

T.C.
GÜMÜŞHANE ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

YAPAY ZEKA VE AKILLI SİSTEMLER ANA BİLİM DALI

SOLUNUM SESİ SINIFLANDIRMASI İÇİN KLASİK VE DERİN ÖĞRENME
MODELLERİNİN KARŞILAŞTIRILMASI

YÜKSEK LİSANS

Eren AYGÜN

ŞUBAT-2026
GÜMÜŞHANE



**T.C.
GÜMÜŞHANE ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ**

YAPAY ZEKA VE AKILLI SİSTEMLER ANA BİLİM DALI

**SOLUNUM SESİ SINIFLANDIRMASI İÇİN KLASİK VE DERİN ÖĞRENME
MODELLERİNİN KARŞILAŞTIRILMASI**

**COMPARISON OF CLASSICAL AND DEEP LEARNING MODELS FOR
RESPIRATORY SOUND CLASSIFICATION**

YÜKSEK LİSANS

Eren AYGÜN

**ŞUBAT-2026
GÜMÜŞHANE**



**T.C.
GÜMÜŞHANE ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ**

YAPAY ZEKA VE AKILLI SİSTEMLER ANA BİLİM DALI

**SOLUNUM SESİ SINIFLANDIRMASI İÇİN KLASİK VE DERİN ÖĞRENME
MODELLERİNİN KARŞILAŞTIRILMASI**

**COMPARISON OF CLASSICAL AND DEEP LEARNING MODELS FOR
RESPIRATORY SOUND CLASSIFICATION**

YÜKSEK LİSANS

Eren AYGÜN

Danışman: Dr. Öğr. Üyesi Hayati TÜRE

**ŞUBAT-2026
GÜMÜŞHANE**

BİLİMSEL ETİĞE UYGUNLUK BEYANI

Yüksek Lisans Tezi olarak hazırlamış olduğum “**Solunum Sesi Sınıflandırması İçin Klasik ve Derin Öğrenme Modellerinin Karşılaştırılması**” isimli bu tezimin, tamamen kendi çalışmam olduğunu, her alıntıya kaynak gösterdiğimi, alıntı yaptığım tüm çalışmaları kaynakçada belirttiğimi ve Gümüşhane Üniversitesi’nin lisanslı kullanıcısı olduğu intihal yazılım programı ile Lisansüstü Eğitim Enstitüsü’nün belirlediği kısıtlara uygun olarak raporladığımı taahhüt ederim. Tezimin kâğıt ve elektronik kopyalarının Gümüşhane Üniversitesi Lisansüstü Eğitim Enstitüsü arşivinde saklanmasına izin verdiğimi onaylarım.

Lisansüstü Eğitim ve Öğretim Yönetmeliği’nin ilgili maddeleri uyarınca gereğinin yapılmasını arz ederim.

04/02/2026

.....
Eren AYGÜN

TEŞEKKÜR

Bu çalışma süresince bilgi ve deneyimlerini benimle paylaşan, her aşamada yol gösteren ve desteğini esirgemeyen değerli danışmanım Dr. Öğr. Üyesi Hayati TÜRE'ye en içten teşekkürlerimi sunarım.

Tez çalışmam boyunca katkı sağlayan bölüm hocalarıma ve araştırma sürecinde yardımlarını gördüğüm tüm akademik ve idari personele teşekkür ederim.

Ayrıca eğitim hayatım boyunca maddi ve manevi desteklerini her zaman yanımda hissettiğim sevgili aileme ve sabırları için arkadaşlarıma şükranlarımı sunarım.

Eren AYGÜN
GÜMÜŞHANE – 2026

ÖZET

Bu çalışmada, solunum sesi sinyallerinin otomatik sınıflandırılması amacıyla ICBHI 2017 veri kümesi üzerinde klasik makine öğrenmesi ve modern derin öğrenme yaklaşımları karşılaştırılmıştır. Ön işleme adımlarının ardından Mel-spektrogram ve tayfsal öznitelikler çıkarılmış; bu öznitelikler klasik yöntemler (SVM, KNN, Random Forest, XGBoost) ile, Mel-spektrogramlar ise evrişim tabanlı mimariler (GhostNet ailesi, ResNet-50, EfficientNet-B0, MobileNetV3) ile değerlendirilmiştir. Değerlendirmede hasta-bağımsız (subject-wise) veri bölmesi kullanılmış; sınıf dengeleme amaçlı veri artırma yalnızca eğitim katlarına uygulanarak doğrulama/sınama sızıntısı engellenmiştir. Bulgular, GhostNetV4 mimarisinin %89.9 doğruluk ve makro-F1 = 0.898 ile diğer derin ve klasik yaklaşımları aştığını göstermektedir. Sonuçlar, hesaplama yükü görece düşük derin öğrenme modellerinin, solunum sesi analizi için klinik karar destek sistemlerine entegrasyonda güçlü adaylar olduğunu ortaya koymaktadır. Ayrıca çalışma, hasta-bağımsız protokol ve şeffaf veri artırma politikasıyla literatürdeki değerlendirme farklılıklarına karşı kıyaslanabilirlik sağlamaktadır.

Anahtar Kelimeler: Derin öğrenme, GhostNet, Makine öğrenmesi, Mel-spektrogram, Solunum sesi analizi

SUMMARY

This study compares classical machine learning and modern deep learning approaches for the automatic classification of respiratory sound signals on the ICBHI 2017 dataset. After preprocessing, Mel-spectrograms and spectral features were extracted. Classical algorithms (SVM, KNN, Random Forest, XGBoost) were trained on handcrafted features, while convolutional architectures (the GhostNet family, ResNet-50, EfficientNet-B0, MobileNetV3) were trained on Mel-spectrograms. A patient-independent (subject-wise) split was adopted, and class-balancing data augmentation was applied only to the training folds to avoid validation/test leakage. The results show that GhostNetV4 achieved the best performance with 89.9% accuracy and macro-F1 = 0.898, outperforming both deep and classical baselines. These findings indicate that computationally efficient deep learning models are strong candidates for integration into clinical decision support systems for respiratory sound analysis. The study also contributes a transparent evaluation protocol that improves comparability across works by coupling a patient-independent split with a clearly defined augmentation policy.

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

İÇİNDEKİLER

KABUL VE ONAY	III
BİLİMSEL ETİĞE UYGUNLUK BEYANI.....	IV
TEŞEKKÜR.....	V
ÖZET.....	VI
SUMMARY	VII
İÇİNDEKİLER	VIII
TABLolar DİZİNİ	XIII
ŞEKİLLER DİZİNİ.....	XIV
SİMGELER VE KISALTMALAR DİZİNİ	XVI
1. GİRİŞ	1
1.1. Solunum Sesi Analizinin Önemi	1
1.2. Solunum Hastalıklarında Erken Teşhis ve Dijital Yaklaşımlar	2
1.3. Veri Tabanlı Yaklaşımların Avantajları.....	3
1.4. Bu Çalışmanın Amacı ve Kapsamı	4
2. TEMEL KAVRAMLAR.....	7
2.1. Solunum Sistemi ve Fizyolojisi	7
2.2. Sesin Fiziksel Özellikleri	9
2.3. Solunum Seslerinin Tarihsel Gelişimi	12
2.4. Solunum Seslerinin Oluşumu.....	15
2.5. Normal Solunum Sesleri	18
2.5.1. Veziküler Ses	18
2.5.2. Bronşiyal Ses.....	19
2.5.3. Bronkoveziküler Ses	20
2.6. Patolojik Solunum Sesleri	21
2.6.1. Raller (Crackles)	21
2.6.2. Hışıltı (Wheeze)	22
2.6.3. Ronküs (Rhonchi)	22
2.6.4. Stridor.....	22
2.7. Solunum Seslerinin Frekans Özellikleri.....	23
2.8. Solunum Seslerinin Klinik Önemi	24
2.9. Veri Ön İşleme Yöntemleri.....	25
2.9.1. Normalizasyon	25

2.9.2. Gürültü Giderme ve Filtreleme	25
2.9.2.1. Band-Pass Filtre	26
2.9.2.2. Wiener Filtre	26
2.9.2.3. Empirical Mode Decomposition (EMD).....	26
2.9.3. Segmentasyon Teknikleri.....	26
2.9.4. Artefaktların Temizlenmesi	27
2.10. Veri Artırma (Data Augmentation) Teknikleri	27
2.10.1. Zaman Kaydırma.....	28
2.10.2. Gürültü Enjeksiyonu	28
2.10.3. Ton Değiştirme (Pitch Shifting).....	28
2.10.4. Zaman Ölçekleme (Time Stretching).....	29
2.10.5. SMOTE ve Yapay Örnek Üretimi.....	29
2.10.6. Zaman ve Frekans Maskelenmesi	29
2.10.7. Ses Seviyesi Ayarı ve Kazanç Değişimi	29
2.10.8. Ortam Gürültüsü ve Yankı Eklenmesi	30
2.10.9. Jeneratif Yöntemler (GAN / Diffusion Models)	30
2.11. Öznitelik Çıkarımı Yöntemleri	31
2.11.1. Mel Frekans Kepstrum Katsayıları (MFCC).....	31
2.11.2. Mel-spektrogram	33
2.11.3. Kısa Zamanlı Fourier Dönüşümü (STFT).....	33
2.11.4. Dalgacık Dönüşümleri (Wavelet)	33
2.11.5. İstatistiksel Özellikler.....	33
2.11.6. Chroma Özellikleri.....	33
2.11.7. Spektral Kontrast.....	34
2.12. Klasik Makine Öğrenmesi Yaklaşımları	35
2.12.1. Destek Vektör Makineleri (SVM).....	35
2.12.2. k-En Yakın Komşu (KNN)	35
2.12.3. Rastgele Ormanlar (RF)	36
2.12.4. XGBoost.....	36
2.12.5. Naive Bayes	36
2.12.6. Doğrusal Ayırıcı Analiz (LDA)	37
2.12.7. AdaBoost.....	37
2.12.8. Karar Ağaçları (Decision Tree).....	37
2.13. Derin Öğrenme Yaklaşımları	38
2.13.1. Uzamsal Tabanlı Modeller (CNN ve Türevleri)	39

2.13.1.1. GhostNet	39
2.13.1.2. ResNet	39
2.13.1.3. EfficientNet	40
2.13.1.4. MobileNet	40
2.13.1.5. DenseNet	41
2.13.1.6. InceptionNet	41
2.13.1.7. SqueezeNet	41
2.13.1.8. ShuffleNet	41
2.13.1.9. CBAM / SE Attention Bloklarıyla Donatılmış CNN'ler	41
2.13.2. Zamana Duyarlı Modeller	42
2.13.2.1. Tekrarlayan Sinir Ağları (RNN) ve LSTM	42
2.13.2.2. Dikkat Mekanizmaları (Attention Mechanisms)	42
2.13.2.3. Tipik Kombinasyonlar ve Uygulamalar	42
2.13.3. Transformer Tabanlı Yaklaşımlar	43
2.13.3.1. Vision Transformer (ViT)	43
2.13.3.2. Audio Spectrogram Transformer (AST)	43
2.13.3.3. Self-Supervised AST (SSAST)	43
2.13.3.4. Conformer	44
2.13.3.5. Swin Transformer	44
2.13.3.6. Perceiver / Perceiver IO	44
2.14. Solunum Sesi Veri Tabanları	45
3. LİTERATÜR TARAMASI	47
3.1. Giriş	47
3.2. Klasik Makine Öğrenmesi Yöntemlerine Dayalı Çalışmalar	48
3.2.1. MFCC Tabanlı Çalışmalar	49
3.2.2. Destek Vektör Makinelerine (SVM) Dayalı Çalışmalar	49
3.2.3. K-En Yakın Komşu (KNN) ve Bayes Sınıflandırıcıları	50
3.2.4. Dalgacık, Zaman-Frekans Dönüşümleri ve Hibrit Yaklaşımlar	50
3.2.5. Klasik Yöntemlerin Genel Değerlendirmesi	50
3.3. Derin Öğrenmeye Dayalı Çalışmalar	51
3.3.1. CNN Tabanlı Yaklaşımlar	52
3.3.2. Transfer Öğrenme ve Hibrit Modeller	53
3.3.3. RNN/LSTM ve Dikkat Mekanizmaları	53
3.3.4. Dönüşümcü (Transformer) Tabanlı ve Kokleogram Yaklaşımları	54

3.3.5. Alan Uyumlama (Domain Adaptation), Federated Learning ve Öz-Denetimli (Self-Supervised) Öğrenme.....	54
3.3.6. Derin Öğrenme Yaklaşımlarının Genel Değerlendirmesi.....	55
3.4. Veri Setleri ve Değerlendirme Protokolleri	56
3.5. Performans Ölçütleri	58
3.6. Literatürün Ortak Noktaları.....	59
3.7. Literatürdeki Boşluklar	61
3.8. Genel Değerlendirme	63
4. MATERYAL VE YÖNTEM	65
4.1. Veri Seti	65
4.2. Segmentasyon ve Etiketleme	67
4.3. Ön İşleme	68
4.3.1. Örnekleme Hızı Birleştirme ve DC Offset Kaldırma.....	69
4.3.2. Bant Geçiren Filtreleme	69
4.3.3. Gürültü Azaltma.....	69
4.3.4. Normalizasyon	70
4.3.5. Görselleştirme ve Doğrulama	70
4.4. Özellik Çıkarma	72
4.4.1. Mel Frekans Kepstral Katsayıları (MFCC).....	72
4.4.2. Mel-Spektrogram	74
4.4.3. Delta ve Delta-Delta Katsayıları	75
4.4.4. Özelliklerin Kaydedilmesi ve Eğitim İçin Hazırlık	77
4.4.5. Özelliklerin Görselleştirilmesi	78
4.4.6. Veri Dengesizliği ve Veri Artırma (Augmentation) Uygulaması	78
4.5. Modelleme ve Sınıflandırma Yaklaşımları	82
4.5.1. Eğitim ve Test Ayrımı.....	82
4.5.2. Klasik Makine Öğrenmesi Yöntemleri	82
4.5.3. Derin Öğrenme Yöntemleri.....	84
4.5.3.1. Evrişimli Sinir Ağları (CNN).....	85
4.5.3.2. Tekrarlayan Sinir Ağları (RNN – LSTM).....	86
4.5.3.3. Karma Mimari (CRNN)	86
4.5.4. Model Eğitim Süreci	86
4.5.5. Değerlendirme Metrikleri.....	87
5. BULGULAR VE TARTIŞMA	89
5.1. Deneysel Sonuçların Genel Özeti	89

5.2. Derin Öğrenme Modellerinin Performansı	90
5.2.1. Önerilen GhostNetV4 Model Mimarisi	92
5.3. Klasik Makine Öğrenmesi Modellerinin Performansı	94
5.4. Veri Artırma (Data Augmentation) Etkisinin Analizi.....	96
5.5. Karışıklık Matrisi Analizi	97
5.6. ROC Eğrisi, AUC ve Eğitim Dinamiklerinin Analizi.....	102
5.7. Genel Değerlendirme ve Literatürle Karşılaştırma	107
6. SONUÇ, BULGULARIN ÖZETİ VE GENEL KATKILAR.....	109
7. GELECEK ÇALIŞMALAR.....	111
7.1. Harici Doğrulama ve Gerçek Dünya Verisiyle Genelleme.....	111
7.2. Uçtan Uca Gerçek Zamanlı Entegrasyon ve Model Sıkıştırma	111
7.3. Sınıflandırma Kapsamının Genişletilmesi ve Klinik Senaryolara Uyum	112
7.4. Açıklanabilir Yapay Zekâ ve Klinik Geri Bildirim	112
7.5. Etik, Adalet ve Regülasyon Uyumluluğu	112
7.6. Dayanıklılık Öğrenme Yöntemleri ve Veri Eksikliğiyle Baş Etme	113
7.7. Klinik Akışlara Uyum, Sürdürülebilirlik ve Model Güncellenebilirliği.....	113
KAYNAKÇA.....	114
ÖZGEÇMİŞ	124

TABLÖLAR DİZİNİ

Tablo 1. Dünya genelinde başlıca solunum sistemi hastalıkları, görülme sıklıkları, ölüm oranları ve karakteristik solunum sesleri (GBD, 2019; IARC, 2022; WHO, 2023; WHO, 2024).....	2
Tablo 2. Solunum sesi veri artırma yöntemlerinin karşılaştırılması	31
Tablo 3. Yaygın öznelilik çıkarım yöntemlerinin karşılaştırılması	34
Tablo 4. Klasik makine öğrenmesi algoritmalarının güçlü ve zayıf yönleri.....	38
Tablo 5. Derin öğrenme modellerinin avantajları ve dezavantajları.....	44
Tablo 6. Yaygın Kullanılan Solunum Sesi Veri Setlerinin Karşılaştırması.....	46
Tablo 7. ICBHI 2017 veri kümesinde dosya adı adlandırma şeması ve örnek anotasyon kayıtları	66
Tablo 8. Sınıf etiketleri, açıklamaları ve segment sayıları.....	68
Tablo 9. Ön işleme hattı ve temel parametreler	71
Tablo 10. Veri artırma öncesi ve sonrası sınıf dağılımları.....	79
Tablo 11. Solunum sesi özellik çıkarma; adımlar, ayarlar ve tasarım gerekçesi	81
Tablo 12. Klasik makine öğrenmesi modelleri: girdi temsili, başlıca hiperparametreler ve seçime dair kısa gerekçe	84
Tablo 13. Mel-spektrogram girişi için CNN omurgalarının mimari özeti ve pratik kurulum	85
Tablo 14. Derin öğrenme modellerinin sınıflandırma performanslarının karşılaştırılması	91
Tablo 15. GhostNetV4 katman özeti ve parametre bilgisi	94
Tablo 16. Klasik makine öğrenmesi modellerinin sınıflandırma performansları	95
Tablo 17. Veri artırma sonrasında her sınıfa ait F1-skorundaki değişim.....	96
Tablo 18. GhostNetV4 modeline ait sınıf bazlı performans metrikleri	98
Tablo 19. Farklı modeller için sınıf bazlı AUC (ROC) değerleri	103

ŞEKİLLER DİZİNİ

Şekil 1. Tez çalışmasında uygulanan metodolojik akış diyagramı	6
Şekil 2. İnsan solunum sisteminin temel anatomik yapısı (T.C. Milli Eğitim Bakanlığı – OGM, t.y.).....	9
Şekil 3. Ses dalgasının temel bileşenleri genlik ve dalga boyu (Yokuş, 2011).	11
Şekil 4. René Laennec tarafından 1816 yılında tasarlanan ilk stetoskop (Science Museum Group Studio, 2019).....	12
Şekil 5. Erken dönem binaural stetoskop (Science Museum Group, 2019).....	13
Şekil 6. Elektronik Stetoskop (Webtekno, 2021).....	14
Şekil 7. Anterior solunum oskültasyon noktaları (Pomietlo, 2020).....	17
Şekil 8. Posterior solunum oskültasyon noktaları (Pomietlo, 2020).....	18
Şekil 9. Solunum sesi ön işleme ve veri artırma sürecinin genel akış diyagramı	30
Şekil 10. MFCC çıkarım sürecinin şematik gösterimi	32
Şekil 11. Ses kaydının segmentlere ayrılması.....	67
Şekil 12. Bir segmentin orijinal ve ön işlenmiş (filtrelenmiş) dalga formu karşılaştırması	72
Şekil 13. Bir segmentin orijinal ve ön işlenmiş (filtrelenmiş) spektrogram karşılaştırması	72
Şekil 14. Dört farklı sınıfa ait MFCC ısı haritaları	73
Şekil 15. Dört farklı sınıfa ait örnek Mel-spektrogram görselleri.....	74
Şekil 16. Bir solunum segmentine ait Mel-spektrogram, MFCC, Delta ve Delta-Delta katsayılarının karşılaştırmalı görselleştirilmesi	76
Şekil 17. Çalışmada kullanılan mel_specs.h5 dosyasının içeriğine ait yapısal analiz	77
Şekil 18. Veri artırma öncesi ve sonrası sınıf dağılımı	79
Şekil 19. Veri artırma sonrası orijinal ve üretilmiş ses segmentlerinin dalga formu karşılaştırması	80
Şekil 20. Zaman kaydırma ve gürültü ekleme teknikleri kullanılarak gerçekleştirilen veri artırma sonrasında, orijinal ve üretilmiş ses segmentlerinin spektrogram karşılaştırması	80
Şekil 21. Rastgele kazanç ve zaman esnetme teknikleri kullanılarak gerçekleştirilen veri artırma sonrasında, orijinal ve üretilmiş ses segmentlerinin spektrogram karşılaştırması	81

Şekil 22. XGBoost modelinin dört sınıf (Normal, Wheeze, Crackle, Both) için oluşturulan karışıklık matrisi	98
Şekil 23. KNN modelinin dört sınıf (Normal, Wheeze, Crackle, Both) için oluşturulan karışıklık matrisi	99
Şekil 24. Random Forest modelinin dört sınıf (Normal, Wheeze, Crackle, Both) için oluşturulan karışıklık matrisi	99
Şekil 25. SVM modelinin dört sınıf (Normal, Wheeze, Crackle, Both) için oluşturulan karışıklık matrisi	100
Şekil 26. GhostNetV4 modelinin dört sınıf (Normal, Wheeze, Crackle, Both) için oluşturulan karışıklık matrisi	100
Şekil 27. GhostNetV2 modelinin dört sınıf (Normal, Wheeze, Crackle, Both) için oluşturulan karışıklık matrisi	101
Şekil 28. EfficientNet – B0 modelinin dört sınıf (Normal, Wheeze, Crackle, Both) için oluşturulan karışıklık matrisi	101
Şekil 29. CNN modelinin dört sınıf (Normal, Wheeze, Crackle, Both) için oluşturulan karışıklık matrisi	102
Şekil 30. GhostNetV4 modelinin dört sınıf (Normal, Wheeze, Crackle, Both) için ROC eğrileri ve sınıf bazlı ortalama AUC değerleri.....	104
Şekil 31. GhostNetV4 modelinin eğitim süreci boyunca kayıp (Loss), doğruluk (Accuracy), F1 skoru ve birleşik performans (Combined Performance) metriklerinin epoch bazlı değişimi.....	105
Şekil 32. GhostNetV3 modelinin dört sınıf (Normal, Wheeze, Crackle, Both) için ROC eğrileri ve sınıf bazlı ortalama AUC değerleri.....	105
Şekil 33. GhostNetV3 modelinin eğitim süreci boyunca kayıp (Loss), doğruluk (Accuracy), F1 skoru ve birleşik performans (Combined Performance) metriklerinin epoch bazlı değişimi.....	106
Şekil 34. XGBoost modelinin dört sınıf (Normal, Wheeze, Crackle, Both) için ROC eğrileri ve sınıf bazlı ortalama AUC değerleri.....	106

SİMGELER VE KISALTMALAR DİZİNİ

ACC	: Accuracy (Doğruluk)
ADASYN	: Adaptive Synthetic Sampling
AGM	: Adaptive Ghost Module
AMP	: Automatic Mixed Precision (Otomatik Karışık Hassasiyet)
AUC	: Area Under the Curve (Eğri Altı Alan)
BN	: Batch Normalization (Yığın Normalizasyonu)
CMVN	: Cepstral Mean and Variance Normalization (Kepstrum Ortalama–Varyans Norm.)
CNN	: Convolutional Neural Network (Evrişimsel Sinir Ağı)
CRNN	: Convolutional Recurrent Neural Network (Evrişimsel Tekrarlayan Sinir Ağı)
CWT	: Continuous Wavelet Transform (Sürekli Dalgacık Dönüşümü)
DCT	: Discrete Cosine Transform (Ayrık Kosinüs Dönüşümü)
DWT	: Discrete Wavelet Transform (Ayrık Dalgacık Dönüşümü)
EMD	: Empirical Mode Decomposition (Ampirik Kip Ayırıştırma)
ERS	: European Respiratory Society (Avrupa Solunum Derneği)
F	: Frekans
F1	: F1 Skoru (Harmonik Ortalama)
FFT	: Fast Fourier Transform (Hızlı Fourier Dönüşümü)
FN	: False Negative (Yanlış Negatif)
FP	: False Positive (Yanlış Pozitif)
GBD	: Global Burden of Disease
GAN	: Generative Adversarial Networks
GAP	: Global Average Pooling (Küresel Ortalama Havuzlama)
GMM	: Gaussian Mixture Model (Gauss Karışım Modeli)
HDF5	: Hierarchical Data Format v5
IARC	: International Agency for Research on Cancer
ICBHI	: International Conference on Biomedical Health Informatics
KNN	: K-Nearest Neighbors (K-En Yakın Komşu)
KOAH	: Kronik Obstrüktif Akciğer Hastalığı
L2	: L2 ağırlık çürümesi / weight decay
LR	: Learning Rate (Öğrenme Oranı)
LSTM	: Long Short Term Memory (Uzun Kısa Süreli Bellek)

MCC	: Matthews Correlation Coefficient (Matthews Korelasyon Katsayısı)
MFCC	: Mel Frekans Kepstrum Katsayıları
MSFA	: Multi-Scale Feature Aggregation (Çok Ölçekli Özellik Birleştirme)
NPV	: Negative Predictive Value (Negatif Öngörü Değeri)
OST	: Optimized/Modified S-Transform
PPV	: Positive Predictive Value / Kesinlik
PR	: Precision–Recall (Kesinlik–Duyarlılık)
ReLU	: Rectified Linear Unit (Doğrultulmuş Doğrusal Birim)
RF	: Random Forest (Rastgele Orman)
RNN	: Recurrent Neural Network (Tekrarlayan Sinir Ağı)
ROC	: Receiver Operating Characteristic (Alıcı İşletim Özelliği Eğrisi)
s	: Saniye
SA	: Split-Attention (Bölünmüş Dikkat)
Se/Sp	: Sensitivity (Duyarlılık) / Specificity (Özgüllük)
SNR	: Signal-to-Noise Ratio (Sinyal–Gürültü Oranı)
SMOTE	: Synthetic Minority Over-sampling Technique
STFT	: Short-Time Fourier Transform (Kısa-Zamanlı Fourier Dönüşümü)
SVM	: Support Vector Machines (Destek Vektör Makineleri)
t	: Zaman
TN	: True Negative (Doğru Negatif)
TP	: True Positive (Doğru Pozitif)
t.y.	: Tarih yok
URL	: İnternet Adresi (Uniform Resource Locator)
vd.	: Ve diğerleri
WHO	: World Health Organization (Dünya Sağlık Örgütü)
XGBoost	: Extreme Gradient Boosting
ZCR	: Zero-Crossing Rate (Sıfır Geçiş Oranı)

1. GİRİŞ

1.1. Solunum Sesi Analizinin Önemi

Solunum sistemi; oksijenin akciğerlere alınması, dokularda kullanılması ve oluşan karbondioksitin vücuttan uzaklaştırılması gibi yaşamsal görevleri yerine getirir. Havanın solunum yollarından geçerken oluşturduğu akış hareketi, doku titreşimleri ve basınç değişimleri ise duyulabilir solunum seslerini ortaya çıkarır. Bu sesler, akciğer ve hava yollarının durumu hakkında önemli bilgiler taşıdığından, uzun yıllardır tıbbi değerlendirmede kullanılmaktadır.

Stetoskopun 19. yüzyılda geliştirilmesiyle birlikte solunum seslerinin dinlenmesi, yani oskültasyon, klinik muayenenin temel parçası hâline gelmiştir (Zimmerman, 2023). Normal solunum sesleri veziküler, bronşiyal ve bronkoveziküler seslerden oluşur. Hastalıklara bağlı durumlarda ise raller (crackles), hışıltı (wheeze), ronküs (rhonchi) ve stridor gibi ek sesler duyulabilir. Örneğin ince raller genellikle zatürre ya da akciğer dokusunu etkileyen hastalıklarda görülürken; hışıltı genellikle Kronik Obstrüktif Akciğer Hastalığı (KOAH) ve astım gibi hava yolu daralmasıyla seyreden durumlarda ortaya çıkar.

Günümüzde solunum sistemi hastalıkları, dünya genelinde hem ölüm nedenleri arasında üst sıralarda yer almakta hem de toplum sağlığını önemli ölçüde etkilemektedir. Dünya Sağlık Örgütü (WHO), Global Burden of Disease (GBD) ve Uluslararası Kanser Araştırma Ajansı (IARC) raporları; KOAH, astım, tüberküloz, alt solunum yolu enfeksiyonları ve akciğer kanserinin kapsamlı bir sağlık sorunu olmaya devam ettiğini göstermektedir (GBD, 2019; IARC, 2022; WHO, 2023; WHO, 2024).

Bu hastalıklarla ilgili güncel epidemiyolojik veriler ve onlara eşlik eden karakteristik solunum sesleri Tablo 1’de sunulmaktadır. Tablo, hem küresel düzeyde hastalık yükünü hem de her bir solunum hastalığına özgü ses örüntülerini özetleyerek klinik değerlendirme ve otomatik analiz yöntemleri için temel bir referans sağlamaktadır. Veriler, solunum hastalıklarının yüksek prevalans ve mortaliteyle seyreden önemli bir halk sağlığı sorunu olduğunu açık biçimde ortaya koymaktadır.

Bu çerçevede solunum seslerinin doğru, nesnel ve tekrarlanabilir biçimde değerlendirilmesi kritik bir gerekliliktir. Geleneksel oskültasyon her ne kadar hızlı ve düşük maliyetli bir yöntem olsa da dinleyiciye bağımlı olması ve yorumlayıcı farklılıklara açık olması nedeniyle tanılabilirliği sınırlayabilir. Bu nedenle, solunum seslerinin

daha objektif, ölçülebilir ve standardize edilmiş yaklaşımlarla analiz edilmesine yönelik modern yöntemlerin önemi giderek artmaktadır.

Tablo 1. Dünya genelinde başlıca solunum sistemi hastalıkları, görülme sıklıkları, ölüm oranları ve karakteristik solunum sesleri (GBD, 2019; IARC, 2022; WHO, 2023; WHO, 2024).

Hastalık	Küresel Prevalans (yaklaşık)	Yıllık Ölüm (yaklaşık)	Karakteristik Solunum Sesleri
KOAH	≈ 384 milyon kişi	≈ 3.2 milyon	Wheeze (hışıltı), ronküs, azalmış solunum sesleri
Astım	≈ 262 milyon kişi	≈ 0.46 milyon	İnce ısığa benzer wheeze
Alt solunum yolu enfeksiyonları	Her yıl yüz milyonlarca vaka	≈ 2.5 milyon	Ral (krepitan), ronküs, hırıltı
Tüberküloz	≈ 10.6 milyon yeni vaka	≈ 1.3 milyon	Bronşiyal solunum, ral, yer yer ses azalması
Akciğer kanseri	≈ 2.5 milyon	≈ 1.8 milyon	Stridor, tek taraflı wheeze
COVID-19	> 700 milyon vaka	≈ 7 milyon	Yaygın ral, wheeze, azalmış solunum sesi

1.2. Solunum Hastalıklarında Erken Teşhis ve Dijital Yaklaşımlar

Solunum sesi değerlendirmesinin en temel yöntemi olan oskültasyon, düşük maliyetli, hızlı uygulanabilir ve hastaya herhangi bir girişim yapılmadan gerçekleştirilebilen bir klinik inceleme tekniğidir. Bu özellikleri nedeniyle uzun yıllardır yaygın bir şekilde kullanılmaktadır (Pasterkamp vd., 1997). Ancak oskültasyonun etkinliği büyük ölçüde hekim deneyimine bağlıdır. Dinleyenin işitsel hassasiyeti, klinik tecrübesi ve çevresel gürültü koşulları, solunum seslerinin doğru yorumlanmasını doğrudan etkileyebilir (Zimmerman, 2023). Bu nedenle, özellikle sınırda ya da karma solunum seslerinin bulunduğu durumlarda, subjektif değerlendirmeye dayalı geleneksel yöntemlerin tanısal doğruluğu sınırlı kalabilmektedir.

Dijital tıp alanındaki gelişmeler, solunum seslerinin değerlendirilmesinde daha objektif yaklaşımların kullanılmasını mümkün hâle getirmiştir. Dijital stetoskoplar; daha geniş frekans aralıklarında, yüksek örnekleme hızlarında ve düşük gürültü seviyelerinde kayıt alınmasına olanak tanımakta, böylece akciğer seslerinin bilgisayar ortamına

aktarılmasını sağlamaktadır (Huang vd., 2023). Kaydedilen sinyaller üzerinde gürültü azaltma, filtreleme, normalizasyon ve zaman–frekans dönüşümleri gibi ön işleme adımları uygulanarak, solunum seslerinin daha anlamlı biçimlerde temsil edilmesi mümkün olmaktadır (Rocha vd., 2019). Bu sayede düşük genlikli raller gibi insan kulağının zor algılayabildiği ses bileşenleri, dijital yöntemlerle daha güvenilir şekilde tespit edilebilmektedir.

Dijital yaklaşımlar yalnızca kayıt kalitesini artırmakla kalmayıp uzaktan sağlık hizmetlerinin gelişmesine de önemli katkı sağlamaktadır. Özellikle teletıp uygulamalarında, hastaların ev ortamında kaydettikleri solunum seslerinin bulut tabanlı analiz sistemlerine iletilmesi ve bu sistemler tarafından değerlendirilmesi, erken tanı ve izlem süreçlerini güçlendirmektedir (Huang vd., 2023). Bu dijital izlem altyapısı, KOAH alevlenmeleri, astım kontrol düzeyindeki bozulmalar ve pnömoni gibi klinik açıdan kritik durumların erken aşamada saptanmasına olanak tanıyarak sağlık hizmetlerinin etkinliğini anlamlı ölçüde artırmaktadır (WHO, 2024).

Sonuç olarak dijital teknolojiler, solunum sesi analizini subjektif bir değerlendirme alanı olmaktan çıkarıp daha nesnel, ölçülebilir ve tekrarlanabilir bir yapıya dönüştürmektedir. Bu gelişme, hem klinisyenlerin karar süreçlerine katkı sağlamakta hem de akıllı sağlık sistemlerinin yaygınlaşması ile birlikte solunum hastalıklarının erken tanı ve izleminde önemli bir potansiyel sunmaktadır.

1.3. Veri Tabanlı Yaklaşımların Avantajları

Solunum seslerinin değerlendirilmesinde veri tabanlı yaklaşımlar, geleneksel oskültasyon yöntemine kıyasla daha nesnel, ölçülebilir ve tekrarlanabilir sonuçlar sunması bakımından önemli bir avantaj sağlamaktadır. Dijital olarak kaydedilen solunum sesleri, bilgisayar ortamında işlenerek hem zaman hem de frekans alanında ayrıntılı biçimde analiz edilebilmekte; böylece insan kulağının algılamakta zorlanabileceği düşük genlikli bileşenler daha net bir şekilde ortaya çıkarılabilmektedir (Rocha vd., 2019).

Veri tabanlı analiz sürecinin temel adımlarından biri, ses sinyalinden anlamlı özellikler çıkarmaktır. Bu kapsamda Mel Frekans Kepstrum Katsayıları (MFCC), Kısa Zamanlı Fourier Dönüşümü (STFT), Mel-spektrogram temsilleri ve dalgacık tabanlı yöntemler uzun yıllardır solunum sesi araştırmalarında yaygın olarak kullanılmaktadır (Zimmerman, 2023). Bu yöntemler, ses sinyalinin frekans yapısını, zaman içindeki değişimini ve solunum döngüsü boyunca ortaya çıkan karakteristik örüntüleri sayısal olarak ifade etmeyi mümkün kılar.

Son yıllarda derin öğrenme yöntemlerinin gelişmesiyle birlikte solunum seslerinin otomatik sınıflandırılmasında Evrişimsel Sinir Ağları (CNN) ve dönüştürücü (transformer) tabanlı modeller öne çıkmaya başlamıştır. CNN tabanlı yaklaşımlar, zaman-frekans temsillerinden otomatik olarak özellik öğrenebilmekte ve manuel özellik çıkarımı ihtiyacını büyük ölçüde azaltmaktadır (Huang vd., 2023). Bu durum, özellikle büyük veri setlerinde modelin başarısını artırmakta ve daha tutarlı sonuçlar elde edilmesini sağlamaktadır. Ayrıca son çalışmalarda hafif yapıda tasarlanan CNN mimarilerinin mobil cihazlarda dahi yüksek doğruluk sağlayabildiği gösterilmektedir (Ethala, 2025).

Veri tabanlı yaklaşımların önemli bir zorluğu, solunum sesi kayıtlarının çeşitliliğinden kaynaklanmaktadır. Farklı klinik ortamlar, farklı kayıt cihazları, hastalar arası değişkenlik ve çevresel gürültü; modellerin genelleme başarımını etkileyen önemli faktörlerdir (Rocha vd., 2019). Buna ek olarak solunum sesi veri setlerinde sınıf dengesizliği sık görülen bir durumdur. Örneğin "normal" solunum sesi kayıtları çoğu zaman bol miktarda bulunurken, hışıltı veya ral içeren kayıtlar daha az sayıdadır. Bu nedenle veri artırma, sınıf ağırlıklandırma ve hasta-bağımsız eğitim-test bölme yaklaşımları modellerin daha adil değerlendirilmesi için yaygın olarak tercih edilmektedir (Ethala, 2025; Rocha vd., 2019).

Veri tabanlı yöntemlerin en önemli avantajlarından biri, farklı çalışmalardan elde edilen sonuçların karşılaştırılabilirliğini artırmasıdır. Açık erişimli veri setlerinin kullanılması ve standart değerlendirme protokollerinin uygulanması, solunum sesi sınıflandırma araştırmalarında ortak bir değerlendirme zemini oluşturmaktadır. Bu durum hem yeni yöntemlerin evrensel ölçekte test edilebilmesini sağlamakta hem de literatürdeki ilerlemelerin daha hızlı ve güvenilir olmasına katkı sunmaktadır (GBD, 2019).

Sonuç olarak veri tabanlı yaklaşımlar, solunum sesi analizini daha nesnel, tekrarlanabilir ve hesaplamaya dayalı bir yapıya dönüştürerek hem klinik uygulamalarda hem de yapay zekâ destekli tanı sistemlerinde önemli bir temel oluşturmaktadır.

1.4. Bu Çalışmanın Amacı ve Kapsamı

Bu tez çalışmasının temel amacı, solunum seslerinin otomatik olarak sınıflandırılmasında kullanılan yöntemlerin performansını karşılaştırmak ve bu yöntemlerin klinik açıdan uygulanabilirliğini değerlendirmektir. Bu kapsamda tez çalışması, hem klasik makine öğrenmesi modellerini hem de güncel derin öğrenme mimarilerini aynı veri seti üzerinde sistemli bir biçimde analiz etmektedir. Böylece, farklı

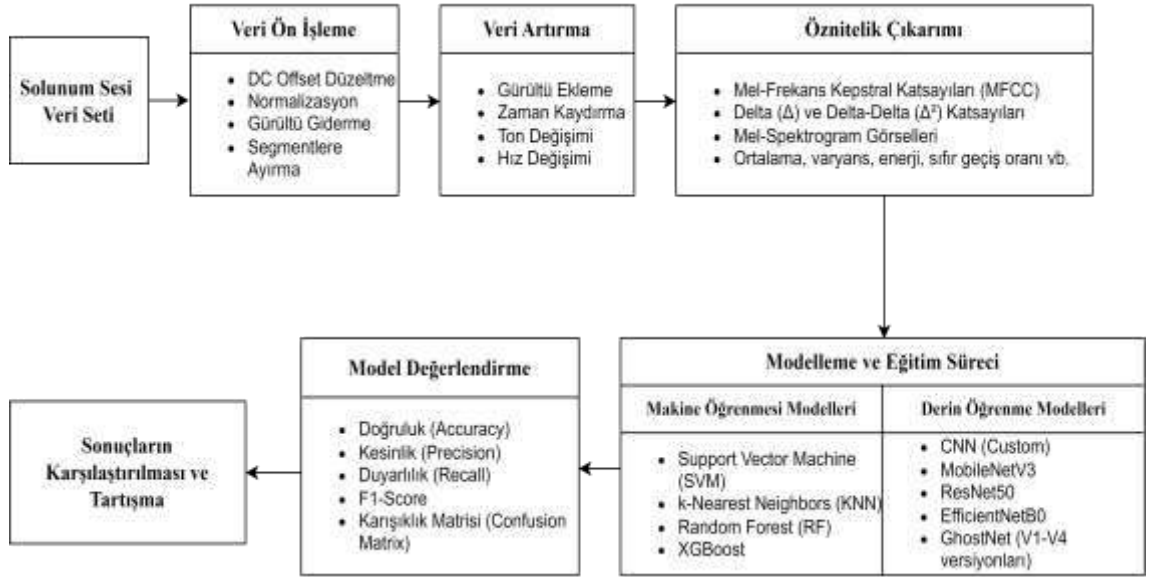
model türlerinin güçlü ve zayıf yönleri karşılaştırılabilmekte ve solunum sesi analizi için en uygun yaklaşım belirlenebilmektedir.

Tezde kullanılan veri seti, uluslararası bir değerlendirme platformu olarak kabul edilen ICBHI 2017 Respiratory Sound Database olup, 126 bireye ait 920 ses kaydı ve 6.898 solunum döngüsü içermektedir (ICBHI Challenge, 2017). Farklı cihazlar ve farklı klinik ortamlarda kaydedilmiş olması, veri setine yüksek çeşitlilik kazandırmakta ve geliştirilen modellerin gerçek dünya koşullarında test edilmesine olanak sağlamaktadır (Rocha vd., 2019).

Bu çalışmada solunum seslerinden Mel Frekans Kepstrum Katsayıları (MFCC), Delta, Delta-Delta ve Mel-spektrogram temsilleri çıkarılmış; bu temsiller hem klasik makine öğrenmesi yöntemlerinde hem de Evrişimsel Sinir Ağı (CNN) tabanlı derin öğrenme modellerinde giriş verisi olarak kullanılmıştır. Klasik yöntemler arasında Destek Vektör Makineleri (SVM), En Yakın Komşu Sınıflandırıcısı (KNN), Rastgele Orman (RF) ve XGBoost yer almaktadır. Derin öğrenme tarafında ise ResNet-50, EfficientNet-B0, MobileNetV3 ve hafif yapısıyla öne çıkan GhostNet gibi modeller değerlendirilmiştir (Ethala, 2025; Huang vd., 2023).

Model geliştirme sürecinde hasta-bağımsız eğitim-test ayrımı uygulanmış, veri dengesizliğinden kaynaklanan hataları azaltmak için veri artırma ve sınıf ağırlıklandırma stratejileri kullanılmıştır. Modeller doğruluk (accuracy), duyarlılık (recall), kesinlik (precision) ve makro F1 skorları gibi yaygın performans ölçütleriyle değerlendirilmiştir. Elde edilen bulgular, özellikle hafif yapıda tasarlanmış CNN mimarilerinin hem yüksek doğruluk sunduğunu hem de hesaplama maliyetinin düşük olması nedeniyle taşınabilir cihazlara ve klinik karar destek sistemlerine uyumlu olduğunu göstermektedir (Ethala, 2025).

Bu yönleriyle çalışma, solunum sesi sınıflandırma alanında kullanılan yöntemlere bütüncül bir bakış sunmakta; veri işleme adımlarından model karşılaştırmasına kadar geniş bir çerçeve ortaya koymaktadır. Tezin genel iş akışı, veri toplama, ön işleme, özellik çıkarımı, model eğitimi ve performans değerlendirme aşamalarını içermekte olup, bu aşamalar Şekil 1’de özetlenmiştir.



Şekil 1. Tez çalışmasında uygulanan metodolojik akış diyagramı

2. TEMEL KAVRAMLAR

2.1. Solunum Sistemi ve Fizyolojisi

Solunum, organizmanın yaşamını sürdürebilmesi için gerekli olan oksijenin çevreden alınması ve metabolizma sonucu oluşan karbondioksitin dışarı atılması sürecidir. Bu işlev solunum sistemi tarafından yürütülür ve burun, ağız, yutak, gırtlak, trakea, bronşlar, bronşiolle ve alveoller gibi pek çok yapının bir arada uyumlu çalışmasını gerektirir. Solunum sistemi anatomik olarak üst solunum yolları (burun, ağız, yutak) ve alt solunum yolları (gırtlak, trakea, bronşlar, bronşiolle, alveoller) olmak üzere iki bölümde incelenir. Üst solunum yolları havayı süzer, ısıtır ve nemlendirir; alt solunum yolları ise bu havayı akciğerlerin derin bölgelerine taşır ve gaz değişiminin gerçekleşmesine aracılık eder (Bohadana vd., 2014; Pasterkamp vd., 1997; West, 2012).

Akciğerlerin uç birimlerini oluşturan alveoller, oksijen ve karbondioksit değişiminin gerçekleştiği temel yapılarıdır. Alveol duvarı ile kılcal damarlar arasındaki ince bariyer, oksijenin kana, karbondioksitin ise alveol içine geçişine olanak sağlar. Alveol yüzeyini kaplayan sürfaktan tabakası, yüzey gerilimini azaltarak küçük hava keseciklerinin nefes verme sırasında çökmesini engeller ve solunumun daha az eforla gerçekleşmesine katkıda bulunur (West, 2012). Deneysel ve klinik çalışmalar, özellikle küçük hava yollarının açılıp kapanma döngülerinin hem solunum mekaniği hem de solunum seslerinin akustik özellikleri üzerinde önemli etkisi olduğunu, ince rallerin büyük ölçüde bu ani açılma olayları ile ilişkili olduğunu göstermektedir (Alencar vd., 1999; Pasterkamp vd., 1997; Vyshedskiy vd., 2009).

Solunum kaslarının kasılması ile göğüs boşluğunun hacmi değişir. Diyafram kası kasıldığında aşağı doğru hareket eder, kaburgalar arası kasların kasılması ile göğüs boşluğu genişler ve akciğer içi basınç atmosfer basıncının altına düşer. Böylece hava içeri çekilir ve inspirasyon gerçekleşir. Nefes verme sırasında kaslar gevşer, göğüs boşluğu daralır ve akciğer içi basınç yükselerek hava dışarı atılır. İstirahat hâlinde ekspirasyon büyük ölçüde pasif bir süreçtir; egzersiz, solunum sıkıntısı ve hava yolu direncinin arttığı durumlarda yardımcı solunum kasları devreye girer (Pasterkamp vd., 1997; West, 2012).

Solunum sırasında hava akımının farklı çaplarda ve farklı uzunluklarda hava yollarından geçerken oluşturduğu akım düzeni, solunum seslerinin temel kaynağını oluşturur. Geniş hava yollarında akım çoğunlukla düzenli iken, bronşiolle doğru ilerledikçe akım çizgileri bozulur, türbülans artar ve çevre dokularda titreşimler meydana

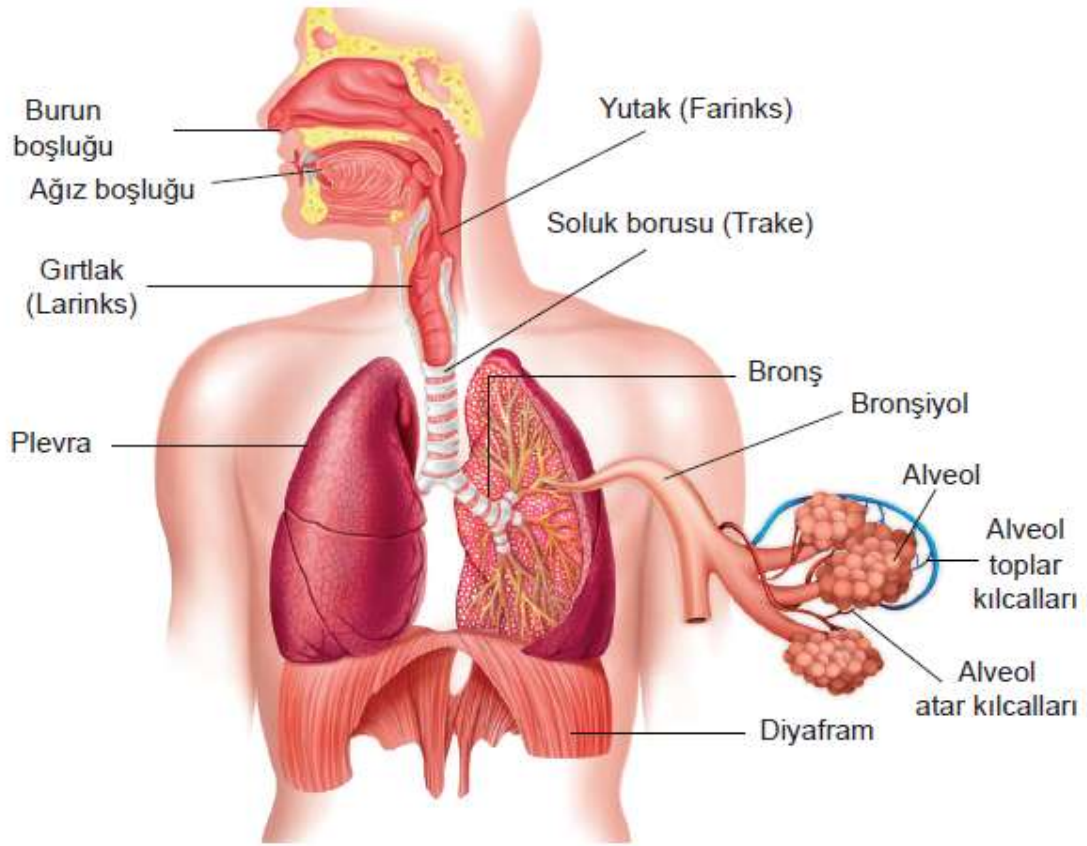
gelir. Akciğer dokusu, plevra ve göğüs duvarı bu titreşimleri filtreleyerek göğüs yüzeyine iletir ve stetoskop ya da elektronik sensörler aracılığıyla kayıtlanan solunum sesleri ortaya çıkar (Bohadana vd., 2014; Pasterkamp vd., 1997).

