Gümüşhane Üniversitesi Fen Bilimleri Dergisi

GUFBD / *GUJS* (2024) 14(1): 164-178 doi: 10.17714/gumusfenbil.1288281 Araştırma Makalesi / *Research Article*

Derin özniteliklerin farklı atak senaryolarındaki yüz sahteciliği tespiti başarımlarının incelenmesi

Investigation of face spoofing detection performances of deep features in different attack scenarios

Asuman GÜNAY YILMAZ*[1], Fırat ŞAKAR[1]

[1]*Karadeniz Teknik Üniversitesi, Of Teknoloji Fakültesi, Yazılım Mühendisliği Bölümü, 61080, Trabzon*

• Geliş tarihi / *Received*: 26.04.2023 • Kabul tarihi / *Accepted*: 03.11.2023

Öz

Günümüzde yüz tanıma sistemlerinin kullanımının çoğalmasıyla birlikte bu sistemlere karşı yapılan saldırılar da artmıştır. Özellikle artan sosyal medya kullanımı ile yüz görüntü ve videolarının paylaşımının artışı, saldırganların bu içeriği kullanarak yüz tanıma sistemlerini daha kolay kandırmasına imkân sağlamaktadır. Bu nedenle yüz sahteciliği tespiti (YST) konusu oldukça önemli bir çalışma alanı haline gelmiştir. Yüz sahteciliği saldırıları çeşitli türlerde gerçekleştirilmektedir. Genellikle çalışmalarda tüm atak türlerinin birlikte değerlendirildiği senaryolar üzerinde başarım değerlendirilmesi yapılmaktadır. Bu nedenle bu çalışmada Replay-Attack veri setindeki Basılı Fotoğraf (Printed Photo), Dijital Fotoğraf (Digital Photo) ve Video Oynatma (Replay Video) saldırı türlerinde derin öğrenme yöntemlerinin YTS başarımları değerlendirilmiştir. Bu amaçla ilk aşamada VGG16, DenseNet121 ve MobileNet derin ağ mimarilerinin bu saldırı türlerindeki YST başarımları incelenmiştir. İkinci aşamada her bir ağın ürettiği derin özniteliklerin klasik makine öğrenmesi yöntemi olan destek vektör makineleri (Support Vector Machines – SVM) ile sınıflandırılması sonucu YST başarımlarındaki değişim incelenmiştir. Son olarak VGG16, DenseNet121 ve MobileNet ağlarının ürettikleri derin öznitelikler birleştirilerek (öznitelik seviyesinde birleştirme - feature level fusion) tüm saldırı türleri için SVM ile gerçek/sahte sınıflandırması gerçekleştirilmiştir. Yapılan deney sonuçlarına göre derin özniteliklerin ya da birleşimlerinin SVM ile sınıflandırılması saldırı türüne göre YST başarımını artırmaktadır.

Anahtar kelimeler: Derin öğrenme, Sınıflandırma, Yüz sahteciliği tespiti

Abstract

Today, the use of face recognition systems and attacks against these systems have increased. Especially with the increasing use of social media, the increase in the sharing of facial images and videos allows attackers to deceive facial recognition systems more easily. For this reason, face spoofing detection has become a very important field of study. Face spoofing attacks are carried out in various types. Generally, performance evaluations are made on scenarios in which all attack types are evaluated together. Therefore, in this study, face spoofing detection performances of deep learning methods in Printed Photo, Digital Photo and Replay Video attack types in the Replay-Attack dataset were evaluated. For this purpose, the face spoofing detection performances of VGG16, DenseNet121 and MobileNet deep network architectures were examined. Then, the change in face spoofing detection performances because of classification of deep features produced by each network with support vector machines (SVM) was examined. Finally, the deep features produced by VGG16, DenseNet121 and MobileNet networks were combined (feature level fusion) and real/fake classification was performed for all attack types with SVM. According to the results, the classification of deep features or their combinations with SVM increases the performance of face spoofing detection according to the attack type.

Keywords*: Deep learning, Classification, Face spoofing detection*

*Asuman GÜNAY YILMAZ; gunaya@ktu.edu.tr

1. Giriş
1. Introduction

Yüz tanıma sistemleri günümüzde en çok kullanılan biyometrik tanıma sistemleridir. Bu sistemlerin kullanımının yaygınlaşması beraberinde, sistemleri kandırmaya yönelik saldırılar da artırmıştır. Bu nedenle yüz tanıma sistemleri günümüzde iyi bir başarıma ulaşmış olsa da bu sistemlerin güvenilirliği sorgulanmaya başlanmıştır. Bu nedenle yüz sahteciliği tespiti (YST) alanındaki çalışmalar oldukça önem taşımaktadır. Yüz sahteciliği saldırıları standart görüntüleme cihazları üzerinden gerçekleştirilmektedir. Saldırganlar kişilerin basılmış fotoğraflarını yüz tanıma sistemine sunabileceği gibi kişilere ait videoları bir mobil cihaz ya da tablet üzerinden de kameraya gösterip sistemi kandırmaya çalışmaktadır. Daha ileri saldırı türleri ise 3D maskelerin kullanımını ya da plastik cerrahi ile kişilerin kimliklerine bürünmeyi içermektedir. Dolayısı ile atak türleri oldukça çeşitlidir. Bunun dışında YST sistemleri, ortam ışığı, görüntü kalitesi, gürültü, yüz görüntülerindeki tutarsızlık (gözleri kısma veya kapama, yüz ifadesini değiştirme) ve yaşlanmaya bağlı yüz değişiklikleri gibi çok sayıda değişkenden etkilenmektedir. Bu nedenle problemin çözümü zorlaşmaktadır.

Mevcut yüz sahteciliği tespit sistemleri 4 gruba ayrılabilir: 1) hareket analizine dayalı yöntemler, 2) canlılık tespitine dayalı yöntemler ve 3) görüntü kalitesi analizine dayalı yöntemler ve 4) doku analizine dayalı yöntemler. Hareket analizi tabanlı yöntemler video dizilerinden hesaplanan değişikliklerin hesaplanmasına dayalıdır. Bu yöntemlerin taklit edilmesi zordur ve düşük kullanıcı iş birliği gerektirir. Ancak, yüksek hareket etkinliğine sahip video dizilerine ihtiyaç duyulması ve yüksek hesaplama karmaşıklığı, bu yaklaşımların ana dezavantajlarıdır. Anjos vd. (2014), optik akış kullanarak ön plan/arka plan hareket korelasyonuna dayanan bir yöntem önermiştir. Canlılık tespitine dayalı yöntemler, videolardaki göz kırpmaları, yüz ifadeleri gibi fizyolojik yaşam belirtilerini tespit etmeye çalışmaktadır. Ancak bu yöntemler yüksek kullanıcı işbirliğine, ekstra cihazlara ve video dizilerine ihtiyaç duyar. Ayrıca zaman alıcı ve hesaplama açısından karmaşıktırlar. Alotaibi ve Mahmood (2017), derinlik bilgisi elde etmek ve sınır konumlarını korumak için doğrusal olmayan difüzyon kullanan bir yüz canlılığı tespit yöntemi önermiştir. Ardından, görüntülerin ayırt edici ve yüksek seviyeli özniteliklerini çıkarmak için, bir derin evrişim sinir ağı kullanılmıştır. Gerçek erişimlerin ve yasadışı saldırıların görüntü kalite özellikleri farklı olduğundan, görüntü kalitesi analizine dayalı yöntemler, renk çeşitliliği, bulanıklık, kenar bilgisi, kromatik moment gibi öznitelikleri kullanmaktadır. Bu yöntemlerin uygulanması kolay, hesaplama maliyetleri düşüktür ve kullanıcı iş birliğine ihtiyaç duymazlar. Fakat performansları büyük ölçüde görüntülerin kalitesine bağlıdır. Wen vd. (2015), görüntü bozulma analizi (Image Distortion Analysis-IDA) tabanlı yüz sahteciliği tespiti yöntemi önermiştir. Yöntemde yüz görüntülerinden 4 farklı IDA özelliği (aynasal yansıma, bulanıklık, renk momenti ve renk çeşitliliği) elde edilmiş ve SVM sınıflandırıcısı kullanılarak görüntü için gerçek veya sahte kararı verilmiştir. Galbally vd. (2014), gerçek ve sahte yüzleri birbirinden ayırt etmek için 25 görüntü kalitesi özelliği (ortalama kare hatası, tepe sinyali / gürültü oranı, maksimum fark, ortalama fark, vb.) kullanmıştır. Görüntüler lineer ve ikinci derece diskriminant analizi ile gerçek ya da sahte olarak sınıflandırılmıştır. Doku analizine dayalı yöntemler, sahtecilik girişimlerini belirlemek için gerçek ve sahte yüzlerin doku desenleri (baskı hataları, görüntü bulanıklığı, ...) arasındaki farkları kullanır. Bu yaklaşımların uygulanması kolaydır ve kullanıcı iş birliğine ihtiyaç duymazlar. Ancak gerçek ve sahte yüzleri ayırt etmek için iyi öznitelik vektörlerine ihtiyaç duyarlar. Bu çalışmada doku analizine dayalı bir YST yöntemi önerilmiştir.

