



Derin Öğrenme ve Sınıflandırma Yaklaşımları ile BT görüntülerinden Covid-19 Tespiti

Covid-19 Detection from CT images with Deep Learning and Classification Approaches

Erdal Özbay^{1*}, Feyza Altunbey Özbay²

¹ Fırat Üniversitesi, Bilgisayar Mühendisliği Bölümü, Elazığ, erdalozbay@firat.edu.tr, 0000-0002-9004-4802

² Fırat Üniversitesi, Yazılım Mühendisliği Bölümü, Elazığ, faltunbey@firat.edu.tr, 0000-0003-0629-6888

MAKALE BİLGİLERİ

Makale geçmişi:

Geliş: 19 Ekim 2020
Düzeltilme: 12 Kasım 2020
Kabul: 13 Kasım 2020

Anahtar kelimeler:

Covid-19, Makine öğrenmesi, Derin öğrenme, Evrişimli Sinir Ağı, Göğüs BT görüntüleri

ÖZET

İlk olarak Çin ülkesinin Wuhan eyaletinde, Aralık 2019 tarihinde görülen ve oldukça bulaşıcı bir hastalık olan yeni tip Koronavirüs (Covid-19), sadece birkaç ay içerisinde tüm dünyaya yayılmış ve bir pandemi haline gelmiştir. Covid-19, dünya ekonomik yapısını, insanların dini, siyasi, sosyal yaşamını, halk sağlığı yapısını, insanların günlük yaşam yapısını değiştirmiş ve milyonlarca insanı işsiz bırakmıştır. Bu salgınla mücadele etmenin öncelikli yolu, enfekte olan kişinin mümkün olan en kısa sürede teşhis edilmesi ve onun sağlıklı bireylerden uzaklaştırılmasıdır. Şu anda, dünya çapında Covid-19 hastalarını tespit etmek için Ters Transkripsiyon-Polimeraz Zincir Reaksiyonu (TT-PZR) kullanılmaktadır. Ancak Dünya Sağlık Örgütü'nce (DSÖ), TT-PZR'nin erken evre vakalarının tespitinde düşük duyarlılık ve düşük özgüllükten muzdarip olduğu vurgulanmıştır. Son araştırmalar göstermiştir ki, göğüs Bilgisayarlı Tomografi (BT) taraması görüntüleri, Covid-19 vakalarını belirlemede yararlı bir rol oynamaktadır. Bu çalışmada, Covid-19 vakalarının sınıflandırma sonuçlarına dayalı tahmin modeli için son teknolojik gelişmelere uygun birçok sınıflandırma algoritmaları ile birlikte Evrişimli Sinir Ağı (ESA) performansları karşılaştırılmıştır. Sonuç, önerilen ESA modelinin, diğer gelişmiş sınıflandırma algoritmalarından daha iyi performans gösterdiği ve %98.1 doğruluk elde ettiği vurgulanmıştır.

Doi: 10.24012/dumf.812810

ARTICLE INFO

Article history:

Received: 19 October 2020
Revised: 12 November 2020
Accepted: 13 November 2020

Keywords:

Covid-19, Machine learning, Deep learning, Convolutional Neural Network, Chest CT images

ABSTRACT

The new type of Coronavirus (Covid-19), which was first seen in Wuhan province of the Chinese country in December 2019 and was a highly contagious disease, spread all over the world in just a few months and became a pandemic. Covid-19 has changed the world economic structure, people's religious, political, social life, public health structure, people's daily life structure and left millions of people unemployed. The primary way to combat this epidemic is to diagnose the infected person as soon as possible and remove him from healthy individuals. Currently, Reverse Transcription-Polymerase Chain Reaction (RT-PCR) is used to detect Covid-19 patients worldwide. However, it has been emphasized by the World Health Organization (WHO) that RT-PCR suffers from low sensitivity and low specificity in the detection of early stage cases. Recent research has shown that chest Computed Tomography (CT) scan images play a useful role in identifying Covid-19 cases. In this study, Convolutional Neural Network (CNN) performances were compared with many classification algorithms suitable for the latest technological developments for the prediction model based on the classification results of Covid-19 cases. As a result, it was emphasized that the proposed CNN model performs better than other advanced classification algorithms and achieves 98.1% accuracy.

* Sorumlu yazar / Correspondence

Erdal ÖZBAY

✉ erdalozbay@firat.edu.tr

Giriş

Son bir yılda tüm dünyayı etkisi altına alan yeni tip Koronavirüs (Covid-19) hastalığı 2019 Aralık ayında Çin ülkesinin Wuhan kenti merkezli olarak ortaya çıkmış ve kısa bir zaman içerisinde de Çin'in her yerine yayılmıştır. Çok kısa bir sürede ise Çin dışına çıkarak tüm dünyada yayılmış ve dünya pandemisi haline gelmiştir. 17 Ekim 2020 itibariyle dünya çapında enfekte vaka sayısı 39,644,861 ve ölüm vakası 1,110,182'dur [1]. Bu rakamlar, tüm dünyayı tehdit altına alan Covid-19'un ciddiye alınması gerektiğini göstermektedir. Bu bakımdan, Covid-19 vakaları erken tespit edilebilirse, bu hastalar tecrit edilebilir, böylelikle enfekte olmayan sağlıklı bireyler güvende kalabilir. Şu anda, Covid-19 hastalarını teşhis etmenin küresel yöntemi, ters transkripsiyon-polimeraz zincir reaksiyonudur (TT-PZR). Ancak bu yöntemin birincil sorunu, düşük duyarlılık ve özgüllükten muzdarip olmasıdır [2]. Bunun yanında, uzak kırsal bölgelerde TT-PZR test kitlerinin azlığı nedeniyle, doktorlar Covid-19 taraması için tıbbi görüntülerin kullanılmasını önermektedir [3]. Bilgisayarlı tomografi (BT) tarama görüntüsü, pozitif Covid-19 hastalarının tespit edilmesi için önemli ayrıntılar taşımaktadır [4]. BT tarama görüntüsünün faydalarına rağmen, Covid-19 ve diğer akciğer hastalıkları arasında benzer özellikler bulunmaktadır. Bu bakımdan taramanın yürütülmesi oldukça zordur. Son zamanlarda, radyolojik görüntüler üzerinden, makine öğrenmesi ve derin öğrenme teknikleri kullanılarak bazı özelliklerin çıkarılması ve tespit edilmesi kullanışlı hale gelmiştir. Bu çalışmada, Covid-19 pozitif hastalarını, BT tarama görüntülerinden tespit etmek için birkaç makine öğrenimi ve derin öğrenme tekniği kullanılmaktadır. Veri seti, 1252 adet Covid-19 pozitif yani hasta, 1230 adet Covid-19 negatif yani sağlıklı olmak üzere toplam 2482 görüntüden oluşan nispeten büyük bir veri kümesinden oluşmaktadır. Sınıflandırma için bilinen en iyi sınıflandırıcı algoritmalarından Rastgele Orman, Destek Vektör Makinesi, Rastgele Ağaç ve Naive Bayes ile Evrişimli Sinir Ağı modelinin sonuçları karşılaştırılmış ve Covid-19 hastalarını en iyi doğrulayabilecek model belirlenmiştir. Makalenin geri kalan bölümü şu şekilde düzenlenmiştir; Bölüm 2, ilgili

