



Makine Öğrenmesi Algoritmalarıyla Akciğer Tomografi Görüntülerinden COVID-19 Tespiti

Özgür Kart^{1*}, Fatih Başçiftçi²

^{1*} Selçuk Üniversitesi, Teknoloji Fakültesi, Bilgisayar Mühendisliği, Konya, Türkiye, (ORCID: 0000-0002-7004-8419), "ozgurkart42@hotmail.com"

² Selçuk Üniversitesi, Teknoloji Fakültesi, Bilgisayar Mühendisliği, Konya, Türkiye, (ORCID: 000-0003-1679-7416), "basçiftci@selcuk.edu.tr"

(1st International Conference on Applied Engineering and Natural Sciences ICAENS 2021, November 1-3, 2021)

(DOI: 10.31590/ejosat.1009611)

ATIF/REFERENCE: Kart, Ö., & Başçiftçi, F. (2021). Makine Öğrenmesi Algoritmalarıyla Akciğer Tomografi Görüntülerinden COVID-19. European Journal of Science and Technology, (28), 630-637.

Öz

Koronavirüs (COVID-19), solunum yolu enfeksiyonuna neden olan ve insandan insana geçebilen bulaşıcı bir virüsdür. Bu virüs dünyada kısa sürede etkili olmuş ve bir salgına dönüşmüştür. Bu tür bulaşıcı hastalıkların erken teşhisi ve gerekli tedavinin erken süreçte başlatılması gerekmektedir. COVID-19 hastalığı tespiti için akciğer görüntülerinden ve ağız yoluyla alınan tükürük ile tespit edilmektedir. COVID-19 hastasını RT-PCR (Reverse Transcription- Polymerase Chain Reaction) ile tespit etmek için yaklaşık 4-6 saat sürmektedir. Pandeminin büyüklüğüne bakıldığında çok ta hızlı sayılmamaktadır. Aynı zamanda test kitinin de bir maliyeti bulunmaktadır. Ekonomik olarak güçlü olmayan ülkeler RT-PCR kitlerine erişmekte sorun yaşamaktadır. Pandemi döneminde zorlu süreçlerden bir tanesi her raporu manuel olarak incelemek için, birden fazla radyoloji uzmanı gerekmektedir. Bu çalışmada makine öğrenmesi algoritmaları ile farklı kategorilerdeki akciğer tomografisi görüntülerinden COVID-19 olan görüntü tespit edilmiştir. Orange Data Mining Veri analizi programında makine öğrenmesi algoritması olan K-En Yakın Komşuluk, Yapay Sinir Ağları, Rastgele Orman ve Destek Vektör algoritmaları ile Akciğer veri setinden COVID-19 hastalığına ait görüntüler sınıflandırılmış, en iyi sonucu Destek Vektör Algoritması ile elde edilmiştir.

Anahtar Kelimeler: Yapay Zeka ile hastalık teşhisi, Covid-19 teşhisi, Makine öğrenmesi Algoritmaları, KNN Algoritması, SVM Algoritması, Rastgele Orman Algoritması

COVID-19 Detection from Lung Tomography Images with Machine Learning Algorithms

Abstract

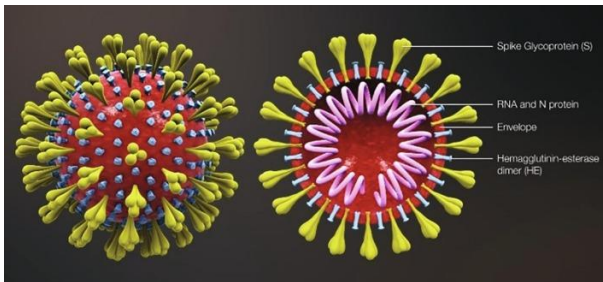
Coronavirus (COVID-19) is a contagious virus that causes respiratory tract infection and can be passed from person to person. This virus became effective in the world in a short time and turned into an epidemic. Early diagnosis of such infectious diseases and the necessary treatment should be initiated in the early period. For the detection of COVID-19 disease, it is detected by lung images and oral saliva. It takes approximately 4-6 hours to detect a COVID-19 patient by RT-PCR (Reverse Transcription- Polymerase Chain Reaction). Considering the size of the pandemic, it is not considered very fast. At the same time, the test kit has a cost. Countries that are not economically strong have problems accessing RT-PCR kits. One of the challenging processes during the pandemic period is to manually review each report, requiring multiple radiologists. In this study, the images with COVID-19 were detected from different categories of lung tomography images with machine learning algorithms. In the Orange Data Mining data analysis program, the images of the COVID-19 disease from the Lung data set were classified with the machine learning algorithm K-Nearest Neighborhood, Artificial Neural Networks, Random Forest and Support Vector algorithms, and the best result was obtained with the Support Vector Algorithm.

Keywords: Disease diagnosis with artificial intelligence, Covid-19 diagnosis, Machine learning Algorithms, KNN Algorithm, SVM Algorithm, Random Forest Algorithm

1. Giriş

Yapay zeka, insan zekasının belli bir parçasını taklit etmeye çalışan sistemlerin genel adıdır. Yapay zeka sayesinde bir çok karmaşık işlem hızlı bir şekilde çözülmektedir. Yapay zekanın gelişimi ile birlikte, bilgisayarların insan gibi düşünmesini sağlamak için yapılan çalışmalar makine öğrenimini ve uygulamalarını açığa çıkarmıştır. Makine öğrenmesi, diğer yapay zeka uygulamalarından farklı olarak, bir yandan insan zekasını taklit ederken, diğer yandan bizim yorumlayıp elle gireceğimiz kurallara ihtiyaç duymayan algoritmalar bütünüdür. Makine öğrenimin gelişmesiyle birçok hastalığın teşhisi ve tedavisinde önemli bir rol oynamaktadır. Makine öğrenimi ve yapay zekanın gelişimiyle birlikte bazı yapılacak işlemleri otomatikleştirdi ve insan müdahalesini sınırlandırdı. Sağlık alanında da makine öğrenme yöntemleri ile hastalıkların tespiti yapılmaktadır. Öğrenilecek veriler de arttığı için makine öğrenimi zamanla çok akıllı ve kullanışlı hale gelmektedir. Bu veriler arttıkça, hastalıkları tespit etmek, sınıflandırmak ve bölümlere ayırmak için otomatik bir sistem uygulama şansı da oluşmaktadır. Makine öğrenimi ve yapay zeka kullanılarak sağlık alanında bir çok çalışma gerçekleştirilmektedir. COVID-19 hastalığının tespit edilmesinde bu yöntemler kullanılabilir.