Son yıllarda doku analizine dayalı YST alanında, elle üretilen özniteliklere (hand-crafted features) dayalı ve derin özniteliklere (deep-features) dayalı çeşitli çalışmalar yapılmaktadır. Määttä vd. (2011), yüz sahteciliği tespitinde, yüz dokusunu analiz etmek için yerel ikili örüntüler (Local Binary Patterns-LBP) kullanmıştır. Gerçek ve sahte yüzler arasındaki farkları yakalamak için çok ölçekli LBP operatörleri ile çıkarılan öznitelikler, bir destek vektör makinesi (Support Vector Machine-SVM) ile sınıflandırılmıştır. Başka bir çalışmada (Määttä vd., 2012), hem doku tabanlı (LBP, Gabor) hem de gradyan tabanlı (Gradyanların Histogramı-HoG) yüz tanımlayıcılarının yüz sahteciliği tespit performansı incelenmiştir. Agarwal vd. (2016), ayrık dalgacık dönüşümü uygulanmış görüntü dizilerinden blok tabanlı Haralick doku özniteliklerini (korelasyon, kontrast, entropi, fark varyansı, toplam ortalaması, vb.) çıkarmış ve bunları SVM ile sınıflandırarak yüz sahteciliği tespitinde kullanmıştır. Boulkenafet vd. (2016), renk dokusu analizini kullanarak bir yüz sahteciliğini önleme yöntemi önermişlerdir. Farklı renk uzaylarının (HSV ve YCbCr) parlaklık ve renklilik kanallarından doku özniteliklerini hesaplamak için farklı tanımlayıcılar (LBP, LPQ, BSIF ve SID) kullanılmıştır. Diğer bir çalışmalarında ise (Boulkenafet vd., 2017) farklı renk uzaylarından (HSV, YCbCr) (Speeded-Up Robust Features-SURF) özniteliklerini çıkarıp birleştirmiş ve öznitelik vektörlerini doğrusal sınıflandırmaya daha uygun yüksek boyutlu bir alana gömmek için Fisher vektör kodlaması

kullanmıştır. Zhao vd. (2018), dinamik öznitelikleri temsil etmek için yeni bir doku tanımlayıcısı (Volume Local Binary Count - VLBC) önermiştir. Yöntemde herhangi bir t çerçevesindeki bir merkez piksel için için, t-1, t ve t+1 çerçevelerinde, R yarıçap mesafesinde eşit olarak konumlanmış P komşu piksel birlikte kullanılarak, merkez piksele göre eşiklenmekte ve bir kod üretilmektedir. Gan vd., (2017), CASIA ve Replay-Attack gibi veri setlerindeki video saldırı veri kümeleri üzerinde 3D-CNN'i (Convolutional Neural Networks-CNN) kullanmıştır. Liu vd., (2019), baskı, yeniden oynatma, 3D maske vb. gibi 13 tür sahte saldırı türü için yeni bir Deep Tree Network (DTN) önermiştir. Einy vd. (2021), Replay Attack ve Rose-Youtu veri setlerindeki görüntüleri RGB, HSV ve YCbCr renk uzaylarına dönüştürmüş, RGB görüntüler için VGG-Face, HSV ve YCbCr görüntüler için VGG-16 modelini kullanmıştır. Korkmaz vd. (2023), DFDC veri tabanında bulunan deep fake videolarındaki oynamaları tespit etmek için EfficientNet ağını kullanmıştır.

Yüz sahteciliği yöntemleri genel olarak Basılı Fotoğraf saldırısı (printed photo), Video Oynatma (video replay) saldırısı ve 3D maske saldırısı olarak üçe ayrılır. 3D modellerin oluşturulması daha zor ve pahalı olduğu için, saldırganlar ilk iki yöntemi daha kolay gerçekleştirebilir. Video Oynatma saldırıları kişinin fotoğrafının ya da videosunun telefon/tablet gibi cihazlardan kameraya tutularak sistemin kandırılmaya çalışılmasını içermektedir. Bunun dışında farklı çözünürlük, aydınlatma koşulları, giriş cihazı türü ve konumu gibi etkenler de saldırı senaryolarını çeşitlendirmektedir. Dolayısıyla, YST sisteminin gerçek zamanlı kullanımında ne tür bir senaryo ile karşılaşacağı tam olarak bilenememektedir. Literatürde yaygın kullanılan veri kümeleri birçok saldırı senaryosu içermektedir. Ancak bunları kullanan çoğu çalışmada genellikle tek bir saldırı senaryosu için (tüm saldırı türlerinin bir arada olduğu senaryo) sonuçlarının paylaşıldığı görülmektedir. Bu durum, önerilen modelin farklı senaryolar altındaki davranışını kestirmeyi ve genelleştirilmiş bir çözüm üretmeyi zorlaştırmaktadır. Diğer yandan derin ağların YST başarımları çeşitli çalışmalarda incelense de bu ağlardan üretilen derin özniteliklerin birleştirilmesinin YST başarımına etkisi incelenmemiştir. Bu nedenlerle bu çalışmada üç farklı saldırı türü için üç farklı CNN mimarisinin YST başarımı değerlendirilmiştir. İkinci aşamada CNN ağlarından üretilen derin öznitelikler SVM ile sınıflandırılmış ve herhangi bir performans artışı olup olmadığı gözlemlenmiştir. Son olarak üç farklı ağın ürettiği derin özniteliklerin birleşiminin YST başarımına etkisi incelenmiştir.

2. Materyal ve metot
2. Material and method

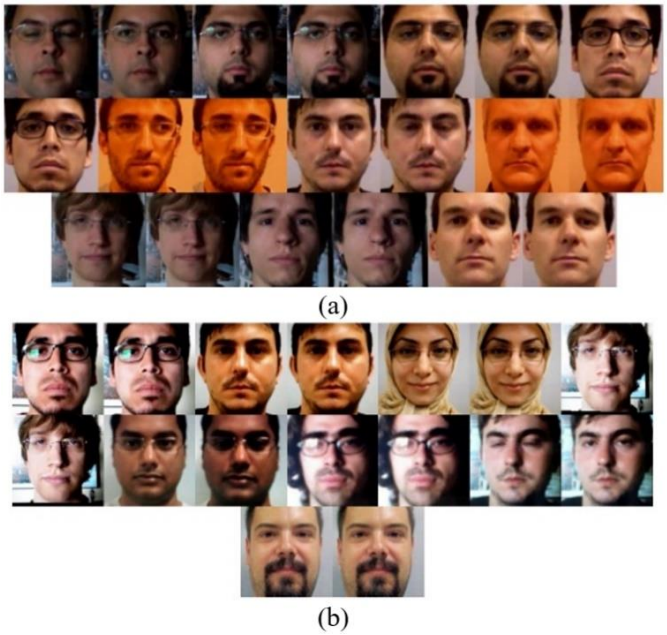
2.1. Veri seti
2.1.Dataset

Deneylerde yüz sahteciliği saldırı türlerini barındıran Replay-Attack (Chingovska vd., 2012) veri seti kullanılmıştır. Replay-Attack veri seti 50 kişiye ait gerçek ve sahte erişim videolarından oluşmaktadır. Videolar MacBook Air 13" kamerası ile iki farklı aydınlatma koşulunda elde edilmiştir. Kontrollü ortamda florasan lambalarıyla ışıklandırma yapılmış ve tek bir arka plan kullanılmıştır. Kontrolsüz ortamda ise videolar gün ışığında farklı arka planlarda çekilmiştir. Yüksek çözünürlüklü videolar iPhone 3GS ve Canon PowerShot SX150 IS cihazları ile elde edilmiştir. Veri setinde üç tür atak bulunmaktadır. Bunlar Basılı Fotoğraf, Mobil ve Yüksek Çözünürlük ataklarıdır. Görüntülerin kameraya elle ya da sabit bir yerde tutularak gösterilmesi açısından da iki atak türü bulunmaktadır.

Bu çalışmada Replay-Attack veri setindeki Basılı Fotoğraf (fotoğrafın yüksek kalite basılarak kameraya sunulması), Dijital Fotoğraf (Fotoğrafın dijital bir cihazdan gösterilmesi) ve Video Oynatma saldırıları için YST başarımları incelenmiştir. Burada bu üç saldırı türü, içerisinde, yüksek/normal çözünürlüklü görüntüleri, kontrollü/kontrolsüz ortamda üretilen görüntüleri ve sabit bir yerde tutularak kameraya sunulan görüntüleri içermektedir. Veri kümesinde her biri ortalama 10 snlik çok sayıda video bulunmaktadır. Ağların ezberlemesini engellemek ve Google Colab ortamının sınırlı kaynaklarını kullanabilmek adına videolardan 125ms lik aralıklarla (sn de 8 çerçeve) görüntüler elde edilmiştir. Bu görüntülerden yüz bölgesi düzgün bir biçimde elde edilenler ayrılmıştır. Daha sonra 50 kişiden 20 kişi rasgele olarak eğitim kümesi, 10 kişi doğrulama kümesi, 10 kişi ise test kümesi olarak seçilmiştir. Buna göre eğitim, doğrulama ve test kümelerinde kullanılan görüntü sayıları Tablo 1'de verilmiştir. Eğitim kümesinden seçilen gerçek ve sahte görüntü örnekleri ise Şekil 1'de görülmektedir.

