T.C.
FIRAT ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

ÖZ DENETİMLİ ÖĞRENME YAKLAŞIMLARI İLE DERİN SAHTE SES VE GÖRÜNTÜ MANİPÜLASYONUNUN TESPİTİ

Merve YILDIRIM

Yüksek Lisans Tezi

BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI

TEMMUZ 2025

T.C.
FIRAT ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

Bilgisayar Mühendisliği Anabilim Dalı

Yüksek Lisans Tezi

ÖZ DENETİMLİ ÖĞRENME YAKLAŞIMLARI İLE DERİN SAHTE SES VE GÖRÜNTÜ MANİPÜLASYONUNUN TESPİTİ

Tez Yazarı
Merve YILDIRIM

Danışman
Prof. Dr. İlhan AYDIN

TEMMUZ 2025
ELAZIĞ

T.C.
FIRAT ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

Bilgisayar Mühendisliği Anabilim Dalı

Yüksek Lisans Tezi

Başlığı:	Öz Denetimli Öğrenme Yaklaşımları ile Derin Sahte Ses ve Görüntü Manipülasyonunun Tespiti
Yazarı:	Merve YILDIRIM
İlk Teslim Tarihi:	17.06.2025
Savunma Tarihi:	18.07.2025

TEZ ONAYI

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

İmza

Danışman: Prof. Dr. İlhan AYDIN — Onayladım
Fırat Üniversitesi, Mühendislik Fakültesi

Başkan: Prof. Dr. Mehmet KARAKÖSE — Onayladım
Fırat Üniversitesi, Mühendislik Fakültesi

Üye: Dr. Öğr. Üyesi Mehmet Umut SALUR — Onayladım
Gaziantep İslam Bilim ve Teknoloji Üniversitesi, Mühendislik ve Doğa Bilimleri Fakültesi

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

İmza

Prof. Dr. Burhan ERGEN
Enstitü Müdürü

BEYAN

Fırat Üniversitesi Fen Bilimleri Enstitüsü tez yazım kurallarına uygun olarak hazırladığım "Öz Denetimli Öğrenme Yaklaşımları ile Derin Sahte Ses ve Görüntü Manipülasyonunun Tespiti" Başlıklı Yüksek Lisans Tezimin içindeki bütün bilgilerin doğru olduğunu, bilgilerin üretilmesi ve sunulmasında bilimsel etik kurallarına uygun davrandığımı, kullandığım bütün kaynakları atıf yaparak belirttiğimi, maddi ve manevi desteği olan tüm kurum/kuruluş ve kişileri belirttiğimi, burada sunduğum veri ve bilgileri unvan almak amacıyla daha önce hiçbir şekilde kullanmadığımı beyan ederim.

18.07.2025

Merve YILDIRIM

ÖNSÖZ

Günümüzde, derin sahte teknolojisi medya içeriğini manipüle ederek güvenlik ve doğruluk açısından ciddi tehditler oluşturmaktadır. Özellikle sosyal medya platformlarının yaygın kullanımıyla birlikte, derin sahte içerikleri hızla yayılmakta ve toplumsal, siyasi ve etik sorunlara yol açmaktadır. Bu bağlamda, derin sahte teknolojisinin tespiti ve önlenmesi büyük önem taşımaktadır.

Bu tez, derin sahte ses ve görüntü manipülasyonlarını tespit etmeye yönelik derin öğrenme tabanlı modeller geliştirmeyi amaçlamaktadır. Çalışma, medya güvenliğini koruma ve doğru bilgi akışını sağlamaya katkıda bulunmayı hedeflemektedir.

Yüksek Lisans sürecim boyunca, rehberliği ve akademik desteğiyle çalışamama yön veren değerli danışmanım Prof. Dr. İlhan Aydın'a, teşekkür ederim.

Bu süreçte ve hayatımın her anında yanımda olan aileme teşekkür ederim.

Merve YILDIRIM
ELAZIĞ, 2025

İÇİNDEKİLER

Sayfa

ÖZET

Öz Denetimli Öğrenme Yaklaşımları ile Derin Sahte Ses ve Görüntü Manipülasyonunun Tespiti

Merve YILDIRIM

Yüksek Lisans Tezi

FIRAT ÜNİVERSİTESİ
Fen Bilimleri Enstitüsü

Bilgisayar Mühendisliği Anabilim Dalı

Temmuz 2025, Sayfa: xiii + 55

Derin sahte teknolojisi, ses ve görüntü manipülasyonları yoluyla hızla yaygınlaşan ve ciddi güvenlik riskleri taşıyan bir tehdit unsuru haline gelmiştir. Bu tür sahte içeriklerin tespiti, medya güvenilirliğinin korunması ve sahte içeriklerin toplum üzerindeki olumsuz etkilerinin önlenmesi açısından kritik bir öneme sahiptir. Bu tez çalışmasında, derin öğrenme tabanlı iki farklı yaklaşım kullanılarak derin sahte video içeriklerinin tespiti amaçlanmıştır. İlk yaklaşımda öz-denetimli bir yöntem uygulanmıştır. Bu yöntemde, "deepfake-and-real-images" veri seti kullanılarak, ResNet18 tabanlı bir RotationNet modeliyle görüntülerin döndürülme açısını tahmin etmeye yönelik ön eğitim gerçekleştirilmiş, sonrasında bu modelin özellik çıkarıcı katmanları sabitlenerek gerçek/sahte sınıflandırması için ince ayar yapılmıştır. İkinci yaklaşımda ise çoklu modaliteye dayalı bir derin öğrenme yaklaşımı önerilmektedir. Çalışmada, yaygın olarak kullanılan DeepFake Detection Challenge veri setinden alınan videolar kullanılmıştır. Önerilen yöntemde, video akışlarından yüz görüntülerinin elde edilmesi için MTCNN tabanlı bir yüz tespit ve hizalama yöntemi kullanılmıştır. Bu yüz görüntüleri, daha sonra CNN ve LSTM tabanlı modele girdi olarak verilmiştir. Ses akışlarından çıkarılan Mel spektrogramları ise bir CNN tabanlı modelle analiz edilmiştir. Her iki modaliteden elde edilen özellikler ile bu özelliklerin birleşimini işleyen görsel-işitsel füzyon modeli geliştirilmiştir. Görüntü, ses ve füzyon modellerinden elde edilen tahminler, öğrenilebilir ağırlıklarla birleştirilerek bir ensemble model oluşturulmuş ve nihai karar bu model üzerinden verilmiştir.

Yapılan deneysel çalışmalar sonucunda, geliştirilen öz-denetimli yaklaşımın derin sahte videoları, %88.89 doğruluk, %83.15 geri çağırma ve %88.10 F1 skoru ile tespit ettiği gözlenmiştir. Çok modlu modelin ise %92.17 doğruluk, %90.36 geri çağırma ve %92.02 F1 skoru ile başarılı bir şekilde tespit edebildiği gözlemlenmiştir. Bu çalışmalar, kullanılan yöntemlerin derin sahte tespitindeki etkinliğini ortaya koyarak, bu alandaki çalışmalara katkı sunmayı hedeflemektedir.

Anahtar Kelimeler: Derin sahte tespiti, Derin öğrenme, Öz-denetimli öğrenme, Ses ve görüntü manipülasyonu

ABSTRACT

Detection of Deep Fake Audio and Image Manipulation with Self-Supervised Learning Approaches

Merve YILDIRIM

Master's Thesis

FIRAT UNIVERSITY
Graduate School of Natural and Applied Sciences

Department of Computer Engineering

July 2025, Pages: xiii + 55

Deepfake technology has become a rapidly spreading threat through audio and video manipulations and poses serious security risks. Detection of such fake content is of critical importance in terms of preserving media credibility and preventing the negative effects of fake content on society. In this thesis, it is aimed to detect deepfake video content using two different deep learning-based approaches. In the first approach, a self-supervised method is applied. In this method, pre-training is performed to estimate the rotation angle of the images with a ResNet18-based RotationNet model using the "deepfake-and-real-images" dataset, and then the feature extractor layers of this model are fixed and fine-tuned for real/fake classification. In the second approach, a multi-modality deep learning approach is proposed. In the study, videos taken from the widely used DeepFake Detection Challenge dataset are used. In the proposed method, an MTCNN-based face detection and alignment method is used to obtain face images from video streams. These face images are then given as input to the CNN and LSTM-based model. Mel spectrograms extracted from audio streams were analyzed with a CNN-based model. An audiovisual fusion model was developed that processes features derived from both modalities and their combination. Predictions from the image, audio, and fusion models were combined with learnable weights to create an ensemble model, and the final decision was made based on this model.

As a result of the experimental studies, it was observed that the developed self-supervised approach detected deepfake videos with 88.89% accuracy, 83.15% recall and 88.10% F1 score. It was observed that the multimodal model could successfully detect 92.17% accuracy, 90.36% recall and 92.02% F1 score. These studies aim to contribute to the studies in this field by revealing the effectiveness of the methods used in deepfake detection.

Keywords: Deepfake detection, Deep learning, Self-supervised learning, Audio and image manipulation

ŞEKİLLER LİSTESİ

Sayfa

TABLOLAR LİSTESİ

Sayfa

SİMGELER

α	: Ek kayıp terimlerini dengeleyen hiperparametre
e^{-x}	: Çıkış fonksiyonu için standart üstel fonksiyonu
L_{audio}	: Ses tabanlı modelin kayıp fonksiyonu
L_{av}	: Ses-görüntü birleşik kayıp fonksiyonu
$L_{ensemble}$	: Ağırlıklı kayıplardan elde edilen bileşik kayıp fonksiyonu
L_{total}	: Nihai toplam kayıp fonksiyonu
L_{video}	: Görüntü tabanlı modelin kayıp fonksiyonu
w_0, w_1, w_2	: Kayıp bileşenleri için ağırlık katsayıları
W_i	: Ağ katmanlarında öğrenilen ağırlık parametre kümesi

KISALTMALAR

AUC	: Area Under Curve
ASV	: Automatic Speaker Verification
CEW	: Closed Eye in the Wild
Celeb-DF	: Celebrity DeepFake Veri Seti
CNN	: Convolutional Neural Networks
DNN	: Deep Neural Networks
DFDC	: Deepfake Detection Challenge
DT	: Decision Trees
ERR	: Equal Error Rate
FBank	: Filter Bank
FFHQ	: (Flickr-Faces-HQ)
GAN	: Generative Adversarial Networks
HQ	: High Quality
LA	: Logical Access
LBP	: Local Binary Pattern
LCNN	: Light Convolutional Neural Network
LQ	: Low Quality
LSTM	: Long Short-Term Memory
MDS	: Multi-dimensional Scaling
MHA	: Multi-Head Attention
MFCC	: Mel-frequency Cepstral Coefficients
MesoNet	: Mesoscopic Network
MTCNN	: Multi-task Cascaded Convolutional Networks
OpenCV	: Open Source Computer Vision Library
PA	: Physical Access
PCA	: Principal Component Analysis
ProGAN	: Progressive Growing of GANs
RFF	: Real Fake Face
RFFD	: Real and Fake Face Detection
ResNet	: Residual Network
RGB	: Red, Green, Blue
SCNN	: Set Convolutional Neural Networks

StyleGAN3	: Style-Based Generative Adversarial Network 3
SVM	: Support Vector Machines
TCN	: Temporal Convolutional Network
ViT	: Vision Transformer
VGG16	: Visual Geometry Group 16-layer Network
wCNN	: Width-Extended Convolutional Neural Network
WaveletCNN	: Wavelet Convolutional Neural Network
YCbCr	: Y (Luminance), Cb (Chroma Blue), Cr (Chroma Red)

1. GİRİŞ

Yapay zeka ve derin öğrenme, günümüzün en hızlı gelişen teknolojilerinden biri olup, dijital teknolojilerdeki bu gelişmeler, medya üretimi ve tüketimi üzerindeki etkilerini her geçen gün daha fazla hissettirmektedir. Son yıllarda, bu teknolojilerin ses ve görüntü işleme alanındaki uygulamaları dikkat çekici bir şekilde yaygınlaşmış ve gelişmiştir.

Deepfake (derin sahte) olarak bilinen ve gerçekçi sahte içeriklerin üretimini mümkün kılan yöntemler, ses ve görüntü manipülasyonlarında önemli bir etki yaratmıştır. Bu terim, genellikle bir kişinin yüzünü veya sesini, mevcut bir görüntü veya videodaki başka bir kişinin yüzü veya sesiyle değiştirmek ya da tamamen sentetik olarak yeni, gerçekçi görünen ama var olmayan insan yüzleri ve sesleri oluşturmak için kullanılan yapay zeka tabanlı teknikleri ifade eder. Derin sahte teknolojisi ilk olarak 2017 yılının sonlarında internet platformlarında ortaya çıkmıştır. [1] Kısa sürede büyük bir hızla gelişmiş ve yaygınlaşmıştır. Bu gelişmeler hem olumlu hem de olumsuz sonuçlar doğurmuştur. Başlangıçta eğlence amaçlı veya zararsız görünen uygulamalarla gündeme gelse de, teknolojinin erişilebilirliğinin artması ve kullanımının kolaylaşmasıyla birlikte kötü niyetli kullanım potansiyeli de ortaya çıkmıştır. Olumlu yönleri arasında yaratıcı projelere katkı sağlama, eğitim materyalleri geliştirme ve görsel efekt teknolojilerini ileriye taşıma gibi çeşitli faydalar sağlamaktadır. Bunun yanı sıra, tarihi olayların canlandırılması, dil bariyerlerini aşmak için senkronize çeviri içeriklerinin oluşturulması veya kaybolmuş kültürel eserlerin yeniden canlandırılması gibi faydalı uygulamalara sahiptir. Olumsuz yönleri ele alındığında ise, kötüye kullanımı ciddi tehditler yaratmaktadır. Özellikle sahte içeriklerin üretilmesi ve bu içeriklerin hızla yayılması, medya manipülasyonlarına, siyasi müdahalelere ve toplumsal güvenlik risklerine zemin hazırlamaktadır. Yanlış bilgilendirme, bireylerin itibarını zedeleme ve sosyal kutuplaşmayı tetikleme gibi çeşitli olumsuz sonuçlara yol açabilmektedir. [2] Örneğin derin sahte videoları, politik figürler veya kamuya mal olmuş kişiler hakkında yanıltıcı içerik oluşturmak için kullanılabilmekte ve bu durum, toplumda kaosa ve güvensizliğe neden olabilmektedir. Bir diğer örnek ise, bireylerin özel hayatlarına zarar verebilecek şekilde kötüye kullanılabilmekte ki bu da, kişilerin rızası olmadan oluşturulan sahte videolar, psikolojik travmalara ve sosyal itibar kaybına yol açabilmektedir.

Bu bağlamda bakıldığında, güvenilir ve etkili derin sahte tespit yöntemlerine olan ihtiyaç gün geçtikçe artmaktadır ve bu teknolojinin denetlenmesinin ve kontrol altına alınmasının ne derece önemli olduğu görülmektedir. Buna dayanarak, bu tür içeriklerin hızlı ve doğru bir şekilde tespit edilmesi, yalnızca dijital dünyanın güvenliğini sağlamak açısından değil, aynı zamanda bireysel mahremiyetin korunması ve toplumsal düzenin sürdürülebilirliği için kritik bir gereklilik haline gelmiştir. Derin sahte içeriklerin zarara yol açacak etkilerinin önlenmesi adına bu içeriklerin

tespitine yönelik yöntemlerin geliştirilmesi ve bu yöntemlerin yaygınlaştırılması büyük bir öneme sahiptir**.**

1.1. Tezin Amacı ve Kapsamı

Son yıllarda yapay zeka destekli içerik üretme teknolojilerinin gelişimiyle birlikte, gerçekçi görünen ama tamamen sahte olan video ve ses dosyaları yaygınlaşmıştır. Bu durum, özellikle bilgi güvenliği, medya manipülasyonu ve toplum güvenliği açısından ciddi tehditler oluşturmaktadır. Derin sahte tespitinde kullanılan derin öğrenme algoritmaları, bu alandaki en kritik teknolojilerden biri olarak öne çıkmakta ve kötüye kullanımı minimize etmek için umut verici bir çözüm sunmaktadır. Bu algoritmaların etkin bir şekilde uygulanması, hem bireysel hem de kurumsal dijital güvenliğin artırılmasına olanak tanırken, doğru tespit edilen sahte içerikler, toplumu yanıltacak gerçek olmayan bilgilerin yayılmasını engelleyerek, medya platformlarının güvenilirliğini hem korur hem de güçlendirir. Bu bağlamda, derin sahte tespiti, toplumsal güvenliği sağlamak için büyük bir gereklilik haline gelmiştir.

Bu tez çalışmasının temel amacı, derin sahte tespit problemini ele almak üzere iki farklı derin öğrenme metodolojisi geliştirmek, uygulamak ve bu metodolojilerin performansını analiz etmektir. Çalışmanın kapsamı, öz-denetimli yaklaşım ve çok modlu yaklaşım olarak bu iki temel yaklaşım etrafında şekillenmektedir.

İlk yöntemde, etiketsiz görüntü verilerinden anlamlı temsiller öğrenmeyi amaçlayan öz-denetimli öğrenme (self-supervised learning) stratejisi kullanılmıştır. ResNet18 mimarisi tabanlı bir omurga modeli oluşturularak ve bu model RotationNet yöntemi ile görüntülerin döndürülme açıları tahmin edilerek ön eğitim gerçekleştirilmiştir. Bu sayede model, herhangi bir etikete ihtiyaç duymadan görsel temsilleri öğrenir. Daha sonra, bu modelin özellik çıkarıcı katmanları sabitlenerek gerçek/sahte sınıflandırması yapılmıştır.

İkinci yöntem olan çok modlu (multimodal) yaklaşımda, hem görsel hem de işitsel verileri birlikte işleyen bir derin öğrenme mimarisi geliştirilmiştir. Önerilen bu sistem, sahte içeriklerin daha güvenilir şekilde tespit edilmesini hedeflerken, görüntü ve ses içeriklerinden elde edilen verileri birleştirerek detaylı bir analiz sunmuştur. Çalışma kapsamında, öncelikle video içeriklerindeki görsel manipülasyonları yakalamak için yüz bölgelerinden elde edilen verileri işleyen, Evrişimli Sinir Ağları (CNN) ve Uzun Kısa Süreli Bellek (LSTM) ağlarını temel alan bir video analiz modeli tasarlanmıştır. Yüz görüntülerinin çıkarılmasında doğruluğu yüksek ve literatürde yaygın olarak kullanılan MTCNN (Multi-task Cascaded Convolutional Networks) yüz tespit algoritmasından faydalanılmıştır. Aynı şekilde, ses verisi üzerinden oluşturulan Mel spektrogramları, görsel forma dönüştürülerek CNN tabanlı ses sınıflandırma modeline girdi olarak sunulmuştur. Çoklu modalitelerin (görsel ve işitsel) ayrı ayrı işlenmesiyle elde edilen özellikler, daha sonra ensemble (topluluk) öğrenme temelli bir füzyon modeli ile birleştirilmiş ve nihai karar

bu model üzerinden verilmiştir. Bu yaklaşım sayesinde, her bir modalitenin güçlü yönlerinden faydalanılarak, daha güvenilir bir karar mekanizması oluşturulması amaçlanmıştır.

Tez kapsamında geliştirilen iki farklı yaklaşımın hedeflerine ve teknik gereksinimlerine uygun olarak, bu alanda kullanılan iki ayrı veri seti kullanılmıştır. Öz-denetimli yaklaşım için "deepfake-and-real-images" veri seti kullanılmıştır. Bu veri seti, statik görüntülerdeki ince manipülasyon izlerini ve temel görsel özellikleri öğrenmek için zengin bir kaynak sunar. Çok modlu model için kullanılan veri seti, bu alanda yaygın olarak kullanılan ve çeşitli manipülasyon teknikleri içeren Deepfake Detection Challenge (DFDC) veri setinin bir alt kümesinden oluşmaktadır. Veri seti, gerçek ve sahte videolardan oluşmakta ve hem yüz ifadeleri hem de ses örüntüleri açısından zengin bir çeşitlilik sunmaktadır.

Model performansları, doğruluk (accuracy), geri çağırma (recall), F1 skoru gibi yaygın değerlendirme metrikleri ile ölçülerek, değerlendirilecektir.

Ayrıca, bu modellerin performansını karşılaştırmalı olarak değerlendirebilmek amacıyla, popüler derin sahte veri setlerinden biri olan Celeb-DF üzerinde görüntü tabanlı tek modlu ek bir deneysel çalışma gerçekleştirilmiştir.

Bu tez çalışması, hem öz-denetimli yaklaşım hem de çok modlu yaklaşımdan elde edilecek başarılı sonuçlarla derin sahte tespitindeki etkinliğini ortaya koyarak, akademik literatüre kavramsal ve teknik katkılar sunmayı hedeflenmektedir. Bunun yanı sıra, sahte içeriklerin otomatik olarak tespit edilmesine yönelik güvenilir sistemlerin geliştirilmesine katkıda bulunmayı ve medya güvenliği, dijital doğrulama sistemleri, adli bilişim ve içerik denetimi gibi pratik uygulama alanları için de somut bir temel oluşturmayı hedeflemektedir. Bu yönüyle tez, hem araştırma dünyasına hem de medya endüstrisine, sosyal medya platformlarına, kamu güvenliği politikaları gibi birçok alandaki uygulamalara katkı sağlayabilecektir.

1.2. Tezin Yapısı

Tezin birinci bölümünde, derin sahte kavramı tanımlanmış; bu teknolojinin toplumsal etkileri ile araştırmanın amacı ve kapsamı ele alınmıştır. Aynı zamanda çalışmanın genel yapısı ve temel problemine dair açıklamalar yapılmıştır. Tez yapısı şu şekilde devam edecektir:

İkinci bölümde, derin sahte tespiti ve derin öğrenme algoritmaları ile ilgili literatür taraması yapılacak, bu alanda yapılmış önceki çalışmalar değerlendirilecektir. Literatür incelemesi, derin sahte teknolojisinin gelişimi, kullanılan derin öğrenme yöntemleri ve bu yöntemlerin doğruluk oranları üzerine yapılmış araştırmalar ele alınacaktır.

Üçüncü bölümde, tezin temelini oluşturan derin öğrenme yönteminin temel prensipleri ele alınacak ve nasıl çalıştığına dair temel bilgiler sunulacaktır. Aynı zamanda Evrişimli Sinir Ağı ve Uzun Kısa Süreli Bellek modeli detaylı bir şekilde incelenecektir.

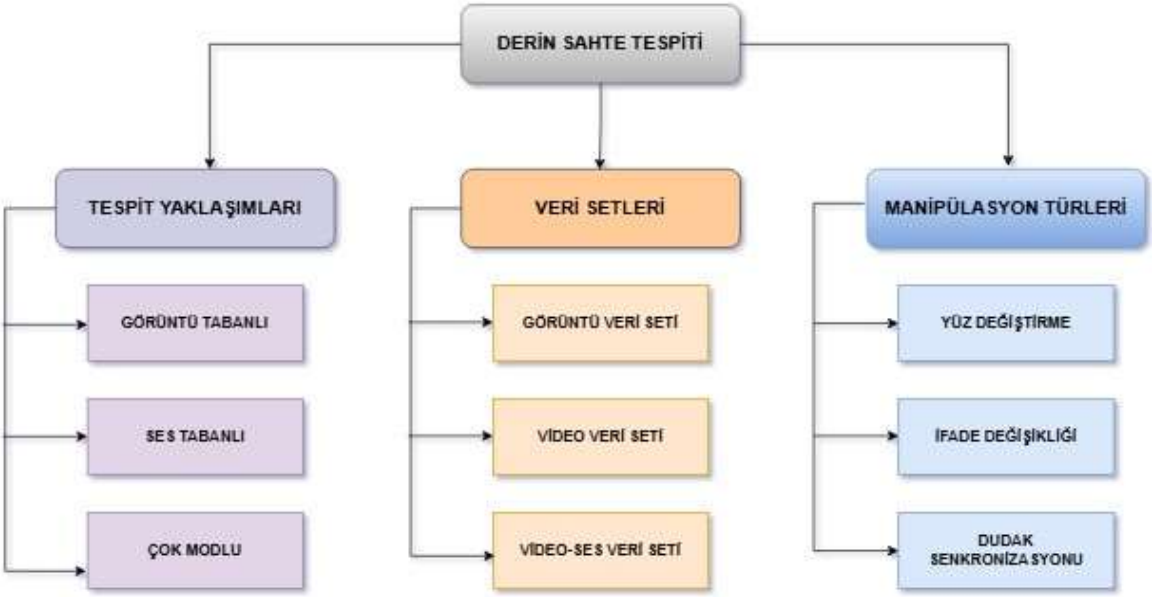
Dördüncü bölümde, öz-denetimli öğrenme yaklaşımı ile çok modlu model (ses ve görüntü) tespiti için kullanılan veri setleri, derin öğrenme teknikleri ve uygulanan metotlar açıklanacaktır. Modellerin eğitilmesinde tercih edilen algoritmalar, kullanılan test veri setleri ve performans değerlendirmeleri detaylandırılacaktır.