COVID-19, 2019 yılı Aralık ayında Çin'in Wuhan kentinde ilk olarak görülmüş, kısa bir sürede tüm dünyaya yayılmıştır. Bu virüs salgını hızlı bir şekilde ilerlemiş, birçok ülkenin sağlık sisteminde ciddi sıkıntılar oluşturmuştur. Oluşturduğu küresel salgından dolayı pandemi olarak tanımlanmaktadır [1,2]. COVID-19, Yüksek ateş ve nefes darlığı ile tanımlanan yeni viral solunum yolu hastalığıdır. Hastalığın damlacık ve temas yoluyla bulaştığı bilinmektedir. Bulaştığı canlılarda önemli düzeyde solunum yetmezliğine sebep olmaktadır. koronavirüs enfeksiyonu belirtileri arasında ateş, öksürük, nefes darlığı ve solunum güçlüğü bulunmaktadır. Daha ağır vakalarda enfeksiyon zatürreye, akut solunum yetmezliğine, böbrek yetmezliğine ve hatta ölüme neden olabilmektedir. Özellikle bu hastalık, direnci veya bağışıklık sistemi zayıf olanlar, kronik rahatsızlıkları olanlar ve yaşlılar için çok daha tehlikeli olmaktadır [3]. Şekil 1'de Koronavirüsün şeklini temsil eden üç boyutlu tıbbi görüntü gösterilmektedir [4].



Şekil 1. Koronavirüsün şeklini temsil eden üç boyutlu tıbbi görüntü [4]

COVID-19, genellikle virüs hasta bireylerden öksürme, hapşırma yoluyla ortaya saçılan damlacıklarla ve hastaların solunum salgılarının temas ettiği yüzeylere dokunulmasından sonra ellerin göz, ağız, burun mukozası teması ile bulaşmaktadır [5]. Günümüzde COVID-19'u tespit etmek için, ters transkripsiyon-polimeraz zincir reaksiyonu (RT-PCR) en yaygın kullanılan yöntemdir. Yalnız bu testin süresi uzundur

[6]. Erken dönemlerde COVID-19'u teşhis etmek için X-ışını (Röntgen) ve Bilgisayarlı Tomografi (BT) gibi radyolojik görüntüleme yöntemleri tercih edilmektedir. Pandemi döneminde zorlu süreçlerden bir tanesi her raporu manuel olarak incelemek için, birden fazla radyoloji uzmanı gerekmektedir [7]. Koronavirüsün hızlı bir şekilde yayılması birçok ülkede ölüm oranlarını da artırmıştır. Bu sebeplerden dolayı etkili bir tedavi yönteminin geliştirilmesi, teşhis, karantina ve erken tedavinin başlatılması olmak üzere hastalığın kontrol edilmesi gerekmektedir [8]. Sağlık sistemindeki etki de uzun süre yoğun bakım ünitesine ve mekanik ventilatöre ihtiyaç duyan insan sayısı nedeniyle yüksektir [9]. Bu senaryoda, sağlık sistemindeki stresi olası bir şekilde azaltmak için doğru tedavi için erken teşhis çok önemlidir. Bu bağlamda, yapay zekâ destekli otomatik bir sistem bu sürece katkıda bulunabilmektedir [10,11,12].

2. Materyal ve Metot

2.1. Literatür Taraması

Yapay zeka teknikleri kullanılarak COVID-19 virüsü bulaşmış hastaları teşhis etmek için birçok çalışma gerçekleştirilmiştir.

Li ve arkadaşları yaptıkları makale çalışmasında, COVID-19 hastalığını tespit etmek için, göğüs BT görüntülerinden üç boyutlu bir derin öğrenme modeli tasarlamışlar. Bu Derin öğrenme modeli, COVID-19'u doğru bir şekilde tespit edebildiğini, COVID-19'u zatürre ve diğer akciğer hastalıklarından ayırabildiğini belirtmişlerdir [12].

Uçar ve arkadaşı, Yapay Zeka tabanlı COVIDiagnosis-Net olarak adlandırdıkları bir yapı önermişlerdir. SqueezeNet, Bayesian optimizasyon algoritmaları kullanarak, COVID-19 teşhis çalışması yapmışlar ve sonucunda 0,983 test doğruluğuna ulaşmışlardır. Önerilen model Bayes - SqueezeNet, göğüs röntgeni görüntülerini kullanarak COVID-19 tanısında, diğer modellere göre çok daha iyi performans gösterdiğini ve daha yüksek COVID-19 tanı doğruluğu elde ettiklerini belirtmişlerdir [13].

Zheng ve arkadaşları tarafından yapılan makale çalışmasında, COVID-19'u tespit etmek için 3 boyutlu BT kullanılarak derin öğrenme tabanlı evrişimli sinir ağı (DeCoVNet) önermişlerdir. Derin öğrenme algoritması kullanılarak yapılan çalışmada 0,90 doğruluk oranı ve 0,91 hassasiyet oranına ulaşmışlardır [14].

Rahimzadeh ve arkadaşları yaptıkları çalışmada, 48260 BT tarama görüntüsünü COVID-19'u tespit etmek için kullanılmışlardır. Görüntüleri ImageNet veri kümesini kullanarak eğitmişlerdir. ResNet50V2 ile CT görüntülerini sınıflandırmışlar ve 98,49% doğruluk elde ettiklerini belirtmişlerdir [15].

