T.C.
KASTAMONU ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
BİLGİSAYAR MÜHENDİSLİĞİ ANA BİLİM DALI

DERİN ÖĞRENME YÖNTEMLERİ KULLANILARAK DEEPFAKE MEDYA DOSYALARININ TESPİTİ

RIFAT KÖSE

YÜKSEK LİSANS TEZİ

DR. ÖĞR. ÜYESİ MURAT MERİÇELLİ

ARALIK - 2024
KASTAMONU

TEZ ONAYI

Rıfat KÖSE tarafından hazırlanan "**Derin Öğrenme Yöntemleri Kullanılarak Deepfake Medya Dosyalarının Tespiti**" adlı tez çalışmasının savunma sınavı **27.12.2024** tarihinde yapılmış olup aşağıda verilen jüri tarafından oy birliği / oy çokluğu ile Kastamonu Üniversitesi Fen Bilimleri Enstitüsü **Bilgisayar Mühendisliği Ana Bilim Dalı Yüksek Lisans Tezi** olarak kabul edilmiştir.

Danışman	Dr. Öğr. Üyesi Murat MERİÇELLİ Kastamonu Üniversitesi	.....
Jüri Üyesi	Doç. Dr. Salih GÖRGÜNOĞLU Kastamonu Üniversitesi	.....
Jüri Üyesi	Doç. Dr. Mürsel Ozan İNCETAŞ Alanya Alaaddin Keykubat Üniversitesi	.....

Jüri üyeleri tarafından kabul edilmiş olan bu tez Kastamonu Üniversitesi Fen Bilimleri Enstitüsü Yönetim Kurulunca onanmıştır.

Enstitü Müdürü Doç. Dr. Selçuk MEMİŞ .....

TAAHHÜTNAME

Bu tezin tasarımı, hazırlanması, yürütülmesi, araştırmalarının yapılması ve bulgularının analizlerinde bütün bilgilerin etik davranış ve akademik kurallar çerçevesinde elde edilerek sunulduğunu; ayrıca tez yazım kurallarına uygun olarak hazırlanan bu çalışmada bana ait olmayan her türlü ifade ve bilginin kaynağına eksiksiz atıf yapıldığını, bilimsel etiğe uygun olarak kaynak gösterildiğini bildirir ve taahhüt ederim.

Rıfat KÖSE

ÖZET

YÜKSEK LİSANS TEZİ

DERİN ÖĞRENME YÖNTEMLERİ KULLANILARAK DEEPFAKE MEDYA DOSYALARININ TESPİTİ

RIFAT KÖSE

KASTAMONU ÜNİVERSİTESİ FEN BİLİMLERİ ENSTİTÜSÜ
BİLGİSAYAR MÜHENDİSLİĞİ ANA BİLİM DALI
DANIŞMAN:DR. ÖĞR. ÜYESİ MURAT MERİÇELLİ

Deepfake medyalar, insanların görüntülerinin ve/veya seslerinin değiştirildiği, taklit edildiği her türlü görsel işitsel verilerdir. Genellikle insanların yüzlerinin değiştirildiği deepfake medya türü ile daha sık karşılaşılmaktadır. Teknolojik gelişmelerin paralelinde yapay zeka algoritmalarında görülen gelişmeler çok daha gerçekçi deepfake medyalar üretilmesine olanak sağlamıştır. Deepfake medyaların birçok farklı alanda iyi niyetli ya da kötü niyetli olarak kullanım örnekleri görülmektedir. Bireylerin deepfake medya teknolojisinin kötüye kullanımına maruz kalmaması için sosyal medya gibi platformlarda herkese açık olarak görsel ve işitsel medyalarını paylaşmaması gerekmektedir. Deepfake medyaların kötüye kullanımını engellemek adına devletlerin de gerekli önlemleri alması çok önemlidir. Son dönemlerde deepfake medyaların tespit edilmesi hakkında akademik çalışmaların sayısında artış görülmektedir. Çalışmamızın ana konusu deepfake medyaların derin öğrenme mimarileri kullanılarak tespit edilmesine yöneliktir. Bu kapsamda 5 farklı ön eğitimli model (VGG16, EfficientNetB4, DenseNet201, InceptionV3, ResNet50V2) Google Colab ortamında FaceForensics++ veri seti üzerinde test edilmiştir. 0,93 AUC değeri ile en başarılı model EfficientNetB4 olmuştur.

ANAHTAR KELİMELER: Derin Sahte Video, Yapay Zeka, Derin Öğrenme, Öğrenme Aktarımı, Yapay Sinir Ağı, Evrişimli Sinir Ağı

Aralık 2024, 73 Sayfa

ABSTRACT

MSC THESIS

DETECTION OF DEEPFAKE MEDIA FILES USING DEEP LEARNING METHODS

RIFAT KÖSE

KASTAMONU UNIVERSITY INSTITUTE OF SCIENCE
DEPARTMENT OF COMPUTER ENGINEERING
SUPERVISOR:ASST. PROF. DR. MURAT MERİÇELLİ

Deepfake media is any kind of audiovisual data where people's images and/or voices are changed or imitated. Deepfake media, where people's faces are changed, is more frequently encountered. Developments in artificial intelligence algorithms in parallel with technological developments have made it possible to produce much more realistic deepfake media. There are examples of deepfake media being used in many different areas, both well-intentioned and malicious. In order for individuals not to be exposed to the misuse of deepfake media technology, they should not share their visual and audio media publicly on platforms such as social media. It is also very important for states to take the necessary measures to prevent the misuse of deepfake media. In recent years, there has been an increase in the number of academic studies on the detection of deepfake media. The main subject of our study is to detect deepfake media using deep learning architectures. In this context, 5 different pre-trained models (VGG16, EfficientNetB4, DenseNet201, InceptionV3, ResNet50V2) were tested on the FaceForensics++ dataset in the Google Colab environment. EfficientNetB4 was the most successful model with an AUC value of 0,93.

KEYWORDS: Deepfake Video, Artificial Intelligence, Deep Learning, Transfer Learning, Artificial Neural Network, Convolutional Neural Network

December 2024, 73 Page

TEŞEKKÜR

Bu tez çalışmasının hazırlanması süresince bilgi ve tecrübeleri ile beni destekleyen ve yardımlarını esirgemeyen danışman hocam sayın Dr. Öğr. Üyesi Murat MERİÇELLİ'ye saygılarımla teşekkür ederim.

Lisans ve Yüksek Lisans eğitimim süresince derslerde bilgi ve tecrübelerini paylaşarak bana bu konuları sevdiren öncelikle Doç. Dr. Salih GÖRGÜNOĞLU ve Doç. Dr. Kemal AKYOL hocam olmak üzere Kastamonu Üniversitesi Bilgisayar Mühendisliği Bölümünde görev yapan tüm öğretim üyelerine saygılarımı sunar ve teşekkür ederim.

Son olarak tüm eğitim hayatım süresince ve özellikle bu tez çalışmasının oluşturulması sürecinde bana inanarak desteğini hiç esirgemeyen sevgili eşim Gülşen KÖSE'ye sonsuz teşekkürlerimi sunarak bu süreçte yeterince vakit ayırarak ilgilenemediğim kızlarım Zeynep Hüma KÖSE ve Duru Mina KÖSE'den özür diliyor ve bu çalışmayı onlara ithaf ediyorum.

Rıfat KÖSE

Kastamonu, 2024

İÇİNDEKİLER

Sayfa

ŞEKİLLER DİZİNİ

Sayfa

TABLOLAR DİZİNİ

Sayfa

SİMGELER VE KISALTMALAR DİZİNİ

Simgeler

$\sum$	**:** Sigma (Toplam Sembolü)
σ	**:** Sigmoid Aktivasyon Fonksiyonu
∞	**:** Sonsuzluk İşareti

Kısaltmalar

GAN	**:** Generative Adversarial Networks
AI	**:** Artificial Intelligence
CGI	**:** Computer Generated Imagery
NLP	**:** Neuro Linguistic Programming
VAE	**:** Variational Autoencoder
DSSIM	**:** Difference Structure Similarity Index Method
MSE	**:** Mean Squared Error
CV	**:** Computer Vision
CNN	**:** Convolutional Neural Network
RNN	**:** Recurrent Neural Network
LSTM	**:** Long Short Tertm Memory
LRCN	**:** Long Term Recurrent Convolutional Network
TÜBİTAK	**:** Türkiye Bilimsel ve Teknolojik Araştırma Kurumu
TRUBA	**:** Türk Ulusal Bilim e-Altyapısı
SAAS	**:** Software As A Service
MLP	**:** Multi Layer Perceptron
DNN	**:** Deep Neural Network
RGB	**:** Red Green Blue

1. GİRİŞ

Yapay zeka (AI – Artificial Intelligence) yakın geçmişte hemen herkesin hakkında azda olsa bilgisi olduğu bir teknoloji haline gelmiştir. Evlerde kullanılan birçok eşya yapay zeka teknolojisini kullanarak akıllı ev konseptini oluşturmuştur. Ayrıca kullanılan taşıtlar, cep telefonları, kol saatleri v.b. yapay zeka teknolojisini kullanmaktadır. Hayatımızı kolaylaştıran bu teknoloji her ne kadar iyi yönde kullanılsa da kötü niyetli insanlar tarafından kullanılarak bizlere zarar verebilir. Bunun en belirgin örnekleri arasında, son yıllarda insanların adını sıkça duyduğu deepfake teknolojisi yer almaktadır.

Deepfake, temelinde yapay zekanın alt dalı olan makine öğrenmesi ve yapay sinir ağlarını temel alan derin öğrenme teknolojisini kullanmaktadır. Bu teknoloji kullanılarak insanların yüzleri başka insanların yüzü ile değiştirilebilir, yüz ifadeleri ve mimikleri değiştirilebilir, sesi taklit edilebilir. Kısacası insanların görüntüsü ve sesleri değiştirilerek gerçekçi medya dosyaları oluşturulabilir.

Deepfake medyaların oluşturulabilmesi için kullanılan derin öğrenme teknolojisi, temelde binlerce hatta milyonlarca veriye ihtiyaç duymaktadır. Teknoloji kullanımın artması nedeniyle yediden yetmişe herkesin sahip olduğu cep telefonları ve her gün saatlerce vakit harcanılan sosyal medya platformları, video paylaşma platformları derin öğrenme için gerekli olan veriyi fazlasıyla sağlamaktadır. Bu platformlarda paylaşılan resim ve videolardan elde edilen görüntü ve ses dosyaları, derin öğrenme mimarileri kullanılarak işlenmekte ve ortaya aslında hiçbir zaman var olmamış yeni medyalar ortaya çıkmaktadır. Bu tarz sahte medyalar insanların hayatını zora sokabilecek sonuçlar doğurabilmektedir. Bu nedenle insanların görüntülerini paylaşırken herkese açık bir şekilde paylaşmaması ve veri güvenliğini sağlamak için önlem alması hayati önem taşımaktadır.

İnsanların hayatını zora sokabilecek bu teknolojinin üretmiş olduğu medya türleri, önceki zamanlarda gözle görülebilir şekilde ayırt edilebilirken gelişen teknoloji ile birlikte ancak uzmanlar tarafından ayırt edilebilir hale gelmiştir. Uzmanların kullandığı yöntemlerin bile yetersiz kaldığı durumlarda ise deepfake medyalar

oluştururken kullanılan, derin öğrenme mimarileri yardımımıza koşmaktadır. Son yıllarda deepfake medyaların tespiti için derin öğrenme mimarilerinin kullanımı hızla artmakta olup büyük teknoloji firmalarının düzenlediği ödüllü yarışmalar bu alanda yapılacak olan çalışmaları teşvik etmektedir.

Ülkemizde ise bu tarz çalışmalar birkaç yıldır önem kazanmakla birlikte yeteri kadar önem verilmediği düşünülmektedir. Yapılan çalışmalar lisansüstü tez ve makalelerden oluşmakta olup bu alanda literatüre katkı sağlaması bakımından, çalışmanın önemini ortaya çıkarmaktadır.

Çalışmada, deepfake medya dosyalarının tespiti için daha önce, büyük bir veri seti olan ImageNet veri setinde eğitilmiş olan önceden eğitilmiş (pretrained) modeller kullanarak, Öğrenme Aktarımı (Transfer Learning) yöntemi kullanılmıştır. Sahte medyaların tespiti için FaceForensics++ veri seti tercih edilmiştir. Kullanılan hazır modellerden EfficientNetB4 en yüksek başarıyı sağlayan model olmuştur.

2. LİTERATÜR

Deepfake medyaların farklı sektörlerde kullanımı ve artan önemi nedeniyle son yıllarda deepfake tespit çalışmalarında da artış görülmektedir. Araştırmacılar deepfake tespiti için farklı teknikler ve farklı derin öğrenme mimarileri kullanmışlardır.

Afchar ve arkadaşları, deepfake medya oluşturmak için yaygın olarak kullanılan Deepfake ve Face2Face teknikleri ile oluşturulan videolarda deepfake tespiti için görüntülerin mezoskopik özelliklerine odaklanarak Meso-4 ve MesoInception-4 adıyla 2 farklı ağ önermişlerdir. Face2Face tekniği için FaceForensics++ veri seti, Deepfake tekniği için ise kendi oluşturdukları veri seti kullanılmıştır. Face2Face tekniği için %95, Deepfake tekniği için %98 başarı oranı sağlanmıştır (Afchar vd., 2018).

Li ve arkadaşları, deepfake medya tespiti için görüntülerde fizyolojik bir sinyal olan göz kırpma tespitine odaklanmıştır. Kendi oluşturdukları veri setinde yüz hizalaması için yüz işaretlerini çıkararak gözlerin açık ve kapalı durumunu tespit etmek üzere Evrişimli Sinir Ağı (CNN) ve Tekrarlayan Sinir Ağı (RNN) birleşiminden oluşan Uzun Süreli Tekrarlayan Evrişimli Sinir Ağı (LRCN) kullanılmış ve %98 başarı sağlanılmıştır (Li vd., 2018).

Nguyen ve arkadaşları, Deepfake, Face2Face ve FaceSwap teknikleriyle üretilen deepfake medyaların tespiti için kapsül ağları temel alan Kapsül Adli Tıp (Capsule Forensics) ağını geliştirmişlerdir. Kapsül Adli Tıp ağı FaceForensics++ veri setinde eğitilerek Deepfake tekniği için 92,17 Face2Face tekniği için 90,36 FaceSwap tekniği için 92,79 başarı puanı elde edilmiştir (Nguyen vd., 2019).

Guarnera ve arkadaşları, deepfake medya tespiti için görüntü oluşturma sürecinde ortaya çıkan bir tür parmak izi olan evrişimsel izleri kullanmışlardır. Geliştirilen modelin eğitim ve testi için CelebA veri seti ve 5 farklı Çekişmeli Üretici Ağ (GAN) kullanılarak oluşturulan veri setleri olmak üzere 6 farklı veri seti kullanılmıştır. StyleGAN2 yöntemiyle oluşturulan veri seti üzerinde %99,81'lik başarı puanına ulaşılmıştır (Guarnera vd., 2020).

Zhao ve arkadaşları, kullandıkları CNN ağına görüntüleri çoklu dikkat bölgelerine ayırmak için birden fazla mekânsal dikkat başlığı, sığ özelliklerdeki ince eserleri (artefact) yakalamak için dokusal özellikleri geliştirme bloğu ekleyerek sonrasında düşük seviyeli dokusal özellik ve yüksek seviyeli anlamsal özellikleri bir araya getirmeyi hedeflemişlerdir. Celeb-DF, DFDC ve FaceForensics++ veri setleri üzerinde EfficientNet-B4 ve Xception ağlarını kullanıp, eğitimlerini gerçekleştirerek FaceForensics++ veri seti üzerinde EfficientNet-B4 ağıyla %99,29 doğruluk (accuracy) değeri elde etmişlerdir (Zhao vd., 2021).

Wang ve arkadaşları, farklı mekânsal düzeylerdeki görüntülerin yerel tutarsızlıklarını tespit etmek için farklı boyutlarda yamalar üzerinde çalışan Çok Modlu Çok Ölçekli Dönüştürücü (M2TR) modelini geliştirmişlerdir. Bu modeli kullanarak ince manipülasyon eserlerini tespit etmeyi hedeflemişlerdir. Ayrıca, FaceForensics++ veri setinde bulunan orijinal videolar kullanılarak oluşturulan deepfake veri seti olan SR-DF tanıtılmıştır. Yazarlar önerdikleri modeli kendi veri setleri SR-DF ile FaceForensics++, Celeb-DF, ve ForgeryNet veri kümelerinde değerlendirmişlerdir. FaceForensics++'da %99,92; Celeb-DF'de %95,5; SR-DF'de %86,7 ve ForgeryNet'te %82,52'lik bir AUC puanı elde etmişlerdir (J. Wang vd., 2022).

Raza ve Malik, sesli deepfake videolarda, ses ve görüntülerden öğrenilmiş kanalları çıkardıktan sonra IntrAmodality Mixer Layer'da (IAML) bağımsız olarak karıştıran, bunları IntErModality Mixer Layers'da (IEML) birlikte işleyen ve sonuçları çok etiketli sınıflandırma başlığına besleyen birleşik birçok modlu çerçeve olan Multimodaltrace'i önermişlerdir. Model, FakeAVCeleb veri setinde %92,9 doğruluk değeri elde etmiştir (Raza ve Malik, 2023).

Zhang ve arkadaşları, görme dönüştürücü omurgasına dayalı yeni bir Alan Kayma Modelleme (DSM) çerçevesi önermişlerdir. Ayrıca alan kaymalarını modellemek, alan bozulmalarını hafifletmek ve alan kaymaları için daha iyi genelleme yapmak üzere Dikkat Rehberliğinde Yama Maskeleme (AGPM) ve Özellik İstatistik Kayma Tahmini (FSSE) modülleri geliştirilmiştir. Önerilen model FaceForensics++, Celeb-DF, DFDC ve DeeperForensics veri setlerinde değerlendirilmiş ve FaceForensics++ veri setinde %99 AUC değeri yakalamıştır (Zhang vd., 2024).

3. DEEPFAKE TANIMI VE TEMEL KAVRAMLAR

Deepfake terimi ilk olarak 2017 yılında, bir sosyal tartışma platformu olan Reddit'te "deepfakes" isimli bir kullanıcının paylaşımları ile gündeme gelmiştir. Bu kullanıcı ünlülerin yüzlerini, yetişkin içerikli videolarda bulunan insanların yüzüne ekleyerek sahte videolar üretmiştir. 2018 yılında bu kullanıcının hesabı Reddit tarafından engellenmiştir (Kirchengast, 2020). Günümüzde ise Facebook ve Google gibi firmalar deepfake medyaları engellemek için çeşitli yasaklar getirmiştir. Amerika Birleşik Devletleri'nin bazı eyaletleri, Avrupa Birliği ve Çin gibi büyük ülkeler deepfake teknolojisini yasaklayan yasalar çıkarmıştır.

Deepfake kelimesi temel olarak derin öğrenme (deep learning) terimi ve sahte (fake) terimlerinin birleşiminden türemiştir. Genel olarak hedef kişinin yüz görüntülerini, kaynak kişinin yüz görüntüsüne bindirerek hedef kişinin bir şeyler yaptığı veya söylediği videolar yaratmak için kullanılmaktadır (Nguyen vd., 2022).

Deepfake, yapay zeka ve derin öğrenme mimarisinin kullanılarak kişilerin görüntülerinde, hareketlerinde, konuşmalarında, jest ve mimiklerinde gerçekte var olmayan değişiklikler yaparak sentetik medya oluşturmaya yarayan bir teknolojidir.

Deepfake medyalar genellikle, bir derin öğrenme yöntemi olan Çekişmeli Üretici Ağlar (GAN - Generative Adversarial Networks) kullanılarak oluşturulan sahte medyalardır. GAN, Ian Goodfellow tarafından 2014 yılında ortaya çıkarılmıştır. Bu ağ milyonlarca insan resmini kullanarak kendini eğitmiş ve gerçekte olmayan insan yüzleri üretmeyi başarmıştır (Karakoç ve Zeybek, 2022). GAN kullanılarak oluşturulmak istenen deepfake medyalarında hedef kişinin ne kadar çok görseli kullanılırsa ve eğitim süresi ne kadar uzun olursa sonuç o kadar iyi olabilmektedir (Belada, 2024). Görüntü, ses ve metin üzerinde manipülasyonlar yaparak yeni medyalar üreten bu teknoloji gelişen teknoloji ile birlikte hızlı bir şekilde ilerleme katetmektedir.

Bilgisayarların hayatımıza girmesi ile birlikte Paint gibi programlar aracılığıyla resimler üzerinde basit değişiklikler yapılmakta iken Adobe Photoshop gibi daha

profesyonel programlar, görüntüler üzerinde daha kaliteli manipülasyonlar yapılmasına olanak sunmaktadır. FaceApp, gibi yapay zeka tabanlı video düzenleme araçlarındaki son gelişmeler ile birlikte deepfake medya üretimi çok daha kolay hale gelmiştir. Deepfake ile üretilen bu medyalar çok farklı amaçlar için kullanılabilmektedir (Rana vd., 2021).

3.1 Deepfake Kullanım Alanları

Deepfake terimi ortaya çıkmadan önce de medya dosyaları üzerinde çeşitli manipülasyon işlemleri yapılmaktaydı. Bu işlemler için yapay zeka teknolojisinin kullanılmaya başlamasıyla birlikte çok daha hızlı ve kolay bir şekilde, çok daha başarılı sahte medyalar oluşturulabilmesi neticesinde, deepfake teknolojisinin kullanım alanı genişlemiştir. Deepfake teriminin isim babası olan "deepfakes" isimli Reddit kullanıcısı bu teknolojiyi kötü bir hedef doğrultusunda kullanmış olup farklı sektörlerde ve farklı alanlarda iyi yönde kullanım örnekleri de mevcuttur.

3.1.1 Yararlı Kullanım Alanları

Deepfake'in yararlı kullanım alanları sinema ve televizyon, sosyal medya ve telefon, müzik, eğitim ve sağlık olarak sıralanmaktadır.

- **Sinema ve Televizyon:** Deepfake teknolojisi sinema alanında sıkça kullanılır hale gelmiştir. Bu teknoloji sayesinde hayatta olmayan sinema sanatçılarının yüzleri ve sesleri dublörlerin görüntülerinin üzerine bindirilerek, gerçekte hayatta olmayan sanatçının videosu gibi kullanılmaktadır. Şekil 3.1'de gösterilen Yeşilçam Sineması filmlerinde oynamış olan; Münir ÖZKUL, Hülya KOÇYİĞİT, Halit AKÇATEPE, Hulusi KENTMEN, Adile NAŞİT ve bazı Hababam Sınıfı filmi oyuncularının yüzleri kullanılarak ulusal bir bankanın reklam filmlerinde oynatılmıştır. Bir örnek de Hızlı ve Öfkeli film serisinin başrol oyuncularından olan ve film çekimleri devam ederken araba kazasında vefat eden Paul WALKER'ın yüzü Bilgisayar Tabanlı Görüntü (CGI – **C**omputer Generated Imagery) teknolojisi kullanılarak kardeşinin yüzü ile değiştirilmesidir.

Şekil 3.1 Yeşilçam oyuncularının deepfake ile yapılan reklam görüntüleri

- **Sosyal Medya ve Telefon Uygulamaları:** Bazı sosyal medya siteleri ve telefon programlarında yapay zeka teknolojisi kullanılarak filtreler aracılığıyla insanların yüzleri değiştirilmekte, fotoğraflar canlandırılarak videolar oluşturulmaktadır. Özellikle son zamanlarda kullanılan bazı programlarla insanlar kaybettiği yakınlarının resimlerini kullanarak hareketli medya haline getirmekte bu teknolojiyi kullanmaktadır.

- **Müzik:** Deepfake teknolojisi sadece görüntülerde değil ses üzerinde de manüpilasyon yapmak için kullanılmaktadır. Mevcut sanatçıların yeni şarkılarının, hayatta olmayan bazı sanatçıların sesleri kulllanılarak değiştirilmesine son dönemde sıkça rastlanılmaktadır. Buna bir örnek olarak Mabel MATİZ'in şarkıları, Barış MANÇO ve Cem KARACA'nın sesi kullanılarak oluşturulan yeni şarkılar video izleme sitelerinde milyonlarca kişi tarafından izlenme almaktadır.

- **Eğitim:** Telefon uygulamaları ve internet siteleri aracılığıyla online eğitimlerin verilmesinin yaygınlaşması, bu alanda deepfake kullanımını da beraberinde getirmiştir. Bazı online eğitim platformları eğitim sırasında deepfake teknolojisi ile yaratılmış aslında gerçek olmayan insanların görüntülerini kullanarak eğitimlerinde kullanmaktadır.

- **Sağlık:** Amiyotrofik Lateral Skleroz (ALS), Multipl Skleroz (MS) gibi motor nöron hastalıkları yüzünden konuşamayan hastaların sesini, yapay zeka

teknolojisi ile oluşturarak bu bireylerin seslerini yapay olarak kullanabilmesini sağlamaktadır (Çiçek ve Yalçın, 2024).

