
# SES SİNYALLERİNİN GRAF TABANLI TEMSİLLERİNİN YAPAY ZEKÂ YÖNTEMLERİ İLE SINIFLANDIRILMASI

**Serkan AKSU**

**2022**
**DOKTORA TEZİ**

**Tez Danışmanı**
**Doç. Dr. İlker TÜRKER**

# SES SİNYALLERİNİN GRAF TABANLI TEMSİLLERİNİN YAPAY ZEKÂ YÖNTEMLERİ İLE SINIFLANDIRILMASI

**Serkan AKSU**

**T.C.**
**Karabük Üniversitesi**
**Lisansüstü Eğitim Enstitüsü**
**Bilgisayar Mühendisliği Anabilim Dalında**
**Doktora Tezi**
**Olarak Hazırlanmıştır**

**Tez Danışmanı**
**Doç. Dr. İlker TÜRKER**

**KARABÜK**
**Mart 2022**

Serkan AKSU tarafından hazırlanan "SES SİNYALLERİNİN GRAF TABANLI TEMSİLLERİNİN YAPAY ZEKÂ YÖNTEMLERİ İLE SINIFLANDIRILMASI" başlıklı bu tezin Doktora Tezi olarak uygun olduğunu onaylarım.

Doç. Dr. İlker TÜRKER .....
Tez Danışmanı, Bilgisayar Mühendisliği Anabilim Dalı

Bu çalışma, jürimiz tarafından Oy Birliği ile Bilgisayar Mühendisliği Anabilim Dalında Doktora tezi olarak kabul edilmiştir. 23/03/2022

| Ünvanı, Adı SOYADI (Kurumu) | İmzası |
|---|---|
| Başkan : Doç. Dr. Ümit ATİLA (GÜ) | ..... |
| Üye : Prof. Dr. Ergin YILMAZ (BEÜ) | ..... |
| Üye : Doç. Dr. İlker TÜRKER (KBÜ) | ..... |
| Üye : Dr. Öğr. Üyesi Fuat ŞİMŞİR (KBÜ) | ..... |
| Üye : Dr. Öğr. Üyesi Burhan SELÇUK (KBÜ) | ..... |

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

Prof. Dr. Hasan SOLMAZ .....
Lisansüstü Eğitim Enstitüsü Müdürü

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

Serkan AKSU

# ÖZET

**Doktora Tezi**

## SES SİNYALLERİNİN GRAF TABANLI TEMSİLLERİNİN YAPAY ZEKÂ YÖNTEMLERİ İLE SINIFLANDIRILMASI

**Serkan AKSU**

**Karabük Üniversitesi**
**Lisansüstü Eğitim Enstitüsü**
**Bilgisayar Mühendisliği Anabilim Dalı**

**Tez Danışmanı:**
**Doç. Dr. İlker TÜRKER**
**Mart 2022, 98 sayfa**

Bu çalışmada, ses sinyallerinin zaman boyutundaki komşu genlik seviyeleri arasında bulunan ilişkiye dayalı graf temelli yeni bir temsil yöntemi geliştirilmiştir. Karmaşık ağ biliminin sinyal işleme alanına uyarlandığı bu yaklaşımda zaman boyutundaki genlik seviyeleri ve bunların komşuları arasında bağlantı dikkate alınır. İlk olarak ses sinyalleri, önceden belirlenmiş *n-bit* seviyesine göre ölçeklenir. Ölçeklenmiş sinyallere 3 farklı değerde uygulanan alt-örnekleme ile 3 farklı bağlantı matrisi (graf) elde edilir. Daha sonra üst üste yerleştirilen bu 3 matrisin sırasıyla RGB katmanlarını temsil ettikleri bir imaj oluşturulmuş olur. Ses sinyallerinin segmentasyonundan elde edilen sinyal parçalarına ayrı ayrı uygulanan bu yöntem sonucunda her bir zaman çerçevesi için $[2^n \times 2^n \times 3]$ boyutunda RGB-imajı elde edilmiş olur. Bu RGB kare matrisler daha sonra dikey formatta düzleştirilerek $[2^{2n} \times 3]$ büyüklüğünde tek boyutlu RGB dizisine dönüştürülür.

Art arda gelen çerçevelerden elde edilen bu dikey diziler yatay eksende birleştirilir ve *connectogram* adını verdiğimiz $[2^{2n} \times segment\ sayısı \times 3]$ boyutunda bir temsil imajı elde edilmiş olur. Böylece ses sinyalleri zaman-graf eksenli *connectogram* adını verdiğimiz farklı bir yöntemle temsil edilmiş olur.

Sesin *connectogram* şeklinde temsil edilmesi ile elde edilen bu yeni yöntemin çevresel sesler üzerindeki sınıflandırma başarısı, mel-spektrogram (mels) ve MFCC gibi bilinen yöntemler ile karşılaştırılarak test edilmiştir. Bu test işlemi için temsil matrisleri imaja dönüştürülmüş ardından bu imajlar bilinen en yeni transfer öğrenme modellerine girdi olarak verilmiştir. Elde edilen sonuçlar, *connectogram*'ların tek başına kullanıldığında mel-spektrogram ile rekabet edecek şekilde en iyi sonucu vermediğini göstermiştir. Fakat, bu matrisler sesin mel-spektrogram temsili ile RGB formatının bir katmanı olacak şekilde [*mels* + *mels* + *connectogram*] şeklinde birleştirildiğinde sınıflandırma başarısında 2% gibi önemli bir artış sağlandığı görülmüştür. Yapılan sınıflandırma denemelerinde en iyi sonucun 5-fold çapraz doğrulama ile ResNet50 modeli üzerinde 95.59 % olarak elde edilmiştir.

**Anahtar Sözcükler** **:** Graf, ses sınıflandırma, zaman serileri ile sınıflandırma, karmaşık ağlar, derin öğrenme.

**Bilim Kodu** **:** 92410

## ABSTRACT

**Ph. D. Thesis**

**CLASSIFICATION OF GRAPH-BASED REPRESENTATIONS OF AUDIO SIGNALS BY ARTIFICIAL INTELLIGENCE METHODS**

**Serkan AKSU**

**Karabük University**
**Institute of Graduate Programs**
**Department of Computer Engineering**

**Thesis Advisor:**
**Assoc. Prof. Dr. İlker TÜRKER**
**March 2022, 98 pages**

We introduce a graph-inspired representation for sounds capturing temporal convexity characteristics based on deviations in amplitude levels. Assuming the quantized amplitude levels as nodes with a pre-defined bit depth (*n*), a network-theoretic approach is conducted to establish connections between these amplitude levels based on their neighborhood in time domain. This procedure is run for 3 downsampling rates, resulting in a 3-layer adjacency matrix representation for a single time frame after segmentation, that is combined to form an RGB-image of size $[2^n \times 2^n \times 3]$ for each segment. These matrices are further flattened to $[2^{2n} \times 3]$ vertical RGB-arrays, derived from each sound frame. Tiling these vertical arrays from consecutive frames horizontally, we generate a time-graph representation of size $[2^{2n} \times num.segments \times 3]$ named *connectogram*, capturing the temporal convexity characteristics of sound waves.

The representation capability of *connectogram*s is tested in comparison with mel-spectrograms (mels) and MFCCs for an environmental sound classification task, as input to state-of-art transfer learning models. Results indicate that *connectogram*s cannot compete with the best-performer mel-spectrogram representations in standalone format, however they significantly improve their classification performance in case they are combined as single layers of hybrid RGB representations. A combination of [*mels+mels+connectogram*] outperforms either sole representations or their combinations by 2%, with 95.59 % classification accuracy with 5-fold cross validation for ResNet50 classifier model.

**Key Words** **:** Graph representation, sound classification, time-series classification, complex networks, deep learning.

**Science Code** **:** 92410

## TEŞEKKÜR

Bu tez çalışmasının planlanmasında, araştırılmasında, yürütülmesinde ve oluşumunda desteğini esirgemeyen, yönlendirme ve bilgilendirmeleriyle çalışmamı bilimsel temeller ışığında şekillendiren sayın hocam Doç. Dr. İlker TÜRKER'e teşekkürlerimi sunarım.

Tez çalışması sürecinde bilimsel ve akademik anlamda desteklerini esirgemeyen tez izleme komitesindeki değerli hocalarım Doç Dr. Ümit ATİLA'ya ve Dr. Öğr. Üyesi Fuat ŞİMŞİR'e teşekkür ederim.

Hayatımın her aşamasında sabırla bana maddi ve manevi destek veren sevgili eşim ve çocuklarıma da en kalbi teşekkürü borç bilirim.

## İÇİNDEKİLER

**Sayfa**

**Sayfa**

**Sayfa**

## ŞEKİLLER DİZİNİ

**Sayfa**

**Sayfa**

## ÇİZELGELER DİZİNİ

**Sayfa**

# SİMGELER VE KISALTMALAR DİZİNİ

## SİMGELER

$\boldsymbol{\sigma}$ : Standart sapma

$\boldsymbol{T}$ : Pariyod

$\boldsymbol{\omega}$ : Açısal frekans

$\boldsymbol{f}$ : Sinyal frekansı

$\boldsymbol{F}$ : Örnekleme frekansı

$\boldsymbol{\phi}$ : Faz farkı

$\boldsymbol{\Sigma}$ : Toplam Sembolü

## KISALTMALAR

**CNN** : Convolutional Neural Network

**CQT** : Constant Q-Transform

**ÇKASA** : Çok Katmanlı Algılayıcı Sinir Ağlar

**DTFT** : Discreate Time Fourier Transform

**DFT** : Discreate Fourier Transform

**DWT** : Discreate Wavelet Transform

**DYSA** : Derin Yapay Sinir Ağlar

**EEG** : Electro Encephalo Graphy

**ESA** : Evrişimsel Sinir Ağları

**FFT** : Fast Fourier Transform

**GFCC** : Gammatone Frequency Cepstral Coefficients

**GG** : Görünürlük Grafları

**STFT** : Short Time Fourier Transform

**SVM** : Destek Vektör Makineleri

**MFCC** : Mel-Frequency Cepstral Coefficients

**OAS** : Optimum Allocation Sampling

**SSL** : Semi-Supervised Learning

**STFT** : Short-Time Fourer Transform

**TFR** : Time–Frequency Representation

**YSA** : Yapay Sinir Ağları

**YZ** : Yapay Zekâ

# BÖLÜM 1

# GİRİŞ

Derin Yapay Sinir Ağları (DYSA), sınıflama yöntemlerinin gelişmesi ile son yıllarda ses sınıflandırma çalışmaları önemli ölçüde güçlendirilmiştir. Bu alanda özellikle konuşma tanıma, müzik bilgilerinin belirlenmesi ve ortam seslerinin sınıflandırılması gibi çalışmalar yapılmış ve oldukça yüksek başarılar elde edilmiştir. Konuşma ve müzik kayıtlarından farklı olarak çevresel sesler, çeşitli kaynaklardan üretilen farklı gürültülerin de eklenmiş olduğu ses sinyalleridir. Yapay zekâ çalışmaları için ortam seslerinin sınıflandırılması çok hayati bir öneme sahiptir ve gözetleme, ev otomasyonu ve güvenlik gibi farklı alanlarda yaygın olarak uygulanmaktadır [1].

Zaman serileri; biyomedikal sensörlerden elde edilmiş kayıtlar [2,3], finansal kayırlar [4], endüstriyel sensörlerden elde edilmiş kayıtlar [5], hava durumu aktiviteleri [6], ses olayları [7] gibi tek veya çok kanallı [8] verileri işlemek için yaygın olarak kullanılmaktadır. Zaman serileri ile ilgili çalışmalar; örüntü tanıma, sınıflandırma, kümeleme, özetleme gibi amaçlara odaklanmıştır [9]. Ana görev olarak sınıflandırma, özellik çıkarmaya dayalı klasik makine öğrenme yaklaşımları, derin öğrenme modelleri ve çeşitli sınıflandırıcıların birleşimine dayalı birçok sınıflandırma yaklaşımı ile güçlendirilmiştir [10,11].

Ses sınıflandırmayı içeren ana akım çalışmalar, örneklenecek seslerin içerdiği gürültüye karşı daha sağlam temsil verileri elde edilmesini sağlayan *mel-frequency cepstral coefficients (MFCCs)* gibi mel-spektrogram imajları kullanmaktadır. Buradaki coefficient'ler seslerin kısa süreli güç spektrumunun yakalanmasını sağlamaktadır. Spektrogram imajları sinyallerin belli zaman dilimlerine bölünmesi ve bu dilimlere ayrık Fourier dönüşümleri (AFD) uygulanması sonucu elde edilir. AFD işleminden sonra sabit-

uzunluklu pencereler elde edilmiş olur ve son aşamada bu pencerelere MFCC'lerden elde edilmiş mel-filtreler uygulanır [12]. Cochleagram, mel-spektrogram gibi sesin bir zaman-frekans gösterimidir. Bu yöntemde, kulak salyangozu olarak ta bilinen ve insanın iç kulağının işitsel kısmını oluşturan cochlea'nın özelliği kullanılarak sesin frekans dağılımı üzerine logaritmik bir işlem uygulanır. Daha düşük frekans aralığında daha fazla frekans bileşeni ve daha yüksek frekans aralığında daha düşük frekans bileşeni bulunduğu için cochleagramı alınmış seslerden daha fazla ayrıntı elde edilebilmektedir [13].

Graf kaynaklı yöntemler yapılandırılmış biçimleri ile, zaman serisi verilerini temsil etmek için ESA tabanlı makine öğrenmesi yöntemlerine güçlü bir şekilde girdi oluşturacak alternatif yapılar sunmaktadır. Zaman serilerinin karmaşık ağlara dönüştürülmesinde görünürlük grafı, daha düşük karmaşıklığa sahip çözümler sunmaktadır. Bu yöntemle zaman-serileri komşuluk matrisi şeklinde temsil edilebilmektedir. Komşuluk matrisleri, sayısallaştırılmış ve ölçeklenmiş ardışık sinyal seviyeleri arasında bağlantıları temsil etmektedir [14].

Oluşturulan *görünürlük grafı,* özellikle sinyallerin genlik seviyelerinin ardışık sıralaması şeklinde temsil edilen zaman serilerinin bazı özelliklerini miras alır. Sonuç olarak elde edilmiş bu *görünürlük grafları*, sınıflandırma işlemine girdi olarak verilmek üzere işlenmemiş ham graflar veya daha küçük boyutlu graflar için belli proseslere tabi tutulmaya uygun matrisler olarak temsil edilmektedir [15–17].

Doğrusal olmayan dinamiklerin yapısal organizasyonunu karakterize etme yetenekleri nedeniyle zaman serileri için network tabanlı teorik yaklaşımlar gittikçe artan bir ilgi odağı olmuştur. Birbirleriyle etkileşim halinde olan bileşenlerden oluşan her sistem bir network olarak tanımlanma potansiyeline sahiptir [18–21]. *Zaman serisi ağların* (ZSA) oluşturulması için kullanılan metotlar üç ana alt bölüme ayrılmıştır. *Proximity ağlar*, aynı yapı içerisinde bulunan çok sayıdaki zaman serisi veri arasında oluşan benzerlikleri yakalar. İnsan beynine bağlanan birden fazla elektrottan aldığı çok kanallı sinyalleri içeren Electro Encephalo Gram (EEG) verileri bu yöntem için uygun bir örnektir. Bu yöntem ayrıca *fonksiyonel ağlar* olarakta bilinmektedir [21,22]. *Görünürlük grafları*, tek kanallı

sensörlerden elde edilen veriler gibi tek değişkenli bir zaman serisinde bulunan sıralı örneklerin değişkenliği arasındaki ilişkiye göre oluşturulur [14].

Geçiş ağları, dinamik bir sistemin durumları arasındaki geçiş olasılıklarını göstermek için kullanılır. Bu tür bir yapıda bağlantı olasılıklarını tanımlamadan önce sistemlerin bu durumların ayrıklaştırılması gerekir [23]. Tüm bu ağ temsillerinin amacı, zaman serisi kayıtlardan oluşan bir varlığı derin öğrenme modelleri için güçlü bir veri oluşturacak şekilde bir graf gösterimine dönüştürmektir. Bu üst düzey temsillerin ortaya çıkmış ve kullanılmaya başlanmış olması derin öğrenme performansının artması noktasında umut verici gelişmelere yol açmaktadır ve *graf tabanlı öğrenme olarak* tanımlanacak yeni bir yaklaşımın oraya çıkmasını sağlamıştır [24].

Bu çalışmada özellikle ses dalgaları gibi zaman serilerinin graf tabanlı temsillerinin nasıl yapılacağı incelenmiş ve bu graf gösteriminin halen kullanılmakta olan imaj tabanlı temsillere güçlü bir alternatif olabileceği gösterilmiştir. Bir zaman serisinin ardışık örnekleri için sinyaldeki değişkenlik karakteristiklerinin gösterimi amacıyla sinyalden alınan sabit uzunluklu çerçeveler kare matris şeklinde graf gösterimine dönüştürülmüştür. Daha sonra bu matrisler, normalize edilmiş genlik seviyeleri arasındaki komşuluk aktivitesini göstermek amacıyla dikey vektörler şeklinde düzleştirilmiştir. Her çerçeve için oluşturulmuş bu dikey dilimler, bir sonraki adımda *connectogram* olarak adlandırılan *zaman-graf* gösterimi için yatay olarak birleştirilir.

*Connectogram*'da yatay eksen zamanı, dikey eksen ise sinyal dalgalanmalarını göstermektedir. Elde edilen bu gösterim biçimi ESC-10 veri seti kullanılarak ortam seslerinin sınıflandırılmasında test edilmiştir. Bu amaçla *connectogramlar* DYSA'lar için en uygun girdi biçimi olan RGB katmanlarına dönüştürülmüştür. Elde edilmiş olan bu *connectogram* gösterimi özellikle RGB katmanlarından biri olarak spektrogram ile birleştirildiğinde imaj tabanlı gösterimlere ek olarak sinyalin genlik seviyesi ile ilgili spesifik özellikleri de temsil ettiğinden ses ve biyomedikal sinyal sınıflandırmada başarının önemli ölçüde artmasını sağlamaktadır.

Bu metot, ayrıca spesifik boyutların gereksinimlerini karşılayan yüksek seviyeli temsiller için kenar özelliklerinin çıkartılması yerine grafçıklar olarak kodlanmaya elverişlidir.

## 1.1. ÇALIŞMANIN LİTERATÜRDEKİ YERİ

Ses sınıflandırma amacıyla son yıllarda özellikle DYSA'lara dayalı birçok çalışma yapılmıştır. Özellikle canlı ve farklı mekanik ortamlarda üretilen çevresel seslerden oluşan veri tabanları geliştirilen bu yöntemlerin test edilmesi amacıyla kullanılmıştır.

Zaman serileri ve ses sinyalleri ile ilgili literatür incelendiğinde bu alanda gerçekleştirilen çalışmalar aşağıdaki şekilde özetlenebilir.

Özellikle çevresel seslerin sınıflandırılması konusunda literatürde çok fazla sayıda çalışma bulunmaktadır. DYSA'ya dayalı görüntü sınıflandırma yöntemlerinin de gelişmesi ile seslerden elde edilen spektrogram ve MFCC gibi görüntüler derin öğrenme uygulamalarına girdi olarak verilmiştir.

Bu çalışmalardan biri de Boddapati vd. tarafından yapılmıştır [25]. Bu çalışmada, çeveresel seslerin Spektrogram, MFCC ve CRP temelli imaj temsilleri bir RGB imajının katmanları olarak kullanılmış ve elde edilen bu RGB imajı AlexNet [26] ve GoogleNet [27] gibi görüntü tanıma kütüphanelerinde sınıflandırılmıştır. Sonuç olarak ESC-10 veri seti için AlexNet ile 86% ve GoogleNet ile 86% oranında sınıflandırma başarısı elde edilmiştir. Veri seti olarak ESC-50 kullanıldığında ise AlexNet ile 65% ve GoogleNet ile 73% başarı oranları elde edilmiştir. Bu çalışmada en iyi sonuç ise UrbanSound8K veri seti üzerinde GoogleNet ile spektrogramlar kullanılarak 93% olarak elde edilmiştir.

Bagnall ve arkadaşları tarafından COTE (Collective Of Transformation-based Ensembles) olarak adlandırılan derin olmayan üst düzey bir sınıflandırıcı yaklaşımı önerilmiştir. Bu yöntemde sinyaller veya DWT özellikleri şekilcikler olarak dönüştürülmüş ve 35 sınıflandırıcı topluluğundan oluşan bir şekle sokulmuştur [25]. Bu model, hiyerarşik oylama mekanizması ve HIVE-COTE olarak bilinen ilave iki sınıflayıcı ile selefine göre

önemli ölçüde geliştirilmiştir ve şu anda derin olmayan modeller arasında zaman serisi sınıflandırması için en gelişmiş algoritma olarak kabul edilmektedir [26].

Bilgisayarla görme ve örüntü tanımada güçlü bir araç olarak ortaya çıkan Derin Yapay Sinir Ağları (DYSA) zaman serisi sınıflandırma (ZSS) problemlerinde gittikçe daha fazla iyileştirme sağlamıştır. DYSA modelleri olarak Long Short-Term Memory (LSTM) [27], Gated Recurrent Units (GRU) [28], Temporal Convolution Networks (TCN) [29] gibi yöntemler vektör formatında zaman serilerini işleyen DYSA modelleri olarak zamansal modellemede başarılı sonuçlara yol açmıştır. DYSA'ların derin olmayan modellere göre başlıca avantajları, veri ön işlemede yoğun işleme ihtiyacından kaçınması olmuştur. Ayrıca, DYSA'lar derin olmayan makine öğrenimi modellerine göre önemli ölçüde daha iyi performans göstermiş ve COTE ve HIVE-COTE ile karşılaştırılabilir sonuçlar elde edilmesini sağlamıştır [30].

Bilgisayarla görme alanında devrim niteliğinde gelişmelere yol açan Evrişimsel Sinir Ağları (ESA), gizli örüntüleri ortaya çıkarmak için güçlü yetenekler ortaya koymuştur [9]. Son yıllarda yüksek boyutlu verilerdeki problemleri ele almak için ESA'ları kullanma eğiliminin artması ile ses sınıflandırma alanında da önemli başarılar elde edilmiştir. Ortam seslerinin sınıflandırılması, otomatik konuşma tanıma, müzik bilgisi tanıma gibi akustik sınıflandırma problemleri için spektrogram veya cochleagram gibi sesin görüntü şeklinde temsil edilen zaman-frekans gösterimleri ESA'lara giriş verisi olarak verilmektedir [1,31–33].

Khamparia vd.'nin 2019'da yaptıkları çalışmada çevresel seslerin spektrogram imajlarını sınfılandırmak için Convolutional Neural Network (CNN) ve Tensor Deep Stacking Network (TDSN) derin öğrenme modellerini kullanmışladır [37]. Bu çalışamda ESC-10 bve ESC-50 veri setleri kullanılmış ve her iki sistem de bu veri setleri ile eğitilmiştir. Sonuç olarak CNN ile 77% ve 49% başarı oranları elde edillmiştir. TDSN modeli ESC-10 ile eğitildiğinde ise 56% oranında bir başarı sağlanmıştır.

2019'da Strisciuglio vd. tarafından yapılan çalışmada 4 farklı veri seti üzerinde ses sınıflandırma işlemi gerçekleştirilmiştir [11]. Seslerin özelliklerini elde etmek için mel-spektrogram'lardan COPE özellik çıkarma yöntemi kullanılmıştır. COPE yöntemi, seslerin *Gammatonegram* gösterimlerinin lokal enerji piklerini kullanarak bir özellik elde edilmesini sağlar. Bu yöntemi seslere sonradan eklenmiş gürültülere karşı oldukça dayanıklı olduğu tespit edilmiştir [38]. Sınıflayıcı olarak ise birden-hepsine şemasına göre tasarlanmış çok-sınıflı SVM sınıflayıcılar kullanılmıştır. Bu çalışmada veri seti olarak ise MIVIA audio events, MIVIA road events, ESC-10 ve TU Dortmund veri setleri kullanılmış ve sırası ile 91.71%, 94%, 81.25% ve 94.27% oranlarında sınıflandırma başarıları elde edilmiştir.

Tang vd. tarafından 2019'da Evrişimsel Sinir Ağları (ESA) tabanlık bir sınıflandırma çalışması gerçekleştirilmiştir. Çalışmada, sesler 22050 Hz frekansında örneklenmiş her bir ses sinyali 50% örtüşme ile 81 çerçeveye ayrılmıştır [39]. Seslerin özelliklerini elde etmek için *log-scaled mel-spektrogram* temsilleri kullanılmıştır. Bu amaçla 512 bölümden oluşan FFT'ler uygulanmış ve 80 tane *mel-band* tercih edilmiştir. ESA'nın eğitimi için ise Caffe [40] adlı paket kullanılmıştır. Bu çalışmada ESC-10 ve ESC-50 veri setleri için en iyi sonuç AECNet (Acoustic Event Classification Net) sınıflandırıcı mimarisi ile elde edilmiş ve 84.9% ve 68.6% oranlarında başarı elde edilmiştir. Ayrıca DCASE veri seti için de yine en iyi sonuç AecNet ile 86.5% olarak kaydedilmiştir.

Bavu vd. tarafından 2019'da yapılan çalışmada TimeScaleNet YSA modeli ile bir ses sınıflandırma işlemi gerçekleştirilmiş [41]. TimeScaleNet, hem örnek düzeyinde hem de çerçeve düzeyinde zaman bağımlılıklarını öğrenerek bir sesin verimli bir temsilini öğrenmeyi amaçlar. Önerilen yaklaşım, gelişmiş derin öğrenme ve sinyal işleme tekniklerini birleştirerek öğrenme şemasının yorumlanabilirliğinin geliştirilmesine olanak tanımaktadır. Önerilen bu yöntem ile *ses komutları* veri seti üzerinde 94.87% ve ESC-10 veri seti üzerinde ise 85.2% oranında bir başarı elde edilmiştir.

Zhang vd. 2019'da yaptıkları çalışmada, özel bir ESA mimarisi için log-gammatone spektrogramlar girdi olarak kullanılmıştır [42]. Bu çalışmada 44.1 kHz frekansı ile

örneklenen sesler, öncelikle 1024 örneklemeli pencerelere bölünmüş ve *hamming window* kullanılarak Short Time Fourier Transform (STFT)'leri alınmıştır. Daha sonra bu örneklere 128-band *gammatone filter bank* uygulanmış ve $128 \times 128 \times 2$ büyüklüğünde 3-boytlu bir özellik gösterimi elde edilmiştir. Sesleri sınıflandırmak için ise, 8 tane evrişimsel katman ve 2 tane çift yönlü GRU katmanından oluşan evrişimsel RNN modeli kullanılmıştır. Söz konusu çalışmada, ESC-10, ESC-50 ve DCASE2016 veri setleri kullanılmış ve sırası ile 94.2%, 86.5% ve 88.9% oranında sınıflandırma başarıları elde edilmiştir.

Makine öğrenmesinde son yıllarda özellikle transfer öğrenme yöntemleri sıklıkla kullanılmaktadır. Demir vd. 2020'de yaptıkları çalışmada sesleri sınıflandırmak için *pyramidal concatinated CNN* adlı farklı bir DYSA modeli kullanmışladır[35]. Seslerin özelliklerini elde etmek için gürültüden arındırılmış sinyaller STFT yöntemi ile spektrogramlara dönüştürülmüştür. Ardından, ses imajları özellik çıkarma işlemi için önceden eğitilmiş VGG19 ve DenseNet201 gibi önceden eğitilmiş modellere girdi olarak verilmiş. Bu çalışmada, özellik çıkarımı piramidal bir şekilde gerçekleştirilmiştir ve bu işlem özellik boyutunu oldukça büyütmektedir. Önerilen yöntemin son aşamasında ise SVM sınıflandırıcısı kullanılmıştır. Bu çalışmada, ESC-10, ESC-50 ve UrbanSound8K veri setleri kullanılmış ve sırası ile 94.8%, 81.4% ve 78.14% sınıflandırma başarıları elde edilmiştir.

Sharma vd. 2020'de yaptıkları çalışmada çevresel sesleri sınıflandırmak için ESA kullanmışlardır [43]. Seslerin özellik imajlarını elde etmek için MFCC [44], GFCC [45], CQT [46] ve Chromagram [47] tabanlı çoklu özellik kanalları kullanılmıştır. Sınıflandırma aşamasında ise derin ESA modeli kullanılmıştır. Bu çalışmada, UrbanSound8k veri seti için 97.52%, ESC-10 için 94.75% ve ESC-50 için 87.45% sınıflandırma başarıları elde edilmiştir.

Ahmad vd. 2020'de yaptıkları çalışmada çevresele sesleri sınıflandırmak için SVM yöntemini kullanmışlardır [48]. Bu amaçla seslerin özelliklerini elde etmek için her uzunluktaki ses sinyalinin azaltılmış homojen uzunluk dizisini veren OAS yöntemi

kullanılmıştır. Bu sinyaller daha sonra Empirical Mode Decomposition (EMD) [49] yöntemi ile band-sınırlı özeli bir moda dönüştürülmüştür. Sonuç olarak ESC-10 veri seti ile MC-LS-SVM sınıflandırıcı [50] ile 87.25% ve ELM sınıflandırıcı [51] ile ise 77.61% oranlarında başarı elde edilmiştir.

Chandrakala vd. 2020'de yaptıkları çalışmada çevresel sesleri sınıflandırmışlardır [52]. Çalışmada sınıflandırma işlemi için ISAGMMs tabanlı SVM yöntemi kullanılmış; DCASE2013 veri seti ile 90% ve DCASE2016 veri set ile ise 85.9% oranlarında başarı elde edilmiştir.

Esmaeilpour vd. 2020'de yaptıkları çalışmada seslerin özelliklerini elde etmek için Discrete Wavelet Transform (DWT) kullanılarak 2-boyutlu spektrogram imajları oluşturulmuştur. Sınıflandırma aşamasında ise GoogleNet üzerinde ConvNet mimarisi kullanılmıştır. Veri seti olarak ESC-10, ESC-50, UrbanSound8k ve DCASE-2017 kullanılmış ve sırası ile 78.26%, 80.52%, 91.02% ve 81.37% sınıflandırma başarıları elde edilmiştir [53].

Çevresel seslerin sınıflandırılması çalışmalarında öznitelik çıkartılması amacıyla istatistiğe dayalı birçok farklı yöntem de kullanılmıştır. Bunlardan biri de 2020'de Akbal vd. tarafından yapılan ve 3 temel aşamadan oluşan öznitelik çıkarma yöntemidir [54]. Bu aşamalar sırasıyla özellik çıkarma, özellik oluşturma ve sınıflandırma aşamalarıdır. Özellik çıkartma aşamasında, tek boyutlu yerel ikili desen (1D-LBP) [55], tek boyutlu üçlü desen (1D-TP) [56] gibi istatistiksel yöntemler kullanılmıştır. Ardından özellik seçimi aşamasında, seslerin ayırt edici özelliklerini belirlemek için komşuluk bileşeni analizi (NCA) [57] yöntemi kullanılmıştır. Sınıflandırma aşamasında ise SVM'ler kullanılmıştır. Önerilen bu yöntemle ESC-10 veri seti üzerinde yapılan çalışmalarda 90.25% başarı oranı elde edilmiştir.