Alom ve arkadaşları tarafından yapılan çalışmada, COVID-19 hastalarını derin öğrenme yöntemleriyle tanımlamanın hızlı ve verimli bir yolunu önermişler. Hem röntgen hem de CT tarama görüntülerini kullanarak bir çalışma yapmışlar. X-ışını görüntülerinden 84,67% test doğruluğu ve CT görüntülerinden 98,78% doğruluk oranına ulaşmışlar. Yapılan çalışmada, X-ışını ve CT görüntülerinde enfekte olmuş bölgelerin yüzdesini belirlemek için yeni bir analiz stratejisi de önerilmiştir. Niteliksel ve niceliksel sonuçlar, COVID-19

tespiti ve enfekte bölge lokalizasyonu için umut verici sonuçların olduğunu belirtmişlerdir [16].

Salman ve arkadaşları tarafından yapılan çalışmada, COVID-19 hastalığını teşhis etmek için 260 adet X-Ray görüntüsü üzerinde Evrişimli Sinir Ağı (CNN) modelini uygulamışlar ve 100%'lük bir doğruluk oranı ile COVID-19 virüsünün olup olmadığını tespit etmişlerdir[17].

Jaiswal ve arkadaşları, COVID-19 bulaşmış hastaların sınıflandırılması için yaptıkları çalışmada DenseNet201 tabanlı derin transfer öğrenmeyi kullanmışlardır. Sınıflandırma neticesinde 97% doğruluk oranı elde etmiş ve hastaları pozitif ve negatif olarak ayırtmışlardır [18].

Butt ve arkadaşları tarafından yapılan çalışmada, BT göğüs taramalarından elde edilen veri setleri ile çok hızlı ve güvenilir şekilde COVID-19'u tespit etmişlerdir. Derin öğrenme metodu kullanılarak yapılan bu çalışma, RT-PCR testinden daha yüksek bir tespit oranına ulaşmışlar. CNN ile yapılan çalışmada 98,2% duyarlılık ve 92,2% özgüllük elde ettiklerini belirtmişlerdir [19].

Wang ve arkadaşları, COVID-19 hastalığının seyri, hastanın tedaviye nasıl cevap vereceğini tahmin etmek için, BT görüntüsünü kullanarak tam otomatik Derin öğrenme sistemi (DL) önermişler. DL sisteminin, COVID-19'un salgın kontrolünde çok fazla maliyet eklemeyen yardımcı olabileceğini, COVID-19'u tanımlamada iyi performans elde ettiklerini belirtmişlerdir [20].

Pathak ve arkadaşları tarafından yapılan çalışmada, CT görüntülerinden derin transfer öğrenmeye dayalı COVID-19 tespiti yapılmıştır. Bu model ile 96,22% eğitim doğruluğu ve 93,01%'a test doğruluğu sonucunu aldıklarını belirtmişlerdir [21].

Song ve arkadaşları yaptıkları çalışmada, COVID-19 hastalarını belirlemek için derin öğrenmeye dayalı bir BT tanı sistemi geliştirmişlerdir. DRE-Net adını verdikleri modelde COVID-19 hastalarını bakteri pnömonisi hastalarından doğru bir şekilde ayırt edebileceğini belirtmişlerdir. Yapılan çalışmada COVID-19 hastalarını diğerlerinden ayırarak 0.86 doğruluk oranı ve 0.79 hassasiyet oranına ulaşmışlardır [22].

Özyurt ve arkadaşları, Bilgisayarla görme yöntemleri yaygın olarak kullanıldığını belirtmişler. Bu çalışmada, CT görüntülerini kullanarak bir model önermişlerdir. Önerilen modelde Yapay Sinir Ağları (YSA) ve Derin Sinir Ağı (DNN) kullanılmıştır. YSA ve DNN modelleri sırasıyla 94,10% ve 95,84% sınıflandırma doğruluğuna ulaşmışlardır. Yapılan çalışma ile evrişimli sinir ağlarından (CNN) daha yüksek bir sınıflandırma doğruluğu elde ettiklerini belirtmişlerdir [23].

Apostolopoulos ve arkadaşları, Yedi sınıfa ait 3905 adet X-Ray görüntüsünü kullanarak derin öğrenme yöntemlerinden CNN ile COVID-19 hastalığını tespit etmişlerdir. Çalışma sonucunda, derin öğrenme modelinin iyi bir performans gösterdiğini, hastalığın teşhisinde 99,18% doğruluk, 97,36% hassasiyet ve 99,42% özgüllük oranı elde ettiklerini belirtmişlerdir [24].

Serte ve arkadaşı COVID-19 ve normal CT hacimlerini sınıflandırmak için bir yapay zeka modeli önermektedir. Önerilen Yapay Zeka modelinde ResNet-50'yi kullanmışlardır. CT tarama görüntülerinden COVID-19'u tahmin etmek için derin öğrenme modeli ile birlikte birleştirme teknikleriyle karşılaştırmışlardır. Resnet-50 modelinin diğer tüm

modellerden ve birleştirme tekniklerinden daha iyi performans gösterdiğini belirtmişlerdir [25].

2.2. Veri Seti ve Verilerin Hazırlanması

COVID19 hastalığının tespiti için açık kaynaklı bir platform olan "https://www.kaggle.com/" web adresinden güncel ve gerçek verisetleri içinden normal ve COVID-19 X-Ray görüntüleri alınacaktır. Şekil 2'de Normal ve COVID-19 örnek akciğer tomografisi görüntüleri verilmiştir.



