılmaktadır. Dolayısıyla sonraki katman için giriş boyutu (Genişlik x Yükseklik) azalmaktadır. Bu işlem veride derinlik boyutunu etkilemez (inik 2019). Maksimum havuzlama (maximum pooling), görüntünün filtre tarafından kaplanan bölümün maksimum değerini, ortalama havuzlama (average pooling) görüntünün filtre tarafından kaplanan bölümün ortalama değerini ve minimum havuzlama (minimum pooling) de görüntünün filtre tarafından kaplanan bölümün minimum değeri döndürür.

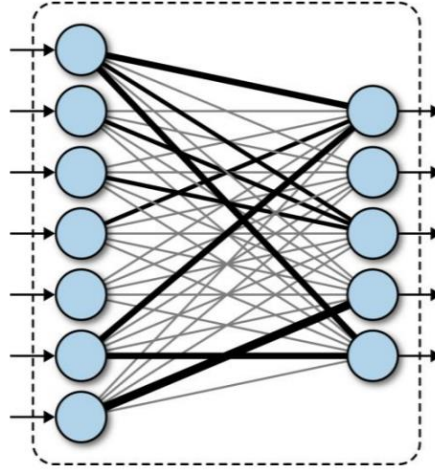
Örneğin Şekil 3.9'da ESA modelinin ortaklama katmanında kullanılan 4x4 boyutundaki görüntünün üzerinden adım kayma sayısı 2 olan 2x2 filtrenin gezdirilerek maksimum havuzlama (maximum pooling), ortalama havuzlama (average pooling) ve minimum havuzlama (minimum pooling) değerleri alınarak havuzlama işlemi gösterilmiştir.



Şekil 3.9. ESA mimarisinde pooling (havuzlama) uygulama işlemi

3.2.5 Tam bağlantılı katman (Fully Connected Layer)

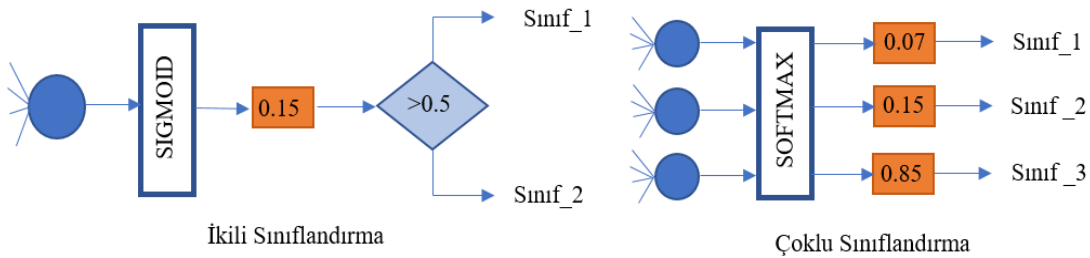
Tam bağlantılı katman ESA mimarilerinde konvolüsyon, aktivasyon ve ortaklama katmanlarından sonra gelip kendinden önceki tüm alanlara bağlıdır. Tam bağlantılı katmanların girdi sayısı artıkça hesaplama maliyeti de artmaktadır. Bu sebeple önceki katmanlarda boyut küçültmeye gidilerek bu katmana girecek özellik boyutlarını makul sayıda örneklenmelidir (Castelluccio ve ark 2015). Tek boyutlu bir matristen oluşan bu bağlantılı katmanlar mimarinin sonlarına doğru ve sınıf ağırlıklarının optimize etmede kullanılır (CINAR ve ark 2020). Tasarlanan farklı ESA mimarilerinde bu katman sayısı değişebilir (Özkan ve Ülker 2017). Şekil 3.10'da tam bağlantılı katman yapısı görülmektedir.



Şekil 3.10. ESA mimarisinde tam bağlantılı katmanın yapısı (Tammina 2019)

3.2.6 Sınıflandırma (softmax) katmanı

Tam bağlantılı katmandan sonra gelen sınıflandırma katmanı her sınıf için (0-1) aralığında bir değer üretir. Ağa verilen giriş verisi ESA'dan geçirildikten sonra sınıflandırma katmanında en yüksek değer sahip olan sınıfa dahil edilir. Bu katmanda belirlenen sınıf sayısı kadar sınıflandırma yapar. Örneğin 10 nesnenin sınıflandırılması yapılacaksa, sınıflandırma katmanı 10 çıkış değerine sahip olmalıdır. Bu katmanda çoğunlukla ikili sınıflandırma işlemi için sigmoid fonksiyonu kullanılırken çoklu sınıflandırma işlemi için de softmax fonksiyonu kullanılmaktadır. Bu fonksiyonlara ait örnekler Şekil 3.11'de verilmiştir.



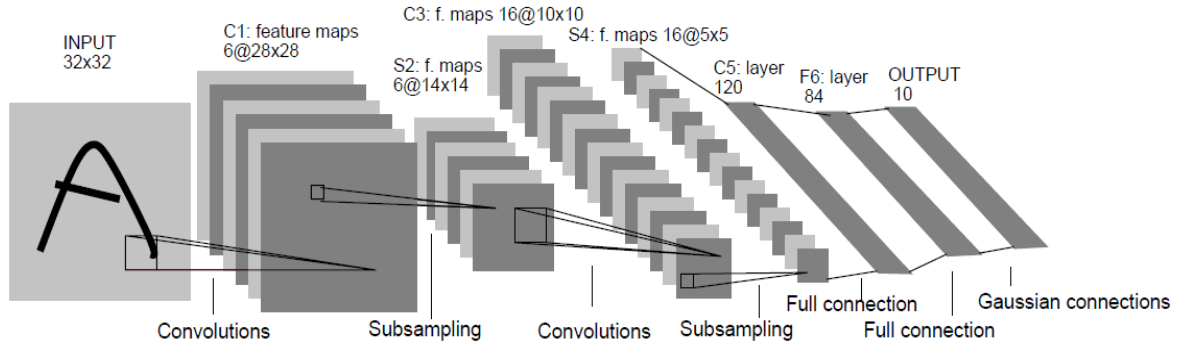
Şekil 3.11. ESA mimarisinde sınıflandırma katmanı

3.3. Görüntü Sınıflandırmada Kullanılan Derin Öğrenme Modelleri

3.3.1. LeNet

LeNet, 1998 yılında Yann LeCun ve ark. tarafından “Doküman Tanımaya Uygulanan Gradyan Temelli Öğrenme” makalesinde tanıtılmıştır (LeCun ve ark 1998). Bu çalışmada amaç elle tasarlanmış özellik çıkarma yöntemlere nazaran otomatik daha iyi görüntü tanıma sistemlerinin oluşturmaktır (LeCun ve ark 1998). Bu model konum ve ölçek değişimlerinden

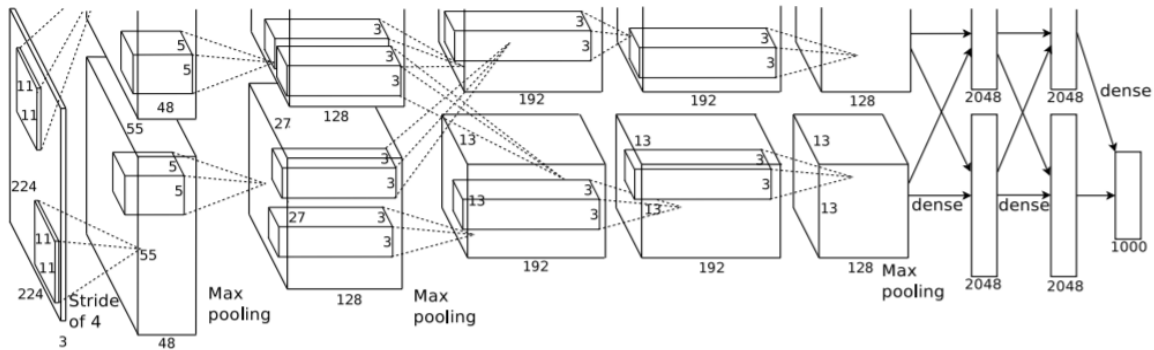
etkilenmeden banka çekleri ve adres numaraları üzerindeki el yazısı rakamlarını tespit etmek için geliştirilmiştir (KAYA 2021). Şekil 3.12’de LeNet-5 mimarisi gösterilmektedir. Bu çalışmada kullanılan veri seti 28x28 boyutunda 60000 resimden oluşan MNIST (el ile yazılmış rakamlar) kullanılmıştır (LeCun ve ark 1998, Furkan 2019).



Şekil 3.12. ESA mimarisinde LeNet-5 mimarisi (LeCun ve ark 1998)

3.3.2. AlexNet

Alex Krizhevsky ve ark. tarafından 2012 yılında yapılan ILSVRC (ImageNet Büyük Ölçekli Görsel Tanıma) (Competition 2012) 1,2 milyon görüntüden oluşan, 1000 farklı sınıfı sınıflandırmak için geliştirilmiş bir modeldir (Krizhevsky ve ark 2012). Bu yarışmadaki başarısı ile etki oluşturan AlexNet nesne tanımda hata oranını %26,2’den %15,3’e düşürmüştür. AlexNet, görüntü sınıflandırma ve tanıma için yeni stratejiler geliştiren ilk derin öğrenme mimarisi olarak kabul edilir. Bu modelin mimarisi Şekil 3.13’te verilmiştir.

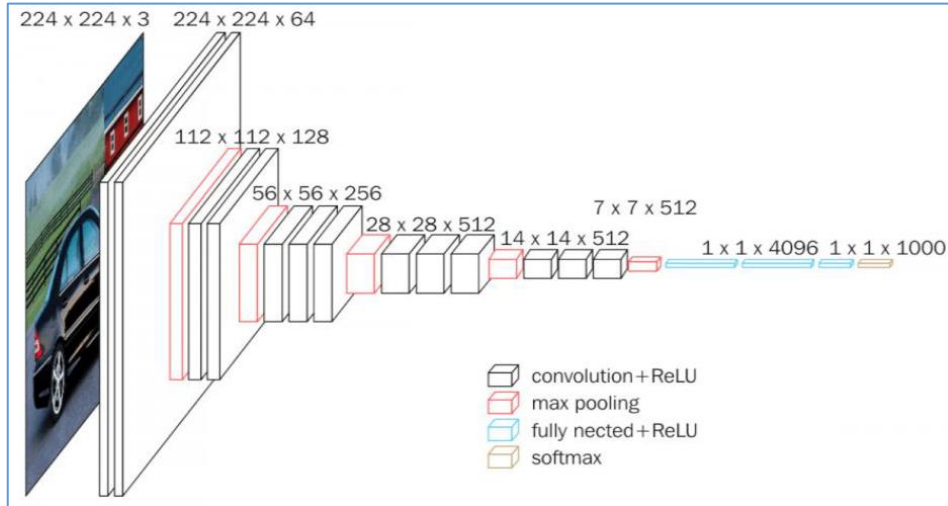


Şekil 3.13. ESA mimarisinde AlexNet mimarisi (Krizhevsky ve ark 2012)

3.3.3. VggNet

VggNet, Karen Simonyan ve Andrew Zisserman tarafından 2014 yılında ILSVRC yarışması için geliştirilmiştir. Bu model bu yarışmayı % 7.3 hata oranı ile sınıflandırma işlemi

gerçekleştirmiştir. Bu model basit ve ağın derinliğini artırarak evrimsel sinir ağının performansını artırmayı hedeflemiştir (Simonyan ve Zisserman 2014). VggNet modelinin mimarisi Şekil 3.14’te verilmiştir.



Şekil 3.14. ESA mimarisinde VggNet ağ mimarisi (Hassan 2018)

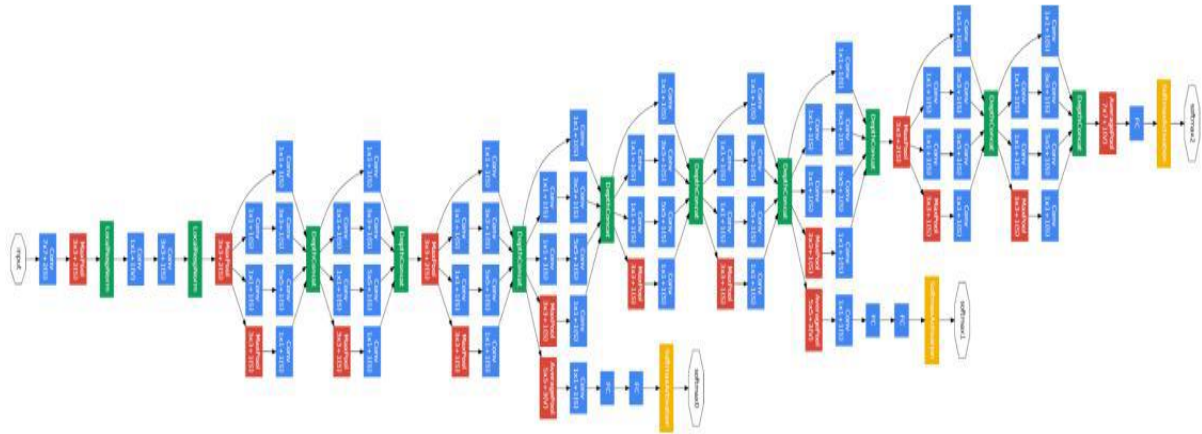
VggNet modeline ait Tablo 3.1’de beş farklı konfigürasyon bulunmaktadır. Bunlar A, B, C, D ve E olarak adlandırılmıştır. Konfigürasyonlar sırasıyla 133, 134, 138 ve 144 milyon parametreye sahiptir. Bunlar katman sayılarına göre isimlendirilmektedir. Örneğin 16 katmanlı olana Vgg16, 19 katmanlı olana Vgg19 denilmektedir.

Tablo 3.1. ESA mimarisinde VggNet modeline ait farklı konfigürasyonlar (Simonyan ve Zisserman 2014)

ConvNet Configuration					
A	A-LRN	B	C	D	E
11 weight layers	11 weight layers	13 weight layers	16 weight layers	16 weight layers	19 weight layers
input (224 × 224 RGB image)					
conv3-64	conv3-64 LRN	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64
maxpool					
conv3-128	conv3-128	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128
maxpool					
conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256 conv1-256	conv3-256 conv3-256 conv3-256	conv3-256 conv3-256 conv3-256 conv3-256
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 conv1-512	conv3-512 conv3-512 conv3-512	conv3-512 conv3-512 conv3-512 conv3-512
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 conv1-512	conv3-512 conv3-512 conv3-512	conv3-512 conv3-512 conv3-512 conv3-512
maxpool					
FC-4096					
FC-4096					
FC-1000					
soft-max					

3.3.4. GoogLeNet

GoogLeNet (Szegedy ve ark 2015), 22 katmandan oluşan bu model katmanları üst üste eklemek yerine paralel çalışan Inception mimarisini ortaya koymuştur. Temel amaç hesaplama maliyetini artırmadan ağın derinliğini ve genişliğini artırmaktır. Şekil 3.15'te gösterilen bu mimariye ait parametreler Tablo 3.2'de verilmiştir. Inception modüllerinden oluşan bu farklı mimari ILSVRC yarışmasının 2014 yılında %6.67 hata oranı ile kazanan modeli olmuştur. Inception mimarisi olarak da bilinen bu model 2014 yılında Szegedy ve ark. tarafından %6.67 hata oranı ile başarı sağlayan ESA modelidir.



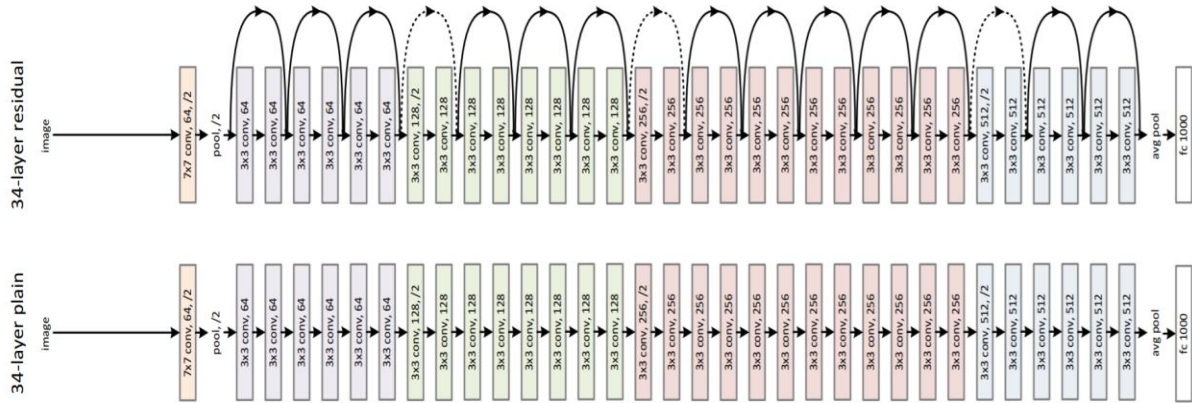
Şekil 3.15. ESA mimarisinde GoogLeNet/ Inception ağı (Szegedy ve ark 2015)

Tablo 3.2. ESA mimarisinde GoogLeNet parametreleri (Szegedy ve ark 2015)

type	patch size/ stride	output size	depth	#1×1	#3×3 reduce	#3×3	#5×5 reduce	#5×5	pool proj	params	ops
convolution	7×7/2	112×112×64	1							2.7K	34M
max pool	3×3/2	56×56×64	0								
convolution	3×3/1	56×56×192	2		64	192				112K	360M
max pool	3×3/2	28×28×192	0								
inception (3a)		28×28×256	2	64	96	128	16	32	32	159K	128M
inception (3b)		28×28×480	2	128	128	192	32	96	64	380K	304M
max pool	3×3/2	14×14×480	0								
inception (4a)		14×14×512	2	192	96	208	16	48	64	364K	73M
inception (4b)		14×14×512	2	160	112	224	24	64	64	437K	88M
inception (4c)		14×14×512	2	128	128	256	24	64	64	463K	100M
inception (4d)		14×14×528	2	112	144	288	32	64	64	580K	119M
inception (4e)		14×14×832	2	256	160	320	32	128	128	840K	170M
max pool	3×3/2	7×7×832	0								
inception (5a)		7×7×832	2	256	160	320	32	128	128	1072K	54M
inception (5b)		7×7×1024	2	384	192	384	48	128	128	1388K	71M
avg pool	7×7/1	1×1×1024	0								
dropout (40%)		1×1×1024	0								
linear		1×1×1000	1							1000K	1M
softmax		1×1×1000	0								

3.3.5. Microsoft RestNet

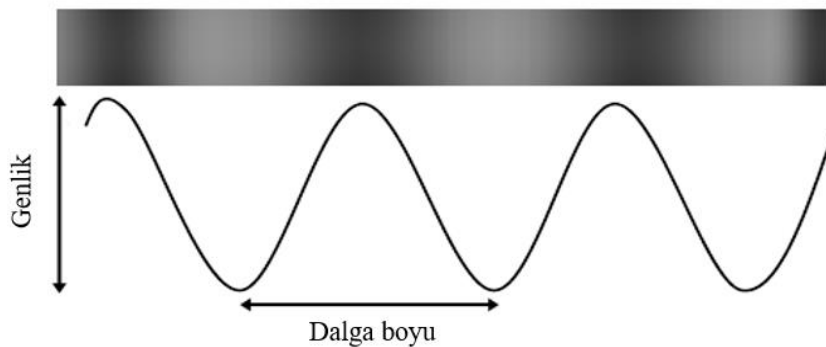
RestNet, Kaiming He ve ark. tarafından tasarlanan bu mimari 152 katmandan oluşan derin bir ağ mimarisine sahiptir. Bu model %3.6 hata oranı ile ILSVRC yarışmasının 2015 kazananı olmuştur. Şekil 3.16’da verilen bu mimari derin ağ mimarilerinde ara bağlantı yolları ile ağın kompleksliğini artırmadan doğruluk oranını arttırdığını sunmuşlardır (He ve ark 2016).



Şekil 3.16. ESA mimarisinde Microsoft RestNet ağı (He ve ark 2016)

3.4. Ses ve Özellikleri

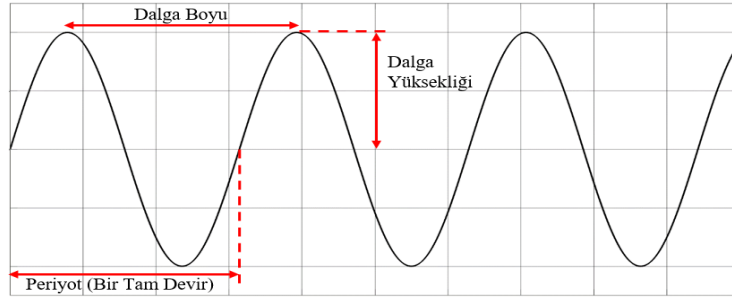
Ses, enerjinin uygulandığı yöne paralel olarak, hava moleküllerinin sıkışması ve seyrekleşmesine bağlı oluşan bir basınç dalgasıdır. Bu doğrultuda, enerji uygulanmasına bağlı olarak hava moleküllerinin normalden daha sıkı duruma geçtiği bölgelere alınmasına sıkıştırma, bu alanın normalden daha az yani sıkılaştırmanın olmadığı durumlar ise seyrekleşme bölgesi olarak tanımlanmaktadır (Huang ve ark 2001). Bu sıkılaştırma ve seyrekleşme durumlarına bağlı olarak ses dalgası Şekil 3.17’de olduğu gibi bir sinüs grafiği ile gösterilmektedir. Sıkılaştırmanın en fazla ve seyrekleşmenin en az olduğu anlar sinüs işaretinin tepe ve çukur noktalarına karşılık gelir.



Şekil 3.17. Hava moleküllerinin sıkılaştırma ve seyrekleşme durumlarına bağlı olarak oluşan ses dalgası (Huang ve ark 2001)

3.4.1. Sesin özellikleri

Ses dalgasının başlıca özellikleri; yayılma hızı, frekans, ses basıncı, dalga boyu ve genlik olarak sıralanabilir.



Şekil 3.18. Fiziksel özellikleri bakımından ses dalgası

3.4.2. Sesin yayılma hızı

Bir dalganın yayılma hızı, $c = \lambda / T$ eşitliği ile tanımlanır. Burada; c yayılma hızını, λ dalga boyunu ve T ise periyodu ifade etmektedir. Pratikte sesin yayılma hızı, ortamın yoğunluğundan ve ortamın sıcaklık ile basınç değişimlerinden etkilenir. Ortamda bulunan moleküllerin sıkıştırılma özelliği azaldıkça sesin yayılma hızı artmaktadır (MOREAU 2019). Örneğin katı bir ortamda moleküller birbirine daha sıkı bir şekilde bağlandığı için dalgalar molekülden moleküle daha hızlı geçecektir, bu nedenle ses daha hızlı yayılacaktır. Ortamdaki sıcaklığın da artmasıyla hava molekülleri daha büyük kinetik enerjiye sahip olacağı için bir ses dalgasının molekülden moleküle geçişi hızlanır. Herhangi bir ortamın yokluğunda ise sesin yayılması mümkün değildir. Bu nedenle, dalgaları bir kaynaktan alıcıya taşıyacak bir ortam olmadığında, örneğin uzay boşluğunda, ses yayılmaz (Karmakar 2011).