Sağlıklı bireylerde akciğerin dış bölgelerinde duyulan yumuşak, düşük ve orta frekans ağırlıklı solunum sesleri veziküler ses olarak adlandırılır. Trakea üzerinde ve büyük hava yollarına yakın bölgelerde daha sert, daha yüksek frekanslı bronşiyal sesler duyulur. Skapula medialinde ve sternum yanlarında yer alan bazı bölgelerde ise bu iki ses tipi arasında özellik gösteren bronkoveziküler ses işitilir (Pasterkamp vd., 1997; Sovijärvi vd., 2000).

Hastalık durumunda bu temel örüntüler bozulur ve patolojik solunum sesleri ortaya çıkar. Raller (crackles), hışıltı (wheeze), ronküs (rhonchi) ve stridor gibi ek sesler; hava yolu daralması, sekresyon birikimi, akciğer parankiminde sertleşme veya üst hava yolu darlığı gibi farklı patolojik süreçlere işaret edebilir. İnce, kısa ve patlamalı karakterli raller genellikle akciğer ödemi, zatürre ve interstisyel akciğer hastalıkları ile ilişkilendirilirken; ısıklı benzeri süregelen hışıltılar çoğunlukla astım ve kronik obstrüktif akciğer hastalığına bağlı hava yolu daralmasını düşündürür (Bohadana vd., 2014; Pasterkamp vd., 1997; Sovijärvi vd., 2000; Vyshedskiy vd., 2009).

Solunum seslerinin güvenilir değerlendirmesi için kayıt ve dinleme koşullarının olabildiğince standardize edilmesi gereklidir. Kayıt ortamının sessiz olması, stetoskop ya da sensörün doğru anatomik noktalara yerleştirilmesi, kıyafet sürtünmesinin en aza indirilmesi ve hastanın düzenli, mümkünse derin soluklar alması hem klinik oskültasyonun hem de sayısal analizlerin doğruluğunu artırır. Ayrıca göğüs duvarı kalınlığı, vücut yapısı, kullanılan cihazın frekans yanıtı ve kayıt protokolü gibi faktörler veriler karşılaştırılırken mutlaka göz önünde bulundurulmalıdır (Sovijärvi vd., 1996; 2000; West, 2012).

Şekil 2’de insan solunum sisteminin temel anatomik yapısı gösterilmektedir. Bu şekilde burun boşluğu, yutak, gırtlak, trakea, bronşlar, bronşiolleler, alveoller ve diyafram kası gibi yapılar bir arada sunularak solunum sisteminin bütüncül organizasyonu vurgulanmaktadır.



Şekil 2. İnsan solunum sisteminin temel anatomik yapısı (T.C. Milli Eğitim Bakanlığı – OGM, t.y.).

2.2. Sesin Fiziksel Özellikleri

Ses, bir kaynağın titreşmesiyle ortaya çıkan ve bir ortam boyunca yayılan mekanik bir dalgadır. Hava, sıvı ve katı ortamlar sesin iletimine olanak tanırken, boşlukta iletim gerçekleşmez. Solunum sistemi içerisinde duyulan sesler, akciğerlerden gelen hava akımının solunum yolu yapıları ile etkileşimi sonucunda oluşan mekanik titreşimlerin göğüs duvarına ve stetoskopa iletilmesiyle ortaya çıkar. Solunum seslerinin bilimsel olarak incelenebilmesi için bu titreşimlerin temel fiziksel özelliklerinin anlaşılması gereklidir (Kinsler ve Frey, 1982; Pasterkamp vd., 1997).

Bir ses dalgasının temel fiziksel bileşenleri frekans, genlik ve dalga boyudur. Frekans, bir saniyede gerçekleşen titreşim sayısını ifade eder ve Hertz (Hz) cinsinden ölçülür. Genlik, sesin algılanan şiddetini belirler ve genellikle desibel (dB) birimiyle ifade edilir. Dalga boyu, ardışık iki tepe noktası arasındaki mesafeyi tanımlar ve frekans ile ters orantılıdır. Sesin zaman içinde açılıp kapanan yapısını gösteren zarf ve sesin karakterini tanımlayan tını da solunum seslerinin değerlendirilmesinde önemli kavramlardır. Şekil 3'te, basit bir sinüzoidal dalga üzerinden genlik ve dalga boyu örneklenmektedir (Kinsler ve Frey, 1982; Yokuş, 2011).

İnsan işitme sistemi yaklaşık 20 ile 20.000 Hz arasındaki frekanslara duyarlıdır; ancak solunum sesleri bu bandın daha dar bir bölümünde yoğunlaşır. Sağlıklı bireylerde akciğerin periferal bölgelerinde duyulan veziküler solunum sesleri çoğunlukla 100 ile 1000 Hz aralığında yer alır. Trakea ve büyük hava yollarına yakın bölgelerde duyulan bronşiyal sesler ise daha geniş bir frekans dağılımı göstererek 100 ile 4000 Hz aralığına kadar uzanabilir. Patolojik ek sesler bu tabloyu önemli ölçüde değiştirir; hışıltılar genellikle 400 ile 1600 Hz arasında dar bantlı, süregelen yapılar olarak izlenirken, raller milisaniyeler düzeyinde süren, geniş bantlı ve patlamalı enerji paternleri ile karakterizedir (Bohadana vd., 2014; Pasterkamp vd., 1997; Sovijärvi vd., 2000).

Solunum seslerinin genliği, konuşma ve kalp sesleri gibi diğer biyolojik sinyallere kıyasla genellikle daha düşüktür. Bu nedenle çevresel gürültü, stetoskopun cilde temas basıncı, kablo ve giysi sürtünmesi gibi etmenler sinyal kalitesini kolaylıkla bozabilir. Sayısal analiz sürecinde bu olumsuz etkileri azaltmak amacıyla bant geçiren filtreler, gürültü bastırma algoritmaları, normalizasyon ve ölçekleme adımları uygulanır. Özellikle düşük genlikli ince raller gibi klinik açıdan önemli bileşenlerin korunabilmesi için filtreleme parametrelerinin veri setinin frekans içeriğine uygun biçimde seçilmesi önemlidir (Rocha vd., 2019; Sovijärvi vd., 2000).

Dalga boyu ile frekans arasındaki ters ilişki, solunum seslerinin göğüs duvarından iletimine dair önemli ipuçları sunar. Yüksek frekanslı bileşenler yumuşak dokular tarafından daha fazla sönmüldüğü için akciğerin periferal bölgelerinde daha çok düşük frekanslı veziküler sesler duyulur. Trakea ve ana bronşlara yakın alanlarda ise yüksek frekans bileşenleri görece daha iyi iletilir ve bronşiyal nitelikte daha sert sesler işitilir. Bu biyofiziksel özellik, aynı hastada farklı anatomik bölgelerde neden farklı ses örüntülerinin duyulduğunu açıklar (Bohadana vd., 2014; Pasterkamp vd., 1997).

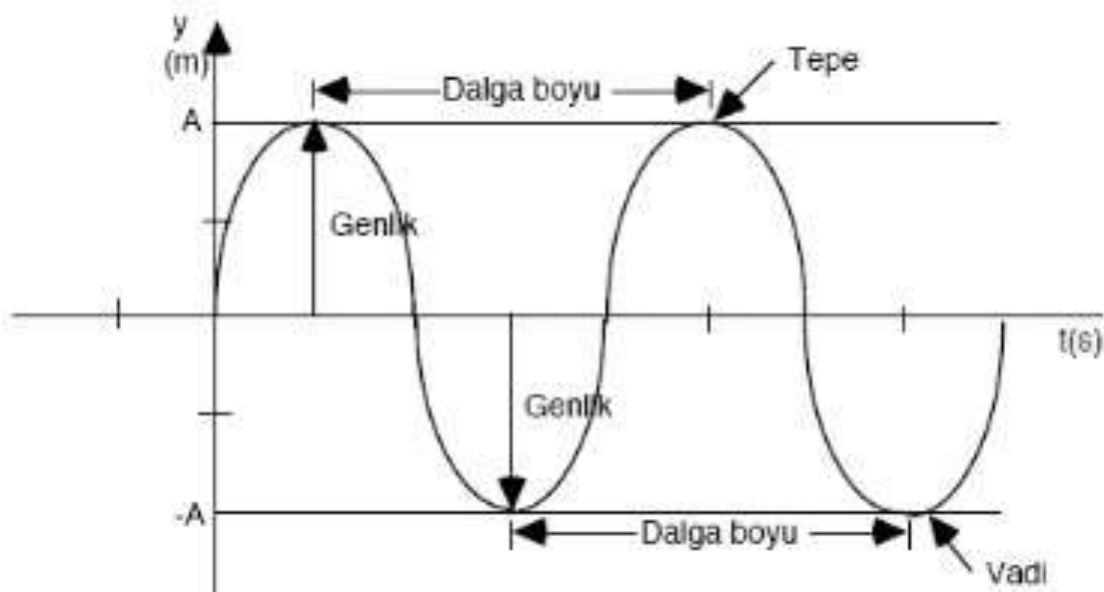
Sesin tınısı, aynı frekans ve genlikte olsalar bile iki sesi birbirinden ayırt etmemizi sağlayan akustik imzadır. Solunum patolojileri tını üzerinde belirgin değişikliklere yol açar. Hışıltılar zaman-frekans düzleminde dar bir frekans bandında uzayan çizgiler olarak izlenirken, raller çok kısa süreli, ani ve geniş bantlı enerji sıçramaları ile görünür. Bu nedenle mel-spektrogramlar, kısa zamanlı fourier dönüşümü (Short-Time Fourier Transform, STFT) ve dalgacık dönüşümleri gibi yöntemler, solunum seslerinin sınıflandırılması için sık kullanılan zaman-frekans temsillerini sağlar (Bohadana vd., 2014; Rocha vd., 2019; Sovijärvi vd., 2000).

Sayısal kayıtların kalitesi, örnekleme hızı gibi teknik parametrelerle de yakından ilişkilidir. Nyquist teoremine göre, bir sinyalin doğrulukla temsil edilebilmesi için içerdiği en yüksek frekansın en az iki katı hızla örneklenmesi gerekir. Solunum sesi

kayıtlarında örnekleme hızı çoğunlukla 4 ile 16 kHz aralığında seçilir; klinik çalışmalarda 8, 11.025 veya 16 kHz değerleri yaygın olarak kullanılır. Örnekleme öncesinde aliasing etkisini azaltmak için düşük geçiren bir süzgeç uygulanması, daha sonra pencereleme ve ön vurgu gibi işlemlerle spektral sızıntının azaltılması, hem işitsel hem de sayısal analizlerin doğruluğunu artırır (Pasterkamp vd., 1997; Sovijärvi vd., 2000).

Psikoakustik açıdan bakıldığında, insan kulağı tüm frekanslara eşit duyarlılıkta değildir. Orta frekanslarda daha yüksek duyarlılık bulunurken, çok düşük ve çok yüksek frekanslarda aynı fiziksel genlikteki sesler daha zayıf algılanabilir. Eşit gürlük eğrileri ile gösterilen bu özellik, klinik dinleme ile sayısal analizler arasında algısal farklılıklar oluşmasına neden olabilir. Bu nedenle solunum sesi analizinde spektral ağırlık merkezi, bant genişliği, sıfır geçiş oranı ve spektral düzlük gibi istatistiksel öznitelikler kullanılarak sinyalin kulağın algısına yakın biçimde temsil edilmesi hedeflenir (Bohadana vd., 2014; Zwicker ve Fastl, 1999).

Sonuç olarak solunum seslerinin fiziksel özellikleri; frekans bileşenleri, genlik düzeyleri, dalga boyu dağılımı ve tını yapısının birlikte oluşturduğu karmaşık bir akustik yapı şeklinde ortaya çıkar. Bu yapının doğru modellenmesi, hem hekimin stetoskopla yaptığı nitel değerlendirmeyi hem de bilgisayar destekli tanı sistemlerinin nicel kararlarını doğrudan etkiler. Uygun örnekleme, dikkatli ön işleme ve yerinde seçilmiş analiz yöntemleri ile solunum seslerinden güvenilir klinik bilgi elde etmek mümkündür (Bohadana vd., 2014; Pasterkamp vd., 1997; Rocha vd., 2019).



Şekil 3. Ses dalgasının temel bileşenleri genlik ve dalga boyu (Yokuş, 2011).

2.3. Solunum Seslerinin Tarihsel Gelişimi

Akciğer seslerini tanı amacıyla dinleme fikri, tıbbın yazılı tarihinden çok daha eskiye dayanır. Antik Yunan hekimlerinden Hipokrat'ın hastalarının göğüslerine kulağını dayayarak solunumu dinlediği, göğüs duvarındaki titreşimleri doğrudan hissederek bazı hastalıkları tanımlamaya çalıştığı bilinmektedir (Sarkar vd., 2015; Andrews ve Badger, 1979). Ancak bu doğrudan dinleme yöntemi hem hijyenik hem de pratik sınırlamalar içerdiğinden yaygınlaşamamıştır.

Modern anlamda akciğer dinleme, Fransız hekim René Laennec'in 1816'da stetoskopu icat etmesiyle başlar. Laennec, genç bir kadın hastayı doğrudan dinlemenin uygunsuz olacağını düşündüğü için defterini rulo hâline getirerek göğüsle kulak arasına yerleştirmiş ve böylece sesi daha iyi duyabildiğini fark etmiştir. Bu basit fikir, kısa süre içinde geliştirilen tek parça ahşap stetoskopla tıbbın ayrılmaz bir parçası hâline gelmiştir (Bohadana vd., 2014; Roguin, 2006). Laennec'in tasarladığı ilk stetoskopa ait görsel Şekil 4'te sunulmuştur (Science Museum Group Studio, 2019).



Şekil 4. René Laennec tarafından 1816 yılında tasarlanan ilk stetoskop (Science Museum Group Studio, 2019).

Laennec'in stetoskopu tanıtmasından sonraki dönemde, cihaz yalnızca bir dinleme aracı olarak kalmamış; aynı zamanda akciğer seslerinin sınıflandırılmasını ve terminolojinin oluşumunu tetiklemiştir. Örneğin, Laennec'in tanımladığı "râles" kavramı zamanla bugünkü "crackles" terimine evrilmiştir. Bu süreç günümüzde de akciğer seslerinin standardizasyonu açısından temel bir başlangıç olarak kabul edilmektedir (Pasterkamp, 2016). Ayrıca güncel çalışmalar, Laennec'in tanımları ve modern dijital

akciğer seslerinin frekans analizleri arasında güçlü bir korelasyon olduğunu göstermektedir (Xu ve Sankar, 2024).

Laennec'ten sonraki yüzyılda, hekimler solunum seslerini sınıflandırmak ve tanımlamak için ortak bir dil geliştirmeye çalıştılar. "Veziküler", "bronşiyal", "hışıltı (wheeze)" ve "ral (crackle)" gibi terimler, zamanla oskültasyon terminolojisinin temel yapıtaşları hâline geldi. Avrupa Solunum Derneği (European Respiratory Society – ERS) tarafından geliştirilen CORSA standardı, bu ses türlerinin tanımını, hangi hastalıklarla ilişkili olduklarını ve nasıl raporlanması gerektiğini sistematik hâle getirmiştir. Bu tür standartlaştırmalar hem klinik karar desteğinde hem de eğitim süreçlerinde öğretilbilirliği ve güvenilirliği önemli ölçüde artırmıştır (Pasterkamp vd., 1997; Sovijärvi vd., 2000).

Stetoskopun gelişiminde ikinci önemli sıçrama, 1850'li yıllarda çift kulaklıklılı dinleme sistemlerinin ortaya çıkmasıyla yaşanmıştır. 1851 yılında Dr. Arthur Leard, ilk binaural stetoskop örneğini Londra'daki büyük sergide tanıtmış; ancak bu modele ait fiziksel örnekler günümüze ulaşamamıştır. Ardından Amerikalı hekim Dr. George P. Cammann, bugünkü klinik uygulamalarda hâlen kullanılan binaural stetoskop tasarımını geliştirerek yaygınlaştırmıştır. Cammann'ın tasarımında metal kollar üzerindeki yay mekanizması kulak uçlarının sabitlenmesini sağlarken, fildişi uçlar sesin netliğini artırmıştır (Yaqub, 2015). Şekil 5'te Cammann tipi erken dönem binaural stetoskopa ait görsel sunulmuştur (Science Museum Group, 2019).



Şekil 5. Erken dönem binaural stetoskop (Science Museum Group, 2019).

20. yüzyılın ortalarından itibaren teknoloji stetoskopa entegre edilmeye başlandı. Önce mikrofonlar eklenerek sesin yükseltilmesi sağlandı; ardından elektronik stetoskoplarla kayıt alma, sayısal olarak işleme ve analiz etme olanakları doğdu. Böylece

dinleme artık sadece o anla sınırlı kalmadı; sesler dijital ortama kaydedilip arşivlenebilir hâle geldi (Kim vd., 2023). Bu gelişme, aynı hastanın farklı zamanlardaki bulgularının ya da farklı hastalar arasındaki karşılaştırmaların daha objektif bir temelde yapılabilmesini mümkün kıldı. Ayrıca sesin genlik ve frekans içeriğini değiştirerek duyulması zor bileşenleri öne çıkaran filtreleme teknikleri sayesinde tanı duyarlılığı artırıldı (Bohadana vd., 2014; Sovijärvi vd., 2000). Elektronik stetoskopa ait bir örnek Şekil 6'da görülmektedir.



Şekil 6. Elektronik Stetoskop (Webtekno, 2021).

21. yüzyıla gelindiğinde dijitalleşme yalnızca sesleri kaydetme aşamasında kalmamış; otomatik analiz ve yapay zekâ (YZ) temelli sınıflandırma sistemlerini de mümkün kılmıştır. Solunum sesleri artık yalnızca duyulmakla kalmıyor, aynı zamanda zaman–frekans düzleminde nicel olarak çözümlenebiliyor ve otomatik olarak sınıflandırılabilir (Abeyratne vd., 2013; Rocha vd., 2017). Son yıllarda yapılan sistematik incelemeler, YZ temelli akciğer sesi analizlerinin özellikle hışıltı ve ral gibi patolojik örüntülerin tanımlanmasında uzman hekimlerin performansına yakın doğruluklara ulaşabildiğini göstermektedir (Seah vd., 2023).

Ayrıca evde takip, teletıp uygulamaları ve giyilebilir stetoskop sistemleri üzerine yayımlanan güncel çalışmalar, stetoskopun değerlendirme şeklini köklü biçimde değiştirmektedir. Özellikle COVID-19 sonrası dönemde, akciğer seslerinin uzaktan otomatik analizi klinik yönetimde önemli bir araç hâline gelmiştir (Xu ve Sankar, 2024). Taşınabilir elektronik stetoskopların düşük maliyetli, kablosuz ve bulut bağlantılı hâle gelmesi sayesinde hastaların uzun dönem izlenmesi mümkün olmuş; bu durum hem kronik akciğer hastalıklarında hem de akut solunum yolu enfeksiyonlarında erken uyarı sistemlerinin geliştirilmesini desteklemiştir (Kim vd., 2023).

Sonuç olarak stetoskopun tarihi; Hipokrat'ın doğrudan dinleme uygulamasından Laennec'in icadına, binaural tasarımın ortaya çıkmasına ve günümüzün dijital ve yapay zekâ destekli modern stetoskoplara kadar uzanan çok katmanlı bir evrimi temsil eder. Bu süreç yalnızca klinik muayene pratiğini değil; aynı zamanda akustik sinyal işleme, uzaktan izlem ve karar destek sistemleri gibi alanların da gelişimini belirgin şekilde etkilemiştir.

2.4. Solunum Seslerinin Oluşumu

Solunum sesleri, akciğerlerdeki hava akımının bronşiyal ve alveoler düzeyde oluşturduğu mekanik titreşimlerin, çevre dokular ve göğüs duvarı aracılığıyla dış ortama iletilmesiyle ortaya çıkar. Bu seslerin özellikleri yalnızca hava akımının varlığına değil, aynı zamanda hava yolu geometrisine, doku elastikiyetine ve mevcut patolojik süreçlere bağlıdır. Farklı genişlik ve uzunluktaki hava yollarından geçen hava, belirli bölgelerde türbülans oluşturarak bu yapıların titreşimini tetikler; akciğer parankimi, plevra ve göğüs duvarı ise bu titreşimleri filtreleyerek göğüs yüzeyinde duyulan nihai ses örüntüsünü şekillendirir (Bohadana vd., 2014; Pasterkamp vd., 1997; Sovijärvi vd., 2000).

Geniş hava yollarında akım çoğunlukla laminar özellik gösterdiği için sessiz veya düşük şiddetli iken, daha dar ve dallanan bölgelere geçildikçe akım çizgileri bozulur ve türbülans artar. Akciğerin periferik bölgelerinde, özellikle küçük bronşoller düzeyinde oluşan bu mikrotürbülanslar, yumuşak ve düşük şiddetli veziküler solunum seslerinin temel kaynağıdır. Hava yolu daralması, sekresyon birikimi, doku elastikiyetinin azalması veya alveollerin geç açılıp erken kapanması gibi durumlarda akım düzeni bozulur; bu da raller, hışıltılar ve ronküsler gibi patolojik seslerin ortaya çıkmasına yol açar. Bu nedenle solunum sesleri, yalnızca anatomik yapı hakkında değil, hava yolu direnci ve parankim sertliği gibi fizyopatolojik süreçler hakkında da dolaylı bilgi sağlar (Bohadana vd., 2014; Kandaswamy vd., 2004; Pasterkamp vd., 1997).

Kaynak bölgede oluşan mekanik titreşimler, akciğer dokusu ve göğüs duvarından geçerken dokular tarafından önemli ölçüde sönümlenir ve filtrelenir. Özellikle yüksek frekanslı bileşenler yumuşak dokularda daha fazla sönümlenir; bu nedenle periferik akciğer bölgelerinde daha düşük frekanslı veziküler sesler baskın hâle gelirken, trakea ve büyük bronşlara yakın alanlarda daha yüksek frekanslı bronşiyal sesler belirginleşir. Göğüs duvarı bu anlamda doğal bir akustik filtre gibi davranır ve stetoskop ya da mikrofon tarafından kaydedilen sinyal, kaynaktan çıkan sesin filtrelenmiş bir sürümü olarak değerlendirilebilir (Bohadana vd., 2014; Pasterkamp vd., 1997).

Solunum seslerinin klinik değerlendirilmesinde en sık kullanılan yöntem stetoskop aracılığıyla yapılan oskültasyondur. Oskültasyon; göğüs ön (anterior), arka (posterior) ve yan (lateral) yüzeylerinde belirlenmiş anatomik noktaların sistematik olarak dinlenmesini içerir. Dinleme sırasında hastanın mümkünse oturur pozisyonda olması, derin ve düzenli nefes alıp vermesi ve stetoskop diyaframının doğrudan cilde temas etmesi önerilir. Kıyafet, saç veya takı gibi unsurlar ile stetoskop başlığının sürtünmesi, özellikle düşük ve orta frekanslı aralıkta istenmeyen gürültü oluşturarak solunum seslerinin değerlendirilmesini güçleştirebilir.

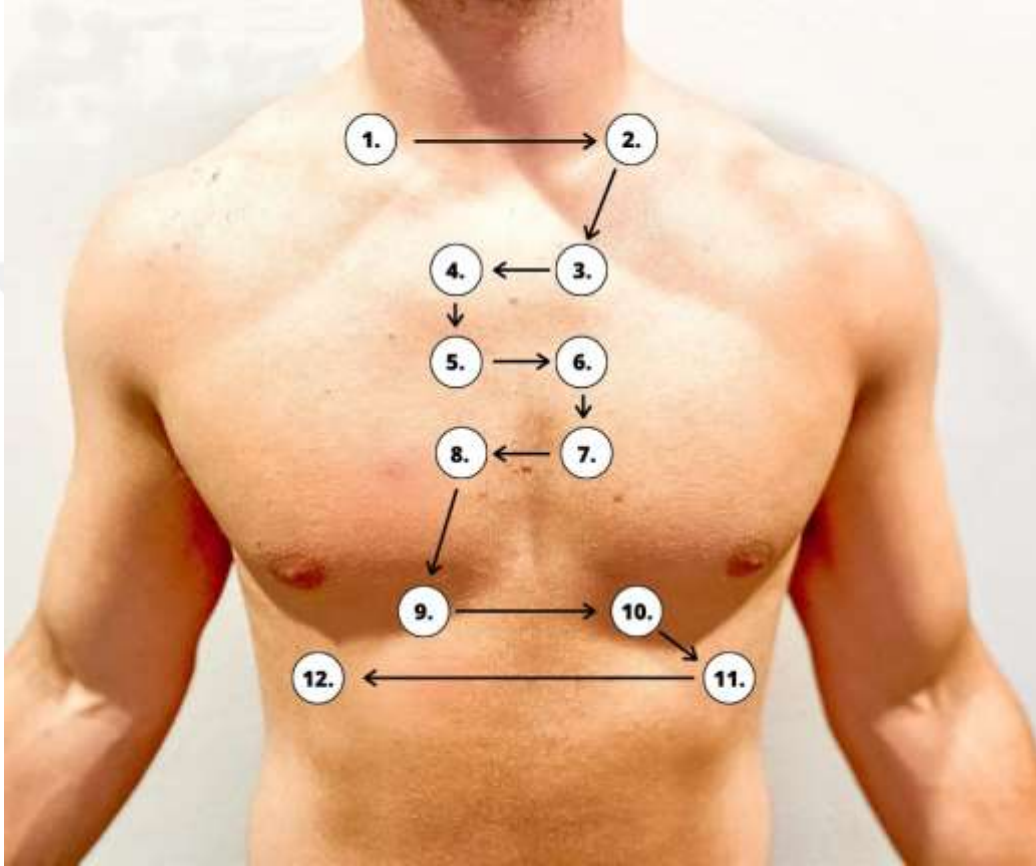
Dinleme noktaları, akciğer lob ve segmentlerini temsil edecek şekilde simetrik olarak planlanır. Klinik uygulamada, her iki hemitoraksta da üst, orta ve alt akciğer alanlarını kapsayan bir dinleme dizisi önerilir; trakea bölgesi de bronşiyal seslerin değerlendirilmesi için ayrı olarak dinlenir. Böylece sağ ve sol akciğer arasında duyulabilecek farklılıklar, hem temel solunum sesleri hem de patolojik ek sesler açısından daha güvenilir biçimde saptanabilir (Pasterkamp vd., 1997; Sovijärvi vd., 2000). Oskültasyon noktalarının sistematik olarak belirlenmesi ve kayıtların bu noktalardan alınması, son yıllarda geliştirilen solunum sesi veri tabanlarının da ortak yaklaşımı hâline gelmiştir.

Uygulamada en sık kullanılan dinleme düzeni, yukarıdan aşağıya doğru ve sağ-sol simetri gözetilerek ilerleyen bir sıra izler. Temel klinik yaklaşımlarda göğüs ön yüzünde yaklaşık 6–8, arka yüzünde ise 8–10 noktanın dinlenmesi önerilmektedir. Bu tez kapsamında kullanılan şekillerde, yaygın kabul gören anterior (Şekil 7) ve posterior (Şekil 8) oskültasyon noktaları şematik olarak gösterilmiştir. Bu haritalama, hem klinik muayenede hem de dijital kayıtların standardizasyonunda solunum seslerinin karşılaştırılabilir biçimde analiz edilmesine olanak tanır (Pomietlo, 2020).

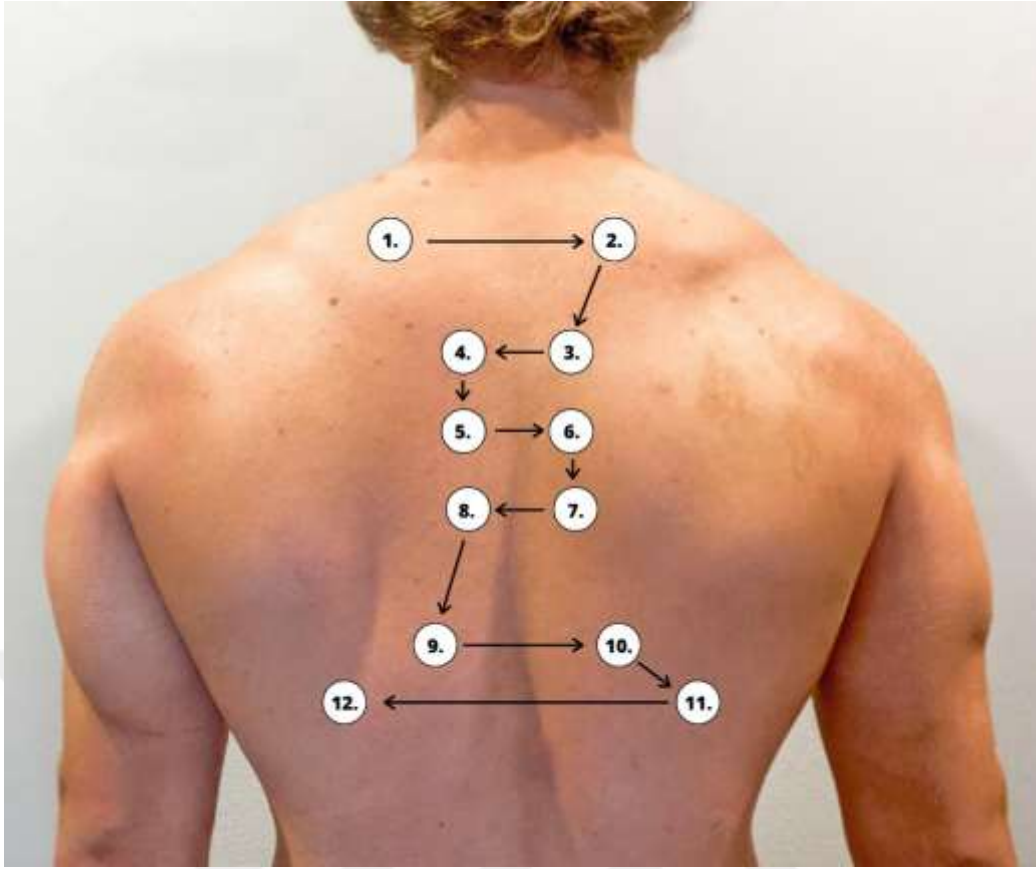
Bununla birlikte solunum seslerine çoğu zaman kalp sesleri, kas aktivitesi ve çevresel gürültü gibi istenmeyen bileşenler de karışır. Kalp sesleri özellikle sol hemitoraksta belirgindir ve düşük frekanslı olmaları nedeniyle bazı solunum seslerini maskeleyebilir. Stetoskop başlığının cilt veya kıyafetle sürtünmesi de geniş bantlı ve düzensiz yapıda artefaktlar oluşturur. Bu nedenle özellikle bilgisayar destekli analizlerde bant geçiren filtreler, gürültü bastırma yöntemleri, kalp sesi ayrıştırma algoritmaları ve solunum döngüsü segmentasyonu gibi ön işleme adımları kritik önem taşır; yeni çalışmalarda bu tür ön işleme stratejilerinin otomatik sınıflandırma başarımını anlamlı ölçüde artırdığı gösterilmiştir (Agnew vd., 2023; Rocha vd., 2019; Sovijärvi vd., 1996).

Sonuç olarak solunum seslerinin oluşumu, hava akımı dinamikleri, hava yolu anatomisi ve göğüs duvarı ile akciğer dokusunun fiziksel özelliklerinin ortak etkisiyle

ortaya çıkan çok bileşenli bir süreçtir. Normal koşullarda veziküler, bronkoveziküler ve bronşiyal sesler dinlenirken; hastalık varlığında bu temel seslere raller, hışıltılar, ronküsler ve stridor gibi patolojik sesler eklenebilir. Bu seslerin doğru şekilde algılanması ve yorumlanması, hem klasik klinik değerlendirme hem de geliştirilen otomatik tanı sistemlerinin başarısı açısından büyük önem taşımaktadır (Bohadana vd., 2014; Kim vd., 2022; Pasterkamp vd., 1997; Sovijärvi vd., 2000).



Şekil 7. Anterior solunum oskültasyon noktaları (Pomietlo, 2020).



Şekil 8. Posterior solunum oskültasyon noktaları (Pomietlo, 2020).

2.5. Normal Solunum Sesleri

Normal solunum sesleri, havanın büyük hava yollarından akciğerin uç kısımlarına doğru düzenli biçimde ilerlediği ve akciğer dokusu ile göğüs duvarının sesi adeta doğal bir süzgeç gibi süzdüğü durumlarda duyulur. Klinik dinlemede üç ana örüntü öne çıkar: veziküler, bronkoveziküler ve bronşiyal ses. Bu üç ses tipi; nefes alma ve verme sürelerinin birbirine göre uzunluğu, şiddeti, sesin rengi/tonu ve duyulduğu bölgeler açısından birbirinden ayrılır. Sağlıklı bireyde hangi bölgede hangi nitelikte sesin duyulmasının beklendiğini bilmek, örneğin akciğerin dış bölgelerinde bronşiyal nitelikte ses işitildiğinde bunu hızla "yer değiştirmiş" bir bulgu olarak fark etmeyi ve doğru yorumu yapmayı kolaylaştırır (Bohadana vd., 2014; Pasterkamp vd., 1997).

2.5.1. Veziküler Ses

Veziküler ses, sağlıklı bireyde en sık duyulan temel sestir ve akciğerin dış (periferik) bölgelerinde stetoskopla dinlendiğinde belirginleşir. Yumuşak, düşük–orta frekanslarda yoğunlaşan ve kesintisiz bir uğultu gibidir. Nefes alma, nefes vermeden daha uzun ve daha yüksek şiddette duyulur; nefes verme daha kısadır ve daha düşük şiddettedir. İki faz arasında keskin bir durak hissi beklenmez; geçiş doğal ve akıcıdır (Pasterkamp vd., 1997).

Bu karakterin temelinde, küçük hava yollarında havanın çoğunlukla düzenli akması ve akciğer dokusunun yüksek frekansları süzen doğal bir ortam gibi davranması yatar. Göğüs duvarının kalın olduğu, plevra boşluğunda sıvı biriktiği ya da amfizemde olduğu gibi akciğerde aşırı hava tutulduğu durumlarda veziküler sesin şiddeti azalabilir ve bu bulgu "azalmış solunum sesi" olarak not edilir. Çocuklarda göğüs duvarının ince olması nedeniyle veziküler ses daha keskin ve nispeten daha yüksek frekanslı algılanabilir; bu fizyolojik farklılık patolojiyle karıştırılmamalıdır (Bohadana vd., 2014; Sovijärvi vd., 2000).

Akciğerin dış (periferik) bölgelerinde bronşiyal nitelikte solunum sesi duyulması normal değildir; böyle bir bulgu çoğunlukla parankimde yoğunlaşma (konsolidasyon) olasılığını düşündürür. Yoğunlaşmış doku, büyük hava yollarından gelen yüksek frekans bileşenlerini periferik alanlara daha az sönümle ilettiğinden periferde bronşiyal tını (sert, boru benzeri ve çoğu kez ekspirasyonun daha belirgin olduğu bir karakter) ortaya çıkabilir. Bu nedenle periferde beklenenden farklı bir temel ses işitildiğinde, bulgu mutlaka aynı tarafın komşu bölgeleri ve karşı akciğerle sistematik olarak karşılaştırılarak yorumlanmalıdır (Pasterkamp vd., 1997; Bohadana vd., 2014).

2.5.2. Bronşiyal Ses

Bronşiyal solunum sesi, sert ve boru tınısına yakın bir sestir. En iyi, nefes borusunun (trakea) üstünde ve göğüs ön duvarının orta hattına yakın bölgede duyulur; buralarda işitilmesi normal kabul edilir. Bu seste nefes alma ve verme evreleri birbirine benzer güçte olur; arada kısa bir durak hissedilebilir ve çoğu zaman nefes verme daha uzun ve daha yüksek perdede duyulur. Bu özellikleriyle, yumuşak ve uğultu şeklindeki veziküler sestten kolayca ayrılır (Pasterkamp vd., 1997; Bohadana vd., 2014).

Uyarıcı olan durum, bu sert tınının akciğerin dış bölgelerinde duyulmasıdır. Normalde bu alanlarda doku sesi yumuşatır ve "veziküler" nitelik beklenir. Dış bölgede boru tınısına yakın, nefes vermenin daha baskın olduğu bir ses işitiliyorsa, akciğer dokusunun yapısında değişiklik olmuş olabilir. Örneğin:

- Akciğerin bir bölümünün iltihaplanıp sertleşmesi (günlük dilde "zatürre odağı" diyebileceğimiz durum),
- Akciğerin bir parçasının sönmesi ama o bölgeye giden hava yolunun açık kalması,
- Akciğer zarları arasında sıvı birikmesi ve bu sıvının üst sınırına yakın bölgeler, gibi tablolar, merkeze ait sert seslerin dış bölgelere daha az kayıpla taşınmasına yol açar; sonuçta dış alanda bronşiyal tını ortaya çıkar (Bohadana vd., 2014; Pasterkamp vd., 1997).

Bazen tını daha da farklı bir hâl alır: boş bir kaba üfler gibi metalik ve çınlayıcı duyulur. Bu, akciğerde genişçe bir boşluk oluştuğunda ya da akciğer ile göğüs duvarı arasına hava kaçtığında görülebilir. Yine bazı hastalarda, sıvı birikiminin üstünde ses ince ve burundan gelen bir tonda iletilir; bu da değerlendirici için ek bir ipucudur. Tüm bu tını farkları, mutlaka sesin duyulduğu yer, karşı akciğerle karşılaştırma ve eşlik eden muayene görüntüleme bulgularıyla birlikte yorumlanmalıdır (Bohadana vd., 2014; Sovijärvi vd., 2000).

Sonuç olarak: Bronşiyal ses merkezî alanlarda normaldir; akciğerin dış bölgelerinde duyuluyorsa altta yatan bir sorun olasılığını gündeme getirir. Dinlemede evrelerin göreceli süresi ve şiddeti, tının sertliği, aradaki kısa durakların varlığı ve sesin duyulduğu yer birlikte ele alındığında, bronşiyal–veziküler ayrımı güvenle yapılır ve klinik karar için sağlam bir zemin oluşturur (Pasterkamp vd., 1997; Sovijärvi vd., 2000).

2.5.3. Bronkoveziküler Ses

Bronkoveziküler ses, veziküler ve bronşiyal sesin arasında kalan bir tınıdır. Ne tamamen yumuşaktır ne de boru gibi serttir. Orta yükseklikte ve orta şiddettedir. Nefes alma ile nefes verme süre ve şiddet olarak birbirine yakın duyulur. Nefes verme, veziküler sese göre biraz daha belirgin olsa da bronşiyalde olduğu kadar baskın değildir. Bu sesin en sık ve normal olarak duyulduğu yerler, omuz bıçaklarının iç kenarına yakın arka bölge ile göğüs ön duvarında göğüs kemiğinin yan taraflarıdır. Bu alanlarda işitildiğinde fizyolojik kabul edilir (Bohadana vd., 2014; Pasterkamp vd., 1997).

Akciğerin dış bölgelerinde bronkoveziküler tını duyulması uyarıcı bir bulgudur. Bu durum, o bölgedeki havalanmanın azaldığını ama büyük hava yollarından gelen sesin yine de daha kolay iletildiğini düşündürür. Günlük pratikte, akciğer dokusunun sertleşip yoğunlaştığı erken dönem bir zatürre odağı ya da sönmüş bir akciğer parçasının sınırı gibi durumlarda bu nitelik ortaya çıkabilir. Böyle bir bulgu mutlaka aynı tarafın diğer bölgeleri ile karşı akciğerle karşılaştırılmalı ve duyulduğu yerin yayılımı ile birlikte değerlendirilmelidir. Gerekirse konuşma sesi iletimi ve diğer muayene işaretleri de gözden geçirilmelidir (Pasterkamp vd., 1997).

Gündelik dinlemede ayrımı kolaylaştıran birkaç ipucu vardır. Veziküler sese göre nefes vermenin biraz daha belirgin oluşuna ve tının daha toplu duyulmasına dikkat edin. Bronşiyal sese göre ise sesin boru gibi sert ya da metalik çınlamaya sahip olmamasına ve nefes alma ile verme arasında keskin bir durak hissedilmemesine bakın. Çocuklarda göğüs duvarı daha ince olduğu için bronkoveziküler tını daha geniş bir alanda fizyolojik

olarak duyulabilir. Erişkinlerde dış bölgelerde işitildiğinde çoğunlukla altta yatan bir sorun olasılığı akla gelmelidir (Bohadana vd., 2014; Sovijärvi vd., 2000).

Sonuç olarak bronkoveziküler ses, belirli bölgelerde normaldir. Bu alanların dışında duyulduğunda havalanmayı bozan ve sesi olduğundan daha sert ileten nedenler araştırılmalıdır. Düzenli bir karşılaştırmalı dinleme ile doğru yer, doğru nitelik ve doğru bağlam birlikte ele alındığında bu ara tip ses tanıda değerli bir yol gösterici olur (Pasterkamp vd., 1997; Sovijärvi vd., 2000).

2.6. Patolojik Solunum Sesleri

Normalde akciğer alanlarında duyulan temel solunum seslerinden farklı karakter taşıyan tüm sesler “patolojik” olarak değerlendirilir. Bu seslerin ortaya çıkışı genellikle hava akımının bozulması, hava yollarında daralma ya da tıkanıklık oluşması, balgam veya sekresyon birikimi ya da akciğer dokusunda gelişen yapısal değişikliklerle ilişkilidir. Klinik değerlendirme sırasında, duyulan sesin solunum döngüsünün hangi fazında (inspirasyon veya ekspirasyon) belirginleştiği, kısa patlamalar hâlinde mi yoksa uzayıp giden bir tonda mı olduğu, frekansının (perdesinin) yüksek mi yoksa düşük mü olduğu ve öksürük ile değişip değişmediği ayrıntılı şekilde not edilmelidir. Bu sistematik yaklaşım, hem ayırıcı tanı sürecinde hem de tedaviye verilen yanıtın izlenmesinde önemli ölçüde güvenilirlik sağlar (Bohadana vd., 2014; Pasterkamp vd., 1997).

2.6.1. Raller (Crackles)

Raller, kısa süreli ve patlama tarzında çıkan çıtırtı sesleridir. Genellikle nefes alma sırasında duyulurlar; bazı durumlarda nefes verirken de işitilebilirler. İnce raller daha tiz ve kısadır; zatürre, akciğer ödemi veya bazı yaygın akciğer hastalıklarında görülür. Kalın raller ise daha pes ve kaba olup bronşektazi, balgam birikimi ya da KOAH gibi hava yollarını etkileyen hastalıklarla ilişkilidir. Nefes alma sonuna doğru belirginleşen raller, hava keseciklerinin geç açıldığına ve akciğer dokusunun sertleştiğine işaret eder. Nefes alma başında veya nefes verme sonunda ortaya çıkan raller, büyük hava yollarında biriken salgıyı düşündürür. Rallerin yer değiştirmemesi ve öksürükle kaybolmaması, akciğer dokusuna ait bir sorun olduğunu destekler. Bu nedenle rallerin zamanı, şiddeti ve bulunduğu yer dikkatle kaydedilmeli ve gerektiğinde takip için kullanılmalıdır (Bohadana vd., 2014; Pasterkamp vd., 1997; Sovijärvi vd., 2000).

2.6.2. Hışıltı (Wheeze)

Hışıltı, ısıklık benzeri, ince ve uzayan bir sestir. Genellikle nefes verme sırasında duyulur; daralmanın ciddi olduğu durumlarda nefes alma sırasında da işitilebilir. Hava yolunun dar bir bölgesinden hızla geçen hava, duvarlarda titreşim oluşturarak kulakta tek perdeli ya da çok perdeli bir ses olarak algılanır. Tek perdeli hışıltı, sınırlı bir daralmayı; çok perdeli hışıltı ise yaygın daralmayı düşündürür. Hışıltı en çok astım ve KOAH alevlenmelerinde görülür. Hastalık şiddetlendikçe, sesin süresi uzar ve tonu yükselir. Bronş genişletici tedavi sonrası azalması, daralmanın geri döndürülebilir olduğunu gösterir. Hışıltının tek taraflı mı, çift taraflı mı olduğu, hangi bölgede duyulduğu ve solunum döngüsünün hangi evresinde belirginleştiği titizlikle değerlendirilmelidir. Tek taraflı hışıltı durumunda, yabancı cisim varlığı, bölgesel tıkanıklık veya tümör gibi nedenler düşünülmelidir (Bohadana vd., 2014; Pasterkamp vd., 1997; Sovijärvi vd., 2000).

2.6.3. Ronküs (Rhonchi)

Ronküs, genellikle kalın, uğultulu ve boğuk karakterde bir sestir. Büyük hava yollarında biriken balgam nedeniyle oluşur. Hışıltıya göre daha pes tondadır ve genellikle dalgalanır gibi bir karakter sergiler. En ayırt edici özelliği, öksürükle azalması veya tamamen kaybolmasıdır; bu da sesin kaynağının hareketli salgılar olduğunu gösterir. Ronküs, özellikle bronşit, KOAH alevlenmeleri ya da üst solunum yolu salgılarının bronşlara ilerlemesi gibi klinik tablolarla ilişkilidir. Ronküs duyulduğunda, sesin nereden geldiği, hangi tarafta olduğu ve öksürükle değişip değişmediği ayrıntılı şekilde değerlendirilmelidir. Öksürük sonrası düzelmeyen ve bölgesel olarak duyulan kalın uğultulu bir ses varsa, hava yolunda tıkanmaya neden olabilecek başka yapılar araştırılmalıdır (Bohadana vd., 2014; Sovijärvi vd., 2000).

2.6.4. Stridor

Stridor, genellikle üst solunum yolunda ciddi daralma olduğunda ortaya çıkan, yüksek perdeli, kaba ve sert bir sestir. Çoğu zaman nefes alma sırasında belirgindir ve hasta konuşurken ya da istirahatleyken dahi dışarıdan duyulabilir. Bu kadar keskin ve dikkat çekici olması, hava yolundaki daralmanın hayati tehlike oluşturabilecek düzeyde olduğuna işaret eder. Gırtlak düzeyinde ödem, yabancı cisim aspirasyonu, larenjit veya yapısal bozukluklar bu sesin başlıca nedenleri arasındadır. Stridor işitildiğinde, hastanın solunum durumu hızla değerlendirilmelidir. Sesin yalnızca nefes alma sırasında mı yoksa nefes verirken de devam edip etmediği ve tek mi çift taraflı mı olduğu dikkatlice

gözlemlenmelidir. Bu bulgular, daralmanın düzeyi ve şiddeti hakkında önemli bilgiler verir (Sovijärvi vd., 2000).

2.7. Solunum Seslerinin Frekans Özellikleri

Solunum sesleri, sesin kaynağını oluşturan yapıların özelliklerine ve hava akımının düzenine bağlı olarak belirli frekans aralıklarında yoğunlaşır. Sağlıklı bireylerde akciğerin dış (periferik) bölgelerinde stetoskopi dinlenen veziküler solunum sesleri genellikle 100–1000 Hz arasında yer alır. Bu nedenle bu sesler yumuşak, hafif uğultu şeklinde ve sürekliliği bozulmayan bir karakter gösterir. Buna karşılık, soluk borusu (trakea) ve büyük bronşlar üzerinden duyulan bronşiyal sesler 150–4000 Hz aralığında, daha sert, daha parlak ve keskin bir ton ile karakterizedir (Pasterkamp vd., 1997; Sovijärvi vd., 2000).

Patolojik ek seslerin frekans içeriği, bu temel tabloyu önemli ölçüde değiştirir. Nefes alma sonunda kısa, patlayıcı şekilde ortaya çıkan ince raller çoğunlukla 350–2000 Hz bandında yoğunlaşır. Daha kalın ve genellikle “ıslak” nitelikte algılanan raller ise 150–1000 Hz aralığında görülür. Hava yollarındaki daralmanın neden olduğu ve süreklilik gösteren hışıltı (wheeze), genellikle 400–2000 Hz aralığında dar bir frekans bandında belirir. Üst solunum yollarındaki ciddi darlığa işaret eden stridor sesi ise çoğunlukla 500–2000 Hz arasında, yüksek perdeli ve keskin bir profil gösterir (Pasterkamp vd., 1997; Sovijärvi vd., 2000).

Bu frekans aralıkları yalnızca tanımlayıcı değil; aynı zamanda solunum seslerinin bilgisayar destekli analizinde uygulanacak sinyal işleme adımlarının tasarımını da yönlendirir. Örneğin, analiz öncesinde uygulanacak bant geçiren süzgecin alt sınırı çoğu zaman 80–100 Hz olarak belirlenir. Bu sayede kalp sesleri, vücut hareketlerinden kaynaklanan düşük frekanslı bozulmalar ve çevresel titreşimler bastırılmış olur. Üst sınır ise konuşma sesleri ve yüksek frekanslı gürültülerin etkisini azaltmak amacıyla genellikle 1800–2000 Hz civarında sınırlanır.

Zaman–frekans çözümlemelerinde, analiz penceresinin süresi ve örtüşme miktarı da ses türüne göre dikkatle belirlenmelidir. Veziküler sesin görece yavaş değişen karakteristiği için daha uzun pencereler tercih edilirken; milisaniyeler içinde oluşan ince rallerin kaçırılmaması adına daha kısa pencereler ve yüksek örtüşme oranları seçilir. Hışıltı seslerinin dar bir frekans bandında ve uzayan bir çizgi şeklinde görünmesi, yeterli frekans çözünürlüğünü zorunlu kılar (Pasterkamp vd., 1997).

Pratikte, bu frekans özelliklerinin kişiden kişiye ve kayıt yapılan bölgeye göre değişebileceği unutulmamalıdır. Göğüs duvarının kalınlığı, stetoskop başlığının

yerleştirildiği nokta ve uygulanan basınç, ölçüm bölgesinin soluk borusuna olan uzaklığı ve kullanılan cihazın frekans yanıtı gibi birçok faktör, özellikle yüksek frekansların iletimini etkileyebilir. Bu nedenle analiz öncesinde kısa bir deneme kaydı alınarak sinyal-gürültü oranı ve hedef seslerin görünürlüğü birlikte değerlendirilmelidir. Böylece filtreleme, özellik çıkarımı ve sınıflandırma gibi adımlarda daha bilinçli ve güvenilir kararlar alınabilir; bu da genel analiz performansını artırır (Pasterkamp vd., 1997; Sovijärvi vd., 2000).