Şekil 2. Akciğer tomografisi görüntüleri

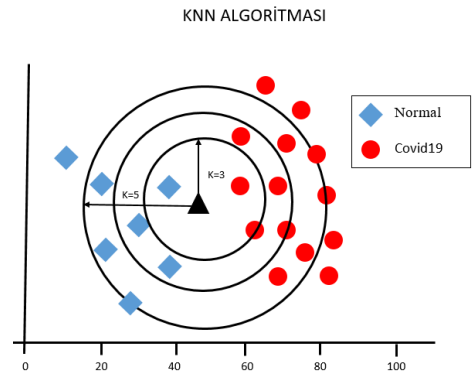
COVID-19'un tespiti için 1500 adet Posteroanterior Chest X-ray (PA-CXR) görüntüsü kullanılmıştır. Bu veri seti içerisinde, 750 adet COVID-19 (enfekte olmuş) ve 750 adet normal veri kullanılmıştır. Tablo 1.'de kullanılacak veri setiyle ilgili sayılar gösterilmektedir.

Tablo 1. Kategoriler ve veri sayıları

Sıra No	Kategori Adı	Eğitim (80%)	Test (20%)	Toplam
1	Normal	600	150	750
2	COVID19	600	150	750
Toplam		1200	300	1500

2.3. K-En Yakın Komşu (KNN) Algoritması

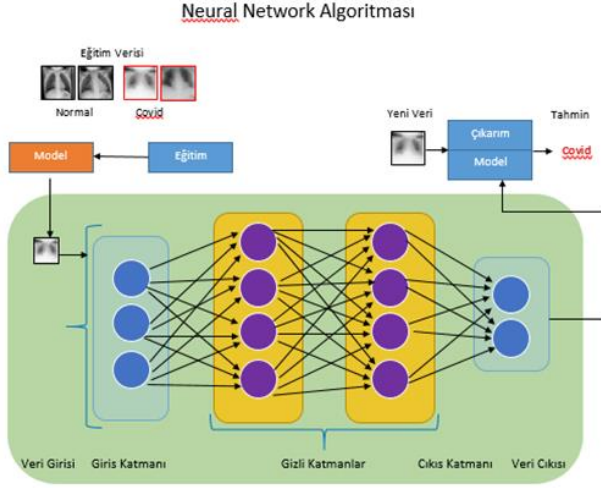
KNN algoritması ya da diğer adıyla K-En yakın komşu algoritması makine öğrenim algoritmaları içerisinde en çok kullanılan algoritmalarından biridir. Sınıfı bilinmeyen veri, eğitim setindeki diğer veriler ile karşılaştırılır ve bir uzaklık ölçümü yapılır. KNN, komşular arasındaki uzaklıkları hesaplar, en yakın komşuları bularak verileri sınıflara göre sınıflandırır [26]. Şekil 3.'de Knn algoritması tasarımı görünmektedir.



Şekil 3. KNN Algoritması Uygulaması

2.4. Neural Network Algoritması

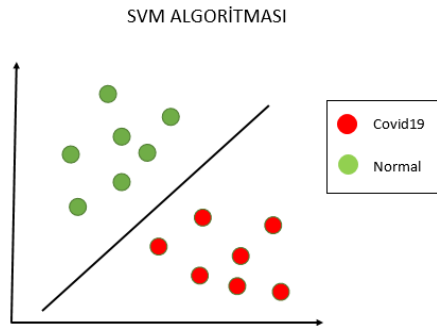
Sinir ağları, kümelenebilir ve sınıflandırmamıza yardımcı olur. Bunları, sakladığınız ve yönettiğiniz verilerin üzerinde bir kümeleme ve sınıflandırma katmanı olarak düşünebilirsiniz. Etiketlenmemiş verileri örnek girdiler arasındaki benzerliklere göre gruplamaya yardımcı olurlar ve üzerinde eğitim alabilecekleri etiketli bir veri kümesine sahip olduklarında verileri sınıflandırır [27]. Şekil 4.'de Neural Network algoritması tasarımı görünmektedir.



Şekil 4. Neural Network Algoritması Tasarımı

2.5. Destek Vektör Makineleri (SVM) Algoritması

Sınıflandırma veya regresyon problemleri için kullanılabilen denetimli bir makine öğrenmesi algoritmasıdır. Bununla birlikte, çoğunlukla sınıflandırma problemlerinde kullanılmaktadır [28]. Bu algoritmada, her bir veri maddesini belirli bir koordinatın değeri olan her özelliğin değeri ile birlikte n-boyutlu boşluğa bir nokta olarak çizilir. Ardından, iki sınıftan oldukça iyi ayırım yapan hiper-düzlemi bularak sınıflandırma gerçekleştirilir. Şekil 5.'de SVM algoritması tasarımı görünmektedir.



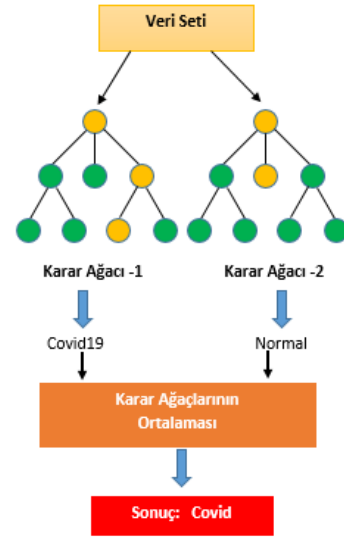
Şekil 5. SVM Algoritması Tasarımı

2.6. Random Forest Algoritması

Sınıflandırma işlemi sırasında birden fazla karar ağacı üreterek sınıflandırma oranını yükselten algoritmadır. Karar ağaçlarının birleşiminden oluşan bir algoritma olup, kullanılan karar ağaçları arasında doğruluğu ve bağımsızlığı en yüksek ağaçlar tercih edilmektedir. Ağaçların her birisi veri setindeki özelliklere göre dallanmaktadır. Tek bir düğüm ile başlar ve

yeni sonuçlara dallanarak bir ağaç yapısı oluşturulur. Algoritma çalıştığında girilen değer düğümlere bakılarak belli bir yolda ilerler ve bir sonuç verir [29]. Şekil 6.'da Random Forest Algoritması tasarımı gösterilmektedir.

