

Boruta Öznitelik Seçimi Algoritması ve Derin Öğrenme Yöntemleri Kullanılarak Covid-19 Hastalığının Prognozunun Tahmini

Nedim MUZOĞLU¹, Melike KAYA KARAASLAN², Ahmet Mesrur HALEFOĞLU³, Bekir Sıddık Binboğa YARMAN⁴

¹ İstanbul Üniversitesi-Cerrahpaşa, Lisansüstü Eğitim Enstitüsü, Biyomedikal Mühendisliği, İstanbul, Türkiye,

² Kocaeli Üniversitesi, Lisansüstü Eğitim Enstitüsü, Biyomedikal Mühendisliği, Kocaeli, Türkiye

³ Sağlık Bilimleri Üniversitesi, Şişli Hamidiye Etfal Eğitim ve Araştırma Hastanesi, Radyoloji Departmanı, İstanbul, Türkiye,

⁴ İstanbul Üniversitesi-Cerrahpaşa, Mühendislik Fakültesi, Elektrik Elektronik Bölümü, İstanbul, Türkiye.

Sorumlu yazar e-posta: nmuzoglu@gmail.com

ORCID ID: <http://orcid.org/0000-0003-1591-2806>

melikekaya08@gmail.com

ORCID ID: <http://orcid.org/0000-0001-9078-8468>

halefoglu@hotmail.com

ORCID ID: <http://orcid.org/0000-0002-2054-3550>

yarman@istanbul.edu.tr

ORCID ID: <http://orcid.org/0000-0003-1562-5524>

Geliş Tarihi: 09.05.2022

Kabul Tarihi: 23.06.2022

Öz

Covid-19 pandemisi nedeniyle milyonlarca insan hayatını kaybetmiş ve birçok ülkede yetersiz sağlık sistemleri hizmet veremez hale gelmiştir. Covid-19 hastalarının yoğun bakım ve ventilasyon ihtiyaçlarının belirlenerek hastalığın prognozu hakkında tahminlerde bulunulması, hastanın sağlık durumu ve sağlık sistemlerinin etkin kullanımı açısından önemlidir. Bu amaçla oluşturulan Covid-19 akciğer bilgisayarlı tomografi (BT) bulguları veri seti buzlu cam opasitesi, konsolidasyon, kaldırım taşı paterni, konsolidasyon ve buzlu cam, nodül ve buzlu cam sınıflarını içermektedir. Bu çalışmada önerilen yaklaşım dört adımdan oluşmaktadır. Birinci adımda VGG-16 modeli akciğer BT bulguları veri seti ile eğitilmiştir. İkinci adımda elde edilen en ayırt edici öznitelikler BORUTA algoritması kullanılarak seçilmiştir. Üçüncü adımda sıralama yöntemiyle her görüntü için en değerli ilk 200, 300 ve 400 öznitelikler elde edilmiştir. Son adımda ise Destek Vektör Makineleri ve Lineer Diskriminant Analizi ile bu özellikler sınıflandırılmıştır. Akciğer BT bulguları veri seti için elde edilen genel doğruluk %97,02'dir. Derin Öğrenme yöntemleri ile Covid-19 hastalık prognozunu tahmin etmek için oluşturulan veri seti kullanılarak elde edilen bu başarılı sonuç, viral pnömoni türlerinin akciğer BT bulgularının sınıflandırılmasında çok önemli bir yeniliktir.

Anahtar kelimeler

Covid-19; Akciğer bt bulguları; Boruta; Buzlu cam opasitesi; Konsolidasyon; Kaldırım taşı paterni

Prediction of the Prognosis of Covid-19 Disease Using Deep Learning Methods and Boruta Feature Selection Algorithm

Abstract

Millions of people have lost their lives due to the Covid 19 pandemic, and inadequate health systems have been overwhelmed in many countries. Determining the intensive care and ventilation needs of Covid-19 patients and thus making predictions about the prognosis of the disease is crucial in terms of the patient's health status and the effective use of health systems. The Covid-19 chest computed tomography (CT) findings dataset created for this purpose consists of ground-glass opacity (GGO), consolidation, crazy paving pattern (CPP), consolidation and ground glass (GGOC), nodule and ground glass classes (GGON). The approach proposed in this study consists of four steps. The VGG16 model was trained with the chest CT findings dataset in the first step. The most discriminative features obtained in the second step were selected using the BORUTA algorithm. In the third step, the most valuable top 200, 300 and 400 features for each image were obtained by ranking method. In the last step these features were classified with Support Vector Machines and Linear Discriminant Analysis. The overall accuracy obtained for the chest CT findings dataset is 97.02%. This successful result, obtained using the dataset to predict Covid 19 disease prognosis with Deep Learning methods, is a crucial innovation in the classification of chest CT findings in viral types of pneumonia.

Keywords

Covid-19; Chest ct findings; Boruta; Ground glass opacities; Consolidation; Crazy paving pattern

1. Giriş

Aralık 2019'da Çin'in Wuhan kentinde ilk SarsCov2(Covid-19) enfeksiyon vakası bildirilmiştir. SarsCov2 (Covid-19) virüsü dünya çapında hızla yayılmış ve Mart 2020'de Dünya Sağlık Örgütü (DSÖ) tarafından pandemi olarak ilan edilmiştir (Özdemir, 2021). Covid-19 genellikle birbirine yakın kişiler arasında virüsü içeren aerosoller veya damlacıklar solunduğunda veya doğrudan burun, ağız veya gözlerle temas ettiğinde bulaşmaktadır. Diğer bir çok virüs gibi Covid-19'da bir RNA virüsüdür (Batirel 2020, Türken ve Köse 2020, Gayaf vd. 2021). Covid-19 virüsü vücudumuza girdikten sonra RNA'sını serbest bırakır(Togay ve Yılmaz 2020). Enfekte olan hücreler bu RNA'yı okur ve virüsün yeni kopyalarının bir araya getirilmesine yardımcı olan proteinleri üretir böylece hastalığın tüm vücuda ve diğer insanlara yayılmasına neden olmaktadır. Virüs bulaşan hastada semptomların ortaya çıkması ortalama olarak 5-6 gün sürmektedir. Hastalığın en yaygın semptomları arasında vücut sıcaklığının artması, öksürük, yorgunluk, tat veya koku kaybı ve vücut ağrısı şeklinde görülmektedir (Çelik ve Köse 2020, Kıral 2021). Covid-19 yaygın bulaşıcı bir salgın haline gelerek dünya çapında 532.887.351 doğrulanmış vakaya ve 6.307.021 ölüme neden olurken ekonomik ve sosyal problemleri de beraberinde getirmiştir (Int Kyn. 1). Hastalığın tespitinde ters transkripsiyon polimeraz zincir reaksiyonu (rt-PCR) testi, hastalığın erken evrelerinde veya düşük yüklü örneklerde yeterli duyarlılığa (% 60 -70) sahip olmasa da Covid-19'un tanısında kabul edilen standart referanstır (Fang *et al.*, 2020). Öte yandan, yüksek duyarlılığı (% 98) nedeniyle bilgisayarlı tomografi (BT) görüntüleme ile de Covid-19'un akciğerdeki şiddeti ve seyri takip edilebilmektedir (Ai *et al.*, 2020, Akbaş 2021, Ozer 2020, Cömert 2020). Bu nedenle Covid-19 hastalığının hızlı ve yüksek hassasiyette tespitinde, pulmoner evresinin radyologlar tarafından değerlendirilmesinde ve hastaların uygun triyajında BT görüntüleme çok önemli bir rol oynamaktadır. Covid-19'un akciğerlerdeki BT bulguları ve hastalığın prognozu ile arasındaki ilişki birçok çalışmada incelenmiş olup hastalığın evreleri aşağıda verilmiştir (Altıntaş ve Şenol 2021, Akdaş vd. 2021, Karahacıoğlu vd. 2022).

