

# HorrorFace: Derin Öğrenme Tabanlı Korkutucu Yüzlerin Tespiti ve Sınıflandırılması

*Araştırma Makalesi/Research Article*

 Betül AY

Bilgisayar Mühendisliği, Fırat Üniversitesi, Elazığ, Türkiye

[betulay@firat.edu.tr](mailto:betulay@firat.edu.tr)

(Geliş/Received:07.02.2021; Kabul/Accepted:22.10.2021)

DOI: 10.17671/gazibtd.875816

**Özet**— Korkutucu yüzler, özellikle sağlıklı ergen ve çocuklarda beynin güçlü amigdala tepkisi ile birlikte duygusal hafıza gelişimini olumsuz etkilemektedir. İnternet kullanımının katlanarak arttığı ve denetimsiz görsel materyallerin hızla çoğaldığı günümüzde, korku riski oluşturan yüzlerin otomatik filtrelenmesi önemli bir problem olmuştur. Bu çalışmada derin öğrenme korkuyu öğrenebilir mi sorusunun cevabını arıyoruz; normal yüzleri korkunç yüzlerden ayıran bir ikili sınıflandırıcı inşa etmeyi hedefliyoruz. Literatürde bildiğimiz kadarıyla, açık kaynaklı bir veri kümesi olmadığı için HorrorFace adını verdiğimiz yeni ve erişilebilir bir veri kümesi sunuyoruz. HorrorFace veri seti korkutucu ve normal olmak üzere iki sınıfla etiketlenmiş 19,600 yüz görüntüsünden oluşmaktadır. Önerilen veri setinin doğruluğunu, güvenilirliğini ve genelleme yeteneğini kanıtlamak için, çeşitli yüz sınıflandırma görevlerinde başarısını kanıtlamış olan omurga konvolüsyonel sinir ağlarını kullanarak öğrenme aktarımı yönteminin gücünden faydalanıyoruz. Deneysel sonuçlar, en iyi derin öğrenme modelinin gerçekleştirdiği % 99.30 doğrulukla etkili ve sağlam bir tanıma performansına ulaşıldığını göstermektedir.

**Anahtar Kelimeler**— korkutucu yüzlerin sınıflandırılması, derin öğrenme, HorrorFace veri seti

## HorrorFace: Deep Learning-based Detection and Classification of Scary Faces

**Abstract**— Scary faces negatively affect emotional memory development, especially in healthy adolescents and children, with the brain's strong amygdala response. In today's world where internet usage is increasing exponentially and uncontrolled visual materials proliferate, automatic filtering of faces that pose a risk of fear has become an important issue. In this study, we are looking for the answer to the question of whether deep learning can learn fear; we aim to build a binary classifier that distinguish normal faces from horror faces. As far as we know in the literature, there is no open-source dataset related this domain. We introduce a new and publicly dataset that we call HorrorFace. HorrorFace dataset consists of 19,600 face images labeled with two classes, namely horror and normal. To prove the accuracy, reliability, and generalization ability of the proposed dataset, we are harnessing the power of the transfer learning technique using convolutional neural networks (CNN), which have proven successful in various face classification tasks. Experimental results show that an effective and robust recognition performance has been achieved with an accuracy of 99.30% carried out by the best deep learning model.

**Keywords**— classification of scary faces, deep learning, HorrorFace dataset

### 1. GİRİŞ (INTRODUCTION)

Amigdala, duygusal ve sosyal uyaranlara karşı duygusal hafızanın ve tepkilerin oluşmasında önemli bir rol oynayan beynin bade m şeklindeki bir bölümünü temsil etmektedir.

Beynin medial temporal lobunda bir dizi nöronlardan oluşan bu bölüm, başta korku denetimi olmak üzere çeşitli duyguların denetimden sorumludur [1]. Literatürde gerçekleştirilen beyin görüntüleme çalışmalarından elde edilen sonuçlar, sağlıklı yetişkin, ergen ve çocuklarda

amigdala'nın korkulu (korkutucu ve maskeli) yüzlere diğer ifadelerden (nötr veya mutlu yüzler gibi) daha güçlü tepki verdiğini ortaya koymuştur [2, 3]. Dahası anksiyete bozukluğu olan çocuklar, sağlıklı çocuklara kıyasla korkulu yüzlere abartılı bir amigdala tepkisi gösterebilmektedir [4].

korku ve stres gibi olumsuz uyaran durumlarda gözbebeğinin genişlediğini vurgulayarak, göz bebeği hareketleri temelli üç sınıf duygu (olumlu, olumsuz ve nötr) sınıflandırmasını, kNN, Naive Bayes, Destek Vektör Makinaları, Doğrusal Diskriminant analizi, karar ağacı ve lojistik regresyon teknikleri ile gerçekleştirmişlerdir [8].

Tablo 1. Derin öğrenme tabanlı örnek yüz sınıflandırma çalışmaları  
(Sample of face classification studies based on deep learning)

Referans Çalışma	Yıl	Sınıflandırma Türü	Kullanılan Yöntem	Kullanılan Veri Seti	Görüntü Sayısı	Sınıflar	Doğruluk Oranı
[8]	2015	Duygu	AlexNet, VGGNet	Fer2013 EmotiW	32,298 921	Sinirli, Üzgün, Şaşkın, Şaşkın, Natürel, Tiksinç, Mutlu, Korkmuş	%55.6
[9]	2017	Cinsiyet	CNN, Xception	IMDB	460,723	Kadın, Erkek	%96,00
[10]	2018	İrk	VGGNet	4race-50K	50,000	Siyahi, Beyaz, Hint, Asyalı	%97.6
[11]	2019	Yaş	GoogleLeNet, ResNet, VGGNet	Adience	26,000	Yaş Grupları: (0-2), (4-6), (8-13), (15-20), (25-32), (38-43), (60+)	%62.26
[14]	2019	Sahte Yüz	ResNet, InceptionV3	MFSD	15,545	Gerçek, Sahte	%97.52
[13]	2020	Medikal Maske	InceptionV3	SMFD	1570	Maskeli, Maskesiz	%100
Önerilen	2021	Korkutucu Yüz	ResNet, DenseNet VGGNet, MobileNet InceptionV3	HorrorFace	19,600	Korkutucu, Normal	%99.30