Beşinci bölümde, gerçekleştirilen çalışmaların sonuçları detaylı bir şekilde sunulacak ve geliştirilen modellerin performansı değerlendirilecektir. Bu kapsamda, öncelikle tez kapsamında geliştirilen öz-denetimli öğrenme yaklaşımı modeli ve çok modlu (ses ve görüntü) derin öğrenme modeli incelenecektir. Her bir modelin doğruluk, duyarlılık, ve F1 skoru gibi performans metrikleri üzerinden değerlendirilmesi de yapılacaktır.

Altıncı bölümde ise, çalışmadan elde edilen genel sonuçlara yer verilecek, tez kapsamında önerilen yöntemin katkıları özetlenecek ve ileride yapılabilecek çalışmalara yönelik öneriler sunulacaktır.

2. LİTERATÜR ARAŞTIRMASI

Son yıllarda dijital manipülasyon teknolojilerindeki hızlı gelişmeler, görüntü ve ses içeriklerindeki sahteciliklerin tespit edilmesini giderek daha önemli hale getirmiştir. Derin sahte teknolojisinin yaygınlaşması, bu sahte içeriklerin tespit edilmesi için çeşitli yöntemlerin geliştirilmesine ve bu alandaki araştırmaların artmasına neden olmuştur. Literatürde yer alan çalışmalar, sahte içeriklerin güvenilir bir şekilde tespit edilmesine yönelik önemli katkılar sunmaktadır. Şekil 2.1'de bu tespit sistemi örneği gösterilmiştir.



Şekil 2.1. Derin sahte tespit sistemi

2.1. Görüntü Tabanlı Derin Sahte Tespiti

P. Nandini ve M. Vikram [3] , derin sahte videoları tespit etmek için Inception_ResNet_v2 modeline dayalı bir yöntem önermişlerdir. Çalışmada DFDC veri seti kullanılmış, veriler ön işlemden geçirilerek modelin genelleştirme kapasitesi artırılmıştır. Inception_ResNet_v2, transfer öğrenme yaklaşımıyla iyileştirilmiş ve %97.73 doğruluk oranı elde edilmiştir. Model, özellikle yüksek kaliteli videolardaki manipülasyonları tespit etmekte etkili bulunmuş, düşük ışık veya bulanık ortamlardaki performansı konusunda sınırlı kalmıştır.

Abdul Kareem ve arkadaşları [4], makine öğrenimi tabanlı bir yaklaşım kullanarak sahte yüz görüntülerinin tespitini ele almıştır. Çalışmada, görüntüler RGB formatından YCbCr formatına dönüştürülerek ön işleme yapılmıştır. Canny filtresi ile kenar tespiti yapıldıktan sonra iki farklı yöntem denenmiştir: SVM+PCA ve yalnızca SVM. SVM+PCA ile %96,8 doğruluk elde edilirken, yalnızca SVM kullanılarak %72,2 doğruluk oranına ulaşılmıştır.

Vurimi ve arkadaşları [5], derin sahte video tespiti için derin öğrenme modelleri kullanarak, bir yaklaşım önermiştir. Çalışmada, ResNext adlı bir Evrişimli Sinir Ağı algoritması ve Uzun Kısa Süreli Bellek ağı kullanılarak derin sahte videolarının tespiti yapılmıştır. Önerilen yöntem, Celeb-Df veri kümesi üzerinde %91 doğruluk oranı elde etmiştir.

Khalil ve arkadaşları [6], derin sahte videoların tespitini iyileştirmek için transfer öğrenme yöntemine dayalı modeller önermiştir. EfficientNet tabanlı model, DFDC ve Face Forensics++ veri kümelerinde sırasıyla %94 ve %98 AUC skorlarına ulaşmıştır. Bu çalışmada, gelecekte video tabanlı derin sahte tespiti için daha az parametreye sahip derin öğrenme modelleri ve birden fazla sahte yüz içeren videoların tespiti üzerine çalışmalar önerilmiştir.

Suratkar ve Kazi [7], transfer öğrenme yaklaşımı kullanarak bir model önermişlerdir. Çalışmada, pre-trained deep learning models kullanarak derin sahte videoların tespiti yapılmıştır. Bu yaklaşım, daha önce eğitilmiş modellerin güçlü özelliklerini transfer ederek daha hızlı ve doğru tespit yapılmasını sağlamaktadır. Araştırmada, farklı veri setleriyle yapılan deneyler, transfer öğrenme yöntemlerinin derin sahte video tespitindeki doğruluğu önemli ölçüde artırdığını göstermiştir. Özellikle, modelin ResNet ve VGG gibi derin öğrenme ağlarıyla eğitilmesinin başarı sağladığı belirtilmiştir.

Yang ve arkadaşları [8], derin sahte saldırılarına karşı konuşmacı doğrulama sistemlerini korumak amacıyla dinamik dudak hareketi analizi yöntemini kullanmışlardır. Yöntem, CNN ve LSTM mimarileri ile derin öğrenme tekniklerini birleştirerek konuşmacı doğrulama sistemlerini daha güvenli hale getirmeyi hedeflemiştir. Evrişimli Sinir Ağı mimarisi, videolarda yüz hareketlerinin özelliklerini çıkarmak için kullanılmış, Uzun Kısa Süreli Bellek ise bu hareketlerin zaman içindeki dizisini analiz etmiştir. Bu sayede, dudak hareketlerinin doğal akışı ve değişimleri tespit edilerek, sahte videolarda görülen yapaylıklar belirlenmiştir. Çalışmada, bu yöntem ile derin sahte saldırılarını engellemeye yönelik yüksek bir başarı sağlanmıştır. Ayrıca, dinamik dudak hareketi analizi kullanılarak sesli kimlik doğrulama sistemlerinde %2,1 EER (Eşit Hata Oranı) değerine ulaşılmıştır.

Mishra ve arkadaşları [9], derin sahte tespitinde kullanılan tekniklerin evrimini ve mevcut yaklaşımları derinlemesine inceleyen bir çalışma sunmuşlardır. Çalışma, deepfake, FaceSwap, Face2Face, ve GAN tabanlı yöntemler gibi video manipülasyon tekniklerini analiz etmektedir. Çalışmada, derin öğrenme algoritmaları ve RNN gibi modellerin tespit süreçlerinde etkin olduğunu vurgulamışlardır. Ayrıca, çalışmada kullanılan veri setleri arasında Celeb-DF, UADFV, Deepfake-TIMIT, DFDC, ve FaceForensics++ gibi büyük ve çeşitli kaynaklar yer almaktadır. Özellikle, Phoneme-Viseme tutarsızlıklarının analiz edilmesi ve RNN+LSTM kombinasyonu ile zamansal manipülasyonların tespiti, yenilikçi yaklaşımlar arasında gösterilmiştir.

Xia ve arkadaşları, [10] derin sahte videoları tespit etmek için yeni bir yöntem geliştirmişlerdir. Çalışmalarında, MesoNet adlı klasik bir modeli alıp, bir ön işleme modülü

eklemişlerdir. Bu modül, sahte ve gerçek görüntüler arasındaki farkları daha belirgin hale getirmek için düşük frekanslı sinyalleri filtreleyip yüksek frekanslıları korumaktasdır. Bu yöntemle, görüntüdeki detayları sıkıştırılmış videolarda bile başarılı bir şekilde analiz etmişlerdir. Testlerde kullandıkları veri setleri FaceForensics++ ve Celeb-DF'dir. Ağır sıkıştırılmış videolarda bile %88'in üzerinde doğruluk sağlamışlardır. Bu, yöntemin dayanıklılığını göstermektedir. Gelecek çalışmalarda, bu yöntemi daha geniş veri setlerinde test etmeyi veya ses ve görüntü verilerini birleştirerek çok modlu analizler yapmayı önermişlerdir. Bu tür kombinasyonların, derin sahte tespitinde daha da sağlam sonuçlar verebileceğini gösteriyor.

Matern ve arkadaşları [11], görsel kusurların analizine dayalı bir yöntem önermişlerdir. Çalışma, özellikle sınır hataları, doku uyumsuzlukları, ve ışıklandırma farklılıkları gibi görsel tutarsızlıklara odaklanmaktadır. Bu yöntem, sahte içeriklerdeki ince görsel hataları belirleyerek sahte videoları gerçeklerden ayırt etmekte etkili bulunmuştur. Araştırmada, farklı veri setleri üzerinde gerçekleştirilen testler, yöntemin yüksek bir doğruluk oranıyla derin sahte içerikleri tespit edebildiğini göstermiştir. Özellikle, düşük çözünürlüğe sahip videolarda bile görsel kusurların kolayca tespit edilebildiği vurgulanmıştır. Çalışma, derin sahte manipülasyonlarının temel özelliklerini analiz ederek, bu içeriklerin sahte olduğunu güvenilir bir şekilde ayırt edebilmek için etkili bir araç sunmaktadır.

Li ve arkadaşları [12], tespit için göz kırpma davranışlarını analiz eden bir yöntem geliştirmiştir. Çalışmalarında, gerçek videolardaki göz kırpma oranlarının doğal bir ritme sahip olduğu, ancak yapay videolarda bu davranışın genellikle eksik veya düzensiz olduğu gözlemlenmiştir. Bu farklılıkları belirlemek için bir Evrişimli Sinir Ağı tabanlı model kullanılmıştır. Yöntem, göz bölgesindeki hareketleri analiz ederek sahte videoları gerçeklerden ayırt etmede oldukça etkili olmuştur. Bu çalışmada, CEW (Closed Eye in the Wild**)** veri setini kullanmışlardır. Yapılan testlerde, modelde %95'in üzerinde doğruluk oranı elde edilmiş, ayrıca göz kırpma hareketlerindeki tutarsızlıkların tespitiyle, sahte videoları güvenilir bir şekilde ayırt etmişlerdir. Özellikle, göz kırpma davranışının zamanlaması ve sıklığındaki bozulmalar, derin sahte videolarının belirlenmesinde kritik bir işaret olarak kullanılmıştır.

Peng Zhou ve arkadaşları [13], yüz manipülasyonlarını tespit etmek için iki akışlı (two-stream) bir ağ geliştirmişlerdir. Bu yöntemde, birinci akışta, GoogLeNet modeli yüz manipülasyonlarını tespit etmek için kullanılmıştır. İkinci akış ise, triplet ağ (patch-based triplet network) kullanarak, yerel gürültü ve kamera özelliklerini yakalamaya çalışmaktadır. Bu iki akış birlikte çalışarak daha doğru ve güvenilir sonuçlar elde edilmiştir. Çalışma, yüz manipülasyonlarını belirlemek için görsel hataları ve yerel gürültü kalıntılarını kullanarak sahte yüzlerin tespitini daha sağlam bir şekilde yapmaktadır.

Yang ve arkadaşları [14], tutarsız baş pozları analiz eden bir yöntem geliştirmiştir. Bu yöntemde, UADFV ve DARPA GAN Challenge veri kümelerindeki gerçek ve sahte videolar

kullanılmıştır. Model, SVM sınıflandırıcılarıyla eğitilmiş ve Area Under ROC (AUROC) değeri 0.94 ile UADFV veri kümesinde, 0.87 ile DARPA GAN Challenge veri kümesinde doğruluk sağlamıştır. Baş pozları farkı, deepfake videolarını tanımlamak için güçlü bir özellik olarak kullanılmıştır.

Tolosana ve arkadaşları [15], çalışmalarında sahte yüz tespiti tekniklerini kapsamlı bir şekilde ele almışlardır. Çalışmada, yüz manipülasyonlarını dört ana kategoriye ayırmışlardır: tam yüz sentezi, kimlik değiştirme, özellik manipülasyonu ve ifade değiştirme. Bu kategorilerde kullanılan yöntemler detaylı bir şekilde açıklanmış ve tespit teknikleri tartışılmıştır. Derin sahte tespitinde kullanılan çeşitli veri kümelerini de incelemişlerdir. Bu veri kümeleri arasında FaceForensics++, CelebA ve 100K-Generated-Images gibi yaygın olarak kullanılan kaynaklar yer almaktadır. Yöntemlerin zamanla gelişen manipülasyon tekniklerine karşı adaptasyonunun artırılması gerektiğini vurgulamışlardır. Ayrıca, tespit için kullanılan derin öğrenme yöntemleri ve geleneksel kameraya özgü parmak izleri gibi teknikler incelenmiştir.

Korshunov ve Marcel [16] tarafından yapılan çalışmada, sahte videoların tespitine yönelik bir güvenlik açığı değerlendirmesi yapılmıştır. Çalışmanın temel amacı, derin sahte videoların insan gözlemcileri ve otomatik tanıma sistemleri tarafından ne kadar gerçekçi algılandığını ölçmek ve bu tür manipülasyonların tespitini sağlamak için etkili yöntemler geliştirmektir. Yapılan çalışmada, GAN (Generative Adversarial Networks) tabanlı yüz değiştirme tekniklerinin yüz tanıma sistemlerini nasıl yanıltabileceğini incelemiş ve mevcut tespit yöntemlerinin bu tür sahte içerikleri nasıl ayırt edebileceğini değerlendirmiştir. Çalışmada, VidTIMIT veri kümesinden yararlanılarak düşük kalite (LQ) ve yüksek kalite (HQ) Deepfake videoları üretilmiştir. VGG ve Facenet gibi yüz tanıma sistemlerinin bu videolara karşı zayıf kaldığı ve %88.75 (LQ) ve %95.00 (HQ) gibi yüksek false acceptance rate (FAR) değerlerine sahip olduğu bulunmuştur. Sahte videoların bu sistemleri ne kadar yanıltabileceğini ve bu tür videoların tespit edilmesinin neden bu kadar zor olduğunu göstermektedir.

Raj ve arkadaşları [17], derin sahte ve propaganda videolarının tespitine yönelik FSSpotter, CNN tabanlı teknikler ve YOLO-CNN Extreme gradyan güçlendirme gibi gelişmiş yapay zeka yöntemlerini incelemişlerdir. Çalışma, mekansal ve zamansal analiz ile yüz manipülasyonu tespiti yaparak, bu tür içeriklerin yayılmasını engellemeye yönelik yenilikçi çözümler geliştirilmesi gerektiğini vurgulamaktadır. Derin sahte teknolojilerinin yol açtığı zorluklarla başa çıkabilmek için, bu alanda daha etkili yöntemler geliştirilmesinin önemini ortaya koymaktadır.

Kosarkar ve arkadaşları [18], sahte video tespitleri için özelleştirilmiş bir CNN modeli geliştirmişlerdir. Çalışma, derin sahte video görüntülerinin sınıflandırılması ve açığa çıkarılması için kullanılan yeni bir yaklaşımı sunmaktadır. Çalışmada öne çıkan özellik ise, geleneksel CNN yöntemlerinden farklı olarak daha özelleştirilmiş bir modelin kullanılması şeklindedir. Özelleştirilmiş CNN modeli, %91,4 doğruluk, 0,342 kayıp değeri ve 0,92 AUC ile en iyi sonuçları

elde etmiştir. Ayrıca, test aşamasında CNN modelin %85,2 doğruluk**,** MLP-CNN modelin ise %95,5 doğruluk sağladığı görülmüştür.

Gustavo ve arkadaşları [19], biyometrik yüz tanıma sistemlerinin sahtecilik saldırılarına karşı güvenliğini artırmayı hedefleyen bir çalışma yapmışlardır. wCNN (width-extended CNN) adı verilen yeni bir mimari önermişlerdir. Bu mimari, yüksek performanslı yüz sahteciliği tespiti yapabilmek amacıyla tasarlanmıştır. wCNN'in mevcut yöntemlerden daha verimli olduğu gözlemlenmiştir ve çoğu durumda daha yüksek doğruluk sağladığı anlaşılmıştır. Buna ek olarak, wCNN'in gereksinim duyduğu donanım kaynaklarının oldukça düşük olduğu görülmüştür, bu da özellikle mobil cihazlar gibi sınırlı kapasiteli ortamlar için avantaj sağlamaktadır.

Usmani ve arkadaşları [20], derin sahte tespiti için görüntü dönüştürücü model önermişlerdir: Shallow Vision Transformer (ViT). Bu model, diğerlerine kıyasla daha az parametrete sahiptir. Çoklu başlık dikkat mekanizması (MHA), görsellerin önemli kısımları vurgulanmıştır. Model, Real Fake Face (RFF) ve Real and Fake Face Detection (RFFD) veri setlerinde test edilmiştir. ViT, sadecce yarı RFF veri setinde eğitim aldığında dahi oldukça iyi bir doğruluk oranı gösterdiği görülmüştür.

Taeb ve Chi [21], yaptıkları çalışmada, derin öğrenme tabanlı VGG-19, DenseNet-121 ve özel tasarlanmış bir CNN modelini karşılaştırmıştır. Real and Fake Face-Detection ve 140K Real and Fake Faces datasetini kullanmışlardır. Bu veri seti, gerçek ve sahte yüz görüntüleri içeren geniş kapsamlı bir veri setidir, gri tonlamalı analizler için kullanılmıştır. Sonuçlara göre, VGG-19 modeli %95 doğruluk oranıyla diğer modellerden daha iyi bir performans göstermiştir. DenseNet-121 modeli %94 doğruluk oranıyla ikinci iyi performansı gösteren model olmuştur. Özel tasarlanmış CNN modeli ise %89 doğruluk elde etmiştir. Çalışmanın temel hedefi, adli bilişimde delil güvenilirliğini artırmak için bu tür modellerin önemini vurgulamaktır.

Patel ve arkadaşları [22], geliştirdikleri derin öğrenme tabanlı yöntemde, ResNext CNN ve LSTM modelleri kullanmışlardır. Bu model, sahte videolarda ortaya çıkan çözünürlük tutarsızlıklarını ve zamansal düzensizlikleri tespit etmek için çerçeveleri analiz etmektedir. Çerçeveler, ResNext kullanılarak 2048 boyutundaki özellik vektörleri çıkarmıştır. %91,5 doğruluk oranı, 80 çerçeveli analizle elde edilmiştir. Daha düşük çerçeve sayısında doğruluk oranları %84,2 - %90,5 arasında değişmiştir. Bu çalışmada, yaygın olarak kullanılan FaceForensics++ ve YouTube videolarından seçilen karışık bir veri seti kullanılmıştır. Çalışmada, ResNext CNN ve LSTM kombinasyonun, zamansal tutarsızlıkları analiz ederek yüksek doğruluk oranları sağladığı gözlemlenmiştir.

Bonomi ve arkadaşları [23], dinamik doku analizi yöntemini kullanarak sahte görüntü tespiti için yenilikçi bir yaklaşım geliştirmişlerdir. Çalışmada, uzamsal ve zamansal özelliklerin çıkarılması hedeflenmiştir. Yöntemde, FaceForensics++ veri setini kullanmışlardır. Destek vektör

makineleri kullanılarak sınıflandırmalar yapılmıştır. Dinamik doku analizi, sahte görüntülerin belirlenmesinde etkili bir sonuç oluşturmuştur ve yüksek doğruluk oranları elde edilmiştir.

Xu ve arkadaşları [24], sahte videoların tespiti için Set Convolutional Neural Networks (SCNN) tabanlı bir yöntem geliştirmişlerdir. SCNN kullanarak Deepfake-TIMIT, FaceForensics++, ve DFDC Preview veri setleri üzerinde deepfake videoların tespiti için oldukça kapsamlı bir analiz gerçekleştirmişlerdir. Çalışmada, MesoNet, XceptionNet, ve bunların optimize edilmiş türevlerini karşılaştırmıştır. Deepfake-TIMIT verisetinde, düşük kalitedeki videolar için MesoNet doğruluk: %76.12, XceptionNet doğruluk: %88.24, XceptionNet AUC: %99.81 iken yüksek kalitedeki videolar üzerinde, MesoNet doğruluk: %77.49 oranına ulaşmıştır. FaceForensics++ veri setinde, t_XceptionNet, sıkıştırılmamış videolar için %98'in üzerinde doğruluk elde etmiştir. Genel olarak, SCNN'in oldukça iyi bir performans sergilediği görülmüştür.

Kingra ve arkadaşları [25], LBPNet adlı bir yöntem önermişlerdir, Bu çalışma, Local Binary Pattern (LBP) kullanarak sahte ve gerçek yüzler arasındaki dokusal tutarsızlıkları analiz ederek, bu bilgiyi CNN tabanlı bir modelle entegre etmektedir. Önerilen yöntemde, Celeb-DF, DFDC ve DeeperForensics veri setleri kullanılmıştır. LBPNet, bu veri setlerinde üzerinde Celeb-DF'de %92,38 doğruluk, DFDC'de %80 doğruluk ve diğer veri setinde ise %86 doğruluk oranlarına ulaşarak, üstün bir performans göstermiştir.

Cao ve arkadaşları [26], çalışmalarında yüz manipülasyonun tespiti için üç sınıflandırmalı bir yöntemi (TFMD) önermişlerdir. Bu yöntem dikkat tabanlı bir özellik ayrıştırma modelidir. Önerilen çalışmada, manipülasyon için yaptıkları sınıflandırma şu şekildedir**:** yüz değiştirme ve yüzü yeniden canlandırma. Çalışmada, görsel tutarsızlık tespitleri için dikkat mekanizmaları ve derin özellik ayrıştırma teknikleri birleştirilmiştir. FaceForensics++ ve Celeb-DF veri setleri kullanılmıştır. Tespit için yüksek doğruluk elde edilerek, etkili bir yöntem sunulduğu gözlemlenmiştir.

Dang ve arkadaşları [27], yüz manipülasyonu tespitinde kullanılan geleneksel yöntemlerin yetersizliği konusuna dikkat çekerek, derin öğrenme yöntemlerini önermişlerdir. Bu çalışmada, manipüle edilmiş yüz görüntülerini tespit etmek için özelleştirilmiş bir Evrişimsel Sinir Ağı modeli olan MANFA (Manipulated Face Analysis) geliştirilmiştir. Bununla beraber, MANFA'ya Adaptive Boosting (AdaBoost) ve Extreme Gradient Boosting (XGBoost) algoritmalarını entegre eden hibrit bir model olan HF-MANFA tasarlamışlardır. Yapılan deneylerde, MANFA %84.7 doğruluk ve 0.81 AUC puanı elde etmiştir. Hibrit model olan ADA-MANFA %85.4 doğruluk ve 0.89 AUC, XGB-MANFA ise %87.1 doğruluk ve 0.90 AUC puanı almıştır.Dengeli olmayan veri kümelerinde ise, en yüksek performansı XGB-MANFA göstermiş, AUC değeri 0.89'un üzerindedir. Bu çalışmada ayrıca, MANFA modelinin özellik çıkarmadaki etkinliği ve hibrit modellerin performans artırıcı etkisini vurgulanmıştır.

Rössler ve arkadaşları [28], dijital medya adli bilişim alanında sahte yüz tespiti için FaceForensics++ veri kümesini kullanmışlardır. Çalışmada, bu veri setiyle sahtecilik tespiti için bir otomatik değerlendirme kriteri ve derin öğrenme tabanlı yöntemler geliştirmişlerdir. Yöntemler arasında XceptionNet ve destek vektör makineleri yer almıştır. Sonuçlara bakıldığında, XceptionNet'in özellikle düşük kaliteli videolarda gözlemcilerden çok daha iyi performans gösterdiği görülmüştür. Bu çalışmada, manipülasyon yöntemlerinin ve veri setinin büyüklüğünün model performansı üzerindeki etkisi detaylı bir şekilde incelenmiştir.

2.2. Ses Tabanlı Derin Sahte Tespiti

Ousama ve arkadaşları [29], makine öğrenimi ve derin öğrenme algoritmalarını kullanarak sahte sesleri tespit etmek için çeşitli yöntemler kullanmıştır. Bu yöntemler arasında SVM, DT, CNN, Siyam CNN, DNN ve CNN-RNN kombinasyonları yer almaktadır. Yöntemler arasında SVM %99 doğruluk oranıyla en etkili yöntem olurken, DT %73,33 en düşük performansı göstermiştir. Ayrıca, EER oranları %2 (Deep-Sonar) ile %12,24 (DNN-HLL) arasında değişmiştir. Siyam CNN'lerin, min-t-DCF ve EER'de %55'lik iyileşme sağladığı belirtilmiştir.

