

COVID-19 Detection with Deep Learning Methods Under Cross-Dataset Evaluation

Yahya DOĞAN^{1*} 

¹Siirt University, Faculty of Engineering, Department of Computer Engineering, Siirt, Turkey

Article Info

Research article
Received: 26/11/2022
Revision: 28/04/2023
Accepted: 17/05/2023

Keywords

Deep Learning
COVID-19
Convolutional Neural
Networks
Grad-CAM
Transfer Learning

Makale Bilgisi

Araştırma makalesi
Başvuru: 26/11/2022
Düzeltilme: 28/04/2023
Kabul: 17/05/2023

Anahtar Kelimeler

Derin Öğrenme
COVID-19
Evrimsel Sinir Ağları
Grad-CAM
Öğrenme Aktarımı

Graphical/Tabular Abstract (Grafik Özet)

Recently, deep learning algorithms have been utilized as an effective method for the rapid detection of COVID-19. However, the limited geographical coverage of the datasets used in these studies raises the possibility that the model contains racial biases and cannot demonstrate the same performance across multiple datasets. In this study, the performance of transfer learning models was examined through cross-data evaluation. / Son zamanlarda, derin öğrenme algoritmaları, COVID-19'un hızlı tespiti için etkili bir yöntem olarak kullanılmıştır. Ancak, çalışmalarda kullanılan veri setlerinin sınırlı coğrafi kapsamı, modelin etnik bağımlılığa sahip olabileceği ve farklı veri setlerinde aynı performansı gösterememe riskini artırabilir. Bu çalışmada, çapraz veri değerlendirilmesiyle öğrenme aktarma modellerinin performansı incelenmiştir.

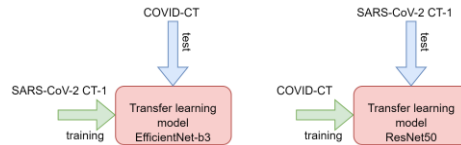


Figure A: Model comparison under cross-dataset setting / **Şekil A:** Çapraz veri seti ayarında model karşılaştırması

- The performances of state-of-the-art deep learning models were compared using popular COVID-19 datasets. / Yaygın COVID-19 veri setleri kullanılarak güncel derin öğrenme modellerinin performansları karşılaştırıldı.
- Cross-dataset evaluation was performed using the two most common datasets collected for COVID-19 detection. / COVID-19 tespiti için toplanan en yaygın iki veri seti kullanarak çapraz veri seti değerlendirilmesi yapılmıştır.
- Grad-CAM and the top-performing ResNet50 model were used to detect the focused areas in CT images. / Grad-CAM ve en iyi performans gösteren ResNet50 modeli, BT görüntülerinde odaklanılan alanları tespit etmek için kullanıldı.

Amaç (Aim): The aim of this study is to investigate whether the datasets collected within the scope of COVID-19 exhibit ethnic biases. / Bu çalışmanın amacı, COVID-19 kapsamında toplanan veri setlerinin etnik bağımlılık taşıyıp taşımadığını araştırmaktır.

Özgünlük (Originality): This study demonstrates the limited success of COVID-19 detection under cross-dataset evaluation and reveals that even with fewer training samples, utilizing a general dataset yields higher scores. / Bu çalışma, çapraz veri seti değerlendirmesi altında COVID-19 tespitindeki sınırlı başarıları ve daha az örnekle eğitim yapıldığında bile genel bir veri setinin daha yüksek skorlar sağladığını göstermektedir.

Bulgular (Results): Experimental results indicate that the successes achieved in COVID-19 datasets are not generalizable and exhibit ethnic dependency. / Deneysel sonuçlar, COVID-19 veri setlerindeki başarıların genellenebilir olmadığını ve etnik bağımlılık taşıdığını göstermektedir.

Sonuç (Conclusion): COVIT-CT, when used as the training set, produces low training performance but high test scores. This can be attributed to the broader and more diverse nature of the COVIT-CT dataset. The SARS-CoV-2 CT dataset, acquired from hospitals in the same state, demonstrates high performance on the training set but poor performance on the test set. This indicates a weak generalization ability of the model trained on this dataset. In the conducted experimental investigations, the ResNet50 model achieved the highest score (71.47%) in cross-dataset evaluation. / COVIT-CT eğitim seti olarak kullanıldığında, eğitim performansı düşük olmasına rağmen test skorları yüksektir. Bunun nedeni, COVIT-CT veri setinin daha genel ve çeşitli olmasıdır. SARS-CoV-2 CT veri seti ise aynı eyaletteki hastanelerden toplandığı için eğitim seti başarısı yüksek olmasına rağmen test başarısı düşüktür. Bu durum, bu veri setiyle eğitilen modelin genelleme yeteneğinin zayıf olduğunu göstermektedir. Yapılan deneysel çalışmalarda, çapraz veri seti değerlendirmesi altında en yüksek skor (%71.47) ResNet50 modeliyle elde edilmiştir.



Derin Öğrenme Yöntemleriyle Çapraz Veri Seti Değerlendirmesi Altında COVID-19 Tespiti

Yahya DOĞAN^{1*}

¹Siirt University, Faculty of Engineering, Department of Computer Engineering, Siirt, Turkey

Makale Bilgisi

Araştırma makalesi
Başvuru: 26/11/2022
Düzeltilme: 28/04/2023
Kabul: 17/05/2023

Anahtar Kelimeler

Derin Öğrenme
COVID-19
Evrimsel Sinir Ağları
Grad-CAM
Öğrenme Aktarımı

Öz

COVID-19 salgını tüm dünyayı etkilemiş ve son yüz yılın en şiddetli rahatsızlıklarından biri haline gelmiştir. Yüksek bulaşıcılığı nedeniyle, COVID-19'un erken aşamada tespiti ve enfekte olan hastaların diğerlerinden izole edilmesi pandemiyi kontrol etmede en önemli aşamalardan biridir. Revers-Transkriptaz Polimeraz Zincir Reaksiyonu (RT-PCR) enfekte olan hastaları teşhis etmek için kullanılan en yaygın yöntemdir; ancak, bu yöntem zaman alıcı ve yoğun emek gerektirmektedir. Yakın zamanda, COVID-19'un hızlı tespiti için bilgisayarlı tomografi görüntüleri kullanılarak yapılan derin öğrenme tabanlı birçok çalışma mevcuttur. Bu çalışmalarda, enfekte olan hastalar hızlı ve yüksek doğruluk oranları (>%97) ile tespit edilebilmektedir. Fakat bu kapsamda hazırlanan veri setleri incelendiğinde, verilerin genellikle aynı ülke veya aynı ildeki hastanelerden elde edildiği gözlenmektedir. Bu durum, kurulan modelin etnik bağımlılığa sahip olmasına ve farklı veri setlerinde aynı performansı göstermemesine neden olabilir. Bu çalışmada, çapraz veri değerlendirme (eğitim ve test için farklı veri setleri) altında güncel derin öğrenme modellerinin performansı deneysel olarak incelenmiştir. İncelenen modeller arasında en yüksek tespit skoru %71.47 ile ResNet50 modeli kullanılarak elde edilmiştir. Ayrıca, eğitilen ResNet50 modelinin görüntünün sınıfına (Covid/Non-Covid) karar verirken odaklandığı alanları göstermek için Grad-CAM sonuçları sağlanmıştır.