2.8. Solunum Seslerinin Klinik Önemi

Oskültasyon, solunum sistemi hastalıklarının değerlendirilmesinde ilk başvuru olan yöntemlerden biridir ve doğru yorumlandığında yalnızca ses var mı, yok mu? sorusunun ötesine geçer. Hekim, sesin tınısına, hangi soluk evresinde ortaya çıktığına ve göğüs üzerinde nasıl yayıldığına bakarak hastalığın yaygınlığı, evresi ve şiddeti hakkında çıkarımlarda bulunabilir. Örneğin, astım vakasında hışıltının duyulduğu alanın genişlemesi, birden çok perdede işitilmesi ve nefes vermenin sonuna kadar sürmesi, hava yolu daralmasının arttığını ve alevlenmenin yaklaştığını düşündürür. Benzer şekilde, kalp yetmezliği ya da akciğer ödemi olan bir hastada alt akciğer alanlarında duyulan ince raller, akciğerde sıvı birikiminin arttığını işaret eder; uygun tedavi sonrası bu rallerin zayıflaması, erken yanıt işareti olabilir. Öte yandan, tek bir bölgeyle sınırlı hışıltı ya da belirgin solunum sesi azalması ise çoğu kez yabancı cisim, mukus tıkaçı ya da büyük hava yolunu tutan bir kitle gibi odaklanmış bir soruna işaret eder; bu durumda ileri görüntüleme ve gerekirse girişimsel değerlendirme gündeme gelir (Bohadana vd., 2014; Sovijärvi vd., 2000).

Güncel yaklaşımlar, bu klinik gözlemi dijital kayıt ve sayısal analiz ile desteklemektedir. Sesin standardize koşullarda kaydedilebilmesi, tek seferlik bir muayeneyi zaman içinde karşılaştırılabilir ölçümlere dönüştürür. Böylece aynı hastada günler veya haftalar içinde ortaya çıkan değişim daha objektif biçimde izlenebilir. Örneğin, KOAH hastalarında alevlenme öncesinde hışıltı yoğunluğunun artması ya da rallerin daha sık hâle gelmesi, otomatik sistemler tarafından “erken uyarı” olarak işaretlenebilir (Sarkar vd., 2015; Zulfıkar vd., 2021). Dijital yaklaşım ayrıca hekimin subjektif yorumlarını dengeleyerek gözlemciler arası farklılığı azaltma potansiyeline sahiptir (Zimmerman ve Williams, 2023). Klinik karar destek sistemleriyle entegre edilen bir oskültasyon ve ses-analiz hattı, özellikle birinci basamakta ve evde izlemde, eşik değerlerin aşılması durumunda hekimi zamanında uyarmaya ve başvuru gecikmelerini azaltmaya yardımcı olabilir (Kim vd., 2022).

2.9. Veri Ön İşleme Yöntemleri

Solunum sesi kayıtları genellikle klinik ortamda alınır ve bu kayıtlar ortam gürültüsü, konuşmalar, stetoskop kablo sürtünmesi, hasta hareketi ve kalp sesleri gibi istenmeyen bileşenlerle birlikte gelir. Ayrıca kullanılan kayıt cihazının türü, göğüs üzerindeki yerleşim noktası ve oda akustiği değişkendir. Bu çeşitlilik, doğrudan sınıflandırma metoduna geçildiğinde hata oranını artırabilir. Ön işleme adımlarının temel amacı, ham veriyi daha tutarlı ve karşılaştırılabilir bir yapıya dönüştürmektir. Temel hedefler şöyle özetlenebilir: genlik ölçeklerinin hizalanması, anlamlı frekans içeriğinin korunarak gürültünün azaltılması ve ilgili solunum döngülerinin temiz biçimde ayrılması. İyi tasarlanmış bir ön işleme hattı, hem klinik dinlenebilirliği artırır hem de çıkarılacak özniteliklerin kararlılığını yükseltir (Sovijärvi vd., 1996; Rocha vd., 2019; Fava, 2024).

2.9.1. Normalizasyon

Farklı cihazlar ve kayıt koşulları altında elde edilen aynı tip sesler büyük genlik farklılıkları gösterebilir. Normalizasyon, bu genlik farklarını azaltır ve algoritmanın “sesin ne kadar yüksek olduğuna” değil, “sesin yapısına” odaklanmasını sağlar. Uygulamada iki yaygın yöntem vardır:

- Min-Maks ölçekleme: Tüm değerler önceden belirlenmiş bir aralığa (örneğin $[0, 1]$ ya da $[-1, 1]$) ölçeklenir; özellikle mel-spektrogram gibi görüntü temsillerine geçerken pratiktir.
- Z-puan standardizasyonu: Her kaydın ortalaması sıfıra, standart sapması bir e eşitlenir; istatistiksel olarak daha dengeli bir hizalanma sunar.

Hangi yöntem seçilirse seçilsin, normalizasyon parametrelerinin yalnızca eğitim kümesinden hesaplanıp doğrulama ve test kümelerine aynı şekilde uygulanması bilgi sızıntısını önlemek için temel bir kuraldır. Ayrıca, dinamik aralığın çok geniş olduğu durumlarda logaritmik ölçekleme ile en güçlü ve en zayıf bileşenler dengelenerek normalizasyona destek verilebilir (Rocha vd., 2019; Fava, 2024).

2.9.2. Gürültü Giderme ve Filtreleme

Solunum sesi sinyali genellikle düşük genlikli olup istenmeyen gürültüyle kolay karışabilir; bu yüzden gürültünün kontrolü sınıflandırma başarımı açısından kritik önemdedir (Sovijärvi vd., 1996).

2.9.2.1. Band-Pass Filtre

Uygulamada en yaygın adım, enerjinin büyük kısmının bulunduğu yaklaşık 100–2000 Hz aralığını geçiren bir süzgeç uygulanmasıdır. Alt kesim genellikle 80–100 Hz düzeyinde belirlenir (kalp sesleri ve düşük frekanslı hareket bileşenlerini bastırmak için); üst kesim ise yaklaşık 1800–2000 Hz civarında seçilir (yüksek frekanslı ortam gürültüsünü ve konuşmayı seyreltmek için). Elektrik şebekesi kaynaklı 50/60 Hz uğultu varsa dar bantlı bastırma (notch) eklenebilir. Ancak kesim değerlerinin çok agresif seçilmesi, ince raller ya da hışıltı gibi tanısal sinyal bileşenlerini de kırabileceği için filtre ayarları veri setinin spektral dağılımına göre dikkatle yapılmalıdır (Rocha vd., 2019).

2.9.2.2. Wiener Filtre

Bu yöntem, sinyal ve gürültünün istatistiksel özelliklerini kullanarak her frekans bileşeni için uyarlanabilir bir zayıflatma katsayısı hesaplar. Özellikle oda uğultusu gibi sabit arka plan gürültülerinde solunum sesinin anlaşılabilirliğini önemli ölçüde artırabilir. Ancak parametreler aşırı bastırmaya ayarlanırsa dar bantlı hışıltı bileşenleri istenmeden silinebilir; bu yüzden dinleyerek ve görsel incelemeyle eş zamanlı kontrol önerilir (Fava, 2024).

2.9.2.3. Empirical Mode Decomposition (EMD)

Durağan olmayan biyomedikal sinyalleri daha basit salınımlara ayırmak için uygundur. Bazı kiplerin gürültü ağırlıklı olduğu durumlarda bu kipler zayıflatılıp sinyal yeniden birleştirilebilir. Kısa süreli raller ile geniş tabanlı parazitin ayrıştırılmasında avantaj sağlar; ancak hangi kiplerin bırakılıp hangilerinin bastırılacağı kararı öznel ayarlara duyarlıdır ve çapraz doğrulama ile görsel kontrol desteği gerektirir (Demirci vd., 2022).

2.9.3. Segmentasyon Teknikleri

Analizde tüm kaydı kullanmak yerine, soluk alma ve verme evrelerine ait bölümlerin ayıklanması hem klinik hem de sayısal açıdan faydalıdır. Segmentasyon genellikle iki şekilde yapılır:

- Manuel işaretleme: Uzman tarafından soluk fazları işaretlenir, en güvenilir ancak zaman alıcıdır.
- Otomatik yöntemler: Sinyalin zarfı (envelope), enerji profili, sıfır geçiş yoğunluğu veya kısa süreli güç haritaları kullanılarak soluk evreleri otomatik olarak belirlenir.

Örneğin, enerjisi nefes alma sırasında yükselen, nefes verme sırasında azalan bir desen gösteren sinyal, bu değişime göre ayrılabilir. Zaman-frekans düzleminde yüksek enerji kümelerinin takibiyle döngü sınırları saptanabilir. Kalp sesi ve hareketin yoğun olduğu bölümlerde kısa koruma aralıkları bırakmak ve komşu pencere sınırlarında yumuşak geçiş yapmak, yanlış kesimleri azaltır. Segmentasyon doğruluğu hem dinlenerek hem de spektrogram üzerinde görsel incelemeyle kontrol edilmelidir; zira kötü bir segmentasyon, en iyi sınıflandırıcıyı bile yanıltabilir (Rocha vd., 2019; Sovijärvi vd., 1996).

2.9.4. Artefaktların Temizlenmesi

Kayıt sırasında hastanın konuşması, öksürmesi, hareket etmesi, stetoskop başlığının kıyafetle sürtünmesi ya da başlığın yer değiştirmesi gibi durumlar “gerçek” solunum sesine benzemeyen ani tepe ve çukurlara yol açabilir. Bu tür artefaktlar çoğunlukla kısa süreli, geniş bantlı ve düzensiz bir enerji profiline sahiptir. Tespitte iki pratik ölçüt kullanılır:

- Zamansal olarak çok yüksek genlikli tekil sıçramalar
- Frekans alanında “her yere yayılan” düzensiz güç dağılımı

Tanımlanan bölümler ya tamamen dışlanır ya da komşu temiz kısımlardan kısa bir ara değerle doldurulur. Ancak aşırı temizlik, özellikle raller gibi zaten kısa ve patlayıcı yapıda olan tanısız işaretleri yanlışlıkla silme riski taşır. Bu nedenle eşik değerleri veri-odaklı seçilmeli ve şüpheli kısımlar mutlaka dinlenerek kontrol edilmelidir (Fava, 2024).

2.10. Veri Artırma (Data Augmentation) Teknikleri

Solunum sesi sınıflandırma modelleri geliştirirken karşılaşılan temel zorluklardan biri, veri sayısının sınırlı olması ve sınıflar arasındaki dengesizliktir. Özellikle “hem hışıltı hem ral içeren” gibi nadir rastlanan örüntüler, makine öğrenmesi algoritmalarının bu örnekleri yeterince tanımasını zorlaştırır. Bu durumda model, sık görülen sınıflara yönelerek azınlık sınıflarda düşük başarı gösterir.

Veri artırma (data augmentation) teknikleri, mevcut kayıtların yapay fakat gerçeğe yakın biçimlerde türetilmesiyle eğitim veri setini genişletir. Böylece model aynı sınıfın farklı varyasyonlarını tanıma yeteneği kazanır, aşırı öğrenme (overfitting) riski azalır ve genelleme gücü artar (Goodfellow vd., 2016; Rocha vd., 2019).

Veri artırma yalnızca eğitim kümesine uygulanmalı; doğrulama ve test kümeleri orijinal hâliyle korunmalıdır. Ayrıca hasta bazında ayırım yapılması gereken durumlarda,

aynı kişiye ait kayıtların hem eğitim hem test kümesinde yer almaması sağlanmalı, artırılmış veriler de bu kurala uygun şekilde gruplanmalıdır.

Yaygın kullanılan veri artırma yöntemleri, avantajları ve sınırlılıkları ile birlikte Tablo 2'de özetlenmiştir. Bu yöntemler arasında zaman kaydırma, gürültü enjeksiyonu, ton değiştirme, zaman ölçekleme ve SMOTE gibi istatistiksel yöntemlerin yanı sıra, zaman-frekans maskelenmesi ve jeneratif modellerle sentetik örnek üretimi gibi daha gelişmiş yaklaşımlar da yer almaktadır. Her bir yöntem, sinyalin farklı yönlerini hedef alarak çeşitli varyasyonlar sunar; dolayısıyla kullanım bağlamına uygun seçim yapmak büyük önem taşır.

2.10.1. Zaman Kaydırma

Zaman kaydırma, solunum sinyalinin tamamını zaman eksenini boyunca az miktarda ileri veya geri hareket ettirerek yeni örnekler oluşturur. Dalga biçiminin karakteristik yapısı korunur ancak olayların başlangıç noktası değişmiş olur. Tipik kaydırma miktarı, döngü süresinin %5–15'i ya da sabit 50–250 ms aralığında seçilir. Zaman kaydırma özellikle kısa süreli hışıltı veya ral gibi sinyallerin pozisyonunu değiştirmeksizin varyasyon yaratmak için etkilidir. Ancak kaydırma sonucunda soluk döngüleri arasında taşma olmamasına ve etiketlerin uygun şekilde güncellenmesine dikkat edilmelidir (Rocha vd., 2019).

2.10.2. Gürültü Enjeksiyonu

Eğitim verisine düşük seviyede gürültü eklenerek modelin gerçek dünya koşullarında daha dayanıklı hâle gelmesi sağlanabilir. Beyaz, pembe veya klinik ortamdan alınmış gerçek arka plan gürültüleri kullanılabilir. Gürültü seviyesi, sinyal-gürültü oranı (SNR) kullanılarak kontrol edilir; 5–20 dB aralığı genellikle klinik senaryoları kapsar. Düşük SNR değerleri zorlu ortamları, yüksek SNR değerleri ise sessiz klinik koşulları simüle eder. Ancak çok yüksek gürültü düzeyleri tanısal önemi olan ince raller gibi düşük genlikli yapıları örtebilir (Rocha vd., 2019).

2.10.3. Ton Değiştirme (Pitch Shifting)

Ton değiştirme, sinyalin frekans içeriğini yukarı veya aşağı kaydırarak yaş, cinsiyet ve göğüs duvarı yapısı gibi fizyolojik varyasyonları taklit eder. ± 0.5 –2 yarım seslik bir değişim genellikle yeterli olur. Ancak büyük kaydırmalar sesin klinik karakterini değiştirebilir; örneğin veziküler sesi bronşiyal sese dönüştürebilir. Bu nedenle formant

koruyucu algoritmalarla ve dikkatli görsel–işitsel kontrolle uygulanmalıdır (Goodfellow vd., 2016).

2.10.4. Zaman Ölçekleme (Time Stretching)

Zaman ölçekleme, sesin hızını artırarak veya azaltarak yeni varyasyonlar üretir. %5–15 oranındaki değişiklikler, klinik olarak makul solunum hızı varyasyonlarını yansıtır. Faz-vokoder temelli yöntemler zaman uzunluğunu değiştirirken frekansı sabit tuttuğu için tercih edilir. Ancak aşırı yavaşlatma veya hızlandırma, kısa patlamaları yapay biçimde uzatabilir veya bastırabilir. Bu nedenle işlem sonrası görsel ve işitsel kontrol önerilir (Rocha vd., 2019).

2.10.5. SMOTE ve Yapay Örnek Üretimi

Sınıf dengesizliğini azaltmak amacıyla kullanılan SMOTE (Synthetic Minority Oversampling Technique), azınlık sınıfındaki örnekler arasında yeni, ara-değerli vektörler üretir. Ancak bu teknik ham ses sinyali yerine MFCC, Chroma veya Spectral Contrast gibi öznel vektörleri üzerinde uygulanmalıdır. SMOTE uygulanan veri setlerinde bireye özgü örneklerin test setine sızmasını önlemek gerekir; aksi takdirde modelin genel başarımı olduğundan fazla görünebilir. Alternatif olarak ADASYN veya SMOTE gibi teknikler de kullanılabilir (Chawla vd., 2002; He vd., 2008).

2.10.6. Zaman ve Frekans Maskelenmesi

Zaman ve frekans maskelenmesi, spektrogram üzerinde belirli zaman veya frekans bantlarının rastgele kapatılması esasına dayanır. Bu sayede modelin belirli bir frekans bileşenine veya zaman aralığına aşırı bağımlı hale gelmesi engellenir. Özellikle hışıltı gibi dar bantlı sinyallerin yer değiştirmesi, modelin daha genellenebilir özellikler öğrenmesini sağlar (Park vd., 2019). Ayrıca, küçük varyasyonlara karşı toleransı artırır.

2.10.7. Ses Seviyesi Ayarı ve Kazanç Değişimi

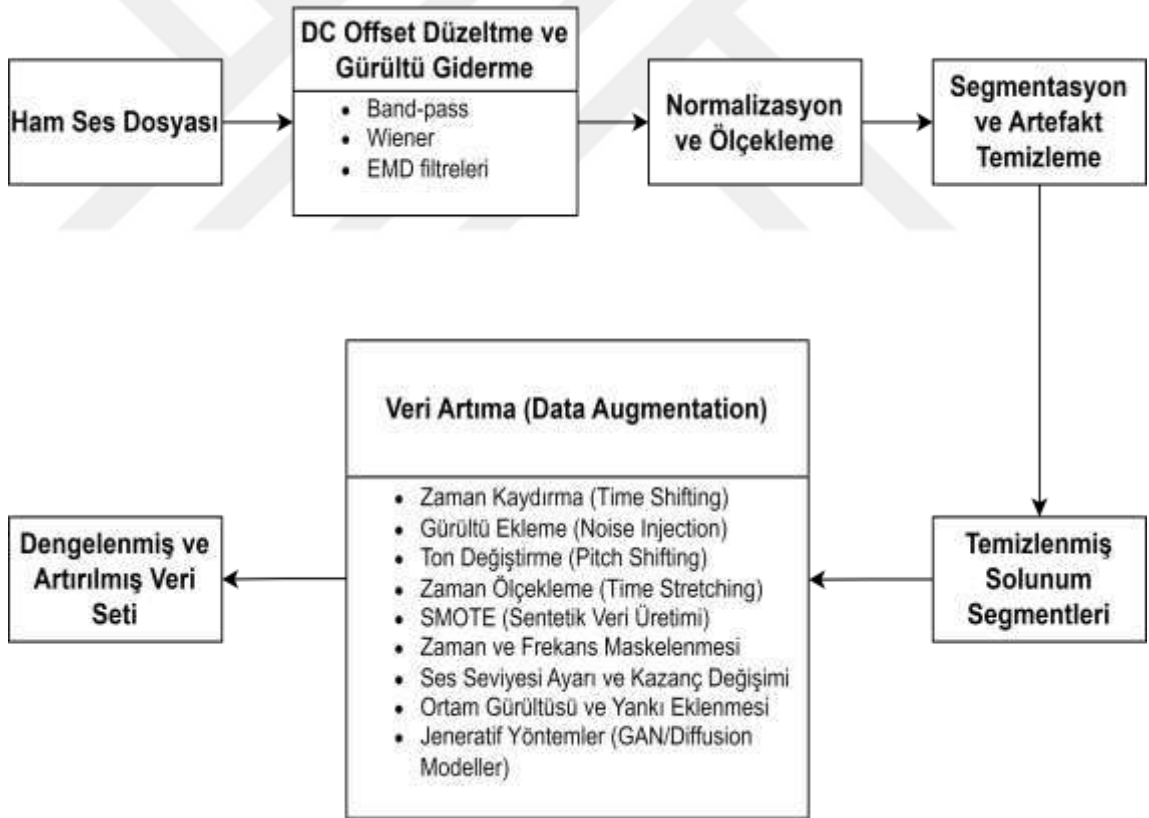
Sesin genliğinin artırılması veya azaltılması, farklı kayıt cihazlarının duyarlılığı ya da hastayla mikrofon arasındaki mesafenin taklit edilmesine olanak tanır. Bu yöntem, sesin yapısını bozmadan modelin farklı genliklere alışmasını sağlar (Salamon ve Bello, 2017). Ancak genlik artışının distorsiyona neden olmaması için dikkatli uygulanmalıdır.

2.10.8. Ortam Gürültüsü ve Yankı Eklenmesi

Mekân yansıması, yankı (reverberation) ve oda tepki fonksiyonları (Room Impulse Response – RIR) ile veri setine akustik çeşitlilik kazandırılabilir. Böylece model yalnızca sessiz klinik ortamda değil, ev veya gürültülü ortamlarda da dayanıklı performans gösterir (Ko vd., 2017). RIR uygulamaları özellikle konuşma tanıma sistemlerinde yaygın olmakla birlikte, klinik seslerde de anlamlı katkı sağlayabilir.

2.10.9. Jeneratif Yöntemler (GAN / Diffusion Models)

Veri sayısının çok az olduğu durumlarda, Generative Adversarial Networks (GAN) ya da Diffusion Models gibi yöntemlerle yapay ama gerçeğe yakın yeni örnekler üretilebilir. Bu teknikler solunum sesi gibi karmaşık biyolojik sinyallerde, varyasyonları zenginleştirerek eğitim veri kümesini güçlendirir (Kim vd., 2023). Ancak klinik kullanıma girmeden önce uzman onayı ve etik kontrol büyük önem taşır.



Şekil 9. Solunum sesi ön işleme ve veri artırma sürecinin genel akış diyagramı

Tablo 2. Solunum sesi veri artırma yöntemlerinin karşılaştırılması

Yöntem	Güçlü Yönleri	Sınırlılıkları / Dikkat Edilmesi Gerekenler
Zaman Kaydırma	Basit ve etkili, Raller ve hışıltı gibi kısa olayları bozmadan varyasyon sağlar	Etiket güncellemesi gerekebilir Büyük kaydırmalar zaman yapısını bozabilir
Gürültü Enjeksiyonu	Gerçek dünya dayanıklılığını artırır, Klinik ortam koşullarını taklit eder	Fazla gürültü, tanısal ipuçlarını bastırabilir
Ton Değiştirme	Fizyolojik varyasyonları (yaş, cinsiyet) taklit eder	Aşırı kaydırma, sesin klinik karakterini değiştirebilir
Zaman Ölçekleme	Farklı solunum hızlarını simüle eder, Faz-vokoder algoritmaları ile kaliteli sonuç verir	Aşırı yavaşlatma kısa olayları çarpıtabilir
SMOTE / ADASYN	Düşük temsil edilen sınıfları sayısal olarak dengeler	Ham ses yerine öznitelik düzeyinde uygulanmalıdır Aşırıya kaçılırsa yapay örnekler “anlamsız” olabilir
Zaman / Frekans Maskelenmesi	Modelin belirli bölgelere aşırı bağımlı olmasını önler	Maskleme genişliği dikkatle seçilmeli; bilgi kaybı yaratabilir
Ses Seviyesi Ayarı	Mikrofon uzaklığı / cihaz farklılıklarını taklit eder	Genlik distorsiyonuna dikkat edilmeli
Yankı / Oda Tepkisi (RIR)	Gerçek ortamlardaki yankıyı simüle eder	Uygun RIR verisi gerektirir; aşırı yankı ses kalitesini bozabilir
Jeneratif Modeller (GAN vb.)	Gerçeğe yakın yeni örnekler üretir Az veriyle çalışmalarda büyük destek sağlar	Karmaşık ve maliyetli- Klinik geçerlilik dikkatle denetlenmeli

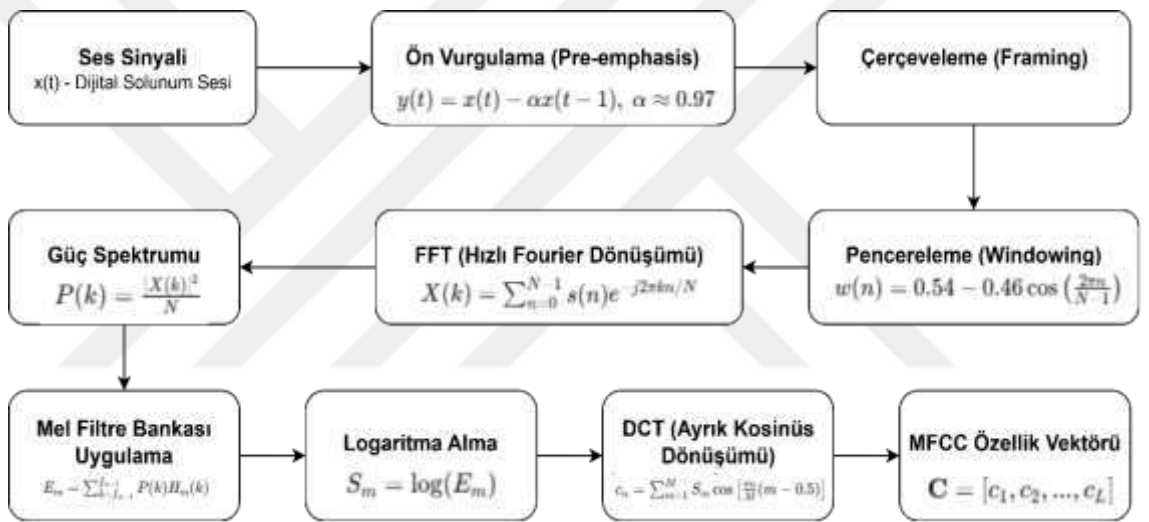
2.11. Öznitelik Çıkarımı Yöntemleri

2.11.1. Mel Frekans Kepstrum Katsayıları (MFCC)

Öznitelik çıkarımı, ham ses dalgasının doğrudan sınıflandırıcıya verilmesi yerine, tanı açısından anlamlı özelliklerle temsil edilmesi sürecidir. Böylece model, tüm sinyale değil, ayırt edici ipuçlarına odaklanabilir. Solunum sesi özelinde bu ipuçları çoğu kez iki

grupta değerlendirilir: zamana dayalı örüntüler (sinyalin süresi, genlik değişimi, tepe yapılarının sıklığı gibi) ve frekansa dayalı örüntüler (enerji yoğunluğunun frekans dağılımı gibi). Klinik doğruluğu artırmak için zaman ve frekans bilgilerini birlikte barındıran temsiller sıklıkla tercih edilir (Rocha vd., 2017).

Bu bağlamda en yaygın kullanılan yöntemlerden biri Mel Frekans Kepstrum Katsayılarıdır (MFCC). İnsan kulağının frekansı logaritmik algılayışına benzer bir ölçüm temeline dayanır ve solunum gibi biyolojik sinyallerde etkili bir şekilde kullanılır. MFCC çıkarımı sürecinde; ses sinyali ön vurgulama, pencereleme, Fourier dönüşümü, mel filtreleme ve logaritmik dönüşüm adımlarından geçerek, son olarak Ayrık Kosinüs Dönüşümü (DCT) uygulanarak öznitelikler elde edilir. MFCC çıkarım sürecine dair adımlar Şekil 10’da şematik olarak gösterilmektedir.



Şekil 10. MFCC çıkarım sürecinin şematik gösterimi

Şekil 10’da yer alan eşitliklerde:

$x(t)$: Zamana bağlı dijital solunum sesi sinyalini,

$y(t)$: Ön vurgulama (pre-emphasis) uygulanmış sinyali,

α : Ön vurgulama katsayısını (genellikle 0.95–0.97 aralığında),

$w(n)$: Pencereleme (windowing) fonksiyonunu,

$X(k)$: Hızlı Fourier Dönüşümü (FFT) sonucunda elde edilen frekans bileşenlerini,

$P(k)$: Güç spektrumunu,

$H_m(k)$: Mel filtre bankasındaki m ’inci filtreyi,

E_m : m ’inci mel filtresine ait enerji değerini,

S_m : Logaritmik enerji değerini,

c_n : Ayrık Kosinüs Dönüşümü (DCT) sonucu elde edilen n ’inci MFCC katsayısını,

L : Kullanılan toplam MFCC katsayısı sayısını ifade etmektedir.

2.11.2. Mel-spektrogram

Mel-spektrogram, kısa zaman dilimlerinde sinyalin mel-frekans eksenindeki enerji dağılımını sunar. Zaman ve frekans bilgisini birlikte içeren bu görsel temsil, evrimsel sinir ağlarında doğrudan girdi olarak kullanılır. Hışıltı, dar bant çizgiler; raller ise parlak adacıklar olarak görünür. dB dönüşümü ve dinamik aralık sıkıştırması, model performansını artırır (Rocha vd., 2017).

2.11.3. Kısa Zamanlı Fourier Dönüşümü (STFT)

STFT, her pencere için sinyalin frekans bileşenlerini hesaplar. Zamansal çözünürlük ve frekans çözünürlüğü arasında ödünleşim vardır: kısa pencereler zaman çözünürlüğünü artırırken frekans çözünürlüğünü azaltır. Solunum sesleri için Hamming penceresi, %50–75 bindirme oranı ve 4–8 kHz örnekleme hızı önerilir (Sovijärvi vd., 2000).

2.11.4. Dalgacık Dönüşümleri (Wavelet)

Dalgacık dönüşümü, farklı zaman ölçeklerinde analiz yaparak hem kısa hem uzun süreli örüntüleri yakalar. Ayırık dalgacık dönüşümü (DWT) enerji ve entropi gibi ölçümleri çıkarırken, sürekli dalgacık dönüşümü (CWT) zaman-ölçek düzleminde sinyalin detaylarını sunar. Ana dalgacık seçimi ve ayrıştırma seviyesi, hedef örüntülere göre belirlenmelidir (Sovijärvi vd., 2000).

2.11.5. İstatistiksel Özellikler

Temel zaman ve frekans istatistikleri hâlen etkili özniteliklerdir. Ortalama genlik, varyans, çarpıklık (skewness), basıklık (kurtosis), sıfır geçiş oranı (ZCR), spektral ağırlık merkezi (centroid), bant genişliği ve düzlük (flatness) gibi öznitelikler frekans yoğunluğunun konumu ve dağılımı hakkında bilgi verir. Bu öznitelikler tek başına sınırlı olsa da MFCC ve dalgacık gibi temsillerle birleştirildiğinde sınıflandırma performansını artırabilir (Rocha vd., 2017).

2.11.6. Chroma Özellikleri

Chroma temelli temsiller, frekans bileşenlerini 12 yarım seslik sınıfa indirger. Müzikal ses analizi kökenli olsa da bazı solunum seslerinde ton temelinde örüntülerin ayrımını kolaylaştırabilir. Ancak klinik solunum sinyalleri gibi geniş spektrumlu ve düşük tonlu sinyallerde dikkatli kullanılmalıdır (Tzanetakis ve Cook, 2002).

2.11.7. Spektral Kontrast

Spektral kontrast, spektrumun zirve ve çukurları arasındaki farkı ölçerek ton ve gürültü ayrımı yapmaya yardımcı olur. Solunum sesi gibi frekans yapısı değişken sinyallerde, spektral kontrast hışıltı ve raller gibi örüntülerin ayrımında etkili olabilir (Jiang vd., 2002).

Tablo 3. Yaygın öznelite çıkarm yöntemlerinin karşılaştırılması

Yöntem	Zaman Bilgisi	Frekans Bilgisi	Kullanım Alanı	Avantaj	Dezavantaj
MFCC	Orta	Orta–Yüksek	Klasik makine öğrenmesi	İnsan işitmesine benzer yapı, yaygın kullanım	Gürültüye duyarlı, kısa pencerelerde bilgi kaybı olabilir
Mel-Spektrogram	Yüksek	Yüksek	Derin öğrenme (CNN, RNN)	Görsel temsile uygun, zaman–frekans çözümlemesi sağlar	Büyük veri ve yüksek hesaplama gücü gerektirir
STFT	Orta	Orta	Temel sinyal analizi	Basit, yaygın araçlarla uyumlu	Sabit pencere boyutu nedeniyle sınırlı çözünürlük
Dalgacık	Yüksek	Yüksek	Patolojik ses analizi	Çok ölçekli analiz, kısa süreli olayları yakalar	Hesaplama maliyeti yüksek, parametrik hassasiyet taşır
İstatistiksel	Düşük	Düşük	Hızlı prototipleme	Hesaplama açısından hafif, yorumlanabilir	Frekans bilgisini doğrudan taşımaz, tek başına sınırlı olabilir
Chroma	Düşük–Orta	Orta	Ton analizi, özel durumlar	Ton temelli ayırt edici özellikler	Solunum sesinde sınırlı başarı, geniş spektrumlu sinyallerde düşük ayırım
Spektral Kontrast	Orta	Orta–Yüksek	Ton/gürültü ayrımı	Hışıltı ve raller ayrımında etkili, frekans yapısını iyi yansıtır	Gürültüde bozulabilir, yorumlama için görsel analiz gerekebilir

2.12. Klasik Makine Öğrenmesi Yaklaşımları

Klasik makine öğrenmesi yaklaşımında işlem hattı üç temel aşamadan oluşur. İlk olarak, kayıtlar temizlenir, parazitler giderilir ve nefes alma–verme döngüleri ayrıştırılır. İkinci aşamada, her döngüden anlamlı öznitelikler (features) çıkarılır — örneğin Mel Frekans Kepstrum Katsayıları (MFCC), dalgacık enerjisi, sıfır geçiş oranı, spektral ağırlık merkezi ve benzeri göstergeler. Son aşamada ise bu öznitelikler, seçilen bir sınıflandırma algoritmasına girdi olarak verilerek etiket (örneğin “normal”, “hışıltılı”, “ral”) tahmini yapılır.

Model başarımını belirleyen iki temel unsur vardır:

- Çıkarılan özniteliklerin gerçekten ayırt edici olması,
- Hasta-bağımsız bir doğrulama düzeninin (örneğin hasta bazlı çapraz doğrulama) izlenmesidir.

Farklı kayıt cihazları, mikrofon yerleşimleri veya hastalar arası biyolojik değişkenlik, modelin “kolay ama aldatıcı” ilişkiler öğrenmesine neden olabilir. Bu nedenle aynı kişiye ait döngülerin eğitim ve test kümesinde karışmaması büyük önem taşır (Rocha vd., 2017; Bishop, 2006; Hastie vd., 2009).

2.12.1. Destek Vektör Makineleri (SVM)

Destek Vektör Makineleri (Support Vector Machines – SVM), farklı sınıfları ayıran en geniş marjine sahip karar sınırını (hiper düzlemi) bulmayı amaçlar. Doğrusal olarak ayrılmayan verilerde “çekirdek (kernel)” işlevleri kullanılarak veriler daha yüksek boyutlu bir uzaya taşınır ve bu uzayda doğrusal bir ayırım yapılabilir (Cortes ve Vapnik, 1995).

Solunum sesi sınıflandırmalarında özellikle RBF (Radial Basis Function) çekirdeği tercih edilir; MFCC, dalgacık enerjileri veya spektral istatistiklerden türetilen özniteliklerle oldukça yüksek doğruluk sağlar (Bishop, 2006).

C (ceza katsayısı) ve gamma parametrelerinin uygun seçilmesi, modelin hem duyarlılığını hem özgüllüğünü etkiler. Küçük veri setlerinde genellikle yüksek performans gösterirken, çok büyük veri kümelerinde eğitim süresi uzayabilir (Hastie vd., 2009).

2.12.2. k-En Yakın Komşu (KNN)

K-En Yakın Komşu algoritması, yeni bir örneğin sınıfını en yakın k komşusunun etiketlerine göre belirler (Cover ve Hart, 1967). Mesafe ölçüsü olarak genellikle öklit veya kosinüs benzerliği kullanılır.

Solunum sesi analizinde KNN'in avantajı, model varsayımı gerektirmemesi ve küçük veri setlerinde hızlıca uygulanabilmesidir. Ancak yüksek boyutlu özniteliklerde (örneğin 40+ MFCC katsayısı) “boyutsallık laneti” etkisiyle performans düşebilir. Gürültülü veya dengesiz veri kümelerinde, k ve mesafe metriği özenle seçilmelidir (Hastie vd., 2009).

2.12.3. Rastgele Ormanlar (RF)

Rastgele Orman (Random Forest), çok sayıda karar ağacının bir araya gelerek (bagging yöntemiyle) oluşturduğu güçlü bir topluluk (ensemble) modelidir (Breiman, 2001).

Her ağaç, eğitim verisinin farklı bir alt kümesiyle eğitilir; sonuçta bu ağaçların ortalaması veya çoğunluk oyu alınır. Bu yöntem, aşırı öğrenmeye (overfitting) karşı dayanıklıdır.

Solunum sesi sınıflandırmasında MFCC ve dalgacık enerjisi gibi heterojen öznitelikler birlikte kullanılabilir.

Yorumlanabilirliği tek bir ağaca göre zayıf olsa da, model kararlılığı ve genel doğruluğu oldukça yüksektir (Hastie vd., 2009).

2.12.4. XGBoost

XGBoost (Extreme Gradient Boosting), art arda eğitilen küçük karar ağaçlarının, önceki hataları düzelterek oluşturduğu gradyan artırma bir topluluk algoritmasıdır (Friedman, 2001; Chen ve Guestrin, 2016).

Solunum sesi verilerinde özellikle dengesiz sınıf dağılımlarında başarılıdır; azınlık sınıflara (örneğin yalnızca ral içeren kayıtlar) daha fazla ağırlık verebilir.

Parametreleri arasında öğrenme hızı (η), maksimum derinlik, alt örnekleme oranı ve L1/L2 düzenleme katsayıları yer alır. Hiperparametre ayarı grid search veya Bayesian optimization ile yapılabilir.

Yüksek doğruluk sağlamasına karşın, karmaşık yapısı nedeniyle yorumlanabilirlik sınırlıdır.

2.12.5. Naive Bayes

Naive Bayes, Bayes teoremini temel alır ve öznitelikler arasındaki koşullu bağımsızlık varsayımıyla olasılıksal sınıflandırma yapar (Murphy, 2012).

Solunum sesleri gibi karmaşık sinyallerde bu varsayım tam olarak sağlanmasa da, küçük veri setlerinde ve yüksek boyutlarda sağlam bir temel karşılaştırma modeli oluşturur.

Gaussian Naive Bayes versiyonu, MFCC veya dalgacık katsayılarının dağılımını normal varsayarak hızlı ve açıklanabilir sonuçlar üretir (Bishop, 2006).

2.12.6. Doğrusal Ayırıcı Analiz (LDA)

LDA (Linear Discriminant Analysis), sınıflar arası farkı maksimize ederken, sınıf içi varyansı minimize etmeye çalışır (Fisher, 1936).

Bu nedenle hem bir boyut indirgeme tekniği hem de sınıflandırıcı olarak kullanılabilir.

Solunum seslerinde MFCC ve spektral istatistikler gibi çok boyutlu özniteliklerin 2–6 boyuta indirgenmesi, daha kompakt ve ayırt edici bir temsil sunar.

Ancak doğrusal sınır varsayımı nedeniyle, hışıltı gibi doğrusal olmayan örüntülerde yetersiz kalabilir (Hastie vd., 2009).

2.12.7. AdaBoost

AdaBoost (Adaptive Boosting), zayıf sınıflandırıcıları (örneğin karar ağaçları) art arda eğitip, yanlış sınıflandırılan örneklere daha fazla ağırlık vererek güçlü bir birleşik model üretir (Freund ve Schapire, 1997).

Solunum seslerinde özellikle azınlık sınıfların (örneğin sadece hışıltı içeren kayıtlar) yakalanmasında etkilidir. Ancak etiket hatalarına karşı oldukça hassastır ve aşırı uyuma eğilimli olabilir (Hastie vd., 2009).

2.12.8. Karar Ağaçları (Decision Tree)

Karar ağaçları, veriyi öznitelik değerlerine göre dallara ayırarak tahmin yapan sezgisel ve yorumlanabilir modellerdir (Breiman vd., 1984; Quinlan, 1993).

Her düğümde en yüksek bilgi kazancını sağlayan öznitelik seçilir ve verinin farklı sınıflara ayrılması hedeflenir.

Solunum seslerinde MFCC, dalgacık enerjisi, spektral ağırlık merkezi ve sıfır geçiş oranı gibi karışık türde öznitelikler kolayca kullanılabilir.

Aşırı derin ağaçlar overfitting riski taşır; bu nedenle budama (pruning) işlemi önerilir.

Karar ağaçları, yorumlanabilirlik ve hız açısından avantajlı olup genellikle RF, AdaBoost veya XGBoost gibi topluluk yöntemlerinin temel bileşenidir (Hastie vd., 2009).

Tablo 4. Klasik makine öğrenmesi algoritmalarının güçlü ve zayıf yönleri

Algoritma	Güçlü Yönleri	Zayıf Yönleri
SVM	Küçük veri setlerinde yüksek doğruluk	Büyük veri setlerinde yavaş, parametreye duyarlı
KNN	Basit, parametrik olmayan	Gürültüye duyarlı, yüksek hesaplama maliyeti
RF	Aşırı öğrenmeye dayanıklı, kararlı	Yorumlanabilirliği sınırlı
XGBoost	Dengesiz veri setlerinde etkili	Parametre seçimine duyarlı
Naive Bayes	Hızlı, düşük hesaplama maliyeti	Özellik bağımsızlığı varsayımı zayıf
LDA	Boyut indirgeme sağlar	Doğrusal olmayan yapılarda yetersiz
AdaBoost	Azınlık sınıflarda başarılı	Gürültüye duyarlı, aşırı öğrenme riski
Decision Tree	Kolay yorumlanabilir	Aşırı öğrenmeye yatkın

2.13. Derin Öğrenme Yaklaşımları

Klasik makine öğrenmesi algoritmalarının aksine, derin öğrenme yaklaşımlarında öz nitelik çıkarımı süreci büyük oranda modelin kendisi tarafından gerçekleştirilir. Bu sayede insan müdahalesine gerek kalmadan, verinin iç yapısından anlamlı temsiller otomatik olarak öğrenilebilir.

Solunum sesi analizinde bu tür modeller, genellikle Mel-spektrogram gibi zaman-frekans düzlemindeki temsiller üzerinde uygulanır. Bu temsiller, görüntü verilerine benzedikleri için, görüntü işleme alanında başarılı olmuş derin öğrenme mimarileri doğrudan veya transfer öğrenme yöntemiyle solunum sesi sınıflandırmalarına adapte edilebilir.

Ancak derin öğrenme algoritmaları genellikle büyük miktarda veri gerektirdiğinden, veri artırma, transfer öğrenme, dropout, erken durdurma gibi düzenleme stratejileri yaygın biçimde kullanılmaktadır (Goodfellow vd., 2016; LeCun vd., 2015).

2.13.1. Uzamsal Tabanlı Modeller (CNN ve Türevleri)

CNN, Mel-spektrogramlar üzerindeki yerel desenleri algılayarak sınıflandırma yapan temel derin öğrenme mimarileridir. Solunum sesi örüntüleri olan hışıltı (yüksek frekansta devamlı çizgiler) ve ral (kısa süreli, ani parlamalar) gibi bileşenlerin tespitinde etkilidirler.

CNN'ler, küçük çekirdek filtreleri kullanarak spektrogramın farklı bölgelerinde aynı desenleri arar. Havuzlama katmanları ise bu bilgileri özetleyerek modelin genelleme kabiliyetini artırır.

Uygulamada, genellikle 4–6 evrişim bloğunun ardından küresel ortalama havuzlama (global average pooling) ve tam bağlantılı (fully connected) katmanlar kullanılır. Bu yapı, parametre sayısını azaltarak aşırı öğrenmeyi (overfitting) sınırlamaya yardımcı olur (LeCun vd., 2015; Rocha vd., 2017).

2.13.1.1. GhostNet

GhostNet, derin sinir ağlarının hesaplama yükünü azaltmak amacıyla geliştirilmiş bir Evrişimsel Sinir Ağı (CNN) türevidir.

Modelin temel yaklaşımı, her filtreleme işlemi için yeni öznitelikler üretmek yerine, düşük maliyetli doğrusal işlemler aracılığıyla bu özniteliklerin “hayalet (ghost)” kopyalarını oluşturmaktır (Han vd., 2020).

Solunum sesi sınıflandırma uygulamalarında GhostNet, özellikle mobil stetoskop cihazları, sahada çalışan tanı sistemleri ve enerji kısıtlı donanımlar için uygun bir seçenek olarak öne çıkmaktadır.

Ghost blokları, Mel-spektrogram temsilleri üzerinde minimum bilgi kaybıyla ve düşük parametre sayısı ile hafif öznitelik temsilleri üretebilmektedir (Han vd., 2020).

2.13.1.2. ResNet

ResNet (Artık Ağ, Residual Network), çok katmanlı derin sinir ağlarında eğitim sürecini kararlı hâle getirmek amacıyla geliştirilen bir CNN mimarisidir.

Modelin temel özelliği, katmanlar arasında artık bağlantılar (skip connections) kullanmasıdır. Bu bağlantılar sayesinde, her katman gerekirse girdiyi hiçbir dönüşüm uygulamadan doğrudan bir sonraki katmana aktarabilir.

Bu yapı, gradyan sönümlenmesi (gradient vanishing) gibi derin ağlarda görülen yaygın problemleri azaltır ve daha derin mimarilerin etkin biçimde eğitilebilmesini mümkün kılar (He vd., 2016).

Solunum sesi sınıflandırmalarında, özellikle Mel-spektrogram temelli girişler üzerinde ResNet mimarisinin 34 veya 50 katmanlı önceden eğitilmiş sürümleri (ResNet-34, ResNet-50) sıkça tercih edilmektedir.

Bu uygulamalarda genellikle ağın erken katmanları sabit tutulur (frozen); yalnızca son katmanlar solunum sesi verilerine özel olarak yeniden eğitilir (fine-tuning) (He vd., 2016; Zhang vd., 2020).

2.13.1.3. EfficientNet

EfficientNet, derin öğrenme modellerinde derinlik (depth), genişlik (width) ve giriş çözünürlüğü (resolution) gibi boyutları dengeli bir şekilde ölçekleyerek performansı artırmayı hedefleyen optimize bir Evrişimsel Sinir Ağı (CNN) mimarisidir (Tan ve Le, 2019).

Bu dengeli ölçekleme stratejisi sayesinde, geleneksel modellere kıyasla çok daha az parametreyle daha yüksek doğruluk elde etmek mümkün olmaktadır.

Solunum sesi sınıflandırmalarında, özellikle Mel-spektrogram çözünürlüğü adım adım artırılarak EfficientNet-B0 veya EfficientNet-B1 gibi daha hafif sürümlerle başlanması önerilmektedir.

Bu sürümler, düşük hesaplama maliyetine rağmen yüksek sınıflandırma başarımı sunmaları nedeniyle mobil cihazlar ve sınırlı kaynaklı sistemler için uygun seçeneklerdir (Rocha vd., 2017; Tan ve Le, 2019).

2.13.1.4. MobileNet

MobileNet, hesaplama yükünü azaltmak amacıyla geliştirilen hafif bir CNN mimarisidir. Model, geleneksel evrişim işlemi yerine derinlik-ayrık evrişim (depthwise separable convolution) yaklaşımını kullanarak parametre sayısını ve işlem maliyetini önemli ölçüde düşürmektedir (Howard vd., 2017).

Bu yapısı sayesinde MobileNet, düşük güçlü donanımlarda dahi gerçek zamanlı sınıflandırma yeteneği sunmakta ve mobil cihazlarda yaygın olarak kullanılmaktadır.

Solunum sesi sınıflandırma uygulamalarında, özellikle taşınabilir teşhis sistemleri ve kenar bilişim (edge computing) senaryolarında tercih edilmektedir.

Ayrıca model, 8-bit nicemleme (quantization) ve budama (pruning) gibi tekniklerle daha da hafifletilerek bellek ve işlem gücü açısından optimize edilebilmektedir (Howard vd., 2017).

2.13.1.5. DenseNet

DenseNet (Densely Connected Convolutional Networks), her katmanın tüm önceki katmanlardan gelen özellik haritalarını giriş olarak kullanmasıyla öne çıkar (Huang vd., 2017).

Bu yaklaşım, bilgi kaybını önler, gradyan akışını iyileştirir ve parametre verimliliğini artırır.

Solunum sesi sınıflandırmalarında veri azlığı problemlerinde avantaj sağlar.

2.13.1.6. InceptionNet

Inception mimarisi, farklı boyutlardaki filtreleri (1x1, 3x3, 5x5) aynı anda çalıştırarak çok ölçekli (multi-scale) bilgi çıkarmayı hedefler (Szegedy vd., 2015).

Solunum seslerinin hem kısa hem uzun süreli bileşenleri yakalanmak istendiğinde özellikle etkilidir.

2.13.1.7. SqueezeNet

SqueezeNet, parametre sayısını önemli ölçüde azaltmak için “sıkıştırma (squeeze)” ve “genişletme (expand)” blokları kullanır (Iandola vd., 2016).

MobileNet gibi gömülü sistemlerde veya mikrodenetleyicilerde kullanılmak üzere uygundur.

Solunum seslerinin uzaktan izlem sistemlerine entegre edilmesinde öne çıkar.

2.13.1.8. ShuffleNet

ShuffleNet, grup evrişimi (group convolution) ve kanal karıştırma (channel shuffling) teknikleriyle hesaplamayı azaltırken doğruluğu korumayı hedefler (Zhang vd., 2018).

Özellikle mobil tabanlı sağlık uygulamaları için uygundur. GhostNet’e benzer bir sınıfa girer ama veri dağılımında daha dengeli performans sağlar.

2.13.1.9. CBAM / SE Attention Bloklarıyla Donatılmış CNN’ler

CNN tabanlı mimarilere Channel Attention, Spatial Attention veya Squeeze and Excitation (SE) blokları entegre edilerek modelin hangi özelliklere odaklanacağı iyileştirilebilir (Woo vd., 2018; Hu vd., 2018).

Özellikle ral vs. hışıltı gibi benzer örüntüleri ayırmada performansı artırır.

Bu yapılar CBAM (Convolutional Block Attention Module) gibi hazır modüllerle uygulanabilir.

2.13.2. Zamana Duyarlı Modeller

Solunum sesleri, sadece belirli bir anda değil, zaman boyunca değişen yapılar barındırır. Bu nedenle zaman eksenindeki örüntüleri modelleyebilen zamana duyarlı derin öğrenme mimarileri, solunum sesi analizinde önemli bir rol oynamaktadır.

