

**T.C.
FIRAT ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**



**DENETİMLİ ÖĞRENME YÖNTEMLERİ KULLANILARAK BİR
SES OLAY TESPİT SİSTEMİNİN GERÇEKLEŞTİRİLMESİ**

Hakan GÜLER

Yüksek Lisans Tezi

YAZILIM MÜHENDİSLİĞİ ANABİLİM DALI

TEMMUZ 2021

T.C.
FIRAT ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

Yazılım Mühendisliği Anabilim Dalı

Yüksek Lisans Tezi

DENETİMLİ ÖĞRENME YÖNTEMLERİ KULLANILARAK BİR SES OLAY TESPİT SİSTEMİNİN GERÇEKLEŞTİRİLMESİ

Tez Yazarı
Hakan GÜLER

Danışman
Dr. Öğr. Üyesi Mustafa ULAŞ

TEMMUZ 2021
ELAZIĞ

T.C.
FIRAT ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

Yazılım Mühendisliği Anabilim Dalı

Yüksek Lisans Tezi

Başlığı:	Denetimli Öğrenme Yöntemleri Kullanılarak Bir Ses Olay Tespit Sisteminin Gerçekleştirilmesi
Yazarı:	Hakan GÜLER
İlk Teslim Tarihi:	15.06.2021
Savunma Tarihi:	26.07.2021

TEZ ONAYI

Fırat Üniversitesi Fen Bilimleri Enstitüsü tez yazım kurallarına göre hazırlanan bu tez aşağıda imzaları bulunan jüri üyeleri tarafından değerlendirilmiş ve akademik dinleyicilere açık yapılan savunma sonucunda OYBİRLİĞİ ile kabul edilmiştir.

Danışman:	Dr. Öğr. Üyesi Mustafa ULAŞ Fırat Üniversitesi, Mühendislik Fakültesi	<i>İmza</i> Onayladım
Başkan:	Doç. Dr. Murat KARABATAK Fırat Üniversitesi, Teknoloji Fakültesi	Onayladım
Üye:	Doç. Dr. Aytuğ BOYACI Milli Savunma Üniversitesi, Mühendislik Fakültesi	Onayladım

Bu tez, Enstitü Yönetim Kurulunun/...../20..... tarihli toplantısında tescillenmiştir.

İmza
Doç. Dr. Kürşat Esat ALYAMAÇ
Enstitü Müdürü

BEYAN

Fırat Üniversitesi Fen Bilimleri Enstitüsü tez yazım kurallarına uygun olarak hazırladığım “Denetimli Öğrenme Yöntemleri Kullanılarak Bir Ses Olay Tespit Sisteminin Gerçekleştirilmesi” Başlıklı Yüksek Lisans Tezimin içindeki bütün bilgilerin doğru olduğunu, bilgilerin üretilmesi ve sunulmasında bilimsel etik kurallarına uygun davrandığımı, kullandığım bütün kaynakları atıf yaparak belirttiğimi, maddi ve manevi desteği olan tüm kurum/kuruluş ve kişileri belirttiğimi, burada sunduğum veri ve bilgileri unvan almak amacıyla daha önce hiçbir şekilde kullanmadığımı beyan ederim.

26.07.2021

Hakan GÜLER



ÖNSÖZ

Bu seminer çalışması, Fırat Üniversitesi Fen Bilimleri Enstitüsü Yazılım Mühendisliği Anabilim Dalı Tezli Yüksek Lisans programı için hazırlanmıştır.

Çalışmalarım boyunca beni yönlendiren, bana araştırma olanağı sunan, bilgi ve geniş hoşgörüsü ile benden hiçbir desteğini esirgemeyen değerli hocam Dr. Öğr. Üyesi Mustafa ULAŞ'a, Dr. Öğr. Üyesi Aytuğ BOYACI ve Öğretim Görevlisi Yunus KORKMAZ hocalarıma teşekkürlerimi borç bilirim.