COVID-19 Detection with Deep Learning Methods Under Cross-Dataset Evaluation

Article Info

Research article
Received: 26/11/2022
Revision: 28/04/2023
Accepted: 17/05/2023

Keywords

Deep Learning
COVID-19
Convolutional Neural Networks
Grad-CAM
Transfer Learning

Abstract

The COVID-19 epidemic has affected the whole world and has become one of the most severe diseases of the last century. Due to its high prevalence, early detection of COVID-19 and isolating infected patients from others is one of the most important points in controlling the pandemic. The reverse polymerase chain reaction (RT-PCR) is the most common method used to diagnose infected patients; however, this method is time-consuming and labor-intensive. Recently, there are many deep learning-based studies using computer tomography images for rapid detection of COVID-19. In these studies, infected patients can be detected rapidly and with high accuracy i.e. > 97%. When the datasets collected in this context are examined, it is observed that the data are generally obtained from hospitals in the same country or in the same province. This may cause the established model to have an ethnic dependency and not perform the same performance in different datasets. In this study, the performance of state-of-the-art deep learning models under cross-data evaluation, i.e., different datasets for training and testing, is experimentally investigated. The highest detection score was obtained by using the ResNet50 model with 71.47% among the models examined. Moreover, Grad-CAM results are provided to show the areas that the trained ResNet50 model focuses on when deciding on the class of the image (Covid/Non-Covid).

1. GİRİŞ (INTRODUCTION)

Koronavirüs hastalığı (COVID-19) ilk olarak Aralık 2019'da Çin'in Wuhan kentinde bildirildi ve dünya çapında hızla yayılarak her gün binlerce ölüme sebep oldu. Bu virüsün yüksek enfeksiyon oranları

nedeniyle, enfekte kişiyi izole etmek için erken teşhis önemlidir; sonradan tanıtılan virüs varyantları daha yüksek enfeksiyon oranlarına sahiptir. Günümüzde Covid-19'u tespit etmek için kabul edilen standart ve en yaygın kullanılan yöntem Ters

Transfer Polimeraz Zincir Reaksiyonudur (Reverse Transfer Polymerase Chain Reaction-RT-PCR) [1]. Bu yöntem diğer yöntemlere göre daha yüksek başarı oranı ile tanı sağlar; ancak RT-PCR yöntemi zaman alıcıdır ve hastalığın erken evrelerinde duyarlılığı düşüktür [2]. Son bulgular, COVID-19 ile enfekte hastaların, yapay zeka yaklaşımları kullanılarak göğüs röntgeni [3-5] veya bilgisayarlı tomografi (BT) görüntüleri [6-8] ile hızlı bir şekilde tespit edilebileceğini göstermektedir. Hem BT hem de X ışınları, iç vücut yapılarının görüntülerini sağlar. Bu yöntemler karşılaştırıldığında, geleneksel X-ışınlarında yapılar üst üste geldiğinden BT, X-ışınlarından daha fazla bilgi sağlar; kaburgalar göğsü örterek tanıyı kısmen zorlaştırır. Bu bağlamda birçok araştırmacı, BT görüntülerini kullanarak COVID-19 teşhisi için yapay zekâ teknikleriyle bir karar destek aracı sağlamayı hedeflemektedir. Yapay zekâ, özellikle yapay zekânın bir alt dalı olan derin öğrenmedeki hızlı ilerlemeler, organ segmentasyonu [9], görüntü iyileştirme veya onarımı [10], hastalık tespiti veya sınıflandırması [11] gibi tıbbi alanlarda ve farklı alanlardaki birçok çalışmada [12-18] önemli katkılar sağladığı gözlenmektedir.

BT göğüs görüntülerinden oluşan birkaç açık çevrimiçi veri seti mevcuttur. Bildiğimiz kadarıyla, teşhis için en yaygın kullanılanları COVID-CT [19] ve SARS-CoV-2 CT-1 tarama [20] veri setleridir. Şekil 1'de, COVID-CT veri setinden alınan bazı örnek görüntüler verilmiştir. Literatürde, COVID-19 hastalığını tespit eden çalışmaların başarısı, veri setini bölme yöntemine bağlı olarak önemli ölçüde değişmektedir. Genel olarak araştırmacılar 2 farklı bölme yöntemi kullanmaktadır; rastgele (random) ve kişiden bağımsız. Rastgele yöntemde, veri seti rastgele olarak eğitim, doğrulama ve test olarak ayrılmaktadır; diğer bir deyişle, aynı kişiye ait görüntüler 3 parçanın tamamında olabilir. Bu yöntemde, test veri setindeki hastaların farklı görüntüleri daha önce modele girdi olarak verildiğinden, ilgili hastalara yönelik bir eğilim (bias) oluşmaktadır. Genel olarak, rastgele bölme işlemi yapılığında, modellerin başarı oranları yüksek olmakta (>90), ancak genelleme yeteneği düşük kalabilmektedir; diğer bir deyişle, model eğitiminde görüntüleri olmayan hastaların teşhisi zor olabilmektedir. Kişiden bağımsız yöntemde ise veri seti hasta bazlı bölünmektedir. Bu bölme yönteminde, eğitim, validasyon ve test veri setinde farklı hastalara ait görüntüler kullanılmamaktadır. Bu yöntemde, modelin eğitim işlemi daha zor olabilmekte ve rastgele yöntemle göre performansı düşük kalabilmektedir. Fakat rastgele bölme yöntemi ile kıyaslandığında, hastalara karşı bir

eğilim oluşmadığı için modelin genelleme yeteneği daha iyi olmaktadır.

Belirtilen bu bölme yöntemlerinde aynı veri seti kullanılmaktadır. COVID-19 kapsamında hazırlanan veri setleri incelendiğinde genellikle aynı ülke veya aynı ildeki hastanelerden elde edildiği görülmektedir. Bu, elde edilen modelin etnik bağımlılığa sahip olmasına ve farklı veri setlerinde aynı performansı göstermemesine sebep olabilir. Bu çalışmada temel amaç, çapraz veri seti değerlendirmesi altında COVID-19 tespiti için güncel transfer öğrenme modellerinin performansını incelemektir.