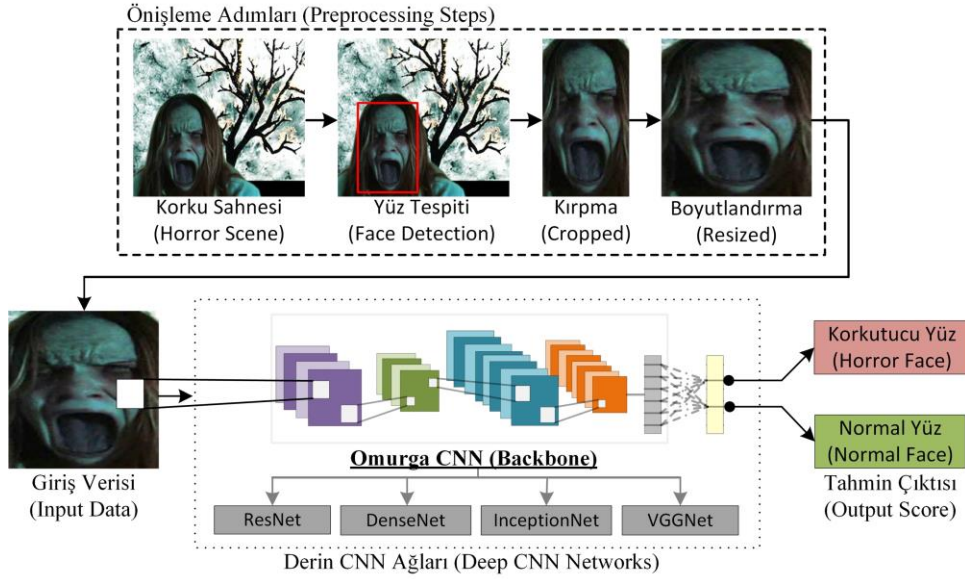
İnternetin ve kamera teknolojilerinin hızlı gelişimi birlikte dünyadaki görsel veri miktarı son on yılda büyük bir artış göstermiştir. Son yıllarda üretilen cep telefonlarının birden fazla kamera ile donatılmaktadır. Bu sensor artışından doğan doğal bir sonuç olarak dünyadaki ortalama insan sayısından daha fazla kameranın olduğu görülmektedir. Sensor sayısındaki bu artış ile birlikte muazzam bir genişlikte ve büyüklükte görsel veri üretmektedir. Yapılan istatistikler, toplam internet trafiğinin %80'inden fazlasının videolardan oluştuğunu göstermektedir, yalnızca Netflix film portalı ABD internet trafiğinin %20'sinden fazlasını kaplamaktadır [5]. 2019 yılında Youtube'a her dakikada yüklenen video sayısı 500 saatten fazla olmuştur [6]. Bir video görüntüsünün hersaniye 25 kare ya da resim aldığını farz edersek, yani 25 FPS (frame per second), 500 saatlik video 45 milyon görüntüye eş değer sayılmaktadır. Görüntü miktarındaki büyük artış ile birlikte, video görüntülerindeki korku sahneleri içeriklerinin özellikle korkulu yüzlerin özellikle çocuklar üzerindeki güçlü amigdala tepkisi endişe verici olmaktadır. Bu nedenle, korku içeriklerinin etkili bir şekilde tespit edilerek filtrelenmesi büyük önem arz etmektedir.

Duygular günlük rutinleri, dikkati, sosyal etkileşimleri, algıyı ve hafızayı etkileyen en önemli psikofizyolojik değişimdir. Duygular için en güçlü göstergelerden biri ise "yüzdür"; gözler, kaşlar, göz kapakları ve dudaklar gibi temel yüz özelliklerindeki değişikliklere dayanarak yüz ifadelerini incelemek, eylemlerin nedenleri hakkında bir fikir vermektedir. Bilgisayar tabanlı yüz ifadelerinin analizi, her türlü içeriğe yönelik filtrelenmemiş duygusal tepkiler yakaladığı için günümüzde çok güncel ve popüler bir çalışma alanıdır [7]. Mete ve arkadaşları tehlike halinde

Bilgisayar görmesi algoritmaları kullanarak otomatik yüz görüntülerinin analizi ve sınıflandırılması yapay zeka teknolojilerindeki ilerlemeler ile birlikte mümkün olmaktadır. Literatürde yüz görüntülerini kullanılarak duygu, yaş, cinsiyet, ırk gibi farklı sınıflandırma görevlerini gerçekleştiren pek çok çalışma yapılmıştır (Tablo 1'e bakınız). N.g ve arkadaşları CNN mimarilerini kullanarak öğrenme transferi (transfer learning) ile yüz görüntülerinden duygu sınıflandırması gerçekleştirmişlerdir [9]. Doğrulama setinde genel sınıflandırma performansları %48.5 ve test setinde ise %55.6 olarak hesaplanmıştır. Çalışmalarında ince nüanslı yüzlere doğru etiketin atanmasının zorluğundan ve bu zorluğun model performanslarını etkilediğini belirtmişlerdir. Duygu sınıflandırması ile birlikte cinsiyet sınıflandırması gerçekleştiren bir diğer çalışmada [10], IMDB cinsiyet veri seti üzerinde %96 doğruluk başarıları kaydedilmiştir. Eğitim setlerindeki yüzlerin çoğunluğu batılılardan oluştuğu için, yeniden eğitilen popüler CNN modellerinin batı yüz özelliklerine eğilim gösterdiğini ve yanlış bulunan yüzlerin bu eğilimden kaynaklandığını belirtmişlerdir. Önceden eğitilmiş CNN modellerini kullanan benzer çalışmada [11], 4 farklı ırk sınıfından oluşturulan 50,000 yüz görüntülerinden yeni bir veri seti önerilmiştir ve 16 katmanlı VGGNet mimarisi ile ortalama %97.6 sınıflandırma başarıları elde edilmiştir. Mallouh, Qawaqneh ve Barkana [12] gerçek dünya yüz görüntülerinden ile yaş sınıflandırma gerçekleştirmişler ve en iyi modeli elde etmek için önceden eğitilmiş CNN modellerinin performanslarını karşılaştırmıştır. Tüm modelleri birleştirmede tüm sınıflar için ortalama %62,26 tanıma başarıları kaydederken, eğitilen modeller (0-2) ve (65 yaş ve

üzeri) yaş gruplarında en iyi performansı sağlamıştır. Yüz tanıma sistemleri temel olarak, görüntülerdeki yüz bölgelerinin makine öğrenmesi algoritmaları kullanılarak tespit edilip daha önce sistemde kayıtlı kişiler ile eşleştirme yapılarak otomatik kimlik doğrulama sistemi olarak görev yapmaktadır. Ancak yüz tanıma sistemleri, kişinin yüz bilgisinin fotoğraf ya da video olarak ele geçirilmesi ve kişinin kayıtlı olduğu bu sisteme kişiyi taklit edecek şekilde yüz saldırılarına maruz kalabilmektedir. Bu problemin üstesinden gelmek için yapılan bir diğer çalışmada [13] gerçek ve sahte yüzleri ayırt etmede kullanılan ResNet ve InceptionV3 CNN mimarilerinin performansları değerlendirilmiştir. Çeşitli parametre