Hakan GÜLER

ELAZIĞ, 2021



İÇİNDEKİLER

	Sayfa
ÖNSÖZ.....	iv
İÇİNDEKİLER	v
ÖZET	vii
ABSTRACT	viii
ŞEKİLLER LİSTESİ	ix
TABLolar LİSTESİ	x
EKLER LİSTESİ	xi
KISALTMALAR	xii
1. GİRİŞ	1
2. SES VE KONUŞMA	4
2.1. Ses	4
2.2. Akustik	5
2.2.1. Akustik Sahne Sınıflandırma Geçmişi	5
2.3. İşitsel Uyarılar	6
2.4. Konuşma Süreci ve Tanımlanması	7
2.4.1. İnsanda Konuşmada Üretim	7
2.4.2. Konuşma Üretimi	7
2.4.3. Konuşma Algısı.....	8
Şekil 2.3'te de görüldüğü gibi Konuşmada Algı Sürecinin Şematik Diyagramı konuşmacı ve dinleyici merkezli olarak vücut organlarından beyinde sinir aktivitelerinin dil koduna dönüşümü gerçekleşir.....	9
2.4.4. Sesli ve Sessiz Konuşma	9
2.4.5. İnsan Konuşma Modeli	9
3. ÖNİŞLEMLER.....	11
3.1. Ön Vurgu Filtresi.....	11
3.2. Çerçeveleme	11
3.3. Pencereleme.....	12
3.4. Sıfır Geçiş Oranı	12
3.5. Enerji	12
3.6. Hızlı Fourier Dönüşümü	13
3.7. Mel Ölçeği Filtre Bankası.....	13
3.8. Öznitelik Çıkarma Yöntemleri	13
4. DENETİMLİ ÖĞRENME YÖNTEMLERİ	14
4.1. Denetimli Öğrenme Algoritmaları.....	14
4.1.1. K-En Yakın Komşular Yöntemi (k-NN)	14
4.1.2. Naïve Bayes (NB) Sınıflandırıcı	15
4.1.3. Karar Ağacı ve Konuşma Sınıflandırılması.....	16
4.1.4. Destek Vektör Makinesi (SVM).....	17
4.1.5. Random Forest Algoritması	18
4.1.6. SMO	20
4.1.7. J48 Algoritması	21
4.1.8. IB1 Algoritması.....	21

4.1.9. Çok Katmanlı Yapay Sinir Ağları	21
4.2. Konuşma Sınıflandırmasında Sinir Ağları.....	22
4.3. Konuşmada Derin Sinir Ağları Tanıma ve Sınıflandırma	23
4.4. Denetimli Öğrenme Yöntemleri Kullanan Ses Çalışmalarının İncelenmesi.....	23
5. MATERYAL VE METOT	31
5.1. Materyal.....	31
5.2. Metot	34
5.2.1. Özellik Çıkarma Yöntemleri	34
5.2.2. MFCC.....	34
5.2.3. Wavelet Packet Transform Shannon Entropi	36
5.2.4. Zero Crossing Rate.....	37
5.2.5. Time Domain Energy	37
6. BULGULAR VE TARTIŞMA	38
7. SONUÇLAR.....	53
KAYNAKLAR.....	54
EKLER	58
ÖZGEÇMİŞ	

ÖZET

Denetimli Öğrenme Yöntemleri Kullanılarak Bir Ses Olay Tespit Sisteminin Gerçekleştirilmesi

Hakan GÜLER

Yüksek Lisans Tezi

FIRAT ÜNİVERSİTESİ
Fen Bilimleri Enstitüsü

Yazılım Mühendisliği Anabilim Dalı

Temmuz 2021, Sayfa: xii + 57

Günümüzde ses ile ilgili çok farklı analiz çalışmaları yapılmaktadır. Bunların başında konuşmayı tanıma, konuşmanın metne dönüştürülmesi, ses transkripsiyonu(sesin ayrıştırılması) gibi çalışma alanları gelmektedir. Yapılan çalışmalar ile ilgili literatür incelendiğinde çalışmaların önemli bir bölümünün bir veri setindeki ses verilerinin öğrenilmesi ve sınıflandırılması üzerine yapılmış olan çalışmalar olduğu görülmektedir. Bilindiği üzere doğada milyarlarca ses bulunmaktadır ve sınıflandırma yöntemleri ile gerçekleştirilen çalışmalar sınırlı veri kümesi üzerinden öğrenimin gerçekleşmesinden dolayı yetersiz kalabilmektedir. Tez çalışmasının ana hedefi “denetimli bir öğrenme modeli ortaya koyarak ses verisi üzerindeki kişi seslerinin ve arka plan sesleri üzerinden çıkarımlarda bulunulması” konusunda araştırma gerçekleştirmektir. Tezde ses verisi ve makine öğrenmesi algoritmaları ile işlenmesi için gerekli olan çalışma yapılmıştır. Tezde akustik ve inceleme yöntemleri ve denetimli makine öğrenmesi yöntemleri hakkında yapılan araştırmalara yer verilmiştir.

