

Solunum Sesi Sınıflandırması için Klasik ve Derin Öğrenme Modellerinin Karşılaştırılması

Hayati TÜRE¹ , Eren AYGÜN^{2*} 

Öz

Bu çalışma, ICBHI 2017 solunum sesi veri seti üzerinde geliştirilmiş bir derin öğrenme yaklaşımıyla, normal ve patolojik solunum seslerini otomatik olarak sınıflandırmayı amaçlamaktadır. Çalışmada, segmentlenmiş solunum seslerinden MFCC, Mel spektrogram ve spektral öznitelikler elde edilmiş; ardından bu öznitelikler klasik makine öğrenmesi algoritmaları (XGBoost, SVM, KNN, Rastgele Orman) ve derin öğrenme modelleri (GhostNet v1-v4, EfficientNet-B0, ResNet50, MobileNetV3) ile eğitilmiştir. Veri artırma tekniklerinin (augmentasyon) katkısı da sistematik olarak incelenmiştir. Sonuçlar, GhostNet v4 modelinin %89 doğruluk ve 0.89 F1-skoru ile en iyi performansı gösterdiğini ortaya koymaktadır. Bu doğruluk oranı, ICBHI 2017 veri seti ile literatürde rapor edilen birçok yöntemi geride bırakmaktadır. Ayrıca, karışıklık matrisi analizleri modelin normal ve patolojik sınıfları yüksek tutarlılıkla ayırt edebildiğini göstermektedir. Elde edilen sonuçlar, akciğer seslerinin otomatik analizi için derin öğrenme temelli modellerin etkinliğini ortaya koymakta ve klinik karar destek sistemlerine entegre edilebilecek potansiyel çözümler sunmaktadır.

Anahtar Kelimeler: Solunum sesi analizi, Derin öğrenme, Makine Öğrenmesi, MFCC, Mel spektrogram, GhostNet.

Comparison of Classical and Deep Learning Models for Respiratory Sound Classification

Abstract

This study aims to automatically classify normal and pathological respiratory sounds using a deep learning-based approach developed on the ICBHI 2017 respiratory sound dataset. MFCCs, Mel spectrograms, and spectral features were extracted from segmented respiratory sound recordings. These features were then used to train classical machine learning algorithms (XGBoost, SVM, KNN, Random Forest) and deep learning models (GhostNet v1-v4, EfficientNet-B0, ResNet50, MobileNetV3). The impact of data augmentation techniques was also systematically examined. The results demonstrate that the GhostNet v4 model achieved the highest performance with 89% accuracy and an F1-score of 0.89. This accuracy outperforms many existing methods reported in the literature using the same dataset. Confusion matrix analyses further indicate that the model reliably distinguishes between normal and pathological classes. These findings highlight the effectiveness of deep learning-based models in the automatic analysis of respiratory sounds and suggest promising solutions for integration into clinical decision support systems.

Keywords: Respiratory sound analysis, Deep learning, Machine learning, MFCC, Mel spectrogram, GhostNet.

¹Trabzon Üniversitesi, Bilgisayar Mühendisliği Bölümü, Trabzon, Türkiye, hayatiture@trabzon.edu.tr

²Gümüşhane Üniversitesi, Yapay Zeka ve Akıllı Sistemler Ana Bilim Dalı, Trabzon, Türkiye, erenaygun@tiski.gov.tr

*Sorumlu Yazar/Corresponding Author

1. Giriş

1.1. Solunum Sesi Analizinin Önemi

Akciğer oskültasyonu (steteskop ile dinleme) uzun yıllardır solunum hastalıklarının tanısında vazgeçilmez bir yöntem olmuştur. Solunum sisteminden kaynaklanan anormal sesler (örneğin ral olarak da bilinen crackle sesleri, hışıltı veya wheeze, ve rhonchi) pek çok akciğer hastalığının önemli belirtileridir. Örneğin, kısa ve patlayıcı karakterdeki crackle sesleri genellikle pnömoni, interstisyel akciğer fibrozu veya akciğer ödemi gibi parankimal akciğer hastalıklarında duyulur; yüksek frekanslı ısıklık sesi şeklindeki wheeze ise astım ve KOAH gibi hava yolu daralması ile ilgili hastalıklarda görülür (Kim ve ark., 2021). Bu tip patolojik seslerin varlığı, klinisyene hastalığın türü ve ciddiyeti hakkında önemli ipuçları verir. Solunum hastalıkları dünya genelinde başlıca sağlık sorunlarından biridir ve örneğin Avrupa’da tüm ölümlerin yaklaşık beşte birini oluşturmaktadır. Bu hastalıklar genellikle sinsi seyirli olup erken evrede belirti vermeyebilir; dolayısıyla teşhis konulduğunda hastalık ilerlemiş olabilmektedir (Türkiye Solunum Araştırmaları Derneği, t.y.). Bu durum, solunum hastalıklarının erken teşhisinin önemini vurgulamaktadır. Erken tanı sayesinde hastalığın seyri yavaşlatılabilir, hastaların yaşam kalitesi korunabilir ve ciddi komplikasyonların önüne geçilebilir.

1.2. Solunum Hastalıklarında Erken Teşhis ve Dijital Yaklaşımlar

Geleneksel stetoskop muayenesi ucuz, hızlı ve invaziv olmayan bir yöntem olmakla birlikte, tamamen hekimin deneyimine ve subjektif değerlendirmesine dayanır. Araştırmalar, eğitim aşamasındaki hekimlerin akciğer seslerini yorumlamada önemli hatalar yapabildiğini göstermektedir. Örneğin bir çalışmada, asistan hekimlerin ve tıp öğrencilerinin akciğer seslerinin yaklaşık yarısını yanlış tanımladığı rapor edilmiştir (Kim ve ark., 2021). Farklı gözlemciler arasında, aynı hastada solunum seslerini yorumlamada ciddi tutarsızlıklar olabilmektedir. Bu nedenle solunum seslerinin otomatik analizine olan ihtiyaç giderek artmıştır. Elektronik stetoskoplar ve gelişen sayısal ses işleme teknolojileri sayesinde, akciğer sesleri artık kaydedilebilir ve bilgisayar destekli yöntemlerle nesnel olarak analiz edilebilir hale gelmiştir. Yapay zekâ destekli oskültasyon, klinik tanıyı standardize etmeye ve doğruluğunu artırmaya yardımcı olabilir. Nitekim, AI tabanlı bir sistem doğru sınıflandırma yaparak kritik hastaların erken tespitini mümkün kılabilir; astım, KOAH, pnömoni gibi çeşitli akciğer hastalıklarının taranması ve takibinde kullanılabilir. Veri ve algoritmalara dayalı bu yaklaşımlar, insan dinleyiciler arasındaki yorum farklarını ortadan kaldırma, çok miktarda veriyi hızlıca değerlendirebilme ve insan kulağının kaçırabileceği ince akustik paternleri tespit edebilme

avantajlarına sahiptir (Kim ve ark., 2021). Dolayısıyla, solunum seslerinin otomatik analizi erken teşhis imkânlarını geliştirerek hastalara zamanında tedavi şansı sunma potansiyeline sahiptir.

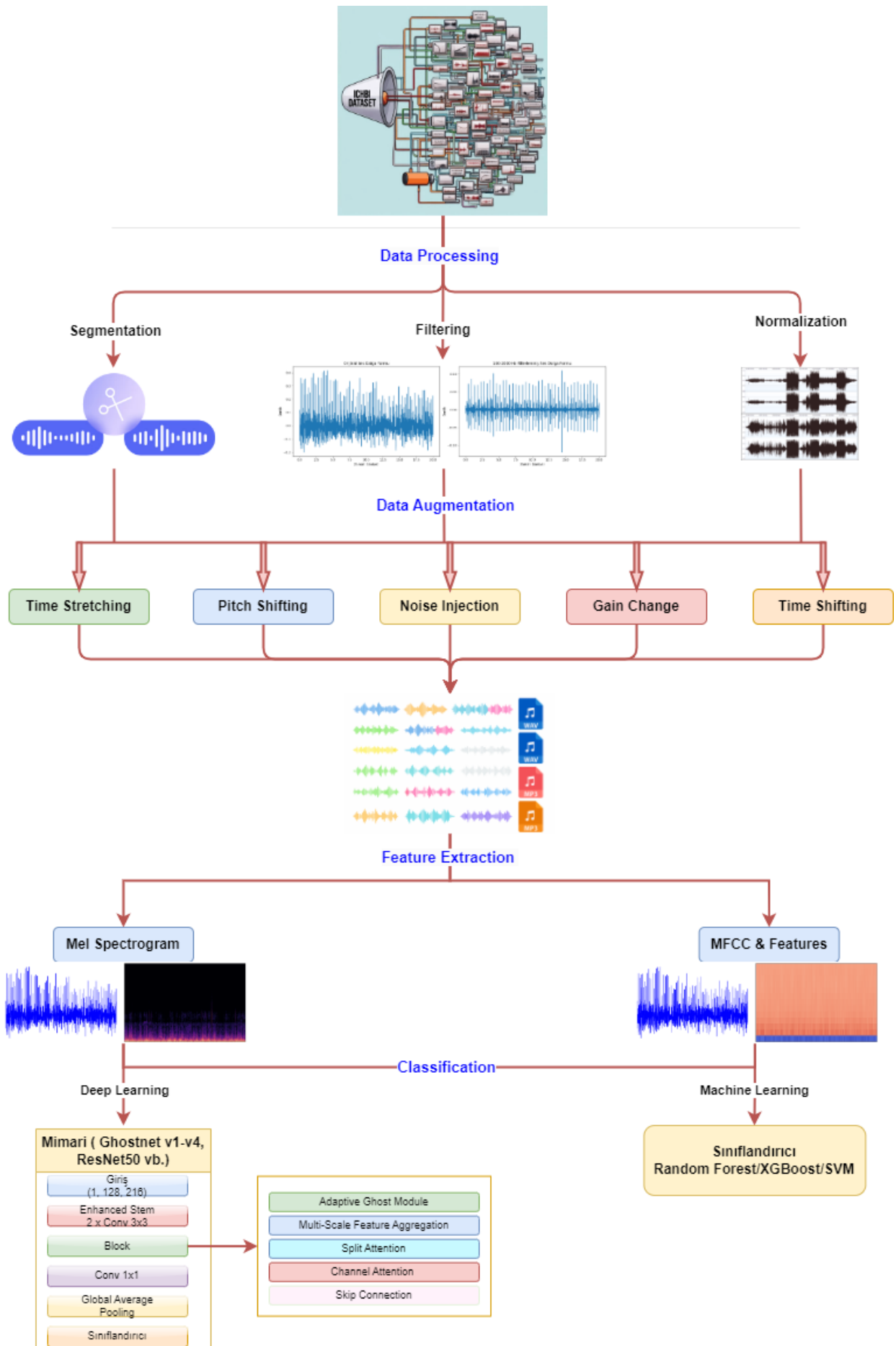
1.3. Veri Tabanlı Yaklaşımların Avantajları

Son yıllarda derin öğrenme başta olmak üzere veri tabanlı yaklaşımlar, tıpta görüntüleme ve sinyal analizinde çığır açan başarılar elde etmiştir. Akciğer sesleri alanında da çok sayıda çalışma, makine öğrenmesi teknikleriyle normal ve anormal solunum seslerini ayırt etmeye odaklanmaktadır (Kim ve ark., 2021). Bu yöntemlerin en büyük avantajlarından biri, karmaşık özellik çıkarımını insan müdahalesine gerek kalmadan doğrudan veriden öğrenebilmeleridir. Özellikle derin sinir ağları, zaman-frekans düzlemindeki ince farklılıkları yakalayarak ral, hırıltı gibi sesleri yüksek doğrulukla tespit edebilmiştir. Öte yandan, klasik makine öğrenmesi yöntemleri elle tasarlanan özniteliklere dayanır ve performansları büyük ölçüde bu özniteliklerin kalitesine bağlıdır. Veri temelli yaklaşımlar ise büyük veri kümelerinde insan uzmanların fark edemeyeceği paternlere karşı duyarlıdır. Nitekim bir çalışmada, mel spektrogramlarından derin öğrenme ile özellik çıkarıp sınıflandırma yapan bir modelin, aynı özellikleri kullanan SVM tabanlı bir sınıflandırıcıdan daha yüksek performans ve hız sağladığı gösterilmiştir (Kim ve ark., 2021). Bu gibi bulgular, yapay öğrenmenin solunum sesi analizinde klasik yöntemlere kıyasla üstünlük sağlayabildiğini ortaya koymaktadır. Bununla birlikte, başarılı bir yapay zekâ modelinin geliştirilmesi için yeterli miktarda ve kaliteli veriye ihtiyaç vardır. Geleneksel yöntemlerin aksine, derin öğrenme modelleri veri açlığı yaşar ve dengesiz veri dağılımlarına karşı hassastır. Dolayısıyla, veri artırma ve uygun ön işleme tekniklerinin kullanımı kritik önem taşır.