Tablo 1. Eğitim, doğrulama ve test kümelerindeki örnek sayıları
Table 1. *Number of samples in training, validation and test sets*

	Eğitim		Doğrulama		Test	
Saldırı türü	**Gerçek**	**Sahte**	**Gerçek**	**Sahte**	**Gerçek**	**Sahte**
Basılı fotoğraf	1200	800	600	342	600	400
Dijital fotoğraf	1185	800	600	400	600	400
Video oynatma	1185	797	600	384	597	379



Şekil 1. Replay-Attack veri setinden a) gerçek, b) saldırı sınıfından örnek görüntüler (Chingovska vd., 2012)
Figure 1. *Sample images from a) real, b) attack classes of the Replay-Attack dataset (Chingovska vd., 2012)*

2.2. Ön işleme
2.2. Preprocessing

Çalışmada öncelikle veri setinden alınan görüntülerde yüz bölgesinin belirlenmesi işlemi gerçekleştirilmiştir. Bu amaçla giriş görüntülerinde Dlib (King, 2009) kütüphanesinin ön eğitimli 5 nokta işaretleyicisi (gözlerin iç dış kenarları (4 nokta) ve burun-ağız arası (1 nokta)) kullanılarak yüz bölgeleri belirlenmiş ve göz pozisyonlarına göre normalize edilmiştir. Noktalar belirlendikten sonra noktalar arasındaki açıya göre görüntü döndürülmüş ve daha sonra görüntü kesilerek 128×128 boyutlarına ölçeklenmiştir. Böylece tüm görüntüler hizalanmış ve aynı boyuta sahip olmuştur.

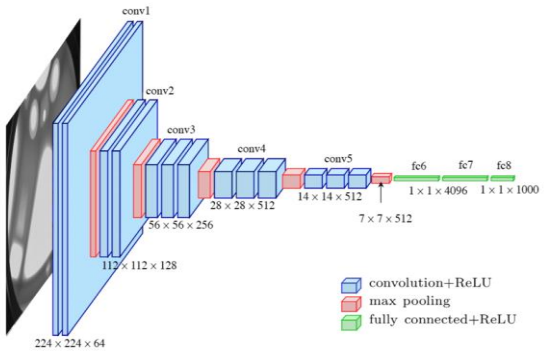
2.3. Evrişimsel sinir ağları
2.3. Convolutional neural networks

Evrişimsel sinir ağları (Convolutional Neural Networks-CNN), çok katmanlı derin öğrenme mimarilerinden biridir. Bu ağlar, evrişim katmanlarında görüntülere uygulanan filtreler ile görüntülerden öznitelik çıkarım işlemini gerçekleştirmektedir. Evrişim katmanının sayısı arttıkça görüntüye ait spesifik öznitelikler daha belirgin hale gelir. Evrişim katmanları ile üretilen öznitelik matrisleri, vektörel forma dönüştürülür ve tam bağlantılı katmana iletilir. Tam bağlantılı katmanda sınıflandırma işlemi gerçekleştirilir. Bu çalışmada CNN modellerinden VGG16, DenseNet121 ve MobileNet kullanılmıştır.

2.3.1. VGG16
2.3.1. VGG16

VGG16, Oxford Üniversitesi'nde Karen Simonyan ve Andrew Zisserman tarafından önerilen bir CNN modelidir. Bu model tek piksel adımlı 3×3'lük filtreler kullanmakta ve 13 evrişim ve 3 tam bağlı katman olmak

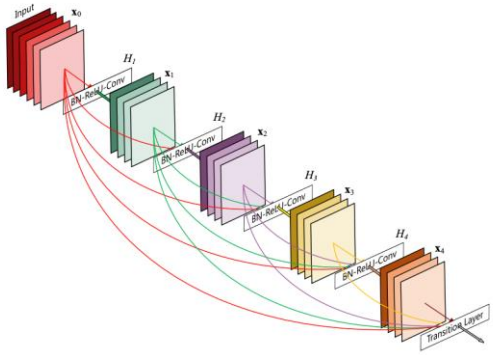
üzere toplam 16 katmandan oluşmaktadır. VGG16 mimarisi, genel olarak 224×224 boyutlu görüntüleri giriş olarak alır. Görüntüler ilk yığında bulunan 1 adımlı 3×3'lük 2 evrişim katmanından geçirilir ve hemen ardından ReLU aktivasyon fonksiyonu devreye girer. İki evrişim katmanı da 64 filtre içerir. Daha sonra görüntü 2 adımlı 2×2 'lik bir maksimum havuzlama katmanından geçirilir. Bu işlemlerin sonunda ilk katmandan çıkan öznitelik 112×112×64 boyutlarına sahip olacaktır. Aynı şekilde filtre sayısı iki katına çıkarılarak devam eder ve en sonunda 7×7×512 boyutuna sahip görüntü öznitelikleri elde edilir. VGG16 ağ mimarisi Şekil 2'de görülmektedir.



Şekil 2. VGG16 ağ mimarisi (Le, 2021)
Figure 2*. VGG16 architecture (Le, 2021)*

2.3.2. DenseNet121
2.3.2. DenseNet121

DenseNet121, Gluon CV tarafından ortaya çıkarılmış bir CNN modelidir. DenseNet ailesi içinde yer aldığından dolayı bağlantı ağırlıklarının birleştirilmesi (concatenation) prensibi üzerine inşa edilmiştir. DenseNet121 modelinin mimarisi, giriş katmanı, yoğun bloğu ve geçiş katmanından oluşmaktadır. Giriş katmanı girdi görüntüsünün öznitelik haritalarına çevrilmesini sağlar. Her bir yoğun blok, bir dizi katmanı içerir ve her bir katman, önceki katmanların öznitelik haritalarının birleştirilmesi ile oluşur. Bu bağlantılar, ağırlık paylaşımını ve veri akışını optimize etmeyi amaçlar. Her bir yoğun bloğun sonunda, geçiş katmanı bulunur. Bu katman, yoğun bloğun son katmanından gelen öznitelik haritalarının boyutlarını azaltmak ve veri akışını optimize etmek için kullanılır. Bu tasarım sayesinde, veri yoğunluğunu ve öznitelik akışını arttırır. Bu durum daha iyi performans ve daha az overfitting (aşırı uyum) riski gibi avantajları beraberinde getirir. DenseNet121 ağ mimarisi Şekil 3'te görülmektedir.

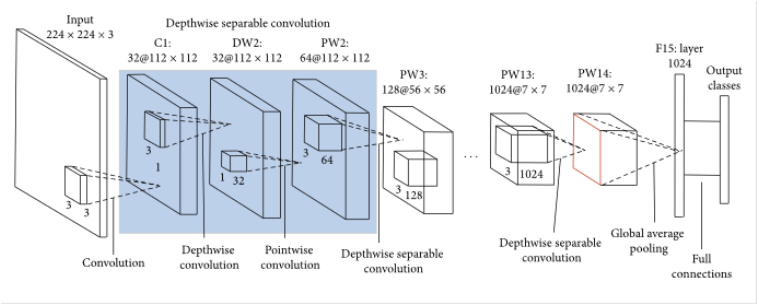


Şekil 3. DenseNet121 ağ mimarisi (Sarkar, 2020)
Figure 3. *DenseNet121 architecture (Sarkar, 2020)*

2.3.3. MobileNet
2.3.3. MobileNet

MobileNet, mobil cihazlarda yüksek performansla görüntü tanıma veya nesne tanıma gibi uygulamaları desteklemek amacıyla tasarlanmış bir CNN modelidir. Ağırlıklardan ve katmanlardan tasarruf yaparak, bellek ve işlem gereksinimlerini azaltır ve bu nedenle mobil cihazlarda etkin bir şekilde çalışmasına olanak tanır.

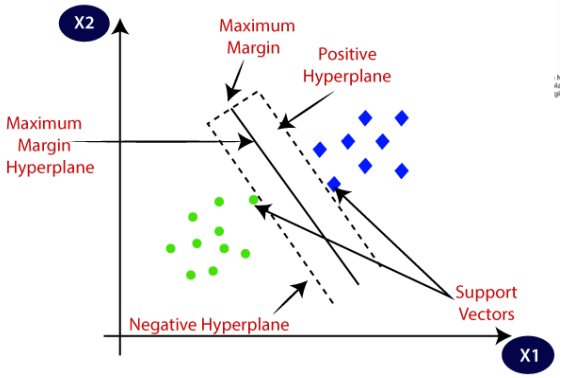
MobileNet mimarisi, klasik evrişim katmanları yerine mobil katmanlardan oluşur. Mobil katmanlar, derinlemesine ayrılabilir evrişim ve noktasal evrişim işlemlerini gerçekleştirir. Derinlemesine ayrılabilir evrişim, her kanal için ayrı ayrı bir filtre uygular ve bu filtrelerin sonuçlarını toplar. Noktasal evrişim ise, tüm kanallar üzerinde aynı filtreyi uygular. MobileNet ağ mimarisi Şekil 4'te görülmektedir.



