

TOBB EKONOMİ VE TEKNOLOJİ ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

**PATOLOJİK SESLERİN TANISI İÇİN
DERİN ÖĞRENME TABANLI
TIBBİ KARAR DESTEK SİSTEMİNİN GELİŞTİRİLMESİ**

YÜKSEK LİSANS TEZİ

İrem BİGAT

Biyomedikal Mühendisliği Anabilim Dalı

Tez Danışmanı: Prof. Dr. Osman EROĞUL

AĞUSTOS 2022

ÖZET

Yüksek Lisans Tezi

PATOLOJİK SESLERİN TANISI İÇİN

DERİN ÖĞRENME TABANLI

TIBBİ KARAR DESTEK SİSTEMİNİN GELİŞTİRİLMESİ

İrem BİGAT

TOBB Ekonomi ve Teknoloji Üniversitesi
Fen Bilimleri Enstitüsü
Biyomedikal Mühendisliği Anabilim Dalı

Danışman: Prof. Dr. Osman EROĞUL

Tarih: Ağustos 2022

Patolojik duruma bağlı olarak normal konuşma akışının bozulması, ses bozukluğu olarak bilinir. Bu nedenle, mevcut herhangi bir bozukluk, konuşma üretim sisteminin işleyişini bozar ve dolayısıyla bozuk bir ses üretir. Bazı laringeal patolojiler hayatı tehdit eder, bu nedenle ses bozukluğunun erken tespiti önemlidir. Patolojik seslerin tespitinde bir karar destek sisteminin geliştirilmesi hayati önem taşımaktadır. Patolojik seslerin belirlenmesi amacıyla seslerden çıkarılan özniteliklerin değerlendirilmesinde istatistiksel yöntemlerin grup bazında bir sonuç vermesi nedeniyle bireysel düzeyde bir cevap elde edilebilmesi amacıyla son yıllarda makine öğrenme yöntemleri araştırmacılar tarafından ilgi çekici bir konu olmuştur. Bununla birlikte makine öğrenmesinin özniteliklerin manuel çıkarılmasına ihtiyaç duyması nedeniyle optimal özniteliklerin otomatik olarak çıkarılabildiği derin öğrenme teknikleri araştırmacıların güncel araştırma konuları arasına girmiştir. Ancak henüz patolojik ses bozukluklarının tespiti alanında derin öğrenme tekniklerinin kullanımı ile ilgili az sayıda araştırma çalışması bulunmaktadır. Bu tez çalışmasında, patolojik seslerin belirlenmesi amacıyla derin öğrenme yöntemleri kullanılmıştır. Çalışmada

Saarbruecken Ses Veritabanından vokal kordlardaki yapısal deęişikliklerin neden olduęu organik disfoniye sebep olan patolojilere sahip hastaların ses kayıtları seçilmiştir. Bu patolojiler arasında larenjit, lökoplazi, Reinke ödemi, rekürren laringeal sinir felci, vokal kord karsinomu ve vokal kord polibi bulunmaktadır. Her bir bireyin nötr perdesinde sürekli sesli /a/ sesi kayıtları seçilmiştir. 380'i sağlıklı ve 380'i patolojik olmak üzere 760 ses kaydı kullanılmıştır. Veriler, sırasıyla %75 ve %25 örnek içeren eğitim seti ve test seti olarak ayrılmıştır. Ses sinyallerine öncelikle dalgacık gürültü giderme işlemi uygulanmıştır. Daha sonrasında ses sinyallerinin spektrogram görüntüleri alınarak dört farklı Evrişimsel Sinir Ağı (ESA) mimarisine girildi olarak verilmiştir. Tez kapsamında ESA mimarisi olarak GoogleNet, ResNet-50, AlexNet ve SqueezeNet çalışılmıştır. İlk aşamada patolojik seslerin belirlenmesi amacıyla seçilen Evrişimsel Sinir Ağı mimarileri kendi sınıflandırıcılarıyla birlikte kullanılmıştır. Daha sonra aynı Evrişimsel Sinir Ağı mimarileri bu kez sadece öznelik çıkarımında kullanılmıştır ve Komşuluk Bileşen Analizi ile öznelik seçimi yapıldıktan sonra farklı sınıflandırma algoritmalarıyla sınıflandırılarak oluşturulan modellerin performans analizleri yapılmıştır. Kullanılan sınıflandırma algoritmaları Karar ağaçları, Destek Vektör Makineleri, k-En yakın komşuluk, Ensemble ve bu çalışma için tasarlanmış bir karar ağacı yöntemidir. En başarılı performans SqueezeNet mimarisinden çıkarılan özneliklerin Ensemble algoritması ile sınıflandırılması sonucu elde edilmiştir. Gözlemlenen bulgular, önerilen bu modelin patolojik seslerin belirlenmesinde umut verici olduğunu göstermektedir.

Anahtar Kelimeler: Patolojik ses belirleme, Derin öğrenme, Evrişimsel sinir ağları, GoogleNet, ResNet-50, AlexNet, SqueezeNet.

ABSTRACT

Master of Science

**DEVELOPMENT OF A DEEP LEARNING-BASED
MEDICAL DECISION SUPPORT SYSTEM
FOR THE DIAGNOSIS OF PATHOLOGICAL VOICES**

İrem BİGAT

TOBB University of Economics and Technology
Institute of Natural and Applied Sciences
Biomedical Engineering Science Programme

Supervisor: Prof. Dr. Osman EROĞUL

Date: August 2022

The disruption of normal speech flow due to pathological conditions is known as a voice disorder. Therefore, any existing disorder disrupts the speech production system's functioning and produces a distorted voice. Since some laryngeal pathologies are life-threatening, the early detection of voice disorders is important. For this purpose, there is a need to develop a decision support system in the detection of pathological voices. In recent years, machine learning methods have become an interesting research topic to determine pathological voices in order to obtain an individual-level answer, since statistical methods give a group-based result in the evaluation of features extracted from voices. However, since machine learning requires manual extraction of features, deep learning techniques, in which optimal features can be extracted automatically, have become one of the current research topics. However, there are only few research studies on the use of deep learning techniques in the detection of pathological voice disorders. In this thesis study, deep learning methods were used to identify pathological voices. The voice recordings of patients with pathologies causing organic dysphonia due to structural changes in the

vocal cords were selected from the Saarbruecken Voice Database. These pathologies included laryngitis, leukoplakia, Reinke's edema, recurrent laryngeal nerve paralysis, vocal cord carcinoma, and vocal cord polyps. The sustained vowel /a/ at the neutral pitch of each individual was selected. The sample included a total of 760 recordings, of which 380 belonged to healthy voices and 380 belonged to pathological voices. The data were divided into training and test sets containing 75% and 25% of the samples, respectively. In the analysis of the samples, first, wavelet noise denoising was applied to the voice signals. Then, the spectrogram images of the voice signals were taken and utilized as inputs in four different Convolutional Neural Network (CNN) architectures, namely GoogleNet, ResNet-50, AlexNet, and SqueezeNet. The selected CNN architectures were used with their own classifiers to determine the pathological voices. Subsequently, the same architectures were employed only for feature extraction, then the Neighborhood Component Analysis employed for feature selection. The performance analyses of the models were undertaken by classifying the selected features with the following classification algorithms: Decision trees, Support Vector Machines, k-Nearest Neighborhood, Ensemble, and a decision tree method designed for this study. The most successful performance was obtained from the method in which the features had been extracted by the SqueezeNet architecture and classified with the Ensemble algorithm. According to the results, the proposed model is promising for the identification of pathological voices.

Keywords: Pathological voice detection, Deep learning, Convolutional neural networks, GoogleNet, ResNet-50, AlexNet, SqueezeNet.

İÇİNDEKİLER

	<u>Sayfa</u>
ÖZET	iv
ABSTRACT	vi
TEŞEKKÜR	viii
İÇİNDEKİLER	ix
ŞEKİL LİSTESİ	xi
ÇİZELGE LİSTESİ	xiii
KISALTMALAR	xiv
SEMBOL LİSTESİ	xv
1. GİRİŞ	1
1.1 Tezin Amacı ve Önemi.....	2
1.2 Tezin Kapsamı.....	3
2. SES VE KONUŞMA	5
2.1 Ses ve Konuşmanın Tanımı.....	5
2.2 Konuşmayı uyaran etkenler.....	5
2.3 Konuşmanın nörofizyolojisi	6
2.4 Konuşmanın laringeal ve solunum sistemi komponenti	7
2.4.1 Larenksin fonksiyonel anatomisi	7
2.4.2 Larenksin çatısı	8
2.4.3 Vokal kordların fizyolojisi	17
2.5 Ses fizyolojisi	19
2.5.1 Ses fiziği ile ilgili temel kavramlar	19
2.5.2 Sesin özellikleri.....	21
2.5.3 Ses perdesi ve şiddetinin belirlenmesi	22
2.5.4 Ses aralığı ve rejisterler	22
2.6 Disfoni	23
2.6.1 Disfonin klinikopatolojik özellikleri.....	23
2.6.2 Vokal kord patolojileri ve disfoni	23
2.6.3 Disfonide kullanılan tanı yöntemleri.....	27
2.6.3.1 Subjektif ses analizi.....	27
2.6.3.2 Objektif ses analizi.....	28
2.6.3.3 Videolaringoskopi.....	31
2.6.3.4 Elektromiyografi	32
3. KURAMSAL TEMELLER	33
3.1 Dalgacık Eşiği Gürültü Giderme Algoritması.....	33
3.2 Spektrogram	34
3.3 Derin Öğrenme	36
3.4 Evrişimsel Sinir Ağları (ESA).....	37
3.4.1 ESA'nın temel basamakları	37
3.4.1.1 Giriş katmanı	37
3.4.1.2 Evrişim katmanı	38
3.4.1.3 Doğrusal olmayan aktivasyon fonksiyonu	40
3.4.1.4 Havuzlama katmanı.....	41

3.4.1.5 Tam bağlantılı katman.....	42
3.4.1.6 Dizi normalizasyonu	42
3.4.1.7 DropOut.....	42
3.4.1.8 Sınıflandırma katmanı	43
3.4.2 GoogleNet	44
3.4.3 ResNet-50.....	45
3.4.4 SqueezeNet.....	46
3.4.5 AlexNet	48
3.5 Komşuluk Bileşen Analizi.....	49
3.6 Sınıflandırıcılar.....	52
3.6.1 Karar ağaçları	52
3.6.2 Destek Vektör Makineleri	53
3.6.3 k-En Yakın Komşuluk Sınıflaması	55
3.6.4 Topluluk öğrenmesi yöntemleri	56
3.6.4.1 Artırma	57
3.7 Patolojik Seslerin Belirlenmesinde Öne Çıkan Çalışmalar	58
4. MATERYAL VE YÖNTEM	61
4.1 Kullanılan Veri Seti.....	62
4.2 Ön İşleme.....	64
4.2.1 Ses sinyali gürültü giderme	64
4.2.2 Spektrogram görüntülerinin üretilmesi	65
4.3 Farklı ESA Mimarileri ile Patolojik Seslerin Sınıflandırılması	68
4.4 Seçilen ESA Mimarilerinden Öznitelik Çıkarımı ve Öznitelik Seçimi.....	68
4.5 Diğer Sınıflandırma Algoritmaları	70
5. BULGULAR, DEĞERLENDİRME VE TARTIŞMA	73
5.1 Bulgular	74
5.2 Değerlendirme ve Tartışma	77
6. SONUÇ VE ÖNERİLER	83
KAYNAKLAR.....	85
ÖZGEÇMİŞ	93

ŞEKİL LİSTESİ

Sayfa

Şekil 2.1 : Beyindeki kontrol merkezleri [Url-1].	6
Şekil 2.2 : Larinks anatomisi [Url-2].	8
Şekil 2.3 : Larinks anatomisinde kıkırdaklar [Url-3].	9
Şekil 2.4 : Vokal kord kesiti [13].	14
Şekil 2.5 : Vokal kord mikro anatomisi ve histolojisi [Url-4].	15
Şekil 2.6 : Bilgisayar ortamında üretilmiş frekansı 100 Hz olan saf sese ait dalga şekli [12].	20
Şekil 2.7 : Glottis düzeyindeki işlenmemiş sesin spektrumu [12].	20
Şekil 2.8 : [i] sesine ait dalga şekli [12].	20
Şekil 2.9 : [i] sesine ait HFD. F0, temel frekans; F1, birinci formant; F2, ikinci formant; F3, üçüncü formant, F4, dördüncü formant [12].	21
Şekil 2.10 : Fonasyon sırasında posterior glottik açıklık [13].	24
Şekil 2.11 : Fonasyon sırasında bowing tipi glottik kapanma yetmezliği [13].	24
Şekil 2.12 : Fonasyon sırasında kum saati şeklinde açıklık [13].	24
Şekil 2.13 : Tek taraflı vokal kord paralizisi [13].	25
Şekil 3.1 : (a) bir ses sinyali, (b) ses sinyalinin spektrogramı.	35
Şekil 3.2 : Örnek bir ESA şeması [49].	37
Şekil 3.3 : Bilgisayar bir görüntüyü bir dizi sayı olarak görür. Sağdaki matris, her biri sol görüntüdeki piksel parlaklığına karşılık gelen 0 ile 255 arasında sayılar içerir. Her ikisi de ortadaki görüntüde kaplanmıştır [51].	38
Şekil 3.4 : 5x5x3 boyutundaki bir giriş görüntüsüne 3x3'lük bir filtrenin uygulandığı evrişim işlemi [52].	39
Şekil 3.5 : Sinir ağlarına yaygın olarak uygulanan aktivasyon fonksiyonları: (a) ReLU, (b) sigmoid ve (c) tanh [51].	40
Şekil 3.6 : Bir ESA modelinde evrişim işlemi ve ReLU işleminin giriş görüntüsüne yaptığı etki [52].	40
Şekil 3.7 : Maksimum, ortama havuzlama ve küresel ortalama havuzlama örneği [54].	41
Şekil 3.8 : Havuzlama katmanının uygulanması sonucu oluşan görüntü [52].	42
Şekil 3.9 : (a) ağın orijinal yapısı, (b) DropOut işleminden sonraki hali [55].	43
Şekil 3.10 : Sigmoid fonksiyonu [49].	43
Şekil 3.11 : Softmax fonksiyonu [56].	44
Şekil 3.12 : Google Block'un temel yapısı [54].	45
Şekil 3.13 : GoogLeNet mimarisi [57].	45
Şekil 3.14 : Artık öğrenmeye ilişkin basit bir yaklaşım [60].	46
Şekil 3.15 : ResNet-50 mimarisi [61].	46
Şekil 3.16 : SqueezeNet modelinde bir ateşleme modülünün genel yapısı [62].	47
Şekil 3.17 : Temel SqueezeNet modelinin makro mimarisi [62].	47
Şekil 3.18 : AlexNet mimarisi [65].	49
Şekil 3.19 : Karar ağacı örnek şeması.	52

Şekil 3.20 : 2 boyutlu uzayda ayrılabilir bir örnek [71].	53
Şekil 3.21 : Topluluk öğrenmesi genel şeması [75].	56
Şekil 4.1 : Tez çalışması akış şeması.	61
Şekil 4.2 : Örnek birer (a) sağlıklı ve (b) patolojik ses için dalgacık gürültü giderme.	65
Şekil 4.3 : Örnek birer (a) sağlıklı ve (b) patolojik ses sinyalinin gürültü giderme ardından KZFD uygulandıktan sonra elde edilen spektrogramlar.	66
Şekil 5.1 : ESA mimarileri ile birlikte kullanılan sınıflandırma algoritmalarının patolojik seslerin belirlenmesinde doğruluk performansları.	77
Şekil 5.2 : İlk aşama ile ikinci aşamanın doğruluk performansları.	80



ÇİZELGE LİSTESİ

Sayfa

Çizelge 4.1 : Dalgacık gürültü giderme uygulaması ve spektrogram görüntülerinin oluşturulması için parametre değerlerinin belirlenmesi.....	67
Çizelge 4.2 : Karar ağacı parametrelerinin belirlenmesi.....	70
Çizelge 4.3 : DVM parametrelerinin belirlenmesi.....	71
Çizelge 4.4 : k-NN parametrelerinin belirlenmesi.....	71
Çizelge 4.5 : Ensemble parametrelerinin belirlenmesi.....	72
Çizelge 4.6 : Tasarım karar ağacı parametrelerinin belirlenmesi.....	72
Çizelge 5.1 : Farklı ESA mimarileri ile patolojik seslerin sınıflandırma sonuçları...	74
Çizelge 5.2 : GoogleNet ile elde edilen ve seçilen özneliklerin sınıflandırma sonuçları.....	75
Çizelge 5.3 : ResNet-50 ile elde edilen ve seçilen özneliklerin sınıflandırma sonuçları.....	75
Çizelge 5.4 : AlexNet ile elde edilen ve seçilen özneliklerin sınıflandırma sonuçları.....	76
Çizelge 5.5 : SqueezeNet ile elde edilen ve seçilen özneliklerin sınıflandırma sonuçları.....	76
Çizelge 5.6 : ESA ile patolojik seslerin belirlenmesi için benzer literatür karşılaştırması.....	82

KISALTMALAR

KBB	: Kulak Burun Boğaz
dB	: Desibel
F₀	: Temel Frekans
HNR	: Harmonik/Gürültü oranı (Harmonic to Noise Ratio)
Hz	: Hertz
NNE	: Normalize Gürültü Enerjisi (Normalized Noise Energy)
VHI	: Ses Handikap Endeksi (Voice Handicap Index)
VPQ	: Ses Performans Anketi (Vocal Performance Questionnaire)
RLN	: Rekküren Laringeal Sinir (Recurrent Laryngeal Nerve)
SLN	: Superior Laringeal Sinir (Superior Laryngeal Nerve)
ESSLN	: Superior Laringeal Sinirin Eksternal Dalı (External Branch of the Superior Laryngeal Nerve)
IBSLN	: Superior Laringeal Sinirin İç Dalı (Inner Branch of the Superior Laryngeal Nerve)
STA	: Superior Tiroid Arteri (Superior Thyroid Artery)
HFD	: Hızlı Fourier Dönüşümü
VLS	: Videolaringoskopi
EMG	: Elektromiyografi
KZFD	: Kısa Zamanlı Fourier Dönüşümü
ESA	: Evrişimli Sinir Ağları
ReLU	: Doğrultulmuş Doğrusal Birim (Rectified Linear Unit)
tanh	: hiperbolik tanjant
ILSVRC	: ImageNet Large Scale Visual Recognition Challenge
RGB	: Kırmızı Yeşil Mavi (Red Green Blue)
ResNet	: Residual Network
KBA	: Komşuluk Bileşen Analizi
k-NN	: k-En Yakın Komşuluk (k-Nearest Neighbor)
DVM	: Destek Vektör Makineleri
MEEI	: Massachusetts Göz ve Kulak Kliniği (Massachusetts Eye and Ear Infirmary)
SVD	: Saarbrücken Ses Veritabanı (Saarbruecken Voice Database)
MDVP	: Çok Boyutlu Ses Programı (Multi-Dimensional Voice Program)
AVPD	: Arapça Ses Patolojisi Veritabanı (Arabic Voice Pathology Database)
GNR	: Glottal-Gürültü Oranı (Glottal-to-Noise Ratio)
PPV	: Pozitif Tahmin Değeri (Positive Predictive Value)
CDBN	: Convolutional Deep Belief Network (Evrişimsel Derin İnanç Ağı)
LR⁺	: Positive Likelihood Ratio (Pozitif Olabilirlik Oranı)

SEMBOL LİSTESİ

Bu çalışmada kullanılmış olan simgeler açıklamaları ile birlikte aşağıda sunulmuştur.

Simgeler	Açıklama
$\hat{\omega}_{m,n}$	Tahmini dalgacık katsayısı
$\omega_{m,n}$	Orijinal sinyalin dalgacık ayrıştırma katsayısı
μ	Kaldırılan gürültü eşiği
m	Zaman
ω	Frekans
n	Parça
g	Görüntü boyutu
p	Kenarlık değeri
f	Filtre boyutu
a	Adım sayısı
x	Mini dizi
μ_B	Mini dizi ortalaması
σ_B^2	Mini dizi varyansı
γ	Optimizasyon katsayısı
β	Optimizasyon katsayısı
z	Girdi vektörü
S	Eğitim seti
c	Sınıf sayısı
n	Gözlem sayısı
d	Uzaklık
w	Ağırlık vektörü
k	Çekirdek işlevi
λ	Düzenleme terimi
N_t	Toplam gözlem sayısı
T	Eşik
τ	Sabitlenmiş tolerans
H	Entropi
b	Eşik değeri
a_i	I indisindeki Lagrange çarpanı
k	Kontrol edilen komşuların sayısı

1. GİRİŞ

Konuşma, insanın günlük yaşamdaki doğal iletişim yolu olup; Anlaşılır bir konuşma üretebilmek için gırtlak da dahil olmak üzere birincil fonetik sistem organları önemli bir rol oynamaktadır. Herhangi bir patolojik duruma bağlı olarak normal konuşma akışının bozulması, ses bozukluğu olarak bilinmektedir [1].

Black ve arkadaşlarının yaptığı çalışmada 2011'de 3 ila 17 yaş arasındaki ABD'li çocukların yaklaşık %7,7'sinde ses, konuşma ve yutma ile ilgili bozukluklar görüldüğü bildirilmektedir [2]. Başka bir çalışmada da 2013'te 18 yaş ve üzeri ABD'li yetişkinlerin yaklaşık %7,6'sı ses bozukluğu şikayeti olduğu bildirilmektedir [3]. Literatürde ABD'li öğretmenlerin %20'si ses bozukluklarından muzdardır olduğu belirtilmektedir [4]. Ses bozukluğunun sosyal, profesyonel ve kişisel etkileşim biçimleri üzerinde önemli bir etkisi olmakla birlikte [5], bunun da düşük benlik saygısı ve kendini soyutlama gibi sorunlara yol açabilmektedir [6]. Larinks gibi ses oluşumunda etkili olan organlarda gelişebilen onkolojik durumlarda ilk semptom ses bozuklukları olması nedeniyle ses bozukluklarının erken dönemde tanınması bu hastaların prognozları açısından önemli olmaktadır.

Vokal kord patolojilerinin teşhisi Kulak Burun Boğaz (KBB) uzmanları tarafından endoskopi ve videolaringoskopi gibi invaziv prosedürler kullanılarak gerçekleştirilmekte olup, bu yöntemler zaman alıcı, pahalı ve invazivdirler. Bu olumsuzlukları gidermek amacıyla son dönemlerde geliştirilen bilgisayar destekli tıbbi sistemler ilgi odağı olmaktadır. Bu tamamlayıcı tanı yöntemleri genellikle non-invaziv ve objektif olup, ses bozuklukları ile ilgilenen uzmanlara avantaj sağlamaktadır.

Objektif ses analizi hastalardan alınan ses örneklerinin çeşitli yazılımlar kullanılarak incelenmesi olup; bu testlerden jitter, shimmer, habitüel ve ortalama temel frekans (Fo), Normalize Gürültü Enerjisi (Normalized Noise Energy, NNE), Harmonik/Gürültü oranı (Harmonic to Noise Ratio, HNR), tırmalayıcı ses, kısık ses ve soluklu ses karakteri, habitüel fonasyon süresince temel frekansın değişkenliği,

ötümlülük ve ötümsüzlük zamanının yüzdesi, en pes ve en tiz sesler arasındaki frekans farkı ve daha birçok parametre elde edilmektedir.

Son yıllarda ses patolojilerinin otomatik tespiti ile ilgili birçok araştırma yapılmakta olup, bu çalışmalarda, konuşma kayıtlarından öznitelikler çıkarılmakta ve daha sonra normal ses örneklerini patolojik ses kayıtlarından ayırt etmek için sınıflandırıcılar tarafından işlenmektedir. Optimal özniteliklerin manuel olarak çıkarılmasındaki zorluklar nedeniyle özniteliklerin otomatik tespitini yapabilmek amacıyla derin sinir ağlarının performansı güncel araştırma konuları arasına girmiştir.

1.1 Tezin Amacı ve Önemi

Onkolojik laringeal patolojiler insan hayatını tehdit etmekte olup, bu olgularda en erken semptom çoğu zaman ses bozuklukları olmaktadır. Bu olguların seslerindeki bozuklukların erken dönemde tanınması hastalığın tanı ve tedavisi açısından önemli avantajlar sağlayabilmektedir. Erken tanı konulması tedavinin başarısı için oldukça önemli olup, bu olgularda patolojik seslerin tespitinde bir karar destek sisteminin geliştirilmesi hayati önem taşımaktadır.

Patolojik seslerin belirlenmesi amacıyla seslerden çıkarılan özniteliklerin değerlendirilmesinde istatistiksel yöntemlerin zaman alıcı olması nedeniyle son yıllarda makine öğrenme yöntemleri araştırmacılar tarafından ilgi çekici bir konu olmuştur. Bununla birlikte makine öğrenmesinin özniteliklerin manuel çıkarılmasına ihtiyaç duyması nedeniyle optimal özniteliklerin otomatik olarak çıkarılabildiği derin öğrenme teknikleri araştırmacıların güncel araştırma konuları arasına girmiştir. Ancak henüz patolojik ses bozukluklarının tespiti alanında derin öğrenme tekniklerinin kullanımı ile ilgili literatürde az sayıda araştırma çalışması bulunmaktadır.

Bu tez çalışmasındaki hipotezimiz, son dönemlerde sinyal işleme ve makine öğrenmesi alanında yaşanan bilimsel gelişmelerden faydalanarak; ses bozuklukları şikâyeti olan olguların 2. ve 3. basamakta invaziv yöntemler ile değerlendirilmesinden önce, patolojik sesleri olan olguların ilk başvuru yaptıkları aile hekimlikleri ve diğer ses profesyoneli olmayan merkezlerde tanınmasının, onkolojik olgulara erken tanı konulmasının sağlanabileceği ve bu sayede hem KBB kliniklerinde iş yükü oluşturan sağlıklı sesi olan olgulara yönelik uygulanan gereksiz

invaziv işlemlerin önüne geçilebileceği hem de erken tanı konulması sağlanarak onkolojik olgularda prognozun daha iyi olması sağlanabileceğidir. Sonuç olarak; geliştirilecek bu yöntemlerle hem hastaların sağlığına hem de sosyal güvenlik kurumlarına önemli katkılar sağlanabileceğini düşünmekteyiz.

1.2 Tezin Kapsamı

Bu tez çalışmasında, derin öğrenme yöntemi kullanılarak patolojik seslerin belirlenmesi için ilk aşamada dört farklı ESA mimarisi kendi sınıflandırıcılarıyla kullanılıp daha sonra bu mimariler dört farklı makine öğrenme yöntemi ve bunların bir arada kullanılması ile oluşturulan bir karar ağacı algoritması ile birleştirilerek sınıflandırma gerçekleştirilerek, bu modellerin performansları değerlendirilmiştir.

Tezin ikinci bölümünde, ses ve konuşmanın tanımına, konuşmayı uyaran etkenler, konuşmanın laringeal ve solunum sistemi komponenti ile birlikte nörofizyolojisi, ses fiziği, disfoninin tanımı ve tanısında kullanılan yöntemleri özetlenmeye çalışılmıştır.

Üçüncü bölümde ise; çalışmada kullanılan dalgacık eşiği gürültü giderme algoritmasına, spektrograma, evrişimsel sinir ağlarının yapılarına, komşuluk bileşen analizine ve kullanılan sınıflandırma algoritmalarına özetlendikten sonra; literatürde son dönemlerde patolojik seslerin belirlenmesi konusunda yapılan çalışmalar derlenmiştir.

Dördüncü bölümde, çalışmada kullanılan veri setine ve bu setin seçilme kriterlerine değinildikten sonra, çalışmada uygulanan ön işleme metotları ve bunların kullanım nedenleri, kullanılan ESA mimarileri ve parametre seçimleri, öznitelik seçimi aşaması ve kullanılan sınıflandırma metotları seçilen parametre değerleriyle birlikte verilmiştir.

Beşinci bölümde ise; oluşturulan modellerden elde edilen bulgular, performanslarına bağlı olarak önerilen modellerin değerlendirmeleri ve literatürdeki benzer çalışmalar ile karşılaştırmaların sunulduğu tartışma bölümüne yer verilmiştir.

Tez çalışmasının genel değerlendirmesi ve öneriler ise; altıncı bölümde sunulmuştur.

2. SES VE KONUŞMA

2.1 Ses ve Konuşmanın Tanımı

Ses, doğadaki esnek cisimlerin titreşimi ile oluşan fiziksel bir enerji olup, İnsanda bu enerji çeşitli perde ve tınılarda kişiye özgü olarak konuşma şeklinde ortaya çıkmaktadır [7]. Konuşma düşüncelerin sözlerle anlatımı olmakla birlikte, çeşitli komponentlerden oluşan kompleks fonetik bir olaydır [8]. Konuşma, işitme sonucu ortaya çıkmakta olup, majör komponenti ses çıkarmaktır. Ses oluşumunda boyun, farenks, göğüs, karın ve pelvis ile birlikte, birçok diğer vücut bölgeleri rol oynamaktadır [9].

2.2 Konuşmayı uyaran etkenler

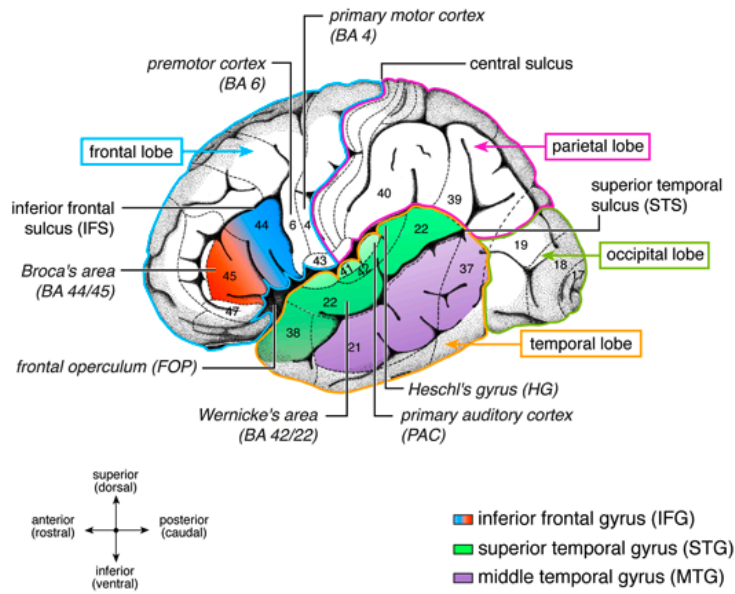
Ses oluşumunda ses çıkarmayı uyaran işitsel ve görsel uyarılar başta olmak üzere birçok duyuşal uyarı ile birlikte, bu uyarılara cevap veren sistemler rol oynamaktadır [7].

İşitsel uyarılar iç kulaktan alınmakta ve temporal lob transvers Heschl's gyrus merkezinde son bulmaktadırlar. Her iki işitme merkezinden gelen uyarılar sol hemisferdeki Wernicke alanına iletilmektedir. Burada değerlendirildikten sonra cevap oluşturulması için Fasciculus Arcuatus ile Brocca Merkezine iletilmektedir ve bu merkezde de gelen uyarıya uygun motor cevap düzenlenmektedir [7].

Görsel ve yazılı uyarılar retinada bulunan rod ve koni hücrelerinde başlayan, retinanın bipolar hücreleri tarafından alınan bu uyarılar, beyinde oksipital lobda bulunan kuneus ve lingual gyrusdaki görme merkezinde son bulmaktadır. Bu uyarılar yardımcı görme merkezlerini de uyardıktan sonra gyrus angularis ve gyrus supramarginalise ulaşır ve sonra bu merkezlerde değerlendirilen bilgiler, önce Wernicke alanına daha sonra Fasciculus Ancuatus ile motor cevap üretilmesi için Brocca alanına iletilir [7].

Diğer duysal impulslar tad, koku ve temas gibi duylara ait uyarılar pariyetal veya temporal korteksteki merkezlerde son bulmaktadır. Daha sonra öncelikle Wernicke alanına ve daha sonra da Brocca alanına iletilmektedirler (Şekil 2.1).

Görsel uyarılar oksipital, işitsel uyarılar temporal ve diğer duysal impulslar pariyeto-temporal bölgede algılanmaktadır ve sonra bilgiler entegrasyon merkezinde (gyrus angularis) değerlendirilerek Wernicke alanına iletilmektedir. Daha sonra motor cevap oluşulması için uyarılar Brocca alanına gönderilmektedir [7] (Şekil 2.1).



Şekil 2.1 : Beyindeki kontrol merkezleri [Url-1].

2.3 Konuşmanın nörofizyolojisi

Gyrus presentralisin alt kısmında konuşma ile ilgili organların kumanda merkezleri bulunmaktadır. Konuşmanın gerçekleşebilmesi için motor konuşma merkezi (Brocca) ile gyrus presentralisin alt kısmını kapsula eksternanın içindeki lifler bağlamaktadır [7].

Konuşma esnasında kasların iki yönlü koordine harekti gerekmektedir, bu nedenle sol serebral hemisferdeki sorumlu kortikal merkezler korpus kallozum ile sağ serebral hemisferdeki sorumlu merkezler ile iletişime geçmektedir. Brocca alanından çıkan iletiler gyrus presentralisin alt kısmındaki motor merkezlere iletildiğinde konuşma hazırlanmış olur. Solunum ile alınan havanın ekspirasyonu ile fonatuar ve artikülator kasların da devreye girmesi sayesinde konuşma ortaya çıkmaktadır.

Konuşma sırasında kelimeler ilk olarak zihinde düşünce veya kavram olarak gelişmektedir ve sonrasında gelişen bu kavram sembolize edilmektedir. Sonrasında da düşünce telaffuz edilmektedir. Yüksek serebral merkezler ve solunum işlevi ile oral kavitede bulunan yumuşak dokular, çene eklemi ile yüzde bulunan birtakım anatomik yapıların birlikte çalışması ile ortaya çıkan konuşma, nöromuskuler bir işlemdir [8].

2.4 Konuşmanın laringeal ve solunum sistemi komponenti

İnsanda sesin oluşabilmesi için anatomik nörokontrol mekanizmaları ile çalışan, enerji kaynağı ile titreşen iki yapı gerekmektedir. Akciğerlerden gelen hareketli, basınçlı hava enerji kaynağıdır ve titreşen yapılar da larenksteki vokal kordlardır. Ses oluşumunda sırasıyla akciğerler, larenks ve oral boşluğun birlikte ve uyumlu bir şekilde çalışması gerekmektedir.

