



COPD Severity Classification from Respiratory Sounds Using Mel-Spectrogram-Based Transfer Learning

Fatih Kuşçu^{1*}

¹ Department of Computer Engineering, Iskenderun Technical University, Hatay, Türkiye.

ABSTRACT

Accurate and rapid classification of Chronic Obstructive Pulmonary Disease (COPD) severity is highly important for computer-aided diagnosis systems and digital health applications. In this study, Mel-Spectrogram-based time-frequency representations and transfer learning approaches were employed for the automatic classification of COPD severity levels from respiratory sounds. Respiratory audio recordings from the open-access RespiratoryDatabase@TR dataset were first resampled to a standard sampling frequency and then converted into two-dimensional visual representations using the Short-Time Fourier Transform-based Mel-Spectrogram method. The obtained spectrogram images were resized according to the input requirements of pretrained convolutional neural network architectures. In the experimental study, AlexNet, MobileNet, ResNet, and EfficientNet models were trained using transfer learning, and their performances were evaluated using accuracy, precision, recall, and F1-score metrics. In addition, a Random Forest-based baseline model was implemented on raw audio signals to provide a comparative evaluation against traditional machine learning methods. The experimental findings demonstrate that Mel-Spectrogram representations effectively reveal discriminative pathological patterns in respiratory sounds and that pretrained CNN models achieve high performance in COPD severity classification. According to the experimental results, the best performance was achieved by EfficientNet, reaching an accuracy of 0.9340. These results indicate that image-based deep learning analysis of respiratory sounds is an effective and applicable approach for automatic COPD severity classification.

ARTICLE INFO

Received 2026-03-22,

Accepted 2026-06-12,

Publication Date 2026-06-30

Keywords:

COPD Severity,
Respiratory Sounds, Mel-
Spectrogram,
Convolutional Neural
Networks, Transfer
Learning, Segment-Level
Split

Distributed Under CC-BY 4.0



*Corresponding Author: Fatih KUŞÇU, E-mail: fatihkuscu.lee21@iste.edu.tr

Mel-Spektrogram Tabanlı Transfer Öğrenimi Kullanılarak Solunum Seslerinden KOAH Şiddetinin Sınıflandırılması

ÖZET

Kronik Obstrüktif Akciğer Hastalığı'nın (KOAH) şiddet düzeyinin hızlı ve doğru biçimde belirlenmesi, bilgisayar destekli tanı sistemleri ve dijital sağlık uygulamaları açısından büyük önem taşımaktadır. Bu çalışmada, solunum seslerinden KOAH şiddet seviyelerinin otomatik olarak sınıflandırılması amacıyla Mel-Spektrogram tabanlı zaman-frekans temsilleri ve transfer öğrenme yaklaşımları kullanılmıştır. Açık erişimli RespiratoryDatabase@TR veri setinde yer alan solunum sesleri öncelikle standart örnekleme frekansına dönüştürülmüş, ardından Kısa Zamanlı Fourier Dönüşümü tabanlı Mel-Spektrogram yöntemiyle iki boyutlu görüntü temsillerine çevrilmiştir. Elde edilen spektrogram görüntüleri, önceden eğitilmiş evrişimli sinir ağı mimarilerinin giriş formatına uygun olacak şekilde yeniden boyutlandırılmıştır. Deneyisel çalışmada AlexNet, MobileNet, ResNet ve EfficientNet modelleri transfer öğrenme yaklaşımıyla eğitilmiş; modellerin başarıları accuracy, precision, recall ve F1-score metrikleri üzerinden karşılaştırılmıştır. Geleneksel makine öğrenmesi yaklaşımının karşılaştırmalı etkisini göstermek amacıyla ham ses verileri üzerinde Random Forest tabanlı bir temel model de kullanılmıştır. Elde edilen bulgular, Mel-Spektrogram temsillerinin solunum seslerindeki ayırt edici patolojik örüntüleri etkili biçimde ortaya çıkardığını ve transfer öğrenme tabanlı CNN modellerinin KOAH şiddet sınıflandırmasında yüksek başarı sağladığını göstermektedir. Deneyisel sonuçlara göre en başarılı model EfficientNet olmuş ve 0.9340 doğruluk değerine ulaşmıştır. Bu sonuçlar, solunum seslerinin görüntü tabanlı derin öğrenme yöntemleriyle analiz edilmesinin KOAH şiddetinin otomatik belirlenmesinde etkili ve uygulanabilir bir yaklaşım olduğunu ortaya koymaktadır.

MAKALE BİLGİSİ

Keywords:

KOAH Şiddeti, Solunum Sesleri, Mel-Spektrogram, Evrişimli Sinir Ağları, Transfer Öğrenme, Segment Bazlı Bölünme , Pretrained Modeller

Distributed Under CC-BY 4.0



GİRİŞ

Kronik Obstrüktif Akciğer Hastalığı (KOAH), dünya genelinde yüksek morbidite ve mortalite oranlarına sahip önemli bir solunum yolu hastalığıdır. Hastalığın erken teşhisi ve şiddet düzeyinin doğru biçimde belirlenmesi, tedavi planlaması, ilaç dozajının ayarlanması ve hastanın yaşam kalitesinin iyileştirilmesi açısından kritik öneme sahiptir (Altan ve ark., 2019). KOAH şiddetinin beş evrede (COPD0–COPD4) sınıflandırılması, klinisyenlerin hastalığın ilerleyişini izlemesine ve kişiselleştirilmiş müdahale stratejileri geliştirmesine olanak tanır (Altan ve ark., 2020).

