

**SOLUNUM SESLERİNİN EVRİŞİMSEL SİNİR AĞLARIYLA
SINIFLANDIRILMASI**

**CLASSIFICATION OF LUNG SOUNDS WITH
CONVOLUTIONAL NEURAL NETWORKS**

FUNDA CİNYOL

PROF. DR. UĞUR BAYSAL

Tez Danışmanı

Hacettepe Üniversitesi
Lisansüstü Eğitim-Öğretim ve Sınav Yönetmeliğinin
Elektrik ve Elektronik Mühendisliği Anabilim Dalı için Öngördüğü
DOKTORA TEZİ olarak hazırlanmıştır.

2022

ÖZET

SOLUNUM SESLERİNİN EVRİŞİMSEL SİNİR AĞLARIYLA SINIFLANDIRILMASI

FUNDA CİNYOL

Doktora, Elektrik ve Elektronik Mühendisliği Bölümü

Tez Danışmanı: Prof. Dr. Uğur BAYSAL

Haziran 2022, 92 sayfa

Hekimler klinik muayenede hem non-invaziv olmasından hem de kolay erişilebilir aynı zamanda ucuz bir teşhis ve tespit aracı olmasından dolayı stetoskop kullanmaktadırlar. Ancak bu aşamada ortam gürültüsü, hekimin işitme kabiliyeti, yaşı vs gibi bir çok olumsuz neden teşhisi zorlaştırmaktadır. Ayrıca solunum sesleri insan işitme kabiliyetinin alt sınırlarına yakın frekanslarda önemli bilgiler içermektedir. Daha niceliksel bir yaklaşım sunulabilmesi açısından elektronik stetoskopla bu seslerin kayıt edilmesi önem taşımaktadır. Bu tez çalışmasıyla tamamen klinik ortamda toplanmış 294 adet solunum sesi 3 farklı ses grubu (Normal, Ral, Ronküs) için sınıflandırılmıştır.

Solunum seslerinin yapay zeka yöntemleriyle sınıflandırılması ile ilgili birçok çalışma yapılsa da kullanılan veri sayılarındaki yetersizlik sağlam bir mimari oluşturulmasına engel olmuştur. Bu eksiklikle baş etmek amacıyla ICBHI 2017 veri kümesi oluşturulmuş ve solunum seslerinin sınıflandırılması çalışmaları hız kazanmıştır.

Evrışimsel Sinir Ağları (ESA); görüntü ve video tanıma, görüntü sınıflandırma, doğal dil işleme ve tıbbi görüntü analizi gibi birçok alanda kullanılan Derin Sinir Ağlarına ait solunum ses sınıflandırmada da sıklıkla kullanılan bir yöntemdir. ESA'larda genellikle son katmanda sınıflandırma amacıyla softmax fonksiyonu kullanılmaktadır. Ancak sa-

dece softmax fonksiyonunun kullanılması yerine Destek Vektör Makinelerinin (DVM) kullanımı da literatürde yer almaktadır. Bu tez çalışmasında hem ESA'nın son katmanda Softmax kullanmasıyla oluşturulan mimariler (ESA-Softmax), hem de ESA'nın son katmanda DVM ile birlikte kullanıldığı (ESA-DVM) mimariler oluşturulmuştur. Ayrıca aktarmalı öğrenme yardımıyla VGG16 modeli ile birlikte oluşturulan ESA-Softmax, ESA-DVM modelleri birleştirilerek VGG16-ESA-Softmax ve VGG16-ESA-DVM mimarileri oluşturulmuştur. Bu mimariler literatürde başarısı kanıtlanmış öğrenme aktarımı yöntemleri ile (VGG16, DenseNet201, InceptionV3 ve ResNet101) de kıyaslanmıştır. Eğitim ve test verilerinin %80- %20 oranında ayrılarak 10 kat çapraz doğrulama uygulanması sonucunda elde edilen en yüksek sınıflandırma doğruluğu VGG16-ESA-DVM modeli ile gerçekleştirilmiştir. Önerilen yöntemin sınıflandırma başarı metrikleri sırasıyla ; ROC AUC Skoru: %88.4, Maksimum Doğruluk: %83, kesinlik= %82, duyarlılık= %83, F1 skoru= %82 olarak elde edilmiştir. Ayrıca üç grup ses verisi için sırasıyla Normal %84 , Ral %80 ve Ronküs ise %86 doğrulukla sınıflandırılmıştır.

Anahtar Kelimeler: Evrişimsel Sinir Ağları, Solunum Sesleri, Sınıflandırma, DVM

ABSTRACT

CLASSIFICATION OF LUNG SOUNDS WITH CONVOLUTIONAL NEURAL NETWORKS

FUNDA CİNYOL

**Doctor of Philosophy, Department of
Electrical and Electronics Engineering**

Supervisor: Prof. Dr. Uğur BAYSAL

June 2022, 92 pages

Physicians use stethoscope in clinical examination because it is both non-invasive and easily accessible and inexpensive diagnostic and detection tool. However, at this stage, many negative reasons such as ambient noise, hearing ability of the physician, age, etc. make the diagnosis difficult. In addition, respiratory sounds contain important information at frequencies close to the lower limits of human hearing ability. It is important to record these sounds with an electronic stethoscope in order to present a more quantitative approach. In this thesis, 294 lung sounds collected in a clinical setting were classified for 3 different sound groups (Normal, Crackle, Rhonchi).

Although many studies have been carried out on the classification of respiratory sounds with artificial intelligence methods, the inadequacy of the data used has prevented the creation of a solid architecture. In order to cope with this shortcoming, the ICBHI 2017 dataset was created and the classification studies of lung sounds gained momentum.

Convolutional Neural Networks (CNNs); It is a method frequently used in breathing sound classification of Deep Neural Networks, which is used in many areas such as image and video recognition, image classification, natural language processing and medical image analysis. In CNNs, softmax function is generally used for classification in the last layer. However, instead of using only the softmax function, the use of Support Vector

Machines (SVM) is also included in the literature. In this thesis, both the architectures created by using Softmax in the last layer of CNN (CNN-Softmax) and the architectures in which CNN is used together with SVM in the last layer (CNN-SVM) were created. In addition, VGG16-CNN-Softmax and VGG16-CNN-SVM architectures were created by combining the CNN-Softmax, CNN-SVM models created with the VGG16 model with the help of transfer learning. These architectures have also been compared with proven learning transfer methods in the literature (VGG16, DenseNet201, InceptionV3 and ResNet101). These architectures have also been compared with proven learning transfer methods in the literature (VGG16, DenseNet201, InceptionV3 and ResNet101). The highest classification accuracy obtained as a result of 10-fold cross validation by separating the training and test data by 80%- 20% was achieved with the VGG16-CNN-SVM model. The classification success metrics of the proposed method are obtained as respectively; ROC AUC Score: 88.4%, Maximum Accuracy: 83%, precision= 82%, recall= 83% and F1 score= 82%. In addition, for three groups of voice data, Normal 84%, Crackle 80% and Rhonchi were classified with 86% accuracy, respectively.

Keywords: Convolutional Neural Network, Lung Sounds, Classification, SVM

TEŞEKKÜR

Akademik hayatım boyunca bilgi ve deneyimleri ile bana yol gösteren, bu süreçte birçok fırsat sunan tez danışmanım Sayın Prof. Dr. Uğur BAYSAL'a,

Tez İzleme Komitesi üyeleri olarak çalışmalarımı takip eden ve önerilerini benimle paylaşan Sayın Prof. Dr. Ethem GELİR'e ve Sayın Doç. Dr. Seniha YÜKSEL ERDEM'e,

Klinik verilerin toplanması için çalışan, tıbbi bilgi ve deneyimleri ile bana yol gösteren Prof.Dr. Deniz Köksal, Doç. Dr. Elif Babaoğlu ve Doç.Dr. Sevinç Sarıncı Ulaşlı'ya,

Hayatım boyunca desteklerini esirgemeyen, güven ve sevgilerini her an hissettiğim canım aileme,

Karşılaştığım tüm zorlukları benimle birlikte göğüsleyen, bu zorlu süreçte de anlayışını ve desteğini benden hiç eksik etmeyen sevgili eşim Sedat CİNYOL'a,

sonsuz teşekkürlerimi sunuyorum.

Funda CİNYOL

Haziran 2022, Ankara

İÇİNDEKİLER

ÖZET	i
ABSTRACT.....	iii
TEŞEKKÜR	v
İÇİNDEKİLER	vi
ÇİZELGELER.....	ix
ŞEKİLLER.....	x
KISALTMALAR.....	xiii
1. GİRİŞ	1
1.1. Tezin Akışı.....	2
2. SOLUNUM SESLERİ	3
2.1. Solunum Sesi Türleri	4
2.2. Solunum Seslerini Sınıflandırmanın Önemi	9
3. MAKİNE ÖĞRENMESİ	11
3.1. Derin Öğrenme	12
3.1.1. Algılayıcılar	12
3.2. Sınıflandırıcılar	13
3.2.1. Lineer Sınıflandırıcılar	13
3.2.2. Destek Vektör Makineleri	16
3.2.3. Lojistik Regresyon.....	19
3.3. EVRİŞİMSEL SİNİR AĞLARI	20
3.3.1. Ağın Eğitilmesinde Kullanılan Parametreler	20
3.3.1.1. Maliyet Fonksiyonları	20
3.3.2. Aktivasyon Fonksiyonları	21
3.3.2.1. Sigmoid Fonksiyonu	21
3.3.2.2. Tanh Fonksiyonu	22
3.3.2.3. ReLU Fonksiyonu	23
3.3.3. Eniyileme (Optimizasyon) Türleri	24

3.3.3.1.	Paket Eğitim Düşümü (Batch Gradient Descent)	25
3.3.3.2.	Olasılıksal Eğitim Düşümü (Stochastic Gradient Descent - SGD)	26
3.3.3.3.	Mini Paket Eğitim Düşümü (Mini-Batch Gradient Descent)	26
3.3.3.4.	Eğitim Düşümü Optimizasyonu (Gradient Descent Optimization)	26
4.	ÖN İŞLEMLER	28
4.1.	Gürültü Bastırma	28
4.2.	Solunum Sesi Sinyali'nin Görüntüye Dönüştürülme Yöntemleri	30
4.2.1.	Gabor Dönüşümü ve Spektrogram	30
4.2.1.1.	Ayrık Gabor Dönüşümü	34
4.2.2.	Mel Spektrogramı	37
4.3.	Veri Artırma (Data Augmentation) Teknikleri	41
5.	AKTARMALI ÖĞRENME (TRANSFER LEARNING)	45
5.1.	VGG	46
5.2.	ResNet	49
6.	SOLUNUM SESLERİNİN DERİN ÖĞRENME YÖNTEMLERİ İLE SINIFLANDIRILMASI UYGULAMALARI	55
6.1.	Veri Kümeleri	55
6.2.	Değerlendirme Yöntemleri	58
6.3.	ESA ile Sınıflandırma	60
6.4.	Aktarmalı Öğrenme ve ESA ile Sınıflandırma	63
6.4.1.	VGG16-ESA-Softmax Mimarisi ile Elde Edilen Sonuçlar	65
6.4.2.	VGG16-ESA-DVM Mimarisi ile Elde Edilen Sonuçlar	69
6.5.	Başarısı Kanıtlanmış ESA Modellerinin Aktarmalı Öğrenme ile Solunum Sesi Sınıflandırmada Kullanılması	74
6.5.1.	İki Sınıf İçin VGG16-ESA-DVM Modeli ile Elde Edilen Sonuçlar	76
6.6.	VGG16-ESA-DVM Mimarisinin Açık Kaynak Veri Kümesi Üzerindeki Performansının İncelenmesi	78
7.	TARTIŞMA	81

8. SONUÇ	84
KAYNAKLAR	85
9. EKLER.....	91
9.1. Tezde Kullanılan Yazılımsal ve Donanımsal Özellikler	91
9.2. Tezden Üretilen Yayınlar	91
9.3. Tez Kapsamında Tamamlanmış Projeler.....	91
ÖZGEÇMİŞ	92



ÇİZELGELER

Çizelge 2.1.	Solunum Seslerinin Türleri [13]	4
Çizelge 3.1.	Eniyileme Algoritmaları	27
Çizelge 4.1.	PSD Sözde Kod (pseudocode) Parçacığı	28
Çizelge 6.1.	Oluşturulan Veri Grupları	57
Çizelge 6.2.	ESA-Softmax Mimarisi Sınıflandırma Başarıları (Gabor Spektrogramı)	61
Çizelge 6.3.	ESA-Softmax Sınıflandırma Başarıları (Mel Spektrogramı)	62
Çizelge 6.4.	ESA-DVM Sınıflandırma Başarıları (Gabor Spektrogramı)	62
Çizelge 6.5.	ESA-DVM Mimarisi Sınıflandırma Başarıları (Mel Spektrogramı) ...	62
Çizelge 6.6.	VGG16-ESA-Softmax ve VGG16-ESA-DVM Mimarilerinin Sınıflandırma Başarıları (Gabor Spektrogramı)	64
Çizelge 6.7.	VGG16-ESA-Softmax ve VGG16-ESA-DVM Modelleri ile Elde Edilen Sınıflandırma Başarıları (Mel Spektrogramı)	64
Çizelge 6.8.	Aktarmalı Öğrenme ile Elde Edilen Sınıflandırma Başarıları	75
Çizelge 6.9.	VGG16-ESA-DVM Mimarisi ile Elde Edilen İkili Sınıflandırma Başarıları Gabor Spektrogramı	77
Çizelge 6.10.	ICBHI Verileri ile Tez Çalışmasında Kullanılan Verilerin Genel Çerçevesi	78
Çizelge 7.1.	Literatürde Solunum Sesi Sınıflandırması Yapılan Çalışmalar	81

ŞEKİLLER

Şekil 2.1.	Klinik Pratikte Oskültasyon Analizinin Doğruluğu [31]	7
Şekil 2.2.	Klinik Pratikte Oskültasyon Analizi ile Ronküs Doğruluğu [31]	8
Şekil 3.1.	Algılayıcılar	13
Şekil 3.2.	Lineer Sınıflandırıcı Karar Sınırı	14
Şekil 3.3.	Sınıflandırıcı Katsayıları	15
Şekil 3.4.	Lineer Sınıflandırıcı Modeli	16
Şekil 3.5.	Veriyi Hiper Düzlem ile Sınıflandırma	17
Şekil 3.6.	Hiper Düzlemi Ayıran Maksimum Marjın Gösterimi	18
Şekil 3.7.	Sigmoid Fonksiyonu	22
Şekil 3.8.	Tanh Fonksiyonu	23
Şekil 3.9.	ReLU Fonksiyonu	24
Şekil 3.10.	Eğimin ters yönünde ağırlık güncelleme işlemi	25
Şekil 4.1.	Solunum Sesi Sinyali	28
Şekil 4.2.	Ses Sinyalinin Güç Spektral Yoğunluğu	29
Şekil 4.3.	Gürültülü ve Filtrelenmiş Solunum Sesinin Güç Spektral Yoğunluğu ...	29
Şekil 4.4.	Gürültülü ve Filtrelenmiş Solunum Sesi	30
Şekil 4.5.	Kısa Zamanlı Fourier Dönüşümü'nün Gösterimi	31
Şekil 4.6.	Hamming Penceresi	32
Şekil 4.7.	Gabor dönüşümünün Gauss penceresiyle elde edilmesi [42]	33
Şekil 4.8.	İkinci dereceden cıvıltı sinyalinin güç spektral yoğunluğu [42]	34
Şekil 4.9.	İkinci dereceden cıvıltı sinyalinin spektrogramı [42]	35
Şekil 4.10.	Normal Solunum Sesi Sinyali ve Spektrogram Görüntüsü.	35
Şekil 4.11.	Ral Solunum Sesi Sinyali ve Spektrogram Görüntüsü.	36
Şekil 4.12.	Ronküs Solunum Sesi Sinyali ve Spektrogram Görüntüsü.	36
Şekil 4.13.	Çerçeveleme	38
Şekil 4.14.	Mel Ölçeği Filtre Bankasının Gösterimi	40
Şekil 4.15.	Mel Spektrogram	41

Şekil 4.16. Kazanç Katsayısı ile Veri Artırma	42
Şekil 4.17. Zaman Genleşme ile Veri Artırma	43
Şekil 4.18. Perde Kaydırma	43
Şekil 4.19. Beyaz Gürültü Ekleme Yöntemi ile Veri Artırma	44
Şekil 5.1. Aktarmalı Öğrenme	45
Şekil 5.2. VGG’de kullanılan konfigürasyonlar	48
Şekil 5.3. VGG-16’nın Mimarisi	49
Şekil 5.4. Artık Öğrenme Bloğu (Block of residual learning)	50
Şekil 5.5. ResNet’in Kısmi Mimarisi	51
Şekil 5.6. Resnet-18 Modeli ve Res Blok Yapısı	52
Şekil 5.7. Resnet-18 Modeli Evrişim ve Identity Blok Yapısı	53
Şekil 6.1. Veri Kümesinin Dağılımı.....	55
Şekil 6.2. Normal, Ral ve Ronküs Seslerinin Ortalama Değerleri	56
Şekil 6.3. Normal, Ral ve Ronküs Seslerinin Maksimum Değerleri	57
Şekil 6.4. Hata Matrisi Örneği	58
Şekil 6.5. ROC Eğrisi	59
Şekil 6.6. Solunum Seslerinin ESA ile Softmax veya DVM Kullanarak Sınıflandırılma Mimarisi	60
Şekil 6.7. Öğrenim Aktarımı Kullanılarak Solunum Seslerinin Sınıflandırılma Mimarisi	63
Şekil 6.8. VGG16-ESA-Softmax Mimarisi Hata Matrisi (294 Adet Veri)	65
Şekil 6.9. VGG16-ESA-Softmax Mimarisi ROC Eğrisi (294 Adet Veri)	66
Şekil 6.10. VGG16-ESA-Softmax Mimarisi Hata Matrisi (1470 Adet Veri)	67
Şekil 6.11. VGG16-ESA-Softmax Mimarisi ROC Eğrisi (1470 Adet Veri)	67
Şekil 6.12. VGG16-ESA-Softmax Mimarisi Hata Matrisi (Gürültüsüzleştirilmiş Veriler)	68
Şekil 6.13. VGG16-ESA-Softmax Mimarisi ROC Eğrisi (Gürültüsüzleştirilmiş Veriler)	69
Şekil 6.14. VGG16-ESA-DVM Mimarisi Hata Matrisi (294 Adet Veri)	70
Şekil 6.15. VGG16-ESA-DVM Mimarisi ROC Eğrisi (294 Adet Veri)	70

Şekil 6.16. VGG16-ESA-DVM Mimarisi Hata Matrisi (1470 Adet Veri)	71
Şekil 6.17. VGG16-ESA-DVM Mimarisi ROC Eğrisi (1470 Adet Veri)	72
Şekil 6.18. VGG16-ESA-DVM Mimarisi Hata Matrisi (Gürültüsüzleştirilmiş Veriler)	73
Şekil 6.19. VGG16-ESA-DVM Mimarisi ROC Eğrisi (Gürültüsüzleştirilmiş Veriler)	73
Şekil 6.20. Aktarmalı Öğrenme ile Solunum Seslerinin Sınıflandırılması	75
Şekil 6.21. VGG16-ESA-DVM Mimarisi Hata Matrisi (2 Sınıf)	76
Şekil 6.22. VGG16-ESA-DVM Mimarisi ROC Eğrisi (2 Sınıf)	77
Şekil 6.23. VGG16-ESA-DVM Mimarisi Hata Matrisi (ICBHI verileri)	79
Şekil 6.24. VGG16-ESA-DVM Mimarisi ROC Eğrisi (ICBHI verileri)	79

KISALTMALAR

Kısaltmalar

YSA	Yapay Sinir Ağı
ESA	Evrişimsel Sinir Ağı
PSD	Power Spectral Density (Güç Spektral Yoğunluğu)
DVM	Destek Vektor Makineleri
KZFD	Kısa Zamanlı Fourier Dönüşümü
SGD	Stochastic Gradient Descent (Olasılıksal Eğitim Düşümü)
ROC	Receiver Operating Characteristics (Karar Değerlendirme Grafiği)
DSÖ	Dünya Sağlık Örgütü
KOAH	Kronik Obstrüktif Akciğer Hastalığı

1. GİRİŞ

İnsanoğlu var olduğundan beri birçok hastalık ile mücadele etmiştir. Hastalık ve türleri değiştikçe modern tıp bilimi de bu değişimlere uygun yeni tedavi yöntemleri geliştirmektedir. Dünya Sağlık Örgütü (DSÖ); elde ettiği verilere dayanarak en önemli ölüm nedenlerinden olan Kronik Solunum Yolu Hastalıklarının üçüncü sırada olduğunu açıklamıştır [1]. Kronik Solunum Yolu Hastalıklarında ise ilk sırayı Akciğer hastalıkları almaktadır. Akciğer hastalıklarının; hasta yatışları ve hasta ölüm nedenlerinin temel unsurlarından en önemlisi olduğu belirtilmektedir [2, 3]. Bu bilgiler doğrultusunda, solunum seslerinin tıbbi önemi açıktır ve tıp araştırmacılarının yanı sıra mühendislerin de son yıllarda daha fazla üzerinde durduğu bir alan olmuştur.

Akciğer oskültasyonu akciğer hastalıklarının gözlem, teşhis ve tedavi sürecinde klinisyenler tarafından yüz yıllardır kullanılan bir yöntemdir. Akciğer oskültasyonu, en basit non-invaziv yani cerrahi işlem gerektirmeyen, hızlı, ucuz ve kolay erişilebilir bir yöntemdir. Oskültasyon belirtilen bu avantajlarının yanı sıra yorumlayan hekimin öznelliğine bağlı kalması ile büyük bir dezavantaj oluşturmaktadır. Ancak solunum seslerini analiz etmek kolay değildir. Oldukça düzenli olan ve vücut üzerinde 4 farklı odaktan toplanan kalp seslerinden farklı olarak, anormal akciğer sesleri aralıklıdır, hem göğüs hem de sırt olmak üzere birçok farklı odaktan duyulabilir ve hastalar arasında oldukça değişkendir. Sesler soluk aldıkça değişkenlik gösterebilir bazen çok kısa süre duyulup kaybolabilir [4]. Hem zaman hem de frekans alanındaki bu yüksek değişkenlik gerçek bir hasta popülasyonunda akciğer seslerini tanımlamaya çalışan algoritmalar için büyük zorluklara neden olmaktadır [5].

Bu tez çalışmasının amacı tamamen klinik ortamda elde edilmiş veri kümelerinden sadece etiket (bir girdi kümesi için sistemin çıktılarının sınıflandırılması) bilgisine sahip olarak sınıflandırma yapabilecek bir model oluşturmaktır. Bu tez çalışmasında, solunum seslerinin, başarısı kanıtlanmış ESA modelleri (VGG16, DenseNet201, InceptionV3, ResNet101 gibi) ile sınıflandırılması gerçekleştirilmiştir. Klasik ESA mimarisine ek olarak sınıflandırma aşamasında Destek Vektör Makineleri (DVM) kullanılmasıyla başarıyı etkilediği gözlemlenmiş ve VGG16-ESA-DVM modeli ile Normal, Ral ve Ronküs seslerinin sınıflandırılmasında % 83 sınıflandırma başarısı sağlanmıştır. Özellikle teşhiste

ayırt edilmesi en zor seslerden olan Ronküs sesinin % 86 doğrulukla sınıflandırılması da literatüre katkı sunmaktadır.

1.1 Tezin Akışı

Bu tez sekiz ayrı bölümden oluşmaktadır. Tezin organizasyon yapısı şu şekilde sıralanabilir;

Bölüm-1, çalışma hakkında kısa bilgi verir ve çalışmanın temel amacını açıklar.

Bölüm-2, solunum sesleri hakkında bilgi verir, literatürdeki mevcut tekniklerin ve çalışmaların bir incelemesini sunar. Ayrıca kullanılan veri kümesi hakkında bilgi sağlar.

Bölüm-3, çalışmada kullanılan makine öğrenmesi algoritmaları için teorik bilgiler vermektedir.

Bölüm-4, solunum seslerine uygulanan ön işlemler hakkında bilgi vermektedir.

Bölüm-5, aktarmalı öğrenme yöntemleri ve mimarilerini içermektedir.

Bölüm-6, tez çalışmasında uygulanan deneyler ve sonuçları hakkında detaylı bilgi sunmaktadır.

Bölüm-7, yapılan deneylerin sonuçlarını karşılaştırılmalı bir şekilde sunmakta ve gelecekteki çalışmalar için önerilerde bulunmaktadır.

Bölüm-8, tez çalışmasından elde edilen sonucu ve literatüre katkısını özet olarak sunmaktadır.

2. SOLUNUM SESLERİ

Solunumun iki farklı anlamı vardır, bu anlamlardan ilki; oksijenin hücreler tarafından organik moleküllerin metabolizması için kullanılması (iç/ hücresel solunum) ve diğer anlamı ise; oksijen ve karbondioksitin bir organizma ile dış ortam arasında değiş tokuşudur ki buna pulmoner fizyoloji denir.

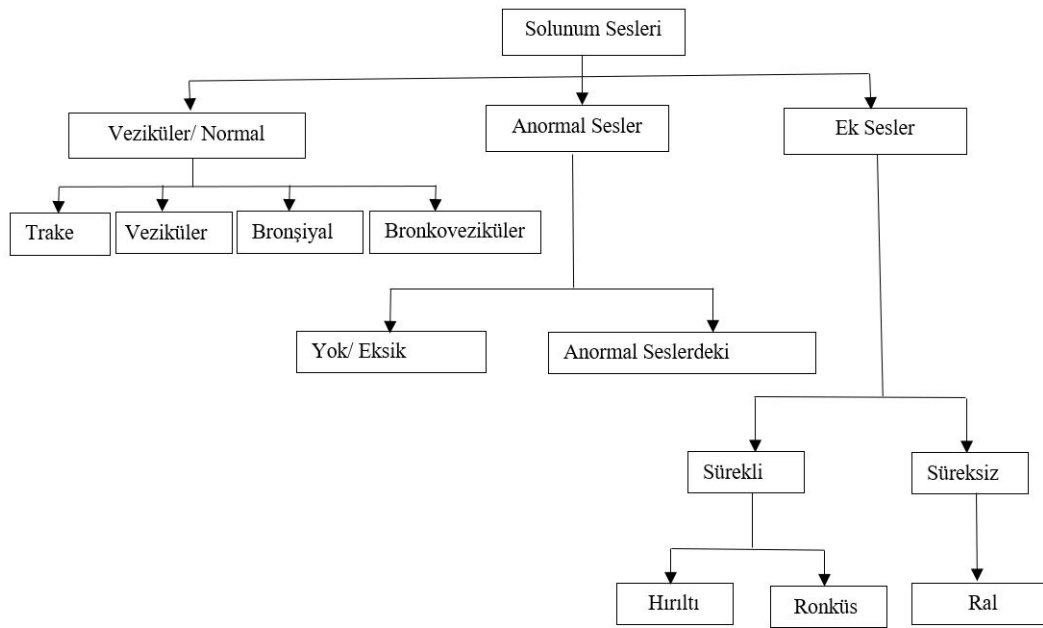
Akciğerler, vücudun boyun ile karın arasında yer alan bölümü olan toraks (göğüs) içine yerleşmiştir. Akciğerler alveol adı verilen hava keseciklerinden oluşmaktadır. Yetişkin bir bireyde 300 milyona yakın alveol bulunmaktadır. Alveollerde kan ile gaz alış veriş yapılmaktadır. Soluk alma (inspirasyon) havanın dış ortamdan soluma esnasında hava yolları üzerinden geçerek alveollere girmesidir. Soluk verme (ekspirasyon) esnasında ise bu sürecin tersi meydana gelir. Bir soluk alma ve bir soluk verme süreci solunum döngüsü olarak adlandırılır. Göğüs duvarının ön, yan ve arka yüzeylerinden stetoskop ile solunum sesleri dinlenir. Solunum sesleri geniş bir frekans aralığına sahiptir, ancak çoğu insan işitmesinin alt eşiğine yakındır [6]. İdeal durumda insan işitme alt eşiği 20Hz, üst eşiği 20kHz'dir. İnsan solunum sesleri, hareket eden hava ile toraks dokuları ile çevrili akciğerler arasındaki mekanik etkileşimden kaynaklanır ve solunum akışına, toraks geometrisine, doku akustik özelliklerine ve kayıt konumuna bağlıdır. Solunum sesleri, solunum sisteminin geometrisinin yapısı ve doğası ve durumu hakkında bilgi taşır.

Akciğer fonksiyonlarının klinik muayenesi, göğüs duvarında önceden belirlenmiş bir noktada stetoskop kullanılarak dinleme (oskültasyon) ile başlar. Hastanın göğsüne hafifçe vurarak vokal titreşimlerin göğüsten nasıl geçtiğini değerlendiren perküsyon, palpasyon ve inspeksiyon gibi başka fizik muayene yöntemleri de vardır [7]. 200 yılı aşkın bir süredir oskültasyon klinik muayenede etkin, ucuz ve kullanışlı bir yöntem olarak görülmüş ve stetoskop hekimler için profesyonel bir ekipman haline gelmiştir. Oskültasyon, solunum sağlığı durumlarını izlemek için basit, kolay ve ucuz bir yöntem olmasına rağmen, bazı sınırlamaları ve sakıncaları vardır. Sınırlamalar vücudun içsel yapısından gelir, örneğin göğüs duvarı duyulan seslerin frekans aralığını sınırlayan bir alçak geçiren filtre görevi görür. Diğer bir kısıtlılık ise ekspirasyon evresinde göğüsten duyulan seslerin sınırlı olması ve ayrıca konvansiyonel oskültasyonun sıklıkla yapılamamasıdır [8]. Klasik bir stetoskop ile oskültasyon, verilerin gelecekte referans veya daha fazla danışma için

kaydedilmesine izin vermeyebilir. Öte yandan, oskültasyonun tanı doğruluğu klinisyenin deneyimine bağlıdır, bu nedenle hekimin yeterli işitme becerisinden daha fazlasına sahip olması ve iyi eğitilmiş olması gerekir [9–11].

2.1 Solunum Sesi Türleri

Solunum sesleri farklı alt gruplara ayrılır [7,12]. Solunum seslerinin türleri Çizelge 2.1’de gösterilmiştir.



Çizelge 2.1. Solunum Seslerinin Türleri [13]

Bu çalışmada kullanılan veriler Normal, Ral ve Ronküs solunum seslerinden oluşmaktadır. Ral sesi patlayıcı ve süreksiz seslerdendir genellikle inspiratuar fazda (nefes alırken) ortaya çıkar. Akciğer seslerinde ral oluşumu, genellikle akciğer dokularında veya hava yollarında patolojik bir oluşumun göstergesidir [14]. Raller, küçük hava yollarının açılması nedeniyle meydana gelen ve genellikle 5-40 ms arasında değişen kısa süreli patlama sesleridir. Akciğerin çoğu yerinde oluşabilirler ve aynı anda bir (tek taraflı raller) veya her iki akciğerde (iki taraflı raller) bulunabilirler. Bazal raller, akciğerlerin alt kısmından kaynaklanan ralleri tanımlamak için kullanılır. Raller, sesin özelliğine göre birkaç ana türe ayrılabilir; kaba, orta ve ince, ayrıca ıslak veya kuru. İnce Raller genellikle yumuşak, yüksek perdeli ve kısa sürelidir, kaba Raller ise daha yüksek, daha düşük perdeli ve daha

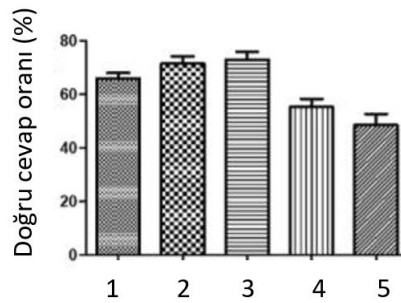
uzun sürelidir. Raller bir inspirasyon evresinde Ral'in türüne bağlı olarak geç veya erken duyulabilirler. Kaba raller erken inspiratuar fazda, ince raller ise geç inspiratuar fazda meydana gelirler. Ral sesi sağlıklı akciğerlerde de oluşabilir, ancak sürekli bir hırıltı varlığı, akciğerlerde küçük hava yollarının açıldığını ve küçük boşlukların (alveoller) sıvı, eksüda veya nefes verme sırasında havalandırma eksikliği nedeniyle çöktüğünü gösterir. Bu semptomlar genellikle pnömoni, pulmoner fibroz, akut bronşit ve diğer rahatsızlıkları olan hastalarda ortaya çıkar [15]. Ral sesleri çok ince seslerdir, bu nedenle kumaşa ve göğüs kıllarına sürtünen bir mikrofona benzer sesler üretebilir. Ronküs 100 ms'den uzun süren ve 300 Hz'den düşük frekanslı, hızlı sönümlü periyodik dalga formları içeren pes (düşük perdeli) bir sestir [14]. Hava yolu darlıkları olması durumunda Ronküs sesleri meydana gelmektedir. Priferik hava yollarından kaynaklanan Ronküslerin yüksek frekanslı, büyük hava yollarından kaynaklanan Ronküslerin ise düşük frekanslı olduğu bilinmektedir [16].

