

T.C.
FIRAT ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ



**GÖRÜNTÜ MANİPÜLASYONLARININ DERİN ÖĞRENME
YAKLAŞIMI İLE BELİRLENMESİ**

Semih YAVUZKILIÇ

Doktora Tezi

ELEKTRİK ELEKTRONİK MÜHENDİSLİĞİ ANABİLİM DALI

Elektrik Elektronik Mühendisliği Teknolojileri Programı

Telekomünikasyon

OCAK 2022

T.C.
FIRAT ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

Elektrik Elektronik Mühendisliği Anabilim Dalı
Elektrik Elektronik Mühendisliği Teknolojileri Programı

Doktora Tezi

GÖRÜNTÜ MANİPÜLASYONLARININ DERİN ÖĞRENME YAKLAŞIMI İLE BELİRLENMESİ

Tez Yazarı

Semih YAVUZKILIÇ

Danışman

Prof. Dr. Abdulkadir ŞENGÜR

OCAK 2022

ELAZIĞ

T.C.
FIRAT ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

Elektrik Elektronik Mühendisliği Anabilim Dalı
Elektrik Elektronik Mühendisliği Teknolojileri Programı

Doktora Tezi

Başlığı:	Görüntü Manipülasyonlarının Derin Öğrenme Yaklaşımı İle Belirlenmesi
Yazarı:	Semih YAVUZKILIÇ
İlk Teslim Tarihi:	16.12.2021
Savunma Tarihi:	21.01.2022

TEZ ONAYI

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

Danışman:	Prof. Dr. Abdulkadir ŞENGÜR Fırat Üniversitesi, Teknoloji Fakültesi	<i>İmza</i> Onayladım
Başkan:	Doç. Dr. Erkan DENİZ Fırat Üniversitesi, Teknoloji Fakültesi	Onayladım
Üye:	Doç. Dr. Çetin GENÇER Fırat Üniversitesi, Teknoloji Fakültesi	Onayladım
Üye:	Doç. Dr. Ömer Faruk ALÇİN Malatya Turgut Özal Üniversitesi, Mühendislik ve Doğa Bilimleri Fakültesi	Onayladım
Üye:	Dr. Öğr. Üyesi Muzaffer ASLAN Bingöl Üniversitesi, Mühendislik-Mimarlık Fakültesi	Onayladım

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

İmza
Prof. Dr. Kürşat Esat ALYAMAÇ
Enstitü Müdürü

BEYAN

Fırat Üniversitesi Fen Bilimleri Enstitüsü tez yazım kurallarına uygun olarak hazırladığım “Görüntü Manipülasyonlarının Derin Öğrenme Yaklaşımı İle Belirlenmesi” Başlıklı Doktora Tezimin içindeki bütün bilgilerin doğru olduğunu, bilgilerin üretilmesi ve sunulmasında bilimsel etik kurallarına uygun davrandığımı, kullandığım bütün kaynakları atıf yaparak belirttiğimi, maddi ve manevi desteği olan tüm kurum/kuruluş ve kişileri belirttiğimi, burada sunduğum veri ve bilgileri unvan almak amacıyla daha önce hiçbir şekilde kullanmadığımı beyan ederim.

21.01.2022

Semih YAVUZKILIÇ



ÖNSÖZ

Makine öğrenmenin bir dalı olan derin öğrenme yöntemleri popülerliği gün geçtikçe artan uygulamalardan biri olmuştur. Günümüzde bu teknolojilerin farklı uygulamaları mevcuttur. Bu uygulamaların çoğu insanlık yararına kullanılıyor olmasına rağmen bu uygulamalardan bazıları insanlık için tehdit veya tehlike arz etmektedir. Tehlikeli olduğu düşünülen uygulamalardan birisi görüntü manipülasyonu tekniklerinden biri olan deepfake'dir. Bu tez çalışmasında, görüntü manipülasyonlarını tespit yöntemlerini geliştirmek için derin öğrenme tabanlı mimarilerden olan Evrişimsel Sinir Ağlarının farklı yapıları kullanılarak sınıflandırma problemleri çözülmeye çalışılmış, yeni bir veri seti oluşturulmuş ve manipülasyon tespiti için yeni yaklaşımlar geliştirilmiştir.

Tez çalışmalarımın her sürecinde bana yol gösteren, önüme çıkan her sorunda sihirli bir dokunuşla çözüm üreten, her türlü zorlukta varlığını hissettiren, hayata ve bilime bakış açımı değiştiren, hem örnek kişiliği hem de akademik karakteriyle ufkumu açan değerli danışman hocam Prof. Dr. Abdulkadir ŞENGÜR'e sonsuz teşekkürlerimi sunarım.

Tez çalışmaları boyunca ilgilerini ve değerli bilgilerini paylaşan Doç. Dr. Erkan DENİZ'e ve Doç. Dr. Murat KARABATAK'a katkılarından dolayı teşekkür ederim. Tezin hazırlanması aşamasında yardımlarını esirgemeyen her zaman yanımda olan değerli arkadaşlarım Bube TURAN ve Ahmet Erdem UĞUR'a teşekkür ederim.

Hayatımın her anında yanımda olan, her zaman ve her şartta beni destekleyen başta ilham kaynağım olan annem, babam olmak üzere tüm kardeşlerime ve aileme teşekkürü bir borç bilirim.

Hayatıma yeni anlamlar katan, hiçbir desteğini esirgemeyen her zaman yanımda olan, moral ve motivasyon kaynağım sevgili eşim Yasemin'e en içten teşekkürlerimi sunarım.

Semih YAVUZKILIÇ
ELAZIĞ, 2022

İÇİNDEKİLER

	Sayfa
ÖNSÖZ.....	iv
İÇİNDEKİLER	v
ÖZET	vii
ABSTRACT	viii
ŞEKİLLER LİSTESİ	ix
TABLolar LİSTESİ	xi
1. Giriş	1
1.1. Tezin Literatüre Katkısı.....	4
1.2. Tezin Organizasyonu	4
2. GÖRÜNTÜ MANİPÜLASYONLARI.....	5
2.1. Yüz Manipülasyon Kategorileri	5
2.1.1. Yüz Sentezi.....	5
2.1.2. Yüz Değiştirme.....	8
2.1.3. Yüz Özellikleri	10
2.1.4. Yüz İfadeleri.....	11
2.2. Deepfake Manipülasyonlarının Tespiti.....	13
2.2.1. Doku Tabanlı Yöntemler	13
2.2.2. Doğal Özellik Bozulmaları Tabanlı Yöntemler	14
2.2.3. Derin Öğrenme Tabanlı Yöntemler	14
3. ÖZNİTELİK ÇIKARMA VE SINIFLANDIRMA AŞAMALARINDA KULLANILAN YÖNTEMLER	21
3.1. Yapay Sinir Ağları (YSA)	21
3.1.1. Tek Katmanlı YSA	22
3.1.2. Çok Katmanlı YSA.....	22
3.1.3. İleri Beslemeli YSA.....	23
3.1.4. Geri Beslemeli YSA	23
3.2. YSA Öğrenme Yöntemleri	24
3.2.1. Derin İnanç Ağları	25
3.2.2. Kısıtlı Boltzmann Makineleri	26
3.2.3. Tekrarlı YSA	27
3.3. Evrimsel Sinir Ağları (ESA).....	28
3.3.1. ESA Mimarisi	33
3.4. Otomatik Kodlayıcılar	41
3.4.1. Seyrek Otomatik Kodlayıcı	41
3.4.2. Gürültü Giderici Otomatik Kodlayıcı	42
3.4.3. Varyasyonel Otomatik Kodlayıcı	43
3.5. ESA'da Kullanılan Optimizasyon Yöntemleri	43
3.5.1. Uyarlanabilir Öğrenme Oranları Kullanan Optimizasyon Yöntemleri	45
3.6. Önceden Eğitilmiş ESA Modelleri	47
3.6.1. Artık Sinir Ağı (ResNet).....	47
3.6.2. Uzun Kısa Süreli Bellek (UKSB)	48
3.6.3. VGGNet.....	49

3.6.4. Transfer Öğrenme	51
3.7. Sınıflandırmada Eğitim ve Test Verileri	52
3.8. Sınıflandırma Performansı Ölçümü	54
4. DERİN ARTIK AĞ VE UKSB MİMARİLERİNİN HİBRİT MODELİ KULLANILARAK DEEPFAKE YÜZ VİDEOLARININ TESPİTİ.....	55
4.1. Giriş	55
4.2. Önerilen Yöntem	55
4.3. Veri Seti.....	56
4.3.1. DeepFakeTIMIT Veri Seti.....	56
4.3.2. Celeb-DF Veri Seti	56
4.4. Deneysel Çalışmalar	58
5. ÇOKLU AKIŞ ESA MODELLERİNİN ÖZELLİKLERİ BİRLEŞTİREREK DEEPFAKE VE YÜZ MANİPÜLASYONLARINI TESPİT ETME	62
5.1. Giriş	62
5.2. Önerilen Yöntem	62
5.3. Veri Seti.....	63
5.3.1. DPDVS Veri Seti.....	64
5.4. Deneysel Çalışmalar	65
5.4.1. Birinci Senaryo	66
5.4.2. İkinci Senaryo	68
5.4.3. Üçüncü Senaryo: Yeni Tip Yüz Manipülasyonu Altındaki Performans	69
5.4.4. Önerilen Yöntemi Mevcut Deepfake Tespit Yöntemleri ile Karşılaştırma.....	72
5.4.5. Önerilen Yöntemin Mevcut Veri Setleri Üzerindeki Performansı.....	73
6. SONUÇLAR.....	75
ÖNERİLER	76
KAYNAKLAR	77
ÖZGEÇMİŞ	

ÖZET

Görüntü Manipülasyonlarının Derin Öğrenme Yaklaşımı ile Belirlenmesi

Semih YAVUZKILIÇ

Doktora Tezi

FIRAT ÜNİVERSİTESİ

Fen Bilimleri Enstitüsü

Elektrik Elektronik Mühendisliği Anabilim Dalı
Elektrik Elektronik Mühendisliği Teknolojileri Programı
Ocak 2022, Sayfa: xiii + 85

Makine öğrenmesinin popüler bir alt dalı olan derin öğrenme yaklaşımları, günden güne artan bir ilgiyle birçok alanda kullanılmaktadır. Derin öğrenme yaklaşımları; görüntü işleme, nesne tanıma, sinyal işleme gibi birçok alanda oldukça başarılı bir şekilde kullanılmaktadır. Benzer şekilde; insansız hava araçları, bilgisayarlı görüntüleme, akıllı telefon gibi alanlarda da etkili bir şekilde kullanılmaktadır.

Günümüzde bu teknolojilerin çok çeşitli uygulamaları mevcuttur. Ancak bu uygulamalardan bazıları tehdit veya tehlike arz etmektedir. Bu tez çalışmasında, görüntü manipülasyonlarını tespitinde derin öğrenme tabanlı yöntemlerden olan Evrişimsel Sinir Ağları (ESA)'nın farklı mimarileri kullanılarak sınıflandırma problemi çözülmeye çalışılmıştır. Özellikle, görüntü manipülasyonları türlerinden olan ve son yıllarda kullanımı yaygınlaşan yüz manipülasyonu (deepfake) tespiti üzerinde durulmuştur. Bu kapsamda yeni bir deepfake veri seti oluşturulmuştur.

Önerilen yöntemlerde, genel olarak derin öznitelik çıkarımı tabanlı kullanılmıştır. Görüntü manipülasyonlarının tespiti amacıyla farklı derin öğrenme mimarileri kullanılmıştır. Önceden eğitilmiş bazı derin öğrenme mimarileri görüntülerin sınıflandırılması probleminde kullanılmıştır. Ayrıca yeni bir derin öğrenme mimarisi görüntü manipülasyonlarının tespiti için sunulmuştur. Önerilen yöntemler, aynı veri setlerini kullanan mevcut yöntemlere göre daha iyi bir doğruluk performansı sergilemiştir.

Anahtar Kelimeler: Makine Öğrenme, Derin Öğrenme, Evrişimsel Sinir Ağı, Görüntü İşleme, Görüntü Manipülasyonları, Deepfake

ABSTRACT

Detection of Image Manipulations with Deep Learning Approach

Semih YAVUZKILIÇ

Ph.D. Thesis

FIRAT UNIVERSITY

Graduate School of Natural and Applied Sciences

Department of Electrical and Electronics Engineering
Electrical and Electronics Engineering Technologies

January 2022, Pages: xiii + 85

Deep Learning approaches, which are popular sub-branch of machine learning, are used in many fields with increasing interest day by day. It has been successfully applied in many areas such as image processing, object recognition and signal processing. Similarly, it is also used effectively in areas such as unmanned aerial vehicles, computer imaging and smart phones.

Nowadays, these technologies have a wide variety of applications. However, some of these applications offer a threat or danger. In this thesis, the classification problem has been tried to be solved by using different architectures of Convolutional Neural Networks (CNN), which is one of the deep learning-based methods to detect image manipulations. In particular, face manipulation (deepfake) detection, which is one of the types of image manipulations and has become widespread in recent years, has been emphasized. In this context, a new deepfake dataset was created.

In the proposed methods, studies have been carried out with deep features-based approaches in general. Different deep learning architectures have been used to detect image manipulations. Some pre-trained deep learning architectures have been used in the image classification problem. Also, a new deep learning architecture is presented for detection of image manipulations. The proposed methods showed better accuracy performance than existing methods using the same datasets.

Keywords: Machine Learning, Deep Learning, Convolutional Neural Network, Image Processing, Image Manipulations, Deepfake

ŞEKİLLER LİSTESİ

	Sayfa
Şekil 2.1. ÜÇA'nın yapısı	6
Şekil 2.2. Sentezlenmiş yüzler ile oluşturulan çeşitli veri setlerinin örnekleri	8
Şekil 2.3. Yüz değiştirme manipülasyon tekniği için FaceForensics++ [1] veri setinde yer alan gerçek ve sahte yüz görüntüleri örnekleri	9
Şekil 2.4. Yüz özelliklerinin oluşturulması manipülasyon türü için gerçek ve sahte yüz görüntüleri örnekleri.....	10
Şekil 2.5. Yüz ifadeleri manipülasyon türü için FaceForensics ++ [1] veri setinde yer alan bazı örnek resimler.	12
Şekil 3.1. Tek katmanlı YSA mimarisi.....	22
Şekil 3.2. Çok katmanlı YSA mimarisi	23
Şekil 3.3. Geri beslemeli YSA'nın temsili gösterimi	24
Şekil 3.4. Derin İnanç Ağının temsili gösterimi	25
Şekil 3.5. Boltzmann Makinesi mimarisinin temsili gösterimi.....	26
Şekil 3.6. Kısıtlı Boltzmann Makinesi.....	27
Şekil 3.7. Klasik bir sinir ağında çıkış katmanının giriş birimleri ile etkileşimi	29
Şekil 3.8. Temel bir evrişim ağında çıkış katmanının giriş birimi ile etkileşimi	29
Şekil 3.9. Klasik bir sinir ağında giriş biriminin çıkış katmanları ile etkileşimi	30
Şekil 3.10. Temel bir evrişim ağında giriş biriminin çıkış katmanı ile etkileşimi	30
Şekil 3.11. İki katmanlı temel bir evrişim ağında çıkış katmanının giriş birimleri ile etkileşimi	31
Şekil 3.12. Temel bir sinir ağının parametre paylaşımı	32
Şekil 3.13. Bir Evrişimsel Sinir Ağında parametre paylaşımı	33
Şekil 3.14. Temel bir ESA mimarisi örneği	33
Şekil 3.15. Sigmoid fonksiyonunun grafiği.....	35
Şekil 3.16. DDB katmanında kullanılan aktivasyon fonksiyonu	36
Şekil 3.17. Maksimum ve ortalama havuzlama örnek gösterimi	37
Şekil 3.18. Tam bağlı katmanın temsili gösterimi	39
Şekil 3.19. İki katmalı bir ağ için seyreltme katmanındaki seyreltme senaryoları	40
Şekil 3.20. Otomatik kodlayıcının temel yapısı.....	41
Şekil 3.21. Örnek bir seyrek otomatik kodlayıcı modeli	42
Şekil 3.22. Gürültü giderici otomatik kodlayıcının temsili gösterimi	43
Şekil 3.23. Dışbükey olmayan bir fonksiyonun temsili gösterimi	46
Şekil 3.24. ResNet yapısını gösteren blok diyagram	48

Şekil 3.25. VGG16 mimarisi	50
Şekil 3.26. VGG19 mimarisi	51
Şekil 3.27. Klasik makine öğrenmesi modeli (a) ve transfer makine öğrenme modeli (b).....	52
Şekil 3.28. K değeri 10 olarak seçilmiş 10- Kat çapraz geçerlemenin temsili gösterimi	53
Şekil 3.29. Karmaşıklık matrisi	54
Şekil 4.1. Önerilen deepfake tespit mimarisi	56
Şekil 4.2. Gerçek videolardan alınan bazı örnek çerçeveler. İlk satırda Celeb-DF [14] veri setinden alan bazı çerçeveler yer alırken ikinci satırda ise DeepFakeTIMIT [10] veri setinden bazı çerçeveler gösterilmektedir.	57
Şekil 4.3. Verilen çerçevelerde algılanan yüz bölgeleri. İlk satır Celeb-DF veri seti için tespit edilen bazı yüzleri gösterirken ikinci satır ise DeepFakeTIMIT veri seti için tespit edilen bazı yüzleri göstermektedir.	57
Şekil 4.4. Sahte videolardan alınan bazı örnek çerçeveler. İlk satır Celeb-DF veri setinden alınan bazı sahte çerçeveleri, ikinci satır ise DeepFakeTIMIT veri setinden alınan bazı sahte çerçeveleri göstermektedir.	58
Şekil 4.5. Verilen sahte çerçevelerde algılanan yüz bölgeleri. İlk satır, Celeb-DF veri seti için tespit edilen bazı sahte yüzleri gösterirken ve ikinci satır ise DeepFakeTIMIT veri seti için tespit edilen bazı sahte yüzleri göstermektedir.	58
Şekil 5.1. Deepfake sınıflandırması için önerilen çoklu akış ESA mimarisinin temsili gösterimi.	63
Şekil 5.2. DPDVS veri setinde yer alan örnek çerçeveler. İlk satırda, farklı politikacıların gerçek görüntüleri gösterilmektedir. Takip eden satırlarda ise sırasıyla makyaj (yalnızca kadınlar için)/sakal (yalnızca erkekler için), saç modeli, saç rengi, yaş, gülüms.....	65
Şekil 5.3. Birinci senaryonun eğitimi için eğitim doğruluğu eğrisi (a) ve eğitim kayıp eğrisi (b)	67
Şekil 5.4. Birinci senaryo için elde edilen karmaşıklık matrisi	67
Şekil 5.5. İkinci Senaryonun eğitimi için eğitim doğruluğu eğrisi (a) ve eğitim kayıp eğrisi (b)	68
Şekil 5.6. İkinci senaryo için elde edilen karmaşıklık matrisi	69

TABLÖLAR LİSTESİ

	Sayfa
Tablo 4.1. Eğitim parametre değerleri	59
Tablo 4.2. Önerilen yöntemin her iki veri kümesinde EER (%) ve FRR@FAR10% (%) cinsinden performans değerlendirmesi	59
Tablo 4.3. Önerilen yöntemin DeepFakeTIMIT veri setindeki mevcut yöntemlerle EER (%) ve FRR@FAR10% (%) cinsinden performans karşılaştırması	60
Tablo 4.4. Önerilen yöntemin Celeb-DF veri setindeki mevcut yöntemlerle EER (%) cinsinden performans karşılaştırması.	61
Tablo 5.1. Birinci senaryo için hesaplanan performans ölçüleri	68
Tablo 5.2. İkinci senaryo için hesaplanan performans ölçüleri	69
Tablo 5.3. Önerilen derin akış yönteminin eğitim seti olarak yalnızca bir manipülasyon türü dikkate alındığında diğer manipülasyon türleri test seti olarak kullanıldığında elde edilen doğruluk oranları	71
Tablo 5.4. Önerilen yöntemin mevcut yöntemler ile DPDVS veri seti kullanılarak karşılaştırıldıktan sonra elde edilen doğruluk oranları	73
Tablo 5.5. Çeşitli veri setleri kullanılarak önerilen derin akış yönteminin doğruluk oranları	74

SİMGELER VE KISALTMALAR

Simgeler

$\dot{x}_i$	: Normalleştirilmiş etkileşim
L_i	: Gerçek sınıf etiketleri
M_i	: Momentum parametresi
S_i	: Eşiksiz en büyük işlev çıkışı
b_i	: Yanılgı
m_b	: Mini toptan ortalama
v_b	: Mini toptan değişim
x_i	: Giriş
ω_i	: Ağırlık değeri
η	: Öğrenme oranını

Kısaltmalar

3B	: Üç Boyutlu (Three Dimensional)
AUC	: Eğri Altında Kalan Alanı (Area Under Curve)
BGP	: İkili Gabor Modeli (Binary Gabor Pattern)
BM	: Boltzmann Makinesi
BSIF	: İkili İstatistiksel Görüntü Özellikleri (Binarised Statistical Image Features)
CENTRIST	: CENSus Transform Histogramı
DDB	: Düzeltilmiş Doğrusal Birim
DFD	: Deep Fake Tespiti (Deep Fake Detection)
DFDC	: Deepfake Tespit Yarışması (Deepfake Detection Challenge)
DFFD	: Farklı Sahte Yüz Veri Seti (Diverse Fake Face Dataset)
DİA	: Derin İnanç Ağları (Deep Belief Network)
DP	: Doğruluk Performansı
DPDVS	: Dünya Politikacılar Deepfake Veri Seti
DVM	: Destek Vektör Makinesi
EER	: Eşit Hata Oranı (Equal Error Rate)
ESA	: Evrimsel Sinir Ağları
FAR	: Yanlış Kabul Oranı (False Acceptance Rate)
FDLBP	: Yerel İkili Model (Frequency Decoded Local Binary Pattern)
FRR	: Yanlış Ret Oranı (False Rejection Rate)
HOG	: Gradyanların Histogramı (Histogram of Gradients)
IcGAN	: Tersine Çevrilebilir Koşullu ÜÇA
KBM	: Kısıtlı Boltzmann Makineleri (Restricted Boltzmann Machine)
LBP	: Yerel İkili Modeller (Local Binary Patterns)
MCT	: Hareket Dengeleme Transformatörü (Motion Compensation Transformer)
PHOG	: Yönlendirilmiş Gradyanların Histogramı Piramidi

PSNR	: Tepe Sinyal-Gürültü Oranı (Peak Signal-To-Noise Ratio)
ReLU	: Düzeltilmiş Doğrusal Birim (Rectified Linear Unit)
ResNet	: Artık Sinir Ağı (Residual Neural Network)
RGİ	: Rastgele Gradyan İnişi
RGİM	: Rastgele Gradyan İnişi Momentumu
SFE	: Mekansal Özellik Çıkarıcı (Spatial Feature Extractor)
SIFT	: Ölçek Değişmeyen Özellik Dönüşümü (Scale Invariant Feature Transform)
SPMC	: Alt Piksel Hareket Dengeleme Katmanını (Sub-Pixel Motion Compensation)
SRM	: Mekansal Zengin Model (Spatial Rich Model)
SSIM	: Yapısal Benzerlik İndeksi (Structural Similarity Index)
SURF	: Hızlandırılmış Güçlü Özellikler (Speeded Up Robust Features)
TC-Gan	: Doku Tamamlama ÜÇA (Texture Completion GAN)
TÖ	: Transfer Öğrenme
TSA	: Tekrarlı Sinir Ağı (Recurrent Neural Network)
UKSB	: Uzun Kısa Süreli Bellek
ÜÇA	: Üretken Çekişmeli Ağlar (Generative Adversarial Networks, GAN)
YSA	: Yapay Sinir Ağları

1. GİRİŞ

Günümüzde derin öğrenme ve yapay zeka kullanım alanları önemli ölçüde genişlemiştir. Bunlar birçok günlük yaşam uygulamalarında ve sistemlerinde yaygın olarak kullanılmaktadırlar. Benzer şekilde sosyal ağlar da günümüzde bilgi ve haberlerin yayılmasında önemli bir rol oynamaktadır. Sosyal medya ile birlikte makine öğrenmesi ve yapay zekanın ortaya çıkışı, özellikle dijital dünyada gerçeklik algısını değiştirmiştir. Örneğin; mevcut teknolojiler, gerçek olanlardan neredeyse ayırt edilemez hale getirilen sahte multimedya örneklerini oluşturmak için kullanılmaktadır. Bununla birlikte, çok fazla teknik bilgiye ihtiyaç duymayan hazır görüntü ve video işleme yazılımları ve uygulamaları da (Face2Face, AgingBooth, FaceApp, PotraitPro Studio ve Adobe Photoshop vb.) mevcuttur. Bu nedenle, manipüle edilmiş görüntülerin ve videoların sayısı oldukça fazladır. Hatta internette veya sosyal medyada yayınlanmış olan yanıltıcı ve manipüle edilmiş birçok resim ve video bulunmaktadır. Manipüle edilen resimler veya videolar, görüntüleri güzelleştirmek için rötuşlama yapmak gibi iyi niyetli nedenlerle veya sahte haber kampanyaları gibi kötü amaçlar için kullanılabilir. Özellikle, derin öğrenme teknikleri kullanılarak oluşturulan sahte yüz resimleri ve videoları son zamanlarda kamuoyunda büyük bir ilgiyle beraber endişeye de yol açmıştır.

Yapay zeka, son yıllarda kullanımı en yaygın teknolojik uygulamalardan biri haline gelmiştir. Günümüzde bu teknolojilerin çok çeşitli uygulamaları mevcuttur. Ancak bu uygulamalardan bazıları tehdit veya tehlike olarak tanımlanmaktadır. Tehlikeli olduğu düşünülen uygulamalardan birisi de deepfake'dir. Deepfake, bir resim veya videoda bir kişinin görüntüsünün yapay sinir ağları (YSA) aracılığıyla değiştirilmesi veya başka bir kişinin yüzü ile değiştirildiği bir manipülasyon türüdür[1–3]. Deepfake manipülasyonu, farklı algoritmalar veya uygulamalar kullanılarak gerçekleştirilebilmektedir. Bu yöntemlerin başında Üretken Çekişmeli Ağlar (ÜÇA) (Generative Adversarial Networks, GAN) [4] gelmektedir. Bu algoritmalar ile gerçek yüzlerden neredeyse ayırt edilemeyen insan yüzleri üretilmektedir. Ayrıca sosyal medya kullanımının hızla artmasıyla DeepFake, Face2Face, FaceApp ve Face Swap Live gibi halka açık birçok uygulama ortaya çıkmıştır. Söz konusu uygulamalar ve yazılımlar kişinin yaşı, cinsiyeti, saç rengi gibi karakteristik özelliklerini değiştirmek veya farklı iki kişiye ait iki yüzü birbiriyle değiştirmek veya gerçekte olmayan bir kişiye ait sentetik yüzler oluşturmak için kullanılabilmektedir. Şekil 1'de farklı manipülasyon türleri uygulanarak elde edilen deepfake örnekleri gösterilmiştir. İlk satırda gerçek yüz örnekleri; ikinci satırda ise manipüle edilmiş yüz örnekleri yer almaktadır. Son sütunda yer alan yüz ise sentetik olarak üretilmiş gerçekte var olmayan bir kişiyi göstermektedir.



Şekil 0.1. Çeşitli manipülasyon sınıflarına ait örnek resimler

Deepfake teknolojisi, gerçek gibi görünen sahte görüntüler ve sesler üreten gelişmiş yapay zeka ve derin öğrenme teknikleri kullanılarak üretilen manipüle edilmiş videolar veya diğer dijital temsillerdir. Deepfake uygulamaları çoğunlukla araştırma veya eğlence gibi zararsız amaçlar için kullanılabilirler. Ancak, kötü niyetli kişiler tarafından hazırlanan uygulama ve yazılımların kullanılması nedeniyle önemli risk ve tehlikelere yol açabilmektedir. Örneğin; hırsızlık, dolandırıcılık veya intikam için sahte ses ve video kayıtları üretmek için kullanılabilir. Günümüzde siber savaşlarda, sosyal olayları tetiklemede ve algı yönetiminde deepfake kullanılabilir. Deepfake resim veya videolar kullanılarak gerçek dışı bilgiler kolay bir şekilde yayılabilir. Bu resim veya videolar ile politikacılara yönelik yanlış algılar oluşturulabilir. Politikacılar tarafından gerçekte hiç söylenmeyen sözler söylenmiş gibi gösterilerek politikacılar zan altında bırakılarak yapılacak seçimlerde seçmenleri manipüle edebilecek bir siyasi saldırı aracı olarak kullanılabilir. Benzer şekilde deepfake teknikleri ile oluşturulan sahte deliller mahkemede kişilerin aleyhine kullanılabilir ve bu yüzden masum kişiler işlemediği suçlardan dolayı itham edilebilmektedir. Öte yandan, suçlular sahte delillere dayanarak serbest bırakılabilmektedir. Ayrıca kötü niyetli aktörler; tüketici dolandırıcılığı, rüşvet, cinsel taciz gibi suçlara karışmayan bir şirket veya şirket yöneticisi hakkında sahte haberler üretmek ve yaymak için deepfake algoritmalarını bir araç olarak kullanabilir. Bu tür gerçekçi olmayan haberler, şirketlerin itibarını zedeleyebilmektedir. Ek olarak, yüz yaşlanmasının ve yüz sahteciliğinin otomatik yüz tanıma ve tanımlama sistemlerini olumsuz yönde etkilediği tespit edilmiştir [5–9]. Bunun sonucunda, insanlar yaşa dayalı erişim kontrollerini atlatmak için yüzlerini daha genç veya daha yaşlı görünecek şekilde değiştirebilirler. Bunun yanında deepfake manipülasyonları insanları aldatmakla beraber aynı zamanda yüz tanıma sistemlerinin doğruluğunu da düşürmektedir. Korshunov ve Marcel [10], deepfake'in VGG ve FaceNet sinir ağları tabanlı yüz tanıma sistemlerinin hata oranlarını sırasıyla %85,62 ve %95,00 oranında artırdığını göstermiştir.

Deepfake içeriklerine karşı alınan en iyi önlemlerin başında gerçek yüz örneklerini sahte yüzlerden ayırt etmeyi amaçlayan deepfake tespit yöntemleri yer almaktadır [11–13]. Bu tespiti

yapmak için araştırmacılar tarafından birçok deepfake tespit algoritmaları geliştirilmiştir. Ancak deepfake tespit algoritmalarını eğitmek için çok sayıda gerçek ve sahte resim veya video gerekmektedir. Bu ihtiyacı gidermek için araştırmacılar tarafından birkaç farklı veri seti hazırlanmıştır. Bu veri setlerine Celeb-DF [14], UADFV [15] ve DeepfakeTIMIT [10] ve CelebA [16] gibi deepfake veri setleri örnek olarak gösterilebilir. Ayrıca, bazı büyük şirketler de deepfake tespiti için büyük ölçekli veri setlerini araştırmacıların hizmetine sunmuşlardır. Örneğin; Google, Deep Fake Tespiti (Deep Fake Detection, DFD) [17] adında yeni bir veri setini oluşturmuştur. Benzer şekilde Facebook, Deepfake Tespit Yarışması (Deepfake Detection Challenge, DFDC) [18] 'de yer alan yeni bir veri setini sunmuştur.

Lyu [19], çeşitli zorlukları ve deepfake alanındaki araştırma fırsatlarını araştırmıştır. Araştırmacıya göre mevcut deepfake oluşturma yöntemlerinin kritik bir dezavantajı, cilt ve yüz kılları gibi iyi detayları üretememeleridir. Bu dezavantaj üretimin kodlama adımındaki bilgi kaybından kaynaklanmaktadır. Yazara göre deepfake oluşturma yöntemi üç farklı şekilde yapılmaktadır. Kafa kuklası, kaynak kişinin başının ve omuz üst kısmının kopyalanmasını ve ardından hedef kişinin vücuduna yapıştırılmasını içermektedir. Böylece hedef, kaynağa benzer şekilde davranmaktadır. İkinci yöntem, yalnızca kaynak kişinin yüzünü hedefinkiyle değiştiren yüz değiştirmedir. Aynı zamanda yüz ifadelerinin değişmeden kalmasını sağlar. Üçüncü yöntem, hedefin gerçekte konuşmadığı bir şeyi konuşuyormuş gibi görünmesi için yalnızca dudak bölgesini manipüle ederek tahrif edilmiş bir video oluşturmak için kullanılan dudak senkronizasyonudur. Algılama yöntemleri, çerçeve düzeyinde ikili sınıflandırma problemleri olarak formüle edilmiştir. Yaygın olarak kullanılan üç tespit yönteminden birinci kategori, deepfake videolarında fiziksel/fizyolojik yönlerde sergilenen tutarsızlıkları dikkate alır. İkinci algoritma, sinyal seviyesindeki sahtelikleri kullanır. Son algılama kategorisi ise veri odaklıdır, doğrudan orijinal ve sahte videolar üzerinde eğitilmiş birden çok derin sinir ağı yöntemlerini kullanmaktadır, ancak yalnızca belli yapaylıklar yakalamaktadır. Ayrıca deepfake veri setlerinin kalitesi, bu yöntemlerin sınırlamalarına da ışık tutmaktadır.

Deepfake tespiti için birçok çalışma yapılmıştır. Örneğin; Neubert vd. [20], yüz görüntülerindeki manipülasyonu tespit etmek için frekans ve mekansal alan özelliklerini tasarlamışlardır. Bunun yanında, derin öğrenme tabanlı yapıların nesne tanıma ve otonom araç gibi çeşitli uygulamalardaki başarılarından esinlenen araştırmacılar, derin öğrenme tekniklerinin yüz manipülasyon tespitine karşı etkinliğini araştırmışlardır. Buna istinaden son birkaç yılda deepfake görüntü manipülasyonlarını tespit etmek için derin öğrenme mimarileri başarıyla kullanılmıştır. Özellikle manipülasyon tespitinde her çerçeveden öznitelikleri çıkarmak için Evrimsel Sinir Ağları (ESA) kullanılmaktadır. Önceden eğitilmiş ESA mimarilerini öznitelik çıkarıcı olarak kullanmak, yüz manipülasyon tespitinin doğruluğunu arttırmanın etkin bir yoludur [21,22].

Bu tez çalışmasında görüntü manipölasyonu türlerinden olan ve son yıllarda popüler hale gelen yüz görüntüsü manipölasyonu (deepfake) tespitinde yeni algoritmalar geliştirmek için derin öğrenme tabanlı yöntemlerden olan ESA mimarileri kullanılarak sınıflandırma problemi çözülmeye çalışılmıştır.

1.1. Tezin Literatüre Katkısı

Bu tez çalışması kapsamında dünyaca ünlü politikacılara ait videolardan 320.499 adet resim ve bu resimlere çeşitli manipölasyonlar uygulanarak aynı sayıda sahte resim elde edilerek yeni bir deepfake veri seti oluşturulmuştur.

Tez süresince bir adet uluslararası kitap bölümü yayınlanmış; bir adet uluslararası Science Citation Index Expanded indeksli makale yayınlanmıştır. Ayrıca bir adet ulusal Science Citation Index Expanded indeksli makale yayınlaması için sunulmuştur. Çalışmalar ile ilgili bilgiler aşağıda belirtilmiştir:

- S. Yavuzkiliç, Z. Akhtar, A. Sengür, K. Siddique, DeepFake Face Video Detection Using Hybrid Deep Residual Networks and LSTM Architecture, AI Deep Learn. Biometric Secur. (2021) 81–104. <https://doi.org/10.1201/9781003003489-4>.
- S. Yavuzkiliç, A. Sengür, Z. Akhtar, K. Siddique, Spotting deepfakes and face manipulations by fusing features from multi-stream cnns models, Symmetry (Basel). 13 (2021). <https://doi.org/10.3390/sym13081352>.

1.2. Tezin Organizasyonu

Bu tez çalışması altı bölümden oluşmaktadır. Birinci bölümde tez çalışması ile ilgili genel bilgiler verilmiştir. İkinci bölümde görüntü manipölasyonları ile ilgili bilgiler verilmiştir. Üçüncü bölümde öznitelik çıkarma ve sınıflandırma aşamalarında kullanılan yöntemler tanıtılmıştır. Dördüncü bölümde tez ile ilgili olarak yapılan ilk çalışma olan hibrit derin öğrenme mimarisi kullanılarak deepFake videolarında yüz tespiti yapılmıştır. Beşinci bölümde, çoklu akış ESA modellerinin özellikleri birleştirilerek deepfake ve yüz manipölasyonlarının tespiti yapılmış ve önerilen yöntemlerin deneysel sonuçları verilmiştir. Son olarak altıncı bölümde önerilen yöntemlerden elde edilen sonuçlar irdelenmiştir.

2. GÖRÜNTÜ MANİPÜLASYONLARI

Görüntü manipülasyon türleri genel olarak içeriği korunan ve içeriği değiştirilen görüntüler olmak üzere iki kategoriye ayrılabilir. İçeriği korunan manipülasyon türünde; sıkıştırma, bulanıklaştırma, kontrast geliştirme gibi görüntüye sonradan yapılan işlemler uygulanır ve görüntüye uygulanan işlemler herhangi bir anlamsal içeriği değiştirmediği için daha az zararlı olarak kabul edilir. Diğer türde ise kopyala – yapıştırma, nesnenin kaldırılması yüz manipülasyonları gibi işlemler uygulanarak görüntü içeriğini rastgele yeniden şekillendirir ve görüntünün anlamsal içeriği önemli ölçüde değiştirir. Bu çalışmada görüntü manipülasyonlarında türlerinden yüz manipülasyonları tespiti ele alınmıştır.

