



Derin Öğrenme ile Pnömoni Hastalığının Tespiti

Diagnosis of Pneumonia Via Deep Learning Algorithms

Gamze AKPOLAT

Türk Ekonomi Bankası, IT Veri Yönetimi Bölümü
İstanbul, Türkiye
gamzeakpolat@hotmail.com

Hülya YALÇIN

İTÜ Elektronik ve Haberleşme Mühendisliği Anabilim Dalı
İstanbul, Türkiye
hulyayalcin@itu.edu.tr

Özetçe— Pnömoni hastalığı dünya da en yaygın ve ölümlü sonuçlanabilir hastalıklar arasında yer almaktadır. Erken tanı ölüm oranını azaltmadaki en önemli etkidir. Teşhis de en çok tercih edilen yöntem ise görüntüleme tekniklerinden biri olan göğüs röntgenleridir. Göğüs röntgenleri ile hastalık tek başına çok büyük oranda tespit edilebilmektedir. Teşhiste doğruluk ve hızın artmasına katkıda bulunmak için birçok mühendislik uygulamalarına başvurulmaktadır. Bu uygulamalardan biri de makine öğrenmesi yöntemleri ile hastalık teşhis etmektir. Çalışmada NIH klinik merkezinin kamu ile paylaştığı kapsamlı göğüs röntgeni veri seti kullanılmıştır. İlgili veri setinde pnömoni hastalığı teşhisi üzerinde çalışılmıştır. CNN ve VGG16 derin öğrenme algoritmaları incelenmiş olup çalışmada Python programlama dili kullanılmıştır. Derin öğrenmedeki gizli katman sayısını arttırmak, bir noktaya kadar olumlu bir etki yapsa da sonrasında parametre artışından dolayı sınıflandırma hızını düşürmesi nedeni ile performansı olumsuz yönde etkilemektedir. İncelenen yöntemler ile derin öğrenmenin sınıflandırma konusundaki alternatif modelleri, özellik haritaları ve başarımları sunulmuştur.

Anahtar Kelimeler — derin öğrenme, pnömoni hastalığı, makine öğrenmesi, göğüs röntgeni, akciğer hastalığı, sinir ağları, cnn ile hastalık tespiti, vgg16 ile hastalık tespiti, görüntü işleme, zatürre

Abstract— Pneumonia is one of the most common and fatal diseases in the world. Early diagnosis is the most important factor in reducing mortality. Chest X-rays, which are one of the imaging techniques, are the most commonly used diagnostic methods. The disease alone can be detected to a great extent by chest x-ray. Many engineering applications are used to help improve accuracy and speed in diagnostics. One of these applications is to diagnose disease with machine learning methods. In this study, a comprehensive chest x-ray data set shared with the public by the NIH clinic center was used. Pneumonia diagnosis was studied in the related data set. CNN and VGG16 deep learning algorithms were examined and Python programming language was used in the study. Increasing the number of hidden layers in deep learning has a positive effect up to a point, but it has a negative effect on performance as it decreases the classification rate due to the increase in parameters. Alternative models of classification, feature maps

and performance results of classification of deep learning are presented with the examined methods.

Keywords — deep learning, pneumonia disease, machine learning, chest x-ray, lung disease, neural networks, disease detection with cnn, disease detection with vgg16, image processing, pneumonia.

I. GİRİŞ

Röntgen teknolojisi vücudun birçok bölümünü görüntüleme ve incelemede kullanılır. Bunlardan biri de göğüs röntgenleri ile akciğer sorunlarını tespit etmedir. Zatürre (Akciğer İltihabı), tüberküloz veya akciğer kanseri gibi sorunların belirtileri göğüs röntgenlerinde görülebilmektedir.

Halk arasında zatürre olarak bilinen akciğer iltihaplanmasının tıbbi ismi pnömonidir [1]. Esasen, akciğerin en uç yapısını oluşturan ve havadaki oksijenin vücuda girdiği bölge olan hava keseciklerinin iltihaplanması hastalığıdır [2].

