



GÖRÜNTÜ SINIFLANDIRMADA DERİN ÖĞRENME YÖNTEMLERİNİN KARŞILAŞTIRILMASI

¹İlker DAĞLI , ^{2,3}Ali ÖZTÜRK 

¹Necmettin Erbakan Üniversitesi, Bilgi İşlem Daire Başkanlığı, Konya, TÜRKİYE

²KTO Karatay Üniversitesi, Uygulamalı Bilimler Yüksekokulu, Pilotaj Bölümü, Konya, TÜRKİYE

³Havelsan A.Ş., Ankara, TÜRKİYE

¹idagli@erbakan.edu.tr, ²ali.ozturk@karatay.edu.tr, ³aliozturk@havelsan.com.tr

(Geliş/Received: 22.04.2020; Kabul/Accepted in Revised Form: 06.10.2021)

ÖZ: Bu çalışmada ESA (Evrşimsel Sinir Ağları), ResNet ve AİA (Ağ İçinde Ağ) yaklaşımları kullanılarak oluşturulan ve E-Model, R-Model, A-Model şeklinde adlandırılan derin öğrenme modellerinin farklı veri kümeleri üzerinde performansları karşılaştırılmıştır. CIFAR-10 veri kümesi için derin öğrenme modelleri sadece MİB (Merkezi İşlem Birimi) içeren bir makinede ve MİB ile GİB (Grafik İşlem Birimi) içeren bir makinede ayrı ayrı çalıştırılmıştır. Sadece MİB içeren makinede R-Model, A-Model ve E-Model için sırasıyla yaklaşık 415 saatlik, 129 saatlik ve 3.5 saatlik eğitim aşamaları sonucunda doğrulama veri seti üzerinde sırasıyla %82.76, %87.64 ve %83.47 doğruluk oranları elde edilmiştir. MİB ve GİB içeren makinede ise R-Model, A-Model ve E-Model için sırasıyla yaklaşık 4.45 saatlik, 2.20 saatlik ve 1.82 saatlik eğitim aşamaları sonucunda doğrulama veri seti üzerinde sırasıyla %82.61, %87.95 ve %82.43 doğruluk oranları elde edilmiştir. Diğer veri kümeleri için ise modeller MİB ve GİB içeren makinede çalıştırılarak deneysel sonuçlar elde edilmiştir. Oluşturulan derin öğrenme modellerinin yapıları, eğitim için kullanılan parametre değerleri, doğrulama verileri için elde edilen karmaşıklık matrisleri, doğruluk ve kayıp grafikleri ayrıntılı olarak verilmiştir.

Anahtar Kelimeler: Derin öğrenme, Görüntü Sınıflandırma, ESA, CIFAR-10, SVHN, Parmaklar, Çiçekler, Ürün Kalitesi, Balıklar

Comparison of Deep Learning Models in Image Classification

ABSTRACT: In this study, various experiments have been performed via deep learning models based on CNN (Convolutional Neural Networks), ResNet (Residential Energy Services Network) and NIN (Network In Network) approaches and their performances on various datasets have been investigated. The deep learning models were named as E-Model, R-Model and A-Model, respectively. The deep learning models were trained with CIFAR-10 dataset on a machine having only CPU (Central Processing Unit) and a machine having both CPU and GPU (Graphical Processing Unit). On the machine having only CPU, the training time of the R-Model, A-Model and E-Model were approximately 415 hours, 129 hours and 3.5 hours, respectively. The percentage correct values on the validation data set were %82.76, %87.64 ve %83.47, respectively. On the machine having both CPU and GPU, the training time of the R-Model, A-Model and the E-Model were approximately 4.45 hours, 2.20 hours and 1.82 hours, respectively. The percentage correct values on the validation data set were %82.61, %87.95 ve %82.43, respectively. The experimental results for the other data sets were obtained by training the models on the machine having both CPU and GPU. The structures of the constructed deep learning models, the parameters used for the training, the obtained confusion matrices for the validation data, the accuracy and loss graphics are given in detail.

Key Words: Deep learning, Image Classification, CNN, CIFAR-10, SVHN, Fingers, Flowers, Product Quality, Fishes

GİRİŞ (INTRODUCTION)

Derin öğrenme teknikleri günümüzde resim, ses, el yazısı, trafik işaretleri gibi verilerin tanınmasında veya sınıflandırılmasında kullanılmakta ve yüksek doğruluk oranlarıyla dikkat çekmektedir. McCulloch ve Pitts (1943) insan sinir sisteminin temel yapısı olan sinir hücresinin modelini tanımlamıştır. Yapay sinir ağlarının (YSA) da başlangıcı olan bu model, doğrusal problemler üzerinde çok iyi sonuçlar vermektedir. Bununla birlikte, aynı avantaj, XOR problemleri gibi doğrusal olmayan problemlerde elde edilemez. Bu nedenle yapay sinir ağlarına olan ilgi azalmıştır. 1980'lerde insan beyninden ilham alan araştırmalar paralel dağıtık bir model önermiş ve derin öğrenmenin temelleri atılmıştır. Geri yayılma algoritması, geçmişte sinir ağlarını eğitmek için yaygın olarak kullanılmıştır (Rumelhart ve diğ., 1986). 2012 yılında derin öğrenme büyük bir sıçrama yapmıştır. Dünya çapında büyük ilgi çeken ImageNet (Deng ve diğ., 2009) yarışmasında, bir ESA modeli birinci olmuştur (Krizhevsky ve diğ., 2012). Bu derin öğrenmenin ilk başarısı olmakla birlikte, günümüzde derin öğrenmenin popüler olmasının nedenlerinden biri, yeterli miktarda veri üretilebilmesidir. Dünyanın dijitalleşmesiyle veri miktarı daha da büyümekte ve bu büyüyen veriyle birlikte derin öğrenme algoritmaları daha yüksek performans sonuçları elde etmektedir. Derin öğrenme algoritmaları yüksek başarımlar elde ettiği için günümüzde ses, resim, robotik, otonom araçlar gibi birçok alanda kullanılmaktadır. Hatta bazı çalışmalarda insanın başarısının üstüne çıktığı görülmekte ve bu çalışmalar insanların resim tanıma konusundaki bakış açılarının değişmesine sebep olmuştur (Ganin ve diğ., 2016). Görüntü sınıflandırma ve tanıma, günümüzün personel takibi gibi sıradan bir otomasyon sisteminde de kullanılmaktadır. Bu sistemlerde arka planda derin öğrenme ve görüntü işleme kullanılabilmektedir.

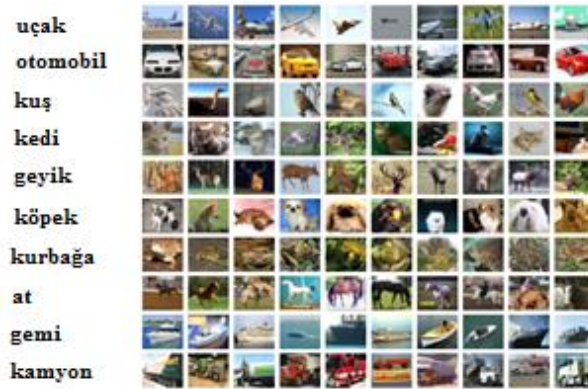
Literatürde farklı derin öğrenme modelleri önerilmiş ve performanslarının karşılaştırılmasında belirli veri kümeleri kullanılmıştır. CIFAR-10 bu amaçla yaygın olarak kullanılan veri kümelerinden birisidir. Ciregan ve arkadaşlarının önerdiği çok kolonlu ESA modelinde birden fazla ESA modelinin eş zamanlı olarak ürettiği sonuçların ortalaması alınarak genel sınıflandırma sonucu elde edilmektedir (Ciregan ve diğ., 2012). CIFAR-10 veri kümesi için uyguladıkları 8 kolonlu ve 10 katmanlı ESA modeliyle %88.8 doğruluk oranı elde etmişlerdir. Fakat kullandıkları her bir ESA modeli 1.210,710 parametre içermektedir. Bu değeri 8 ile çarpınca yaklaşık 9 milyon parametrelili bir model ortaya çıkmaktadır. Goodfellow ve arkadaşları, önerdikleri ESA modelinde girişlerinden en büyüğünü seçen aktivasyon fonksiyonuna sahip sinir hücreleri kullanmışlardır (Goodfellow ve diğ., 2013). CIFAR-10 veri kümesi için kullandıkları modelde toplam parametre sayısı 1,034,298 olup %90.6 doğruluk oranı elde etmişlerdir. Graham tarafından önerilen ESA modeli çok fazla boş değer içeren veri kümelerinde verimli çalışma hedefiyle ortaya çıkmıştır (Graham, 2014). Modelin toplam parametre sayısı yaklaşık olarak 11,250,000 olup CIFAR-10 veri kümesi üzerindeki doğruluk oranı %93.7'dir. Springenberg ve arkadaşları ESA modelinde kullanılan havuz katmanlarının uygun evrimsel katmanlarla değiştirilerek sadece evrimsel katmanlardan oluşan daha sade ESA modeli oluşturulabileceğini öne sürmüşlerdir (Springenberg ve diğ., 2015). Bu şekilde oluşturdukları toplam 1,300,000 parametrelili ESA modelinin CIFAR-10 veri seti üzerindeki doğruluk oranı %90.9 olarak bulunmuştur. He ve arkadaşlarının önerdikleri toplam 10,200,00 parametrelili ResNet-1001 modelinin CIFAR-10 veri seti üzerindeki doğruluk oranı %95.4'tür (He ve diğ., 2016). Huang ve arkadaşları, çok fazla katmandan oluşarak derinleşen ESA modellerinde katmandan katmana geçişte oluşan bilgi kaybını azaltmak üzere katmanlar arası fazladan bağlantılar içeren yoğun biçimde bağlı ESA modeli önermişlerdir (Huang ve diğ., 2017). CIFAR-10 üzerinde uyguladıkları model 190 katmandan oluşmakta ve toplam 25,600,000 parametre içermektedir. Bu modelin doğruluk oranı %96.5 olarak bulunmuştur. Liu ve Mukhopadhyay, ImageNet ile önceden eğitilmiş ResNet-50 (He ve diğ., 2016) modelinin çıkış katmanından önceki havuz katmanında bulunan resim özelliklerinin bir Hopfield ağında depolanmasına dayanan bir model önermişlerdir (Liu ve Mukhopadhyay, 2018). Önerdikleri model CIFAR-10 veri kümesi üzerinde %83.1 doğruluk oranına sahip olup, modelde yararlandıkları ResNet-50 toplam 25,636,712 parametre içermektedir. Wang ve arkadaşlarının önerdiği ve özellikli düğümleri arasındaki bağlantı durumunun adaptif olarak belirlendiği ESA modeli yaklaşık 19,800,000