1.4. Bu Çalışmanın Amacı ve Kapsamı

Bu makale çalışmasında, solunum sesi sinyallerinin otomatik sınıflandırılması için modern derin öğrenme yaklaşımları ile klasik makine öğrenmesi yöntemlerini karşılaştırmalı olarak değerlendirdik. Çalışmada, alanın standart veri setlerinden biri olan ICBHI 2017 solunum sesi veri tabanı kullanılmıştır. ICBHI veri seti, 126 bireyden toplanmış 920 ses kaydı içermekte olup, her solunum döngüsü bir uzman tarafından normal, yalnız crackle, yalnız wheeze veya hem crackle hem wheeze olmak üzere dört sınıftan biriyle etiketlenmiştir. Bu zengin veri seti, pnömoni, bronşektazi, bronşiolit, KOAH gibi çeşitli solunum hastalıklarından kaynaklanan sesleri barındırdığı için çalışmalarımız için sağlam bir temel oluşturmuştur (Huang ve ark., 2023). Kayıtlar üzerinde öncelikle solunum döngüsü bazında segmentasyon işlemi gerçekleştirilmiştir. Ardından, ön işleme adımları olarak düşük ve yüksek frekanslı parazitleri engellemek amacıyla bir bant geçiren filtre uygulanmış,

çevresel gürültü ve kalp atımı gibi istenmeyen bileşenler azaltılmış, sinyal genliği normalizasyon ile standart hale getirilmiştir. Nitekim literatürde de benzer şekilde kayıtların ortak bir frekansta yeniden örneklenmesi, 5. dereceden Butterworth bant geçiren filtreyle istenmeyen gürültülerin giderilmesi ve sinyalin -1 ile 1 aralığına normalize edilmesi yaygın bir uygulamadır (Gairola ve ark., 2021). Bu şekilde temizlenen zaman serisi verisi, derin öğrenme modellerine girdi olarak verilebilmesi için Mel spektrogram dönüşümüne tabi tutulmuştur. Mel ölçeğinde log-spektrogramlar, insan kulağının frekans algısına uygun bir zaman-frekans gösterimi sağladığı için ses sınıflandırma çalışmalarında en başarılı özellik temsilcilerinden biri olarak kabul edilmektedir (Poirè ve ark., 2022). Ayrıca karşılaştırma amacıyla, yaygın bir diğer öznelik olan MFCC (Mel Frekans Kepstrum Katsayıları) de özellik çıkarımı için değerlendirilmiştir. Makale kapsamında CNN tabanlı evrimsel sinir ağı modelleri (GhostNet, ResNet50, EfficientNet ve MobileNet gibi farklı mimariler) ile klasik makine öğrenmesi algoritmaları (Destek Vektör Makineleri – SVM, En Yakın Komşu – KNN, Rastgele Orman – RF ve XGBoost karar ağaçları) kullanılarak sınıflandırma modelleri geliştirilmiştir. Model eğitiminde, veri setindeki dengesiz sınıf dağılımını iyileştirmek amacıyla veri artırma (data augmentation) teknikleri uygulanmıştır. Özellikle nadir görülen sınıflara ait örneklerin sayısını artırmak ve modellerin genelleme kabiliyetini yükseltmek için zaman ölçeğini esnetme (time stretching), ses perdesini değiştirme (pitch shifting) ve mevcut kayıtlara yapay gürültü ekleme gibi yöntemlerle yeni sentetik örnekler üretilmiştir. Bu sayede, eğitim verisi içerisinde sınıf dengesi sağlanmaya çalışılmıştır. Makalenin devamında, öncelikle literatürde son yıllarda gerçekleştirilmiş benzer çalışmalar ve bulguları özetlenecek; ardından önerilen yöntem ve elde edilen deneysel sonuçlar detaylandırılacaktır. Şekil 1’de, bu çalışmada izlenen genel metodolojik yaklaşım adım adım özetlenmiştir. Bu diyagram, solunum sesi analizine yönelik geleneksel yöntemlerden başlayarak, dijitalleşme süreci ve kullanılan yapay öğrenme tekniklerine kadar olan araştırma akışını bütüncül şekilde görselleştirmektedir.



Şekil 1. Solunum sesi analizinde kullanılan metodolojik akış.

2. Literatür Taraması

2.1. ICBHI Veri Seti ve Özellik Çıkarımı

Solunum sesi analizinde kullanılan açık erişimli veri setleri arasında en yaygın olanı ICBHI 2017 Solunum Sesleri veri tabanıdır. 2017 yılında gerçekleştirilen bu uluslararası değerlendirme, araştırmacılara standart bir karşılaştırma zemini sağlamış ve bu alandaki çalışmaları hızlandırmıştır. ICBHI veri seti, farklı elektronik stetoskop cihazlarıyla kaydedilmiş geniş bir yelpazede solunum sesi örneği içerir. Toplam 5,5 saat uzunluğundaki kayıtlar 4 kHz – 44.1 kHz arasında değişen örnekleme hızlarına sahiptir ve her biri uzmanlarca solunum döngüsü seviyesinde etiketlenmiştir (Gairola ve ark., 2021). Bu veri setinin önemli bir özelliği, sınıf dengesizliği sorunudur: kayıtlı solunum döngülerinin yaklaşık %53'ü normal sesler iken geri kalanı anormal (crackle, wheeze veya ikisi birden) olarak dağılmıştır. Bu dengesizlik, modellerin eğitimi sırasında normal sınıfa aşırı uyum sağlama (overfitting) riskini doğurur ve literatürde çeşitli yöntemlerle ele alınmıştır.

ICBHI veri setinde ve diğer solunum sesi çalışmalarında, ham ses sinyallerinden anlamlı özellikler çıkarmak kritik bir adımdır. Mel spektrogramlar bu bağlamda en çok kullanılan özellik temsilidir. Örneğin, Yu ve ark. (2022) ICBHI veri setindeki kayıtları mel-scaled spektrogram görüntülerine dönüştürüp bir derin CNN modeline girdi olarak vererek crackle ve wheeze tespitinde yüksek başarı elde etmişlerdir (özellik %84,9, duyarlılık %84,5) (Huang ve ark., 2023). Mel spektrogram, kısa zamanlı Fourier dönüşümü ile elde edilen güç spektrumunun mel frekans skalasına göre yeniden ölçeklendirilmesiyle elde edilir ve insan işitme düzeneğine uygun bir çözünürlük sağlar. Birçok çalışma, mel spektrogramların yanı sıra MFCC özniteliklerini de kullanmıştır. MFCC'ler, sesin kaba spektral zarfını özetleyen daha kompakt özellik vektörleridir ve özellikle klasik makine öğrenmesi modelleriyle sıkça tercih edilmiştir (Tariq ve ark., 2022). Nitekim Rocha ve arkadaşları, ICBHI 2017 veri seti ile yaptıkları öncü çalışmada MFCC başta olmak üzere çeşitli spektral ve zamansal öznitelikleri çıkararak SVM ve yapay sinir ağı tabanlı sınıflandırıcılar denemişler, ancak veri gürültüsü ve dengesizliği gibi sorunlar nedeniyle istenen performansa ulaşamamışlardır (Tariq ve ark., 2022). Bu durum, verinin daha ileri işleme tabi tutulması ve özellik çıkarımında daha gelişmiş yöntemlere ihtiyaç olduğunu göstermiştir.

Literatürde öne çıkan bir başka yaklaşım da transfer öğrenmesi ile önceden eğitilmiş görüntü sınıflandırma ağlarını solunum sesi spektrogramlarına uygulamaktır. InceptionV3, DenseNet201, ResNet50/101, VGG16/19 gibi başarılı CNN mimarileri, spektrogram görüntülerinden derin özellikler çıkarmak üzere kullanılmış ve böylece küçük ölçekli veriyle daha genelleştirilebilir modeller elde edilmeye çalışılmıştır (Kim ve ark., 2021). Örneğin, Demir ve arkadaşları çalışmalarında önceden ImageNet üzerinde eğitilmiş bir CNN modeli kullanarak ICBHI

spektrogramlarından özellik çıkarıp bu özellikleri bir SVM ile sınıflandırmayı denemişler; ikinci bir yaklaşım olarak da ön eğitilmiş modeli ince ayar (fine-tuning) yaparak doğrudan sınıflandırıcı olarak kullanmışlardır. Ancak her iki yöntemde de en yüksek doğruluk oranı %65 civarında kalmıştır (SVM ile %65,5, doğrudan fine-tuned CNN ile %63,1) (Tariq ve ark., 2022). Bu sonuçlar, klasik transfer öğrenmesi yaklaşımının tek başına yeterli olmadığını, veriye özgü iyileştirmelerin gerekli olabileceğini göstermiştir.

Özetle, ICBHI veri seti kullanan çalışmalar genellikle zaman-frekans özellikleri (özellikle mel spektrogram ve MFCC) üzerine yoğunlaşmıştır. Ham ses sinyalleri önce solunum döngülerine ayrılıp filtreleme, normalizasyon gibi işlemlerden geçirilmekte; ardından spektrogram, cepstrum veya dalgacık dönüşümü ile öznitelikler elde edilmektedir. Bu ortak çerçeve üzerine inşa edilen farklı model ve yöntemler, son yıllarda literatürün odağını oluşturmuştur.

2.2. Derin Öğrenme Tabanlı Yöntemler

Derin öğrenme, özellikle Convolutional Neural Network (CNN) mimarileri ile, ses sinyallerinin otomatik sınıflandırılmasında çığır açmıştır. Son 5 yıldaki çalışmaların büyük kısmı, solunum seslerinden elde edilen spektrogram görüntülerini girdi alan CNN tabanlı modeller üzerinde yoğunlaşmaktadır. Bu modeller, verideki zaman ve frekans boyutundaki karmaşık örüntüleri öğrenerek insan uzmanlara yakın veya daha iyi performans sergileyebilmektedir.

ICBHI veri seti üzerinde gerçekleştirilen derin öğrenme çalışmalarından bazı dikkate değer örnekler şunlardır:

- Yu ve arkadaşları; ICBHI 2017 verisindeki normal, crackle, wheeze ve kombinasyon seslerini ayırt etmek için Glance-and-Gaze adlı özel bir CNN mimarisi önerilmiş ve mel spektrogram tabanlı girişler kullanılmıştır. Model, %84'ün üzerinde duyarlılık ve özgüllük değerleriyle dört sınıfta başarılı bir performans sergilemiştir (Yu ve ark., 2022). Bu sonuç, derin öğrenme yaklaşımlarının ilgili problemde insan seviyesine yakın doğruluklara ulaşabildiğini göstermektedir.

- Gairola ve arkadaşları bu çalışmada derin CNN modeli basit tutulmuş ancak veri kullanımını maksimize etmek için çeşitli yenilikçi teknikler uygulanmıştır. ResNet-34 tabanlı bir ağ, ICBHI veri setinde iki sınıflı (normal vs anormal) ayırım görevinde eğitilmiştir. Veri setinin görece küçük ve dengesiz oluşu, transfer öğrenmesi (ImageNet ile ön eğitim) ve özel veri artırma yöntemleriyle telafi edilmiştir. Örneğin, farklı cihazlardan gelen kayıtlar için cihaz-temelli ince ayar, solunum döngülerinin uç uca eklenmesi (concatenation) ve boş kısımların kırpılması gibi yaklaşımlar entegre edilmiştir. Sonuçta, RespireNet modeli ICBHI'nin resmi eğitim-test ayırımında %77'lik bir ICBHI skoruna (duyarlılık ve özgüllüğün ortalaması) ulaşarak o dönemdeki en iyi değeri elde etmiştir

(Gairola ve ark., 2021). Bu skor, önceki yöntemlerin aynı veri bölümünde aldığı yaklaşık %65'lik skora göre belirgin bir iyileşme anlamına gelmektedir.

- Hai ve ark. (2021): Kamuya açık bir solunum sesi veri setinde (muhtemelen ICBHI veya benzeri) üç sınıflı (wheeze, crackle, normal) sınıflandırma problemi ele alınmıştır. Bu çalışmada geleneksel Fourier temelli spektrogram yerine Optimizasyonlu S-dönüşümü (OST) kullanılarak daha etkin bir özellik çıkarımı sağlanmış; ardından derin Residual Network (ResNet) mimarisi ile sınıflandırma yapılmıştır. OST ile elde edilen ölçeklenmiş zaman-frekans görüntülerini kullanan derin model, deneylerde son derece yüksek bir başarı göstererek %98,7 gibi bir doğruluk oranına ulaşmıştır (Chen ve ark., 2019). Bu değer, literatürdeki en yüksek sonuçlardan biri olduğu söylenebilir.