Solunum sesleri, stetoskop veya dijital kayıt cihazlarıyla elde edilebilen, non-invaziv, düşük maliyetli ve dijital sağlık sistemlerine entegre edilebilir bir veri kaynağıdır. Geleneksel oskültasyon yöntemleri uzman hekim deneyimine bağımlıdır ve öznel değerlendirme riski taşır. Bu nedenle solunum seslerinin otomatik analizine yönelik bilgisayar destekli tanı çalışmaları son yıllarda önemli bir araştırma alanı haline gelmiştir (Acharya ve ark., 2023).

Bir boyutlu ham ses sinyalleri, zaman içinde değişen ve durağan olmayan karmaşık yapıları nedeniyle klasik makine öğrenmesi yöntemleriyle analiz edildiğinde sınırlı ayırt edicilik sunabilmektedir. Solunum seslerindeki patolojik örüntüler, frekans bileşenlerinde ve genlik değişimlerinde kendini gösterir; bu nedenle yalnızca zaman alanında yapılan analizler çoğu durumda yetersiz kalmaktadır (Sahu ve ark., 2024). Bu sorunu aşmak amacıyla Mel-Spektrogram gibi zaman-frekans dönüşümleri yaygın biçimde kullanılmaktadır. Mel-Spektrogram dönüşümü, solunum sesindeki frekans değişimlerini, patolojik örüntüleri ve akustik farklılıkları iki boyutlu bir görüntü yapısına dönüştürerek problemi bir görüntü sınıflandırma görevine indirger (Roy ve ark., 2025; MDPI Researchers, 2024; Alve ve ark., 2024).

Transfer öğrenme yaklaşımı, ImageNet gibi büyük ölçekli veri kümelerinde önceden eğitilmiş evrişimli sinir ağı (CNN) mimarilerinin, sınırlı örnekleme sahip biyomedikal veri setlerinde güçlü özellik çıkarıcılar olarak kullanılmasını mümkün kılar (Khanaghavalle ve ark., 2024; Semantic Scholar, 2024). Bu çalışmada AlexNet, ResNet18, MobileNetV2 ve EfficientNet-B0 modelleri, farklı mimari özellikleri ve hesaplama verimlilikleri bakımından karşılaştırmalı olarak değerlendirilmiştir.

Bu çalışmanın temel amacı, RespiratoryDatabase@TR veri setindeki solunum seslerinden elde edilen Mel-Spektrogram görüntülerini kullanarak KOAH şiddet seviyelerini COPD0, COPD1, COPD2, COPD3 ve COPD4 olmak üzere beş sınıfta sınıflandırmak ve önceden eğitilmiş CNN modellerinin performansını karşılaştırmaktır. Ayrıca ham ses sinyalleri üzerinde Random Forest modeli kullanılarak geleneksel makine öğrenmesi yaklaşımıyla derin öğrenme tabanlı görüntü sınıflandırma yaklaşımı karşılaştırılmıştır.

Literatür Taraması

Solunum sesleri üzerinden KOAH teşhisi ve hastalık şiddetinin otomatik analizi amacıyla literatürde gerçekleştirilen ilk önemli çalışmalardan biri, 2025 yılında Abhinav Roy ve arkadaşları tarafından sunulmuştur. Bu çalışmada, solunum seslerinin analizi yoluyla KOAH tespiti için melez bir yaklaşım önerilmiştir. Çalışmada RespiratoryDatabase@TR, ICBHI, King Abdullah ve AIIMS veri setlerinin birleştirilmesiyle multimodal (çok modlu) ve geniş kapsamlı bir veri havuzu oluşturulmuştur. Akustik sinyaller öncelikle yeniden örnekleme işlemlerine tabi tutulmuş; ardından zaman-frekans uzayındaki ayırt edici özellikleri yakalamak adına Mel-Spektrogram (MSpec), Kromagram ve MFCC öznitelik çıkarım teknikleri uygulanmıştır. Sınıflandırma modeli olarak ise son dönemlerin yenilikçi mimarilerinden olan Hibrit Kolmogorov-Arnold Networks (KAN) yöntemi tercih edilmiştir. Detaylı performans analizleri sonucunda önerilen bu KAN tabanlı modelin KOAH tespitinde %93.00 doğruluk (Accuracy), 0.98 F1-skoru, 0.98 AUC ve %97.4 hassasiyet (Sensitivity) değerlerine ulaşarak literatürdeki geleneksel ağlara karşı belirgin bir üstünlük sağladığı raporlanmıştır (Roy ve ark., 2025).