parametreye sahip olup CIFAR-10 veri kümesi üzerinde %94'lük doğruluk oranına sahiptir (Wang ve diğ., 2019). Chu ve arkadaşlarının önerdiği gürültülü ayırt edilebilir mimari aramasına dayanan model, toplam 5,500,000 parametre içermekte olup %98.28 doğruluk oranına sahiptir (Chu ve diğ., 2020). Kabir ve arkadaşlarının önerdiği ve gizli katmanlara gelen ağırlık girişlerinin az olduğu dereceli girişe dayanan SpinalNet modeli 20,160,000 parametre içermekte olup, %91.4 doğruluk oranına sahiptir (Kabir ve diğ., 2020). Lu ve arkadaşlarının önerdiği sinirsel mimari transferine dayanan model 6,900,000 parametreye sahip olup %98.4 doğruluk oranına sahiptir (Lu ve diğ., 2021).

ESA, ResNet ve AİA yaklaşımları ile oluşturulan derin öğrenme modellerinin performanslarının karşılaştırıldığı bu çalışmada, bu amaçla yaygın olarak kullanılan veri kümeleri yanında daha önce derin öğrenme modelleri uygulanmamış geniş ölçekli balık veri kümesi (Ulucan ve diğ., 2020) üzerinde sınıflandırma işlemleri gerçekleştirilmiştir. Oluşturulan bu üç modelin parametre sayısının literatürde kullanılan modellerden daha az olmasına dikkat edilmiştir. Parametre sayısının az olması, modelin depolama ve bellek gereksinimini azaltan bir durum olduğu için bu modelleri kullanacak uygulamaların etkinliği için avantaj sağlayacağı düşünülmüştür.

VERİ KÜMELERİNİN TANIMI (DESCRIPTION OF DATA SETS)

Cifar-10 (Krizhevsky, 2009) görüntü tanıma ve sınıflandırma çalışmalarında kullanılan 32x32 RGB biçiminde 60,000 adet görüntü içeren bir veri kümesidir. Bu görüntüler uçaklar, arabalar, kuşlar, kediler, geyikler, köpekler, atlar, gemiler, kamyonlar olmak üzere 10 sınıftan oluşur. Bu çalışmada, CIFAR-10 veri kümesindeki 50000 görüntü eğitim, 10000 görüntü test için kullanılmıştır. Şekil 1'de örnek resimler ile CIFAR-10 veri kümesine ait sınıflar gösterilmektedir.



Şekil 1. CIFAR 10 Veri Kümesi

Figure 1. CIFAR-10 Dataset

SVHN (Google Street View House Numbers) veri kümesi (Netzer ve diğ., 2011), 0 ila 9 arasındaki 10 farklı sınıfa ait rakamların görüntülerini içermektedir. Görüntüler, değişik çözünürlüklerde, renkli ev numaralarına ait rakamlardan oluşmaktadır. SVHN veri kümesi 73,257 adet eğitim verisi ve 26,023 test verisi içermektedir.

Bu çalışmada kullanılan diğer veri kümeleri kaggle platformunda yayınlanmış bulunan Parmaklar (Koryakin, 2019), Çiçekler (Nilsback ve Zisserman, 2006), Ürün kalitesi (Dabhi, 2020) ve Balıklar (Ulucan ve diğ., 2020) veri kümeleridir.

Parmaklar veri kümesi, işaret dili ile 0 ila 5 arasındaki rakamların parmaklar yardımıyla gösterimini içeren görüntülerden oluşmaktadır. Eğitim kümesi 18,000 görüntü ve test kümesi 36,000 görüntü içermektedir.

Çiçekler veri kümesi, 17 farklı çiçeğe ait görüntülerden oluşmaktadır. Eğitim kümesinde her çiçek için 80 görüntü, test kümesinde ise her çiçek için 10 görüntü bulunmaktadır. Dolayısıyla, eğitim kümesi toplam 1360 test kümesi ise 170 görüntüden oluşmaktadır.

Ürün kalitesi veri kümesi, metal döküm sürecinde ürünlerin yüzeylerinde oluşan istenmeyen düzensizliklerin tespit edilmesi amacıyla oluşturulmuştur. Veri kümesinde, yüzeyin düzgün olduğu ve arızalı olduğu ürün görüntüleri yer almaktadır. Eğitim kümesinde 6633 görüntü, test kümesinde ise 715 görüntü bulunmaktadır.

Balıklar veri kümesi (Ulucan ve diğ., 2020), 9 farklı deniz ürününe ait farklı açılardan çekilmiş görüntülerden oluşmaktadır. Eğitim kümesi 9000 görüntü, test kümesi ise 430 görüntü içermektedir.

DERİN ÖĞRENME YÖNTEMLERİ (DEEP LEARNING METHODS)

ESA (Evrşimsel Sinir Ağları)

Evrşimsel sinir ağları ilk olarak 2012 yılında nesne tanıma alanında yapılan ImageNet yarışmasında Krizhevsky'nin AlexNet (Krizhevsky ve diğ., 2012) modeli ile birinci olmuş ve daha önceki hata oranlarında yaklaşık %10'luk bir düşüş yaratarak modelin potansiyelini ortaya koymuştur (İnik ve Ülker, 2017). Günümüzde artan miktarda veri ve grafik işlemci üniteler gibi donanımlarla bilgisayar performanslarındaki artış, ESA gibi derin öğrenme yöntemlerinin başarısı ve popülerliği üzerinde büyük bir etkiye sahiptir. ESA, klasik YSA'lardan farklı olarak herhangi bir öznitelik girdisi beklemeyerek, girilen nesnelere göre kendi özniteliklerini hiyerarşik olarak basitten karmaşığa doğru çıkarmaktadır. Örneğin ilk katmanlarda kenar, köşe bilgisi gibi basit öznitelikler elde edilirken daha sonraki katmanlarda görüntüye ait daha detaylı öznitelikler elde edilmektedir (Kızrak ve Bolat, 2018). Bu dinamik yapı, ESA'nın sınıflandırma başarısında önemli bir artış elde edilmesini sağlamıştır. Dahası, ESA mimarisinde kullanılan bırakma yöntemi, ağı eğitim verilerinin ezberlenmesini ve aşırı kullanılma sorununu önlemiştir. ESA mimarileri, Şekil 2'de gösterildiği gibi Giriş (Input), Özellik Öğrenimi (Feature Learning) ve Sınıflandırma (Classification) olmak üzere 3 ana bölümden oluşmaktadır. Bu bölümler, işlevsel olarak sırasıyla Girdi (Input), Konvolüsyon (Convolution), Havuz (Pooling), Tam Bağlı (Fully Connected), Bırakma (Dropout), Düzleştirme (Flattening) ve Sınıflandırma (Classification) katmanı olmak üzere bazıları tekrarlı olmak üzere 7 ana katmandan oluşmaktadır. Bu katmanlar, çalışmada kullanılan veri kümesi için aşağıda açıklanmıştır.