Yapılan önceki araştırmalar incelendiğinde derin sinir ağı modellerinin öğrenilmiş özelliklerini yeni bir problem için uyarlayarak öğrenme transferi gerçekleştirmek yüz görüntülerini sınıflandırmada oldukça yüksek bir başarı sağlamaktadır. Bu çalışmada, farklı sınıflandırma görevleri için eğitilmiş derin öğrenme modellerinin öğrenilmiş ağırlıkları kullanılarak yeni bir korkutucu yüz sınıflandırma modelinin inşa edilmesi amaçlanmıştır. Literatürde bildiğimiz kadarıyla korkutucu yüzlerin sınıflandırmasını yapan bir çalışma yoktur. MIT Media Lab. [18] tarafından bu çalışma konusuna benzer kabus makinesi (nightmare machine) isimli proje gerçekleştirilmiştir. Projede derin öğrenme algoritmaları



Şekil 1. Önerilen sistem için iş akış diyagramı  
(Workflow for proposed system)

ayarlamaları ile oluşturulan modellerden test seti üzerindeki en yüksek doğruluk oranına %97.52 ile ResNet152 mimarisi ulaşmıştır. Covid-19 salgını ve gelişen teknolojiler ile birlikte kamera görüntülerinden derin öğrenme yöntemleri ile yüz maskesinin tanınması sıcak bir çalışma alanı olmuştur. Chowdary ve arkadaşları [14], InceptionV3 modelini kullanarak öğrenme transferi ile ikili yüz sınıflandırma gerçekleştirmişler ve çok yüksek bir başarı elde etmişlerdir. Yadav ve Vishwakarma [15] film türlerini otomatik olarak tespit edebilmek amacıyla bir derin öğrenme çözümü sunmuşlardır. Yüksek seviyeli özellikleri çıkarabilmek için Inception V4, Bi-LSTM ve LSTM katmanlarının birleştirildiği bir derin mimari yardımıyla Hint filmlerinin otomatik olarak sınıflandırılması gerçekleştirilmiştir. Çalışmada ayrıca 6 tür film için 100 adet klipten oluşan bir veri seti sunulmuştur. Filmlerin türlerinin otomatik olarak tespiti konusunda yapılan çalışmaları inceleyen [16] makalede geleneksel ve derin öğrenme yaklaşımları ile bu alanda yapılan çalışmalar özetlenmiştir. Shambharkar ve Doja [17] yaptıkları çalışmada konvolüsyonel sinir ağları ve geyik avı optimizasyonu tekniği ile film kliplerinin sınıflandırmasını gerçekleştirmişlerdir. Çalışmada filmler 8 farklı sınıfta değerlendirilmiş ve insan hareketleri temelli olarak sınıflandırılmıştır.

kullanılarak korkutucu yüz ve korku sahneleri üretilmiştir. Açık kaynak bir veri seti olmadığı için, bu çalışmanın ilk aşaması geniş ölçekli ve çeşitliliği sağlayan yeni bir veri seti oluşturmak olmuştur. Korkutucu yüzlerin sınıflandırılması ve oluşturulan veri setinin doğrulanması için karşılaştırma performansı sağlamak amacıyla literatürde kullanılan omurga CNN mimarileri kullanılmıştır.

## 2. MATERYAL VE METOD (MATERIAL AND METHOD)

### 2.1. Korkutucu Yüzlerin Sınıflandırılması (Classification of Scary Faces)

Son yıllarda, ImageNet gibi mevcut veri setlerindeki artış, Tensorflow gibi derin öğrenme platformlarının ortaya çıkması, Grafiksel İşleme Birimi (GPU) ve Merkezi İşleme Birimlerinin (CPU) hesaplama gücündeki gelişmeler ile birlikte daha karmaşık görsel veriyi tanıma problemlerini çözmek ve daha derin katmanları eğitmek mümkün hale gelmiştir. Böylece pek çok derin sinir ağları (örneğin Konvolüsyonel Sinir Ağı-CNN) mimarilerinin oluşturulmasının önündeki engeller kalkarak, yapay zeka algoritmaları hızlı bir gelişim göstermiştir. 2010 yılından

beri düzenlenen ImageNet görsel tanıma yarışması [19] ile birlikte bugün bilinen ve öğrenme aktarımı için yaygın olarak kullanılan en yeni mimariler ortaya çıkmıştır. 2012 yılındaki en iyi buluş ve Toronto Üniversitesi'ndeki Alex Krizhevsky ve takım arkadaşları tarafından geliştirilen AlexNet mimarisidir [20]. 2014 yılında ise yarışmada iki mimari ön plana çıkmıştır. İlki Oxford Üniversitesi'nde görsel geometri grubu tarafından sunulan VGG mimarisidir [21]. Bu mimarinin VGG-16 ve VGG-19, sırasıyla 16 ve 19 katman sayısına sahip iki ağ mimarisi vardır. Yarışmanın ikinci kazananı ise Inception olarak bilinen GoogLeNet mimarisidir [22]. 2015 yılında ise Derin Kalıntı Ağları (Residual Network-ResNet) olarak adlandırılan ve Microsoft araştırma grubu tarafından akademik olarak yayınlanan bir ağ mimarisi bu yarışmayı kazanmıştır [23]. ResNet mimarisi, önceki CNN mimarilerinde yaşanan sıfırlanan gradyan (vanishing gradient) problemine kısa yol bağlantısı sunarak çözüm getirmiştir. DenseNet [24] kısa yol bağlantıların kullanmak yerine, önceki tüm özellik haritalarının bir sonraki katmanın girdisi olacak şekilde inşa ederek, ResNet'i daha da geliştirmiş ve derin ağlarda parametre sayısını düşük tutmayı başarmıştır.

Bu çalışmada video karelerinden korkutucu ve normal yüzlerin sınıflandırılması için literatürde başarısını ispatlamış ResNet, VGGNet, MobileNet, InceptionNet ve DenseNet omurga CNN mimarilerini kullanan bir sistem öneriyoruz. Önerilen sistemin iş akışı diyagramı Şekil 1'de verilmiştir. Sisteme verilen video karesinde öncelikle yüz algılama işlemi gerçekleştirilmektedir. Yüz algılama yani yüz tespit işleminde her bir yüzün koordinatları yani sınırlayıcı kutusu (bounding box) bulunur ve bu yüz kutusu kaydedilir. Sinir ağının girdi boyutu gereksinime göre kaydedilen yüzler yeniden boyutlandırılır, normalize edilir ve eğitilen derin öğrenme modeline girdi olarak verilir. CNN tabanlı ikili sınıflandırıcı model, verilen girdi (yüz görüntüsü) için tahmini sınıf etiketi ve olasılık puanı üretir.