Bu çalışmanın temel katkıları şu şekilde özetlenebilir:

- Derin öğrenme alanında ImageNet [21] gibi büyük veri setlerinde önemli başarılar sağlayan modellerin COVID-19 veri setindeki performansları deneysel olarak karşılaştırılmıştır.
- COVID-19 tespiti için toplanan en yaygın iki veri seti kullanılarak çapraz veri seti değerlendirilmesi yapılmıştır.
- En yüksek performans gösteren derin öğrenme modeli kullanılarak, COVID-19 tespitinde, BT görüntülerinde hangi noktalara odaklanıldığı incelenmiştir.

Makalenin geri kalanı aşağıdaki gibi yapılandırılmıştır. Bölüm 2'de ilgili alandaki önceki çalışmalar kısa ve öz bir şekilde ele alınmıştır. Bölüm 3'te çalışma kapsamında kullanılan veri setleri ve modeller hakkında kapsamlı bir araştırma yapılmıştır. Bölüm 4'te deneysel çalışmalar sonucu elde edilen bulgular ve değerlendirmeler sunulmuştur. Bölüm 5'te ise sonuçlar kısaca özetlenmiş ve gelecekte yapılması planlanan çalışmalar belirtilmiştir.

2. MATERYAL VE METOD (MATERIALS AND METHODS)

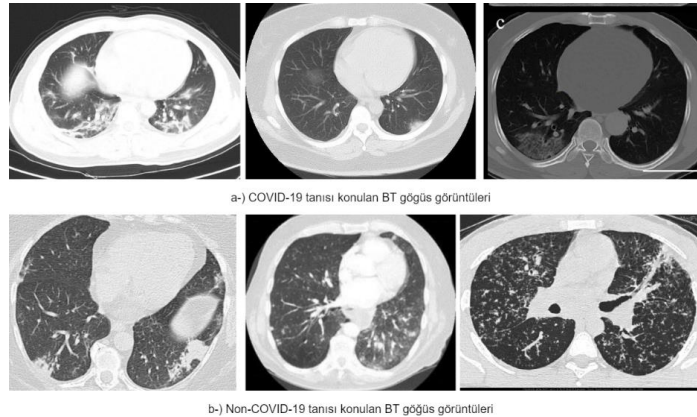
2.1. Veri Setleri (Datasets)

Bu çalışmada, COVID-19 teşhisi için hazırlanan iki yaygın veri seti kullanılmıştır. COVID-19 için önerilen yöntemler incelendiğinde genellikle aynı veri setinden görüntü setleri ile eğitilmekte ve test edilmektedir. Bu çalışmada, temel hedef derin öğrenme yöntemlerini kullanarak bireylerden ve etnik bağımlılıktan bağımsız olarak COVID-19 tespiti yapmaktır. Bu kapsamda iki aşamalı yol izlenmiştir. İlk aşamada, en yaygın veri setlerinden biri olan SARS-CoV-2 CT-scan veri seti eğitim ve farklı makalelerden toplanan görüntülerden oluşan

COVID-CT test olarak kullanılarak güncel derin öğrenme modellerinin performansı deneysel olarak karşılaştırılmıştır. İkinci aşamada ise eğitim ve test veri setlerinin yeri değiştirilip model performansları incelenmiştir.

COVID-CT [19]: Bu veri seti COVID-19 klinik bulgularını içeren 216 hastadan toplanan 349 BT görüntüsünden ve farklı hastalık türleri içeren 397 negatif BT görüntüsünden oluşmaktadır. İlgili veri setindeki görüntüler medRxiv, bioRxiv, NEJM, JAMA, Lancet vb. kaynaklı COVID-19 ile ilgili bilimsel makalelerden toplanmıştır. COVID-19 anormallikleri içeren BT'ler, makalelerdeki şekil başlıkları okunarak seçilmiştir.

SARS-CoV-2 CT-scan [20]: Bu veri seti SARS-CoV-2 enfeksiyonu pozitif olan 1252 BT taraması ve SARS-CoV-2 ile enfekte olmayan hastalar için 1230 BT taraması olmak üzere toplamda 2482 BT taramasından oluşmaktadır. Enfekte olanların 32'si erkek ve 28'i kadın olmak üzere 60 hasta, enfekte olmayanların 30'u erkek ve 30'u kadın olmak üzere 60 hasta verisi mevcuttur. Veri setindeki görüntülerin çözünürlükleri farklıdır ve kontrast farklılıkları vardır. Bu veri seti, tek bir eyaletteki, yani Brezilya'daki Sao Paulo hastanelerinden toplanmıştır.



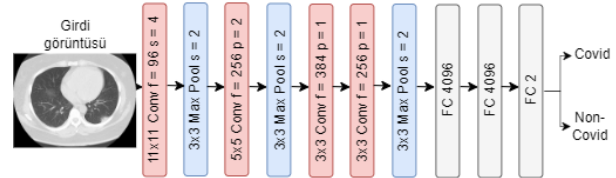
Şekil 1. SARS-CoV-2 CT tarama veri kümesinden COVID-19 (üstte) ve Non-COVID-19 (altta) için BT örnek görüntüleri (CT images of COVID-19 (top) and Non-COVID-19 (bottom) from the SARS-CoV-2 CT scan dataset)

2.2. Mimariler (Architectures)

Bu kısımda çalışmada kullanılan derin öğrenme modelleri ele alınmıştır

AlexNet [22]: Krizhevsky ve diğerleri tarafından önerilen AlexNet mimarisi, 2012 ImageNet yarışmasını top-5 kategorisinde %15,3'lik hata oranıyla kazanmıştır. Bu mimari, öğrenme yoluyla otomatik olarak elde edilen öznelikleri kullanmanın, manuel olarak elde edilen öznelikleri kullanmaktan daha iyi performans sağladığını göstermiştir. Şekil 2'de COVID-19 problemine uyarlanan AlexNet mimarisi verilmiştir. Mimari, beş evrişim (convolution) katmanından ve üç tam bağlantılı (fully connected) katmandan oluşmaktadır. AlexNet mimarisinde CNN'lerde günümüzde de kullanılan bazı yaklaşımlar kullanılmıştır; ReLU aktivasyon fonksiyonu, çoklu GPU yaklaşımı ve örtüşen havuzlama (pooling) tekniği. Model yaklaşık 60 milyon öğrenebilir parametre barındırmaktadır. Kapasitesi yüksek olduğundan aşırı öğrenme (overfitting) problemi meydana gelebilmektedir. Bunun üstesinden

gelmek için veri artırma ve dropout teknikleri kullanılmıştır.

