



Video Duygu Analizi

Emre Arıĝ^{1*}, Metin Turan²

¹ İstanbul Ticaret Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği Bölümü, İstanbul, Türkiye (ORCID: 0000-0003-3645-8761)

² İstanbul Ticaret Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği Bölümü, İstanbul, Türkiye (ORCID: 0000-0002-1941-6693)

(Bu yayın 26-27 Haziran 2020 tarihinde HORA-2020 kongresinde sözlü olarak sunulmuştur.)

(DOI: 10.31590/ejosat.779059)

ATIF/REFERENCE: Arıĝ, E. & Turan M. (2020). Video Duygu Analizi. *Avrupa Bilim ve Teknoloji Dergisi*, (Special Issue), 59-67.

Öz

Bu çalışmada, videodaki görüntülerden tespit edilen insan yüzleri üzerinde CNN derin öğrenme modeli ile duygu analizi yapılmıştır. Bu analize ait sonuçlar saniye saniye kayıt edilerek bir duygu analizi grafiği çıkarılmıştır.

Çalışma 3 ana safhadan oluşmaktadır. İlk CNN modeli için gerekli duygu yüklü görsellerin bulunup etiketlenmesi, ikincisi duygu analizi yapabilecek bir CNN derin öğrenme modelinin oluşturulması ve üçüncüsü de videolardan yüz görüntülerinin tespit edilmesidir.

Eğitim veri seti oluşturmak amacıyla, seçilen 61 adet filmde binlerce yüz fotoğrafı analiz edilmiştir. Bunların arasında Bay Evet, Karabasan, Yaralı Yüz, Yedi Yaşam gibi farklı duyguların ağırlıklı olduğu filmler bulunmaktadır. İlk olarak 7 duygu türü için yüzler toplanmıştır. Bu duygular bıkkınlık, korku, mutluluk, sakinlik, şaşkınlık, sinirlilik ve üzünlüktür. Yüz tespiti kısmında Haarcascade tekniği kullanılmıştır. Tespit edilen yüzlerin duygulara göre etiketlenmesinde, Amazon webservisi olan Face Recognition'dan yardım alınmıştır. Çalışmada, 50 bin civarı yüz örnekleme elde edilmiştir. Ancak daha sonra yapılan kontrollerde Haarcascade ile bulunmuş görüntüler arasında yüz olmayan birçok görsel tespit edilerek çıkarılmıştır. Ayrıca, Amazon web servisinden dönen duygu analizlerinde %40 civarında yanlış duygu tespiti olduğu belirlenerek, eğitim veri setinden çıkarılmıştır. Tüm veri seti temizleme çalışmaları sonucunda 7 duygu için etiketlenmiş 20 bin fotoğraf elde edilmiştir.

Derin öğrenme sonucu, yapılan sınamalarda en çok karıştırılan 4 duygudan 2'sinin bıkkınlık ve şaşkınlık olduğu gözlemlenmiştir. Bıkkınlık sakinlikle, şaşkınlık ise korku yüz ifadeleri ile karışmaktadır. Elimizde kalan 5 duygu ile yapılan analizde, önerilen model ile %60'lık doğruluk değerine ulaşılmıştır. Videodan yüzleri çıkarıp modele gönderen ve bu sonuçlar ile bir duygu analizi grafiği çıkaran yazılımda, yüz tespitinin daha doğru olması için gerçek zamanlı analizde Haarcascade yöntemi yerine bir DNN modeli kullanılmıştır.

Anahtar Kelimeler: Duygu Analizi, Derin Öğrenme, Video İşleme, Yapay Sinir Ağları

Video Emotion Analysis

Abstract

In this study, emotional analysis was carried out with the CNN deep learning model on the human faces detected from the images in the video. The results of this analysis were recorded in seconds and an emotion analysis graph was created.

The study consists of 3 main stages. The first is to find and label the emotional images required for the CNN model, the second is to create a CNN deep learning model that can conduct emotion analysis, and the third is to identify the facial images from the videos.

In order to create a training data set, thousands of photographs from 61 selected films were analyzed. These include films with different feelings such as Yes Man, The Babadook, Scarface, Seven Pounds. First, faces were collected for 7 types of emotions. These feelings are boredom, fear, happiness, calmness, confusion, irritability and sadness. Haarcascade technique is used in the face detection section. Assistance was received from the Amazon Face Recognition web service for tagging detected faces according to emotions. In the study, about 50 thousand face samples were obtained. However, in the subsequent controls, many non-facial images were detected and removed from the images found with Haarcascade. In addition, approximately 40% false emotion detection was determined in the emotional analysis returned from the Amazon web service and removed from the training data set. As a result of the clearing of the whole data set, 20 thousand photos were tagged for 7 emotions.

As a result of the deep learning, it was observed that 2 of the 4 emotions most confused during the tests were boredom and confusion. Boredom is confused with calmness, and confusion with fear facial expressions. In the 5 emotion analysis that we have, 60% accuracy value has been reached with the proposed model. In the software that extracts faces from the video and sends them to the model and displays an emotional analysis graph with these results, a DNN model is used instead of Haarcascade method in real-time analysis to make the face detection more accurate.

Keywords: Emotion Analysis, Deep Learning, Video Processing, Artificial Neural Networks

1. Giriş

Gelişen teknoloji ve dijital hayatla birlikte veri kavramı giderek daha önemli hale gelmektedir. Veri artık sadece metin anlamına gelmemekte, fotoğraf, video gibi nesnelere artık dijital yapılarından dolayı işlenebildiklerinden bir tür veri olarak kabul edilmektedir. Bu nedenle günümüzün önemli çalışma ve araştırma alanlarından birisini, tartışmasız görüntü işleme ve yine bu temelden geliştirilen çalışmalar oluşturmaktadır.

Görüntü işleme, dijital ortamdaki görüntülerden yararlanarak ihtiyaca veya istenilen bir veriye ulaşmak için kullanılan yöntemlerdir. Bu yöntemler son zamanlarda artan işlemci güçleri ile daha da popülerleşmiştir. Görüntü işlemenin en popüler kütüphanesi OpenCV'dir. OpenCV, Intel tarafından 1999 yılında geliştirilen açık kaynak kodlu görüntü işleme kütüphanesidir [1]. C, C++, Python ve Java arayüzleri bulunan OpenCV, Windows, Linux, Mac OS, iOS, Android gibi farklı platformlarda çalışabilmektedir [2]. BSD (Berkeley Software Distribution) lisansı sayesinde her türlü projede ücretsiz olarak kullanılabilir. [3].