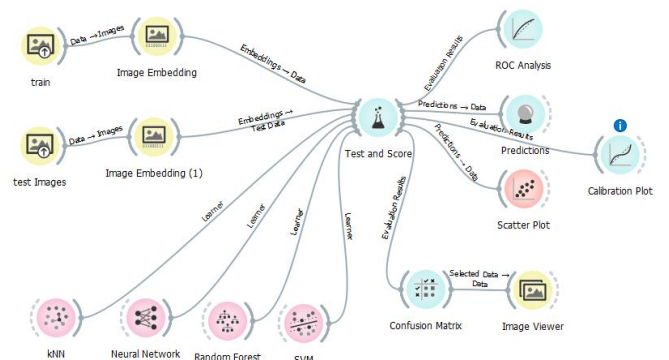
Random Forest Algoritması



Şekil 6. Random Forest Algoritması Tasarımı

2.7. Çalışmanın Veri Analiz Programında Tasarımı

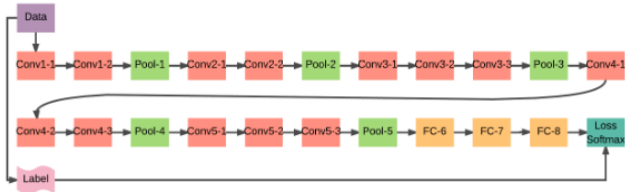
COVID-19 ve normal akciğer görüntüleri ile modelin eğitilmesi sağlanmıştır. Bundan dolayı modele ait girdilerin değerleri tanımlanması yapılmıştır. Şekil 7.'de Çalışmanın COVID-19 teşhisi için Orange veri analiz ekranı verilmektedir. Train ve test olmak üzere iki bölümden görüntüler işlenmek üzere Embedding kısmına gelmektedir. Visual Geometry Group (VGG-16) ağ modeli ile işlendikten sonra seçilen 4 adet algoritma ile modeller oluşturulmaktadır. Bu modeller ile test ve score kısmında başarı değerleri tablolar ile verilmektedir. Test ve Score'dan sonraki kısımda ise test verilerinin algoritmalara göre görselleştirilmesi yapılmıştır. Bunlar her algoritmaya ait yüzdelik doğruluk değerleri, çıkış katmanına göre doğruluk, hassasiyet, duyarlılık vb. değerleri verilmiştir.



Şekil 7. Çalışmanın Orange Veri analizi programında tasarımı

VGG16 modeli Oxford Üniversitesi tarafından geliştirilmiştir. VGG16, tamamen bağlı ağı eğiten yaklaşık 138

milyon parametreye sahip büyük ölçekli görüntü tanıma modelidir [30]. Bu modelde görüntüdeki düşük ve yüksek seviyeli özellikleri çıkarmak için resme bazı filtreler uygulamaktadır. VGG16 içerisinde 13 konvolüsyon katman, 3 tam bağlı katman ile birlikte 2x2 boyutlu 5 adet havuzlama katmanı bulunmaktadır. Son katmanda ise softmax bulunmaktadır, bu katmanda gelen giriş verisi sınıflandırılmaktadır [31]. Şekil 8.'de VGG16 mimarisinin görünümü verilmiştir [32].



Şekil 8. VGG16 mimarisi [18].

3. Araştırma Sonuçları ve Tartışma

3.1. Deneysel Sonuçlar

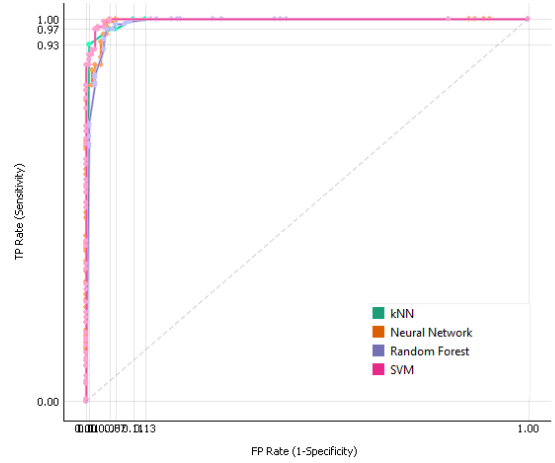
Bu bölümde, normal ve COVID-19 X-Ray görüntülerine ait eğitim ve test işlemi yapılmıştır. KNN, SVM, Random Forest ve Neural Network makine öğrenme algoritmaları kullanılmıştır. 80% eğitim ve 20% test verisi olmak üzere veri seti iki gruba ayrılmıştır. Algoritmalarla göre test başarımları ve doğruluk oranı Tablo 2.'te görülmektedir. En yüksek test başarımları 99,8%'lük bir oranla SVM algoritmasından elde edilmiştir.

Tablo 2. Algoritmalarından elde edilen sonuçlar

Model	Test	Duyarlılık	F1	Hasasiyet	Doğruluk
KNN	99,3%	95,3%	95,3%	95,4%	95,3%
SVM	99,8%	96,6%	96,6%	96,9%	96,7%
Random	98,9%	95%	94,9%	95,2%	95,0%
Neural Network	99,4%	96,3%	96,3%	96,5%	96,3%

3.2. Algoritmaların Performans Grafiği

Makine öğrenmesinde performans ölçümü, sınıflandırma probleminin performansının değerlendirilmesinde AUC(Area Under Curve) - ROC (Receiver Operating Characteristic Curve) eğrisinden yararlanılmaktadır. Herhangi bir sınıflandırma modelinin performansını kontrol etmek için en önemli değerlendirme ölçütlerinden biridir [33]. ROC bir olasılık eğrisidir ve AUC derece veya ölçüyü temsil etmektedir. Modelin sınıflar arasında ne kadar ayırım yapabildiğini belirtmektedir. AUC ne kadar yüksekse, model hastalığı olan hastaları ayırt etmede o derecede başarılıdır [34]. Şekil 9'da uygulanan 4 model ile ilgili grafik verilmiştir. Yapılan çalışmada, en iyi oranı SVM algoritmasından elde edilmiştir.