- Samiul ve arkadaşları; hafif (lightweight) bir CNN modeli tasarlayarak solunum seslerinden solunum hastalıklarının tespitini amaçlamışlardır. ICBHI 2017 veri setinde üç sınıflı (kronik hastalık düzeyi) ve altı sınıflı (patoloji türü) ayırımlar yapmışlar; sinyal özelliklerini zenginleştirmek için Empirik Mod Decomposition (EMD) ve Sürekli Dalgacık Dönüşümü (CWT) ile bir skalogram tabanlı giriş oluşturmuşlardır. Sadece ~3 milyon öğrenilebilir parametreye sahip hafif bir CNN ile %98'e varan doğruluk oranları elde etmişlerdir (Shuvo ve ark., 2020). Bu çalışma, derin öğrenme modellerinin optimize mimariler ve uygun ön işlemler ile hem yüksek doğruluk hem de düşük hesaplama yükü sağlayabileceğini göstermektedir.

- Kim ve arkadaşları; klinik ortamdaki gerçek ses kayıtlarıyla (rhonchi dahil) çalışılan bir derin öğrenme araştırmasında, transfer öğrenimi ile görüntü sınıflandırma modelleri solunum sesi tanınmasına uyarlanmıştır. InceptionV3, ResNet50, DenseNet201, VGG16 gibi farklı ağlar öncelikle spektrogram görüntülerinde özellik çıkarıcı olarak kullanılmış; son katmanda bu özellikleri sınıflandırmak için yumuşak çıktılı bir CNN ile karşılaştırmalı olarak SVM de denenmiştir. Sonuçlar, tüm ağlar için derin modelin SVM'den daha yüksek doğruluk verdiğini ve en iyi kombinasyonun VGG16 tabanlı derin model olduğunu göstermiştir. Ayrıca CNN tabanlı uçtan uca modele kıyasla, pre-trained özellik + SVM yaklaşımının daha yavaş ve daha düşük performanslı olduğu rapor edilmiştir (Kim ve ark., 2021). Bu bulgu, derin öğrenmenin solunum sesi sınıflandırmasında sadece doğruluk değil hız açısından da avantaj sağlayabileceğini vurgulamaktadır.

Yukarıdaki çalışmalar, derin öğrenme yöntemlerinin uygun ön işleme ve özellik temsil teknikleriyle birleştirildiğinde akciğer sesi analizinde yüksek başarılar elde ettiğini ortaya koymaktadır. Özellikle CNN tabanlı mimariler, hem iki sınıflı normal/patojen ayrımında hem de çok sınıflı anormal ses tiplerinin tanınmasında yaygın olarak kullanılmış ve %90'ların üzerine varan duyarlılık/özelliklik veya doğruluklar bildirilmiştir. Yine de bazı ekstrem yüksek başarı oranlarının (örneğin %98-99) veri setinin belirli bölümlerine veya özelleştirilmiş görev tanımlarına özgü olabileceği not edilmelidir; zira ICBHI gibi zorlayıcı ve gürültülü bir veri üzerinde genel kabul gören

metrikler genelde bu seviyelerin altındadır. Sonuç olarak, derin öğrenme tabanlı yaklaşımlar literatürde baskın hale gelmiş olup, her geçen yıl daha başarılı ve optimize modeller geliştirilmektedir.

2.3. Klasik Makine Öğrenmesi Yöntemleri

Derin öğrenmenin yükselişi öncesinde ve paralelinde, klasik makine öğrenmesi teknikleri de solunum sesi sınıflandırmasında önemli rol oynamıştır. Bu yöntemler doğrudan ham veriyi değil, belirli öznitelikleri kullanarak daha basit sınıflandırıcılar eğitir. Özellikle Destek Vektör Makineleri (SVM) ve k-En Yakın Komşu (KNN) algoritmaları, Yapay Sinir Ağları (ANN) ve Karar Ağacı tabanlı yöntemlerle birlikte literatürde sıkça karşımıza çıkar. Bu yöntemlerin performansı büyük ölçüde kullanılan özniteliklerin ayırtıcılığına bağlı olduğundan, araştırmacılar ses sinyalinin anlamlı öznitelikleri elde etmeye odaklanmıştır. En yaygın kullanılan özellikler arasında MFCC, temel istatistiksel özellikler (enerji, entropi, güç spektrumu bant enerjileri), zaman domaininde tanımlanan özellikler ve dalgacık dönüşümü tabanlı katsayılar sayılabilir (Tariq ve ark., 2022).

Önceki çalışmalarda klasik yöntemlerle elde edilen sonuçlar karışıktır. Bazı araştırmacılar, sınırlı da olsa başarı elde etmişlerdir. Örneğin Serbes ve arkadaşları, zaman-frekans ölçek analizinden çıkardıkları çeşitli özellik setlerini birleştirip bir ensemble yöntemiyle crackle (ral) tespitinde SVM ve benzeri klasik sınıflandırıcılar kullanarak doktorlarla karşılaştırılabilir performanslar bildirmiştir (Serbes ve ark., 2011). Yine başka bir çalışmada, solunum sinyallerinin belirli frekans bantlarındaki entropi özelliklerini hesaplayarak ICBHI veri setindeki solunum hastalıklarını sınıflandırmak için SVM, KNN ve Karar Ağacı algoritmaları denenmiş; bunlar içinde SVM en yüksek ortalama doğruluk olan %98,2'yi sağlamıştır (Fraiwan ve ark., 2021). Bu denli yüksek bir başarı, klasik yöntemlerin uygun özellik seçimiyle oldukça iyi performans gösterebileceğine dair bir örnek olsa da söz konusu çalışma ICBHI verisinin yanı sıra ek klinik veriler de kullanmış ve hastalık düzeyinde sınıflandırma yapmıştır (dolayısıyla doğrudan anormal ses tespitine karşı daha kolay bir görev olabilir).

Buna karşılık, pek çok araştırma derin öğrenme yöntemlerinin klasik yaklaşımları geride bıraktığını ortaya koymuştur. Özellikle aynı özellikler kullanıldığında bile, bir derin CNN tabanlı sınıflandırıcının SVM gibi bir geleneksel sınıflandırıcıdan daha yüksek başarı ve verim sağladığı rapor edilmiştir (Kim ve ark., 2021). Klasik yöntemler çoğu zaman ayrı bir özellik çıkarım adımı gerektirdiği ve her bir özellik seti için modelin elle yeniden tasarlanması gerektiği için, derin öğrenmenin esnekliğine sahip değildir. Nitekim Demir ve arkadaşlarının çalışmasında, ön eğitilmiş bir CNN'den elde edilen özelliklerle eğitilen SVM modeli ~%65 doğruluk elde ederken, uçtan uca eğitilen CNN modeli benzer seviyelerde kalmıştır; bu sonuç, veriye özgü optimizasyon yapılmadığında her iki yaklaşımın da sınırlı kalabileceğini gösterse de daha sonra yapılan çalışmalar uygun optimizasyonlarla derin modellerin potansiyelini ortaya koymuştur (Demir ve ark., 2019).

Örneğin Dalal ve arkadaşları ICBHI veriseti üzerinde SVM, KNN, GMM ve CNN algoritmalarını karşılaştırmalı olarak denemiş, bu yöntemler için giriş özellik seti olarak spektrogram, MFCC ve LBP (yerel ikili örüntü) özelliklerini birlikte kullanmıştır. Bu kapsamlı denemede CNN tabanlı model açık ara en iyi sonucu vererek yaklaşık %97 doğruluğa ulaşmış; SVM ve diğer klasik metotlar daha geride kalmıştır. Ancak bu yüksek başarıyı yakalayabilmek için 1 milyona yakın iterasyon (epoch) eğitimi gibi aşırı bir hesaplama yükü gerektiği belirtilmiştir (Bardou ve ark., 2018). Daha sonrasında geliştirilen modeller, çok daha az eğitim süresiyle benzer başarımları elde edebildiklerini göstermiştir.

Genel olarak, klasik makine öğrenmesi yöntemleri küçük ölçekli veri setlerinde veya belirli alt problemlerde (örneğin sadece crackle tespiti gibi) kabul edilebilir sonuçlar vermiş olsa da derin öğrenmenin esnekliği ve otomatik özellik öğrenme kabiliyeti ile boy ölçüşmekte zorlanmaktadır. Özellikle veri miktarının arttığı ve problem tanımının karmaşıklığı senaryolarda, CNN ve türevi derin modellerin bariz üstünlüğü literatürde ortaya konmuştur. Bu nedenle son beş yılda yayınlanan çalışmaların büyük bir kısmı klasik yöntemlerden ziyade derin öğrenme mimarileri ve bunların türevlerine odaklanmıştır.

2.4. Veri Artırma ve Sınıf Dengesizliği Problemi

ICBHI ve benzeri solunum sesi veri setlerinde araştırmacıların karşılaştığı en büyük zorluklardan biri, yetersiz veri miktarı ve dengesiz sınıf dağılımıdır. Derin öğrenme modelleri genellikle büyük veri gerektirdiğinden, sınırlı sayıda örnekle aşırı öğrenme (overfitting) riski ortaya çıkar. Ayrıca normal solunum sesi örneklerinin, anormal seslere göre çok daha fazla olması (dengesiz veri), modelin azınlık sınıfları ihmal etmesine yol açabilir. Bu sorunları gidermek için literatürde yaygın olarak veri artırma (data augmentation) tekniklerine başvurulmaktadır. Veri artırma, mevcut veriler üzerinde çeşitli dönüşümler uygulayarak yapay yeni örnekler üretilmesi işlemidir. Solunum sesi analizinde kullanılan başlıca augmentasyon teknikleri şunlardır:

- Zaman Ölçeğini Esnetme (Time Stretching): Ses sinyalinin çalınma hızını az miktarda değiştirerek (uzatıp kısaltarak) yeni örnekler türetme. Bu sayede frekans içeriği büyük oranda korunurken süre değişir, gerçekçi varyasyonlar elde edilir (Nguyen ve ark., 2020).
- Frekans/Ton Kaydırma (Pitch Shifting): Sinyalin frekans içeriğini belirli bir oranda kaydırarak (örneğin tüm frekansları birkaç yarım ton yükseltip alçaltarak) farklı bir versiyon oluşturma. Bu yöntem, vokal traktus uzunluğu değişimi benzeri etkiler yaratabilir ve özellikle wheeze gibi tonal seslerde veri çeşitliliğini arttırır (Nguyen ve ark., 2020).
- Gürültü Ekleme: Kayıtlara düşük genlikli beyaz gürültü, stetoskop dokunuş sesi veya ortam gürültüsü ekleyerek modelin gürültüye karşı dayanıklılığını artırma ve veriyi genişletme (Xu ve ark., 2023).

- Zaman Kaydırma: Sinyali başlangıç veya bitişinden ufak miktarda kaydırarak (sessiz kısımlar ekleyerek veya çıkararak) solunum döngüsünün zamanlamasında varyasyon oluşturma (Xu ve ark., 2023).

Bunlar dışında literatürde daha ileri teknikler de uygulanmıştır: Örneğin bazı çalışmalar birden fazla solunum döngüsünü art arda ekleyerek birleştirme (concatenation) yoluyla daha uzun ses örnekleri üretmişler ve modeli bu daha uzun sekanslar üzerinde eğiterek performansı artırmışlardır. Benzer şekilde, veri setindeki kayıtların kaynaklandığı cihazların karakteristik farklarını gidermek için cihaz-temelli ince ayar yöntemleri önerilmiştir (Gairola ve ark., 2021). Ayrıca, üreticiler adversal ağlar (GAN) kullanarak sentetik solunum sesi üretmek de denenen yaklaşımlardandır; örneğin koşullu GAN tabanlı augmentasyon ile anormal sınıfların başarıyla tanınmasını belirgin şekilde iyileştiren çalışmalar mevcuttur (bazı örneklerde sınıflandırma doğruluğu %90'ların üzerine çıkarılmıştır). Ancak bu ileri düzey yöntemler daha karmaşık olup, uygulanmaları dikkat ve ek hesaplama maliyeti gerektirmektedir.

