



Evrişimli Sinir Ağları Kullanarak Yüz Belirleme ve Tanıma Uygulaması

Özge Gürbüz¹, Tolga Yılmaz

(Alınış / Received: 07.09.2023, Kabul / Accepted: 21.12.2023, Online Yayınlanma / Published Online: 24.01.2024)

Anahtar Kelimeler
Derin Öğrenme,
Evrişimli Sinir Ağı,
Yüz Tanıma

Öz: Son yıllarda gelişen teknolojilerin başarı oranının artması ve günlük hayatı kolaylaştırması nedeniyle yüz tanıma tabanlı uygulamaların kullanımı giderek artmaktadır. Yüz tanıma uygulamaları telefonların kilit ekranlarında, öğrenci ve personel devam takibinde, güvenlik kontrolü vb. alanlarda kullanılmakta ve gün geçtikçe kullanım alanı genişlemektedir. Yüz tanıma işlemi için farklı yöntemler bulunmaktadır. Bu yöntemlerden bir tanesi derin sinir ağı çeşidi olan Evrişimli Sinir Ağı kullanmaktır. Bir görüntü içerisindeki nesnelere belirleme ve tanıma alanında yüksek başarı gösterdiği için Evrişimli Sinir Ağları yüz tanıma işleminde ön plana çıkmaktadır. Bu sebeple çalışma kapsamında yüz belirleme ve tanıma için Evrişimli Sinir Ağları kullanılmıştır. Bu çalışmada, en üstün performansı sergileyen ESA modelinin belirlenmesi, ESA modellerinin eğitiminde kullanılan verilerin çeşitli yöntemlerle artırılması ve yüz tanıma sistemine dönüştürülmesi gerçekleştirilmiştir.

Face Detection and Recognition Application Using Convolutional Neural Networks

Keywords
Deep Learning,
Convolutional Neural
Networks,
Face Recognition

Abstract: In recent years, the use of face recognition-based applications has been increasing due to the increasing success rate of developing technologies and facilitating daily life. Face recognition applications can be used on the lock screens of phones for student and staff attendance tracking, security control, etc. It is used in certain areas, and its usage area is expanding day by day. There are different methods for face recognition. One of these methods is to use a convolutional neural network, which is a type of deep neural network. Convolutional neural networks come to the fore in face recognition because they show high success in identifying and recognizing objects in an image. Therefore Convolutional neural networks (CNNs) were used for face detection and recognition within the scope of this study. In this study, the determination of the ESA model with the highest performance, the data used in the training of the ESA models were increased by various methods and converted into a face recognition system.

1. Giriş

Son yıllarda kişisel bilgilerin işlenmesi ile ortaya çıkan güvenlik açıkları artan yüz tanıma tabanlı uygulamaların yaygın olarak kullanılmaya başlanmasına vesile olmuştur. Bu uygulamalar birçok alanda uygulanmaya başlanmış ve başarılı sonuçlar ortaya çıkmıştır. Bu alanların başında sağlık, endüstriyel uygulamalar, eğitim ve askeri alanlar yer almaktadır. Yüz tanıma (FR), kimlik tespiti için yüzün biyolojik karakteristiğini işleyerek bunu kullanan teknolojik bir süreçtir. Literatürde yaygın olarak kullanılan bu yöntemin diğer biyometrik yöntemlerle karşılaştırıldığında; parmak izleri, iris ve DNA gibi özelliklerin yerine yüz, üstün uygulama potansiyeli sunan güvenilirlik, güvenlik ve toplumsal kabul açısından ön plana çıkmaktadır. [1]. Yüz tanıma uygulamalarının kullanım durumu farklı alanlarda da kabul görmektedir. Özellikle Covid-19 salgını sebebiyle temas gerektirmeyen kontroller giderek önem kazanmıştır. Bunlardan bazıları, suçluların tespiti, çalışanların şirket içi uygulamalardaki fotoğraf gönderilerinden duygu tespiti, öğrencilerin online yoklama ile varlık yokluk kontrolü ve insan duygu analizleri gibi çalışmalar yaygınlaşmaktadır [1].

Bu çalışmada yaygınlaşan uygulamalardan biri olan okullarda yoklama sistemi için veri çoğaltma ve derin sinir ağı modelinin seçimi yapılmıştır. Bir görüntü içerisindeki nesnelere belirleme ve tanıma alanında yüksek başarı gösterdiği için derin sinir ağı çeşidi olan Evrişimli Sinir Ağları yüz tanıma işleminde ön plana çıkmaktadır [2]. Evrişimli Sinir Ağları (CNN'ler), bilgisayarlı görme görevlerinde son derece başarılı bir performans sergileyen ve geniş çapta kullanılan bir sinir ağı tipidir [3]. Yüksek başarısından dolayı bu çalışmada Evrişimli Sinir Ağı modelleri kullanılmıştır.

Şu anki yüz tanıma sistemleri, kontrol edilen ortamlarda yüksek bir başarı düzeyine ulaşmış olsa da çeşitli zorluklarla karşılaşmaya devam etmektedir. Bu zorluklar, poz değişiklikleri, aydınlatma koşulları, ifade farklılıkları ve kısmi yüz kapanmaları gibi durumları içerir. Özellikle kısmi yüz kapanmaları, gerçek dünya uygulamalarında en zorlu sorunlardan birini oluşturur. Bu tür kapanmalar, gözlük, güneş gözlüğü, şapka, eşarp gibi yaygın nesnelere veya yüzün bir kısmının örtülmesiyle ortaya çıkabilir. Bu nedenle, gerçek dünya koşullarında güvenilir bir şekilde çalışabilen bir yüz tanıma sisteminin, bu tür kapanmalara karşı dirençli olması gerekmektedir. Geleneksel olarak, bu tür kapanmalarla başa çıkmak için kullanılan temel yaklaşımlardan biri, kapanmış yüz görüntülerini yerel olarak analiz ederek veya test görüntülerini eğitim verilerinde bulunan kapanmamış yüzlerle karşılaştırarak çalışır [4].

Derin sinir ağı eğitimlerinde çok sayıda veri gerekmektedir. Evrişimli Sinir Ağı eğitimlerinde görsel veri kullanıldığı için bu durum veri bulma problemi doğurmaktadır. Yoklama sisteminde küçük okullar için öğrencilerin az sayıda fotoğrafı yeterli olsa da büyük okullar, fakülteler vb. yerlerde fotoğraf sayısı kişiler arasındaki farkı ağı daha iyi öğrenebilmesi için daha fazla olmalıdır. Ayrıca öğrenciler yıl boyunca hep aynı görünmedikleri durumlar oluşabilmektedir. Öğrencinin sakal veya bıyık bırakması, saçını uzatması ya da kesmesi, yüzüne dövme yaptırması, kamera önüdeyken şapka-gözlük takması vb. birçok olay bu durumlara örnektir. Bu durumlardan biri olan makyaj, genellikle yüzün çekiciliğini artırmak amacıyla yaygın bir şekilde kullanılan ve genel olarak toplum tarafından kabul gören bir uygulamadır. Ancak, farklı makyaj stilleri, yüzün görünümünde belirgin değişikliklere yol açabilir. Makyajlı ve makyajsız yüz görüntülerini başarılı bir şekilde eşleştirmek hala büyük bir zorluk oluşturmaktadır [5].

Bu çalışmada, yüz tanıma ve belirleme sistemi için öğrenci fotoğraflarına şapka, gözlük, farklı saç modelleri vb. durumlar eklenerek veriler çoğaltılmış ve elde edilen sonuçlar ile derin sinir ağı eğitilmiştir. İşlenecek verilerin ortaya çıkması ile analiz çalışmaları yapılarak çalışmanın önemi vurgulanmıştır. Makale içeriği olarak kullanılmış araçlar, yöntemler ve veri setleri detaylandırılarak aktarılmıştır.

2. Materyal ve Metot

Yöntem bölümünde çalışmanın geliştirilme aşamasında takip edilmiş adımlar verilmiştir. Bu adımlarda sırasıyla veri kümesi oluşturma, Evrişimli Sinir Ağları, veri çoğaltma ve Evrişimli Sinir Ağları Modellerinin Eğitilmesi anlatılmıştır.

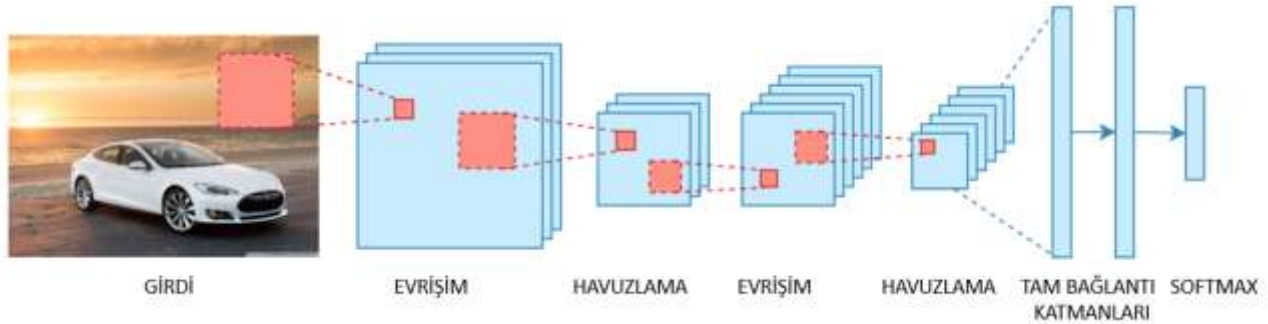
Bu çalışmada iki bilgisayar kullanılmıştır. Bilgisayarlardan birisi i7-13700H 2.40 GHz işlemci, 16GB DDR4 bellek, Windows 11 Home Single Language işletim sistemi ve 6GB Nvidia RTX4050 ekran kartına sahiptir. Diğeri ise Ryzen 7 5700X 4.60 Ghz işlemci, 32GB DDR4 bellek, Windows 10 Pro 64-bit işletim sistemi ve 12GB Nvidia RTX3060 ekran kartına sahiptir.