3.4.3. Frekans

Şekil 3.18’de gösterildiği gibi periyot, ses dalgasının kendini tekrarladığı süreye denir. Birimi saniyedir. Ses sinyali için periyodu ifade edecek olursak, hava moleküllerinin iki ardışık sıkışma veya seyrekleşme durumları arasında geçen süredir (MOREAU 2019). Frekans ise bir saniyedeki titreşim veya periyot sayısına denir. Birimi Hertz (Hz)’dir. Bir başka tanımla frekans, bir sinyalin bir saniyede kendini tekrarlama sayısı olarak adlandırılır. Ses sinyali için ifade edecek olursak, bir ses dalgasının frekansı, bir noktadan bir saniyede geçen ses dalgalarının sayısıdır (Karmakar 2011). 20-20.000 Hz arasındaki ses dalgaları duyulabilir ses frekans aralığındadır. Periyot ve frekans arasındaki ilişki Eşitlik 3.4’deki gibi tanımlanır.

$$f=1/T \quad (3.4)$$

3.4.4. Dalga boyu

Dalga boyu, ses hızı (c) ve periyodu (T) veya frekansı (f) ile formüle edilir ve Şekil 3.18’de gösterildiği gibi ses dalgasının ardışık iki maksimum veya minimum noktası arasındaki mesafeyi ifade eder (Karmakar 2011). Frekansla ters ilişkili olan dalga boyu λ ile gösterilir ve Eşitlik 3.5’teki gibi bir bağıntıya sahiptir:

$$\lambda = c/f = cT \quad (3.5)$$

3.4.5. Ses basıncı

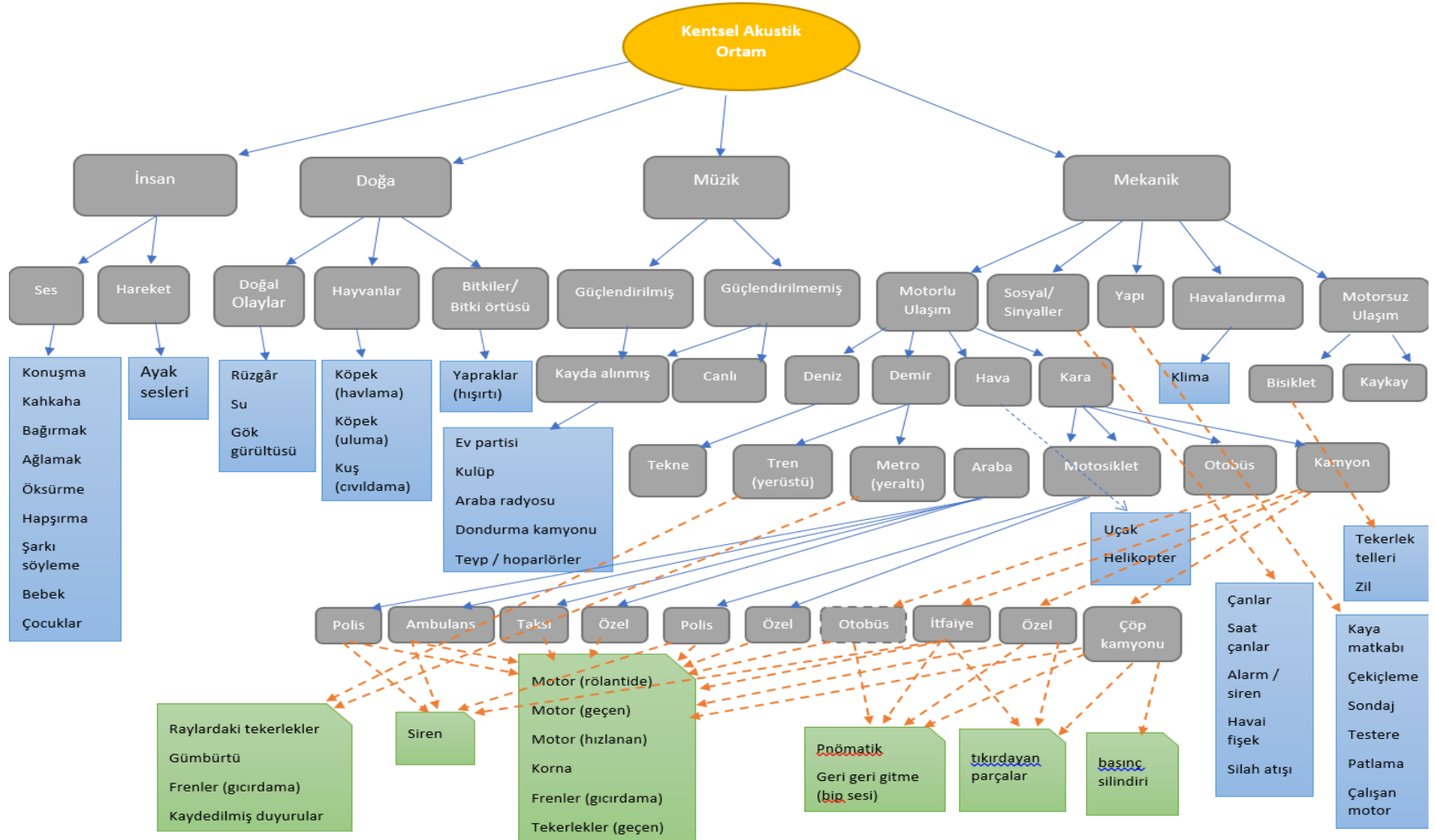
Ses basıncı, uzayda ses dalgası tarafından oluşturulan bir basınç değeridir. Birimi bar’dır. Newton/m² ya da milibar cinsinden ifade edilir.

3.4.6. Ses genliği ve şiddeti

Şekil 3.18’te gösterildiği gibi sinüs formunda ifade edilen bir ses dalgasının genliği, üst tepe ve alt tepe noktaları arasındaki mesafenin yarısına karşılık gelmektedir. Genlik, bir ses dalgasındaki enerjinin ölçüsüdür ve dalganın genliği ne kadar büyük olursa, ses şiddeti o kadar yüksek olur (Karmakar 2011). Ses şiddetinin birimi desibel (dB) olarak ifade edilir.

3.5. Kentsel Sesler

Kentsel sesler, şehir ortamlarındaki multimedya sensör ağları (Steele ve ark 2013) tarafından ve ayrıca çevresel sahneleri tasvir eden büyük miktarda çevrimiçi multimedya içeriği tarafından sağlanan seslerdir. Literatürde yapılan çalışmalara bakıldığında diğer ses sınıflandırmalara nispeten ÇSS’ler üzerinde daha az çalışmalar olduğu görülmektedir. Bunun temel sebepleri gerçek dünyadan alınmış yeterli veri setlerinin olmaması ve kentsel seslerin taksonomisinin belirlenememesidir (Salamon ve ark 2014). ÇSS’lerin taksonomisi Salamon ve ark. tarafından Şekil 3.19’da gösterildiği gibi elde edilmiştir. Yapılan bu çalışma ile beraber ÇSS’ler hakkında yeni veri setleri oluşturulmaya başlanmıştır. ÇSS’ler için yeni bir veri seti oluşturulurken, ses kayıtlarının çevresel gürültülerle beraber alınması ve belirli bir standartta bu verilerin oluşturulması çok önem arz etmektedir. Örneğin sokaktaki bir köpeğin sesinin sınıflandırılması için köpek ile aynı ortamda bulunan diğer nesnelerin de seslerinin var olması gerekmektedir. Böylelikle otomatik olarak sokaktaki havlayan bir köpek sesinin sınıflandırılması aşamasında kullanılacak modelin eğitimi için bütün sahne seslerinin köpek sesiyle beraber sisteme verilmesi gerekmektedir.



Şekil 3.19. ÇSS'lerin taksonomisi (Salamon ve ark 2014)

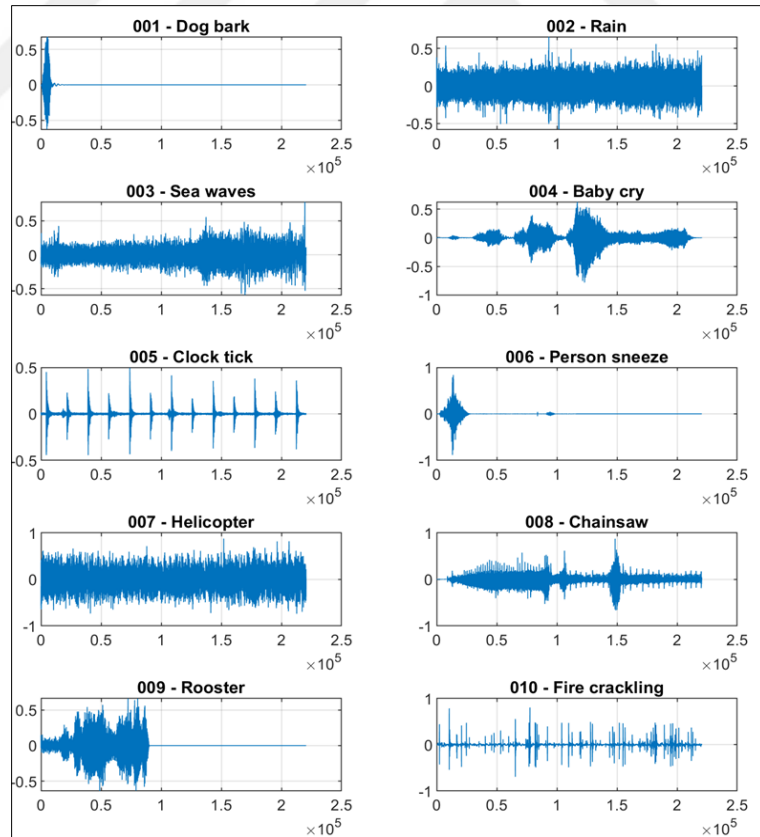
3.6. Veri Setleri

Literatürde son yıllarda yapılan çevresel seslerin sınıflandırılması için yoğun olarak kullanılan üç farklı veri seti bulunmaktadır. Bu veri setleri sırasıyla ESC10, ESC50 ve UrbanSaund8K'dır. Bu veri setleri hakkında detaylı bilgileri aşağıda verilmiştir.

3.6.1. ESC10 veri seti

ESC10 veri seti Karol J. Piczak tarafından oluşturulmuştur (Piczak 2015). Veri setinde toplam 10 sınıf bulunmaktadır bu sınıflar sırasıyla köpek havlama sesi, yağmur sesi, deniz dalga sesi, bebek ağlama sesi, saat tik-tok sesi, hapşırma sesi, helikopter sesi, elektrikli testere sesi, horoz sesi, ateş kıvılcım sesidir. Her bir sınıfta 40 tane ses kaydı bulunmaktadır. Her bir kayıt ortalama 5 sn uzunluktaki kayıttan oluşmaktadır. Toplamda her bir sınıf için ortalama 40 ses kayıt bulunmaktadır. Veri setindeki her bir sınıfa ait seslerin örnek bir gösterimi Şekil 3.20'de verilmiştir. Bu veri seti ESC50 veri setinin bir alt kümesidir.

Belli bir standardın elde edilmesi için ilk olarak bu veri seti oluşturulmuştur.



Şekil 3.20. ESC-10 veri setindeki sınıflara ait kayıtların grafiksel gösterimi

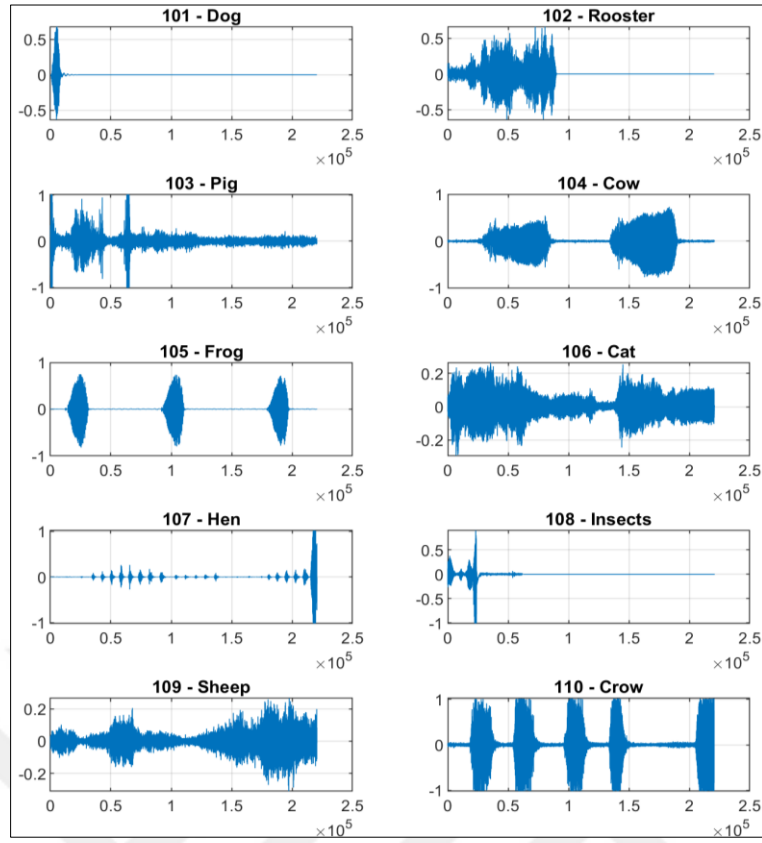
3.6.2. ESC50 veri seti

ESC50 veri seti ESC10 veri setinin geliştirilmiş halidir. Bu veri setinde 50 tane sınıf bulunmaktadır. Her bir sınıf 5 saniye uzunlukta 40 tane kayıttan oluşmaktadır. 2000 tane kentsel ortamlardan elde edilmiş ses kliplerine sahiptir. Bu sınıflar hayvanlar, doğal ses manzaraları ve su sesleri, insan konuşma dışı sesler, iç/dış mekân sesleri ve dış şehir gürültüsü ana kategorilerinde oluşmaktadır. Bu ana kategoriler altındaki her bir sınıfın ismi Tablo 3.3'te verilmiştir.

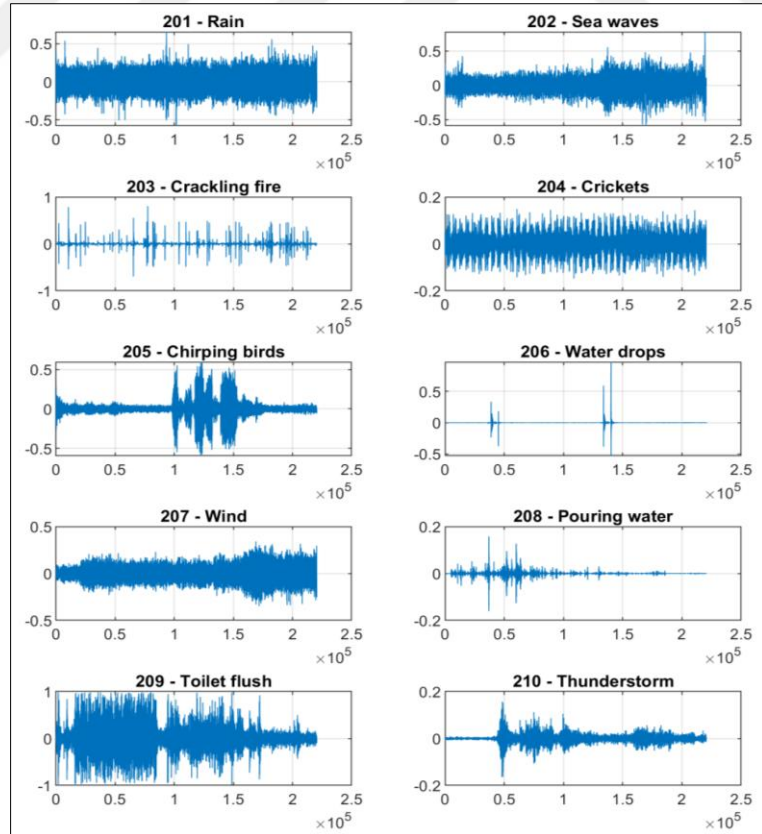
Tablo 3.3. ESC50 veri setindeki kategoriler ve her bir kategorideki sınıfların isimleri

Animals	Natural soundspaces & water sounds	Human, Non-speech sounds	Interior/domestic sounds	Exterior/urban noises
Dog	Rain	Crying baby	Door knock	Helicopter
Rooster	Sea waves	Sneezing	Mouse click	Chainsaw
Pig	Crackling fire	Clapping	Keyboard typing	Siren
Cow	Crickets	Breathing	Door, wood creaks	Car horn
Frog	Chirping birds	Coughing	Can opening	Engine
Cat	Water drops	Footsteps	Washing machine	Train
Hen	Wind	Laughing	Vacuum cleaner	Church bells
Insects (flying)	Pouring water	Brushing teeth	Clock alarm	Airplane
Sheep	Toilet flush	Snoring	Clock tick	Fireworks
Crow	Thunderstorm	Drinking, sipping	Glass breaking	Hand saw

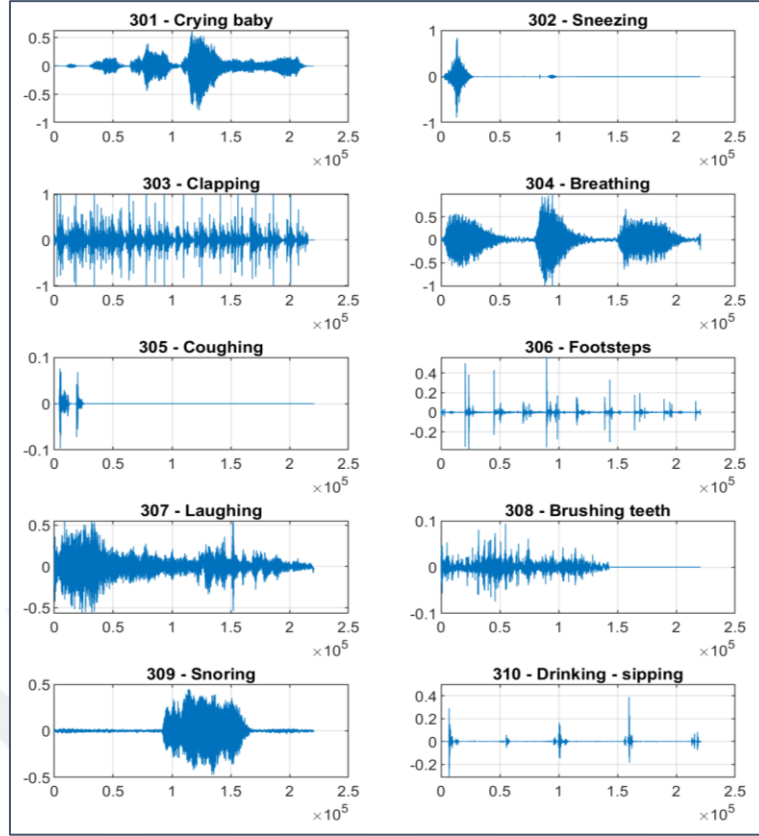
ESC50 veri setindeki her bir sınıfın ses kayıtlarına ait örnek bir grafik gösterimi Şekil 3.21'de verilmiştir.



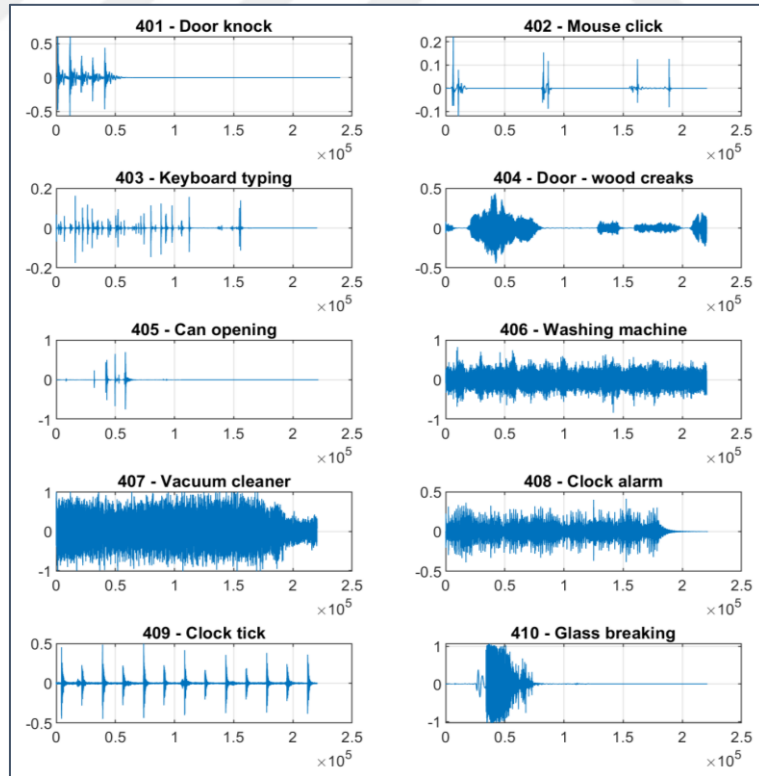
Şekil 3.21. (a) ESC50 veri setindeki her bir sınıfın ses kayıtlarına ait örnek bir grafik gösterimi



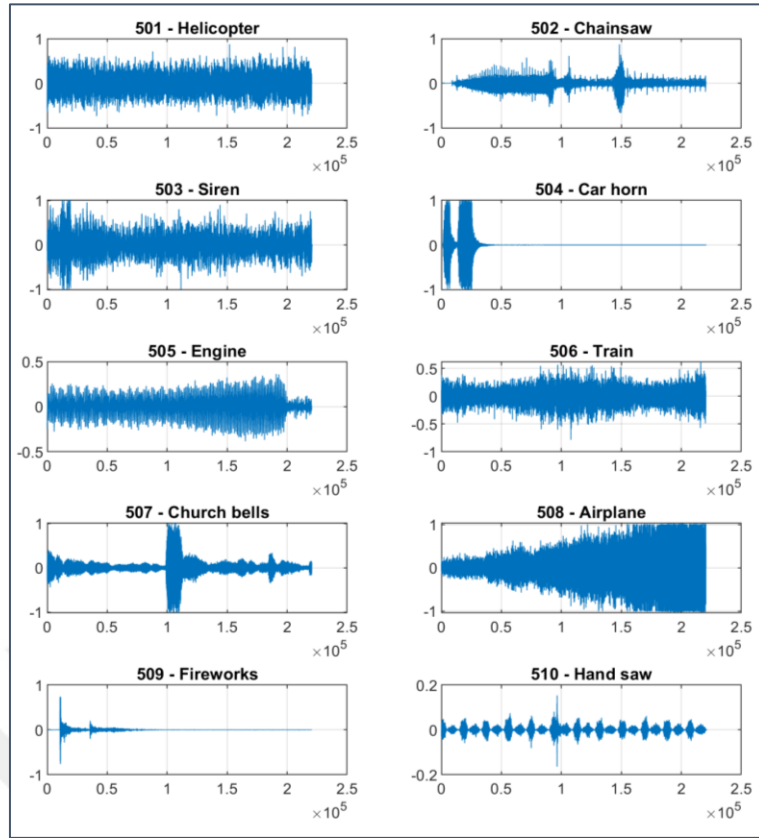
Şekil 3.21. (b) ESC50 veri setindeki her bir sınıfın ses kayıtlarına ait örnek bir grafik gösterimi



Şekil 3.21. (c) ESC50 veri setindeki her bir sınıfın ses kayıtlarına ait örnek bir grafik gösterimi