I. Erken evrede, ilk semptomun başlamasından 0-4 gün sonra buzlu cam opasitesi (GGO) genellikle alt loblarda tek veya çift taraflı olarak görülür.

II. İlk semptomdan 5-8 gün sonra ilerleyici aşamaya gelinir ve hastalık ilerledikçe bilateral GGO, kaldırım taşı paterni (CPP) ve konsolidasyon görülür.

III. İlk semptomdan 9-13 gün sonra hastalığın zirvesine ulaşılır ve yoğun konsolidasyon, GGO ve CPP alanları gözlenir.

IV. Absorpsiyon fazı olarak da adlandırılan ilk semptomdan sonraki 14 günlük dönemde konsolidasyon yavaş yavaş kaybolur ve CPP artık görülmez, ancak GGO gözlemlenebilir.

Yapılan çalışmalarda, sıradan hastalar ile ağır/kritik hastaların Covid-19 BT bulguları incelendiğinde hastaların yoğun bakım ünitesine (YBÜ) yatırılmasına ve/veya mekanik ventilasyon gerekip gerekmediğine karar verecek şekilde önemli farklılıklar olduğu yapılan çalışmalarla tespit edilmiştir(Cau *et al.*, 2021). Dolayısıyla bu bulguların tespitiyle, Covid-19 hastalarının prognozu tahmin edilerek hem sağlık hizmetlerinin yeterliliği artırılacak hem de hastaların sağlığı açısından önemli kazanımlar sağlanabilecektir (Pekçevik ve Belet, 2020). Covid-19 hastalığının çeşitli evrelerinde göğüs BT bulgularını inceleyerek hastalığın prognozunu tahmin etmeye yönelik bazı çalışmalar aşağıdaki gibidir.

Sanli vd., (2021) 7 günden az dışardan tedavisine devam edilen (grup1), yedi günden fazla hastanede tedavi olan (grup2) ve en az bir gün yoğun bakımda kalan (grup3) toplam 231 Covid-19 hastası ile yapılan çalışmada grup2 ve grup3'te kaldırım taşı paternleri ve konsolidasyon paternlerinin oranlarının ayrıca etkilenen lob sayısının yüksek olduğunu tespit etmişlerdir. Konsolidasyon paterni ile sağ orta veya üst lob tutulumunun hâkim olduğu infiltrasyon durumunda YBÜ ihtiyacının artabileceğini belirtmişlerdir. Li *et al.*, (2020) akciğer tomografi bulgularının mortalite ile ilişkisini araştırdıkları çalışmalarında konsolidasyon bulgularının hayatını kaybeden grupta daha sıklıkla karşılaşıldığını tespit etmişlerdir. Ding *et al.*, (2020) 112 Covid-19 hastalığının devam ettiği süre esas alınarak beş farklı grupta yaptığı araştırmada hastalığı daha uzun sürede geçiren 3., 4. ve 5. grupta konsolidasyon ve CPP'nin daha sıklıkla görüldüğünü tespit etmişlerdir. Cau *et al.*, (2021) 23 hasta YBÜ'den ve 195 hasta kliniklerde yatan Covid-19 hastalarına ait BT bulguları ile yaptıkları çalışmalarında YBÜ hastalarının, YBÜ olmayan hastalara göre GGO ile karışık akciğer konsolidasyonu ve bilateral akciğer tutulumu insidansının daha yüksek olduğunu bulmuşlardır. Yapay zekâ kullanılarak tıbbi görüntülerden

hastalıkların teşhis edilmesi konusunda çok başarılı çalışmalar olduğu gibi Covid-19'un tespitinde ve hastalığın prognozunda da birçok başarılı çalışma yapılmıştır (Toğaçar *et al.*, 2020). Kauczor *et al.*, (2000) 84 hasta BT görüntüsünde buzlu cam opasitesinin tespiti için bir model önermişlerdir. Buzlu cam opasitesinin (GGO) segmentasyonunda bir yoğunluk maskesi kullanarak önerilen model ile %99 hassasiyet ve %89 doğruluk elde etmişlerdir. Covid-19'un şiddetini değerlendirmek için klinik ve görüntüleme verilerini kullanan Quiroz *et al.*, (2021) önerdikleri model için toplam 346 hastaya ait görüntüleri kullanmışlardır. Ayrıca, EfficientNetB7 U-Net mimarisini göğüs BT bulguları GGO, konsolidasyon ve diğer anormallikleri içeren veri setini segmentlere ayırmak için kullanmışlardır. İlave, sentetik azınlık aşırı örnekleme tekniği (SMOTE), ciddi ve kritik durumlarda dengesizlik oranını aşırı derecede örnekleme için kullanılmıştır. Sonuç olarak, Shapley Katkı Açıklamaları (SHAP) ile değerlendirmede, şiddeti belirleyen en önemli özelliklerin başında BT bulgularının geldiğini ortaya koymuşlardır. Doğrusal regresyon ile elde edilen en iyi sonuçlar ise %95 AUC, %76 duyarlılık ve %91,19 özgüllük olmuştur. Toğaçar *et al.*, (2022) ilk kez, Covid-19 BT taramalarının radyologlar tarafından BT bulgularına göre sınıflandırıldığı akciğer BT bulguları Veri Setini kullanmıştır. Veri seti GGO, nodül ve GGO (GGON), CPP, konsolidasyon, konsolidasyon ve GGO (GGOC) içeren BT bulgu sınıflarını içerir. Bu çalışmalarında baskın aktivasyon setlerini çıkarmak için ResNet modellerini kullanıp ardından yerel yorumlanabilir model agnostik açıklamalar (LIME) yöntemini kullanarak en iyilerini seçip ardından Softmax ile sınıflandırılma işlemi yapılmıştır. Bu öneri ile %99,62 genel doğruluk elde edilmiştir. Çalışmalar Covid-19 BT bulgularının tespitinin hastalığın ciddiyetini ve seyrini tahmin etmede kullanılabileceğini böylece hastaların ayaktan tedavi, hastane klinik ya da yoğun bakım servislerinde tedavilerine karar vermede sağlık çalışanlarına yardımcı olabileceğini göstermektedir.