3.1.2 Zararlı Kullanım Alanları

Deepfake'in zararlı kullanım alanları siyaset, dolandırıcılık ve hukuk olarak sıralanmaktadır.

- **Siyaset:** Deepfake kullanımının en yoğun olduğu alanlardan birisi şüphesiz siyasettir. Ülkelerin en üst düzey yöneticilerinin manipüle edilmiş videolar sıkça karşımıza çıkmaktadır. Özellikle savaş durumunda olan ülkelerin düşmanları tarafından, savaşın seyrini değiştirebilecek açıklamalar yapıldığı deepfake ile oluşturulan videolar karşımıza çıkabilmektedir. Şekil 3.2'de gösterilen, ABD eski başkanları Donald Trump, Barack Obama, Rusya Devlet Başkanı Vladimir Putin, Ukrayna Devlet Başkanı Volodimir Zelenski ve diğer birçok devlet başkanının liderlerinin deepfake videolarına sıkça rastlanılmaktadır (Yılmaz, 2024). Deepfake teknolojisinin bu alanda kullanımı siyasilerin itibarlarını zedelemekte, savaşların seyirlerini değiştirmektedir.

Şekil 3.2 Devlet başkanlarının sahte görüntüleri

- **Dolandırıcılık:** Kötü niyetli insanlar tarafından görüntü ve ses manipülasyonu yoluyla bireylerin ve hatta şirketlerin bile dolandırıldığı haberler gün geçtikçe artış göstermektedir. Hong Kong'da bir şirketin çalışanları ile yöneticileri arasında yapılan online görüşmede, yöneticinin deepfake ile yapılmış görüntüsü ile şirket çalışanlarına yüklü miktarda para havalesi talimatı vererek dolandırması bunun en büyük örneklerindendir (Tunçer, 2024).

- **Hukuk:** Deepfake ile yapılan videolarda özel hayatın gizliliği konusunda birçok ihlal bulunmaktadır. Özellikle ünlülerin görüntüleri kullanılsa da bu teknolojinin yaygınlaşması ile birlikte tüm insanları tehdit eder duruma gelmiştir. İnsanların çıplak videoları yapılarak şantaj aracı haline getirilmesi, deepfake teknolojisinin kötüye kullanımı konusunda en sık karşılaştığımız durumdur. Kadınların görüntülerini kullanarak üzerlerindeki kıyafetleri yok edip kişinin çıplak görüntülerini oluşturan "DeepNude" isimli uygulama sonralarda yasaklansa da kötü niyetli insanlar tarafından bu teknolojinin tekrar kullanılması olasıdır (Yeh vd., 2020).

3.2 Deepfake Türleri

Teknolojinin hızlı ilerlemesi ve bilgisayar donanımlarının gelişimiyle birlikte yazılım sektörü de büyük bir ivme kazanmıştır. Özellikle derin öğrenme algoritmalarının daha etkili sonuçlar vermesi için gerekli olan altyapı hızla olgunlaşmaktadır. Bu durum, deepfake teknolojisinin kullanım alanlarını genişletmekte, metin, ses ve görüntü üzerinde çeşitli deepfake medya içeriklerinin oluşturulmasını mümkün kılmaktadır.

3.2.1 Metin Üzerinde Yapılan Deepfake

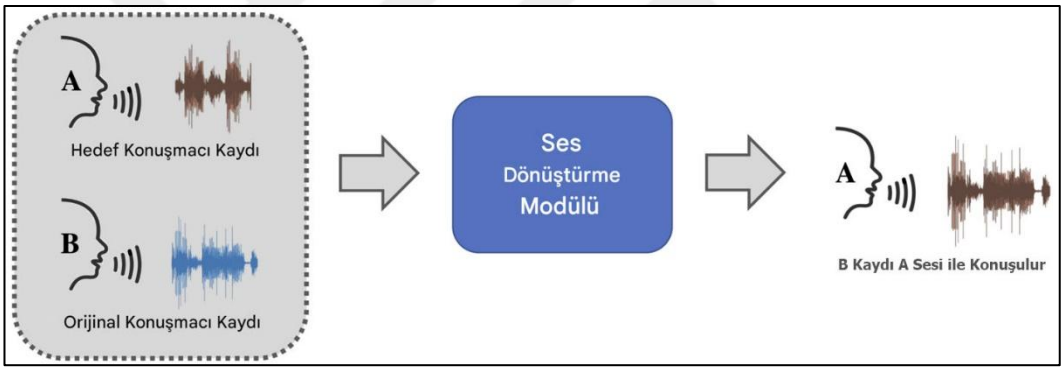
Yapay zekanın alt dalı olan Doğal Dil İşleme (NLP - Neuro Linguistic Programming), ses üzerinde oynama yapmaktan ziyade seslerin yazıya dökülerek metin olarak işlenmesini kapsamaktadır (Şeker, 2015). Deepfake ile oluşturulan videolarda bulunan konuşmaları üretmek için metin sentezi kullanılabilmektedir. Metinsel deepfakeler, yapay zeka tarafından oluşturulan sentetik metinlerdir. Son zamanlarda adını sıkça duyduğumuz ChatpGPT, metin sentezi yapabilen bir sohbet robotu olarak, kullanıcılarına istedikleri bilgilerle ilgili metinler sunabilmektedir (Kırık ve Özkoçak, 2023).

3.2.2 Ses Üzerinde Yapılan Deepfake

Ses sentezi, birinin konuşma kalıplarını ve tonlamalarını taklit eden sentetik ses kayıtları oluşturma sürecidir. Bu süreç Şekil 3.3'te açıkça görülebilmektedir. Ayrıca deepfake kullanılarak insanların sesleri de taklit edilebilmektedir. Bunun için gerekli

olan ise sesi taklit edilecek olan kaynak kişinin ses dosyalarından oluşan verilerdir. Deepfake bu verileri işleyerek kaynak kişinin sesini taklit eder ve istenilen metini kaynak kişi konuşmuş gibi işleyerek medyalar oluşturabilir. Lyrebird ve Deep Voice gibi ticari yazılımlar deepfake teknolojisini kullanarak kullanıcıdan aldığı sesleri taklit ederek verilen metni kullanıcının ses tonuyla okuyan medyalar oluşturmaktadır (BasuMallick C., 2022).

Ses üzerinde deepfake kullanımı müzik sektöründe de kendine yer bulmuştur. Hayatını yitiren bazı sanatçıların ses tonu kullanılarak sentetik şarkılar oluşturulmaktadır. Ayrıca WaveAI gibi yazılımlar aracılığıyla şarkı sözleri, melodi ve hatta yapay ses oluşturarak tamamen sentetik şarkılar oluşturabilmektedir (Yurdigül ve Yıldırım, 2021). Deepfake videolar oluştururken kişilerin seslerinin taklit edilmesi de ses üzerinde yapılan deepfake örneği olarak gösterilebilir.



Şekil 3.3 Ses üzerinde deepfake kullanımı(Wikimedia, 2022)

3.2.3 Görüntü Üzerinde Yapılan Deepfake

Görüntü üzerinde yapılan deepfake, genellikle insanların yüzleri üzerinde işlem yapmaktadır. Yüz üzerinde yapılan manipülasyonlar hakkında literatürde yer alan çalışmalar incelendiğinde 4 ana başlık altında incelenmesinin daha doğru olduğu düşünülmektedir (Dang ve Nguyen, 2023).

3.2.3.1 Yüz sentezi

Yüz sentezi temelinde binlerce hatta milyonlarca insan yüzü görüntüsünü kullanarak gerçekte var olmayan insan yüzleri oluşturmak olarak tanımlanabilir. Bu şekilde

sadece insan yüzleri değil istediğimiz tüm canlı cansız varlıkların gerçekte olmayan sentezleri üretilebilir. Şekil 3.4'te gösterilen yüz sentezi oluşturma işlemi GAN kullanılarak yapılmaktadır.

GAN'lar üretken modelleme problemlerinde çözüm sağlamak için ortaya atılan yapay zeka algoritmalarıdır. Bu ağların temel amacı, sentetik yüz oluşturabilmek için bir yüz veri setini inceleyerek bunların üreten olasılık dağılımını öğrenmek ve bu olasılık dağılımına göre yeni örnek yüzler üretmektir (Goodfellow vd., 2014).

Şekil 3.4 Yüz sentezi ile oluşturulan resim

3.2.3.2 Yüz değişimi

Yüz takası ismiyle de bilinen bu uygulama deepfake teknolojisinin ortaya çıktığı ilk uygulama biçimidir. Şekil 3.5'te gösterilen yüz değişimi uygulamalarında hedef kişisinin yüzü alınarak kaynak kişisinin yüzünün üzerine bindirilir. Çıktı olan medyada kaynak kişisinin ifadeleri ve arka planı bulunmakta iken yüzü tamamen hedef kişinin yüzü ile değiştirilmiştir.

Günümüzde 2 farklı şekilde yüz değişimi yapılmaktadır. Bunlar; geleneksel Bilgisayarlı Görü (CV – Computer Vision) tabanlı yöntemler (FaceSwap v.b.) ve daha karmaşık mimarilere sahip derin öğrenme tabanlı yöntemler (Deepfake). Yüz değiştirme uygulamaları birçok alanda kullanılmakla birlikte genellikle film sektörü, maddi dolandırıcılık, reklam sektörü gibi alanlarda daha fazla örneğine rastlanmaktadır (Dang ve Nguyen, 2023).

Şekil 3.5 Yüz değişimi ile oluşturulan resim

3.2.3.3 Yüz yeniden canlandırma

Literatürde ifade değişimi olarak da bilinen bu yüz değiştirme biçiminde hedef kişisinin kimliği ve yüz özellikleri korunarak kaynak kişisinin yüz ifadelerini hedef yüze aktararak yeni bir çıktı oluşturmaktadır. Burada yüz değişiminden farklı olarak çıktı medyasında hedef kişisinin arka planı ve yüzü kullanılırken kaynak kişisinin sadece ifadeleri kullanılmaktadır. Yüz yeniden canlandırma ile ilgili bir örnek Şekil 3.6'da verilmiştir.

Yüz yeniden canlandırmada temel amaç hedef kişisinin aslında hiç olmayan konuşmaları yaptığı medyalar oluşturmaktır. Günümüzde sahtecilik, dolandırıcılık ve siyasi alanlarda çokça karşımıza çıkan bu tür, kötü niyetli insanların kullanımı sonucunda maddi ve manevi çok büyük zararlar oluşturabilen bir uygulama haline gelebilmektedir.

Rusya – Ukrayna savaşı sırasında iki ülkenin liderlerinin konuşmaları üzerine birçok kez uygulanan yüz değiştirme yöntemi ile liderlerin aslında hiç yapmadığı, savaş ile ilgili tehdit ve teslim olma hakkındaki konuşmalara ilişkin sahte videolara, internet sitelerinde yaygın olarak karşılaşılmaktadır. Yine eski ABD başkanları; Barack Obama ve Donald Trump'ın yüz yeniden canlandırma tekniğiyle oluşturulmuş videoları ve bazı ünlü ve sanatçıların bu teknikle oluşturulmuş onlarca videosuna rastlamak mümkündür.

Şekil 3.6 Yüz yeniden canlandırma ile oluşturulan resim

3.2.3.4 Yüz niteliği değişimi

Yüz niteliği değişimi gerçek bir yüz görüntüsünün anlamsal niteliklerini değiştirmek için kullanılır. Burada hedef kaynak yüzün korunarak hedef özniteliğin yani değiştirilmek istenen niteliğin çıktı yüz görüntüsüne doğru bir şekilde aktarılmasıdır (G. Yang vd., 2021).

Şekil 3.7'de gösterildiği gibi genellikle kişilerin göz rengi, saçları, yaşı, cinsiyeti gibi özelliklerinin değişimi olarak karşımıza çıkmaktadır. Özellikle sosyal medya programlarında kamera filtreleri ile kullanımı sıkça rastlanılan bir durumdur.

Şekil 3.7 Yüz niteliği değişimi ile oluşturulan resim

3.3 Deepfake Oluşturmada Kullanılan Yöntem ve Araçlar

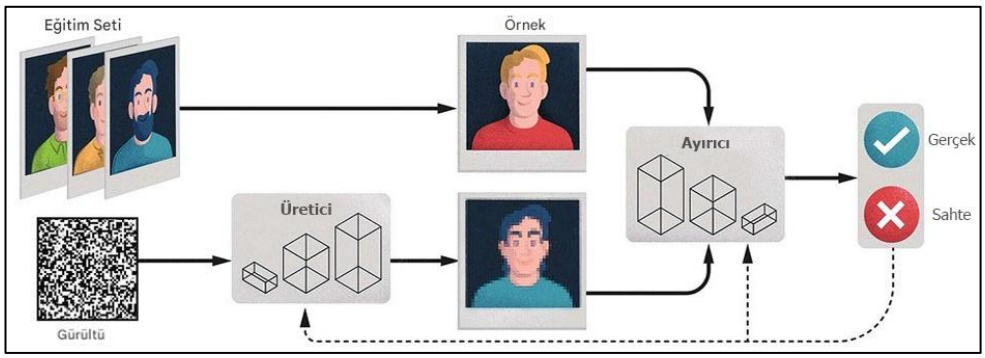
Deepfake oluşturmak için kullanılan birçok mobil uygulama, online web sitesi, masaüstü yazılımı, açık kaynaklı yazılım ve bunları oluşturmak için derin öğrenme ve görüntü işleme tabanlı mimariler bulunmaktadır.

3.3.1 Deepfake Oluşturma Yöntemleri

Deepfake medyaları (video, fotoğraf, ses) oluşturmak için kullanılan 2 temel yöntem vardır. Bunlar; Varyasyonel Otomatik Kodlayıcılar (VAE - Variational Autoencoder) ve Çekişmeli Üretici Ağlar (GAN – Generative Adversarial Nets)'dır (Maksutov vd., 2020).

3.3.1.1 GAN ile deepfake oluşturma

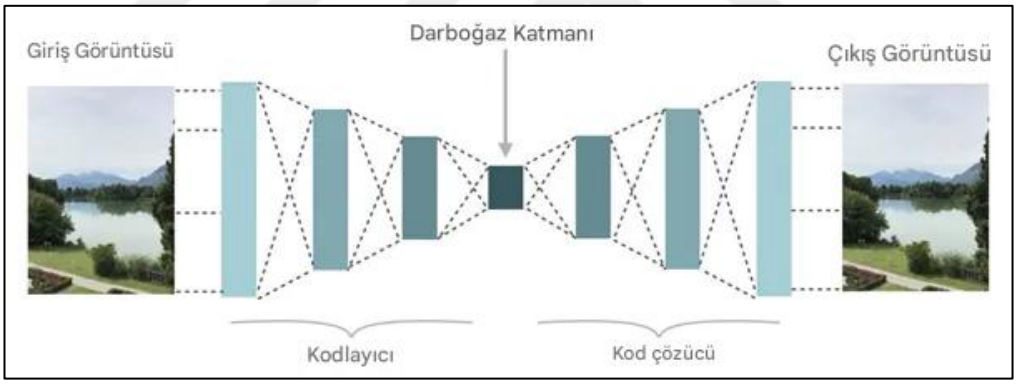
GAN, temelde 2 farklı sinir ağının birleşiminden oluşur. Bunlar üretici (generator) ve ayırıcı (discriminator) ağlardır. GAN ile deepfake oluşturma süresi Şekil 3.8'de verilmiştir. Üretici, verileri işleyip yeni örnekler üreterek ayırıcıya gönderirken ayırıcı bunları gerçek ya da sahte diye ayırmaya çalışır (Goodfellow vd., 2020). Ayırıcının cevabına göre üretici tekrar verileri işleyerek ayırıcıya gönderir. Bir döngü şeklinde üreticinin ayırıcıyı kandırma çabası devam ederek mükemmel sonuca ulaşılmaya çalışılır. Bu ağın eğitimi zaman maliyeti açısından pahalı bir mimari olmakla birlikte sonuçta insan gözüyle ayırt edilemeyecek kadar gerçekçi sahte görüntüler sunmaktadır. Faceswap-GAN(Shaonlu, 2018), CycleGAN altyapısını kullanan açık kaynaklı bir deepfake yazılımıdır.



Şekil 3.8 GAN ile deepfake oluşturma süreci(Moğulkoç, 2024)

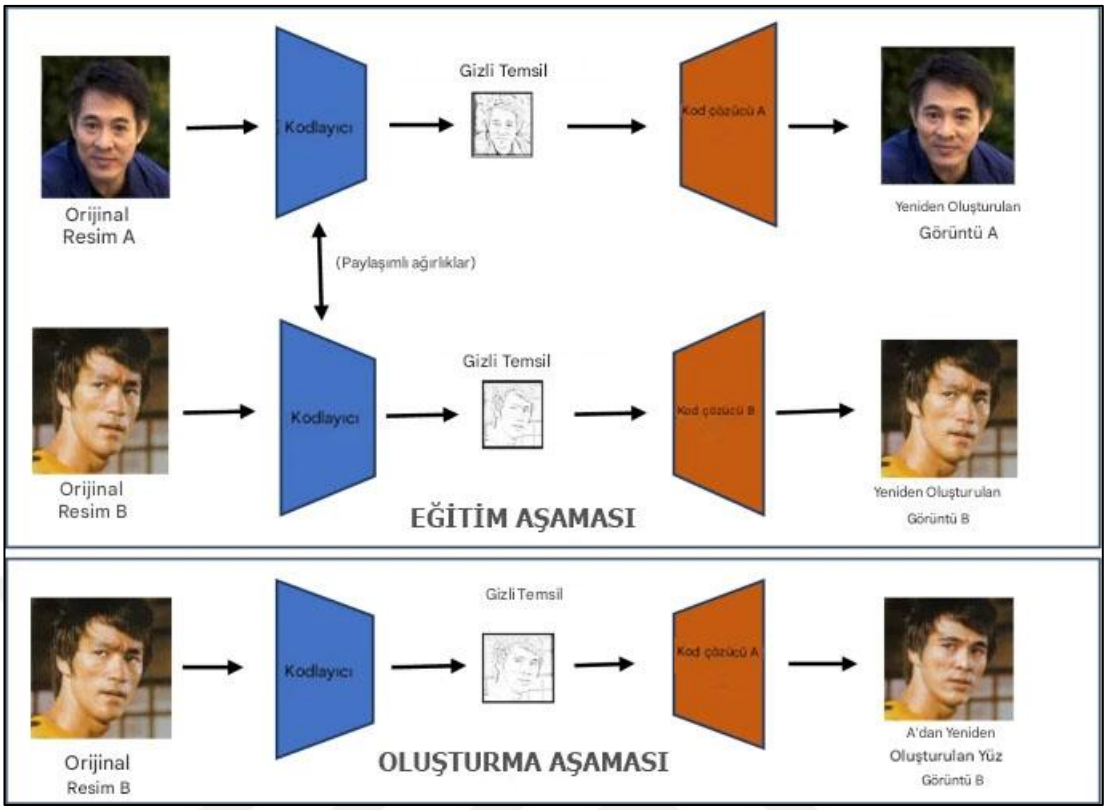
3.3.1.2 VAE ile deepfake oluşturma

Otomatik Kodlayıcı (AE – Autoencoder)'lar girdi olarak verilen veriyi sıkıştırarak otomatik olarak koda dönüştürmeyi ve en az kayıpla tekrar üretmeyi amaçlar. Kodlayıcı (encoder) ve kod çözücü (decoder) olmak üzere iki parçadan oluşan bu mimaride eğitim aşamasında kodlayıcı ve kod çözücü beraber eğitilir. Kodlayıcı, girdi olarak verilen veriyi bottlenecek layer (darboğaz katmanı) denilen bir katmanda sıkıştırarak bir kod (latent vector) oluşturur. Burada üretilen kod ile girdi verisi her eğitimde farklıdır ve arasındaki ilişki bilinmezdir (Öngün C, 2020). Kod çözücü ise bu kodu kullanarak orijinal verileri tekrar oluşturur ve kodlayıcıyı eğitim verilerinden anlamlı temsiller öğrenmeye zorlar (Kingma & Welling, 2019). AE'ler öznitelik çıkarımı, boyut azaltma, gürültü giderme, resim renklendirme ve tamamlama problemlerinde sıkça kullanılmaktadır. Otomatik kodlayıcıların işleyiş şeması Şekil 3.9'da verilmiştir.



Şekil 3.9 Otomatik kodlayıcı şeması (Bakır vd., 2024)

VAE, AE'den farklı olarak darboğaz katmanında belirli bir olasılıksal dağılım öğrenip bu dağılımdan rastgele kodlar kullanılarak yeni veriler üretilebilir (Öngün C, 2020). VAE'ler tıpkı GAN'lar gibi gerçekte olmayan sentetik veriler üretmekte sıkça kullanılmaktadırlar. Faceswap, DFaker, DeepFakeLab gibi açık kaynaklı yazılımlar VAE mimarisini kullanmaktadır (Maksutov vd., 2020). VAE ile deepfake oluşturma süreci Şekil 3.10'da verilmiştir.



Şekil 3.10 VAE ile deepfake oluşturma süreci (Dagar & Vishwakarma, 2022)

3.3.2 Deepfake Oluşturma Araçları

Deepfake medyalar oluşturabilmek için farklı mimarileri kullanan çeşitli araçlar bulunmaktadır. Bu araçlar masaüstü yazılımı, açık kaynak kodlu yazılım, mobil uygulama ve online web siteleri olarak karşımıza çıkmaktadır.

3.3.2.1 Masaüstü yazılımları

Deepfake oluşturmak için genellikle en az tercih edilen yazılım çeşididir. Ticari olarak geliştirilen yazılımların yanında ücretsiz olarak sunulan masaüstü deepfake oluşturma araçları bulunmaktadır. Bunlar;

- **Adobe Premiere:** Bu ticari program ücretli olarak kullanıcıya video düzenleme ve efektler eklemesine olanak sağlamaktadır. Ayrıca konuşmayı metne çevirme ve otomatik yeniden kareleme gibi yapay zeka destekli araçlar sunmaktadır. Ses ve video manipülasyonu gerçekleştirmeniz olanak sağlayan gelişmiş bir programdır.

- **Adobe Photoshop:** Adobe firmasının ücretli programlarından birisidir. Fotoğraflar üzerinde her türlü manipülasyon imkanı sunan program aynı zamanda yapay zeka destekli araçlar sunmaktadır.

- **Corel Video Studio:** Corel firmasının ücretli olarak sunduğu bu program ile video, ses ve fotoğraflar üzerinde her türlü düzenleme yapılabilmektedir.

- **PowerDirector 365:** Cyberlink firması tarafından sunulan bu ücretli video düzenleme programı videolar üzerinde yapay zeka destekli düzenlemeler yapmak için kullanılmaktadır.

- **FakeApp:** Deepfake teriminin isim babası olarak kabul edilen "deepfakes" isimli Reddit kullanıcısının oluşturmuş olduğu ücretsiz programdır. Deepfake oluşturmak için kullanılan en popüler yazılımlardan biri olmakla birlikte ününü ilk ortaya çıktığında, ünlü kadın sanatçıların yüzlerini yetişkin içerikli videolar yapmak için kullanmasından almaktadır. Uygulama internet üzerinde farklı kaynaklardan indirilebilmektedir.

Bunlar gibi onlarca yapay zeka destekli video ses ve fotoğraf düzenleme programları bulunmaktadır.

3.3.2.2 Açık kaynak kodlu yazılımlar

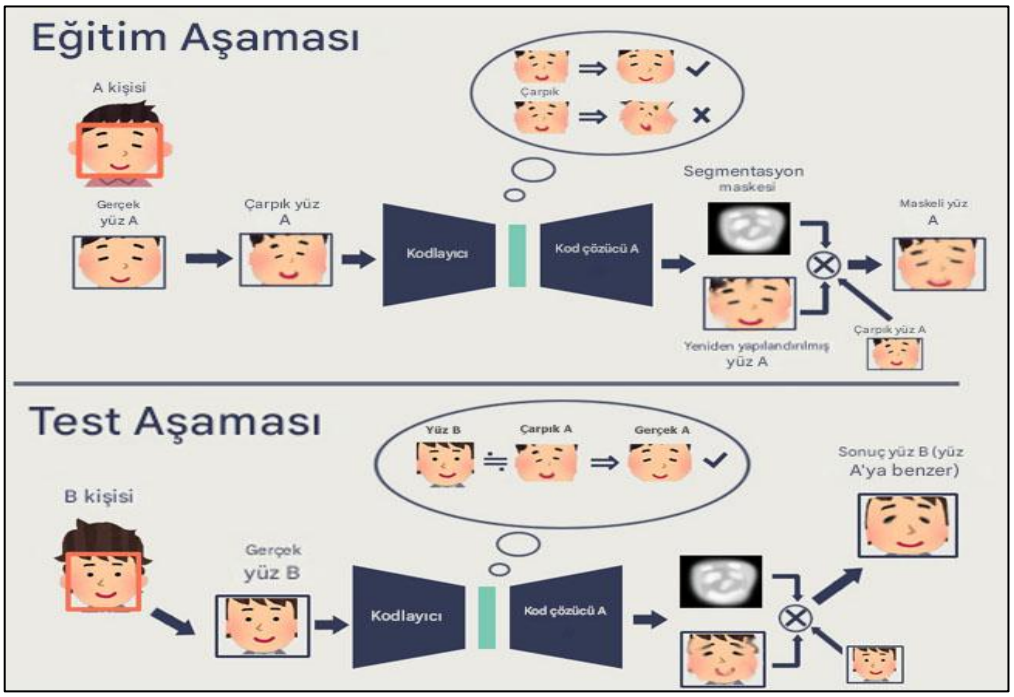
Açık kaynak kodlu yazılımlar, oluşturucuları tarafından yazılımı ve kaynak kodlarının ücretsiz olarak paylaşıldığı yazılımlardır. Genellikle bu yazılımlar GitHub, GitLab, AWS CodeCommit gibi web tabanlı depolama servislerinde halka açık olarak sunulmaktadır. Deepfake oluşturmak için kullanılan en popüler yazılımlar genellikle açık kaynak kodlu yazılımlardır. Bu şekilde onlarca yazılım bulunmakla birlikte en popüler olanları şunlardır;

- **Wav2Lip:** Dudak senkronizasyonu yaparak yüz yeniden canlandırma deepfake türünde sahte medyalar üretilebilen GAN mimarisini kullanan açık kaynak kodlu yazılımdır (Prajwal vd., 2020), (Wav2Lip, 2020).