Ses sınıflandırmada daha başarılı sonuçlar elde etmek için veri çoğaltma yöntemleri de sıklıkla kullanılmaktadır. Bu alanda yapılan en kapsamlı çalışmalardan birinde Mushtaq vd. 2021'de spektrogram imajlarına dayalı transfer öğrenme yöntemi kullanarak veri

çoğaltmaya dayalı bir ses sınıflandırma gerçekleştirmişlerdir [58]. Bu çalışmada, sesler öncelikle spektrogram imajlarına dönüştürülmüş, ardından sınıflandırma işlemi gerçekleştirilmiştir. Bu işlemler sırasında zaman boyutunda ve imajlar üzerinde olmak üzere iki farklı veri çoğaltma işlemi gerçekleştirilmiştir. Yapılan deneylerde, ResNet-152 modeli kullanılarak UrbanSound8k veri seti üzerinde en iyi sonuç 99.49% ve ESC-10 üzerinde ise 99.04% olarak elde edilmiştir. ESC-50 veri seti ile yapılan testlerde ise en iyi sonuç sDenseNet-161 modeli ile 97.57% olarak elde edilmiştir.

Kentsel ses olaylarının kompakt ve verimli bir tanımını sunan tanımlayıcılar önermek amacıyla Luz vd. 2021'de ESA kullanarak kentsel seslerin sınıflandırılması üzerinde çalışmışlardır [36]. Söz konusu çalışmada, ses sinyalleri üzerinde 50% örtüşme ile *Blackman-Harris pencereleme fonksiyonu* kullanarak mel-spektrogramlar elde edilmiştir. Sınıflandırma aşamasında ise hem el yapımı hem de derin öğrenme ile elde edilmiş özellikler kullanılmış ve UrbanSouund8K veri seti üzerinde 96.16% oranında bir sınıflandırma başarısı başarı elde edilmiştir.

Çevresel seslerin sınıflandırma performansının bir sesten prototipik özellikleri çıkarmak için kullanılan tekniğine ne kadar bağlı olduğunu araştırmak amacıyla Tripathi ve Mishra 2021'de ESC'lerdeki sınıf içi tutarsızlık problemlerini çözen bir vurgulama modülü geliştirmiş ve bu modeli ResNet ile kullanarak test etmişlerdir [59]. Sonuç olarak ESC-10 veri seti ile 88.70% ve DCASE 2019 veri seti ile 82.39% doğrulukta sınıflandırma başarıları elde edilmiştir.

Son yıllarda derin öğrenmede kullanılan bir diğer sınıflandırma yöntemi de *Semi-Supervised Learning* (SSL) metodudur. Bu yöntemi kullanarak Cances ve Pellegrini 2021'de çok sınırlı sayıda etiket kullanılan veri setleri üzerinde sınıflandırma çalışması gerçekleştirmişlerdir [60]. Çalışmada farklı boyutlarda ve çok farklı kaynaklardan gelen sesleri içeren üç veri seti üzerinde deneyler yapılmıştır. Google Speech Commands veri seti için 50% etiketlenmiş veri ile 94.17% ve 100% etiketlenmiş data ile 95.58%, UrbanSound8k veri seti için 50% etiketlenmiş veri ile 73.66% ve 100% etiketlenmiş veri

ile 74.44% ve ESC-10 veri seti için ise 50% etiketlenmiş veri ile 88.28% ve 100% etiketlenmiş veri ile 91.72% başarı oranları elde edilmiştir.

Tripathi ve Mishra 2021'de yaptıkları diğer bir çalışmada ise yine çevresel seslerin sınıflandırılması üzerinde durmuşlardır [61]. Sesler, 44.1 kHz'de örneklenmiş ve spektrogram imajları çıkartılmıştır. Sınıflandırma aşamasında ise SSL tabanlı bir derin sınıflayıcı kullanılmıştır. Çalışmada ESC-10 ve DCASE 2019 veri setleri kullanılmış ve sırası ile 91.67% ve 75.09% başarı oranları elde edilmiştir.

Görünürlük Grafları (GG) kullanılarak farklı alanlarda sınıflandırma çalışmaları yapılmıştır. Zhu vd. tarafından 2012'de GG kullanarak uyku seviyelerinin belirlenmesi amacıyla bir çalışma yapmışlardır [62]. Söz konusu çalışmada, ham olarak kaydedilen EEG ve EOG sinyallerinin özniteliklerini elde etmek için Visibility Graph Similarity (VGS) yöntemi kullanılmıştır [63]. Sleep-EDF veri seti üzerinde 7-seviye için yapılan sınıflandırmada VGS+LIBSVM kullanılarak 10-fold çapraz doğrulama uygulanmış ve sonuç olarak 81.11% oranında bir başarı elde edilmiştir.

Zaman serileri üzerinde karmaşık ağların kullanılması ile ilgili önemli çalışmalardan birini de Demir ve Türker 2021'de biyomedikal sinyal işleme alanında karmaşık ağ yaklaşımını kullanarak gerçekleştirmiştir [64]. Bu çalışmada, Zyma vd. tarafından PhysioBank [65] veri tabanı temel alınarak geliştirilen yeni bir EEG veri seti kullanılmıştır [66]. Gerçekleştirilen çalışmada, beynin EEG sinyalleri kullanılarak insanların dinlenme/çalışma ve aritmetik başarı seviyelerinin cinsiyete dayalı ayırt edici özellikleri incelenmiştir. Karmaşık ağlara dayalı olarak gerçekleştirilen teorik çalışma sonucunda zeki bireylerin beyinlerindeki bağlantıların dinlenme durumunda daha yoğun olduğu sonucuna varılmıştır. Bunun yanında, dinlenme durumunda bağlantı gücü ve verimliliği daha düşük olan erkek beyninin, zihinsel iş yükü altında bağlantıyı artırma yeteneği sergilediği gözlemlenmiştir.

Zaman serisi verilerin yanı sıra metin tabanlı verilerde de karmaşık ağ yaklaşımı oldukça verimli bir şekilde kullanılmaktadır. Türker vd. 2016'da gerçekleştirdikleri çalışmada

kitap ve sosyal medya metinlerini analiz etmek için ağ yaklaşımını kullanmışlardır [67]. Çalışmada kitap ve sosyal medya metinlerindeki farklılıkları dilsel tipoloji açısından araştırmak için kitap, Facebook ve Twitter metinlerinin hem sıralı hem de cümle sıralama ağları yönlendirilmemiş ve ağırlıklı kenarlarla oluşturulmuştur. Karşılaştırmalar yapılırken, ortalama derece, modülerlik, ortalama kümeleme katsayısı, ortalama yol uzunluğu, çap, ortalama bağlantı ağırlığı gibi temel parametreler kullanılmıştır. Ayrıca, eşleştirme düğümlerinin düğüm dereceleri, kenar ağırlıkları ve maksimum derece farklılıkları için dağılım grafları incelenmiştir. Çalışma sonucunda, dilsel tipolojinin, kitaplarda daha resmi bir kullanıma sahip olduğu, Twitetr'da ise daha çok resmi olmayan bir kullanıma saptığı bunun da karakter sınırlamasından kaynaklandığı tespit edilmiştir. Ayrıca, Facebook'un ağ parametreleri aracılığıyla bu medyalar arasında enterpolasyon yaptığı gözlemlenmiştir.

Türker ve Sulak 2018'de gerçekleştirdikleri çalışmada sosyal ağlarda karmaşık ağ yaklaşımını kullanmışlardır [68]. Bu çalışmada, Twitter içeriklerinden elde edilen büyük miktardaki veri seti üzerindeki etiketlerden oluşan ağın iki katmanlı bir analizi gerçekleştirilmiştir. Sonuçta; "*small world*" [69] yapısına ait kümelenme ve çeşitli ağ parametrelerindeki güç yasası dağılımları gibi gerçek ağların evrensel özelliklerinin, çok katmanlı hashtag ağlarında da belirgin olduğu gözlemlenmiştir. Ayrıca, tweetlerdeki hashtaglerin birlikte oluşumlarının çoğunlukla semantik ilişkilerle birleştiği belirlenmiştir.

Newman'ın yapmış olduğu bilimsel işbirliği veri tabanlarına dayalı çalışmalar da karmaşık ağların dinamiklerini açıklamada önemli katkılar sağlamıştır [70].

Yukarda bahsedilen çalışmalara ek olarak zaman serileri üzerinde Görünürlük Grafları yöntemleri kullanılarak enerji yayılım oranları [71], finansal veriler [72], fizyolojik zaman serileri [73], EEG sinyalleri ile hastalık nöbetlerinin tespiti [74], kardiyolojik ve solunum etkileşim sinyalleri [75], EEG sinyalleri ile alkolizm teşhisi [76] gibi alanlarda da çalışmalar yapılmıştır [77].

## 1.2. BU TEZ ÇALIŞMASININ LİTERATÜRE KATKILARI

Bu tez çalışması kapsamında gerçekleştirilen Connectogram tabanlı ses sınıflandırma alanında literatüre sunulan makale SCI-Expanded kapsamındaki Q2 kalite sınıfında olan hakemli bir yayınlanmıştır.

Makale; 7 Eylül 2021'de dergiye gönderilmiş, 27 Ocak 2022'de kabul edilmiş ve 15 Şubat 2022'de online olarak erişime açılmıştır.

Türker, İ. and Aksu, S., "Connectogram – A graph-based time dependent representation for sounds", Applied Acoustics, 191: 108660 (2022).

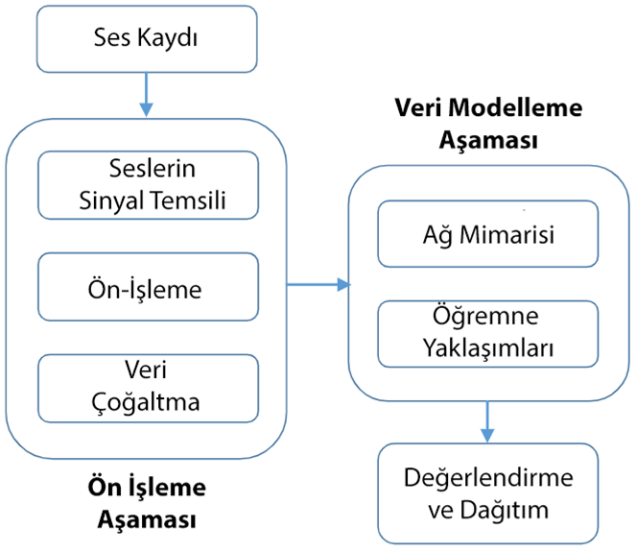
DOI Numarası: https://doi.org/10.1016/j.apacoust.2022.108660

# BÖLÜM 2

# SESLERİN ÖZNİTELİKLERİNİN ÇIKARTILMASI

Ses sinyalleri üzerinde özellik çıkartma işlemi, ses işleme araştırma ve geliştirmesinin temel taşlarından biridir. Ses özellikleri, bir ses sinyalinden çıkarılabilen ve o sesi diğerlerinden ayırt etmeyi sağlayan bağlamsal bilgilerdir. Bu bölümde, seslerin makine öğrenmesi yöntemleri ile sınıflandırılmasından önce özniteliklerini belirlenmesi için gerçekleştirilen farklı özellik çıkartma yöntemler ele alınacaktır.

Ses sinyali işleme algoritmaları genellikle sinyalin analizini, özelliklerini çıkarmayı, davranışını tahmin etmeyi, sinyalde herhangi bir model olup olmadığını belirlemeyi ve belirli bir sinyalin diğer benzer sinyallerle nasıl ilişkilendirildiğini içerir [78].




Şekil 2.1. Akustik sınıflandırma algoritmasının akış diyagramı.

Şekil 2.1'de tipik bir derin öğrenme tabanlı akustik sınıflandırma algoritmasının adımları gösterilmiştir [79]. Burada da görüldüğü gibi sınıflandırma işleminden önce ses sinyalleri üzerinde gürültüden arındırma, çerçevelere ayırma, veri çoğaltma gibi bir takım ön işlemlerin yapılması ve ayırt edici özelliklerin çıkartılması gerekmektedir.

İnsanlar fazladan bir çaba harcamadan çeşitli sesler arasında kolayca sınıflandırma yapabilirler; konuşma ve müzik, araba ve kamyon sesleri, bebek ve yetişkin konuşma kalitesi, çeşitli hoparlörler, gürültü ve kullanışlı sesler gibi sesleri ayırt edebilirler. Fakat Daha detaylı sınıflandırma için özellik çıkarmaya dayalı bir makine öğrenmesi algoritması kullanılması gerekmektedir [80].


Şekil 2.2. Sınıflandırmada kullanılan ses çeşitleri.

Herhangi bir özel amaç ne olursa olsun, bir makine öğrenme sistemi, bir makinenin doğru ve hızlı bir şekilde öğrenmesine yardımcı olan sağlam ve ayrıştırıcı özelliklerin verilmesine dayanır. Normalde tüm veri kümesi, özelliklerini öğrenmek için eğitim sırasında makineye bütün veri ham olarak değil, bunun yerine sinyallerin boyut olarak küçültülmüş bir temsili verilir. Şekil 2.2'de görüldüğü gibi ses alanındaki çalışmalar başlıca konuşma tanıma, müzik sesleri ve çevresel seslerin sınıflandırılması gibi üç temel alanda incelenebilir [78].

Literatüre bakıldığında, ses sinyallerinin özelliklerini elde etmek için zamansal alan, frekans alanı, cepstral alanı, dalgacık alanı ve zaman-frekans alanı gibi farklı yöntemler kullanıldığı görülmektedir [78]. Aşağıda kullanılan bu yöntemler üzerinde durulmuş,

ayrıca bu tez çalışması kapsamında özellik çıkartma işlemi için kullanılan *Görünürlük Grafları* (GG) detaylı olarak açıklanmaya çalışılmıştır.

## 2.1. SESİN ZAMAN ALANINDAKİ FİZİKSEL ÖZELLİKLERİNİN ÇIKARTILMASI

Bu yöntem, başka herhangi bir dönüşüm işlemi gerçekleştirmeden hesaplamaların doğrudan ses sinyali üzerinde gerçekleştirilmesine dayanır. Ses öznitelik çıkarımına yönelik bu yaklaşım, en temel ve klasik olanlardan birini oluşturur. Bu konu ile ilgili önemli çalışmalardan birini Mitrovi´ vd. gerçekleştirmiştir [81].

Sesin zaman alanındaki fiziksel özellikleri zero crossing tabanlı, genlik tabanlı, güç tabanlı ve ritim tabanlı olmak üzere farklı kategorilerde incelenebilir [82].

### 2.1.1. Short-Time Enerji Fonksiyonu

Çerçeve tabanlı bir yöntem kullanılarak elde edilen STE, sinyal çerçevesi başına ortalama enerji olarak tanımlanır. Bu yöntem, MPEG-7 ses sinyalinin gücünün tanımlayıcısında da kullanılmaktadır. Ayrıca sesin spektral alanının gücünü hesaplamak için başka STE tanımlamaları da bulunmaktadır [55]. STE bir ses sinyalindeki sessiz bölgelerden sesli bölgelere geçiş noktalarını belirlemek ve sessiz bölümlerle sesli bölümleri ayırmak için kullanılmaktadır [56].

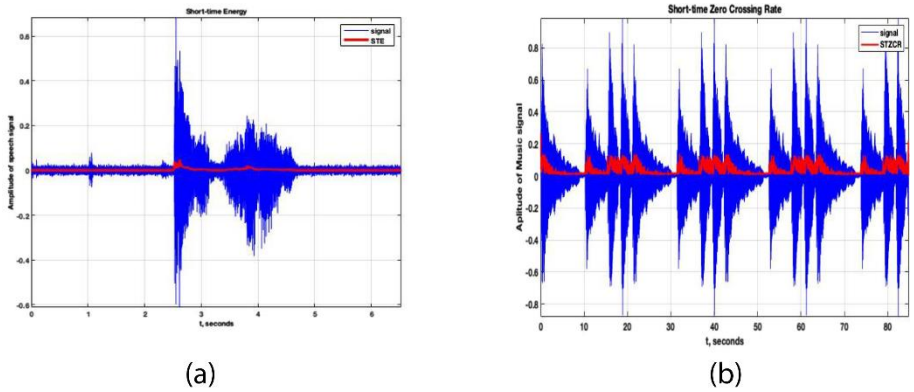
Ses sinyalinin küçük bir aralıkta nispeten yavaş değiştiğini varsayarsak $E_n$ enerji fonksiyonu Eşitlik 2.1'deki gibi hesaplanır.

$$E_n = \frac{1}{N}\sum_{m}[x(m)w(n-m)]^2 \tag{2.1}$$

$x(m)$ : Ayrık zamanlı ses sinyali

$n$ : Short-time enerjinin zaman indeksi

$w(m)$ : $N$ uzunluğu için dikdörtgensel pencere




(a) (b)

Şekil 2.3. İki farklı ses sinyalinin STE grafikleri a) konuşma b) müzik parçası.

Şekil 2.3'de biri konuşma ve diğeri müzik parçası olmak üzere iki farklı ses sinyaline ait STE grafikleri gösterilmiştir [78].

### 2.1.2. Zero-Crossing Rate (ZCR)

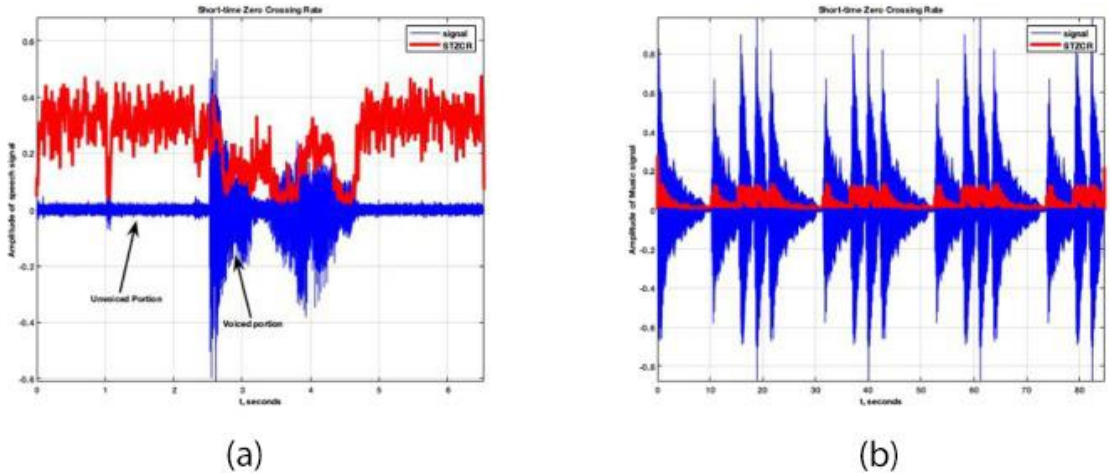
ZCR bir ses sinyali çerçevesinde konuşma olup olmadığının tespit edilmesinde veya bu çerçevenin sessiz bölgelerinin algılanmasında önemli bilgiler vermektedir. ZCR, konuşmanın olduğu kısmına kıyasla olmadığı kısımları için daha yüksektir [78].

Ses sinyalinin sıfırdan geçiş sayısı ZCR olarak bilinmektedir. Ayrık zaman sinyalinde ardışık örneklerin farklı işaretleri varsa bu işaretler arasında sıfır-geçişi olacaktır. Sıfır geçişlerin meydana gelme hızı, bir sinyalin frekans içeriğinin basit bir ölçüsüdür. Short-time ortalaması ve ZCR değerleri Eşitlik 2.2'ye göre hesaplanır [84].

$$Z_n = \frac{1}{2}\sum_m |sgn[x(m)] - sgn[x(m-1)]|w[x(n-m)] \qquad (2.2\ )$$

$$sgn[x(n)] = \begin{cases} 1, & x(n) \geq 0 \\ -1, & x(n) < 0 \end{cases}$$

Burada $w(n)$ pencere fonksiyonunu ifade etmektedir.




Şekil 2.4. İki farklı ses sinyalinin ZCR grafikleri a) konuşma, b) müzik.

Şekil 2.4'de konuşma ve müzik olmak üzere iki farklı ses sinyalinin grafikleri karşılaştırmalı olarak gösterilmiştir [78]. Şekilde seslendirilmemiş bölümler için ZCR'nin sesli bölümlere göre çok yüksek olduğu görülmektedir.

ZCR, aynı zamanda konuşmanın temel frekansını (FF: Fundamental Frequency) tahmin etmek için oldukça elverişli bir tekniktir [85]. ZCR, sinyalin frekansının iki katına eşittir. Dolayısıyla ZCR'nin sinyalin frekansı hakkında dolaylı bilgi verdiğini söylemek mümkündür. Bu özelliğinden dolayı ZCR ayırıcı ve sınıflandırıcı tasarlamak için kullanılabilmektedir [86].

### 2.1.3. Root Mean Square (RMS)

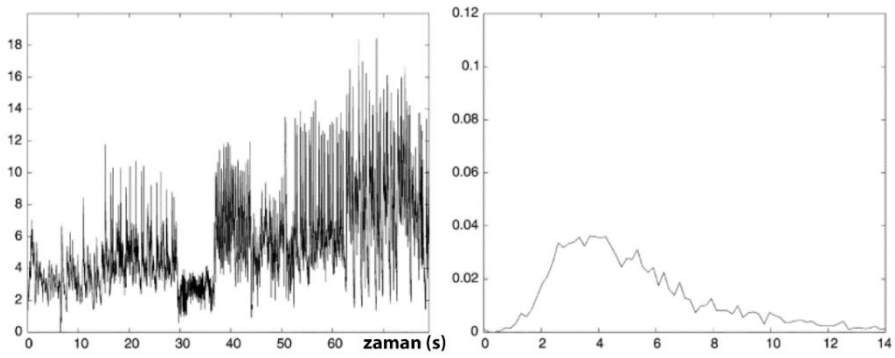
Bir çerçevedeki dalga formunun bütünlüğü olarak tanımlanan RMS, bir büyüklüğün zaman içindeki ortalama değişimini göstermek için kullanılmaktadır. Bir ses dosyasında tüm sinyaller pozitif ve negatif bileşenlerden oluşmaktadır. Bu özellik RMS ölçümünü ve hesaplamayı daha geçerli kılmaktadır.

$$RMS \triangleq \sqrt{\sum_{n=1}^{N} x^2(n)} \tag{2.3}$$

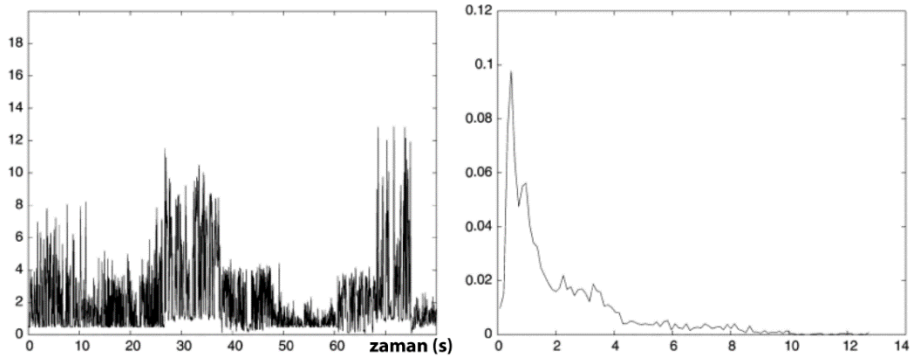
$$ZC \triangleq \frac{1}{2}\sum_{n=2}^{N} \left|sign\big(x(n)\big) - sign(x(n-1))\right| \tag{2.4}$$

$$sign(a) = f(x) = \begin{cases} 1, & a > 0 \\ 0, & a = 0 \\ -1, & \alpha < 0 \end{cases}$$

Sinyal genlikleri, RMS ve ZC değerleri Eşitlik 2.3 ve Eşitlik 2.4'e göre hesaplanır [87]. Böylece hesaplamayı basitleştirmek için, dikkate alınan aralığın tüm örneklerinin ortalaması, herhangi bir veri kaybı olmadan alınmış olur.




Şekil 2.5. Bir müzik dosyasının RMS ve histogram dağılımı.

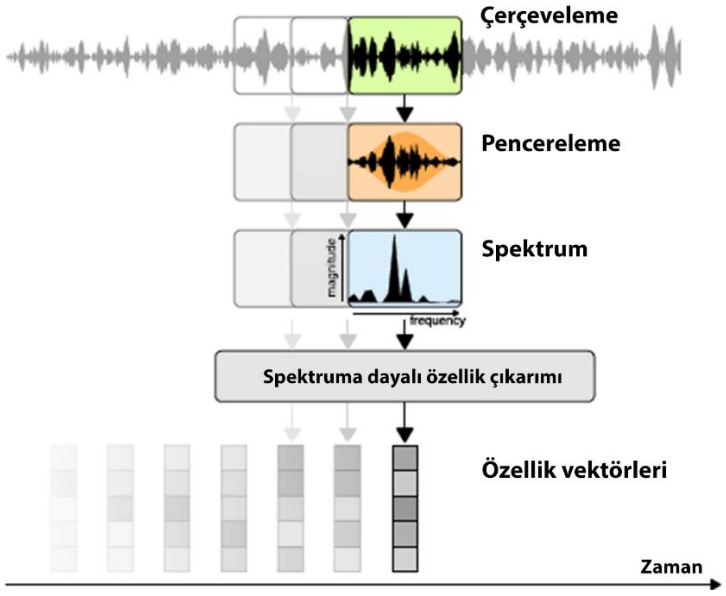



Şekil 2.6. Bir konuşma dosyasının RMS ve histogram dağılımı.

Şekil 2.5 ve Şekil 2.6 ayrı ayrı müzik ve ses dosyalarının RMS ve histogram dağılımlarını göstermektedir. RMS dağılımı ile bir ses dosyasındaki konuşma ve müzik, genlik değerlerinin dağılımı ile ayırt edilebilmektedir. Müzik ve konuşma sinyallerinin genlik değerlerinin dağılımları farklı olduğundan bu özellik hem segmentasyon hem de sınıflandırma için kullanılabilmektedir [87].

## 2.2. SESİN FREKANS ALANINDAKİ FİZİKSEL ÖZELLİKLERİNİN ÇIKARTILMASI

Frekans alanında sesin fiziksel özelliklerinin çıkartılması literatürde önemli bir yer kaplamaktadır. Frekans tabanlı özellikler, genellikle Kısa-Zamanlı Fourier Dönüşümü (STFT: Short-Time Fourer Transform) ile veya Autoregression analizi ile elde edilir.




Şekil 2.7. Ses işleme adımlarının akış diyagramı.

Ayrık Fourier Dönüşümü hesaplanarak her kısa zaman çerçevesi sinyalinin frekans alanı gösterimi elde edilir. Frekans alanındaki ses özelliği çıkarma aşamaları Şekil 2.7'da görselleştirilmiştir [88].

### 2.2.1. Ön-Vurgulama

Sayısal sinyallerde yüksek frekanslar düşük frekanslara göre daha anlamlı bilgiler içermektedir. Frekansların elde edilmesinden önce yüksek frekansları daha belirgin hale getirmek amacıyla sinyal üzerinde bir ön-vurgulama işlemi gerçekleştirilir. Bu işlem, yüksek frekansların düşük frekanslara oranla daha düşük boyutlu bir güce sahip olmasından kaynaklanan dengesizliği gidermek için yüksek frekans bantlarının genliğini artırıp daha düşük frekanslı bantların genliğini azaltmaktadır. Bu işlem sayesinde Fourier dönüşümü sırasındaki sayısal problemlerden kaçınmak ve Signal-to-Noise Ratio (SNR) değerini yükseltmek mümkün olmaktadır.

$$y(k,t) = x(t) - \alpha x(t-1) \tag{2.5}$$

Bir x sinyali üzerinde ön-vurgulama işlemi, α filtre katsayısı olmak üzere Eşitlik 2.5'e göre yapılmaktadır.

### 2.2.2. Çerçeveleme ve Pencereleme

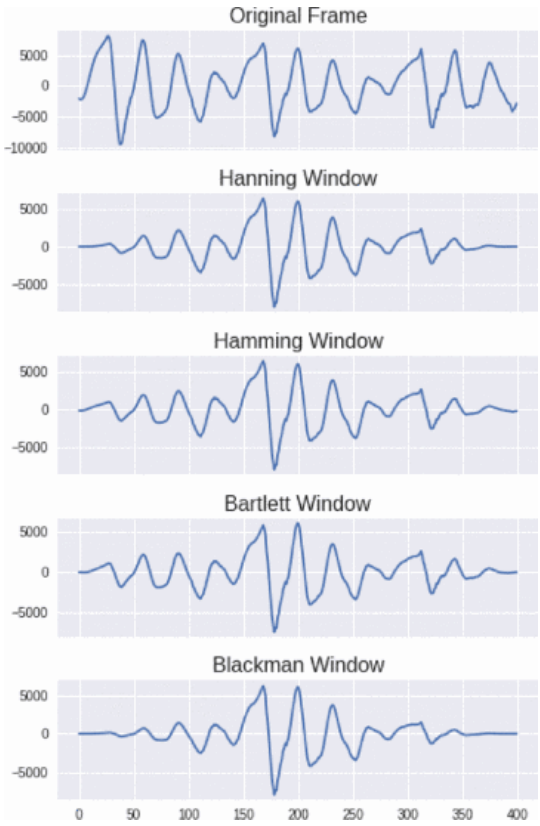
Bir sinyaldeki frekanslar ve her bir frekans bandının gücü zaman boyunca sabit olmayıp sürekli değişmektedir. Sinyalin gücündeki bu değişim sinyalin tamamı üzerinde bir Fourier dönüşümü ve otokorelasyon işlemini imkânsız hale getirmektedir. Sürekli değişen frekans konturlarını daha hassas bir şekilde elde etmek ve daha anlamlı bilgiler çıkartabilmek amacıyla sinyallerin küçük dilimler halinde değerlendirilmesi gerekmektedir. Bu amaçla çerçeveleme işlemi gerçekleştirilir ve Fourier dönüşümü bu çerçeveler üzerinde uygulanır.

Çerçeve uzunluğu genellikle 10-30 ms arasında seçilmektedir. Çünkü bu aralıkta alınan ses sinyalleri genellikle sabit bir akustik yapıya sahiptir. Sinyal bölündükten sonra komşu iki çerçeve arasında örtüşme uygulanır. Örtüşme oranı, 30% ile 75% arasında değişmekle beraber ses sinyallerinin işlenmesinde bu oran genellikle pencere uzunluğunun yarısı olarak (50%) seçilmektedir [88].

Frekansı ve içeriği dinamik olarak değişen ses verilerinden frekanslara ait sağlam özellikler çıkarmak ve nispeten statik ses klipleri elde etmek için çerçevelere ayrılmış ses dalga biçimlerine bir takım pencereleme işlemi uygulanmalıdır. Bu amaçla rectangular window hamming window, hanning window ve Blackman window gibi farklı pencereleme fonksiyonları bulunmaktadır. Bu konuda yapılan çalışmalar hanning pencerelerin diğer pencerelere göre daha iyi sonuç verdiğini göstermiştir [89].