Bu çalışmada, Covid-19 hastalığının akciğerlerdeki tutulumunun tespiti için oluşturulan beş sınıflı (GGO, GGON, GGOC, CPP ve Konsolidasyon) BT akciğer bulguları veri seti kullanılarak hastalığın prognozunun tahmininin yapılması ve hastaların gerektiğinde yoğun bakım ihtiyacının hızlıca tespit edilmesi amaçlanmıştır. Böylece Covid-19 ve olası benzer pandemilerde BT akciğer bulguları kullanılarak yapılan tespitlerin sağlık sistemlerinin etkin olarak kullanılmasına katkıda bulunması amaçlanmıştır. Bu çalışmada derin öğrenme modelinden elde edilen özniteliklerden en etkin olanlar seçim algoritması yardımıyla seçilerek makine öğrenme sınıflandırıcılarıyla sınıflandırılıp BT bulgularının tespiti yöntem olarak tercih edilmiştir.

Bu çalışma şu şekilde düzenlenmiştir; veri seti, derin öğrenme modeli ve uygulama detayları Bölüm 2'de açıklanmıştır. Bulgular Bölüm 3'te tartışma Bölüm 4'te ve çalışmanın sonunda ise sonuç bölümü yer almaktadır.

2. Materyal ve Metot

2.1. Veri Seti

Veri seti, gerekli Etik Kurul izinleri, alındıktan sonra İstanbul Hamidiye Etfal Eğitim ve Araştırma Hastanesi PACS sisteminde kayıtlı ve hastane radyologları tarafından incelenen 196 hastaya ait 614 Covid-19 BT görüntüsünden elde edilmiştir. Ayrıca bu veri setine Yang *et al.*, (2020) tarafından yayınlanan COVID-CT Veri seti de radyologlar tarafından BT bulgularına göre analiz edilip eklenerek CovCT Bulgular Veri Seti oluşturulmuştur. Tüm resimler .jpg, .jpeg formatındadır ve 512x512 ile 1096x1289 piksel arasında değişen boyutlara sahiptir. Ancak Derin Öğrenme modelinin girdi boyutuna göre 224x224 olarak yeniden boyutlandırılmıştır. Bu çalışmada kullanılan CovCT Bulgular Veri Setinden her bir sınıfa ait örnek görüntüler Şekil 1'de ve CovCT-Bulgular Veri Setinde yer alan her bir sınıftaki BT görüntü sayıları ile ilgili ayrıntılar ise Çizelge 1'de verilmiştir.



Şekil 1. CovCT Bulguları Veri Seti sınıflarından örnek BT görüntüleri a) Buzlu Cam ve Nodül (GGON) b) Kaldırım Taşı Paterni (CCP) c) Konsolidasyon d) Buzlu Cam ve Konsolidasyon (GGOC) e) Buzlu Cam Opasitesi (GGO).

Çizelge 1. CovCT Bulgular Veri Setinde her bir sınıfa ait toplam BT görüntü sayıları.

CovCT-Bulgular Veri Seti	COVID-CT Veri Seti	COVID-19 Veri Seti	BT Görüntü Sayısı	Toplam BT Görüntü Sayısı
Buzlu Cam Opasitesi (GGO)	0	337	337	
Konsolidasyon (Consolidation)	108	36	144	
Buzlu Cam ve Nodül (GGON)	7	191	198	785
Buzlu Cam ve Konsolidasyon (GGOC)	27	23	50	
Kaldırım Taşı Paterni (CPP)	29	27	56	

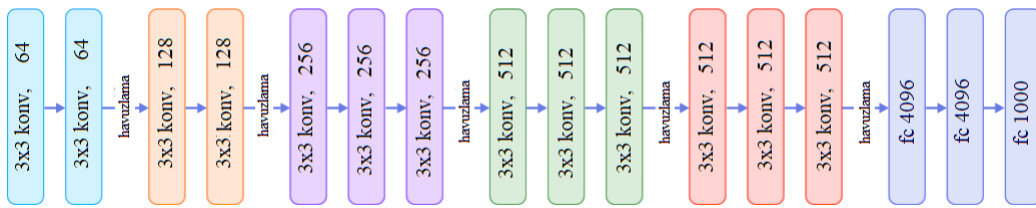
2.2 Konvolüsyonel Derin Öğrenme Ağları

VGG-16 derin öğrenme ağı ile Covid-19 hastalığının teşhisinde birçok çalışmada başarılı sonuçlar elde edildiğinden bu çalışmada da tercih edilmiştir. VGG-16'nın giriş görüntülerin boyutu 224x224 olmalıdır. Derin öğrenme ağının eğitiminde VGG-16'nın girişindeki görüntüler çekirdek adı verilen matrislerle eleman bazında çarpılarak ve konum bazında toplanıp öznitelik haritaları elde edilir. Elde edilen her bir öznitelik haritası ise sonrasında lineer olmayan aktivasyon fonksiyonuna işlemine tabi tutulur. İki konvolüsyon katmanı arasında ise genellikle bir havuzlama katmanı kullanılarak önemli özellikler kaybedilmeden aktivasyon haritalarının boyutu düşürülür. Havuzlama katmanından elde edilen öznitelikler tek boyutlu vektöre dönüştürülerek tüm nöronlarının birbirine bağlı olduğu tam bağlantılı katmana iletilir. Tam bağlantılı katmanın en sonunda sınıflandırma katmanı olarak kullanılan Softmax fonksiyonu bulunur. Softmax fonksiyonu girdilerini, giriş

verisinin sınıfları sayısında, çıkışlarında olasılık değerlerine dönüştürerek maliyet fonksiyonuna aktarır. Giriş değeri ile çıkışta tahmin edilen sınıfa ait hata değeri belirlenen değere ulaşana kadar derin öğrenme ağının parametreleri güncellenir ve bu derin öğrenme ağının eğitimi olarak tanımlanır. VGG-16 ağının ağırlıklarının eğitim tamamlanana kadar optimize edilmesi Stokastik gradyan inişi (SGD) ile gerçekleştirilmiştir. SGD optimizasyonu her bir eğitim verisi x^i ve etiketi y^i için parametre güncellemesi yapar ve α öğrenme katsayısı ile gradyanların çarpılması sonucunda güncel ağırlıklar (w_t) Eşitlik 1'de ifade edilmiştir.

$$w_t = w_t - \alpha \frac{\partial L}{\partial w_t} \quad (1)$$

Bu çalışmada eğitim sırasında derin öğrenme ağının öğrenme oranı 10^{-5} , yığın boyutu 64, döngü sayısı 50 ve veri setinin eğitim test oranı 70:30 seçilmiştir. VGG-16'nın özniteliklerin çıkartıldığı fc8 katmanına kadar ki genel gösterimi Şekil 2'de verilmiştir.