Şekil 4. MobileNet ağ mimarisi (Singhal, 2020)
Figure 4. *MobileNet architecture (Singhal, 2020)*

2.4. Destek vektör makineleri
2.4. Support vector machines

SVM, bir sınıflandırma ve regresyon yöntemidir. Yöntemde veri noktaları arasındaki en büyük marjı bulan dolayısı ile veri noktalarını en iyi şekilde ayrıştıran hiper düzlem belirlenir (Şekil 5). SVM, bazı veri noktalarını destek vektörleri olarak tanımlar. Destek vektörleri, hiper düzlem ile en yakın olan veri noktalarıdır ve hiper düzlemin belirlenmesinde önemli rol oynarlar. Ayrıca SVM, doğrusal ve doğrusal olmayan veri kümelerinde kullanılabilir. Doğrusal veri kümelerinde, hiper düzlem doğrusal bir şekilde belirlenir ve sınıflandırma yapılır. Doğrusal olmayan veri kümelerinde ise, verilerin daha iyi ayrışması için önce veriler dönüştürülür, sonrasında hiper düzlem belirlenir.



Şekil 5. Destek vektörleri
Figure 5. *Support vectors*

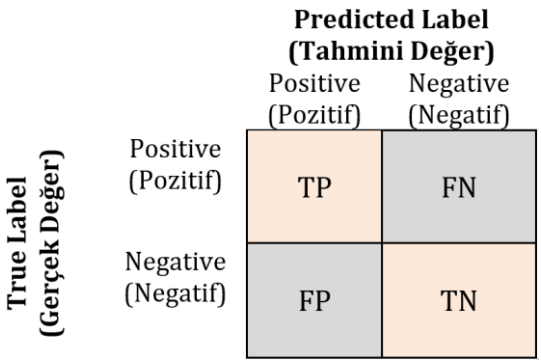
3. Bulgular
3. Results

Bu çalışmada VGG16, DenseNet121 ve MobileNet CNN mimarileri kullanılarak Replay-Attack veri setindeki Basılı Fotoğraf (Print Photo), Dijital Fotoğraf (Digital Photo) ve Video Oynatma (Replay Video) saldırıları için YST gerçekleştirilmiştir. Bu amaçla öncelikle her bir saldırı senaryosu için VGG16, DenseNet121 ve MobileNet ağlarının gerçek/sahte sınıflandırma başarımları incelenmiştir. İkinci aşamada CNN ağlarının ürettiği derin öznitelikler SVM ile sınıflandırılarak saldırı tespiti yapılmıştır. Son aşamada ise CNN ağlarının ürettiği derin öznitelikler birleştirilmiş ve SVM ile sınıflandırma gerçekleştirilmiştir. Deneysel bulgular alt bölümlerde verilmiştir.

Çalışmada ağların eğitiminde batch-size 128, öğrenme katsayısı 0.0001, ve Adam optimizer kullanılmıştır. Kayıp fonksiyonu ise categorical cross entropy'dir. Uygulamalar Google Colab ortamında çalıştırılmıştır.

Transfer öğrenme yöntemi ile ImageNet veriseti ile ön eğitilmiş CNN ağları YST için 25 epoch boyunca tekrar eğitilmiştir.

Performans ölçütü olarak Doğruluk (Accuracy), Kesinlik (Precision), Duyarlılık (Recall) ve F1-skoru kullanılmıştır. Bu ölçütlerin hesaplanmasında ise karmaşıklık matrisi kullanılmıştır. Karmaşıklık matrisi Şekil 6'da görülmektedir. Bu matris sayesinde gerçek değerler ile tahmin edilen değerlerin kıyaslanması mümkün olmaktadır. Matriste TP (True Positive) doğru sınıflandırılan pozitif örnek sayısını, TN (True Negative) doğru sınıflandırılan negatif örnek sayısını, FP (False Positive) ve FN (False Negative) ise sırasıyla yanlış sınıflandırılan negatif ve pozitif örnek sayılarını ifade etmektedir. Bu çalışmada Pozitif gerçek görüntüleri, Negatif ise sahte görüntüleri temsil etmektedir.



Şekil 6. Karmaşıklık matrisi
Figure 6*. Confusion matrix*

Bu değerler çalışmada Doğruluk, Kesinlik, Duyarlılık ve F1-skoru ölçütlerinin hesaplanmasında aşağıdaki şekilde kullanılmıştır. Doğruluk bir modelin başarısını ölçmek için kullanılan ancak tek başına yeterli olmadığı görülen bir metriktir. Doğruluk değeri modelde doğru tahmin edilen örnek sayısının toplam örnek sayısına oranı ile hesaplanmaktır. Kesinlik doğru sınıflandırılan verilerin oranı iken Duyarlılık sadece pozitif değerlerden doğru sınıflandırılanların oranıdır. F1-skoru ise Kesinlik ve Duyarlılık değerlerinin harmonik ortalaması olup Denklem 4'e göre hesaplanmaktadır.

$$Doğruluk = \frac{TP+TN}{TP+FP+TN+FN} \quad (1)$$

$$Kesinlik = TP/\ (TP + FP) \quad (2)$$

$$Duyarlılık = TP/\ (TP + FN) \quad (3)$$

$$F1\_skor = \frac{2*Kesinlik*Duyarlılık}{Kesinlik+Duyarlılık} \quad (4)$$

3.1. VGG16, DenseNet121 ve MobileNet ağları ile YST
3.1. Face spoofing detection with VGG16, DenseNet121 and MobileNet

3.1.1. Basılı fotoğraf saldırısı için YST başarımları
3.1.1. Face spoofing detection performances for printed photo attack

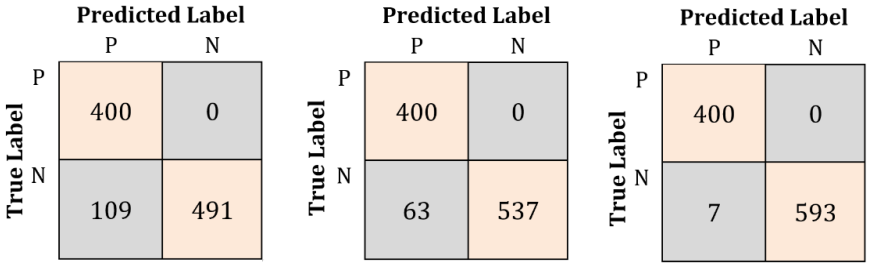
Basılı Fotoğraf saldırısı senaryosunda VGG16, DenseNet121 ve MobileNet ağlarıyla elde edilen başarım değerleri Tablo 2'de verilmiştir. Tablodan görüldüğü gibi bu saldırı senaryosunda MobileNet ağı ile %99.30 doğruluk ve %99.13 F1-skor değeri ile en iyi başarım değeri elde edilmiştir. MobileNet ağı ile elde edilen Kesinlik ve Duyarlılık değerleri sırasıyla %98.28 ve %100 şeklindedir. Bu atak türünde DenseNet121 ağı ile %93.63, VGG16 ağı ile %89.10 doğruluk sonucuna ulaşılmıştır. Eğitilen ağların test kümesindeki örnekler için yaptığı tahmin değerlerine göre oluşturulan karmaşıklık matrisleri ise Şekil 7'de görülmektedir. Karmaşıklık matrisleri incelendiğinde MobileNet ağı ile 400 gerçek 600 basılı fotoğraf görüntüden oluşan test kümesinde, gerçek görüntülerin hepsinin, sahte görüntülerin 593'ünün doğru sınıflandırıldığı görülmektedir.

Ayrıca genel olarak tüm ağlar test kümesindeki gerçek görüntüleri sınıflandırmada başarılı iken, VGG16 ve DenseNet121 ağları sahte görüntülerin sınıflandırılmasında MobileNet ağından daha başarısızdır.

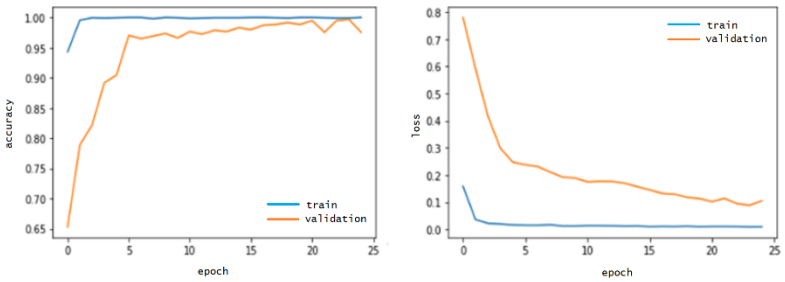
Tablo 2. Basılı fotoğraf saldırısı için VGG16, Dense Net121 ve MobileNet ağlarının başarımları (%)
***Table 2**. Performances of VGG16, Dense Net121 ve MobileNet networks for printed photo attack (%)*

Yöntem	Doğruluk	Kesinlik	Duyarlılık	F1-skoru
VGG16	89.10	78.58	100	88.00
DenseNet121	93.69	86.39	100	92.69
MobileNet	**99.30**	**98.28**	**100**	**99.13**

Basılı Fotoğraf saldırısı için en başarılı sonucu veren MobileNet ağından, eğitim (train) ve doğrulama (validation) setleri için elde edilen doğruluk ve kayıp grafikleri Şekil 8'de verilmiştir. Şekilden görüldüğü gibi MobileNet ağı bu saldırı türünün belirlenmesinde kararlı bir yapıya sahiptir ve epochlar ilerledikçe yakınsama sağlanmıştır. Kayıp değeri ise başlangıç epochlarında düşmeye başlamış ve eğitimin sonuna doğru belli bir seviyede kalmıştır.



