

PAPER DETAILS

TITLE: X-Ray Görüntülerinden Covid-19 Tespiti için Hibrit Bir Derin Öğrenme Yaklaşımı

AUTHORS: Gizem Öter, Emrah Haçer

PAGES: 27-36

ORIGINAL PDF URL: <https://dergipark.org.tr/tr/download/article-file/2879011>

X-Ray Görüntülerinden Covid-19 Tespiti için Hibrit Bir Derin Öğrenme Yaklaşımı

Gizem Öter^{1*}, Emrah Hançer²

^{1*} Burdur Mehmet Akif Ersoy Üniversitesi, Sosyal Bilimler Enstitüsü, Yönetim Bilişim Sistemleri Bölümü, Burdur, Türkiye, (ORCID: 0000-0002-1627-3543), gizemmoterr@gmail.com

² Burdur Mehmet Akif Ersoy Üniversitesi, Teknoloji Fakültesi, Yazılım Mühendisliği Bölümü, Burdur, Türkiye (ORCID: 0000-0002-3213-5191), ehancer@mehmetakif.edu.tr

(İlk Geliş Tarihi 8 Ocak 2023 ve Kabul Tarihi 24 Eylül 2023)

(DOI: 10.5281/zenodo.10254260)

ATIF/REFERENCE: Öter, G. & Hançer, E. (2023). X-Ray Görüntülerinden Covid-19 Tespiti için Hibrit Bir Derin Öğrenme Yaklaşımı. *Avrupa Bilim ve Teknoloji Dergisi*, (52), 27-36.

Öz

Covid-19 salgını tüm dünyada hayatı durma noktasına getirmiş ve binlerce can almıştır. Bu salgın ile mücadelede en önemli unsur erken teşhistir. Erken teşhisle enfekte olan kişi karantinaya alınarak bulaş hızı ve ölüm oranı düşürülebilir. Günümüzde Covid-19'un teşhisinde PCR (Polymerase Chain Reaction- Polimeraz Zincir Reaksiyonu) testlerinden ve akciğer X-Ray görüntülerinden faydalanılmaktadır. PCR testlerinden sonuç almak zaman alabilen bir süreç olduğundan dolayı erken teşhiste X-Ray görüntüleri önem arz etmektedir. Bu çalışmada Covid-19 teşhisinde uzmanlara yardımcı olabilmek amacıyla X-Ray görüntüleri kullanılarak bir teşhis metodolojisi tasarlanmıştır. Metodolojinin ilk aşamasında ön eğitilmiş SqueezeNet mimarisi ile X-Ray görüntülerinden oluşan veri setinden öznitelikler çıkarılmıştır. İkinci aşamada, çıkarılan özniteliklere en yakın bileşen analizi (neighborhood component analysis (NCA)) uygulanarak mevcut öznitelik setinden gürültülü olanları elimine edilmiştir. Son aşamada ise, destek vektör makineleri ile gürültüden arındırılmış öznitelik seti üzerinde Covid-19 tespiti gerçekleştirilmiştir. Önerilen metodolojinin performans analizi 5 farklı veri seti üzerinde birden çok derin öğrenme mimarileri ile karşılaştırılarak yapılmıştır. Elde edilen sonuçlar incelendiğinde önerilen metodolojinin GoogLeNet, ResNet18, MobileNetV2 ve VGG16 mimarilerinden daha başarılı sonuçlar aldığı gözlemlenmiştir.

Anahtar Kelimeler: Derin Öğrenme, SqueezeNet, Covid-19, NCA.

A Hybrid Deep Learning Approach for Diagnosis of Covid-19 from X-Ray Images

Abstract

The Covid-19 epidemic has brought life to a standstill all over the world and claimed thousands of lives. The most important element in combating the epidemic is early diagnosis. It is possible to reduce the rate of transmission and mortality by quarantining the infected person with early diagnosis. Today, PCR (Polymerase Chain Reaction) tests and lung X-Ray images are used in the diagnosis of Covid-19. The purpose of this study is to develop a diagnostic methodology based on X-Ray images in order to assist experts in diagnosing Covid-19. Since getting results from PCR tests is a lengthy process, X-Ray images are important in early diagnosis. In the first stage of the methodology, features are extracted from the dataset consisting of X-Ray images by using a pre-trained SqueezeNet architecture. In the second stage, nearest component analysis (NCA) is applied to the extracted feature set, and noisy ones are eliminated from this feature set. In the final stage, Covid-19 detection is performed on the noise-free feature set with support vector machines. The performance analysis of the proposed methodology is made by comparing it with a variety of deep pre-trained architectures on 5 different datasets. It is obtained from the results that the proposed methodology achieves more successful results than GoogLeNet, ResNet18, MobileNetV2 and VGG16 methodologies.

Keywords: Deep Learning, SqueezeNet, Covid-19, NCA.

* Sorumlu Yazar: gizemmoterr@gmail.com

1. Giriş

Zamanının küresel sağlık krizi olarak tanımlanan Covid-19 tüm dünyada etkisi göstermektedir. Covid-19 sadece Türkiye’de 16.896.522 vaka sayına ulaşmış ve 101.179 kişinin ölümüne neden olmuştur (Covid-19, 2022). Her ülke virüsle mücadele sürecinde sokağa çıkma kısıtlamaları, seyahat kısıtlamaları, sosyal mesafe, maske zorunluluğu ve hijyen bilincini arttırmaya yönelik çalışmalar yapmış ve yapmaya devam etmektedir. Ancak, virüsün etkilerini tam anlamıyla bitirilebilmiş değildir (Chowdhury, v.d, 2020).

Kuşkusuz Covid-19 ile mücadelede en önemli hususlardan birisi hızlı tanı ve hızlı müdahaledir. Günümüzde yapılan tanı yöntemlerinden en yaygın PCR (Polimeraz Zincir Reaksiyon) testleridir. Yapılan bu testlerin sonuçlanması 3-5 saat sürebilmektedir. Bu durum ilgili süreci olumsuz etkilemektedir. Bu test yönteminin kitlere bağımlı olması ve kırsal alanlarda erişim sıkıntısı olması gibi olumsuz yönleri vardır. Bu testlerin temel hedefinde, sonucu daha hızlı bir şekilde verebilmek olmasına rağmen maliyetleri fazladır ve test sonucunu alabilme süresi X-ray görüntülerinden alınan sonuçlara oranla daha uzundur.