Şekil 2. VGG -16 Derin Öğrenme Ağının Mimarisi.

2.3 Öznitelik Seçme Yöntemi

Yapay zekâ için kullanılan veri setleri aslında bir grup öznitelikten oluşmaktadır. Veri setleri içerisindeki gereksiz ve yararsız özniteliklerin belirlenerek veri setinden çıkartılarak boyutun düşürülmesiyle daha iyi sınıflandırma performansı sağlanacaktır. Kurşa, *et al.*, (2010) önemli ve önemsiz nitelikleri birbirlerinden ayırt etmek için öncelikle orijinal veri setine rastgele gölge öznitelikler ekleyerek genişletip sonrasında ise genellikle rastgele karar ormanları tarafından eğitildiği Boruta'ya

önermişlerdir. Bu yöntemde belirli bir özelliği kullanan ormandaki tüm karar ağaçları için sınıflandırma doğruluğu kaybı ayrı ayrı hesaplanırken her seferinde niteliklerin önemi hesaplanır. Önem ölçüsü, ortalama kaybı standart sapmasına bölerek hesaplanan Z puanıdır. Her bir öznitelik için ve gölge özniteliklerin en yüksek olduğu Z puanı hesaplanarak (MZS) belirlenen değerden çok yüksek ve çok düşük olanlar elenir. Rastgele değişkenlerden daha yüksek öneme sahip değişkenler

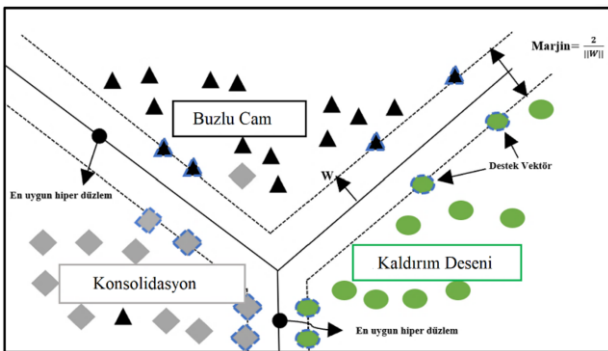
önemli kabul edilmiştir. Gölge olanlardan çok daha kötü öneme sahip özellikler ise elenir. Bu süreç belirlenen bir iterasyon sayısı kadar ya da özniteliklerin reddedilme ve kabul edilme işlemleri tamamlanana kadar devam eder. Bu çalışmada Boruta öznitelik seçimi algoritması rastgele karar ormanları ile çalıştırılarak en değerli öznitelikler elde edilmiştir. Rastgele Ormanlar birbirinden bağımsız karar ağaçlarından oluşur ve bu yapıda her bir karar ağacının sınıf tahmini değerlendirilerek en çok oyu alan tahmin modelin tahmini olacağından herhangi bir ağacın bireysel performansından daha yüksek performans gösterirler.

2.4 Destek Vektör Makineleri

Destek Vektör Makineleri (DVM), sınıflar arasında en büyük marjla optimal hiperdüzlemi bulmak için geliştirilmiş bir sınıflandırma yöntemidir (Cortes *et al.*, 1995). Sınıflardan hiperdüzleme en yakın olan ve hiperdüzlem/karar sınırını bir sütun olarak destekleyen noktalara destek vektörleri denir. N boyutlu uzayda giriş örnekleri ve çıkış değerleri $\{x_i, y_i\}$ olmak üzere $i=1, 2, \dots, r$, ve x hiperdüzlem üzerinde uzanan bir noktayı, $y_i = \{1, -1\}$ ile temsil edilen eğitim verilerini ve $(w \cdot x_i + b) \geq \pm 1$ ile tanımlanan hiperdüzlemi, w hiperdüzlemin katsayı vektörünü, b hiperdüzlemin orijinden uzaklığı olmak üzere birleştirilen hiperdüzlemlerin denklemi Eşitlik 2'de ifade edilmiştir.

$$y_i(w \cdot x_i + b) - 1 \geq 0 \quad (2)$$

Doğrusal olarak ayrılabilen sınıflar için iki boyutlu uzayda DVM temsili gösterimi Şekil 3'te verilmiştir.



Şekil 3. DVM temsili gösterimi.

Burada destek vektörleri ile hiperdüzlem arasındaki maksimum mesafenin genişliği marjin olarak tanımlanmıştır. DVM iki sınıflı sınıflandırma esasına göre yapılanmış bir metot olsa da çok sınıflı sınıflandırmalar için sınıflı sınıflandırma düzeyine düşürülerek çoklu

sınıflandırmalar için birçok çalışmada kullanılmaktadır. Sınıfların arasında doğrusal olmayan karar yüzeyleri olduğu durumlarda ise doğrusal olmayan karar yüzeylerini doğrusal bir fonksiyona dönüştürerek sınıflandırmak için çekirdek fonksiyonları kullanılır. Bu çalışmada bazı denemeler sonucunda en yüksek başarıyı verdiği tespit edilen doğrusal DVM metodu tercih edilmiştir.