Solunumdan sorumlu olan karın, pelvis, boyun gibi anatomik yapılar gibi emosyonel olaylar da ses oluşumunu dolaylı veya doğrudan etkilemektedir. Ekspirasyon havası ses için gereken enerjiyi taşımakla birlikte sesin çevreye yayılmasını da sağlamaktadır. Bunun için bu havanın yeterli basınca ve akım hızına sahip olması gerekmektedir, bunu sağlamak amacıyla diyafram ile karın kasları devreye girer [7].

2.4.1 Larenksin fonksiyonel anatomisi

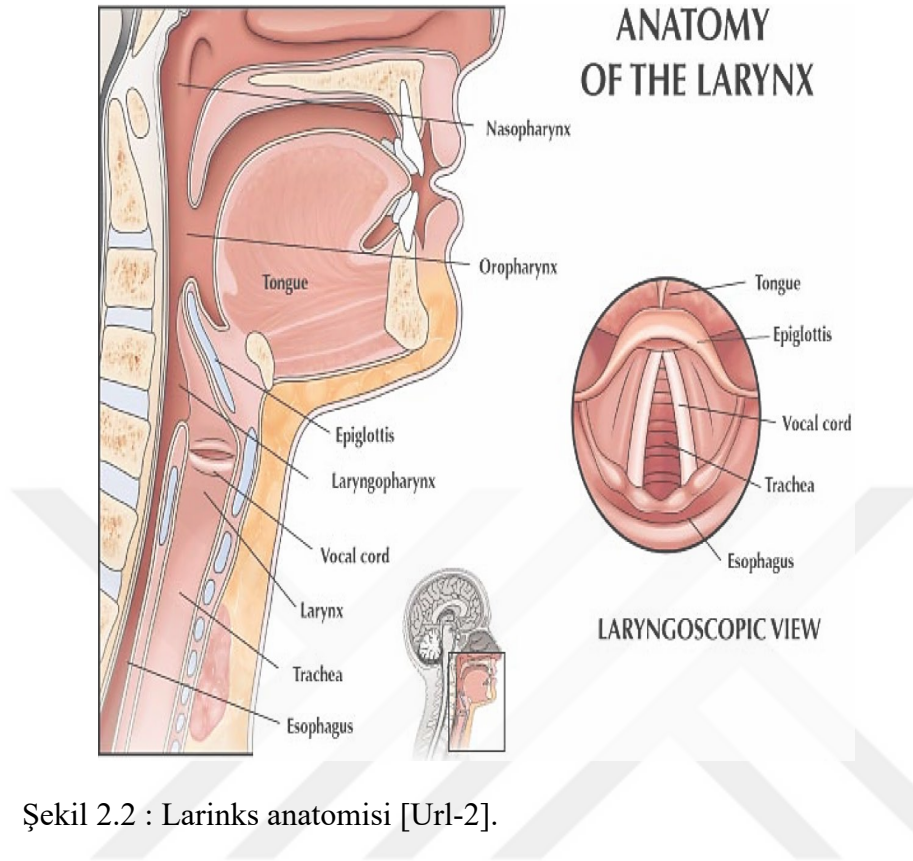
Larinks fonksiyonel olarak, yutma sırasında hava yolunu koruma amacıyla kapanma ve fonasyona izin verme işlevleri olan bir sfinkterdir [10].

Konuşmada larenksin dört primer fonksiyonu vardır:

- a) Hava akımını açıp kapaması sayesinde fonatuar kontrolü sağlamak
- b) Sesin temel frekansının kontrolünü sağlamak
- c) Ses şiddetinin kontrolünü sağlamak
- d) Ses kalitesini düzenlenmesini sağlamak [11].

Larinks solunum ve koruma işlevlerinin yanısıra, fonasyon işlevi ile de sözlü iletişimi sağlayan bir organdır. Larinksin çatısı; kıkırdaklar, kıkırdakları birbirine ve çevredeki dokulara bağlayan bağlardan ve membranlardan oluşmaktadır. Larinks çatısı dışında, larenksin iç yüzünü iç ve dış kaslarla örten ve ventriküler bant ile ses

kıvrımlarını oluşturan mukoza da larinks işlevleri açısından önemli diğer anatomik yapılardır [12]. Larinks anatomisi Şekil 2.2’de gösterilmiştir.



Şekil 2.2 : Larinks anatomisi [Url-2].

2.4.2 Larenksin çatısı

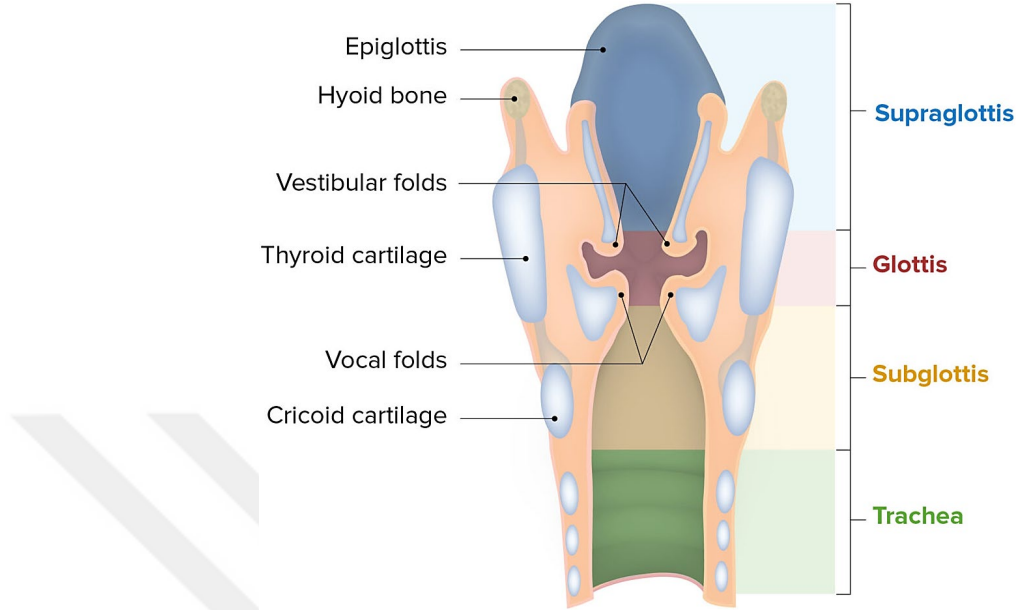
Kıkırdaklar

Larenks dokuz kıkırdaktan oluşmaktadır. Bu kıkırdaklardan tiroid, krikoid ve epiglot kıkırdaklar tek; ariteooid, komikülat ve kuneiform kıkırdaklar çifttir. Larinks anatomisinde kıkırdakların gösterimi Şekil 2.3’te verilmiştir.

- Tiroid kıkırdak: Larenks kıkırdakları arasında en büyüğüdür. Dörtgen şekilli iki laminadan oluşmaktadır. Laminaların ön kenarı, erkeklerde orta hat üzerinde 90°lik, kadınlarda ise 120°lik açı ile birleşmektedir. Bu açının dar olması erkeklerde larenksin boyunda çıkıntı yapmasına ve vokal kordların daha uzun olması nedeniyle ses perdesinin kadınlara göre daha kalın olmasına neden olur.
- Krikoid kıkırdak
- Aritenoid kıkırdak: İşlevi açısından en önemli kıkırdaktır. Vokal çıkıntı tabandan öne doğru uzanır ve buraya vokal ligaman ile vokal kas

yapışmaktadır. Laterale doğru uzanan çıkıntı da kas çıkıntısıdır ve buraya da posterior ile lateral aritenoit kaslar yapışmaktadır.

d) Kornikülat ve kuneiform kıkırdaklar [12].



Şekil 2.3 : Larinks anatomisinde kıkırdaklar [Url-3].

Eklemler

Krikotiroid ve krikoaritenoit eklemler larinkste bulunan iki çift sinovyal eklemdir.

Krikotiroid eklem: Aritenoit kıkırdakların tabanı ve krikoid lamina arasındadır. Bu eklem üç farklı hareketi vardır. Krikoid kıkırdak ile tiroid kıkırdağın alt kornusu arasında bulunan bu eklem birincil hareketi, tiroid kıkırdağın kayma ve rotasyon hareketidir:

- a) Rotasyon hareketi: Krikotiroid kastaki pars recta kısmının kasılması ile olur. Tiroid kıkırdak ile krikoid kıkırdak arasındaki açı daralmakta ve bundan dolayı vokal kordların uzunluğu ile gerginliği artarken kalınlığı azalmaktadır. Rotasyon hareketinde aritenoitler dikey eksen üzerinde lateral veya medyal rotasyon hareketi yapmaktadırlar. Lateral krikoaritenoit kasın kasılması sonrasında aritenoitler vokal çıkıntılar birbirlerine yaklaşacak şekilde (medyal), posterior krikoaritenoit kasın kasılması ile birbirlerinden uzaklaşacak şekilde (lateral) hareket etmektedirler. Bu hareketler sonucunda, vokal kordlar birbirlerine yaklaşmakta veya uzaklaşmaktadırlar.

- b) Eşik hareketi: Bu hareket de iki farklı doğrultuda görülebilmektedir. Tiroaritenoit kasın kasılması ile aritenoit kıkırdaklar 30°'lik bir aralıkta öne ve arkaya doğru hareket etmektedirler. İnteraritenoit kasın kasılması sonucu da aritenoitler birbirlerine yaklaşarak medyale doğru eğilme hareketi yapmaktadırlar.
- c) Kayma hareketi: Krikotiroid kasın pars obliqua kısmının kasılması ile meydana gelir. Krikotiroid eklemin sublükse olmasıyla tiroid kıkırdak öne veya medyale kaymaktadır. Tiroaritenoit kasın kasılması sonucu aritenoitler yaklaşık 2 mm öne kaymakta ve vokal kordlar kısalmaktadır. İnteraritenoit kasın kasılması sonucu da aritenoitler medyale kayarak birbirlerine yaklaşmaktadırlar [12].

Ligamanlar ve membranlar

Ligaman ve membranların bazıları ekstrensektir ve çevre dokulara larenksi bağlamaktadırlar. Bazıları da intrensektir, bunların da tamamı larinksin içindedir [12].

Kaslar

Ekstrensek ve intrensek olmak üzere ikiye ayrılmaktadırlar.

- a) Ekstrensek larenks kasları: larenksi çevre dokulara bağlamakta ve larenksin boyundaki pozisyonunu belirlemektedirler. Larenksin yukarı aşağı hareketi, kıkırdaklar arasındaki açılı ve vokal kordların gerginlik durumunu değiştirmektedir. Ayrıca intrensek kasların düzgün faaliyet gösterebilmesi için ekstrensek kaslar ile larenksin boyunda sabitlenmesi gerekmektedir.

Ekstrensek larenks kasları, suprahiyoit ve infrahiyoit olacak şekilde iki gruptur. Bu iki grup kasların birlikte uyumlu bir şekilde kasılması ile larenks boyunda sabitlenmektedir [12].

- b) İntrensek larenks kasları: Larenks işlevlerinden birinci derecede sorumludurlar ve vokal kordların hareketlerini etkilemektedirler.
- Abdüktör kaslar: vokal kordların birincil abdüktörü posterior krikoiaritenoit kasıdır. Bunun kasılması ile aritenoitler rotasyon hareketi yaparlar ve böylece vokal çıkıntılar birbirinden uzaklaşırlar. Glottis açılır, vokal kordlar birbirlerinden uzaklaşır, uzar ve gerginleşir.

- Addüktör kaslar: Koruma veya fonasyon işlevi için vokal kordları birbirlerine yaklaştıran kaslardır.
- Lateral krikoaritenoid kas: Birincil addüktördür ve kasılmasıyla aritenoidler rotasyon yapar, vokal çıkıntılar birbirlerine yaklaşır. Sonucunda membranöz glottis kapanır ve vokal kordlar uzayarak gerginleşir.
- Interaritenoid kas: Kasılması ile aritenoidler, birbirlerine yaklaşır ve kartilajinöz glottis kapanır.
- Vokal kordları geren kaslar: Sadece krikotiroid kas bu gruptadır. Oblik ve düz olarak iki parçası vardır. Oblik parça kasılınca krikotiroid ekleme sublüksasyon meydana gelir ve vokal kordlar alçalır, gerilerek incilir ve paramedyan pozisyona geçerler. Düz parça kasılınca ise vokal kordlar uzayarak gerilir. Krikotiroid kas ses perdesi ve şiddetinin düzenlenmesinde kritik roledir. Bu kasın kasılması ile ses perdesi incilir yani ses frekansı yükselir.
- Vokal kordları gevşeten kaslar: Tiroaritenoid kas bu gruptaki esas kastır. Eksternus ile internus olarak iki kısma ayrılmaktadır. İnternal kısmın lifleri vokal ligamanın hemen alt bölgesinde yer alarak aritenoidin vokal çıkıntısına yapışır, kastaki bu bölüme musculus vokalis denmektedir.

Tiroaritenoid kastaki kasılma aritenoidlerin öne doğru hareket etmesini ve ses kıvrımlarının kısalarak kalınlaşmasını sağlamaktadır. Bu esnada, gövde yani vokal kas tabakası sertleşmekte, örtü yani epitel ve Reinke boşluğu ve geçiş tabakası yani vokal ligaman gevşemektedir. Sonuç olarak ses perdesi kalınlaşmakta yani temel frekans düşmektedir [12].

Mukoza ve kıvrımlar

Mukoza larenksin iç yüzünü örtmektedir ve üzerinde ventriküler bant ve vokal kord olarak isimlendirilen iki çift kıvrım bulunmaktadır. Bu yapı ses fizyolojisi açısından kritiktir [12].

Vokal kordların anatomisi

Larenksteki vokal kordlar (titreşen elemanlar) insan sesi oluşumunda diğer bir komponenttir. Vokal kordlar sağ ve sol olmak üzere 2 tanedir ve ön tarafta tiroid kıkırdak iç yüzüne yapışarak ön komissürü oluşturur; arka tarafta ise aritenoid

kıkırdağa yapışan ve dinlenme durumunda ayrı duran, fonasyon sırasında yaklaşan bölümler arka komissürü oluşturur. Vokal kordların arasındaki açıklık rima glottis olarak adlandırılır. Bu bölge derin inspiryumda maksimum açıklıkta, normal solunum ve fısıltıda kısmi açıklıktadır; fonasyon sırasında ise neredeyse tamamen kapalı durumdadır.

Alt solunum yollarından basınçlı hava akımı gelir glottis seviyesinde kesilip kordların açılmasını sağlar. Bu esnada kordların titreşimi ile “Primer Glottik Ses” oluşur. Burada subglottik basıncın, vokal kordların kitlesinin, uzunluğunun, gerginliğinin ve mukoza mobilitesinin önemi büyüktür. Üst rezonatör ve artikülör yapılar ile modüle edilen primer glottik ses, harf, hece ve kelime olarak ağızdan çıkarak konuşma şeklinde çevreye yayılmaktadır [7].

Vokal kordlar erkeklerde yaklaşık 17-20 mm, kadınlarda ise yaklaşık 12-17 mm uzunluğundadır. Vokal kordun 2/3 ön kısmı membranöz yani vibratuar, 1/3 arka kısmı kartilajinöz yani nonvibratuardır. Membranöz kısım, fonasyon esnasında birincil olarak titreşim yapan bölge olmakla birlikte vokal kordların 2/3 ön kısmını oluşturmaktadır. Kartilajinöz kısım vokal kordların 1/3 arka kısmını oluşturmaktadır ve fonasyon esnasında titreşim yapmamaktadır. Vokal kordların ön 3/5 kısmına intermembranöz, 2/5 arka kısmına ise interkartifajinöz bölüm denilmektedir. Vokal kordların arasındaki açıklığa ise daha önce de belirtildiği gibi rima glottidis denilmektedir. Rima glottidis erkeklerde kadınlara oranla daha geniştir. Erkeklerde ortalama 23 mm olan bu açıklık kadınlarda ortalama 16 mm'dir. Glottisin, membranöz kısımları arasında kalan kısma anterior glottis yani fonatuar glottis, kartilajinöz kısımları arasında kalan kısmına ise posterior glottis yani respiratuar glottis denilmektedir. Vokal kordların yukarı kısmında bulunan ventriküler bant yani yalancı ses kıvrımlarının normal koşullarda fonasyonla bir ilgisi yoktur. Ancak hiperfonksiyonel disfonilerde intrensek larenks kaslarında oluşan gerginlik ile birbirlerine yaklaşabilmekte ve ventriküler fonasyon denilen ses problemine neden olabilmektedirler [12].

Histolojik olarak vokal kordlar mukoza ve kas dokusundan meydana gelmektedir. Mukoza, epitel ile lamina propria olarak ikiye ayrılmaktadır. Lamina propria dıştan içe doğru olacak şekilde sırasıyla yüzeysel, ara ve derin olarak üç tabaka şeklindedir. Çok az sayıda elastik ve kollajen liflerle yine çok az miktarda fibroblastı içeren yüzeysel tabaka gevşek bir dokudur. Bu tabakaya Reinke boşluğu da denilmektedir.

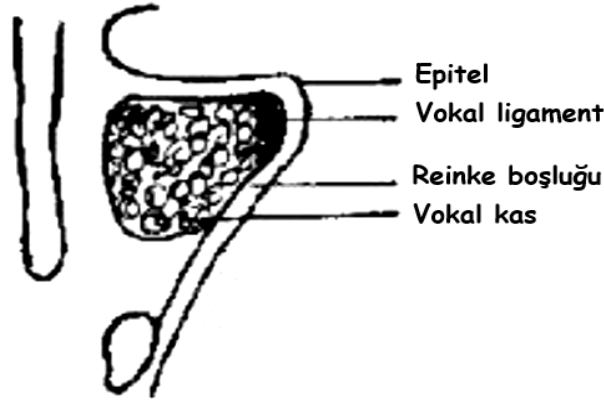
Vokal kordların yüzey kısmını çok katlı nonkeratinized bir yassı epitel örtmektedir. Histolojik olarak vokal kordların yapısı 5 tabakadır [7].

Bu tabakalar:

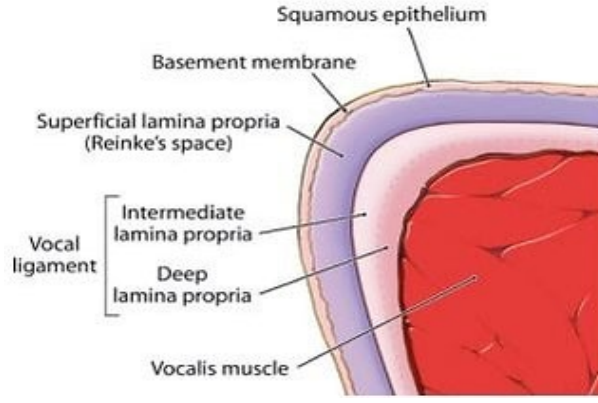
- a) Yassı epitel: Kordları koruma işlevi olan bir kabuk şeklindedir. Nonkeratinize çok katlı bir yassı epitelden oluşmaktadır.
- b) Lamina propria superfisialis: Hücreler arasındaki matriks ve gevşek bağ dokusu ile meydana gelmiştir ve jelatinöz özelliktedir. Buraya Reinke boşluğu da denilmektedir. Bu boşlukta gelişen patolojiler mukozal vibratuar dalgaları bozmak suretiyle disfoniye neden olmaktadır.
- c) Lamina propria medius: Lamina propria superfisialis'de bulunan gevşek bağ dokusu ile lamina propria profundus'da bulunan sıkı bağ dokusu arasındaki geçiş bölgesidir. Çoğunlukla elastik liflerden ve az miktarda da kollajen liflerden oluşmuştur. Fibroblastlarca zengin bir yapıdır.
- d) Lamina propria profundus: Kollajen liflerin çoğunlukta olması sayesinde sıkı bağ dokusundadır. Skar oluşumundan ara ve derin tabakasında bolca bulunan fibroblastlar sorumludur. Kist, polip ve benzeri lezyonların eksizyonu esnasında bu tabakaların travmatize edilmesi, skar oluşumuna yol açarak fonasyon işlevini bozabilmektedir [12].
- e) M. vokalis: Vokal kordun asıl yapısını oluşturmaktadır. Tiroaritenoid kasın medyal kısmıdır ve vokal kordun gövdesini oluşturmaktadır. Bu kasın kasılması vokal kordun adduksiyon hareketine yardımcı olmakla birlikte yine vokal kordun kısalıp kalınlaşmasında ve seviyesinin düşürülmesinde rol oynamaktadır. Vokal kas yapısında pek çok küçük kas demeti bulunmaktadır. Kas demetlerinin her biri birbirinden ayrı sinir lifi sayesinde kontrol edilmektedir ve bu sayede birbirlerinden ayrı hareket edebilmekte ve kasılabilmektedirler. Vokal ligaman ile larenks duvarına komşu lifler birbirlerinden ayrı hareket edip kasılabilmektedir. Bu nedenle farklı seslerin çıkarılması esnasında vokal kordların serbest kenarları kalın veya ince, keskin veya kunt hale geçebilmektedir [7].

Şekil 2.4 ve Şekil 2.5'te gösterilen bu tabakalı yapının pek çok önemi bulunmaktadır:

- Her bir tabaka mekanik açıdan farklı özellikler taşımaktadır.
- Vokal kasa kadar yer alan ilk dört tabaka mukoz yapıda olup pasif hareketlidir. Vokal kas ise pasif olarak da aktif olarak da hareket edebilmektedir.
- Vokal patolojiler her zaman için bu tabakalardan birinde bağımsız olarak başlamaktadır.
- Bu tabakaların vokal kasa yapışma dereceleri, seyri ve kalınlıkları vokal kord boyunca homojen durumda değildir. Bu tabakaların özelliklerine bağlı olarak yüzeyden gövdeye doğru vokal kordlar giderek artan sertlikte çok tabakalı birer vibratördürler. Vokal kordların vibratuar aktivitesinin çoğundan yüzey örtüsü sorumludur. Mukoza dalgaları fonasyon esnasında örtü tabakasının gövde üzerinde kayması sonucu oluşmaktadır. Mukoza dalgaları olmaz ise titreşim ile fonasyon da olmayacaktır. Düşük tonda çıkarılan seslerde ilk 4 tabakanın, yüksek tonda çıkarılan seslerde de tüm tabakaların işlevinin olduğu kabul edilmektedir. Vokal bölge sürekli şekil değiştirerek farklı dalgalar oluşturabilen bir bölgedir. Vokal patolojiler her zaman için bu tabakalardan birinden başlamaktadır. Artan yaşla orantılı olarak elastik lifler ile vokal kas atrofiye olmakta ve kollajenöz lifler kalınlaşmaktadır [7].



Şekil 2.4 : Vokal kord kesiti [13].



Şekil 2.5 : Vokal kord mikro anatomisi ve histolojisi [Url-4].

Laringeal sinir anatomisi

Larinksin motor ile duyuşal innervasyonu N.Vagustan ayrışan Rekküren Laringeal Sinir (Recurrent Laryngeal Nerve, RLN) ve Superior Laringeal Sinirden (Superior Laryngeal Nerve, SLN) sağlanmaktadır. SLN vagal gangliondan çıkarak superior servikal sempatik gangliondan dal alır. Farinks boyunca inerek iç ile dış laringeal sinirlere bölünür. Larinks mukozasının duyuşu vokal kord seviyesine kadar iç laringeal sinir tarafından sağlanır. Krikotiroid kası SLN'nin dış dalı innerve eder [14,10,15]. Superior laringeal sinir tarafından vokal kordların sensoriyal innervasyonu ve krikotiroid adalenin motor innervasyonu sağlanır. İnfior laringeal sinir (N.rekürrens) diđer bütün intrinsek larenks kaslarını innerve eder [16].

- a) Rekküren laringeal sinir anatomisi: Aortik arkların önünde bulunurlar. Embriyonik gelişim esnasında kalp ve büyük damarlar aşağı inip boyun uzadıkça aortik ark tarafından RLN'ler aşağı doğru çekilir [17]. RLN motor ve duyuşal bileşenlere sahiptir ve larinksin intrinsek kaslarını innerve etmektedir. Glottik larinksin duyuşal innervasyonundan sorumlu olan RLN, krikotiroid kas haricinde dört intrinsek laringeal kasa da motor işlev sağlamakta olan bir dış dal ve vokal kordlar ile subglottik bölgeye duyu sağlamakta olan iç dallara ayrılır [18].
- b) Non rekküren sinir anatomisi: Sağ ortak karotis ile sağ subklavian arasındaki dördüncü sıradaki sağ aortik arkın bir kısmı embriyolojik gelişim esnasında kaybolduğunda, ilkel arter halkalarında kırılma meydana gelebilmektedir. Halkadaki kırılma sonucu sol aortik ark oluşmakta ve sol subklavyen arterin altından sağ subklavyen arter çıkmaktadır [17]. Bu çıkış sebebiyle, sağ kola ulaşmak amacıyla sağ subklavyen arter özofagusun arkasında bulunan orta

çizgiyi geçebilmektedir [19]. Bu durum da özofagusun sıkışmasına ve disfajiye sebebiyet verebilmektedir. Bu atrezi sonucunda da sağ RLN'nin rekküren olarak boyundan yukarı çıkabilmesi için altından geçmesi gereken innominat arter yoktur. Bu sebeple sağ RLN servikal bölgede bulunan vagustan kaynaklanmaktadır [20].

- c) Superior Laringeal Sinir Anatomisi: Krikotiroid kasa Superior Laringeal Sinirin Eksternal Dalı (External Branch of the Superior Laryngeal Nerve, ESSLN) motor lifler göndermekte ve laringeal mukoza zarının bölümlerini innerve etmektedir [21].

Superior Laringeal Sinirin İç Dalı (Inner Branch of the Superior Laryngeal Nerve, IBSLN), Superior Tiroid Arteri (Superior Thyroid Artery, STA) ile birlikte tiroid membranı delip larinks mukozasının duysal innervasyonunu sağlamaktadır. IBSLN üç bölümde incelenmektedir: üst, orta ve alt bölümler [22]. Epiglotun laringeal yüzeyinin mukozasını üst bölüm innerve eder. Gerçek ve yalancı vokal kordların mukozası ile ariepiglottik kıvrımını orta bölüm innerve eder. Aritenoid bölgenin mukozasının, subglottisin, hipofarenksin ön duvarının ve üst özofagus sfinkterinin innervasyonundan ise alt bölüm sorumludur [23]. IBSLN'nin bazı liflerinin de interaritenoid kasa motor innervasyon sağladığı belirtilmiştir [24].

Larenksin damarlanması

Larenksin beslenmesi tiroid arterlerden (superior ve inferior) sağlanmaktadır. Superior lareneal arterin kaynaklandığı arter superior tiroid arterdir. İnfierior lareneal arterin kaynaklandığı arter inferior tiroid arterdir ve rekürren lareneal sinirle çaprazlaşmaktadır [13].

Larenksin lenfatik drenajı

Supraglottik bölge en yoğun lenfatik drenajın olduğu bölgedir. Glottik bölgedeki lenfatik drenaj çok azdır. Vokal kordun Reinke boşluğunda lenfatik drenaj yoktur. Yalnızca ön ile arka komissürlerde birkaç tane lenfatik kapiller bulunmaktadır. Subglottik bölgenin lenfatik drenajı altta bulunan derin servikal zincire doğrudur [13].

2.4.3 Vokal kordların fizyolojisi

Ses vokal kordda oluşan vibrasyondur. Sesin karakteristikleri (tını ve şiddet) vokal kord üzerindeki larengeal kasların etkileri ve hava akımının şiddetiy ile değişiklik göstermektedir. Sesin larenkste oluşmasında ekspiratuar hava basıncı, vokal kordların gerginlik durumu ve vibratuar özellikleri, solunum yollarının genişliği, rima glottidis şekli gibi çeşitli durumlar rol oynamaktadır.

Vokal kordlar ses oluşumunda addüksiyon, ekstansiyon ve vibrasyon olmak üzere üç temel hareket yapmaktadırlar. Addüksiyon, vokal kordların orta hatta gelmesidir. Ekstansiyon, orta hatta gelen vokal kordların gerginliğinin artmasıdır [13].

Vibrasyonun nasıl oluştuğu henüz tam açıklanamamıştır. Vokal kord vibrasyonlarına ilişkin teoriler şunlardır [7,25]:

- a) Myo-elastik aerodinamik teori
- b) Muko-elastik aerodinamik teori
- c) İki kitle modeli teorisi
- d) Bir kitle modeli teorisi
- e) Vücut kaplayan katman teorisi

Literatürde bu teoriler arasında en çok Myo-Elastik Aerodinamik Teori kabul görmektedir. Bu teoriye göre larenkesteki ses üretimi Bernoulli fenomenine dayanmaktadır. Teoriye göre, vokal kordların periyodik bir şekilde açılması ve kapanması, vokal kordların kütlesi ve gerginliği ile ekspiryum esnasında dışarıya atılan havanın oluşturduğu aerodinamik güçlerin bir etkileşimi sonucu oluşmaktadır.

Miyoeplastik aerodinamik teori daha detaylı olarak ele alınırsa:

1. Vokal kordlar addüksiyonla orta hatta gelir ve gergin bir şekilde orada durur.
2. Akciğerlerden gelen hava glottis kapalı olduğu için subglottik basıncı artırır. Larenks bir sfinkterdir ve fonasyon esnasında ekspiratuar basıncın etkisiyle ses kıvrımlarının gerginliğini yenecek düzeye gelen subglottik basınç ile vokal kordlar açılıp az miktarda hava da yukarıya supraglottise geçer.
3. Larenks kum saatine benzetilirse, kum saatinin dar bölgesi hava hareketinin olduğu bölgedir. Geniş bölgeler ise, aşağıda subglottis ile trakea, yukarıda supraglottis ile farinks olarak düşünülebilir.

4. Havanın glottisten hızlı geiři sırasında meydana gelen negatif basın Bernoulli etkisi nedeni ile vokal kordlarda emme etkisi oluřturur ve bu blgede basıncın dūřmesine sebep olarak glottisin kapanmasını saęlar.
5. Bernoulli etkisi ile birlikte vokal kordların elastikiyetiyle subglottik basıncın dūřmesi de glottisin tekrar kapanmasında rol oynamaktadır. Bu řekilde glottisin aılıp sonra tekrar kapanması arasında geen zamana glottal (vibratuar) siklus denilmektedir.

Art arda oluřan glottal siklus sayısı sesin temel frekansını belirlemektedir. Vokal kordlarda kiřinin normal fonasyonu esnasında bir saniyede oluřmakta olan vibrasyon sayısına temel frekans denilmektedir. Normalde erkeklerde 100-150 Hz bandında ortalama 130 Hz, kadınlarda ise 200-300 Hz bandında ortalama 250 Hz olan temel frekans fizyolojik faktrlere gre de ğiřmektedir.

Fonasyon sırasında glottisten hızla geen hava bu blgede basıncın azalmasına ve vokal kordların birbirine yaklařmasına yol amaktadır. Vokal kordların osilasyonları myoelastik yapılar ve aerodinamik kuvvetler tarafından oluřturulmaktadır. Derin inspiryumda kordlar birbirlerinden ayrılmakta ve rima geniřlemektedir. Bunu saęlayan kas posterior krikoaritenoid kastır. Ses oluřumu iin ramanın kapanması gerekmektedir. Bu duruma fonasyon durumu denilmektedir. Vokal kordların gerginlięi ile subglottik blge basıncında oluřan deęiřimler sesin oluřmasını saęlamaktadır. Vokal kord gerginlięinin artması sonucu daha yksek tonda ses oluřmaktadır. Subglottik basınta artma olması durumunda da sesin řiddeti artmaktadır. Sesin frekanstan sonra gelen en nemli zellięi de řiddetidir. Sesin řiddeti, havanın miktarına ve glottik seviyeden geme hızına baęlı olarak deęiřmektedir. Artan subglottik basın rimayı daha fazla amaktadır bu da canlı kapanmaya neden olmaktadır ve sonu olarak da sesin řiddeti artmaktadır. Heyecan faktrlerinin bulunmadıęı sakin durumlarda vcut ve beyin istirahatleyken O₂ ile CO₂ deęiřim hızı ve buna baęlı olarak da solunum frekansı ile derinlięi de dūřk olmaktadır. İstirahat durumundaki bu hava akımı gevřek ve esnek vokal kordları titreřtirerek kaliteli, frekans ve řiddetinin kontrol edilebildięi bir ses oluřurmaktadır [13].

Ses oluřması iin krikotiroid ve vokal kas vokal kordların gergin olmasını saęlamaktadır. Vokal kordların ok ok kısa aralıklarla yaptıkları ritmik hareketler

sonucunda rimaın açılıp kapanması rimadan geen hava akımında dalgalanmalar oluřturmaktadır. Rimadan geen hava dalgaları da kordlara arparak onları titreřtirmektedir. n kommisr ve aritenoid blgesine gre vokal kord mukozasının hareketlilięi kordların orta noktasında daha fazla olmaktadır. Vibrasyonu etkileyen birincil fiziksel zellikler kordun kitle ve katılıęıdır. Kitle ve katılıkta artıřa sebep olan btn patolojiler vibrasyonun amplitdnde, frekansında, simetrisinde ve sreklilięinde bozulmalara neden olmaktadır [13,7].

Rezonans ve artiklasyon: Supraglottik yapılar konuřma sesinin karakterini oluřturaktan sorumludur. Vokal kord vibrasyonu ile larenkte oluřan ses, yalnızca supraglottik vokal traktusta rezonans kazanır ve artikle edilirse konuřma haline gelebilmektedir. Supraglottik larenks, farenks, paranasal sinsler, nazal ve oral kavitede primer glottik ses rezonansa uğrar ve bu sayede bazı frekanslarda gçlenirken, bazılarında da sner. Damak, dil ve dudaęın dinamik hareketleri ile ses artiklasyonu tamamlanır ve konuřma ortaya ıkar [7].

2.5 Ses fiziyojisi

Ses kelimesi ile iki farklı durum anlatılmaktadır:

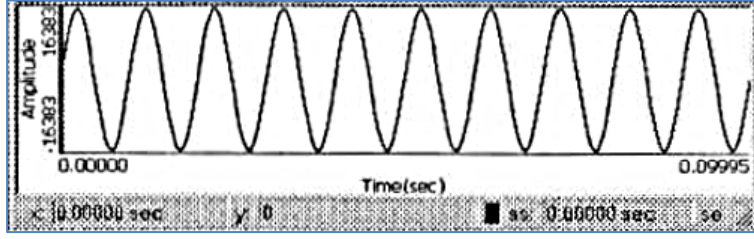
1. konuřma sesinin ham maddesi
2. kulak tarafından algılanabilen her trl titreřim

İnsan sesinin oluřumunda nefesli bir mzik enstrmanında olduęu gibi gç kaynaęı, vibratr ve rezonatr gerekmektedir. Ekspiratuvar hava akımını saęlayan diyafram, abdominal kaslar, torakal kaslar ile alt solunum yolları gç kaynaęı grevini grmektedir. Vokal kordlar vibratr grevini grmektedir. Supraglottik larenks, larinks, aęız ve burun bořluęu rezonatr grevini grmektedir [7].