- **DFaker:** Yeniden canlandırma türünde deepfake yapılmasına olanak sağlayan bir yazılımdır (DFaker, 2018). Keras kütüphanesini kullanan ve yüz yeniden yapılandırma için yüzün etrafındaki değerleri sıfır olarak döndüren ve alakasız özellikleri eğitmeyen bir DSSIM (Difference Structure Similarity Index Method) kaybı ve Maske için Ortalama Kare Hatası (MSE - Mean Squared Error) kullanmaktadır (Alheeti vd., 2021). Uygulama ile yapılan bir deepfake örneği Şekil 3.11'de gösterilmiştir.

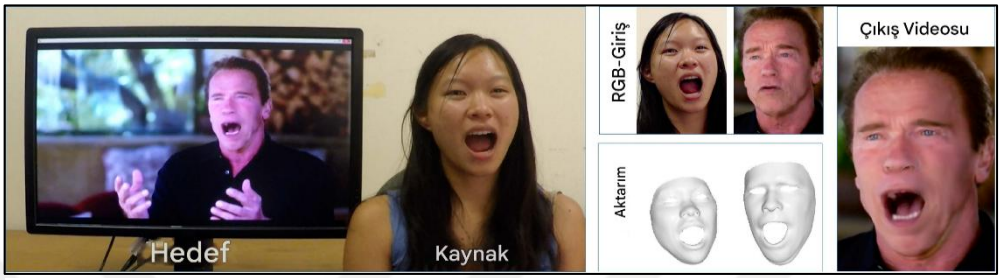
Şekil 3.11 DFaker ile oluşturulan deepfake görüntüsü

- **Faceswap-GAN:** "deepfakes" isimli Reddit kullanıcısının kullanmış olduğu AE mimarisine Adversarial Loss ve Perceptual Loss (VGGface) ekleyerek yeni bir mimari geliştiren deepfake yazılımıdır (Faceswap-GAN, 2018). Şekil 3.12'de Faceswap-GAN ile deepfake oluşturma süreci gösterilmiştir.



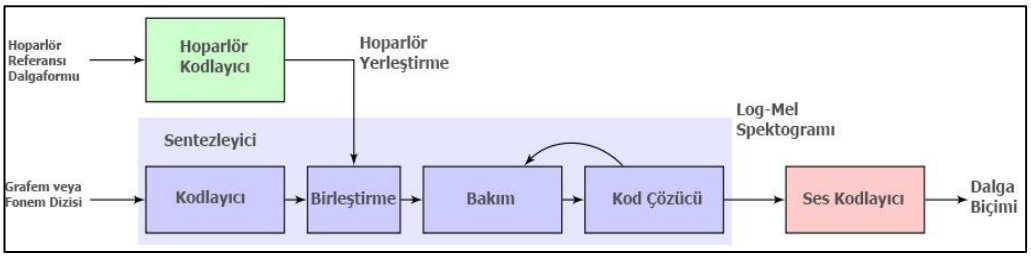
Şekil 3.12 Faceswap-GAN ile deepfake oluşturma süreci

- **Face2Face:** Geliştiricileri arasında Standford Üniversitesinin de bulunduğu, kamera görüntüsü üzerinden gerçek zamanlı yüz yakalama ve RGB videolardan yüz yeniden canlandırma tekniği ile deepfake videolar oluşturmayı sağlayan açık kaynak kodlu yazılımdır (Thies vd., 2016a). Programa ve ilgili makaleye Standford Üniversitesi'nin resmi web sitesi üzerinden ulaşılabilmektedir. Uygulamanın bir örneği Şekil 3.13'de verilmiştir.



Şekil 3.13 Face2Face ile deepfake oluşturma süreci (Thies vd., 2016)

- **SV2TTS:** Geliştiricileri arasında Google'ın da bulunduğu ses klonlama ve metinden sese (TTS - Text-To-Speech) dönüşüm için öğrenme aktarımı yöntemiyle WaveNet derin sinir ağını kullanan yazılımdır (Jia vd., 2018). SV2TTS ile konuşma sentezi oluşturma süreci Şekil 3.14'te verilmiştir.



Şekil 3.14 SV2TTS ile konuşma sentezi oluşturma süreci (Theiler, 2019)

3.3.2.3 Mobil uygulamalar

Google Play Store, Apple App Store, Aptoide gibi mobil uygulama indirme sitelerinde indirme sayıları milyonları aşan onlarca deepfake oluşturma uygulamaları bulunmaktadır. Genellikle bulut tabanlı çalışan bu uygulamalardan en popüler olan ve en çok indirilenlerden bazıları şunlardır;

- **ReFace:** *Neocortext* firmasının sunduğu bu program AI teknolojini kullanarak yüz rötuşlama, yüz değiştirme, fotoğraf hareketlendirme, ses değiştirme gibi

birçok özelliği bünyesinde barındıran ve 100 milyondan fazla indirmesi olan uygulamadır (Reface, 2024).

- **FaceApp:** *FaceApp Technology Ltd* şirketi tarafından sunulan ve 500 milyondan fazla indirmesi bulunan bu program, yüz niteliği değiştirme türünde deepfake videolar oluşturmak için oldukça popüler bir uygulamadır (FaceApp, 2024).

- **FaceHub:** *Creative Hive* firmasının sunmuş olduğu yüz değiştirme türünde deepfake videolar oluşturmaya yarayan mobil uygulamadır (FaceHub, 2022).

- **Deepfake Studio:** *Deep Work* firması tarafından sunulan bu uygulama ile yüz değiştirme türünde deepfake videolar oluşturulabilmektedir (Deepfake Studio, 2024).

- **ZAO:** Çin menşeili *Momo* firması tarafından sunulan ve oldukça popüler olan mobil uygulamadır (Zao, 2019).

3.3.2.4 Online web siteler

Ses ve görüntü medyaları üzerinde farklı deepfake medya oluşturma tekniklerinin uygulanmasını kolaylaştıran, bulut tabanlı onlarca web sitesi bulunmaktadır. Bunlardan bazıları;

- **wavel.ai:** Ses ve görüntü üzerinde ücretsiz olarak çevrimiçi deepfake videolar yapılmasını sağlayan web sitesidir (Wavel AI, 2024).

- **deepfakesweb.com:** Videolarda yüz değiştirme türünde deepfake yapılabilen ücretli bir web sitesidir (Deepfakesweb, 2024).

- **voicery.com:** Yapay zeka destekli, TTS çevirme uygulamaları yapılabilen web sitesidir (Voicery, 2024).

3.4 Deepfake Tespitinde Kullanılan Yöntem ve Araçlar

Bilgisayar donanımları, teknolojik gelişmeler eşliğinde hızlı bir şekilde gelişmektedir. Bununla birlikte güçlü donanımlara ihtiyaç duyan karmaşık yazılımlar ortaya çıkmaktadır. Deepfake videolar bu karmaşık yazılımların en önemli örneklerindendir. Çünkü deepfake medya oluşturabilmesi için güçlü donanımı olan bir bilgisayara ihtiyaç duyulmaktadır.

Görsel ve işitsel medyalar üzerinde yapılan manipülasyonlar, önceleri adli bilişim uzmanları tarafından, geleneksel yöntemler kullanılarak yapılmıştır. Yapay zekanın hızlı gelişimi ile birlikte GAN, VAE gibi derin öğrenme yöntemleri ile yapılan deepfakeler, geleneksel adlı bilişim teknikleriyle tespit edilemez hale gelmiştir.

Deepfake tespit yöntemlerinden bahsetmeden önce bu tespitlerin kullanımına ilişkin yaklaşımlardan bahsedilmesinin daha doğru olacağı düşünülmektedir. Bu yaklaşımları şu şekilde sıralanabilir;

- **Uçtan Uca (End-to-End) Yaklaşım:** Girdi olarak verilen veri ile çıktı verisinin, birbiriyle bağlantılı ve etkileşimli parçalar halinde geliştirildiği, tüm sürecin tek bir sistem içinde entegre bir şekilde çalıştığı yaklaşımdır. Girdi olarak bir video alındığında, model bu videonun içindeki yüzleri ve dudak hareketlerini otomatik olarak analiz eder, özellikleri çıkarır ve ardından bu bilgiyi kullanarak videonun gerçek mi yoksa sahte mi olduğunu belirler. Bu yaklaşım derin öğrenme modelleri kullanılarak geliştirilen yöntemlerde en çok kullanılan yaklaşımdır.

- **Hibrit Yaklaşım:** Bu yaklaşımda farklı deepfake tespit teknikleri bir arada kullanılarak sonuca ulaşılmaya çalışılmaktadır. Derin öğrenme yöntemleri kullanılarak deepfake tespiti yaparken farklı aşamalar bulunmaktadır. Bunlar; veri toplama, veri işleme, özellik çıkarımı, model eğitimi ve testi, sonuçların yorumlanması ve geribildirim döngüsü. Bu aşamaların modüler bir şekilde farklı tespit teknikleri kullanarak gerçekleştirmesi hibrit yaklaşıma örnektir. Uçtan uca yaklaşım kadar tercih edilmese de literatürde örnekleri bulunmaktadır. Örnek olarak resimlerde bulunan yüz görüntülerin, geleneksel yöntemlerle belirlenip

kaydettikten sonra derin öğrenme yöntemleri ile deepfake tespiti yapılması hibrit bir yaklaşımdır.

- **Birleştirme (Ensemble) Yaklaşım:** Birden fazla yöntemin çıktılarının birlikte yorumlanarak karar verilmesine dayanan yaklaşımdır. Örnek olarak baş pozisyonunu dikkate alan bir teknik ve dudak senkronizasyonunu kullanan bir teknik kullanıldığında bu iki tekniğin vermiş olduğu çıktıların beraber değerlendirilerek bir karar verilmesi ensemble yaklaşımdır. Zaman maliyeti açısından dezavantajlı bir yaklaşım olduğu için fazla tercih edilmemektedir.

3.4.1 Deepfake Tespit Yöntemleri

Deepfake medyaların tespiti için birden fazla yöntem bulunmaktadır. Daha önce deepfake hakkında yapılan çalışmalar incelendiğinde bu yöntemlerin farklı isimlerle ve farklı gruplarda incelendiği görülmektedir. Bu yöntemler tekil olarak kullanılabildiği gibi hibrit olarak kullanımı da mümkündür.Mevcutta kullanılan yöntemler çoğunlukla temel özellikleri hedef almaktadır. Tespit yöntemlerini 5 ana başlık altında toplamak mümkündür (Yu vd., 2021).

3.4.1.1 Genel ağ tabanlı yöntemler

Genel ağ tabanlı yöntemler, literatürde bütünsel yöntemler olarak da karşımıza çıkan genellikle uçtan uca yaklaşım kullanılarak yapılan ve çok fazla tercih edilen deepfake tespit yöntemidir. Bu yöntemde mekânsal özellikler analiz edilmek için yüz görüntüsü içeren kareler çıkarılarak bu kareleri analiz etmek ve sınıflandırmak üzere Evrişimli Sinir Ağları (CNN - Convolutional Neural Networks) kullanılır. Görüntülerin CNN kullanılarak sınıflandırması görevinde 2 tür yaklaşım bulunmaktadır.

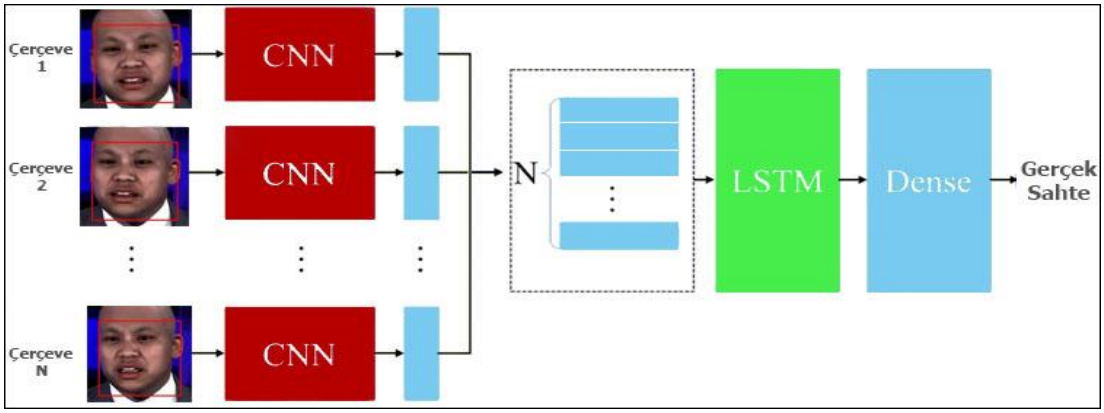
- **Özel tasarlanmış ağlar:** Deepfake tespit görevinde verilerin CNN kullanılarak sınıflandırılması için literatürde bazı çalışmalarda araştırmacılar kendi ağlarını sıfırdan oluşturmuşlardır. (Afchar vd., 2018), görüntülerin mezoskopik özelliklerine odaklanarak Mesonet isimli bir ağ oluşturmuşlardır. (H. H. Nguyen vd., 2018), mevcut ağların performansını iyileştirmek için bir kapsül ağı geliştirmişlerdir.

- **Öğrenme aktarımı (Transfer Learning) kullanımı:** Öğrenme aktarımı, daha önce başka veriler üzerinde eğitilerek ağırlıkları kaydedilen bazı ağların (XceptionNet, InceptionNet, ExceptionNet, VGG16 v.b.) ağırlıklarının yeniden kullanılarak verilerin sınıflandırmasında kullanılması yaklaşımıdır.

Genel ağ tabanlı yöntemler deepfake tespit görevinde sıklıkla kullanılmakta olup zaman maliyeti açısından avantajlı modeller olsalar da belirli veri kümelerine aşırı uyum (overfitting) sağlamaktadır. Aşırı uyum, bir modelin, eğitim için kullanılan verileri doğru sınıflandırırken yeni verilerde aynı başarıyı sağlayamaması durumudur. Genel ağ tabanlı yöntemlerin dezavantajı da eğitildikleri deepfake oluşturma türü haricinde bir türle oluşturulan deepfake medya içeren veri setlerinde başarı oranlarının düşük olmasıdır.

3.4.1.2 Zamansal tutarsızlık tabanlı yöntemler

Genel ağ tabanlı yöntemlerde videolardaki yüz içeren kareler belirli aralıklarla ya da rastgele seçilerek kaydedilmektedir. Araştırmacılar bu yaklaşımın zamansal tutarsızlıkları tespit edemeyeceği için ardışık kareler arasındaki sürekliliği kontrol etmek ve tutarsızlıkları belirlemek üzere Tekrarlayan Sinir Ağları (RNN - Recurrent Neural Network)'nı kullanmaya başlamışlardır.



Şekil 3.15 Zaman serisi analizi ile deepfake medya tespiti

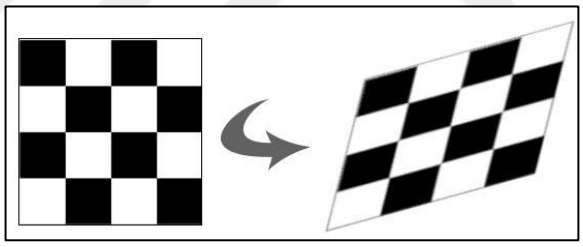
Şekil 3.15'te iş akışı verilen bu tekniğin genel yapısı ardışık karelerin özniteliklerini çıkarmak için CNN kullanılması ve zamansal dizi analizi için RNN'in bir türü olan Uzun-Kısa Süreli Bellek (LSTM – Long-Short Term Memory) kullanılmasını öngörmektedir. Genel ağ tabanlı yaklaşımlara göre bitişik kareler arasındaki

tutarsızlıkları kullanarak başarıyı arttıran bu modeller orijinal karelerdeki mekânsal özellikleri yakalayamazlar (Yu vd., 2021).

3.4.1.3 Görsel yapay eser (artefakt) tabanlı yöntemler

Deepfake videoların temel prensibi hedef görselde bulunan yüzün kaynak videoda bulunan yüz ile değiştirilmesidir. Sonuçta ortaya çıkan videoda bulunan yüzler hedefe ait iken arka plan kaynak videoya aittir. Bu işlemler sırasında yüz görüntüsü ile arka plan arasında sınır anomalisi ve renk tonu farklılıkları oluşmaktadır. Bu farklılıklar artefakt yani yapay eser olarak adlandırılmaktadır. Görsel yapay eser tabanlı yöntemlerde bu anomalileri kullanarak deepfake tespiti yapılmaktadır.

- **Yüz çarpıtma yapay eserleri:** Basit deepfake videolarda hedef yüz görüntüsü kaynak videoya aktarılırken afin dönüşüme tabi tutulur. Afin dönüşüm, geometrik dönüşümler arasında çizgi, paralellik ve mesafeleri koruyan, çevirme ölçekleme dönme gibi işlemlerin yapıldığı dönüşümdür. Afin işlem sonrası geometrik dağılım bozulmaktadır (Feizabadi vd., 2015).

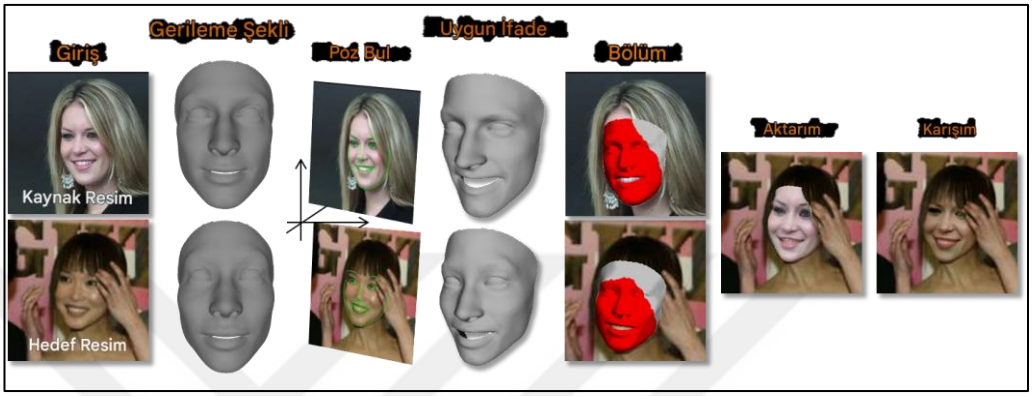


Şekil 3.16 Afin dönüşüm (Rayaguru, 2023)

Deepfake videolarda afin dönüşüm ile yüz takasının yapılması sonucunda yüz ile arka plan arasında renk farkları ve çözünürlük tutarsızlıkları oluşmaktadır. Bu yapay eserleri tespit ederek deepfake tespiti yapılmaktadır.

Şekil 3.17 Yüz çarpıtma tutarsızlıkları

Yüz çarpıtma sonucu oluşan yapay eserler (artefact) Şekil 3.17'de açıkça görülmektedir. Teknolojik gelişmelerle birlikte Deepfake videolarda afin dönüşüm yerine Şekil 3.18'de gösterilen yüz işaretlerine dayalı dışbükey bir maske oluşturularak yüz takası gerçekleştirilmekte ve sınırlardaki renk tonları eşleştirilmekte olduğundan bu teknikler bu tür Deepfake videolarını doğru sınıflandırma konusunda başarılı olamamaktadır.



Şekil 3.18 3D yüz değişimi (Nirkin vd., 2018)

- **Baş pozisyonu tutarsızlığı:** Bu teknikte kaynak videoda manipüle edilen alanın 3 boyutlu dönüm noktaları ile hedef yüzün 3 boyutlu dönüm noktaları arasındaki farka odaklanılmaktadır. Hedef yüz kaynak videoya aktarılırken yüzün pozisyonunun belirlenmesinde sadece yüzün merkez dönüm noktalarının kullanıldığı ve kaynak videoda bulunan kafanın tamamının 3 boyutlu dönüm noktalarının kullanıldığı varsayılmaktadır. Hedef ve kaynak yüz arasındaki baş pozisyonları vektörleri arasındaki farklar hesaba katılarak sahte ve gerçek videolar sınıflandırılmaktadır.

3.4.1.4 Dijital parmak izi (finger print) tabanlı yöntemler

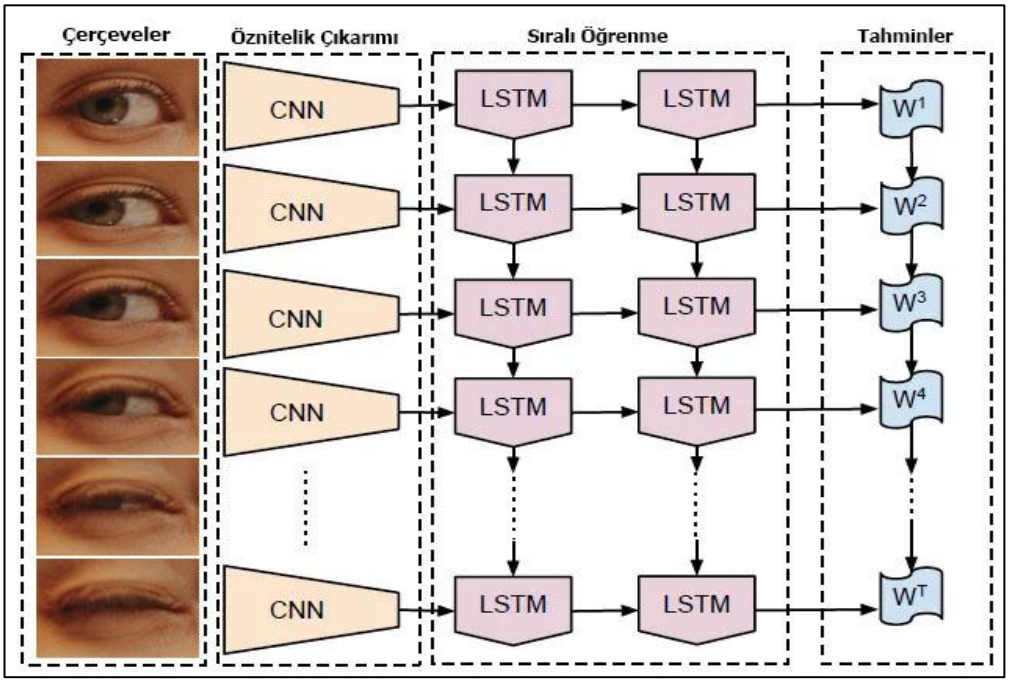
- **Kamera parmak izi (PRNU):** Kamera ve diğer optik cihazlarda bulunan dijital görüntü sensörlerine ait benzersiz bir gürültü bileşeni olan PRNU (Photo Response Non-Uniformity) (Vatansever & Dirik, 2023), benzersizliği nedeniyle deepfake algılama çalışmalarında kullanılan bir cihaz parmak izidir (Chen vd., 2008). Bu teknikler GAN tarafından oluşturulan deepfake türlerinde fazla başarı gösterememektedir.

- **Video gürültü desenleri:** Deepfake videolarda hedef yüzün bulunduğu medya ile kaynak yüzün bulunduğu medyanın farklı şekilde oluşturuldukları ve içinde bulunan gürültülerinde farklı olacağı görüşüne dayanan deepfake tespit yöntemidir (Yu vd., 2021).

