

**T.C.
HARRAN ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

YÜKSEK LİSANS TEZİ

**EVRIŞİMLİ SİNİR AĞLARINDA TRANSFER ÖĞRENMESİ İLE GAN
TARAFINDAN ÜRETİLEN SAHTE GÖRÜNTÜ TESPİTİ**

Ece ECEMİŞ ELMACI

BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI

**ŞANLIURFA
2023**

**T.C.
HARRAN ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

YÜKSEK LİSANS TEZİ

**EVRIŞİMLİ SİNİR AĞLARINDA TRANSFER ÖĞRENMESİ İLE GAN
TARAFINDAN ÜRETİLEN SAHTE GÖRÜNTÜ TESPİTİ**

Ece ECEMİŞ ELMACI

BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI

**ŞANLIURFA
2023**

İÇİNDEKİLER

Sayfa No

ÖZET	i
ABSTRACT	ii
TEŞEKKÜR	iii
ŞEKİLLER DİZİNİ	iv
ÇİZELGELER DİZİNİ.....	v
SİMGELER ve KISALTMALAR DİZİNİ.....	vi
1. GİRİŞ.....	1
2. ÖNCEKİ ÇALIŞMALAR	3
3. MATERYAL ve YÖNTEM	12
3.1. Materyal	12
3.1.1. Veri seti	12
3.2. Yöntem.....	14
3.2.1. Evrişimli Sinir Ağları	15
3.2.2. DenseNet-121 modeli.....	17
3.2.3. Kullanılan makine öğrenmesi algoritmaları.....	19
3.2.3.1. Tepe (Ridge) Sınıflandırıcı	19
3.2.3.2. Rastgele Orman Sınıflandırıcı	21
3.2.3.3. Lojistik Regresyon.....	24
3.2.3.4. K-en yakın komşular	26
3.2.4. DensetNet-121 ile elde edilen özniteliklerin sınıflandırılması.....	28
3.2.5. Gradient-weighted Class Activation Mapping (Grad-CAM).....	30
4. ARAŞTIRMA BULGULARI ve TARTIŞMA	32
4.1. Değerlendirme Ölçütleri	32
4.2. Deneysel Sonuçlar	32
5. SONUÇLAR ve ÖNERİLER	41
KAYNAKLAR.....	43
ÖZGEÇMİŞ.....	50

ÖZET

Yüksek Lisans Tezi

EVRIŞİMLİ SINIR AĞLARINDA TRANSFER ÖĞRENMESİ İLE GAN TARAFINDAN ÜRETİLEN SAHTE GÖRÜNTÜ TESPİTİ

Ece ECEMİŞ ELMACI

**Harran Üniversitesi
Fen Bilimleri Enstitüsü
Bilgisayar Mühendisliği Anabilim Dalı**

**Danışman: Dr. Öğr. Üyesi Kemal GÜNER
Yıl: 2023, Sayfa: 42**

Derin öğrenme yöntemlerinin son dönemlerdeki popüler ağlarından olan çekişmeli üretken ağ (Generative Adversarial Network, GAN), üretken bir derin öğrenme modeli olarak bilinir. Üretici (generator) ve ayırt edici (discriminator) olmak üzere iki kısımdan oluşmaktadır. Sentetik ya da yapay veri olarak bilinen GAN üretimi verilerin, literatürde oldukça başarılı örnekleri yer almaktadır. Birbirinden oldukça farklı sahalarda kullanıldığı bilinen sentetik yüz verilerinin, insan algısı ile tespit edilebilmesi pek mümkün olmamaktadır. Bu tez çalışması kapsamında literatürde yaygın olarak kullanılan Evrişimli Sinir Ağı (CNN) modellerinin öznetelik çıkarıcı olarak kullanıldığı, sentetik ve gerçek yüz görüntülerini ayırt eden bir ağ için Laplace filtresi ve benzemezlik tabanlı yeni bir evrişim katmanı içeren bir model önerilmiştir. Bu yöntemde, Densenet121 modeli öznetelik çıkarıcı olarak kullanılıp, öznetelikler sınıflandırmak amacıyla makine öğrenmesi modellerine verilmiştir. Tez çalışması kapsamında oluşturulan yeni veri kümesi farklı makine öğrenmesi modelleri üzerinde denenmiş olup, en yüksek başarı LR ve KNN modellerinden %99 olarak elde edilmiştir. Elde edilen sonuçlar, önerilen yöntemin sahte yüzlerin tespitinde kullanılabileceğini göstermektedir.

ANAHTAR KELİMELER: Çekişmeli üretken ağ, evrişimli sinir ağı, sentetik veri tespiti.

ABSTRACT

MSc Thesis

DETECTION OF FAKE IMAGES GENERATED BY GAN USING TRANSFER LEARNING IN CONVOLUTIONAL NEURAL NETWORKS

Ece ECEMİŞ ELMACI

**Harran University
Graduate School of Natural and Applied Sciences
Department of Computer Engineering**

**Supervisor: Assist. Prof. Dr. Kemal GÜNER
Year: 2022, Page: 42**

Generative Adversarial Network (GAN), which is one of the popular networks of deep learning methods, is known as a productive deep learning model. It consists of two parts, the generator and the discriminator. There are very successful examples of GAN production data, known as synthetic or artificial data, in the literature. It is not possible to detect synthetic face data, which are known to be used in quite different fields, by human perception. Within the scope of this thesis, a model including a Laplace filter and a new dissimilarity-based convolution layer is proposed for a network that distinguishes synthetic and real face images, in which Convolutional Neural Network (CNN) models, which are widely used in the literature, are used as feature extractors. In this method, Densenet121 model is used as feature extractor and features are given to machine learning models in order to classify them. The new dataset created within the scope of the thesis study was tested on different machine learning models, and the highest success was obtained from LR and KNN models at 99%. The obtained results show that the proposed method can be used to detect fake faces.

KEY WORDS: Generative adversarial network, convolutional neural network, synthetic image.

TEŞEKKÜR

Yüksek lisans eğitimim boyunca bilgi birikim ve deneyimlerini bana aktaran çok değerli danışman hocam Sayın Dr. Öğr. Üyesi Kemal Güner'e, yüksek lisans çalışmalarım sırasında desteğini esirgemeyen başta Arş. Gör. Emre Kuran olmak üzere tüm hocalarıma, hayatımın her döneminde her türlü desteği veren çok kıymetli aileme ve eşim Arş. Gör. Mehmet Elmacı'ya çok teşekkür ederim.



ŞEKİLLER DİZİNİ

	Sayfa No
Şekil 3.1. Önerilen yöntemin akış çizelgesi	11
Şekil 3.2. Gerçek ve sahte görüntüler.....	13
Şekil 3.3. DenseNet-121 mimarisi	17
Şekil 3.4. RF sınıflandırıcısı.....	22
Şekil 4.1. Tepe sınıflandırıcı için karmaşıklık matrisi	31
Şekil 4.2. LR için karmaşıklık matrisi	32
Şekil 4.3. RF sınıflandırıcısı için karmaşıklık matrisi.....	33
Şekil 4.4. KNN sınıflandırıcısı için karmaşıklık matrisi	34
Şekil 4.5. Grad-CAM sonuçları	37



ÇİZELGELER DİZİNİ

	Sayfa No
Çizelge 3.1. Veri setinin detayları	12
Çizelge 3.2. Veri setinin detayları	13
Çizelge 3.3. DenseNet-121 mimarisinin detayları	7
Çizelge 4.1. Farklı sınıflandırıcılar için ana veri setinde sınıflandırma sonuçları	34
Çizelge 4.2. Farklı sınıflandırıcılar için ana veri setinde sınıflandırma sonuçları (BFKL olmadan)...	35
Çizelge 4.3. Farklı sınıflandırıcılar için ana veri setinde sınıflandırma sonuçları (Grad-CAM ile)	35
Çizelge 4.4. Farklı sınıflandırıcılar için veri setinde sınıflandırma sonuçları	36
Çizelge 4.5. Farklı sınıflandırıcılar için veri setinde sınıflandırma sonuçları (BFKL olmadan)	36
Çizelge 4.6. Farklı sınıflandırıcılar için veri setinde sınıflandırma sonuçları (Grad-CAM ile)	36



SİMGELER ve KISALTMALAR DİZİNİ

CNN	Evrişimli Sinir Ağları
FC	Tam Bağlı
GAN	Çekişmeli Üretken Ağlar
GLCM	Gri-Seviye Birlikte Oluşum Matrisi
KA	Karar Ağacı
KNN	K-en Yakın Komşular
LR	Lojistik Regresyon
RF	Rastgele Orman
SVM	Destek Vektör Makineleri



1. GİRİŞ

Yapay zeka ve derin öğrenme teknolojilerinin son yıllardaki hızlı gelişimi, hayatımızın birçok alanında devrim niteliğinde etkiler yaratmış olup oldukça farklı alanlarda kullanılmaktadır (Benet ve Pellicer-Valero, 2022). Bu ilerlemeler, genellikle bilim ve teknoloji alanlarında beklenmedik sorunların çözülmesine yardımcı olan yeni yaklaşımlar ve teknikler ortaya çıkarmıştır. Bu problemlerden bir tanesi de, insanların gizliliğini ve güvenliği tehlikeye atan DF (DeepFake) teknolojisidir. DF ile gerçek veya sahte olduğu ayırt edilemeyen görüntüler oluşturulmaktadır. Bu görüntüler sahte bilginin gerçek gibi yayılabilmesine sebep olmaktadır. Özellikle ünlü insanlar bu tehlikeye daha çok maruz kalabilmektedir (Malik ve ark., 2022). Bu kapsamda, yapay zekanın oldukça fazla olumlu sonuçlarının yanında çeşitli yeni zorlukları da beraberinde getirdiği anlaşılmaktadır. Özellikle, GAN'lar tarafından üretilen sahte görüntülerin etkin ve güvenilir bir şekilde tespit edilmesi, günümüzde önemli bir zorluk haline gelmiştir (Dong ve ark., 2022). CNN'ler görüntü işleme sahasında büyük bir potansiyel göstermiş olsa da, son araştırmalar başka derin öğrenme tekniklerini de incelemiştir. Örneğin, Generative Adversarial Networks (GAN'lar) (Goodfellow ve ark., 2020). GAN'ların kullanımı, çeşitli, yüksek kaliteli sahte görüntülerin oluşturulmasını sağlar. Bazı problemler için GAN ile üretilen veriler de eğitimde kullanılarak modellerinin sağlamlığını potansiyel olarak artırabilir (Fezza ve ark., 2022; Ranjith Kumar ve ark., 2020).

Bir başka yaklaşım, veri kodlamasını denetimsiz bir şekilde öğrenmek üzere tasarlanmış sinir ağları olan otomatik kodlayıcıların (autoencoders) kullanılmasıdır. Otomatik kodlayıcılar, sahte yüz tespitinde gerçek yüzlerin normal özelliklerini öğrenmek ve ardından sahte olanlardaki sapmaları belirlemek için kullanılmıştır (Balasubramanian ve Trojovský, 2022). Önceden eğitilmiş bir modelin yeni bir sorunda ince ayarlamalar yapılarak yeniden kullanılması işlemi olan transfer öğrenme, sahte yüz tespitine de uygulanmıştır. Büyük ölçekli görüntü veya yüz tanıma görevlerinde zaten eğitilmiş modellerle başlayarak, araştırmacılar öğrenilmiş özellikleri kullanabilir. Böylece büyük miktarda eğitim verisi ihtiyacını

karşılayabilirler. Nitekim, transfer öğrenmenin etkinliği, orijinal ve yeni görevler arasındaki benzerliğe bağlı olabilir (Elhassan ve ark., 2022; Mishra ve Samanta, 2022).

Bu çalışma, bilimsel literatüre önemli bir katkı sağlamaktadır. Hem teorik hem de pratik yönleriyle, sahte görüntü tespiti ve transfer öğrenme tekniklerine ilişkin bilimsel anlayışı genişletmeyi hedeflemektedir. Bu amaçla makine öğrenmesi teknikleriyle, transfer öğrenmesi birleştirilerek, CNN tabanlı transfer öğrenmesi modelinin başarısının daha da ileriye taşınması hedeflenmiştir.

Literatürde makine öğrenmesi modelleri ile derin öğrenme modelleri birlikte (hibrit) kullanılarak her iki yöntemin avantajlarından faydalanılmıştır (Qaid ve ark., 2021). Bu çalışmada da, bu çalışmanın bir ön çalışması olan BFKL katmanına sahip bir transfer öğrenme modeli kullanılmıştır (Ecemiş ve ark., 2023). Bu model, öznetelik çıkarıcı olarak kullanılıp öznetelikler makine öğrenmesi algoritmalarıyla sınıflandırılmıştır. Elde edilen sonuçlar, önceki modelin doğruluk oranını (accuracy) yaklaşık %2 oranında artırmıştır.

Bu çalışmanın devamı aşağıdaki bölümler ile düzenlenmiştir. İkinci bölümde, güncel literatürde sentetik veri tespiti yapan çalışmalar hakkında yapılan araştırmalara ilişkin bilgiler verilmiştir. Üçüncü bölümde, kullanılan materyaller, önerilen yöntem ve bu yöntemin detayları açıklanmıştır. Dördüncü bölümde, önerilen yöntemin nicel olarak sergilemiş olduğu performansı gösterebilmek amacıyla deneysel sonuçlar, veri seti hakkında bilgiler ve yardımcı görseller aktarılmıştır. Beşinci ve son bölümde ise, sonuçlar hakkında yorumlamalar ve genel bir değerlendirme yer almaktadır.

2. ÖNCEKİ ÇALIŞMALAR

DF, "Derin öğrenme" ve "sahte" kelimelerinin birleşiminden türetilen "derin sahte" terimi, genellikle bireylerin hiç yapmadıkları şeyleri söyledikleri veya yaptıkları hiper-gerçekçi manipüle edilmiş görüntüleri veya videoları ifade eder. Bu manipülasyonlar, eğlence ve sanat gibi alanlarda zararsız kullanımlar bulmuş olsalar da, yanıltıcı bilgi, kimlik hırsızlığı ve iftira gibi önemli tehditler de oluştururlar. Bu nedenle, güvenilir ve sağlam sahte yüz tespit tekniklerinin geliştirilmesi, oldukça önemli bir araştırma çabası haline gelmiştir.

Bu literatür incelemesi, sahte yüz tespiti için derin öğrenme yöntemlerini kullanan son çalışmaları kapsamlı bir şekilde incelemeyi amaçlamaktadır. İnceleme, çeşitli teknikleri kapsayarak, anahtar bulguları vurgulayarak, mevcut zorlukları tartışıp alandaki potansiyel gelecek yönlerini önermektedir.

Favorskaya ve Yakimchuk (2021), tarafından yapılan bir çalışma, sahte yüz resimlerinin tespiti için iki aşamalı bir yöntem önermektedir. İlk olarak arka plan analizi ile bir yüz resminin küresel parlaklık ve renk parametrelerinin tüm sahneye ne ölçüde uyduğunu ya da ondan ne ölçüde ayrıldığını değerlendirmek için gereklidir. Bu amaçla, bir Gauss karışım modeli kullanılır. İkinci aşamasında, yani yerel ve global eşleştirme aşamasında ise, yüz resmi bir 3x3 öge boyutunda ızgara temsiline ayrılır. Ayrıca, global eşleştirme, tüm yüz resminin genel bir değerlendirmesini gerçekleştirir. Bu aşama, özellikle 3D özelliklerin belirlenmesi amacıyla kullanılır. Bu yöntem, OULU-NPU veri seti ve kendi veri seti üzerinde yapılan deneylerle test edilmiştir. Sunum saldırıları ve ters pertürbasyon saldırılarına karşı doğruluk oranları sırasıyla %82.4-89.1 ve %69.5-75.2 olarak bulunmuştur. Veri artırma (data augmentation) kullanılarak, sunum saldırılarına karşı sahte yüz resmi tespitindeki doğruluk oranı %85.7-92.5'e çıkarılmıştır. Ancak, tanıma sürecinin zaman tahminleri gerçek zamanla uyumlu olmadığı için algoritmaların daha da geliştirilmesi gerekmektedir.