Girdi Katmanı (Input Layer)

Bu katman ESA mimarisinin ilk katmanı olup, verilerin diğer katmanlara aktarıldığı bölümdür. Örneğin, bu çalışmada giriş katmanını oluşturmak için 32x32 boyutlu renkli görüntülerden oluşan bir veri kümesi kullanılmaktadır. Performans ve doğruluk oranını artırmak ve aşırı eğitimi (overfitting) önlemek için bu katmana girdi olan veriler üzerinde veri ön işleme (data preprocessing) ve veri çoğaltma (data augmentation) işlemleri de yapılmaktadır (İnik ve Ülker, 2017).

Konvolüsyon Katmanı (Convolution Layer)

Bu katman, ESA mimarilerinin temelidir ve bu katmandaki filtrelerle özellik çıkarımı yapılır. Adım değerine göre filtreler, x ve y eksenine göre simetriği alınarak veya alınmadıkça (çapraz korelasyon) giriş görüntüleri üzerinde kaydırılır. Giriş değerleri ve filtre değerleri her adımda çarpılmakta, toplanmakta ve çıkış matrisine yazılmaktadır. Ardından aktivasyon fonksiyonuna girmeden önce bir eğilim (bias) değeri ile toplanmaktadır (Kızrak ve Bolat, 2018). Konvolüsyon işleminde giriş matrisi ile çıkış matrisi arasında boyut farklılıkları olabilir. Eğer girdi boyutu $(n \times n) = (32 \times 32)$, filtre boyutu $(f \times f) = (3 \times 3)$, ve adım kaydırma değeri (stride = 1) ise, çıktı matrisi boyutu denklem 1 kullanılarak elde edilir.

$$\left[\frac{(n-f)}{s} + 1 \right] \times \left[\frac{(n-f)}{s} + 1 \right] \quad (1)$$

Çıktı matrisinin boyutu küçültmek istenmiyorsa, giriş matrisinin etrafında "dolgu" adı verilen piksel ekleme yöntemi kullanılabilir. Bu amaçla Denklem 2'de görülen hesaplama ile çıktı boyutu hesaplanabilir.

$$\left[\frac{(n + 2p - f)}{s} + 1 \right] \times \left[\frac{(n + 2p - f)}{s} + 1 \right] \quad (2)$$

Önceki örnek için aynı değerlere 1 dolgu değeri ($p = 1$) eklenirse, sonucun giriş boyutuna eşit (32×32) olarak değiştiği görülecektir (Kızrak ve Bolat, 2018).

Havuz Katmanı (Pooling Layer)

Bu katmanda, ağ üzerinde özellik çıkartma veya öğrenme gerçekleştirilmez. Performans iyileştirme ve maliyet azaltma amacıyla, giriş matrisinin genişliği ve uzunluğu kanal numarası sabit tutularak azaltılır. Örneğin, giriş değeri $32 \times 32 \times 64$ iken çıkış değeri $16 \times 16 \times 64$ olarak hesaplanabilir. Bu, bilginin azalmasına yol açar, bu da maliyetleri azaltmak için bilgilerin tehlikeye atılması anlamına gelir. ESA mimarilerinde sıkça kullanılan havuzlama yöntemleri maksimum havuzlama (max pooling) ve ortalama (average) havuzlamadır (İnik ve Ülker, 2017).

Tam Bağlı Katman (Full Connected Layer)

Bu katman, önceki bölümlerde anlatılan ve ardı sıra tekrar eden öznelik çıkarım katmanlarından sonra gelmektedir. Kendisinden önceki katmandaki tüm alanlar ile bağlantılıdır (İnik ve Ülker, 2017). Örneğin, $16 \times 16 \times 256$ boyutundaki çıktı, öncelikle 4096×1 matris ile tam bir bağlantı yaparak, 65536×1 'e dönüştürülür. Sınıflandırma katmanına gelmeden önce, tüm nitelikler 65536×4096 boyutundaki geleneksel bir sinir ağına aktarılır.

Bırakma Katmanı (Dropout Layer)

Ağın bazı düğümlerini kaldırarak ağın aşırı yüklenmesini önlemek için kullanılan katmandır (İnik ve Ülker, 2017).

Sınıflandırma Katmanı (Classification Layer)

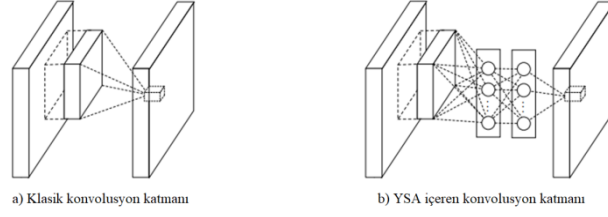
Tam bağlı katmanı takip eden ve ağın sınıflandırılmasını ve tahminini yapan katmandır. CIFAR-10 veri kümesi için tam bağlı katmandan 4096 nöron 10 nörona indirgenir ve bir 4096×10 matris yapısı oluşur. Sınıflandırma katmanı genellikle başarısından dolayı softmax fonksiyonunu kullanır. CIFAR-10 veri kümesi için Softmax, 0 ile 1 arasında 10 sınıflandırma değeri oluşturur ve 1'e en yakın değer ağımızın tahminidir (İnik ve Ülker, 2017).

Düzleştirme Katmanı (Flattening Layer)

Bu katmanın görevi Tam bağlı katmanın girişindeki verileri hazırlamaktır. Matris halinde bulunan görüntünün tek boyutlu vektöre dönüştürülmesini sağlamaktadır. Düzleştirme katmanı bu işlemi otomatik olarak yapmaktadır. Bu işlem aynı zamanda boyut değiştirme işlemi olarak düşünülebilir. Düzleştirme işleminde yapay sinir ağını hızlıca eğitmek amaçlanmaktadır.

Ağ İçinde Ağ (Network In Network)

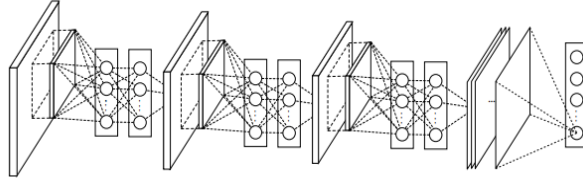
Lin ve arkadaşları (2013) sonradan Google'ın Inception modeline de ilham veren şekil 2'de görülen yeni bir ESA mimarisi ortaya attı. Bu mimari, klasik ESA mimarilerinde konvolüsyon katındaki giriş ve çıkış arasına iki katmanlı bir MLP yapısı eklenerek oluşturulur.



Şekil 2. a) Klasik Konvolüsyon Katmanı b) YSA içeren Konvolüsyon Katmanı (Convolution Layer with MLP) (Liu ve diğ., 2017)

Figure 2. a) Classical Convolution Layer b) Convolution Layer with MLP

Klasik ESA mimarilerinde önerilen bir diğer değişiklik ise konvolüsyon işlemlerinden sonra tamamen birleştirilmiş tabaka kullanmak yerine, softmax sınıflandırma işlemi için gerekli boyut küçültmenin, küresel ortalama havuzlama tabakası kullanılarak gerçekleştirilmesidir (Lin ve diğ., 2013). Doğrusal olmayan bir YSA yapısına sahip olan AİA mimarisi, ESA mimarilerindeki yerel filtrelerde özütleme ve soyutlama üzerine düşük seviyeli doğrusal filtreler geliştirmiştir. Çünkü karmaşık özelliklerin her zaman kolayca doğrusal olarak ayıramadığı belirtilmektedir. Bu çalışmada kullanılan AİA yapısı oluşturulurken, Şekil 3’de verilen ve Lin ve diğ., (2013) tarafından geliştirilmiş olan AİA yapısından esinlenilmiştir. Lin ve diğ., (2013) tarafından önerilen orijinal AİA modelinin CIFAR-10 üzerindeki doğruluk oranı veri çoğaltma olmadan %89.59 iken veri çoğaltma uyguladıklarında %91.19 olarak bulunmuştur.