2.5.1 Ses fizięi ile ilgili temel kavramlar

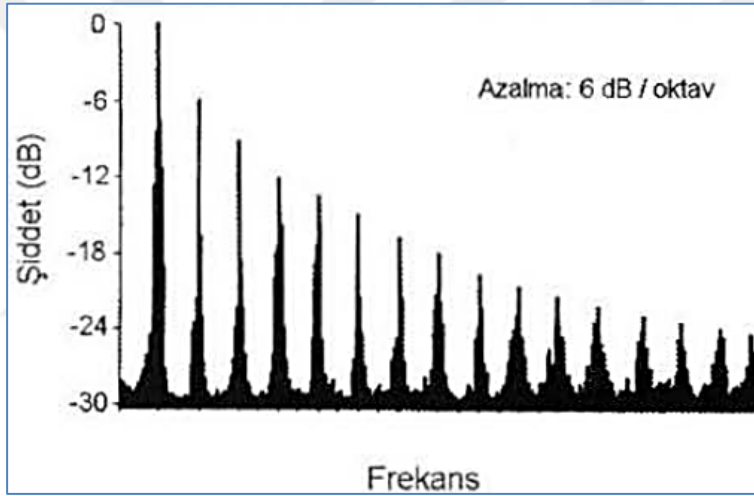
Havada ve dięer ortamlarda ses sıkıřma-seyrelme dalgaları olarak ilerler. Sıkıřma, basıntaki artıřı gsterirken, seyrelme ise basıntaki azalmayı gsterir. Dalga rneęinin tekrar ediyorsa periyodik ses, etmiyorsa aperiyodik ses olarak adlandırılır. Grlt sesleri aperiyodik sesler iken; diyaazon sesi, insan sesi gibi sesler aynı dalga rneęinin tekrar edildięi periyodik seslerdir. Bir saniyede iinde tekrar eden dalga rneęi sayısı o sesin frekansını verir, birimi Hertz (Hz)'tir [7]. Diyaazon sesi

gibi saf yani tek frekansı olan seslerde dalga şekli sinüs eğrisi şeklindedir (Şekil 2.6) [12].



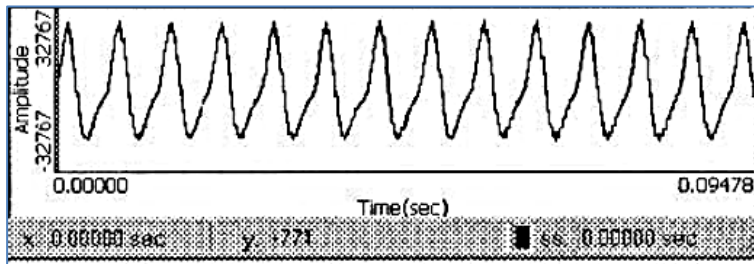
Şekil 2.6 : Bilgisayar ortamında üretilmiş frekansı 100 Hz olan saf sese ait dalga şekli [12].

İnsan sesi gibi doğadaki seslerin çoğu farklı frekans ve şiddetteki birçok sesin birleşmesi ile meydana gelmektedir. Glottis düzeyinde oluşan ses, bir temel frekans ile bu temel frekansın harmonik adı verilen katlarından oluşmaktadır (Şekil 2.7) [12].

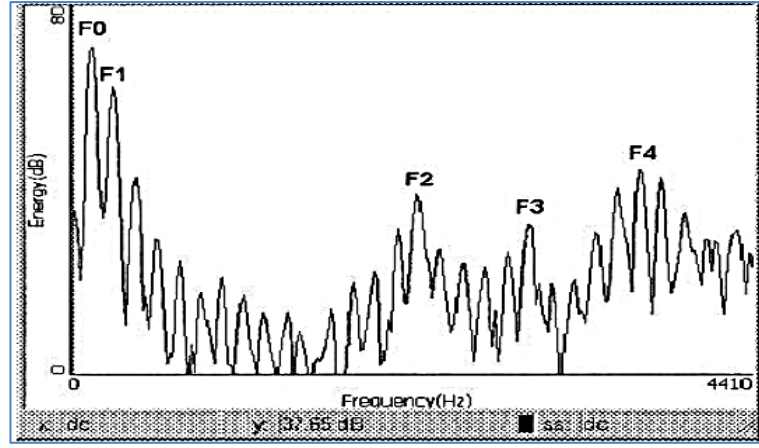


Şekil 2.7 : Glottis düzeyindeki işlenmemiş sesin spektrumu [12].

Şekil 2.8'de [i] sesinin dalga şekli, Şekil 2.9'da ise bu sesin Hızlı Fourier Dönüşümü (HFD) alınarak çıkarılmış frekans bileşenleri (harmonikleri) gösterilmektedir [12].



Şekil 2.8 : [i] sesine ait dalga şekli [12].



Şekil 2.9 : [i] sesine ait HFD. F0, temel frekans; F1, birinci formant; F2, ikinci formant; F3, üçüncü formant, F4, dördüncü formant [12].

Örneğin, temel frekansı 120 Hz olan bir sesin frekans bileşenleri (harmonikleri) 240, 360, 480... Hz şeklindedir. Şiddeti en yüksek olan harmonik temel frekanstır. Frekansın her ikiye katlanmasında (oktav başına) ses şiddeti 6 dB azalmaktadır. Aynı örnekte temel frekansın şiddetinin 60 dB olduğunu varsayarsak, iki katı frekansa sahip olan birinci harmonik 240 Hz'tir ve şiddeti 54 dB'dir, üçüncü harmonik 480 Hz'tir ve şiddeti 48 dB'dir. Burada bahsedilen özellikler glottis düzeyindeki sesle ilgilidir. Yalnızca ters filtreleme gibi yöntemlerle ses yolunun filtreleme özelliği ortadan kaldırılırsa gösterilebilir. Ses yolundaki rezonatör etki nedeniyle ağızdan çıkan ses değişikliğe uğramaktadır bu nedenle de akustik özellikleri çok daha farklıdır. Rezonatör etki sebebiyle glottis düzeyinde oluşan sesin bazı harmoniklerinin şiddeti artmakta iken bazılarınınki azalmaktadır. Şiddeti artan harmonikler formantları oluşturmaktadır.

Oktav: Bir müzik terimidir ve sekizli anlamına gelmektedir. Perte üzerine yazılan herhangi bir nota ilk nota olarak kabul edilirse ve tiz yöne doğru notalar sayılarak çıkılırsa, sayılan sekizinci nota ilk notanın oktavı (sekizlisi)dir. Frekansın ikiye katlanması da bunun fiziksel karşılığıdır.

Semiton: Bir oktavda sekiz nota ve yedi aralık bulunmaktadır. Bu aralıklardan beşi tam, ikisi yarıdır. Bu yarım aralığa semiton denilmektedir.

2.5.2 Sesin özellikleri

Sesin dört özelliği vardır, bunlar aşağıda açıklanmıştır:

1. Perde: Perde algısal bir terimdir ve sesin ince veya kalın oluşunu bildirir. İki perde arasındaki aralığa semiton denilmektedir. Frekans perdenin fiziksel

karşılığdır. Temel frekans (vokal kordların saniyedeki titreşim sayısı) genel anlamda insan sesinin frekansı denildiğinde akla ilk gelendir. Bu titreşim sayısının değişiklik miktarı şarkı söylerken büyükken konuşurken daha azdır. Konuşma sesinde ortalama temel frekans erkeklerde yaklaşık 125 Hz iken kadınlarda yaklaşık 215 Hz'dir.

2. Şiddet: Birimi dB'dir. Ses düzey ölçeri ile ölçüm yapıldığında insan sesinin şiddeti dB(A) şeklinde gösterilir.
3. Kalite: Solunum organları ve vokal kordların birlikte bir uyum içinde çalışması sonuç olarak da supraglottik bölgede hava türbülansı oluşmayacak biçimde eşit aralıklarla ve düzgün bir şekilde vokal kordların titreşmesidir. Anormal ses kalitesi algısal karşılığı, ses kısıklığı, ses düzensizliği ve solukluluk; fiziksel karşılığı, frekans pertürbasyonu (jitter), amplitüt pertürbasyonu (shimmer), NNE, HNR gibi akustik parametrelerin normalden yüksek olmasıdır.
4. Rezonans: Glottis düzeyinde oluşan ses geçtiği boşluklarda (farinks, ağız) boşlukların hacmine ve boşluk duvarlarının gerginliğine bağlı olarak değişime uğrar. Bazı frekans bölgelerinde rezonatör organların şekline göre ses şiddeti artarak formantlar oluşur [12].

2.5.3 Ses perdesi ve şiddetinin belirlenmesi

Konuşma sırasında sesin perdesi, şiddeti, kalitesi intrensek larenks kaslarının kasılması ile değişmektedir. Ses perdesini belirleyen titreşen kenarın kalınlığı ve ses kıvrımlarının gerginlik durumudur. Vokal kordların gerginliği artar ve kalınlığı azalırse perde incelikir. Buna ek olarak nefesli enstrümanlardaki gibi subglottik basınç artarsa da ses perdesi incelikir. Subglottik basınç ile glottik direnç artarsa ses şiddeti artar. Ses şiddetinin sabit kalabilmesi için subglottik basınç sabit olmalıdır [26].

2.5.4 Ses aralığı ve rejisterler

Bir kişiden çıkabilen en kalın ses ile en ince ses arasındaki aralığa ses aralığı denir.

Farklı frekanslarda vokal kordlar farklı titreşim yaparlar. Ses aralığında, aynı titreşimi yapan ve aynı kalitede olan seslerin oluşturduğu gruba rejister denilir.

Vokal kordlarda fonasyon sırasındaki davranışlarına göre üç farklı rejisterden bahsedilir. Kalından inceye doğru bu rejisterler, cızırtı rejisteri, modal rejister ve falsetto rejisteridir. Sınırlarda üst üste binmeler olsa da her rejisterin kendine özgü frekans bölgesi vardır.

Cızırtı rejisterinin frekans aralığı yaklaşık olarak 25-80 Hz arasındadır. Modal rejisterin frekans aralığı erkeklerde yaklaşık olarak 75 - 450 Hz, kadınlarda ise yaklaşık olarak 130 - 520 Hz arasında olmakla birlikte şarkı ve konuşma seslerinin bulunduğu rejisterdir. Falsetto rejisterin frekans aralığı erkeklerde yaklaşık olarak 275-620 Hz, kadınlarda ise yaklaşık olarak 490 - 1130 Hz arasındadır [12].

2.6 Disfoni

2.6.1 Disfonin klinikopatolojik özellikleri

Disfoni, sesi korumada zorluk, ses yorgunluğu, temel frekansta değişiklik, ses kısıklığı, ses hacmi eksikliği, ses etkinliği kaybı ve konuşma esnasında güçsüzlük gibi semptomlarla kendini göstermektedir.

2.6.2 Vokal kord patolojileri ve disfoni

Vokal kord hareket bozuklukları

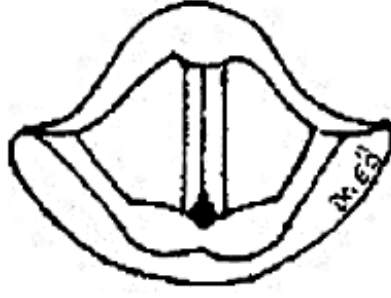
Vokal kordların açılmasında kısıtlılık var ise sınırlı abduksiyon, kısıtlılık kapanmasında var ise sınırlı adduksiyon vardır. Vokal kordlardan biri veya ikisi de hiç hareket etmiyor ise vokal kord fiksasyonu vardır [13].

Periyodisite bozuklukları

Vibratuvar siklusun düzenliliği durumuna periyodisite denir. Vibrasyon süresinin siklustan siklusa stabilitesini ifade eder. Vokal kordların mekanik özellikleri ile subglottik hava akımında oluşan herhangi bir değişiklik aperiodyositeye neden olur [13].

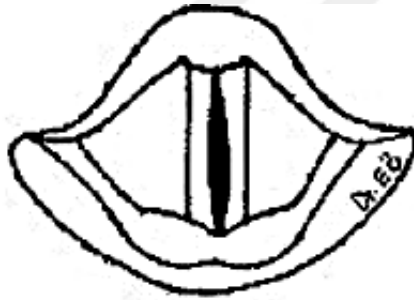
Glottik kapanma sorunu

Fonasyon esnasında glottisin normalde tüm uzunluk boyunca tam bir şekilde kapanması gerekmektedir. Glottisin ön kısmında fonasyon sırasında açıklık kalıyorsa buna anterior açıklık denilir. Glottisin posteriorunda fonasyon sırasında açıklık kalması genellikle kadınlarda normal olarak değerlendirilmektedir (Şekil 2.10) [13].



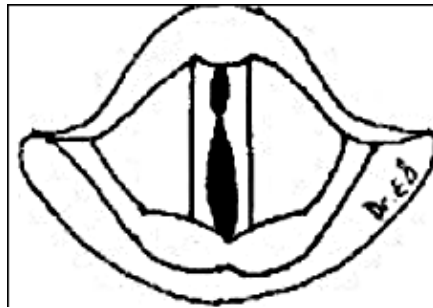
Şekil 2.10 : Fonasyon sırasında posterior glottik açıklık [13].

Endolarengeal mikrocerrahi sonrası oluşan skar formasyonuna bağlı olarak veya yaşlılık nedeniyle bilateral olacak şekilde vokal kas kitlesinin azalması nedeniyle vokal kordlarda anterior ve posteriorde kapanma olurken ortada bir açıklık kalabilir. Bu duruma bowing tipi glottik kapanma yetmezliği denir (Şekil 2.11) [13].



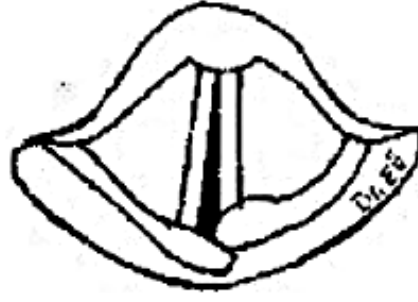
Şekil 2.11 : Fonasyon sırasında bowing tipi glottik kapanma yetmezliği [13].

Fonasyon sırasında hem anteriorda hem de posteriorde açıklık kalıyor ve sadece orta bölümde kapanma oluyorsa buna kum saati şeklinde açıklık denir. Vokal kordlardan polip, kist veya nodül bulunması halinde bu tür açıklık oluşabilir (Şekil 2.12) [13].



Şekil 2.12 : Fonasyon sırasında kum saati şeklinde açıklık [13].

Hastalıklı vokal kord orta hatta doğru kapanamıyorsa ve sağlam vokal kordun da kompanzasyonu yeterli olamıyorsa glottis boyunca fonasyon sırasında açıklık kalır. Bu duruma tek taraflı vokal kord paralizisi denir (Şekil 2.13) [13].



Şekil 2.13 : Tek taraflı vokal kord paralizisi [13].

Non-vibratuar bölgelerin bulunması (adinamik segmentler)

Normalde iki vokal kordun da tüm membranöz kısımları vibrasyona katılmaktadır ve hiçbir yerde adinamik segment bulunmamaktadır. Videolarinoskopi (VLS) muayenesinde adinamik segment tespit edilirse bu epitelyum ile lamina proprianın anatomisinde bozulmaya neden olan ciddi vokal kord lezyonlarının işaretidir [13].

Horizontal hareket amplitüdü

Vibratuar siklusun maksimal açılma fazında vokal kordun serbest kenarının orta hattın uzaklığını amplitüd ifade eder. Başka bir deyişle vibrasyon esnasında vokal kordun orta membranöz bölgesinin laterale doğru maksimum hareketinin miktarıdır ve normal olarak her iki kord için de aynı olmalıdır. Vokal kordların arasında amplitüd farkı olması asimetrik vibrasyonu işaret eder. Horizontal hareket amplitüdü sıfır, azalmış, normal, ve artmış olarak değerlendirilmektedir. Vokal kordun kitlesi ve/veya sertliğinde artış olursa ve vibratuar segment kısalığında azalmış amplitüd görülür (karsinom, papillom vs). Yüksek sesli fonasyon sırasında (subglottik basınç arttığında) amplitüd artar. Spasmodik disfoni gibi çok sıkı glottik kapanma durumunda amplitüd azalır [13].

Mukozal dalga anomalileri

Kordun superior yüzünde görüntülenen mukozal dalga ilerlemesi vokal kord yüzeyi boyunca mediolateral tarafta bir kayma hareketidir. Lamina proprianın yüzeysel tabakasını dolduran, vokal ligamanı infiltrate eden, vokal ligamanı iten lezyonlar mukozal dalgayı sınırlayabilir ya da yok edebilir. Artan subglottik basınç ile beraber ve ciddi sıvı içeriği bulunan lezyonlarda mukozal dalga da artmaktadır [13].

Vokal kord serbest kenarı düzensizlikleri

Mukoza hasarına neden olan tüm lezyonlarda vokal kordun serbest kenarı düzenli değildir. Vokal kordların serbest kenarlarının yanı sıra üst yüzeyleri de vibrasyonu etkileyen çeşitli lezyonları bulundurabilir (sulkus vokalis, lökoplazi vs.) [13].

Supraglottik vibrasyon

Normalde supraglottik yapılarda fonasyon esnasında vibrasyon görülmemektedir. Vokal kordların vibrasyonu bozulduğunda (kordektomi, stripping vs) bu yapıların titreşmesi kompensatuar bir mekanizma olarak görülmektedir. Buna ek olarak VLS muayenesinde fonasyon esnasında supraglottik yapılarda antero-posterior tarafta ve ventriküler bantlarda lateromedial tarafta kontraksiyon görülmesi hiperfonksiyonel patolojilerin işaretidir [13].

Vokal kord polipleri

Ses bozukluğuna neden olan ve sıklıkla görülen bir hastalık grubu olan bu lezyonlar genellikle ses kısıklığı olarak ortaya çıkarlar, ses kalitesinde bozulmaya neden olurlar. Vokal kord poliplerinin oluşumunda sesin aşırı ve/veya yanlış kullanımı birincil faktörlerdendir. Hastaların %80'i sigara içenlerden oluşmaktadır. Daha çok erkeklerde rastlanmaktadır.

Genellikle unilateral olarak görülen polipler bilateral de olabilmektedir ve bilateral olan polipler genellikle asimetriktirler. Genellikle lamina proprianın süperfisiyel tabakasında bulunan polipler hemorajik, ödematöz ve fibröz şekillerdedir. Poliplerin lokalizasyonu nodüllere göre değişik varyasyonlarda olsa da, sıklıkla vokal kordlarda membranöz kısımda orta bölümünün serbest kenarında yerleşik olarak görülürler çünkü bu bölgeler irritasyona en sık maruz kalan bölgelerdir.

Polipler geniş tabanlı veya saplı olabilmektedirler. Saplı olanlar serbest hareketli olabileceklerinden dolayı pozisyon değişikliğinde seste ani değişimler veya havayolu semptomları görülebilir.

Vokal kordun üst yüzünde bulunan polipler ses kalitesini fazla etkilemeyebilir. Ancak vokal kordların birbirlerine temas eden iç kenarlarına yerleşmiş polipler vibrasyon esnasında tam kapanmayı engelleyerek disfoniye sebep olurlar. Bu durumda ön ve arkada açıklık kalacağından dolayı kısık ve havalı özellikte bir ses oluşur. Saplı polipler hareket özellikleri sayesinde sesi daha az bozarlar. Glottisi

tıkayacak kadar yer kaplamayan bir polip ise tek semptom ses kısıklığı olabilmektedir. Buna ek olarak hasta ses tonunu hızlı ve heyecanlı konuşmalar sırasında sabit tutamaz, polipler vokal korda göre daha geç hareket ettiğinden dolayı diplofoni oluşur. Bazı hastalarda da başvuru semptomu boğazda gıcıklanma hissi olmaktadır [13].

2.6.3 Disfonide kullanılan tanı yöntemleri

2.6.3.1 Subjektif ses analizi

Subjektif ses analizi hastanın kendisi tarafından yapılan değerlendirmeler ile gerçekleştirilir. Subjektif ses analizinin gerçekleştirilebilmesi amacıyla en yaygın kullanılan yöntemler Ses Engellilik İndeksi (Voice Handicap Index, VHI) ve Ses Performans Anketidir (Vocal Performance Questionnaire, VPQ). Ayrıca derece, pürüzlülük, nefes alma, asteni, zorlanma ölçeği (Grade, Roughness, Breathiness, Asthenia, Strain, GRBAS Scale) ve Uzlaşmış İşitsel Algısal Ses Değerlendirmesi (Consensus Auditory-Perceptual Evaluation of Voice, CAPE-V) testleri de kullanılmaktadır.

- Ses Performans Anketi: Anket şeklinde bir analiz yöntemi olan VPQ testi 12 sorudan oluşmaktadır. Ses şikayetinin günlük yaşamdaki etkilerini ve şiddetini ölçmek amacıyla hastalara ameliyat öncesi ve ameliyat sonrası 1. ayda olmak üzere 2 kez uygulanmaktadır. Testten alınacak en yüksek puan 60, en düşük puan ise 12 olmaktadır [27].
- Ses Engellilik İndeksi: Bu indeks, bireyin kendi kendine algıladığı ses bozukluğunun öznel bir indeksidir ve ses değişikliğinin geçerli ve güvenilir bir göstergesi olduğu bildirilmektedir. Anketin amacı farklı patolojileri birbirinden ayırmak olmayıp, hastanın kendi sorununu kendisinin değerlendirmesini sağlamaktır. VHI anketi, hastanın kendi ses bozukluğunun günlük aktiviteleri üzerindeki göreceli etkisi hakkındaki yargısını değerlendirmektedir. Bu nedenle VHI, hastanın sesiyle ilgili şikayetini yansıtmaktadır ve bu da hastanın bir ses kliniğine profesyonel yardım amacıyla başvurma nedenidir. VHI, ses problemleriyle ilgili fiziksel, işlevsel ve duygusal faktörleri ölçmek için tasarlanmıştır ve bu faktörler için her biri 10'ar maddeden oluşan üç alt grubu vardır. VHI'da her yanıt 0'dan 4'e kadar derecelendirilmektedir. Toplam skor maksimum 120'dir ve bu skor ne kadar

yüksekse sesle ilgili sorun da o kadar büyüktür. Yüksek seviyeler daha fazla işlev bozukluğunu göstermektedir ve her 18 puanlık artış genellikle klinik olarak anlamlı kabul edilmektedir. İstatistiksel olarak sağlamdır, yüksek bir iç tutarlılığa sahiptir ve test-tekrar test kararlılığı güçlüdür [28].

- GRBAS ölçeği: Bu ölçek, sesin algısal yönden değerlendirilebilmesi amacıyla geliştirilmiş en eski yöntemdir [29]. Bu ölçekte derece (Grade, G), genel olarak algılanan disfoni derecesidir; pürüzlülük (Roughness, R), temel frekansın düzensiz dalgalanmasıdır; nefes alma durumu (Breathiness, B) yetersiz glottik kapatma yoluyla havanın sızmasına bağlı türbülanstır; asteni (Asthenia, A) sesin zayıflığı ve zorlanması (Strain, S) algılanan aşırı çabadır. Her parametre, 0'ın normal, 1'in hafif rahatsızlık, 2'nin orta derecede rahatsızlık ve 3'ün şiddetli rahatsızlık olduğu 0'dan 3'e kadar bir skala kullanılarak puanlanmaktadır [30].
- Uzlaşmış İşitsel Algısal Ses Değerlendirmesi: Bu değerlendirme ses örnekleri elde etmek ve ses kalitelerini derecelendirmek amacıyla uzman görüş birliği ve sesin işitsel algısal niteliklerinin psikofiziksel olarak uygun ölçümüne dayanan standart bir ses protokolü olarak geliştirilmiştir [31]. CAPE-V, dinleyicilerin sesin çeşitli algısal özelliklerini değerlendirdiği bir dizi 100 mm'lik görsel analog skaladır. Algısal ses özellikleri, disfoninin derecesi veya genel şiddeti, pürüzlülük, nefes alma, gerginlik, perde ve ses yüksekliğini içermektedir. Her ses özelliği için puanlar, ses işlev bozukluğunun ciddiyetini yansıtmaktadır [32].

2.6.3.2 Objektif ses analizi

Periyodik bir ses üç temel özelliğe sahiptir [33].

- Periyot: Sinyalde birbirini takip eden iki geçiş arasındaki süreye periyot denir ve milisaniye ile ölçülmektedir.
- Frekans: Saniyedeki periyod sayısıdır ve Hertz olarak ölçülmektedir.
- Amplitüd: Sesin şiddetidir ve insan sesinde amplitüd dB(A) olarak ölçülmektedir.

Objektif ses analizinin yapılabilmesi için yaygın olarak hastalardan alınan ses örneğinin çeşitli yazılımlar kullanılarak incelenmesi ile akustik ses analizi gerçekleştirilmektedir.

Akustik ses analizi aşağıda açıklanan parametreler ile yapılmaktadır.

- Temel frekans (F_0): Vokal kordların bir saniyedeki titreşim sayısıdır ve Hertz olarak ifade edilmektedir. Vokal kordlarda oluşan vibrasyonun hızının değişmesi ile temel frekans değişmektedir.
- Frekans pertürbasyonu parametreleri: Düz fonasyonda ideal şartlarda frekansta değişiklik olmaması gerekir. Fakat fonasyonda temel frekansta istem dışı değişiklikler görülmektedir. Bu düzensizliklere “frekans perturbasyonu” veya “jitter” denir.
 - Mutlak jitter: Ardışık periyotlar arasında ölçülen mutlak farkın ortalaması mutlak jitteri verir ve mikrosaniyeler ile ölçülmektedir.
 - Jitter (%): İki ardışık periyot arasındaki mutlak farkın, ortalama periyot sayısına bölünmesi ile de Jitter (%) elde edilir. Mutlak jitter temel frekansa bağlı değişiklik gösterebilir bu durumu ortadan kaldırmak amacıyla kullanılmaktadır. Jitter (%), ses örneğinde perdedeki kısa zamanlı değişikliklerin göreceli bir değerlendirmesini vermektedir.
- Amplitüd perturbasyonları (Shimmer): Amplitüdde sikluslar arası kısa zamanlı pertürbasyonları göstermektedir ve iki şekilde ifade edilmektedir.
 - Shimmer (dB), ardışık periyotlar arasındaki amplitüd farkının ortalamasının 10. dereceden logaritmasının 20 ile çarpılması sonucu elde edilmekte ve dB ile ifade edilmektedir. Her bir periyodun tepe amplitüdü takip eden periyodun tepe amplitüdü ile karşılaştırılarak dB cinsinden hesaplanmaktadır.
 - Shimmer (%): birbirini takip eden periyotların amplitüdüleri farkının ortalamasının, amplitüdün ortalamasına bölünmesi sonucu elde edilmektedir. Bu değer ile analiz edilen sesin pik değerleri arasındaki kısa zamanlı değişimlerin göreceli bir değerlendirmesi yapılabilmektedir.
- Spektral parametreler: Ses spektrumunda bulunan, temel frekansın katlarından meydana gelen harmonikler ve harmonikler dışında kalan gürültü

sesleriyle ilgili parametrelerdir. Kısa süreli periyot ve amplitüd değişimlerine ek olarak, vokal kordların kapanmaması veya değişken yapısı glottis boyunca hava kaçağına sebep olmaktadır ve bu halde akustik karakter gürültü olarak adlandırılmaktadır.

- Harmonik gürültü oranı: harmonik spektral enerjinin harmonik olmayan enerjiye oranı bu değeri vermektedir. Analiz edilen sinyalin genel bir değerlendirmesidir. HNR amplitüdden ve frekanstaki pertürbasyonlardan etkilenmektedir. Bu ölçüm sinyalin gürültülüğü ve kabalığının en iyi korelasyonunu gösteren parametredir.
- Normalize gürültü enerjisi: Total vokal enerjiden harmonik enerjinin çıkarılması sonucu elde edilmektedir.
- Diğer parametreler:
 - Maksimum Fonasyon Süresi: Bir nefeste en uzun çıkartılabilecek ses süresidir. Yaş, cinsiyet, akciğer performansı ve ek hastalıkların olup olmamasına bağlı olarak değişmektedir ve ortalama değerler erkeklerde yaklaşık 20 saniye, kadınlarda yaklaşık 15 saniye ve çocuklarda ise yaklaşık 10 saniyedir.
 - S/Z Oranı: Bir nefeste maksimum uzunlukta çıkartılabilecek “s” sessiz harfinin süresinin “z” sessiz harfinin süresine oranı bu değeri vermektedir. Sağlıklı kişilerde bu sürelerin yaklaşık olarak birbirine eşit olması, bu durumda da oranının yaklaşık 1 olması beklenmektedir. Vokal kord vibrasyonunun bozulması ya da glottal kapanmanın tam gerçekleşmediği durumlarda z süresinin azalması ve sonuç olarak da S/Z oranının artması beklenmektedir.
 - Ses kısıklığı ve sesteki kabalaşmalar: Glottisin tam kapanmaması sonucu oluşan hava akımı türbülansı veya vibrasyondaki düzensizlikten dolayı ortaya çıkmaktadır. Başka bir deyişle vokal kordun vibratuar özelliklerinde bozulma sonucu ses kısıklığı ve seste kabalaşma meydana gelmektedir.
 - Havalı ses: Fonasyon esnasında hava kaçıışı olur ve dışarıdan da duyulabilmektedir. Ses net değildir. Tam olarak gerçekleşemeyen vokal kord kapanmasına sebep olan nörolojik sorunlarda (periferik ve santral) veya kitlesel lezyonların olması durumunda ortaya çıkmaktadır.

- Ötümsüz Zaman Yüzdesi: Total fonasyon zamanı içindeki ötümsüz zamanın yüzdesidir.
- Frekans Değişim Aralığı: Hem habitüel ton ve frekansta hemde sesin en pesten en tize çıkarılması sırasında en düşük ile en yüksek frekanslar arasında Hz ya da semitone cinsinden farktır.
- Tremor: Kişinin kontrolü dışında çıkan ritmik ses perdesi ile ses yüksekliği değişikliklerini tanımlar. Genellikle santral konuşma merkezini etkileyen nörolojik bozuklukların olması halinde görülmektedir [13].

2.6.3.3 Videolaringoskopi

Normal ses oluşumunun temelini vokal kordların vibratuar kenarındaki kompleks hareket teşkil etmektedir. Vokal kordlar fonasyon perdesine bağlı olarak saniyede 60 ile 1500 kez açılıp kapanmaktadır. Bir objenin hızlı periyodik hareketinin hareketsiz ya da yavaş hareket eden şekildeki optik görüntüsünü Stroboskopi ifade etmektedir. Retinaya gelen bir görüntü 0.2 saniye kalıcı olabilir. Görüntüler 0.2 saniyeden daha sık aralıklarla gelirse, bu görüntüler hareketli gibi algılanırlar [13]. Stroboskopide vokal kordun titreşim frekansı ve illüminasyon frekansı arasındaki senkronizasyon, kişinin boynuna yerleştirilen mikrofonla sağlanmaktadır.

Videolaringostroboskopik değerlendirmede kullanılan parametreler şunlardır [34,13]:

- Vokal kord hareketleri
- Periyodisite
- Glottik kapanma
- Non-vibratuar bölgeler (adinamik segmentler)
- Horizontal hareket amplitüdü
- Mukozal dalga
- Vokal kord serbest kenarı
- Supraglottik değerlendirme

2.6.3.4 Elektromiyografi

İntrinsik ve ekstrinsik laringeal kasların Elektromiyografisi (EMG) yapılarak RLN ve/veya SLN hasarlarında tanı konulup, prognostik bilgi sağlanabilmektedir. Ancak bu prosedür invazivliği nedeniyle klinik olarak nadiren tercih edilmektedir. Krikotiroid ve tiroaritenoid kaslar bu yöntemde değerlendirilmesi gereken temel kaslardır [35]. Disfonili olgularda kullanılan bir diğer tanı aracı olan fiber optik esnek endoskopi, vokal kordların mükemmel şekilde görüntülenmesini ve solunum ve fonasyon egzersizleri yapan hastaların gözlemlenmesini sağlamaktadır [36]. Ayrıca diğer bir tanı yöntemi olan laringeal elektromiyografi, ESSLN yaralanması şüphesinde yaralanmanın teşhisi için altın standarttır [22].

3. KURAMSAL TEMELLER

3.1 Dalgacık Eşiği Gürültü Giderme Algoritması

Dalgacık eşiği gürültü giderme algoritması, dalgacık algoritmasının önemli bir parçasıdır. Algoritma Eşitlik 3.1’de gösterilmiştir.

$$f(t) = s(t) + n(t) \quad (3.1)$$

Denklem, gürültülü tek boyutlu bir sinyal modelidir, burada $f(t)$ gürültülü sinyaldir, $s(t)$ faydalı sinyaldir ve $n(t)$, $N(0, \sigma^2)$ dağılımının Gauss beyaz gürültüsüne uyan gürültüdür [37,38].

Algoritma kullanılırken öncelikle dalgacık tabanı ve dalgacık dönüşüm katmanı için uygun değerler seçilmektedir. Ortalama kısmını ve detay kısmını elde etmek için $f(t)$ üzerinde dalgacık ayrıştırması yapılmaktadır. Daha sonra, her katman için uygun eşik değeri ve eşik işlevi seçilmekte ve her katmanın tahmini dalgacık katsayıları elde edilmektedir. Daha sonra, orijinal yararlı sinyal $s(t)$ 'nin tahmini değerini elde etmek için tüm katmanların ölçek faktörü ve dalgacık katsayıları yeniden yapılandırılmaktadır.

Dalgacık dönüşümü temeli nedeniyle, dalgacık seçimi sabit değildir ve tüm sinyal türleri üzerinde iyi bir gürültü giderme etkisine sahip olabilecek dalgacık tabanlı bir işlev yoktur. Bu nedenle, dalgacık fonksiyonunun kalitesi, dalgacık analiz yönteminin konuşma sinyali işleme sonuçları ile teorik sonuçlar arasındaki hataya göre belirlenmektedir ve dalgacık temel işlevi buna göre seçilmektedir. Eşik değerinin seçimi, eşik boyutu tahmininin doğruluğa önemli ölçüde bağlıdır. Tahmin edilen eşik değeri çok küçükse, sinyallerdeki gürültü tamamen giderilemez; tahmin edilen eşik değeri çok büyükse, faydalı konuşma bilgileri filtrelenebilir ve bu da sinyallerin gerçekliğinin azalmasına yol açmaktadır [38].