Fathan ve arkadaşları [30], Mel-spektrogram tabanlı bir yaklaşım ile kanal uyumsuzluklarına rağmen sahte ses tespiti üzerine odaklanmıştır. VGG16 ve WaveletCNN mimarileri kullanılarak, ASVspoof2019 ve ASVspoof2021 veri kümelerinde değerlendirme yapmıştır. Önerilen model, Eşit Hata Oranı (EER) açısından %2,8 gibi düşük bir değer elde ederek mevcut yöntemlerden daha iyi performans göstermiştir. Ayrıca, ses verileri için özel olarak tasarlanmış artırma yöntemlerinin ve mel-spektrogram özelliklerinin, sahte verilerdeki frekans ve zaman bilgilerini daha iyi analiz ederek genel tespit başarısını artırdığı gözlemlenmiştir.

M.Nafees ve arkadaşları [31], sahte ses tespiti için bir ML-DL SafetyNet modeli önermişlerdir. Model, ASVspoof 2019 veri kümesinden alınan mantıksal erişim (LA) ve fiziksel erişim (PA) seslerini, sahte veya gerçek olarak sınıflandırmak için geliştirilmiştir. Derin öğrenme ve makine öğrenimi algoritmalarını birleştiren model, %90 doğruluk oranı elde etmiştir. Ayrıca, Mel-spektrogramlardan özellik çıkarma ve sınıflandırma işlemleriyle modelin performansı artırılmıştır. SVM, %90 doğruluk oranıyla en iyi sınıflandırıcı olarak belirlenmiştir.

Kwak ve arkadaşları [32], sahte sesleri tespit etmek için "ResMax" adlı bir mimari önermişlerdir. ResMax, ResNet'ten alınan atlama bağlantı konseptini ve Light CNN (LCNN)'den alınan maksimum özellik haritasını birleştirerek optimize edilmiştir. ASVspoof 2019 veri setinde, fiziksel erişim (PA) ve mantıksal erişim (LA) verileri üzerinde sırasıyla %0,30 ve %2,19 EER elde edilmiştir. Ayrıca, daha hafif bir model olan ResMaxSep, parametre sayısını %84 oranında azaltarak %0,36 EER ile rekabetçi bir performans göstermiştir.

Agarwal ve arkadaşları [33], fonem-visem uyuşmazlıklarını kullanarak bir yöntem geliştirmiştir. Çalışma, sesli videolarda söylenen kelimelerle dudak hareketlerinin uyumsuzluğuna

dayanır. Geliştirilen model, video ses ve dudak hareketlerini analiz ederek bu uyumsuzlukları tespit etmeye çalışmaktadır. Yöntem, lip-sync (dudak senkronizasyonu) tekniklerinin, video manipülasyonları için önemli bir ipucu sunduğunu vurgulamaktadır. Testlerde, MBP (M, B, P fonem grubu) gibi kelimelerle uyumsuzlukların en çok görüldüğü yerler tespit edilmiştir. Yöntem, %76.8 doğruluk oranı ile doğru eşleşmeler elde ederken, in-the-wild deepfakes üzerinde %90 doğruluk oranı elde edilmiştir. Ayrıca, dudak hareketlerinin ve sesin senkronizasyonunu sağlayan otomatik algoritmaların yüksek doğruluk oranları sağladığı görülmüştür.

Mcuba ve arkadaşları [34] çalışmalarında, sahte ses tespiti üzerine derin öğrenme yöntemlerinin etkilerini incelemişlerdir. Çalışmada, ses klonlama teknolojilerinin dijital adli soruşturmalarda nasıl bir tehdit oluşturduğuna ve bu tür manipülasyonların tespitine yönelik kullanılan mevcut yöntemlere odaklanılmıştır. Ses dosyalarındaki özellikler, MFCC, Mel-spectrum, Chromagram ve spectrogram gibi temsil yöntemleriyle görselleştirilmiştir. Yapılan testlerde, VGG-16 mimarisinin MFCC görüntü özellikleriyle en iyi sonuçları verdiği, özel bir Custom Architecture modelinin ise Chromagram, Spectrogram ve Mel-Spectrum görüntülerinde daha iyi performans gösterdiği bulunmuştur. Çalışma, dijital adli tıp uzmanlarının gerçek ve sahte sesleri ayırt etmelerine yardımcı olmayı amaçlamaktadır.

Conti ve arkadaşları [35], duygu tanıma (SER) tekniklerini kullanan yeni bir yaklaşım geliştirmişlerdir. Bu yaklaşım, TTS algoritmaları ile üretilen sahte konuşma parçalarının tespiti üzerine odaklanmaktadır. Yöntem, VGGish ve RawNet2'yi geride bırakarak AUC = 0,98 gibi yüksek performans sergilemiş ve CNN tabanlı yöntemlere kıyasla daha iyi sonuçlar elde edilmiştir.

Khochare ve arkadaşları [36], sahte ve gerçek ses verilerini sınıflandırmak için iki yaklaşım geliştirmişlerdir: özellik tabanlı ve görüntü tabanlı. Çalışmada, bu yaklaşımlar karşılaştırılmış ve TCN (Temporal Convolutional Network), %92 doğruluk oranı ile en iyi performansı göstermiştir. Bu, derin öğrenme modellerinin, özellikle sıralı verilerde (ses verisi gibi) etkili olduğunu ve geleneksel CNN modelleriyle karşılaştırılabilir doğruluk seviyelerine ulaşabileceğini kanıtlamaktadır.

Nugroho ve Winarno [37], ses sahteliği tespiti için Derin Sinir Ağı (DNN) algoritmalarını kullanmışlardır. Yapılan testlerde, DNN modeli sahte konuşma tespiti konusunda yüksek doğruluk oranları elde etmiştir. Bu model, özellikle sahte seslere karşı etkili olup, %96,5 doğruluk oranı ile başarılı performans sergilemiştir. Çalışma, DNN tabanlı modellerin, sahte sesleri tespit etmede çok güçlü bir araç olduğunu ve bu modellerin gelişen sahte seslere karşı sağlam bir çözüm sunduğunu vurgulamaktadır.

2.3. Çok Modlu Tabanlı Derin Sahte Tespiti

Hashmi ve arkadaşları [38], sahte video içeriklerini tespit etmek için geliştirdikleri yöntemde FakeAVCeleb veri seti üzerinde bir multimodal derin öğrenme yaklaşımı önermişlerdir.

Unimodal ve çok modlu ağları birlikte kullanılarak toplu bir tahmin mekanizması tasarlamışlardır. FakeAVCeleb veri seti üzerinde %89 doğruluk elde etmişlerdir. Bunun sonucunda, mevcut yöntemlerden daha iyi performans elde ettikleri gözlemlenmiştir.

Zhang ve arkadaşları [39], sahte video tespiti için yapılan deneylerde, MFCC özellik çıkarıcısı FBank ile değiştirilerek DFDC veri kümesinde test edilmiştir. Sonuçlar, önerilen yöntemin DFDC veri kümesinin bir alt kümesinde derin sahte videoları tespit etmede %84,4 doğruluk elde ettiğini göstermiştir. Ayrıca, uyarlanabilir MDS skoru hesaplama yöntemi de AUC'de, %1'lik bir artış sağlanmış ve uyarlanabilir eşiklerin sabit eşiklerle karşılaştırıldığında daha iyi sonuçlar elde ettiği görülmüştür.

Mittal ve arkadaşları [40] çalışmalarında ses ve görüntü modaliteleri arasındaki benzerlik ve duygusal tutarsızlıkları analiz ederek derin sahte tespiti yapan bir yöntem önermişlerdir. Siamese ağ mimarisinden esinlenen bir model geliştirmişlerdir. Deepfake-TIMIT (DF-TIMIT) ve DFDC olmak üzere iki büyük ölçekli derin sahte veri seti üzerinde Alan Eğrisi Altındaki Alan (AUC) metriğini raporlamışlardır. Yaklaşımlarını çeşitli güncel (SOTA) derin sahte tespit yöntemleriyle karşılaştıran araştırmacılar, DFDC veri setinde %84.4 AUC ve DF-TIMIT veri setinde ise %96.6 AUC elde ettiklerini belirtmişlerdir.

3. DERİN ÖĞRENME

Derin öğrenme, çok katmanlı yapay sinir ağlarını kullanan makine öğrenmesinin bir alt kümesidir. Derin öğrenme ve makine öğrenmesi arasındaki temel fark, kullanılan sinir ağı mimarisinin yapısıdır. Geleneksel makine öğrenmesi modelleri, genellikle bir ya da iki katmandan oluşur. Derin öğrenme modelleri ise, genellikle üç veya daha fazla katmanla çalışır, bazen yüzlerce hatta binlerce katman içerebilir. Bu da, derin öğrenme modellerinin daha karmaşık veri setlerini işlemekteki başarısına büyük katkı sağlar.

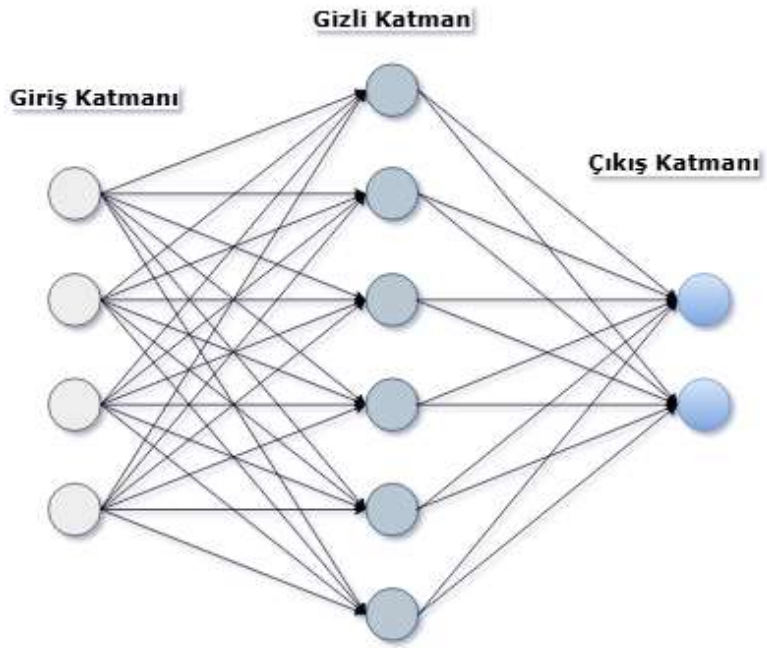
Derin öğrenme algoritmalarının temel amacı, büyük miktarda veriyi işleyerek, bu verilerdeki karmaşık olan bağlantıları otomatik olarak öğrenmektir. Bu işlem yapılırken, algoritmalar insan müdahalesine ihtiyaç duymadan, veri içerisindeki soyut ve karmaşık özellikleri kendiliğinden tanımlayabilir. Örneğin sınıflandırma, tahmin, örüntü tanımlama ve nesne algılama gibi karmaşık görevlerde oldukça etkili bir performans gösterir. Derin öğrenme, birbirine bağlı yapay sinir hücrelerinden oluşan çok katmanlı bir mimari kullanır. Katmanlar genel olarak üçe ayrılır:

Giriş katmanı: Verinin modele alındığı kısımdır.

Gizli Katman**:** Her düğümde matematiksel hesaplamaların yapıldığı bölümdür. Bu katmanlar, girdilerden gelen veriyi işler ve katmanlar arasında dönüşümler yaparak modelin verilerdeki karmaşık yapıları öğrenmesini sağlar.

Çıkış Katmanı : Model sonucunun üretildiği katmandır.

Katman yapısı Şekil 3.1'de gösterilmektedir.



Şekil 3.1. Sinir ağı model katmanları

Giriş verisinin, modelin ağı boyunca ilerleyerek bir çıktı oluşturması ileri besleme olarak adlandırılır. Geri yayılımda (backpropagation), modelin ürettiği çıktıyla gerçek değer karşılaştırılarak hata hesaplanır. Bu hata, optimizasyon algoritmaları kullanarak ağın ağırlıklarının güncellenmesi için geriye doğru yayılır.

Bu süreçler bir araya getirilerek, derin öğrenme modelleri büyük veri setlerini etkili bir şekilde analiz etmesi sağlanır ve karmaşık problemleri çözmek için güçlü bir araç haline gelir.

3.1. Derin Öğrenmenin Kısa Tarihçesi

Derin öğrenmenin temelleri, yapay sinir ağlarının gelişimiyle atılmaya başlamıştır. 1940'larda Warren McCulloch ve Walter Pitts beyin hücrelerinin işlevini taklit eden modeli önermişlerdir. Bu model, Mcp-neuron olarak adlandırılmıştır ve derin öğrenmenin temelini oluşturmuştur [41]. 1957'de, Frank Rosenblatt'ın geliştirdiği perceptron modeli, sinir ağlarının eğitilmesi için ilk başarılı algoritmayı sunmuştur. 1960'larda bu modelin sınırlamaları ortaya çıkınca, derin öğrenme araştırmaları duraklamıştır. 1986'da, Geoffrey Hinton ve arkadaşları, geri yayılım algoritmasını (backpropagation) geliştirerek, çok katmanlı sinir ağlarının eğitilmesini sağlamıştır [42]. 1990'ların sonlarında, Yann LeCun çığır açan çalışmalar ortaya koymuştur. Lecun, evrişimli sinir ağı (CNN) LeNet-5 mimarisini geliştirmiştir. 1999'da, Nvidia ilk GPU'yu geliştirmiştir. 2006 yılında, Geoffrey Hinton ve arkadaşları, derin öğrenme terimini ortaya atmıştır. 2012 yılında AlexNet'in ImageNet yarışmasında büyük başarı elde etmesiyle, derin öğrenmenin önemi artmıştır. Ardından, doğal dil işleme ve ses tanıma gibi alanlarda da derin öğrenme algoritmalarının başarısı artmıştır.

Bugüne geldiğimizde ise derin öğrenme, otonom araçlar, sağlık, güvenlik ve finans gibi birçok alanda kullanılmakta ve gelişmeye devam etmektedir.

3.2. Derin Öğrenmenin Kullanım Alanları

Derin öğrenme, çok çeşitli veri türleri üzerinde karmaşık örüntüleri öğrenme kapasitesi sayesinde birçok alanda etkin şekilde kullanılmaktadır. Görüntü işleme alanında, nesne tanıma, yüz tanıma, tıbbi görüntü analizi, otonom araçların çevre algılama sistemleri gibi uygulamalar öne çıkmaktadır. Ses ve konuşma işlemede ise sesli asistanlar, konuşma tanıma, dil modelleme, duygusal analiz ve sahte ses tespiti gibi pek çok uygulama alanı bulunmaktadır. Doğal dil işleme alanında metin sınıflandırma, çeviri, özetleme, duygu analizi ve sohbet robotları gibi uygulamalarda da derin öğrenme teknikleri yaygın olarak kullanılmaktadır. Ayrıca, metinden konuşma sentezleme (text-to-speech) sistemleri giderek daha doğal sesler üretebilmekte, ses klonlama ve konuşmacı tanıma gibi alanlarda da derin öğrenme modelleri başarılı bir şekilde kullanılmaktadır. Bu kullanım alanları arasında otonom sistemler için de büyük öneme sahiptir;

özellikle sürücüsüz araçlar, çevrelerini algılamak, şerit takibi yapmak ve trafik işaretlerini tanımak için derin öğrenme tabanlı algılama ve kontrol sistemlerine güvenmektedir.

Sağlık ve biyoinformatik alanları da, derin öğrenmenin en umut veren uygulama alanlarından biridir. Tıbbi görüntülerden (MR, röntgen vb.) hastalıkların erken teşhisi, kanserli hücrelerin tespiti, genetik verilerin analizi kişiye özel tedavi yöntemlerinin geliştirilmesi gibi konularda derin öğrenme modelleri önemli katkılar sağlar.

Finans sektöründe ise dolandırıcılık tespiti, kredi risk değerlendirmesi ve müşteri hizmetleri otomasyonu gibi alanlarda derin öğrenme algoritmaları etkin bir şekilde kullanılmaktadır. E-ticaret ve içerik platformlarında kullanıcı davranışlarını analiz ederek kişiselleştirilmiş ürün veya içerik önerileri sunan öneri sistemleri, büyük ölçüde derin öğrenme tekniklerinden faydalanmaktadır.

Bu yaygın ve etkili kullanım alanlarına bakıldığında derin öğrenme, günümüzde birçok alanda tercih edilen güçlü bir yapay zeka tekniği haline gelmiştir.

3.3. Derin Öğrenme Yaklaşımları

Derin öğrenme yaklaşımları, model yapılarına ve öğrenme yöntemlerine göre farklılık gösterirler. Bu yaklaşımlar: Denetimli öğrenme (supervised learning)**,** yarı denetimli öğrenme (semi-supervised learning), denetimsiz öğrenme (unsupervised learning) olarak gruplara ayrılır.

3.3.1. Denetimli Öğrenme

Bu yaklaşımda, model bir eğitim veri seti ile eğitilir ve bu veri setinde her örneğe karşılık bir etiket vardır. Model, giriş verilerinin doğru etiketlerle eşleşmesini öğrenir. Sınıflandırma ve regresyon problemlerinde yaygın olarak kullanılır. Örnek olarak, bir modelin bir fotoğrafı kediler ve köpekler arasında sınıflandırmayı öğrenmesini verebiliriz. Verilerin etiketli olmasından dolayı öğrenme sürecinin daha kontrollü ve doğrulaması kolay olması bu yaklaşımın avantajıdır.

3.3.2. Yarı Denetimli Öğrenme

Genel olarak sınıflandırmada kullanılır. Bu yaklaşımda, veriler hem etiketli hem de etiketlenmemiş olabilir. Etiketlenmiş verilerin genellikle sınırlı olduğu durumlarda etiketsiz verilerden faydalanarak model performansını arttırabilir. Derin öğrenme alanında veri etiketleme işlemini optimize ederek daha etkili ve verimli modellerin geliştirilmesine yardımcı olabilir.

3.3.3. Denetimsiz Öğrenme

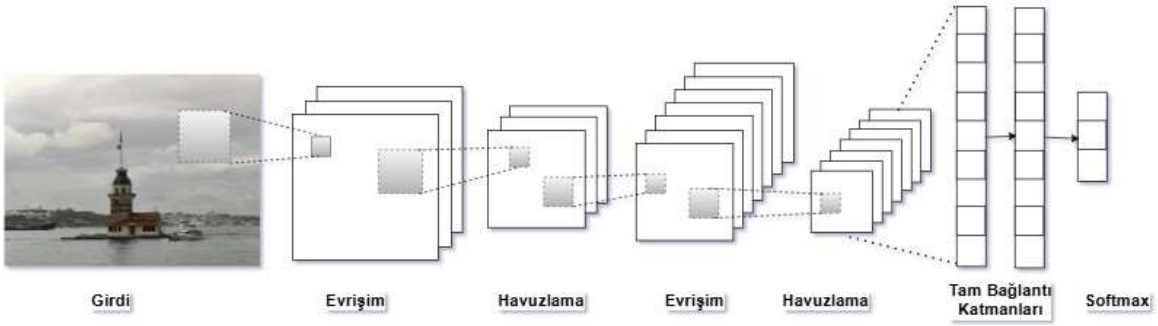
Bu yaklaşım eğitim verilerinin etiketlenmediği durumlarda kullanılır. Bu öğrenmenin en yaygın uygulama alanları arasında kümeleme (clustering) ve boyut indirgeme (dimensionality

reduction) vardır. Bu yaklaşımda, model veriler arasındaki benzerlikleri, grupları ya da gizli yapıları anlamaya çalışır. Genellikle veri kümesindeki kalıpları, ilişkileri ya da sınıflandırmaları bulmak için kullanılır.

3.4. Evrişimli Sinir Ağları (Convolutional Neural Networks – CNN)

Evrişimsel Sinir Ağları, derin öğrenmenin en yaygın kullanılan modellerinden biridir. Özellikle görüntü ve video işleme, nesneleri algılama ve sınıflandırma işlemlerinde sıklıkla kullanılır ve oldukça etkili sonuçlar verir. CNN'lerin önemli avantajları arasında, Yapay Sinir Ağlarındaki parametre sayısını önemli ölçüde azaltarak daha verimli bir öğrenme süreci sağlamasıdır. Bu avantaj, araştırmacıların ve geliştiricilerin klasik yapay sinir ağlarıyla başa çıkılamayan karmaşık problemler için daha büyük ve karmaşık modellere yönelmesinin önünü açmıştır. [43] Bu nedenle, CNN'ler görüntü işleme ve derin öğrenme uygulamaları gibi alanlarda standart olarak kullanılan bir yöntem haline gelmiştir.

Şekil 3.2'de CNN mimarisine ait genel yapı gösterilmektedir.



Şekil 3.2. Evrişimsel sinir ağı mimarisi

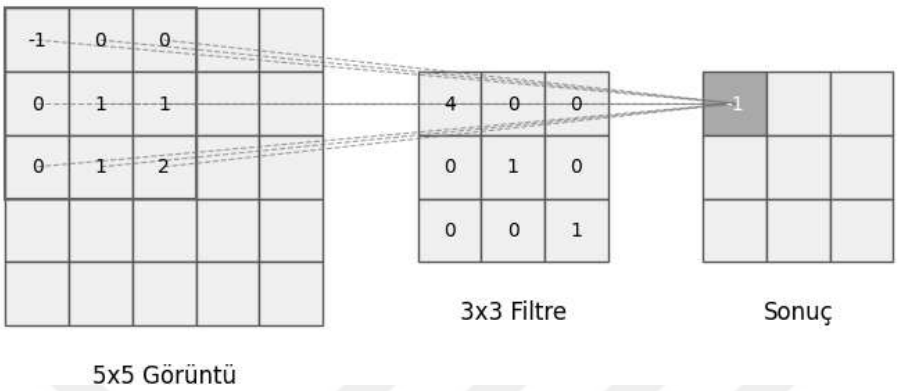
3.4.1. CNN Katmanları ve Yapısı

Evrişimli Sinir Ağları (CNN), derin öğrenme modellerinin temel yapılarından biridir. Bu model bir dizi katmandan oluşur ve bunların her biri özel bir işlevi yerine getirir. Bu katmanlar, modelin öğrenme sürecinde farklı seviyelerde özellikleri çıkarmasını sağlar. CNN'nin temel katmanları genellikle evrişim (convolutional), havuzlama (pooling), tam bağlantılı (fully connected) ve dropout katmanlarını içerir. Bu katmanlar, modelin giriş verisinden anlamlı özellikler öğrenmesini sağlar, sınıflandırma ve regresyon gibi görevlerde yüksek başarı elde edilmesine yardımcı olur.

- **Evrişim Katmanı**

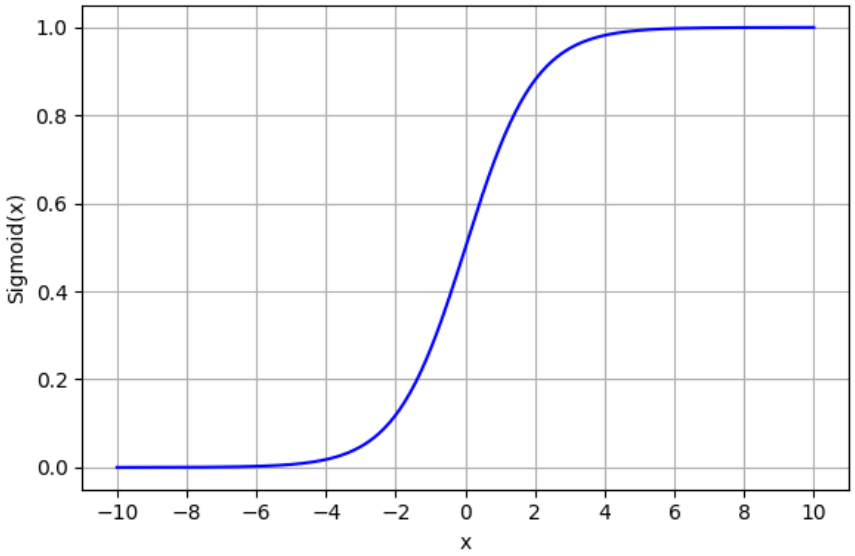
Evrişim katmanında, CNN'lerin temel yapı taşlarından biri olan çekirdek adı verilen küçük boyutlu matrisler kullanılır (3x3 ya da 5x5 gibi) ve özellikle görsel verilerde önemli bir rol oynar. Bu çekirdekler, girdinin belirli bir bölgesiyle çarpılarak özellik haritaları oluşturur. Örneğin, bir resim üzerinde dikey kenarları algılamak için tasarlanmış bir çekirdek kullanıldığında, evrişim

işlemi bu özellikleri tanımlamak için çıkarımda bulunur. Şekil 3.3'te gösterildiği gibi Filtre, resim üzerinde kaydırılarak, her konumda bir çarpma ve toplama işlemi gerçekleştirir ve bu şekilde özellikler özetlenir.