Günümüzde solunum seslerini dinlemek ve kayıt etmek mümkündür, bu da solunum verilerinin özellikleri hakkında daha ayrıntılı değerlendirmeler yapılabilmesini sağlar. Derin öğrenme gibi bilgisayar destekli sistemler, tıbbi personele yardımcı olarak teşhis kararlarının daha hızlı ve doğru şekilde verilmesine olanak sağlarlar. Solunum seslerinin işlenmesindeki zorluk, solunum seslerinin durağan olmayan sinyaller olmasıdır, bu nedenle analizleri ve birbirlerinden ayırt edilmeleri önemlidir [17]. Elektronik stetoskopların icadı sonrasında biyolojik sesleri kayıt edebilme imkanı doğmuş ve 1950'li yıllarda bu amaçla ilk çalışmalar [18] başlamış, 1960'larda ve sonrasında da devam etmiştir [15, 19]. Yıllar içinde, objektif kriterleri tanımlamak için CORSA projesi [20, 21] ve ASAP Projesi [22] tarafından standart terminoloji geliştirilmiştir. Daha fazla araştırma için kanıt temeli, sağlam çalışma tasarımları, farklı popülasyonlarla yapılan denemeler ile bilgisayarlı nefes sesi analizi kılavuzları izlenerek oluşturulmalıdır [23]. Sağlıklı kişilerde solunum sesleri hakkında sistematik bilgi eksikliğinin olduğu ve bunun sağlıklı hastaların solunum seslerinin ayırt edilmesini engelleyerek tanı ve tedaviyi etkileyebileceğini gösteren çalışmalar mevcuttur [23]. Literatürde akciğer seslerinin sınıflandırılmasıyla ilgili yapılan çalışmalarda ESA'nın kullanılması ilk olarak [24] 2016 gibi yakın tarihlerde başlanmış olup, çok yeni bir konudur. ESA'lar öznitelik çıkarılmasına ihtiyaç duymazlar, kendi özniteliklerini eğitim aşamasında kendileri çıkarırlar. Bu açıdan zaman kaybına engel olmaktadır. Ayrıca yeterli veri sayısı ile yüksek oranda başarılar elde edilmesi söz konusudur, ESA'nın

kullanılması bu anlamda önemlidir. Ancak kullanılabilecek yeterli miktarda veri kayıtlarının olmaması bu alanda çalışma yapmayı kısıtlamıştır. Son zamanlarda 126 kişiden elde edilen 920 kayıttan oluşan ICBHI 2017 veri setinin de yayınlanmasıyla solunum sesi sınıflandırmak amacıyla açık kaynak veri kümesi paylaşılmış oldu. 126 kişiden toplanan bu veriler oldukça kısıtlı sayıda verilerle çalışma sorununu ortadan kaldırmıştır. Serbes ve ark. [25], seslerin ayırt edici özelliklerini temsil eden spektral özellikler çıkararak dört sınıftan oluşan veri setini sınıflandırmak için RBF çekirdeğe sahip DVM kullanarak %57,88 doğruluk elde etmişlerdir. Yapılan bir başka çalışmada [26] solunum seslerinin spektrogram görüntüleri VGG-16 modeli (ImageNet) transfer edilerek sınıflandırma yapılmış ve bir diğer yöntemde ise VGG-16'nın çıkışı DVM'e giriş olarak verilmiştir. Bu yöntemler kullanılarak sırasıyla %63.09 ve %65.5'luk sınıflandırma başarısı elde edilmiştir. ICBHI 2017 veri setindeki ses kayıtları, sağlıklı, astım, zatürre, kalp yetmezliği, bronşektazi veya bronşit ve kronik obstrüktif akciğer hastalığı (KOAH) dahil olmak üzere etiketler içermektedir. Fraiwan ve ark. [27] yaptıkları çalışma ile bu hastalıkları sınıflandırmayı amaçlamışlardır. Ses sinyallerinin özellikleri Shannon entropisi, logaritmik enerji entropisi ve spektrogram tabanlı spektral entropi kullanarak elde edilmiştir. Karar ağaçları ve ayırt edici sınıflandırıcılar kullanılan çalışmada DVM sınıflandırıcı ile, ortalama %98.20 sınıflandırma doğruluğu sağlanmıştır. Ayrıca, artırılmış karar ağaçlarının hastalık tahmininde %98.27 ile daha yüksek doğruluğa ulaştığı gözlemlenmiştir. Yapılan çalışma ile hastalık teşhisinde de makine öğrenmesinin gösterdiği başarı açıktır.

Bu çalışmada literatürden farklı olarak Normal, Ral ve Ronküs solunum sesleri olmak üzere üç farklı ses sınıflandırılması yapılmaktadır. Literatürde bu üç grup ses için sınıflandırma yapan bir çalışmada [28] solunum seslerinin spektrogram görüntüleri elde edilerek sınıflandırma işlemi ESA mimarisi ile gerçekleştirilmiştir, ve elde edilen sınıflandırma doğruluğu bu üç ses için %76 olmuştur. Bu üç ses grubuna çalışmamızda yer verilmesinin nedenlerinden önce Hırıltı sesine yer verilmemesinin nedenini ifade etmek gerekirse; Hırıltı sesi stetoskop kullanılmadan insan kulağıyla dışarıdan duyulabilen bir ses olduğundan [11,29] sınıflandırma için veri kümesi oluşturulurken bu ses grubuna yer verilmemiştir. Tez çalışmasında Ronküs sesine özellikle yer verilmesinin nedenini anlamak içinse Wilkins ve arkadaşlarının [30] solunum terminolojisiyle ilgili yaptıkları araştırma çok çarpıcıdır. Öyle ki kayıtlı sekiz akciğer sesi örneğini tanımlamak için 223 göğüs hekimi ve 54 diğer uzmanlık alanlarındaki hekimin bu terminolojiye hakimiyetini rapor

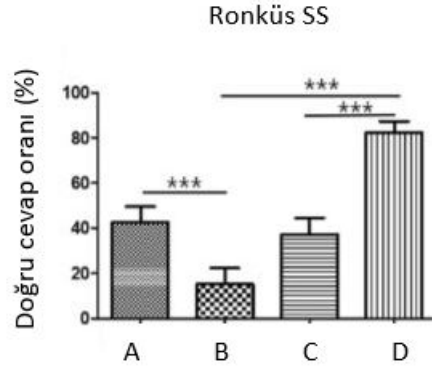
etmişlerdir. Katılımcılar akciğer seslerini dinlemiş ve çoğu normal nefes seslerini tanımış ancak abnormal solunum seslerini tanımlamak için nitelikli bir sıfat kullanmamışlardır, kullananlar arasında ise çok az uyuşma olmuştur. Ronküs sesine ait bir kayıt hekimlere dinletilmiş ve %50'sinden fazlası bu sesi yanlış tarif etmiştir. Bu çalışma solunum seslerinin doğru tespitinin zorluğunun ve öneminin bir ölçütüdür. Yapılan son çalışmalar da bunu desteklemektedir. Klinik uygulamalarda solunum seslerinin sınıflandırılmasında özellikle Ronküs sesinin ayırt edilmesi güçtür. Bunun en güncel örneği 2021 yılında yapılan çalışmada açıkça ortaya konulmuştur [31]. Yapılan çalışmada, gerçek klinik uygulamada oskültasyon analizinin doğruluğu Şekil 2.1'de gösterilmiştir.



Şekil 2.1. Klinik Pratikte Oskültasyon Analizinin Doğruluğu [31]

1: Tümü (overall), 2: Normal, 3: Ral, 4: Hırıltı, 5: Ronküs

Çalışmalarında [31] genel olarak dört grup solunum sesi (Normal, Ral, Hırıltı, Ronküs) Tıp Fakültesi öğrencileri, stajyerler, asistanlar ve akademisyenlerine dinletilmiş ve sesleri sınıflandırmaları istenmiştir. Yapılan deney sonucunda bu dört ses grubu içerisinde Ronküs sesinin ayırt edilmesi en zor ses olduğu Şekil 2.1'deki gibi gözlemlenmiştir. Öğrenciler, stajyerler, asistanlar ve akademisyenlerin Ronküs sesini tespit etme yüzdeleri de Şekil 2.2'de gösterilmiştir.



Şekil 2.2. Klinik Pratikte Oskültasyon Analizi ile Ronküs Doğruluğu [31]

A:Öğrenci , B: Stajyer (Intern), C: Asistan (Resident) , D: Akademisyen(Fellow)

Şekil 2.2’den de görüldüğü gibi yapılan deneylerde öğrenci, stajyer ve asistan gruplarının Ronküs sesini % 60’ın altında doğrulukla tespit edebildiği görülmektedir. Yapılan çalışmada 3 sınıf (Ral, Hırıltı ve Ronküs) için %85.7, 2 sınıf (normal/ abnormal) için ise %86.5 sınıflandırma doğruluğu elde edilmiştir. Aslında veri kümelerinde normal sesi olmasına rağmen 3 sınıf (Normal, Ral, Ronküs) için bir sınıflandırma yapmamışlardır.

Solunum seslerini sınıflandırma ile ilgili bir çalışma da John Hopkins Üniversitesi tarafından yapılmıştır [32]. Yapılan çalışmada zatürre teşhisi koymada %87 başarı sağlayan (normal-zatürre sınıflandırması) bir makine öğrenmesi algoritması geliştirdiklerini ve algoritma üzerinde çalışmaya devam ederek gerekli testler tamamlandıktan sonra geliştirdikleri stetoskopu kullanıma sunmayı hedeflediklerini belirtmişlerdir. Önerdikleri algoritmayı beş Afrika ülkesinden ve iki Asya ülkesinden yaklaşık 1500 hastalık bir veri kümesi ile eğitmişlerdir. Yaptıkları çalışmayla yüzbinlerce hayatın kurtulmasına katkı sağlanabileceğini belirtmişlerdir.

Pulmoner semptomlarla başvuran hastalarda solunum sesleri tamamen normal olabileceği gibi; hava yolu hastalığı olanlarda ronküsler, akciğer parankim hastalığı olanlarda raller en sık işitilen seslerdir. Tüm bu nedenlerle temel olarak bu üç ses türünün (Normal, Ral, Ronküs) ayırt edilmesi önemlidir ve bu tez çalışmasında bu üç ses grubuna yer verilmiştir. Solunum sinyalleri, Littmann 3M elektronik stetoskop ile toplanmıştır. Solunum verileri, 15 saniyelik sabit bir süre boyunca toplanmış ve kayıpsız bir WAV formatında saklanmıştır. Sinyal 4 kHz frekansta sayısallaştırılmıştır. Böylece her hasta için 60.000 örnekten

oluşan bir veri kümesi elde edilmiştir. Veri kümesi 294 adet hastaya ait 294 adet nefes alış ve nefes verişlerinden oluşan solunum döngülerine ait ses kayıtlarından oluşmaktadır. Bu tez çalışmasında kullanılan veriler Hacettepe Üniversitesi Etik kurul onayı (Karar No: GO 18/631 -45) alınarak Hacettepe Üniversitesi Göğüs Hastalıkları Ana Bilim Dalı'nda görev yapan Prof.Dr. Deniz Köksal, Doç.Dr. Sevinç Sarıncı Ulaşlı, Doç. Dr. Elif Babaoğlu tarafından toplanmıştır.

2.2 Solunum Seslerini Sınıflandırmanın Önemi

Normal ile anormal solunum sesleri arasındaki ayrım, kesin bir tıbbi tanı için önemlidir. Solunum sesleri, akciğer ve solunum yolu tıkanıklığının fizyolojisi ve patolojileri hakkında paha biçilmez bilgiler içerir. Seslerin spektral yoğunluğu ve genliği, akciğer parankiminin durumunu, hava yollarının boyutunu ve patolojik değişikliklerini gösterebilir [33].

Geleneksel stetoskop oskültasyonu subjektif ve zorlukla paylaşılabılırken, yeni sistemler sonuçların tekrarlanabilirliği ve verilerin paylaşılabilirliği gibi özellikler ile objektif ve erken teşhise yardımcıdırlar. Ayrıca, tanı koyma, izleme ve internet üzerinden veri alışverişi gibi uygulamalar objektif ve otomatik oskültasyon ses analizi için tamamlayıcı araçlardır. Aynı zamanda az gelişmiş ülkeler ve uzak topluluklar için de faydalı bir çözüm olabilir [34]. Ek olarak, bu tür sistemler, oskültasyonun noninvazif olması ve düşük maliyetli olması gibi özellikleri nedeniyle büyük avantaja sahiptir. Akciğerin ses repertuarı, bir stetoskopla duyulduğunda gerçekten sınırlı olabilir, ancak dijital olarak analiz edildiğinde çok daha geniş bir bilgi içeriği yelpazesi sergilediği açıkça görülür. Bilgisayar analizi insan kulağının kabiliyetlerinin ötesine geçmektedir. Solunum sesleri, solunum döngüsü sırasında akciğer hacminin değişmesi ve kalp sesi nedeniyle durağan olmayan (non-stationary) stokastik sinyallerdir. Bu, solunum seslerinin analizini zor bir görev haline getirir.

Görsel veya işitsel bilgilerle ilgili sağlık problemini doğru bir şekilde tanımlayabilmek tecrübe ve zaman gerektirir, DSÖ istatistiklerine göre, DSÖ üye Devletleri'nin %45'i, her 1000 nüfus için 1'den az hekime sahip olduğunu bildirmektedir [35]. Tüm bunlar hekimlerin iş yükünü azaltacak, hekimlere teşhiste yardımcı olabilecek, hatta hekimlerin olmadığı bölgelerde veya tıp eğitiminde kullanılabilecektir. Bu da, solunum sesi tanıma

sistemlerine olan ihtiyacı göstermektedir. Evrişimsel Sinir Ağları (ESA) 1980’lerden beri önerilmektedir ve son yıllarda çeşitli nesne sınıflandırma görevlerinde kullanılmak üzere etkili bir yöntem olarak kabul edilmiştir [36]. Solunum sesi tanıma sistemlerine olan ihtiyaç ve ESA’daki başarılar dikkate alındığında solunum seslerinin ESA ile sınıflandırılması fikri oluşmuştur. Tez kapsamındaki ESA ile ilintili matematiksel bağıntılara geçmeden önce, genel olarak yapay zeka ve makine öğrenimi konularından bir sonraki bölümde bahsedilecektir.



3. MAKİNE ÖĞRENMESİ

Yapay Zeka (AI), bilgisayarların okuma, görme, karar verme gibi insan görevlerini yapabilmesi üzerine çalışan bir bilim dalıdır. Makine öğrenmesi ise makineye bir grup veri verilmesiyle makinenin o verilerle öğrenmesi anlamına gelip, yapay zekanın bir alt kümesidir. Makine öğrenmesi makineler nasıl davranacakları öğretilmeden yapılmaktadır. Sınıflandırma problemi de bir makine öğrenmesi problemidir. Makine öğrenimi, uygun maliyetlidir, ölçeklenebilirdir ve sınıflandırma görevlerinde neredeyse uzman düzeyinde kesinlik elde etme potansiyeline sahiptir [37]. Bilgisayar tabanlı sınıflandırma, gözetimli (supervised) ve gözetimsiz (unsupervised) öğrenme olmak üzere iki temel strateji kullanılarak yapılır [38].

Gözetimli öğrenme, bilinen veri kümelerinin sisteme uygulanması anlamına gelir. Bu sonuçların gözetmen tarafından doğrulandığı anlamına gelmez, girdilerin bir çıktıyla doğru eşlendiği anlamına gelir. Sınıflandırıcı ilk önce doğru sınıfların bulunduğu verileri kullanarak eğitilir [38]. Bir girdi kümesi için sistemin çıktılarını sınıflandırma işlemine etiketleme denir.

$$Y = f(X) \quad (3.1)$$

Y çıktı değişkeni yani insan kaynaklı etiketlerdir, X ise bu çıktı değişkenleri ile eşlemek için kullanılacak girdi değerleridir. Eşleme sonucunda doğruluk yeterince iyi olduğunda, öğrenme aşaması durdurulabilir ve yeni veriler üzerinde tahminler yapılabilir [39]. Bir makine öğrenmesi uygulaması yapılırken ilk olarak, sorunun çerçevelendirilmesi gerekir: gözetimli mi yoksa gözetimsiz öğrenme yöntemi mi kullanılacak karar vermek gerekir [40]. Bu çalışmada kullanılan veriler Bölüm 2 ve Bölüm 5'te anlatıldığı üzere, hekimler tarafından etiketleme işlemlerinden geçtiği için yapılan sınıflandırma işlemi gözetimli öğrenmedir.

Gözetimsiz öğrenmede ise, sınıflandırıcı, girdi verilerinin kendi içindeki kalıpları bulmaya çalışır: sınıflandırıcının veri kalıpları hakkında ön bilgisi yoktur ve belki de mevcut sınıfların sayısı bile yoktur [38]. Eşitlik 3.1'de görüldüğü gibi eşleme çıktısı olmayan verilerdeki örüntüyü tespit etmek için gözetimsiz öğrenme kullanılır.

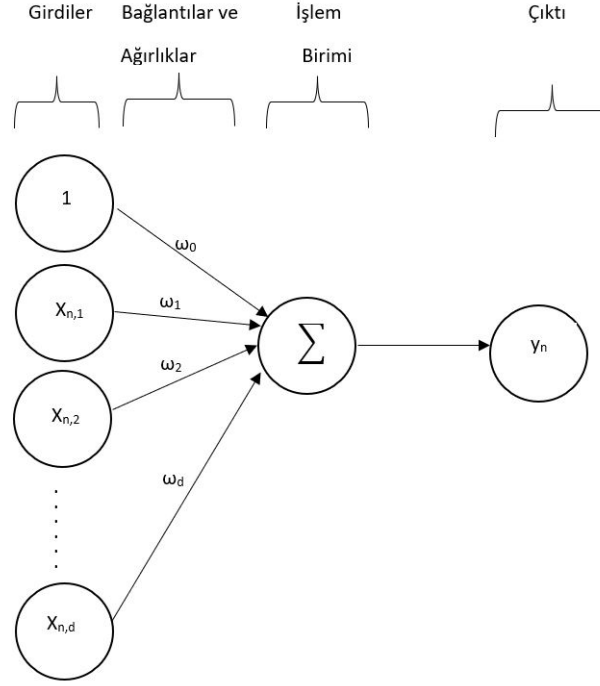
3.1 Derin Öğrenme

Derin öğrenme, özneliliklerin çıkarılması ve dönüştürülmesi için birçok doğrusal olmayan bilgi işlem katmanından yararlanan bir makine öğrenmesi türüdür. Ard arda gelen her katman, önceki katmandaki çıktıyı girdi olarak alır [41]. Ayrıca derin öğrenmenin matematiksel eşitlikler kullanarak insan beyninin işleyişini kopyaladığı da söylenebilir.

3.1.1 Algılayıcılar

Algılayıcılar (perceptron), YSA'nın temel işlem birimleridirler. İnsan karar verme ve düşünce yetisini modellemek araştırmacıların ilgisini çeken bir konudur. Bu amaçla yapılan ilk çalışma McCulloch-Pitts modeliyle 1943 yılında McCulloch ve Pitts tarafından gerçekleştirilmiştir. McCulloch-Pitts modeli yapay sinir ağlarının (YSA) temelini teşkil etmektedir. 1958'de Rosenblatt'ın Perceptron ardından Widrow ve Hoff'un 1960'da Adaptive Linear Element (ADALINE) kavramlarını ortaya atmasıyla çalışmalar hız kazanmıştır.

Tek katmanlı algılayıcılar, yalnızca girdi ve çıktı katmanına sahiptir. Yalnızca tek katmana sahiptir, bu nedenle tek katmanlı algılayıcı adı verilir. Çok katmanlı algılayıcıdaki gibi gizli katmanlar içermez.



Şekil 3.1. Algılayıcılar

Yukarıdaki şekilde x_n 'ler algılayıcının girişleri, w_j 'ler algılayıcının ağırlıkları ve y_n ise çıkışı temsil etmektedir. Algılayıcının çıkışı Eşitlik 3.2'deki gibi hesaplanır.

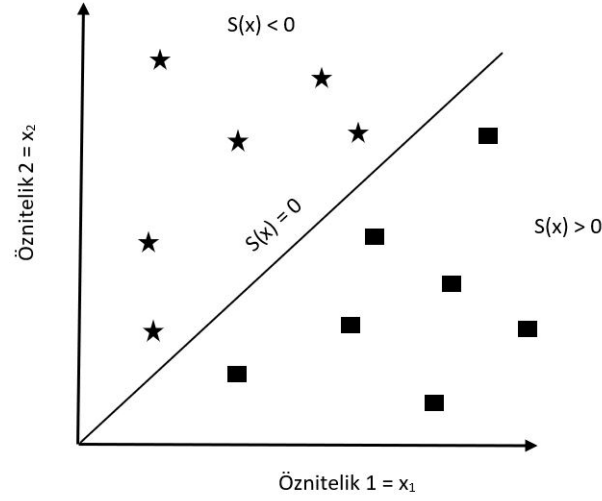
$$y_n = w_0 + \sum_{j=1}^D x_{n,j} w_j \quad (3.2)$$

ESA'lar, çok katmanlı perceptron modelinin basit bir uzantısıdır. Öncelikle görsel tanıma alanlarında kullanılmasına rağmen, evrimsel mimariler konuşma algılama ve müzik analizi gibi alanlarda da uygulanmıştır. Algılayıcıların genel yapısı Şekil 3.1'de gösterilmiştir.

3.2 Sınıflandırıcılar

3.2.1 Lineer Sınıflandırıcılar

Şekil 3.2, iki farklı sınıfın birbirinden $S(x) = 0$ karar sınırı ile ayrılmasını göstermektedir. Karar sınırının oluşturulması önemlidir. Karar sınırı oluşturulurken Eşitlik 3.3'ten



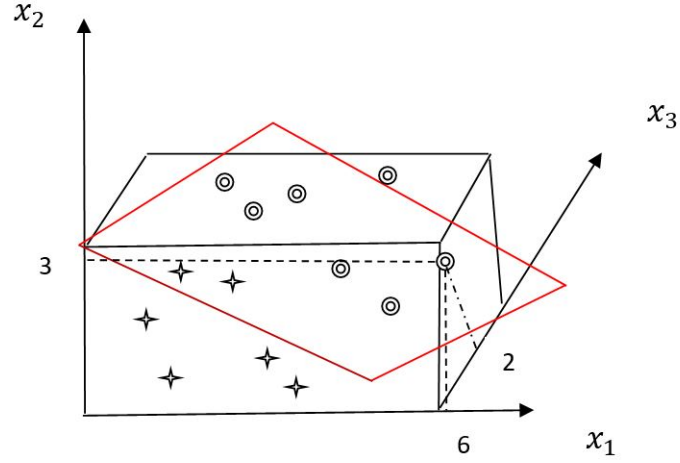
Şekil 3.2. Lineer Sınıflandırıcı Karar Sınırı

yararlanılır.

$$S(x) = f(\mathbf{w} \cdot \mathbf{x}) = f\left(\sum_j w_j x_j\right) \quad (3.3)$$

Lineer sınıflandırıcıda çıkış, giriş değerlerinin ağırlıklı toplamlarından oluşur. Karar sınırı pozitif ve negatif tahmin sonuçlarını birbirinden ayırır. Şekil 3.2’de görüldüğü gibi sınıflandırıcı iki katsayıdan oluşuyorsa karar sınırı çizgi şeklindedir. Üç katsayıdan oluşuyorsa düzlem, dört katsayı ise hiper düzlem şeklindedir.

Basit bir hiper düzlem ve sınıflandırıcının katsayıları örneği Şekil 3.3’te verilmiştir.



Şekil 3.3. Sınıflandırıcı Katsayıları

Şekil 3.3'te yıldız ve içi boş daireler iki farklı sınıfı, eksenler ise öznitelikleri temsil etmektedir. Hiper düzlem ise kırmızı renkle çizilmiştir. Bu bilgilere bağlı olarak uygun hiper düzlemi oluşturmak için bir skor değeri hesaplanır böylece model oluşturulmuş olur.

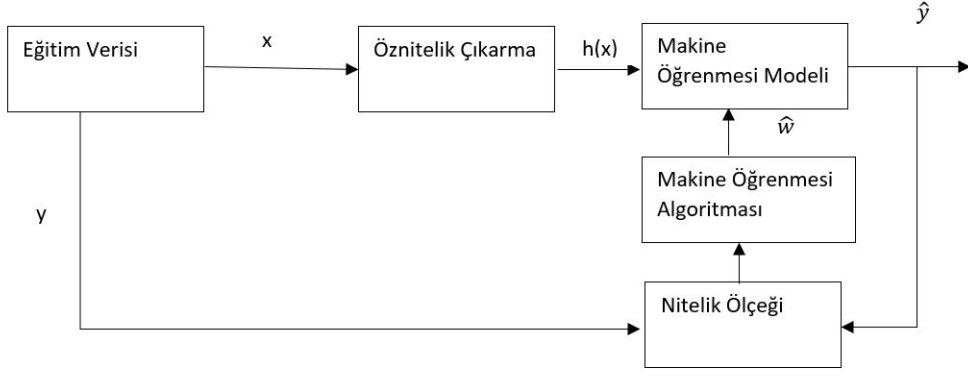
$$\hat{y}_i = \text{sign}(\text{score}(x_i))$$

$$\begin{aligned} \text{score}(x_i) &= w_0 + w_1 x_i[1] + \dots + w_d x_i[d] \\ &= w^T x_i \end{aligned} \quad (3.4)$$

Eşitlik 3.4'te basit 2 boyutlu bir hiper düzlem için oluşturulan model ve skor denklemi verilmiştir. Şekil 3.3'te belirtilen örnek değerler için skor değeri:

$$\text{score}(x) = w_0 + 6w_1 + 3w_2 + 2w_3 \quad (3.5)$$

şeklinde hesaplanır. Lineer bir sınıflandırıcı modeli Şekil 3.4'te görüldüğü gibidir.



Şekil 3.4. Lineer Sınıflandırıcı Modeli

Şekil 3.4’te görüldüğü gibi eğitim verileri makine öğrenmesi modeline gelerek bir tahminde bulunur, bu tahmin $\hat{y}$ ’dır. Makine öğrenmesi algoritması ile tahmin ($\hat{y}$) ve gerçek değer (y) kıyaslanarak ağırlıklar $\hat{w}$ güncellenir. Bu işlemler nitelik ölçeği ile belirlenen ana kadar devam ederek optimum tahmin gerçekleştirilmeye çalışılır.

3.2.2 Destek Vektör Makineleri

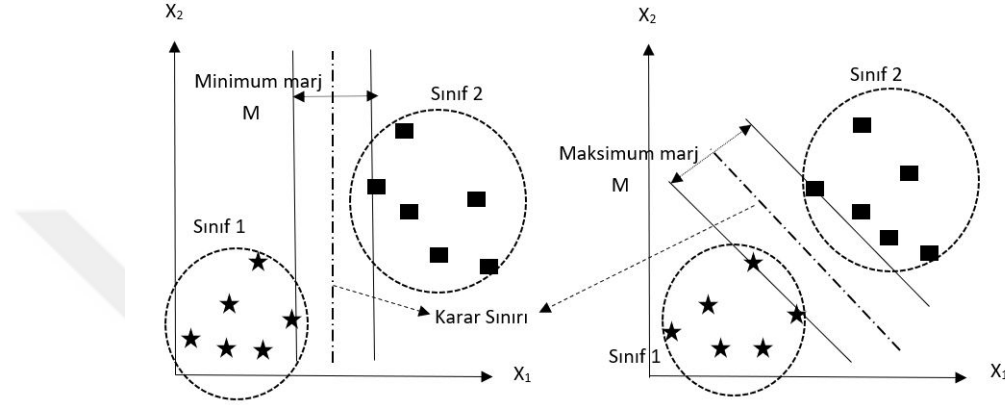
Destek Vektör Makineleri (DVM), endüstride ve bilimde yaygın olarak kullanılan ve genellikle rakip yöntemlerden daha iyi sonuçlar sağlayan temel bir makine öğrenimi aracıdır. Rastgele orman algoritmasıyla birlikte, son yıllarda makine öğreniminin temel taşları olmuşlardır. DVM ve rastgele orman, en iyi sınıflandırma puanlarının gerekli olduğu uygulamalar için sıklıkla kullanılan algoritmalarıdır [42]. Vapnik ve Chervonenkis’in DVM algoritması [43], verileri farklı kümelerle bölmek için hiper düzlemlerin optimize edildiği 1963’teki istatistiksel öğrenme literatüründen geliştirilmiştir. Yaklaşık otuz yıl sonra, Boser, Guyon ve Vapnik, çekirdek hilesini maksimum marjlı hiper düzlemlere uygulayarak doğrusal olmayan sınıflandırıcılar üretmişlerdir [42]. Mevcut standart, 1990’ların ortalarında Cortes ve Vapnik tarafından önerilmiştir [44]. 2000’li yıllara gelince DVM konuşmacı tanıma ve ses tanıma gibi problemlerin çözümünde de kullanılmaya başlanmıştır.

DVM gözetimli öğrenme yöntemlerindendir. DVM’lerde yüksek boyutlu girdi vektörü X ile skaler/vektör çıktısı y arasındaki ilişki Eşitlik 3.1’de görüldüğü gibidir. DVM’ler, optimizasyon ile marjı maksimize eden bir hiper düzlem belirleyerek iki sınıfı ayırmayı

amaçlarlar. İkili sınıflandırma problemi düşünülürse, eğitim verileri de şu şekilde ifade edilir:

$$(x_1, y_1), (x_2, y_2), (x_3, y_3), \dots, (x_n, y_n), x \in \mathbb{R}^2, y \in \{+1, -1\} \quad (3.6)$$

$x \in \mathbb{R}^2$ olmak üzere iki boyutlu giriş uzayında basitçe Şekil 3.5'teki gibi gösterilebilir.



Şekil 3.5. Veriyi Hiper Düzlem ile Sınıflandırma

Şekil 3.5'te de görüldüğü gibi veriler lineer olarak ayrılabilir ve farklı hiper düzlemler de bulunabilir. Şekil 3.5'teki hiper düzlem şöyle tanımlanabilir:

$$\omega_1 x_1 + \omega_2 x_2 + b = 0 \quad (3.7)$$

En iyi hiper düzlemi nasıl bulacağımız önemlidir. $\omega = [\omega_1, \omega_2, \dots, \omega_n]^T$ ağırlıklar olmak üzere ayırt edici veya karar fonksiyonu $d(x, \omega, b)$ aşağıdaki şekildedir.

$$d(x, \omega, b) = \omega^T x + b = \sum_{i=1}^n \omega_i x_i + b, \text{ burada } x, \omega \in \mathbb{R}^d \quad (3.8)$$

Şekil 3.5'teki karar sınırı veya ayrıştırıcı çizgi için $d(x, \omega, b) = 0$ 'dır. Eğitim aşamasından sonra, ağırlıklar kullanılarak makine, aşağıdaki şekilde verilen bir gösterge (indicator) fonksiyonuna göre, x_p 'ler üzerinde çıktı üretir.

$$i_F = o = \text{sign}(d(x_p, \omega, b)) \quad (3.9)$$

Yukarıdaki denklemde i_F gösterge fonksiyonudur ve o (output), makine öğrenmesinin çıktısıdır ve karar kuralı (decision rule) aşağıdaki gibidir:

Eğer $d(x_p, \omega, b) > 0$ ise, örüntü x_p 1 sınıfına aittir, $o = y_1 = +1$

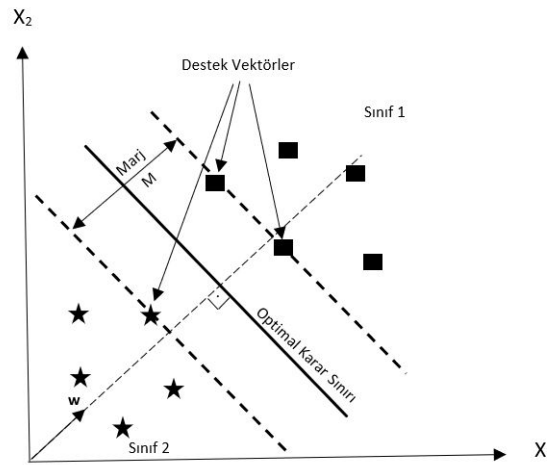
Eğer $d(x_p, \omega, b) < 0$ ise, örüntü x_p 2 sınıfına aittir, $o = y_2 = -1$

Bir hiperdüzlemin uzaklığının $\frac{1}{\|\omega\|}$ olduğu bilinerek, iki toplam marj şu şekilde ifade edilebilir:

$$M = \frac{2}{\|\omega\|} \quad (3.10)$$

Marjın maksimum olabilmesi için $\|\omega\|$ ağırlık vektörünün minimum olması gerekir. Minimizasyon problemi Eşitlik 3.11'deki gibi olur.

$$\omega^T \omega = \sum_{i=1}^n w_i^2 = \omega_i^2 = \omega_1^2 + \omega_2^2 + \dots + \omega_n^2 = \text{minimize } \frac{1}{2} \omega^T \omega \quad (3.11)$$



Şekil 3.6. Hiper Düzlemi Ayıran Maksimum Marjın Gösterimi

Şekil 3.6 optimal marjı M olan ve standart hiperdüzlem olarak da adlandırılan bir ayırma hiperdüzlemini göstermektedir. Kesik çizgiler iki farklı sınıfla kesişir ve bunlara destek vektörleri denir. Optimal hiperdüzlemler aşağıdaki koşulu sağlarlar:

$$y_i [\omega^T x_i + b] \geq 1, i = 1, l \quad (3.12)$$

burada l , veri noktalarının sayısını gösterir. Bu nedenle, optimal ayırma hiperdüzlemini bulmak için, $\|\omega\|^2$ 'nin Eşitlik 3.12'deki eşitsizlik kısıtlamalarına bağlı olarak minimize

edilmesi gerekir. Bu, Lagrange fonksiyonunun eyer noktası (saddle point) tarafından çözülen eşitsizlik kısıtlamaları ile klasik bir ikinci dereceden optimizasyon problemiyle sonuçlanır.

$$L(\omega, b, \alpha) = \frac{1}{2}\omega^T\omega - \sum_{i=1}^l \alpha_i \{y_i [\omega^T x_i + b] - 1\} \quad (3.13)$$

Eşitlik 3.13'te α_i 'ler Lagrange çarpanlarıdır. Lagrange L 'nin ω ve b 'ye göre minimize edilmesi için optimal eyer noktası gereklidir. Bu, kısıtlı (constrained) bir fonksiyonun optimumu için Karush-Kuhn-Tucker (KKT) koşulları uygulanarak yapılır. Bu durumda Eşitlik 3.12 dışbükeydir (convex) ve Eşitlik 3.13 için KKT koşulları gerekli ve yeterlidir [45].