2.5 Doğrusal Diskriminant Analizi

Doğrusal Diskriminant Analizi (DDA) ile veri sınıflandırılması yapılırken veri setinde sınıflar arası varyansın sınıf içi varyansa oranı maksimize edilerek maksimum ayrılabirlik sağlanır (Park and Park, 2008). Veri seti ve test setinin ortalama değerleri μ_1 ve μ_2 olarak ve sınıfların bağımsız olasılık değerleri p_1 ve p_2 alınarak μ_3 her iki kümenin birleşiminin ortalaması olmak üzere Eşitlik 3 ile ifade edilmektedir. Her bir sınıfın beklenen kovaryans matrisi ve dağılımı Eşitlik 4 ve Eşitlik 5 ile ifade edildiğinde sınıflar arası beklenen dağılım ise Eşitlik 6 ile ifade edilir. Burada S_b her bir sınıftaki verilerin ortalama değerlerinin kovaryansı olmak üzere optimizasyon kriterini hesaplamak için sınıflar arası dağılımın sınıf içi dağılıma oranı hesaplanır. Bu kriteri maksimum yaparak dönüştürülmüş uzayın eksenlerini elde edip Eşitlik 7'de verilmiştir.

$$\mu_3 = p_1 \mu_1 + p_2 \mu_2 \quad (3)$$

$$cov_j = (x_j - \mu_j)(x_j - \mu_j)^T \quad (4)$$

$$S_w = \sum_j P_j x(cov_j) \quad (5)$$

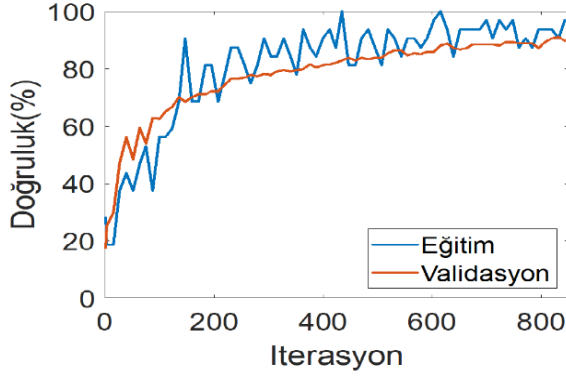
$$S_b = \sum_j (\mu_j - \mu_3) x (\mu_j - \mu_3)^T \quad (6)$$

$$kriter_{opt} = inv(S_w) x S_b \quad (7)$$

DDA kullanılarak dönüşüm sağlanan yeni uzayda hesaplanan Euclidian mesafesi ile her bir sınıf elemanı sınıflandırılabilir. Şekil 4'te ikili sınıflandırmada verilerin saçılım ve varyansa göre ayrıştırılması gösterimi verilmiştir.

3. Bulgular

CovCT Bulguları Veri Seti kullanılarak yapılan çalışmada COVID-19 hastalığının akciğerlerdeki seyri beş farklı sınıfta (GGO, GGON, GOOC, CPP ve Konsolidasyon) sınıflandırılmıştır. Çalışmanın ilk adımında CovCT Bulguları Veri Seti ile eğitilen VGG-16 modeline ait doğruluk grafiği ve eğitim sonrasında sınıflara ait



performans metriklerinin elde edildiği karmaşıklık matrisi Şekil 6'da verilmiştir. CovCT Bulguları Veri Seti VGG-16 modeli ile eğitilip modelin son katmanında mevcut olan Softmax sınıflandırıcısıyla sınıflandırıldığında her bir sınıfa ait duyarlılık, özgüllük, hassasiyet, F-skor ve doğruluk değerleri Çizelge 2'de verilmiştir.

		Tahmin Etiket					Gerçek Etiket
		CPP	Consolidation	GGO	GGOC	GGON	
Gerçek Etiket	CPP	14	3				
	Consolidation		41	1	1		
	GGO			95			6
	GGOC		3		12		
	GGON			4			55

Şekil 6. VGG-16 modelinin CovCT Bulguları Veri Seti eğitim validasyon grafiği ve ilgili karmaşıklık matrisi.

Çizelge 2. CovCT Bulguları Veri Setinin VGG-16 ile eğitiminden elde edilen performans metrik değerleri.

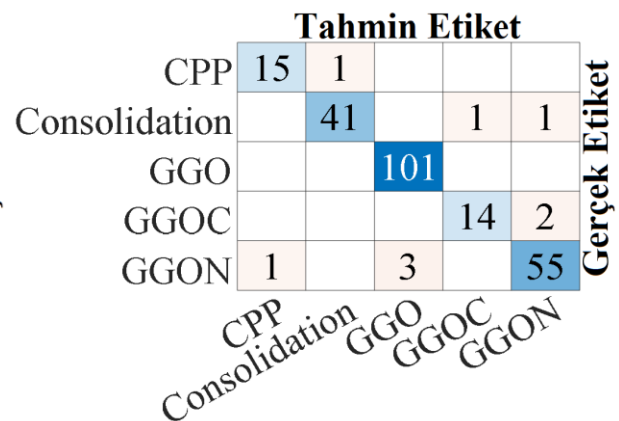
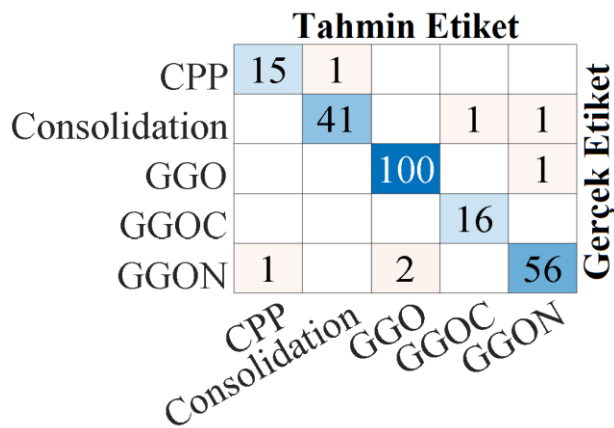
Öznitelik Sayısı	Sınıflandırıcı	Sınıflar	Duyarlılık (%)	Özgüllük (%)	Hassasiyet (%)	F-Skor (%)	Doğruluk (%)	Genel Doğruluk (%)
1000	Softmax	CCP	82.35	100	100	90.32	98.72	92.34
		Konsolidasyon	95.35	96.88	87.23	91.11	96.60	
		GGO	94.06	96.18	95	94.53	95.26	
		GGOC	80	99.55	92.31	85.71	98.30	
		GGON	93.22	96.59	90.16	91.67	95.74	

VGG-16 modeli fc8 katmanından her bir görüntü için elde edilen öznitelikler tam bağlantılı katmana ileterek yapılan eğitim sonunda elde edilen genel doğruluk değeri %92,34 olarak elde edilmiştir. Çalışmanın ikinci adımında VGG-16 modelinin fc8 katmanından elde edilen özniteliklerden sistem performansına katkısı az olan ilgisiz öznitelikleri belirleyip en verimli öznitelikleri seçmek için üçüncü adımda Boruta öznitelik seçme algoritması rastgele orman ağaçlarıyla birlikte kullanılmış olup 100 karar ağacı ve 100 iterasyon değerleri seçilmiştir. Optimizasyon sonucunda toplam 1000 öznitelikten 473 adeti algoritma tarafından verimli ve faydalı olduğu tespit edilmiştir. Elde edilen 473 öznitelik verimlilik önem sırasına göre

sıralandıktan sonra en verimli 200, 300 ve 400 adeti seçilerek ayrı ayrı DVM ve DDA sınıflandırıcılarıyla sınıflandırılmıştır. Önerilen yaklaşımın performans metrik değerleri ayrıntılı olarak Çizelge 3'te verilmiştir. Yapılan çalışma sonrasında toplam 1000 öznitelik kullanılarak elde edilen %92,34 genel doğruluk oranı, Boruta seçim algoritmasıyla seçilen en değerli 300 öznitelik kullanılarak DVM sınıflandırıcısıyla yapılan çalışma sonrasında %97.02'ye yükseltilmiştir. En başarılı sonuçların elde edildiği 300 özniteliğin DVM ve DDA sınıflandırıcıları kullanılarak elde edilen karışıklık matrisleri Şekil 7'de verilmiştir.