son çalışmaları ele almaktadır. Bölüm 3, bu çalışma için kullanılan veri setini ve yöntemleri tartışmaktadır. 4. Bölümde deneysel sonuçlar analiz edilmektedir ve önerilen Covid-19 tahmin modelinin deneysel sonuçlarının tartışılması ele alınmaktadır.

İlgili Çalışmalar

Pandeminin ortaya çıkışından itibaren, Covid-19 hastalığının tespiti için otomatik tarama sistemi, araştırma topluluğu için en önemli öncelik haline gelmiştir. Akciğerin BT taramalarının sınıflandırılması için otomatik bir sistemin geliştirilmesi, görsel bir incelemede bulaşıcı ve enflamatuar akciğer hastalıklarının teşhisinin karmaşıklığı nedeniyle zorlu olmaya devam etmektedir. Görsel muayene kabul edilebilir bir standart olmasına rağmen, teşhis edilmesi gereken çok sayıda hastadan kaynaklanan hatalara maruz kalma eğilimindedir. Bu bakımdan araştırmacılar, çeşitli yöntemlerle Covid-19'un benzersiz özelliklerini otomatik olarak belirlemek için birçok çalışma önermişlerdir.

Kang vd., Covid-19 için otomatik olarak teşhis koyabilen çok görüntülü bir temsil öğrenme tekniği önermişlerdir [5]. Önerdikleri modeli, doğrulamak için 2522 BT tarama görüntüsüne uygulamışlardır. Uyguladıkları yöntemle, sırasıyla %95.5, %96.6 ve %93.2 doğruluk, duyarlılık ve özgüllüğe ulaşmışlardır. Li vd., göğüs BT'leriyle Covid-19'u doğru bir şekilde tanımlamak için bir derin öğrenme otomatik çerçevesi olan COVNet'i önermişlerdir [6]. Modellerini oluştururken 4356 görüntüden oluşan göğüs BT'si kullanmışlardır. Bu modelle, diğer pnömoni hastalarından Covid-19 hastalarını tespit etmede % 87'lik bir duyarlılık ve 0.95'lik bir Eğri Altındaki Alan (EAA) değeri elde edilmiştir. Xu vd., Covid-19'un erken taranması için ResNet adlı bir derin öğrenme modeli tasarlamışlardır [7]. Model oluşturulurken toplam 618 pulmoner BT örneği kullanılmıştır. Bu çalışmada, Covid-19'u influenza-A pnömonisinden ve sağlıklı vakalardan ayıran % 86,7'lik nihai bir doğruluk elde edilmiştir. Ardakani vd., Covid-19'u ayırt etmek için on adet evrişimli sinir ağını, yani VGG-16, VGG-19, AlexNet, GoogleNet,

SqueezeNet, ResNet-18, ResNet-50, ResNet-101, MobileNet-V2 ve Xception'ı 1020 adet BT görüntüsü kullanarak diğer pnömonilerle (yani Covid-19 olmayan) karşılaştırmışlardır [8]. Bu bakımdan, ResNet 101 ve Xception'ın en yüksek EAA değeri olan 0.994'ü elde ettiğini gözlemlemişlerdir ve Covid-19 hastalarını karakterize etmek ve saptamak için Resnet 101'i önermişlerdir. Bir diğer çalışmada, Bai X. vd., derin bir sinir ağı mimarisi olan EfficientNet'i ortaya koymuş ve 1186 hastadan elde edilmiş BT karelerini bu mimariye uygulamışlardır [9]. Üretilen sistemde, Covid-19 ve Covid-19 olmayanlar arasında ayırım yapılırken, piyasaya sürülen sistemde %96 doğruluk, %95 duyarlılık ve %96 özgüllük elde edilmiştir. Shi vd., Covid-19'u taramak için bir makine öğrenimi algoritması olan Rastgele Orman'ı (RO) uygulamışlardır [10]. Bu çalışma için sundukları modellerini değerlendirmek adına 2685 hastanın BT görüntülerini kullanmışlardır. Modelde, 5 kat çapraz doğrulama tekniği değerlendirildikten sonra, model sırasıyla %87.9, %90.7 ve %83.3 doğruluk, duyarlılık ve özgüllüğe ulaşmıştır. Bir başka çalışmada, Özkaya vd., 150 adet BT görüntüsünden 3000 adet yama görüntü üretmiş ve bu görüntüler üzerinden daha fazla sıralama ve füzyon teknikleri uygulamışlardır [11]. Sınıflandırma için Destek Vektör Makinesi (DVM) kullanılmıştır ve bundan önce, transfer öğrenme yönteminin bir parçası olarak da önceden eğitilmiş olan bir ESA modeli kullanılmıştır. Sunulan prosedürde %98,27 doğruluk, %97,63 kesinlik ve %97,6 duyarlılık elde edilmiştir. Alom vd., [12], Covid-19'u tespit etmek için verimli bir derin öğrenme yaklaşımı olan transfer öğrenmeli, Bölgesel-Evrişimli Sinir Ağı'nı (B-ESA) sunmuşlardır. Araştırmacılar daha önceleri, sınıflandırmanın sonucunu artırabilmek için enfekte olan alan bölümlendirme için NABLA-N ağını kullanmışlardır. Önerilen yöntemin değerlendirilebilmesi için hem X-ışını hem de BT görüntüleri üzerinde çalışılmıştır. X-ışını ve BT görüntülerinden sırasıyla %84.67 ve %98.78 doğruluk değerleri elde etmişlerdir.