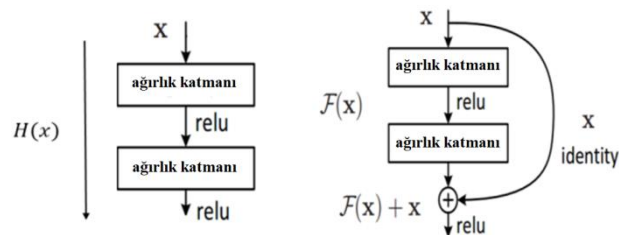


Şekil 3. Lin ve diğ. (2013) tarafından önerilen AİA yapısı

Figure 3. The structure of Network In Network proposed by Lin et al. (2013)

ResNet (Deep Residual Network)

ResNet, ESA mimarisinde küçük değişiklikler ile oluşturulmuştur. Bu değişiklikler, derin ağın hızlı bir şekilde eğitilmesini, büyük bir ağ yerine derin bir ağ oluşturarak parametrelerin sayısının azaltılmasını, yokuş geçiş dereceleri sorununun giderilmesini ve resim sınıflandırma başarısının artmasını sağlamıştır. ResNet modeli ağın derinliğini artırırken, kaybolan gradyanlar ve optimizasyon zorlaşması bu problemlerden çıkan olumsuzlukları engellemektedir. Şekil 4’de artık bağlantılı konvolüsyonun işleyişi gösterilmiştir.



Şekil 4. Klasik Konvolüsyon ve Artık Bağlantılı Konvolüsyon (He ve diğ., 2016)

Figure 4. Classical Convolution and Residual Connected Convolution

ResNet mimarisinde $F(x) = H(x) \cdot x$ düzlemsel olmayan (non-linear) fonksiyonu ile haritalandırma yapmaktadır (He ve diğ., 2016). Giriş değerinden çıkış değerine ek bir kısayol bağlantısı ile giriş değeri olan x , $F(x)$ fonksiyonuna eklenmekte, $F(x)+x$ olarak Relu fonksiyonuna geçirilmektedir (He ve diğ., 2016). 2 katmanın sonunda yeni katmana geçerken giriş değeri olarak bu değer eklenmekte ve bu şekilde ağı eğitmek daha kolay hale gelmektedir. Evrimsel sinir ağını daha düzenli hale getirmek amacıyla kuyruk normalizasyonu kullanılmaktadır. Düzenleyici bir etkiyle birlikte eğitim esnasında sinir ağının yok olma gradyanına bir direnç göstermesini sağlamaktadır. Bu işlem modelin eğitim süresini kısaltarak daha iyi bir performans göstermesine yardımcı olur. Global Ortalama Havuzlama işlemi önceki katmandaki özellik haritasının ortalama çıktısını hesaplayan bir işlemdir. Bu oldukça basit işlem, verileri önemli ölçüde azaltarak modeli son sınıflandırma katmanı için hazırlar.

DERİN ÖĞRENME MODELLERİNİN OLUŞTURULMASI (IMPLEMENTATION OF DEEP LEARNING MODELS)in

Bu bölümde, ESA, AİA ve ResNet yaklaşımlarıyla oluşturulan ve sırasıyla E-Model, A-Model ve R-Model olarak adlandırılan modellerin uygulanmasında kullanılan yöntem ve parametreler açıklanmaktadır. Bu modeller çeşitli açık kaynak kütüphaneler aracılığıyla oluşturulabilmektedir. Bu çalışmada, Python programlama dili üzerinde geliştirilmiş olan tensorflow kütüphanesi ile modeller gerçekleştirilmiştir. Tüm ön işleme, eğitim ve test işlemleri bu kütüphanenin sınıf ve yöntemleriyle gerçekleştirilmiş, sonuç analizleri de tensorboard, sklearn ve matplotlib kütüphaneleri yardımıyla oluşturulmuştur.

E-Model (E-Model)

Çizelge 1’de, E-Modelin parametreleri listelenmektedir. Sınıflandırma katmanındaki ve bir önceki tam bağlı katmandaki düğüm sayısı kullanılan veri kümesindeki sınıf sayısına bağlı olarak değişmektedir. Bu çizelgede CIFAR-10 veri kümesindeki 10 farklı sınıfla uygun olacak şekilde 10 düğümden oluşmaktadır. Diğer veri kümelerindeki sınıf sayılarına göre bu katmanlardaki düğüm sayıları uygun biçimde değiştirilmiştir. Bu tablodaki parametreler ilgili modelin literatürde kabul edilmiş girdileri baz alınarak oluşturulmuştur.

Çizelge 1. CIFAR-10 Veri Kümesi için E-Model

Table 1. E-model for CIFAR10 dataset

Katman	Giriş Boyutu	Filtre Boyutu	Filtre	Dolgu	Kaydırma	Aktivasyon
Konvolüsyon 1	32x32	3x3	32	same	1	Relu
Havuzlama (Maks)-1	32x32	2x2			2	
Konvolüsyon 2	16x16	3x3	64	same	1	Relu
Havuzlama (Maks)-2	16x16	2x2			2	
Konvolüsyon 3	8x8	3x3	64	same	1	Relu
Havuzlama (Maks)-3	8x8	2x2			2	
Düzleştirme	2x2x64 (256’ dan 1’e)					
Tam Bağlı	512 (256’ dan 512’ ye)					
Bırakma	0.1					
Tam Bağlı	256 (512’ ten 256’ ya)					
Bırakma	0.1					
Tam Bağlı	10 (256’ ten 10’ a)					
Sınıflandırma	10					

Çizelge 1’de görüldüğü gibi E-Model art arda gelen 3’er adet konvolüsyon ve maksimum havuzlama katmanı ile bunların ardından gelen düzleştirme katmanı içermektedir. Düzleştirme katmanından sonra sırasıyla 512, 256 ve 10 düğüm içeren 3 adet tam bağlı katman gelmekte, bunların arasında ise 2 adet

birakma katmanı yer almaktadır. Sınıflandırma katmanı ise Softmax fonksiyonunu kullanan 10 adet düğümden oluşmaktadır.

A-Model (A-Model)

Bu çalışmada kullanılan A-Modelinin parametreleri Çizelge 2'de gösterilmiştir. Bu çizelgede de yine sınıflandırma katmanındaki düğüm sayısı CIFAR-10 veri kümesine göre 10 olarak verilmiştir. Diğer veri kümelerindeki sınıf sayılarına göre bu katmanlardaki düğüm sayıları uygun biçimde değiştirilmiştir.

Çizelge 2. CIFAR10 Veri Kümesi için A-Model

Table 2. A-model for CIFAR10 dataset

Katman	Giriş Boyutu	Filtre Boyutu	Filtre Sayısı	Dolgu	Kaydırma	Aktivasyon
Konv1	32x32	5x5	192	same	1	Relu
YSA1-1	32x32	1x1	160	0	1	Relu
YSA1-2	32x32	1x1	96	0	1	Relu
Havuzlama (Maks)-1	32x32	3x3		same	2	
Bırakma-1	0.5, 16x16					
Konv2	16x16	5x5	192	same	1	Relu
YSA2-1	16x16	1x1	160	0	1	Relu
YSA2-2	16x16	1x1	96	0	1	Relu
Ortalama Havuz	16x16	3x3		same	2	
Bırakma-2	0.5, 8x8					
Konv3	8x8	3x3	192	same	1	Relu
YSA3-1	8x8	1x1	192	0	1	Relu
YSA3-2	8x8	1x1	10	0	1	Relu
Ortalama Havuz	8x8	8x8		same	8	
Düzleştirme	1x1x10 to 10					
Softmax	10					

Çizelge 2'de yapısı verilen A-Model'de 3'er adet konvolüsyon ve havuzlama katmanı arasında 2'şer adet YSA yapısı bulunmaktadır. Bu YSA'ların giriş boyutları sırasıyla 32x32, 16x16 ve 8x8 olup, filtre boyutları 1x1 şeklindedir. Konvolüsyon ve YSA katmanlarındaki düğümlerde Relu aktivasyon fonksiyonu, sınıflandırma katmanında ise Softmax aktivasyon fonksiyonu kullanılmıştır. Sınıflandırma katmanından hemen önce Düzleştirme katmanı yer almaktadır.

R-Model (R-Model)

Bu çalışmada kullanılan R-Model'in parametreleri Çizelge 3'de verilmiştir. Bu çizelgede de yine sınıflandırma katmanındaki düğüm sayısı CIFAR-10 veri kümesine uygun olacak şekilde 10 olarak verilmiştir. Diğer veri kümelerindeki sınıf sayılarına göre bu katmandaki düğüm sayıları uygun biçimde değiştirilmiştir. Çizelge 3'te yapısı verilen R-Model'de 2 adet bırakma katmanı arasında 5 adet artık bağlantılı konvolüsyon (ABK) bloğu yer almaktadır. Bu blokların filtre sayısı sırasıyla 16, 32, 32, 64 ve 64 olup, ikinci ve dördüncü bloklarda aşağı örnekleme yapılmaktadır. Aktivasyon fonksiyonu olarak sınıflandırma katmanında Softmax, diğer katmanlarda Relu kullanılmaktadır. Son ABK bloğunu takip eden bırakma katmanından sonra kuyruk normalizasyonu uygulanmakta, daha sonra Global Ortalama Havuzlama katmanı gelmektedir. Sınıflandırma katmanından hemen önce 10 adet düğüm içeren tam bağlı katman yer almaktadır. Diğer veri kümelerindeki sınıf sayılarına bağlı olarak bu katmandaki düğüm sayıları farklıdır.