Şekil 9. ROC-AUC Eğrisi

3.3. Tartışma

Bu bölümde kullanılan makine öğrenme algoritmalarından elde edilen sonuçların karşılaştırılması yapılmıştır. Makine öğrenmesinde kullanılan sınıflandırma modellerinin performansını değerlendirmek için karmaşıklık matrisi (Confusion Matrix) kullanılmaktadır. Hedef niteliğe ait tahminlerin ve gerçek değerlerin karşılaştırıldığı karmaşıklık matrisi sıklıkla kullanılmaktadır [35].

3.3.1. KNN Algoritması modelinden elde edilen sonuçlar

KNN modelinden test için kullanılan 300 adet normal ve covid X-Ray görüntülerinden 99,3% test başarımları oranı, 95,3% duyarlılık oranı 95,3% F1 oranına karşılık geldi. 95,4% hassasiyet oranı ve 95,3% doğruluk oranı elde edilmiştir. Tablo 3.'de KNN Algoritmasının Confusion Matrix sonuçları gösterilmektedir. Gerçek Covid X-Ray görüntülerini 93,6% ve gerçek normal X-Ray görüntülerini 97,2% oranında doğru tahmin etmiştir. KNN Algoritması ile yapılan çalışmada ortalama 95,3% tahmin oranı elde edilmiştir. Tablo 4.'de KNN Algoritmasının test sayıları Confusion Matrix tablosunda gösterilmiştir. Buna göre 150 adet gerçek Covid X-Ray görüntülerini 146 tanesini ve 150 adet gerçek normal X-Ray görüntülerinin 140 tanesini doğru tahmin etmiştir. İki kategori toplamında ortalama 143 tanesini doğru tahmin etmiştir.

Tablo 3. KNN Algoritmasının Confusion Matrix sonuçları

Gerçek Veriler	Tahmin Edilen Veriler			Σ
		Covid	Normal	
Covid		93,6%	2,8%	150
Normal		6,4%	97,2%	150
Σ		156	144	300

Tablo 4. KNN Algoritması Test Sayılarının Confusion Matrix Tablosu

Gerçek Veriler	Tahmin Edilen Veriler			Σ
		Covid	Normal	
Covid		146	4	150
Normal		10	140	150
Σ		156	144	300

3.3.2. SVM Algoritması modelinden elde edilen sonuçlar

1500 adet X-Ray görüntüsü içerisinde 1200 adet eğitim verisi, 300 adet test verisi olarak belirlenmiştir. SVM algoritması ile yapılan testte X-Ray görüntülerinden 99,8% test başarımları ile 4 algoritma içerisinde en yüksek oranı elde etmiştir. 96,6% duyarlılık oranı 96,6% F1 oranına karşılık geldi. 96,9% hassasiyet oranı ve 96,7% doğruluk oranı elde etmiştir. Tablo 5.'de SVM Algoritmasının Confusion Matrix sonuçları gösterilmektedir. Gerçek Covid X-Ray görüntülerini 93,8% ve gerçek normal X-Ray görüntülerini 100% oranında doğru tahmin etmiştir. SVM Algoritması ile yapılan çalışmada ortalama 96,9% tahmin oranı elde edilmiştir. Tablo 6.'da SVM Algoritmasının test sayıları Confusion Matrix tablosunda gösterilmiştir. Buna göre 150 adet gerçek Covid X-Ray görüntülerini 150 tanesini ve 150 adet gerçek normal X-Ray görüntülerinin 140 tanesini doğru tahmin etmiştir. İki kategori toplamında ortalama 145 tanesini doğru tahmin etmiştir.

Tablo 5. SVM Algoritmasının Confusion Matrix sonuçları

Gerçek Veriler	Tahmin Edilen Veriler			Σ
		Covid	Normal	
Covid		93,8%	0%	150
Normal		6,2%	100%	150
Σ		160	140	300

Tablo 6. SVM Algoritmasının Test Sayılarının Confusion Matrix Tablosu

Gerçek Veriler	Tahmin Edilen Veriler			Σ
		Covid	Normal	
Covid		150	0	150
Normal		10	140	150
Σ		160	140	300

3.3.3. Random Forest Algoritması modelinden elde edilen sonuçlar

300 adet test için kullanılan Normal ve COVID-19 X-Ray görüntülerinden 98,9% test başarımları, 95% duyarlılık oranı 94,9% F1 oranına karşılık geldi. 95,2% hassasiyet oranı ve 95% doğruluk oranı elde etmiştir. Tablo 7.'de Random Forest Algoritmasının Confusion Matrix sonuçları gösterilmektedir. Gerçek Covid X-Ray görüntülerini 91,2% ve gerçek normal X-Ray görüntülerini 96,5% oranında doğru tahmin etmiştir. Random Forest Algoritması ile yapılan çalışmada ortalama

93,85% tahmin oranı elde edilmiştir. Tablo 8.'de Random Forest Algoritmasının test sayıları Confusion Matrix tablosunda gösterilmiştir. Buna göre 150 adet gerçek Covid X-Ray görüntülerini 145 tanesini ve 150 adet gerçek normal X-Ray görüntülerinin 136 tanesini doğru tahmin etmiştir. İki kategori toplamında ortalama 141 tanesini doğru tahmin etmiştir.

Tablo 7. Random Forest Algoritmasının Confusion Matrix sonuçları

Gerçek Veriler	Tahmin Edilen Veriler			Σ
		Covid	Normal	
Covid		91,2%	3,5%	150
Normal		8,8%	96,5%	150
Σ		159	141	300

Tablo 8. Random Forest Algoritması Test Sayılarının Confusion Matrix Tablosu

Gerçek Veriler	Tahmin Edilen Veriler			Σ
		Covid	Normal	
Covid		145	5	150
Normal		14	136	150
Σ		159	139	300

3.3.4. Neural Network Algoritması modelinden elde edilen sonuçlar

300 adet güncel ve gerçek X-Ray görüntülerinden 99,4% test başarımları, 96,3% duyarlılık oranı %96,3 F1 oranına karşılık geldi. 96,5% hassasiyet oranı ve 96,3% doğruluk oranı elde etmiştir. Tablo 9.'da Neural Network Algoritmasının Confusion Matrix sonuçları gösterilmektedir. Gerçek Covid X-Ray görüntülerini 93,7% ve gerçek normal X-Ray görüntülerini 99,2% oranında doğru tahmin etmiştir. Neural Network Algoritması ile yapılan çalışmada ortalama 96,5% tahmin oranı elde edilmiştir. Tablo 10.'da Neural Network Algoritmasının test sayıları Confusion Matrix tablosunda gösterilmiştir. Buna göre 150 adet gerçek Covid X-Ray görüntülerini 149 tanesini ve 150 adet gerçek normal X-Ray görüntülerinin 140 tanesini doğru tahmin etmiştir. İki kategori toplamında ortalama 145 tanesini doğru tahmin etmiştir.