Röntgen bir diğer adıyla X-ray bir nesnenin içini görselleştirmek için belirli radyasyon türlerini kullanan bir görüntüleme tekniğidir. Tıpta X-ray görüntüleri oldukça erişilebilir ve akciğer koşullarını değerlendirmede kullanılan yaygın bir yöntemdir. Göğüs röntgeni ise göğüs bölümünde bulunan kemik, doku ve organların görüntülenme şeklidir. Covid-19, X-ray görüntülerinde opasifikasyon (Buzlu cam görünümleri) şeklinde belirtilen beyaz lekeler şeklinde görüntülenebilir (Ardakani, v.d, 2020). Bu bağlamda X-ray görüntüleri, Covid-19’un tespitinde kullanılabilir. Buradan yola çıkılarak son zamanlarda bu görüntüler üzerinde derin öğrenme algoritmaları kullanılarak Covid-19 tespiti üzerinde çalışmalar yoğunlaşmıştır (Ismael & Şengür, 2021; Demir & Yılmaz, 2021).

(Toğaçar, v.d, 2019) evrişimsel sinir ağlarını kullanarak zatürre hastalığını erken tespit etmeye çalışmışlardır. Çalışmada kullanılan veri seti 1-5 yaş arası çocukların göğüs X-ray görüntülerinden oluşmaktadır. AlexNet mimarisi kullanılarak öznetelik çıkarımı yapılmıştır. Elde edilen öznetelikler karar ağaçları, destek vektör makinesi, lineer regresyon, doğrusal ayırmacılık analizi, en yakın komşu ve softmax kullanılarak sınıflandırılmıştır. Çalışmada en iyi sonuçların destek vektör makinesiyle alındığı gözlemlenmiştir.

(Minaee, v.d, 2020) evrişimsel sinir ağlarını kullanarak Covid-19 tespitine yardımcı olmayı amaçlamışlardır. Birden fazla kaynaktan toplanan 5000 X-ray görüntüsünden yeni veri seti oluşturulmuştur. Veri seti SqueezeNet, ResNet50, ResNet18 ve DenseNet-161 mimarileri kullanılarak eğitilmiştir. Sonuçlara bakıldığında çok az bir farkla SqueezeNet mimarisiyle daha başarılı sonuçlar alındığı söylenebilmektedir.

(Rahman, v.d, 2020) bakteriyel zatürre ve zatürrenin otomatik olarak tespit edebilmesine yönelik bir çalışma ortaya koymuşlardır. Çalışmada 5247 X-ray görüntüsünden oluşan veri seti ve AlexNet, ResNet18, SqueezeNet, DenseNet201 mimarileri kullanılmıştır. DenseNet201 mimarisi ile en başarılı sonuçlar elde edilmiştir.

(Eryılmaz & Karacan, 2021) COVID-19 ve diğer akciğer hastalıklarına ait X-Ray görüntüleriyle çoklu sınıflandırma yapılmasını amaçlamıştır. Veri setinde yer alan 11.293 X-Ray görüntüsü MobileNetV2, NASNetMobile, Xception ve DenseNet121 mimarileri kullanılarak sınıflandırılmış ve sonuçların karşılaştırmalı analizi yapılmıştır. Çalışmada en iyi sonuçlar DenseNet121 ve MobileNet mimarileri kullanılarak elde edilmiştir.

(Özcan, 2020) GoogleNet, ResNet18 ve ResNet50 mimarilerini kullanarak Covid-19 tespitine yardımcı olacak bir yöntem geliştirilmiştir. Çalışmada Grid Search ve öğrenme aktarımı destekli evrişimsel sinir ağına dayalı bir yöntem önerilmiştir. Kullanılan veri seti %50 eğitim, %20 doğrulama ve %30 test olacak şekilde ayrılmıştır. Sonuçlar karşılaştırıldığında, en başarılı sonucun Grid Search ve ResNet50 kullanılarak sağlandığı gözlemlenmiştir. Bir başka çalışmada, (Özcan, 2021) X-ray görüntülerinden çıkarılan derin öznetelikler kullanılarak Covid-19 vakalarının tespit edilmesinde yeni bir birleşik model yaklaşımı önerilmiştir. Yaklaşımın single layer-based (SLB) ve feature fusion-based (FFB) olmak üzere iki varyansı vardır. SLB modeli, ön işleme, derin öznetelik çıkarımı, son işleme ve sınıflandırma aşamalarından oluşmaktadır. FFB modeli ise ön işleme, derin öznetelik çıkarımı, öznetelik birleştirme, son işleme ve sınıflandırma aşamalarından oluşmaktadır. Öznetelik çıkarma aşamasında kullanılan katmanların sayısı ve ikili kombinasyonuna göre dört farklı SLB ve altı farklı FFB modeli geliştirilmiştir. Her model için ikili ve çoklu sınıflandırma işlemi yapılmıştır. FFB3 modeli ile daha başarılı sonuçlar elde edilmiştir.

Literatürdeki çalışmalardan görüleceği üzere, X-Ray görüntülerinden Covid-19 başta olmak üzere çeşitli hastalıkların otomatik olarak tespit edilmesine yönelik çalışmalar hızla devam etmektedir. Ancak mevcut çalışmaların çoğunluğu belirli bir veri seti üzerinde yoğunlaşmakta ve bundan kaynaklı olarak farklı veri setine uygulandığında istenilen performansı göstermeyebilir. Dolayısıyla farklı veri setleri üzerinde uygulanarak geliştirilmiş hastalık tespit metodolojilerine ihtiyaç vardır. Literatürdeki çalışmalarda görülen bir başka husus da evrişimsel sinir ağı mimarilerinin kullanımına ilişkindir. Her ne kadar ilgili derin öğrenme mimarileri görüntülerden ilgili özneteliklerin çıkarılmasında çok iyi yöntemler olsa da elde edilen özneteliklerin tamamının kullanılması, oluşturulacak olan sınıflandırma modelinin performansına ve verimliliğine olumlu katkı sağlaması mümkün olmayabilir. Bir diğer söylemle çıkarılan özneteliklerde faydalı olanlar ve gereksiz veya alakasız olanlarla beraber bulunabilmektedir. Bu sebepten ötürü gereksiz veya alakasız olan öznetelikleri elimine eden öznetelik seçme algoritmalarının kullanımı ihtiyaç söz konusudur.