Şekil 3.3. Evrişim işlemi

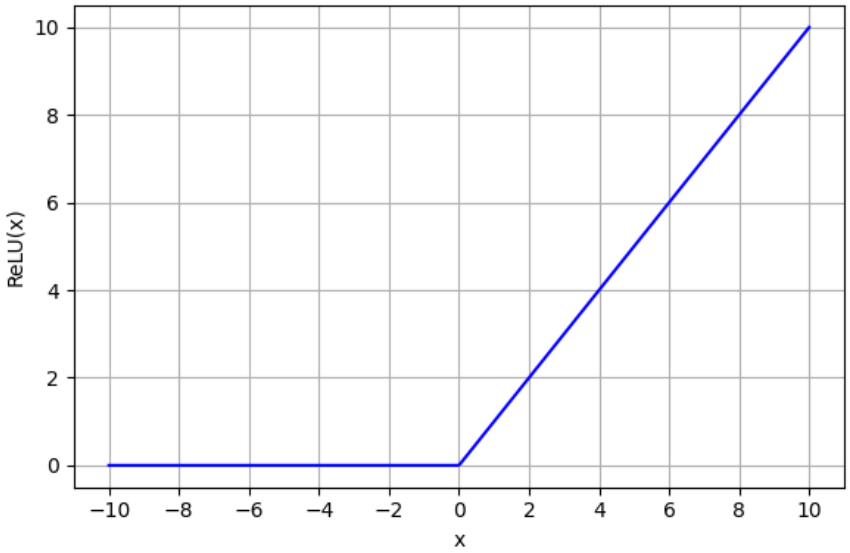
Kaydırma işleminde çekirdeğin, giriş üzerinde ne kadar hareket ettiği belirlenir. Örneğin, kaydırma değeri 1 olan bir filtre her pikselde bir hareket ederken, kaydırma değeri 2 olan bir filtre her iki pikselde bir hareket eder. Daha büyük kaydırma, daha az özellik haritası pikseli oluşturur. Görüntü kenarlarına filtreyi tam uygulayabilmek için yapılan işlem ise doldudur. Evrişim işlemi sonrasında elde edilen değerler, bir aktivasyon fonksiyonuyla işlenir. ReLU (Rectified Linear Unit) en yaygın kullanılan aktivasyon fonksiyonudur. Bu fonksiyon, negatif değerleri sıfırlayarak doğrusal olmayan bir yapı ekler. Bu da, modelin daha karmaşık ilişkileri öğrenmesini sağlar.



Şekil 3.4. Sigmoid aktivasyon fonksiyonu

$$f(x) = \frac{1}{1 + e^{-x}}$$

Şekil 3.4'te Sigmoid aktivasyon fonksiyonunun grafiği sunulmuştur.



Şekil 3.5. ReLU aktivasyon fonksiyonu

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

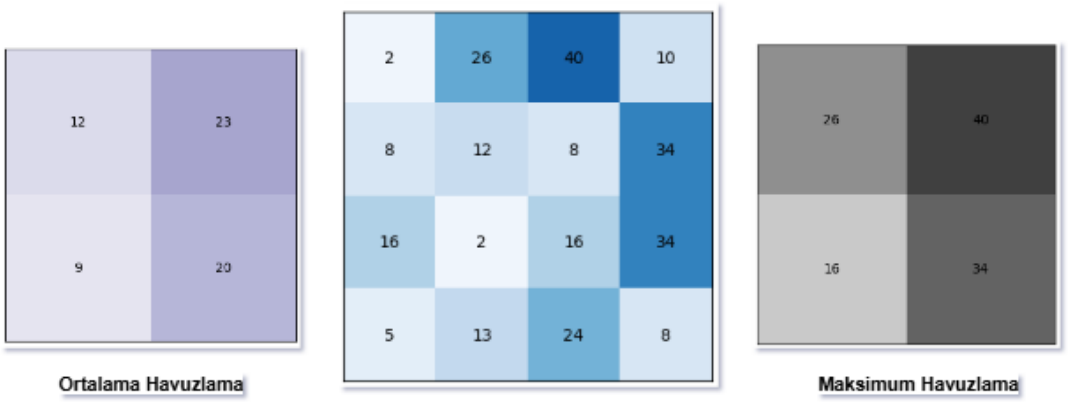
Özellik haritası, evrişim katmanı tarafından çıkarılan belirli bir özelliği temsil eder (kenar, köşeler gibi). Örneğin, bir çekirdek dikey kenarları algılarken, başka bir çekirdek yatay kenarları öğrenebilir.

ReLU aktivasyon fonksiyonuna ait grafik Şekil 3.5'te gösterilmektedir.

- **Havuzlama Katmanı**

Havuzlama, evrişim katmanları tarafından çıkarılan özellik haritalarının boyutlarını küçültmek için kullanılır. Bu işlemin örnek bir görsel temsili Şekil 3.6'da sunulmuştur. İşlem, öğrenilen özelliklerin daha genel hale gelmesini sağlar. Böylece modelin daha az parametreyle çalışmasını sağlar ve işlem süresi kısalır. Bu da modelin daha verimli çalışmasını olanak tanır.

En yaygın kullanılan havuzlama türü maksimum havuzlamadır. Bir bölgedeki (genellikle 2x2 veya 3x3 boyutlarında) en büyük değeri seçer. Böylece, belirli bir bölgedeki önemli özellikler korunmuş olur. Ortalama havuzlamada ise, bir bölgedeki tüm değerlerin ortalamasını alır. Maksimum havuzlamaya göre, daha az hassasiyet gerektirir.



Şekil 3.6. Havuzlama işlemi

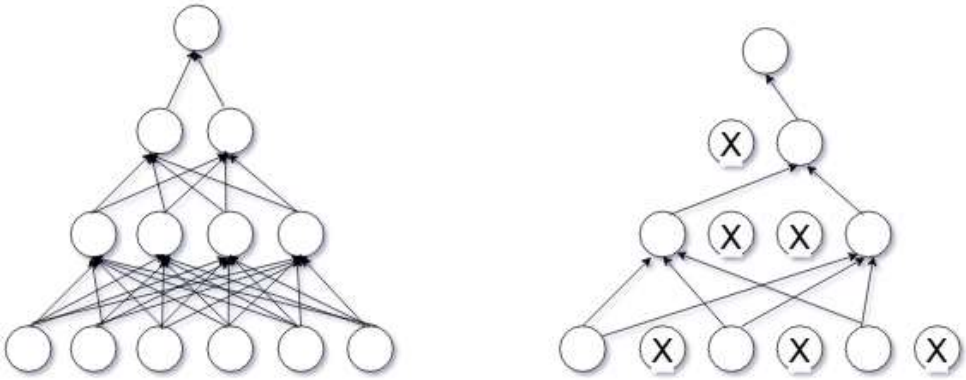
Havuzlama, modelin farklı ölçeklerdeki özellikleri tanımasını sağlar ve aşırı uyum riskini azaltarak modelin genelleme kapasitesini artırır.

- **Tam Bağlantılı Katman**

Tam bağlı katman, havuzlama katmanını takip eder ve son adım bu katmanda gerçekleşir. Tam bağlı katman, özellik haritasını tek boyutlu bir vektöre dönüştürmekle görevlidir [44]. Bu katmandaki, düğümler, kendisinden önce ve sonra gelen katmandaki diğer düğümlerle doğrudan bir bağlantıya sahiptir. Bu şekilde model daha karmaşık ilişkileri öğrenebilir.

- **Dropout Katmanı**

Çok fazla parametreye sahip olan sinir ağları aşırı uyuma sebep olur. Bu katman, derin öğrenme modellerinde aşırı öğrenmeyi önlemek amacıyla kullanılır. Kısaca sinir ağlarında düzenleme tekniğidir denilebilir. Katmanın temel amacı, modelin yalnızca belirli nöronlara bağımlı olmasını engellemektir. Eğitim sırasında rastgele nöronların devre dışı bırakılması, modelin genel performansını artırmayı sağlar. Dropout, özellikle çok büyük ve karmaşık derin öğrenmede etkili bir düzenleme tekniğidir. Dropout uygulamasının görsel karşılaştırması Şekil 3.7'de sunulmuştur.



Şekil 3.7. Normal sinir ağı ve dropout uygulanan sinir ağı

3.4.2. CNN Mimarileri ve Uygulamaları

Evrimsel Sinir Ağları, derin öğrenme yöntemleri arasında özellikle görsel veri analizi için çok önemli bir yere sahiptir. Başlarda basit görüntü sınıflandırma problemleri için geliştirilmiş olan bu model, yıllar içisinde daha karmaşık mimarilerle zenginleştirilmiştir. Bu mimariler, günümüzde çeşitli alanlarda kullanılmakta, özellikle derin sahte tespiti gibi karmaşık problemlerin çözümünde önemli bir temel oluşturmaktadır. Bu bölümde, CNN'nin öne çıkan mimarileri tanıtılarak, bu yapıların derin öğrenme uygulamalarındaki rolü incelenecektir.

- **Xception**

Xception, 2017'de ilk kez François Chollet tarafından önerilmiştir. Google tarafından geliştirilmiş olan ve Extreme Inception olarak adlandırılan bir mimaridir. Yani Inception modelinin evrimidir. Inception mimarisiyle benzerlikleri olsa da, Xception, çok daha derin bir yapıya sahiptir [45]. Bu mimari, özellikle derin öğrenme uygulamalarında evrişim katmanlarını daha verimli hale getirmek için derinlik ayrıştırması kavramını kullanır. Geleneksel evrişim katmanlarının yerine derinlik ayrıştırmalı evrişimler, iki aşamalı işlem içerir.

Derinlik ayrıştırmalı evrişim işleminde, her giriş kanalındaki filtreler bağımsız olarak işlenir. Bu da, çok daha az parametre kullanılarak işlemin yapılmasını sağlar. Puanlama evrişimi aşamasında ise, her bir kanalın birleşimi üzerinde işlem yapılır, bu da kanal sayısının artırılmasını sağlar.

Bu iki aşamalı işlem, modelin daha hızlı çalışmasını sağlar ve daha az hesaplama gücü gerektirir. Bu şekilde daha büyük veri setlerinde de hızlı bir şekilde eğitim yapılabilir.

Xception mimarisi, konvolüsyon katmanları, derinlik ayrıştırmalı evrişimler ve tam bağlantılı katmanlar içerir.

Derinlik ayrıştırmalı evrişimlerle, Xception çok daha az parametreyle yüksek doğruluk oranlarına ulaşabilir. Derin öğrenme modelleri arasında en hızlı eğitim süreçlerinden birine sahiptir. Bu mimari, görüntü sınıflandırmanın yanında nesne tespiti, yüz tanıma gibi çok farklı görsel analiz işlemlerinde de kullanılabilir.

- **EfficientNet**

Bu mimari, Google tarafından 2019 yılında Mingxing Tan ve Quoc V. Le tarafından önerilen bir mimaridir [46]. EfficientNet, büyük veri setlerinde yüksek doğruluk elde etmek için çok verimli bir yapıya sahiptir. Mimarinin amacı, daha az parametreyle daha yüksek performans elde etmektir. Böylece işlem süresini ve hesaplama gereksinimlerini azaltır. EfficientNet, klasik mimarilerde olduğu gibi evrişimsel katmanlar kullanır. Bu mimaride, bileşik ölçeklendirme (compound scaling) tekniği kullanılmaktasdır. Bu teknik, modelin derinlik, genişlik ve çözünürlük olmak üzere üç temel boyutunu eşit şekilde ölçeklendirir. Bu sayede modelin, verimliliği artar ve en iyi performansı gösterir.

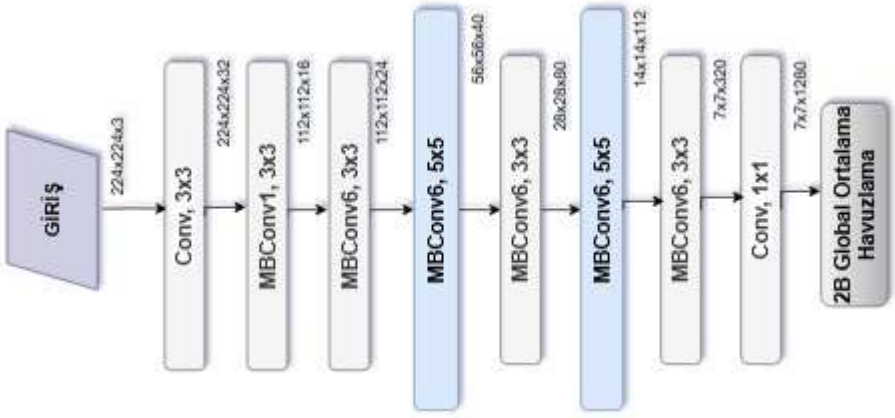
EfficientNet, 7 farklı model versiyonuna sahiptir (EfficientNet-B0, EfficientNet-B1, ..., EfficientNet-B7). Bu versiyonlar, bileşik ölçeklendirme kullanılarak farklı ihtiyaçlara göre optimize edilmiştir. Modelin her versiyonunun temel yapısı ve tasarımı benzer olmasına rağmen katman sayısı, filtre sayısı ve giriş çözünürlüğü bakımından farklıdır.

EfficientNet-B0 modeli, en hızlı çalışan, küçük boyutlu ve daha az hesaplama gerektiren bir modeldir. EfficientNet-B7 ise, yüksek doğruluk gerektiren işlemler için uygun, daha karmaşık ve büyük bir modeldir.

EfficientNet, geleneksel evrişimsel sinir ağı mimarilerine göre çok daha az parametreyle yüksek doğruluk elde edebilir. Bu, özellikle büyük veri setleri ve karmaşık işlemler için önemli bir avantaj sağlar. Diğer derin öğrenme modellerine göre daha az hesaplama gücü ve bellek gerektirir. Bu da, daha düşük maliyetli uygulamalarda ya da sınırlı donanıma sahip sistemlerde de kullanılmasını sağlar. Modelin farklı versiyonları arasında, ihtiyaca göre ölçeklendirme yaparak, işlemler için en uygun modelin seçilmesini sağlar.

Bu mimari, görüntü sınıflandırma, nesne tespiti, segmentasyon gibi görsel işlem görevlerinde etkili şekilde kullanılabilir.

Bu mimarinin genel yapısı Şekil 3.8'de sunulmuştur.



Şekil 3.8. EfficientNet mimarisi

- **ResNet**

ResNet 2015'te Microsoft Research tarafından derin ağlardaki öğrenme problemini çözmek için tasarlanmıştır [47]. Geleneksel CNN'lerde katmanlar ardışık eklenirken, ResNet'te bir katmanın çıktısı sonraki bir veya daha fazla katmanın çıktısına doğrudan eklenir (skip connection). Bu da, gradyan kaybını (vanishing gradient) önler.

$$y = F(x, \{W_i\}) + x \tag{3.1}$$

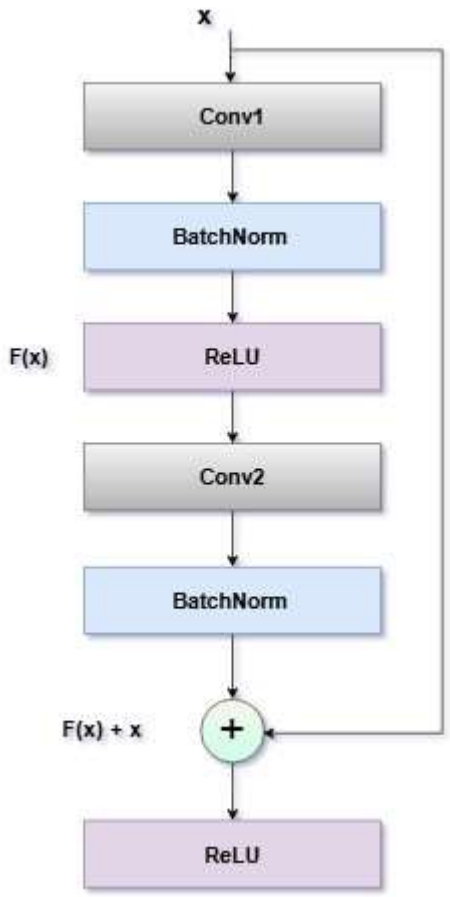
Residual Blocks denilen atlamalı bağlantılarla daha derin ağların eğitilmesini sağlar. ResNet-18, ResNet-34, ResNet-50, ResNet-101, ResNet-152 gibi farklı derinlikte versiyonları vardır

(sayılar katman sayısını belirtir). Residual bağlantılar, gerektiğinde katmanların kimlik fonksiyonunu öğrenmesini sağlar.

Basic Block: ResNet-18/34'te kullanılır. İki 3x3 konvolüsyon katmanı ve artık bağlantı içerir.

Bottleneck Block: ResNet-50/101/152'de kullanılır. 1x1, 3x3, 1x1 konvolüsyonlar ile hesaplama maliyetini düşürür. Bu, derin ağlarda performans düşüşünü engeller. Geleneksel CNN'lerde 20+ katmandan sonra performans düşerken, ResNet ile 100+ katman başarıyla eğitilebilir.

Görüntü sınıflandırma, nesne algılama ve birçok derin öğrenme uygulamasında kullanılmaktadır. Bu mimarinin yapısı ise Şekil 3.9'da gösterilmektedir.



Şekil 3.9. ResNet mimarisi

- **VGGNet**

VGGNet, 2014 yılında Oxford Üniversitesi tarafından geliştirilen ve daha derin ağların gücünü göstermek için tasarlanan bir CNN mimarisidir [48]. Görüntü sınıflandırma ve nesne tespiti gibi alanlarda kullanılır. Bu mimaride, katmanlar küçük 3x3 filtreler kullanılarak birleştirilir. 16-19 katmanlı versiyonlarıyla bilinir. Bunlar VGG-16 ve VGG-19'dur. VGGNet'te daha derin ağlar kullanılarak doğruluğun artırılabileceği gösterilmektedir.

- **AlexNet**

2012 yılında Alex Krizhevsky ve arkadaşları tarafından geliştirilmiş ve ImageNet yarışmasında büyük başarı elde etmiştir [49]. Bu mimari büyük veri setleri üzerinde görüntü sınıflandırma alanlarında kullanılır. 5 evrişim katmanı ve 3 tam bağlantılı katman olmak üzere 8 katmandan oluşur. İlk defa ReLU aktivasyon fonksiyonu kullanılarak daha hızlı öğrenme sağlanmıştır. Dropout ve veri artırma yöntemleri, aşırı öğrenmeyi önlemek için uygulanmıştır. Derin öğrenmenin geniş veri setlerinde kullanılmasını sağlamıştır. GPU'ların kullanımını popüler hale getirmiştir.

- **GoogleNet**

Bu mimari, 2014 yılında Google tarafından geliştirilen ve evrişim katmanlarının farklı boyutlarda aynı anda çalışmasını sağlayan bir mimaridir. Bu yapı, modelin hesaplama verimliliğini artırır. Görüntü sınıflandırma ve nesne algılama alanlarında kullanılır. Inception Module adı verilen birimlerle aynı katmanda birden fazla çekirdek boyutu kullanılır (1x1, 3x3, 5x5). Ağın genişlemesini kontrol altına almak için 1x1 filtrelerle parametre sayısı azaltılır.

3.4.3. CNN Modelleri için Kullanılan Kütüphaneler

Evrişimsel Sinir Ağı modellerinin geliştirilmesinde ve eğitiminde kullanılan kütüphaneler, derin öğrenme süreçlerini kolaylaştıran ve hızlandıran araçlardır. TensorFlow, PyTorch ve Keras gibi kütüphanaler, modelin yapısını oluşturmada ve eğitim sürecinde önemli bir yere sahiptir. OpenCV, NumPy, Pandas ve Scikit-Learn ise veri işleme ve analiz aşamalarında etkili bir şekilde kullanılır. Bu kütüphaneler, büyük veri setleriyle çalışmayı sağlar ve model performansını arttırmaya yardımcı olur.

- **TensorFLow**

2015 yılında Google tarafından geliştirilmiş, açık kaynaklı bir derin öğrenme kütüphanesidir. [50] Evrişimli Sinir Ağları gibi derin öğrenme modellerini tasarlamak ve eğitmek için yaygın olarak kullanılır. TensorFlow, hem yüksek seviyeli API'leri hem de düşük seviyeli API'ler sunar ve basit projelerden karmaşık olanlara kadar birçok gereksinimi karşılar. GPU ve TPU hızlandırmalarıyla büyük veri setlerinde yüksek performans sağlar. Mobil cihazlar ve web tarayıcıları gibi farklı platformlarda da kullanılabilir. Büyük veri kümeleri ve karmaşık CNN mimarileri için dağıtık hesaplama yapabilir.

- **PyTorch**

2017 yılında Facebook tarafından geliştirilen açık kaynaklı bir derin öğrenme kütüphanesidir. PyTorch, dinamik bir yapıya sahiptir. Özellikle araştırma ve geliştirme süreçlerinde sıkça tercih edilen bir kütüphanedir. CNN modellerini oluşturma ve eğitme sürecinde sunduğu esneklikle öne çıkar. [51] Bu kütüphane, GPU hızlandırmasını da destekler.

- **Keras**

Keras, derin öğrenme modellerini hızlı ve kolay bir şekilde oluşturmak için tasarlanmış, açık kaynaklı kütüphanedir. TensorFlow ile entegre bir şekilde çalışarak Evrişimli Sinir Ağları gibi modelleri kolayca tasarlama imkanı sağlaması en büyük avantajlarından biridir. Yüksek seviyeli bir API sağlar böylece kodlama sürecini basitleştirir. [52] Keras, GPU hızlandırmasını destekler aynı zamanda geniş bir topluluk desteği sunar.

- **OpenCV**

OpenCV (Open Source Computer Vision Library), görüntü işleme ve bilgisayarla görme uygulamaları için geliştirilmiş açık kaynaklı bir kütüphanedir. [53] Evrişimli Sinir Ağları ile çalışırken veri ön işleme aşamasında yaygın olarak kullanılmaktadır. Görüntüleri yeniden boyutlandırma, filtreleme, kırpma, dönüştürme gibi işlemleri gerçekleştirebilir. Bununla beraber video analizi, nesne takibi ve yüz algılama gibi birçok farklı işlev sağlar. Hızlı yapısı sayesinde büyük veri kümeleriyle çalışırken performans avantajı sağlar.

- **Numpy**

NumPy, bilimsel hesaplamalar için kullanılan açık kaynaklı bir Python kütüphanesidir. Evrişimli Sinir Ağlarıyla çalışırken veri yapılarını matrislere dönüştürmek için ve sayısal işlemleri verimli bir şekilde gerçekleştirmek amacıyla kullanılabilen bir araçtır. [54] NumPy, çok boyutlu dizileri ve matrisleri destekler ve çalışmayı kolaylaştırır.

- **Pandas**

Pandas, veri analizi ve işlemesi için geliştirilmiş açık kaynaklı bir Python kütüphanesidir. [55] Evrişimli Sinir Ağları ile çalışırken veri kümelerinin düzenlenmesi, analiz edilmesi gibi süreçlerde sıkça kullanılır. Özellikle tablo biçimindeki veriler üzerinde esnek işlemler yapmayı sağlar. Veri çerçeveleri ve diziler gibi yapılar sunar, eksik değerlerle çalışmak ya da veri dönüştürmek gibi işlemleri kolaylaştırır.

- **Scikit-Learn**

Scikit-Learn, makine öğrenimi ve veri analizi için kullanılan açık kaynaklı bir Python kütüphanesidir. Evrişimli Sinir Ağlarıyla çalışırken model doğrulama, regresyon ve kümeleme gibi görevlerde yaygın olarak kullanılır. Özellikle veri ölçeklendirme, özellik seçimi ve çapraz doğrulama gibi ön işleme adımları için güçlü araçlar sağlar. [56] Geniş algoritma yelpazesi ve kullanım kolaylığı sağlar, bu sayede araştırma ve uygulama geliştirme süreçlerinde de tercih edilir.