Pnömoni, Türkiye'de ölüme sebep olan hastalıklar arasında beşinci sırada yer alırken enfeksiyona bağlı ölümler arasında birinci sıradadır [1]. Aynı şekilde bu hastalık dünya çapında yaklaşık 450 milyon kişiyi etkilemektedir (nüfusun% 7'si) ve yılda yaklaşık 4 milyon ölümlü sonuçlanmaktadır [1, 2].

Bu sağlık sorununun tedavisi için erken tanı büyük önem taşımaktadır [8]. Teşhiste doğruluk ve hızın artmasına katkıda bulunmak için birçok mühendislik uygulamalarına başvurulmaktadır. Bu uygulamalardan biri de makine öğrenmesi ile hastalık teşhis etmektir.

Bu bildiri çalışmasında da derin öğrenme metotları kullanılarak göğüs röntgen görüntülerinden pnömoni hastalığını teşhis etmek için, NIH klinik merkezinin sağladığı veriler kullanılarak bir sınıflandırma modeli oluşturulmuştur. Bu model oluşturulurken aynı veri seti üzerinde yapılmış çalışmalar incelenmiş ve o çalışmaların üzerine ek çalışmalar yapılarak başarımları artırılmış ve modellerin özellik haritaları da çıkarılmıştır. Görüntü sınıflandırmada sık tercih edilen CNN ve VGG16 sinir ağları kullanılarak sıfırdan bir model oluşturulmuş ve Imagenet veri seti üzerinde eğitilmiştir.

II. İLGİLİ ÇALIŞMALAR

Pnömoni hastalığını teşhis etmek için daha önce çalışmalar yapılmıştır. İlgili veri seti üzerinde yapılan çalışmaların ilki Wang'a aittir [3]. Wang çalışmasında veri setinin %70 'ini eğitim, %10'unu validasyon ve %20'sini test için kullanmıştır. Yöntem olarak Konvolüsyonel Sinir Ağları'nı (CNN) uygulamış, başarı oranını arttırmak için daha sonra RESNet-50 üzerine yoğunlaşmıştır. Eğitim setini ImageNet üzerinden eğitmiştir. Çalışmalarında önceden eğitilmiş modellerden AlexNet, GoogLeNet, VGGNet-16 ve ResNet-50 kullanmış, sonuçlarını karşılaştırmıştır. AlexNet ile 0.54, GoogLeNet ile 0.59, VGGNet-16 ile 0.51 ve en son ResNet-50 ile 0.63 başarı oranı yakalamıştır [3]. Wang'dan sonra aynı veri seti üzerinden Li Yao çalışma yapmıştır [4]. Çalışmasında Tekrarlayan Sinir Ağları'nı (RNN) kullanmıştır. Yao'da Wang gibi eğitim setini rastgele olarak %70 'ini eğitim, %10'unu validasyon ve %20'sini test için ayırmıştır. Modelini OpenI veri setinde eğitmiştir. Sigmoid aktivasyon fonksiyonunu kullanmıştır. Li Yao'da kendisinden önce aynı veri setinde çalışan Wang'ın çalışmasını inceleyip, kendi bulguları ile karşılaştırmıştır [3, 4]. Li Yao çalışmaları sonucunda 0.71 başarı oranı yakalamıştır [4]. Yine aynı veri seti üzerinde Pranav Rajpurkar'da çalışmıştır [5]. Konvolüsyonel Sinir Ağı'nın gelişmiş versiyonu olan Yoğun Evrişim Ağı (Dense Convolutional Network (DenseNet)) ile oluşturulan modelini CheXNet olarak adlandırmıştır [5]. CheXNet modelinde sigmoid aktivasyon fonksiyonu kullanılmıştır [5]. Ağırlıklar, ImageNet'te önceden tanımlanmış bir modelin ağırlıkları ile başlatılmıştır. İyileştirici fonksiyon (optimizer) olarak ADAM kullanılmıştır. Grup sayısı (batch size) 16, öğrenme oranı (learning rate) 0.001 ve epoch değerini 10 almıştır. Bu parametreler ışığında elde ettiği başarı oranı 0.76 olmuştur [5]. Rajpurkar'da bu başarı oranını yakalayabilmek için kendisinden önceki ve yukarıda aktarılan Wang ve Li Yao'nun çalışmalarını incelemiş ve bu çalışmalardan elde edilen başarı oranının üzerine çıkabileceği bir model oluşturmuştur.