Gelişmiş yüz manipülasyon teknikleri hem insanları hem de otomatik yüz tanıma sistemlerini aldatabilmektedir. Günümüzde deepfake teknikleri kullanılarak oluşturulmuş yanlış bilgilerle manipüle edilmiş içerikler mesajlaşma ve sosyal ağ platformları kullanılarak hızla yayılabileceğinden deepfake önemli bir sorun haline gelmiştir. Yanlış bilgilerin bu şekilde yayılması insanlar için sıkıntılı sonuçlar doğurabilmektedir. Örneğin; daha önce tartıştığınız bir kişi, deepfake teknikleri kullanarak yüzünüzü başka bir kişinin yüzüyle değiştirebilir ve bunun sonucunda işlemediğiniz bir suçtan dolayı yargılanabilirsiniz.

2.1. Yüz Manipülasyon Kategorileri

Manipüle edilen yüzler resim veya videolarda genellikle saç rengi ve yaş gibi yüz özelliklerinin değiştirilmesi, iki yüzün birbiriyle yer değiştirmesi (face swap) ve yapay olarak yüzler oluşturulmasıyla üretilmektedir [2]. Mevcut yüz manipülasyon teknikleri sistematik olarak analiz edildiğinde tüm manipülasyonlar genel olarak dört kategoride gruplanabilmektedir. Bunlar yüz sentezi, yüz değiştirme, yüz özellikleri ve yüz ifadesi manipülasyon teknikleridir [3,23].

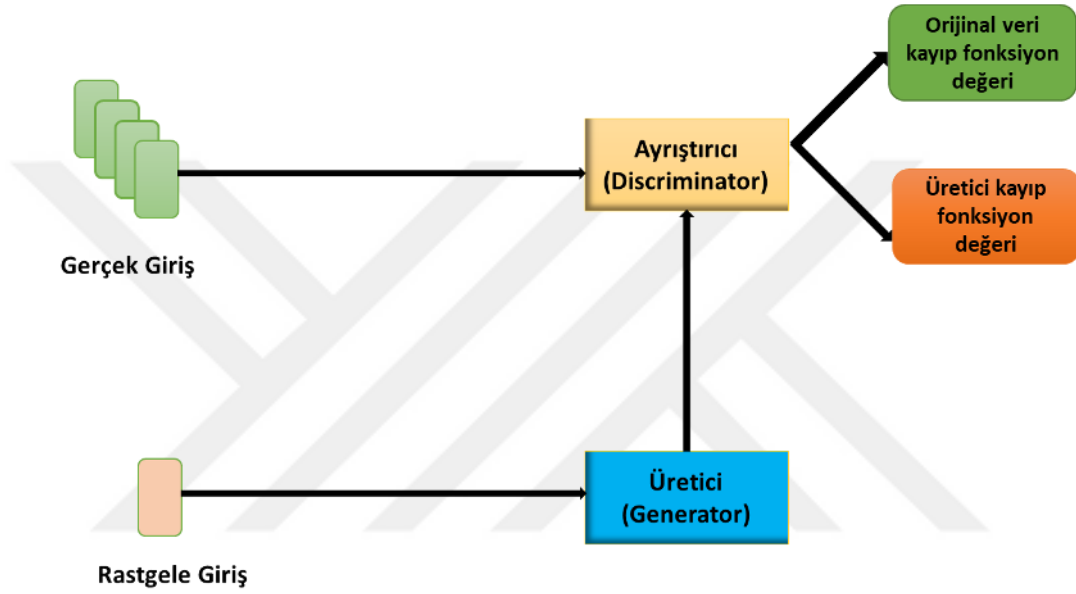
2.1.1. Yüz Sentezi

Yüz sentezi manipülasyon tekniği ile genellikle ÜÇA algoritmaları kullanılarak gerçek olanlardan neredeyse ayırt edilemeyen, gerçekte dünyada var olmayan insan yüz örnekleri elde edilmektedir [4]. Örneğin; Karras vd. [24], ProGAN [25] yaklaşımının geliştirilmiş bir modeli olan ve gerçekte olmayan yüzleri oluşturabilen StyleGAN mimarisini sunmuşlardır.

Yüz Sentezi Oluşturma Yöntemleri ve Veri Setleri

ÜÇA'ların algoritmik mimarileri, üretici (generator) ve ayrıştırıcı (discriminator) olarak adlandırılan iki sinir ağına dayanmaktadır. İlk olarak, üretici gerçekçi kalitede sahte yüz görüntüleri oluştururken, ayrıştırıcı yüz görüntülerini gerçek ve sahte örnekler arasında ayırt eder. Ayrıştırıcı,

gerçek ve sahte görüntüler arasında ayrım yapamadığında, sonuçta gerçekte olmayan ama görünüşte gerçekle aynı olan görüntüler ortaya çıkar. Şekil 2.1’de ÜÇA’ların çalışma şekli gösterilmektedir. ÜÇA’dan yararlanarak, görüntüden görüntüye çeviriyi öğrenen CycleGAN [26] önerilmiştir. Shen vd. [27] yüz kimliği sınıflandırmasını üçüncü aktör olarak kabul eden gerçek ve sentezlenmiş yüzlerin kimliklerini ayırt ederek üretici ağ ile mücadele eden FaceID-GAN’ı önermişlerdir. ÜÇA algoritmaları tamamen bağlı, ESA ile oluşturulan ve koşullu ÜÇA’lar olmak üzere üç farklı şekilde sınıflandırılabilir.



Şekil 0.1. ÜÇA'nın yapısı

Tamamen bağlı ÜÇA'lar ilk ÜÇA mimarilerindendir. Üretici ve ayırmacı tarafında tam bağlı katman ağı kullanılarak sentezlenmiş görüntüler oluşturulmuştur. Tam bağlı ÜÇA'lar kullanılarak rakam ve el yazılarından oluşan MNIST [28] veri seti, birbirinden farklı on sınıftan oluşan CIFAR-10 [29] veri seti gibi bazı veri setleri oluşturulmuştur.

ESA'ların oldukça etkili görüntü tanıma özelliğinden mütevellit ESA'lar, ÜÇA mimarilerinde de kullanılmıştır. Ancak ESA ile üretici ve ayırtıcı eğitiminin zor olmasından dolayı bu yöntemin kullanılmasını sınırlandırmıştır. ESA mimarisi kullanılarak sunulan Laplace Piramit ÜÇA [30], üreticiyi ayırtarak çoklu ölçekler oluşturmuştur. ÜÇA ilk olarak bir görüntünün oldukça düşük çözünürlüğe sahip bir şeklini elde etmek için eğitilmiştir. Daha sonra elde edilen bu görüntüye yeni özellikler eklenerek eğitimler gerçekleştirilmiştir. Ayrıca Laplace Piramit ÜÇA, aşağı örnekleme ve yukarı örnekleme operatörlerinin eğitim süresince öğrenmesini sağlayan aşamalı ve kesirli aşamalı evrişimlerden yararlanmaktadır. Örnekleme operatörleri

görüntü uzayına ve gizli alandan bir ayrıştırıcıya haritalamada önemli bir gereksinim olan örnekleme hızlarını ve konumlardaki değişimleri takip eder [30,31].

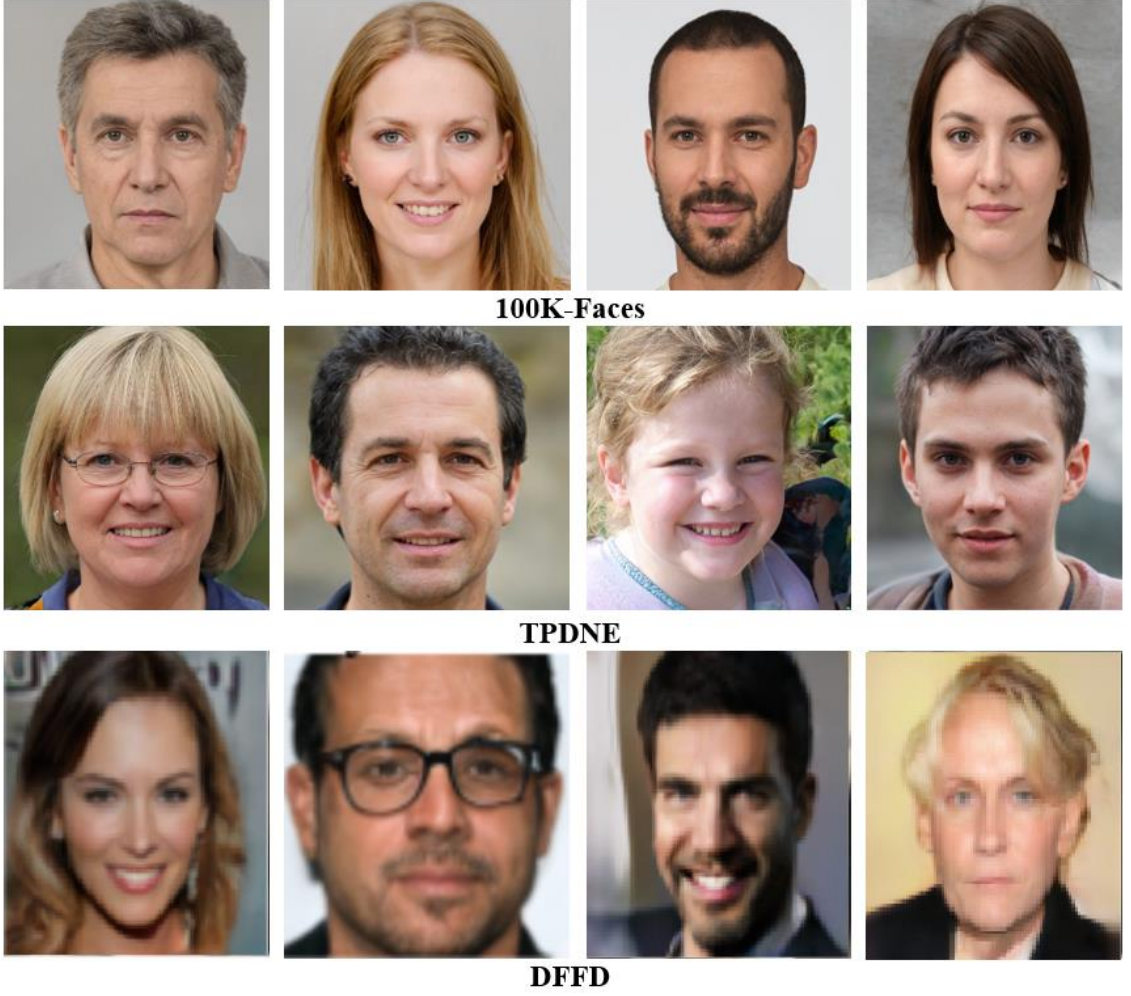
Koşullu ÜÇA, çok modlu veri sentezi için daha iyi görüntüler üretebilmektedir. Üretici ile gizli alanın çıkışları arasında yer alan karşılıklı bilgiyi en üst seviyeye çıkartarak değişkenliğin gizli etkilerini bulmaya çalışır. Gürültü kaynağını gizli bir koda ayrıştırabilen ve sıkıştırılabilir kaynağa çevirebilen InfoGAN [32] ile koşullu ÜÇA arasında bir paralellikten bahsedilebilir. InfoGAN, pozlar, renkler, yüz resimlerinin duygusal içeriğinde bulunan çeşitlikler ve ışık da dahil olmak kaydıyla resimlerin iç içe geçmiş karmaşık özelliklerini öğrenerek görüntüler oluşturmaktadır [31].

Araştırmacılar için halka açık birkaç adet yüz sentezi veri seti bulunmaktadır. Şekil 2.2 sentezlenmiş yüzlerin üç farklı veri setinde yer alan resimlerin örneklerini gösterilmektedir. Bu veri setlerinin ortak özelliği, hiçbirinin gerçek bir kişi resmini içermemesidir. Bu nedenle, bu tür manipülasyon tespitine veya tanımaya odaklanan araştırmacılar, sistemlerini eğitmek için genellikle halka açık veri setlerinde yer alan gerçek yüzleri kullanırlar. Aşağıda, yüz sentezi manipülasyon tekniği ile oluşturulan bazı veri setlerinin kısaca açıklaması yer almaktadır.

100K-Faces [33]: Bu veri seti, StyleGAN [24] kullanılarak oluşturulan 100.000 sentetik resimden oluşmaktadır. Veri setinde StyleGAN, 69 farklı kimliğe ait yaklaşık 29.000 görüntü ile eğitilmiş ve düz bir arka plana sahip yüz örnekleri üretmiştir.

TPDNE: Bu veri seti, [34]'deki web sitesinde toplanan 150.000 sentetik yüz görüntüsünü içermektedir. Sentetik yüz görüntüleri, FFHQ veri seti [35] ile eğitilmiş StyleGAN mimarisine dayanmaktadır.

DFFD: Stehouwer vd. [23], Farklı Sahte Yüz Veri Seti (Diverse Fake Face Dataset, DFFD) olarak adlandırılan bir veri seti sunmuşlardır. Yazarlar, yapay yüz sentezi manipülasyonu için önceden eğitilmiş iki model kullanmışlardır. 100.000 ve 200.000 sahte görüntü, sırasıyla ProGAN ve StyleGAN modelleri kullanılarak oluşturulmuştur.



Şekil 0.2. Sentezlenmiş yüzler ile oluşturulan çeşitli veri setlerinin örnekleri

2.1.2. Yüz Değiştirme

Yüz değiştirme manipülasyonu tekniğinde bir kişinin yüzü bir başkasının yüzü ile değiştirilmektedir. Yüz değiştirme, geleneksel bilgisayar grafik tabanlı uygulamalar veya derin öğrenme teknikleri ile yapılabilmektedir. Ayrıca yüz değiştirme işlemini gerçekleştirmek için Snapchat gibi kullanımı yaygın birkaç farklı mobil uygulama da bulunmaktadır. Yüz değiştirme için son zamanlarda literatürde birçok çalışma yayınlanmıştır. Örneğin, Korshunov ve Marcel [10], ÜÇA tabanlı bir yüz değiştirme mimarisi kullanarak deepfake veri seti oluşturmuşlardır. Bazı yüz değiştirme örnekleri Şekil 2.3'de gösterilmiştir.



Şekil 0.3. Yüz değiştirme manipölasyon tekniđi için FaceForensics++ [1] veri setinde yer alan gerçek ve sahte yüz görüntüleri örnekleri

Yüz Deđiřtirme Oluřturma Yöntemleri ve Veri Setleri

Yüz deđiřtirme, kullanımı gün geçtikçe artan manipölasyon tekniklerinden biridir. Ařađıda, halka açık bazı veri setleri özetlenmektedir.

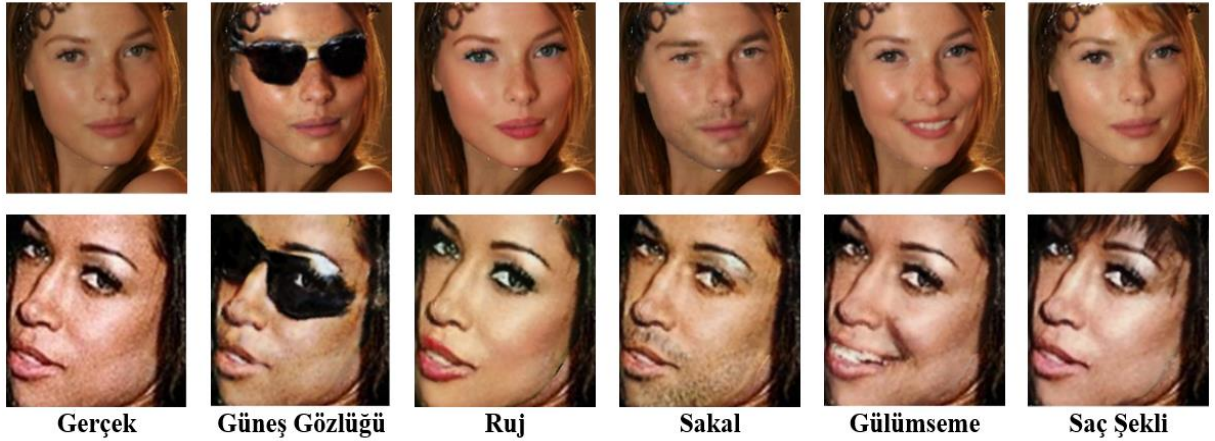
UADFV [15]: YouTube'dan indirilen 49 gerçek video ve bu videolardan FakeApp uygulaması kullanılarak elde edilen 49 manipüle edilmiř videodan oluřmaktadır. Her video, 294×500 piksel çözünürlüđe ve ortalama 11,14 sn. süreye sahip bir kiřiyi temsil etmektedir.

FaceForensics++ [1]: Bu veri seti, YouTube'da seçilen 1.000 gerçek video ve 1.000 manipüle edilmiř video içermektedir. Faceswap [36] uygulaması kullanılarak manipölasyonlar gerçekleřtirilmiřtir.

DFDC-preview [37]: Deepfake Tespit Yarışması (DeepFake Detection Challenge, DFDC) veri seti, Facebook tarafından yayınlanmıřtır. Veri seti 66 aktöre ait 1.131 gerçek video ve 4.119 sahte video içermektedir. Sahte videolar iki farklı yaklařım kullanılarak oluřturulmuřtur. Ancak bu algoritmaların detayları açıklanmamıřtır.

2.1.3. Yüz Özellikleri

Bu manipülasyon tekniği kategorisinde ten rengi, cinsiyet gibi yüz özellikleri değiştirilmektedir. Adobe Photoshop, AgeingBooth ve FaceApp bu tür manipülasyonlar için kullanılan popüler uygulamalardan bazılarıdır. Ayrıca, He vd. [38], kişinin kimliğini ve diğer yüz ayrıntılarını korurken sakal, yaş, saç rengi ve ağız yüz özelliklerini değiştirebilen attGAN adlı bir ÜÇA yapısı geliştirmişlerdir. Şekil 2.4'te farklı görüntülere uygulanan bu yöntemin sonuçları gösterilmektedir.



Şekil 0.4. Yüz özelliklerinin oluşturulması manipülasyon türü için gerçek ve sahte yüz görüntüleri örnekleri [39]

Yüz Özellikleri Oluşturma Yöntemleri ve Veri Setleri

Çoğu ÜÇA yapısının kodu araştırmacılar tarafından herkese açık olarak paylaşılmıştır. Bu tür ÜÇA tekniklerinden yararlanarak yüz özellikleri manipülasyon tekniği kategorisiyle ilgili bilinen birkaç veri seti mevcuttur. Chang vd. [39] yüz özellikleri ile ilgili manipülasyon oluşturmak için iki aşamalı bir teknik önermişlerdir. Bunlar Doku Tamamlama ÜÇA (Texture Completion GAN, TC-GAN) ve 3D özellik ÜÇA (3D Attribute GAN, 3DA-Gan)'dır. TC-GAN, eksik görünümü otomatik olarak kaldırır ve normalleştirilmiş bir UV dokusu sağlar. 3DA-GAN, maksimum korunan özne kimliği ile hedef nitelikleri oluşturmak için UV doku alanı üzerinde çalışır. Ayrıca, karmaşık resim değişikliği için Perarnau vd. [40], koşullu GAN ile birlikte toplu olarak kullanılan bir kodlayıcının bir kombinasyonuna dayanan tersine çevrilebilir koşullu ÜÇA (IcGAN) olarak adlandırılan bir yaklaşım sunmuşlardır. Bu yöntem, kişinin yüz kimliğini kritik bir şekilde değiştirmemesine rağmen, yüz özelliklerini değiştirmek için iyi sonuçlar vermiştir. Araştırma amacıyla kullanılabilecek bazı yüz özellikleri manipülasyon veri setleri halka açıktır. Bunlardan bazıları aşağıda detaylandırılmıştır.

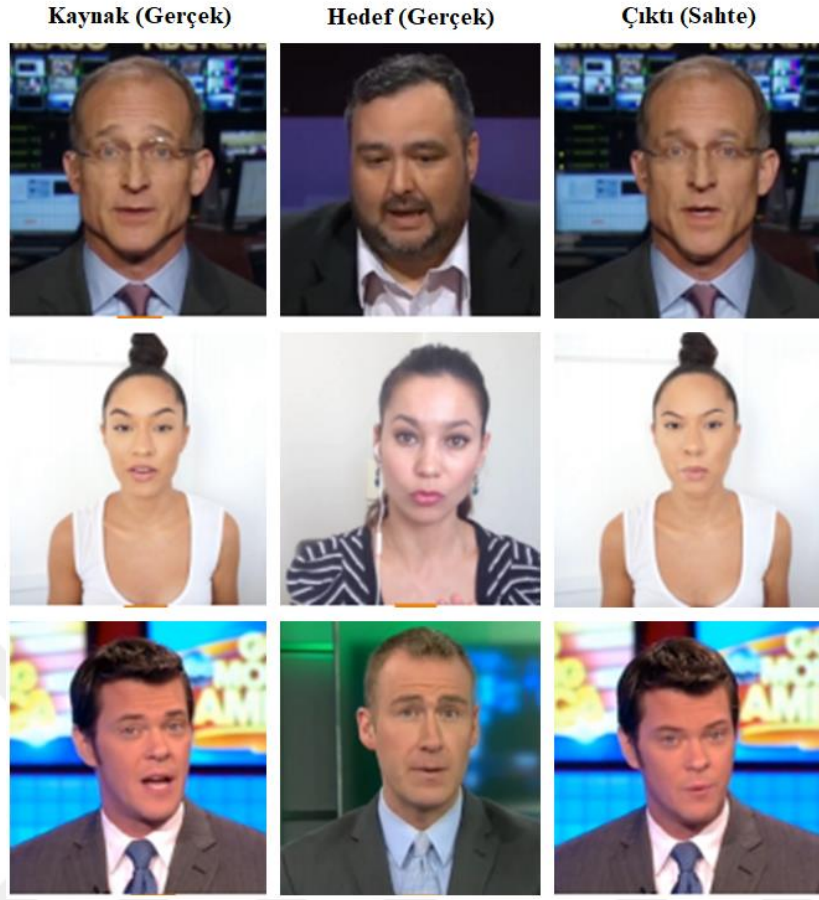
CelebA [16]: Büyük ölçekli bir yüz özelliği veri seti olan CelebFaces [38]'ten seçilen görüntülerin etiketlenmesiyle elde edilmiştir. Veri seti, 10.177 kişiden, beş belirli noktanın (landmark) konumu ve her görüntü için 40 adet ikili ilişkilendirme ile 202.000'den fazla yüz görüntüsünden oluşmaktadır.

PubFig [39]: Bu veri seti 200 kişiye ait 58.797 resimden oluşmaktadır. İnternette kontrolsüz koşullar altında elde edildiğinden farklı duruş ve ifade gibi görsel özelliklerde dikkate değer varyasyonlardan oluşmaktadır. 73 yüz özelliği sınıfı içermektedir.

Attribute 25K [41]: Bu veri seti, Facebook'tan toplanan 24.943 kişinin görüntüsünü içermektedir. Görüntüler; perspektif ve poz gibi özellikler açısından büyük farklılıklar gösterdiğinden, tüm özellikler her görüntü için etiketlenememektedir.

2.1.4. Yüz İfadeleri

Bu manipülasyon tekniği ile bir kişinin yüz ifadesi başka birinin yüz ifadesi ile değiştirilmektedir. Thies vd. [42] yüz ifadesi manipülasyonu için gerçek zamanlı videolar üzerinde çalışan bir teknik sunmuşlardır. Sunulan teknik, Face2Face olarak adlandırılmaktadır. Şekil 2.5'te çeşitli görüntülere uygulanan bu manipülasyon üretim yönteminin sonuçları gösterilmiştir.



Şekil 0.5. Yüz ifadeleri manipülasyon türü için FaceForensics ++ [1] veri setinde yer alan bazı örnek resimler.

Yüz İfadesi Oluşturma Yöntemleri ve Veri Setleri

Yüz ifadesi manipülasyonu türüne odaklandığı bilinen tek veri seti FaceForensics ++ [1]'dir. Bu veri seti FaceForensics'in [22] bir uzantısıdır. İlk başta, FaceForensics veri seti, yalnızca hedef kimliğini korurken kaynak kimlik ifadesini hedef kimliğe aktaran bir bilgisayar grafik yöntemi olan Face2Face'e odaklanmıştır. Bu işlem manuel olarak anahtar çerçeve seçilerek gerçekleştirilmiştir. Daha sonra her karenin kaynak ifadeleri hedef videoya aktarılarak sahte örnekler oluşturulmuştur. Aynı araştırmacılar FaceForensics ++'da sinirsel dokulara (NeuralTextures) dayanan yeni bir öğrenme yaklaşımını tanıtmışlardır [43]. FaceForensics++, bir işleme ağı kullanılarak hedef kişinin sinirsel görünümünü öğrenen gerçek video verilerini kullanan işleme tabanlı (rendering-based) bir yaklaşımdır. FaceForensics ++ araştırmacılar tarafından, uygulamalarında yama tabanlı olan Pix2Pix'te [44] kullanılan bir ÜÇA kaybı olarak değerlendirilmiştir. Bu yaklaşım ile sadece ağıza karşılık gelen yüz ifadesi değiştirilmiştir. Veri seti, YouTube'dan indirilen 1.000 gerçek videoyu içermektedir. Göz önünde bulundurulan her yaklaşım için her biri 1.000 olmak üzere toplam 2.000 sahte video yer almaktadır.

Yüz ifadelerini manipüle etmek için kullanılabilecek birkaç uygulama mevcuttur. Yüz ifadelerinin kolayca değiştirilmesine izin veren mevcut ÜÇA algoritmik yapılarına dayanan Face2Face bu uygulamalara örnek gösterilebilir. Benzer şekilde, StarGAN [45] yaklaşımı ile yazarlar, bir kişinin yüz görüntüsünün farklı ifadelerle değiştirilebileceğini göstermiştir. Bu manipülasyon oluşturma tekniği ile alakalı olarak araştırma amaçlı oluşturulan şu ana kadar bilinen tek veri seti FaceForensics ++ [1]'dir.

2.2. Deepfake Manipülasyonlarının Tespiti

Deepfake tespiti, belirli bir yüz örneğinden göze çarpan özelliklerin çıkarıldığı ve daha sonra ikili (binary) bir sonuç elde etmek için bir sınıflandırma yapısına verildiği, gerçek ve sahte olmak üzere iki sınıflı bir sınıflandırma problemi olarak düşünülebilir. Deepfake tespit teknikleri doku (textural), doğal özellik bozulmaları ve derin öğrenme tabanlı tespit yöntemi olmak üzere genel olarak üç ana sınıfa ayrılabilir.

2.2.1. Doku Tabanlı Yöntemler

Bu bölümdeki yaklaşımlar, verilen örneğin gerçekliğini öğrenmek için örneğin dokusal özelliklerini analiz etmektedir. Örneğin, Zahid ve Dasgupta [46] deepfake tespiti için Frekans Çözülmüş Yerel İkili Model (Frequency Decoded Local Binary Pattern, FDLBP), İkili İstatistiksel Görüntü Özellikleri (Binarised Statistical Image Features, BSIF), CENSus Transform Histogramı (CENTRIST) ve İkili Gabor Modeli (Binary Gabor Pattern, BGP) gibi on yerel görüntü özelliği tanımlayıcısının etkisini incelemişlerdir. Benzer şekilde, Scherhag vd. [47], manipüle edilen yüzleri tespit etmek için Yerel İkili Modeller (Local Binary Patterns, LBP), Gradyanların Histogramı (Histogram of Gradients, HOG) ve Ölçek Değişmeyen Özellik Dönüşümü (Scale Invariant Feature Transform, SIFT) gibi yerel tanımlayıcılar kullanmışlardır. Sun vd. [48], daha iyi tespit doğruluğu için frekans alanındaki kenar geometrik özelliklerin görüntü özellik kümesine tanıtıldığı ve mekansal alanda üst düzey anlamsal özelliklerle kaynaştırıldığı alanlar arası füzyona dayalı yeni bir deepfake tespit yöntemi önermişlerdir. Araştırmacılar, özellikle yüz kenar özelliklerini bulmaya odaklanmışlardır. Ancak önerilen yöntem dudak, göz, burun gibi özelliklere daha az odaklandığı için FaceForensics++ veri setinde yer alan sinirsel dokular veri kümesinde daha az tespit doğruluğu göstermiştir. Önerilen Yöntem, Afchar vd. [12] tarafından hazırlanan veri setinde ve FaceForensics++ veri setlerinde eğitilerek sırasıyla %90,50 doğruluk oranı ve en yüksek 0,97 AUC oranına ulaşmaya başarmıştır.

2.2.2. Doğal Özellik Bozulmaları Tabanlı Yöntemler

Bu algoritmalar, resim veya videonun kalite, gürültü gibi doğal özelliklerindeki değişiklikleri analiz eder. Örneğin, Zhang vd. [49], sensör desen gürültüsünün Fourier spektrumunun analizine dayanan manipüle edilmiş yüzler için bir yapı geliştirmiştir. Çiftçi vd. [50] deepfake tespiti için biyolojik sinyaller (kafa hareketine dayalı balistokardiyogram ve fotopletismografi) kullanmışlardır.

2.2.3. Derin Öğrenme Tabanlı Yöntemler

Bu kategorideki yaklaşımlar deepfake tespiti için yapay sinir ağlarını kullanmışlardır. Örneğin, Zhou vd. [13], yüz manipülasyon tespiti için temeldeki GoogleNet modeliyle [51] iki akışlı ESA'lar kullanmışlardır. İlk akış yüzdeki sahtelikleri tespit eden bir ağ iken ikinci akış eğitilmiş bir yama tabanlı üçlü ağıdır. Ağ mimarisi, yazarlar tarafından SwapMe ve FaceSwap uygulamalarının yardımıyla oluşturulan yayınlanmamış veri seti üzerinde eğitilmiştir. Önerilen yöntem %99,9 tespit doğruluğu elde etmiştir. Afchar vd. [12] ESA tabanlı bir tespit yöntemi önermişlerdir. Önerilen yöntem, birkaç katmana sahip derin bir sinir ağı kullanarak görüntülerin mezoskopik özelliklerine odaklanmıştır. Yöntem, yazarlar tarafından toplanan ve oldukça gerçekçi sahte videolardan oluşturulan veri setleri üzerinde eğitilmiştir. Yazarlar, Deepfake ve Face2Face veri setleri için sırasıyla %98,4 ve %95,3 doğruluk oranı elde etmişlerdir. Guera ve Delp [11], deepfake tespiti için zamana duyarlı bir ardışık düzen önermişlerdir. Çerçeve öznitelikleri, daha sonra Tekrarlı Bir Sinir Ağını (Recurrent Neural Network, TSA) eğitmek için kullanılan ESA kullanılarak çıkarılmıştır. Deneyin sonuçları, yazarlar tarafından toplanan bir veri setinde %97,1'lik doğruluk oranı rapor edilmiştir. Li ve Lyu [52] deepfake videolarını tespit etmek için yeni bir derin öğrenme tabanlı yaklaşım önermişlerdir. Önerilen yöntem, sentezlenen yüz bölgelerini ve komşuluk alanlarını bir ESA modeli ile eşleştirerek sahtelikleri tespit etmiştir. Özellikle, temel deepfake oluşturma algoritmalarında yeniden ölçeklendirme ve enterpolasyon işlemlerinin neden olduğu yüz manipülasyonlarını tespit etmek için 50 katmanlı artık ağ (ResNet-50) [53] modelini kullanmışlardır. UADFV [15] ve DeepfakeTIMIT LQ ve DeepfakeTIMIT HQ [10] veri setleri üzerinde deneyler yapılarak sırasıyla %97,4, %99,9 ve %93,2 doğruluk oranlarını elde etmişlerdir.

McCloskey ve Albright [54], gerçek ve sahte resimler arasında çeşitli manipülasyonların tespiti için ÜÇA veri hattını (pipeline) incelemişlerdir. Yazarlar, renk özelliklerine dayanan bir tespit yöntemi ve sınıflandırmak için doğrusal bir Destek Vektör Makinesi (Support Vector Machine, DVM) sunmuşlardır. Önerilen yöntem, NIST MFC2018 [55] veri setinde eğitilmiştir. Yazarlar, önerilen yaklaşım için ROC eğrisinin altında kalan alanı (Area Under Curve, AUC) %70,0 olarak elde etmişlerdir. Nataraj vd. [56], doğal görüntü istatistikleri ve steganalizden yararlanan bir tespit yaklaşımı önermişlerdir. Önerilen teknik, özellikle piksel kümelenme

matrisleri ve ESA'ların bir bileşiğine dayanmaktadır. Yöntemleri öncelikle CycleGAN yaklaşımı kullanılarak oluşturulan farklı nesnelerden oluşan bir veri kümesi üzerinde test edilmiştir. Ek olarak, araştırmacılar, iki GAN yapısı (CycleGAN ve StarGAN) aracılığıyla oluşturulan sahte görüntüler için önerilen yöntemin geçerliliğini iyi genelleme sonuçlarıyla gözlemlemek için dikkat çekici bir analiz yapmışlardır. Daha sonra Neves vd. [57] bu tespit yöntemini 100K-Faces veri setinde uygulamış ve %7,2'lik bir eşit hata oranı (Equal Error Rate, EER) sonucunu elde etmişlerdir.

Tariq vd. [58], yüz özellikleri manipülasyonlarını tespit etmek için bir ESA mimarisi önermişlerdir. VGG16 [59], VGG19 [59], Artık Sinir Ağı (Residual Neural Network, ResNet) [53] ve XceptionNet [60] gibi çeşitli ESA yaklaşımlarını kullanmışlardır. Önerilen mimaride gerçek yüz görüntüleri için CelebA veri seti kullanılmıştır. Sahte görüntüler için ÜÇA'lara dayanan makine yaklaşımları ve makyaj, gözlük gibi manipülasyonları içeren Adobe Photoshop CS6 tabanlı manuel yaklaşım olmak üzere iki farklı yaklaşım dikkate alınmıştır. Makine tarafından oluşturulan senaryo ve manuel olarak oluşturulan senaryolar için sırasıyla %99,99 ve %74,9 AUC oranlarını elde etmişlerdir. Stehouwer vd. [23], sınıflandırıcı modeli için özellik haritalarını kullanmak ve geliştirmek için dikkat teknikleriyle ESA'ya dayanan bir tespit yöntemi sunmuşlardır. Düşünülen dikkat haritası rahatlıkla uygulanabilmekte ve mevcut omurga ağlarına yerleştirilebilmektedir. Yazarlar tespit yöntemini DFFD veri seti üzerinde uygulamışlardır. Ele alınan yöntem ile özellikle yüz değişimi tespiti için %99,43 AUC ve %3,1 EER oranları elde edilmiştir. Bharati vd. [61], bir derin öğrenme yönteminin dijital rötuşla yüz görüntülerinin tespiti için kısıtlı Boltzmann makinesine bağlı olduğunu iddia etmişlerdir. Tespit yöntemi girişi, her bir resmi orijinal veya sahte olarak sınıflandırmak ve ayırt edici özellikleri öğrenmek için yüz yamalarını içermektedir. Deneyler, ND-IIITD veri seti (koleksiyon B) ve internetten indirilen çok sayıda ünlünün yüz resimlerini içeren iki farklı sahte yüz veri seti üzerinde gerçekleştirilmiştir. Oluşturulan veri seti için %96,2 doğruluk oranı ve ND-IIITD veri seti için ise %87,1 doğruluk oranı elde etmişlerdir.

Nguyen vd. [62], sahte yüz görüntülerini ve videolarını tespit etmek için bir VGG19 [59] ağına dayanan kapsül yapılarını [63] kullanmışlardır. Önerilen yöntem, FaceForensics ++ [1] üzerinde %97,05 doğruluk performansı sergilemiştir. Matern vd. [64], sahte yüzlerin gözlerinde, dişlerinde ve yüz çevrelerinde manipülasyonların tespitine dayanan bir tespit yaklaşımı sunmuşlardır. Çok katmanlı ileri beslemeli sinir ağı ve lojistik regresyon (LogReg) model sınıflandırıcıları olmak üzere iki farklı yaklaşım kullanılmıştır. YouTube'dan deepfake videolarından ve CelebA veri setinden alınan görüntüler kırılarak yayınlanmamış bir veri seti elde edilmiştir. Önerilen yöntem ile 0,866 AUC oranı elde edilmiştir. Dolhansky vd. [18] DFDC veri seti aracılığıyla üç basit tespit yaklaşımı kullanarak temel sonuçlar elde etmiştir. Düşük seviyeli görüntü manipülasyonlarını tespit etmek için ESA mimarisini kullanmışlardır. Ardından, yalnızca yüz çerçeveleri kullanılarak bir XceptionNet ağı eğitilmiştir. Son olarak, görüntünün tamamı kullanılarak başka bir XceptionNet ağı eğitilmiştir. Önerilen yöntem ile, %93,0 hassasiyet

(precision) ve %8,4 geri çağırma (recall) oranları elde edilmiştir. Sabir vd. [65], akışta var olan geçici verilere dayanan sahte örneklerin tespiti için bir yöntem sunmuşlardır. Önerilen yöntemin arkasındaki anlayış, çerçeveler arasındaki zamansal farklılıklardan yararlanmaktır. Bu nedenle, önceden eğitilmiş bir model kullanmak yerine, uçtan uca eğitilmiş benzer tekrarlı bir evrimsel ağ kullanmışlardır. Yazarlar FaceForensics ++ veri seti üzerinde deneylerini gerçekleştirmişlerdir. Veri setindeki düşük kaliteli videoları kullanarak yüz değiştirme ve deepfake sınıfları için sırasıyla %96,3 ve %96,9 AUC oranlarını elde etmişlerdir.