2.1. Veri kümesi

Çalışmanın veri çoğaltma ve Evrişimli Sinir Ağı eğitimlerinde “FEI Face Database” [6] ve “Labeled Faces in the Wild” [7] yüz veri setleri kullanılmıştır. FEI Face Database, 2005 ve 2006 yılları arasında Brezilya’daki FEI Yapay Zekâ Laboratuvarında çekilmiş yüzlerin veri tabanıdır. 200 farklı kişinin 14 farklı pozisyonda (yüzün sağından, solundan, 2 önünden, gülümserken ve gülümsemezken, düşük ışık altında) fotoğraflarını içermektedir. Yüzler 19 ila 40 yaşları arasındaki FEI’deki öğrenci ve personellerden oluşmaktadır [6]. LFW, web’den toplanan 13.000’den fazla yüz görüntüsü içermektedir. Bu görüntülerdeki yüzler Viola Jones algoritması tarafından tespit edilebilmiştir [7]. Bu veri setinde birçok kişinin az miktarda fotoğrafı bulunmaktadır. Bu sebeple veri setinde fotoğraf sayısı sekiz ve üstü olan kişilerin yüz verileri çalışmada kullanılmıştır. FEI yüz veri setindeki 14 farklı yüz pozisyonundaki fotoğrafların 8 tanesi bu çalışmada kullanılmıştır. Eğitim aşamasında kişilerin veri sayısı eksik kalmaması için LFW veri setinde fotoğraf sayısı minimum 8 olan kişiler seçilmiş ve bunların 8 adet fotoğrafı kullanılmıştır.

2.2. Evrişimli sinir ağları

Evrişimli Sinir Ağları (Convolutional Neural Networks – CNN), bir Derin Sinir Ağlar (DSA) türüdür. Ve çok katmanlı algılayıcıların mimarisini kullanır. Evrişimli Sinir Ağları çoğunlukla görüntü analizi, görüntülerdeki nesnelere tanımak ve Bilgisayar görüşündeki (CV) sorunları çözmek için kullanılır. Evrişimli Sinir Ağları, insanların ve hayvanların görsel korteksinin işleyişini taklit eder. Son zamanlarda görüntü analizi, görüntülerdeki nesnelere tanımak ve Bilgisayar görüşündeki (CV) sorunları çözmek konusunda yüksek başarı elde etmişlerdir [8]. Bundan dolayı bu çalışmada Evrişimli Sinir Ağı kullanılmıştır. Bu ağlar, bir görüntüdeki nesneyi tanımak için birlikte çalışan birçok farklı katmana sahiptirler, Şekil 1’te ESA’nın yapısı gösterilmektedir.



Şekil 1. ESA'nın yapısı [9]

Konvolüsyon (convolutional) ve havuzlama (pooling), Evrişimli Sinir Ağları mimarisinin ilk aşamalarındandır. Tam bağlı katman ve hemen ardındaki sınıflandırma katmanı ise Evrişimli Sinir Ağları mimarisinin son aşamalarındandır. Evrişimli sinir ağları birden fazla eğitilebilir bölümlerden oluşur ve devamında eğitici bir sınıflandırıcı ile devam eder. Giriş verileri alınır, katman katman işlemler yapılarak eğitim süreci gerçekleştirilir. Son olarak da doğru sonuç ile karşılaştırma yapabilmek için bir final çıktısı verir. Üretilen sonuç ile istenen sonuç arasındaki fark kadar hata oluşur. Geriye yayılım algoritması sayesinde bu hata bütün ağırlıklara dağıtılır. Her bir iterasyonla birlikte de ağırlıklar güncellenerek hatanın azaltılması sağlanır [10]. Evrişim katmanı ve havuz katmanı, sonraki bölümlerde açıklanan hiper parametrelere göre ince (fine-tuned) ayarlanabilir [10]. ESA girdi görüntüsünü alarak çalışmaya başlar, ilk gizli katmanda nöronlar kenarlar, kulakların şekli veya burun gibi basit kalıpları öğrenirler. Sonra, görüntü bir sonraki katmana iletilir ki bu katman bir önceki katman ile kıyaslandığında daha belirgin özellikleri öğrenebilmektedir. Katmanlar arttıkça nöronlar, yüz veya araba gibi üst düzey özellikleri öğrenmeye başlar [8]. Görüntülerin boyutu, tam bağlı katmana varıncaya kadar düşmektedir. Çünkü girdi verisi tam bağlı katmana kadar konvolüsyon ve havuzlama katmanlarından geçer. Konvolüsyon (evrişim) katmanında görüntü üzerinde filtreler benzer operasyonlar uygulanır ve farklı özellikleri vurgulanır. Bu katmanlar görüntü boyutunu azaltmazlar, ancak çıkış olarak elde edilen özellik haritaları giriş boyutuna kıyasla daha küçük olabilmektedir. Havuzlama katmanları konvolüsyon katmanının çıktısını 3x3, 5x5 gibi filtreler kullanarak özetler ve bu özetleme sırasında girdinin boyutu küçülür. Bu işlem daha yüksek seviye özetleme ve hesaplama maliyetinin düşürülmesi için kullanılır.

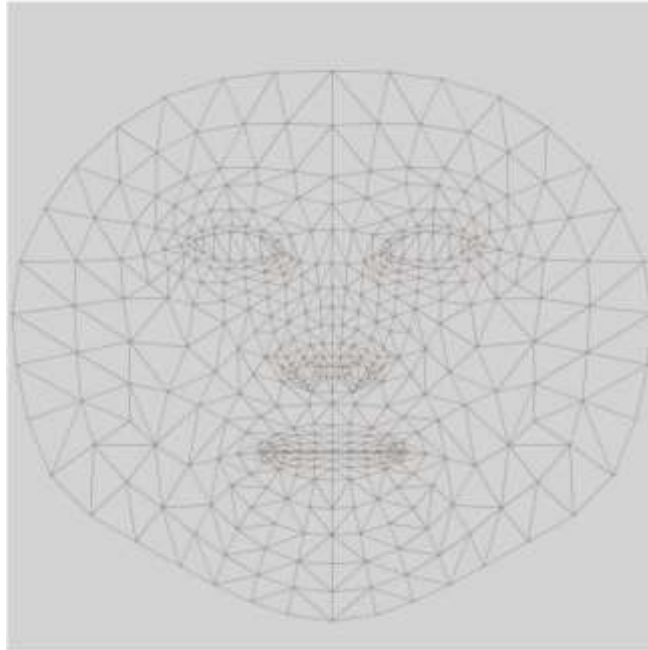
2.3. Veri çoğaltma

ESA modellerinin eğitimlerinde ne kadar çok iyi veri kullanılırsa eğitimin başarısı o kadar artar. Burada çok iyi veriden kastedilen veri setinin homojen olmasıdır. Örneğin veri setindeki bir kişinin kırmızı gözlük ile olan bir fotoğrafı eğitimde kullanılmadığı, veri setindeki diğer kişilerin ise kırmızı gözlük taktığı fotoğraflar eğitimde kullanıldığı varsayalım. Kişinin kırmızı gözlük taktığı bir fotoğraf ESA modeline verildiğinde eğitimde kırmızı gözlük takan kişilere benzerlik oranı daha fazla çıkar ve hatalı sonuç döndürür. Aynı zamanda yoklama sisteminde kullanıcıların zamanla yüz görüntüleri değişebilmektedir. Kişilerin fiziksel olarak değişimlerinin yanında aksesuar kullanımı da veri işleme sürecinde olumsuz durumlar ortaya çıkarabilmektedir. Bunlar, şapka takması, saçın durumu, gözlük kullanma, sakal-bıyık uzatma vb. durumlarda eğitilmiş olan ağ daha önce çoğaltılmış veriler ile eğitildiği için bu değişiklik karşısında da başarılı olması bu tip sistemler için son derece önemlidir. Yapılan çalışmada bu tip durumlar için başarılı sonuçlar ortaya çıktığı analizler sonucunda gösterilmiştir. Bu nedenle verilerin çoğaltılması ve uygun sonuçlar alınması son derece önemli bir konudur.

Veri çoğaltmak için iki yöntem kullanılmıştır. Biri Python ortamında Mediapipe Frameworkü yardımıyla resimler üzerine PNG formatındaki başka bir resmin birleştirilmesi. Diğer yöntem ise yapay zekâ yöntemi olan "Stable Diffusion" yöntemidir.