Anahtar Kelimeler: Ses Tespiti, Akustik Olay Tespiti, Sınıflandırma

ABSTRACT

Realization of a Sound Event Detection System Using Controlled Learning Methods

Hakan GÜLER

Master's Thesis

FIRAT UNIVERSITY
Graduate School of Natural and Applied Sciences
Department of Software Engineering

July 2021, Pages: xii + 57

Today, very different analysis studies related to sound are carried out. These include work areas such as speech recognition, converting speech into text, voice transcription (parsing of sound). When examining the literature related to the studies, an important part of the studies is the studies on learning and classification of sound data in a data set is observed. As is known, there are billions of voices in nature, and studies with classification methods may be insufficient due to the fact that learning is realized through a limited set of data. The main objective of the thesis study is to “reveal a supervised learning model to make inferences of person voices and background sounds on voice data” is to perform. The thesis carried out the work necessary for processing with sound data and machine learning algorithms. The thesis included researches about acoustics and examination methods and supervised machine learning methods.

Keywords: Sound Detection, Acoustic Event Detection, Classification

ŞEKİLLER LİSTESİ

	Sayfa
Şekil 1.1 Ses Etkinliği Algılama Şeması	3
Şekil 2.1 İşitmenin Nöroanatomisi	6
Şekil 2.2 İnsan Konuşma Üretim Sisteminin Diyagramı [23]	8
Şekil 2.3 Konuşmada Algı Sürecinin Şematik Diyagramı [24]	8
Şekil 2.4 a-) İnsanda Konuşma Sistemi	9
Şekil 3.1 Ses Sinyalinde Çerçeveleme	11
Şekil 5.1 AudioSet Ontolojisi.....	32
Şekil 5.2 Önerilen Denetimli Öğrenme Yöntemi Kullanılarak Gerçekleştirilen Ses Tespit Sistemi.....	34
Şekil 5.3 MFCC özellik çıkarmının akış diyagramı	35
Şekil 5.4 WPT'nin ayrışimsal yapısı.[59]	36
Şekil 6.1 SVM Sınıflandırıcısına Ait Karışıklık Matrisi (Data Set 1)	39
Şekil 6.2 NN Sınıflandırıcısına Ait Karışıklık Matrisi (Data Set 1)	40
Şekil 6.3 NN Sınıflandırıcısına Ait Karışıklık Matrisinin Yüzdelik Gösterimi (Data Set 1).....	41
Şekil 6.4 KNN Sınıflandırıcısına Ait Karışıklık Matrisi (Data Set 1)	42
Şekil 6.5 Ensemble Sınıflandırıcısına Ait Karışıklık Matrisi (Data Set 1)	43
Şekil 6.6 Ensemble Sınıflandırıcısına Ait Karışıklık Matrisinin Yüzdelik Gösterimi (Data Set 1).....	44
Şekil 6.7 SVM Sınıflandırıcısına Ait Karışıklık Matrisi (Data Set 2)	46
Şekil 6.8 SVM Sınıflandırıcısına Ait Karışıklık Matrisinin Yüzdelik Gösterimi (Data Set 2).....	46
Şekil 6.9 NN Sınıflandırıcısına Ait Karışıklık Matrisi (Data Set 2)	48
Şekil 6.10 NN Sınıflandırıcısına Ait Karışıklık Matrisinin Yüzdelik Gösterimi (Data Set 2).....	48
Şekil 6.11 KNN Sınıflandırıcısına Ait Karışıklık Matrisi (Data Set 2)	50
Şekil 6.12 KNN Sınıflandırıcısına Ait Karışıklık Matrisinin Yüzdelik Gösterimi (Data Set 2).....	50
Şekil 6.13 Ensemble Sınıflandırıcısına Ait Karışıklık Matrisi (Data Set 2)	51
Şekil 6.14 Ensemble Sınıflandırıcısına Ait Karışıklık Matrisinin Yüzdelik Gösterimi (Data Set 2).....	52

TABLÖLAR LİSTESİ

	Sayfa
Tablo 6.1 SVM Sonuçları (Data Set 1)	38
Tablo 6.2 Neural Network Sonuçları (Data Set 1)	40
Tablo 6.3 KNN Sonuçları (Data Set 1)	42
Tablo 6.4 Ensemble Sonuçları (Data Set 1)	43
Tablo 6.5 SVM Sonuçları (Data Set 2)	45
Tablo 6.6 Neural Network Sonuçları (Data Set 2)	47
Tablo 6.7 KNN Sonuçları (Data Set 2)	49
Tablo 6.8 Ensemble Sonuçları (Data Set 2)	51