ST ve ark. (2022), tarafından yapılan bir çalışmada DF görüntülerinin tespiti için yeni bir teknik olan Fisherface Doğrusal İkili Desen Histogramı ve Derin İnanç Ağı sınıflandırıcısı (FF-LBPH DBN) kullanılmıştır. Bu metodoloji ile gerçek ile manipüle edilmiş görüntüler arasında ayırım yapılmaya çalışılmıştır. FF-LBPH DBN modeli, DF yüz manipülasyonlarını incelemek için kullanılmış ve başarılı bir performans göstermiştir. İşlem hızını artırmak için, özellik boyutlarının azaltılması Fisherface-LBPH algoritmasının birleşimi ile gerçekleştirildi. Bu yaklaşım, manipüle edilmiş yüz görüntülerinin tespitini kolaylaştırmıştır. FF-LBPH yöntemi doğruluk açısından, önerilen FF-LBPH-DBN, CASIA-WebFace görüntü veri seti ile %98.82 başarı oranı elde ederken, DFFD veri seti ile %97.82 doğruluk oranı elde etti.

Arunkumar ve ark. (2022), tarafından yapılan çalışmada Fuzzy Fisher Face ile Capsule Dual Graph (FFF-CDG) kullanılarak farklı türlerde sahte görüntülerin veya videoların tespitine odaklanılmaktadır. Bu çalışmada FFHQ veri setinden yararlanılmış olup, önerilen sistemler tarafından elde edilen doğruluk oranları sırasıyla %81.5, %89.32, %91.35 ve %95.82 olmuştur.

Kim ve Cho (2021), tarafından yapılan çalışmada, manipüle edilmiş yüz medyasının adli analizi için tasarlanmış yeni bir derin öğrenme modeli önermişlerdir. Model, çok kanallı kısıtlı konvolüsyon adı verilen bir yöntem kullanarak içerik ve iz özelliklerini aynı anda çıkararak sahte yüzleri belirlemektedir. Bu teknik, renkli giriş görüntülerinden türetilen içerik-dışı bir görüntüden iz özelliklerini çıkarırken, içerik özellikleri, ince ayarlanmış bir ResNet-18 modeli aracılığıyla giriş görüntüsünden elde edilmektedir. Önerilen modelin performansı, Face2Face ve DF tarafından manipüle edilen iki veri seti kullanılarak değerlendirilmiş ve çeşitli video sıkıştırma düzeylerinde temel modellere göre daha üstün bir doğruluk göstermiştir. İlgili çalışmada, video sıkıştırma oranı arttıkça doğrulukta bir azalma olduğu belirtilmektedir. Sırasıyla sıkıştırma olmadan, hafif sıkıştırma ve ağır sıkıştırma ile DF ile üretilen görüntülerde %99,96, 97,14 ve 77,94 doğruluk elde edilirken; Face2Face ile üretilen görüntülerde %99,46, %97,49 ve %85.50 doğruluk oranları elde edilmiştir.

Xue ve ark. (2023), tarafından yapılan çalışmada, yerel fizyolojik ve global alıcı özelliklerin tam anlamıyla kullanılmasını sağlayan, yani GLFNet olarak adlandırılan global-yerel yüz bütünleştirme ağı önerilmiştir. GLFNet modeli yerel bölge tespit bölümü ve global tespit bölümü olmak üzere iki bölümden oluşmaktadır. İlk bölüm, iris ve göz bebekleri gibi yüz parçalarından sahteciliğin izlerini tespit eder. İkinci bölüm ise, gerçek görüntüleri sahte olanlardan ayırt edebilmek için bir artıklar bağlantısını benimser. GLFNet, fizyolojik karakteristikleri derin öğrenme ile birleştirerek çeşitli yollarla sahte izlere ulaşır. Bu yöntem, derin öğrenme özelliklerini öğrenirken fizyolojik özelliklere dayanıklıdır. GLFNet ile, çeşitli GAN modelleri tarafından üretilen görüntüler sınıflandırılmış ve doğruluk oranları ölçülmüştür. ProGAN için %99,8, StyleGAN için %98,4 ve StyleGAN2 için ise %99,1 doğruluk oranı elde edilmiştir.

Sharma ve ark. (2022), tarafından yapılan çalışmada, yazarlar gerçek ve sahte yüz görüntülerini sınıflandırmak için derin bir CNN modeli önermişlerdir. Önceden var olan modelleri, özellikle VGG16 ve ResNet50'yi kullanarak bir transfer öğrenme yaklaşımı benimsemişlerdir. Önerilen yaklaşımın etkinliği, 140000 Real and Fake Faces, Real and Fake Face Detection, ve FakeFaces olmak üzere üç tanınmış veri seti kullanılarak doğrulanmıştır. Önerilen model, üç referans veri seti üzerinde sırasıyla %95,85, %53,25 ve %88,63'lük test doğrulukları elde etmiştir. Önerilen modelin, önceden eğitilmiş VGG16 ve ResNet50 modelleri ile entegre edilerek derin topluluklar oluşturulduğunda performansında önemli bir artış gözlemlenmiştir. Bu, üç veri setinde sırasıyla %98,79, %75,79 ve %95,52'lik iyileştirilmiş doğruluklara sonuçlanmıştır.

Şafak ve Barişçi (2022), tarafından yapılan çalışmada, sahte yüz görüntülerinin tespitinde hafif CNN modellerini (MobileNet, MobileNetV2, EfficientNetB0, NASNetMobile) denemişlerdir. Eğitim setinde gerçek ve StyleGAN2 ile üretilmiş sahte görüntüler yer almaktadır. Transfer öğrenme yaklaşımı ile gerçekleştirilen bu çalışmada EfficientNetB0, %93,64 başarı oranı ile en yüksek doğruluk oranına ulaşmaktadır.

Guo ve ark. (2022), tarafından yapılan çalışmada, gözlerdeki tutarsızlıkları inceleyerek GAN tarafından üretilen yüzleri tespit eden sağlam ve dikkatli bir uçtan uca model önermiştir. Model, iki göz arasındaki kalıntıları belirleyip karşılaştırarak tutarsız göz bileşenlerini otomatik olarak belirlemeyi öğrenmek için tasarlanmıştır. İris bölgeleri Mask-RCNN ile çıkarıldıktan sonra, her iki gözdeki kornea parıltıları arasındaki tutarlılığı incelemek için bir Artık Dikkat Ağı (Residual Attention Network, RAN) uygulanmıştır. Model, geleneksel çapraz entropi kaybını, Wilcoxon-Mann-Whitney (WMW) istatistikleri üzerinden ROC-AUC kaybının bir gevşetmesiyle eşleştiren birleşik bir kayıp fonksiyonunu kullanarak dengesiz veriden öğrenme gerçekleştirmiştir. Hem dengeli hem de dengesiz durumlarda, yeni oluşturulan FFHQ-GAN veri seti üzerinde yapılan kapsamlı değerlendirmelerde, model başarılı bir performans sergilemiş olup, dengeli olan veri setinde %97lik bir doğruluk oranı elde edilmiştir.

Liu ve ark. (2018), tarafından yapılan çalışmada, DF tespiti yapılmaya çalışılmıştır. Gerçek ve sahte yüzler arasındaki doku farklılıklarının kolayca tespit edilebilir olduğunun vurgulandığı çalışmada, sıklıkla yerel ikili desen (LBP) doku operatörleri ve optik akış algoritmalarını birleştirmişlerdir. Sabit merkez piksel metodu yerine en yakın komşu ikili hesaplama metodu dikkate alınarak temel LBP yöntemleri de modifiye edilmiştir; ayrıca, geleneksel optik akış algoritması, resmin gürültüsünü azaltan çoklu füzyon özelliği üst üste binme metodu uygulanarak da değiştirilmiştir. Piramit modelinde, çoklu blok resimleri oluşturan blok hesaplamaları kullanılarak her katmanda görüntü işleme gerçekleştirilmiştir. Görüntünün özellikleri, iki füzyon algoritması (MOLF) aracılığıyla elde edilmiş ve daha sonra bir SVM sınıflandırıcı tarafından ayrı ayrı eğitilmiş ve test edilmiştir. Deneysel sonuçlar, bu yöntemin tespit doğruluğunu artırırken aynı zamanda hesaplama karmaşıklığını azaltabileceğini göstermiştir. Yazarlar, optik akışı ve füzyon algoritmalarını içeren çeşitli LBP algoritmalarını karşılaştırmak için CASIA, PRINT-ATTACK ve REPLAY-ATTACK veri tabanlarını kullanmışlardır. Genel olarak, çeşitli saldırılarda %83 ve üzeri doğruluk oranı vardır.

Raza ve ark. (2022), tarafından yapılan çalışmada, araştırmacılar, VGG16 ve evrişimsel sinir ağı mimarisi hibritine dayalı yeni bir derin sahte öngörücü (Deep Fake Predictor, DFP) tekniği önermektedir. Gerçek ve sahte yüzleri içeren bir derin sahte veri seti, sinir ağı stratejilerini oluşturmak için kullanılmıştır. İlgili çalışma, Xception, NAS-Net, Mobile Net ve VGG16 gibi transfer öğrenme yöntemlerini de karşılaştırmaktadır. DFP yöntemi, derin sahte tespiti için %95 doğruluk ve %94 hassasiyet elde etmiştir.

Wang ve ark. (2022), tarafından yapılan çalışmada, GAN tarafından oluşturulan sahte görüntüleri tespit etmek için yeni bir çerçeve - iki akışlı CNN - önerilmektedir. Önerilen bu çerçeve bir RGB akışı ve Fotoğraf Tepki Eşitsizliği (PRNU) akışını içerir. RGB akışının ön işleme aşamasında, rastgele silme özelliği örneklemin çeşitliliğini artırır ve ağı, görüntü içeriğindeki GAN parmak izlerindeki farklara daha çok odaklanmasına yardımcı olur. PRNU akışının oluşturulması, gerçek görüntülerdeki eşsiz PRNU özelliklerine ve bu özelliklerin görüntü dönüşümüne olan dayanıklılığına dayanmaktadır. PRNU'nun varlığı, ağı, görüntü piksel değerindeki değişikliklere odaklanmaya yönlendirir ve ağın genelleme performansını artırır. Birden çok veri setinde yapılan deneysel sonuçlar, önerilen yöntemin doğruluk ve genelleme açısından başarılı olduğunu göstermektedir. Model 5000 ImageNet ve 5000 PGGAN görüntüsü üzerinde test edilmiş ve %81,48 doğruluk oranı elde ettiği belirtilmiştir..

Li ve ark. (2022), tarafından yapılan çalışmada ilk aşamada, bilinen GAN ağları tarafından üretilen tanınabilir sahte yüz özelliklerinin filtrelenmesini sağlamak için ikili bir sınıflandırma kayıp fonksiyonu kullanılır. İkinci aşamada, çeşitli türde GAN ağları veya bilinmeyen sahte yüz oluşturma yöntemleriyle başa çıkmak için bir sınıf sınıflandırma kayıp fonksiyonu kullanılır. MCCNN modeli ana ağ olarak kullanılmış olup genelde çeşitli GAN'lar tarafından üretilen görüntülerde yaklaşık %80 ve üstü doğruluk oranı, ve ProGAN veri setinde %98,90 doğruluk oranı elde edilmiştir.

S. Lee ve ark. (2021), tarafından yapılan çalışmada benzersiz bir El İşi Yüz Manipülasyonu (HFM) görüntü veri seti ve Shallow-FakeFaceNets olarak bilinen yumuşak hesaplama sinir ağı modelleri tanıtılmaktadır. Bunlar, yüksek etkili bir yüz manipülasyonu algılama hattına entegre edilmiştir. Sinir ağı sınıflandırıcı modeli olan Shallow-FakeFaceNet (SFFN), manipüle edilmiş yüz simgelerine odaklanabilme kabiliyetine sahiptir. Böylece sahte görüntüleri tespit eder. Yöntem, küçük GAN üretimi sahte görüntülerini tespit etmede yüzde 93,99 doğruluğa ulaşmıştır.

(Remya ve ark., 2022), tarafından yapılan çalışmada, birleşik renk dokusu özelliği ve derin öğrenme tekniğini kullanarak görüntüleri gerçek veya sahte olarak sınıflandırmak için bir yöntem önerilmektedir. Üç renk kanalından (R, G, B) kanallar-arası ortak renk dokusu bilgisini veren OC-LBP haritası, sahte yüz görüntüsü algılamada RGB görüntülere kıyasla çok etkili bulunmuştur. Üç farklı, öneğitimli CNN modeli; GoogLeNet, MobileNet-v2 ve ResNet-18, OC-LBP haritaları kullanılarak eğitilmiş ve performansları karşılaştırmalı olarak analiz edilmiştir. Bu yöntem, PGGAN ile CelebA-HQ veri seti ve StyleGAN2 ile FFHQ veri seti üzerinde çalıştırılmıştır. İnce ayarlı üç CNN modeli arasında, MobileNet-v2 en yüksek test doğruluğunu elde etmiştir. Özellikle, MobileNet-v2 modeli, PGGAN ile CelebA-HQ veri seti için yüzde 99,73 ve StyleGAN2 ile FFHQ veri seti için yüzde 99,93 test doğrulukları elde etmiştir. İlgili yöntem, JPEG sıkıştırmasına karşı dayanıklılığının yanı sıra veri setlerinin karşılıklı test doğruluğu açısından iyi genelleme yeteneği göstermektedir.

Fu ve ark. (2022), tarafından yapılan çalışmada, GAN tarafından oluşturulan yüz görüntülerini tespit etmek için kör bir yöntem önerilmiştir. Doğal oluşum mekanizmasındaki farklılıklar nedeniyle, doğal ve GAN tarafından oluşturulan yüz görüntüleri farklı doku ve sensör gürültülerini sergiler, bu da GAN tarafından oluşturulan yüz görüntülerini açığa çıkarmak için ipuçları olarak kullanılmıştır. Özellikle, gerçek yüz ve oluşturulan yüz görüntülerinden lokal ikili örüntü (LBP) öznitelikleri çıkartılmıştır ve sensör gürültülerinde çıkarıcı piksel bitişiklik matrisi (SPAM) özellikleri elde edilir. Her iki öznitelik de, yüz görüntülerinin otantikliğini

doğrulamak için bir SVM sınıflandırıcısına beslenmiştir. Sonuçlar, önerilen yöntemin GAN tarafından oluşturulan sahte yüz görüntülerini %97,60'a kadar bir doğrulukla başarıyla tespit edebildiğini göstermektedir.

Deng ve ark. (2023), sahte ve gerçek yüzleri ayırt etmek için GAN tarafından oluşturulan sahte yüzleri incelemiştir. Sahte görüntülerin çeşitliliğinin gerçek görüntüler kadar zengin olmadığını tespit etmişlerdir. PGGAN, StyleGAN, StyleGAN2, StarGAN v2, Deepfakes ve AttGAN tarafından oluşturulan sahte yüz veri setlerini deneylerinde kullanışlardır. Gürültüsü giderilmiş gerçek ve sahte görüntülerin analizi, sahte yüzlerin doğal yüzlerden daha güçlü yerel olmayan benzerlikler gösterdiğini açıkça ortaya koymuştur. Yerel olmayan özellik çıkarma (NFE) modülü ve çok aşamalı bir sınıflandırma alt ağı içeren yerel olmayan dikkat tabanlı sahte görüntü algılama ağı (NAFID) ağı ve sınıflandırma alt ağı olarak ResNeXt'i kullanmışlardır. NAFID'in başarısı, çeşitli sahte yüz veri setlerinde test edilmiş ve özellikle StyleGAN2 veri setinde %99.07'lik bir algılama doğruluğu sergilediği belirtilmiştir.

B. Chen ve ark. (2022), tarafından yapılan çalışmada sonradan işlenmiş GAN görüntüleri üzerinde çalışılmıştır. Bulgular, hem lüminans hem de krominans bileşenlerinin önemli olduğunu ve RGB ve YCbCr renk uzaylarının, HSV ve Lab renk uzaylarına göre daha üstün performans gösterdiğini vurgulamaktadır. Robustluğu artırmak için, renk bilgisini etkili bir şekilde kullanmak amacıyla, çift renkli uzayların (RGB ve YCbCr) hem lüminans hem de krominans bileşenleri dikkate alınmıştır. Makale ayrıca, özelliği temsil etme gücünü artırmak ve çok katmanlı özellikleri birleştirmek amacıyla, Xception modeline evrişimli blok dikkat modülü ve çok katmanlı özellik toplama modülünü eklemektedir. Son olarak, çift renkli uzaylar olan RGB ve YCbCr'yi entegre eden ve geliştirilmiş bir Xception modeli kullanan robust bir çift akışlı ağ tasarlanmıştır. Deneysel sonuçlar, ilgili yöntemin %99.7'ye kadar varan doğruluk oranları verdiğini belirtmektedir.