## 2.2. Konvolüsyonel Sinir Ağları (Convolutional Neural Networks-CNN)

CNN özel bir sinir ağı türüdür; konvolüsyon işlemi ve sinir ağının birleşiminden oluşmaktadır. Katmanlı sinir ağı mimarisinin yapısındaki farklılıkların yanı sıra geleneksel sinir ağından farklı olarak konvolüsyon işlem adımını kullanmaktadır. Model girdisi için özellik vektörlerini değil iki boyutlu uzayda piksel değerleri daha iyi yakalanabilen matrisleri kullanmaktadır. Konvolüsyon işleminin temel amacı, bir değerler matrisi olarak düşünülebilen herhangi bir görüntüden özelliklerin (özellik haritaları ya da aktivasyon haritaları) çıkarılmasıdır. Bu işlem bir ya da birden fazla filtrenin, matrisler üzerindeki taranması ile gerçekleştirilir. Filtre (çekirdek), görüntü verilerinden daha küçük olan sabit değerlere ve boyutlara sahip bir matristir. Daha fazla filtre ile daha fazla özellik çıkarılabilmek mümkündür, ayrıca rastgele ağırlıklar ile değişen filtreler farklı özellik haritaları oluşturmaktadır [25]. Verilen bir  $x$  girdisi ve  $w$  filtresi için konvolüsyon işlemi sonrası elde edilen  $a$  özellik haritası, matematiksel ifade ile Denklem 1'de açıklanmıştır:

$$a(t) = (x \otimes w)(t) = \int_{-\infty}^{\infty} x(b)w(t-b)db \quad (1)$$

Tipik bir CNN mimarisinde, her konvolüsyon katmanını doğrusal olmayan bir aktivasyon fonksiyonu (örneğin ReLU) ve havuzlama (pooling) katmanı takip etmektedir. Konvolüsyon katmanı sonrasında oluşan özellik haritaları, aktivasyon katmanından geçerek havuzlama katmanına girdi olarak verilmektedir. Havuzlama katmanı gelen girdi boyutunu yani veri boyutunu değerlerin toplanması ve ortalamasının alınması gibi toplama işlemleriyle azaltır. En popüler havuzlama işlemi giriş matrisindeki en büyük değeri seçen maksimum havuzlama yöntemidir. Verilen girdi değerleri havuzlama katmanı sonrasında azaltılarak ağın parametre sayısı dolayısıyla hafıza kullanımı indirgenmiş olur. İşlenen ve indirgenen görüntü verileri düzleştirilir (flatten) ve düzleştirilen vektörler son adım olarak gelenek bir sinir ağına iletilir. Sınıf vektöründen oluşan nihai çıktı katmanındaki nöron sayısı sınıf sayısına eş değerdir ve tahmini sınıf olasılığını vermektedir. Özetle verilen bir giriş görüntüsüne ait değerler matrisi bir dizi katmandan geçirilerek, son çıktı katmanında sınıf puanlarına dönüştürülür.

## 2.3. Öğrenme Aktarımı (Transfer Learning)

İnsan beyi her şeyi sıfırdan öğrenmez, önceden öğrendiklerini aktararak yeni şeyler öğrenir. İnsan beynini taklit eden yapay sinir ağlarında ise her bir problem için yeni bir eğitim seti oluşturularak sıfırdan eğitim gerçekleştirilir. Sıfırdan eğitilen modeller yalnızca eğitildiği veri setine ve probleme bağlı olduğu için, genelleme yeteneğinde eksik kalırlar. Bir sinir ağının öğrendiği özellikleri, başka bir sinir ağına aktarma yöntemine öğrenme aktarımı denilmektedir. İnsan beynine benzer mantıkta çalışmayı amaçlayan öğrenme aktarımı yöntemi, popüler ve gelişmekte olan bir çalışma alanıdır. Dahası, başka görevler için eğitilmiş ağların öğrenilmiş özellikleri (ağırlıkları) yeni probleme aktarıldığı için, yeni geliştirilen modellerin ve oluşturulan veri setlerinin genelleme ve doğruluk yeteneğini ölçmede bir karşılaştırma (benchmarking) ölçütü sunmaktadır. Bu çalışmada sunulan ikili sınıflandırma hipotezini doğrulamak ve oluşturulan modellerin güvenilirliğini ispatlamak amacıyla literatürde karşılaştırma CNN modelleri olarak bilinen önceden eğitilmiş mimarileri kullanıyoruz. Öğrenme aktarımı ile daha az veri ve daha az hesaplama kaynağı kullanarak yeni bir sınıflandırıcı model inşa ediyoruz.

## 2.4. HorrorFace Veri Seti (HorrorFace Dataset)

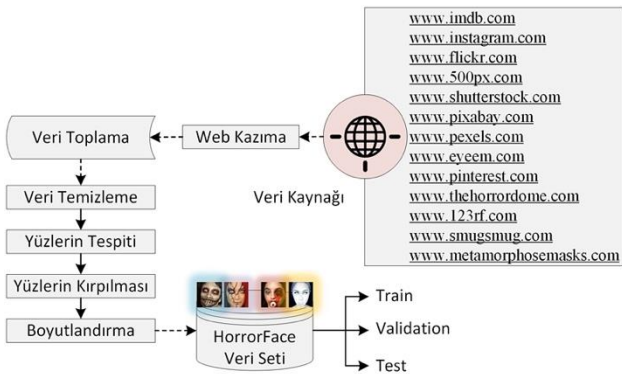
Optimal bir model, aşırı öğrenme ile ağ karmaşıklığı arasındaki dengeyi sağlayacak şekilde tasarlanmıştır. Derin sinir ağlarının karmaşıklığı arttıkça ise görüntü özelliklerini yakalamak ve daha başarılı sonuç elde etmek yüksek olasıdır. Dahası bir sinir ağının öğrenmesi ile insan beynin öğrenmesi arasındaki en temel farklardan birisi sinir ağının çok miktarda veriye ihtiyaç duymasıdır. Eğitilecek derin sinir ağlarında daha yüksek sınıflandırma

doğru elde etmek için bu çalışmada, büyük ölçekli ve kapsamlı yeni bir veri seti sağlıyoruz. HorrorFace adı verilen önerilen veri setinin istatistiksel dağılımı Tablo 2’de sunulmuştur. Veri seti, eğitim seti, eğitim süresince modellerin performansını takip etmek için doğrulama seti ve eğitilen modellerin gerçek doğruluk performansını ölçmek için test seti olmak üzere üç bölüme ayrılmıştır. Veri setinin hazırlanma süreçleri Şekil 2’de açıklanmıştır. Normal yüzler IMDb film sitesinden web kazıma aracı ile toplanırken, korkutucu yüzler sınıfı Şekil 2’de verilen çeşitli web adreslerinden çekilen görüntülerden oluşmaktadır.