Eşik işlevi, sert eşik işlevi ve yumuşak eşik işlevi olarak ikiye ayrılmaktadır. Sert eşik fonksiyonunun hesaplama formülü Eşitlik 3.2’de verilmiştir:

$$\hat{\omega}_{m,n} = \begin{cases} \omega_{m,n} & |\omega_{m,n}| \geq \mu \\ 0 & |\omega_{m,n}| < \mu \end{cases} \quad (3.2)$$

Burada $\hat{\omega}_{m,n}$ tahmini dalgacık katsayısıdır, $\omega_{m,n}$ orijinal sinyalin dalgacık ayrıştırma katsayısıdır ve μ kaldırılan gürültü eşığıdır. Sert eşik işlevi, farklı ölçeklerdeki gürültülü sinyallerin dalgacık ayrıştırma katsayısını eşik değeriyle karşılaştırır. Dalgacık ayrıştırma katsayısı eşik değerinden küçükse doğrudan 0'a çekilir ve eşik değerinden büyükse nokta değişmez [38].

Yumuşak eşik fonksiyonunun hesaplama formülü Eşitlik 3.3'te verilmiştir:

$$\hat{\omega}_{m,n} = \begin{cases} \text{sgn}(\omega_{m,n})(|\omega_{m,n}| - \mu) & |\omega_{m,n}| \geq \mu \\ 0 & |\omega_{m,n}| \leq \mu \end{cases} \quad (3.3)$$

Fonksiyonun süreksizliğinin gürültü giderme üzerindeki etkisini ortadan kaldırmak için Eşitlik 3.3 uygulanır. Farklı ölçeklerdeki gürültü sinyalinin dalgacık ayrıştırma katsayısı eşik değeri ile karşılaştırılır. Dalgacık ayrıştırma katsayısı eşik değerinden küçükse doğrudan 0'a çekilir ve büyükse nokta değişmez [38].

3.2 Spektrogram

Spektrogram, bir sinyalin zaman içindeki sinyal gücünün görsel olarak temsil etmektedir. Genellikle ses sinyalleri için kullanılmaktadır [39-41].

Bir spektrogram görüntüsünde dikey eksen frekansları, yatay eksen zamanı ve renklerin yoğunluğu ise frekans bileşenlerinin genliğini temsil etmektedir. Yüksek genlikler kırmızı tonlarında gösterilmekte, düşük genlikler ise mavi tonlarında gösterilmektedir. Spektrogram Kısa Zamanlı Fourier Dönüşümü (KZFD) ile elde edilmektedir [41].

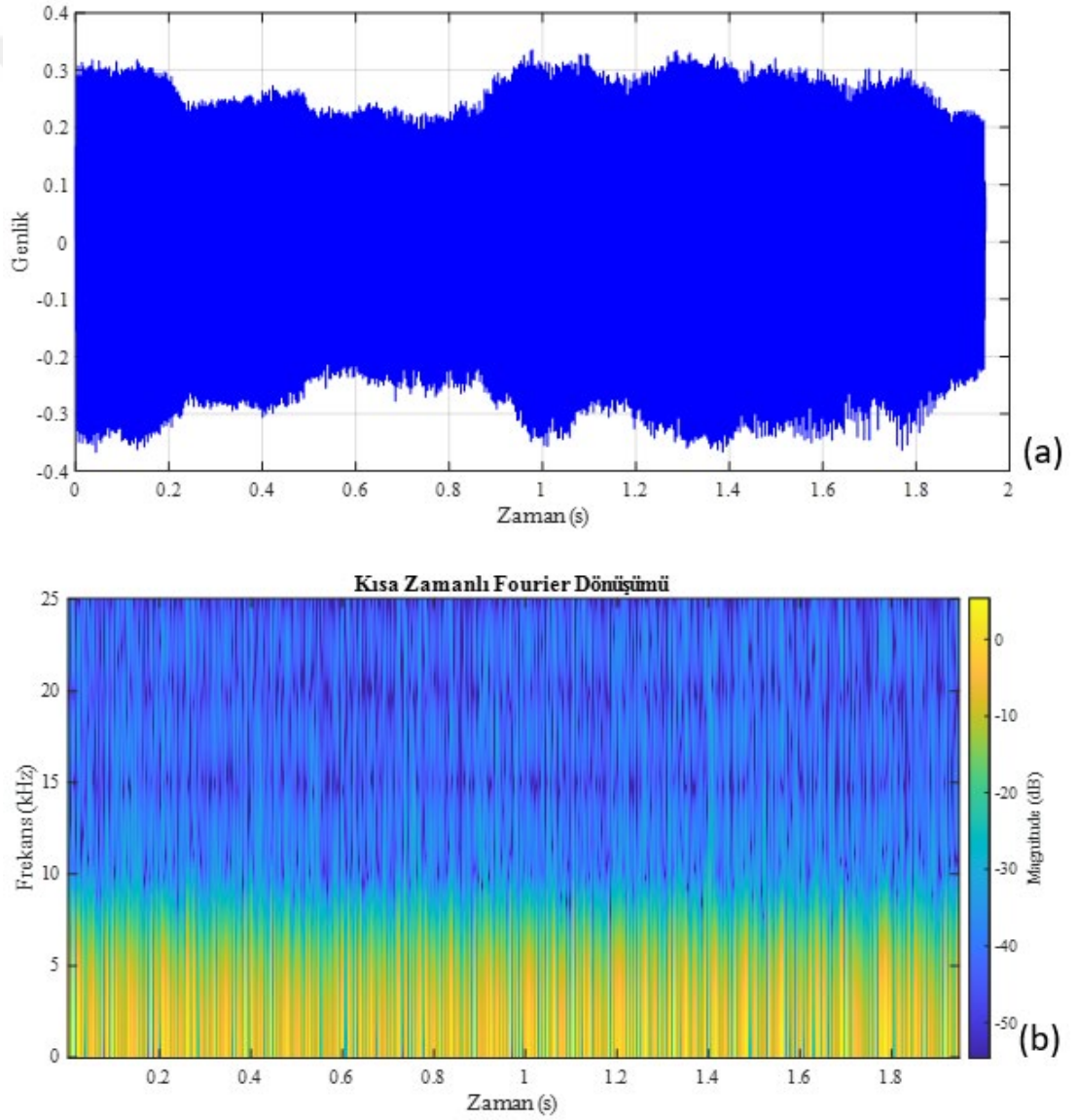
KZFD, bir sinyalin zamanla değişen yerel bölümlerinin frekansını ve fazını belirtmektedir. Giriş sinyalini zaman eksenini boyunca eşit olarak küçük parçalara bölmekte ve ardından frekans bileşenlerinin büyüklüğünü hesaplamak için her parça için HFD hesaplamaktadır. Böylece her parça kendi frekans dağılımını almaktadır. Her parçanın görüntü için karşılık gelen dikey çizgisi vardır ve bu parçalar genellikle sınırdan üst üste gelerek önceki yığından küçük bilgiler taşımaktadır. Her bir parçaya karşılık gelen dikey çizgiler, dizideki sırasına göre konumlandırılmakta ve bu şekilde

spektrogram görüntüsü oluşturulmaktadır. KZFD ve spektrogram formülleri sırasıyla Eşitlik 3.4 ve Eşitlik 3.5'te verilmiştir:

$$STFT\{x[n]\}(m, \omega) = \sum_{n=-\infty}^{\infty} x[n]w[n - m]e^{-j\omega n} \quad (3.4)$$

$$\text{spectrogram } \{x[n]\}(m, \omega) = |STFT\{x[n]\}(m, \omega)|^2 \quad (3.5)$$

Burada $x[n]$ dönüştürülecek sinyali, $w[n]$ pencereleme işlevini, m zamanı, ω frekansı ve n parçayı temsil etmektedir [41]. Şekil 3.1 (a)'da bir ses sinyalinin zaman – genlik grafiği, (b)'de aynı ses sinyalinin spektrogramı gösterilmiştir.



Şekil 3.1 : (a) bir ses sinyali, (b) ses sinyalinin spektrogramı.

3.3 Derin Öğrenme

Makine öğrenmesinin bir alt dalı olan derin öğrenme, yapay zekâ uygulamalarından biridir ve günümüzde sıklıkla tercih edilmektedir. Yeni bir yöntem olarak değerlendirilse de ilk çalışmalar 1950’li yıllarda başlamıştır. Derin öğrenmenin temelleri yapay sinir ağları araştırmalarına dayanmaktadır. Yapay sinir ağları 1940-1960 yılları arasında tek bir nöronun eğitilmesi ile ortaya çıkmıştır [42]. Daha sonraları 1980-1995 yılları arasında bir veya birden çok gizli katman ile yapay sinir ağının eğitilmesi çalışılmıştır [43]. 2006 yılından itibaren “Derin Öğrenme” adıyla gerçekleştirilen çalışmalarda derin ağların eğitilmesi konusunda araştırmalar başlamıştır [44-46].

Temel olarak insan beyninin çalışma mekanizmasının örnek alındığı sinir ağı modellerinde insan biyolojik sinir sisteminde yer alan nöronların uyarılması ve sinirler arasındaki bağlantı mekanizmalarının makineye öğretilmesi çalışılmıştır [47]. Derin öğrenme algoritmalarının 2006’ya kadar gündemde yer alamaması söz konusu algoritmaların çalıştırılabileceği yeterli donanımın 1980’lerde mevcut olmamasına bağlanmaktadır. Sonraki yıllarda grafik işlemede gerçekleşen ilerlemeler sayesinde daha hızlı veri işleme ile derin öğrenme yeniden çalışma konusu olmuştur. Veri setindeki özneliklerin derin öğrenmede katmanlar tarafından otomatik olarak çıkarılabilmesi derin öğrenmenin geleneksel makine öğrenmesine göre en büyük avantajıdır, ayrıca büyük veri setlerini sınıflandırmadaki başarısı da derin öğrenmenin avantajlarından biridir [48]. Geleneksel makine öğrenmesinde problem bölünüp daha sonra tekrar sonuçlar en son aşamada birleştirilmektedir bu nedenle makine öğrenmesi nesnelere tanıyabilmek amacıyla bazı algoritmalarından çıkarılan özneliklere ihtiyaç duymaktadır, derin öğrenmede ise problem baştan sona başka herhangi bir algoritmaya ihtiyaç duyulmadan çözülebilmektedir, ancak bu nedenle derin öğrenmede eğitim süresi daha uzundur [49].

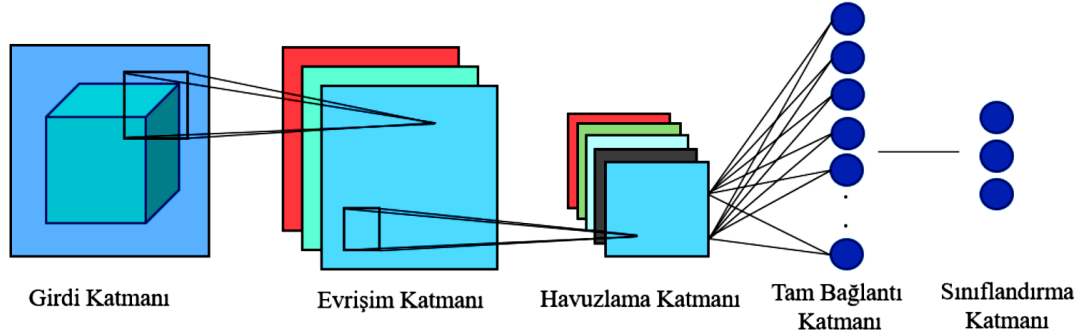
Derin öğrenme nesne tanımda, ses tanımda, video işlemede, hastalık teşhisinde ve pek çok endüstriyel alanda sağladığı avantajlar sayesinde kullanılmaya başlanmıştır. İlerleyen yıllarda gelişen teknoloji ile daha da yaygınlaşması beklenmektedir [49].

3.4 Evrişimsel Sinir Ağları (ESA)

ESA, son yıllarda örüntü tanıma ile ilgili görüntü işlemeden ses tanımaya kadar çeşitli alanlarda çığır açan sonuçlara sahiptir [50].

ESA'da girdi görüntüleri çeşitli katmanlarla işlenmektedir ve kullandığı evrişim katmanları ile öznitelikleri otomatik bir şekilde çıkarmaktadır. Klasik makine öğrenmesi metodlarında ayırt edici özniteliklerin belirlenmesi hem zaman almakta hem de konu hakkında bilgi gerektirmektedir. Yanlış veya yetersiz sayıda öznitelik çıkarımı, sınıflandırıcı performansını büyük oranda etkileyebilmekte ve başarıyı düşürmektedir. Evrişimsel sinir ağları bu probleme farklı boyutlarda filtreler ve evrişim katmanları ile çözüm bulmaktadır [49].

ESA'nın temel katmanları giriş, evrişim, havuzlama ve tam bağlantılı katmandır. Evrişim ve havuzlama katmanları, öznitelik çıkarma işlemini gerçekleştirirken, tam bağlantılı katman çıkarılan öznitelikleri sınıflandırarak bir çıktı elde eder [51]. Örnek bir ESA şeması Şekil 3.2'de gösterilmiştir.



Şekil 3.2 : Örnek bir ESA şeması [49].

3.4.1 ESA'nın temel basamakları

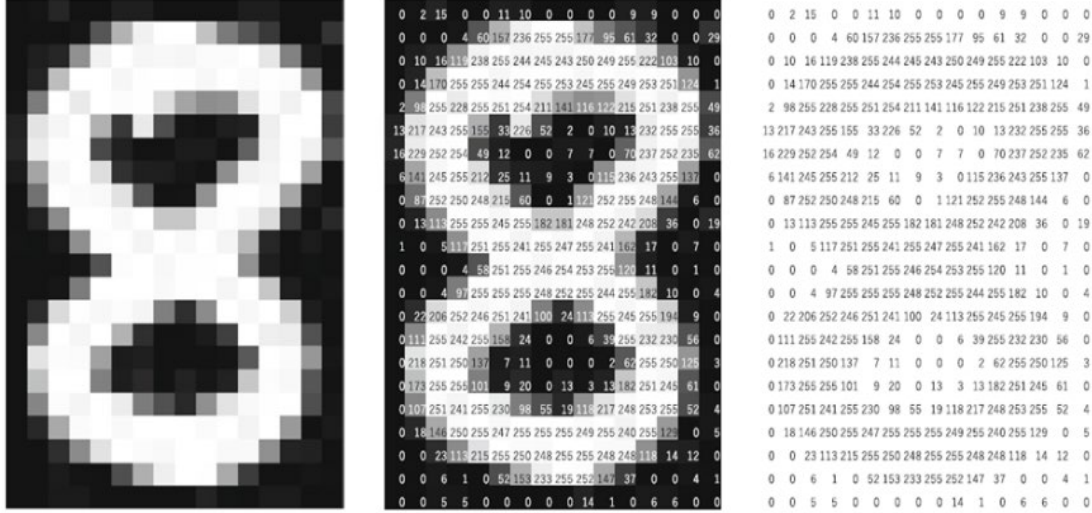
3.4.1.1 Giriş katmanı

Giriş katmanında veri ham olarak ağına verilmektedir. Tasarlanacak modelin başarımı için bu katmandaki verinin boyutu önem kazanmaktadır. Veri boyutunun yüksek seçilmesi yüksek bellek kapasitesine ihtiyaç doğurabilir ve eğitim süresi ile test süresini uzatabilir ancak başarıyı artırabilir. Veri boyutunun düşük seçilmesi de düşük bellek ihtiyacı, eğitim ve test sürelerinin azaltılmasını sağlar ancak ağın performansında da düşme görülebilir [52].

3.4.1.2 Evrişim katmanı

Evrişim katmanı, tipik olarak doğrusal ve doğrusal olmayan işlemlerin bir kombinasyonundan oluşan özellik çıkarımını gerçekleştiren ESA mimarisinin temel bir bileşenidir.

Şekil 3.3'de gösterildiği gibi dijital görüntülerde piksel değerleri iki boyutlu bir matriste saklanır [51].

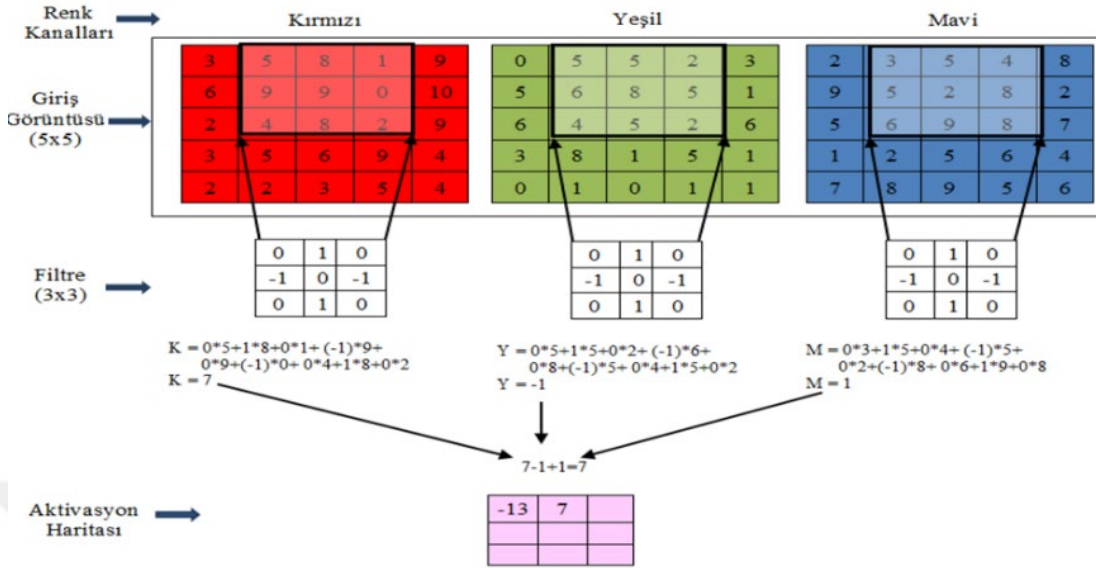


Şekil 3.3 : Bilgisayar bir görüntüyü bir dizi sayı olarak görür. Sağdaki matris, her biri sol görüntüdeki piksel parlaklığına karşılık gelen 0 ile 255 arasında sayılar içerir. Her ikisi de ortadaki görüntüde kaplanmıştır [51].

Evrişim işlemi belirli bir filtrenin tüm görüntü üzerinde dolaştırılması işlemine dayanmaktadır. Her bir katmada farklı boyutta filtreler bulunur ve bu filtreler ile her seferinde bir önceki katmandan gelen görüntülere evrişim işlemi uygulanarak çıkış verisi oluşturulur. Bu evrişim işlemi sonucu özellik (aktivasyon) haritası elde edilir ve bu harita ile her bir filtreye özgü özellikler çıkarılır. Ağın eğitimi sırasında bu filtrelerin boyutları, eğitim kümesindeki her öğrenme yinelemesiyle değişir. Bu şekilde ağ, özelliklerin belirlenmesi için, verinin hangi bölgelerinin önemli olduğuna karar verir.

Evrişim işleminde bir filtre ile giriş görüntüsü üzerinde sağa veya sola doğru belirli bir adım kaydırılarak dolaşılır ve matris sınırına gelindiğinde ise bir basamak aşağı kayarak devam edilir. Bu işlem matrisin tümü üzerine uygulanır. Filtre katsayıları her bir renk kanalındaki değerlerle çarpılıp toplanır. Her üç kanal üzerinde bu işlem yapılarak toplanarak aktivasyon haritası oluşturulur. 5x5x3 boyutundaki bir giriş

görüntüsüne 3x3'lük bir filtrenin uygulandığı evrişim işlemi Şekil 3.4'te gösterilmiştir [52].



Şekil 3.4 : 5x5x3 boyutundaki bir giriş görüntüsüne 3x3'lük bir filtrenin uygulandığı evrişim işlemi [52].

Filtrenin adım sayısı ve kenarlık parametreleri ile aktivasyon haritasının boyutları belirlenmektedir (Eşitlik 3.6). Filtre penceresinin kayma miktarı adım sayısı ile ifade edilmektedir. Kaynak görüntünün sınırları dışına sıfır değerinde yeni piksellerin eklenmesi halinde belirlenen ek kenar sayısı da kenarlık değeri olarak ifade edilmektedir. Kenarlık değeri sayesinde orijinal matris boyutu ile aynı boyutta bir çıktı matrisi elde edilebilmektedir [49].

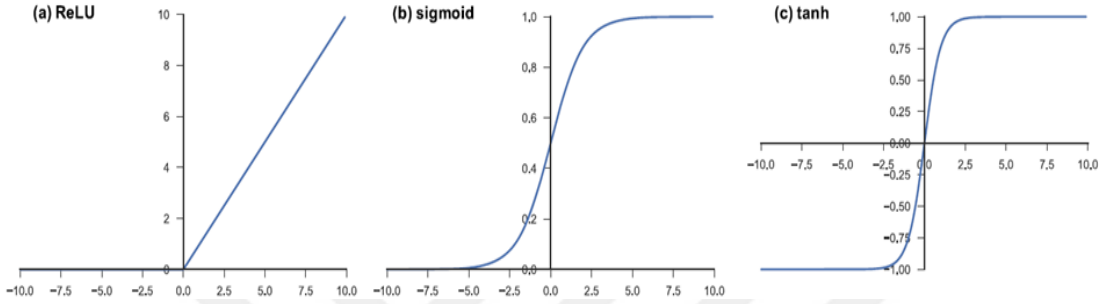
$$(m, n) = \frac{g + 2p - f}{a} + 1 \quad (3.6)$$

Eşitlik 2.6'da g görüntü boyutunu, p kenarlık değerini, f filtre boyutunu ve a adım sayısını ifade etmektedir. Eşitlik 3.6'dan da anlaşılacağı gibi evrişim işlemi sırasında kenarlık değeri işleme katılmaz ise çıktı görüntüsünün boyutu azalmaktadır [49].

Evrişim katmanı sayısı arttıkça daha detaylı özellikler elde edilmektedir [53]. Ancak daha fazla özellik çıkarımı her zaman daha iyi sınıflandırma performansı alınabileceği anlamına gelmemektedir. Evrişim katmanı sayısının düşük olması da eğitim verisinde yeterli sayıda ayırt edici özelliğin çıkarılamamasına ve modelin başarısının düşmesine neden olabilmektedir. Yüksek performans başarısı sağlayabilmek için yeterli sayıda katman kullanılması gerekmektedir.

3.4.1.3 Doğrusal olmayan aktivasyon fonksiyonu

Evrişim gibi doğrusal bir işlemin çıktıları daha sonra doğrusal olmayan bir aktivasyon fonksiyonundan geçirilir. Sigmoid veya hiperbolik tanjant (tanh) işlevi gibi düzgün doğrusal olmayan fonksiyonlar, biyolojik bir nöron davranışının matematiksel temsilleri oldukları için daha önce kullanılmış olsa da, şu anda kullanılan en yaygın doğrusal olmayan aktivasyon fonksiyonu, ReLU'dur. Aktivasyon fonksiyonları Şekil 3.5'te gösterilmiştir [51].



Şekil 3.5 : Sinir ağlarına yaygın olarak uygulanan aktivasyon fonksiyonları: (a) ReLU, (b) sigmoid ve (c) tanh [51].

Bu fonksiyonun giriş verisine yapmış olduğu etki negatif değerleri sıfıra çekmesidir. Öncesinde evrişim katmanında belirli matematiksel işlemler yapıldığı için ağ doğrusal bir yapıdadır. Bu derin ağı doğrusal olmayan bir yapıya sokmak için bu fonksiyon uygulanır. Bu sayede ağın daha hızlı öğrenmesi sağlanır. Matematiksel olarak Eşitlik 3.7'deki gibi tanımlanır ve görsel olarak da Şekil 3.6'da gösterilmiştir [52].

$$f(x) = \begin{cases} 0 & \text{eğer } x < 0 \\ x & \text{eğer } x \geq 0 \end{cases} \quad (3.7)$$

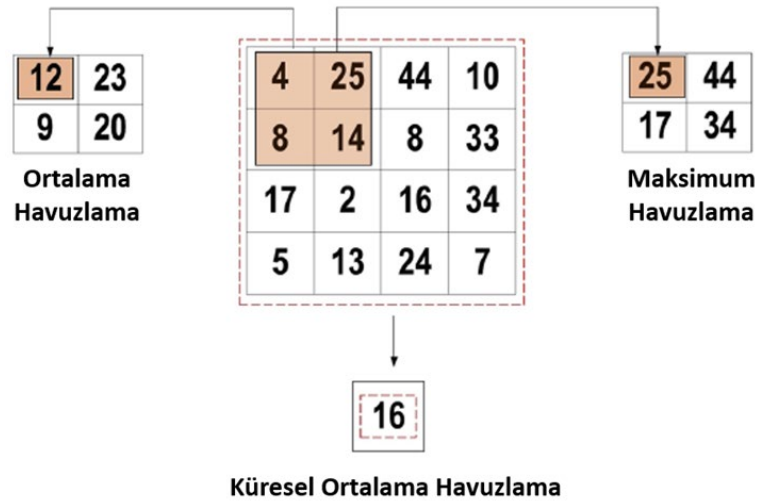


Şekil 3.6 : Bir ESA modelinde evrişim işlemi ve ReLU işleminin giriş görüntüsüne yaptığı etki [52].

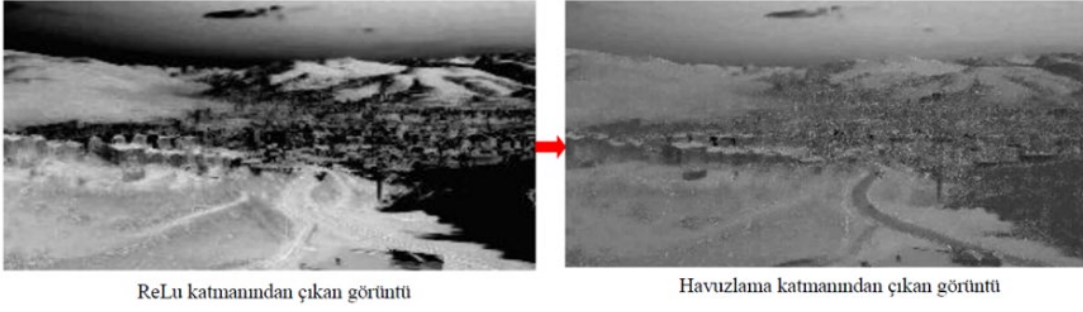
3.4.1.4 Havuzlama katmanı

Havuzlama katmanının amacı, sonraki evrişim katmanı için giriş boyutunu düşürmektir. Boyuttaki bu azalma bilgi kaybına yol açar. Bu kayıp ile bir sonraki ağ katmanları için daha az hesaplama yükü oluşturulur ve sistemin ezberlemesini önler. Bu katmanda da evrişim katmanındaki gibi belirli filtreler tanımlanır. Filtreler görüntü üzerinde kaydırılarak görüntüdeki piksellerin maksimum değerleri (maksimum havuzlama) veya değerlerin ortalaması (ortalama havuzlama) alınarak işlem gerçekleştirilir. Genellikle daha iyi performans gösterdiği için maksimum havuzlama tercih edilir. Havuzlama işlemi, evrişim katmanı sonucu oluşan görüntülerin tümünde gerçekleştirilir [52].

Maksimum havuzlama, girdi özellik haritalarını bölütlendirip, her bir bölütteki maksimum değeri çıkararak diğer tüm değerleri atar. Pratikte yaygın olarak 2 adımlı 2×2 boyutunda bir filtreye sahip maksimum havuzlama kullanılır. Bu, özellik haritalarının düzlem içi boyutunu 2 faktörüyle altörnekler. Yükseklik ve genişlikten farklı olarak, özellik haritalarının derinlik boyutu değişmeden kalır. Ortalama havuzlama, girdi özellik haritalarını bölütlendirip, her bir bölütteki ortalama değeri çıkarır. Küresel ortalama havuzlama, yükseklik \times genişlik boyutuna sahip bir özellik haritasının tüm bölütlerinin ortalaması alınır altörneklendirilir [51]. Maksimum, ortama havuzlama ve küresel ortalama havuzlama örneği Şekil 3.7'de ve havuzlama katmanının uygulanması sonucu oluşan bir görüntü örneği Şekil 3.8'de gösterilmektedir.



Şekil 3.7 : Maksimum, ortama havuzlama ve küresel ortalama havuzlama örneği [54].



Şekil 3.8 : Havuzlama katmanının uygulanması sonucu oluşan görüntü [52].

3.4.1.5 Tam bağlantılı katman

Son evrişim veya havuzlama katmanının çıktısı özelliği haritaları tipik olarak düzleştirilir, yani tek boyutlu bir sayı dizisine (veya vektöre) dönüştürülür ve her girdinin her çıktıya öğrenilebilir bir ağırlıkla bağlandığı tam bağlantılı katmana bağlanır.

3.4.1.6 Dizi normalizasyonu

Dizi normalizasyonundaki amaç önceki katmandaki ağırlıkların, sonraki katmana oranla oldukça büyük olduğu durumlarda katmanlar arasındaki kovaryansı azaltmaktır. Böylece eğitimin hızlanması sağlanmakla birlikte eğitim esnasında aşırı öğrenme durumu da engellenebilmektedir.

ESA'nın eğitiminde kullanılan en kritik hiperparametrelerden olan mini dizi boyutu ile tüm verinin tek seferde eğitime verilmesi yerine, verinin küçük paketlere ayrılarak eğitime verilerek işlem yükünün azaltılması sağlanır. Dizi normalizasyonu bu mini dizilere uygulanmaktadır (Eşitlik 3.8).

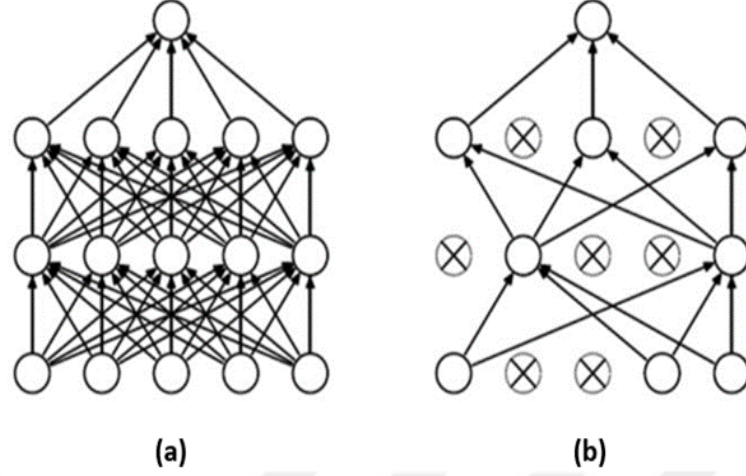
$$BN_{\gamma, \beta}(x_i) = \gamma \frac{x_i - \mu_B}{\sqrt{\sigma_B^2 + \epsilon}} + \beta \quad (3.8)$$

Eşitlik 3.8'de x mini dizi, μ_B mini dizi ortalaması, σ_B^2 mini dizi varyansı, γ ve β katsayıları ise eğitim sırasında optimize edilen katsayılardır [49].

3.4.1.7 DropOut

ESA'da büyük veriler ile eğitim işlemi yapıldığı için bazen ağ ezberleme yapabilir. Ezberlemenin önüne geçmek amacıyla bu DropOut işlemi kullanılır. Bu sayede ağdaki bazı düğümleri kaldırılır. Bu yöntemde nöronlar olasılık tabanlı seçilmektedir

ve modele olasılık değerinin belirtilmesi gerekmektedir. Şekil 3.9(a)'da ağın orijinal yapısı, Şekil 3.9(b)'de DropOut işleminden sonraki hali gösterilmiştir [49,55].



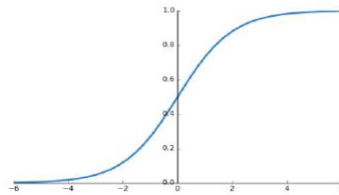
Şekil 3.9 : (a) ağın orijinal yapısı, (b) DropOut işleminden sonraki hali [55].

3.4.1.8 Sınıflandırma katmanı

Mimarinin son katmanı olan bu katmanda sınıf olasılığı sigmoid (Eşitlik 3.9) ve softmax (Eşitlik 3.10) gibi fonksiyonlar kullanılarak belirlenmektedir. Bu katmanda girilmesi gereken parametre sınıf sayısıdır, kullanılan fonksiyon ile çıktıda sınıf sayısı kadar olasılık değeri alınır ve en yüksek değer hangi sınıfta elde edilirse girdi o sınıfa atanır [49].

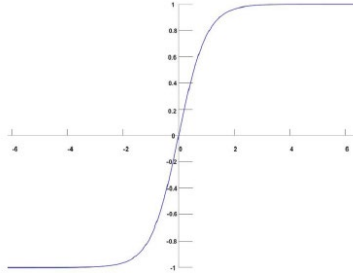
$$f(z)_i = \frac{1}{1 + e^{-z_i}} \quad (3.9)$$

Eşitlik 3.9'da z girdi vektörü, i girdi vektörünün her bir elemanı, $f(z)$ çıktı olasılığıdır. Sigmoid fonksiyonu, uygulamada geçen süreyi azaltmaktadır. Ancak fonksiyonun türevinin kısa bir aralığa sahip olmasından dolayı önemli ölçüde bilgi kaybına neden olmaktadır, bu durum da bu fonksiyonun kullanılması açısından bir dezavantajdır. Şekil 3.10'da Sigmoid fonksiyonu gösterilmiştir [49].



Şekil 3.10 : Sigmoid fonksiyonu [49].

Softmax fonksiyonu, girdi vektöründeki her bir elemanı üsteliyle orantılı bir olasılık dağılımına normalize eden bir aktivasyon fonksiyonudur (Şekil 3.11).



Şekil 3.11 : Softmax fonksiyonu [56].