Hu ve ark. (2021), tarafından yapılan çalışma GAN ile sentezlenmiş yüzlerin, iki göz arasındaki korneal spekülere parlaklık tutarsızlığı nedeniyle ortaya çıktığına

vurgu yapmaktadır. Bu tutarsızlık, GAN modellerindeki fiziksel/fizyolojik kısıtlamaların eksikliğinden kaynaklanmaktadır. Yüksek kaliteli GAN ile sentezlenmiş yüzlerde bu tür kalıntıların (görüntüdeki bozulmaların) yaygın olarak bulunduğunu ve iki gözden korneal speküler parlaklığı çıkarma ve karşılaştırma için otomatik bir yöntemi sunmaktadır. Yöntem, 0.94 AUC skoru üretmiştir.

Dang ve ark. (2020), tarafından yapılan araştırmada, manipüle edilmiş yüz görüntülerini ve GAN ile üretilmiş resimleri ayırt edebilen derin öğrenme mimarisi sunmaktadır. Bu, manipüle edilmiş bölgelerde çok katmanlı soyutlama kullanarak özellik çıkarabilen birçok evrişimli katmanın potansiyelini araştırmaktadır. Dengesiz veri sorununu ele almak için, çalışma, önerilen modeline veri tabanlı, maliyet duyarlı öğrenme tabanlı (sınıf ağırlığı) ve topluluk tabanlı (eXtreme Gradient Boosting) stratejileri uygulamaktadır. Özellikle dengesiz veri problemi (IDP) ile başa çıkmada önerilen modelin avantajı, bir dizi senaryo altında manipüle edilmiş yüz veri setine ve bir GAN üretilmiş yüz veri setine uygulanarak doğrulanmıştır. Modelin performansı, manipüle edilmiş yüz görüntülerini ve GAN ile üretilen resimleri tanımlama görevinde mevcut uzman sistemleri, hesaplama karmaşıklığı, eğri altında alan (AUC) ve dayanıklılık açısından aşmaktadır.

Tan vd. (2023), Gan üretimi görüntülerin tespiti için gradyen bilgisinden yararlanmışlardır. Bu yöntemi kullanan literatürde ilk çalışma olduğu belirtilmiştir. GAN tarafından oluşturulan görüntüleri algılamak için Learning on Gradient (LGrad) adlı yeni bir algılama çerçevesi geliştirilmiştir. Görüntüleri gradyanlara dönüştürmek için dönüşüm modeli olarak önceden eğitilmiş bir CNN modelini kullanılmıştır. Elde edilen gradyen bilgileri sınıflandırıcı yapıları kullanılarak giriş görüntüsünün Gan üretimi olup olmadığı tespit edilmiştir.

Corvi ve ark. (2023), difüzyon modelleri kullanarak üretilen sentetik görüntüleri bozulmamış olanlardan ayırmanın zorluğunu ve mevcut son teknoloji dedektörlerin bu görev için uygun olup olmadığını anlamaya çalıştıkları bir çalışma gerçekleştirmişlerdir. Bu amaçla, önce difüzyon modellerinin bıraktığı adli tıp izlerini açığa çıkarmış, ardından GAN tarafından oluşturulan görüntüler için

geliştirilen mevcut dedektörlerin, özellikle görüntü sıkıştırma ve yeniden boyutlandırmayı içeren zorlu sosyal ağ senaryolarında bu yeni sentetik görüntüler üzerinde nasıl performans gösterdiklerini inceleyen bir çalışma gerçekleştirmişlerdir.

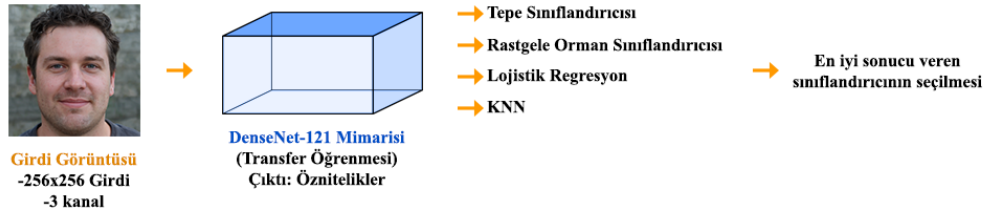
İncelenen yöntemlerde büyük veri setleri ve karmaşık modeller kullanılmıştır, bu çalışmada hem yeni ve düşük boyutlu bir veri seti kullanılmış, hem de yeni önerilen bir model makine öğrenmesi algoritmaları ile hibritlenmiştir.



3. MATERYAL ve YÖNTEM

Bu çalışma kapsamında GAN tarafından her saniyede yeni bir insan yüzü görüntüsü üreten thispersondoesnotexist platformundan elde edilen görüntüler ile flickrdan elde edilen telifsiz, gerçek insan görüntüleri arasında bir sınıflandırma işlemi gerçekleştirilmiştir.

Bu bölümde ve alt bölümlerinde önerilen yöntem ve yöntemde kullanılan alt yöntemler açıklanacaktır. Çalışma, önceden eğitilmiş bir derin öğrenme modeli olan DenseNet-121'yi kullanarak transfer öğrenme yaklaşımını benimsemektedir. Bu yaklaşım, modelin genel görüntü tanıma yeteneklerini, özellikle sahte yüzleri tespit etmek için kullanır. Makine öğrenmesi algoritmaları, modelden sonra, modelin sağladığı (çıkarttığı) öznitelikleri kullanarak sınıflandırma yapılmasını sağlar. Bu bölümde, kullanılan veri setinin hazırlanması, modelin eğitilmesi ve kullanılan araçlar da dahil olmak üzere açıklanacaktır. Şekil 3.1.'de yönteme genel bir bakış verilmiştir.



Şekil 3.1. Önerilen yöntemin modeli: Giriş görüntüsünden ilk olarak Densenet121 mimarisinde bir özellik çıkarıcı kullanılarak özellik çıkarılır, ardından tepe sınıflandırıcısı, rastgele orman sınıflandırıcısı, lojistik regresyon ve KNN kullanılarak sınıflandırma işlemi gerçekleştirilir. Sınıflandırma performanslarına göre en başarılı sınıflandırıcı tespit edilir.

3.1. Materyal

3.1.1. Veri seti

Bu araştırmada ele alınan ana veri seti, sentetik ve gerçek insan görüntülerini içermektedir. İlgili görüntüler bir dizi farklı kaynaktan toplanmıştır. Sentetik

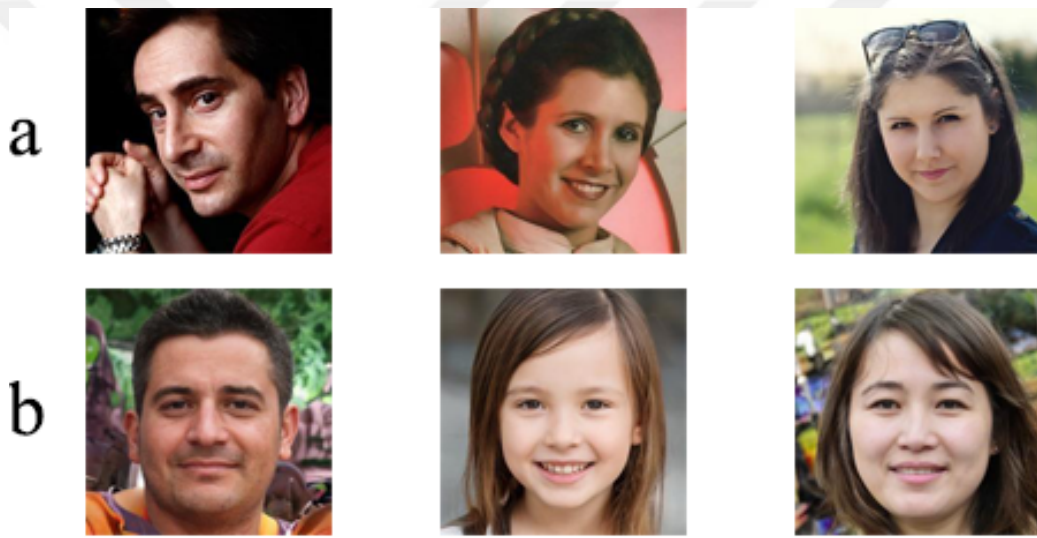
görüntüler, thispersondoesnotexist (“This Person Does Not Exist”, 2022) web portalından alınmıştır, gerçek görüntüler ise Flickr (“Flickr”, 2022) platformundaki telifsiz fotoğraflardan derlenmiştir. Kullanılan veri seti, 1000 sentetik ve 1000 gerçek görüntü olmak üzere toplam 2000 görüntüyü kapsamaktadır. Bu görüntülerin %80'i (1600 görüntü) modelin eğitiminde, kalan %20'si (400 görüntü) ise test sürecinde kullanılmıştır. Görüntüler PNG formatında ve 3 kanallı RGB (Renkli) şeklindedir. Boyutları 256x256 piksel olarak belirlenmiştir ve her kanal, 8-bitlik bir derinliğe sahiptir. Her bir kanal için piksel değerleri [0-255] arasında değişkenlik gösterir. Görüntü etiketleri, dosya adlarından çıkarılmıştır. Gerçek görüntülerin dosya adında "g", sahte görüntülerin dosya adında ise "s" harfi bulunmaktadır. Veri seti, dosyalardan görüntülerin okunması ile kullanıma hazır hale getirilmiştir. İkinci veri seti için (“Kaggle”, 2019) veri setine ve Karras ve ark. (2019) tarafından yapılan çalışmaya bakılabilir. Bu veri seti, 2000 görüntüden oluşmakta olup 1000 görüntü gerçek görüntü, 1000 görüntü ise sahte görüntüdür. Görüntüler PNG formatında ve 3 kanallı RGB (Renkli) şeklindedir. Benzer şekilde bu görüntüler de 256x256 piksel boyutlarında olup, her bir kanal için piksel değerleri [0-255] arasından değişkenlik göstermektedir. Şekil 3.2., ana veri setindeki örnek görüntüleri sergilemektedir. Veri seti sadece test amaçlı kullanılmıştır. Çizelge 3.1.'de kullanılan ana veri setinin detayları verilmiştir. Çizelge 3.2.'de ise veri setine ait detaylar verilmiştir.

Çizelge 3.1. Veri setinin detayları.

Karakteristikler	Sentetik Görüntüler	Gerçek Görüntüler	Toplam
Kaynak	thispersondoesnotexist	Flickr	-
Görüntü Sayısı	1000	1000	2000
Eğitim için Kullanılan Görüntüler (%80)	800	800	1600
Test için Kullanılan Görüntüler (%20)	200	200	400
Görüntü Formatı	PNG	PNG	-
Görüntü Kanal Sayısı	3 (RGB)	3 (RGB)	-
Görüntü Boyutu	256 x 256	256 x 256	-
Kanalların Bit Derinliği	8	8	-
Piksel Değer Aralığı	[0-255]	[0-255]	-

Çizelge 3.2. Veri setinin detaylar (devam)

Karakteristikler	Sentetik Görüntüler	Gerçek Görüntüler	Toplam
Görüntü Sayısı	1000	1000	2000
Görüntü Formatı	JPG	JPG	-
Görüntü Kanal Sayısı	3 (RGB)	3 (RGB)	-
Görüntü Boyutu	256 x 256	256 x 256	-
Kanalların Bit Derinliği	8	8	-
Piksel Değer Aralığı	[0-255]	[0-255]	-



Şekil 3.2. Gerçek ve sahte görüntüler. a: Flickr platformundan elde edilmiş, gerçek insan görüntüleri, b: GAN üretimi sentetik insan görüntüleri.

3.2. Yöntem

Gan üretimi görüntülerin tespiti için, literatürde oldukça farklı sahalarda başarılı sonuçlar ürettiği bilinen CNN modellerinden yararlanılmıştır. Özellik çıkarıcı olarak kullanılan CNN modellerinden elde edilen özellikler farklı makine öğrenmesi teknikleri kullanılarak sınıflandırılmış, bu sınıflandırıcılar arasında bir karşılaştırma gerçekleştirilmiştir. Bu bölümde yöntemle ilgili detaylı açıklamalar verilmektedir.

3.2.1. Evrişimli Sinir Ağları

CNN, Yann LeCun tarafından 1990'ların başında geliştirilmiş bir yapay sinir ağı türüdür (LeCun ve ark., 1989). CNN'ler, özellikle görüntü sınıflandırma ve nesne tanıma gibi görsel verilerle çalışırken büyük başarılar elde etmiştir ve tıbbi görüntü analizi (S. Chen ve ark., 2022), otomatik sürüş (D. H. Lee, Chen, Liou, Liu ve Liu, 2021) ve videoların anlaşılması (Suneetha ve ark., 2021) gibi birçok alanda yaygın olarak kullanılmaktadır. CNN'ler, bir veya daha fazla evrişim katmanı ve ardından genellikle bir veya daha fazla tam bağlantılı katman (fully connected layer) içerir. Evrişim katmanları, öznitelik haritaları (feature maps) oluşturmak için bir giriş görüntüsü üzerinde bir evrişim işlemi uygular. Evrişim işlemi, çekirdek veya filtre adı verilen bir matrisle giriş verileri arasında bir çarpma işlemi gerçekleştirir. Bu işlem matematiksel olarak Denklem 3.1 ile ifade edilebilir.

$$f(x, y) * g(x, y) = \sum \sum f(i, j)g(x - i, y - j) \quad (3.1)$$

Denklem 3.1'de $f(x, y)$ giriş görüntüsü, $g(x, y)$ ise çekirdektir. Evrişim işleminin uygun bir şekilde uygulanması için, ESA kendi içinde özelleşmiş bazı katmanlar barındırmaktadır.

Evrişim Katmanı (Convolutional Layer): Giriş görüntüsünün boyutunu azaltmak ve öznitelikleri çıkarmak için kullanılır. Her evrişim katmanı, giriş görüntüsü üzerinde belirli bir özellik setini tespit etmek için farklı filtreler içerir (Cao ve ark., 2022).

Havuzlama Katmanı (Pooling Layer): Ağırlıklı olmayan bir işlem olan havuzlama, öznitelik haritasının boyutunu azaltırken özelliklerin korunmasına yardımcı olur. MaxPooling ve AvgPooling popüler havuzlama türleridir. Havuzlama katmanı, genellikle bir evrişim katmanının çıktısının boyutunu azaltmak için kullanılır, böylece modelin hesaplama yükünü azaltır ve aşırı uyumu (overfitting) önler (Jie ve Wanda, 2020).

Tam Bağlantılı Katman (Fully Connected Layer): CNN'nin sonunda yer alır ve özniteliklerin sınıflandırma için kullanıldığı yerdir. Özellikler, tam bağlantılı katmana geldiğinde, her düğümün çıktıları, bir sonraki katmandaki her düğüme bağlanır. Bu, ağıın öğrenmesi ve tahmin yapması için gerekli olan final sınıflandırma işlemini sağlar (Basha ve ark., 2020).

Dropout: Aşırı uyumu önlemeye yönelik bir düzenleme tekniği olan dropout, rastgele seçilen nöronların eğitim süresince "kapatılmasına" (dropout) yol açar. Bu, modelin farklı nöronlar arasındaki bağımlılığı azaltır ve aşırı uyuma karşı daha dayanıklı hale getirir (Norhikmah ve ark., 2022).

Aktivasyon Fonksiyonları: CNN'lerde, ReLU (Rectified Linear Unit) gibi aktivasyon fonksiyonları, evrişim ve tam bağlantılı katmanların çıktılarına uygulanarak ağıın doğrusal olmayan özelliklerini öğrenmesine yardımcı olur. Aktivasyon fonksiyonları, nöronların çıktılarını sınırlar ve gradyanların düzgün bir şekilde yayılmasını sağlar. CNN'ler, geriye yayılım (backpropagation) algoritması kullanılarak eğitilir. Bu, ağırlıkları ve sapmaları (biases) güncellemek için bir optimizasyon sürecidir. Stokastik gradyan inişi (SGD) gibi optimizasyon yöntemleri, ağırlıkların ve sapmaların güncellenmesi için kullanılır (Tian ve ark., 2023). Ayrıca, öğrenme hızı, momentum ve ağırlık düşüşü gibi hiperparametreler, modelin eğitim sürecini düzenlemeye yardımcı olur.