III. VERİ SETİNİ İNCELEME VE MODELLEME

A. Veri Seti

Veri kümesi Pnömoni hastalığına sahip olan ve olmayan insanların akciğer röntgen görüntülerini içermektedir. Veri kümesindeki görüntüler yüksek çözünürlüklü olup, jpeg formatındadır. Yaklaşık 6000 görüntü içermekte olup, toplam klasör boyutu 5,5 GB'ın üzerindedir [6]. Önceki çalışmalarda olduğu gibi veri setinin %70 'ini eğitim, %10'unu validasyon ve %20'si test için kullanılmıştır. Görüntülerin bir kısmının kanal boyutları bozulmuştur. İlgili veri kümesinin tamamının kullanılabilmesi veri temizleme aşamasında tek boyutlu (gray scale) olarak verilen kanal boyutları, çoklu kanal yapısına (multi-channel) dönüştürülmüştür. Veri kümesi The NIH Clinical Center tarafından paylaşılmıştır [6]. Paylaşılan veri setinde Pnömoni ile normal-sağlıklı akciğer görüntüleri

bulunmaktadır. Veri kümesi içerisinde, pnömoni olanlar "1", normal-sağlıklı olanlar "0" şeklinde etiketlenmiştir.

Veri kümesindeki röntgen görüntüleri ilgili klinik doktorları tarafından etiketlenmektedir [6]. Etiketlenen görüntü sayısı arttıkça yayınlanan veri setinin sürümleri ve boyutları değişmektedir.

Medikal veri setlerinde genelde karşılaşılmakta olan veri dağılım dengesizliği sorunu bu veri kümesinde de yaşanmıştır.



Şekil 1. Veri Kümesi dağılım durumu

Şekil 1'de üzerinde çalışılan veri kümesinin dağılımı gösterilmiştir, mavi renkli küme normal-sağlıklı, turuncu ise pnömoninin dağılımıdır. Etiketler arasında yaklaşık 2,8 kat fark bulunmaktadır. Kurulacak modelin doğru çalışabilmesi için sağlıklı veriler üzerinden sentetik veri artırma işlemi uygulanmıştır.

Şekil 2'de ise veri kümesi içerisinde alınmış, sol tarafta pnömoni, sağ tarafta sağlıklı bir akciğer röntgen görüntüsü bulunmaktadır.



Şekil 2. Pnömoni-Sağlıklı Akciğer Röntgen Görüntüleri

B. Modeller

Bu bildiri çalışmasında iki farklı model kullanılmıştır. Modellerde, kaydırma adımı 1, piksel ekleme 0, öğrenme oranı 0.001, erken durdurma değeri 5, grup sayısı (batch size) olarak ise 16 değeri verilmiştir. Eğitim tur sayısı ise 20 epoch olarak belirlenmiştir. Optimizasyon fonksiyonu olarak Adam, aktivasyon fonksiyonu olarak ise Relu-Softmax kullanılmıştır. Her iki model de Imagenet veri seti üzerinden eğitilmiştir. CNN modeli ilk tamamen bağlı katmanında 0.70, ikinci tamamen bağlı katmanında ise 0.50 oranında veri seyreltme uygulanmıştır. Ayrıca CNN modelinde havuzlama olarak sadece max-pooling kullanılmıştır. VGG16 modelinde ise

havuzlama olarak hem max-pooling hem de global average pooling uygulanmıştır. Modellerin başarımlarını ölçtüğü için aşağıdaki oran ve hesaplamalardan faydalanılmıştır.