3.4.1.5 Biyolojik sinyal tabanlı yöntemler

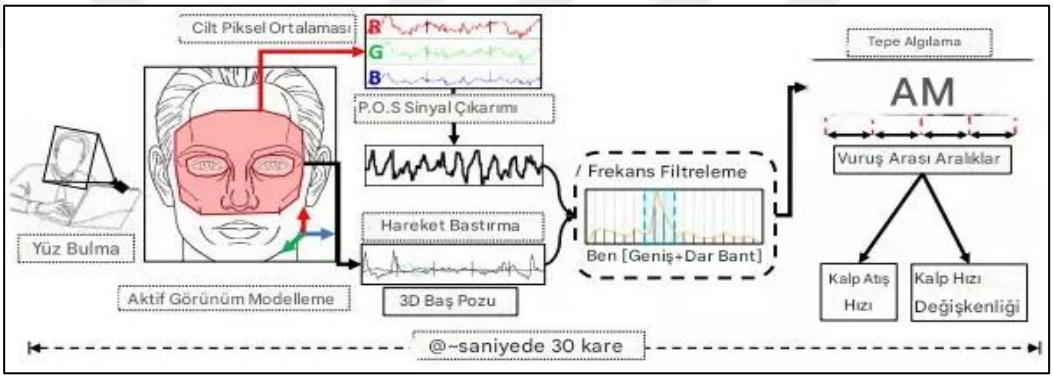
Göz kırpma, kalp atış hızı gibi insanların belli fizyolojik sinyallerinin RNN kullanılarak tespitine dayalı yöntemlerdir.

- **Göz kırpma:** Doğal olmayan göz hareketleri ve göz kırpma hareketinin hiç olmaması deepfake video olma ihtimalini artırmaktadır. Göz kırpma sıklığı önceki deepfake tespit çalışmalarında ayırt edilebilir bir özellik olarak kabul edilmiştir (Li vd., 2018b). Bir insanın ortalama göz kırpma süresinin 100-400 ms olduğu varsayımı kullanılarak videolardaki göz kırpma hareketleri arasındaki süreler dikkate alınmaktadır. Şekil 3.19'da gösterilen gözün açık ve kapalı durumlarını analiz etmek için CNN ve LSTM'i birlikte kullanan Uzun Dönemli Tekrarlayan Evrişimli Ağlar (LRCN - Long-term Recurrent Convolutional Networks) kullanılmaktadır (Donahue vd., 2015).



Şekil 3.19 LRCN ile göz kırpma tabanlı deepfake tespiti (P. Wang vd., 2018)

- **Kalp atış hızı:** Kalp atış hızı (HR - Heart Rate) kalbin dakikadaki atış sıklığı olarak bilinir. Kan akış hızını ölçmeye yarayan cihazlara ise Fotopletismogram (PPG - Photoplethysmogram) denilmektedir. PPG'lerin kamera yardımıyla uzaktan kalp atış hızını tespit edebilen versiyonlarına Uzaktan Fotopletismogram (rPPG) denilmektedir. rPPG'ler webcam, termal kamera yada RGB kameraları kullanarak uzak PPG sinyali oluşturmak için ciltteki renk değişikliklerini tespit edebilmektedirler (Xiao vd., 2024). Deepfake videoların tespitinde, rPPG cihazı kullanarak oluşturulan bu sinyallerin, Şekil 3.20'de genel görünümü verilen sahte videolardaki zaman-mekan tutarsızlığının ortaya çıkartılması hedeflenmiş ve başarılı modeller oluşturulmuştur (Feng vd., 2014).



Şekil 3.20 Önerilen kalp hızı değişkenliği tahmin hattı genel görünümü(Karthick vd., 2023)

3.4.1.6 Deepfake tespit çalışmalarında karşılaşılan zorluklar

Deepfake uygulamalarının kötü niyetli insanlar tarafından kullanılmaya başlaması sonucunda deepfake tespit çalışmalarının önemi ortaya çıkmıştır. Deepfake tespiti için global teknoloji şirketleri büyük yatırımlar yaparak kendi çözümlerini sunmakta, akademik çalışmalar sonucunda ortaya çıkan çözümler literatürde yerini almakta, bireysel kullanıcıların kendi çabaları ile geliştirdikleri uygulamalar çeşitli platformlarda açık kaynaklı olarak kullanıma sunulmakta, yine büyük teknoloji şirketleri tarafından ödüllü yarışmalar düzenlenmektedir. Bu çalışmalar sırasında karşılaşılan bazı zorluklar bulunmaktadır. Bu zorlukları farklı başlıklar altında sıralamak mümkündür.

- **Veri seti kaynaklı zorluklar:** Deepfake tespit çalışmalarının en önemli argümanı şüphesiz ki üzerinde çalışmak üzere ihtiyaç duyulan veri setleridir. Medyalar

üzerinde bulunan manipülasyonları tespit etmek için geliştirilen uygulamalar, bu medyalar üzerinde yapılan deepfake türü ile daha önce oluşturulmuş medyaların bulunduğu veri setlerine ihtiyaç duymaktadır. Mevcut veri setleri üzerinde yapılan çalışmalar, sadece o türde yapılan deepfake medyalarını tespit etmekte iken deepfake yapmak için geliştirilen yazılımlar, bu çalışmalarda tespit edilen eksiklikleri gidermiş olduğu için güncel medyalar üzerinde başarı oranları büyük oranda azalmaktadır. Mevcutta kullanılan güncel veri setleri genellikle Alphabet, Meta gibi global teknoloji şirketleri tarafından sunulmaktadır. Bireysel ve akademik çalışmalar sonucunda oluşturularak paylaşılan veri setleri ise genellikle dengesiz veri dağılımı, etiketsiz veriler ve eski deepfake oluşturma algoritmaları tarafından oluşturulan sahte medyalar içerdiğinden dolayı bu veri setleri kullanılarak yapılan çalışmalar mevcut verilere aşırı uyum (overfitting) sağlama sorunu ile karşı karşıya kalmaktadır.

- **Teknolojik altyapı kaynaklı zorluklar:** Deepfake tespit çalışmalarının ihtiyaç duyduğu diğer bir argüman ise bu çalışmaların yürütüleceği teknolojik altyapıdır. Bu çalışmaların standart bilgisayarlarla yapılması mümkün değildir.

  Deepfake tespitinde kullanılan derin öğrenme algoritmalarının başarı oranı, kullanılan verilerin büyüklüğü ve uzun süren eğitim süreleri ile doğru orantılıdır. İçerisinde bin tane görsel veya işitsel veri olan bir veri seti ile birkaç saatlik eğitim süresi sonunda ortaya çıkan algoritmaların güncel deepfake medyalar üzerinde başarı sağlaması mümkün değildir. Bireysel ve akademik çalışmalarda Google Colab, Kaggle gibi bulut tabanlı Jupyter Notebook servislerin yaygın olarak kullanıldığı görülmektedir. Bunun başlıca sebebi bu platformların gelişmiş CPU, GPU ve TPU kullanımına erişim vermesidir. Google Colab bu platformlar arasında en popüler uygulama olmasına rağmen avantajlı ve dezavantajlı yönleri bulunmaktadır.

  Ülkemizde ise bu platformlara alternatif olarak 2003 yılında Türkiye Bilimsel ve Teknik Araştırma Kurumu (TÜBİTAK) tarafından temelleri atılan, hesaplama ve veri depolama kaynağı olarak kullanabilen Türk Ulusal Bilim e-Altyapısı (TRUBA), araştırma kurumları ve araştırmacıların kullanımına sunulmuştur

(TRUBA, 2003). Deepfake tespitine ilişkin çalışmaların artması için bu tarz altyapıların sayılarının artması, mevcut platformların ise araştırmacılara ücretsiz ve herhangi bir kısıtlama olmadan sunulması gerekmektedir.

3.4.1.7 Deepfake tespit araçları

Deepfake tespit araçları; global teknoloji şirketleri, siber güvenlik firmaları, akademik ve bireysel araştırmacılar tarafından oluşturulan hizmet olarak yazılım (saas - software as a service), masaüstü yazılımları, mobil uygulama, online web siteleri ve açık kaynak kodlu yazılımlar olarak sunulan araçlardır.

3.4.1.7.1 Hizmet olarak yazılım (Saas - Software as a Service)

Tüketicilerin, bulut altyapısında çalışan ve çeşitli platformlardan sağlayıcının uygulamalarına erişim sağlayabildiği yazılımlara saas denilmektedir (Cloud, 2011). Genellikle siber güvenlik üzerine çalışan firmalar müşterilerine saas yöntemi ile destek sağlamaktadır. Farklı platformlardan erişim sağlanabildiği için örnekleri diğer başlıklar altında verilmiştir.

3.4.1.7.2 Masaüstü yazılımları

Bu yazılımlar genellikle bulut tabanlı uygulamalara erişim sağlanabilmesi için masaüstü uygulamaları sadece arayüz olarak kullanan yazılımlardır.

- **Oz Liveness:** Oz Forensics firmasının çok platformlu, yüz tanıma ve kimlik doğrulama yazılımıdır. Uygulama Linux, Android, İphone gibi birçok platformdan erişim sağlanabilen saas tabanlı bir yazılımdır.

- **Sumsub:** Sumsub firmasının sunmuş olduğu Windows, Mac, Android, Linux İphone gibi birçok platformda uyumlu uygulaması bulunan saab tabanlı bir siber güvenlik uygulamasıdır. Sumsub AI destekli yüz doğrulama teknolojisi sahte kimlikleri tespit etmektedir.

3.4.1.7.3 Açık kaynak kodlu yazılımlar

- **DeepFake-o-meter:** UB Media Forensics Lab tarafından geliştirilen, birden fazla tespit algoritması kullanarak deepfake içeriğini analiz etmeye yarayan açık kaynaklı yazılımdır (Li vd., 2021).

- **DeepFake-Detect:** Keras ve Tensorflow kütüphaneleri kullanılarak öğrenim aktarımı yöntemiyle EfficientNet mimarisi kullanılarak geliştirilen ve bir web sitesi arayüzüne sahip açık kaynaklı yazılımdır.

3.4.1.7.4 Mobil uygulamalar

- **Deepware:** Deepware firması tarafından sunulan deepfake videoları tespit edebilen online tabanlı mobil uygulamadır.

- **Phocus:** DuckDuckGoose firması tarafından geliştirilen Phocus, çalışan profillerini yönetmek ve analiz için video veya görüntü yüklemek için mobil uygulaması bulunan saas tabanlı bir deepfake tespit yazılımıdır.

3.4.1.7.5 Online web siteleri

- **deepfakedetector.ai:** DeepFake Detector firması tarafından sunulan ses ve video Deepfake medyalarını tespit edebilen web tabanlı uygulamadır.

- **AI or Not:** Görüntü ve ses dosyalarında bulunan Deepfake tespit eden web tabanlı yazılımdır.

Deepfake tespit uygulamalarının karşılaştırması Tablo 3-1'de verilmiştir.

Tablo 3-1 Deepfake tespit uygulamalarının karşılaştırması.

Yazılım	Kullanım		Medya				Platform Desteği				Kaynak
	Açık	Ücretsiz	Resim	Video	Ses	Metin	Saas	Bilgisayar	Mobil	Web	
Deepware	✓	✓	✗	✓	✗	✗	✗	✗	✗	✓	https://scanner.deepware.ai/
DeepFake-Detect	✓	✓	✓	✓	✗	✗	✗	✗	✗	✓	https://deepfake-detect.com/
Phocus	✗	✗	✓	✓	✗	✗	✓	✗	✓	✗	https://www.duckduckgoose.ai/phocus
WeVerify	✓	✓	✓	✓	✗	✓	✓	✗	✗	✗	https://weverify.eu/verification-plugin/
AI or Not	✗	✗	✓	✗	✓	✗	✗	✗	✗	✓	https://www.aiornot.com/
Oz Liveness	✗	✗	✓	✓	✗	✗	✓	✓	✓	✓	https://ozforensics.com/
Sumsub	✗	✗	✓	✓	✗	✗	✓	✓	✓	✓	https://sumsub.com/
Deepfakedetector	✗	✗	✗	✓	✓	✗	✗	✗	✗	✓	https://deepfakedetector.ai/
Content at Scale	✗	✓	✓	✗	✗	✓	✗	✗	✗	✓	https://contentatscale.ai/ai-image-detector/
AI Voice Detector	✗	✗	✗	✗	✓	✗	✗	✗	✗	✓	https://aivoicedetector.com/
Sensity	✗	✗	✓	✓	✓	✗	✓	✗	✗	✗	https://sensity.ai/deepfake-detection/
Resemble.AI	✗	✓	✗	✗	✓	✗	✓	✗	✓	✗	https://resemble.ai/free-deepfake-detector/
TrueMedia	✓	✓	✓	✓	✓	✗	✗	✗	✗	✗	https://truemedia.org/
FakeCatcher	✗	✗	✓	✓	✗	✗	✗	✗	✗	✓	Intel
Pindrop Pulse	✗	✗	✗	✗	✓	✗	✓	✗	✗	✗	https://www.pindrop.com/deepfake/
Sentinel	✗	✗	✓	✓	✓	✗	✗	✗	✓	✓	https://thesentinel.ai/
DeepFake-o-meter	✓	✓	✓	✓	✓	✗	✓	✗	✗	✗	zinc.cse.buffalo.edu/ubmdfl/deep-o-meter/

3.4.1.8 Deepfake tespitinde kullanılan veri setleri

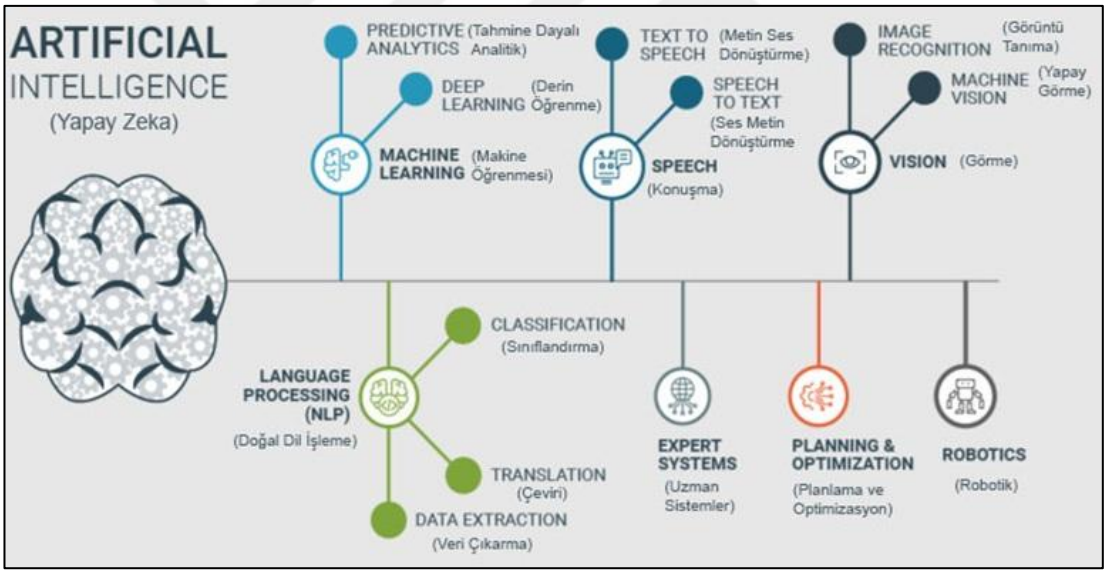
Deepfake tespit çalışmalarının karşılaştığı ortak problem, bu alanda oluşturulan veri setlerinin sayısı ve niteliğinin yetersiz olmasıdır. Bu alanda bir veri seti oluşturmak için kişisel verilerin korunması ile ilgili yasalara dikkat etmek, kişilerin özel izni ya da gönüllü katılımını sağlamak çok önemlidir. Bu zorluklar, veri setindeki benzersiz kişi sayılarının yetersiz kalması nedeniyle bu veri setleri kullanılarak yapılan çalışmaların genelleme yeteneğinin yetersiz kalmasına sebep olmaktadır. Günümüzde deepfake tespitinin artan önemi oranında yeni veri setleri oluşturulmakla birlikte yine de gerekli çeşitliliği sağlayabilen veri seti sayısı yeterli değildir. Deepfake tespit veri setleri içerik olarak video, sesli video, ses ve fotoğraf medyalarından oluşturulmalarına göre ayrı kategorilerde incelenmektedir. Çalışmamızın konusu ses deepfake tespiti olmadığı için sadece videolardan oluşan veri setleri incelenmiştir. Tablo 3-2'de bu alanda sıkça kullanılan veri setleri ve güncel veri setlerinin özeliklerine yer verilmiştir.

Tablo 3-2 Deepfake veri setlerinin karşılaştırması

Yıl	Veri Seti	Video Sayısı		Aktör Sayısı	KAYNAK
		Gerçek	Sahte		
2018	UADFV	49	49	49	https://paperswithcode.com/dataset/uadfv
2018	DF-TIMIT	320	640	32	https://conradsanderson.id.au/vidtimit/
2018	FaceForensics(FF)	1004	2008	-	https://github.com/ondyari/FaceForensics/tree/original
2019	FaceForensics++ (FF++)	1000	4000	977	https://github.com/ondyari/FaceForensics
2019	Deep Fake Detection (DFD)	363	3068	28	https://github.com/ondyari/FaceForensics/tree/master/dataset
2019	Diverse Fake Face Dataset (DFFD)	1000	3000	-	https://cvlab.cse.msu.edu/dffd-diverse-fake-face-dataset.html
2020	DeepFake Detection Challenge (DFDC)	23.564	104.500	3426	https://ai.meta.com/datasets/dfdc/
2020	Celeb-DF v1	408	795	13	https://github.com/yuezunli/celeb-deepfakeforensics/tree/master/Celeb-DF-v1
2020	Celeb-DF v2	590	5639	59	https://github.com/yuezunli/celeb-deepfakeforensics/tree/master/Celeb-DF-v2
2020	Deeper Forensics (DF) 1.0	50.000	10.000	100	https://github.com/EndlessSora/DeeperForensics-1.0
2020	Face Forensics in the Wild (FFIW10K)	10.000	10.000	-	https://github.com/tfzhou/FFIW
2021	ForgeryNet	99.630	121.617	5400+	https://github.com/yinanhe/forgerynet
2021	Korean DeepFake Detection Dataset (KoDF)	62.166	175.776	403	https://deepbrainai-research.github.io/kodf/
2022	FakeAVCeleb	500	19.500	500	https://sites.google.com/view/fakeavcelebdash-lab/
2022	DFDM	590 CelebDF	6450	59	https://github.com/shanface33/Deepfake_Model_Attribution
2022	Localized Audio Visual DeepFake Dataset (LAV-DF)	36.431	99.873	153	https://github.com/ControlNet/LAV-DF
2022	DeePhy	100	5.040	100	https://iab-rubric.org/deephy-database
2023	DF-Platter	764	132.496	454	https://iab-rubric.org/df-platter-database
2023	AV-Deepfake1M	286.721	860.039	2.068	https://github.com/ControlNet/AV-Deepfake1M
2024	DeepSpeak v1.0	6.226	6.799	220	https://huggingface.co/datasets/faridlab/deepspeak_v1
2024	DF40	FF++, CelebDF	100.000+	-	https://github.com/YZY-stack/DF40

4. DERİN ÖĞRENME TANIMI VE TEMEL KAVRAMLAR

Derin öğrenme kavramını anlayabilmek için önce yapay zeka ve makine öğrenmesi gibi kavramların iyi anlaşılması gereklidir. Çünkü bu terimler arasında hiyerarşik bir ilişki vardır. Yapay zeka, en temel anlamda insan zekasını taklit ederek hareket etme, bilgi toplama, konuşma, analiz etme ve bu analizler sonucunda karar verme gibi insan yetilerine sahip olmayı hedefleyen bir yaklaşımdır. Bu yaklaşımın en temel ve geniş tanımı olan yapay zekanın amaçları doğrultusunda zamanla alt türleri de gelişmiştir. Şekil 4.1'de görüldüğü üzere yapay zekanın insanın, görme yetisini taklit eden görüntü tanıma, konuşma yetisini taklit eden konuşma sentezi, hareket yetisini taklit eden robotik, anlama ve bilgi toplama yetisini taklit eden doğal dil işleme ve karar verme yetisini taklit eden makine öğrenmesi gibi farklı alt dalları bulunmaktadır.



Şekil 4.1 Yapay zekanın alt türleri (Rosunee, 2021)

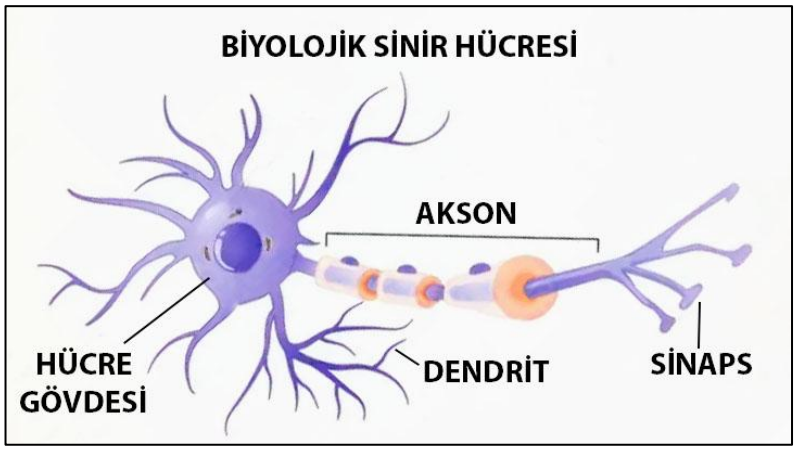
Yapay zekanın alt dalı olan makine öğrenmesi, kendisine verilen her türlü veriyi işleyerek bu verileri yorumlama, sınıflandırma yeteneğine sahip bir disiplindir. Bu disiplinde algoritmaya nasıl öğrenmesi gerektiği açıkça belirtilmez. Algoritma kendisine verilen verileri işleyerek bu verilerdeki kalıp ve korelasyonları bulur, analiz eder ve tahminlerde bulunur (Türkmenoglu & Tantug, 2014).

Makine öğrenmesi algoritmalarının, öğrenme işlemini yapmak için kendisine verilen verilerin etiketli olup olmamasına göre denetimli öğrenme, denetimsiz öğrenme ve pekiştirmeli öğrenme olarak 3 farklı türü vardır. Etiket nitel veri setleri içinde bulunan veri türlerini birbirlerinden ayırmak için kullanılan ve farklı tekniklerle oluşturulan sayısal değerlerdir (Şahinaslan vd., 2023). Hangi verinin hangi türe ait olduğunu belirten etiketli verilerle çalışan algoritmalara denetimli öğrenme, etiketsiz veri ile çalışan algoritmalara ise denetimsiz öğrenme algoritmaları denilmektedir.

Makine öğrenmesinin bir alt disiplini olarak karşımıza çıkan derin öğrenme ise Yapay Sinir Ağları (YSA, ANN – Artificial Neural Network) kullanarak verilerden öznitelik çıkarmayı ve bunları yorumlamayı sağlayan katmanlı ve karmaşık yapıda algoritmalardır. Derin öğrenme algoritmaları denetimli veya denetimsiz olabilirler. Derin öğrenme algoritmalarını anlayabilmek için temelini oluşturan YSA'ları anlamak önemlidir.

4.1 Yapay Sinir Ağları

YSA'lar insan beyninde bulunan sinirlerin çalışma prensibini taklit ederek öğrenme, hatırlama ve ilişki kurma gibi yetiler kazanmayı hedefleyen algoritmalardır.

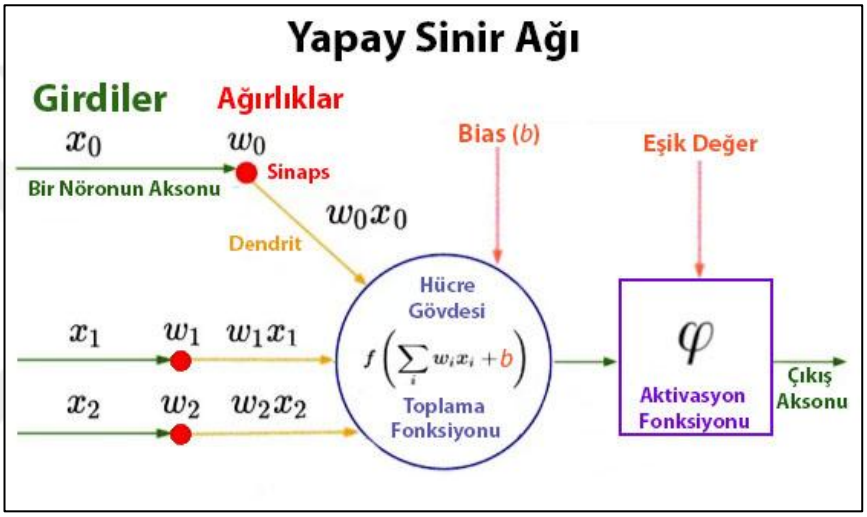


Şekil 4.2 Biyolojik sinir hücresi (Psikolog, 2023)

Şekil 4.2'de verilen insan beyninde bulunan biyolojik sinir hücreleri (nöron) arasında iletilen sinyaller akson aracılığı ile hedef nörona iletilir. Sinyalleri ileten aksonların uç kısımlarına sinaps, sinapstan gelen sinyalleri girdi olarak alan hedef nöronun uç

kısımlarına ise dendrit denilmektedir. Hücre gövdesi ise dendritlerden gelen sinyalleri analog bir yöntemle işlemektedir (Vikipedi, 2024). Dendritler tarafından alınan sinyaller tetikleyici ya da engelleyici olabilir. Sinyaller belirli bir eşik değerini aşarsa akson vasıtasıyla diğer nöronlara sinyal iletilir. İnsan beyninin işleyişi basit anlamda bu şekilde işlemektedir.