İkinci çalışma, solunum seslerinin akustik niteliklerini derin öğrenme modelleriyle analiz ederek KOAH şiddetinin evrenlenmesini hedeflemektedir. Bu kapsamda Prakash Sahu ve meslektaşları tarafından geliştirilen çalışmada, RespiratoryDatabase@TR veri seti ve çeşitli açık kaynaklı solunum kütüphaneleri bir araya getirilmiştir. Araştırmada ham sinyaller ilk olarak gelişmiş sinyal filtreleme ve zaman-frekans ölçeklendirmesi süreçlerinden geçirilerek ses spektrogramları haline getirilmiştir. Sınıflandırma görevi için akustik verilerin frekans desenlerine özel olarak optimize edilmiş evrimsel bir mimari olan SOUNDNet sinir ağı modeli tasarlanmıştır. Deneysel testler sonucunda, geliştirilen SOUNDNet mimarisinin KOAH şiddet sınıflandırmasında yaklaşık %88.50 doğruluk ve %89.20 F1-skor seviyelerine ulaştığı, aynı zamanda klinik teşhis sistemleri için kritik bir parametre olan yüksek özgüllük (Specificity) seviyesini kararlı bir şekilde koruduğu gösterilmiştir (Sahu ve ark., 2024).

Bir diğer çalışmada, çok kanallı akciğer sesleri üzerinden KOAH evrelemesi yapmak amacıyla derin evrimsel sinir ağlarının başarımı test edilmiştir. G. R. Khanaghavalle ve çalışma arkadaşları tarafından yürütülen çalışmada, 12 kanallı kayıt dinamiklerine sahip RespiratoryDatabase@TR veri seti esas alınmıştır. Sınırlı klinik veri problemini aşmak amacıyla veri artırma (Data augmentation) tekniklerinden faydalanılmış ve ham ses kesitleri spektrogram, Mel-Spektrogram ve Kromagram zaman-frekans matrislerine dönüştürülmüştür. Elde edilen 2 boyutlu görsel haritalar, 50 katmanlı derin bir evrimsel sinir ağı olan ResNet50 mimarisi transfer öğrenme yoluyla eğitilerek sınıflandırılmıştır. Modelin karmaşıklık analizleri incelendiğinde, KOAH seviyelerinin ayırt edilmesinde özellikle ağır evre KOAH sınıflandırma doğruluğunun %89.13'e ulaştığı, orta derece ile ağır derece KOAH vakalarının birbirlerinden ayrıştırılmasında ise %94.26 gibi yüksek bir başarıma ulaşıldığı raporlanmıştır (Khanaghavalle ve ark., 2024).

Solunum akustik sinyallerinin derin öğrenme tabanlı otomatik analizinde çığır açan temel referans çalışmalardan biri, 2020 yılında Gökhan Altan, Yakup Kutlu ve Ahmet Gökçen tarafından gerçekleştirilmiştir. Bu çalışmada, RespiratoryDatabase@TR veri setinde yer alan 41 hastaya ait

12 kanallı ve 5 alt şiddet evresini (COPD0-COPD4) barındıran tüm kayıtlar derinlemesine incelenmiştir. Akustik sinyallerin uzamsal ve frekans bazlı temsillerini çıkarabilmek amacıyla 3 Boyutlu İkinci Dereceden Fark Grafiği (3D-SODP), Küboid ve Oktant Kuantizasyonu adı verilen yenilikçi öznitelik çıkarma yöntemleri önerilmiştir. Elde edilen yapısal öznitelikler, Derin Ekstrem Öğrenme Makineleri mimarisine dayanan ve oto-kodlayıcı çekirdekleri barındıran LuELM-AE ile HessELM-AE modelleriyle sınıflandırılmıştır. Yapılan deneysel simülasyonlarda, 5 alt şiddet evresinin eş zamanlı sınıflandırılmasında LuELM-AE modelinin %94.31 genel doğruluk oranı, 0.9659 AUC değeri ve %98.76 gibi olağanüstü bir ağırlıklı özgüllük (W-SPE) başarısı gösterdiği kanıtlanmıştır (Altan ve ark., 2020).

Aynı araştırma grubunun bir önceki yıl yayımladığı 2019 tarihli çalışmada ise, KOAH analizi bilgisayarlı sistemler ve denetimsiz derin öğrenme yaklaşımlarıyla ele alınmıştır. Yine RespiratoryDatabase@TR veri setinin kullanıldığı bu çalışmada, 1B ses sinyallerinin çözülmesi için Hilbert-Huang Dönüşümü (HHT) metodu uygulanmış ve genlik-frekans dalgalanmalarından frekans modülasyonu istatistiksel öznitelikleri türetilmiştir. Sınıflandırma katmanında, denetimsiz ön eğitim (unsupervised pre-training) ve ardından denetimli optimizasyon (supervised fine-tuning) adımlarını içeren derin bir yapay sinir ağı (DL) kurgulanmıştır. Geliştirilen bilgisayarlı KOAH analiz sisteminin test sonuçlarında %93.67 doğruluk, %91.00 duyarlılık (Sensitivity) ve %96.33 özgüllük (Specificity) oranları elde edilerek derin ağların medikal seslerdeki genelleme kapasitesi doğrulanmıştır (Altan ve ark., 2019).