Önerilen veri setinden örnek görseller Şekil 3’te verilmiştir: Korkutucu yüz sınıfındaki yüzler bilgisayar üretimi animasyon karakterlerini, makyajlı gerçek yüzleri, gerçekçi silikon ve lateks maskelerini içermektedir. Normal yüz sınıfındaki yüzler farklı cinsiyetten (kadın ve erkek), farklı yaş gruplarından (çocuk, genç, yetişkin, yaşlı), farklı aksesuarlardan (gözlük, şapka) ve farklı yüz özelliklerinden (sakal, bıyık, uzun, kısa saç ve benzeri) oluşmaktadır. Böylece veri setinin çeşitliliği sağlanmıştır. HorrorFace veri seti daha fazla araştırmaya teşvik etmek amacıyla Fırat Üniversitesi, Büyük Veri ve Yapay Zeka Laboratuvarı (BVYZLab) web adresinde [26] erişime açıktır.

Tablo 2. HorrorFace veri seti istatistiksel dağılımı (HorrorFace dataset statistics)

Özellik	Sınıf (Label)	
	Korkutucu Yüz (Horror)	Normal Yüz (Normal)
Eğitim	7440	7440
Doğrulama	1860	1860
Test	500	500



Şekil 2. HorrorFace veri seti hazırlama süreçleri (HorrorFace dataset preparation pipeline)

### 3. DENEYSEL SONUÇLAR (EXPERIMENTAL RESULTS)

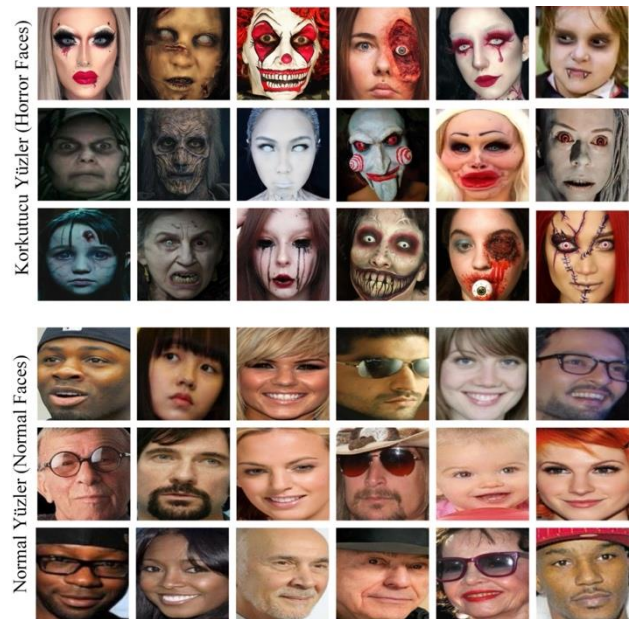
Bu çalışma kamera görüntülerinden korkutucu yüzlerin tespiti için omurga CNN mimarileri kullanılarak ikili sınıflandırma ağı eğitilmiştir. Model eğitimi ve parametre optimizasyonları için Tensorflow 2.3.0 kütüphanesi, Keras Fonksiyonel API ve Python programlama dili kullanılmıştır. Öğrenme aktarımı için kullanılan modeller Tensorflow Hub [27] geliştirme adresinde projelendirilmiştir. Modellerin eğitim sonuçlarını gözlemleyerek en iyi parametrelerin ve

optimizasyonların seçilmesinde web tabanlı TensorBoard uygulaması kullanılmıştır. Yüz tespiti için OpenCV Kütüphanesi, görselleştirme araçları için Matplotlib ve Scikit-learn kütüphaneleri tercih edilmiştir. Tüm sinir ağlarının eğitimi ve performans sonuçları Ubuntu 16.04 işletim sistemi yüklü, 24-core Intel Xeon E5-2628L CPU, 256 GB RAM, 8 NVidia GTX 1080-Ti GPU özellikleri sahip bir sunucu üzerinde gerçekleştirilmiştir.

#### 3.1. Performans Sonuçları (Performance Results)

Derin öğrenme modellerinin yanlılık/varyans kontrolünü sağlamak için bırakılma (drop-out) yöntemi kullanılmıştır. Bırakılma değeri 0.5 seçilerek, gizli katmandaki nöronların yaklaşık yarısı pasif hale getirilmiştir. Gizli katmanlardaki nöron (düğüm) sayısı 256’dır. Optimizasyon için hız ve doğruluk açısından genellikle SGD algoritmasından daha iyi performans gösteren Adam algoritması tercih edilmiştir. Öğrenme katsayısı (learning rate) değeri 0.001 alınmıştır. Eğitilen modellerin boyutları ResNet, InceptionNet, DenseNet ve VGGNet için sırasıyla 95MB, 88MB, 30MB ve 58MB.

Modellerin performansını değerlendirmek ve karşılaştırmak için bu çalışmada kullanılan performans metrikleri şunlardır: Precision (Kesinlik), Recall (Hassasiyet), TPR (Doğru Pozitif Oranı-True Positive Rate), FPR (Yanlış Pozitif Oranı- False Positive Rate), ROC Eğrisi (Curve of Receiver Operating Characteristic), AUC (Eğri Altında Kalan Alan- Area Under the Curve), F1-score (F1-ölçütü), Accuracy (doğruluk). Recall ve TPR aynı ölçüm değerini ifade etmektedir. Performans ölçümleri için hesaplanan TP (Doğru Pozitif-True Positive), FP (Yanlış Pozitif-False Positive), TN (Doğru Negatif-True Negative) ve FN (Yanlış Negatif-False Negative) temel metriklerdir.