Çizelge 3. CIFAR10 Veri Kümesi için R-Model

Table 3. R-Model for CIFAR10 dataset

Katman	Giriş Boyutu	Filtre Boyutu	Filtre Sayısı	Aşağı Örnekleme	Aktivasyon
Konv1	32x32	3x3	16		Relu
Bırakma	0.2				
Res-blok1* x5	32x32	3x3	16		Relu
Res-blok2* x1	32x32	3x3	32	Var*	Relu
Res-blok3* x4	16x16	3x3	32		Relu
Res-blok4* x1	16x16	3x3	64	Var*	Relu
Res-blok5* x4	8x8	3x3	64		Relu
Bırakma	0.2				
Kuyruk Normalizasyonu + Aktivasyon(Relu)					
Global Ortalama	8x8x64 'den 64'a				
Tam Bağlı	10 (64 'ten 10'a)				
Softmax	10				

SONUÇLAR VE ANALİZ (RESULTS AND ANALYSIS)

Bu bölümde farklı veri kümeleri üzerinde E-Model, A-Model ve R-Model modellerinin uygulanması sonucu elde edilen karmaşıklık matrisleri, başarı oranları, kayıp oranları ve çalışma süreleri karşılaştırmalı olarak listelenmektedir.

Şekil 5'te verilen karmaşıklık matrisi E-Modelin CIFAR-10 veri kümesinin doğrulama kısmı üzerine uygulanması ile elde edilmiştir. E-Model en iyi otomobil (923) ve kurbağa (924) sınıflarını doğru tahmin etmiştir.

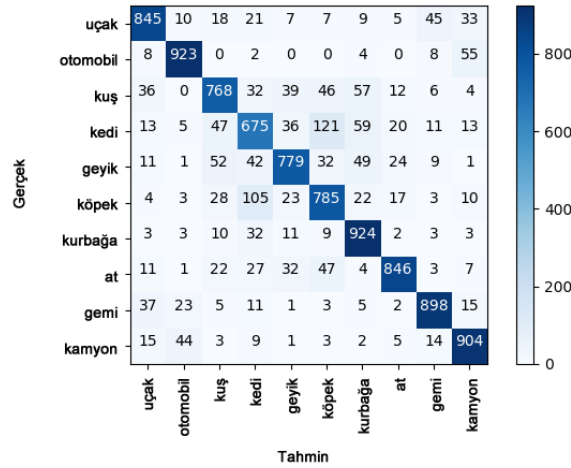
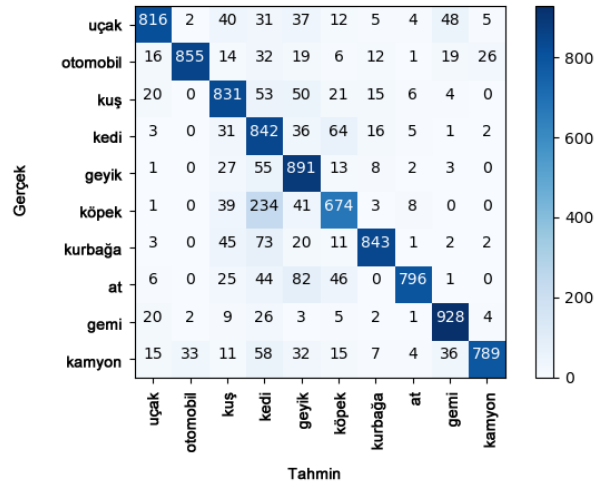
**Şekil 5.** E-Model Karmaşıklık Matrisi

Figure 5. E-Model Confusion Matrix

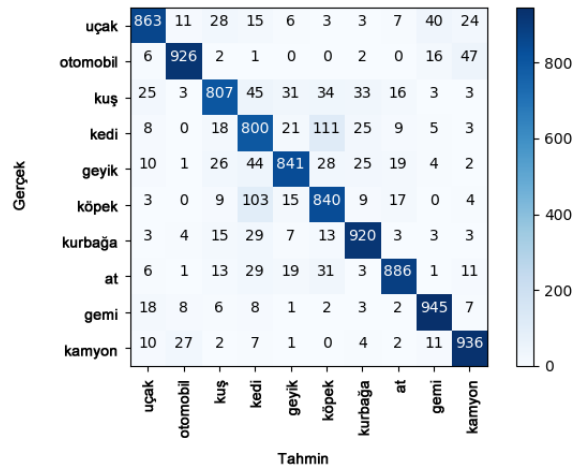
Şekil 5'i incelediğimizde denenen E-Modelin hatalarının genellikle gerçek hayatta birbirine yakın olan sınıflarda olduğunu görebiliriz. Örneğin, kedi sınıfı bir görüntüyü köpek veya kurbağa olarak tahmin etme olasılığı, bir araç olarak tahmin etme olasılığından daha büyüktür.



Şekil 6. R-Model Karmaşıklık Matrisi

Figure 6. R-Model Confusion Matrix

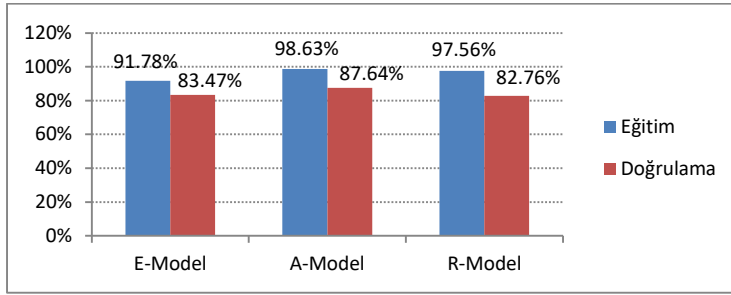
Şekil 6'daki karmaşıklık matrisini incelendiğinde, R-Modelin gemi sınıfında (928) iyi performans gösterdiğini görüyoruz. Aksine, en yanlış sınıflandırılan resimler birbirine yakın oldukları bilinen kedi, köpek ve kurbağa sınıfındaki resimlerdir.



Şekil 7. A-Model Karmaşıklık Matrisi

Figure 7. A-Model Confusion Matrix

Şekil 7'ye göre, denenen A-Model gemi ve kamyon sınıflarında (945 ve 936) yüksek başarı oranı elde etmiştir. A-Modelin en düşük başarı oranı, kedi sınıfı görüntülerini köpekler olarak sınıflandırmaktır.



Şekil 8. Derin Öğrenme Modellerinin Performansları

Figure 8. Performances of The Deep Learning Models

Şekil 8’de derin öğrenme modellerinin sadece MİB içeren bilgisayarda CIFAR-10 veri kümesi üzerinde karşılaştırma grafiği verilmiştir. Bu grafiğe göre, eğitim ve doğrulama veri kümeleri üzerinde en iyi başarımla sırasıyla %98.63 ve %87.64 doğruluk ile A-Model’e aittir. Eğitim veri kümesi üzerindeki en başarılı ikinci model %97.56 doğruluk ile R-Model’dir. E-Model’in eğitim doğruluğu %91.78 ve doğrulama veri kümesi doğruluğu %83.47 olarak bulunmuştur. R-Model’inin doğrulama veri kümesi üzerindeki doğruluğu ise %82.76’dır.

Çizelge 4’te Intel(R) Xeon(R) CPU E5-2630 V4 2.20 GHz 2 işlemcili, 24 GB RAM’e sahip sunucu üzerinde her bir modelin CIFAR-10 eğitim kümesi üzerinde çalışma süreleri ve doğrulama veri seti üzerindeki doğruluk oranları verilmiştir. Modeller ayrıca, Intel(R) Core(TM) i7-9750H CPU 2.60GHz işlemcili, 16 GB RAM ve 1920 Cuda çekirdekli 8 GB paylaşımlı sistem belleğine ve 6 GB ayrılmış video belleğe sahip NVIDIA GeForce RTX 2060 grafik kartlı bir dizüstü bilgisayarda çalıştırılmıştır. Çizelge 4’te bu denemelerle ilgili sonuçlar yer almaktadır. Sadece CPU içeren birinci makine için “Sadece MİB”, CPU ve GPU içeren ikinci makine için ise “MİB+GİB” kısaltmaları kullanılmıştır. GPU kullanımı ile modellerin eğitim süresi dramatik biçimde azalmıştır.