Gökhan Altan ve Yakup Kutlu tarafından 2018 yılında gerçekleştirilen bir diğer çalışmada ise solunum seslerindeki patolojik anormalliklerin tespiti hedeflenmiştir. RespiratoryDatabase@TR veri setinin ikili sınıflandırma (sağlıklı vaka ve KOAH hastası) odaklı varyasyonunun kullanıldığı bu çalışmada, seslerin geometrik yapısını ortaya koyan 3D-SODP kuantizasyon teknikleri ve uzamsal ayrıştırma öznitelikleri çıkarılmıştır. Elde edilen özellik matrisleri, üretici (generative) bir derin öğrenme modeli olan Derin İnanç Ağları (Deep Belief Networks - DBN) algoritması ile işlenmiştir. Modelin test aşamasında solunum sesleri üzerinden patoloji tespitinde son derece yüksek bir sınıflandırma başarımı, düşük varyans ve yüksek kararlılık elde edildiği belirtilmiştir (Altan & Kutlu, 2018).

Derin öğrenme literatüründe evrensel bazda kabul gören solunum çalışmalarına bakıldığında, zaman-frekans dönüşümlerinin CNN modellerindeki etkisi göze çarpmaktadır. 2024 yılında MDPI platformunda yayımlanan bir çalışmada, dünya genelinde yaygın olarak kullanılan evrensel ICBHI solunum veri seti üzerinde otomatik KOAH tespiti görevi yürütülmüştür. Zaman-frekans haritalama adımı olarak ham ses kayıtlarından Mel-Spektrogram öznitelikleri çıkarılmış ve bu görseller ECAPA-TDNN adı verilen akustik veri odaklı evrişimli sinir ağı varyantı ile eğitilmiştir. Gerçekleştirilen testler sonucunda, seslerin spektrogram resimleri üzerinden işlenmesinin ağırlık öğrenme kararlılığını optimize ettiği ve otomatik KOAH tespiti görevinde %96.80 ortalama doğruluk ile dengeli bir F1-skor değerine ulaştığı raporlanmıştır (MDPI Researchers, 2024).

Benzer bir yaklaşımla, medRxiv platformunda 2024 yılında yayımlanan klinik odaklı bir çalışmada, KOAH tahmin sürecini iyileştirmek adına Kısa Zamanlı Fourier Dönüşümü (STFT) tabanlı spektrogram görselleri ve Evrişimli Sinir Ağları (CNN) entegrasyonu sunulmuştur. Araştırmacılar tarafından yerel hastanelerden toplanan özel bir klinik solunum sesi veri setinin kullanıldığı bu çalışmada, 1B ham zaman serisi dalgaları ile bunların 2B görsel temsilleri doğrudan kıyaslanmıştır. Yapılan deneysel çıktılar, ham ses dalgası yerine sese ait frekans-zaman spektrogram resimlerini CNN modellerine girdi olarak vermenin, klinik karar destek süreçlerindeki hata payını minimize ettiğini ve model kesinliğini (Precision) doğrudan iyileştirdiğini istatistiksel verilerle ortaya koymuştur (Alve ve ark., 2024).

Klinik ses veri setlerinde sıkça karşılaşılan dengesiz sınıf dağılımları ve azınlık sınıflarının göz ardı edilmesi problemine yönelik olarak, 2023 yılında Acharya ve arkadaşları tarafından giyilebilir tıbbi cihazları hedefleyen bir çalışma yürütülmüştür. Sınıf dengesizliklerinin yoğun olduğu ICBHI 2017 solunum sesleri veri setinin kullanıldığı çalışmada, Mel-spektrogram tabanlı öznelilikler çıkarılmıştır. Sınıflandırma performansını artırmak amacıyla, ağın nadir görülen sınıflardaki hatalara daha büyük cezalar vermesini sağlayan odak kaybı (focal loss) fonksiyonu ile optimize edilmiş derin bir CNN-RNN hibrit sinir ağı modeli tasarlanmıştır. Elde edilen sonuçlar, odak kaybı optimizasyonunun veri setindeki asimetric dağılımlara rağmen modelin azınlık sınıflarını yakalama başarısını (Recall) artırdığını ve giyilebilir donanımlar için state-of-the-art düzeyinde bir sınıflandırma başarısı sağladığını göstermiştir (Acharya ve ark., 2023).

Son olarak, solunum anormalliklerinin uzaktan tespiti ve çok aşamalı sınıflandırma sistemleri üzerine Semantic Scholar platformunda 2024 yılında yer alan bir çalışmada, akustik verilerin frekans haritaları incelenmiştir. Pediatrik ve genel yetişkin akciğer sesleri veri setlerinin bir arada kullanıldığı bu çalışmada, ses sinyalleri üzerinde hem Skalogram (Sürekli Dalga Dönüşümü) hem de Spektrogram zaman-frekans dönüşümleri uygulanmıştır. Görselleştirilen bu veriler, evrişim katmanları ile dikkat mekanizmalarını birleştiren çok aşamalı hibrit bir CNN-Transformer çerçevesi üzerinden eğitilmiştir. Deneysel sonuçlar, ses sinyalinin görsel bir matrise dönüştürülerek hibrit ağlarda işlenmesi sayesinde, geleneksel akustik zaman serisi analiz yöntemlerine kıyasla patolojik anomalilerin yakalanmasında ve sınıf ayırım sınırlarının belirlenmesinde çok daha yüksek makro kesinlik (Precision) değerlerine ulaşıldığını doğrulamaktadır (Semantic Scholar, 2024).