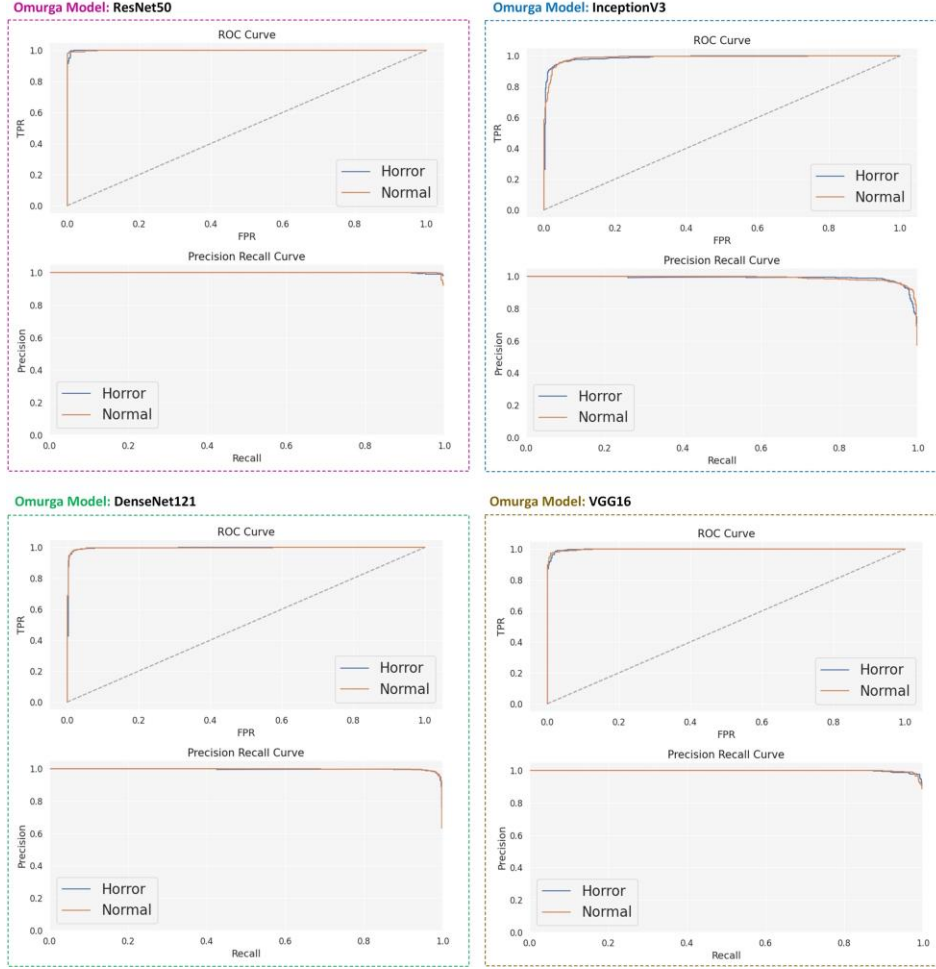
Şekil 3. HorrorFace veri seti örnek görseller (HorrorFace dataset samples)

Tablo 3. Performans sonuçlarının karşılaştırılması  
(Comparison of performance results)

Performans Ölçütü	Omurga Modeller (CNN Backbone)			
	ResNet	InceptionNet	DenseNet	VGGNet
Precision	0.996	0.950	0.982	0.978
Recall	0.990	0.956	0.982	0.982
AUC	0.999	0.989	0.997	0.998
F1-score	0.993	0.953	0.982	0.980
Accuracy (%)	99.30	95.30	98.20	98.00

ROC eğrisi farklı eşik ayarlarında sınıflandırma problemleri için bir performans ölçümü olarak kullanılır.

CNN tabanlı ResNet modeli %99 sınıflandırma başarısı ile en iyi model olarak seçilmiştir, InceptionNet modeli %95 sınıflandırma başarısı ile en kötü sınıflandırma performansı sergileyen model olmuştur, DenseNet ve VGGNet modelleri ise ortalama %98 başarı göstermiştir. Tablo 3 incelediğinde AUC değerleri ResNet, InceptionNet, DenseNet ve VGGNet için sırasıyla %99.3, %95.3, %98.2 ve %98'tir. Hem doğrulama veri setinde (hem de test veri setinde yüksek doğruluğa sahip olmak, genelleme yeteneğini yani modellerin görünmeyen örneklerle yüksek uyum sağlama yeteneğini ortaya koymaktadır. 30MB model boyutu ile DenseNet modeli



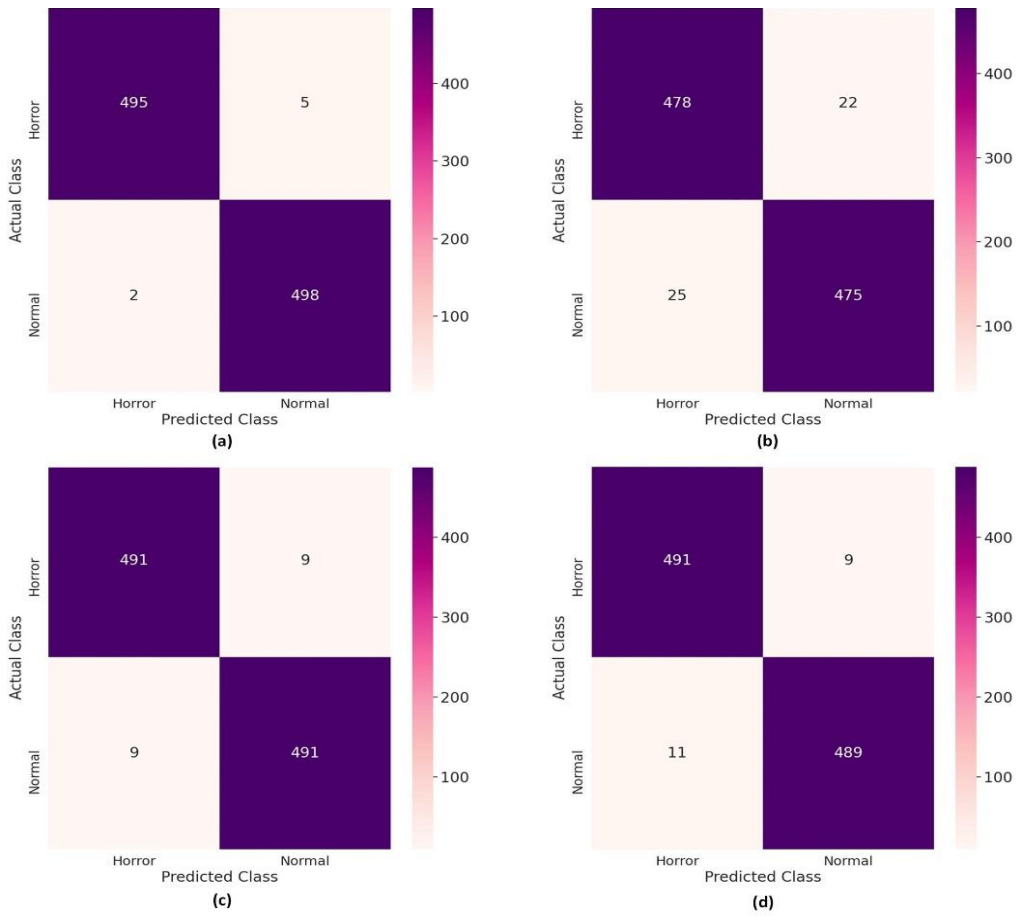
Şekil 4. Eğitilen modellerin değerlendirme seti üzerinde sınıflandırma sonuç grafikleri  
(Classification results of the trained models on the validation set)