Transfer öğrenme, önceden eğitilmiş bir modelin ağırlıklarının ve mimarisinin, yeni ve benzer bir görev için kullanılmasına olanak tanır. Bu, özellikle büyük ölçekli modellerin eğitimi için büyük miktarda hesaplama gücü ve zaman gerektiğinde faydalıdır. Transfer öğrenme sayesinde, önceden eğitilmiş bir modelin ağırlıkları, yeni göreve özgü ince ayarlar yapılarak kullanılabilir (Zhuang ve ark., 2021).

3.2.2. DenseNet-121 modeli

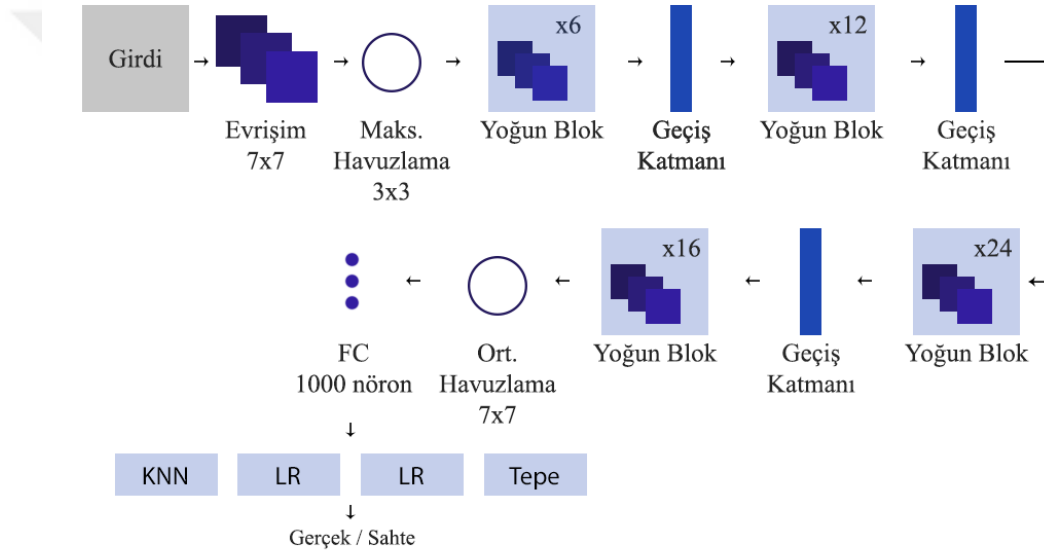
CNN'ler, görüntü sınıflandırma, nesne algılama ve anlamsal bölütleme gibi çeşitli görevlerde en iyi sonuçları elde ederek bilgisayarlı görü alanında ciddi bir yenilik olmuştur (He ve ark., 2016; Krizhevsky ve ark., 2012; Simonyan ve Zisserman, 2015). Bu alandaki en önemli gelişmelerden biri, Dense Convolutional Networks (DenseNet) 2017 yılında ortaya çıkmıştır ve birkaç farklı varyasyonu vardır (Huang ve ark., 2017). Bu bölümde, DenseNet'in temel kavramları, ana avantajları ve derin öğrenme alanındaki etkisi incelenecektir.

Ecemiş ve ark. (2023), Gan üretimi görüntülerin tespiti için aynı veri kümesi üzerinde Densenet mimarisinin en başarılı performansı sergilediklerini göstermişlerdir. Bundan dolayı çalışmada Densenet121 mimarisi özellik çıkarıcı olarak kullanılmıştır. DenseNet'in ana fikri, her katmanı ileri besleme yöntemiyle birbirine bağlamak ve yoğun bir bağlantı düzeni oluşturmaktır (Huang ve ark., 2017). Bu yoğun bağlantı, önceki katmanlar tarafından üretilen öznetelik haritalarının toplama yerine birleştirilmesiyle uygulanır; bu da Residual Networks (ResNet)'lerde yaygın olarak kullanılır (He ve ark., 2016). Bu tasarım seçiminin arkasındaki motivasyon, özelliklerin yeniden kullanımını teşvik etmek ve gradyan akışını iyileştirmek, böylece kaybolan gradyan sorununu hafifletmek ve eğitim verimliliğini artırmaktır. DenseNet'in önemli bir yönü olan büyüme oranı (k olarak belirtilir), her katman tarafından üretilen öznetelik haritalarının sayısını kontrol eder. Küçük bir büyüme oranı, daha az parametre içeren kompakt bir ağa yol açar ve DenseNet'i hesaplamalı açıdan verimli hale getirirken yüksek performansını sürdürür. Ayrıca, DenseNet'in parametre verimliliği, önceki katmanlardan özellikleri yeniden kullanma yeteneğinden kaynaklanır ve böylece yeni özellik çıkarıcılarına olan ihtiyacı azaltır (Huang ve ark., 2017).

Ecemiş ve ark., 2023 Gan üretimi görüntüler ile gerçek görüntüler arasında sınıflandırma yaptıkları çalışmada, aynı veri kümesi üzerinde Densenet121 modelinin en başarılı performansı sergilediğini göstermişlerdir. Buradan elde edilen tecrübeler özellik çıkarıcı modeli olarak Densenet121 modeli tercih edilmiştir.

Hesaplamalı verimliliği artırmak için DenseNet, her evrişimli katmandan önce darboğaz katmanları ekler. Darboğaz katmanında, 1×1 evrişim kullanılarak giriş öznitelik haritalarının sayısı azaltılır ve ağ parametrelerinin büyümesi sınırlandırılır. Ek olarak, DenseNet yoğun bloklar arasında uzamsal boyutları ve öznitelik haritalarının sayısını azaltmak için geçiş katmanları kullanır.

Bu çalışmada kullanılan DenseNet-121 ($k = 32$) mimarisidir ve bu mimarinin detayları Çizelge 3.3. ile verilmiştir. Mimarisi ise Şekil 3.3. ile verilmiştir. Bu mimaride son katmanda kullanılan sınıflandırıcılar da gösterilmiştir.



Şekil 3.3. DenseNet-121 mimarisi. Giriş görüntüsü ilk olarak 7×7 filtre boyutlarındaki evrişim katmanına iletilir, ardından maksimuma havuzlama işlemi ile görüntünün boyutu küçültülürken öznitelik sayısı artırılır. Ardından üç farklı yoğun blok-geçiş katmanı yapısından geçtikten sonra son yoğun bloğa iletilir. Buradan elde edilen özellikler 7×7 filtre boyutlarındaki ortalamaya havuzlama katmanında boyutları küçüldükten sonra 1000 nöronluk tam bağlı katman vasıtasıyla sınıflandırma işlemi gerçekleştirilir.

Çizelge 3.3. DenseNet-121 mimarisinin detayları.

Katmanlar	Çıktı Boyutu	Özellikleri
Evrişim Katmanı	112 x 112	7 x 7 konvolüsyon, kaydırma:2
Havuzlama Katmanı	56 x 56	3 x 3 maks. Havuzlama, kaydırma:2
Yoğun Blok (1)	56 x 56	1 x 1 evrişim 3 x 3 evrişim [x 6]
Geçiş Katmanı (1)	1 x 1 evrişim için: 56 x 56 2 x 2 ort. havuzlama için: 28 x 28	1 x 1 evrişim 2 x 2 ort. havuzlama, kaydırma: 2
Yoğun Blok (2)	28 x 28	1 x 1 evrişim 3 x 3 evrişim [x 12]
Geçiş Katmanı (2)	1 x 1 evrişim için: 28 x 28 2 x 2 ort. havuzlama için: 14 x 14	1 x 1 evrişim 2 x 2 ort. havuzlama, kaydırma: 2
Yoğun Blok (3)	14 x 14	1 x 1 evrişim 3 x 3 evrişim [x 24]
Geçiş Katmanı (3)	1 x 1 evrişim için: 14 x 14 2 x 2 ort. havuzlama için: 14 x 14	1 x 1 evrişim 2 x 2 ort. havuzlama, kaydırma: 2
Yoğun Blok (4)	7 x 7	1 x 1 evrişim 3 x 3 evrişim [x 16]
Sınıflandırma Katmanı	1 x 1	7 x 7 global ort. havuzlama 1000 Nöron FC katman, Softmax

3.2.3. Kullanılan makine öğrenmesi algoritmaları

Giriş görüntüleri Densenet121 mimarisine verilerek elde edilen özellik vektörleri literatürde yaygın olarak kullanılan sınıflandırıcılardan Tepe, Rastgele Orman, Lojistik Regresyon ve K-En yakın komşular algoritmaları kullanılarak sınıflandırılarak giriş görüntüsünün Gan üretimi sentetik görüntü ya da gerçek insan görüntüsü olup olmadığı tespit edilmektedir. Bu bölümde kullanılan sınıflandırıcı yapılardan detaylı olarak bahsedilmektedir.

3.2.3.1. Tepe (Ridge) Sınıflandırıcı

Tepe sınıflandırıcı, doğrusal bir modelleme yöntemi olan Tepe Regresyonu (Hoerl ve Kennard, 1970) temel alınarak geliştirilmiş, sınıflandırma problemleri için kullanılan bir yöntemdir. Bu bölümde, Tepe sınıflandırıcı yöntemini detaylı bir

şekilde inceleyeceğiz. Tepe sınıflandırıcı, Tepe regresyonu adı verilen doğrusal bir modelleme yöntemi temel alınarak geliştirilmiştir. Tepe regresyonu, özellikle çoklu doğrusal bağlantı ve yüksek boyutlu veri kümesi gibi problemlerde başarılı sonuçlar veren, doğrusal regresyonun bir türevi olan bir yöntemdir. Tepe sınıflandırıcı ise, bu regresyon yöntemi kullanılarak yapılan sınıflandırma problemleri için uyarlanmış bir algoritmadır.

Tepe sınıflandırıcı, hedef değişken ve özellikler arasındaki ilişkiyi modelleyerek çalışır. Yüksek boyutlu ve çoklu doğrusal bağlantılı veri kümelerinde, Tepe regresyonu, tahmin edici katsayıları küçültmek ve düzenlemek için bir düzeltme terimi kullanır (Singh ve ark., 2017). Bu düzeltme terimi, modelin karmaşıklığını ve çoklu doğrusal bağlantı problemini azaltır. Tepe sınıflandırıcı da, benzer bir düzeltme terimi kullanarak sınıflandırma problemlerini çözmeye çalışır. Düzenleştirme (regularization) özelliği sayesinde modelin karmaşıklığını ve ezberlemeyi (overfitting) azaltır (Ullah ve ark., 2022). Tepe sınıflandırıcı, doğrusal modellerle çalıştığı için hesaplamaları hızlı bir şekilde gerçekleştirir ve büyük veri kümeleriyle çalışırken zaman tasarrufu sağlar (Arpit ve ark., 2013). Tepe sınıflandırıcı, farklı alanlardaki sınıflandırma problemlerinde kullanılabilir. Örneğin, tıbbi teşhis (Deepa ve ark., 2021) ve metin sınıflandırması (Elgeldawi ve ark., 2021) gibi alanlarda başarılı sonuçlar elde edilmiştir. Tepe sınıflandırıcı performansının değerlendirilmesine yönelik araştırmalar da bulunmaktadır.

Ridge sınıflandırıcı, doğrusal bir model olan Ridge regresyonuna dayanmaktadır. Ridge regresyonu, doğrusal regresyonun düzenleştirilmiş bir versiyonudur. Doğrusal regresyon modeli Denklem 3.2 ile verilebilir.

$$y = X\beta + \varepsilon \quad (3.2)$$

Burada y , bağımlı değişkeni (tahmin etmek istediğimiz hedef); X , bağımsız değişkenlerin (özelliklerin) matrisi; β , tahmin edici katsayıları ve ε , hata terimidir.

Ridge regresyonu, β katsayılarını düzenlileştirerek, özellikle çoklu doğrusal bağlantı ve yüksek boyutlu veri kümesi gibi problemlerde daha doğru ve istikrarlı tahminler yapmayı amaçlar. Ridge regresyonunda, β katsayıları şu Denklem 3.3'te verildiği gibi tahmin edilir.

$$\beta_r = (X^T X + \lambda I)^{-1} X^T y \quad (3.3)$$

Burada λ , düzenlileştirme parametresi (pozitif bir sayı) ve I , birim matristir. λ değeri, düzenlileştirme teriminin etkisini kontrol eder. λ arttıkça, β katsayılarının büyüklüğü azalır ve model daha düşük karmaşıklığa sahip olur.

Tepe sınıflandırıcısı, Tepe regresyonunu sınıflandırma problemlerine uyarlar. İki sınıflı bir sınıflandırma problemi için, Tepe sınıflandırıcısının tahmini Denklem 3.4 ile hesaplanır.

$$\hat{y} = \text{sign}(X\beta_r) \quad (3.4)$$

Burada sign , işaret fonksiyonudur ve $X\beta_{\text{ridge}}$ değeri sıfırdan büyükse 1, sıfırdan küçükse -1 değerini döndürür. Bu şekilde, Tepe sınıflandırıcısı, gözlemleri iki sınıfa (1 ve -1) ayırır. Çok sınıflı sınıflandırma problemleri için, Tepe sınıflandırıcısı, her sınıf için ayrı bir Tepe regresyon modeli eğitir ve her bir modelin tahminlerini karşılaştırarak en yüksek tahmin değerine sahip sınıfı seçer.

3.2.3.2. Rastgele Orman Sınıflandırıcı

RF Sınıflandırıcısı (Breiman, 2001), veri madenciliğinde ve makine öğreniminde yaygın olarak kullanılan güçlü ve esnek bir yöntemdir . Bu yöntem, birden fazla karar ağacını bir araya getirerek oluşturulmuş bir topluluk öğrenme modelidir. Bu yazıda, RF sınıflandırıcısının temel ilkeleri, matematiksel modeli ve denklemleri üzerinde duracağız. RF sınıflandırıcısının temel fikri, birden fazla karar ağacını eğitmek ve bu ağaçların her birinin tahminlerini birleştirerek daha doğru ve

kararlı bir sınıflandırma elde etmektir (Breiman, 2001). Bu yaklaşımın temel bileşenleri şunlardır:

Önyükleme Toplama (Bagging): Her ağaç için orijinal veri kümesinden tekrarlamalı örnekleme yapılır. Bu, ağaçların farklı veri alt kümelerine dayalı olarak eğitilmesini sağlar ve modelin varyansını azaltır (Breiman, 1996).

Öznitelik Alt Kümesi Seçimi: Her düğümde, rastgele bir öznitelik alt kümesi seçilir ve bu alt kümedeki öznitelikler, düğümü bölmek için kullanılır. Bu, ağaçların korelasyonunu azaltır ve modelin doğruluğunu artırır (Ho, 1998).

Her ağaç için bir "bootstrap" örnekleme yapılır. Bu örneklem, orijinal veri kümesinin yaklaşık %63'ünü içerir ve kalan %37'si "out-of-bag" (OOB) örnekleme olarak adlandırılır. Her düğümde, özniteliklerin bir alt kümesi seçilir ve bu alt küme, düğümü bölmek için kullanılır. Alt kümenin boyutu genellikle, öznitelik sayısının karekökü olarak seçilir. Ağaçlar, gini impurity veya entropi gibi ölçütlere göre oluşturulur ve düğümler en iyi bölünmeyi sağlayacak şekilde ayrılır (Ismanto ve ark., 2020). Her ağaç, bootstrap örnekleme üzerinde eğitilir ve OOB örnekleme üzerinde test edilir. Ağaçların performansı, OOB hata oranı ile ölçülür. Son olarak, tahminler, ağaçların çoğunluk oyu ile birleştirilir ve son sınıflandırma elde edilir (Breiman, 2001).

RF sınıflandırıcısının performansını değerlendirmek için kullanılan matematiksel denklemler şunlardır:

Gini Safsızlığı (Gini Impurity): Gini safsızlığı, bir düğümdeki örneklerin sınıf dağılımının homojenliğini ölçer. Düğümdeki sınıfların dağılımı ne kadar dengeliyse, Gini safsızlığı o kadar yüksek olur. Denlem 3.5. ile Gini safsızlığı verilmiştir.

$$\text{Gini_safsızlığı}(D) = 1 - \sum_j [p(j|D)]^2 \quad (3.5)$$

Burada, $p(j|D)$ D düğümündeki örnekler arasında j sınıfının oranıdır.