3.2.3 Lojistik Regresyon

En yaygın sınıflandırma yöntemlerinden biri Lojistik Regresyon'dur. İkili bir sınıflandırma yöntemi olmasına rağmen, Bire Karşı Diğerleri (One Versus Rest) yöntemi kullanılarak çok sınıflı sınıflandırma da gerçekleştirilebilir. Öncelikle net girdi, ağırlıkların ve örnek özelliklerin lineer kombinasyonu ile hesaplanır. Daha sonra hesaplanan bu değer, aktivasyon fonksiyonu tarafından değerlendirilir. Aktivasyon fonksiyonunun çıktısı, sınıf etiketlerini belirlemek için kullanılır. Maliyet fonksiyonunun minimize edilmesi ağırlıkların daha doğru belirlenmesine yardımcı olur. Lojistik fonksiyon Eşitlik 3.14'te belirtilmiştir.

$$p(x) = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}} \quad (3.14)$$

Eşitlik 3.14'te $p(x)$ ifadesi $E(Y|x)$ 'in yerine kullanılmıştır. Çıkış aralığı, lojistik regresyon için 0 ile 1 arasında tanımlanmıştır. $E(Y|x)$, verilen x değeri için Y 'nin beklenen değeri olarak okunur. β_0 bir sapmadır ve β_1 tek giriş değeri x 'in katsayısıdır. Eşitlikte α_i Lagrange çarpanıdır. Eşitlik 3.14'ün düzenlenmesiyle Eşitlik 3.15 elde edilir.

$$p(x) = \ln \left[\frac{\pi(x)}{1 - \pi(x)} \right] = \beta_0 + \beta_1 x \quad (3.15)$$

Eşitlik 3.15, lojistik regresyon modelinin parametresine göre doğrusal bir model olduğunu

kanıtlar ($y = ax + b$). Daha ayrıntılı olarak, lojistik regresyon parametreleri doğrusal bir kombinasyondur ve yapılan dönüşüm logit dönüşümü olarak adlandırılır [39].

3.3 EVRİŞİMSEL SİNİR AĞLARI

Evrışimsel Sinir Ağları (ESA), çok katmanlı algılayıcıların (multi layer perceptron) özel bir halidir. Tam bağlantılı katmanda (fully connected layer) matris çarpımı gerçekleştirilirken evrişim katmanında doğrusal bir matematik operatörü olan evrişim işlemi gerçekleştirilir. Evrişim işleminin sürekli zamandaki ifadesi Eşitlik 3.16'da verilmiştir.

$$y(t) = u(t) * x(t) = \int u(\tau)x(t - \tau)d\tau \quad (3.16)$$

Evrışim işleminin ayrık zamandaki ifadesi ise Eşitlik ??'deki gibidir.

$$y[k] = \sum_{j=-\infty}^{\infty} u[j]x[k - j] \quad (3.17)$$

3.3.1 Ağın Eğitilmesinde Kullanılan Parametreler

3.3.1.1 Maliyet Fonksiyonları

Modelin eğitimi süresince kullanılacak ağırlıkların manuel olarak tek tek denenmesi için gereken maliyet ve zaman kaybını en aza indirmek için bu ağırlıkların eğitim süresince en iyisinin bulunması için kullanılan fonksiyonlara maliyet fonksiyonları denilmektedir. Eğitim süresince yaygın olarak kullanılan maliyet fonksiyonlarından karesel maliyet fonksiyonu (quadratic cost function) aşağıdaki denklemde gösterilmektedir.

$$c(w, b) = \frac{1}{n} \sum_x \|y(x) - a\|^2 \quad (3.18)$$

Bilindiği üzere denklemde ağırlıklar (weight) w , bias değerleri b , x inci eğitim örneği için gerçek ağ çıktısı a ve ağın çıktısı y simgesiyle ifade edilmiştir. Tahmin edilen $y(x)$ değeri gerçek değer olan a 'ya ne kadar yakınsa tahmin o kadar doğrudur ve bu da maliyetin de

bir o kadar az olduğu anlamına gelmektedir. Bir diğer maliyet fonksiyonu olan Çapraz Entropi Fonksiyonu [2] ise aşağıdaki gibidir:

$$C = -\frac{1}{n} \sum_x [y \ln a + (1 - y) \ln(1 - a)] \quad (3.19)$$

n ; eğitim verisindeki elemanların toplamını, x ; eğitim verisindeki girdilerin toplamını, y ise çıktıyı ifade etmektedir. Nöron, x girdi değerleri için hedeflenen y değerine yaklaştıkça çapraz-entropi sonucu sıfıra yaklaşır. Çapraz-entropi fonksiyonu çıkışında her zaman pozitif değerler elde edilir. Maliyet fonksiyonlarından karesel maliyet, öğrenme yavaşlığına yol açtığı için çapraz entropi tercih edilmektedir [46].

3.3.2 Aktivasyon Fonksiyonları

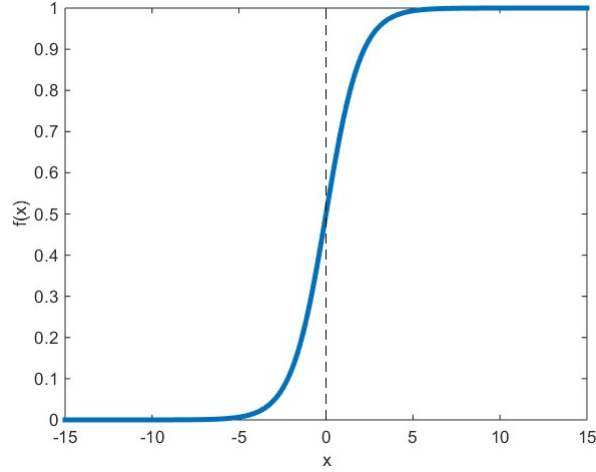
Sinir ağındaki her bir nöron düğümü, bir önceki katmanın nöronunun çıkış değerini girdi değeri olarak kabul eder ve girdi değerini bir sonraki katmana iletir. Girdi katmanına ait nöron düğümü, girdi öznitelik değerini doğrudan bir sonraki katmana geçirecektir. Çok katmanlı bir sinir ağında, üst düğümün çıkışı ile alt düğümün girdisi arasında işlevsel bir ilişki vardır. Bu fonksiyona aktivasyon fonksiyonu denir. Özetle aktivasyon fonksiyonu, sinir ağının sonuna veya arasına yerleştirilerek nöronun ateşlenip ateşlenmeyeceğine karar vermeye yardımcı olan bir düğümdür.

3.3.2.1 Sigmoid Fonksiyonu

Sigmoid fonksiyonu, fonksiyon çıkışını 0 ile 1 arasına sıkıştırır [0,1]. 0 odaklı değildir, 0'dan uzaklaştıkça fonksiyon çıkışının doyuma ulaşması da eğimin sıfırlanması problemine yol açmaktadır. Bu gibi olumsuz nedenlerden dolayı fazla tercih edilmemektedir. Sigmoid fonksiyonun matematiksel ifadesi Eşitlik 3.20'deki gibi yazılır.

$$f(x) = \frac{1}{1 + e^{-x}} \quad (3.20)$$

Sigmoid fonksiyonunun grafiği Şekil 3.7'deki gibidir.



Şekil 3.7. Sigmoid Fonksiyonu

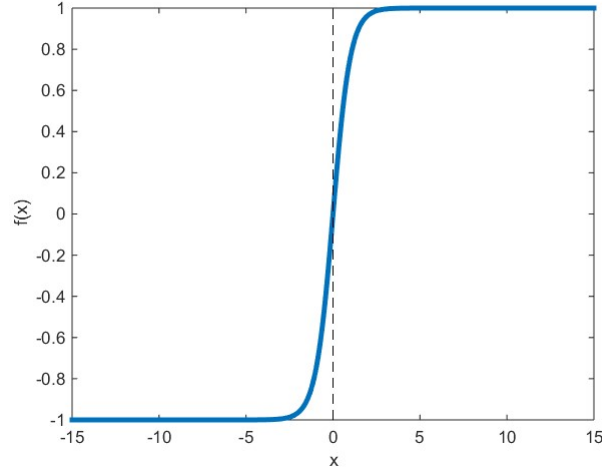
Değişken büyük bir pozitif değer veya küçük bir negatif değer aldığıda doymunluk meydana gelir ve artık girdi verilerindeki küçük değişikliklere duyarlı değildir. Sigmoid fonksiyonunun giriş değeri arttıkça çıkış üst eşik değeri olan 1'e yaklaşır, ve bir noktadan sonra çıkış sabit kalır. Geri yayılımda bu kaybolan gradyan problemine sebep olur.

3.3.2.2 Tanh Fonksiyonu

Fonksiyon çıkışını -1 ile 1 arasına sıkıştırır $[-1,1]$. 0 odaklıdır, Sigmoid fonksiyonunda olduğu gibi eğim sıfırlanabilmektedir. 0 odaklı olmasından ötürü sigmoid fonksiyonuna oranla daha sık tercih edilmektedir. Tanh fonksiyonun matematiksel ifadesi Eşitlik 3.21'deki gibi yazılır.

$$f(x) = \frac{2}{1 + e^{-2x}} - 1 \quad (3.21)$$

Tanh fonksiyonunun grafiği Şekil 3.8'deki gibidir.



Şekil 3.8. Tanh Fonksiyonu

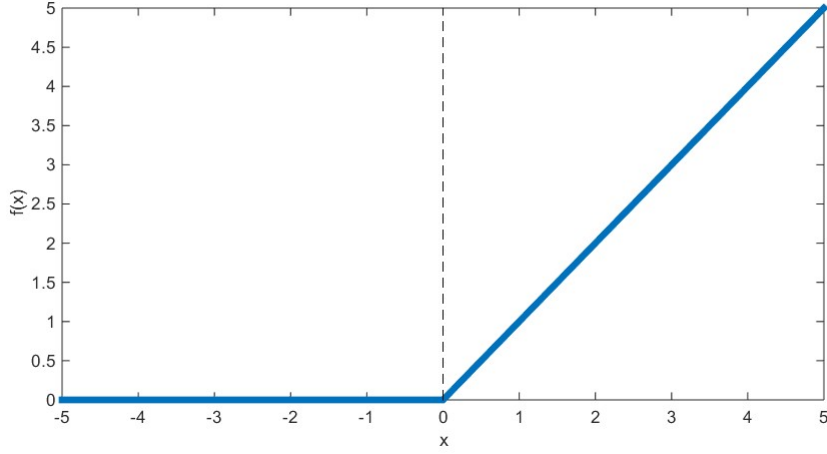
Tanh fonksiyonunun grafiği sigmoid fonksiyonunun grafiğine benzer, burada da değişken büyük bir pozitif değer veya küçük bir negatif değer aldığı anda doygunluk meydana gelir.

3.3.2.3 ReLU Fonksiyonu

ReLU fonksiyonunda sıfırdan büyük değerlerde eğim her zaman sıfırdan farklıdır. Ancak negatif değerler için fonksiyon çıkışının sıfır olması ve eğimin de sıfır olması olumsuz özelliğidir. ReLU fonksiyonun matematiksel ifadesi 3.22'deki gibi yazılır.

$$f(x) = \begin{cases} 0 & \text{for } x < 0 \\ x & \text{for } x \geq 0 \end{cases} \quad (3.22)$$

ReLU fonksiyonunun grafiği Şekil 3.9'daki gibidir.

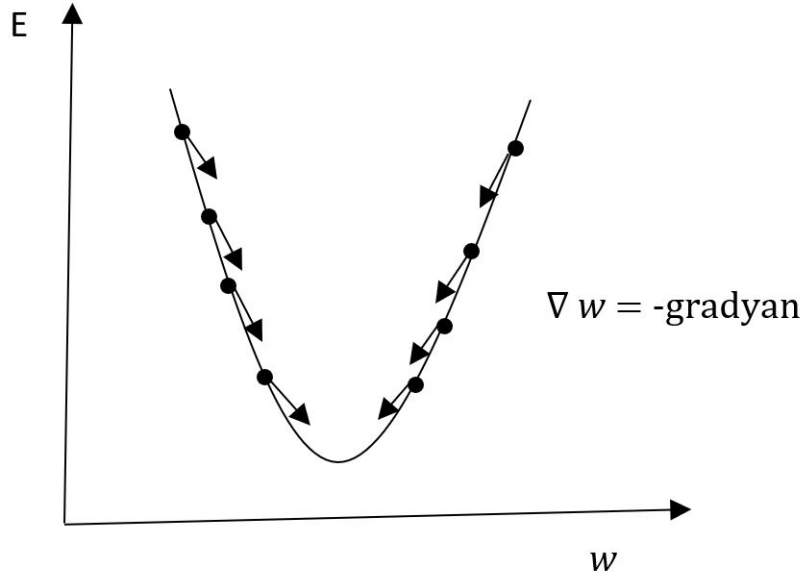


Şekil 3.9. ReLU Fonksiyonu

Sigmoid ve Tanh fonksiyonlarındaki sorunu çözmek için ReLu aktivasyon fonksiyonu önerilmiştir [47]. Hesaplama karmaşıklığı düşüktür ve üstel hesaplamalar gerektirmez. Aktivasyon değeri, bir eşik değeri olduğu sürece elde edilebilir [48]. Dezavantajı ise ReLu fonksiyonunun sadece değişken değeri pozitif olduğunda gradyanın kaybolması sorununu çözebilmesidir.

3.3.3 Eniyileme (Optimizasyon) Türleri

Maliyet fonksiyonun herhangi bir noktada türevini almak, o noktada fonksiyonun teğetinin eğimini ifade etmektedir yani Şekil 3.10'da görüldüğü gibi aşağıya doğru giden yönü göstermektedir. Eğim düşümü yönteminde her bir ağırlığa göre maliyetin kısmi türevi, her bir ağırlıktan çıkarılır. Bu yöntemle her adımda en dip noktaya yani ağırlığın en iyi olduğu değere ulaşılmaktadır.



Şekil 3.10. Eğimin ters yönünde ağırlık güncelleme işlemi

Hata (E) azaldıkça o noktadaki ağırlık değerleri güncellenerek adım adım yerel minimum noktasına ulaşılır. Hata fonksiyonu Eşitlik 3.23'teki gibidir.

$$E = \frac{1}{2} \left(y - f \left(\sum w_i x_i \right) \right)^2 \quad (3.23)$$

Eniyileme (optimizasyon) işlemi ağırlık eğitimi aşamasında oldukça önemli yere sahiptir. Eğitim süresince ağırlık ve bias değerlerinin güncellenerek eniyi sonuca gidilirken maliyeti en aza indirebilmek için bu fonksiyonlara ihtiyaç duyulmaktadır. Sinir ağlarını eniyilemek için en sık kullanılan algoritmalarından biri eğim düşümü (gradient descent) algoritmasıdır. Verinin boyutuna bağlı olarak kullanılabilecek 3 farklı eğim düşümü türü vardır.

3.3.3.1 Paket Eğim Düşümü (Batch Gradient Descent)

Paket eğim düşümü bütün veri kümesi üzerinden eğim düşümünü hesaplar. Kararlı olması sebebiyle tercih edilmektedir. Olumsuz tarafı veri kümesini bir bütün olarak işlemesinden kaynaklı olarak bellek yetersizliğine sebep olmasıdır. Maliyet fonksiyonunun denklemi Eşitlik 3.24'te verilmiştir.

$$\theta_{t+1} = \theta_t - \eta \nabla_{\theta} J(\theta_t) \quad (3.24)$$

Yukarıdaki eşitlikte η öğrenme katsayısını, $\theta \in \mathbb{R}^d$ model parametrelerini, $\nabla_{\theta} J(\theta)$ ise θ parametrelerine bağlı hedef fonksiyonu J 'nin eğimini ifade eder. Paket eğim düşümünde bütün veri kümesi tek seferde işlendiği için bu, hız açısından olumsuzluklara yol açmaktadır. Buna rağmen bu yöntemin tercih edilme sebebi konveks (içbükey) fonksiyonlarda global minimum noktasını, konveks olmayan fonksiyonlarda yerel minimum noktasını bulmayı garanti etmesidir [49].

3.3.3.2 Olasılıksal Eğim Düşümü (Stochastic Gradient Descent - SGD)

Bu yöntemde her eğitim örneği ve etiket bilgisi için eğim düşümü hesaplanarak ağırlıkların güncellenmesi sağlanmaktadır [49]. SGD algoritmasının denklemleri Eşitlik 3.25'te verildiği gibidir.

$$\theta_{t+1} = \theta_t - \eta \nabla_{\theta} J(\theta_t; x^{(i)}; y^{(i)}) \quad (3.25)$$

Bu yöntem paket eğim düşümünde olduğu gibi bütün eğitim kümesini tek seferde değerlendirip ağırlık güncellemesi yapmak yerine sadece eğitim örneği için ağırlık güncellemesi yaparak hatanın minimum noktasına ulaşmaya çalışır. Bu nedenle Paket Eğim Düşümü'nden daha verimli bir yöntemdir [50].

3.3.3.3 Mini Paket Eğim Düşümü (Mini-Batch Gradient Descent)

Mini paket eğim düşümü ile eğitim setindeki her bir mini paket için bir güncelleme gerçekleştirir. Modelin güncellenme sıklığı paket eğim düşümü yönteminden daha fazladır. Bu yöntem daha fazla yerel minimum noktasından kaçınarak global minimum noktasına hızlı bir yakınsama sağlamaktadır. Hesaplama işlemi Eşitlik 3.26'da olduğu gibi ifade edilebilir [49].

$$\theta_{t+1} = \theta_t - \eta \nabla_{\theta} J(\theta_t; x^{(i,i+n)}; y^{(i,i+n)}) \quad (3.26)$$

3.3.3.4 Eğim Düşümü Optimizasyonu (Gradient Descent Optimization)

Eğim düşümleri farklı eniyileme algoritmaları ile birlikte kullanılabilir. Bu algoritmalar-dan sıklıkla kullanılanlar Çizelge 3.1'de güncelleme denklemleri ile birlikte verilmiştir.

Eniyileme Algoritmaları	Parametre Güncelleme Fonksiyonu
Olasılıksal Eğim Düşümü (SGD)	$\theta_{t+1} = \theta_t - \eta \nabla_{\theta} J(\theta_t; x^{(i)}; y^{(i)})$
Momentum	$\mathcal{V}_{t+1} = \gamma \mathcal{V}_t + \eta \nabla_{\theta} J(\theta)$ $\theta_{t+1} = \theta_t - \mathcal{V}_{t+1}$
Adam	$\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{\mathcal{V}'_t + \epsilon}} m'_t$
Adagrad	$\theta_{t+1,i} = \theta_{t,i} - \frac{\eta}{\sqrt{G_{t,ii} + \epsilon}} g_{t,i}$
RMSprop	$\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{E[g^2]_t + \epsilon}} g_t$
Nadam	$\theta_{t+1} = \theta_t - \frac{a_t}{\sqrt{\hat{n}_t + \epsilon}} \hat{m}_t$

Çizelge 3.1. Eniyileme Algoritmaları

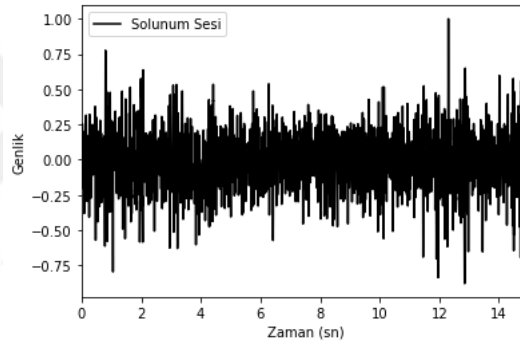
Günümüzde yaygın olarak kullanılan eniyileme algoritmalarının başında Adam (ADAPtive Moment) algoritması gelmektedir. Hem Adam hem de Adagrad, eğitimin ilk aşamasında maliyeti düşürmek için hızlı ilerleme kaydetmelerine rağmen, epok (algoritmanın tamamladığı tüm eğitim veri kümesini geçiş sayısı) arttıkça Adam ESA'lar için Adagrad'dan önemli ölçüde daha hızlıdır [51]. Solunum seslerinin sınıflandırılmasında da eniyileme algoritması olarak Adam kullanılmıştır.

4. ÖN İŞLEMLER

Veriler çoğu zaman direkt oluşturulan modele giriş olarak sunulabilecek mükemmellikte olmazlar. Kimi zaman veriyi elde ederken çeşitli bozulmalar (cihaz kaynaklı veya ortam gürültüsü gibi) meydana gelebilir ya da verinin kullanılacak modele uygun biçime dönüştürülmesi gerekebilir. İşte bu gibi durumlarda veri kümeleri üzerinde yapılan biçim değiştirme (ses sinyalinin görüntüye dönüştürülmesi gibi), normalleştirme, eksik veriyi tamamlama, gürültü bastırma vb. işlemlerine ön işlemler denilmektedir.

4.1 Gürültü Bastırma

Veri kümemizdeki solunum sesleri aşağıdaki Şekil 4.1’de de görüldüğü gibi 15 saniyelik sinyallerden oluşmaktadır.



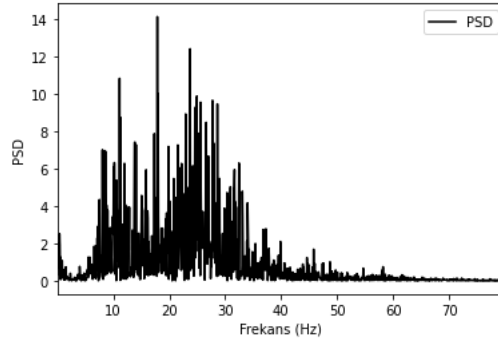
Şekil 4.1. Solunum Sesi Sinyali

Şekil 4.1’de görüldüğü gibi solunum sesi sinyalimiz zamanın bir fonksiyonudur. Bu sinyal $f(t)$ olarak isimlendirilebilir. $f(t)$ gürültülü sinyalinin Hızlı Fourier Dönüşümü yardımıyla Güç Spektral Yoğunluğu (GSY) elde edilir. GSY’nun Sözde Kod (pseudocode) parçacığı Çizelge 4.1’de verilmiştir.

```
t = zaman
f = sinyal
n = sinyalin uzunluğu
f hat = FFT(f,n)
PSD = f hat *conj(f hat )/n
```

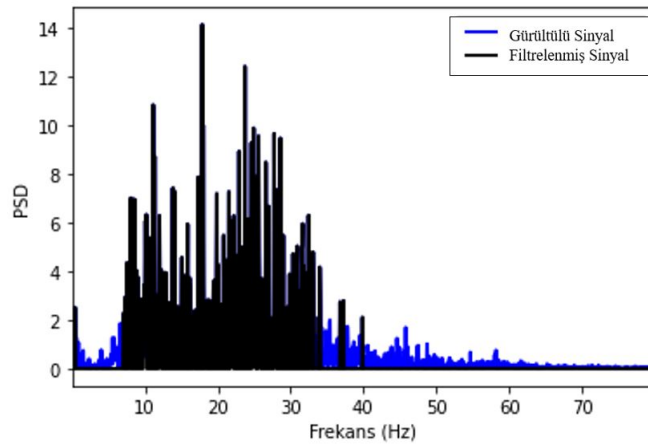
Çizelge 4.1. PSD Sözde Kod (pseudocode) Parçacığı

İnspirasyon ve ekspirasyon sırasında ortaya çıkan sağlıklı akciğer sesleri 100 Hz'in altındaki frekans aralığında büyük bir güç spektrumuna sahiptir. Sağlıklı akciğer sesleri spektrumunun enerjisi 100 Hz'in üzerinde keskin bir şekilde azalır, ancak yine de 2.000 Hz'e kadar tespit edilebilir [52].



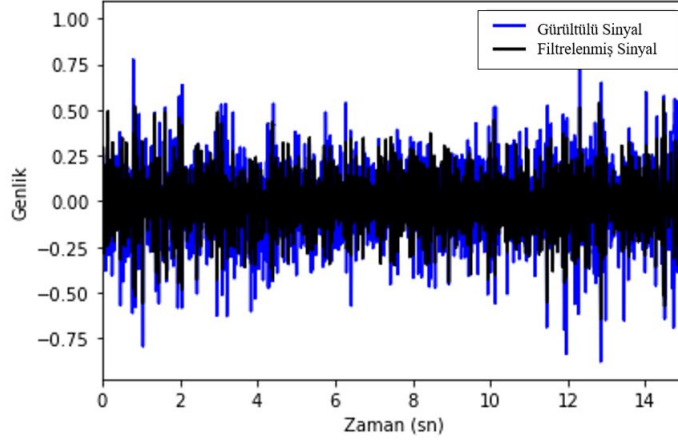
Şekil 4.2. Ses Sinyalinin Güç Spektral Yoğunluğu

Şekil 4.2'de solunum sesinin güç spektral yoğunluğu verilmiştir. x eksen frekansı belirlerken y eksen gürültülü sinyalde belirtilen frekanslarda ne kadar enerji olduğunu ifade etmektedir. Solunum sesleri incelendiğinde eşik değeri $PSD = 2$ olarak belirlenmiştir. PSD değerinin eşik değerinden düşük olduğu noktaları filtreleyerek elde edilen sinyal Şekil 4.3'te siyah renkle gösterilmiştir.



Şekil 4.3. Gürültülü ve Filtrelenmiş Solunum Sesinin Güç Spektral Yoğunluğu

Daha sonra elde edilen sinyalin hızlı Fourier Dönüşümü'nün tersi (iFFT) alınarak filtrelenmiş sinyal Şekil 4.4'te de görüldüğü gibi elde edilir .



Şekil 4.4. Gürültülü ve Filtrelenmiş Solunum Sesi

Şekil 4.4'te önışlemden geçmemiş gürültülü sinyal mavi renkle, filtrelenmiş sinyal ise siyah renkle çizilmiştir.

4.2 Solunum Sesi Sinyali'nin Görüntüye Dönüştürölme Yöntemleri

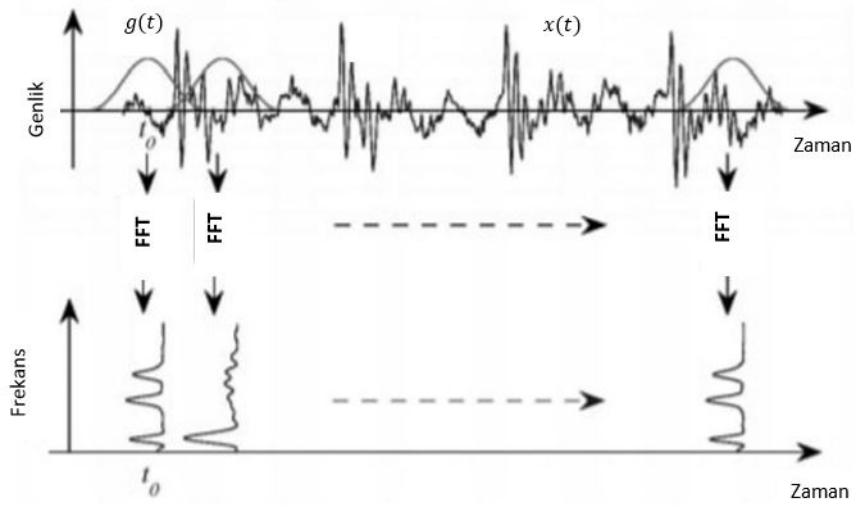
Tıbbi görüntülerle yapılan çalışmalarda klasik görüntü sınıflandırma işlemleri uygulamak her zaman mümkün olmayacaktır. Örneğin görüntü sınıflandırmada sıklıkla kullanılan kedi ve köpek görüntüleri ele alınacak olursa bu görüntüler ters çevrilse de görüntü anlamını yitirmez. Ancak spektrogram görüntüsü ters çevrilirse ya da herhangi bir yöne çevrilirse bu görüntüyü anlamsızlaştırır. Dolayısıyla çalışılacak veri kümesinin tanınması oldukça önemlidir. Solunum sesleri görüntüye dönüştürölürken görüntüler, Gabor Spektrogramı ve Mel Spektrogramı olmak üzere 2 farklı yöntemle elde edilmiştir.

4.2.1 Gabor Dönüşümü ve Spektrogram

Spektrogram bir sinyalin spektral yoğunluğunun zamanla değişiminin bir göstergesidir [53]. Sinyallerin çoğu özelliğini zaman ekseninde görmek mümkün olmadığından sinyalleri farklı eksenlere aktararak analiz etmek mümkündür. Genellikle seçilen yöntem frekans eksenine sinyali transfer etmek olmaktadır. Bu amaçla Hilbert dönüşümleri, güç spektrumları gibi spektral analiz yöntemleri kullanılmaktadır. Spektrogramlar bir sinyalin zaman içerisinde gücünü görsel olarak ifade ederler. Genellikle ses sinyalleri için kullanılırlar [54,55]. Konuşma sentezi, çevirisi, geliştirme, bölütleme ve kodlama gibi konuşma

işleme sistemlerinde KZFD spektrogramları akustik temsilin kilit noktası olarak oldukça yaygındır [56].

Bir spektrogram görüntüsünde, dikey eksen frekansları, yatay eksen ise zamanı temsil eder ve renklerin yoğunluğu, frekans bileşenlerinin genliğini temsil eder. Aşağıda normal, ral ve ronküs seslerinin spektrogram görüntüleri sıralanmıştır. Daha yüksek genlikler açık renklerden kırmızıya doğru bir renk aralığından oluşurken, düşük genlikler koyu renklerden maviye doğru bir renk aralığından oluşur. Spektrogramlar farklı yollarla üretilir. En yaygın yöntemlerden biri Fourier dönüşümüdür. Sadece FFT kullanılırsa zaman hakkında bir bilgi olamazdı. Bu yüzden spektrogramlarda KZFD kullanılır. KZFD’ler verilerin pencereleli FFT bölümleridir. Herhangi bir $x(t)$ sinyalinin KZFD dönüşümü Şekil 4.5’teki gibidir.

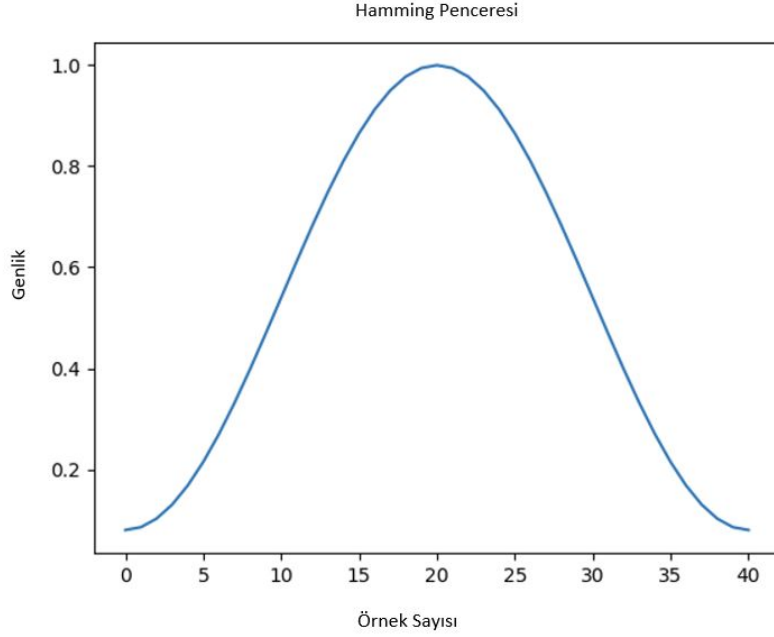


Şekil 4.5. Kısa Zamanlı Fourier Dönüşümü'nün Gösterimi

Şekil 4.5'te görüldüğü gibi $x(t)$ sinyalinin $g(t)$ pencereleri boyunca FFT'si alınarak sinyal zaman- frekans bölgesine taşınmış olur. Uygulanacak pencerenin boyutu sinyalin KZFD gösterimini değiştirir. Dar bir pencere kullanılması durumunda zaman açısından daha geniş bir pencereye göre zaman açısından daha doğru sonuçlar verecektir. Daha geniş pencere kullanılması ise frekans açısından daha doğru sonuçlar verecektir. Kullanılacak pencere türlerinden en yaygını Hamming penceresidir ve denklemi aşağıda verilmiştir.

$$w(n) = 0.54 - 0.46 \cos\left(\frac{2\pi n}{M-1}\right) \quad 0 \leq n \leq M-1 \quad (4.1)$$

Hamming penceresinin gösterimi Şekil 4.6'daki gibidir.