Şekil 2.8'de orijinal bir ses sinyaline farklı pencereleme fonksiyonlarının uygulanması ve sonuçta elde edilen yeni sinyal formları gösterilmiştir [90]. Örnek bir hanning pencereleme fonksiyonun sinyale uygulanması Eşitlik 2.6'da gösterilmiştir.

$$\mathrm{w(n)} = 0{,}5 - 0.46\cos\left(\frac{2\pi \mathrm{n}}{\mathrm{N}-1}\right) \qquad 0 \leq n \leq N-1 \tag{2.6}$$




Şekil 2.8. Farklı pencereleme fonksiyonlarının bir ses sinyale uygulanması.

Genel anlamda, fiziksel frekans alanı özellikleri, sinyal frekans içeriğinin fiziksel özelliklerini tanımlar. Frekans tabanlı özellikler aşağıdaki gibi ayrıştırılabilir [82]:

- Autoregression-tabanlı
- STFT-tabanlı
- Brightness-tabanlı
- Tonality-tabanlı
- Chroma-tabanlı
- Spectrum-tabanlı özellikler

### 2.2.3. Autoregression-Tabanlı Frekans Özellikleri

Autoregression tabanlı özellikler tipik bir sinyalin lineer tahmin analizinden elde edilmektedir. Bu tahmin analizi, genellikle ses sinyallerindeki spektral üstünlük noktalarını elde etmek için kullanılmaktadır.

Autoregressif (AR) spektral tahmin, kısa veri segmentlerinde daha kararlı spektrumlar üretir ve böylece AR yaklaşımı periyodikliği varsaymadığı için FFT tekniklere bir alternatif sunmaktadır. AR parametrelerinin kestirimi lineer denklemler çözülerek kolaylıkla yapılabildiğinden AR yöntem en sık kullanılan parametrik yöntemdir [91].

AR modeline göre $x_{norm}(t)$'in şimdiki değeri, bir önceki $x_{norm}(t-1)$ değeri ve beyaz gürültü girişinin lineer kombinasyonu olarak hesaplanır. Böylece $p$'inci dereceden AR modeli Eşitlik 2.7 ile göre hesaplanır [91].

$$x_{norm}(t) = -\sum_{k=1}^{p} a_k x_{norm}(t-k) + e \tag{2.7}$$

Bu denklemde, $a_k$ AR katsayılarını göstermektedir. Hanning penceresi ile Ardışık 256 tane $x_{norm}(t)$ serisi kullanıldığında, PSD tahmini Eşitlik 2.8 ile elde edilebilir.

$$P_{AR}(f) = \frac{T}{\left|1 + \sum_{k=1}^{p} a_k e^{-j2\pi f k T}\right|^2} \tag{2.8}$$

Bu denklemde, $\sigma_w^2$ giren gürültü işaretinin varyansını, $T$ ise örnekleme periyodunu göstermektedir. Sonuç olarak AR-PSD tahmini $a_k$ ve $\sigma_w^2$'den hesaplanabilir. Sinyallerin yaygın olarak kullanılan autoregression özellikleri aşağıda açıklanmıştır [58].

- **Linear Predictive Coefficients (LPC):** LPC, bir ses sinyalinde formatlar veya spektral rezonanslardan oluşan spektral zarfları (SZ) yakalamak için kullanılır. Bu özelliğinden dolayı LPC, genellikle konuşma kodlama, tanıma, ses bölümlendirme ve kaybolan sesleri kurtarma gibi uygulamalarda kullanılmaktadır [59].
- **Line Spectral Frequencies (LSF):** LSF'ler, LPC parametrelerinin kuantalama ve interpolasyon amacıyla kullanılan güçlü bir temsilidir. LSF'ler, LPC polinomal temsili oluşturan palindromik ve antipalindromik polinomların kök fazları olarak hesaplanabilir. Bu metotlar özellikle konuşmacıların ayrıştırılması ve müzik enstrümanlarının tanınması problemlerinde yaygın olarak kullanılmaktadır [60].
- **Code Excited Linear Prediction Features (CELP):** Schroeder ve Atal tarafından ortaya konan ve konuşma kodlama standartları için kullanılan en önemli özelliklerden biridir [61]. Bu özellikler, LSP'lerin yanında sinyalin pitch ve öngörü residueleri ile ilgili iki farklı kod-bankası katsayılarını da içerir. CELP özellikleri, ayrıca ortam seslerinin tanınması içinde yaygın olarak kullanılmaktadır [62].

### 2.2.4. Kısa Zamanlı Fourier Dönüşümü (STFT: Short Time Fourier Transform)

Fourier dönüşümü ve ayrık sinyal versiyonları, bir sinyali zaman alanından frekans alanına dönüştürmek için kullanılır. Frekans içeriği zamanla değişen sinyalleri bazı ek gereksiz bilgiler sunmasına rağmen zaman-frekans alanında analiz etmek daha uygundur [96]. Zaman frekans gösterimlerinin (TFR) temel bazı özellikleri sinyale bağlı olma biçimleri ile ilgilidir. Bu bağımlılık doğrusal, ikinci dereceden veya doğrusal olmayan

olabilir şeklinde olabilir [97]. Kısa Zamanlı Fourier Dönüşümü (STFT) ve dalgacıklar gibi tüm doğrusal TFR'ler süperpozisyon veya doğrusallık ilkesini yerine getirmektedir [96].

STFT, görüntü işleme, ses tanıma, mühendislik, biyoloji ve top gibi birçok alanda kullanılmaktadır. STFT, bir pencere işlevi kullanarak daha uzun olan bir sinyalden kısa bir sinyal periyodu alır ve ardından bu kısa periyoda Fourier dönüşümü uygulayarak sesin frekans özelliklerinin çıkartılmasını sağlar [98]. Bu süreç orijinal sinyalin başından sonuna kadar tekrarlanır.

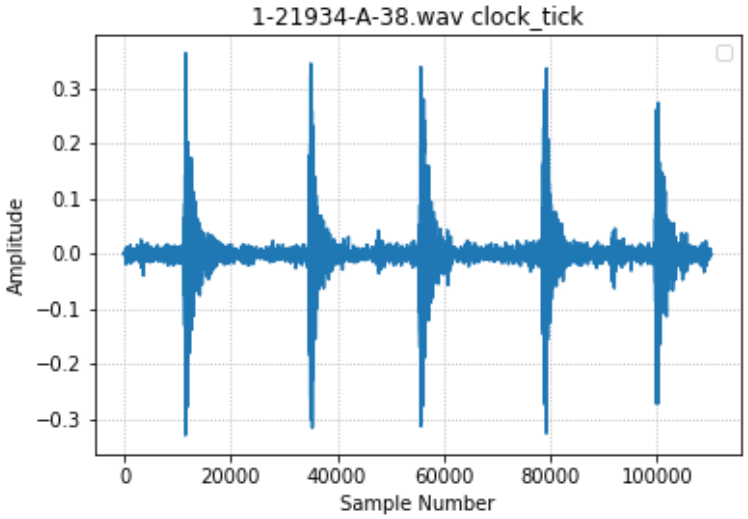
Eşitlik 2.6 ile elde edilen pencereler üzerinde Eşitlik 2.9 ile $STFT$'ler elde edilir.

$$STFT\{x[n]\}(m,\omega) = X(m,\omega) = \sum_{n=0}^{N-1} x[n]w[n-m]e^{-i\omega n} \tag{2.9}$$

Bu denklemde $x[n]$ sayısal sinyali ve $w[n]$ ise Eşitlik 2.6 ile gösterilen pencereleme sinyalini ifade etmektedir. Bu denklemde kullanılan $n$ değişkeni ayrık, $w[n]$ ise süreklidir. Sinyal işleme uygulamalarında STFT, genellikle bilgisayar kullanılarak Hızlı Fourier Dönüşümü (FFT) yöntemi ile yapıldığından bütün değişkenler ayrıktır.

## 2.3. SPEKTROGRAM TABANLI AKUSTİK ÖZELLİKLER

Sinyallerin zaman-frekans alanında gösterilmesi ses sınıflandırma için birçok bakımdan oldukça önemli faydalar sağlamaktadır. Birincisi; zaman-frekans dönüşümü geriye dönüştürülebilme özelliğine sahiptir ve bu gösterim, ses verisinin bütün özelliklerini içermektedir. Daha da önemlisi sesin zaman-frekans gösterimi, genellikle ses sinyalinin birçok farklı karakteristikten oluşan ayırt edici özelliklerini içermektedir [99]. Böylece DYSA'lar, belirli spektro-zamansal şekillerden sesle ilgili ayırt edici verileri alabilir önemli düzeyde bir sağlamlıkla sınıflandırma gerçekleştirebilir.




Şekil 2.9. Ses sinyalinin zaman boyutunda gösterilmesi.

Görsel olarak seslerin özelliklerinin çıkartılması için Şekil 2.9'de gösterilen zaman alanındaki orjinal ses sinyalinin imaj alanında yeniden temsil edilmesidir. Spektrogram, pencerelere ayrılmış ses sinyalinin spektrumlarının zaman ekseninde diziliminden oluşmuş iki boyutlu bir temsildir. Ses sinyali, sonuç olarak karmaşık değerler veren spektrogram imajlarını oluşturmak için kullanılacak ayrık fourier dönüşümü (DFT) Eşitlik 2.10'e göre hesaplanır.

$$X(k,r) = \sum_{n=0}^{N-1} x(n)w(n)e^{\frac{-2\pi ikn}{N}}, \quad k = 0,1,\dots,N-1 \tag{2.10}$$

Burada; $N$: pencere uzunluğu, $x(n)$: sinyalin zaman alanındaki temsili, $X(k,r)$: $r$'inci çerçeve için $f(k) = kF_s/N$ frekansına uygun gelen $k$'inci harmonik, $F_s$: örnekleme frekansı, $w(n)$ pencere fonksiyonu.

Spektrogram değerleri DFT değerlerinin büyüklüğünün logaritmasından Eşitlik 2.11'e göre elde edilir.

$$S(k,r) = log|X(k,r)| \tag{2.11}$$

Tüm sinyaller için aynı zaman-frekans imaj çözünürlüğünü elde etmek için her se sinyali belli uzunlukta çerçevelere bölünür ve bu çerçeveler arasında 50% örtüşme uygulanır.

## 2.4. SMOOTHED SPEKTROGRAM

Smoothed spektrogram elde etmek için, örtüşmeyen ve hareketli ortalama hesaplama işlemi her bir çerçevedeki spektrogramın frekans bileşenleri üzerinde Eşitlik 2.12'ye göre gerçekleştirilir.

$$X_s(a,r) = \sum_{v=(a-1)W+1}^{aW} |X(k_v,r)| \ , \quad a = 1,2,\dots,N/2W \tag{2.12}$$

Burada $W$: hareketli ortalama penceresinin uzunluğunu ifade etmektedir.

## 2.5. MEL-SPEKTROGRAM

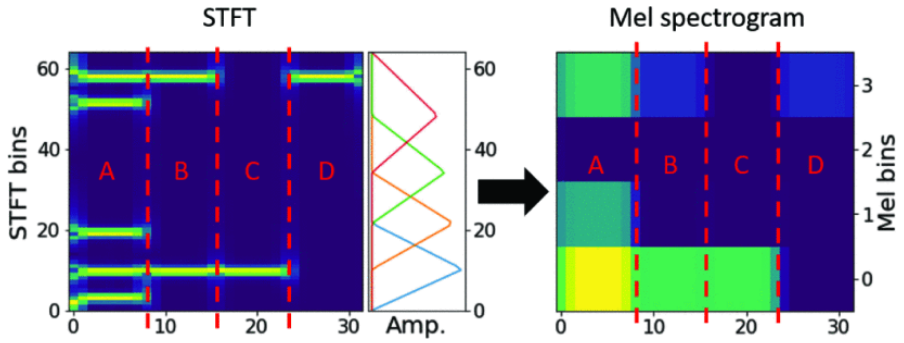
İnsan kulağı bütün frekanslara eşit derecede duyarlı değildir ve çözünürlüğü frekans ekseni boyunca değişir. Sesi insanın işitsel sisteminde olduğu gibi analiz etmek için frekansları doğrusal olmayan bir eksene dönüştürmek daha iyi bir çözüm olacaktır. Bu amaçla DFT'den tüm frekansları ayrı ayrı almak yerine belirli frekans bantlarında mevcut olan enerji veya enerji oranları gibi içeriği analiz etmek faydalı olabilir. Mel ölçeği, algısal davranışla eşleşen bir frekans ölçeğidir ve seslerin insan kulağı tarafından algılanabilir hale gelmesini sağlar [100].

Mel-spektrogram görüntü yoğunluğu değerleri, ayrık cosinus dönüşümleri (DCT) uygulanmadan filtre bankası enerjilerini kullanılarak MFCC'ye benzer şekilde hesaplanır. $m$'inci filtrenin mel-filtre bankası çıkışı Eşitlik 2.13'a göre hesaplanır [101].

$$E(m,r) = \sum_{k=0}^{\frac{N}{2}-1} V(m,k)\,|X(k,r)| \quad , \qquad m = 1,2,\dots,M \tag{2.13}$$

Bu eşitlikte; $E(m,r)$: $r$'inci çerçevedeki $m$'inci filtrenin filtre bankası çıkışı, $V(m,k)$: Mel ölçeğinde eşit eşit aralıklı dağıtılmış olan üçgensel filtre bankalarının normalize edilmiş filtre yanıtı, $M$: Mel-filtrelerinin toplam sayısı [102].

Mel-spektrogram için, düzleştirilmiş spektrogramda kullanılan spektrogram imajı hesaplanır ve mel filtrelerinin sayısı olan $M$, örneğin $32 \times 15$'lik bir imaj için 32 olarak ayarlanır.




Şekil 2.10. Mel filtre bankaları, bir STFT çıktısını mel-spektrogram'a dönüştürmek için kullanılabilir.

STFT'de bulunan k frekans bölmelerinin Mel bölmelerine eşlemek için Mel filtre bankalarının kullanılması ile elde edilen sonuç

Şekil 2.10'da gösterilmiştir [103]. Burada her bir Mel filtresi bankası birden fazla STFT'i kapsar ve bu bloğu tek bir Mel bloğuna indirger. Mel ölçeği $mel(f)$ ile frekans ölçeği $f$ arasındaki ilişki Eşitlik 2.14'de verilmiştir.

$$mel(f) = \frac{1000}{log(2)} log\ (1 + \frac{f}{1000}) \tag{2.14}$$

Burada, delta mesafesine sahip iki çift mel-frekans eşit uzaklıkta olarak algılanır. Mel-spektrogram, her banttaki spektrumun enerjisini hesaplayan bir dizi örtüşen üçgen filtre uygulanarak elde edilir. Burada filtre bant genişliği frekansla birlikte artmaktadır.

## 2.6. COCHLEGRAM (GAMMATONEGRAM)

Cochleagram gösteriminde, zaman-frekans görüntüsündeki frekans bileşenleri, insan kokleasının frekans seçicilik özelliğini temel alır ve Eşitlik 2.15'de verilen bir *gammatone* filtresi ile modellenir [104].

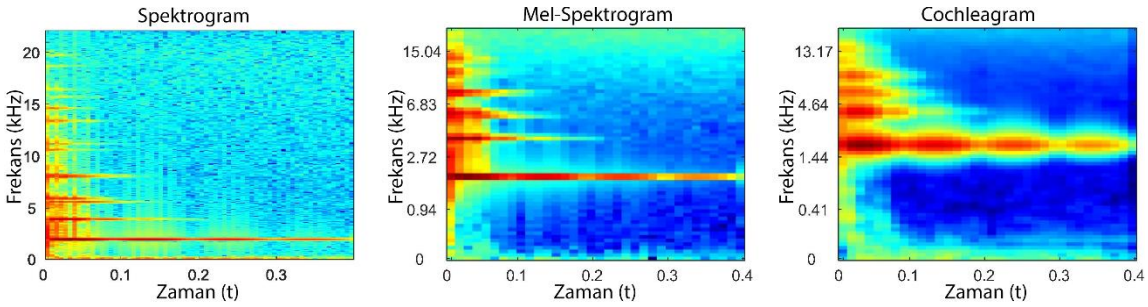
$$h(t) = At^{j-1}e^{-2\pi Bt}\cos(2\pi f_c t + \phi) \tag{2.15}$$

Bu eşitlikte; $A$: genlik, $j$: filtrenin sırası, $B$: filtrenin genişliği, $f_c$: filtrenin merkez frekansı, $\phi$: faz farkı, $t$: zaman değerlerini ifade etmektedir.

*Gammatone* filtresini kullanarak sinyali filtreledikten sonraki sinyallerin cochleagram temsili, Eşitlik 2.16'de gösterilen her bir frekans kanalı için pencerelenmiş sinyallere enerji ekledikten sonraki spektrogram temsiline benzemektedir [105,106].

$$C(g,r) = \sum_{n=0}^{N-1} |\hat{x}(g,n)| w(n), \qquad g = 1,2,\dots,G \tag{2.16}$$

Bu eşitlikte; $\hat{x}(g,n)$: gammatone filtrelenmiş sinyal, $C(g,r)$: $r$'inci çerçevede bulunan $f_{cg}$ merkez frekansına karşılık gelen $g$'inci harmonik, $G$: gammatone filtrenin numarası.




Şekil 2.11. 22.050 Hz'de alınmış örnek bir ses sinyalinin spektrogram, mel-spektrogram ve cochleagram imajları.

Şekil 2.11'de Örnek bir ses sinyali için spektrogram, mel-spektrogram ve cochleagram imaj görüntüleri verilmiştir [101]. Her üç gösterim de 0 Hz'den 22.050 Hz'e kadar Nyquist

frekansına sahiptir. Ancak, cochleagram gösterimde alt frekans aralığındaki frekans bileşenleri daha iyi çözünürlüğe sahip olduğu için daha fazla spektral bilginin elde edilmesini sağlar.

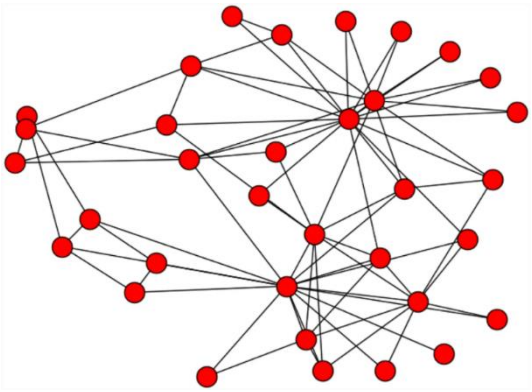
## 2.7. GÖRÜNÜRLÜK GRAFLARI (VISIBILITY GRAPHS)

### 2.7.1. Karmaşık Ağlar

Karmaşık Ağlar, geleneksel graf teorisinin çalışma alanları arasında bulunmaktadır ve gerçek ağların ampirik analizinden esinlenilerek geliştirilmiştir. Bir sistemi graf olarak modellemenin avantajı, graflarla problemlerin daha basit ve daha izlenebilir hale gelmesidir. Gerçekten de karmaşık ağlar, teknolojik ağlardan biyolojik ağlara kadar çeşitli gerçek sistemleri anlamamızı sağlamaktadır [107,108]. Örneğin; beyindeki nöron bağlantılarını oluşturan sinapslar, internet cihazları arasındaki bağlantılar, toplumsal ilişkiler, sosyal medya bağlantıları ve profesyonel iş birlikleri gibi pek çok bağlantı karmaşık ağların çalışma alanına girmektedir. [107,109,110].

Bireysel bileşenlerinden topluluk davranışlarını tahmin etmek mümkün olmadığından yukarıda sayılan sistemlere karmaşık sistemler denmektedir. Ancak bu sistemlerin matematiksel tanımını anlamak, onları tahmin etmemizi ve kontrol etmemizi sağlamaktadır. Kompleks sistemlerin davranışını analiz etmek, günlük hayatımızda önemli bir rol oynadığı için, günümüzün en büyük bilimsel çalışmalarından biri durumuna gelmiştir [108].

Şekil 2.12'de bir sosyal ağın en yaygın olarak kullanılan karmaşık ağ modeli ile temsili verilmiştir [108]. Bu örnekte, düğümler bir gruptaki bireyleri ve aralarındaki bağlantılar ise bu bireyler arasındaki etkileşimi temsil etmektedir.



Şekil 2.12. Bir gruptaki ilişkileri gösteren karmaşık ağ yapısı.

Mevcut hesaplama kapasitelerindeki son artış ve bilimin çeşitli alanlarında artan veri hacimleri ile karmaşık ağlar, karşılıklı etkileşim içindeki birimler arasında oluşan yapısal bağımlılıkları tanımlamak için ilginç ve çok yönlü bir araç haline gelmiştir. Klasik araştırma alanlarının yanında, ölçülen birimlerin fiziksel olarak açıkça tanımlanabildiği durumlarda karmaşık ağ teorisi başarılı sonuçların elde edilmesine yol açmıştır [77].

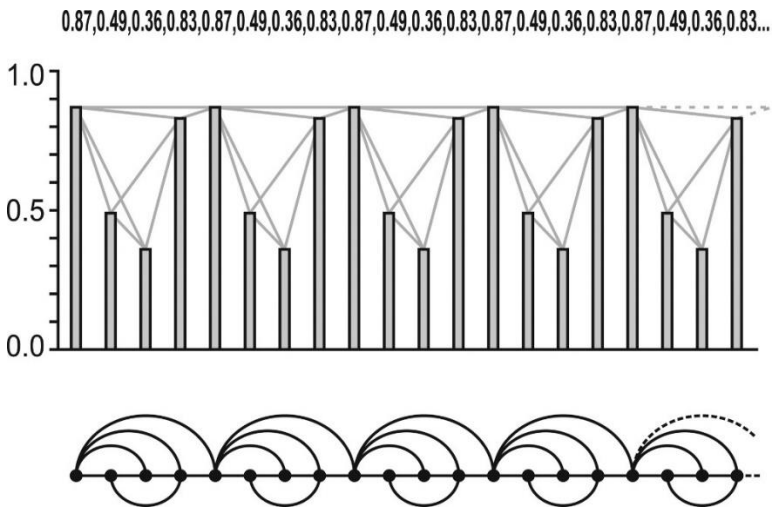
Karmaşık ağ teorisinin geleneksel olmayan uygulama alanlarından biri de fonksiyonel ağlardır. Bu fonksiyonel ağlar, dikkate alınan bağlantının fiziksel köşelerine ve kenarlarına atıfta bulunmamakta, ancak incelenen sistemin farklı bölümleri tarafından sergilenen dinamikleri arasındaki karşılıklı istatistiksel ilişkileri yansıtmaktadır [77].

Fonksiyonel ağlar ilk olarak, farklı beyin bölgelerindeki eşzamanlı nöronal aktivitenin genellikle bir dizi standartlaştırılmış EEG kanalı kullanılarak kaydedildiği nörobilimsel uygulamalarda kullanılmıştır. Bu veriler, belirli görevleri yerine getirirken farklı beyin bölgeleri arasındaki istatistiksel karşılıklı ilişkileri incelemek için kullanılabilir, en güçlü istatistiksel bağımlılıkların yansıttığı fonksiyonel bağlantıları içermektedir [111–113].

### 2.7.2. Zaman Serilerinin Görünürlük Grafları ile Temsili

Görünürlük Grafları (GG), zaman serilerini bir ağa modeli ile temsil etmek için kullanılan en elverişli yöntemlerden birisidir. GG ilk olarak Lacasa vd tarafından önerilmiştir [14].

Bu yöntemin ana fikri, zaman serilerini karakterize etmek için graf teorisi tekniklerini kullanmaktır.




Şekil 2.13. 20 veriden oluşan bir zaman serisi ve bu seriden görünürlük algoritması ile türetilen ilişkili görünürlük grafı örneği.

Şekil 2.13'de 20 veriden oluşan bir zaman serisi ve bu seriden görünürlük algoritması ile türetilen ilişkili görünürlük grafı örneği görülmektedir [14]. Bu grafta, her düğüm zaman serisindeki aynı sıraya karşılık gelir. Veriler arasındaki görünürlük çizgileri, graftaki düğümleri birbirine bağlayan bağlantıları tanımlamaktadır. GG yönsüzdür ve bir grafta bulunan her bir düğüm en azından en yakın komşu düğümle bağlantılıdır. Ayrıca, GG'nın bağlantısı düğümlerin görünürlüğü ile tanımlandığından bu graflar, zaman serisi üzerinde yatay ve dikey eksende yapılan ölçekleme, döndürme ve üst üste bindirme gibi işlemlerden etkilenmez [14].

Zaman serilerinin analizi aslında bir veri sıkıştırması olarak tanımlanabilir [114]. Karmaşık ağlarda son yıllarda sağlanan gelişmeler ve bazı önemli fiziksel ve doğal sistemlere yakınlığı zaman serilerinin analizinde de bu yöntemlerin önemli ölçüde kullanılmasına yol açmıştır [115].

Bir zaman serisi verildiğinde, sistemin temel dinamikleri büyük bir ölçüm örneğinden elde edilen birkaç karakteristik sayı ile hesaplanabilmektedir. Bu nedenle, karakteristik sayılarla temsil edilen azaltılmış bilgi, sistemin bazı spesifik özelliklerini vurgulamaktadır [77].

Doğrusal olmayan zaman serisi analizi, sistem dinamiklerini bir dizi doğrusal olmayan fark denklemi veya doğrusal olmayan sıradan diferansiyel denklemlerle araştıran ve *kaos teorisi* olarak adlandırılan hızlı gelişmeden kaynaklanmaktadır. Daha da önemlisi, bu kavram güç spektrumu veya spektral tutarlılık gibi klasik lineer teknikler kullanılarak çözülemeyen bilgilerin çıkarılmasını sağlar [77].

Her ne kadar bütün zaman serileri için en iyi yöntem olmasa da karmaşık ağlar ile zaman serisi analizi, tekil zaman serilerinin sistem karakteristiğini kestirme işlemi için etkili bir yaklaşımdır [114]. Zaman serilerinin analizinde karmaşık ağlar aşağıdaki amaçlar için kullanılmaktadır [77].

- Tekil bir zaman serisinden sistem karakteristiğinin tespiti.
- Bir sinyali diğer bazı sinyallerden ayırma.
- Sinyallerin temel dinamik özelliklerindeki genel karakter değişimlerinin veya dönüşüm noktalarının tespiti.
- Zaman serilerinin tersine çevrilebilirliğinin tespiti.
- Gürültü giderme ve filtreleme.
- Sonraki zaman serilerinin tahmin edilmesi.

Zaman serilerini erişilebilir kılmak için karmaşık ağ analizi teknikleri ile bu serilerin uygun bir ağ temsiline dönüştürülmesi gerekmektedir. Bu işlem için ilk olarak, ağın düğüm noktalarını ve kenarlarını tanımlayan bir algoritma gerektirir. Bu tanımlara bağlı olarak, zaman serilerinin karmaşık ağlar ile analizine yönelik en az üç ana yaklaşım modeli bulunmaktadır ve bu yaklaşımlar Çizelge 2.1'de gösterilmiştir [77].

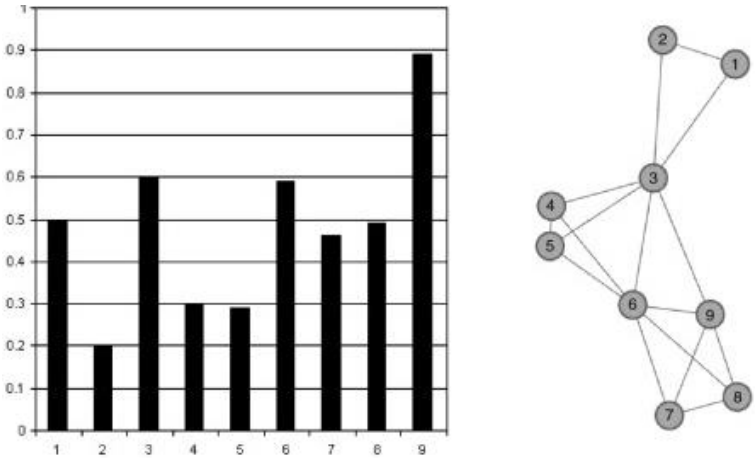
Çizelge 2.1. Mevcut karmaşık ağ yaklaşımlarında düğüm ve kenarların varlığına ilişkin kriterlerin özeti.

| Yöntem | Düğüm | Kenar | Yönlülük |
|---|---|---|---|
| **Yakınlık Ağları** | | | |
| Çevrim ağları | Çevrim | Döngüler arasındaki korelasyon veya faz alanı mesafesi | Yönsüz |
| Korelasyon ağları | Durum vektörü | Durum vektörleri arasındaki korelasyon katsayısı. | Yönsüz |
| **Yineleme Ağları** | | | |
| K en yakın komşu ağları | Durum (vektör) | Durum tekrarı (sabit komşuluk kütleri) | Yönlü |
| Uyarlanabilir en yakın komşu ağları | Durum (vektör) | Durum tekrarı (sabit kenar sayısı) | Yönsüz |
| ε-tekrar ağları | Durum (vektör) | Durum tekrarı (sabit komşuluk yoğunluğu) | Yönsüz |
| **Görünürlük grafları** | | | |
| Doğal görünürlük grafları | Sayısal durum | Durumların karşılıklı görünürlüğü | Yönsüz |
| Yatay görünürlük grafları | Sayısal durum | Durumların yatayda karşılıklı görünürlüğü | Yönsüz |
| **Geçiş Ağları** | | | |
| Eşik tabanlı ağlar | Faz uzayı ayrımı | Zamansal ardıllık | Yönlü |
| Soralı desen ağları | Sıralı desenler | Zamansal ardıllık | Yönlü |

Zaman serilerini karmaşık ağlar ile temsil etmek için yukarıdaki karakteristikler dışında Lyapunov exponent, entropiler ve korelasyon boyutları gibi yöntemler de kullanılmaktadır [116].

GG aşağıdaki görünürlük kriterleri ile tanımlanmıştır. $(t_a, y_a)$ ve $(t_b, y_b)$ görünürlüğe sahip olan ve dolayısı ile birleşmiş graf üzerinde bağlı iki düğüm olsun. Eğer bu iki düğüm arasına yerleştirilen bir $(t_c, y_c)$ düğümü varsa; bir zaman serisinden çıkartılan ilişki grafı modeli Eşitlik 2.17'e şeklinde olacaktır.

$$y_c < y_b + (y_a - y_v)\frac{t_b - t_c}{t_b - t_a},\ \forall t_c \in (t_a, t_b) \tag{2.17}$$




Şekil 2.14. Görünürlük graflarının bir başka temsili.

Değerleri 0.5, 0.2, 0.6, 0.3, 0.29, 0.59, 0.46, 0.49, 0.89 olan bir başka zaman serisinin görünürlük grafı Şekil 2.14'da verilmiştir. Bu örnekte, göreceli değerleri daha büyük olduğu için 3. ve 9. sıradaki düğümlerin görünürlükleri en iyi olacağından bu düğümlerin derecelerinin en yüksek değerde olması beklenir [117].

GG ile ilgili çalışmalardan birini de Zhang ve Small gerçekleştirmiştir. Bu amaçla zaman serileri ve karmaşık ağlar asarında farklı bir haritalama yöntemi geliştirmişlerdir. Bu yöntemde graf teorisine dayalı olarak zaman serisi bir graf ile temsil edilmiştir. Bu çalışmada araştırmacılar karmaşık ağların *small world*'e uygunluk ve ölçeksiz olma özelliklerinden dolayı kaotik dinamikler sergileyen bir zaman serisinin ikilisi olarak düşünülebileceğini ortaya koymuşlardır. Bu nedenle kaotik dinamik sistemler, karşılık gelen karmaşık bir ağ topolojisi imzasına sahiptir [115].