Şekil 7. Basılı fotoğraf saldırısı için karmaşıklık matrisleri (sırasıyla: VGG16, DenseNet121, MobileNet)
***Figure 7.** Confusion matrixs of printed photo attack (VGG16, DenseNet12, MobileNet)*



Şekil 8. Basılı fotoğraf saldırı senaryosu için MobileNet ile elde edilen doğruluk ve kayıp grafikleri
***Figure 8.** Accuracy and loss graphics of printed photo attack with MobileNet*

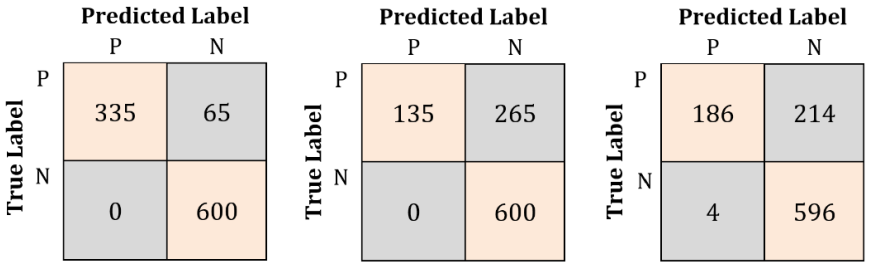
3.1.2. Dijital fotoğraf saldırısı için YST başarımları
3.1.2. Face spoofing detection performances for digital photo attack

Dijital Fotoğraf saldırısı senaryosunda VGG16, DenseNet121 ve MobileNet ağlarıyla elde edilen başarım değerleri Tablo 3'te verilmiştir. Tablodan görüldüğü gibi bu saldırı senaryosunda VGG16 ağı ile %93.5 doğruluk ve %91.15 F1-skor değeri ile en iyi başarım değeri elde edilmiştir. VGG16 ağı ile elde edilen Kesinlik ve Duyarlılık değerleri sırasıyla %100 ve %83.75 şeklindedir. Bu atak türünde DenseNet121 ağı ile %73.50, MobileNet ağı ile %78.20 doğruluk sonucuna ulaşılmıştır.

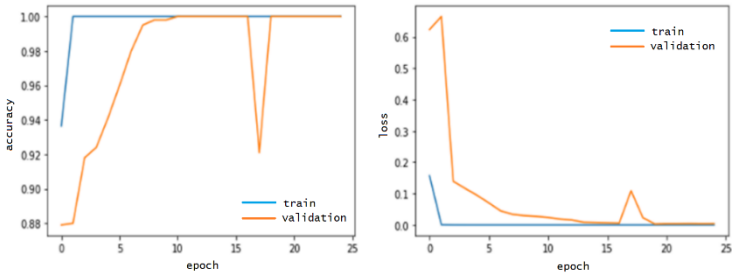
Tablo 3. Dijital fotoğraf saldırısı için VGG16, Dense Net121 ve MobileNet ağlarının başarımları (%)
***Table 3**. Performances of VGG16, Dense Net121 ve MobileNet networks for digital photo attack (%)*

Yöntem	Doğruluk	Kesinlik	Duyarlılık	F1-skoru
VGG-16	**93.50**	**100**	**83.75**	**91.15**
DenseNet121	73.50	100	33.75	50.46
MobileNet	78.20	97.89	46.50	63.05

Eğitilen ağların test kümesindeki örnekler için yaptığı tahmin değerlerine göre oluşturulan karmaşıklık matrisleri ise Şekil 9'da görülmektedir. Karmaşıklık matrisleri incelendiğinde bu saldırı türünde VGG16 ağı ile 400 gerçek 600 basılı fotoğraf görüntüden oluşan test kümesinde, gerçek görüntülerin 335'inin, sahte görüntülerin hepsinin doğru sınıflandırıldığı görülmektedir. Ayrıca genel olarak tüm ağlar test kümesindeki sahte görüntüleri sınıflandırmada daha başarılı iken, DenseNet121 ve MobileNet ağları gerçek görüntülerin sınıflandırılmasında VGG16 ağına göre daha başarısızdır. Dijital Fotoğraf saldırısı için en başarılı sonucu veren VGG16 ağından, eğitim (train) ve doğrulama (validation) setleri için elde edilen doğruluk ve kayıp grafikleri Şekil 10'da verilmiştir. Şekilden görüldüğü gibi VGG16 ağı bu saldırı türünün belirlenmesinde kararlı bir yapıya sahiptir ve epochlar ilerledikçe yakınsama sağlanmıştır. Kayıp değeri ise başlangıç epochlarında düşmeye başlamış, ara bir adımda yükselmiş olsa da eğitimin sonuna doğru belli bir seviyede kalmıştır.



Şekil 9. Dijital fotoğraf saldırısı için karmaşıklık matrisleri (sırasıyla: VGG16, DenseNet121, MobileNet)
Figure 9. *Confusion matrixs of digital photo attack (VGG16, DenseNet12, MobileNet)*



Şekil 10. Dijital fotoğraf saldırı senaryosu için VGG16 ile elde edilen doğruluk ve kayıp grafikleri
Figure 10. *Accuracy and loss graphics of digital photo attack with VGG16*

3.1.3. Video oynatma saldırısı için YST başarımları
3.1.3. Face spoofing detection performances for video replay attack

Video Oynatma saldırısı senaryosunda VGG16, DenseNet121 ve MobileNet ağlarıyla elde edilen başarım değerleri Tablo 4'te verilmiştir. Tablodan görüldüğü gibi bu saldırı senaryosunda VGG16 ağı ile %88.01 doğruluk ve %86.62 F1-skor değeri ile en iyi başarım değeri elde edilmiştir. VGG16 ağı ile elde edilen Kesinlik ve Duyarlılık değerleri sırasıyla %76.41 ve %100 şeklindedir. Bu atak türünde DenseNet121 ağı ile %84.52, MobileNet ağı ile %62.90 doğruluk sonucuna ulaşılmıştır.

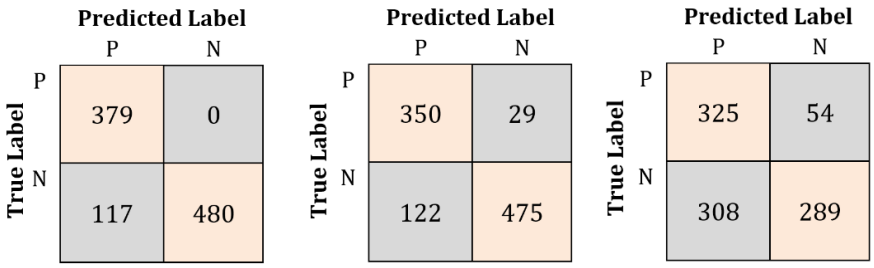
Tablo 4. Video oynatma saldırısı için VGG16, Dense Net121 ve MobileNet ağlarının başarımları (%)
Table 4*. Performances of VGG16, Dense Net121 ve MobileNet networks for video replay attack (%)*

Yöntem	Doğruluk	Kesinlik	Duyarlılık	F1-skoru
VGG-16	**88.01**	**76.41**	**100**	**86.62**
DenseNet121	84.52	74.15	92.34	82.25
MobileNet	62.90	51.34	85.75	64.22

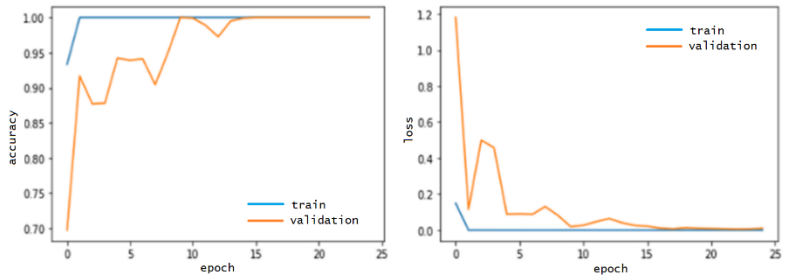
Eğitilen ağların test kümesindeki örnekler için yaptığı tahmin değerlerine göre oluşturulan karmaşıklık matrisleri ise Şekil 11'de görülmektedir. Karmaşıklık matrisleri incelendiğinde bu saldırı türünde VGG16 ağının diğer ağlara göre gerçek/sahte sınıflandırmada daha başarılı olduğu anlaşılmaktadır. VGG16 ağı ile test

kümesindeki 379 gerçek görüntünün hepsi doğru sınıflandırılmışken, 597 sahte görüntünün 480'i doğru sınıflandırılmıştır. Bu saldırı türünde genel olarak ağlar gerçek görüntülerin sınıflandırılmasında daha başarılı olmaktadır. DenseNet121 ve MobileNet ağları sahte görüntülerin sınıflandırılmasında VGG16 ya kıyasla daha başarısızdır.