3.5. Uzun Kısa Süreli Bellek (Long Short-Term Memory – LSTM)

Derin öğrenme mimarilerinden biri olan Uzun Kısa Süreli Bellek, özellikle ardışık veri analizi ve zaman serisi problemlerinde kullanılan özel bir yinelenen sinir ağı (RNN) çeşididir. Schmidhuber ve Hochreiter tarafından 1997 yılında geliştirilen bu mimari, geleneksel RNN'lerin

karşılaştığı gradyan sönmesi (vanishing gradient) gibi problemleri çözmek amacıyla tasarlanmıştır [57]. LSTM'ler, hafıza hücreleri ve kapı mekanizmaları sayesinde bilgi akışını kontrol edebilir ve uzun vadeli bağımlılıkları etkili şekilde öğrenebilir.

3.5.1. LSTM'nin Temel Yapısı ve Kapı Mekanizmaları

LSTM ağları, her biri "hücre" olarak adlandırılan birimlerden oluşur. Bu hücreler, üç temel kapıdan oluşan karmaşık bir yapıya sahiptir: unutma kapısı (forget gate), giriş kapısı (input gate) ve çıkış kapısı (output gate) [58]. Bu kapılar, sigmoid ve hiperbolik tanjant (tanh) aktivasyon fonksiyonları kullanarak bilgi akışını düzenler. Hücre durumu, LSTM'nin uzun vadeli belleğini temsil eder ve kapılar tarafından kontrol edilen bir bilgi otoyolu görevi görür.

- **Unutma Kapısı**

Hücre durumunda bulunan hangi bilgilerin silinmesi gerektiğini belirler. Geçmiş durum ve mevcut giriş bilgilerine göre sigmoid aktivasyon fonksiyonu aracılığıyla çalışır.

- **Giriş Kapısı**

Hücreye hangi yeni bilgilerin ekleneceğini kontrol eder. Sigmoid ve tanh katmanları birlikte çalışarak yeni bilgi oluşturur ve bunu hücreye ekler.

- **Çıkış Kapısı**

Hücrenin çıkışını yani sonraki zaman adımına neyin aktarılacağını belirler. Bu çıkış, hücre durumunun bir kısmını temsil eder.

Bu mekanizmalar sayesinde LSTM ağları, önceki zaman adımlarından gelen bilgiyi unutmadan, yeni bilgiyle birlikte etkili bir şekilde işler.

4. MATERYAL VE METOT

Bu bölümde, derin sahte tespiti için geliştirilen öz-denetimli yaklaşım ile çok modlu derin öğrenme modelinin detayları ve kullanılan metodoloji sunulmaktadır. Çalışmada kullanılan veri setleri, ön işleme adımları, kullanılan derin öğrenme modelleri, eğitim süreci ve performans değerlendirmesi detaylandıracaktır.

4.1. Veri Seti

Bu tez çalışmasında, geliştirilen iki farklı metodolojinin (öz-denetimli ve çok modlu) kendine özgü gereksinimlerini karşılamak üzere, alanda yaygın olarak kullanılan ve farklı özelliklere sahip iki ayrı veri seti tercih edilmiştir.

Öz-denetimli öğrenme yaklaşımında kullanılan "Deepfake and Real Images" veri seti [59], çeşitli kaynaklardan elde edilmiş sahte ve gerçek yüz görsellerinden oluşmaktadır. Çalışma için 1200 görüntü verisi kullanılmıştır. Veri yükleme sürecinde, sınıf dengesizliğini önlemek amacıyla otomatik balanslama yapılmış ve her sınıftan eşit sayıda örnek seçilmiştir. Veri seti, özellikle sınırlı etiketli veriyle derin temsiller öğrenmek için uygundur. Görüntüler farklı sahne yapılarından ve yüz ifadelerinden oluşur; bu da modele çeşitli açılardan genelleme kabiliyeti kazandırmayı amaçlar.

Çok modlu yaklaşımda kullanılan veri seti, derin sahte tespiti alanında yaygın olarak bilinen ve geniş ölçekli bir veri seti olan DeepFake Detection Challenge (DFDC)'dır. [60] DFDC veri seti, Meta (eski Facebook) şirketi tarafından desteklenen ve bir yarışma kapsamında oluşturulmuştur. Bu veri seti, yüz manipülasyon teknikleriyle oluşturulmuş sahte videolar ile orijinal videolardan oluşmaktadır. Çalışma için, DFDC veri setinin belirli bir alt bölümü kullanılmıştır (dfdc_train_part_35). Bu veri kümesi, JSON formatında meta veri dosyası (metadata.json) ile birlikte sunulmaktadır. Meta veri dosyası, her bir videonun 'REAL' (gerçek) veya 'FAKE' (sahte) olarak işaretlenmiş etiket bilgisini içermektedir.

4.2. Veri Ön İşleme

Bu bölümde, çalışmada kullanılan veri setleri farklı biçimlerde ve modalitelerde olduğu için her bir deneysel yaklaşım için özel veri ön işleme adımları uygulanmıştır.

4.2.1. Öz-Denetimli Yaklaşım için Veri Ön İşleme

Öz-denetimli yaklaşımda kullanılan "Deepfake and Real Images" veri setinde kullanılan görseller, model girişine uygun olacak şekilde ön işlemden geçirilmiştir. Görseller, 224×224 piksel boyutuna yeniden ölçeklendirilmiştir ve görüntülere rastgele kırpma, yatay çevirme, renk tonu ve parlaklık değiştirme gibi çeşitli geometrik ve renk tabanlı dönüşümler dahil edilmiştir. Bu veri

artırma adımları, modelin genelleme kapasitesini artırmayı amaçlamaktadır. Tüm görseller, son olarak normalize edilerek tensör biçimine dönüştürülmüştür. Ön eğitim aşamasında, bu ön işlemden geçmiş görüntüler, dört farklı açıyla (0°, 90°, 180°, 270°) döndürülmüş ve bu açılar sınıf etiketi olarak atanmıştır. Böylece model, herhangi bir etiketli veri kullanmadan görsellerin uzamsal yapısını öğrenmeye yönelik öz-denetimli bir görevle eğitilmiştir. Ön eğitim tamamlandıktan sonra, aynı görsel önişleme adımları kullanılarak model, gerçek/sahte sınıflandırması görevine aktarılmıştır.

4.2.2. Çok Modlu Yaklaşım için Görsel Veri Ön İşleme

DFDC veri setindeki her bir video, sahte veya gerçek yüz içeren karelerden oluşmaktadır. Görsel verilerin işlenmesinde ilk adım olarak, videolardan kareler çıkarılmış ve bu kareler üzerinde yüz algılama işlemi gerçekleştirilmiştir. Yüz tespiti için MTCNN (Multi-task Cascaded Convolutional Networks) modeli kullanılmıştır. Bu model, PyTorch tabanlı FaceNet-pytorch kütüphanesi ile uygulanmıştır. Her video için 10 adet kare eşit aralıklarla seçilmiştir. Tespit edilen yüz bölgeleri 128×128 piksel boyutuna yeniden ölçeklendirilmiştir. Normalizasyon işlemi için ImageNet standartları referans alınmıştır.(ortalama: [0.485, 0.456, 0.406], standart sapma: [0.229, 0.224, 0.225]). Modelin genelleme yeteneğini artırmak veri artırma işlemi uygulanmıştır. Video dosyasının açılamaması, okunamaması, MTCNN'in belirli bir karede yüz tespit edememesi veya herhangi bir işleme hatası durumunda, ilgili kare yerine 128x128x3 boyutunda sıfırlardan oluşan bir NumPy dizisi (placeholder) kullanılmıştır. Eğer bir videodan toplamda 10 adetten az sayıda geçerli yüz karesi çıkarılmışsa, eksik kalan kareler bu placeholder dizilerle tamamlanarak her video için sabit sayıda (10) kare olması sağlanmıştır.

4.2.3. Çok Modlu Yaklaşım için İşitsel Veri Ön İşleme

Ses verilerini işleme adımında her videodan 3 saniyelik ses verisi alınmıştır. Yüklenen ses sinyalinden librosa kütüphanesi kullanılarak 128 mel bandlı spektrogramlar oluşturulmuştur. Mel spektrogramı hesaplanmasında spektrogram penceresi 25ms, adım aralığı 10ms olarak ayarlanmıştır. Elde edilen spektrogram, daha yaygın kullanılan logaritmik bir ölçek olan desibel (dB) ölçeğine dönüştürülmüştür. Spektrogramın zaman eksenindeki uzunluğunun tüm örnekler için aynı olması amacıyla, 128×128 boyutunda yeniden ölçeklendirilmiştir. Spektrogram bu hedeften kısaysa, spektrogramdaki en düşük desibel değeri kullanılarak sonuna dolgu (padding) eklenmiş; hedeften uzunsa sondan kırpılmıştır. Spektrogram değerleri 0-1 aralığına min-max normalizasyonu ile ölçeklendirilmiştir. Minimum ve maksimum değerler her bir spektrogram için ayrı ayrı hesaplanmıştır. Ses dosyasının bulunamaması, yüklenememesi veya librosa işlemleri sırasında bir

hata oluşması durumunda, 128x128 boyutunda sıfırlardan oluşan bir NumPy dizisi (placeholder) döndürülmüştür.

4.3. Model Mimarisi

Bu tez çalışmasında, derin sahte tespit problemine yönelik geliştirilen iki farklı yaklaşım için özgün model mimarileri tasarlanmış ve kullanılmıştır. Bu bölümde, her bir çalışmanın mimari yapısı ayrı başlıklar altında detaylandırılmaktadır.

4.3.1. Öz-Denetimli Öğrenme Tabanlı Mimari

Bu çalışmada önerilen öz-denetimli yaklaşım, iki aşamalı bir mimariden oluşmaktadır. İlk aşamada, ResNet18 mimarisi tabanlı bir model, döndürülmüş görüntüleri tanımaya yönelik bir görevle (rotation prediction) öz-denetimli olarak eğitilmiştir. Bu aşamanın temel amacı, etiketli veriye ihtiyaç duymadan görsel özellikler öğrenmektir. Bu amaçla kullanılan RotationNet modeli iki ana kısımdan oluşur:

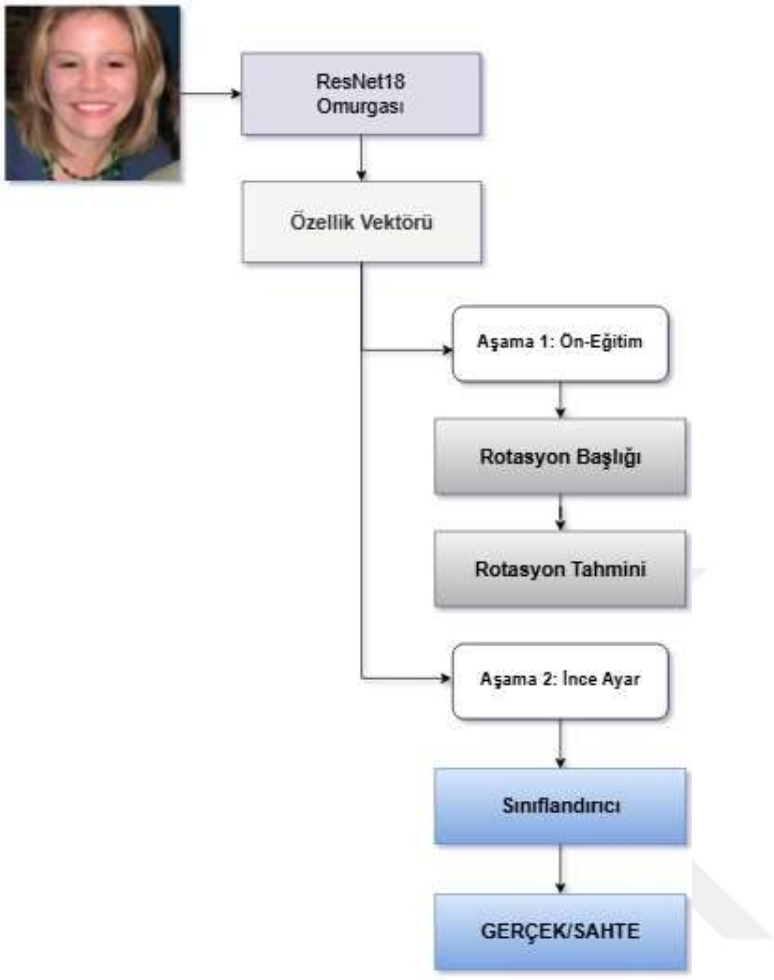
Omurga (Backbone): ImageNet üzerinde ön-eğitimli bir ResNet-18 mimarisi temel özellik çıkarıcı olarak kullanılmıştır. Modelin son tam bağlantılı katmanı (fully connected layer) çıkarılarak, önceki evrişimsel katmanlardan 512 boyutlu bir özellik vektörü elde edilmesi sağlanmıştır.

Rotasyon Başlığı (Rotation Head): Omurgadan gelen özellik vektörünü girdi olarak alan ve görüntünün hangi açıyla (0°, 90°, 180°, 270°) döndürüldüğünü tahmin eden Çok Katmanlı Algılayıcı (MLP) başlığıdır. Bu başlık, ardışık bir tam bağlı katman, ReLU aktivasyonu ve dropout katmanından oluşmaktadır.

Ön eğitim tamamlandıktan sonra, aynı omurga modeli gerçek/sahte sınıflandırması yapmak üzere ikinci bir başlık ile birlikte yeniden eğitilmiştir. Bu aşamada modelin genel yapısı aşağıdaki gibidir:

Transfer Öğrenme: RotationNet modelinde eğitilmiş olan ResNet-18 omurgası, öğrenilmiş ağırlıklarıyla birlikte bu yeni modele transfer edilir. Rotasyon başlığı ise atılır.

Sınıflandırıcı Başlığı (Classifier Head): Transfer edilen omurganın üzerine, derin sahte tespiti için sıfırdan tasarlanmış yeni bir sınıflandırıcı başlığı eklenir. Bu başlık, ReLU aktivasyon fonksiyonu ve %50 oranında bir Dropout katmanı içeren birinci tam bağlantılı katman, yine ReLU aktivasyon fonksiyonu ve %30 oranında bir Dropout katmanı içeren ikinci tam bağlantılı katman, "Gerçek" ve "Sahte" olmak üzere 2 sınıf için olasılık skorları üreten bir çıktı katmanından oluşur. Özellikle sahte içeriklerin tespitini iyileştirmek için ağırlıklı çapraz entropi kaybı kullanılmış ve sahte sınıfına 5 kat daha fazla önem verilmiştir. Eğitim sürecinde, AdamW optimizasyon algoritması ve dinamik öğrenme oranı ayarlama (ReduceLROnPlateau) kullanılmıştır.

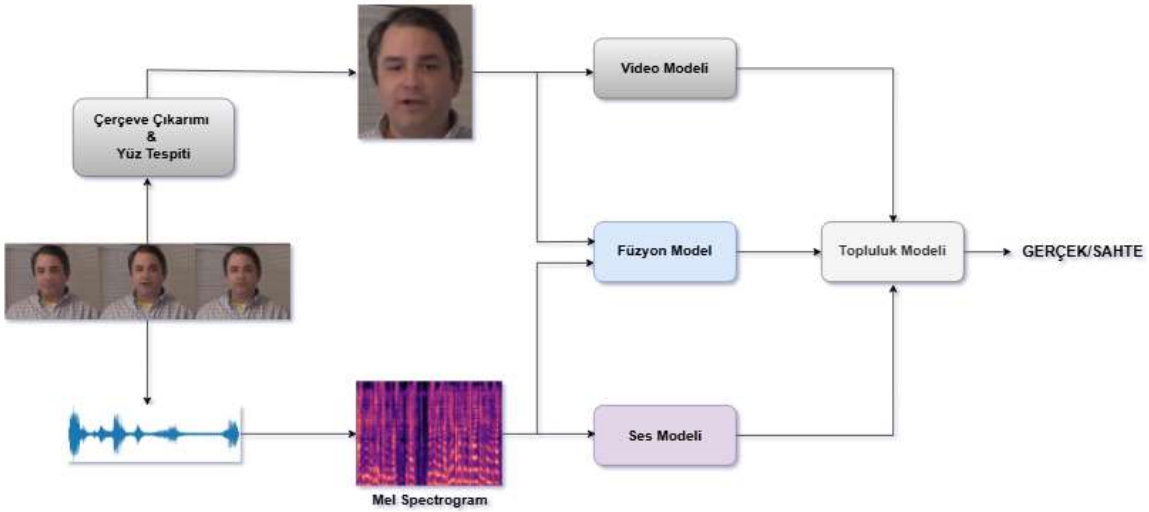


Şekil 4.1. Öz-denetimli model mimarisi

Bu iki aşamalı mimari sayesinde, model önce görsel temsilleri etiket olmadan öğrenir, daha sonra bu temsilleri sahte içerik tespitinde kullanır. Bu yaklaşımda kullanılan modelin genel mimarisi Şekil 4.1'de sunulmuştur.

4.3.2. Çok Modlu Topluluk Model Mimarisi

Bu çalışmada geliştirilen model, üç ana bileşenden oluşan hibrit bir ensemble mimarisine sahiptir. Bu bileşenler görsel, işitsel ve çok modlu özellikleri ayrı ayrı işleyerek sonuçları birleştirmektedir. Bu modeller şunlardır: yalnızca video karelerini işleyen model (VideoModel), yalnızca ses spektrogramlarını işleyen model (AudioModel) ve her iki modaliteden çıkarılan özellikleri birleştirerek işleyen bir füzyon modeli (AudioVisualModel). Son olarak, bu üç modelin çıktıları öğrenilebilir ağırlıklarla birleştirilerek topluluk modeli (EnsembleModel) oluşturulmuştur. Bu bölümde her bir bileşenin detaylı mimarisi ve işleyiş mekanizması açıklanmaktadır.



Şekil 4.2. Topluluk model mimarisi

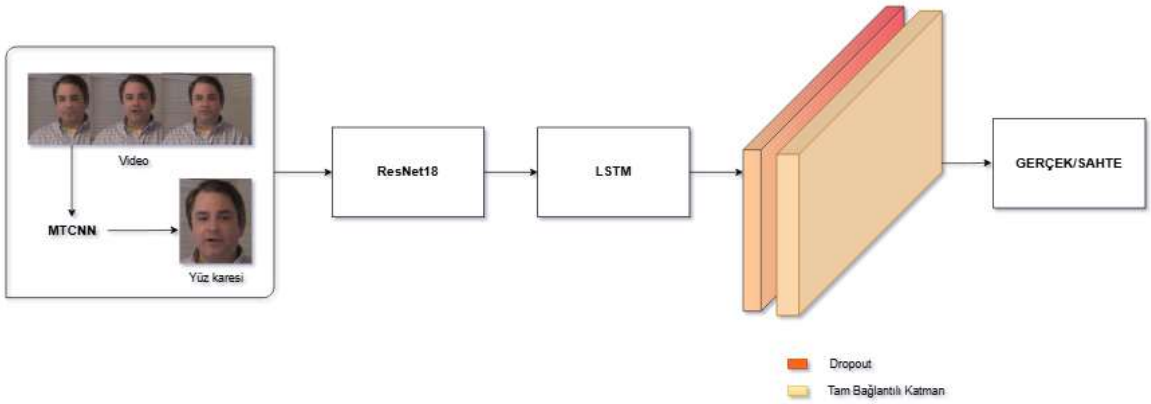
4.3.3. Video Modeli

Modelin ilk aşamasında, her videodan eşit aralıklarla seçilen 10 adet video karesi MTCNN (Multi-Task Cascaded Convolutional Neural Network) ile yüz bölgelerini tespit ederek, önceden eğitilmiş bir ResNet18 mimarisi ile işlenmektedir. ResNet18 modeli, ImageNet veri seti üzerinde eğitilmiş olup genel görsel özellikleri tanımada yüksek başarı göstermektedir. Ancak bu çalışmada modelin orijinal yapısında bir değişiklik yapılarak son tam bağlantılı katman çıkarılmış ve 512 boyutlu bir özellik vektörü elde edilmiştir. Bu değişiklik sayesinde model, yüz karelerindeki derin özellikleri daha etkin bir şekilde çıkarabilmektedir.

Özellik çıkarımı sonrasında, video kareleri arasındaki zaman bağımlılıklarını modellemek için bir LSTM ağı kullanılmaktadır. LSTM katmanı, 256 birimli video kareleri arasındaki tutarsız yüz hareketleri, anormal mimik değişimleri ve yapay görünen ifadeler gibi derin sahtelere özgü artefaktları yakalayabilmektedir. LSTM'nin bu yeteneği, özellikle yüz ifadelerinin doğal akışındaki anormallikleri tespit etmede kritik rol oynamaktadır. Zamansal analiz katmanının çıktıları, 0.5 dropout oranına sahip bir tam bağlantılı katmana beslenmekte ve burada gerçek ve sahte videolar arasındaki ayrım yapılmaktadır. Dropout katmanı, modelin aşırı öğrenme problemini engelleyerek, genelleme yeteneğini artırmaktadır.

Modelin mimarisi, video karelerindeki hem uzamsal hem de zamansal bilgiyi birlikte işleyebilme yeteneğine sahiptir. CNN katmanları her bir karedeki statik manipülasyon izlerini (örneğin renk tutarsızlıkları) yakalarken, LSTM katmanları bu izlerin zaman içindeki değişim şekillerini analiz etmektedir. Bu iki bileşenin entegrasyonu, modelin tutarlı bir şekilde karar verebilmesini sağlamaktadır.

Modelin genel yapısı Şekil 4.3'te gösterilmiştir.



Şekil 4.3. Video modeli

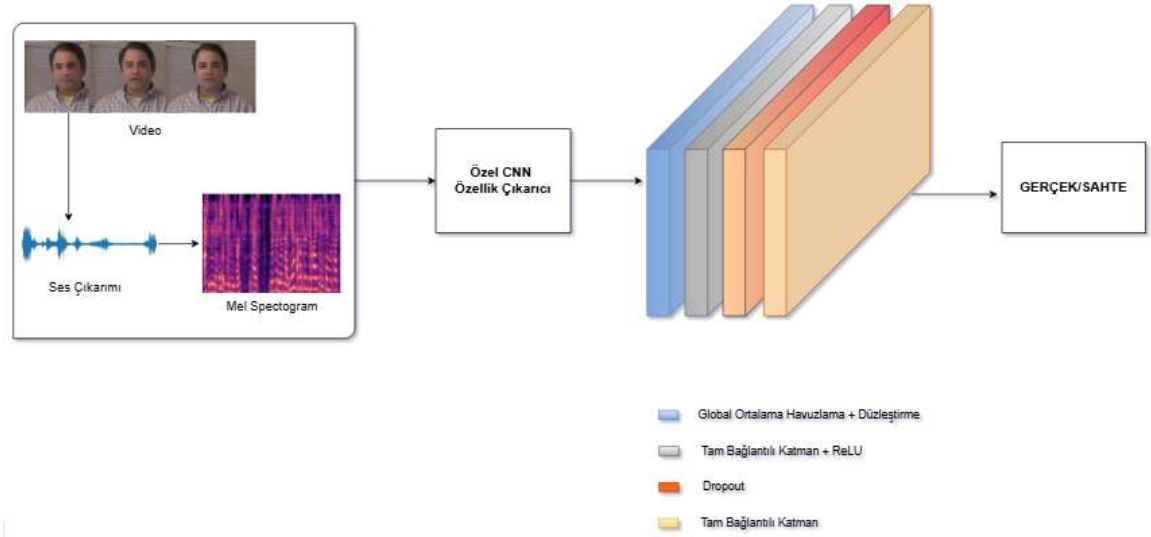
4.3.4. Ses Modeli

Ses akış modeli, derin sahte videolardaki işitsel manipülasyon izlerini tespit etmek üzere tasarlanmış özel bir derin öğrenme mimarisidir. Bu model, ses sinyalinin spektral özelliklerini analiz ederek yapay ses sentezi veya ses manipülasyonu gibi anormallikleri belirlemeyi amaçlamaktadır.

Modelin temel girdisi olarak ön işleme adımında oluşturulan mel spektrogramları kullanılmaktadır. Mel spektrogramlar, ses sinyalinin zaman-frekans dağılımını gösteren ve insan işitme sisteminin frekans algısına uygun şekilde ölçeklendirilmiş iki boyutlu temsillerdir.

