Bilgisayar Bilimleri ve Mühendisliği Dergisi

Cilt-17
Sayı-2
2024

Güncel Düzmece Video Algılama Yöntemleri Üzerine Bir Kaynak Derlemesi

A Literature Review on Current Deepfake Video Detection Methods

Suzan AYDIN
Haliç Üniversitesi
İşletme Fakültesi
İstanbul, Türkiye
demirsuzann@gmail.com
ORCID: 0009-0005-7337-1851

Zeki ÖZEN
İstanbul Üniversitesi
İktisat Fakültesi
İstanbul, Türkiye
zekiozen@istanbul.edu.tr
ORCID: 0000-0001-9298-3371

Öz

Son yıllarda yapay zekâ ve derin öğrenme teknolojilerindeki hızlı gelişmeler, düzmece (Deepfake) gibi yeni ve yenilikçi uygulamaların ortaya çıkmasını sağlamıştır. Düzmece görsel ve işitsel içeriklerin düzenlenmesine olanak tanır ve özellikle bireylerin görüntü ve seslerini taklit etmek için kullanılır. Düzmece teknolojisi sağladığı olanak ve avantajların yanında kişisel bilginin güvenliği, mahremiyeti ve oluşturulan içeriklerin güvenilirliği gibi konularda ciddi endişelere yol açmaktadır. Bu endişeler, Düzmece içeriklerinin algılanması ve doğrulanması amacıyla yapılan araştırmalara ivme kazandırmıştır. Bu kaynak incelemesi, düzmece türlerini, düzmece video içerikleri algılayan algoritmaların eğitiminde kullanılan veri kümelerini ve güncel düzmece video algılama yöntemlerini ele almaktadır.

Anahtar sözcükler: Düzmece, Düzmece Türleri, Düzmece Görselleri Algılama Yöntemleri, Derin Öğrenme, Yapay Zekâ

Abstract

"In recent years, rapid advancements in artificial intelligence and deep learning technologies have led to the emergence of new and innovative applications such as Deepfake. Deepfake allows for the manipulation of visual and auditory content and is particularly used to imitate individuals' images and voices. Alongside the possibilities and advantages provided by Deepfake technology, it raises serious concerns regarding the security of personal information, privacy, and the reliability of the created content. These concerns have accelerated research aimed at the perception and verification of Deepfake content. This literature review addresses types of Deepfake, datasets used in training algorithms that detect Deepfake video content, and current methods for detecting Deepfake videos.

Keywords: Deepfake, Deepfake Types, Deepfake Video Detection Methods, Deep Learning, Artificial Intelligence

Gönderme, düzeltme ve kabul tarihi: 28.04.2024 - 08.06.2024 – 30.07.2024
Makale türü: Derleme

1. Giriş

2024 itibariyle dünya genelinde 5,35 milyar internet kullanıcısı bulunmaktadır [1]. Statista raporuna göre son bir yılda 97 milyon yeni kullanıcı sayesinde %1,8 oranındaki artış, internet kullanımının küresel ölçekte yayılmaya devam ettiğini göstermektedir. Bu devasa kullanıcı kitlesinin sosyal medya etkileşimleri ve sayısal içerik sunan platformlardaki paylaşımları kişisel verilerin kötü niyetli kullanımı veya sızdırılması gibi riskleri de beraberinde getirmektedir. ITRC (Identity Theft Resource Center) 2023 Veri İhlali Raporu'na göre 2023 yılında sadece Amerika Birleşik Devletleri'nde 353 milyondan fazla internet kullanıcısı veri ihlalinden etkilenmiştir. Yine aynı rapora göre bu oran 2022 yılına kıyasla %16 artmıştır. Veri ihlali olayı sayısındaki bu artış bilişim güvenlik tehditlerinin ne kadar kritik bir duruma ulaştığını göstermektedir [2]. Özellikle veri ihlalleri sonucunda elde edilen kişisel bilgiler, daha gelişkin saldırılar için kullanılabilir. Bu bağlamda, veri güvenliği endişeleri arasına yeni nesil bir tehdit olarak ön plana çıkan "düzmece" teknolojisinin giderek artan kullanımı girmiştir. Düzmece sistemleri, insanların gerçeğinden ayırt edemediği sahte görüntüler oluşturabilir [3]. Veri ihlalleri sonucu sızdırılan kişisel veriler, özellikle ses ve görüntü kayıtları, düzmece içeriklerin oluşturulmasında

kullanılabilir. Bu durum, sadece bireylerin kimliklerinin kötüye kullanılmasına yol açmakla kalmaz, aynı zamanda gerçek olmayan bilginin yayılmasına ve toplumda güven kaybına neden olabilir. Dolayısıyla düzmece içeriklerin üretimi ve tespit edilmesi, teknikleri, yöntemleri ve mekanizmaları hakkında kapsamlı bir anlayış gerektirir [4].

2. Düzmece Tanımı ve Zararları

Düzmece (Deepfake), yapay zekâ ve derin öğrenme algoritmalarını kullanarak ses ve görüntü içeriklerini isteğe göre değiştirme teknolojisidir [5]. Bu teknolojinin özellikle sosyal medya platformlarında video, görüntü ve ses içerikleri üzerinde kullanılması siber zorbalık ve oynama riskini artırmaktadır. Bu teknoloji kullanılarak dolandırıcılar tarafından üretilen sahte videolar, toplumu kolayca aldatabilmekte ve endişeye yol açmaktadır [6, 7]. Bu sebeple düzmece teknolojisiyle oluşturulmuş içerikleri etkili bir şekilde algılayabilen yöntemlere duyulan artan ihtiyaç artmaktadır [8].

Sayısal medya ve platformlarda düzmece teknolojisinin ortaya çıkışı çift taraflı bir kılıç olarak belirmiştir [9]. Eğlence ve iletişimde yenilikler getirmiş olmasına karşın düzmece teknolojisinin kötüye kullanım potansiyeli önemli endişelere yol açmıştır. 2017 yılında 'deepfakes' adlı bir Reddit kullanıcısı tarafından bir ünlünün düzmece yöntemi ile oluşturulan pornografik videosu, bu teknolojinin kötüye kullanılmasının kaçınılmaz olduğunu gösteren ilk örnek olmuştur ve düzmece içerikleri algılama çalışmalarını tetiklemiştir [10]. Gerçek ile değiştirilmiş içerikleri ayırt etmede makine öğrenmesi ve derin öğrenme yöntemleri kullanılmaktadır. Düzmece teknolojileri ve özellikle düzmece video algılama yöntemleri hakkında bir kaynak taraması sunan bu araştırma bu alanda çalışan araştırmacılara ve siber güvenlik önlemlerinin sürekli gelişimine katkıda bulunmayı amaçlamaktadır.

3. Düzmece Türleri ve Veri Setleri

Düzmece içerikler; görüntüler ve video kliplerinin birleştirme, harmanlama, değiştirme ve üst üste bindirme gibi çeşitli teknikler kullanarak oluşturulur [10]. Bu klasik yöntemlerin dışında, ilk kez 2014'te Goodfellow'un tanıttığı Çekişmeli Üretken Ağlar (*Generative Adversarial Networks - GAN*) düzmece üretimini bir başka seviyeye taşımıştır. GAN, rekabet eden iki sinir ağından oluşur: üretici ağ rastgele gürültüden gerçekçi görünen veriler üretmeye çalışırken, ayırt edici ağ gerçek veriler ile üreticinin oluşturduğu sahte verileri ayırt etmeye odaklanır. GAN eğitim sürecinde, üretici daha inandırıcı veriler üretmeyi öğrenirken, ayırt edici de bu verileri gerçeklerden ayırt etme becerisini geliştirir. Bu karşılıklı rekabet, her iki ağın da gelişimini teşvik eder ve üreticinin giderek daha gerçekçi veriler oluşturmasını sağlar [11].

Düzmece içeriklerinin yaygınlığı, karmaşık ve gerçeğe çok yakın sonuçlar üreten birçok gelişmiş GAN tabanlı tekniğin gelişmesine olanak sağlamıştır. AttGAN [12], özellik farkındalığı, hassas ve yüksek kaliteli özellik değişikliklerine olanak sağlar ve yüz özelliklerini değiştirerek yüz değiştirme, yaşlandırma ya da gençleştirme gibi uygulamalarda etkilidir. Benzer şekilde StyleGAN [13], değişik yüz özelliklerinin değiştirilmesine izin vererek gerçeğe çok yakın ve ayrıntılı görüntüler üretmede başarılı olan bir GAN mimarisidir. Aynı şekilde STGAN [14], özellik değişikliklerini doğru bir şekilde kontrol ederek kişinin sadece belirli yüz özelliklerini değiştirebilen etiketli ve etiketsiz verilerle çalışabilen bir araçtır. Başka bir teknik olan StarGANv2 [15], birçok görevi ayrı ayrı eğitmek yerine tek bir büyük modelle tüm bu dönüşümleri sağlayan yeteneğiyle aynı modelle hem saç rengini değiştirebilir hem kişiyi güldürebilir hem de canlandırma karakterini gerçek bir aktörmüş gibi gösterebilir.

Düzmece teknolojisinin görsel, sese ve hatta metin tabanlı uygulamaları bulunmaktadır. Metin tabanlı düzmece örnekleri, sahte veya yanıltıcı içerik oluşturan GAN veya Büyük Dil Modelleri yardımıyla üretilen metinleri ifade eder. Toplumu ve insanları yanıltma amacıyla yayınlanan bu içerikler yanıltıcı bilgilendirmenin bir türüdür [16]. Ses düzmece örnekleri, belirli bir kişinin ses kayıtları üzerinde derin öğrenme modelleri kullanarak ses kopyalama ve ses değiştirme olanağı sağlamaktadır. Düzmece resimler bireylerin görünüşünü değiştirirken, düzmece videolar bireylerin görünüşünü veya eylemlerini değiştiren gerçekçi videolar yaratmak için en sık kullanılan medya formatıdır. Salman ve ark. [4]'e göre düzmece video oluşturulmasında dört ana tip bulunmaktadır. Birinci tip, Gerçek Sesli Yüz Değişimi (Fake Video/Face-Swapped With Real Audio - Type I), bir videodaki A Kişisinin yüzünü B Kişisinin yüzüyle değiştirmeyi ve A Kişisinin sesini korumayı hedefler. İkinci tip, Sentetik Konuşmalı Gerçek Video (Real Video With Synthetic Speech - Type II), videodaki kişinin yüzünü korurken sesini değiştirmeyi amaçlar. Üçüncü tip olan Ses Kopyalanmış Gerçek Video (Real Video With Voice-Cloned - Type III), hedef kişinin gerçek sesini taklit ederek söylemediği bir şeyi söylemiş gibi göstermeyi amaçlar. Dördüncü ve son tip, Sahte Sesli Sahte Video (Fake Audio With Fake Video - Type IV), hem yüzün hem de sesin tamamen oynanmış olduğu durumlardır [4].