Bu mimariler, özellikle zamansal bağımlılıkların öğrenilmesi gereken durumlarda CNN tabanlı uzamsal modellerle birlikte veya onların yerine tercih edilmektedir.

2.13.2.1. Tekrarlayan Sinir Ağları (RNN) ve LSTM

Tekrarlayan Sinir Ağları (Recurrent Neural Networks, RNN), sıralı verilerdeki geçici ilişkileri öğrenebilen yapılardır. Ancak geleneksel RNN'ler, uzun süreli bağımlılıkları öğrenmede gradyan sönümlenmesi (vanishing gradient) gibi problemler yaşar.

Bu nedenle Uzun Kısa Süreli Bellek (Long Short Term Memory, LSTM) yapıları geliştirilmiştir (Hochreiter ve Schmidhuber, 1997). LSTM hücreleri, bilgiyi uzun süre saklayabilme yeteneği sayesinde solunum sesi döngülerindeki zamanla değişen özellikleri başarıyla yakalayabilir.

Solunum sesi sınıflandırma uygulamalarında yaygın olarak çift yönlü LSTM (Bidirectional LSTM, Bi-LSTM) yapıları tercih edilmektedir. Bu yapılar, bir ses örneğinin hem öncesini hem de sonrasını dikkate alarak sınıflandırma doğruluğunu artırmaktadır.

2.13.2.2. Dikkat Mekanizmaları (Attention Mechanisms)

Dikkat mekanizmaları (attention mechanisms), modelin uzun bir giriş dizisi içinde hangi bölgelere odaklanması gerektiğini öğrenmesini sağlar. Bu yaklaşım, özellikle önemli zaman dilimlerinin ve frekans-zaman bölgelerinin belirlenmesinde ve modelin yorumlanabilirliğinin artırılmasında kritik bir rol oynar (Bahdanau vd., 2014).

Spektral temsiller üzerine uygulanan dikkat mekanizmaları, modelin sınıflandırmada etkili olan frekans-zaman örüntülerine yoğunlaşmasını sağlar. Bu özellikle ral, wheeze gibi kısa süreli ya da düşük genlikli solunum seslerinin tanınmasında fayda sağlamaktadır.

2.13.2.3. Tipik Kombinasyonlar ve Uygulamalar

Zamana duyarlı modeller genellikle uzamsal öznitelik çıkarımı yapan CNN yapıları ile entegre edilir.

Literatürde sıklıkla karşılaşılan mimari kombinasyonlar şunlardır:

- CNN + LSTM / Bi-LSTM
- CNN + Attention
- CNN + Bi-LSTM + Attention

Bu mimariler, önce Mel-spektrogram gibi temsillerden uzamsal özellikleri çıkarır, ardından zamansal örüntüleri RNN veya LSTM ile işler ve dikkat mekanizmasıyla en anlamlı bölgelere odaklanır (Rocha vd., 2019; Ethala, 2025).

2.13.3. Transformer Tabanlı Yaklaşımlar

Transformer mimarileri, özellikle uzun süreli bağımlılıkları modelleme ve dikkat mekanizmaları sayesinde girişin anlamlı bölümlerine odaklanma konularında son derece başarılı yapılar sunmaktadır. Vaswani ve arkadaşları (2017) tarafından geliştirilen temel Transformer yapısı, görüntü ve ses sınıflandırması gibi alanlarda geniş çapta kullanılmaya başlanmıştır. Solunum sesi analizinde de Transformer tabanlı modellerin kullanımı son yıllarda dikkat çekici biçimde artmıştır.

2.13.3.1. Vision Transformer (ViT)

Vision Transformer (ViT), görüntü verilerini sabit boyutlu parçalara (patch) ayırarak doğrudan Transformer yapısında işlemektedir. Bu model, Mel-spektrogram gibi görüntü temsillerini işlerken evrişimli yapılara kıyasla daha esnek öğrenme kapasitesi sunar (Dosovitskiy vd., 2020).

2.13.3.2. Audio Spectrogram Transformer (AST)

AST modeli, ses sınıflandırmasına özel olarak geliştirilmiş ilk Transformer mimarilerindendir. Google tarafından tanıtılan bu model, AudioSet veri kümesi üzerinde önceden eğitilmiş olup, solunum sesi gibi spektral temelli seslerin sınıflandırılmasında başarılı sonuçlar sunmaktadır (Gong vd., 2021).

2.13.3.3. Self-Supervised AST (SSAST)

AST modelinin kendi kendine öğrenen versiyonu olan SSAST, etiketlenmemiş ses verilerinden anlamlı temsiller öğrenebilmekte ve az veriyle yüksek doğruluk elde edebilmektedir. Bu özellik, medikal ses verilerinin sınırlı olduğu durumlarda önemli avantaj sağlar (Gong vd., 2022).

2.13.3.4. Conformer

Conformer mimarisi, CNN ve Transformer bloklarının hibrit bir birleşimidir. Hem yerel örüntülerin öğrenilmesi hem de uzun süreli bağımlılıkların modellenmesi açısından başarılıdır. Solunum sesi gibi zaman-frekans temsillerinde bu tür ikili öğrenme süreci oldukça etkilidir (Gulati vd., 2020).

2.13.3.5. Swin Transformer

Swin Transformer, “kayan pencere” (sliding window) tekniğini kullanarak yüksek çözünürlüklü görüntülerde verimli işlem yapar. Mel-spektrogram üzerinde yer alan ince detayları kaybetmeden analiz yapması, bu yapıyı solunum sesi sınıflandırmalarında da potansiyel bir aday hâline getirmektedir (Liu vd., 2021).

2.13.3.6. Perceiver / Perceiver IO

Perceiver mimarisi, farklı veri türlerini (görüntü, ses, metin) aynı çatı altında işleyebilme yeteneğine sahiptir. Özellikle çok modlu (multimodal) sağlık uygulamalarında solunum sesi ile birlikte diğer biyobelirteçlerin de analizinde kullanılabilmektedir (Jaegle vd., 2021).

Tablo 5. Derin öğrenme modellerinin avantajları ve dezavantajları

Model	Avantajları	Dezavantajları
CNN	Otomatik öznitelik çıkarımı, yüksek doğruluk	Büyük veri ihtiyacı
GhostNet	Hafif yapı, mobil uyumlu	Literatürde sınırlı kullanım
ResNet	Derin ağlarda stabil öğrenme	Yüksek hesaplama maliyeti
EfficientNet	Optimize yapı, az parametreyle yüksek doğruluk	Eğitim süreci karmaşık
MobileNet	Gerçek zamanlı, düşük bellek kullanımı	Karmaşık desenleri öğrenmede zayıf
DenseNet	Gradyan akışını iyileştirme, veri azlığında etkili	Hafıza ve işlem yükü
InceptionNet	Çok ölçekli bilgi çıkarımı	Mimari karmaşıklık

Tablo 5. (Devamı)

Model	Avantajları	Dezavantajları
Transformer (ViT)	Uzun bağımlılıkları yakalama, yorumlanabilirlik	Çok veri gerektirir, eğitim maliyeti yüksek
Swin Transformer	Hiyerarşik yapı, düşük çözünürlükte bile güçlü performans	Yapısal karmaşıklık
RNN / LSTM	Zaman bağımlılıklarını öğrenir	Uzun eğitim süresi
Attention Mek.	Önemli bölgelere odaklanma, yüksek temsil gücü	Yüksek hesaplama maliyeti

2.14. Solunum Sesi Veri Tabanları

Solunum sesi sınıflandırması ve analizi alanında güvenilir veri kümeleri, yöntemlerin karşılaştırılabilirliğini ve klinik geçerliliğini doğrudan etkiler. Yalnızca kayıt sayısı değil; kayıt cihazı ve örnekleme hızı, dinleme/boylama bölgesi, etiketleme düzeyi (döngü, olay, kayıt), hasta-bağımsız bölme ve erişim/lisans koşulları gibi metodolojik ayrıntılar da kritik önem taşır. Aşağıda literatürde öne çıkan üç veri kümesi, kapsam ve sınırlılıklarıyla özetlenmiş; karşılaştırmalı görünüm Tablo 6’da sunulmuştur.

ICBHI 2017 Respiratory Sound Database, 126 bireyden toplanmış 920 kayıt ve yaklaşık 6.898 solunum döngüsü içerir. Veriler farklı elektronik stetoskop/mikrofonlarla; trakea, ön ve arka akciğer alanlarından kaydedilmiştir. Döngüler normal, hışıltı (wheeze), ral (crackle), hışıltı+ral olarak etiketlenmiştir. Gerçek dünya çeşitliliği genellenebilirliği artırırken; sınıf dengesizliği ve cihaz/ortam kaynaklı akustik değişkenlik, ön işleme (filtreleme, normalizasyon), veri artırma ve hasta-bağımsız eğitim-test bölmelerini zorunlu kılar. (ICBHI Challenge, 2017; Rocha vd., 2019).

Coswara (IISc Bangalore), COVID-19 döneminde oluşturulan bu küme, nefes alma (sığ/derin), öksürük (sığ/derin), sesli harf fonasyonu ve sayma gibi mikrofon tabanlı sesleri içerir; eşlik eden yaş, cinsiyet, semptom ve test sonucu gibi zengin meta-veri sağlar. Çeşitli cihaz/ortamlardan uzaktan toplandığı için gürültü kontrolü daha zordur; ancak solunumla ilişkili (oskültasyon değil) geniş ölçekli bir referans olarak değerlidir (Bhattacharya vd., 2023; Sharma vd., 2020).

Jordan Lung Sound Dataset, Ürdün’de bir klinikte 3M Littmann 3200 elektronik stetoskopla göğüs duvarından kaydedilen veri setidir. Kaynakta P1–P112 olmak üzere 112 kişiye ait kayıtlar; Bell (20–200 Hz), Diaphragm (100–500 Hz), Extended (50–500 Hz) olmak üzere üç frekans filtresi ile alınmıştır. Etiketler; hışıltı, ral, normal işaretleri ve tanı (astım, KOAH, pnömoni, bronşit, kalp yetmezliği, plevral efüzyon, akciğer fibrozu)

gibi klinik başlıkları içerir. Mendeley Data üzerinde CC BY 4.0 lisansı ile paylaşıldığından erişilebilirlik üst düzeydedir (Fraïwan vd., 2021).

Değerlendirme açısından hatırlatmalar. Kayıt (dosya) sayısı ile etiketli döngü sayısı aynı değildir; yöntem karşılaştırmalarında doğrulama düzeyi açıkça belirtilmelidir. Birey bağımsız bölme esas olmalı; sonuçlar yalnızca doğrulukla değil duyarlılık, özgüllük, F1 gibi dengesizliği yansıtan ölçülerle raporlanmalıdır. ICBHI literatüründe sık kullanılan “ICBHI skoru” (normal-anormal düzeyinde, duyarlılık/özgüllük ortalaması) klinik tarama senaryolarında pratik bir özet sağlar (ICBHI Challenge, 2017; Rocha vd., 2019).

Tablo 6. Yaygın Kullanılan Solunum Sesi Veri Setlerinin Karşılaştırması

Veri Seti	Kayıt Sayısı	Sınıflar	Kayıt Özelliği	Notlar
ICBHI 2017	920	Normal, Wheeze, Crackle, Combo	Döngü düzeyi, çok cihazlı	Sınıf dengesizliği, çoklu dinleme bölgesi
Coswara	2.600+	Normal, COVID+, farklı ses tipleri	Uzaktan, çok modlu	Meta-veri zengin, ortam kontrolü sınırlı
Jordan	112	7+ hastalık tipi	3 frekans filtresi	Cihaz kontrollü, hastalık çeşitliliği geniş

3. LİTERATÜR TARAMASI

3.1. Giriş

Solunum sistemi hastalıkları, dünya genelinde yüksek bir hastalık yükü oluşturur ve başlıca ölüm nedenleri arasında yer alır. Astım, Kronik Obstrüktif Akciğer Hastalığı (KOA), pnömoni ve bronşektazide erken tanı ile izlemde solunum seslerinin değerlendirilmesi kritik önemdedir. Stetoskopla yapılan oskültasyon hızlı ve düşük maliyetlidir; ancak gözlemciye bağımlı öznel niteliği, dinleyiciler arası farklılık ve kayıt koşullarına duyarlılık gibi nedenlerle tanısallık güvenirliği sınırlandırabilir (Bohadana vd., 2014; Pasterkamp vd., 1997).

Son yirmi yılda elektronik stetoskoplar ve sayısal sinyal işleme yöntemlerinin yaygınlaşmasıyla bilgisayar destekli solunum sesi analizi belirgin bir ivme kazanmıştır. Standartlara uygun kayıt alma, zaman–frekans temelli çözümleme ve otomatik karar destek sistemleri, oskültasyonun öznel doğasını azaltmayı amaçlar. Araştırmaların karşılaştırılabilirliğini artırmak için ortak veri tabanlarının (örneğin ICBHI 2017) kullanımı ve hasta düzeyinde değerlendirme protokollerine yönelim dikkat çekmektedir (ICBHI Challenge, 2017; Pramono vd., 2017; Rocha vd., 2017; Rocha vd., 2019).

Bu veri seti üzerinde özellikle dört sınıflı (normal, wheeze, crackle, wheeze+crackle) kurulum, alan için fiilî bir karşılaştırma zemini hâline gelmiş; CNN ve LSTM tabanlı derin ağlardan transfer öğrenme, dikkat mekanizmaları, hibrit makine öğrenmesi ve transformer temelli modellere kadar geniş bir yöntem ailesi aynı referans üzerinden raporlanmıştır (Liu vd., 2019; Demir vd., 2020; Petmezas vd., 2022; Nguyen ve Pernkopf, 2022; Xu vd., 2023; Prabhakar vd., 2023; Khan vd., 2024; Kim vd., 2024; Wu vd., 2024; Zhang vd., 2024; Wang vd., 2024; Mang vd., 2024). Bununla birlikte aynı veri seti ve aynı sınıf tanımı kullanılsa dahi; segmentleme yaklaşımı, veri artırma stratejileri ve özellikle hasta düzeyi bölme protokolündeki farklılıklar performansın doğrudan karşılaştırılabilirliğini etkileyebileceğinden, sonuçların bu bağlamda temkinli yorumlanması gerekir (Chambres vd., 2018; Rocha vd., 2019).

Literatürde iki ana yaklaşım öne çıkar. Klasik makine öğrenmesi çizgisinde ham sesten Mel Frekans Kepstrum Katsayıları (MFCC), dalgacık katsayıları, enerji ve entropi gibi öznitelikler çıkarılır ve bu öznitelikler Destek Vektör Makineleri (Support Vector Machines – SVM), k-En Yakın Komşu (K-Nearest Neighbors – KNN) veya Naive Bayes gibi sınıflandırıcılara verilerek karar üretilir; bu işlem hattı şeffaf ve yorumlanabilir bir

yapı sunar (Bahoura, 2009; Bishop, 2006; Davis ve Mermelstein, 1980; Hastie vd., 2009; Pramono vd., 2017). Derin öğrenme çizgisinde ise Evrişimsel Sinir Ağları (Convolutional Neural Networks – CNN) ve Tekrarlayan Sinir Ağları / Uzun Kısa Süreli Bellek (Recurrent Neural Networks – RNN, Long Short Term Memory – LSTM) gibi mimariler öznitelikleri doğrudan veriden öğrenir; özellikle büyük ve heterojen veri kümelerinde daha güçlü genelleme potansiyeli sergiler (Goodfellow vd., 2016; LeCun vd., 2015; Rocha vd., 2019; Sfayyih vd., 2023).

Bu tez, klasik öznitelik tabanlı yöntemlerle CNN/RNN tabanlı derin modelleri aynı veri kümesi ve hasta düzeyinde protokoller altında sistematik biçimde karşılaştırarak klinik geçerliliği yüksek, adil bir değerlendirme sunmayı amaçlamaktadır. İzleyen alt bölümlerde sırasıyla klasik yöntemlere dayalı çalışmalar, derin öğrenme tabanlı yaklaşımlar ve kullanılan veri setleri ile değerlendirme protokolleri özetlenecek; bulgular kıyaslanarak literatürdeki boşluklar ve geleceğe yönelik araştırma yönleri tartışılacaktır.

3.2. Klasik Makine Öğrenmesi Yöntemlerine Dayalı Çalışmalar

Klasik yaklaşımın başarısı, doğru öznitelik mühendisliği ile hasta düzeyinde (patient-level) doğrulama protokolünün titizlikle uygulanmasına dayanır. Gürültü giderme (çoğunlukla 100–2000 Hz bant geçiren ve 50/60 Hz notch filtre), kalp sesi bastırma, nefes alma–nefes verme evrelerinin segmentasyonu ve artefakt temizliği, neredeyse tüm klasik hatların ortak ön işleme çerçevesini oluşturur. Bu adımların ardından MFCC, dalgacık tabanlı enerji/entropi ölçüleri, istatistiksel özetler (varyans, çarpıklık, basıklık), spektral ağırlık merkezi/roll-off/düzlük, sıfır geçiş oranı gibi özellikler çıkarılır ve SVM, KNN, Gauss Karışım Modeli (Gaussian Mixture Model – GMM), Naive Bayes, Karar Ağaçları ya da artırmalı/topluluk temelli sınıflandırıcılarla karar verilir. Literatür, küçük ve kontrollü veri senaryolarında iki ya da üç sınıflı görevlerde güçlü sonuçlar bildirirken, cihaz/merkez çeşitliliği ve sınıf sayısı arttığında performansın özellikle veri bölme protokolüne (aynı hastanın kayıtlarının eğitim ve testte karışmaması) duyarlı olduğunu gösterir (Pramono vd., 2017; Rocha vd., 2019).

2020 sonrasında klasik öğrenme çizgisi tamamen terk edilmemiş; aksine özellikle hibrit öznitelik setleri ve gelişmiş öznitelik seçimiyle güncellenmiştir. Örneğin Tuncer vd. (2021), yerel ikili örüntü (local binary pattern) türevi kalıntı tabanlı betimleyiciler ile dalgacık enerjisini birleştirerek istiflenmiş bir özellik kümesi oluşturmuş, ardından klasik sınıflandırıcılarla (SVM, KNN vb.) çok sınıflı solunum sesi sınıflandırmasında yüksek doğruluklar rapor etmiştir. Bu tür çalışmalar, özenli öznitelik mühendisliği yapıldığında

klasik hattın hâlâ rekabetçi olabileceğini göstermektedir (Pramono vd., 2017; Tuncer vd., 2021).

3.2.1. MFCC Tabanlı Çalışmalar

Konuşma tanımadan miras MFCC, solunum sesi için adeta başlangıç standartıdır. Bahoura (2009), dar bant gürültüsü ve kalp sesi etkisini azaltan bir bant geçiren filtre ile ön vurgulama ardışığını takiben kısa pencerelerde MFCC ve türevlerini (Δ , $\Delta\Delta$) çıkarıp GMM ile özellikle hışıltı (wheeze) tespitini iki sınıflı kurulumda göstermiştir. Bu hatta ortalama–varyans düzeltmesi (Cepstral Mean and Variance Normalization – CMVN) gibi basit istatistiksel eşitlemeler cihazlar arası farklılığı bir ölçüde dengeler; nefes alma ve nefes verme evrelerinin ayrı işlenmesi hışıltının sürekliliğini daha görünür kılar (Bahoura, 2009).

Çok sınıflı düzeneklerde MFCC’nin tek başına yetmediği görüldüğünde, log-enerji, bant yoğunluk oranları ve şekil betimleyicileri (örneğin spektral kontrast) ile zenginleştirilmiş MFCC setleri SVM ya da Çok Katmanlı Algılayıcı (Multilayer Perceptron – MLP) ile birleştirilmiştir (Bahoura, 2009; Pramono vd., 2017). Jin vd. (2014), MFCC’yi Yüksek Boyutlu Model Temsili (High Dimensional Model Representation – HDMR) ile zenginleştirerek korelasyonu azaltıp gürültüye dayanıklılığı artırmış; ardından SVM ve hafif sınıflandırıcılarla karar vermiştir. Yeni dönemde Tuncer vd. (2021), MFCC ve yerel dokusal öznitelikleri birlikte kullanarak, öznitelik seçimi için modern sezgisel yöntemlerden yararlanmış ve bu hibrit temsilin klasik sınıflandırıcılarla birleştirildiğinde çok sayıda solunum sesi sınıfını ayırt etmede yüksek performans sunduğunu göstermiştir (Jin vd., 2014; Pramono vd., 2017; Tuncer vd., 2021).

3.2.2. Destek Vektör Makinelerine (SVM) Dayalı Çalışmalar

SVM, orta boy öznitelik vektörlerinde (yaklaşık 20–200 öznitelik) Radyal Taban Fonksiyonu (Radial Basis Function – RBF) çekirdeğiyle güvenilir bir başlangıç noktasıdır. İçer ve Gengeç (2014), durağan olmayan adventif sesleri tanımlamak için adaptif zaman–frekans öznitelikleri tasarlamış, nefes alma/nefes verme için ayrı pencerelemeyi ve kalp sesi bastırmayı standardize ederek RBF-SVM ile ral/ronküs ayrımını göstermiştir. Jin vd. (2014) ise spektro-zamansal öznitelikleri (örneğin dar bant izlerin sürekliliği, tepe yoğunluğu, bant-enerji oranları) bir araya getirip SVM’de çok sınıflı kurulumda karar vermiş; özellikle bant genişliği ve “çizgi takip” tabanlı istatistiklerin hışıltı için kilit olduğuna dikkat çekmiştir. Bu çalışmalar, C ve gamma parametrelerinin ızgara taraması ile ayarlanmasını ve sınıf ağırlıkları kullanılarak azınlık

sınıfların duyarlılığının dengelenmesini önermektedir (İçer ve Gengeç, 2014; Jin vd., 2014; Cortes ve Vapnik, 1995).

3.2.3. K-En Yakın Komşu (KNN) ve Bayes Sınıflandırıcıları

K-En Yakın Komşu, iyi ölçeklenmiş MFCC- $\Delta/\Delta\Delta$ + istatistik setlerinde kosinüs veya öklit uzaklığıyla beklenenden güçlü performans verebilir; ancak boyut arttıkça “yakınlık” kavramı bulanıklaşır. Bu nedenle literatürde KNN çoğu kez Doğrusal Ayırıcı Analiz (Linear Discriminant Analysis – LDA) ya da Temel Bileşenler Analizi (Principal Component Analysis – PCA) ile boyut indirgeme sonrasında kullanılır. Naive Bayes ise öznitelikler arası bağımsızlık varsayımına rağmen küçük–orta veri setlerinde hızlı ve istikrarlı bir taban çizgisi sunar. Sistemik derlemeler, genetik algoritmalarla öznitelik seçimi ve Fisher doğrusal ayırımıyla sadeleştirme sonrası KNN/Naive Bayes’in, hışıltı/ral ayrımlarında hasta düzeyi bölmelerle rekabetçi olduğunu bildirmektedir (Bishop, 2006; Cover ve Hart, 1967; Hastie vd., 2009; Pramono vd., 2017).

3.2.4. Dalgacık, Zaman-Frekans Dönüşümleri ve Hibrit Yaklaşımlar

Kandaswamy vd. (2004), dalgacık paket ayrışımıyla bant enerjisi/entropi öznitelikleri çıkarıp küçük yapay sinir ağlarıyla erken dönem akciğer sesi sınıflandırması yapmış; “hangi bantta enerji birikiyor?” sorusunu doğrudan öznitelik uzayına taşımıştır. Dokur vd. (2003) ise benzer dalgacık–enerji betimlerinden beslenen sinir ağı ve bulanık temelli kararlarla klasik çizgiye öncülük etmiştir. Mendes vd. (2016), müzik enformasyon erişimi (Music Information Retrieval – MIR) kökenli özellikler, Teager enerji operatörü (Teager Energy Operator – TEO) ile dar bantlı titreşim göstergeleri, dalgacık kırıntıları ve istatistiklerini bir “çoklu öznitelik” çerçevesinde birleştirip lojistik regresyon ve küçük karar ağaçları gibi hafif sınıflandırıcılarla ral tespitini işlemiştir; çerçeve süresi ve bindirme oranının milisaniye ölçekli patlamaları kaçırmamak için kritik olduğu vurgulanır. Hilbert–Huang dönüşümünü (Empirical Mode Decomposition – EMD + Hilbert spektrumu) kullanan çalışmalarda ise anlık frekans ve enerji izleri çıkarılarak hışıltı ve ral için olay temelli eşikleme ve makine öğrenmesi kararı üretilmiştir (Mendes vd., 2016; Reichert vd., 2008; Taplidou ve Hadjileontiadis, 2007).

3.2.5. Klasik Yöntemlerin Genel Değerlendirmesi

Klasik hatta bildirilen “çok yüksek” başarılar, kimi zaman dosya/segment düzeyinde rastgele bölmelerden kaynaklanır; aynı hastaya ait döngülerin hem eğitim hem testte yer alması modeli “kişiye özgü izlere” duyarlılaştırıp genel performansı olduğundan

yüksek gösterebilir. Sistematik derlemeler ve ICBHI çevresindeki metodoloji yazıları bu nedenle hasta düzeyi ayrımları, en azından resmî ICBHI bölmesini ya da “leave-one-subject-out” şemalarını önermekte; metrik olarak doğruluğun yanı sıra sınıf dengesizliğini daha iyi yansıtan duyarlılık, özgüllük ve ICBHI skorunun raporlanmasını istemektedir (Pramono vd., 2017; Rocha vd., 2017; Rocha vd., 2019; Rocha vd., 2021).

Özetle klasik makine öğrenmesi hattı; (i) filtreleme–segmentasyon–artefakt temizliği, (ii) MFCC $\pm$ türevler + dalgacık/bant-enerji + istatistiksel/spektral betimleyiciler, (iii) SVM/KNN/GMM ve hafif topluluklar üçlüsüne dayanır. Ral gibi kısa ve geniş bant olaylar için dalgacık ve olay temelli ölçüler; hışıltı gibi dar bant, süreklilik gösteren olaylar için de MFCC’ye ek bant izleme ve spektral kontrast türü ölçüler öne çıkar. Hasta düzeyi bölmeler ve cihaz/ortam çeşitliliğini hesaba katan veri artırma/dengeleme stratejileri uygulanmadığında sonuçların klinik genellemesi sınırlı kalabilir. Buna karşılık aynı ön işleme ve protokoller altında klasik hat; yorumlanabilirliği, düşük hesap yükü ve küçük veriyle çalışabilmesi sayesinde derin öğrenme yaklaşımlarına güçlü ve pratik bir kıyas çizgisi sunmaya devam eder (Bahoura, 2009; Pramono vd., 2017; Reichert vd., 2008; Rocha vd., 2021; Tuncer vd., 2021).

Son yıllarda klasik çizginin modern varyantlarının da ICBHI 2017 üzerinde rekabetçi sonuçlar bildirdiği görülmektedir. Örneğin Prabhakar vd. (2023), varyasyonel mod ayrışımı (VMD) ile elde edilen bileşenleri Manhattan mesafesi temelli bir temsil ve Aşırı Öğrenme Makinesi (Extreme Learning Machine – ELM) benzeri hafif bir sınıflandırma yaklaşımıyla birleştirerek dört sınıflı solunum sesi sınıflandırmasında yüksek doğruluklar rapor etmiştir. Bu bulgu, uygun öznitelik ayrıştırması ve dengeli bir karar mekanizması ile klasik ya da yarı-klasik hatların, derin öğrenme modellerine karşı hâlen güçlü bir referans çizgisi oluşturabildiğini göstermektedir (Prabhakar vd., 2023).

3.3. Derin Öğrenmeye Dayalı Çalışmalar

Derin öğrenme yaklaşımında solunum sesleri çoğunlukla zaman–frekans temsillerine (özellikle Mel-spektrogram, Kısa Zamanlı Fourier Dönüşümü – STFT tabanlı spektrogram, Sürekli Dalgacık Dönüşümü – CWT/skalogram) dönüştürülerek görüntü sınıflandırma mimarilerine verilir; böylece alt katmanlar yerel çizgisel dokuları ve bant izlerini, üst katmanlar ise hışıltının (wheeze) dar bant boyunca “uzayan çizgi” yapısını veya rallerin (crackle) milisaniyelik geniş bant parlamalarını ayırt edebilecek soyut temsilleri öğrenir. Veri azlığı ve heterojen cihaz/ortam değişkenliği nedeniyle çalışmalarda sıklıkla veri artırma (zaman/frekans maskesi, gürültü enjeksiyonu, zaman ölçekleme), sınıf ağırlıkları ve erken durdurma kullanılır; ayrıca hasta düzeyi bölme

protokolleriyle (aynı kişiye ait kayıtlar eğitim ve testte karışmayacak şekilde) genelleme gücü izlenir (Goodfellow vd., 2016; LeCun vd., 2015; Park vd., 2019; Pramono vd., 2017; Rocha vd., 2019). Son dönem derlemeler, özellikle ICBHI 2017 veri tabanı üzerinde çalışan derin modellerin büyük kısmının CNN, hibrit CNN–RNN mimarileri ve son yıllarda dönüştürücü (transformer) ve öz denetimli öğrenme (self-supervised learning) ailesinde toplandığını göstermektedir (Hsu vd., 2021; Sfayyih vd., 2023).

Son yıllarda ICBHI 2017 veri seti üzerinde gerçekleştirilen deneysel çalışmalar, mimari çeşitliliğin arttığını ve transfer öğrenme ile topluluk (ensemble) stratejilerinin yaygınlaştığını göstermektedir. Örneğin Demir vd. (2020) paralel havuzlama yapısına sahip derin CNN'den çıkarılan özellikleri LDA ve Random Subspace Ensemble ile birleştirerek dört sınıflı problemde %63,09 doğruluk bildirmiştir. Liu vd. (2019) ise Mel tabanlı temsiller üzerinde CNN kullanarak segment düzeyinde %81,62 doğruluğa ulaşmıştır. Daha yakın dönemde Xu vd. (2023), AudioSet üzerinde ön-eğitilmiş Wavegram-LogMel CNN ve snapshot ensemble yaklaşımıyla ICBHI skorunu %81,1'e yükseltmiş; demografik değişkenlerin modele etkisini ayrıca incelemiştir.

3.3.1. CNN Tabanlı Yaklaşımlar

CNN çizgisinde, özelleştirilmiş mimarilerle dikkat (attention) ve çok ölçekli alıcı alanlar birleştirilerek wheeze/crackle imzaları daha belirgin hâle getirilmeye çalışılır. Gairola vd. (2021) RespireNet ile ResNet-34 omurgasını solunum sesi görevine uyarlamış; veri artırmada zaman/frekans maskesi, dengelemede sınıf ağırlıkları ve hasta düzeyi protokolün önemini vurgulamıştır. Mimari, hem normal–anormal hem de çok sınıflı kurulumlarda yeniden kullanılabilir bir iskelet olarak sunulmuştur (Gairola vd., 2021).

Chen vd. (2019), optimize edilmiş S-dönüşümü (Optimized S-Transform – OST) ile elde edilen spektral haritaları ResNet tarzı CNN'lerle birleştirerek wheeze/crackle ayrımında klasik MFCC tabanlı hatlara kıyasla belirgin kazanımlar rapor etmiş, bu yaklaşım sonraki çalışmalarda referans bir yöntem olarak anılmıştır (Chen vd., 2019; Sfayyih vd., 2023).

Son yıllarda Mel-spektrogram temelli girişlerle ses görevleri için önceden eğitilmiş CNN tabanlı ses ağlarının (örneğin Pretrained Audio Neural Networks – PANNs) solunum sesi sınıflandırmasına uyarlandığı çalışmalar da artmıştır. Bu çalışmalarda gürültü dayanıklılığını artırmak için zaman/frekans maskesi ve gürültü enjeksiyonu gibi veri artırma stratejilerinin, özellikle düşük sinyal gürültü oranlı solunum kayıtlarında

performansı anlamlı biçimde iyileştirebildiği gösterilmiştir (Kong vd., 2020; Park vd., 2019; Sfayyih vd., 2023).

Bununla birlikte literatürde CNN omurgalarının farklı makine öğrenmesi sınıflandırıcıları ile hibrit kullanımına yönelik çalışmalar da bulunmaktadır. Zhang vd. (2024), ImageNet ön-eğitilmiş CNN + CBAM dikkat modülü ile çıkarılan özellikleri CatBoost sınıflandırıcısına vererek ICBHI veri setinde %75,73 ICBHI skoru elde etmiş ve boosting tabanlı yöntemlerin tam bağlı katmanlara kıyasla daha iyi genelleme sağlayabildiğini göstermiştir. Benzer şekilde Wu vd. (2024) Bi-ResNet ve CNN bileşimini kullanarak %77,81 doğruluk raporlamıştır.

3.3.2. Transfer Öğrenme ve Hibrit Modeller

Veri kısıtı olan senaryolarda, görüntü görevlerinde önceden eğitilmiş CNN omurgalarının (VGG, ResNet, DenseNet, EfficientNet) erken katmanları sabitlenip son katmanlar solunum sesi için yeniden eğitilir. Bu hat, Mel-spektrogram tabanlı girişlerde özellikle etkili bulunmuştur. Çeşitli çalışmalarda ResNet-50, DenseNet-201 gibi omurgalardan alınan özellikler ya doğrudan uçtan uca eğitilerek ya da SVM/LDA gibi klasik sınıflandırıcılarla sonlandırılarak değerlendirilmiştir; bildirilen sonuçlar veri kalitesi, bölme protokolü ve dengeleme stratejilerine oldukça duyarlıdır (Chen vd., 2019; Gairola vd., 2021; Sfayyih vd., 2023).

Karışık mimarilerde CNN katmanlarından çıkarılan uzamsal özellik haritaları, LSTM veya Çift Yönlü LSTM (Bidirectional LSTM – Bi-LSTM) yapısına verilerek döngü içi süreklilik yakalanmaktadır. Perna ve Tagarelli (2019), bu tür CNN + Bi-LSTM kombinasyonunun inspirasyon/ekspirasyon boyunca bant izlerinin evrimini modelleyerek wheeze/crackle ayırımında tek başına CNN’e göre daha yüksek doğruluk sunduğunu göstermiştir.

Güncel çalışmalarda transfer öğrenmenin farklı alanlardan kazanılan temsillerle genişletildiği görülmektedir. Khan vd. (2024) LSTM tabanlı derin mimari ile %79,61 doğruluk bildirirken, Kim vd. (2024) çapraz alan uyumlama stratejileri kullanmış ve veri dağılımı farklılıklarının performans üzerindeki etkisini ortaya koymuştur.

3.3.3. RNN/LSTM ve Dikkat Mekanizmaları

Solunum sinyallerinde bilgi yalnızca tek karede değil, döngü boyunca birikir. Bu nedenle CNN’den sonra zaman boyutunu modelleyen Bi-LSTM/Gated Recurrent Unit (GRU) katmanları ve dikkat (attention) mekanizmaları yaygındır. Dikkat, modelin Mel-spektrogram üzerinde “tanısal” zaman–frekans bölgelerine ağırlık vermesine izin vererek

hem performansı hem de açıklanabilirliği artırır. Ağın odaklandığı bölgeler Grad-CAM benzeri haritalarla görselleştirilerek hışıltı çizgilerinin dar bant boyunca izlenmesi veya ral adacıklarının kümelenmesi gibi klinik olarak anlaşılır ipuçları sunulur (Bahdanau vd., 2014; Perna ve Tagarelli, 2019; Selvaraju vd., 2017; Vaswani vd., 2017).

Pediyatrik gruplarda yapılan yeni çalışmalarda da, wheeze tespiti için CNN + LSTM ve dikkat mekanizmasının birlikte kullanılmasının, yalnız CNN tabanlı modellere kıyasla daha yüksek duyarlılık sunduğu bildirilmiştir (Hsu vd., 2021; Sfayyih vd., 2023).

3.3.4. Dönüşümcü (Transformer) Tabanlı ve Kokleogram Yaklaşımları

Son yıllarda Mel-spektrogram yerine işitsel koklea modelini taklit eden kokleogram temsilleriyle Vision Transformer (ViT) veya ses odaklı dönüştürücü omurgalarının (örneğin Audio Spectrogram Transformer – AST) denendiği çalışmalar artmıştır. Kokleogram, düşük frekans bantlarında daha ince çözünürlük vererek wheeze'nin dar bant imzasını ve ralın yüksek frekans bileşenlerini ayırt etmeyi kolaylaştırır; ViT tarafında ise çok başlı dikkat katmanları uzun menzilli bağımlılıkları yakalar (Dosovitskiy vd., 2020; Latif vd., 2020).

Bae ve Kim (2023), AST tabanlı bir mimariyi ICBHI 2017 veri kümesi üzerinde değerlendirerek, Mel-spektrogram yamalarını (patch) giriş alan bir dönüştürücü modelin, klasik CNN'lere kıyasla çok sınıflı solunum sesi sınıflandırmasında rekabetçi veya üstün performans sağlayabildiğini göstermiştir. Bu tür transformer tabanlı mimariler, özellikle çok merkezli ve gürültülü veri koşullarında CNN'lere alternatif olarak öne çıkmaktadır (Bae ve Kim, 2023; Sfayyih vd., 2023).

3.3.5. Alan Uyumlama (Domain Adaptation), Federated Learning ve Öz-Denetimli (Self-Supervised) Öğrenme

ICBHI gibi çok cihazlı veri kümelerinde modelin bir hastane/cihazdan diğerine genellemesi kritik bir sorundur. Alan uyumlama (domain adaptation) çalışmaları, kaynak alan (örneğin belirli bir stetoskop veya klinik) ile hedef alan arasında spektral istatistikleri hizalayan düzenleyici terimler ve karşıt-öğrenme (adversarial learning) şemaları kullanarak wheeze/crackle karışıklığını azaltmayı hedefler (Hsu vd., 2021; Sfayyih vd., 2023).

Veri gizliliği ve çok merkezli dağıtık eğitim gereksinimleri için federated learning çizgisi öne çıkmaktadır: kenar cihazlarda (taşınabilir elektronik stetoskoplar) yerel CNN–LSTM modelleri eğitilip yalnızca ağırlık güncellemeleri merkezi sunucuya aktarılmakta;

böylece ham veriyi paylaşmadan ortak bir model elde edilmektedir (Hsia vd., 2021; Kairouz vd., 2021).

Etiketli veri kısıtını hafifletmek için öz-denetimli ve karşıt (contrastive) ön eğitim stratejileri de benimsenmiştir. Özellikle solunum sesinin farklı artırılmış görünümlerini (zaman/frekans maskesi, kazanç değişimi, gürültü ekleme) aynı örneğe ait olarak tanıtan contrastive learning yaklaşımları, etiket gereksinimini düşürmekte ve son katman ince ayarıyla klinik etiketlere aktarılmaktadır (Hsu vd., 2021; Sfayyih vd., 2023).

3.3.6. Derin Öğrenme Yaklaşımlarının Genel Değerlendirmesi

Derin öğrenme tabanlı çalışmalarda performansı belirleyen başlıca unsurlar; kullanılan temsil (Mel-spektrogram, skalogram, kokleogram vb.), ağ mimarisi (ResNet, EfficientNet, hafif CNN + LSTM, Vision Transformer vb.) ve doğrulama protokolünün hasta düzeyinde kurgulanıp kurgulanmadığıdır. Literatürde bildirilen yüksek doğruluk oranlarının önemli bir kısmının, iki sınıflı (normal–anormal) senaryolarda veya kayıt/segment düzeyinde rastgele bölmelerle elde edildiği; aynı hastaya ait kayıtların hem eğitim hem test kümelerinde yer almasının “bilgi sızıntısı”na yol açabildiği vurgulanmaktadır (Chambres vd., 2018; Gairola vd., 2021; Pramono vd., 2017; Rocha vd., 2019). Bu nedenle güncel iyi uygulama yaklaşımı; hasta düzeyi bölme, resmî ICBHI ayrımının kullanılması ve doğruluğa ek olarak duyarlılık, özgüllük, F1 ve ICBHI skorunun birlikte raporlanmasıdır (Chambres vd., 2018; ICBHI Challenge, 2017; Rocha vd., 2019).

Çalışmalar, derin ağların özellikle heterojen ve çok sınıflı solunum sesi görevlerinde klasik yöntemlere göre daha güçlü bir genelleme potansiyeline sahip olduğunu göstermektedir. OST tabanlı CNN’ler (Chen vd., 2019), RespireNet (Gairola vd., 2021) ve CNN + Bi-LSTM kombinasyonları (Perna ve Tagarelli, 2019), wheeze/crackle ayrımında klasik MFCC tabanlı hatlara göre belirgin iyileşmeler bildirmektedir. Buna karşılık, sınıf dengesizliği, cihazlar arası frekans yanıt farklılıkları ve düşük sinyal-gürültü oranına (SNR) dayanıklılık hâlen önemli zorluk alanlarıdır (Rocha vd., 2019; Sfayyih vd., 2023).

Temsil tarafında Mel-spektrogramlara ek olarak çok ölçekli CWT tabanlı skalogramlar ve insan kokleasını taklit eden kokleogramlar denenmiştir. Bu temsiller, özellikle düşük frekanslı bantlarda daha yüksek çözünürlük sağlayarak wheeze’in dar bant boyunca uzayan çizgi yapısını ve crackle’ın milisaniye düzeyindeki geniş bant parlamalarını daha belirgin hâle getirebilmektedir (Latif vd., 2020; Sfayyih vd., 2023).

Öz-denetimli ve çok görünüm (multi-view) temelli yaklaşımlar, etiketleme maliyetini düşürme potansiyeli nedeniyle önem kazanmaktadır. Çok görünümlü çerçevelerde solunum seslerinin farklı artırılmış versiyonları yakın temsil uzayında bir araya getirilmekte; etiketlenmemiş veriden öğrenilen bu gömüler, daha sonra küçük etiketli kümeler üzerinde ince ayar yapılarak sınıflandırma performansını artırmaktadır (Hsu vd., 2021; Sfayyih vd., 2023). Bununla birlikte derin öğrenme temelli yaklaşımların klinik ortama aktarımında açıklanabilirlik, model kalibrasyonu ve alt grup (yaş, cinsiyet, komorbidite) performans analizleri gibi başlıklar hâlâ sınırlı düzeyde ele alınmaktadır (Doshi-Velez ve Kim, 2017; Rudin, 2019; Sfayyih vd., 2023).

2022–2024 dönemindeki sonuçlar incelendiğinde, ICBHI üzerinde raporlanan doğrulukların yaklaşık %70–80 bandında yoğunlaştığı, ancak kullanılan protokole ve segment/hasta düzeyi ayırımına bağlı olarak değişkenlik gösterdiği görülmektedir. Örneğin Nguyen vd. (2022) CNN-MoE ile %78,6, Petmezas vd. (2022) federated CNN-LSTM ile %76,39 doğruluk bildirirken, Xu vd. (2023) transfer öğrenme ve ensemble stratejileri ile bu seviyeyi aşmıştır. Bu tablo, mimari karmaşıklığın tek başına yeterli olmadığını; veri artırma, adil bölme ve özellik temsillerinin en az ağ derinliği kadar kritik olduğunu göstermektedir.

3.4. Veri Setleri ve Değerlendirme Protokolleri

Literatürde en yaygın kullanılan kaynak ICBHI 2017 Respiratory Sound Database'dir. Veri tabanı 126 bireyden toplanmış 920 kayıt ve bu kayıtlardan çıkarılmış 6.898 solunum döngüsü içerir; döngüler dört sınıfta etiketlenmiştir: normal, wheeze (hışıltı), crackle (ral), wheeze+crackle. Kayıtlar farklı merkez ve cihazlardan (çeşitli elektronik stetoskop/mikrofonlar) ve önden/arkadan/trakeal bölgelerden alınmıştır; bu çok merkezli yapı gerçek yaşama yakın bir heterojenlik sağlarken cihaz yanıtı ve ortam farklılıklarından doğan varyansı da artırır (ICBHI Challenge, 2017; Rocha vd., 2019). Bu nedenle çalışmaların önemli bir kısmı, filtreleme–segmentasyon–artefakt temizliği yanında hasta düzeyi bölme ve dengeleme/artan veri stratejileriyle model genellemesini güçlendirmeye odaklanır (Chambres vd., 2018; Pramono vd., 2017; Rocha vd., 2019).

ICBHI yarışmasının değerlendirme protokolünde araştırma topluluğunun benimsediği resmî eğitim–test bölmesi yer alır; bu bölme özünde hasta temelli olup (aynı kişiye ait kayıtların eğitim ve testte birlikte bulunmasını engeller) literatürde sıkça görülen “dosya/segment rastgele ayırım” ile kıyaslandığında bilgi sızıntısı riskini belirgin biçimde azaltır. Birçok çalışma bu resmî bölmeyi veya hasta düzeyi çapraz doğrulamayı benimser; ayrıca dört sınıflı kurulumda yalnız doğruluk değil, ICBHI skoru (wheeze ve

crackle sınıfları için duyarlılık ve özgüllüğün ortalaması) de raporlanır (Chambres vd., 2018; ICBHI Challenge, 2017; Rocha vd., 2019).

ICBHI dışındaki yaygın veri kaynaklarından Coswara (Indian Institute of Science – IISc Bangalore) uzaktan toplanmış büyük ölçekli bir derlemedir. Sharma vd. (2020), COVID-19 döneminde web tabanlı bir platform aracılığıyla nefes alma (sığ/derin), öksürük (sığ/derin), sesli harf fonasyonu ve sayma gibi görevlerden oluşan çok modlu ses kayıtlarını, yaş, cinsiyet, semptom ve COVID-19 test sonucu gibi zengin meta verilerle birlikte toplamışlardır. Coswara, oskültasyon verisi içermese de solunumla ilişkili sesler için yaygın bir referans hâline gelmiş; özellikle COVID-19 tarama ve solunum bozukluklarının uzaktan izleminde sıkça kullanılmıştır (Sharma vd., 2020).

Jordan Lung Sound Dataset ise Ürdün’de bir klinikte 3M Littmann 3200 elektronik stetoskopla göğüs duvarından kaydedilen 112 birey ve çoklu hastalık sınıflarını (astım, KOAH, pnömoni, bronşit, kalp yetmezliği, plevral efüzyon, akciğer fibrozu vb.) kapsar. Kayıtlar Bell (20–200 Hz), Diaphragm (100–500 Hz) ve Extended (50–500 Hz) olmak üzere üç frekans filtresi ile yinelenerek alınmıştır. Etiketler; hışıltı, ral, normal ve tanı kategorilerini içerir. Veri sayısı ICBHI’ye kıyasla sınırlı olsa da hastalık çeşitliliği ve cihaz kontrolü nedeniyle yöntem geliştirme ve karşılaştırma çalışmaları için değerlidir (Fraïwan vd., 2021).

Son yıllarda, wheeze ve crackle gibi olayların zaman damgaları ile işaretlendiği ve dinleme noktasına, cihaza, hatta filtre ayarına ilişkin ayrıntılı meta veri içeren veri kümeleri de yayımlanmaya başlamıştır. Örneğin HF_Lung_V1 ve devamında tanımlanan HF_Lung_V2 gibi olay düzeyinde anotasyonlu derlemeler, kısa süreli rallerin ve süreklilik gösteren hışıltıların hem zaman serisi hem de olay düzeyinde değerlendirilmesine olanak tanır; ancak örnek ölçeği ve toplama protokollerindeki farklılıklar nedeniyle çapraz-veri genellemesi hâlâ güç bir problemdir (Hsu vd., 2021).

Çalışmalar arası adil karşılaştırma için üç ilke öne çıkar: (i) görev tanımının açık yazılması (iki, üç ya da dört sınıf; kayıt düzeyi mi döngü düzeyi mi), (ii) bölme protokolünün hasta düzeyi olması veya resmî ICBHI bölmesinin kullanılması, (iii) doğruluk yanında duyarlılık, özgüllük, F1 ve varsa ICBHI skorunun raporlanması. Özellikle dosya/segment rastgele bölmelerin, aynı kişiye ait kayıtların eğitim ve testte birlikte yer alması yoluyla performansı yapay biçimde yükseltebildiği vurgulanmalıdır (Chambres vd., 2018; ICBHI Challenge, 2017; Pramono vd., 2017; Rocha vd., 2019).

3.5. Performans Ölçütleri

Solunum sesi sınıflandırma çalışmalarında en sık kullanılan ölçütler doğruluk (accuracy), duyarlılık (sensitivity/recall), özgüllük (specificity), F1 skoru ve ICBHI skorudur. Bu ölçütlerin birlikte raporlanması, model performansının çok boyutlu ve dengeli değerlendirilmesini sağlar (Bishop, 2006; Hastie vd., 2009; Powers, 2020; Sokolova ve Lapalme, 2009).

Doğruluk, modelin tüm örnekler üzerindeki genel başarı oranını ifade eder ve aşağıdaki şekilde hesaplanır (Eşitlik 1).

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (\text{Eşitlik 1})$$

Duyarlılık, modelin pozitif sınıfları doğru tanıma oranını gösterir (Eşitlik 2).

$$\text{Sensitivity} = \frac{TP}{TP + FN} \quad (\text{Eşitlik 2})$$

Özgüllük, modelin negatif sınıfları doğru ayırt etme yeteneğini tanımlar (Eşitlik 3).

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (\text{Eşitlik 3})$$

F1 skoru, duyarlılık ve kesinlik (precision) arasındaki harmonik ortalamadır. Dengesiz veri setlerinde dengeyi sağlar (Eşitlik 4).