Çizelge 3. Boruta Algoritmasıyla seçilen özneliklerin DVM ve LDA ile sınıflandırılması.

Öznitelik Sayısı	Sınıflandırıcı	Sınıflar	Duyarlılık (%)	Özgüllük (%)	Hassasiyet (%)	F-Skor (%)	Doğruluk (%)	Genel Doğruluk (%)
200	Destek Vektör Makineleri	CCP	100	98.6	84.21	91.43	98.72	94.04
		Konsolidasyon	88.64	100	100	93.98	97.87	
		GGO	98.02	94.78	93.40	95.65	96.17	
		GGOC	80	100	100	88.89	98.72	
		GGON	93.32	97.73	93.22	93.22	96.60	
	Doğrusal Diskriminant Analizi	CCP	87.50	98.17	77.78	82.35	97.45	92.34
		Konsolidasyon	88.64	97.91	90.7	89.66	96.17	
		GGO	95.05	96.27	95.05	95.05	95.74	
		GGOC	93.33	99.55	93.33	93.33	99.15	
		GGON	91.53	97.73	93.1	92.31	96.17	
300	Destek Vektör Makineleri	CCP	93.75	99.54	93.75	93.75	99.15	97.02
		Konsolidasyon	95.35	99.48	97.62	96.47	98.72	
		GGO	99.01	98.51	98.04	98.52	98.72	
		GGOC	100	99.54	94.12	99.97	99.57	
		GGON	94.92	98.86	96.55	95.73	97.87	
	Doğrusal Diskriminant Analizi	CCP	93.75	99.54	93.75	93.75	99.15	96.17
		Konsolidasyon	95.35	99.48	97.62	96.47	98.72	
		GGO	100	97.76	97.12	98.54	98.72	
		GGOC	87.5	99.54	93.33	90.32	98.72	
		GGON	93.22	98.3	94.83	94.02	97.02	
400	Destek Vektör Makineleri	CCP	100	99.54	94.12	96.97	99.57	96.59
		Konsolidasyon	95.35	100	100	97.62	99.15	
		GGO	100	94.74	93.58	96.68	97.02	
		GGOC	93.33	100	100	96.55	99.57	
		GGON	91.53	100	100	95.58	97.87	
	Doğrusal Diskriminant Analizi	CCP	100	99.54	94.12	96.97	99.57	95.74
		Konsolidasyon	95.35	100	100	97.62	99.15	
		GGO	99.02	93.98	92.66	95.73	96.17	
		GGOC	93.33	100	100	96.55	99.57	
		GGON	89.83	99.43	98.15	93.81	97.02	



Şekil 7. Boruta öznelik seçim algoritmasıyla seçilen 300 özneliğin DVM ve LDA sınıflandırması sonucunda elde edilen karmaşıklık matrisleri.

4. Tartışma

Covid-19 hastalığının akciğerlerdeki seyrinin hastalığın prognozu ile ilgili bilgiler içerdiği ve bu bilgilerin Covid-19 ve benzer pandemilerde ilgili hekimler için yol gösterici olduğu ayrıca hasta sağlığı açısından etkin kullanımında ise önemli bir rol oynayabileceği konuyla ilgili birçok tıp makalesinde ortaya konmuştur (Cau *et al.*, 2021). Ancak bildiğimiz ve araştırdığımız kadarıyla derin öğrenme çalışmalarında kullanılmak üzere Covid-19 Akciğer BT Bulgularının incelendiği bir veri seti, CovCT Bulguları Veri Seti dışında henüz sunulmadığından bu çalışmanın sonuçları bu anlamda değerlidir. Bu çalışma ile CovCT Bulguları Veri Seti kullanılarak Covid-19 hastalığının akciğerlerdeki seyri beş farklı sınıfta (GGO, GGON, GOOC, CPP ve Konsolidasyon) sınıflandırılmış böylece hem hastalığın prognozunun tahmin edilmesi hem de gerektiğinde hastanın yoğun bakım ve/veya ventilasyon sürecinin başlatılmasına karar verilmesinde ilgili hekimlere yardımcı bir uygulama olarak önerilmiştir. CovCT Bulguları Veri Setini kullanarak Toğaçar vd., (2022) tarafından yapılan önceki çalışmada ResNet modelleri özellik çıkarıcı olarak kullanılmış ve SoftMax'tan önce her modelden baskın aktivasyonları seçen bir yaklaşım önerilmiştir. 5 sınıflı bir sınıflandırma problemi için %99,15'lik bir genel doğruluk elde etmişlerdir. Toğaçar *et al.*, (2022) tarafından önerilen yaklaşımda özneteliklerin üç farklı ResNet modelinden alınmasında yaşanan manuel işlem fazlalığı göz önünde bulundurularak birçok Covid-19 derin öğrenme çalışmasında kullanılan VGG-16 modeli kullanılmıştır. Önceki çalışmada seçilen özneteliklerden baskın olanların seçilme işlemi bu