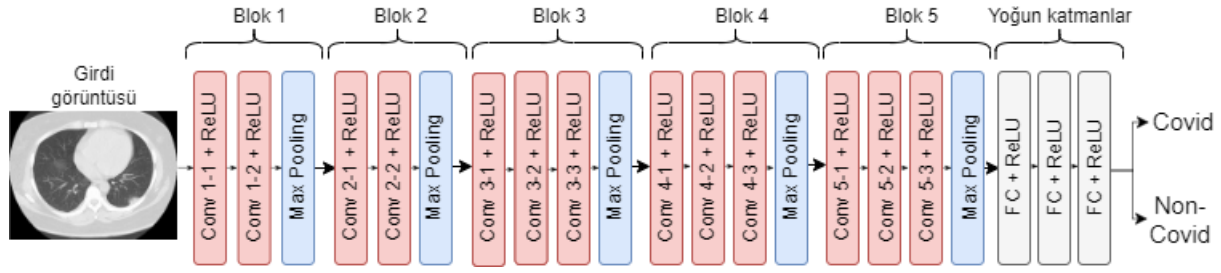


Şekil 2. AlexNet mimarisi. s stride, p padding FC ise tam bağlantılı katmanları temsil etmektedir (The AlexNet architecture. s represents stride, p represents padding, and FC represents fully connected layers)

VGG16 [23]: Simonyan ve Zisserman tarafından önerilen bu mimari 2014 yılında ImageNet veri setinde top-5 kategorisinde %92.7 test doğruluğu ile ILSVR-2014 yarışmasında en yüksek skoru almıştır. VGG16 mimarisinde AlexNet mimarisinde kullanılan yüksek çekirdek (kernel) boyutları azaltılmıştır. 3x3 filtre kullanan evrişim bloklarına ve 2x2 max havuzlama katmanlarından oluşan bir mimariye sahiptir. Tüm mimari boyunca tutarlı bir şekilde evrişim ve maksimum havuzlama

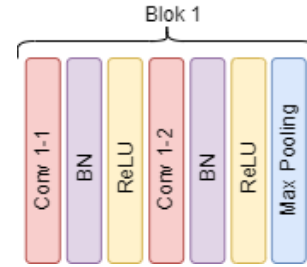
katmanları kullanılmaktadır. Sınıflandırma bloğunda ise softmax katmanı ile biten tam bağlantılı katmanlar içermektedir. VGG16'nın temel dezavantajı ise model kapasitesinin yüksek olmasıdır; yaklaşık 138 milyon parametreye

sahiptir. Bu nedenle, eğitilmesi çok yavaştır, oldukça fazla disk alanı ve bant genişliği alır ve bu da onu verimsiz kılar. Şekil 3'te VGG16 mimarisi verilmiştir.



Şekil 3. VGG16 mimarisi. FC katmanları arasında dropout katmanları kullanılmaktadır (The VGG16 architecture. Dropout layers are used between the FC layers)

VGG16_BN: CNN modellerinde ağırlık derinliği arttıkça gradyan azalması (vanishing gradient) olarak adlandırılan bir problem ortaya çıkmaktadır; gradyanlar son katmanlardan ilk katmanlara doğru aktarılırken azalmakta veya sönümlenmektedir. Bu durum, modelin ilk katmanlardaki parametrelerin güncellenmemesine ve dolayısıyla model performansının düşük kalmasına sebep olmaktadır. VGG16_BN modelinde, VGG16'daki her konvolüsyon katmanından sonra batch normalization (BN) [24] katmanı kullanılmaktadır. BN, verileri aktivasyon fonksiyonuna (yani ReLU) vermeden önce ortalama ve standart sapmayı kullanarak veriyi normalize etmektedir. Bu yöntem ile aktivasyonlar doyuma ulaşmaz ve dolayısıyla türev değeri istenen bir aralıkta kalır; gradyanların çok küçük olması engellenir ve gradyan sinyalinin ilk katmanlara ulaşılmasına olanak sağlanır. Şekil 4'te VGG16_BN modelinin ilk blok mimarisi verilmiştir. Diğer bloklarda benzer bir yapıya sahiptir.



Şekil 4. VGG16_BN blok 1 mimarisi (The architecture of VGG16_BN block 1)

VGG16_GAP: VGG16'nın orijinal mimarisinde, sınıflandırma kısmı olarak adlandırılan son katmanlarda, üç tam bağlantılı katman mevcuttur. Bu katmanlardan önce, özellik haritaları düzleştirilmekte ve ardından tam bağlantılı katmanlar ile sınıflandırma yapmak için kullanılmaktadır. Ancak, bu katmanlar birçok parametre içerir ve aşırı öğrenmeye eğilimlidir. Global average pooling (GAP) kullanımı son yıllarda popüler hale gelmiştir. GAP, son özellik haritasındaki her öznelik haritasının ortalamasını alır ve böylece düzleştirme işlemine ihtiyaç duymadan, her bir özellik haritası için tek bir çıktı üretir. Bu, aşırı öğrenmeye karşı daha dirençli ve daha düşük parametrelili bir model sağlamaktadır. Özetle, VGG16-GAP modelinde, VGG16 mimarisindeki 3 tam bağlantılı katman yerine, GAP katmanı kullanılmıştır.