Düzmece teknolojisi, farklı medya formatlarında çeşitli uygulamaları kapsadığı için her bir medya formatının benzersiz zorlukları bulunmaktadır [18]. Ancak sosyal medya ve sayısal ortamlarda özellikle video içerikler daha çok tüketildiği için bu kaynak incelemesinde düzmece video algılama çalışmaları değerlendirilecektir.

Tüm ortam formatlarındaki düzmece içeriklerin algılanabilmesini ve gerçek içeriklerden ayırt edilebilmesini sağlayan en önemli etken veri kümelerinin büyüklüğü ve verimliliğidir [17]. Düzmece içerik algılayan modellerin eğitim ve sınama aşamalarında kullanılan açık kaynak güncel veri kümelerinin bazıları şunlardır:

UADFV, YouTube'dan alınmış 49 özgün video ve bunların değiştirilmesiyle oluşturulan 49 video ile toplamda 98 videodan oluşan düzmece tespiti için ilk yayınlanmış veri kümesidir [19].

FaceForensics++ (FF++) veri kümesi, FaceForensics veri kümesinin bir uzantısı olarak araştırmacıların denetimli bir şekilde derin öğrenme yaklaşımları geliştirmelerine olanak tanıyan Face2Face, FaceSwap, Deepfakes ve NeuralTextures gibi dört alt veri kümesi içeren YouTube videoları kullanılarak oluşturulmuştur [20].

DeepfakeTIMIT veri kümesi, 32 öznenin toplam 620 videosunu içerir. Her konu için iki farklı kalitede 20 düzmece videosu bulunmaktadır; bunların 10 tanesi düşük kalite düzmece video içeren Deepfake-TIMIT-LQ alt veri kümesine, kalan 10 tanesi ise yüksek kalite düzmece video içeren Deepfake-TIMIT-HQ alt veri kümelerine aittir [21].

Facebook tarafından oluşturulan DFDC (*Deepfake Detection Challenge)* veri kümesi 100.000'den fazla video içerir ve Kaggle yarışmalarında kullanılmak üzere yayınlanmıştır. DFDC-preview veri kümesinin bir uzantısıdır [22].

Deepfake Detection (DFD) veri kümesi, 363 oyuncu kullanılarak çekilmiş 3068 sahte ve 363 gerçek video içeren düzmece tespit yöntemlerinin geliştirilmesine destek olmak amacıyla Google tarafından yayınlanmıştır [23].

Celeb-DF, YouTube'dan derlenen çeşitli yaş, etnik grup ve cinsiyetlerden kişilere ait gerçek videolar ve iki milyondan fazla kareye ilişkin bir sentez süreci kullanılarak oluşturulmuştur [24].

VoxCeleb2 veri kümesi, YouTube'a yüklenen videolardan çıkarılan, 6.000'den fazla ünlüye ait değişik görsel ve işitsel ortamlarda çekilmiş 150.480 video ve 1 milyondan fazla konuşma içermektedir [25].

FakeAVCeleb, VoxCeleb2 veri kümesinden çıkarılan 500 gerçek videonun çeşitli düzmece üretim yöntemleri ile oynanmış 20.000 düzmece video içeren veri kümesidir [26]. Aşağıda verilen Çizelge-1'de literatürde düzmece video tespit algoritmalarının sıklıkla kullandığı bazı veri kümeleri özet olarak verilmiştir [27, 28].

Çizelge-1: Düzmece Video Tespit Çalışmalarında Kullanılan Bazı Veri Setleri

Veri Kümesi	Yıl	İçerik	Toplam	Yöntem	İçerik Türü ve Sayısı	Kaynak
UADFV [19]	2018	Video	98	FakeApp	49 Sahte 49 Gerçek	YouTube
FaceForensics++ (FF++) [20]	2019	Video	4,000	Düzmece Bilgisayar Grafikleri	4,000 Sahte 1,000 Gerçek	YouTube
DeepfakeTIMIT [21]	2018	Video	640	GAN Faceswap	320 Düşük Kalite Sahte 320 Yüksek Kalite Sahte	32 Oyuncu
DFDC (Deepfake Detection Challenge) [22]	2020	Video	120,000	Bilinmiyor	100,000 Sahte 20,000 Gerçek	1,311 Oyuncu
DFDC- preview [22]	2020	Video	5,250	Bilinmiyor	4119 Sahte 1131 Gerçek	1,311 Oyuncu
Celeb-DF [24]	2020	Video	1,203	Düzmece	795 Sahte 408 Gerçek	YouTube
Deepfake Detection (DFD) [23]	2019	Video	3,431	Faceswap	3068 Sahte 363 Gerçek	363 Oyuncu
VoxCeleb2 [25]	2018	Video Ses	150,480 1,128.246	-	Gerçek	YouTube
FakeAVCeleb [26]	2021	Video	20,000	Faceswap FSGAN Wav2Lip	20,000 Sahte	VoxCeleb2
TIMIT-TTS [27]	2022	Ses	20,000	Dynamic Time Warping (DTW)	20,000 Sahte	VidTIMIT
DeeperForensics (DF) [27]	2020	Video	60,000	DF-VAE	10,000 Sahte 50,000 Gerçek	100 Oyuncu
WildDeepfakes (WDF) [27]	2020	Video	7,314	Bilinmiyor	3,509 Sahte 3,805 Gerçek	Bilinmiyor
Presidential Deepfakes Dataset [28]	2020	Video	32	Düzmece	16 Sahte 16 Gerçek	YouTube
World Politicians Deepfake Dataset (WPDD) [28]	2020	Video	135,251	Düzmece Faceswap	31,016 Sahte 104,235 Gerçek	YouTube

4. Düzmece Video Algılama Yöntemleri

Rana ve ark. [29], düzmece algılama çalışmalarına yönelik yaptıkları kapsamlı literatür taramasında düzmece algılama çalışmalarını Derin Öğrenme (*Deep Learning - DL*), Makine Öğrenimi (*Machine Learning - ML*), istatistiksel teknikler ve öbek zinciri (blockchain) tabanlı yöntemler olarak dört kategori altında değerlendirmiştir. Bu dört kategoriye insan

odaklı düzmece algılama çalışmaları [30, 31, 32] da dahil edilerek düzmece algılama yöntemleri beş ulam olarak sınıflandırılacaktır.

Ancak düzmece algılama çalışmalarında baskın ulam olan derin öğrenme tabanlı yöntemler, bu çalışmanın özellikle yöneldiği alan olmuştur. Derin öğrenme, makine öğrenmesinin bir alt dalı olarak verilerden özellik çıkarma ve dönüştürme işlemlerini gerçekleştirmek için çok katmanlı doğrusal olmayan işlem birimlerini kullanır. Bu katmanlar, birbirini takip eden şekilde düzenlenmiş olup, her bir katman önceki katmanın çıktısını girdi olarak alır [33]. Düzmece içerikleri algılamak için kullanılan üç derin öğrenme yaklaşımı dikkat çekmektedir: Evrişimli Sinir Ağı (*Convolutional Neural Networks - CNN*), Tekrarlayan Sinir Ağları (*Recurrent Neural Networks -RNN*) ve Transformatör (*Transformer*) modelleri çalışmalarda yoğunlukla yer almakta ve yüksek performans sunmaktadır. Çizelge-2 çoğunlukla kullanılmakta olan CNN mimarilerini içermektedir.

Düzmece içerik algılama yöntemleri arasında uzamsal kalıntı (*spatial artefacts*) analizleri, biyolojik ve fizyolojik göstergelerin değerlendirilmesi, ses ve görüntü verileri arasındaki uyumsuzlukların incelenmesi, evrişimsel izlerin saptanması, kimlik doğrulama bilgilerinin analizi, zaman içindeki tutarsızlıkların izlenmesi, yüz ifadelerinin detaylı incelenmesi ve uzamsal-zamansal özelliklerin değerlendirilmesi yer almaktadır [34]. Çalışmanın devamında burada sayılan düzmece video algılama yöntemleri tek tek ele alınmış ve her bölümün sonunda ilgili yöntemle yapılan çalışmaları özetleyen tablolar verilmiştir. Bu çizelgelerde düzmece algılama çalışmalarında en çok kullanılan AUC (*Area Under Curve - Eğri Altı Alan*) ve ACC (*Accuracy – Doğruluk*) metriklerinden elde edilen en yüksek performans değeri dikkate alınmıştır.

Çizelge-2: CNN Mimarileri

Model	Yıl	Özellikler	Avantajlar	Dezavantajlar	Kullanım Alanları
AlexNet [35]	2012	Basit ve hızlı Katman: 8	Görüntü sınıflandırmada başarılı	Daha az doğruluk	Görüntü sınıflandırma, nesne algılama
VGGNet [36]	2014	Daha derin ve daha doğru Katman: 16/19	AlexNet [35]'e göre daha yüksek doğruluk	Daha fazla parametre ve işlem gücü	Görüntü sınıflandırma, nesne algılama
Inception [37]	2014	Daha az parametre ile yüksek doğruluk Katman: 22	Farklı boyutlarda filtreler kullanır	Hesaplama açısından daha karmaşık	Görüntü sınıflandırma, nesne algılama, görüntüyü metne çevirme
ResNet [38]	2015	En yüksek doğruluk Katman: 0/101/152	Kalan bağlantıları kullanır	Daha fazla parametre	Görüntü sınıflandırma, nesne algılama, insan pozlama tahmini
DenseNet [39]	2016	Yoğun bağlantı modeli	Daha az parametre ile yüksek doğruluk	Eğitmek daha zor	Görüntü sınıflandırma, nesne algılama, görüntü segmentasyonu
MobileNet [40]	2017	Mobil cihazlar için optimize edilmiş	Daha az parametre ve işlem gücü	Daha az doğruluk	Görüntü sınıflandırma, nesne algılama, yüz tanıma
EfficientNet [41]	2019	Otomatik model arama ve seçme	En yüksek doğruluk ve verimlilik dengesini sunar	Karmaşık mimari	Görüntü sınıflandırma, nesne algılama, görüntü segmentasyonu

4.1. Uzamsal Kalıntılara Dayalı Tespit Çalışmaları

Düzmece görüntü üretilince resim veya videoda düzenlemelerin gerçekleşmesiyle sıklıkla uzamsal kalıntılar oluşur. Görüntüdeki sıkıştırma izleri, blokların bozulmaları, tekrar eden örüntü desenleri veya benekleşme gibi görsel bozulmaların varlığı videonun düzmece olduğunun anlaşılmasına dair ipuçları verir.

Afchar ve ark. [42] görüntülerdeki uzamsal detaylara ve ince farklılıklara odaklanarak düzmece algılama için mezoskopik özellikleri inceleyen MesoInception-4 adlı bir CNN modeli önerdiler. Bu model Inception [37] modüllerinin bir varyasyonunu kullanarak %98,4 doğruluk oranı ile ön plana çıkmıştır. Bu çalışma, düzmece algılamada derin ağların etkinliğini göstererek bu alanın genişlemesine katkıda bulunmuştur.