YSA'ların işleyişi temel olarak biyolojik sinir hücrelerinin işleyişi ile benzer yapıya sahiptir. Şekil 4.3'te biyolojik sinir hücresi ve yapay sinir ağlarının benzerlikleri gösterilmektedir.



Şekil 4.3 Yapay sinir ağı (Kızrak & Bolat, 2018)

4.1.1 Yapay Sinir Ağlarının Bileşenleri

YSA'ların temel olarak 5 temel bileşeni bulunmaktadır. Bunlar; girdiler, ağırlıklar, toplama fonksiyonu, aktivasyon fonksiyonu ve çıktılardır (Baş, 2006).

- **Girdiler**: Yapay sinir hücrelerine işlenmek üzere gelen verilere girdi denilmektedir. Girdiler ağa dış ortamdan gelen veriler ya da ağın çıkışında aktivasyon fonksiyonu tarafından gönderilen çıktılar olabilmektedir. Girdiler Şekil 4.3'te (X_0,X_1,X_2) olarak gösterilmektedir.

- **Ağırlıklar:** Yapay sinir hücresine girdi olarak gelen verilerin önem oranına göre etkisini belirleyen katsayılara ağırlık denilmektedir. Ağırlıklar aynı zamanda yapay sinir ağlarında bulunan katmanların parametresi olarak

adlandırılabilmektedir (Chollet, 2019). YSA'larda bulunan bütün hücreler arasındaki bağlantılara ait farklı ağırlık değerleri bulunmaktadır. Şekil 4.3'te (W_0,W_1,W_2) olarak gösterilmektedir.

- **Toplama Fonksiyonu**: Toplama fonksiyonu bütün hücrelerden gelen girdi ve ağırlıkların çarpım değerlerini toplayıp sonuca bias (sapma) değerini ekleyerek aktivasyon fonksiyonuna göndermek üzere net girdi değerini oluşturur. Transfer fonksiyonu olarak da adlandırılan bu fonksiyona eklenen bias değerinin temel amacı ise toplama sonucunun 0 olması durumunda bile esnekliği sağlayarak öğrenmenin gerçekleştirilmesidir. Toplama fonksiyonu genel olarak;

$$f\left(\sum X_i W_i\right)+b \quad (1)$$

biçiminde ifade edilir.

- **Aktivasyon Fonksiyonu**: Toplama fonksiyonundan gelen net girdi değerlerini işleyerek belirli bir eşik değerine göre çıktı yapay sinir hücresinin çıktı değerini belirleyen fonksiyona aktivasyon fonksiyonu denilmektedir. Doğrusal (Linear), Adım (Step), Sigmoid, Hiperbolik Tanjant, ReLU, Leaky ReLU, Parameterized ReLU, Swish, Softmax gibi birçok farklı aktivasyon fonksiyonu bulunmaktadır.

- **Çıktılar**: Aktivasyon fonksiyonu tarafından belirlenmiş olan değerlere çıktı değeri denilmektedir. Çıktı değerleri aktivasyon fonksiyonları tarafından genellikle [0,1], [-1,1] [-∞,+∞] gibi aralıklarda belirlenmektedir.

4.1.2 Yapay Sinir Ağlarının Türleri

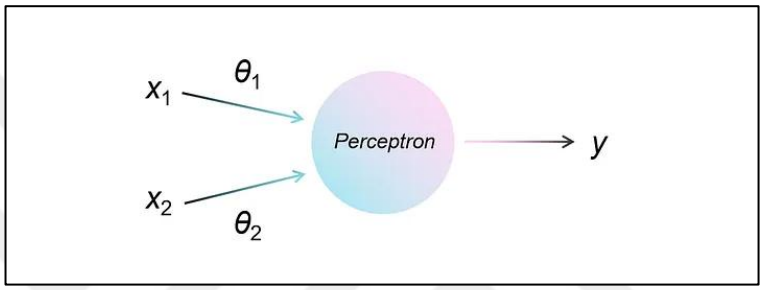
Yapay sinir ağları katmanlar arasındaki bağlantıların yönlerine göre 2 sınıfa ayrılmaktadırlar.

4.1.2.1 İleri beslemeli ağlar

İleri beslemeli ağlarda yapay sinir hücreleri giriş katmanından çıkış katmanına doğru düzenli olarak sıralanır ve sinyaller sadece tek yönlü olarak hareket ederler. Literatürde

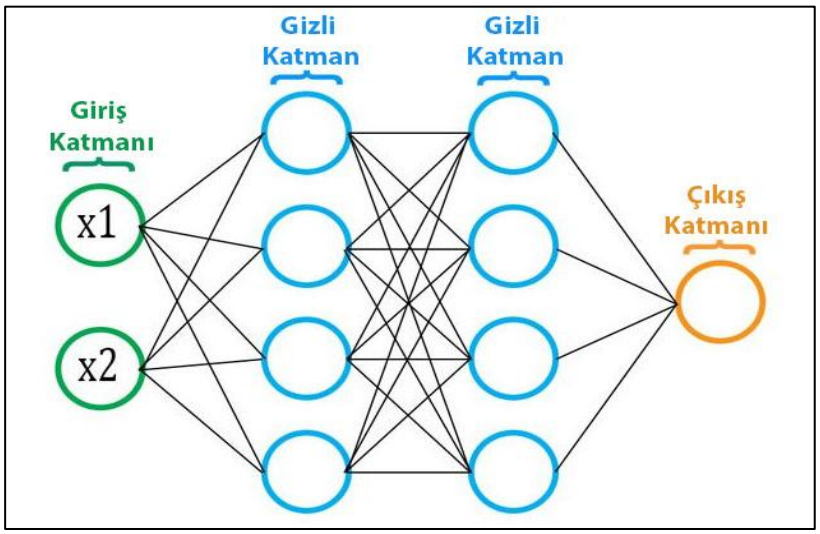
tekrarlanamayan ağlar olarak da bilinen bu ağlar kendi aralarında 2 farklı sınıfa ayrılırlar. Bunlar, tek katmanlı ileri beslemeli ağlar ve çok katmanlı ileri beslemeli ağlardır.

- **Tek katmanlı ileri beslemeli ağlar**: Sadece giriş ve çıkış birimlerinden oluşan Şekil 4.4'te gösterilen Tek Katmanlı Algılayıcılar (Perceptron) olarak ta bilinen bu ağlar en temel sinir ağı türüdür. Çıktı değeri (-1) ya da (+1) olabilen ve doğrusal bir ağ olan bu ağ türü XOR problemi gibi doğrusal olmayan problemlere çözüm sunamamaktadır.



Şekil 4.4 Tek Katmanlı Algılayıcı (Perceptron) mimarisi

- **Çok katmanlı ileri beslemeli ağlar:** Giriş ve çıkış katmanları arasında bir veya birden fazla gizli katman bulunan ve mimarisi Şekil 4.5'te gösterilen ağlardır. Çok Katmanlı Algılayıcılar (MLP- Multi Layer Perceptron) olarak da bilinen bu ağ XOR gibi doğrusal olmayan problemlerin çözümü için oluşturulmuştur. CNN'ler MLP'lerin bir türüdür.



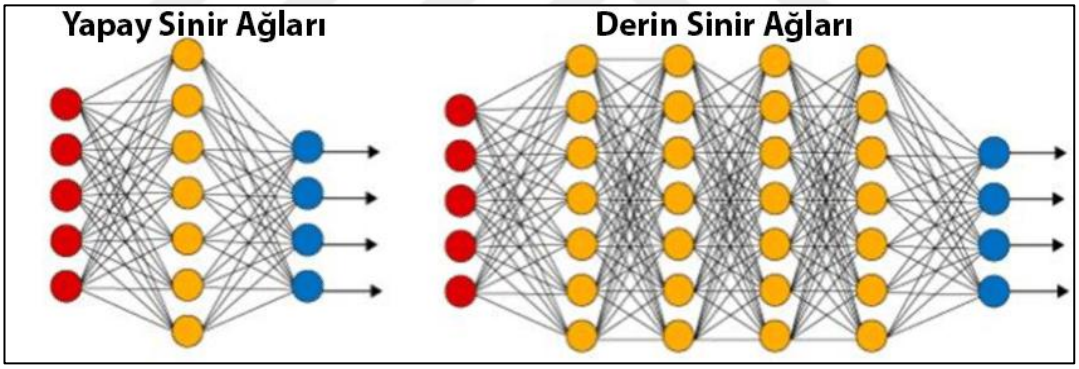
Şekil 4.5 Çok Katmanlı Algılayıcı (Multi Layer Perceptron) mimarisi

4.1.2.2 Geri beslemeli ağlar

İleri beslemeli ağların aksine sinyallerin sadece tek yönlü olmadığı, gizli katman ve çıkış katmanlarından çıkan sinyallerin giriş katmanı ya da gizli katmanlara geri beslendiği ağlardır. Literatürde tekrarlanan, yinelenen ağlar olarak da bilinen bu ağlar doğrusal olmayan ve dinamik yapılara sahip ağlardır. RNN ve LSTM gibi ağlar Geri Beslemeli Ağlar'ın bir türüdür.

4.2 Derin Sinir Ağları (DNN – Deep Neural Network)

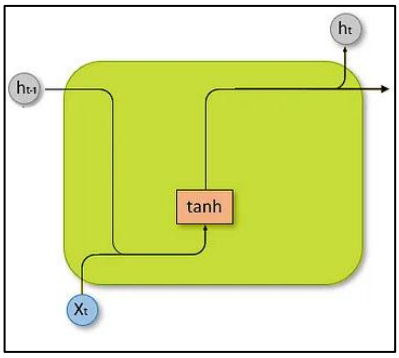
Çok Katmanlı Yapay Sinir Ağlar'da bulunan gizli katman sayılarının arttırılması ile oluşan Derin Sinir Ağları derin öğrenme mimarileri olarak da adlandırılmaktadır. Çok Katmanlı Yapay Sinir Ağları birkaç tane gizli katmana sahip iken DNN'ler çok sayıda ve karmaşık yapılarda gizli katmanlara sahiptirler. Yapay Sinir Ağları ve Derin Sinir Ağları'nın mimarileri Şekil 4.6'da verilmiştir.



Şekil 4.6 ANN ve DNN mimarileri

4.2.1 Tekrarlayan Sinir Ağı (RNN – Recurrent Neural Network)

Düğümler arasındaki bağlantıların yönlendirilmiş bir döngü oluşturarak dinamik zamansal davranış sergilediği, Geri Beslemeli Yapay Sinir Ağları'nın gelişmiş bir versiyonu olan ağlardır (Şeker vd., 2017). RNN'ler sıralı bilgileri işlemek için kullanılan hafızaya sahip ağlardır. RNN işlem akışında gizli katmanlardan çıkan sonuçlar hem içerik birimlerinde (content unit) tutulur hem de bir sonraki katmanda bulunan düğümlere iletilir.

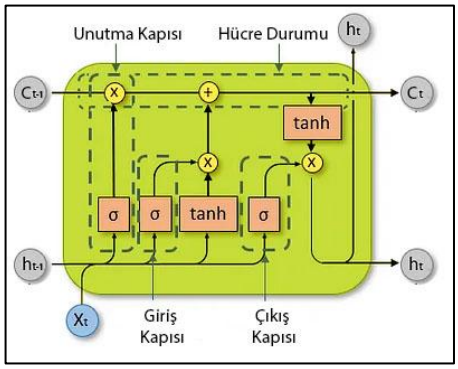


Şekil 4.7 RNN

RNN işlem döngüsünde Şekil 4.7'de görüldüğü gibi t-1 zamanda üretilen ve bir önceki gizli katmandan gelen değerin (h_{t-1}) belirli bir katsayı (U) ile çarpımı ve t zamanında üretilen güncel bilgilerin (X_t) başka bir ağırlık değeri (W) ile çarpımının sonuçları toplanarak tanh aktivasyon fonksiyonuna sokulur. Aktivasyon fonksiyonun ürettiği değer (h_t), t anındaki gizli katmanın sonuç değeridir ve bir sonraki yapay sinir hücresine girdi olarak verilmek üzere içerik biriminde tutulur (Ergüder, 2018). RNN'ler zaman serisi analizi, doğal dil işleme, görüntü işleme gibi alanlarda sıklıkla kullanılan ağlardır. Deepfake tespit çalışmalarında RNN'ler sıklıkla kullanılmaktadır.

4.2.2 Uzun Kısa Süreli Bellek (LSTM – Long Short Term Memory)

RNN'lerde, uzun metinlerde ortaya çıkan kaybolan gradyan problemini çözmek için LSTM ağları ortaya çıkarılmıştır. RNN'ler kısa süreli bir hafızaya sahipken LSTM'ler daha büyük bir hafızaya sahiptir ve daha uzun metinlerde bulunan eksik kelimeleri tahmin edebilirler. LSTM'ler de, Şekil 4.8'de görüldüğü gibi giriş, çıkış, unutma kapıları ve hücre durumu (cell state) bulunmaktadır.



Şekil 4.8 LSTM

- **Unutma Kapısı (Forget Gate):** Unutma kapısı hangi bilgilerin unutulacağını ya da hafızada tutulacağına karar veren kapıdır. Bir önceki gizli katmandan gelen değer (h_{t-1}) ve güncel bilgiler (X_t) sigmoid fonksiyonundan geçirilir. Değer 0'a yakınsa bilgi unutulurken 1'e yakınsa hafızada tutulur.

- **Giriş Kapısı (Input Gate)**: Giriş kapısı hücre durumunu (cell state) güncellemek için kullanılır. Unutma kapısında olduğu gibi bir önceki gizli katmandan gelen değer (h_{t-1}) ve güncel bilgiler (X_t) sigmoid fonksiyonundan geçirilerek hangi bilginin tutulacağına karar verilir. Sonra aynı bilgiler tanh fonksiyonundan geçirilip (-1,1) arasına indirgenerek çıkan sonuçlar çarpılır.

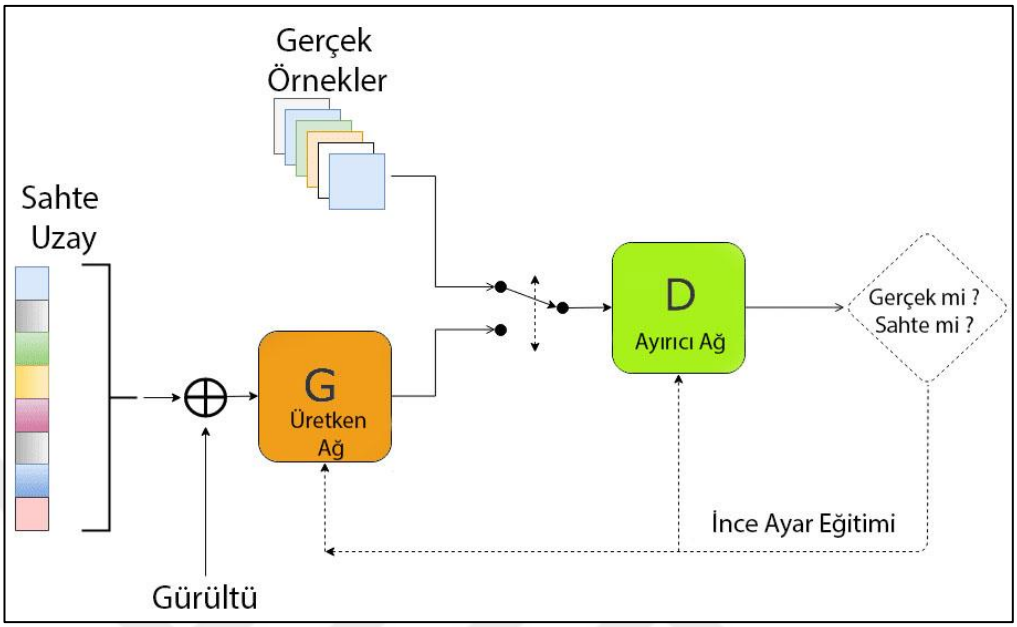
- **Hücre Durumu (Cell State)**: Ağ üzerindeki veri akışının sağlandığı, iletişim hattına cell state denilmektedir. Bir önceki gizli katmanın cell state değeri (C_{t-1}) ile unutma kapısından gelen değer çarpılır. Çıkan sonuç, hem bir sonraki katmana, bir önceki cell state değeri (C_t) olarak gönderilir, hem de giriş kapısından gelen değer ile toplanarak Tanh fonksiyonundan geçirilir.

- **Çıkış Kapısı (Output Gate)**: Sonraki katmana gönderilecek olan değerin belirlendiği kapıdır. Bir önceki gizli katmandan gelen değer (h_{t-1}) ve güncel bilgiler (X_t) sigmoid fonksiyonundan geçirilerek ortaya çıkan değer ile Cell State'de Tanh fonksiyonundan geçirilerek ortaya çıkan değer çarpılarak bir sonraki katmana, bir önceki gizli katmandan gelen değer (h_t) olarak gönderilir.

4.2.3 Üretken Çekişmeli Ağlar (GAN - Generative Adversarial Networks)

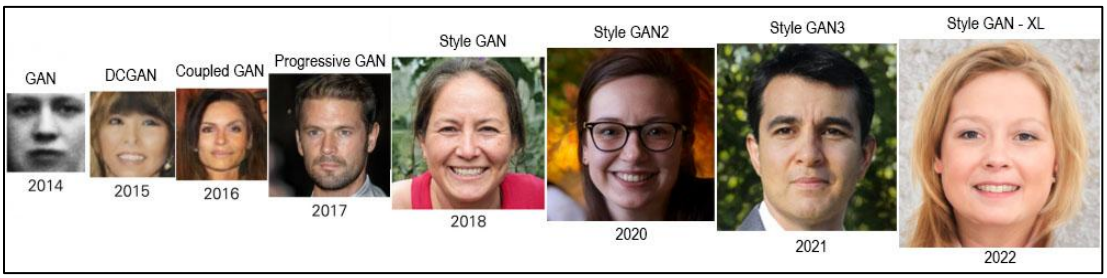
2014 yılında Ian Goodfellow ve arkadaşları tarafından ortaya atılan denetimsiz öğrenme türündeki bir Derin Sinir Ağı'dır. 2 farklı ağın birleşiminden oluşmaktadır. Bunlar; Üretici Ağ (Generative Network) ve Ayırt Edici Ağ (Discriminative Network) olarak adlandırılmaktadır. Mimarisi Şekil 4.9'da verilen GAN'larda üretici sahte resim üreterek bu resmi ayırıcıya gönderir. Ayırıcı kendisine verilen gerçek resimlerle üreticiden gelen resimler arasındaki farkları inceleyerek, üreticiden gelen resim için bir puan hesaplayıp ince ayar yaparak resmi, gerçeğe daha yakın yapabilmesi için üreticiye geri gönderir. Üretici ayırıcıdan gelen puan doğrultusunda hatalarını görüp resmi düzeltir ve tekrar ayırıcının takdirine sunar. Bu döngü, ayırıcı gerçek ve sahte

resim arasındaki farkı ayırt edemeyene kadar devam eder ve neticesinde insan gözüyle sahteliği ayırt edilemeyen sentetik resimler ortaya çıkar.



Şekil 4.9 GAN mimarisi (Bayram, 2021)

GAN'lar deepfake medyalar üretiminde sıklıkla kullanılan derin sinir ağlarıdır. Şekil 4.10'da görüldüğü gibi 2014 yılından bu yana kendini oldukça geliştiren bu ağın çıktıları öncelerde bariz şekilde ayırt edilebilirken, gelişen teknoloji ile beraber daha fazla eğitim süreleri ve daha fazla örnek resimler üzerinde çalışan GAN'lar, mükemmele yakın kalitede sentetik veri üretebilmektedir.

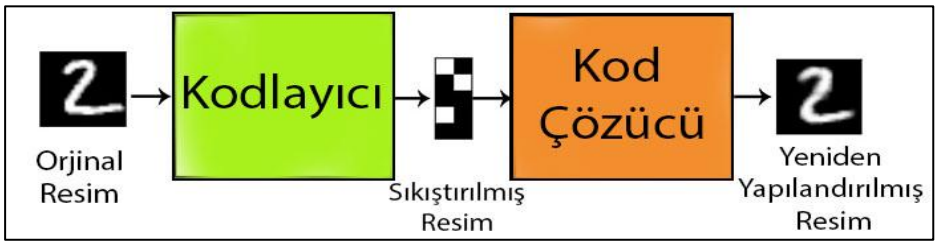


Şekil 4.10 GAN ile oluşturulan insan resimlerinin gelişimi (Goodfellow vd., 2020)

4.2.4 Otokodlayıcı (Autoencoder)

Otokodlayıcı, denetimsiz öğrenme türündeki sinir ağlarının gelişmiş bir örneğidir. Otokodlayıcılar'ın çalışma prensibi, deepfake oluşturma yöntemleri başlığı altında detaylı anlatıldığı için bu bölümde tekrar detay verilmemiştir. GAN'larda olduğu gibi

Otokodlayıcı'lar; kodlayıcı (encoder) ve kod çözücü (decoder) olmak üzere 2 parçadan oluşmaktadır. Otokodlayıcılar'ın işleyişinde, giriş olarak verilen veriyi önce kodlayıcı işleyerek sıkıştırır. Kod çözücü de bu sıkıştırılmış veriyi kullanarak orijinal veriyi yeniden oluşturmaya çalışır. Otokodlayıcı iş akışı Şekil 4.11 de görülmektedir.



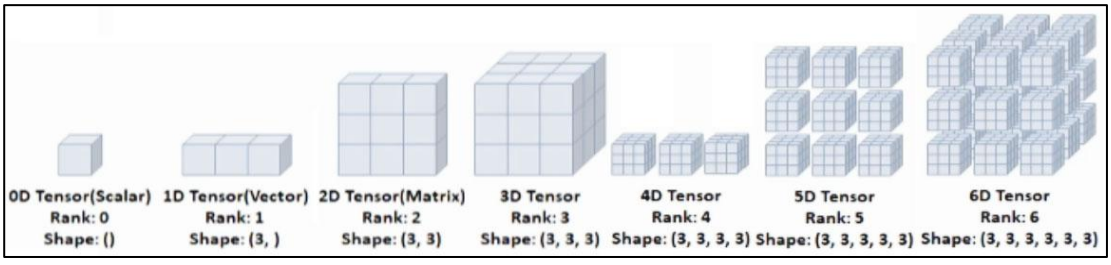
Şekil 4.11 Otokodlayıcı

4.2.5 Evrişimli Sinir Ağları (CNN - Convolution Neural Network)

İleri beslemeli çok katmanlı yapay sinir ağlarının bir türü olan CNN'ler görüntü verileri üzerinde çalışmaktadır. YSA'lara göre çok daha fazla gizli katmana sahip olan CNN'lerde ek olarak 3 farklı katman bulunmaktadır. Bunlar; Evrişim Katmanı (Convolutional Layer), Havuzlama Katmanı (Pooling Layer) ve Tam Bağlantılı Katman (Dense - Fully Connected Layer)'dır.

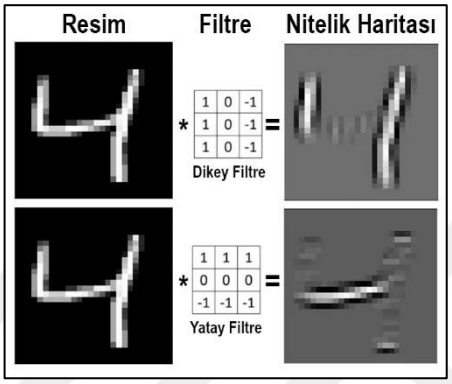
4.2.5.1 Evrişim katmanı

Evrişim bir tür doğrusal işlem türüne verilen isim olup CNN katmanlarının en az bir tanesinde genel matris çarpımı yerine evrişim işlemi uygulanmaktadır (Goodfellow vd., 2016). CNN'ler görüntüleri işlerken sayısal veriler olarak tensörler halinde işlerler. Şekil 4.12'de farklı çeşitleri görülen tensör, içeriğinde sayısal değerlerin barındırıldığı farklı boyutlardaki kümeler olarak adlandırılabilir.