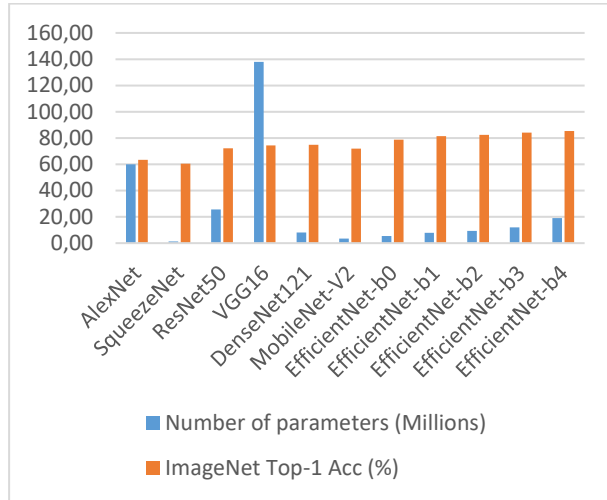
SqueezeNet [25]: Bu mimari yapısında 1x1 evrişim katmanlarını kullanarak model parametre sayısını azaltmayı sağlamaktadır. AlexNet ile kıyaslandığında model parametre sayısını 50 kat azaltmakta ve ayrıca top-1 ve top-5 kategorisinde aynı veya daha yüksek doğruluk oranları sağlamaktadır. Parametre sayısının azalması bazı avantajlar sağlamaktadır: (1) Modelin hesaplama yükü, dolayısıyla hesaplama süresi azalmaktadır, (2) Bir modeli buluttan bir araca aktarmak için daha az bant genişliği gerektirir ve (3) Sınırlı belleğe sahip FPGA'lere ve diğer donanımlara dağıtmak için daha uygundur. SqueezeNet mimarisi tasarlanırken 3 strateji izlenmiştir. İlk olarak, 3x3 filtreler 1x1 filtreler ile değiştirilmiştir; bu parametre sayısının 9 kat azalmasını sağlamıştır. İkinci olarak, sıkıştırma katmanları (squeeze layers) kullanılarak giriş kanallarının sayısı azaltılmıştır. Son olarak, diğer parametreler sabit tutulduğunda, ilk katmanlarda büyük aktivasyon haritaları (stride değeri düşük tutularak bu durum sağlanmaktadır) kullanmanın yüksek sınıflandırma performansı gösterdiği belirtilmiştir. Burada ilk iki strateji model parametre sayısını azaltmaya yönelikken, 3. strateji sınırlı bir parametreye sahip bir modelin

sayısının azaltılması sağlanmıştır; küçük filtre boyutu kullanımından kaynaklanan azalan receptive field'ı telafi etmek için katman sayısı artırılmıştır. Bu mimarinin B0 ve B7 arasında değişen farklı versiyonları önerilmiştir. Şekil 8'de modellerin

ImageNet veri setindeki top-1 başarıları ve parametre sayıları verilmiştir. Dikkat edilirse, EfficientNet modelleri parametre sayısı bakımından daha küçük olmasına rağmen doğruluk oranı bakımından diğer modellerin performansını geçmektedir.

Tablo 1. MobileNet-V2 mimarisi (The architecture of MobileNet-V2)

Girdi boyutu	Katmanlar	Genişleme faktörü	Kanal sayısı	Tekrar sayısı	Stride
224x224x3	Conv2d	-	32	1	2
112x112x3	Bottleneck	1	16	1	1
112x112x3	Bottleneck	6	24	2	2
56x56x24	Bottleneck	6	32	3	2
28x28xx32	Bottleneck	6	64	4	2
14x14x64	Bottleneck	6	96	3	1
14x14x96	Bottleneck	6	160	3	2
7x7x160	Bottleneck	6	320	1	1
7x7x320	Conv2d 1x1	-	1280	1	1
7x7x1280	AvgPool 7x7	-	-	1	-
1x1x1280	Conv2d 1x1	-	k	-	-



Şekil 8. Çalışma kapsamında kullanılan modellerin ImageNet verisetinde top-1 doğruluk oranları ve parametre sayısı bakımından karşılaştırılması (Comparison of the models used in the scope of the study in terms of top-1 accuracy rates and parameter numbers on the ImageNet dataset)

2.3. Eğitim Detayları (Training Details)

Çalışma kapsamında kullanılan CNN modelleri çok fazla parametre barındırmaktadır. Veri miktarının az olması durumunda genel olarak aşırı öğrenme (overfitting) problemini meydana gelmektedir. Aşırı öğrenme, bir modelin (yani fonksiyonun) bir veri setine (yani eğitim veri setine) çok yakından hizalanmasından (aligned) kaynaklı ortaya

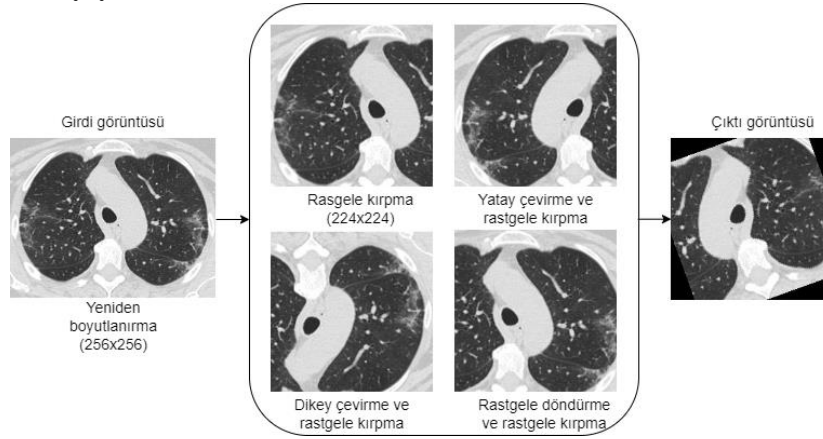
çıkılmaktadır. Sonuç olarak, model eğitim veri setinde yüksek performans göstermekte ve test veri setinde başarısız olmaktadır. Diğer bir deyişle, genelleme yeteneği zayıf kalmaktadır. Bu problemin üstesinden gelmenin bir yolu veri artırma (data augmentation) tekniğini kullanmaktır. Veri artırma ilave eğitim verileri üreterek ve modeli farklı veri sürümlerine maruz bırakarak bir düzenleyici (regularizer) görevi görür. Bu kapsamda eğitim veri setinde bazı işlemler uygulanmıştır:

- Veri setindeki imgeler 256x256 çözünürlüğüne yeniden boyutlandırılmıştır (resize).
- İmgeleri modeller için tavsiye edilen 224x224 çözünürlüğüne getirmek için kırpma (crop) işlemi uygulanmıştır.
- 0.5 olasılığında yatayda çevirme işlemi uygulanmıştır.
- 0.5 olasılığında dikeyde çevirme işlemi uygulanmıştır.
- [0-20] derecelik döndürme işlemi uygulanmıştır.

Şekil 9'da bir imge üzerinden uygulanan veri artırma işlemleri örneklenmiştir. Tüm modeller sıfırdan 50 epoch boyunca eğitilmiştir; fine-tuning işlemi uygulanmamıştır. Optimizer olarak stochastic gradient descent [30] algoritması kullanılmış, batchsize 16 ve öğrenme oranı 2-e3

olarak alınmıştır. Kayıp fonksiyonu olarak CrossEntropyLoss kullanılmıştır. 2 epoch boyunca validation loss'ta düşüş olmaması durumunda

ağırlık azalması (weight decay) işlemi uygulanmış ve öğrenme oranı yarıya düşürülmüştür.