Agarwal vd. [66], kafa hareketlerine ve yüz ifadelerine dayanan bir tespit yaklaşımını önermişlerdir. Görüntülerdeki öznitelikleri çıkarmak için OpenFace2 araç takımını [67] kullanarak yoğunluk ve oluşum ile 18 farklı yüz eylemi elde etmişlerdir. Ayrıca baş hareketleri için dört öznitelik göz önünde bulundurulmuştur. Son sınıflandırma için DVM kullanılmıştır. Araştırmacılar, YouTube'dan videolar indirerek kendi veri setlerini oluşturmuşlardır. En iyi performans için %96,3 AUC oranını elde etmişlerdir. Nguyen vd. [68], manipüle edilmiş resimleri veya videoları eşzamanlı olarak tespit etmek ve verilen resimde veya videoda manipüle edilen bölgeleri bulmak için çok görevli bir öğrenme ESA modeli tasarlamışlardır. FaceForensics ve FaceForensics ++ veri setleri için sırasıyla %92,77 ve %92,50 sınıflandırma doğruluk oranlarına ulaşmışlardır. Amerini vd. [69] iki aşamalı bir yöntem önermişlerdir. İlk olarak, uygulanabilir çerçeveler arası farklılık özellikleri için optik akış alanları [70] uygulamışlardır. Daha sonra bu öznitelikler ESA sınıflandırıcılarını beslemiştir. ESA eğitiminden sonra test için Resnet-50 [53] ve VGG16 [59] kullanılmıştır. Önerilen model, FaceForensics++ veri setinde Resnet-50 ve VGG16 için sırasıyla %75,46 ve %81,61 doğruluk oranına ulaşmıştır. Yang vd. [15], bir DVM modeliyle kafa pozlarındaki tutarsızlıkları inceleyerek deepfake videolarını tespit etmişlerdir. Videolara ve resimlere dayalı alıcı çalışma karakteristik eğrisi altındaki alanlar, UADFV veri seti kullanılarak sırasıyla 0,974 ve 0,89 olarak elde edilmiştir. Jung vd. [71], göz kırpma davranışını değerlendiren DeepVision adlı el yapımı bir deepfake tespit yöntemi sunmuşlardır. İlk olarak, potansiyel bir deepfake videosundan yüz bölgesini çıkartmışlardır. Daha sonra, bir kişinin göz bölgesini tespit etmek için bir göz takip cihazı kullanmışlardır. Bu adımdan sonra, her çerçevenin göz bölgesini kapalı veya açık gözler için kontrol edilerek geçen göz kırpma sürelerini ve göz kırpma periyotlarını hesaplayarak deepfake videoları tespit etmişlerdir.

Liu vd. [72], basit bir Üç Boyutlu (Three Dimensional, 3B) ESA önermiştir. Kanal dönüştürme modülü, çok daha az parametrelili özellikleri daha üst düzeyde çıkarmak için tasarlanmıştır. Mekansal-zamansal modül olarak hizmet veren 3B ESA'lar, mekansal özellikleri zaman boyutunda birleştirmek için benimsenmiştir. Çerçeve içeriğini önlemek ve çerçeve dokusunu vurgulamak için, mekansal-zamansal modülün daha iyi performans elde etmesine yardımcı olan giriş çerçevelerinden Mekansal Zengin Model (Spatial Rich Model, SRM) özellikleri çıkartılmıştır. Önerilen yöntem için yazarlar, FaceForensics++ DeepfakeTIMIT, DFDC ve Celeb-

DF veri setlerini kullanmışlardır. Önerilen yöntemde FaceForensics++ veri seti için %99,83 DeepfakeTIMIT HQ ve LQ veri setleri için sırasıyla %99,28 ve %99,60 DFDC veri seti için %93,98 son olarak Celeb-DF veri seti için %98,07 doğruluk oranları elde edilmiştir. Stanciu ve Ionescu [73], seçilen 3 yüz bölgesini kullanarak deepfake tespiti için mekansal-zamansal ESA ve UKSB yaklaşımını sundular. Bu bölgelerde ayrı ayrı çalıştırılan modelin performansını geç füzyon kombinasyonu ve 2 son teknoloji veri setinde tüm yüzü kullanan modelle karşılaştırmalar yapılmıştır. Önerilen yöntem ile %97,06 ve %99,95 AUC oranları sırasıyla Celeb-DF ve FaceForensics++ veri setleri için elde edilmiştir. Pranjal vd. [74], videoları sahte veya orijinal olarak sınıflandırmak için ESA ve UKSB kombinasyonuna dayalı bir sistem önermişlerdir. Faceforensics++, Celeb-DF, DFDC veri setlerini kullanmışlardır. Tek veri setli eğitim modellerinden özel test setinde en iyi performans gösteren, %70,57 doğrulukla DFD modelidir, birleşik eğitim bölünmüş modelde ise %73,20'ye doğruluk oranına ulaşmıştır. En kötü performans gösteren çapraz test çifti, DFD'de test edilen Celeb-DF modelidir. DFD üzerinde eğitilen ve DFDC üzerinde test edilen model için en yüksek çapraz test doğruluğu %66,23'tür. Birleşik test seti performansı, tek dağıtım tabanlı model arasında %74.03 ile DFD modeli için en yüksektir ve birleşik eğitim modeli, %86.49 ile en yüksek sınıflandırma doğruluğunu kaydetmiştir. Bu, XceptionNet'in, üç dağıtımda da mevcut olan çeşitli manipülasyonları öğrenebildiğinden ve etkili, birleşik test doğruluğundan dolayı hala fazla uygun olmadığı için geniş bir öğrenme kapasitesine sahip olduğunu göstermektedir. Yazarlar, sistemin doğruluğunu artırmak için sistemlerinde transfer öğrenmeyi kullanmışlardır.

Chen vd. [75], videolardaki mekansal ve zamansal bilgileri aynı anda keşfedebilen birleşik bir mimari olan FSSPOTTER'ı geliştirdiler. Mekansal Özellik Çıkarıcı (Spatial Feature Extractor, SFE) önce videoları her biri belirli sayıda çerçeve içeren birkaç ardışık kliplere bölmektedir. SFE, girdi olarak klipleri alır ve çerçeve düzeyinde özellikler oluşturur. Çerçevelerdeki mekansal özellikleri çıkaran omurga ağı olarak toplu normalleştirme (batch normalization) ile VGG16'nın evrişim katmanları kullanılmıştır. Ayrıca, omurga ağını daha fazla özellik çıkarmaya teşvik etmek için süper piksel bazında ikili sınıflandırma biriminden (superpixel-wise binary classification unit) yararlanılmıştır. Geçici özellik toplayıcı (temporal feature aggregator), çerçevedeki zamansal tutarsızlıkları bulmak için çift yönlü bir UKSB kullanılmıştır. Ardından, klibin gerçek veya sahte olma olasılıklarını hesaplamak için tamamen bağlı bir katman ve bir softmax katmanından yararlanılmıştır. Değerlendirme için FaceForensics++, DeepfakeTIMIT, UADFV ve Celeb-DF veri setleri kullanılmıştır. Önerilen yöntem ile yazarlar, FaceForensics++ için %100 UADFV için %91,1 Celeb-DF için %77,6 ve DeepfakeTIMIT HQ, LQ için %98,5, %99,5 AUC oranlarını elde ettiğini öne sürmüştür.

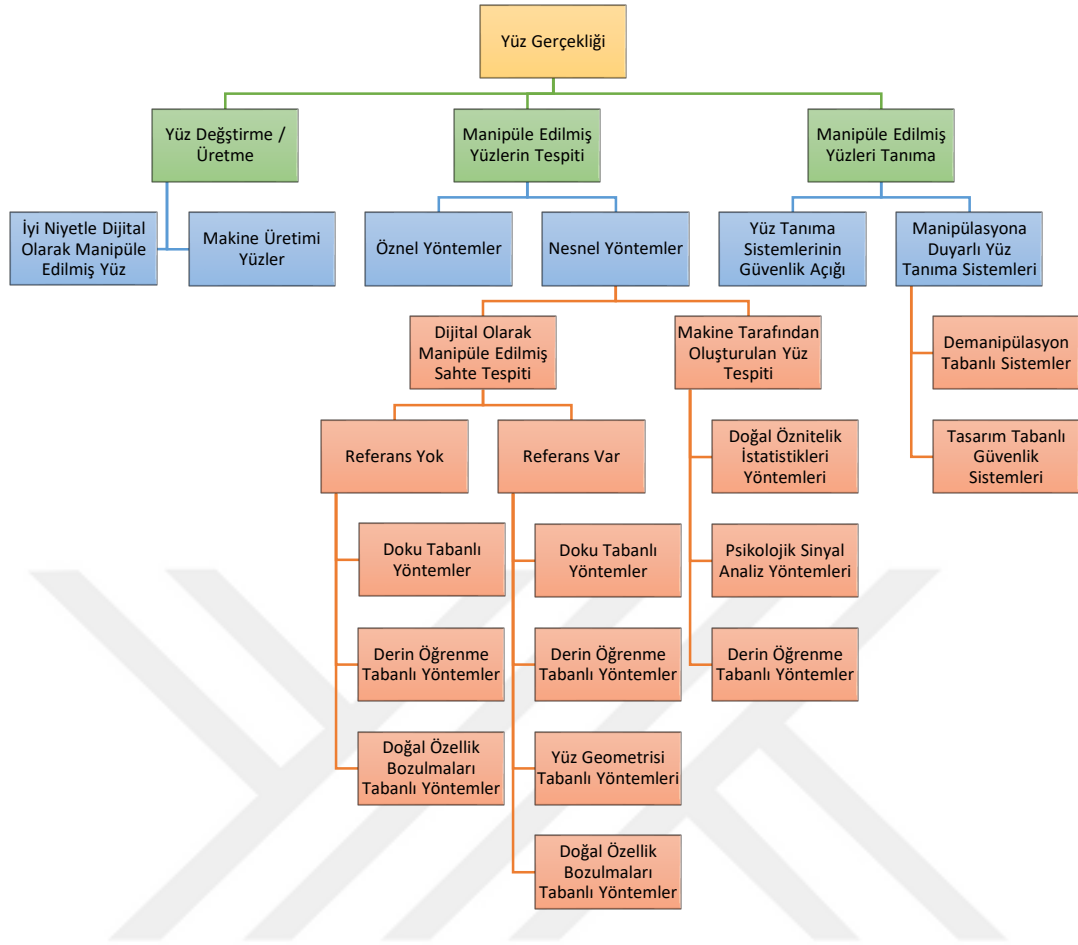
Luo vd. [76], çerçevede bulunan yüzleri algılayabilen bir sistem önermişlerdir. ESA algoritmasındaki sorun, görüntünün sabit bir boyuta dönüştürülmesi ve girdi olarak verilmesidir,

ancak bu durum görüntüde yer alan bilgilerin kaybolmasına neden olmaktadır ve sistemin genel doğruluğunu etkilemektedir. Yazarlar, bilgi kaybı sorununu ortadan kaldırmak için evrişim katmanı ile tam bağlı katman arasındaki mekansal piramit havuzlamayı tanıttılar. Önerilen sistemde kullanılan veri setleri CelebA, AFW [77] ve FDDB [78]'dir. Önerilen yöntemde görüntü önce 11x11 ızgaraya bölünmektedir ve her ızgarada iki tahmin çerçevesi bulunmaktadır. Önerilen sistemin ayrık doğruluğu %95, sürekli doğruluğu için ise %74'tür. Önerilen yöntem, belirli faktörler nedeniyle ESA algoritmasının doğruluğunun nasıl düşebileceğine dair bir anlayış sağlamıştır. Kullanılan sistem aynı zamanda ESA'nın doğruluğunun düşmemesi için kullanılabilecek çözümü de sağlamıştır. Ayrıca, bu sistem, sistemden maksimum doğruluk elde etmelerini sağlayan gelişmiş kayıp fonksiyonuna sahiptir. Tao vd. [79], daha iyi sonuçlar elde etmek için uygun çerçeve hizalaması ve hareket dengelemenin yapılması gerektiğini vurgulayan bir sistem önerdiler. Yazarlar, bir ESA mimarisinde bir Alt Piksel Hareket Dengeleme Katmanını (Sub-Pixel Motion Compensation, SPMC) tanıttılar. FlowNet-S CNN ile birlikte, çerçeve hizalama ve hareket dengeleme, Hareket Dengeleme Transformatörü (Motion Compensation Transformer, MCT) modülü kullanılarak elde edilmektedir. Araştırmacılar, daha iyi sonuçlar elde etmek için çoklu çerçeve girdilerinin nasıl organize edileceğine dair bir fikir vermişlerdir. Ayrıca, verilerin ESA modeline beslenmeden önce nasıl örnekleneceği hakkında bir ön bilgi verilmiştir. Ayrıca, internette herkese açık olarak bulunan yüksek kaliteli 1080p HD video kliplerden 975 dizi topladılar ve orijinal çerçeveleri 540×960 piksele indirdiler. Önerilen yöntemi ile 36.71'lik bir Tepe Sinyal-Gürültü Oranına (Peak Signal-To-Noise Ratio, PSNR) ve 0.96'lık bir Yapısal Benzerlik İndeksi (Structural Similarity Index, SSIM) değerine ulaşılmıştır. Yu vd. [80], yüz değişiklikleri olan videolar için düşük doğruluk oranı elde eden süper çözünürlük algoritmasının bir dezavantajına odaklanan bir sistem önerdiler. Süper çözünürlük için önce her video birden çok çerçeveye bölünerek ve her çerçeveye farklı şekilde ESA uygulanmıştır. Yazarlar, videonun yüz ifadesi değiştiğinde, süper çözünürlüklü sistemler için daha yüksek doğruluklar üretmenin çok zor olduğunu fark etmişlerdir. Bu sorunun üstesinden gelmek için tanıtılan sistem ile bir yüz görüntüsünün esnek bir şekilde nasıl ele alınabileceğini önerilmiştir. Önerilen sistem küresel izleme, yerel hizalama, süper çözünürlük algoritması olmak üzere üç adımda çalışmaktadır. Sistemin performans ölçümü için yazarlar, yaklaşık 3 dakika süren farklı 10 kişiye ait 10 video dizisi kaydetmişlerdir. Önerilen sistem, yüz değişiklikleri olduğunda süper çözünürlük algoritmasının karşılaştığı sorunlar hakkında fikir vermiştir. Eğer yüz ifadesi değişiklikleri süper çözünürlüğün doğruluğunu engellerse bu durum sistemi çok kötü etkileyecektir çünkü süper çözünürlük aşamasının çıktısı sistemde ESA aşamasına girdi olarak verilmektedir. Bu problemin üstesinden gelmek için yüz görüntüsünün esnek bir şekilde ele alınmasını kullanan bir çözüm sunulmuştur. PSNR değeri de bu yaklaşım kullanılarak mevcut çalışmalara göre artmıştır. PSNR değeri 26,6261 olarak bulunmuştur.

Birkaç deepfake manipölasyon tespit yöntemi için, kullanılan yöntem kullanılan veri setleri ve önerilen yöntemin performansının kısa bir özeti Tablo 2.1'de verilmiştir. Ayrıca [2]'de de yer alan deepfake oluşturma, tespit ve tanıma taksonomisi Şekil 2.6'da yer almaktadır.

Tablo 2.1. Bazı deepfake manipölasyon tespit yöntemlerinin özeti. Burada AUC, ROC eğrisinin altında kalan alanı, D.O. ise doğruluk oranını temsil etmektedir.

Çalışma	Yöntem	Veri Seti	En İyi Performans
Guera ve Delp [11]	TSA	Özel veri seti	D.O. = %97,1
Afchar vd. [12]	Tasarlanmış ESA	Özel veri seti	D.O. = %98,4 AUC = %97,4
Zhou vd. [13]	GoogleNet	Özel veri seti	AUC = %99,9 AUC = %93,2
McCloskey ve Albright [54]	ÜÇA veri hattı (pipeline) Özellikleri	NIST MFC2018	AUC = %70,0
Nguyen vd. [62]	Kapsül Ağı	FaceForensics++	D.O. = %97,05
Matern vd. [64]	ESA ve Lojistik Regresyon	Özel veri seti	AUC = %86,6
Sabir vd. [65]	ESA + TSA	FaceForensics++ [1]	AUC = %96,9
Agarwal vd. [66]	OpenFace2 + DVM	Özel veri seti	AUC = %96,3
Amerini vd. [69]	ESA	FaceForensics++	D.O. = %81,61



Şekil 0.6. Deepfake oluşturma, tespit ve tanıma taksonomisi [2]

3. ÖZNİTELİK ÇIKARMA VE SINIFLANDIRMA AŞAMALARINDA KULLANILAN YÖNTEMLER

Bu tez çalışmasında öznitelik çıkarmak için ESA tabanlı derin öğrenme yaklaşımları kullanılmıştır. Sınıflandırma işlemleri için ESA tabanlı yaklaşımlar ile birlikte popüler sınıflandırma algoritmaları kullanılmıştır.

3.1. Yapay Sinir Ağları (YSA)

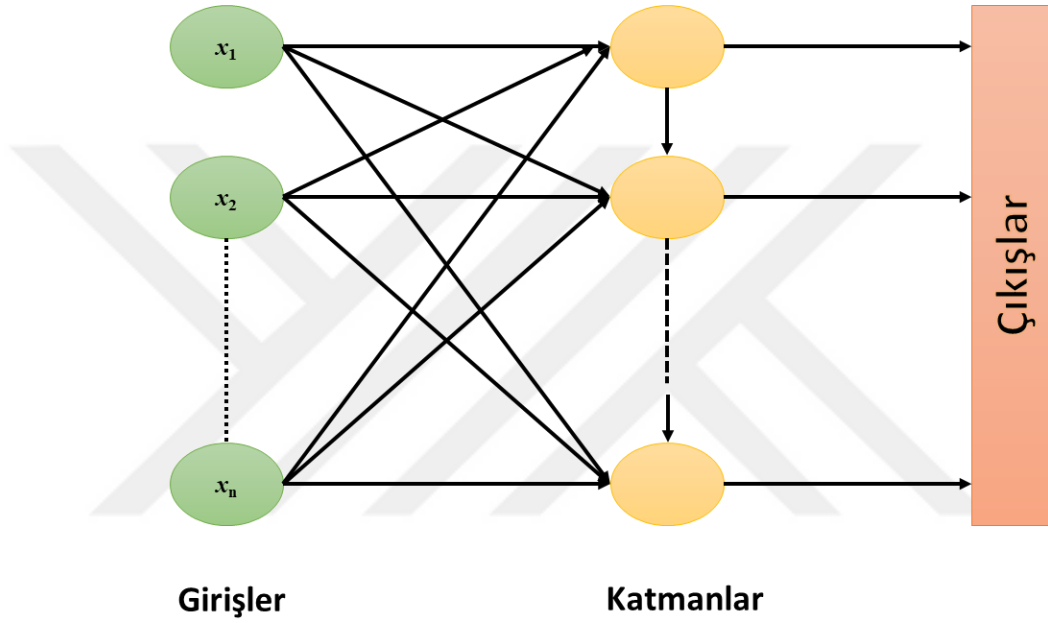
Beyinde bilgilerin öğrenimi veya muhafaza edilmesi sinir hücreleri arasında meydana gelen kimyasal ve elektriksel tepkimeler ile gerçekleşmektedir. Sinir hücresi ya da nöron, sinir sisteminde beynin temel yapı taşlarını oluşturan ve bilgi taşıma görevini yerine getirmektedir. Bir nöron dentrit, akson ve hücre gövdesi olarak adlandırılan üç bölümden oluşmaktadır. Bilgi nöron üzerinde elektriksel bir sinyal olarak aktarılmaktadır. Nöron üzerinde iletilen bu elektrik sinyali aksiyon potansiyeli olarak adlandırılmaktadır. Dentritler, hücre gövdesine gelen bütün girişleri toplayarak hücre gövdesinde birleştirir. Dentritler tarafından gerçekleştirilen bu birleştirme işleminden sonra oluşan aksiyon potansiyeline bakılarak bir çıkış sinyalinin üretilmesine veya üretilmemesine karar verilir. Eğer bir çıkış sinyali üretilecekse aksiyon potansiyeli aksonlar üzerinden taşınarak başka nöronlara aktarılır. Nöronlar arasındaki bağlantılar sinaps olarak adlandırılan özelleşmiş bağlantı noktalarıyla gerçekleştirilir.

Bir YSA'yı bilgisayar ortamında elde etmek için yukarıda bahsedilen nöronların temel bölümlerinin işlevleri düşünülerek bahsi geçen bölümlerin aslına en uygun biçimde taklit edilmesi gerekmektedir. Literatürde ilk yapay sinir hücresi McCulloch ve Pitts [81] tarafından sunulmuştur. Yalın bir elektrik devresi ile girişlerin belli bir eşik değerini aşması ile aktif hale gelen sinir hücresini oluşturmuşlardır. Hebb [82], 1949 yılında sunduğu çalışmasıyla nöronların algılama düzeylerinin kullanımlarına ve beraber faal gelmelerinden dolayı olumlu yönde bir artış kaydettiğini ispatlamıştır. Yapılan bu çalışmaların ışığında Rosenblatt ilk algılayıcının (perceptron) modellemesini gerçekleştirmiştir. İkili sınıflandırma problemleri için önerilen bu modelde girişler ve ağırlıklar önerilen öğrenme yöntemleri ile beraber görüntü sınıflandırma problemlerinde kullanılmıştır. Minsky ve Papert [83], en basit sinir ağı yapısını tanıtmışlardır. Önerilen mimaride sinir ağı tek katman algılayıcı bir girdi ve bir katman çıktı düğümleri içerirken sonradan yapılan çalışmalarda tek katman algılayıcıların yalnızca doğrusal sınıflandırma görevlerinde başarılı olduğunu ortaya çıkarmıştır [84]. Giriş olarak tek katmalı algılayıcılara verilen özellikler ağı sınıflandırma yeteneğini önemli bir ölçüde etkilemektedir. Tek katmanlı algılayıcılarda bulunan bu sorunu çözmek için doğrusal olmayan gizli katmanlar sinir ağı

modellerine eklenmiştir [85]. Böylece çok katmanlı ağ mimarileri veya derin ağ mimarileri ortaya çıkmıştır.

3.1.1. Tek Katmanlı YSA

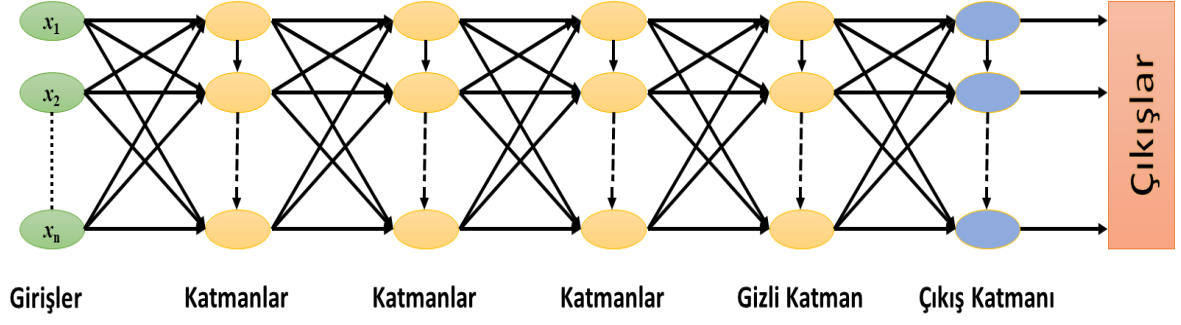
Tek katmanlı YSA'da tüm nöronlar tek katmanda ve aynı düzeyde bulunmaktadır. YSA, girdi sinyallerini alarak nöronlara iletir. Buradan da çıktı sinyalleri üretilir. Şekil 3.1'de tek katmanlı YSA'nın mimarisi gösterilmektedir.



Şekil 0.1. Tek katmanlı YSA mimarisi

3.1.2. Çok Katmanlı YSA

Çok katmanlı YSA'da aynı girdileri kullanan paralel nöron düzeninde çoklu katmanlar kullanılmaktadır. Bu ağ mimarisinde gerçek bilgileri temsil etmek için oluşturulan bir fonksiyona yaklaşımda oldukça başarılıdır. Ek olarak, bu ağlar çok katmana sahip olduğu için doğrusal olmayan veriler için de kullanılabilir [86]. Şekil 3.2'de çok katmanlı YSA mimarisi yer almaktadır.



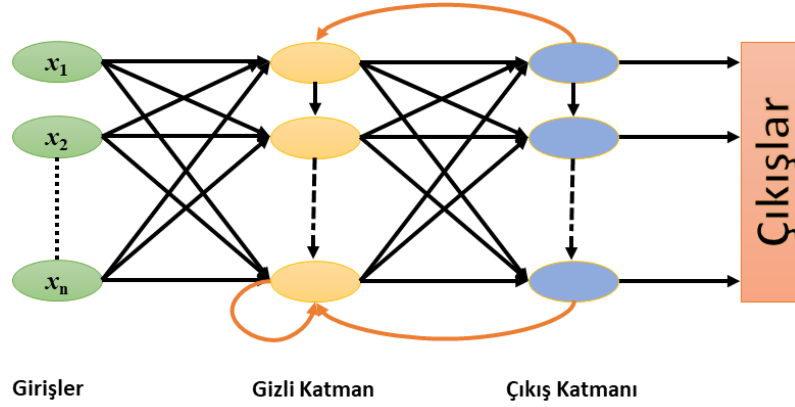
Şekil 0.2. Çok katmanlı YSA mimarisi

3.1.3. İleri Beslemeli YSA

İleri beslemeli YSA'da sinyal iletimi tek bir doğrultuda veya tekrarlanabilir bir şekilde gerçekleştirilebilmektedir. İlk olarak girdi sinyalleri giriş katmanına iletiildiğinden ileri beslemeli YSA olarak adlandırılmaktadır. Çok katmanlı YSA mimarisinde olduğu gibi giriş sinyalleri bir sonraki katmana iletilir.

3.1.4. Geri Beslemeli YSA

Geri beslemeli YSA mimarilerinde ileri beslemeli YSA mimarilerinden farklı olarak bir nöron çıkış sinyali yalnızca kendisinden sonra gelen nöron katmanına giriş sinyali olarak verilmemektedir. Geri beslemeli YSA, ağda bir iç tekrarlamamanın olması, iletilen sinyallerin daha önce sinyali almış ve işlemiş bir nörona ya da katmana tekrar iletiildiği ağlardır. Ağda geri beslemenin kullanılmasının temel sebebi ağın öğrenme kabiliyetini güçlendirmek için bir iç hatırlatma gerektiren zaman serileri veya örüntü tanımayı gerçekleştirmek için dinamik bir davranış sergilemesini sağlamaktır. Şekil 3.3'te temsili geri beslemeli YSA mimarisi gösterilmektedir.



Şekil 0.3. Geri beslemeli YSA'nın temsili gösterimi

3.2. YSA Öğrenme Yöntemleri

YSA'da diğer makine öğrenmesi yöntemlerinde benzer şekilde iki çeşit öğrenme vardır. Bunlar denetimli ve denetimsiz öğrenmedir. İnsan beyninin öğrenme şekli de bu biçimdedir. İnsanlar, herhangi bir hedef gözetmeden gözlemlerinden bilgi sahibi olabilmekte veya bir öğreticiden öğrenerek bilgi sahibi olabilmektedir.

Denetimli öğrenme, görüntü sınıflandırma, nesne tanıma, nesne tespiti gibi tanımı belli olan desene sahip görevler için uygun olmaktadır. Denetimli öğrenme yaklaşımında YSA mimarisine hem girdi hem de çıktıya bağlı değerler için veri sağlanmalıdır. Bu veriler genellikle etiket olarak adlandırılmaktadır.

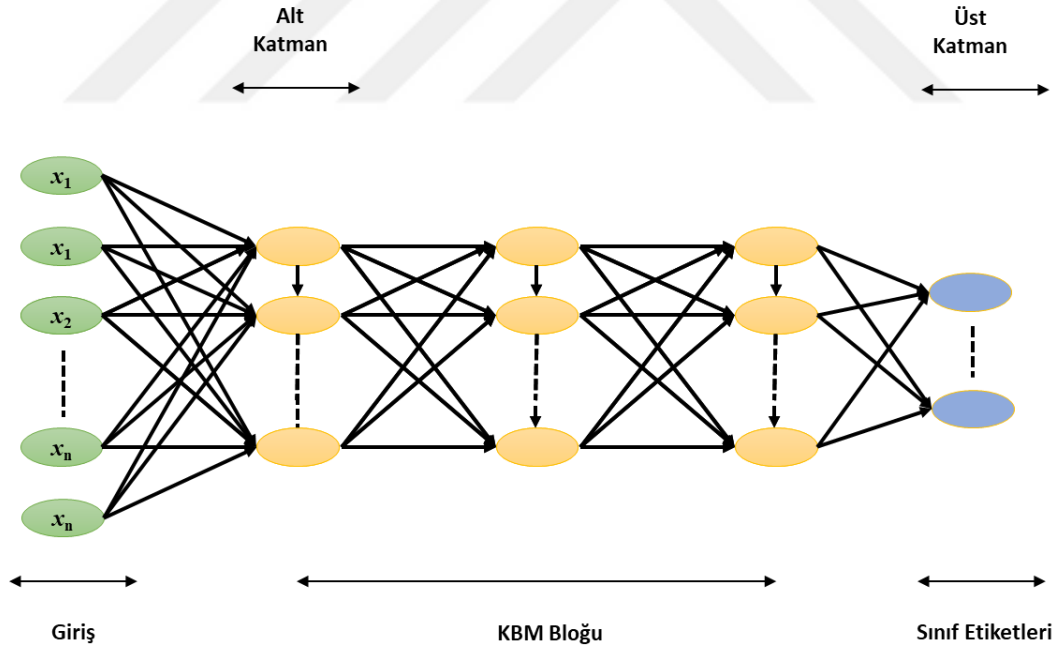
Denetimsiz öğrenmede ise herhangi bir sınıfı veya etiketi olmayan verilerle ilgilenilmektedir. Bu yaklaşımda herhangi bir etikete veya sınıfa sahip olmayan giriş verileri alınarak bilgi sahibi olunmaya çalışılmaktadır. Görüntüden özniteliklerin çıkartılması, kümeleme gibi işlemler denetimsiz öğrenmeye örnek gösterilebilir.

Takviyeli öğrenme (Reinforcement Learning), denetimsiz öğrenme metotlarından biri olmakla beraber etiketleri bilinmeyen verileri bir ödül fonksiyonu vasıtasıyla doğru ve yanlış seçimi yaparak veriyi öğrenebilmektedir. Temelinde deneme yanılma öğrenme yöntemi olan takviyeli öğrenme, seçimlerini geçmişte öğrendiği bilgileri yeni seçimlere dayandırarak seçim yapmaktadır [87]. Bu bölümde sıklıkla kullanılan YSA öğrenme yöntemlerinden Derin İnanç Ağları (Deep Belief Network, DİA), Kısıtlı Boltzmann Makineleri (Restricted Boltzmann Machine, KBM) ve Tekrarlı Yapay Sinir Ağları (Recurrent Neural Networks, TSA) tanıtılacaktır.

3.2.1. Derin İnanç Ağları

DİA, derin sinir ağı mimarisinde evrişim olmayan başarılı ilk yapılarından biridir. DİA mimarisi tanıtılmadan önce derin mimarilerin optimizasyonu oldukça güç bir sorun teşkil etmekteydi. DİA'lar MNIST [28] veri seti üzerinde oldukça iyi sonuçlar elde etmiştir [88].

KBM, DİA'nın temel bloğunu oluşturmaktadır. Gibbs örnekleme ve yeniden yapılandırma hatasını temelinde barındıran Zıt Iraksama (Contrastive Divergence) [89] algoritması kullanılarak KBM'nin ağırlıkları eğitilmektedir. DİA'da bulunan katmanlardan olan gizli katmanlar ikili değerli, görünür katmanlarsa ikili ya da gerçek değerli olabilmektedir. Bu ağlarda arabirim bağlantıları bulunmamaktadır. Katmanlarda bulunan her birim ekseriyetle komşu katmanda yer alan birimlere bağlanmaktadır. Bununla birlikte daha seyrek bağlı DİA'lar da oluşturulabilmektedir. DİA mimarisinde yer alan en üst iki katman arasında bulunan bağlantılar doğrudan yapılmamaktadır. Geriye kalan katmanlar arasında bulunan bağlantılar yönlendirilmektedir. Oklar, verinin en yakındaki katmana işaret etmektedir. Son yıllarda, DİA'lar popülaritesini kaybederek diğer öğrenme yaklaşımlarına nazaran kullanımı daha az tercih edilmektedir [90,91]. Şekil 3.4'te DİA mimarisinin temsili gösterimine yer verilmiştir.



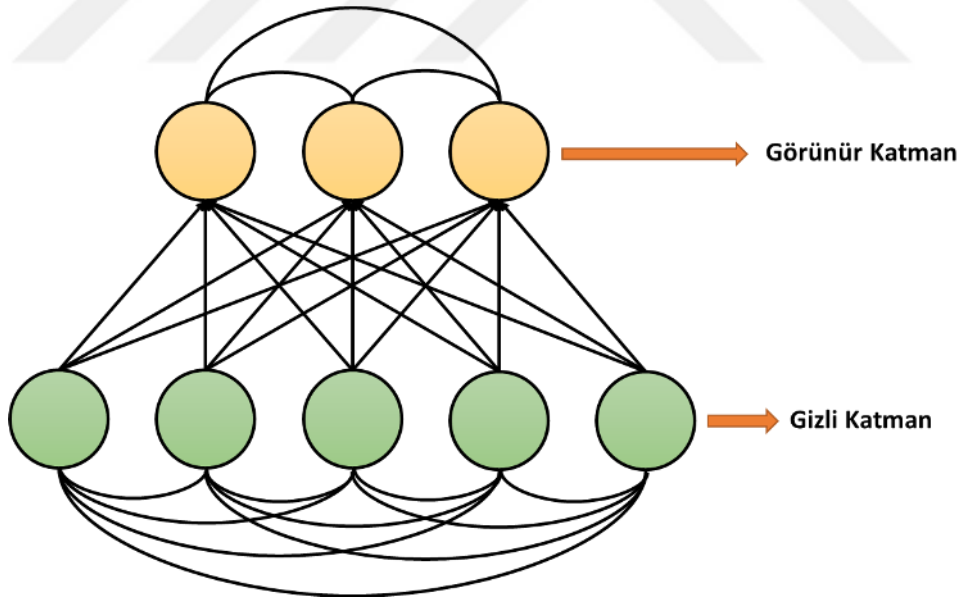
Şekil 0.4. Derin İnanç Ağının temsili gösterimi

3.2.2. Kısıtlı Boltzmann Makineleri

DİA'ların katman – katman eğitiminde kullanılan yöntemle Kısıtlı Boltzmann Makinesi denilmektedir. Boltzmann Makinesi (BM) bu yöntemin temel bir şeklidir. KBM, BM'nin özelleştirilmiş bir yapısı olup her iki yöntemde Hinton tarafından sunulmuştur [88,92].

BM, stokastik sonuçlar üreten ve yapay sinirlerden meydana gelen simetrik bağlantılı ağ yapısıdır. BM'ler pozitif ve negatif olmak suretiyle iki fazdan oluşan ve özyinelemeli yapısı bulunmaktadır. Pozitif fazda YSA'ya girdi katmanından girdi verileri verilirken negatif fazda herhangi bir girdi verisi verilmemektedir. BM, karmaşık verilerdeki gizli öznelilikleri bulmak için kullanılan bir YSA modelidir.

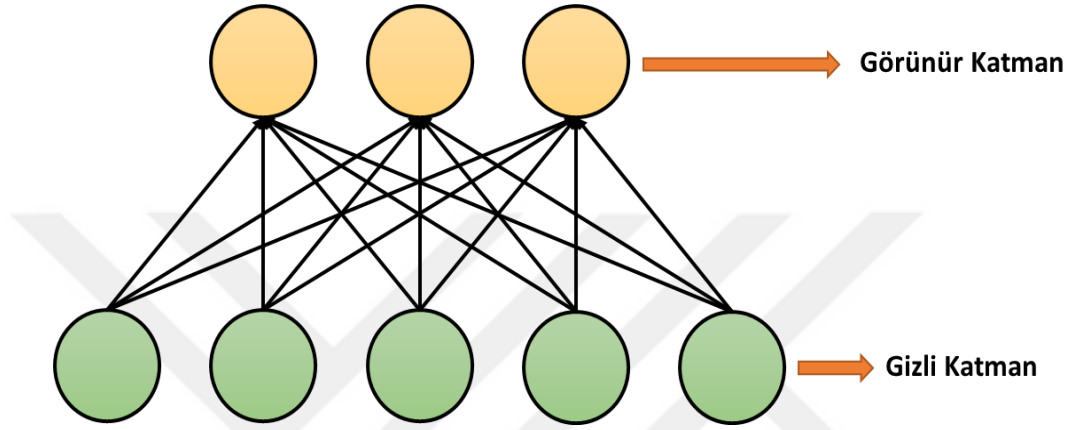
BM'lerin mimarisi klasik sinir ağlarında farklı olmamakla beraber çalışma metodolojisinde önemli bir özellik barındırmaktadır. YSA mimarisinde enerji kavramını benimsemesi BM'nin önemli olan özelliğidir. BM'nin yapısında bulunan her bir birim bir stokastik durum barındırmaktadır. Bu birimlerin durumu bağlı olduğu ağların enerjileri hesap edilerek belirlenmektedir. Şekil 3.5'te BM'nin yapısı yer almaktadır. Üst kısımda yer alan nöronlar aynı zamanda görünür katman şeklinde de adlandırılan girdi katmanını temsil ederken alt kısımda yer alan nöronlar gizli katmanları ifade etmektedir.



Şekil 0.5. Boltzmann Makinesi mimarisinin temsili gösterimi

BM'ler kullanılmaya başlandığı zamanlarda performansı oldukça iyi iken BM'ler pratikte kullanıldıkça farklı sorunlar ortaya çıkmaya başlamıştır. Bu problemlerin başında BM'lerin ağ yapısında bulunan bütün ağların tamamen birbirine bağlı olmasından dolayı uzun bir işlem

zamanının alması gelmektedir. Bilinmeyen olasılık dağılım yönlerinin öğrenilmesi güç olduğundan dolayı mimarinin verimini arttırmak ve ağın topolojisini sadeleştirmek için KBM geliştirilmiştir. KBM, BM kullanımını azaltarak daha kısa süreler içerisinde sorunları aşan bir yapı haline gelmiştir. KBM, iki farklı katmandan oluşmaktadır. Bunlar; görünür düğümlerin bulunduğu ve gizli düğümlerin bulunduğu katmanlardır. KBM de tıpkı BM yapılarında olduğu gibi ağın enerjisini temel almaktadır. KBM temsili mimarisine Şekil 3.6’da yer verilmiştir.