2.3.1. Mediapipe kullanarak veri çoğaltma

Mediapipe, Google tarafından geliştirilen makine öğrenimi çözümleri oluşturmak için kullanılan bir framework'tür. Mediapipe, video analizi, nesne algılama, yüz izleme, beden izleme, el izleme gibi birçok çoklu ortam işleme görevini gerçekleştirebilen açık kaynak kodlu bir projedir. Video içerisinde insan yüzünün, gözünün, burnunun, alınının nerede olduğunu; elin parmak uçlarından avuç içine kadar nerelerde olduğunu kullanıcıya geri döndürür. Bu çalışmada Mediapipe insan yüzünün fotoğraf üzerinde nerede olduğunu ve koordinatlarının neler olduğunu bulmakta kullanılmıştır. Platformdan bağımsız olarak farklı işletim sistemlerinde kullanılabilir [11]. Bu çalışmada Mediapipe ile veri çoğaltma temelinde iki resmin birleştirilmesi mantığına dayanmaktadır. Yüz fotoğrafı ile eklenecek olan fotoğrafın doğru koordinatlar ile birleştirilmesi sonucu veriler çoğaltılır. Doğru koordinatlar yüz landmark noktaları ile tespit edilmektedir. Yüz landmark noktaları, insan yüzünün uzuvları olan göz, burun, kaş, dudak, çene, yanak vb. bölgelerin noktalar ile işaretlendiği halidir. Farklı yöntemler ile bu noktalar tespit edilebilse de bu çalışmada bu noktaların tespitinde Mediapipe kullanılmıştır. Mediapipe ile bir fotoğraftaki insan yüzünün numaralandırılmış bölgelerinin koordinatlarını elde edilebilmektedir. Şekil 2'de 468 yüz landmark noktasının görseli verilmiştir.



Şekil 2. 468 Yüz landmark noktası [12]

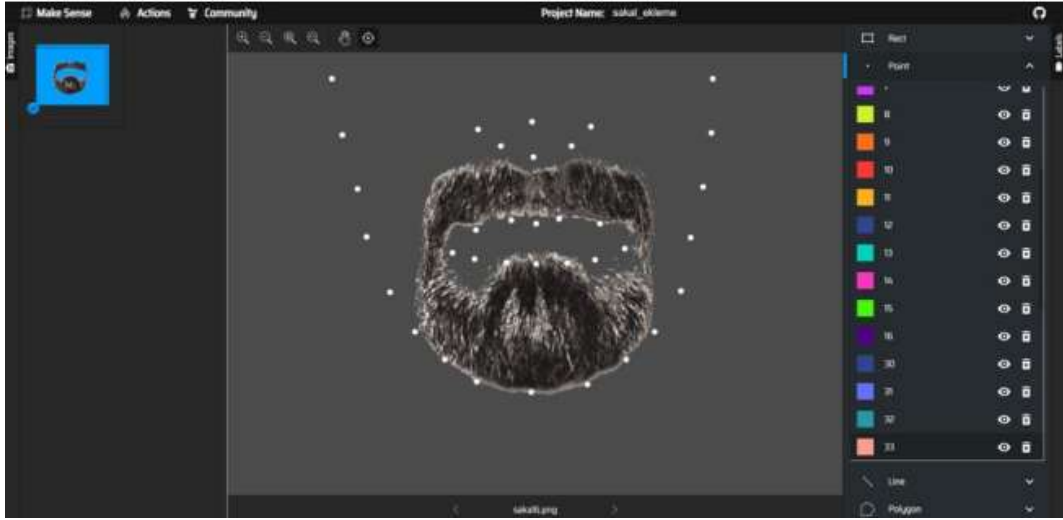
Şekil 2'de görülen her çizgi çakışması insan yüzü üzerindeki 1 landmark noktasını temsil etmektedir. Bu çalışmada sadece 468 yüz landmark noktası değil 74 yüz landmark noktası da kullanılmıştır. Mediapipe yardımıyla 2 farklı birleştirme işlemi yapılabilmektedir. Bunlardan bir tanesi iki resmi doğrudan birleştirmektir. Bu yöntemle 468

yüz landmark noktasından doğru olanları seçilip PNG formatındaki fotoğrafın bu noktalara göre konumlandırılarak resimlerin birleştirilmesi. Şekil 3'te bu konumlandırma işleminin gözlük ile yapıldığında elde edilen sonuçlardan üç adet örnek verilmiştir.



Şekil 3. Gözlük ekleme örnekleri

Bir diğeri ise 74 yüz landmark noktasını kullanarak yüze yapışık şekilde eklemeler yapmak. Mediapipe sayesinde bir PNG resmi insan yüzü üzerinde yapışık şekilde eklenebilmektedir. Bu "PNG" formatındaki resmin etrafında bulunan landmark noktalarının etiketlenmesi ile ".csv" uzantılı dosya elde edilmektedir. Bunun için MakeSense isimli web tabanlı uygulama ücretsiz olarak kullanılabilir [13]. MakeSense bir veri etiketleme uygulamasıdır. MakeSense yardımıyla fotoğraf üzerindeki bir noktayı ya da bir alanı işaretleyerek o koordinatlara sınıf etiketi verebiliriz. Şekil 4'te MakeSense web sitesinin arayüzü ve bir PNG sakal resminin etiketlenmesi gösterilmektedir.



Şekil 4. MakeSense arayüzü [13]

Etiketlenmiş PNG resmi ile Python geliştirme ortamında 74 yüz landmark noktası kullanılarak birleştirme işlemi gerçekleştirilir. Bu birleştirme sonrası PNG resim insan yüzünde belirtilen noktalar arasında yapıştırılmış gibi görünmektedir. Şekil 5'te örnek birleştirilmiş resimler verilmektedir.



Şekil 5. Örnek eklemeli resimler

Python geliştirme ortamında tek bir fotoğraf girdisi ile yukarıda bahsedilen iki yöntem kullanılarak, bir fonksiyonun çıktısı diğer ekleme fonksiyonunun girdisi haline getirilir, sadece cinsiyet bilgisi verilerek verileri çoğaltan kod yazılmıştır. Bu kod yoklama sistemine de entegre edilebilir hale getirilmiştir. Bu aşamada 30 farklı yüz yapısına sahip kişi FEI yüz veri setinden seçilmiştir. Her bir PNG dosyası bu yüzler üzerinde denenerek X eksenindeki kayma, Y eksenindeki kayma ve PNG resminin kaç kat büyütülüp ya da küçültüleceği hesaplanmıştır.

Tek bir fotoğraftan otomatik çoğaltma ile yaklaşık 50.000 ayrı fotoğraf elde edilebilmektedir fakat bu sayı eğitim için hem aşırı fazla hem de eğitimin süresini aşırı büyütüp gereksiz zaman kaybetmeye sebep olabilmektedir. Ayrıca tek bir fotoğraftan çok fazla veri çıkarılmasından dolayı depolama alanı sorunu yaşanabilmektedir. Bu sebeple fotoğraflara resim ekleme kısıtı getirilerek tek bir fotoğraftan yaklaşık 500 ayrı çoğaltılmış veri elde edilen şekle çevrilmiştir. Eklenerek gözlük, şapka, sakal vb. resimlerin tamamı bir fotoğrafa uygulanmayıp, rastgele seçilerek belli bir kısım ekleme yapılarak aşırı fazla veri üretiminin önüne geçilmiştir. Şekil 6'da Mediapipe ile başarılı şekilde çoğaltılmış fotoğraf örnekleri verilmiştir.



Şekil 6. Mediapipe başarılı çoğaltılmış fotoğraf örnekleri

Genel olarak kişinin sağ ve sola dönük fotoğraflarında başarısızlık oranı daha fazla olmaktadır. Fakat kişinin düz baktığı fotoğraflarda da başarısızlık durumuyla karşılaşılabilir. Kişinin sağa ve sola dönük olduğu fotoğrafları çoğaltılmaz, sadece düz şekilde durduğu fotoğrafları çoğaltılırsa hatalı veri oranı büyük oranda düşmektedir. Yoklama sisteminde öğrenciler sınıfa girerken kameraya doğrudan baktığı varsayılırsa bu yöntem başarılı olmaktadır. Fakat öğrencilerin sınıfa kameraya doğrudan bakmadan hafif sağa ya da sola dönüş şeklinde girmeleri durumunda bu açılarından olan fotoğrafları da çoğaltma ihtiyacı doğmaktadır. Ya da veriler çoğaltıldıktan sonra başarısız olan çoğaltma işlemleri veri setinden çıkarılmalıdır.

2.3.2. Stable diffusion ile veri çoğaltma

Stable Diffusion, özgün bir veri setini kullanarak yeni örnekler oluşturmak için kullanılan bir veri çoğaltma yöntemidir. Bu yöntem, veriye gürültü ekleyerek veya varyasyonlar oluşturarak mevcut örnekleri çeşitlendirmek için kullanılır. Stable Diffusion, genellikle derin öğrenme ve yapay zekâ alanlarında kullanılan ve özellikle sınırlı veri durumlarında performansı arttırmaya yardımcı olan bir tekniktir.

Stable Diffusion Inpainting, eksik veya hasarlı bir resmin boş veya zarar görmüş bölgelerini doldurmak için ısı yayılma özelliklerini kullanan bir iç boyama tekniğidir. Bu yöntem, eksik veya hasarlı bölgenin etrafındaki resim piksellerine ısı yayılma işlemi uygulayarak, doğal bir şekilde bütünleşen pürüzsüz bir yama oluşturur [14].