OpenCV, hesaplama verimliliğini artırmak ve gerçek zamanlı uygulamalar için tasarlanmıştır [3]. Nesnelere tespit edilmesi veya tanınması amacıyla yapılan çalışmalarda farklı yöntemler önerilmektedir. Nesnelere ait basit özellikler kullanılarak hızlı ve etkili nesne tanıma yönelik çalışmalar karmaşık arka plan çıkarımı ile tanıma, şekil tanıma, renk tanıma, kenar ve köşe tanıma, istatistiksel görüntü tanıma, şablon eşleme gibi çeşitli yöntemler kullanılmaktadır [4][5][6].

Kameradan alınan görüntülerin bilgisayarlarda işlenmesi ile nesne tanıma (object recognition), yüz bulma (face detection), yüz takibi (face tracking), yüz tanıma (face recognition), duygu tanıma (emotion recognition), hareket takibi ve tanıma (human action recognition and tracking) [7], cinsiyet/yaş tespiti (gender/age classification), işaret dili algılama (gesture recognition for sign language) [8], sürücü uykusu durumu tespiti (driver drowsiness detection)[9], bakış yönü tespiti (gaze detection) ve kafa yönü tahmini (head pose estimation) gibi birçok uygulama geliştirilmiştir. İnsan davranışlarının görüntü işleme teknikleri ile tespit edilip analiz edilmesi konusu son zamanlarda popülerlik kazanmış ve farklı alanlarda önemli çalışmalara zemin hazırlamıştır [10].

Görüntü işlemeye yönelik en popüler çalışmaların başında şüphesiz yüz tanıma teknolojisi bulunmaktadır. Yüz tanıma, kişinin yüzünü otomatik olarak algılayan ve analiz eden bilgisayar tabanlı bir güvenlik sistemidir. Yapılan çalışmalar, günümüzde sadece elektronik kartlara veya şifrelere bağımlı sistemlerin yeterli olmadığını, bu nedenle güvenliği arttırmak için biyometrik tanıma sistemlerinin iyi bir alternatif olacağını göstermektedir. Biyometrik sistemler kişiye özgü ham verileri işleyerek bireyi en iyi tanımlayan elektronik bilgiye çevirmektedir [11]. Bilgisayar destekli yüz tanıma konusunda yapılan ilk çalışma, 1964 ve 1965 yıllarında, Woodrow Wilson ve arkadaşları tarafından bilgisayarı insan yüzlerini tanıma kullanma üzerine yaptıkları araştırmadır [12]. Başta güvenlik uygulamaları olmak üzere, hukuki bilgi toplama, takip sistemleri, kimlik tespiti ve eğlence sektörü olmak üzere hızla artan bir kullanım alanına sahip olan yüz tanıma sistemleri, gelişen akıllı telefon ve tablet uygulamalarıyla da bütünleştirilerek yaygınlaşmaktadır [11]. Gerçek zamanlı görüntü işleme donanımlarının maliyetler açısından daha kolay ulaşılabilir olması, işlemci kapasitelerinin artması yüz tanıma sistemlerinin otomatikleşmesine ve diğer yazılım ağları ile bütünlük şeklinde çalışabilmesine olanak sağlamıştır [13].

Yüz tanıma konusunda son yıllarda yaygın olarak kullanılmaya başlanan bir diğer yöntem ise derin öğrenmedir. Derin öğrenme, makine öğrenme algoritmaları kullanılmak suretiyle doğrusal olmayan dönüşümler vasıtasıyla tanınacak olan veriyi en iyi temsil edecek özellikleri bulmayı amaçlayan bir yaklaşımdır. Derin öğrenme isimlendirmesindeki "derin" kavramı klasik yapay sinir ağlarına göre daha çok katmanlı bir yapıya dayalı bir öğrenmeyi belirtmektedir. Derin ağlar sayesinde görüntü işleme alanında oldukça büyük ilerlemeler kaydedilmiştir. Çok katmanlı sinir ağları görüntülerde bulunan nesnelere sınıflandırılması ve tespiti ile ilgili yapılan yarışmalarda normal makine öğrenmesi yöntemlerinden çok daha iyi derecelere alanının en iyisi olduğunu kanıtlamıştır [14]. Temelinde yapay sinir ağları olan derin öğrenme insan beyninin çalışmasını yapay sinir ağlarına göre daha iyi modelleyen yöntemler önerir. Derin öğrenmede kullanılan algoritmalarından birçoğu Geoffrey Hinton ve Toronto Üniversitesi'ndeki araştırmacıların yaptıkları çalışmalara dayanmaktadır. 1980'lerde öne sürülen bu algoritmalar yoğun matris işlemleri gerektirdiğinden ve büyük verilerin işlenmesinde yüksek işlemci gücüne ihtiyaç duyduğundan o yıllarda yaygın bir uygulama alanına kavuşmamıştır. Özellikle görüntü işleme alanında derin öğrenmenin özelleşmiş bir mimarisi olan Evrişimsel Sinir Ağları (Convolutional Neural Networks-CNN) başarılı sonuçlar vermekte ve son zamanlarda yaygın olarak kullanılmaktadır. ESA'da iki temel katman bulunmaktadır. Bunlardan biri evrişim diğeri ise havuzlama katmanıdır. ESA'lar bu iki katmanda temel bazı işlemler ile görüntünün önemli olan özelliklerini çıkarmayı amaçlamaktadır [15].