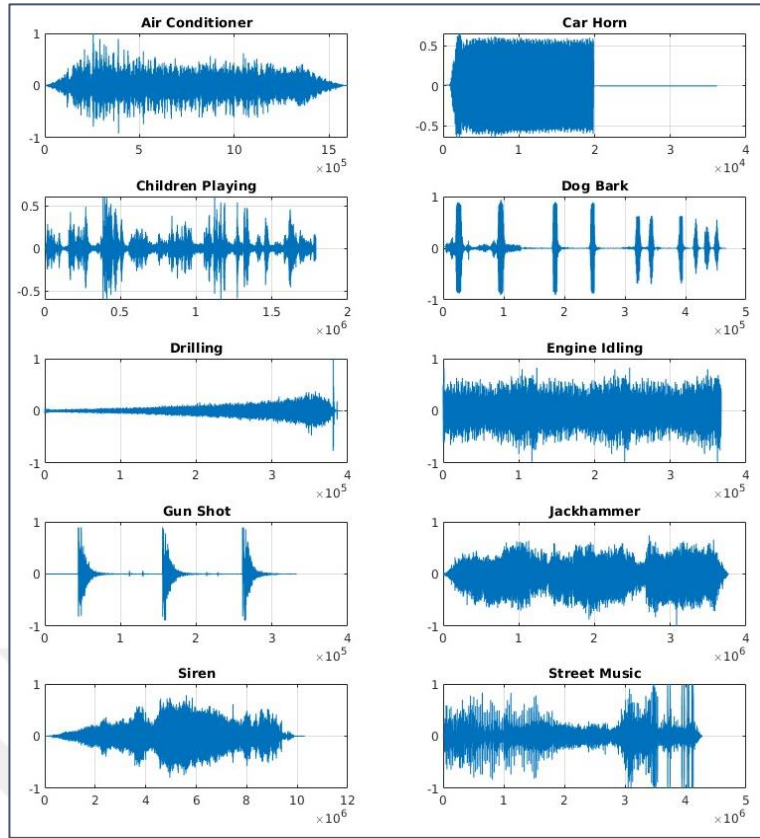
Şekil 3.21. (d) ESC50 veri setindeki her bir sınıfın ses kayıtlarına ait örnek bir grafik gösterimi



Şekil 3.21. (e) ESC50 veri setindeki her bir sınıfın ses kayıtlarına ait örnek bir grafik gösterimi

3.6.3. UrbanSound8K

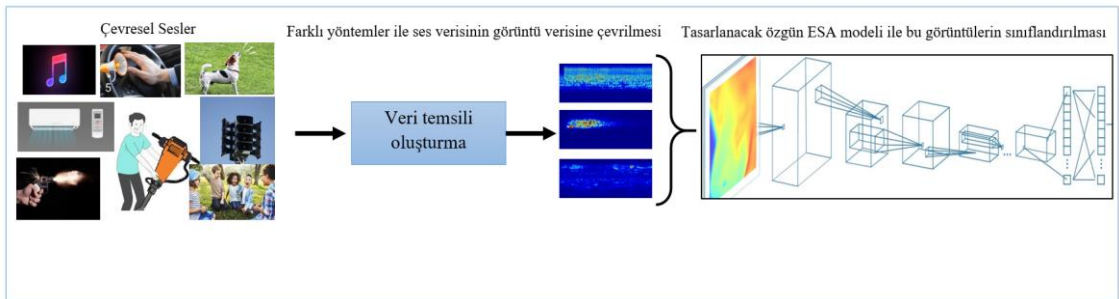
UrbanSaund8K veri seti Salamon ve ark. tarafından hazırlanan 8732 etiketli seslerden oluşan bir veri setidir (Salamon ve ark 2014). Veri setindeki her bir kayıt yaklaşık 4 saniye uzunluktadır. Bu veri seti www.freesound.org sitesinde yüklenen kayıtlar içerisinde elde edilmiştir. Veri seti toplamda 10 sınıfa sahiptir. Bu sınıflar sırasıyla; air_conditioner (klima), car_horn (araba kornası), children_playing (oynayan çocuklar), dog_bark (köpek havlaması), drilling (delme), enginge_idling (motor rölantisi), gun_shot (silah sesi), jackhammer (delici-kırıcı), siren (siren), ve street_music (sokak müziği). Bu veri setindeki her bir sınıfa ait seslerin grafiksel bir gösterimi Şekil 3.22’de verilmiştir.



Şekil 3.22. UrbanSaund8k veri setindeki sınıflardaki her bir kayda ait örnek gösterim

4. ÖNERİLEN YÖNTEM

Yapılan tez çalışması için önerilen yöntemin mantıksal olarak tasarımı Şekil 4.1’de verilmiştir. Önerilen yöntem üç aşamadan oluşmaktadır. Birinci aşamada veri setinin belirlenmesi, ikinci aşamada ESA modellerinde kullanılmak üzere veri setlerinde ön işlemlerin yapılması son aşamada ise bu veri setlerinin sınıflandırılması için özgün ESA modellerinin tasarlanması, eğitilmesi ve test edilmesini kapsamaktadır. Birinci aşamada Bölüm 3’te belirtildiği gibi literatürde sıklıkla kullanılan ve belli standartlarda elde edilmiş üç farklı veri seti kullanılmıştır. İkinci aşamada ESA modellerinin eğitilmesi için bu veri setlerindeki her bir kayıt sinyal formatında görüntü formatına çevrilmiştir. Çevirim işlemi Matlab Wavelet modülü ile gerçekleştirilmiştir. Ses sinyalleri bu modülde ki scalogram yöntemi ile gerçekleştirilmiştir. Scalogram, zaman ve frekansın bir fonksiyonu olarak çizilen bir sinyalin Sürekli Dalgacık Dönüşümünün (CWT) mutlak değeridir. ESC10 veri setine ait sınıflardaki ses kayıtlarının scalogram karşılıkları için örnek bir gösterim Şekil 3.20’de verilmiştir. Üçüncü aşamadaki ESA modellerinin eğitilmesi için her üç veri setindeki görüntüler 32x32x3, 224x224x3 boyutlarında olmak üzere iki farklı veri seti oluşturulmuştur. Böylelikle ESA modellerinin eğitilmesi için altı farklı veri seti elde edilmiştir. Üçüncü aşamada ikinci aşamada elde edilen veri setlerine uygun olarak ESA modelleri tasarlanmıştır. Her bir veri seti için birden fazla ESA modelleri tasarlanmış olup bu modeller içerisinde en iyi performansı sergileyen model kaydedilmiştir.

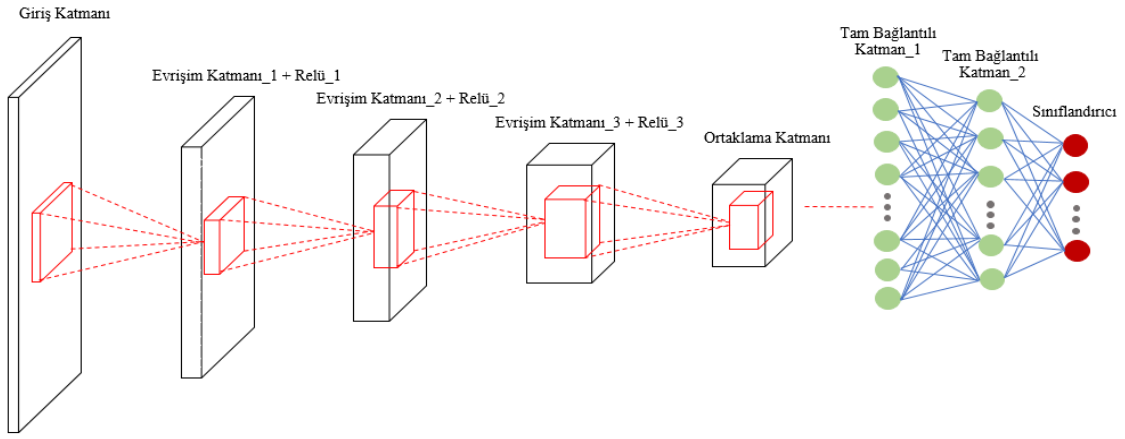


Şekil 4.1. Tezde uygulanacak yöntemin mantıksal tasarımı

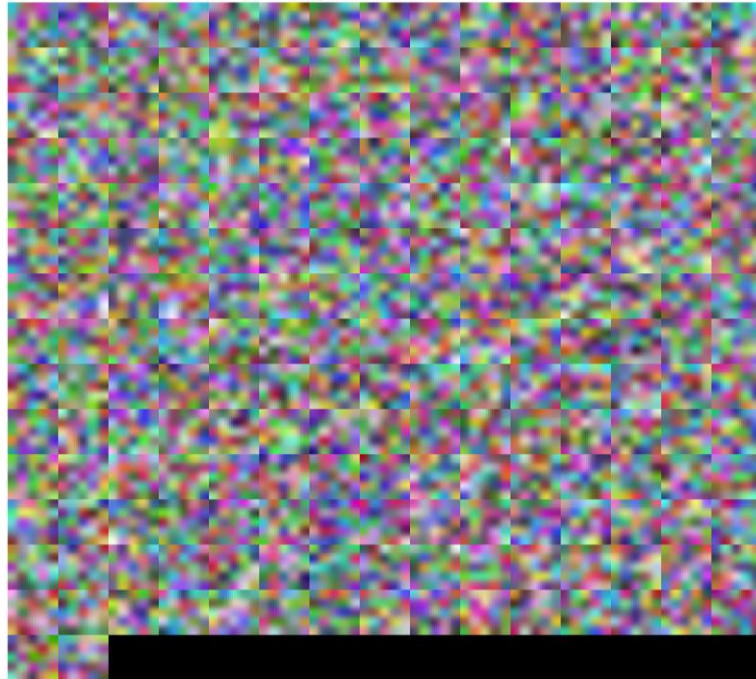
Şekil 4.1’de gösterilen yöntemin üçüncü aşaması için, en iyi sonucu veren altı adet ESA modeli elde edilmiştir. Bu modeller sırasıyla ESC10_ESA32, ESC10_ESA224, ESC50_ESA32, ESC50_ESA224, URBANSOUND8K_ESA32, URBANSOUND8K_ESA224 şeklinde isimlendirilmiştir. Bu modellerin tasarım yapıları ve parametreleri hakkında detaylı bilgi aşağıda verilmiştir.

4.1. ESC10_ESA32

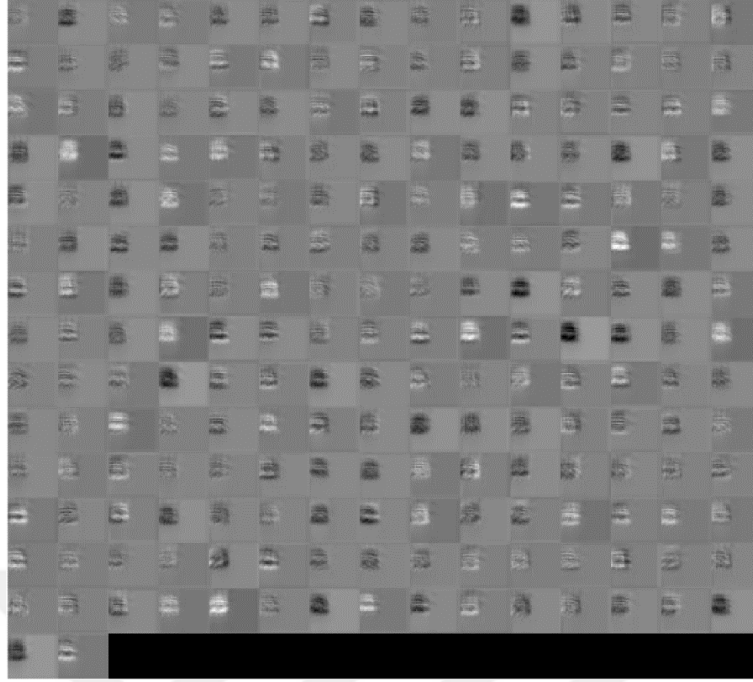
Bu model ESC10 veri setinin 32x32x3 giriş görüntü boyutuna sahip veriler için tasarlanmıştır. Modelin tasarım mimarisi Şekil 4.2’de verilmiştir. Bu mimarideki birinci evrişim katmanındaki filtreler Şekil 4.3 ve bu filtrelerin giriş görüntüsüne uygulamasından sonra elde edilen özellik haritaları Şekil 4.4’te verilmiştir. Bu mimariye ait bütün parametreler Tablo 4.1’de verilmiştir.



Şekil 4.2. ESC10_ESA32 modelinin mimarisi



Şekil 4.3. ESC10_ESA32 modelinin birinci evrişim katmanındaki filtreler



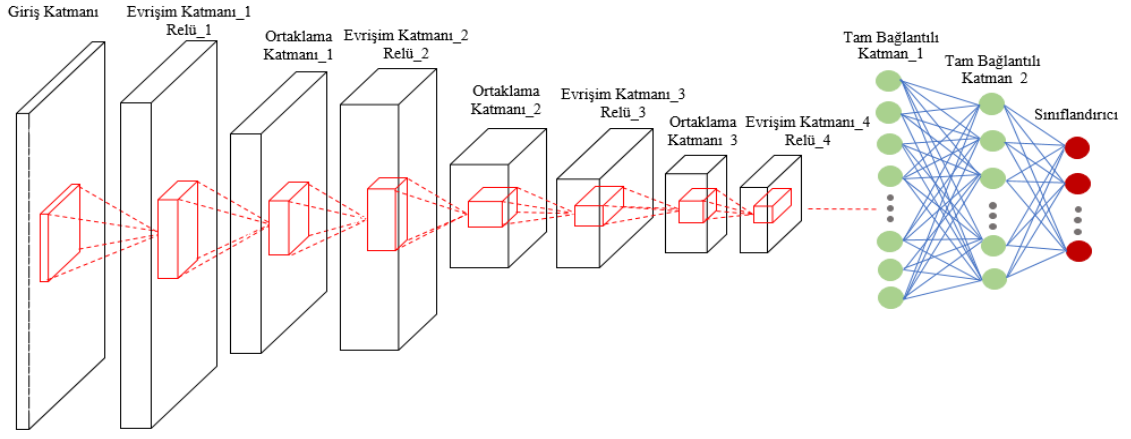
Şekil 4.4. ESC10_ESA32 modelinin birinci evrişim katmanındaki filtrelerin giriş görüntüsüne uygulandıktan sonra elde edilen özellik haritaları

Tablo 4.1. ESC10_ESA32 modelindeki her bir katmana ait parametre değerleri

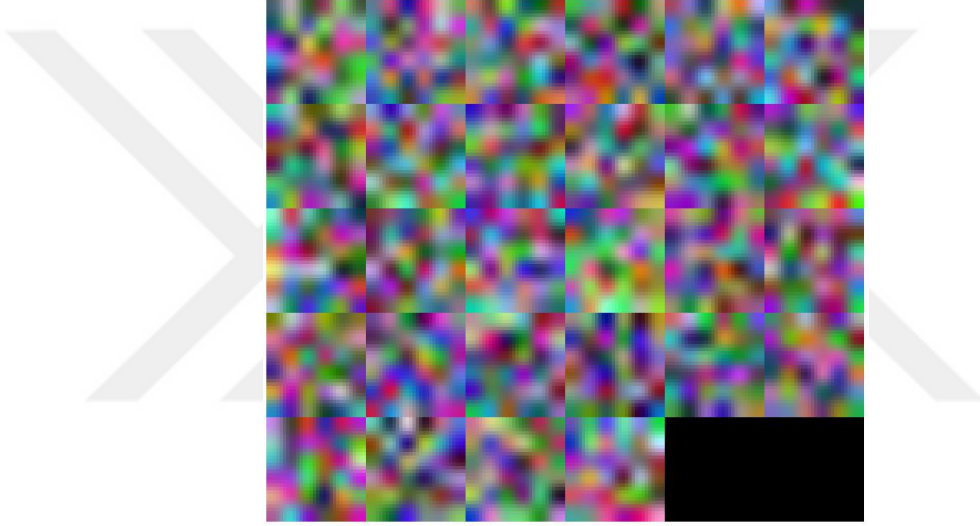
Katman Sırası	Katman Adı	Boyut	Filtre Sayısı	Filtre Boyutu	Adım Sayısı	Parametreler	Toplam Parametre Sayısı
1	Giriş Katmanı	32x32x3	-	-	-	-	0
2	Evrişim Katmanı_1	32x32x212	212	5x5	1x1	5x5x3x212	15900
3	Relü katmanı_1	32x32x212	-	-	-	-	0
4	Evrişim Katmanı_2	32x32x71	71	4x4	1x1	4x4x212x71	240832
5	Relü katmanı_2	32x32x71	-	-	-	-	0
6	Evrişim Katmanı_3	32x32x191	191	6x6	1x1	6x6x71x191	488196
7	Relü katmanı_3	32x32x191	-	-	-	-	0
8	Ortalama Ortaklama	8x8x191	-	2x2	4x4	-	0
9	Tam Bağlantılı Katman_1	1x1x893	-	-	-	893x12224	10916032
10	Düğüm seyreltme katmanı	1x1x893	-	-	-	-	0
11	Relü katmanı_4	1x1x893	-	-	-	-	0
12	Tam Bağlantılı Katman_2	1x1x50	-	-	-	50x893	44650
13	Softmax	1x1x50	-	-	-	-	0
14	Sınıflandırma Katmanı	1x1x50	-	-	-	-	0

4.2. ESC10_ESA224

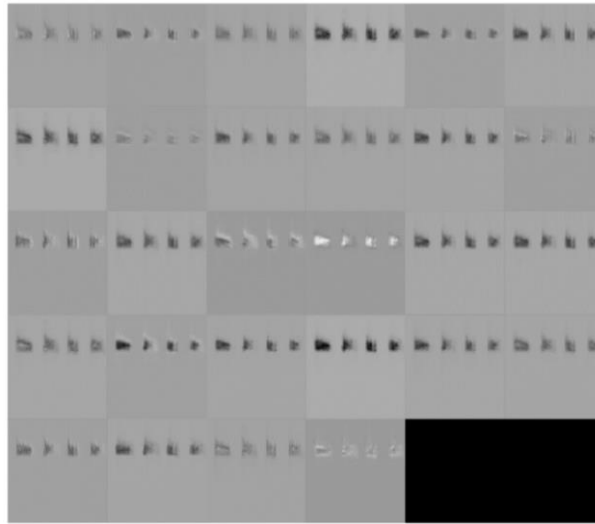
Bu model ESC10 veri setinin 224x224x3 giriş görüntü boyutuna sahip veriler için tasarlanmıştır. Modelin tasarım mimarisi Şekil 4.5'te verilmiştir. Bu mimarideki birinci evrişim katmanındaki filtreler Şekil 4.6 ve bu filtrelerin giriş görüntüsüne uygulanmasından sonra elde edilen özellik haritaları Şekil 4.7'de verilmiştir. Bu mimariye ait bütün parametreler Tablo 4.2'de verilmiştir.



Şekil 4.5. ESC10_ESA224 modelinin mimarisi



Şekil 4.6. ESC10_ESA224 modelinin birinci evrişim katmanındaki filtreler



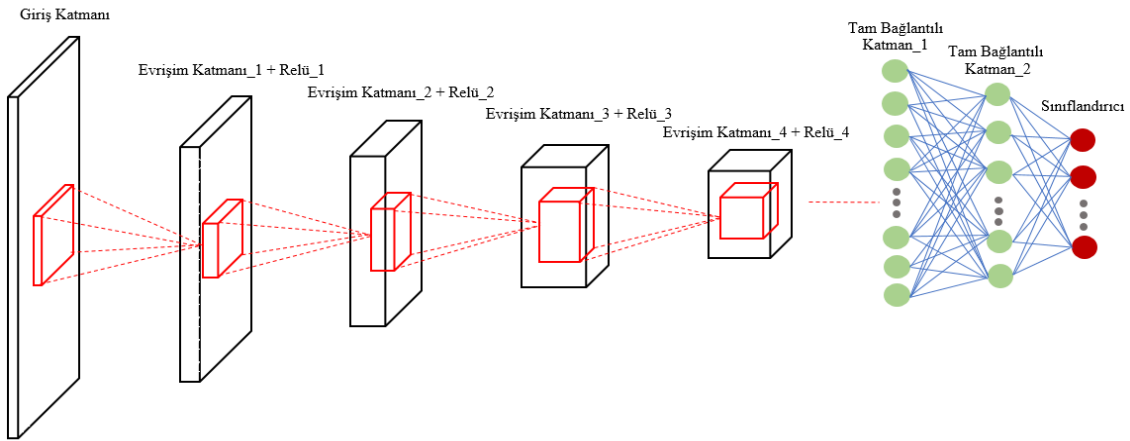
Şekil 4.7. ESC10_ESA224 modelinin birinci evrişim katmanındaki filtrelerin giriş görüntüsüne uygulandıktan sonra elde edilen özellik haritaları

Tablo 4.2. ESC10_ESA224 modelindeki her bir katmana ait parametre değerleri

Katman Sırası	Katman Adı	Boyut	Filtre Sayısı	Filtre Boyutu	Adım Sayısı	Parametreler	Toplam Parametre Sayısı
1	Giriş Katmanı	224x224x3	-	-	-	-	0
2	Evrişim Katmanı_1	224x224x28	28	6x6	1x1	6x6x3x28	3024
3	Relü katmanı_1	224x224x28	-	-	-	-	0
4	Maksimum Ortaklama Katmanı_1	111x111x28	-	4x4	2x2	-	0
5	Evrişim Katmanı_2	111x111x174	174	2x2	1x1	2x2x28x174	19488
6	Relü katmanı_2	111x111x174	-	-	-	-	0
7	Maksimum Ortaklama Katmanı_2	36x36x174	-	6x6	3x3	-	0
8	Evrişim Katmanı_3	36x36x129	129	6x6	1x1	6x6x174x129	808056
9	Relü katmanı_3	36x36x129	-	-	-	-	0
10	Maksimum Ortaklama Katmanı_3	5x5x129	-	4x4	7x7	-	0
11	Evrişim Katmanı_4	5x5x80	80	5x5	1x1	5x5x129x80	258000
12	Relü katmanı_4	5x5x80	-	-	-	-	0
13	Tam Bağlantılı Katman_1	1x1x897	-	-	-	897x2000	1794000
14	Düğüm Seyreltme Katmanı	1x1x897	-	-	-	-	0
15	Relü katmanı_5	1x1x897	-	-	-	-	0
16	Tam Bağlantılı Katman_2	1x1x10	-	-	-	10x897	8970
17	Softmax	1x1x10	-	-	-	-	0
18	Sınıflandırma Katmanı	1x1x10	-	-	-	-	0

4.3. ESC50_ESA32

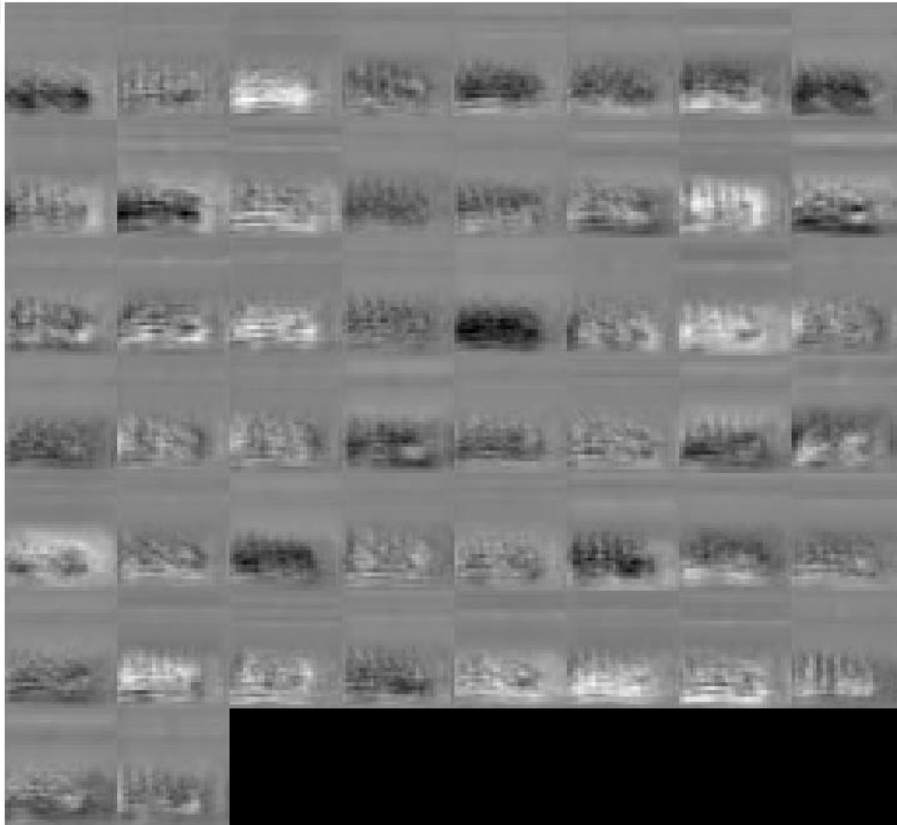
Bu model ESC50 veri setinin 32x32x3 giriş görüntü boyutuna sahip veriler için tasarlanmıştır. Modelin tasarım mimarisi Şekil 4.8’de verilmiştir. Bu mimarideki birinci evrişim katmanındaki filtreler Şekil 4.9 ve bu filtrelerin giriş görüntüsüne uygulanmasından sonra elde edilen özellik haritaları Şekil 4.10’da verilmiştir. Bu mimariye ait bütün parametreler Tablo 4.3’te verilmiştir.