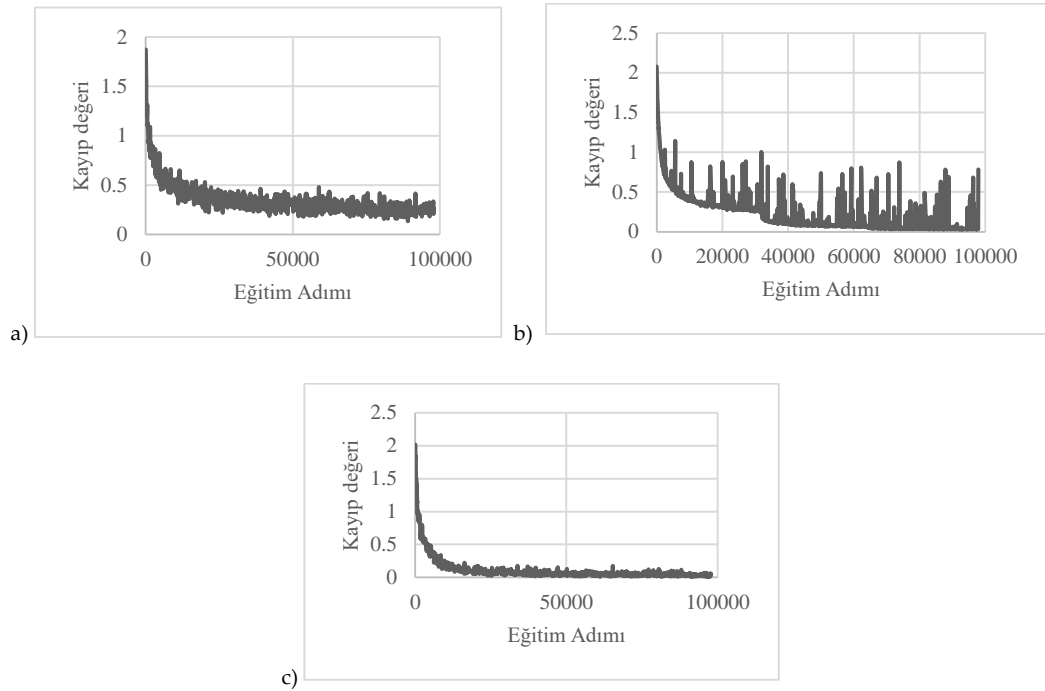
Çizelge 4. Modellerin Karşılaştırılması

Table 4. Comparison of the models

Model	Parametre Sayısı	Sadece MİB		MİB+GİB	
		Çalışma Süresi	Doğruluk Oranı	Çalışma Süresi	Doğruluk Oranı
A-Model	966,986	129 saat	%87.64	2.2 saat	%87.95
R-Model	550,058	415 saat	%84.72	4.45 saat	%87.60
E-Model	715,018	3.5 saat	%83.47	1.8 saat	%82.43

Çizelge 4’e göre modellerin doğrulama veri kümesi üzerindeki doğruluk oranları birbirine çok yakındır. Eğitim ve test doğruluğunu göz önüne alarak, Şekil 8’deki en yüksek doğruluk oranının A-Model’de olduğu görülmektedir. Fakat, bu modelin parametre sayısı 966,986’dır. Eğitim süresi 415 saat ve parametre sayısı 550,058 olan R-Modelin doğruluğunun A-Model’in altında kaldığı görülmektedir. En kısa eğitim süresine sahip E-Model’in R-Model’den biraz daha iyi doğruluk oranına sahip olduğu gözlenmiştir.

Şekil 9, 10 ve 11’deki grafikler sadece MİB içeren makine üzerinde modellerin CIFAR-10 veri kümesi üzerinde çalıştırılması ile elde edilmiştir. Şekil 9’da modellerin eğitim adımı sayılarına karşılık kayıp grafikleri görülmektedir.



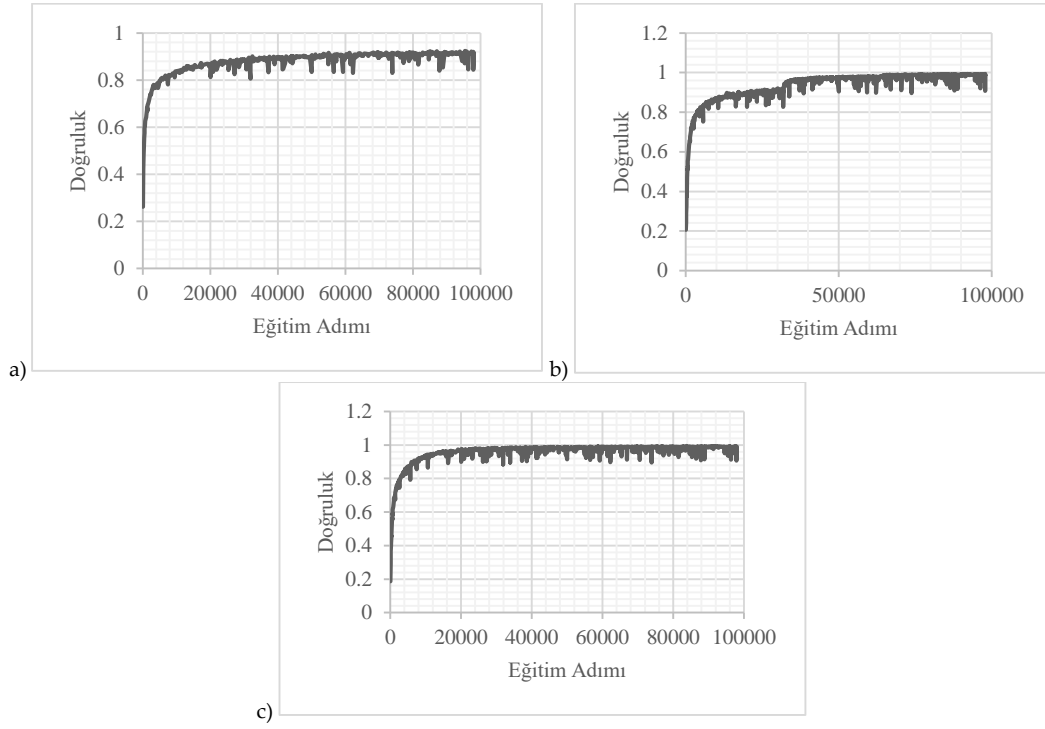
Şekil 9. Modellerin eğitim aşaması kayıp grafikleri a) E-Model b) R-Model c) A-Model

Figure 9. Loss plots of the models during training a) E-Model b) R-Model c) A-Model

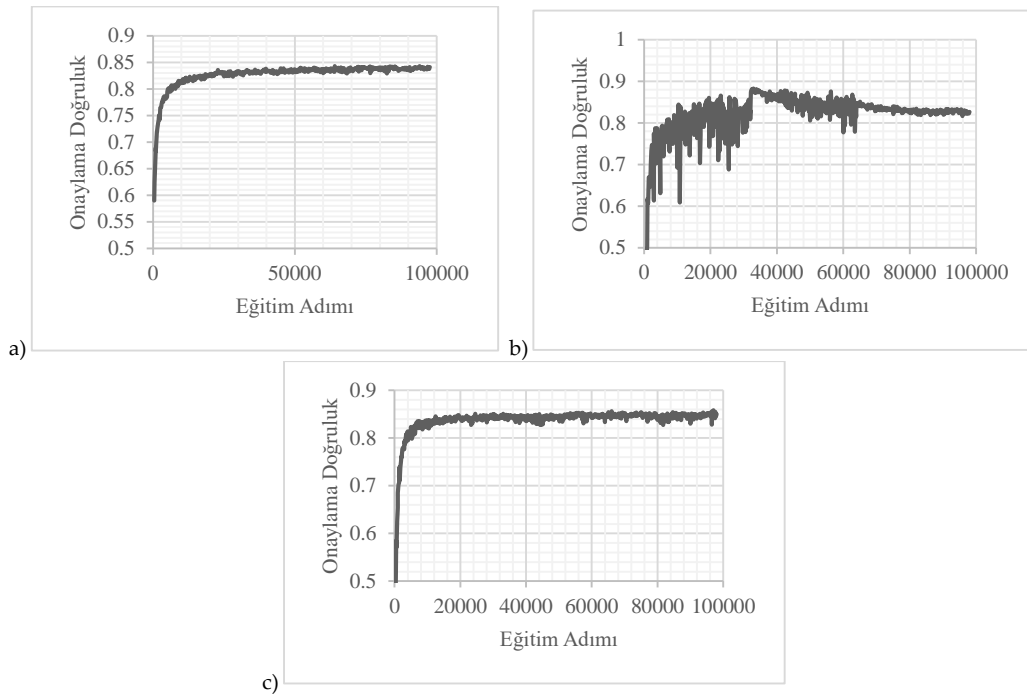
Şekil 9’da eğitim adımı sayısına göre hata oranlarındaki düşüş karşılaştırıldığında, en hızlı düşüş oranına A-Model’in sahip olduğu gözlemlenebilmektedir. Fakat E-Model’in de A-Model’e yakın bir sonuç ürettiği gözlemlenmiştir.