Şekil 4.12 Farklı boyutlardaki tensörler (Exem, 2022)

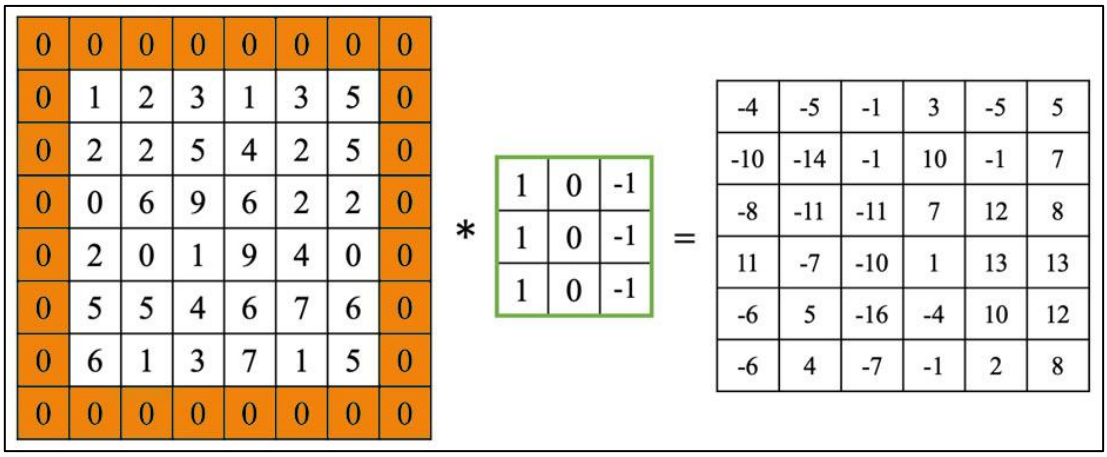
Siyah-beyaz resimler, [0-255] değerleri arasında tek kanala (gri tonlama) sahip olduğu için 2D tensör (matris), renkli resimler ise [0-255] değerleri arasında 3 kanala (Red, Green, Blue) sahip olduğu için 3D tensör halinde işleme tabi tutulur. Evrişim katmanında resimlerdeki bölgesel örüntüleri öğrenmek üzere nitelik haritaları (feature map) çıkarılır. Nitelik haritalarının çıkarılması için pikseller üzerinde, filtre (filter) ya da çekirdek (kernel) denilen, genellikle 3x3, 5x5, 7x7 gibi farklı boyutlardaki matrisler gezdirilir. Nitelik haritası çıkarma işlemi Şekil 4.13'te görülmektedir.



Şekil 4.13 Nitelik haritası çıkarma işlemi

Öğrenilen nitelik haritaları yön değiştirmeyen özelliğe sahiptir. Eğer bu özellik resmin sol alt tarafını temsil ediyor ise benzer ya da aynı özellik, başka bir yerde bulunursa öğrenme işlemi tekrar edilir. CNN'lerin ilk evrişim katmanlarında resimlerin kenar köşe gibi temel parçaları öğrenilirken daha derin katmanlarda kulak, burun, göz gibi daha genel özellikler öğrenilir ve bu şekilde devam eder (Chollet, 2019).

Resmin pikselleri üzerinde filtreler gezdirildiğinde oluşturulan nitelik haritası, boyut olarak girdi resimden daha küçük olur. Resim 7x7 boyutunda bir girdi resim üzerinde 3x3'lük bir filtre gezdirildiğinde oluşan nitelik haritası 5x5 boyutuna, 5x5'lik bir filtre gezdirildiğinde 3x3 boyutuna küçülmektedir. Bu durum bazı özelliklerin öğrenilememesine neden olmaktadır. Çözüm olarak Şekil 4.14'te görüldüğü üzere resmin kenarına, gezdirilen filtrenin boyutuna göre 0'lardan oluşan ekstra piksel eklenmektedir. 3x3 filtre için tüm kenarlara 1'er tane piksel 5x5 için 2'şer tane piksek eklenmelidir. Bu işleme kenar doldurma (padding) denilmektedir.



Şekil 4.14 Kenar Doldurma

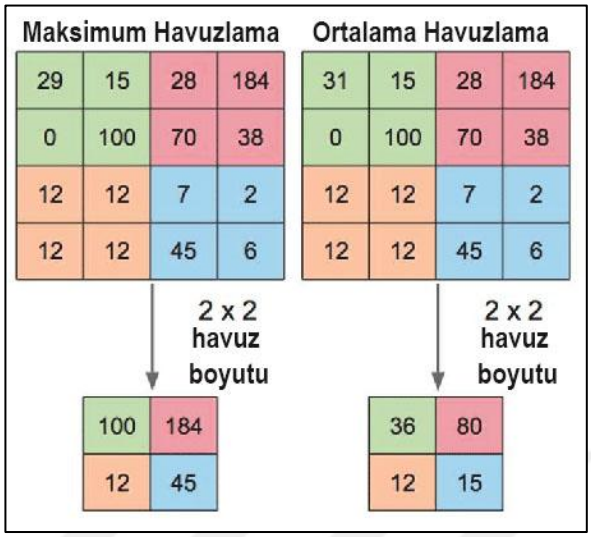
Çıktı büyüklüğüne etki eden başka bir durum da adım aralığı (stride) olarak adlandırılan, filtrenin girdi resim üzerinde kaç piksel atlayarak işlem yaptığıdır. Varsayılan olarak 1 değerini alan bu değer 2 olduğunda oluşan nitelik haritası ½ oranında küçülür. Bazı önemli özelliklerin öğrenilememesine neden olabilen bu özellik nadiren kullanılsa da bu işlem yerine daha çok maksimum havuzlama işlemi tercih edilmektedir (Chollet, 2019).

4.2.5.2 Havuzlama katmanı

Literatürde ortaklama katmanı olarak da geçen havuzlama katmanında temel amaç boyut azaltma (down sampling) işlemidir. Evrişim katmanında yapılan, adım aralığının 2 olarak belirlenmesi ile benzer işi yapar. Aradaki fark ise gezdirilen filtrenin içinde bir değer olmamasıdır. Örnek verecek olursak 3x3'lük boş bir matris pikseller üzerinde gezinir. Bu matrisin üzerinde olduğu tüm piksellerin en büyüğü değer olarak alınıyorsa bu işleme maksimum havuzlama, piksellerin ortalaması alınıyorsa ortalama havuzlama denilmektedir.

Şekil 4.15'te örneği verilen havuzlama katmanında boyut indirgeme yapılmasının 2 temel amacı bulunmaktadır. İlki uzamsal hiyerarşilerin öğrenilebilmesidir. Örnek verecek olursak 3x3'lük bir pencere başlangıç girdisi 7x7'lik bir pencereden gelen girdiyi içerdiğinde, havuzlama işlemi yapılıp boyut indirgeme yapılmazsa sadece 7x7lik bir görüntüye bakarak öğrenilen bilgiyi tahmin etmeye çalışır. Oysaki gerçekte öğrenilen bilgiler farklı girdilerde farklı boyutlarda olabilir. İkinci amaç ise üzerinde

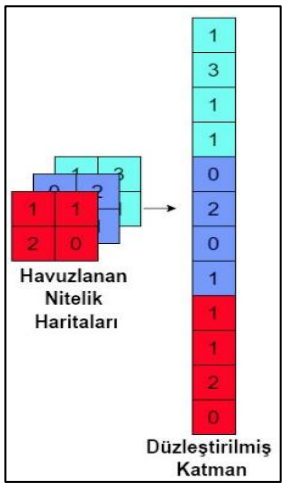
çalışılan katsayı boyutunu azaltarak ağın hesaplayacağı parametre sayısını azaltmaktır (Chollet, 2019).



Şekil 4.15 Havuzlama

4.2.5.3 Tam bağlantılı katman

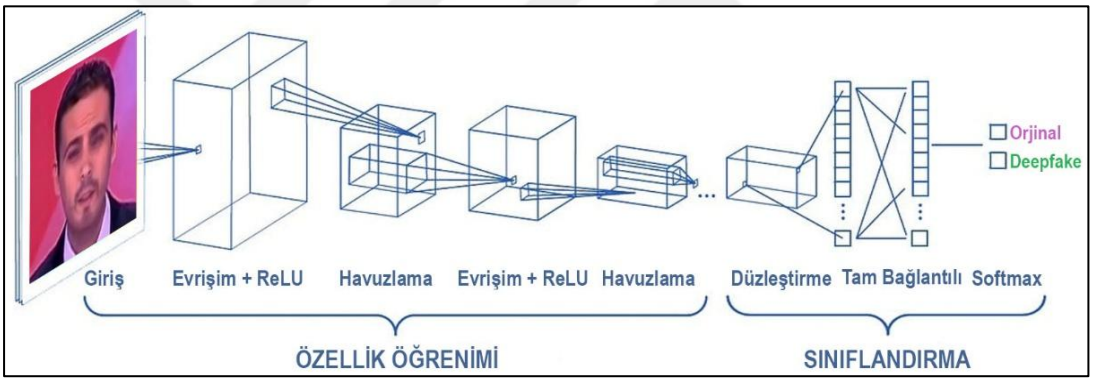
Yoğun (Dense) katman olarak da adlandırılan tam bağlantılı katman CNN'lerin son katmanıdır. Küresel ilişkilerin anlaşılabilmesi için havuzlama katmanından gelen tüm nitelik haritaları birleştirilir. Birleştirilen 2 boyutlu matris şeklindeki nitelik haritaları tam bağlantılı katmanda işlenebilmesi için Şekil 4.16'da görüldüğü gibi düzleştirilerek tek eksenli tensör (vektör) haline getirilir.



Şekil 4.16 Düzleştirme

Tam bağlantılı katman içinde yer alan katmanlar, içindeki düğümleri bir sonraki katmandaki tüm düğümlere bağlar. Tam bağlantılı katmanın ismi buradaki bağlantı yapısından gelmekte olup temel amaç çıkarılan tüm özelliklerin her bir düğüme katkıda bulunması ve ağ tarafından verilerin birleştirilmesidir. Böylelikle özellikler arasındaki ilişkiler ortaya çıkarılarak genel özellikler ve örüntüler öğrenilir (Aramendia, 2024).

Tam bağlantılı katmanların içerisindeki tüm düğümlerde kullanılan ReLU gibi doğrusal olmayan aktivasyon fonksiyonları ile birlikte, önceki katmanlarda öğrenilemeyen doğrusal olmayan ilişkiler öğrenilir. Bu katmandan sonra sınıflandırma başlığı olarak da adlandırılan çıktı katmanında kullanılan softmax, sigmoid gibi aktivasyon fonksiyonları sayesinde veriler ayrıştırılarak sınıflandırma işlemi yapılır. Tipik bir CNN mimarisi Şekil 4.17'de gösterilmiştir.



Şekil 4.17 CNN mimarisi (Ergün & Kılıç, 2021)

4.3 Öğrenme Aktarımı (Transfer Learning)

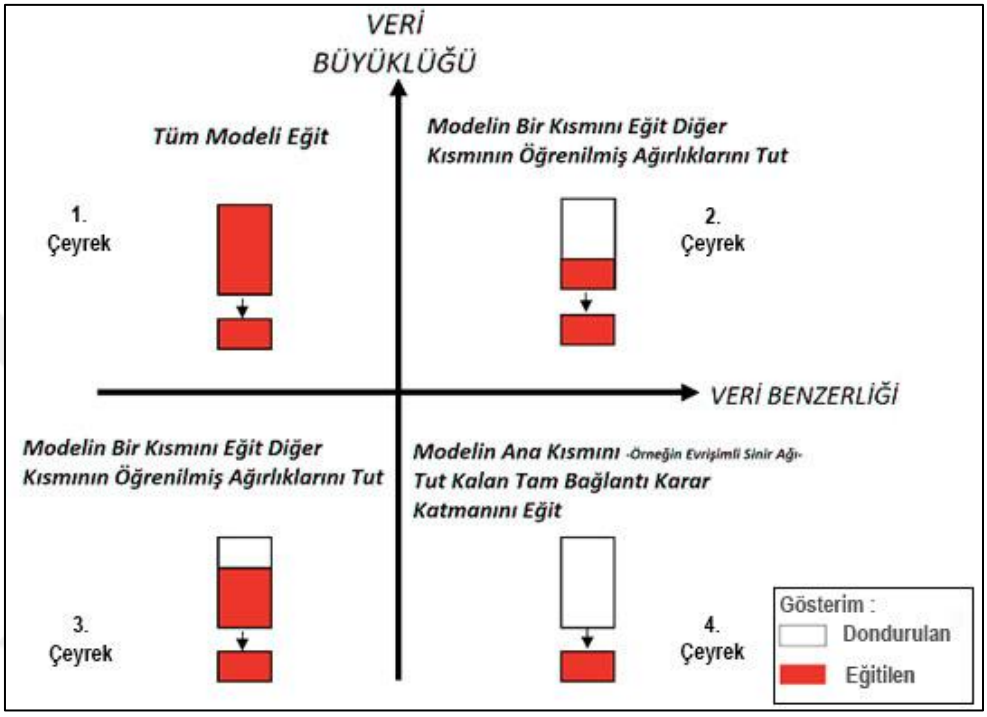
Bir görüntü veri seti üzerinde sınıflandırma problemi için CNN kullanılmak istenildiğinde sıfırdan bir mimari geliştirmenin maliyeti yüksek olabilmektedir. Oluşturulan CNN mimarisinin başarımını arttırmak için derin tecrübe ve zaman gerekmektedir. Özellikle zaman ve tecrübe eksikliğinin giderilmesi için en çok tercih edilen yöntem, ImageNet gibi çok sınıflı ve içerisinde milyonlarca görsel veri içeren veri setlerinde eğitilerek birçok farklı sınıflandırma probleminde yüksek başarı sağlamış olan önceden eğitilmiş (pretrained) modellerin kullanılmasıdır.

Tablo 4-1 Keras kütüphanesinde bulunan önceden eğitilmiş modeller(Keras, 2015)

Model	Boyut (MB)	En iyi (%)		Parametreler	Derinlik	Çıkarım adımı başına zaman (ms)	
		1 Doğruluk	5 Doğruluk			CPU	GPU
Xception	88	79,0	94,5	22,9 milyon	81	109,4	8,1
VGG16	528	71,3	90,1	138,4 milyon	16	69,5	4,2
VGG19	549	71,3	90,0	143,7 milyon	19	84,8	4,4
ResNet50	98	74,9	92,1	25,6 milyon	107	58,2	4,6
ResNet50V2	98	76,0	93,0	25,6 milyon	103	45,6	4,4
ResNet101	171	76,4	92,8	44,7 milyon	209	89,6	5,2
ResNet101V2	171	77,2	93,8	44,7 milyon	205	72,7	5,4
ResNet152	232	76,6	93,1	60,4 milyon	311	127,4	6,5
ResNet152V2	232	78,0	94,2	60,4 milyon	307	107,5	6,6
InceptionV3	92	77,9	93,7	23,9 milyon	189	42,2	6,9
InceptionResNetV2	215	80,3	95,3	55,9 milyon	449	130,2	10,0
MobileNet	16	70,4	89,5	4,3 milyon	55	22,6	3,4
MobileNetV2	14	71,3	90,1	3,5 milyon	105	25,9	3,8
DenseNet121	33	75,0	92,3	8,1 milyon	242	77,1	5,4
DenseNet169	57	76,2	93,2	14,3 milyon	338	96,4	6,3
DenseNet201	80	77,3	93,6	20,2 milyon	402	127,2	6,7
NASNetMobile	23	74,4	91,9	5,3 milyon	389	27,0	6,7
NASNetLarge	343	82,5	96,0	88,9 milyon	533	344,5	20,0
EfficientNetB0	29	77,1	93,3	5,3 milyon	132	46,0	4,9
EfficientNetB1	31	79,1	94,4	7,9 milyon	186	60,2	5,6
EfficientNetB2	36	80,1	94,9	9,2 milyon	186	80,8	6,5
EfficientNetB3	48	81,6	95,7	12,3 milyon	210	140,0	8,8
EfficientNetB4	75	82,9	96,4	19,5 milyon	258	308,3	15,1
EfficientNetB5	118	83,6	96,7	30,6 milyon	312	579,2	25,3
EfficientNetB6	166	84,0	96,8	43,3 milyon	360	958,1	40,4
EfficientNetB7	256	84,3	97,0	66,7 milyon	438	1578,9	61,6
EfficientNetV2B0	109.42	81,3	-	28,6 milyon	-	-	-
EfficientNetV2B1	192.29	82,3	-	50,2 milyon	-	-	-
EfficientNetV2B2	338,58	85,3	-	88,5 milyon	-	-	-
EfficientNetV2B3	755.07	86,3	-	197,7 milyon	-	-	-
EfficientNetV2S	1310	86,7	-	350,1 milyon	-	-	-

2015 yılında François Chollet tarafından oluşturulan ve bir derin öğrenme kütüphanesi olan Keras kütüphanesinde onlarca hazır model kullanıma sunulmuş olup bu modellerin bazıları Tablo 4-1'de özellikleri ile birlikte verilmiştir. Bu modellerin haricinde farklı kaynaklardan indirilerek kullanılabilen hazır modellerde bulunmaktadır.

Önceden eğitilmiş modeller, mevcut ağırlıkları ile kullanılabildiği gibi ağırlıkları olmadan da kullanılabilmektedir. Modelin ağırlıkları kullanılarak eğitim yapılması durumunda, model üzerinde ufak değişiklikler yaparak ağın eğitildiği veri seti ile mevcut veri seti arasında, veri boyutu ve veri benzerliği açısından değerlendirilme yapılarak hazır modelin hangi katmanlarının eğitilmesi gerektiğine karar verilmelidir.



Şekil 4.18 Öğrenme aktarımında yöntemin seçilmesi (Kizrak, 2019)

Şekil 4.18 incelendiğinde 1. çeyrekte mevcut veri setinin büyük ve ağın eğitildiği veri setinde bulunan sınıfla benzerliğinin az olduğu durumda tüm modelin eğitilmesi gerektiği anlaşılmaktadır. Bu yaklaşıma sıfırdan eğitim denilmektedir. Aynı şekilde 2, 3 ve 4. çeyreklere bakılarak veri büyüklüğü ve benzerliği kıstasına göre modelin katmanlarının hangi oranda dondurulacağına karar verilerek eğitimin gerçekleşmesi durumuna ince ayar (fine tuning) denilmektedir.

5. ÖĞRENME AKTARIMI İLE DEEPFAKE MEDYA TESPİTİ

Deepfake medya tespitine ilişkin literatür çalışmalarının incelenmesi, mevcut veri setleri, çalışılacak olan fiziksel donanım, çalışma ortamı gibi etmenlerin değerlendirilmesi sonucunda önceden eğitilmiş CNN modelleri kullanarak öğrenme aktarımı ile deepfake medya tespiti yapılmasına karar verilmiştir. Bu bölümde 3 ve 4. bölümlerde derinlemesine tanıtımı yapılan yöntem ve araçların hangilerinin tercih edildiği ve nasıl kullanıldığı anlatılmaya çalışılmıştır.

5.1 Geliştirme Ortamı ve Fiziksel Donanım

Bu çalışma kapsamında geliştirme ortamı olarak Juppyter notebook tabanlı Google Colab Pro+ platformu kullanılmıştır. Colab'ın ücretli versiyonu olan Pro+ sürümünde kullanıcılara 500 işlem birimi, NVIDIA'nın; A100, L4 ve T4 tensör çekirdekli ekran kartlarına erişim, arka planda yürütme ve aynı anda 3 makineye erişim özellikleri sunulmaktadır. Özellikle derin öğrenme mimarileri ile uzun süren eğitimler gerektiren çalışmalarda NVIDIA'nın ekran kartlarının kullanımı araştırmacılara zaman avantajı kazandırmaktadır. Colab platformunun sunmuş olduğu CPU (Merkezi İşlem Birimi), TPU (Tensör İşlem Birimi) ve GPU (Grafik İşlem Birimi) donanımları kullanılarak, 2000 resim ve 32 batch size değeri ile yapılan deneysel çalışma sonucunda elde edilen zaman ve işlem birimi maliyetlerinin kıyaslaması Tablo 5-1'de verilmiştir.

Tablo 5-1 Google Colab platformuna ait donanımların karşılaştırılması

	A100 GPU	**L4 GPU**	**T4 GPU**	**TPU**	**CPU**
Maliyet (işlem birimi / saat)	10,59	3	1,67	1,76	0,07
Ortalama Eğitim Süresi	3 ms	7 ms	12 ms	132 ms	1020 ms

Çalışmamızda genellikle NVIDIA A100 ekran kartı ve L4 ekran kartı kullanımı tercih edilmiştir. Belirtilen donanımlar kullanılarak deneysel çalışmalar haricinde sadece veri setinden çerçeve çıkartılması ve modellerin eğitimi süreçleri ortalama 40 saat sürmüştür.

Çalışma kapsamında programlama dili olarak Python kullanılmıştır. Kullanılan modellerin oluşturulması ve eğitim aşamaları için Google tarafından geliştirilen açık kaynaklı makine öğrenimi kütüphanesi olan Tensorflow kullanılmıştır. Tensorflow ile uyumlu olması ve bünyesinde onlarca önceden eğitilmiş ağ barındırması nedeniyle Uygulama Programlama Arabirimi (API) olarak Keras tercih edilmiştir. Çalışmada klasörler üzerinde kolay çalışabilmek için os kütüphanesi, resimler üzerinde farklı işlemler yapmak için open-cv kütüphanesi, video karelerinden yüz görüntülerini çıkarmak için mediapipe kütüphanesi, tensörler üzerinde işlem yapabilmek için numpy kütüphanesi ve modellerin sonuç grafiklerinin gösterimi için matplotlib kütüphanesi tercih edilmiştir.

5.2 Veri Seti Tercihi ve Verilerin Hazırlanması

CNN modelleri ile çalışılmak için veri seti tercihinde dikkat edilmesi gereken hususlardan bir tanesi sınıflar arasındaki veri sayılarının dengeli dağılımıdır. Dengesiz bir veri seti erken aşamada aşırı öğrenme (overfitting) problemine yol açabilmektedir. Çalışmamız için veri seti tercih ederken sınıflar arası dengeli dağılım ve yeterli sayıda veri (her sınıf için 1000 video) barındırması, açık ve kolay erişimin bulunması ve yeterli veri çeşitliliğine (977 benzersiz aktör) sahip olması nedeniyle, bu konu ile ilgili literatürdeki birçok çalışmada kullanılan FaceForensics++ veri seti tercih edilmiştir. Veri setinde 5 farklı yöntem ile yapılan deepfake videolar bulunmaktadır. Çalışma da bu yöntemlerden Deepfakes ve Faceswap ile yapılan deepfake videolar kullanılmıştır.

Veri setinde bulunan videolardan open-cv kütüphanesi yardımıyla 60 adet çerçeve çıkarımı işlemi yapılmıştır. Videolardan çıkarılacak olan çerçevelerin, videonun tümünden örnekler taşıması amacıyla eşit aralıklarda olmasına karar verilmiştir. Videoların süreleri, çıkarılacak olan çerçeve sayısına (60) bölünerek çıkan zaman değerlerinde bulunan kareler Google Drive sürücüsüne kaydedilmiştir. Toplam 2000 adet videodan 120.000 çerçeve çıkartılmıştır. Veri setinde bulunan her 2 sınıfa ait örnek görseller Şekil 5.1'de verilmiştir.

Şekil 5.1 Veri setinde bulunan sınıflara ait görseller

Google Drive sürücüsüne kaydedilen 120.000 çerçevede bulunan yüz görüntülerini çıkarmak için MediaPipe kütüphanesinin face detection modülü kullanılmıştır. Yüz görüntüleri çıkarmak için farklı teknikler kullanılabilmektedir. OpenCV kütüphanesinde bulunan Haar Cascade sınıflandırıcıları, Dlib kütüphanesi ve MediaPipe kütüphanesi bunlardan bazılarıdır. Çalışma kapsamında farklı teknikler denenerek MediaPipe kütüphanesinin bu konuda daha başarılı olduğu gözlemlenmiştir.

Videoların çözünürlüklerinin farklı olması nedeniyle tüm çerçevelerin yüksekliği 720p olacak şekilde en boy oranını koruyarak yeniden ölçeklendirilmiştir. Çerçevelerden çıkarılacak olan yüzler için yüzün orta noktasına odaklanılarak etrafında kare olacak şekilde yüz görüntüsü tespit edildikten sonra üstten 90, alttan 10, sağ ve soldan 50'şer piksel fazlalık bırakılarak yüz görüntüleri kırpılmıştır. Mediapipe kütüphanesi yardımıyla çıkarılan yüz görüntüleri için güven puanı hesaplanarak görüntüler bu puana göre sıralanmıştır. Bu aşamada bazı videolarda farklı nedenlerle yüz tespit edilememesi veya birden fazla yüz tespit edilmesi gibi durumlar nedeniyle 120.000 görüntü incelenerek deepfake'in doğasına aykırı olan yüz görüntüleri ile insan yüzü haricinde tespit edilen nesne ve canlı görüntüleri manuel olarak silinmiştir. En yüksek

güven puanına sahip 10'ar tane yüz görüntüsü veri setinde kullanılmak için seçilmiştir. Bazı videolardan çıkarılan yüz görüntü sayısı 10'dan daha az olması nedeniyle yaklaşık 20.000 yüz görüntüsü elde edilmiştir. Videolardan yüz görüntülerinin çıkarılması iş akışı Şekil 5.2'de yer almaktadır.

Şekil 5.2 Videolardan yüz görüntüsünün çıkarılması

Elde edilen yaklaşık 20.000 yüz görüntüsünün (10.000 gerçek ve 10.000 deepfake) %80'i eğitim, %10'u doğrulama ve %10'u test için ayrılmıştır. Modelin genelleme başarısının doğru bir şekilde değerlendirilebilmesi için test ve doğrulama için ayrılan görüntülerde bulunan yüzlerin, eğitim setinde bulunan yüzlerden farklı olması özellikle dikkat edilmiştir.