Bir yüz görüntüsü değiştirdikten sonra kafaya yerleştirildiğinde yüz ve baş açısının tutarsızlığına dikkat çeken Yang ve ark. [19] yüzdeki iki boyutta (2D) işaretlerden başın üç boyutlu (3D) pozisyonunu tahmin ederek sahte içeriği algılamayı hedeflemişlerdir. Çalışmada gerçek içerikle sahte içeriği ayırt etmek için Destek Vektör Makineleri (*Support Vector Machine - SVM*) kullanılmıştır. Bu teknik düzmece algılamada %89 AUC skoru verse de bulanık görüntülerde yüz işaretlerini tahmin etmekte zorlanmıştır.

H. Zhao ve ark. [43] düzmece algılamaya bir sınıflandırma problemi olarak yaklaşarak çoklu dikkat ağı (*Multiattentional Network*) önerdiler. Bu yöntem, FF++ veri kümesinde %97,60 doğruluk sağlamış olsa da yüksek sıkıştırmaya sahip görüntülerde uzamsal alandaki faydalı bilgilerin çoğu bulanıklaştığı için düşük kaliteli görüntülerde iyi sonuçlar alınamamıştır. Aynı yıl Kohli ve Gupta [44], düzmece videolarını analiz etmek için frekans odaklı bir CNN

(*Frequency Convolutional Neural Network - FCNN*) kullanmıştır ve FF++ veri kümesinde %85,24 doğruluk oranı elde etmişlerdir. Luo ve ark. [45] ise mekânsal dikkat ve yüksek-düşük frekans etkileşimlerine odaklanan bir model kullanarak yaptıkları çalışmada %99,5 AUC skoru elde etmişlerdir. İsmail ve ark. [46] ise çeşitli modelleri birleştiren YOLO (You Only Look Once)-InceptionResNetV2-XGBoost (YIX) mimarisini geliştirerek düzmece algılamada %90,73 doğruluk elde etmişlerdir.

Das ve Sebastian [47] düzmece videoları algılamak için video karelerine yönelik bir yöntem geliştirdiler. Bu yöntemde, videodaki yüz algılanıp kırpıldıktan sonra her bir video karesinin bir makine öğrenimi algoritmasıyla özellik çıkarımı ve sınıflandırması yapılmaktadır. Her karenin görüntü özellikleri üç farklı CNN modeli kullanılarak elde edilerek Temel Bileşen Analizi (*Principal Component Analysis - PCA)* ile seçilen ve boyutları küçültülen özellik vektörlerinde birleştirilmektedir. Ardından bir SVM, her kareyi gerçek veya sahte olarak sınıflandırmaktadır. Bu yöntem DFDC alt veri kümesinde %96,50 doğruluk elde etmiştir.

Güncel bir çalışma olarak Dhanaraj ve Sridevi [48] transfer öğrenimi kullanarak videolardaki yüz bozulma bölgelerini tespit eden bir yöntem geliştirdiler. ImageNet [35] veri kümesi ve bir CNN sınıflandırıcı kullanılarak geliştirilen önceden eğitilmiş bir model, bilgisini bu amaca özel yeni bir sınıflandırıcıya aktarmaktadır. İşlenmiş videoya uygulanan Xception [49] CNN sınıflandırıcısı yüz bozulma bölgelerini verimli bir şekilde tespit etmektedir. Önerilen model, bozulmuş yüz bölgelerini %89,25 doğrulukla tespit edebilmektedir.

Çizelge-3'te uzamsal kalıntılara dayalı düzmece algılama çalışmalarının özeti bulunmaktadır.

Çizelge-3: Uzamsal Kalıntılara Dayalı Düzmece Video Tespit Çalışmaları

Çalışma	Algoritma/ Önerilen Model	Veri Setleri	Başarım Oranı
[42]	CNN	VidTIMIT, FF++	%98,4 ACC
[19]	SVM	MFC, UADFV	%89 AUC
[43]	Multiattentional Framework	FF++, CelebDF, DFDC	%97,60 ACC
[44]	fCNN	FF++, Celeb-DF(v2)	%85,24 ACC
[45]	CNN	DeeperForensics, FF++, DFD, DFDC, Celeb-DF	%99,5 AUC
[46]	CNN	CelebDF, FF++, WiderFace	%90,73 ACC
[47]	CNN-SVM-PCA	DFDC	%96,50 ACC
[48]	CNN	Celeb-DF, Youtube videoları, ImageNet	%89,25 ACC

4.2. Biyolojik/Fizyolojik İşaretlere Dayalı Tespit Çalışmaları

Matern ve ark. [50] düzmece videolardaki göz rengi, yansıma eksiklikleri ve göz ile dişlerdeki bulanıklıklar gibi belirgin hataları incelediler ve bu biyolojik farklılıkların düzmece tespitinde yararlı olabileceğini belirtmişlerdir.

Göz kırpma, göz yüzeyini nemlendirmek ve tozlardan arındırmak için gerekli bir fizyolojik aktivitedir. GAN ile çalışan modellerin eğitim veri kümesi çoğunlukla açık gözlü insan görüntüleri içerdiği için bu modellerin ürettiği sahte yüzler, göz kırpma simülasyonunda genellikle başarısız olmaktadır [51]. Bu nedenle bu modellerin kapalı göz durumunu etkin bir şekilde taklit edemediğini belirten Li ve ark. [51], göz kırpma sayısını ölçmek için LRCN (*Uzun Dönemli Yinelemeli CNN - Long-term Recurrent Convolutional Networks*) modeli kullanmayı önerdiler. Önce gerçek videolarda göz kırpma oranlarını inceleyerek göz kırpmanın fark edilmesi gereken ortalama süreyi hesaplamışlardır. Ardından, düzmece videolardaki göz kırpma süreleri incelenmiş ve videolarda göz kırpma oranının, sentezlenmiş veya sahte bir videoyu tespit edebildiği sonucuna varılmıştır. Benzer bir yaklaşımda Jung ve ark. [52] göz kırpma düzenlerinin zamansal dinamiklerini analiz etmek için Uzun Kısa Süreli Bellek (*Long Short-Term Memory – LSTM*) ağlarından yararlanmışlardır. Bu yöntem göz kırpma periyodu, göz kırpma tekrarlama sayısı ve göz kırpma süresini değerlendirerek çoğu düzmece video türünü etkili bir şekilde tespit edebilmiştir. Ancak GAN tabanlı düzmece modelleri, kapalı göz resimleriyle eğitildiklerinde göz kırpma temelli düzmece algılama sistemlerini kandırabilecek gerçekçi göz kırpma hareketleri üretebilmektedir [53].

Kare kare üretilmiş düzmece videolar (özellikle her karenin tek başına manipüle edildiği türler) doğal kalp ritmi değişiminden gelen ipuçları içermemektedir [54]. Qi ark. [55] tarafından geliştirilen DeepRhythm, uzaktan görsel fotopletismografi (*PPG*) tekniği ile yüz derisindeki minik ve ritmik renk değişimlerini analiz ederek kan akışını ve dolayısıyla kalp atışını ölçmektedir. Gerçek yüzlerde ölçülebilen doğal kalp atış ritminin düzmece videolarda eksik olacağını varsayarak kalp atışı sinyalindeki bozulmaları hem uzamsal hem de zamansal olarak analiz etmişlerdir. Çalışma, yüksek doğruluk oranı ve JPEG bozulması gibi zorluklara karşı dayanıklılık iddialarıyla öne çıkmakta ve kalp atış ritmi bazlı ilk düzmece algılama yöntemi olma özelliğini taşımaktadır. Çiftci ve ark. [56] çalışmasında önerdikleri FakeCatcher yöntemi yüz bölümünden altı farklı biyolojik sinyali mekânsal ve zamansal olarak izleyerek, bu sinyaller arasındaki tutarlılığı gerçeklik belirtisi olarak değerlendirmiştir. Yine fizyolojik sinyallere odaklandıkları bir başka çalışmada Çiftci ve ark. [57] Düzmece algılama yöntemine ek olarak analiz edilen düzmece içeriğin hangi GAN modeliyle üretildiği sorusuna da cevap aramış ve bu amaçla PPG (*fotopletismografi*) temelli kalp atışı ölçümü yöntemini kullanmışlardır.

Hernandez-Ortega ve ark. [58] DeepfakesON-Phys uzaktan fotopletismografi (*rPPG*) kullanarak video dizilerindeki ince cilt rengi değişiklikleri aracılığıyla kalp atış hızı bilgilerini analiz

ettikleri çalışmada Celeb-DF ve DFDC veri kümelerinde %98'in üzerinde AUC oranı elde etmişlerdir.

Biyolojik sinyallerdeki farklılıklara odaklanan Wang ve ark. [59] Siamese ağ çerçevesi içinde Xception [49] ağını kullanarak UADFV veri kümesinde %99,94 oranında yüksek bir doğruluk elde etmişler ve oylama mekanizmaları aracılığıyla bu yöntemi daha da iyileştirmişlerdir.

Khurana ve ark. [60], GAN modellerinin her kişiye özel olan kalp atış hızını taklit edemeyeceğine dikkat çekerek düzmece algılamada kalp atış hızı analiz yöntemini kullanmışlardır. Önerilen metodoloji PPG hücreleri veya uzamsal-zamansal hücreler şeklinde bir biyolojik sinyal çıkarıcıyı farklı makine öğrenimi modelleri aracılığıyla işlemeyi içermektedir. MBConv Bloklarını kullanıldığı çalışmada yazarlar ResNet-18 üzerinde %95'e varan bir doğruluk artışı gözlemlemişlerdir.

He ve ark. [54], GazeForensics yöntemi ile özellikle kare kare manipülasyon teknikleriyle yapılmış olan videolardaki göz hareketlerinin doğal akıcılığındaki tutarsızlıkları (uzamsal tutarsızlık) ve göz bölgesindeki biyometrik özelliklerindeki (iris rengi, ışık yansıması, göz şekli vb.) tutarsızlıkları görüntünün manipüle edildiğine dair işaret kabul etmiş ve yaptıkları çalışmada %99,64 doğruluk elde etmişlerdir.

Son olarak Liang ve ark. [61], düzmece algılama için yüzün değiştirilmiş ve değiştirilmemiş bölgelerinin özelliklerini ayırt etmek için yüz haritalarını ve frekans alanındaki korelasyonları analiz ettiler. Bu yöntemde göz çevresindeki farklılıkların geometrik özellikleri kullanılmıştır. Çalışma, FF++ veri kümesinde %99,6 doğruluk yakalamıştır.

Çizelge-4'te biyolojik ve fizyolojik işaretlere dayalı düzmece video algılama çalışmalarının özeti verilmiştir.