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (\text{Eşitlik 4})$$

Son olarak, ICBHI skoru, wheeze ve crackle sınıflarının dengeli olarak değerlendirilmesini amaçlayan özel bir ölçüttür. ICBHI skoru, her iki patolojik sınıfın duyarlılık ve özgüllük ortalamalarının aritmetik ortalaması alınarak hesaplanır (Eşitlik 5).

$$\text{ICBHI Score} = \frac{\text{Sensitivity}_{\text{avg}} + \text{Specificity}_{\text{avg}}}{2} \quad (\text{Eşitlik 5})$$

Bu eşitliklerde:

TP (True Positive): Modelin doğru biçimde pozitif (örneğin patolojik ses) olarak sınıflandırdığı örnek sayısı,

TN (True Negative): Modelin doğru biçimde negatif (örneğin normal ses) olarak sınıflandırdığı örnek sayısı,

FP (False Positive): Modelin hatalı biçimde pozitif (patolojik) olarak sınıflandırdığı, aslında negatif olan örnek sayısı,

FN (False Negative): Modelin hatalı biçimde negatif (normal) olarak sınıflandırdığı, aslında pozitif olan örnek sayısı,

Accuracy (Doğruluk): Toplam doğru tahminlerin (TP + TN) tüm örnekler içindeki oranı,

Sensitivity (Duyarlılık / Recall): Gerçek pozitif örneklerin ne kadarının doğru tespit edildiği,

Specificity (Özgüllük): Gerçek negatif örneklerin ne kadarının doğru tanımlandığı,

Precision (Kesinlik): Pozitif tahminlerin ne kadarının gerçekten pozitif olduğu,

F1 Skoru: Duyarlılık ve kesinliğin harmonik ortalaması,

$Sensitivity_{avg}$: Wheeze ve crackle sınıflarına ait duyarlılık (sensitivity) değerlerinin ortalaması,

$Specificity_{avg}$: Wheeze ve crackle sınıflarına ait özgüllük (specificity) değerlerinin ortalaması,

ICBHI Score: Duyarlılık ve özgüllük değerlerinin ortalamasına dayalı, solunum sesi sınıflandırması için özel olarak tanımlanmış değerlendirme metriğini göstermektedir (Bishop, 2006; Powers, 2020; Sokolova ve Lapalme, 2009).

3.6. Literatürün Ortak Noktaları

Literatürde genel eğilim, solunum seslerinin bilgisayar destekli analiziyle elde edilen bulguların klasik stetoskopla yapılan oskültasyona kıyasla daha nesnel, tekrarlanabilir ve eğitim/izlem süreçlerine daha elverişli olduğudur. Dijital kayıt ve standartlaştırılmış değerlendirme yaklaşımları, gözlemciler arası değişkenliği azaltırken patolojik örüntülerin zaman içinde karşılaştırmalı olarak izlenmesine olanak tanır (Bohadana vd., 2014; Pasterkamp vd., 1997; Rocha vd., 2017). Sistematik derlemeler de bu tabloyu doğrulamakta; özellikle derin öğrenme tabanlı yöntemlerin ICBHI 2017 gibi veri kümelerinde giderek baskın hâle geldiğini, ancak klasik yöntemlerin hâlâ önemli bir kıyas çizgisi olarak kullanıldığını vurgulamaktadır (Pramono vd., 2017; Sfayyih vd., 2023).

Klasik makine öğrenmesi hattında başarının kritik belirleyicisi öznitelik çıkarımıdır: MFCC, dalgacık dönüşümleri ve enerji/entropi ile sıfır geçiş oranı gibi

istatistiksel ölçüler, hışıltı (wheeze) ve ral (crackle) gibi örüntülerin ayırt edilmesinde en sık başvurulmuş temsillerdir. Bu doğrultuda MFCC tabanlı temsilin GMM ya da yapay sinir ağları gibi sınıflandırıcılarla iki veya üç sınıflı görevlerde yüksek doğruluklar üretebildiği; zaman–frekans ve zaman–ölçek temelli dalgacık özniteliklerinin ise özellikle kısa süreli çıkırtıları yakalamakta güçlü olduğu raporlanmıştır (Bahoura, 2009; Dokur vd., 2003; Kandaswamy vd., 2004; Serbes vd., 2017). Sınıflandırıcı seçimi de performansı anlamlı biçimde etkiler: SVM, KNN ve GMM gibi yöntemlerin iyi ölçeklenmiş ve uygun boyut indirgeme ile desteklenmiş öznitelik uzaylarında rekabetçi olduğu gösterilmiştir (Bishop, 2006; Hastie vd., 2009; Pramono vd., 2017). Tuncer vd. (2021) gibi güncel çalışmalar ise hibrit öznitelik setleri ve gelişmiş öznitelik seçimi kullanıldığında klasik hattın çok sınıflı senaryolarda da güçlü performans gösterebildiğini ortaya koymuştur.

Derin öğrenme tarafında, zaman–frekans temsillerini doğrudan işleyen CNN tabanlı mimariler ile artık bağlantılı yapılar (örneğin ResNet) ve ölçekleme kurallarıyla optimize edilmiş ağlar (örneğin EfficientNet) literatürde dikkat çekici sonuçlar vermiştir (Chen vd., 2019; Gairola vd., 2021; He vd., 2016; Tan ve Le, 2019). Yeni nesil çalışmalar, transformer tabanlı ağlar ve kokleogram temsilleriyle daha da ileri gitmiş; özellikle çok merkezli ve gürültülü veri koşullarında CNN'lere göre avantajlar bildirmiştir (Bae ve Kim, 2023; Latif vd., 2020; Sfayyih vd., 2023). Buna karşılık yüksek hesaplama maliyeti, veri gereksinimi ve yorumlanabilirlik eksikliği, bu modellerin klinik ölçeğe taşınmasını güçleştirmektedir. Bu nedenle açıklanabilir yapay zekâ araçlarının (örneğin sınıf etkinlik haritaları, dikkat/saliency analizleri) entegrasyonu ve hafif mimarilerin tercih edilmesi (mobil/gerçek zaman kısıtları için) giderek daha fazla önem kazanmaktadır (Doshi-Velez ve Kim, 2017; Rudin, 2019).

Veri boyutunda, ICBHI 2017 Respiratory Sound Database fiili bir kıyaslama zemini işlevi görmektedir. Ancak çok merkezli ve çok cihazlı yapısı, sınıf dengesizliği ile cihaz/koşul kaynaklı akustik farklılıklar nedeniyle genelleme gücünü zorlayabilir; aynı hastaya ait kayıtların eğitim ve test alt kümelerine birlikte düşmesi, başarıyı yapay biçimde yükseltebilir. Bu nedenle hasta düzeyi bölmelerin, resmî ICBHI ayrımının veya hasta düzeyi çapraz doğrulamanın kullanılması ve sonuçların doğruluk, duyarlılık, özgüllük, F1 ve ICBHI skoru gibi metriklerle birlikte açıkça raporlanması, adil ve tekrarlanabilir kıyaslama için temel bir gerekliliktir (Chambres vd., 2018; ICBHI Challenge, 2017; Rocha vd., 2019; Sfayyih vd., 2023).

3.7. Literatürdeki Boşluklar

Son yıllardaki ilerlemelere karşın, bilgisayar destekli solunum sesi analizinde bazı yapısal kısıtlar sürmektedir. En temel sorun, veri seti kısıtlılığı ile heterojenliğin aynı anda bulunmasıdır. Alanın fiilî standardı sayılan ICBHI 2017 veri seti farklı merkezler, cihazlar ve yaş gruplarından toplanmış geniş bir çeşitlilik sunsa da bu çeşitlilik sınıf dengesizliği, cihaz/koşul kaynaklı spektral farklılık ve kayıt protokolü değişkenliği nedeniyle genelleme gücünü zorlayabilmektedir (ICBHI Challenge, 2017; Pramono vd., 2017; Rocha vd., 2019). Kapsayıcı derlemeler, daha geniş ölçekli, iyi dengelenmiş ve çok merkezli yeni veri tabanlarının; aynı bireyden tekrarlı ölçümlerin, çocuk–erişkin ayrımını ve ek hastalıkları kapsayan anotasyonların gerekliliğini vurgulamaktadır (Fraïwan vd., 2021; Hsu vd., 2021; Sfayyih vd., 2023).

Görev tanımlarındaki sınırlılık da dikkat çekicidir. Literatürde iki sınıflı normal–anormal ayrımı veya ikili patolojik sınıflar yaygındır; oysa klinikte wheeze, crackle, stridor ve plevral sürtünme gibi bulgular aynı döngü içinde eşzamanlı görülebilir. Bu nedenle çok sınıflı ve çok etiketli (multi-label) çerçevelere, hatta “şiddet” ve “yaygınlık” benzeri derecelendirmeyi de içeren yarı-sürekli etiketlemelere duyulan ihtiyaç açıktır (Chen vd., 2019; Gairola vd., 2021; Sfayyih vd., 2023). Bugüne dek kullanılan değerlendirme metriklerinin çoğu nihai sınıf kararına odaklıdır; oysa klinikte olayı konum ve faz bazında (inspirasyon/ekspirasyon) raporlamak, tekrarlı döngülerde kalıcılık ve yayılımı göstermek kadar önemlidir. Bu da anotasyon protokollerinin standardizasyonunu ve uzmanlar arası uzlaşma (consensus) etiketlerinin raporlanmasını gerektirir (Hsu vd., 2021; Rocha vd., 2017).

Gerçek zamanlı ve taşınabilir uygulamaların görece azlığı, klinik etkiye giden yolu geciktiren bir başka boşluktur. Çok sayıda yöntem çevrimdışı senaryolarda değerlendirilmiş; yatak başında anlık geri bildirim sağlayan, mobil/ev içi izlem sistemleriyle bütünleşik ve enerji etkin çözümler sınırlı kalmıştır. Hesap yükü ve gecikme gereksinimleri bu noktada belirleyicidir; MobileNet, EfficientNet ve GhostNet gibi hafif mimarilerin ses görevlerine uyarlanması, model budama/nicemleme ve gerekirse bilgi damıtma (knowledge distillation) ile birlikte düşünüldüğünde pratik bir araştırma eksenini oluşturmaktadır (Han vd., 2020; Howard vd., 2017; Tan ve Le, 2019; Sfayyih vd., 2023). Ancak bu tür dağıtımlar için model kalibrasyonu, belirsizlik (uncertainty) tahmini ve alarm eşiği yönetimi gibi güvenlik katmanlarının da ölçülüp raporlanması gerekir; aksi hâlde yüksek doğruluklu modeller dahi klinik iş akışlarına güvenle entegre edilemez (Doshi-Velez ve Kim, 2017; Rudin, 2019).

Genelleme kapasitesi bağlamında veri sızıntısı riski hâlâ önemli bir sorun olarak öne çıkmaktadır. Aynı hastaya ait kayıtların hem eğitim hem test bölmelerine düşmesi performansı yapay olarak yükseltebilir ve klinik geçerliliği zayıflatabilir. Bu nedenle hasta düzeyi ayrımlar, resmî ICBHI bölmeleri ve leave-one-subject-out gibi protokoller standart uygulama hâline gelmeli; raporlar açık görev tanımı ve metrik setleriyle birlikte sunulmalıdır (Chambres vd., 2018; Pramono vd., 2017; Rocha vd., 2019). Ayrıca “haricî doğrulama”nın — farklı merkez/cihazdan bağımsız bir test kümesiyle denemenin — nadiren yapılması, yöntemin sahaya taşınabilirliğini değerlendirmeyi güçleştirir.

Açıklanabilirlik ve hibrit tasarımlar alanında da boşluk göze çarpar. Klasik yöntemlerin yorumlanabilirliği ile derin ağların temsil gücünü birleştiren hibrit yaklaşımlar sınırlıdır. Klinik kararın yalnız doğrulukla değil, anlaşılabilir bir gerekçeyle desteklenmesi beklenmektedir; bu kapsamda zaman–frekans temsilleri üzerinde sınıf etkinlik haritaları (Grad-CAM vb.) ve dikkat/saliency analizleri gibi açıklanabilir yapay zekâ (Explainable Artificial Intelligence – XAI) tekniklerinin hem eğitimde model denetimi hem de klinisyen-içi döngü (clinician-in-the-loop) kullanımında sistematik olarak raporlanması gereklidir (Doshi-Velez ve Kim, 2017; Rudin, 2019; Selvaraju vd., 2017; Sfayyih vd., 2023).

Etiket kalitesi ve gözlemciler arası uyum da çoğu çalışmada yeterince nicelleştirilmemiştir. Çok merkezli veri tabanlarında uzmanlar arası anlaşma katsayılarının (örneğin kappa) ve etiket belirsizliğinin modele nasıl yansıtıldığının raporlanmaması, üst sınır performansın yorumunu zorlaştırır. Anotasyonların faz bilgisiyle (inspirasyon/ekspirasyon), anatomik dinleme noktasıyla ve eşlik eden klinik bağlamla (yaş, ek hastalıklar, cihaz filtresi) birlikte verilmesi, hem eğitim hem değerlendirme için daha sağlam bir zemin sağlayacaktır (Bohadana vd., 2014; Hsu vd., 2021; Rocha vd., 2017).

Son olarak, veri adaleti ve önyargı (bias) başlığı henüz yeterince ele alınmamıştır. Farklı yaş, cinsiyet, beden yapısı ve etnik köken gruplarında göğüs duvarı iletimi ve temel ses tınısı belirgin farklılık gösterebilir; cihaz filtreleri ve dinleme noktası dağılımları da buna eklenince, modellerin belirli alt gruplarda görece performans kaybı yaşaması mümkündür. Bu nedenle alt grup metriklerinin sistematik raporlanması, dengesiz gruplar için yeniden ağırlıklandırma ve veri artırma stratejilerinin açıkça belirtilmesi; ayrıca model kararlarının kalibrasyon ve belirsizlikle birlikte sunulması önerilmektedir (Doshi-Velez ve Kim, 2017; Gairola vd., 2021; Rocha vd., 2019; Sfayyih vd., 2023).

Özetle, literatürün başlıca açıkları; ölçek ve denge açısından zengin, çok merkezli ve standart anotasyonlu veri kaynaklarının eksikliği; çok etiketli/çok sınıflı ve faz–konum

duyarlı etiketleme gereksinimi; gerçek zamanlı, enerji etkin ve kalibre edilebilir dağıtım senaryolarının azlığı; hasta düzeyi ve haricî doğrulama protokollerinin yeterince yaygınlaşmaması; açıklanabilirliğin ve klinisyen-içi döngü kullanımının sınırlı entegrasyonu; etiket kalitesi ile adalet/önyargı analizlerinin sistematik raporlanmamasıdır. Bu boşlukların giderilmesi hem yöntemlerin klinik geçerliliğini hem de sahadaki sürdürülebilir etkisini belirgin biçimde artıracaktır (Bohadana vd., 2014; Chambres vd., 2018; Doshi-Velez ve Kim, 2017; Gairola vd., 2021; Pramono vd., 2017; Rocha vd., 2017; Rocha vd., 2019; Rudin, 2019; Sfayyih vd., 2023).

3.8. Genel Değerlendirme

Literatürün genel görünümü, solunum sesi sınıflandırmasının klinik tanı ve izlem süreçlerinde güçlü bir tamamlayıcı araç olabileceğini göstermektedir. Klasik makine öğrenmesi yaklaşımları; özenli bir ön işleme (gürültü giderme, soluk evresi segmentasyonu, artefakt temizliği) ve ayırt edici öznitelik mühendisliğiyle (MFCC, dalgacık tabanlı enerji/entropi, istatistiksel betimler) küçük ve dengeli veri kümelerinde yüksek doğruluklara ulaşabilmektedir. MFCC ve dalgacık temelli çalışmalarda bildirilen bulgular, özellikle hışıltının dar bantlı sürekliliği ile ralın kısa ve geniş bantlı doğasının yakalanmasında bu hattın etkinliğini destekler niteliktedir (Bahoura, 2009; Dokur vd., 2003; Kandaswamy vd., 2004; Serbes vd., 2017). Bununla birlikte bu yöntemler çoğu kez iki sınıflı ya da sınırlı sınıf sayılı kurulumlarda öne çıkmakta; cihaz/merkez çeşitliliği ve hasta düzeyi ayrımların kullanıldığı daha heterojen koşullarda performansın seçilen öznitelik setine ve doğrulama protokolüne duyarlı hâle geldiği görülmektedir (Hastie vd., 2009; Pramono vd., 2017; Rocha vd., 2019).

Derin öğrenme yaklaşımları, öznitelik çıkarımını verinin kendisinden otomatik olarak öğrenmeleri sayesinde çok sınıflı ve heterojen veri koşullarında genellikle daha üstün bir genelleme gücü sergiler. Zaman-frekans temsillerini doğrudan işleyen CNN tabanlı mimariler ile artık bağlantılı yapılar (ResNet) ve ölçekleme kurallarıyla optimize edilmiş ağlar (EfficientNet) literatürde dikkat çekici sonuçlar vermiştir (Chen vd., 2019; Gairola vd., 2021; He vd., 2016; Tan ve Le, 2019). Transformer tabanlı AST/ViT mimarileri ve kokleogram temelli yaklaşımlar da, özellikle ICBHI 2017 gibi çok merkezli veri setlerinde yeni başarı düzeyleri (state-of-the-art) bildirmektedir (Bae ve Kim, 2023; Latif vd., 2020; Sfayyih vd., 2023). Öte yandan yüksek hesaplama maliyeti, veri gereksinimi ve yorumlanabilirlik eksiği, bu modellerin klinik ölçeğe taşınmasını güçleştirmektedir. Bu nedenle açıklanabilir yapay zekâ araçlarının entegrasyonu ve hafif mimarilerin (MobileNet, EfficientNet-B0, GhostNet gibi) tercih edilmesi, mobil ve

gerçek zaman kısıtları altında giderek daha önemli hâle gelmektedir (Han vd., 2020; Howard vd., 2017; Doshi-Velez ve Kim, 2017; Rudin, 2019).

Veri boyutunda ICBHI 2017 Respiratory Sound Database fiilî bir kıyaslama zemini işlevi görmektedir. Ancak çok merkezli ve çok cihazlı yapısı, sınıf dengesizliği ile cihaz/koşul kaynaklı akustik farklılıklar nedeniyle genelleme gücünü zorlayabilir; ayrıca aynı hastaya ait kayıtların eğitim ve test alt kümelerine birlikte düşmesi başarıyı yapay biçimde yükseltebilir. Bu nedenle hasta düzeyi bölmelerin, resmî ICBHI ayrımının veya hasta düzeyi çapraz doğrulamanın kullanılması; sonuçların doğruluk, duyarlılık, özgüllük, F1 ve ICBHI skoru gibi metriklerle birlikte açıkça raporlanması, adil ve tekrarlanabilir kıyaslama için temel bir gerekliliktir (Chambres vd., 2018; ICBHI Challenge, 2017; Rocha vd., 2019; Sfayyih vd., 2023).

Genel olarak klasik ve derin öğrenme yaklaşımları farklı güçlü yanlar sunar ve bağlama göre birbirini tamamlayabilir. Klasik yöntemler sınırlı veride daha hızlı, daha hafif ve daha yorumlanabilir çözümler üretirken; derin öğrenme modelleri büyük ve karmaşık veri ortamlarında daha yüksek doğruluk ve esneklik sağlayabilmektedir. Bu tez kapsamında iki, üç ve dört sınıflı görevlerde ortak ön işleme ve hasta düzeyi protokoller altında yapılacak sistematik karşılaştırmaların; yöntemlerin güçlü/zayıf yönlerini nesnel biçimde ortaya koyması, literatürdeki değerlendirme düzenekleriyle kıyaslanabilirliği güvence altına alması ve gelecekteki araştırmalara — örneğin çok etiketli kurulumlar, açıklanabilirlik, hafif/gerçek zaman dağıtımlar ve haricî doğrulama — yön verecek bütüncül bir çerçeve sunması beklenmektedir (Bahoura, 2009; Chen vd., 2019; Gairola vd., 2021; He vd., 2016; Pramono vd., 2017; Tan ve Le, 2019; Doshi-Velez ve Kim, 2017; Rudin, 2019; Sfayyih vd., 2023).

4. MATERYAL VE YÖNTEM

4.1. Veri Seti

Bu çalışmada Uluslararası Biyomedikal ve Sağlık Enformatik Konferansı kapsamında yayımlanan ICBHI 2017 Respiratory Sound Database kullanılmıştır. ICBHI, solunum sesi sınıflandırması alanında çok merkezli–çok cihazlı yapısıyla gerçek yaşam koşullarındaki çeşitliliği yansıtması ve standardize edilmiş etiket şeması nedeniyle alanın başat açık veri kaynağı olarak kabul edilmektedir (Rocha vd., 2017; ICBHI Challenge, 2017).

Kayıtların önemli bir bölümü Portekiz’de Aveiro Üniversitesi Sağlık Bilimleri Fakültesi (ESSUA) bünyesindeki Solunum Araştırma ve Rehabilitasyon Laboratuvarı (Lab3R) ile Hospital Infante D. Pedro’da, diğer kısmı ise Yunanistan’da Aristotle Üniversitesi/Coimbra Üniversitesi iş birliğiyle Papanikolaou Genel Hastanesi (Selanik) ve Imathia Genel Hastanesi (Naousa) kliniklerinde toplanmıştır. Böylece farklı hasta profilleri, klinik akışlar ve oda akustikleri tek bir havuzda birleştirilmiştir (Rocha vd., 2017; ICBHI Challenge, 2017).

Veri seti 126 bireye ait 920 ses dosyasından oluşur; tek bir dosyanın süresi yaklaşık 10–90 saniye arasında değişir ve toplam süre yaklaşık 5,5 saattir. Uzman anotasyonlarıyla her dosya içinde solunum döngülerinin başlangıç–bitiş anları işaretlenmiş, toplam 6.898 döngü tanımlanmıştır. Döngü seviyesinde etiket dağılımı; 1.864 crackle, 886 wheeze, 506 crackle+wheeze ve kalanının normal olacak şekildedir. Bu belirgin sınıf dengesizliği, yöntemlerin hasta düzeyi (patient level) doğrulama ve uygun dengeleme/artan veri stratejileriyle değerlendirilmesini zorunlu kılar (Rocha vd., 2017).

Kayıtlar trakea ile akciğerin ön/arka ve lateral sahalarından toplanmıştır. Dosya adlarında göğüs bölgesi ve cihaz bilgisi kısaltmalarla verilir; örneğin Tc (trakea), Pr (posterior sağ), kanal modu mc (çok kanallı) veya sc (tek kanallı) ve cihaz kodu (ör. LittC2SE, Meditron) gibi bilgiler adlandırma şablonunun parçasıdır. Her ses dosyasına eşlik eden .txt anotasyon dosyasında, her bir döngünün başlangıç zamanı, bitiş zamanı ile crackle ve wheeze varlığı ikili (0/1) olarak satır satır verilir. Bu yapının özeti ve gerçek dosyalardan alınan örnek satırlar Tablo 7’de sunulmaktadır. Tez boyunca veri hazırlama aşamalarında kullanılan alan adları ve kısaltmalar, aynı tabloda standartlaştırılmış biçimde verilmektedir (Rocha vd., 2017; ICBHI Challenge, 2017).

Kullanılan sensörler arasında AKG C417L mikrofon, 3M Littmann Classic II SE, 3M Littmann 3200 ve Welch Allyn Meditron Master Elite gibi stetoskop/mikrofon sistemleri yer alır. Cihaz çeşitliliği ve bölgeye göre farklı iletim özellikleri, özellikle yüksek frekans bileşenleri ve zemin gürültüsü profillerinde değişkenliğe yol açar; bu nedenle çalışmamızda bant geçiren ve oluk (notch) filtreleri, kalp sesi bastırma ve artefakt temizliği gibi ön işleme adımları tüm deneylerde aynı parametrelerle standardize edilmiştir (Rocha vd., 2017).

Son olarak, ICBHI topluluğunca önerilen değerlendirme kuralları gereği, aynı kişiye ait kayıtların eğitim ve test bölümlerine eşzamanlı düşmemesi için hasta düzeyi bölme uygulanmış; performans ölçütleri doğrulukla birlikte duyarlılık, özgüllük, F1 ve ICBHI skoru ile raporlanmıştır. Bu uygulamalar, heterojen veri koşullarında genellenebilirlik ve adil karşılaştırma için kritik önemdedir (ICBHI Challenge, 2017; Rocha vd., 2017).

Tablo 7. ICBHI 2017 veri kümesinde dosya adı adlandırma şeması ve örnek anotasyon kayıtları

Dosya Adı	Göğüs Bölgesi	Kanal	Cihaz	Döngü Başlangıç (s)	Döngü Bitiş (s)	Crackle (0/1)	Wheeze (0/1)	Açıklama
135_2b2_Tc_mc_LittC2SE.txt	Tc (trakea)	mc	LittC2SE	0.036	1.979	0	0	Normal
135_2b2_Tc_mc_LittC2SE.txt	Tc	mc	LittC2SE	1.979	4.693	0	0	Normal
135_2b2_Tc_mc_LittC2SE.txt	Tc	mc	LittC2SE	4.693	7.593	0	0	Normal
078_1b1_Pr_sc_Meditron.txt	Pr (arka sağ)	sc	Meditron	0.000	2.210	1	0	Sadece crackle
078_1b1_Pr_sc_Meditron.txt	Pr	sc	Meditron	2.210	4.480	0	1	Sadece wheeze

Tablo 7. (Devamı)

Dosya Adı	Göğüs Bölgesi	Kanal	Cihaz	Döngü Başlangıç (s)	Döngü Bitiş (s)	Crackle (0/1)	Wheeze (0/1)	Açıklama
078_1b1_Pr_sc_Meditron.txt	Pr	sc	Meditron	4.480	6.735	1	1	Crackle + wheeze
078_1b1_Pr_sc_Meditron.txt	Pr	sc	Meditron	6.735	9.015	0	0	Normal

4.2. Segmentasyon ve Etiketleme

ICBHI veri kümesindeki her ses kaydı birden çok solunum döngüsü içerdiğinden, ilk adım kayıtların döngü düzeyinde kesitlere ayrılmasıdır. Bu işlem, veri setiyle birlikte sağlanan ve her döngünün başlangıç–bitiş zamanlarını saniye cinsinden veren anotasyon dosyaları kullanılarak gerçekleştirildi; böylece ham .wav dosyaları zaman ekseninde dilimlenip her biri tek bir solunum döngüsünü temsil eden segmentlere dönüştürüldü (ICBHI Challenge, 2017; Rocha vd., 2017). Örnek bir kayıt üzerinde segmentasyon akışı Şekil 11’de gösterilmektedir.



Şekil 11. Ses kaydının segmentlere ayrılması

Segmentlerin sınıf etiketleri, aynı anotasyon dosyalarında döngü bazında işaretlenen “crackle” ve “wheeze” göstergeleri esas alınarak atandı. Sınıf atama kuralı şu şekilde tanımlandı: yalnızca crackle işaretli olan döngüler Crackles, yalnızca wheeze işaretli olan döngüler Wheezes, her iki göstergenin de işaretlenmediği döngüler Normal ve her iki göstergenin de birlikte işaretli olduğu döngüler Both (crackle+wheeze) olarak etiketlendi. Uygulama kolaylığı açısından bu dört sınıf sırasıyla 0, 1, 2 ve 3 kodlarıyla ifade edildi (ICBHI Challenge, 2017). Sınıfların açıklamaları ve segment sayıları Tablo 8’de sunulmaktadır.

Model eğitiminde anlamsız örneklerin etkisini azaltmak için çok kısa ve çok düşük genlikli kesitler önceden elendi. Bu amaçla, süresi 0.1 s’den kısa olan ya da RMS genliği

ön incelemelerle belirlenen eşiğin altında kalan segmentler analiz dışında bırakıldı; böylece mekanik tıklamalar, kayda giren sessiz boşluklar ve solunum dışı mikro-artefaktların modele taşınması önleildi. Kalite güvencesi kapsamında, her segmentin dalga formu ve Mel-spektrogramı görsel olarak incelendi; seçilen örnekler gerektiğinde işitsel olarak da doğrulandı. Segmentlere ilişkin yardımcı bilgiler (hasta kimliği, kayıt bölgesi, kullanılan cihaz, sınıf etiketi, süre) tek bir metadata tablosunda tutuldu; bu yapı, ilerleyen bölümlerde raporlanan istatistiksel özetler ve sınıf dengesizliği analizleri için referans olarak kullanıldı (Rocha vd., 2017).

Segmentasyon sonrasında sınıf dağılımı belirgin biçimde dengesizdir: 1.864 adet Crackles, 886 adet Wheezes, 3.642 adet Normal ve 506 adet Both segmenti elde edilmiştir. Normal sınıfın sayısal üstünlüğü ile özellikle Both sınıfının azlığı, bir sonraki bölümde ayrıntıları verilen veri artırma adımlarını gerekli kılmıştır (Rocha vd., 2017; ICBHI Challenge, 2017).

Tablo 8. Sınıf etiketleri, açıklamaları ve segment sayıları

Etiket Kodu	Sınıf Adı	Açıklama	Segment Sayısı
0	Crackles	Yalnızca crackle sesi içeren döngü	1.864
1	Wheezes	Yalnızca wheeze sesi içeren döngü	886
2	Normal	Patolojik ses içermeyen döngü	3.642
3	Both	Hem crackle hem de wheeze seslerini içeren döngü	506

4.3. Ön İşleme

Klinik ortamda kaydedilen solunum seslerinde cihaz, yerleşim ve ortam koşullarına bağlı değişkenlik yüksek olduğundan, sınıflandırma modellerine girdi verilmeden önce gürültünün azaltılması, bilgi içeriği yüksek bantların korunması ve genlik ölçeklerinin hizalanması amacıyla standart bir ön işleme hattı uygulanmıştır. Solunum bilgisinin büyük kısmı 100–2000 Hz aralığında yoğunlaştığından, alt bantta kalp/hareket artefaktları ve üst bantta çevresel parazitlerin etkisi sistematik biçimde bastırılmıştır. Boru hattının adımları ve temel parametreleri Tablo 9’da özetlenmiş, dalga formu ve zaman-frekans temsili üzerindeki etkiler Şekil 12 ve Şekil 13’te gösterilmiştir.

4.3.1. Örnekleme Hızı Birleştirme ve DC Offset Kaldırma

Kayıtlar tek bir hedef örnekleme hızında (8 kHz) birleştirildi; yeniden örnekleme sırasında aliasing'i önlemek için düşük geçiren ön filtre kullanıldı. DC ofset (sinyalin sıfır çizgisine kaymış olması) her segment için ortalamasının çıkarılmasıyla giderildi (Eşitlik 6)

$$x_{corr}(t) = x(t) - \mu_x \quad (\text{Eşitlik 6})$$

Bu eşitlikte:

$x(t)$: orijinal sinyali,

μ_x : sinyalin ortalamasını temsil etmektedir.

$x_{corr}(t)$: DC bileşeni giderilmiş sinyali ifade etmektedir.

Tek örnekleme hızı, tüm filtre ve spektrogram parametrelerini sabitler. DC bileşeni temizlemek, düşük frekansta yapay enerji oluşmasını ve crackle/wheeze imzalarının bozulmasını engeller.

4.3.2. Bant Geçiren Filtreleme

Solunum bilgisini korumak ve artefaktları bastırmak amacıyla 5. dereceden Butterworth bant geçiren filtre 100–2000 Hz kesimleriyle uygulanmıştır. Faz bozulmasını engellemek için ileri-geri (sıfır fazlı, filtfilt) uygulama tercih edilmiştir. Kayıtlarda elektrik şebekesi uğultusu izlendiğinde 50/60 Hz için dar bantlı notch filtresi eklenmiş, gerekiyorsa 100/120 Hz harmonikleri de bastırılmıştır. Kesim frekansları ve filtre derecesi, wheeze gibi dar bantlı çizgisel yapıları ve ince rallerin üst harmoniklerini zayıflatmayacak biçimde ayarlanmıştır.

4.3.3. Gürültü Azaltma

Bant geçiren filtreden sonra kalan düzensiz bileşenleri azaltmak için iki aşamalı yaklaşım kullanılmıştır. Öncelikle stetoskop sürtünmesi gibi kısa süreli, geniş bant darbeleri 5–15 ms pencereli medyan filtre ile yumuşatılmıştır. Ardından arka plan gürültüsü için kayıt içi koşullara göre aşağıdaki yöntemlerden biri uygulanmıştır:

- Spektral çıkarma: Nefes aralarında kestirilen gürültü spektrumu, her çerçevenin genlik spektrumundan aşırı çıkarma katsayısı $\alpha=1.0-1.5$ ile düşülmüş; spektral zemin yaklaşık -20 dB tabanla sınırlandırılmıştır.
- Wiener tabanlı zayıflatma: Frekans-bağımlı SNR kestirimi ile kazanç katsayıları hesaplanmıştır.

STFT sızıntısını azaltmak ve segment sınırlarında tıslama oluşumunu engellemek üzere baş ve son kısımlara 5–10 ms Hann fade-in/out uygulanmıştır. Parametreler, wheeze’in dar bant çizgilerini ve crackle’ın kısa süreli geniş bant vuruşlarını koruyacak şekilde görsel (spektrogram) ve işitsel denetimle ayarlanmıştır.

Kalite güvencesi kapsamında, süresi < 0.1 s olan, RMS değeri dağılımın alt %10’unda kalan veya kliplenme ($|x| > 0.98$) saptanan segmentler analize dâhil edilmemiştir.

4.3.4. Normalizasyon

Kaynak ve yerleşim farklılıklarından doğan genlik değişkenliğini azaltmak için iki kademeli ölçekleme uygulanmıştır:

- z-skor ile standart sapmaya göre ölçekleme (istatistikler yalnızca eğitim kümesinden hesaplanıp doğrulama/testte sabitlenmiştir)
- hiperbolik tanjant ile yumuşak sıkıştırma
İşlem aşağıdaki eşitlikle verilmiştir (Eşitlik 7)

$$x_{\text{norm}}(t) = \tanh\left(\frac{x(t)}{\sigma_x}\right) \quad (\text{Eşitlik 7})$$

Bu eşitlikte:

$x(t)$: Orijinal ses sinyalini,

σ_x : Sinyalin standart sapmasını,

$x_{\text{norm}}(t)$: Normalizasyon uygulanmış sinyali ifade etmektedir.

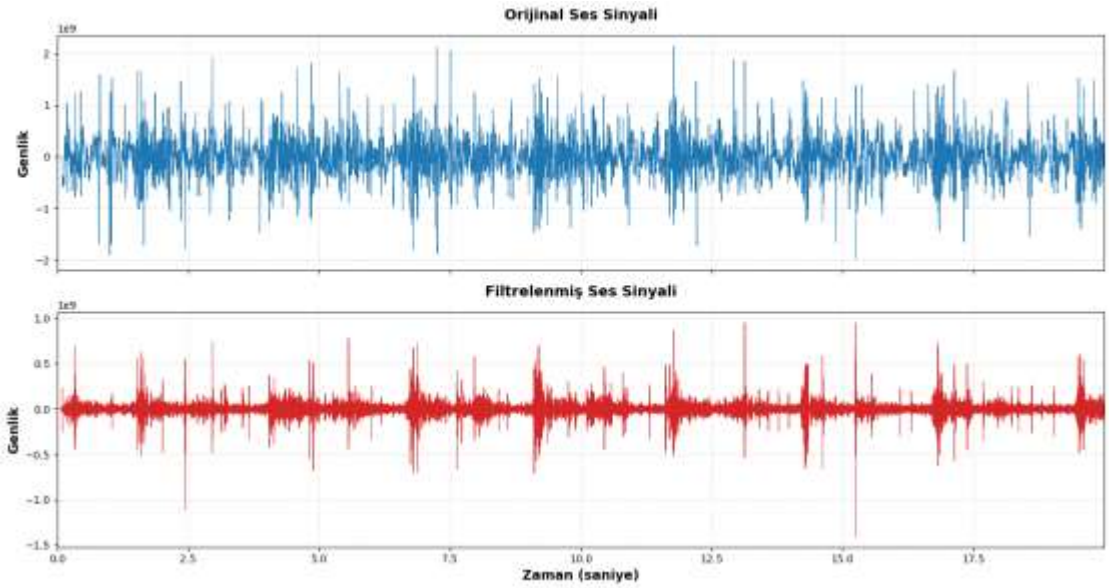
Bu yöntem sayesinde, ani yüksek genlikli gürültüler bastırılmış, düşük genlikli segmentler ise daha belirgin hale getirilmiştir.

4.3.5. Görselleştirme ve Doğrulama

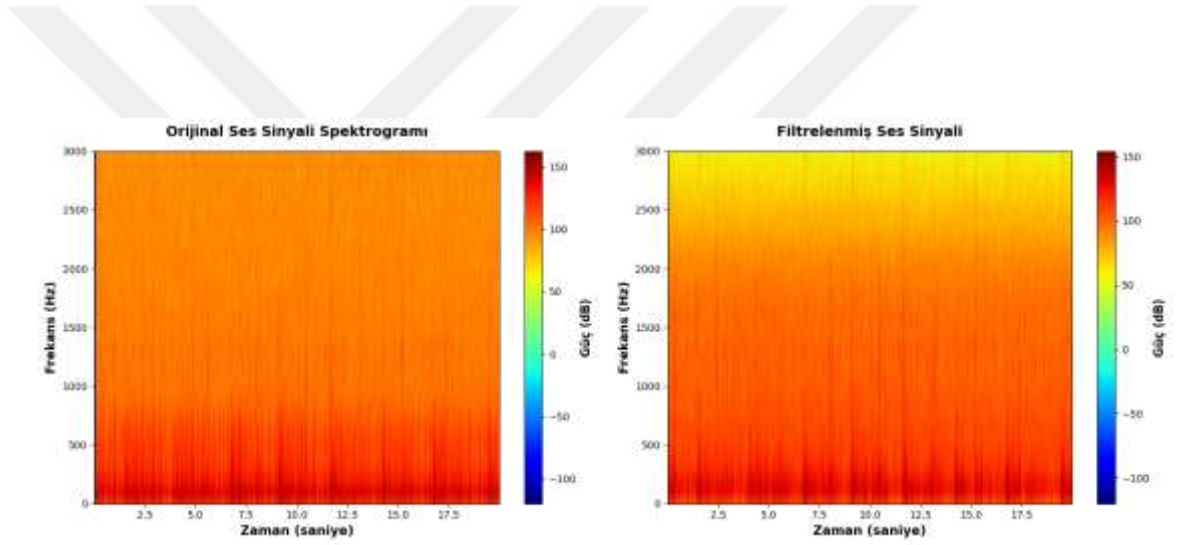
Ön işlemenin etkileri dalga formu ve spektrogram üzerinde sistematik olarak doğrulanmıştır. Zaman alanında taban çizgisi kayıklığının giderildiği ve darbe-tipi parazitlerin azaldığı; zaman-frekans alanında ise düşük frekans uğultuları ve yüksek frekans parazitlerinin zayıfladığı gözlenmiştir. Rastgele seçilen örneklerde işitsel doğrulama yapılarak wheeze çizgilerinin sürekliliği ve crackle patlarının zamansal keskinliği korunmuştur. Tüm ön işleme parametreleri ve sürüm bilgileri tekrarlanabilirlik amacıyla proje metadata dosyalarında saklanmıştır.

Tablo 9. Ön işleme hattı ve temel parametreler

Adım	Amaç	Ayar / Parametre	Uygulama / Not
Örnekleme birleştirme	Tutarlılık	8 kHz; aliasing'i önleyici alçak geçiren filtre	Tüm kayıtlar tek örnekleme hızına indirgenir.
DC ofset giderimi	Düşük frekans sızıntısını önleme	Ortalama çıkarma (Eşitlik 6)	Her segmentin ortalaması çıkarılarak taban çizgisi sıfıra yaklaştırılır.
Bant geçiren filtre	Klinik bantta odaklanma	5. dereceden Butterworth, 100–2000 Hz, sıfır fazlı (filtfilt)	Solunum bilgisinin yoğun olduğu bant korunur; faz bozulması engellenir, harmonikler zayıflamaz.
Notch filtre	Şebeke uğultusunu bastırma	50/60 Hz, $Q \approx 30$; gerekirse 100/120 Hz harmonikleri	Kayıt-bağımlı olarak elektrik şebekesi gürültüsü zayıflatılır.
Darbe yumuşatma	Sürtünme/klik artefaktlarını azaltma	Medyan filtre (5–15 ms pencere)	Stetoskop sürtünmesi ve kısa süreli geniş bant darbeler yumuşatılır.
Arka plan gürültüsü	Zemin gürültüsünü bastırma	Spektral çıkarma (aşırı çıkarma katsayısı ≈ 1.0 – 1.5 ; taban ≈ -20)	Kayıt koşullarına göre arka plan gürültüsü azaltılır; yöntem kayıt-bağımlı seçilir.
Sınır yumuşatma	STFT sızıntısını azaltma	5–10 ms Hann “fade-in/fade-out”	Segment başı ve sonunda oluşan keskin geçişler yumuşatılır.
Kalite filtresi	Anlamsız örnekleri eleme	Süre < 0.1 s; RMS alt %10 (P10); kliplenme kontrolü	Çok kısa, çok düşük genlikli veya kliplenmiş segmentler analiz dışı bırakılır.
Normalizasyon	Genlik ölçeklerini hizalama	z-skor (eğitim istatistikleriyle) + tanh sıkıştırma (Eşitlik 7)	Cihaz/yerleşim kaynaklı genlik farkları azaltılır; aşırı yüksek genlikler yumuşatılır.
İzlenebilirlik	Tekrarlanabilirlik	Parametre ve sürüm bilgisinin kaydı	Tüm ön işleme adımları ve kullanılan ayarlar metadata dosyalarında saklanır.



Şekil 12. Bir segmentin orijinal ve ön işlenmiş (filtrelenmiş) dalga formu karşılaştırması



Şekil 13. Bir segmentin orijinal ve ön işlenmiş (filtrelenmiş) spektrogram karşılaştırması

4.4. Özellik Çıkarma

Bu bölümde solunum sinyalinin klinik açıdan ayırt edici öğelerini hem kompakt vektörler (MFCC, Δ , Δ^2) hem de ayrıntılı zaman–frekans temsilleri (Mel-spektrogram) üzerinden görünür kılan bir öznitelik kümesi oluşturulmuştur. Tasarımın temel hedefleri; (i) bilgi yoğun bantların vurgulanması, (ii) zamansal geçişlerin yakalanması, (iii) cihaz ve yerleşim farklılıklarından kaynaklanan ölçek sapmalarının azaltılmasıdır. Tüm işlem hattı sabit örnekleme hızında çalışacak biçimde standardize edilmiştir.

4.4.1. Mel Frekans Kepstral Katsayıları (MFCC)

Sinyal kısa zamanlı çerçevelere ayrılmış (pencere 64 ms, Hamming; örtüşme $\approx$ %64; atım $\approx$ 23 ms), her çerçevede güç tayfı elde edilmiştir. Tayf, Mel ölçekli filtre bankası üzerinden log-enerjilere dönüştürülmüş; ardından Ayırık Kosinüs Dönüşümü

(DCT) uygulanarak 40 kepsral katsayı çıkarılmıştır. Kepsral alanın aşırı yüksek sıralı bileşenleri dengelemek üzere liftering (L=22) kullanılmış; log-enerjiler $[-80, 0]$ dB aralığına kırılarak aykırı piklerin etkisi sınırlanmıştır. Çerçeve bazındaki 40 katsayı, segment düzeyinde kararlı bir temsil elde etmek amacıyla zaman boyunca ortalama ile havuzlanmış ve 40 boyutlu bir vektöre indirgenmiştir. Matematiksel olarak, k 'nıncı MFCC katsayısı aşağıdaki eşitlik yardımıyla hesaplanmıştır (Eşitlik 8).

$$c_k = \Sigma_{m=1}^M \left[\log(E_m) \cos \left(\left(\frac{\pi k}{M} \right) \left(m - \frac{1}{2} \right) \right) \right], k = 1, 2, \dots, K \quad (\text{Eşitlik 8})$$

Bu eşitlikte:

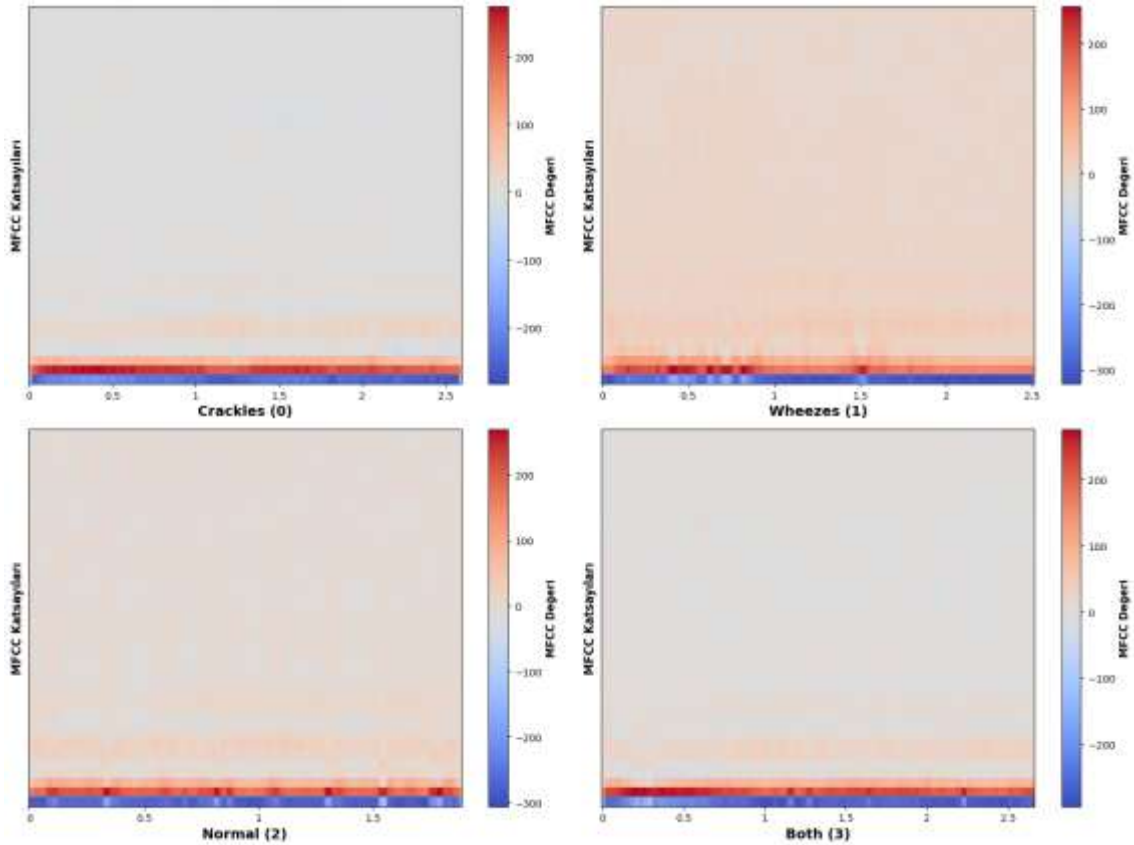
E_m : m 'inci Mel filtresinden elde edilen enerjiyi,

M : Toplam Mel filtre sayısını,

K : Çıkarılan MFCC katsayılarının sayısını,

c_k : k 'nıncı MFCC katsayısını ifade etmektedir.

Çalışmada $K=40$ olarak seçilmiştir. Sınıflar arası tipik kepsral örüntüler Şekil 14'te sunulmuştur.



Şekil 14. Dört farklı sınıfa ait MFCC ısı haritaları

4.4.2. Mel-Spektrogram

Zaman–frekans çözümlemesi için Kısa Zamanlı Fourier Dönüşümü (STFT) gerçekleştirilmiş ($n_fft=512$, pencere 64 ms, atım ≈ 23 ms), elde edilen güç tayfı 128 bantlı Mel ölçeğine projekte edilmiştir. Elde edilen değerler log-güç alanına aktarılmıştır. Segmentler tekdüzelik sağlamak amacıyla 5 saniyeye hizalanmış; kısa segmentler sıfır doldurma ile genişletilmiş, uzun segmentler merkezden kırpılmıştır. Böylece tüm örnekler için 128×216 boyutunda zaman–frekans tensörleri elde edilmiştir. Bu temsil, CNN tabanlı mimarilere doğrudan giriş sağlayacak düzenliliği ve çözünürlüğü sunmaktadır (Şekil 15). Mel-spektrogram şu şekilde tanımlanır (Eşitlik 9).

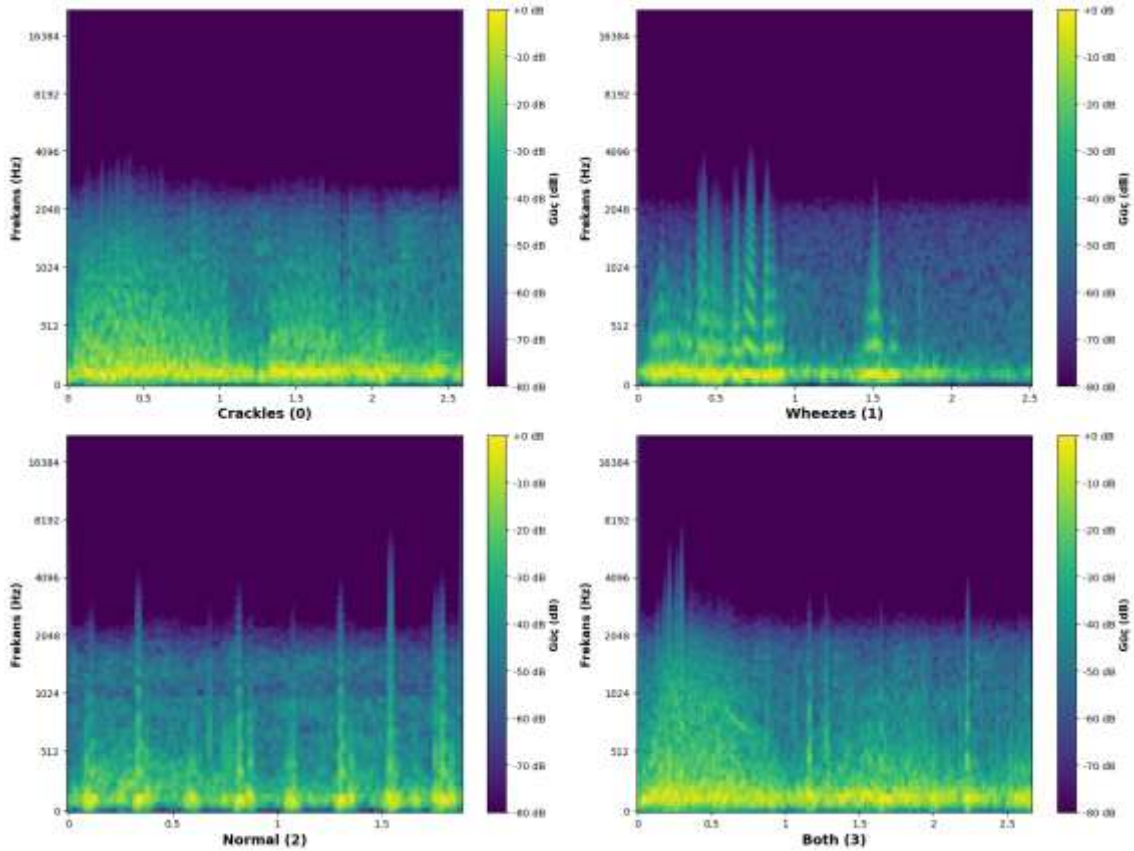
$$S_{mel}(f, t) = 10 \cdot \log_{10} \left(\sum_{k=f_{min}}^{f_{max}} |X(k, t)|^2 \cdot H_{mel}(f, k) \right) \quad (\text{Eşitlik 9})$$

Bu eşitlikte:

$X(k, t)$: kısa süreli Fourier dönüşümünden elde edilen karmaşık spektrum,

$H_{mel}(f, k)$: Mel filtresinin ağırlık fonksiyonu,

$S_{mel}(f, t)$: logaritmik Mel-spektrogram değeridir.