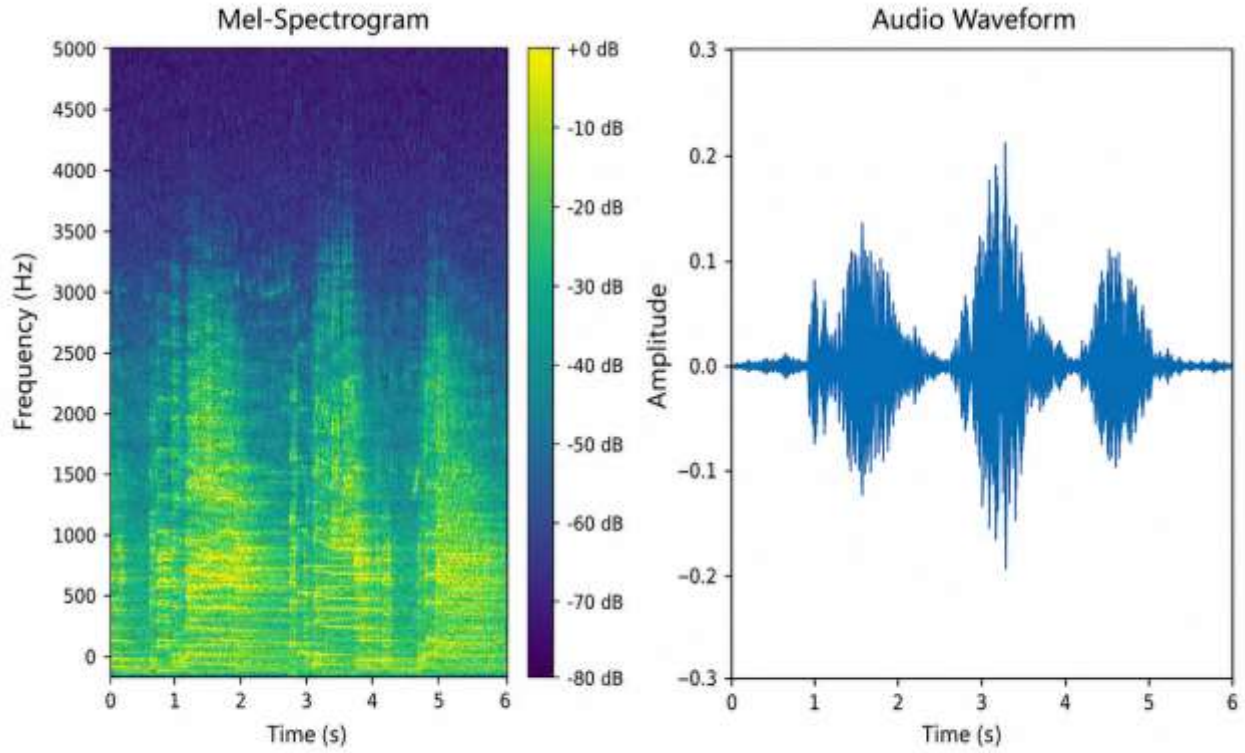
MATERYAL VE METHOD

Veri Seti ve Ön İşleme

Bu çalışmada, KOAH (Kronik Obstrüktif Akciğer Hastalığı) şiddetinin derin öğrenme tabanlı otomatik analizi amacıyla, Mendeley Data platformu üzerinde açık erişimli olarak yayımlanan RespiratoryDatabase@TR veri seti kullanılmıştır [4]. Veri setinde, katı klinik protokollerle kayıt altına alınmış 42 farklı hastadan elde edilen toplam 504 adet akustik solunum sesi (.wav formatında) dosyası yer almaktadır. Hastalar klinik evrelerine ve semptom şiddetlerine göre beş farklı sınıfa ayrılmıştır: COPD0 (Sağlıklı), COPD1 (Hafif KOAH), COPD2 (Orta KOAH), COPD3 (Ağır KOAH) ve COPD4 (Çok Ağır KOAH).

Orijinal haliyle 1 boyutlu (1B) zaman serisi olan akustik sinyaller, derin öğrenme modellerine uygun hale getirilmek için öncelikle 4 kHz örnekleme hızına (downsampling) getirilmiştir. Ardından, bilgisayarlı görü tekniklerinin güçlü öznelilik çıkarma kabiliyetlerinden faydalanabilmek amacıyla, Kısa Zamanlı Fourier Dönüşümü (STFT) tabanlı Mel-Spektrogram algoritması kullanılarak 2 boyutlu (2B) frekans-zaman görsellerine dönüştürülmüştür. Tüm spektrogram görselleri, önceden eğitilmiş CNN mimarilerinin girdi katmanlarına uyum sağlaması amacıyla 224x224 piksel boyutlarında ölçeklenmiştir. Deneysel kurgunun temelini oluşturan veri bölme işleminde, literatürdeki benzer çalışmalarla karşılaştırılabilir sonuçlar elde edebilmek amacıyla stratified segment-level train-test split yaklaşımı uygulanmıştır. Bu yaklaşımda Mel-Spektrogram görüntüleri, sınıf dağılımları korunacak şekilde %80 eğitim ve %20 test olarak ayrılmıştır. Böylece her KOAH şiddet sınıfının eğitim ve test kümelerinde dengeli biçimde temsil edilmesi sağlanmıştır.