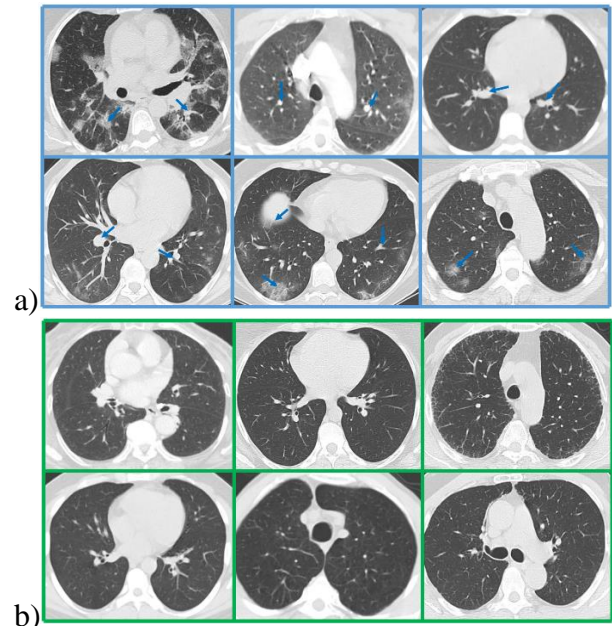
Mevcut sınıflandırma modelleri, özellik çıkarım karmaşıklığı açısından bazı sınırlamalar göstermektedir. Çeşitli özellik çıkarım algoritmaları, görüntü piksellerinin uzamsal

dağılımındaki önemli değişikliklerini yakalamada önemli bir role sahiptir. Son zamanlarda, kesirli analiz ve uygulamaları farklı uygulama alanlarında kullanılmıştır [13]. Bu çalışmada, çalışmanın katkılarında biri olarak kabul edilen, görüntü sınıflandırma görevleri için akciğer BT taraması sınıflandırması için mevcut modeller, özellik çıkarma için yalnızca derin öğrenmeye dayanmaktadır. Bu nedenle, klinik bulgular ile derin öğrenme özelliklerini birleştirmek, Covid-19'lu ve sağlıklı vakalar arasındaki sınıflandırma performansını daha da artırarak hastalık hikayesinin seyrini olumlu biçimde etkileyecektir. Bu çalışmanın motivasyonu, derin öğrenme kullanarak BT taramalarında Covid-19'lu ve sağlıklı akciğerlerin verimli bir sınıflandırmasını önermektir.

Materyal ve Metot

Veri Kümesi

Bu araştırmada, Covid-19 hastalarını sınıflandırmak için, Kaggle adlı platform üzerinden halka açık şekilde paylaşılmış, BT tarama görüntülerinden oluşan bir veri seti kullanılmıştır [14]. Bu veri kümesinde Brezilya, Sao Paulo'dan toplanan 2482 göğüs BT görüntüsü bulunmaktadır.

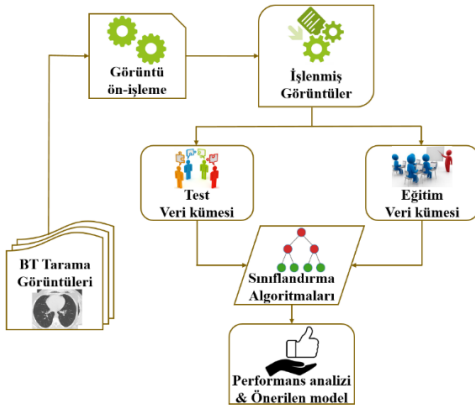


Şekil 1. Covid-19 vakaların BT tarama görüntü örnekleri, a) pozitif-mavi çerçeveli (mavi oklar kontamine bölgeyi göstermektedir), b) negatif-yeşil çerçeveli

Veri kümesi içerisinde, 1252 tane pozitif Covid-19 vakalı göğüs BT taraması görüntüsü ve 1230 negatif Covid-19 vakası anlamına gelen başka akciğer hastalıkları olan göğüs BT taraması görüntüleri bulunmaktadır. Bu veri kümesindeki göğüs BT tarama görüntülerinden bir kesit şekil 1’de gösterilmiştir.

Görüntü İşleme

Veri kümesindeki görüntülerin tümü farklı boyutlardadır. Veri kümesindeki tüm görüntüleri aynı boyuta getirmek için Python Open CV kullanılarak, yeniden boyutlandırma işlemi yapılmıştır. Tüm görüntüler tek bir boyuta getirildikten sonra renk uzayı dönüşümü gerçekleştirilmiştir. Bu işlem ile, görüntüler RGB renk uzayından gri renk uzayına dönüştürülmüştür. Ön işleme aşaması, görüntülerin sınıflandırma algoritmalarında kullanılabilmesi için dizilere dönüştürülmesiyle tamamlanmıştır. Araştırma çalışmamızın akış diyagramı şekil 2’de gösterilmektedir.