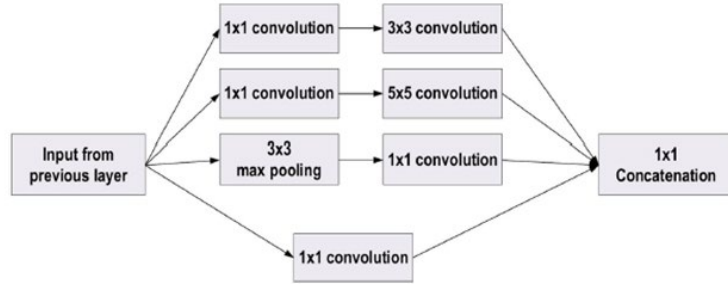
Bazı girdi verileri 0'dan küçük veya 1'den büyük olabilir. Softmax uygulandıktan sonra, her bir elemanın değeri 0 ile 1 aralığında çekilmekte ve toplamı 1 olmaktadır. Böylece çıktı verileri olasılık dağılımı şeklinde değerlendirilebilmektedir.

$$f(z)_i = \frac{e^{z_i}}{\sum_{i=1}^K e^{z_i}} \quad (3.10)$$

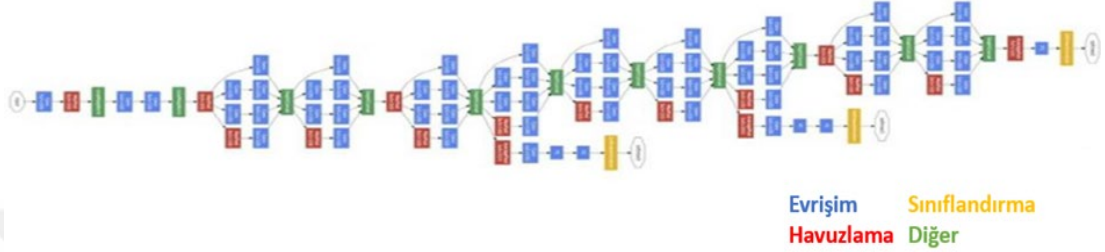
Eşitlik 3.10'da z girdi vektörü, i girdi vektörünün kaçınıcı değeri için işlem yapıldığı ve K vektördeki eleman sayısıdır (sınıf sayısı) [49].

3.4.2 GoogleNet

2014-ILSVRC yarışmasında GoogleNet (Inception-V1) kazanan olarak ortaya çıkmıştır. GoogleNet, 7 milyon öğrenilebilir parametreye sahip 22 katmanlı bir ESA'dır. Bu önceden eğitilmiş ağ, görüntüleri 1000 kategoride sınıflandırabilmektedir. Ağ, ImageNet veritabanı kullanılarak eğitilir ve 224×224 Kırmızı Yeşil Mavi (Red Green Blue – RGB) görüntü girişini kabul etmektedir [57]. GoogleNet'in amacı, ESA parametrelerinin verimliliğini artırmanın yanı sıra öğrenme kapasitesini artırmak ve yüksek düzeyde doğruluk elde etmektir. GoogleNet, özellik çıkarımı için birleştirme, dönüştürme ve bölme işlevlerini kullanarak çok ölçekli evrimsel dönüşümleri birleştirdiği için ESA yapısında yeni bir başlangıç bloğu konsepti sunmuştur. Şekil 3.12'de başlangıç bloğu mimarisi gösterilmiştir [54]. Elde edilen özellik haritaları daha sonra her bir başlangıç modülünün çıktısında birleştirilmektedir. GoogleNet ayrıca, geleneksel ESA'larda son evrim katmanından sonra gelen tam bağlantılı katmanı, küresel bir ortalama havuzlama ile değiştirerek ağdaki parametre sayısını önemli ölçüde azaltmaktadır. GoogleNet mimarisi Şekil 3.13'de gösterilmiştir [57].



Şekil 3.12 : Google Block'un temel yapısı [54].

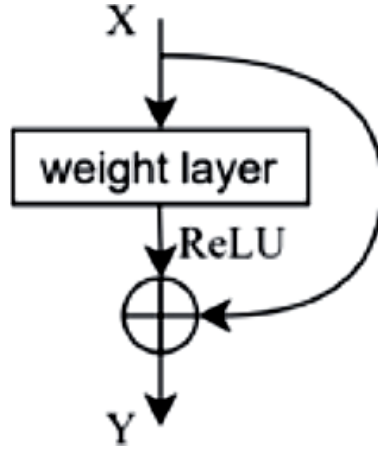


Şekil 3.13 : GoogLeNet mimarisi [57].

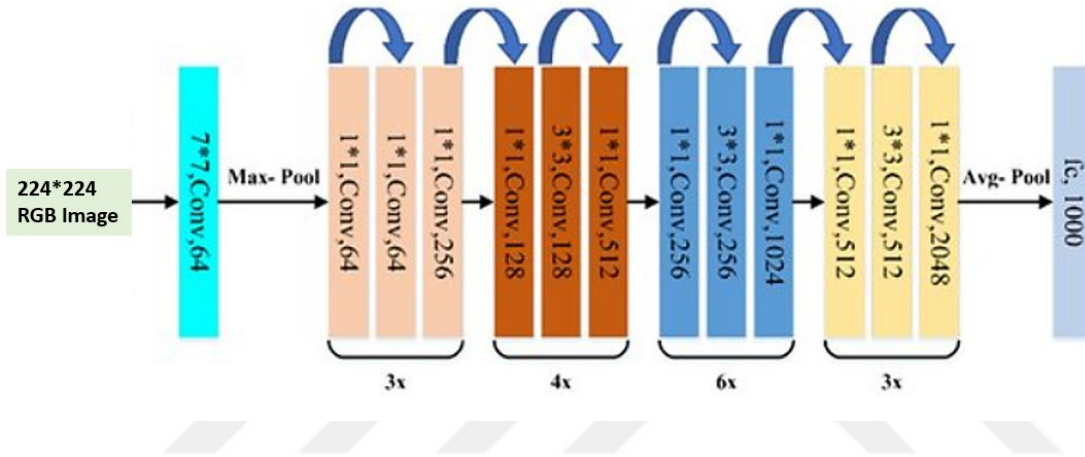
3.4.3 ResNet-50

He ve arkadaşları [58] ILSVRC 2015'in galibi olan ResNet'i (Residual Network) geliştirmiştir. Amaçları, önceki ağlara kıyasla, kaybolan gradyan sorunu olmayan ultra derin bir ağ tasarlamaktır. Katman sayısına göre 34 katmandan 1202 katmana kadar çeşitli ResNet türleri geliştirilmiştir. En yaygın tür ResNet-50'dir [54].

ResNet-50, 48 evrişim katmanından oluşan bir ESA tabanlı modeldir. Diğer iki katman, 1 Maksimum havuzlama katmanı ve 1 Ortalama havuzlama katmanıdır. ResNet mimarisi, ESA'nın birden çok katmanla çalışmasına olanak sağlamıştır. Birden çok yığılmış katmana sahip derin sinir ağları, daha az katmana sahip modellere göre daha yüksek eğitim hatası yüzdeleri üretmektedir. ResNet, kısayol bağlantılarının eklenmesine ve artık işlevlerin kullanılmasına izin vererek yığılmış derin sinir ağları katmanlarının eğitim hatalarının azaltılmasını sağlamıştır. Bu mimaride doğrudan bağlantı seti bazı katmanların atlanmasına yardımcı olur [59]. Şekil 3.14'te artık öğrenmeye ilişkin basit bir yaklaşımı gösterilmektedir. Bir önceki X katmanının çıktısı ağırlıklarla çarpılmaktadır ve doğrusal olmayan çıktı elde etmek için ReLU aktivasyon fonksiyonu uygulanmaktadır. Son olarak, sonuca X girişi eklenmekte ve bir sonraki katman için Y çıkışı elde edilmektedir [60]. Şekil 3.15'te ResNet-50 mimarisi gösterilmiştir.



Şekil 3.14 : Artık öğrenmeye ilişkin basit bir yaklaşım [60].

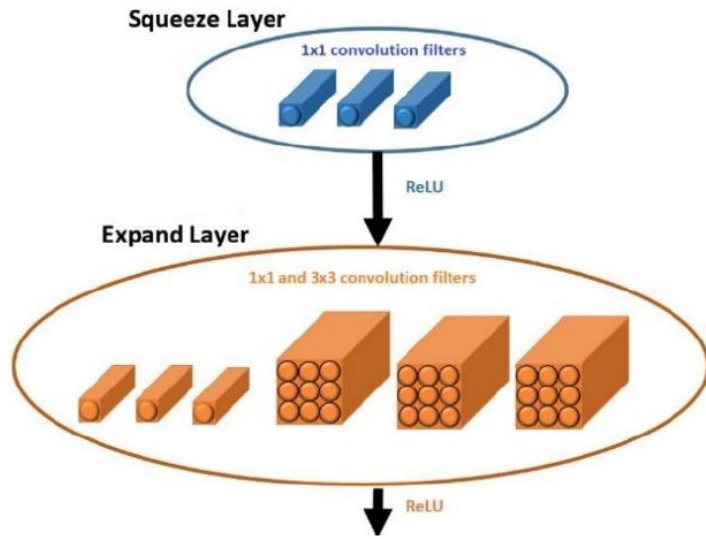


Şekil 3.15 : ResNet-50 mimarisi [61].

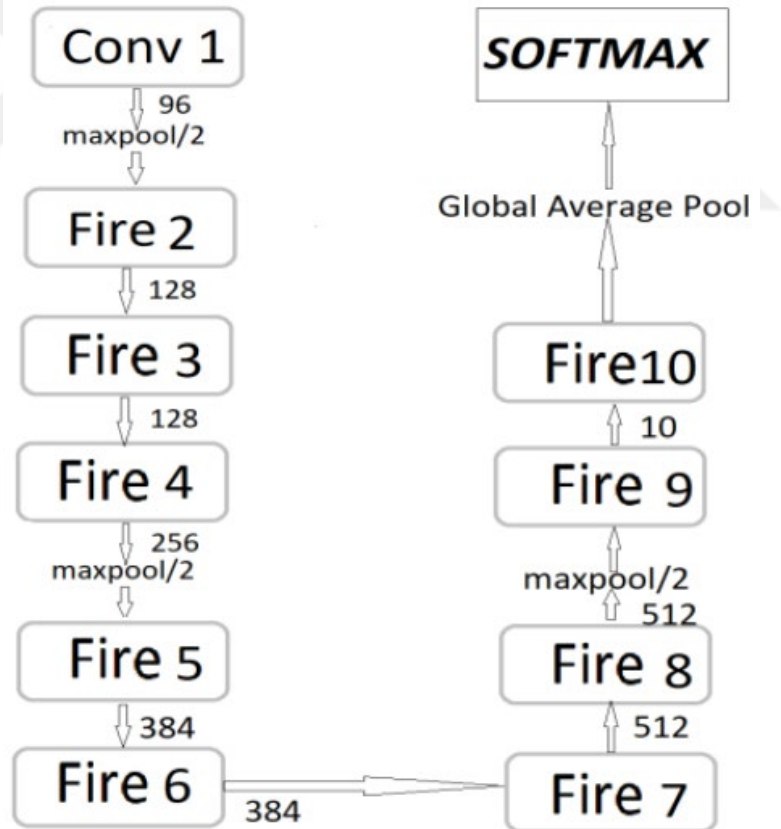
3.4.4 SqueezeNet

Bu mimari 2016 yılında Iandola ve arkadaşları tarafından önerildi [62]. SqueezeNet ağının kendisi, Şekil 3.16’da gösterildiği gibi, ateşleme (Fire) modülleri adı verilen yapı taşlarından oluşur. Her ateşleme modülü, temelde bir sıkıştırma (squeeze) ve ardından bir genişletme (expand) katmanıdır. Sıkıştırma katmanı, 1x1 evrişim filtrelerinden oluşan bir katmandır ve genişletme katmanı 1x1 ve 3x3 filtrelerinin bir kombinasyonudur. Sıkıştırma katmanındaki özellik haritalarının sayısı, genişletme katmanı özellik haritalarının sayısından daha az veya buna eşit sayıdadır, bu nedenle ağ ağırlıklarının sayısı azaltırken çıkarılan özellik haritaları üzerinde bir tür sıkıştırma gerçekleştirir. Bu ateşleme modülleri daha sonra Şekil 3.17’de görüldüğü gibi SqueezeNet modelinin makro mimarisini oluşturmak için bir araya toplanır. Ateşleme modülünün önemli bir hiperparametresi sıkıştırma oranıdır. Sıkıştırma katmanı özellik haritalarının sayısı genişletme katmanı özellik haritalarının sayısına

bölünür. Bu oranın 1/2'ye yükseltilmesi genellikle ağın boyutunu artırmayla sonuçlansa da ağı doğruluk oranlarını artırır [63].



Şekil 3.16 : SqueezeNet modelinde bir ateşleme modülünün genel yapısı [62].



Şekil 3.17 : Temel SqueezeNet modelinin makro mimarisi [62].

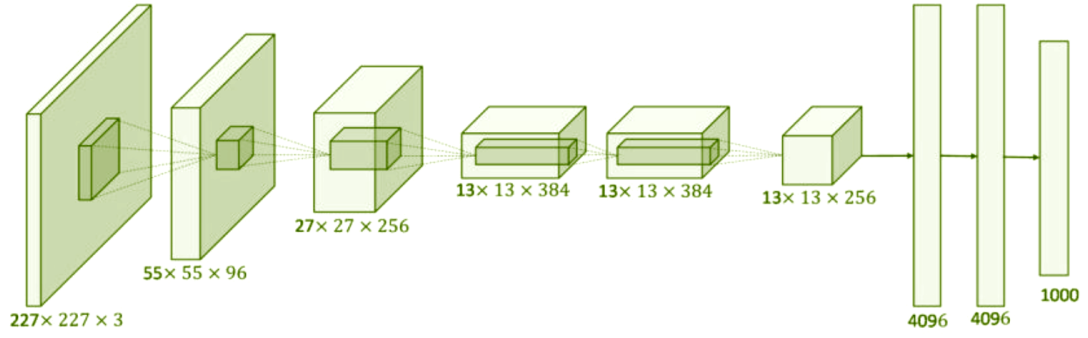
SqueezeNet, geleneksel ESA performansını iyileştirmek için üç ana stratejiyi benimser. İlk olarak, ağıda kullanılan filtrelerin çoğu 3x3 yerine 1x1'dir, böylece ağı ağırlıklarının sayısı büyük ölçüde azaltılır. İkincisi, giriş kanallarının sayısı 3x3

filtrelere düşürülmüştür. Bu yaklaşım aynı zamanda ağ ağırlıklarının sayısını da büyük ölçüde azaltır. Üçüncüsü, altörnekleme daha sonra ağda daha büyük aktivasyon haritalarında gerçekleştirilir. Buradaki mantık, üzerinde altörneklemenin gerçekleştirildiği aktivasyon haritalarının boyutu ile nihai sınıflandırma doğruluğu sonuçları arasında doğrudan bir ilişki olduğudur.

Ağ ağırlıklarının sayısını büyük ölçüde azaltan bir diğer önemli strateji, ağın sonunda sıklıkla kullanılan tam bağlantılı katmanların kaldırılmasıdır. Bu katman, çıktı kanallarının sayısının veri sınıflarının sayısına eşit olduğu ve ardından bir dropout katmanı ve softmax aktivasyon fonksiyonu tarafından takip edilen bir evrişim katmanı ile değiştirilir [63].

3.4.5 AlexNet

AlexNet, ImageNet veritabanı ile eğitilmiş ve 150.000 ImageNet verisi üzerinde 60 milyon öğrenilebilir parametre ile test edilmiş sıralı yapıda tasarlanmış bir ağ modelidir [64]. Girdi verisi olarak 227x227 boyutlarında RGB görüntüler kullanılmaktadır. 5 evrişim ve 3 tam bağlantılı katmandan meydana gelmektedir. İlk evrişim katmanında 11x11x3 boyutunda adım sayısı 4 olan 96 adet filtre uygulanmaktadır. Bu katmanın çıktılarına ReLU ve kanal bazlı normalizasyon amacıyla çapraz kanal normalizasyonu uygulanmaktadır. Maksimum havuzlama katmanı, ReLU veya normalizasyon çıktısındaki parametre sayısının düşürülmesi amacıyla uygulanmaktadır. İkinci evrişim katmanı 5x5x48 boyutunda, adım sayısı 1 ve kenarlık sayısı 2 olan 2 tip 128 filtre içermektedir. Üçüncü evrişim katmanı adım sayısı 1 ve kenarlık sayısı 1 olan 3x3x256 boyutunda 384 adet filtreden oluşmaktadır ve dördüncü evrişim katmanı da 3x3x192 boyutunda adım sayısı 1 ve kenarlık sayısı 1 olan 192 adet filtreden oluşmaktadır. Son evrişim katmanı ise dördüncü evrişim katmanıya aynı özelliklerdedir fakat 128 adet filtre içermektedir. Birinci tam bağlantılı katman 4096 boyutlu bir öznitelik vektörü çıkarmaktadır. Aşırı öğrenmeyi engellemek için 0.5 olasılıklı dropout uygulanmaktadır. İkinci tam bağlantılı katman birinci ile aynı özelliklere sahiptir; üçüncü tam bağlantılı katman da çıktıyı 1000 adet düğümle eşlemektedir, bu sayede son katman itibarıyla görüntüleri 1000 sınıfa ayırabilmektedir. Son aşamada sınıflandırmada softmax fonksiyonu kullanılmaktadır. Şekil 3.18'de AlexNet mimarisi gösterilmiştir [65].



Şekil 3.18 : AlexNet mimarisi [65].

3.5 Komşuluk Bileşen Analizi

Özellik seçimi, uygun olmayan ve fazlalık özellikleri atarak belirli bir özellik kümesinden küçük bir alt küme seçme tekniğidir. Uygun özellik seçimi sadece özelliklerin boyutlarını azaltmakla kalmaz, aynı zamanda algoritmaların genelleme performansını ve yürütme hızını da iyileştirir [66].

Komşuluk Bileşen Analizi (KBA), k-En Yakın Komşuluk (k-Nearest Neighbor, k-NN) algoritmasına dayalı olarak geliştirilmiş, parametrik olmayan ve gömülü bir yöntemdir. KBA, bir değişken alt kümesinin kullanılabilirliğini değerlendirmek için tahmin performansını kullanan bir öğrenme algoritması içeren sarmalayıcı yöntemdir. Sarmalayıcı yaklaşım, öğrenme algoritmasına en uygun optimal alt kümeyi seçer [67]. KBA algoritmasının amacı, optimize edilmiş bir düzenleme parametresi ile bir dışarıda bırakılan sınıflandırma doğruluğunu maksimize ederek özellik ağırlık vektörünü öğrenmektir [68]. KBA'nın avantajı, özellik sıralaması ile birlikte önemli özellikler hakkında bilgi sağlamasıdır. KBA, boyut azaltma işlemi sırasında herhangi bir bilgi kaybetmez [66].

Eğitim seti S olsun [68]:

$$S = \{(x_i, y_i), i = 1, 2, \dots, n\} \quad (3.11)$$

Burada x_i d -boyutlu özellik vektörüdür, $y_i \in \{1, 2, \dots, c\}$ karşılık gelen sınıf etiketleridir, c sınıf sayısıdır ve n gözlem sayısıdır.

x_i ve x_j iki gözlemi arasındaki d_w uzaklık fonksiyonu w ağırlık vektörü cinsinden ifade edilir [68]:

$$d_w(x_i, x_j) = \sum_{r=1}^d w_r^2 |x_{ir} - x_{jr}| \quad (3.12)$$

d_w özellik ağırlıklarıdır. S eğitim kümesinde sınıflandırma doğruluğunu en üst düzeye çıkarmak için bir dışarıda bırakma tekniği düşünülür. Referans noktası olarak S 'den rastgele bir nokta seçilir ve buna göre etiketlenir. Referans noktası x , bir olasılık dağılımı ile S 'den seçilir. Referans noktası şu şekilde tanımlanır [68]:

$$p_{ij} = \frac{k(d_w(x_i, x_j))}{\sum_{j=1}^n k(d_w(x_i, x_j))} \quad (3.13)$$

burada k çekirdek işlevidir, $k(z) = \exp\left(-\frac{z}{\sigma}\right)$ ve σ çekirdek genişliğidir.

Doğru sınıflandırmanın ortalama bir dışarıda bırakma olasılığı, rasgele sınıflandırıcının S^{-i} kullanarak gözlem i 'yi doğru şekilde sınıflandırma olasılığı p_i 'dir [68]:

$$p_i = \sum_{j=1, j \neq i}^n p_{ij} y_{ij} \quad (3.14)$$

$$y_{ij} = \begin{cases} 1, & \text{if } y_i = y_j \\ 0, & y_i \neq y_j \end{cases} \quad (3.15)$$

Bu nedenle, yaklaşık bir dışarıda bırakılan sınıflandırma doğruluğu şu şekilde yazılabilir [68]:

$$F(w) = \sum_{i=1}^n p_i \quad (3.16)$$

KBA'nın amacı, düzenleme terimi (λ) ile w 'ye göre $F(w)$ 'yi maksimize etmektir [68]:

$$F(w) = \sum_{i=1}^n p_i - \lambda \sum_{r=1}^d w_r^2 \quad (3.17)$$

Burada $\lambda > 0$, çapraz doğrulama yoluyla ayarlanabilen bir düzenleme parametresidir [68].

$$F(w) = \sum_{i=1}^n \sum_{j=1, j \neq i}^n p_{ij} y_{ij} - \lambda \sum_{r=1}^d w_r^2 \quad (3.18)$$

$F(w)$ türevlenebilir olduğundan, w_r 'ye göre türevi şu şekilde hesaplanabilir [68]:

$$\begin{aligned} \frac{\partial F(w)}{\partial w_r} &= \sum_{i=1}^n \sum_{j=1}^n y_{ij} \left[\frac{2}{\sigma} p_{ij} \left(\sum_{k \neq i} p_{ik} |x_{ir} - x_{kr}| - |x_{ir} - x_{jr}| \right) w_r \right] - 2\lambda w_r \\ &= \frac{2}{\sigma} \sum_{i=1}^n \left(p_i \sum_{k \neq i} p_{ik} |x_{ir} - x_{kr}| - \sum_{j=1}^n y_{ij} p_{ij} |x_{ir} - x_{jr}| \right) w_r - 2\lambda w_r \\ &= 2 \left(\frac{1}{\sigma} \sum_{i=1}^n \left(p_i \sum_{k \neq i} p_{ij} |x_{ir} - x_{jr}| - \sum_{j=1}^n y_{ij} p_{ij} |x_{ir} - x_{jr}| \right) - \lambda \right) w_r \quad (3.19) \end{aligned}$$

Yukarıdaki türevi kullanmak, karşılık gelen gradyan yükseliş güncelleme denklemine yol açar. En iyi λ , minimum sınıflandırma kaybına karşılık gelir. Çapraz doğrulama kısımlarındaki ortalama kayıp, λ seçimine bağlıdır. λ ayarı, minimum sınıflandırma kaybıyla sonuçlanır ve $1/N_t$ ile hesaplanır; burada N_t , eğitim setindeki toplam gözlem sayısıdır. Aşağıdakiler tarafından verilen görel bir eşik (T) tanımlamaya dayalı özellik ağırlıklarını kullanarak önemli özellikleri seçme kriterleri [68]:

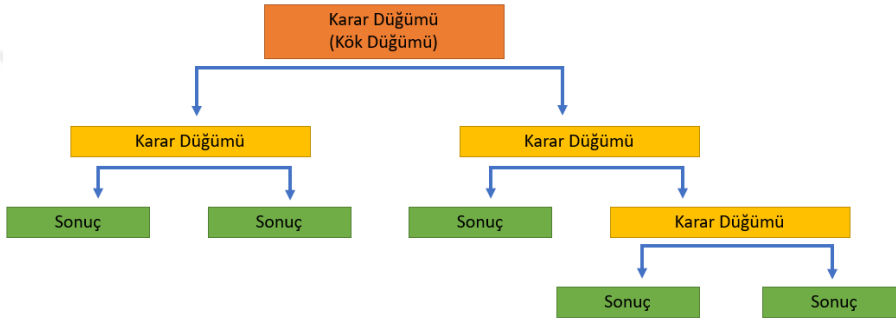
$$T = \tau * \max(w) \quad (3.20)$$

Burada τ , 0.02'ye sabitlenmiş toleranstır ve w , özelliklerin güncellenmiş ağırlığıdır. T 'den daha büyük özellik ağırlığı, önemli özellikler olarak sayılır ve kalan özellikler S kümesinden çıkarılır. λ çok büyük olduğunda, tüm özellik ağırlıkları sıfıra yaklaşır ve sonuçta tümü alakasız özelliklerdir. Bu nedenle, sınıflandırma kaybını en aza indiren λ 'yı ayarlamak esastır [66,68].

3.6 Sınıflandırıcılar

3.6.1 Karar ağaçları

Karar Ağaçları, ağaç yapısına sahip bir karar verme yöntemidir ve sınıflandırmaya tamamen farklı bir yaklaşım sunmaktadır. Bir karar ağacı, bir dizi eğer-o zaman kuralı geliştirir. Her kural, ağacın bir bölümüne iletir ve bu noktada başka bir eğer-o zaman kuralı uygulanır. Tüm veri setini içeren ilk bölüm, karar ağacının kök düğümüdür. Karar ağaçları, ikiden fazla olası sonucu olan nominal sonuçlarla ve sıralı sonuç değişkenleriyle çalışır [69]. Şekil 3.19'da bir karar ağacı şematik olarak gösterilmiştir.



Şekil 3.19 : Karar ağacı örnek şeması.

Sınıflandırma ve regresyon sırasında ayrıştırma uygulanacak kriterler genellikle değişkendir ve algoritmanın performansını önemli ölçüde etkiler. Ayrıştırmanın nasıl uygulanacağı karar ağaçlarında metodun doğruluğunu etkileyen en önemli faktörlerdendir. Bir düğümü birçok alt düğümde ayırabilmek amacıyla pek çok metod kullanır. Uygulanacak algoritmanın seçimi, hedef değişkenin türüyle ilgilidir. Kategorik tabanlı değişkenler için Gini, Entropi ve Sınıflandırma Hatası metodları yüksek oranda tercih edilmektedir. Sürekli değişkenlerde ise En Küçük Kareler algoritması oldukça sık kullanılmaktadır [70].

Entropi, elementlerin kirlilik ölçüsüdür. Bu nedenle mevcut verilerin yüksek doğrulukta sınıflandırılabilmesi amacıyla entropi değeri mümkün olan en düşük seviyede alınarak bölme işlemi yapılmalıdır. Karar sisteminin çalışma verimliliği bölme işleminin doğruluğuna bağlıdır. Ögelerin olasık hesabı Eşitlik 3.21'de verilmiştir [70].

$$H = -\sum p(x) \log p(x) \quad (3.21)$$

Burada, $p(x)$ belirli bir sınıfa dahil bir grubun yüzdesini ve H de entropiyi belirtmektedir. Entropi değerinin minimuma düşürülmesi amacıyla en uygun bölünmenin bulunması gerekmektedir. En uygun bölünmenin seçimi amacıyla Eşitlik 3.22'deki bilgi kazancı hesabı yapılmaktadır [70].

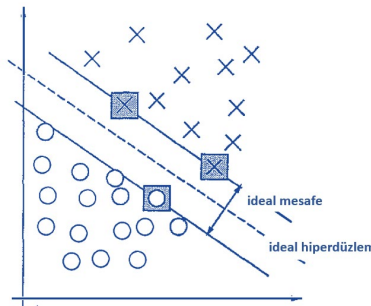
$$\text{Bilgi Kazancı } (S, D) = H(S) - \sum_{V \in D} \frac{|V|}{|S|} H(V) \quad (3.22)$$

Burada, S toplam veri kümesini, D veri setinin bölünmüş bir parçasını göstermektedir. Her V , S 'nin ayrı bir alt kümesidir. V değerlerinin tamamı ayrıktır, bunların birleşimi S veri kümesini oluşturmaktadır. Bilgi kazancı, bölünmeden önceki orijinal S 'nin entropisi ile her öznitelik için hesaplanan entropinin farkı ile bulunmaktadır [70].

3.6.2 Destek Vektör Makineleri

Destek Vektör Makineleri (DVM), problemi temsil eden özellikler ile sınıflar arasındaki farklılığı belirleyen bir sınıflandırma algoritmasıdır. DVM ile eğitilmiş bir modele farklı bir test verisi verilerek bunun hakkında tahmin veya genelleme yapılabilmektedir. İki sınıflı problemlerde olduğu kadar çok sınıflılarda da doğrusal olmayan sınıflandırıcının gerekli olması halinde kullanılabilir ve performans başarısı oldukça yüksektir [49].

DVM ile ikili sınıflandırma yapılırken, Şekil 3.20'de gösterildiği gibi iki sınıfı ayıracak hiperdüzlem tespit edilmektedir ve test edilerek sınıflandırma sonuçları elde edilmektedir [71].



Şekil 3.20 : 2 boyutlu uzayda ayrılabilir bir örnek [71].

Burada kesikli çizgi hiperdüzlemi, düz çizgiler sınır düzlemlerini, sınır düzlemleri arasındaki mesafe ideal mesafeyi ve halkalar ile çarpılar da farklı sınıfları ifade etmektedir. Hiper düzlem sınır düzlemlerine eşit mesafededir ve $w^T x + b = 0$

denklemini gösterilir. Bu denklemde w ağırlık vektörü, b eşik değeridir. İdeal mesafe ise $2/\|w\|$ değerindedir ve bu değerdeki $\|w\|$, w 'nin normudur.

İki sınıflı bir problemi ele alırsak, n adet x girdisi ve bu girdileri ifade eden y etiketinin bildiğimizi varsayalım. Etiket değerinin -1 ve 1 olduğu düşünülürse, sınır düzlemlerinin koşulları Eşitlik 3.23'te verilmiştir [49].

$$y_i(w \cdot x_i + b) \geq 1 \quad (3.23)$$

İdeal mesafenin maksimum olabilmesi için $\|w\|$ 'nin minimum değeri hesaplanmalıdır. Bu problemin çözümü Eşitlik 3.24'teki Lagrange Çarpanları yöntemi ile gerçekleştirilmektedir [49].

$$L(w, b, a) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^n a_i y_i (w \cdot x_i + b) + \sum_{i=1}^n a_i \quad (3.24)$$

Burada, a_i Lagrange çarpanlarıdır. Denklem çözümünde Eşitlik 2.24'ün w ve b 'ye göre kısmi türevleri alınmalı sıfıra eşitlenmelidir. Türev işlemi yapıldığında Eşitlik 3.25 ve Eşitlik 3.26'deki koşullar elde edilir [49].

$$W = \sum_{i=1}^n a_i y_i x_i \quad (3.25)$$

$$\sum_{i=1}^n a_i y_i = 0 \quad (3.26)$$

Bu koşulları Eşitlik 2.24'te yerine koyarsak Eşitlik 2.27 elde edilecektir. Böylece artık w ve b 'ye bağımlılık durumu ortadan kalkmaktadır. Yeni elde ettiğimiz Eşitlik 3.27'de a_i çözümü bulunarak w ağırlıkları hesaplanabilmektedir [49].

$$L(a) = \sum_{i=1}^n a_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n a_i a_j y_i y_j (x_i, x_j) \quad (3.27)$$

Son olarak karar fonksiyonu Eşitlik 3.28'deki gibi bulunmaktadır ve yeni bir test verisinin ilgili olduğu sınıfa atanması bu eşitlik ile yapılmaktadır [49].

$$f(x) = \text{sgn} \left[\left(\sum_{i=1}^n a_i y_i x_i \cdot u \right) + b \right] \quad (3.28)$$

Doğrusal olmayan problemlerde DVM uygulamalarında, eğitim verisine uyumlu bir dönüşüm yapılmaktadır. Farklı kernel kullanılarak doğrusal yaklaşımlar uygulanmakta ve bu şekilde sınıflandırma yapılmaktadır [49].

3.6.3 k-En Yakın Komşuluk Sınıflaması

Denetimli bir sınıflandırma algoritması olan k-NN çıktı bir sınıf üyeliğidir. Etiketlenmiş veriler üzerinden diğer verilerin etiketlenmesi gerçekleştirilmektedir. Algoritma yeni bir veriyi en yakın olan daha önceden etiketlenmiş verileri (en yakın komşuluklar) inceleyerek etiketlemektedir. Komşulardan aynı etikete uyanlar bir sayılarak oylama sistemi şeklinde sistem çalıştırılmaktadır. Yeni noktanın etiketi, komşularının içinde en fazla hangisi ise ona göre belirlenmektedir. Bu sistemde “ k ” kontrol edilen komşuların sayısını ifade etmektedir [72]. “ k ” değeri algoritmanın performansını büyük ölçüde etkilemektedir. Uzaklık ölçümünde kullanılan yöntemler de bu algoritmanın performansını belirleyen bir diğer özelliktir. Literatürde yaygın olarak kullanılan uzaklık hesaplama yöntemleri arasında Öklid uzaklığı, Chebyshev uzaklığı ve Minkowski uzaklığı bulunmaktadır [73]. x ve y noktalarının uzaklığı D , $x = (x_1, x_2, x_3, \dots, x_k)$ ve $y = (y_1, y_2, y_3, \dots, y_k)$ olarak alınırsa bu üç yöntem ile hesaplamalar aşağıdaki gibi yapılır.

Öklid uzaklığı, Öklid uzayında iki nokta arasındaki düz çizgi mesafesidir ve Eşitlik 3.29 ile hesaplanır [74].

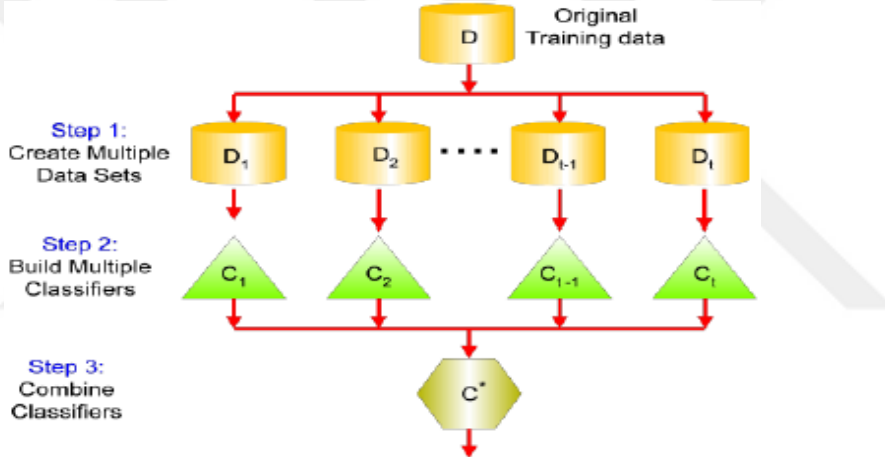
$$D = \sqrt{\sum_{i=1}^k (x_i - y_i)^2} \quad (3.29)$$

Minkowski uzaklığı için Eşitlik 3.30 kullanılır. Bu hesaplamadaki p değişkeninin farklı değerleri için farklı hesaplamalar belirlenmesi amacıyla Minkowski uzaklığı kullanılır. Farklı p değerleri için Öklid uzaklığı formülüne veya aynı p değeri için $k \rightarrow \infty$ durumu gerçekleşirse Chebyshev uzaklığı hesaplamasına eşitlenmektedir [73].

$$D = \left(\sum_{i=1}^k |x_i - y_i|^p \right)^{1/p} \quad (3.30)$$

3.6.4 Topluluk öğrenmesi yöntemleri

Topluluk öğrenmesi yöntemleri (Ensemble yöntemler), aynı problemi çözmek için birden fazla öğrencinin eğitilmesine dayanan bir makine öğrenmesi modelidir. Şekil 3.21'de gösterildiği gibi, eğitim verilerinden bir hipotez öğrenmeye çalışan sıradan makine öğrenmesi yaklaşımlarından ayrı olarak, topluluk yöntemleri bir dizi hipotez oluşturmaya ve bunları kullanmak üzere birleştirmeye çalışmaktadır. Hepsisi, çoğu zaman tek bir model üzerinde tahmin performansını artırabilmektedir. Sınıflandırma görevlerine ve sayısal tahmin problemlerine uygulanabilen genel tekniklerdir [75].