Entropi: Entropi, bir düğümdeki örneklerin sınıf dağılımının düzensizliğini ölçer. Düğümdeki sınıfların dağılımı ne kadar dengeliyse, entropi o kadar yüksek olur. Entropi denklemi Denklem 3.6 ile verilmiştir.

$$\text{Entropi}(D) = - \sum_j [p(j|D) \cdot \log_2(p(j|D))] \quad (3.6)$$

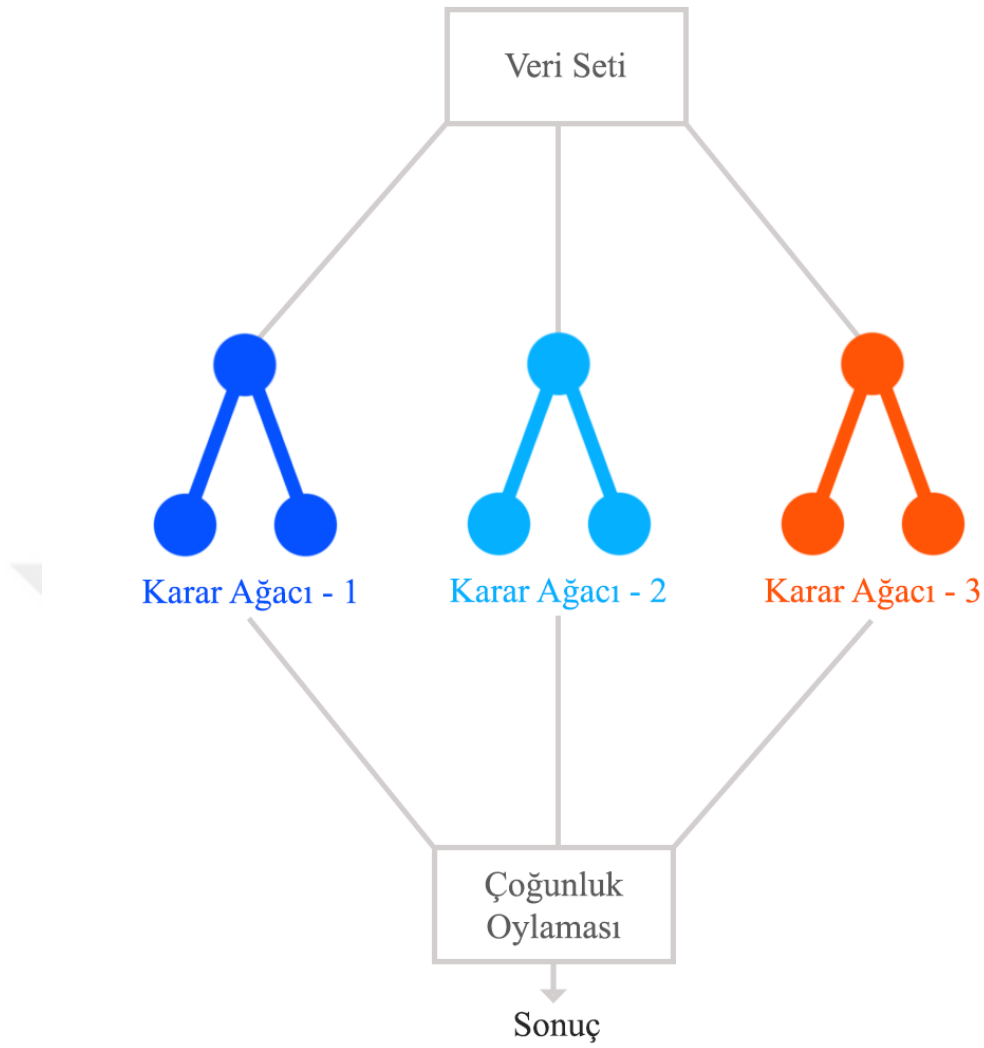
Burada, $p(j|D)$ D düğümündeki örnekler arasında j sınıfının oranıdır.

OOB Hata Oranı: OOB hata oranı, her ağacın OOB örnekleme üzerindeki performansını ölçer. Daha düşük bir OOB hata oranı, daha iyi bir model performansı anlamına gelir. Denklem 3.7 ile OOB Hata Oranı verilmiştir.

$$\text{OOB Hata Oranı} = \frac{\sum_i (y_i \neq \hat{y}_i)}{N} \quad (3.7)$$

Burada, y_i gerçek sınıf etiketi, $\hat{y}_i$ tahmini sınıf etiketi ve N toplam örnek sayısıdır.

RF Sınıflandırıcısı, veri madenciliği uygulamalarında yaygın olarak kullanılan etkili bir yöntemdir (Tang, 2022). Şekil 3.4. ile RF sınıflandırıcısı temsili olarak gösterilmiştir.



Şekil 3.4. RF sınıflandırıcısı. Farklı ağaç sınıflandırıcılarından meydana gelen bir yapı ihtiva etmektedir. Son aşamasında yer alan çoğunluk oylaması işlemi ile sınıflandırıcının karar mekanizması olarak hangi ağacın seçileceğine karar verilmektedir.

3.2.3.3. Lojistik Regresyon

LR, bağımlı değişkenin iki kategori arasında olduğu durumlarda kullanılan istatistiksel bir analiz yöntemidir (Cramer, 2003). Bu yöntem, bağımsız değişkenlerin etkisini ölçmek ve bağımlı değişkenin hangi kategoriye ait olduğunu tahmin etmek için kullanılır. LR, 1958 yılında David Cox tarafından geliştirilmiştir ve o zamandan beri sosyal bilimler, tıp ve ekonomi gibi birçok alanda yaygın olarak kullanılmaktadır (Cox, 1959).

LR modeli, doğrusal regresyon modelinin bir uzantısıdır. Doğrusal regresyon modelinde, bağımlı değişkenin değeri, bağımsız değişkenler ve hata terimi arasındaki ilişkiye dayalı olarak tahmin edilir. LR modelinde ise, bağımlı değişkenin değeri, bağımsız değişkenler ve hata terimi arasındaki ilişkiye dayalı olarak tahmin edilen olasılık değeridir.

LR modelinin temel denklemi Denklem 3.8 ile verilmiştir.

$$P(Y = 1|X) = \frac{1}{1+e^{-(\beta_0+\beta_1X_1+\beta_2X_2+\dots+\beta_kX_k)}} \quad (3.8)$$

Bu denklemde, $P(Y=1|X)$ bağımlı değişkenin 1 kategorisine ait olma olasılığıdır ve $X_1, X_2, \dots, X_k$ bağımsız değişkenlerdir. $\beta_0, \beta_1, \dots, \beta_k$ ise regresyon katsayılarıdır.

LR modelinde, logit dönüşümü kullanılır. Logit dönüşümü, bağımlı değişkenin iki kategori arasındaki oranının doğal logaritmasını alarak, olasılık değerlerini doğrusal bir yapıya dönüştürür. Bu yapı Denklem 3.9 ile gösterilir.

$$\text{logit}(P) = \ln\left(\frac{P}{1-P}\right) \quad (3.9)$$

Bu dönüşüm sonucunda, LR modeli Denklem 3.10 ile verilebilir.

$$\text{logit}(P(Y = 1|X)) = \beta_0 + \beta_1X_1 + \beta_2X_2 + \dots + \beta_kX_k \quad (3.10)$$

LR modelinin parametreleri, maksimum olabilirlik yöntemiyle tahmin edilir (Scott ve ark., 1991). Bu yöntem, gözlemlenen verilere en uygun modeli seçmek için olabilirlik fonksiyonunun maksimum değerini bulmayı amaçlar. LR katsayıları, bağımsız değişkenlerin bağımlı değişken üzerindeki etkisini ölçer. Özellikle, her bir katsayı, bağımsız değişkenin bir birim artışının, bağımlı değişkenin log-odds oranında ne kadar değişiklik yapacağını gösterir. Katsayılar, şu şekilde yorumlanabilir: Katsayı pozitif ise, bağımsız değişkenin değeri arttıkça bağımlı değişkenin 1 kategorisine ait olma olasılığı artar. Katsayı negatif ise, bağımsız

değişkenin değeri arttıkça bağımlı değişkenin 1 kategorisine ait olma olasılığı azalır. Katsayı sıfıra eşit ise, bağımsız değişkenin değeri arttıkça bağımlı değişkenin 1 kategorisine ait olma olasılığında değişiklik olmaz.

LR modelinde, katsayıların etkisini daha iyi anlamak için, katsayıların e^β şeklinde ifade edilen oranları kullanılır. Bu oranlar, olasılık oranı (odds-ratio) olarak adlandırılır ve bağımsız değişkenlerin etkisini yorumlamada daha kolaydır.

3.2.3.4. K-en yakın komşular

KNN, gözetimli öğrenme algoritmaları arasında popüler ve basit bir yöntemdir (Altman, 1992). Bu algoritma, önceden etiketlenmiş verilere dayalı olarak yeni veri noktalarının sınıflandırılmasında kullanılır. KNN'nin temel mantığı, bir veri noktasının sınıfının, en yakın komşularının çoğunluk sınıfına dayandığıdır (Cover ve Hart, 1967).

Teorik Temeller

KNN algoritması, örneklerin bir özellik uzayında temsil edildiği ve benzer örneklerin birbirine yakın olduğu varsayımına dayanır. Bu varsayım, özellik uzayındaki komşuluk kavramının, sınıflandırma için önemli olduğunu belirtir.

KNN, özellik uzayında yeni bir veri noktasının etiketini tahmin etmek için en yakın 'K' komşusunu kullanır. K değeri, kullanıcı tarafından belirlenen bir hiperparametredir ve genellikle çapraz doğrulama ile seçilir.

Algoritmanın temel adımları şu şekildedir: İlk olarak, veri noktaları arasındaki mesafeyi hesaplamak gereklidir. Bu mesafe genellikle Öklidyen, Manhattan veya Minkowski gibi belirli bir mesafe ölçüm formülü kullanılarak belirlenir. Daha sonra, bu hesaplanan mesafelere dayanarak, bir veri noktasının 'K' en yakın komşuları belirlenir. Bu noktalar, hesaplanan mesafeye göre en yakın olan 'K' tane veri noktasıdır. Üçüncü adımda, bu 'K' en yakın komşuların hangi sınıflara ait olduklarına

bakılır ve bir oylama yapılır. Oylama genellikle basit bir çoğunluk sistemine dayanır, yani hangi sınıfın en çok temsil edildiği belirlenir. Son adımda ise, bu oylama sonucuna göre yeni veri noktasının sınıfı belirlenir. Yani, yeni veri noktası, 'K' en yakın komşusunun en çok ait olduğu sınıfa atanır. Bu, KNN algoritmasının temel işleyiştir.

KNN algoritması, Öklidyen mesafesi, Manhattan mesafesi, Minkowski mesafesi veya başka bir mesafe ölçümü kullanarak veri noktaları arasındaki mesafeleri hesaplamaktadır (Choi, 2020). Öklidyen mesafesi Denklem 3.11 ile, Manhattan mesafesi Denklem 3.12 ile ve Minkowski mesafesi Denklem 3.13 ile verilmiştir. Bu denklemlerde x ve y farklı noktalar olmak üzere, i ilgili nokta iken N ise toplam nokta sayısıdır. Minkowski mesafesindeki p, metriğin derecesini belirler.

$$d(x,y)=\sqrt{\sum_{i=0}^N(x_i - y_i)^2} \quad (3.11)$$

$$d(x,y) = \sum_{i=0}^N |x_i - y_i| \quad (3.12)$$

$$d(x,y) = (\sum_{i=0}^N |x_i - y_i|^p)^{\frac{1}{p}} \quad (3.13)$$

Yeni veri noktasına en yakın 'K' komşular, hesaplanan mesafelere göre sıralanır ve seçilir. K en yakın komşuların sınıfları, çoğunluk oyu ile belirlenir ve yeni veri noktasının sınıfı bu şekilde atanır. KNN algoritması, farklı mesafe ölçümlerine dayalı olarak çalışabilir. Öklidyen mesafesi, en yaygın kullanılan mesafe ölçümüdür ve iki nokta arasındaki doğrusal uzaklığı hesaplar. Manhattan mesafesi ise, iki nokta arasındaki ızgara yolu uzunluğunu ölçer ve özellikle yüksek boyutlu verilerde etkili olabilir. Minkowski mesafesi, Öklidyen ve Manhattan mesafelerini genelleştiren bir mesafe ölçümüdür. Minkowski mesafesinde 'p' parametresi değiştirilerek farklı mesafe ölçümleri elde edilir. Bu mesafe ölçümleri arasında seçim yaparken, verinin doğasına ve özelliklerine bağlı olarak en uygun olanı belirlemek önemlidir.

KNN algoritmasında K değeri, kullanıcı tarafından belirlenen bir hiperparametredir. K değeri, modelin karmaşıklığı ve genelleme yeteneği üzerinde

önemli bir etkiye sahiptir. Küçük K değerleri, modelin gürültüye daha duyarlı olmasına ve aşırı uyum riskinin artmasına neden olabilir. Büyük K değerleri ise, modelin daha düşük varyans ve daha yüksek yanlılığa sahip olmasına yol açar. K değerinin seçimi için çapraz doğrulama yöntemi kullanılabilir.

KNN algoritması, yüksek boyutlu veri kümelerinde düşük performans gösterme eğilimindedir. Bu durum, "boyut laneti" olarak adlandırılır ve yüksek boyutlu veri kümelerinde mesafelerin anlamlı olmaktan çıkması nedeniyle ortaya çıkar. Bu sorunu çözmek için, boyut indirgeme teknikleri kullanılabilir. Özellik seçimi ve özellik çıkarımı yöntemleri, veri kümesinin boyutunu azaltarak KNN algoritmasının etkinliğini artırabilir.

KNN Sınıflandırıcısı, gözetimli öğrenme algoritmaları arasında basit ve etkili bir yöntemdir. Uygulaması kolaydır ve farklı mesafe ölçümleri ile kullanılabilen esnek bir yapıya sahiptir. Teorik temelleri ve matematiksel modeli sayesinde, KNN algoritması birçok farklı veri kümesinde başarılı sınıflandırma sonuçları elde etmeye olanak tanır. Bununla birlikte, yüksek boyutlu veri kümelerinde etkinliğinin düşmesi ve hiperparametre seçiminin önemli olması gibi dezavantajları da bulunmaktadır. Bu dezavantajların üstesinden gelmek için, boyut indirgeme teknikleri ve doğru hiperparametre seçimi uygulanarak KNN algoritmasının performansı optimize edilebilir.

3.2.4. DensetNet-121 ile elde edilen özniteliklerin sınıflandırılması

Yapay sinir ağlarındaki öğrenme ve optimizasyon işlemleri, çeşitli hiper parametrelerin ayarlanması ve modele göre değiştirilmesi ile yönetilir. DenseNet modelinde, bu işlemler özellikle adım sayısı ve öğrenme katsayısı gibi önemli parametreler üzerinden gerçekleştirilir. Bu çalışmada, daha yüksek doğruluk seviyeleri elde etmek için adım sayısının artırılması ve öğrenme katsayısının düzenlenmesi gibi çeşitli yöntemler kullanılmıştır. Öğrenme katsayısı, bir ağın optimum ağırlıklarını bulmasını sağlar. Ancak, aşırı büyük öğrenme katsayıları optimum ağırlıkları kaçırabilir ve modelin genel performansını olumsuz etkileyebilir.