Şekil 2. Araştırma yöntem bilimimize ait akış diyagramı

Sınıflandırma

Bu çalışmada BT görüntülerini sınıflandırmak için beş adet farklı sınıflandırma algoritması yürütülmüştür.

a. Evrişimsel Sinir Ağları (ESA)

Son yıllarda, derin öğrenme, nesne tanıma, beyin tümörü segmentasyonu ve sınıflandırma, meme kanseri tespiti, rahim ağzı kanseri tanıma gibi tıbbi problemlerin çözümünde oldukça ilgi çekmiştir. ESA, derin öğrenmenin bir parçasıdır ve bilgisayarla görme problemlerine sıklıkla uygulanmaktadır. ESA mimarisi, bir evrişim

katman, bir havuzlama katmanı ve tamamen bağlantılı bir katman olmak üzere yapısında üç katmanın kombinasyonunu barındırmaktadır. İlk iki katman, giriş görüntüsünden derin özellikler çıkarır ve tamamen bağlantılı katman, çıkarılan özellikleri çıktı katmanına eşlemektedir. Böylelikle görüntünün gereksiz kısımlarının yapay sinir ağına gitmesi engellenerek, sistemin hızlı ve doğru sonuç vermesi sağlanmaktadır.

Evrişimli Katman: Evrişimli katman, özellik çıkarma işlemini gerçekleştirmektedir. Doğrusal evrişimli işlem ve doğrusal olmayan etkinleştirme fonksiyonu, evrişim adımını gerçekleştirmek için gereken iki temel işlemdir.

Doğrusal evrişimli süreçte, giriş görüntüsünden öznelikleri çıkarmak için bir öznelik algılayıcı veya çekirdek kullanılmaktadır. Bir özellik haritası veya aktivasyon haritası olarak da bilinen kıvrımlı bir görüntü oluşturmak için giriş tensörü ve çekirdek arasında eleman bazlı ürün operasyonu gerçekleştirilmektedir. Bu evrişimin birincil amacı, giriş görüntüsünün boyutunu azaltmaktır. Matematiksel olarak, aşağıdaki denklem evrişimli işlemi temsil etmektedir:

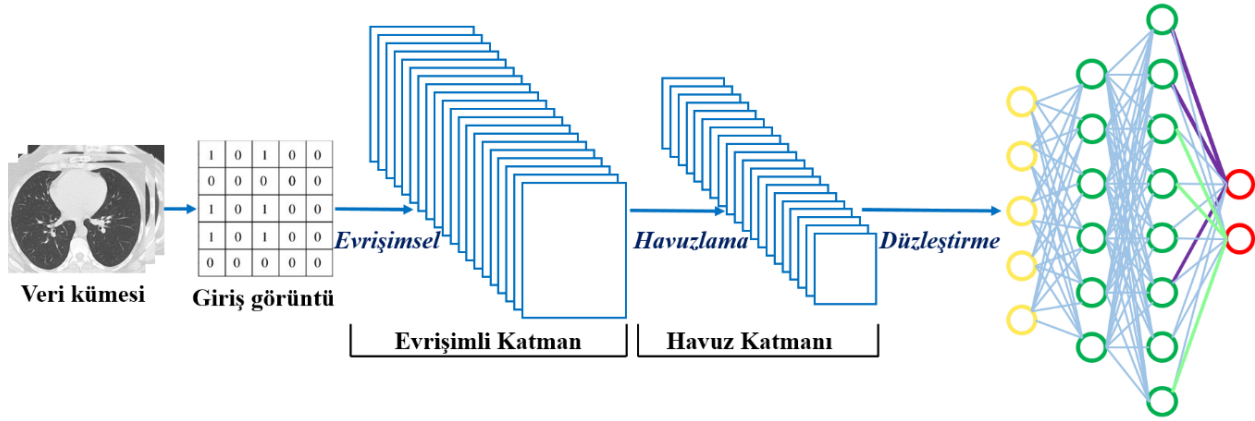
$$(f * g)(t) = \int_{-\infty}^{+\infty} f(T)g(t - T)dT \quad (1)$$

Evrişimden sonra, Doğrultulmuş Doğrusal Birim (Rectified Linear Unit- RELU) doğrusallığı kırmakta ve ağdaki doğrusal olmayışı geliştirmektedir. Fonksiyonun çıktısı aşağıdaki gibi özetlenmiştir:

$$f(x) = \max(0, x) \quad (2)$$

Havuz Katmanı: Havuzlama, havuzlanmış özellik haritası oluşturmak için özellik haritasının boyutunu azaltmak için kıvrılmış görüntüden özellikler çıkarmaktadır. Maksimum havuzlama, ortalama havuzlama gibi farklı havuzlama türleri kullanılmaktadır. Daha sonra, havuzlanmış özellik haritası tek boyutlu bir sütun haline getirilmekte ve ek işlemler için yapay sinir ağına beslenmektedir.

Tamamen Bağlantılı Katman: Bu tam bağlantı, şekil 3’te gösterilen, tamamen bağlı bir katman aracılığıyla bir çıktı katmanına sahip bir giriş katmanından oluşan bir birleşimden oluşmaktadır.



Şekil 3. Evrişimli Sinir Ağının (ESA) mimarisi

Burada, tamamen bağlı katmanlar, tüm düğümler tamamen bağlı olduğundan özel bir gizli katman olarak da bilinmektedir. Bu katmanlarda, RELU bir aktivasyon fonksiyonu olarak kullanılmaktadır.