Şekil 0.6. Kısıtlı Boltzmann Makinesi

3.2.3. Tekrarlı YSA

Temel olarak tekrarlı bir mimariye sahip oldukları için TSA olarak adlandırılmaktadır. Ağda işlemler gerçekleştirilirken bilgiler geriye doğru ve aynı seviyede bulunan nöronlar arasında bilgiler işlenebilmektedir. Geriye doğru işlem kabiliyeti sayesinde TSA’nın giriş bilgileri işlenirken dahili hafıza kullanılabilmektedir.

Evrişim işlemi, ağı zamanla değişkenlerini paylaşmasına olanak tanımakla beraber sıgıdır. Evrişimin çıktısı, çıktının her bir biriminin girişin az sayıda biriminin bir işlevi olan bir dizidir. Değişken paylaşımı zaman adımlarında aynı evrişim çekirdeğinin uygulanmasıyla meydana gelmektedir. TSA’lar değişkenleri farklı bir biçimde paylaşmaktadır. Çıkışın her bir birimi çıkışın daha önceki birimlerine ait bir işlevdir. Daha önceki çıkışlara uygulanmış olan güncelleme kuralı uygulanarak çıkışın her bir birim üretilmektedir. Yinelenen bu yöntem ile değişkenlerin çok derin bir hesaplama grafiği ile paylaşılmasına olanak tanınmaktadır. Her bir durum bahsi geçen bu tekrarlama bağlantısı üzerinden önceden hesaplanan bütün hesaplamalar ile ilişkilidir. Bu durumu önemli kılan sebep, TSA’nın zaman ile bir hafızaya sahip olmasıdır. Çünkü mevcut haller daha önceki evrelerde bulunan bilgileri taşımaktadır. Teoride TSA, uzun bir zaman dilimi için veriyi hatırlayabilmekteyken uygulamada yalnızca birkaç basamak önceki bilgilere bakmak ile sınırlı olmaktadır.

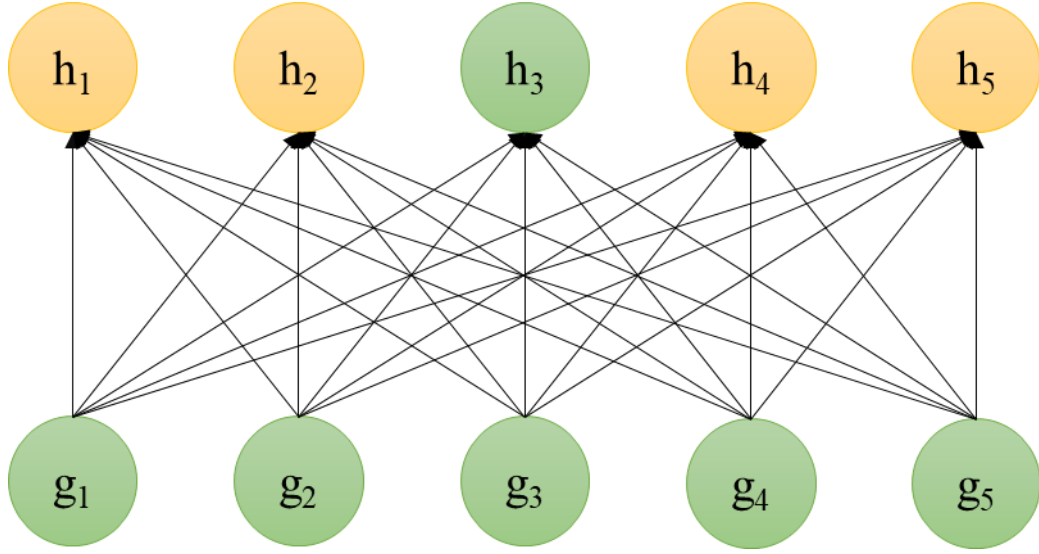
3.3. Evrişimsel Sinir Ağları (ESA)

ESA görsel görüntüleri analiz etmek için en yaygın olarak kullanılan derin öğrenme tabanlı yapay sinir ağı sınıfıdır. ESA bilgisayarlı görme alanında ilk defa 1998 yılında kullanılmıştır. ESA esas popülerliğini Krizhevski vd. [93]'nin ImageNet nesne tanıma yarışmasından sonra kazanmıştır. ESA'nın popülerliğinin artması üç temel sebeple özetlenebilir: İlk olarak, gün geçtikçe gelişen üretim teknolojileriyle bilgisayar ekran kartlarının hesaplama güçlerinin katlanarak artması, ikincisi büyük verilerin toplanması ve bu verilere kolaylıkla ulaşılması, son olarak geliştirilen işlevsel ESA modellerinin ortaya çıkmasıdır. Bir ESA modeli temelde iki bölümden oluşmaktadır. Bunlar evrişimsel ve sınıflandırma katmanlarıdır. Evrişimsel katmanın kullanım amacı görüntülerin özniteliklerini çıkarmak iken sınıflandırma katmanı çıkarılan öznitelikleri kullanarak sınıflandırma problemine uygun olarak sınıflandırma işlemini gerçekleştirir.

Geleneksel bir makine öğrenme yönteminin geliştirilmesiyle ESA ortaya çıkmıştır. ESA'nın klasik makine öğrenme yöntemlerine göre birçok avantajı bulunmaktadır. Parametre paylaşımı ve seyrek etkileşimler ESA'nın en önemli avantajlarından. Ayrıca ESA, farklı büyüklükteki girişler için kolaylık sağlamaktadır.

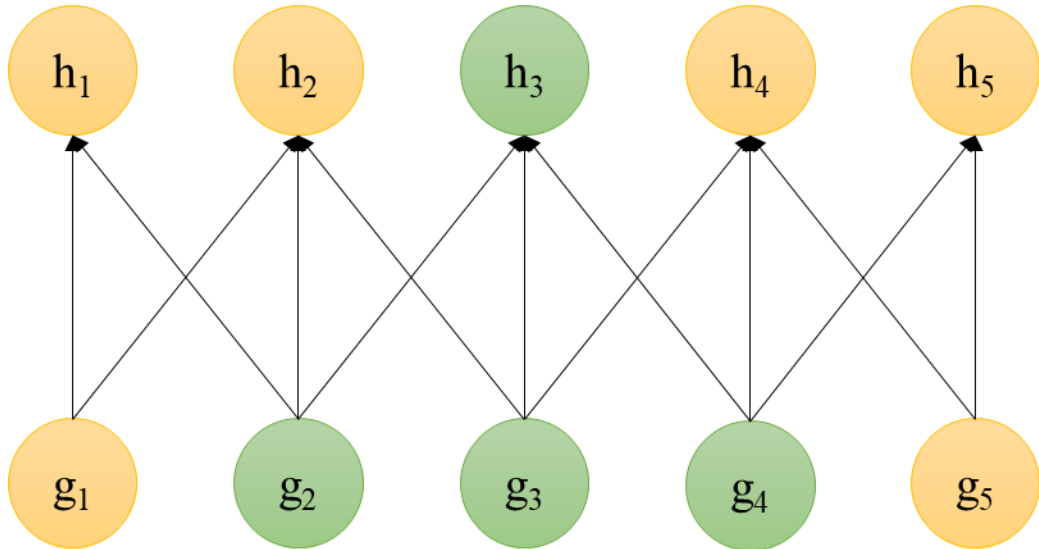
Klasik sinir ağı katmanlarında, her girdi elemanının her girdi elemanı ile etkileşime girdiği anlamına gelen ve her bir girdi elemanı ve her bir çıktı elemanı arasındaki etkileşimi açıklayan ayrı bir parametre içeren bir ağırlık matrisi ile matris çarpımı yapılmaktadır. Bunun yanında ESA mimarisinde geleneksel YSA mimarilerine göre daha seyrek etkileşimler vardır. Bu durum, seyrek bağlantı ya da seyrek ağırlıklar olarak tanımlanmaktadır. Bu işlev, çekirdeğin girdiden daha küçük boyuta getirilmesiyle sağlanır. Örneğin; binlerce ya da milyonlarca pikselden oluşan bir giriş görüntüsünü işlerken milyonlarca piksel içerisinden sadece onlarca ya da yüzlerce pikseli barındıran çekirdeklerden küçük ve anlamlı öznitelikler tespit edilmektedir. Bu da hem oluşturulan modelin bellek gereksinimlerini düşürmekte hem de istatistiksel verimliliği arttırmaktadır. Böylece depolanması gereken parametre azalmaktadır. Aynı zamanda çıkışa kadar yapılan hesaplarda daha az işleme ihtiyaç duyulmaktadır. ESA'da genellikle verimliliği arttıran bu gelişmeler büyük oranda bulunmaktadır. Örneğin; m tane girdisi ve n adet çıktısı olan bir YSA yapısında matris çarpımı için $m \times n$ adet parametreye ihtiyaç duyulmaktadır. Bu çarpım işleminde kullanılan algoritmalar belli bir işlem zamanı gerektirmektedir. Diğer taraftan, ESA'da ise k ($k < m$) gibi bir girdi ile bağlanması gereken bağlantı sayısı sınırlandırılırsa, o zaman seyrek bağlı yaklaşımda yalnız $k \times n$ adet parametreye ihtiyaç duyulmaktadır. Bu sayede, YSA yapısından daha az bir işlem zamanına ihtiyaç duyulmaktadır [67,68].

Girdileri g_1, g_2, g_3, g_4 ve g_5 olan ve gizli katmanları h_1, h_2, h_3, h_4 ve h_5 olan 5 girişli ve 5 gizli katmanlı bir sinir ağının giriş ve çıkış katmanlarının etkileşimleri Şekil 3.7'de verilmiştir.



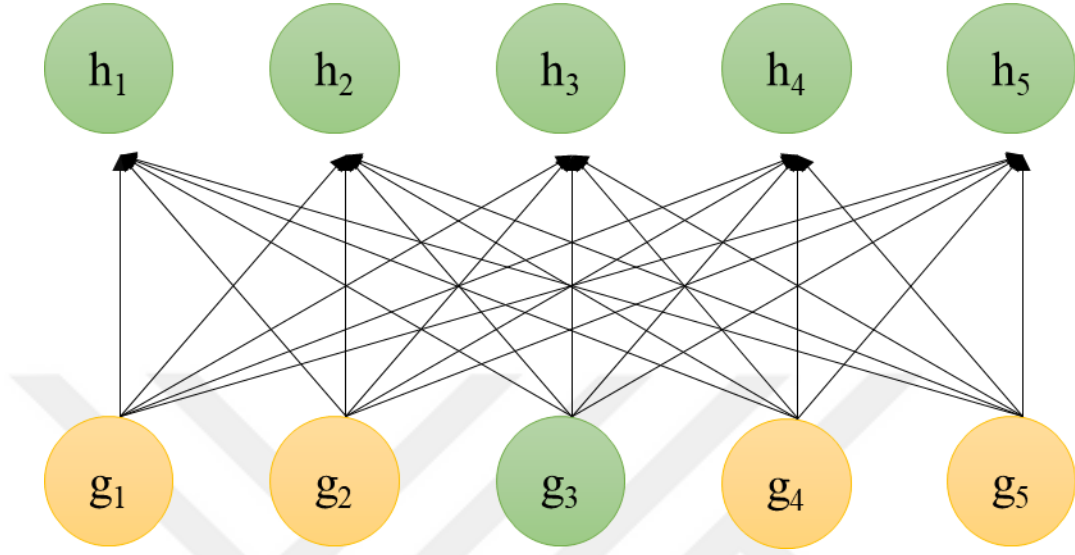
Şekil 0.7. Klasik bir sinir ağında çıkış katmanının giriş birimleri ile etkileşimi

Şekil 3.7’de de görülebileceği gibi sinir ağında h_3 çıktı biriminin tüm girdi birimleri arasında etkileşim vardır. Çıktı birimleri ile etkileşimi bulunan girdi birimleri alıcı alan olarak adlandırılmaktadır. 5 girdi ve 5 çıktı birimi olan bir evrişim ağının bir çıkış biriminin hangi çıktı birimleri ile etkileşimde olduğu Şekil 3.8’de gösterilmektedir. Şekilde yer alan evrişim ağında filtre boyutu 3 seçildiğinden h_3 çıkış biriminin g_1 , g_2 ve g_3 girdi birimleri ile etkileşimi bulunmaktadır.



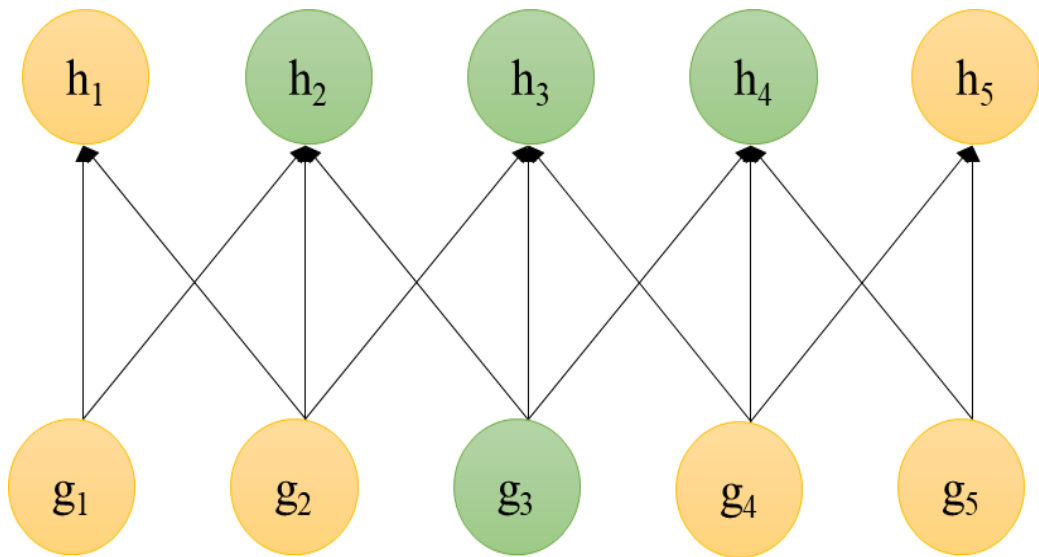
Şekil 0.8. Temel bir evrişim ağında çıkış katmanın giriş birimi ile etkileşimi

Şekil 3.9’da 5 girdi ve 5 çıktı katmanı olan bir sinir ağında bir girdinin hangi çıktılarla etkileşimde olduğu temsil edilmektedir. Şekilden anlaşılabacağı üzere g_3 girdisinin tüm çıktılar ile etkileşimi bulunmaktadır.



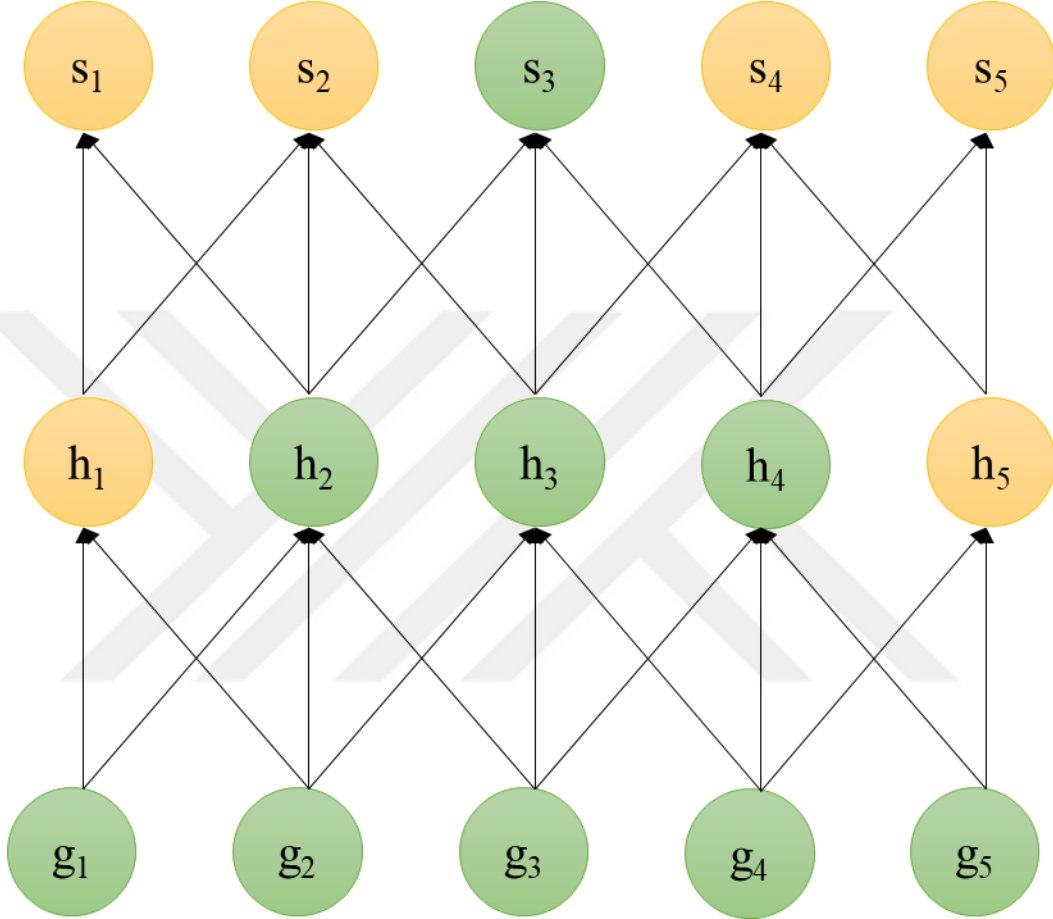
Şekil 0.9. Klasik bir sinir ağında giriş biriminin çıkış katmanları ile etkileşimi

Şekil 3.10’da ise temel bir evrişim ağında bir girdinin hangi çıktılar ile etkileşimde olduğu temsil edilmektedir. Şekilden anlaşılabacağı üzere filtre boyutu üç olduğundan g_3 girdi birimi ile h_2 , h_3 ve h_4 çıktı birimleri arasında etkileşim bulunmaktadır.



Şekil 0.10. Temel bir evrişim ağında giriş biriminin çıkış katmanı ile etkileşimi

Şekil 3.8’de yer alan evrişimsel ağda h_3 çıktı birimi yalnızca filtre boyutu kadar girdi birimi ile etkileşimde bulunsa da daha derin bir evrişimsel ağda daha derinde bulunan çıktı birimleri giriş birimlerinin hepsi veya tamamına yakını ile etkileşimde bulunabilir. Şekil 3.11’de bahsi geçen durum etkileşim mimarisine yeni bir katman ilave edilerek gösterilmiştir.

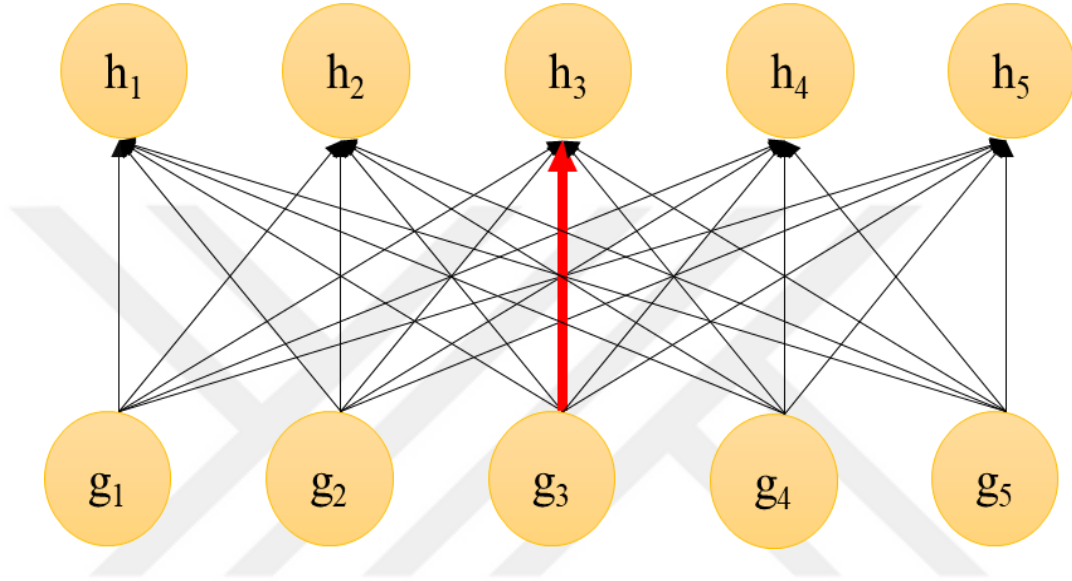


Şekil 0.11. İki katmanlı temel bir evrişim ağında çıkış katmanının giriş birimleri ile etkileşimi

Şekil 3.11’de görüleceği gibi s_3 çıktı birimi ile h_2 , h_3 ve h_4 çıktı birimleri arasında etkileşim vardır. Bu çıktı birimlerinin de tüm girdi birimleri arasında etkileşim bulunmaktadır.

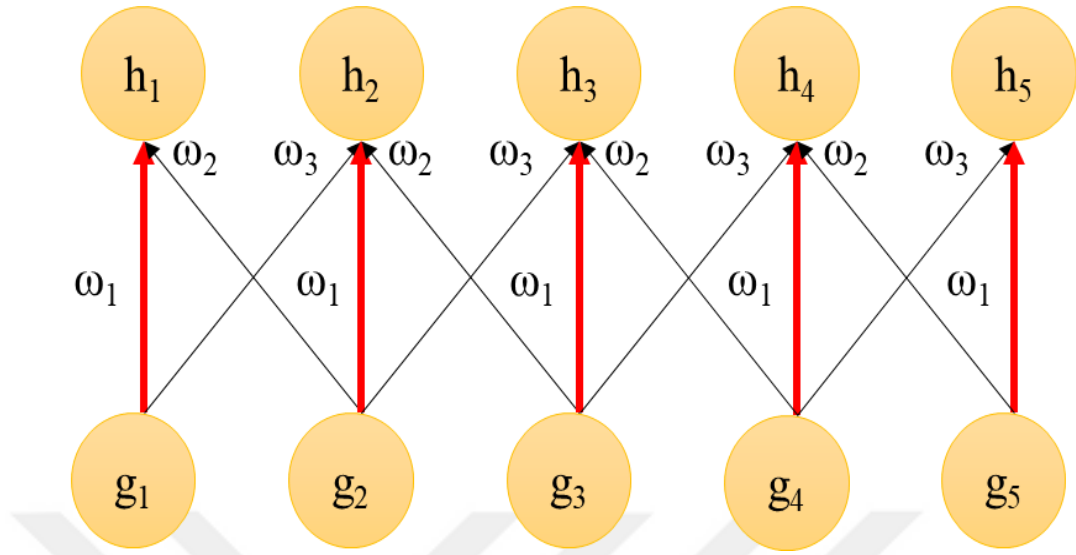
Aynı parametrenin bir modelde birden fazla işlem için kullanılmasına parametre paylaşımı denilmektedir. Klasik bir sinir ağında bir katmanın çıktısını hesaplamak için ağırlık matrisinin her elemanı yalnızca bir defa kullanılmaktadır. ESA’da ise çekirdeğin her birimi giriş biriminin her bir konumunda kullanılır. Parametre paylaşımının evrişim işleminde kullanılmasından dolayı her bir yerde farklı bir parametre setinin öğrenilmesi yerine sadece bir ağırlık matris setinin öğrenilmesi yeterli olmaktadır. Bahsi geçen bu durum, ileri yayılım çalışma süresini etkilemez. Bunun yanında, evrişim işlemleri; geri yayılımdaki hesaplama, bellek ihtiyaçları ve istatistiksel verimlilik

bakımından klasik bir sinir ağı yapısına göre daha fazla verimlilik sağlamaktadır. Şekil 3.12’de yer alan basit bir sinir ağı için parametre paylaşımına bakıldığında koyu renk ile ifade edilen ok yalnızca g_3 (ağırlık) girdi birimi ile h_3 çıktı birimi arasında kullanılmaktadır. Benzer şekilde, diğer ağırlıklar içinde aynı durum geçerlidir. Bu sebeple, parametre paylaşımı bulunmamaktadır. Şekil 3.12’de gösterilen 5 girdi birimi ve 5 çıktı birimi olan sinir ağında $5 \times 5 = 25$ adet farklı ağırlık bulunmaktadır [94].



Şekil 0.12. Temel bir sinir ağının parametre paylaşımı

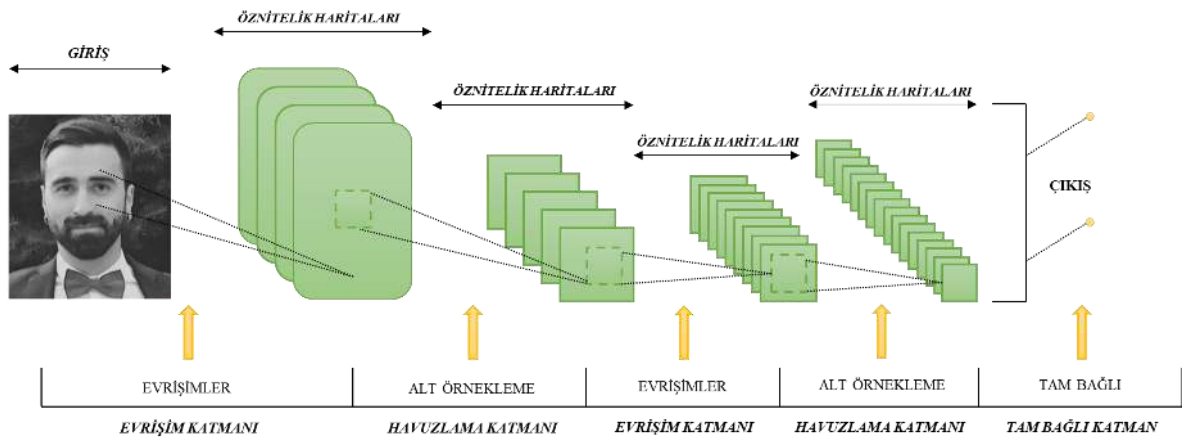
Şekil 3.13’te yer alan evrişimsel ağda parametre paylaşımı sayesinde çok daha az sayıda ağırlık kullanılmıştır. Aynı indislere sahip bir girdiden gelen ağırlıkta ω_1 , kendi indisinden ardışık olarak sonra gelen girdi biriminden gelen ağırlıkta ω_2 kendisinden bir önceki indisteki girdi biriminden gelen ağırlıkta ω_3 olmak üzere 3 adet parametre kullanılmıştır [95]. Kayda değer bir şekilde sayıları azalan ağırlık parametreleri bilhassa geri yayılım algoritmalarının hesaplama maliyetini oldukça önemli bir oranda düşürmektedir.



Şekil 0.13. Bir Evrişimsel Sinir Ağında parametre paylaşımı

3.3.1. ESA Mimarisi

Temel bir ESA mimarisi evrişimsel katmanlar ve sınıflandırıcı katmanlar olmak üzere iki bölümden oluşmaktadır. Evrişimsel katmanlar, görüntüden öz nitelikler çıkartmak amacıyla bir dizi evrişim ve birleştirme işlemleri için kullanılmaktadır. Sınıflandırıcı katmanlar ise sıralı bir tam bağlı katmandan oluşmaktadır. Evrişimsel katmanlar ile çıkarılan öz nitelikler sınıflandırıcı katmanlar tarafından kullanılarak çözülmek istenen probleme uygun kategori değerleri üretmektedir. Şekil 3.14'te temel bir ESA'ya ait ağ mimarisi yer almaktadır.



Şekil 0.14. Temel bir ESA mimarisi örneği

Evrişim Katmanı

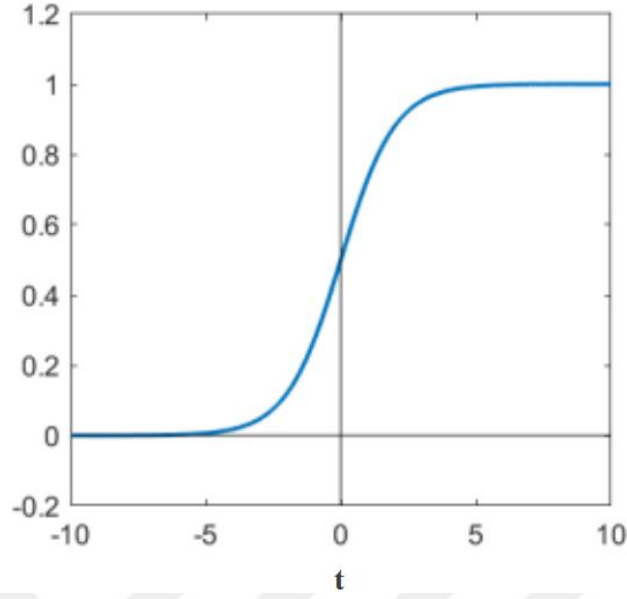
ESA'nın en temel katmanı olarak kabul edilen evrişim katmanının amacı, evrişim işlemini uygulamak ve görüntülerin öznitelikleri çıkarmaktır. Evrişim işleminde, girdi görüntü matrisi üzerinde boyutları değişebilen filtre veya kernel olarak adlandırılan bir matris gezdirilmektedir. Giriş görüntüsü gri seviyede (gray scale) ise, matrisin derinliği 1; şayet RGB renkli görüntü ise, matrisin derinliği 3'tür. Giriş görüntü matrisi üzerinde gezdirilen filtre matrisi ile çarpılıp sonucu toplanır ve filtre matrisi görüntünün tamamında kaydırılarak evrişim işlemi yinelenir.

Evrişim işlemi sonucunda elde edilen yeni görüntü özellik haritası (feature map) veya aktivasyon haritası olarak adlandırılmaktadır. Filtre sayısı değiştikçe özellik haritalarının sayısı da doğru orantılı olacak şekilde değişmektedir. Kullanılan her bir filtre özellik tanımlamak için kullanıldığından bu filtrelere özellik tanımlayıcı denilmektedir. Evrişim işlemi gerçekleştirmek için piksel ekleme (padding) ve kaydırma adımı parametreleri de kullanılmaktadır.

Aktivasyon Katmanı

Evrişim katmanında çıkartılan özniteliklerden sonra genellikle aktivasyon fonksiyonları kullanılmaktadır. Doğrusal aktivasyon fonksiyonları, bilinen en basit aktivasyon fonksiyonları olup bu fonksiyonlar kullanıldığında kullanılan ağda hiçbir değişim uygulamamaktadır. Bu tür aktivasyon fonksiyonları kullanılan bir ağın eğitimi oldukça basit olmasına rağmen veride bulunan karmaşık özellikleri bulamamaktadır. Ek olarak *sigmoid*, *tanh* gibi doğrusal olmayan aktivasyon fonksiyonları da kullanılmakta olup veride yer alan karmaşık özellikleri daha iyi belirleyebilmektedir.

Ağırlık ve eşik değerlerdeki küçük değişimler bile doğru sınıflandırmayı etkileyen faktörler arasında yer almaktadır. Değişim miktarı çok küçük olsa da algılayıcının doğru bir karar verme durumunu etkileyebilmektedir. Bunun sonucunda bütün ağın durumunda değişiklikler meydana gelebilmektedir. Bu sorunu çözmek için *sigmoid* fonksiyonu kullanılmaktadır. Şekil 3.15'te *sigmoid* fonksiyonunun grafiği gösterilmektedir.



Şekil 0.15. Sigmoid fonksiyonunun grafiği

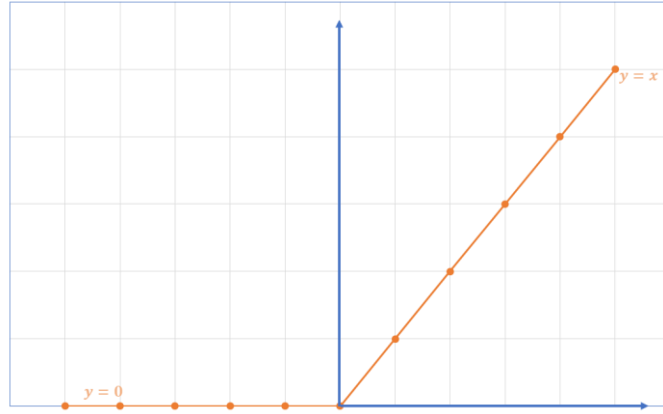
Sigmoid fonksiyonu algılayıcılar ile benzerlikler göstermesine rağmen bu fonksiyonların ağırlık ve eşik değerlerinde yapılan olası küçük değişiklikler çıkışlarına da daha az etkilemektedir [96]. *Sigmoid* fonksiyonlarının da algılayıcılara benzer şekilde $x_1, x_2 \dots$ gibi girişleri bulunmaktadır. Ancak bu girişlerin değeri 0 ya da 1 olmasının yanında 0 ile 1 arasında yer alan değerlerden oluşmaktadır. *Sigmoid* fonksiyonları da algılayıcı da yer alan nöronlar gibi bir ağırlık ve eşik değerine sahiptir. *Sigmoid* fonksiyonunun formülü Denklem 3.1 ile ifade edilmektedir.

$$sigmoid = \frac{1}{1+e^{-t}} \quad (3.1)$$

Yukarıda bahsedilen *sigmoid*, *tanh* gibi doğrusal olmayan aktivasyon fonksiyonların ağırlık katmanlarına doğru türevleri sıfır olduğundan ağırlıklarda güncellenememe gibi dezavantajlar bulunmaktadır. Bu nedenle ESA’larda genellikle kullanımı tercih edilen aktivasyon fonksiyonu Düzeltilmiş Doğrusal Birim (DDB) fonksiyonu ya da bilinen adı ile ReLU (Rectified Linear Unit)’dur. Denklem 3.2 ve Şekil 3.16’da gösterildiği gibi aktivasyon fonksiyonu pozitif bir değer aldığı anda alınan değeri geri döndürürken negatif bir değer aldığı anda ise sonuç olarak sıfır değerini üretmektedir.

$$f(x) = \max(0, x) \quad (3.2)$$

Burada; x DDB’nin girişi iken $f(x)$ düzeltilmiş çıkışıdır.



Şekil 0.16. DDB katmanında kullanılan aktivasyon fonksiyonu

Havuzlama Katmanı

Havuzlama işlemi genellikle aktivasyon katmanından sonra gelmektedir. Havuzlama işleminin temel kullanım amacı kendisinden sonra gelen evrişim katmanı için giriş matrisinin sütun ve satır boyutunun azaltılmasıdır. Havuzlama işlemi yapıldıktan sonra çıkışta matrisin derinlik boyutu değişmemektedir. Havuzlama katmanında gerçekleştirilen bu işlem aşağı örnekleme olarak adlandırılmaktadır. Havuzlama işlemi sonucunda kullanılan veride bilgi kaybı meydana gelmektedir. Yaşanan bu bilgi kaybının iki yönden avantajı bulunmaktadır. İlk olarak, kendisinden sonra gelen katmanlar için daha az hesaplama işlemi yapılmasını sağlamaktadır. İkinci avantajı ise ağıın ezberlemesi önlenmektedir. Havuzlama işleminde, evrişim katmanına benzer şekilde belli filtreler giriş matrisine uygulanmaktadır. Filtreler giriş matrisinin üzerinde seçilen belli bir adım kaydırma değerine istinaden gezdirilmektedir [97]. Giriş matrisinin filtre sınırları içindeki maksimum, minimum ve ortalama değerlerinden bir tanesi kullanılarak havuzlama işlemi yapılır [97]. Yapıldığı işlemlere göre bu katmanda maksimum, minimum ve ortalama havuzlama katmanı gibi farklı havuzlama yöntemleri bulunmaktadır. Maksimum havuzlama yöntemi ile üzerinde filtre gezdirilen her alt bölgeden maksimum değer alınırken ortalama havuzlama yönteminde ise her alt bölgenin ortalama değerini alınmaktadır. Benzer şekilde minimum havuzlama yönteminde üzerinde filtre gezdirilen her bölgenin minimum değeri alınmaktadır.

Havuzlama işlemi kendisinden önceki evrişim katmanında kullanılmış olan filtre sayısı kadar matriste gerçekleştirilmektedir. Bunun yanında ESA mimarilerinde DDB katmanından sonra havuzlama katmanının kullanılması zorunlu değildir. Havuzlama katmanının boyutu girdi ve kullanılan filtre boyutuna göre çıktı boyutu aşağıdaki denklemlere göre hesaplanmaktadır.

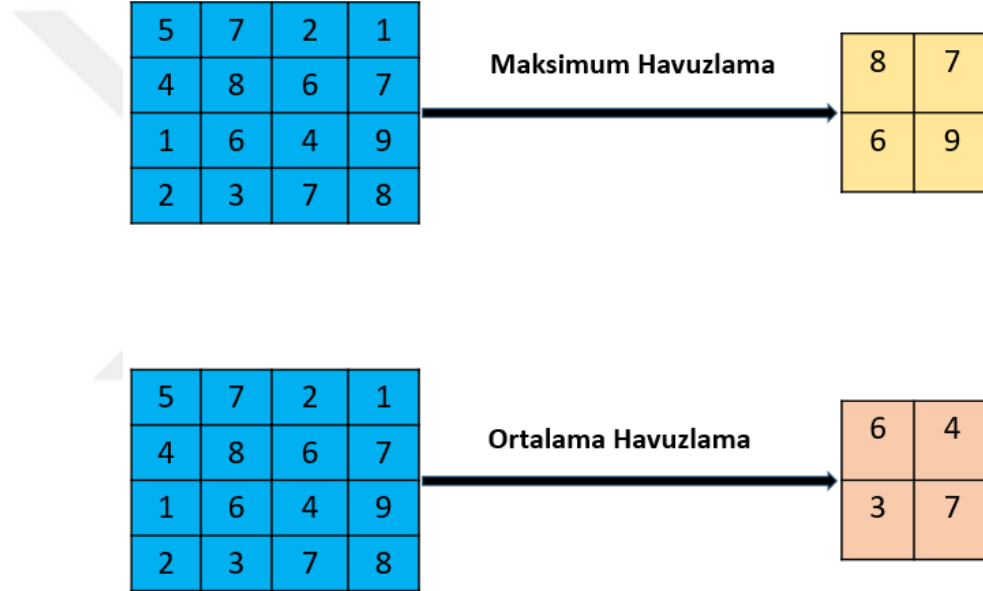
$$\text{Çıkış boyutu} = X_2 \times Y_2 \times Z_2 \quad (3.3)$$

$$X_2 = \frac{(X_1 - F)}{S} + 1 \quad (3.4)$$

$$Y_2 = \frac{(Y_1 - F)}{S} + 1 \quad (3.5)$$

$$Z_2 = Z_1 \quad (3.6)$$

Burada giriş matrisinin genişlik, yükseklik ve derinlik boyutu sırasıyla X_1 , Y_1 ve Z_1 olup havuzlama işlemi gerçekleştirildikten sonra çıkış matrisinin yükseklik ve derinlik boyutu sırasıyla X_2 , Y_2 ve Z_2 'dir. S kaydırılacak adımı gösterirken F ise filtre boyutunu temsil etmektedir. Şekil 3.17'de 4x4 boyutundaki girişe 2x2 boyutundaki 2 adım kayma ile maksimum havuzlama ve ortalama havuzlama işlemi uygulanmıştır. Gerçekleştirilen bu işlem sonucunda çıkış matrisinin boyutu 2x2 olmuştur.