ROC olasılık eğrileri eğitilen modellerin korkutucu ve normal sınıflar arasında ne kadar iyi ayrım yapabildiğini göstermektedir. Şekil 4'te, eğitilen modellerin doğrulama setinde sınıflandırma sonuçları ROC ve Kesinlik-Hassasiyet (Precision-Recall) Eğrileri ile grafiksel olarak sunulmuştur. ROC eğrilerinden gözlenebileceği gibi tüm sınıflandırıcıların doğru pozitif oranları sol üst köşede 1'e yakınsamaktadır. Sınıflandırıcılar arasında en iyi ayırt edebilme yeteneğini, yanlış pozitif oranı ile neredeyse 0 değerine yaklaşan ResNet50 modeli göstermiştir.

Eğitilen modellerin belirtilen değerlendirme ölçütleri ile performans karşılaştırma tablosu Tablo 3'te sunulmuştur.

diğer modellere oranla en hafif (light) modeldir. ResNet (95MB), InceptionNet (88MB) ve VGGNet (58MB) mimarilerine göre daha hafif bir model olmasına rağmen yüksek sınıflandırma doğruluğu elde etmiştir. Bu nedenle zaman maliyeti ve performans açısından en optimal model olarak değerlendirilebilir. Bu sonuçlara karşı en yüksek doğruluğa sahip olan 50 katmanlı ResNet modeli en ağır ve çıkarım süresi (inference time) en uzun sınıflandırıcıdır.

Toplam 1000 örnekten oluşan test seti üzerindeki hata matris görsel sonuçları Şekil 5'te gösterilmiştir. Şekil 5'te



(a) ResNet, (b) InceptionNet, (c) DenseNet, (d) VGGNet

Şekil 5. Test seti üzerindeki hata matris sonuçları  
(Confusion matrices on test set)

her bir test örneği için bilinen sınıf değerleri (actual class) ve bu bilinen sınıf değerlerine karşılık (ground-truth) her bir tahmini sınıf (predicted class) çıktısı görsel hata matrisi ile verilmiştir. Hatalı pozitif ve negatif sayısı en fazla InceptionNet mimarisi ile eğitilen sınıflandırıcı modelde görülmüştür. Görsel hata matrisi incelendiğinde en güvenilir ve doğru modelin %0.7 hata oranı ile ResNet mimarisi olduğu gözlemlenebilmektedir.

### 3.2 Çalışmanın Kısıtları (Limits of the Study)

Bu çalışmada kullanılan yaklaşım video karelerinin çıkarılarak her bir karede yüz tespiti yapılması ve bunların sınıflandırılması şeklindedir. Çalışmanın birinci kısıtı belirgin olmayan, çok küçük boyutta veya net olmayan sahnelerde yüzlerin tespit edilememesidir. Bu problemi aşabilmek için çok daha yüksek başarıyla yüz tespiti yapabilen modellerin araştırılması ve sisteme entegre edilmesi gerekecektir.

Çalışmanın ikinci kısıtı, videoların karelerine ayrıştırılarak işlenmesi esnasında geçen süredir. Özellikle yapay zeka modelinin her bir kare üzerinde çalıştırılması, bir videonun işlenmesi için gereken süreyi arttırmaktadır. Bu problemin çözümü için karelerin paralel olarak işlenmesi gerçekleştirilecektir. Aynı anda onlarca karenin paralel olarak yapay zeka modelinden geçirilebileceği bir mimari

yardımıyla videoların gerçek zamanından çok daha kısa bir sürede işlenebilmesi hedeflenmektedir. Ayrıca çok daha yüksek performanslı işleme için GPU'lardan yararlanılması planlanmaktadır.

### 4. GELECEK ÇALIŞMALAR (FUTURE WORK)

Bu çalışmada video karelerinin otomatik olarak çıkarılarak CNN tabanlı mimariler yardımıyla korkunç yüzlerin tespit edilmesi süreci açıklanmıştır. Yöntemin başarısı, çalışmanın farklı alanlarda yüksek doğrulukla kullanılabileceğini göstermektedir. Çalışmanın temel amaçlarından birisi özellikle çocukların internet ve televizyon üzerinden karşılaştıkları film-videoların uygunluğunun tespitinin yapılmasıdır. Bu amaçla elde edilen yapay zeka modelleri yardımıyla film veya kısa videoların uygunluğunun tespitini yapabilecek bir mimari oluşturulması hedeflenmektedir. Bu mimari, ilgili videonun alınarak karelere ayrıştırılması ve her bir karenin sınıflandırma modelinden geçirilerek puanlanması esasına dayanmaktadır. Video içerisinde korkunç yüzlerin tespit edilmesi durumunda veritabanına ilgili karenin saniyesi, içeriği ve tespit edilen yüzün korkunçluk derecesi veritabanına yazılacaktır. Böylece analiz edilen videonun yaş sınıflandırması sağlanacaktır. Televizyon yayınlarında kullanılan reyting sistemleri ülkelere göre farklılık gösterse

de hemen hepsinde çocukların izlemesine uygun içerikler özellikle belirtilmektedir [28]. Ancak ülkeler arası reyting sistemlerindeki farklılıklardan dolayı bir takım videolarda çocuklar için rahatsız edici içeriklerin çocuklara uygun olarak işaretlendiği veya az da olsa korku veya şiddet içeren sahneleri barındırdığı gözlenmektedir. Önerilen sistem ile bu tür problemlerin önüne geçilebilecektir.

Çalışmanın devamında sadece korkunç yüzlerin değil, korkutucu, cinsel veya çocuklar için uygun olmayan diğer tür içeriklerin de tespitini yapacak yapay zeka modellerinin geliştirilmesi ve sisteme entegre edilmesi planlanmaktadır.