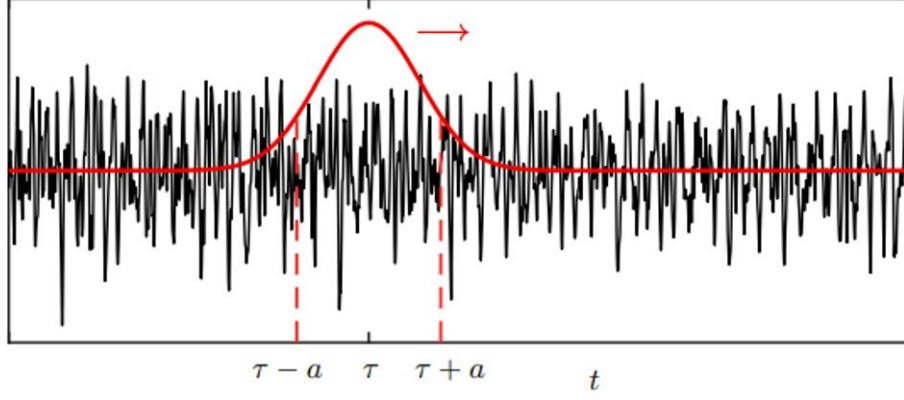


Şekil 4.6. Hamming Penceresi

Şekil 4.6, 40 örneklilik bir Hamming penceresini temsil etmektedir. Bu pencere sinyal boyunca kaydırılmaktadır. Fourier dönüşümü, verilen bir sinyalin frekans içeriği hakkında ayrıntılı bilgi vermesine rağmen, bu frekansların ne zaman meydana geldiği hakkında herhangi bir bilgi vermez. Fourier dönüşümü yalnızca periyodik ve durağan sinyalleri Eşitlik 4.2'deki integral ile karakterize edebilir [42].

$$f(x) = F^{-1}(\widehat{f(\omega)}) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \widehat{f(\omega)} e^{i\omega x} d\omega \quad (4.2)$$

Müzik bestesi gibi durağan olmayan frekans içeriğine sahip bir sinyal için, frekans içeriğini ve zaman içindeki gelişimini aynı anda karakterize etmek önemlidir. Kısa zamanlı Fourier dönüşümü (KZFD) olarak da bilinen Gabor dönüşümü, Şekil 4.7'de gösterildiği gibi hareketli bir pencerede FFT hesaplar [42].



Şekil 4.7. Gabor dönüşümünün Gauss penceresiyle elde edilmesi [42]

Şekil 4.7’den de görüldüğü gibi kullanılacak pencere zaman eksenı boyunca kaydırılarak KZFD hesaplanır. KZFD, Şekil 4.5’te gösterildiğı gibi, frekansa karşı zamanın bir grafiğı olan spektrogramla sonuçlanan frekans içeriğinin zaman içinde gösterimini sağlar. Kısa zamanlı Fourier dönüşümü 4.3’te verildiğı gibidir:

$$G(f)(t, \omega) = \hat{f}_g(t, \omega) = \int_{-\infty}^{\infty} f(\tau) e^{-i\omega\tau} \bar{g}(\tau - t) d\tau = \{f, g_{t,\omega}\} \quad (4.3)$$

Bu denklemde $g_{t,\omega}(\tau)$ şöyle tanımlanır:

$$g_{t,\omega}(\tau) = e^{i\omega t} g(\tau - t) \quad (4.4)$$

Burada $g(t)$ fonksiyonu kerneldir ve genellikle Gaussian olarak seçilir:

$$g(t) = e^{-(t-\tau)^2/a^2} \quad (4.5)$$

a parametresi Fourier dönüşümü için kısa zaman penceresinin yayılmasını belirler ve τ hareketli pencerenin merkezini belirler. KZFD’nin tersi şu şekilde ifade edilir:

$$f(t) = G^{-1} \left(\hat{f}_g(t, \omega) \right) = \frac{1}{2\pi \|g\|^2} \hat{f}_g(\tau, \omega) g(t - \tau) e^{i\omega t} d\omega dt \quad (4.6)$$

4.2.1.1 Ayırık Gabor Dönüşümü

Genel olarak, Gabor dönüşümü, FFT’de olduğu gibi ayırık sinyaller üzerinde gerçekleştirilir. Bu durumda, hem zamanı hem de frekansı ayırklaştırmak gerekir:

$$\begin{aligned} v &= j\Delta\omega \\ \tau &= k\Delta t \end{aligned} \quad (4.7)$$

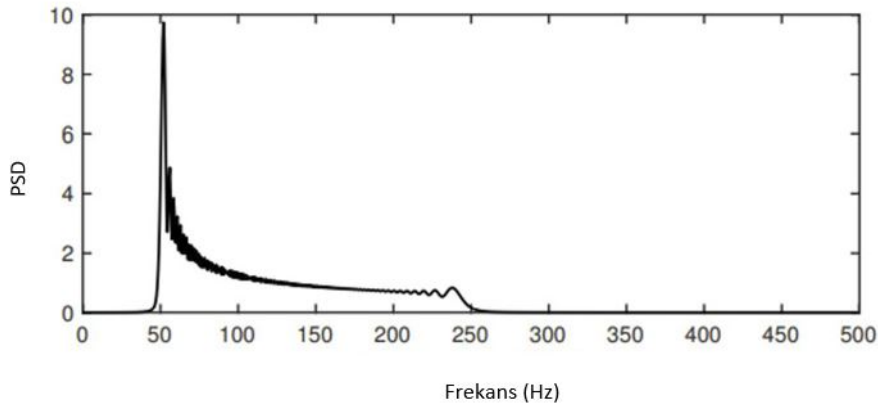
Ayrıklaştırılmış çekirdek işlevi (discretized kernel function) Eşitlik 4.8’da gösterildiği gibidir.

$$g_{j,k} = e^{ik\pi_j\Delta\omega t} g(t - k\Delta t) \quad (4.8)$$

Ayrık Gabor dönüşümü ise Eşitlik 4.9’da gösterilmiştir.

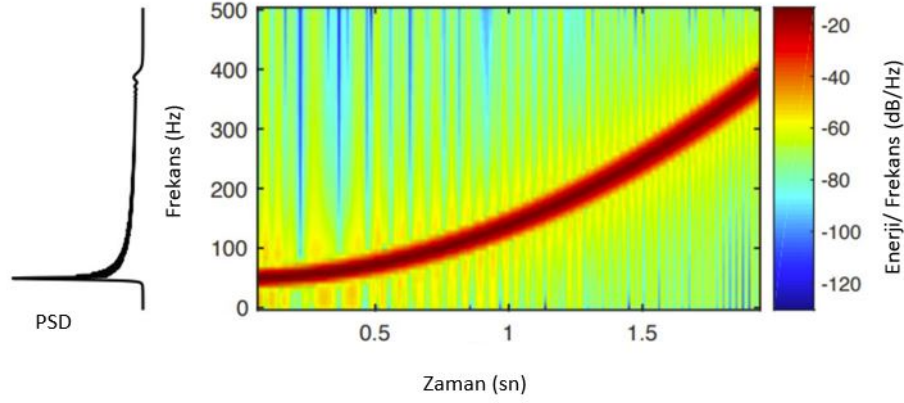
$$\hat{f}_{j,k} = \{f, g_{j,k}\} = \int_{-\infty}^{\infty} f(\tau) \bar{g}_{j,k}(\tau) d\tau \quad (4.9)$$

Bu integral daha sonra sonlu bir Riemman toplamı kullanılarak ayrıklaştırılmış f ve $\bar{g}_{j,k}$ fonksiyonlarına benzetilebilir [42]. Görsel olarak güç spektral yoğunluğunun önemini ifade etmesi için bir cıvıltı sinyali örneği seçilmiş ve bu örnek üzerinden güç spektral yoğunluğu Şekil 4.8’de gösterilmiştir.



Şekil 4.8. İkinci dereceden cıvıltı sinyalinin güç spektral yoğunluğu [42]

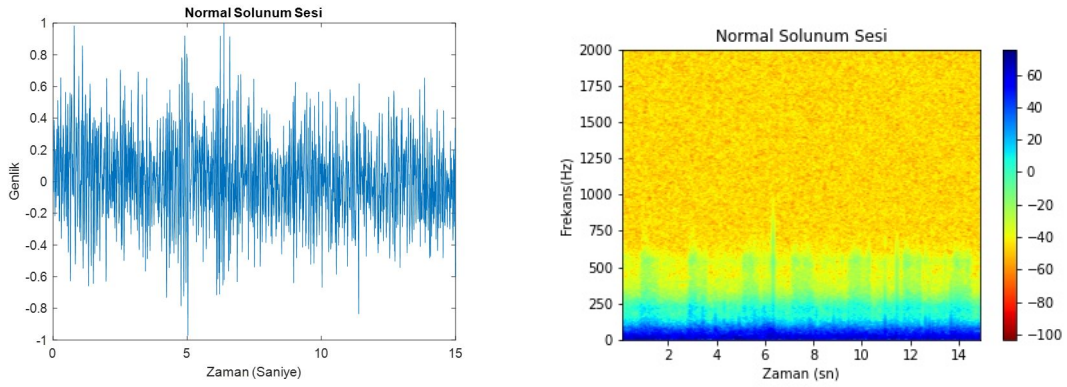
Şekil 4.8’den görüldüğü gibi örnek olarak verilen sinyalin yaklaşık 50-250 Hz bandında enerjisinin yoğun olduğu anlaşılmaktadır. Aynı cıvıltı sinyali örneği için spektrogram görüntüsü Şekil 4.9’da verilmiştir.



Şekil 4.9. İkinci dereceden cıvıltı sinyalinin spektrogramı [42]

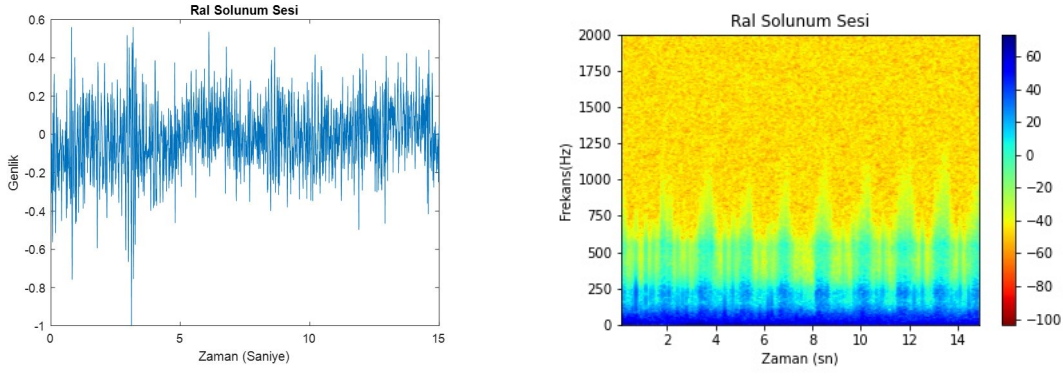
Şekil 4.9’da, spektrogramın satırları (row) boyunca entegre güce karşılık gelen PSD, solda gösterilmiştir.

Normal, Ral ve Ronküs seslerine ait Spektrogram görüntüleri sırasıyla Şekil 4.10, Şekil 4.11 ve Şekil 4.12’de gösterilmiştir.



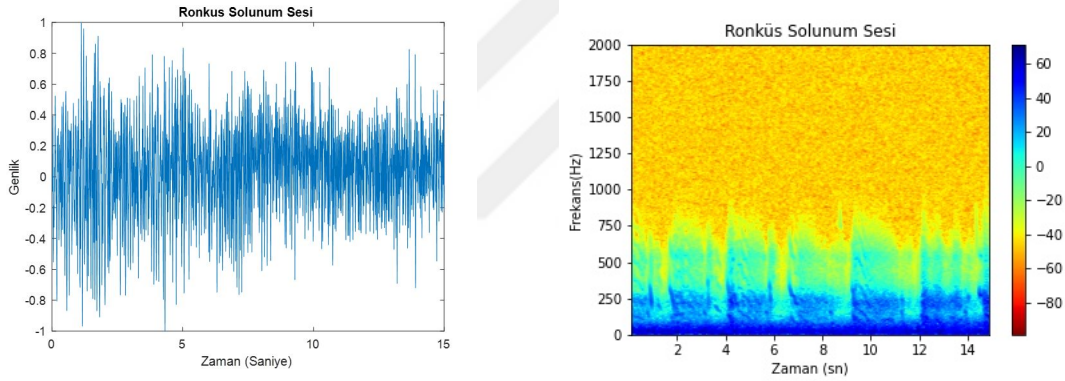
Şekil 4.10. Normal Solunum Sesi Sinyali ve Spektrogram Görüntüsü.

Sağlıklı insanlarda normal sesler $200 - 600\text{ Hz}$ frekans aralığında baskındır [23]. Normal olarak etiketlenen bir solunum sesinin zaman-genlik ve zaman-frekans yani spektrogram görüntüsü gösterilmiştir.



Şekil 4.11. Ral Solunum Sesi Sinyali ve Spektrogram Görüntüsü.

Ral solunum sesinin baskın frekansının $200Hz - 2kHz$ aralığında olduğu bilinmektedir [13]. Ral olarak etiketlenen bir solunum sesinin zaman-genlik ve zaman-frekans yani spektrogram görüntüsü de benzer olarak Şekil 4.11’de gösterilmiştir.



Şekil 4.12. Ronküs Solunum Sesi Sinyali ve Spektrogram Görüntüsü.

Ronküs solunum sesinin baskın frekansının $200Hz$ ’den düşük olduğu belirtilmiştir [13]. Ronküs olarak etiketlenen bir solunum sesinin zaman-genlik ve zaman-frekans yani spektrogram görüntüsü de incelendiğinde zaman ekseninde bir ses sinyalinin hangi gruba dahil olduğunu anlamak oldukça güçtür. Ancak spektrogram görüntülerine bakarak baskın frekansları hakkında bir yorum yapılabilir.

Bu tezde solunum seslerinden Gabor Transformu ile elde edilmiş Spektrogram görüntüleri kullanılmıştır. Ayrıca Mel ölçeği ve MFKK (Mel Frekans Kepstral Katsayıları) ile elde edilen Mel Spektrogram görüntüleri için de deneyler yapılmıştır.

4.2.2 Mel Spektrogramı

Mel frekansının kepsral özelliklerini hesaplamak için her pencereye sırasıyla şu adımlar uygulanır [57]: Sinyalin Ayırık Fourier Dönüşümü hesaplanarak, dönüşümden elde edilen katsayılar Mel filtre bankasının genlik frekans cevabına göre ağırlıklandırılır ve logaritmik enerjileri hesaplanarak, Ayırık Kosinüs Dönüşümü (DCT) bulunur. Kısaca sinyal önvurgu (pre-emphasis) filtresinden geçer; sonra (örtüşen) çerçevelere ayrılır ve her çerçeveye bir pencereleme işlevi uygulanır; daha sonra, her çerçevede bir Fourier dönüşümü yapılır (veya daha özel olarak Kısa Zamanlı Fourier Dönüşümü) ve güç spektrumu hesaplanır; ve daha sonra filtre bankaları hesaplanır. Mel frekansına ait kepsral katsayıları elde etmek için, filtre bankasına Ayırık Kosinüs Dönüşümü uygulanır son adım ise ortalama normalizasyonudur.

- Önvurgu (pre-emphasis)

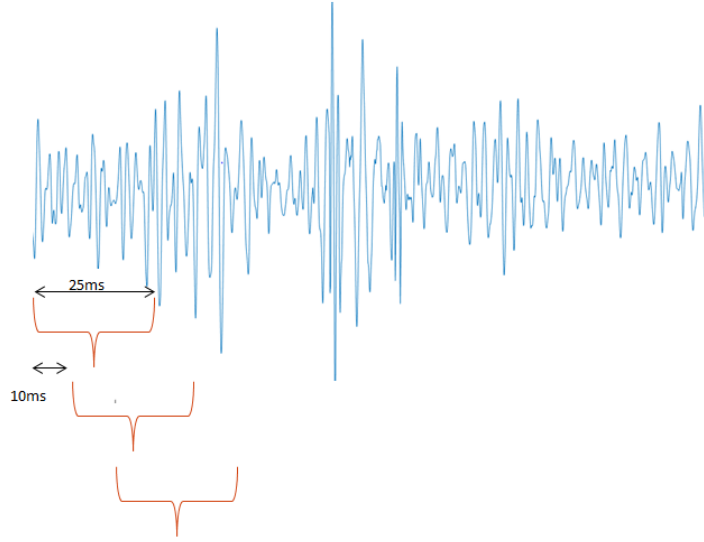
Yüksek frekansları yükseltmek için sinyale önvurgu filtresi uygulanmaktadır. Ön-vurgu filtresi, aşağıdaki denklemde birinci dereceden filtre kullanılarak x sinyaline uygulanabilir:

$$y(t) = x(t) - \alpha x(t - 1) \quad (4.10)$$

Böylelikle sinyalin yüksek frekanslı bölgeleri vurgulanırken düşük frekanslı bölgelerinin etkisi ise azaltılmış olur.

- Çerçeveleme (Framing)

Önvurgu işleminden sonra, sinyalinin çerçevelere ayrılması gerekir. Bu aşamada ses sinyalinin kısa zamanlı spektral analizini yapabilmek amacıyla sürekli ses sinyali, N adet örnekten oluşan çerçevelere bölünür. İlk çerçeve N örnekten oluşurken ikinci çerçeve ilk çerçeveden M örnek sonra başlar ve ilk çerçevenin $N-M$ örnek kadar üzerine biner. Aynı şekilde üçüncü çerçeve ilk çerçeveden $2M$ örnek, ikinci çerçeveden M örnek sonra başlar ve ilk çerçevenin $N-2M$ örnek kadar üstüne biner [58]. Bu işlem tüm ses sinyali boyunca yapılır. Çerçeveleme işlemi Şekil 4.13'te görülmektedir.



Şekil 4.13. Çerçeveleme

Bu adımın temel mantığı, bir sinyaldeki frekansların zaman içinde değişmesi, ve aynı zamanda sinyalin frekans konturlarını kaybedeceğimiz için Fourier dönüşümünü tüm sinyal boyunca uygulamanın mantıklı olmamasıdır. Bunu önlemek için, sinyaldeki frekansların çok kısa bir süre içinde sabit olduğunu varsayabiliriz. Dolayısıyla, bu kısa süreli çerçeve üzerinde bir Fourier dönüşümü yapıp, bitişik kareleri birleştirerek sinyalin frekans sınırlarını iyi bir şekilde yakalayabiliriz.

- Pencereleme (Windowing)

Sinyalin çerçevelere bölünmesinden sonra, her çerçeveye pencereleme işlemi uygulanır. Pencereleme işlemi her çerçeveyi pencereleyerek, sinyalin başındaki ve sonundaki sürekli bölümleri minimize eder. Buradaki amaç, pencereyi kullanarak çerçevelerin başı ve sonundaki bilgi içermeyen bölümleri kırmak, dolayısıyla spektral bozulmayı engellemektir. Eğer pencereyi $w(n)$, $0 \leq n \leq N - 1$ olarak kabul edip, her çerçevedeki örnek sayısını da N olarak kabul edersek, sinyalin pencerelenmiş hali aşağıdaki şekilde olur [58].

$$y(n) = x(n).w(n) \quad (4.11)$$

Pencere çeşitleri; Dikdörtgen, Blackman, Hamming ve Hanning pencereleridir. Hamming

penceresinin gösterimi Şekil 4.6’da verilmiştir.

- Fourier Dönüşümü ve Güç Spektrumu

Pencereleme işleminden sonra, Kısa Zamanlı Fourier Dönüşümü (KZFD) olarak da adlandırılan frekans spektrumunu hesaplamak için her çerçevede bir N (256 veya 512) noktalı FFT uygulayabiliriz. x sinyalinin i . çerçevesi $\mathcal{X}_i$ olmak üzere:

$$P = \frac{|FFT(x_i)|^2}{N} \quad (4.12)$$

Şeklinde verilir. Yukarıdaki denklemi kullanarak güç spektrumunu (periodogram) hesaplayabiliriz. Gabor spektrogramı için yapılan işlemler aynı şekilde Mel spektrogramını elde etmek için de yapılmıştır. Bundan sonra Mel Spektrogramının farkı olan filtre bankasının nasıl oluştuğu ifade edilecektir.

- Filtre Bankası

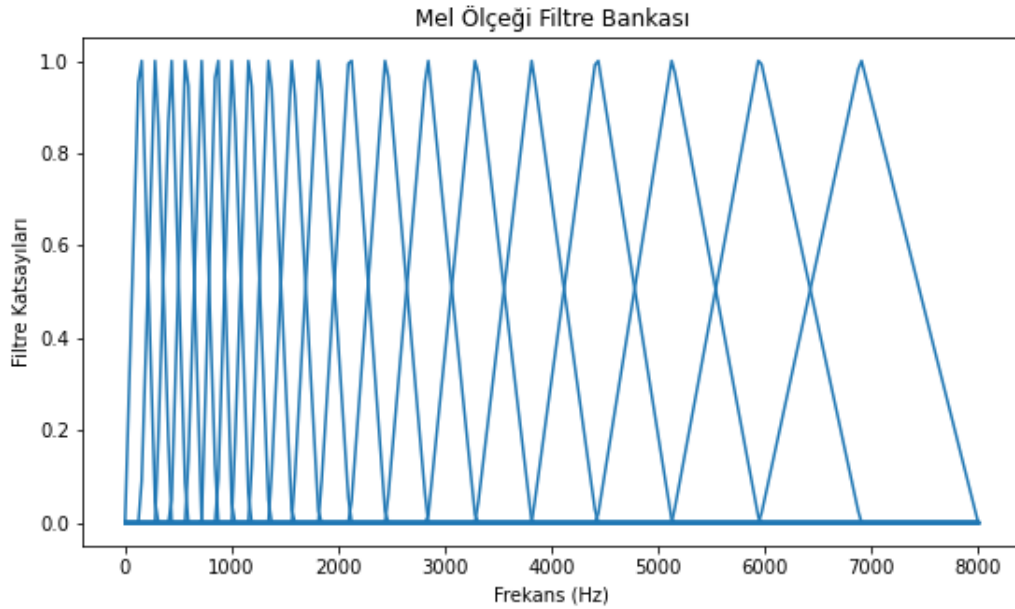
Filtre bankalarını hesaplamak için, güç spektrumuna bir Mel ölçeğinde tipik olarak filtre sayısı 40 olan üçgen filtreler uygulanmaktadır. Mel ölçeği, lineer olmayan insan kulağının ses algısını taklit etmeyi, düşük frekanslarda daha ayırt edici ve daha yüksek frekanslarda daha az ayırt edici olmayı amaçlamaktadır.

$$m = 2595 \log_{10} \left(1 + \frac{f}{700} \right), f = 700 (10^{m/2595} - 1) \quad (4.13)$$

Yukarıdaki denklemleri kullanarak Hertz (f) ve Mel (m) arasında dönüşüm yapılabilir. Filtre bankasındaki her filtre, merkez frekansında 1’lik bir cevaba sahip üçgen şeklindedir ve cevabın 0 olduğu iki bitişik filtrenin merkez frekanslarına ulaşana kadar doğrusal olarak 0’a doğru düşer, Eşitlik 4.14 ile modellenenebilir:

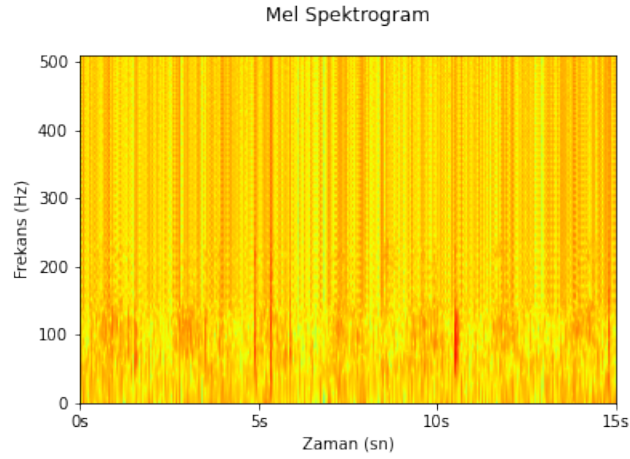
$$H_m(k) = \begin{cases} 0, & k < f(m-1) \\ \frac{k - f(m-1)}{f(m) - f(m-1)}, & f(m-1) \leq k < f(m) \\ 1, & k = f(m) \\ \frac{f(m+1) - k}{f(m+1) - f(m)}, & f(m) < k \leq f(m+1) \\ 0, & k > f(m+1) \end{cases} \quad (4.14)$$

Eşitlik 4.14’te formülü verilen filtre bankası ifadesinin 0 – 8000 Hz aralığı için gösterimi Şekil 4.14’te verilmiştir.



Şekil 4.14. Mel Ölçeği Filtre Bankasının Gösterimi

Filtre bankası sinyalin güç spektrumuna (periodogram) uygulandıktan sonra, spektrogramı Şekil 4.15’teki gibi elde edilir.



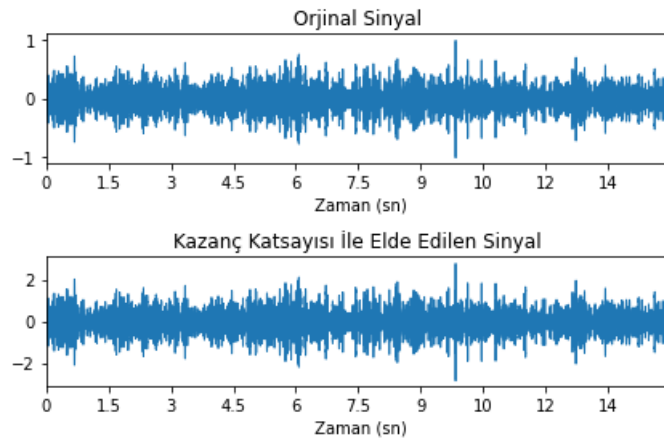
Şekil 4.15. Mel Spektrogram

Şekil 4.15'te Normal solunum sesine ait bir sinyalden elde edilen spektrogram görüntüsü görülmektedir. Solunum sesine ait frekans bileşenlerinin özellikle 0-100 Hz arasında baskın olduğu da spektrogram yardımıyla görülebilmektedir. İki farklı yöntemle elde edilmiş spektrogram görüntüsü kullandığımız için karışıklık olmaması adına Mel ölçeği kullanılarak elde edilen bu spektrogram görüntüsü Mel Spektrogramı olarak belirtilecektir.

4.3 Veri Artırma (Data Augmentation) Teknikleri

Derin öğrenmenin başarılı bir şekilde uygulanması, yeterli eğitim verisi olması durumunda sağlanabilir. Tıpkı insan beyninin yeni bir bilgiyi öğrenirken o bilgiye daha fazla maruz kaldıkça daha kolay öğrenmesi gibi makineler de, bir bilgiyi öğrenirken fazla sayıda veri ile daha kolay öğrenme işlemini gerçekleştirerek daha gerçekçi tahminlerde bulunabilirler. Yeterince büyük bir eğitim kümesi, sınıflar arasında bir denge ve doğru olarak etiketlenmiş gözlemler olması durumunda başarı sağlamak mümkündür. Düşük kaliteli veriler, aşırı uyumlanma (overfitting) gibi sorunlara yol açabilir. Ancak birçok gerçek dünya uygulamasında (özellikle medikal veriler için) bu tür ideal verilerin yeterince fazla miktarda toplanması mümkün olamayabilir. Bu nedenle, yeni veri toplamadan sınırlı veri kümelerini genişletmek için çeşitli artırma teknikleri geliştirilmiştir ve kullanılmaktadır. Veri artırma, tipik olarak, mevcut verileri değiştirerek veri kümesinin boyutunu artırmak için kullanılır. Daha büyük veri kümeleri her yıl daha erişilebilir hale geliyor, ancak Makine Öğrenmesi modelleri milyonlarca veri parametresi gerektiriyor. Veri artırmanın arkasındaki fikir, etiketlenmiş veriyi anlamsal ifadesini değiştirmeden deforme

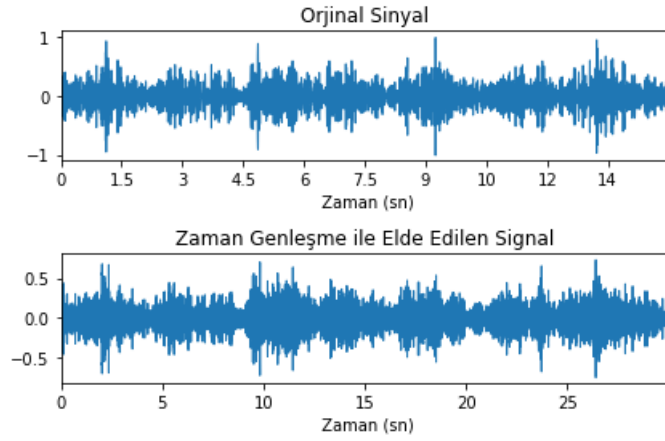
etmektedir [59]. Ağın görünmeyen veriler üzerindeki genellemesi, ağın bu ek veri örnekleri üzerinde eğitilmesiyle iyileştirilir. Görüntü veri alanında, eğitim örneklerinin sayısını artırmak için döndürme (rotation), öteleme (flipping) ve çevirme (translation) gibi basit yöntemler kullanılmaktadır [60]. Diğer veri türleri için büyütme o kadar basit değildir. Ses verilerinde, veri büyütmenin en yaygın uygulaması örnekleme süresini uzatmak, perdesini kaydırmak ve zamanını değiştirmektir [59]. Veri artırma ile eğitim veri setinin boyutunu artırmanın birçok örneği vardır [59, 61–63], ancak uygulamaları eğitim veri seti ile sınırlı olmak zorunda değildir. Aşağıdaki şekillerde, veri artırmanın etkisini göstermek için veri kümemizdeki solunum sesi verileri ile veri artırımı ile elde ettiğimiz veri birlikte gösterilmektedir.



Şekil 4.16. Kazanç Katsayısı ile Veri Artırma

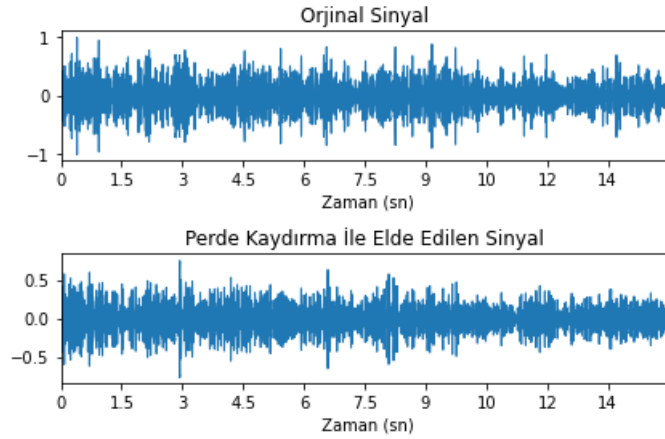
Şekil 4.16’da orjinal solunum sesi sinyali ile sesin kazanç katsayısı ile çarpımından elde edilen artırılmış solunum sesi sinyali görülmektedir.

Uygulanan katsayı değerine göre sinyalin perdesi değişir.



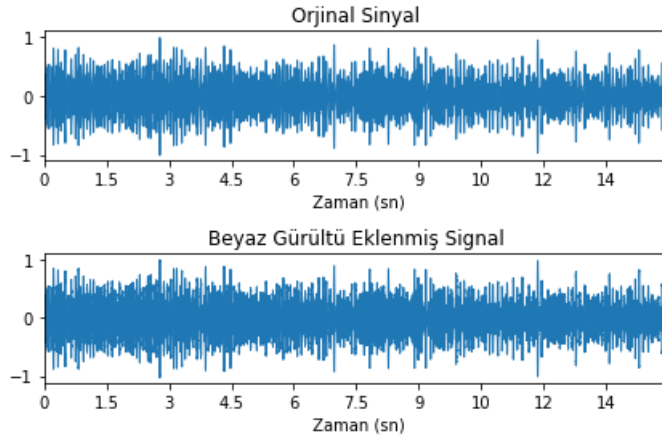
Şekil 4.17. Zaman Genleşme ile Veri Artırma

Şekil 4.17’de orjinal solunum sesi sinyaline zaman genleşme (time stretching) uygulanarak elde edilen artırılmış solunum sesi sinyali görülmektedir. Zaman genleşme ile sinyalin perdesi korunurken zamanda bir genleşme olur. Zaman genleşme (time stretching) uygulanmış sinyal, daha hızlı okunarak perde kaydırma (pitch shifting) uygulanmış bir sinyale dönüşür. Dijital anlamda bu, örnekleri dışarıda bırakmak anlamına gelir.



Şekil 4.18. Perde Kaydırma

Şekil 4.18’de orjinal solunum sesi sinyali ile sesin perdesi kaydırılarak elde edilen artırılmış solunum sesi sinyali görülmektedir. Perde kaydırmada orjinal dalganın perdesi yükseltilip azaltılabilirken uzunluğu korunur. Sıklıkla kullanılan bir diğer veri artırımı tekniği olan beyaz gürültü ekleme yöntemi ile elde edilen orjinal sinyal ve artırılmış sinyal Şekil 4.19’da gösterilmiştir.