# BÖLÜM 3

# YAPAY ÖĞRENME

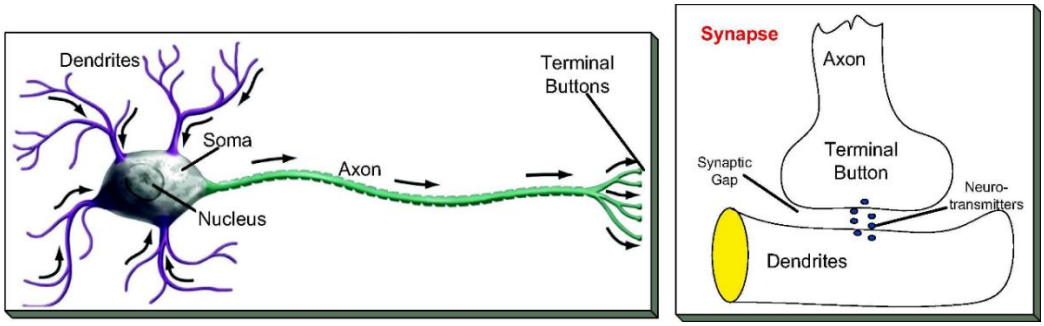
Sınıflandırma problemi, bir nesnenin o nesneyle ilgili bir dizi gözlenen özniteliğe dayalı olarak önceden tanımlanmış bir grup veya sınıfa atanması gerekliliğinden ortaya çıkar. Parmak izi ve yüz tanıma gibi biyometrik güvenlik sistemleri, ortam sesi tanıma, konuşma tanıma, kredi puanlama, tıbbi teşhis, kalite kontrol, el yazısı karakter tanıma gibi işletme, bilim, endüstri ve tıp alanlarındaki birçok problem, sınıflandırma problemleri olarak ele alınabilir [118].

Haykin'e göre, bir YSA, insan beyninin ilgilenilen belirli bir görevi yerine getirmesiyle aynı şekilde çalışmak üzere üretilmiş karşılaştırılabilir bir makine olabilir [119]. Beynin biyolojik sinir sisteminin çalışmasını modelleyerek geliştirilen YSA, yukarıda bahsedilen problemleri çözmek için oldukça kullanışlı bir yaklaşımdır. YSA, genellikle yapay zekanın alt dalları olan makine öğrenmesi ve optimizasyon gibi alanlarda yaygın olarak kullanılmaktadır.

YSA uygulamasının en önemli avantajlarından biri, büyük girdiler için karmaşık doğal sistemlerden daha kolay ve daha doğru hesaplamalar yapabilen modeller getirebilmesidir. Son yıllarda geliştirilen derin öğrenme modelleri ile de YSA'nın problem çözme ve makine öğrenimine sağladığı katkılar daha da geliştirilmiştir.

Yapay Sinir Ağları (YSA), biyolojik sinir sistemlerinin matematiksel simülasyonu ile tasarlanan doğrusal olmayan öğrenme modellerdir. İnsan beyni, bir görevi yerine getirmek için kolayca koordine edilebilen çeşitli karmaşık sinyal ve veri hesaplama fonksiyonlarına sahip son derece verimli, büyük ve benzersiz bir makinedir [120].

Son zamanlarda, beyin işlevselliğine yönelik araştırmalar bütün dünyada hızla artmaktadır. İnsan beyni yaklaşık olarak 86 milyar nörondan oluşan büyük bir ağ yapısına sahiptir [121]. Beynin yapısı detaylı olarak biliniyor olsa bile tüm gelişmelere rağmen günümüzde bu yapıyı tam olarak simüle edecek bir teknoloji bulunmamaktadır.




Şekil 3.1. Bir biyolojik nöronda sinyal akışının yönü ve bir sinaps yapısı.

Biyolojik öğrenme sistemlerinde bu, hücresel düzeyde nöronlarda gerçekleşir.
Şekil 3.1'de biyolojik bir sinir sisteminde bulunan bir nöronda sinyal akışının yönü ve bir sinaps gösterilmiştir [122]. Burada her nöron elektrik sinyalini üç farklı işlem üzerinden iletir.

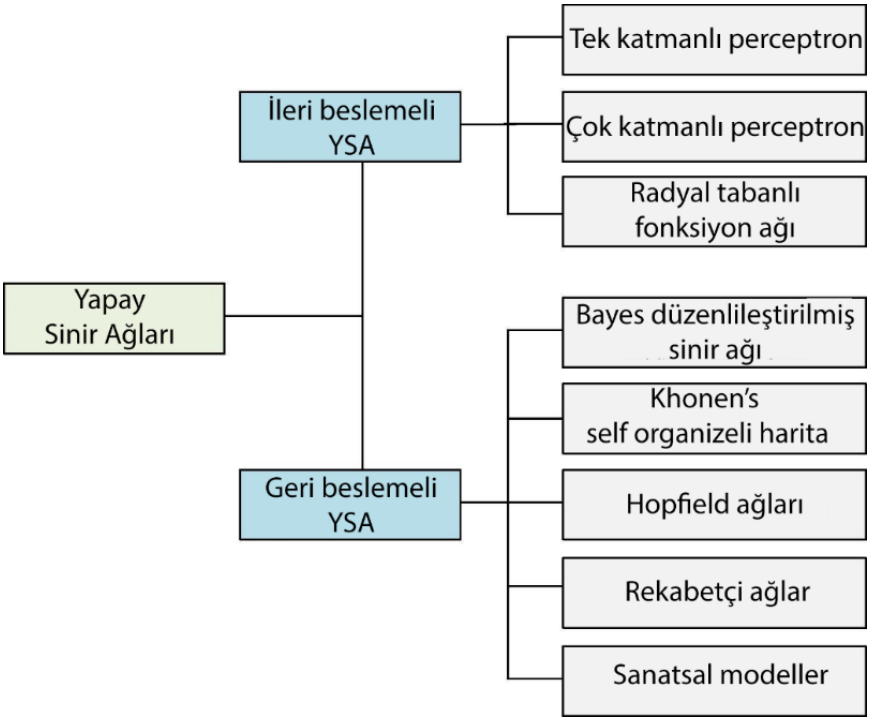
- **Dendritlerdeki sinaptik bağlantılar:** Giriş nöronunun akson terminallerinin sinaps bağlantısından gelen her giriş sinyali üzerinde ayrı bir işlem gerçekleştirilir.
- **Sinyal akışı:** Soma'da uzamsal ve zamansal sinyal entegratörü aracılığıyla çalıştırılan giriş sinyallerinin bir havuzlama işlemi ile birleştirilmesi.
- **Akson'un tepeciğinin ilk bölümünde bir aktivasyon:** Eğer havuzlanmış potansiyeller belirli bir limiti aşarsa bu durum "*aksiyon potansiyelleri*" olarak adlandırılan bir dizi tetiklemeyi etkinleştirir.

Her sinapsın fiziksel ve nörokimyasal özellikleri, yeni giriş sinyalinin sinyal gücü ve polaritesi ile birlikte genel olarak doğrusal olmayan sinyal çalışmasını belirler [123]. Bilgi depolama veya işleme, hücrelerin sinaptik bağlantılarında veya daha kesin olarak bu

bağlantıların belirli işlemleriyle oluşan örneğin ağırlık gibi bağlantı güçleri ile birlikte yoğunlaşır [123].

Sinyal iletim modlarına göre YSA'lar Şekil 3.2'de gösterildiği gibi ileri beslemeli ve geri beslemeli sinir ağları olarak iki alt grupta sınıflandırılmıştır [124]. Örneğin Elman ağları [125] gibi geri beslemeli sinir ağları YZ alanında önemli bir rol oynamaktadır.

Bazı araştırmacılar çok katmanlı algılayıcı gibi ileri beslemeli YSA'lar ile Elman ağları gibi geri beslemeli YSA modellerini biyosorpsiyon kapasitesindeki uygulamada karşılaştırmış ve ileri beslemeli sinir ağlarında $\pm 10\%$ gibi daha küçük tahmin hatalarını rapor etmişlerdir. Bu hatalar geri beslemeli ağlarda $\pm 12\%$ olarak tespit edilmiştir [126].



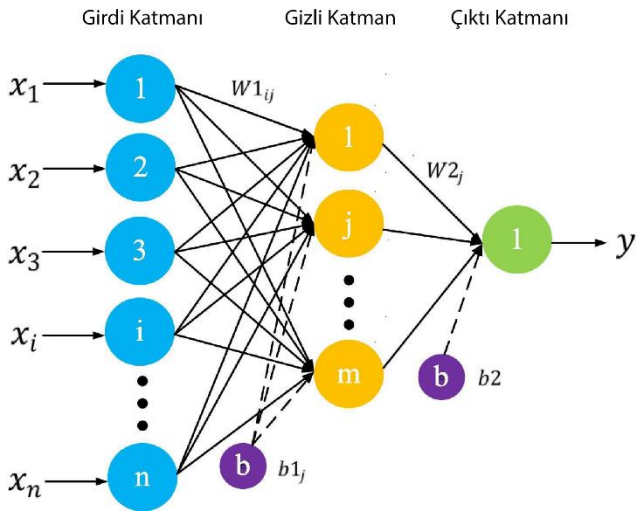

Şekil 3.2. YSA'ların sınıflandırma modelleri.

İleri beslemeli sinir ağları nispeten daha basittir ve ilgili çalışmalarda yaygın olarak kullanılmaktadır. Yetilmezsoy vd.'nin 2011'de yaptığı bir review çalışmasına göre

2010'den önce yapılan çalışmaların 97%'sinde ileri beslemeli YSA'lar kullanılmıştır [127].

## 3.1. ÇOK KATMANLI ALGILAYICI SİNİR AĞLARI (MULTILAYER PERCEPTRON NEURAL NETWORK)

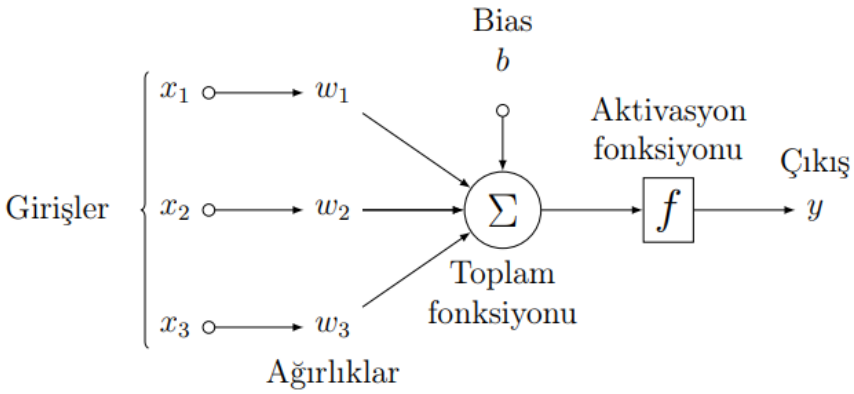
Çok Katmanlı Algılayıcı Sinir Ağları (ÇKASA), hızlı çalışması, uygulama kolaylığı ve daha küçük eğitim seti gereksinimleri nedeniyle en yaygın kullanılan ileri beslemeli sinir ağlarıdır [128–130]. Çok katmanlı ileri beslemeli bir sinir ağında, bir katmandaki nöronlar, farklı ağırlıklardaki farklı bağlantılarla bir sonraki katmanda bulunan nöronlara bağlanır.




Şekil 3.3. Çok katmanlı algılayıcı bir sinir ağının yapısı.

Şekil 3.3'de görüldüğü gibi ÇKASA katmanları girdi, gizli ve çıktı olmak üzere birbirini takip eden 3 katmandan oluşmaktadır [124]. Gizli katman, girdi katmanından aldığı verileri işler ve çıktı katmanına iletir. Gizli katmanda yetersiz veya aşırı sayıda nöron içeren bir ÇKASA modeli, büyük olasılıkla kötü genelleme ve aşırı öğrenme gibi sorunlara neden olur ve gizli katmandaki nöron sayısını belirlemek için analitik bir yöntem yoktur [131]. Bu nedenle, nöron sayısı yalnızca deneme yanılma yoluyla bulunur [119,129].

YSA'lar genellikle yapay sinir hücrelerinin birleşiminden oluşmaktadır. Şekil 3.4'de girişler, aktivasyon fonksiyonları ve çıkışlardan oluşan tek bir YSA nöronu görülmektedir [105]. Formal olarak bir YSA nöronunun girişi, $\vec{x} \in \mathbb{R}^n$ şeklinde ifade edilen $n$ büyüklüğünde bir vektördür. Girişler biyolojik nörondaki dentritlere, çıkış ise akson'a karşılık gelmektedir. Katmanlar arasında $1 < i < n$ girişlerine $w_1, w_2, \dots, w_n$ ağırlıkları atanmıştır. Girişlerin toplamı aktivasyon fonksiyonuna bir girdi olarak verilir. Aktivasyon fonksiyonundan elde edilen çıktı ise nöronun çıkışı olarak alınmaktadır.




Şekil 3.4. Girişler, aktivasyon fonksiyonu ve çıkıştan oluşan tek bir sinir hücresi.

Bir YSA için 6 tane tanımlama grubu bulunmaktadır.

$N$ : Boş olmayan sonlu nöronlar kümesi

$C \subseteq NxN$ : Nöronlar arasındaki boş olmayan yönlendirilmiş kenarlar

$X \subset N$ : Giriş katmanındaki boş olmayan nöronlar

$Y \subset N$ : Çıkış katmanındaki boş olmayan nöronlar

$w$ : $C \longmapsto \mathbb{R}$ ağırlıklandırma fonksiyonu

$t$ : $C \longmapsto \mathbb{R}$ ağ biasları için fonksiyon

Bir YSA'nın girdi katmanındaki nöronlar, dış ortamdan gelen verilerle beslenir. Çıktı katmanındaki nöronlar ise, ara katmanlardaki işlemlerden sonra değer üretir.

Gizli katmandaki her bir $j$'inci nöron, kendisine çarpan $x_i$ giriş sinyallerini, ilgili bağlantıya karşılık gelen $w_{ij}$ ağırlıklarıyla çarptıktan sonra toplar. Her bir nöronun çıkışı Eşitlik 3.1'e göre hesaplanır.

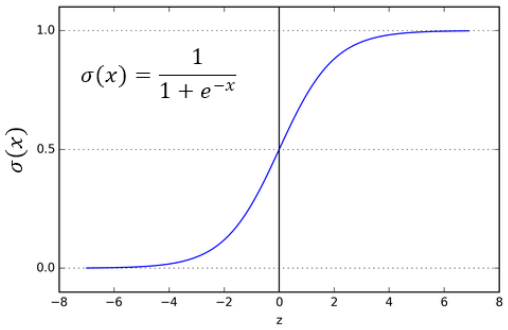
$$y_i = f\left(\sum w_{ji} x_i\right) \tag{3.1}$$

Burada; $f$: girdilerin ağırlıklı toplamlarını kullanan aktivasyon fonksiyonunu göstermektedir. Bir aktivasyon fonksiyonu basit bir eşikleme, sigmoid veya hiperbolik tanjant fonksiyonu olabilir [133,134].

### 3.1.1. Sigmoid Aktivasyon Fonksiyonu

Sigmoid fonksiyonu, doğrusal olmayan bir fonksiyon olduğu için en yaygın kullanılan aktivasyon fonksiyonudur. Eşitlik 3.2 ile elde edilen sigmoid fonksiyonu çıkış değerlerini [0..1] aralığına dönüştürür ve türevlenebilir bir fonksiyondur.

$$\sigma(x) = \frac{1}{1 + e^{-x}} \tag{3.2}$$



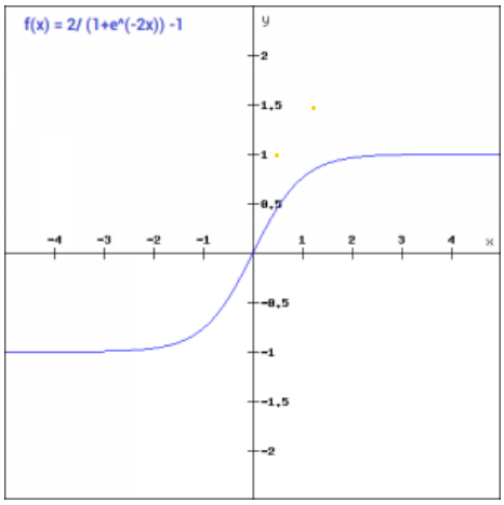

Şekil 3.5. Sigmoid aktivasyon fonksiyonu.

Şekil 3.5'de sigmoid aktivasyon fonksiyonu gösterilmiştir [135]. Sigmoid fonksiyonu, girişleri [0..1] aralığına dönüştürdüğünden olasılık hesaplamalarına uygundur ve herhangi bir fonksiyonun çıkışını tahmin etme problemlerinde yaygın olarak kullanılmaktadır.

### 3.1.2. Hiperbolik Tanjant (tanh) Aktivasyon Fonksiyonu

Hiperbolik tanjant fonksiyonu sigmoid fonksiyonuna benzemektedir ancak orijine göre simetriktir. Bu durum, bir sonraki katmana girdi olarak beslenecek olan önceki katmanlardan farklı çıktı işaretleri ile sonuçlanır. Eşitlik 3.3 ile elde edilen hiperbolik tanjant fonksiyonu $[-1..1]$ aralığında süreklidir ve türevlenebilir bir fonksiyondur.

$$\tanh(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}} \tag{3.3}$$




Şekil 3.6. Hiperbolik tanjant fonksiyonu.

Şekil 3.6'da hiperbolik tanjant aktivasyon fonksiyonu gösterilmiştir [135]. Sigmoid fonksiyonu ile karşılaştırıldığında, hiperbolik tanjant fonksiyonunun gradyanı daha diktir. Hiperbolik tanjant, belirli bir yönde değişmekle sınırlı olmayan gradyanlara sahip olduğu için, ayrıca sıfır merkezli olduğu için sigmoid fonksiyona tercih edilmektedir [135].

YSA'nın çıkış nöronlarının istenen ve gerçek değerleri arasındaki farkların karelerinin toplamı $E$, Eşitlik 3.4'e göre hesaplanır.

$$E = \frac{1}{2}\sum_{j}\left(y_{dj} - y_j\right)^2 \tag{3.4}$$

Bu eşitlikte, $y_{dj}$: $j$'inci çıkış nöronunun istenen değerini, $y_j$ ise bu nöronun hesaplanan değerini ifade etmektedir. Burada her $w_{ji}$ ağırlığı, kabul edilen eğitim algoritmasına göre $E$'nin değerini optimize etmek için ayarlanır [119,129,131]. Bu bağlamda, geri yayılım algoritması, bir YSA modelinin birincil parçası olarak yaygın olarak kullanılmaktadır. Ancak, geri yayılımın yavaş yakınsama gibi bazı kısıtlamaları olduğundan veya hata fonksiyonunun global minimum değerini bulamadığından geri yayılım için farklı birçok varyasyon önerilmektedir [129].

## 3.2. DERİN ÖĞRENME VE EVRİŞİMSEL SİNİR AĞLARI

Yapay Sinir Ağları (YSA) alanında yapılan çalışmalar sonucunda günümüzde gelinen son aşamada Derin Yapay Sinir Ağları (DYSA) modelleri kullanılmaktadır. DYSA'lara dayalı Evrişimsel Sinir Ağları (ESA), makine öğrenmesi yöntemlerinin bir alt dalıdır ve son yıllarda özellikle bilgisayarlı görme, çevresel seslerin tanınması, konuşma tanıma, doğal dil işleme gibi alanlarda çok yaygın olarak kullanılmaktadır [136,137].

Bir DYSA, nöronların bir önceki katmandan nöron aktivasyonlarını girdi olarak aldığı ve basit bir hesaplama gerçekleştirdiği çoklu katman dizisinde düzenlenen nöronların bir koleksiyonudur [137]. Ağın nöronları, girdiden çıktıya kadar karmaşık ve doğrusal olmayan bir eşlemeyi ortaklaşa uygular. Bu haritalama, hata geri yayılımı adı verilen bir teknik kullanılarak her bir nöronun ağırlıklarının uyarlanmasıyla elde edilen verilerden öğrenilir.

Derin öğrenme alanındaki ilk uygulama, 1998 yılında LeCun ve arkadaşlarının yayınlamış oldukları bir makale ile yapılmış ve bu kavram duyurulmuştur [138]. Bu alanındaki ikinci önemli gelişme ise 2012'de yapılan ImageNet [139] yarışmasında gerçekleşmiştir. Bu yarışmayı derin öğrenme mimarisi ile gerçekleştirilen AlexNet modeli kazanmış ve bu çalışma aynı yıl Krizhevsky ve Sutskever tarafından bir makale olarak yayınlanmıştır [26].

### 3.2.1. DYSA Çeşitleri

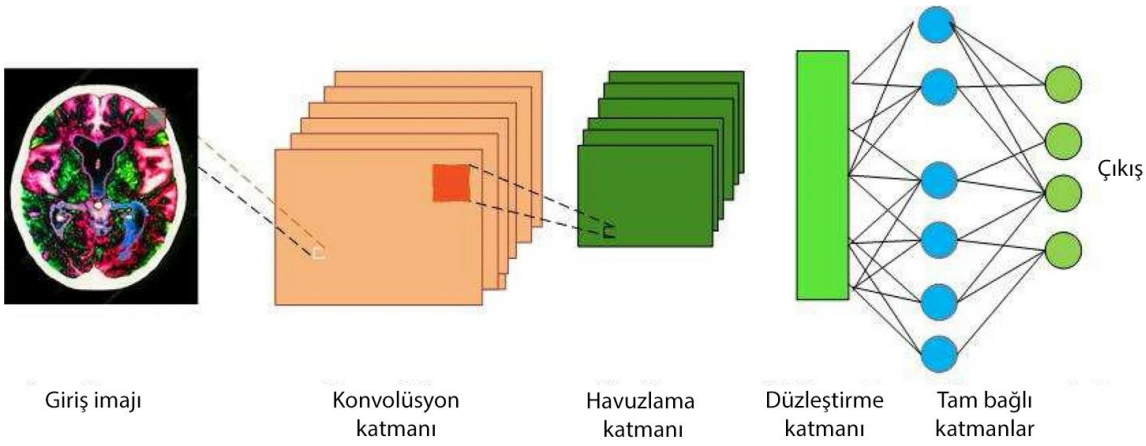
Doğrusal olmayan birçok sinir ağı katmanından oluşan DYSA'lar ileri-beslemeli olarak çalışmaktadır. Derin öğrenme yöntemleri, özellik çıkartma ve dönüşüm işlemleri için danışmanlı veya danışmansız olarak modellenmektedir. Derin ağlar, geleneksel makine öğrenmesi tekniklerinin aksine öğrenme işlemini ham veri üzerinde yapmakta ve ham veriyi işlerken gerekli bilgiyi farklı katmanlardan çıkartmış olduğu sonuçlarla elde etmektedir.

Donanım ve yazılım teknolojilerinin ilerlemesi ve makine öğrenmesi alanında yeni optimizasyon algoritmalarının geliştirilmesi ile derin öğrenme için sürekli yeni yöntemler geliştirilmiştir. Kullanılmakta olan derin öğrenme çeşitleri aşağıda açıklanmıştır [140].

- **Danışmansız (Unsupervised) Derin Ağlar:** Üretken Öğrenme olarak nitelendirilen bu ağlar, etiketlenmiş sınıflar olmadan çalışmaktadır. Danışmansız öğrenme modeli, veride bulunan yüksek dereceli korelasyondan faydalanarak pikseller arasındaki örüntülere göre çalışır [141].
- **Danışmanlı derin Ağlar:** Bu model, girişten aldığı etiketlenmiş verilerin ayırt edici özelliklerini kullanarak bu özelliklere göre bir sınıflandırma yapar. Bu model, makine öğrenmesinde kullanılan en yaygın yöntemdir. Danışmanlı öğrenme, eğitim ve test için oldukça etkili bir modeldir ve kurulum için çok esnek bir yapıya sahiptir. Ayrıca bu model, karmaşık sistemlerin eğitimi için oldukça elverişlidir [142].
- **Melez Derin Ağlar:** Yarı-danışmanlı olarak adlandırılan bu ağlar, veriler üzerinde üretilmiş özellikler ve mevcut bulunan ayırt edici özelliklerin birlikte kullanılması yöntemi ile çalışır. Bu ağlar, (örneğin generative adversarial networks or GANs [143]) etiketli ve etiketsiz verilerle çalışır. Yarı-denetimli öğrenme, sınırlı eğitim örnekleri problemiyle başa çıkmak için hiperspektral görüntü sınıflandırmasında çok faydalıdır [144].

### 3.2.2. Evrişimsel Sinir Ağlarının Yapısı

Evrişimsel Sinir Ağları (ESA), görüntü içeriğini anlamak için kullanılan en iyi öğrenme algoritmalarından biridir ve görüntü bölümlendirme, sınıflandırma, algılama ve alma ile ilgili görevlerde örnek teşkil eden performans göstermiştir [145–147]. Bir ESA'nın mimarisi, girdi görüntüsünün iki boyutlu yapısından yararlanmak için tasarlanmıştır [148].




Şekil 3.7. Evrişimsel bir derin öğrenme modelinin blok diyagramı.

Şekil 3.7'de gösterildiği gibi, evrişimsel katmanlar, aktivasyon fonksiyonları, havuzlama ve tam bağlantılı katmanlar, ESA'ların temel yapı taşlarıdır [149]. ESA'lar, nöronlara benzer şekilde öğrenilebilir ağırlıklara ve biaslara sahip evrişim katmanlarından oluşmaktadır [149].

ESA'lar, danışmalı doğrusal olmayan sinir ağlarıdır ve derin öğrenme modelinin özel bir çeşididir. ESA'lar, birçok dizi setinden oluşan veriyi işlemek amacıyla YSA'lar temel alınarak geliştirilmiştir [150]. Literatüre bakıldığında ESA için kullanılan üç farklı sınıflandırma bilgisi bulunduğu görülmektedir.

- **Sadece Spektral Bilgilerin Çıkartılması:** Spektral tabanlı sınıflandırma yaklaşımları, genel olarak gerçekleştirilmesi basit yöntemlerdir. ve belirli bir

piksel için en iyi sınıfı bulmak amacıyla istatistiksel terminolojileri kullanmaktadır. Fakat bu yöntem uzaysal bileşenleri ihmal eder. Bu yöntemde her pikselin saf ve tek bir bölge olarak etiketlenmiş olduğu varsayılmaktadır. ESA'lar ile spektral özellik sınıflandırması için orjinal imaj verisinin spektral özelliğinin giriş vektörü olarak ayarlanmış olması gerekmektedir. Böylece 1-boyutlu ESA mimarisi elde edilir ve $N \times 1$ uzunluğunda giriş vektörlerini alır [150].

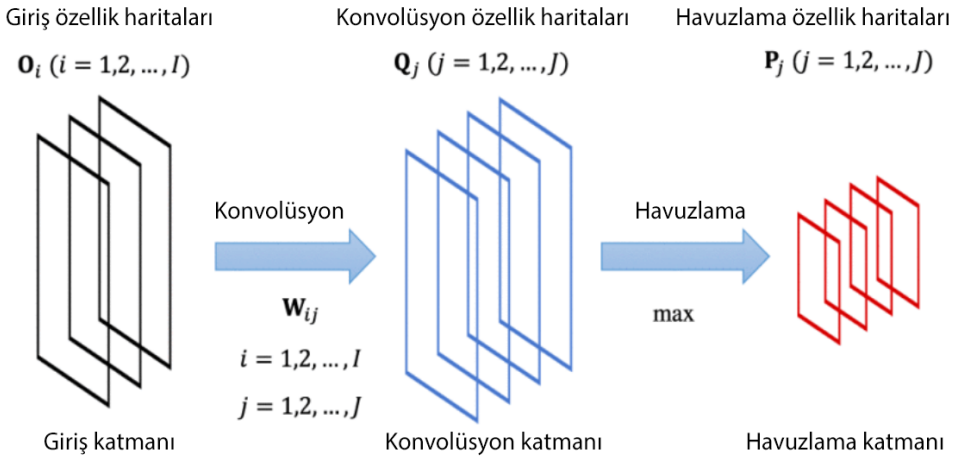
- **Sadece Uzaysal Bilgilerin Çıkartılması:** Bu model, girişten alınan orjinal imajdaki komşu pikseller arasında bulunan ilişkiyi kullanarak uzaysal 24 bit derinliğinde özellik çıkartır. Bu özellik 2-boyutlu ESA yapısına uygun olduğundan $P \times P$ boyutunda komşu piksellerden oluşan imajları giriş olarak almaktadır [151].
- **Spektral ve Uzaysal Bilgilerin Çıkartılması:** Bu yöntemde spektral ve uzaysal verilerin birleştirilmiş hali kullanılmaktadır ve modelin sınıflandırma başarımını önemli ölçüde artırdığı tespit edilmiştir. Bu modelde, 1-Boyutlu ve 2-Boyutlu ESA'lar için spektral ve uzaysal özellik çıkartma yöntemleri bir arada kullanılmakta veya uzaysal ESA'lar ile birçok farklı spektral özellik bir arada kullanılabilmektedir [152].

3-boyutlu mimariler, $B$ adet spektral bant için $P \times P$ boyutlu uzaysal imajda bulunan piksellerin komşuluk ilişkilerini hesaplar ve böylece $P \times P \times B$ boyutunda bir giriş verisi oluşmuş olur. Bu yapıya sahip modeller uzaysal ve spektral boyutunda bulunan yerel sinyal değişimlerini öğrenmek için 3-boyutlu kerneller kullanır.

ESA'lar, bloklardan oluşmakta ve bu bloklar imaj, ses, video sinyalleri gibi verileri uzay ve zaman alanında işlemektedir. Her blok, giriş verisini nöron aktivasyonlarının bir çıkış verisine dönüştürür. Bu çıkışlar bir sonraki bloğun girişi olarak kabul edilmektedir.

Evrişimsel sinir ağları eğitilebilen birçok katmandan oluşmaktadır. Her katmanın içerisinde konvolüsyon katmanı, öznitelik havuzlama katmanı, filtre bankası katmanı ve doğrusal olmayan katman olmak üzere dört katman bulunur. Filtre bankası katmanında değişik özniteliklerin çıkartılması için çekirdekler bulunur. Havuzlama katmanında, elde

edilen öznitelik haritaları bu katmanda ayrı ayrı ele alınır. ESA'ların katmanlar halindeki yapısı aşağıda açıklanmıştır.




Şekil 3.8. Evrişimsel sinir ağlarının bir katmanının temsili blok yapısı.

Derin bir ESA art arda konvolüsyon ve havuzlama çiftlerinden iki veya daha fazlasından oluşmaktadır. Bir konvolüsyon katmanından ve art arda bir havuzlama katmanından oluşan ESA'nın bir katmanının blok yapısı Şekil 3.8'de gösterilmiştir [153]. ESA terminolojisinde, art arda bir çift evrişim ve havuzlama katmanı genellikle bir ESA katmanı olarak adlandırılır.