Şekil 15. Dört farklı sınıfa ait örnek Mel-spektrogram görselleri

Her bir alt görsel, Normal, Crackles, Wheezes ve Crackles+Wheezes sınıflarına ait birer örnek solunum segmentinin Mel-spektrogram temsillerini göstermektedir. Renk yoğunlukları, zaman-frekans düzleminde enerjinin dağılımını ifade etmektedir. Normal solunum seslerinde enerji dağılımı genellikle homojen bir yapı sergilerken, wheeze sınıfında orta ve yüksek frekans bantlarında süreklilik gösteren paralel tonlar görülmektedir.

4.4.3. Delta ve Delta-Delta Katsayıları

MFCC katsayıları, sesin frekans bilgisini temsil etmesine rağmen zamansal değişimleri yeterince yakalayamaz. Bu nedenle, sinyalin zaman içindeki eğilimini modellemek amacıyla Delta (Δ) ve Delta-Delta (Δ^2) katsayıları hesaplanmıştır. Bu katsayılar, sırasıyla birinci ve ikinci türev bilgilerini temsil eder ve özellikle solunum seslerindeki ani değişimlerin yakalanmasında önemli rol oynar.

Delta katsayıları, her bir zaman anındaki MFCC katsayılarının doğrusal değişimini ifade eder. Delta katsayıları aşağıdaki eşitlik yardımıyla hesaplanmıştır (Eşitlik 10).

$$\Delta c_t = \frac{\sum_{n=1}^N n(c_{t+n} - c_{t-n})}{2 \sum_{n=1}^N n^2} \quad (\text{Eşitlik 10})$$

Delta-Delta katsayıları ise Delta değerlerinin zaman içerisindeki değişimini, yani ivmeyi temsil eder. Delta-Delta katsayıları aşağıdaki eşitlik yardımıyla hesaplanmıştır (Eşitlik 11).

$$\Delta^2 c_t = \frac{\sum_{n=1}^N n(\Delta c_{t+n} - \Delta c_{t-n})}{2 \sum_{n=1}^N n^2} \quad (\text{Eşitlik 11})$$

Bu eşitliklerde:

c_t : t anındaki MFCC katsayısını,

N : Hesaplama kullanılan pencere yarıçapını,

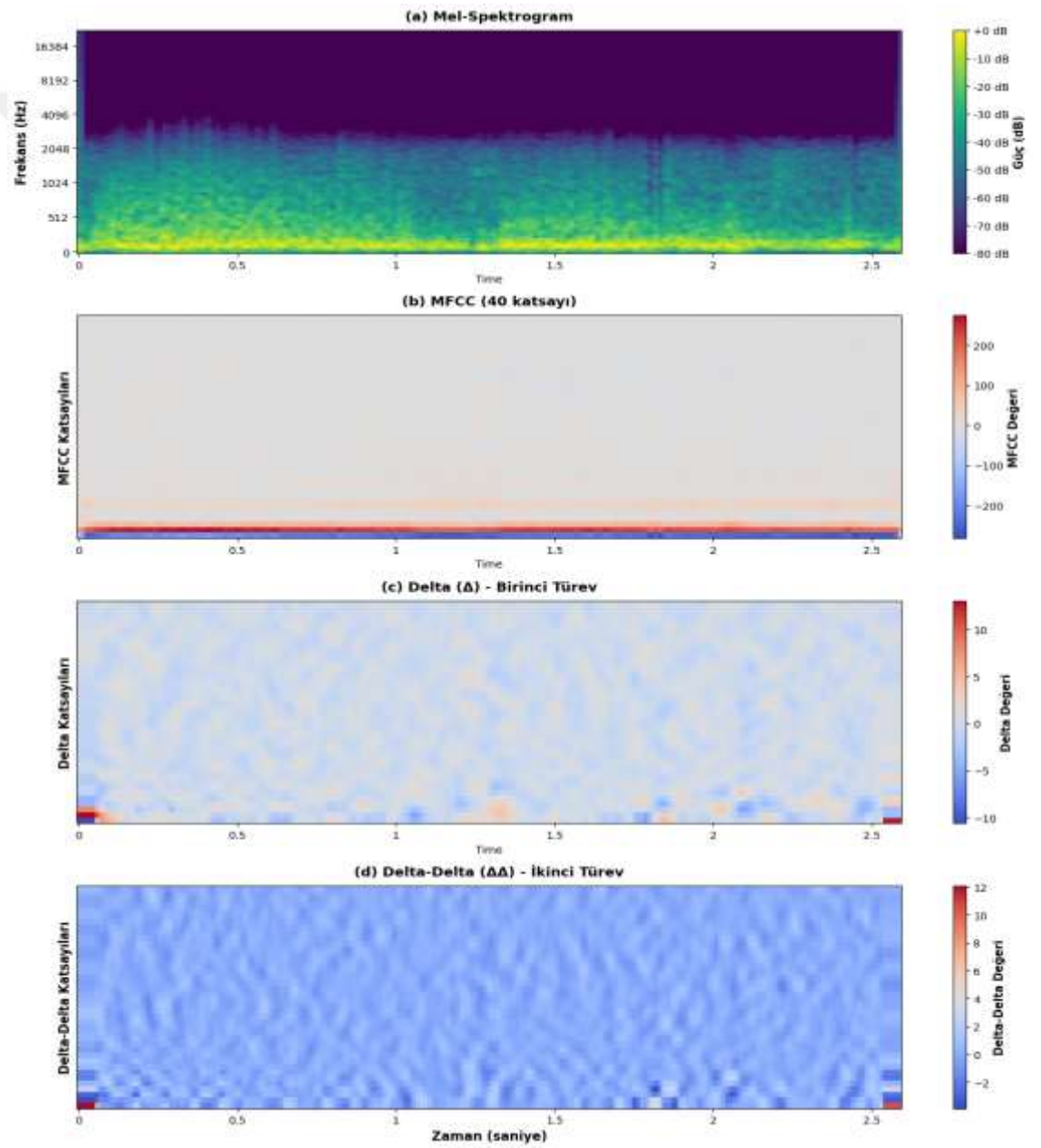
Δc_t : MFCC katsayılarının birinci türevini,

$\Delta^2 c_t$: MFCC katsayılarının ikinci türevini ifade etmektedir.

Delta ve Delta-Delta katsayıları, solunum seslerindeki spektral değişim hızını ve yönünü temsil ederek modelin dinamik bileşenleri daha iyi öğrenmesini sağlar. Özellikle crackle gibi ani ve kısa süreli anomalilerin tespitinde, Delta özelliklerinin sınıflandırma başarısını önemli ölçüde artırdığı gözlemlenmiştir.

Görselde, solunum sesinin zaman-frekans yapısı ve bu yapıdan türetilen özniteliklerin hiyerarşik dönüşümü sunulmaktadır.

Mel-spektrogram, sinyalin enerji yoğunluğunu frekans bandına göre gösterirken; MFCC, bu enerjiyi insan işitme algısına uygun bir logaritmik ölçeğe dönüştürmektedir. Delta katsayıları MFCC değerlerindeki kısa süreli değişimleri (eğimleri) temsil ederken, Delta-Delta katsayıları bu değişimlerin ivmesini göstermektedir. Bu yapı, özellikle patolojik seslerdeki ani frekans dalgalanmalarının (örneğin crackle) ve süreklilik gösteren harmonik yapıların (örneğin wheeze) modellenmesine olanak tanımaktadır.

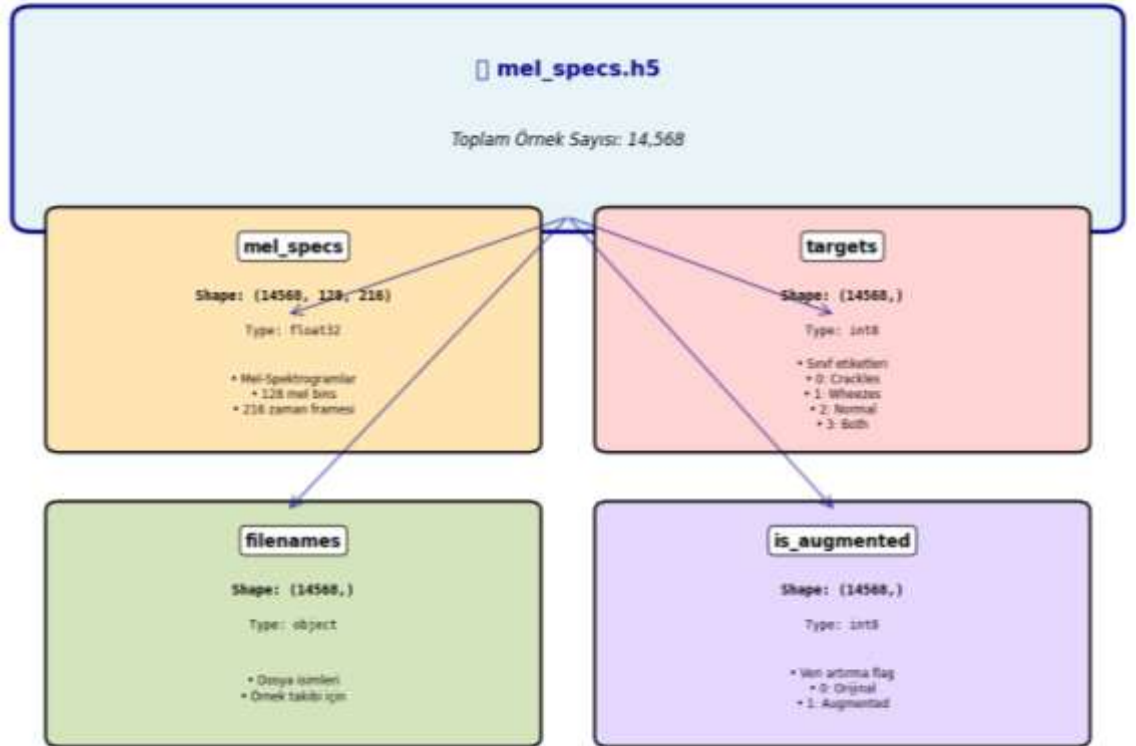


Şekil 16. Bir solunum segmentine ait Mel-spektrogram, MFCC, Delta ve Delta-Delta katsayılarının karşılaştırmalı görselleştirilmesi

4.4.4. Özelliklerin Kaydedilmesi ve Eğitim İçin Hazırlık

Öznitelikler çıkarıldıktan sonra doğrudan eğitime geçmek yerine, veriyi hızlı okunabilir ve tekrarlanabilir bir düzene sokmak amacıyla tek bir HDF5 dosyası altında birleştirildi. HDF5, büyük veriyle çalışırken sunduğu sıkıştırma, bütünlük ve rastgele erişim olanakları nedeniyle tercih edildi. Dosya, özünde iki ana bölümden oluşur: “özellikler” ve “metadata”. Özellikler bölümünde her segment için log-dB ölçeğinde normalize edilmiş Mel-spektrogramlar ile MFCC+ Δ + Δ^2 vektörleri saklanır; metadata bölümünde ise segment etiketleri (0: Crackles, 1: Wheezes, 2: Normal, 3: Crackles+Wheezes), kaynak dosya adları ve örneğin orijinal mi yoksa artırma ile üretilmiş mi olduğunu belirten bir bayrak yer alır.

Bu düzen, eğitim sırasında tek merkezden ve ardışık okumayla verinin GPU’ya beslenmesini kolaylaştırır, gereksiz disk erişimlerini azaltır ve deneylerin aynı veriye aynı sırayla erişmesini sağlayarak tekrarlanabilirliği güçlendirir. Çalışmada kullanılan tekil dosya, toplam 14.568 segment içerir; Mel-spektrogram temsilleri 128 frekans bandı ve 5 saniyelik sabit uzunluk için 216 zaman penceresiyle tutulmuştur. Böylece hem klasik makine öğrenmesi hem de CNN tabanlı derin öğrenme modelleri için tutarlı ve verimli bir giriş biçimi elde edilmiştir.



Şekil 17. Çalışmada kullanılan mel_specs.h5 dosyasının içeriğine ait yapısal analiz

4.4.5. Özelliklerin Görselleştirilmesi

Elde edilen öznelikler yalnızca eğitim için değil, ön işleme hattının beklenen biçimde çalıştığını doğrulamak amacıyla da görselleştirildi. Dalga formu incelemeleri, DC düzeltme ve bant geçiren filtre sonrasında düşük frekanstaki dalgalanmaların ve yüksek frekans parazitlerinin belirgin biçimde azaldığını; solunum döngüsü sınırlarının ise daha net seçildiğini gösterdi. Mel-spektrogramlar sınıflar arasındaki tipik desenleri ortaya koydu: normal solunumda enerji dağılımı daha homojen seyrederken, wheeze örneklerinde orta–yüksek bantlarda süreklilik gösteren dar çizgiler, crackle içeren örneklerde ise kısa süreli, geniş bantlı parlak adacıklar izlendi. MFCC ısı haritalarında düşük sıralı katsayılar genel tınıyı, üst sıralı katsayılar daha ince ayrıntıları yansıtırken; Δ ve Δ^2 türevleri zamansal geçişleri ve ani değişimleri vurguladı.

Karşılaştırılabilirlik için tüm görsellerde eksen aralıkları, pencere uzunluğu/örtüşme ve renk haritası sabit tutuldu; güç değerleri log-dB ölçeğine dönüştürülerek dinamik aralıklar hizalandı. Her sınıftan rastgele seçilen örnekler hem görsel hem işitsel olarak kontrol edildi; özellikle veri artırma ile üretilen segmentlerde yapay artefakt oluşumu olup olmadığı ayrı bir incelemeyle değerlendirildi. Bu görsel doğrulamalar, filtreleme–normalizasyon–özellik çıkarımı zincirinin tutarlı çalıştığını ve sınıflar arası ayırt edici ipuçlarının zaman-frekans temsillerinde yeterince görünür hâle geldiğini doğruladı.

4.4.6. Veri Dengesizliği ve Veri Artırma (Augmentation) Uygulaması

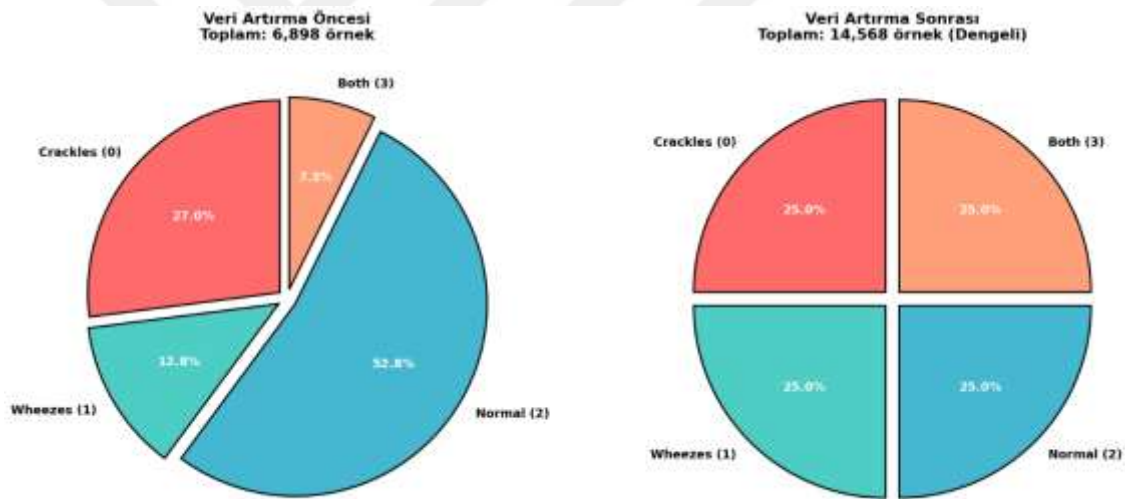
ICBHI veri setinde sınıflar arasındaki belirgin dengesizlik, özellikle “Normal” sınıfının fazlalığı ve “Crackles+Wheezes” sınıfının azlığı nedeniyle, modellerin çoğunluk sınıfına eğilim göstermesine yol açmaktadır. Bu etkiyi azaltmak için veri artırma yalnızca eğitim kümesine uygulanmış; doğrulama ve test kümeleri dokunulmamış bırakılmıştır. Ayrıca hasta düzeyi ayırım korunmuş, aynı kişiye ait kayıtların artırılmış kopyaları eğitim dışındaki kümelere aktarılmamıştır.

Uygulanan artırmalar dalga-biçimi (waveform) alanında, öznelik çıkarımından önce gerçekleştirilmiş ve her artırılmış örnek, özgün etiketini koruyacak şekilde kaydedilmiştir (metadata’da `is_augmented=1`). Kullanılan teknikler ve tipik ayar aralıkları aşağıda özetlenmiştir:

- Zaman kaydırma (time shifting): Segment başlangıçları ± 50 – 250 ms aralığında rastgele ötelenmiştir; taşan kısım uçtan uca sarmalanmış veya kısa sessizlikle doldurulmuştur.

- Gürültü ekleme: Karışım SNR'si 10–20 dB aralığında olacak şekilde düşük seviyeli beyaz/pembe gürültü eklenmiştir.
- Pitch shifting: Perde ± 0.5 –2 yarım ses (semitone) aralığında küçük kaydırmalarla değiştirilmiş, süre korunmuştur.
- Zaman ölçeklendirme (time stretching): Faz-vokoder temelli yöntemle hız 0.9 – $1.1\times$ arasında değiştirilmiş, perde sabit tutulmuştur.
- Rastgele kazanç (random gain): Genlik ölçeği 0.8 – 1.2 aralığında rastgele ayarlanmış, doygunluk oluşmaması için kırpma/normlizasyon uygulanmıştır.

Her segment için en fazla iki artırma zincirlenmiş, işlemlerin uygulanma olasılıkları denge gözetilerek rastgeleştirilmiştir (örn. her operasyon için $p \approx 0.5$). Bu süreç sonunda her sınıf 3.642 örneğe tamamlanmış, toplam segment sayısı 14.568'e yükselmiştir. Böylelikle eğitim verisi sınıflar arasında dengelenmiş, modellerin azınlık sınıflarındaki duyarlılığı artıracak bir dağılım elde edilmiştir.



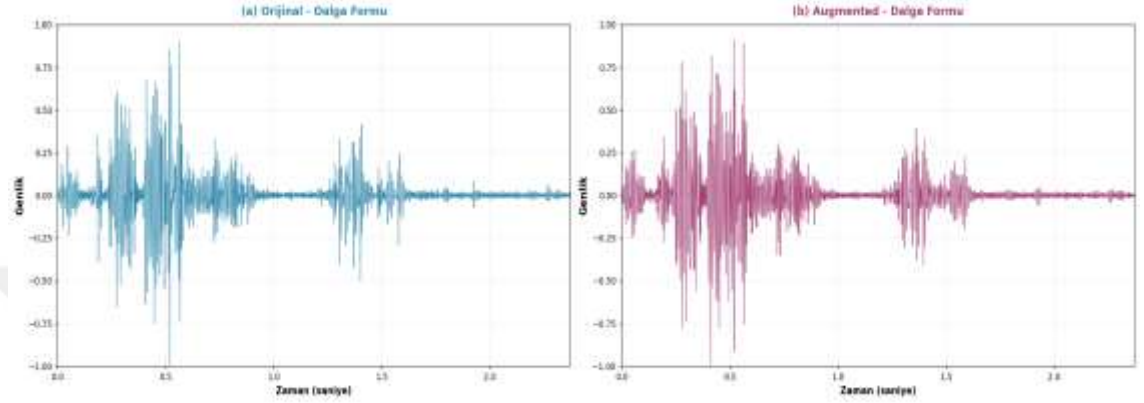
Şekil 18. Veri artırma öncesi ve sonrası sınıf dağılımı

Tablo 10. Veri artırma öncesi ve sonrası sınıf dağılımları

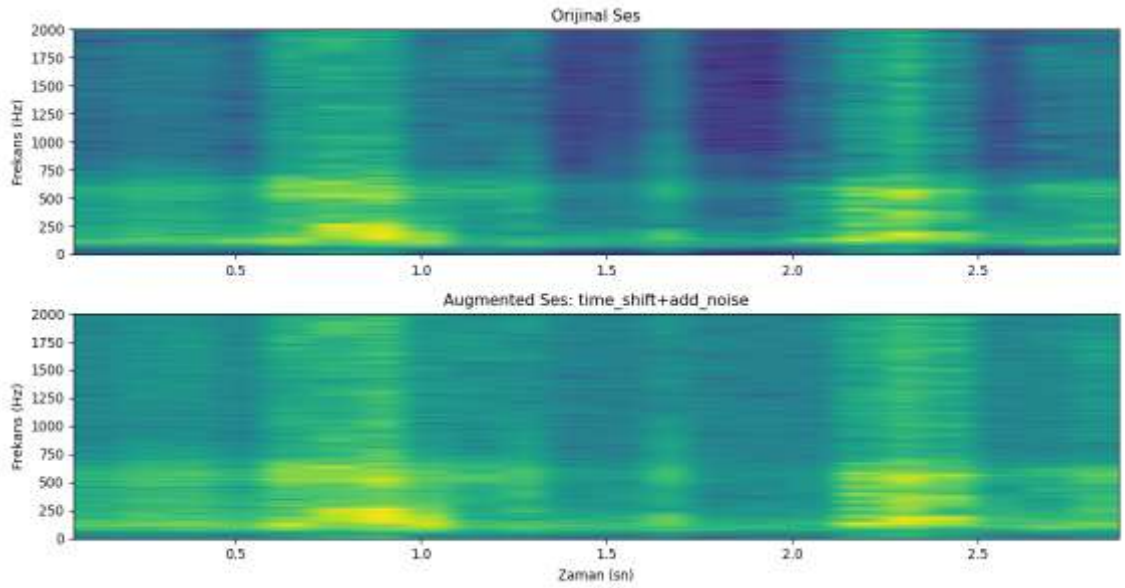
Sınıf	Tanım	Artırma Öncesi (Segment Sayısı)	Artırma Sonrası (Segment Sayısı)	Artırma Sonrası Oran (%)
0	Crackles	1.864	3.642	%25.00
1	Wheezes	886	3.642	%25.00
2	Normal	3.642	3.642	%25.00

Tablo 10 (Devamı)

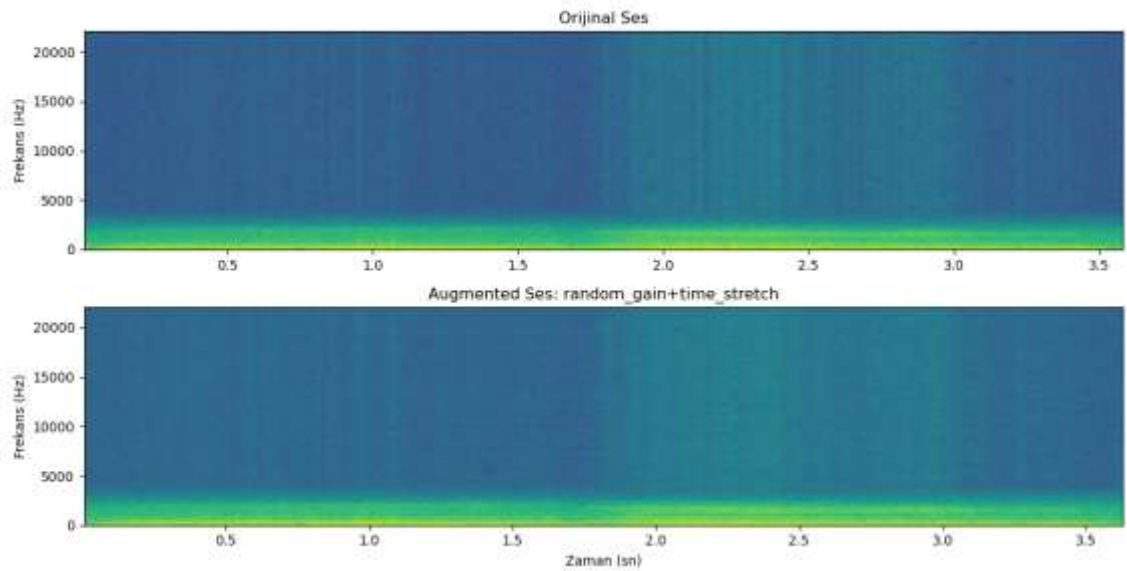
Sınıf	Tanım	Artırma Öncesi (Segment Sayısı)	Artırma Sonrası (Segment Sayısı)	Artırma Sonrası Oran (%)
3	Crackles + Wheezes	506	3.642	%25.00
Toplam		6.898	14.568	%100



Şekil 19. Veri artırma sonrası orijinal ve üretilmiş ses segmentlerinin dalga formu karşılaştırması



Şekil 20. Zaman kaydırma ve gürültü ekleme teknikleri kullanılarak gerçekleştirilen veri artırma sonrasında, orijinal ve üretilmiş ses segmentlerinin spektrogram karşılaştırması



Şekil 21. Rastgele kazanç ve zaman esnetme teknikleri kullanılarak gerçekleştirilen veri artırma sonrasında, orijinal ve üretilmiş ses segmentlerinin spektrogram karşılaştırması

Tablo 11. Solunum sesi özellik çıkarma; adımlar, ayarlar ve tasarım gerekçesi

Aşama	Seçilen ayar	Gerekçe / Ne sağlıyor
Örnekleme standardizasyonu	Tüm kayıtlar tek örnekleme hızına dönüştürüldü	Cihaz farklarından doğan süre/frekans ölçek kaymalarını azaltır; tüm adımların tutarlı çalışmasını sağlar.
STFT (zaman-frekans analizi)	$n_{fft}=512$; pencere: 64 ms Hamming; atım ≈ 23 ms	Zaman ve frekans çözünürlüğü dengeler; kısa olaylar (crackle) ile süreklilik gösteren bantları (wheeze) birlikte görünür kılar.
Mel-spektrogram	128 Mel bant; 100–2000 Hz; log-dB ölçeği; 5 s $\rightarrow 128 \times 216$	Solunumun bilgi yoğun bandını öne çıkarır; log-dB ile dinamik aralık hizalanır; CNN’ler için sabit boyutlu 2B giriş üretir.
MFCC	40 katsayı; DCT + liftering (L=22)	Spektral zarfı kompakt biçimde temsil eder; yüksek sıralı gürültülü bileşenler liftering ile dengelenir.
Delta / Delta-Delta	Pencere yarıçapı N=2	Zamansal eğim ve ivmeyi yakalar; kısa süreli ani olaylara ve sürekliliğe duyarlılığı artırır.
Çerçeveden segmente havuzlama	Zaman ortalaması (frame $\rightarrow$ segment)	Çerçeve gürültüsünü yumuşatır; segment başına $40 + 40 + 40 = 120$ boyutlu kararlı vektör üretir.

Tablo 11. (Devamı)

Aşama	Seçilen ayar	Gerekçe / Ne sağlıyor
Normalizasyon	CMVN (yalnız eğitim istatistikleri)	Cihaz/yerleşim kaynaklı genlik farklarını törpüler; veri sızıntısı olmadan testte aynı dönüşüm uygulanır.
Depolama biçimi	mel_specs.h5 (features + metadata)	Hızlı/tekrarlanabilir I/O; Mel-spektrogram (128×216) ve 120-D MFCC- Δ - Δ^2 vektörleri tek dosyada, etiket ve artırma bayraklarıyla birlikte saklanır.

4.5. Modelleme ve Sınıflandırma Yaklaşımları

Ön işleme ve özellik çıkarımı adımlarından elde edilen temsil-ler, bu bölümde hem klasik makine öğrenmesi hem de derin öğrenme tabanlı sınıflandırıcılarla değerlendirilmiştir. Tüm kurulumlarda veri artırma yalnızca eğitim kümesine uygulanmış; doğrulama ve test kümeleri değiştirilmeden bırakılmıştır. Hasta düzeyi ayırım korunarak aynı kişiye ait segmentlerin farklı kümelere düşmesi engellenmiştir.

4.5.1. Eğitim ve Test Ayırımı

ICBHI'deki 126 birey hasta-bağımsız biçimde ayrılmış; katılımcıların yaklaşık %80'i eğitime, %20'si teste atanmıştır. Sınıf dağılımı her iki kümede de benzer tutulacak şekilde atama birey düzeyinde stratifiye edilmiştir. Eğitim kümesinin %10'u doğrulama için ayrılmış; ölçekleme/normalizasyon istatistikleri sadece eğitim verisinden hesaplanıp aynen doğrulama ve test verisine uygulanmıştır. Veri artırma yalnızca eğitim kümesinde etkinleştirilmiş, böylece olası bilgi sızıntısı önlenmiştir. Bu düzenek, test üzerinde tarafsız genelleme değerlendirmesi yapmamıza imkân verir.

4.5.2. Klasik Makine Öğrenmesi Yöntemleri

MFCC + Δ + Δ^2 vektörleri girdi olarak kullanıldı. Tüm modeller çalıştırılmadan önce özellikler, yalnızca eğitim kümesinden hesaplanan istatistiklerle z-puanı (StandardScaler) kullanılarak ölçeklendirildi; aynı dönüşüm doğrulama ve test bölmelerine aynen uygulandı. Hiperparametre seçimi hasta düzeyi stratifiye 10-katlı çapraz doğrulama ile yalnızca eğitim verisi üzerinde gerçekleştirildi.

Destek Vektör Makineleri (SVM). Doğrusal ve RBF çekirdekler sınıandı; nihai kurulumda RBF (C=10, gamma="scale") tercih edildi. ROC/AUC ve kalibrasyon gereksinimleri için probability=True ile Platt ölçeklemesi etkinleştirildi. Böylece yüksek

boyutlu uzayda esnek ve pürüzsüz karar yüzeyleri elde edilerek sınıf sınırları ayrıntılı biçimde modellendi.

Lojistik Regresyon (LR). Çok sınıflı softmax (multinomial) formülasyon ve lbfgs çözücü ile `max_iter=1000` altında çalıştırıldı. Yorumlanabilir katsayılar ve kararlı optimizasyon sayesinde güçlü bir taban çizgisi oluşturdu.

XGBoost. Doğrusal olmayan ayrımları ağaç tabanlı güçlendirme ile yakaladı. Nihai kurulum: `n_estimators=200`, `learning_rate=0.1`, `max_depth=6`, `subsample=0.8`, `colsample_bytree=0.8`, değerlendirme ölçütü `mlogloss`. Az örnekli sınıflarda dahi ayrıştırıcı kurallar üretmeyi hedefledi.

Gradient Boosting (GB). Art arda eklenen zayıf öğrencilerle artık hatayı azalttı. Nihai kurulum: `n_estimators=300`, `learning_rate=0.05`, `max_depth=5`, `subsample=0.8`. Küçük öğrenme oranı ve örnekleme ile kapasite kontrollü tutuldu.

k-En Yakın Komşu (KNN). Parametrik olmayan yapısıyla yerel komşuluk bilgisinden yararlanıldı. `n_neighbors=7`, `weights='distance'`, `metric='minkowski'` ($p=2$, Öklit) ayarlarıyla, karar marjının sınıf yoğunluğu tarafından aşırı etkilenmemesi için mesafe ağırlıklı oylama kullanıldı. Bu seçim, wheeze gibi süreklilik içeren örüntülerde yakın komşu kümelerini etkili biçimde değerlendirme amacı taşıdı.

Naive Bayes (GaussianNB). Basit ve hızlı bir olasılıksal temel çizgi sağlamak üzere tercih edildi. Özelliklerin sınıf koşullu Gauss varsayımı altında GaussianNB ile olasılık tahminleri üretildi; ROC/AUC analizleri için `predict_proba` doğrudan kullanıldı. Özellikle düşük hesaplama maliyeti ve beklenmedik derecede güçlü ayırım gücüyle karşılaştırma referansı sundu.

LightGBM. Büyük gradyan kazanımlı ağaçları yaprak odaklı büyütme stratejisiyle verimli biçimde eğitti. Nihai kurulum: `boosting_type='gbdt'`, `n_estimators=300`, `learning_rate=0.05`, `max_depth=-1`, `subsample=0.8`, `colsample_bytree=0.8`, `random_state=42`. Derin ağaçlara izin veren `max_depth=-1` ile birlikte örnek ve özellik alt-örnekleme, genelleme performansını dengeledi.

Rastgele Orman (RF). Birbirinden bağımsız çok sayıda karar ağacının topluluk kararı kullanıldı. Nihai kurulum: `n_estimators=300`, `max_depth=25`, `min_samples_split=5`, `min_samples_leaf=2`, `max_features='sqrt'`, `n_jobs=-1`. Ağaç çeşitliliği ve özellik alt-örnekleme, gürültüye dayanıklılığı artırırken; özellik önemi (feature importance) analiziyle yorumlanabilirlik sağlandı.

DeneySEL kurulum özeti, kullanılan girdi temsilleri, başlıca hiperparametreler ve tercih gerekçeleri Tablo 12’de güncellenmiş hâliyle sunulmuştur.

Tablo 12. Klasik makine öğrenmesi modelleri: girdi temsili, başlıca hiperparametreler ve seçime dair kısa gerekçe

Model	Girdi	Temel Ayarlar	Gerekçe
SVM (RBF)	MFCC+ Δ + Δ^2	C=10, gamma=scale, probability=True	Esnek karar yüzeyi, yüksek ayırım gücü
Lojistik Reg.	MFCC+ Δ + Δ^2	multinomial, lbfgs, max_iter=1000	Hızlı, kararlı, katsayıları yorumlanabilir
XGBoost	MFCC+ Δ + Δ^2	n_est=200, lr=0.1, max_depth=6, subs=0.8, colsample=0.8, eval=mlogloss	Güçlü doğrusal-olmayan ayırım, kademeli hata azaltımı
Grad. Boosting	MFCC+ Δ + Δ^2	n_est=300, lr=0.05, max_depth=5, subs=0.8	Kontrollü kapasiteyle ardışık iyileştirme
KNN	MFCC+ Δ + Δ^2	n_neighbors=7, weights=distance, metric=minkowski (p=2)	Parametrik olmayan, yerel komşuluk bilgisiyle esnek
Naive Bayes	MFCC+ Δ + Δ^2	GaussianNB (varsayılan ayarlar)	Çok hızlı temel çizgi, olasılıksal yorum
LightGBM	MFCC+ Δ + Δ^2	gbdt, n_est=300, lr=0.05, max_depth=-1, subs=0.8, colsample=0.8, random_state=42	Verimli yaprak-odaklı büyüme, güçlü genelleme
Random Forest	MFCC+ Δ + Δ^2	n_est=300, max_depth=25, min_split=5, min_leaf=2, max_features='sqrt'	Gürültüye dayanıklı topluluk, özellik önemiyle açıklanabilir

4.5.3. Derin Öğrenme Yöntemleri

Bu aşamada zaman-frekans temsilleri doğrudan iki boyutlu giriş olarak kullanıldı ve derin ağların otomatik öznitelik çıkarma kapasitesinden yararlanıldı. Tüm modellerde girişler, ön işleme bölümünde üretilen log-Mel-spektrogramlar (128 bant $\times$ 216 zaman çerçevesi) ve MFCC haritaları olacak şekilde standartlaştırıldı; her kanal eğitim kümesinin ortalaması ve standart sapmasıyla normalize edildi. Eğitim sırasında erken durdurma ve model-seçimi yalnızca doğrulama bölmesi üzerinden yapıldı; test kümesi

hiçbir aşamada görülmedi. Optimizasyon için Adam kullanıldı, başlangıç öğrenme hızı $1e-3$ seçildi ve doğrulama kaybı iyileşmediğinde öğrenme hızı kademeli olarak düşürüldü. Aşırı uyumu sınırlamak amacıyla bırakma (dropout), L2 düzenleme ve veri artırmadan gelen dengeli dağılım birlikte kullanıldı; eğitimler sabit rastgele tohumlarla tekrarlanabilir şekilde yürütüldü.

4.5.3.1. Evrişimli Sinir Ağları (CNN)

Mel-spektrogramlardan uzamsal (frekans–zaman) örüntü çıkarımı için 2B konvolüsyon katmanları kullanıldı. Blok yapısı Conv(3×3, padding=“same”)–BatchNorm–ReLU dizilerinden oluştu; blok sonlarında MaxPool(2×2) ile boyutsal indirgeme yapıldı. Zaman çözünürlüğünü korumak gereken ilk aşamalarda stride=1 ve/veya asimetrik havuzlama (2×1) tercih edildi; böylece kısa süreli geçici bileşenler silinmeden üst katmanlara aktarıldı. Tam bağlantılı katman yerine Global Average Pooling kullanılarak parametre sayısı düşürüldü; sınıflandırma başlığı Dropout(≈0.4)–Dense(4, softmax) şeklinde kuruldu. Ağırlık başlatma He normal, aktivasyon ReLU olarak bırakıldı. Üç kanallı girişte ilk katman 3 kanala, tek kanalda 1 kanala göre uyarlandı.

CNN yalnız tek bir kalıp yerine birden fazla omurga (backbone) ailesiyle sınılandı: ResNet-18/34 (rezidüel bloklar), MobileNetV2/V3 (depthwise separable + inverted residual; V3’te SE blokları), EfficientNet-B0 (MBConv + SE, bileşik ölçekleme), GhostNet (ghost module ile düşük maliyetli özellik üretimi). Tüm omurgalarda ilk katman giriş kanalına göre düzenlendi; orta katmanlar değişmeden bırakıldı, başlık kısmı standartlaştırıldı (GAP–Dropout–Softmax). Giriş boyutlarına ve stride seçimlerine bağlı gerçek FLOPs/parametre farklılık göstermekle birlikte karşılaştırmalı özet Tablo 13’te verilmiştir.

Tablo 13. Mel-spektrogram girişi için CNN omurgalarının mimari özeti ve pratik kurulum

Omurga	Blok tipi (öz)	Stride/Havuzlama (erken evre)	Başlık (sınıflandırıcı)	Kısa not
ResNet-18 / 34	Rezidüel Conv–BN–ReLU + skip	stride=1; 2×1 → 2×2	GAP → Dropout(0.4) → Softmax	Zaman çözünürlüğünü korur, orta maliyet

Tablo 13. (Devamı)

Omurga	Blok tipi (öz)	Stride/Havuzlama (erken evre)	Başlık (sınıflandırıcı)	Kısa not
MobileNetV2 / V3	Depthwise separable, inverted residual (V3:	Düşük stride; 2x2 / 2x1	GAP → Dropout(0.4) → Softmax	Düşük hesap maliyeti, gerçek-zaman
EfficientNet-B0	MBConv + SE, bileşik ölçekleme	İlk blok stride=1; 2x2	GAP → Dropout(0.4) → Softmax	Parametre/verim dengesi iyi
GhostNet	Ghost module + pointwise	2x2 / 2x1	GAP → Dropout(0.4) → Softmax	En düşük hesap maliyeti

4.5.3.2. Tekrarlayan Sinir Ağları (RNN – LSTM)

Zamansal bağımlılıkları daha iyi yakalamak için MFCC dizileri ardışık giriş olarak iki katmanlı, iki yönlü LSTM mimarisine verildi. Her LSTM katmanında 128 gizli birim kullanıldı; girdiye ve yinelemeli bağlantılara ayrı bırakma (ör. 0.2/0.2) uygulanarak genelleme iyileştirildi. LSTM çıkışları zaman üzerinde ortalama alınarak sıkıştırıldı ve sınıflandırma başlığına aktarıldı. Bu yapı özellikle süreklilik taşıyan wheeze örüntülerinde ve nefes fazları arasında uzanan geçişlerin yakalanmasında avantaj sağladı.

4.5.3.3. Karma Mimari (CRNN)

Spektral dokuları yerel olarak yakalayıp, ardından zaman boyunca bağımlılıkları modellemek için konvolüsyonel özek çıkarıcı ile tekrarlayan katmanları birleştiren bir CRNN tasarlandı. İlk aşamada iki ya da üç konvolüsyon–havuzlama bloğu ile frekans eksenini kısmen toplanıp gürültü seyreldi; ara özellik haritaları zaman eksenini korunacak biçimde yeniden şekillendirilerek iki yönlü LSTM katmanına aktarıldı. LSTM çıkışlarının üzerine genel ortalama havuzlama ve tek gizli katmanlı bir sınıflandırıcı eklendi. Bu yaklaşım, crackle’ın kısa patlamaları ile wheeze’in sürekliliğini aynı çatı altında bir arada modelleyerek CNN ve LSTM’nin tamamlayıcı güçlü yanlarından faydalanmayı amaçladı.

4.5.4. Model Eğitim Süreci

Bu çalışmada dört sınıflı (Normal, Crackles, Wheezes, Crackles+Wheezes) problem kategorik çapraz entropi kaybı ile uçtan uca eğitildi; etiketler tek-sıcak (one-hot)

biçiminde kodlandı. Optimizasyon için Adam kullanıldı ve başlangıç öğrenme oranı 0.001 olarak ayarlandı. Öğrenme oranı doğrulama kaybı iyileşmediğinde yarıya düşecek şekilde otomatik olarak kademeli azaltıldı; böylece eğitim kararlılığı artırıldı. Parti boyutu 32 seçildi. Maksimum 100 epoch sınırı kondu ve erken durdurma ile doğrulama kaybı 10 ardışık epoch boyunca iyileşmezse eğitim sonlandırılıp en iyi ağırlıklar geri yüklendi. Aşırı öğrenmeyi sınırlamak için derin modellerde bırakma (dropout) ve gerektiğinde hafif L2 düzenleme uygulandı; gradyan patlamalarını önlemek amacıyla gradyan kırpması kullanıldı. Eğitim verisi her epoch başında karıştırıldı; tüm veri akışı tekrarlanabilirlik için sabit bir tohumla başlatıldı.

Veri dengesizliği iki aşamalı ele alındı. Veri artırma öncesi denemelerde sınıf ağırlıkları kullanılarak azınlık sınıflarına ek önem verildi. Nihai deneylerde veri artırma ile sınıflar dengelendikten sonra sınıf ağırlıkları kapatıldı ve doğrulama kümesi üzerindeki iyileşme yalnızca mimari/hiperparametre etkisini yansıtacak biçimde izlendi. Artırma işlemleri yalnızca eğitim kümesine uygulandı; doğrulama ve test kümeleri dokunulmadan bırakıldı.

Değerlendirme, hasta-bağımsız ayırım korunarak yalnızca bağımsız test kümesi üzerinde yapıldı; model seçimi ve tüm hiperparametre ayarları yalnız doğrulama kümesine bakılarak belirlendi. Raporlanan ölçütler, genel doğruluk ile birlikte sınıf başına duyarlılık (recall) ve özgüllük, makro-F1 ve ICBHI skorudur; çok sınıflı karışıklık matrisi ayrıca sunuldu. Çok sınıflı ROC/AUC gerektiren karşılaştırmalarda bir-e-karşı-tümü yaklaşımıyla sınıf başına eğriler çıkarıldı. Tüm klasik ve derin öğrenme modelleri aynı eğitim-doğrulama-test düzeni, aynı ön işleme/özelliğe çıkarım hattı ve aynı artırma politikasıyla eğitilerek adil karşılaştırma sağlandı.

4.5.5. Değerlendirme Metrikleri

Model performansı yalnızca genel doğrulukla değil; kesinlik (precision), duyarlılık/geri çağırma (recall/sensitivity), F1 skoru ve karışıklık matrisi üzerinden çok boyutlu olarak değerlendirildi. Çok sınıflı kurulumda (Normal, Crackles, Wheezes, Crackles+Wheezes) tüm metrikler hem sınıf başına hem de makro ortalama (sınıflara eşit ağırlık) olarak raporlandı; ek olarak sınıf dağılımını yansıtmak amacıyla ağırlıklı ortalama değerler de verildi. Karışıklık matrisi “satırlar gerçek, sütunlar tahmin” düzeninde hem mutlak sayılarla hem de gerçek sınıfa göre normalize edilmiş yüzdelerle sunuldu; böylece hangi sınıflar arasında karışma olduğuna ilişkin ayrıntılı bir görünüm sağlandı.

ICBHI ile kıyaslanabilirlik için ayrıca ICBHI Challenge Score hesaplandı. Bu ölçüt, yalnızca wheeze ve crackle sınıfları için elde edilen duyarlılık ve özgüllüklerin

ortalamalarından türetilir ve bu iki ortalamanın aritmetik ortalaması olarak raporlanır; normal ve “wheeze+crackle” sınıfları ICBHI skoruna dâhil edilmez.

Tüm metriklerin resmi tanımları ve formülleri Bölüm 3.5’te (Eşitlik 1–5) verilmiştir. Değerlendirme, hasta-bağımsız ayırım korunarak yalnızca bağımsız test kümesi üzerinde tek seferde yapılmış; model seçimi ve hiperparametre ayarı ise yalnız doğrulama kümesi sonuçlarına göre gerçekleştirilmiştir.



5. BULGULAR VE TARTIŞMA

Bu bölümde, ICBHI 2017 solunum sesi veri seti üzerinde yürütülen deneylerin bulguları sunulmakta ve yorumlanmaktadır. Derin öğrenme tabanlı yaklaşımlar ile klasik makine öğrenmesi algoritmaları aynı eğitim–doğrulama–test protokolü altında karşılaştırılmış; veri artırma uygulamalarının etkileri ayrı ayrı incelenmiştir. En yüksek başarıyı veren GhostNetV4 modeline ait karışıklık matrisi ile ROC/PR eğrileri üzerinden ayrıntılı davranış analizi de bölümün ilerleyen kısımlarında yer almaktadır.

5.1. DeneySEL Sonuçların Genel Özeti

Tüm deneyler, kişi-bağımsız (patient-independent) olacak şekilde düzenlenmiştir: aynı bireye ait kayıtlar eğitim ve testte bir araya getirilmemiş; katılımcıların yaklaşık %80'i eğitime, %20'si teste ayrılmış, eğitim kümesinin %10'u doğrulama amaçlı kullanılmıştır. Kayıtlar önce solunum döngülerine ayrılmış, ön işleme hattında DC ofset giderme, 100–2000 Hz bant geçiren filtreleme, gürültü azaltma ve tanh tabanlı normalizasyon uygulanmıştır. Böylece cihaz/ortam kaynaklı değişkenlik sınırlanmış ve sınıflandırmaya uygun, tutarlı genlik ve bant içerikleri elde edilmiştir.

Özellikle iki koldan yürütülmüştür. Klasik öğrenciler için her segmente MFCC, Δ ve Δ^2 katsayıları çıkarılmış; istatistiksel özetlerle tek bir vektöre dönüştürülerek sınıflandırıcılara verilmiştir. Derin öğrenme tarafında her segment, 128 bantlı log-Mel-spektrogram biçimine getirilmiş ve 2B giriş olarak evrişimsel mimarilere sunulmuştur. Veri dengesizliği eğitim evresinde, yalnızca eğitim kümesine uygulanan ılımlı artırımlarla (zaman kaydırma, düşük düzeyli gürültü enjeksiyonu, küçük perde kaydırma, zaman ölçeklendirme ve eğitim içi karışımlar) giderilmiş; doğrulama ve test kümeleri hiçbir şekilde değiştirilmemiştir. Tüm ölçekleme/normalizasyon istatistikleri yalnız eğitim kümesinden hesaplanıp doğrulama ve teste aynen uygulanmıştır.

Karşılaştırmaya dâhil edilen klasik yöntemler: RBF çekirdekli SVM, Rastgele Orman, K-En Yakın Komşu ve Lojistik Regresyon. Bu modellerde hiperparametre seçimi, eğitim verisi üzerinde hasta düzeyi stratifiye 10-katlı çapraz doğrulama ile yapılmış; test kümesi yalnızca nihai değerlendirme için kullanılmıştır. Derin öğrenme cephesinde temel bir CNN, CRNN (CNN+LSTM) hattı, ResNet-50, EfficientNet-B0, MobileNetV3 ve GhostNet ailesinin V1–V4 sürümleri değerlendirilmiştir. Tüm derin modellerde Adam optimizasyonu, öğrenme oranı zamanlaması ve erken durdurma

kullanılmış; aşırı uyumu sınırlamak için bırakma ve hafif L2 düzenleştirme uygulanmıştır.

Değerlendirme; genel doğruluk, kesinlik, duyarlılık, F1 (makro ve mikro), karışıklık matrisi, ROC/PR eğrileri ve ICBHI skoru üzerinden yapılmıştır. Genel tablo, derin öğrenme mimarilerinin dört sınıflı görevde klasik yöntemlerin üzerine çıktığını göstermektedir. GhostNetV4, makro-F1 ve sınıflar arası denge açısından en istikrarlı sonuçları üretmiş; “Wheeze” ve “Crackle” sınıflarında yüksek duyarlılığı korurken “Both” sınıfında da güçlü ayırım sağlamıştır. CRNN, süreklilik gösteren wheeze örüntülerinde ek kazanım sunarken; parametre sayısı-başarım dengesi açısından MobileNetV3 ve EfficientNet-B0 uygulanabilir seçenekler olarak öne çıkmıştır. Klasik tarafta RBF-SVM en güçlü taban çizgisini oluşturmuş, Rastgele Orman normal sınıfta yüksek özgülük sergilemiş; ancak azınlık sınıflarda duyarlılık kayıpları gözlenmiştir. Aşağıdaki alt bölümlerde, bu genel özetten hareketle model bazlı ayrıntılar, sınıf bazlı metrikler ve ROC/PR analizleri sırasıyla sunulmaktadır.