çalışmada ise seçilen özneteliklerden daha verimli olanlar kullanılarak en başarılı sınıflandırma doğruluğu elde edilmiştir. Toğaçar *et al.*, (2022) tarafından önerilen yaklaşımda ResNet modellerinden elde edilen en düşük sınıflandırma genel doğruluğu olan %96,80 önerilen yaklaşımla yaklaşık %2,5'lik bir artış sağlanarak %99,15'e yükseltilmiştir. Önerilen yaklaşımda VGG-16 modelinden elde edilen %92,34 genel doğruluk oranı ise önerilen yaklaşımla 97,02'ye yükseltilerek yaklaşık %5 artış sağlanmıştır. Dolayısıyla Boruta öznetelik seçim algoritmasıyla elde edilen en etkin özneteliklerle birbirinden farklı modellerin ve uygulamaların birlikte kullanılmasına gerek kalmadan başarılı bir sonuç elde edilmiştir. Her ne kadar Covid-19 akciğer bulguları ile ilgili sınıflandırma çalışması henüz bu çalışmalar dışında tespit edebildiğimiz kadarıyla yapılmamış olsa da farklı akciğer hastalıklarının BT bulgularıyla ilgili nadir de olsa bazı çalışmalar ile karşılaşılmıştır. Hashimoto *et al.* (2018) diffüz akciğer hastalığının BT bulgularını ayırt etmek için konsolidasyon, GGO ve normal opasiteli dokudan oluşan bir BT veri seti oluşturmuşlardır. BT görüntülerini sınıflandırmak için iki konvolüsyonel sinir ağını birliktelik yöntemiyle kullanarak %93,3 sınıflandırma doğruluğu elde etmişlerdir. Gauss filtre kullanılarak opaklıkların sınırlarının yumuşatıldığı çalışmada toplam 50 görüntü dilimi kullanılmıştır. Önceki çalışmalar ve önerilen yaklaşıma ait veriler Çizelge 4'te özetlenmiştir. Bu çalışmada elde edilen başarılı sonuçlara rağmen sistemin manuel işlemlere gerek duyması önerilen yaklaşımın bir dezavantajıdır. Ancak Covid-19 BT Bulgularının önerilen yaklaşımla yüksek doğrulukta tespiti bu alanda önemli bir yenilik olmuştur.

Çizelge 4. Covid-19 Akciğer BT Bulgularıyla ilgili yapılan derin öğrenme çalışmaları.

Önceki Çalışma	Model/Metot	Veri Seti	Eğitim/Validasyon Oranı	Görüntü Kaynağı	Yıl	Genel Doğruluk (%)
Hashimoto <i>et al.</i>	Derin Konvolüsyonel Sinir Ağı, Gauss Filtreleme,	50 GGO 50 Konsolidasyon 50 Normal Opasiteli Doku	80:20	BT	2018	93.3
Togacar <i>et al.</i>	Transfer Öğrenme, Öznetelik Seçimi, LIME	436 GGO 144 Konsolidasyon 198 GGON 50 GGOC 56 CCP 337 GGO	70:30	BT	2022	99.15
Önerilen Yaklaşım	Transfer Öğrenme, Öznetelik Seçimi, Boruta	144 Konsolidasyon 198 GGON 50 GGOC 56 CCP	70:30	BT	2022	97.02

5. Sonuç

Covid-19 hastalığının derin öğrenme modelleriyle tespitiyle ilgili birçok başarılı çalışma yapılmıştır. Hastalığın prognozunun BT bulgularıyla tahmin edilebileceğinin ise alanında birçok tıbbi çalışma ile ortaya konmasına rağmen akciğer BT bulgularının sınıflandırılması ve tespitiyle ilgili çalışmalar CovCT Bulguları Veri Seti oluşturulana kadar henüz yapılamamıştır. Covid-19 Bulgularının tespitinin hastalığın prognozundaki önemi çalışma içerisinde belirtilmiş olup elde edilen sonuçlar transfer öğrenme ve öznetelik seçilmesi yöntemiyle yapılan çalışmaların bu alanda da oldukça başarılı sonuçlar verebileceğini göstermiştir. Önerilen çalışmada oluşturulan CovCT Bulgular Veri Seti, VGG-16 modeli ile eğitildikten sonra Boruta öznetelik seçim algoritması ve DVM sınıflandırıcı kullanılarak en başarılı genel doğruluk başarıları %97,02 olarak elde edilmiştir. Yapılan çalışma ile derin öğrenme yapılarıyla Covid-19 tespiti yapıldıktan sonra hastalığın akciğerlerdeki tutulumunun seyrinin de başarılı tespit edileceği gösterilmiştir. Elde edilen başarılı sonuçlar ile önerilen çalışmanın bilgisayar destekli yazılım olarak radyoloji departmanlarında yardımcı bir uygulama olarak önerilebileceğini böylece Covid-19 ve benzer pandemilerde sağlık tesislerinde hastalığın hızlı tespiti ve etkin yönetimi için kullanılabileceğini de ayrıca göstermiştir.

6. Kaynaklar

- Ai, T., Yang, Z., Hou, H., Zhan, C., Chen, C., Lv, W., ... Xia, L., 2020. Correlation of Chest CT and RT-PCR Testing for Coronavirus Disease 2019 (COVID-19) in China: A Report of 1014 Cases. *Radiology*, **296**(2), E32–E40. <https://doi.org/10.1148/radiol.2020200642>
- Akbaş, İ., 2021. Emergency departments in the detection of COVID-19 cases; multi-centered data from Turkey. *Haydarpaşa Numune Training and Research Hospital Medical Journal*, **61**(3), 314–324. <https://doi.org/10.14744/hnhj.2021.26121>
- Akdaş Tekin, E., Meke, A., Küçükkepeci, H., Önel, S. D., Şimşek, F., Arıca, S., & Turgut, N., 2021. Prediction of Clinical Results with the First Thoracic CT Findings in COVID-2019 Patients; Survey Study. *European Archives of Medical Research*, **37**(4), 268–272. <https://doi.org/10.4274/eamr.galenos.2021.87004>
- Altıntaş, D. D., & Şenol, A., 2021. Hastaneye Yatırılan COVID -19 H astalarında A akciğer B ilgisayarlı Tomografi Parankimal B ulguları ile C Reaktif Protein A rasındaki ilişki The

Relationship Between Parenchymal Findings of Chest Computed Tomography and C Reactive Protein in COVID-19 Pati. *Kocaeli Medical Journal*, **10**(2), 160–166.