Çizelge-4: Biyolojik/Fizyolojik İşaretlere Dayalı Düzmece Video Tespit Çalışmaları

Çalışma	Önerilen Model	Veri Setleri	Başarım Oranı
[51]	LRCN	CEW Dataset, Özel Veri Kümesi	%99,00 ACC
[52]	CNN-LTSM	Özel Veri Kümesi	%87,5 ACC
[55]	CNN	FF++, DFDC	%100 ACC
[56]	CNN	FF++, Celeb-DF, FF, FakeCatcher	%96 ACC
[57]	CNN	FF++, Celeb-DF	%94,66 ACC
[58]	Convolutional Attention Network	Celeb-DF, DFDC-Preview	%98,7 AUC
[59]	CNN	UADFV, Celeb-DF (v2), FF++	%99,94 ACC
[60]	CNN-RNN	Celeb-DF	%95 ACC
[54]	CNN	FF++, Celeb-DF, WDF	%99,64 ACC
[61]	CNN-LTSM	CelebA, FF++, Celeb-DF, DFD	%99,6 ACC

4.3. Ses-Görsel Tutarsızlıklarına Dayalı Tespit Çalışmaları

Bu çok yönlü yaklaşım, yüz veya ses manipülasyonu gerçekleşmiş düzmece videolarda özellikle etkili bir metodoloji sunmaktadır [23].

Zhou ve Lim [62] konuşma sırasındaki hece ve dudak hareketleri arasında güçlü bir ilişkinin var olduğunu gözlemlediler. Bu ilişki, düzmece videolarda sıklıkla bozulmaktadır. Dil bağımsız bu modeli FF++ ve DFDC veri kümelerinde test ederek %81,96 doğruluk elde ettiler. Model, video ve ses akışlarındaki düşük ve yüksek seviye özellikler arasındaki ilişkiyi öğrenerek ses ve görsel bölgeler arasındaki bağlantıyı kurmaktadır.

Cai ve ark. [63] geliştirdiği Sınır Algılamalı Zamansal Sahtecilik Tespiti (*Boundary Aware Temporal Forgery Detection*) tekniği, iki boyutlu evrişimsel sinir ağı (*2DCNN*) ile ses verilerinden bilgi çıkartarak çerçeve seviyesindeki uzamsal-zamansal bilgileri bir dizi kare olarak öğrenmek için de üç boyutlu evrişimsel sinir ağını (*3DCNN*) kullanmaktadır. Bu tekniğin tespit doğruluk oranı en yüksek %99 bulunmuştur.

Ilyas ve ark. [64] tarafından sunulan AVFakeNet tekniğinde özellik çıkarmak için Transformer modeli olan Yoğun Swin Dönüştürücü Ağı (*Dense Swin Transformer Net - DST-Net*) kullanılmıştır. Model çeşitli veriş etlerinde test edilmiş %93,40 ACC performans değeri elde edilmiştir.

Anas Raza ve Mahmood Malik [65] tarafından geliştirilen Multimodaltrace tekniği, ses ve görüntü düzmece içeriği algılayan için yeni bir metodolojidir. Modelde, IntrAmodality Mixer katmanları aynı türdeki veri kanallarını (ses veya görüntü gibi) işlerken, IntErModality Mixer katmanları ise farklı türdeki veri kanallarını (hem ses hem de görüntü) birleştirir. Böylece videodaki ses uyumsuzluğu incelenebilmektedir. FakeAVCeleb veri kümesi ve diğerleri üzerinde test edilen Multimodaltrace tekniğinin elde ettiği %92,9 doğruluk oranının mevcut yöntemlerden daha iyi olduğu ifade edilmiştir.

Hashmi ve ark. [66] ses ve görüntüyü birlikte analiz eden Transformer tabanlı bir ensemble modeli (toplu öğrenme) önermişlerdir. Mevcut yöntemlerin aksine bu yöntem hem ses hem de görsel manipülasyonları tespit edebilmektedir. Transformer modelinin güçlü modelleme, paralel işleme ve dikkat mekanizması sayesinde videodaki uzun vadeli bağımlılıklar ve küresel bilgi yakalanabilmektedir. Ayrıca ensemble öğrenme ile birden fazla modelin tahminleri birleştirilerek daha sağlam bir sonuç elde edilmektedir. FakeAVCeleb veri kümesi üzerinde yapılan analizlerde bu yöntemin mevcut tüm yöntemlere göre performansının daha yüksek olduğunu belirtmişlerdir.

Aşağıda verilen Çizelge-5'te ses ve görsel tutarsızlıklara dayalı düzmece video algılama çalışmalarının özeti verilmiştir.

Çizelge-5: Ses-Görsel Tutarsızlıklarına Dayalı Düzmece Video Tespit Çalışmaları

Çalışma	Önerilen Model	Veri Setleri	Başarım Oranı
[62]	ResCNN+GRU	FF++, DFDC	%81,96 ACC
[63]	3DCNN	DFDC, Celeb-DF	%99 ACC
[64]	AVFakeNet	FakeAVCeleb, Celeb-DF, ASVSpoof2019 LA, WLD, PDD	%93,40 ACC
[65]	Multimodaltrace	FakeAVCeleb, PDD, WLD	%92,9 ACC
[66]	AVTENetf	FakeAVCeleb	%99 ACC

4.4. Evrişimsel İzlerin Algılanmasına Dayalı Çalışmalar

Düzmece üretiminden kaynaklanan bir kalıntı türü olan ve insan gözü ile fark edilmesi zor olan Evrişimsel izler (*convolutional traces*) aynı zamanda düzmece algılama için kullanılmaktadır. Bu yöntemin temel prensibi düzmece oluşturma işleminin görüntünün frekans spektrumunda belirli izler bırakmasıdır.

Li ve Lyu [67], çalışmasında düzmece video yapılırken yeniden boyutlandırma ve değiştirme işlemlerinin yarattığı yapay izlerin sebep olduğu yüzdeki bozulma kalıntılarını dikkate almaktadır. MesoNet [42]'i temel alan yöntemin tanıdık olmayan kaynaklardan gelen videolarda daha az etkili olduğu belirtilmiştir.

Frank ve ark. [68], farklı GAN mimarileri ve veri kümeleri üzerinde frekans alanında ortaya çıkan kalıntıları kapsamlı bir şekilde incelediler. GAN modellerinin örnekleme yükseltme teknikleri nedeniyle ciddi kalıntıların ortaya çıktığını gözlemlemişlerdir. Yapılan analizler, basit bir doğrusal model ve bir CNN tabanlı model içeren bir sınıflandırıcının, tüm frekans spektrumu üzerinde yüksek sonuçlar elde edebileceğini göstermektedir. Younus ve Hasan [69], Düzmece videolarını algılamak için Haar Dalga Dönüşümü odaklı bir yöntem geliştirdiler. Bu yöntem, videodaki yüzlerin boyut ve çözünürlüğündeki tutarsızlıkları ve farklılıkları analiz ederek sahte yüzlerin videolara eklenirken oluşturduğu belirgin bulanıklıkları tespit etmektedir. Sahte yüz sonradan yerleştirildiği için arka planla (yani gerçek vücut) arasında belli noktalarda bir bulanıklık oluşur. Haar Dalga Dönüşümü hem bu bulanıklığı ayırt edebilmektedir hem de dönüştürülen görüntünün "kenarlarını" daha iyi belirginleştirerek örneğin yüz çenesindeki geçişin keskinliğinin orijinale göre fark edilmesini sağlamaktadır. Yazarlar UADFV veri kümesi %90,5 oranında tespit başarısı elde etmişlerdir. Huang ve ark. [70], GAN üretimi sırasında video karesinin değiştirilmiş alanlarında renk, doku gibi özelliklerde anormallikleri tespit etmek için FakeLocator adında sahte alanları doğru bir şekilde bulan özel bir ağ ve gri tonlamalı bir harita kullanmıştır.

Görüntülerde yapılan değişiklikler gradyan bilgisinde (renklerin geçişlerinde, bir anlamda doku üzerindeki dalgalanma) içeriğin düzmece olduğunu belli edecek özellikler bırakır. Gri-Tonlama işlemi kullanılarak bu gradyan bilgisinin "görünürlüğü" artırılır [71]. Gri tonlamalı görüntülerin içinde gizlenmiş gradyan verilerini kullanan Xiao ve ark. [71] geliştirdikleri yöntemle DeeperForensics veri kümesinde %98,17 oranında doğruluk elde etmişlerdir.

Lin ve Sun [72], az sayıda eğitim örneğiyle görüntülerdeki sahte alanları tespit edebilen GAN tabanlı bir yöntem geliştirdiler. Bu yöntem, gözetimsiz ve yarı gözetimli öğrenme kullanarak sahte yüzleri belirleyerek %93 doğrulukla sınıflandırma yapmıştır.

Çizelge-6'da evrişimsel izlere dayalı düzmece video algılama çalışmalarının özeti yer almaktadır.

Çizelge-6: Evrişimsel İzlerin Algılanmasına Dayalı Düzmece Video Tespit Çalışmaları

Çalışma	Önerilen Model	Veri Setleri	Başarım Oranı
[67]	CNN	UADFV, Deepfake-TIMIT	%99,9 AUC
[68]	CNN	Özel Veri Kümesi	%99,91 ACC
[69]	GAN	UADFV	%90,5 ACC
[70]	GAN	FF++, DFFD	%99,95 ACC
[71]	GAN	DeeperForensics	%98,17 ACC
[72]	GAN	FF++	%93 ACC

4.5. Kimlik Bilgilerine Dayalı Çalışmalar

Agarwal ve ark. [73] yaptığı çalışma, düzmece videoları algılamak için statik yüz tanıma, yüz ifadelerinde gözlemlenen zamansal davranışlar ve baş hareketleri gibi çeşitli biyometrik özellikleri kullanmaktadır. Bu özelliklerin bütünleşik bir şekilde analiz edilmesi düzmece içeriklerin tespitinde kritik bir rol oynamaktadır. Çalışmada kullanılan CNN metrik-öğrenme amaç fonksiyonu aracılığıyla bu davranışsal özelliklerin entegrasyonunu öğrenmektedir.

Bu araştırmanın devamı niteliğindeki başka bir çalışmada Agarwal ve ark. [74] özellikle dudak senkronizasyonuna dayalı düzmece algılamaya odaklanmışlardır. Bu teknik, ağız şeklinin dinamiklerinin bazen söylenen seslerle (fonem) uyumsuz olabileceği gerçeğini göz önünde bulundurarak bu tür uyumsuzlukları tespit etmeye yöneliktir.

Cozzolino ve ark. [75], gerçek videolardan öğrenerek düzmece videoları algılayan bir yöntem geliştirdiler. Yüzdeki değişimleri ve hareketleri inceleyerek düzmece görüntüleri algılayan bu yüz okuma yöntemi, düşük kaliteli videolarda bile %81,8 doğruluk sağlayarak çalışmada kıyaslanan diğer yöntemlerden daha başarılı olmuştur.

Dong ve ark. [76], düzmece algılama için yüzdeki kimlik bilgilerini kullanan ve farklı veri kümelerindeki zorluklara rağmen kimlik tutarlılığını vurgulayan Kimlik Tutarlılığı Dönüştürücüsü (*Identity Consistency Transformer*) adında yeni bir model sundular. Bu çalışma benzerlik ölçüleri için bir eşik belirleme zorluğuna ve genellemeyi etkileyebilecek farklı veri kümelerinde değişen benzerlik dağılımları sorununa dikkat çekmiştir.