Şekil 0.17. Maksimum ve ortalama havuzlama örnek gösterimi

Toptan Normalleştirme Katmanı

Çok derin ağ mimarileri, çeşitli katman veya fonksiyonların birleşiminden oluşmaktadır. Diğer katmanların değişmediği varsayılarak kullanılan her bir parametrenin ne şekilde güncelleneceği gradyan inme algoritması ile yapılmaktadır. Bütün katmanlar uygulamada eşzamanlı bir biçimde güncellenmektedir. Ancak diğer fonksiyonların (diğer katmanlar ve girdileri) aynı kaldığı varsayılarak elde edilen güncellemeler kullanılır ve oluşturulan birçok fonksiyon eşzamanlı olarak değiştirilir. Bunun sonucunda ESA gibi çok katmanlı mimarilerde ağırlıkların olması gereken düzeyden kaymasına yol açar. Bu duruma iç değişken değişme problemi denilmektedir. Bahsi geçen bu sorunu çözmek için Ioffe ve Szegedy [98] toptan normalleştirme algoritmasını sunmuşlardır. Toptan normalleştirme katmanı, ağın ilk çalışmaya başladığındaki

ağırlık değerlerine karşı hassasiyetiyle birlikte ağırlık eğitimi süresini de azaltmaktadır. Toptan normalleştirme algoritması aşağıda yer alan denklemlerle gerçekleştirilmektedir. Denklem 3.7 ile normalleştirilmiş etkileşim $\dot{x}_i$, girdi x_i , mini toptan ortalama m_b ve mini toptan değişim v_b değerleri ile hesaplanmaktadır.

$$\dot{x}_i = \frac{x_i - m_b}{\sqrt{v_b^2 + \epsilon}} \quad (3.7)$$

Burada ϵ sabit değerli bir sayıdır ve v_b 'nin küçük olduğu durumlarda sayısal kararlılığın korunmasına yardımcı olur.

$$m_b = \frac{1}{n} \sum_{i=1}^n x_i \quad (3.8)$$

$$v_b = \frac{1}{n} \sum_{i=1}^n (x_i - m_b)^2 \quad (3.9)$$

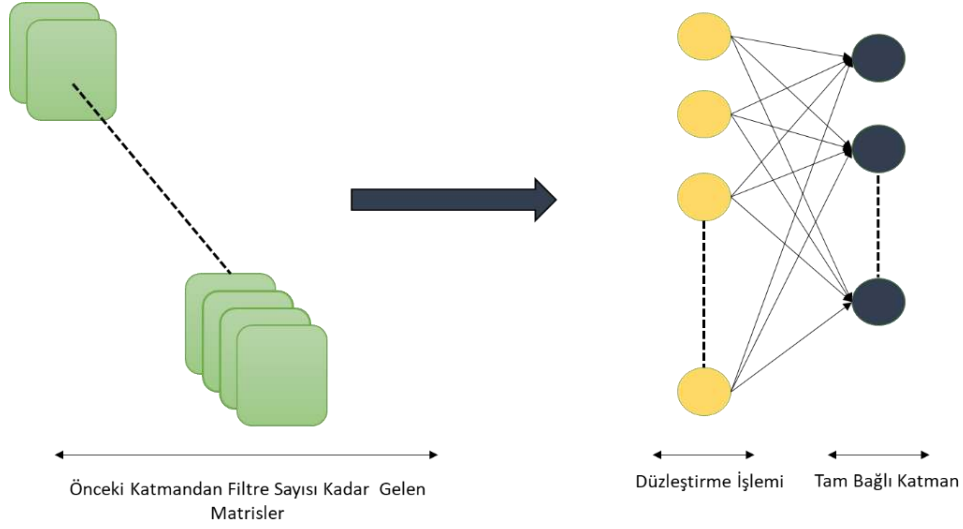
Son olarak toptan normalleştirme katmanındaki etkileşim y_i , kayma ve ölçek değerleri hesaplanmaktadır.

$$y_i = a\dot{x}_i + b \quad (3.10)$$

Burada kayma ve ölçek faktörleri sırasıyla a ve b olarak gösterilmektedir. Aynı zamanda bu faktörler eğitim boyunca kullanılan öğrenilebilir değerlerdir.

Tam Bağlı Katman

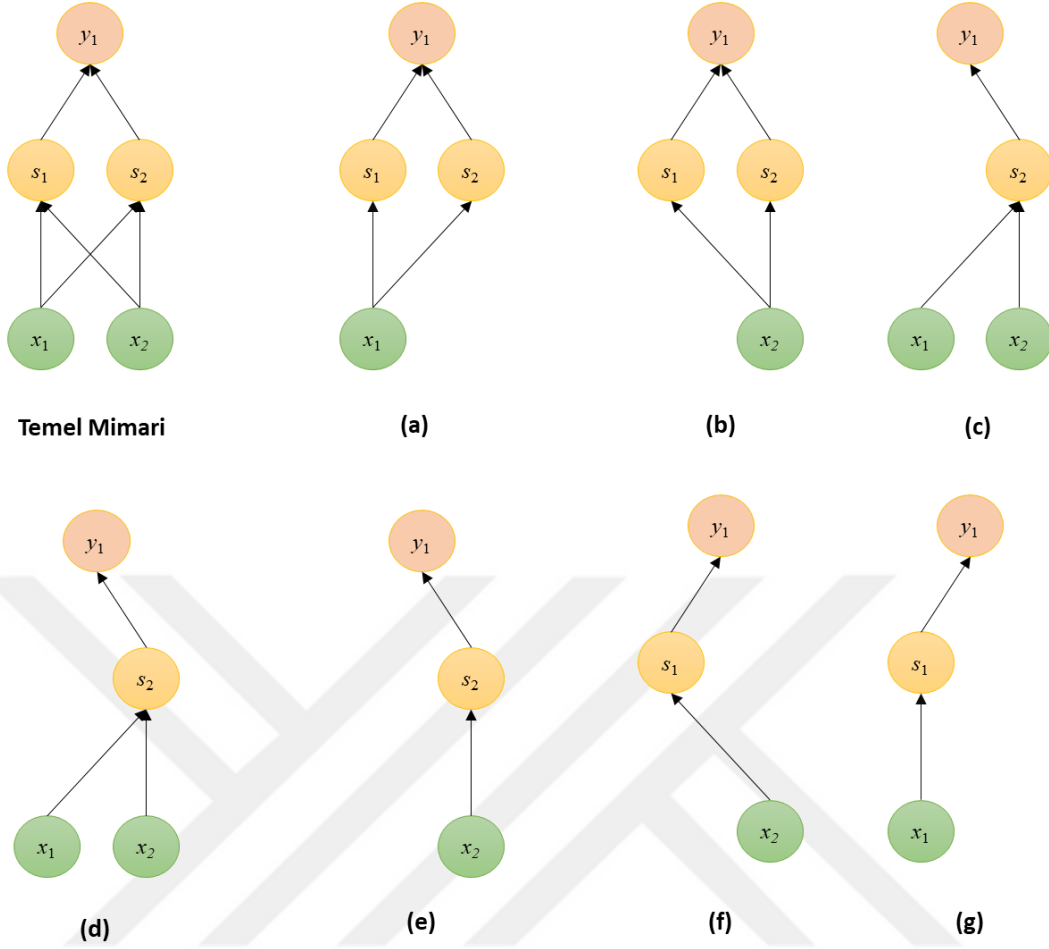
Tam bağlı katman ESA'ların genellikle son katmanıdır. Evrişim, aktivasyon ve havuzlama katmanları ile öznitelikler elde edildikten sonra bu bilgiler tam bağlı katmana aktarılır. Tam bağlı katmandan önceki katmandan gelen matris verilerinin yükseklik, genişlik ile derinlik değerleri vektör yapısında olacak şekilde düzeltilmektedir. Tam bağlı katmanda bulunan nöronlar önceki katmanda bulunan tüm nöronlara tamamen bağlıdır. Elde edilen öznitelikleri tek boyut halinde alıp sınıflandırma katmanına iletmesi tam bağlı katmanın en önemli özelliğidir. Ağ yapısının en sonunda kullanılan tam bağlı katmanda, her girdi için toplam sınıf sayısı adedince sınıf olasılık değerleri içermektedir. Örnek olarak; 5 sınıflı 1000 eğitim verisinden oluşan bir makine öğrenmesi problemi için tam bağlı katmanın çıkışında her bir veri için 5 adet olmak üzere 1000x5 boyutunda öznitelikleri temsil eden matris elde edilir. Şekil 3.18'de tam bağlı katmanın temsili gösterimi yer almaktadır.



Şekil 0.18. Tam bağlı katmanın temsili gösterimi

Seyreltme Katmanı

Seyreltme işlemi, çok katmanlı YSA mimarilerini daha düzenli bir hale getirmek için basit ve kuvvetli bir yöntem sağlamaktadır [99]. Seyreltme işlemi sinir ağı topluluklarında aşırı öğrenmenin engellenmesi ve torbalamanın basit bir hale getirilmesi için önerilmiştir. Torbalama ile birden fazla model eğitilmek ile beraber her test verisinde birden fazla model değerlendirilmektedir. Her modelin büyük bir sinir ağı olduğu hallerde ağı eğitilmesi, test edilmesi, çalışma süresi ve bellek kullanımı yönünden büyük bir maliyet ortaya çıkmaktadır. Eğitim verileri, ESA mimarilerinde de toptan eğitilmemektedir. Eğitimi gerçekleştirmek için veriler öncelikle belli bir boyuta düşürülmüş toptan modellere ayrılmaktadır. Bu işlemden sonra eğitim gerçekleştirilerek her grubun çıktısından elde edilen değerlerin ortalaması alınıp bir çıktı belirlenmektedir. Beş ile on sinir ağından oluşan toplulukların kullanımı genellikle yaygındır. Szegedy vd. [51] 2014 yılında gerçekleştirilen ILSVRC görüntü sınıflandırma yarışmasında 6 adet sinir ağı mimarisini kullanmışlardır. Fakat önerilen mimaride seyreltme katmanını kullanmadıkları için ağ eğitimi hantallaşmıştır. Bu olaydan sonra ESA mimarilerinde seyreltme katmanı yaygın bir şekilde kullanılmaya başlanmıştır. Her bir sinir ağı mimarisinde bazı düğüm ve bağlantılar rastgele bir şekilde seyreltme katmanında kullanılmaz. Şekil 3.19’da geleneksel iki katmanlı YSA modelinde rastgele seyreltme işlemine ait temsili gösterimler yer almaktadır.



Şekil 0.19. İki katmalı bir ağ için seyreltme katmanındaki seyreltme senaryoları

Eşiksiz En Büyük İşlev Katmanı

ESA tam bağlı katman çıkışında bulunan her girdi için veride bulunan sınıf sayısı adedince vektör değerleri mevcuttur. ESA’da bu vektör değerlerini uygun olan sınıf değerine aktarmak için ‘eşiksiz en büyük işlev’ kullanılmaktadır [100]. Eşiksiz en büyük işlev kullanımı, her bir veri için olasılık değeri ürettiğinden dolayı daha çok çok sınıflı sınıflandırma görevlerinde tercih edilmektedir. Eşiksiz en büyük işlev aşağıda yer alan denklem ile hesaplanmaktadır.

$$S_i = \frac{e^{y_i}}{\sum_j e^{y_j}} \quad (3.11)$$

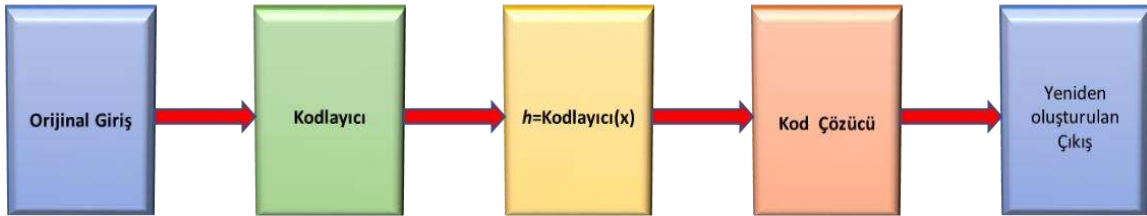
Burada; S_i , her bir değer için eşiksiz en büyük işlev çıkışı olmakla beraber her bir j değeri için sınıflara ait olma olasılığı hesaplanmaktadır. Tam bağlı katmandan alınan değerleri y_i temsil etmektedir. Eşiksiz en büyük işlev çıkışından elde edilen tahmini sınıf verileri ile gerçek sınıf verileri arasında bulunan hata, çapraz düzensizlik maliyet işlevi ile hesaplanmaktadır. Çapraz düzensizlik maliyet işlevi aşağıda yer alan Denklem 3.12 ile hesaplanmaktadır.

$$D(S, L) = - \sum_i L_i \log(S_{(i)}) \quad (3.12)$$

Burada; L_i gerçek sınıf etiketlerini temsil etmektedir. Çapraz düzensizlik maliyet işlevinden elde edilen değer düşük ise doğru orantılı olarak tahmini hata oranı değeri de düşük olmaktadır. Elde edilen hata oranı istenilen değere ulaşınca kadar farklı optimizasyon teknikleri kullanılarak ağ mimarisindeki tüm öğrenilebilir parametre değerleri güncellenir.

3.4. Otomatik Kodlayıcılar

Otomatik kodlayıcılar (autoencoder), bir ağa giriş olarak verilen veriyi o ağın çıkışına kopyalama görevi görmektedir [101]. Genellikle kodlayıcı (encoder) ve kod çözücü (decoder) olmak üzere iki bölümden oluşmaktadır. Dahası, yapısında giriş verisini temsil eden bir gizli katman da bulunmaktadır. Otomatik kodlayıcının temel yapısı Şekil 3.20’de yer almaktadır. Kodlayıcı, giriş verisini alıp bu veriden çıkış verisi üretmektedir. Daha sonra, bu çıkış verisi gizli katmana iletilir. Gizli katman da çıkış verisini kod çözücüye giriş verisi olarak aktarır. Bu işlemin sonucunda kod çözücü tekrardan üretilmiş bir giriş verisi üretir. Giriş verisi kopyasının oluşturulma sürecinde gizli katmanda oluşan değerler ile öznitelik çıkarımı, boyut küçültme gibi problemlerin çözülmesinde de kullanılabilir [102]. Giriş verisinin h gibi bazı öznitelikleri kodlayıcı birimi kullanılarak elde edilmektedir. Elde edilen h öznitelikleri kod çözücüye aktarılıp giriş olarak verilen veri yeniden üretilmektedir. Seyrek, gürültü giderici ve varyasyonel otomatik kodlayıcı en yaygın kullanılan otomatik kodlayıcı türleridir.

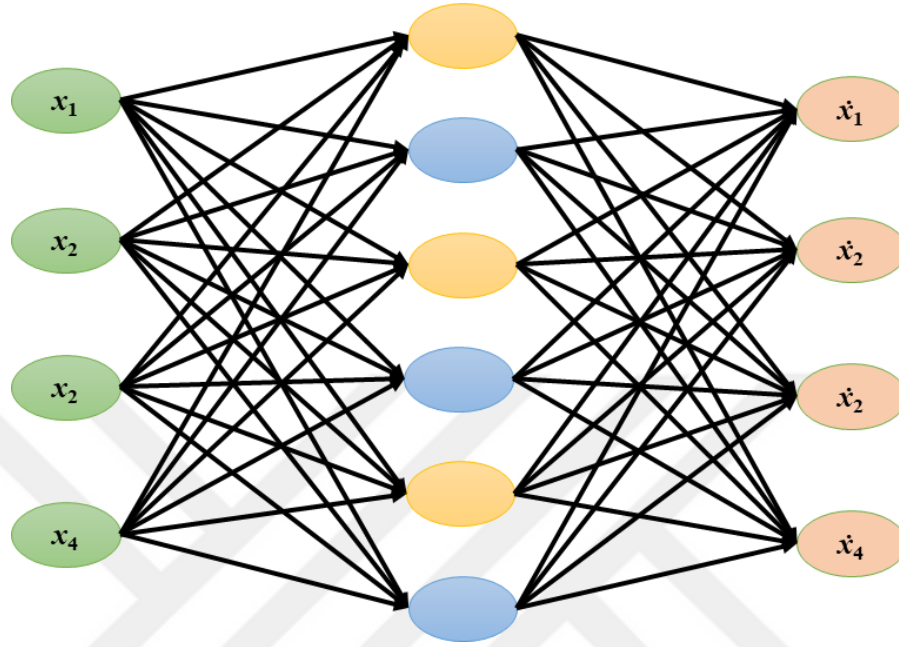


Şekil 0.20. Otomatik kodlayıcının temel yapısı

3.4.1. Seyrek Otomatik Kodlayıcı

Seyrek otomatik kodlayıcılar, ağ giriş boyutunun gizli katman boyutundan daha küçük olabildiği otomatik kodlayıcı türüdürler. Ancak bu otomatik kodlayıcı türü daha az gizli katmanın bir giriş verisi için etkin olmasını sağlamaktadır. Böylece ağ modeli giriş verisinin bazı özelliklerine yoğunlaşabilmektedir. Bu sayede girişin benzersiz özniteliklerini çıkartmaya olanak tanımaktadır [101,103,104]. Örnek bir seyrek otomatik kodlayıcı mimarisi Şekil 3.21’de yer almaktadır.

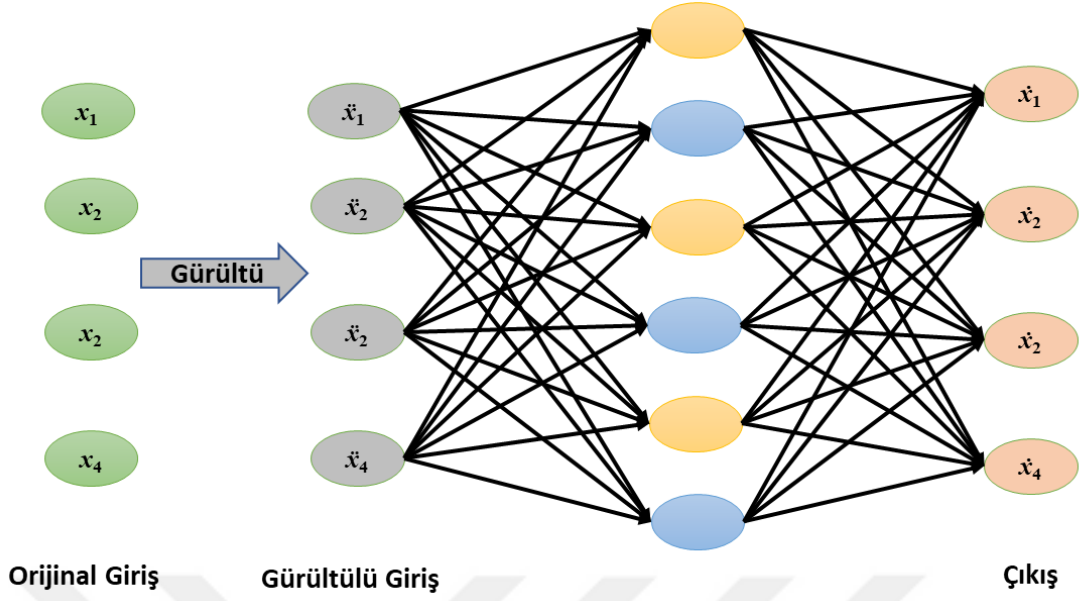
Şekilden de anlaşılacağı üzere giriş verisinde gizli alandan daha az sayıda birim bulunmaktadır. Birimlerin sadece bir kısmı aktif hale geldiğinden dolayı giriş verisinin benzersiz öznitelikleri öğrenilmektedir.



Şekil 0.21. Örnek bir seyrek otomatik kodlayıcı modeli

3.4.2. Gürültü Giderici Otomatik Kodlayıcı

Gürültü giderici otomatik kodlayıcılar, gürültülü girdi verilerin çıktısını orijinal ve gürültüsüz olarak üretmek için eğitilen otomatik kodlayıcılardır [105–107]. Gürültü giderici otomatik kodlayıcı, giriş verisine gürültü ekleyip giriş verisinin bozuk bir kopyasını üretir. Bu sayede, otomatik kodlayıcının verilerden çıkartılacak öznitelikleri öğrenmeden giriş verilerinin çıkış verilerine kopyalanmasının önüne geçilmektedir. Giriş verilerinin bir bölümünün sıfır yapılması ile girişin bozulması rastgele bir şekilde yapılabilmektedir. Gürültü giderici otomatik kodlayıcı, bozulmuş giriş verisini tekrar oluştururken çıktıyı orijinal giriş verisine benzetmektedir ve bu sebeple bozulan verileri düzeltmeye çalışmaktadır. Çıkış verileri gürültülü giriş yerine orijinal giriş verileriyle karşılaştırılır. Bu işlem, kayıp fonksiyonunu en düşük seviyeye indirene kadar sürmektedir. Gürültü giderici otomatik kodlayıcının temsili gösterimi Şekil 3.22’de gösterilmektedir. Gürültülü veri orijinal girişe eklendikten sonra gürültülü veriden orijinal giriş verisi elde edilmektedir.



Şekil 0.22. Gürültü giderici otomatik kodlayıcının temsili gösterimi

3.4.3. Varyasyonel Otomatik Kodlayıcı

Varyasyonel otomatik kodlayıcılar, giriş verilerini olasılık temsiline benzetirler. Varyasyonel otomatik kodlayıcı, giriş verilerinin örüntüsünü oluştururken örüntüye karşılık gelen bir değer üretmek yerine çıkarılan özneliklerin olasılıksal bir temsili bulmaktadır. Varyasyonel otomatik kodlayıcının amaç fonksiyonu maksimum olabilirlik (log-likelihood) ve Kullback-Leibler uzaklığı (Kullback-Leibler divergence) olmak üzere iki farklı terimden oluşmaktadır. Maksimum olabilirlik giriş verilerini yeniden üretirken Kullback-Leibler uzaklığı giriş verilerinin temsillerini farklı tutmak için değerlendirilmektedir. Kullback-Leibler uzaklığı, yapı kalitesiyle beraber elde edilen verilerin farklılığını bilhassa etkileyebilmektedir [108,109].

3.5. ESA’da Kullanılan Optimizasyon Yöntemleri

Sinir ağı modellerinde istenilen doğruluk değerlerini elde etmek için ağların katmanlarında kullanılan ağırlıklar, istenilen hata skorunu elde edilinceye kadar ağırlıkların sürekli olarak güncellenmesi gerekmektedir. Güncelleme işlemi belli kurallar gözetilerek yapılmaktadır. Ağ mimarisinde kullanılan ağırlık değerlerinin birbirleriyle olan ilişkisi değerlendirilerek ağırlık değerlerinin güncellenmektedir. Ağırlıkların birbiriyle ilişkisi sinir ağının giriş katmanından çıkış katmanı boyunca sürdüğünden dolayı ağ mimarisinde ileri ve geri yönde devamlı cebirsel işlemler yapılmaktadır. Geri yönde hesaplamalar ‘zincir kuralı’ ile yapılmaktadır. Bir çıkış değerinin kendisini doğrudan etkilemeyen bir giriş değerine bağlı olarak değişimi bulunmak istediğinde zincir kuralı uygulanmaktadır. Örneğin; $z = a_1y + b_1 = a_2x + b_2$ olan bir eşitlikte z değişkeninin

x değişkenine göre değişimi z 'nin y 'ye göre ve y 'nin x 'e göre kısmi türevlerinin çarpımı $(\frac{\Delta z}{\Delta x} = \frac{\Delta z}{\Delta y} \frac{\Delta y}{\Delta x})$ ile bulunmaktadır. Sinir ağı mimarilerinde ise istenilen değeri elde etmek için ağ mimarisinde yer alan öğrenilebilir değerler yitim fonksiyonunun (L) temel alınmasıyla güncellenmektedir.

Makine öğrenmesi mimarilerinde sıklıkla kullanılan optimizasyon yöntemlerinin başında Rastgele Gradyan İnişi (RGİ) optimizasyon yöntemi gelmektedir. RGİ optimizasyon değerleri Denklem 3.13 ile hesaplanmaktadır.

$$\omega_{i+1} = \omega_i - \eta \frac{\partial L}{\partial \omega_i} \quad (3.13)$$

Burada; ω_i ağırlıkları η öğrenme oranını temsil etmekle beraber öğrenme oranı istenilen bir değer olarak seçilebilmektedir. Ancak tüm eğitim boyunca yalnızca bir öğrenme oranı değeri kullanmak uygun değildir. Örneğin; öğrenme oranı büyük bir değer seçilirse yitim fonksiyonunun en az hata skorunun elde edildiği global minimum noktası es geçilebilir. Dahası, öğrenme oranı küçük bir değer seçilirse eğitim için alan süre uzayabilir. Bu sorunun üstesinden gelebilmek için ağırlık parametreleri güncellendiğinde öğrenme oranı için iki farklı yol izlenmektedir. İlk yol olarak, eğitim sürecinin başında, ortasında ve sonunda sırasıyla yüksek, düşük ve daha düşük öğrenme oranları kullanılmaktadır. İkinci ve kullanımı daha uygun olan yol ise öğrenme oranı değerini, bir önceki öğrenme oranı ile birlikte içinde bulunulan zamandaki döngü sayısı ile bağlantılı olarak üstel bir şekilde azaltmaktır. Denklem 3.14'te bu yolla hesaplanan öğrenme oranı değişimi yer almaktadır.

$$\eta_i = \eta_{i-1} e^{-\frac{i}{20}} \quad (3.14)$$

Öğrenme oranında yapılan bu düzenlemelere rağmen öğrenilebilir parametrelerde gürültü bulunuyorsa ilk ağ eğitim değerleri hedef değerlerden çok uzak belirlenmişse eğitim süresi tekrardan uzun zaman almaktadır. Bu sebeple Sutskever vd. [110] RGİ optimizasyon yöntemine bir momentum algoritması entegre etmiştir. Önerilen bu yöntem Rastgele Gradyan İnişi Momentumu (RGİM) optimizasyon yöntemi olarak adlandırılmaktadır. RGİM optimizasyon tekniği derin öğrenme modellerinde sıklıkla kullanılan optimizasyonlardandır. Ağırlıklar, RGİM optimizasyon yöntemi ile Denklem 3.15'teki gibi hesaplanmaktadır.

$$\omega_{i+1} = \omega_i - \eta \frac{\partial L}{\partial \omega_i}(t) - \alpha \frac{\partial L}{\partial \omega_i}(t-1) \quad (3.15)$$

Burada; t , içinde bulunulan döngüde olan değerleri gösterirken $t-1$ ise bir önceki döngüye ait değerleri göstermektedir. α değeri 0 ile 1 arasında değişen bir katsayı olmakla beraber genellikle 0,9 olarak alınmaktadır. Denklem 3.15'ten de anlaşılacağı üzere içinde bulunulan döngünün değişim değerleri, bir önceki döngüde yer alan değişim değerlerinden belli bir oranda etkilenmektedir. Bu sayede, öğrenilebilir değerler hızlı değiştiğinde değişim aniden yavaşlamaz ya da yavaş değiştiğinde değişim aniden hızlanmaz.

3.5.1. Uyarlanabilir Öğrenme Oranları Kullanan Optimizasyon Yöntemleri

Araştırmacılar, ağ mimarisi performansında önemli bir etkiye sahip olan öğrenme hızının düzgün bir şekilde belirlenmesinin en çok zorlayıcı üst değişkenlerden biri olduğunun farkına varmışlardır [110]. Yitim fonksiyonu, parametre uzayında yer alan bazı yönler oldukça duyarlı olabilmektedir. Momentum algoritması ile problemler bir nebze aşılmış olsa da bu işlem gerçekleştirilirken yeni bir üst değişken açığa çıkmıştır. Bu sebeple öğrenme oranları üzerinde yapılan araştırmalar artmıştır. Duyarlılık yönünün verilerde yer alan eksen hizalarına bağlı olduğu düşünüldüğünde her bir değişken için farklı bir öğrenme oranı üreten algoritmalar önerilmiştir.

Delta-bar-delta algoritması eğitim esnasında ağ modelinin değişkenlerini bireysel öğrenme oranlarına adapte etmeye yönelik kullanılan sezgisel bir yaklaşımdır [111]. Bu algortmada verilen bir ağ modeli değişkenine göre yitim fonksiyonunun kısmi türevi alındığında işaret değişmiyorsa öğrenme oranı arttırılmalıdır. Aksi takdirde öğrenme oranı azaltılmalıdır. Bu yaklaşım sadece ‘*Tam Toptan Optimizasyon*’ yöntemlerinde uygulanabilmiştir. Büyük ölçekli veriler kullanıldığında eğitim süresinin çok uzaması bu yaklaşımın dezavantajıdır.

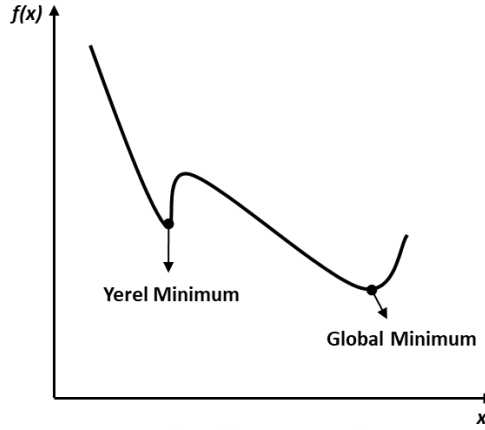
Son yıllarda eğitimleri belirli gruplar şeklinde yapıp değişkenleri adapte eden optimizasyon yöntemleri geliştirilmiştir. AdaGrad algoritması bu yöntemlerden biridir [112]. Algoritma, modelde yer alan bütün değişkenlerin öğrenme oranlarını bütün önceki kare değerlerinin toplamı kareköküyle ters orantılı olacak şekilde uyarlanmaktadır. Yitim fonksiyonunun en büyük kısmi türev değerine sahip olan değişkenlerin öğrenme oranları hızlı bir düşüşe sahiptir. Bununla birlikte küçük kısmi türev değerine sahip değişkenler ise öğrenme oranlarında daha yavaş bir düşüş bulunmaktadır. AdaGrad yaklaşımında ağırlıkların değişimi hesaplanırken Denklem 3.16 ve 3.17 kullanılmaktadır.

$$\Delta\omega_i(t) = -\frac{\eta}{\sqrt{G_i(t)+\epsilon}} \frac{\partial L}{\partial \omega_i}(t) \quad (3.16)$$

$$G_i(t) = G_i(t-1) + \left(\frac{\partial L}{\partial \omega_i}(t) \right)^2, G_i(0) = 0 \quad (3.17)$$

Burada, ϵ küçük değere sahiptir ve $G_i(t)$ fonksiyonunun sıfır olma olasılığına karşı denklemde yer almaktadır. Dışbükey optimizasyon tekniklerinde AdaGrad yaklaşımı teoride istenilen özelliklere sahip olmasına rağmen uygulamada derin sinir ağı mimarilerinin eğitiminde, eğitim başlangıcından sonra gradyan karelerin birikmesinden dolayı etkin öğrenme hızında erken ve aşırı bir düşüşe yol açabilmektedir. Bundan dolayı AdaGrad optimizasyon yöntemi kullanılan bazı derin öğrenme mimarilerinde iyi sonuçlar elde edilirken bazılarında ise tam tersi bir durum hasıl olmaktadır. Bu problemi çözmek amacıyla AdaGrad yaklaşımına ağırlıklı ortalamalar entegre edilerek RMSProp algoritması sunulmuştur [113]. Bir sinir ağı eğitildiğinde dışbükey olmayan bir fonksiyon ile karşılaşıldığı zaman öğrenme yörüngesi çeşitli yapılardan geçerek bir yerel minimum

bölgesine erişebilmektedir. Şekil 3.23’de dışbükey olmayan bir fonksiyonun temsili gösterimi yer almaktadır.



Şekil 0.23. Dışbükey olmayan bir fonksiyonun temsili gösterimi

AdaGrad yaklaşımı, öğrenme oranını gradyan karesinin tüm geçmişine temel alınarak düşürürken öğrenme oranı global minimuma gelmeden çok düşürülmüş olabilmektedir. Bu sorunu çözmek için RMSProp yaklaşımında, geçmişin üzerindeki aşırı geçmişin etkisini silmek için genellikle azalan bir ortalama kullanılmaktadır. RMSProp yaklaşımında Denklem 3.17 güncellenerek hesaplamalar Denklem 3.18’e göre yapılmaktadır.

$$G_i(t) = \beta G_i(t-1) + (1-\beta) \left(\frac{\partial L}{\partial \omega_i}(t) \right)^2, G_i(0) = 0 \quad (3.18)$$

Burada, β 0 ile 1 arasında değişmekte olan değişken ortalamanın skaler değişimini tayin eden bir sayı olmakla beraber genelde 0,9 olarak seçilmektedir.

RMSProp optimizasyon yönteminde momentum değerine ayarlayan bir algoritma mevcut değildir. RMSprop yaklaşımına yeni bir momentum algoritması adapte edilerek etkili bir optimizasyon yöntemi olan Adam algoritması geliştirilmiştir [114]. Adam algoritması, RMSProp algoritmasında Denklem 3.16’da kullanılan $\frac{\partial L}{\partial \omega_i}(t)$ gradyan denklemi yerine $M_i(t)$ momentum parametresinin güncellenmesiyle oluşturulmuştur. Adam algoritması Denklem 3.19 3.20 ile hesaplama yapılmaktadır.

$$\Delta \omega_i(t) = -\frac{\eta}{\sqrt{G_i(t)+\epsilon}} M_i(t) \quad (3.19)$$

$$G_i(t) = \alpha M_i(t-1) + (1-\alpha) \left(\frac{\partial L}{\partial \omega_i}(t) \right)^2, M_i(0) = 0 \quad (3.20)$$

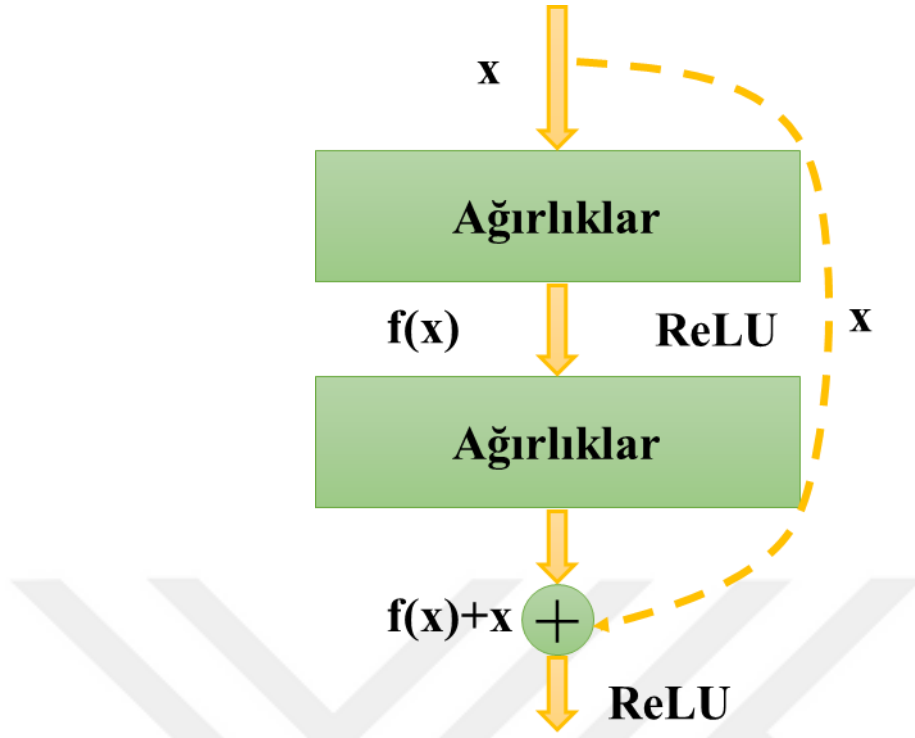
Böylece Adam optimizasyon yöntemi kullanılarak hem momentum değeri hem de öğrenme oranı değeri uyarlanabilir bir biçimde ayarlama yapılabilmektedir.

3.6. Önceden Eğitilmiş ESA Modelleri

Bilgisayarlı görüntü işleme problemlerinin çözümünde ESA'lar 2012 yılından sonra her sene düzenlenen ImageNet yarışmasından sonra popüler hale gelerek geleneksel makine öğrenmesi yöntemlerinin yerini almışlardır. Sınıflandırma, bölütleme, nesne tespiti ve yapay veri üretme gibi görüntü temelli işlemler için farklı ESA mimarileri geliştirilmiştir. Bu bölümde son yıllarda geliştirilen ve bu tez çalışmasında kullanılan etkili ESA modelleri özetlenecektir.