Bu nedenle, öğrenme katsayısı, modelin daha tutarlı bir şekilde çalışmasını sağlamak için dikkatlice ayarlanmalıdır. DenseNet modeli, bu tür bir optimizasyonu etkin bir şekilde gerçekleştirmeye olanak sağlar. Bu çalışmada, DenseNet modeli, Gri Düzey Eş Oluşum Matrisi (GDEOM) öznitelikleri ve Laplace filtresi ile birleştirilerek güçlendirilmiştir. GDEOM, bir görüntüdeki piksel çiftleri arasındaki uzamsal ilişkiyi tanımlayan bir matristir. Laplace filtresi ise, bir görüntüdeki kenarları ve ince ayrıntıları vurgulamak için kullanılır. Geliştirilen yeni katman olan Benzemezlik Filtreli Konvolüsyonel Laplace (BFKL) katmanı, ağırlık, sentetik yüz görüntüleri ve gerçek yüz görüntüleri arasındaki ayrımı yapabilme yeteneğini geliştirmek için tasarlanmıştır. BFKL katmanı, basitçe iki ayrı öznitelikten ve ağırlık yapılandırmasından oluşur: girdiye tamamen duyarsız olan konvolüsyon ve girdiye duyarlı olan konvolüsyon. Bu çalışmada, önceden eğitilmiş DenseNet modeli kullanılmış ve modelin son katmanı BFKL katmanı ile değiştirilmiştir. DenseNet, önceden eğitilmiş bir model olması ve özellikle görüntü sınıflandırma görevlerinde yüksek başarı göstermesi nedeniyle seçilmiştir. DenseNet, ResNet'e benzer bir yapıya sahiptir, ancak katmanlar arasında daha yoğun bağlantılar sağlar. Bu, modelin daha derin bir öğrenme kapasitesine sahip olmasını ve daha karmaşık özellikler öğrenmesini sağlar. DenseNet'in bu özellikleri, özellikle görüntü sınıflandırma gibi görevlerde, modelin daha karmaşık özellikleri yakalamasına yardımcı olur. Kullanılan bu mimari, daha önce yapılan bir çalışmada detaylıca açıklanmıştır (Ecemiş ve ark., 2023).

Önceden eğitilmiş DenseNet modelinin son katmanı, BFKL katmanı ile değiştirilmiştir. BFKL katmanı, giriş görüntüsünün özelliklerini daha iyi anlamak ve modelin gerçek yüz görüntüleri ile sentetik yüz görüntüleri arasında daha iyi ayrım yapabilmesini sağlamak için tasarlanmıştır. BFKL katmanı, girdiye duyarsız bir konvolüsyon ve girdiye duyarlı bir konvolüsyon olmak üzere iki ayrı konvolüsyon işlemi gerçekleştirir. Duyarsız konvolüsyon, giriş görüntüsünün özelliklerini modelin diğer katmanlarından bağımsız olarak analiz ederken, duyarlı konvolüsyon, giriş görüntüsünün modelin diğer katmanlarına göre nasıl değiştiğini analiz eder. Bir sonraki adım, BFKL katmanının DenseNet modeline entegrasyonudur. Bu, modelin öğrenme sürecine BFKL katmanının eklenmesi ve modelin gerçek yüz görüntüleri ile

sentetik yüz görüntüleri arasında daha iyi ayırım yapabilmesi için BFKL katmanının eğitilmesi anlamına gelir. BFKL katmanı, her iterasyonda modelin giriş verisi üzerinde çalışır ve modelin çıktısını elde etmek için gerekli hesaplamaları yapar. Bu, modelin hem gerçek yüz görüntülerini hem de sentetik yüz görüntülerini doğru bir şekilde sınıflandırabilmesini sağlar. Sonuç olarak, bu çalışmada, DenseNet ve BFKL katmanının birleştirilmesi yoluyla, gerçek ve sentetik yüz görüntüleri arasında daha iyi ayırım yapabilen bir görüntü sınıflandırma modeli önerilmiştir. BFKL katmanı, giriş görüntülerinin özelliklerini daha etkili bir şekilde analiz etmeye ve modelin gerçek yüz görüntüleri ile sentetik yüz görüntüleri arasında daha doğru sınıflandırmalar yapabilmesine yardımcı olmaktadır. Bu yöntem, özellikle derin sahte (DF) teknolojilerine karşı önlem almak için kullanılabilir. DF teknolojisi, gerçek görüntülerin üzerine başka kişilere ait yüz özelliklerinin yerleştirilerek sahte görüntüler üretmeye yarayan bir yöntemdir. Bu tür sahte görüntüler, sosyal medya ve diğer platformlarda yayılarak yanıltıcı bilgilerin yayılmasına ve kişilerin itibarına zarar verebilir.

DenseNet-121 modelinden elde edilen 15 öznitelik Bölüm 3.2.3.'te verilen makine öğrenmesi algoritmalarına verilmiş ve böylece hibrit bir yöntem tercih edilmiştir.

3.2.5. Gradient-weighted Class Activation Mapping (Grad-CAM)

Grad-CAM, derin öğrenme modellerinin yüksek seviye yorumlanabilirliğini sağlamak için kullanılan bir tekniktir. Grad-CAM, konvolüsyonel sinir ağlarının (CNN) aktivasyon haritalarını kullanarak modelin hangi özelliklerine odaklandığını gösteren görselleştirmeler üretir (Selvaraju ve ark., 2020). Bu tür görselleştirmeler, modelin çeşitli sınıflara neden yüksek veya düşük puan verdiğini anlamak için kullanılabilir. Yorumlanabilirlik, özellikle tıbbi teşhis, otomatik sürüş ve güvenlik uygulamaları gibi kritik alanlarda önemli bir yer tutar (Carvalho ve ark., 2019). Bu bölüm, Grad-CAM'ın matematiksel formülasyonunu, avantajlarını ve dezavantajlarını ve farklı alanlarda nasıl uygulanabileceğini ayrıntılı bir şekilde incelemektedir.

Yorumlanabilirlik, bir modelin tahminlerini insanların anlayabileceği şekilde açıklayabilmesi anlamına gelir. Bu, modelin kararlarının güvenilirliğini ve doğruluğunu artırabilir. Dahası, yorumlanabilirlik sayesinde, modelin hatalı tahminler yapmasına neden olan faktörler anlaşılabilir, böylece modelin performansı iyileştirilebilir (Ribeiro ve ark., 2016). Grad-CAM genel olarak Denklem 3.14 Denklem 3.15 ve Denklem 3.16 ile verilebilir.

$$\text{CAM} = \sum_k w_k A^k \quad (3.14)$$

$$w_k = \frac{1}{Z} \sum_i \sum_j \frac{\partial y}{\partial A_{ij}^k} \quad (3.15)$$

$$L_{\text{Grad-CAM}}^c = \text{ReLU}(\text{CAM}) \quad (3.16)$$

Bu denklemlerde, w_k sınıf aktivasyonu ağırlıklarıdır ve A^k özellik haritalarıdır. Z normalizasyon sabitidir ve y modelin çıktı katmanının çıktısıdır. ReLU fonksiyonu, sadece pozitif aktivasyonları korur, bu sayede modelin "dikkat ettiği" bölgeler belirlenir. İlk olarak, bir konvolüsyonel sinir ağı (CNN) modeli eğitilir. Modelin son konvolüsyon katmanından özellik haritası A^k elde edilir. Öznitelik haritası ile modelin çıktısı olan y arasında kısmi türev alınır ve w_k hesaplanır. w_k ve A^k kullanılarak aktivasyon haritası oluşturulur ve bu harita ReLU fonksiyonuna verilir.

4. ARAŞTIRMA BULGULARI ve TARTIŞMA

4.1. Değerlendirme Ölçütleri

Bu bölümde, analiz sonuçlarını değerlendirmek ve sınıflandırıcıların başarılarını birbirleriyle karşılaştırmak için kullanılan belirli değerlendirme kriterleri belirtilmiştir. Bu amaçla kullanılan metrikler doğruluk, kesinlik, özgüllük ve hassasiyet olup, bu metriklerin hesaplanma yöntemleri sırayla Denklem 4.1 Denklem 4.2 Denklem 4.3 ve Denklem 4.4'te ifade edilmiştir.

$$\text{Doğruluk} = \frac{TP+TN}{TP+TN+FP+FN} \quad (4.1)$$

$$\text{Kesinlik} = \frac{TP}{TP+FP} \quad (4.2)$$

$$\text{Hassaslık} = \frac{TP}{TP+FN} \quad (4.3)$$

$$F1 = 2 * \frac{(\text{Kesinlik} * \text{Hassaslık})}{(\text{Kesinlik} + \text{Hassaslık})} \quad (4.4)$$

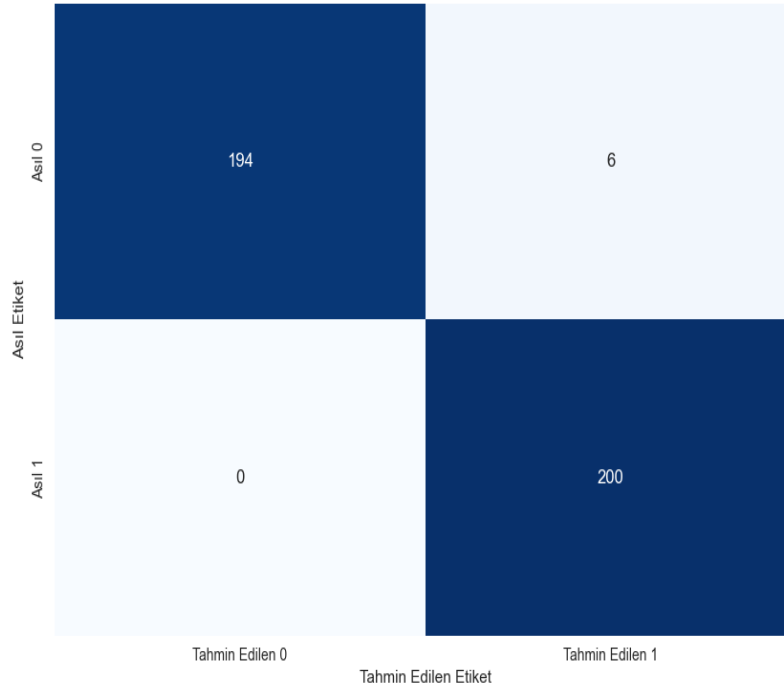
Formülleri belirtilen Denklem 4.1, Denklem 4.2, Denklem 4.3. ve Denklem 4.4'te, TP doğru pozitif olarak, TN doğru negatif olarak, FP yanlış pozitif olarak ve FN yanlış negatif olarak tanımlanmıştır.

4.2. Deneysel Sonuçlar

Önerilen yöntemin deneysel uygulaması Intel (R) Core (TM) i7-6500U 2.50Ghz işlemci, 8GB bellek ve Intel HD Graphics 520 grafik kartına sahip bir bilgisayar ile yapılmıştır. Kaynak kodu Python ile yazılmıştır. Python dili için geliştirilen TensorFlow ve NumPy kütüphaneleri kullanılmıştır.

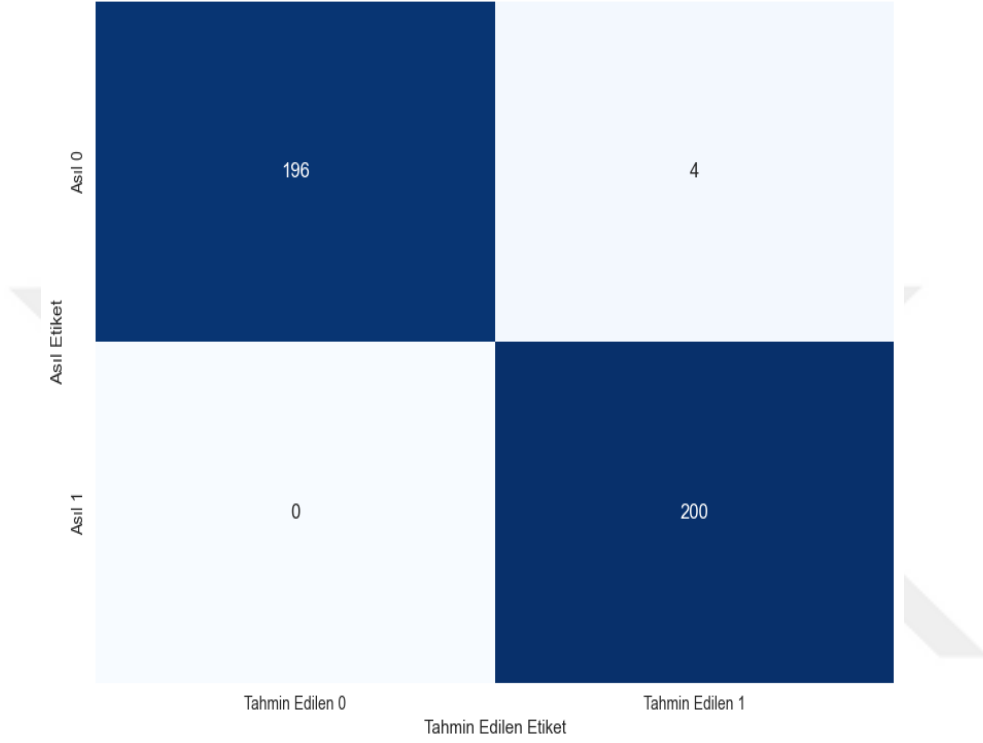
Bu makalede, verilen bir veri kümesinde dört popüler sınıflandırıcı olan Tepe Sınıflandırıcısı, LR, RF Sınıflandırıcısı ve KNN Sınıflandırıcısı, DensetNet-121 modelinden çıkartılan özniteliklerin sınıflandırılmasında kullanılmıştır. Modellerin performansları doğruluk, kesinlik, hassasiyet ve F1 skoru gibi çeşitli değerlendirme ölçütleri kullanılarak ölçülmüş, güçlü ve zayıf yönlerini kapsamlı bir şekilde karşılaştırılmıştır. Bu yöntemler ana veri setinde sınanmakla birlikte, sadece test için kullanılan veri setinde de sınanmıştır. Bu bölümde önce ana veri setindeki sonuçlardan bahsedilip, daha sonra veri seti sonuçlarına değinilecektir.

Tepe Sınıflandırıcısı, %98,50 doğruluk oranına ulaşmıştır, bu da modelin veri kümesindeki örneklerin %98,5'ini doğru bir şekilde sınıflandırdığını göstermektedir. Model, %100 hassasiyet değeri ile tüm pozitif durumları belirleyebilmiştir. Ayrıca, %97,09'luk kesinlik, pozitif olarak tahmin edilen örneklerin gerçekten pozitif olduğunu göstermektedir. F1 skoru %98,52'dir ve bu da modelin hem yanlış pozitifler hem de yanlış negatifler açısından neredeyse %100'e yakın bir performans gösterdiğine işaret eder. Şekil 4.1 Şekil 4.2 Şekil 4.3 ve Şekil 4.4 ile karmaşıklık matrisleri verilmiştir.



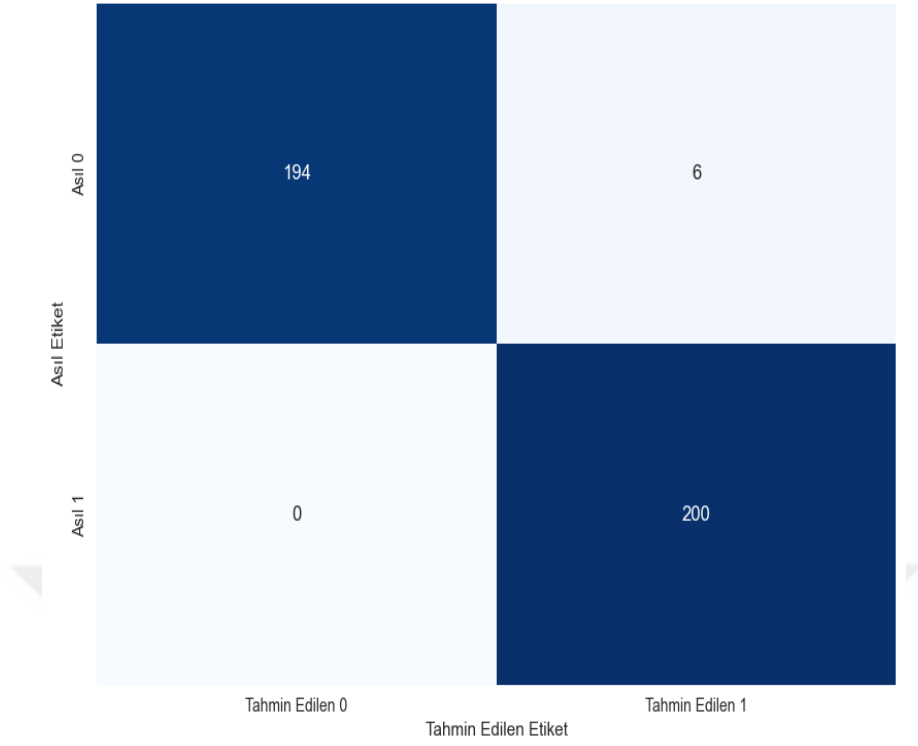
Şekil 4.1. Tepe sınıflandırıcı için karmaşıklık matrisi. TN:194,TP: 200, FN:0, FP:6

LR modeli, %99.00 doğruluk, %98.04 kesinlik, %100.00 hassasiyet ve %99.01 F1 skoru ile sonuçlanmıştır. Bu, modelin genel olarak mükemmel performans gösterdiğini, özellikle tüm pozitif tahminlerin doğru olduğunu (hassasiyet %100) belirtir.