Shen ve ark. [77], gerçek ve sahte kimlikler arasında ayrım yapma becerisini geliştirmek için Dong ve ark. [76] yaptığı çalışmadan esinlenerek oynanmış ve edilmemiş yüzlerin farklarını öğrenmeye dayalı bir yöntem önerdiler. CNN temelli algılama yönteminin görüntünün merkezine odaklanmasına çözüm olarak giriş görüntülerinde rastgele maskeler kullandılar. Bu yaklaşım, zıt öğrenme görevlerinin aksine aynı kişiden kimliğe göre pozitif ve negatif örnekler almaktadır. Shen ve ark. [77] yaptıkları çalışma Celeb-DF veri kümesinde %91,76 doğruluğa ulaşmıştır.

Liu ve ark. [78], bir video içeriğinde yer alan bir kişinin yüzünde farklı kareler arasındaki farklılıkları tespit eden bir model geliştirdiler. Bu model, tüm karelerden elde ettiği kimlik bilgilerini kimlik vektörlerine dönüştürmekte ve ardından vektörlerden zamansal özellikler öğrenerek tutarsızlıkları tespit etmektedir.

Çizelge-7'de kimlik bilgilerine dayalı düzmece video algılama çalışmalarının özetine yer verilmiştir.

Çizelge-7: Kimlik Bilgilerine Dayalı Düzmece Video Tespit Çalışmaları

Çalışma	Önerilen Model	Veri Setleri	Başarım Oranı
[73]	CNN	FF++, DFDC, Celeb-DF, WLDR, DFD	%98,9 ACC
[74]	CNN	Özel Veri Kümesi	%99,6 ACC
[75]	CNN	FF++, DFDC, DFD	%96 AUC
[76]	Transformer	FF++	%94,43 AUC
[77]	CNN	FF++, Celeb-DF	%91,76 ACC
[78]	Encoder-RNN	FF++, DFD, DeeperForensics, Celeb-DF	%99,95 AUC

4.6. Zaman Uyumsuzluğuna Dayalı Çalışmalar

Bu yöntemin amacı, videonun kareleri arasındaki tutarsızlıkları analiz edilerek düzmece görüntüleri algılamaktır.

Düzmece teknolojisinin hızla gelişmesiyle birlikte bu alandaki algılama metodolojileri de sürekli evrim geçirmektedir. Bu süreçte Guera ve Delp [79] RNN kullanarak düzmece algılamada önemli bir adım atmışlardır. Bu çalışmanın yaklaşımı, CNN ile çerçeve düzeyinde özellikler çıkararak RNN modeli eğitmeye dayanmaktadır. Bu yöntem sadece kısa süreli videolardaki uzamsal bilgileri yakalamada %94'ün üzerinde bir başarı elde ederek zamansal dinamiklerin düzmece algılamadaki kritik rolünü vurgulamıştır.

Montserrat ve ark. [80] videolardaki görsel değişiklikleri tespit etmek için Otomatik Yüz Ağırlığı (*Automatic Face Weighting - AFW*) mekanizması ile CNN ve RNN kullanarak videonun karelerindeki sahte yüz olasılıklarını ağırlıklandıran ve karelerin özellik vektörlerini zamansal (temporal) öğreniminde kullanılan bir GRU katmanı içermektedir. Geliştirilen model DFDC veri kümesinde %91,88 doğruluk elde etmişse de ses içeriği için bir analiz verilmemiştir.

Zheng ve ark. [81] gerçek ve sahte videolar arasındaki uzun vadeli bağımlılıkları yakalamak için zamansal transformatör kullanmayı önermişlerdir. Zamansal transformatör videonun her bir karesini ve kareler arasındaki ilişkileri dikkate alan bir evrişimsel ağ türüdür. Yazarlar, transformatörlerin küresel bağımlılıkları yakalamaya çok daha uygun olduğunu ancak zamana dayalı evrişimsel bir ağ kullanarak genelleme yeteneğinin geliştirilebileceğine dikkat çekmişlerdir. Çalışmada önerilen yöntemin %94,2 doğruluk oranıyla gerçek ve sahte videoları ayırt edebildiği gösterilmiştir.

Saikia ve ark. [82] videodaki nesnelerin ve kameranın hareketini takip eden "optik akış" tekniği ile görüntülerden anlamlı özellikler çıkarmak için CNN ve zamansal sıralardaki bilgileri işleme ve öğrenme için RNN bir karışımı bir model kullanmışlardır. Model çeşitli veri kümelerinde test edilmiş ve FF++ veri kümesinde en yüksek %91,21 doğruluk oranı elde etmiştir.

Rahman ve ark. [83] düzmece videoları algılayabilecek zamana duyarlı bir çerçeve geliştirmişlerdir. Araştırmacılar InceptionResNetV2, MobileNet ve DenseNet modelleriyle basit CNN kullanmış ve analizlerde MobileNet en iyi performansı göstermiştir. Kullanılan CNN modeli özellikle düşük çözünürlüklü ve kısa süreli videolarda DFDC ve FF++ veri kümeleri üzerinde sınandığında yüksek doğruluk oranları elde etmiştir. Model DFDC veri kümesinde %94,93 ve FF++ veri kümesinde %93,2 doğruluk oranına ulaşmıştır.

Kolagati ve ark. [84], CNN ve Çok Katmanlı Algılayıcı (*Multi-Layer Perceptron - MLP*)'yı birleştirerek düzmece videoları sınıflandıran hibrit bir model geliştirmişlerdir. Model CNN ile videodaki yüzün belirgin noktalarını analiz ederek özellikler çıkarmakta ve MLP ile ilk sınıflandırmayı gerçekleştirmektedir. Son karar ise CNN ve MLP'nin ortak bilgisine dayanmaktadır. 199 sahte ve 119 gerçek videodan oluşan test veri kümesinde model %87 doğruluk elde ederek sadece CNN kullanan modelleri geride bırakmıştır. Modelin aynı zamanda aşırı öğrenmeyi (*overfitting*) azaltarak daha hızlı eğitim aldığı belirtilmiştir.

Thing [85] çalışmasında düzmece algılamada zamansal tutarsızlıklara ve çoklu-veri kümesi değerlendirmesine odaklanmıştır. Araştırmada CNN ve Transformer modellerinin düzmece tespitindeki etkinlikleri incelenmiş ve veri kümesi bağımlılığı, mimari seçimi ve veri dengesizliği gibi zorluklar ele alınmıştır. Çalışma kapsamında FF++, Google DFD, Celeb-DF, DeeperForensics ve DFDC gibi köklü veri kümeleri üzerinde düzmece tespit modelleri test edilmiştir. DeeperForensics veri kümesinde %99,73 doğruluk elde edilmiştir.

Gu ve ark. [53], CNN ve kareler arası analiz için LSTM ağlarını kullandıkları yöntemde FF++ veri kümesinde %92,4 algılama doğruluğu elde edilmiştir.

Çizelge-8'de zaman uyumsuzluğuna dayalı düzmece video algılama çalışmaları verilmiştir.

Çizelge-8: Zaman Uyumsuzluğuna Dayalı Düzmece Video Tespit Çalışmaları

Çalışma	Önerilen Model	Veri Setleri	Başarım Oranı
[79]	CNN-RNN	İnternet Videoları	%94,00 ACC
[80]	CNN-GRU	DFDC	%91,88 ACC
[81]	Transformatör	FF++, FaceShifter, DeeperForensics, DFDC	%94,2 ACC
[82]	CNN-RNN	DFDC, FF++, Celeb-DF	%91,21 ACC
[83]	CNN	DFDC,FF++	%94,93 ACC
[84]	CNN ve MLP	DFDC,Youtube	%87 ACC
[85]	CNN ve Transformer	FF++, DFD, Celeb-DF, DeeperForensics, DFDC	%99,73 ACC
[53]	CNN-LSTM	FF++	%92,4 ACC

4.7. Yüz İfadelerine Dayalı Çalışmalar

Mittal ve ark. [86], yüz ifadeleri ve sesteki duygusal ifadelerden yola çıkarak Siyam (*Siamese*) sinir ağları ve modelin benzer örnekleri ayırt etmeyi ve farklı örnekleri birbirinden ayırmayı öğrenmesini sağlayan triplet kayıp fonksiyonundan esinlenen bir derin öğrenme yaklaşımı ortaya koydular. Özgünlükleri, bu modelde hem ses hem görüntü içeriklerinin yanında kişinin bu içeriklerdeki duygularını da dahil etmeleridir. Bu yöntem, DFDC veri kümesinde %84,4 ve DF-TIMIT veri kümesinde %96,6 AUC skoru sağlamıştır. Ancak yazarlar, insan duygularının karmaşıklığı nedeniyle oluşabilecek algılama başarısızlıklarına da dikkat çekmektedir.

Hosler ve ark. [87] düzmece tespiti için sesteki ve görüntüdeki duygusal ifadeler arasındaki tutarsızlıklara odaklanan bir yöntem geliştirdiler. Yöntem, zaman içinde duyguları temsil etmek için kişinin konuşmasından ve yüz ifadelerinden temel özellikler çıkarmayı ve ardından bu özellikleri eğitilmiş bir RNN modeli ile analiz etmeyi içermektedir. RNN modelinde, konuşmadaki temel duygusal özellikleri incelemek için LSTM ağları ve tahmin edilen duygu bilgisine dayanarak düzmece videoları ayırt etmek için bir gözetimli sınıflandırıcı kullanılmıştır. Çalışmada, düzmece videolarda yüz ifadelerinin taklit edilmesine odaklanarak sahteciliği yakalayabilmekten ziyade düzmece ile sesteki duyguların taklit edilmesinin daha başarısız olduğu vurgulanmıştır. Öte yandan daha uzun video örneklerinde bu yöntemin başarısı %99,5 seviyesine kadar çıkabilmektedir.

Pei ve ark. [88] video içeriklerindeki düzmece yüzlerin tespiti için zamansal özelliklere odaklanan Çift Yönlü LSTM yöntemi önermişlerdir. Bu yöntem, yüz ifadelerindeki zamansal değişimleri analiz ederek kaş kaldırma, göz kırpma ve gülümseme gibi ince yüz ifadelerini dikkate alarak sahte ve gerçek yüzlerin ayırt edilmesi sağlamaktadır. Çift Yönlü LSTM yöntemi daha az eğitim süresi gerektirerek DFDC veri kümesi üzerinde %82,65 doğruluk oranı elde etmiştir. Bu yöntemin özellikle de gürültülü sıkıştırılmış videoları tespit etmede kayda değer başarılı olduğu belirtilmiştir.

Güncel ve önemli düzmece tespit çalışmalarından biri olarak Haq ve ark. [89], psikolojik bilginin ve sembolik muhakeme yeteneğinin birleştirildiği psikolojik analiz temelinde yeni bir biyolojik sinyal geliştirdiler. Bu sinyal, kişinin duygusal geçişlerindeki tutarlılığa ve farklı ifade araçları (mimik, ses, vücut dili) arasındaki uyuma odaklanmaktadır. Analiz sonuçlarında PDD veri kümesinde %93,7 doğruluk oranı, WLDR veri kümesinde ise %75,34 doğruluk oranı elde edilmiştir. Psikoloji ve duygu bilimi alanındaki temel bilgileri kullanan bu yöntemin daha genellenebilir olması oldukça yüksektir.