Son tahmin, çıktı katmanında aktivasyon fonksiyonu olarak softmax veya sigmoid kullanılarak yapılmaktadır. Bilgi bu şekilde yapay sinir ağından geçmektedir. Ağ performansının değerlendirilmesi tahmin hatası veya kayıp fonksiyonunun 0'a yakınlığı ile ölçülmektedir. Ağ optimize etmek için bu işlevin en aza indirilmesi gerekmektedir. Mükemmel bir tahmin elde etmek için, tahmin hatası ağ üzerinden geri yayılmaktadır.

b. Rastgele Orman (RO)

Popüler makine öğrenme modellerinden biri olan denetimli bir sınıflandırma yöntemidir, bu algoritmayı popüler yapan en önemli özelliği, hiper parametre kestirimi yapılmasına gerek duymadan iyi sonuçlar üretmesi ayrıca regresyon ve sınıflandırma problemlerine uygulanabilmesidir. Algoritmanın temel çalışma prensibinde, modelin temel bloğu olan karar ağaçları kullanılmaktadır. En basit şekliyle algoritma rastgele olarak bir orman yaratmaktadır, bu algoritmadaki ağaç sayısı ve elde edebileceği sonuç arasında doğrudan bir ilişki bulunmaktadır. Ağaç sayısı artırdıkça daha kesin bir sonuç elde edilebilmektedir. Rastgele Orman algoritması ile Karar Ağacı algoritması arasındaki en önemli fark, Rastgele Orman yöntemindeki kök düğümün (Root Node) bulunması ve düğümlerin bölünmesi işlemlerinin rastgele çalışıyor olmasıdır. Rastgele Orman

algoritmasında kullanılan model farklı veri setleri üzerinde eğitim gerçekleştirdiği için karar ağaçlarının en büyük problemlerinden olan varyans yani aşırı uyumluluk, diğer bir deyişle overfitting azalmaktadır [15].

c. Destek Vektör Makinesi (DVM)

Destek Vektör Makinesi (DVM), sınıflandırma ve regresyon problemlerini çözmek için kullanılan denetimli bir makine öğrenme algoritmasıdır. Destek vektörü makine modeli, çok boyutlu uzayda bir hiper düzlemdeki birkaç sınıfı temsil etmektedir. Hataları azaltmak için, hiper düzlem algoritma tarafından yinelemeli olarak oluşturulmaktadır. Algoritmanın temel amacı, veri kümelerini farklı sınıflara bölerek maksimum marjinal hiper düzlem elde etmektir. Bunlar iki adımda yapılmaktadır: İlk olarak, seviyeleri en iyi şekilde ayıran hiper düzlemler yinelemeli bir şekilde oluşturulmaktadır. İkinci olarak, sınıfları doğru bir şekilde ayıran alt düzlem seçilmektedir.

d. Rastgele Ağaç (RA)

Rastgele Ağaç, birçok karar ağacından gelen tahminleri birleştiren bir toplu makine öğrenimi algoritmasıdır. Yaygın olarak kullanılan Rastgele Orman algoritması ile ilişkilidir. Grubun üyeleri olarak kullanılan karar ağaçlarını oluşturmak için daha basit bir algoritma kullanılmasına rağmen, genellikle rastgele orman algoritmasından daha iyi performans elde edilebilmektedir. Rastgele ağaç sınıflandırıcı, sınıflandırma ve regresyon için kullanılan, denetimli bir makine öğrenme algoritmasıdır. Rastgele ağaç sınıflandırıcı, eğitim veri setinden çok sayıda budanmamış karar ağacı

oluşturmaktadır. Tahmin yapmak için, her karar ağacının oy verdiği yerlerde çoğunluk oylama tekniğini kullanmaktadır ve en yüksek oylanan tahmin, nihai sınıflandırma sonucu olarak kabul edilmektedir. Her karar ağacı, bir ağaç ormanı oluşturmak için orijinal eğitim setinden oluşturulmaktadır. Her karar ağacı, özellik alt kümesinden verilen her test düğümündeki rastgele k özellik örneğinden verileri ayırmak için en iyi özelliği seçmektedir. Daha sonra, birden fazla ilintisiz karar ağaçları olarak da bilinen fazladan ağaç ormanı, rastgele bir özellik örneği kullanılarak üretilmektedir.

e. Naive Bayes (NB)

Naive Bayes sınıflandırıcıları, Bayes teoreminin özellikleri arasında güçlü (naif) bağımsızlık varsayımlarıyla uygulamaya dayanan basit “olasılıksal sınıflandırıcılar” ailesidir. Naive Bayes sınıflandırıcıları, bir öğrenme problemindeki değişkenlerin (özellikler / öngörücüler) sayısında doğrusal bir dizi ölçeklendirilebilir parametre gerektirmektedir. Maksimum olasılık eğitimi, diğer birçok sınıflandırıcı türü için kullanılan pahalı yinelemeli yaklaşımdan ziyade doğrusal zaman alan kapalı bir form ifadesinin değerlendirilmesiyle yapılabilmektedir. İstatistik ve bilgisayar bilimi literatüründe saf Bayes modelleri, basit Bayes ve bağımsız Bayes gibi çeşitli isimler adı altında bilinmektedir. Ancak tüm bu isimler, sınıflandırıcının karar kuralında Bayes teoreminin kullanımına atıfta bulunmaktadır. En basit Bayes ağ modelleri kernel yoğunluğu tahmini ile birleştirilebilmekte ve daha yüksek doğruluk seviyeleri elde edilebilmektedir.

Performans Değerlendirme Ölçütleri

Bu çalışmada, k -kat çapraz doğrulama tekniği kullanılarak veri kümesi k farklı alt kümeye bölünerek önerilen sistemin verimliliği test edilmiştir. Herbir adımda, $k-1$ alt küme modeli eğitmek için, geri kalanı ise test kümesi olarak kullanılmıştır. Model oluşturma sırasında, bu işlem k kez tekrarlanarak, modelin performansı, bağımsız k alt kümenin test sonuçlarının ortalaması alınarak ölçülmektedir. Bu çalışmada k değeri, 10 olarak belirlenmiştir. Bu çalışmada,

altı (6) farklı değerlendirme ölçütü doğruluk, kesinlik, hassasiyet, F-ölçütü, özgüllük ve ROC eğrisi olarak bilinen Eğri Altındaki Alan (EAA) kullanılmıştır. Tablo 1’de verilen karışıklık matrisi, herhangi bir tahmin modelinin genel performansını göstermek için kullanılmaktadır. Bu karışıklık matrisi kullanılarak, belirlenen altı performans ölçütü hesaplanabilmektedir [16].