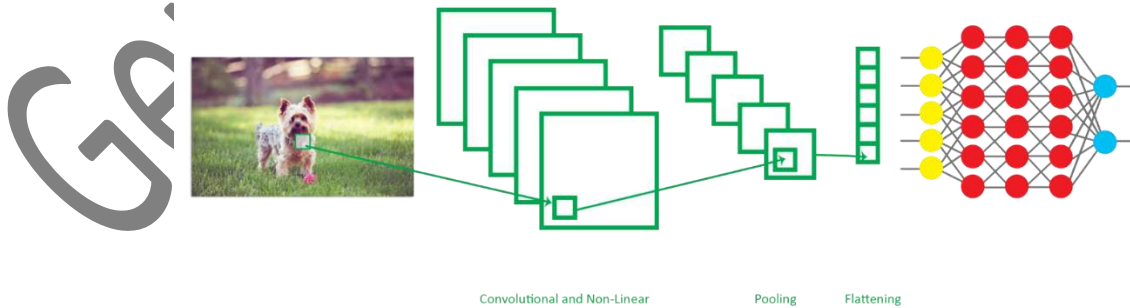
Videodan duygu analizi çalışması, yüz tespiti, görüntü işleme teknikleri ve sınıflandırıcıdan yararlanarak görüntüdeki insan yüzlerinden duygularının analizi yapabilmektir. İnsan duygularının tahmini oldukça zor bir konudur. İnsanların genel duygu durumları ya da bir olaya karşı verdikleri tepkiler yüzlerine yansımaktadır. Videolarda bu yüzler ise video içerikleri hakkında bizlere çeşitli bilgiler vermektedir. Bu analizler sayesinde çeşitli istatistiki bilgilere de ulaşılabilmektedir. Bu alan Amazon, Microsoft ve Google gibi büyük firmalarında öncülüğünü ettiği bir çalışma alanıdır.

Duygu analizine yönelik çalışmalar kronolojik olarak incelendiğinde, analizinin gelişimine önemli katkıları bulunan bazı çalışmalar öne çıkmaktadır. Valsar M. 2012 yılında yapmış olduğu çalışmada, SVM (Destek Vektör Makineleri) ve HMM (Gizli Markov Modeli) kullanarak geometrik öznitelik çıkarma yapmıştır. Yaptığı bu çalışma sonucunda %50,4 oranında tutarlılığa ulaşmıştır [16]. Bu sayede tespit edilen yüzlere ait duygular geometrik öznitelik çıkarma ile koordinat sisteminde ayrıştırılmaktadır. Gudi A. ise 2015 yılında, makine öğrenmesi temelli yaptığı çalışmada CNN mimarisini kullanmıştır. SEMAINE ve BP4D olarak iki ayrı veri seti üzerinde yaptığı çalışmada yüzden duygu analizi çıkarımında %52,2 ve %34,1 değerlerinde tutarlılık oranına ulaşmıştır [17]. Çalışmamız bu çalışma ile CNN mimarisinin kullanımı açısından benzerlik göstermektedir. Pooya K. 2016 yılında yaptığı çalışmada, CNN ve RNN (Giriş Geri Beslemeli) mimarilerini birlikte kullanarak %50,5 oranında bir tutarlılığa ulaşmıştır [18]. Çalışmasında yüzdeki belirli noktaları algılayarak duyguları tespit etmiştir. Nguyen ve arkadaşları tarafından 2018 yılında yapılan çalışmada ise önerdikleri model, çok seviyeli 18 katmanlı CNN modelidir ve VGG'ye (Çok katmanlı relu aktivasyon fonksiyonu tanımlı derin öğrenme modeli) benzemektedir. Bu modelde yüksek düzeyde özellikler bulunmasıyla beraber orta düzeyde özelliklerde bulunmaktadır. Çok seviyeli CNN modeli önerilmiş ve %69,21 hassasiyete ulaşmıştır. Duygu analizinde ulaştıkları bu oranı yükseltmek için bu modelde Fer2013 veri seti kullanmış ve %73,03 hassasiyete kadar ulaşmayı başarmıştır [19]. Güncel çalışmalardan bir diğeri ise Cao ve arkadaşlarının 2019 yılında yapmış olduğu çalışmadır. Bu çalışma, K-Means kümeleme ve Support Vector Machine (SVM) sınıflandırıcısını birleştirerek bir model oluşturmuştur. Bu modelde K-Means, CNN'nin başlangıç değerini belirler. SVM katmanları ise Fer2013 veri kümesindeki görüntüleri sınıflandırmak için eğitilmiş CNN modelinden özellikler almaktadır [20]. Bu çalışmayı öne çıkaran SVM ve CNN mimarilerinin bir arada kullanılmasıdır. Kronolojik olarak 2012 ve 2019 yılları arasında yapılan bu çalışmalar incelendiğinde, kullanılan veri setleri ve mimariye göre tutarlılık değerlerinin değişkenlik gösterdiği görülmektedir.

Videolar saniyede en az 24 fotoğraf karesinin arka arkaya gösterilmesi ile yapılır. Böylece göz bu sıralı görselleri hareketli olarak algılayarak video görüntüsü oluşur. Bu çalışmanın konusu olan videodan duygu analizinde amaç, her saniye akan bu karelerde yer alan insan yüzleri analiz edilerek bir duygu analizi grafiği ortaya koymaktır. Projede videodan duygu analizi 5 temel duygu için uygulanmıştır. Bu duygular, Korku, Mutluluk, Sakinlik, Sinirlilik ve Üzgünlüktür. Aslında insan yüzünde birden fazla duygu aynı anda betimlenebilmektedir. Bu nedenle bu çalışma en baskın olan duyguyu istatistiğe almak üzerine yapılmıştır. Bu sayede filmdeki karakterlerin istatistiki olarak duygularına ulaşılmıştır ve işlenen filmin türü tahmin edilmektedir.

2. Bölümde konu ile ilgili kavramlar, 3. Bölümde veri toplama, modelleme ve analiz aşamalarının detayları ve son olarak 4. Bölümde ise sonuç, kullanım alanları ve geliştirilmesine yönelik öneriler bulunmaktadır.

2. İlgili Kavramlar



Şekil 1. Convolutional Neural Network Mimarisi [21]

CNN yapısının genel olarak işleyişi, yapay sinir ağının öncesinde bulunan özellik çıkarımı katmanı, görüntü üzerinde belirlediğimiz parametrelere göre yapay sinir ağına gönderilecek ayrıştırıcı özellikleri elde etmektedir. Bu özellikler yapay sinir ağının eğitilmesini sağlamaktadır. Özellik çıkarım aşaması genellikle birden çok katmandan oluşur. Söz konusu katmanlar bu bölümde incelenecektir. GPU tabanlı yapılan bu uygulamada sonuçların doğruluk oranlarının yanında, yüksek öğrenme hızı da büyük bir avantajdır. CPU

tabanlı yapılan uygulamalara göre GPU tabanlı uygulamalar yaklaşık 10 ile 60 kat arasında daha hızlı bir öğrenme hızına sahiptir [22].