Çizelge-9'da yüz ifadelerine dayalı düzmece video algılama çalışmaları özetlenmiştir.

Çizelge-9: Yüz İfadelerine Dayalı Düzmece Video Tespit Çalışmaları

Çalışma	Önerilen Model	Veri Setleri	Başarım Oranı
[86]	IQIEA-FS	DFDC, DF-TIMIT	%96,6 AUC
[87]	RNN	Celeb-DF	%99,5 ACC
[88]	LSTM	DFDC	%82,65 ACC
[89]	ResNet50	WLDR, PDD	%93,7 ACC

4.8. Uzamsal-Zamansal Özelliklere Dayalı Tespit

Uzamsal-zamansal özelliklerden yararlanma yöntemi, düzmece algılamada yaygın kullanılan çok modlu (farklı veri kaynağı kullanan) bir yaklaşımdır. Burada video kareleri içindeki görsel düzensizlikler (*intra-frame inceleme*) ve video akışları boyunca zamansal özellikler (*inter-frame inceleme*) analiz edilir [28].

Nguyen ve ark. [90] çoklu görev metodu CNN tabanlı Y-shaped Autoencoder (*Y-şekilli otokodlayıcı*) modeli, videolardan oynanmış içeriğin hem uzamsal (*spatial*) hem de zamansal (*temporal*) özelliklerini analiz ederek tespit ve yerelleştirme işlemleri için tasarlanmıştır. Bu model düzmece algılamada FF+ veri kümesinde %92,77 doğruluk oranı sağlasa da görülmemiş senaryolarda değerlendirme doğruluğunun düştüğü belirtilmiştir.

De Lima ve ark. [91] düzmece videolardaki karelerden uzamsal ve zamansal bilgileri (*3D input*) öğrenerek manipülasyonları tespit etmeyi hedeflemişlerdir. Çalışmada CNN modeli VGG-11 ile videolardan görsel özellik çıkarılmış ardından bu özelliklerin zaman içindeki sırasını analiz etmek için LSTM kullanılmıştır. R3D, ResNet, I3D gibi farklı CNN modelleri LSTM'in ürettiği zaman sıralı verilerle eğitilerek düzmece tespiti için zor olan Celeb-DF veri kümesine üzerinde test edilmiştir. Bu çalışmada R3D modeli en iyi sonuç vererek düzmece videolarını %98,26 doğruluk oranıyla sınıflandırabilmiştir. Ancak bu yaklaşım yüksek hesaplama maliyeti gerektirmektedir.

Li ve ark. [92], birden fazla video karesini tek seferde işledikleri çalışmada zaman içerisinde yüzdeki kısmi değişikleri algılamayı amaçlamışlardır. Model FF++ veri kümesinde %99,82 doğruluk elde etmiştir.

Hubalovsky ve ark. [93] CNN tabanlı YOLO dedektör ve yerel ikili desen histogramı (*local binary pattern histogram*) kullanarak, uzamsal ve zamansal özellikleri bir araya getirdiler. Bu yöntemle öncelikle video karelerinde veya görüntülerde yüz tespiti yapılmakta ve ardından uzamsal özellikler çıkarılmaktadır. Çalışmada %98,12 doğruluk elde edilmiştir.

Lu ve ark. [94], iCapsNet-TSF olarak adlandırdıkları gelişmiş bir Kapsül Ağı ile zamansal-uzamsal özellikleri yakalayan optik akış algoritmasını birleştiren bir yöntem geliştirmişlerdir. Ayrıca zamansal-uzamsal verileri analiz etmek için Dinamik Yönlendirme Algoritması kullanılmıştır. Kapsül ağı içindeki ağırlık başlatma ve güncelleme gibi iyileştirmeler sayesinde bu yaklaşım düzmece algılama doğruluğunu önemli ölçüde artırarak %98,83 doğruluk yakalamıştır. Ek olarak, yöntem farklı veri kümeleri arasında çalışırken de doğruluğu koruma zorluğunun üstesinden gelmiştir.

Dolla ve ark. [95], düzmece algılamayı iyileştirmek için uzamsal-zamansal özellik piramit ağı geliştirdiler. Bu ağ, video karelerinden gelen uzamsal ve zamansal özellikleri güçlendirerek çalışmaktadır. Model, doku bazlı özniteliklerle yüksek seviye görsel detayları bir araya getirerek yüzleri doğru bir şekilde temsil etmeye odaklanır. CNN ve RNN tabanlı modellerin kullanıldığı ve FF++, DeepForensics ve CelebDF gibi farklı veri kümelerinin sınandığı analizlerde FF++ veri kümesinde %99,99 AUC değeri elde edilmiştir.

Kaddar ve ark. [96] düzmece video algılama için uzamsal-zamansal özelliklere odaklanarak HCiT adını verdikleri mimariyi kullanmışlardır. HCiT modelinde bir CNN katmanı ilk olarak girdi görüntüsündeki kenar ve köşe gibi alt seviye görsel özellikleri çıkarmaktadır. Bu öznitelikler daha sonra ViT (*Vision Transformer*) mimarisine beslenir ve dikkat mekanizmasını kullanarak bu detaylar arasındaki ilişkiyi analiz etmektedir. Yöntem, özellikle yüz değiştirme tekniklerıyle oynanmış görüntülerde yüksek değerler sağlarken FaceSwap ve Face2Face alt veri kümelerinde %95,85'in üzerinde doğruluk göstermiştir.

Çizelge-10'da uzamsal-zamansal özelliklere dayalı düzmece video algılama çalışmalarının özeti yer almaktadır.

Çizelge-10: Uzamsal-Zamansal Özelliklere Dayalı Düzmece Video Tespit Çalışmaları

Çalışma	Önerilen Model	Veri Setleri	Başarım Oranı
[90]	CNN	FF++	%92,77 ACC
[91]	CNN	Celeb-DF	%98,26 ACC
[92]	S-MIL	Celeb-DF, FF++, DFDC	%100 ACC
[93]	CNN-YOLO	CelebDF-FF++, DFFD, CASIA-Web Face Dataset	%98,12 ACC
[94]	iCapsNet-TSF	Celeb-DF, FaceSwap, Deepfakes	%98,83 ACC
[95]	CNN-RNN	DeepForensics, Celeb-DF, FaceForensics++	%99,99 AUC
[96]	ViT	FF++, DFDC preview, Celeb -DF	%95,85 ACC

5. Sonuç

Bu literatür incelemesi, düzmece video algılama amacıyla yapılan son yıllarda yapılan çalışmaları ele almaktadır. Bu alanda yapılan çalışmalar kayda değer başarılar sergilese de düzmece algılama hala birçok açıdan geliştirilmesi gereken önemli bir konudur.

Düzmece üretimiyle ilgili endişe verici bir nokta, GAN modellerinin hızla gelişmesiyle inandırıcılığı yüksek video görüntülerinin ortaya çıkmasıdır. Bu eğilim, düzmece tespitini daha da zorlaştırdığı için mevcut yöntemlerden daha etkili algoritmaların ortaya çıkması veya mevcut algoritmaların evrimleşmesi ile bu sorunun çözülmesi sağlanabilecektir [34]. Çalışmadan da anlaşılacağı üzere günümüz tespit yöntemleri görülmemiş veri kümelerinde %100 algılama doğruluk oranı sağlayamamaktadır.

Veri kümeleri arasındaki tutarsızlıklar ve modelin hiç karşılaşmadığı veri türleri mevcut düzmece algılama modellerinin başarısını etkilemektedir. Düzmece algılama modellerinin etkinliği, eğitim için çeşitli ve çok sayıda örnek içeren veri kümelerine büyük ölçüde bağlıdır. Modeller, bilinmeyen manipülasyonlara sahip ortamlarla karşılaştığında bu tür manipülasyonları doğru bir şekilde tanımlama yetenekleri düşmektedir. Düzmece algılama sistemlerini aldatmak amacıyla video üzerinde bulanıklaştırma, yumuşatma, kırpma gibi işlemler uygulanır. Manipülasyon tekniklerindeki bu çeşitlilik kapsamlı veri kümelerinin eksikliğiyle birleştiğinde etkili tespit modelleri tasarlamada önemli bir zorluk oluşturmaktadır [23]. Bu nedenle veri kümelerinin geliştirilmesi ve bu konuda gerekli desteklerin belirli kuruluşlar tarafından sağlanması önemlidir.

Bir başka düzmece algılama kısıtı olan hesaplama kısıtları hem düzmece üretme hem de algılamayı önemli ölçüde etkilemektedir. Genellikle GAN modelleri gibi gelişmiş derin öğrenme modelleri kullanan düzmece oluşturma süreci yüksek hesaplama gücü gerektirmektedir. Siber suç analizi veya içerik denetimi gibi uygulamalar için gerçek zamanlı analiz gerektiren düzmece algılama algoritmaları yüksek çözünürlüklü videoları işlemek için yüksek işlem gücü gerektirdiği gibi çok miktarda enerji tüketerek yüksek maliyetlere yol açmaktadır. Bu durum sınırlı kaynaklara sahip olanlar için bu teknolojinin kullanımını zorlaştırmaktadır. Sonuç olarak, bu hesaplama sorunlarını gidermek için daha verimli yapay zekâ modellerine ve özel donanım geliştirmeye artan bir ihtiyaç vardır [27].

Kaynakça

[1] Statista. Global internet user penetration 2024, https://0311b0kku-y-https-www-statista-com.halic.proxy.deepknowledge.io/statistics/325706/global-internet-user-penetration/, Erişim tarihi:15.02.2024.

[2] ITRC. 2023 Annual Data Breach Report, https://www.idtheftcenter.org/publication/2023-data-breach-report/, Erişim tarihi:15.02.2024.