3.6.1. Artık Sinir Ağı (ResNet)

ResNet, He ve ark. [53] tarafından 152 katmanlı derin ESA mimarisi olarak geliştirilmiştir. ResNet, ESA'nın geri yayılımı sırasında meydana gelen gradyan problemini çözmeye çalışmaktadır. ResNet mimarisi, derin ağ eğitimi sırasında bilgi kaybını önlemek için artık bağlantılar (bağlantıyı atlama) sunmuştur. Bağlantıyı atlama tekniği, model performansını iyileştirmek için çok derin ağların eğitilmesini sağlar. Artık bloklar, ResNet mimarisinin ana yapı taşlarıdır. ResNet mimarisi, art arda gelen gizli katmanlar derin olmayan sinir ağlarında bir diğerine bağlanırken, artık bloklar aracılığıyla bağlantılar içermektedir. Eğitim boyunca kazanılan bilgilerin korunması ve modelin eğitim süresinin hızlandırılmasıyla sonuçlanan ağ kapasitesinin artırılması, ResNet mimarisindeki artık bağlantıların en önemli avantajlarından ikisidir. Bu çalışmada ResNet-18 ve ResNet-50 modelleri kullanılmıştır. ResNet-18, 71 katmana sahip bir derin ESA modelidir. ResNet-18 modeli giriş katmanı, 20 adet evrişim katmanı, 17 adet ReLU katmanı, 20 adet toplu normalize katmanı, 8 adet ekleme (addition) katmanı, maksimum havuzlama katmanı, global havuz katmanı, softmax ve sınıflandırma çıktı katmanından oluşmaktadır. ResNet-50 modeli ise 177 katmana sahiptir. Bunlar; giriş katmanı, 53 evrişim katmanı, 49 ReLU katmanı, 53 toplu normalize katmanı, 15 ekleme katmanı, maksimum havuzlama katmanı, ortalama havuzlama katmanı, tam bağlı katman, softmax katmanı ve sınıflandırma çıktı katmanıdır. ResNet'in temel yapısı Şekil 3.24'de gösterilmiştir.



Şekil 0.24. ResNet yapısını gösteren blok diyagram

3.6.2. Uzun Kısa Süreli Bellek (UKSB)

UKSB, özel bir TSA türüdür [16,41,115]. UKSB, tekrarlı gizli katmandaki bellek blokları ve tekrarlı bağlantılar nedeniyle genellikle ileri beslemeli sinir ağlarından ve TSA'lardan çok daha güçlü olduğu kabul edilmektedir [42]. UKSB, sınıflandırma ve regresyon problemlerinde de oldukça etkilidir [41,115].

UKSB'nin bellek blokları, her adımda ağı geçici durumlarını depolayan, kendi kendine bağlanan bellek hücrelerine sahiptir. Bilgi akışı, bellek birimlerine/hücrelerine bir giriş yoluyla sağlanır. Daha sonra oradan kapıları kullanarak diğer birimlere geçer. Bir giriş olarak bellek hücresine eklemeyen önce hücrenin iç durumunu ayarlamak için bir unutma birimi (kapı) kullanılır. Bu işlem hafıza hücresinin kendisini tekrar ederek ve gerekirse hücrenin hafızasını yeniden ayarlayarak veya unutmasıyla gerçekleştirilir. Unutma kapısı, aşağıdaki gibi tanımlanan tek katmanlı bir sinir ağı ile bir etkinleştirme fonksiyonu tarafından kontrol edilir;

$$f_t = \alpha(W[x_t, h_{t-1}, C_{t-1}] + b_f) \quad (3.21)$$

Burada C_{t-1} , h_{t-1} , x_t ve b_f sırasıyla önceki UKSB blok belleği, önceki bloğun çıkışı, giriş dizisi ve yanılğı (bias) vektörüdür. Lojistik sigmoid fonksiyonu ve her giriş için atanan ağırlık vektörü sırasıyla α ve W olarak gösterilmektedir. Aktivasyon operatörü, önceki bellek yapısına/bloğa uygulanır. Aktivasyon operatörü, öge bazında birikim (çarpma) ile devam eden

UKSB üzerindeki önceki bellek yapısını/blok etkisini belirler. Aktivasyon seti/vektör çıktısının değeri kontrol edilir ve sifıra yakın ise önceki hafıza unutulur.

Giriş kapısında, basit bir sinir ağı, önceki bellek bloğunun etkisini ve $\tanh$ aktivasyon fonksiyonunu hesaba katarak yeni bir bellek üretir. İlgili süreç Denklem 3.22-23 ile hesaplanmaktadır.

$$i_t = \alpha(c[x_t, h_{t-1}, C_{t-1}] + b_i) \quad (3.22)$$

$$C_t = f_t C_{t-1} + i_t \tanh(W[x_t, h_{t-1}, C_{t-1}] + b_c) \quad (3.23)$$

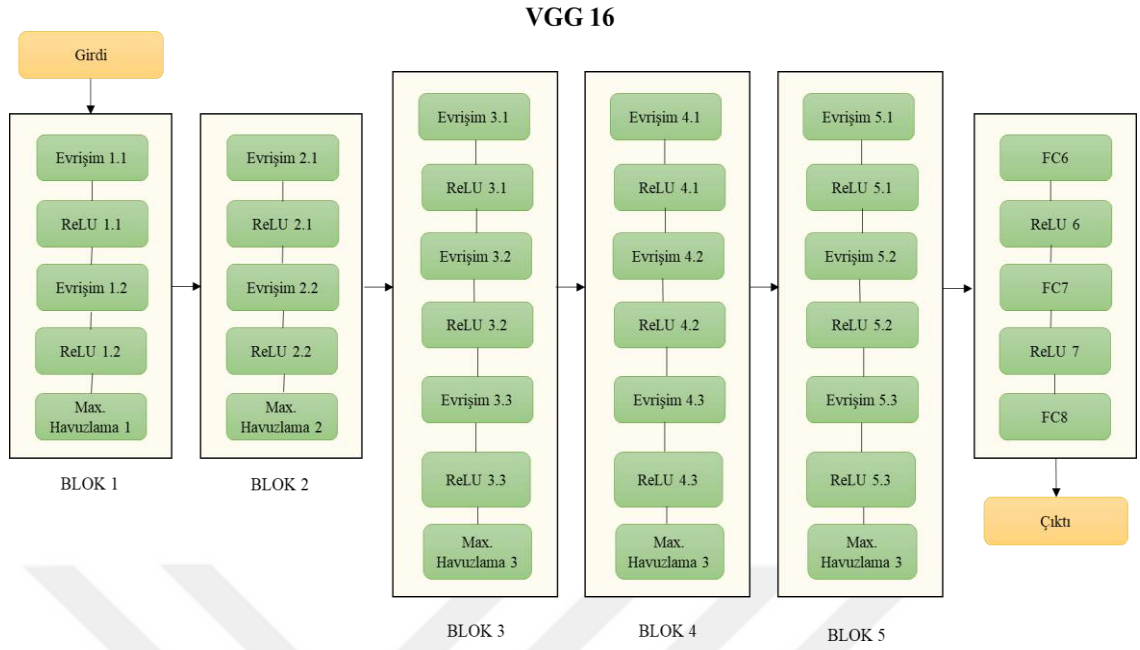
Burada, b_i ve W , sırasıyla giriş kapısının, yanılğı (bias) vektörünün ve ağırlıkların sonucunu ifade etmektedir. h_{t-1} önceki blokun sonucunu, C_{t-1} önceki UKSB hafızasını α parametresi ise aktivasyon fonksiyonunu göstermektedir [41,43]. Çıktı (sonuç) kapısı için devam eden UKSB yapısının/blokunun sonucunun aşağıdaki Denklem 3.24-25 kullanılarak üretildiği düşünülebilir.

$$o_t = \alpha(W[x_t, h_{t-1}, C_{t-1}] + b_o) \quad (3.24)$$

$$h_t = \tanh(C_t).o_t \quad (3.25)$$

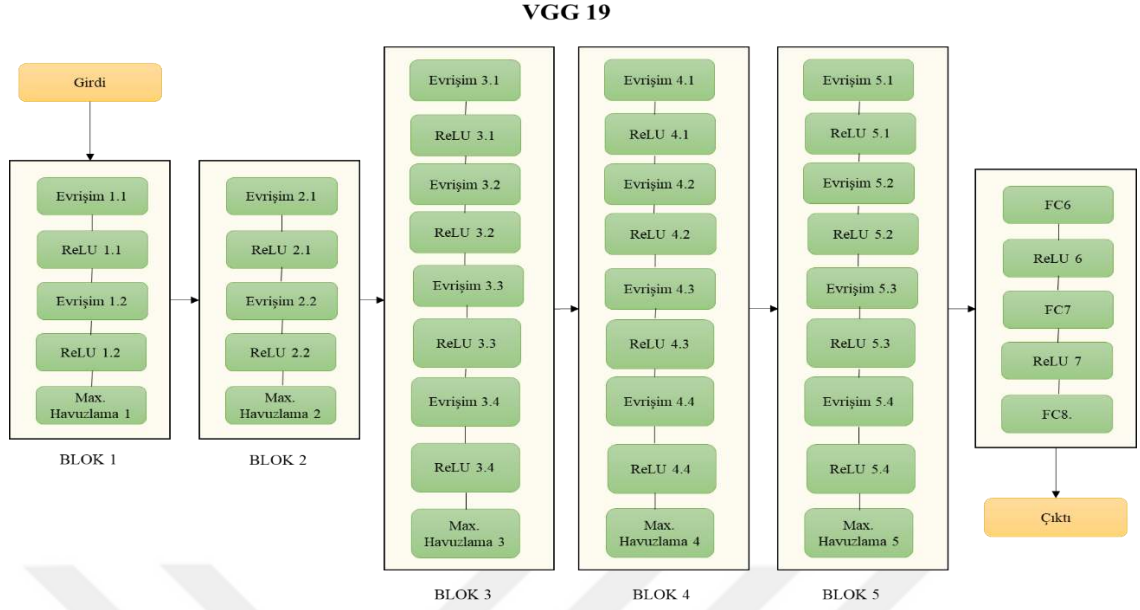
3.6.3. VGGNet

VGGNet, 2014 yılında düzenlenen ILSVRC görüntü tanıma yarışmasında Simonyan vd. [59] tarafından yeni bir ESA modeli olarak tasarlanmıştır. Blok yapısı bakımından AlexNet [66] modeline benzemektedir. Bunun yanında VGGNet modelinde 3×3 boyutundaki evrişim filtreleri kullanılmıştır. Bu filtreler birer adımlık ilerlemelerle kullanılmaktadır. VGG 16 ve VGG19 olmak üzere iki farklı VGGNet modeli bulunmaktadır. VGG16 modeli, 16'sı öğrenilebilir ağırlıklara sahip toplam 41 katmana sahiptir. Öğrenilebilir katmanlar, 13 evrişim katmanından ve 3 tam bağlı katmandan oluşmaktadır. Maksimum havuzlama katmanları, evrişim katmanlarından sonra gelmektedir. VGG 16'nın yapısı Şekil 3.25'de verilmiştir.



Şekil 0.25. VGG16 mimarisi

VGG19 modeli ise VGG16 modelinin daha derin bir versiyonu olarak geliştirilmiştir. VGG19 modeli 16 evrişim katmanı, 18 ReLU katmanı, 5 maksimum havuzlama katmanı, 2 bırakma (dropout) katmanı, 3 tam bağlı katman, softmax ve sınıflandırma çıktı katmanının bulunduğu 47 katman içermektedir. Şekil 3.26’da VGG 19 modelinin yapısı yer almaktadır.

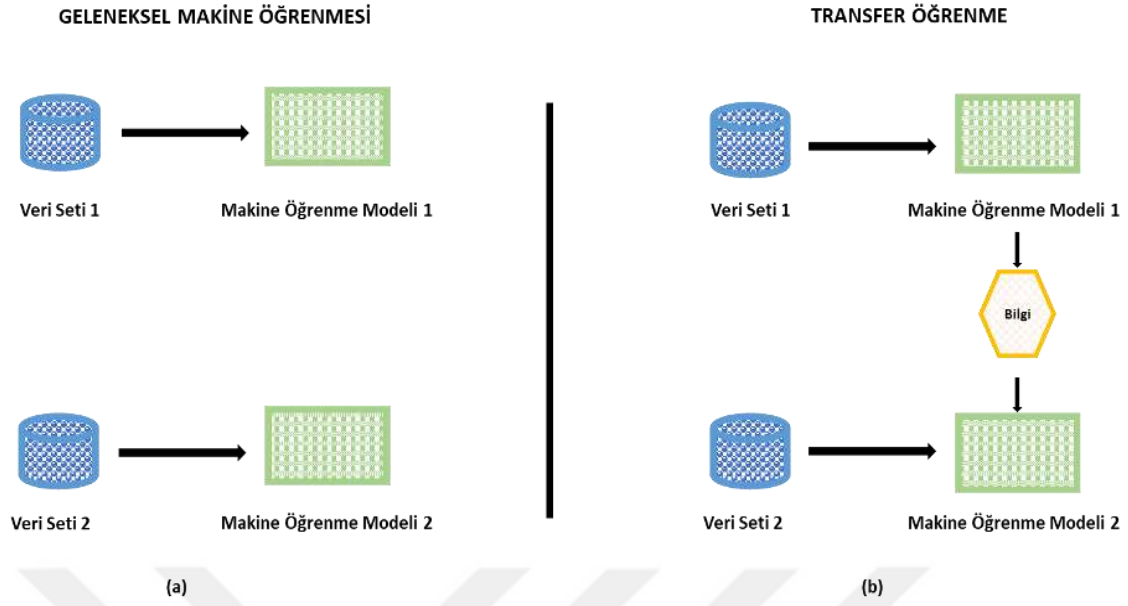


Şekil 0.26. VGG19 mimarisi

3.6.4. Transfer Öğrenme

Klasik makine öğrenmesi ve veri madenciliği yöntemleri sınıfı belli olan veya sınıfı belli olmayan veri setleri ile eğitilmiş istatistiksel algoritmaları kullanarak yeni veriler üzerinde tahminde bulunmaktadır. Transfer Öğrenme (TÖ), temelini çoklu görev öğrenme, bilgi aktarımı ve öğrenmeyi öğrenme gibi çeşitli isimler ile yapılan çalışmalardan almaktadır. Ortak olan gizli özellikleri ortaya çıkartmak için kullanılan çoklu görev öğrenme mimarisi TÖ ile yakından ilişkilidir. TÖ'nün kullanım amacı, bir bilgiyi bir ya da daha fazla kaynaktan görevden elde etmek ve elde edilen bu bilgiyi yeni bir göreve tatbik etmektir. Çoklu görev öğrenmede var olan tüm kaynak ve hedef görevlerinin aynı anda öğrenilmesine karşın TÖ yeni görevle alakalıdır [116].

Şekil 3.27'te klasik makine öğrenme modeli ile TÖ modeli gösterilmektedir. Şekil 3.27a 'da görülebileceği gibi klasik makine öğrenme yapıları birbirinden ayrı olarak belli bir veri setine uygulanmaktadır. Bu yapılar arasında bilgi aktarımı yapılamamaktadır. Bununla birlikte Şekil 3.27b'de anlaşılabacağı üzere TÖ'de bilgi aktarımı bulunmaktadır. TÖ yöntemi ile yeni mimarileri eğitmek için daha önceden eğitilen mimarilerden elde edilen öznitelikler yeni mimarilerde kullanılabilir. Yeni modele iletilen bu özniteliklerden dolayı daha iyi ve hızlı sonuçlara ulaşılmaktadır. Özellikle küçük ölçekli veri setlerinde iletilen bu özniteliklerin önemi oldukça büyüktür.



Şekil 0.27. Klasik makine öğrenmesi modeli (a) ve transfer makine öğrenme modeli (b)

Büyük ölçekli veri setlerine uygulanan mimarilerden elde edilen öznitelikler başka mimarilerde de kullanılabilir. Derin sinir ağları, daha önceden eğitilmiş mimarilerden çıkartılan öznitelikleri farklı görevler içinde kullanabilmektedir. Örneğin; ImageNet [117] veri seti yaklaşık bir milyon görüntü ve 1000 farklı sınıftan oluşmaktadır. ImageNet gibi büyük ölçekli veri setlerinde kullanılan farklı ESA mimarileri bu veri setlerine ait farklı öznitelikleri çıkartabilmektedir. Çıkartılan bu öznitelikler farklı veri setleri içinde TÖ ile kullanılabilir. Daha önceden farklı nesnelerin öğrenilmesi için uygulanan algoritmalar ile her katmanından çıkartılan farklı öznitelikleri daha küçük ölçekli veri setlerine iletilerek daha iyi sonuçlar elde edilebilir [118].

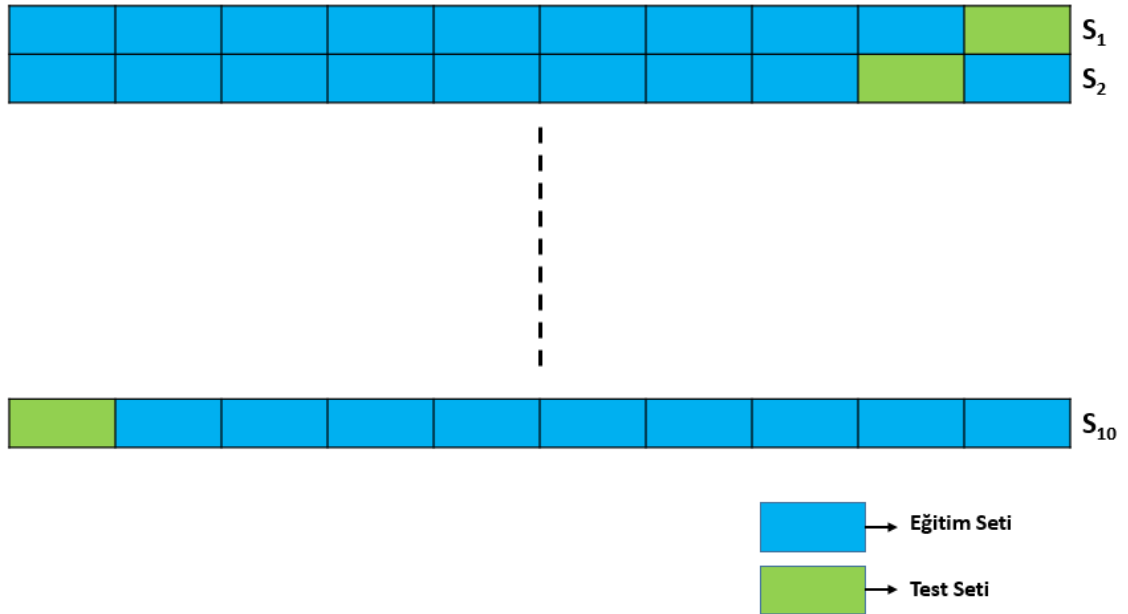
3.7. Sınıflandırmada Eğitim ve Test Verileri

Makine öğrenmesi için geliştirilen yöntemlerin performansının denemesi için verilerin eğitim ve test verisi şeklinde ayrılmalıdır. Eğitim verileri ile en iyi sınıflandırmayı yapabilecek öğrenilebilir değişkenler hesaplanmaktadır. Öte yandan test verileri ile öğrenilmiş olan bu değişkenlerin sınıflandırma doğruluk performansı ölçülmektedir.

Eğitim ve test verilerinin hangi şekilde ayrılacağı ile alakalı olarak birkaç farklı yöntem mevcuttur. Eğitim ve test için belli bir oranda verilen bölünmesi bu yöntemlerden bir tanesidir. Örnek olarak %75'i eğitim ve %25'i test verisi olacak şekilde ayırma yapılabilir. Ayrıca bölme işleminin rastgele bir şekilde yapılması önerilen yöntemin geçerliliğini arttırmaktadır. Dahası, bu rastgele bölmenin birkaç kez yinelenmesi de önerilen yöntemin performans geçerliliği

hakkında bilgi vermektedir. Eğitim ve test verilerinin hangi şekilde ayrılacağına dair bir başka yöntemde ise bölünme işlemi veriyi oluşturanlar tarafından yapılmaktadır. Bu yöntemde verilerin bölünmesi belli bir plana göre yapıldığından dolayı performans ölçümünde geçerliliği bulunmaktadır. Bazı veri setleri bu yöntemle göre hazırlanmıştır.

K-Kat çapraz geçerleme metodu önerilen ağ mimarisinin performans ölçümünde eğitim ve verileri ayırma yöntemlerindedir. Bu metotta veri seti K adet rastgele eşit parçaya bölünmektedir. K adet parçadan bir adedi her katta test işleminde kullanılmak üzere ayrılmaktadır. Veri setinden geriye kalan K-1 adet parça birleştirilip o katın eğitiminde kullanılmaktadır. Her kat için elde edilen sınıflandırma doğruluklarının ortalaması alınarak modelin doğruluk performansı elde edilmektedir. Bu yöntemde K sayısı oldukça büyük önem arz etmektedir. K'nin çok büyük seçildiği durumlarda eğitim setinin yüzdesi oldukça büyük olurken test setinin yüzdesi çok küçük olmaktadır. Bu sebeple etkili bir test işlemi gerçekleşmez. Ek olarak, öğrenilebilir değişken sayısı eğitim setinin büyüklüğüyle doğru orantılı olduğu için öğrenme süresinin uzamasına yol açmaktadır. K değerinin çok küçük seçildiği durumlarda ise sınıflandırma performansı düşmektedir. Bu yöntemde kat sayısı genellikle 10 alınmaktadır. Son yıllardaki performansı yüksek işlemcilerin gelişimiyle bir modelde K-Kat çapraz geçerleme yönteminin kullanımı birden fazla yapılmaya başlanmıştır [119]. K-Kat çapraz geçerlemenin birden fazla defa yapılması ile daha doğru bir hata tahmini yapılabilecektir. Şekil 3.28'de K değeri 10 olarak seçilmiş 10- Kat çapraz geçerlemenin temsili gösterimi yer almaktadır.



Şekil 0.28. K değeri 10 olarak seçilmiş 10- Kat çapraz geçerlemenin temsili gösterimi

3.8. Sınıflandırma Performansı Ölçümü

Sınıflandırma işleminden sonra uygulanan yöntemlerin performans ölçümleri için karmaşıklık matrisi ve Yanlış Ret Oranı (False Rejection Rate, FRR), Yanlış Kabul Oranı (False Acceptance Rate, FAR) ve Eşit Hata Oranı (Equal Error Rate, EER) kullanılmıştır. Şekil 3.29'da karmaşıklık matrisinin temsili gösterimi yer almaktadır.

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

Şekil 0.29. Karmaşıklık matrisi

Karmaşıklık matrisinde yer alan değerlerde Doğru Pozitif (DP) doğru bir şekilde sınıflandırılan değerleri, Yanlış Pozitif (YP) başka sınıfta yer alması gerekirken ilgili sınıfta olan değerleri, Yanlış Negatif (YN) ilgili sınıfta yer alması gerekirken başka sınıfta olan değerleri göstermektedir. Doğru Negatif (DN) ise doğru sınıflandırılan diğer sınıftaki değerleri ifade etmektedir. Karmaşıklık matrisinde yer alan değerlerden faydalanılarak en sık kullanılan doğruluk, duyarlılık ve özgüllük bu tez çalışmasında sınıflandırma performans ölçümleri için kullanılmıştır. Doğruluk, duyarlılık ve özgüllük ölçüm kriterlerinin denklemleri aşağıda verilmiştir.

$$\text{Doğrluk} = \frac{DP+DN}{DP+DN+YP+YN} \quad (3.26)$$

$$\text{Duyarlılık} = \frac{DP}{DP+YN} \quad (3.27)$$

$$\text{Özgüllük} = \frac{DN}{DN+YP} \quad (3.28)$$

Ayrıca, geliştirilen mimarının etkinliği, FRR, FAR ve EER kullanılarak değerlendirilmiştir. FRR, sahte olarak sınıflandırılan gerçek örneklerin yüzdesidir, FAR ise yanlış bir şekilde gerçek örnekler olarak sınıflandırılan sahte örneklerin yüzdesidir. EER ölçüsü, FAR, FRR'ye eşit olduğunda, yani %FAR = %FRR olduğunda sistemin doğruluğunu tanımlamaktadır.

4. DERİN ARTIK AĞ VE UKSB MİMARİLERİNİN HİBRİT MODELİ KULLANILARAK DEEPFAKE YÜZ VİDEOLARININ TESPİTİ

Bu uygulamada deepfake videolarının tespiti için derin öğrenme tabanlı hibrit bir yöntem geliştirilmiştir. Önceden eğitilmiş bir ESA modeliyle derin özniteliklerin çıkarılması ve Uzun Kısa Süreli Bellek (UKSB) sınıflandırma için kullanılmıştır.

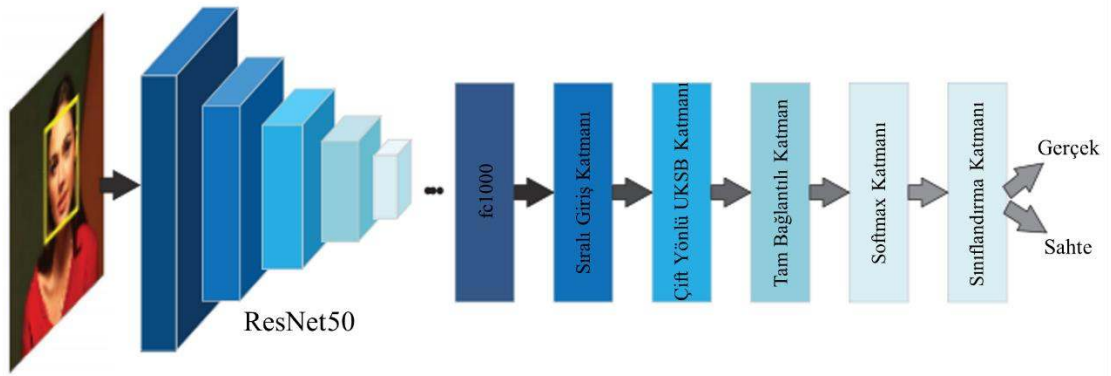
4.1. Giriş

Son yıllarda yapay zeka teknolojilerinin hızlı gelişimiyle birlikte sayıları gün geçtikçe artan yeni uygulamalar ve hizmetler sunulmaktadır. Günümüzde bu teknolojilerin çok çeşitli uygulamaları bulunmaktadır. Bunlardan biri yapay olarak değiştirilen veya makine öğrenimi üretim modellerinden yararlanılarak oluşturulan tüm multimedya içeriklerini ifade eden deepfake'dir. Deepfake, bir görüntü veya videoda bir kişinin yüz görüntüsü üzerinde yapılan oynamalara verilen addır. Bu aynı zamanda yüz manipülasyonu olarak da bilinir. Yapay öğrenme tabanlı farklı algoritmalar veya uygulamalar kullanılarak bu manipülasyonlar gerçekleştirilebilir. Bu bölümde deepfake videoları için derin öğrenme tabanlı hibrit bir yöntem geliştirilmiştir. Önerilen yöntem yüz algılama, derin özniteliklerin çıkarılması ve sınıflandırma işlemlerinden oluşmaktadır. Önerilen yöntemde belirli bir video için ilk olarak her çerçevede yüz bölgeleri tespit edilmektedir. Tespit edilen yüz bölgeleri, öznitelik çıkarmak için önceden eğitilmiş bir ESA modeline (yani önceden eğitilmiş derin artık ağ modelinin FC1000 katmanına) giriş olarak verilmektedir. Çıkarılan öznitelikler daha sonra sınıflandırma için yedi katmanlı UKSB modelinde [yani giriş, iki adet çift yönlü UKSB (biUKSB), bırakma, tam bağlı, softmax ve çıkış katmanları] kullanılmaktadır. Deneyler için DeepFakeTIMIT ve Celeb-DF veri setleri kullanılmıştır. Önerilen yöntem literatürde yer alan diğer yöntemlerle karşılaştırılmıştır. Önerilen yöntem, daha önce önerilen deepfake tespit yöntemlerinden daha iyi performans göstermiştir.

4.2. Önerilen Yöntem

Yüz deepfake tespiti, iki sınıflı bir sınıflandırma problemi olarak düşünülebilir. Bu problemde, giriş videosu gerçek ya da sahte olarak belirtilmelidir. Şekil 4.1 geliştirilen deepfake tespit yönteminin genel mimarisini göstermektedir. Önerilen yöntem, verilen çerçevelerde yüz tespiti, tespit edilen yüz bölgelerinde öznitelik çıkarımı ve öznitelik sınıflandırması olmak üzere üç aşamadan oluşmaktadır. Bu çalışmada, yüzleri algılamak için kademeli yüz dedektörü kullanılmıştır [120]. Özellikle yüzleri, burunları, gözleri ve ağzı tespit eden kademeli yüz dedektörü olan Viola-Jones algoritması [121] kullanılmıştır. Tespit edilen yüz bölgeleri daha sonra kırpılmış ve ResNet-50 modelinin girişiyle uyumlu olması için 224×224 olarak yeniden boyutlandırılmıştır.

ResNet-50 modelinin tam bağı katmanı (fc1000), kırılmış yüz bölgeleri için derin özellikleri çıkarmak için kullanılmıştır. En sonunda sınıflandırma işlemini gerçekleştirmek için UKSB sınıflandırıcısı kullanılmıştır. Bu çalışmada, geliştirilen mimarinin etkinliği yanlış ret oranı (False Rejection Rate, FRR), yanlış kabul oranı (False Acceptance Rate, FAR) ve eşit hata oranı (Equal Error Rate, EER) kullanılarak değerlendirilmiştir. FRR, sahte olarak sınıflandırılan gerçek örneklerin yüzdesi, FAR ise yanlış bir şekilde gerçek örnekler olarak sınıflandırılan sahte örneklerin yüzdesidir. EER ölçüsü, FAR, FRR'ye eşit olduğunda, yani $FAR\% = FRR\%$ olduğunda sistemin doğruluğunu tanımlamaktadır.



Şekil 0.1. Önerilen deepfake tespit mimarisi

4.3. Veri Seti

Bu çalışmada halka açık iki farklı veri seti kullanılmıştır. Bunlar DeepFakeTIMIT ve Celeb-DF veri setleridir.

4.3.1. DeepFakeTIMIT Veri Seti

DeepFakeTIMIT veri seti [10], VidTIMIT [122] veri seti kullanılarak oluşturulan düşük kaliteli (LQ) ve yüksek kaliteli (HQ) olmak üzere deepfake görüntüleri içeren iki eşit boyutlu alt kümeden oluşmaktadır. Hem LQ hem de HQ alt kümeleri, sırasıyla 64×64 ve 128×128 pikseli 320 video içermektedir. Bu çalışmada, daha zorlayıcı alt küme olduğu için HQ alt kümesi kullanılmıştır.

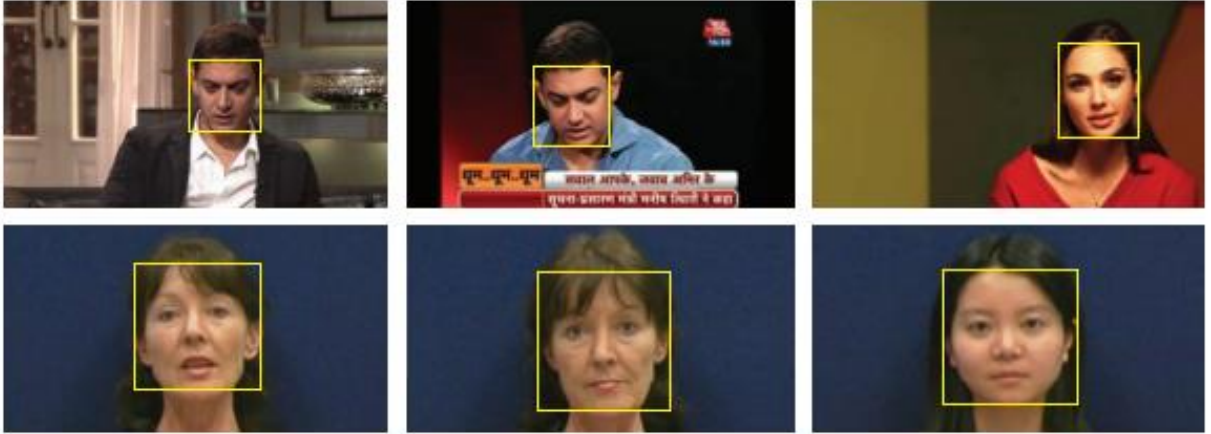
4.3.2. Celeb-DF Veri Seti

Celeb-DF veri seti [14], 590 gerçek ve 5.639 sahte deepfake videodan oluşmaktadır. İki milyondan fazla video çerçevesi içermektedir. Gerçek videolar, farklı cinsiyet, yaş grupları ve etnik gruplara sahip 59 ünlünün YouTube videolarından alınmıştır. Deepfake videoları, belirtilen 59

deneğin her bir çifti için yüzlerinin birbirleriyle değiştirilmesiyle oluşturulmuştur. Şekil 4.2 her iki veri setinden gerçek video çerçevelerinin bazı örneklerini gösterirken Şekil 4.3 tespit edilen yüz bölgelerini göstermektedir. Benzer şekilde, Şekil 4.4 her iki veri setinden alınan sahte video çerçevelerinin örneklerini gösterirken ve Şekil 4.5 tespit edilen sahte yüz bölgelerini göstermektedir.



Şekil 0.2. Gerçek videolardan alınan bazı örnek çerçeveler. İlk satırda Celeb-DF [14] veri setinden alan bazı çerçeveler yer alırken ikinci satırda ise DeepFakeTIMIT [10] veri setinden bazı çerçeveler gösterilmektedir.



Şekil 0.3. Verilen çerçevelerde algılanan yüz bölgeleri. İlk satır Celeb-DF veri seti için tespit edilen bazı yüzleri gösterirken ikinci satır ise DeepFakeTIMIT veri seti için tespit edilen bazı yüzleri göstermektedir.



Şekil 0.4. Sahte videolardan alınan bazı örnek çerçeveler. İlk satır Celeb-DF veri setinden alınan bazı sahte çerçeveleri, ikinci satır ise DeepFakeTIMIT veri setinden alınan bazı sahte çerçeveleri göstermektedir.



Şekil 0.5. Verilen sahte çerçevelerde algılanan yüz bölgeleri. İlk satır, Celeb-DF veri seti için tespit edilen bazı sahte yüzleri gösterirken ve ikinci satır ise DeepFakeTIMIT veri seti için tespit edilen bazı sahte yüzleri göstermektedir.

4.4. Deneysel Çalışmalar

Önerilen yöntemde ön işleme aşamasında daha önce bahsedildiği gibi, tespit edilen yüz bölgeleri ResNet-50 modelinin girişi ile uyumlu olması için 224×224 olarak yeniden boyutlandırılmıştır. Yeniden ölçeklenen yüz bölgeleri, öznetelikleri çıkarmak için ResNet-50 modeline girdi olarak verilmiştir. Girdi katmanının boyutu 1.000 olarak ayarlanmıştır. biUKSB katmanının gizli elemanlarının boyutu 100 olarak seçilmiştir ve iki adet tam bağlı katman kullanılmıştır. Tablo 4.2'de eğitim kriteri, değerleri ve parametreleri sunulmuştur. Veri setlerinin %75'i önerilen yöntemin eğitiminde, kalan %25'i ise önerilen yöntemin test edilmesinde kullanılmıştır. Tablo 4.1 ve burada açıklanan parametreler kullanılarak eğitim prosedürü 7.560 tekrar üzerinden gerçekleştirilmiştir. Her devirde (epoch) yinleme sayısı 63 olarak belirlenmiştir.

UKSB ağına uygulanmadan önce verileri normalize etmek için sıfır merkez normalizasyon kullanılmıştır. Bu amaçla, tüm dizi hesaplamasının her bir öznitelik açısından ortalaması ve standart sapması gerçekleştirilmiştir. Daha sonra, her eğitim girdisi/örneği için, standart sapma ile bölme ile ortalama değer çıkarılmıştır. Bu çalışmada UKSB ağı için eğitim prosedürü olarak “Adam” seçilmiştir.

Tablo 0.1. Eğitim parametre değerleri

Parametreler	Değerler
Maksimum Devir	120
Mini-Batch Boyutu	20
İlk Öğrenme Oranı	0,001
Öğrenme Oranı Programı	Bölümlü
Öğrenme Bırakma Süresi	100
Öğrenme Bırakma Faktörü	0,001
Gradyan Eşik	1

Tablo 4.2'de, FAR = %10 eşiği kullanarak her iki veri seti için EER ve FRR@FAR10% açısından tasarlanmış deepfake video tespit mimarisinin sonuçları yer almaktadır. Tablo 4.2'de her iki veri seti için önerilen yöntemin makul sonuçlar üretebildiği görülmektedir. Örneğin DeepFakeTIMIT veri seti için 2.4217 EER(%) ve 0.0795 FRR@FAR10% (%) oranları elde edilirken, Celeb-DF veri seti için 0.5014 EER(%) ve 0FRR@FAR10% (%) oranları elde edilmiştir. Bu sonuçlar literatürdeki diğer yöntemlerle karşılaştırıldığında diğer yöntemlerden daha iyi bir sonuç elde edildiği görülmüştür.

Tablo 0.2. Önerilen yöntemin her iki veri kümesinde EER (%) ve FRR@FAR10% (%) cinsinden performans değerlendirmesi

Veri Seti	EER (%)	FRR@FAR10% (%)
DeepFakeTIMIT	2.4217	0,0795
Celeb-DF	0.5014	0,00

Ayrıca önerilen deepfake video tespit yönteminin performansı literatürdeki mevcut yöntemlerden bazıları ile karşılaştırılmıştır. Tablo 4.3, [46]'de benimsenen çeşitli yerel tanımlayıcılar kullanılarak DeepFakeTIMIT veri seti kullanılarak elde edilen bazı sonuçları göstermektedir. Bunlar; LBP, Yönlendirilmiş Gradyanların Histogramı Piramidi (Pyramid of Histogram of Oriented Gradients, PHOG), SIFT, CENTRIST, BSIF, Yerel Faz Örnekleme (Local Phase Quantization, LPQ), BGP, Kuaterniyonik Yerel Sıralama İkili Modeli (Quaternionic Local

Ranking Binary Pattern, QLRBP), FDLBP ve Hızlandırılmış Güçlü Özellikler (Speeded Up Robust Features, SURF)'dir [46].