Ses analizi için geliştirilen model, dört katmanlı derin bir evrişimli sinir ağı (CNN) mimarisine sahiptir. Her evrişim katmanı 3×3 boyutunda filtreler içermekte ve ardışık olarak 32, 64, 128 ve 256 filtre sayısına sahip katmanlardan oluşmaktadır. Evrişim işlemlerinin ardından batch normalizasyon ve ReLU aktivasyon fonksiyonları uygulanmakta, son olarak 2×2 boyutunda max pooling ile boyut küçültme yapılmaktadır. Evrişimsel blokların çıktısı, değişken boyutlu girişleri sabit boyutlu vektörlere dönüştüren global ortalama havuzlama katmanına iletilmektedir. Bu katman sayesinde model, farklı uzunluktaki ses kayıtlarından aynı boyutta özellik vektörleri çıkarabilmektedir. Havuzlama sonrasında elde edilen 256 boyutlu özellik vektörü, 128 birimli tam bağlantılı bir katmana beslenmekte ve burada daha yüksek seviyeli özellik temsilleri öğrenilmektedir. Modelin son katmanında, 128 boyutlu vektör 2 çıkış birimine (gerçek ve sahte) sahip bir sınıflandırıcıya bağlanmaktadır.

Model mimarisi ise Şekil 4.4'te sunulmuştur.



Şekil 4.4. Ses modeli

4.3.5. Çok Modlu Füzyon Modeli

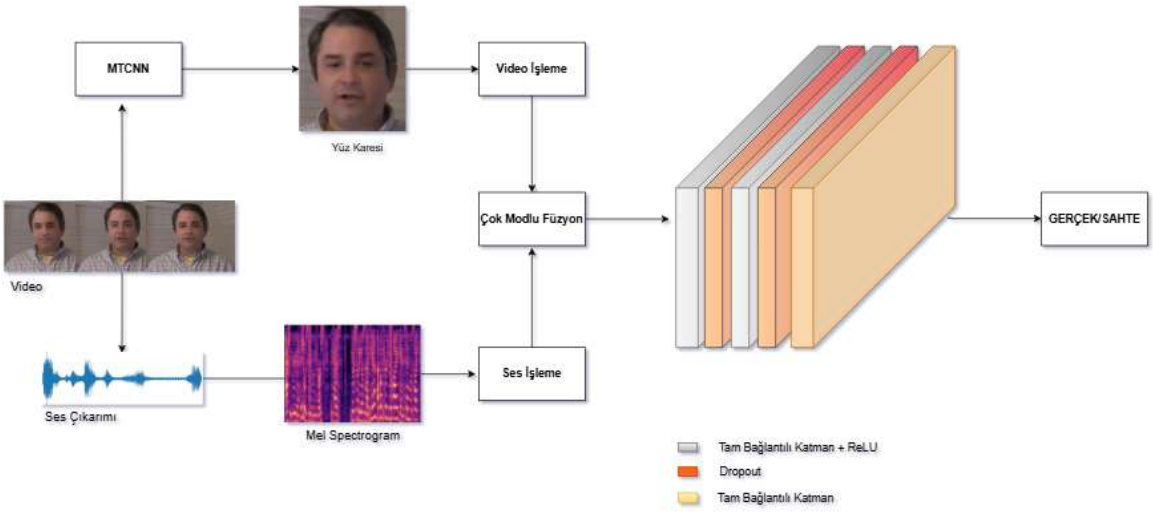
Çok modlu füzyon (Audio-Visual) modeli, derin sahte tespitinde hem görsel hem de işitsel verileri birleştiren güçlü bir çok modlu yaklaşımdır. Bu model, videoların manipüle edildiğini tespit etmek için yüz görüntüleri ve ses verilerinin eşzamanlı analizini gerçekleştirir. Bu model, kendi içerisinde görsel ve işitsel özellik çıkarıcı tabanlarını barındırır.

Görsel için, video akışından seçilen kare ResNet-18 mimarisi kullanılarak transfer öğrenme yaklaşımı ile son sınıflandırma katmanı çıkarılarak, 512 boyutlu yüksek seviyeli görsel özellik vektörü elde edilmiştir. Ses için, dört katmanlı bir CNN mimarisi kullanılmaktadır. Ses verisi, mel-spektrogramlara dönüştürülerek CNN'e giriş olarak verilmektedir. Burada 256 boyutlu özellik vektörü ek bir tam bağlantılı katmanla 128 boyuta indirgenmektedir. Her iki modalite için de batch normalizasyon ve ReLU aktivasyonları uygulanarak özellik dağılımlarının stabilize edilmesi sağlanmaktadır.

Çıkarılan özellik vektörleri, özel bir birleştirme katmanında entegre edilmektedir. Füzyon katmanları, ResNet-18'den elde edilen 512 boyutlu görsel özellikler ile özel CNN'den gelen ve 128 boyutuna indirgenen işitsel özellikleri birleştirir. Bu birleştirme işlemi ile 640 boyutlu (512+128) birleşik bir özellik vektörü oluşturulur. Birleştirme sonrası, birleşik özellikler bir dizi tam bağlantılı katmandan geçer. İlk tam bağlantılı katman, 640 boyutlu girdi özelliklerini 512 boyutlu katmana dönüştürür. Bu dönüşüm, modaliteler arası ilişkilerin öğrenilmesini sağlar. ReLU aktivasyon fonksiyonu bu aşamada uygulanarak doğrusal olmayan bir dönüşüm gerçekleştirilir, böylece model daha karmaşık ve soyut desenleri yakalayabilir. Ardından, modelin genelleme yeteneğini artırmak ve aşırı öğrenmeyi engellemek için 0.5 oranında dropout katmanı uygulanır. İkinci tam bağlantılı katman 512 boyuttan 256 boyuta indirgenir. Burada da ReLU aktivasyonu kullanılır ve bir kez daha dropout uygulanır.

Son füzyon katmanı, 256 boyutlu özellikleri iki sınıflı çıktıya (gerçek/sahte) indirger. Bu katmanda aktivasyon fonksiyonu kullanılmaz çünkü çıktı, sınıflandırma için gerekli olan logit değerleridir. Bu logitler daha sonra softmax fonksiyonu ile olasılıksal tahminlere dönüştürülür. Füzyon katmanlarının önemli bir özelliği, her iki modaliteden gelen özellikleri sadece birleştirmekle kalmayıp, aralarındaki ilişkileri öğrenme yeteneğidir. Bu katmanlar, ses ve görüntü arasındaki tutarsızlıkları tespit ederek derin sahte içeriğinin karakteristik özelliklerini yakalayabilir. Örneğin, bir videoda dudak hareketleri ile ses arasındaki senkronizasyon bozukluğu bu katmanlar tarafından öğrenilebilir.

Bu modelin yapısı Şekil 4.5'te verilmektedir.



Şekil 4.5. Çok modlu füzyon modeli

4.3.6. Topluluk Modeli

Bu çalışmada, çok modlu füzyon modelin genelleme yeteneğini artırmak amacıyla bir topluluk (ensemble) model yaklaşımı benimsenmiştir. Topluluk modeli, farklı modalitelerden (video ve ses) ve farklı mimarilerden yararlanan üç temel modelin (video, ses ve çok modlu füzyon) tahminlerini birleştirerek nihai kararı vermektedir. Tanımlanan bu üç temel modelin (VideoModel, AudioModel, AudioVisualModel) tahminlerini ağırlıklı olarak birleştiren bir meta-model yaklaşımı benimsenmiştir ve bu mimari, PyTorch çerçevesinde uygulanmıştır. Ağırlıklar, softmax fonksiyonu kullanılarak normalize edilir. Nihai ensemble tahmini şu şekilde hesaplanır:

$$L_{ensemble} = w_0 \cdot L_{video} + w_1 \cdot L_{audio} + w_2 \cdot L_{av} \quad (4.1)$$

Topluluk model eğitimi için ana kayıp (ensemble çıktısı) ve yardımcı kayıplar (alt model çıktıları) olarak iki seviyeli bir kayıp fonksiyonu kullanılmıştır. Toplam kayıp fonksiyonu şu formül ile hesaplanmaktadır:

$$L_{total} = L_{ensemble} + \alpha(L_{video} + L_{audio} + L_{av}) \quad (4.2)$$

Burada α yardımcı kayıpların ağırlığını belirleyen hiperparametredir ve deneysel çalışmalarda 0.1 olarak belirlenmiştir.

- L_{video} : Video tabanlı model kaybı.
- L_{audio} : Ses tabanlı modelin kaybı.
- L_{av} : Görsel-işitsel füzyon modelinin kaybı.
- $L_{ensemble}$: Video, ses ve füzyon alt modellerin kayıplarının ağırlıklı ortalamasıdır.
- L_{total} : Toplam kayıp fonksiyonudur.
- w_0, w_1, w_2 : Kayıplar için öğrenilebilir ağırlık katsayılarını ifade eder.

Topluluk model yaklaşımı video, ses ve çok modlu (görsel-işitsel) olmak üzere üç farklı bilgi kaynağından yararlanılarak daha zengin bir özellik temsili elde eder. Her bir modelin katkısının sabit olmak yerine veri üzerinden öğrenilmesi, modelin adaptasyon yeteneğini artırır. Farklı modellerin hatalarını dengeleme potansiyeli ile genellikle tekil modellere kıyasla daha yüksek genelleme performansı gösterir. Şekil 4.2'de gösterilen mimari bu modeli temsil etmektedir.

4.4. Eğitim ve Deneysel Yapılandırma

Bu çalışmada geliştirilen modellerin eğitimi belirli bir yapılandırma ve strateji izlenerek gerçekleştirilmiştir.

4.4.1. Veri Seti Kullanımı

Bu tez kapsamında kullanılan veri setleri, ilgili yöntemlerin gereksinimlerine uygun şekilde alt kümelere ayrılarak kullanılmıştır.

Öz-denetimli öğrenme yaklaşımında, "Deepfake and Real Images" veri seti kullanılmıştır. Bu veri setinden toplam 1200 görüntü verisi seçilmiştir. Bu görüntülerin 600'ü gerçek, 600'ü ise sahte olmak üzere dengeli bir alt küme oluşturulmuştur. Veriler eğitim, doğrulama ve test olmak üzere üç alt kümeye ayrılmıştır. Bu verinin %70'i eğitim, %15'i doğrulama (validation), %15'i ise test amaçlı kullanılmıştır.

Çok modlu yaklaşım için kullanılan DFDC veri setinin tamamı oldukça büyük olduğundan, hesaplama kaynaklarını etkin kullanmak ve eğitim sürelerini makul seviyelerde tutmak amacıyla veri setinin bir alt kümesi alınmıştır. Bu alt kümeden toplam 830 dengeli veri kullanılmıştır. Videoların gerçek veya sahte olduğuna dair etiket bilgileri, veri seti ile birlikte sunulan metadata.json dosyasından elde edilmiştir. Bu dengeleme adımı, modelin herhangi bir sınıfa aşırı odaklanmasını engellemeyi amaçlar. Örneklem oluşturulduktan sonra video listesi rastgele karıştırılmıştır. Bu veri seti, modelin eğitimi, doğrulanması ve test edilmesi amacıyla eğitim (train),

doğrulama (validation) ve test olarak ayrılmıştır. Toplam veri setinin %20'si test için kullanılmıştır. Tablo 4.1'de veri seti dağılımı gösterilmiştir.

Tablo 4.1. Veri seti dağılımı

Veri Seti	Modalite	Gerçek Veri	Sahte Veri	Toplam Veri
Deepfake and Real Images	Görsel	600	600	1200
DFDC	Görsel + Ses	415	415	830

Deepfake and Real Images veri seti, yüksek çözünürlüklü ve çeşitli bir koleksiyon olan FFHQ (Flickr-Faces-HQ) veri setinden alınan gerçek yüz fotoğraflarından oluşur ve yaş, cinsiyet, etnik köken gibi çeşitlilik içerir. Bu veri setindeki sahte içerikler ise, StyleGAN3 ve ProGAN gibi Üretken Çekişmeli Ağlar (GAN) kullanılarak, sentetik olarak üretilmiş yüzlerdir.

DFDC veri seti, birden fazla kaynaktan (oyuncular, kamuya açık figürler) toplanan videolardan oluşur ve yaş, cinsiyet, etnik çeşitlilik içerir. Veri setinde yüz değiştirme, ağız hareketi manipülasyonu, yüz ifadesi transferi, derin öğrenme tabanlı sahte üretim yöntemleri gibi çeşitli manipülasyon teknikleri yer almaktadır.

4.4.2. Eğitim Parametreleri

Öz-denetimli öğrenme yaklaşımı parametre ve değerleri sırasıyla Tablo 4.2 ve Tablo 4.3'te gösterilmiştir.

Tablo 4.2. RotationNet eğitim parametreleri

Hiperparametre	Değer
Optimizasyon Algoritması	AdamW
Öğrenme Oranı	1e-3
Batch Boyutu	32
Epoch Sayısı	20
Scheduler Faktörü	0.2
Scheduler Sabrı	3
Ağırlık Düşümü	1e-5

Tablo 4.3. Sınıflandırma eğitim parametreleri

Hiperparametre	Değer
Optimizasyon Algoritması	AdamW
Öğrenme Oranı	Backbone: 0.00002, Classifier Head: 0.001
Batch Boyutu	32
Epoch Sayısı	45
Scheduler Faktörü	0.2
Scheduler Sabrı	3
Ağırlık Düşümü	1e-4

Geliştirilen ensemble modelin eğitimi de, belirli optimizasyon stratejileri, kayıp fonksiyonları ve hiperparametreler kullanılarak gerçekleştirilmiştir. Tablo 4.4'te parametre ve değerler gösterilmiştir.

Tablo 4.4. Topluluk model eğitim parametreleri

Hiperparametre	Değer
Optimizasyon Algoritması	Adam
Öğrenme Oranı	0.0005
Batch Boyutu	16
Epoch Sayısı	15
Scheduler Faktörü	0.5
Scheduler Sabrı	2
Ağırlık Düşümü	1e-5

4.4.3. Değerlendirme Metrikleri

Model performansı aşağıdaki metriklerle ölçülmüştür:

- **Doğruluk (Accuracy)**

Tüm doğru tahminlerin toplam tahmin sayısına oranıdır.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (4.3)$$

Burada TP (True Positive) doğru tespit edilen sahte videoları, TN (True Negative) doğru tespit edilen gerçek videoları, FP (False Positive) yanlışlıkla sahte olarak tespit edilen gerçek videoları ve FN (False Negative) yanlışlıkla gerçek olarak tespit edilen sahte videoları ifade eder.

- **Kesinlik (Precision)**

Pozitif olarak tahmin edilen örneklerin ne kadarının gerçekten pozitif olduğunu ölçer.

$$Precision = \frac{TP}{TP + FP} \tag{4.4}$$

- **Duyarlılık (Recall)**

Gerçek pozitif örneklerin ne kadarının doğru tahmin edildiğini gösterir.

$$Recall = \frac{TP}{TP + FN} \tag{4.5}$$

- **F1 Skoru**

Precision ve Recall'un harmonik ortalamasıdır.

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{4.6}$$

- **Karmaşıklık Matrisi (Confusion Matrix)**

Modelin hangi sınıflarda hata yaptığını analiz etmek için kullanılır.

Sonuçlar sonraki bölümde detaylı olarak paylaşılacaktır.

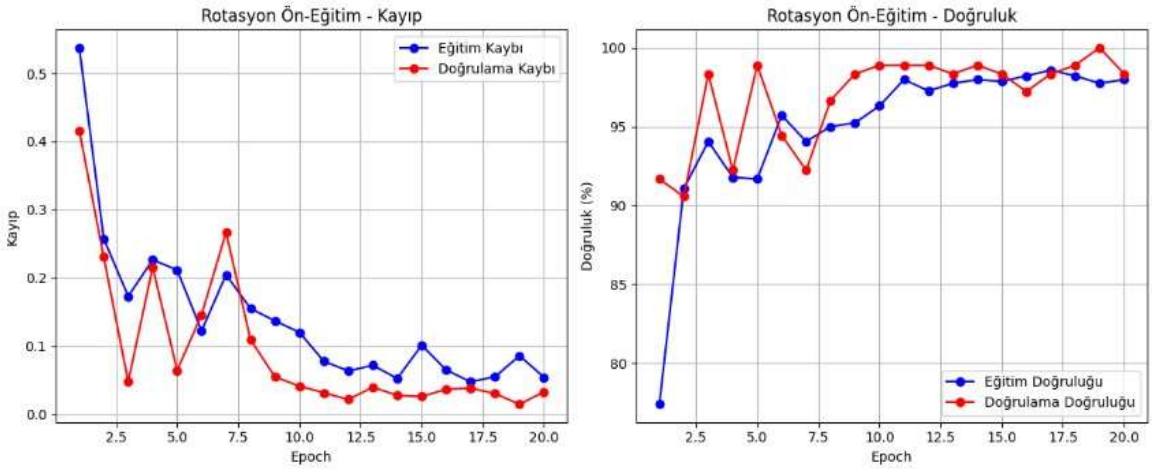
5. BULGULAR VE TARTIŞMA

Bu bölümde, tez kapsamında geliştirilen öz-denetimli ve çok modlu derin sahte tespit modellerinin deneysel sonuçları sunulacaktır. Her modelin eğitim süreci ve test performansı ayrı ayrı incelenecek, daha sonra bu yaklaşımın karşılaştırmalı bir analizi yapılarak elde edilen bulgular tartışılacaktır.

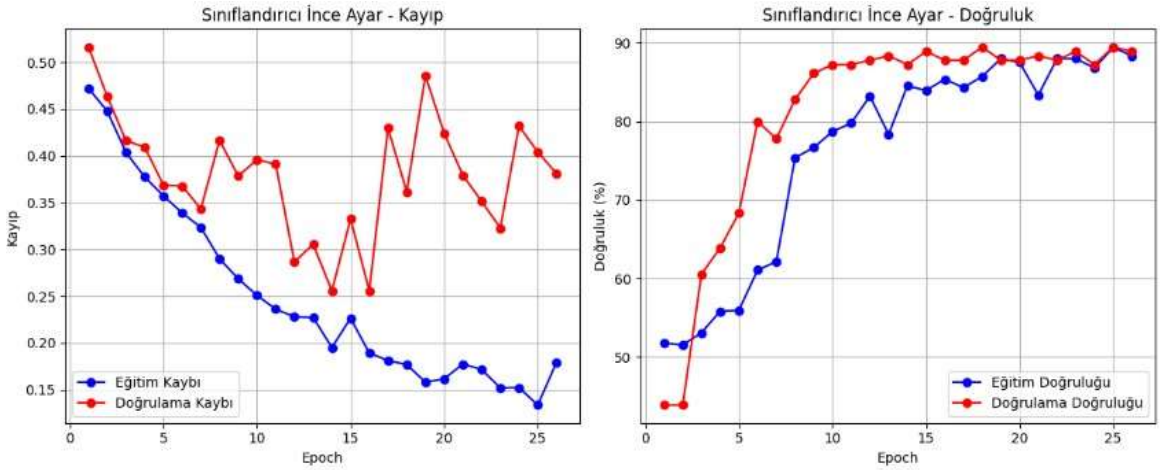
5.1. Öz-Denetimli Öğrenme Yaklaşımlı Model

Bu bölümde, geliştirilen öz-denetimli öğrenme modelinin performansı sunulmaktadır. Çalışmada Kaggle'da bulunan "deepfake-and-real-images" veri setinde 1200 görüntü verisi kullanılmıştır. Veri yükleme sürecinde, sınıf dengesizliğini önlemek amacıyla otomatik balanslama yapılmıştır.

Öz-denetimli öğrenme yaklaşımı iki aşamalı bir eğitim stratejisi ile uygulanmıştır. Bu kapsamında eğitilen modelin eğitim süreci boyunca elde edilen kayıp (loss) ve doğruluk (accuracy) değerlerinin değişimi Şekil 5.1 ve Şekil 5.2'de sunulmuştur.



Şekil 5.1. Rotasyon kayıp ve doğrulama grafiği



Şekil 5.2. Sınıflandırma kayıp ve doğrulama grafiği

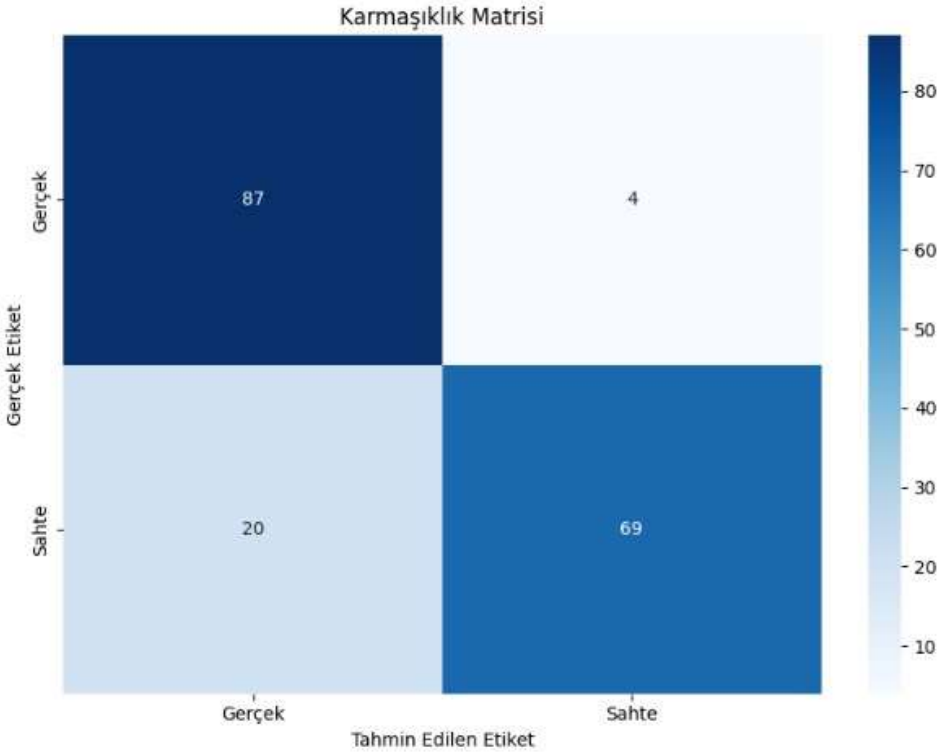
Model performansı, doğruluk (accuracy), kesinlik (precision), duyarlılık (recall), F1 skoru ve karmaşıklık matrisi kullanılarak değerlendirilmiştir. Modelin test aşamasında 0.3-0.5 aralığında eşik değer optimizasyonu yapılarak recall-precision dengesi incelenmiştir. Sonuçlar, karmaşıklık matrisi ve yanlış sınıflandırılan örneklerin görselleştirilmesiyle raporlanmıştır. Sistem, özellikle düşük eşik değerlerinde (0.3) sahte içeriklerin tespitinde daha yüksek duyarlılık değerine ulaşmıştır. Özellikle SAHTE sınıfına ait metrikler ve farklı karar eşik değerlerindeki (0.5, 0.4, 0.3) performans Tablo 5.1'de gösterilmiştir.