[3] Heidari, A., Jafari Navimipour, N., Dag, H., & Unal, M. Deepfake detection using deep learning methods: A systematic and comprehensive review, WIREs Data Mining and Knowledge Discovery, 2023. https://doi.org/10.1002/widm.1520

[4] Salman, S., Shamsi, J. A., & Qureshi, R., Deep Fake Generation and Detection: Issues, Challenges, and Solutions. IT Professional, 2023, 25(1), 52-59. https://doi.org/10.1109/MITP.2022.3230353

[5] Kietzmann, J., Lee, L. W., McCarthy, I. P., & Kietzmann, T. C., Deepfakes: Trick or treat? Business Horizons, 2020, 63(2), 135-146. https://doi.org/10.1016/j.bushor.2019.11.006

[6] Ajao, O., Bhowmik, D., & Zargari, S., Sentiment Aware Fake News Detection on Online Social Networks, IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2019, 2507-2511. https://doi.org/10.1109/ICASSP.2019.8683170

[7] Caldelli, R., Galteri, L., Amerini, I., & Bimbo, A., Optical Flow based CNN for detection of unlearnt Deepfake manipulations, Pattern Recognition Letters, 2021, 146. https://doi.org/10.1016/j.patrec.2021.03.005

[8] Van Der Sloot, B., & Wagensveld, Y., Düzmeces: Regulatory challenges for the synthetic society, Computer Law & Security Review, 2022, 46, 105716. https://doi.org/10.1016/j.clsr.2022.105716

[9] Neethirajan, S., Is Seeing Still Believing? Leveraging Deepfake Technology for Livestock Farming, Frontiers in Veterinary Science, 2021, 8. https://api.semanticscholar.org/CorpusID:244715980

[10] Yu, P., Xia, Z., Fei, J., & Lu, Y., A Survey on Deepfake Video Detection, IET Biometrics, 2021, 10(6), 607-624. https://doi.org/10.1049/bme2.12031

[11] Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y., Generative Adversarial Networks, 2014. https://doi.org/10.48550/ARXIV.1406.2661

[12] He, Z., Zuo, W., Kan, M., Shan, S., & Chen, X., AttGAN: Facial Attribute Editing by Only Changing What You Want, IEEE Transactions on Image Processing, 2019, 28(11), 5464-5478. https://doi.org/10.1109/TIP.2019.2916751

[13] Karras, T., Laine, S., & Aila, T., A Style-Based Generator Architecture for Generative Adversarial Networks, 2018. https://doi.org/10.48550/ARXIV.1812.04948

[14] Liu, M., Ding, Y., Xia, M., Liu, X., Ding, E., Zuo, W., & Wen, S., STGAN: A Unified Selective Transfer Network for Arbitrary Image Attribute Editing, 2019. https://doi.org/10.48550/ARXIV.1904.09709

[15] Choi, Y., Uh, Y., Yoo, J., & Ha, J.-W., StarGAN v2: Diverse Image Synthesis for Multiple Domains, 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020, 8185-8194. https://doi.org/10.1109/CVPR42600.2020.00821

[16] Temnikova, I., Marinova, I., Looking for Traces of Textual Deepfakes in Bulgarian on Social Media, Proceedings of the Conference Recent Advances in Natural Language Processing - Large Language Models for Natural Language Processings, 2023, 1151-1161. https://doi.org/10.26615/978-954-452-092-2_122

[17] Conti, E., Salvi, D., Borrelli, C., Hosler, B., Bestagini, P., Antonacci, F., Sarti, A., Stamm, M. C., & Tubaro, S., Deepfake Speech Detection Through Emotion Recognition: A Semantic Approach, ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2022, 8962-8966. https://doi.org/10.1109/ICASSP43922.2022.9747186

[18] Patel, Y., Tanwar, S., Bhattacharya, P., Gupta, R., Alsuwian, T., Davidson, I. E., & Mazibuko, T. F., An Improved Dense CNN Architecture for Deepfake Image Detectio, IEEE Access, 2023, 11, 22081-22095. https://doi.org/10.1109/ACCESS.2023.3251417

[19] Yang, X., Li, Y., & Lyu, S., Exposing Deep Fakes Using Inconsistent Head Poses, 2018. https://doi.org/10.48550/ARXIV.1811.00661

[20] Li, L., Bao, J., Zhang, T., Yang, H., Chen, D., Wen, F., & Guo, B., Face X-ray for More General Face Forgery Detection, 2019. https://doi.org/10.48550/ARXIV.1912.13458

[21] Tolosana, R., Vera-Rodriguez, R., Fierrez, J., Morales, A., & Ortega-Garcia, J., Deepfakes and Beyond: A Survey of Face Manipulation and Fake Detection, 2020. https://doi.org/10.48550/ARXIV.2001.00179

[22] Zi, B., Chang, M., Chen, J., Ma, X., & Jiang, Y.-G., WildDeepfake: A Challenging Real-World Dataset for Deepfake Detection, Proceedings of the 28th ACM International Conference on Multimedia, 2020.

[23] Masood, M., Nawaz, M., Malik, K. M., Javed, A., & Irtaza, A., Deepfakes Generation and Detection: State-of-the-art, open challenges, countermeasures, and way forward, 2021, https://doi.org/10.48550/ARXIV.2103.00484

[24] Waseem, S., Abu-Bakar, S., Omar, Z., Ahmed, B., Baloch, S., & Hafeezallah, A., Multi-attention-based approach for Deepfake face and expression swap detection and localization, EURASIP Journal on Image and Video Processing, 2023. https://doi.org/10.1186/s13640-023-00614-z

[25] Chung, J. S., Nagrani, A., & Zisserman, A., VoxCeleb2: Deep Speaker Recognition, 2018. https://doi.org/10.48550/ARXIV.1806.05622

[26] Salvi, D., Liu, H., Mandelli, S., Bestagini, P., Zhou, W., Zhang, W., & Tubaro, S., A Robust Approach to Multimodal Deepfake Detection, Journal of Imaging, 2023, 9(6), 122. https://doi.org/10.3390/jimaging9060122

[27] Gupta, G., Raja, K., Gupta, M., Jan, T., Whiteside, S. T., & Prasad, M., A Comprehensive Review of Deepfake Detection Using Advanced Machine Learning and Fusion Methods, Electronics, 2023, 13(1), 95. https://doi.org/10.3390/electronics13010095

[28] Juefei-Xu, F., Wang, R., Huang, Y., Guo, Q., Ma, L., & Liu, Y., Countering Malicious Deepfakes: Survey, Battleground, and Horizon, International Journal of Computer Vision, 2022, 130, 1-57. https://doi.org/10.1007/s11263-022-01606-8

[29] Rana, M. S., Nobi, M. N., Murali, B., & Sung, A. H., Deepfake Detection: A Systematic Literature Review, IEEE Access, 2022, 10, 25494-25513. https://doi.org/10.1109/ACCESS.2022.3154404

[30] Groh, M., Epstein, Z., Firestone, C., & Picard, R., Deepfake detection by human crowds, machines, and machine-informed crowds, Proceedings of the National Academy of Sciences, 2022, 119(1), e2110013119. https://doi.org/10.1073/pnas.2110013119

[31] Köbis, N. C., Doležalová, B., & Soraperra, I., Fooled twice: People cannot detect Deepfakes but think they can, iScience, 2021, 24(11), 103364. https://doi.org/10.1016/j.isci.2021.103364

[32] Somoray, K., & Miller, D. J., Providing detection strategies to improve human detection of Deepfakes: An experimental study, Computers in Human Behavior, 2023, 149, 107917. https://doi.org/10.1016/j.chb.2023.107917

[33] Deng, L., Deep Learning: Methods and Applications, Foundations and Trends® in Signal Processing, 2014, 7(3-4), 197-387. https://doi.org/10.1561/2000000039

[34] Naitali, A., Ridouani, M., Salahdine, F., & Kaabouch, N., Deepfake Attacks: Generation, Detection, Datasets, Challenges, and Research Directions, Computers, 2023, 12(10), 216. https://doi.org/10.3390/computers12100216

[35] Krizhevsky, A., Sutskever, I., & Hinton, G. E., ImageNet classification with deep convolutional neural networks, Communications of the ACM, 2017, 60(6), 84-90. https://doi.org/10.1145/3065386

[36] Simonyan, K., & Zisserman, A., Very Deep Convolutional Networks for Large-Scale Image Recognition, 2014. https://doi.org/10.48550/ARXIV.1409.1556

[37] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., & Rabinovich, A., Going Deeper with Convolutions, 2014. https://doi.org/10.48550/ARXIV.1409.4842

[38] He, K., Zhang, X., Ren, S., & Sun, J., Deep Residual Learning for Image Recognition, 2015. https://doi.org/10.48550/ARXIV.1512.03385

[39] Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q., Densely Connected Convolutional Networks, 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, 2261-2269. https://doi.org/10.1109/CVPR.2017.243

[40] Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., & Adam, H., MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications, 2017. https://doi.org/10.48550/ARXIV.1704.04861

[41] Tan, M., & Le, Q. V., EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks, 2019. https://doi.org/10.48550/ARXIV.1905.11946

[42] Afchar, D., Nozick, V., Yamagishi, J., & Echizen, I., MesoNet: A Compact Facial Video Forgery Detection Network, 2018. https://doi.org/10.48550/ARXIV.1809.00888

[43] Zhao, H., Wei, T., Zhou, W., Zhang, W., Chen, D., & Yu, N., Multi-attentional Deepfake Detection, 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021, 2185-2194. https://doi.org/10.1109/CVPR46437.2021.00222

[44] Kohli, A., & Gupta, A., Detecting Deepfake, FaceSwap and Face2Face facial forgeries using frequency CNN, Multimedia Tools and Applications, 2021, 80(12), 18461-18478. https://doi.org/10.1007/s11042-020-10420-8

[45] Luo, Y., Zhang, Y., Yan, J., & Liu, W., Generalizing Face Forgery Detection with High-frequency Features, 2021. https://doi.org/10.48550/ARXIV.2103.12376

[46] Ismail, A. A., Elpeltagy, M. S., Zaki, M. S., & Eldahshan, K. A., A New Deep Learning-Based Methodology for Video Deepfake Detection Using XGBoost, Sensors, 2021, 21.

[47] Das, A., & Sebastian, L., A Comparative Analysis and Study of a Fast Parallel CNN Based Deepfake Video Detection Model with Feature Selection (FPC-DFM), 2023 Advanced Computing and Communication Technologies for High Performance Applications (ACCTHPA), 2023, 1-9. https://doi.org/10.1109/ACCTHPA57160.2023.10083340

[48] Dhanaraj, R., & Sri̇Devi̇, M., Face Warping Deepfake Detection and Localization in a Digital Video using Transfer Learning Approach, Journal of Metaverse, 2023, 4(1), 11-20. https://doi.org/10.57019/jmv.1338907

[49] Chollet, F., Xception: Deep Learning with Depthwise Separable Convolutions, 2017. http://arxiv.org/abs/1610.02357

[50] Matern, F., Riess, C., & Stamminger, M., Exploiting Visual Artifacts to Expose Deepfakes and Face Manipulations, 2019 IEEE Winter Applications of Computer Vision Workshops (WACVW), 2019, 83-92. https://doi.org/10.1109/WACVW.2019.00020

[51] Li, Y., Chang, M.-C., & Lyu, S., In Ictu Oculi: Exposing AI Created Fake Videos by Detecting Eye Blinking, 2018, 1-7. https://doi.org/10.1109/WIFS.2018.8630787

[52] Jung, T., Kim, S., & Kim, K., DeepVision: Derin Deepfakes Detection Using Human Eye Blinking Pattern, IEEE Access, 2020, 8, 83144-83154. https://doi.org/10.1109/ACCESS.2020.2988660

[53] Gu, J., Xu, Y., Sun, J., & Liu, W., Exploiting Deepfakes by Analyzing Temporal Feature Inconsistency, International Journal of Advanced Computer Science and Applications, 2023, 14(12). https://doi.org/10.14569/IJACSA.2023.0141291

[54] He, Q., Peng, C., Liu, D., Wang, N., & Gao, X., GazeForensics: Deepfake Detection via Gaze-guided Spatial Inconsistency Learning, 2023. https://doi.org/10.48550/ARXIV.2311.07075

[55] Qi, H., Guo, Q., Juefei-Xu, F., Xie, X., Ma, L., Feng, W., Liu, Y., & Zhao, J., DeepRhythm: Exposing Deepfakes with Attentional Visual Heartbeat Rhythms, 2020. https://doi.org/10.48550/ARXIV.2006.07634