Tablo 0.3. Önerilen yöntemin DeepFakeTIMIT veri setindeki mevcut yöntemlerle EER (%) ve FRR@FAR10% (%) cinsinden performans karşılaştırması

Yöntem	EER (%)	FRR@FAR10% (%)
IQM + SVM [10]	8.97	9,05
LPB [46]	17.16	43,02
FDLBP [46]	37.19	88,30
QLRBP [46]	27.70	59,49
BGP [46]	13.33	15,99
LPQ [46]	13.69	16,53
BSIF [46]	60.88	93,69
CENTRIST [46]	11.43	13,12
PHOG [46]	89.70	100
SIFT [46]	57.58	95,43
SURF [46]	67.26	98,17
Önerilen Yöntem	2.42	0,080

Tablo 4.3'te, [46]'de kullanılan yerel görüntü tanımlayıcıları arasında, CENTRIST tanımlayıcısının %11,43 EER(%) ve %13,12 FRR@FAR10% (%) gibi daha iyi sınıflandırma doğruluğuna ulaştığı görülebilir. Ayrıca, BGP ve LPQ tanımlayıcıları sırasıyla %13.33 EER(%) ve %15.99 FRR@FAR %10 (%), %13.69 EER(%) ve %16.53 FRR@FAR %10 (%) oranlarını elde etmişlerdir. En kötü başarı, %89.70 EER(%) ve %100 FRR@ FAR10 (%) oranları ile PHOG tanımlayıcıları tarafından üretilmiştir. Bununla birlikte, önerilen yöntem ile %2,43 EER(%) ve %0,080 FRR@FAR10 (%) oranları elde edilerek diğer dikkate alınan yerel tanımlayıcılardan daha iyi performans göstermiştir. Benzer şekilde, önerilen yöntem, [10]'da geliştirilen görüntü kalitesi öznetelik tabanlı yöntemden daha iyi sonuçlar elde etmiştir. Bu çalışmada sunulan yöntemin bu yüksek performansı, derin ESA modelinde hem renk hem de doku özelliklerinin birleştirildiği derin özellikler sayesinde elde edilmiştir.

Tablo 4.4'de, önerilen mimarinin Celeb-DF veri seti kullanılan önceki yöntemlerle EER (%) cinsinden bir karşılaştırması yapılmıştır. Li vd. [123] Face X-ray-Blended ve son teknoloji manipülasyonlar için Face X-rayBlended+FaceForensics gibi farklı eğitim prosedürlerine sahip ESA tabanlı yöntemler önermişlerdir. Öte yandan Rossler vd. [22], deepfake video tespiti için ince ayarlı önceden eğitilmiş XceptionNet'i [60] kullanmışlardır. Tablo 4.4'de, geliştirilen mimarinin deepfake videolarını tespit etmede önceki yöntemlerden daha iyi performans gösterdiğini görülebilmektedir. Örneğin, önerilen yöntem %0,5014 EER oranına ulaşırken Face X-rayBlended

[123] ve Xception tabanlı yöntem [22] sırasıyla %31,16 ve %59,64 oranlarına ulaşmıştır. Önerilen yöntemle karşılaştırıldığında, [22] ve [123]'de geliştirilen mimariler, yazarlar tarafından da belirtildiği gibi, daha düşük hata oranları elde etmek için büyük eğitim veri setleri gerektirmektedir. Ancak, bu çalışmada geliştirilen mimari, iyi performanslar elde etmek için nispeten daha az sayıda eğitim örneğine ihtiyaç duymaktadır.

Tablo 0.4. Önerilen yöntemin Celeb-DF veri setindeki mevcut yöntemlerle EER (%) cinsinden performans karşılaştırması.

Yöntem	EER (%)
Face X-ray-Blended [22]	31,16
Xception [60]	59,64
Face X-ray-Blended+FaceForensics [22]	26,70
Önerilen Yöntem	0,5014

5. ÇOKLU AKIŞ ESA MODELLERİNİN ÖZELLİKLERİ BİRLEŞTİREREK DEEPPAKE VE YÜZ MANİPÜLASYONLARINI TESPİT ETME

Bu çalışmada, belirli bir videoda sahte yüz tespiti için yeni bir yaklaşım önerilmiştir. Önerilen yaklaşımda, çoklu akış bir derin ESA mimarisi geliştirilmiştir. Çoklu akış ESA mimarisinde, önceden eğitilmiş üç ESA modeli, bir füzyon katmanı aracılığıyla birleştirilerek sınıflandırma işlemi yapılmıştır.

5.1. Giriş

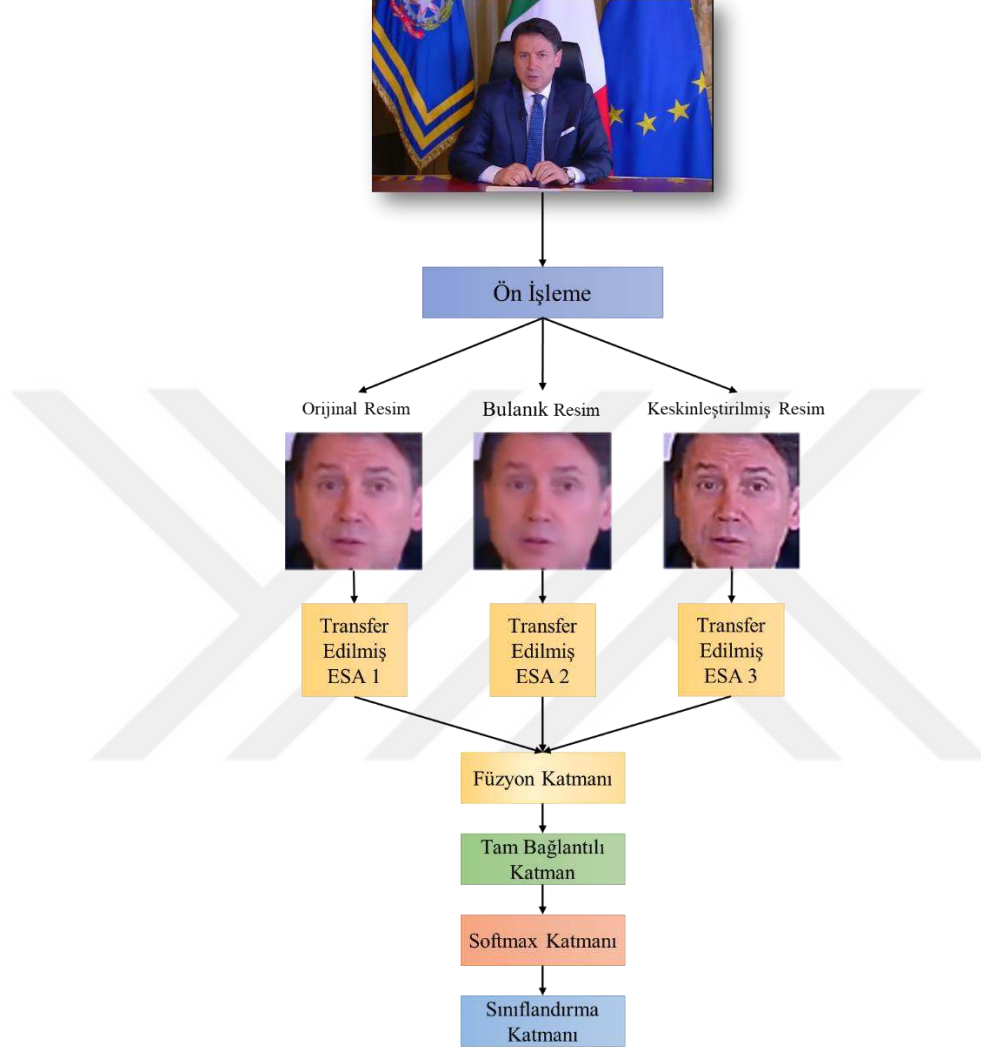
Bu bölümde, deepfake tespit algoritmalarını geliştirmek için büyük ölçekli yeni bir veri seti olan Dünya Politikacılar Deepfake Veri Seti (DPDVS, World Politicians Deepfake Dataset (WPDD)) tanıtılmıştır. Belirli bir videoda sahte yüz tespiti için yeni bir çoklu akış (multi-stream) derin öğrenme yöntemi önerilmiştir. Önerilen mimari, füzyon katmanında birleştirilen üç akıştan oluşmaktadır. Füzyon katmanından sonra, sırasıyla tam bağlı, softmax ve sınıflandırma katmanları üzerinden sınıflandırma gerçekleştirilmektedir. Önceden eğitilmiş VGG16 modeli, birinci akış için benimsenmiştir. İkinci akışta, önceden eğitilmiş VGG19 modeli kullanılmıştır. Üçüncü akışta ise önceden eğitilmiş ResNet-18 modeli kullanılmıştır.

5.2. Önerilen Yöntem

Önerilen çoklu akış derin öğrenme yöntemi Şekil 5.1'de gösterilmektedir. Birinci akışta kafa şekli ve saç rengi gibi genel özellikleri öğrenmek için veri setinde yer alan ön işleme tabii tutulmamış gerçek ve sahte görüntüler kullanılmıştır. İkinci akışta veri setinde yer alan politikacıların ten rengini öğrenmek için veri setinde yer alan görüntülere ön işlem uygulanarak elde edilen bulanık görüntüler girdi olarak kullanılmıştır. Son akışta ise keskinleştirilmiş gerçek ve sahte resimler kişilerin yüzlerinin yerel özelliklerini öğrenmek için kullanılmıştır. Ön işleme aşamasında yüz tanıma, yüz bulanıklaştırma ve yüz keskinleştirme işlemleri uygulanmıştır. Tespit edilen yüz bölgeleri başlangıçta kırpılmış ve ardından 224×224 olarak yeniden boyutlandırılmıştır. Keskinlik, iki veya daha fazla renk arasındaki kontrast olarak tanımlanır. Görüntüler keskinleştirildiğinde, farklı renklerin kesiştiği kenarlardaki kontrast artar. Tespit edilen yüz görüntülerinin bulanıklaştırılması için kılavuzlu filtre (guided filter) kullanılmıştır. Kılavuzlu (guided) filtrenin kare penceresinin yarıçapı 12 olarak seçilmiştir.

Önceden eğitilmiş VGG16 modeli, birinci akış için benimsenmiştir. Transfer öğrenmede, önceden eğitilmiş ESA modelinin ağırlıkları, yeni sınıflandırma problemini eğitmek için daha fazla kullanılmıştır. İkinci akışta önceden eğitilmiş VGG19 modeli kullanılmıştır. Son olarak üçüncü

akışta ise önceden eğitilmiş ResNet-18 modeli dikkate alınmıştır. Şekil 5.1’de deepfake sınıflandırması için önerilen çoklu akış ESA mimarisinin temsili gösterimi yer almaktadır.



Şekil 0.1. Deepfake sınıflandırması için önerilen çoklu akış ESA mimarisinin temsili gösterimi.

5.3. Veri Seti

Bu çalışmada deepfake tespit algoritmalarını geliştirmek için büyük ölçekli yeni bir veri seti olan Dünya Politikacılar Deepfake Veri Seti (DPDVS, World Politicians Deepfake Dataset (WPDD)) oluşturulmuştur. Ayrıca önerilen yöntem, DeepFakeTIMIT HQ ve Celeb-DF veri setleri üzerinde de değerlendirilmiştir.

5.3.1. DPDVS Veri Seti

DPDVS veri setini oluşturmak için YouTube'dan dokuzu kadın, on biri erkek olan dünyaca ünlü yirmi politikacıya ait videolar indirilmiştir. Gerçek uzunlukları 5 ile 16 dakika arasında değişen bu videolar, uzunlukları 10 ile 20 saniye arasında değişen 946 videoya bölünmüştür. Elde edilen bu videolar her bir manipülasyon sınıfı için gruplandırılmıştır. Daha sonra her bir videonun çerçeveleri çıkartılarak toplamda 320.499 çerçeve elde edilmiştir. Gruplandırılan bu çerçevelere makyaj (sadece kadınlar için), saç modeli, saç rengi, yaş, gülümseme, gözlük, sakal (sadece erkekler için) ve yüz değiştirme olmak üzere her bir cinsiyet için 7 farklı sınıf olacak şekilde çeşitli manipülasyonlar uygulanmıştır. Her bir manipülasyon sınıfı için, ilgili sınıfa ait olan farklı tipte manipülasyonlar uygulanmıştır. Örneğin, saç rengi manipülasyon sınıfı için farklı politikacılara siyah, kahverengi, sarı ve kumral gibi farklı saç renkleri uygulanmıştır. Benzer şekilde saç modeli manipülasyon sınıfı için kıvrıkcık, düz ve dalgalı gibi farklı saç modelleri kullanılmıştır. Şekil 5.2'de DPDVS veri setine ait örnek çerçeveleri göstermektedir.



Şekil 0.2. DPDVS veri setinde yer alan örnek çerçeveler. İlk satırda, farklı politikacıların gerçek görüntüleri gösterilmektedir. Takip eden satırlarda ise sırasıyla makyaj (yalnızca kadınlar için)/sakal (yalnızca erkekler için), saç modeli, saç rengi, yaş, gülüms

5.4. Deneysel Çalışmalar

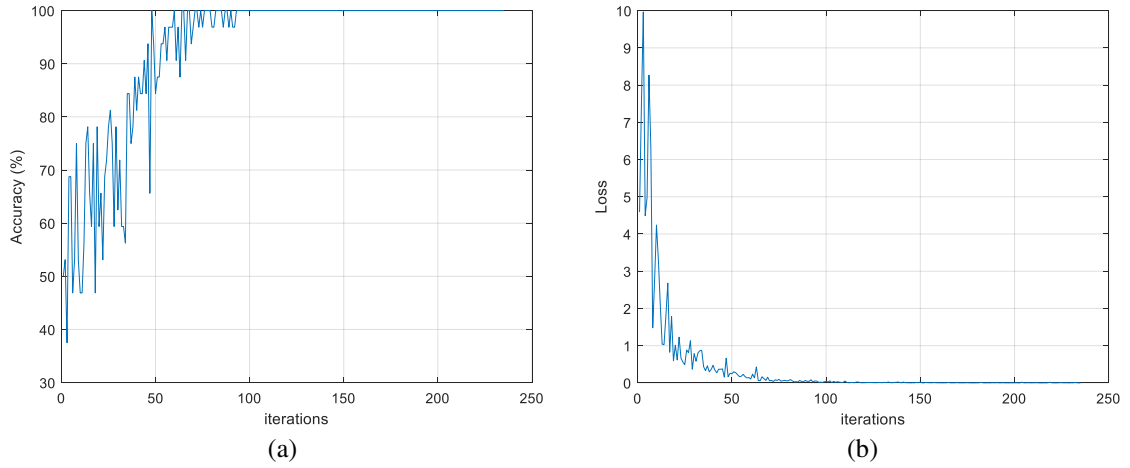
Bu bölümde, önerilen deepfake ve yüz manipülasyon tespit yönteminin deneysel değerlendirmeleri sunulmaktadır. Ayrıca, önerilen yöntemin önceki tekniklerle karşılaştırılması, yeni yüz işleme türleri altındaki performans ve mevcut veri setleri üzerindeki performansı sunulmaktadır.

Deneysel çalışmalar ön işlem aşamasında tespit edilen yüz bölgeleri çerçeveleri, önceden eğitilmiş derin modellerin girişi ile uyumlu olması için 224×224 olarak yeniden boyutlandırılmıştır. Yüz bölgesi tespiti için iyi bilinen Viola-Jones algoritması kullanılmıştır. Veri setinden rastgele seçilen görüntülerin %70'i önerilen yöntemin eğitiminde kullanılmış geriye kalan %30'u ise önerilen yöntemin test edilmesi için kullanılmıştır. Önceden eğitilmiş ESA modellerinin ince ayarında, devir sayısı 5 olarak ayarlanmıştır. Bırakma olasılığı (dropout probability) 0,2 olarak seçilmiştir. MiniBatch boyutu, Gradyan Eşiği ve başlangıç öğrenme oranı sırasıyla 32, 2 ve 0,001 olarak seçilmiştir.

Deneysel çalışmalarda üç farklı senaryo düşünülmüştür. Birinci senaryoda gerçek yüz ile sahte yüz ayrımı gerçekleştirilmiş, ikinci senaryoda ise yedi sahte yüz sınıfı ile gerçek yüz arasındaki ayrım (sekiz sınıflı sınıflandırma problemi) ele alınmıştır. Sınıflandırma performansı; doğruluk, duyarlılık ve özgüllük kriterleri kullanılarak ölçülmüştür. Üçüncü senaryoda ise yeni tip yüz manipülasyonu altında deepfake tespit sisteminin performansı değerlendirilmiştir.

5.4.1. Birinci Senaryo

Birinci senaryo için gerçek ve sahte yüz olmak üzere iki sınıflı sınıflandırma problemi üzerinde durulmuştur. Şekil 5.3 birinci senaryo için çoklu akış ESA modelinin eğitimini ve kayıp süreçlerini göstermektedir. Şekil 5.3'a da görüldüğü gibi, ilk senaryo için çoklu akış ESA, eğitim boyunca geçen sürenin 54 dakika olduğu 250 yineleme (iteration) için eğitilmiştir. ESA modelinin başlangıç performansı yaklaşık %50 iken birkaç eğitim yinelemesinden sonra %40'a düşmüştür. ESA modelinin performansı 50. yinelemede %90'a yükselmiş ve 100. yinelemeden sonra eğitim doğruluğu %100 olmuştur. Kayıp değeri eğitimin başında beş civarındayken birkaç yinelemeden sonra 10'a yükselmiştir. Kayıp değeri, 50. yinelemeden sonra 0,1 civarına düşmüş ve 100. yinelemeden sonra kayıp değeri sıfıra oldukça yaklaşmıştır. Şekil 5.4, birinci senaryo için elde edilen karmaşıklık matrisini göstermektedir. Karmaşıklık matrisinin sütunları tahmin edilen sınıfları gösterirken, satırlar gerçek sınıfları göstermektedir. Şekil 5.4'ten 120 sahte yüzün gerçek, 160 gerçek yüzün ise sahte yüz olarak sınıflandırıldığı görülebilmektedir.



Şekil 0.3. Birinci senaryonun eğitimi için eğitim doğruluğu eğrisi (a) ve eğitim kayıp eğrisi (b)

Gerçek Sınıflar	Sahte	47915	120
	Gerçek	160	47955
		Sahte	Gerçek
		Tahmin Edilen Sınıflar	

Şekil 0.4. Birinci senaryo için elde edilen karmaşıklık matrisi

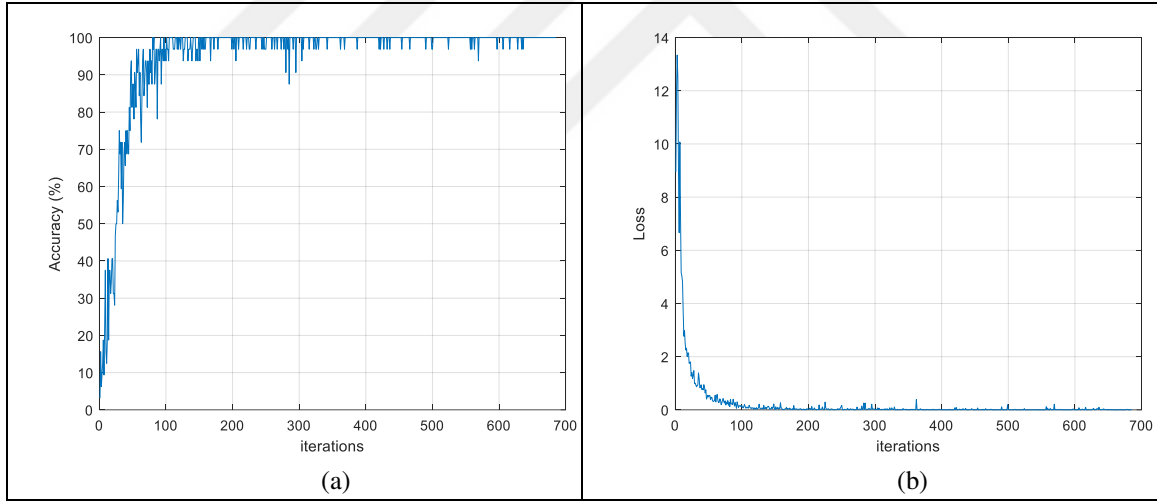
Ayrıca 47.915 sahte ve 47.955 gerçek yüz, önerilen çoklu akış ESA yaklaşımı ile doğru bir şekilde sınıflandırılmıştır. Tablo 5.1'de birinci senaryo için performans değerlendirme ölçüleri verilmiştir. Tablo 5.1'de görüldüğü gibi hesaplanan doğruluk oranı %99,71, duyarlılık oranı %99,67 ve özgüllük oranı ise %99,75 olarak elde edilmiştir.

Tablo 0.1. Birinci senaryo için hesaplanan performans ölçüleri

	Doğruluk	Duyarlılık	Özgüllük
Birinci Senaryo	%99,71	%99,67	%99,75

5.4.2. İkinci Senaryo

İkinci senaryoda yedi sahte yüz manipülasyon sınıfı ile gerçek görüntüler olmak üzere toplamda sekiz sınıflı sınıflandırma problemi ele alınmıştır. Şekil 5.5 ikinci senaryo için çoklu akış ESA modelinin eğitimini ve kayıp süreçlerini göstermektedir. Şekil 5.5a, eğitim boyunca doğruluk eğrisini, Şekil 5.5b ise eğitim boyunca kayıp eğrisini göstermektedir. Şekil 5.5a'da görüldüğü gibi, ikinci senaryo için çoklu akış ESA, eğitim boyunca geçen sürenin 81 dakika olduğu 700 yineleme için eğitilmiştir. ESA modelinin ilk başlangıç performansı %10 civarındayken 100. yinelemede %90'a yükselmiştir. Ayrıca 100. yinelemeden sonra eğitim doğruluğu yaklaşık %100 olmuştur. Eğitimin başında kayıp değeri 13 civarında iken 100. yinelemede yaklaşık sıfıra düşmüştür. 100. yinelemeden sonra kayıp değeri sıfıra oldukça yaklaşmıştır.



Şekil 0.5. İkinci Senaryonun eğitimi için eğitim doğruluğu eğrisi (a) ve eğitim kayıp eğrisi (b)

Şekil 5.6 ikinci senaryo için elde edilen karmaşıklık matrisini göstermektedir. Karmaşıklık matrisinin sütunları tahmin edilen sınıfları, satırlar ise gerçek sınıfları göstermektedir. Şekil 5.6'dan da görüleceği gibi önerilen çoklu akış ESA yaklaşımı için yaş, sakal, yüz değiştirme, gözlük, saç rengi, saç modeli, makyaj, gülümseme ve gerçek sınıfları için sırasıyla 9212, 11494, 20083, 10304, 8505, 9191, 5130, 11781 ve 11.739 yüz doğru şekilde sınıflandırılmıştır.

Gerçek Sınıflar	Yaş								
	Sakal	11494							
	Yüz Değiştirme	14	20083						
	Gözlük			10304					
	Saç Rengi				8505				
	Saç Modeli					9191			
	Makyaj						5130		
	Gülümseme	14						11781	14
	Gerçek	14				14	14	14	11739
Tahmin Edilen Sınıflar									

Şekil 0.6. İkinci senaryo için elde edilen karmaşıklık matrisi

Tablo 5.2'de ikinci senaryo için performans değerlendirme ölçüleri verilmiştir. Tablo 5.2.'de görüldüğü gibi, hesaplanan doğruluk, duyarlılık ve özgüllük oranları sırasıyla %99,82 , %99,83 ve %99,97'dir.

Tablo 0.2. İkinci senaryo için hesaplanan performans ölçüleri

	Doğruluk	Duyarlılık	Özgüllük
İkinci Senaryo	%99,82	%99,83	%99,97

5.4.3. Üçüncü Senaryo: Yeni Tip Yüz Manipülasyonu Altındaki Performans

Önerilen yöntemin farklı manipülasyon türleri arasında genelleme kapasitesini değerlendirmek için başka deneyler gerçekleştirilmiştir. Önerilen derin akış yönteminin doğruluğu, eğitim seti olarak yalnızca bir manipülasyon türü dikkate alınarak değerlendirilirken, diğer manipülasyon türleri tarafından test seti olarak kullanılmıştır. Örneğin, eğitim seti olarak sadece yaş manipülasyonu uygulanan görüntüler için test seti olarak sakal manipülasyonu uygulanan görüntüler kullanılmıştır. Tablo 5.3'te elde edilen sonuçlar rapor edilmiştir. Tablo 5.3'te görüldüğü gibi eğitimde yaş, yüz değiştirme, gözlük, saç rengi, saç modeli ve gülümseme manipülasyon

çeşitleri kullanılmış ve diğer manipülasyonlarla test edilmiştir. Eğitimde yaş manipülasyon türü kullanıldığında ve diğer manipülasyon türleri üzerinde test edildiğinde, ortalama %84,84 tespit doğruluk oranı elde edilmiştir.

Yüz değiştirme manipülasyon türü için, %86,47'lik bir ortalama tespit doğruluğu elde edilmiştir. Eğitim için yüz değiştirme manipülasyon türü kullanıldığında test için kullanılan sakal manipülasyon türü için en yüksek doğruluk oranı olan %97,81 elde edilmiştir. Gözlük manipülasyon türü için ortalama doğruluk puanı %77,01 olarak kaydedilmiştir. Gözlük manipülasyon türü için en iyi doğruluk puanı %86,30 ile yaş manipülasyon türü için elde edilmiştir. Saç rengi manipülasyon tipi için, elde edilen ortalama tespit doğruluk oranı %73,57'dir. Saç rengi manipülasyon türü eğitimde kullanıldığında saç modeli manipülasyon türü için %88,37 doğruluk oranı elde edilmiştir. Benzer şekilde saç modeli manipülasyon türü için en yüksek doğruluk oranı %83,26 saç rengi manipülasyon türü için elde edilmiştir. Saç modeli manipülasyon türü için ortalama test doğruluk oranı %75.37 olarak elde edilmiştir. Son olarak, gülümseme manipülasyon türü eğitim için kullanıldığında diğer manipülasyon sınıfları üzerinde test için kullanıldığında %77,12 ortalama doğruluk oranı elde edilmiştir. Tablo 5.3'teki sonuçlar, önerilen yöntemin, topluluk öğrenmesinden ilham alan basit bir yöntem olsa bile, dikkate değer bir genelleme yeteneği elde etme kapasitesine sahip olduğunu göstermektedir. Önerilen yöntem, eğitimde bulunmayan manipülasyon türleri için çeşitli ve daha ayırt edici öznitelikleri çıkartabilir. Ayrıca, önerilen yöntem, manipüle edilmiş video örnekleri üzerinde, manipüle edilmiş görüntü örnekleri için [124]'de benimsenen derin modellerden daha yüksek genelleme doğruluk oranları elde edebilir.

Tablo 0.3. Önerilen derin akış yönteminin eğitim seti olarak yalnızca bir manipülasyon türü dikkate alındığında diğer manipülasyon türleri test seti olarak kullanıldığında elde edilen doğruluk oranları

Üzerinde Eğitilen Manipülasyon Türü	Üzerinde Test Edilen Manipülasyon Türü	Performans (%)
Yaş	Sakal	88,20
	Yüz Değiştirme	86,25
	Gözlük	89,60
	Saç Rengi	76,21
	Saç Modeli	75,23
	Makyaj	86,30
	Gülümseme	92,10
	Ortalama	84,84
Yüz Değiştirme	Yaş	95,26
	Sakal	97,81
	Gözlük	94,30
	Saç Rengi	75,46
	Saç Modeli	77,28
	Makyaj	69,80
	Gülümseme	95,40
	Ortalama	86,47
Gözlük	Yaş	86,30
	Sakal	82,14
	Yüz Değiştirme	80,63
	Saç Rengi	75,26
	Saç Modeli	72,44
	Makyaj	71,50
	Gülümseme	70,78
	Ortalama	77,01
Saç Rengi	Yaş	84,65
	Sakal	66,36
	Yüz Değiştirme	69,82
	Gözlük	75,35
	Saç Modeli	88,37
	Makyaj	67,20
	Gülümseme	63,27
	Ortalama	73,57
Saç Modeli	Yaş	82,36
	Sakal	73,26
	Yüz Değiştirme	78,20
	Gözlük	75,60
	Saç Rengi	83,26
	Makyaj	65,46
	Gülümseme	69,45
	Ortalama	75,37
Gülümseme	Yaş	77,85
	Sakal	79,80
	Yüz Değiştirme	82,99
	Gözlük	73,21
	Saç Rengi	68,47
	Saç Modeli	75,80
	Makyaj	81,70
	Ortalama	77,12

5.4.4. Önerilen Yöntemi Mevcut Deepfake Tespit Yöntemleri ile Karşılaştırma

Önerilen yöntem, mevcut yöntemlerden bazıları ile karşılaştırılmıştır. Karşılaştırmalar Tablo 5.4'te görülebilir. Örneğin, Yang vd. [15], her videodan tahmini 3D kafa yönelimleri üzerinde bir SVM yapısına dayalı olarak sentezlenen videoların baş pozlarındaki tutarsızlıkları kullanarak deepfake videoları tespit etmek için yeni bir yöntem önermişlerdir. UADFV veri seti, bu yöntemde SVM modelini eğitmek için kullanılmıştır. Videolara ve çerçevelere dayanan alıcı işletim karakteristik eğrisi (areas under receiver operating characteristic curve, AUROC) altındaki en iyi alanlar sırasıyla %0.974 ve %0.89 olarak elde edilmiştir. Yu vd. [125] tarafından bir girdi görüntüsünü ilgili parmak izi (finger print) görüntüsüyle eşleştirmek için bir ağ yapısı sunulmuştur. Sonuç olarak, her kaynak (her ÜÇA örneği ve gerçek görüntüler) için bir model parmak izi bularak bir resim parmak izi ile her model parmak izi arasındaki korelasyon indeksi; sınıflandırmak için ise softmax logit kullanılmıştır. Önerilen yöntem, bazı ÜÇA teknikleri (ProGAN [125], SNGAN [126], CramerGAN [127] ve MMDGAN [128]) kullanılarak elde edilen sentetik yüzlerin yanı sıra CelebA veri setindeki gerçek yüzler üzerinde denenmiştir.

Tablo 5.4'te görüldüğü gibi, önerilen yöntem doğruluk açısından diğer yöntemlerden daha yüksek performans göstermiştir. Yang vd. [15], özellikle merkezi yüz bölgesinin dikkate alındığı deepfake tespit yöntemi için 3D kafa pozlarını kullanmıştır. [15]'teki çalışma, yüz değiştirme tespiti için iyi sonuçlar elde etmesine rağmen, diğer manipülasyon türleri dikkate alınmadığı için bu çalışmada önerilen yöntemden daha iyi sonuçlar elde edememiştir. Afchar vd. [12] yüz manipülasyon tespiti için iki derin ağ önermiştir. Araştırmacılar, özellikle sadece iki sınıf sınıflandırması ile az sayıda katmana sahip görüntülerin mezoskopik özelliklerine odaklanmış, ancak çok sınıflı deepfake tespiti dikkate alınmamıştır. Li ve Lyu [52] deepfake tespiti için derin bir ESA modeli sunmuşlardır. Ancak yazarlar, metodolojilerini geliştirmek için yalnızca videonun çözünürlüğüne odaklanmışlardır. Herhangi bir zorlayıcı veri seti kullanmamışlardır. Yu vd. [125], manipüle edilmiş yüz örneklerini tespit etmek için ÜÇA'nın parmak izlerini kullanmışlardır. Ancak, [125] 'deki çalışmada, çok sınıflı yüz manipülasyon tespiti incelenmemiştir. Performansla ilgili olarak, Yu vd. [125], daha önce sunmuş oldukları yöntemi DPDVS veri seti üzerinde eğitim test ederek %99,63 doğruluk puanını elde etmiş ve en iyi ikinci doğruluk oranına ulaşmıştır. Ayrıca [12]'deki çalışma için %99,21 doğruluk puanı elde edilmiştir. Li vd. [52] ve Yang vd. [15] önerdiği yöntemler ile DPDVS veri seti üzerinde sırasıyla %99,36 ve %98,62 doğruluk oranlarını elde etmişlerdir.

Tablo 0.4. Önerilen yöntemin mevcut yöntemler ile DPDVS veri seti kullanılarak karşılaştırıldıktan sonra elde edilen doğruluk oranları

Yöntem	Doğruluk
Yang vd. [15]	%98,62
Afchar vd. [12]	%99,21
Li ve Lyu [52]	%99,36
Yu vd. [125]	%99,63
Önerilen Yöntem	%99,80

Bu çalışmada, hem ikili hem de çok sınıflı deepfake tespiti için çoklu akış yaklaşımı önerilmiştir. Önerilen yöntemin doğruluğunu artırmak için çoklu akış yapısına üç ESA modeli dahil edilmiştir. Elde edilen sonuçlar, önerilen yapının elde edilen doğruluğu iyileştirmek için yeterince verimli olduğunu göstermiştir. Doğruluğun bu şekilde iyileştirilmesi, milyonlarca örneğin orijinallik kontrolüne tabi tutulduğu gerçek dünya uygulamalarında oldukça önemlidir. Önerilen yöntemin daha yüksek doğruluk elde etmesinin ana nedenlerinden biri, tamamlayıcı bilgi veya özellikler kullanan topluluk öğrenme mekanizmasına dayanmasıdır. Çeşitli temel özellikleri veya algoritmaları birleştiren topluluk öğrenme sınıflandırma mekanizmaları, özellikle tek tür özellik veya algoritma kullanan geleneksel yöntemlere kıyasla daha iyi performans göstermiştir [129].

5.4.5. Önerilen Yöntemin Mevcut Veri Setleri Üzerindeki Performansı

Önerilen yöntem ayrıca mevcut veri setlerinden bazılarında (DeepFakeTIMIT HQ ve Celeb-DF) analiz edilmiştir. Doğruluk oranları açısından performanslar Tablo 5.5'de rapor edilmiştir. Çiftçi vd. [50] deepfake tespiti için FakeCatcher'ı sunmuşlardır. FakeCatcher, sahte videoları gerçek videolardan ayırt etmek için videolarda gizlenen biyolojik sinyallerin farklılığından yararlanmaktadır. Araştırmacılar, önerilen yöntemi Face Forensics, Face Forensics++, CelebDF ve yeni bir deepfake veri setinde uygulamış ve sırasıyla %96,0, %94.65, %91.50 ve %91.07 doğruluk oranları elde etmişlerdir. Jafar vd. [130], deepfake videolarını tespit etmek için ağız biyolojik bir sinyal olarak kullanan ağız özelliklerine sahip bir deepfake tespit modeli tanıtmışlardır. Önerilen yöntem DeepfakeTIMIT LQ, DeepfakeTIMIT HQ ve Celeb-DF veri setlerinde değerlendirilerek sırasıyla %98,7, %73,1 ve %71,25 doğruluk oranına ulaşmıştır. Bu çalışmada önerilen yöntem, DeepFakeTIMIT HQ ve Celeb-DF veri setleri için sırasıyla %99,9 ve %98,2 doğruluk değerleri elde etmiştir. Bu çalışmada elde edilen doğruluk oranları DeepFakeTIMIT HQ ve Celeb-DF veri setleri için [50] ve [130]'daki mevcut son teknoloji deepfake tespit yöntemlerinden daha iyi performans göstermiştir.

Tablo 5.5'de görüldüğü gibi önerilen yöntem DeepFakeTIMIT HQ ve Celeb-DF veri setleri için sırasıyla %99,98 ve %99,95 doğruluk oranları elde etmiştir. Önerilen yöntem çoklu akış bir ESA yapısı üzerinde giriş yüz görüntülerinin orijinal, bulanık ve keskinleştirilmiş versiyonlarını kullandığından, elde edilen sonuçlar, karşılaştırılan yöntemlere göre biraz daha iyileştirilmiştir.

Veri setinde mevcut olan manipölasyon türleri aynı zamanda deepfake tespit yönteminin doğruluğunu da etkilemektedir. Örneğin; DeepFakeTIMIT HQ, manipüle edilmiş örnekleri oluşturmak için Celeb-DF veri setinden nispeten daha düşük kalite ve daha yüksek yapaylıklara sahip olan basit ÜÇA tabanlı bir algoritmayı kullanmıştır. Bu nedenle önerilen yöntem DeepFakeTIMIT HQ veri seti için daha iyi bir doğruluk performansı sergilemiştir.

Tablo 0.5. Çeşitli veri setleri kullanılarak önerilen derin akış yönteminin doğruluk oranları

Veri Seti	Doğruluk Oranı
DeepFake-TIMIT HQ	%99,98
Celeb-DF	%99,95

6. SONUÇLAR

Bu tez çalışmasında görüntü manipölasyonlarını tespit yöntemlerini geliştirmek için derin öğrenme tabanlı yöntemlerden olan ESA'nın farklı mimarileri kullanılarak sınıflandırma problemi çözülmeye çalışılmıştır. Görüntü manipölasyonları türlerinde yer alan ve son yıllarda popüler hale gelen yüz manipölasyonu (deepfake) tespiti üzerinde özellikle durulmuştur. Tüm deneysel analizler, Intel(R) Xeon(R) CPU E5-1650 @3.60 GHz 64 GB bellek ve NVIDIA Quadro M4000 GPU içeren bir bilgisayarda Matlab programı kullanarak gerçekleştirilmiştir.