Giriş katmanındakilere benzer şekilde, evrişim ve havuzlama katmanlarının birimleri de haritalar halinde düzenlenebilir. Burada konvolüsyon katmanı havuzlama katmanına haritalanmaktadır.

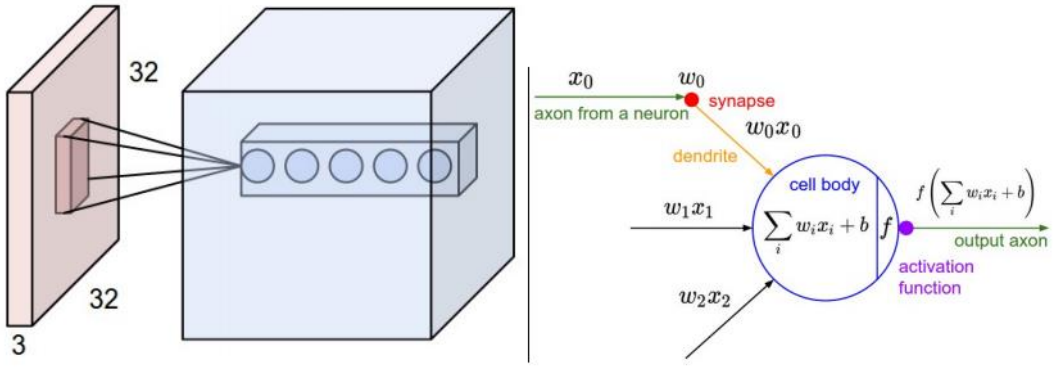
#### 3.2.2.1. Giriş (Input) Katmanı

Bu katman, ESA'nın ilk katmanını oluşturmaktadır ve veriyi dış ortamdan ham olarak alır. Bu katmanın yapısı tasarlanan modelin başarımı ve kaynak maliyeti açısından oldukça önemlidir. Giriş görüntü boyutunun büyük seçilmesi durumunda ağın başarısının

artmasına karşılık sistem kaynak ihtiyacı, eğitim ve test süreci gibi maliyetler yükselmektedir.

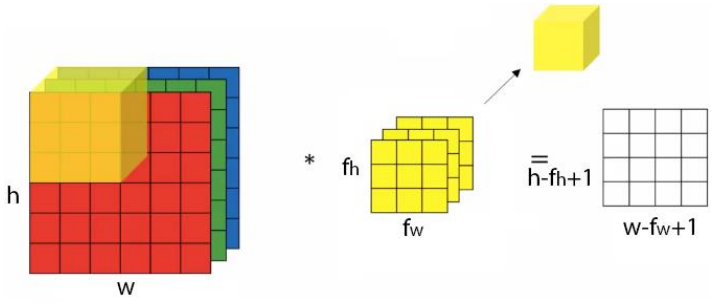
### 3.2.2.2. Konvolüsyon (Convolution) Katmanı

ESA'ların temelini oluşturan bu katman, dönüşüm katmanı olarakta bilinmektedir. Örneğin RGB gibi üç katmanlı bir imaj girdi olarak verildiğinde 3-boyutlu bir katman kullanılmış olur. Bu işlemde her nöron, giriş verisi üzerinde $3 \times 3$, $5 \times 5$, $7 \times 7$, $11 \times 11$ piksel boyutlarında çekirdekler kullanarak piksek ağırlıkları arasındaki ilişkiden bir değer üretir. Bu işlemin amacı bir önceki katmandan gelen imaj üzerinde belirleyici bir özellik çıkartmaktır.




Şekil 3.9. Konvolüsyon katmanına dış ortamdan $32 \times 32 \times 3$ boyutunda örnek bir imajın giriş olarak verilmesi.

Şekil 3.9'de ESA'ya dış ortamdan $32 \times 32 \times 3$ boyutunda bir RGB imaj giriyor ve ilk konvolüsyon katmanındaki nöronlara girişi görülmektedir [154]. Konvolüsyon katmanındaki her nöron giriş imajı üzerinde sadece bir bölgeye bağlanmıştır. Fakat, bir imajın 3 kanalı gibi bu bölgedeki bütün katmanları işlemektedir. Katmandaki derinlik boyunca bütün nöronlar giriş verisinde aynı bölgeye bakmaktadır.




Şekil 3.10. 3 katmandan oluşan $w \times h$ boyutunda bir imaja konvolüsyon işlemi.

Şekil 3.10'de görüldüğü gibi 3 katmandan oluşan $w\ x\ h$ boyutunda bir imaja konvolüsyon işlemi uygulandığında elde edilecek çıktı imajının boyutu $(w - f_w + 1) \times (h - f_h + 1)$ şeklinde olacaktır.

Konvolüsyon katmanına giren 3-boyutlu bir imajdan 2-boyutlu $m \times n$ büyüklüğünde bir özellik haritası elde edilir. Burada her bir bileşen $x_{m,n}^{i}$ ve her bir özellik haritası $x^i$ şeklinde ifade edilir. Ayrıca, $m_1 \times n_1$ büyüklüğünde $k$ adet özellik haritası kullanılması durumunda çıkış, $m_1 \times n_1 \times k$ büyüklüğünde 3-boyutlu da olabilmektedir. Konvolüsyon katmanı, $l \times l \times q$ boyutunda k adet eğitilebilir filtreye sahiptir ve bu filtre bankası $(W)$ giriş özellik haritasını çıkış özellik haritasına dönüştürür.

$$z^s = \sum_{t=1}^{q} W_i^s * x^i + b_s \tag{3.5}$$

Konvolüsyon katmanının çıkış için hesapladığı özellik haritası Eşitlik 3.5'de gösterildiği gibidir. Bu denklemde $*$ konvolüsyon operatörü ve $b_s$ bias operatörüdür.

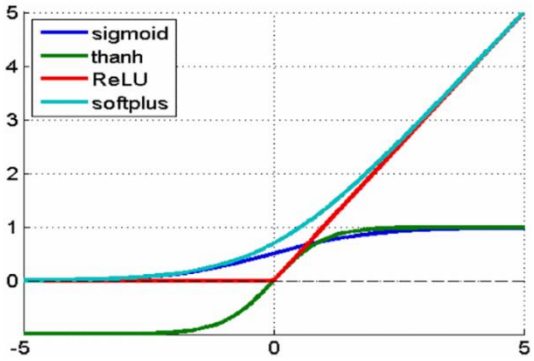
### 3.2.2.3. Nonlinearity Katmanı

Konvolüsyon katmanından sonra gelen katman nonlinearity katmanıdır ve üretilen çıktıyı ayarlamak veya kesmek için kullanılabilir. Bu katman, çıktıyı doyurmak veya üretilen çıktıyı sınırlamak için uygulanır.

Bu katmanda, doğrusal birimleri düzeltmek için lineer olmayan fonksiyonlar bulunmaktadır. Geleneksel YSA'larda bu katman her bir giriş bileşenini özellik haritasına dönüştüren bir nonlinearity fonksiyonudur.

$$ReLU(x) = \max(0, x) \tag{3.6}$$

$$\frac{d}{dx}(x) = \begin{cases} 1, & x > 0 \\ 0, & diğer \end{cases}$$



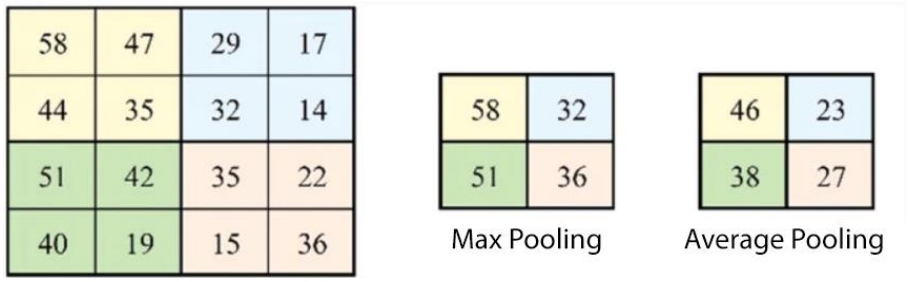

Şekil 3.11. Yaygın olarak kullanılan nonlinearity türleri.

Şekil 3.11'de yaygın olarak kullanılan nonlinearity türleri gösterilmiştir [154]. Uzun süre *sigmoid* ve *hiperbolik tanjant*, en yaygın kullanılan nonlinearity fonksiyonları olmuştur. Fakat, ESA'larda hem işlev hem de gradyan olarak daha basit tanımlamalara sahip olduğundan ve bazı avantajlarından dolayı Rectified Linear Unit (ReLU) daha yaygın olarak kullanılmaktadır [154].

#### 3.2.2.4. Havuzlama (Pooling) Katmanı

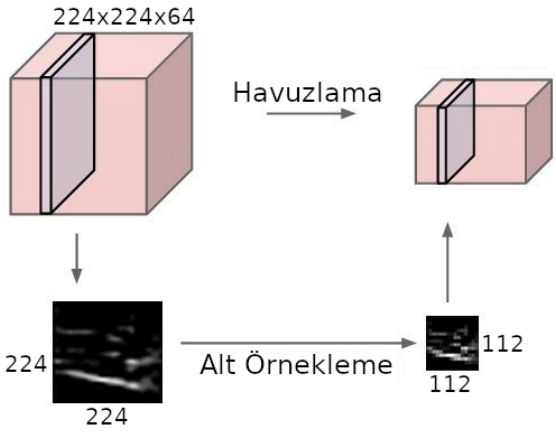
Havuzlama katmanının temel amacı bir sonraki konvolüsyon katmanı için giriş boyutunu azaltmaktır. Bu işlem verideki derinlik boyutunu etkilememektedir. Bu işlem sonucunda imajda bilgi kaybı ve boyutta azalma meydana gelir. Bu azalma, sonraki katmanlar için daha az hesaplama yükü oluşturacağından ve sistemin ezberlemesini önlemeye yardımcı olduğundan öğrenme için oldukça faydalıdır.

Bir önceki basamakta gerçekleştirilen Konvolüsyon işleminde olduğu gibi havuzlama katmanında da bazı filtreler kullanılır. Bu filtreler ile görüntü üzerindeki bölgelerde maksimum havuzlama (max-pooling) ve ortalama havuzlama (average pooling) gibi işlemler yapılır. ESA'larda genellikle maksimum havuzlamanın daha iyi sonuç verdiği tespit edilmiştir.




Şekil 3.12. Bir imaj üzerinde max havuzlama ve average havuzlama işlemleri.

Şekil 3.12'da $2 \times 2$ boyutlu bir filtre ile $4 \times 4$ boyutlu bir matris üzerinde maksimum havuzlama ve average havuzlama işlemlerinin işleminin gerçekleştirilmesi aşamaları gösterilmiştir [155]. Şekil 3.13'de ise $224 \times 224$ piksel çözünürlüğündeki bir görüntünün havuzlama işlemi sonucunda $112 \times 112$ piksele düşürüldüğü görülmektedir.




Şekil 3.13. Havuzlama işlemi sonucu görüntü boyutunun düşürülmesi.

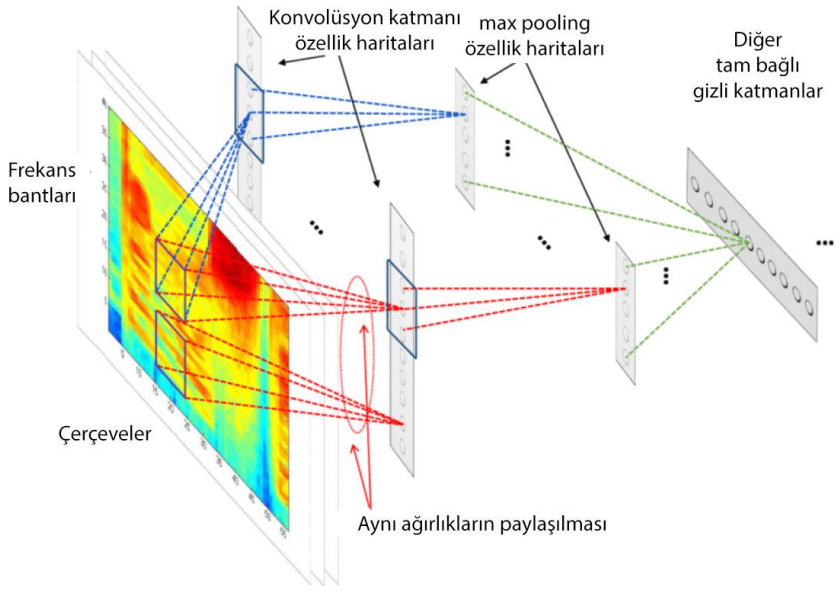
Veri üzerindeki uzaysal ve geçici değişmezliği sağlamak için sinir ağları genellikle havuzlama katmanlarını birleştirmektedir. $l_p$ 'inci havuzlama katmanının çıkışı Eşitlik 3.7'e göre olur [109].

$$a_k = \sqrt[p]{\sum_j a_j^p} \tag{3.7}$$

Burada $a_j$, $j$ havuz içerisindeki aktivasyonları ifade etmektedir. Bu katmandaki mümkün yayılma kuralı ise Eşitlik 3.8'e göre olacaktır.

$$R_{j \leftarrow k} = \frac{a_j}{\sum_j a_j} R_k \tag{3.8}$$

### 3.2.2.5. Tam Bağlantılı (Fully Connected) Katman




Şekil 3.14. Tam bağlantılı katman.

Tam bağlantılı katmandaki bütün nöronlar bir önceki katmanın bütün nöronlarına tam olarak bağlıdır. Bu katmanda matris çarpma işlemi yapılır ve bu çarpımda çıkan değerlere

bias değeri eklenir. Konvolüsyon katmanındaki nöronları giriş imajındaki sadece belli bölgelere bağlı olması ve konvolüsyon katmanındaki birden fazla nöronun aynı parametreleri paylaşması bu katmanı tam bağlı katmandan ayırmaktadır.

Şekil 3.14'de ağırlıkların tamamının paylaşıldığı tam bağlantılı katmanın ESA içerisindeki konumu gösterilmiştir [153]. Burada frekans bantları boyunca 1-boyutlu bir evrişim uygulanmıştır. Tam bağlantılı katman imajla ilgili global özellikleri çıkartırken softmax katmanı ise sınıflandırma işlemini gerçekleştirmektedir.

Tam bağlantılı katman, kendinden önce gelen havuzlama katmanının tüm alanlarına bağlıdır. Örnek bir ESA mimarisinde Havuzlama katmanından çıkan matrisin boyutu $25 \times 25 \times 256 = 160.000 \times 1$ ve tam bağlantılı katmandaki matrisin boyutu $4.096 \times 1$ olarak alınırsa toplam olarak $160.000 \times 4.096$ ağırlıklı bir matris oluşacaktır. Bunun anlamı, her bir $160.000$ nöronun $4.096$ nöron ile tam olarak bağlanması demektir.

#### 3.2.2.6. Seyreltme Katmanı

Tam bağlantılı katmanlarda belli eşik değerlerin altındaki düğümlerin seyreltilmesinin başarımı arttırdığı gözlenmiştir. Bunun anlamı zayıf bilgilerin unutulması öğrenimi arttırmaktadır [151].

### 3.2.3. Sınıflandırma Performans Ölçütleri

Değerlendirme ölçüleri, ESA'ların sınıflandırma performansını ölçmek için kullanılmaktadır. Bu ölçütler, göre sinir ağının modelini belirlemede ve mimarisini belirlemede önemli kriterlerdir. Sınıflandırıcının bir metrikte mükemmel, diğerinde kötü performans göstermesinden dolayı sınıflandırma algoritmalarının performansını değerlendirmek için birçok değerlendirme ölçütü geliştirilmiştir [156].

- **True Positive (TP):** Sınıflandırıcının pozitif bir sınıfı pozitif olarak doğru bir şekilde tahmin ettiği tahminlerin sayısını ifade eder.
- **True Negative (TN):** Sınıflandırıcının negatif sınıfı negatif olarak doğru bir şekilde tahmin ettiği tahminlerin sayısını ifade eder.
- **False Positive (FP):** Sınıflandırıcının negatif bir sınıfı pozitif olarak yanlış tahmin ettiği tahminlerin sayısını ifade eder.
- **False Negative (FN):** Sınıflandırıcının pozitif bir sınıfı negatif olarak yanlış tahmin ettiği tahminlerin sayısını ifade eder.

| | | Assigned Class | |
|---|---|---|---|
| | | Positive | Negative |
| Actual Class | Positive | TP | FN |
| | Negative | FP | TN |

Şekil 3.15. ESA'da değerlendirme ölçütleri için konfüzyon matrisi.

Şekil 3.15'deki ESA'ları değerlendirmek içim kullanılan konfüzyon matrisinden $Accuracy$, $Precision$, $Recall$ ve $F_1 score$ değerleri aşağıdaki gibi elde edilir.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{3.9}$$

$Accuracy$, modelin tahminlerinin modellenen gerçeklikle ne ölçüde eşleştiğinin bir ölçüsünü ifade etmektedir. Doğru tahmin sayısının toplam girdi örneği sayısına oranıdır.

$$Precision = \frac{TP}{TP + FP} \tag{3.10}$$

$Precision$, genellikle, tanımı gereği alınan ilgili belgelerin bir parçası olan Pozitif Tahmini Değer (PTD) olarak da adlandırılır [157].

$$Recall = \frac{TP}{TP + FN} \tag{3.11}$$

$Recall$, genellikle, doğru şekilde alınan ilgili numunenin bir parçası olan duyarlılık olarak da adlandırılır [157].

$$F_1 score = 2\frac{Precision * Recall}{Precision + Recall} \tag{3.12}$$

$F_1 score$, $Precision$ ve $Recall$ arasındaki harmonik ortalamayı belirler ve $[0,1]$ aralığında bir değer alır. Bu değer, kaç örneğin doğru sınıflandırıldığını saptayarak sınıflandırıcının ne kadar hassas ve sağlam olduğunu gösterir. $F_1 score$ değerinin büyük olması modelin performansının iyi olduğu anlamına gelmektedir.

Yüksel $Precision$'a karşılık düşük $Recall$ değerleri yüksek sınıflandırma başarısı fakat çok sayıda örneğin kaçırılması anlamına gelir. Bu durum sınıflandırma işlemini güçleştirmektedir.

#### 3.2.3.1. Top-k Sınıflandırma Ölçütü

Sınıflandırma performansını ölçmek için kullanılan kriterlerden birisi de top-k başarı ölçütüdür. Burada k değeri 1 veya 5 gibi bir değer alabilir. Örneğin, $k=5$ alındığında bir uygulayıcı 5 tahmini sıralamak ve bu sıralanmış tahminler arasında gerçek sınıfın sıralamasına bağlı olarak bir puan atamak isteyebilir. Böylece bulunan nesnenin sıralamadan bağımsız olmak üzere bu 5 tahminden biri içerisinde olması beklenir [158].

ImageNet veri seti üzerinde 5 farklı TÖ modelinin Top-1 ve Top-5 başarı değerleri Çizelge 3.1'de gösterilmiştir. Buradaki derinlik değeri ağın topolojik derinliğini ifade etmektedir [159].

Çizelge 3.1. Transfer Öğrenme modellerinin Top-1 ve Top-5 başarı değerleri.

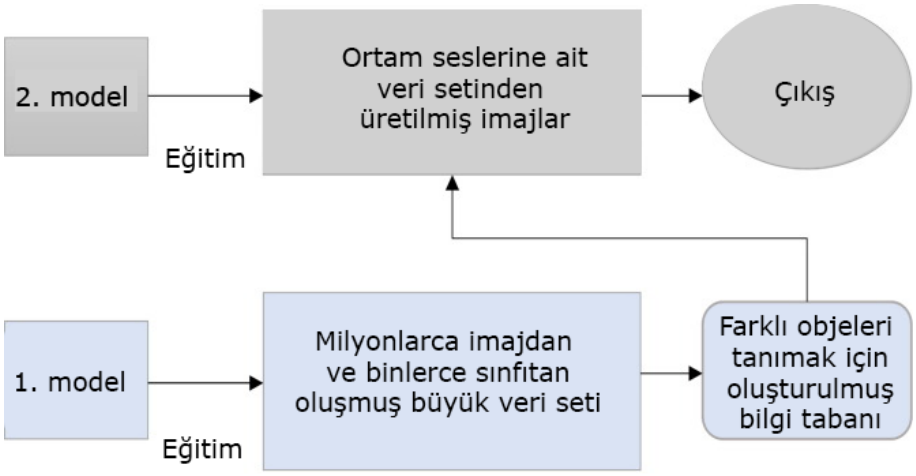
| Model | Top-1 Başarı | Top-5 Başarı | Parametre Sayısı | Derinlik |
|---|---|---|---|---|
| DenseNet201 | 0.773 | 0.936 | 20,242,984 | 201 |
| Inception-V3 | 0.779 | 0.937 | 23,851,784 | 159 |
| ResNet-50 | 0.749 | 0.921 | 25,636,712 | - |
| VGG-19 | 0.713 | 0.900 | 143,667,240 | 26 |
| Xception | 0.790 | 0.945 | 22,910,480 | 126 |

## 3.3. TRANSFER ÖĞRENME MODELLERİ

Transfer Öğrenme (TÖ) çalışmaları, yeni sorunları daha hızlı veya daha iyi sonuçlarla çözmek için daha önce öğrenilen bilgileri uygulayabilmektedir. TÖ, eğitim ve testte kullanılan örneklerin ve dağılımların farklı olmasına olanak tanır. Önceden eğitilmiş bu modeller zaten çok derin ve yoğun katmanlara sahiptir. Böylece imajlarla eğitilmiş bir TÖ modeli sesleri sınıflandırmak için de kullanılabilmektedir. Bu teknik sayesinden ESA sınıflandırma başarısında önemli gelişmeler sağlanmıştır.

TÖ, ilk olarak 1995'de NIPS-95 "Öğrenmeyi Öğrenme" konulu seminerde sunulmuştur ve bir alanda öğrenilen bilgiyi farklı ancak ilgili bir alana uygulamaya odaklanmıştır [160]. 2005 yılında, Savunma İleri Araştırma Projeleri Ajansı transfer öğrenmenin yeni bir tanımını yapmış ve sistemin önceki görevlerde öğrenilen bilgi ve becerileri yeni bir göreve aktarma ve uygulama yeteneği olarak açıklamıştır [160].

TÖ yönteminde önceden eğitilmiş ilk katmanların bir kısmı dondurulur. Bu ilk katmanlar daha az parametre içermesinin yanında yüksek bir hesaplama maliyeti gerektirmektedir. Sonraki adımlarda, modeller, döngüsel öğrenme tekniğine dayalı olarak optimal öğrenme oranları ile başlangıç katmanları çözülürken eğitilmektedir. Ayrıca, yapılan çalışmalarda TÖ modellerinin spektral imajların tanımasında düşük epoch döngülerinde bile oldukça iyi sonuçlar verdiği görülmüştür [58]. TÖ modelinin genel blok diyagramı Şekil 3.16'de gösterilmiştir. Şekilde de görüldüğü gibi TÖ modelleri, milyonlarca imajdan ve binlerce sınıftan oluşan ImageNet gibi veri setleri ile önceden eğitilmiştir.




Şekil 3.16. Transfer öğrenme modelinin genel blok diyagramı.

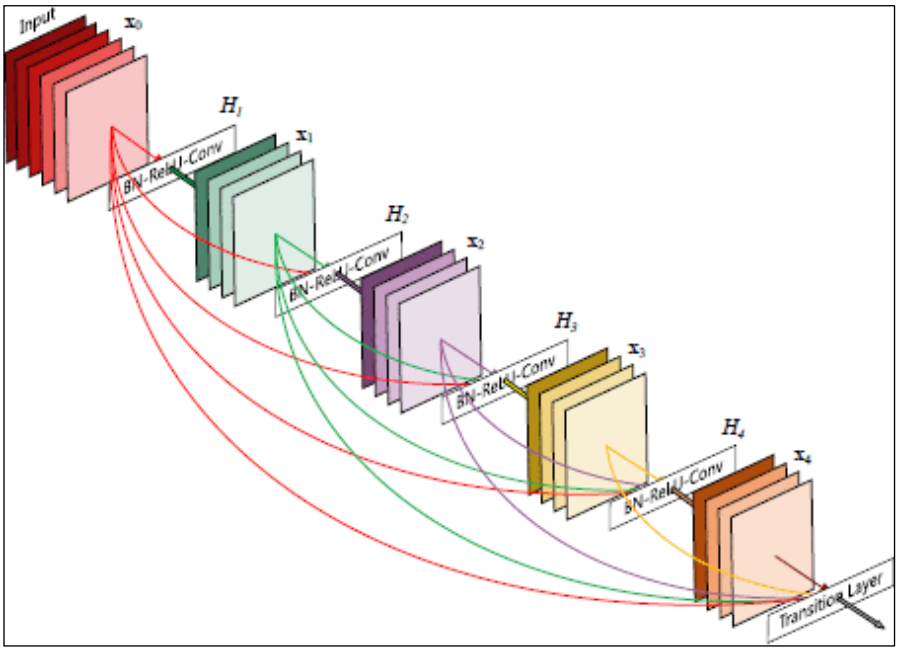
Çok çeşitli TÖ modelleri olmakla beraber burada bu çalışmada kullanılan 5 farklı model açıklanmıştır.

### 3.3.1. DenseNet201

DenseNet mimarisi Huang ve arkadaşları tarafından 2017'de önerilmiştir [161]. Bu ağ mimarisinde, ağdaki katmanlar arasında maksimum bilgi akışını sağlamak üzere benzer özellikli graf boyutuna sahip herhangi iki katman arasında doğrudan bir bağlantı oluşturulur. Böylece, daha kompakt ve doğru bir modelin öğrenilebilmesi için ağ özelliklerinin yeniden kullanılmasına izin verir. Aynı zamanda, sinir ağındaki örnek görüntünün özellik yayılımı güçlendirilir ve gradyan difüzyonu azalır. Bu mimaride, ileri beslemenin doğal yapısını korumak amacıyla her katman, önceki tüm katmanlardan ek girdiler alır ve sonraki tüm katmanlara kendi özellik haritalarını iletir.

DenseNet modelinin blok diyagramı
Şekil 3.17'de gösterilmiştir [161]. Sonuç olarak, $\ell$'inci katman, $x_0, \ldots x_{\ell-1}$ şeklinde girdi olarak önceki bütün katmanlardan gelen özellik haritalarını alır. Şekil 3.17'de görüldüğü gibi mimarideki alt örneklemeyi kolaylaştırmak için ağ, *dense* bloklara bağlanan birden

çok *dense* katmana bölünmüştür. Bu mimaride bloklar arasındaki katmalar, konvolüsyon ve havuzlama yapan *transition layers* olarak adlandırılmıştır.




Şekil 3.17. Büyüme oranı k = 4 olan 5 katmanlı DenseNet mimarisinin blok diyagramı. Her katman, önceki tüm özellik haritalarını girdi olarak alır.

$$x_\ell = H_\ell([x_0, x_1, \dots x_{\ell-1}]) \tag{3.13}$$

Eşitlik 3.13'de $[x_0, x_1, \dots x_{\ell-1}]$; $0,1, \dots, \ell - 1$ katmanları ile kurulan özellik haritaları arasındaki bağlantıları göstermektedir. Bağlantının yoğunluğundan dolayı bu ağ mimarisine Dense *Convolutional Network (DenseNet)* olarak adlandırılmıştır. Uygulamayı kolaylaştırmak için, Eşitlik 3.13'deki $H_\ell(\cdot)$'e ait birden fazla olan giriş birleştirilerek tek bir tensöre indirgenmiştir.

Eşitlik 3.13'de kullanılan birleştirme işlemi özellik haritalarının boyutu değiştirildiğinde uygulanabilirlik özelliğini yitirir. Bununla birlikte, evrişimsel ağların önemli bir kısmı özellik haritalarının boyutunu değiştiren katmanlar üzerinde bir alt örnekleme uygular.

Eğer her $H_\ell(\cdot)$ fonksiyonu $k$ özellik haritaları üretirse; bu fonksiyon $l$'inci katmana ait $k_0 + k \times (\ell - 1)$ giriş özellik haritaları olduğunu gösterir. Burada $k_0$, giriş katmanındaki kanalların sayısını ifade etmektedir.

Çizelge 3.2. ImageNet için oluşturulmuş DenseNet mimarisi. Buradaki büyüme oranı ilk 3 ağ için $k = 32$ ve DenseNet-161 için $k = 48$ dir. Tabloda gösterilen her konvolüsyon katmanı, BN-ReLU-Conv dizilimine karşılık gelir.

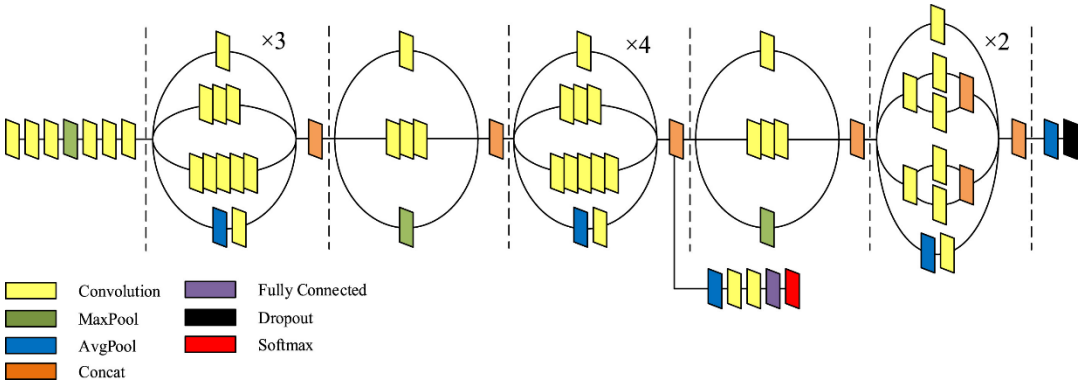
| Katmanlar | Çıkış Boyutu | DenseNet-121($k = 32$) | DenseNet-169($k = 32$) | DenseNet-201($k = 32$) | DenseNet-161($k = 48$) |
|---|---|---|---|---|---|
| Convolution | $112 \times 112$ | $7 \times 7$ conv, stride 2 | | | |
| Pooling | $56 \times 56$ | $3 \times 3$ max pool, stride 2 | | | |
| Dense Block (1) | $56 \times 56$ | $\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 6$ | $\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 6$ | $\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 6$ | $\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 6$ |
| Transition Layer (1) | $56 \times 56$ | $1 \times 1$ conv | | | |
| | $28 \times 28$ | $2 \times 2$ average pool, stride 2 | | | |
| Dense Block (2) | $28 \times 28$ | $\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 12$ | $\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 12$ | $\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 12$ | $\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 12$ |
| Transition Layer (2) | $28 \times 28$ | $1 \times 1$ conv | | | |
| | $14 \times 14$ | $2 \times 2$ average pool, stride 2 | | | |
| Dense Block (3) | $14 \times 14$ | $\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 24$ | $\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 32$ | $\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 48$ | $\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 36$ |
| Transition Layer (3) | $14 \times 14$ | $1 \times 1$ conv | | | |
| | $7 \times 7$ | $2 \times 2$ average pool, stride 2 | | | |
| Dense Block (4) | $7 \times 7$ | $\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 16$ | $\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 32$ | $\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 32$ | $\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 24$ |
| Classification Layer | $1 \times 1$ | $7 \times 7$ global average pool | | | |
| | | 1000D fully-connected, softmax | | | |

Çizelge 3.2'de örnek bir ImageNet veri tabanı üzerinde $224 \times 224$ boyutunda giriş imajları için 4 dense bloktan oluşan DenseNet mimarisinin kullanılması gösterilmiştir. Başlangıçtaki konvolüsyon katmanı, 2 adım kayma gerçekleştiren $7 \times 7$ boyutunda $2k$ konvolüsyon içermektedir. Diğer bütün katmanlardaki özellik haritalarının sayısı $k$'ya göre belirlenmektedir.