2.1. Konvolüsyon Katmanı (Convolutional Layer)

Özellik çıkarımı için en önemli katmandır denilebilir. Görüntü üzerinde önemli ayrıştırıcı bölgelerin belirlendiği katmandır. Belirlenen bir matris penceresi kullanılarak, örneğin (3x3) matrisi, görüntü üzerinde dolaşarak özellik bölgeleri tanımlanmaktadır. Bu aşamanın sonucunda yeni bir matris oluşmaktadır. Bu yeni matrise ise özellik haritası (Feature Map) denir. Genellikle birden çok konvolüsyon katmanı uygulanabilir.

2.2. Aktivasyon Katmanı (Activation Layer)

Aktivasyon katmanı, elde edilen özellik haritası üzerine bir aktivasyon fonksiyonu uygulanarak, özelliklerin regüle edilmesi sağlanır. Bu aşamada en yaygın kabul gören aktivasyon fonksiyonu 'relu' dur.

2.3. Havuzlama (Pooling Layer)

Havuzlama katmanında, bir matris filtresi bulunmaktadır. Bu filtre görüntü üzerinde gezerek matris değerlerini oluşturur ve belirlediğimiz havuz türüne göre maksimum ya da ortalama değeri elde edilir.

2.4. Vektörel Dönüşüm (Flattening Layer)

Vektörel dönüşüm katmanının temel görevi sinir ağının uçlarına iletilecek verinin uygun formata getirilmesidir. Buraya kadar elde edilen veri matris formatındadır, ancak sinir ağları genellikle tek boyutlu dizi ile beslenir. Bu katmanda matristen, vektöre dönüşüm sağlanır.

2.4. Tamamen Bağlı Ağ Katmanı (Fully Connected Layer)

Tamamen bağlı ağ katmanı son katmandır. Birbirlerine tamamen bağlı nöronlardan oluşan bir sinir ağı oluşur. Bu katmandan sonra sinir ağı karar verebilir hale gelmektedir.

3. Yapılan Çalışma

3.1. Eğitim Verilerinin Toplanması

Çalışmamızdaki eğitim ve test verilerimiz görsellerden oluşmaktadır. Bu görsellerin içeriği duyguların belirgin olduğu yüzlerdir. Bu görsellerin toplanabileceği en büyük kaynaklardan biri de film sektörüdür. Yüzler için belirgin duyguların baskın olduğu düşünülen filmlerden yüzler kesilerek alınmıştır. Bu aşamada süreci hızlandırmak adına bir çalışma yapılmıştır. Bu çalışmada filmler seçildikten sonra film sahnelerinden Haarcascade ile yüzler tespit edilmiştir. Bu tespit edilen yüzler çerçevelenerek kesilmiş ve film ismine açılan klasör içine kopyalanmıştır. Aynı zamanda bir dosya oluşturup bu yüzlerin filmin kaçınıcı saniyesinde olduğu ve filminden kesilen kaçınıcı fotoğraf olduğu görsel adı ile birlikte JSON objesi olarak kayıt edilmiştir.

Yüz tespiti ve kesim aşamaları tamamlandıktan sonrasında, duygu kategorizasyonu için bir Amazon Webservisi olan Face Recognition kullanılmıştır. Bu aşamada, yüz tespiti sırasında saniye saniye kayıt edilen görsel bilgilerinin bulunduğu dosyalar okunmaya başlanır ve kesilen yüzler ilgili web servisine gönderilir. Servisten dönen duygu sonuçlarına göre iki işlem yapılır. İlk olarak, görsel adı ve sırası aynı yüz tespitindeki gibi JSON objesi olarak bir dosyaya kayıt edilir. Böylelikle daha sonra başka bir analiz için kullanılabilmesi mümkün olan bir arşiv oluşur. İkinci olarak, bu sonuçlara göre en baskın duygu seçilerek ilgili duygu klasörüne kopyalanır. Bu klasör analiz yapılan tüm filmlerden görselleri barındırır. Görseller rastgele isimlendirilmiştir. Bu arşivde toplamda 61 adet filminden 50 bin civarı görsel bulunmaktadır.

Daha sonra görseller incelendiğinde bu kaba ayrımın yetersiz olduğu anlaşılmıştır. Haarcascade yüz olmayan bazı cisimleri yüz olarak algılayıp ayırmış ve servise göndermiştir. Bunun sonucunda duygu klasörlerinde yer alan bu hatalı görseller tek tek temizlenmiştir. Ayrıca, Amazon Servisi Olan Face Recognition duyguları %40 civarının yanlış etiketlediği tespit edilmiştir. Bunlarda arşivlerden bulunarak çıkarılmıştır. Son olarak karanlık, düşük çözünürlük ve bulanık olan görsellerde temizlendikten sonra, toplam 7 duygu için 20 bin civarı görsel kalmıştır.

En sonunda, eğitim ve test için gerekli olan ayrıştırma yapılmıştır. Bu amaçla, veri kümeleri duygulara göre kategorize edilmiş ve oluşturulan bu duygu kategorileri altında eğitim ve sınav için farklı klasörler oluşturulmuştur. Eğitim klasörleri altında yer alan örnekler modelin eğitimi için, sınav klasörleri altında yer alan örnekler modelin sınavması için kullanılır. Oluşturulan veri setinin her duygu türü için ortalama %80'i eğitim, %20'si test olarak ayrılan görseller gönderilmiştir. Örneklerin duygu bazında eğitim ve sınav amaçlı sayısal dağılımı Tablo 1'de görülmektedir.