Şekil 9. Eğitim veri setinde uygulanan veri arttırma işlemleri ve üretilen sentetik imge (Data augmentation techniques applied in the training dataset and the generated synthetic images)

3. DENEYSSEL SONUÇLAR (EXPERIMENTAL RESULTS)

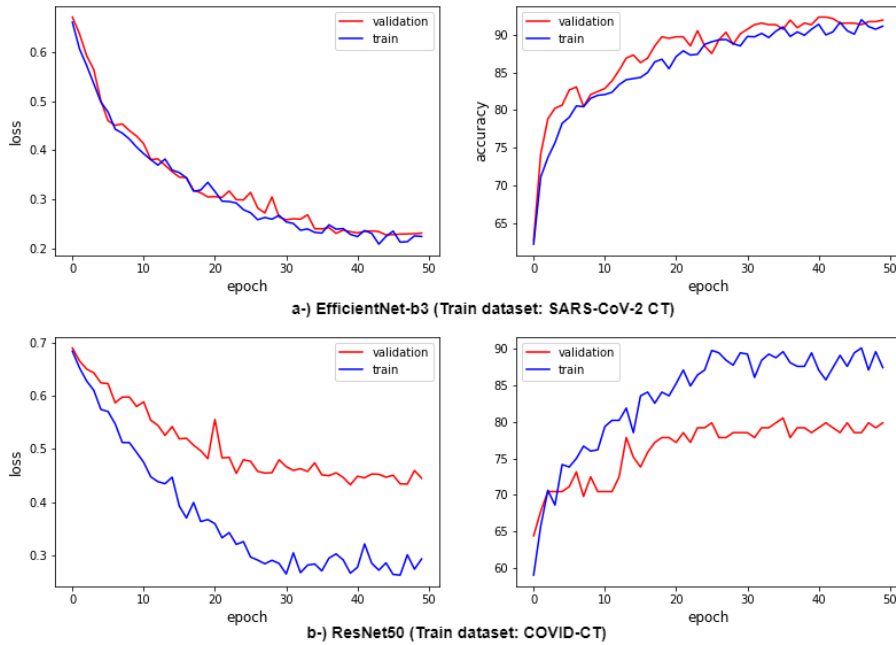
Bu çalışmada temel amacımız çapraz veri seti altında COVID-19 teşhisi için derin öğrenme modellerinin performansını karşılaştırmaktır. Bu kapsamda, ilk senaryoda tek bir hastaneden toplanan SARS-CoV-2 CT veri seti eğitim ve farklı makalelerden toplanan imgelerden oluşan COVID-CT veri seti test olacak şekilde model performansları deneysel olarak incelenmiştir. Ayrıca, hiper-parametre optimizasyonu yapmak ve modelin en iyi performansı hangi epoch'ta sağladığını belirlemek için eğitim veri seti %80 eğitim ve %20 validasyon olarak ayrılmıştır. Eğitim detayları kısmında belirtildiği gibi validasyon kayıp değerindeki değişime bağlı olarak öğrenme oranı ayarlanmıştır. Ayrıca validasyon loss değerinin en düşük olduğu epoch'taki model ağırlık değerleri kaydedilmiş ve bu model kullanılarak COVID-CT test seti ile performansı incelenmiştir. Tablo 2'in ilk kolonunda belirtilen ayarlamalar (setting) için her bir modelin eğitim, validasyon ve test sonuçları verilmiştir. Dikkat edilirse, tüm modellerin eğitim ve validasyon performanslarının yüksek olduğu, ancak test performanslarının düşük olduğu gözlenmektedir. En iyi test skoru %68.93 ile EfficientNet-b3 modeli ile elde edilmiştir. Modeller genel olarak değerlendirildiğinde, eğitim performansları %85.13 ile %97.17 arasındayken test performansları %63.56 ile %68.93 arasında değişkenlik göstermektedir. Sonuçlar incelendiğinde, SARS-CoV-2 CT veri seti eğitim seti olarak kullanıldığında, modelin genelleme yeteneğinin zayıf kaldığı görülmektedir.

İkinci senaryoda ise COVID-19 ile ilgili bilimsel makalelerden toplanan imgelerden oluşan COVID-CT eğitim ve tek bir hastaneden toplanan SARS-CoV-2 CT test seti olacak şekilde model performansları benzer şekilde karşılaştırılmıştır. Tablo 2'in ikinci kolonunda her bir modelin bu ayarlama için performansları verilmiştir. En iyi test skoru %71.47 ile ResNet50 modeli ile elde edilmiştir. Modeller genel olarak değerlendirildiğinde, eğitim performansları %74.66 ile %90.10 arasındayken test performansları %60.18 ile %71.47 arasında değişkenlik göstermektedir. Dikkat edilirse, daha genel bir yapıya sahip olan COVID-CT veri seti eğitim için kullanıldığında test performansının %68.93'dan %71.47'ye yükseldiği görülmektedir. COVID-CT veri setindeki çeşitliliğin ve etnik farklılığın fazla olmasından kaynaklı modellerin eğitim başarısı nispeten düşük olmasına rağmen test performansının daha yüksek olduğu gözlenmiştir. Şekil 10'da her iki senaryo için en yüksek performans gösteren modellerin eğitim aşamasındaki kayıp (loss) (solda) ve doğruluk (accuracy) (sağda) grafikleri verilmiştir. Şekil 10'da üstte eğitim seti olarak SARS-CoV-2 CT'in kullanıldığı ve en yüksek test skorunun elde edildiği EfficientNet-b3 modelinin eğitim/validasyon kayıp ve doğruluk grafikleri verilmiştir. Dikkat edilirse veri setinin çeşitliliğinin düşük olmasından kaynaklı dengeli (stable) bir eğitim söz konusudur. Ancak, modelin genelleme yeteneği düşüktür. Diğer taraftan, altta eğitim seti olarak COVID-CT'nin kullanıldığı ve en yüksek test skorunun elde edildiği ResNet50 modelinin eğitim/validasyon kayıp ve doğruluk grafikleri verilmiştir. Veri setinin

zorluğundan kaynaklı eğitim ve validasyon grafikleri arasında boşluk oluşmakta ve validasyon başarısı düşük kalmaktadır. Test performansına bakıldığında ise Senaryo 1'e göre daha yüksek bir skor elde edilmektedir. Bu da modelin genelleme yeteneğinin daha başarılı olduğunu göstermektedir.