### 3.3.2. Inception-V3

Inception-V3, GoogleNet için bir modül olarak başlatılmış, görüntü analizine ve nesne algılamaya yardımcı olmak için tasarlanmış evrişimsel bir sinir ağı mimarisidir [162]. GoogLeNet ağı, 2014 yılında Google tarafından önerilen bir ESA'dır. Bu model yalnızca ağ parametrelerinin miktarını azaltmakla kalmayıp aynı zamanda ağ derinliğini de artıran Inception ağ yapısını benimser [163].

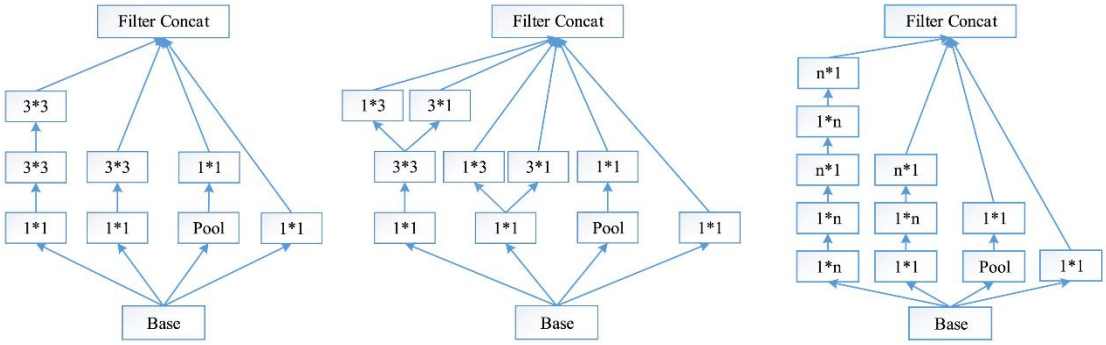
Inception modülü tipik olarak üç farklı boyutta konvolüsyon ve bir maksimum havuzlama içerir. Önceki katmanın ağ çıkışı için, konvolüsyon işleminden sonra kanal toplanır ve ardından doğrusal olmayan füzyon gerçekleştirilir. Bu sayede ağın anlatımı ve farklı ölçeklere uyarlanabilirliği geliştirilebilmekte ve aşırı sığmanın önüne geçilebilmektedir.




Şekil 3.18. Inception-V3 mimarisinin blok diyagramı.

Inception v3, öncelikle Keras tarafından geliştirilen ve ImageNet ile önceden eğitilmiş bir ağ yapısıdır. Varsayılan olarak $299 \times 299$ boyutundaki imajları girdi olarak alır. Tipik bir Inception-V3 ağ yapısı
Şekil 3.18'de gösterilmiştir [163]. Bu modelin temel özellikleri aşağıda belirtilmiştir.

- Çok sayıda sinyal birbirine yakın yerleştirilmiştir. Bu yapı, minimum konvolüsyon oluşturmak için kullanılabilmektedir. Komşu sinyaller birbiriyle ilişkili olduğu için konvolüsyonu uygulamadan önce boyutları küçültme işleminin kayıpsız bir şekilde mümkün olmasını sağlar.
- Başarılı sonuçlar elde etmek için kaynakların serbest ağırlığını arttırırken, aynı zamanda sinir ağının derinliğini ve genişliğini artırmak gerekir.
- Özellikle evrişimli bir sinir ağının ilkinde, parametre değerlerini keskin bir şekilde en aza indiren katmanları kullanmak etkin değildir.
- Geniş katmanlar çok hızlı öğrenir, bu da büyük seviyelerde önemlidir.




Şekil 3.19. Inception-V3 içerisinde bulunan Inception modülleri.

Aynı zamanda Inception-V3, Şekil 3.19'de görüldüğü gibi Inception ağ yapısını $35 \times 35$, $17 \times 17$ ve $8 \times 8$ şeklinde 3 farklı boyuta sahi gridleri kullanarak optimize eder [163].

### 3.3.3. ResNet-50

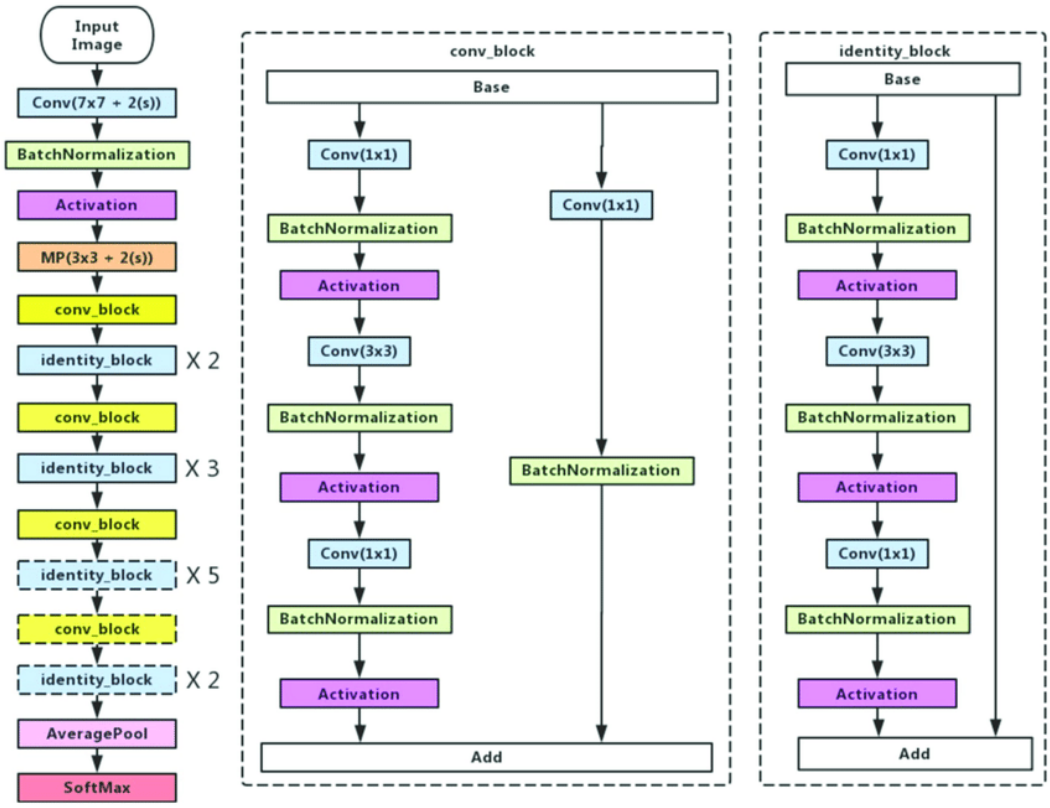
ResNet'in varyasyonlarından biri olan ResNet-50, 50 katmanlı bir ESA modelidir [164]. 48 Konvolüsyon katmanı, 1 MaxPooling katmanı ve 1 Average Pooling katmanı bulunmaktadır. ResNet mimarisi *deep residual learning fremawork* temeline dayanmaktadır. Bu model son derece derin ağlarda bile *vanishing gradient (kaybolan gradyan)* problemlerini çözmektedir [165]. ResNet-50, 23 milyon eğitilebilir katmana sahiptir ve bu sayı mevcut CNN mimarisine göre oldukça küçüktür.

Resnet-50 modelinde fark (residual=artık) değeri Eşitlik 3.14'e göre aşağıdaki gibi hesaplanır.

$$R(x) = Output - Input = H(x) - x \tag{3.14}$$

Girişler, $x$ olarak alınacak olursa gerçek dağılım fonksiyonu $H(x)$ olur. Eşitlik 3.14 yeniden düzenlenecek olursa. Eşitlik 3.15 elde edilmiş olur. ResNet-50 mimari Şekil 3.20'da gösterilmiştir [165].

$$H(x) = R(x) - x \tag{3.15}$$



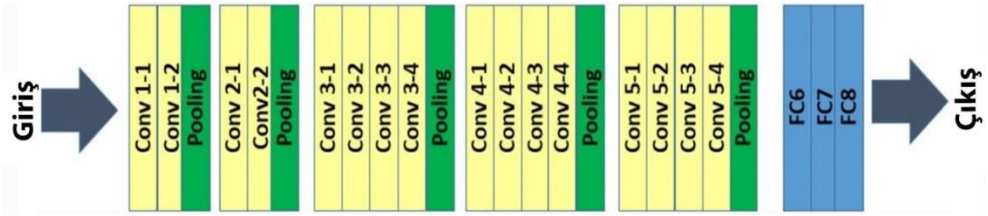

Şekil 3.20. ResNet-50 mimarisinin akış diyagramı.

### 3.3.4. VGG-19

VGG-19, Simonyan ve Zisserman tarafından 2014'de Oxford Üniversitesi'nde VGG-16 modeli geliştirilerek önerilmiş bir Transfer Öğrenme modelidir [166]. VGG (Visual Geometry Group) ImageNet ILSVRC veri seti üzerinde, 1000 sınıftan oluşan 1.3 milyon imaj kullanılarak eğitilmiştir. Bu veri setindeki imajların 100.000 tanesi eğitin için ve 50.000 tanesi ise test için kullanılmıştır. VGG-19, diğer son teknoloji modellere kıyasla sürekli olarak daha iyi performans elde eden VGG mimarisinin 19 derin bağlantılı katman içeren bir varyantıdır.

VGG-19, yüksek bağlantılı evrişimli ve tam bağlantılı katmanlardan oluşan daha iyi özellik çıkarma yeteneğine sahip bir modeldir. Bu modelde, SoftMax aktivasyon

fonksiyonu kullanarak önceki sınıflandırmayı alt örnekleme işlemi için average pooling (ortalama havuzlama) yerine maxpooling kullanılmaktadır.

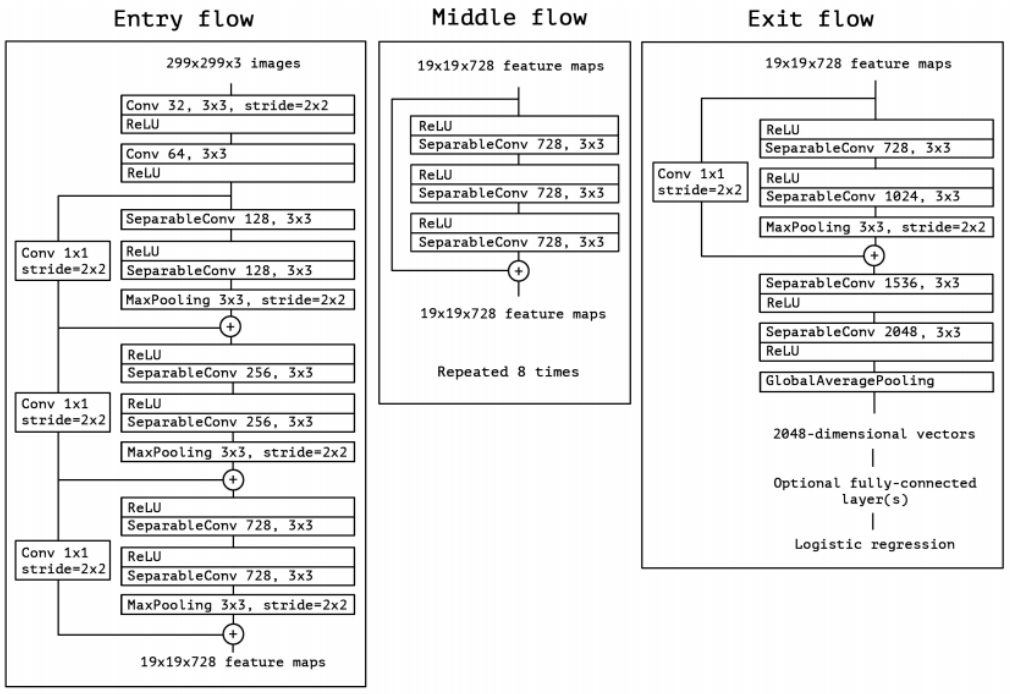



Şekil 3.21.VGG-19 mimarisinin blok diyagramı.

VGG-19 mimarisinin blok diyagramı

Şekil 3.21'de gösterilmiştir [167]. Bu modelde, Max havuzlama, tam bağlı katman, ReLu, Dropout ve Softmax katmanlarından oluşan 41 tane katman bulunmaktadır. VGG-16'ya benzer olarak $224 \times 224 \times 3$ boyutunda RGB imajları girdi olarak alır. VGG-19'da son katman sınıflama katmanı olarak çalışır.

### 3.3.5. Xception




Şekil 3.22. Xception mimarisi.

Xception mimarisi, ağın özellik çıkarma tabanını oluşturan 36 evrişim katmanına sahiptir. 36 evrişim katmanı, ilk ve son modüller hariç, hepsinin çevresinde doğrusal artık bağlantılara sahip olan 14 modül halinde yapılandırılmıştır.

Şekil 3.22'den görüleceği üzere Xception mimarisi 3 ana bölümden oluşmaktadır [168]. Giriş akışı (Entry flow), 8 kez tekrarlanan Orta akış (Middle flow) ve Çıkış Akışı (Exit flow) [168]. Bu modelde ayrılabilir evrişim katmanları kullanılmaktadır.

**Filtre boyutları:**

Geleneksel Evrişim Katmanı $= 3 \times 3 \times 3 \times 64 = 1{,}728$

Ayrılabilir Evrişim Katmanı $= (3 \times 3 \times 1 \times 3) + (1 \times 1 \times 3 \times 64) = 27 + 192 = 219$

Ayrılabilir evrişim katmanları, diğer hesaplama maliyeti ve bellek kullanımı açısından geleneksel evrişim katmanlarına göre çok daha avantajlı olduğu görülmektedir. Bu

modelde tüm evrişim ve ayrılabilir evrişim katmanları *batch normalization* katmanı tarafından takip edilmiştir [168].

# BÖLÜM 4

# VERİ VE METOT

## 4.1. VERİ SETİ VE VERİ ÇOĞALTMA

Ortam Seslerinin sınıflandırılması çalışmalarında genellikle ESC-10, ESC-50 [169] ve UrbanSound8k [170] olmak üzere genel kullanıma açık farklı veri setleri kullanılmaktadır. ESC-50 veri seti, 2000 tane kısa ortam sesinden oluşmaktadır ve bu sesler 50 tane farklı sınıftan oluşan 5 ana kategoriye ayrılmıştır. ESC-10 ise, *dog bark, rain, sea waves, baby cry, clock tick, person sneeze, helicopter, chainsaw, rooster, fire crackling* seslerinden oluşan 10 farklı sınıfa ait 5 farklı kategoride 400 ses kaydı içermektedir [1].

Bu tez kapsamında yapılan çalışmada, ESC-10 veri-seti kullanılmış ve sesler 60 dB'in altında filtrelenerek 2050 Hz frekansla örneklenmiş ve elde edilen bu sinyaller daha sonra [-1, +1] aralığına normalize edilmiştir. Önerilen yöntem dahil tüm imaj dönüşümleri için ses sinyalleri 1024 örnekleme genişliğinde pencerelere bölünmüş ve pencereler arasındaki örtüşme için 512 hop-length uygulanmıştır. Böylece, ardışık gelen pencereler arasında 50% örtüşme sağlanmıştır.

Bu çalışma için iki seviyeli bir veri çoğaltma prosedürü uygulanmıştır. İlk olarak ses sinyallerine zaman alanındaki *shifting positive pitch, shifting negative pitch, stretch time quickly, stretch time slowly, adding white noise* gibi işlemler uygulanmış ve her ses örneği için 5 yeni veri elde edilmiştir [171]. Böylece 400 olan ses örneği 2400'e çıkartılmıştır. İkinci veri çoğaltma yönteminde ise, *TensorFlow* kütüphanesi ile seslerden elde edilen RGB imajlar üzerinde görüntüye dayalı yöntemler uygulanmış ve *rotation, horizontal and vertical shift, brightness, shear, zoom* gibi işlemler gerçekleştirilmiştir [58,169,170].

## 4.2. SES SİNYALLERİNİN GRAFLAR İLE TEMSİLİ

Önerilen grafa dönüştürme metodu orijinal sinyalin her bir penceresini ayrı ayrı işlemektedir. Belirlenmiş komşu pencereler arasındaki normalize edilmiş ve sayısallaştırılmış genlik seviyelerini bu seviyelerin düğümleri ve ardışık komşuları olarak kabul eden bir ağ teorisi yaklaşımı ile düğümler arasında bağlantılar oluşturulur. Sonsuz sayıda katman üretebilme kapasitesine sahip bu yöntem, RGB imajları ile birleştirilmek üzere 3 ağ katmanı üretecek şekilde optimize edilmiştir. Bu eşitlik, çok katmanlı bir ağın formal tanımına bağlı olarak Eşitlik 4.1'deki gibi tanımlanmıştır.

$$\mathcal{M} = (\mathcal{G}, \mathcal{C}) \tag{4.1}$$

Burada $\mathcal{G}$, ağın katmanlarını oluşturmaktadır.

$$\mathcal{G} = \{G_\alpha \,; \alpha = 1,2,3, \dots, M\} \tag{4.2}$$

Bu ifade, ayrıca aşağıdaki şekilde bir graf ailesi olarak tanımlanır;

$$G_\alpha = (V_\alpha, E_\alpha) \tag{4.3}$$

Sonuç olarak;

$$\mathcal{C} = \{E_{\alpha\beta} \subseteq V_\alpha \times V_\beta;\ \alpha, \beta \in [1,2, \dots ., M], \alpha \neq \beta\} \tag{4.4}$$

Eşitlik 4.4, $\alpha \neq \beta$ için $G_\alpha$ gibi $G_\beta$ gibi özel katmanlar içindeki yönlendirilmiş bağlantıları tanımlar [18,23]. Bu çalışmada önerilen yaklaşım, Eşitlik 4.4'de tanımlanan çok katmanlı ağların katmanları arasındaki bağlantı özelliklerini kullanmaz.

### 4.2.1. Normalizasyon ve Sayısallaştırma

Ses sinyallerini de içeren zaman serileri, farklı genlik aralıklarına sahip, orijinal olarak gerçek değerli ayrık sinyallerdir. Ek olarak, bu sinyallerin tepe değerleri, pozitif veya

negatif alternatifler için ortaya çıkabilir. Burada uygulanan normalizasyon prosedürü ilk olarak ayrık bir zaman serisi olan verinin mutlak değerini tespit eder ve bu veriyi [0-1] gibi tepe değeri aralığına eşler.

Belirli bir zaman serisinin ardışık çerçeveleri için sabit tutulan bu değişim aralığı, bu zaman serisinden alınan çerçevelerden üretilen graf temsillerinin bütünlüğünü güvence altına almaktadır. Niceleme prosedürü, öncelikle graf temsillerini tanımlayan bitişik matrislerin satır ve sütun sayısını tanımlayan sabit bir *bit derinliğinin* belirlenmesini gerektirir.

Şekil 4.1'deki Zaman-serisinden grafa dönüştürme prosedürünün gösteriminde 3-bit derinliği kullanılmıştır. 3-bit için elde edilen sayısallaştırma derinliği $2^3 = 8$ olur ve bu derinlik değeri ana fonksiyona *ölçekleme* parametresi olarak gönderilir. Bu bitişik matris temsilleri, *connectogram*'ın tek bir sütunu olacak şekilde daha sonra düzleştirildiğinden dolayı bu gösterim *connectogram*'da sonuç olarak $2^3 \times 2^3 = 64$ boyutunda bir kare-matris şeklinde temsil edilmiş olur.

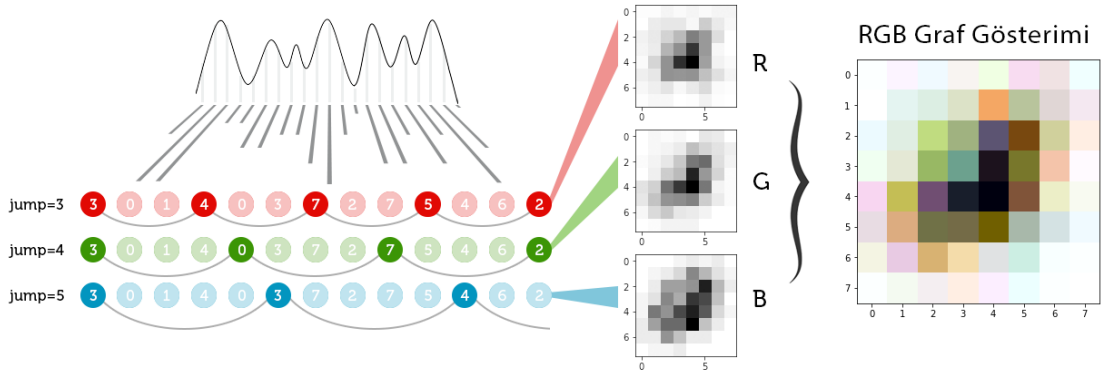
Örnekleme amacıyla bit derinliği değişkenini küçük tuttuktan sonra bu çalışmada söz konusu değişkenin optimal değeri 4 olarak belirlenmiştir. Sonuçta elde edilen graf boyutu $16 \times 16$ ve *connectogram* yüksekliği 256 piksel olacaktır.

### 4.2.2. Zaman Serisinden Grafa Dönüştürme İşlemi

Sayısallaştırma işleminden sonra, ardışık tüm sinyal seviyeleri 0 ve $scale - 1$ ($2^{bit} - 1$) aralığında tam sayılara dönüştürülür. Bu genlik seviyeleri, bir graf temsilinin düğümleri olarak tanımlanırken ardışık düğümler arasındaki komşuluklar graf gösteriminin kenarlarını ifade eder. Ancak, ses sinyalleri çeşitli frekanslara sahip segmentlerden oluştuğu için graflarda kodlanmış değişkenlik bilgileri orijinal sinyale uygulanan *alt örnekleme* oranlarından etkilenir. Bu çalışmada *jump* parametresi olarak etiketlenmiş olan çeşitli *alt örnekleme* oranları, bir ses çerçevesi için farklı graf temsilleri şeklinde

sonuçlandırılır. Bu çalışmada *jump* parametresi 3 farklı sayıdan oluşan bir dizi olarak tanımlanmış ve sonuç olarak 3 farklı bağlantı grafı elde edilmiştir ve bu graflardan her biri için oluşturulan *connectogram,* RGB imajının katmanları olarak ele alınmıştır.

Aynı çerçeveye uygulanan her jump parametresi, ara düğümler göz ardı edilerek Şekil 4.1'deki gibi bir alt örnekleme gerçekleştirilmiş olur [172]. Bu jump parametresinin değerleri, aynı anda üretilen *alt-örnekleme* değerleri için çakışmayı önlemek amacıyla asal sayılardan seçilmiştir.




Şekil 4.1. Zaman Serisinden grafa dönüştürme prosedürü.

*neighDiffThreshold* olarak adlandırılan kenar biçimlendirme ile ilgili bir diğer parametre ise eşikleme işlemi için kullanılır ve default olarak -1 verilmiştir. Bu parametre ile o anki bağlantıyı kurmak üzere komşu genlik seviyeleri arasında minimum farka karar verilir. *neighDiffThreshold* değerinin default olarak -1 seçilmesi ile, eşit değerlere sahip düğümler olsa bile bu düğümlerden her biri diğeri ile bağlanabilir. Böylece komşuluk matrisinin diyagonal elemanlarının boş olmaması sağlanmış olur.

Kenar biçimlendirme ile ilgili son parametre ise *windowSize* olarak adlandırılmıştır. Segmentasyondaki pencereleme parametreleriyle ilgisi bulunmayan bu parametre, birbiri ile bağlantı kuran komşu düğümler arasındaki mesafeyi belirler. Bu parametre değerinin 2 olması, geçerli düğümün uygulanan atlama parametresine göre en yakın 2 komşusuna bağlanacağı anlamına gelir.

Bağlantı prosedürü yönlendirilmiş bir şekilde yürütülür ve böylece sadece ileri yöndeki komşuluklarda bulunan genlik düğümleri arasında bir bağlantı kurulur. Bu işlem, mevcut seviyenin üstünden görülebilmeleri durumunda, devam eden sayısallaştırılmış genlik seviyeleri arasında bağlantılar kurar ve yukarıda bahsedilen *Görünürlük Grafı* yaklaşımına benzer.

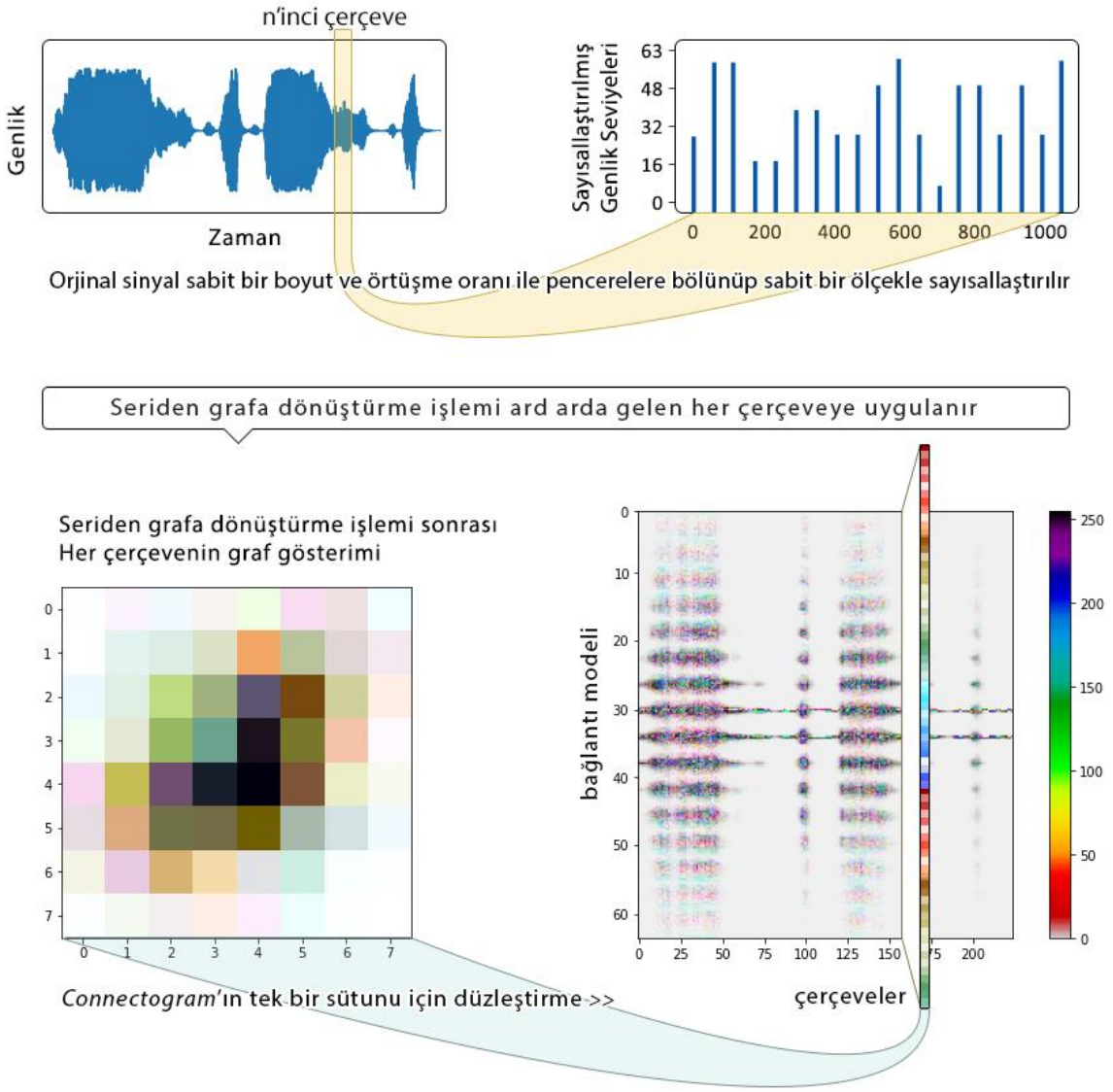
*Görünürlük Grafı* ile *grafik dönüştürme* prosedürü arasındaki temel farklılıklar şunlardır: (i) Değişken alt örnekleme oranları ile komşu düğümler arasında bağlantı kurma ile elde edilen graflar daha sonra RGB imajı formunda birleştirilebilir. (ii) Görünürlük durumunu göz ardı etmek ve düğüm mesafesi içerisindeki düğümleri bağlamak için yalnızca sabit sayıda düğümü atlamayı ele almak. Aşağıda tanımlanan, bu bağlantı şemasının zamana bağımlılığı mevcut çalışma için benzersizdir.

### 4.2.3. Connectogram Formunu Oluşturmak İçin Graf Temsillerinin Birleştirilmesi

Yukarıda açıklanan zaman serisinden grafa dönüştürme işlemi, her bir ses çerçevesi için matris temsillerinin üretilmesinden sorumludur. Bir sonraki adımda ise, bir ses sinyalinin tamamının zamana bağlı gösterimini elde etmek amacıyla her çerçevedeki graf temsillerinin birleştirilmesi işlemi gerçekleştirilir. Böylece *spektrogram* veya *cochleagram* imajları gibi zaman derinliğine sahip bir bağlantı şeması elde edilmiş olur. Bu yöntem, karışık zamansal yapıda olan ve bütün çerçevelerinde aynı içeriğin bulunmadığı farklı ortamlarda kaydedilmiş çevresel seslerin özelliklerini elde çıkarmak için oldukça uyumludur.

Önceden belirlenmiş Pencere uzunluğu ve örtüşme parametrelerine sahip ardışık ses çerçevelerinin RGB graf temsillerini türetme işlemi, kare boyutlu graflar tek bir sütun oluşturacak şekilde düzleştirilerek ve bu sütunlar bir *Connectogram* imajları oluşturacak şekilde dizilerek elde edilmesi aşamaları Şekil 4.2'de gösterilmiştir. Burada, ilk aşamada bir zaman serisi üzerinde örnekleme, sayısallaştırma ve alt pencerelere ayırma işlemleri gerçekleştirilir. Ardından her çerçeve RGB-graf gösterimine dönüştürülür. Bir sonraki

aşamada ise *Connectogram* formunu elde etmek üzere ise bu RGB-graflar düzleştirilir ve yatay olarak dizilir.




Şekil 4.2. Zaman serisinden bir connectogram oluşturma prosedürünün aşamaları.

Bu sonuç imajı, karşılık gelen ses sinyali için zaman (yatay) ve sinyaldeki değişkenlik bilgisi (düşey) olmak üzere iki veriyi temsil etmektedir [172].

Şekil 4.1 ve Şekil 4.2, verileri görsel olarak daha iyi temsil etmek amacıyla parametrelerin düşük değerleri için çizilmiştir.