Video Oynatma saldırısı için en başarılı sonucu veren VGG16 ağı ile gerçek/sahte sınıflandırması sonucunda eğitim (train) ve doğrulama (validation) setleri için elde edilen doğruluk ve kayıp grafikleri Şekil 12'de verilmiştir. Şekilden görüldüğü gibi VGG16 ağı bu saldırı türünün belirlenmesinde kararlı bir yapıya sahiptir ve epochlar ilerledikçe yakınsama sağlanmıştır.



Şekil 11. Video oynatma saldırısı için karmaşıklık matrisleri (sırasıyla: VGG16, DenseNet121, MobileNet)
Figure 11. *Confusion matrixs of video replay attack (VGG16, DenseNet12, MobileNet)*



Şekil 12. Video oynatma saldırı senaryosu için VGG16 ile elde edilen doğruluk ve kayıp grafikleri
Figure 12. *Accuracy and loss graphics of video replay attack with VGG16*

3.2. VGG16, DenseNet121 ve MobileNet ağları ile çıkarılan derin özniteliklerin SVM ile sınıflandırılması ile YST

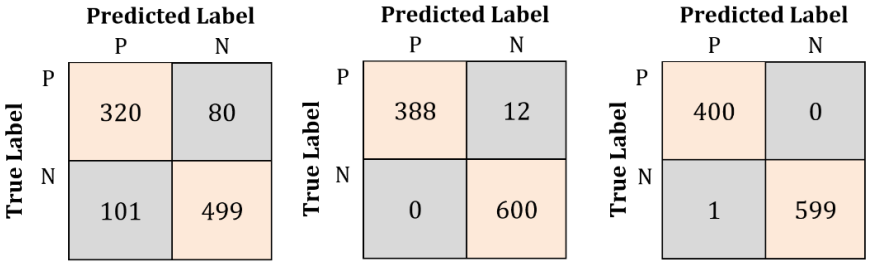
3.2. Face spoofing detection with SVM classification of deep features extracted using VGG16, DenseNet121 and MobileNet networks

Çalışmanın ikinci bölümünde VGG16, DenseNet121 ve MobileNet ağlarının eğitimi sonucunda üretilen derin öznitelikler alınarak SVM ile sınıflandırılmış ve YST başarımı incelenmiştir. Yine Replay-Attack veri kümesindeki Basılı Fotoğraf, Dijital Fotoğraf ve Video Oynatma saldırı senaryoları için derin özniteliklerin başarımları incelenmiştir. Buna göre Basılı Fotoğraf saldırısı için VGG16+SVM, DenseNet121+SVM ve MobileNet+SVM yöntemlerinin karmaşıklık matrisleri Şekil 13'te başarım sonuçları ise Tablo 5'te görülmektedir. Tablo 5'teki başarım sonuçlarından DenseNet121+SVM ve MobileNet+SVM yöntemlerinin ağların tek başına sınıflandırma başarımını sırasıyla %93.69'dan %98.8'e ve %99.30'dan %99.90'a yükselttiği görülmektedir. Bu iki ağ için derin özniteliklerin SVM ile sınıflandırılması gerçek/sahte sınıflandırma başarımını artırmıştır. VGG16 ağı ise tek başına VGG16+SVM yönteminden %7.2 daha başarılıdır.

Tablo 5. Basılı fotoğraf saldırısı için VGG16+SVM, DenseNet121+SVM ve MobileNet+ SVM yöntemlerinin başarım değerleri
Table 5. *Performances of VGG16+SVM, DenseNet12+SVM and MobileNet+SVM methods for printed photo attack*

Yöntem	Doğruluk	Kesinlik	Duyarlılık	F1-skoru
VGG-16+SVM	81.90	76.00	80	77.94
DenseNet121+SVM	98.80	100	97	98.47
MobileNet+SVM	**99.90**	**99.75**	**100**	**99.87**

Şekil 13 incelendiğinde ise basılı fotoğraf saldırısında DenseNet121+SVM ve MobileNet+SVM yöntemlerinin sahte görüntülerin sınıflandırılmasında VGG16+SVM yöntemine göre daha başarılı olduğu görülmektedir. Gerçek görüntülerin sınıflandırılmasında ise MobileNet+SVM yöntemi diğer iki modele göre daha başarılıdır.

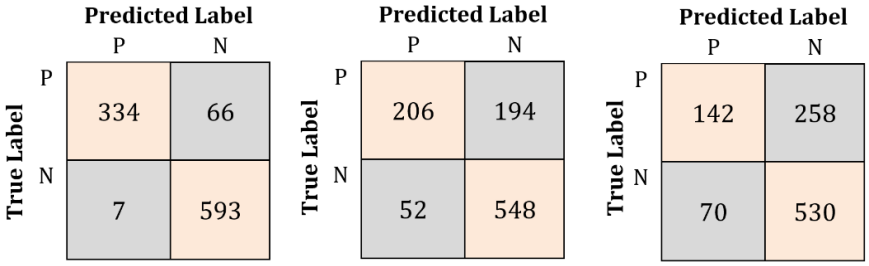


Şekil 13. Basılı fotoğraf saldırısı için VGG16+SVM, DenseNet121+SVM ve MobileNet+ SVM yöntemlerinin karmaşıklık matrisleri
Figure 13. *Confusion matrixs of printed photo attack with VGG16+SVM, DenseNet12+SVM and MobileNet+SVM*

Dijital Fotoğraf saldırısı için VGG16+SVM, DenseNet121+SVM ve MobileNet+SVM yöntemlerinin karmaşıklık matrisleri Şekil 14'te başarım sonuçları ise Tablo 6'da görülmektedir. Şekil 14 incelendiğinde dijital fotoğraf saldırısında VGG16+SVM yönteminin gerçek ve sahte görüntülerin sınıflandırılmasında DenseNet121+SVM ve MobileNet+SVM yöntemlerine göre daha başarılı olduğu görülmektedir. Genel olarak üç yöntem sahte görüntü sınıflandırmada gerçek görüntülere göre daha başarılıdır. Tablo 6'daki başarım sonuçlarından ise sadece DenseNet121+SVM yönteminin, ağın tek başına sınıflandırma başarımını %73.50'den %75.40'a yükselttiği görülmektedir. Bu ağ için derin özniteliklerin SVM ile sınıflandırılması gerçek/sahte sınıflandırma başarımını artırmıştır. VGG16 ve MobileNet ağları ise, VGG16+SVM ve MobileNet+SVM yöntemlerinden sırasıyla %0.8 ve %11 daha başarılıdır. Bu saldırı türünde en iyi başarım %92.70 ile VGG16+SVM yöntemi ile elde edilmiştir.

Tablo 6. Dijital fotoğraf saldırısı için VGG16+SVM, DenseNet121+SVM ve MobileNet+ SVM yöntemlerinin başarım değerleri (%)
Table 6. *Performances of VGG16+SVM, DenseNet12+SVM and MobileNet+SVM methods for digital photo attack (%)*

Yöntem	Doğruluk	Kesinlik	Duyarlılık	F1-skoru
VGG-16+SVM	**92.70**	**97.94**	**83.50**	**90.14**
DenseNet121+SVM	75.40	79.84	51.50	62.61
MobileNet+SVM	67.20	66.98	35.50	46.40



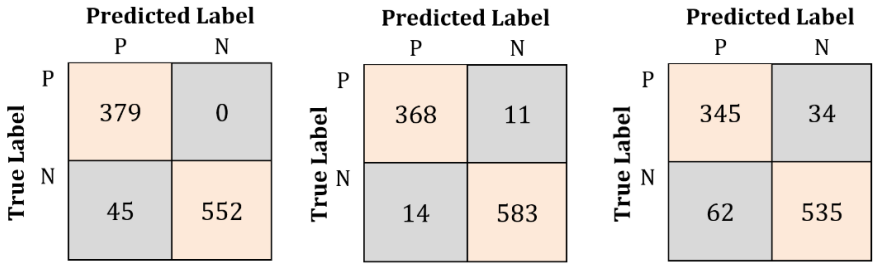
Şekil 14. Dijital fotoğraf saldırısı için VGG16+SVM, DenseNet121+SVM ve MobileNet+ SVM yöntemlerinin karmaşıklık matrisleri
Figure 14. *Confusion matrixs of digital photo attack with VGG16+SVM, DenseNet12+SVM and MobileNet+SVM*

Video Oynatma saldırısı için VGG16+SVM, DenseNet121+SVM ve MobileNet+SVM yöntemlerinin karmaşıklık matrisleri Şekil 15'te başarım sonuçları Tablo 7'de görülmektedir. Şekil 15 incelendiğinde video oynatma saldırısında DenseNet121+SVM yönteminin gerçek ve sahte görüntülerin sınıflandırılmasında VGG16121+SVM ve MobileNet+SVM yöntemlerine göre daha başarılı olduğu görülmektedir. Genel olarak üç yöntem gerçek görüntü sınıflandırmada sahte görüntü sınıflandırmaya göre daha başarılıdır. Tablo 7'deki

başarım sonuçlarından ise her üç yöntemin de ağların tek başına sınıflandırma başarımlarını artırdığı görülmektedir. VGG16+SVM yöntemi ile %95.38, DenseNet121+SVM yöntemi ile %97.43, MobileNet+SVM yöntemi ile ise %90.16 doğruluk değeri elde edilmiştir. Bu saldırı türünde derin özniteliklerin SVM ile sınıflandırılması gerçek/sahte sınıflandırma başarımını artırmıştır. En fazla başarım artışı %27.26 ile MobileNet+SVM yöntemi ile elde edilmiştir.