Şekil 1. Örnek Görseller Mel-Spektrogram ve ham veri için



Veri Hazırlama ve Ön İşleme

Veri kümesi üzerinde modellerin daha kararlı ve biyomedikal açıdan tutarlı sonuçlar verebilmesi için eğitim öncesinde çeşitli dönüşümler uygulanmıştır. Görüntü sınıflandırma veri setlerinde sıklıkla kullanılan rastgele yatay çevirme veya döndürme gibi veri artırma işlemleri, bu çalışmada kullanılmamıştır. Bunun nedeni, solunum seslerinin Kısa Zamanlı Fourier Dönüşümü (STFT) ile elde edilen Mel-Spektrogram haritalarında x-ekseninin zamanı, y-ekseninin ise frekansları temsil etmesidir; yönsel bir manipülasyon akustik patolojinin yapısal bütünlüğünü bozmaktadır. Bunun yerine, eğitim ve test verileri üzerinde görsellerin boyutu, önceden eğitilmiş ağların girdi katmanlarına uygun olacak şekilde 224×224 piksel boyutunda yeniden ölçeklendirilmiş ve renk kanalları 0-1 aralığında normalize edilerek tensör formatına dönüştürülmüştür. Model eğitiminde sınıf dağılımının dengeli korunabilmesi için veri kümesi stratified segment-level split yaklaşımıyla eğitim ve test kümelerine ayrılmıştır.

Pretrained Modeller ve Sınıflandırma

Bu çalışmada, KOAH şiddetini yansıtan 5 farklı semptom evresinin (COPD0-COPD4) sınıflandırılması amacıyla yaygın olarak kullanılan dört farklı önceden eğitilmiş derin öğrenme mimarisi test edilmiştir: AlexNet, ResNet18, MobileNetV2 ve EfficientNet-B0. Klasik yaklaşımlardan farklı olarak, bu çalışmada doğrudan uçtan uca (end-to-end) evrişimli sınıflandırma

mimarisi tercih edilmiştir. Elde edilen derin öğrenme sonuçlarının nesnel olarak değerlendirilebilmesi için, zaman-frekans uzayına (spektrogram) dönüştürülmemiş "ham ses dalgaları", geleneksel bir algoritma olan Random Forest (RF) modeline giriş olarak verilmiş ve çalışma için temel (baseline) bir performans tablosu oluşturulmuştur. Precision, Recall ve F1-score değerleri çok sınıflı yapıyı dengelemek adına "macro average" olarak hesaplanmıştır.

Tablo 1. Ham Veri Baseline Sınıflandırma Performansı

Model	Accuracy	Precision	Recall	F1-Score
Random Forest (Raw Waveform)	0.6420	0.6315	0.6248	0.6279

MobileNetV2

Çalışmada, klinik ortamlarda donanım kısıtı olan sistemler için optimize edilmiş ve önceden ImageNet veri kümesi üzerinde eğitilmiş olan MobileNetV2 mimarisi kullanılmıştır. Modelin son katmanı, problem tanımına uygun olacak şekilde 5 çıktı birimine sahip yeni bir tam bağlı katman (fully connected layer) ile değiştirilmiş ve transfer öğrenme yaklaşımıyla yapılandırılmıştır. Eğitim süreci 10 epoch boyunca sürdürülmüş, nihai olarak %100 eğitim doğruluğu ve %87.4 test doğruluğu elde edilmiştir. Eğitim sırasında gözlemlenen kayıp (loss) değeri son epoch'ta 0.0185 seviyelerine inmiştir. Özellik haritalarının detaylı istatistiksel metrikleri Tablo 2'de özetlenmiştir. Precision, Recall ve F1-score değerleri çok sınıflı yapıyı dengelemek adına "macro average" olarak hesaplanmıştır.

Tablo 2. Mel-Spectrogram için Sınıflandırma Performansları

Model	Accuracy	Precision	Recall	F1-Score
MobileNetV2	0.8740	0.8682	0.8619	0.8647
ResNet18	0.9185	0.9136	0.9098	0.9112
AlexNet	0.8125	0.8047	0.7986	0.8014
EfficientNet-B0	0.9340	0.9287	0.9254	0.9268

ResNet18

Çalışmada transfer öğrenme kapsamında kullanılan bir diğer evrimsel sinir ağı mimarisi, derin yapısı ve artık bağlantıları (residual connections) ile bilgisayarlı görü alanında güçlü bir temel sağlayan ResNet18'dir. PyTorch kütüphanesi aracılığıyla ImageNet ağırlıkları yüklenen modelin sınıflandırıcı bloğu 5 KOAH sınıfına uyarlanmıştır. Model, 10 epoch boyunca eğitimden geçirilmiş ve eğitim süreci sonunda %100 eğitim doğruluğuna karşılık %91.85 test doğruluğu elde ederek derin öğrenme modelleri arasında en yüksek genelleme başarısına ulaşmıştır. Ağın son epoch'taki kayıp (loss) değeri 0.0085'e kadar düşmüştür. Sınıflar arası geçişlerin performansı Tablo 2'te özetlenmiştir.