Tablo 5.1. Öz-denetimli öğrenme ile farklı karar eşiklerinde test performansı

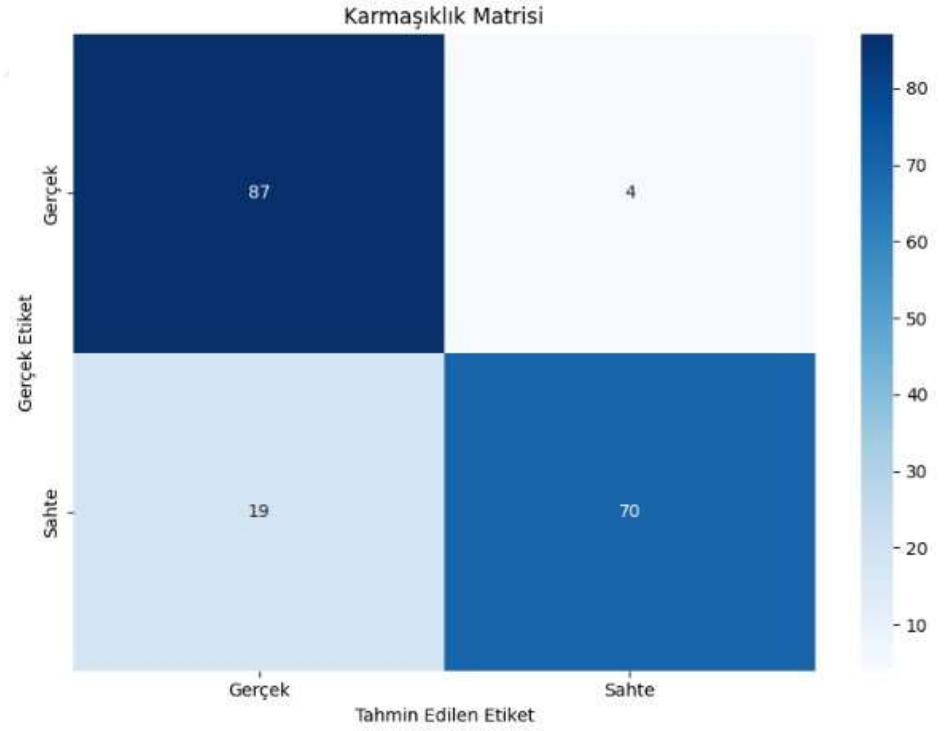
Karar Eşiği	Kesinlik(Sahte)	Duyarlılık(Sahte)	F1-Skoru(Sahte)	Doğruluk
0.5	%94.52	%77.53	%85.19	%86.67
0.4	%94.59	%78.65	%85.89	%87.22
0.3	%93.67	%83.15	%88.10	%88.89

Tablo 5.1'e bakıldığında, karar eşiği değeri düşürüldükçe SAHTE sınıfı için duyarlılık (recall) değerinin arttığı gözlemlenmektedir. Eşik değeri 0.5'te %77.53 olan duyarlılık, eşik 0.3'e çekildiğinde %83.15'e yükselmiştir. Bu, modelin daha fazla sahte videoyu doğru bir şekilde tespit edebildiği anlamına gelmektedir.

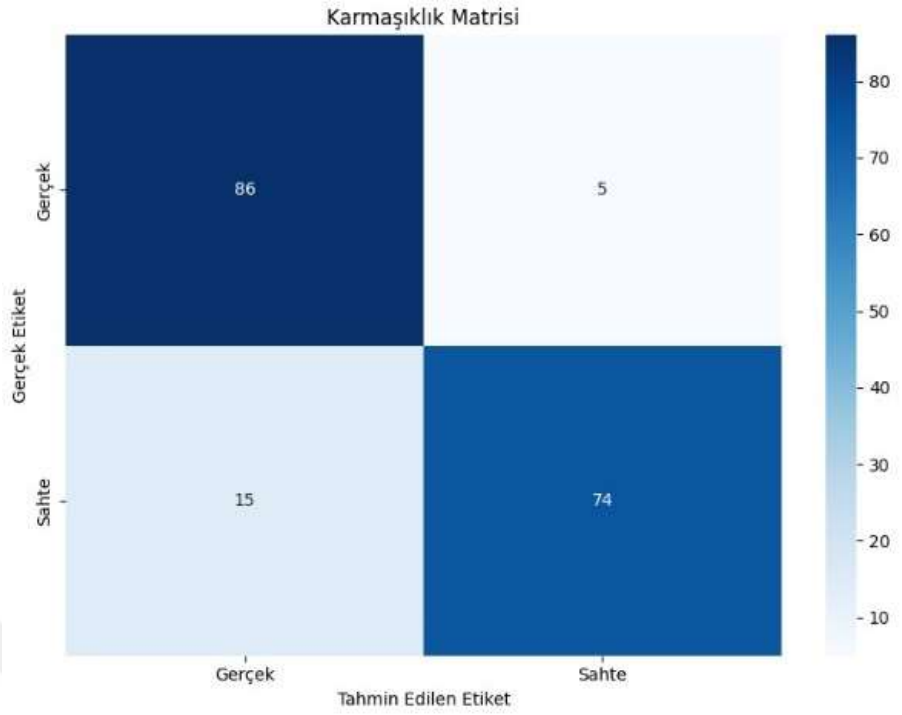
Farklı eşik değerleri için elde edilen karmaşıklık matrisleri Şekil 5.3, Şekil 5.4 ve Şekil 5.5'te sırasıyla 0.5, 0.4 ve 0.3 eşik değerleri için gösterilmektedir.



Şekil 5.3. 0.5 Eşik değeri için karmaşıklık matrisi



Şekil 5.4. 0.4 Eşik değeri için karmaşıklık matrisi



Şekil 5.5. 0.3 Eşik değeri için karmaşıklık matrisi

5.2. DFDC Veri Seti ile Çok Modlu Model

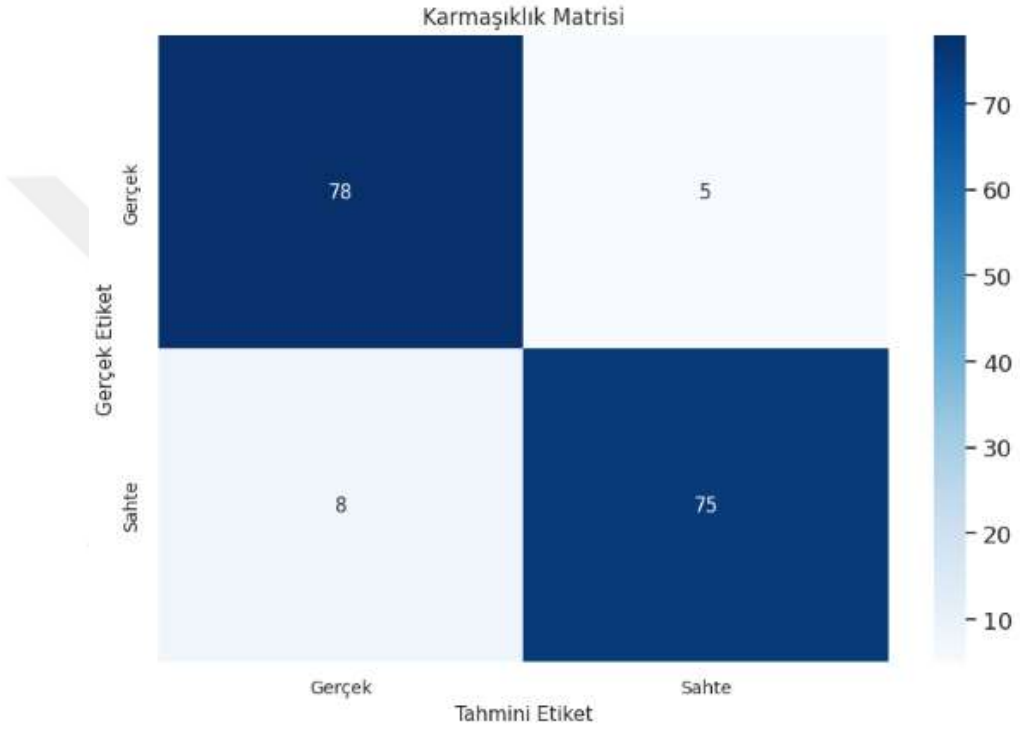
Bu bölümde, geliştirilen çok modlu derin sahte tespit modelinin performansı sunulmaktadır. Model, DeepFake Detection Challenge (DFDC) veri setinin alt kümesi (dfdc_train_part_35) kullanılarak değerlendirilmiştir. Veri seti, toplam 830 video örneği içerecek şekilde, dengeli sayıda örnek alınarak oluşturulmuştur. Bu veri setinin %20'si test verisi olarak ayrılmıştır. Modelin test seti üzerindeki genel performansı Tablo 5.2'de özetlenmiştir.

Tablo 5.2. Çok modlu modelin DFDC test performansı

Metrik	Değer
Test Kaybı	0.1985
Doğruluk	%92.17
SAHTE:	
Duyarlılık	%90.36
Kesinlik	%93.75
F1-Skoru	%92.02
GERÇEK:	
Duyarlılık	%93.98
Kesinlik	%90.70
F1-Skoru	%92.31

Tablo 5.2'de de görüldüğü gibi, geliştirilen model test verileri üzerinde %92.17'lik bir genel doğruluk oranına ulaşmıştır. Özellikle sahte videoların tespiti için önemli bir metrik olan SAHTE sınıfı duyarlılığı %90.36 olarak ölçülmüştür.

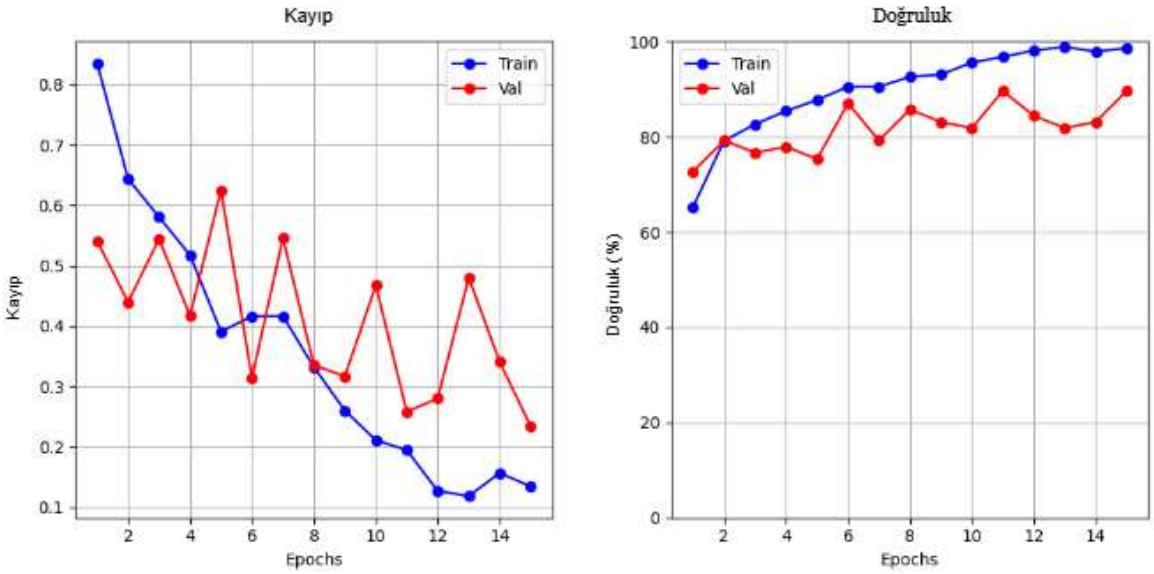
Modelin sınıflandırma performansını daha ayrıntılı incelemek ve yapılan hata türlerini analiz etmek amacıyla bir karmaşıklık matrisi oluşturulmuştur. Bu karmaşıklık matrisi Şekil 5.6'da gösterilmiştir.



Şekil 5.6. DFDC veri seti karmaşıklık matrisi

Karmaşıklık matrisine göre model, 83 Gerçek videonun 78'ini doğru bir şekilde gerçek olarak sınıflandırmış (Doğru Negatif - TN), ancak 5 gerçek videoyu yanlışlıkla sahte olarak etiketlemiştir (Yanlış Pozitif - FP). 83 sahte videonun 75'ini doğru bir şekilde sahte olarak sınıflandırmış (Doğru Pozitif - TP), ancak 8 sahte videoyu gözden kaçırarak gerçek olarak yanlış sınıflandırmıştır (Yanlış Negatif - FN). Bu değerler, modelin sahte videoları büyük oranda tespit edebildiğini göstermektedir.

Modelin eğitim sürecine ait kayıp ve doğruluk değişim grafikleri Şekil 5.7'de gösterilmiştir. Grafiklerde, modelin eğitim sürecinde istikrarlı bir şekilde optimizasyon sağladığı gözlemlenmektedir.



Şekil 5.7. Kayıp ve doğrulama grafikleri

Eğitim kaybı düzenli olarak düşerken, doğrulama kaybında zaman zaman dalgalanmalar görülmekle birlikte, genel eğilim düşüş yönündedir. Bu durum, modelin karmaşık yapısına rağmen aşırı uyum göstermediğini ve genelleme kapasitesinin makul düzeyde olduğunu göstermektedir.

Her sınıf (GERÇEK ve SAHTE) için kesinlik (precision), duyarlılık (recall) ve F1 skoru değerleri ise Tablo 5.3'te sunulan sınıflandırma raporunda detaylandırılmıştır.

Tablo 5.3. Çok modlu modelin sınıflandırma raporu

	Kesinlik	**Duyarlılık**	**F1-Skoru**
GERÇEK	0.91	0.94	0.92
SAHTE	0.94	0.90	0.92
Doğruluk	-	-	0.92

5.2.1. Çok Modlu Yaklaşım İçin Örnek Veri Gösterimi

Bu bölümde, çok modlu derin sahte tespit modelinin analiz ettiği görsel ve işitsel verilerin niteliğini göstermek amacıyla, kullanılan DFDC (Deepfake Detection Challenge) veri setinden seçilmiş gerçek ve sahte örnekler sunulacaktır.

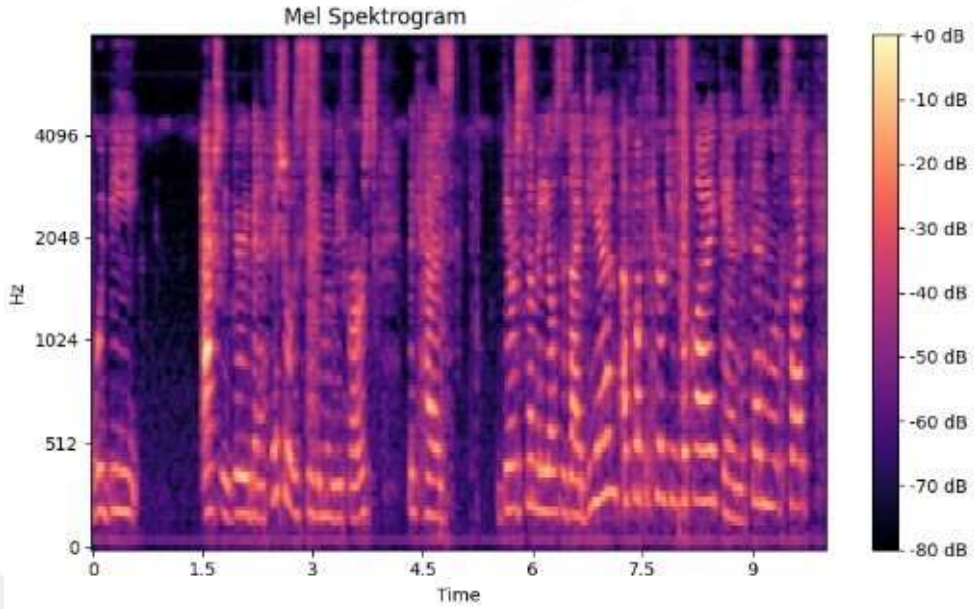
Modelin video akışı, videolardan çıkarılan yüz karesi dizilerini analiz etmektedir. Şekil 5.8'de DFDC veri setinden alınmış orijinal (gerçek) bir videodan çıkarılan ardışık birkaç yüz karesi gösterilmektedir. Bu kareler, doğal yüz ifadelerini, mimikleri ve kafa hareketlerini içermektedir. Modelin bu tür doğal değişimleri sahte manipülasyonlardan ayırt etmesi beklenir.

Şekil 5.8. Gerçek görüntü kareleri

Şekil 5.9'da ise aynı veri setinde, yüz değiştirme (face-swapping) veya ifade manipülasyonu (expression manipulation) gibi bir derin sahte tekniğiyle oluşturulmuş sahte bir videodan çıkarılan ardışık yüz kareleri sunulmaktadır. Bu karelerde, kenarlarda bulanıklık, renk uyumsuzluğu, doğal olmayan göz hareketleri, mimiklerdeki yapaylık gibi artefaktlar dikkat çekmektedir.

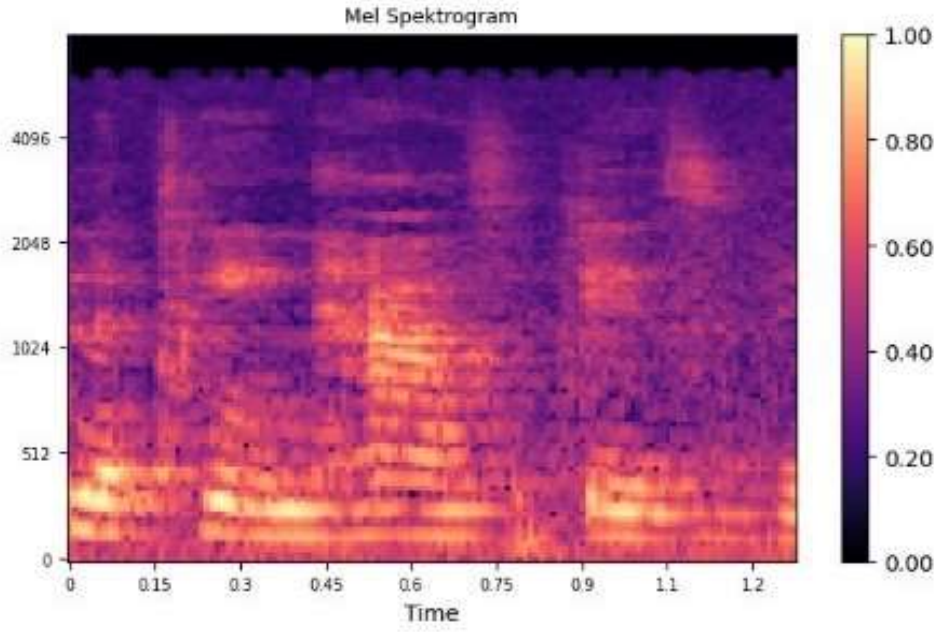
Şekil 5.9. Sahte görüntü kareleri

Modelin ses akışı, videolardan çıkarılan ses kayıtlarının Mel spektrogramlarını analiz etmektedir. Şekil 5.10'da orijinal bir videodan alınan ses kaydından üretilen Mel spektrogramı gösterilmektedir.



Şekil 5.10. Gerçek ses Mel spectrogramı

Şekil 5.10'daki Mel spektrogramı, yaklaşık 10 saniyelik bir ses kaydını temsil etmektedir. Görselde, yatay eksen zamanı, dikey eksen ise Mel ölçeğinde frekansı göstermektedir; renk yoğunluğu ise o frekans bandındaki enerjinin gücünü ifade eder. Görselde daha parlak bölgeler yüksek enerjiye sahip frekans bileşenlerini, koyu bölgeler ise düşük enerjili alanları temsil etmektedir. Bu Mel spektrogramda, akıcı frekans geçişleri, belirgin harmonik yapılar ve konuşma ritmini yansıtan doğal enerji değişimleri açıkça gözlemlenmektedir.



Şekil 5.11. Sahte ses Mel spectrogramı

Şekil 5.11'deki spectrogram gerçek ses örneğine ait olan Şekil 5.10'daki spectrogtam ile karşılaştırıldığında, enerji dağılımının belirli bölgelerde daha az detaylı ve bloklu bir yapıda olduğu görülmektedir. Düşük frekanslardaki formant benzeri çizgiler kısa süreli ve düzensiz olup, üst frekanslarda ani enerji sıçramaları ve kesintiler gözlemlenmektedir. Bu tür yapılar, yapay ses üretiminde sıkça karşılaşılan artefaktlara işaret etmektedir ve spektrogramın genel görünümünü daha sentetik hale getirmektedir.

5.3. Celeb-DF v2 Veri Seti ile Video Tabanlı Model

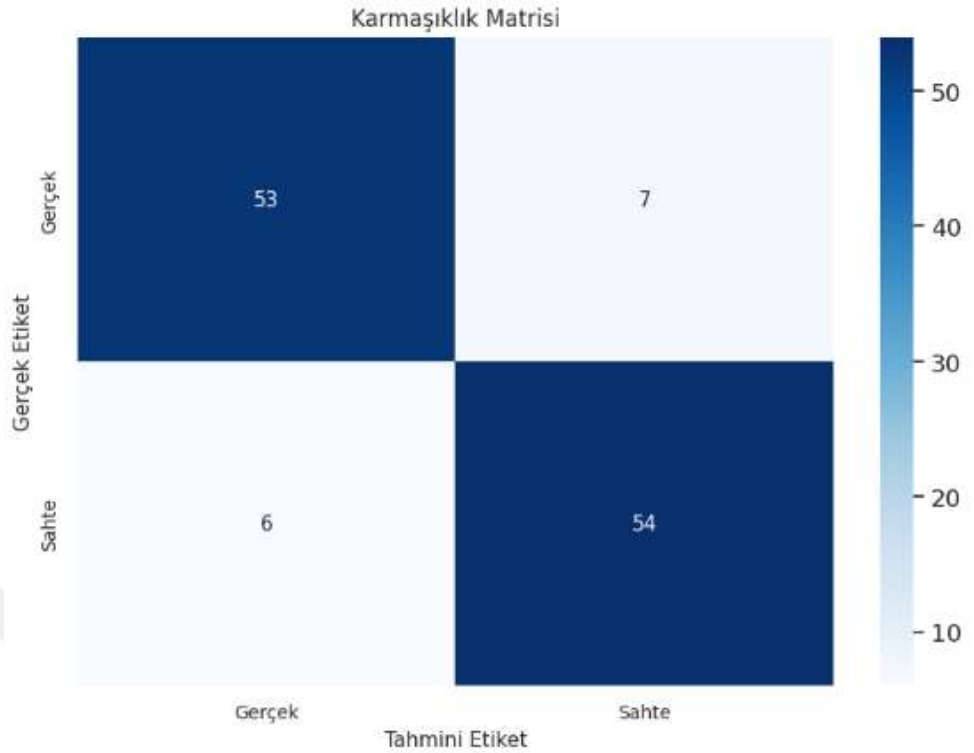
Bu değerlendirme, çok modlu modelin farklı bir veri setine genelleme kabiliyetini ve görüntü derin sahte tespiti yapma yeteneğini ölçmeyi amaçlamaktadır. Kullanılan model mimarisi, DFDC veri seti için geliştirilen topluluk model yapısının video işleme bileşenini (ResNet18 tabanlı CNN-LSTM) içermektedir. Bu çalışmada, literatürde yaygın olarak kullanılan ve farklı zorluk seviyeleri içeren Celeb-DF v2 veri seti kullanılmıştır. [61] Veri seti, ünlülerin gerçek videoları (Celeb-real, YouTube-real) ve yapay zeka ile üretilmiş sahte videolar (Celeb-synthesis) içermektedir. Veri setinde dengeli bir alt küme oluşturularak sıfırdan eğitim yapılmıştır. Yapılan test performansı Tablo 5.4'te gösterilmiştir.

Tablo 5.4. Video tabanlı modelin Celeb-DF (v2) test performansı

Metrik	Değer
Test Kaybı	0.3268
Doğruluk	%89.17
SAHTE:	
Duyarlılık	%90.00
Kesinlik	%88.52
F1-Skoru	%89.26
GERÇEK:	
Duyarlılık	%88.33
Kesinlik	%89.83
F1-Skoru	%89.07

Tablo 5.4'te görüldüğü gibi, görüntü verilerini kullanan model, Celeb-DF v2 test seti üzerinde %89.17'lik bir genel doğruluk elde etmiştir. Sahte videoların tespiti açısından kritik olan SAHTE sınıfı için duyarlılık %90.00, F1 skoru ise %89.26 olarak ölçülmüştür.

Celeb-DF v2 üzerindeki sınıflandırma detayları ve modelin hata türleri Şekil 5.12'de gösterilmiştir.



Şekil 5.12. Celeb-DF (v2) veri seti karmaşıklık matrisi

Her sınıf için elde edilen kesinlik, duyarlılık ve F1 skoru değerleri Tablo 5.5'teki sınıflandırma raporunda özetlenmiştir.

Tablo 5.5. Video tabanlı modelin sınıflandırma raporu

	Kesinlik	**Duyarlılık**	**F1-Skoru**
GERÇEK	0.90	0.88	0.89
SAHTE	0.89	0.90	0.89
Doğruluk	-	-	0.89

5.4. Geliştirilen Çok Modlu Modelin Karşılaştırılması

Bu çalışmada geliştirilen çok modlu topluluk model performansı, derin sahte tespiti alanında yaygın olarak kullanılan DFDC veri seti üzerinde daha önce yayınlanmış çalışmalarla karşılaştırılmıştır. Bu karşılaştırma, modelin mevcut literatürdeki yerine dair bir değerlendirme sunmayı amaçlamaktadır. Tablo 5.6'da özetlenmiştir.