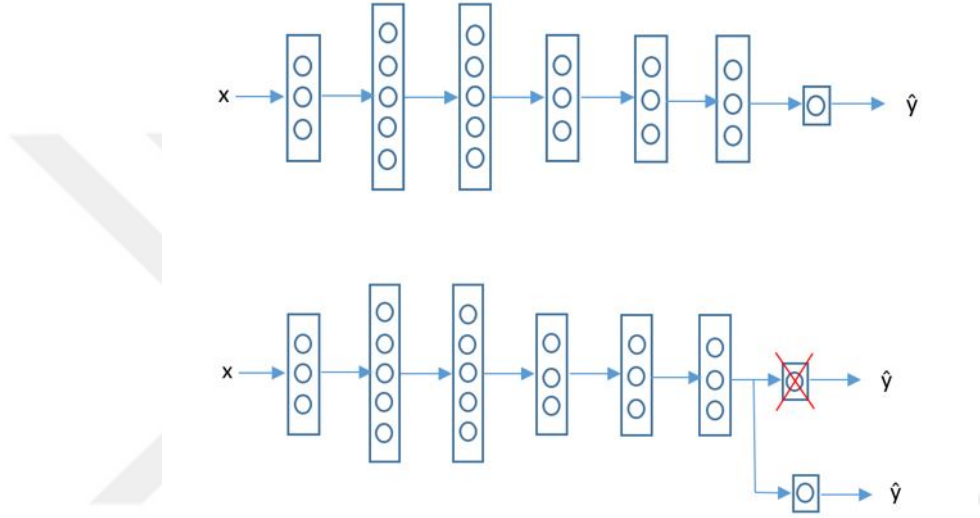


Şekil 4.19. Beyaz Gürültü Ekleme Yöntemi ile Veri Artırma

Şekil 4.19’da orjinal solunum sesi sinyaline beyaz gürültü ekleyerek elde edilen artırılmış solunum sesi sinyali görülmektedir. Solunum sesi gibi toplanması ve etiketlenmesi zaman alan veriler için veri artırımı yöntemleri veri eksikliğini gidermesi açısından düşünülebilir. Bu tezde de yukarıda bahsedilen veri artırımı yöntemleri kullanılarak 294 adet solunum sesi verisinden 1176 adet artırılmış solunum sesi verisi olmak üzere toplam 1470 adet solunum sesi verisi oluşturulmuştur. Sınıflandırma uygulamaları hem 294 adet veri için hem de 1470 adet veri için gerçekleştirilmiştir.

5. AKTARMALI ÖĞRENME (TRANSFER LEARNING)

Aktarmalı öğrenme, önceden edinilen bilgileri yeni bir alana aktaran bir makine öğrenimi tekniğidir. Yeni alan bir şekilde öncekine benzer olmalıdır. Motivasyonu yeni sorunları daha hızlı, daha iyi ve daha verimli çözmektir [64]. Yeni görevlerin öğrenilmesi için önceden öğrenilen görevlerden bilgilerin yeniden kullanılması veya aktarılması, bir takviye öğrenme aracının verimliliğini önemli ölçüde geliştirme potansiyeline sahiptir. Öğrenme aktarımının nasıl gerçekleştiği basitçe Şekil 5.1’de gösterilmiştir.



Şekil 5.1. Aktarmalı Öğrenme

Şekil 5.1’de görüldüğü gibi x girişlerini alıp bundan bir kestirim ($\hat{y}$) yapan bir modelin genel mimarisi alınarak yeni bir problemin çözümünde kullanılabilmektedir. Öğrenmeyi transfer etmenin (veya aktarmanın) iki yolu vardır.

1. Önceden eğitilmiş modelden özellik çıkarma ve ardından bunun üzerine bir sınıflandırıcı eğitimi
2. Modelin öğrenilen ağırlıklarını başlangıç parametresi olarak koruyarak önceden eğitilmiş bir modelin ince ayarının yapılması

Literatürde aktarmalı öğrenmenin metin sınıflandırma için kullanıldığı ve başarı sağladığı görülmektedir [65, 66]. Aktarmalı öğrenme bölütleme işlemlerinde de kullanılabilmektedir. Hsiao ve ark. [67] tarafından solunum seslerinde aktarmalı öğrenme kullanılarak yapılan çalışmada solunum seslerinin spektrogram görüntülerine ResNet-101 modeli transfer edilerek nefes alıp verme bölümlerinin ayrılması % 92’lik bir başarı sağlamıştır. Çalış-

malarında 22 hastadan dijital stetoskop veya gürültü önleyici mikrofon ile topladıkları solunum seslerini kullanılmıştır. Bunun yanında, aktarmalı öğrenme sınıflandırma problemlerini çözmek için de kullanılabilir. Yapılan bir çalışmada [26] solunum sesleri ICBHI veri kümesinden elde edilen solunum sesleri kullanılarak dört farklı sınıf için (1: Ral, 2: Ral + Hırıltı, 3: Normal, 4: Hırıltı) iki farklı yöntemle sınıflandırma yaparak % 65.5 ve % 63.09'luk sınıflandırma başarısı elde etmişlerdir. Kullandıkları yöntemlerde solunum seslerinin spektrogram görüntüleri VGG-16 modeli (ImageNet) transfer edilerek sınıflandırma yapılmış ve bir diğer yöntemde ise VGG-16'nın çıkışı DVM'e giriş olarak verilmiş ve DVM ile sınıflandırmada başarı oranının % 63.09'ten % 65.5'e çıktığı görülmüştür. Yapılan bir başka çalışmada [68] ise yine öğrenme aktarımı ile Normal, Ral ve Hırıltı sınıfına ait seslerin %98 doğrulukla sınıflandırıldığı belirtilmiştir.

5.1 VGG

VGG adını, Oxford Üniversitesi'nde kurulan Visual Geometry Group adındaki araştırma grubunun baş harflerinden almıştır. VGG'nin [69] boyutları 11 ile 19 katman arasında değişen çeşitli konfigürasyonlarla tasarlanmıştır. VGG, 2014'te ISLVR (ImageNet Large Scale Visual Recognition Challenge)'de iyi bir performans göstermiştir, ancak kazanan olamamıştır. Kazanan, VGG için % 7,3'lük ilk 5 hata oranına kıyasla % 6,7'lik ilk 5 hata oranına sahip olan GoogLeNet olmuştur. Yine de, VGG önemlidir çünkü gelecekteki mimarilerde standart hale gelen birkaç önemli tasarım ilkesine sahiptir [70]. VGG yenilik olarak, filtre boyutlarını azaltmış, ancak derinliği artırmıştır. Azaltılmış filtre boyutu artırılmış derinliği gerektirmektedir. Bunun nedeni, ağ derin olmadığı sürece küçük bir filtrenin görüntünün yalnızca küçük bir bölümünü yakalayabilmesidir. Bununla birlikte, birbirini izleyen üç evrişim, genellikle tek bir evrişimden daha ilginç ve karmaşık özellikleri yakalayabilir ve tek bir evrişimle ortaya çıkan aktivasyonlar ilkel kenar özellikleri gibi görünecektir. Bu nedenle, 7×7 filtreli ağ, daha küçük bölgelerdeki karmaşık şekilleri yakalayamayacaktır. Genel olarak, daha büyük derinlik, daha fazla doğrusal olmayışı (nonlinearity) ve daha fazla düzenliliği (regularization) gerektirir. Daha derin bir ağ, daha fazla ReLU katmanının varlığından dolayı daha fazla doğrusal olmayacak ve daha fazla düzenliliğe sahip olacaktır. Özetle daha fazla derinliğe ve azaltılmış filtre boyutuna sahip mimariler daha az parametre gerektirir. Artan derinlik, aynı zamanda, modelin ayırt etme gücünü artıran daha fazla sayıda doğrusal olmayan aktivasyonun kullanımına izin

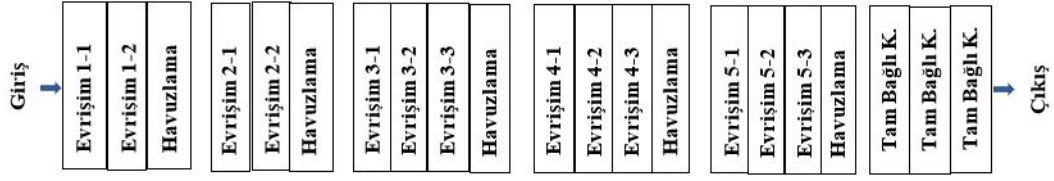
verir. Bu nedenle VGG her zaman 3×3 uzamsal ayak izine (footprint) ve 2×2 boyutunda havuzlamaya sahip filtreler kullanır. VGG'nin bir başka ilginç tasarım seçeneği, filtre sayısının her maksimum havuzlamadan sonra genellikle 2 kat artırılması ile elde edilebilir. Bu tasarım seçimi, katmanlar arasında işlemsel maliyette bir miktar denge sağlamıştır ve ResNet gibi bazı sonraki mimariler tarafından da kullanılmıştır. Derin yapılarda, artan derinlik kararsızlığa neden olmuştur. Bu sorun, önce daha sık bir mimarinin eğitildiği ve daha sonra başka katmanların eklendiği ön eğitim kullanılarak çözülmüştür. Bununla birlikte, ön eğitim katman katman temelinde yapılmamıştır. Bunun yerine, ilk önce mimarinin 11 katmanlı bir alt kümesi eğitilmiş, bu eğitilmiş katmanlar, daha derin mimaride katmanların bir alt kümesini başlatmak için kullanılmıştır. VGG'nin farklı konfigürasyonları Şekil 5.2'de gösterilmiştir. Bunlar arasında D sütunu ile gösterilen mimari, kazanan mimari olmuştur. Her maksimum havuzlamadan sonra filtre sayısının 2 kat artması nedeniyle, maksimum havuzlama, uzamsal yükseklik ve genişliğin 2 kat azalmasına neden olur, ancak bu, derinliği 2 kat artırarak telafi edilir. 3×3 filtreler ve 1 dolgu (padding) ile evrişimler gerçekleştirmek, uzamsal ayak izini değiştirmez. Bu nedenle, Şekil 5.2'nin D sütunundaki farklı maksimum havuzlama katmanları arasındaki bölgelerdeki her bir uzamsal boyutun (yani, yükseklik ve genişlik) boyutları sırasıyla 224, 112, 56, 28 ve 14'tür. Tam bağlantılı katmanın oluşturulmasından hemen önce, uzamsal ayak izini (footprint) 7'ye kadar azaltan son bir maksimum havuzlama gerçekleştirilir. Bu nedenle, ilk tam bağlı katman, 4096 nöron ile $7 \times 7 \times 512$ hacim arasında yoğun bağlantılara sahiptir [70].

ADI:	A	A-LRN	B	C	D	E
Katman Sayısı	11	11	13	16	16	19
	C3D64	C3D64	C3D64	C3D64	C3D64	C3D64
		LRN	C3D64	C3D64	C3D64	C3D64
	M	M	M	M	M	M
	C3D128	C3D128	C3D128	C3D128	C3D128	C3D128
			C3D128	C3D128	C3D128	C3D128
	M	M	M	M	M	M
	C3D256	C3D256	C3D256	C3D256	C3D256	C3D256
	C3D256	C3D256	C3D256	C3D256	C3D256	C3D256
				C1D256	C3D256	C3D256
						C3D256
	M	M	M	M	M	M
	C3D512	C3D512	C3D512	C3D512	C3D512	C3D512
	C3D512	C3D512	C3D512	C3D512	C3D512	C3D512
				C1D512	C3D512	C3D512
						C3D512
	M	M	M	M	M	M
	C3D512	C3D512	C3D512	C3D512	C3D512	C3D512
	C3D512	C3D512	C3D512	C3D512	C3D512	C3D512
				C1D512	C3D512	C3D512
						C3D512
	M	M	M	M	M	M
	FC4096	FC4096	FC4096	FC4096	FC4096	FC4096
	FC4096	FC4096	FC4096	FC4096	FC4096	FC4096
	FC1000	FC1000	FC1000	FC1000	FC1000	FC1000
	S	S	S	S	S	S

Şekil 5.2. VGG’de kullanılan konfigürasyonlar

Şekil 5.2’de C3D64 terimi, 3×3 (ve bazen 1×1) uzaysal boyuta sahip 64 filtre ile evrişimlerin gerçekleştirildiği durumu ifade eder. Filtrenin derinliği ilgili katmanla eşleşir. Tüm evrişimleri ReLU takip eder. Maksimum havuz katmanı M olarak adlandırılır ve yerel yanıt normalizasyonu LRN olarak adlandırılır. Softmax katmanı S ile belirtilir ve FC4096, 4096 birim ile tamamen bağlı (fully connected) bir katmanı ifade eder. Son katman kümesinin dışında, filtre sayısı her maksimum havuzlamadan sonra her zaman artar [70].

Şekil 5.3’te ise VGG mimarilerinden VGG16 gösterilmiştir.



Şekil 5.3. VGG-16'nın Mimarisi

Giriş katmanında 224x224 boyutunda RGB görüntüler (224x224x3) kullanılır. VGG16 her evrişim işlemi sonrasında görselden daha küçük ancak daha derin öznetelik haritaları (feature maps) çıkarır.

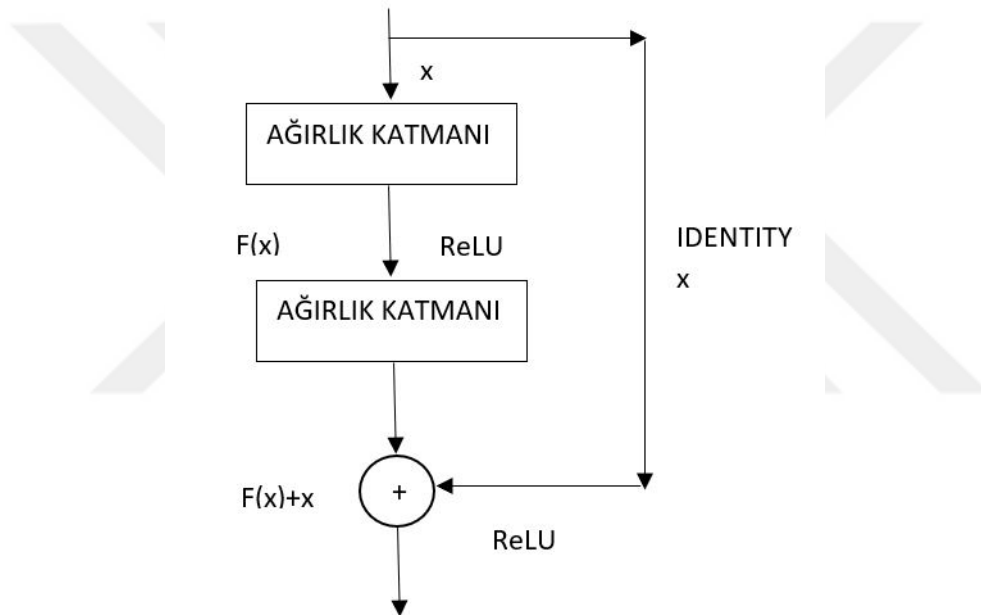
5.2 ResNet

ResNet'te, daha önce diğer mimariler tarafından kullanılan katman sayısından daha fazla katman sayısı (152 katman) kullanılmıştır. Bu mimari, 2015 yılında ILSVRC yarışmasının galibi olmuş ve % 3,6'lık hata elde ederek insan düzeyinde performansa sahip ilk sınıflandırıcı olmuştur. 152 katmanlı bir mimariyi eğitmek, bazı önemli yenilikler dahil edilmedikçe genellikle mümkün değildir [70]. Bu kadar derin ağların eğitimindeki ana sorun, katmanlar arasındaki gradyan akışının, gradyanların boyutunu artırabilen veya azaltabilen derin katmanlardaki çok sayıda işlem tarafından engellenmesidir. Kaybolan ve patlayan (vanishing and exploding) gradyanlar gibi sorunlara artan derinlik neden olur [70]. Bununla birlikte, yapılan bir çalışma [71], bu tür derin ağlardaki ana eğitim sorununun, özellikle yığın normalleştirme (batch normalization) kullanıldığında, bu gibi sorunlara neden olmayabileceğini göstermektedir. Ana sorun, öğrenme sürecinin makul bir süre içinde düzgün bir şekilde yakınsamasını sağlamanın zorluğundan kaynaklanmaktadır. Bu tür yakınsama sorunları, karmaşık kayıp yüzeyleri (complex loss surface) olan ağlarda yaygındır. Bazı derin ağlar, eğitim ve test hatası arasında büyük farklar göstermese de, eğitim ve test verilerindeki hata birçok derin ağda yüksektir [70]. Bu, eniyileme sürecinin yeterli ilerleme kaydetmediği anlamına gelir.

Bazı kavramlar sığ ağlar kullanılarak öğrenilebilirken, diğerleri ince detaylı bağlantılar gerektirir. Örneğin, kare bir çerçeve üzerinde duran karmaşık bir objeyi ele alırsak, objenin karmaşık özelliklerinden bazıları, özellik mühendisliği için çok sayıda katman ge-

rektirebilirken, kare çerçevenin özellikleri çok daha az katman gerektirebilir. Kavramları öğrenmek için tüm yollarda sabit derinliğe sahip çok derin bir ağ kullanıldığında yakınsama gereksiz yere yavaş olacaktır, bunların çoğu sığ mimariler kullanılarak da öğrenilebilir. ResNet, her bir özelliği öğrenmek için kaç katman kullanılacağına neden sinir ağının karar vermesine izin vermeyelim fikrinden yola çıkmıştır.

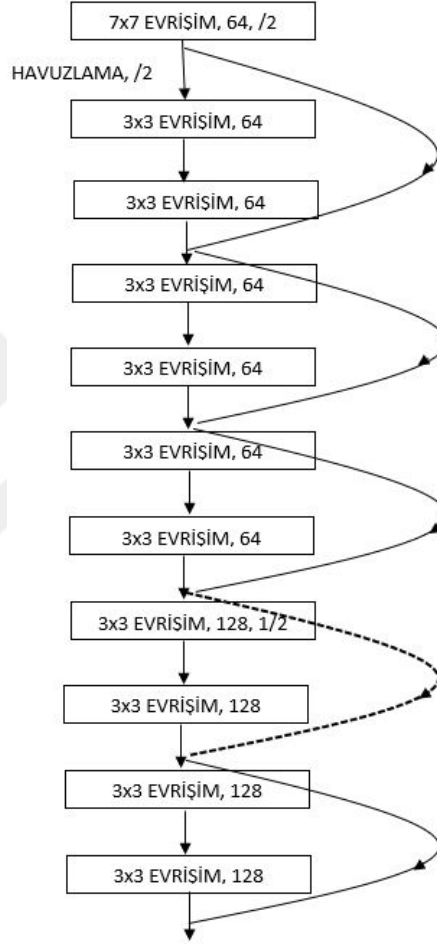
ResNet, katmanlar arasında atlama bağlantılarını kullanır. Çoğu ileri beslemeli ağ, yalnızca i ve $(i + 1)$ katmanları arasındaki bağlantıları içerirken, ResNet, $r > 1$ için i ve $(i + r)$ katmanları arasındaki bağlantıları içerir [70]. ResNet'in temel birimini oluşturan bu tür atlama bağlantılarının örnekleri, Şekil 5.4'te $r = 2$ ile gösterilmiştir [71].



Şekil 5.4. Artık Öğrenme Bloğu (Block of residual learning)

Bu atlama bağlantısı (skip connection), basitçe katman i girdisini kopyalar ve bunu $(i + r)$ katmanının çıktısına ekler. Böyle bir yaklaşım, gradyan akışının etkili olmasını sağlar. Bu temel birim, artık (residual) modül olarak adlandırılır ve tüm ağ, bu temel modüllerin bir araya getirilmesiyle oluşturulur. Çoğu katmanda, uygun şekilde doldurulmuş bir filtre kullanılır, böylece girdinin uzamsal boyutu ve derinliği katmandan katmana değişmez. Bu gibi durumlarda, basitçe i 'inci katmanın girdisini $(i + r)$ 'inci katmanın girdisine eklemek kolaydır. Bununla birlikte, bazı katmanlar, her bir uzaysal boyutu 2 kat azaltmak için aşamalı evrişimler kullanır. Aynı zamanda, derinlik, daha fazla sayıda filtre kullanılarak

2 kat artırılır. Böyle bir durumda, özdeşlik fonksiyonu (identity function) atlama bağlantısı üzerinden kullanılamaz. Bu nedenle, boyutluluğu ayarlamak için atlama bağlantısının üzerine doğrusal bir projeksiyon matrisinin uygulanması gerekebilir. ResNet'in orijinal fikrinde, yalnızca i ve $(i + r)$ katmanları arasına bağlantılar eklenir. Örneğin, $r = 2$ için, yalnızca ardışık tek katmanlar arasında atlama bağlantıları kullanılır [70]. Mimarinin ilk birkaç katmanının bir örneği Şekil 5.5'te gösterilmiştir.

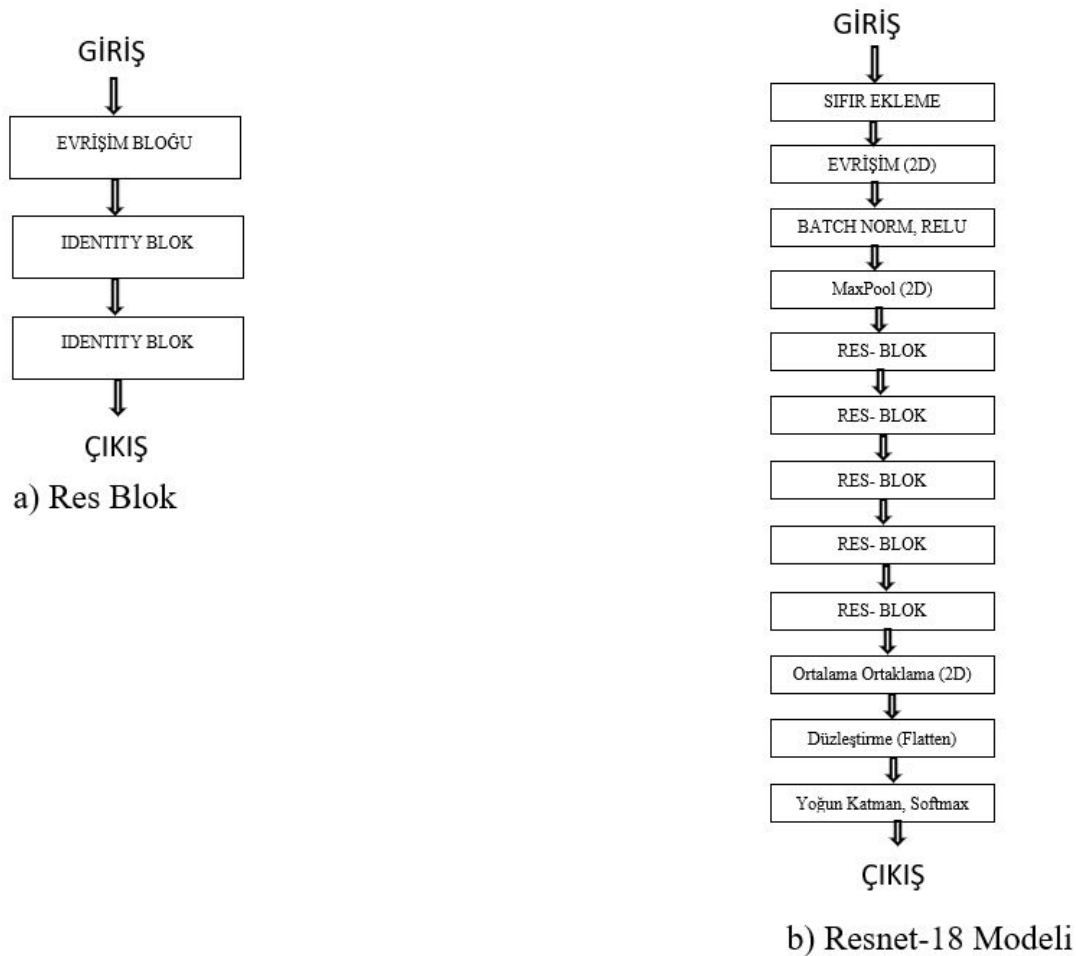


Şekil 5.5. ResNet'in Kısmi Mimarisi

Sırasıyla 34, 50, 101 ve 152 katman içeren, dört farklı mimarinin test edildiği bir çalışmada [71] en iyi performansı 152 katmanlı mimari sağlamıştır, ancak 34 katmanlı mimari bile, önceki yıllarda ILSVRC'de gösterilen performanstan daha iyi performans göstermiştir.

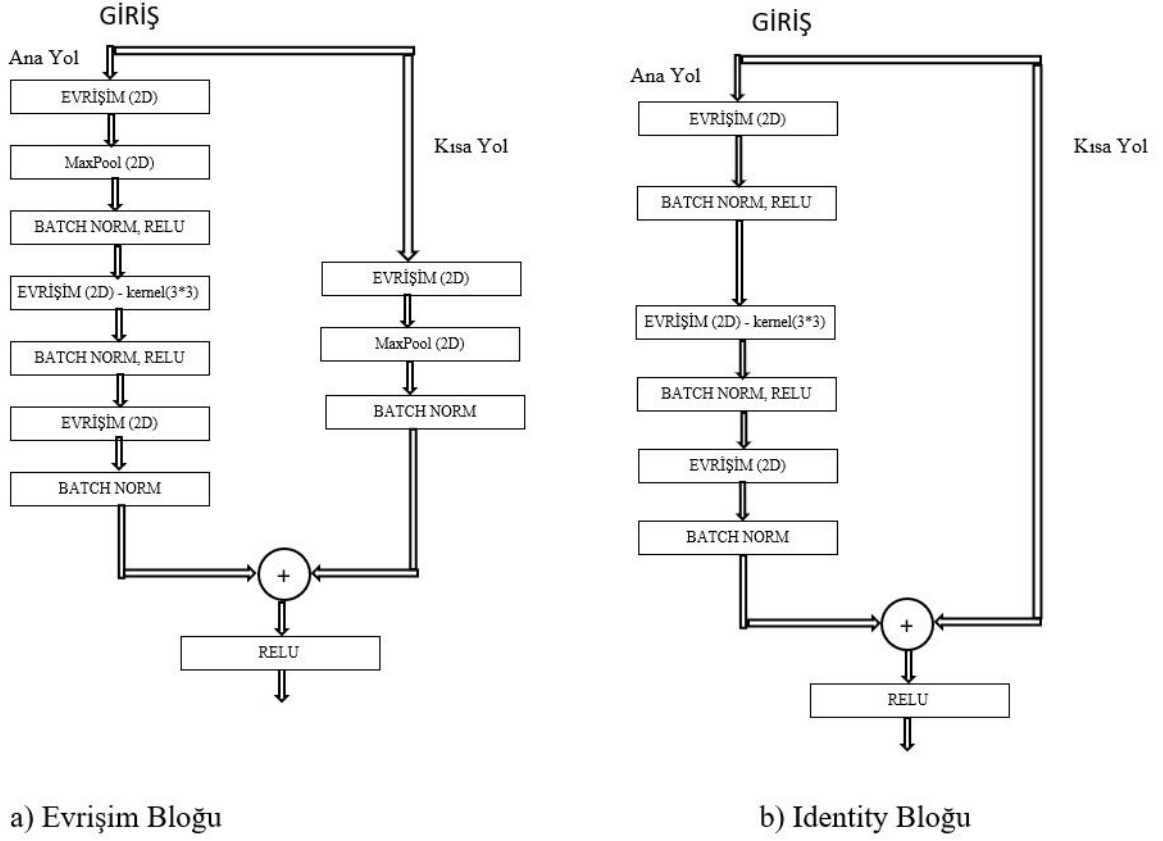
Atlama bağlantılarının kullanımı, engelsiz gradyan akışı yolları sağlar ve bu, geri yayılım algoritmasının davranışı için önemli sonuçlar verir. Geniş artık ağlar üzerine yapılan bir

araştırmada [72], artık ağın derinliğinin artmasının her zaman iyi olmadığı, çünkü aşırı derin yolların çoğunun zaten kullanılmadığı gösterilmiştir. Ayrıca bu çalışma [72], toplam katman sayısını bir dereceye kadar sınırlayarak (diyelim ki 150 yerine 50) ve her katmanda artan sayıda filtre kullanarak daha iyi sonuçların elde edilebileceğini göstermektedir. ResNet öncesi standartlara göre 50 büyüklüğünde derinlik hala oldukça büyüktür, ancak artık ağlarla yapılan son deneylerde kullanılan derinlikle karşılaştırıldığında bu derinlik küçüktür [70]. Res Blok ve Resnet-18 modeli Şekil 5.6’da gösterilmiştir.



Şekil 5.6. Resnet-18 Modeli ve Res Blok Yapısı

Şekil 5.6’da görüldüğü gibi Res Blok bir evrişim bloğu ve peş peşe iki adet identity blok-tan oluşmaktadır. ResNet-18 modeli de bu Res bloklarının beşinin art arda sıralandığı bir yapıdadır. Res Blok yapısının içinde bulunan evrişim ve identity blokları sırasıyla Şekil 5.7’de gösterildiği gibidir.



Şekil 5.7. Resnet-18 Modeli Evrişim ve Identity Blok Yapısı

Şekil 5.6 ve Şekil 5.7 ile Resnet 18 modelinin mimarisi açıkça gösterilmiştir. Oldukça fazla sayıda evrişim katmanı barındıran VGG ve ResNet dışında DenseNet201, InceptionV3 gibi mimariler de bu çalışmada kullanılmıştır.

Inception ağı sürekli evrimleşmiş ve bu ağın bir dizi versiyonu oluşmuştur. En popüler versiyonları: InceptionV1, InceptionV2 ve InceptionV3'tür. Inception'ın her sürümü, önceki sürüme göre önemli bir gelişmedir. InceptionV3 ağ modeli derin bir sinir ağıdır, düşük yapılandırılmış bir bilgisayarla doğrudan eğitmek çok zordur, eğitilmesi en az birkaç gün sürer [73]. DenseNet201, yoğunlaştırılmış ağdan yararlanır ve eğitilmesi kolaydır [74]. DenseNet201, ImageNet ve CIFAR-100 gibi çeşitli veri kümelerinde başarı sağlamıştır.

ResNet modeli solunum seslerini sınıflandırma çalışmalarında da kullanılmıştır. Yapılan bir çalışmada ICBHI veri kümesi kullanılarak ResNet modeli ile [75] solunum seslerinin spektrogram görüntüleri ResNet mimarisinden geçirilirken ResNet'e NL (Non-Local)

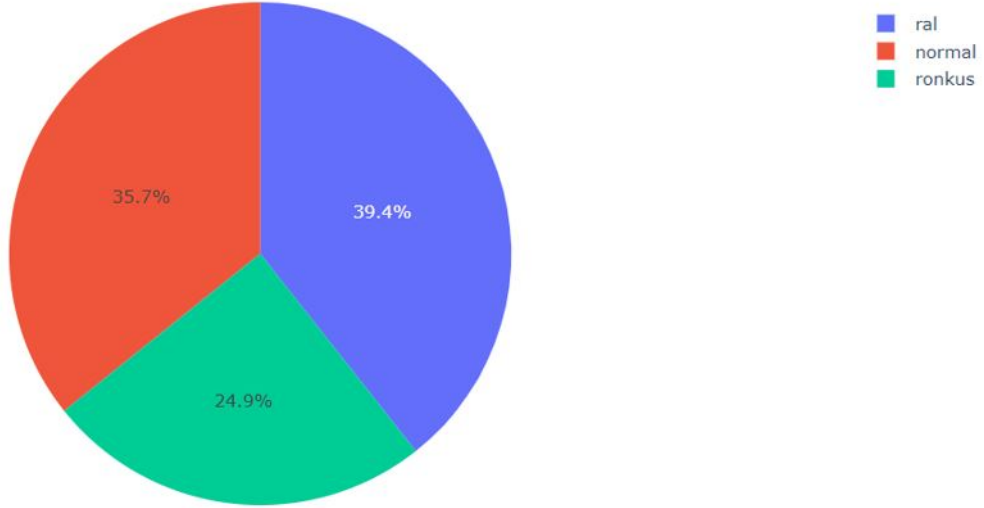
olarak isimlendirdikleri bir ara katman ekleyerek %52.26'lık başarı sağlamışlardır. Bundan önceki çalışmalarında [76] ise, oluşturdukları başka bir ResNet modeli ile % 50.16'lık başarı elde etmişlerdir.

15 saniyelik wav formatında kayıt edilmiş 294 adet solunum sesi verisi 4kHz örnekleme frekansı ile sayısallaştırılmıştır. 3 farklı solunum ses grubuna ait bu seslerin makine öğrenimi teknikleri ile sınıflandırılmasına geçmeden önce veri kümesini tanımak önemlidir. Solunum ses verileri içerisinde kan dolaşım sisteminden gelen sesler, kalp atış hızı, ortam gürültüsü gibi sesleri de barındırdığından stokastik sinyallerdendir. Bu yüzden ses sinyalleri işlenirken gürültü bastırma (Bölüm 4.1) yapılmasına karar verilmiştir. Bunun yanında makine öğrenimi yöntemlerinde veri sayısının büyüklüğü eğitim ve test aşamasında başarıyı artıran bir faktördür. Bu nedenden solunum seslerinin veri artırma teknikleri ile (Bölüm 4.3) çoğaltılmasıyla elde edilen artırılmış verilerle de deneyler yapılmıştır. Sınıflandırma problemlerinde gösterdiği başarı nedeniyle ESA mimarileri kullanılarak sınıflandırma işlemleri gerçekleştirilmiştir. Veri kümesinin boyutu küçükse ESA modeli aşırı uyma problemi yüzünden başarı sağlayamaz, bu nedenle solunum sesi sınıflandırma için de öğrenme aktarımı yaklaşımı (Bölüm 5) kullanılmıştır. Başarısı kanıtlanmış bir çok ESA mimarisi mevcuttur (VGG, ResNet, Inception,...), bu mimarilerin de solunum verileri üzerindeki başarıları öğrenme aktarımı ile incelenecektir.