Şekil 4.8. ESC50_ESA32 modelinin mimarisi



Şekil 4.9. ESC50_ESA32 modelinin birinci evrişim katmanındaki filtreler



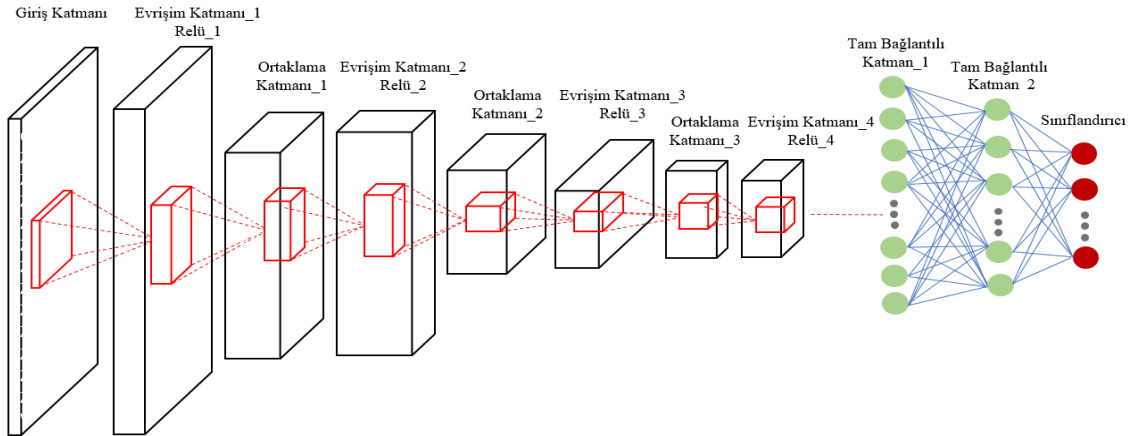
Şekil 4.10. ESC50_ESA32 modelinin birinci evrişim katmanındaki filtrelerin giriş görüntüsüne uygulandıktan sonra elde edilen özellik haritaları

Tablo 4.3. ESC50_ESA32 modelindeki her bir katmana ait parametre değerleri

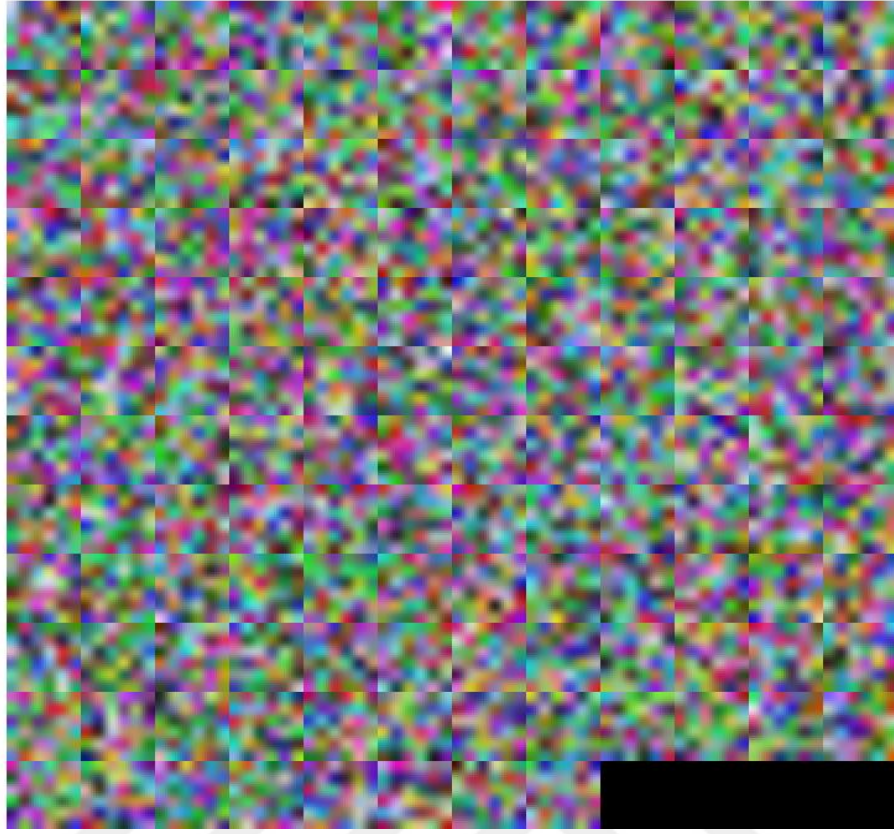
Katman Sırası	Katman Adı	Boyut	Filtre Sayısı	Filtre Boyutu	Adım Sayısı	Parametreler	Toplam Parametre Sayısı
1	Giriş Katmanı	32x32x3	-	-	-	-	0
2	Evrişim Katmanı_1	32x32x50	50	7x7	1x1	7x7x3x50	7350
3	Relü katmanı_1	32x32x50	-	-	-	-	0
4	Evrişim Katmanı_2	32x32x86	86	5x5	1x1	5x5x50x86	107500
5	Relü katmanı_2	32x32x86	-	-	-	-	0
6	Evrişim Katmanı_3	30x30x90	90	3x3	1x1	3x3x86x90	69660
7	Relü katmanı_3	30x30x90	-	-	-	-	0
8	Evrişim Katmanı_4	25x25x255	255	6x6	1x1	6x6x90x255	826200
9	Relü katmanı_4	25x25x255	-	-	-	-	0
10	Tam Bağlantılı Katman_1	1x1x574	-	-	-	574x159375	91481250
11	Düğüm Seyreltme Katmanı	1x1x574	-	-	-	-	0
12	Relü katmanı_5	1x1x574	-	-	-	-	0
13	Tam Bağlantılı Katman_2	1x1x50	-	-	-	50x574	28700
14	Softmax	1x1x50	-	-	-	-	0
15	Sınıflandırma Katmanı	1x1x50	-	-	-	-	0

4.4. ESC50_ESA224

Bu model ESC50 veri setinin 224x224x3 giriş görüntü boyutuna sahip veriler için tasarlanmıştır. Modelin tasarım mimarisi Şekil 4.11’de verilmiştir. Bu mimarideki birinci evrişim katmanındaki filtreler Şekil 4.12 ve bu filtrelerin giriş görüntüsüne uygulanmasından sonra elde edilen özellik haritaları Şekil 4.13’te verilmiştir. Bu mimariye ait bütün parametreler Tablo 4.4’te verilmiştir.



Şekil 4.11. ESC50_ESA224 modelinin mimarisi



Şekil 4.12. ESC50_ESA224 modelinin birinci evrişim katmanındaki filtreler



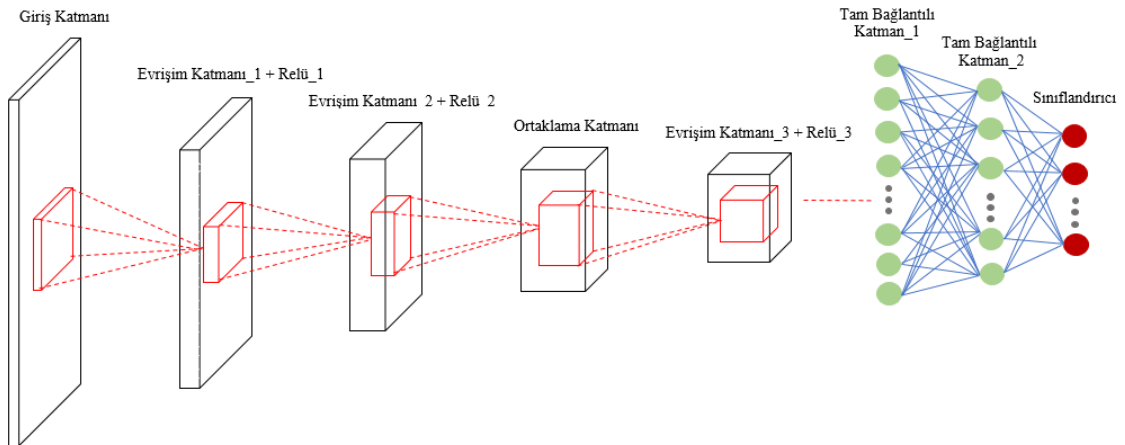
Şekil 4.13. ESC50_ESA224 modelinin birinci evrişim katmanındaki filtrelerin giriş görüntüsüne uygulandıktan sonra elde edilen özellik haritaları

Tablo 4.4. ESC50_ESA224 modelindeki her bir katmana ait parametre değerleri

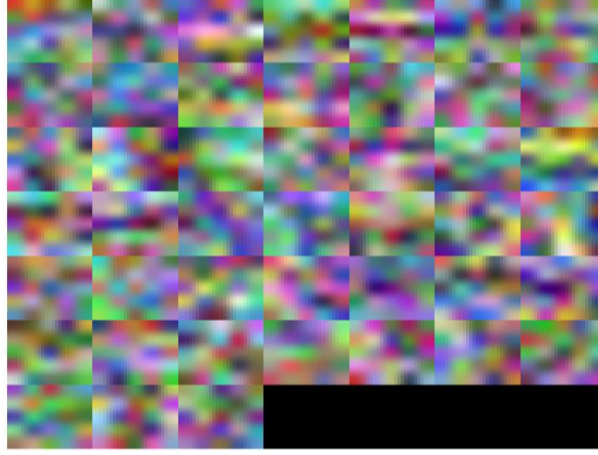
Katman Sırası	Katman Adı	Boyut	Filtre Sayısı	Filtre Boyutu	Adım Sayısı	Parametreler	Toplam Parametre Sayısı
1	Giriş Katmanı	224x224x3	-	-	-	-	0
2	Evrişim Katmanı 1	224x224x140	140	6x6	1x1	6x6x3x140	15120
3	Yığın Normalizasyon Katmanı 1	224x224x140				1x1x140	140
4	Relü katmanı 1	224x224x140	-	-	-	-	0
5	Ortalama Ortaklama Katmanı 1	45x45x140	-	4x4	5x5	-	0
6	Evrişim Katmanı 2	45x45x205	205	10x10	1x1	10x10x140x205	2870000
7	Yığın Normalizasyon Katmanı 2	45x45x205				1x1x205	205
8	Relü katmanı 2	45x45x205	-	-	-	-	0
9	Maksimum Ortaklama Katmanı 2	21x21x205	-	4x4	2x2	-	0
10	Evrişim Katmanı 3	21x21x183	183	9x9	1x1	9x9x205x183	3038715
11	Yığın Normalizasyon Katmanı 3	21x21x183				1x1x183	183
12	Relü katmanı 3	21x21x183	-	-	-	-	0
13	Maksimum Ortaklama Katmanı 3	3x3x183	-	5x5	7x7	-	0
14	Evrişim Katmanı 4	3x3x103	103	10x10	1x1	10x10x183x103	1884900
15	Yığın Normalizasyon Katmanı 4	3x3x103				1x1x103	103
16	Relü katmanı 4	3x3x103	-	-	-	-	0
17	Tam Bağlantılı Katman 1	1x1x763	-	-	-	763x927	707301
18	Relü katmanı 5	1x1x763	-	-	-	-	0
19	Düğüm seyreltme katmanı	1x1x763	-	-	-	-	0
20	Tam Bağlantılı Katman 2	1x1x50	-	-	-	50x763	38150
21	Softmax	1x1x50	-	-	-	-	0
22	Sınıflandırma Katmanı	1x1x50	-	-	-	-	0

4.5. URBANSOUND8K_ESA32

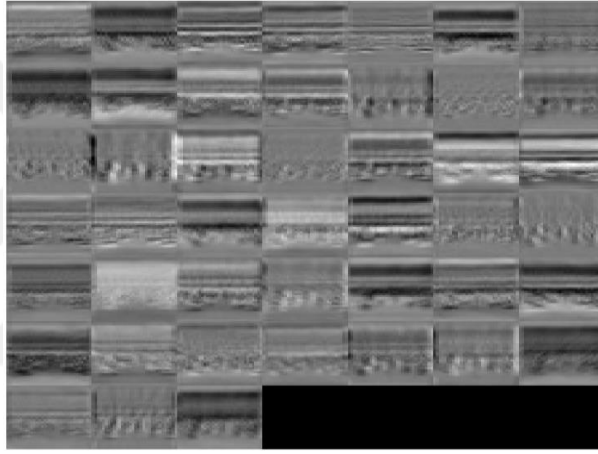
Bu model URBANSOUND8K_ESA32 veri setinin 32x32x3 giriş görüntü boyutuna sahip veriler için tasarlanmıştır. Modelin tasarım mimarisi Şekil 4.14'te verilmiştir. Bu mimarideki birinci evrişim katmanındaki filtreler Şekil 4.15 ve bu filtrelerin giriş görüntüsüne uygulanmasından sonra elde edilen özellik haritaları Şekil 4.16'da verilmiştir. Bu mimariye ait bütün parametreler Tablo 4.5'te verilmiştir.



Şekil 4.14. URBANSOUND8K_ESA32 modelinin mimarisi



Şekil 4.15. URBANSOUND8K_ESA32 modelinin birinci evrişim katmanındaki filtreler



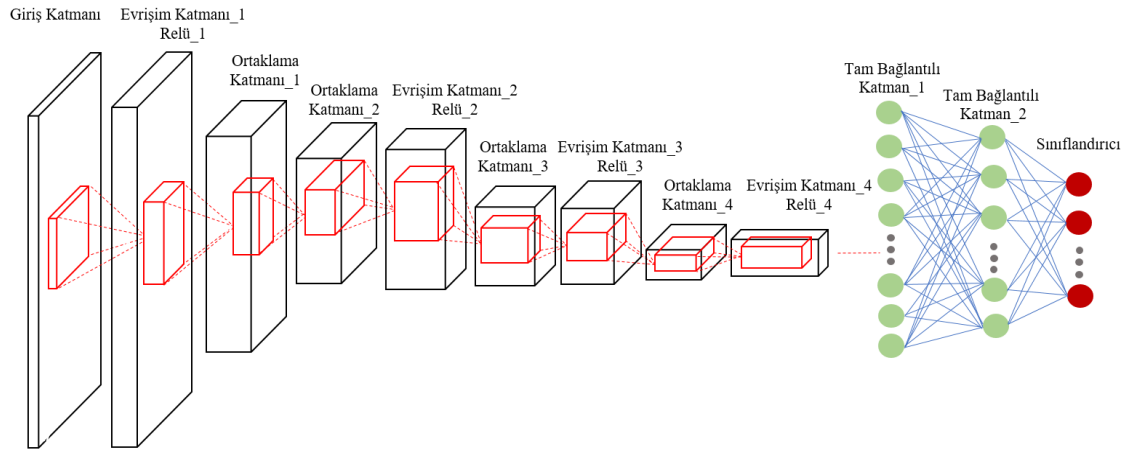
Şekil 4.16. URBANSOUND8K_ESA32 modelinin birinci evrişim katmanındaki filtrelerin giriş görüntüsüne uygulandıktan sonra elde edilen özellik haritaları

Tablo 4.5. URBANSOUND8K_ESA32 modelindeki her bir katmana ait parametre değerleri

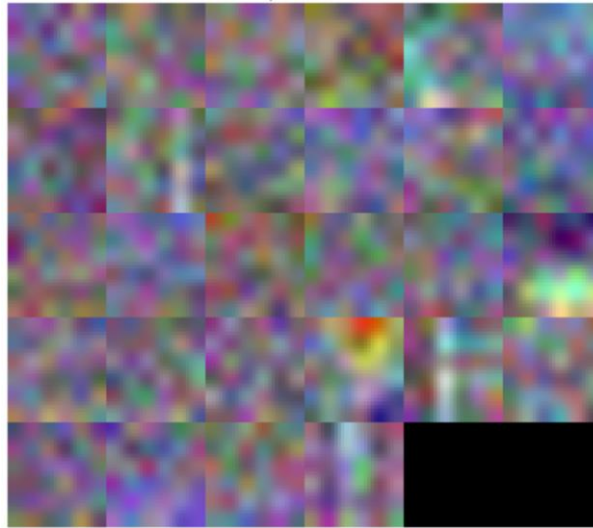
Katman Sırası	Katman Adı	Boyut	Filtre Sayısı	Filtre Boyutu	Adım Sayısı	Parametreler	Toplam Parametre Sayısı
1	Giriş Katmanı	32x32x3	-	-	-	-	0
2	Evrişim Katmanı_1	32x32x45	45	5x5	1x1	5x5x3x45	3375
3	Relü katmanı_1	32x32x45	-	-	-	-	0
4	Evrişim Katmanı_2	32x32x193	193	8x8	1x1	8x8x45x193	555840
5	Relü katmanı_2	32x32x193	-	-	-	-	0
6	Ortalama Ortaklama Katmanı	10x10x193	-	3x3	3x3	-	0
7	Evrişim Katmanı_3	10x10x139	139	10x10	1x1	10x10x193x139	2682700
8	Relü katmanı_3	10x10x139	-	-	-	-	0
9	Tam Bağlantılı Katman_1	1x1x891	-	-	-	891x13900	12384900
10	Düğüm Seyreltme Katmanı	1x1x891	-	-	-	-	0
11	Relü katmanı_4	1x1x891	-	-	-	-	0
12	Tam Bağlantılı Katman_2	1x1x10	-	-	-	10x891	8910
13	Softmax	1x1x10	-	-	-	-	0
14	Sınıflandırma Katmanı	1x1x10	-	-	-	-	0

4.6. URBANSOUND8K_ESA224

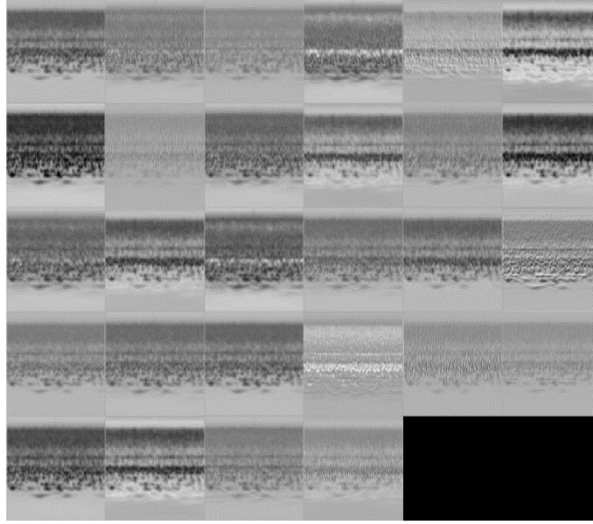
Bu model URBANSOUND8K_ESA224 veri setinin 224x224x3 giriş görüntü boyutuna sahip veriler için tasarlanmıştır. Modelin tasarım mimarisi Şekil 4.17’de verilmiştir. Bu mimarideki birinci evrişim katmanındaki filtreler Şekil 4.18 ve bu filtrelerin giriş görüntüsüne uygulanmasından sonra elde edilen özellik haritaları Şekil 4.19’da verilmiştir. Bu mimariye ait bütün parametreler Tablo 4.6’da verilmiştir.



Şekil 4.17. URBANSOUND8K_ESA224 modelinin mimarisi



Şekil 4.18. URBANSOUND8K_ESA224 modelinin birinci evrişim katmanındaki filtreler



Şekil 4.19. URBANSOUND8K_ESA224 modelinin birinci evrişim katmanındaki filtrelerin giriş görüntüsüne uygulandıktan sonra elde edilen özellik haritaları

Tablo 4.6. URBANSOUND8K_ESA224 modelindeki her bir katmana ait parametre değerleri

Katman Sırası	Katman Adı	Boyut	Filtre Sayısı	Filtre Boyutu	Adım Sayısı	Parametreler	Toplam Parametre Sayısı
1	Giriş Katmanı	224x224x3	-	-	-	-	0
2	Evrişim Katmanı_1	224x224x28	28	6x6	1x1	6x6x3x28	3024
3	Relü katmanı_1	224x224x28	-	-	-	-	0
4	Maksimum Ortaklama Katmanı_1	111x111x28	-	4x4	2x2	-	0
	Maksimum Ortaklama Katmanı_2	36x36x28	-	6x6	3x3	-	0
5	Evrişim Katmanı_2	36x36x174	174	2x2	1x1	2x2x28x174	19488
6	Relü katmanı_2	36x36x174	-	-	-	-	0
7	Maksimum Ortaklama Katmanı_3	11x11x174	-	6x6	3x3	-	0
8	Evrişim Katmanı_3	11x11x129	129	6x6	1x1	6x6x174x129	808056
9	Relü katmanı_3	11x11x129	-	-	-	-	0
10	Maksimum Ortaklama Katmanı_4	4x4x129	-	4x4	2x2	-	0
11	Evrişim Katmanı_4	4x4x180	180	5x5	1x1	5x5x129x180	580500
12	Relü katmanı_4	4x4x180	-	-	-	-	0
13	Tam Bağlantılı Katman_1	1x1x897	-	-	-	897x2880	2583360
14	Düğüm Seyreletme Katmanı	1x1x897	-	-	-	-	0
15	Relü katmanı_5	1x1x897	-	-	-	-	0
16	Tam Bağlantılı Katman_2	1x1x10	-	-	-	10x897	8970
17	Softmax	1x1x10	-	-	-	-	0
18	Sınıflandırma Katmanı	1x1x10	-	-	-	-	0

5. DENEYSEL ÇALIŞMALAR

Yapılan tez kapsamında kentsel seslerin sınıflandırılması için üç farklı veri seti üzerinde farklı ESA modelleri giriş görüntü boyutuna göre eğitim işlemleri gerçekleştirilmiştir. 32x32x3 ve 224x224x3 giriş görüntü boyutuna sahip veri setleri üzerinde 10-Kat Çapraz Doğrulama (10-Fold Cross Validation) yapılarak modeller eğitilmiştir. ESA modellerinin eğitilmesi için Matlab R2020a yazılımının Derin Öğrenme kütüphanesi kullanılmıştır. Yapılan tüm eğitim ve test işlemleri üzerinde Intel Core i9-7900X 3.30GHz×20 işlemci, 64 GB Ram ve 2 x GeForce RTX2080Ti ekran kartı bulunduran özellikte bir bilgisayarda gerçekleştirilmiştir. Modellerin başarı ölçütleri doğruluk değerine göre yapılmıştır. Bu bölümde her bir modelin eğitim aşamasındaki elde ettikleri yakınsama grafikleri, performans değerleri ve literatürdeki diğer çalışmalarla karşılaştırılması verilmiştir.