TABLO I. BAŞARI ÖLÇÜM OLASILIKLARI

		Tahmin Edilen	
		Olumsuz	Olumlu
Gerçek Durum	Olumsuz	Pozitif Olumsuz (tn)	Negatif Olumlu (fp)
	Olumlu	Negatif Olumsuz (fn)	Pozitif Olumlu (tp)

Denklem (1)'de doğruluk, makine öğrenme modelinin ne kadar iyi olduğunu inceleyen bir ölçüdür. [7].

$$\text{Doğruluk, ACC} = \frac{tp + tn}{tp + tn + fn + fp} \quad (1)$$

Denklem (2), Gerçek Pozitif Oran (Recall, (TPR)), pozitif tahminlerin kaç tanesinin doğru olduğunu ölçüsü olarak tanımlanmaktadır [7].

$$\text{Hassasiyet, TPR} = \frac{tp}{tp + fn} \quad (2)$$

Denklem (3), Pozitif Öngörü Değeri (Precision, positive predictive value (PPV)), toplam pozitif durumdan, kaç tane pozitifin doğru tahmin edildiğini ölçer [7].

$$\text{Pozitif Öngörü Değeri, PPV} = \frac{tp}{tp + fp} \quad (3)$$

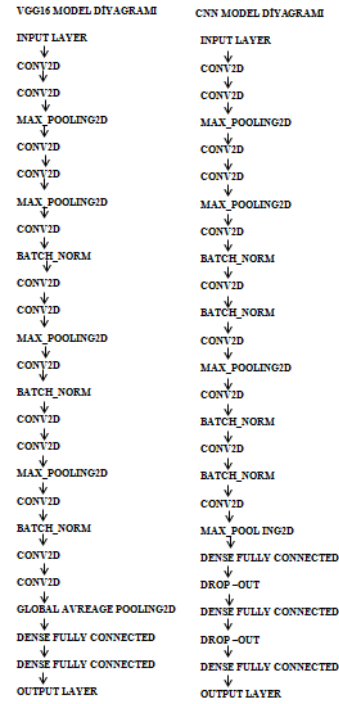
Denklem (4), F-skor değerlendirme ölçütü hem pozitif öngörü değerini hem de gerçek pozitif oranını sınıflandırma etkinliğinin bir ölçüsü olarak birleştirir [7].

$$F_1 = \frac{2 \cdot PPV \cdot TPR}{PPV + TPR} = \frac{2 \cdot tp}{2tp + fp + fn} \quad (4)$$

C. Modellerin Akış Diyagramı

Her iki modelde de aşırı öğrenmeyi engellemek için havuzlama, toplu normalleştirme (batch norm), rasgele veri seyreltme (drop out), erken durdurma vb. gibi yöntemler uygulanmıştır [9, 10]. Toplu normalleştirme, sinir ağlarının performansını ve dengesini arttırmak için uygulanan bir tekniktir [10,11]. Toplu normalleştirmenin arkasındaki asıl amaç ağ eğitimini optimize etmektir. Toplu normalleştirme, gizli birim değerlerinin çevresinde kaydığı miktarı (kovaryans değişimi) azaltır [14,15]. Ayrıca, ağın her katmanının diğer katmanlardan bağımsız olarak biraz daha fazla öğrenmesini sağlamaktadır [11, 12, 13].

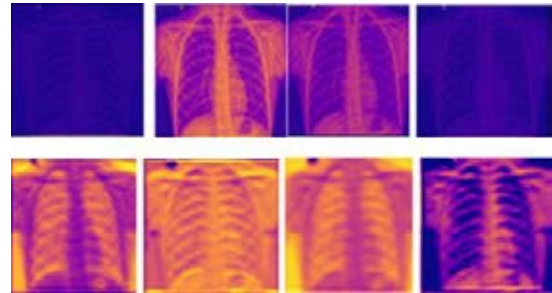
Şekil 3'te oluşturulan modellerin diyagramı verilmiştir.