Şekil 3.21 : Topluluk öğrenmesi genel şeması [75].

Bir topluluk yöntemi, temel öğrenciler olarak adlandırılan birkaç öğrenciden oluşur. Bir topluluğu genelleştirme yeteneği, genellikle temel öğrencilerden çok daha güçlüdür. Topluluk öğrenme yöntemi, doğru kararlar verebilen güçlü öğrenciler olmak için rastgele tahminden daha iyi performans gösteren zayıf öğrencileri geliştirmek açısından çok güçlüdür. Bununla birlikte, teorik analizlerin çoğu zayıf öğrenciler üzerinde çalışsa da, pratikte kullanılan temel öğrencilerin, o kadar da zayıf olmayan temel öğrencilerin kullanılması genellikle daha iyi performansla sonuçlandığından, mutlaka zayıf olmamaları dikkate alınmalıdır [76].

Bir topluluk sınıflandırıcısının tek bir sınıflandırıcıdan daha iyi performans göstermesi için 2 temel koşul vardır:

- 1) temel sınıflandırıcılar birbirinden bağımsız olmalıdır.

- 2) temel sınıflandırıcılar, rastgele tahmin gerçekleştiren bir sınıflandırıcıdan daha iyisini yapmalıdır.

Topluluk sınıflandırıcıları, her biri en azından rastgele tahminden daha doğru olan birden fazla bağımsız ve çeşitli kararı birleştirmektedir, bunun için rastgele hatalar birbirini götürmekte ve doğru kararlar pekiştirilmektedir.

Veriler bazı örnekleme dağılımlarına göre yeniden örneklenecek çoklu eğitim setleri oluşturulabilmektedir. Örnekleme dağılımı, eğitim için bir örneğin seçilme olasılığını belirlemekte; bir denemeden diğerine değişebilmektedir. Sınıflandırıcı, belirli bir öğrenme algoritması kullanılarak eğitim setlerinin her birinden oluşturulmaktadır [77].

3.6.4.1 Artırma

Artırma, başlangıçta, sınıflandırma performansını iyileştirmek için birkaç zayıf sınıflandırıcıyı birleştirmek için önerilmiştir. Daha sonra, “AdaBoost” olarak adlandırılan daha yetenekli ve pratik bir artırma algoritması önerilmiştir. Adaptive Boosting'in kısaltması olan Ada-Boost, bir meta-öğrenme algoritmasıdır. Önceki zayıf sınıflandırıcıların performansına göre diğerleri üzerinde yinelemeli olarak zayıf bir sınıflandırıcı oluşturmaya çalışır. Buna göre AdaBoost, önceki zayıf sınıflandırıcılarla doğru bir şekilde sınıflandırılmayan örneklere daha fazla ağırlık vererek odaklanmaya yönlendirilir. AdaBoost, zayıf sınıflandırıcılar rastgeleden daha iyi performans gösterdiği sürece eğitim hatalarını azaltabilir [78]. AdaBoost'un çok iyi bir genellemeye (yeni verileri sınıflandırma yeteneği) sahip olduğu gözlemlenmiştir [79,80]. Ancak, diğer sınıflandırıcıların çoğu gibi AdaBoost da çok gürültülü verilerle uğraşırken aşırı uyum problem yaşamaktadır [81]. Bu durumla başa çıkmak için Friedman ve arkadaşları [82], LogitBoost kullanmanın eğitim hatalarını doğrusal olarak azaltabileceğini ve dolayısıyla daha iyi genelleme sağlayabileceğini bulmuştur.

İkili LogitBoost

AdaBoost, etkin bir şekilde çözülebilen üstel kayıp fonksiyonu $ELOSS(F) = E(e^{-yF(\vec{x})})$ beklentisini en aza indirmek için bir eklemeli lojistik regresyon modeli $F(\vec{x}) = \sum_{t=1}^T \alpha_t f_t(\vec{x})$ olarak düşünülebilir [82]. Ancak, üstel kayıp işlevi, sınıflandırma hatasıyla katlanarak değişmekte ve gürültülü verileri işlerken

AdaBoost algoritmasını işlevsiz hale getirmektedir. Bu problemi çözmek için Friedman ve arkadaşları [82], sınıflandırma hatasıyla doğrusal olarak değişen ve gürültü ve aykırı değerlere karşı daha az duyarlı olduğu ortaya çıkan bir binom log-olasılık kayıp fonksiyonu $LLOSS(F) = E[-\log(1 + e^{-yF(\vec{x})})]$ önermiştir. Optimizasyon, eklemeli simetrik bir lojistik modele uyması için Newton adımları kullanılarak gerçekleştirilebilmektedir [83].

Zayıf sınıflandırıcıların yapısı, güçlendirme algoritmalarının performansını etkileyen kilit faktörlerden biridir. Karar ağaçları, güçlendirme için zayıf sınıflandırıcılar için oldukça uygun olan en küçük kareler hatasını en aza indirmek için girdi alanını iç içe bölgelere, genellikle dikdörtgenlere bölmeye çalışmaktadır [83].

3.7 Patolojik Seslerin Belirlenmesinde Öne Çıkan Çalışmalar

Patolojik ses bozukluklarının otomatik olarak saptanması, zorlu ve önemli bir tıbbi sınıflandırma sorunudur ve araştırmacılar tarafından geniş çapta araştırılan bir alandır. Patolojik seslerin belirlenmesine yönelik çalışmalar yalnızca teknik açıdan değil, aynı zamanda hem analiz için seçilen ses kayıtları hem de sınıflandırma için patolojilerin alt kümelerinin nasıl kullanıldığı açısından farklılık göstermektedir.

Araştırma çalışmalarının çoğu, Massachusetts Göz ve Kulak Kliniği (Massachusetts Eye and Ear Infirmary, MEEI) veritabanını kullanır. Ancak bu veri tabanındaki sağlıklı ses kayıtları ve patolojik ses kayıtları iki farklı ortamda [84] kaydedilmektedir ve bu da ayırt edici olanın ortamlar mı yoksa ses özellikleri mi sorusunu doğurmaktadır. Saarbrücken Ses Veritabanı (Saarbruecken Voice Database, SVD), 50 kHz'de ve 16 bit çözünürlükte örneklenen tüm kayıtları içeren indirilebilir bir veritabanıdır. Bu veritabanında kayıtlar aynı ortamda kaydedildiği için bu iş için ideal bir veri tabanıdır.

Muhammad ve arkadaşları [85] yaptıkları çalışmada ses üretim teorisine dayalı bir otomatik ses patolojisi tespit sistemi geliştirmiştir. Ses yolu alanından çıkarılan özellikleri kullandıkları bu çalışmada sürekli sesli harf örnekleriyle MEEI veri tabanından kist, felç ve polip tanısı almış hastaların ses kayıtları üzerinde çalışmışlardır. Sınıflandırıcı olarak DVM kullandıkları bu çalışmada yaklaşık %99 doğruluk elde etmişlerdir.

Hammami ve arkadaşları [86] yaptıkları çalışmada DVM'ye dayalı ses bozukluklarının tanımlanması için bir yöntem önermiştir. SVD veri tabanını kullandıkları bu çalışmada 80 sağlıklı 80 patolojik tanı almış kişinin sürekli /a/ sesi kayıtlarını kullanmışlardır. Ses sinyallerinden HFD ile F0 ve ilk üç formantı çıkarıp DVM sınıflandırıcısı için girdi vektörü olarak kullanılmışlardır. %86 doğruluk elde ettikleri bu çalışmada DVM sınıflandırıcısının iyi performans gösterdiğini gözlemlemişlerdir.

Al-nasheri ve arkadaşları [87] yaptıkları çalışmada ses patolojilerini otomatik olarak tespit etmek ve sınıflandırmak için Çok Boyutlu Ses Programı (Multi-Dimensional Voice Program, MDVP) ile akustik analiz yaparak elde ettikleri parametreleri kullanmışlardır. Çalışmada üç farklı veri tabanından (AVPD, MEEI ve SVD) ortak patolojilerin sürekli /a/ sesi kayıtlarını kullanmışlardır. Sınıflandırıcı olarak DVM kullandıkları bu çalışmada Fisher ayrım oranına göre düzenlenmiş en yüksek dereceli üç MDVP parametresini kullanarak SVD'de %99.68, MEEI'de %88.21 ve AVPD'de %72.53 doğruluk elde etmişlerdir.

Literatürde önerilen çözümlerin çoğu, makine öğrenmesi algoritmaları ile birleştirilen sağlam öznitelik tanımlayıcılarına dayanmaktadır. Başarılarına rağmen, belirli sınıflandırma görevleri için optimal olan özellikleri manuel olarak tasarlamak pratikte zordur. ESA, sınıflar arasındaki seçici özelliklerin otomatik tespitini yapabilen bir derin sinir ağıdır. Öznitelikler bu ağlarda evrişim katmanları sayesinde otomatik olarak çıkarıldığından, öznitelik çıkarımının manuel olarak yapılmasının neden olduğu sorunlara çözüm sunmaktadır.

Derin öğrenme modelleri genellikle iki boyutlu sinyalleri işlemek amacıyla kullanılmaktadır, ancak genel olarak biyolojik sinyaller tek boyutludur. Bazı araştırmalar, tek boyutlu sinyallere KZFD [88-90] uygulayarak spektrogram görüntülerini elde etmekte ve ağı beslenmesini bu görüntüler ile sağlamaktadır. Böylelikle sadece zaman-genlik bilgisi yerine zaman ve frekans özellikleri birlikte değerlendirilerek tek boyutlu biyolojik sinyallerin çok boyutlu değerlendirmesinin yapılması mümkün olmaktadır.

Ses sinyalleri de tek boyutlu biyolojik sinyallerdir ve bu sinyaller için de spektrogram görüntülerinin elde edilmesi ESA'nın beslenmesi amacıyla mantıklı olmaktadır. Aynı zamanda zaman alanı gösterimi ile karşılaştırıldığında, spektral

alan sinyalleri daha fazla patolojik bilgi içermektedir. Bazı araştırma çalışmaları [91,92], perde, formantlar, HNR, NNE ve Glottal-Gürültü Oranı (Glottal-to-Noise Ratio, GNR) gibi özelliklerin tümünün spektrogramlarda görülebilen ses kısıklığı, nefes alma ve pürüzlülük gibi patolojik sesin bazı özelliklerini temsil edebileceğini göstermiştir.

Derin öğrenme teknikleri konuşma tanıma alanında önemli ilerleme kaydetmiş olsa da patolojik ses bozukluklarının tespiti alanında henüz az araştırma çalışması yapılmıştır.

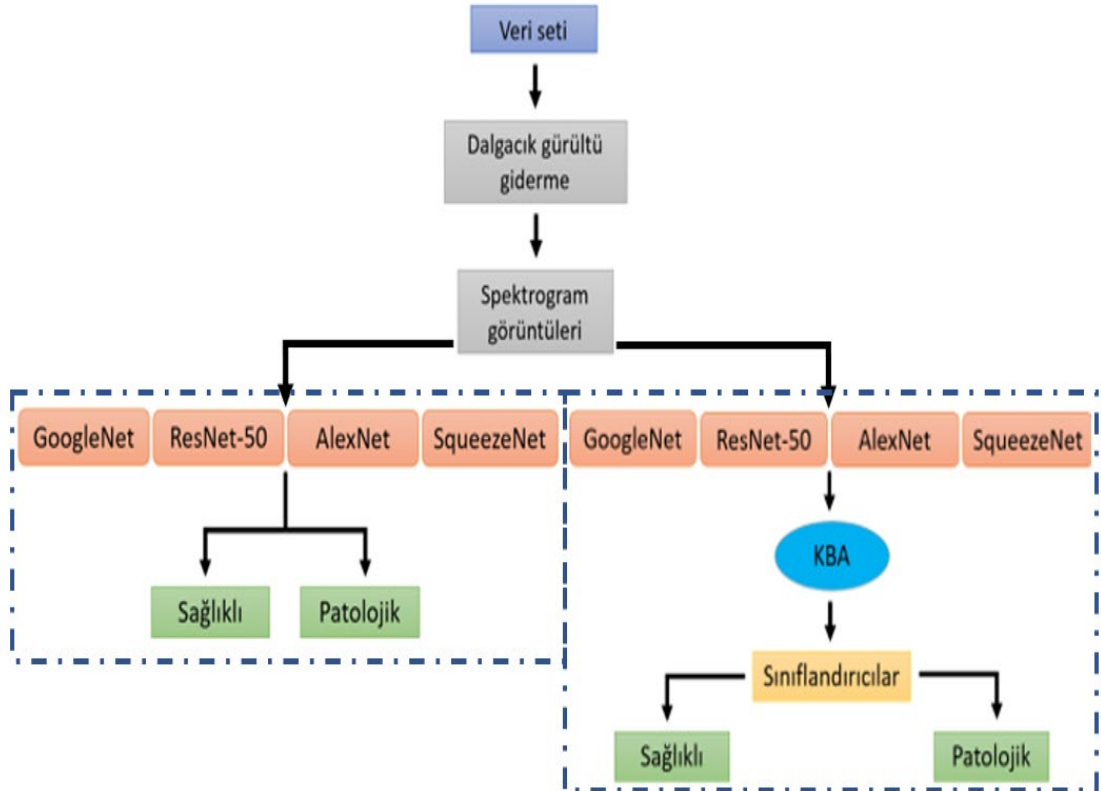
Wu ve arkadaşları [90] yaptıkları çalışmada patolojik sesi ve sağlıklı sesi otomatik olarak ayırt etmek için SVD'den 482 sağlıklı ve 482 patoloji teşhisi olan nötr perdede /a/ sesi kayıtlarını kullanmışlardır. Temel mimari olarak ESA kullandıkları bu çalışmada girdi olarak normal ve patolojik ses kayıtlarının spektrogramlarını kullanmışlardır. Uyguladıkları mimari ile %71 doğruluk elde etmişler ve derin öğrenme yaklaşımı ile sınıflandırmada iyi sonuçlar elde etmek için az miktarda verinin kullanılabileceği göstermişlerdir.

Huckvale ve Buciuleac yaptıkları çalışmada [93] patolojik ve sağlıklı sesi ayırt etmek için SVD'den ses kayıtları kullanmışlardır. Mimari olarak ResNet-50 ve girdi olarak ses kayıtlarının spektrogramlarını kullanmışlardır. Tüm patolojileri dahil ederek patolojilerin yalnız bir sesli harf kayıtlarını kullandıklarında %70 doğruluk elde etmişlerdir. Tüm patolojiler yerine yalnız organik patolojileri kullandıklarında doğrulukta yaklaşık %2'lik bir artış olduğunu gözlemlemişlerdir. Ayrıca yaptıkları denemelerde cinsiyetin performans üzerinde çok az etkisi olduğunu gözlemlemişlerdir.

Alhussein ve Muhammad yaptıkları çalışmada [94] patolojik seslerin belirlenmesi amacıyla SVD'den ses kayıtları kullanmışlardır. 1616 patolojik ve 686 sağlıklı ses kaydı ile gerçekleştirdikleri çalışmada nötr perdede sürekli /a/ sesi kayıtlarını kullanmışlardır. Çalışma kapsamında VGG16 ve CaffeNet mimarilerini karşılaştırmışlar ve sırasıyla %93.50 ile %93.90 doğruluk elde etmişlerdir.

4. MATERYAL VE YÖNTEM

Bu bölümde ilk olarak kullanılan veri setine, veri setinden seçilen patolojilere ve seçim kriterlerine değinilmektedir. Ardından, sinyaller üzerinde uygulanan ön işleme metodları, neden bu metodların seçildiği ve kullanıldığı konusunda bilgi verilmektedir. Sonrasında, literatürde sıklıkla kullanılan dört farklı ESA mimarisinin çalışmada patolojik seslerin belirlenmesinde kullanımı ve seçilen parametreler verilmektedir. Daha sonra öznelik elde edilmesinde kullanılan ESA mimarileri seçilen parametreleriyle birlikte verilmekte ve öznelik seçiminde kullanılan yöntem açıklanmaktadır. Son olarak ESA mimarilerinden elde edilen ve indirgenen özneliklerin sınıflandırılması amacıyla kullanılan makine öğrenme yöntemleri kullanılan fonksiyonlar ve seçilen parametreleriyle verilmektedir. Tez çalışması kapsamında önerilen modeller için genel akış şeması Şekil 4.1’de verilmektedir.



Şekil 4.1 : Tez çalışması akış şeması.

4.1 Kullanılan Veri Seti

Bu tez kapsamında, Almanya'daki Saarland Üniversitesi Fonetik Enstitüsü tarafından kaydedilen Saarbruecken ses veritabanı kullanılmıştır (Url-5). Bu veritabanı 2000'den fazla kişiden alınan konuşma kayıtları ile 71 farklı patolojiyi içermektedir. Her katılımcı dosyası için nötr, alçak, yüksek ve alçak-yüksek-alçak tonlamalarda uzun süreli /a/, /i/ ve /u/ sesli harflerinin kayıtlarını ve sürekli bir konuşma cümlesi olan “Guten Morgen, wie geht es Ihnen?” (“Günaydın nasılsın?”) ses kayıtlarını içermektedir. Bu veritabanındaki ses kayıtlarının örnekleme frekansı 50 kHz ve çözünürlüğü 16 bittir. Sürekli /a/ ünlüsü, zaman içinde durağan olduğu ve değişiklikleri daha kolay gördüğü için bu çalışmada kullanılmıştır.

Patolojik grup olarak larenjit, lökoplazi, Reinke ödemi, rekürren laringeal sinir felci, vokal kord karsinomu ve vokal kord polipleri olmak üzere 6 patoloji seçilmiştir. Bu patolojilerin tümü, vokal kordlardaki yapısal değişikliklerin neden olduğu organik disfoniye sebep olmaktadır. 380'i sağlıklı ve 380'i patolojik (82 larenjit, 15 lökoplazi, 37 Reinke ödemi, 197 rekürren laringeal sinir felci, 20 vokal kord karsinomu ve 29 vokal kord polibi) her bireyin nötr perdesinde sürekli sesli /a/ sesi kullanılmıştır. Veriler, sırasıyla %75 ve %25 örnek içeren eğitim seti ve test seti olarak ayrılmıştır.

- a) Larenjit: Gırtlak dokularının iltihaplanmasıdır. Akut ve subakut formları ani başlangıçlıdır ve 3 haftadan kısa sürmektedir. Kronik larenjit genellikle yavaş gelişmektedir [95].
 - 1) Bulaşıcı larenjit: Virüsler, bakteriler, mantarlar veya küflerin gırtlak enfekte etmesidir.
 - 2) Viral Larenjit: Enfekte kişiler tarafından hava damlacıkları yoluyla bulaş gerçekleşir [96].
 - 3) Mekanik Larenjit: Ses üretimi esnasında vokal kordların çarpışma kuvvetlerinin yüksek olması sonucu ortaya çıkmaktadır.

Bu hastalarda laringeal travma, alışılmış boğaz temizleme davranışları ve kronik öksürük görülür. Boğuk-nefesli, gergin, sert ve düşük perdeli ses anormallikleri görülmektedir.

- 4) Kronik Tahriş Edici ve Kontakt Larenjit: Haftalar/aylar süren ve bu süre esnasında şiddeti değişkenlik gösterebilen larenjit belirtileri görülmektedir. Kalıcı veya aralıklı boğaz ağrısı, globus hissi, boğuk

sesler, odinofaji ve öksürük nöbetlerine neden olmaktadır. Laringofaringeal reflü, zararlı/toksik kimyasalların solunması/yutulması, oral inhale ilaçlar ve solunan/yutulmuş alerjenler sebep olmaktadır [97].

- b) Lökoplazi: Müköz membranlardaki beyaz lezyonları tanımlayan klinik bir terimdir. Laringeal lökoplazi, skuamöz hücreli karsinom için prekanserözdür. Sigara, alkol, reflü, viral enfeksiyonlar, sesin kötü kullanımı, toksik gaz inhalasyonu nedeniyle ortaya çıkmaktadır. Tanısında direkt laringoskopi kullanılmaktadır. Patoloji tanısında en önemli parametre ise displazi varlığı ile displazinin şiddetidir.
- c) Reinke Ödemi: Reinke boşluğu, jelatinöz özellikte bir gevşek bağ dokusundan oluşmaktadır. Bu alanda lenfatik drenaj bulunmamaktadır. Buradaki patolojiler mukozal vibratuar dalgaları bozacağı için disfoniye neden olmaktadır. Sigara içmek ve sesin kötüye kullanılması başlıca nedenleridir.
- d) Rekürren Laringeal Sinir Felci: Tiroidektomi sonrası en önemli komplikasyonlardan biri vokal disfonksiyondur. Bazı ses değişiklikleri geçicidir, bazıları ise uzun süreli kalmaktadır. Vokal disfonksiyonun alışılmış bir nedeni rekürren laringeal sinirin yaralanmasıdır [98-100]. Bu da vokal kordlarda mobilitede ve atrofiye bozulma [101], fonasyon süresinde azalma, pertürbasyon parametrelerinde artış [102] ile sonuçlanmaktadır. Tiroid kanseri, Graves hastalığı ve tekrarlayan guatr için yapılan operasyonlar, önemli ölçüde daha yüksek rekürren laringeal sinir felcine yol açmaktadır.
- e) Vokal Kord Kanseri: Vokal kord kanserleri histolojik olarak incelendiğinde en büyük bir oranda yassı hücreli karsinomlar görülmektedir. Ana risk faktörü sigaradır [103]. Sebepleri arasında alkol, silika veya Asbeste maruziyet, nikel buharı / tekstil / kömür tozu maruziyeti bulunmaktadır [104]. Tanı genellikle önce aralıklı, sonra kalıcı olan ve giderek daha şiddetli büyüyen disfoniye takiben tanı konur [105]. Disfoni başlangıçta vokal kord mukozasındaki kusurlu vibrasyon sonucu görülmektedir ve vokal ligaman ve kasın tutulumu ile birlikte şiddetlenerek vokal kordların hareketliliğini bozmaktadır [106].
- f) Vokal Kord Polipleri: Vokal kordların benign lezyonlarından biridir. En sık ses kısıklığı semptomu ile karşımıza çıkarlar. Ses kalitesinde bozulmaya yol açarlar. İdiopatik, geçirilmiş bir üst solunum yolu enfeksiyonu kaynaklı

rezidüel ödem, akut vokal travma, sesin uzun süreli yanlış / kötü kullanımı vokal kord poliplerine sebebiyet vermektedir. Larinkste kronik enflamasyona sebep olan tüm faktörler de poliplere neden olabilmektedirler. Hastaların çoğunda sigara ve hatalı ses kullanımı görülmektedir. Poliplere en sık erkeklerde rastlanılmaktadır. Geniş tabanlı ya da saplı olabilmektedirler. Saplı olanlar serbestçe hareket edebilirler ve pozisyonları değişikçe seste ani değişiklikler / havayolu semptomları oluşabilir. Endolarengeal mikrocerrahi vokal kord poliplerinin en temel tedavisidir.

4.2 Ön İşleme

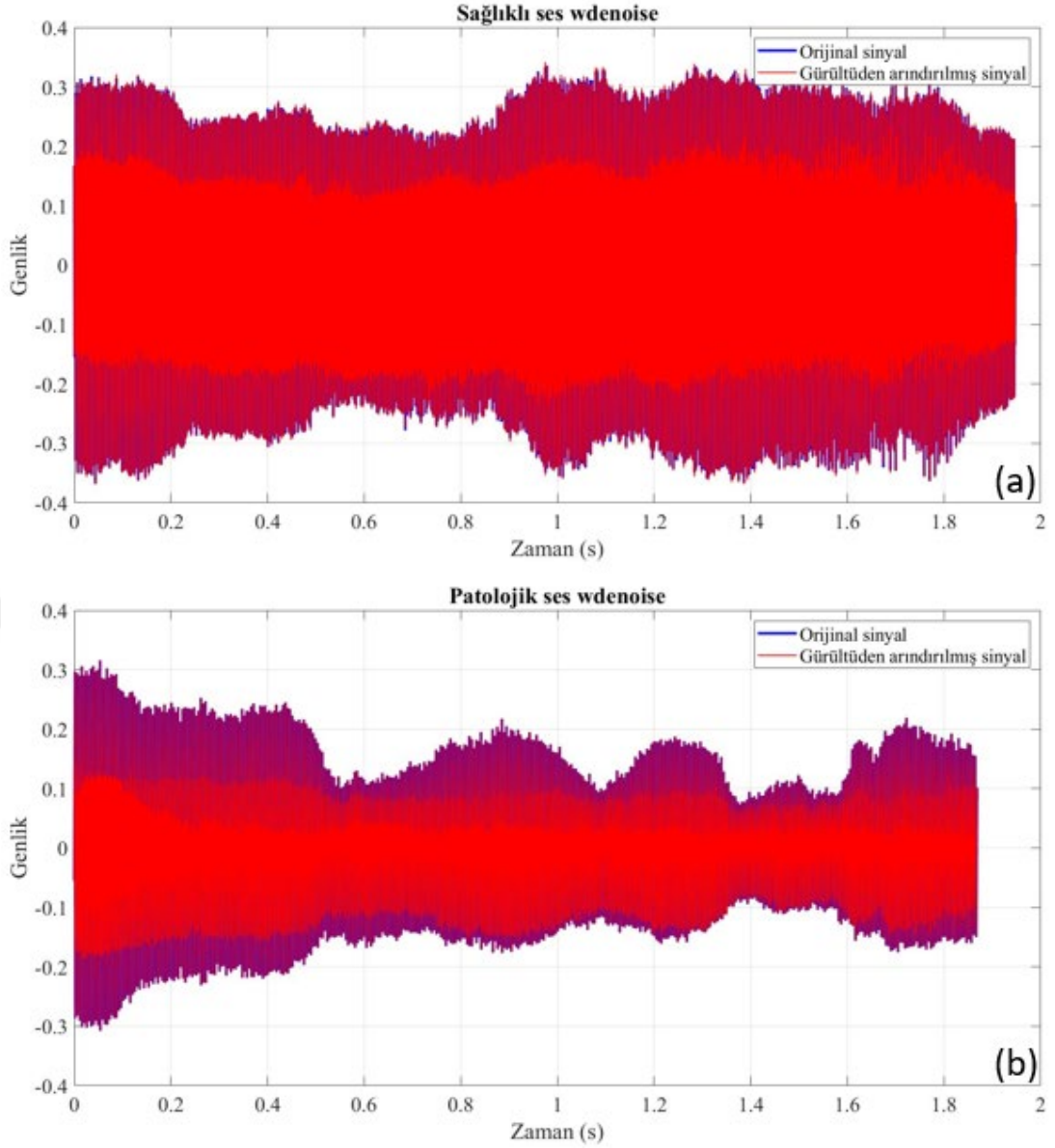
4.2.1 Ses sinyali gürültü giderme

Dautov ve Özerdem [107] ses sinyalleri üzerinde yaptıkları bir dizi deney sonucunda sinyal gürültü giderme uygulamalarının başarısı için doğru eşikleme tekniği ve dalgacık ailesinin seçilmesinin hayati önem taşıdığını gözlemlemişlerdir.

Mihov ve arkadaşları [108] dalgacık gürültü gidermede bütün uygulamalar için belirli bir dalgacığın uygun olmadığını ve her bir uygulama için uygun deneyler yapıldıktan sonra seçilecek dalgacığın ve eşikleme tekniğinin gürültü gidermede etkili yöntem olduğunu gözlemlemişlerdir.

Zhang ve arkadaşları [109], dalgacık tabanlı gürültü gidermenin, laringeal patolojilerden kaynaklanan fizyolojik bozuklukları teşhis etmek için analizin geliştirilmesinde faydalı olduğunu ileri sürmüştür. Daubechies ailesinden db10'nun 2. seviye ayrıştırmasının yaptıkları çalışma için en iyi gürültü gidermeyi sağladığını gözlemlemişlerdir.

Bu bilgiler ışığında çalışmada kullanılacak olan seslere ilk önce dalgacık gürültü giderme uygulanmıştır. Literatürde kullanılan dalgacıklar [107,109] ve çalışma için denenen bir başka dalgacık ile yapılan bir dizi deney sonucunda en iyi sonuç Daubechies ailesinden db2'nin 2. seviye ayrıştırması ile elde edilmiştir (Çizelge 4.1). Uygulama Matlab'ın wdenoise fonksiyonu ile yapılmıştır. Şekil 4.2 (a)'da örnek bir sağlıklı sese ve (b)'de örnek bir patolojik sese uygulanan dalgacık gürültü giderme işleminden kesitler gösterilmiştir.



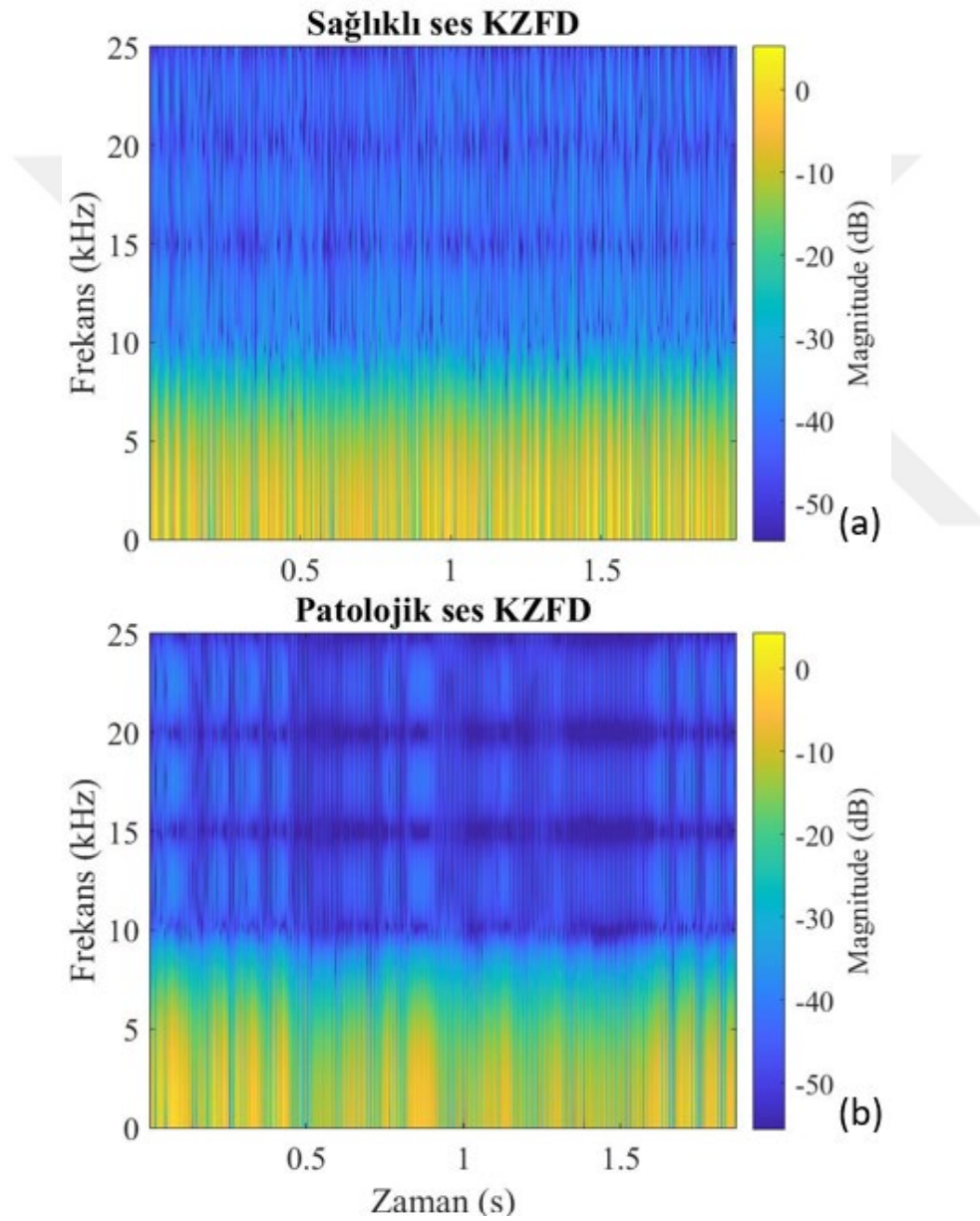
Şekil 4.2 : Örnek birer (a) sağlıklı ve (b) patolojik ses için dalgacık gürültü giderme.

4.2.2 Spektrogram görüntülerinin üretilmesi

Ses sinyalleri tek boyutlu sinyallerdir ve bu sinyaller için spektrogram görüntülerinin elde edilmesi ESA'nın beslenmesi amacıyla mantıklı olmaktadır. Aynı zamanda zaman alanı gösterimi ile karşılaştırıldığında, spektral alan sinyalleri daha fazla patolojik bilgi içermektedir. Bazı araştırma çalışmaları [91,92], perde, formantlar, HNR, NNE ve GNR gibi özelliklerin tümünün spektrogramlarda görülebilen ses kısıklığı, nefes alma ve pürüzlülük gibi patolojik sesin bazı özelliklerini temsil edebileceğini göstermiştir.