Tablo 1. Karışıklık Matrisi

		Tahmin Edilen	
		Pozitif	Negatif
Gerçek	Pozitif	Doğru Pozitif (DP)	Yanlış Negatif (YN)
	Negatif	Yanlış Pozitif (YP)	Doğru Negatif (DN)

$$\text{Doğruluk} = \frac{DP+DN}{DP+YP+DN+YN} \quad (3)$$

$$\text{Kesinlik} = \frac{DP}{DP+YP} \quad (4)$$

$$\text{Duyarlılık} = \frac{DP}{DP+YN} \quad (5)$$

$$F_1 - \text{Ölçütü} = \frac{2*DP}{YN+YP+2*DP} \quad (6)$$

$$\text{Özgüllük} = \frac{DN}{YP+DN} \quad (7)$$

Bu çalışmada EAA eğrisi, Covid-19 pozitif ve negatif iki durumun ne kadar doğru bir şekilde ayırdığını ölçmek için kullanılmaktadır. EAA eğrisi, DP oranının YP oranına göre değişiminin çizilmesiyle elde edilmektedir. Eğrinin Altındaki Alan’ın (EAA) fazlalığı tanı testinin başarısını göstermektedir. EAA, tüm olası sınıflandırma eşiklerinde toplu bir performans ölçüsü sağlamaktadır. EAA’ı yorumlamanın bir diğer şekli, modelin rastgele bir pozitif örneğinin, rastgele bir negatif örnekten daha yüksek bir şekilde sıralanması olasılığıdır.

Deneysel Sonuçlar ve Tartışma

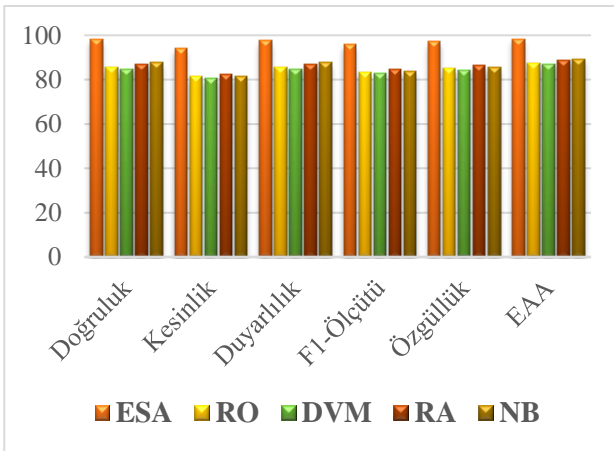
Bu çalışmada, çeşitli uygulamalar geliştirmek için Google tarafından sağlanan Python programlama diline dayalı bir bulut hizmeti olan Google Colab kullanılmıştır [17]. Daha hızlı

işlem için sanal Tensör İşleme Birimi (TİB) kullanılmıştır. Deneysel veri setimiz %75 : %25 oranına bölünmüştür. Bu bakımdan veri kümemizdeki verilerin %75'i eğitim verisi olarak ve %25'i ise test verisi olarak kullanılmıştır. Modelin aşırı uyumlu olmasını önlemek için 10 kat çapraz doğrulama tekniği kullanılmıştır. Bu çalışmada, Covid-19 vakalarını sınıflandırmak için beş farklı sınıflandırma tekniği karşılaştırılmıştır. Deneysel sonuçlardan Covid-19 hastalarının sınıflandırılmasında ESA, RO, DVM, RA ve NB'in sırasıyla %98,1, %85,45, %84,6, %86,93 ve %87,77 doğruluk sağladığı gösterilmiştir. Performansı değerlendirmek için dikkate alınan diğer istatistiksel değerlendirme ölçümleri tablo 2'de gösterilmektedir.

Tablo 2. Sınıflandırıcı algoritmaların performans değerlendirme sonuçları

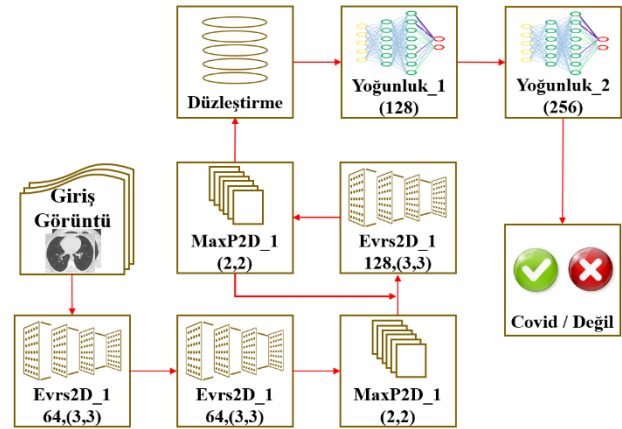
Sınıflandırıcı / Ölçütler	ESA	RO	DVM	RA	NB
Doğruluk	98,10	85,45	84,6	86,93	87,77
Kesinlik	94	81,4	80,73	82,5	81,64
Duyarlılık	97,72	85,63	84,91	87,01	87,81
F1-Ölçütü	95,81	83,46	82,76	84,68	83,67
Özgüllük	97,37	85,32	84,34	86,63	85,41
EAA	98,35	87,51	86,72	88,96	89,38

Şekil 4'te, ESA sınıflandırıcısının Covid-19 sınıflandırması için tahmin performansı açısından diğer son teknoloji sınıflandırma algoritmalarından daha iyi performans gösterdiğini görebilmekteyiz.