Bu çalışmanın amacı, mevcut X-Ray görüntülerinden Covid-19 hastalığı tespitine yönelik olarak genelleştirilmiş bir metodoloji ortaya koymaktır. Bu amacı gerçekleştirmek için öncelikle üç aşamalı bir derin öğrenme tabanlı Covid-19 tespit metodolojisi geliştirilmiştir. İlk aşamada ön eğitilmiş SqueezeNet mimarisi (Iandola, v.d, 2016) ile X-Ray görüntülerinden özneteliklerin çıkarılması sağlanır. İkinci aşamada çıkarılan özneteliklerden komşuluk bileşen analizi (neighborhood component analysis (NCA)) algoritması (Goldberger, v.d, 2004) ile ilgili olan özneteliklerin seçilmesi sağlanır. Son aşamada ise seçilen öznetelik alt seti üzerinde destek vektör makineleri (support vector machines (SVM)) algoritması ile bir sınıflandırma modeli oluşturularak Covid-19 tespiti gerçekleştirilir. Geliştirilen metodoloji spesifik bir veri setine göre dizayn edilmeyip 5 farklı Covid-19 ile ilişkili veri setleri üzerinde uygulanmıştır.

Çalışmanın diğer bölümlerinin tasarımı şu şekilde tasarlanmıştır: Bölüm 2’de kullanılan veri setleri ve önerilen Covid-19 tespit yöntemi anlatılmıştır. Bölüm 3’te metodolojide kullanılan algoritmaların parametre ayarları ve değerlendirme için kullanılan metrikler verilmiştir. Ardından elde edilen deneysel sonuçlar tartışılmıştır. Son bölümde ise çalışmanın genel değerlendirmesi yapılacak olup ilerde yapılması muhtemel çalışmalara değinilmiştir.

2. Materyal ve Metot

2.1. Kullanılan Veri Setleri

Dataset 1, GitHub’tan (Covid chestxray dataset, 2022) alınmıştır. Pozitif olan veya şüphelenilen hastaların akciğer röntgen görüntüleri farklı hastanelerden alınarak veri seti oluşturulmuştur. Veri seti Montreal Üniversitesi Etik Kurulu tarafından onaylanmıştır. Veri setinde Covid-19, normal ve zatürre bireylerin X-ray görüntüleri yer almaktadır (Tablo 1).

Tablo 1. Dataset 1 (Table 1. Dataset 1)

Sınıf Etiketi	Sayı
Covid-19	125
Normal	500
Zatürre	500
Toplam	1125

Dataset 2, Kaggle’dan (Curated chest X-ray image dataset for Covid-19, 2022) alınmıştır. Veri seti 15 farklı kaynaktan alınan verilerin harmanlanmasıyla oluşturulmuştur. Veri setinde Covid-19, zatürre, bakteriyel zatürre ve normal bireylerin X-ray görüntüleri yer almaktadır (Tablo 2).

Tablo 2. Dataset 2 (Table 2 Dataset 2)

Sınıf Etiketi	Sayı
Covid-19	1281
Bakteriyel Zatürre	3001
Normal	3270
Zatürre	1656
Toplam	9208

Dataset 3, Kaggle’dan (Covid-19 radiography database, 2022) alınmıştır. Katar Üniversitesi, Dakka Üniversitesi, Bangladeş'ten bir araştırma ekibi, Pakistan ve Malezya'dan işbirlikçiler ve doktorların yardımıyla 43 farklı kaynaktan yararlanılarak bu veri seti oluşturulmuştur. Veri setinde Covid-19 ,zatürre, akciğer opaklığı ve normal bireylerin X-ray görüntüleri yer almaktadır (Tablo 3).

Tablo 3. Dataset 3 (Table 3. Dataset 3)

Sınıf Etiketi	Sayı
Covid-19	3616
Akciğer Opaklığı	6012
Normal	10192
Zatürre	1345
Toplam	21165

Dataset 4, Kaggle’dan (Covid-19 detection X-ray dataset, 2022) alınmıştır. Farklı kaynaklardan alınan veriler derlenerek hazırlanmış bir veri setidir. Veri setinde Covid-19, bakteriyel zatürre, zatürre ve normal bireylerin X-ray görüntüleri yer almaktadır (Tablo 4).

Tablo 4. Dataset 4 (Table 4. Dataset 4)

Sınıf Etiketi	Sayı
Covid-19	60
Bakteriyel Zatürre	650
Normal	880
Zatürre	412
Toplam	2002

Dataset 5, Kaggle’dan (Covid X-ray dataset, 2022) platformundan alınmıştır. Farklı kaynaklardan alınan veriler derlenerek hazırlanmış bir veri setidir. Veri setinde Covid-19 ve normal bireylerin X-ray görüntüleri yer almaktadır (Tablo 5).

Tablo 5. Dataset 5 (Table 5. Dataset 5)

Sınıf Etiketi	Sayı
Covid-19	1790
Normal	1301
Toplam	3091

2.2. Önerilen Metodoloji

Önerilen derin öğrenme tabanlı Covid-19 tespit metodolojisi Şekil 1’de sunulmuştur. Şekil 1’den görüleceği üzere öncelikli olarak verilerin toplaması gerçekleştirilmiştir. Ardından SqueezeNet ile öğrenme aktarımı (transfer learning) işlemi uygulanarak öz niteliklerin çıkartılması sağlanmıştır. Çıkarılan öz nitelik setine daha sonra NCA öz nitelik seçme algoritması uygulanarak gürültülü öz niteliklerin elimine edilmesi sağlanmıştır. Son olarak geriye kalan alt öz nitelik seti ile SVM algoritması kullanılarak bir Covid-19 tespit modeli oluşturulmuştur.