Yapılan denemelerde, farklı graf kombinasyonları ve derin öğrenme modellerini kıyaslama amacıyla bu parametreler sabit tutulmuştur. Varsayılan bu parametreler ESC-10 veri setindeki 5 saniyelik ses örneklerine uygulanarak $(256 \times 213 \times 3)$ boyutunda imajlar elde edilmiş ve ardından bu imajlar, görüntü işleme yöntemi ile ResNet50 Transfer Learning modeli için en uygun boyut olan $(224 \times 224 \times 3)$ değerlerine ölçeklenmiştir [173]. Söz konusu parametrelerin kullanılan veri seti için en optimum değerleri Çizelge 4.1'de verilmiştir.

Çizelge 4.1. Deneylerde kullanılan en uygun değerleri ile birlikte *connectogram* parametreleri.

| **Parametre** | **Değer** | **Açıklama** |
|---|---|---|
| bit-depth | 4 | Orjinal ses sinyalinden elde edilen çerçeveler, ilk olarak $2^4 \times 2^4$ boyutunda komşuluk matrislerinden oluşan grafa dönüştürülür. |
| windowSize | 2 | Her düğüm, kendinden sonra gelen ileri yöndeki 2 komşu düğüme bağlanır. |
| sr | 22050 | Orijinal ses sinyalinin örnekleme frekansı (Hertz: Hz) |
| jump | [1,3,5] | 3 katmanlı bir graf oluşturacak şekilde alt örneklemeler için atlama değerleri. |
| neighDiffThreshold | -1 | Komşu düğümler arasındaki minimum fark değerini belirler.<br>-1: düğümlerin mevcut değerlerini kabul eder. |
| winLength | 1024 | Ses çerçevelerinin uzunluğu. |
| hopLength | 512 | Ardışık çerçeveler arasındaki uzaklık 1024/2=512 olması 50% örtüşme olmasını sağlar. |

# BÖLÜM 5

# DENEYSEL ÇALIŞMA

Mevcut ESC veri setlerinin kullanıma sunulmasından bu yana, se sınıflandırma çalışmalarında kayda değer bir ilerleme sağlandı. Bu ilerleme daha çok, farklı ses kayıt teknikleri, gürültüden ayıklama gibi çeşitli ön işleme teknikleri, yapay ses kaynakları ile karıştırma, pencereleme ve filtreleme yaklaşımları, YSA gibi farlı sınıflandırma grupları temelli çalışmalara dayanmaktadır. Bu tez kapsamında *connectogram* imajları ile elde edilen sonuçlardan önce ESC-10 veri seti ile bugüne kadar yapılmış çalışmaların bir özeti Çizelge 5.1'de aktarılmıştır.

Çizelge 5.1. ESC-10 veri seti ile ilgili son yıllarda yapılan sınıflandırma çalışmalarında elde edilen sonuçlar ve kısa açıklamaları.

| Yıl | Kaynak | Metot | Başarı (%) |
|---|---|---|---|
| 2019 | [37] | Çevresel seslerin spektrogram imajları, convolutional neural network (CNN) ve tensor deep stacking network (TDSN) derin öğrenme modellerinin eğitimi için kullanılmıştır. | **77.0** |
| 2019 | [11] | mel-spektrogram'lardan COPE özellik çıkarma yöntemi kullanılmıştır ve sınıflandırma için Çok-Sınıflı Support Vector Machines (SVM) tercih edilmiştir. | **81.25** |
| 2019 | [39] | AecNet mimarisinde Multi-Scale CNN için mel-spektrogramlar girdi olarak verilmiştir. | **84.9** |
| 2019 | [41] | Spektrogram imajlarına çoklu-çözünürlük yaklaşımı uygulanmış ve sınıflandırma aşamasında ise TimeScaleNet mimarisi kullanılmıştır. | **69.71** |
| 2019 | [42] | Özel bir CNN mimarisi için Log-gammatone spektrogramlar girdi olarak verilmiştir. | **94.2** |
| 2020 | [35] | Gürültüden arındırılmış sinyaller STFT yöntemi ile spektrogramlara dönüştürülmüş, ardından pyramidal fashion ile özellik çıkarma işlemi için VGGNet16, VGGNet19 ve DenseNet201modelleri kullanılmıştır. Son aşamada ise sınıflandırma için SVM tercih edilmiştir. | **94.8** |
| 2020 | [43] | MFCC, GFCC, CQT ve Chromagram yöntemleri ile elde edilen özellikler birleştirilmiş ve CNN sınıflandırıcıya çok-kanallı girdi olarak verilmiştir. | **97.25** |
| 2020 | [48] | Özellik çıkarma işlemi için Optimum Allocation Sampling (OAS) temelli deneysel bir metot kullanılmış, sınıflandırma aşamasında ise Multi-Class Least Squares Support Vector Machine (MC-LS-SVM) tercih edilmiştir. | **87.25** |

Çizelge 5.2. (devam ediyor).

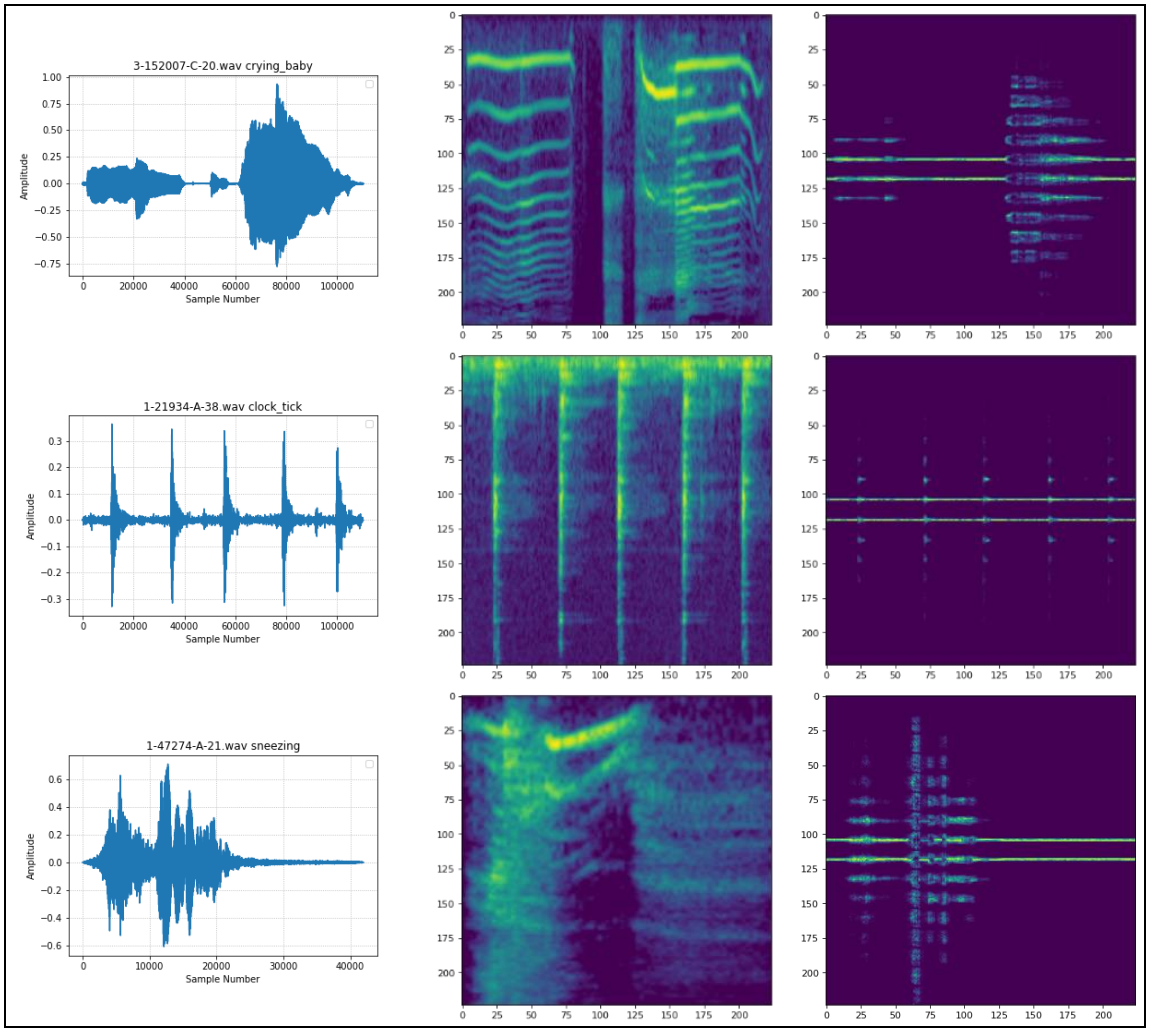
| | | | |
|---|---|---|---|
| 2020 | [52] | Instance-Specific Hidden Markov Models (ISHMMs) temelli Support Vector Machine (SVM) sınıflayıcısı kullanılmıştır. | **74.0** |
| 2020 | [53] | Discrete Wavelet Transform (DWT) spektrogram imajları, DNN (GoogleNet) model kullanılarak sınıflandırılmıştır. | **78.26** |
| 2021 | [58] | Ses kliplerine uygulanan anlamlı veri çoğaltma varyasyonları ile ResNet ve DenseNet gibi önceden eğitilmiş 11 tane model kullanılmıştır. | **99.04** |
| 2021 | [36] | 50% örtüşme ile Blackman-Harris pencereleme fonksiyonu kullanılarak oluşturulan mel-spektrogramlar sınıflandırılmıştır. Söz konusu çalışmada hem el yapımı hem de derin öğrenme ile elde edilmiş özellikler kullanılmıştır. | **97.6** |
| 2021 | [59] | ESC'lerdeki sınıf içi tutarsızlık problemlerini çözen bir vurgulama modülü, ResNet ile beraber kullanılmıştır. | **92.16** |
| 2021 | [61] | ESC için self-supervised learning (SSL) temelli bir derin sınıflayıcı kullanılmıştır. Buradaki SSL mekanizması, modeli spektrogram formatındaki verilerden prototip özellikleri etkin bir şekilde öğrenmeye yönlendirir. | **91.67** |
| 2021 | [96] | Ses Eğitimi için uyarlanmış Deep-Co-Training algorithm (DCT) kullanılmış ve Mean Teacher (MT) olarak adlandırılan farklı bir SSL yaklaşımı ile karşılaştırılmıştır. Bu çalışmada veri olarak mel-spektrogramlar kullanılmıştır. | **91.72** |
| *Mevcut Çalışma* | | Mel-spektrogramlar ile farklı kombinasyonlarda hibrit bir gösterim olarak kullanılmak üzere yeni bir zaman serisi grafı önerildi. Çıkış imajları, Transfer Learning mimarisi olan ResNet50 için giriş olarak kullanıldı. | **95,59** |

Çizelge 5.1'de görüldüğü gibi son yıllarda yapılan çalışmaların çoğu, short time power spektrum kalıbını elde edecek şekilde mel-spektogram imajlarının Mel-Frequency Cepstral Coefficients (MFCC)'lerle birlikte kullanılması temeline dayanmaktadır. *Connectogram*'ların ses sinyallerini temsil kapasitesini belirleyebilmek için mel-spektrogram ve MFCC'lerden elde edilen özellik imajları ile karşılaştırma yapılmış, ayrıca daha iyi sonuçlar elde edebilmek amacıyla bunların kombinasyonları ile kullanılması üzerinde durulmuştur.

## 5.1. Mel-spektrogram İmajları ile Karşılaştırmalı Olarak Verilen Örnek Connectogram İmajları

*Connectogram*, esas olarak genlik seviyelerindeki sapmaları yakalar ve böylece bir sinyalin zaman alanı karakteristiği olarak değerlendirilebilir. Ancak bu sapmanın frekans karakteristiği ile ilişkisi nedeniyle *Connectogram* grafların frekans alanından da temellere sahip olduğu söylenebilir. Ses sinyallerinden elde edilen örnek *Connectogram* imajların

*mel-spektrogram* imajlarla bir aradaki temsili, bu gösterimin orijinal sinyalin özelliklerine nasıl benzediğine dair bir referans sağlamak üzere
Şekil 5.1'de gösterilmiştir [172].

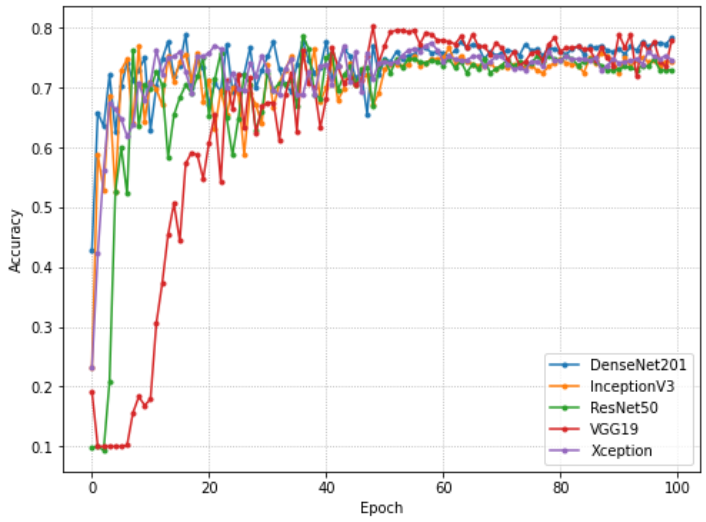



Şekil 5.1. Ses sinyallerinden elde edilen örnek connectogram imajları, aynı sinyalin mel-spektrogram imajları ile gösterilmiştir. Soldaki sütunda ses sinyallerinin zaman alanındaki grafikleri verilmiştir. Ortadaki sütunda aynı sinyallerin mel-spektrogram imajları ve sağdaki sütunda ise bu sinyallerin connectogram temsilleri gösterilmiştir.

Şekil 5.1'de görüleceği üzere spektrogram imajlarından farklı olarak, *connectogram* grafları sinyal zarfına bağlı genlik tabanlı bilgileri içermektedir. Bu bilgiler, sinyaldeki değişkenlik bilgisine bağlı olarak parlaklık ve renklerle temsil edilmiştir. Daha yüksek

yoğunluklu bazı yatay çizgiler, ses çerçevelerinin düzleştirilmemiş graf gösteriminin diyagonal baskınlığı tarafından yönlendirilmiştir. Daha koyu pikseller ise komşuluk matrisinin diyagonal çizgisine olan uzaklığına bağlı olarak ortaya çıkmaktadır ve genellikle daha düşük yoğunluğu temsil etmektedir.

Şekil 5.2'de *connectogram* imajların herhangi başka bir yöntemle birleştirilmeden bağımsız olarak kullanılması ile elde edilen sonuçlar gösterilmiştir. Bu yöntemde 5-fold çapraz doğrulama ile elde edilen en iyi sonuçlar: DenseNet201: 78.75%, InceptionV3: 77.08%, ResNet50: 78.75%, VGG19: 80.00%, Xception: 77.50% olarak tespit edilmiştir. Bu sonuçlardan *connectogram* temelli imajların çok-sınıflı seslerin ayrıştırılmasında ayırt edici özellikler sağladığı görülmektedir. Bu sonuçların Çizelge 5.1'de gösterilen günümüzdeki rakiplerine göre düşük olmasına karşılık 2019'daki spektrogram temelli sonuçlara oldukça yakın olduğu görülmektedir [172].




Şekil 5.2. 5 farklı Transfer Learning modeli için connectogram imajların sınıflandırma başarıları. En yüksek başarı VGG19 TM modeli ile elde edildiği ve 80%'i aşmadığı görülmektedir.

Bu çalışmada mevcut sınıflandırma teknolojilerini de dikkate alarak *connectogram* imajlarını, mel-spektrogramlar ve MFCC'lerin farklı kombinasyonları ile birlikte kullandık. Bu kombinasyonun temel özelliği, aynı pencereleme parametrelerini

kullanarak elde edilen mel-spektrogram, MFCC ve *connectogram* imajlarının aynı ses sinyalinden alınan aynı temsil şemalarını göstermesidir. Böylece bu görüntüleri bir RGB görüntüsünün farklı katmanları olarak birleştirmek, bu sınıflandırma çalışmasında dikkate değer bir husustur.

Çalışmada, öncelikle en iyi sonucun elde edildiği ResNet50 TÖ modeli kullanılarak farklı parametreler ile oluşturulan grafların sınıflandırma performansı değerlendirilmiştir. Bu amaçla *bit-depth*, *jump* ve *windowSize* gibi parametrelerin bütün kombinasyonları denenmiş ve en iyi sonucun hangi parametrelerle elde edildiği bulunmaya çalışılmıştır. Bu aşamada sadece Fold-1 test için kullanıldığında Çizelge 5.3'de verilen sonuçlar elde edilmiştir. Bu kapsamda; öncelikle mel-spektrogramlar ve connectogramlar tek katman oluşturmak amacıyla gri imajlara dönüştürülmüş, ardından bu imajlar [mels, mels, conn] şeklinde birleştirilerek RGB katmanları elde edilmiştir. Bu çizelgeye göre en optimum parametreler; *bit-depth*: 4, *jump*: [1,3,5], *windowSize*:2 olarak belirlenmiştir. Bu değerler ile elde edilen en iyi sınıflandırma başarısının test için fold-1 kulalnıldığında **96.46%** olduğu tespit edilmiştir.

Çizelge 5.3. Mel-spektrogramlar (mels) ve connectogramlar (conn) kullanılarak yapılan çalışmalarda connectogram grafları üretmek için kullanılan farklı parametrelerin sınıflandırma başarısına etkisi.

| | | **windowSize** | | |
|---|---|---|---|---|
| **bit-depth** | **jump** | **1** | **2** | **3** |
| 3 | [1,3,5] | 92.50 | 95.42 | 92.08 |
| | [3,5,7] | 93.75 | 94.79 | 94.58 |
| | [5,7,11] | 92.08 | 93.75 | 91.67 |
| 4 | [1,3,5] | 93.54 | **96.46** | 94.58 |
| | [3,5,7] | 93.13 | 95.42 | 93.75 |
| | [5,7,11] | 96.04 | 94.37 | 93.75 |
| 5 | [1,3,5] | 92.92 | 91.46 | 94.58 |
| | [3,5,7] | 96.06 | 94.75 | 93.75 |
| | [5,7,11] | 95.20 | 93.33 | 93.96 |

Mel-spektrogramlar (mels) ve *connectogramlar* (conn) kullanılarak yapılan çalışmalarda *connectogram* grafları üretmek için kullanılan farklı parametrelerin sınıflandırma

sonuçlarına etkisi. Bu kapsamda; öncelikle mel-spektrogramlar ve *connectogramlar* tek katman oluşturmak amacıyla gri imajlara dönüştürülmüş, ardından bu imajlar [mels, mels, conn] şeklinde birleştirilerek RGB katmanları elde edilmiştir.

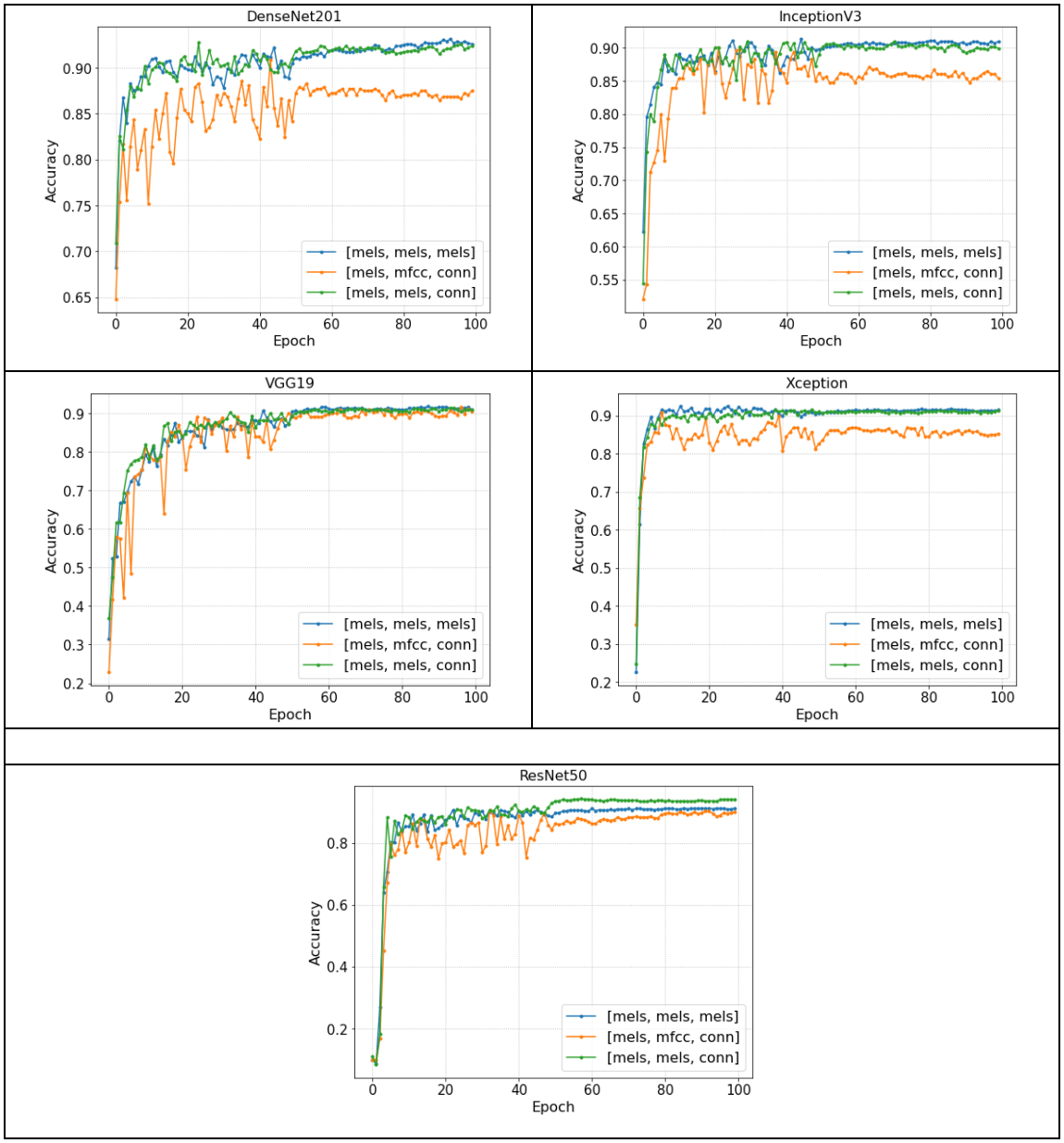
Bu çalışmanın devamında graf oluşturma için kullanılan parametreler, yukarıda bulunan en optimum değerlerde sabitlenmiş ve böylece elde edilen RGB temsilleri, aynı sabit hiper-parametrelerle tasarlanmış farklı TÖ modelleri ile test edilmiştir. Her RGB kombinasyonu için 5-fold çapraz doğrulama ile sınıflandırma işlemi gerçekleştirilmiş ve elde edilen en iyi sonuçlar Çizelge 5.4'de gösterilmiştir. Bu amaçla tek katman elde etmek için her temsil imajı önce gri seviye imajına dönüştürülür ve bu katmanlar sonuçta RGB imajı olarak birleştirilir. Çizelgede de görüldüğü üzere mel-spektrogram ve *connectogram* kombinasyonu mel-spektrogram gösterimin tek başına kullanılmasına göre sınıflandırma başarısını önemli ölçüde arttırmaktadır.

Çizelge 5.4. mel-spektrogram (mels), mel-frequency cepstral coefficients (mfcc) ve *connectogram* (conn) gösterimlerinin farklı kombinasyonları için 5 farklı Transfer Öğrenme modelinde elde edile sınıflandırma başarıları.

| **Model** | **[mels, mels, mels]** | **[mels, mfcc, conn]** | **[mels, mels, conn]** |
|---|---|---|---|
| **ResNet50** | 93,79 % | 90.63 % | **95,59 %** |
| **DenseNet201** | 94,37 % | 90.83 % | 94,75 % |
| **InceptionV3** | 93,33 % | 90.00 % | 92,92 % |
| **VGG19** | 92,79 % | 91.67 % | 92,79 % |
| **Xception** | 94,04 % | 90.83 % | 93,63 % |

Bu gelişme, yatay (zaman) eksende temsil imajlarının senkron olması nedeni ile farklı temsil yöntemlerinin uygunluğunun bir göstergesidir. Dikey eksende ise mel-spektrogram için frekans ve connectogram için sinyaldeki değişkenlik ölçüsü birbirlerini tamamlayıcı olarak temsil edilmektedirler. Mel-spektrogram, MFCC ve connectogram temsillerinin kombinasyonun mel-spektrogram imajlarının yalnız kullanılması ile elde edilen sonuçları arttırmadığı görülmektedir. Fakat mel-spektrogram ve *connectogram* kombinasyonunun

sınıflandırma başarısını diğer Derin Öğrenme modellerinde de önemli ölçüde arttırdığı tespit edilmiştir.




Şekil 5.2. Mel-spektrogram, MFCC ve connectogram temsillerinin hibrit olarak kullanıldığı imajlarla 5 farklı Transfer Öğrenme modeli (ResNet50, DenseNet201, InceptionV3, Vgg19, Xception) için örnek Accuracy grafikleri.

.

Uygulanan ilk aşamasında mel-spektrogram temsilinin yalnız kullanılması durumunda elde edilen imajlar ile transfer öğrenme modelleri arasında VGG19 modelinin en başarılı sonucu verdiği görülmektedir. Bununla beraber bütün seçenekler için 5-fold çapraz doğrulama kullanılarak yapılan denemelerde; **[*mels, mels, conn*]** kombinasyonunun **ResNet50** TÖ modeli ile en başarılı sonucu verdiği tespit edilmiş ve **95.59 %** oranında bir sınıflandırma başarısı elde edilmiştir. Belirlenen modeller için aynı TÖ parametreleri ile 100 epoch döngünün her seferinde farklı bir fold'u test verisi olarak belirleyip 5 kez çalışması sonucu elde edilen sınıflandırma başarılarının ortalamaları Çizelge 5.4**'**de gösterilmiştir. Aynı sonuçların Accuracy grafikleri ise Şekil 5.2'de gösterilmiştir.

Bu sonuçları kısaca Çizelge 5.1'de belirtilen mevcut en iyi çalışmalarla karşılaştırdığımızda elde edilen sınıflandırma performansının önerilen hibrit temsil yöntemleri sayesinde son yıllarda gerçekleştirilen aynı alandaki çalışmalarla rekabet edebilecek başarıyı sağladığını söylemek mümkündür. Çizelge 5.1'de bahsedilen daha başarılı sonuçlar, el yapımı özellik kümeleri, gelişmiş veri çoğaltma teknikleri veya çok-seviyeli sınıflandırıcılar gibi bazı yardımcı yöntemlere dayanmaktadır.

# BÖLÜM 6

# SONUÇLAR VE ÖNERİLER

## 6.1. SONUÇLAR

Bu tez çalışması kapsamınsa seslerin akustik özelliklerinin çıkartılması için yeni bir yöntem geliştirilmiştir. Karmaşık ağlara dayalı bu yöntemde ses sinyallerinin zaman alanındaki gösterimlerinin genlik seviyeleri *görünürlük grafı* olarak adlandırılan belli varyasyonlarla bir ağ yapısına dönüştürülmüştür.

Görünürlük grafları elde edilirken öncelikle 22.050 Hz frekansında örneklenen ses sinyalleri, 40 ms'lik çerçevelere bölünmüştür. Elde edilen bu çerçevelerden en iyi sınıflandırma performansını verecek görünürlük grafları elde etmek için farklı parametreleri ile denemeler yapılmış ve en iyi sonucun *bit-depth*: 4, *jump*: [1,3,5], *windowSize*: 2 kombinasyonu ile elde edildiği tespit edilmiştir.

Sonraki aşamada yukarıdaki parametreler ile oluşturulan graflar [*mels, mels, conn*] şeklinde 3-katmanlı RGB imajı olarak transfer öğrenme yönteminin farklı modellerinde test edilmiştir. Yapılan sınıflandırma çalışmalarında en başarılı sonucun [*mels, mels, conn*] kombinasyonu kullanılarak **ResNet50** transfer öğrenme modeli üzerinde 5-fold çapraz doğrulama ile **95.59 %** oranında bir sınıflandırma başarısı elde edildiği görülmüştür. Aynı veri seti ve aynı öğrenme modelleri ile [*mels, mels, mels*] kullanıldığında **94.37 %** oranında bir sınıflandırma başarısı elde edilmiştir.

Ayrıca, sadece graflardan oluşan [*conn, conn, conn*] imajları ile yapılan denemelerde 80.00% oranında bir sınıflandırma başarısı elde edilmiş ve 2019'daki sonuçlara oldukça yakın olduğu görülmüştür.

## 6.2. ÖNERİLER

Bu tez çalışması kapsamında sadece 3 farklı adım sayısı kullanılarak yapılan alt örnekleme, 3'ten fazla sayıda uygulanarak çok katmanlı (örneğin 4 katman ile CMYK kodlu imajlar) imaj gösterimleri elde edilebilir. Sinyallere uygulanan pencereleme parametreleri ve fonksiyonu değiştirilerek sınıflandırma üzerindeki etkileri incelenebilir. Ayrıca sinyallere dönüşüm öncesinde uygulanacak çeşitli filtrelerin sınıflandırma performansına etkileri de incelenebilir.

Bunlara ek olarak sesin zaman alanı üzerinde geliştirdiğimiz ve *connectogram* olarak adlandırılan bu yenilikçi yaklaşımın; EEG sinyalleri, EKG verileri, finansal veriler ve sensörlerden elde edilen veriler gibi zaman serisi kullanılan farklı veri setleri üzerinde de başarılı sonuçlar vereceği düşünülmektedir.

## KAYNAKLAR


1. Zhang Zhichao and Xu, S. and C. S. and Z. S., "Deep Convolutional Neural Network with Mixup for Environmental Sound Classification", (2018).

2. Pourbabaee, B., Roshtkhari, M. J., and Khorasani, K., "Deep Convolutional Neural Networks and Learning ECG Features for Screening Paroxysmal Atrial Fibrillation Patients", ***IEEE Transactions On Systems, Man, And Cybernetics: Systems***, 48 (12): 2095–2104 (2018).

3. Gharehbaghi, A. and Lindén, M., "A Deep Machine Learning Method for Classifying Cyclic Time Series of Biological Signals Using Time-Growing Neural Network", ***IEEE Transactions On Neural Networks And Learning Systems***, 29 (9): 4102–4115 (2018).

4. Bao, W., Yue, J., and Rao, Y., "A deep learning framework for financial time series using stacked autoencoders and long-short term memory", ***PLoS ONE***, 12: (2017).

5. Canizo, M., Triguero, I., Conde, A., and Onieva, E., "Multi-head CNN–RNN for multi-time series anomaly detection: An industrial case study", ***Neurocomputing***, 363: 246–260 (2019).

6. Soares, E., Costa, P., Costa, B., and Leite, D., "Ensemble of evolving data clouds and fuzzy models for weather time series prediction", ***Applied Soft Computing***, 64: 445–453 (2018).

7. Dafna, E., Tarasiuk, A., and Zigel, Y., "Sleep staging using nocturnal sound analysis", ***Scientific Reports***, 8 (1): 13474 (2018).