Negative Prompt (Olumsuz Bilgi İstemi), Stable Diffusion'ı kullanıcının herhangi bir ekstra giriş yapmadan görmek istemediğini belirtmesine izin verecek şekilde kullanmanın bir yoludur. Stable Diffusion modeline, oluşturulan görüntüye nelerin dahil edilmeyeceğini söyleyen bir parametredir.

Prompt (Bilgi İstemi), yazılı metni girdi olarak kullanarak görüntü üretici ile iletişim kurma yöntemidir. Birincil işlevi, yapay zekâ (AI) modeline verilen istemi temel alarak görüntüler oluşturma talimatı vermektir [16]. Şekil 7'te Stable Diffusion modelinin Inpaint yöntemi kullanılarak üretilmiş fotoğraflardan örnekler yer verilmiştir.



Şekil 7. Stable Diffusion modeli Inpaint yöntemi kullanılarak üretilmiş örnekler

2.4. ESA modellerinin eğitilmesi

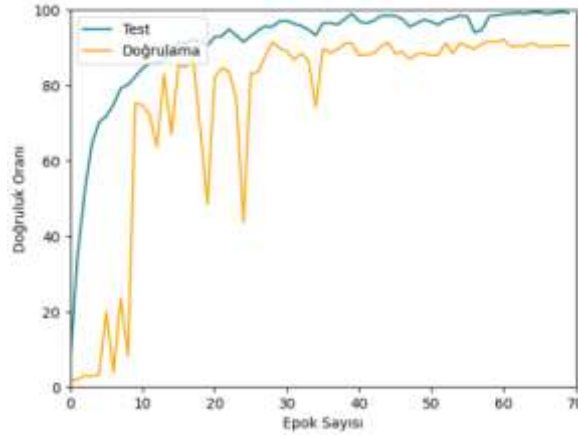
Bu kısımda veri çoğaltma aşamasında üretilen veri setleri ile ESA modelleri eğitilmiştir. Mediapipe ve Stable Diffusion yöntemleri ile çoğaltılan verilerden 50 kişilik ayrı bir veri seti oluşturulmuştur. Bu veri setinde başarısız çoğaltma işlemi yapılan veriler ayıklanmıştır. ESA modellerinin eğitimleri uzun sürdüğü ve Google Colab 12 saat çalışma sınırı verdiği için veri setindeki başarısız veriler çıkarılarak eğitim süresi düşürülmeye çalışılmıştır. Google Colab'ta AlexNet, VGG-16, Darknet ve GoogLeNet modelleri oluşturulan veri seti ile eğitilmiştir. Ek olarak her fotoğraf için bir ekleme yapılmış, daha sade bir veri seti ile de eğitim yapılmıştır.

2.4.1. AlexNet

2012 yılında Alex Krizhevsky, Ilya Sutskever ve Geoffrey E. Hinton tarafından geliştirilen Evrişimli Sinir Ağı mimarisidir. Girdi olarak 224x224x3 (bazı kaynaklarda 227x227x3 olarak geçer fakat 224x224x3 doğru olanıdır) boyutunda görüntü alır. Beş adet evrişim katmanı ve üç adet tam bağlı katmandan oluşur. Havuzlama katmanlarında maxpooling kullanır [17]. ILSRVC (ImageNet Büyük Ölçekli Görsel Tanıma Yarışması) 2012 kazananıdır [18].

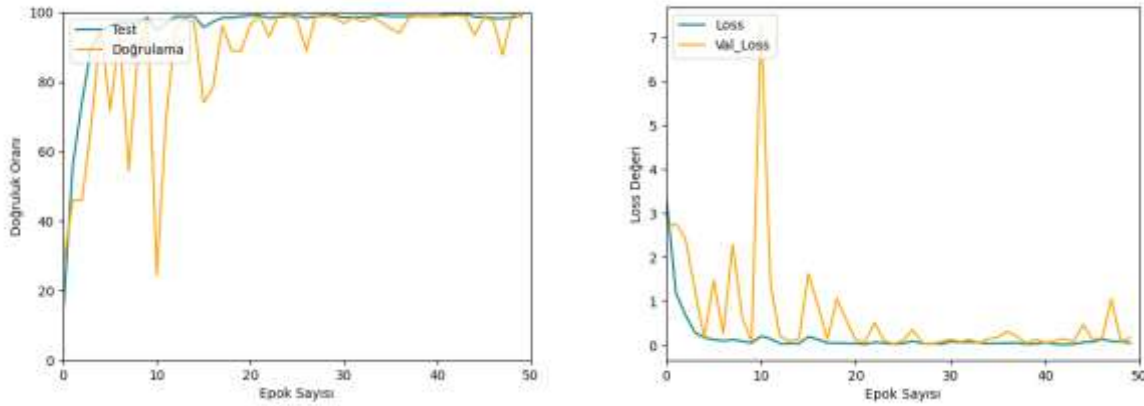
Eğitilen ilk derin sinir ağı modeli AlexNet. Bu modelin eğitimlerinde düşük doğruluk oranı problemi ile karşılaşmıştır. Öğrenme katsayısı, aktivasyon fonksiyonu vb. hiper parametrelerde değişiklikler yapılarak bu

problem çözülmüştür. Bu modelin eğitimi 70 epok, öğrenme katsayısı "0.001" ve batch boyutu "32" olarak ayarlanmıştır. Batch Boyutu, ağa verilen alt örneklerin sayısını belirtmektedir [19].



Şekil 8. AlexNet doğruluk oranı grafiği

Şekil 8'da AlexNet modelinin bir fotoğrafta birden fazla eklemenin yapıldığı veri seti ile eğitiminin doğruluk oranı grafiği verilmiştir. Bu grafikte mavi renk ile belirtilen test, sarı renk ile belirtilen ise doğruluk oranının grafiğidir. Eğitim veri setinin tamamının modele verilip bir kere işlenmesine bir epok denir. Her epok sonunda model test verileri ile test edilir ve doğruluk oranına bağlı olarak eğitim yön kazanır. Her epok sonunda doğrulama veri seti de modele verilir. Eğitimin performansını izlemek ve hiperparametreleri ayarlamak için doğrulama verisi ile doğruluk oranı kullanılır. Aynı zamanda modelin ezber yapıp yapmadığı da kontrol edilir. Modelin eğitim ve test verilerini ezberleyip doğruluk oranını yüksek verir ancak modele verilmeyen verilerde düşük başarı elde edilir. Grafikte 20 ve 24. epoklarda bu durum gözlemlenmiştir. Doğrulama doğruluk oranı yüksek değerlerden birden çok fazla düşmüştür. Grafikte görüldüğü gibi doğrulama verisi doğruluk oranı test verisinin altında kalmıştır. Epok sayısı arttıkça test ve doğrulama doğruluk oranları birbirlerine benzer hareket etmiştir. Bu modelin başarı oranı eğitimlerde %85 seviyelerine çıkmıştır.



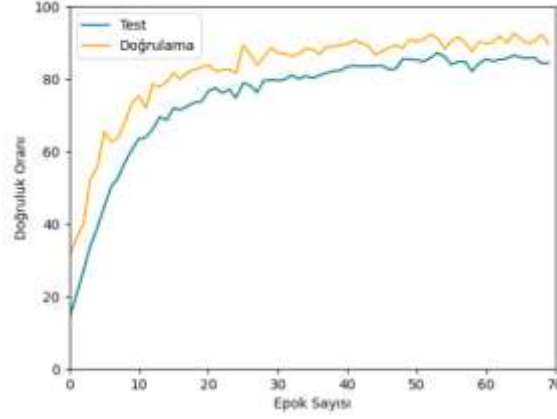
Şekil 9. Sadeleşmiş veri seti ile eğitilmiş AlexNet doğruluk ve loss oranı grafiği

Şekil 9 'da ise daha sade veri seti ile yapılmış eğitimin doğruluk oranları ve loss değeri grafiği verilmiştir. Loss değeri modelin gerçek verilere ne kadar yakın veya ne kadar uzak olduğunu belirtir. Sağdaki grafikte "Loss" olarak belirtilen test verisi loss değeri ve "Val_Loss" olarak belirtilen ise doğrulama verisi loss değeridir. Bu veri seti 50 epok uzunluğunda eğitilmiştir. İki veri setinde de benzer özellikler görülmektedir.