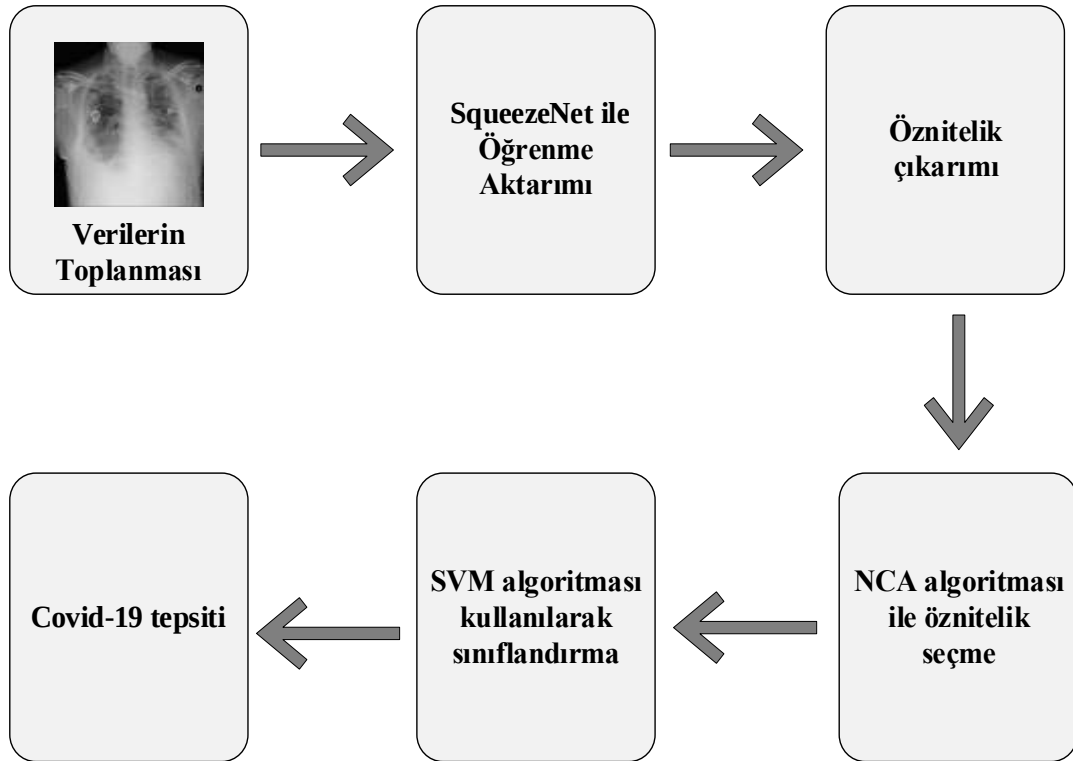
2.2.1. SqueezeNet

Iandola ve arkadaşları evrişimli sinir ağları mimarileri içerisinde yer alan SqueezeNet modelini 2016 yılında önermişlerdir. SqueezeNet, AlexNet’e kıyasla 50 kat daha az parametre kullanarak aynı işlemleri yapabilmektedir. Bunun yanı sıra model sıkıştırma teknikleri ile SqueezeNet 0,5 MB’den daha az veya AlexNet’ten 510 kat daha küçük olacak şekilde sıkıştırılabilmektedir. Böylece bellekte daha az yer kaplayan ve AlexNet’e kıyasla daha kısa öğrenme süresi olan bir mimarıdır. Günümüzde bu gibi olumlu yanlarından dolayı sıklıkla kullanılmaktadır (Iandola, v.d, 2016; Polat & Özerdem, 2021; Öter, 2023).

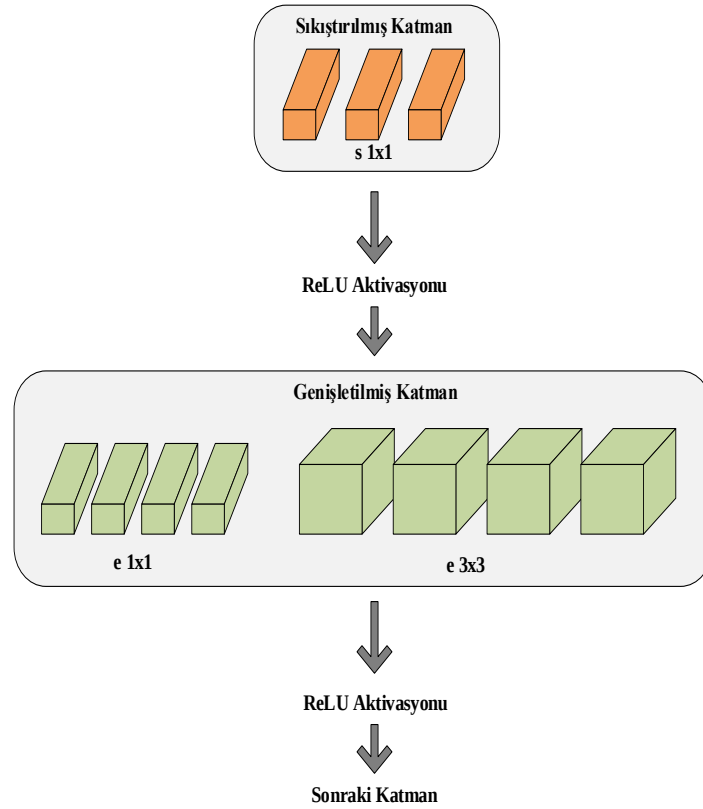
SqueezeNet önerilen ateşleme modüllerinden oluşmaktadır. Ateşleme modülünde genişletilmiş ve sıkıştırılmış katmanlar yer almaktadır. Genişletilmiş katman, $1 \times 1(e_{1 \times 1})$ $3 \times 3(e_{3 \times 3})$ filtrelerinden oluşmaktadır. Sıkıştırılmış katman ise $1 \times 1(s_{1 \times 1})$ filtresinden oluşmaktadır. SqueezeNet modelinde genişletilmiş katmandaki toplam filtre sayısı, sıkıştırılmış katmandaki toplam filtre sayısından büyük olarak ayarlanmıştır ($\text{num}[(e_{1 \times 1}) + (e_{3 \times 3})] > \text{num}[(s_{1 \times 1})]$).