5.2. Derin Öğrenme Modellerinin Performansı

Dört sınıflı kurulumda derin mimariler hem doğruluk hem de makro F1 açısından klasik yöntemlerin önüne geçmiştir. GhostNet ailesi içinde sürümden sürüme kademeli bir iyileşme gözlenmiş, en yüksek başarı GhostNetV4 ile elde edilmiştir. Bu model %89.90 doğruluk ve 0.898 makro F1 ile çalışmanın tepe performansını üretmiştir. İlerleme, “ghost” özellik üretimiyle kanalı verimli genişleten hafif modüllerin, çok ölçekli birleştirme katmanlarının ve kanal-dikkat mekanizmalarının birlikte çalışmasına dayanmaktadır; sonuç, düşük parametre sayısına karşılık yüksek ayırt ediciliktir.

Karşılaştırmalı görünümde EfficientNet-B0 %87.00 doğruluk ve 0.868 makro F1 ile ikinci sırada yer alırken, MobileNetV3 %84.80 doğrulukla hesaplama maliyeti-başarım dengesi açısından güçlü bir pratik seçenek sunmuştur. Temel CNN hattı %86.40 doğrulukla sağlam bir taban çizgisi üretmiş, ResNet-50 ise bu veri ölçeğinde %76.70 doğrulukta kalmıştır. Bu tablo, solunum seslerinin zaman-frekans uzayında aynı anda hem dar bant boyunca süreklilik (wheeze) hem de kısa süreli geniş bant patlamalar (crackle) içeren karmaşık dokular barındırdığını; modern ve hafif mimarilerin bu dokuyu ağır mimarilere kıyasla daha verimli temsil edebildiğini göstermektedir.

Sınıf bazında GhostNetV4; “Wheeze” ve “Both” kategorilerinde yüksek F1’i korurken “Crackle” sınıfında da belirgin bir iyileşme sağlamıştır. Bu durum, modelin hem süreklilik gösteren harmonik çizgileri hem de milisaniye ölçeğindeki kesikli olayları aynı ağ içinde yakalayabilen çok ölçekli alıcı alanına ve kanal dikkatine sahip olmasından

kaynaklanmaktadır. Karışıklık matrislerinde görülen sınırlı “Wheeze↔Both” karışımı, ortak harmonik bileşenleri; “Crackle” tarafındaki nispi zorluk ise olayların kısa süreli doğası ve zaman çözünürlüğüne duyarlılığı işaret etmektedir.

Genelleme açısından, hasta-bağımsız bölme ve yalnızca eğitim bölümünde uygulanan ılımlı veri artırma adımları belirleyici olmuştur. Zaman kaydırma, düşük seviyeli gürültü ekleme, küçük perde kaydırma ve zaman esnetme sayesinde model, cihaz/bölge/birey kaynaklı varyasyonları daha iyi kapsamış; bu da özellikle azınlık sınıflarda duyarlılık artışı olarak yansımıştır. Eğitim dinamikleri (kayıp/başarım eğrileri) erken durdurma ve öğrenme oranı zamanlaması ile birlikte dengeli seyretmiş; test kümesindeki kazanımlar aşırı uyum pahasına oluşmamıştır.

Hesaplama maliyeti bakımından GhostNetV4 yaklaşık 1.3M parametreyle çalışmaktadır; bu değer, klasik ResNet-50’nin parametre sayısından çok daha düşüktür. Pratikte bu fark daha hızlı eğitim/çıkartım, düşük bellek gereksinimi ve gömülü/mobil kullanıma uygunluk anlamına gelir. Böylece klinik ortamlarda taşınabilir sistemler üzerinde gerçek zamana yakın uygulama olanağı doğmaktadır.

Tablo 14. Derin öğrenme modellerinin sınıflandırma performanslarının karşılaştırılması

Model	Doğruluk (%)	Makro F1	Normal F1	Wheeze F1	Crackle F1	Both F1
ResNet-50	76.7	0.767	0.72	0.80	0.74	0.81
EfficientNet-B0	87.0	0.868	0.81	0.91	0.82	0.92
MobileNetV3	84.8	0.847	0.79	0.87	0.82	0.90
CNN	86.4	0.864	0.81	0.90	0.82	0.92
GhostNetV1	84.4	0.844	0.78	0.89	0.80	0.90
GhostNetV2	86.7	0.867	0.81	0.90	0.82	0.94
GhostNetV3	85.1	0.851	0.80	0.88	0.81	0.91
GhostNetV4 (önerilen)	89.9	0.898	0.86	0.93	0.87	0.93

5.2.1. Önerilen GhostNetV4 Model Mimarisi

GhostNetV4, Mel-spektrogramları ($128 \text{ frekans} \times 216 \text{ zaman}$, tek kanal) doğrudan iki boyutlu evrişim katmanlarıyla işler. Mimari; düşük kanallı bir giriş (stem) katmanı, ardışık GhostV4 bloklarından oluşan çok ölçekli öznitelik çıkarıcı omurga ve küresel ortalama havuzlama ile tam bağlantılı sınıflandırma başlığından oluşur. Tüm konvolüsyonların ardından Batch Normalization uygulanır; etkinleştirme Hardswish'tir. Çözünürlük indirgemeleri gereken yerlerde stride=2 kullanılır; kanal ya da uzamsal boyut değişimlerinde 1×1 konvolüsyonlu projeksiyon ile artık bağlantılar korunur.

Genel akışta tek kanallı giriş spektrogram, stem aşamasında $3 \times 3 / \text{stride}=1$ konvolüsyon ile 8 kanala yükseltilir ($8 \times 128 \times 216$). Bunu izleyen GhostV4 blokları kademeli olarak kanal sayısını artırırken uzamsal boyutları iki aşamada yarıya indirir. İlk iki blok tam çözünürlükte çalışır ve $24 \times 128 \times 216$ boyutunda zenginleştirilmiş bir temsil üretir. Üçüncü ve dördüncü bloklarda stride=2 ile 64×108 çözünürlüğe inilir ve kanallar 60'a çıkar. Beşinci–yedinci bloklarda tekrar stride=2 uygulanarak 32×54 çözünürlüğe geçilir, kanal sayısı 100'e ulaşır. Sekizinci–dokuzuncu bloklar genişletilmiş bir GhostBottleneck düzeniyle kanalları 336'ya taşır (16×27 çözünürlükte). Son aşamada 1×1 konvolüsyonla kanal projeksiyonu 288'e ayarlanır; adaptif ortalama havuzlamayla $288 \times 1 \times 1$ 'e sıkıştırılan temsil, sınıflandırıcı başlığına aktarılır. Başlık bölümünde Dropout($p=0.20$) ve iki katmanlı doğrusal yapı ($288 \rightarrow 144 \rightarrow 4$) yer alır; çıkış katmanı dört sınıf için softmax üretir.

Her GhostV4 bloğunun içyapısında üç mekanizma birlikte çalışır. Adaptive Ghost Module (AGM), girişi önce 1×1 veya 3×3 konvolüsyonla “intrinsic” bir öznitelik alt-kümesine dönüştürür; aynı girdiden derinlik-ayrık (depthwise) 3×3 konvolüsyonlar ve 1×1 lineer projeksiyonlar üzerinden ek (“ghost”) öznitelikler türetilir. Intrinsic ve ghost öznitelikler kanal boyutunda birleştirilip 1×1 konvolüsyonla hedef kanal sayısına getirilir. Pratikte intrinsic/ghost oranı yaklaşık 1:1 seçilir; böylece yoğun konvolüsyonların maliyeti olmadan benzer kanal çeşitliliği elde edilir. Multi-Scale Feature Aggregation (MSFA) belirlenen ara noktalarda öznitelikleri paralel kollara ayırır (örneğin 1×3 , 3×3 ve dilated $3 \times 3 (d=2)$); her kol farklı alıcı alanıyla dar bant çizgisel süreklilikleri (wheeze) ve kısa süreli geniş bant patlamaları (crackle) eşzamanlı örnekler. Kolların çıktıları kanal boyutunda birleştirilir, 1×1 konvolüsyonla yeniden ölçeklenir ve BatchNorm sonrası ana akışa döner. Split-Attention (SA) ise kanal uzayını gruplara ayırarak (ör. $g=2$) global ortalama havuzlama ile grup başına bir sıkıştırma vektörü üretir; daraltma-genişletme ($r \approx 4$) ve yumuşak kapılama ile bağlamsal önemi yüksek kanallar güçlendirilir, gürültü baskılanır. Blok sonunda artık bağlantı eklenir; kanal veya çözünürlük değişimi varsa 1×1

projeksiyonla hizalama yapılır. Stride=2 indirgemeler, bilgi zenginleştirmesinden sonra uygulanacak şekilde MSFA/AGM çıkışında gerçekleştirilir.

Normalizasyon ve etkinleştirme düzeni, eğitim kararlılığına odaklanır. Her konvolüsyon katmanını BatchNorm izler ve Hardswish etkinleştirme, düşük genlik bölgelerinde yumuşak türev sağlayarak gradyan akışını dengeler. Sınıflandırıcı başında Dropout($p=0.20$) kullanılır; ağırlık çürümesi ($L2 \approx 1e-4$) ve global-norm gradyan kırpması (≈ 1.0) ile aşırı uyum sınırlanır. Girdi kanalı (log-Mel güç haritası) eğitim istatistikleriyle z-puan normalize edilir; çıkarımda kararlı ve hafif yapı sayesinde karışık kesinlik (AMP/fp16) güvenle uygulanabilir.

Eğitim ve çıkarım ayarları, mimarinin hafifliğini destekleyecek şekilde seçilmiştir. Optimizasyon Adam ile yürütülür, başlangıç öğrenme oranı $1e-3$ 'tür ve doğrulama kaybı iyileşmediğinde öğrenme oranı otomatik olarak yarıya indirilir. Erken durdurma etkinleştirilerek en iyi doğrulama ağırlıkları korunur; veri akışı sabit tohumlarla tekrarlanabilir biçimde düzenlenir. Bu kurgu, hem hız hem de genelleme açısından dengeli bir eğitim süreci sağlar.

Parametre bütçesi ve bellek izi, gömülü ve mobil kurulumlara uygun düşük kaynak tüketimini hedefler. Toplam parametre sayısı yaklaşık 1.3 milyon olup, tek duyarlılıkta ağırlık belleği ~ 5.2 MB mertebesinde. Aynı giriş boyutunda yaklaşık 25 milyon parametrelili bir ResNet-50 ile karşılaştırıldığında, %95'e varan bir parametre tasarrufu elde edilir; bu da daha düşük bellek gereksinimi, daha hızlı eğitim/çıkartım ve enerji verimliliği anlamına gelir.

Zaman-frekans çözünürlüğünün aşamalı planlanması, farklı akustik ipuçlarını uygun ölçekte yakalamaya dönüktür. 128×216 çözünürlükte çalışan ilk bloklar, yüksek frekans ayrıntılarını ve ince dokuları kayıpsız işler. 64×108 çözünürlüğe indirgenen orta katmanlar, wheeze benzeri çizgisel süreklilikleri ve formant benzeri yapıları daha geniş bağlamda modeller. 32×54 çözünürlükteki derin katmanlar, kısa süreli patlamaların çevresel bağlamını ve sınıflar arası ayrımı güçlendirir. Son olarak yüksek kanallı dar uzayda seçici bir özet üretilir; 1×1 projeksiyon ve küresel ortalama havuzlama, konumdan bağımsız, sınıf-ayırıştırıcı bir gövde vektörü oluşturur ve bu vektör dört sınıflı başlıkta tutarlı kararlar üretir.

Bu bütünleşik düzen, tek bir omurga içinde dar bant boyunca süreklilik gösteren çizgisel izler ile milisaniye ölçekli geniş bant patlamaları birlikte temsil eder. Parametre ekonomisi, eğitim ve çıkarımda düşük hesap/bellek yükü sağlarken; MSFA ve SA entegrasyonu, farklı bant ve bant genişliklerindeki ipuçlarını seçici olarak birleştirerek sınıflar arası ayrımı kuvvetlendirir. Sonuç olarak GhostNetV4, 128×216 Mel-

spektrogram girişlerinden çıkarılan gövde öznitelikleriyle dört sınıfta dengeli ve yüksek doğrulukta bir sınıflandırma çıktısı üretir.

Tablo 15. GhostNetV4 katman özeti ve parametre bilgisi

Katman / Blok	Açıklama (çekirdek/işlev)	Çıktı Boyutu*	Parametre ($\approx$)
Giriş (Stem)	3×3 Conv, BN, Hardswish (stride=1, same pad)	8×128×216	~0,5K
GhostV4 Blok 1–2	AGM + derinlik-ayrık 3×3 + noktasal 1×1, hafif artık bağlantı	24×128×216	~60K
GhostV4 Blok 3–4	Stride=2 ile zamansal/uzamsal indirgeme + MSFA entegrasyonu	60×64×108	~160K
GhostV4 Blok 5–7	3×3 derinlik-ayrık + Split-Attention (gruplu kanal ağırlama)	100×32×54	~320K
GhostV4 Blok 8–9	Genişletilmiş GhostBottleneck (kanal 120→336), artık bağlantı	336×16×27	~620K
Conv Last	1×1 Conv, BN, Hardswish (kanal projeksiyonu)	288×16×27	~180K
Küresel Ortalama Havuzlama	AdaptiveAvgPool2d → (1,1)	288×1×1	—
Sınıflandırıcı	Dropout(p=0.2) + Linear(288→144→4), Hardswish arakatmanı	4	~118K
Toplam	—	—	≈1,3M

5.3. Klasik Makine Öğrenmesi Modellerinin Performansı

Klasik yöntemler, dört sınıflı solunum sesi görevinde derin öğrenme mimarilerinin altında kalmıştır. En yüksek değer XGBoost ile elde edilmiş; doğruluk %76,8, makro F1 ise 0.766'dır. KNN ve Random Forest %74–75 bandında birbirine yakın sonuçlar üretirken SVM %73.7 doğruluk düzeyinde kalmıştır. Doğrusal (Logistic Regression) ve basit olasılıksal (Naive Bayes) yaklaşımlar belirgin biçimde geride kalmış; özellikle Naive Bayes'in sınıf-içi kovaryans varsayımları ve MFCC dağılımlarına duyarlılığı, düşük skorlarla sonuçlanmıştır.

Hata örüntüsü sınıf-özel bir karakter göstermektedir. Normal ve Wheeze sınıflarında görece daha yüksek doğruluk ve duyarlılık elde edilirken, Crackle ve özellikle

iki patolojik yapıyı birlikte içeren Both sınıfında başarımlar düşmektedir. MFCC + Δ + Δ^2 tabanlı öznitelikler kısa süreli, geniş bantlı patlamaları (crackle) ve eşzamanlı bileşenleri (both) özet istatistiklere indirger; bu indirgeme, sınıf sınırlarının keskinleşmesini zorlaştırır. Ağaca dayalı güçlendirme (XGBoost) bu noktada diğer klasik yöntemlere göre avantaj sağlar: doğrusal-olmayan ayrımları kural seviyesinde biriktirerek karma bölgelerde daha keskin kararlar üretebilir. Bununla birlikte Matthews korelasyon katsayısının (MCC) orta düzeylerde seyretmesi (≈ 0.65 – 0.69), sınıf dağılımı dengelenmiş olsa dahi öznitelik uzayında tam ayrışmanın sağlanamadığını göstermektedir.

Modeller arası karşılaştırma, aynı hasta-bağımsız bölme, aynı ön işleme (standardizasyon) ve aynı özellik seti üzerinde yapılmıştır; bu nedenle sonuçlar doğrudan kıyaslanabilir. XGBoost, derin karar yapıları ve alt-örnekleme stratejileri sayesinde en yüksek taban çizgisini oluştururken; Random Forest istikrarlı ancak daha sık kurallarla benzer bir bantta kalmıştır. KNN, ölçeklendirilmiş özniteliklerle rekabetçi görünse de yerel yoğunluk dalgalanmalarına duyarlı olduğundan azınlık sınıflarda dalgalanma göstermiştir. RBF çekirdekli SVM’de gamma/C ayarları geniş aralıkta aranmış olmasına rağmen, öznitelik özetlerinin sınıf ayrımını yeterince keskinleştirememesi nedeniyle XGBoost seviyesine çıkılamamıştır. Bu tablo, klasik yaklaşımların yorumlanabilir ve hafif bir referans zemini sunduğunu; ancak zaman–frekans dokuyu doğrudan işleyen derin mimariler karşısında üst sınırlarının daha erken geldiğini göstermektedir.

Tablo 16. Klasik makine öğrenmesi modellerinin sınıflandırma performansları

Model	Doğruluk (%)	Makro F1	Precision	Recall	MCC
XGBoost	76.8	0.766	0.768	0.767	0.691
Random Forest	74.7	0.745	0.749	0.747	0.665
KNN	75.3	0.753	0.757	0.753	0.672
SVM	73.7	0.737	0.738	0.737	0.650
Logistic Regression	53.0	0.529	0.530	0.530	0.374
Naive Bayes	40.6	0.380	0.440	0.406	0.222

5.4. Veri Artırma (Data Augmentation) Etkisinin Analizi

Deneylerde en çarpıcı bulgulardan biri, veri artırmanın model performansı üzerindeki belirgin katkısıdır. Augmentasyon uygulanmadan eğitilen GhostNetV4 yaklaşık yüzde 80 doğruluk ve 0.79 makro F1 üretirken, zaman kaydırma, gürültü enjeksiyonu, pitch shifting, zaman ölçeklendirme ve eğitim sırasında uygulanan mixup ile çeşitlendirilen eğitim seti sayesinde doğruluk yüzde 89'a, makro F1 ise 0.89'a yükselmiştir. Bu artış, modelin farklı cihaz, bölge ve birey kaynaklı değişkenliklere karşı daha dayanıklı temsiller öğrendiğini göstermektedir. Özellikle patolojik sınıfların spektral–zamansal varyasyonlarının daha iyi kapsanması, klinik açıdan kritik yanlış negatiflerin azalmasına katkı sağlamıştır.

Sınıf bazında en büyük iyileşme Both kategorisinde gözlenmiştir. Crackle ve wheeze örüntülerinin aynı segmentte bir araya gelmesi nedeniyle bu sınıf doğal olarak daha karmaşık bir dağılıma sahiptir; çoklu bozulma senaryolarını taklit eden augmentasyonlar bu karmaşıklığın daha doğru modellenmesine imkân vermiştir. Normal, Wheeze ve Crackle sınıflarında da benzer biçimde anlamlı iyileşmeler elde edilmiştir. Uygulanan artırma genlik ve zaman ekseninde “ılımlı” aralıklarda tutulmuş, böylece rallerin milisaniye ölçeğindeki yapısının bulanıklaşması ya da wheeze çizgilerinin fizyolojik aralığın dışına itilmesi önlenmiştir.

Tablo 17. Veri artırma sonrasında her sınıfa ait F1-skorundaki değişim

Sınıf	F1-Skor (Augmentasyonsuz)	F1-Skor (Augmentasyonlu)	Fark ($\Delta F1$)
Normal	0.77	0.86	+0.09
Wheeze	0.82	0.93	+0.11
Crackle	0.76	0.87	+0.11
Both	0.79	0.93	+0.15
Makro Ortalama	0.79	0.89	+0.10

5.5. Karışıklık Matrisi Analizi

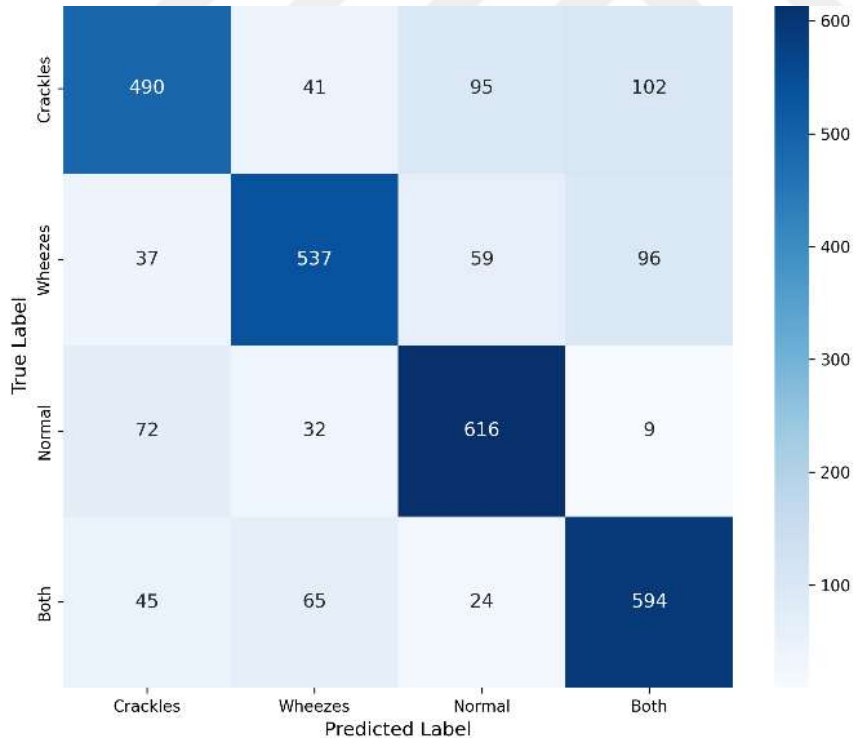
Modellerin hangi sınıflarda daha başarılı olduğunu ve hata örüntülerinin nerede yoğunlaştığını görmek amacıyla klasik makine öğrenmesi algoritmaları ve derin öğrenme mimarilerine ait karışıklık matrisleri incelenmiştir. Klasik yöntemlerde ortak eğilim, Normal ve Wheeze sınıflarında görece yüksek doğruluk; Crackle ve özellikle Both (Crackle + Wheeze) sınıflarında ise belirgin hata oranları yönündedir. XGBoost'ta Normal–Crackle ve Wheeze–Both çiftleri arasında sık karışma görülmüş; Random Forest ve SVM'de de benzer biçimde sınıflar arası geçiş bölgelerinde hatalar artmıştır. Bu tablo, MFCC + Delta + Delta² gibi öznitelik setlerinin kısa süreli olayları ve karma örüntüleri sınırlı temsil edebilmesine bağlanabilir.

Derin öğrenme tarafında karışıklık matrisi dağılımları belirgin biçimde iyileşmiştir. ResNet-50'de Crackle örneklerinin bir kısmı Normal ile yer değiştirmeye eğilimliken, EfficientNet-B0 ve MobileNetV3 bu hataları azaltarak daha dengeli bir sınıflandırma sağlamıştır. Wheeze ve Both ayrımı bu iki modelde daha başarılı görünmektedir. En yüksek ve dengeli başarı GhostNetV4'te elde edilmiştir. Karışıklık matrisi, modelin dört sınıfta da tutarlı duyarlılık ürettiğini göstermektedir: Normal için %84.6; Wheeze için %93.3; Crackle için %85.9; Both için %95.7 geri çağırma değerleri kaydedilmiştir. Bu sonuçlar, hem süreklilik gösteren harmonik yapıları (wheeze) hem de milisaniye ölçekli patlamaları (crackle) içeren karma segmentlerde sınırların daha doğru çizildiğini düşündürmektedir. Wheeze ile Both arasındaki sınırlı karışıklık, ortak harmonik bileşenlerin varlığı ile; Crackle'daki kısmi hata oranı ise olayların kısa süreli doğası ve zaman çözünürlüğüne duyarlılık ile açıklanabilir.

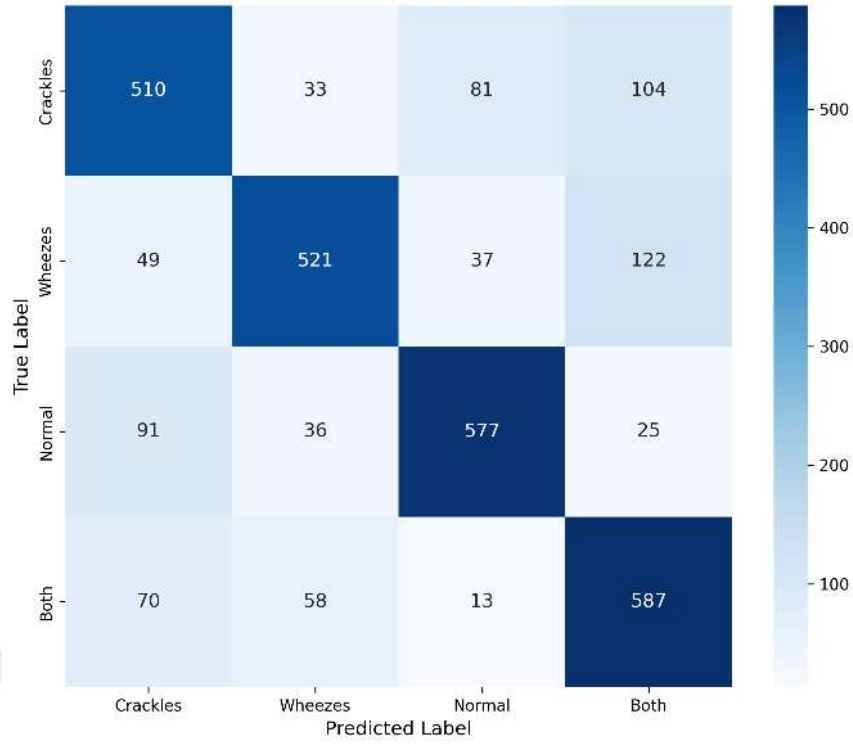
Ayrıntılı sınıf metrikleri Tablo 18'de özetlenmiştir. GhostNetV4 modeli dört sınıfta da dengeli ve yüksek bir performans sergilemiştir: Wheeze ve Both sınıflarında elde edilen %93'ün üzerindeki F1-skorları, modelin benzer patolojik solunum seslerini güvenilir biçimde ayırt edebildiğini göstermektedir. Crackle sınıfındaki F1-skoru %87 düzeyinde kalmakla birlikte, bu değer dahi klasik yöntemlerin belirgin biçimde üzerindedir ve sınıfın kısa süreli, düşük genlikli yapısından kaynaklanan zorluğu yansıtmaktadır. Genel doğruluk oranı %89.9 ve makro F1-skoru 0.898 olup, literatürde raporlanan birçok çalışmanın üzerinde bir performansa işaret etmektedir. Karışıklık matrislerinden elde edilen bu bulgular, bir sonraki bölümde sunulan ROC/AUC analizleriyle tamamlanarak modelin eşik-bağımlı davranışıyla birlikte değerlendirilecektir.

Tablo 18. GhostNetV4 modeline ait sınıf bazlı performans metrikleri

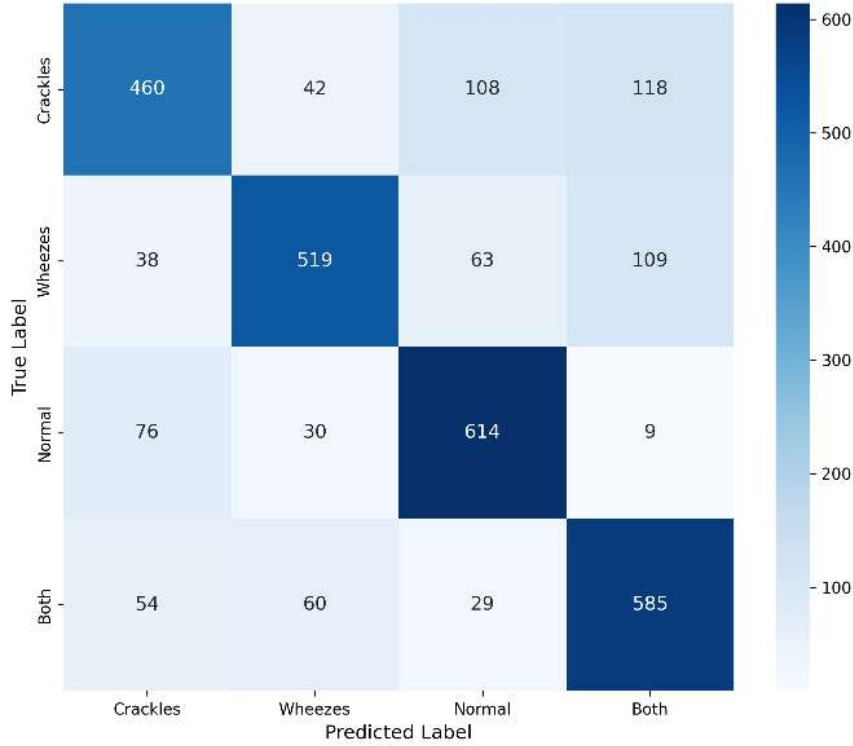
Sınıf	Precision	Recall (Se)	F1-Skoru	AUC-ROC	Destek (n)	Specificity (Sp)	ICBHI
Normal	0.867	0.846	0.856	0.958	729	0.957	0.898
Wheeze	0.928	0.933	0.930	0.987	728	0.976	0.954
Crackle	0.885	0.859	0.872	0.969	729	0.963	0.908
Both	0.914	0.957	0.935	0.989	728	0.970	0.964
Makro Ortalama	0.898	0.899	0.898	0.976	—	0.966	0.931
Genel Doğruluk	—	—	0.899	—	2914	—	—



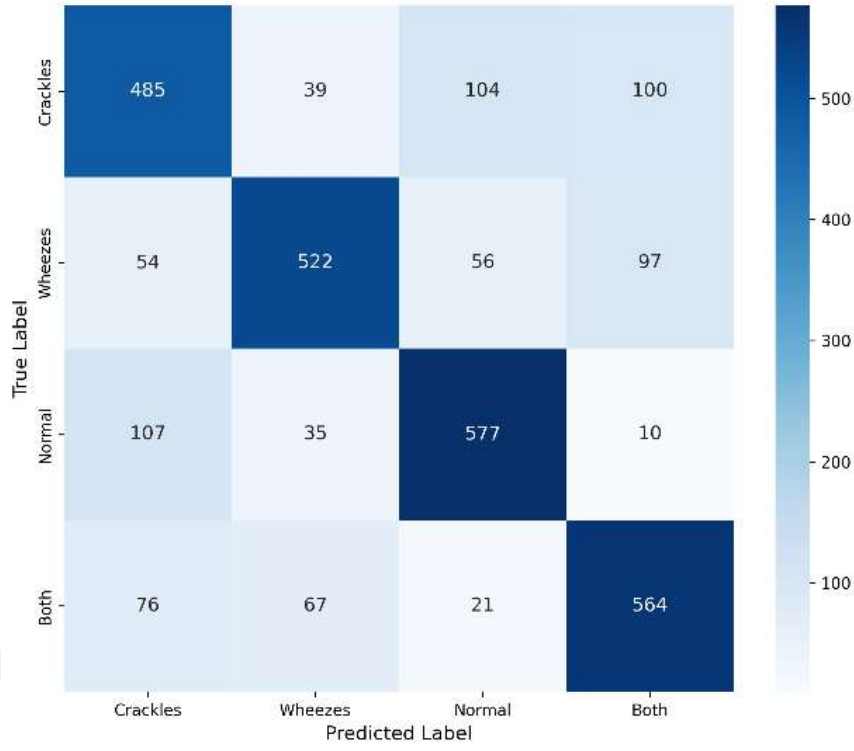
Şekil 22. XGBoost modelinin dört sınıf (Normal, Wheeze, Crackle, Both) için oluşturulan karışıklık matrisi



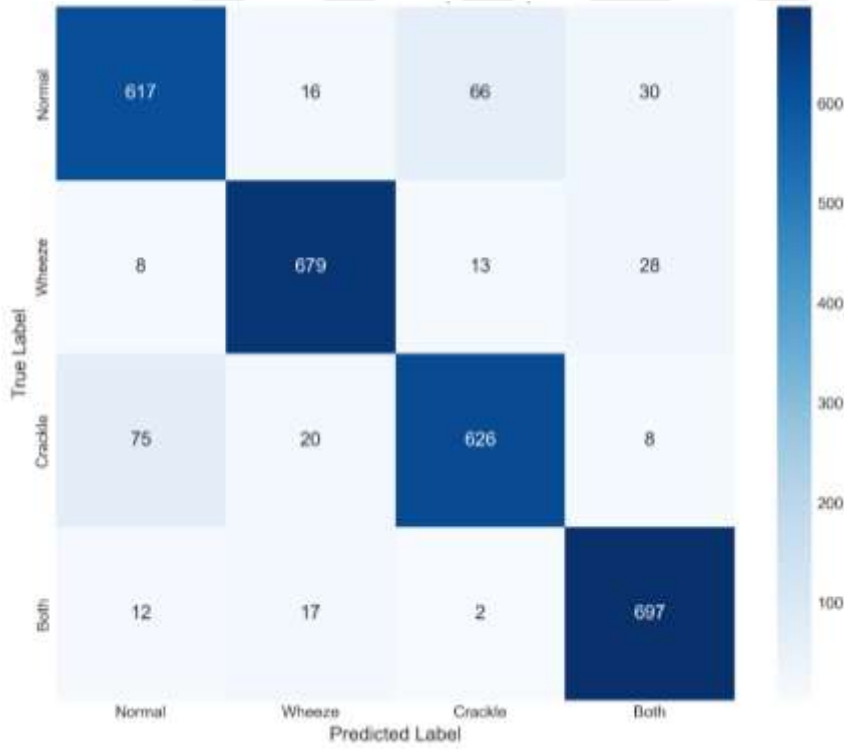
Şekil 23. KNN modelinin dört sınıf (Normal, Wheeze, Crackle, Both) için oluşturulan karışıklık matrisi



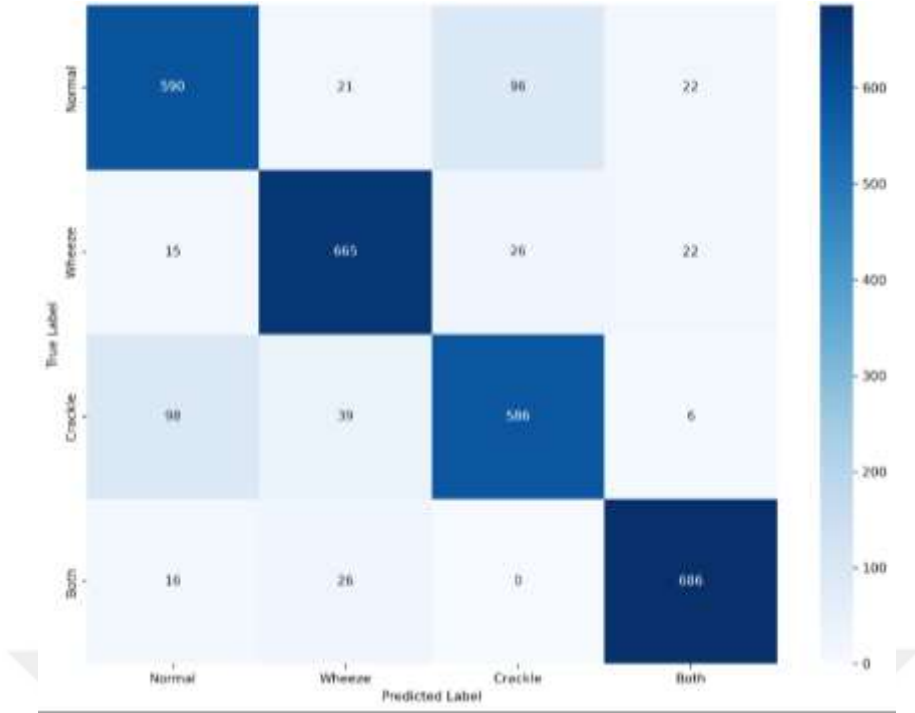
Şekil 24. Random Forest modelinin dört sınıf (Normal, Wheeze, Crackle, Both) için oluşturulan karışıklık matrisi



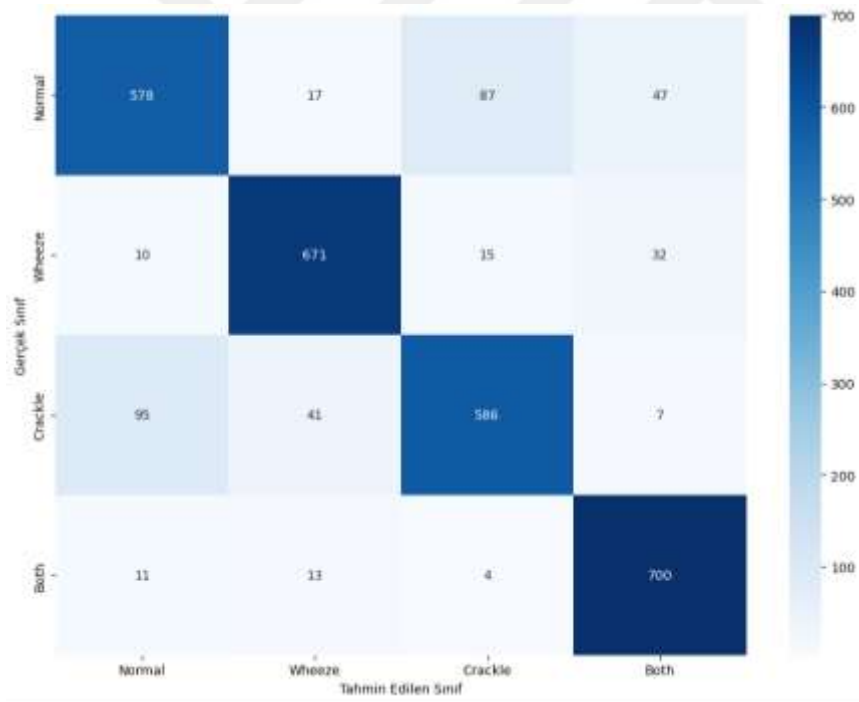
Şekil 25. SVM modelinin dört sınıf (Normal, Wheeze, Crackle, Both) için oluşturulan karışıklık matrisi



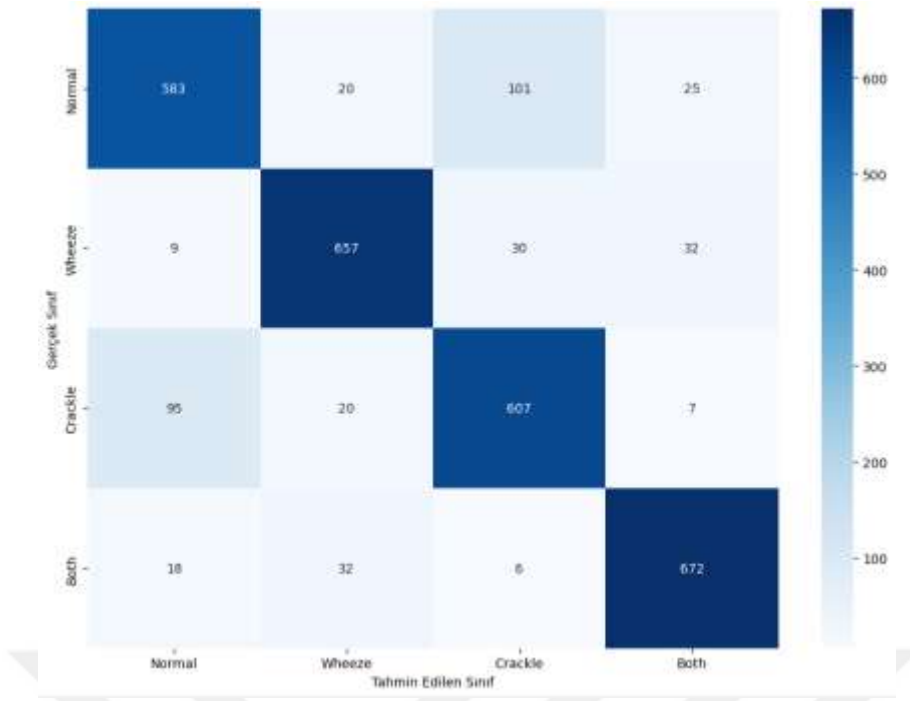
Şekil 26. GhostNetV4 modelinin dört sınıf (Normal, Wheeze, Crackle, Both) için oluşturulan karışıklık matrisi



Şekil 27. GhostNetV2 modelinin dört sınıf (Normal, Wheeze, Crackle, Both) için oluşturulan karışıklık matrisi



Şekil 28. EfficientNet – B0 modelinin dört sınıf (Normal, Wheeze, Crackle, Both) için oluşturulan karışıklık matrisi



Şekil 29. CNN modelinin dört sınıf (Normal, Wheeze, Crackle, Both) için oluşturulan karışıklık matrisi

5.6. ROC Eğrisi, AUC ve Eğitim Dinamiklerinin Analizi

Tablo 19 ve Şekil 30–31’de sunulan bulgular, GhostNetV4 mimarisinin hem eşik-bağımsız ayırt edicilikte hem de eğitim kararlılığında diğer yaklaşımların üzerinde performans sergilediğini göstermektedir. Ortalama AUC değerinin 0.976’ya ulaşması, modelin karar eşiği değiştiğinde dahi sınıflar arasında yüksek ayırt ediciliği koruduğunu ortaya koymaktadır. Sınıf bazında değerlendirildiğinde, Normal için 0.958, Wheeze için 0.987, Crackle için 0.970 ve Both için 0.989 AUC değerleri elde edilmiştir. Özellikle Both sınıfında AUC’nin ≈ 0.99 düzeyine çıkması, crackle ve wheeze bileşenlerini aynı anda içeren karma örneklerin güvenilir biçimde ayrıştırılabildiğine işaret etmektedir. EfficientNet-B0 ikinci en yüksek ortalama AUC’yi (0.945) sağlamış; MobileNetV3 ve temel CNN daha düşük ancak tutarlı sonuçlar üretmiştir. ResNet-50’nin parametrece ağır yapısı, bu veri ölçeğinde görece zayıf genelleme ile sonuçlanmıştır. XGBoost, SVM ve Random Forest gibi klasik yöntemlerde ortalama AUC’lerin 0.70–0.82 bandında kalması, bu yaklaşımların özellikle karma ve kısa süreli örüntülerde sınıf sınırlarını yeterince keskinleştiremediğini göstermektedir.

GhostNet ailesi içinde sürümler arası ilerleme de belirgindir. GhostNetV3’ün ROC eğrileri yüksek bir taban sunmakla birlikte (Şekil 32), GhostNetV4’te eğrilerin sol-üst köşeye daha da yaklaşması özellikle Wheeze ve Both ayrımında ek kazanç sağlamıştır. Bu iyileşme, çok ölçekli özellik birleştirme ve kanal-dikkat mekanizmalarının farklı bant ve bant genişliklerindeki ipuçlarını bütüncül biçimde kullanabilmesine bağlanmaktadır.

Sınıf bazlı F1 artışlarıyla birlikte okunduğunda, AUC'deki bu sıçrama modelin hem süreklilik gösteren harmonik yapıları (wheeze) hem de kısa ve kesikli olayları (crackle) aynı mimaride başarılı biçimde temsil edebildiğini doğrulamaktadır.

Eğitim dinamikleri, elde edilen AUC sonuçlarının aşırı uyum pahasına oluşmadığını göstermektedir. Şekil 31'de, eğitim ve doğrulama kayıplarının birlikte ve düzenli biçimde azaldığı; doğruluk ve makro F1 skorlarının ise epoch boyunca istikrarlı şekilde yükseldiği izlenmektedir. Erken durdurma, öğrenme oranı zamanlaması ve kapsamlı veri artırma adımları sayesinde eğitim-doğrulama eğrileri arasında belirgin bir ayrışma oluşmamış; “Combined Performance” göstergesinin zamanla yüksek bir plato düzeyine yerleşmesi, farklı ölçütlerde eşzamanlı ve dengeli bir iyileşmeye işaret etmiştir. Bu bulgular, GhostNetV4'ün test verisi üzerinde sergilediği yüksek AUC ve F1 değerlerinin sağlam bir genelleme kapasitesine dayandığını göstermektedir.

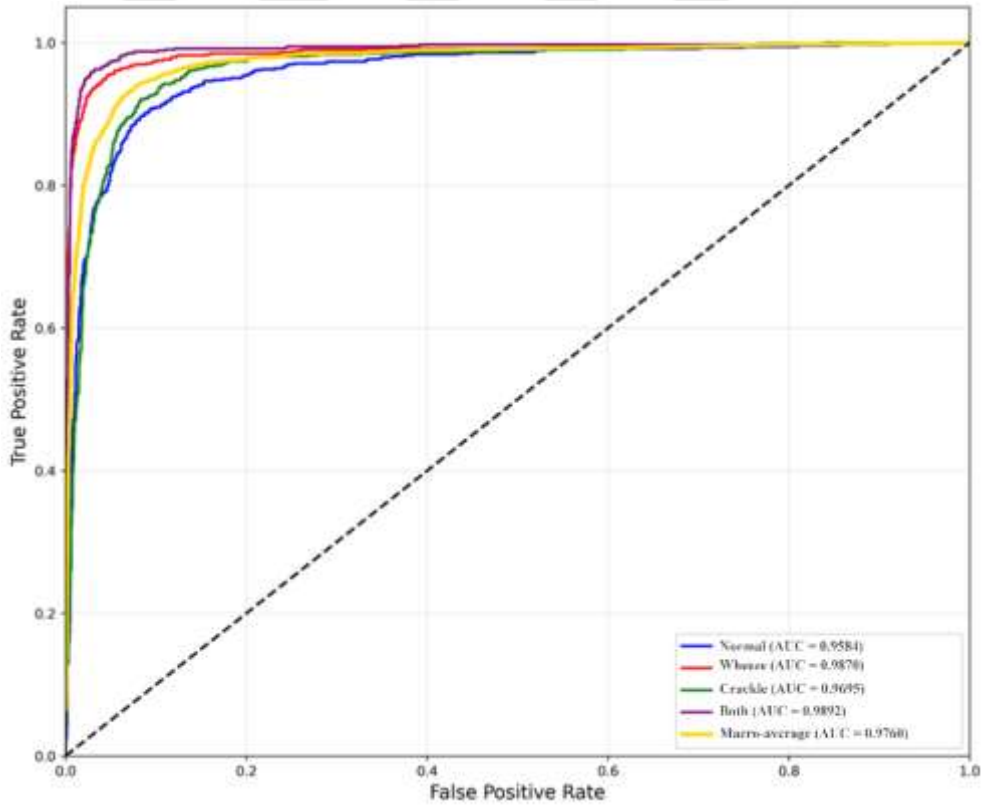
Toplam değerlendirmede, GhostNetV4'ün çok ölçekli ve dikkat temelli hafif mimarisi, solunum seslerinin zaman-frekans uzayındaki heterojen ipuçlarını yüksek duyarlılık ve özgüllükle işleyebilmekte; özellikle klinik olarak kritik öneme sahip karma (Both) örneklerde belirgin üstünlük sunmaktadır. EfficientNet-B0 güçlü bir alternatif olarak öne çıkarken, MobileNetV3 ve temel CNN hız-performans dengesi açısından uygulanabilir seçeneklerdir. Klasik yöntemler yorumlanabilirlik ve taban çizgisi karşılaştırmaları için yararlı olmakla birlikte, gerçekçi gürültü koşulları ve karma sınıf yapılarında derin öğrenme tabanlı yaklaşımların gerisinde kalmaktadır. Bu çerçevede, çalışmanın önceki bölümlerinde sunulan karışıklık matrisleri ve sınıf bazlı metriklerle tutarlı bir bütün oluşturmaktadır.