2.4.2. VGG-16

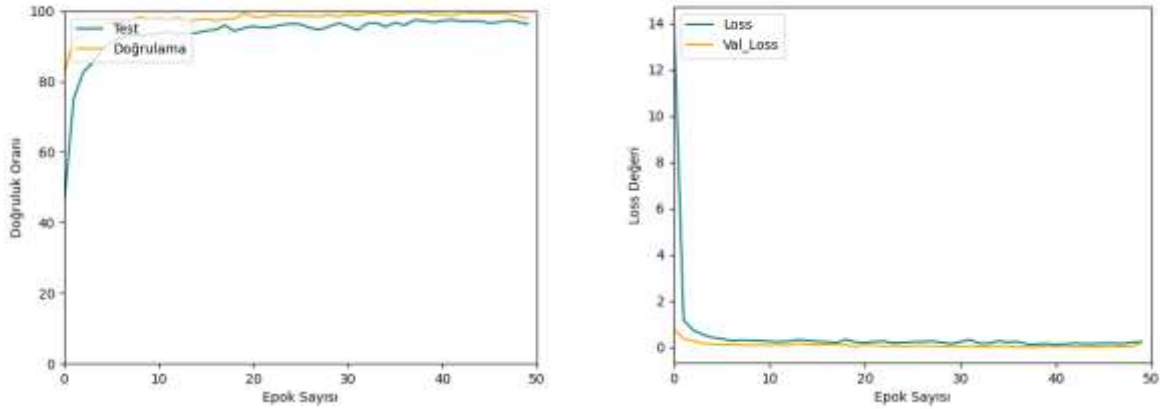
Bir diğer derin sinir ağı modeli VGG-16 popüler bir ESA modelidir. 2014 yılında Karen Simonyan ve Andrew Zisserman tarafından Oxford üniversitesinde tasarlanmıştır. Altı farklı model olarak ortaya çıkmıştır. Bu modeller 11 (Vgg-11), 13 (Vgg-13), 16 (Vgg-16) ve 19 (Vgg-19) katmandan oluşmaktadır. 11 ve 16 katmandan oluşan modellerin iki farklı çeşidi vardır. Her bir modelde evrişim katmanlarından sonra üç adet tam bağlı katman bulunmaktadır. Bu modeller evrişim katmanlarında 3x3 boyutunda filtre uygular. Girdi olarak 224x224x3 boyutunda görüntü alır. Havuzlama katmanlarında 2x2 boyutunda filtreye max-pooling uygulanır. Bu model doğrudan Tensorflow kütüphanesinin içerisinde bulunmaktadır. Modelin her bir katmanını tek tek oluşturmak yerine doğrudan fonksiyon olarak çağrılarak kullanılabilir.

Çalışmada test edilmiş model Vgg-16 modelidir. Bu model 13 adet evrişim katmanı ve üç adet tam bağlı katmandan oluşmaktadır [20]. Modelin eğitimi 70 epok, batch boyutu 32 ve öğrenme katsayısı "0.001" olarak ayarlanmıştır.



Şekil 10. Vgg-16 doğruluk oranı grafiği

Şekil 10'de da görüldüğü gibi test ve doğrulama verisi doğruluk oranları birbirlerine çok benzer ilerlemiştir. Doğrulama verisi doğruluk oranı eğitim boyunca test verisinin üstünde kalmıştır. Farklı veri setleri ile yapılan eğitim denemelerinde de doğrulama verisi doğruluk oranı test verisinden yüksek kalmıştır.



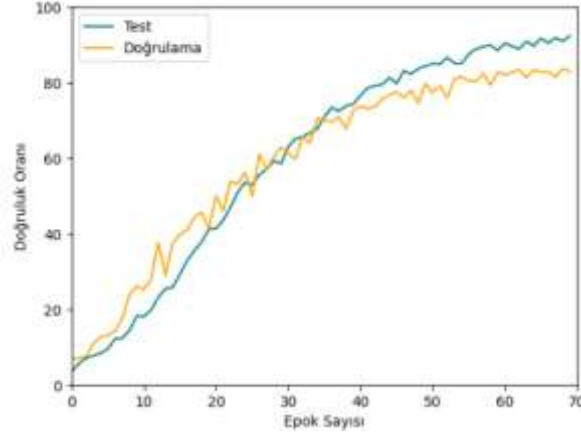
Şekil 11. Sadeleşmiş veri seti ile eğitilmiş Vgg-16 doğruluk ve loss oranı grafiği

Şekil 11'de sade veri seti ile eğitilmiş Vgg-16'nın doğruluk oranları ve loss değeri grafiği verilmiştir. Sade veri seti 50 epok uzunluğunda eğitilmiştir. Bu veri setinin eğitiminin doğruluk değeri daha erken epoklarda yüksek başarı oranlarına çıkmıştır.

2.4.3. GoogLeNet

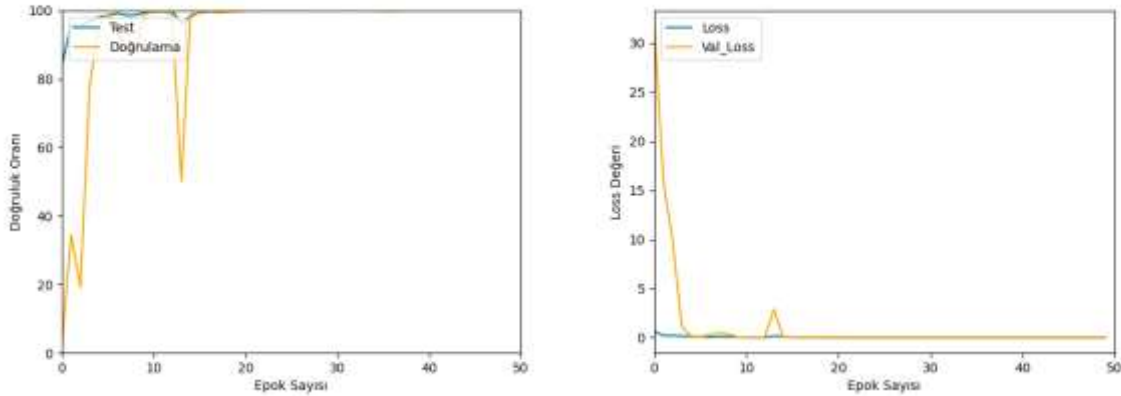
2014 yılında yapılan ILSRVC (ImageNet Büyük Ölçekli Görsel Tanıma Yarışması) kazananıdır. Girdi olarak gelen görüntüyle aynı anda birden fazla evrişim ve havuzlama işlemi gerçekleştirilir. Elde edilen sonuçlar birleştirilir ve sonraki katmana gönderilir [21]. GoogLeNet mimarisi 22 katman derinliğindedir. Giriş olarak 224x224x3 boyutunda görüntü alır. Havuzlama katmanlarında max-pooling kullanır [22].

Katman yapısı diğer modellere göre daha karmaşık ve uzun olan bir model olmasına rağmen eğitim süreleri genel olarak VGG-16'dan daha kısa sürmüştür. Modelin eğitimi 70 epok, batch boyutu 32 ve öğrenme kat sayısı "0.0001" olarak ayarlanmıştır.



Şekil 12. GoogLeNet doğruluk oranı grafiği

GoogLeNet modelinin diğer modeller gibi tek çıkışlı değildir. Bu modelin üç adet çıktısı bulunmaktadır. Şekil 12'de verilen grafik son katmandaki çıkışa aittir. Grafikte görüldüğü gibi test ve doğrulama verisi doğruluk oranı birbirlerine benzer ilerlemiştir. Eğitimin başlangıcında doğrulama verisi doğruluk oranı test verisinin üstünde kalmış fakat eğitim ilerledikçe test verisi doğruluk oranı doğrulama verisinininkini geride bırakmıştır.



Şekil 13. Sadeleşmiş veri seti ile eğitilmiş GoogLeNet doğruluk ve loss oranı grafiği

Sade veri seti ile GoogLeNet modeli 50 epok uzunluğunda eğitilmiştir. Test doğruluk oranı başlangıçta çok yüksek seviyeden başlamıştır. Veri seti daha sade olduğu için dalgalanma az olmuş ve çok hızlı bir şekilde %100 doğruluk oranını yakalamıştır. Doğrulama doğruluk oranı ise bazı epoklarda hızlı düşüşler yapıp tekrar yükselmiştir.

2.4.4. Model-A

Tensorflow ile derin sinir ağı modeli oluşturulabilmektedir. Bu sebeple yapılan çalışma için basit bir Evrişimli Sinir Ağı modeli oluşturulmuştur. İsmine sembolik olarak Model-A verilmiştir. Bu model 5 adet evrişim katmanından oluşmakta ve her evrişim katmanından sonra bir adet maksimum havuzla bulunmaktadır. Giriş verisi olarak

(256,256,3) boyutunda fotoğraflar almakta, aktivasyon fonksiyonu olarak "ReLU", filtre boyutu (3x3) ve çıkış katmanında "softmax" fonksiyonu kullanmaktadır. Tablo 1'de bu modelin katmanlarının özeti verilmiştir.

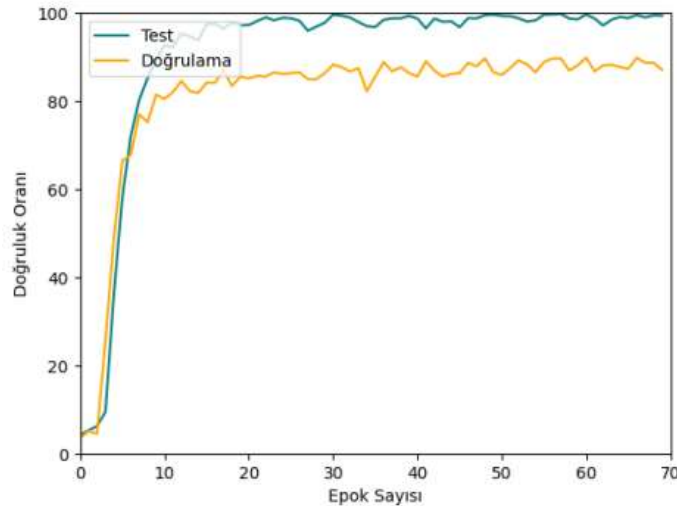
Tablo 1. Model-A modelinin katman bilgileri