Deneysel çalışmalar yapmak üzere 20.000 görselden oluşan veri seti haricinde, çıkarılan 120.000 çerçeveden ayrıca 100.000 görselden oluşan başka bir veri seti daha oluşturulmuş ve bazı modeller bu veri setinde de eğitilmiştir. 100.000 görselli veri seti için bir videodan aynı kişiye ait 50 görüntü olması nedeniyle tüm modeller 1. dönemde (epoch) bile aşırı öğrenme problemi yaşamıştır. Bu nedenle 20.000 görselli veri setinin yeterli olduğu kanaatine varılmıştır. Deneysel çalışmalar aynı videodan çıkarılan kare sayısının arttırılması ile daha büyük veri seti oluşturmanın başarıyı arttırmadığını, farklı videolardan farklı kişilere ait yüz görüntüleri kullanılarak veri sayısı ile birlikte veri çeşitliliğinin artması gerektiğini göstermiştir.

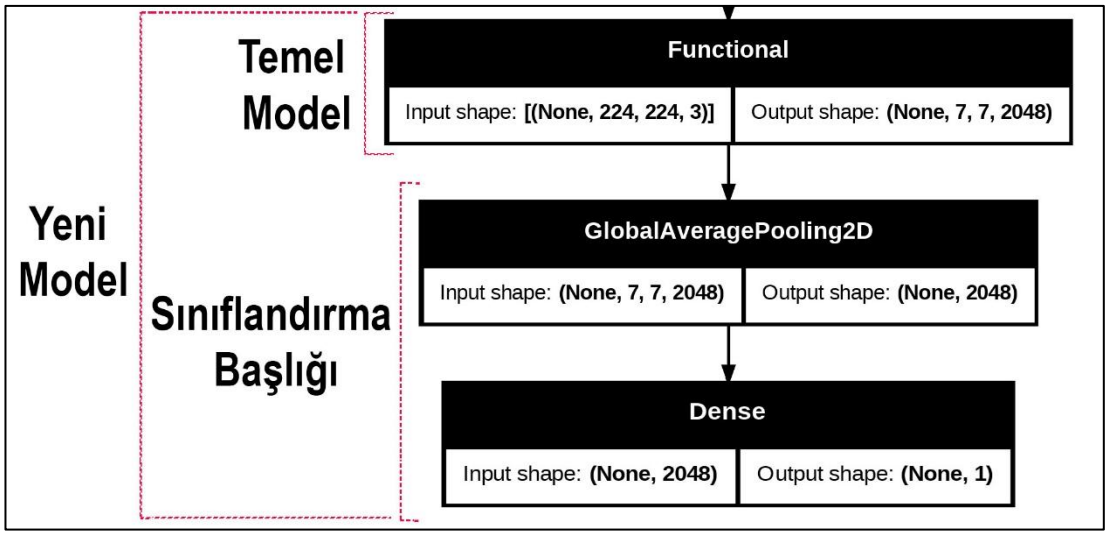
5.3 Kullanılacak Modellerin ve Yöntemin Seçilmesi

Deepfake tespiti yapmak üzere sıfırdan bir CNN oluşturmak yerine önceden eğitilmiş modellerin mimarisinin ve öğrenme aktarımı yönteminin kullanılması tercih edilmiştir. Bölüm 4'te yer alan Tablo 4-1'de detayları verilen ve Keras kütüphanesinden kolaylıkla indirilerek kullanılabilen modellerden, EfficienNetB4 modeli, bu modellin eğitilebilir parametre sayılarına yakın parametreleri (yaklaşık 20 milyon parametre) bulunan; ResNet50V2, DenseNet201 ve InceptionV3 modelleri ile parametre sayısı 138,4 milyon olmasına rağmen çıkarım adımı başına milisaniye cinsinden zamanı (GPU için) diğer modellerle yakın değeri bulunan ve sınıflandırma problemlerinde sıklıkla kullanılan VGG16 modeli tercih edilmiştir.

Tablo 5-2 Kullanılan modellerin özellikleri

Model	Parametreler	Derinlik	Çıkarım adımı başına zaman GPU (ms)
VGG16	138,4 milyon	16	4,2
ResNet50V2	25,6 milyon	103	4,4
InceptionV3	23,9 milyon	189	6,9
DenseNet201	20,2 milyon	402	6,7
EfficientNetB4	19,5 milyon	258	15,1

Tablo 5-2'de detayları verilen modellerin mimarisi 1000 sınıflı ImageNet veri setinde eğitilmek üzere oluşturulduğu için modeli mevcut 2 sınıflı veri setine uyarlamak adına modellerin, sınıflandırma başlığı olarak adlandırılan üst katmanları; GlobalAveragePooling2D ve Dense katmanları çıkarılarak temel modeller oluşturulmuştur. Kullanılan veri setinde bulunan sınıfların her ikisi de insan yüzü olduğu için deepfake görüntülerde bulunan küçük eserlere (artefact) odaklanmak üzere modellerin sıfırdan eğitilmesi gerektiğine, deneysel çalışmalar sonucunda karar verilmiştir. Bunun için temel model yüklenirken ağırlıkları hariç tutulmuştur. Temel model üzerine GlobalAveragePooling2D katmanı ve sonrasında 2 sınıflı veri setini sınıflandırabilmesi için Dense katmanına sigmoid aktivasyon fonksiyonu parametre olarak eklendikten sonra Şekil 5.3'te görüldüğü gibi modellere tekrar dahil edilmiştir.



Şekil 5.3 Temel model kullanılarak oluşturulan yeni model

5.4 Ön İşlemlerin Yapılması ve Parametrelerin Ayarlanması

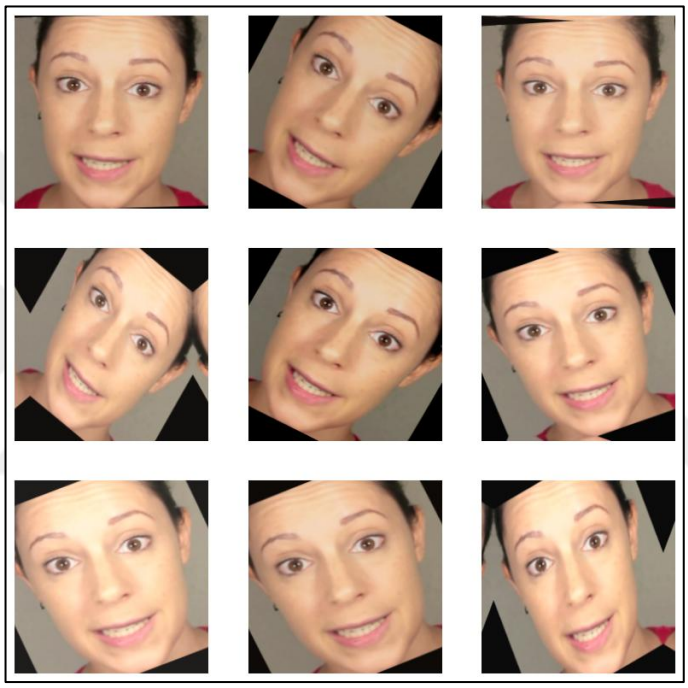
Modellerin başarımının arttırılması için veriler üzerinde bazı ön işlemler yapılmıştır. Önceden eğitilmiş modellere verilecek olan giriş görüntülerinin boyutlarının bazı aralıklarda ya da standart ölçülerde olması gerekmektedir. Bu dönüşümleri Keras API'si bünyesinde yer alan *image_dataset_from_directory* fonksiyonu ile eğitim, test ve doğrulama verilerinin hazırlanması aşamasında yapılmıştır.

Bazı modellere verilecek olan görüntülerin piksel değerlerinin [0,255] aralığından [0,1] ya da [-1,1] aralığına dönüştürülmesi gerekmektedir. Bu işlem için Keras API'sinde yer alan *Rescaling* fonksiyonu kullanılabilmektedir. Fakat bu çalışma kapsamında modellerin kendi bünyesinde yer alan *preprocess_input* özelliği kullanılmıştır.

Tablo 5-3 Modellerin giriş görüntüleri için istediği değerler

Model	**Modele Özgü Değerler**	
	Giriş Boyutu	**Piksel Değer Aralığı**
VGG16	224x224	[0-255] - BGR
ResNet50V2	224x224	[-1,1] - RGB
InceptionV3	299x299	[-1,1] -RGB
DenseNet201	224x224	[0,255] - RGB
EfficientNetB4	380x380	[0,255] - RGB

Görüntü verilerinin, modellere özgü değerlere dönüşümü yapıldıktan sonra eğitim verilerinin çeşitliliğini arttırmak ve modellerin genelleme yeteneğini arttırmak için veri arttırma (data augmentation) işlemi yapılmıştır. Veri arttırma işlemi için Keras API'si bünyesinde yer alan özellikler kullanılarak görüntülere %10 oranında döndürme, yakınlaştırma, kontrast ekleme, parlaklık değiştirme ve gauss gürültüsü eklenmiştir. Döndürme işlemi uygulanırken oluşan boş pikseller 0 (siyah) değeri ile doldurulmuştur. Veri arttırma işlemi uygulanan bir görselin sonucu Şekil 5.4'te görülmektedir.



Şekil 5.4 Veri arttırma işlemi yapılan görüntü

Verilerin hazırlanması ve modellerin kurulmasından sonra hiper parametre adı verilen, modelin başarımını doğrudan etkileyen ve başarıma olan etkisi deneme yanılma yolu ile öğrenilebilen değişkenlerin değerleri ayarlanmıştır.

Yığın boyutu (batch size) değeri, modele her iterasyonda verilecek olan görüntülerin, grup halinde modele verilerek paralel işlem sayesinde eğitim süresini kısaltmaya olanak sağlayan bir değerdir. Bu değer genellikle 16, 32, 64 gibi 2'nin kuvvetleri şeklinde belirlenmektedir. Bu çalışma kapsamında 16 ve 32 yığın boyutu değeri kullanılarak sonuçlar kıyaslanmıştır.

Öğrenme oranı (learning rate) parametresi, optimizasyon algoritmasının her eğitim döneminde (epoch) ağırlıkların güncellenmesi için kullandığı bir katsayıdır. Genellikle 0,1; 0,01; 0,001 gibi 10'un negatif kuvvetleri şeklinde verilen değerler alır. Optimizasyon algoritması olarak *Adam Optimizasyon Algoritması* tercih edilmiştir. Çalışma kapsamında başlangıç öğrenme oranı olarak 0,0001 ($10^{-4)}$ değeri verilmiştir. Eğitim esnasında *ReduceLROnPlateau* geri çağırma (callback) fonksiyonu kullanılarak kayıp fonksiyonunun hesapladığı doğrulama kaybı (validation loss) değerinin 3 dönem boyunca azalmaması durumunda öğrenme oranı yarıya indirilmiştir. Kayıp fonksiyonu olarak *Binary Crossentropy* fonksiyonu kullanılmıştır.

Modeller 30 dönem (epoch) boyunca eğitime tabi tutulmuştur. Eğitim sırasında *EarlyStopping* (erken durdurma) geri çağırma fonksiyonu kullanılarak doğrulama kaybı değerinin 5 dönem boyunca azalma göstermemesi durumunda eğitime son verilerek, model doğrulama kaybı değerinin en az olduğu döneme ait ağırlıklarla kaydedilmektedir.

Modellerin aşırı öğrenme problemi ile karşılaşmaması ve performansının iyileştirilmesi için kullanılan düzenleme (regularization) tekniklerinden olan dışlama (dropout) tekniği çalışmaya dahil edilerek etkisi test edilmiştir. Dropout tekniği eğitim sırasında bazı nöronları belirtilen oranda rastgele olarak devre dışı bırakmaktadır. Çalışma kapsamında dışlama oranı olarak 0,2 değeri kullanılmıştır.

Tablo 5-4 Kullanılan hiper parametre ve fonksiyonlar

Hiper Parametre	**Değer**
Yığın Boyutu (Batch Size)	16, 32
Eğitim Dönemi (Epoch)	30
Öğrenme Oranı (Learning Rate)	0,0001 (10^{-4}, 1e-4)
Dışlama (Dropout)	0,2
Optimizisyon Algoritması	Adam
Kayıp Fonksiyonu (Loss Function)	Binary Crossentropy
Geri Çağırmalar (Callback)	ReduceLROnPlateau, Early Stoping

6. BULGULAR VE TARTIŞMA

Verilerin hazırlanması, modellerin parametrelerinin ayarlanmasının ardından 5 farklı model, dışlama katmanı eklenmesi, 16 yığın boyutu ve 32 yığın boyutu ile birlikte 3 farklı varyasyonla birlikte eğitime tabi tutularak toplam 15 tane eğitim işlemi yapılmıştır. Yapılan her işlem için karışıklık matrisleri çizdirilerek Doğru Pozitif (TP), Doğru Negatif (TN), Yanlış Pozitif (FP) ve Yanlış Negatif (FN) değerleri hesaplanmıştır. Hesaplanan değerler kullanılarak Şekil 6.1'de formülleri verilen, Doğruluk (Accuracy), Kesinlik (Precision), Duyarlılık (Sensitivity-Recall-True Positive Rate), Özgüllük (Specificity-True Negative Rate) ve F1 Puanı değerleri hesaplanmıştır. Son olarak ROC eğrileri çizdirilerek AUC (Area Under the Curve-Eğri Altındaki Alan) değeri hesaplatılmıştır. AUC değeri modelin performansının değerlendirilmesinde sıklıkla kullanılan bir değerdir.

$$\textbf{Kesinlik} = \frac{TP}{TP + FP}$$

$$\textbf{Duyarlılık} = \frac{TP}{TP + FN}$$

$$\textbf{F1 Score} = \frac{2 \times precision \times recall}{precision + recall}$$

$$\textbf{Doğruluk} = \frac{TP + TN}{TP + FN + TN + FP}$$

$$\textbf{Özgüllük} = \frac{TN}{TN + FP}$$

Şekil 6.1 Başarı ölçüm metrikleri

Dışlama (dropout) katmanı eklenerek eğitilen modellerin başarı sonuçları Tablo 6-1'de verilmiştir. Değerler incelendiğinde VGG16, DenseNet201 ve EfficientNetB4 modellerinin 0,90 AUC değeri ile başarılı bir sonuç elde ettiği görülmektedir. Yapılan deneysel çalışmalar sonucunda 0,2 değerinin dışlama katmanı için en ideal oran olduğu belirlenmiştir. Sonuçlar dışlama katmanının doğru oranlarda kullanılmasının modelin başarısını etkileyebileceğini ortaya çıkarmıştır.

Tablo 6-1 Dropout katmanının başarıma etkisi

Model	Başarı Ölçüm Metrikleri (Dropout(0.2) + 16 Yığın Boyutu)					
	AUC	Doğruluk (Acc)	Kesinlik (Pre)	Duyarlılık (Sen)	Özgüllük (Spe)	F1 Puanı
VGG16	**0,90**	**0,82**	0,81	**0,84**	**0,80**	**0,82**
ResNet50V2	0,87	0,77	0,77	0,78	0,76	0,77
InceptionV3	0,81	0,73	0,75	0,69	0,77	0,72
DenseNet201	**0,90**	0,80	**0,85**	0,72	0,87	0,78
EfficientNetB4	**0,90**	**0,82**	0,80	**0,84**	0,79	**0,82**

Dışlama katmanı tüm modellerden çıkarılarak 16 yığın boyutu ile tekrar eğitim gerçekleştirilmiştir. Tablo 6-2'de verilen sonuçlar incelendiğinde 0,93 AUC değeri ile EfficientNetB4 modeli yüksek bir başarım gösterdiği ortaya çıkmıştır. ResNet50V2 ve DenseNet201 modelleri de 0,89 AUC değeri ile başarılı sonuçlar elde ettiği söylenebilir.

Tablo 6-2 (16) yığın boyutunun başarıma etkisi

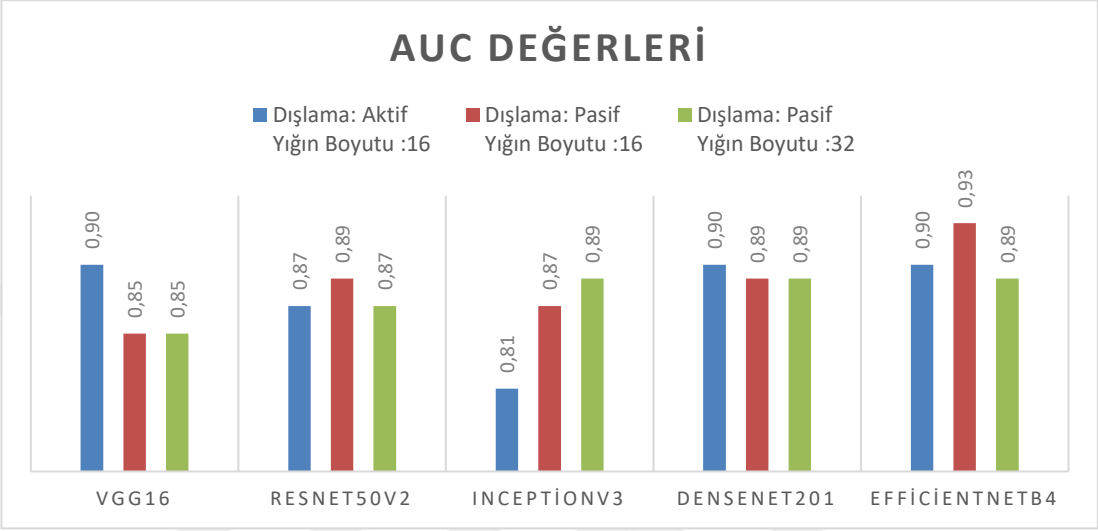
Model	Başarı Ölçüm Metrikleri (16 Yığın Boyutu)					
	AUC	Doğruluk (Acc)	Kesinlik (Pre)	Duyarlılık (Sen)	Özgüllük (Spe)	F1 Puanı
VGG16	0,85	0,77	0,78	0,75	0,79	0,77
ResNet50V2	0,89	0,80	**0,86**	0,74	**0,87**	0,79
InceptionV3	0,87	0,77	0,82	0,69	0,84	0,75
DenseNet201	0,89	0,80	0,85	0,74	0,86	0,79
EfficientNetB4	**0,93**	**0,84**	0,85	**0,84**	0,85	**0,84**

Dışlama katmanı eklenmeden yığın boyutu 32 olarak değiştirilerek yapılan eğitimlerin sonuçları Tablo 6-3'te verilmiştir. Sonuçlar incelendiğinde AUC değerlerinin tüm modeller için birbirine yakın değerler olduğu görülmüştür. InceptionV3, DenseNet201 ve EfficientNetB4 modelleri 0,89 puan ile en yüksek AUC değerini yakalamıştır.

Tablo 6-3 (32) yığın boyutunun başarıma etkisi

Model	Başarı Ölçüm Metrikleri (32 Yığın Boyutu)					
	AUC	Doğruluk (Acc)	Kesinlik (Pre)	Duyarlılık (Sen)	Özgüllük (Spe)	F1 Puanı
VGG16	0,85	0,78	0,79	0,77	0,79	0,78
ResNet50V2	0,87	0,77	0,76	0,81	0,73	0,78
InceptionV3	**0,89**	0,80	**0,85**	0,72	**0,87**	0,78
DenseNet201	**0,89**	**0,81**	0,79	**0,85**	0,77	**0,82**
EfficientNetB4	**0,89**	0,77	0,84	0,68	**0,87**	0,75

Modellerin farklı değerlerle eğitim işlemi sonucunda ortaya çıkan değerler incelendiğinde farklı değerlerin modeller için farklı sonuçlar ortaya çıkardığı görülmektedir. Hangi değişikliğin hangi modele ne şekilde etki ettiğini görebilmek adına bütün eğitimlerin sonucunda hesaplanan AUC değerleri incelenmiştir.



Şekil 6.2 Farklı parametrelerin başarıma etkisi

Şekil 6.2'de yer alan AUC değerleri incelendiğinde dışlama katmanı eklenmesinin DenseNet201 modelinde 0,1 puan, VGG16 modelinde ise 0,5 gibi yüksek bir puanla olumlu artış gösterdiği, diğer modeller için başarımı azalttığı görülmektedir. Yığın boyutunun arttırılması VGG16 ve DenseNet201 için nötr, InceptionV3 modeli için 0,2 puanlık olumlu etki, ResNet50V2 modeli için 0,2 puanlık olumsuz etki, EfficientNetB4 modeli için ise 0,4 puanlık yüksek bir olumsuz etki göstermiştir. Sonuçlar incelendiğinde hiper parametrelerin değiştirilerek modelin başarısına etkisinin ne kadar önemli olduğu açıkça görülmektedir.

Modellerin AUC değerleri incelendiğinde 3 farklı varyasyonun hepsinde de en başarılı modelin EfficientNetB4 modeli olduğu açıkça görülmektedir. EfficientNetB4 modelinin dışlama (dropout) katmanı olmadan 16'lı yığın halinde eğitildiği seçenek 0,93 puan ile en yüksek başarımı sağlamıştır. Diğer modellerin AUC puanlarının da literatürdeki birçok çalışmaya göre yüksek olduğu görülmektedir.

Çalışmada kullanılan hiper parametrelerin modellere etkisinin görülmesi için yapılan deneme çalışmaları sonucunda bir model için başarıyı arttıran parametrenin farklı modeller için başarıyı azaltabileceği görülmüştür. Videolardan çıkarılan çerçevelerden, yüz görüntüleri çıkartılması sırasında MediPipe kütüphanesi kullanılarak yüz görüntüsü dışında farklı görüntüler üzerinde çalışılması engellenmiş ve sonrasında manuel olarak aşırı gürültülü ve yüz görüntüsü içermeyen, deepfake medyanın doğasına aykırı olduğu düşünülen verilerin çıkartılmasının başarı oranına olumlu etki gösterdiği gözlemlenmiştir. Ayrıca veri setinde bulunan sınıfların ikisinde de insan yüzü görüntüsü bulunması nedeniyle aşırı uyum problemi ile karşılaşmamak ve deepfake görüntülerde bulunan küçük eserlere odaklanmak için modellerin tüm veriler üzerinde sıfırdan eğitilmesi de başarı oranını arttırdığı gözlemlenmiştir.

Çalışmanın başarısı üzerinde olumlu etki gösteren diğer faktör ise veri çeşitliliğini arttırmak için yüz görüntüleri üzerinde veri arttırma işlemi yapılmasıdır. Veri arttırma işlemi sırasında farklı katsayılar kullanılarak başarıyı en çok arttıran oranın 0.2 olduğu gözlemlenmiş ve çalışmalarda bu oran kullanılmıştır. Başarı oranını arttıran bir diğer faktör ise modele verilecek olan görüntüler üzerinde önişlem yapılmak için modellerin kendi bünyesinde yer alan *preprocess_input* özelliğinin kullanılmasıdır. Yapılan deneysel çalışmalar sonucunda piksel değerlerinin değiştirilmesi için kullanılan farklı tekniklerin model başarısı üzerinde farklı sonuçlar doğurduğu ve *preprocess_input* özelliğinin kullanılmasının başarıyı olumlu yönde etkilediği görülmüştür.

Tablo 6-4 AUC değerlerinin karşılaştırılması

Çalışma	**Yöntem (Odak Noktası)**	**Veri Seti**	**AUC Puanı**
Afchar vd.	Mezoskopik özellikler	FaceForensics++	0,91
Li vd.	Göz kırpma sinyalleri	Özgün veri seti	0,99
Sabir vd.	Uzay zamansal özellikler	FaceForensics++	0,96
Li ve Lyu	Yüz çarpıtma yapıları	FaceForensics++	0,93
X. Yang vd.	Baş pozları	UADFV	0,89
Coccomini vd.	EfficientNet ve vision dönüştürücü birleşimi	DFDC	0,95
Korkmaz ve Alkan	Çerçeve tabanlı	DFDC	0,91
Yan vd.	Ortak Özellikler (ConvNet)	FaceForensics++	0,84
Mevcut çalışma	Çerçeve tabanlı	FaceForensics++	**0,93**

Tablo 6-4'te çalışmanın sonucu ile literatürdeki diğer çalışmaların sonuçlarının karşılaştırılması verilmiştir. Tablodaki değerler incelendiğinde çalışmanın kabul edilebilir ve benzer bir değer aralığında sonuç çıkardığı açıkça görülmektedir. Sadece AUC puanına göre karşılaştırma yapılmasının modelin başarımını kıyaslayabilmek için yeterli olmadığı düşünülmektedir. Ayrıca literatürde yer alan çalışmalarda kullanılan veri setlerinin farklılığı, çalışmaların odak noktası, başarı ölçüm metriklerinin farklılığı gibi faktörler çalışmaların kıyaslanmasını zorlaştırmaktadır. Tablo 6-4 incelendiğinde Li ve arkadaşlarının göz kırpma sinyallerine odaklanarak kendi oluşturdukları veri setinde 0,99 başarı oranı sağladığı görülmektedir. Güncel tekniklerle oluşturulan veri setlerinde göz kırpma sinyalleri gerçeğe yakın olarak yapıldığı için önerilen modelin bu veri setleri üzerinde başarımının düşmesi muhtemeldir.