6. SOLUNUM SESLERİNİN DERİN ÖĞRENME YÖNTEMLERİ İLE SINIFLANDIRILMASI UYGULAMALARI

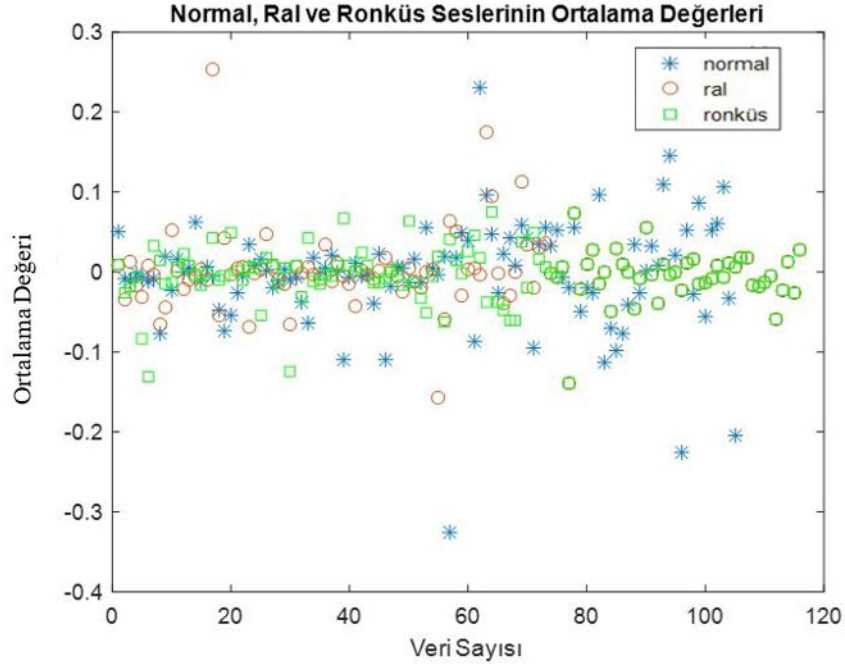
6.1 Veri Kümeleri

Bu çalışmada kullanılan solunum sesi verileri Hacettepe Üniversitesi Göğüs Hastalıkları AD’de çalışan 3 farklı uzman hekimin (Bölüm 2.2) de önerileri doğrultusunda belirlenen veri grupları için Klinik Araştırmalar Etik Kurulu (Karar No: GO 18/631 -45) onayı alınarak bir yıl süreyle (1 Eylül 2018- 2019) toplanmıştır. Normal olarak etiketlenmiş 105, Ral olarak etiketlenmiş 116 ve Ronküs olarak etiketlenmiş 73 adet solunum sesi olmak üzere toplam 294 adet solunum sesi veri kümesi elde edilmiştir. Solunum sesleri veri kümesinin dağılımı Şekil 6.1’de görüldüğü gibidir.



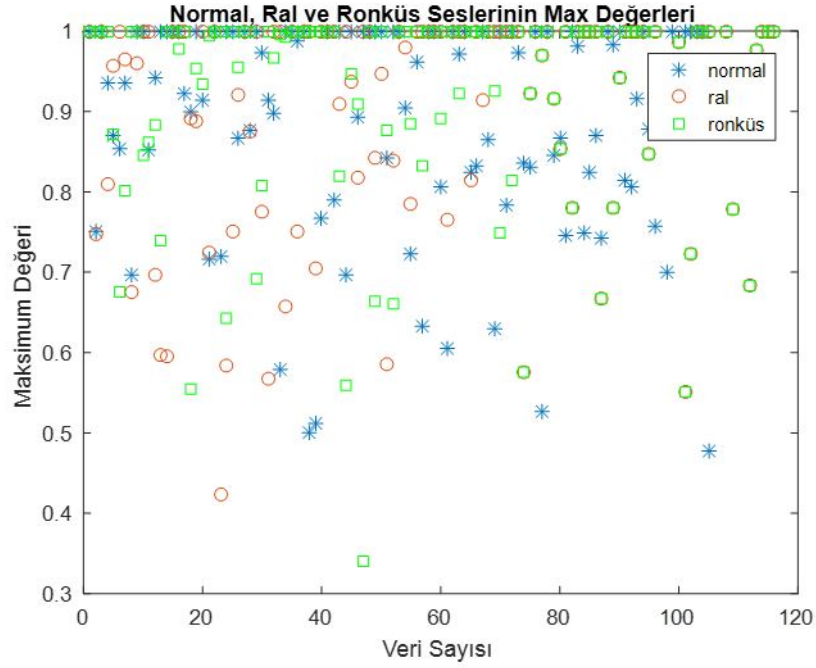
Şekil 6.1. Veri Kumesinin Dağılımı

Şekil 6.1’deki dağılım aynı zamanda gürültü bastırma uygulanan veriler ve veri artırma ile elde edilen veriler için de geçerlidir. Veri artırma ile elde edilen veri sayısı 1470 olup, 525 adet Normal, 580 adet Ral ve 365 adet Ronküs sesi içermektedir. Yapılan tüm deneylerde k katlamalı çapraz doğrulama uygulanmıştır. K katlamalı çapraz doğrulama verileri eğitim ve test kümesine ayırırken modelin sürekli aynı veri kümeleri ile karşılaşarak tesadüfi sonuçlar elde etmesini engellemeye yönelik bir yöntemdir. Bunun için verileri seçilen k değerine bağlı olarak kümelere ayırır ve her defasında veri kümelerinden biri test amacıyla kullanılır, kalan $k - 1$ adet veri kümesi ise eğitim amacıyla kullanılır. Veri kümesindeki 294 adet sesin ortalama değerleri Şekil 6.2’de gösterildiği gibidir.



Şekil 6.2. Normal, Ral ve Ronküs Seslerinin Ortalama Değerleri

Solunum sesleri incelendiğinde ortalama değerlerinin 0 civarında kümelenmiş olduğu gözlemlenmiştir (Şekil 6.2). Veri kümesindeki 294 adet sesin maksimum değerleri ise Şekil 6.3'te gösterildiği gibidir.



Şekil 6.3. Normal, Ral ve Ronküs Seslerinin Maksimum Değerleri

Solunum sesleri incelendiğinde maksimum değerlerinin genellikle 1 civarında kümelenmiş olduğu gözlemlenmiştir (Şekil 6.3). Bu tez çalışması kapsamında toplanmış olan 294 adet veri, ve veriye uygulanan önışlemler sonucunda 3 farklı grup Çizelge 6.1'deki gibi elde edilmiştir.

Veri Kümesi Grupları	Veri Sayısı	Önişlemler
Orijinal Veriler	294 adet SS	Önişlem yok
Artırılmış Veriler	1470 adet SS	Veri Artırımı
Gürültüsüzleştirilmiş Veriler	294 adet SS	Gürültü Bastırma

Çizelge 6.1. Oluşturulan Veri Grupları

Bu tez çalışmasında, Çizelge 6.1'de gösterildiği üzere hem 294 adet veriden oluşan orijinal veriler, hem de önışlemler sonucunda elde edilen artırılmış ve gürültüsüzleştirilmiş veriler için deneyler yapılmıştır. Sınıflandırma işlemleri Bölüm 4.2'de anlatıldığı gibi hem

Gabor ve Mel Spektrogram görüntülerine dönüştürülerek yapılmıştır. Bu verilerin %80'i modelin eğitimi aşamasında, geriye kalan %20'si ise modelin başarısını ölçmek için test aşamasında kullanılmıştır.

6.2 Değerlendirme Yöntemleri

Bir modelin başarısının belirlenmesi için modelin değerlendirilmesi gerekir. Kullanılan verilere ve kullanılan makine öğrenimi modellerinin türüne bağlı olarak, farklı değerlendirme ölçütleri ve yöntemleri kullanılabilir. Bu bölümde, solunum seslerini sınıflandırmak amacıyla kullandığımız yöntemleri değerlendirmek için faydalandığımız metriklere yer verilmiştir. İkili sınıflandırma yapılırken, bir modelin performansını değerlendirmek için genellikle aşağıdaki temel metrikler hesaplanır.

TP (Gerçek Pozitifler): Pozitif olarak sınıflandırılan pozitif örneklerin sayısı

TN (Gerçek Negatifler): Negatif olarak sınıflandırılan negatif örneklerin sayısı

FP (Yanlış Pozitifler): Negatif olarak sınıflandırılan pozitif örneklerin sayısı

FN (Yanlış Negatifler): Pozitif olarak sınıflandırılan negatif örneklerin sayısı

	Pozitif Tahmin	Negatif Tahmin
Gerçek Pozitif	TP	FN
Gerçek Negatif	FP	TN

Şekil 6.4. Hata Matrisi Örneği

Bu dört temel değer, daha açıklayıcı metrikleri hesaplamak için kullanılabilir. En yaygın olarak hesaplanan metrik, doğruluktur:

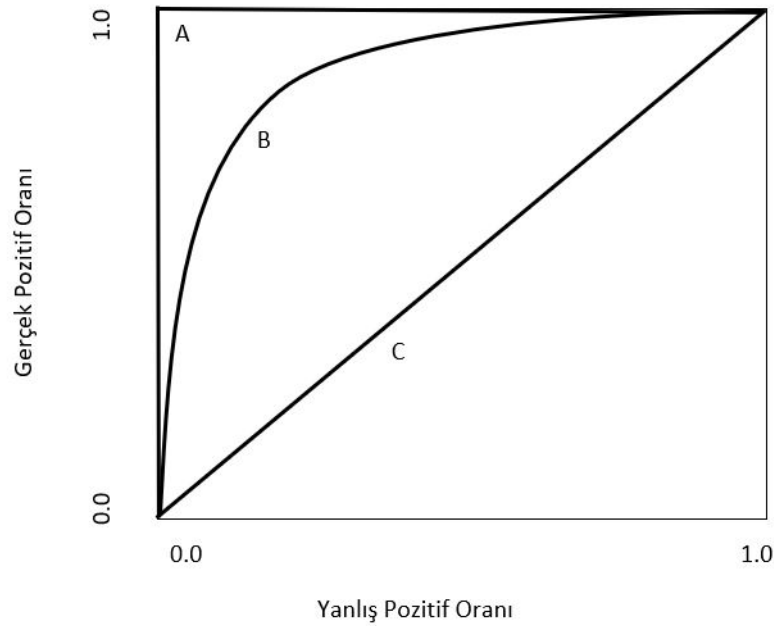
$$\text{Doğruluk} = \frac{TP + TN}{TP + TN + FP + FN} \quad (6.1)$$

Dolayısıyla doğruluk, yapılan doğru tahmin sayısının toplam tahmin sayısına oranı olarak hesaplanır. Doğruluk çok yaygın kullanılan bir ölçümdür, ancak çoğu problemde bir modelin performansını doğruluk kullanarak değerlendirmek yeterli değildir. Örneğin, veri kümesi dengesiz ise bir modelin doğruluğu yanıltıcı bir ölçüm olabilir.

Bu nedenle, model değerlendirmesi yaparken aşağıdaki ek metrikleri hesapladık:

$$\begin{aligned} \text{Kesinlik} (P = \text{Precision}) &= \frac{TP}{TP + FP} \\ \text{Duyarlılık} (Recall = \text{Sensitivity}) &= \frac{TP}{TP + FN} \\ \text{F1 Skoru} &= \frac{2PR}{P + R} \end{aligned} \quad (6.2)$$

Bu metriklerin dışında ROC (karar değerlendirme grafiği) eğrisi de sınıflandırmanın başarısı hakkında bilgiler içermektedir. ROC eğrisini analiz ederek sistemin ne kadar iyi performans gösterdiğine dair bir gösterge elde edebiliriz. Bununla birlikte, model değerlendirmesi sırasında, model performansını tanımlamak için tek bir metriğe sahip olmak istenir. ROC eğrisinin altındaki alan böyle bir bilgi verir. Resmi olarak ifade edildiğinde AUC, bir sınıflandırıcının rastgele seçilmiş bir pozitif numuneyi rastgele seçilmiş bir negatif numunenin üzerinde sıralama olasılığıdır.

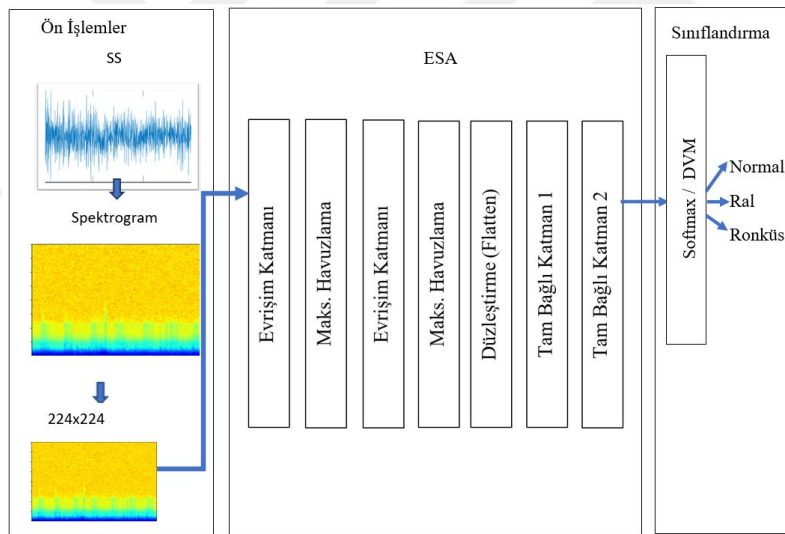


Şekil 6.5. ROC Eğrisi

Şekil 6.5, bir ROC eğrisi örneğini göstermektedir. A eğrisi, duyarlılığı (Gerçek Pozitif Oranı) ve özgüllüğü (Yanlış Pozitif Oranı) 1.0 olan ideal bir sistemi temsil eder; C eğrisi, eşit duyarlılık ve özgüllük ile rastgele seçimler yapan bir sistemi temsil ederken, B tipik bir ROC eğrisini gösterir. ROC eğrisi sol üst köşeye ne kadar yakınsa sistemin o kadar iyi olduğunu ifade eder. Bir ROC eğrisi C eğrisine yakınsa, sistem performansı rastgele bir sınıflandırıcının performansına doğru eğilim göstermiş demektir bu da performansın kötülüğüne işaret eder.

6.3 ESA ile Sınıflandırma

Solunum seslerinin sınıflandırılması amacıyla; görüntü sınıflandırmada sıklıkla kullanılan bir yöntem olan ESA mimarileri kullanılmıştır. Öncelikle sınıflandırma işlemi temel bir ESA modeli ile yapılmıştır. 2 katmanlı ESA mimarisinin solunum seslerini sınıflandırmadaki başarısı bilindiği için [24, 28], 2 katmanlı ESA mimarisi kullanılmıştır.



Şekil 6.6. Solunum Seslerinin ESA ile Softmax veya DVM Kullanarak Sınıflandırılma Mimarisi

Şekil 6.6’da görüldüğü gibi giriş 224 x 224 spektrogram görüntüsüdür. ESA mimarisinde iki evrişim katmanı ve havuzlama katmanından sonra düzleştirme ve tam bağlı katmanlar kullanılır. Yapılan deneylerde ortalama havuzlama katmanı da kullanıldı ancak sonuçların maksimum havuzlama ile daha iyi olması nedeniyle maksimum havuzlama katmanı tercih edildi. Kullanılan 2-katmanlı ESA mimarisinin katmanları aşağıdaki gibidir:

1. Giriş: 224 X 224 görüntüler

2. EVRİŞİM KATMANI: filtre sayısı: 64 ,filtre boyu: 3 x 3, aktivasyon fonksiyonu: ReLU
3. HAVUZLAMA KATMANI (POOLING LAYER): Havuzlama türü: Maksimum Havuzlama , filtre boyu: 2 x 2
4. EVRİŞİM KATMANI: filtre sayısı: 32, filtre boyu: 3 x 3, aktivasyon fonksiyonu: ReLU
5. HAVUZLAMA KATMANI (POOLING LAYER): Havuzlama türü: Maksimum Havuzlama , filtre boyu: 2 x 2
6. DÜZLEŞTİRME KATMANI (FLATTEN LAYER):
7. TAM BAĞLI KATMAN (FULLY CONNECTED LAYER): Nöron sayısı: 100 , (Hidden Neurons)
8. TAM BAĞLI KATMAN (FULLY CONNECTED LAYER): 3 Sınıf

solunum seslerinin 2 katmanlı ESA mimarisi ile sınıflandırılması iki farklı durum için test edilmiştir. Öncelikle ilk olarak klasik ESA'da olduğu gibi ESA-Softmax mimarisi ile sınıflandırma yapılmıştır. ESA mimarisinde softmax fonksiyonu ve çapraz entropi fonksiyonu kullanılarak bir sınıflandırma yapılmıştır. Sınıflandırma işlemi üç farklı veri kümesi (Çizelge 6.1) için de gerçekleştirilmiş olup elde edilen başarı oranları Çizelge 6.2'de gösterilmiştir.

Çizelge 6.2. ESA-Softmax Mimarisi Sınıflandırma Başarıları
(Gabor Spektrogramı)

Veri Kümesi	Minimum Doğruluk	Maksimum Doğruluk	Ortalama Doğruluk	Standart Sapma
294 adet veri	% 41	% 69	% 58.6	9.6
1470 adet artırılmış veri	% 35	% 75	% 68.8	11.9
294 adet gürültüsüzleştirilmiş veri	% 42	% 63	% 55.7	7.4

Çizelge 6.2'te 2 katmanlı ESA mimarisi ile solunum seslerinin Gabor spektrogram görüntüleri kullanıldığında elde edilen sonuçlar verilmiştir. 10 kat çapraz doğrulama ile elde edilen maksimum doğruluk %75 olsa da elde edilen minimum doğruluk oranının %35'te kalması istenen bir durum değildir.

Çizelge 6.3. ESA-Softmax Sınıflandırma Başarıları
(Mel Spektrogramı)

Veri Kümesi	Minimum Duyarlılık	Maksimum Duyarlılık	Ortalama Duyarlılık	Standart Sapma
294 adet veri	% 51	% 59	% 56	2.26
1470 adet artırılmış veri	% 35	% 77	% 67.9	17.36
294 adet gürültüsüzleştirilmiş veri	% 41	% 59	% 52.3	5.12

Çizelge 6.3’de 2 katmanlı ESA mimarisi ile solunum seslerinin Mel spektrogram görüntüleri kullanıldığında elde edilen sonuçlar verilmiştir. Burada da benzer şekilde elde edilen minimum doğruluklar oldukça düşüktür. Daha sağlam sonuçlar elde edebilmek adına deneylere devam edilmiştir ve ESA-DVM mimarisi de sınıflandırma için kullanılmıştır. ESA mimarisinde kayıp fonksiyonu olarak menteşe kaybı (hinge loss) kullanılarak DVM ile sınıflandırma yapılmış ve elde edilen başarı oranları Çizelge 6.4’te gösterilmiştir.

Çizelge 6.4. ESA-DVM Sınıflandırma Başarıları
(Gabor Spektrogramı)

Veri Kümesi	Minimum Doğruluk	Maksimum Doğruluk	Ortalama Doğruluk	Standart Sapma
294 adet veri	% 49	% 59	% 55.7	3.33
1470 adet artırılmış veri	% 64	% 69	% 67	1.69
294 adet gürültüsüzleştirilmiş veri	% 44	% 49	% 46.8	1.22

Çizelge 6.4’te solunum seslerinin Gabor spektrogram görüntüleri bu defa ESA-DVM mimarisi ile sınıflandırıldığında elde edilen sonuçlar verilmiştir. Elde edilen maksimum doğruluk %69 olmuştur.

Giriş olarak Mel Spektrogram görüntülerinin kullanıldığı durumda ESA-DVM mimarisi ile elde edilen sonuçlar ise Çizelge 6.5’teki gibidir.

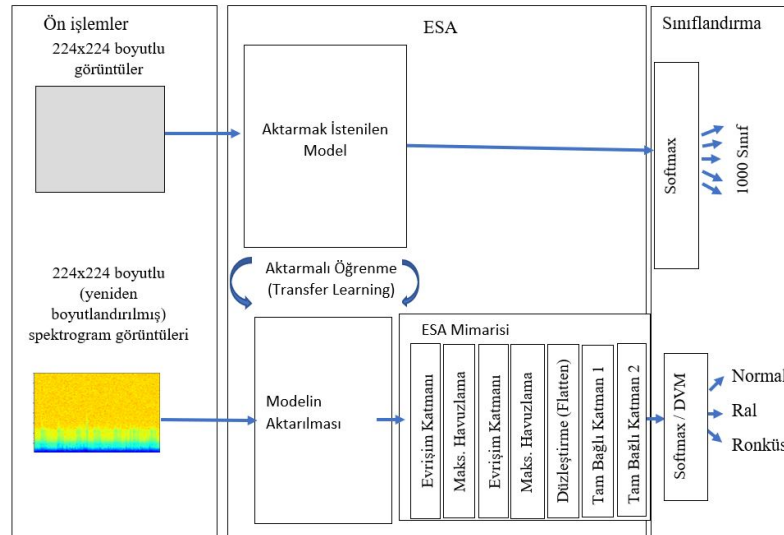
Çizelge 6.5. ESA-DVM Mimarisi Sınıflandırma Başarıları
(Mel Spektrogramı)

Veri Kümesi	Minimum Doğruluk	Maksimum Doğruluk	Ortalama Doğruluk	Standart Sapma
294 adet veri	% 44	% 59	% 50.4	4.24
1470 adet artırılmış veri	% 59	% 67	% 63.1	2.20
294 adet gürültüsüzleştirilmiş veri	% 53	% 61	% 57.5	2.46

Çizelge 6.5’te görüldüğü gibi ESA-DVM mimarisi ile yapılan test sonucunda elde edilen maksimum doğruluk %67 olmuştur. ESA-Softmax ve ESA-DVM mimarileri ile elde edilen test başarıları karşılaştırıldığında ESA-DVM ile yapılan sınıflandırma işlemlerinde maksimum doğruluk oranı daha az olsa da minimum doğruluk oranlarının (ESA-Softmax’ e göre) arttığı söylenebilir. Daha sağlam ve daha daha yüksek test başarıları elde edebilmek için deneylere devam edilmiştir.

6.4 Aktarmalı Öğrenme ve ESA ile Sınıflandırma

Aktarmalı öğrenme herhangi bir görev için geliştirilen bir modelin farklı bir görev için yeniden kullanılmasıdır. Bir problem üzerinde daha önce eğitilmiş ve başarı sağlamış modeller öğrenme aktarımı ile kendi problemimize aktarılabilir. Solunum seslerini sınıflandırma problemimize ImageNet yarışmasında 1000 sınıf için hazırlanan VGG16 modeli aktararak deneyler yapılmıştır. 1000 sınıf için hazırlanan VGG16 modeli üç sınıf için transfer edilirken, VGG16’nın son üç tam bağlantılı katmanı alınmamıştır çünkü bu son üç katman 1000 sınıf için ayarlanmıştır ve yeni bir çıktı katmanı ekleyebilmek için bu katmanlar kullanılmaz. Bu şekilde yeni katmanların veya yeni bir mimarinin birleştirilmesi mümkün olmaktadır. VGG16 modeli ESA modeli ile Şekil 6.7’deki gibi birleştirilmiştir.



Şekil 6.7. Öğrenim Aktarımı Kullanılarak Solunum Seslerinin Sınıflandırılma Mimarisi

Şekil 6.7’de görüldüğü gibi aktarılmak istenilen model (VGG16) aktarmalı öğrenmeyle solunum sesi spektrogram görüntülerini sınıflandırmak için transfer edilmiş ve ardından

Şekil 6.6'daki 2 katmanlı ESA mimarisi ile birleştirilmiş ve yeni bir mimari elde edilmiştir. Elde edilen bu mimari (VGG16-ESA), sınıflandırma performansı açısından hem Softmax hem de DVM için ayrı ayrı test edilmiştir. 3 sınıfa ait 294 veri ile ESA-Softmax, ESA-DVM, VGG16-ESA-Softmax, VGG16-ESA-DVM olmak üzere dört farklı deney yapılmıştır. Modeller, 10 kat çapraz doğrulama kullanılarak eğitilip test edilmiştir. Hata matrisleri ve performans değerlendirme metrikleri, her bir test seti için bağımsız olarak elde edilmiştir. Test sonuçları Çizelge 6.6'da belirtilmiştir.

Çizelge 6.6. VGG16-ESA-Softmax ve VGG16-ESA-DVM Mimarilerinin Sınıflandırma Başarıları (Gabor Spektrogramı)

Veri Kümesi	Yöntem	Minimum Doğruluk	Maksimum Doğruluk	Ortalama Doğruluk	Standart Sapma
294 adet veri	VGG16-ESA-Softmax	% 71	% 78	% 74.4	1.95
294 adet veri	VGG16-ESA-DVM	% 69	% 76	% 73.6	2.27
1470 adet artırılmış veri	VGG16-ESA-Softmax	% 70	% 73	% 71.7	1.05
1470 adet artırılmış veri	VGG16-ESA-DVM	% 70	% 74	% 71.6	1.17
294 adet gürültüsüzleştirilmiş veri	VGG16-ESA-Softmax	% 76	% 80	% 78.8	1.39
294 adet gürültüsüzleştirilmiş veri	VGG16-ESA-DVM	% 76	% 83	% 79.8	1.81

Çizelge 6.6'da görüldüğü gibi en yüksek sınıflandırma başarısı VGG16-ESA-DVM mimarisi ile %83 olarak elde edilmiştir.

Çizelge 6.7. VGG16-ESA-Softmax ve VGG16-ESA-DVM Modelleri ile Elde Edilen Sınıflandırma Başarıları (Mel Spektrogramı)

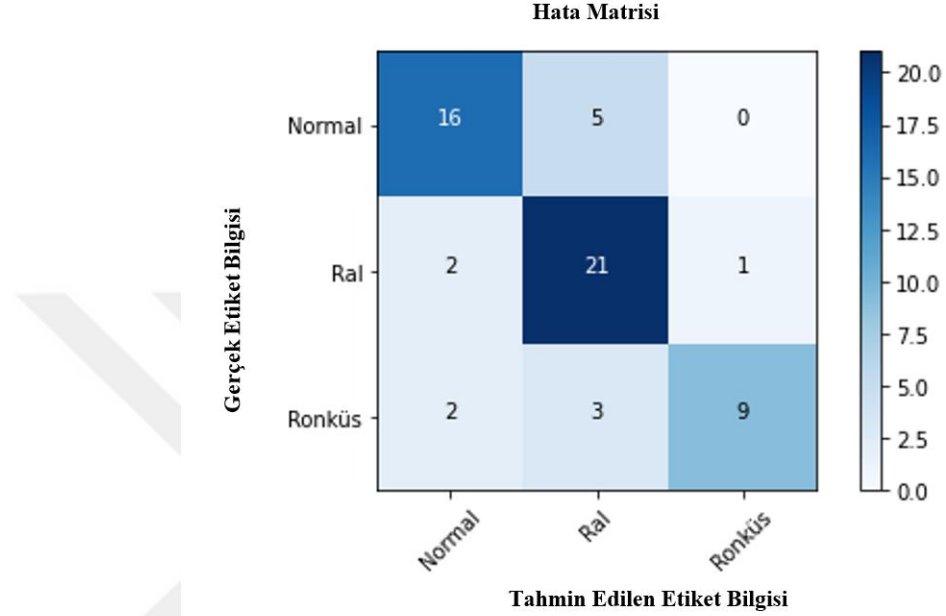
Veri Kümesi	Yöntem	Minimum Doğruluk	Maksimum Doğruluk	Ortalama Doğruluk	Standart Sapma
294 adet veri	VGG16-ESA-Softmax	66%	71%	69.5%	1.77
294 adet veri	VGG16-ESA-DVM	68%	73%	71.7%	2.67
1470 adet artırılmış veri	VGG16-ESA-Softmax	72%	75%	73.3%	1.22
1470 adet artırılmış veri	VGG16-ESA-DVM	73%	76%	74.2%	1.06
294 adet gürültüsüzleştirilmiş veri	VGG16-ESA-Softmax	68%	75%	71.7%	2.83
294 adet gürültüsüzleştirilmiş veri	VGG16-ESA-DVM	66%	71%	69.4%	1.64

Çizelge 6.7'de görüldüğü gibi en yüksek sınıflandırma başarısı 1470 adet veri ve VGG16-ESA-DVM mimarisi ile %76 olarak elde edilmiştir. Ancak bu başarı Gabor spektrogramı ile elde edilen başarılarından (Çizelge 6.6) düşüktür.

VGG16-ESA-Softmax ve VGG16-ESA-DVM mimarileri ile yapılan sınıflandırma işlemlerinin sonuçları 2 katmanlı ESA mimarilerine göre oldukça başarılı olmuştur ve daha sağlam sonuçlar elde edilmiştir. Bu nedenle bu sonuçlara daha detaylı bakmak gerekir (Bölüm 6.4.1- 6.4.2).

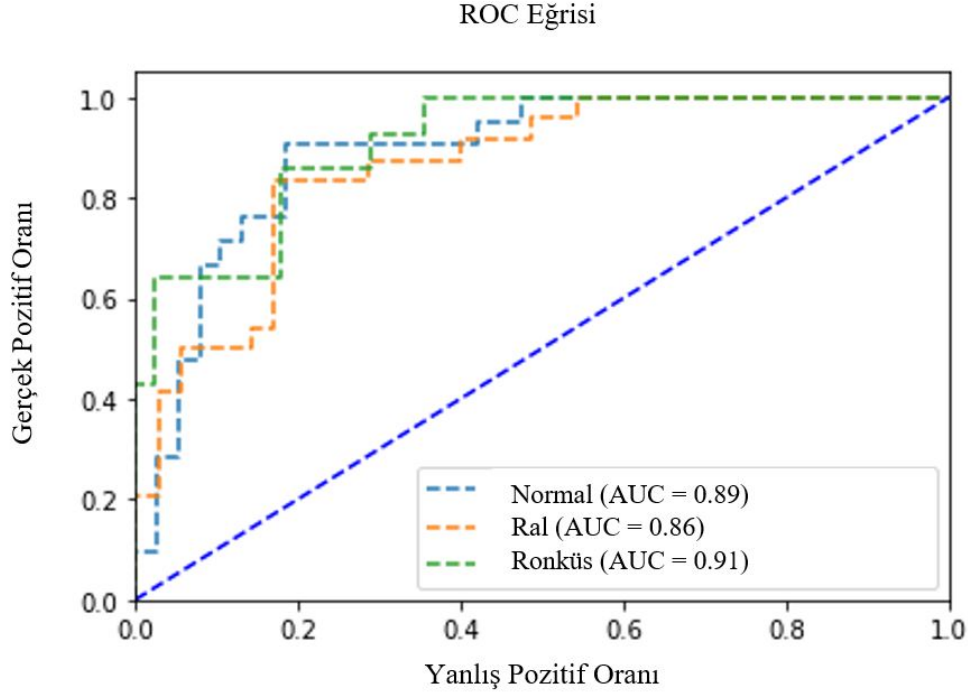
6.4.1 VGG16-ESA-Softmax Mimarisi ile Elde Edilen Sonuçlar

294 adet solunum sesinin Gabor dönüşümü ile elde edilen spektrogram görüntülerini giriş olarak alan VGG16-ESA-Softmax mimarisinin hata matrisi ve Roc eğrileri sırasıyla Şekil 6.8 ve Şekil 6.9’da gösterilmiştir.



Şekil 6.8. VGG16-ESA-Softmax Mimarisi Hata Matrisi (294 Adet Veri)

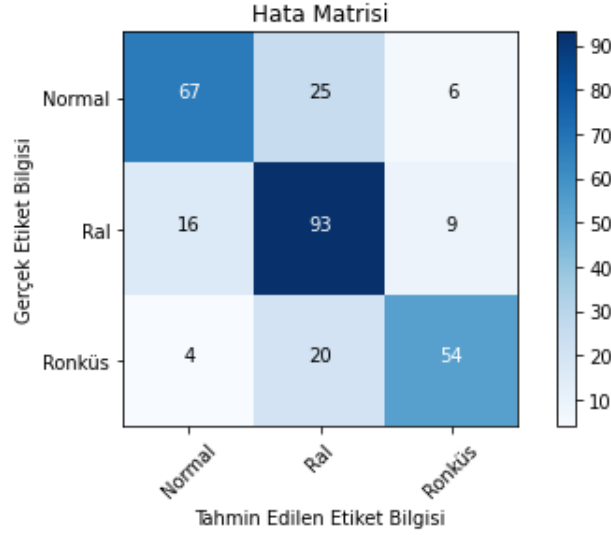
Şekil 6.8’de test grubundaki seslerin hata matrisleri görselleştirilmiştir. Sınıflandırma doğruluğu %78 olurken, Normal seslerin 16’sı (%76), Ral seslerinin 21’i (%88) ve Ronküs seslerinin 9’u (%64) doğru şekilde sınıflandırılmıştır.