Genişletilmiş katmandaki toplam filtre sayısı, sıkıştırılmış katmandaki toplam filtre sayısından büyük olabilmesi için giriş kanal sayısının 3×3 filtrelerle sınırlandırılması gerçekleştirilmiştir (Şekil 2).

Metodolojide kullanılan yöntemler sırasıyla devam eden alt bölümlerde açıklanmıştır.



Şekil 1. Çalışmanın Metodolojisi (Figure 1. Methodology of the Study)



Şekil 2. Ateşleme modülündeki evrimci filtrelerin organizasyonu. Bu örnekte; $s1 \times 1 = 3$, $e1 \times 1 = 4$ ve $e3 \times 3 = 4$ olarak ayarlanmıştır. (Figure 2. Organization of convolution filters in the Fire module. In this example, $s1 \times 1 = 3$, $e1 \times 1 = 4$, and $e3 \times 3 = 4$.)

2.2.2. En Yakın Bileşen Analizi (NCA)

Günümüzde çok sayıda büyük boyutlu verilerin ortaya çıkmasıyla birlikte makine öğrenmesi uygulamalarında öznitelik seçimi önem kazanmıştır. Öznitelik seçimi, belirli bir veri kümesindeki gereksiz ve aykırı öznitelikleri çıkararak alt küme oluşturma tekniğidir (Hançer, v.d, 2015). Öznitelik seçme yöntemlerinden biri olan Neighborhood Component Analysis (Komşuluk Bileşen Analizi), sınıflandırma doğruluğunu maksimize edecek şekilde öznitelik ağırlıklandırma vektörünü öğrenen en yakın komşu tabanlı öznitelik ağırlıklandırma algoritmasıdır. Eğitim veri setindeki her bir örneğin kendi bulunduğu sınıftaki örneklerle komşu olabilmeye ihtimali artırılmaktadır. Komşuluk bileşen analizinde özniteliklerin arasındaki mesafe ölçümleri ve ağırlıkları baz alınarak yapılmaktadır (Goldberger, v.d, 2004; İsmailoğlu, 2021; Hançer, 2020). NCA algoritması non-parametrik ve denetimli öğrenme algoritmaları içerisinde yer almaktadır (Özyurt, 2019; Tuncer & Ertam, 2020).

$T = \{(x_1, y_1), \dots, (x_i, y_i), \dots, (x_N, y_N)\}$, bir eğitim kümesi olduğu varsayılırsa, x_i d boyutlu bir değişken vektörünü, $y_i \in \{1, \dots, C\}$ sınıf etiketini ve N örnek sayısını ifade eder. Amaç, en yakın komşu algoritmasını optimize eden öznitelik alt kümesini seçmek için uygun olan ağırlık (w) vektörünün bulunmasıdır (Yang, v.d, 2012).

2.2.3. Destek Vektör Makineleri (SVM)

Destek vektör makinelerinin temeli 1963'te Vladimir Vapnik ve Alexey Chervonenkis tarafından ortaya atılmıştır. Daha sonra 1995 yılında Vladimir Vapnik, Isabelle Guyon ve Bernhard Boser tarafından (Boser, Guyon & Vapnik, 1992) destek vektör makinesinin (SVM) temeli geliştirilmiştir. İlk olarak sınıflandırma problemlerine çözüm olarak ortaya çıkarılmış daha sonra regresyon problemlerine uyarlanmıştır. Temel gayesi iki sınıfın ayrımını en iyi şekilde belirleyecek hiper-düzlemi bulmaktır (Cortes & Vapnik, 1995). Destek vektör makineleri verilerin türüne bağlı olarak çekirdek fonksiyonlarda kullanılabilir. Bu sayede hem doğrusal hem de doğrusal olmayan sınıflandırmayı yapabilmektedir. Eğer tam ayrıştırılabilen veriler kullanılırsa, hiper düzlem ile sınıflandırma yapılabilir. Tam ayrıştırılamayan veriler kullanılırsa da aynı boyutta tek bir düzlem ile sınıflandırma yapılmaktadır. Destek vektör makineleri, işlem basamaklarını azaltmasından dolayı çözümlemeyi hızlı bir şekilde yapabilmektedir. Bu gibi özelliklerinden dolayı büyük verilerin analizinde sıklıkla tercih edilmektedir (Ayhan & Erdoğmuş, 2014; Öter, 2023).

3. Sonuçlar ve Tartışma

3.1. Deneysel Dizayn

Çalışmada ilk olarak SqueezeNet ile öznitelik çıkarımı yapılmıştır. SqueezeNet'in 66. katmanı olan "pool10" kullanılmış ve katmandan alınan öznitelikler SVM algoritması kullanılarak sınıflandırılmıştır. Veri setleri %80 eğitim %20 test olacak şekilde ayrılmıştır. Başarıyı arttırmak içinde SqueezeNet ile öznitelik çıkarımı yapıldıktan sonra NCA algoritması kullanılarak öznitelik seçme yapılmıştır. NCA algoritmasının parametreleri eşik değeri 0.02 ve iterasyon sayısı 100 olarak belirlenmiştir. Öznitelik seçme işlemi ile elde edilen öznitelikler SVM algoritması kullanılarak sınıflandırılmıştır.