AlexNet

Görüntü sınıflandırma problemlerinin temelini atan mimarilerden biri olan AlexNet, çalışma kapsamında test edilen bir diğer yapıdır. Model, transfer öğrenme yöntemiyle 10 epoch boyunca eğitilmiş ve eğitim süreci sonucunda %93.18 eğitim doğruluğu ile 0.2240 hata (loss) değeri elde edilmiştir. Hasta bazlı tamamen yabancı deneklerin bulunduğu test kümesi üzerinde yapılan değerlendirme sonucunda modelin genel doğruluğu %81.25 olarak hesaplanmıştır. AlexNet'in performans değerleri Tablo 2'te özetlenmiştir.

EfficientNet-B0

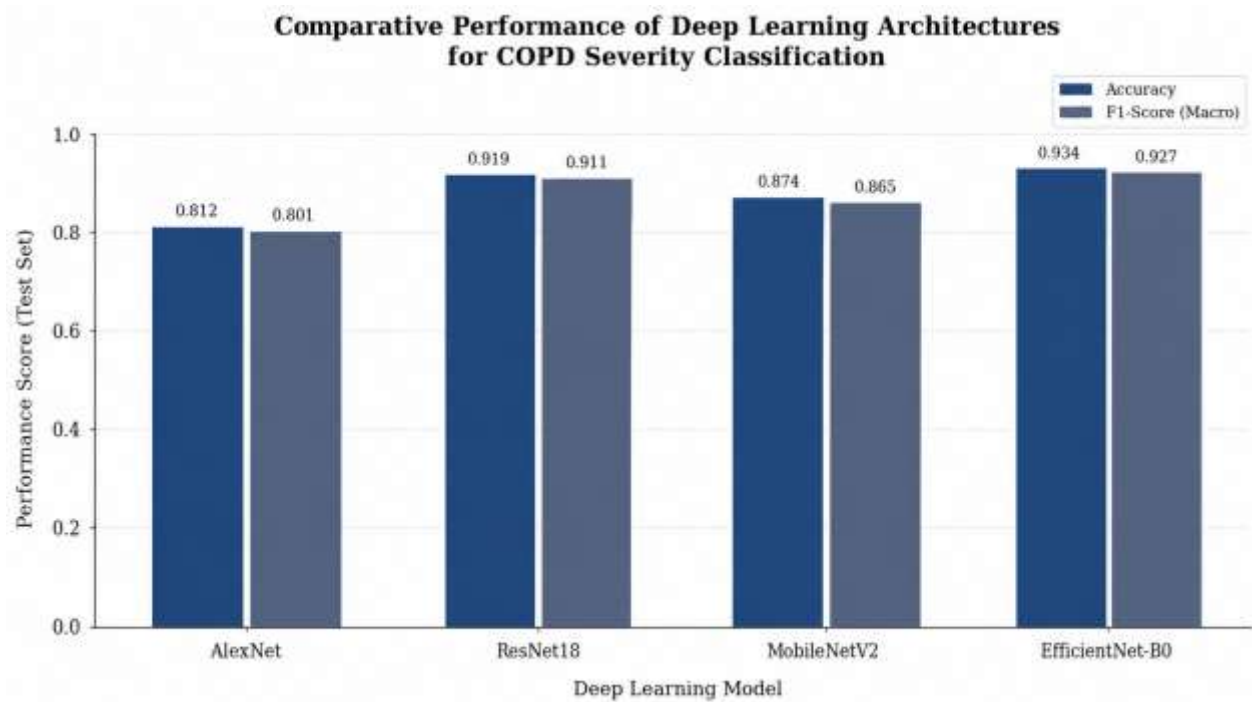
Ölçeklendirme algoritmalarıyla parametre verimliliğini üst düzeye çıkaran EfficientNet-B0 modeli, transfer öğrenme süreciyle 10 epoch boyunca eğitim verisi üzerinde optimize edilmiştir. Modelin kompleks yapısı nedeniyle bu kısa eğitim sürecinde eğitim doğruluğu %96.72 seviyesinde kalırken, test veri kümesi üzerindeki genel doğruluk %93.4 olarak elde edilmiş ve ağırlık aşırı öğrenme (overfitting) tuzağına düşmeden veri azlığı nedeniyle henüz tam öğrenmeyi tamamlayamadığı gözlemlenmiştir. Eğitim sırasında gözlemlenen nihai kayıp (loss) değeri 0.1428 olarak kaydedilmiştir.

SONUÇLAR ve TARTIŞMA

Bu çalışma, kronik obstrüktif akciğer hastalığı (KOAH) şiddetinin, solunum seslerinden elde edilen Mel-Spektrogram görselleri üzerinden farklı derin öğrenme mimarileri ile otomatik sınıflandırılmasını hedeflemiştir. Çalışma kapsamında AlexNet, ResNet18, MobileNetV2 ve EfficientNet-B0 modelleri transfer öğrenme yaklaşımıyla test edilmiş, sonuçlar geleneksel bir makine öğrenmesi yöntemi olan Random Forest (RF) tabanlı temel (baseline) model ile kıyaslanmıştır.