Tablo 9. Neural Network Algoritmasının Confusion Matrix sonuçları

Gerçek Veriler	Tahmin Edilen Veriler			Σ
		Covid	Normal	
Covid		93,7%	0,7%	150
Normal		6,3%	99,3%	150
Σ		159	141	300

Tablo 10. Neural Network Test Sayılarının Confusion Matrix Tablosu

Gerçek Veriler	Tahmin Edilen Veriler			Σ
	Covid	Normal	Σ	
Covid	149	1	150	
Normal	10	140	150	
Σ	159	141	300	

4. Sonuç

COVID-19 tüm dünyaya hızla bir şekilde yayılmış ve bulaştığı insanlar üzerinde ciddi sıkıntılar oluşturmuştur. COVID-19'un erken evrelerde tespit edilmesi, zamanında hızlı ve doğru müdahalenin yapılması, salgının hem yayılımını azaltacak hem de oluşabilecek sağlık sorunlarını en aza indirebilecektir. Son yıllarda makine öğrenmesi birçok alanda kullanılmaktadır. Makine öğrenmesinin, hastalıkların erken teşhisinde kullanılabilme potansiyeli de bulunmaktadır. Bu çalışmada Makine öğrenmesi teknikleriyle "https://www.kaggle.com/" web adresinden alınan güncel ve gerçek veriler (X-Ray görüntüleri) kullanılarak COVID-19'u tespiti yapılmıştır. Yapılan uygulamada, 4 farklı makine öğrenme algoritması kullanılmıştır. Yapılan çalışmada En yüksek 99,8% test başarı oranı, 96,9% hassasiyet oranı, 96,9% duyarlılık oranı ve 96,7% doğruluk oranı ile SVM algoritması ile elde edilmiştir. Sonuç olarak, yapılan çalışmada, yüksek oranda doğru tahmin elde edilmiştir. Makine öğrenme algoritmalarının sağlık alanında verimli bir şekilde kullanılabileceğini göstermektedir.

Kaynakça

- [1] Sohrabi, C., Alsafi, Z., O'Neill, N., Khan, M., Kerwan, A., Al-Jabir, A., ... & Agha, R. (2020). World Health Organization declares global emergency: A review of the 2019 novel coronavirus (COVID-19). *International journal of surgery*, 76, 71-76.
- [2] Knight, T. E. (2020). Severe acute respiratory syndrome coronavirus 2 and coronavirus disease 2019: a clinical overview and primer. *Biopreservation and Biobanking*, 18(6), 492-502.
- [3] Lai, C.-C., Shih, T.-P., Ko, W.-C., Tang, H.-J., & Hsueh, P.-R. (2020). Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and coronavirus disease-2019 (COVID-19): The epidemic and the challenges. *International Journal of Antimicrobial Agents*, 55(3), 105924. <https://doi.org/10.1016/j.ijantimicag.2020.105924>
- [4] Wikimedia Commons, 3D medical animation corona virus.jpg, https://commons.wikimedia.org/wiki/File:3D_medical_animation_corona_virus.jpg, [Ziyaret Tarihi: 15 Mayıs 2021].
- [5] Toğaçar, M., Ergen, B., & Cölmert, 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*, 1-12, 2020.
- [6] Franquet, T. (2011). Imaging of pulmonary viral pneumonia. *Radiology*, 260(1), 18-39.
- [7] Öztürk, T., Talo, M., Yıldırım, E. A., Baloğlu, U. B., Yıldırım, Ö., & Acharya, U. Automated detection of COVID-19 cases using deep neural networks with X-ray images, *Computers in Biology and Medicine*, 1-11, 2020.
- [8] Tolksdorf, K., Buda, S., Schuler, E., Wieler, L. H., ve Haas, W., 2020, Influenza-associated pneumonia as reference to assess seriousness of coronavirus disease (COVID-19). *Euro Surveill*, 25(11). doi:10.2807/1560-7917.ES.2020.25.11.2000258
- [9] Grasselli, G., Pesenti, A., & Cecconi, M. (2020). Critical care utilization for the COVID-19 outbreak in Lombardy, Italy: early experience and forecast during an emergency response. *Jama*, 323(16), 1545-1546.
- [10] Mei, X., Lee, H. C., Diao, K. Y., Huang, M., Lin, B., Liu, C., Xie, Z., Ma, Y., Robson, P. M., Chung, M., Bernheim, A., Mani, V., Calcagno, C., Li, K., Li, S., Shan, H., Lv, J., Zhao, T., Xia, J., Long, Q., ... Yang, Y., 2020, Artificial intelligence-enabled rapid diagnosis of patients with COVID-19, *Nature medicine*, 26(8), 1224-1228. <https://doi.org/10.1038/s41591-020-0931-3>
- [11] Gozes, O., & Siegel, E. (2020). Rapid AI development cycle for coronavirus, pandemic: Initial results for automated detection & patient monitoring, using deep learning CT image analysis. *arXiv preprint arXiv:2003.05037*.
- [12] Li, L., Qin, L., Xu, Z., Yin, Y., Wang, X., Kong, B., ... & Xia, J. (2020). Artificial intelligence distinguishes COVID-19 from community acquired pneumonia on chest CT. *Radiology*.
- [13] Ucar F., Korkmaz D., 2020, COVIDiagnosis-Net: Deep Bayes-SqueezeNet based diagnosis of the coronavirus disease 2019 (COVID-19) from X-ray images, *Medical Hypotheses*, 2020 Jul;140:109761. DOI: 10.1016/j.mehy.2020.109761.
- [14] Zheng, C., Deng, X., Fu, Q., Zhou, Q., Feng, J., Ma, H., Liu, W., ve Wang, X., 2020, Deep Learning-based Detection for COVID-19 from Chest CT using Weak Label, *MedRxiv*, 2020.03.12.20027185. <https://doi.org/10.1101/2020.03.12.20027185>
- [15] Rahimzadeh, M., Attar, A., 2020, A modified deep convolutional neural network for detecting COVID-19 and pneumonia from chest X-ray images based on the concatenation of Xception and ResNet50 V2, *Informatics in Medicine Unlocked* 19, 100360, <https://doi.org/10.1016/j.imu.2020.100360>
- [16] Alom, M. Z., Rahman, M. M., Nasrin, M. S., Taha, T. M., & Asari, V. K., 2020, COVID_MTNNet: COVID-19 detection with multi-task deep learning approaches. *arXiv preprint arXiv:2004.03747*.
- [17] Salman, F. M., Abu-Naser, S. S., Alajrami, E., Abu-Nasser, B. S., ve Alashqar, B. A., 2020, Covid-19 detection using artificial intelligence.
- [18] Jaiswal, A., Gianchandani, N., Singh, D., Kumar, V., ve Kaur, M., 2020, Classification of the COVID-19 infected patients using DenseNet201 based deep transfer learning. *Journal of Biomolecular Structure and Dynamics*, 1-8.