Katman	Çıkış Boyutu	Parametre Sayısı
1.Evrişim Katmanı	(None, 254,254,16)	448
Maksimum Havuzlama	(None, 127,127,16)	0
2.Evrişim Katmanı	(None, 125,125,32)	4640
Maksimum Havuzlama	(None, 62,62,32)	0
3.Evrişim Katmanı	(None, 60,60,64)	18496
Maksimum Havuzlama	(None, 30,30,64)	0
4.Evrişim Katmanı	(None,28,28,32)	18464
Maksimum Havuzlama	(None, 14,14,32)	0
5.Evrişim Katmanı	(None,12,12,16)	4624
Maksimum Havuzlama	(None, 6,6,12)	0
Flatten	(None, 576)	0
Tam Bağlantı Katmanı	(None, 256)	147712
Tam Bağlantı Katmanı	(None, 50)	7710

Toplam Parametre = 202,094

Tablo 1’de parametre sayısı olarak belirtilen değerler o katmanda bulunan öğrenilebilir parametreleri ifade eder. Öğrenilebilir parametreler Evrişimli Sinir Ağının nöronları arasında bulunan ağırlıklar ve önyargıdır (bias değerleri). ESA modelinin eğitilmesi sırasında bu parametrelerin değerleri değiştirilir. ESA modelleri eğitilerek bu ağırlıklar ve önyargı değerlerinin en iyi kombinasyonunu bulmaya çalışır. ESA modeline eğitim sonrası verilen girdi verisi bu ağırlık ve önyargı değerleri ile çarpılarak doğru sınıflandırma yapmaya çalışılır.

Bu model de diğer modellerin eğitildiği veri seti ile eğitilmiştir. Eğitim uzunluğu 70 epok, öğrenme katsayısı "0.0001" ve batch boyutu 32 olarak ayarlanmış ve eğitime sokulmuştur. Diğer modellere oranla çok daha az katman ve parametre sayısına sahip olduğu için eğitimi yaklaşık 7 ila 8 kat hızlı gerçekleşmiştir. Bu modelin doğruluk oranı diğer modeller ile benzer çıkmıştır. Şekil 14’te Model-A doğruluk oranı grafiği verilmiştir.

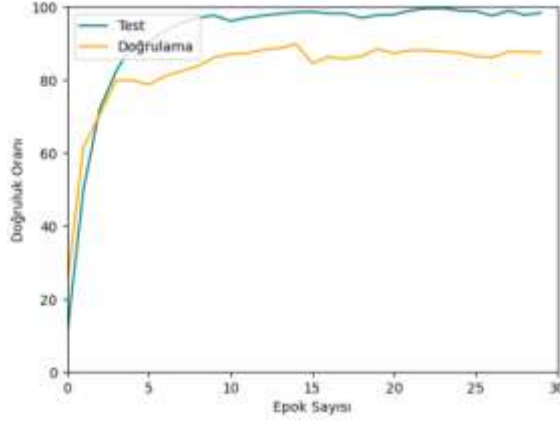


Şekil 14. Model-A doğruluk oranı grafiği

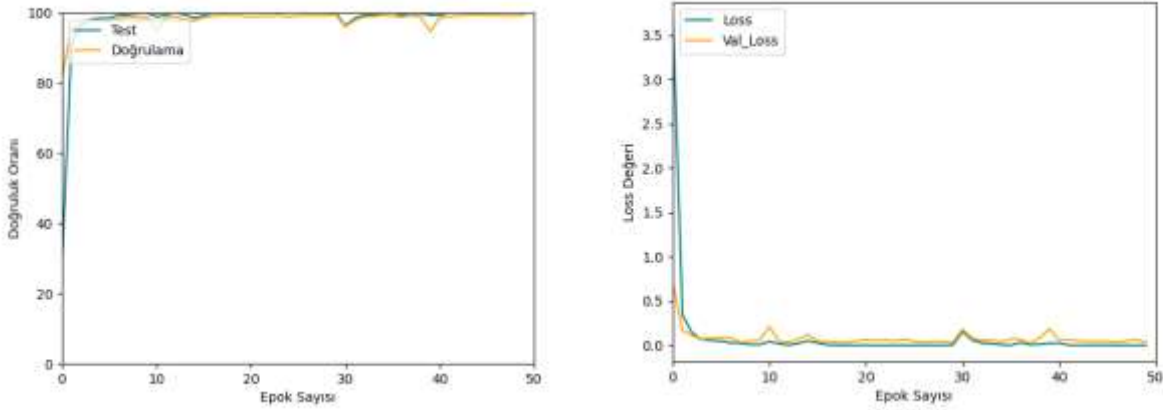
Model-A ESA modeli diğer eğitilen modellerden daha hızlı eğitim gerçekleştirdiği için veri azaltma işlemi uygulamadan, başarısız verilerle birlikte tekrar eğitilmiştir. Elde edilen eğitim sonucu, azaltılmış veri ile yapılan eğitim kadar olmasa da diğer modellerden daha yüksek doğruluk oranı yakalanmıştır.

Şekil 14'te de görüldüğü gibi başlangıçta düşük doğruluk oranı ile başlayıp hızlı bir artış göstermiştir. Doğrulama verisi genel olarak test verisinden düşük seviyede kalmıştır.

Şekil 15'te Model-A'nın azaltılmamış ve başarısız verilerin silinmediği veri seti ile olan eğitiminin doğruluk oranı grafiği verilmiştir.



Şekil 15. Model-A azaltılmamış veri ile eğitimin doğruluk oranı grafiği



Şekil 16. Sadeleşmiş veri seti ile eğitilmiş Model-A doğruluk ve loss oranı grafiği

Bu oluşturulan model sade veri seti ile 50 epok uzunluğunda tekrar eğitilmiştir. Yukarıdaki grafikte görüldüğü gibi başlangıç epoklarında %95 in üzerinde doğruluk oranı göstermiş ve devamında %100 doğruluk oranına ulaşmıştır. Genel olarak sade veri seti ile yapılan eğitimler, oluşturulan veri setinden çok daha yüksek doğruluk oranına sahip olmuştur.

Düşük sayıda sınıf olduğu için bu basit model yüksek doğruluk oranlarını yakaladığını, daha fazla sınıf olduğunda GoogLeNet, VGG-16 ve AlexNet gibi karmaşık modellerin daha başarılı olacağı tahmin edilmektedir. Test aşamasında, eğitilen ModelA test edilmemiştir. Doğru bir test olması için diğer modellere verilen veri seti ile eğitilen ilk model test edilmiştir.

2.4.5. Darknet

Darknet modeli, C ve CUDA dilleri kullanılarak geliştirilmiş bir açık kaynaklı nesne algılama framework'üdür, açık kaynak kodlu bir projedir ve GitHub üzerinden erişilebilir [23].

YOLO (You Only Look Once) algoritması da dahil olmak üzere derin öğrenme teknikleri kullanır [24]. Darknet, nesne algılama ve sınıflandırma, yüz tanıma, video analizi gibi görsel analiz problemlerini çözmek için özellikle etkilidir. Önceden eğitilmiş ağ modellerini kullanma ve yeni veri setlerine özelleştirme yeteneklerine sahiptir. Derin öğrenme modelinin eğitimini gerçekleştirmek için gerekli olan birtakım dosyaları da içerir. YOLO algoritması için Darknet, önceden eğitilmiş bir modeli kullanarak nesne tespiti işlemini gerçekleştirebilir [25].

Bu derin sinir ağını eğitmek için diğerlerinden farklı bir yöntem kullanılmıştır. Veri seti olarak Stable Diffusion ile çoğaltılan yüz verileri, model olarak da Yolov4 kullanılmıştır. Bu yöntemin eğitiminin diğerlerinden farkı yüz verileri fotoğraf üzerinde seçilerek kime ait olduğunun etiketlenmesidir. Etiket kullanılmadan eğitim yapılabilir fakat bu çalışmada etiketli veriler ile eğitim yapılmıştır. Elde edilen veri setindeki her bir fotoğrafta yüzler işaretlenip kime ait oldukları etiketlenmiştir. Yolov4 modeline bu etiketlenmiş yüz verileri verilerek derin sinir ağı eğitilmiştir.