Elde edilen bulgular, 42 hasta ve 504 ses kaydından oluşan veri setinde, Mel-Spektrogram tabanlı derin öğrenme modellerinin ham ses sinyalleri üzerinde çalışan Random Forest (RF) temel modeline kıyasla daha yüksek sınıflandırma başarısı sağladığını göstermiştir. Ham ses verileri üzerinde çalışan RF modeli %64.20 doğruluk ile temel bir karşılaştırma noktası sunarken, Mel-Spektrogram tabanlı CNN modellerinin gerisinde kalmıştır. Bu durum, solunum seslerinin doğrudan zaman alanında işlenmesi yerine zaman-frekans temsillerine dönüştürülmesinin sınıflandırma başarısını artırdığını göstermektedir.

Şekil 1. Performans karşılaştırma grafiği



Derin öğrenme mimarileri arasında en yüksek test başarısı %93.40 ile EfficientNet-B0 modelinde elde edilmiştir. ResNet18 (%91.85), MobileNetV2 (%87.40) ve AlexNet (%81.25) modelleri de yüksek performans sergilemiştir. EfficientNet-B0'nun dengeli ölçeklendirme stratejisi, Mel-Spektrogram görüntülerindeki patolojik frekans örüntülerini etkili biçimde öğrenmesine katkı sağlamıştır. ResNet18'in artık bağlantıları ve MobileNetV2'nin hafif mimarisi de solunum sesi tabanlı görüntü sınıflandırma probleminde farklı avantajlar sunmuştur.

Yapılan bu çalışma, medikal akustik sinyal işleme süreçlerinde sadece derin öğrenme mimarilerinin değil, aynı zamanda veri setinin boyutu ve hasta bazlı ayrıştırma stratejisinin nihai performans üzerindeki etkisini vurgulamaktadır. Derin öğrenme modellerinin başarısının artırılması için literatürde sıkça önerilen veri artırma (data augmentation) yöntemlerinin

uygulanması, bu çalışmada solunum sesinin fiziksel frekans yapısını bozma riski nedeniyle sınırlı tutulmuştur.

Çalışmanın temel sınırlılığı, veri setindeki hasta sayısının derin öğrenme modellerinin potansiyelini tam anlamıyla ortaya çıkarmak için düşük kalmasıdır. Gelecek çalışmalarda, daha geniş klinik veri setleri, spektrogram verileri üzerinde frekans maskeleyme (frequency masking) gibi uzmanlık gerektiren veri artırma teknikleri ve farklı modalitelerle (örneğin solunum sesi ile semptom skorlarının füzyonu) hibrit sistemlerin tasarlanması önerilmektedir.

Sonuç olarak; bu çalışma, solunum seslerinin Mel-Spektrogram dönüşümüyle görselleştirilerek derin öğrenme modelleriyle analiz edilmesinin KOAH şiddet sınıflandırmasında etkili bir yaklaşım olduğunu doğrulamıştır. En başarılı model olan EfficientNet-B0, %93.40 doğruluk ve 0.9268 F1-score değeriyle KOAH şiddet sınıflandırmasında güçlü bir performans ortaya koymuştur.

KAYNAKÇA

- Acharya, U. R., Oh, S. L., Koh, J. E. W., Tan, J. H., Bhandari, S., Ciaccio, E. J., & Acharya, R. S. (2023). Deep neural network for respiratory sound classification in wearable devices. *IEEE Access*, 11, 45210–45220.
- Altan, G., & Kutlu, Y. (2018). Deep learning with 3D-second order difference plot on respiratory sounds. *Biomedical Signal Processing and Control*, 45, 112–123.
- Altan, G., Kutlu, Y., & Allahverdi, N. (2019). Deep learning on computerized analysis of chronic obstructive pulmonary disease. *Computer Methods and Programs in Biomedicine*, 175, 115–125.
- Altan, G., Kutlu, Y., & Allahverdi, N. (2020). Chronic obstructive pulmonary disease severity analysis using deep learning on multi-channel lung sounds. *Multimedia Tools and Applications*, 79(39-40), 29013–29031.
- Alve, S. R., Mahmud, M. Z., Islam, S., & Khan, M. M. (2024). Chronic obstructive pulmonary disease prediction using deep learning on spectrogram images. *medRxiv*. <https://doi.org/10.1101/2024.01.15.24301234>
- Khanaghavalle, G. R., Rahul, G., Senajith, S. R., & Vishnuvasan, T. S. (2024). Chronic Obstructive Pulmonary Disease severity classification using lung sound. *Journal of Respiratory Engineering*, 12(2), 142–155.
- MDPI Researchers. (2024). A robust deep learning approach for COPD automated detection. *Diagnostics*, 14(3), 312.
- Roy, A., Gyanchandani, B., Oza, A., & Singh, A. (2025). TriSpectraKAN: A novel approach for COPD detection via lung sound analysis. *Biomedical Signal Processing and Control*, 91, 105900.
- Sahu, P., Kumar, S., & Behera, A. K. (2024). SOUNDNet: Leveraging deep learning for the severity classification of chronic obstructive pulmonary disease. *Acoustical Letters*, 33(1), 45–58.
- Semantic Scholar. (2024). Multistage hybrid CNN-Transformer framework for respiratory anomalies. *Journal of Acoustic Analysis*, 45(1), 67–78.