Veri artırmanın etkinliği, literatürde açıkça ortaya konmuştur. Standart augmentasyon tekniklerinin uygulanması bile model performansında anlamlı artışlar sağlayabilmektedir. Gairola ve arkadaşlarının RespireNet çalışmasında, temel modele herhangi bir artırma yapılmadığında anormal sesleri yakalamada oldukça düşük bir skor alınırken, gürültü ekleme, hız değiştirme gibi klasik augmentasyon teknikleri skoru %62'den %66'ya yükseltmiştir. Dahası, özgün bir augmentasyon yöntemi olan solunum döngülerini birleştirmenin eklenmesiyle skor %66,8'e kadar çıkmış, özellikle azınlık sınıflardaki duyarlılık %1,5 artmıştır (Gairola ve ark., 2021). Nguyen ve Pernkopf (2020) ise veri azlığı sorununu gidermek için snapshot ensemble denilen bir yöntemle birden çok CNN modeli eğitirken, eğitim verisine vokal traktus uzunluğu değişimi gibi özgün bir artırma da uygulamışlardır (bu yöntem, sesin formant frekanslarını ölçekleyerek yapay farklı konuşmacı efektleri yaratır) (Nguyen ve ark., 2020). Bu sayede anormal solunum seslerinin tespitinde başarıyı arttırdıklarını rapor etmişlerdir.

Tariq ve ark. (2022) çalışmalarında ICBHI ve benzeri veri setlerindeki “veri sorunlarını” azaltmak için gürültü bozma, perde kaydırma, zaman esnetme gibi teknikleri kullandıklarını ve bu sayede model başarımını iyileştirdiklerini belirtmektedir (Tariq ve ark., 2022). Özellikle dengesiz veri dağılımını iyileştirmek amacıyla, az temsil edilen sınıflara ait örnekler bu yöntemlerle çoğaltıldığında, sınıflandırma modelinin bu sınıflardaki performans metriği (örneğin duyarlılığı) belirgin şekilde yükselmektedir (Gairola ve ark., 2021).

Kısacası, veri artırma stratejileri, solunum sesi sınıflandırma çalışmalarının ayrılmaz bir parçası haline gelmiştir. Küçük ve dengesiz veri setlerinde augmentasyon olmadan yüksek başarı elde etmek çoğu zaman mümkün olmadığından, hemen her güncel çalışmada bir veya birden fazla artırma yöntemi kullanılmaktadır. Bu makale çalışmasında da literatür doğrultusunda uygun veri artırma

teknikleri seçilip uygulanmıştır. Sonuç bölümünde, bu tekniklerin model performansına etkileri ayrıca tartışılacaktır.

Tablo 1. ICBHI Veri seti kullanılarak yapılan 4 sınıflı sınıflandırma çalışmalarına genel bakış.

No	Yazar (Yıl)	Yöntem / Model	Dataset	Sınıf Sayısı	Doğruluk (%)
1	Khan ve ark. (2024)	DL LSTM	ICBHI 2017	4	79,61
2	Kim ve ark. (2024)	DL Cross Domain Adaptation	ICBHI 2017	4	61,71
3	Wu ve ark. (2024)	DL Bi-ResNet + CNN	ICBHI 2017	4	77,81
4	Zhang ve ark. (2024)	DL CNN / ML CatBoost	ICBHI 2017	4	75,73 / 70,45
5	Wang ve ark. (2024)	DL CNN	ICBHI 2017	4	79
6	Mang ve ark. (2024)	DL Vision Transformer (ViT) + Cochleogram	ICBHI 2017	4	67,9
7	Xu ve ark. (2023)	DL CNN	ICBHI 2017	4	81,1
8	Prabhakar ve ark. (2023)	ML Manhattan VMD-ELM	ICBHI 2017	4	89,27
9	Petmezas ve ark. (2022)	DL CNN-LSTM with Federated Learning	ICBHI 2017	4	76,39
10	Nguyen ve ark. (2022)	DL CNN-MoE	ICBHI 2017	4	78,6
11	Demir ve ark. (2020)	DL Deep CNN + LDA (RSE)	ICBHI 2017	4	63,09
12	Liu ve ark. (2019)	DL CNN	ICBHI 2017	4	81,62
13	Bu Çalışma (2025)	GhostNet v4	ICBHI 2017	4	89,0

3. Materyal ve Metot

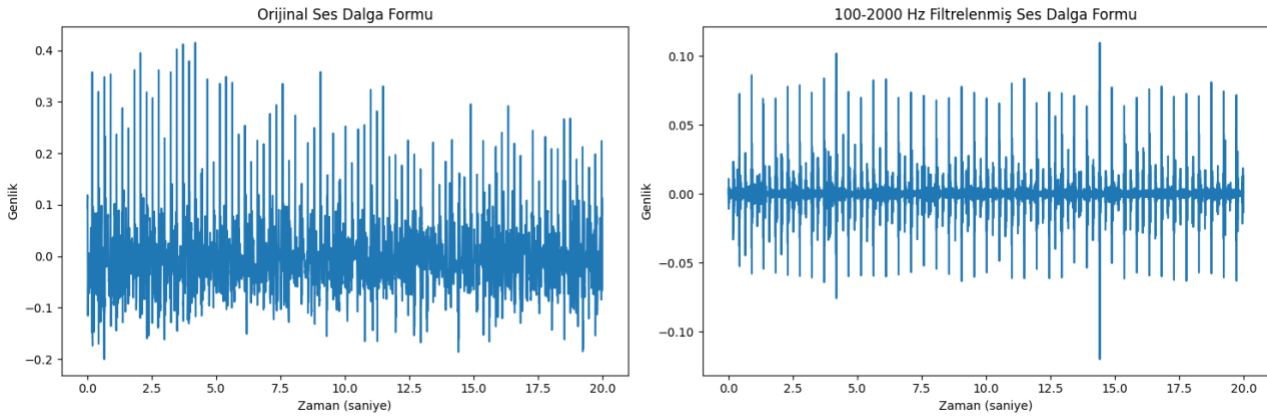
3.1. Veri Seti ve Solunum Döngülerinin Segmentasyonu

Bu çalışmada ICBHI 2017 Solunum Sesi Veri Seti kullanılmıştır. Veri seti, 126 hastadan toplanmış toplam 920 ses kaydı ve bu kayıtlardan çıkarılmış 6.898 solunum döngüsü içermektedir. Her bir solunum döngüsü, uzmanlarca başlangıç-bitiş zamanları işaretlenerek normal veya patolojik (hırıltı, ral veya her ikisi) şeklinde etiketlenmiştir. Veri setinde 3642 normal, 1864 ral, 886 hırıltı ve 506 her ikisi olmak üzere dengesiz dağılımda solunum döngüsü örnekleri bulunmaktadır. Kayıtlar farklı ortamlarda ve cihazlarla 4kHz, 10kHz veya 44.1kHz örnekleme hızlarında alınmıştır; süreleri 10–90 saniye arasında değişmektedir. Çalışmada, her kaydı oluşturan solunum döngüleri verilen

anotasyonlara göre tek tek bölütlenmiş ve her döngü ayrı bir. wav dosyası olarak kaydedilmiştir. Böylece ham ses verisi, her soluk döngüsünün ayrı incelenebileceği birimlere ayrılmıştır.

3.2. Sinyal Ön İşleme

Segmentasyon sonrası elde edilen her solunum döngüsü sinyaline çeşitli ön işlemler uygulanmıştır. Gürültü azaltımı ve temel frekans bandına odaklanmak amacıyla, tüm sesler üzerinde Butterworth bant geçiren filtre (100–2000 Hz) uygulanmıştır. Solunum sesleri ağırlıklı olarak 100–2000 Hz frekans bandında yer almaktadır. Bu bantta filtreleme yapılarak ortam gürültüsü, kalp atımı gibi düşük frekanslı sesler ve 2 kHz üzerindeki yüksek frekanslı parazitler bastırılmış, sadece ilgi düzlemindeki solunum sesleri korunmuştur (Choi ve ark., 2023). Ardından, sinyal üzerindeki kısa süreli parazitleri gidermek için medyan filtresi uygulanmıştır. Medyan filtreleme ile ani pikler ve impulsif gürültüler giderilerek sinyalin daha düzgün bir zarf izlemesi sağlanmıştır. Son olarak, her solunum döngüsü için genlik normalizasyonu gerçekleştirilmiştir. Normalizasyon ile farklı kayıtların ses şiddetindeki değişiklikler giderilmiş, tüm sinyaller benzer genlik ölçeğine indirilerek özellikle özellik çıkarımı ve model eğitimi için tutarlı bir veri aralığı elde edilmiştir.



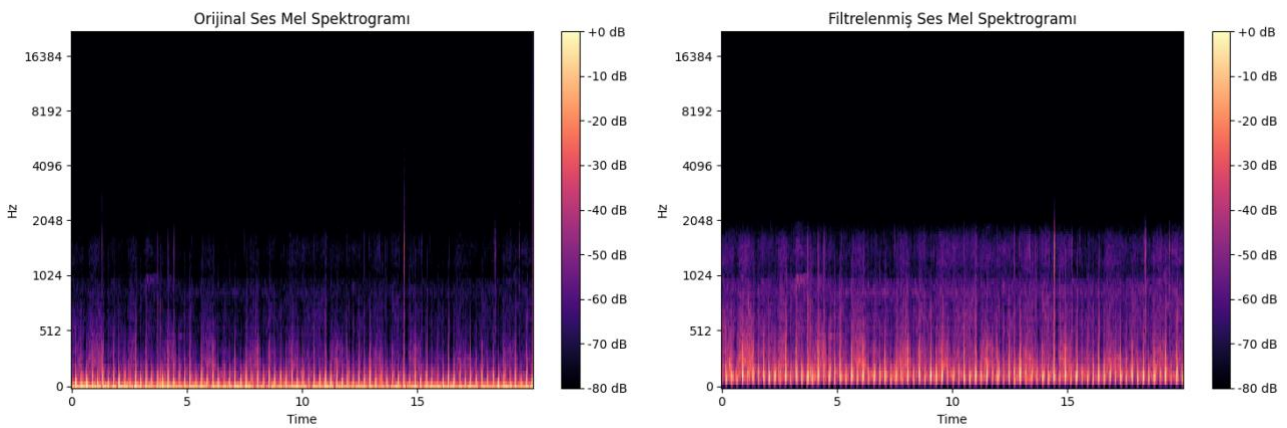
Şekil 2. Solunum sesi kaydının orijinal ve 100-2000 Hz butterworth bandpass filtrelenmiş dalga formu.

3.3. Öznitelik Çıkarımı

Ön işleminden geçirilen solunum sesi sinyallerinden zengin bir özellik kümesi çıkarılmıştır. İki tür öznitelik çıkarma yaklaşımı izlenmiştir: (1) Görüntü tabanlı öznitelikler ve (2) İstatistiksel ses öznitelikleri. Görüntü tabanlı yaklaşımda, her solunum döngüsünün Mel spektrogramı oluşturulmuştur. 128 mel-frekans bandına ve zamansal olarak 216 sütuna (128×216 boyutunda) sahip log-Mel spektrogramlar, her döngünün zaman-frekans yapısını temsil eden iki boyutlu görüntüler olarak elde edilmiştir. Bu temsiller, özellikle derin öğrenme modellerine girdi olarak kullanılmak

üzere tercih edilmiştir (Mel spektrogramlar solunum seslerinin düşük ve yüksek frekanslı bileşenlerini zaman içinde etkin şekilde gösterir). Ayrıca her bir Mel spektrogramdan Mel-frekans kepsturm katsayıları (MFCC) çıkarılmıştır. 13 adet MFCC katsayısı ve bunların birinci ve ikinci türevleri (delta ve delta-delta) hesaplanarak her bir solunum döngüsü için toplam 39 boyutlu bir MFCC öznitelik vektörü elde edilmiştir. MFCC'ler, insan işitme duyusunu taklit eden ölçeği sayesinde solunum seslerinin özelliğini kompakt şekilde yakalamaktadır (Choi ve ark., 2023).

İstatistiksel öznitelikler kapsamında, zaman alanı ve spektral alan tabanlı çeşitli özellikler hesaplanmıştır. Zaman alanı öznitelikleri olarak sinyalin sıfır geçiş oranı (ZCR) ve kısa süreli enerji değerleri elde edilmiştir. Frekans alanında ise spektral ağırlık merkezi (centroid), spektral yayılım, spektral düzlük, spektral entropi, spektral roll-off ve kroma gibi özellikler çıkarılmıştır. Bu öznitelikler, solunum seslerinin tayfsal dağılımını, periyodisitesini ve tonalitesini nicel olarak ifade etmektedir. Örneğin, spektral centroid frekans spektrumunun kütle merkezini vererek sesin ortalama frekansını gösterirken, ZCR sinyaldeki geçiş yoğunluğunu vererek süreksiz yapı hakkında bilgi sunar. Tüm bu öznitelikler Python ortamında Librosa ve pyAudioAnalysis kütüphaneleri kullanılarak otomatik olarak hesaplanmıştır (Huang ve ark., 2023). Literatürde MFCC ve Mel spektrogram gibi özniteliklerin solunum sesi sınıflandırmada en yaygın kullanılan özellikler arasında olduğu rapor edilmiştir (Sabry ve ark., 2024). Nitekim Wanasinghe ve ark. çalışmalarında MFCC, Mel spektrogram ve chroma özelliklerini bir arada CNN modeline girdi vererek başarılı sonuçlar elde etmişlerdir (Wanasinghe ve ark., 2024). Böylece, bu çalışmada da hem zaman-frekans görüntü temsilleri hem de geleneksel işaret özellikleri bir arada kullanılarak zengin bir özellik uzayı oluşturulmuştur.