5.1. Performans Ölçütleri

Bu çalışmada tasarlanan ESA modellerinin performans ölçütleri karışıklık matrisi (confusion matrix) (Sammur ve Webb 2011) kullanılarak elde edilmiştir. Karışıklık matrisi yapılacak bir sınıflandırma işleminde gerçekte olan ve tahmin edilen sınıflar hakkında bilgi verir. İki boyutta olan karışıklık matrisi bir boyutu nesnenin gerçekte olan sınıfını gösterirken diğer boyutu sınıflandırıcı tarafından tahmin edilen boyutu gösterir (Deng ve ark 2016). Algoritmaların, performans başarıları değerlendirilirken; gerçekte olan ve tahmin edilen değerlerin arasındaki ilişkiyi veren karışıklık matrisinden faydalanılır. Karışıklık matrisindeki örneklem ifadeleri Tablo 5.1’de verilmiştir.

Tablo 5.1. Karışıklık matrisi

KARIŞIKLIK MATRİSİ		TAHMİN EDİLEN DEĞER	
		Pozitif	Negatif
GERÇEK DEĞER	Pozitif	Doğru Pozitif (TP)	Yanlış Pozitif (FP)
	Negatif	Yanlış Negatif (FN)	Doğru Negatif (TN)

TP: Pozitif örneğin doğru sınıflandırılması işlemidir.

TN: Negatif örneğin doğru sınıflandırılması işlemidir.

FP: Negatif örneğin yanlış sınıflandırılması işlemidir.

FN: Pozitif örneğin yanlış sınıflandırılması işlemidir.

Tablo 5.1'deki karışıklık matrisine göre; Doğruluk (Accuracy), Hassaslık (Precision), Duyarlılık (Recall) ve F-ölçüsü (F-measure) sırasıyla Eşitlik 5.1-5.4'e göre hesaplanmaktadır.

$$\text{Doğruluk} = \frac{TP + TN}{TP + FP + FN + TN} \quad (5.1)$$

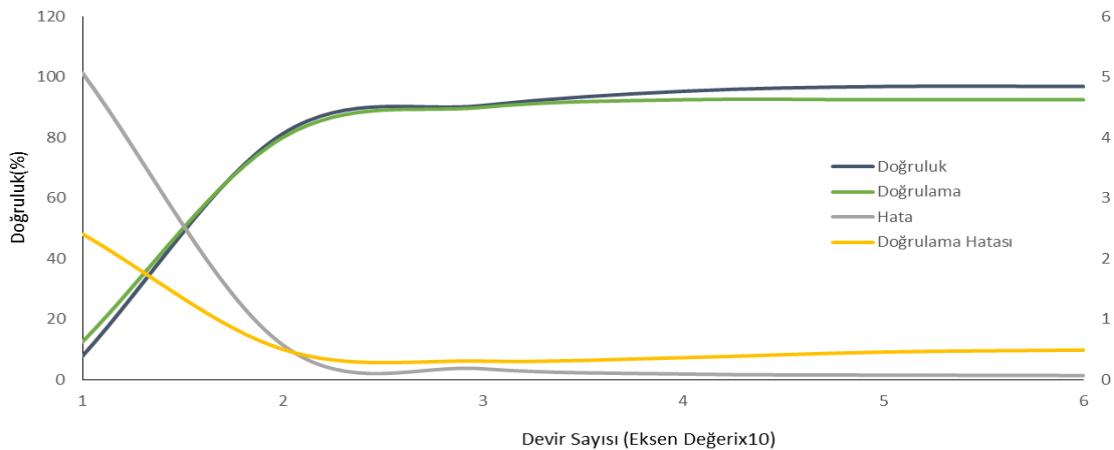
$$\text{Hassaslık} = \frac{TP}{TP + FP} \quad (5.2)$$

$$\text{Duyarlılık} = \frac{TP}{TP + FN} \quad (5.3)$$

$$F - \text{Ölçüsü} = \frac{2 * \text{Duyarlılık} * \text{Hassaslık}}{\text{Duyarlılık} + \text{hassaslık}} \quad (5.4)$$

5.2. ESC10 Veri Seti Üzerinde Yapılan Çalışmalar

ESC10 veri seti üzerinde ESC10_ESA32 modelinin eğitim aşamasında elde edilen yakınsama grafiği Şekil 5.1'de verilmiştir. Bu modelin elde ettiği karışıklık matrisleri Tablo 5.2'de verilmiştir. 10 Kat çaprazlama sonucunda elde edilen ortalama doğruluk karışıklık matrisi ise Tablo 5.3'te verilmiştir.



Şekil 5.1. ESC10_ESA32 modelinin eğitim aşamasındaki yakınsama grafiği

Şekil 5.1'e bakıldığında modellerin eğitim esnasında ezberleme yapmadan daha iyi yakınsama gerçekleştirdikleri görülmüştür.

Tablo 5.2. (a) ESC10_ESA32 modelinin test verisi üzerinde elde ettiği karışıklık matrisi (Fold 1-6)

GERÇEK	TAHMİN									
	Dog Bark	Rain	Sea Waves	Baby Cry	Clock Tik	Person Sneeze	Helicopter	Chainsaw	Rooster	Fire Cracking
	Dog Bark	4	0	0	0	0	0	0	0	0
	Rain	0	4	0	0	0	0	0	0	0
	Sea Waves	0	0	2	0	0	0	2	0	0
	Baby Cry	1	0	0	2	0	0	0	1	0
	Clock Tik	0	0	0	0	3	0	1	0	0
	Person Sneeze	0	0	0	0	0	4	0	0	0
	Helicopter	0	0	0	0	1	0	3	0	0
	Chainsaw	0	0	1	0	0	0	0	3	0
	Rooster	0	0	0	0	0	0	1	0	3
	Fire Cracking	0	0	0	0	0	0	0	0	4
FOLD-1: Doğruluk %80										
GERÇEK	TAHMİN									
	Dog Bark	Rain	Sea Waves	Baby Cry	Clock Tik	Person Sneeze	Helicopter	Chainsaw	Rooster	Fire Cracking
	Dog Bark	4	0	0	0	0	0	0	0	0
	Rain	0	4	0	0	0	0	0	0	0
	Sea Waves	0	0	3	0	0	0	1	0	0
	Baby Cry	0	0	0	3	0	0	0	1	0
	Clock Tik	0	0	0	0	2	0	0	0	2
	Person Sneeze	0	0	0	0	0	4	0	0	0
	Helicopter	0	1	0	0	0	0	3	0	0
	Chainsaw	0	0	0	0	0	0	0	4	0
	Rooster	0	0	0	0	0	0	0	0	4
	Fire Cracking	0	0	0	0	0	0	0	0	4
FOLD-2: Doğruluk %87.5										
GERÇEK	TAHMİN									
	Dog Bark	Rain	Sea Waves	Baby Cry	Clock Tik	Person Sneeze	Helicopter	Chainsaw	Rooster	Fire Cracking
	Dog Bark	2	0	0	1	0	1	0	0	0
	Rain	0	3	0	0	0	0	1	0	0
	Sea Waves	0	0	3	0	0	0	0	1	0
	Baby Cry	0	0	0	4	0	0	0	0	0
	Clock Tik	0	0	0	0	2	1	0	0	1
	Person Sneeze	0	0	0	0	0	4	0	0	0
	Helicopter	0	0	0	0	0	0	4	0	0
	Chainsaw	0	1	1	0	0	0	0	2	0
	Rooster	0	0	0	1	0	0	0	0	3
	Fire Cracking	0	0	0	0	0	0	0	0	4
FOLD-3: Doğruluk %77.5										
GERÇEK	TAHMİN									
	Dog Bark	Rain	Sea Waves	Baby Cry	Clock Tik	Person Sneeze	Helicopter	Chainsaw	Rooster	Fire Cracking
	Dog Bark	4	0	0	0	0	0	0	0	0
	Rain	0	4	0	0	0	0	0	0	0
	Sea Waves	0	1	3	0	0	0	0	0	0
	Baby Cry	0	0	0	4	0	0	0	0	0
	Clock Tik	2	0	0	0	0	0	0	0	2
	Person Sneeze	0	0	0	0	0	4	0	0	0
	Helicopter	0	1	0	0	0	0	3	0	0
	Chainsaw	0	0	0	0	0	0	0	4	0
	Rooster	0	0	0	0	0	0	0	0	4
	Fire Cracking	0	0	0	0	0	0	0	0	4
FOLD-4: Doğruluk %85										
GERÇEK	TAHMİN									
	Dog Bark	Rain	Sea Waves	Baby Cry	Clock Tik	Person Sneeze	Helicopter	Chainsaw	Rooster	Fire Cracking
	Dog Bark	4	0	0	0	0	0	0	0	0
	Rain	0	3	1	0	0	0	0	0	0
	Sea Waves	0	0	3	0	0	0	0	1	0
	Baby Cry	0	0	0	3	0	0	0	0	1
	Clock Tik	1	0	0	0	3	0	0	0	0
	Person Sneeze	0	0	0	1	0	3	0	0	0
	Helicopter	0	0	0	0	0	0	3	0	1
	Chainsaw	0	0	0	0	0	0	0	4	0
	Rooster	0	0	0	0	0	0	0	0	4
	Fire Cracking	0	0	0	0	1	0	0	0	3
FOLD-5: Doğruluk %82.5										
GERÇEK	TAHMİN									
	Dog Bark	Rain	Sea Waves	Baby Cry	Clock Tik	Person Sneeze	Helicopter	Chainsaw	Rooster	Fire Cracking
	Dog Bark	4	0	0	0	0	0	0	0	0
	Rain	0	4	0	0	0	0	0	0	0
	Sea Waves	0	1	3	0	0	0	0	0	0
	Baby Cry	0	0	0	3	0	0	0	0	1
	Clock Tik	0	0	0	0	4	0	0	0	0
	Person Sneeze	1	0	0	0	1	2	0	0	0
	Helicopter	0	1	0	0	1	0	2	0	0
	Chainsaw	0	0	0	0	0	0	0	4	0
	Rooster	0	0	0	0	0	0	0	0	4
	Fire Cracking	0	1	0	0	0	0	0	0	3
FOLD-6: Doğruluk %82.5										

Tablo 5.2. (b) ESC10_ESA32 modelinin test verisi üzerinde elde ettiği karışıklık matrisi (Fold 7-10)

GERÇEK	TAHMİN									
	Dog Bark	Rain	Sea Waves	Baby Cry	Clock Tik	Person Sneez	Helicopter	Chainsaw	Rooster	Fire Cracking
	Dog Bark	3	0	0	0	1	0	0	0	0
	Rain	0	3	0	0	1	0	0	0	0
	Sea Waves	0	0	3	0	0	1	0	0	0
	Baby Cry	0	0	0	4	0	0	0	0	0
	Clock Tik	0	0	0	0	2	0	0	0	2
	Person Sneez	1	0	0	0	0	3	0	0	0
	Helicopter	0	1	0	0	0	0	3	0	0
	Chainsaw	0	0	0	0	0	0	0	4	0
	Rooster	0	0	0	0	0	0	0	0	4
	Fire Cracking	0	0	0	0	1	0	0	0	3
FOLD-7: Doğruluk %80										
GERÇEK	TAHMİN									
	Dog Bark	Rain	Sea Waves	Baby Cry	Clock Tik	Person Sneez	Helicopter	Chainsaw	Rooster	Fire Cracking
	Dog Bark	2	0	0	1	0	1	0	0	0
	Rain	0	3	0	0	0	0	1	0	0
	Sea Waves	0	1	3	0	0	0	0	0	0
	Baby Cry	1	0	0	3	0	0	0	0	0
	Clock Tik	0	0	0	1	2	0	0	0	1
	Person Sneez	0	0	0	0	1	3	0	0	0
	Helicopter	0	0	0	0	0	0	4	0	0
	Chainsaw	0	0	0	0	0	0	0	4	0
	Rooster	1	0	0	0	0	0	0	3	0
	Fire Cracking	0	0	0	0	0	0	0	0	4
FOLD-8: Doğruluk %77.5										
GERÇEK	TAHMİN									
	Dog Bark	Rain	Sea Waves	Baby Cry	Clock Tik	Person Sneez	Helicopter	Chainsaw	Rooster	Fire Cracking
	Dog Bark	1	0	0	1	0	1	0	1	0
	Rain	0	2	2	0	0	0	0	0	0
	Sea Waves	0	0	4	0	0	0	0	0	0
	Baby Cry	1	0	0	2	0	0	0	1	0
	Clock Tik	0	0	0	0	2	0	0	0	2
	Person Sneez	0	0	0	0	0	4	0	0	0
	Helicopter	0	0	0	0	0	0	4	0	0
	Chainsaw	0	1	1	0	0	0	2	0	0
	Rooster	0	0	0	0	0	0	0	4	0
	Fire Cracking	0	0	0	0	0	1	0	0	3
FOLD-9: Doğruluk %70										
GERÇEK	TAHMİN									
	Dog Bark	Rain	Sea Waves	Baby Cry	Clock Tik	Person Sneez	Helicopter	Chainsaw	Rooster	Fire Cracking
	Dog Bark	3	0	0	0	1	0	0	0	0
	Rain	0	3	0	0	0	0	1	0	0
	Sea Waves	0	1	2	0	0	0	1	0	0
	Baby Cry	0	0	0	4	0	0	0	0	0
	Clock Tik	0	0	0	0	4	0	0	0	0
	Person Sneez	1	0	0	0	0	3	0	0	0
	Helicopter	0	0	0	0	0	0	4	0	0
	Chainsaw	0	0	0	0	0	0	0	4	0
	Rooster	0	0	0	0	0	0	0	0	4
	Fire Cracking	0	0	0	0	0	1	0	0	3
FOLD-10: Doğruluk %85										

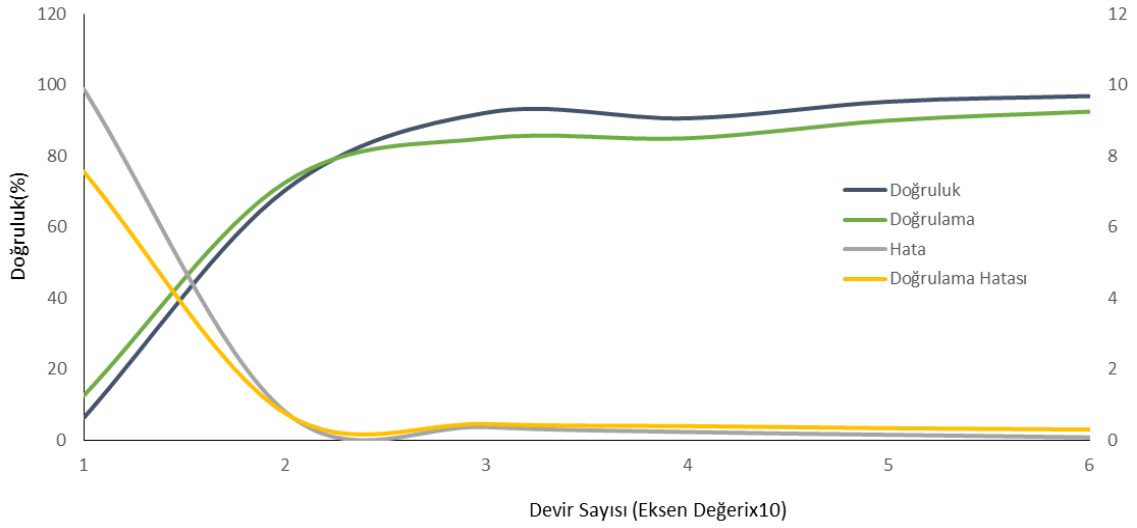
Tablo 5.2’de verilen 10-Kat Çapraz Doğrulama içerisinde en yüksek doğruluk oranı %87.50 değerine sahip Fold-2 olmuştur. En düşük sınıflama doğruluk oranı ise Fold-9 da %70 olarak elde edilmiştir.

Tablo 5.3. ESC10_ESA32 modelinin 10-kat aprazlama ile ESC10 veri seti zerinde %80.75 ortalama doęruluęun elde edildięi karışıklık matrisi

True Class	001 - Dog bark	31			3	2	3		1			77.5%	22.5%
	002 - Rain		33	3		1		3				82.5%	17.5%
	003 - Sea waves		4	29				3	3	1		72.5%	27.5%
	004 - Baby cry	3			32				1	4		80.0%	20.0%
	005 - Clock tick	3			1	24	1	1			10	60.0%	40.0%
	006 - Person sneeze	3			1	2	34					85.0%	15.0%
	007 - Helicopter		4			2		33			1	82.5%	17.5%
	008 - Chainsaw		2	3					35			87.5%	12.5%
	009 - Rooster	1			1			1		37		92.5%	7.5%
	010 - Fire crackling		1			2		2			35	87.5%	12.5%
		75.6%	75.0%	82.9%	84.2%	72.7%	89.5%	76.7%	87.5%	88.1%	76.1%		
		24.4%	25.0%	17.1%	15.8%	27.3%	10.5%	23.3%	12.5%	11.9%	23.9%		
		001 - Dog bark	002 - Rain	003 - Sea waves	004 - Baby cry	005 - Clock tick	006 - Person sneeze	007 - Helicopter	008 - Chainsaw	009 - Rooster	010 - Fire crackling		
		Predicted Class											

ESC10_ESA32 modelinin 10-Kat apraz Doğrulama (cross-validation) sonucunda Tablo 5.3'te elde edilen ortalama doęruluk oranına sahip karışıklık matrisine bakıldığında, ortalama doęruluk oranı %80.75 olduęu görlmştr. Tablo 5.3'te en yksek doęruluk sınıfının 40 test grntsnn 37 tanesini doęru etiketleyerek %92.50 başarı oranı elde edilen horoz sesi olduęu anlaşılmıştır. En dřk sınıflandırma ise %60 ile saat tik-tak sesi olduęu görlmştr. Yapılan deęerlendirmede en ok bu sınıfa ait kentsel sesin dięer ses sınıflarıyla karıştırdıldığı anlaşılmıştır. Genel olarak ESC10_ESA32 modeli, saat tik-tok sesi, dalga sesi ve kpek havlama sesi dıřındaki dięer sınıflarda %80 zeri bir doęruluk oranı ile sınıflandırma iřlemi yaptıęı görlmştr.

ESC10 veri seti zerinde ESC10_ESA224 modelinin eęitim ařamasında elde edilen yakınsama grafięi řekil 5.2'de verilmiştir. Bu modelin elde ettięi karışıklık matrisleri Tablo 5.4'te verilmiştir. 10 Kat aprazlama sonucunda elde edilen ortalama doęruluk karışıklık matrisi ise Tablo 5.5'te verilmiştir.



Şekil 5.2. ESC10_ESA224 modelinin eğitim aşamasındaki yakınsama grafiği

Şekil 5.2'ye bakıldığında modellerin eğitim esnasında ezberleme yapmadan daha iyi yakınsama gerçekleştirdikleri görülmüştür.

Tablo 5.4. (a). ESC10_ESA224 modelinin test verisi üzerinde elde ettiği karışıklık matrisi (Fold 1-2)

		TAHMİN									
		Dog Bark	Rain	Sea Waves	Baby Cry	Clock Tik	Person Sneeze	Helicopter	Chainsaw	Rooster	Fire Cracking
GERÇEK	Dog Bark	3	0	0	0	0	1	0	0	0	0
	Rain	0	4	0	0	0	0	0	0	0	0
	Sea Waves	0	1	2	0	0	1	0	0	0	0
	Baby Cry	0	0	0	4	0	0	0	0	0	0
	Clock Tik	0	0	0	0	3	0	0	0	0	1
	Person Sneeze	0	0	0	0	0	4	0	0	0	0
	Helicopter	0	1	1	0	0	0	2	0	0	0
	Chainsaw	0	0	0	0	0	0	0	4	0	0
	Rooster	0	0	0	0	0	0	0	0	4	0
	Fire Cracking	0	0	0	0	0	0	0	0	0	4
		FOLD-1: Doğruluk %85									
		TAHMİN									
		Dog Bark	Rain	Sea Waves	Baby Cry	Clock Tik	Person Sneeze	Helicopter	Chainsaw	Rooster	Fire Cracking
GERÇEK	Dog Bark	2	0	0	0	1	1	0	0	0	0
	Rain	0	3	0	0	0	0	0	0	0	1
	Sea Waves	0	1	3	0	0	0	0	0	0	0
	Baby Cry	0	0	0	4	0	0	0	0	0	0
	Clock Tik	0	0	0	0	3	0	0	1	0	0
	Person Sneeze	0	0	0	0	0	4	0	0	0	0
	Helicopter	0	0	0	0	0	0	4	0	0	0
	Chainsaw	0	0	1	0	0	0	0	3	0	0
	Rooster	0	0	0	2	0	0	0	0	2	0
	Fire Cracking	0	0	0	0	0	1	0	0	0	3
		FOLD-2: Doğruluk %77.5									

Tablo 5.4. (b). ESC10_ESA224 modelinin test verisi üzerinde elde ettiği karışıklık matrisi (Fold 3-8)