Şekil 4. Covid-19 tespiti için sınıflandırma algoritmalarının performans analizi

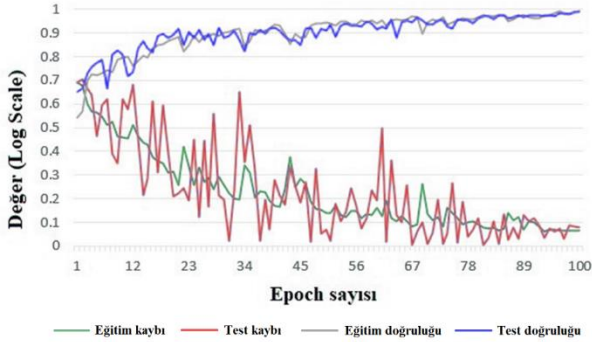
Önerilen tahmin modelinin detaylarıyla tartışılması gerekirse deneysel sonuçlar bölümünde, ESA sınıflandırıcısının, performans açısından diğer algoritmalarından daha iyi performans gösterdiği sayısal ve görsel olarak gösterilmiştir. Bu çalışmayla, önerilen ESA modelimizin göğüs BT taraması görüntüleri aracılığıyla Covid-19 vakalarının tespit edilmesi için kullanılabileceği söylenebilmektedir. Önerilen ESA modelinin tam yapısı şekil 5'te gösterilmektedir.



Şekil 5. Önerilen ESA modelinin yapısı

Deneysel sonuçlar için kullanılan veri kümesindeki görüntüler farklı boyutlarda olduğundan dolayı, görüntüler 64x64 piksellik sabit bir boyuta indirgenmiştir. Ardından, girdi görüntülerinden öznetelikleri çıkarmak için ilk evrişim katmanında 3x3 boyutunda 64 filtre kullanılmıştır. Birinci evrişim katmanının konfigürasyonu aynı şekilde ikinci bir evrişim katmanı daha eklenmiş ve aktivasyon fonksiyonu olarak ReLU kullanılmıştır. Daha sonra, 2x2 boyutunda Maksimum havuzlama, özellik haritasının boyutunu azaltmak için kullanılmıştır. 3x3 boyutunda 128 çekirdeğin bulunduğu bir başka evrişim katmanı daha eklenmiştir. Daha sonra, özellik haritasının boyutunu küçültmek için daha önce olduğu gibi Maksimum havuzlama eklenmiştir. Ardından son evrişim ve maksimum havuz katmanı tekrar eklenmiş ve ardından tek boyutlu bir diziye düzleştirilmiştir. Daha sonra 128 birim ve 256 birimlik toplam iki yoğun / gizli katman eklenmiştir, burada ReLU bir aktivasyon işlevi olarak kullanılmıştır. Nihai çıktı katmanında Covid-19 sınıflandırmasıyla ilgili son tahmini almak için yalnızca bir düğüm kullanılmıştır.

İkili bir sınıflandırma problemi olduğu için sigmoid fonksiyonunun bir aktivasyon fonksiyonu olarak benimsenmesinin nedeni budur. Optimize edici olarak önerilen ESA modeli, kayıp fonksiyonunu hesaplamak için ikili çapraz entropiyi kullanmaktadır. Önerilen ESA modelini daha sağlam hale getirmek ve aşırı uyumluluğu ve yetersiz uyumluluğu önlemek için, tüm operasyonlar devir olarak 100 adet, parti boyutu ise 32 olarak gerçekleştirilmiştir. Şekil 6'da, devrin başlarında eğitim kaybının son derece yüksek olduğu ve doğruluk oranının oldukça düşük olduğu gösterilmektedir. Ancak devir 100'e doğru yaklaştıkça, eğitim ve test kayıplarının 0'a yaklaşmasına karşılık, eğitim ve test doğruluklarının ise son derece yüksek değere ulaştığı görülmektedir.



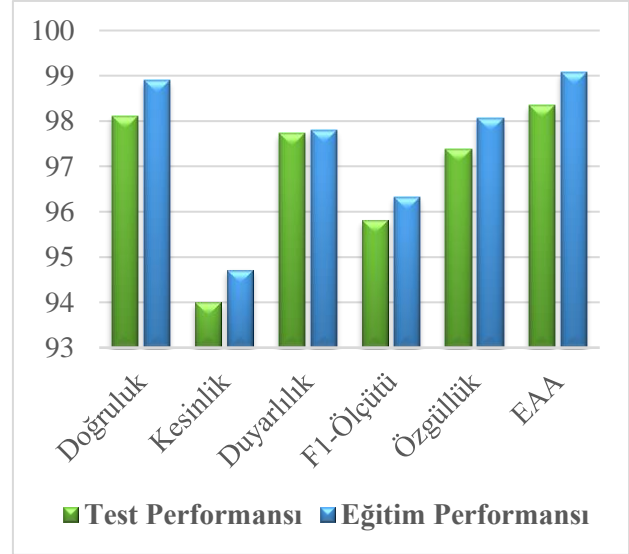
Şekil 6. Önerilen ESA modelin Eğitim, Test kaybı ve Eğitim, Test doğruluk eğrileri

Her adımda, ağırlıklar parti boyutlarına göre sürekli güncellenerek ağa iletilmektedir. Her devir dönemi tamamlandıktan sonra, kayıp değeri kademeli olarak azalmakta ve doğruluk oranı artmaktadır. 100 devir döneminin tamamı tamamlandıktan sonra, önerilen ESA modeli %98,1 doğruluk elde etmiştir. ESA'nın genel test ve eğitim performansları tablo 3'te gösterilmektedir.

Tablo 3. Önerilen ESA modelinin test ve eğitim performans sonuçları

Sınıflandırıcı / Ölçütler	Test Performansı	Eğitim Performansı
Doğruluk	98,10	98,9
Kesinlik	94	94,72
Duyarlılık	97,72	97,8
F1-Ölçütü	95,81	96,33
Özgüllük	97,37	98,07
EAA	98,35	99,07

Ek olarak, önerilen ESA'nın test ve eğitim performansları şekil 7'de görsel olarak karşılaştırılmıştır.