[56] Ciftci, U., Demir, I., & Yin, L., FakeCatcher: Detection of Synthetic Portrait Videos using Biological Signals, IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, PP, 1-1. https://doi.org/10.1109/TPAMI.2020.3009287

[57] Ciftci, U. A., Demir, I., & Yin, L., How Do the Hearts of Deep Fakes Beat? Deep Fake Source Detection via Interpreting Residuals with Biological Signals, 2020. https://doi.org/10.48550/ARXIV.2008.11363

[58] Hernandez-Ortega, J., Tolosana, R., Fierrez, J., & Morales, A., DeepfakesON-Phys: Deepfakes Detection based on Heart Rate Estimation, 2020. https://doi.org/10.48550/ARXIV.2010.00400

[59] Wang, B., Li, Y., Wu, X., Ma, Y., Song, Z., & Wu, M., Face Forgery Detection Based on the Improved Siamese Network, Security and Communication Networks, 2022, 1-13. https://doi.org/10.1155/2022/5169873

[60] Khurana, P. S., Sudarshan, T. B., Natarajan, S., Nagesh, V., Lakshminarayanan, V., Bhat, N., & Vinay, A., AFMB-Net: Deepfake Detection Network Using Heart Rate Analysis, Tehnički glasnik, 2022, 16(4), 503-508. https://doi.org/10.31803/tg-20220403080215

[61] Liang, P., Liu, G., Xiong, Z., Fan, H., Zhu, H., & Zhang, X., A facial geometry based detection model for face manipulation using CNN-LSTM architecture, Information Sciences, 2023, 633, 370-383. https://doi.org/10.1016/j.ins.2023.03.079

[62] Zhou, Y., & Lim, S.-N., Joint Audio-Visual Deepfake Detection, 2021 IEEE/CVF International Conference on Computer Vision (ICCV), 2021, 14780-14789. https://doi.org/10.1109/ICCV48922.2021.01453

[63] Cai, Z., Stefanov, K., Dhall, A., & Hayat, M., Do You Really Mean That? Content Driven Audio-Visual Deepfake Dataset and Multimodal Method for Temporal Forgery Localization, 2022. https://doi.org/10.48550/ARXIV.2204.06228

[64] Ilyas, H., Javed, A., & Malik, K. M., AVFakeNet: A unified end-to-end Dense Swin Transformer deep learning model for audio–visual Deepfakes detection, Applied Soft Computing, 2023, 136, 110124. https://doi.org/10.1016/j.asoc.2023.110124

[65] Anas Raza, M., & Mahmood Malik, K, Multimodaltrace: Deepfake Detection using Audiovisual Representation Learning, 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2023, 993-1000. https://doi.org/10.1109/CVPRW59228.2023.00106

[66] Hashmi, A., Shahzad, S. A., Lin, C.-W., Tsao, Y., & Wang, H.-M., AVTENet: Audio-Visual Transformer-based Ensemble Network Exploiting Multiple Experts for Video Deepfake Detection, 2023. https://doi.org/10.48550/ARXIV.2310.13103

[67] Li, Y., & Lyu, S., Exposing Deepfake Videos By Detecting Face Warping Artifacts, 2018. https://doi.org/10.48550/ARXIV.1811.00656

[68] Frank, J., Eisenhofer, T., Schönherr, L., Fischer, A., Kolossa, D., & Holz, T., Leveraging Frequency Analysis for Deep Fake Image Recognition, 2020. https://doi.org/10.48550/ARXIV.2003.08685

[69] Younus, M. A., & Hasan, T. M., Effective and Fast Deepfake Detection Method Based on Haar Wavelet Transform, 2020 International Conference on Computer Science and Software Engineering (CSASE), 2020, 186-190. https://doi.org/10.1109/CSASE48920.2020.9142077

[70] Huang, Y., Juefei-Xu, F., Guo, Q., Liu, Y., & Pu, G., FakeLocator: Robust Localization of GAN-Based Face Manipulations, 2020. https://doi.org/10.48550/ARXIV.2001.09598

[71] Xiao, S., Yang, J., & Lv, Z., Protecting the trust and credibility of data by tracking forgery trace based on GANs, Digital Communications and Networks, 2022, 8(6), 877-884. https://doi.org/10.1016/j.dcan.2022.07.010

[72] Lin, Y.-K., & Sun, H.-L., Few-Shot Training GAN for Face Forgery Classification and Segmentation Based on the Fine-Tune Approach, Electronics, 2023, 12(6), 1417. https://doi.org/10.3390/electronics12061417

[73] Agarwal S, Farid H, El-Gaaly T, Lim SN., Detecting deep-fake videos from appearance and behavior, In: 2020 IEEE International Workshop on Information Forensics and Security (WIFS), IEEE, 2020, pp 1–6.

[74] Agarwal, S., Farid, H., Fried, O., & Agrawala, M., Detecting Deep-Fake Videos from Phoneme-Viseme Mismatches, 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2020 2814-2822. https://doi.org/10.1109/CVPRW50498.2020.00338

[75] Cozzolino, D., Rössler, A., Thies, J., Nießner, M., & Verdoliva, L., ID-Reveal: Identity-aware Deepfake Video Detection, 2020. https://doi.org/10.48550/ARXIV.2012.02512

[76] Dong, X., Bao, J., Chen, D., Zhang, T., Zhang, W., Yu, N., Chen, D., Wen, F., & Guo, B., Protecting Celebrities from Deepfake with Identity Consistency Transformer, 2022. https://doi.org/10.48550/ARXIV.2203.01318

[77] Shen, D., Zhao, Y., & Quan, C., Identity-Referenced Deepfake Detection with Contrastive Learning, Proceedings of the 2022 ACM Workshop on Information Hiding and Multimedia Security, 2022, 27-32. https://doi.org/10.1145/3531536.3532964

[78] Liu, B., Liu, B., Ding, M., Zhu, T., & Yu, X., TI 2 Net: Temporal Identity Inconsistency Network for Deepfake Detection, 2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), 2023, 4680-4689. https://doi.org/10.1109/WACV56688.2023.00467

[79] Guera, D., & Delp, E., Deepfake Video Detection Using Recurrent Neural Networks, 2018, 1-6. https://doi.org/10.1109/AVSS.2018.8639163

[80] Montserrat, D. M., Hao, H., Yarlagadda, S. K., Baireddy, S., Shao, R., Horváth, J., Bartusiak, E., Yang, J., Güera, D., Zhu, F., & Delp, E. J., Deepfakes Detection with Automatic Face Weighting, 2020. https://doi.org/10.48550/ARXIV.2004.12027

[81] Zheng, Y., Bao, J., Chen, D., Zeng, M., & Wen, F., Exploring Temporal Coherence for More General Video Face Forgery Detection, 2021. https://doi.org/10.48550/ARXIV.2108.06693

[82] Saikia, P., Dholaria, D., Yadav, P., Patel, V., & Roy, M., A Hybrid CNN-LSTM model for Video Deepfake Detection by Leveraging Optical Flow Features, 2022. https://doi.org/10.48550/ARXIV.2208.00788

[83] Rahman, A., Siddique, N., Moon, M. J., Tasnim, T., Islam, M., Shahiduzzaman, Md., & Ahmed, S., Short And Low Resolution Deepfake Video Detection Using CNN, 2022 IEEE 10th Region 10 Humanitarian Technology Conference (R10-HTC), 2022, 259-264. https://doi.org/10.1109/R10-HTC54060.2022.9929719

[84] Kolagati, S., Priyadharshini, T., & Mary Anita Rajam, V., Exposing Deepfake using a deep multilayer perceptron – convolutional neural network model, International Journal of Information Management Data Insights, 2022, 2(1), 100054. https://doi.org/10.1016/j.jjimei.2021.100054

[85] Thing, V. L. L., Deepfake Detection with Deep Learning: Convolutional Neural Networks versus Transformers, 2023. https://doi.org/10.48550/ARXIV.2304.03698

[86] Mittal, T., Bhattacharya, U., Chandra, R., Bera, A., & Manocha, D., Emotions Don't Lie: An Audio-Visual Deepfake Detection Method using Affective Cues, Proceedings of the 28th ACM International Conference on Multimedia, 2022, 2823-2832. https://doi.org/10.1145/3394171.3413570

[87] Hosler, B., Salvi, D., Murray, A., Antonacci, F., Bestagini, P., Tubaro, S., & Stamm, M. C., Do Deepfakes Feel Emotions? A Semantic Approach to Detecting Deepfakes Via Emotional Inconsistencies, 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2021, 1013-1022. https://doi.org/10.1109/CVPRW53098.2021.00112

[88] Pei, S., Wang, Y., Xiao, B., Pei, S., Xu, Y., Gao, Y., & Zheng, J., A bidirectional-LSTM method based on temporal features for deep fake face detection in videos, 2nd International Conference on Information Technology and Intelligent Control, 2022, 28. https://doi.org/10.1117/12.2653461

[89] Haq, I. U., Malik, K. M., & Muhammad, K., Multimodal Neurosymbolic Approach for Explainable Deepfake Detection, ACM Transactions on Multimedia Computing, Communications, and Applications, 2023, 3624748. https://doi.org/10.1145/3624748

[90] Nguyen, H. H., Fang, F., Yamagishi, J., & Echizen, I., Multi-task Learning For Detecting and Segmenting Manipulated Facial Images and Videos, 2019. https://doi.org/10.48550/ARXIV.1906.06876

[91] de Lima, O., Franklin, S., Basu, S., Karwoski, B., & George, A., Deepfake Detection using Spatiotemporal Convolutional Networks, 2020. https://doi.org/10.48550/ARXIV.2006.14749

[92] Li, X., Lang, Y., Chen, Y., Mao, X., He, Y., Wang, S., Xue, H., & Lu, Q., Sharp Multiple Instance Learning for Deepfake Video Detection, 2020. https://doi.org/10.48550/ARXIV.2008.04585

[93] Hubálovský, Š., Trojovský, P., Bacanin, N., & K, V., Evaluation of deepfake detection using YOLO with local binary pattern histogram, PeerJ Computer Science, 2022, 8, e1086. https://doi.org/10.7717/peerj-cs.1086

[94] Lu, T., Bao, Y., & Li, L., Deepfake Video Detection Based on Improved CapsNet and Temporal–Spatial Features, Computers, Materials & Continua, 2023, 75(1), 715-740. https://doi.org/10.32604/cmc.2023.034963

[95] Dolla, M. S., Ruan, L., Zhu, K., & Xiao, L., Spatio-Temporal Feature Pyramid Network for Deepfake Detection, SSRN, 2023. https://doi.org/10.2139/ssrn.4507991

[96] Kaddar, B., Fezza, S. A., Akhtar, Z., Hamidouche, W., Hadid, A., & Serra-Sagristà, J., Deepfake Detection Using Spatiotemporal Transformer, ACM Transactions on Multimedia Computing, Communications, and Applications, 2024, 3643030. https://doi.org/10.1145/3643030