		TAHMİN									
		Dog Bark	Rain	Sea Waves	Baby Cry	Clock Tik	Person Sneeze	Helicopter	Chainsaw	Rooster	Fire Cracking
GERÇEK	Dog Bark	4	0	0	0	0	0	0	0	0	0
	Rain	0	4	0	0	0	0	0	0	0	0
	Sea Waves	0	0	4	0	0	0	0	0	0	0
	Baby Cry	0	0	0	3	0	0	0	0	1	0
	Clock Tik	0	0	0	0	4	0	0	0	0	0
	Person Sneeze	0	0	0	0	1	3	0	0	0	0
	Helicopter	0	0	0	0	0	0	4	0	0	0
	Chainsaw	0	1	0	0	0	0	0	3	0	0
	Rooster	0	0	0	0	0	0	0	0	4	0
	Fire Cracking	0	0	0	0	0	0	0	0	0	4
FOLD-3: Doğruluk %92.5											
		TAHMİN									
		Dog Bark	Rain	Sea Waves	Baby Cry	Clock Tik	Person Sneeze	Helicopter	Chainsaw	Rooster	Fire Cracking
GERÇEK	Dog Bark	4	0	0	0	0	0	0	0	0	0
	Rain	0	2	0	0	0	1	1	0	0	0
	Sea Waves	0	0	2	0	0	0	2	0	0	0
	Baby Cry	0	0	0	4	0	0	0	0	0	0
	Clock Tik	0	0	0	0	3	0	0	0	0	1
	Person Sneeze	0	0	0	0	0	4	0	0	0	0
	Helicopter	0	0	0	0	0	0	3	0	0	1
	Chainsaw	0	0	0	0	0	0	0	4	0	0
	Rooster	0	0	0	0	0	0	0	0	4	0
	Fire Cracking	0	1	0	0	0	0	0	0	0	3
FOLD-4: Doğruluk %82.5											
		TAHMİN									
		Dog Bark	Rain	Sea Waves	Baby Cry	Clock Tik	Person Sneeze	Helicopter	Chainsaw	Rooster	Fire Cracking
GERÇEK	Dog Bark	3	0	0	0	0	0	0	0	0	1
	Rain	0	3	1	0	0	0	0	0	0	0
	Sea Waves	0	0	4	0	0	0	0	0	0	0
	Baby Cry	0	0	0	4	0	0	0	0	0	0
	Clock Tik	0	0	0	0	3	0	1	0	0	0
	Person Sneeze	0	0	0	0	0	3	0	0	1	0
	Helicopter	0	2	0	0	0	0	2	0	0	0
	Chainsaw	0	1	0	0	0	0	0	3	0	0
	Rooster	1	0	0	0	0	1	0	0	2	0
	Fire Cracking	0	0	0	0	1	0	0	0	0	3
FOLD-5: Doğruluk %75											
		TAHMİN									
		Dog Bark	Rain	Sea Waves	Baby Cry	Clock Tik	Person Sneeze	Helicopter	Chainsaw	Rooster	Fire Cracking
GERÇEK	Dog Bark	3	0	0	0	0	0	1	0	0	0
	Rain	0	3	0	0	0	0	1	0	0	0
	Sea Waves	0	1	3	0	0	0	0	0	0	0
	Baby Cry	0	0	0	4	0	0	0	0	0	0
	Clock Tik	0	0	0	0	4	0	0	0	0	0
	Person Sneeze	0	0	0	0	2	2	0	0	0	0
	Helicopter	0	0	0	0	0	0	4	0	0	0
	Chainsaw	0	0	1	0	0	0	1	2	0	0
	Rooster	0	0	0	0	0	0	0	0	4	0
	Fire Cracking	0	0	0	0	0	0	1	0	0	3
FOLD-7: Doğruluk %80											
		TAHMİN									
		Dog Bark	Rain	Sea Waves	Baby Cry	Clock Tik	Person Sneeze	Helicopter	Chainsaw	Rooster	Fire Cracking
GERÇEK	Dog Bark	3	0	0	0	0	0	0	0	1	0
	Rain	0	4	0	0	0	0	0	0	0	0
	Sea Waves	0	0	4	0	0	0	0	0	0	0
	Baby Cry	0	0	0	4	0	0	0	0	0	0
	Clock Tik	0	0	0	0	3	1	0	0	0	0
	Person Sneeze	0	0	0	0	0	4	0	0	0	0
	Helicopter	0	0	0	0	0	0	3	0	0	1
	Chainsaw	0	0	1	0	0	0	0	3	0	0
	Rooster	0	0	0	1	0	0	0	0	3	0
	Fire Cracking	0	0	0	0	0	0	0	0	0	4
FOLD-8: Doğruluk %87.5											

Tablo 5.4. (c). ESC10_ESA224 modelinin test verisi üzerinde elde ettiği karışıklık matrisi (Fold 9-10)

		TAHMİN									
		Dog Bark	Rain	Sea Waves	Baby Cry	Clock Tik	Person Sneeze	Helicopter	Chainsaw	Rooster	Fire Cracking
GERÇEK	Dog Bark	3	0	0	0	1	0	0	0	0	0
	Rain	0	3	1	0	0	0	0	0	0	0
	Sea Waves	0	0	3	0	0	0	1	0	0	0
	Baby Cry	0	0	0	3	0	0	0	0	1	0
	Clock Tik	0	0	0	0	3	0	0	0	0	1
	Person Sneeze	1	0	0	0	0	3	0	0	0	0
	Helicopter	0	1	0	0	0	0	2	1	0	0
	Chainsaw	0	0	0	0	0	0	0	4	0	0
	Rooster	1	0	0	0	0	0	0	0	3	0
	Fire Cracking	0	0	0	0	0	0	0	0	0	4
		FOLD-9: Doğruluk %77.5									
		TAHMİN									
		Dog Bark	Rain	Sea Waves	Baby Cry	Clock Tik	Person Sneeze	Helicopter	Chainsaw	Rooster	Fire Cracking
GERÇEK	Dog Bark	4	0	0	0	0	0	0	0	0	0
	Rain	0	3	0	0	0	0	1	0	0	0
	Sea Waves	0	1	2	0	0	0	0	1	0	0
	Baby Cry	0	0	0	2	1	0	0	0	1	0
	Clock Tik	0	0	0	0	4	0	0	0	0	0
	Person Sneeze	0	0	0	0	1	3	0	0	0	0
	Helicopter	0	0	0	0	0	0	3	1	0	0
	Chainsaw	0	0	1	0	0	0	0	3	0	0
	Rooster	0	0	0	0	0	0	0	0	4	0
	Fire Cracking	0	0	0	0	0	0	0	0	0	4
		FOLD-10: Doğruluk %80									

Tablo 5.4'te verilen 10-Kat Çapraz Doğrulama içerisinde en yüksek doğruluk oranı %92.50 değerine sahip Fold-3 olmuştur. En düşük sınıflama doğruluk oranı ise Fold-5 da %75 olarak elde edilmiştir.

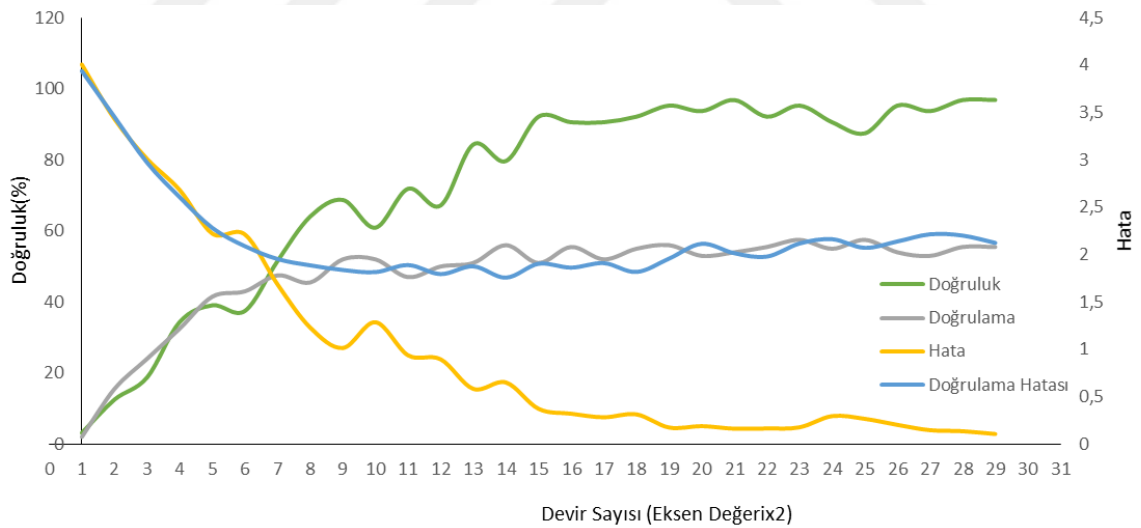
Tablo 5.5. ESC10_ESA224 modelinin 10-kat çaprazlama ile ESC10 veri seti üzerinde %82.25 ortalama doğruluğun elde edildiği karışıklık matrisi

True Class	001 - Dog bark	32			1	2	2	1		1	1	80.0%	20.0%
	002 - Rain		32	2			1	4			1	80.0%	20.0%
	003 - Sea waves		4	31			1	3	1			77.5%	22.5%
	004 - Baby cry				35	1				4		87.5%	12.5%
	005 - Clock tick	1				33	1	1	1		3	82.5%	17.5%
	006 - Person sneeze	1				4	33			1	1	82.5%	17.5%
	007 - Helicopter		5	1				30	2		2	75.0%	25.0%
	008 - Chainsaw		2	4				1	33			82.5%	17.5%
	009 - Rooster	2			3		1			34		85.0%	15.0%
	010 - Fire crackling		1			1	1	1			36	90.0%	10.0%
		88.9%	72.7%	81.6%	89.7%	80.5%	82.5%	73.2%	89.2%	85.0%	81.8%		
		11.1%	27.3%	18.4%	10.3%	19.5%	17.5%	26.8%	10.8%	15.0%	18.2%		
		001 - Dog bark	002 - Rain	003 - Sea waves	004 - Baby cry	005 - Clock tick	006 - Person sneeze	007 - Helicopter	008 - Chainsaw	009 - Rooster	010 - Fire crackling		
		Predicted Class											

ESC10_ESA224 modelinin 10-Kat Çapraz Doğrulama (cross-validation) sonucunda Tablo 5.5'te elde edilen ortalama doğruluk oranına sahip karışıklık matrisine bakıldığında, ortalama doğruluk oranı %82.25 olduğu görülmüştür. Tablo 5.5'te en yüksek doğruluk sınıfının 40 test görüntüsünü 36 tanesini doğru etiketleyerek %90 başarı oranı elde edilen ateş kıvılcım sesleri olduğu anlaşılmıştır. En düşük sınıflandırma ise %75 ile helikopter sesi olduğu görülmüştür. Yapılan değerlendirmede en çok bu sınıfa ait kentsel sesin diğer ses sınıflarıyla karıştırıldığı anlaşılmıştır. Genel olarak ESC10_ESA224 modeli, deniz dalga sesi ve helikopter sesi dışındaki diğer sınıflarda %80 üzeri bir doğruluk oranı ile sınıflandırma işlemi yaptığı görülmüştür.

5.3. ESC50 Veri Seti Üzerinde Yapılan Çalışmalar

ESC50 veri seti üzerinde ESC50_ESA32 modelinin eğitim aşamasında elde edilen yakınsama grafiği Şekil 5'te verilmiştir. Bu modelin elde ettiği karışıklık matrisleri Tablo 5.6'da verilmiştir. 10 Kat çaprazlama sonucunda elde edilen ortalama doğruluk karışıklık matrisi ise Tablo 5.7'de verilmiştir.



Şekil 5.3. ESC50_ESA32 modelinin eğitim aşamasındaki yakınsama grafiği

Şekil 5.3'e bakıldığında modelin eğitim esnasında ezberleme yaptığı görülmüştür. Bunun sebebi ise giriş görüntü boyutunun 32x32x3 olması sebebi ile modelin son katmanlarına aktarılan verinin özelliklerinin yeterince keşfedilememesindendir.

[illegible][illegible]

[illegible]52

Tablo 5.6. (c) ESC50_ESA32 modelinin test verisi üzerinde elde ettiği karışıklık matrisi (Fold 5-6)

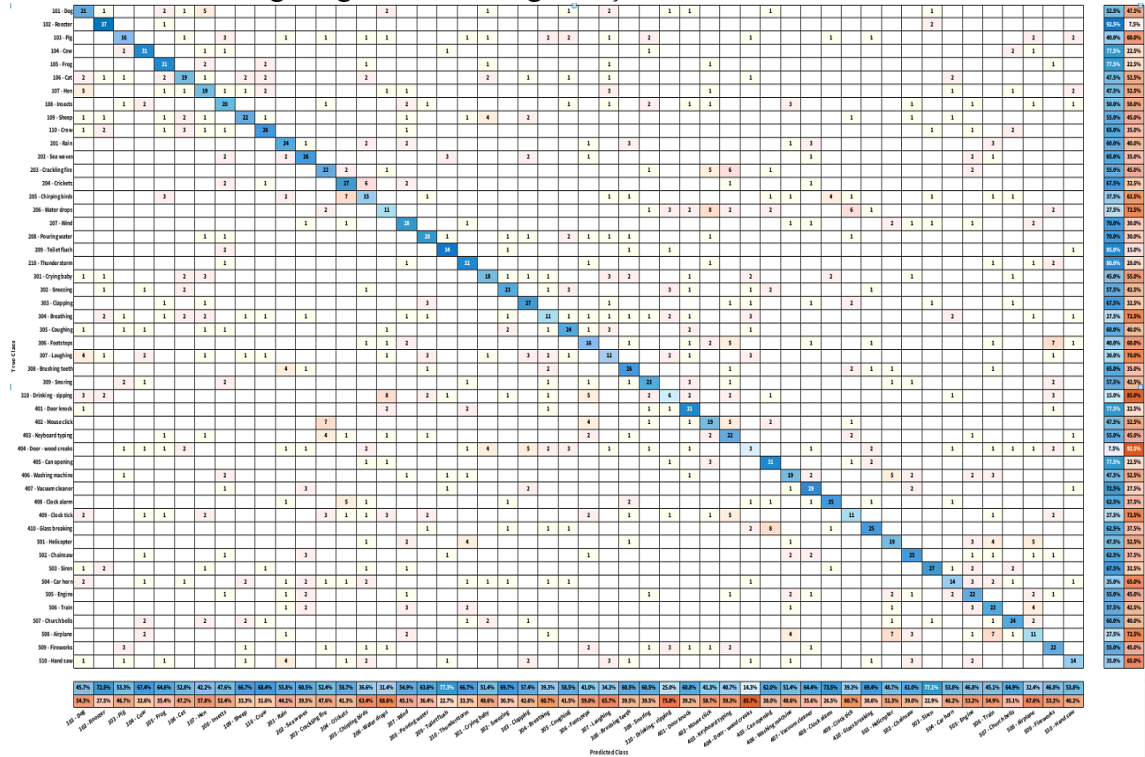
[illegible]

Tablo 5.6. (e) ESC50_ESA32 modelinin test verisi üzerinde elde ettiği karışıklık matrisi (Fold 9-10)

[illegible]

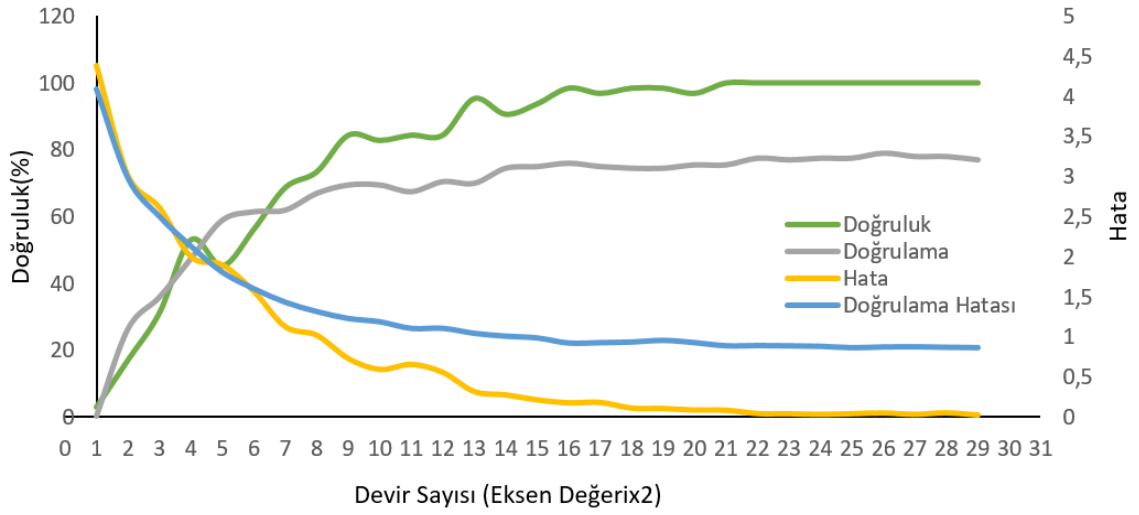
Tablo 5.6’da verilen 10-Kat Çapraz Doğrulama içerisinde en yüksek doğruluk oranı %58.50 değerine sahip Fold-9 olmuştur. En düşük sınıflama doğruluk oranı ise Fold-6 da %51 olarak elde edilmiştir.

Tablo 5.7. ESC50_ESA32 modelinin 10-kat çaprazlama ile ESC50 veri seti üzerinde %54.55 ortalama doğruluğun elde edildiği karışıklık matrisi



ESC50_ESA32 modelinin 10-Kat Çapraz Doğrulama (cross-validation) sonucunda Tablo 5.7’de elde edilen ortalama doğruluk oranına sahip karışıklık matrisine bakıldığında, ortalama doğruluk oranı %54.55 olduğu görülmüştür. Tablo 5.7’de en yüksek doğruluk sınıfının 40 test görüntüsünü 37 tanesini doğru etiketleyerek %92.50 başarı oranı elde edilen horoz sesi olduğu anlaşılmıştır. En düşük sınıflandırma ise %7.5 ile ahşap kapı gıcırta sesi olduğu görülmüştür. Yapılan değerlendirmede en çok bu sınıfa ait kentsel sesin diğer ses sınıflarıyla karıştırıldığı anlaşılmıştır.

ESC50 veri seti üzerinde ESC50_ESA224 modelinin eğitim aşamasında elde edilen yakınsama grafiği Şekil 5.4’te verilmiştir. Bu modelin elde ettiği karışıklık matrisleri Tablo 5.8’de verilmiştir. 10 Kat çaprazlama sonucunda elde edilen ortalama doğruluk karışıklık matrisi ise Tablo 5.9’da verilmiştir.



Şekil 5.4. ESC50_ESA224 modelinin eğitim aşamasındaki yakınsama grafiği

Şekil 5.4'e bakıldığında ESC50_ESA32 modeline göre daha iyi yakınsama grafiği verdiği görülmüştür. Diğer veri setleri için geliştirilen modellere nispeten ezberleme yapma eğilimi gösterdiği anlaşılmıştır.

Tablo 5.8. (a) ESC50_ESA224 modelinin test verisi üzerinde elde ettiği karışıklık matrisi (Fold 1-2)

[illegible][illegible]

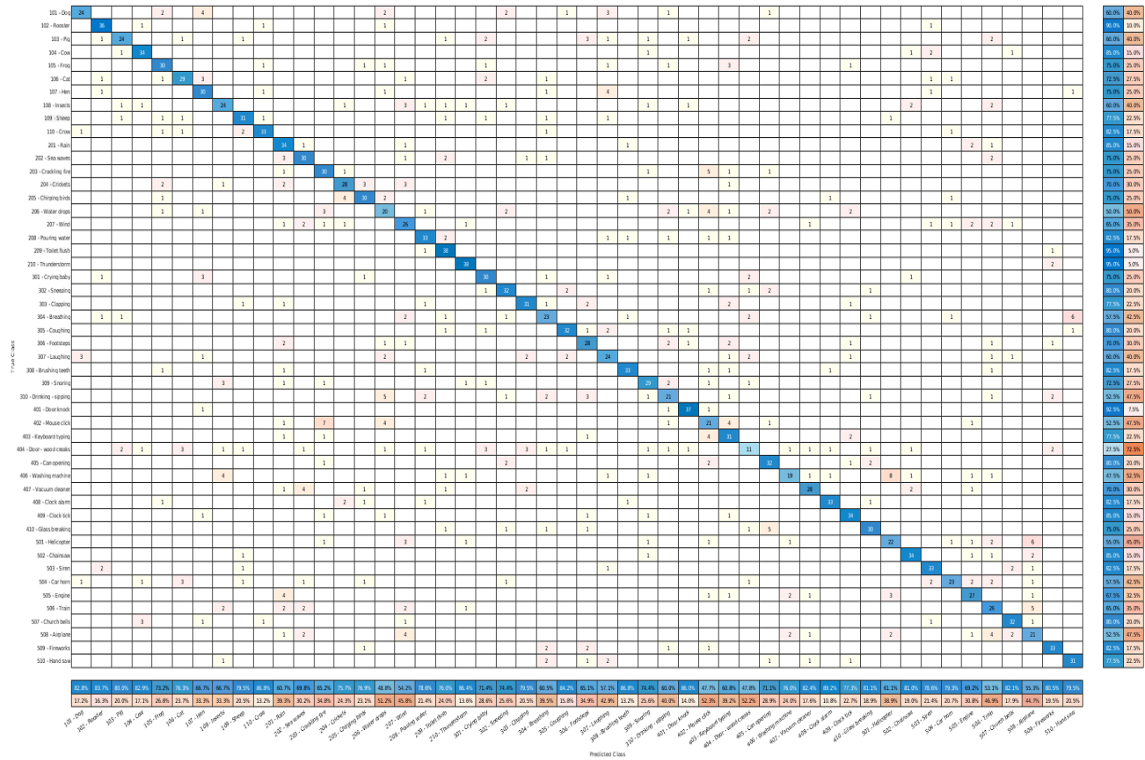
[illegible]60

Tablo 5.8. (e) ESC50_ESA224 modelinin test verisi üzerinde elde ettiği karışıklık matrisi (Fold 9-10)

[illegible][illegible]

Tablo 5.8’de verilen 10-Kat Çapraz Doğrulama içerisinde en yüksek doğruluk oranı %77.50 değerine sahip Fold-2 olmuştur. En düşük sınıflama doğruluk oranı ise Fold-5 da %66.50 olarak elde edilmiştir.

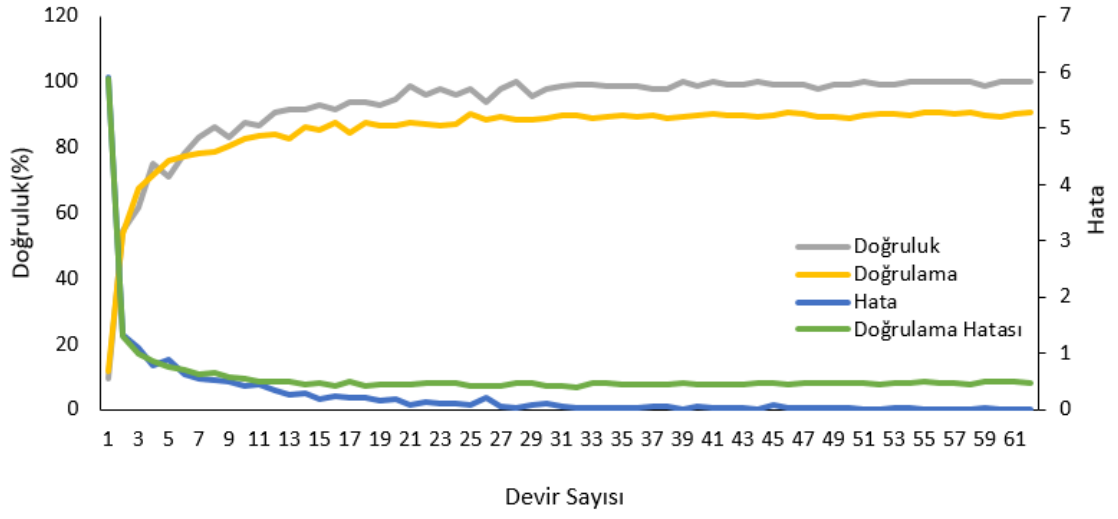
Tablo 5.9. ESC50_ESA224 modelinin 10-kat çaprazlama ile ESC50 veri seti üzerinde %72.15 ortalama doğruluğun elde edildiği karışıklık matrisi



ESC50_ESA224 modelinin 10-Kat Çapraz Doğrulama (cross-validation) sonucunda Tablo 5.9’da elde edilen ortalama doğruluk oranına sahip karışıklık matrisine bakıldığında, ortalama doğruluk oranı %72.15 olduğu görülmüştür. Tablo 5.9’da en yüksek doğruluk sınıfının 40 test görüntüsünü 38 tanesini doğru etiketleyerek %95 başarı oranı elde edilen tuvalet fırça sesi ve gök gürültüsü sesi olduğu anlaşılmıştır. En düşük sınıflandırma ise %27 ile ahşap kapı gıcırta sesi olduğu görülmüştür. Yapılan değerlendirmede en çok bu sınıfa ait kentsel sesin diğer ses sınıflarıyla karıştırıldığı anlaşılmıştır.