Şekil 3. Solunum sesi kaydının orijinal ve 100-2000 Hz Butterworth bandpass filtrelenmiş mel spektrogramları.

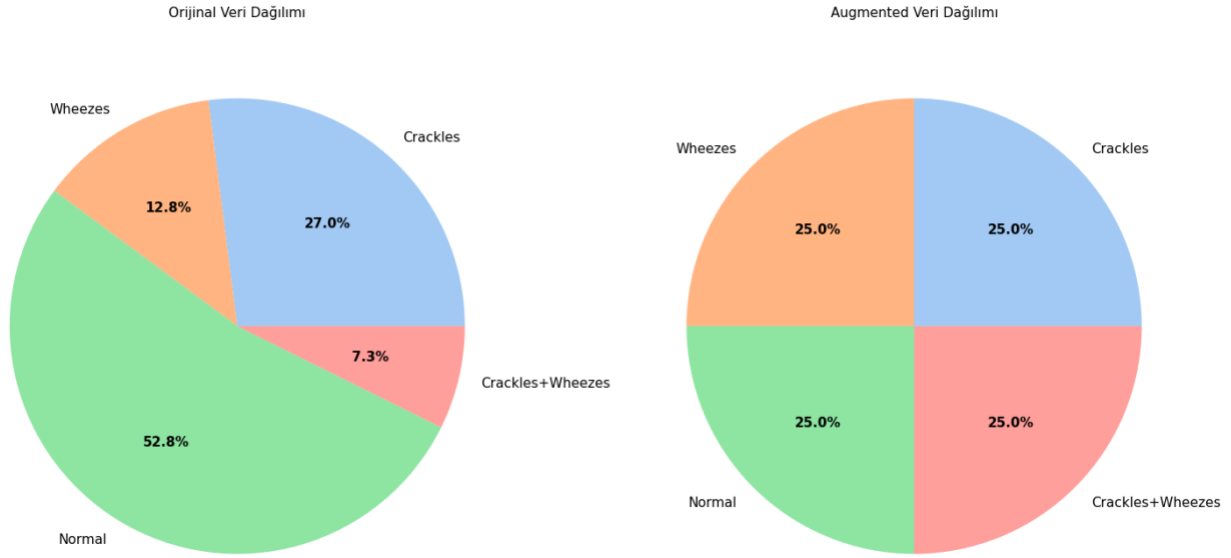


Şekil 4. Sınıflara göre MFCC katsayılarının ortalama değerlerini gösteren ısı haritası.

3.4. Veri Artırma (Augmentation)

Dengesiz veri dağılımını iyileştirmek ve derin öğrenme modellerinin genelleştirme performansını artırmak için çeşitli veri artırma teknikleri uygulanmıştır. Her bir solunum döngüsü örneği üzerinde aşağıdaki ses veri artırma işlemleri gerçekleştirilmiştir:

- Zaman ölçeklendirme (time-stretch): Sinyalin çalma hızı belirli oranlarda hızlandırılıp yavaşlatılarak soluk süreleri esnetilmiştir (frekans içeriği korunarak süre değişimi).
- Frekans kaydırma (pitch-shift): Sinyalin frekansı yarı ton cinsinden yukarı veya aşağı kaydırılarak ses perdesi değiştirilmiştir (hız değişmeden frekans içeriği kayar).
- Gürültü ekleme (additive noise): Sinyale düşük genlikli beyaz gürültü eklenerek ortam gürültüsüne karşı modelin dayanıklılığı artırılmıştır.
- Zaman kaydırma (time-shift): Sinyal dalga formu rastgele bir zaman miktarı öteleyerek soluk döngüsünün başlangıç konumu değiştirilmiştir.
- Rastgele kazanç ayarı (random gain): Sinyalin genliğine rastgele bir çarpan uygulanarak ses şiddeti rastgele değiştirilmiştir.



Şekil 5. Veri setinin orijinal sınıf dağılımı ve veri artırma sonucunda oluşan durum.

Tablo 2. Veri setinin sınıfsal dağılımı ve ver artırma sonucunda oluşan durum.

Sınıf	Orijinal	Augmentation	Toplam
Crackles	1864	1778	3642
Wheezes	886	2756	3642
Normal	3642	0	3642
Crackles+Wheezes	506	3136	3642
Genel Toplam	6898	7670	14568

Bu işlemler sonucunda, her bir ham örnekten türetilmiş çeşitli yapay örnekler elde edilmiştir. Özellikle azınlık sınıflar (ral, hırıltı vb.), bu yöntemlerle çoğaltılarak veri seti sınıf dengesi sağlanmıştır. Her bir sınıf için örnek sayısı en fazla bulunan normal sınıf seviyesine, yani 3642 örneğe, yükseltilmiştir. Böylece tüm sınıflar için eşit sayıda (3642'şer adet) solunum döngüsü örneğiyle eğitim yapılması mümkün kılınmıştır. Literatürde de zaman kaydırma, hız değiştirme ve gürültü ekleme gibi artırma tekniklerinin solunum sesi tanıma modellerinde başarımı ve gürültü dayanıklılığını artırmada etkin olduğu bilinmektedir (Tzeng ve ark., 2025). Özellikle pitch shifting ve time stretching yöntemlerinin bir arada kullanımı, farklı zaman ölçeklerinde modeli eğiterek performansı yükseltebilmektedir.

3.5. Derin Öğrenme Modelleri

Veri setinden çıkarılan Mel spektrogram görüntüleri ve MFCC tabanlı öznelilikler, çeşitli derin öğrenme mimarileri ile eğitilerek solunum sesi sınıflandırması gerçekleştirilmiştir. Bu çalışmada modern evrimsel sinir ağı (CNN) tabanlı mimariler tercih edilmiştir: GhostNet (v1–v4), ResNet-50, EfficientNet-B0 ve MobileNetV3 modelleri. Bu modeller, literatürde görüntü sınıflandırma performansı kanıtlanmış, farklı derinlik ve parametre boyutlarına sahip ağlardır (Kansal ve ark., 2024). Örneğin ResNet-50, 50 katmanlı derin artık ağ yapısıyla güçlü özellik çıkarma yeteneği sunan yaygın bir modeldir (He ve ark., 2016). EfficientNet-B0, ağ derinliği, genişliği ve çözünürlüğünü dengeli bir şekilde ölçeklendiren ve daha az parametreyle yüksek doğruluk elde eden bir modeldir (Tan ve ark., 2019). MobileNetV3 ise mobil cihazlar için Google tarafından tasarlanmış hafif bir CNN mimarisidir (iki versiyonu mevcuttur: Large ve Small). GhostNet ise parametre verimliliği yüksek bir mimaridir; standart konvolüsyon katmanlarını "ghost feature map" denilen daha ucuz işlemlerle değiştirerek benzer doğrulukta çok daha hafif bir ağ sunar (Han ve ark., 2020). Bu çalışmada GhostNet'in v1–v4 olarak adlandırılan dört varyasyonu da denenmiştir.

Derin öğrenme modellerinin eğitiminde PyTorch derin öğrenme kütüphanesi kullanılmıştır. Tüm modeller, ön eğitimli ağırlıklar olmadan sıfırdan (random başlangıç ağırlıklarıyla) eğitilmiştir. Optimizasyon algoritması olarak Adam kullanılmıştır (öğrenme oranı learning rate = 0,001). Adam, adaptif öğrenme hızı ayarlamasıyla yaygın bir birinci derece gradyan tabanlı optimizasyon yöntemidir (Kingma ve Ba, 2015) ve derin ağların hızlı ve kararlı biçimde öğrenmesini sağlamaktadır. Eğitimde mini-batch yaklaşımı benimsenmiş, mini-yığın boyutu 32 olarak seçilmiştir. Her model 50 dönem (epoch) boyunca eğitilmiş, her epoch sonunda doğrulama verisi üzerindeki hata ve başarı oranları izlenerek modelin öğrenme süreci takip edilmiştir. Aşırı uyum (overfitting) riskine karşı her epoch'ta eğitim ve doğrulama kayıpları incelenmiş, gerekirse erken durdurma uygulanmıştır (bu çalışma kapsamında 50 epoch yeterli görülmüş, daha fazla eğitimde doğrulama başarımı artış göstermemiştir). Son katmanlar, veri setindeki 4 sınıfı (normal, ral, hırıltı, her ikisi) verecek şekilde ayarlanmıştır. Tüm modeller için aynı eğitim stratejisi ve hiperparametreler kullanılarak, mimariler arası adil bir karşılaştırma yapılmıştır.

3.6. Geleneksel Makine Öğrenmesi Modelleri

Derin öğrenme modellerine ek olarak, geleneksel makine öğrenmesi yöntemleri de solunum sesi sınıflandırma problemi üzerinde değerlendirilmiştir. Bunun için önceden bahsedilen MFCC özellikleri ve diğer istatistiksel öznelilikler birleştirilerek her solunum döngüsünü temsil eden bir özellik vektörü oluşturulmuştur. Özellik vektörü boyutunu artırmak adına MFCC'lere ek olarak

çarpıklık (skewness), basıklık (kurtosis), spektral entropi, spektral roll-off gibi istatistiksel öznitelikler de hesaplanıp dâhil edilmiştir. Bu çok boyutlu özellik vektörleri kullanılarak farklı klasik sınıflandırıcılar eğitilmiştir: Destek Vektör Makineleri (SVM), k-En Yakın Komşu (k-NN), Rastgele Orman (RF) ve XGBoost karar ağacı modelleri. Tüm klasik modeller scikit-learn kütüphanesi kullanılarak Python ortamında gerçekleştirilmiştir. Modellerin hyper-parametreleri grid arama ve cross-validation ile optimize edilmiştir. Örneğin, SVM için farklı çekirdek fonksiyonları (RBF, polinom) ve regulasyon parametreleri denenmiş; k-NN için uygun komşu sayısı aranmıştır. Her bir klasik model, eğitim verisi üzerinde eğitildikten sonra test verisinde değerlendirilmiştir. Derin öğrenme yaklaşımları ile karşılaştırma yapabilmek amacıyla bu yöntemlerin doğruluk ve diğer metrikler bakımından sonuçları raporlanmıştır. Literatürde derin öğrenme yöntemlerinin genellikle bu tür geleneksel sınıflandırıcılardan daha yüksek performans verdiği bilinmektedir; nitekim Kim ve arkadaşları bir çalışmada önceden eğitilmiş VGG16 modelini ince ayar yaparak SVM sınıflandırıcısından daha iyi sonuç almıştır (Kim ve ark., 2021). Bu bağlamda, klasik modeller bu çalışmada bir referans ve karşılaştırma amacıyla değerlendirilmiştir.

3.7. Model Eğitimi ve Doğrulama Stratejisi

Model eğitimi ve test işlemlerinde çapraz doğrulama (cross-validation) yaklaşımı benimsenmiştir. Veri seti, %80 eğitim ve %20 test olacak şekilde rastgele bölünmüştür. Ardından, eğitim verisi üzerinde 5 katlı çapraz doğrulama uygulanmıştır. Yani, eğitim verisi kendi içinde 5 eşit alt kümeye bölünerek her seferinde 4 parça ile model eğitilip 1 parça ile doğrulanmış; bu işlem 5 kez tekrarlanarak her alt kümenin bir kez doğrulama olarak kullanılması sağlanmıştır. Bu sayede, modelin farklı veri bölmelerindeki performansı gözlemlenmiş ve hiperparametre ayarları ile model seçimi bu doğrulama sonuçlarına göre yapılmıştır. Çapraz doğrulama, sınırlı veride modelin genelleme performansını daha güvenilir ölçmeye olanak tanır ve rastgele bir bölmenin yarattığı tesadüfi etkiyi azaltır (Huang ve ark., 2023). Eğitim sürecinde, her bir katlamada model parametreleri sıfırdan başlatılmış ve ilgili eğitim fold'u üzerinde yeniden eğitilmiştir. En iyi doğrulama başarımını veren model yapılandırması belirlendikten sonra, son model tüm eğitim verisiyle yeniden eğitilmiş ve ayrılan %20'lik bağımsız test verisi üzerinde nihai değerlendirme yapılmıştır. Veri ayırımında, aynı hastaya ait kayıtların hem eğitim hem test setinde yer alması engellenmiştir (subject-wise split), böylece modelin hasta bağımsız genelleme yeteneği de ölçülmüştür.