Bu kapsamda yapılan ilk çalışmada, sahte ve gerçek yüz videolarının ayrımı için yetkin bir mimari geliştirilmiştir. Önerilen yaklaşım, ESA ile UKSB mimarilerini birleştirerek derin ESA özelliklerinin ayırt edici güçlerini kullanan hibrit bir yönteme dayanmaktadır. Önerilen yöntemde öznetelik çıkarmak için önceden eğitilmiş ResNet-50 modeli ve sınıflandırma için UKSB kullanılmıştır. Deneysel analizde iki farklı halka açık deepfake video veri seti kullanılmıştır. Derin özelliklerin ve UKSB sınıflandırıcının, sahte yüz videolarını gerçek olanlardan ayırt etmede büyük potansiyele sahip olduğu gösterilmiştir. Önerilen deepfake tespit yöntemi, mevcut tekniklerden daha iyi performans göstermiştir. Derin özellikler hem renk hem de doku kullandığından önceki yöntemlerden oldukça iyi performans göstermiştir.

Yapılan ikinci çalışmada, belirli bir videoda sahte yüz tespiti için yeni bir yaklaşım önerilmiştir. Önerilen yaklaşımda, yeni bir çoklu akış derin ESA mimarisi geliştirilmiştir. Çoklu akış ESA mimarisinde, önceden eğitilmiş üç ESA modeli, bir füzyon katmanı kullanılarak birleştirilmiştir. Ayrıca bu çalışma için yeni bir deepfake veri seti olan 320.499 adet gerçek ve 320.499 adet sahte görüntü içeren DPDVS veri seti oluşturulmuştur. Önceden eğitilmiş ESA modellerinin girdilerine veri setinde yer alan görüntülerin sırasıyla orijinal, bulanık ve keskinleştirilmiş görüntüleri verilmiştir. Deneysel çalışmalarda üç adet sahte yüz tespit senaryosu ele alınmıştır. İlk olarak sadece sahte ve gerçek yüz ayrımı ele alınmıştır. İkinci olarak, erkek ve kadınlar için toplamda sekiz farklı yüz manipölasyonu (yaş, sakal (sadece erkeler için), makyaj (sadece kadınlar için), yüz değiştirme, gözlük, saç rengi, saç modeli, gülümseme ve gerçek yüz ayrımı) ayrımı gerçekleştirilmiştir. Üçüncü senaryoda ise yeni tip yüz manipölasyonu altında deepfake tespit sisteminin performansı değerlendirilmiştir. Ayrıca önerilen yöntem farklı mevcut veri setleri üzerinde de test edilmiştir. Önerilen yöntem mevcut yöntemlere göre daha iyi bir doğruluk performansı sergilemiştir. Ek olarak, farklı mevcut yöntemler önerilen yöntem ile birlikte DPDVS veri seti üzerinde denenmiştir. Deneysel sonuçlar, önerilen yöntemin mevcut yöntemler arasında en iyi doğruluk performansını gösterdiğini ortaya koymuştur.

ÖNERİLER

Tez çalışmasında önerilen yöntemlerde, derin öğrenmenin alt kollarından olan ESA'yı temel alan yöntemler kullanılmıştır. Önerilen yöntemler ile mevcut görüntü manipülasyonu tespit algoritmaları kullanan yöntemlere göre daha iyi bir performans elde edilmiştir. Bununla birlikte; günden güne artan geliştirilmiş tespit mekanizmalarına rağmen, önceki mekanizmaların çoğu, manipüle edilmiş yüzleri tespit etme yetenekleriyle sınırlıdır. Yani, eğitimin bir parçası olmayan farklı manipülasyonlar türleri ile karşılaştıklarında, mevcut deepfake tespit yöntemlerinin performansı önemli ölçüde düşmektedir. Sonuç olarak, kullanılan tespit algoritmaları düşük genelleme yeteneklerine sahiptirler. Eğitim aşamasında bulunmayan yeni manipülasyonlar, araçlar ve veri setleri için daha yüksek genelleştirme yeteneklerine ve daha düşük hata oranlarına sahip deepfake tespit mimarilerine büyük bir ihtiyaç vardır. Araştırmalar, daha çok yeni geliştirilmiş deepfake tespit yöntemlerini geliştirmeye yoğunlaşabilir.

Manipüle edilmiş yüz örnekleri kullanıldığında yüz tanıma sistemlerinin doğruluk performansının düştüğü gösterilmiştir. Ayrıca literatürde deepfake konusunu ele almaya çalışan çok sınırlı sayıda çalışma olduğu görülmektedir. Manipülasyon tespit tabanlı sistemler ve güvenlik geliştirme tabanlı sistemleri birlikte tasarlamaya yönelik çalışmalar yapılabilir.

Çok az sayıda büyük ölçekli deepfake veri seti herkesin kullanımına açıktır. Derin öğrenme yapılarının eğitimlerinin daha güvenilir ve yüksek doğruluk performansı göstermesi için çeşitli türde manipülasyonlara sahip büyük ölçekli veri setlerine ihtiyaç vardır. Ayrıca, veri setleri oluşturmak için kullanılacak yüksek dereceli sentetik yüz oluşturma teknikleri zorlayıcı bir sorundur. Bu tür zorluklar deepfake tespiti alanındaki ilerlemeyi engellemektedir. Bunun için tüm manipülasyon türlerini içeren büyük ölçekli veri setleri oluşturulabilir.

KAYNAKLAR

- [1] A. Rossler, D. Cozzolino, L. Verdoliva, C. Riess, J. Thies, M. Niessner, FaceForensics++: Learning to detect manipulated facial images, Proc. IEEE Int. Conf. Comput. Vis. 2019-Octob (2019) 1–11. <https://doi.org/10.1109/ICCV.2019.00009>.
- [2] Z. Akhtar, D. Dasgupta, B. Banerjee, Face Authenticity: An Overview of Face Manipulation Generation, Detection and Recognition, SSRN Electron. J. (2019). <https://doi.org/10.2139/ssrn.3419272>.
- [3] R. Tolosana, R. Vera-Rodriguez, J. Fierrez, A. Morales, J. Ortega-Garcia, Deepfakes and beyond: A Survey of face manipulation and fake detection, Inf. Fusion. 64 (2020) 131–148. <https://doi.org/10.1016/j.inffus.2020.06.014>.
- [4] I.J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial nets, Adv. Neural Inf. Process. Syst. 3 (2014) 2672–2680. https://doi.org/10.3156/jsoft.29.5_177_2.
- [5] Z. Akhtar, A. Rattani, A. Hadid, M. Tistarelli, Face recognition under ageing effect: A comparative analysis, Lect. Notes Comput. Sci. (Including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics). 8157 LNCS (2013) 308–318. https://doi.org/10.1007/978-3-642-41184-7_32.
- [6] Z. Akhtar, G. Fumera, G.L. Marcialis, F. Roli, Robustness evaluation of biometric systems under spoof attacks, Lect. Notes Comput. Sci. (Including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics). 6978 LNCS (2011) 159–168. https://doi.org/10.1007/978-3-642-24085-0_17.
- [7] Z. Akhtar, G. Fumera, G.L. Marcialis, F. Roli, Robustness analysis of likelihood ratio score fusion rule for multimodal biometric systems under spoof attacks, Proc. - Int. Carnahan Conf. Secur. Technol. (2011). <https://doi.org/10.1109/CCST.2011.6095935>.
- [8] Z. Akhtar, A. Rattani, A Face in any Form: New Challenges and Opportunities for Face Recognition Technology, Computer (Long. Beach. Calif). 50 (2017) 80–90. <https://doi.org/10.1109/MC.2017.119>.
- [9] Z. Akhtar, A. Hadid, M.S. Nixon, M. Tistarelli, J.L. Dugelay, S. Marcel, Biometrics: In search of identity and security (Q A), IEEE Multimed. 25 (2018) 22–35. <https://doi.org/10.1109/MMUL.2018.2873494>.
- [10] P. Korshunov, S. Marcel, DeepFakes: a New Threat to Face Recognition? Assessment and Detection, (2018). <http://arxiv.org/abs/1812.08685>.
- [11] D. Guera, E.J. Delp, Deepfake Video Detection Using Recurrent Neural Networks, Proc. AVSS 2018 - 2018 15th IEEE Int. Conf. Adv. Video Signal-Based Surveill. (2019). <https://doi.org/10.1109/AVSS.2018.8639163>.
- [12] D. Afchar, V. Nozick, J. Yamagishi, I. Echizen, MesoNet: A compact facial video forgery detection network, 10th IEEE Int. Work. Inf. Forensics Secur. WIFS 2018. (2019). <https://doi.org/10.1109/WIFS.2018.8630761>.
- [13] P. Zhou, X. Han, V.I. Morariu, L.S. Davis, Two-Stream Neural Networks for Tampered Face Detection, IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Work. 2017-July (2017) 1831–1839. <https://doi.org/10.1109/CVPRW.2017.229>.

- [14] Y. Li, X. Yang, P. Sun, H. Qi, S. Lyu, Celeb-DF: A New Dataset for DeepFake Forensics, Arxiv. 1 (2019) 1–6. <http://arxiv.org/abs/1909.12962>.
- [15] X. Yang, Y. Li, S. Lyu, Exposing Deep Fakes Using Inconsistent Head Poses, ICASSP, IEEE Int. Conf. Acoust. Speech Signal Process. - Proc. 2019-May (2019) 8261–8265. <https://doi.org/10.1109/ICASSP.2019.8683164>.
- [16] Z. Liu, P. Luo, X. Wang, X. Tang, Deep learning face attributes in the wild, Proc. IEEE Int. Conf. Comput. Vis. 2015 Inter (2015) 3730–3738. <https://doi.org/10.1109/ICCV.2015.425>.
- [17] Deep Fake Detection Dataset, <https://ai.googleblog.com/2019/09/contributing-data-to-deepfake-detection.html>. Eriřim: 29 Ekim 2019., (n.d.).
- [18] B. Dolhansky, J. Bitton, B. Pflaum, J. Lu, R. Howes, M. Wang, C.C. Ferrer, The DeepFake Detection Challenge (DFDC) Dataset, (2020). <http://arxiv.org/abs/2006.07397>.
- [19] S. Lyu, Deepfake detection: Current challenges and next steps, 2020 IEEE Int. Conf. Multimed. Expo Work. ICMEW 2020. (2020). <https://doi.org/10.1109/ICMEW46912.2020.9105991>.
- [20] T. Neubert, C. Kraetzer, J. Dittmann, A face morphing detection concept with a frequency and a spatial domain feature space for images on eMRTD, IH MMSec 2019 - Proc. ACM Work. Inf. Hiding Multimed. Secur. (2019) 95–100. <https://doi.org/10.1145/3335203.3335721>.
- [21] R. Raghavendra, K.B. Raja, S. Venkatesh, C. Busch, Transferable Deep-CNN Features for Detecting Digital and Print-Scanned Morphed Face Images, IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Work. 2017-July (2017) 1822–1830. <https://doi.org/10.1109/CVPRW.2017.228>.
- [22] A. Rössler, D. Cozzolino, L. Verdoliva, C. Riess, J. Thies, M. Nießner, FaceForensics: A Large-scale Video Dataset for Forgery Detection in Human Faces, (2018). <http://arxiv.org/abs/1803.09179>.
- [23] H. Dang, F. Liu, J. Stehouwer, X. Liu, A.K. Jain, On the Detection of Digital Face Manipulation, Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (2020) 5780–5789. <https://doi.org/10.1109/CVPR42600.2020.00582>.
- [24] T. Karras, S. Laine, T. Aila, A style-based generator architecture for generative adversarial networks, Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. 2019-June (2019) 4396–4405. <https://doi.org/10.1109/CVPR.2019.00453>.
- [25] T. Karras, T. Aila, S. Laine, J. Lehtinen, Progressive growing of GANs for improved quality, stability, and variation, 6th Int. Conf. Learn. Represent. ICLR 2018 - Conf. Track Proc. (2018).
- [26] J.Y. Zhu, T. Park, P. Isola, A.A. Efros, Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks, Proc. IEEE Int. Conf. Comput. Vis. 2017-Octob (2017) 2242–2251. <https://doi.org/10.1109/ICCV.2017.244>.
- [27] Y. Shen, P. Luo, J. Yan, X. Wang, X. Tang, FaceID-GAN: Learning a Symmetry Three-Player GAN for Identity-Preserving Face Synthesis, Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (2018) 821–830. <https://doi.org/10.1109/CVPR.2018.00092>.
- [28] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition, Proc. IEEE. 86 (1998) 2278–2323. <https://doi.org/10.1109/5.726791>.
- [29] A. Krizhevsky, Learning Multiple Layers of Features from Tiny Images, ... Sci. Dep. Univ. Toronto, Tech. (2009) 1–60. <https://doi.org/10.1.1.222.9220>.

- [30] E. Denton, S. Chintala, A. Szlam, R. Fergus, Deep generative image models using a laplacian pyramid of adversarial networks, *Adv. Neural Inf. Process. Syst.* 2015-Janua (2015) 1486–1494.
- [31] A. Creswell, T. White, V. Dumoulin, K. Arulkumaran, B. Sengupta, A.A. Bharath, Generative Adversarial Networks: An Overview, *IEEE Signal Process. Mag.* 35 (2018) 53–65. <https://doi.org/10.1109/MSP.2017.2765202>, (n.d.).
- [32] X. Chen, X. Chen, Y. Duan, R. Houthoof, J. Schulman, I. Sutskever, P. Abbeel, InfoGAN: Interpretable Representation Learning by Information Maximizing Generative Adversarial Nets, *Adv. Neural Inf. Process. Syst.* (2016) 2172–2180. <http://arxiv.org/abs/1606.03657> <http://papers.nips.cc/paper/6399-infoGAN-interpretable-representation-learning-by-information-maximizing-generative-adversarial-nets.pdf>.
- [33] 100K-Faces, <https://generated.photos/>, (n.d.).
- [34] <https://thispersondoesnotexist.com>, (n.d.).
- [35] ffhq-dataset, <https://github.com/NVLabs/ffhq-dataset>, Eriřim: 2 řubat 2019, (n.d.).
- [36] Faceswap, <https://github.com/deepfakes/faceswap>, Eriřim: 12 Kasım 2018, (n.d.).
- [37] Deepfake Detection Challenge Dataset, <https://ai.facebook.com/datasets/dfdc/>, Eriřim: 25 Haziran 2020, (n.d.).
- [38] Z. He, W. Zuo, M. Kan, S. Shan, X. Chen, AttGAN: Facial attribute editing by only changing what you want, *ArXiv*. (2017).
- [39] F.-J. Chang, X. Yu, R. Nevatia, M. Chandraker, Pose-variant 3D Facial Attribute Generation, (2019). <http://arxiv.org/abs/1907.10202>.
- [40] G. Perarnau, J. van de Weijer, B. Raducanu, J.M. Álvarez, Invertible Conditional GANs for image editing, (2016). <http://arxiv.org/abs/1611.06355>.
- [41] N. Zhang, M. Paluri, M. Ranzato, T. Darrell, L. Bourdev, PANDA: Pose aligned networks for deep attribute modeling, *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.* (2014) 1637–1644. <https://doi.org/10.1109/CVPR.2014.212>.
- [42] J. Thies, M. Zollhofer, M. Stamminger, C. Theobalt, M. Niebner, Face2Face: Real-Time Face Capture and Reenactment of RGB Videos, *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.* 2016-Decem (2016) 2387–2395. <https://doi.org/10.1109/CVPR.2016.262>.
- [43] J. Thies, M. Zollhöfer, M. Nießner, Deferred neural rendering: Image Synthesis using Neural Textures, *ACM Trans. Graph.* 38 (2019). <https://doi.org/10.1145/3306346.3323035>.
- [44] P. Isola, J.Y. Zhu, T. Zhou, A.A. Efros, Image-to-image translation with conditional adversarial networks, *Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017.* 2017-Janua (2017) 5967–5976. <https://doi.org/10.1109/CVPR.2017.632>.
- [45] Y. Choi, M. Choi, M. Kim, J.W. Ha, S. Kim, J. Choo, StarGAN: Unified Generative Adversarial Networks for Multi-domain Image-to-Image Translation, *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.* (2018) 8789–8797. <https://doi.org/10.1109/CVPR.2018.00916>.
- [46] Z. Akhtar, D. Dasgupta, A Comparative Evaluation of Local Feature Descriptors for DeepFakes Detection, 2019 IEEE Int. Symp. Technol. Homel. Secur. HST 2019. (2019). <https://doi.org/10.1109/HST47167.2019.9033005>.

- [47] U. Scherhag, C. Rathgeb, C. Busch, Performance variation of morphed face image detection algorithms across different datasets, IWBF 2018 - Proc. 2018 6th Int. Work. Biometrics Forensics. (2018) 1–6. <https://doi.org/10.1109/IWBF.2018.8401562>.
- [48] Z. Fan, Sun; Zhang, Niuniu; Xu, Pan; Song, Deepfake Detection Method Based on Cross-Domain Fusion, Secur. Commun. Networks. 2021 (2021) 1–11. <https://doi.org/10.1155/2021/2482942>.
- [49] L.B. Zhang, F. Peng, M. Long, Face Morphing Detection Using Fourier Spectrum of Sensor Pattern Noise, Proc. - IEEE Int. Conf. Multimed. Expo. 2018-July (2018). <https://doi.org/10.1109/ICME.2018.8486607>.
- [50] U.A. Ciftci, I. Demir, L. Yin, FakeCatcher: Detection of Synthetic Portrait Videos using Biological Signals, IEEE Trans. Pattern Anal. Mach. Intell. (2020) 1–1. <https://doi.org/10.1109/tpami.2020.3009287>.
- [51] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, Going deeper with convolutions, Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. 07-12-June (2015) 1–9. <https://doi.org/10.1109/CVPR.2015.7298594>.
- [52] Y. Li, S. Lyu, Exposing DeepFake Videos By Detecting Face Warping Artifacts, (2018). <http://arxiv.org/abs/1811.00656>.
- [53] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. 2016-Decem (2016) 770–778. <https://doi.org/10.1109/CVPR.2016.90>.
- [54] S. McCloskey, M. Albright, Detecting GAN-generated Imagery using Color Cues, ArXiv. (2018).
- [55] H. Guan, M. Kozak, E. Robertson, Y. Lee, A.N. Yates, A. Delgado, D. Zhou, T. Kheyrkhan, J. Smith, J. Fiscus, MFC datasets: Large-scale benchmark datasets for media forensic challenge evaluation, Proc. - 2019 IEEE Winter Conf. Appl. Comput. Vis. Work. WACVW 2019. (2019) 63–72. <https://doi.org/10.1109/WACVW.2019.00018>.
- [56] L. Nataraj, T.M. Mohammed, B.S. Manjunath, S. Chandrasekaran, A. Flenner, J.H. Bappy, A.K. Roy-Chowdhury, Detecting GAN generated Fake Images using Co-occurrence Matrices, IS T Int. Symp. Electron. Imaging Sci. Technol. 2019 (2019). <https://doi.org/10.2352/ISSN.2470-1173.2019.5.MWSF-532>.
- [57] J.C. Neves, R. Tolosana, R. Vera-Rodriguez, V. Lopes, H. Proença, Real or Fake? Spoofing State-Of-The-Art Face Synthesis Detection Systems, ArXiv. 13 (2019) 1–8. <http://arxiv.org/abs/1911.05351>.
- [58] S. Tariq, S. Lee, H. Kim, Y. Shin, S.S. Woo, Detecting both machine and human created fake face images in the wild, Proc. ACM Conf. Comput. Commun. Secur. (2018) 81–87. <https://doi.org/10.1145/3267357.3267367>.
- [59] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, 3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc. (2015). <http://www.robots.ox.ac.uk/>.
- [60] F. Chollet, Xception: Deep learning with depthwise separable convolutions, Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017. 2017-Janua (2017) 1800–1807. <https://doi.org/10.1109/CVPR.2017.195>.

- [61] A. Bharati, R. Singh, M. Vatsa, K.W. Bowyer, Detecting Facial Retouching Using Supervised Deep Learning, *IEEE Trans. Inf. Forensics Secur.* 11 (2016) 1903–1913. <https://doi.org/10.1109/TIFS.2016.2561898>.
- [62] H.H. Nguyen, J. Yamagishi, I. Echizen, Use of a Capsule Network to Detect Fake Images and Videos, (2019). <http://arxiv.org/abs/1910.12467>.
- [63] G.E. Hinton, S. Sabour, N. Frosst, Dynamic routing between capsules, *Adv. Neural Inf. Process. Syst.* 2017-Decem (2017) 3857–3867.
- [64] F. Matern, C. Riess, M. Stamminger, Exploiting visual artifacts to expose deepfakes and face manipulations, *Proc. - 2019 IEEE Winter Conf. Appl. Comput. Vis. Work. WACVW 2019.* (2019) 83–92. <https://doi.org/10.1109/WACVW.2019.00020>.
- [65] E. Sabir, J. Cheng, A. Jaiswal, W. AbdAlmageed, I. Masi, P. Natarajan, Recurrent Convolutional Strategies for Face Manipulation Detection in Videos, (2019). <http://arxiv.org/abs/1905.00582>.
- [66] S. Agarwal, H. Farid, Y. Gu, M. He, K. Nagano, H. Li, Protecting world leaders against deep fakes, *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Work.* 2019-June (2019) 38–45.
- [67] T. Baltrusaitis, A. Zadeh, Y.C. Lim, L.P. Morency, OpenFace 2.0: Facial behavior analysis toolkit, *Proc. - 13th IEEE Int. Conf. Autom. Face Gesture Recognition, FG 2018.* (2018) 59–66. <https://doi.org/10.1109/FG.2018.00019>.
- [68] H.H. Nguyen, F. Fang, J. Yamagishi, I. Echizen, Multi-task Learning for Detecting and Segmenting Manipulated Facial Images and Videos, 2019 IEEE 10th Int. Conf. Biometrics Theory, Appl. Syst. BTAS 2019. (2019). <https://doi.org/10.1109/BTAS46853.2019.9185974>.
- [69] I. Amerini, L. Galteri, R. Caldelli, A. Del Bimbo, Deepfake video detection through optical flow based CNN, *Proc. - 2019 Int. Conf. Comput. Vis. Work. ICCVW 2019.* (2019) 1205–1207. <https://doi.org/10.1109/ICCVW.2019.00152>.
- [70] S.S. Beauchemin, J.L. Barron, The Computation of Optical Flow, *ACM Comput. Surv.* 27 (1995) 433–466. <https://doi.org/10.1145/212094.212141>.
- [71] T. Jung, S. Kim, K. Kim, DeepVision: Deepfakes Detection Using Human Eye Blinking Pattern, *IEEE Access.* 8 (2020) 83144–83154. <https://doi.org/10.1109/ACCESS.2020.2988660>.
- [72] J. Liu, K. Zhu, W. Lu, X. Luo, X. Zhao, A lightweight 3D convolutional neural network for deepfake detection, *Int. J. Intell. Syst.* 36 (2021) 4990–5004. <https://doi.org/10.1002/int.22499>.
- [73] D.C. Stanciu, B. Ionescu, Deepfake Video Detection with Facial Features and Long-Short Term Memory Deep Networks, *ISSCS 2021 - Int. Symp. Signals, Circuits Syst.* (2021). <https://doi.org/10.1109/ISSCS52333.2021.9497385>.
- [74] P. Ranjan, S. Patil, F. Kazi, Improved generalizability of deep-fakes detection using transfer learning based CNN framework, *Proc. - 3rd Int. Conf. Inf. Comput. Technol. ICICT 2020.* (2020) 86–90. <https://doi.org/10.1109/ICICT50521.2020.00021>.
- [75] P. Chen, J. Liu, T. Liang, G. Zhou, H. Gao, J. Dai, J. Han, FSSPOTTER: Spotting face-swapped video by spatial and temporal clues, *Proc. - IEEE Int. Conf. Multimed. Expo. 2020-July* (2020). <https://doi.org/10.1109/ICME46284.2020.9102914>.

- [76] M. Luo, Y. Xiao, Y. Zhou, Multi-scale face detection based on convolutional neural network, Proc. 2018 Chinese Autom. Congr. CAC 2018. (2019) 1752–1757. <https://doi.org/10.1109/CAC.2018.8623411>.
- [77] X. Zhu, D. Ramanan, Face detection, pose estimation, and landmark localization in the wild, Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (2012) 2879–2886. <https://doi.org/10.1109/CVPR.2012.6248014>.
- [78] V. Jain, E. Learned-Miller, Fddb: A benchmark for face detection in unconstrained settings, UMass Amherst Tech. Rep. (2010). http://works.bepress.com/erik_learned_miller/55/.
- [79] X. Tao, H. Gao, R. Liao, J. Wang, J. Jia, Detail-Revealing Deep Video Super-Resolution, Proc. IEEE Int. Conf. Comput. Vis. 2017-Octob (2017) 4482–4490. <https://doi.org/10.1109/ICCV.2017.479>.
- [80] Y. Jiangang, B. Bhanu, Super-resolution of facial images in video with expression changes, Proc. - IEEE 5th Int. Conf. Adv. Video Signal Based Surveillance, AVSS 2008. (2008) 184–191. <https://doi.org/10.1109/AVSS.2008.72>.
- [81] W.S. McCulloch, W. Pitts, A logical calculus of the ideas immanent in nervous activity, Bull. Math. Biophys. 5 (1943) 115–133. <https://doi.org/10.1007/BF02478259>.
- [82] D.O. Hebb, The organization of behavior: a neuropsychological theory, Science Ed, 1962.
- [83] M.L. Minsky, S. Papert, Perceptrons (An introduction to computational geometry): Epilogue, Handb. Attach. Theory, Res. Clin. (1988). <http://cdsweb.cern.ch/record/114106>.
- [84] D. Riesel, A brief Introduction on Neural Networks, Springer-Verlag, Berlin. (2007) Second edition.
- [85] N. J. Smelser and P. B. Baltes, International Encyclopedia of the Social & Behavioral Sciences, 2001. https://www0.gsb.columbia.edu/faculty/bkogut/files/Chapter_in_smelser-Baltes_2001.pdf.
- [86] Learning and Soft Computing: Support Vector Machines, Neural Networks, and Fuzzy Logic Models | MIT Press eBooks | IEEE Xplore, (n.d.). <https://ieeexplore.ieee.org/book/6267294> (accessed May 3, 2021)., (n.d.).
- [87] H.U. Dike, Y. Zhou, K.K. Deveerasetty, Q. Wu, Unsupervised Learning Based On Artificial Neural Network: A Review, 2018 IEEE Int. Conf. Cyborg Bionic Syst. CBS 2018. (2019) 322–327. <https://doi.org/10.1109/CBS.2018.8612259>.
- [88] Y.W.T. G.E. Hinton, S. Osindero, A fast learning algorithm for deep belief nets, Neural Comput. 18 (2006) 1527–1554. <https://doi.org/https://doi.org/10.1162/neco.2006.18.7.1527>.
- [89] H.G. E., Training Products of Experts by Minimizing Contrastive Divergence, Neural Comput. 14 (2002) 1771–18000.
- [90] V. Nair, G.E. Hinton, 3D object recognition with Deep Belief Nets, Adv. Neural Inf. Process. Syst. 22 - Proc. 2009 Conf. (2009) 1339–1347.
- [91] X. Chen, S. Xiang, C.L. Liu, C.H. Pan, Aircraft detection by deep belief nets, Proc. - 2nd IAPR Asian Conf. Pattern Recognition, ACPR 2013. (2013) 54–58. <https://doi.org/10.1109/ACPR.2013.5>.
- [92] W. Liu, Z. Wang, X. Liu, N. Zeng, Y. Liu, F.E. Alsaadi, A survey of deep neural network architectures and their applications, Neurocomputing. 234 (2017) 11–26. <https://doi.org/10.1016/j.neucom.2016.12.038>.

- [93] A. Krizhevsky, I. Sutskever, G.E. Hinton, ImageNet classification with deep convolutional neural networks, *Adv. Neural Inf. Process. Syst.* 2 (2012) 1097–1105.
- [94] Goodfellow, I., Bengio, Y. ve Courville, A. (2016). *Deep learning*, MIT press, (n.d.).
- [95] F. Demir, (2020). *Konuşma Olmayan Sinyallerin Evrimsel Sinir Ağları Tabanlı Yaklaşımlar İle Sınıflandırılması*, Doktora Tezi, Fırat Üniversitesi, Fen Bilimleri Enstitüsü.
- [96] M.A. Nielsen, *Neural networks and deep learning*, Determinat, San Francisco, CA, USA, 2015.
- [97] G.E. Dahl, T.N. Sainath, G.E. Hinton, Improving deep neural networks for LVCSR using rectified linear units and dropout, *ICASSP, IEEE Int. Conf. Acoust. Speech Signal Process. - Proc.* (2013) 8609–8613. <https://doi.org/10.1109/ICASSP.2013.6639346>.
- [98] S. Ioffe, C. Szegedy, Batch normalization: Accelerating deep network training by reducing internal covariate shift, *32nd Int. Conf. Mach. Learn. ICML 2015.* 1 (2015) 448–456.
- [99] N. Srivastava, (2013), *Improving neural networks with dropout*, PhD Thesis, University of Toronto.
- [100] F. Zang, J.S. Zhang, Softmax discriminant classifier, *Proc. - 3rd Int. Conf. Multimed. Inf. Netw. Secur. MINES 2011.* (2011) 16–19. <https://doi.org/10.1109/MINES.2011.123>.
- [101] J. Sander, B.D. de Vos, I. Išgum, Unsupervised super-resolution: creating high-resolution medical images from low-resolution anisotropic examples, (2021) 44. <https://doi.org/10.1117/12.2580412>.
- [102] A. Makhzani, B. Frey, k-Sparse autoencoders, *2nd Int. Conf. Learn. Represent. ICLR 2014 - Conf. Track Proc.* (2014).
- [103] A. Ng, *Lecture Notes on Sparse Autoencoder*, (2011) 1–19. <http://www.stanford.edu/class/cs294a/>.
- [104] P.A.M. P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, Stacked Denoising Autoencoders: Learning Useful Representations in a Deep Network with a Local Denoising Criterion, *J. Mach. Learn. Res.* 11 (2010) 3371–3408. <file:///Files/5B/5B0A667A-A790-4979-A679-9D6976727DDF.pdf>.
- [105] Z. Meng, X. Zhan, J. Li, Z. Pan, An enhancement denoising autoencoder for rolling bearing fault diagnosis, *Meas. J. Int. Meas. Confed.* 130 (2018) 448–454. <https://doi.org/10.1016/j.measurement.2018.08.010>.
- [106] P. Jiang, M. Maghrebi, A. Crosky, S. Saydam, Unsupervised Deep Learning for Data-Driven Reliability and Risk Analysis of Engineered Systems, *Handb. Neural Comput.* (2017) 417–431. <https://doi.org/10.1016/B978-0-12-811318-9.00023-5>.
- [107] X. Hou, L. Shen, K. Sun, G. Qiu, Deep feature consistent variational autoencoder, *Proc. - 2017 IEEE Winter Conf. Appl. Comput. Vision, WACV 2017.* (2017) 1133–1141. <https://doi.org/10.1109/WACV.2017.131>.
- [108] W. Wang, X.-G. Xia, C. He, Z. Ren, J. Lu, T. Wang, B. Lei, A deep network for sinogram and CT image reconstruction, (2020). <http://arxiv.org/abs/2001.07150>.
- [109] H.Z. Shao, S. Yao, D. Sun, A. Zhang, S. Liu, D. Liu, J. Wang, T. Abdelzaher, ControlVAE: Controllable variational autoencoder, *37th Int. Conf. Mach. Learn. ICML 2020. PartF16814* (2020) 8614–8623.
- [110] I. Sutskever, J. Martens, G. Dahl, G. Hinton, On the importance of initialization and momentum in deep learning, *30th Int. Conf. Mach. Learn. ICML 2013.* (2013) 2176–2184.

- [111] R.A.Jacobs, Increased Rates of Convergence Through Learning Rate Adaptation, *Neural Networks*. 1 (1988) 295–307.
- [112] J. Duchi, E. Hazan, Y. Singer, Adaptive subgradient methods for online learning and stochastic optimization, *J. Mach. Learn. Res.* 12 (2011) 2121–2159.
- [113] T. Tieleman, G. Hinton, Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude, *COURSERA Neural Networks Mach. Learn.* 4 (2012) 26–31.
- [114] J.L. Ba, D.P. Kingma, Adam: A method for stochastic optimization, 3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc. (2015) 1–15.
- [115] Y. Sun, Y. Chen, X. Wang, X. Tang, Deep learning face representation by joint identification-verification, *Adv. Neural Inf. Process. Syst.* 3 (2014) 1988–1996.
- [116] S.J. Pan, Q. Yang, A survey on transfer learning, *IEEE Trans. Knowl. Data Eng.* 22 (2010) 1345–1359. <https://doi.org/10.1109/TKDE.2009.191>, (n.d.).
- [117] L.-J. Li, K. Li, F.F. Li, J. Deng, W. Dong, R. Socher, L. Fei-Fei, ImageNet: a Large-Scale Hierarchical Image Database Human trajectory forecasting View project Shrimp Project View project ImageNet: A Large-Scale Hierarchical Image Database, 2009 IEEE Conf. Comput. Vis. Pattern Recognit. (2009) 248–255. <https://www.researchgate.net/publication/221361415>.
- [118] S. Akçay, M.E. Kundegorski, M. Devereux, T.P. Breckon, Transfer learning using convolutional neural networks for object classification within X-ray baggage security imagery, *Proc. - Int. Conf. Image Process. ICIP. 2016-Augus (2016)* 1057–1061. <https://doi.org/10.1109/ICIP.2016.7532519>.
- [119] E. Alpaydın, *Yapay Öğrenme*, Boğaziçi Üniversitesi Yayınevi. (2018) 486.
- [120] R. Lienhart, A. Kuranov, V. Pisarevsky, Empirical analysis of detection cascades of boosted classifiers for rapid object detection, *Lect. Notes Comput. Sci. (Including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*. 2781 (2003) 297–304. https://doi.org/10.1007/978-3-540-45243-0_39.
- [121] P. Viola, M. Jones, Rapid object detection using a boosted cascade of simple features, *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.* 1 (2001). <https://doi.org/10.1109/cvpr.2001.990517>.
- [122] F.A. Gers, N.N. Schraudolph, J. Schmidhuber, Learning precise timing with LSTM recurrent networks, *J. Mach. Learn. Res.* 3 (2003) 115–143. <https://doi.org/10.1162/153244303768966139>.
- [123] L. Li, J. Bao, T. Zhang, H. Yang, D. Chen, F. Wen, B. Guo, Face X-ray for more general face forgery detection, *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.* (2020) 5000–5009. <https://doi.org/10.1109/CVPR42600.2020.00505>.
- [124] Z. Akhtar, M.R. Mouree, D. Dasgupta, Utility of Deep Learning Features for Facial Attributes Manipulation Detection, *Proc. - 2020 IEEE Int. Conf. Humaniz. Comput. Commun. with Artif. Intell. HCCAI 2020.* (2020) 55–60. <https://doi.org/10.1109/HCCAI49649.2020.00015>.
- [125] N. Yu, L. Davis, M. Fritz, Attributing fake images to GANs: Learning and analyzing GAN fingerprints, *Proc. IEEE Int. Conf. Comput. Vis.* 2019-Octob (2019) 7555–7565. <https://doi.org/10.1109/ICCV.2019.00765>.
- [126] T. Miyato, T. Kataoka, M. Koyama, Y. Yoshida, Spectral normalization for generative adversarial networks, 6th Int. Conf. Learn. Represent. ICLR 2018 - Conf. Track Proc. (2018).

- [127] M.G. Bellemare, I. Danihelka, W. Dabney, S. Mohamed, B. Lakshminarayanan, S. Hoyer, R. Munos, The Cramer Distance as a Solution to Biased Wasserstein Gradients, (2017). <http://arxiv.org/abs/1705.10743>.
- [128] M. Binkowski, D.J. Sutherland, M. Arbel, A. Gretton, Demystifying MMD GANs, 6th Int. Conf. Learn. Represent. ICLR 2018 - Conf. Track Proc. (2018).
- [129] M.A. Ganaie, M. Hu, M. Tanveer*, P.N. Suganthan*, Ensemble deep learning: A review, (2021). <http://arxiv.org/abs/2104.02395>.
- [130] M.T. Jafar, M. Ababneh, M. Al-Zoube, A. Elhassan, Digital Forensics and Analysis of Deepfake Videos, 2020 11th Int. Conf. Inf. Commun. Syst. ICICS 2020. (2020) 53–58. <https://doi.org/10.1109/ICICS49469.2020.239493>.



ÖZGEÇMİŞ

Semih YAVUZKILIÇ

[illegible]

AKADEMİK FAALİYETLER

Makaleler:

1. S. Yavuzkiliç, Z. Akhtar, A. Sengür, K. Siddique, DeepFake Face Video Detection Using Hybrid Deep Residual Networks and LSTM Architecture, *AI Deep Learn. Biometric Secur.* (2021) 81–104. <https://doi.org/10.1201/9781003003489-4>.

Bildiriler:

1. A. . Sengur, B.N. Akilolu, S.A. Tuncer, Z. Kadiroglu, S. Yavuzkilic, U. Budak, E. Deniz, Optic disc determination in retinal images with deep features, 26th IEEE Signal Process. Commun. Appl. Conf. SIU 2018. (2018) 1–4. <https://doi.org/10.1109/SIU.2018.8404339>.
2. S. Yavuzkilic, A. Sengur, Z. Cambay, I. Emre, Video Analysis of Morris Water Maze Experiments, 8th Int. Adv. Technol. Symp. (2017) 1941–1946.
3. S. Yavuzkiliç, M. Aslan, A. Şengür, A Matlab Tool for Morris Water Maze Test Implementation, 2018 Int. Conf. Artif. Intell. Data Process. IDAP 2018. (2019). <https://doi.org/10.1109/IDAP.2018.8620841>.

Kitap Bölümleri:

1. S. Yavuzkiliç, Z. Akhtar, A. Sengür, K. Siddique, DeepFake Face Video Detection Using Hybrid Deep Residual Networks and LSTM Architecture, *AI Deep Learn. Biometric Secur.* (2021) 81–104. <https://doi.org/10.1201/9781003003489-4>.