Tablo 1. Duygu Bazında Eğitim ve Test Örneklerinin Sayısal Dağılımı

Duygular	Eğitim (Adet)	Test (Adet)
Korku	2582	682
Mutluluk	2806	610
Sakinlik	2666	668
Sinirlilik	2114	315
Üzgünlük	2328	576

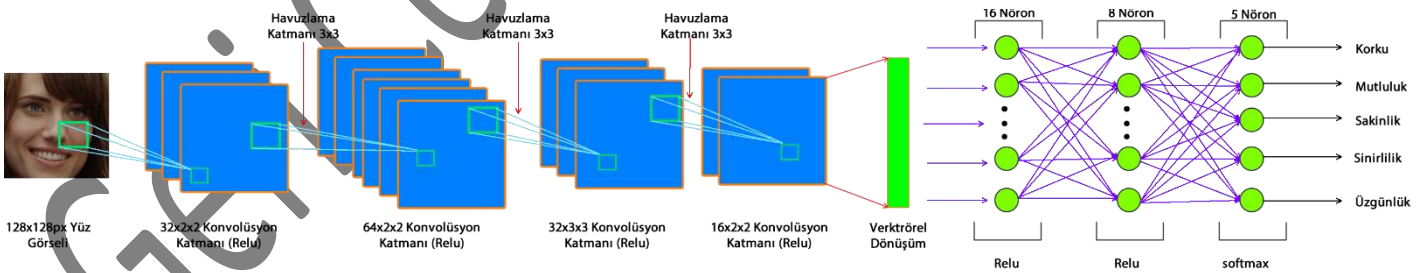
3.2. CNN Modeli Oluşturulması

Bu aşamada oluşturduğumuz yüz arşivi veri setimiz için CNN yapay sinir ağı oluşturulmuştur. Proje yazılımı Python dilinde geliştirilmiştir. Python dilinin tercih edilme sebeplerinin başında Tensorflow ve Keras gibi kütüphanelerinin kolayca eklenebilmesi, esnek kodlama imkanı ve verimli çalışması gelmektedir.

İlk olarak bir yapay sinir ağı modeli oluşturulmuştur. Model oluşturulurken katmanları biriktireceğimiz bir ardışık yığın oluşturulur. Daha sonra amacımıza yönelik olarak konvolüsyon katmanları eklenir. Belirlediğimiz parametrelerde, katmandaki düğüm sayısı, aktivasyon fonksiyonu ve ilk konvolüsyon katmanındaki giriş görsellerinin boyutu verilir. Modelde 32, 64, 34, 16 şeklinde giriş katmanı kullanılmıştır. Bu katmanlar arasında ise havuz katmanları eklenmiştir. Bu sayede farklı oranlara göre yeterince özellik çıkarılabilmektedir. Ara katmanlarda aktivasyon fonksiyonu olarak “relu”, çıkış katmanında ise “softmax” kullanılmıştır.

Görseller sinir ağına 128x128 piksel olarak gönderilmiştir. Bunun sebebi olarak, daha küçük görsellerde detayın kaybolabileceği, daha büyük boyutlarda ise eğitim için gerekli donanımın ihtiyacının artması ve performansın düşmesi verilebilir. Çıkış katmanlarında ise aşırı öğrenmeyi (overfit) önlemek için “dropout” yaklaşımı kullanılmıştır. Bu fonksiyon sinir ağının her adımında, fonksiyona girilen oran kadar sinir ağını kapatarak aşırı öğrenmeyi önlemektedir. Çıkış katmanları ise 16, 8, 5 şeklinde oluşturulmuştur. “loss” fonksiyonu için ise “categorical_crossentropy” kullanılmıştır. Bu yaklaşım sonucu, 5 farklı duygunun kayıp değeri en ideal şekilde hesaplanmaktadır. “optimizer” fonksiyonunda ise “sgd” kullanılmıştır. Tutarlılık ve kayıp değerlerini en verimli hesaplayan fonksiyonun bu olduğu tespit edilmiştir. Modelin başarı metriği olarak doğruluk (accuracy) hedeflenmiştir.

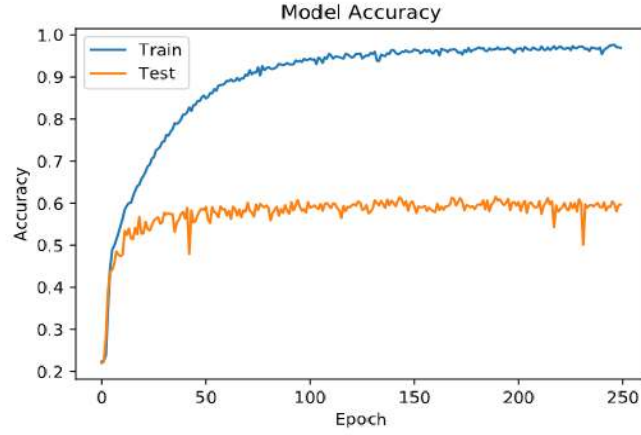
Bu parametrelerle yapılan denemelerden en başarılı sonuç veren 58 tanesi kodu ile birlikte kayıt altına alınmıştır. Bu denemelerin ışığında en yüksek tutarlılık ve en düşük kayıp (loss) değerine sahip sinir ağ modeli seçilmiştir. Bu model, oluşturduğumuz veri setinin eğitim kümesi ile 250 tekrar (epochs) yapılarak eğitilmiştir. Eğitim bittiğinde yapay sinir ağı ağırlıklarını içeren bir JSON dosyası ve doğruluk (accuracy) ve kayıp (loss) grafikleri kayıt edilmiştir. Elde edilen modelin, sonuç grafiklerine bakıldığında tutarlılık değeri %60 lara kadar ulaşmıştır.