Önerilen metodolojinin performansını değerlendirmek için kullandığımız başlıca metrikler şunlardır:

a) *Doğruluk (Accuracy)*: Doğruluk değeri, doğru tahmin edilmiş verilerin toplam veri setine oranını ifade etmektedir.

$$\frac{TN + TP}{TN + TP + FP + FN} \quad (1)$$

Burada TP ve TP gerçek pozitif (true positive) ve gerçek negatifi (true negative) temsil ederken, FP ve FN yanlış pozitif (false positive) ve yanlış negatifi (false negative) temsil etmektedir.

b) *Kesinlik (Precision)*: Doğru tespit edilen pozitif sınıfların toplam pozitif değerlere oranıdır.

$$\frac{TP}{TP + FP} \quad (2)$$

c) *Duyarlılık (Recall)*: Doğru tespit edilen pozitif sınıfların gerçek pozitif değerlere oranıdır.

$$\frac{TP}{TP + FN} \quad (3)$$

d) *F1-Skoru (F1-Score)*: F1-skoru, kesinlik ve duyarlılık ölçütlerinin harmonik ortalamasıdır.

$$2 * \frac{(Recall * Precision)}{(Recall + Precision)} \quad (4)$$

e) *Cohen'in Kappa Katsayısı (Cohen Kappa Score)*: Cohen Kappa, iki veya daha fazla gözlemci arasındaki değişimleri hesaplayarak uyumu ölçen bir yöntemdir. Hesaplanırken P_0 ve P_c kullanılmaktadır. P_0 , uyumların toplamının oranını ifade eder. P_c ise uyumların şansa bağlı ortaya çıkma olasılığını ifade etmektedir. Kappa -1 ile +1 arasında değerler almaktadır. Değer +1 ise uyumun tam olduğu, 0 ise uyumun şansa bağlı olduğu, -1 ise uyumsuzluğun tam olduğu ifade edilmektedir.

$$k = \frac{P_0 - P_c}{1 - P_c} \quad (5)$$

Önerilen metodolojinin karşılaştırmalı performans analizi için şu derin öğrenme mimarileri kullanılmıştır:

a) *GoogleNet*, havuzlama katmanları da dahil olmak üzere 27 katmandan oluşmaktadır. ILSVR 2014 yarışmasında %5.7 hata oranıyla birinci olmuştur (Szegedy, v.d, 2015).

b) *ResNet18*, 18 katman derinliğinden oluşan bir mimaridir. ResNet18, 50, 101 ve 152 olmak üzere farklı mimarileri de bulunmaktadır. Bu mimariler aynı mantığı takip eder tek farkları ağıdaki katman sayılarıdır (He, v.d, 2016).

c) *MobileNetV2*, 53 katman derinliğinden oluşan bir mimaridir. MobileNetV2'in hesaplama gücü diğer mimarilere göre daha düşüktür. Böylece düşük donanım gücüne sahip bilgisayarlara ve mobil cihazlara daha kolay adapte olabilen bir mimari olmaktadır (Sandler, v.d, 2018).

d) *VGG16*, 2014 yılında ILSVR yarışmasında tanıtılmıştır. 13 evrişim ve 3 tam bağlı katmanlardan oluşan evrişimsel bir sinir ağıdır. ImageNet veri setindeki görüntüler ile eğitilmiş ve %89 oranında başarı elde etmiştir (Simonyan & Zisserman, 2014; Bozkurt, 2021).

3.2. Deneysel Sonuçlar

Deneysel sonuçlar doğruluk, f1-skor, kappa, duyarlılık, kesinlik ve boyut üzerinden Tablo 6 ve 7'de sunulmuştur. İlgili tablolarda 'boyut' elde edilen öznitelik sayısını temsil etmektedir. Tablo 6'da iki metodoloji bulunmaktadır: 'SqueezeNet' mevcut özniteliklerin ilgili derin öğrenme yöntemi ile çıkarılarak öznitelik seçme yapılmadan elde edilen sonuçları ve 'SqueezeNet-NCA' ise, önerilen metodoloji ile elde edilen sonuçları ifade etmektedir. Bir diğer söylemle mevcut metodolojilerin birbirinden tek farkı, birinin öznitelik seçme aşamasını uygulamamasıdır.

Tablo 6'ya göre, önerilen metodoloji diğer metodolojiye oranla daha iyi sonuçlar elde etmiştir. Üç ve daha fazla sınıf içeren ilk dört veri setinde sınıflandırma performansı artarken veri setinin boyutu da önemli ölçüde azaltılmıştır. Örneğin, Dataset 1 ve Dataset 2'de boyut yaklaşık olarak 5 buçuk kat azalmasına rağmen sınıflandırma doğruluk oranı %6-%7 oranında artmıştır. Diğer iki sınıf bilgisi içeren Dataset 5 ise yine önemli ölçüde öznitelik sayısı azaltılmış ve sınıflandırma performansı genel olarak korunmuştur. Her ne kadar Dataset 5'te ufak bir performans kaybı olsa da öznitelik sayısındaki yaklaşık 7 kat azalma bu durumu kabul edilebilir kılmaktadır.

Tablo 7'ye göre, SqueezeNet ile oluşturulan Covid-19 tespit metodolojisi genel olarak diğer derin öğrenme mimarilerinden daha iyi sonuçlara ulaşmıştır. Sadece Dataset 5'te SqueezeNet-NCA GoogLeNet-NCA'nın yaklaşık olarak %1 oranında gerisinde kalmıştır. Elde edilen öznitelik sayılarına bakıldığında ise, bir metodolojinin diğerine keskin bir üstünlüğü söz konusu olmayıp genel olarak tüm metodolojiler gürültülü öznitelikleri elimine etmekte başarılı bir performans sergilemiştir. Sonuç olarak önerilen Covid-19 tespit metodolojisi üç ve daha fazla sınıf içeren veri setlerinde ortalama %83.49 gibi bir oran elde ederken, bu oran iki sınıflı veri setinde

%93.85'e kadar çıkmıştır. Bu durum öznelilik seçmenin derin öğrenmede ne kadar hayati bir öneme sahip olduğunu ve önerilen metodolojinin doğru bir şekilde dizayn edildiğini göstermektedir.

4. Sonuç

Bu çalışmada Covid-19 tespiti için geliştirilmiş bir metodoloji önerilmiştir. Önerilen metodoloji spesifik bir veri seti yerine X-Ray görüntülerinden oluşturulmuş 5 farklı veri seti üzerinde ve 5 farklı metodoloji ile performans analizi değerlendirmesi yapılmıştır. Bu değerlendirmeler yapılırken sınıflandırıcı olarak SVM, öznelilik seçme için ise NCA algoritmalarından yararlanılmıştır. Çalışmada önerilen 'SqueezeNet-NCA' metodolojinin gerek veri setleri ile gerekse farklı mimariler ile karşılaştırılması sonucunda başarılı sonuçların elde edildiği gözlemlenmiştir. Literatürde her ne kadar bu alanla ilgili birçok çalışma yapılmış olsa da birden fazla veri setinin kullanıldığı, öznelilik seçme ve NCA algoritmasının ele alındığı çalışmaların son derece az olduğu saptanmıştır. Bu çalışma ile bir nebze de olsa değinilmeyen bu konular ele alınarak literatürdeki boşlukların giderilmesi amaçlanmıştır. Bundan sonraki çalışmada geliştirilmiş bir Covid-19 veri seti oluşturulması üzerinde çalışılması düşünülmektedir.