#### 4. SONUÇ (CONCLUSION)

Bu çalışmada video karelerinde korkutucu yüzlerin tespit ve sınıflandırılması için derin öğrenme tabanlı bir yöntem sunulmuştur. Önerilen veri setinin ve sınıflandırıcı yöntemlerin gerçek başarısını ve genelleme yeteneğini değerlendirmek amacıyla omurga CNN mimarileri kullanılarak öğrenme aktarımı gerçekleştirilmiştir. Deneysel sonuçlar incelendiğinde, önceden eğitilmiş ve başarısını çeşitli sınıflandırma görevleri için ispatlamış CNN modellerinin öğrenilmiş özellikleri, korkutucu yüz sınıflandırma görevi için aktarıldığında omurga ResNet, InceptionNet, DenseNet ve VGGNet mimarilerinde, %99.3, %95.3, %98.2 ve %98 doğruluk oranları ile yüksek tanıma performansları elde edilmiştir. Performans sonuçları bu alandaki derin öğrenme modellerinin etkinliğinin yanı sıra, önerilen veri setinin güvenilirliğini ve doğruluğunu ortaya koymaktadır. Omurga DenseNet modeli ikinci yüksek sınıflandırma başarısına sahip olmasına rağmen, en hafif modeldir. Bu nedenle bu sınıflandırma çalışmasında en optimal model olarak düşünülmektedir. Bir sonraki çalışmada daha basit ve çıkarım süresi daha hızlı olan daha hafif mimariler ile daha yüksek başarı elde edebilen mimariler ve yöntemler üzerinde çalışılacaktır. Önerilen HorrorFace veri seti daha fazla çalışmaya teşvik etmek ve literatüre katkıda bulunmak amacıyla araştırmacılara açık olarak sunulmuştur.

#### KAYNAKLAR (REFERENCES)

- [1] T. Dalgleish, "The emotional brain", *Nature Reviews Neuroscience*, 5(7), 583–589, 2004.
- [2] J. S. Morris et al., "A differential neural response in the human amygdala to fearful and happy facial expressions", *Nature*, 383(6603), 812–815, 1996.
- [3] P. J. Whalen, S. L. Rauch, N. L. Etcoff, S. C. McInemey, M. B. Lee, and M. A. Jenike, "Masked presentations of emotional facial expressions modulate amygdala activity without explicit knowledge", *Journal of Neuroscience*, 18(1), 411–418, 1998.
- [4] K. M. Thomas et al., "Amygdala response to fearful faces in anxious and depressed children", *Arch. Gen. Psychiatry*, 58(11), 1057–1063, 2001.
- [5] X. Liu et al., "A case for a coordinated internet video control plane", *ACM SIGCOMM Conference on Applications, Technologies, Architectures, And Protocols For Computer Communication*, 359–370, 2012.
- [6] Internet: <http://www.statista.com/statistics/259477/hours-of-video-uploaded-to-youtube-every-minute>, 20.01.2021.
- [7] Y.-L. Tian, T. Kanade, and J. F. Cohn, "Facial expression analysis", *Handbook Of Face Recognition*, Springer, 247–275, 2005.
- [8] S. Mete, O. Çakır, O. Bayat, D.G. Duru, A.D. Duru, "Gözbebeği Hareketleri Temelli Duygu Durumu Sınıflandırılması", *Bilişim Teknolojileri Dergisi*, 13(2), 137-144, 2020.
- [9] H.-W. Ng, V. D. Nguyen, V. Vonikakis, and S. Winkler, "Deep learning for emotion recognition on small datasets using transfer learning", *ACM International Conference On Multimodal Interaction*, 443–449, 2015.
- [10] O. Arriaga, M. Valdenegro-Toro, and P. Plöger, "Real-time convolutional neural networks for emotion and gender classification", *arXiv Prepr. arXiv 1710.07557*, 2017.
- [11] M. Taló, B. Ay, S. Makinist, and G. Aydın, "Bigailab-4race-50K: Race Classification with a New Benchmark Dataset", *International Conference on Artificial Intelligence and Data Processing (IDAP)*, 1–4, 2018.
- [12] A. A. Mallouh, Z. Qawaqneh, and B. D. Barkana, "Utilizing CNNs and transfer learning of pre-trained models for age range classification from unconstrained face images", *Image Vis. Comput.*, 88, 41–51, 2019.
- [13] C. Nagpal and S. R. Dubey, "A performance evaluation of convolutional neural networks for face anti spoofing", *International Joint Conference on Neural Networks (IJCNN)*, 1–8, 2019.
- [14] G. J. Chowdary, N. S. Punn, S. K. Sonbhadra, and S. Agarwal, "Face mask detection using transfer learning of inceptionv3", *International Conference on Big Data Analytics*, 81–90, 2020.
- [15] A. Yadav, D.K. Vishwakarma, "A unified framework of deep networks for genre classification using movie trailer", *Applied Soft Computing*, 96, 2020, <https://doi.org/10.1016/j.asoc.2020.106624>.
- [16] P. G. Shambharkar, A. Anand and A. Kumar, "A Survey Paper on Movie Trailer Genre Detection", *2020 International Conference on Computing and Data Science (CDS)*, 2020, 238-244, doi: 10.1109/CDS49703.2020.00055.
- [17] P.G. Shambharkar, M.N. Doja, "Movie trailer classification using deer hunting optimization based deep convolutional neural network in video sequences", *Multimedia Tools and Applications*, 79, 21197–21222, 2020.
- [18] Internet: Nightmare Machine, <http://nightmare.mit.edu/>, 15.12.2020.
- [19] O. Russakovsky et al., "Imagenet large scale visual recognition challenge", *Int. J. Comput. Vis.*, 115(3), 211–252, 2015.
- [20] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks", *Commun. ACM*, 60(6), 84–90, 2017.
- [21] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition", *arXiv Prepr.* 1409.1556, 2014.



- [22] C. Szegedy et al., “Going deeper with convolutions”, **IEEE Conference on Computer Vision And Pattern Recognition**, 1-9, 2015.
- [23] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition”, **IEEE Conference on Computer Vision And Pattern Recognition**, 770-778, 2016.
- [24] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, “Densely connected convolutional networks”, **IEEE Conference on Computer Vision And Pattern Recognition**, 4700-4708, 2017.
- [25] B. Ay Karakus, **Derin öğrenme ve büyük veri yaklaşımları ile metin analizi**, Doktora Tezi, Fırat Üniversitesi, 2018.
- [26] Internet: B. Ay, BVYZLab, <http://buyukveri.firat.edu.tr/verisetleri/>, 01.02.2021.
- [27] Internet: Tensorflow Hub, <https://tfhub.dev/>, 20.12. 2021 .
- [28] Internet: Motion Picture Content System, Wikipedia Article, [https://www.wikiwand.com/en/Motion\\_picture\\_content\\_rating\\_system](https://www.wikiwand.com/en/Motion_picture_content_rating_system), 20.10.2021 .