Tablo 19. Farklı modeller için sınıf bazlı AUC (ROC) değerleri

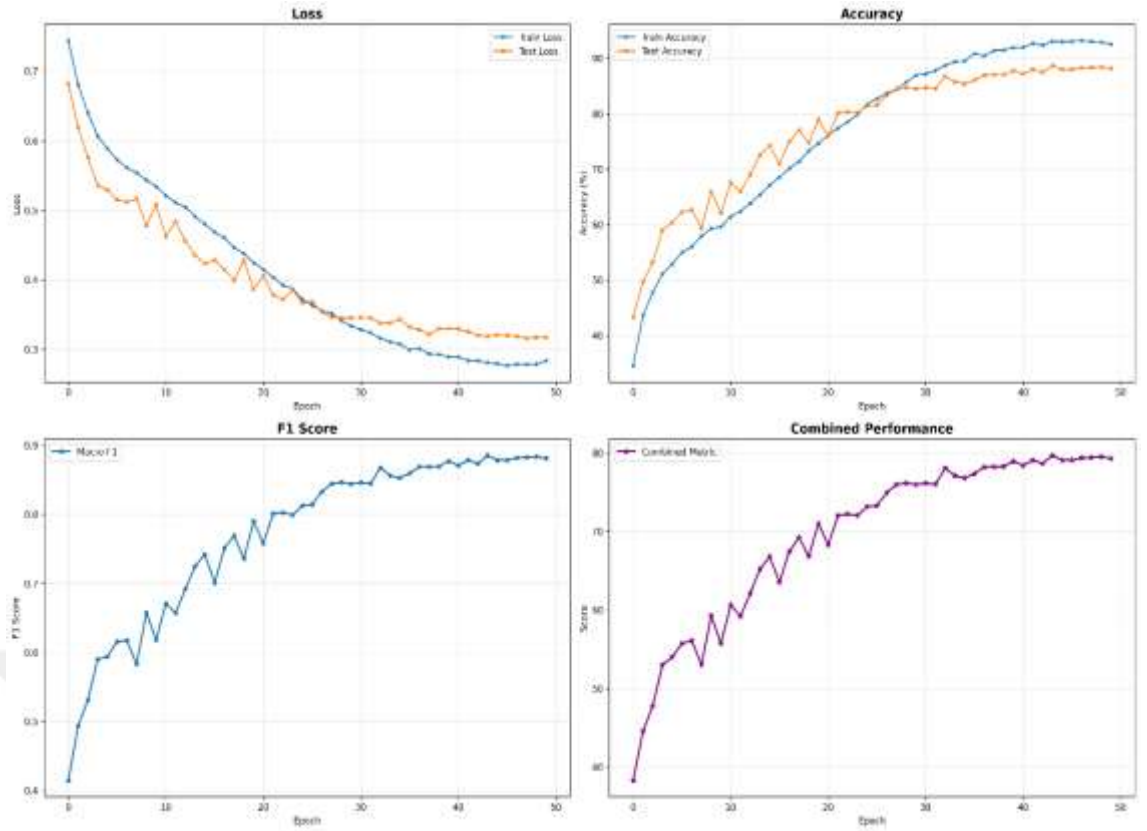
Model	Normal	Wheeze	Crackle	Both	Ortalama AUC
SVM	0.74	0.78	0.67	0.75	0.735
Random Forest	0.75	0.80	0.73	0.79	0.767
XGBoost	0.78	0.81	0.76	0.82	0.793
ResNet-50	0.77	0.82	0.74	0.81	0.785
MobileNetV3	0.83	0.87	0.82	0.90	0.855

Tablo 19. (Devamı)

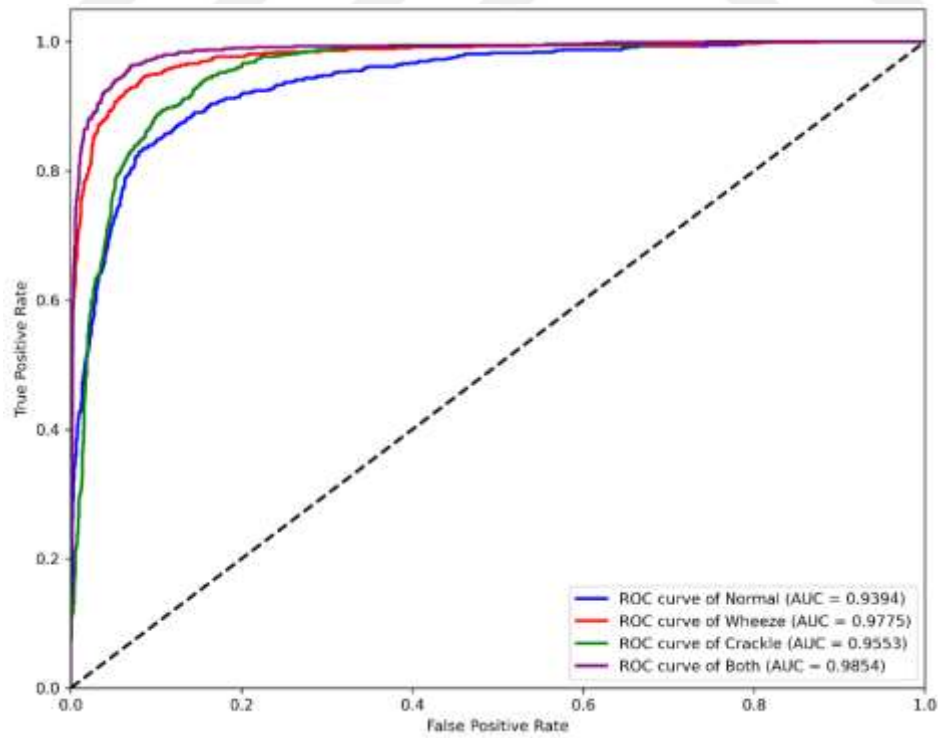
Model	Normal	Wheeze	Crackle	Both	Ortalama AUC
CNN (temel)	0.83	0.90	0.82	0.92	0.868
EfficientNet-B0	0.93	0.96	0.92	0.97	0.945
GhostNetV1	0.84	0.89	0.80	0.90	0.857
GhostNetV2	0.87	0.91	0.82	0.94	0.885
GhostNetV3	0.95	0.97	0.95	0.98	0.963
GhostNetV4 (Önerilen)	0.958	0.987	0.970	0.989	0.976



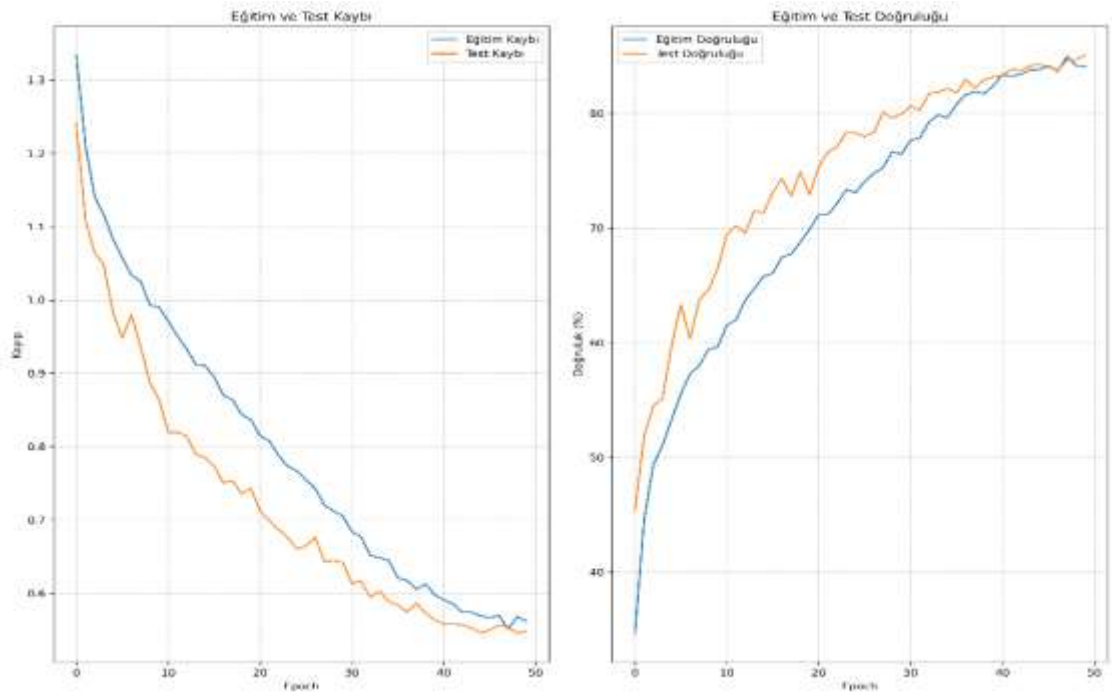
Şekil 30. GhostNetV4 modelinin dört sınıf (Normal, Wheeze, Crackle, Both) için ROC eğrileri ve sınıf bazlı ortalama AUC değerleri



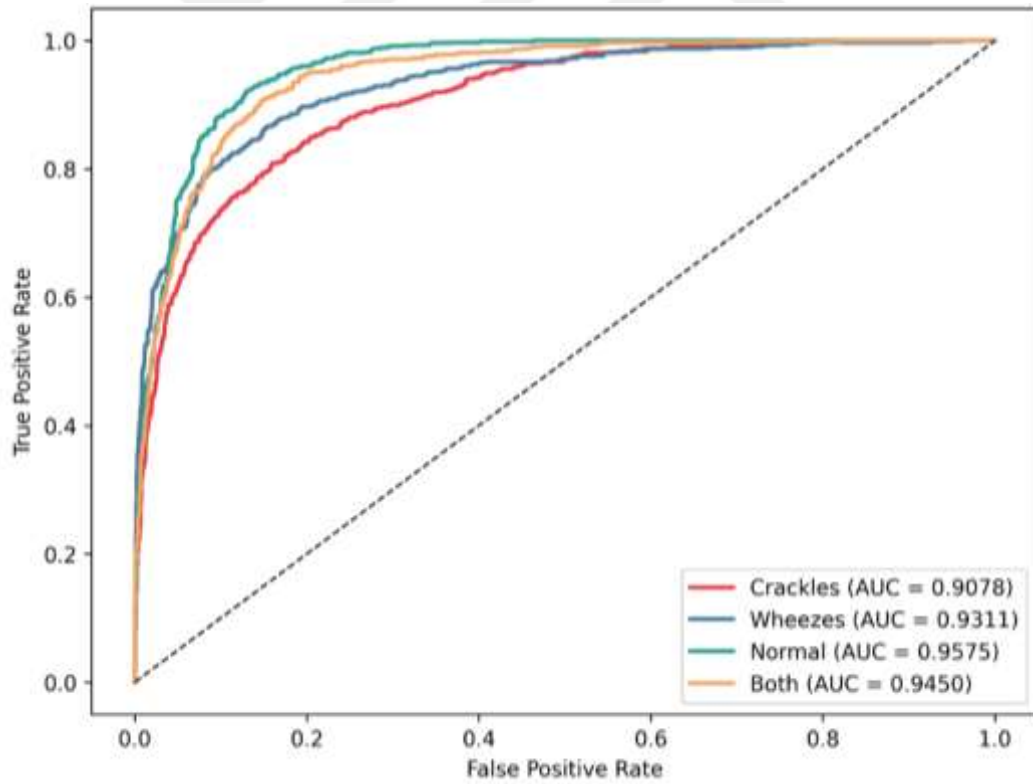
Şekil 31. GhostNetV4 modelinin eğitim süreci boyunca kayıp (Loss), doğruluk (Accuracy), F1 skoru ve birleşik performans (Combined Performance) metriklerinin epoch bazlı değişimi



Şekil 32. GhostNetV3 modelinin dört sınıf (Normal, Wheeze, Crackle, Both) için ROC eğrileri ve sınıf bazlı ortalama AUC değerleri.



Şekil 33. GhostNetV3 modelinin eğitim süreci boyunca kayıp (Loss), doğruluk (Accuracy), F1 skoru ve birleşik performans (Combined Performance) metriklerinin epoch bazlı değişimi



Şekil 34. XGBoost modelinin dört sınıf (Normal, Wheeze, Crackle, Both) için ROC eğrileri ve sınıf bazlı ortalama AUC değerleri.

5.7. Genel Değerlendirme ve Literatürle Karşılaştırma

Elde edilen bulgular, solunum sesi sınıflandırmasında derin öğrenme tabanlı yaklaşımların klasik makine öğrenmesi yöntemlerine göre belirgin bir üstünlük sağladığını göstermektedir. Önerilen GhostNetV4 mimarisi; doğruluk, makro ortalama F1-skoru ve ROC eğrisi altında kalan alan (AUC) ölçütlerinde yüksek ve dengeli sonuçlar üretmiş; %89.9 doğruluk, 0.898 makro F1 ve sınıflar arası ortalama 0.976 AUC değerlerine ulaşmıştır. Bu performans düzeyi, ICBHI veri seti üzerinde raporlanan güncel ve güçlü sonuçlarla aynı aralıkta yer almakta; özellikle çok sınıflı ayrımın söz konusu olduğu senaryolarda modelin genellenebilirliğini desteklemektedir.

Literatürde klasik yöntemlerle (MFCC tabanlı SVM, KNN, RF vb.) çoğunlukla %70–75 doğruluk bandının, temel CNN modelleriyle ise %82–86 bandının raporlandığı dikkate alındığında, GhostNetV4’ün sağladığı iyileşmenin özellikle karma örüntüler açısından anlamlı olduğu görülmektedir. “Both (Crackle + Wheeze)” sınıfında %93’ün üzerinde F1-skoru elde edilmesi, spektral olarak benzer ve birlikte görülen patolojik işaretlerin ayrımında modelin güçlü bir ayırıştırma kapasitesine sahip olduğunu göstermektedir.

Parametre sayısı yaklaşık 1.3 milyon olan GhostNetV4’ün hesaplama gereksinimi düşüktür. Bu nitelik, taşınabilir elektronik stetoskoplar, giyilebilir aygıtlar ve mobil uygulamalar üzerinde gerçek zamanlı çalıştırma olanağı sunarak klinik açıdan pratik bir kullanım vadeder. Ayrıca uygulanan veri artırma stratejileri (zaman kaydırma, gürültü enjeksiyonu, perde kaydırma, zaman esnetme ve mixup) modelin farklı kayıt koşullarına uyumunu kuvvetlendirerek gürültülü ve heterojen klinik ortamlarda dahi kararlı bir performans sergilemesine katkı sağlamıştır.

Sonuç olarak, bu çalışma hem doğruluk hem de genellenebilirlik açısından rekabetçi bir düzeyi yakalamış; solunum sesi analizi alanında hafif, hızlı ve klinik uygulamaya geçirilebilir bir derin öğrenme mimarisi sunmuştur. Modelin daha geniş, çok merkezli ve çok cihazlı veri setlerinde; ayrıca teletıp akışlarına entegre gerçek zamanlı kurulumlarda test edilmesi, pratik kullanıma geçişi hızlandıracaktır.

6. SONUÇ, BULGULARIN ÖZETİ VE GENEL KATKILAR

Bu tez, solunum seslerinin otomatik sınıflandırılması problemini klasik makine öğrenmesi ve derin öğrenme yaklaşımlarını aynı deneysel protokol altında karşılaştırarak kapsamlı biçimde ele almıştır. Çalışma, farklı cihazlar, çok merkezli kayıt koşulları ve değişken ortam gürültüsü içeren ICBHI 2017 veri seti üzerinde yürütülmüş; bu yönüyle gerçek yaşamı temsil gücü yüksek ve klinik uygulamalara aktarımı mümkün bir kurulum tercih edilmiştir.

Klasik yöntemler (SVM, KNN, Rastgele Orman, XGBoost) MFCC, Mel-spektrogram ve temel spektral öznitelikler üzerinden değerlendirildiğinde, belirli bir taban performans sağlamakla birlikte karmaşık akustik örüntülerin ayrıştırılmasında sınırlı kalmıştır. En başarılı klasik yaklaşım olan XGBoost dahi yaklaşık %79 doğruluk ve 0.79 makro F1 düzeyini aşamamıştır. Buna karşılık, görüntü benzeri temsiller üzerinde uçtan uca öğrenme yapan derin mimariler belirgin üstünlük sergilemiştir. EfficientNet-B0 ResNet-50 ve MobileNetV3 ile yapılan adil karşılaştırmalarda, önerilen GhostNetV4 mimarisi %89.9 doğruluk, 0.898 makro F1 ve tüm sınıflarda ≥ 0.96 AUC değerleriyle en yüksek başarıyı elde etmiştir. Karışıklık matrisi ve ROC eğrilerinin birlikte okunması, GhostNetV4'ün normal ve patolojik sınıfları dengeli biçimde ayırtabildiğini; özellikle Wheeze ve Both (Crackle + Wheeze) sınıflarında belirgin ayırt edicilik sağladığını göstermektedir. Daha ağır bir yapı olan ResNet-50'nin Crackle sınıfında görece daha fazla hata üretmesi, kısa süreli ve düşük genlikli olayların çok-ölçekli ve seyrek özellik çıkarımına yatkın hafif mimarilerle daha etkili yakalanabildiğine işaret etmektedir.

Çalışmanın mühendislik açısından kritik bulgularından biri, veri artırma (augmentasyon) stratejilerinin katkısının nicel olarak ortaya konmasıdır. Augmentasyon uygulanmadığında doğruluk ~%80 civarında seyrederken; zaman kaydırma, gürültü enjeksiyonu, perde (pitch) değişimi, zaman esnetme/ölçekleme, mixup ve zaman/frekans maskeleri gibi tekniklerin birlikte kullanımı doğruluğu %89.9 seviyesine taşımış; sınıf bazlı F1 skorlarında %10–15 aralığında artışlar sağlamıştır. Özellikle dengesiz sınıflar olan Wheeze ve Both için gözlenen iyileşmeler, sınırlı ve heterojen klinik veri koşullarında çeşitliliğin yapay olarak artırılmasının genelleme kabiliyetini güçlendirdiğini doğrulamaktadır.

Bu bulgular yalnızca performans üstünlüğünü değil, pratik uygulanabilirliği de desteklemektedir. GhostNetV4'ün düşük parametre bütçesi ve hesaplama verimliliği, taşınabilir elektronik stetoskoplar ve gömülü donanım üzerinde gerçek zamanlı

uygulamaları mümkün kılar. Böyle bir tasarım; gecikme, enerji tüketimi ve bakım maliyetleri açısından avantaj sunarken, hasta başında ya da evde izlem senaryolarında sürekliliği destekler. Modelin dengeli duyarlılık-kesinlik profili, klinik karar destek amaçlı kullanımda risk yönetimini kolaylaştırmakta; yanlış negatifleri azaltmaya dönük eşik ayarlamaları ile operasyonel esneklik sağlamaktadır.

Sınırlılıklar da geleceğe dönük yol haritasını şekillendirmektedir: ICBHI 2017 veri setinin sınıf dağılımı dengesizdir ve kapsamı dört sınıfla sınırlıdır; demografik çeşitlilik, komorbidite ve nadir patolojiler için genelleme yapılırken temkin gereklidir. Kayıt cihazları arasındaki akustik farklılıklar ve etiketlemede gözlemci değişkenliği modele dışsal belirsizlikler katmaktadır. Bu nedenle daha geniş, çok merkezli ve standartlaştırılmış veri koleksiyonları önemlidir.

Toparlarsak, bu tez; klasik ve derin öğrenme yaklaşımlarını aynı çerçevede sistematik karşılaştırması, hafif ve verimli bir mimari olarak GhostNetV4'ü önermesi ve üstünlüğünü göstermesi, veri artırmanın sınıf dengesizliği ve genelleme üzerindeki etkisini ölçülebilir biçimde kanıtlaması, karar davranışlarını makro F1, sınıf-bazlı AUC, karışıklık matrisi ve ROC analizleriyle şeffaf biçimde ortaya koyması ve bulguları mobil/taşınabilir sistemlere uygun bir yapay zekâ tabanlı stetoskop altyapısına dönüştürülebilir biçimde sunması bakımından alana özgün ve uygulanabilir katkılar getirmiştir. Elde edilen %89.9 doğruluk, 0.898 makro F1 ve ≥ 0.96 AUC profili, solunum sesi temelli karar destek sistemlerinin erken tanı, uzaktan izlem ve teletıp uygulamalarında ölçeklenebilir biçimde konumlandırılabileceğini göstermektedir.

7. GELECEK ÇALIŞMALAR

Bu tez kapsamında elde edilen bulgular, GhostNetV4 mimarisinin solunum sesi sınıflandırma görevlerinde yüksek doğruluk ve dengeli performans sunduğunu ortaya koyarak hem akademik çalışmalar hem de klinik uygulamalar açısından güçlü bir temel oluşturmuştur. Bununla birlikte, modelin gerçek dünya koşullarında daha geniş kapsamlı, güvenilir ve sürdürülebilir bir yapıya kavuşturulabilmesi için çeşitli geliştirme alanları mevcuttur. Bu doğrultuda, gelecekteki araştırmalara yön verebilecek yedi ana başlık aşağıda bütüncül bir çerçevede sunulmuştur.

7.1. Harici Doğrulama ve Gerçek Dünya Verisiyle Genelleme

Modelin klinik ortamlara uyarlanabilirliğini artırmak için farklı yaş grupları, cinsiyetler, etnik kökenler ve komorbiditelere sahip geniş hasta popülasyonlarını içeren çok merkezli çalışmalar gerekmektedir. Farklı cihaz ve ortam koşullarında toplanmış solunum sesi kayıtlarının dâhil edilmesi, modelin cihazlar arası uyumluluğunu ve gürültüye karşı dayanıklılığını değerlendirmeye olanak sağlayacaktır. Bu süreçte, birden fazla uzmanın bağımsız anotasyonları kullanılarak etiket güvenilirliği artırılmalı, anotasyon tutarlılığı istatistiksel yöntemlerle ölçülmelidir. Ayrıca cihaz türü, ortam gürültüsü ve kayıt tekniği gibi değişkenlerin model performansına etkisi ayrıntılı analiz edilmelidir.

7.2. Uçtan Uca Gerçek Zamanlı Entegrasyon ve Model Sıkıştırma

Solunum sesi sınıflandırma sistemlerinin sahada kullanılabilirliği, yalnızca doğruluk metrikleriyle değil, aynı zamanda hesaplama verimliliği, enerji tüketimi ve donanımsal kısıtlarla baş edebilme kapasitesiyle de doğrudan ilişkilidir. Bu bağlamda, GhostNetV4 mimarisinin düşük gecikme süresi ve sınırlı bellek gereksinimi ile uç cihazlarda (örneğin mobil cihazlar, taşınabilir stetoskoplar, mikrodenetleyici tabanlı sistemler) çalıştırılabilmesi için modelin optimizasyonuna yönelik ileri adımlar atılmalıdır.

Bu süreçte kuantizasyon (ağırlıkların düşük bit çözünürlüğe indirgenmesi), bilgi damıtma (knowledge distillation ile küçük modelin büyük modelden öğrenmesi), kanal budama (gereksiz nöronların çıkarılması) ve ağırlık paylaşımı (parameter sharing) gibi model sıkıştırma ve hızlandırma teknikleri kullanılabilir. Ek olarak, çevrimdışı çalışma modları, veri güvenliği açısından güvenli güncellemeleri mümkün kılan OTA (Over-the-

Air) mekanizmaları ve hata toleranslı mimari düzenlemeler de uç sistemlerin stabil çalışması için göz önünde bulundurulmalıdır.

7.3. Sınıflandırma Kapsamının Genişletilmesi ve Klinik Senaryolara Uyum

Mevcut çalışmada kullanılan dört temel sınıf (normal, wheeze, crackle, both) üzerine kurulu sınıflandırma sistemleri, bazı klinik durumları yetersiz şekilde temsil edebilir. Bu nedenle, modelin sınıflandırma kapsamı genişletilerek daha fazla hastalık tipi veya semptom örüntüsü tanımlanmalıdır. Özellikle astım ve KOAH alevlenmeleri, pnömoni, interstisyel akciğer hastalıkları, bronşektazi, stridor gibi yaygın ve önemli solunum patolojileri için tekli veya çok etiketli (multi-label) sınıflandırma senaryoları oluşturulabilir.

Buna ek olarak, sadece solunum seslerine dayalı analiz yerine multimodal veri yaklaşımları da gelecekteki çalışmalara entegre edilebilir. Örneğin; semptom günlükleri, nabız, ateş, SpO₂ gibi vital bulgular ya da görüntüleme verileri (BT, X-ray) ile birleştirilmiş veri kümeleri kullanılarak daha zengin ve klinik bağlamda anlamlı sınıflandırma sistemleri geliştirilebilir. Erken ve geç füzyon yöntemleri, ortak temsil öğrenme (shared representation learning) gibi tekniklerle multimodal veri entegrasyonu sağlanabilir.

7.4. Açıklanabilir Yapay Zekâ ve Klinik Geri Bildirim

Derin öğrenme modellerinin karar süreçleri çoğunlukla "kara kutu" (black box) niteliği taşımakta ve bu durum klinik ortamlarda güven sorununa yol açmaktadır. Bu nedenle, geliştirilen sistemlerin kararlarını açıklayabilir hâle getirmek hem kullanıcı güvenini artırmak hem de hataların kökenini anlayabilmek açısından büyük önem taşır.

Gelecekteki çalışmalarda Grad-CAM, SHAP (SHapley Additive exPlanations), LIME, Integrated Gradients gibi yöntemler aracılığıyla modelin dikkat ettiği zaman-frekans bölgeleri görselleştirilmeli ve örnek bazlı yorumlamalar sağlanmalıdır. Ayrıca, modelin belirsizlik düzeyi (örneğin epistemik ve aleatorik belirsizlik) ölçülerek sistemin hangi durumlarda daha az güvenilir kararlar verdiği tanımlanmalı, bu bilgiler kullanıcı arayüzlerine anlamlı geri bildirim olarak yansıtılmalıdır.

7.5. Etik, Adalet ve Regülasyon Uyumluluğu

Tıbbi yapay zekâ sistemlerinin sahaya inebilmesi yalnızca teknik başarıyla değil, aynı zamanda etik sorumluluklar ve düzenleyici gerekliliklerin karşılanmasıyla mümkündür. Bu kapsamda:

- Farklı demografik gruplarda (ör. yaş, cinsiyet, etnik köken) modelin performansının adalet (fairness) açısından test edilmesi gereklidir.
- Hasta verilerinin gizliliği, anonimleştirme süreçleri ve veri yönetişimi ilkeleri (ör. GDPR, HIPAA) gözetilmelidir.
- Modelin bir tıbbi cihaz olarak regülasyonlara (örneğin FDA, CE, ISO 13485, IEC 62304) uygunluk göstermesi için kalite yönetimi, risk analizi, kullanım sonrası izleme (post-market surveillance) gibi süreçlerin entegrasyonu sağlanmalıdır.

7.6. Dayanıklılı Öğrenme Yöntemleri ve Veri Eksikliğiyle Baş Etme

Gerçek klinik veriler genellikle etiket eksikliği, anotasyon hataları, düşük örnekleme gibi zorluklar içerir. Bu koşullarda derin öğrenme sistemlerinin başarısını sürdürebilmesi için dayanıklı öğrenme yöntemleri geliştirilmelidir. Özellikle:

- Yarı denetimli öğrenme (semi-supervised), zayıf denetimli (weakly supervised) ve öz denetimli (self-supervised) yöntemler araştırılmalıdır.
- wav2vec 2.0, HuBERT gibi ön eğitilmiş modeller, küçük veri ile yüksek doğruluk sağlamada avantaj sunabilir.
- Crackle gibi kısa süreli ve zayıf genlikli olayların algılanabilmesi için olay tabanlı segmentasyonlar, çok çözünürlüklü zaman-frekans analizleri ve dikkat mekanizmaları (attention modules) entegre edilebilir.

7.7. Klinik Akışlara Uyum, Sürdürülebilirlik ve Model Güncellenebilirliği

Klinik entegrasyon sürecinde sistemin sürdürülebilirliği ve uzun vadeli güvenilirliği kadar, kullanım kolaylığı ve otomatik güncelleme yetenekleri de büyük önem taşır. Kullanım senaryolarına göre (örneğin tarama, tanı, izlem), hassasiyet-özgüllük dengesine özel eşik optimizasyonu yapılmalıdır. Ayrıca:

- Sürekli öğrenme (continual learning) yöntemleriyle sistem, zaman içinde yeni verilerle güncellenmeli ve "catastrophic forgetting" gibi sorunlara karşı dayanıklı hâle getirilmelidir.
- Veri kaymalarının (data drift) tespiti, versiyon yönetimi, otomatik uyarı sistemleri gibi MLOps araçları aktif şekilde devreye alınmalıdır.

Sonuç olarak, bu tezde önerilen GhostNetV4 tabanlı yaklaşımın daha geniş kapsamlı, açıklanabilir, sürdürülebilir ve etik çerçevede uygulanabilir hâle gelmesi için önerilen bu stratejik yol haritası; yapay zekâ destekli solunum sesi analizinin klinik ortamlarda yaygın olarak benimsenmesi ve güvenle kullanılabilmesi açısından hayati bir temel oluşturacaktır.

KAYNAKÇA

- Abeyratne, U. R., Swarnkar, V., Setyati, A. ve Triasih, R. (2013). Cough sound analysis can rapidly diagnose childhood pneumonia. *Annals of Biomedical Engineering*, 41(11), 2448-2462.
- Agnew, C. C., Agreda, P., Aguib, H., Ahad, M., Ahearn, T. S., Ahmad, M. M. ... Ataollahi, A. (2011). 2011 Index IEEE Transactions on Biomedical Engineering. *IEEE Transactions on Biomedical Engineering*, 58(12), 1.
- Andrews, J. L. ve Badger, T. L. (1979). Lung sounds through the ages: From Hippocrates to Laënnec to Osier. *JAMA*, 241(24), 2625-2630.
- Bae, S., Kim, J. W., Cho, W. Y., Baek, H., Son, S., Lee, B. ... Yun, S. Y. (2023). Patch-mix contrastive learning with audio spectrogram transformer on respiratory sound classification. arXiv preprint arXiv:2305.14032.
- Bahdanau, D., Cho, K. ve Bengio, Y. (2014). Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv:1409.0473.
- Bahoura, M. (2009). Pattern recognition methods applied to respiratory sounds classification into normal and wheeze classes. *Computers in Biology and Medicine*, 39(9), 824-843.
- Bardou, D., Zhang, K. ve Ahmad, S. M. (2018). Lung sounds classification using convolutional neural networks. *Artificial Intelligence in Medicine*, 88, 58-69.
- Bhattacharya, D., Sharma, N. K., Dutta, D., Chetupalli, S. R., Mote, P., Ganapathy, S. ... Alagesan, M. (2023). Coswara: A respiratory sounds and symptoms dataset for remote screening of SARS-CoV-2 infection. *Scientific Data*, 10(1), 397.
- Bishop, C. M. ve Nasrabadi, N. M. (2006). *Pattern recognition and machine learning*. New York: Springer.
- Bohadana, A., Izbicki, G. ve Kraman, S. S. (2014). Fundamentals of lung auscultation. *New England Journal of Medicine*, 370(8), 744-751.
- Breiman, L., Friedman, J., Olshen, R. A. ve Stone, C. J. (2017). *Classification and regression trees*. Chapman and Hall/CRC.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5-32.
- Chambres, G., Hanna, P. ve Desainte-Catherine, M. (2018, Eylül). Automatic detection of patient with respiratory diseases using lung sound analysis. 2018 International Conference on Content-Based Multimedia Indexing (CBMI), (ss. 1-6). IEEE.

- Chawla, N. V., Bowyer, K. W., Hall, L. O. ve Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 16, 321-357.
- Chen, T. ve Guestrin, C. (2016, Ağustos). XGBoost: A scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, (ss. 785-794).
- Chen, H., Yuan, X., Pei, Z., Li, M. ve Li, J. (2019). Triple-classification of respiratory sounds using optimized s-transform and deep residual networks. *IEEE Access*, 7, 32845-32852.
- Chu, Y., Wang, Q., Zhou, E., Fu, L., Liu, Q. ve Zheng, G. (2025). CycleGuardian: A framework for automatic respiratory sound classification based on improved deep clustering and contrastive learning. *Complex & Intelligent Systems*, 11(4), 200.
- Cortes, C. ve Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3), 273-297.
- Cover, T. ve Hart, P. (1967). Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, 13(1), 21-27.
- Davis, S. ve Mermelstein, P. (1980). Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 28(4), 357-366.
- Demir, F., Ismael, A. M. ve Sengür, A. (2020). Classification of lung sounds with CNN model using parallel pooling structure. *IEEE Access*, 8, 105376-105383.
- Demirci, B. A., Koçyiğit, Y., Kızıllırmak, D. ve Havlucu, Y. (2022). Adventitious and normal respiratory sound analysis with machine learning methods. *Celal Bayar University Journal of Science*, 18(2), 169-180.
- Dokur, Z. ve Ölmez, T. (2003). Classification of respiratory sounds by using an artificial neural network. *International Journal of Pattern Recognition & Artificial Intelligence*, 17(4).
- Doshi-Velez, F. ve Kim, B. (2017). Towards a rigorous science of interpretable machine learning. arXiv preprint arXiv:1702.08608.
- Dosovitskiy, A. (2020). An image is worth 16×16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929.
- Ethala, S., Kodipunjula, B. S., Appalaneni, U. K., Narsagari, M. ve Wahaz, N. B. (2025, Şubat). A comprehensive review of computerized respiratory sound analysis and deep learning techniques for acoustic signal-based disease classification. *AIP Conference Proceedings*, 3237(1), 040035.

- Fava, A., Dianat, B., Bertacchini, A., Manfredi, A., Sebastiani, M., Modena, M. ve Pancaldi, F. (2024). Pre-processing techniques to enhance the classification of lung sounds based on deep learning. *Biomedical Signal Processing and Control*, 92, 106009.
- Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, 7(2), 179-188.
- Fraiwan, M., Fraiwan, L., Khassawneh, B. ve Ibnian, A. (2021). A dataset of lung sounds recorded from the chest wall using an electronic stethoscope. *Data in Brief*, 35, 106913.
- Freund, Y. ve Schapire, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55(1), 119-139.
- Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, 29, 1189-1232.
- Gairola, S., Tom, F., Kwatra, N. ve Jain, M. (2021, Kasım). RespireNet: A deep neural network for accurately detecting abnormal lung sounds in limited data setting. 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), (ss. 527-530). IEEE.
- Global Burden of Disease Study (GBD). (2019). Global burden of 369 diseases and injuries in 204 countries and territories, 1990-2019: A systematic analysis for the Global Burden of Disease Study 2019. *The Lancet*, 396(10258), 1204-1222.
- Gong, Y., Chung, Y. A. ve Glass, J. (2021). AST: Audio spectrogram transformer. arXiv preprint arXiv:2104.01778.
- Gong, Y., Lai, C. I., Chung, Y. A. ve Glass, J. (2022). SSAST: Self-supervised audio spectrogram transformer. *Proceedings of the AAAI Conference on Artificial Intelligence*, 36(10), 10699-10709.
- Goodfellow, I., Bengio, Y. ve Courville, A. (2016). *Deep learning*. Cambridge: MIT Press.
- Grenier, M. C., Gagnon, K., Genest, J., Durand, J. ve Durand, L. G. (1998). Clinical comparison of acoustic and electronic stethoscopes and design of a new electronic stethoscope. *American Journal of Cardiology*, 81(5), 653-656.
- Gulati, A., Qin, J., Chiu, C. C., Parmar, N., Zhang, Y., Yu, J. ... Pang, R. (2020). Conformer: Convolution-augmented transformer for speech recognition. arXiv preprint arXiv:2005.08100.

- Han, K., Wang, Y., Tian, Q., Guo, J., Xu, C. ve Xu, C. (2020). GhostNet: More features from cheap operations. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, (ss. 1580-1589).
- Hastie, T., Tibshirani, R., Friedman, J. ve Franklin, J. (2005). The elements of statistical learning: Data mining, inference and prediction. *The Mathematical Intelligencer*, 27(2), 83-85.
- He, H., Bai, Y., Garcia, E. A. ve Li, S. (2008). ADASYN: Adaptive synthetic sampling approach for imbalanced learning. 2008 IEEE International Joint Conference on Neural Networks, (ss. 1322-1328). IEEE.
- He, K., Zhang, X., Ren, S. ve Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (ss. 770-778).
- Hochreiter, S. ve Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735-1780.
- Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T. ... Adam, H. (2017). MobileNets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861.
- Hsu, F. S., Huang, S. R., Huang, C. W., Huang, C. J., Cheng, Y. R., Chen, C. C. ... Lai, F. (2021). Benchmarking of eight recurrent neural network variants for breath phase and adventitious sound detection on a self-developed open-access lung sound database—HF_Lung_V1. *PLOS ONE*, 16(7), e0254134.
- Huang, D. M., Huang, J., Qiao, K., Zhong, N. S., Lu, H. Z. ve Wang, W. J. (2023). Deep learning-based lung sound analysis for intelligent stethoscope. *Military Medical Research*, 10(1), 44.
- International Agency for Research on Cancer (IARC). (2022). *Global Cancer Observatory: Cancer today (GLOBOCAN 2022)*. Lyon, France: WHO/IARC.
- ICBHI Challenge. (2017). Respiratory Sound Database. Erişim adresi: https://bhichallenge.med.auth.gr/ICBHI_2017_Challenge
- İçer, S. ve Gengeç, Ş. (2014). Classification and analysis of non-stationary characteristics of crackle and rhonchus lung adventitious sounds. *Digital Signal Processing*, 28, 18-27.
- Jaegle, A., Gimeno, F., Brock, A., Vinyals, O., Zisserman, A. ve Carreira, J. (2021). Perceiver: General perception with iterative attention. *International Conference on Machine Learning*, (ss. 4651-4664).

- Jiang, D. N., Lu, L., Zhang, H. J., Tao, J. H. ve Cai, L. H. (2002). Music type classification by spectral contrast feature. *IEEE International Conference on Multimedia and Expo*, 1, 113-116.
- Jin, F., Sattar, F. ve Goh, D. Y. (2014). New approaches for spectro-temporal feature extraction with applications to respiratory sound classification. *Neurocomputing*, 123, 362-371.
- Kandaswamy, A., Kumar, C. S., Ramanathan, R. P., Jayaraman, S. ve Malmurugan, N. (2004). Neural classification of lung sounds using wavelet coefficients. *Computers in Biology and Medicine*, 34(6), 523-537.
- Kairouz, P., McMahan, H. B., Avent, B., Bellet, A., Bennis, M., Bhagoji, A. N. ... Zhao, S. (2021). Advances and open problems in federated learning. *Foundations and Trends in Machine Learning*, 14(1-2), 1-210.
- Khan, R., Khan, S. U., Saeed, U. ve 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. ve Jung, H. Y. (2024). RepAugment: Input-agnostic representation-level augmentation for respiratory sound classification. 2024 46th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), (ss. 1-6). IEEE.
- Kim, Y., Hyon, Y., Woo, S. D., Lee, S., Lee, S. I., Ha, T. ve Chung, C. (2023). Evolution of the stethoscope: Advances with the adoption of machine learning and development of wearable devices. *Tuberculosis and Respiratory Diseases*, 86(4), 251.
- Kim, Y., Hyon, Y., Lee, S., Woo, S. D., Ha, T. ve Chung, C. (2022). The coming era of a new auscultation system for analyzing respiratory sounds. *BMC Pulmonary Medicine*, 22(1), 119.
- Kim, J. W., Yoon, C., Toikkanen, M., Bae, S. ve Jung, H. Y. (2023). Adversarial fine-tuning using generated respiratory sound to address class imbalance. arXiv preprint arXiv:2311.06480.
- Kinsler, L. E., Frey, A. R., Coppens, A. B. ve Sanders, J. V. (2000). *Fundamentals of acoustics*. John Wiley & Sons.
- Ko, T., Peddinti, V., Povey, D., Seltzer, M. L. ve Khudanpur, S. (2017). A study on data augmentation of reverberant speech for robust speech recognition. *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, (ss. 5220-5224). IEEE.

- Kong, Q., Cao, Y., Iqbal, T., Wang, Y., Wang, W. ve Plumbley, M. D. (2020). PANNs: Large-scale pretrained audio neural networks for audio pattern recognition. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 28, 2880-2894.
- Latif, S., Rana, R., Khalifa, S., Jurdak, R., Qadir, J. ve Schuller, B. W. (2020). Deep representation learning in speech processing: Challenges, recent advances, and future trends. arXiv preprint arXiv:2001.00378.
- LeCun, Y., Bengio, Y. ve Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436-444.
- Liu, R., Cai, S., Zhang, K. ve Hu, N. (2019). Detection of adventitious respiratory sounds based on convolutional neural network. 2019 International Conference on Intelligent Informatics and Biomedical Sciences (ICIIBMS), (ss. 298-303). IEEE.
- Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z. ... Guo, B. (2021). Swin transformer: Hierarchical vision transformer using shifted windows. Proceedings of the IEEE/CVF International Conference on Computer Vision, (ss. 10012-10022).
- Makhoul, J. (2005). Linear prediction: A tutorial review. *Proceedings of the IEEE*, 63(4), 561-580.
- Mendes, L., Vogiatzis, I. M., Perantoni, E., Kaimakamis, E., Chouvarda, I., Maglaveras, N. ... Paiva, R. P. (2016). Detection of crackle events using a multi-feature approach. 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), (ss. 3679-3683). IEEE.
- Murphy, K. P. (2012). *Machine learning: A probabilistic perspective*. MIT Press.
- Nguyen, T. ve Pernkopf, F. (2022). Lung sound classification using co-tuning and stochastic normalization. *IEEE Transactions on Biomedical Engineering*, 69(9), 2872-2882.
- Oppenheim, A. V. (1999). *Discrete-time signal processing*. Pearson Education India.
- Park, D. S., Chan, W., Zhang, Y., Chiu, C. C., Zoph, B., Cubuk, E. D. ve Le, Q. V. (2019). SpecAugment: A simple data augmentation method for automatic speech recognition. arXiv preprint arXiv:1904.08779.
- Pasterkamp, H., Brand, P. L., Everard, M., Garcia-Marcos, L., Melbye, H. ve Priftis, K. N. (2016). Towards the standardisation of lung sound nomenclature. *European Respiratory Journal*, 47(3), 724-732.
- Pasterkamp, H., Kraman, S. S. ve Wodicka, G. R. (1997). Respiratory sounds: Advances beyond the stethoscope. *American Journal of Respiratory and Critical Care Medicine*, 156(3), 974-987.

- Petmezas, G., Cheimariotis, G. A., Stefanopoulos, L., Rocha, B., Paiva, R. P., Katsaggelos, A. K. ve Maglaveras, N. (2022). Automated lung sound classification using a hybrid CNN-LSTM network and focal loss function. *Sensors*, 22(3), 1232.
- Perna, D. ve Tagarelli, A. (2019). Deep auscultation: Predicting respiratory anomalies and diseases via recurrent neural networks. 2019 IEEE 32nd International Symposium on Computer-Based Medical Systems (CBMS), (ss. 50-55). IEEE.
- Pomietlo, M. (2020). *Open RN: Nursing skills*. Chippewa Valley Technical College.
- Powers, D. M. (2020). Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation. arXiv preprint arXiv:2010.16061.
- Pramono, R. X. A., Bowyer, S. ve Rodriguez-Villegas, E. (2017). Automatic adventitious respiratory sound analysis: A systematic review. *PLOS ONE*, 12(5), e0177926.
- Quinlan, J. R. (2014). *C4.5: Programs for machine learning*. Elsevier.
- Rabiner, L. R. (1978). *Digital processing of speech signals*. Pearson Education India.
- Reichert, S., Gass, R., Brandt, C. ve Andrès, E. (2008). Analysis of respiratory sounds: State of the art. *Clinical Medicine: Circulatory, Respiratory and Pulmonary Medicine*, 2, CCRPM-S530.
- Rocha, B. M., Filos, D., Mendes, L., Vogiatzis, I., Perantoni, E., Kaimakamis, E. ... Maglaveras, N. (2017). A respiratory sound database for the development of automated classification. *International Conference on Biomedical and Health Informatics*, (ss. 33-37). Springer.
- Rocha, B. M., Filos, D., Mendes, L., Serbes, G., Ulukaya, S., Kahya, Y. P. ... De Carvalho, P. (2019). An open access database for the evaluation of respiratory sound classification algorithms. *Physiological Measurement*, 40(3), 035001.
- Roguin, A. (2006). René Théophile Hyacinthe Laënnec (1781–1826): The man behind the stethoscope. *Clinical Medicine & Research*, 4(3), 230-235.
- Rossing, T. (Der.). (2007). *Springer handbook of acoustics*. Springer Science & Business Media.
- Rossing, T. D., Moore, R. F. ve Wheeler, P. A. (2013). *The science of sound*. Pearson Higher Ed.
- Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, 1(5), 206-215.
- Salamon, J. ve Bello, J. P. (2017). Deep convolutional neural networks and data augmentation for environmental sound classification. *IEEE Signal Processing Letters*, 24(3), 279-283.

- Sarkar, M., Madabhavi, I., Niranjana, N. ve Dogra, M. (2015). Auscultation of the respiratory system. *Annals of Thoracic Medicine*, 10(3), 158-168.
- Science Museum Group Studio. (2019). Laennec's stethoscope [Photograph]. Science Museum Group Collection. Erişim adresi: <https://collection.sciencemuseumgroup.org.uk/objects/co90986>
- Seah, J. J., Zhao, J., Wang, D. Y. ve Lee, H. P. (2023). Review on the advancements of stethoscope types in chest auscultation. *Diagnostics*, 13(9), 1545.
- Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D. ve Batra, D. (2017). Grad-CAM: Visual explanations from deep networks via gradient-based localization. Proceedings of the IEEE International Conference on Computer Vision, (ss. 618-626).
- Serbes, G., Ulukaya, S. ve Kahya, Y. P. (2017). An automated lung sound preprocessing and classification system based on spectral analysis methods. International Conference on Biomedical and Health Informatics, (ss. 45-49). Springer.
- Sfayyih, A. H., Sulaiman, N. ve Sabry, A. H. (2023). A review on lung disease recognition by acoustic signal analysis with deep learning networks. *Journal of Big Data*, 10(1), 101.
- Sharma, N., Krishnan, P., Kumar, R., Ramoji, S., Chetupalli, S. R., Ghosh, P. K. ve Ganapathy, S. (2020). Coswara—A database of breathing, cough, and voice sounds for COVID-19 diagnosis. arXiv preprint arXiv:2005.10548.
- Sokolova, M. ve Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing & Management*, 45(4), 427-437.
- Sovijärvi, A., Vanderschoot, J., Earis, J., Heliö, P. ve Dalmaso, F. (1996). Computerized respiratory sound analysis (CORSAs): Techniques, standardization and clinical evaluation. 21st International Conference on Lung Sounds, Chester, England.
- Sovijärvi, A. R. A., Vanderschoot, J. ve Earis, J. E. (Der.). (2000). *Computerized respiratory sound analysis (CORSAs): Recommended standards for terms and techniques*. Munksgaard.
- Tan, M. ve Le, Q. (2019). EfficientNet: Rethinking model scaling for convolutional neural networks. International Conference on Machine Learning, (ss. 6105-6114).
- Taplidou, S. A. ve Hadjileontiadis, L. J. (2007). Wheeze detection based on time-frequency analysis of breath sounds. *Computers in Biology and Medicine*, 37(8), 1073-1083.

- Tuncer, T., Aydemir, E., Özyurt, F., Doğan, S., Belhaouari, S. B. ve Akbal, E. (2021). An automated COVID-19 respiratory sound classification method based on novel local symmetric Euclidean distance pattern and ReliefF iterative MRMR feature selector. *International Advanced Researches and Engineering Journal*, 5(3), 334-343.
- Tzanetakis, G. ve Cook, P. (2002). Musical genre classification of audio signals. *IEEE Transactions on Speech and Audio Processing*, 10(5), 293-302.
- URL-1, <https://www.webtekno.com/stetoskoplar-nasil-calisir-h145052.html> 15 Aralık 2025.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N. ... Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 30.
- West, J. B. (2012). *Respiratory physiology: The essentials*. Lippincott Williams & Wilkins.
- World Health Organization (WHO). (2023). *Asthma fact sheet*. Erişim adresi: <https://www.who.int/news-room/fact-sheets/detail/asthma>
- World Health Organization (WHO). (2023). *Chronic obstructive pulmonary disease (COPD) fact sheet*. Erişim adresi: [https://www.who.int/news-room/fact-sheets/detail/chronic-obstructive-pulmonary-disease-\(copd\)](https://www.who.int/news-room/fact-sheets/detail/chronic-obstructive-pulmonary-disease-(copd))
- World Health Organization (WHO). (2024). *Coronavirus (COVID-19) dashboard*. Erişim adresi: <https://covid19.who.int/>
- World Health Organization (WHO). (2024). *Global Health Estimates 2021: Deaths by cause, age, sex, by country and by region, 2000–2021*. Erişim adresi: <https://www.who.int/data/gho/data/themes/mortality-and-global-health-estimates/ghe-leading-causes-of-death>
- World Health Organization (WHO). (2024). *Pneumonia fact sheet*. Erişim adresi: <https://www.who.int/news-room/fact-sheets/detail/pneumonia>
- World Health Organization (WHO). (2023). *Global tuberculosis report 2023*. Geneva: World Health Organization.
- Wu, C., Ye, N. ve Jiang, J. (2024). Classification and recognition of lung sounds based on improved Bi-ResNet model. *IEEE Access*, 12, 73079-73094.
- Xia, T., Han, J. ve Mascolo, C. (2022). Exploring machine learning for audio-based respiratory condition screening: A concise review of databases, methods, and open issues. *Experimental Biology and Medicine*, 247(22), 2053-2061.

- Xu, M. ve Wiese, L. (2023). Application and performance improvement of transfer learning on ICBHI lung sound dataset. SAI Intelligent Systems Conference, (ss. 156-173). Cham: Springer Nature Switzerland.
- Xu, X. ve Sankar, R. (2024). Classification and recognition of lung sounds using artificial intelligence and machine learning: A literature review. *Big Data and Cognitive Computing*, 8(10), 127.
- Yokuş, A. (2011). *Bazı özel lineer olmayan diferensiyel denklemlerin çözümlerinin elde edilmesi ve bu çözümlerin karşılaştırılması*. Doktora tezi, Fen Bilimleri Enstitüsü, Elazığ.
- Yu, S., Ding, Y., Qian, K., Hu, B., Li, W. ve Schuller, B. W. (2022). A glance-and-gaze network for respiratory sound classification. ICASSP 2022 IEEE International Conference on Acoustics, Speech and Signal Processing, (ss. 9007-9011).
- Zhang, W. ve Liu, X. (2024). Decoding breath: Machine learning advancements in diagnosing pulmonary diseases via lung sound analysis. *Scientific and Engineering Letters*, 12(1), 1-11.
- Zimmerman, B. ve Williams, D. (2023). Lung sounds. In *StatPearls*. StatPearls Publishing.
- Zulfiqar, R., Majeed, F., Irfan, R., Rauf, H. T., Benkhelifa, E. ve Belkacem, A. N. (2021). Abnormal respiratory sounds classification using deep CNN through artificial noise addition. *Frontiers in Medicine*, 8, 714811.
- Zwicker, E. ve Fastl, H. (1999). The ear's own nonlinear distortion. In *Psychoacoustics: Facts and models* (ss. 277-292). Berlin, Heidelberg: Springer.

ÖZGEÇMİŞ

Eren AYGÜN lise öğrenimini Tevfik Serdar Anadolu Lisesinde tamamladıktan sonra lisans eğitimini Karadeniz Teknik Üniversitesi Bilgisayar Mühendisliği Bölümünde 2017 yılında tamamlamıştır. 2018 yılından itibaren Trabzon İçmesuyu ve Kanalizasyon İdaresi Genel Müdürlüğünde Bilgisayar Mühendisi olarak görev yapmaktadır.

Yayınlar:

Türe, H., ve Aygün, E. (2025). Solunum Sesi Sınıflandırması için Klasik ve Derin Öğrenme Modellerinin Karşılaştırılması. Karadeniz Fen Bilimleri Dergisi, 15(4), 1668–1695.