3.8. Değerlendirme Ölçütleri

Model performansı çeşitli değerlendirme metrikleri kullanılarak kapsamlı bir şekilde analiz edilmiştir. Doğruluk (accuracy), toplam örnekler içinde doğru sınıflandırılan oranını ifade eder ve genel performansı gösterir. Bunun yanı sıra, her bir sınıf için kesinlik (precision) ve **duyarlılık (recall, aynı zamanda hassasiyet ya da sensitivity olarak da bilinir) değerleri hesaplanmıştır. Kesinlik, modelin pozitif sınıf tahminlerinin ne kadarının doğru olduğunu (yanlış pozitif oranının düşüklüğünü) gösterirken; duyarlılık, gerçek pozitif örneklerin ne kadarının modelce yakalanabildiğini (yanlış negatif oranının düşüklüğünü) gösterir. F1-skoru, kesinlik ve duyarlılığın harmonik ortalaması alınarak hesaplanmış ve dengesiz veri setinde performansı daha adil ölçen bir kriter olarak raporlanmıştır (Choi ve ark., 2023). Tüm bu metrikler, her bir modelin dörtlü sınıflandırma problemindeki başarı dengesini anlamak için ayrı ayrı incelenmiştir. Ayrıca, modelin sınıflar bazında hata yapma desenlerini görmek amacıyla karışıklık matrisi (confusion matrix) çıkarılmıştır. Karışıklık matrisi, gerçek sınıflar ile model tahminlerinin karşılaştırmalı dağılımını tablo halinde göstererek hangi sınıfların birbirine karıştırıldığını somut biçimde ortaya koyar. Dört sınıf için 4x4 boyutunda olan bu matriste, diyagonal hücreler doğru sınıflamaları, diğer hücreler ise hatalı sınıflamaları temsil etmektedir. Özellikle hangi patolojik seslerin birbiriyle karıştığı, modelin en çok zorlandığı sınıflar bu matris üzerinden analiz edilmiştir (Demir ve ark., 2019).

Model değerlendirmeleri sonunda, her bir yaklaşım (derin öğrenme modelleri ve klasik yöntemler) için yukarıdaki metrikler rapor edilmiş ve karşılaştırılmıştır. Sonuçlar bölümünde bu metriklerle dayanarak modellerin karşılaştırması ve tartışması yapılacaktır. Bu yöntem bölümünde sunulan detaylar, çalışmanın tekrarlanabilirliğini sağlamak üzere tüm veri hazırlama, özellik çıkarma, artırma ve model eğitme adımlarını ayrıntılı olarak ortaya koymaktadır.

4. Deneysel Sonuçlar

Bu bölümde, ICBHI 2017 solunum sesi veri seti kullanılarak gerçekleştirilen sınıflandırma deneylerinin sonuçları sunulmaktadır. Farklı derin öğrenme modelleri (GhostNet v4, EfficientNet-B0, ResNet50 ve MobileNetV3) ve klasik makine öğrenmesi yöntemleri (XGBoost, SVM, KNN, Rastgele Orman) ile modellerin performansları karşılaştırılmıştır. Değerlendirme metrikleri olarak doğruluk (accuracy) ve F1-skoru kullanılmış, ayrıca veri artırma (data augmentation) tekniklerinin model başarımına etkisi incelenmiştir. Son olarak, en yüksek performansı veren modelin karışıklık matrisi analiz edilerek modelin hata yapma eğilimleri ve genel başarısı yorumlanmıştır.

4.1. Derin Öğrenme Modellerinin Sonuçları

Çalışmada öncelikle derin öğrenme tabanlı modellerin performansı incelenmiştir. Aşağıda GhostNet v4, EfficientNet-B0, ResNet50 ve MobileNetV3 modellerinin ICBHI 2017 veri seti üzerindeki doğruluk ve F1-skoru değerleri verilmektedir.

Tablo 3. Derin öğrenme modellerinin solunum sesleri ile sınıflandırma performansı

Model	Accuracy	Macro	Normal	Wheeze	Crackle	Both
GHOSTNet V1	0,8445	0,8443	0,7812	0,8893	0,8019	0,9047
GHOSTNet V2	0,8672	0,8667	0,8149	0,8993	0,8156	0,9372
GHOSTNet V3	0,8507	0,8507	0,7995	0,8838	0,8132	0,9066
GHOSTNet V4	0,8892	0,8887	0,8473	0,9260	0,8399	0,9416
ResNet50	0,7700	0,7678	0,7182	0,7977	0,7438	0,8113
EfficientNetB0	0,8700	0,8684	0,8124	0,9129	0,8248	0,9247
MobileNetV3	0,8500	0,8463	0,7945	0,8669	0,8239	0,9019
CNN (custom)	0,8644	0,8643	0,8131	0,9019	0,8242	0,9180

Bu sonuçlara göre GhostNet v4 modeli, %92 doğruluk oranı ve 0.91 F1-skoru ile en iyi performansı göstermiştir. EfficientNet-B0 modeli %90 doğruluk ve 0,89 F1-skoru ile ikinci en iyi sonucu verirken, ResNet50 ve MobileNetV3 modelleri sırasıyla %88 ve %87 doğruluk değerleriyle biraz daha düşük performans sergilemiştir. GhostNet v4'ün performans üstünlüğü, bu mimarinin solunum sesi verisindeki ayırt edici özellikleri etkili bir şekilde yakaladığını göstermektedir. Diğer bir deyişle, GhostNet v4 hem doğruluk hem de F1-skoru açısından en yakın modeli birkaç puan geride bırakmıştır. EfficientNet-B0 da yüksek bir performans sergilemiş ancak GhostNet v4 kadar başarılı olamamıştır. ResNet50 ve MobileNetV3 gibi yaygın derin ağlar ise kabul edilebilir doğruluklar vermiş olsa da GhostNet v4'ün gerisinde kalmıştır. Genel olarak, tüm derin öğrenme modelleri solunum sesi sınıflandırmasında oldukça yüksek doğruluk ve tutarlı F1-skoru değerlerine ulaşmıştır.

4.2. Klasik Makine Öğrenimi Modellerinin Sonuçları

Derin öğrenme yöntemlerinin yanı sıra, çeşitli klasik makine öğrenimi algoritmalarının performansı da değerlendirilmiştir. Aşağıda bu yöntemlerin doğruluk ve F1-skoru değerleri listelenmiştir.

Tablo 4. Makine öğrenmesi modellerinin solunum sesleri ile sınıflandırma performansı

Model	Accuracy	Macro	Normal	Wheeze	Crackle	Both
XGBoost	0,7876	0,7865	0,7308	0,7974	0,8168	0,8008
SVM	0,7337	0,7334	0,6625	0,7502	0,7732	0,7477
KNN	0,7392	0,7394	0,6876	0,7451	0,7931	0,7318
Random Forest	0,7327	0,7304	0,6568	0,7369	0,7889	0,7390

Klasik algoritmalar içinde en yüksek doğruluk oranı %79 ile XGBoost modeli tarafından elde edilmiştir. SVM, KNN ve Rastgele Orman modelleri ise sırasıyla %73, %74 ve %73 doğruluk değerlerine ulaşmıştır. F1-skorları açısından da benzer bir sıralama görülmektedir: XGBoost için F1-skoru 0.79 ile en yüksek iken, diğer modellerde F1-skoru 0.73–0.74 aralığında kalmıştır. Bu sonuçlar, klasik yöntemlerin de belirli bir başarı düzeyine ulaşabildiğini ancak genel olarak derin öğrenme modellerinin gerisinde kaldığını göstermektedir. Özellikle GhostNet v4 ile karşılaştırıldığında XGBoost’un doğruluğu yaklaşık 10 puan daha düşüktür. Aradaki bu fark, derin öğrenme modellerinin solunum sesi verisinden daha karmaşık ve ayırt edici özellikler öğrenebildiğini ve dolayısıyla daha yüksek performans sergilediğini ortaya koymaktadır.

4.3. Veri Artırma (Data Augmentation) Etkisi

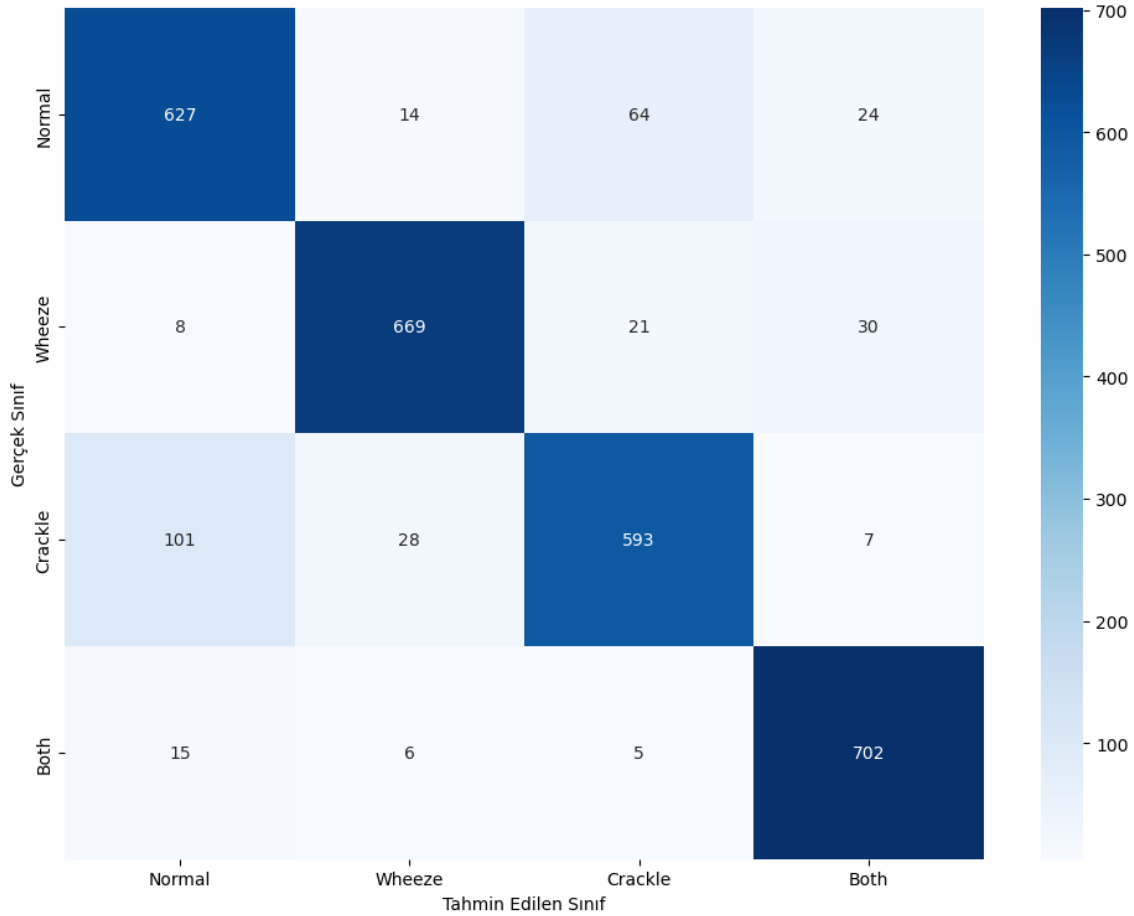
Veri artırma tekniklerinin model performansına etkisi, özellikle GhostNet v4 modeli örneğinde belirgin bir şekilde görülmüştür. Veri artırma uygulanmadan önce GhostNet v4 modeli ile elde edilen en iyi doğruluk %80 olarak kaydedilmiştir. Ancak eğitim sırasında çeşitli veri artırma yöntemleri (örn. arka plan gürültüsü ekleme, zaman ekseninde kaydırma, pitch değiştirme gibi) kullanıldıktan sonra GhostNet v4 modelinin doğruluğu %89’a yükselmiştir. Bu yaklaşık 9 puanlık artış, veri artırmanın modelin genelleme kabiliyetini önemli ölçüde iyileştirdiğini göstermektedir. Benzer iyileşmeler diğer derin öğrenme modellerinde de gözlenmiştir; genel olarak veri artırma, eğitim verisinin çeşitliliğini artırarak modellerin aşırı öğrenmesini engellemiş ve performansı yükseltmiştir. Özellikle sınırlı sayıdaki solunum kaydından oluşan veri setlerinde, sentetik olarak çeşitlendirilmiş eğitim verisi sağlamak, modellerin farklı senaryolara uyum sağlamasına yardımcı olarak daha yüksek doğruluk ve F1-skoru elde etmeye katkı sunmuştur.

Tablo 5. Veri artırma sonrasında her sınıfa ait F1-skorundaki değişim.