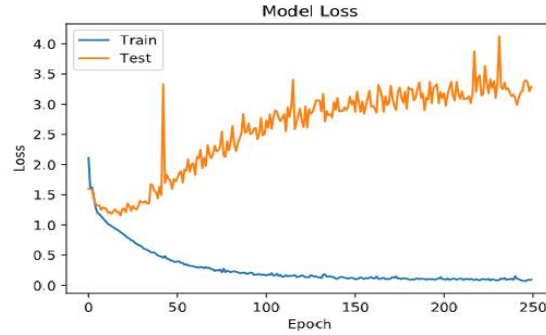
Şekil 2. Cnn Modeli

Projede kullanılan CNN modeli mimarisi Şekil 2’de görülmektedir. İlk olarak 128x128px çözünürlüğündeki yüz görsellerine, 32 filtrede 2x2 matrislik konvolüsyon katmanı uygulanmıştır. Bu aşamada bir eşik değeri belirlenebilmesi için en sık kullanılan aktivasyon fonksiyonlarından ‘Relu’ kullanılmıştır. Bu sayede bu ‘Relu’ ile hesaplanan değerlere göre özellikler çıkarılmıştır. Bu özelliklerin azaltılıp en etkin olanları alınabilmesi için maximum havuzlama işlemi uygulanmıştır. Farklı özellik çıkarımları elde edebilmek için bu işlem 64x2x2, 32x3x3 ve 16x2x2 şeklinde tekrarlanmıştır. Bu sayede bir özellik haritası oluşmuştur. Çıkarılan bu özelliklerin, yapay sinir ağına gönderilebilmesi için tek boyutlu bir diziyeye çeviren vektörel dönüşüm uygulanmıştır. Yapay Sinir Ağı, 3 katmandan oluşmaktadır. Her katmanda sırası ile 16, 8, 5 nöron bulunmaktadır. Daha fazla katman ve nöron kullanıldığında aşırı öğrenmeye sebep olduğu gözlemlenmiştir. 16 ve 8 nöronun bulunduğu gizli katman için ‘relu’ aktivasyon fonksiyonu en tutarlı öğrenim oranlarını verdiği için tercih edilmiştir. Çıkış katmanı olan 5 nöronlu katmanda ise kesin bir olasılık verdiğinden ‘softmax’

fonksiyonu kullanılmıştır. Model eğitimi tamamlandığında 5 duygu için elde edilen en yüksek doğruluk ve kayıp grafikleri surasıyla Şekil 3 ve Şekil 4'te görülmektedir.



Şekil 3. Model Doğruluk Grafiği



Şekil 4. Model Kayıp Grafiği

3.3. Video Duygu Analizi



Şekil 5. Duygu Yüklü Yüzler

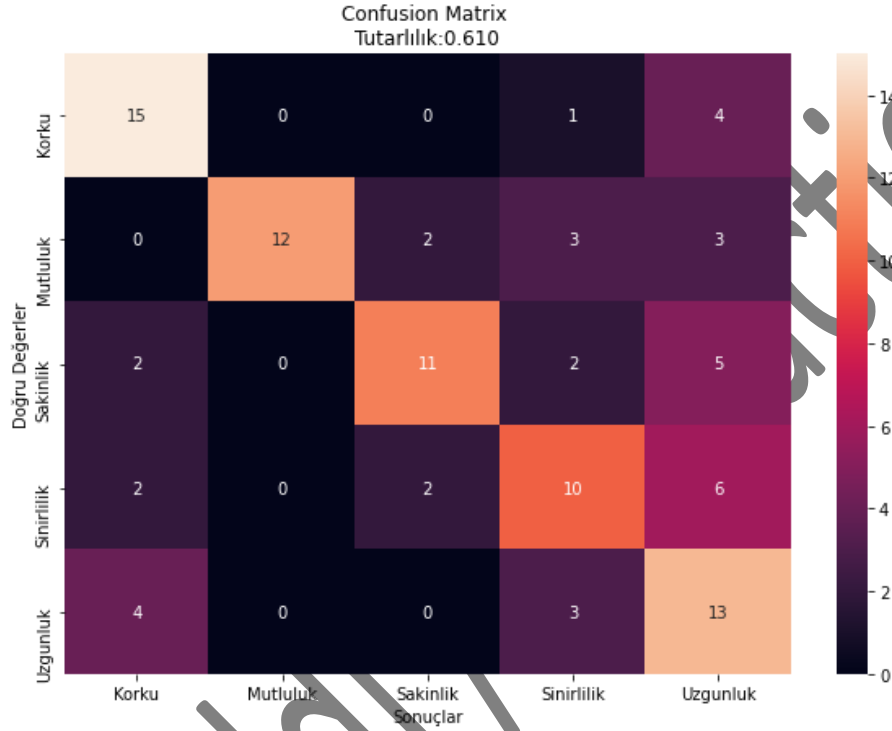
Eğitim sonucu oluşan yapay sinir ağı modeli başarılı bir biçimde sınıflandırmayı öğrenebildiyse, kendine verilen yüz fotoğraflarının hangi duyguya ait olduklarını belirleyecektir. Eğitimi yapılmış modelin ağırlık değerleri, JSON formatında bir dosyada sınama esnasında kullanılmak üzere saklanmıştır.

Örnek videonun işlenmesinden önce özellikleri belirlenir. Kaç frame olduğu veya toplam kaç saniye olduğu gibi değerler kullanılarak algoritma işletilir. Projede, her saniyede bir kare (frame) alınarak üzerinde yüz bulma işlemi gerçekleştirilir. Yüz bulma işleniminde ilk denemelerde Haarcascade methodu denenmiş, ancak daha yüksek başarı elde edildiği için bir caffe modeli kullanılmıştır (DNN) [23]. Alınan karelerdeki yüzler tespit edilmekte, resimden kırılmakta ve klasöre kaydedilmektedir. Ayrıca, kayıt edilen yüz fotoğrafının adı, kaçınıcı saniyede olduğu ve toplamda kaçınıcı fotoğrafa denk geldiği gibi bilgilerde bir başka dosyada saklanır. Bu dosyada veriler JSON nesnesi olarak tutulmaktadır. Yüzün bulunduğu kısmın duygu analizine gönderilmesi ile yapay sinir ağının duygu analiz süresi kısalmakta ve ayrıca daha doğru kararlar vermesi sağlanmaktadır.

Oluşan yüz kesitlerinin analizi, yüz bulmaktan çok daha kısa sürmektedir. İlgili yüz kesitleri dizinden okunarak, duygu analizi yapılmak üzere eğitilen modele girdi olarak verilir. Bu görseller, sinir ağı eğitilirken olduğu gibi tekrardan 128x128px olarak boyutlandırılır. Her ne kadar zaman kaybına sebep olsa da, standart bir boyut kullanıldığından doğruluk oranı artmaktadır. Modelin çıktısı, 5 duygu için bir değer döndürür. Bu değerlerden hangisi yüz görüntüsü en yüksek yüzdelikte tanımlıyor ise, o duygu ile

etiketlenir. Bu sınaama işleminin sonuçları önce bir diziye, daha sonra ise bir JSON nesnesi olarak dosyaya kaydedilir. Bu JSON nesnesi içerisinde yine görselin adı, saniyesi ve sırası gibi bilgilerin yanı sıra, ayrıca yapay sinir ağı modelinden dönen en büyük duygu değerinin yüzdesi de saklanır. Bu dosya ile model değerlendirme verisi oluşturulmaktadır. Bu sayede, bir kez işlenen video tekrar işlenmeden dosyadaki veriler okunarak grafik tekrar elde edilebilmektedir.