- BATIREL, A., 2020. SARS-CoV-2: Ways of Transmission and Methods of Prevention. *Southern Clinics of Istanbul Eurasia*, **31**, 1–7. <https://doi.org/10.14744/scie.2020.00378>
- Cau, R., Falaschi, Z., Paschè, A., Danna, P., Arioli, R., Arru, C. D., ... Saba, L., 2021. CT findings of COVID-19 pneumonia in ICU-patients. *Journal of Public Health Research*, **10**, 515–521. <https://doi.org/10.4081/jphr.2021.2270>
- COMERT, S. S., 2020. Radiological findings of COVID-19 pneumonia. *Southern Clinics of Istanbul Eurasia*, **31**, 16–22. <https://doi.org/10.14744/scie.2020.96158>
- Cortes, C., Vapnik, V., & Saitta, L., 1995. Support-Vector Networks Editor. In *Machine Learning* (Vol. 20). Kluwer Academic Publishers.
- Çelik, D., & Köse, Ş., 2020. COVID-19 in Adults: Clinical Findings. *The Journal of Tepecik Education and Research Hospital*, **30**, 43–48. <https://doi.org/10.5222/terh.2020.88896>
- Ding, X., Xu, J., Zhou, J., & Long, Q., 2020. Chest CT findings of COVID-19 pneumonia by duration of symptoms. *European Journal of Radiology*, (January), 127.
- Fang, Y., Zhang, H., Xie, J., Lin, M., Ying, L., Pang, P., & Ji, W., 2020. Sensitivity of chest CT for COVID-19: Comparison to RT-PCR. *Radiology*, Vol. **296**, E115–E117. Radiological Society of North America Inc. <https://doi.org/10.1148/radiol.2020200432>
- Gayaf, M., Anar, C., Güldaval, F., Karadeniz, G., Polat, G., Ayrancı, A., ... Tatar, D., 2021. Clinical Characteristics and Transmission Routes of COVID-19 in the Early Period of the Pandemic in a Non-Covid Ward of Chest Diseases Hospital. *Journal of İzmir Chest Hospital*, **35**(3), 140–148. <https://doi.org/10.5222/igh.2021.24633>
- Hashimoto N, Suzuki K, Liu J, et al., 2018. Deep neural network convolution (NNC) for three-class classification of diffuse lung disease opacities in high-resolution CT (HRCT): Consolidation, ground-glass opacity (GGO), and normal opacity. *Med Imaging* 2018 **10575**,113. <https://doi.org/10.1117/12.2293550>
- Hatipoğlu, N., 2020. The “New” Problem of Humanity: New Coronavirus (2019-nCoV / COVID-19) Disease. *Medical Journal of Bakirkoy*, **16**(1), 1–8. <https://doi.org/10.5222/BMJ.2020.22931>
- Karahacıoğlu, D., Önel, S. D., Bayraktarlı, R. Y., & Şimşek, F. (2022). COVID-19 Pneumonia: Variation of Chest Computed Tomographic Findings at Different Phases of

- Disease. *European Archives of Medical Research*, **38(1)**, 61–66. <https://doi.org/10.4274/eamr.galenos.2022.27676>
- Kauczor, H. U., Heitmann, K., Heussel, C. P., Marwede, D., Uthmann, T., & Thelen, M., 2000. Automatic detection and quantification of ground-glass opacities on high-resolution CT using multiple neural networks: Comparison with a density mask. *American Journal of Roentgenology*, **175(5)**, 1329–1334. <https://doi.org/10.2214/ajr.175.5.1751329>
- Kiral, N., 2021. The effect of frequency of comorbidity on the severity and prognosis of hospitalized patients with SARS-Cov-2 infection. *Southern Clinics of Istanbul Eurasia*, **32** (March 2020), 245–252. <https://doi.org/10.14744/scie.2021.35467>
- Kursa, M. B., Jankowski, A., & Rudnicki, W. R., 2010. Boruta - A system for feature selection. *Fundamenta Informaticae*, **101(4)**, 271–285. <https://doi.org/10.3233/FI-2010-288>
- Li, K., Wu, J., Wu, F., Guo, D., Chen, L., Fang, Z., & Li, C., 2020. The Clinical and Chest CT Features Associated with Severe and Critical COVID-19 Pneumonia. *Investigative Radiology*, **55(6)**, 327–331. <https://doi.org/10.1097/RLI.0000000000000672>
- OZER, K. B., 2020. The Effect of Radiological and Laboratory Parameters on Prognosis in COVID 19 Disease. *Southern Clinics of Istanbul Eurasia*, **31(3)**, 203–207. <https://doi.org/10.14744/scie.2020.87609>
- Özdemir, Ö., 2021. Pathogenesis of Imaging in COVID-19 (narrative review). *Southern Clinics of Istanbul Eurasia*, **33(1)**, 92–97. <https://doi.org/10.14744/scie.2021.97658>
- Park, C. H., & Park, H., 2008. A comparison of generalized linear discriminant analysis algorithms. *Pattern Recognition*, **41(3)**, 1083–1097. <https://doi.org/10.1016/j.patcog.2007.07.022>
- Pekçevik, Y., & Belet, Ü., 2020. Patient Management in the Radiology Department, the Role of Chest Imaging During the SARS-CoV-2 Pandemic and Chest CT Findings Related to COVID-19 Pneumonia. *The Journal of Tepecik Education and Research Hospital*, **30**, 195–212. <https://doi.org/10.5222/terh.2020.13549>
- Quiroz, J. C., Feng, Y. Z., Cheng, Z. Y., Rezazadegan, D., Chen, P. K., Lin, Q. T., ... Cai, X. R., 2021. Development and validation of a machine learning approach for automated severity assessment of COVID-19 based on clinical and imaging data: Retrospective study. *JMIR Medical Informatics*, **9(2)**, 1–14. <https://doi.org/10.2196/24572>
- Tekcan Sanli, D. E., Yildirim, D., Sanli, A. N., Erozan, N., Husmen, G., Altundag, A., ... Erel Kirisoglu, C., 2021. Predictive value of CT imaging findings in COVID-19 pneumonia at the time of first-screen regarding the need for hospitalization or intensive care unit. *Diagnostic and Interventional Radiology*, **27(5)**, 599–606. <https://doi.org/10.5152/dir.2020.20421>
- Toğaçar, M., Ergen, B., & Cömert, Z., 2020. COVID-19 detection using deep learning models to exploit Social Mimic Optimization and structured chest X-ray images using fuzzy color and stacking approaches. *Computers in Biology and Medicine*, **121**. <https://doi.org/10.1016/j.combiomed.2020.103805>
- Toğaçar, M., Muzoğlu, N., Ergen, B., Yarman, B. S. B., & Halefoğlu, A. M., 2022. Detection of COVID-19 findings by the local interpretable model-agnostic explanations method of types-based activations extracted from CNNs. *Biomedical Signal Processing and Control*, **71**. <https://doi.org/10.1016/j.bspc.2021.103128>
- Togay, A., & Yılmaz, N., 2020. Laboratory Diagnosis of SARS-CoV-2. *The Journal of Tepecik Education and Research Hospital*, **30**, 70–75. <https://doi.org/10.5222/terh.2020.13007>
- Türken, M., & Köse, Ş., 2020. COVID-19 Transmission and Prevention. *The Journal of Tepecik Education and Research Hospital*, **30**, 36–42. <https://doi.org/10.5222/terh.2020.02693>
- Yang, X., He, X., Zhao, J., Zhang, Y., Zhang, S., & Xie, P., 2020. COVID-CT-Dataset: A CT Scan Dataset about COVID-19. Retrieved from <http://arxiv.org/abs/2003.13865>

İnternet kaynakları

1 -<https://covid19.who.int/>, (04.05.2022)