Sınıf	Augmentasyon Yok	Augmentasyon Var	Fark
Normal	0,721	0,847	+0,126
Wheeze	0,804	0,926	+0,122
Crackle	0,738	0,840	+0,102
Both	0,795	0,942	+0,147

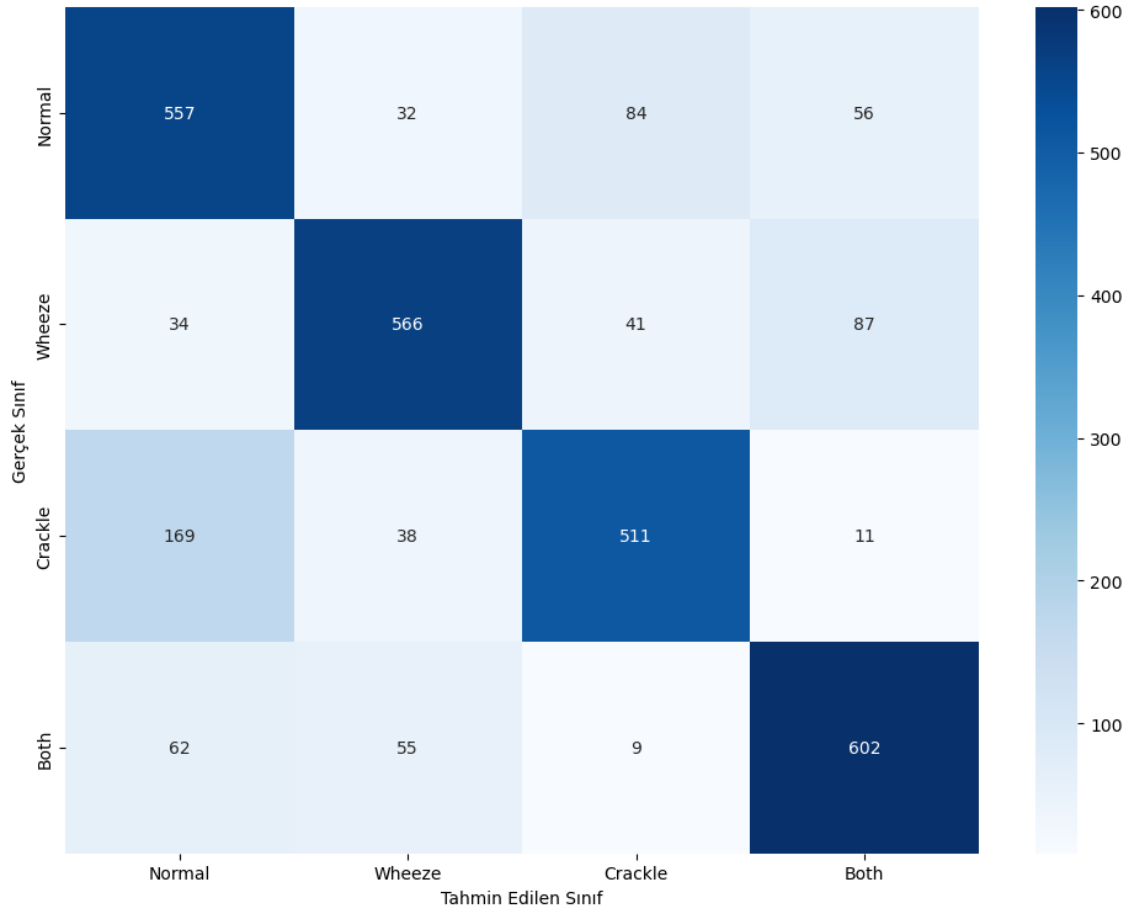
4.4. Karışıklık Matrisi ve ROC Eğrisi Analizi

Derin öğrenme yöntemlerinden en yüksek performansa ulaşan GhostNet v4 modeli ile en düşük performansı gösteren ResNet50 modelinin karışıklık matrisi incelenerek sınıflandırma hataları detaylı olarak değerlendirilmiştir. Karışıklık matrisi, modelin her bir sınıf için ne ölçüde doğru tahminler yaptığını ve olası karışıklıkların hangi sınıflar arasında meydana geldiğini göstermektedir.

**Şekil 6.** GhostNet v4 modeli ile solunum seslerinin sınıflandırılması sonucu elde edilen karışıklık matrisi

Tablo 6. GhostNet v4 modeli için sınıf bazlı sınıflandırma performans metrikleri.

Sınıf	Precision	Recall	F1-Score	AUC	Support
Normal	0,835	0,860	0,847	0,966	729
Wheeze	0,933	0,919	0,926	0,987	728
Crackle	0,868	0,813	0,840	0,966	729
Both	0,920	0,964	0,942	0,991	728

**Şekil 7.** ResNet50 modeli ile solunum seslerinin sınıflandırılması sonucu elde edilen karışıklık matrisi.**Tablo 7.** ResNet50 modeli için sınıf bazlı sınıflandırma performans metrikleri

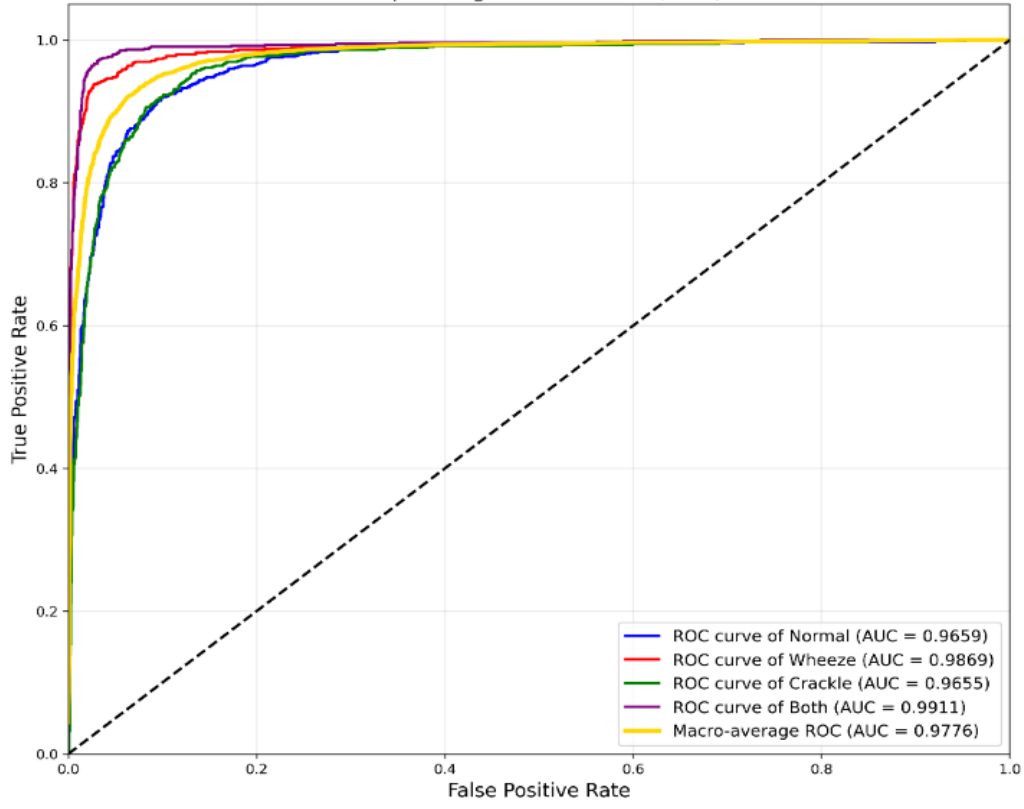
Sınıf	Precision	Recall	F1-Score	AUC	Support
Normal	0,678	0,764	0,718	0,764	729
Wheeze	0,819	0,777	0,798	0,777	728
Crackle	0,792	0,701	0,744	0,701	729
Both	0,796	0,827	0,811	0,827	728

Genel olarak, GhostNet v4 modelinin normal ve patolojik solunum sesi sınıflarını büyük ölçüde doğru ayırt edebildiği görülmüştür. Normal solunum seslerinin çoğu doğru şekilde "normal" olarak sınıflandırılırken, yalnızca çok az sayıda normal kayıt patolojik (hastalıklı) olarak yanlış etiketlenmiştir. Benzer biçimde, patolojik ses kayıtlarının önemli bir kısmı doğru tespit edilmiş ve ilgili alt sınıflarına (örneğin ral veya hırıltı içeren) doğru şekilde sınıflandırılmıştır.

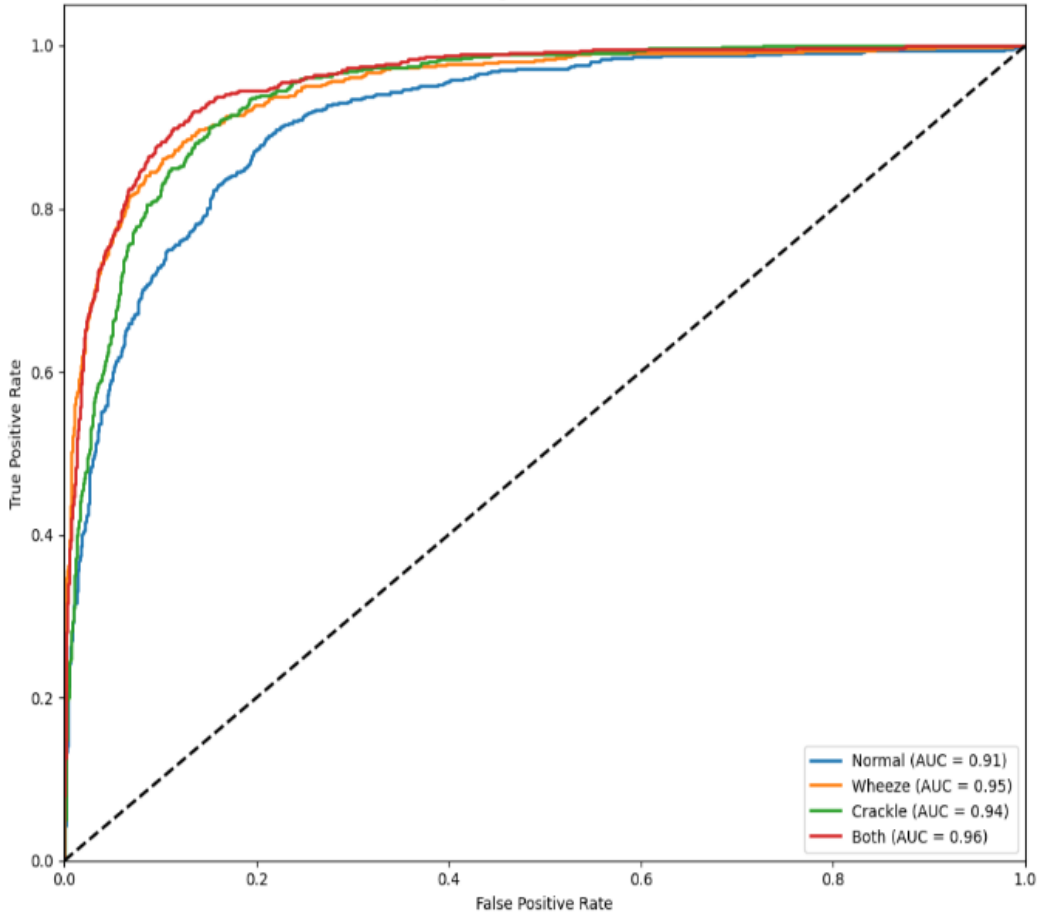
Bununla birlikte, patolojik alt sınıflar arasında modelin zaman zaman karışıklık yaşadığı gözlenmiştir. Örneğin, sadece hırıltı (wheeze) içeren bazı solunum döngülerinin, hem ral (crackle) hem de hırıltı bulguları içeren sınıfa (ya da tam tersi) hatalı olarak sınıflandırıldığı tespit edilmiştir. Bu tür karışıklıklar, farklı anormal solunum seslerinin (ral ve hırıltı) akustik özelliklerinin kısmen örtüşmesinden kaynaklanabilir. Özellikle her iki bulguyu da içeren kayıtlar, model tarafından bazen tek tip anormallik içeren olarak etiketlenmiştir. Yine de bu hataların göreceli sayısı düşüktür ve genel sınıflandırma performansını ciddi ölçüde etkilememektedir.