Şekil 17'de Yolov4 modeli ile eğitim sonrası sonuçları verilmiştir. Bu eğitim sonucundaki genel başarı oranı %92,52 olarak elde edilmiştir. Daha detaylı veriler incelenecek olursa ise kişi bazlı olarak (name karşılığında) ortalama performans başarıları görülebilmektedir.

```

Calculation over (mean average precision)...
Detection layer: 130 type = Z8
Detection layer: 130 type = Z8
Detection layer: 141 type = Z8

200
-detections_count = 1829; unique_truth_count = 200
class_id = 0, name = Akshay Kumar, ap = 100.00% [TP = 10, FP = 3]
class_id = 1, name = Alexandra Daddario, ap = 96.69% [TP = 8, FP = 1]
class_id = 2, name = Alia Bhatt, ap = 68.81% [TP = 2, FP = 3]
class_id = 3, name = Amitabh Bachchan, ap = 100.00% [TP = 10, FP = 0]
class_id = 4, name = Andy Samberg, ap = 100.00% [TP = 10, FP = 3]
class_id = 5, name = Anushka Sharma, ap = 84.67% [TP = 10, FP = 6]
class_id = 6, name = Billie Eilish, ap = 100.00% [TP = 10, FP = 5]
class_id = 7, name = Brad Pitt, ap = 99.89% [TP = 9, FP = 0]
class_id = 8, name = Camila Cabello, ap = 94.87% [TP = 8, FP = 1]
class_id = 9, name = Courtney Cox, ap = 84.18% [TP = 8, FP = 5]
class_id = 10, name = Ellen DeGeneres, ap = 100.00% [TP = 10, FP = 2]
class_id = 11, name = Henry Cavill, ap = 87.73% [TP = 8, FP = 2]
class_id = 12, name = Hrithik Roshan, ap = 91.72% [TP = 9, FP = 1]
class_id = 13, name = Hugh Jackman, ap = 88.43% [TP = 9, FP = 0]
class_id = 14, name = Jessica Alba, ap = 85.36% [TP = 8, FP = 5]
class_id = 15, name = Katalie Portman, ap = 81.94% [TP = 8, FP = 4]
class_id = 16, name = Priyanka Chopra, ap = 96.81% [TP = 10, FP = 3]
class_id = 17, name = Robert Downey Jr, ap = 88.71% [TP = 10, FP = 5]
class_id = 18, name = Roger Federer, ap = 85.45% [TP = 8, FP = 4]
class_id = 19, name = Virat Kohli, ap = 100.00% [TP = 10, FP = 4]

for conf_thresh = 0.25, precision = 0.72, recall = 0.90, F1-score = 0.80
for conf_thresh = 0.25, TP = 100, FP = 09, FN = 30, average IoU = 58.48 %

IoU threshold = 50 %, used Area-under-Curve for each unique Recall
mean average precision (mAP@0.50) = 0.925182, or 92.52 %
total Detection Time: 10 seconds
    
```

Şekil 17. Yolov4 modeli ile eğitim sonrası sonuçları

Test için ayrılan veri setleri eğitilmiş ağa verildiğinde yüksek doğruluk oranı elde edilmiştir. Şekil 19'de test aşamasında alınan çıktılar verilmektedir.



Şekil 18. Yolov4 modeli örnek çıktı sonuçları

3. Bulgular

Bu aşamada eğitilen ESA modelleri test edilmiştir. Toplamda 1000 adet fotoğraf her bir modele verilerek fotoğrafların kime ait oldukları sorulmuştur. Bu 1000 fotoğrafın 750 adeti eğitime sokulan kişilerin eğitimde kullanılmamış fotoğrafları, 250 adeti ise eğitime hiç sokulmamış kişilerden oluşan fotoğraflardır.

GoogLeNet'in testi sırasında bir hata farkedilmiştir. GoogleNet eğitim sırasında yüksek doğruluk oranı vermesine rağmen test sırasında neredeyse hiçbir fotoğrafı doğru sınıflandırmamıştır. Bir kişinin eğitim yapılmış ve eklemeye yapılmamış 5 adet fotoğrafı modele verilmiştir. Model bu 5 fotoğrafı 4 farklı kişi ile eşleştirmiş ve bu 4 kişinin 4'ü de yanlış kişilerdir. Bu hata sebebiyle GoogLeNet ilk veri setinin test aşamasında kontrol edilmemiştir.

AlexNet, VGG-16 ve Model-A ESA modellerine hazırlanan 1000 fotoğraflık test verisi verilmiştir. Benzerlik oranı %85'in altında olan cevaplarda kişi tahmini doğru olsa dahi yanlış kabul edilmiştir.

Tablo 2. ESA Modellerinin Test Veri Seti Başarı Oranları

	AlexNet	VGG-16	Model-A
Test Veri Setindeki Yüzde (%) Başarı Oranı	72,3	66,2	68,7

Tablo 2'de AlexNet, VGG-16 ve Model-A modellerinin eğitimde kullanılmayan ve bir fotoğraf üzerinde birden fazla eklemenin yapıldığı test verisinde doğruluk oranları yüzde olarak verilmiştir. Bu modellerin eğitimde de bir fotoğrafa birden fazla eklemenin yapıldığı sade olmayan veri seti kullanılmıştır. AlexNet diğer modellerden yaklaşık yüzde beş oranında daha fazla doğruluk oranı vermiştir.

Bu değerler beklenen başarı oranlarına göre çok düşüktür. Eğitime sokulmayan fotoğraflara yapılan eklemeler ile eğitime sokulmuş fotoğraflardaki eklemeler birbirleri ile aynı olduğu için %85-90 benzerlik oranı ile eşleştirme yapmaktadır. Az sayıda eklemeye yapılmış fotoğraflar ise %60-65 oranında benzerlik yakalamışlardır. Bunun dışında hata oranının yüksek olması sebep olan iki şey daha vardır. Bunlardan ilki modelin döndürdüğü sınıfın doğru olduğu ama benzerlik sınırının altında kaldığı için yanlış sayılan fotoğraflar. Bu durumu düzeltmek için benzerlik sınırı düşürülebilir fakat bazı test verilerinde sınıf doğru olmasına rağmen %40-45 gibi düşük benzerlik oranı vermektedir.

Diğer bir hata ise veri setindeki dengesizlik. Veriler çoğaltılırken depolama alanı sorunu sebebiyle eklenen filtreler belli sayıda ve rastgele olarak seçilmiştir. Bir gözlük üzerinden örnek verilecek olursa, bir kişi veri çoğaltma aşamasında gözlüğü sadece 1 yerde ve başka bir kişi aynı gözlüğü 3 farklı yerde çoğaltmış varsayalım. Eğitime ilk kişinin bu gözlük ile 10 adet fotoğrafı ikinci kişinin ise 30 fotoğrafı verilmiştir. Test aşamasında ilk kişinin bu gözlük ile olan farklı bir fotoğrafı modele verildiğinde model bu kişiyi ikinci olarak işaretlemektedir. Veri setinin homojen ve doğru olmasının önemi bu kısımda tekrardan anlaşılmıştır.

Bu sebeple sade bir veri seti oluşturulmuştur. Bu veri setinde fotoğraflara tüm eklemeler uygulanmış fakat aynı anda sadece bir adet eklemeye kullanılmıştır. Böylece eğitim aşamasında homojen bir veri seti kullanılmıştır.

Yeni eğitilen modellerin test verileri önceki eğitimin testinde kullanılan 1000 fotoğraf ile aynı fotoğraflardır. Önceki veri setinde olduğu gibi doğru cevap için %85'in üstünde bir benzerlik oranı beklenmektedir. Aşağıdaki Tablo 3'te ModelA, AlexNet, VGG-116 ve GoogleNet'in test sonucu doğruluk oranları verilmektedir.

Tablo 3.ESA Modellerinin sadeleşmiş veri seti ile eğitilmiş Test Veri Seti Başarı Oranları

	AlexNet	VGG-16	Model-A	GoogLeNet
Test Veri Setindeki Yüzde (%) Başarı Oranı	65,0	69,0	72,2	84,2

Tablo 3'te verilen doğruluk oranları eğitim veri setinde bir fotoğrafa maksimum bir adet eklemenin yapıldığı sade veri seti ile eğitilmiş modellerin test sonucu verilmiştir. Bu testte eğitim veri setinin aksine bir fotoğrafta birden fazla eklemenin yapıldığı fotoğraflar kullanılmıştır. Tablo 3'te görüldüğü gibi VGG-16 ve Model-A'nın doğruluk oranında artış gözlemlenirken AlexNet'te düşüş görülmüştür. AlexNet yeni veri setinde bazı kişilerin tüm fotoğraflarını doğru bilirken bazı kişilerin fotoğraflarını eğitime katılmış başka biriyim gibi tek bir kişi olarak

hatalı bilmektedir. Bu sebeple doğruluk oranı düşmüştür. GoogleNet, bu veri setinde hatasız çalışmış ve diğer modellerden daha yüksek doğruluk oranına ulaşmıştır; ayrıca, eğitime sokulmamış kişilerin sınıflandırılmasında da diğerlerine göre önemli ölçüde daha başarılı olmuştur. Test sonuçları ele alındığında sadeleştirilmiş veri seti ile yapılan eğitimin, eklemeleri sınırlandırarak homojen olmayan veri setinin eğitiminden çok daha başarılı olduğu görülmüştür. GoogleNet hem eğitime katılmış kişilerin hem de eğitime katılmamış kişilerin sınıflandırılmasında diğerlerinden başarılı olmuştur. Ayrıca sadeleştirilmiş veri seti ile eğitilen modeller birden fazla ekleme olan fotoğrafları da doğru bilmiştir. Bu sebeple bir fotoğrafa birden fazla ekleme yapmaya gerek görülmemektedir.

4. Tartışma ve Sonuç

Mediapipe yöntemi kullanılarak kodlar ile otomatize edilebilen veri çoğaltma işlemi, yapılan çalışmanın tamamlanmış ve sonuçlanmış bir aşamasını oluşturmaktadır. Test aşamasında meydana gelen homojenlik probleminden dolayı her fotoğrafa bir ekleme yapılmıştır. Veri seti oluşturulurken hatalı veriler çıkarılmamış, modeli güçlendirmek için veri seti içerisinde bırakılmıştır.

Stable Diffusion ile fotoğraflara eklenen gözlük, şapka, sakal vb. eklentilerin görüntüsü Mediapipe ile yapılan eklemelere göre çok daha başarılı olmuştur. Fakat bu işlemde çıktı almak çok uzun sürmektedir. Ayrıca tek seferde çok fazla çıktı alınamamakta, alınan çıktılar bazılarının da bozuk olabilmektedir. Stable Diffusion, kurumun kendi ağlarını daha güçlü hale getirmek için veri setine takviye yapmak isterlerse kullanılabilir bir yöntem olabilir.