Şekil 6.9. VGG16-ESA-Softmax Mimarisi ROC Eğrisi (294 Adet Veri)

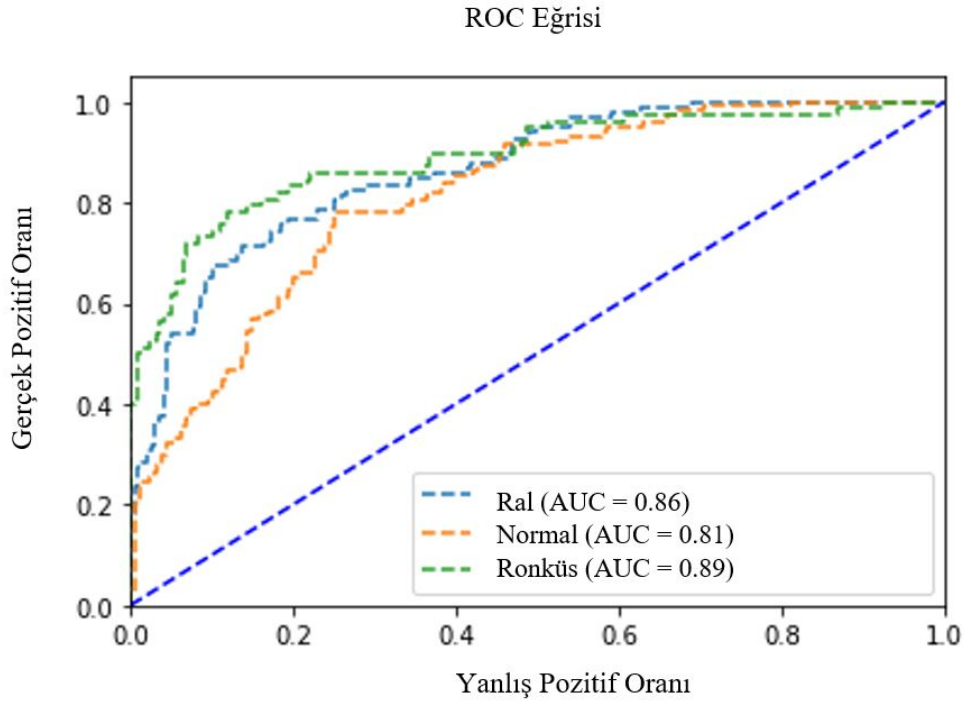
VGG16-ESA-Softmax Modeli ile 294 adet solunum sesi için elde edilen sınıflandırma başarısı (ROC-AUC Skoru) Şekil 6.9’da görüldüğü gibi Normal sesi için %89, Ral sesi için %86 ve Ronküs sesi için %91 olmuştur. Genel olarak ROC-AUC skoru %88.7 ve MCC katsayısı da 0.66 olarak elde edilmiştir.

1470 adet solunum sesinin Gabor dönüşümü ile elde edilen spektrogram görüntülerini giriş olarak alan VGG16-ESA-Softmax mimarisinin hata matrisi ve Roc eğrileri sırasıyla Şekil 6.10 ve Şekil 6.11’de gösterilmiştir.



Şekil 6.10. VGG16-ESA-Softmax Mimarisi Hata Matrisi (1470 Adet Veri)

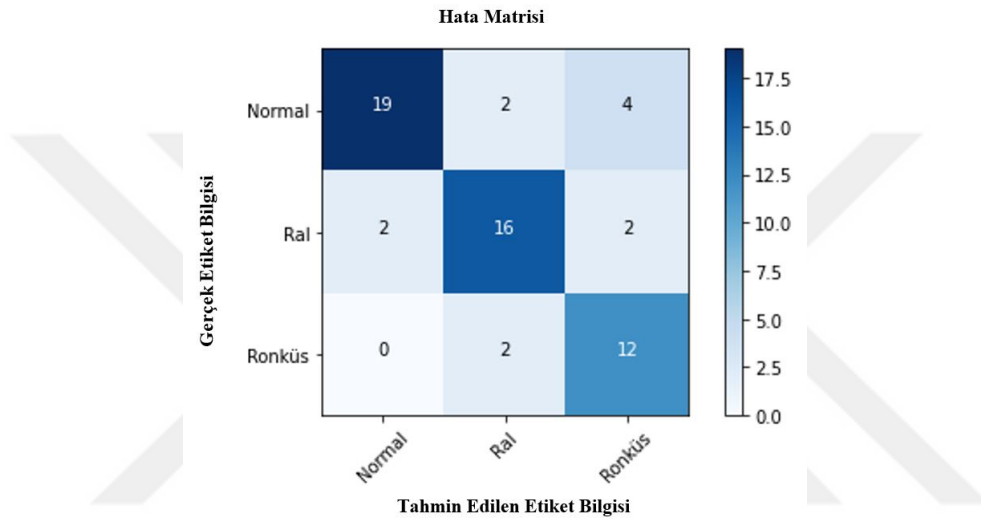
1470 adet veri için VGG16-ESA-Softmax mimarisi ile elde edilen maksimum sınıflandırma doğruluğu %73 olurken, Şekil 6.10’da görüldüğü gibi Normal seslerin 67’si (%68), Ral seslerinin 93’ü (%79) ve Ronküs seslerinin 54’ü (%69) doğru şekilde sınıflandırılmıştır.



Şekil 6.11. VGG16-ESA-Softmax Mimarisi ROC Eğrisi (1470 Adet Veri)

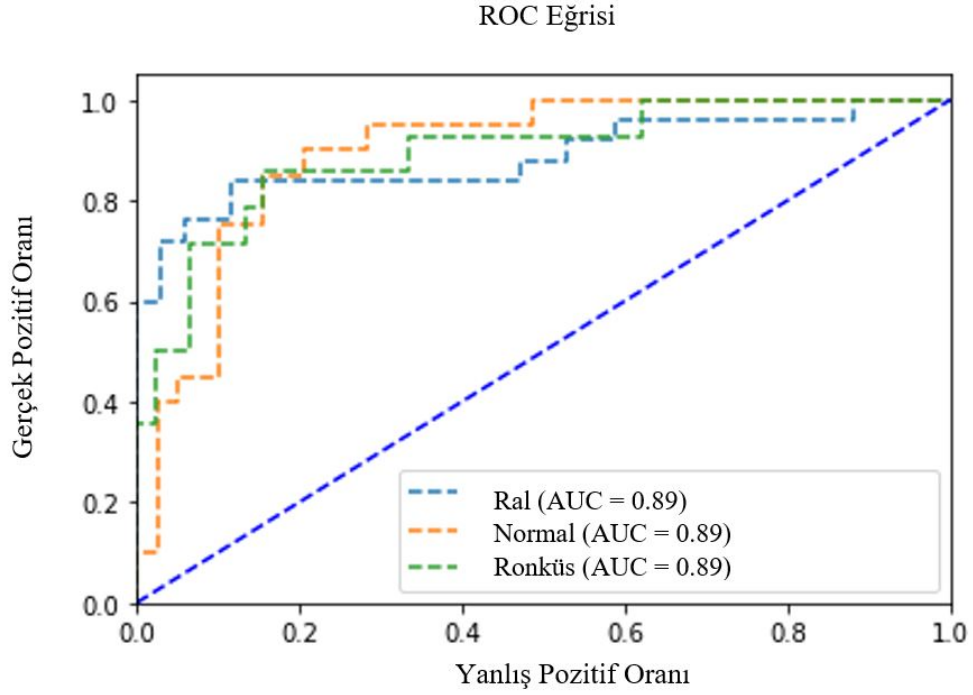
VGG16-ESA-Softmax mimarisi ile veri artırımı ile elde edilen 1470 adet solunum sesi için elde edilen sınıflandırma başarısı (ROC-AUC Skoru) Şekil 6.11’den da görüldüğü gibi Normal sesi için %86, Ral sesi için %81 ve Ronküs sesi için %89 olmuştur. Genel olarak ROC-AUC skoru % 85.6 ve MCC katsayısı da 0.58 olarak elde edilmiştir.

294 adet gürültüsüzleştirilmiş solunum sesinin Gabor dönüşümü ile elde edilen spektrogram görüntülerini giriş olarak alan VGG16-ESA-Softmax mimarisinin hata matrisi ve Roc eğrileri sırasıyla Şekil 6.12 ve Şekil 6.13’te gösterilmiştir.



Şekil 6.12. VGG16-ESA-Softmax Mimarisi Hata Matrisi (Gürültüsüzleştirilmiş Veriler)

Gürültüsüzleştirilmiş 294 adet solunum sesinin VGG16-ESA-Softmax mimarisi ile sınıflandırılmasından elde edilen maksimum sınıflandırma doğruluğu %80 olurken, Şekil 6.12’de görüldüğü gibi Normal seslerin 19’u (%76), Ral seslerinin 16’sı (%80) ve Ronküs seslerinin 12’si (%86) doğru şekilde sınıflandırılmıştır.

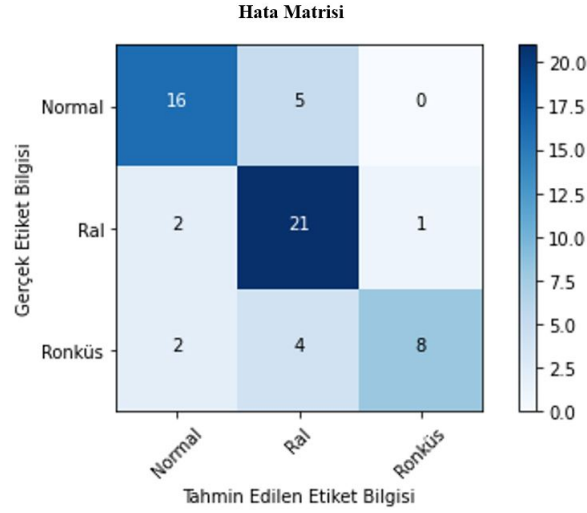


Şekil 6.13. VGG16-ESA-Softmax Mimarisi ROC Eğrisi (Gürültüsüzleştirilmiş Veriler)

VGG16-ESA-Softmax Modeli ile gürültüsüzleştirilmiş 294 adet solunum sesi için elde edilen sınıflandırma başarısı (ROC-AUC Skoru) Şekil 6.13'te görüldüğü gibi Normal sesi için %89, Ral sesi için %89 ve Ronküs sesi için %89 olmuştur. Genel olarak ROC-AUC skoru % 89.1 ve MCC katsayısı da 0.69 olarak elde edilmiştir.

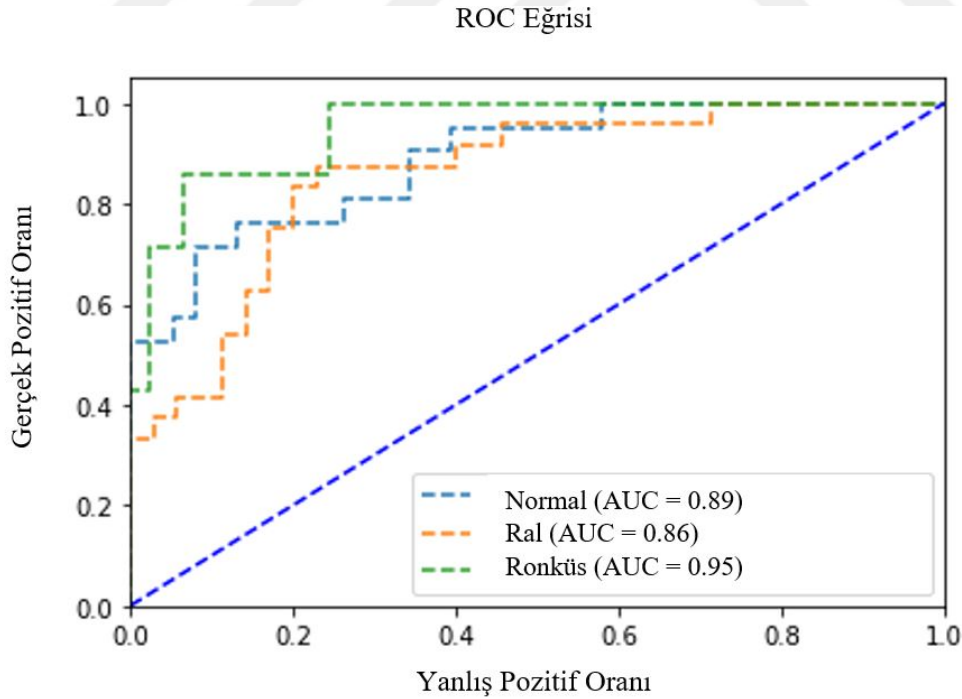
6.4.2 VGG16-ESA-DVM Mimarisi ile Elde Edilen Sonuçlar

294 adet solunum sesinin Gabor dönüşümü ile elde edilen spektrogram görüntülerini giriş olarak alan VGG16-ESA-DVM mimarisinin hata matrisi ve Roc eğrileri sırasıyla Şekil 6.14 ve Şekil 6.15'te gösterilmiştir.



Şekil 6.14. VGG16-ESA-DVM Mimarisi Hata Matrisi (294 Adet Veri)

294 adet solunum sesinin VGG16-ESA-DVM mimarisi ile sınıflandırılmasından elde edilen maksimum sınıflandırma doğruluğu %76 olurken, Şekil 6.14'te görüldüğü gibi Normal seslerin 16'sı (%76), Ral seslerinin 21'i (%88) ve Ronküs seslerinin 8'i (%57) doğru şekilde sınıflandırılmıştır.

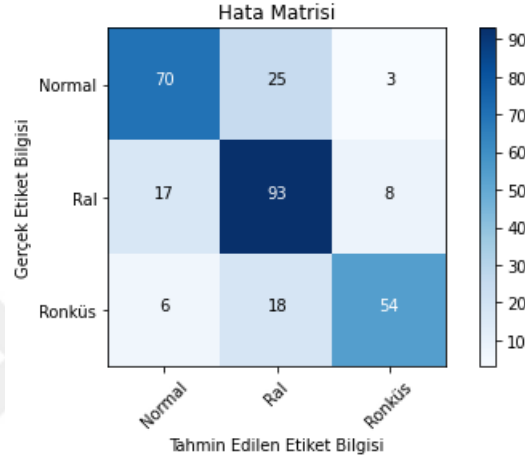


Şekil 6.15. VGG16-ESA-DVM Mimarisi ROC Eğrisi (294 Adet Veri)

VGG16-ESA-DVM Modeli ile 294 adet solunum sesi için elde edilen sınıflandırma başa-

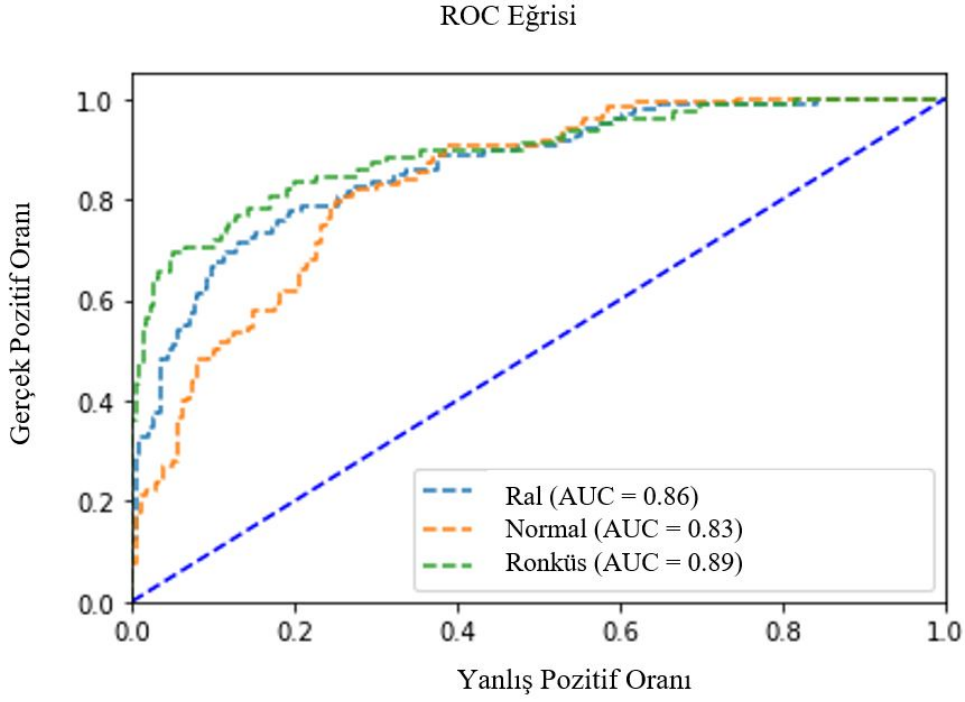
rısı (ROC-AUC Skoru) Şekil 6.15’te görüldüğü gibi Normal solunum sesi için %89, Ral solunum sesi için %86 ve Ronküs solunum sesi için %95 olmuştur. Genel olarak ROC-AUC skoru % 89.8 ve MCC katsayısı da 0.63 olarak elde edilmiştir.

1470 gürültüsüzleştirilmiş solunum sesinin Gabor dönüşümü ile elde edilen spektrogram görüntülerini giriş olarak alan VGG16-ESA-DVM mimarisinin hata matrisi ve Roc eğri-leri sırasıyla Şekil 6.16 ve Şekil 6.17’de gösterilmiştir.



Şekil 6.16. VGG16-ESA-DVM Mimarisi Hata Matrisi (1470 Adet Veri)

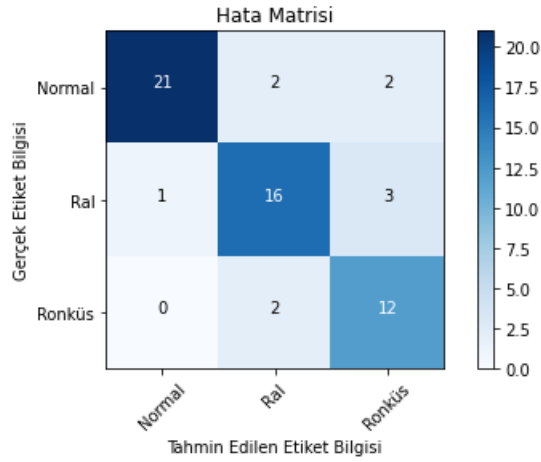
1470 adet veri ile solunum sesinin VGG16-ESA-DVM mimarisi ile sınıflandırılmasından elde edilen maksimum sınıflandırma doğruluğu %74 olurken, Şekil 6.16’da görüldüğü gibi Normal seslerin 70’i (%71), Ral seslerinin 93’ü (%79) ve Ronküs seslerinin 54’ü (%69) doğru şekilde sınıflandırılmıştır.



Şekil 6.17. VGG16-ESA-DVM Mimarisi ROC Eğrisi (1470 Adet Veri)

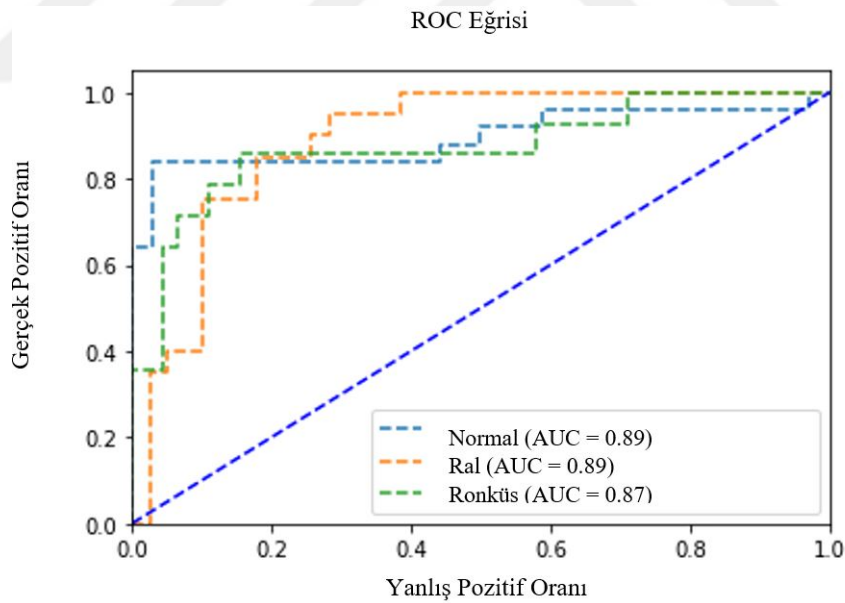
VGG16-ESA-DVM Modeli ile veri artırımı ile elde edilen 1470 adet solunum sesi için elde edilen sınıflandırma başarısı (ROC-AUC Skoru) Şekil 6.17’de görüldüğü gibi Normal sesi için %86, Ral sesi için %83 ve Ronküs sesi için %89 olmuştur. Genel olarak ROC-AUC skoru % 86.3 ve MCC katsayısı da 0.60 olarak elde edilmiştir.

En yüksek başarı 294 adet gürültüsüzleştirilmiş veri ile VGG16-ESA-DVM modelinden elde edilmiştir (Çizelge 6.6). Doğruluk oranı ise 83% olarak elde edilmiştir. 294 gürültüsüzleştirilmiş solunum sesinin Gabor dönüşümü ile elde edilen spektrogram görüntülerini giriş olarak alan VGG16-ESA-DVM mimarisinin hata matrisi ve Roc eğrileri sırasıyla Şekil 6.18 ve Şekil 6.19’da gösterilmiştir.



Şekil 6.18. VGG16-ESA-DVM Mimarisi Hata Matrisi (Gürültüsüzleştirilmiş Veriler)

Gürültüsüzleştirilmiş 294 adet solunum sesinin VGG16-ESA-DVM mimarisi ile sınıflandırılmasından elde edilen maksimum sınıflandırma doğruluğu %83 olurken, Şekil 6.18’de görüldüğü gibi Normal seslerin 21’i (%84), Ral seslerinin 16’sı (%80) ve Ronküs seslerinin 12’si (%86) doğru şekilde sınıflandırılmıştır.



Şekil 6.19. VGG16-ESA-DVM Mimarisi ROC Eğrisi (Gürültüsüzleştirilmiş Veriler)

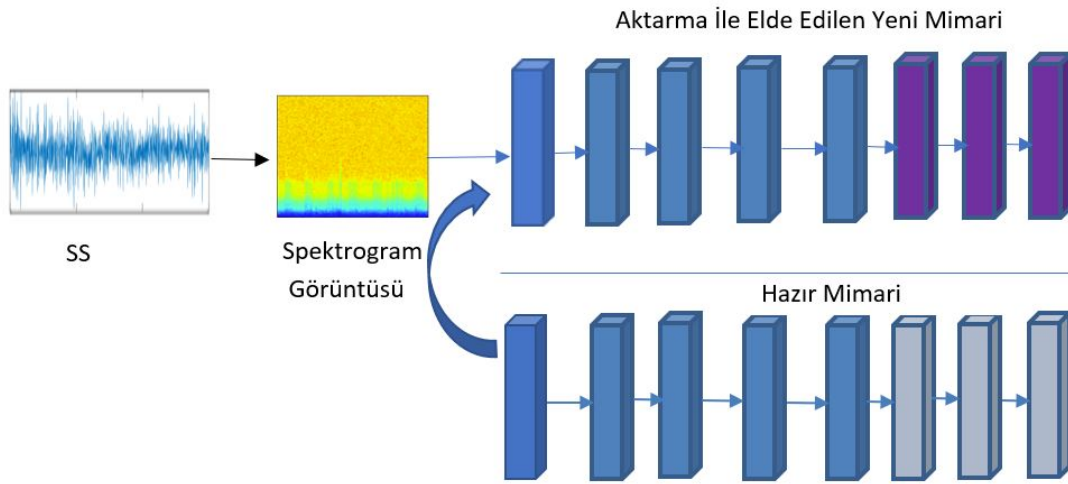
VGG16-ESA-DVM mimarisi ile gürültüsüzleştirilmiş 294 adet solunum sesi için elde edilen sınıflandırma başarısı (ROC-AUC Skoru) Şekil 6.19’da görüldüğü gibi Normal sesi için %89, Ral sesi için %89 ve Ronküs sesi için %87 olmuştur. Genel olarak ROC-AUC

skoru % 88.4 ve MCC katsayısı da 0.74 olarak elde edilmiştir. Ayrıca kesinlik= %82, duyarlılık= %83 ve F1 skoru= %82 olarak elde edilmiştir.

Giriş görüntülerinin, Gabor Spektrogramı (Çizelge 6.6) ve Mel Spektrogramı (Çizelge 6.7) olduğu durumlar kıyaslandığında açıkça Gabor Spektrogramı'ndan elde edilen sınıflandırma başarılarının daha yüksek olduğu görülmüştür. Elde edilen ortalama sınıflandırma doğrulukları Mel Spektrogram görüntüleri kullanıldığında %69 – %74 aralığında iken Gabor görüntüleri kullanıldığında %71 – %79'lara ulaşmıştır. Ayrıca Çizelge 6.6'da görüldüğü gibi tüm yöntemler arasında en iyi sınıflandırma performansı VGG16-ESA-DVM modeli ile elde edilmiştir. Normal, Ral ve Ronküs seslerinin 10 kat çapraz doğrulama ile elde edilen maksimum sınıflandırma doğruluğu %83 olmuştur.

6.5 Başarısı Kanıtlanmış ESA Modellerinin Aktarmalı Öğrenme ile Solunum Sesi Sınıflandırmada Kullanılması

Gabor Spektrogram görüntülerinin VGG16-ESA-DVM mimarisi ile ulaştığı sınıflandırma başarısını kıyaslayabilmek adına, başarısı kanıtlanmış mimarilerle de deneyler yapılmış ve sınıflandırma başarıları incelenmiştir. VGG16, ResNet101, DenseNet201 ve InceptionV3 modellerini problemimizde kullanırken son üç katman (Fully Connected Layer) dışında kalan kısımlar aynen alınmıştır, yani 14 milyondan fazla veri kümesinden oluşan ImageNet verileri ile eğitilmiş modelin ağırlıkları kullanılmıştır. Bu eğitilen katmanların çıkışı sırasıyla Düzleştirme, Relu ve Softmax işlemlerinden geçirilerek sınıflandırma gerçekleştirilmiştir. Sınıflandırma yöntemi olarak Aktarmalı Öğrenme işleminde, eğitim aşamasında ESA mimarilerinin (VGG16, DenseNet201, InceptionV3 ve ResNet101) ağırlıkları ile başlanıp eğitim süresince parametreler güncellenmektedir. Kayıp fonksiyonu olarak crossentropy, eniyileme algoritması (optimizer) olarak Adam algoritması seçilmiştir. Aktarmalı Öğrenme ile solunum seslerinin sınıflandırılmasında kullandığımız model Şekil 6.20'de gösterilmiştir.



Şekil 6.20. Aktarmalı Öğrenme ile Solunum Seslerinin Sınıflandırılması

Şekil 6.20’de görüldüğü gibi solunum sesleri, spektrogram görüntüleri elde edilerek önceden eğitilmiş modellerden aktarmalı öğrenme yardımıyla Normal, Ral ve Ronküs seslerinin sınıflandırılması probleminde kullanılmıştır. Çizelge 6.8’de öğrenme aktarımı ile 3 farklı solunum ses grubunun sınıflandırma işleminden elde edilen test sonuçları belirtilmiştir.

Çizelge 6.8. Aktarmalı Öğrenme ile Elde Edilen Sınıflandırma Başarıları

Veri Kümesi	Yöntem	Minimum Doğruluk	Maksimum Doğruluk	Ortalama Doğruluk	Standart Sapma
294 adet veri	VGG16	% 63	% 71	% 67.3	2.16
294 adet veri	ResNet101	% 68	% 73	% 70.5	1.71
294 adet veri	DenseNet201	% 69	% 75	% 73.2	2.39
294 adet veri	InceptionV3	% 59	% 63	% 60.8	1.56
1470 adet artırılmış veri	VGG16	% 65	% 68	% 66.1	1.10
1470 adet artırılmış veri	ResNet101	% 66	% 69	% 66.9	0.99
1470 adet artırılmış veri	DenseNet201	% 64	% 67	% 65.8	1.22
1470 adet artırılmış veri	InceptionV3	% 51	% 53	% 51.7	0.67
294 adet gürültüsüzleştirilmiş veri	VGG16	% 71	% 78	% 74.6	2.63
294 adet gürültüsüzleştirilmiş veri	ResNet101	% 66	% 76	% 70.3	3.16
294 adet gürültüsüzleştirilmiş veri	DenseNet201	% 71	% 75	% 73.8	1.68
294 adet gürültüsüzleştirilmiş veri	InceptionV3	% 66	% 69	% 67.8	1.03

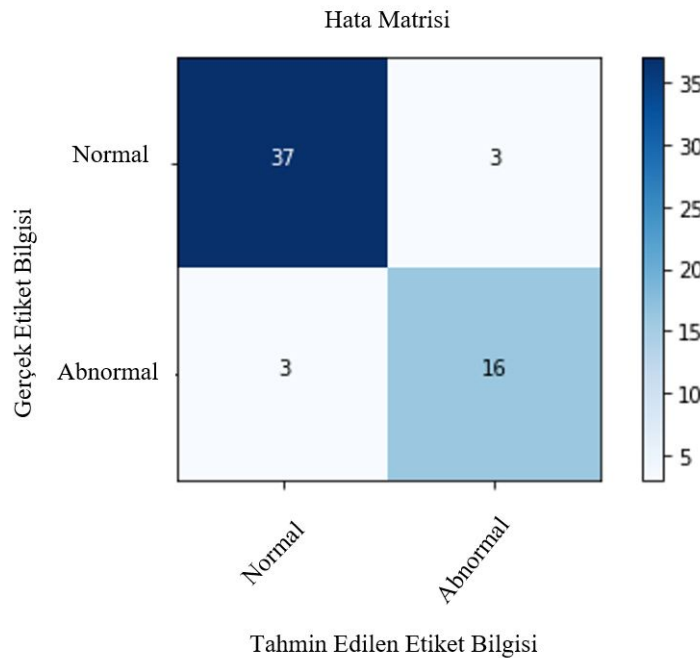
Dört farklı ESA modeli (VGG16, DenseNet201, InceptionV3, ResNet101) kullanılarak öğrenme aktarımı yapıldığında en yüksek doğruluk oranının 294 adet solunum sesi veri kümesi ile VGG16 mimarisi kullanılarak elde edildiği gözlemlenmiştir (Çizelge 6.8). Sınıflandırmada VGG16 mimarisi ile gerçekleştirilen maksimum doğruluk %78 gibi yüksek

bir değere ulaşsa da önerilen mimari (VGG16-ESA-DVM) ile solunum sesleri gürültüsüzleştirilerek elde edilen verilerle yapılan sınıflandırmada %83 ile daha yüksek sınıflandırma başarısı sağlamıştır.

6.5.1 İki Sınıf İçin VGG16-ESA-DVM Modeli ile Elde Edilen Sonuçlar

Bu bölüme kadar çoklu sınıf (Normal, Ral, Ronküs) sınıflandırma işlemleri gerçekleştirildiğinde en yüksek başarının sağlandığı mimarinin VGG16-ESA-DVM mimarisi olduğu gözlemlenmiştir. Aynı mimariyi bu kez ikili sınıf (Normal, Abnormal/ Patolojik) sınıflandırma için kullanarak 294 adet gürültüsüzleştirilmiş veri kümesinden elde edilen sonuçlar gözlemlenmiştir. 294 adet solunum sesi verisinin 105'i Normal, geriye kalan 189'u Abnormal solunum seslerinden oluşmaktadır. Daha önceki deneylerde olduğu gibi veri kümesi %80 eğitim ve %20 test kümesi olacak şekilde ayrıştırılmıştır.

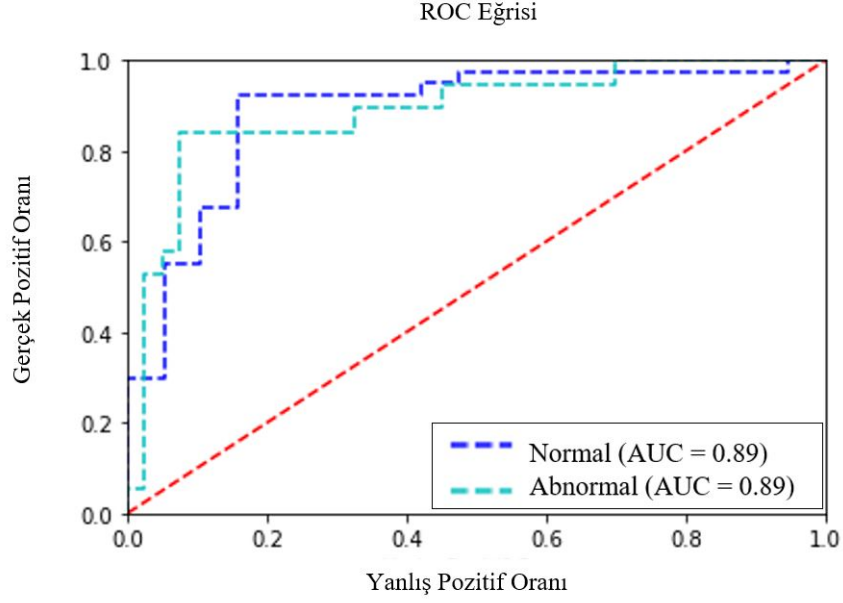
294 gürültüsüzleştirilmiş solunum sesinin Gabor dönüşümü ile elde edilen spektrogram görüntülerini giriş olarak alan VGG16-ESA-DVM mimarisinin hata matrisi ve Roc eğrileri sırasıyla Şekil 6.21 ve Şekil 6.22'de gösterilmiştir.



Şekil 6.21. VGG16-ESA-DVM Mimarisi Hata Matrisi (2 Sınıf)

10 kat çapraz doğrulama sonucunda elde edilen maksimum sınıflandırma performansı, Şe-

Şekil 6.21’de görüldüğü gibi Normal seslerin 37’si (%93) ve Abnormal seslerin 36’sı (%84) olarak elde edilmiştir.



Şekil 6.22. VGG16-ESA-DVM Mimarisi ROC Eğrisi (2 Sınıf)

VGG16-ESA-DVM Mimarisi ile gürültüsüzleştirilmiş 294 adet solunum sesi için elde edilen sınıflandırma başarısı (ROC-AUC Skoru) Şekil 6.22’de görüldüğü gibi Normal sesi için ve Abnormal sesi için %89 ve olmuştur. Genel olarak ROC-AUC skoru % 88.3 ve MCC katsayısı da 0.76 olarak elde edilmiştir.

10 kat çapraz doğrulama sonucunda elde edilen minimum, maksimum ve ortalama sınıflandırma doğrulukları Çizelge 6.9’da belirtilmiştir.