Tablo 6. Deneysel Çalışma Sonuçları (Table 6. Results of Experimental Study)

Dataset	Yöntem	Doğruluk	F1-Skor	Kappa	Duyarlılık	Kesinlik	Boyut
Dataset 1	SqueezeNet	74.22	78.16	55.61	72.67	84.55	1000
	SqueezeNet-NCA	80.44	84.15	67.33	85.33	83	184
Dataset 2	SqueezeNet	79.74	78.98	71.96	80.32	77.68	1000
	SqueezeNet-NCA	86.22	88.77	76.87	89.67	87.89	184
Dataset 3	SqueezeNet	75.40	79.31	64.22	81.30	77.41	1000
	SqueezeNet-NCA	85.78	85.83	77.78	84.68	87.02	548
Dataset 4	SqueezeNet	80.25	75.86	69	68.45	85.06	1000
	SqueezeNet-NCA	81.50	80.33	71.36	78.69	82.04	214
Dataset 5	SqueezeNet	95.15	95.02	90.04	95.02	95.02	1000
	SqueezeNet-NCA	93.85	93.69	87.39	93.69	93.69	149

Tablo 7. Deneysel Çalışmanın Karşılaştırma Sonuçları (Table 7. Comparative Results of the Experimental Study)

Dataset	Yöntem	Doğruluk	F1-Skor	Kappa	Duyarlılık	Kesinlik	Boyut
Dataset 1	SqueezeNet-NCA	80.44	84.15	67.33	85.33	83	184
	GoogLeNet-NCA	71.11	72.13	49.61	65.33	80.50	183
	ResNet18-NCA	79.11	78.41	64.57	77.33	79.52	228
	MobileNetV2-NCA	75.56	74.81	58.44	73.67	75.99	229
	VGG16-NCA	77.33	79.59	61.36	77	82.36	166
Dataset 2	SqueezeNet-NCA	86.22	88.77	76.87	89.67	87.89	184
	GoogLeNet-NCA	81.21	79.32	73.48	79.11	79.52	291
	ResNet18-NCA	83.60	81.76	76.71	80.44	83.12	366
	MobileNetV2-NCA	65.78	70.38	40.32	63.33	79.20	345
	VGG16-NCA	53.83	70.12	40.71	61.68	81.23	294
Dataset 3	SqueezeNet-NCA	85.78	85.83	77.78	84.68	87.02	548
	GoogLeNet-NCA	80.86	81.32	69.41	77.44	85.60	458
	ResNet18-NCA	78.76	81.01	69.11	84.16	78.09	582
	MobileNetV2-NCA	81.78	81.35	70.50	74.83	89.13	557
	VGG16-NCA	80.51	81.14	68.57	75.74	87.37	474
Dataset 4	SqueezeNet-NCA	81.50	80.33	71.36	78.69	82.04	214
	GoogLeNet-NCA	78.75	76.57	67.52	73.72	73.66	210
	ResNet18-NCA	81.25	78.59	71.21	75.60	81.82	256
	MobileNetV2-NCA	78.75	76.33	66.51	68.61	86.01	249
	VGG16-NCA	77.75	73.97	65.52	75.80	72.23	188
Dataset 5	SqueezeNet-NCA	93.85	93.69	87.39	93.69	93.69	149
	GoogLeNet-NCA	95.15	95.03	90.05	95.07	94.98	146
	ResNet18-NCA	93.37	93.22	86.42	93.33	93.11	148
	MobileNetV2-NCA	94.50	94.36	88.72	94.41	94.32	179
	VGG16-NCA	94.34	94.29	88.47	94.59	94	145