Şekil 10’daki doğruluk grafiklerinde x eksenini eğitim verileri üzerindeki başarı oranını göstermekte ve y eksenini modelin kaç eğitim adımı çalıştığını göstermektedir. Grafik incelendiğinde eğitim adımı sayısıyla doğru orantılı olarak aşırı iniş ve çıkma olmadığı gözlemlenmektedir. Grafikte aşırı iniş ve çıkışın olmaması ve eğitim adım sayısıyla artan bir doğruluk grafiğinin olması modellerin eğitim verilerini başarıyla öğrendiğini göstermektedir.

Şekil 11’de derin öğrenme modellerinin onaylama verileri üzerindeki doğruluk oranları gösterilmektedir. Bu grafiklere göre E-Modelin doğruluğunun %85’in altında kaldığı görülmektedir. R-Model’in yaklaşık 36,000 adımdan sonra düşüş yaşadığı gözlemlenmiştir. A-Modelin doğruluğunun ise %85’in üzerine çıktığı görülmektedir. R-Model’in eğitim doğruluğunun %100’e yakın olmasına rağmen, doğrulama doğruluğunun %82-85 Aralığında olması R-Model’in eğitim verilerini ezberlediğini göstermektedir.



Şekil 10. Modellerin eğitim verileri üzerinde doğruluk grafikleri a) E-Model b) R-Model c) A-Model
Figure 10. Accuracy plots of the models on training data a) E-Model b) R-Model c) A-Model



Şekil 11. Modellerin test verileri üzerinde doğruluk grafikleri a) E-Model b) R-Model c) A-Model
Figure 11. Accuracy plots of the models on validation data a) E-Model b) R-Model c) A-Model

Bu çalışmada kullanılan modellerin, literatürde yer alan bazı modeller ile CIFAR-10 veri kümesi üzerindeki doğruluk oranı ve içerdikleri parametre sayısına göre karşılaştırılması Çizelge 5'te verilmiştir.

Çizelge 5. Literatürdeki bazı modellerle bu çalışmada önerilen modellerin karşılaştırılması*Table 5. Comparison of the proposed models with some of the models in the literature*

Model	Doğruluk Oranı	Parametre Sayısı
MCDNN (Ciregan ve diğ., 2012)	%88.8	1,210,710
Maxout DNN (Goodfellow ve diğ., 2013)	%88.32	1,034,298
SSCNN (Graham, 2014)	%93.7	11,250,000
ACN (Springenberg ve diğ., 2015)	%90.92	1,300,000
ResNet-1001 (He ve diğ., 2016)	%95.4	10,200,000
DenseNet (Huang ve diğ., 2017)	%96.54	25,600,000
UL-Hopfiled (Liu ve Mukhopadhyay, 2018)	%83.1	25,600,000
ACNet (Wang ve diğ., 2019)	%94	19,800,000
NoisyDARTS (Chu ve diğ., 2020)	%98.28	5,500,000
SpinalNet (Kabir ve diğ., 2020)	%91.40	20,160,000
NAT (Lu ve diğ., 2021)	%98.40	6,900,000
A-Model	%87.64	966,986
R-Model	%84.04	550,058
E-Model	%83.47	715,018
(Bu çalışmada kullanılan modeller)		

Çizelge 5'te görüldüğü gibi Liu ve Mukhopadhyay (2018)'ın önerdiği UL-Hopfield modeli hariç, modellerin karmaşıklığı ve dolayısıyla parametre sayısı arttıkça doğruluk oranı artmaktadır. Bu çalışmada önerilen üç modelin de parametre sayısı literatürde kullanılan modellerden daha azdır. Bu durum, modellerin depolama ve bellek gereksinimini azaltan bir durum olacağı için bu modelleri kullanacak uygulamaların etkinliği için avantaj sağlayabilir. Özellikle, 242,826 parametreye sahip ve eğitim süresi 3.5 saat olan E-Model, bu modelden çok daha fazla parametre sayısına sahip MCDNN (1,210,710 parametre) ve Maxout DNN (1,034,298 parametre) ile karşılaştırılabilir doğruluk oranlarına sahiptir. Ayrıca, yaklaşık 25,600,000 parametreye sahip ResNet-50 modelini içeren UL-Hopfiled modelinden daha yüksek doğruluk oranına sahiptir.

Derin öğrenme modellerinin CIFAR-10 veri kümesi üzerinde eğitilmesi için kullanılan parametre değerleri Çizelge 6'da verilmiştir.

Çizelge 6. Modellerin CIFAR-10 veri kümesi üzerinde eğitilmesinde kullanılan parametre değerleri*Table 6. Parameters used for training the models on CIFAR-10 data set*

Model	Epoch sayısı	Batch boyutu	Optimizasyon algoritması	Öğrenme oranı	Öğrenme oranı azalma
A-Model	500	256	Adam	0.0001	0.000001
R-Model	500	256	Adam	0.0001	0.000001
E-Model	500	256	Adam	0.0001	0.000001

Çizelge 6'da gösterildiği gibi tüm modeller için epoch sayısı, batch boyutu aynıdır. Başlangıç ağırlık değerleri gloriot_uniform yöntemi ile oluşturulmuştur.

CIFAR-10 veri kümesi dışında derin öğrenme modelleri farklı SVHN, Parmaklar, Çiçekler, Ürün Kalitesi ve Balıklar veri kümeleri üzerinde MİB ve GİB içeren makine kullanılarak karşılaştırılmıştır. Çizelge 7'de bu veri kümeleri için modellerin karşılaştırmasına dair sonuçlar verilmiştir. Her bir veri kümesi için modellerin tümünde eğitim aşamasında en düşük doğrulama kaybına sahip model kaydedilmiştir. Modellerin doğruluk veri seti üzerinde değerlendirilmesi eğitim aşamasında kaydedilen en iyi modeller kullanılarak yapılmıştır. Doğruluk değeri, doğrulama veri kümesi üzerinde elde edilen doğruluktur.

Çizelge 7. Modellerin diğer veri kümeleri üzerinde karşılaştırılması

Table 7. Comparison of the models on the other data sets

Veri kümesi	Resim Boyutu	Modeller	Eğitim Süresi	Doğruluk (Doğrulama)	Epoch sayısı	Batch boyutu	Optimizasyon algoritması	Öğrenme oranı	Öğrenme oranı azalma
SVHN	32x32	A-Model	37 dk	%94.88	100	256	Adam	0.001	0.000001
		R-Model	67 dk	%94.79	100	256	Adam	0.0001	
		E-Model	28 dk	%90.00	100	256	Adam	0.001	
Parmaklar	32x32	A-Model	3 dk	%94.78	100	32	Adam	0.001	0.000001
		R-Model	3 dk 20 sn	%95.21	100	32	Adam	0.001	0.0001
		E-Model	3 dk 30 sn	%90.12	100	32	Adam	0.001	0.000001
Çiçekler	32x32	A-Model	33 dk	%86.18	500	64	Adam	0.001	0.000001
		R-Model	38 dk	%89.29	500	64	Adam	0.001	0.000001
		E-Model	30 dk	%87.05	500	64	Adam	0.001	0.000001
Ürün Kalitesi	128x128	A-Model	3 dk 10 sn	%95	200	16	Adam	0.001	0.000001
		R-Model	3 dk 40 sn	%98.75	200	16	SGD	0.01	0.000001
		E-Model	3 dk	%98.12	200	16	Adam	0.001	0.000001
Balıklar	128x128	A-Model	142 dk	%97.09	500	64	Adam	0.0001	0.000001
		R-Model	165 dk	%96.07	500	64	Adam	0.0001	0.000001
		E-Model	137 dk	%96.76	500	64	Adam	0.0001	0.000001

Çizelge 7'de belirtildiği gibi modellerin eğitilmesinde genel olarak optimizasyon algoritması olarak Adam kullanılmıştır. Sadece Ürün Kalitesi veri kümesinde R-Model'in eğitilmesinde SGD (Stochastic Gradient Descent) algoritmasının Adam'dan daha iyi sonuç verdiği görülmüştür. Sadece Çiçekler veri kümesi için her üç modelin doğruluğu %90 altında kalmıştır. Diğer veri kümelerinde ise doğruluk tüm modeller için %90 ve üzerindedir. Balıklar veri kümesi (Ulucan ve diğ., 2020) üzerindeki doğruluk A-Model, R-Model ve E-Model için sırasıyla %97.09, %96.07 ve %96.76 olarak bulunmuştur. Bu veri kümesi üzerinde derin öğrenme modelleri kullanılarak sınıflandırmanın yapıldığı herhangi bir çalışma yoktur. Ulucan ve diğ., (2020) 6 farklı özellik çıkarım yöntemi kullanarak elde ettikleri özellikleri Destek Vektör Makinelerine (DVM) uygulayarak sınıflandırma yapmışlardır. Doğrulama veri kümesi için en iyi ortalama doğruluk değerlerini kontrast ve enerji özellikleri için sırasıyla %97.64 ve %97.47 olarak bulmuşlardır. En kötü doğruluk değerleri ise BoF (Bag of Features) yöntemi ile bulunan özellik ve gri seviye eş oluşum matrisinden yararlanarak buldukları ikinci ve üçüncü momentin toplamı olan özellik ile elde edilmiştir. Bu değerler sırasıyla %81.55 ve %89.57 olarak bulunmuştur.