Çizelge 6.9. VGG16-ESA-DVM Mimarisi ile Elde Edilen İkili Sınıflandırma Başarıları Gabor Spektrogramı

Veri Kümesi	Minimum Doğruluk	Maksimum Doğruluk	Ortalama Doğruluk	Standart Sapma
294 adet gürültüsüzleştirilmiş veri	% 86	% 90	% 89.2	1.39

Yapılan deneyler göstermiştir ki: 294 adet klinik veri için VGG16-ESA-DVM mimarisi, verinin gürültüsüzleştirilmesi sonucunda Gabor Spektrogram görüntüleri elde edilerek 3 sınıfa ait verilerin (Normal, Ral, Ronküs) sınıflandırılmasında en yüksek performansı sağlamıştır. Aynı mimari 294 adet veri kümesi 2 sınıfa (Normal-Abnormal) ayrıldığında, 2

sınıfı sınıflandırma probleminde ise %90 sınıflandırma doğruluğu elde edilmiştir. Bu da önerilen mimarinin performansını kıyaslamak için önemlidir.

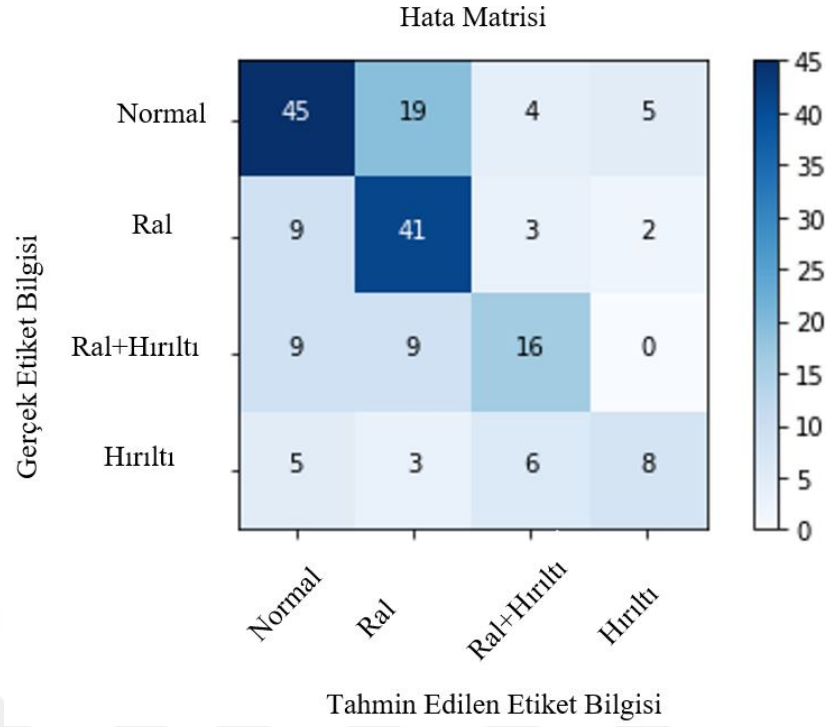
6.6 VGG16-ESA-DVM Mimarisinin Açık Kaynak Veri Kümesi Üzerindeki Performansının İncelenmesi

Kullandığımız veri kümesi ile aynı grup sesleri içeren erişilebilir bir veri kümesi bulunmamaktadır, bu yüzden elde ettiğimiz başarıyı kıyaslayabilmek amacıyla açık erişimli veri kümesi olan ICBHI veri tabanı kullanılarak aynı yöntem ve mimari ile bir sınıflandırma işlemi daha gerçekleştirilmiştir. ICBHI veri tabanı, 126 hastadan elde edilen 920 solunum sesi kaydı içeren açık kaynak bir veri tabanıdır. Ses kayıtları toplam 5.5 saatlik kayıttan oluşmaktadır birlikte her kayıt farklı sürelerde toplanmıştır (7- 9 -11 ... saniye, maks. 86 saniye). 4 ses grubuna (301 adet Normal, 298 adet Ral, 181 adet Ral-Hırıltı, 138 adet Hırıltı) ait sesleri içermektedir.

Çizelge 6.10. ICBHI Verileri ile Tez Çalışmasında Kullanılan Verilerin Genel Çerçevesi

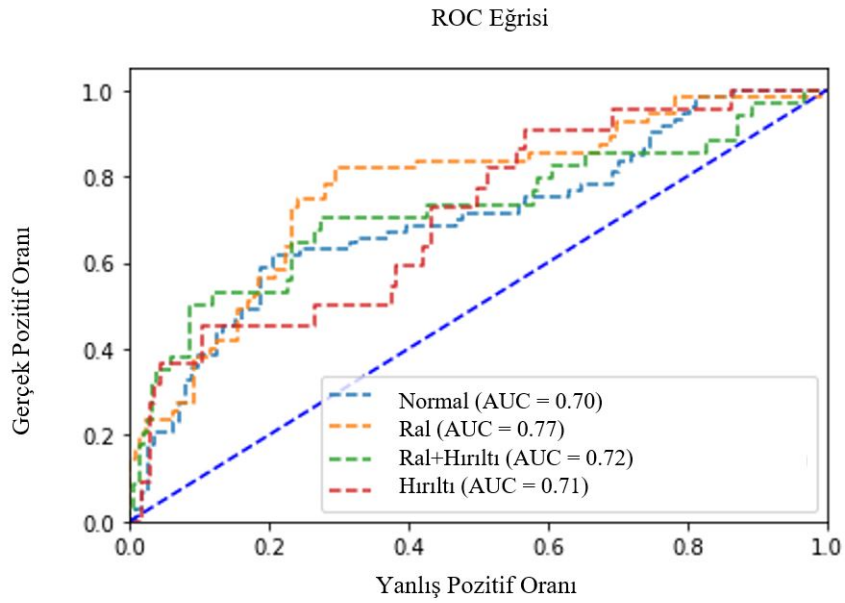
Veri Kümesi	Kayıt Sayısı	İçerdiği Etiket Bilgileri	Yapılan Sınıflandırma Grupları
ICBHI	126 kişiden toplanan 920 ses kaydı 1864 adet Ral 886 adet Hırıltı 506 adet Ral+ Hırıltı 3642 adet Normal	Ses Türüne Göre: Ral Hırıltı Ral+ Hırıltı Normal Hastalık Türüne Göre: KOA: Kronik Obstrüktif Akciğer Hastalığı ASYE: Alt Solunum Yolu Enfeksiyonu ÜSYE: Üst Solunum Yolu Enfeksiyonu	2 sınıf için: Normal- Abnormal 3 sınıf için: KOA, ASYE, ÜSYE 4 sınıf için: Ral, Hırıltı, Hırıltı+ Ral, Normal
Tez Çalışmasında Kullanılan Veri Kümesi	294 kişiden 294 ses kaydı	Normal, Ral, Ronküs	2 sınıf için: Normal- Abnormal 3 sınıf için: Normal, Ral, Ronküs

Normal, Ral ve Ronküs ses gruplarını sınıflandırmada kullandığımız VGG16-ESA-DVM mimarisi bu 4 sınıf için de uygulanmıştır ve Hata Matrisi Şekil 6.23'te belirtilmiştir.



Şekil 6.23. VGG16-ESA-DVM Mimarisi Hata Matrisi (ICBHI verileri)

4 sınıf için %60 sınıflandırma doğruluğu elde edilmiştir. Şekil 6.23'te görüldüğü gibi en yüksek oranda %75 Ral sesi doğru sınıflandırılmıştır. Hırıltı sesi ise %36 doğrulukla en düşük sınıflandırma performansına sahiptir. Bu 4 sınıf için Roc eğrisi Şekil 6.24'te gösterilmiştir.



Şekil 6.24. VGG16-ESA-DVM Mimarisi ROC Eğrisi (ICBHI verileri)

Şekil 6.24'te görüldüğü gibi bu 4 grup ses için %70'in üzerinde AUC skoru elde edilmiştir. 3 grup solunum sesi verisini sınıflandırmak için oluşturulan yöntem ve mimarinin hiçbir değişiklik yapılmadan 4 grup solunum sesi verisi içeren farklı bir veri kümesini sınıflandırmada kullanıldığında da iyi bir performans sağladığını söylemek mümkündür.



7. TARTIŞMA

Bu tezde solunum seslerini sınıflandırmaya çalışırken adım adım en iyi doğruluğa gidilmiştir. Öncelikle çalışmaya basit 2 katmanlı bir ESA mimarisi ile başlanmış daha sonra başka ESA mimarileri ile kıyaslanmış ve bir çok deney sonucunda yeni bir mimari oluşturulmuştur.

Yapılan deneyler sonucunda gürültü azaltma uygulanarak elde edilen 294 adet solunum sesine ait Spektrogram (Gabor) görüntüsünü giriş olarak kullanan VGG16-ESA-DVM mimarisi ile maksimum %83 sınıflandırma doğruluğu ve %88.4 ROC- AUC skoru elde edilmiştir. Bu çalışma ile; Normal seslerin %84, Ral seslerin %80 ve Ronküs seslerinin %86 doğrulukla sınıflandırılması sağlanmıştır. İkili sınıflandırma sonucunda ise Normal sesleri (%93) ve Abnormal sesleri (%84) doğrulukla sınıflandırılmıştır. Tez çalışmasında elde edilen sonuçların literatürdeki yerini belirleyebilmek açısından Çizelge 7.1 oluşturulmuştur.

Çizelge 7.1. Literatürde Solunum Sesi Sınıflandırması Yapılan Çalışmalar

Referans	Veri Bilgisi	Yöntem	Sınıflandırma Başarısı
[28]	1630 hasta 17930 ses kaydı 3 Sınıf Normal,Ral,Ronküs	ESA	% 76
[35]	ICBHI 920 ses kaydı 4 Sınıf 2 Sınıf	Artırılmış Karar Ağacı	% 49.98 % 85
[77]	ICBHI 920 ses kaydı 4 Sınıf	ELM-W-AE ve DVM	% 72.69
[31]	1918 ses kaydı 3 Sınıf (Ral,Hırıltı,Ronküs)	VGG16-DVM	% 86.5
Önerilen Yöntem	294 hasta 294 ses kaydı 3 Sınıf Normal,Ral,Ronküs	VGG16-ESA-DVM	% 83

Çizelge 7.1'deki referans yayınlar belirlenirken, yayınlara; sınıflandırılan veri etiket bilgisi benzerliği (Normal, Ral, Ronküs) [28], literatürdeki güncel çalışma [77], ICBHI veri kümesi ile yapılan sınıflandırma çalışmalarındaki zamanla başarı oranı artışını sunması [35, 77], hem kullanılan veri sayısı fazlalığı (1918) hem de sınıf sayısı (3) eşitliği [31] açılarından bakılarak yer verilmiştir.

Sınıflandırma probleminde niceliksel bir kıyaslamanın yapılabilmesi için kullanılan sınıfların aynı etiket bilgisini taşıması önemlidir. Bu açıdan bakıldığında literatürde aynı sayıda ve türde ses grubunu barındıran sadece bir çalışma mevcuttur. Bu çalışmada [28] Normal, Ral, Ronküs sesleri %76 doğrulukla sınıflandırılmıştır. Yapılan çalışmanın 294 adet veri ile 2 katmanlı ESA mimarisi ile elde ettiğimiz (Çizelge 6.2 ve Çizelge 6.3) maksimum %69'luk sınıflandırma doğruluğunu %76 doğruluk elde ederek aştığı söylenebilir. Bunun temel sebebi veri sayısının oldukça fazla (1630) olmasıdır. Ancak 2 katmanlı ESA mimarisi yerine bu tezde önerilen mimari ile elde edilen sonuçlara bakıldığında, bu tez çalışması %10.8 daha yüksek başarı sağlamıştır. Bunun nedeni de her ne kadar veri sayısı daha az olmuş olsa da kullanılan gürültü azaltma ve önerilen mimarinin başarısıdır.

Bu tezde kullanılan VGG16-ESA-DVM mimarisinin sağlamlığı, bir fikir vermesi açısından ICBHI veri kümesinde de denenmiştir. Bu tez çalışmasında kullanılan veri kümesi ile ICBHI veri kümesi içerdikleri kayıtların sayısı, kullanılan etiket bilgileri ve bu veri kümeleri ile oluşturulan sınıflandırma grupları açısından kıyaslama yapılabilmesi açısından Çizelge 6.10'da sunulmuştur. Bu tez kapsamında elde edilen üç sınıf için en yüksek sınıflandırma performansını gösteren mimari (VGG16-ESA-DVM) aynı şekilde ICBHI veri kümesi için test edildiğinde; 4 sınıf için genel sınıflandırma doğruluğu %60 olarak elde edilmiştir. Hırıltı sesi sınıflandırmada düşük performans (doğruluk= %36) göstermesine rağmen Ral seslerinde (doğruluk= %75) daha yüksek performans elde edilmiştir.

%60'lık sınıflandırma doğruluğu kötü görünse de yapılan çalışmalarda [35] bu doğruluğun altına (%49.98) düşülen çalışmalar da mevcuttur. 2017'den beri üzerinde çalışılan bir konu olduğu için ve bu 4 sınıf özelinde oluşturulan mimariler ve yapılan ön işlemler daha da geliştirilmiş ve en güncel olarak [77] yapılan bir çalışmada %72.69'luk sınıflandırma başarısı elde edilmiştir. Ayrıca çalışmada Aşırı Öğrenme Makinesi Oto Kodlayıcı (ELM-AE) ve altı farklı dalgacık fonksiyonunu kullanarak veri artırımı yapılmış ve sınıflandırma

aşamasında DVM kullanılmıştır.

Elde edilen sonucu yorumlamak gerekirse, Normal, Ral, Ronküs seslerinde başarı gösteren bu tezde önerilen mimarinin ICBHI veri kümesinde Normal ve Ral sınıflarına (%62 ve %75), diğer iki sınıfa (%47 ve %36) oranla daha yüksek doğrulukla sınıflandırması beklenen bir durumdur. Çünkü zaten bu tezde çalışılan ses grupları Normal ve Ral'dir. Farklı bir veri kümesi için yapılan sınıflandırma işlemi bu yeni veriler için farklı ön işlemler gerektirebilir. Dolayısıyla elde edilen başarının (%60) seviyelerini aşması zaten beklenebilir. Yine de tezde önerilen mimarinin başarısı açısından bir fikir edinilmesi açısından önemlidir. Daha fazla veri, daha farklı gruplar ve daha farklı kayıt yöntemlerini kapsayan bir çalışma henüz mevcut değildir. Son yıllarda solunum sesi sınıflandırma ile ilgili çalışmalar her geçen gün artmaktadır, ve bu eksik te literatüre katkı sağlayan çalışmalarla kısa sürede kapanmalıdır. Literatürdeki seslere ek olarak Ronküs sesinin de içinde olduğu daha büyük veri kümeleri oluşturulmalıdır. Hatta gelecekte tüm solunum sesi gruplarını içeren daha kapsamlı veri bankalarının oluşturulması ile sınıflandırma çalışmaları genişletilmelidir.

8. SONUÇ

Bu tez çalışmasında klinik ortamda toplanan 294 adet solunum ses verisinin makine öğrenme teknikleri ile sınıflandırılması üzerine çalışılmıştır. Solunum sesleri gürültü bastırma, veri artırma, Gabor dönüşümü ve Mel ölçeği yardımıyla spektrogram görüntülerinin elde edilmesi gibi ön işlemlerden geçirilmiştir. Yapılan çeşitli deneyler ve kullanılan farklı mimariler sonucunda en yüksek performansın geliştirilen yeni mimari (VGG16-ESA-DVM) ile sağlandığı gözlemlenmiştir. Normal solunum sesi için %84, Ral solunum sesi için %80, Ronküs solunum sesi için %86'lık sınıflandırma doğruluğu ve %88.4 ROC-AUC skoru elde edilmiştir. Ayrıca literatürde başarısı kanıtlanmış mimariler (VGG16, ResNet101, DenseNet201, InceptionV3) kullanılarak yapılan deneyler sonucunda en yüksek sınıflandırma doğruluğunun VGG16 mimarisi (%78) ile sağlandığı gözlemlenmiştir. VGG16 öğrenme aktarımı ile %78 (Çizelge 6.8) doğruluk sağlamıştır. Bunun üzerine bir de 2 katmanlı ESA mimarisi ve DVM sınıflandırıcı eklenince bu başarı oranı %83'e çıkmıştır.

Bu tez çalışması, literatüre solunum sesi verileri açısından niteliksel ve niceliksel olarak katkı sağlamıştır. Ayrıca yapılan çalışmayı özgün kılan ve literatürdeki çalışmalardan ayıran diğer yönleri ise; sınıflandırma performansı açısından gösterdiği başarı ve oluşturulan yeni mimaridir. Oluşturulan yeni mimari ile literatürdeki benzer veri kümesi ile yapılan çalışmadan %10.8 daha yüksek başarı sağlanmıştır. Çalışmada kullanılan veriler ve sınıflandırma performansları detaylı olarak Bölüm 7'de tartışılmıştır.

Bu çalışma ile geliştirilen mimari alternatif tıbbi teşhis yolu olarak uygulanabilir. Bunun yanı sıra teşhis süresini azaltarak hastalıkların erken tanısına imkan sunabilmesi açısından da önemlidir. Sürekli gözlem altında tutulması gereken veya evde bakımı yapılan hastaların verilerinin anlık olarak böyle bir karar destek mekanizmasından geçerek hekime bilgi vermesi zaman açısından bir avantaj sağlayarak hastalıkların erken evrelerini tespit etmek için kullanılabilir. Özellikle uzman hekime ulaşmanın güç olduğu veya hekim sayısının yetersiz olduğu durumlarda bu tür sistemler kullanılabilir.

KAYNAKLAR

- [1] “The top 10 causes of death,” <https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death>, accessed: 2022-5-08.
- [2] A. Mondal, P. Bhattacharya, and G. Saha, “Detection of lungs status using morphological complexities of respiratory sounds,” *The Scientific World Journal*, vol. 2014, **2014**.
- [3] A. Mansoor, U. Bagci, Z. Xu, B. Foster, K. N. Olivier, J. M. Elinoff, A. F. Suffredini, J. K. Udupa, and D. J. Mollura, “A generic approach to pathological lung segmentation,” *IEEE transactions on medical imaging*, vol. 33, no. 12, pp. 2293–2310, 2014.
- [4] E. R. PACHT, J. W. TURNER, M. GAILIUN, L. A. VIOLI, D. RALSTON, H. S. MEKHJIAN, and R. C. S. JOHN, “Effectiveness of telemedicine in the outpatient pulmonary clinic,” *Telemedicine Journal*, vol. 4, no. 4, pp. 287–292, 1998.
- [5] P. Piirila and A. Sovijarvi, “Crackles: recording, analysis and clinical significance,” *European Respiratory Journal*, vol. 8, no. 12, pp. 2139–2148, 1995.
- [6] M. Garingo, “Audio display and environmental sound analysis of diagnostic and therapeutic respiratory sounds,” Master’s thesis, Ryerson University, 2014.
- [7] M. Kompis, H. Pasterkamp, and G. R. Wodicka, “Acoustic imaging of the human chest,” *Chest*, vol. 120, no. 4, pp. 1309–1321, 2001.
- [8] U. Buergi and L. C. Huber, “Lung auscultation—an overview,” *Deutsche medizinische Wochenschrift (1946)*, vol. 140, no. 14, pp. 1078–1082, 2015.
- [9] R. Loudon and R. L. Murphy Jr, “Lung sounds,” *American Review of Respiratory Disease*, vol. 130, no. 4, pp. 663–673, 1984.
- [10] S. Mangione and L. Z. Nieman, “Pulmonary auscultatory skills during training in internal medicine and family practice,” *American journal of respiratory and critical care medicine*, vol. 159, no. 4, pp. 1119–1124, 1999.
- [11] S. Aras, A. Gangal, and Y. Bülbül, “Classification of normalized healthy and pathological lung sounds,” in *2015 23rd Signal Processing and Communications Applications Conference (SIU)*. IEEE, **2015**, pp. 1150–1153.
- [12] A. Cohen, “Signal processing methods for upper airway and pulmonary dysfunction diagnosis,” *IEEE Engineering in Medicine and Biology Magazine*, vol. 9, no. 1, pp. 72–75, 1990.
- [13] R. Palaniappan, K. Sundaraj, and N. U. Ahamed, “Machine learning in lung sound analysis: a systematic review,” *Biocybernetics and Biomedical Engineering*, vol. 33, no. 3, pp. 129–135, **2013**.
- [14] S. Reichert, R. Gass, C. Brandt, and E. Andrès, “Analysis of respiratory sounds: state of the art,” *Clinical medicine. Circulatory, respiratory and pulmonary medicine*, vol. 2, pp. CCRPM–S530, 2008.

- [15] P. Forgacs, "Lung sounds," *British journal of diseases of the chest*, vol. 63, no. 1, pp. 1–12, 1969.
- [16] Ü. Çimen, "Solunum seslerinin yapay zekâ ortamında sınıflandırılması," Master's thesis, 2016.
- [17] S. İçer and Ş. Gengeç, "Classification and analysis of non-stationary characteristics of crackle and rhonchus lung adventitious sounds," *Digital Signal Processing*, vol. 28, pp. 18–27, 2014.
- [18] V. A. MCKUSICK, G. N. WEBB, J. O. Humphries, and J. A. REID, "On cardiovascular sound: further observations by means of spectral phonocardiography," *Circulation*, vol. 11, no. 6, pp. 849–870, 1955.
- [19] S. S. Kraman, G. R. Wodicka, G. A. Pressler, and H. Pasterkamp, "Comparison of lung sound transducers using a bioacoustic transducer testing system," *Journal of Applied Physiology*, vol. 101, no. 2, pp. 469–476, 2006.
- [20] A. Sovijarvi, F. Dalmaso, J. Vanderschoot, L. Malmberg, G. Righini, and S. Stoneman, "Definition of terms for applications of respiratory sounds," *European Respiratory Review*, vol. 10, no. 77, pp. 597–610, 2000.
- [21] A. Sovijarvi, J. Vanderschoot, and J. Earis, "Standardization of computerized respiratory sound analysis," *European Respiratory Review*, vol. 10, no. 77, pp. 585–585, 2000.
- [22] E. Andrès, S. Reichert, R. Gass, and C. Brandt, "A french national research project to the creation of an auscultation's school: the asap project," *European journal of internal medicine*, vol. 20, no. 3, pp. 323–327, 2009.
- [23] A. Oliveira and A. Marques, "Respiratory sounds in healthy people: a systematic review," *Respiratory medicine*, vol. 108, no. 4, pp. 550–570, **2014**.
- [24] Q. Chen, W. Zhang, X. Tian, X. Zhang, S. Chen, and W. Lei, "Automatic heart and lung sounds classification using convolutional neural networks," in *2016 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA)*. IEEE, **2016**, pp. 1–4.
- [25] G. Serbes, S. Ulukaya, and Y. P. Kahya, "An automated lung sound preprocessing and classification system based on spectral analysis methods," in *International Conference on Biomedical and Health Informatics*. Springer, 2017, pp. 45–49.
- [26] F. Demir, A. Sengur, and V. Bajaj, "Convolutional neural networks based efficient approach for classification of lung diseases," *Health information science and systems*, vol. 8, no. 1, pp. 1–8, 2020.
- [27] L. Fraiwan, O. Hassanin, M. Fraiwan, B. Khassawneh, A. M. Ibnian, and M. Alkhodari, "Automatic identification of respiratory diseases from stethoscopic lung sound signals using ensemble classifiers," *Biocybernetics and Biomedical Engineering*, vol. 41, no. 1, pp. 1–14, 2021.

- [28] M. Aykanat, Ö. Kılıç, B. Kurt, and S. Saryal, "Classification of lung sounds using convolutional neural networks," *EURASIP Journal on Image and Video Processing*, vol. 2017, no. 1, p. 65, **2017**.
- [29] M. Syahputra, S. Situmeang, R. Rahmat, and R. Budiarto, "Noise reduction in breath sound files using wavelet transform based filter," in *IOP Conference Series: Materials Science and Engineering*, vol. 190, no. 1. IOP Publishing, 2017, p. 012040.
- [30] R. L. Wilkins, J. R. Dexter, R. L. Murphy, and E. A. DelBono, "Lung sound nomenclature survey," *Chest*, vol. 98, no. 4, pp. 886–889, 1990.
- [31] Y. Kim, Y. Hyon, S. S. Jung, S. Lee, G. Yoo, C. Chung, and T. Ha, "Respiratory sound classification for crackles, wheezes, and rhonchi in the clinical field using deep learning," *Scientific Reports*, vol. 11, no. 1, pp. 1–11, 2021.
- [32] M. Elhilali and J. E. West, "The stethoscope gets smart: Engineers from Johns Hopkins are giving the humble stethoscope an AI upgrade," *IEEE Spectrum*, vol. 56, no. 2, pp. 36–41, 2019.
- [33] A. Sovijarvi, "Characteristics of breath sounds and adventitious respiratory sounds," *Eur Respir Rev*, vol. 10, pp. 591–596, **2000**.
- [34] J. Earis and B. Cheetham, "Future perspectives for respiratory sound research," *European Respiratory Review*, vol. 10, no. 77, pp. 641–646, 2000.
- [35] G. Chambres, P. Hanna, and M. Desainte-Catherine, "Automatic detection of patient with respiratory diseases using lung sound analysis," in *2018 International Conference on Content-Based Multimedia Indexing (CBMI)*. IEEE, 2018, pp. 1–6.
- [36] Y. Liu, Y. Lin, S. Gao, H. Zhang, Z. Wang, Y. Gao, and G. Chen, "Respiratory sounds feature learning with deep convolutional neural networks," in *2017 IEEE 15th Intl Conf on Dependable, Autonomic and Secure Computing, 15th Intl Conf on Pervasive Intelligence and Computing, 3rd Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress (DASC/PiCom/DataCom/CyberSciTech)*. IEEE, 2017, pp. 170–177.
- [37] M. I. Jordan and T. M. Mitchell, "Machine learning: Trends, perspectives, and prospects," *Science*, vol. 349, no. 6245, pp. 255–260, 2015.
- [38] G. B. Semmlow, John L, *Biosignal and medical image processing*. CRC press, 2014.
- [39] A. Hossain, "Diagnosis of autonomous vehicles using machine learning," 2018.
- [40] G. Aurélien, "Hands-on machine learning with scikit-learn & tensorflow," *Geron Aurelien*, 2017.
- [41] L. Deng and D. Yu, "Deep learning: methods and applications," *Foundations and trends in signal processing*, vol. 7, no. 3–4, pp. 197–387, 2014.

- [42] S. L. Brunton and J. N. Kutz, *Data-driven science and engineering: Machine learning, dynamical systems, and control*. Cambridge University Press, 2019.
- [43] B. E. Boser, I. M. Guyon, and V. N. Vapnik, “A training algorithm for optimal margin classifiers,” in *Proceedings of the 5th Annual ACM Workshop on Computational Learning Theory*, pp. 144–152.
- [44] C. Cortes and V. Vapnik, “Support-vector networks,” *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.
- [45] V. Kecman, “Basics of machine learning by support vector machines,” in *Real World Applications of Computational Intelligence*. Springer, 2005, pp. 49–103.
- [46] M. A. Nielsen, *Neural networks and deep learning*. Determination press San Francisco, CA, 2015, vol. 25.
- [47] V. Nair and G. E. Hinton, “Rectified linear units improve restricted boltzmann machines,” in *Icml*, 2010.
- [48] W. Hao, W. Yizhou, L. Yaqin, and S. Zhili, “The role of activation function in cnn,” in *2020 2nd International Conference on Information Technology and Computer Application (ITCA)*. IEEE, 2020, pp. 429–432.
- [49] S. Ruder, “An overview of gradient descent optimization algorithms,” *arXiv preprint arXiv:1609.04747*, 2016.
- [50] “Improving vanilla gradient descent,” <https://towardsdatascience.com/improving-vanilla-gradient-descent-f9d91031ab1d>, accessed: 2021-11-12.
- [51] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” *arXiv preprint arXiv:1412.6980*, 2014.
- [52] A. Abbas and A. Fahim, “An automated computerized auscultation and diagnostic system for pulmonary diseases,” *Journal of medical systems*, vol. 34, no. 6, pp. 1149–1155, 2010.
- [53] S. Haykin, *Advances in spectrum analysis and array processing (vol. III)*. Prentice-Hall, Inc., **1995**.
- [54] L. Deng, M. L. Seltzer, D. Yu, A. Acero, A.-r. Mohamed, and G. Hinton, “Binary coding of speech spectrograms using a deep auto-encoder,” in *Eleventh Annual Conference of the International Speech Communication Association*, **2010**.
- [55] J. Dennis, H. D. Tran, and H. Li, “Spectrogram image feature for sound event classification in mismatched conditions,” *IEEE signal processing letters*, vol. 18, no. 2, pp. 130–133, **2010**.
- [56] T. Kaneko, S. Takaki, H. Kameoka, and J. Yamagishi, “Generative adversarial network-based postfilter for stft spectrograms,” in *INTERSPEECH*, **2017**, pp. 3389–3393.
- [57] K. R. Farrell, R. J. Mammone, and K. T. Assaleh, “Speaker recognition using neural networks and conventional classifiers,” *IEEE Transactions on speech and audio processing*, vol. 2, no. 1, pp. 194–205, **1994**.

- [58] O. Eray, “Destek vektör makineleri ile ses tanıma uygulaması,” Master’s thesis, Pamukkale Üniversitesi Fen Bilimleri Enstitüsü, **2008**.
- [59] J. Salamon and J. P. Bello, “Deep convolutional neural networks and data augmentation for environmental sound classification,” *IEEE Signal processing letters*, vol. 24, no. 3, pp. 279–283, 2017.
- [60] C. Shorten and T. M. Khoshgoftaar, “A survey on image data augmentation for deep learning,” *Journal of big data*, vol. 6, no. 1, pp. 1–48, 2019.
- [61] Y. Aytar, C. Vondrick, and A. Torralba, “Soundnet: Learning sound representations from unlabeled video,” *Advances in neural information processing systems*, vol. 29, pp. 892–900, 2016.
- [62] Z. Mushtaq and S.-F. Su, “Efficient classification of environmental sounds through multiple features aggregation and data enhancement techniques for spectrogram images,” *Symmetry*, vol. 12, no. 11, p. 1822, 2020.
- [63] V. Sze, Y.-H. Chen, T.-J. Yang, and J. S. Emer, “Efficient processing of deep neural networks: A tutorial and survey,” *Proceedings of the IEEE*, vol. 105, no. 12, pp. 2295–2329, 2017.
- [64] S. J. Pan and Q. Yang, “A survey on transfer learning,” *IEEE Transactions on knowledge and data engineering*, vol. 22, no. 10, pp. 1345–1359, 2009.
- [65] W. Dai, G.-R. Xue, Q. Yang, and Y. Yu, “Transferring naive bayes classifiers for text classification,” in *AAAI*, vol. 7, 2007, pp. 540–545.
- [66] E. Eaton, T. Lane *et al.*, “Modeling transfer relationships between learning tasks for improved inductive transfer,” in *Joint european conference on machine learning and knowledge discovery in databases*. Springer, 2008, pp. 317–332.
- [67] C.-H. Hsiao, T.-W. Lin, C.-W. Lin, F.-S. Hsu, F. Y.-S. Lin, C.-W. Chen, and C.-M. Chung, “Breathing sound segmentation and detection using transfer learning techniques on an attention-based encoder-decoder architecture,” in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE, 2020, pp. 754–759.
- [68] S. Gupta, M. Agrawal, and D. Deepak, “Gammatonegram based triple classification of lung sounds using deep convolutional neural network with transfer learning,” *Biomedical Signal Processing and Control*, vol. 70, p. 102947, 2021.
- [69] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” *arXiv preprint arXiv:1409.1556*, 2014.
- [70] C. C. Aggarwal *et al.*, “Neural networks and deep learning,” *Springer*, vol. 10, pp. 978–3, 2018.
- [71] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

- [72] S. Zagoruyko and N. Komodakis, “Wide residual networks,” *arXiv preprint arXiv:1605.07146*, 2016.
- [73] X. Xia, C. Xu, and B. Nan, “Inception-v3 for flower classification,” in *2017 2nd International Conference on Image, Vision and Computing (ICIVC)*. IEEE, 2017, pp. 783–787.
- [74] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, “Densely connected convolutional networks,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4700–4708.
- [75] Y. Ma, X. Xu, and Y. Li, “Lungrn+ nl: An improved adventitious lung sound classification using non-local block resnet neural network with mixup data augmentation,” in *Interspeech*, 2020, pp. 2902–2906.
- [76] Y. Ma, X. Xu, Q. Yu, Y. Zhang, Y. Li, J. Zhao, and G. Wang, “Lungrn: A smart digital stethoscope for detecting respiratory disease using bi-resnet deep learning algorithm,” in *2019 IEEE Biomedical Circuits and Systems Conference (BioCAS)*. IEEE, 2019, pp. 1–4.
- [77] B. ARI, Ö. F. ALÇİN, and A. ŞENGÜR, “A lung sound classification system based on data augmenting using elm-wavelet-ae,” *Turkish Journal of Science & Technology*, vol. 17, no. 1, 2022.
- [78] F. Cinyol, H. E. Mutlu, and U. Baysal, “Classification of lung sounds with convolutional neural network,” in *2017 21st National Biomedical Engineering Meeting (BIYOMUT)*. IEEE, 2017, pp. i–iv.
- [79] F. Cinyol, U. Baysal, E. Gehr, E. Babaoğlu, S. Ulaşlı, and D. Köksal, “Classification of respiratory sounds including normal and crackle/rhonchi pathologies,” in *2020 28th Signal Processing and Communications Applications Conference (SIU)*. IEEE, 2020, pp. 1–4.