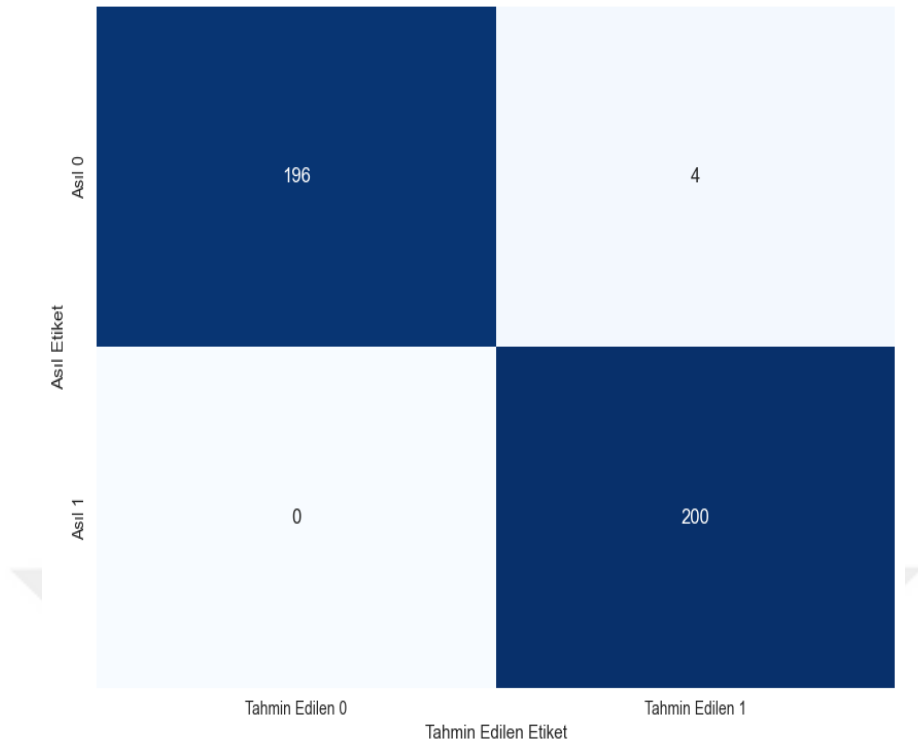
Şekil 4.2. LR için karmaşıklık matrisi. TN:196, TP:200, FN:0 FP:4

RF Sınıflandırıcısı da %98.50 doğruluk, %97.09 kesinlik, %100.00 hassasiyet ve %98.52 F1 skoru ile performans göstermiştir. Yine, model genel olarak yüksek performans göstermiş ve tüm pozitif tahminler doğru çıkmıştır.



Şekil 4.3. RF sınıflandırıcısı için karmaşıklık matrisi. TN:194, TP:200, FN:0 FP:6

Son olarak, KNN modeli %99.00 doğruluk, %98.04 kesinlik, %100.00 hassasiyet ve %99.01 F1 skoru ile sonuçlanmıştır. Bu model de yüksek performans göstermiş ve tüm pozitif tahminler doğru çıkmıştır.



Şekil 4.4. KNN sınıflandırıcısı için karmaşıklık matrisi. TN:196, TP:200, FN:0 FP:4

Genel olarak, bu modellerin hepsi yüksek doğruluk, kesinlik ve hassasiyet oranlarına sahiptir. Bunun yanı sıra, F1 skorları da yüksektir, bu da modellerin hem pozitif hem de negatif sınıfları doğru bir şekilde tahmin etme yeteneğini gösterir. Ancak, hassasiyetin %100 olması, modellerin tüm pozitif tahminleri doğru yapma eğiliminde olduğunu belirtir. Bu durum, modellerin olası yanlışlıkla ilgili sorunları olabileceğine işaret edebilir. Bunun nedeni, tüm pozitif tahminlerin doğru olması, genellikle tüm negatif durumların doğru bir şekilde tahmin edilemediği anlamına gelir. Tüm bu sonuçlar ana veri seti için Çizelge 4.1. ile verilmiştir.

Çizelge 4.1. Farklı sınıflandırıcılar için ana veri setinde sınıflandırma sonuçları.

Modeller	Doğruluk %	Kesinlik %	Hassasiyet %	F1 Skoru %
Tepe Sınıflandırıcısı	98.50	97.09	100.00	98.52
LR	99.00	98.04	100.00	99.01
RF	98.50	97.09	100.00	98.52
KNN	99.00	98.04	100.00	99.01

Çizelge 4.2. Farklı sınıflandırıcılar için ana veri setinde sınıflandırma sonuçları (BFKL olmadan).

Modeller	Doğruluk %	Kesinlik %	Hassasiyet %	F1 Skoru %
Tepe Sınıflandırıcısı	98.75	99.49	98.00	98.74
LR	98.50	98.99	98.00	98.49
RF	97.50	96.57	98.50	97.52
KNN	97.75	96.59	99.00	97.78

Çizelge 4.2. ile ana veri setinde BFKL olmadan DenseNet-121 ve makine öğrenmesi sınıflandırıcıları ile elde edilen başarı tablosu verilmiştir. Sonuçlar incelendiğinde BFKL olmayınca başarı sonuçlarının göreceli olarak düştüğü görülmektedir. Bu sonuçlar da, BFKL katmanının modelin başarısını yükseltmede başarılı olduğunu sergilemektedir. Çizelge 4.3. ana veri seti için verilmiştir. Bu çizelgede Grad-CAM ile DenseNet121'in makine öğrenmesi sınıflandırıcıları ile birlikte kullanılmasıyla alınan sonuçlar mevcuttur.

Çizelge 4.3. Farklı sınıflandırıcılar için ana veri setinde sınıflandırma sonuçları (Grad-CAM ile).

Modeller	Doğruluk %	Kesinlik %	Hassasiyet %	F1 Skoru %
Tepe Sınıflandırıcısı	98.50	96.55	98.00	97.27
LR	97.00	96.53	97.50	97.01
RF	97.00	95.63	98.50	97.04
KNN	96.50	95.59	97.50	96.53

Tüm bunların yanı sıra, modelin ne kadar genel bir kapsamda veriye hitap ettiğini sınamak için 2000 görüntüden oluşan veri seti sadece test amaçlı kullanılmıştır. Bu veri seti için ise sonuçlar Çizelge 4.4.'teki gibidir. Çizelge 4.5. ile veri setinde BFKL olmadan DenseNet-121 ve makine öğrenmesi sınıflandırıcıları ile elde edilen başarı tablosu verilmiştir. Sonuçlar incelendiğinde BFKL olmayınca başarı sonuçlarının farklı bir veri setinde de göreceli olarak düştüğü görülmektedir. Bu sonuçlar da, BFKL katmanının modelin genelleştirme başarısını yükseltmede başarılı olduğunu sergilemektedir. Çizelge 4.6.'de ise Grad-CAM ile

DenseNet121'in makine öğrenmesi sınıflandırıcıları ile birlikte kullanılmasıyla veri setinde elde edilen sonuçlar mevcuttur.

Çizelge 4.4. Farklı sınıflandırıcılar için veri setinde sınıflandırma sonuçları.

Modeller	Doğruluk %	Kesinlik %	Hassasiyet %	F1 Skoru %
Tepe Sınıflandırıcısı	76.25	79.53	70.70	74.85
LR	76.00	79.35	70.30	74.55
RF	93.15	94.53	91.60	93.04
KNN	79.10	80.19	77.30	78.72

Çizelge 4.5. Farklı sınıflandırıcılar için veri setinde sınıflandırma sonuçları (BFKL olmadan).

Modeller	Doğruluk %	Kesinlik %	Hassasiyet %	F1 Skoru %
Tepe Sınıflandırıcısı	66.55	92.33	36.10	51.91
LR	67.85	89.93	40.0	55.56
RF	89.50	89.11	90.00	89.55
KNN	76.20	77.93	73.10	75.44

Çizelge 4.6. Farklı sınıflandırıcılar için veri setinde sınıflandırma sonuçları (Grad-CAM ile).

Modeller	Doğruluk %	Kesinlik %	Hassasiyet %	F1 Skoru %
Tepe Sınıflandırıcısı	69.55	75.23	58.30	65.69
LR	69.65	75.35	58.40	65.80
RF	90.25	91.54	88.70	90.10
KNN	73.85	75.18	71.20	73.14

Çizelge 4.4. incelediğinden en yüksek başarıları veren modellerin sırasıyla RF ve KNN olduğu görülmektedir. %93.15'lik bir doğruluk oranıyla RF, diğer sınıflandırıcılara karşı üstün gelmiştir. Diğer sınıflandırıcılar incelendiğinden doğruluk oranlarının %75-%80 arasında kalması, her ne kadar göreceli olarak ana veri setine göre düşük çıksa da büyük bir veri setinden ve farklı bir veriden kaynaklandığı düşünülebilir. Potansiyel bir sahte görüntü tespit sistemi böyle bir

olasılıkla bile her 10 görüntüden 7'sini rahatlıkla doğru ayırt edebilecektir. RF ile bu sayı 9'a çıkmaktadır. Çizelge 4.5. ve Çizelge 4.6. için de benzer bir durum söz konusudur. Grad-CAM tarafından üretilen haritalar Şekil 4.5.'te verildiği gibidir. Bu haritalarda, yüz ve yüz etrafındaki bölgelerin dikkat mekanizmasına daha çok yakalandığı görülebilmektedir.



Şekil 4.5. Grad-CAM sonuçları.

Makine öğrenmesi (ML), algoritmaların verilerden otomatik olarak öğrenmesini ve deneyimlerinden gelişmesini sağlar, bu da sahte yüz tespiti, modelin daha fazla sahte yüz gördükçe daha doğru tahminler yapma yeteneğine katkıda bulunur. Ayrıca, ML eğitim verilerinden genel modeller oluşturabilir (Albahar, Binsawad, Almalki, El-etriby ve Karali, 2020) ve böylece daha önce görülmemiş veriler üzerinde de tahminlerde bulunabilir. Ancak, ML modelleri doğru ve kaliteli veriye ihtiyaç duyar ve eğer veri seti hatalı veya yanıltıcıysa, bu modeller yanıltıcı sonuçlar verebilir. Ayrıca, bazı ML modelleri (örneğin, rastgele ormanlar) karmaşık olabilir ve bunların parametrelerini ayarlamak zor olabilir. Öte yandan, transfer öğrenme (TL), önceden eğitilmiş modelleri kullanır, bu genellikle daha az veri ve daha hızlı eğitim süreleri gerektirir, çünkü modelin öğrenmesi gereken bilgi miktarı azalır. TL ayrıca genellikle geniş ve çeşitli veri setlerinde eğitilmiş olan modelleri kullanır, bu da modelin daha geniş bir yüz çeşitliliğini tanıma yeteneğini

artırır. Ancak, TL'nin dezavantajları arasında ezberleme (overfitting) riski bulunur; eğer önceden eğitilmiş model, yeni görev için mükemmel bir şekilde uyum sağlamıyorsa, modelin ezberlemeye gitme riski vardır. Ayrıca, TL modelleri önceden eğitilmiş modeller üzerine kurulduğu için, bazen özelleştirme veya ince ayar yapmak zor olabilir, özellikle modelin yeni görevi tam olarak öğrenmesi gerektiğinde bu yaşanabilir.

Bu avantajların ve dezavantajların yanında, model LR ve KNN kullanılarak %99.00 doğruluk oranına ulaşmıştır, bu sebeple sahte yüz tespitinde başarılı olduğu söylenebilir. Ayrıca RF modelinin genelleştirmede %93.15 doğruluk oranı ile daha başarılı olduğu da gözlemlenmiştir.

5. SONUÇLAR ve ÖNERİLER

Sonuçlar incelendiğinde, "sahte yüz tespiti transfer öğrenme ve makine öğrenimiyle" ilgili yöntemin oldukça başarılı olduğu görülmektedir. Tüm modeller yüksek doğruluk ve kesinlik oranları elde etmiştir. Bu, sahte yüzleri tespit etme yeteneğine sahip olduklarını göstermektedir. Hassasiyet ve F1 skoru değerleri ise %100'e yakın, bu da modelin sahte yüzleri doğru bir şekilde tanımladığını ve yanlış pozitif sonuçlar üretmediğini göstermektedir. Transfer öğrenme ve makine öğrenimi kullanmanın bazı avantajları vardır. İlk olarak, transfer öğrenme, daha önce eğitilmiş bir modelin bilgi ve becerilerini hedef probleme uygulayarak zaman kazandırır. Bu sayede daha küçük veri setleriyle bile yüksek performans elde edebiliriz. Bu avantaj, sahte yüz tespiti gibi karmaşık bir görevde oldukça değerlidir, çünkü geniş ve çeşitli bir veri seti toplamak zor olabilir. İkinci olarak, makine öğrenimi modelleri, gerçek zamanlı uygulamalarda kullanılabilecek hızlı ve etkili bir şekilde çalışabilir. Sahte yüz tespiti, özellikle güvenlik sistemlerinde veya otomatik kimlik doğrulama uygulamalarında gerçek zamanlı yanıtlar gerektiren bir alandır. Bu yöntemler, bu tür talepleri karşılamak için uygun olabilir.

Ancak, bu yaklaşımların bazı dezavantajları da vardır. İlk olarak, transfer öğrenme, orijinal modelin farklı bir problem alanıyla ilgili olduğu durumlarda performansını sınırlayabilir. Sahte yüz tespiti gibi görevlerde, genel bir veri seti yerine daha spesifik bir veri setine ihtiyaç duyulabilir. Bu nedenle, uygun veri setini toplamak ve transfer öğrenme modelini uyarlamak için çaba harcamak önemlidir. İkinci olarak, makine öğrenimi modelleri, yanlış pozitif veya yanlış negatif sonuçlar üretebilir. Bu durumda, sahte yüz tespitiyle ilgili sistemlerde yanlış pozitif sonuçlar, gerçek kullanıcılara erişimi engelleyebilir veya hatalı bir şekilde uyarı verebilir. Bu nedenle, modelin doğruluk oranını artırmak için sürekli olarak eğitim ve iyileştirme yapılması önemlidir. Gelecekte, sahte yüz tespiti alanında transfer öğrenme ve makine öğrenimi yaklaşımlarının daha fazla kullanılması önerilebilir. Bu yöntemler, sahte yüzleri tespit etme becerilerini geliştirmek ve gerçek zamanlı uygulamalarda kullanılabilirliklerini artırmak için daha fazla araştırma ve geliştirme

gerektirmektedir. Ayrıca, daha spesifik ve geniş kapsamlı veri setleri toplanması, modelin performansını artırmak için önemlidir. Ayrıca, yanlış pozitif sonuçların azaltılması ve yanlış negatif sonuçların en aza indirgenmesi için sürekli olarak modelin güncellenmesi ve iyileştirilmesi önemlidir.