Farklı çalışmaların başarı puanları incelendiğinde derin öğrenme mimarilerinin farklı odak noktaları kullanılarak kabul görür seviyede başarımlar elde ettiği görülmektedir. Bu sebeple farklı modellerin sonuçlarının kullanılarak birleştirme (ensemble) yöntemiyle sonuçlarının karar verme aşamasında değerlendirmeye alınmasının daha doğru olacağı düşünülmektedir.

Bu çalışma neticesinde elde edilen 0,93 AUC değeri, bu alanda kullanılabilirliğinin bir göstergesi olarak kabul edilmektedir. Deepfake medya tespiti için güncel veri setleri üzerinde farklı odak noktaları ile farklı ön eğitimli modeller kullanılarak farklı sonuçların ortaya çıkabileceği düşünülmekle birlikte kullanılan parametreler ve tekniklerin genel kabul görür sonuçlar elde ettiği açıkça görülmüştür. Çalışmanın sonucunda ön eğitimli modellerin az sayıda veri ile kısa dönem eğitimlerde bile yüksek başarı sağlayabileceği ve bu alanda kullanılabileceği görülmüştür.

7. SONUÇLAR

Bu çalışma kapsamında deepfake medyaların tespiti için 5 farklı ön eğitimli CNN modeli kullanılmıştır. Bu konu ile ilgili oluşturulmuş ve literatürde en çok tercih edilen veri setlerinden biri olan FaceForensics++ veri seti kullanılmıştır. Veri setinde bulunan videolardan çıkarılan yüz görüntüleri ile modeli eğitmek üzere veriler hazırlanmıştır. Çalışma kapsamında veri arttırma kullanımının veri çeşitliliğini arttırma açısından önemli olduğu deneysel çalışmalar sonucunda öğrenilmiştir. 5 model, farklı hiper parametre değerleri, yöntem ve fonksiyonlar kullanılarak eğitime tabi tutulmuştur.

Modellerin karmaşık ve derin ağ yapısı nedeniyle bazı eğitimler çok erken dönemde aşırı uydurma problemi ile karşı karşıya kalmıştır. Aşırı uydurma problemi modelin eğitim verilerini ezberleyerek doğrulama verileri üzerinde genelleme yapamaması durumu olarak tanımlanabilir. Çalışmada aşırı uydurma problemine karşı eğitim verileri üzerinde çeşitli veri arttırma teknikleri uygulanmıştır. Bu sayede veri çeşitliliği arttırılarak modellerin eğitim farklı eğitim verileri ile eğitilmesi sağlanmıştır.

Deneysel çalışmalar sonucunda 3 yöntem ile 5 model eğitilerek 15 farklı sonuç elde edilmiştir. Bu eğitimler sonucunda dışlama (dropout) katmanı olmadan 0,0001 öğrenme oranı ve 16 yığın boyutu ile 30 dönem boyunca eğitime tabi tutulan EfficientNetB4 modeli 0,93 AUC değeri en başarılı model olmuştur. Bu değerin ortaya çıktığı eğitim aşamasında model 20. dönemde erken durdurma yaparak model ağırlıkları en başarılı doğrulama kaybı değerinin görüldüğü 15. dönem ağırlıkları ile kaydedilmiştir. Çalışmanın sonucunda ön eğitimli modellerin deepfake medya tespitinde kullanılabileceği ve yüksek başarımlar elde edebileceğini göstermiştir.

EfficientNetB4 modeli haricinde kullanılan VGG16, DenseNet201, ResNet50V2 ve InceptionV3 modelleri de farklı parametrelerle 0,89 ve 0,90 gibi değerlerle kabul edilebilir sonuçlar yakalamıştır. Çalışma kapsamında kullanılan modeller haricinde birçok ön eğitimli model bulunmaktadır. Bu modeller kullanılarak daha farklı sonuçlar elde edilebilir. Çalışma kapsamında parametre değerlerinin değiştirilmesinin farklı modeller üzerinde farklı sonuçlar verebileceği ortaya çıkmıştır.

Eğitim sonucunda en başarılı model olan EfficientNetB4 için oluşturulan karışıklık matrisi değerleri incelendiğinde test veri setinde bulunan 864 gerçek görüntünün 730'u gerçek 134'ü deepfake olarak, 845 deepfake görüntünün 721'i deepfake 124'ü gerçek olarak sınıflandırıldığı görülmüştür. Çalışma kapsamında elde edilen başarım değerleri incelendiğinde Tablo6-4'te belirtilen, literatürde bulunan bu alandaki çalışmalar ile rekabet edebilir sonuçlar ortaya koymuştur.

8. ÖNERİLER

Deepfake medyaların derin öğrenme mimarileri kullanılarak tespitine ilişkin literatürde birçok farklı çalışma bulunmakla birlikte bu konuda çalışmaların yürütülebilmesi için oluşturulan veri seti sayısı oldukça azdır. Mevcut veri setlerinin birçoğunda sınıflar arası veri dağılımında dengesizlik bulunmakla birlikte bazı veri setleri sadece deepfake medyalardan oluşmaktadır. Bu konuda yürütülecek olan çalışmalarda en önemli konulardan biri veri setinin seçilmesi konusudur. Veri seti tercihi yapılırken benzersiz aktör sayısının fazla olması, videoların çözünürlüklerinin yüksek olması sınıflar arası veri dağılımı ve her sınıf için yeteri kadar veri bulunması gibi faktörler incelenerek en uygun olanı seçilmelidir.

Konu ile ilgili karşılaşılan zorluklardan bir diğeri derin öğrenme mimarilerinin başarımını etkileyen onlarca hiper parametrenin bulunmasıdır. Literatürde yer alan çalışmalarda kullanılan hiper parametrelerin etkisi incelenerek kullanılacak olan hiper parametrelerin sayısının azaltılması daha az deneysel çalışma ile en iyi sonucun yakalanması için doğru bir yaklaşım olabilmektedir. Dolayısıyla bu konuda yapılan çalışmaların incelenmesi araştırmacılara zaman kazandırabilmektedir.

Karşılaşılan bir diğer zorluk ise derin öğrenme mimarileri ile gerçekleştirilen eğitimler için yüksek donanıma sahip bilgisayarlar kullanılması gerekmektedir. Mevcut fiziksel donanımın yetersiz olması çalışma hızını doğrudan etkileyen bir etken olduğu için Google Colab platformunun ücretli versiyonlarının kullanılması araştırmacılara büyük kolaylık sağlamaktadır. Elde edilen tecrübeler ışığında Colab platformu için paket satın alınmasından ziyade kullandığın kadar öde (Pay As You Go) aracılığıyla hesap birimi satın alınmasının daha iyi bir seçenek olduğu gözlemlenmiştir.

Deepfake medya tespitinin zorluklarından birisi de farklı tespit çalışmaları ile yüksek başarı oranı sağlansa da her geçen gün farklı tekniklerle deepfake medyalar oluşturulmasıdır. Tespit çalışmaları genellikle eğitim yapıldığı deepfake türündeki sahte medyaları tespit etmekte başarı sağlamakla birlikte farklı türde oluşturulan deepfake medyalarda başarı değerleri azalmaktadır. Bu sebeple deepfake tespiti yapılırken farklı türde deepfake medyaları tespit eden çalışmaların tahminlerinin

hepsinin birleştirme (ensemble) yöntemiyle bir araya getirilerek değerlendirilmesi ve nihai kararın verilmesi daha doğru sonuçlar verecektir. Unutulmamalıdır ki tespit çalışmalarının sonuçları karar vermek için tek başına yeterli olmamakla birlikte uzmanların karar vermesine destek olmak amacıyla kullanılabileceği düşünülmektedir.

KAYNAKLAR

Afchar, D., Nozick, V., UPEM, F., Yamagishi, J., & Echizen, I. (2018). *MesoNet: a Compact Facial Video Forgery Detection Network*.

Alheeti, K. M. A., Al-Rawi, S. S., Khalaf, H. A., & Al Dosary, D. (2021). Image feature detectors for deepfake image detection using transfer learning. *2021 14th International Conference on Developments in eSystems Engineering (DeSE)*, 499-502.

Aramendia, A. I. (2024). *Convolutional Neural Networks (CNNs) : A Complete Guide*. https://medium.com/@alejandro.itoaramendia/convolutional-neural-networks-cnns-a-complete-guide-a803534a1930

Bakır, H., Demircioğlu, U., Bakır, R., & Adem, K. (2024). *A Deep Learning-Based Approach for Image Denoising: Harnessing Autoencoders for Removing Gaussian and Salt-Peper Noises*.

BasuMallick C. (2022, Mayıs 23). *What Is Deepfake? Meaning, Types of Frauds, Examples, and Prevention Best Practices for 2022*. https://www.spiceworks.com/it-security/cyber-risk-management/articles/what-is-deepfake/

Baş, N. (2006). *Yapay sinir ağları yaklaşımı ve bir uygulama*. Fen Bilimleri Enstitüsü.

Bayram, A. (2021). *GENERATİVE ADVERSARİAL NETWORKS (GAN) nedir*. https://alper-bayram.medium.com/generati%CC%87ve-adversari%CC%87al-networks-gan-nedir-f4ae346e679a

Belada, N. E. S. (2024). DEEPFAKE DEZENFORMASYONU. *Bilişim Hukuku Dergisi*, *6*(1).

Chen, M., Fridrich, J., Goljan, M., & Lukás, J. (2008). Determining image origin and integrity using sensor noise. *IEEE Transactions on information forensics and security*, *3*(1), 74-90.

Chollet, F. (2019). Python ile derin öğrenme. *Baskı ed. Buzdağı Yayınevi, Ankara*.

Cloud, H. (2011). The nist definition of cloud computing. *National institute of science and technology, special publication*, *800*(2011), 145.

Coccomini, D. A., Messina, N., Gennaro, C., & Falchi, F. (2022). Combining efficientnet and vision transformers for video deepfake detection. *International conference on image analysis and processing*, 219-229.

Çiçek, H. K., & Yalçın, N. (2024). Deepfake Bir Tehdit mi Fırsat mı? *Euroasia Journal of Mathematics, Engineering, Natural & Medical Sciences*, *11*(32), 31-46.

Dagar, D., & Vishwakarma, D. K. (2022). A literature review and perspectives in deepfakes: generation, detection, and applications. *International journal of multimedia information retrieval*, *11*(3), 219-289.

Dang, M., & Nguyen, T. N. (2023). Digital face manipulation creation and detection: A systematic review. *Electronics*, *12*(16), 3407.

Deepfake Studio. (2024). *Deepfake Studio*. https://play.google.com/store/apps/details?id=com.deepworkings.dfstudio&hl=tr

Deepfakesweb. (2024). *Deepfake Web | Kendi Deepfake'inizi Oluşturun! [Çevrimiçi Uygulama]*. https://deepfakesweb.com/

DFaker. (2018). *DFaker*. https://github.com/dfaker/df

Donahue, J., Anne Hendricks, L., Guadarrama, S., Rohrbach, M., Venugopalan, S., Saenko, K., & Darrell, T. (2015). Long-term recurrent convolutional networks for visual recognition and description. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2625-2634.

Ergüder, H. (2018). *Recurrent Neural Network Nedir*. https://medium.com/@hamzaerguder/recurrent-neural-network-nedir-bdd3d0839120

Ergün, E., & Kılıç, K. (2021). Derin öğrenme ile artırılmış görüntü seti üzerinden cilt kanseri tespiti. *Black Sea Journal of Engineering and Science*, *4*(4), 192-200.

Exem. (2022). *Bölüm 2. Numpy Bölüm 1: Time Series makine öğrenimi için temel Python kütüphanesi*. https://blog.ex-em.com/1693

FaceApp. (2024). *FaceApp: Yüz Editörü*. https://www.faceapp.com/

FaceHub. (2022). *FaceHub*. https://play.google.com/store/apps/details?id=com.facepeer.facehub&hl=tr

Faceswap-GAN. (2018). *Generative adversarial networks for face swapping*. https://github.com/shaoanlu/faceswap-GAN

Feizabadi, M., Işık, M. S., & İpbüker, C. (2015). Geomatik Mühendisliği Uygulamalarında Dönüşüm Yöntemleri. *TMMOB Harita ve Kadastro Mühendisleri Odası*, *13*.

Feng, L., Po, L.-M., Xu, X., Li, Y., & Ma, R. (2014). Motion-resistant remote imaging photoplethysmography based on the optical properties of skin. *IEEE Transactions on Circuits and Systems for Video Technology*, *25*(5), 879-891.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative adversarial nets. *Advances in neural information processing systems*, *27*.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2020). Generative adversarial networks. *Communications of the ACM*, *63*(11), 139-144.

Guarnera, L., Giudice, O., & Battiato, S. (2020). Deepfake detection by analyzing convolutional traces. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, 666-667.

Jia, Y., Zhang, Y., Weiss, R., Wang, Q., Shen, J., Ren, F., Nguyen, P., Pang, R., Lopez Moreno, I., & Wu, Y. (2018). Transfer learning from speaker verification to multispeaker text-to-speech synthesis. *Advances in neural information processing systems*, *31*.

Karakoç, E., & Zeybek, B. (2022). GÖRMEK İNANMAYA YETER Mİ? GÖRSEL DEZENFORMASYONUN AYIRT EDİCİ BİÇİMİ OLARAK SİYASİ DEEPFAKE İÇERİKLER. *Öneri Dergisi*, *17*(57), 50-72.

Karthick, R., Dawood, M. S., & Meenalochini, P. (2023). Analysis of vital signs using remote photoplethysmography (RPPG). *Journal of Ambient Intelligence and Humanized Computing*, *14*(12), 16729-16736.

Keras. (2015). *Keras Uygulamaları*. https://keras.io/api/applications/

Kırık, A. M., & Özkoçak, V. (2023). MEDYA VE İLETİŞİM BAĞLAMINDA YAPAY ZEKÂ TARİHİ VE TEKNOLOJİSİ: CHATGPT VE DEEPFAKE İLE GELEN DİJİTAL DÖNÜŞÜM. *Karadeniz Uluslararası Bilimsel Dergi*, *58*, 73-99.

Kızrak, M. A., & Bolat, B. (2018). Derin öğrenme ile kalabalık analizi üzerine detaylı bir araştırma. *Bilişim Teknolojileri Dergisi*, *11*(3), 263-286.

Kingma, D. P., & Welling, M. (2019). An introduction to variational autoencoders. *Foundations and Trends® in Machine Learning*, *12*(4), 307-392.

Kirchengast, T. (2020). Deepfakes and image manipulation: criminalisation and control. *Information & Communications Technology Law*, *29*(3), 308-323.

Kizrak, A. (2019). *Udemy_DerinOgrenmeyeGiriş*. https://github.com/ayyucekizrak/Udemy_DerinOgrenmeyeGiris/blob/master/TransferOgrenme_FineTuning/ReadMe.md

Korkmaz, Ş., & Alkan, M. (2023). Derin öğrenme algoritmalarını kullanarak deepfake video tespiti. *Politeknik Dergisi*, *26*(2), 855-862.

Li, Y., Chang, M.-C., & Lyu, S. (2018a). In ictu oculi: Exposing ai created fake videos by detecting eye blinking. *2018 IEEE International workshop on information forensics and security (WIFS)*, 1-7.

Li, Y., Chang, M.-C., & Lyu, S. (2018b). In ictu oculi: Exposing ai created fake videos by detecting eye blinking. *2018 IEEE International workshop on information forensics and security (WIFS)*, 1-7.

Li, Y., & Lyu, S. (2018). *Exposing DeepFake Videos By Detecting Face Warping Artifacts*. http://arxiv.org/abs/1811.00656

Maksutov, A. A., Morozov, V. O., Lavrenov, A. A., & Smirnov, A. S. (2020). Methods of deepfake detection based on machine learning. *2020 IEEE conference of russian young researchers in electrical and electronic engineering (EIConRus)*, 408-411.

Moğulkoç, D. (2024). *Sinir Ağı Kulübü: Derin Öğrenme*. https://denizmogulkoc.medium.com/the-neural-network-club-deep-learning-cabe7013b691

Nguyen, H. H., Yamagishi, J., & Echizen, I. (2018). Capsule-Forensics: Using Capsule Networks to Detect Forged Images and Videos. *arXiv preprint arXiv:1810.11215*.

Nguyen, H. H., Yamagishi, J., & Echizen, I. (2019). Use of a capsule network to detect fake images and videos. *arXiv preprint arXiv:1910.12467*.

Nguyen, T. T., Nguyen, Q. V. H., Nguyen, D. T., Nguyen, D. T., Huynh-The, T., Nahavandi, S., Nguyen, T. T., Pham, Q.-V., & Nguyen, C. M. (2022). Deep learning for deepfakes creation and detection: A survey. *Computer Vision and Image Understanding*, *223*, 103525.

Nirkin, Y., Masi, I., Tuan, A. T., Hassner, T., & Medioni, G. (2018). On face segmentation, face swapping, and face perception. *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*, 98-105.

Öngün C. (2020, Nisan 21). *Autoencoder (Otokodlayıcı) nedir? Ne için kullanılır?* https://cihanongun.medium.com/autoencoder-otokodlay%C4%B1c%C4%B1-nedir-ne-i%C3%A7in-kullan%C4%B1l%C4%B1r-e520a591746a

Prajwal, K. R., Mukhopadhyay, R., Namboodiri, V. P., & Jawahar, C. V. (2020). A lip sync expert is all you need for speech to lip generation in the wild. *Proceedings of the 28th ACM international conference on multimedia*, 484-492.

Psikolog. (2023). *Davranışın Nörobiyolojik Temelleri - Ruh Bilimi - Psikoloji*. https://ruhbilimi.gen.tr/davranisin-norobiyolojik-temelleri/

Rana, M. S., Murali, B., & Sung, A. H. (2021). Deepfake Detection Using Machine Learning Algorithms. *2021 10th International Congress on Advanced Applied Informatics (IIAI-AAI)*, 458-463.

Rayaguru, B. (2023). *Affine Align Transformations: A Practical Guide to Image Alignment and Transformation*. https://medium.com/@babykrishna/affine-align-transformations-a-practical-guide-to-image-alignment-and-transformation-8844bff2aefd

Raza, M. A., & Malik, K. M. (2023). *Multimodaltrace: Deepfake Detection Using Audiovisual Representation Learning* (ss. 993-1000).

Reface. (2024). *Reface – AI Face Swap App & Video Face Swaps*. https://reface.ai/

Rosunee, S. (2021). Leveraging artificial intelligence to foster innovation and inclusive growth in the textile value chain. *Recent Trends in Traditional and Technical Textiles: Select Proceedings of ICETT 2019*, 33-38.

Sabir, E., Cheng, J., … A. J.-I., & 2019, undefined. (2019). Recurrent convolutional strategies for face manipulation detection in videos. *openaccess.thecvf.com*. http://openaccess.thecvf.com/content_CVPRW_2019/papers/Media%20Forensics/Sabir_Recurrent_Convolutional_Strategies_for_Face_Manipulation_Detection_in_Videos_CVPRW_2019_paper.pdf

Shaonlu. (2018, Haziran 6). *faceswap-GAN*. https://github.com/shaoanlu/faceswap-GAN

Şahinaslan, E., Günerkan, M., & Şahinaslan, Ö. (2023). Makine Öğrenmesinde Kategorik Veri Kodlama Tekniğinin Kullanımına Alternatif Bir Çözüm Yöntemi. *Journal of Intelligent Systems: Theory and Applications*, *6*(1), 1-11.

Şeker, A., Diri, B., & Balık, H. H. (2017). Derin öğrenme yöntemleri ve uygulamaları hakkında bir inceleme. *Gazi Mühendislik Bilimleri Dergisi*, *3*(3), 47-64.

Şeker, S. E. (2015). Doğal Dil İşleme (Natural Language Processing). *YBS Ansiklopedi*, *2*(4), 14-31.

Theiler, S. (2019). *The Intuition Behind Voice Cloning (SV2TTS) | Analytics Vidhya*. https://medium.com/analytics-vidhya/the-intuition-behind-voice-cloning-with-5-seconds-of-audio-5989e9b2e042

Thies, J., Zollhofer, M., Stamminger, M., Theobalt, C., & Nießner, M. (2016a). Face2face: Real-time face capture and reenactment of rgb videos. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2387-2395.

Thies, J., Zollhofer, M., Stamminger, M., Theobalt, C., & Nießner, M. (2016b). Face2face: Real-time face capture and reenactment of rgb videos. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2387-2395.

TRUBA. (2003). *Türk Ulusal Bilim e-Altyapısı (TRUBA)*. https://www.truba.gov.tr/index.php/truba-olusumu/

Tunçer, C. (2024, Şubat 5). *Bir şirket "deepfake" yüzünden tam 25 milyon dolar kaybetti.* https://www.log.com.tr/bir-sirket-deepfake-yuzunden-tam-25-milyon-dolar-kaybetti/

Türkmenoglu, C., & Tantug, A. C. (2014). Sentiment analysis in Turkish media. *International Conference on Machine Learning (ICML)*.

Vatansever, S., & Dirik, A. E. (2023). Farklı çözünürlükteki sayısal imge ve videolar için PRNU tabanlı kaynak kamera tespiti üzerine bir çalışma. *Niğde Ömer Halisdemir Üniversitesi Mühendislik Bilimleri Dergisi*, *12*(3), 692-698. https://doi.org/10.28948/ngumuh.1253242

Vikipedi. (2024). *Sinir hücresi*. https://tr.wikipedia.org/wiki/Sinir_h%C3%BCcresi

Voicery. (2024). *Seslendirme Metinden Konuşmaya*. https://www.voicery.com/

Wang, J., Wu, Z., Ouyang, W., Han, X., Chen, J., Lim, S. N., & Jiang, Y. G. (2022). M2TR: Multi-modal Multi-scale Transformers for Deepfake Detection. *ICMR 2022 - Proceedings of the 2022 International Conference on Multimedia Retrieval*, 615-623. https://doi.org/10.1145/3512527.3531415/SUPPL_FILE/ICMR22-158.MP4

Wang, P., Li, W., Ogunbona, P., Wan, J., & Escalera, S. (2018). RGB-D-based human motion recognition with deep learning: A survey. *Computer vision and image understanding*, *171*, 118-139.

Wav2Lip. (2020). *Wav2Lip: Accurately Lip-syncing Videos In The Wild*. https://github.com/Rudrabha/Wav2Lip

Wavel AI. (2024). *Wavel AI | Best Text-to-Speech Voice Solutions for Videos And Localization*. https://wavel.ai/

Wikimedia. (2022). *Imitation-based approach*. https://commons.wikimedia.org/wiki/File:Imitation-based_approach.png

Xiao, H., Liu, T., Sun, Y., Li, Y., Zhao, S., & Avolio, A. (2024). Remote photoplethysmography for heart rate measurement: A review. *Biomedical Signal Processing and Control*, *88*, 105608.

Yan, Z., Zhang, Y., Fan, Y., & Wu, B. (2023). Ucf: Uncovering common features for generalizable deepfake detection. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 22412-22423.

Yang, G., Fei, N., Ding, M., Liu, G., Lu, Z., & Xiang, T. (2021). L2m-gan: Learning to manipulate latent space semantics for facial attribute editing. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2951-2960.

Yang, X., Li, Y., & Lyu, S. (2019). Exposing deep fakes using inconsistent head poses. *ieeexplore.ieee.orgX Yang, Y Li, S LyuICASSP 2019-2019 IEEE International Conference on Acoustics, 2019•ieeexplore.ieee.org*. https://ieeexplore.ieee.org/abstract/document/8683164/

Yeh, C.-Y., Chen, H.-W., Tsai, S.-L., & Wang, S.-D. (2020). Disrupting image-translation-based deepfake algorithms with adversarial attacks. *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision Workshops*, 53-62.

Yılmaz, G. K. (2024). Siyasi Hicvin Evrimi: Siyasi Liderlerin YouTube'daki Deepfake Videolarının Göstergebilimsel Analizi. *REFLEKTİF Sosyal Bilimler Dergisi*, *5*(3), 837-857.

Yu, P., Xia, Z., Fei, J., & Lu, Y. (2021). A survey on deepfake video detection. *Iet Biometrics*, *10*(6), 607-624.

Yurdigül, Y., & Yıldırım, A. (2021). Gerçeklik algısına bir müdahale aracı olarak sentetik medya teknolojileri. *İletişim ve Diplomasi*, *5*, 105-121.

Zao. (2019). *Zao*. https://zaodownload.com/

Zhang, J., Ni, J., & Nie, F. (2024). DSM: Domain Shift Modeling for general deepfake detection. *Signal Processing*, *230*, 109822. https://doi.org/10.1016/J.SIGPRO.2024.109822

Zhao, H., Zhou, W., Chen, D., Wei, T., Zhang, W., & Yu, N. (2021). *Multi-Attentional Deepfake Detection* (ss. 2185-2194).