Tablo 7. Video oynatma saldırısı için VGG16+ SVM, DenseNet121+SVM ve MobileNet+ SVM yöntemlerinin başarım değerleri (%)
Table 7. *Performances of VGG16+SVM, DenseNet12+SVM and MobileNet+SVM methods for video replay attack (%)*

Yöntem	Doğruluk	Kesinlik	Duyarlılık	F1-skoru
VGG-16+SVM	95.38	89.38	100	94.39
DenseNet121+SVM	**97.43**	**96.33**	**97.09**	**96.70**
MobileNet+SVM	90.16	84.76	91.02	87.77



Şekil. 15. Video oynatma saldırısı için VGG16+ SVM, DenseNet121+SVM ve MobileNet+ SVM yöntemlerinin karmaşıklık matrisleri
Figure 15. *Confusion matrixs of video replay attack with VGG16+SVM, DenseNet12+SVM and MobileNet+SVM*

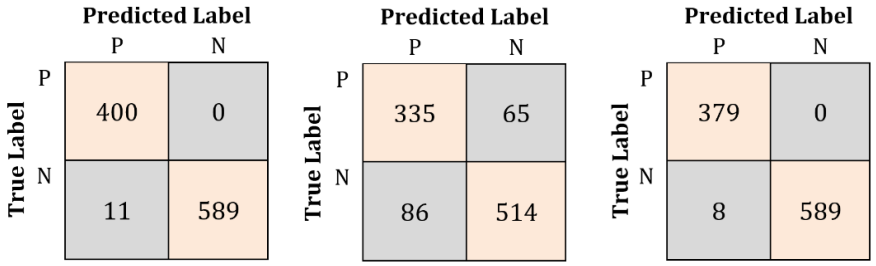
3.3. VGG16, DenseNet121 ve MobileNet ağları ile çıkarılan derin özniteliklerin birleştirilmesi ve SVM ile sınıflandırılması ile YST

3.3. Face spoofing detection with SVM classification of deep feature fusion extracted with VGG16, DenseNet121 and MobileNet networks

Çalışmanın son bölümünde ise VGG16, DenseNet121 ve MobileNet ağlarında üretilen derin öznitelikler birleştirilmiş, ve SVM yöntemi ile gerçek/sahte olarak sınıflandırılmıştır. Replay-Attack veri kümesindeki Basılı Fotoğraf, Dijital Fotoğraf ve Video Oynatma saldırı senaryoları için ayrı ayrı birleştirilmiş derin özniteliklerin başarımları incelenmiştir. Elde edilen sonuçlar karmaşıklık matrisleri Şekil 16'da başarım değerleri ise Tablo 8'de görülmektedir. Bu model ile en yüksek doğruluk değeri Video Oynatma saldırı senaryosu için %99.18 olarak elde edilmiştir. Basılı Fotoğraf ve Dijital Fotoğraf saldırı türleri için sırasıyla %98.9 ve %84.9 doğruluk sonucuna ulaşılmıştır. Bu yöntemde her üç saldırı türü için gerçek görüntüler sahte görüntülere göre daha başarılı şekilde sınıflandırılmaktadır.

Tablo 8. VGG16+DenseNet121+MobileNet+ SVM modeli ile basılı fotoğraf, dijital fotoğraf ve video oynatma saldırılarının YST başarımları (%)
Table 8. *Face spoofing detection performances of VGG16+ DenseNet121+MobileNet+SVM model for printed photo, digital photo, video replay attacks (%)*

Yöntem	Doğruluk	Kesinlik	Duyarlılık	F1-skoru
Basılı fotoğraf	98.90	97.32	100	98.64
Dijital fotoğraf	84.90	79.57	83.75	81.60
Video oynatma	99.18	97.93	100	98.95



Şekil 16. Saldırı senaryoları için VGG16+ DenseNet121+MobileNet+SVM modelinin karmaşıklık matrisleri (sırasıyla: basılı fotoğraf, dijital fotoğraf, video oynatma)
Figure 16. *Confusion matrixs of VGG16+ DenseNet121+MobileNet+SVM model for attack scenarios (printed photo, digital photo, video replay)*

4. Tartışma
4. Discussion

Deneysel çalışmalar sonucunda VGG16, DenseNet121 ve MobileNet ağ modellerinin ve bu ağlarda üretilen derin özniteliklerin SVM ile sınıflandırılması sonucu Basılı Fotoğraf, Dijital Fotoğraf ve Video Oynatma saldırıları için elde edilen tüm doğruluk ve F1-skor değerleri Tablo 9'da verilmiştir. Tablodaki değerler incelendiğinde Basılı Fotoğraf saldırı türü için MobileNet ağının ürettiği derin özniteliklerin SVM ile sınıflandırılması sonucu %99.90 doğruluk ve %99.87 F1-skoru ile en iyi sonucun elde edildiği görülmektedir. Bu saldırı türünde ağların tek başına kullanımında MobileNet %99.3 doğruluk ile en iyi başarıma sahiptir. VGG16 ağı tek başına VGG16+SVM yönteminden daha başarılı iken, DenseNet121+SVM modeli DenseNet121 e göre %5.11 daha başarılı bir sınıflandırma gerçekleştirmiştir. Bu saldırı türünde üç ağdan üretilen derin özniteliklerin birleştirilerek SVM ile sınıflandırılması sonucu (VGG16+DenseNet 121+MobileNet+SVM) %98.9 doğruluk elde edilmiş fakat bu değer MobileNet+SVM modelinden daha düşük kalmıştır.

Tablo 9. Tüm modellerin tüm saldırı senaryoları için doğruluk ve F1-skoru değerleri (%)
Table 9. *Accuracy and F1-score values for all attack scenarios of all models (%)*

Saldırı türü	**Basılı fotoğraf**		**Dijital fotoğraf**		**Video oynatma**	
Yöntem	**Doğruluk**	**F1-skoru**	**Doğruluk**	**F1-skoru**	**Doğruluk**	**F1-skoru**
VGG16	89.10	88,00	**93.50**	**91.15**	88.01	86.62
DenseNet121	93.69	92.69	73.50	50.46	84.52	82.25
MobileNet	99.30	99.13	78.20	63.05	62.90	64.22
VGG16+SVM	81.90	77.94	92.70	90.14	95.38	94.39
DenseNet121+SVM	98.80	98.47	75.40	62.61	97.43	96.70
MobileNet+SVM	**99.90**	**99.87**	67.20	46.40	90.16	87.77
VGG16+DenseNet121+ MobileNet+SVM	98.90	98.64	84.90	81.60	**99.18**	**98.95**

Dijital Fotoğraf saldırı türü için VGG16 ağı tek başına %93.5 doğruluk ve %91.15 F1-skoru başarımı göstermiştir. Bu saldırı türünde DenseNet121 ve MobileNet ağlarının daha düşük bir başarıma sahip olduğu görülmektedir. Diğer yandan derin özniteliklerin ayrı ayrı ya da birleştirilip SVM ile sınıflandırılması herhangi bir performans artışı sağlamamıştır. Diğerlerine kıyasla göz kırpma gibi canlılık belirtileri içerdiği için belirlenmesi daha zor olan Video Oynatma saldırı türünde ise Tablo 9'dan da görüldüğü gibi en iyi başarım VGG16+DenseNet121+ MobileNet+SVM modeli ile elde edilmiştir. Bu saldırı türünde %99.18 doğruluk ve %98.95 F1-skoru sonucuna ulaşılmıştır. Diğer modeller incelendiğinde ağların tek başına düşük başarıma sahip olduğu (VGG16 %88.01, DenseNet121 %84.52, MobileNet %62.90) görülmektedir. Diğer yandan bu saldırı türünde ağlardan üretilen derin özniteliklerin SVM ile sınıflandırılması, genel olarak ağların verdiği doğruluk sonuçlarını büyük ölçüde artırmıştır (VGG16+SVM %95.38, DenseNet121+SVM %97.43, MobileNet+SVM %90.16). Farklı çözünürlük, aydınlatma koşulları, giriş cihazı türü ve konumu gibi etkenler saldırı senaryolarını çeşitlendirmektedir. Dolayısıyla, YST sisteminin gerçek zamanlı kullanımında ne tür bir senaryo ile karşılaşacağı tam olarak bilenememektedir. Literatürde yaygın kullanılan veri kümeleri birçok

saldırı senaryosu içermektedir. Ancak bunları kullanan çoğu çalışmada genellikle tek bir saldırı senaryosu için (tüm saldırı türlerinin bir arada olduğu senaryo) sonuçlarının paylaşıldığı görülmektedir. Bu durum, önerilen modelin farklı senaryolar altındaki davranışını kestirmeyi ve genelleştirilmiş bir çözüm üretmeyi zorlaştırmaktadır. Diğer yandan derin ağların YST başarımları çeşitli çalışmalarda incelense de bu ağlardan üretilen derin özniteliklerin birleştirilmesinin YST başarımına etkisi incelenmemiştir. Bu nedenlerle bu çalışma atak türlerinde YST başarımlarının incelenmesi ve derin özniteliklerin ve birleşimlerinin makine öğrenmesi algoritmasıyla sınıflandırılması konusunda katkı sağlamaktadır.