- [19] Butt, C., Gill, J., Chun, D., Babu, B. A., 2020, Deep learning system to screen coronavirus disease 2019 pneumonia, *Applied Intelligence*, 1–7, Advance online publication. <https://doi.org/10.1007/s10489-020-01714->
- [20] Wang, S., Zha, Y., Li, W., Wu, Q., Li, X., Niu, M., . . . Yu, H., 2020, A fully automatic deep learning system for COVID-19 diagnostic and prognostic analysis. *European Respiratory Journal*, 56(2).
- [21] Pathak, Y., Shukla, P. K., Tiwari, A., Stalin, S., ve Singh, S., 2020, Deep transfer learning based classification model for COVID-19 disease. *Irbm*.
- [22] Song, Y., Zheng, S., Li, L., Zhang, X., Zhang, Xiaodong., Huang, Z., Chen, J., Wang, R., Zhao, H., Zha, Y., Shen, J., Chong, Y., ve Yang, Y., 2021, “Deep learning Enables Accurate Diagnosis of Novel Coronavirus (COVID-19) with CT images”. *IEEE/ACM Transactions on Computational Biology and Bioinformatics* 1-1. doi: 10.1109/TCBB.2021.3065361.
- [23] Ozyurt, F., Tuncer, T., ve Subasi, A., 2021, An Automated COVID-19 Detection Based on Fused Dynamic Exemplar Pyramid Feature Extraction and Hybrid Feature Selection Using Deep Learning, *Computers in Biology and Medicine* 132:104356. doi: 10.1016/j.combiomed.2021.104356.
- [24] Apostolopoulos, I. D., Aznaouridis, S. I., ve Tzani, M. A., 2020, Extracting possibly representative COVID-19 biomarkers from X-Ray images with deep learning approach and image data related to pulmonary diseases. *Journal of Medical and Biological Engineering*, 40, 462-469.
- [25] Serte, S., ve Demirel, H., 2021, Deep learning for diagnosis of COVID-19 using 3D CT scans. *Computers in biology and medicine*, 132, 104306.
- [26] Kılınc, D., Borandağ, E., Yücalar, F., Tunali, V., Şimşek, M., ve Özçift, A., 2016, KNN Algoritması ve R Dili ile Metin Madenciliği Kullanılarak Bilimsel Makale Tasnifi. *Marmara Fen Bilimleri Dergisi*, 28(3). doi:10.7240/mufbed.69674
- [27] Nicholson, C., 2020, A Beginner's Guide to Neural Networks and Deep Learning. Retrieved from, <https://wiki.pathmind.com/neural-network>
- [28] Hsu, C. W., Chang, C. C., ve Lin, C. J., 2010, A Practical Guide to Support Vector Classification.
- [29] Hatipoğlu, E., 2018, Machine Learning — Prediction Algorithms — Decision Tree — Random Forest — Part 5. Retrieved from <https://medium.com/@ekrem.hatipoglu/machine-learning-prediction-algorithms-decision-tree-random-forest-part-5-2970905c021e>
- [30] Zeiler, M. D., & Fergus, R., 2014, September, Visualizing and understanding convolutional networks. In *European conference on computer vision* (pp. 818-833). Springer, Cham.
- [31] Ullah, I., Hussain, M., Qazi, E.-H., Aboalsamh, H., 2018, An automated system for epilepsy detection using EEG brain signals based on deep learning approach, *Expert Syst. Appl.* 107, 61–71.
- [32] Qassim, H., Verma, A., & Feinzimer, D., 2018, Compressed residual-VGG16 CNN model for big data places image recognition. In *2018 IEEE 8th Annual Computing and Communication Workshop and Conference (CCWC)* (pp. 169-175). IEEE.
- [33] Taş, B., 2019, Roc Eğrisi ve Eğri Altında Kalan Alan (Auc), <https://bernatas.medium.com/roc-e%C4%9Frisi-ve-e%C4%9Fri-alt%C4%B1nda-kalan-alan-auc-97b058e8e0cf>
- [34] Narkhede, S., 2018, Understanding auc-roc curve. *Towards Data Science*, 26, 220-227.
- [35] Brownlee, J., 2020, What is a Confusion Matrix in Machine Learning, <https://machinelearningmastery.com/confusion-matrix-machine-learning/>