Şekil 7. Önerilen ESA modelinin Test ve Eğitim performanslarının karşılaştırılması

Sonuçlar

TT-PZR şu anda dünyanın hemen hemen tüm ülkelerinde Covid-19 tespiti için kullanılmaktadır. TT-PZR'nin yanlış negatif sonuçlar verebilmektedir ve aynı zamanda tüm dünyada bu TT-PZR kitlerinde aşırı bir eksiklik bulunmaktadır. Bu nedenle, yapay zeka tekniklerinin yanı sıra, göğüs BT tarama görüntüleri, bu sorunu çözmeye ve insanlığın bu krizi aşmasına yardımcı olmakta aktif bir rol oynayabilmektedir. Bu çalışmada, Covid-19 hastalarını doğru bir şekilde tanımlamak ve sınıflandırmak için beş farklı makine öğrenimi ve derin öğrenme algoritması performansı karşılaştırılmıştır. Bu beş algoritmanın performansı analiz edildiğinde, ESA algoritmasının Covid-19 hastalarını tanımlamak için BT tarama görüntülerinden gizli bilgileri çıkarabildiği söylenebilmektedir. Önerilen ESA modeli %98,10 gibi mükemmel bir doğruluk elde etmiştir. Önerilen bu model, yeterli tanımlama kitlerinin ve uzman hekimlerin olmadığı kırsal alanlarda TT-PZR ile birlikte alternatif bir araç veya yardımcı araç olarak kullanılabilir. Bu uygun maliyetli tahmin modeliyle, Covid-19 hastaları bir dakika içinde tespit edilebilecek, böylelikle hastalıktan etkilenen kişilerin diğer bireylerle olan teması kesilerek topluma

bulaşının ve yayılmanın önüne geçilebilecektir. Covid-19 vakalarının büyük veri kümesine kolayca erişilememesi nedeniyle, bu araştırmada, kullanıma açık olan 2482 görüntüden oluşan bir veri kümesi kullanılmıştır. Boyutu daha büyük bir veri kümesi kullanılabilseydi, Covid-19 hastalarını belirlemek için daha sağlam bir tahmin modeli geliştirilebilirdi. Bu tahmin modeli, göğüs BT tarama görüntülerini girdi olarak almakta ve bir dakika içerisinde sonuç verebilmektedir. Tüm deneysel sonuçları elde etmek için 2.3GHz ve 8GB RAM'e sahip Intel Core i7 işlemcili bir sistem kaynağı kullanılmıştır. Çalışmada, yapılan deneysel sonuçlarla önerilen modelin yüksek doğruluğa sahip değerlere ulaşabildiği ortaya konmuştur.

Kaynaklar

- [1] "Coronavirus Update" [Online] Erişim: <https://www.worldometers.info/coronavirus/>, 17.10.2020.
- [2] Bleve, G., Rizzotti, L., Dellaglio, F., Torriani, S., (2003). Development of reverse transcription (RT)-PCR and real-time RT-PCR assays for rapid detection and quantification of viable yeasts and molds contaminating yogurts and pasteurized food products. *Applied and Environmental Microbiology*, **69**, 7, 4116-4122.
- [3] Long, C., Xu, H., Shen, Q., Zhang, X., Fan, B., Wang, C., Li, H., (2020). Diagnosis of the Coronavirus disease (COVID-19): rRT-PCR or CT?. *European journal of radiology*, 108961.
- [4] Dong, D., Tang, Z., Wang, S., Hui, H., Gong, L., Lu, Y., Jin, R., (2020). The role of imaging in the detection and management of COVID-19: a review. *IEEE reviews in biomedical engineering*.
- [5] Kang, H., Xia, L., Yan, F., Wan, Z., Shi, F., Yuan, H., Shen, D., (2020). Diagnosis of coronavirus disease 2019 (covid-19) with structured latent multi-view representation learning. *IEEE transactions on medical imaging*.
- [6] Li, L., Qin, L., Xu, Z., Yin, Y., Wang, X., Kong, B., Cao, K., (2020). Artificial intelligence distinguishes COVID-19 from community acquired pneumonia on chest CT. *Radiology*.
- [7] Butt, C., Gill, J., Chun, D., Babu, B. A., (2020). Deep learning system to screen coronavirus disease 2019 pneumonia. *Applied Intelligence*, 1.
- [8] Ardakani, A. A., Kanafi, A. R., Acharya, U. R., Khadem, N., Mohammadi, A., (2020). Application of deep learning technique to manage COVID-19 in routine clinical practice using CT images: Results of 10 convolutional neural networks. *Computers in Biology and Medicine*, 103795
- [9] Bai, H. X., Wang, R., Xiong, Z., Hsieh, B., Chang, K., Halsey, K., Mei, J., (2020). AI augmentation of radiologist performance in distinguishing COVID-19 from pneumonia of other etiology on chest CT. *Radiology*, 201491.
- [10] Shi, F., Xia, L., Shan, F., Wu, D., Wei, Y., Yuan, H., Shen, D., (2020). Large-scale screening of covid-19 from community acquired pneumonia using infection size-aware classification. *arXiv preprint arXiv:2003.09860*.
- [11] Ozkaya, U., Ozturk, S., Barstugan, M., (2020). Coronavirus (COVID-19) Classification using Deep Features Fusion and Ranking Technique. *arXiv preprint arXiv:2004.03698*.
- [12] Alom, M. Z., Rahman, M. M., Nasrin, M. S., Taha, T. M., Asari, V. K., (2020). COVID_MNet: COVID-19 Detection with Multi-Task Deep Learning Approaches. *arXiv preprint arXiv:2004.03747*.
- [13] Yang, X. J., Gao, F., Ju, Y., (2020). General fractional derivatives with applications in viscoelasticity. *Academic Press*.
- [14] "SARS-COV-2 Ct-Scan Dataset." Eduardo S., Plamen A., [Online] Erişim: Kaggle, doi: 10.34740/KAGGLE/DSV/1100240, 17.10.2020.
- [15] Pal, M., (2005). Random forest classifier for remote sensing classification. *International journal of remote sensing*, **26**, 1, 217-222.
- [16] Özbay, E., Çinar, A., (2019). A Comparative Study of Object Classification Methods Using 3D Zernike Moment on 3D Point Clouds. *Traitement du Signal*, **36**, 6, 549-555.
- [17] "Google Colab" [Online] Erişim: <https://colab.research.google.com/>, 17.10.2020.