8. Cao, D., Wang, Y., Duan, J., Zhang, C., Zhu, X., Huang, C., Tong, Y., Xu, B., Bai, J., Tong, J., and Zhang, Q., "Spectral Temporal Graph Neural Network for Multivariate Time-series Forecasting", ***CoRR***, abs/2103.07719: (2021).

9. Zhao, B., Lu, H., Chen, S., Liu, J., and Wu, D., "Convolutional neural networks for time series classification", ***Journal Of Systems Engineering And Electronics***, 28 (1): 162–169 (2017).

10. Karim, F., Majumdar, S., Darabi, H., and Chen, S., "LSTM Fully Convolutional Networks for Time Series Classification", ***IEEE Access***, 6: 1662–1669 (2018).

11. Strisciuglio, N., Vento, M., and Petkov, N., "Learning representations of sound using trainable COPE feature extractors", ***Pattern Recognition***, 92: 25–36 (2019).

12. Sharan, R. v and Moir, T. J., "Cochleagram image feature for improved robustness in sound recognition", (2015).

13. Peng, Z., Dang, J., Unoki, M., and Akagi, M., "Multi-resolution modulation-filtered cochleagram feature for LSTM-based dimensional emotion recognition from speech", ***Neural Networks***, 140: 261–273 (2021).

14. Lacasa, L., Luque, B., Ballesteros, F., Luque, J., and Nuño, J. C., "From time series to complex networks: The visibility graph", ***Proceedings Of The National Academy Of Sciences***, 105 (13): 4972–4975 (2008).

15. Lacasa, L., Nicosia, V., and Latora, V., "Network structure of multivariate time series", ***Scientific Reports***, 5 (1): 15508 (2015).

16. Li, D., Lin, J., Bissyande, T. F. D. A., Klein, J., and le Traon, Y., "Extracting statistical graph features for accurate and efficient time series classification", (2018).

17. Mao, S. and Xiao, F., "Time Series Forecasting Based on Complex Network Analysis", ***IEEE Access***, 7: 40220–40229 (2019).

18. Türker, İ. and Sulak, E. E., "A multilayer network analysis of hashtags in twitter via co-occurrence and semantic links", ***International Journal Of Modern Physics B***, 32 (04): 1850029 (2018).

19. Türker, İ., Şehirli, E., and Demiral, E., "Uncovering the differences in linguistic network dynamics of book and social media texts", ***SpringerPlus***, 5 (1): 864 (2016).

20. Baydilli, Y. Y., Bayir, Ş., and Türker, I., "A Hierarchical View of a National Stock Market as a Complex Network.", ***Economic Computation & Economic Cybernetics Studies & Research***, 51 (1): (2017).

21. Demir, S. and Türker, İ., "Arithmetic success and gender-based characterization of brain connectivity across EEG bands", ***Biomedical Signal Processing And Control***, 64: 102222 (2021).

22. Cai, S. M., Chen, W., Liu, D. B., Tang, M., and Chen, X., "Complex network analysis of brain functional connectivity under a multi-step cognitive task", ***Physica A: Statistical Mechanics And Its Applications***, 466: 663–671 (2017).

23. Zou, Y., Donner, R. v., Marwan, N., Donges, J. F., and Kurths, J., "Complex network approaches to nonlinear time series analysis", ***Physics Reports***, 787: 1–97 (2019).

24. Hamilton, W. L., "Graph Representation Learning", ***Synthesis Lectures On Artificial Intelligence And Machine Learning***, 14 (3): 1–159 (2020).

25. Boddapati, V., Petef, A., Rasmusson, J., and Lundberg, L., "Classifying environmental sounds using image recognition networks", ***Procedia Computer Science***, 112: 2048–2056 (2017).

26. Krizhevsky, A., Sutskever, I., and Hinton, G. E., "Imagenet classification with deep convolutional neural networks", ***Advances In Neural Information Processing Systems***, 25: (2012).

27. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., and Rabinovich, A., "Going deeper with convolutions", (2015).

28. Bagnall, A., Lines, J., Hills, J., and Bostrom, A., "Time-Series Classification with COTE: The Collective of Transformation-Based Ensembles", ***IEEE Transactions On Knowledge And Data Engineering***, 27 (9): 2522–2535 (2015).

29. Middlehurst, M., Large, J., Flynn, M., Lines, J., Bostrom, A., and Bagnall, A., .

30. Karim, F., Majumdar, S., Darabi, H., and Harford, S., "Multivariate LSTM-FCNs for time series classification", ***Neural Networks***, 116: 237–245 (2019).

31. Yamak, P. T., Yujian, L., and Gadosey, P. K., "A Comparison between ARIMA, LSTM, and GRU for Time Series Forecasting", (2019).

32. Hewage, P., Behera, A., Trovati, M., Pereira, E., Ghahremani, M., Palmieri, F., and Liu, Y., "Temporal convolutional neural (TCN) network for an effective weather forecasting using time-series data from the local weather station", ***Soft Computing***, 24 (21): 16453–16482 (2020).

33. Ismail Fawaz, H., Forestier, G., Weber, J., Idoumghar, L., and Muller, P.-A., "Deep learning for time series classification: a review", ***Data Mining And Knowledge Discovery***, 33 (4): 917–963 (2019).

34. Su, Y., Zhang, K., Wang, J., and Madani, K., "Environment Sound Classification Using a Two-Stream CNN Based on Decision-Level Fusion", ***Sensors***, 19 (7): (2019).

35. Demir, F., Turkoglu, M., Aslan, M., and Sengur, A., "A new pyramidal concatenated CNN approach for environmental sound classification", ***Applied Acoustics***, 170: 107520 (2020).

36. Luz, J. S., Oliveira, M. C., Araújo, F. H. D., and Magalhães, D. M. V., "Ensemble of handcrafted and deep features for urban sound classification", ***Applied Acoustics***, 175: 107819 (2021).

37. Khamparia, A., Gupta, D., Nguyen, N. G., Khanna, A., Pandey, B., and Tiwari, P., "Sound Classification Using Convolutional Neural Network and Tensor Deep Stacking Network", ***IEEE Access***, 7: 7717–7727 (2019).

38. Wang, A. and others, "An industrial strength audio search algorithm.", (2003).

39. Tang, G., Liang, R., Xie, Y., Bao, Y., and Wang, S., "Improved Convolutional Neural Networks for Acoustic Event Classification", ***Multimedia Tools And Applications***, 78 (12): 15801–15816 (2019).

40. Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama, S., and Darrell, T., "Caffe: Convolutional architecture for fast feature embedding", (2014).

41. Bavu, É., Ramamonjy, A., Pujol, H., and Garcia, A., "TimeScaleNet: A Multiresolution Approach for Raw Audio Recognition Using Learnable Biquadratic IIR Filters and Residual Networks of Depthwise-Separable One-Dimensional Atrous Convolutions", ***IEEE Journal Of Selected Topics In Signal Processing***, 13 (2): 220–235 (2019).

42. Zhang, Z., Xu, S., Zhang, S., Qiao, T., and Cao, S., "Learning Attentive Representations for Environmental Sound Classification", ***IEEE Access***, 7: 130327–130339 (2019).

43. Sharma, J., Granmo, O.-C., and Goodwin, M., "Environment Sound Classification Using Multiple Feature Channels and Attention Based Deep Convolutional Neural Network.", (2020).

44. Davis, S. and Mermelstein, P., "Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences", ***IEEE Transactions On Acoustics, Speech, And Signal Processing***, 28 (4): 357–366 (1980).

45. Shao, Y., Jin, Z., Wang, D., and Srinivasan, S., "An auditory-based feature for robust speech recognition", (2009).

46. Schörkhuber, C. and Klapuri, A., "Constant-Q transform toolbox for music processing", (2010).

47. Shepard, R. N., "Circularity in judgments of relative pitch", ***The Journal Of The Acoustical Society Of America***, 36 (12): 2346–2353 (1964).

48. Ahmad, S., Agrawal, S., Joshi, S., Taran, S., Bajaj, V., Demir, F., and Sengur, A., "Environmental sound classification using optimum allocation sampling based empirical mode decomposition", ***Physica A: Statistical Mechanics And Its Applications***, 537: 122613 (2020).

49. Hassan, A. R. and Bhuiyan, M. I. H., "Computer-aided sleep staging using Complete Ensemble Empirical Mode Decomposition with Adaptive Noise and bootstrap aggregating", ***Biomedical Signal Processing And Control***, 24: 1–10 (2016).

50. Amado, R. G. and Vieira Filho, J., "Pitch detection algorithms based on zero-cross rate and autocorrelation function for musical notes", (2008).

51. Huang, G.-B., Zhu, Q.-Y., and Siew, C.-K., "Extreme learning machine: a new learning scheme of feedforward neural networks", (2004).

52. Chandrakala, S. and Jayalakshmi, S. L., "Generative model driven representation learning in a hybrid framework for environmental audio scene and sound event recognition", ***IEEE Transactions On Multimedia***, 22 (1): 3–14 (2019).

53. Esmaeilpour, M., Cardinal, P., and Koerich, A. L., "Unsupervised feature learning for environmental sound classification using weighted cycle-consistent generative adversarial network", ***Applied Soft Computing***, 86: 105912 (2020).

54. Akbal, E., "An automated environmental sound classification methods based on statistical and textural feature", ***Applied Acoustics***, 167: 107413 (2020).

55. Kaya, Y., Uyar, M., Tekin, R., and Yldrm, S., "1D-local binary pattern based feature extraction for classification of epileptic EEG signals", ***Applied Mathematics And Computation***, 243: 209–219 (2014).

56. Kuncan, M., Kaplan, K., Minaz, M. R., Kaya, Y., and Ertunç, H. M., "A novel feature extraction method for bearing fault classification with one dimensional ternary patterns", ***ISA Transactions***, 100: 346–357 (2020).

57. Tuncer, T. and Ertam, F., "Neighborhood component analysis and reliefF based survival recognition methods for Hepatocellular carcinoma", ***Physica A: Statistical Mechanics And Its Applications***, 540: 123143 (2020).

58. Mushtaq, Z., Su, S. F., and Tran, Q. V., "Spectral images based environmental sound classification using CNN with meaningful data augmentation", ***Applied Acoustics***, 172: 107581 (2021).

59. Tripathi, A. M. and Mishra, A., "Environment sound classification using an attention-based residual neural network", ***Neurocomputing***, 460: 409–423 (2021).

60. Cances, L. and Pellegrini, T., "Comparison of Deep Co-Training and Mean-Teacher approaches for semi-supervised audio tagging", (2021).

61. Tripathi, A. M. and Mishra, A., "Self-supervised learning for Environmental Sound Classification", ***Applied Acoustics***, 182: 108183 (2021).

62. Zhu, G., Li, Y., and Wen, P. P., "An efficient visibility graph similarity algorithm and its application on sleep stages classification", (2012).

63. Rosenblum, M. G., Pikovsky, A. S., and Kurths, J., "Phase synchronization of chaotic oscillators", ***Physical Review Letters***, 76 (11): 1804 (1996).

64. Demir, S. and Türker, İ., "Arithmetic success and gender-based characterization of brain connectivity across EEG bands", ***Biomedical Signal Processing And Control***, 64: 102222 (2021).

65. Goldberger, A. L., Amaral, L. A. N., Glass, L., Hausdorff, J. M., Ivanov, P. C., Mark, R. G., Mietus, J. E., Moody, G. B., Peng, C.-K., and Stanley, H. E., "PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals", ***Circulation***, 101 (23): e215–e220 (2000).

66. Zyma, I., Tukaev, S., Seleznov, I., Kiyono, K., Popov, A., Chernykh, M., and Shpenkov, O., "Electroencephalograms during mental arithmetic task performance", ***Data***, 4 (1): 14 (2019).

67. Türker, İ., Şehirli, E., and Demiral, E., "Uncovering the differences in linguistic network dynamics of book and social media texts", ***SpringerPlus***, 5 (1): 864 (2016).

68. Türker, İ. and Sulak, E. E., "A multilayer network analysis of hashtags in twitter via co-occurrence and semantic links", ***International Journal Of Modern Physics B***, 32 (04): 1850029 (2018).

69. Watts, D. J., "Small Worlds: The Dynamics of Networks between Order and Randomness", ***Princeton University Press Princeton***, (2000).

70. Newman, M. E. J., "The structure of scientific collaboration networks", ***Proceedings Of The National Academy Of Sciences***, 98 (2): 404–409 (2001).

71. Manshour, P., "Complex network approach to fractional time series", ***Chaos: An Interdisciplinary Journal Of Nonlinear Science***, 25 (10): 103105 (2015).

72. Flanagan, R. and Lacasa, L., "Irreversibility of financial time series: A graph-theoretical approach", ***Physics Letters A***, 380 (20): 1689–1697 (2016).

73. Hou, F.-Z., Wang, J., Wu, X.-C., and Yan, F.-R., "A dynamic marker of very short-term heartbeat under pathological states via network analysis", ***EPL (Europhysics Letters)***, 107 (5): 58001 (2014).

74. Zhang, H., Meng, Q., Liu, M., and Li, Y., "A new epileptic seizure detection method based on fusion feature of weighted complex network", (2018).

75. Long, X., Fonseca, P., Aarts, R. M., Haakma, R., and Foussier, J., "Modeling cardiorespiratory interaction during human sleep with complex networks", ***Applied Physics Letters***, 105 (20): 203701 (2014).

76. Zhu, G., Li, Y., Wen, P. P., and Wang, S., "Analysis of alcoholic EEG signals based on horizontal visibility graph entropy", ***Brain Informatics***, 1 (1): 19–25 (2014).

77. Zou, Y., Donner, R. v., Marwan, N., Donges, J. F., and Kurths, J., "Complex network approaches to nonlinear time series analysis", ***Physics Reports***, 787: 1–97 (2019).

78. Sharma, G., Umapathy, K., and Krishnan, S., "Trends in audio signal feature extraction methods", ***Applied Acoustics***, 158: 107020 (2020).

79. Abeßer, J., "A review of deep learning based methods for acoustic scene classification", ***Applied Sciences***, 10 (6): (2020).

80. Lyon, R. F., "Machine hearing: An emerging field [exploratory dsp]", ***IEEE Signal Processing Magazine***, 27 (5): 131–139 (2010).

81. Mitrović, D., Zeppelzauer, M., and Breiteneder, C., "Features for content-based audio retrieval", Advances in Computers, ***Elsevier***, 71–150 (2010).

82. Al\'\ias, F., Socoró, J. C., and Sevillano, X., "A review of physical and perceptual feature extraction techniques for speech, music and environmental sounds", ***Applied Sciences***, 6 (5): 143 (2016).

83. Chu, W.-T., Cheng, W.-H., Hsu, J. Y.-J., and Wu, J.-L., "Toward semantic indexing and retrieval using hierarchical audio models", ***Multimedia Systems***, 10 (6): 570–583 (2005).

84. Zhang, T. and Kuo, C.-C. J., "Audio content analysis for online audiovisual data segmentation and classification", ***IEEE Transactions On Speech And Audio Processing***, 9 (4): 441–457 (2001).

85. Kedem, B., "Spectral analysis and discrimination by zero-crossings", ***Proceedings Of The IEEE***, 74 (11): 1477–1493 (1986).

86. Saunders, J., "Real-time discrimination of broadcast speech/music", (1996).

87. Panagiotakis, C. and Tziritas, G., "A speech/music discriminator based on RMS and zero-crossings", ***IEEE Transactions On Multimedia***, 7 (1): 155–166 (2005).

88. Al-Hattab, Y. A., Zaki, H. F., and Shafie, A. A., "Rethinking environmental sound classification using convolutional neural networks: optimized parameter tuning of single feature extraction", ***Neural Computing And Applications***, 33 (21): 14495–14506 (2021).

89. Wang, W., Yu, X., Wang, Y. H., and Swaminathan, R., "Audio fingerprint based on spectral flux for audio retrieval", (2012).

90. Firmansyah, M. R., Hidayat, R., and Bejo, A., "Comparison of Windowing Function on Feature Extraction Using MFCC for Speaker Identification", (2021).

91. Choi, S. and Jiang, Z., "Cardiac sound murmurs classification with autoregressive spectral analysis and multi-support vector machine technique", ***Computers In Biology And Medicine***, 40 (1): 8–20 (2010).

92. Khan, M. K. S. and Al-Khatib, W. G., "Machine-learning based classification of speech and music", ***Multimedia Systems***, 12 (1): 55–67 (2006).

93. Fu, Z., Lu, G., Ting, K. M., and Zhang, D., "A survey of audio-based music classification and annotation", ***IEEE Transactions On Multimedia***, 13 (2): 303–319 (2010).

94. Schroeder, M. and Atal, B. S., "Code-excited linear prediction (CELP): High-quality speech at very low bit rates", (1985).

95. Tsau, E., Kim, S.-H., and Kuo, C.-C. J., "Environmental sound recognition with CELP-based features", (2011).

96. Mateo, C. and Talavera, J. A., "Short-time Fourier transform with the window size fixed in the frequency domain", ***Digital Signal Processing***, 77: 13–21 (2018).

97. Hlawatsch, F., Boudreaux-Bartels, G. F., and others, "Linear and quadratic time-frequency signal representations", ***IEEE Signal Processing Magazine***, 9 (2): 21–67 (1992).

98. Allen, J., "Short term spectral analysis, synthesis, and modification by discrete Fourier transform", ***IEEE Transactions On Acoustics, Speech, And Signal Processing***, 25 (3): 235–238 (1977).

99. Ćirić, D., Perić, Z., Nikolić, J., and Vučić, N., "Audio Signal Mapping into Spectrogram-Based Images for Deep Learning Applications", (2021).

100. Virtanen, T., Plumbley, M. D., and Ellis, D., "Computational Analysis of Sound Scenes and Events", ***Springer***, (2018).

101. Sharan, R. v. and Moir, T. J., "Acoustic event recognition using cochleagram image and convolutional neural networks", ***Applied Acoustics***, 148: 62–66 (2019).

102. O'shaughnessy, D., "Speech Communications: Human and Machine (IEEE)", ***Universities Press***, (1987).

103. Cheuk, K. W., Agres, K., and Herremans, D., "The impact of audio input representations on neural network based music transcription", (2020).

104. Patterson, R. D., Robinson, K. E. N., Holdsworth, J., McKeown, D., Zhang, C., and Allerhand, M., "Complex sounds and auditory images", Auditory Physiology and Perception, ***Elsevier***, 429–446 (1992).

105. Slaney, M., "Auditory toolbox", ***Interval Research Corporation, Tech. Rep***, 10 (1998): 1194 (1998).

106. Sharan, R. v and Moir, T. J., "Cochleagram image feature for improved robustness in sound recognition", (2015).

107. Albert, R. and Barabási, A.-L., "Statistical mechanics of complex networks", ***Reviews Of Modern Physics***, 74 (1): 47 (2002).

108. Mata, A. S. da, "Complex networks: a mini-review", ***Brazilian Journal Of Physics***, 50 (5): 658–672 (2020).

109. Barrat, A., Barthelemy, M., and Vespignani, A., "Dynamical Processes on Complex Networks", ***Cambridge University Press***, (2008).

110. Dorogovtsev, S. N., Dorogovtsev, S. N., and Mendes, J. F. F., "Evolution of Networks: From Biological Nets to the Internet and WWW", ***Oxford University Press***, (2003).

111. Bullmore, E. and Sporns, O., "Complex brain networks: graph theoretical analysis of structural and functional systems", ***Nature Reviews Neuroscience***, 10 (3): 186–198 (2009).

112. Zhou, C., Zemanová, L., Zamora, G., Hilgetag, C. C., and Kurths, J., "Hierarchical organization unveiled by functional connectivity in complex brain networks", ***Physical Review Letters***, 97 (23): 238103 (2006).

113. Zhou, C., Zemanová, L., Zamora-Lopez, G., Hilgetag, C. C., and Kurths, J., "Structure–function relationship in complex brain networks expressed by hierarchical synchronization", ***New Journal Of Physics***, 9 (6): 178 (2007).

114. Kantz, H. and Schreiber, T., "Nonlinear Time Series Analysis", ***Cambridge University Press***, (2004).

115. Zhang, J. and Small, M., "Complex network from pseudoperiodic time series: Topology versus dynamics", ***Physical Review Letters***, 96 (23): 238701 (2006).

116. Abarbanel, H. D. I., Brown, R., Sidorowich, J. J., and Tsimring, L. S., "The analysis of observed chaotic data in physical systems", ***Reviews Of Modern Physics***, 65 (4): 1331 (1993).

117. Zhuang, E., Small, M., and Feng, G., "Time series analysis of the developed financial markets' integration using visibility graphs", ***Physica A: Statistical Mechanics And Its Applications***, 410: 483–495 (2014).

118. Zhang, G. P., "Neural networks for classification: a survey", ***IEEE Transactions On Systems, Man, And Cybernetics, Part C (Applications And Reviews)***, 30 (4): 451–462 (2000).

119. Haykin, S., "Neural Networks and Learning Machines, 3/E", ***Pearson Education India***, (2010).

120. Abiodun, O. I., Jantan, A., Omolara, A. E., Dada, K. V., Mohamed, N. A. E., and Arshad, H., "State-of-the-art in artificial neural network applications: A survey", ***Heliyon***, 4 (11): e00938 (2018).

121. Herculano-Houzel, S. and Lent, R., "Isotropic fractionator: a simple, rapid method for the quantification of total cell and neuron numbers in the brain", ***Journal Of Neuroscience***, 25 (10): 2518–2521 (2005).

122. Kiranyaz, S., Ince, T., Iosifidis, A., and Gabbouj, M., "Progressive Operational Perceptrons", ***Neurocomputing***, 224: 142–154 (2017).

123. Cochocki, A. and Unbehauen, R., "Neural Networks for Optimization and Signal Processing", ***John Wiley & Sons, Inc.***, (1993).

124. Xu, A., Chang, H., Xu, Y., Li, R., Li, X., and Zhao, Y., "Applying artificial neural networks (ANNs) to solve solid waste-related issues: A critical review", ***Waste Management***, 124: 385–402 (2021).

125. Elman, J. L., "Finding structure in time", ***Cognitive Science***, 14 (2): 179–211 (1990).

126. Liu, Z. W., Liang, F. N., and Liu, Y. Z., "Artificial neural network modeling of biosorption process using agricultural wastes in a rotating packed bed", ***Applied Thermal Engineering***, 140: 95–101 (2018).

127. Yetilmezsoy, K., Ozkaya, B., and Cakmakci, M., "Artificial intelligence-based prediction models for environmental engineering", ***Neural Network World***, 21 (3): 193 (2011).

128. Kocyigit, Y., Alkan, A., and Erol, H., "Classification of EEG recordings by using fast independent component analysis and artificial neural network", ***Journal Of Medical Systems***, 32 (1): 17–20 (2008).

129. Subasi, A., "EEG signal classification using wavelet feature extraction and a mixture of expert model", ***Expert Systems With Applications***, 32 (4): 1084–1093 (2007).

130. Übeyli, E. D., "Combined neural network model employing wavelet coefficients for EEG signals classification", ***Digital Signal Processing***, 19 (2): 297–308 (2009).

131. Orhan, U., Hekim, M., and Ozer, M., "EEG signals classification using the K-means clustering and a multilayer perceptron neural network model", ***Expert Systems With Applications***, 38 (10): 13475–13481 (2011).

132. Pitts, W. and McCulloch, W. S., "How we know universals the perception of auditory and visual forms", ***The Bulletin Of Mathematical Biophysics***, 9 (3): 127–147 (1947).

133. Oğulata, S. N., Şahin, C., and Erol, R., "Neural network-based computer-aided diagnosis in classification of primary generalized epilepsy by EEG signals", ***Journal Of Medical Systems***, 33 (2): 107–112 (2009).

134. Hazarika, N., Chen, J. Z., Tsoi, A. C., and Sergejew, A., "Classification of EEG signals using the wavelet transform", ***Signal Processing***, 59 (1): 61–72 (1997).

135. Sharma, S., Sharma, S., and Athaiya, A., "Activation functions in neural networks", ***Towards Data Science***, 6 (12): 310–316 (2017).

136. Acharya, U. R., Oh, S. L., Hagiwara, Y., Tan, J. H., and Adeli, H., "Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals", ***Computers In Biology And Medicine***, 100: 270–278 (2018).

137. Montavon, G., Samek, W., and Müller, K.-R., "Methods for interpreting and understanding deep neural networks", ***Digital Signal Processing***, 73: 1–15 (2018).

138. Lecun, Y., Bottou, L., Bengio, Y., and Haffner, P., "Gradient-based learning applied to document recognition", ***Proceedings Of The IEEE***, 86 (11): 2278–2324 (1998).

139. Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., and others, "Imagenet large scale visual recognition challenge", ***International Journal Of Computer Vision***, 115 (3): 211–252 (2015).

140. Paoletti, M. E., Haut, J. M., Plaza, J., and Plaza, A., "A new deep convolutional neural network for fast hyperspectral image classification", ***ISPRS Journal Of Photogrammetry And Remote Sensing***, 145: 120–147 (2018).

141. Licciardi, G. A. and del Frate, F., "Pixel unmixing in hyperspectral data by means of neural networks", ***IEEE Transactions On Geoscience And Remote Sensing***, 49 (11): 4163–4172 (2011).

142. LeCun, Y., Bengio, Y., and Hinton, G., "Deep learning", ***Nature***, 521 (7553): 436–444 (2015).

143. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y., "Generative adversarial nets", ***Advances In Neural Information Processing Systems***, 27: (2014).

144. Ma, X., Wang, H., and Wang, J., "Semisupervised classification for hyperspectral image based on multi-decision labeling and deep feature learning", ***ISPRS Journal Of Photogrammetry And Remote Sensing***, 120: 99–107 (2016).

145. Khan, A., Sohail, A., Zahoora, U., and Qureshi, A. S., "A survey of the recent architectures of deep convolutional neural networks", ***Artificial Intelligence Review***, 53 (8): 5455–5516 (2020).

146. Cireşan, D., Meier, U., Masci, J., and Schmidhuber, J., "Multi-column deep neural network for traffic sign classification", ***Neural Networks***, 32: 333–338 (2012).

147. Liu, X., Deng, Z., and Yang, Y., "Recent progress in semantic image segmentation", ***Artificial Intelligence Review***, 52 (2): 1089–1106 (2019).

148. Zhang, L., Zhang, L., and Du, B., "Deep learning for remote sensing data: A technical tutorial on the state of the art", ***IEEE Geoscience And Remote Sensing Magazine***, 4 (2): 22–40 (2016).

149. Sarvamangala, D. R. and Kulkarni, R. v, "Convolutional neural networks in medical image understanding: a survey", ***Evolutionary Intelligence***, 1–22 (2021).

150. Ghamisi, P., Plaza, J., Chen, Y., Li, J., and Plaza, A. J., "Advanced spectral classifiers for hyperspectral images: A review", ***IEEE Geoscience And Remote Sensing Magazine***, 5 (1): 8–32 (2017).

151. Hu, F., Xia, G.-S., Hu, J., and Zhang, L., "Transferring deep convolutional neural networks for the scene classification of high-resolution remote sensing imagery", ***Remote Sensing***, 7 (11): 14680–14707 (2015).

152. Yue, J., Zhao, W., Mao, S., and Liu, H., "Spectral–spatial classification of hyperspectral images using deep convolutional neural networks", ***Remote Sensing Letters***, 6 (6): 468–477 (2015).

153. Abdel-Hamid, O., Mohamed, A., Jiang, H., Deng, L., Penn, G., and Yu, D., "Convolutional neural networks for speech recognition", ***IEEE/ACM***

***Transactions On Audio, Speech, And Language Processing***, 22 (10): 1533–1545 (2014).

154. Albawi, S., Mohammed, T. A., and Al-Zawi, S., "Understanding of a convolutional neural network", (2017).

155. Akhtar, N. and Ragavendran, U., "Interpretation of intelligence in CNN-pooling processes: a methodological survey", ***Neural Computing And Applications***, 32 (3): 879–898 (2020).

156. Liu, Y., Zhou, Y., Wen, S., and Tang, C., "A strategy on selecting performance metrics for classifier evaluation", ***International Journal Of Mobile Computing And Multimedia Communications (IJMCMC)***, 6 (4): 20–35 (2014).

157. Chicco, D. and Jurman, G., "The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation", ***BMC Genomics***, 21 (1): 1–13 (2020).

158. Petersen, F., Kuehne, H., Borgelt, C., and Deussen, O., "Differentiable Top-k Classification Learning", (2021).

159. "İnternet: Keras Applications", .

160. Pan, S. J. and Yang, Q., "A survey on transfer learning", ***IEEE Transactions On Knowledge And Data Engineering***, 22 (10): 1345–1359 (2009).

161. Huang, G., Liu, Z., van der Maaten, L., and Weinberger, K. Q., "Densely connected convolutional networks", (2017).

162. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., and Rabinovich, A., "Going deeper with convolutions", (2015).

163. Dong, N., Zhao, L., Wu, C. H., and Chang, J. F., "Inception v3 based cervical cell classification combined with artificially extracted features", ***Applied Soft Computing***, 93: 106311 (2020).

164. He, K., Zhang, X., Ren, S., and others, "Deep Residual Learning", ***Image Recognition***, (2015).

165. Mandal, B., Okeukwu, A., and Theis, Y., "Masked Face Recognition using ResNet-50", ***ArXiv Preprint ArXiv:2104.08997***, (2021).

166. Simonyan, K. and Zisserman, A., "Very deep convolutional networks for large-scale image recognition", ***ArXiv Preprint ArXiv:1409.1556***, (2014).

167. Özyurt, F., "Efficient deep feature selection for remote sensing image recognition with fused deep learning architectures", ***The Journal Of Supercomputing***, 76 (11): 8413–8431 (2020).

168. Chollet, F., "Xception: Deep Learning With Depthwise Separable Convolutions", (2017).

169. Piczak, K. J., "ESC: Dataset for Environmental Sound Classification", (2015).

170. Salamon, J., Jacoby, C., and Bello, J. P., "A Dataset and Taxonomy for Urban Sound Research", (2014).

171. Mushtaq, Z. and Su, S. F., "Environmental sound classification using a regularized deep convolutional neural network with data augmentation", ***Applied Acoustics***, 167: 107389 (2020).

172. Türker, İ. and Aksu, S., "Connectogram – A graph-based time dependent representation for sounds", ***Applied Acoustics***, 191: 108660 (2022).

173. He, K., Zhang, X., Ren, S., and Sun, J., "Deep Residual Learning for Image Recognition", (2016).

# ÖZGEÇMİŞ

Serkan AKSU, Karadeniz Teknik Üniversitesi, Bilgisayar Mühendisliği Bölümü'nden 2000 yılında mezun oldu. Karabük Üniversitesi, Fen Bilimleri Enstitüsü, Bilgisayar Mühendisliği Anabilim Dalı'nda "Bulut Bilişim Teknolojisinin Bilişim Teknolojilerine Etkileri ve Bir Görüntü İşleme Uygulaması" konulu tez çalışması ile 2013 yılında Yüksek Lisans eğitimini tamamladı. 2002'den beri Bartın Üniversitesi, Bartın Meslek Yüksekokulu, Bilgisayar Teknolojileri Bölümü'nde Öğretim Görevlisi olarak çalışmaya devam etmektedir. Serkan AKSU evli ve üç çocuk babasıdır.