Bu bilgiler ışığında, ESA performansına bağlı olarak en uygun spektrogram görüntülerini elde edebilmek için için literatürde kullanılan [110,89,90] ve çalışma için ayrıca denenilen pencere fonksiyonları, boyutları, örtüşme değerleri seçilerek bir dizi deney yapılmıştır. Yapılan analizler sonucunda, Hamming penceresi tercih edilmiştir. Burada pencere uzunluğu 10 ms ve örtüşme değeri 5 ms olarak belirlenmiştir (Çizelge 4.1).

Şekil 4.3 (a)'da sağlıklı sese ve (b)'de patolojik sese dalgacık gürültü giderme ardından uygulanan KZFD sonucunda oluşan spektrogramlar gösterilmiştir.



Şekil 4.3 : Örnek birer (a) sağlıklı ve (b) patolojik ses sinyalinin gürültü giderme ardından KZFD uygulandıktan sonra elde edilen spektrogramlar.

Üretilen görüntüler hangi ağı girdi olarak verilecek ise o ağın giriş boyutuna göre yeniden boyutlandırılmıştır.

Çizelge 4.1 : Dalgacık gürültü giderme uygulaması ve spektrogram görüntülerinin oluşturulması için parametre değerlerinin belirlenmesi.

Dalgacık gürültü giderme		Spektrogram			GoogleNet
Dalgacık	Seviye	Pencere	Pencere Uzunluğu (ms)	Örtüşme Değeri (ms)	Doğruluk (%)
Db2	2	Hamming	10	5	67.90
Db2	2	Hamming	64	30	65.80
Db2	2	Hanning	10	5	62.10
Db2	2	Hanning	64	30	54.70
Db10	2	Hamming	10	5	65.80
Db10	2	Hamming	64	30	63.20
Db10	2	Hanning	10	5	59.50
Db10	2	Hanning	64	30	60.0
Db8	3	Hamming	10	5	61.60
Db8	3	Hamming	64	30	58.90
Db8	3	Hanning	10	5	60.50
Db8	3	Hanning	64	30	56.80

4.3 Farklı ESA Mimarileri ile Patolojik Seslerin Sınıflandırılması

Tez çalışmasının bu bölümünde patolojik seslerin belirlenmesi için üretilen spektrogram görüntüleri literatürde sıklıkla kullanılan farklı ESA mimarilerine girdi olarak verilerek sınıflandırma yapılmıştır.

İlk olarak GoogleNet mimarisi ile çalışılmıştır. GoogleNet 224x224x3 boyutunda girdi kabul ettiği için üretilen spektrogram görüntüleri bu boyuta göre yeniden boyutlandırılmıştır. GoogleNet mimarisinde eğitim parametreleri Mini-Batch Boyutu 8, maksimum Eğitim Tur (Epoch) Sayısı 1 ve öğrenme hızı 0.0001 olarak seçilmiştir.

İkinci olarak ResNet-50 mimarisi uygulanmıştır. ResNet-50 224x224x3 boyutunda girdi kabul etmektedir bu nedenle üretilen spektrogram görüntüleri bu boyuta göre yeniden boyutlandırılmıştır. ResNet-50 mimarisinde eğitim parametreleri Mini-Batch Boyutu 8, Epoch Sayısı 1 ve öğrenme hızı 0.0001 olarak seçilmiştir.

Üçüncü olarak AlexNet mimarisi ile çalışılmıştır. AlexNet 227x227x3 boyutunda girdi kabul ettiği için üretilen spektrogram görüntüleri bu boyuta göre yeniden boyutlandırılmıştır. AlexNet mimarisinde eğitim parametreleri Mini-Batch Boyutu 10, Epoch Sayısı 6 ve öğrenme hızı 0.0001 olarak seçilmiştir.

Üçüncü olarak SqueezeNet mimarisi uygulanmıştır. SqueezeNet 227x227x3 boyutunda girdi kabul ettiği için üretilen spektrogram görüntüleri bu boyuta göre yeniden boyutlandırılmıştır. SqueezeNet mimarisinde eğitim parametreleri Mini-Batch Boyutu 10, Epoch Sayısı 6 ve öğrenme hızı 0.0001 olarak seçilmiştir.

4.4 Seçilen ESA Mimarilerinden Öznitelik Çıkarımı ve Öznitelik Seçimi

Tez çalışmasının bu bölümünde seçilen ESA mimarilerini bu kez kendi sınıflandırıcılarından farklı algoritmalarla sınıflandırabilmek adına ESA mimarileri yalnızca öznitelik çıkarımı için kullanılmıştır. Öznitelik çıkarımı için üretilen spektrogram görüntüleri uygun boyutlandırma yapılarak seçilen ESA mimarilerine girdi olarak verilerek öznitelikler çıkarılmıştır.

Daha detaylı yani daha karmaşık seviyelerde özniteliklerin elde edilebilmesi için mimarinin katman derinliğinin artırılması gerekmektedir. Ancak öznitelikler karmaşıklıkla sınıflandırıcı başarısı da düşebilmektedir. Bu sebeple önerilen modelde katman derinliği probleme uygun olarak belirlenmelidir. Bunun yanısıra

yeterli sayıda özniteliğin sağlanması kullanılacak sınıflandırıcının performansını da etkileyebilmektedir [49].

Öznitelik çıkarımı için üretilen spektrogram görüntüleri ilk olarak GoogleNet'e girdi olarak verilmiştir. GoogleNet'in kendi sınıflandırmasıyla karşılaştırma yapılabilmesi adına yine aynı eğitim parametreleri seçilmiştir. Yapılan deneyler sonucunda GoogleNet'in birinci tam bağlantılı katmanı öncesindeki Küresel Ortalama Havuzlama katmanından elde edilen özniteliklerin sınıflandırmada daha başarılı sonuçlar sunduğu görülmüş ve bu katmandan 760 x 1024 boyutunda bir özellik haritası elde edilmiştir.

Spektrogram görüntüleri ikinci olarak ResNet-50'ye girdi olarak verilmiştir. ResNet-50'nin kendi sınıflandırmasıyla karşılaştırma yapılabilmesi adına yine aynı eğitim parametreleri seçilmiştir. Yapılan deneyler sonucunda ResNet-50'nin birinci tam bağlantılı katmanı öncesindeki Küresel Ortalama Havuzlama katmanından elde edilen özniteliklerin sınıflandırmada daha başarılı sonuçlar sunduğu görülmüş ve bu katmandan 760 x 2048 boyutunda bir özellik haritası elde edilmiştir.

Spektrogram görüntüleri üçüncü olarak AlexNet'e girdi olarak verilmiştir. AlexNet'in kendi sınıflandırmasıyla karşılaştırma yapılabilmesi adına yine aynı eğitim parametreleri seçilmiştir. Yapılan deneyler sonucunda AlexNet'in ikinci tam bağlantılı katmanından elde edilen özniteliklerin sınıflandırmada daha başarılı sonuçlar sunduğu görülmüş ve bu katmandan 760 x 4096 boyutunda bir özellik haritası elde edilmiştir.

Spektrogram görüntüleri son olarak SqueezeNet'e girdi olarak verilmiştir ve SqueezeNet'in farklı katmanlarından öznitelikler elde edilmiştir. SqueezeNet'in kendi sınıflandırmasıyla karşılaştırma yapılabilmesi adına yine aynı eğitim parametreleri seçilmiştir. Yapılan deneyler sonucunda SqueezeNet'in son evrişim katmanı öncesindeki dropout katmanından elde edilen özniteliklerin sınıflandırmada daha başarılı sonuçlar sunduğu görülmüş ve bu katmandan 760 x 100352 boyutunda bir özellik haritası elde edilmiştir.

ESA mimarilerinden elde edilen özellik haritalarının boyutlarını indirmek için KBA yöntemi kullanılmıştır. KBA yöntemi uygulaması için düzenleme terimi (λ) $1/N_t$ ve sabitlenmiş tolerans (τ) 0.02 olarak alınarak Matlab'ın fscnc fonksiyonu kullanılmıştır. KBA yöntemi uygulandıktan sonra GoogleNet'den elde edilen özellik haritasının boyutu 760 x 40, ResNet-50'den elde edilen özellik haritasının boyutu

760 x 174, AlexNet'ten elde edilen özellik haritasının boyutu 760 x 152 ve SqueezeNet'ten elde edilen özellik haritasının boyutu 760 x 273 olmuştur.

4.5 Diğer Sınıflandırma Algoritmaları

Derin öğrenme avantajıyla otomatik olarak çıkarılan öznelikler, mimarinin kendi sınıflandırıcısı yerine beş farklı sınıflandırıcı ile sınıflandırılmıştır. Bu sınıflandırıcılar Karar ağacı, DVM, k-NN, Ensemble ve bu dört sınıflandırma algoritmasından tasarlanan karar ağacıdır.

Karar Ağacında Matlab'ın fitctree fonksiyonu kullanılmıştır. Bu fonksiyon, girdi değişkenlerine dayalı olarak uygun bir ikili sınıflandırma karar ağacı döndürür. Karar ağacı uygulamasında Gini algoritması seçilmiştir. Yapılan deneyler sonucunda minimum dal düğümü gözlem sayısı 10 ve minimum yaprak düğümü gözlem sayısı 5 seçilmiştir (Çizelge 4.2).

Çizelge 4.2 : Karar ağacı parametrelerinin belirlenmesi.

Karar ağacı		DOĞRULUK (%)			
		GoogleNet	ResNet-50	AlexNet	SqueezeNet
A	B				
1	10	68.90	57.90	60.50	53.70
1	50	63.70	55.80	64.70	58.40
5	10	67.40	58.40	65.30	58.90

*Çizelgede A, Minimum dal düğümü gözlem sayısını; B, minimum yaprak düğümü gözlem sayısını ifade etmektedir.

DVM'de Matlab'ın fitcecoc fonksiyonu kullanılmıştır. Bu fonksiyon, bire-bir kodlama tasarımını kullanan ikili destek vektör makinesi modellerini kullanır. DVM uygulamasında yapılan deneyler sonucunda bir fark gözlenmemiş ve kodlama yöntemi olarak ikili öğrenici için bir sınıfın pozitif, diğerinin negatif ve geriye kalanların yok sayıldığı “onevsone” kodlaması seçilmiştir (Çizelge 4.3).

Çizelge 4.3 : DVM parametrelerinin belirlenmesi.

		DOĞRULUK (%)			
DVM		GoogleNet	ResNet-50	AlexNet	SqueezeNet
Kodlama					
onevsone		69.50	65.80	60.0	58.40
onevsall		69.50	65.80	60.0	58.40
binarycomplete		69.50	65.80	60.0	58.40

K-NN’de Matlab’ın fitcknn fonksiyonu kullanılmıştır. Bu fonksiyon, girdi değişkenlerine dayalı bir k-en yakın komşu sınıflandırma modeli döndürür. Yapılan deneyler sonucunda K-NN uygulamasında öklid uzaklığı seçilmiştir ve komşuluk sayısı 5 alınmıştır (Çizelge 4.4).

Çizelge 4.4 : k-NN parametrelerinin belirlenmesi.

		DOĞRULUK (%)			
k-NN		GoogleNet	ResNet-50	AlexNet	SqueezeNet
Uzaklık	Komşuluk				
Öklid	100	66.30	67.40	65.80	68.40
Öklid	50	67.40	72.60	66.30	69.50
Öklid	5	70.0	71.60	68.90	70.0
Chebyshev	100	64.20	55.30	66.80	64.20
Chebyshev	50	67.90	66.30	67.40	64.20
Chebyshev	5	67.40	63.70	67.40	65.30
Minkowski	100	66.30	67.40	65.80	68.40
Minkowski	50	67.40	72.60	66.30	70.0
Minkowski	5	70.0	71.60	68.90	69.50

Ensemble’da Matlab’ın fitcensemble fonksiyonu kullanılmıştır. Ensemble uygulamasında ikili sınıflandırma için LogitBoost metodu ve 100 sınıflandırma ağacını artırma yöntemi kullanılmaktadır. Bu uygulamada doğruluk performansının artırılabilmesi amacıyla kullanılan karar ağaçlarında dal düğümü sayısı değiştirilebilmektedir. Yapılan deneyler sonucunda dal düğümü sayısı 50 seçilmiştir (Çizelge 4.5).

Çizelge 4.5 : Ensemble parametrelerinin belirlenmesi.

		DOĞRULUK (%)			
Ensemble		GoogleNet	ResNet-50	AlexNet	SqueezeNet
Dal düğümü					
100		66.80	71.60	70.50	66.80
50		70.50	71.60	69.50	70.50
5		64.20	67.40	68.90	65.80

Duman ve arkadaşları yaptıkları çalışmada [111] algoritmaların sonuçlarına göre yeni bir karar ağacı tasarlamışlardır. Bu tez çalışmasında da kullanılan sınıflandırıcıların en az ikisinin patolojik sonucu verdiği durumda sonucun patolojik olarak kabul edildiği ve en az üçünün patolojik sonucu verdiği durumda sonucun patolojik olarak kabul edildiği iki karar ağacı tasarlanmıştır. Yapılan deneyler sonucunda en az ikisinin patolojik sonucu verdiği durumda sonucun patolojik olarak kabul edildiği karar ağacı tasarımı seçilmiştir (Çizelge 4.6).

Çizelge 4.6 : Tasarım karar ağacı parametrelerinin belirlenmesi.

		DOĞRULUK (%)			
Tasarım karar ağacı		GoogleNet	ResNet-50	AlexNet	SqueezeNet
A		70.50	67.90	68.40	68.40
B		68.40	68.90	64.20	63.70

*A, en az iki sınıflandırıcının patolojik sonucu verdiği durumda sonucun patolojik kabul edildiği; B, en az üç sınıflandırıcının patolojik sonucu verdiği durumda sonucun patolojik kabul edildiği karar ağacı tasarımı ifade etmektedir.

5. BULGULAR, DEĞERLENDİRME VE TARTIŞMA

Tüm ön işleme ve ESA modelleri MATLAB (R2021a) ortamında gerçekleştirilmiştir. Kullanılan bilgisayar Windows 10 64 bit işletim sistemine sahip olup, donanımsal olarak NVIDIA GeForce GTX960M ekran kartına, Intel(R) Core™i7-6700HQ işlemcisine, 16GB RAM belleğine sahiptir. Tüm deneyler gözetimli olarak gerçekleştirilmiştir.

Modelinin performansını değerlendirebilmek amacıyla yapılan deneyler on kez tekrarlanmıştır ve bulgularda bu tekrarlamalar sonucunda hesaplanan ortalama performans değerleri değerlendirilmiştir.

Uygulanan testlerin başarıları doğruluk, duyarlılık, özgüllük ve pozitif tahmin değeri (Positive Predictive Value, PPV) ile objektif bir şekilde değerlendirilmiştir (Eşitlik 5.1-Eşitlik 5.4).

$$\text{Doğruluk} = \frac{DP + DN}{DP + DN + YP + YN} \times 100 \quad (5.1)$$

$$\text{Duyarlılık} = \frac{DP}{DP + YN} \times 100 \quad (5.2)$$

$$\text{Özgüllük} = \frac{DN}{DN + YP} \times 100 \quad (5.3)$$

$$\text{PPV} = \frac{DP}{DP + YP} \times 100 \quad (5.4)$$

Doğru pozitif (DP): Doğru şekilde patolojik olarak tespit edilen seslerin sayısını gösterir.

Yanlış pozitif (YP): Sağlıklı olduğu halde patolojik teşhisi konulan seslerin sayısını gösterir. Tip 1 hata olarak kabul edilir.

Doğru negatif (DN): Doğru şekilde sağlıklı teşhisi konulan seslerin sayısıdır.

Yanlış negatif (YN): Patolojik olduğu halde sağlıklı teşhisi konulan seslerin sayısıdır. Tip 2 hata olarak kabul edilir. Patolojik seslerin sağlıklı olarak gösterilmesiyle tedavi

sürecine olumsuz etki edeceğinden Tip 1 hataya göre daha yüksek risk teşkil etmektedir.

Pozitif Olabilirlik Oranı (Positive Likelihood Ratio, LR⁺): Duyarlılık ve özgüllüğü birlikte kullanılmaktadır. Hasta bir bireydeki pozitif test sonucu olasılığını hasta olmayan bir bireydekiyle karşılaştırır. İyi bir test için bu oranın 2'den büyük olması beklenmektedir, ancak aynı zamanda duyarlılık değerine de dikkat edilmelidir.

$$LR^+ = \frac{\text{Duyarlılık}}{1 - \text{Özgüllük}}$$

5.1 Bulgular

İlk aşamada farklı ESA mimarileri ile patolojik seslerin sınıflandırılması yapılmıştır. GoogleNet uygulaması yaklaşık 7 dk, ResNet-50 17 dk, AlexNet 14 dk ve SqueezeNet 16 dk sürmüştür. ESA mimarilerinde gözlenen sonuçlar 10 kez tekrarın ortalaması şeklinde Çizelge 5.1'de verilmiştir. En yüksek doğruluk 66.88 ± 3.75 ile AlexNet'te gözlemlenmiştir. AlexNet'te pozitif olabilirlik oranı 2.94 olarak hesaplanmış olup diğer ESA mimarilerinden yüksektir, ancak AlexNet diğer mimarilerle karşılaştırıldığında en düşük duyarlılık değerine sahiptir. PPV değeri ise diğer mimarilerden yüksektir.

Çizelge 5.1 : Farklı ESA mimarileri ile patolojik seslerin sınıflandırma sonuçları.

	GOOGLENET	RESNET-50	ALEXNET	SQUEEZENET
Doğruluk (%)	62.68 ± 3.23	65.53 ± 3.05	66.88 ± 3.75	66.46 ± 2.87
Duyarlılık (%)	62.0 ± 20.14	59.60 ± 15.17	51.16 ± 9.64	59.58 ± 10.11
Özgüllük (%)	63.37 ± 21.34	69.47 ± 17.15	82.62 ± 8.08	73.37 ± 11.71
PPV (%)	66.19 ± 8.75	69.66 ± 10.93	75.58 ± 6.26	70.59 ± 7.07

İkinci olarak ESA mimarilerinden elde edilen ve seçilen öznelikler farklı sınıflandırıcılar ile sınıflandırılmış ve her bir durum için gözlenen sonuçlar 10 kez tekrarın ortalaması olacak şekilde Çizelge 5.2 – Çizelge 5.5'te verilmiştir.

Çizelge 5.2 sonuçları incelendiğinde GoogleNet'ten elde edilen ve seçilen 40 adet özneteliğin kullanıldığı çalışmada en yüksek doğruluk 66.04 ± 3.22 DVM ile sınıflandırmadan elde edilmiştir. Ancak DVM ile sınıflandırmada pozitif olabilirlik oranı 2.27 olarak hesaplanmasına rağmen duyarlılık değeri düşük görülmüştür. PPV değeri ise diğer sınıflandırıcılardan yüksektir. GoogleNet ile yapılan bu deney yaklaşık 8 dk sürmüştür.

Çizelge 5.2 : GoogleNet ile elde edilen ve seçilen özneteliklerin sınıflandırma sonuçları.

GOOGLENET	Doğruluk (%)	Duyarlılık (%)	Özgüllük (%)	PPV (%)
Ensemble	64.83 ± 3.75	60.73 ± 4.53	68.93 ± 7.26	66.53 ± 4.86
k-NN	63.22 ± 3.13	55.57 ± 4.65	71.05 ± 5.34	65.86 ± 4.12
DVM	66.04 ± 3.22	57.37 ± 3.82	74.73 ± 4.76	69.58 ± 4.43
Karar ağacı	62.0 ± 3.86	61.9 ± 4.84	62.09 ± 6.14	62.17 ± 4.43
Tasarım karar ağacı	65.10 ± 3.96	65.69 ± 4.45	64.53 ± 6.88	65.18 ± 4.63

Çizelge 5.3 sonuçları incelendiğinde ResNet-50'den elde edilip seçilen 174 adet öznetelikle gerçekleştirilen çalışmada en yüksek doğruluk 64.94 ± 2.77 ile Ensemble'dan elde edilmiştir, ancak pozitif olabilirlik oranı 1.75 hesaplanmıştır. ResNet-50 ile yapılan bu deney yaklaşık 21 dk sürmüştür.

Çizelge 5.3 : ResNet-50 ile elde edilen ve seçilen özneteliklerin sınıflandırma sonuçları.

RESNET-50	Doğruluk (%)	Duyarlılık (%)	Özgüllük (%)	PPV (%)
Ensemble	64.94 ± 2.77	69.48 ± 5.10	60.43 ± 4.03	63.73 ± 2.46
k-NN	62.85 ± 3.84	56.10 ± 5.93	69.56 ± 7.31	65.2 ± 5.67
DVM	62.84 ± 2.74	62.21 ± 4.24	63.48 ± 5.88	63.16 ± 3.19
Karar ağacı	56.33 ± 3.80	56.63 ± 4.56	56.01 ± 6.78	56.46 ± 3.92
Tasarım karar ağacı	63.75 ± 2.89	71.99 ± 4.62	55.47 ± 6.60	61.94 ± 2.92

Çizelge 5.4 sonuçları incelendiğinde AlexNet'ten elde edilip seçilen 152 adet öznitelikle gerçekleştirilen çalışmada en yüksek doğruluk 70.69 ± 2.61 ile Ensemble'dan elde edilmiştir. Pozitif olabilirlik oranı 2.56'dır, duyarlılığı da tasarım karar ağacı hariç diğer sınıflandırıcılardan yüksektir. PPV değeri de diğer sınıflandırıcılardan yüksektir. Yapılan bu deney yaklaşık 23 dk sürmüştür.

Çizelge 5.4 : AlexNet ile elde edilen ve seçilen özniteliklerin sınıflandırma sonuçları.

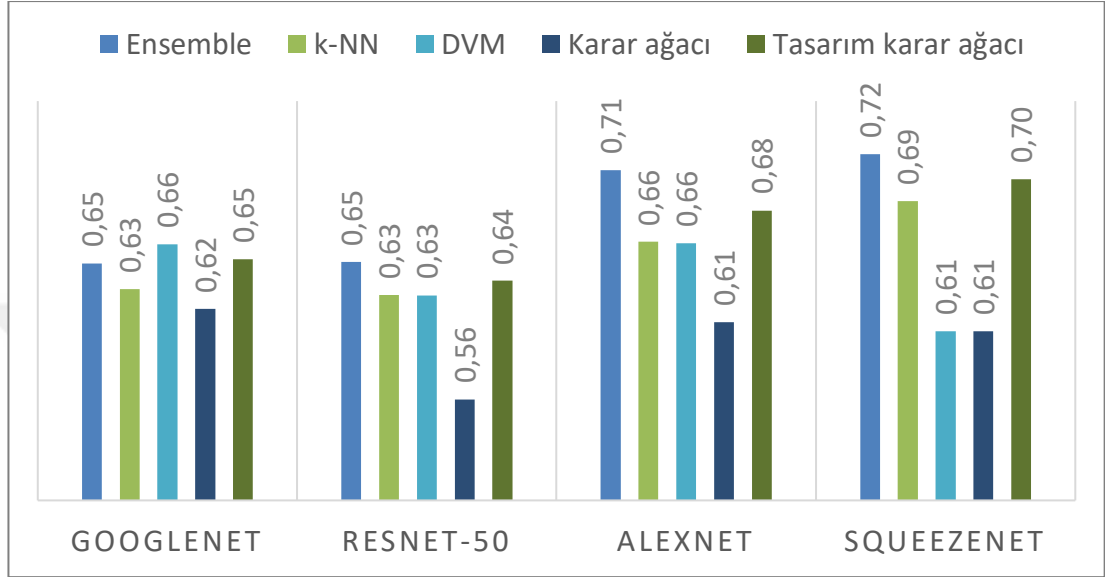
ALEXNET	Doğruluk (%)	Duyarlılık (%)	Özgüllük (%)	PPV (%)
Ensemble	70.69 ± 2.61	68.13 ± 4.23	73.40 ± 5.60	72.74 ± 4.71
k-NN	66.19 ± 2.39	57.63 ± 3.84	75.18 ± 4.35	70.71 ± 4.02
DVM	66.10 ± 2.90	62.53 ± 2.59	69.70 ± 6.37	68.41 ± 3.87
Karar ağacı	61.16 ± 4.08	59.61 ± 5.87	62.83 ± 5.59	62.45 ± 4.80
Tasarım karar ağacı	68.15 ± 2.74	70.63 ± 3.68	65.65 ± 6.43	68.22 ± 4.45

Çizelge 5.5 sonuçları incelendiğinde SqueezeNet'ten edilen ve seçilen 273 adet öznitelikle gerçekleştirilen çalışmada en yüksek doğruluk 71.68 ± 3.45 ile Ensemble'dan elde edilmiştir. Pozitif olabilirlik oranı 2.43'dür, duyarlılığı da diğerlerinden güçlüdür. PPV değeri de diğer sınıflandırıcılardan yüksektir. SqueezeNet ile yapılan bu deney yaklaşık 10 saat sürmüştür.

Çizelge 5.5 : SqueezeNet ile elde edilen ve seçilen özniteliklerin sınıflandırma sonuçları.

SQUEEZENET	Doğruluk (%)	Duyarlılık (%)	Özgüllük (%)	PPV (%)
Ensemble	71.68 ± 3.45	74.01 ± 4.70	69.56 ± 6.24	72.23 ± 5.11
k-NN	68.75 ± 2.99	64.83 ± 6.08	73.42 ± 6.32	72.17 ± 6.33
DVM	60.59 ± 2.93	60.21 ± 6.15	61.15 ± 5.19	62.14 ± 5.14
Karar ağacı	60.59 ± 5.04	59.12 ± 5.93	62.19 ± 6.42	62.45 ± 6.11
Tasarım karar ağacı	70.11 ± 4.25	78.45 ± 5.37	61.69 ± 7.58	68.59 ± 6.38

Ensemble algoritması GoogleNet dışında kullanılan tüm ESA mimarilerinde en yüksek doğruluk performansını göstermektedir. Birlikte kullanıldığında en başarılı sonuçlar verdiği mimari ise SqueezeNet'tir. Karar ağacı algoritması ise kullanılan tüm ESA mimarilerinde en düşük doğruluk performansını göstermektedir (Şekil 5.1).



Şekil 5.1 : ESA mimarileri ile birlikte kullanılan sınıflandırma algoritmalarının patolojik seslerin belirlenmesinde doğruluk performansları.

5.2 Değerlendirme ve Tartışma

Bu tez çalışmasında, derin öğrenme yöntemi kullanılarak patolojik seslerin belirlenmesi için farklı ESA mimarileri çalışılmıştır. Öncelikle kendi sınıflandırıcılarıyla çalıştırılan mimariler daha sonrasında farklı sınıflandırıcılar ile birlikte çalıştırılmıştır ve bu sayede iki yöntemin karşılaştırılması sağlanmıştır.

İkinci modelde seçilen ESA'ların farklı katmanlarından öznetelikler elde edilerek mimarilerin sınıflandırıcılarını kullanmak yerine karar ağacı, k-NN, DVM, ensemble ve çalışma özelinde tasarlanan bir karar ağacı ile sınıflandırma çalışması yapılmış ve bu sayede hem farklı ESA mimarilerinin başarısı hem de derin öznetelikleri sınıflandırmada klasik makine öğrenme yöntemlerinin başarısı araştırılmıştır.

Çalışmada kullanılan sesler SVD veri tabanından alınmıştır. 380'i sağlıklı ve 380'i patolojik olmak üzere 760 ses kaydı kullanılmıştır. Kullanılan patolojik sesler arasında larenjit, lökoplazi, Reinke ödemi, rekürren laringeal sinir felci, vokal kord

karsinomu ve vokal kord polibi bulunmaktadır. Bu çalışma için her bir bireyin nötr perdesinde sürekli sesli /a/ sesi kayıtları seçilmiştir. Veriler, sırasıyla %75 ve %25 örnek içeren eğitim seti ve test seti olarak ayrılmıştır.

Patolojik seslerin belirlenmesine yönelik çalışmalar yalnızca teknik açıdan değil, aynı zamanda hem analiz için seçilen ses kayıtları hem de sınıflandırma için patolojilerin alt kümelerinin nasıl kullanıldığı açısından farklılık göstermektedir. Literatürde günümüze kadar yapılmış çalışmalarda araştırmacıların en büyük handikapı veri azlığı sebebiyle homojen bir veri seti oluşturamamak olmuştur. Bu tez çalışmasında diğer çalışmalara göre oldukça homojen bir veri seti oluşturulmuştur. Çalışma kapsamında seçilen verilerin tümü vokal kordda yapısal değişikliğe sebep olarak organik disfoniye neden olan patolojilerdir, Wu ve arkadaşlarının [90] seçtikleri seslerin aksine bu tez çalışmasında ek patolojisi olan hiçbir ses çalışmaya dahil edilmemiştir. Bu durumda oldukça az bir veri seti ile çalışma tamamlanmış olsa da önerilen yöntem ile Wu ve arkadaşlarının [90] yaptıkları çalışmaya göre daha yüksek performans elde edilmiştir (Çizelge 5.6).

Yapılan tez çalışmasında sağlıklı ve patolojik seslerin sayısı açısından dengeli ve seçilen patolojiler bakımından homojen bir veri seti oluşturulmuştur. Ancak Alhussein ve Muhammad [94] ile Mohammed ve arkadaşlarının [110] yaptıkları çalışmalarda seçilen veri seti dengesiz ve oldukça heterojen olmasına rağmen yüksek performanslar elde edilmiştir. Bu çalışmalarda farklı ESA mimarileri ve farklı patolojiler kullanılmasına rağmen veri sayısının fazla olması nedeniyle bu sonuçların elde edilmesi derin öğrenmede beklenen bir durumdur.

Çalışmada kullanılan ses sinyalleri öncelikle literatürde de ses sinyallerinde gürültü gidermede sıkça kullanılan [112,107,108,89,109,38] dalgacık gürültü giderme yöntemine tabi tutulmuştur. Literatürde kullanılan parametreler ve başka bir parametre kombinasyonu ile yapılan deneyler ile en iyi sonuçların alındığı Daubechies ailesinden db2'nin 2. seviye ayrıştırması seçilmiştir. Daha sonrasında ESA mimarilerinin beslenmesi amacıyla tek boyutlu ses sinyallerin spektrogram görüntüleri elde edilmiştir, spektrogram görüntülerinin elde edilmesi için literatürde kullanılan KZFD parametrelerinden en iyi sonuç veren pencere uzunluğu 10 ms ve örtüşme değeri 5 ms olan Hamming penceresi kullanılmıştır [110,49,89,90]. Üretilen görüntüler hangi ağa girdi olarak verilecek ise o ağın giriş boyutuna göre yeniden boyutlandırılmıştır.

Tez çalışmasında ilk aşamada literatürde sıklıkla kullanılan ESA mimarilerinden en yüksek doğruluk 66.88 ± 3.75 ile AlexNet'te gözlemlenmiştir. AlexNet'te pozitif olabilirlik oranı 2.94 olarak hesaplanmıştır ve bu değer diğer ESA mimarilerinden yüksektir, ancak AlexNet diğer mimarilerle karşılaştırıldığında en düşük duyarlılık değerine sahiptir. PPV değeri ise diğer mimarilerden yüksektir.

İlk aşamada ResNet-50'den elde edilen doğruluk performansı 65.53 ± 3.05 olmuştur. Literatürde bir başka çalışmada ResNet-50'nin yine kendi sınıflandırıcısıyla kullanılmasından daha düşük performans göstermiştir (Huckvale & Buciuileac, 2021) (Çizelge 5.6). Bu durumun büyük ölçüde bu tez çalışmasında seçilen veri setinin bahsedilen çalışmadan farklı patolojiler içermesi ve düşük sayıda veri içermesi ile birlikte ResNet-50 ağına girdi olarak verilen spektrogram görüntülerinin üretilmesinde farklı parametrelerin kullanılmasından kaynaklandığı düşünülmektedir.

Bu aşamada kendi sınıflandırıcısıyla kullanılan ESA mimarilerinin patolojik seslerin belirlenmesinde yeterli başarıyı sağlayamadığı gözlemlenmiştir. Bu durumda ikinci aşama tasarlanmış ve deneyler gerçekleştirilmiştir.

İkinci aşamada öznitelik çıkarımında tüm ESA mimarileri için en başarılı sonuç veren katmanlar seçilmiştir. GoogleNet mimarisi 144 katmanlı bir yapıya sahiptir, en başarılı sonuçlar 140. katman olan Küresel Ortalama Havuzlama katmanından elde edilen öznitelikler ile sağlanmıştır. Bu katmandan her bir veri için 1024 öznitelik elde edilmiştir, daha sonrasında KBA uygulanarak her bir veri için 40 öznitelik seçilmiştir. GoogleNet ile oluşturulan bu modelde en yüksek doğruluk 66.04 ± 3.22 , özgüllük 74.73 ± 4.76 ve PPV 69.58 ± 4.43 DVM ile sınıflandırmadan elde edilmiştir. En yüksek duyarlılık ise 65.69 ± 4.45 olarak tasarım karar ağacında gözlemlenmiştir. GoogleNet mimarisinin kullanılan tüm sınıflandırıcılarda da yeterli başarıyı sağlayamadığı görülmüştür.

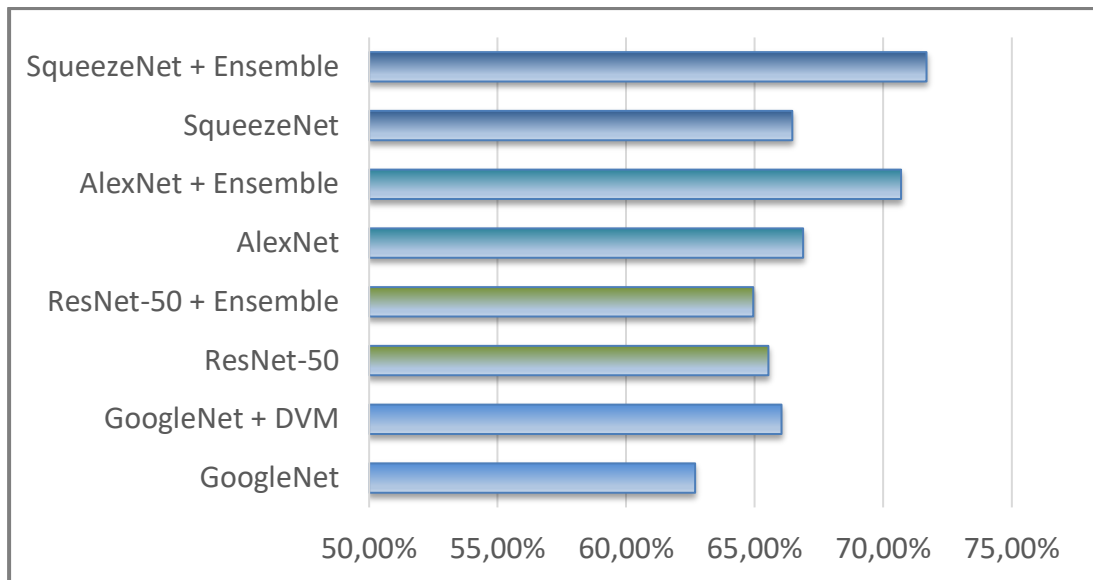
ResNet-50 mimarisi 177 katmanlı bir yapıya sahiptir, en başarılı sonuçlar 174. katman olan Küresel Ortalama Havuzlama katmanından elde edilen öznitelikler ile sağlanmıştır. Bu katmandan her bir veri için 2048 öznitelik elde edilmiştir, daha sonrasında KBA uygulanarak her bir veri için 174 öznitelik seçilmiştir. ResNet-50 ile oluşturulan bu modelde en yüksek doğruluk 66.04 ± 3.22 DVM ile sınıflandırmadan elde edilmiştir. DVM ile sınıflandırmada pozitif olabilirlik oranı 2.27 olarak hesaplanmasına rağmen duyarlılık değeri düşük görülmüştür. PPV değeri ise diğer

sınıflandırıcılardan yüksektir. ResNet-50 mimarisinin kullanılan tüm sınıflandırıcılarda da yeterli başarıyı sağlayamadığı görülmüştür.