Analiz işleminin ardından video içerisindeki yüzlerin saniye saniye duygu analizleri elde edilmiş bulunmaktadır. Dizide tutulan bu bilgiler kullanılarak, bir sonuç grafiği haline getirilir. Bu grafik yorumlanarak, her duygunun diğerlerinden ne kadar ayırt edilebildiği görsel olarak izlenebilir. Ayrıca sınıf bazında karmaşıklık matrisi ile sınaama sonucunda her duygu sınıfının ne kadar başarı ile ayrıştırılabildiği Şekil 6'da görülmektedir. Bu değerler bazı duyguların başarımı açısından çok iyi sonuçlar olarak gözükmese bile, ümit vericidir. Bu tablo ve grafiğin yorumlanması sonuçlar başlığı altında ele alınmıştır

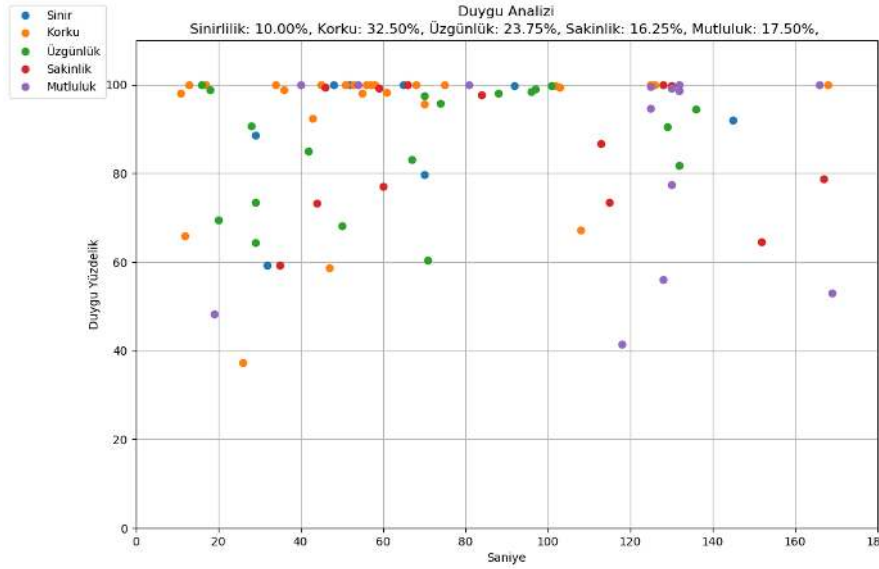


Şekil 6. CNN Modelinin Karmaşıklık Matrisi

Yapılan duygu analizi sonucunda bir grafik oluşmaktadır. Bu grafikte 5 duygunun hangi saniyede yüzde kaç duygu yoğunluğuna sahip olduğu değeri gösterilmektedir. Örneğin, Şekil 7'de Jumanji: The Next Level filminin tanıtım çekiminden yapılan bir analizin sonucu görülmektedir. Bunun yanı sıra, tüm duyguların genel bir yüzde hesapı da yapılmaktadır. Bazı filmlere ait elde edilen duygu yoğunlukları Tablo 2'de verilmiştir. Tek bir yüz karesinden, duygu analizinin yapılması donanıma bağlı olarak 1 ila 3 ms aralığında gerçekleşmektedir.

Tablo 2. Sınanan Bazı Filmlerin Genel Duygu Yüzelikleri

Film Adı	Korku (%)	Mutluluk (%)	Sakinlik (%)	Sinirlilik (%)	Üzgünlük (%)
3 Idiots	17.00	21.51	24.34	18.80	18.34
Glass	28.12	18.23	16.39	10.86	26.39
Joker	17.62	22.88	16.82	12.07	30.60
Shutter Island	22.45	13.05	21.84	15.22	27.44



Şekil 7. Jumanji: The Next Level Trailer Duygu Analizi Grafiği

4. Sonuç ve Öneriler

Bu çalışmanın sonucunda filmler hakkında ortalama bir duygu analizi yapılabilir. Filmlerin içeriklerinde geçen duygu yoğunlukları oransal olarak hesaplanmıştır. Ayrıca, belirli zaman aralıklarında filmde hangi duygu veya duyguların baskın geldiği tespit edilebilmektedir. Toplam film süresi göz önüne alınarak, genel baskın duygunun ne olduğu tespit edilebilmekte, filmin duygu analizi yapılabilmektedir. Bu tür bir bilgi, filmin türünün (örneğin, western, macera, v.b.) yanısıra, duygu ağırlığı hakkında da yıldızsal bir ön değerlendirme yapabilmenin önünü açmaktadır.

Tablo 3. Duygu Analizi Tahminleri

Test Case	Validation Accuracy
VGG 16	39.65%
VGG 16_ovs	40.04%
VGG 16_ru	41.20%
VGG 16_ru_ovs	41.45%
SVM	37.55%
LSTM RNN	38.00%
Hopfield RNN	46.65%
Feature-level Fusion network (MLP)	72.00%
Overall System	54.00%

Yüzden duygu analizi ile ilgili kullanılan yöntemlerin başarı oranları, 2017 de yayınlanan makalede [24] Tablo 3'deki gibi verilmiştir. Duygu analizi ile uygulanan yöntemlerden en iyi sonucun MLP modeli ile alındığı görülmektedir ve kullanılan yöntemlerin ortalama doğruluğu %45 olarak tespit edilmiştir. Yapılan çalışmanın başarı oranı ise %60 olarak elde edilmiştir. Buna rağmen, bu düşük değerler esasen problemin doğası gereği, bazı duyguların ayrıştırılmasında karşılaşılan güçlükten ortaya çıkmaktadır. Duygu analiz yöntemleri, yüz ifadelerinin birkaç duyguyu barındırma veya ifade etmeye sahip karakteristikte olmasından dolayı kesin bir yargıya varmakta zorlanmaktadırlar.