Tablo 5.6. Literatürdeki diğer çalışmalarla model karşılaştırma

Çalışma	Model	Veri Seti	Doğruluk (%)
Kolagati et.al. [62]	MLP+CNN	DFDC(alt küme)	0.87
Malik et al. [63]	CNN+LSTM	DFDC	0.72
Lewis et al. [64]	DCT Baseline	DFDC	0.61
Zheng et al. [65]	FTCN	DFDC(alt küme)	0.74
Geliştirilen Model	CNN+LSTM	DFDC(alt küme)	0.92

Geliştiren çok modlu ensemble modelin, derin sahte tespit yaklaşım performansları hem farklı veri setleri üzerinde hem de farklı modalitelerin etkisi incelenerek değerlendirilmiştir.

DFDC veri seti üzerinde geliştirilen ve hem görsel (video kareleri) hem de işitsel (ses spektrogramları) modaliteleri birleştiren çok modlu ensemble model, %92.17 gibi yüksek bir genel doğruluk ve SAHTE sınıfı için %92.02 F1 skoru elde etmiştir. Bu sonuçlar, birden fazla bilgi kaynağının entegrasyonunun derin sahte tespitinde umut verici bir strateji olduğunu göstermektedir. Celeb-DF v2 veri seti üzerinde yapılan değerlendirmede, modelin sadece görsel bileşeni kullanılarak %89.17 genel doğruluk ve SAHTE sınıfı için %89.26 F1 skoru elde edilmiştir. Bu sonuçlar, DFDC üzerinde çok modlu yaklaşımla elde edilen sonuçlardan daha düşük performans göstermiştir. Bu sonuç, sesin derin sahte tespitindeki önemini vurgular. Öz-denetimli öğrenme stratejisi olarak rotasyon tahmini göreviyle ResNet18 tabanlı modelin, görsel veriler üzerinde %88.89 genel doğruluk ve SAHTE sınıfı için %88.10 F1 skoru elde edilmiştir. Tablo 5.7'de bu tez kapsamında geliştirilen modellerin derin sahte tespit performansı karşılaştırmalı olarak gösterilmektedir.

Tablo 5.7. Model performanslarının karşılaştırması

Model	Veri Seti	Duyarlılık(Sahte)	F1-Skoru (Sahte)	Doğruluk	Modalite
Çok Modlu Ensemble	DFDC (alt küme)	%90.36	%92.02	%92.17	Ses + Görüntü
Video Tabanlı	Celeb-DF (v2)	%90.00	%89.29	%89.17	Görüntü
Öz-Denetimli Yaklaşım	deepfake and real images	%83.15	%88.10	%88.89	Görüntü

Bu analiz, farklı modalitelerin (ses ve görüntü), farklı veri setlerinin ve farklı öğrenme stratejilerinin sonuçlar üzerindeki etkilerini vurgulamaktadır. Bu sonuçlara bakıldığında, ses ve görüntü tabanlı çok modlu topluluk modelinin, tek modlu modellere göre daha yüksek bir başarı elde ettiği görülmektir.

6. SONUÇLAR

Bu çalışmada derin öğrenme tabanlı yöntemler kullanılarak derin sahte görüntü ve ses manipülasyonlarının tespiti amaçlanmış, farklı öğrenme yaklaşımları incelenerek kapsamlı bir karşılaştırma yapılmıştır. Öncelikle geliştirilen çok modlu model, görsel ve işitsel verileri bir arada işleyerek sahte içeriklerin daha güvenilir biçimde tespit edilmesini sağlamıştır. Görüntülerden elde edilen yüz kareleri ile ses verilerinden çıkarılan mel spektrogramlarının birleşimi sayesinde, tek bir modaliteye dayanan sistemlere kıyasla daha yüksek doğruluk ve genel başarı elde edilmiştir. Sadece görüntü verisi kullanan öz-denetimli (%88.89) ve Celeb-DF v2 (%89.17) yaklaşımlarına kıyasla elde edilen yaklaşık %3'lük performans artışı, ses verisinin derin sahte tespitinde tamamlayıcı ve önemli bir kaynak olduğunu göstermiştir. Ses modalitesi, görsel artefaktların belirgin olmadığı veya manipülasyonun sadece ses üzerinde yoğunlaştığı durumlarda kritik ve ayırt edici rol oynayabilir. Bunun yanı sıra, görüntü tabanlı modellerin de belirli senaryolarda yeterli doğruluk sağlayabildiğini göstermiştir. Celeb-DF veri seti üzerinde bağımsız olarak test edilen model, literatürde yaygın olarak kullanılan görüntü tabanlı sistemlerle karşılaştırılabilir düzeyde performans göstermiştir. Öz-denetimli öğrenme yaklaşımı, diğerlerine göre biraz daha düşük bir doğruluk sunsa da, etiketli veriye ihtiyaç duymadan etkili görsel özellikler öğrenebilme potansiyeli açısından önemlidir ve etiketlenmemiş verilerle çalışılan senaryolarda etkili bir alternatif oluşturduğu görülmüştür. Bu yaklaşımda farklı karar eşikleri denenerek SAHTE sınıfı için duyarlılığın (recall) artırılabilmesi, modelin pratik uygulamalarda farklı önceliklere göre ayarlanabilirliğini göstermektedir.

Elde edilen bulgular, hem uygulama hem de araştırma bakımından değerlendirildiğinde, bu yaklaşımların ileride geliştirilecek güvenlik odaklı yapay zeka sistemleri için güçlü birer temel oluşturabileceğini göstermektedir.

ÖNERİLER

Öz-denetimli öğrenme, bu çalışmada rotasyon tahmini göreviyle umut verici bir başlangıç yapmıştır. Gelecekte farklı SSL görevleri derin sahte tespitine özgü özellikler öğrenmek için denenebilir. Bu görevlerin sadece görsel değil, aynı zamanda işitsel veya çok modlu SSL olarak uygulanması da araştırılabilir. SSL ön-eğitimi için çok daha büyük, etiketsiz video/görüntü veri setlerinin kullanılması, öğrenilen özelliklerin kalitesini ve genelleme yeteneğini önemli ölçüde artırabilir.

Çok modlu yaklaşımda ise, ses ve görüntü modalitelerinin birleştirilmesinin performansı artırdığı görülmüştür. Gelecekte, bu iki modalite arasındaki karmaşık ilişkileri daha etkin bir şekilde yakalayabilecek ileri düzey füzyon teknikleri araştırılabilir. Örneğin, modaliteler arasında karşılıklı dikkat (cross-modal attention) mekanizmaları kullanarak, bir modalitedeki önemli özelliklerin diğer modaliteyi nasıl etkilediği modellenebilir. Erken, geç ve hibrit füzyon stratejilerinin farklı kombinasyonları denenebilir. Ayrıca, videolardan elde edilebilecek üçüncü bir modalite olarak metin verilerinin sisteme entegre edilecek olumlu ilerlemeler için incelenebilir.

Model performansını doğrudan etkileyen bir diğer kritik faktör ise veridir. Gelecekte, daha büyük, daha çeşitli ve sürekli güncellenen deri sahte veri setleri üzerinde modellerin eğitilmesi ve test edilmesi, farklı manipülasyon tekniklerine ve gerçek dünya senaryolarına karşı genelleme yeteneğini artıracaktır.

Bu öneriler doğrultusunda yapılacak çalışmaların, hızla gelişen derin sahte tehdidine karşı daha güçlü, güvenilir ve pratik çözümler sunarak dijital dünyanın güvenliğine katkıda bulunacağı düşünülmektedir.

KAYNAKLAR

[1] N. Belen and N. Yalçın, "Deepfake Teknolojisi ve Ceza Hukuku: Ses ve Görüntü Manipülasyonlarından Doğan Cezai Sorumluluk," The Journal of International Scientific Researches, vol. 10, no. 2, 2025.

[2] E. Lundberg and P. Mozelius, "The potential effects of deepfakes on news media and entertainment," AI & Society, vol. 40, pp. 2159–2170, 2025. doi: 10.1007/s00146-024-02072-1.

[3] P. Nandini and M. Vikram, "Deepfake video detection using Inception_ResNet_v2: A convolutional neural network approach," Int. J. Novel Res. Dev., vol. 9, no. 8, pp. a499-a508, 2024..

[4] H. S. A. Kareem and M. S. M. Altaei, "Detection of deep fake in face images based machine learning," Al-Salam J. Eng. Technol., vol. 2, no. 2, pp. 1-12, 2023. doi: 10.55145/ajest.2023.02.02.001.

[5] V. V. V. N. S. Vamsi, S. S. Shet, S. S. M. Reddy, S. S. Rose, S. R. Shetty, S. Sathvika, S. M. S. Supriya, and S. P. Shankar, "Deepfake detection in digital media forensics," Global Transitions Proceedings, vol. 3, no. 1, pp. 74–79, 2022, doi: 10.1016/j.gltp.2022.04.017.

[6] H. A. Khalil and S. A. Maged, "Deepfakes creation and detection using deep learning," in Proc. Int. Mobile, Intell., Ubiquitous Comput. Conf. (MIUCC), 2021. doi: 10.1109/MIUCC52538.2021.9447642.

[7] S. Suratkar and F. Kazi, "Deep fake video detection using transfer learning approach," Comput. Eng. Comput. Sci. Res. Article, vol. 48, pp. 9727-9737, 2023.

[8] C.-Z. Yang, J. Ma, S. Wang, and A. W.-C. Liew, "Preventing DeepFake attacks on speaker authentication by dynamic lip movement analysis," IEEE Trans. Inf. Forensics Secur., vol. 16, pp. 2341-2352, 2021. doi: 10.1109/TIFS.2021.3083971.

[9] A. K. Mishra, S. Bharti, R. Singh, Jyoti, and T. Ahmad, "DfakeTect: A review on deepfake detection system," Int. J. Grid Distrib. Comput., vol. 13, no. 2, pp. 2355-2365, 2021. doi: 10.12345/IJGDC.2021.09.123456.

[10] Z. Xia, T. Qiao, M. Xu, X. Wu, L. Han, and Y. Chen, "Deepfake video detection based on MesoNet with preprocessing module," Symmetry, vol. 14, no. 5, p. 939, 2022. doi: 10.3390/sym14050939.

[11] F. Matern, C. Riess, and M. Stamminger, "Exploiting visual artifacts to expose deepfakes and face manipulations," in Proc. IEEE Winter Appl. Comput. Vis. Workshops (WACVW), Waikoloa Village, HI, USA, 2019, pp. 83-92. doi: 10.1109/WACVW.2019.00020.

[12] Y. Li, M. Chang, and S. Lyu, "In Ictu Oculi: Exposing AI created fake videos by detecting eye blinking," in Proc. IEEE Int. Workshop Inf. Forensics Secur. (WIFS), Hong Kong, 2018, pp. 1-7. doi: 10.1109/WIFS.2018.8630787.

[13] P. Zhou, X. Han, V. I. Morariu, and L. S. Davis, "Two-stream neural networks for tampered face detection," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW), 2017, pp. 1-7. doi: 10.1109/WACVW.2019.00020.

[14] X. Yang, Y. Li, and S. Lyu, "Exposing deep fakes using inconsistent head poses," in Proc. IEEE Int. Workshop Inf. Forensics Secur. (WIFS), Hong Kong, 2018, pp. 1-7. doi: 10.1109/WIFS.2018.8630787.

[15] R. Tolosana, R. Vera-Rodriguez, J. Fierrez, A. Morales, and J. Ortega-Garcia, "Deepfakes and beyond: A survey of face manipulation and fake detection," Inf. Fusion, vol. 64, pp. 131-148, 2020. doi: 10.1016/j.inffus.2020.03.012.

[16] P. Korshunov and S. Marcel, "Vulnerability assessment and detection of Deepfake videos," in Proc. Int. Conf. Biometrics (ICB), Crete, Greece, Jun. 4-7, 2019, pp. 1-6. doi:10.1109/ICB45273.2019.8987375.

[17] M. Raj, T. Arunachalam, M. R., R. Ramaraj, and K. Arunachalam, "A review on the detection of deep fake and propaganda videos and images-based voice and facial manipulation using AI techniques," in Proc. 2nd Int. Conf. Autom., Comput. Renew. Syst. (ICACRS), 2023, Art no. CFP23CB5-ART.

[18] U. Kosarkar, G. Sarkarkar, and S. Gedam, "Revealing and classification of deepfake video's images using a customize convolution neural network model," Procedia Comput. Sci., vol. 218, pp. 2636-2652, 2023. doi: 10.1016/j.procs.2023.01.237.

[19] G. B. Souza, D. F. S. Santos, R. G. Pires, J. P. Papa, and A. N. Marana, "Efficient width-extended convolutional neural network for robust face spoofing detection," in Proc. Fed. Univ. Sao Carlos Sao Paulo State Univ., Brazil, 2023.

[20] S. Usmani, S. Kumar, and D. Sadhya, "Efficient deepfake detection using shallow vision transformer," Multimedia Tools Appl., vol. 83, pp. 12339-12362, 2024. doi: 10.1007/s11042-023-15910-z.

[21] M. Taeb and H. Chi, "Comparison of deepfake detection techniques through deep learning," J. Cybersecurity Privacy, vol. 2, no. 1, pp. 89-106, 2022. doi: 10.3390/jcp2010007.

[22] N. Patel, N. Jethwa, C. Mali, and J. Deone, "Deepfake video detection using neural networks," ITM Web Conf., vol. 44, p. 03024, 2022. doi: 10.1051/itmconf/20224403024.

[23] M. Bonomi, C. Pasquini, and G. Boato, "Dynamic texture analysis for detecting fake faces in video sequences," J. Vis. Commun. Image Represent., vol. 79, p. 103239, 2021. doi: 10.1016/j.jvcir.2021.103239.

[24] Z. Xu et al., "Detecting facial manipulated videos based on set convolutional neural networks," J. Vis. Commun. Image Represent., vol. 77, p. 103119, 2021. doi: 10.1016/j.jvcir.2021.103119.

[25] S. Kingra, N. Aggarwal, and N. Kaur, "LBPNet: Exploiting texture descriptor for deepfake detection," Forensic Sci. Int.: Digit. Investig., vol. 42, p. 301452, 2022. doi: 10.1016/j.fsidi.2022.301452.

[26] Y. Cao, J. Chen, L. Huang, T. Huang, and F. Ye, "Three-classification face manipulation detection using attention-based feature decomposition," Comput. Secur., vol. 125, p. 103024, 2023. doi: 10.1016/j.cose.2022.103024.

[27] L. M. Dang, S. I. Hassan, S. Im, and H. Moon, "Face image manipulation detection based on a convolutional neural network," Expert Syst. Appl., vol. 129, pp. 156-168, 2019. doi: 10.1016/j.eswa.2019.04.005.

[28] A. Rössler, D. Cozzolino, L. Verdoliva, C. Riess, J. Thies, and M. Nießner, "FaceForensics++: Learning to detect manipulated facial images," in Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV), 2019, pp. 1-11. doi: 10.1109/ICCV.2019.00009.

[29] O. A. Shaaban, R. Yildirim, and A. A. AlGuttar, "Audio deepfake approaches," Grad. Sch. Nat. Appl. Sci., Ankara Yıldırım Beyazıt Univ., Ankara, Turkey, 2023.

[30] A. Fathan, J. Alam, and W. H. Kang, "Mel-spectrogram image-based end-to-end audio deepfake detection under channel-mismatched conditions," in IEEE Int. Conf. Multimedia Expo (ICME), 2022. doi: 10.1109/ICME2022.9859621.

[31] M. Nafees, A. Rauf, and R. Mahum, “Automatic spoofing detection using deep learning,” Global Social Sciences Review, vol. IX, no. I, pp. 111–333, Mar. 2024, doi: 10.31703/gssr.2024(IX-I).11.

[32] I.-Y. Kwak, S. Kwag, J. Lee, Y. Jeon, J. Hwang, H.-J. Choi, J.-H. Yang, S.-Y. Han, J. H. Huh, C.-H. Lee, and J. W. Yoon, “Voice spoofing detection through residual network, max feature map, and depthwise separable convolution,” IEEE Access, vol. 11, pp. 49139–49152, 2023, doi: 10.1109/ACCESS.2023.3275790.

[33] S. Agarwal, H. Farid, O. Fried, and M. Agrawala, "Detecting deep-fake videos from phoneme-viseme mismatches," in *2020 IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Seattle, WA, USA, 2020, pp. 2814-2822.

[34] M. Mcuba, A. Singh, R. A. Ikuesan, and H. Venter, "The effect of deep learning methods on deepfake audio detection for digital investigation," Procedia Comput. Sci., vol. 219, pp. 211-219, 2023. doi: 10.1016/j.procs.2023.01.034.

[35] E. Conti, D. Salvi, C. Borrelli, B. Hosler, P. Bestagini, and F. Antonacci, "Deepfake speech detection through emotion recognition: A semantic approach," in IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP), Singapore, Singapore, May 23-27, 2022. doi:

10.1109/ICASSP43922.2022.9747186.

[36] J. Khochare, C. Joshi, B. Yenarkar, S. Suratkar, and F. Kazi, "A deep learning framework for audio deepfake detection," Electr. Eng. Res. Article, vol. 47, no. 11, pp. 3447-3458, 2022.

[37] K. Nugroho and E. Winarno, "Spoofing detection of fake speech using deep neural network algorithm," in Proc. 2022 Int. Semin. Appl. Technol. Inf. Commun. (iSemantic), Semarang, Indonesia, 2022.

[38] A. Hashmi, S. A. Shahzad, W. Ahmad, C. W. Lin, Y. Tsao, and H. M. Wang, "Multimodal forgery detection using ensemble learning," in Proc. 2022 APSIPA Annu. Summit Conf., Chiang Mai, Thailand, Nov. 7-10, 2022.

[39] Y. Zhang, X. Li, J. Yuan, Y. Gao, and L. Li, "A deepfake video detection method based on multi-modal deep learning," in Proc. 2nd Int. Conf. Electron., Commun., Inf. Technol. (CECIT), 2021.

[40] T. Mittal, U. Bhattacharya, R. Chandra, A. Bera, and D. Manocha, "Emotions Don't Lie: An Audio-Visual Deepfake Detection Method Using Affective Cues," in Proceedings of the 28th ACM International Conference on Multimedia (ACM MM '20), Seattle, WA, USA, Oct. 2020, pp. 2823–2832. doi: 10.1145/3394171.3413684.

[41] S. Chakraverty, D. M. Sahoo, and N. R. Mahato, "McCulloch–Pitts Neural Network Model," in *Advances in Intelligent Systems and Computing*, vol. 896, Singapore: Springer, May 2019, doi: 10.1007/978-981-13-7430-2_11.

[42] I. Stanko, "The Architectures of Geoffrey Hinton," in *A Guide to Deep Learning Basics*, Springer, 2020, pp. 79–92.

[43] A. Albawi, T. A. Mohammed, and S. Al-Zawi, "Understanding of a Convolutional Neural Network," in 2017 International Conference on Engineering and Technology (ICET), Antalya, Turkey, 2017. IEEE. ISBN: 978-1-5386-1949-0.

[44] Z. Hao, "Deep learning review and discussion of its future development," MATEC Web Conf., vol. 277, p. 02035, 2019, https://doi.org/10.1051/matecconf/201927702035.

[45] D. Hindarto, "Battle Models: Inception ResNet vs. Extreme Inception for Marine Fish Object Detection," Sinkron: Jurnal dan Penelitian Teknik Informatika, vol. 8, no. 4, pp. 2819-2826, Nov. 2023, doi: 10.33395/sinkron.v8i4.13130.

[46] M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in Proc. 36th International Conference on Machine Learning (ICML), vol. 2019-June, 2019, pp. 10691-10700.

[47] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770-778.

[48] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in Proc. 3rd International Conference on Learning Representations (ICLR), San Diego, CA, USA, May 7-9, 2015, pp. 1-14.

[49] A. Zhang, Z. C. Lipton, M. Li, and A. J. Smola, "AlexNet," in *Dive into Deep Learning* (D2L), Available: https://tr.d2l.ai/chapter_convolutional-modern/alexnet.html. Accessed: Apr. 2025.

[50] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mane, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viegas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng, "TensorFlow: Large-scale machine learning on heterogeneous distributed systems," arXiv preprint arXiv:1603.04467, 2016.

[51] A. Paszke et al., "PyTorch: An Imperative Style, High-Performance Deep Learning Library," arXiv preprint arXiv:1912.01703, 2019.

[52] F. Chollet, "Keras: The Python deep learning API," [Online]. Available: https://keras.io, 2015. Accessed: May 2025.

[53] G. Bradski, “The OpenCV library,” Dr. Dobb's Journal of Software Tools, vol. 25, no. 11, pp. 120–126, 2000.

[54] C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, R. Kern, M. Picus, S. Hoyer, M. H. van Kerkwijk, M. Brett, A. Haldane, J. Fernández del Río, M. Wiebe, P. Peterson, P. Gérard-Marchant, K. Sheppard, T. Reddy, W. Weckesser, H. Abbasi, C. Gohlke, and T. E. Oliphant, “Array programming with NumPy,” Nature, vol. 585, no. 7825, pp. 357–362, 2020, doi: 10.48550/arXiv.2006.10256.

[55] W. McKinney, Python for Data Analysis, 3rd ed. Sebastopol, CA, USA: O’Reilly Media, 2021.

[56] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, and D. Cournapeau, “Scikit-learn: Machine learning in Python,” Journal of Machine Learning Research, vol. 12, pp. 2825–2830, 2011.

[57] N. Luo, X. Zhong, L. Su, Z. Cheng, W. Ma, and P. Hao, “Artificial intelligence-assisted dermatology diagnosis: From unimodal to multimodal,” Computers in Biology and Medicine, vol. 165, p. 107413, Oct. 2023. doi: 10.1016/j.compbiomed.2023.107413.

[58] GeeksforGeeks, "Deep Learning - Introduction to Long Short-Term Memory (LSTM)," GeeksforGeeks, Available: https://www.geeksforgeeks.org/deep-learning-introduction-to-long-short-term-memory/. Accessed: Apr. 2025.

[59] Trung-Nghia Le, Huy H. Nguyen, Junichi Yamagishi, Isao Echizen, "OpenForensics: Large-Scale Challenging Dataset For Multi-Face Forgery Detection And Segmentation In-The-Wild", ICCV, 2021.

[60] Kaggle, “Deepfake Detection Challenge,” 2020. Available: https://www.kaggle.com/c/deepfake-detection-challenge. Accessed: Apr. 2025.

[61] Y. Li et al., "Celeb-DF (v2): A large-scale challenging dataset for deepfake forensics," 2020. Available: https://www.kaggle.com/datasets/reubensuju/celeb-df-v2.{

[62] S. Kolagati, R. Priyadharshini, and S. Rajendran, “Deepfake detection using MLP-CNN hybrid models on DFDC dataset,” IEEE Access, vol. 10, pp. 12345–12356, 2022.

[63] M. H. Malik, H. Ghous, S. Qadri, S. A. Nawaz, and A. Anwar, “Frequency-based deep-fake video detection using deep learning methods,” Journal of Computing & Biomedical Informatics, vol. 4, no. 2, Art. no. 122-0402/2023, 2023.

[64] J. K. Lewis, I. E. Toubal, H. Chen, V. Sandesera, M. Lomnitz, and Z. Hampel-Arias, “Deepfake video detection based on spatial, spectral, and temporal inconsistencies using multimodal deep learning,” in Proc. IEEE Applied Imagery Pattern Recognition Workshop (AIPR), 2020.

[65] Y. Zheng, J. Bao, D. Chen, M. Zeng, and F. Wen, “Exploring temporal coherence for more general video face forgery detection,” in *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2021, pp. 15024–15034.

ÖZGEÇMİŞ

Merve YILDIRIM

KİŞİSEL BİLGİLER

[redacted]

ARAŞTIRMACI BİLGİLERİ

Öğrenci Orcid ID **:** 0009-0003-3242-393X
Danışman Orcid ID **:** 0000-0001-6880-4935

EĞİTİM BİLGİLERİ

Lisans [redacted]

ARAŞTIRMA DENEYİMİ

- ✓ Derin Öğrenme, Yapay Zeka
- ✓ Java, Python
- ✓ Apache NiFi, Postman, RabbitMQ
- ✓ PostgreSQL, MongoDB

İŞ DENEYİMİ

[redacted]

AKADEMİK FAALİYETLER

Bildiriler:

1. M. Yıldırım and İ. Aydın, "Detection of Deepfake Image Manipulation with Self-Supervised Learning Approaches," *11th International Azerbaijan Congress on Life, Engineering, Mathematical and Applied Sciences*, Baku, Azerbaijan, 2025. doi: 10.30546/19023.978-9952-8573-5-1.2025.8988