5.4. UrbanSound8K Veri Seti Üzerinde Yapılan Çalışmalar

UrbanSound8K veri seti üzerinde URBANSOUND8K_ESA32 modelinin eğitim aşamasında elde edilen yakınsama grafiği Şekil 5.5’te verilmiştir. Bu modelin elde ettiği karışıklık matrisleri Tablo 5.10’da verilmiştir. 10 Kat çaprazlama sonucunda elde edilen ortalama doğruluk karışıklık matrisi ise Tablo 5.11’de verilmiştir.



Şekil 5.5. URBANSOUND8K_ESA32 modelinin eğitim aşamasındaki yakınsama grafiği

Şekil 5.5'e bakıldığında modellerin eğitim esnasında ezberleme yapmadan daha iyi yakınsama gerçekleştirdikleri görülmüştür.

Tablo 5.10. (a) URBANSOUND8K_ESA32 modelinin test verisi üzerinde elde ettiği karışıklık matrisi (Fold 1-6)

GERÇEK	TAHMİN									
	Air_conditioner	Car_horn	Children_playing	Dog_bark	Drilling	Engine_idling	Gun_shot	Jackhammer	Siren	Street_music
Air_conditioner	93	0	1	2	1	0	1	0	0	2
Car_horn	1	29	2	1	1	1	1	1	1	4
Children_playing	0	1	81	7	0	0	0	0	0	11
Dog_bark	3	1	4	82	3	1	0	0	0	6
Drilling	0	0	0	0	94	0	0	4	0	2
Engine_idling	0	0	0	1	0	96	0	1	0	2
Gun_shot	0	0	1	0	1	0	35	1	0	0
Jackhammer	1	0	0	1	5	1	0	91	0	1
Siren	0	2	2	2	1	0	0	0	82	4
Street_music	2	0	5	2	0	0	0	1	2	88
FOLD-1: Doğruluk %88.3162										
GERÇEK	TAHMİN									
	Air_conditioner	Car_horn	Children_playing	Dog_bark	Drilling	Engine_idling	Gun_shot	Jackhammer	Siren	Street_music
Air_conditioner	96	0	2	0	0	0	0	0	2	0
Car_horn	0	34	0	4	3	0	0	0	1	1
Children_playing	0	1	78	8	2	0	0	1	3	7
Dog_bark	1	1	5	88	0	0	0	1	1	3
Drilling	0	0	0	0	94	0	0	3	2	1
Engine_idling	1	0	1	1	0	96	0	0	1	0
Gun_shot	0	1	0	0	1	0	36	0	0	0
Jackhammer	0	0	0	0	3	0	0	96	0	1
Siren	0	0	2	0	1	0	0	0	90	0
Street_music	1	3	7	1	0	0	0	3	2	83
FOLD-2: Doğruluk %90.5034										
GERÇEK	TAHMİN									
	Air_conditioner	Car_horn	Children_playing	Dog_bark	Drilling	Engine_idling	Gun_shot	Jackhammer	Siren	Street_music
Air_conditioner	96	0	1	0	2	0	0	1	0	0
Car_horn	2	32	3	2	0	1	0	2	0	1
Children_playing	1	0	83	4	2	2	1	0	1	6
Dog_bark	2	0	7	81	4	1	0	1	2	2
Drilling	2	1	1	1	86	1	1	4	1	2
Engine_idling	0	0	1	2	0	94	0	2	1	0
Gun_shot	0	0	0	0	0	0	38	0	0	0
Jackhammer	1	0	1	0	3	0	0	94	0	1
Siren	0	0	4	1	2	0	0	0	84	2
Street_music	1	0	9	2	3	4	0	0	0	81
FOLD-3: Doğruluk %87.9863										
GERÇEK	TAHMİN									
	Air_conditioner	Car_horn	Children_playing	Dog_bark	Drilling	Engine_idling	Gun_shot	Jackhammer	Siren	Street_music
Air_conditioner	93	0	2	0	1	2	0	2	0	0
Car_horn	0	36	0	1	3	1	0	0	1	1
Children_playing	0	1	84	2	5	0	0	1	2	5
Dog_bark	2	1	8	83	0	1	0	0	1	4
Drilling	0	0	0	0	91	1	0	4	2	2
Engine_idling	3	2	3	0	1	89	0	0	2	0
Gun_shot	0	0	0	0	0	0	35	0	0	2
Jackhammer	0	0	1	0	3	1	0	94	0	1
Siren	1	0	3	2	1	0	0	0	86	0
Street_music	2	0	9	2	2	0	0	2	4	79
FOLD-5: Doğruluk %88.2016										
GERÇEK	TAHMİN									
	Air_conditioner	Car_horn	Children_playing	Dog_bark	Drilling	Engine_idling	Gun_shot	Jackhammer	Siren	Street_music
Air_conditioner	97	0	1	0	0	1	0	1	0	0
Car_horn	2	35	4	0	0	0	0	0	2	0
Children_playing	5	2	80	6	1	2	0	0	1	3
Dog_bark	0	5	4	85	1	0	1	0	0	4
Drilling	0	0	1	1	92	0	0	5	0	1
Engine_idling	0	0	1	0	0	96	0	1	0	2
Gun_shot	0	0	2	0	0	0	34	0	1	0
Jackhammer	0	0	2	0	3	1	0	94	0	0
Siren	0	0	1	1	1	0	0	0	90	0
Street_music	0	0	9	5	1	5	0	3	3	74
FOLD-6: Doğruluk %89.0034										

Tablo 5.10. (b) URBANSOUND8K_ESA32 modelinin test verisi üzerinde elde ettiği karışıklık matrisi (Fold 7-10)

		TAHMİN									
		Air_conditioner	Car_horn	Children_playing	Dog_bark	Drilling	Engine_idling	Gun_shot	Jackhammer	Siren	Street_music
GERÇEK	Air_conditioner	96	0	0	2	0	0	0	1	0	1
	Car_horn	0	34	2	2	2	0	0	0	0	3
	Children_playing	1	1	75	11	1	2	0	0	2	7
	Dog_bark	3	0	4	87	1	1	0	0	0	4
	Drilling	0	1	1	1	87	0	0	6	1	3
	Engine_idling	2	0	0	2	0	95	0	0	0	1
	Gun_shot	0	0	0	0	1	0	35	0	0	1
	Jackhammer	1	0	0	0	2	2	0	94	0	1
	Siren	0	2	8	0	1	0	0	0	80	2
	Street_music	4	1	5	6	0	2	0	0	1	81
FOLD-7: Doğruluk %87.5143											

		TAHMİN									
		Air_conditioner	Car_horn	Children_playing	Dog_bark	Drilling	Engine_idling	Gun_shot	Jackhammer	Siren	Street_music
GERÇEK	Air_conditioner	91	0	0	0	1	1	1	0	0	6
	Car_horn	2	31	2	1	4	2	0	0	0	1
	Children_playing	0	0	76	5	4	3	2	3	1	6
	Dog_bark	0	2	10	82	1	0	0	0	2	3
	Drilling	0	1	0	0	92	0	0	3	1	3
	Engine_idling	1	0	1	0	1	94	0	1	1	1
	Gun_shot	0	0	0	0	0	0	37	0	0	0
	Jackhammer	1	1	1	0	1	1	0	95	0	0
	Siren	0	0	5	0	0	0	0	0	84	4
	Street_music	1	2	2	2	1	2	1	0	4	85
FOLD-8: Doğruluk %87.858											

		TAHMİN									
		Air_conditioner	Car_horn	Children_playing	Dog_bark	Drilling	Engine_idling	Gun_shot	Jackhammer	Siren	Street_music
GERÇEK	Air_conditioner	95	0	2	1	0	0	0	0	2	0
	Car_horn	1	30	3	3	2	0	0	1	2	1
	Children_playing	2	1	84	3	1	1	0	1	3	4
	Dog_bark	2	1	7	85	1	2	0	0	1	1
	Drilling	0	1	1	1	90	0	0	4	1	2
	Engine_idling	1	1	0	0	0	95	0	1	1	1
	Gun_shot	0	0	0	0	0	0	36	0	0	1
	Jackhammer	1	1	0	0	3	0	0	95	0	0
	Siren	0	0	2	0	0	0	0	0	89	2
	Street_music	0	2	9	4	5	1	0	1	6	72
FOLD-9: Doğruluk %88.3162											

		TAHMİN									
		Air_conditioner	Car_horn	Children_playing	Dog_bark	Drilling	Engine_idling	Gun_shot	Jackhammer	Siren	Street_music
GERÇEK	Air_conditioner	96	0	1	0	0	1	0	1	1	0
	Car_horn	0	35	3	3	0	0	0	0	1	1
	Children_playing	0	1	83	2	0	6	0	2	1	5
	Dog_bark	1	3	2	81	2	2	0	1	0	8
	Drilling	0	2	2	0	85	1	0	6	2	2
	Engine_idling	1	0	0	0	0	99	0	0	0	0
	Gun_shot	0	0	0	0	0	0	37	0	0	0
	Jackhammer	2	0	0	0	1	0	0	96	0	1
	Siren	0	5	1	2	0	0	2	0	82	1
	Street_music	1	2	7	1	2	2	0	1	1	83
FOLD-10: Doğruluk %89.0034											

Tablo 5.10’da verilen 10-Kat Çapraz Doğrulama içerisinde en yüksek doğruluk oranı %90.50 değerine sahip Fold-2 olmuştur. En düşük sınıflama doğruluk oranı ise Fold-7 da %87.51 olarak elde edilmiştir.

Tablo 5.11. URBANSOUND8K_ESA32 modelinin 10-kat çaprazlama ile UrbanSound8K veri seti üzerinde %88.60 ortalama doğruluğun elde edildiği karışıklık matrisi

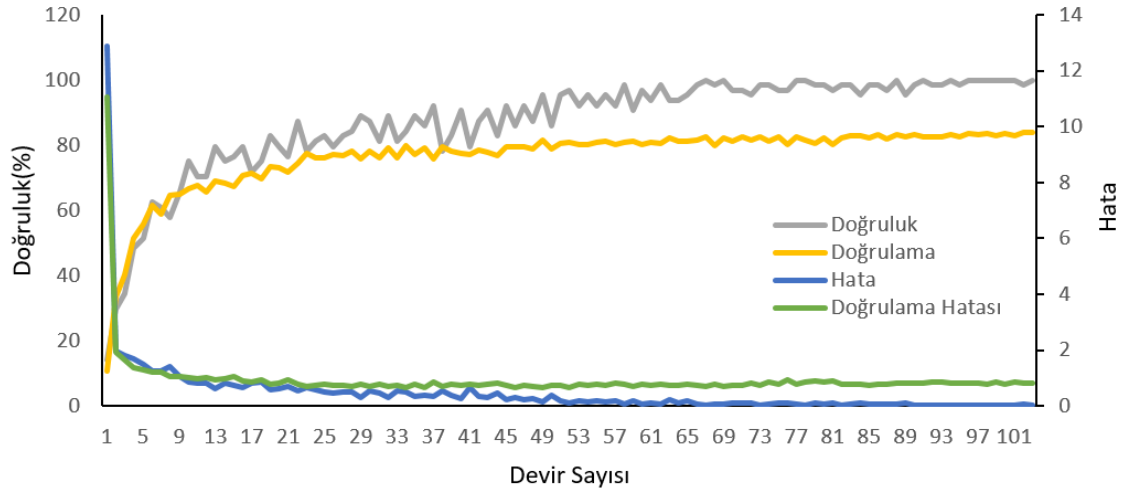
True Class	Air Conditioner	948		12	6	5	5	2	6	6	10	94.8%	5.2%
	Car Horn	10	330	21	18	16	6	1	4	8	15	76.9%	23.1%
	Children Playing	12	9	807	50	16	16	5	10	16	59	80.7%	19.3%
	Dog Bark	15	15	56	843	14	8	3	3	7	36	84.3%	15.7%
	Drilling	2	8	6	5	902	4	1	43	10	19	90.2%	9.8%
	Engine Idling	10	3	9	6	2	948		6	6	10	94.8%	5.2%
	Gun Shot		1	4		3		360	1	1	4	96.3%	3.7%
	Jackhammer	9	2	7	1	24	6		944		7	94.4%	5.6%
	Siren	2	10	31	9	7	2	2		849	17	91.4%	8.6%
	Street Music	13	10	74	29	16	16	1	11	24	806	80.6%	19.4%

92.9%	85.1%	78.6%	87.2%	89.8%	93.8%	96.0%	91.8%	91.6%	82.0%
7.1%	14.9%	21.4%	12.8%	10.2%	6.2%	4.0%	8.2%	8.4%	18.0%
Air Conditioner	Car Horn	Children Playing	Dog Bark	Drilling	Engine Idling	Gun Shot	Jackhammer	Siren	Street Music

Predicted Class

URBANSOUND8K_ESA32 modelinin 10-Kat Çapraz Doğrulama (cross-validation) sonucunda Tablo 5.11’de elde edilen ortalama doğruluk oranına sahip karışıklık matrisine bakıldığında, ortalama doğruluk oranı %88.60 olduğu görülmüştür. Tablo 5.11’de en yüksek doğruluk sınıfının 374 test görüntüsünü 360 tanesini doğru etiketleyerek %96.30 başarı oranı elde edilen silah atış sesi olduğu anlaşılmıştır. En düşük sınıflandırma ise %76.90 ile araç korna sesi olduğu görülmüştür. Yapılan değerlendirmede en çok bu sınıfa ait kentsel sesin diğer ses sınıflarıyla karıştırıldığı anlaşılmıştır. Genel olarak URBANSOUND8K_ESA32 modeli, araç korna sesi dışındaki diğer sınıflarda %80 üzeri bir doğruluk oranı ile sınıflandırma işlemi yaptığı görülmüştür.

UrbanSound8K veri seti üzerinde URBANSOUND8K_ESA224 modelinin eğitim aşamasında elde edilen yakınsama grafiği Şekil 5.6'da verilmiştir. Bu modelin elde ettiği karışıklık matrisleri Tablo 5.12'de verilmiştir. 10 Kat çaprazlama sonucunda elde edilen ortalama doğruluk karışıklık matrisi ise Tablo 5.13'te verilmiştir.



Şekil 5.6. URBANSOUND8K_ESA224 modelinin eğitim aşamasındaki yakınsama grafiği

Şekil 5.6'ya bakıldığında modellerin eğitim esnasında ezberleme yapmadan daha iyi yakınsama gerçekleştirdikleri görülmüştür.

Tablo 5.12. (a) URBANSOUND8K_ESA224 modelinin test verisi üzerinde elde ettiği karışıklık matrisi (Fold 1-6)

GERÇEK	TAHMİN									
	Air_conditioner	Car_horn	Children_playing	Dog_bark	Drilling	Engine_idling	Gun_shot	Jackhammer	Siren	Street_music
Air_conditioner	90	0	4	0	0	2	0	0	3	3
Car_horn	0	35	5	2	3	2	0	0	4	1
Children_playing	1	0	87	7	0	0	0	1	1	5
Dog_bark	2	5	2	81	0	0	0	2	0	3
Drilling	2	0	2	1	103	1	0	2	0	2
Engine_idling	3	1	3	2	0	98	0	1	2	0
Gun_shot	0	0	0	0	0	0	33	0	1	1
Jackhammer	0	0	0	0	4	1	0	78	0	1
Siren	0	2	2	0	0	0	0	0	86	1
Street_music	3	1	15	2	0	1	0	2	4	61
FOLD-1: Doğruluk %86.1397										
GERÇEK	TAHMİN									
	Air_conditioner	Car_horn	Children_playing	Dog_bark	Drilling	Engine_idling	Gun_shot	Jackhammer	Siren	Street_music
Air_conditioner	78	0	2	2	1	2	0	2	1	1
Car_horn	0	30	0	2	1	0	0	0	3	3
Children_playing	3	1	75	7	1	0	0	3	2	15
Dog_bark	2	3	6	88	0	0	0	0	0	4
Drilling	1	0	0	1	89	0	0	4	1	1
Engine_idling	0	0	2	1	1	90	0	0	0	2
Gun_shot	0	0	0	0	0	0	35	1	0	0
Jackhammer	0	0	2	0	5	0	0	102	0	1
Siren	0	4	3	0	0	3	0	0	80	0
Street_music	3	2	10	3	0	1	0	0	7	81
FOLD-2: Doğruluk %85.5835										
GERÇEK	TAHMİN									
	Air_conditioner	Car_horn	Children_playing	Dog_bark	Drilling	Engine_idling	Gun_shot	Jackhammer	Siren	Street_music
Air_conditioner	81	1	1	1	3	0	0	0	1	1
Car_horn	1	37	2	1	2	1	0	3	3	5
Children_playing	4	0	79	7	1	3	2	1	3	11
Dog_bark	1	3	8	83	2	4	0	1	2	2
Drilling	2	0	0	1	110	1	1	5	2	3
Engine_idling	0	1	1	0	2	92	0	2	0	0
Gun_shot	0	0	0	0	0	0	35	0	0	0
Jackhammer	2	0	0	0	5	3	0	75	0	0
Siren	3	3	7	1	0	0	0	0	74	5
Street_music	1	0	10	1	0	3	0	0	3	59
FOLD-3: Doğruluk %82.9519										
GERÇEK	TAHMİN									
	Air_conditioner	Car_horn	Children_playing	Dog_bark	Drilling	Engine_idling	Gun_shot	Jackhammer	Siren	Street_music
Air_conditioner	99	0	1	0	0	0	0	1	0	1
Car_horn	0	32	2	1	2	0	1	2	1	4
Children_playing	6	1	55	13	1	0	0	1	12	14
Dog_bark	3	1	3	73	2	0	0	0	4	5
Drilling	3	2	0	1	90	1	1	5	2	0
Engine_idling	2	0	1	0	0	88	0	1	2	2
Gun_shot	0	1	0	1	0	0	37	0	0	0
Jackhammer	0	0	0	0	1	4	0	93	0	1
Siren	0	3	0	1	1	2	0	0	78	2
Street_music	4	5	10	0	2	1	0	5	1	78
FOLD-5: Doğruluk %82.8179										
GERÇEK	TAHMİN									
	Air_conditioner	Car_horn	Children_playing	Dog_bark	Drilling	Engine_idling	Gun_shot	Jackhammer	Siren	Street_music
Air_conditioner	90	0	5	0	0	3	0	1	0	2
Car_horn	1	28	2	1	2	1	0	0	2	0
Children_playing	2	0	66	10	3	1	1	3	1	11
Dog_bark	1	0	7	97	0	0	0	0	2	1
Drilling	1	3	3	0	80	0	0	5	1	0
Engine_idling	2	1	0	0	2	104	0	0	1	0
Gun_shot	0	0	0	2	0	0	30	0	0	0
Jackhammer	0	1	0	0	4	0	0	98	0	0
Siren	1	1	8	1	2	1	0	3	64	7
Street_music	1	2	10	3	4	0	0	1	1	81
FOLD-6: Doğruluk %84.5361										

Tablo 5.12. (b) URBANSOUND8K_ESA224 modelinin test verisi üzerinde elde ettiği karışıklık matrisi (Fold 7-10)

GERÇEK	TAHMİN									
	Air_conditioner	Car_horn	Children_playing	Dog_bark	Drilling	Engine_idling	Gun_shot	Jackhammer	Siren	Street_music
	Air_conditioner	91	0	7	2	0	2	0	1	1
	Car_horn	0	28	5	0	1	2	0	0	5
	Children_playing	1	1	70	3	1	2	0	0	1
	Dog_bark	2	0	9	67	1	0	2	1	2
	Drilling	1	0	1	1	87	0	1	8	2
	Engine_idling	2	0	4	1	0	88	0	1	0
	Gun_shot	0	0	0	0	0	0	31	0	1
	Jackhammer	3	0	2	0	2	2	0	104	0
	Siren	3	0	5	0	1	1	0	0	88
	Street_music	0	2	12	2	0	1	0	0	2
FOLD-7: Doğruluk %84.8797										
GERÇEK	TAHMİN									
	Air_conditioner	Car_horn	Children_playing	Dog_bark	Drilling	Engine_idling	Gun_shot	Jackhammer	Siren	Street_music
	Air_conditioner	85	0	6	0	2	8	0	2	1
	Car_horn	0	18	3	1	5	0	0	1	1
	Children_playing	1	1	74	4	4	2	0	2	1
	Dog_bark	1	4	11	71	2	1	0	1	3
	Drilling	1	0	0	1	94	1	0	8	0
	Engine_idling	1	0	3	1	0	90	0	1	1
	Gun_shot	0	0	1	0	0	0	41	0	0
	Jackhammer	1	0	4	0	4	6	0	98	0
	Siren	1	2	9	2	0	1	0	2	80
	Street_music	1	3	17	1	1	3	0	2	4
FOLD-8: Doğruluk %81.2142										
GERÇEK	TAHMİN									
	Air_conditioner	Car_horn	Children_playing	Dog_bark	Drilling	Engine_idling	Gun_shot	Jackhammer	Siren	Street_music
	Air_conditioner	89	0	2	1	0	1	0	0	2
	Car_horn	1	33	3	0	1	0	0	1	2
	Children_playing	1	0	92	4	4	1	0	0	3
	Dog_bark	1	7	8	83	0	0	1	0	0
	Drilling	0	0	1	0	77	0	0	3	1
	Engine_idling	3	1	2	0	1	83	0	1	1
	Gun_shot	0	0	0	0	0	0	42	0	0
	Jackhammer	1	0	3	0	6	0	0	85	0
	Siren	0	1	2	0	0	0	0	0	88
	Street_music	3	3	13	5	3	0	0	2	4
FOLD-9: Doğruluk %85.6816										
GERÇEK	TAHMİN									
	Air_conditioner	Car_horn	Children_playing	Dog_bark	Drilling	Engine_idling	Gun_shot	Jackhammer	Siren	Street_music
	Air_conditioner	94	1	1	1	1	2	0	1	1
	Car_horn	1	22	1	5	1	0	0	1	2
	Children_playing	0	3	64	7	1	2	0	2	1
	Dog_bark	2	1	7	84	0	2	0	1	3
	Drilling	0	1	1	0	73	0	2	3	1
	Engine_idling	1	0	1	1	0	97	0	2	1
	Gun_shot	0	0	0	1	1	0	35	0	0
	Jackhammer	0	1	0	0	0	1	0	89	0
	Siren	0	3	2	2	0	1	0	1	84
	Street_music	2	4	13	2	2	0	0	2	2
FOLD-10: Doğruluk %84.1924										

Tablo 5.12’ de verilen 10-Kat Çapraz Doğrulama içerisinde en yüksek doğruluk oranı %86.13 değerine sahip Fold-1 olmuştur. En düşük sınıflama doğruluk oranı ise Fold-8 da %81.21 olarak elde edilmiştir.