SONUÇ ve TARTIŞMALAR (RESULTS and DISCUSSIONS)

Bu çalışmada, AiA, ResNet ve ESA modellerine ait yaklaşımlar kullanılarak A-Model, R-Model ve E-Model olarak adlandırılan özgün derin öğrenme modeller oluşturulmuş ve literatürde yaygın olarak kullanılan veri kümeleri yanında, daha önce derin öğrenme yöntemleri uygulanmamış Balıklar (Ulucan

ve diğ., 2020) veri kümesi üzerinde performansları karşılaştırılmıştır. CIFAR-10 veri kümesi için eğitim sürelerini ve performanslarını karşılaştırma amacıyla modeller MİB ve MİB+GİB içeren bilgisayarlarda ayrı ayrı çalıştırılmıştır. Performanslar arasında çok az fark olduğu, fakat MİB+GİB içeren bilgisayarda eğitim süresinin dramatik biçimde azaldığı tespit edilmiştir. Diğer veri kümelerinin tümü için modeller MİB+GİB içeren bilgisayarda eğitilmiş ve performansları karşılaştırılmıştır. Modellerin eğitim aşamasında en düşük doğrulama kaybına sahip model kaydedilerek ilerlenmiştir. Doğruluk veri seti üzerinde modellerin değerlendirilmesi eğitim aşaması sonunda elde edilen en iyi modeller kullanılarak yapılmıştır.

Ulucan ve diğ., (2020) kendilerinin elde ettiği balık görüntülerini sınıflandırmak için literatürde yaygın olarak kullanılan resim özellik çıkarım yöntemleri ile elde ettikleri özellikleri DVM'ye uygulamışlardır. Dolayısıyla eğitim ve sınıflandırma işlemlerinden önce yoğun işlem gerektiren özellik çıkarımı adımı bulunmaktadır. Öte yandan, derin öğrenme yöntemleri özellik çıkarım adımı gerektirmeden eğitim ve sınıflandırma yapabilmektedir. Bu çalışmada oluşturulan A-Model, R-Model ve E-Model'in Balıklar veri kümesi üzerindeki doğruluğu, Ulucan ve diğ. (2020)'nın en iyi sonucu veren kontrast özelliğiyle elde ettikleri doğruluk oranına çok yakın bulunmuştur. Her modelin bir veri kümesindeki görüntü sınıfları üzerindeki doğruluk değerleri birbirinden farklı ve diğerlerine göre daha iyi olabildiği için, sonraki çalışmalarda genel sınıflandırma performansını artırmak üzere modellerin tahmin sonuçlarının birleştirileceği hibrid yöntemlerin kullanılabilmesi düşünülmektedir.

KAYNAKLAR (REFERENCES)

- Chu, X., Zhang, B., & Li, X., 2020, Noisy differentiable architecture search, arXiv:2005.03566.
- Ciregan, D., Meier, U., & Schmidhuber, J., 2012, Multi-column Deep Neural Networks for Image Classification, *2012 IEEE Conference on Computer Vision and Pattern Recognition*, Providence, RI, USA, 2012, pp. 3642-3649.
- Dabhi, R., 2020, (2020, 03.06.2021). *Casting product image data for quality inspection*. Available: <https://www.kaggle.com/ravirajsinh45/real-life-industrial-dataset-of-casting-product>.
- Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., & Fei-Fei, L., 2009, Imagenet: A large-scale hierarchical image database, in *Computer Vision and Pattern Recognition, 2009, CVPR 2009, IEEE Conference on. IEEE, 2009*, pp. 248-255.
- Ganin, Y., Kononenko, D., Sungatullina, D., & Lempitsky V., 2016, Deepwarp: Photorealistic image resynthesis for gaze manipulation, *European Conference on Computer Vision*. Springer, Cham, pp. 311-326.
- Goodfellow, I., Warde-Farley, D., Mirza, M., Courville, A., & Bengio, Y., 2013, *Proceedings of the 30th International Conference on Machine Learning*, PMLR, 28(3):1319-1327.
- Graham, B., 2014, Spatially-sparse convolutional neural networks, arXiv:1409.6070.
- He, K., Zhang, X., Ren, S. & Sun, J., 2016, Deep residual learning for image recognition, in *Proceedings of the IEEE conference on computer vision and pattern recognition, 2016*, pp. 770-778.
- Huang, G., Liu, Z., Van Der Maaten, L. & Weinberger, K. Q., 2017, Densely Connected Convolutional Networks, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, pp. 2261-2269.
- İnik, Ö. & Ülker, E., 2017, Derin Öğrenme ve Görüntü Analizinde Kullanılan Derin Öğrenme Modelleri, *Gaziosmanpaşa Bilimsel Araştırma Dergisi*, 6(3): 85-104.
- Kabir, H.M.D.; Abdar, M.; Jalali, S.M.J.; Khosravi, A.; Atiya, A.F.; Nahavandi, S.; Srinivasan, D., 2020, SpinalNet: Deep neural net workwith gradual input, arXiv:2007.03347.
- Kızrak, M.A., Bolat, B., 2018, Derin Öğrenme ile Kalabalık Analizi Üzerine Detaylı Bir Araştırma, *Bilişim Teknolojileri Dergisi*, (11) 3: 263-286.
- Koryakin, P., 2019, (2019, 03.06.2021). *Fingers*. Available: <https://www.kaggle.com/koryakinp/fingers>.
- Krizhevsky, A., 2009, Learning multiple layers of features from tiny images, Master's thesis, Department of Computer Science, University of Toronto.

- Krizhevsky, A., Sutskever, I. & Hinton, G. E., 2012, ImageNet Classification with Deep Convolutional Neural Networks, NIPS'12: Proceedings of the 25th International Conference on Neural Information Processing Systems, Lake Tahoe Nevada, USA, 2012, pp. 1097-1115.
- Lin, M., Chen, Q., & Yan, S., 2013, Network in network, arXiv:1312.4400.
- Liu, Q & Mukhopadhyay, S., 2018, Unsupervised Learning using Pretrained CNN and Associative Memory Bank, *International Joint Conference on Neural Networks (IJCNN 2018)*, 8-13 Jul 2018, Rio, Brazil, pp. 1-8.
- Liu, Q., Zhang, N., Yang, W., Wang, S., Cui, Z., Chen, X., & Chen, L., 2017, A Review of Image Recognition with Deep Convolutional Neural Network, In: Huang DS., Bevilacqua V., Premaratne P., Gupta P. (eds) *Intelligent Computing Theories and Application. ICIC 2017. Lecture Notes in Computer Science*, Vol 10361. Springer, Cham.
- Lu, Z., Sreekumar, G., Goodman, E., Banzhaf, W., Deb, K. & Boddeti, V. N., 2021, Neural Architecture Transfer, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (43): 2971-2989.
- McCulloch, W.S. & Pitts, W., 1943, A logical calculus of the ideas immanent in nervous activity, *The Bulletin of Mathematical Biophysics*, 5(4): 115-133.
- Netzer, Y., Wang, T., Coates, A., Bissacco, A., Wu, Bo. & Ng, A.Y., 2011, Reading Digits in Natural Images with Unsupervised Feature Learning, *NIPS Workshop on Deep Learning and Unsupervised Feature Learning*.
- Nilsback, M. & Zisserman, A., 2006, A Visual Vocabulary for Flower Classification, *IEEE Conference on Computer Vision and Pattern Recognition*, Vol. 2, pp.1447-1454.
- Rumelhart, D.E., McClelland, J.L. & Group T.P.R., 1986, *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, MIT Press, Cambridge.
- Springenberg, J.T., Dosovitskiy, A., Brox, T. & Riedmiller, M., 2015, Striving for simplicity: The All Convolutional Net, arXiv:1412.6806.
- Ulucan, O., Karakaya, D. & Turkan, M., 2020, A Large-Scale Dataset for Fish Segmentation and Classification, *2020 Innovations in Intelligent Systems and Applications Conference (ASYU)*, Istanbul, Turkey, 2020, pp. 1-5.
- Wang, G., Wang, K., & Lin, L., 2019, Adaptively Connected Neural Networks, *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, CA, USA, 2019, pp. 1781-1790.