5. Sonuçlar
5. Results

Bu çalışmada VGG16, DenseNet121 ve MobileNet CNN mimarilerinden üretilen derin özniteliklerin Replay-Attack veri setindeki Basılı Fotoğraf (Print Photo), Dijital Fotoğraf (Digital Photo) ve Video Oynatma (Replay Video) saldırıları için YST başarımları incelenmiştir. Çalışmada 7 model 3 saldırı türü olmak üzere 21 farklı deney gerçekleştirilmiştir. Elde edilen sonuçlara göre farklı saldırı türlerinde farklı modeller daha iyi bir başarım göstermektedir. Derin ağların tek başına kullanımı sadece 1 deneyde (Dijital Fotoğraf saldırısı, VGG16 ağı) en yüksek başarıma sahiptir. Derin özniteliklerin SVM ile sınıflandırılması 6 deneyde ağların tek başına kullanımına göre sınıflandırma başarımını artırmıştır. Basılı Fotoğraf saldırı türünde MobileNet+SVM yöntemi ile en iyi sonuç elde edilmiştir. Derin ağlardan üretilen derin özniteliklerin birleştirilip SVM ile sınıflandırılması ise daha karmaşık saldırı türlerinde daha başarılıdır (Video Oynatma saldırısı, VGG16+DenseNet121+MobileNet+SVM modeli). Literatürde genellikle tek bir saldırı senaryosu için (tüm saldırı türlerinin bir arada olduğu senaryo) sonuçları paylaşılmaktadır. Bu durum önerilen modelin farklı senaryolar altındaki davranışını kestirmeyi zorlaştırmaktadır. Diğer yandan derin ağlarda üretilen derin özniteliklerin birleştirilmesinin YST başarımına etkisi incelenmemiştir. Çalışma bu alanda çalışmak isteyenlere farklı bir bakış açısı sunmaktadır. Çalışmada kullanılan Replay-Attack veri seti literatürdeki pek çok çalışmada kullanılmaktadır. Veri setinde farklı etnik kökene sahip kişilerin gerçek ve sahte görüntülerinin da dahil edilmesi ile yöntemin geçerliliği artırılabilir. İleriki çalışmalarda daha güncel veri setleri üzerinde deneyler gerçekleştirilebilir. Çalışmada kullanılan ağlar dışındaki diğer ağların da başarımları incelenebilir. Ayrıca ağların parametreleri değiştirilerek performans artışı sağlamak üzerine çalışmalar yapılabilir. Bu çalışmada öznitelik seviyesinde birleştirme işlemi gerçekleştirilmiş ve YST başarımı değerlendirilmiştir. İleriki çalışmalarda ağların YST sonuçlarının ağırlıklı ya da oylama yoluyla birleştirilmesi ile başarım değerlendirilmesi yapılabilir.

Yazar katkısı
Author contribution

Asuman GÜNAY YILMAZ: literatür taraması, metodoloji, ilgili tabloların/şekillerin yorumlanması, makalenin yazılması ve düzenlenmesi.
Fırat ŞAKAR: deneysel çalışmalar, ilgili tabloların/şekillerin hazırlanması, makalenin yazılması.

Etik beyanı
Declaration of ethical code

Bu makalenin yazarları, bu çalışmada kullanılan materyal ve yöntemlerin etik kurul izni ve / veya yasal-özel izin gerektirmediğini beyan etmektedir.

Çıkar çatışması beyanı
Conflicts of interest

Yazarlar herhangi bir çıkar çatışması olmadığını beyan eder.

Kaynakça
References

Agarwal, A., Singh, R., & Vatsa, M. (2016). Face anti-spoofing using Haralick features. *2016 IEEE 8th International Conference on Biometrics Theory, Applications and Systems* (*BTAS 2016)*, (pp. 1-6), Niagara Falls, NY, USA. https://doi.org/10.1109/ BTAS.2016.7791171

Alotaibi, A., & Mahmood, A. (2017). Deep face liveness detection based on nonlinear diffusion using convolution neural network. *Signal, Image and Video Processing ,11,* 713–720. https://doi.org/10.1007/s11760-016-1014-2

Anjos, A., Chakka, M.M., & Marcel, S. (2014) Motion-based counter-measures to photo attacks in face recognition. *IET Biometrics*, *3*, 147-158. https://doi.org/10.1049/iet-bmt.2012.0071

Boulkenafet, Z., Komulainen, J., & Hadid, A. (2016). Face spoofing detection using colour texture analysis. *IEEE Transactions on Information Forensics and Security, 11*(8), 1818-1830. https://doi.org/10.1109/TIFS.2016.2555286

Boulkenafet, Z., Komulainen, J., & Hadid, A. (2017). Face antispoofing using speeded-up robust features and fisher vector encoding. *IEEE Signal Processing Letters, 24*, 141-145. https://doi.org/10.1109/LSP.2016.2630740

Chingovska, I., Anjos, A., & Marcel, S. (2012). On the effectiveness of local binary patterns in face anti-spoofing, *Proceedings of the International Conference of Biometrics Special Interest Group (BIOSIG),* (pp. 1-7), Darmstadt, Germany.

Einy, S. Oz, C., & Navaei, Y.D. (2021). IoT cloud-based framework for face spoofing detection with deep multicolor feature learning model. *Journal of Sensors*. https://doi.org/10.1155/2021/5047808

Galbally, J., Marcel, S., & Fierrez, J. (2014). Image quality assessment for fake biometric detection: application to iris, fingerprint, and face recognition. *IEEE Transactions on Image Processing, 23*, 710-724. https://doi.org/10.1109/TIP.2013.2292332

Gan, J. Li, S., Zhai, Y., & Liu, C. (2017). 3D Convolutional neural network based on face anti-spoofing, *2017 2nd International Conference on Multimedia and Image Processing (ICMIP)*, (pp. 1-5), Wuhan, China. https://doi.org/10.1109/ICMIP.2017.9

King, D. E. (2009). Dlib-ml: A machine learning toolkit. *The Journal of Machine Learning Research*, 10, 1755–1758. https://dl.acm.org/doi/10.5555/1577069.1755843

Korkmaz, Ş., & Alkan, M. (2023). Derin öğrenme algoritmalarını kullanarak deepfake video tespiti. *Politeknik Dergisi*, *26*(2), 855-862. https://doi.org/10.2339/politeknik.1063104

Le, K. (2021, Mar 25). *An overview of VGG16 and NiN models*. https://medium.com/mlearning-ai/an-overview-of-vgg16-and-nin-models-96e4bf398484

Liu, Y., Stehouwer, J., Jourabloo, A., & Liu, X. (2019). Deep tree learning for zero-shot face anti-spoofing. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, (pp. 4680-4689), Long Beach, CA, USA. https://10.1109/CVPR.2019.00481

Määttä, J., Hadid, A., & Pietikäinen, M. (2011). Face spoofing detection from single images using micro-texture analysis. *2011 International Joint Conference on Biometrics (IJCB*), (pp. 1-7), Washington, DC, USA. *doi: 10.1109/IJCB.2011.6117510.*

Määttä, J., Hadid, A., & Pietikäinen, M. (2012). Face spoofing detection from single images using texture and local shape analysis. *IET Biometrics*, *1*,3-10. https://doi.org/10.1049/iet-bmt.2011.0009

Sarkar, A. (2020, Jul 11). *Creating DenseNet 121 with TensorFlow*. https://towardsdatascience.com/creating-densenet-121-with-tensorflow-edbc08a956d8

Singhal, G. (2020, Nov 16). *Transfer learning in deep dearning using Tensorflow 2.0.* https://www.pluralsight.com/guides/transfer-learning-in-deep-learning-using-tensorflow-2.0

Wen, D., Han, H., & Jain, A.K. (2015). Face spoof detection with image distortion analysis. *IEEE Transactions on Information Forensics and Security, 10*(4), 746-761. https://doi.org/10.1109/TIFS.2015.2400395

Zhao, X., Lin, Y., & Heikkila, J. (2018). Dynamic texture recognition using volume local binary count patterns with an application to 2D face spoofing detection. *IEEE Transactions on Multimedia, 20*, 552-566. https://doi.org/10.1109/TMM.2017.2750415