KAYNAKLAR

- ALBAHAR, M. A., BİNSAWAD, M., ALMALKİ, J., EL-ETRİBY, S., and KARALİ, S. (2020). Improving intrusion detection system using artificial neural network. *International Journal of Advanced Computer Science and Applications*, 11(6), 578–588. doi:10.14569/IJACSA.2020.0110670
- ALTMAN, N. S. (1992). An introduction to kernel and nearest-neighbor nonparametric regression. *American Statistician*, 46(3), 175–185. doi:10.1080/00031305.1992.10475879
- ARPİT, D., WU, S., NATARAJAN, P., PRASAD, R., and NATARAJAN, P. (2013). Ridge Regression based classifiers for large scale class imbalanced datasets. *Proceedings of IEEE Workshop on Applications of Computer Vision* (ss. 267–274). doi:10.1109/WACV.2013.6475028
- ARUNKUMAR, P. M., SANGEETHA, Y., RAJA, P. V., and SANGEETHA, S. N. (2022). Deep Learning for Forgery Face Detection Using Fuzzy Fisher Capsule Dual Graph. *Information Technology and Control*, 51(3), 563–574. doi:10.5755/j01.itc.51.3.31510
- BALASUBRAMANIAN, S., and TROJOVSKÝ, P. (2022). Deep fake detection using cascaded deep sparse auto-encoder for effective feature selection. *PeerJ Computer Science*, 8, e1040. doi:10.7717/peerj-cs.1040
- BASHA, S. H. S., DUBEY, S. R., PULABAİGARI, V., and MUKHERJEE, S. (2020). Impact of fully connected layers on performance of convolutional neural networks for image classification. *Neurocomputing*, 378, 112–119. doi:10.1016/j.neucom.2019.10.008
- BENET, D., and PELLICER-VALERO, O. J. (2022). Artificial intelligence: the unstoppable revolution in ophthalmology. *Survey of Ophthalmology*. doi:10.1016/j.survophthal.2021.03.003
- BREİMAN, L. (1996). Bagging predictors. *Machine Learning*, 24(2), 123–140. doi:10.1007/bf00058655
- BREİMAN, L. (2001). Random Forests. *Machine Learning*, 45(1), 5–32. doi:10.1023/A:1010933404324
- CAO, J., Lİ, Y., SUN, M., CHEN, Y., LİSCHİNSKİ, D., COHEN-OR, D., ... TU, C. (2022). DO-Conv: Depthwise Over-Parameterized Convolutional Layer. *IEEE Transactions on Image Processing*, 31, 3726–3736. doi:10.1109/TIP.2022.3175432

- CARVALHO, D. V., PEREIRA, E. M. ve CARDOSO, J. S. (2019). Machine Learning Interpretability: A Survey on Methods and Metrics. *Electronics*, 8(8), 832. doi:10.3390/electronics8080832
- CHEN, B., LIU, X., ZHENG, Y., ZHAO, G., and SHI, Y. Q. (2022). A Robust GAN-Generated Face Detection Method Based on Dual-Color Spaces and an Improved Xception. *IEEE Transactions on Circuits and Systems for Video Technology*, 32(6), 3527–3538. doi:10.1109/TCSVT.2021.3116679
- CHEN, S., SEDGHİ GAMECHİ, Z., DUBOST, F., Van TULDER, G., and de BRUIJNE, M. (2022). An end-to-end approach to segmentation in medical images with CNN and posterior-CRF. *Medical Image Analysis*, 76, 102311. doi:10.1016/j.media.2021.102311
- COVER, T. M., and HART, P. E. (1967). Nearest Neighbor Pattern Classification. *IEEE Transactions on Information Theory*, 13(1), 21–27. doi:10.1109/TIT.1967.1053964
- CORVÌ, R., COZZOLINO, D., ZINGARINI, G., POGGI, G., NAGANO, K., and VERDOLIVA, L. (2023, June). On the detection of synthetic images generated by diffusion models. In *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 1-5). IEEE.
- COX, D. R. (1959). The Regression Analysis of Binary Sequences. *Journal of the Royal Statistical Society: Series B (Methodological)*, 21(1), 238–238. doi:10.1111/j.2517-6161.1959.tb00334.x
- CRAMER, J. S. (2003). The Origins of Logistic Regression. *SSRN Electronic Journal*. doi:10.2139/ssrn.360300
- DANG, L. M., MIN, K., LEE, S., HAN, D., and MOON, H. (2020). Tampered and Computer-Generated Face Images Identification Based on Deep Learning. *Applied Sciences*, 10(2), 505. doi:10.3390/app10020505
- DEEPA, N., PRABADEVI, B., MADDIKUNTA, P. K., GADEKALLU, T. R., BAKER, T., KHAN, M. A., and TARIQ, U. (2021). An AI-based intelligent system for healthcare analysis using Ridge-Adaline Stochastic Gradient Descent Classifier. *Journal of Supercomputing*, 77(2), 1998–2017. doi:10.1007/s11227-020-03347-2
- DENG, X., ZHAO, B., GUAN, Z., and XU, M. (2023). New Finding and Unified Framework for Fake Image Detection. *IEEE Signal Processing Letters*, 30, 90–94. doi:10.1109/LSP.2023.3243770

- DONG, C., KUMAR, A., and LIU, E. (2022). Think Twice Before Detecting GAN-generated Fake Images from their Spectral Domain Imprints. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (C. 2022–June, ss. 7855–7864). doi:10.1109/CVPR52688.2022.00771
- ECEMİŞ, E., GÜNER, K., KURAN, U., and KURAN, E. C. (2023). Evrişimli Sinir Ağlarında Transfer Öğrenmesi ile GAN tarafından Üretilen Sahte Görüntü Tespiti. Mühendislik Bilimleri ve Araştırmaları Dergisi. doi:10.46387/bjesr.1257332
- ELGELDAWİ, E., SAYED, A., GALAL, A. R., and ZAKİ, A. M. (2021). Hyperparameter Tuning for Machine Learning Algorithms Used for Arabic Sentiment Analysis. Informatics, 8(4), 79. doi:10.3390/informatics8040079
- ELHASSAN, A., AL-FAWA'REH, M., JAFAR, M. T., ABABNEH, M., and JAFAR, S. T. (2022). DFT-MF: Enhanced deepfake detection using mouth movement and transfer learning. SoftwareX, 19, 101115. doi:10.1016/j.softx.2022.101115
- FAVORSKAYA, M., and YAKİMCHUK, A. (2021). Fake Face Image Detection Using Deep Learning-Based Local and Global Matching. CEUR Workshop Proceedings (C. 3047, ss. 133–138). doi:10.47813/sibdata-2-2021-20
- FEZZA, S. A., OUIŞ, M. Y., KADDAR, B., HAMİDOUCHE, W., and HADİD, A. (2022). Evaluation of Pre-Trained CNN Models for Geographic Fake Image Detection. 2022 IEEE 24th International Workshop on Multimedia Signal Processing (MMSP) (ss. 1–6). IEEE. doi:10.1109/MMSP55362.2022.9949282
- FLICKR (Haziran 20, 2023). <https://www.flickr.com/>.
- FU, T., XIA, M., and YANG, G. (2022). Detecting GAN-generated face images via hybrid texture and sensor noise based features. Multimedia Tools and Applications, 81(18), 26345–26359. doi:10.1007/s11042-022-12661-1
- GOODFELLOW, I., POUGET-ABADİE, J., MİRZA, M., XU, B., WARDEFARLEY, D., OZAİR, S., ... BENGİO, Y. (2020). Generative adversarial networks. Communications of the ACM, 63(11), 139–144. doi:10.1145/3422622
- GUO, H., HU, S., WANG, X., CHANG, M.-C., and LYU, S. (2022). Robust Attentive Deep Neural Network for Detecting GAN-Generated Faces. IEEE Access, 10, 32574–32583. doi:10.1109/ACCESS.2022.3157297
- HE, K., ZHANG, X., REN, S., and SUN, J. (2016). Deep Residual Learning for Image Recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (ss. 770–778). IEEE. doi:10.1109/CVPR.2016.90

- HO, T. K. (1998). The random subspace method for constructing decision forests. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(8), 832–844. doi:10.1109/34.709601
- HOERL, A. E., and KENNARD, R. W. (1970). Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics*, 12(1), 55–67. doi:10.1080/00401706.1970.10488634
- HU, S., LI, Y., and LYU, S. (2021). Exposing GaN-generated faces using inconsistent corneal specular highlights. *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings (C. 2021–June*, ss. 2500–2504). doi:10.1109/ICASSP39728.2021.9414582
- HUANG, G., LIU, Z., VAN DER MAATEN, L., and WEINBERGER, K. Q. (2017). Densely connected convolutional networks. *Proceedings of the IEEE conference on computer vision and pattern recognition* (ss. 4700–4708).
- ISMANTO, H., HARTONO, H., and MARFAI, M. A. (2020). Classification tree analysis (CTA) of smoke detection using Himawari 8 satellite data over Sumatera–Borneo Island, Indonesia. *SN Applied Sciences*, 2(9), 1520. doi:10.1007/s42452-020-03310-z
- JIE, H. J., and WANDA, P. (2020). Runpool: A dynamic pooling layer for convolution neural network. *International Journal of Computational Intelligence Systems*, 13(1), 66–76. doi:10.2991/ijcis.d.200120.002
- KIM, E., and CHO, S. (2021). Exposing Fake Faces Through Deep Neural Networks Combining Content and Trace Feature Extractors. *IEEE Access*, 9, 123493–123503. doi:10.1109/ACCESS.2021.3110859
- KRIZHEVSKY, A., SUTSKEVER, I., and HINTON, G. E. (2012). 2012 AlexNet. *Advances In Neural Information Processing Systems*, 1–9.
- LECUN, Y., BOSER, B., DENKER, J. S., HENDERSON, D., HOWARD, R. E., HUBBARD, W., and JACKEL, L. D. (1989). Backpropagation Applied to Handwritten Zip Code Recognition. *Neural Computation*, 1(4), 541–551. doi:10.1162/neco.1989.1.4.541
- LEE, D. H., CHEN, K. L., LIOU, K. H., LIU, C. L., and LIU, J. L. (2021). Deep learning and control algorithms of direct perception for autonomous driving. *Applied Intelligence*, 51(1), 237–247. doi:10.1007/s10489-020-01827-9
- LEE, S., TARIQ, S., SHIN, Y., and WOO, S. S. (2021). Detecting handcrafted facial image manipulations and GAN-generated facial images using Shallow-FakeFaceNet. *Applied Soft Computing*, 105, 107256. doi:10.1016/j.asoc.2021.107256

- LĬ, S., DUTTA, V., HE, X., and MATSUMARU, T. (2022). Deep Learning Based One-Class Detection System for Fake Faces Generated by GAN Network. *Sensors*, 22(20), 7767. doi:10.3390/s22207767
- LĬU, H., ZHENG, S., HAO, S., and LĬ, Y. (2018). Multifeature Fusion Detection Method for Fake Face Attack in Identity Authentication. *Advances in Multimedia*, 2018, 1–10. doi:10.1155/2018/9025458
- SUNEETHA, M., PRASAD, M. V. D., and KĬSHORE, P. V. V. (2021). Multi-view motion modelled deep attention networks (M2DA-Net) for video-based sign language recognition. *Journal of Visual Communication and Image Representation*, 78, 103161.
- MALĬK, A., KURĬBAYASHĬ, M., ABDULLAHĬ, S. M., and KHAN, A. N. (2022). DeepFake Detection for Human Face Images and Videos: A Survey. *IEEE Access*. doi:10.1109/ACCESS.2022.3151186
- MĬSHRA, B., and SAMANTA, A. (2022). Quantum Transfer Learning Approach for Deepfake Detection. *Sparklinglight Transactions on Artificial Intelligence and Quantum Computing*, 02(01), 17–27. doi:10.55011/staiqc.2022.2103
- NORHĬKMAH, LUTFHĬ, A., and RUMĬNĬ. (2022). The Effect of Layer Batch Normalization and Dropout of CNN model Performance on Facial Expression Classification. *International Journal on Informatics Visualization*, 6(2–2), 481–488. doi:10.30630/joiv.6.2-2.921
- QAĬD, T. S., MAZAAR, H., AL-SHAMRĬ, M. Y. H., ALQAHTANĬ, M. S., RAWEH, A. A., and ALAKWAA, W. (2021). Hybrid Deep-Learning and Machine-Learning Models for Predicting COVID-19. *Computational Intelligence and Neuroscience*, 2021, 1–11. doi:10.1155/2021/9996737
- RANJĬTH KUMAR, M. V., PRABHU, A., ASTHANA, S., JAĬSWAL, S., and MADHAVAN, P. (2020). Denet: A deepfake visual media detection network. *Journal of Advanced Research in Dynamical and Control Systems*, 12(2), 792–799. doi:10.5373/JARDCS/V12I2/S20201098
- RAZA, A., MUNĬR, K., and ALMUTAĬRĬ, M. (2022). A Novel Deep Learning Approach for Deepfake Image Detection. *Applied Sciences (Switzerland)*, 12(19). doi:10.3390/app12199820
- REMYA REVĬ, K., ISAAC, M. M., ANTONY, R., and WĬLSCY, M. (2022). GAN-generated Fake Face Image Detection using Opponent Color Local Binary Pattern and Deep Learning Technique. *2022 International Conference on Connected Systems and Intelligence (CSI)* (ss. 1–7). IEEE. doi:10.1109/CSI54720.2022.9924077

- ŞAFAK, E., and BARIŞÇI, N. (2022). Detection of Fake Face Images Using Lightweight Convolutional Neural Networks. *El-Cezeri Journal of Science and Engineering*, 9(4), 1282–1289. doi:10.31202/ecjse.1133527
- SCOTT, A. J., HOSMER, D. W., and Lemeshow, S. (1991). Applied Logistic Regression. *Biometrics*, 47(4), 1632. doi:10.2307/2532419
- SELVARAJU, R. R., COGSWELL, M., DAS, A., VEDANTAM, R., PARİKH, D. ve BATRA, D. (2020). Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. *International Journal of Computer Vision*, 128(2), 336–359.
- SCHÄFER, MATTHİAS. 2021. “This Person Does Exist”. *Temes de Disseny* 37: 214-225. <https://doi.org/10.46467/TdD37.2021.214-225>
- SHARMA, J., SHARMA, S., KUMAR, V., HUSSEİN, H. S., and ALSHAZLY, H. (2022). Deepfakes Classification of Faces Using Convolutional Neural Networks. *Traitement du Signal*, 39(3), 1027–1037. doi:10.18280/ts.390330
- SİMONYAN, K., and ZİSSERMAN, A. (2015). Very deep convolutional networks for large-scale image recognition. 3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings.
- SİNGH, A., PRAKASH, B. S., and Chandrasekaran, K. (2017). A comparison of linear discriminant analysis and ridge classifier on Twitter data. *Proceeding - IEEE International Conference on Computing, Communication and Automation, ICCCA 2016* (ss. 133–138). doi:10.1109/CCAA.2016.7813704
- ST, S., AYOOBKHAN, M. U. A., V, K. K., BACANİN, N., K, V., Štěpán, H., and PAVEL, T. (2022). Deep learning model for deep fake face recognition and detection. *PeerJ Computer Science*, 8, e881. doi:10.7717/peerj-cs.881
- TAN, C., ZHAO, Y., WEİ, S., GU, G., and WEİ, Y. (2023). Learning on Gradients: Generativeized Artifacts Representation for GAN-Generated Images Detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 12105-12114).
- TANG, D. (2022). Optimization of Human Resource Management System Based on Data Mining Technology and Random Forest Algorithm. *Wireless Communications and Mobile Computing*, 2022, 1–10. doi:10.1155/2022/9442369
- This Person Does Not Exist. (2023). This Person Does Not Exist. <https://www.thispersondoesnotexist.com/>.
- TİAN, Y., ZHANG, Y., and ZHANG, H. (2023). Recent Advances in Stochastic Gradient Descent in Deep Learning. *Mathematics*, 11(3), 682. doi:10.3390/math11030682

- ULLAH, A., JAVAİD, N., JAVED, M. U., PAMİR, KİM, B. S., and BAHAJ, S. A. (2022). Adaptive Data Balancing Method Using Stacking Ensemble Model and Its Application to Non-Technical Loss Detection in Smart Grids. *IEEE Access*, 10, 133244–133255. doi:10.1109/ACCESS.2022.3230952
- WANG, J., ZENG, K., MA, B., LUO, X., YİN, Q., Liu, G., and JHA, S. K. (2022). GAN-generated fake face detection via two-stream CNN with PRNU in the wild. *Multimedia Tools and Applications*, 81(29), 42527–42545. doi:10.1007/s11042-021-11592-7
- XHLULU. (2019). Xhlulu. 140k Real and Fake Faces. <http://www.stat-help.com/notes.html>
- XUE, Z., JİANG, X., LİU, Q., and WEİ, Z. (2023). Global–Local Facial Fusion Based GAN Generated Fake Face Detection. *Sensors*, 23(2), 616. doi:10.3390/s23020616
- ZHUANG, F., Qİ, Z., DUAN, K., Xİ, D., ZHU, Y., ZHU, H., and HE, Q. (2021). A Comprehensive Survey on Transfer Learning. *Proceedings of the IEEE*. doi:10.1109/JPROC.2020.3004555