Yapılan ESA modelleri eğitimlerinde doğruluk oranı dört modelde de yüksek sayılara ulaşmıştır. Eğitimlerde doğruluk oranları tüm modellerde birbirine çok yakındır.

Veri seti daha homojen yapıldığında test başarıları daha düzgün seviyelere gelmektedir. Bu modeller arasında yoklama sisteminde kullanmaya en uygunu, test aşamasında diğerlerinden daha yüksek başarı oranına sahip olan GoogleNet olmuştur.

Bu ağ modellerinden farklı olarak eğitilen Yolov4 modeli yoklama sistemi için ana Evrişimli Sinir Ağı olamaz. Çünkü her bir fotoğraf tek tek el ile işaretlenip kime ait olduğu etiketlenmesi gerekmektedir. Fakat bu yöntem dönem ortasında kuruma yeni katılan kişilerin sonraki büyük ESA eğitimine kadar geçici süre yoklama için ufak eğitimler ile kullanılabilir.

Yüz görüntülerine filtre eklemek, literatürde daha önce denenmiş bir yaklaşımdır. Evrişimli Sinir Ağları Kullanarak Yüz Belirleme ve Tanıma Uygulaması, önceki uygulamalardan birkaç önemli noktada farklılık göstermektedir. Önceki uygulamalar genellikle tek bir filtreye odaklanırken, yeni uygulama birden fazla filtre kullanarak insan yüzüne çeşitli kılık değiştirmelerle karşı kimlik tespitinin doğruluğunu artırmayı amaçlamaktadır [26].

İki uygulama arasındaki temel fark, yeni uygulamanın daha kapsamlı olmasıdır. Yeni uygulama, yüz tanımanın çeşitli yönlerini içermektedir ve kişiyi çeşitli durumlarda tanıma potansiyeline sahiptir [26]. Genel olarak, yeni uygulama önceki uygulamalara göre daha iyi sonuçlar sunmaktadır. Bu uygulama, homojen veri seti kullanılarak kişinin filtrelenmiş halinin tespitinde başarılı olmuştur [26].

Bu çalışmada, AlexNet, VGG-16, GoogleNet, DarkNet ve daha basit bir yapıya sahip olan Model-A modellerinin, birden fazla kişinin kullanıldığı bir yüz tanıma uygulamasındaki doğruluk yetenekleri incelenmiştir. Modellerin insan yüzündeki fiziksel değişikliklere karşı daha etkili tahminler yapabilmesi için eğitim verilerine bir veya daha fazla fotoğraf eklenmiştir. Eklenen fotoğrafların modellerin bu değişikliklere tepkilerini değerlendirmek amacıyla başarı oranları analiz edilmiş ve elde edilen sonuçlar, literatüre önemli bir katkı sağlamıştır.

TEŞEKKÜR

Bu araştırma 2209-A Üniversite Öğrencileri Araştırma Projeleri Destekleme Programı Kapsamında 1919B012220788 numarası ile desteklenmiştir.

ETİK STANDARTLARIN BEYANI

Bu makalenin yazar(lar)ı çalışmalarında kullandıkları materyal ve yöntemlerin etik kurul izni ve/veya yasal özel bir izin gerektirmediğini beyan ederler.

YAZARLARIN KATKILARI

Özge Gürbüz-Problemi tanımlamış, konu ile ilgili literatürü taramış, problemi modellemiş, uygulama sonuçlarını analiz etmiş ve çalışmanın yazım işlemini gerçekleştirmiştir.

Tolga Yılmaz-Problemi tanımlamış, konu ile ilgili literatürü taramış, problemi modellemiş, uygulama sonuçlarını analiz etmiş ve çalışmanın yazım işlemini gerçekleştirmiştir.

ÇIKAR ÇATIŞMASI

Yazarlardan Özge Gürbüz Çalık Holding firmasında çalışmaktadır.

KAYNAKLAR

- [1] Li, M., Huang, B., & Tian, G. (2022). A comprehensive survey on 3D face recognition methods. *Engineering Applications of Artificial Intelligence*, 110, 104669.
- [2] Türkoğlu M., Hanbay K., Saraç Sivrikaya I. & Hanbay D., (2020). Derin Evrişimsel Sinir Ağı Kullanılarak Kayısı Hastalıklarının Sınıflandırılması. *BEÜ Fen Bilimleri Dergisi*, 9(1), 334-345.
- [3] Wu, C., Hu, H., Lin, K., Wang, Q., Liu, T., & Chen, G. (2023). Attention-guided and fine-grained feature extraction from face images for gaze estimation. *Engineering Applications of Artificial Intelligence*, 126, 106994.
- [4] Ou, W., You, X., Tao, D., Zhang, P., Tang, Y., & Zhu, Z. (2014). Robust face recognition via occlusion dictionary learning. *Pattern Recognition*, 47(4), 1559-1572.
- [5] Li, Y., Song, L., Wu, X., He, R., & Tan, T. (2019). Learning a bi-level adversarial network with global and local perception for makeup-invariant face verification. *Pattern Recognition*, 90, 99-108.
- [6] <https://fei.edu.br/~cet/facedatabase.html> FEI face database (Erişim tarihi: 25.02.2023)
- [7] <http://vis-www.cs.umass.edu/lfw/> Labeled Faces in the Wild (Erişim tarihi: 25.02.2023)
- [8] Alimovski, E. (2019). Derin öğrenmeye dayalı güçlü yüz tanıma sistemi için gan ile veri çoğaltma (Master's thesis, İstanbul Sabahattin Zaim Üniversitesi, Fen Bilimleri Enstitüsü, Bilgisayar Mühendisliği Anabilim Dalı).
- [9] <https://medium.com/deep-learning-from-deepest/derin-ogrenme-ve-evrisimsel-sinir-aglari-446dc8a8d2f> Derin Öğrenme ve Evrişimsel Sinir Ağları (Erişim Tarihi: 16.09.2023)
- [10] Sert Z., ESA (Evrişimsel Sinir Ağları) <https://zeysert.medium.com/esa-evrisimsel-siniraglari-87d9bd986579> (Erişim tarihi: 07.07.2023)
- [11] <https://developers.google.com/mediapipe> Mediapipe (Erişim tarihi: 14.03.2023)
- [12] <https://github.com/google/mediapipe/tree/a908d668c730da128dfa8d9f6bd25d519d006> Mediapipe (Erişim Tarihi: 7 Temmuz 2023)
- [13] <https://www.makesense.ai> MakeSense (Erişim Tarihi: 16.09.2023)
- [14] <https://neuroflash.com/blog/understanding-stable-diffusion-inpainting-aguide/#:~:text=Stable%20Diffusion%20Inpainting%20is%20a,damaged%20parts%20of%20an%20image> Understanding Stable Diffusion Inpainting: A Guide (Erişim tarihi: : 07.07.2023)
- [15] <https://stablediffusionweb.com/WebUI> Stable Diffusion (Erişim Tarihi: 16.09.2023)
- [16] <https://www.hosting.com.tr/blog/midjourney/> Midjourney Nedir? Yaratıcılığınızı Görsellere Çevirin – Yapay Zekâ ile Mükemmel Görseller Oluşturma (Erişim tarihi: 07.07.2023)
- [17] Krizhevsky A., Sutskever I., & Hinton G. E., (2012). ImageNet Classification with Deep Convolutional Neural Networks. *Advances in Neural Information Processing Systems*, 25(2),1097-1105.
- [18] Ghazi M. M., Yanikoglu B. & Aptoula E., (2017). Plant identification using deep neural networks via optimization of transfer learning parameters. *Neurocomputing*, 235, 228-235.
- [19] <https://towardsdatascience.com/what-are-hyperparameters-andhow-to-tune-the-hyperparameters-in-a-deep-neural-network-d0604917584a> (Erişim tarihi: 07.07.2023)
- [20] Simonyan K. & Zisserman A. (2014). Very Deep Convolutional Networks for LargeScale Image Recognition. *arXiv preprint arXiv:1409.1556*.
- [21] Kurt F., (2018). Evrişimli Sinir Ağlarında Hiper Parametrelerin Etkisinin İncelenmesi, Yüksel Lisans tezi, Hacettepe Üniversitesi.
- [22] Szegedy C., Liu W., Jia Y., Sermanent P., Reed S., Anguelov D., Erhan D., Vanhoucke V. & Rabinovich A., (2015). Going deeper with convolutions, 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), (ss 1-9).
- [23] <https://github.com/pjreddie/darknet> Darknet (Erişim tarihi: 07.07.2023)
- [24] <https://pjreddie.com/darknet/> Darknet: Open Source Neural Networks in C (Erişim tarihi: 07.07.2023)
- [25] Ozel, M. & Baysal, S. & Sahin, M. (2021). Derin Öğrenme Algoritması ile Dinamik Test Süresince Süspansiyon Parçalarında Çatlak Tespiti. *Avrupa Bilim ve Teknoloji Dergisi*, (26), 1-5.
- [26] Yılmaz, A. G., Gedikli, E., & Alhori, O. (2023). Yüz Görüntülerine Morflemeye Dayalı Maske Giydirmeye ve Maskeli Yüz Tanıma. *Süleyman Demirel Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, 27(1), 12-21.