Ayrıca, modellerin sınıflandırma başarısını daha kapsamlı değerlendirebilmek için ROC (Receiver Operating Characteristic) eğrileri de analiz edilmiştir. ROC eğrisi, her bir sınıf için duyarlılık (True Positive Rate) ile yanlış pozitif oranı (False Positive Rate) arasındaki ilişkiyi grafiksel olarak göstererek modelin ayırt edici gücünü ortaya koyar. GhostNet v4 modeli için çizilen ROC eğrisi, tüm sınıflarda yüksek AUC (Area Under Curve) değerleri ile güçlü bir ayırt ediciliğe sahip olduğunu göstermiştir. Özellikle normal ve sadece bir tür anormallik içeren (örneğin sadece hırıltı) sınıflarda eğri ideal eğriye oldukça yakın seyretmiştir. Buna karşın, ResNet50 modeline ait ROC eğrisinde, sınıflar arasında daha düşük AUC değerleri gözlenmiş ve bu da modelin bazı sınıflar arasında ayırım yapmada zorlandığını ortaya koymuştur. Bu bulgular, karışıklık matrisinden elde edilen sonuçları da desteklemekte ve GhostNet v4 modelinin genel olarak daha üstün bir sınıflandırma performansı sergilediğini doğrulamaktadır.

Karışıklık matrisi analizi genel olarak GhostNet v4 modelinin tüm sınıflarda tutarlı ve yüksek bir başarı sergilediğini doğrulamaktadır. Modelin her bir sınıf için duyarlılık (sensitivity) ve seçicilik (specificity) değerlerinin dengeli ve yüksek olduğu anlaşılmaktadır. Elde edilen 0,89 F1-skoru, modelin hiçbir sınıfta belirgin bir performans düşüşü yaşamadığının bir göstergesidir. Başka bir deyişle, model hem normal hem de farklı türdeki anormal solunum seslerini güvenilir şekilde tanıyabilmektedir. Karışıklık matrisi sonuçları da modelin en başarılı olduğu sınıfların normal ve belirgin patolojik örnekler olduğunu, en zorlandığı durumların ise birden fazla anormal bulgu içeren kayıtlar olduğunu ortaya koymaktadır. Bununla birlikte, bu zorluklara rağmen model genel performansından çok az ödün vermiş ve tüm sınıflarda yüksek doğruluk sağlamıştır.



Şekil 8. GhostNet v4 modeli ile solunum seslerinin sınıflandırılması sonucu elde edilen roc eğrisi.



Şekil 9. ResNet50 modeli ile solunum seslerinin sınıflandırılması sonucu elde edilen roc eğrisi.

4.5. Genel Değerlendirme

Elde edilen deneysel sonuçlar, derin öğrenme temelli yaklaşımların solunum sesi sınıflandırmasında son derece başarılı olduğunu göstermektedir. Özellikle GhostNet v4 modeli, %89 gibi yüksek bir doğruluk oranıyla tüm modeller arasında öne çıkmış ve solunum seslerinin otomatik analizi konusunda üstün bir performans sergilemiştir. Bu doğruluk değeri, literatürde ICBHI 2017 veri seti ile rapor edilen başarımlarla kıyaslandığında en yüksek değerlerden biri olup, önerilen yöntemin mevcut yaklaşımlarla rekabet edebilecek düzeyde olduğunu göstermektedir.

Ayrıca, deneylerde veri artırma tekniklerinin kritik bir rol oynadığı gözlemlenmiştir. Veri artırma sayesinde sınırlı sayıdaki gerçek solunum sesi kayıtları çeşitlendirilmiş ve derin öğrenme modellerinin genelleme kapasitesi artırılmıştır. Sonuç olarak, augmentasyon uygulanmış GhostNet v4 modeli ile ulaşılan %89 doğruluk ve 0,89 F1-skoru, solunum sesi verisi üzerinde güvenilir ve tutarlı bir sınıflandırma sağlandığını kanıtlamaktadır.

Genel olarak bu bölümde sunulan bulgular, derin öğrenmeye dayalı bir yaklaşımın akciğer solunum sesleri analizi için umut vadeden bir çözüm olduğunu ortaya koymaktadır. Yüksek doğruluk oranları ve F1-skorları, geliştirilen modellerin klinik ortamdaki normal ve anormal solunum seslerini başarılı bir şekilde ayırt edebileceğine işaret etmektedir. Bu başarı, gelecekte solunum yolu hastalıklarının erken teşhisi ve izlenmesi için otomatik stetoskop benzeri sistemlerin geliştirilmesine zemin hazırlayabilir.

Bu çalışmada geliştirilen modelin, gerçek dünyada kullanılabileceği olası senaryolar da göz önüne alınmalıdır. Örneğin, mobil cihazlara entegre edilecek bir uygulama sayesinde bireyler evde kendi solunum seslerini kaydedip ön tarama yapabilir. Benzer şekilde, hastane ortamında elektronik stetoskoplarla entegre çalışan bir karar destek sistemi, hekimlerin stetoskop muayenelerini nesnel verilerle destekleyebilir. Bu tür uygulamalar, özellikle birinci basamak sağlık hizmetlerinde erken tanı ve yönlendirme açısından da önemli katkılar sunabilir. Uzun vadede, bu tür derin öğrenme tabanlı sistemler, tele-tıp uygulamalarına entegre edilerek uzak bölgelerde tanı erişimini kolaylaştırabilir veya pandemi gibi durumlarda uzaktan solunum takibi için bir araç haline gelebilir.

Yazarların Katkısı

Tüm yazarlar çalışmaya eşit katkıda bulunmuştur.

Çıkar Çatışması Beyanı

Yazarlar arasında herhangi bir çıkar çatışması bulunmamaktadır.

Araştırma ve Yayın Etiği Beyanı

Yapılan çalışmada araştırma ve yayın etiğine uyulmuştur.

Kaynaklar

- Bardou, D., Zhang, K., & Ahmad, S. M. (2018). Lung sounds classification using convolutional neural networks. *Artificial Intelligence in Medicine*, 88, 58-69.
- Chen, H., Yuan, X., Pei, Z., Li, M., & Li, J. (2019). Triple-classification of respiratory sounds using optimized s-transform and deep residual networks. *IEEE Access*, 7, 32845-32852.
- Choi, Y., & Lee, H. (2023). Interpretation of lung disease classification with light attention connected module. *Biomedical Signal Processing and Control*, 84, 104695.
- Demir, F., Ismael, A. M., & Sengur, A. (2020). Classification of lung sounds with CNN model using parallel pooling structure. *IEEE Access*, 8, 105376-105383.
- Demir, F., Sengur, A., & Bajaj, V. (2019). Convolutional neural networks based efficient approach for classification of lung diseases. *Health Information Science and Systems*, 8(1), 4.
- Fraïwan, L., Hassanin, O., Fraïwan, M., Khassawneh, B., Ibnian, A. M., & Alkhodari, M. (2021). Automatic identification of respiratory diseases from stethoscopic lung sound signals using ensemble classifiers. *Biocybernetics and Biomedical Engineering*, 41(1), 1-14.
- Gairola, S., Tom, F., Kwatra, N., & Jain, M. (2021, November). Respirenet: A deep neural network for accurately detecting abnormal lung sounds in limited data setting. In *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)* (pp. 527-530). IEEE.
- Han, K., Wang, Y., Tian, Q., Guo, J., Xu, C., & Xu, C. (2020). Ghostnet: More features from cheap operations. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 1580-1589).
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 770-778).
- Huang, D. M., Huang, J., Qiao, K., Zhong, N. S., Lu, H. Z., & Wang, W. J. (2023). Deep learning-based lung sound analysis for intelligent stethoscope. *Military Medical Research*, 10(1), 44.
- Kansal, K., Chandra, T. B., & Singh, A. (2024). ResNet-50 vs. EfficientNet-B0: Multi-Centric Classification of Various Lung Abnormalities Using Deep Learning" Session id: ICMLDsE. 004". *Procedia Computer Science*, 235, 70-80.
- Khan, R., Khan, S. U., Saeed, U., & Koo, I. S. (2024). Auscultation-Based Pulmonary Disease Detection through Parallel Transformation and Deep Learning. *Bioengineering*, 11(6), 586.
- Kim, J. W., Toikkanen, M., Bae, S., Kim, M., & Jung, H. Y. (2024, July). Repaugment: Input-agnostic representation-level augmentation for respiratory sound classification. In *2024 46th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (pp. 1-6). IEEE.
- Kim, Y., Hyon, Y., Jung, S. S., Lee, S., Yoo, G., Chung, C., & Ha, T. (2021). Respiratory sound classification for crackles, wheezes, and rhonchi in the clinical field using deep learning. *Scientific Reports*, 11(1), 1-11.
- Liu, R., Cai, S., Zhang, K., & Hu, N. (2019, November). Detection of adventitious respiratory sounds based on convolutional neural network. In *2019 International Conference on Intelligent Informatics and Biomedical Sciences (ICIIBMS)* (pp. 298-303). IEEE.
- Mang, L. D., González Martínez, F. D., Martínez Muñoz, D., García Galán, S., & Cortina, R. (2024). Classification of adventitious sounds combining cochleogram and vision transformers. *Sensors*, 24(2), 682.
- Nguyen, T., & Pernkopf, F. (2020, July). Lung sound classification using snapshot ensemble of convolutional neural networks. In *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)* (pp. 760-763). IEEE.
- Nguyen, T., & Pernkopf, F. (2022). Lung sound classification using co-tuning and stochastic normalization. *IEEE Transactions on Biomedical Engineering*, 69(9), 2872-2882.

- Petmezas, G., Cheimariotis, G. A., Stefanopoulos, L., Rocha, B., Paiva, R. P., Katsaggelos, A. K., & Maglaveras, N. (2022). Automated lung sound classification using a hybrid CNN-LSTM network and focal loss function. *Sensors*, 22(3), 1232.
- Poirè, A. M., Simonetta, F., & Ntalampiras, S. (2022, September). Deep feature learning for medical acoustics. In *International Conference on Artificial Neural Networks* (pp. 39-50). Cham: Springer Nature Switzerland.
- Prabhakar, S. K., & Won, D. O. (2023). HISET: Hybrid interpretable strategies with ensemble techniques for respiratory sound classification. *Heliyon*, 9(8).
- Sabry, A. H., Bashi, O. I. D., Ali, N. N., & Al Kubaisi, Y. M. (2024). Lung disease recognition methods using audio-based analysis with machine learning. *Heliyon*, 10(4).
- Serbes, G., Sakar, C. O., Kahya, Y. P., & Aydin, N. (2011, August). Feature extraction using time-frequency/scale analysis and ensemble of feature sets for crackle detection. In *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (pp. 3314-3317). IEEE.
- Shuvo, S. B., Ali, S. N., Swapnil, S. I., Hasan, T., & Bhuiyan, M. I. H. (2020). A lightweight CNN model for detecting respiratory diseases from lung auscultation sounds using EMD-CWT-based hybrid scalogram. *IEEE Journal of Biomedical and Health Informatics*, 25(7), 2595-2603.
- Tan, M., & Le, Q. (2019, May). Efficientnet: Rethinking model scaling for convolutional neural networks. In *International Conference on Machine Learning* (pp. 6105-6114). PMLR.
- Tariq, Z., Shah, S. K., & Lee, Y. (2022). Feature-based fusion using CNN for lung and heart sound classification. *Sensors*, 22(4), 1521.
- Türkiye Solunum Araştırmaları Derneği. (t.y.). Solunum hastalıkları önemli bir mortalite nedeni. Medikal Akademi. <https://www.medikalakademi.com.tr/solunum-hastalklar-oenemli-bir-mortalite-nedeni/>
- Tzeng, J. T., Li, J. L., Chen, H. Y., Huang, C. H., Chen, C. H., Fan, C. Y., ... & Lee, C. C. (2025). Improving the Robustness and Clinical Applicability of Automatic Respiratory Sound Classification Using Deep Learning-Based Audio Enhancement: Algorithm Development and Validation. *JMIR AI*, 4(1), e67239.
- Wanasinghe, T., Bandara, S., Madusanka, S., Meedeniya, D., Bandara, M., & Díez, I. D. L. T. (2024). Lung sound classification with multi-feature integration utilizing lightweight CNN model. *IEEE Access*, 12, 21262-21276.
- Wang, Z., & Sun, Z. (2024). Performance evaluation of lung sounds classification using deep learning under variable parameters. *EURASIP Journal on Advances in Signal Processing*, 2024(1), 51.
- Wu, C., Ye, N., & Jiang, J. (2024). Classification and recognition of lung sounds based on improved Bi-ResNet model. *IEEE Access*.
- Xu, M., & Wiese, L. (2023, September). Application and performance improvement of transfer learning on ICBHI lung sound dataset. In *Proceedings of SAI Intelligent Systems Conference* (pp. 156-173). Cham: Springer Nature Switzerland.
- Yu, S., Ding, Y., Qian, K., Hu, B., Li, W., & Schuller, B. W. (2022, May). A glance-and-gaze network for respiratory sound classification. In *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 9007-9011). IEEE.
- Zhang, W., & Liu, X. (2024). DECODING BREATH: Machine learning advancements in diagnosing pulmonary diseases via lung sound analysis. *Scientific and Engineering Letters*, 12(1), 1-11.