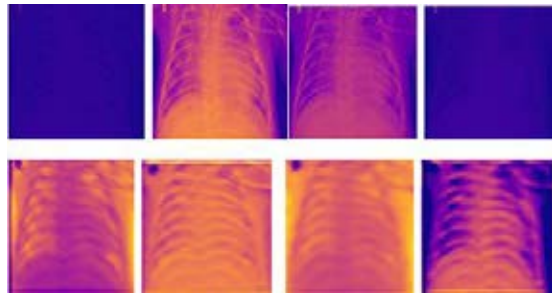
Şekil 3. Modellerin Diyagramı

D. Modellerin Aktivasyon –Özellik Haritaları

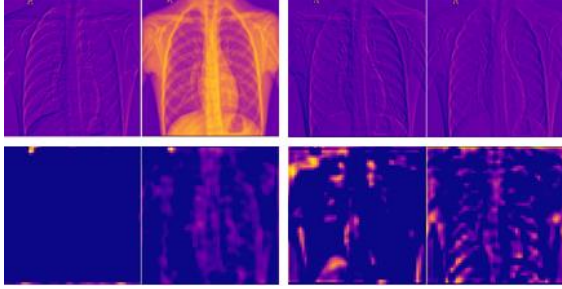
Her iki model için aşağıdaki Şekil 4,5,6,7'de sağlıklı ve pnömoni görüntülerinin birinci ve dokuzuncu katmanlardaki filtrelerin özellik haritası çıkarılmıştır.



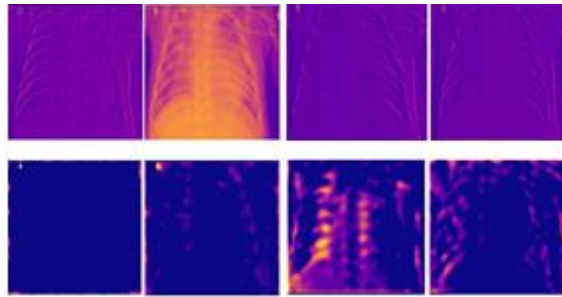
Şekil 4. CNN modelindeki sağlıklı akciğer röntgen görüntüsünün birinci ve dokuzuncu katmandaki filtre çıktıları



Şekil 5. CNN modelindeki pnömonili akciğer röntgen görüntüsünün birinci ve dokuzuncu katmandaki filtre çıktıları



Şekil 6. VGG16 modelindeki sağlıklı akciğer röntgen görüntüsünün birinci ve dokuzuncu katmandaki filtre çıktıları



Şekil 7. VGG16 modelindeki pnömonili akciğer röntgen görüntüsünün birinci ve dokuzuncu katmandaki filtre çıktıları

IV. DENEY VE SONUÇLAR

Çalışılan her iki model ile daha önce gerçekleştirilmiş çalışmaların başarı oranlarının üzerine çıkmıştır. Ayrıca veri seti içerisinde bulunan tek kanallı verilerin çalışma veri setine dahil edilmesi için boyut dönüşümleri yapılmıştır. Az sayıda görüntüyü sınıflandırmak için kullandığımız algoritmalar, başarılı sonuçlar elde etmiştir. Eğitim süreci yaklaşık olarak 12 saat sürmüştür. Görüntü ya da modeldeki katman sayısını arttırmak doğru sonuçlar bulmak açısından olumlu sonuç vermektedir. Fakat veri ya da katman arttıkça eğitim süresi de artış göstermektedir. Bu da verimliliği düşürmektedir. Performans sorununun önüne geçebilmek için gpu programlama tekniklerinin kullanılması önerilmektedir. Bu çalışmada Colab gpu kullanılmıştır. TabloII ve III'te bu bildiri çalışmasının detayı ve başarı oranları gösterilmiştir.