Tablo 5.13. URBANSOUND8K_ESA224 modelinin 10-kat çaprazlama ile UrbanSound8K veri seti üzerinde %84.33 ortalama doğruluğun elde edildiği karışıklık matrisi

True Class	Air Conditioner	900	3	29	7	9	22	1	8	10	11	90.0%	10.0%
	Car Horn	5	295	23	19	19	8	1	8	18	33	68.8%	31.2%
	Children Playing	19	8	728	72	16	12	3	13	27	102	72.8%	27.2%
	Dog Bark	18	25	64	816	8	8	4	6	18	33	81.6%	18.4%
	Drilling	11	6	8	7	882	4	6	48	10	18	88.2%	11.8%
	Engine Idling	15	4	17	7	7	920		9	8	13	92.0%	8.0%
	Gun Shot		1	2	5	1		360	1	2	2	96.3%	3.7%
	Jackhammer	9	2	13		39	17		915		5	91.5%	8.5%
	Siren	8	19	42	8	7	10		6	804	25	86.5%	13.5%
	Street Music	19	23	119	21	16	13		14	31	744	74.4%	25.6%

89.6%	76.4%	69.7%	84.8%	87.8%	90.7%	96.0%	89.0%	86.6%	75.5%
10.4%	23.6%	30.3%	15.2%	12.2%	9.3%	4.0%	11.0%	13.4%	24.5%
Air Conditioner	Car Horn	Children Playing	Dog Bark	Drilling	Engine Idling	Gun Shot	Jackhammer	Siren	Street Music
Predicted Class									

URBANSOUND8K_ESA224 modelinin 10-Kat Çapraz Doğrulama (cross-validation) sonucunda Tablo 5.13'te elde edilen ortalama doğruluk oranına sahip karışıklık matrisine bakıldığında, ortalama doğruluk oranı %84.33 olduğu görülmüştür. Tablo 5.13'te en yüksek doğruluk sınıfının 374 test görüntüsünü 360 tanesini doğru etiketleyerek %96.30 başarı oranı elde edilen silah atış sesi olduğu anlaşılmıştır. En düşük sınıflandırma ise %68.80 ile araç korna sesi olduğu görülmüştür. Yapılan değerlendirmede en çok bu sınıfa ait kentsel seslerin diğer ses sınıflarıyla karıştırıldığı anlaşılmıştır. Genel olarak URBANSOUND8K_ESA32 modeli, araç korna sesi, oyun oynayan çocuk sesleri ve sokak müzik sesleri dışındaki diğer sınıflarda %80 üzeri bir doğruluk oranı ile sınıflandırma işlemi yaptığı görülmüştür.

5.5. Önceki Çalışmalarla Karşılaştırılması

Çevresel seslerin sınıflandırılması ile ilgili literatürde birden fazla çalışmaların olduğu bilinmektedir. Bu tez kapsamında kullanılan veri setleri üzerinde yapılan çalışmalar tablolar halinde verilmiştir. Tablolarda verilen tez kapsamında geliştirilen modeller 224x224x3 giriş görüntü boyutuna sahip modellerdir. Tablo 5.14'te ESC10 veri seti üzerinde yapılan bazı çalışmalar gösterilmiştir. Tablo 5.14'e bakıldığında tez kapsamında önerilen yöntemin diğerlerinden daha başarılı olduğu görülmüştür.

Tablo 5.14. ESC10 veri seti üzerinde yapılan çalışmaların karşılaştırılması

Çalışmayı Yapanlar	Kullanılan Yöntem	Doğruluk Değeri (%)
Piczak (Piczak 2015)	SVM	80
Pillos ve ark. (Pillos ve ark 2016)	Random Forest Multi-Layer	74.5
Su ve ark.(Su ve ark 2019)	MC-Net	72
Salamon ve Bello(Salamon ve Bello 2017)	SB-ConvNets	72
Salamon ve Bello(Salamon ve Bello 2015)	Spherical k-means	74
Karol J. Piczak (Piczak 2015)	CNN	80.5
Karol J. Piczak (Piczak 2015)	Ensemble (Random forest)	72.7
Khamparia ve ark. (Khamparia ve ark 2019)	Spectrogram Images (CNN + TDSN)	56.0
Önerilen Model (ESC10_ESA224)	CNN	82.25

Tez kapsamında ESC50 veri seti için geliştirilen modelin diğer çalışmalarla karşılaştırılması Tablo 5.15'te verilmiştir. Tablo 5.15 incelendiğinde önerilen modelin bir model dışında diğer yöntemlerden daha başarılı olduğu görülmüştür.

Tablo 5.15. ESC50 veri seti üzerinde yapılan çalışmaların karşılaştırılması

Çalışmayı Yapanlar	Kullanılan Yöntem	Doğruluk Değeri (%)
Karol J. Piczak (Piczak 2015)	CNN	64.9
Karol J. Piczak (Piczak 2015)	Ensemble (Random forest)	44.3
Boddapati ve ark.(Boddapati ve ark 2017)	Spectrogram Images (combined features + GoogleNet)	73.0
Khamparia ve ark.(Khamparia ve ark 2019)	Spectrogram Images (CNN + TDSN)	49.0
Önerilen Model (ESC50_ESA224)	CNN	72.15

UrbanSound8K veri seti için geliştirilen modelin literatürdeki diğer modellerle karşılaştırılması Tablo 5.16'da verilmiştir. Tablo 5.16'ya bakıldığında önerilen modelin diğer modellerden çok daha başarılı olduğu görülmektedir.

Tablo 5.16. UrbanSound8K veri seti üzerinde yapılan çalışmaların karşılaştırılması

Çalışmayı Yapanlar	Kullanılan Yöntem	Doğruluk Değeri (%)
Yan Chen ve ark. (Chen ve ark 2019)	Dilated convolution	%78
Karol J. Piczak (Piczak 2015)	Convolutional layers with max-pooling	%74
Justin Salamon ve Juan Pablo Bello (Badrinarayanan ve ark 2017)	Unsupervised feature learning	%73.6
Nelauzjon Maxudov ve ark. (Maxudov ve ark 2016)	Long segments/majority voting	%71.8
Justin Salamon ve ark. (Salamon ve ark 2014)	Baseline system	%68
Karol J. Piczak (Piczak 2015)	CNN	73.7
İnik ve Şeker (Seker ve Inik 2020)	CNN	82.5
Önerilen Model (URBANSOUND_ESA32)	CNN	88.60

6. SONUÇ VE ÖNERİLER

Yapılan bu çalışmada üç farklı ÇSS veri setinin sınıflandırılması için özgün derin öğrenme mimarileri geliştirilmiştir. Bu veri setleri sırasıyla UrbanSound8K, ESC10 ve ESC50 veri setleridir. Bu veri setleri literatürde ÇSS için iyi oluşturulmuş araştırma veri setleridir. Her bir veri seti için özgün ESA modelleri tasarlanmıştır. Model tasarım aşamasında en az parametreye sahip modeller elde edilmeye çalışılmıştır. Bu veri setlerindeki çevresel sesler sinyal formatından görüntü formatına dönüştürülmüştür. Böylelikle ses verilerinden oluşan her bir veri seti için görüntü formatında veriler elde edilmiştir. Bu verilerden 32x32x3 ve 224x224x3 boyutlarında iki farklı veri seti oluşturulmuştur. Üç farklı veri seti için toplamda görüntüden oluşan altı veri seti oluşturulmuştur. Bu altı veri seti için özgün birden fazla ESA modelleri tasarlanmıştır. Tasarlanan bu ESA modelleri içerisinde en yüksek doğruluk oranı veren modeller bu tez kapsamında sunulmuştur. Bu modeller isimlendirilirken veri setinin ismi ve görüntü boyutu bilgileri kullanılarak yapılmıştır. Böylelikle oluşturulan DÖ modellerinin isimleri sırasıyla, ESC10_ESA32, ESC10_ESA224, ESC50_ESA32, ESC50_ESA224, URBANSOUND8K_ESA32 ve URBANSOUND8K_ESA224'dir. Bu modeller 10-Kat Çapraz Doğrulama (10-Fold Cross Validation) yapılarak veri setleri ile eğitilmiştir.

ESC10_ESA32 modeli ortalama doğruluk oranı %80.75 olarak elde edilmiştir. Bu modelde horoz sesi sınıfı %92.50 olarak elde edilmiştir. ESC10_ESA224 modelinin elde ettiği ortalama doğruluk oranı %82.25'dir. Bu model ise ateş kıvılcım seslerini %90 başarı ile en yüksek doğruluk oranı ile sınıflandırmıştır. Yapılan değerlendirmede ESC10 veri seti üzerinde giriş görüntü boyutunun 224x224x3 olması doğruluk oranını artırdığı görülmüştür.

ESC50_ESA32 modeli ortalama doğruluk oranı %54.55 olarak elde edilmiştir. ESC50_ESA224 modelinin elde ettiği ortalama doğruluk oranı %72.15'dir. ESC50 veri seti üzerinde 32x32x3 giriş görüntü boyutuna sahip veri setinde modelin istenen başarıyı elde edemediği görülmüştür. Giriş görüntü boyunun düşük olması sebebiyle modelin ezberleme yaptığı anlaşılmıştır. 224x224x3 giriş görüntü boyutuna sahip veri setinde ise literatürde yapılan çalışmalara nispeten başarılı sonuç elde edilmiştir.

URBANSOUND8K_ESA32 modelinin elde ettiği ortalama doğruluk oranı %88.60'dır. Bu model silah atış sesini %96.30'luk bir doğruluk oranı ile en yüksek seviyede sınıflandırmıştır. URBANSOUND8K_ESA224 modelinin elde ettiği ortalama doğruluk

oranı ise %84.33 olduđu görülmüştür. ESC10 ve ESC50 veri setlerinden farklı olarak UrbanSound8K veri setinde giriş görüntü boyutu düşük olan model daha yüksek doğruluk oranı elde edilmiştir.

Bu tez kapsamında yapılan çalışmalarda çevresel seslerin görüntü formatına çevrilerek yüksek performanslı ESA mimarilerinin tasarlanabileceği anlaşılmıştır. Gelecek çalışmalarda çevresel seslerin sınıflandırılması için yeni bir veri setinin oluşturulması hedeflenmektedir. Ayrıca farklı mimarilerdeki DÖ modelleri tasarlanarak hem sinyal hem de görüntü formatındaki verilerin birleştirilerek kullanılması hedeflenmektedir.



KAYNAKLAR

- Abdoli S, Cardinal P, Koerich AL, 2019. End-to-end environmental sound classification using a 1D convolutional neural network. *Expert Systems with Applications*, 136, 252-63.
- Agha A, Ranjan R, Gan W-S, 2017. Noisy vehicle surveillance camera: A system to deter noisy vehicle in smart city. *Applied Acoustics*, 117, 236-45.
- Akbal E, 2020. An automated environmental sound classification methods based on statistical and textural feature. *Applied Acoustics*, 167, 107413.
- Aumond P, Lavandier C, Ribeiro C, Boix EG, Kambona K, D'Hondt E, Delaitre P, 2017. A study of the accuracy of mobile technology for measuring urban noise pollution in large scale participatory sensing campaigns. *Applied Acoustics*, 117, 219-26.
- Badrinarayanan V, Kendall A, Cipolla R, 2017. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE transactions on pattern analysis and machine intelligence*, 39, 12, 2481-95.
- Bengio Y, Courville A, Vincent P, 2013. Representation Learning: A Review and New Perspectives. *Ieee T Pattern Anal*, 35, 8, 1798-828.
- Bisot V, Serizel R, Essid S, Richard G, 2017. Feature learning with matrix factorization applied to acoustic scene classification. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 25, 6, 1216-29.
- Boddapati V, Petef A, Rasmusson J, Lundberg L, 2017. Classifying environmental sounds using image recognition networks. *Procedia computer science*, 112, 2048-56.
- Cao J, Cao M, Wang J, Yin C, Wang D, Vidal P-P, 2019. Urban noise recognition with convolutional neural network. *Multimedia Tools and Applications*, 78, 20, 29021-41.
- Castelluccio M, Poggi G, Sansone C, Verdoliva L, 2015. Land use classification in remote sensing images by convolutional neural networks. *arXiv preprint arXiv:1508.00092*.
- Chen Y, Guo Q, Liang X, Wang J, Qian Y, 2019. Environmental sound classification with dilated convolutions. *Applied Acoustics*, 148, 123-32.
- Chu S, Narayanan S, Kuo C-CJ, 2009. Environmental sound recognition with time-frequency audio features. *IEEE Transactions on Audio, Speech, and Language Processing*, 17, 6, 1142-58.
- Chu S, Narayanan S, Kuo C-CJ, Mataric MJ. Where am I? Scene recognition for mobile robots using audio features. 2006 IEEE International conference on multimedia and expo, 885-8.
- CINAR I, TASPINAR YS, SARITAS MM, KOKLU M, 2020. FEATURE EXTRACTION AND RECOGNITION ON TRAFFIC SIGN IMAGES. *Selçuk-Teknik Dergisi*, 19, 4, 282-92.
- Competition ILSVR, 2012. Available online: [http://www. image-net. org/challenges. LSVRC/](http://www.image-net.org/challenges/LSVRC/)(accessed on 27 December 2016).
- Crocco M, Cristani M, Trucco A, Murino V, 2016. Audio surveillance: A systematic review. *ACM Computing Surveys (CSUR)*, 48, 4, 1-46.
- Demir F, Turkoglu M, Aslan M, Sengur A, 2020. A new pyramidal concatenated CNN approach for environmental sound classification. *Applied Acoustics*, 170, 107520.

- Deng J, Berg A, Satheesh S, Su H, Khosla A, Fei-Fei L, 2012. Imagenet large scale visual recognition competition 2012 (ILSVRC2012). See net.org/challenges/LSVRC, 41.
- Deng X, Liu Q, Deng Y, Mahadevan S, 2016. An improved method to construct basic probability assignment based on the confusion matrix for classification problem. *Information Sciences*, 340, 250-61.
- Dev D, 2017. *Deep Learning with Hadoop*, Packt Publishing Ltd, p.
- Dhanalakshmi P, Palanivel S, Ramalingam V, 2011. Classification of audio signals using AANN and GMM. *Applied soft computing*, 11, 1, 716-23.
- Furkan A, 2019. YÜZ TANIMA SÜRECİNDE FARKLI CNN MODELLERİNİN PERFORMANS KARŞILAŞTIRMASI Erkan DUMAN Dr. Öğr. Üyesi, Fırat Üniversitesi.
- Geiger JT, Helwani K. Improving event detection for audio surveillance using gabor filterbank features. 2015 23rd European Signal Processing Conference (EUSIPCO), 714-8.
- Green M, Murphy D, 2020. Environmental sound monitoring using machine learning on mobile devices. *Applied Acoustics*, 159, 107041.
- VGG16 – Convolutional Network for Classification and Detection, 2018. Erişim tarihi 13.12.2021. Erişim adresi, <https://neurohive.io/en/popular-networks/vgg16/>.
- He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770-8.
- Huang J. Spatial auditory processing for a hearing robot. *Proceedings. IEEE International Conference on Multimedia and Expo*, 253-6.
- Huang X, Acero A, Hon H-W, Reddy R, 2001. *Spoken language processing: A guide to theory, algorithm, and system development*, Prentice hall PTR, p.
- İnik Ö, 2019. Derin öğrenme kullanarak ovaryum follikülerinin sınıflandırılması.
- İnik Ö, Altıok M, Ülker E, Koçer B, 2021. MODE-CNN: A fast converging multi-objective optimization algorithm for CNN-based models. *Applied Soft Computing*, 107582.
- İnik Ö, Ceyhan A, Balcıoğlu E, Ülker E, 2019. A new method for automatic counting of ovarian follicles on whole slide histological images based on convolutional neural network. *Computers in biology and medicine*, 112, 103350.
- Intani P, Orachon T. Crime warning system using image and sound processing. 2013 13th International Conference on Control, Automation and Systems (ICCAS 2013), 1751-3.
- Karmakar PK, 2011. *Acoustic Waves - From Microdevices to Helioseismology*.
- KAYA V, 2021. OTONOM GÜVENLİK KONTROLLERİNDE KULLANILMAK ÜZERE DERİN ÖĞRENME TABANLI SİLAH TESPİT VE TANIMA SİSTEMİ.
- Kaya V, Tuncer S, Baran A, 2020. *Derin Öğrenme Yöntemleri Kullanılarak Nesne Tanıma*.
- Khamparia A, Gupta D, Nguyen NG, Khanna A, Pandey B, Tiwari P, 2019. Sound classification using convolutional neural network and tensor deep stacking network. *IEEE Access*, 7, 7717-27.
- Krizhevsky A, Sutskever I, Hinton GE, 2012. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25, 1097-105.

- Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 1097-105.
- Laffitte P, Wang Y, Sodoyer D, Girin L, 2019. Assessing the performances of different neural network architectures for the detection of screams and shouts in public transportation. *Expert systems with applications*, 117, 29-41.
- LeCun Y, 1987. Modèles connexionistes de l'apprentissage, Université de Paris VI. 18, 504, 517.
- LeCun Y, Bengio Y, Hinton G, 2015. Deep learning. *nature*, 521, 7553, 436-44.
- LeCun Y, Bottou L, Bengio Y, Haffner P, 1998. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86, 11, 2278-324.
- Lee H, Grosse R, Ranganath R, Ng AY. Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations. *Proceedings of the 26th annual international conference on machine learning*, 609-16.
- Li H, Ishikawa S, Zhao Q, Ebana M, Yamamoto H, Huang J. Robot navigation and sound based position identification. 2007 IEEE International Conference on Systems, Man and Cybernetics, 2449-54.
- Li S, Yao Y, Hu J, Liu G, Yao X, Hu J, 2018. An ensemble stacked convolutional neural network model for environmental event sound recognition. *Applied Sciences*, 8, 7, 1152.
- Lim M, Lee D, Park H, Kang Y, Oh J, Park J-S, Jang G-J, Kim J-H, 2018. Convolutional Neural Network based Audio Event Classification. *KSII Transactions on Internet & Information Systems*, 12, 6.
- Lim M, Lee D, Park H, Kang Y, Oh J, Park J-S, Jang G-J, Kim J-H, 2018. Convolutional neural network based audio event classification. *KSII Transactions on Internet and Information Systems (TIIS)*, 12, 6, 2748-60.
- Ludena-Choez J, Gallardo-Antolin A, 2016. Acoustic Event Classification using spectral band selection and Non-Negative Matrix Factorization-based features. *Expert Systems with Applications*, 46, 77-86.
- Lyon RF, 2010. Machine hearing: An emerging field [exploratory dsp]. *IEEE signal processing magazine*, 27, 5, 131-9.
- Maxudov N, Özcan B, Kırac MF. Scene recognition with majority voting among sub-section levels. 2016 24th Signal Processing and Communication Application Conference (SIU), 1637-40.
- McCulloch WS, Pitts W, 1943. A logical calculus of the ideas immanent in nervous activity. *The Bulletin of Mathematical Biophysics*, 5, 4, 115–33.
- Medhat F, Chesmore D, Robinson J, 2020. Masked Conditional Neural Networks for sound classification. *Applied Soft Computing*, 90, 106073.
- Mesaros A, Heittola T, Virtanen T. TUT database for acoustic scene classification and sound event detection. 2016 24th European Signal Processing Conference (EUSIPCO), 1128-32.
- Minsky MLaP, S. A. , 1969. *Perceptrons*. MIT Press, Cambridge. 15.
- MOREAU R, 2019. The emission, propagation and perception of sound.
- Mulimani M, Koolagudi SG, 2019. Segmentation and characterization of acoustic event spectrograms using singular value decomposition. *Expert Systems with Applications*, 120, 413-25.
- Mushtaq Z, Su S-F, 2020. Environmental sound classification using a regularized deep convolutional neural network with data augmentation. *Applied Acoustics*, 167, 107389.

- Mushtaq Z, Su S-F, Tran Q-V, 2021. Spectral images based environmental sound classification using CNN with meaningful data augmentation. *Applied Acoustics*, 172, 107581.
- Ntalampiras S, 2014. Universal background modeling for acoustic surveillance of urban traffic. *Digital Signal Processing*, 31, 69-78.
- Özkan İ, Ülker E, 2017. Derin öğrenme ve görüntü analizinde kullanılan derin öğrenme modelleri. *Gaziosmanpaşa Bilimsel Araştırma Dergisi*, 6, 3, 85-104.
- Piczak KJ. Environmental sound classification with convolutional neural networks. 2015 IEEE 25th international workshop on machine learning for signal processing (MLSP), 1-6.
- Piczak KJ. ESC: Dataset for environmental sound classification. *Proceedings of the 23rd ACM international conference on Multimedia*, 1015-8.
- Pillos A, Alghamidi K, Alzamel N, Pavlov V, Machanavajhala S, 2016. A real-time environmental sound recognition system for the Android OS. *Proceedings of Detection and Classification of Acoustic Scenes and Events*.
- Radhakrishnan R, Divakaran A, Smaragdis A. Audio analysis for surveillance applications. *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, 2005., 158-61.
- Rosenblatt F, 1958. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological review*, 65, 6, 386–408.
- Rosenblatt F, 1962. *Principles of Neurodynamics*, Spartan, New York. 15, 27, p.
- Rumelhart DE, Hinton GE, Williams RJ, 1986. Learning representations by back-propagating errors. *Nature*, 323, 6088, 533–6.
- Salamon J, Bello JP. Unsupervised feature learning for urban sound classification. 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 171-5.
- Salamon J, Bello JP, 2017. Deep convolutional neural networks and data augmentation for environmental sound classification. *IEEE Signal processing letters*, 24, 3, 279-83.
- Salamon J, Jacoby C, Bello JP. A dataset and taxonomy for urban sound research. *Proceedings of the 22nd ACM international conference on Multimedia*, 1041-4.
- Sammut C, Webb GI, 2011. *Encyclopedia of machine learning*, Springer Science & Business Media, p.
- Seker H, Inik O. CnnSound: Convolutional Neural Networks for the Classification of Environmental Sounds. 2020 The 4th International Conference on Advances in Artificial Intelligence, 79-84.
- Sharma S, Sharma S, Athaiya A, 2017. Activation functions in neural networks. *towards data science*, 6, 12, 310-6.
- Simonyan K, Zisserman A, 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.
- Steele D, Krijnders J, Guastavino C, 2013. The sensor city initiative: cognitive sensors for soundscape transformations. *GIS Ostrava*, 1-8.
- Stowell D, Giannoulis D, Benetos E, Lagrange M, Plumbley MD, 2015. Detection and classification of acoustic scenes and events. *IEEE Transactions on Multimedia*, 17, 10, 1733-46.
- Su Y, Zhang K, Wang J, Madani K, 2019. Environment sound classification using a two-stream CNN based on decision-level fusion. *Sensors*, 19, 7, 1733.

- Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A. Going deeper with convolutions. Proceedings of the IEEE conference on computer vision and pattern recognition, 1-9.
- Takahashi N, Gygli M, Pfister B, Van Gool L, 2016. Deep convolutional neural networks and data augmentation for acoustic event detection. arXiv preprint arXiv:1604.07160.
- Tammina S, 2019. Transfer learning using vgg-16 with deep convolutional neural network for classifying images. International Journal of Scientific and Research Publications (IJSRP), 9, 10, 143-50.
- Tokozume Y, Ushiku Y, Harada T, 2017. Learning from between-class examples for deep sound recognition. arXiv preprint arXiv:1711.10282.
- Widrow B, Hoff ME, 1960. Adaptive switching circuits. Adaptive switching circuits. In 1960 IRE WESCON Convention Record, volume 4, pages 96–104. IRE, New York. 15, 21, 24, 27.
- Xie J, Zhu M, 2019. Investigation of acoustic and visual features for acoustic scene classification. Expert Systems with Applications, 126, 20-9.
- Zhang Q, Bai C, Chen Z, Li P, Yu H, Wang S, Gao H, 2021. Deep learning models for diagnosing spleen and stomach diseases in smart Chinese medicine with cloud computing. Concurrency and Computation: Practice and Experience, 33, 7, 1-.
- Zhang W, Feng X, Goerlandt F, Liu Q, 2020. Towards a Convolutional Neural Network model for classifying regional ship collision risk levels for waterway risk analysis. Reliability Engineering & System Safety, 204, 107127.
- Zhang X, Zou Y, Shi W. Dilated convolution neural network with LeakyReLU for environmental sound classification. 2017 22nd International Conference on Digital Signal Processing (DSP), 1-5.