5. Kaynakça

- Ardakani, A. A., Kanafi, A. R., Acharya, U. R., Khadem, N. ve 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*, 121(103795), 1-9. doi: <https://doi.org/10.1016/j.combiomed.2020.103795>
- Ayhan, S. ve Erdoğan, Ş. (2014). Destek vektör makineleriyle sınıflandırma problemlerinin çözümü için çekirdek fonksiyonu seçimi. *Eskişehir Osmangazi Üniversitesi İktisadi ve İdari Bilimler Dergisi*, 9(1), 175-201. <https://dergipark.org.tr/tr/pub/oguiibf/issue/5712/76473>
- Bozkurt, F. (2021). Derin öğrenme tekniklerini kullanarak akciğer X-ray görüntülerinden Covid-19 tespiti. *Avrupa Bilim ve Teknoloji Dergisi*, (24), 149-156. doi: 10.31590/ejosat.898385
- Chowdhury, M. E., Rahman, T., Khandakar, A., Mazhar, R., Kadir, M. A., Mahbub, Z. B., Islam, K. R., Khan, M. S., Al-Emadi, N., Reaz, M. R ve Islam, T. I (2020). Can AI help in screening viral and Covid-19 pneumonia?. *IEEE Access*, 8, 132665-132676. doi: 10.1109/ACCESS.2020.3010287.
- Cortes, C. ve Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3), 273-297. doi: <http://dx.doi.org/10.1007/BF00994018>
- Covid chestxray dataset. Github. Erişim Tarihi: 7 Ağustos 2022, <https://github.com/ieee8023/covid-chestxray-dataset>
- Covid-19 detection X-ray dataset. Kaggle. Erişim Tarihi: 7 Ağustos 2022, <https://www.kaggle.com/datasets/darshan1504/covid19-detection-xray-dataset>
- Covid-19 radiography database. Kaggle. Erişim Tarihi: 7 Ağustos 2022, <https://www.kaggle.com/datasets/tawsifurrahman/covid19-radiography-database>
- Covid-19. T.C. Sağlık Bakanlığı. Erişim Tarihi: 10 Ekim 2022, <https://covid19.saglik.gov.tr>
- Covid X-ray dataset. Kaggle. Erişim Tarihi: 7 Ağustos 2022, <https://www.kaggle.com/datasets/ahemateja19bec1025/covid-xray-dataset>
- Curated chest X-ray image dataset for Covid-19. Kaggle. Erişim Tarihi: 7 Ağustos 2022, <https://www.kaggle.com/datasets/unaisait/curated-chest-xray-image-dataset-for-covid19>
- Demir, F. B. ve Yılmaz, E. (2021). X-ray görüntülerinden Covid-19 tespiti için derin öğrenme temelli bir yaklaşım. *Avrupa Bilim ve Teknoloji Dergisi*, 32, 627-632. doi: 10.31590/ejosat.1039522
- Eryılmaz, F. ve Karacan, H. (2021). Akciğer X-ray görüntülerinden Covid-19 tespitinde hafif ve geleneksel evrişimsel sinir ağ mimarilerinin karşılaştırılması. *Düzce Üniversitesi Bilim ve Teknoloji Dergisi*, 9(6), 26-39. doi: 10.29130/dubited.1011829
- Goldberger, J., Hinton, G. E., Roweis, S. Ve Salakhutdinov, R. R. (2004). Neighbourhood components analysis. *Advances in Neural Information Processing Systems* 17 (NIPS).
- Hancer, E. New filter approaches for feature selection using differential evolution and fuzzy rough set theory. *Neural Comput & Applic*, 32, 2929–2944 (2020). <https://doi.org/10.1007/s00521-020-04744-7>
- Hançer, E., Xue, B., Karaboğa, D., & Zhang, M. (2015). A binary ABC algorithm based on advanced similarity scheme for feature selection. *Applied Soft Computing*, 36, 334-348. <https://doi.org/10.1016/j.asoc.2015.07.023>
- He, K., Zhang, X., Ren, S. ve Sun, J. (2016). Deep residual learning for image recognition. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, NV, USA, 770-778. doi: 10.1109/CVPR.2016.90 770-778.
- Iandola, F. N., Han, S., Moskewicz, M. W., Ashraf, K., Dally, W. J. ve Keutzer, K. (2016). SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and < 0.5 MB model size. *arXiv preprint arXiv:1602.07360*.
- Ismael, A. M. ve Şengür, A. (2021). Deep learning approaches for Covid-19 detection based on chest X-ray images. *Expert Systems with Applications*, 164 (114054). doi: 10.1016/j.eswa.2020.114054
- İsmailoğlu, F. (2021). Metrik öğrenmesi kullanarak çeşitli kanser dokularına ait mikro dizi gen verilerinin sınıflandırılması. *Düzce Üniversitesi Bilim ve Teknoloji Dergisi*, 9 (5), 1739-1753. doi: 10.29130/dubited.886353
- Minaee, S., Kafieh, R., Sonka, M., Yazdani, S. ve Jamalipour Soufi, G. (2020). Deep-Covid: Predicting Covid-19 from chest X-ray images using deep transfer learning. *Medical image analysis*, 65 (101794). doi: <https://doi.org/10.1016/j.media.2020.101794>
- Öter, G. (2023). *X-ray görüntülerinden covid-19 tespiti için derin öğrenme tabanlı modellerin geliştirilmesi* (Yayın Nu. 790127) [Yüksek lisans tezi, Burdur Mehmet Akif Ersoy Üniversitesi].
- Özcan, T. (2020). A deep learning framework for coronavirus disease (COVID-19) Detection in X-Ray Images. *Research Square*. doi:10.21203/rs.3.rs-26500/v1.
- Özcan T. (2021). A new composite approach for Covid-19 detection in X-ray images using deep features. *Applied Soft Computing*, 111, 107669. doi: <https://doi.org/10.1016/j.asoc.2021.107669>
- Özyurt, F. (2019). Uzaktan Algılama Görüntülerinin Evrişimsel Sinir Ağları ve Komşuluk Bileşen Analizi Tabanlı Özniteliklerinin Sınıflandırılması. *Afyon Kocatepe Üniversitesi Fen ve Mühendislik Bilimleri Dergisi*, 19 (3), 669-675. doi: 10.35414/akufemubid.558311
- Polat, H. ve Özerdem, M. S. (2021). Derin transfer öğrenimi yaklaşımı ile kamusal alanda medikal maske kullanımının otomatik kontrolü. *Türk Doğa ve Fen Dergisi*, 10 (2), 191-198. doi: 10.46810/tdfd.948098
- Rahman, T., Chowdhury, M. E., Khandakar, A., Islam, K. R., Islam, K. F., Mahbub, Z. B., M.A., Kadir ve Kashem, S. (2020). Transfer learning with deep convolutional neural network (CNN) for pneumonia detection using chest X-ray. *Applied Sciences*, 10(9), 3233. doi: <https://doi.org/10.3390/app10093233>
- Sandler, M., Howard, A., Zhu, M., Zhmoginov, A. ve Chen, L. C. (2018). Mobilenetv2: Inverted residuals and linear bottlenecks. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, USA, 4510-4520. doi: 10.1109/CVPR.2018.00474.
- Simonyan, K. ve Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 1-14.

- Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V. ve Rabinovich, A. (2015). Going deeper with convolutions. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, USA, 1-9. doi: 10.1109/CVPR.2015.7298594
- Toğaçar, M., Ergen, B. ve Sertkaya, M. E. (2019). Zatürre hastalığının derin öğrenme modeli ile tespiti. *Fırat Üniversitesi Mühendislik Bilimleri Dergisi*, 31 (1), 223-230. <https://dergipark.org.tr/tr/pub/fumbd/issue/43638/498364>
- Tuncer, T., & Ertam, F. (2020). Neighborhood component analysis and reliefF based survival recognition methods for Hepatocellular carcinoma. *Physica A: Statistical Mechanics and its Applications*, 540, 123143. 10.1016/j.physa.2019.123143.
- Yang, W., Wang, K. ve Zuo, W. (2012). Neighborhood component feature selection for high-dimensional data. *Journal of Computers*, 7(1), 161-168.