Tablo 2. Çapraz veri seti değerlendirmesi altında derin öğrenme modellerinin performansının karşılaştırılması (Comparison of the performance of deep learning models under cross-validation dataset evaluation)

Modeller	Eğitim veri seti: SARS-CoV-2 CT (eğitim %80 ve doğrulama seti %20) Test veri seti: COVID-CT			Eğitim veri seti: COVID-CT (eğitim %80 ve doğrulama seti %20) Test veri seti: SARS-CoV-2 CT		
	Eğitim Perf.(%)	Val Perf.(%)	Test Perf.(%)	Eğitim Perf.(%)	Val Perf.(%)	Test Perf.(%)
AlexNet [19]	92.04	92.94	65.98	84.56	77.85	60.18
SqueezeNet [22]	90.27	92.54	66.78	85.57	76.51	66.84
VGG16 [20]	94.76	92.34	64.91	80.87	75.17	62.16
VGG16-BN	95.81	96.57	68.39	90.10	79.19	67.80
VGG16-GAP	88.0	88.91	63.83	85.03	86.38	68.61
ResNet50 [23]	96.52	96.17	66.78	89.60	80.54	71.47
DenseNet121 [24]	95.66	94.96	66.11	88.93	79.87	67.08
MobileNet-V2 [25]	85.13	86.89	64.90	76.94	77.11	67.56
EfficientNet-b0 [26]	88.10	90.93	64.64	74.66	71.14	69.7
EfficientNet-b1	88.86	91.73	65.97	77.01	76.51	70.26
EfficientNet-b2	89.06	92.54	63.56	83.49	77.18	67.84
EfficientNet-b3	91.38	92.34	68.93	78.36	78.52	69.29
EfficientNet-b4	97.17	98.39	66.78	80.37	77.18	69.37



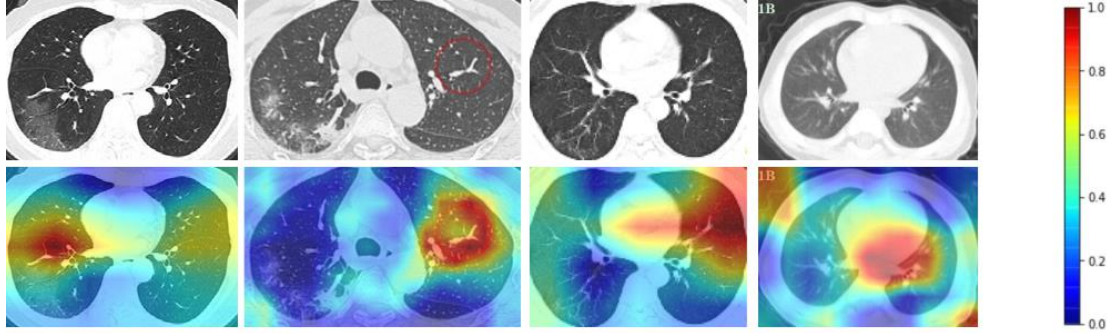
Şekil 10. Kayıp (solda) ve Doğruluk (sağda) grafikleri (Loss (left) and Accuracy (right) plots)

Literatürdeki çalışmalar ile kıyaslandığında, çapraz veri seti değerlendirmesi altında model performanslarının oldukça düşük olduğu gözlenmiştir. Bu, COVID-19 kapsamında toplanan veri setlerinin etnik bağımlılığa sahip olduğunu

göstermektedir. Modellerin genelleme yeteneğini arttırmak için farklı ülke ve hastanelerden toplanan ortak bir veri setinin kullanılmasının daha iyi olacağı gözlenmiştir. Ayrıca, COVID-19 veri seti için model performanslarının, Şekil 8'te verilen

modellerin ImageNet veri setindeki performansları ile paralel olmadığı görülmektedir. Deneysel çalışmalar ile çapraz veri değerlendirmesi altında COVID-19 tespiti için ResNet50 modelinin daha iyi performans sağladığı görülmüştür.

Eğitim işlemi tamamlandıktan sonra, CNN modelinin görüntünün sınıfına (Covid/Non-Covid) karar verirken odaklandığı alanları göstermek için Grad-CAM (Gradient weighted Class Activation Mapping) yöntemi kullanılmaktadır. Genel olarak,



Şekil 11. Grad-CAM Sonuçları (Grad-CAM Results)

4. SONUÇLAR (EXPERIMENTAL RESULTS)

Bu çalışmada, tek bir hastaneden toplanan görüntülerden oluşan SARS-CoV-2 CT veri seti ve literatürde yayınlanan makalelerden toplanan görüntülerden oluşan COVIT-CT veri seti kullanılarak çapraz veri değerlendirmesi altında güncel derin öğrenme modellerinin performansı karşılaştırılmıştır. Deneysel araştırmalar sonucu, eğitim seti olarak COVIT-CT kullanıldığında, eğitim performansı düşük olmasına rağmen test skorlarının daha iyi olduğu görülmüştür. Bunun sebebi, veri setinin daha genel ve çeşitli olmasından kaynaklanmaktadır. Öte yandan SARS-CoV-2 CT veri seti aynı ildeki hastaneden toplandığından, her ne kadar eğitim seti başarısı yüksek olsa da test başarısı düşüktür; bu veri seti kullanılarak eğitilen modelin genelleme yeteneğinin zayıf olduğu gözlemlenmiştir. Yapılan deneysel çalışmalarda, çapraz veri seti değerlendirmesi altında %71.47 ile en yüksek skor ResNet50 modeli ile elde edilmiştir. Çalışmanın devamında, ResNet50 modelinin performansını arttırmaya yönelik araştırmalar yapılacaktır.

TEŞEKKÜR (ACKNOWLEDGEMENTS)

Bu araştırmada yer alan tüm nümerik hesaplamalar TÜBİTAK ULAKBİM, Yüksek Başarım ve Grid Hesaplama Merkezi'nde (TRUBA kaynaklarında) gerçekleştirilmiştir.

ilgili modelin en son evrişim katmanındaki özellik haritasından yararlanılmaktadır. Bu kısımda, en yüksek skorları aldığımız ResNet50 modeli kullanılarak, modelin tahmin yaparken odaklandığı bölgeler görselleştirilmiştir. Şekil 11'de elde edilen sonuçlar ve bölgelerin önemini gösteren ısı haritası (sağda) gösterilmiştir. Burada, kırmızı bölgeler modelin sınıflandırma yaparken odaklandığı kısımları, mavi bölgeler ise sınıflandırmada daha az önemli olan kısımları göstermektedir.

ETİK STANDARTLARIN BEYANI

(DECLARATION OF ETHICAL STANDARDS)

Bu makalenin yazarı çalışmalarında kullandıkları materyal ve yöntemlerin etik kurul izni ve/veya yasal-özel bir izin gerektirmediğini beyan ederler.

The author of this article declares that the materials and methods they use in their work do not require ethical committee approval and/or legal-specific permission.

YAZARLARIN KATKILARI (AUTHORS' CONTRIBUTIONS)

Yahya DOĞAN: Deneyleyi yapmış, sonuçlarını analiz etmiş ve maklenin yazım işlemini gerçekleştirmiştir.