Duygu analizinin oldukça zor bir alan olmasının nedeni birçok duygunun aynı anda var olabilmesi ve yüz mimikleri açısından birbirlerine çok benzeyen duyguların bulunmasıdır. Mevcut sınırlı veri kümesi ile elde edilen başarı oranı, literatürdeki çalışmaların ortalamasından çok daha iyi ve iyiler arasında da ön sıralarda yer aldığı görülmektedir. Bu ayrımların daha doğru yapılabilmesi için, her makine öğrenmesi probleminde olduğu üzere daha büyük veri kümelerine ihtiyaç vardır. Bununla birlikte bahsedilen duygu karmaşıklığından dolayı, bazı duyguların tamamen ayrıştırılmasının mümkün olmayacağı düşünülmektedir.

Bu çalışma kullanılarak, sanatsal ürünlerin duygu analizi veya istatistik bilgileri elde edilerek, tiyatrolar, sinemalar, gösteri salonları gibi yerlerde kullanımı sağlanabilir. Ayrıca bu kullanım ve analizler gerçek zamanlı olarak yapılabilir. Film salonu çıkışında

izleyicilerin kamera görüntüleri işlenerek (örneğin gala sunumlarında), film ile ilgili otomatik duygu yorumu oluşturulabilir. Daha da önemlisi insan etkinliklerinin, kamera görüntüsü alınmış tüm faaliyetlerinin duygu analizinde de kullanılabilir. Örneğin, devlet büyüklerinin yapmış olduğu bir açıklama videosu analiz edilerek duygu analizi yapılabilir. Bunun gibi birçok alanda uygulanması mümkündür. Bunun için gerekli olan, anlık gerçek zamanlı görüntü analizini hızlı yapabilecek güçlü bir donanım desteğidir.

Gelecekte, videolar üzerinde doğal dil işleme teknikleri kullanılarak duygular hakkında bir açıklama metni de hazırlanabilir. Daha da önemlisi, her saniyede alınan yüz kesitleri hakkında bilgiler JSON nesneleri olarak saklandığından, aynı karaktere ait yüz kesitleri ile filmde karakterlerin (artistler) duygusal durumları hakkında yorumlar oluşturulabilir (karakterlere ait önceden oluşturulmuş örneklem yüz veritabanı ile sınıflandırma yapılarak, her yüz kesidinin hangi karaktere ait olduğu ayrıştırılabilir).

Kaynaklar

1. OpenCV. 2017. <http://www.opencv.org> (Erişim Tarihi: 8.7.2017).
2. D. C. Cirean, U. Meier, J. Masci, and L. M. Gambardella, Flexible, High Performance Convolutional Neural Networks for Image Classification, in Proceedings of the Twenty-Second international joint conference on Artificial Intelligence, pp. 1237–1242, 2012.
3. P. Viola and M. Jones, “Rapid object detection using a boosted cascade of simple features,” Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on, vol: 1, pp.:I-511I518, 2001
4. R. Hussin, M.R. Juhari, N.W. Kang, R.C. Ismail and A. Kamarudin, “Digital image processing techniques for object detection from complex background image”. Procedia Engineering, vol:41, pp:340-344, 2012
5. Bradski G., Kaehler A. 2008. Learning OpenCV, O’Reilly Media Inc., USA .
6. Tuncer Ergin, Convolutional Neural Network (ConvNet yada CNN) nedir, nasıl çalışır?, <https://medium.com/@tuncerergin/convolutional-neural-network-convnet-yada-cnn-nedir-nasil-calisir-97a0f5d34cad>
7. Felsen, Agrawal ve Malik, 2017
8. Arı, Uyar ve Akarun, 2008
9. Zhang, Cheng ve Lin, 2012
10. Moore ve Bowden, 2011, Tripathi, Jalal ve Agrawal, 2017
11. Görüntü İşleme Teknikleri ile 3B Yüz Tanıma , Maltepe Üniversitesi Fen Bilimleri Enstitüsü, 2019
12. Bledsoe, 1966a; Bledsoe, 1966b; Bledsoe ve Chan, 1965
13. Romero M., Pears N., “Landmark localisation in 3D face data”, 6th IEEE International Conference on Advanced Video and Signal Based Surveillance, AVSS 2009, s.73–78, 2009.
14. Garcia-Garcia, A., Orts, S., Oprea, S., Villena Martinez, V., Martinez-Gonzalez, P., & Rodriguez, J., A Survey on Deep Learning Techniques for Image and Video Semantic Segmentation, Applied Soft Computing, 70, 41-65, 2018.
15. Tumor detection in MR images of regional convolutional neural networks, Journal of the Faculty of Engineering and Architecture of Gazi University 34:3 (2019) 1395-1408
16. M. F. Valstar and M. Pantic, “Fully automatic recognition of the temporal phases of facial actions,” IEEE Trans. Syst. Man Cybern. Part B Cybern., vol. 42, no. 1, pp. 28–43, 2012.
17. A. Gudi, H. E. Tasli, T. M. Den Uyl, and A. Maroulis, “Deep Learning based FACS Action Unit Occurrence and Intensity Estimation,” vol. 2013, 2015.
18. P. Khorrami, T. L. Paine, K. Brady, C. Dagli, and T. S. Huang, “How Deep Neural Networks Can Improve Emotion Recognition on Video Data,” pp. 1–5, 2016.
19. H.-D. Nguyen, S. Yeom, G.-S. Lee, H.-J. Yang, I. Na, and S. H. Kim, "Facial Emotion Recognition Using an Ensemble of MultiLevel Convolutional Neural Networks," International Journal of Pattern Recognition and Artificial Intelligence, 2018
20. T. Cao and M. Li, "Facial Expression Recognition Algorithm Based on the Combination of CNN and K-Means," presented at the Proceedings of the 2019 11th International Conference on Machine Learning and Computing, Zhuhai, China, 2019.
21. Tuncer Ergin, Convolutional Neural Network (ConvNet yada CNN) nedir, nasıl çalışır?, <https://medium.com/@tuncerergin/convolutional-neural-network-convnet-yada-cnn-nedir-nasil-calisir-97a0f5d34cad>
22. M. Sonka, V. Hlavac and R. Boyle, “Image processing, analysis, and machine vision. Cengage Learning”, 2014
23. Mesut Pişkin, https://github.com/mesutpiskin/computer-vision-guide/tree/master/code/yuz-tanima/python/dnn_yuz_tespiti
24. Emotion recognition and reaction prediction in videos, (Konferans Tarihi:3-5 Kasım.2017), (DOI:10.1109/ICRCICN.2017.8234476), INSPEC Erişim No: 17467617