AlexNet mimarisi 25 katmanlı bir yapıya sahiptir, en başarılı sonuçlar 20. katman olan ikinci Tam Bağlantılı katmanından elde edilen öznetelikler ile sağlanmıştır. Bu katmandan her bir veri için 4096 öznetelik elde edilmiştir, daha sonrasında KBA uygulanarak her bir veri için 152 öznetelik seçilmiştir. AlexNet ile oluşturulan bu modelde en yüksek doğruluk 70.69 ± 2.61 ile Ensemble algoritmasından elde edilmiştir. Pozitif olabilirlik oranı 2.56'dır, duyarlılığı da tasarım karar ağacı hariç diğer sınıflandırıcılardan yüksektir. PPV değeri de diğer sınıflandırıcılardan yüksektir.

SqueezeNet mimarisi 68 katmanlı bir yapıya sahiptir, en başarılı sonuçlar 63. katman olan Dropout katmandan elde edilen öznetelikler ile sağlanmıştır. Bu katmandan her bir veri için 100352 öznetelik elde edilmiştir, daha sonrasında KBA uygulanarak her bir veri için 273 öznetelik seçilmiştir. SqueezeNet ile oluşturulan bu modelde en yüksek doğruluk 71.68 ± 3.45 ile Ensemble algoritmasından elde edilmiştir. Pozitif olabilirlik oranı 2.43'dür ve duyarlılığı da diğerler sınıflandırma algoritmalarından güçlüdür. PPV değeri de diğer sınıflandırıcılardan yüksektir.

İlk aşamada kendi sınıflandırıcısıyla kullanılan ESA mimarileri ile ikinci aşamada farklı sınıflandırıcılarla kullanılan ESA mimarilerinden en yüksek sınıflandırıcı algoritmalarıyla elde edilen doğruluk performansları Şekil 5.2'de verilmiştir.



Şekil 5. 2 : İlk aşama ile ikinci aşamanın doğruluk performansları.

Şekil 5.2 incelendiğinde, ResNet-50 hariç tüm ESA mimarilerinde mimarinin kendi sınıflandırıcısı yerine farklı sınıflandırma algoritmalarının kullanılması ile daha yüksek performansın elde edilebildiği görülmektedir. Fakat farklı sınıflandırıcıların kullanılması çalışma süresini arttırmıştır, bunun sebebi sınıflandırıcılardan daha çok öznelik seçiminde kaybedilen zaman olmuştur.

En başarılı sonuçlar SqueezeNet ile Ensemble algoritmasının kullanıldığı modelden elde edilmiştir. Çalışmada araştırılan tüm ESA mimarileri literatürde en güçlü kabul edilen mimariler olmasına, bununla birlikte her mimari için optimum parametrelerin seçilmesine rağmen SqueezeNet'in en yüksek başarıyı elde etmesindeki temel nedenin katmanlardan elde edilen ve seçilip sınıflandırma algoritmalarına verilen öznelik sayısı olduğu gözlemlenmiştir. Bununla birlikte genel ortalamalara bakıldığında en başarılı sonuçların Ensemble algoritmasından elde edilmesi Ensemble algoritmasının birden fazla öğrencinin eğitilmesine dayanan bir makine öğrenmesi modeli olmasıdır. Literatürde Ensemble algoritmasında LogitBoost kullanmanın eğitim hatalarını doğrusal olarak azalttığı ve dolayısıyla daha iyi genelleme sağladığı gözlemlenmiştir [82].

Literatürdeki patolojik seslerin belirlenmesi amacıyla gerçekleştirilen benzer çalışmaların bulguları ve bu tez çalışmasında en yüksek performansı gösteren SqueezeNet ile Ensemble algoritmasının birleştirildiği modelden elde edilen sonuçların 10 kez tekrarının ortalaması ile karşılaştırması Çizelge 5.6'da verilmiştir.

Çizelge 5.6 : ESA ile patolojik seslerin belirlenmesi için benzer literatür karşılaştırması.

	Kullanılan Veri Seti	Yöntem	Doğruluk (%)	Duyarlılık (%)	Özgüllük (%)	PPV (%)
[90]	SVD, 482 sağlıklı, 482 patolojik, Organik disfoniler, Nötr perde sürekli /a/ sesi	ESA + CDBN	71.0	74.0	68.0	71.0
[93]	SVD, 869 sağlıklı, 597 patolojik, Organik ve organik olmayan disfoniler, Nötr perde sürekli /a/ sesi	Tüm disfoniler (ResNet-50)	69.27	-	-	-
		Organik disfoniler (ResNet-50)	70.87			
[94]	SVD, 686 sağlıklı, 1616 patolojik, Yaş > 15, Karışık patoloji seçimi, Nötr perde sürekli /a/ sesi	VGG16	93.50	94.80	92.40	-
		CaffeNet	93.90	94.80	92.80	-
[110]	SVD, 1304 patolojik, 830 sağlıklı, Karışık patoloji seçimi, Nötr perde sürekli /a/ sesi	ESA	96.11	95.38	95.97	95.41
Önerilen model	SVD, 380 sağlıklı, 380 patolojik, Organik disfoniler, Nötr perde sürekli /a/ sesi	SqueezeNet + Ensemble	71.68	74.01	69.56	72.23

6. SONUÇ VE ÖNERİLER

Patolojik ses bozukluklarının erken tespiti, tedavi süreci ve tedavinin başarısı için oldukça önemlidir, bu nedenle patolojik seslerin tespitinde bir karar destek sisteminin geliştirilmesi hayati önem taşımaktadır.

Son yıllarda patolojik seslerin belirlenmesinde makine öğrenme yöntemleri araştırmacılar tarafından ilgi çekici bir konu olmuştur ancak henüz patolojik ses bozukluklarının tespiti alanında derin öğrenme tekniklerinin kullanımı ile ilgili az sayıda araştırma çalışması yapılmıştır.

Bu tez çalışmasında, patolojik seslerin belirlenmesi amacıyla derin öğrenme yöntemleri kullanılmıştır. Çalışma kapsamında patolojik seslerin bulunduğu bir veri tabanından vokal kordlardaki yapısal değişikliklerin neden olduğu organik disfoniye sebep olan patolojilere sahip hastaların ses kayıtları seçilmiştir. Ses kayıtlarına öncelikle dalgacık tabanlı bir gürültü giderme işlemi uygulanmış ve daha sonrasında spektrogram görüntüleri alınarak literatürde en güçlü olarak belirlenmiş dört farklı Evrişimsel Sinir Ağı mimarisine girdi olarak verilmiştir.

İlk aşamada patolojik seslerin belirlenmesi amacıyla seçilen ESA mimarileri kendi sınıflandırıcılarıyla birlikte kullanılmıştır. Daha sonra aynı ESA mimarileri bu kez öznelik çıkarımından sonra farklı sınıflandırma algoritmalarıyla birlikte kullanılarak oluşturulan modellerin performans analizleri yapılmıştır. En başarılı performans SqueezeNet mimarisi ile Ensemble algoritmasının birlikte kullanılması durumunda gözlemlenmiştir.

Tez çalışmasında en başarılı bulunan model, literatürdeki benzer çalışmalar ile karşılaştırılmış ve diğer çalışmalardan daha homojen bir veri seti kullanılmasına rağmen yine organik disfonilerin çalışıldığı benzer çalışmalara göre daha yüksek performans elde edilmiştir.

SqueezeNet ile Ensemble algoritmasının birleştirildiği modelin umut verici sonuçları ilerleyen çalışmalarda farklı ses veri tabanları ile genişletilerek, kullanılan yöntemlerde farklı parametreler denenerek ve farklı sınıflandırma algoritmaları çalışılarak geliştirilmeye değer bulunmuştur.

Patolojik seslerin belirlenmesi amacıyla bu çalışmada önerilen modelin geliştirilerek aynı zamanda ilerleyen çalışmalarda sinonazal lezyonlar, triodektomi, entübasyon ve gastroösofagiye reflünün ses üzerine etkilerinden dolayı bu hastalıkların belirlenmesinde de çalışılması planlanmaktadır.



KAYNAKLAR

- [1] **Shahnaz, C., Fattah, S. A., Mahbub, U., Zhu, W.-P., and Ahmad, M. O.,** (2012). Detection of voice disorders based on wavelet and prosody-related properties, *2012 IEEE International Symposium on Circuits and Systems (ISCAS)*, Seoul, Korea (South), 20-23 May.
- [2] **Black, L. I., Vahratian, A., and Hoffman, H. J.,** (2015). Communication disorders and use of intervention services among children aged 3-17 years: United states, 2012. Nchs data brief, *Centers for Disease Control and Prevention*, 205, 8.
- [3] **Bhattacharyya, N.,** (2014). The prevalence of voice problems among adults in the united states, *The Laryngoscope*, 124, 10, 2359-2362.
- [4] **Hossain, M. S., Muhammad, G., and Alamri, A.,** (2019). Smart healthcare monitoring: A voice pathology detection paradigm for smart cities, *Multimedia Systems*, 25, 5, 565-575.
- [5] **Hegde, S., Shetty, S., Rai, S., and Dodderi, T.,** (2019). A survey on machine learning approaches for automatic detection of voice disorders, *Journal of Voice*, 33, 6, 947. e911-947. e933.
- [6] **Rufo, M. J., Martín, J., Pérez, C. J., and Paniagua, S.,** (2019). A bayesian decision analysis approach to assess voice disorder risks by using acoustic features, *Biometrical Journal*, 61, 3, 503-513.
- [7] **Gerçeker, M., Yorulmaz, İ., and Ural, A.,** (2000). Ses ve konuşma, *K.B.B. ve Baş Boyun Cerrahisi Dergisi*, 8, 1, 71-78.
- [8] **Erkan, M., Aslan, T.,** (1992). Konuşma ve ses bozuklukları, *Erciyes Tıp Dergisi*, 1, 297-313.
- [9] **Burad, E., Çağlar, Y.,** (1972). Sesiniz, *DSİ Matbaası*, 1, 6, 28-32.
- [10] **Negus, V.,** (1957). The mechanism of the larynx, (in eng), *Laryngoscope*, 67, 10, 961-986.
- [11] **Öğüt, F.,** (1999). Ses fizyolojisi, temel prensipler, sesin aerodinamik ve akustik analizi, *1. Uluslararası Ses Hastalıkları Sempozyumu*, İstanbul, 29–30 Nisan.
- [12] **Kılıç, M.,** (2002). Larenksin fonksiyonel anatomisi ve ses fizyolojisi, *Türkiye Klinikleri KBB Dergisi (Ses Bozuklukları Özel Sayısı)*, 2, 1-8.
- [13] **Erdinç, M.,** (2009). *Vokal kord poliplerinde cerrahi sonrası ses kalitesinin araştırılması*, (tıpta uzmanlık tezi), Adres: <https://www.ogu.edu.tr/>.
- [14] **Mohebatı, A., Shaha, A. R.,** (2012). Anatomy of thyroid and parathyroid glands and neurovascular relations, (in eng), *Clin Anat*, 25, 1, 19-31.
- [15] **Sakorafas, G. H., Kokoropoulos, P., Lappas, C., Sampanis, D., and Smyrniotis, V.,** (2012). External branch of the superior laryngeal nerve: Applied surgical anatomy and implications in thyroid surgery, (in eng), *Am Surg*, 78, 9, 986-991.
- [16] **Arıncı, K., Elhan, A.,** *Anatomi*, Vol. 1, Sf: 366-375, Ankara, Güneş Kitabevi, (1999).
- [17] **Sadler, T. W. L. J.,** *Langman's medical embryology*, Philadelphia, Lippincott William & Wilkins (in English), (2010).

- [18] **Ardito, G., Revelli, L., D'Alatri, L., Lerro, V., Guidi, M. L., and Ardito, F.,** (2004). Revisited anatomy of the recurrent laryngeal nerves, (in eng), *Am J Surg*, 187, 2, 249-253.
- [19] **Myssiorek, D.,** (2004). Recurrent laryngeal nerve paralysis: Anatomy and etiology, (in eng), *Otolaryngol Clin North Am*, 37, 1, 25-44.
- [20] **Sanders, G., Uyeda, R. Y., and Karlan, M. S.,** (1983). Nonrecurrent inferior laryngeal nerves and their association with a recurrent branch, (in eng), *Am J Surg*, 146, 4, 501-503.
- [21] **Moran, R. E., Castro, A. F.,** (1951). The superior laryngeal nerve in thyroid surgery, (in eng), *Ann Surg*, 134, 6, 1018-1021.
- [22] **Sulica, L.,** (2004). The superior laryngeal nerve: Function and dysfunction, (in eng), *Otolaryngol Clin North Am*, 37, 1, 183-201.
- [23] **Sanders, I., Mu, L.,** (1998). Anatomy of the human internal superior laryngeal nerve, (in eng), *Anat Rec*, 252, 4, 646-656.
- [24] **Wu, B. L., Sanders, I., Mu, L., and Biller, H. F.,** (1994). The human communicating nerve. An extension of the external superior laryngeal nerve that innervates the vocal cord, (in eng), *Arch Otolaryngol Head Neck Surg*, 120, 12, 1321-1328.
- [25] **Hirano, M.,** (1974). Morphological structure of the vocal cord as a vibrator and its variations, (in eng), *Folia Phoniatr (Basel)*, 26, 2, 89-94.
- [26] **Hirano, M.,** *Clinical examination of voice*, Wien; New York, Springer-Verlag (in English), (1981).
- [27] **Deary, I. J., Webb, A., Mackenzie, K., Wilson, J. A., and Carding, P. N.,** (2004). Short, self-report voice symptom scales: Psychometric characteristics of the voice handicap index-10 and the vocal performance questionnaire, (in eng), *Otolaryngol Head Neck Surg*, 131, 3, 232-235.
- [28] **Jacobson, B. H., Johnson, A., Grywalski, C., Silbergleit, A., Jacobson, G., Benninger, M. S., and Newman, C. W.,** (1997). The voice handicap index (vhi), *American Journal of Speech-Language Pathology*, 6, 3, 66-70.
- [29] **Kılıç, M.,** (2010). Ses problemi olan hastanın objektif ve subjektif yöntemlerle değerlendirilmesi, *Kulak Burun Boğaz Baş Boyun Cerrahisi'nde Güncel Yaklaşım*, 6, 257-265.
- [30] **Heikkinen, M., Penttilä, E., Qvarnström, M., Mäkinen, K., Löppönen, H., and Kärkkäinen, J. M.,** (2021). Perceptual assessment and acoustic voice analysis as screening tests for vocal fold paresis after thyroid or parathyroid surgery, (in eng), *World J Surg*, 45, 3, 765-773.
- [31] **Nagle, K. F.,** (2016). Emerging scientist: Challenges to cape-v as a standard, *Perspectives of the ASHA Special Interest Groups*, 1, 3, 47-53.
- [32] **Kempster, G. B., Gerratt, B. R., Verdolini Abbott, K., Barkmeier-Kraemer, J., and Hillman, R. E.,** (2009). Consensus auditory-perceptual evaluation of voice: Development of a standardized clinical protocol, (in eng), *Am J Speech Lang Pathol*, 18, 2, 124-132.
- [33] **Öğüt, F.,** (2002). Objektif ses analizi, *Türkiye Klinikleri Journal of ENT*, 2, 2, 121-126.
- [34] **Bless, D. M., Hirano, M., and Feder, R. J.,** (1987). Videostroboscopic evaluation of the larynx, (in eng), *Ear Nose Throat J*, 66, 7, 289-296.

- [35] **Min, Y. B., Finnegan, E. M., Hoffman, H. T., Luschei, E. S., and McCulloch, T. M.**, (1994). A preliminary study of the prognostic role of electromyography in laryngeal paralysis, (in eng), *Otolaryngol Head Neck Surg*, 111, 6, 770-775.
- [36] **Lynch, J., Parameswaran, R.**, (2017). Management of unilateral recurrent laryngeal nerve injury after thyroid surgery: A review, (in eng), *Head Neck*, 39, 7, 1470-1478.
- [37] **Zhenxing, L., Hongzhou, X.**, (2009). A wavelet threshold de-noising algorithm based on empirical mode decomposition, *Computer Simulation*, 26, 9, 192-325.
- [38] **Zhong, X., Dai, Y., Dai, Y., and Jin, T.**, (2018). Study on processing of wavelet speech denoising in speech recognition system, *International Journal of Speech Technology*, 21, 3, 563-569.
- [39] **Badshah, A. M., Ahmad, J., Rahim, N., and Baik, S. W.**, (2017). Speech emotion recognition from spectrograms with deep convolutional neural network, *2017 international conference on platform technology and service (PlatCon)*, Busan, Korea (South), 13-15 February.
- [40] **Deng, L., Seltzer, M. L., Yu, D., Acero, A., Mohamed, A.-r., and Hinton, G.**, (2010). Binary coding of speech spectrograms using a deep auto-encoder, *Eleventh Annual Conference of the International Speech Communication Association*, Makuhari, Chiba, Japan, 26-30 September
- [41] **Top, A. E.**, (2018). *Classification of eeg signals using transfer learning on convolutional neural networks via spectrogram*, (yüksek lisans tezi), Adres: <https://aybu.edu.tr/>.
- [42] **Rosenblatt, F.**, (1958). The perceptron: A probabilistic model for information storage and organization in the brain, *Psychological review*, 65, 6, 386.
- [43] **Rumelhart, D. E., Hinton, G. E., and Williams, R. J.**, (1986). Learning representations by back-propagating errors, *nature*, 323, 6088, 533-536.
- [44] **Bengio, Y., Lamblin, P., Popovici, D., and Larochelle, H.**, (2006). Greedy layer-wise training of deep networks, *Advances in neural information processing systems*, 19.
- [45] **Hinton, G. E., Salakhutdinov, R. R.**, (2006). Reducing the dimensionality of data with neural networks, *science*, 313, 5786, 504-507.
- [46] **Ranzato, M. A., Poultney, C., Chopra, S., and Cun, Y.**, (2006). Efficient learning of sparse representations with an energy-based model, *Advances in neural information processing systems*, 19.
- [47] **Chen, Y., Li, L., Xiao, J., Yang, Y., Liang, J., and Li, T.**, (2018). Particle swarm optimizer with crossover operation, *Engineering Applications of Artificial Intelligence*, 70, 159-169.
- [48] **De Silva, D., Fernando, S., Piyatilake, I. T. S., and Karunaratne, A.**, (2019). Wavelet based edge feature enhancement for convolutional neural networks, *Eleventh International Conference on Machine Vision (ICMV 2018)*, Munich, Germany, 1-3 November.
- [49] **Nasıfoğlu, H.**, (2022). *Obstrüktif uyku apnesinin derin öğrenme kullanılarak tahmin edilmesi*, (doktora tezi), Adres: <http://etu.edu.tr/>.

- [50] **Albawi, S., Mohammed, T. A., and Al-Zawi, S.,** (2017). Understanding of a convolutional neural network, *2017 International Conference on Engineering and Technology (ICET)*, 21-23 Aug. 2017.
- [51] **Yamashita, R., Nishio, M., Do, R. K. G., and Togashi, K.,** (2018). Convolutional neural networks: An overview and application in radiology, *Insights into Imaging*, 9, 4, 611-629.
- [52] **Özkan, İ., Ülker, E.,** (2017). Derin öğrenme ve görüntü analizinde kullanılan derin öğrenme modelleri, *Gaziosmanpaşa Bilimsel Araştırma Dergisi*, 6, 3, 85-104.
- [53] **Hijazi, S., Kumar, R., and Rowen, C.,** (2015). Using convolutional neural networks for image recognition, *Cadence Design Systems Inc.: San Jose, CA, USA*, 9.
- [54] **Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., Santamaria, J., Fadhel, M. A., Al-Amidie, M., and Farhan, L.,** (2021). Review of deep learning: Concepts, cnn architectures, challenges, applications, future directions, *Journal of big Data*, 8, 1, 1-74.
- [55] **Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R.,** (2014). Dropout: A simple way to prevent neural networks from overfitting, *The journal of machine learning research*, 15, 1, 1929-1958.
- [56] **Shen, L., Zhang, Q., Cao, G., and Xu, H.,** (2018). Fall detection system based on deep learning and image processing in cloud environment, *Conference on Complex, Intelligent, and Software Intensive Systems*, Kunibiki Messe, Matsue, Japan, 4-6 July.
- [57] **Szegedy, C., Wei, L., Yangqing, J., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., and Rabinovich, A.,** (2015). Going deeper with convolutions, *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 7-12 June 2015.
- [58] **He, K., Zhang, X., Ren, S., and Sun, J.,** (2016). Deep residual learning for image recognition, *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 27-30 June 2016.
- [59] **Mascarenhas, S., Agarwal, M.,** (2021). A comparison between vgg16, vgg19 and resnet50 architecture frameworks for image classification, *2021 International Conference on Disruptive Technologies for Multi-Disciplinary Research and Applications (CENTCON)*, 19-21 Nov. 2021.
- [60] **Nasifoglu, H., Eroglu, O.,** (2021). Obstructive sleep apnea prediction from electrocardiogram scalograms and spectrograms using convolutional neural networks, *Physiological Measurement*, 42, 6, 065010.
- [61] **Ali, L., Alnajjar, F., Jassmi, H., Gochoo, M., Khan, W., and Serhani, M.,** (2021). Performance evaluation of deep cnn-based crack detection and localization techniques for concrete structures, *Sensors*, 21, 1688.
- [62] **Iandola, F. N., Han, S., Moskewicz, M. W., Ashraf, K., Dally, W. J., and Keutzer, K.,** (2016). Squeezenet: Alexnet-level accuracy with 50x fewer parameters and < 0.5 mb model size, *arXiv preprint arXiv:1602.07360*.
- [63] **Hassanpour, M., Malek, H.,** (2020). Learning document image features with squeezenet convolutional neural network, *International Journal of Engineering*, 33, 7, 1201-1207.

- [64] **Krizhevsky, A., Sutskever, I., and Hinton, G. E.,** (2012). Imagenet classification with deep convolutional neural networks, *Advances in neural information processing systems*, 25.
- [65] **Han, X., Zhong, Y., Cao, L., and Zhang, L.,** (2017). Pre-trained alexnet architecture with pyramid pooling and supervision for high spatial resolution remote sensing image scene classification, *Remote Sensing*, 9, 8, 848.
- [66] **Raghu, S., Sriraam, N.,** (2018). Classification of focal and non-focal eeg signals using neighborhood component analysis and machine learning algorithms, *Expert Systems with Applications*, 113, 18-32.
- [67] **Bastami, Z., Sheikhpour, R., Razzaghi, P., Ramazani, A., and Gharaghani, S.,** (2022). Proteochemometrics modeling for prediction of the interactions between caspase isoforms and their inhibitors, *Molecular Diversity*, 1-13.
- [68] **Yang, W., Wang, K., and Zuo, W.,** (2012). Neighborhood component feature selection for high-dimensional data, *J. Comput.*, 7, 1, 161-168.
- [69] **Das, R.,** (2010). A comparison of multiple classification methods for diagnosis of parkinson disease, *Expert Systems with Applications*, 37, 2, 1568-1572.
- [70] **Kantardzic, M.,** *Data mining: Concepts, models, methods, and algorithms*, John Wiley & Sons,(2011).
- [71] **Cortes, C., Vapnik, V.,** (1995). Support-vector networks, *Machine Learning*, 20, 3, 273-297.
- [72] **Azuaje, F., Witten, I., and E, F.,** (2006). Witten ih, frank e: Data mining: Practical machine learning tools and techniques, *Biomedical Engineering Online - BIOMED ENG ONLINE*, 5, 1-2.
- [73] **Xu, G., Zong, Y., and Yang, Z.,** *Applied data mining*, CRC Press,(2013).
- [74] Nixon, M. S., Aguado, A. S., "8 - introduction to texture description, segmentation and classification," in *Feature extraction and image processing*, M. S. Nixon, A. S. Aguado Eds. Oxford: Newnes, 2002, pp. 291-310.
- [75] **Witten, I. H., Frank, E., Hall, M. A., Pal, C. J., and DATA, M.,** *Practical machine learning tools and techniques*, Vol. 2,(2005).
- [76] **Stan, L., Li, Z., and Jain, A. K.,** "Encyclopedia of biometrics," in *Iz*, vol. 1: Springer Science & Business Media, 2009.
- [77] **Kalabalık, G.,** (2016). *A comparison of the performance of ensemble classification methods in telecom customer churn analysis* (yüksek lisans tezi), Adres: <https://www.yasar.edu.tr/>.
- [78] **Freund, Y., Schapire, R. E.,** (1997). A decision-theoretic generalization of on-line learning and an application to boosting, *Journal of computer and system sciences*, 55, 1, 119-139.
- [79] **Breiman, L.,** (1998). Rejoinder: Arcing classifiers, *The Annals of Statistics*, 26, 3, 841-849.
- [80] **Cortes, C., Drucker, H.,** (1996). Boosting decision trees, *Advances in neural information processing systems*, 8, 479-485.
- [81] **Rätsch, G., Onoda, T., and Müller, K.-R.,** (2001). Soft margins for adaboost, *Machine Learning*, 42, 3, 287-320.
- [82] **Friedman, J., Hastie, T., and Tibshirani, R.,** (2000). Additive logistic regression: A statistical view of boosting (with discussion and a rejoinder by the authors), *The Annals of Statistics*, 28, 2, 337-407.

- [83] **Cai, Y.-D., Feng, K.-Y., Lu, W.-C., and Chou, K.-C.**, (2006). Using logitboost classifier to predict protein structural classes, *Journal of theoretical biology*, 238, 172-176.
- [84] **Muhammad, G., Alsulaiman, M., Ali, Z., Mesallam, T. A., Farahat, M., Malki, K. H., Al-Nasheri, A., and Bencherif, M. A.**, (2017). Voice pathology detection using interlaced derivative pattern on glottal source excitation, *Biomedical Signal Processing and Control*, 31, 156-164.
- [85] **Muhammad, G., Altuwaijri, G., Alsulaiman, M., Ali, Z., Mesallam, T. A., Farahat, M., Malki, K. H., and Al-Nasheri, A.**, (2016). Automatic voice pathology detection and classification using vocal tract area irregularity, *Biocybernetics and Biomedical Engineering*, 36, 2, 309-317.
- [86] **Hammami, I., Salhi, L., and Labidi, S.**, (2016). Pathological voices detection using support vector machine, *2016 2nd International Conference on Advanced Technologies for Signal and Image Processing (ATSIP)*, Monastir, Tunisia, 21-23 March.
- [87] **Al-Nasheri, A., Muhammad, G., Alsulaiman, M., Ali, Z., Mesallam, T. A., Farahat, M., Malki, K. H., and Bencherif, M. A.**, (2017). An investigation of multidimensional voice program parameters in three different databases for voice pathology detection and classification, *Journal of Voice*, 31, 1, 113. e119-113. e118.
- [88] **Verde, L., Brancati, N., De Pietro, G., Frucci, M., and Sannino, G.**, (2021). A deep learning approach for voice disorder detection for smart connected living environments, *ACM Transactions on Internet Technology (TOIT)*, 22, 1, 1-16.
- [89] **Wu, H., Soraghan, J., Lowit, A., and Di Caterina, G.**, (2018). Convolutional neural networks for pathological voice detection, *2018 40th annual international conference of the IEEE engineering in medicine and biology society (EMBC)*, Honolulu, HI, USA, 18-21 July.
- [90] **Wu, H., Soraghan, J., Lowit, A., and Di Caterina, G.**, (2018). A deep learning method for pathological voice detection using convolutional deep belief networks, *Interspeech 2018*.
- [91] **Akbari, A., Arjmandi, M. K.**, (2014). An efficient voice pathology classification scheme based on applying multi-layer linear discriminant analysis to wavelet packet-based features, *Biomedical Signal Processing and Control*, 10, 209-223.
- [92] **Rafael Orozco Arroyave, J., Francisco Vargas Bonilla, J., and Delgado Trejos, E.**, (2012). Acoustic analysis and non linear dynamics applied to voice pathology detection: A review, *Recent Patents on Signal Processing (Discontinued)*, 2, 2, 96-107.
- [93] **Huckvale, M., Buciuileac, C.**, (2021). Automated detection of voice disorder in the saarbrücken voice database: Effects of pathology subset and audio materials, *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, Brno, Czechia, 30 August - 3 September.
- [94] **Alhussein, M., Muhammad, G.**, (2018). Voice pathology detection using deep learning on mobile healthcare framework, *IEEE Access*, 6, 41034-41041.

- [95] **Dworkin, J. P.**, (2008). Laryngitis: Types, causes, and treatments, (in eng), *Otolaryngol Clin North Am*, 41, 2, 419-436, ix.
- [96] **Murray, C. S., Simpson, A., and Custovic, A.**, (2004). Allergens, viruses, and asthma exacerbations, (in eng), *Proc Am Thorac Soc*, 1, 2, 99-104.
- [97] **Hanson, D. G., Jiang, J. J.**, (2000). Diagnosis and management of chronic laryngitis associated with reflux, (in eng), *Am J Med*, 108 Suppl 4a, 112s-119s.
- [98] **Jatzko, G. R., Lisborg, P. H., Müller, M. G., and Wette, V. M.**, (1994). Recurrent nerve palsy after thyroid operations--principal nerve identification and a literature review, (in eng), *Surgery*, 115, 2, 139-144.
- [99] **Kasemsuwan, L., Nubthuenetr, S.**, (1997). Recurrent laryngeal nerve paralysis: A complication of thyroidectomy, (in eng), *J Otolaryngol*, 26, 6, 365-367.
- [100] **Wagner, H. E., Seiler, C.**, (1994). Recurrent laryngeal nerve palsy after thyroid gland surgery, (in eng), *Br J Surg*, 81, 2, 226-228.
- [101] **Abboud, B., Tabchy, B., Jambart, S., Hamad, W. A., and Farah, P.**, (1999). Benign disease of the thyroid gland and vocal fold paralysis, (in eng), *J Laryngol Otol*, 113, 5, 473-474.
- [102] **Hong, K. H., Kim, Y. K.**, (1997). Phonatory characteristics of patients undergoing thyroidectomy without laryngeal nerve injury, (in eng), *Otolaryngol Head Neck Surg*, 117, 4, 399-404.
- [103] **Menvielle, G., Luce, D., Goldberg, P., Bugel, I., and Leclerc, A.**, (2004). Smoking, alcohol drinking and cancer risk for various sites of the larynx and hypopharynx. A case-control study in france, (in eng), *Eur J Cancer Prev*, 13, 3, 165-172.
- [104] **Bozdayi, G., Kemaloglu, Y., Ekinci, O., Dogan, B., Ilhan, M. N., Aydil, U., Akyol, G., Koybasioglu, A., Inal, E., and Rota, S.**, (2009). Role of human papillomavirus in the clinical and histopathologic features of laryngeal and hypopharyngeal cancers, (in eng), *J Otolaryngol Head Neck Surg*, 38, 1, 119-125.
- [105] **Chu, E. A., Kim, Y. J.**, (2008). Laryngeal cancer: Diagnosis and preoperative work-up, (in eng), *Otolaryngol Clin North Am*, 41, 4, 673-695, v.
- [106] **Schultz, P.**, (2011). Vocal fold cancer, (in eng), *Eur Ann Otorhinolaryngol Head Neck Dis*, 128, 6, 301-308.
- [107] **Dautov, Ç. P., Özerdem, M. S.**, (2018). Wavelet transform and signal denoising using wavelet method, *2018 26th Signal Processing and Communications Applications Conference (SIU)*, Izmir, Turkey, 02-05 May.
- [108] **Mihov, S. G., Ivanov, R. M., and Popov, A. N.**, (2009). Denoising speech signals by wavelet transform, *Annual Journal Of Electronics*, 6, 2-5.
- [109] **Zhang, Y., Jiang, J., and Feroze, F.**, (2005). Wavelet-based denoising for improving nonlinear dynamic analysis of pathological voices, *Electronics Letters*, 41, 16, 930-931.
- [110] **Mohammed, M. A., Abdulkareem, K. H., Mostafa, S. A., Khanapi Abd Ghani, M., Maashi, M. S., Garcia-Zapirain, B., Oleagordia, I., Alhakami, H., and Al-Dhief, F. T.**, (2020). Voice pathology detection and classification using convolutional neural network model, *Applied Sciences*, 10, 11, 3723.

- [111] **Duman, F., Erdamar, A., Eroglu, O., Telatar, Z., and Yetkin, S.**, (2009). Efficient sleep spindle detection algorithm with decision tree, *Expert Systems with Applications*, 36, 6, 9980-9985.
- [112] **Brahim, M.**, (2021). Denoising and enhancement speech signal using wavelet, *Journal of Information Systems and Telecommunication*, 9, 33, 37-44.

Url-1 <https://operativeneurosurgery.com/lib/exe/detail.php?id=heschl_s_gyrus&media=heschls_gyrus.png>, alındığı tarih: 02.07.2022.

Url-2 <<https://sborl.es/wp-content/uploads/2018/01/Larynx-anatomy.pdf>>, alındığı tarih: 02.07.2022.

Url-3 <<https://www.lecturio.com/concepts/larynx/>>, alındığı tarih: 02.07.2022.

Url-4 <<https://www.clevelandvoiceandsleep.com/anatomy--physiology.html>>, alındığı tarih: 02.07.2022.

Url-5 <<http://stimmdb.coli.uni-saarland.de/>>, alındığı tarih: 05.02.2022.