He conducted the experiments, analyzed the results and performed the writing process.

ÇIKAR ÇATIŞMASI (CONFLICT OF INTEREST)

Bu çalışmada herhangi bir çıkar çatışması yoktur.

There is no conflict of interest in this study.

KAYNAKLAR (REFERENCES)

- [1] Xu X, Jiang X, Ma C, Du P, Li X, Lv S, Yu L, Ni Q, Chen Y, Su J. A deep learning system to screen novel coronavirus disease 2019 pneumonia. *Engineering*, 2020; 6(10): 1122-1129.
- [2] Ai T, Yang Z, Hou H, Zhan C, Chen C, Lv W, Tau Q, Sun Z, Xia L. Correlation of chest CT

- and RT-PCR testing in coronavirus disease 2019 (COVID-19) in China: a report of 1014 cases. *Radiology*, 2020.
- [3] Panwar H, Gupta P. K, Siddiqui M. K, Morales-Menendez R, Singh V. Application of deep learning for fast detection of COVID-19 in X-Rays using nCOVnet. *Chaos, Solitons & Fractals*, 2020; 138: 109944.
- [4] Brunese, L, Mercaldo F, Reginelli A, Santone A. Explainable deep learning for pulmonary disease and coronavirus COVID-19 detection from X-rays. *Computer Methods and Programs in Biomedicine*, 2020; 196: 105608.
- [5] Das N. N, Kumar N, Kaur M, Kumar V, Singh D. Automated deep transfer learning-based approach for detection of COVID-19 infection in chest X-rays. *Irbm*. 2020.
- [6] Amyar A, Modzelewski R, Li H, Ruan S. Multi-task deep learning based CT imaging analysis for COVID-19 pneumonia: Classification and segmentation. *Computers in Biology and Medicine*. 2020; 126: 104037.
- [7] Silva P, Luz E, Silva G, Moreira G, Silva R, Lucio D, Menotti D. COVID-19 detection in CT images with deep learning: A voting-based scheme and cross-datasets analysis. *Informatics in medicine unlocked*. 2020; 20: 100427.
- [8] Saood A, Hatem I. COVID-19 lung CT image segmentation using deep learning methods: U-Net versus SegNet. *BMC Medical Imaging*. 2021; 21(1): 1-10.
- [9] Gibson E, Giganti F, Hu Y, Bonmati E, Bandula S, Gurusamy K, Davidson B, Pereira S. P, Clarrkson M. J, Barratt D. C. Automatic multi-organ segmentation on abdominal CT with dense V-networks. *IEEE transactions on medical imaging*. 2018; 37(8): 1822-1834.
- [10] Qiu T, Wen C, Xie K, Wen F. Q, Sheng G. Q, Tang X. G. Efficient medical image enhancement based on CNN-FBB model. *IET Image Processing*. 2019; 13(10): 1736-1744.
- [11] Fakoor R, Ladhak F, Nazi A, Huber M. Using deep learning to enhance cancer diagnosis and classification. In *Proceedings of the international conference on machine learning*. ACM, New York, USA. 2013; 28: 3937-3949.
- [12] Atas M, Ozdemir C, Atas İ, Ak B, Özeroğlu E. Biometric identification using panoramic dental radiographic images with few-shot learning. *Turkish Journal of Electrical Engineering and Computer Sciences*. 2022; 30(3): 1115-1126.
- [13] Ahmet, K. A. R. A. Uzun-Kısa Süreli Bellek Ağı Kullanarak Global Güneş Işınımı Zaman Serileri Tahmini. *Gazi University Journal of Science Part C: Design and Technology*. 2019; 7(4): 882-892.
- [14] Atas I, Ozdemir C, Atas M, Dogan Y. Forensic Dental Age Estimation Using Modified Deep Learning Neural Network. *arXiv preprint arXiv:2208.09799*. 2022.
- [15] Yetis A. D, Yesilnacar M. I, Atas M. A machine learning approach to dental fluorosis classification. *Arabian Journal of Geosciences*. 2021; 14(2): 1-12.
- [16] ER M. B. Akçiğer Seslerinin Derin Öğrenme ile Sınıflandırılması. *Gazi University Journal of Science Part C: Design and Technology*. 2020; 8(4): 830-844.
- [17] Ozdemir C, Gedik M. A, Kaya Y. Age Estimation from Left-Hand Radiographs with Deep Learning Methods. *Traitement du Signal*. 2021; 38(6).
- [18] Kontuk R, Turan M. NLP Kullanılarak Haberlerin Yaş Gruplarına Göre Sınıflandırılması. *Gazi University Journal of Science Part C: Design and Technology*. 2020; 8(2): 372-382.
- [19] Yang X, He X, Zhao J, Zhang Y, Zhang S, Xie P. COVID-CT-dataset: a CT scan dataset about COVID-19. *arXiv preprint arXiv:2003.13865*. 2020.
- [20] Soares E, Angelov P, Biaso S, Froes M. H, Abe D. K. SARS-CoV-2 CT-scan dataset: A large dataset of real patients CT scans for SARS-CoV-2 identification. *MedRxiv*. 2020.
- [21] Deng J, Dong W, Socher R, Li L. J, Li K, Fei-Fei L. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*. 2019; 248-255.
- [22] Krizhevsky A, Sutskever I, Hinton G. E. Imagenet classification with deep convolutional neural networks. *Communications of the ACM*. 2017; 60(6): 84-90.
- [23] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*. 2014.
- [24] Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International conference on machine learning*. 2015; 448-456.
- [25] Iandola F. N, Han S, Moskewicz M. W, Ashraf K, Dally W. J, Keutzer K. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and < 0.5 MB model size. *arXiv preprint arXiv:1602.07360*. 2016.
- [26] He K, Zhang X, Ren S, Sun J. Deep residual learning. *Image Recognition*. 2015; 7.
- [27] Huang G, Liu Z, Van Der Maaten L, Weinberger K. Q. Densely connected convolutional networks. In *Proceedings of the*

- IEEE conference on computer vision and pattern recognition. 2017;4700-4708.
- [28] Sandler M, Howard A, Zhu, M, Zhmoginov A, Chen L. C. Mobilenetv2: Inverted residuals and linear bottlenecks. In Proceedings of the IEEE conference on computer vision and pattern recognition. 2018; 4510-4520.
- [29] Tan M, Le Q. Efficientnetv2: Smaller models and faster training. In International Conference on Machine Learning. 2021; 10096-10106.
- [30] Ketkar N. Stochastic gradient descent. In Deep learning with Python. Apress, Berkeley, CA. 2017; 113-132.