TABLO II. YÖNTEMLER VE KARŞILAŞTIRMALARI

YÖNTEMLER				
WANG	LI YAO	RAJPURKAR	Deney-I	Deney-II
CNN	RNN	CNN	CNN	VGG-16
Sigmoid	Sigmoid	Sigmoid	ReLU - Softmax	ReLU - Softmax
ADAM	ADAM	ADAM	ADAM	ADAM
ImageNet	Openl	ImageNet	ImageNet	ImageNet

TABLO III. SONUÇLAR VE KARŞILAŞTIRMALARI

	WANG	LI YAO	RAJPURKAR	Deney-I	Deney-II
Acc	~ 0.63	~ 0.71	~ 0.76	~ 0.83	~ 0.87
Precision	~ 0.66	Belirtilmemiş	Belirtilmemiş	~ 0.81	~ 0.84
Recall	~ 0.93	Belirtilmemiş	Belirtilmemiş	~ 0.98	~ 0.98
F1 Skor	~ 0.77	Belirtilmemiş	Belirtilmemiş	~ 0.88	~ 0.90

BİLGİLENDİRME

Bu bildiri çalışması “Derin öğrenme ile Pnömoni Hastalığının Tespiti” adlı İTÜ Elektronik ve Haberleşme Mühendisliği Anabilim Dalı, yüksek lisans tez çalışmamdan alınmıştır.

KAYNAKLAR

- [1] Tevfik ÖZLÜ, Yılmaz BÜLBÜL, Savaş ÖZSU. Ulusal verilerle toplum kökenli pnömoniler, *Tüberküloz ve Toraks Dergisi* 2007; 55(2): 191-212
- [2] Aydogdu, M, Ozyilmaz, E, Aksoy, Handan, Gursel, G, and Ekim, Numan. Mortality prediction in community-acquired pneumonia requiring mechanical ventilation; values of pneumonia and intensive care unit severity scores. *Tuberk Toraks*, 58(1):25-34, 2010.
- [3] Wang, Xiaosong, Peng, Yifan, Lu, Le, Lu, Zhiyong, Bagheri, Mohammadhadi, and Summers, Ronald M. Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. *arXiv preprint arXiv:1705.02315*, 2017.
- [4] Yao, Li, Poblentz, Eric, Dagunts, Dmitry, Covington, Ben, Bernard, Devon, and Lyman, Kevin. Learning to diagnose from scratch by exploiting dependencies among labels. *arXiv preprint arXiv:1710.10501*, 2017.
- [5] Pranav Rajpurkar, Jeremy Irvin, Kaylie Zhu, Brandon Yang, Hershel Mehta, Tony Duan, Daisy Ding, Aarti Bagul, Robyn L. Ball, Curtis Langlotz, Katie Shpanskaya, Matthew P. Lungren, Andrew Y. Ng. CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning *arXiv:1711.05225v3*, 2017.
- [6] The NIH Clinical Center, <https://www.nih.gov/news-events/news-releases/nih-clinical-center-provides-one-largest-publicly-available-chest-x-ray-datasets-scientific-community>, 2019
- [7] Andriy Burkov - *The Hundred-Page Machine Learning Book*
- [8] *Who Is at Risk for Pneumonia?*. NHLBI. 1 March 2011. Archived from the original on 7 March 2016. Retrieved 3 March 2016.
- [9] Lutz Prechelt- *Early Stopping But When?*- https://page.mi.fu-berlin.de/prechelt/Biblio/stop_tricks1997.pdf
- [10] Sergey Ioffe , Christian Szegedy. *Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift*, 2015.
- [11] CS231n: *Convolutional Neural Networks for Visual Recognition*, 2017, “Neural Networks Part 3: Learning and Evaluation”
- [12] Ashia C. Wilson, Rebecca Roelofs, Mitchell Stern, Nathan Srebro, Benjamin Recht, 2017, “The Marginal Value of Adaptive Gradient Methods in Machine Learning”
- [13] Serwa A, *Studying the Effect of Activation Function on Classification Accuracy Using Deep Artificial Neural Networks*, *Journal of Remote Sensing & GIS* 6: 203., July 2017.
- [14] Diederik P. Kingma, Jimmy Ba. *Adam: A Method for Stochastic Optimization*, 2017
- [15] Tim Salimans, Durk Kingma. *Weight Normalization: A Simple Reparameterization to Accelerate Training of DNN*, 2016.