EKLER LİSTESİ

Sayfa

Ek- 1:	CIEA’2021 Konferansı Katılım Sertifikası.....	58
--------	---	----



KISALTMALAR

ML	: Makine Öğrenmesi
AI	: Yapay Zeka
VoIO	: İnternet Üzerinden Ses Protokolü
VAD	: Ses Etkinliği Algılama
SVM	: Destek Vektör Makinesi
LSTM	: Uzun Kısa Süreli Bellek
OOV	: Kelime Dağarcığı
M2M	: Makineden Makineye
GAN	: Üretken Çekişmeli Ağlar
DWT	: Dalgacık Dönüşümü
rVAD	: Güçlü Ses Etkinliği Algılama
ANN	: Yapay Sinir Ağları
SAD	: Konuşma Etkinliği Algılama
ASR	: Otomatik Konuşma Tanımlama
GMM	: Gauss Karışım Modellemesi
AEC	: Akustik Ortam Sınıflandırma
AEI	: Akustik Ortam Tanımla
LPF	: Düşük Geçiş Filtresi
WPD	: Ayrık Dalgacık Dönüşümleri
MFCC	: Mel-Frekans Cepstral Katsayısı
ZCR	: Sıfır Geçiş Hızı
ID3	: Yinelemeli Dichotomiser 3
DNN	: Derin Sinir Ağı
UBM	: Evrensel Arkaplan Modeli
ROA	: Rastgele Orman Algoritması
CART	: Sınıflandırma ve Regresyon Ağacı
SMO	: Sıralı Minimal Optimizasyon
ASC	: Akustik Sahne Sınıflandırması
NN	: Sinir Ağı
TPR	: Gerçek Pozitif Değerler
FNR	: Hatalı Negatif Değerler
KNN	: K-En Yakın Komşu

1. GİRİŞ

Günümüzde ses ile ilgili çok farklı analiz çalışmaları yapılmaktadır. Bunların başında konuşmayı tanıma, konuşmanın metne dönüştürülmesi, ses transkripsiyonu(sesin ayrıştırılması) gibi çalışma alanları gelmektedir. Yapılan çalışmalar ile ilgili literatür incelendiğinde çalışmaların önemli bir bölümünün bir veri setindeki ses verilerinin öğrenilmesi ve sınıflandırılması üzerine yapılmış olan çalışmalar olduğu görülmektedir. Bilindiği üzere doğada milyarlarca ses bulunmaktadır ve sınıflandırma yöntemleri ile gerçekleştirilen çalışmalar sınırlı veri kümesi üzerinden öğrenimin gerçekleşmesinden dolayı yetersiz kalabilmektedir. Bu çalışmadaki ana hedef denetimli bir öğrenme modeli ortaya koyarak ses verisi üzerindeki kişi seslerinin ve arka plan sesleri üzerinden çıkarımlarda bulunan bir sistem gerçekleştirmektir.

Canlıların işitme organları ile algıladıkları basınç değişimlerine ses denilmektedir. Basit bir mekanik düzensizliktir. Ses, maddedeki moleküllerin titreşimi sonucu meydana gelen bir enerji türüdür. Titreşim sonucu oluşan ses, titreşimlerin enerjiye dönüştürülmesini sağlar. Sesin ölçü aracı desibeldir (db). İnsanlarda ortama desibel frekansı 50 ile 60 arasındadır.

Adli bilişim özellikle sanal ortamdaki etkinliğini göstererek, bilgisayar ortamında bilgilerin ulaşılması, elde edilmesi ve incelenmesi noktasında süreklilik ve netlik sağlayan diğer bilim dalları gibi ilke ve prensiplere bağlı olan bir bilim dalıdır. Ses tespiti ve analizi adli vakalarda anlamlı sonuçlar elde edilmesi açısından önem arz etmektedir.

Dijital medya olarak adlandırdığımız; ses, video ve görsellerin günümüzde adli bilişimde önem kazandığı görülmektedir. Kullanılan bu dosyaların delil niteliği taşıması için, bütünlüğüyle beraber gerçekliği de kabul edilebilir seviye de olmalıdır. Bu ihtiyaçlar gereği, parmak izleri, filigranlar veya yardımcı datalar yok ise zorlu bir görev haline gelir. Güçlü ve kullanımı kolay dijital medya işleme araçlarının mevcudiyeti, dijital medyanın bütünlüğünün doğrulanmasını daha da zorlaştırmıştır. Bunlardan yola çıkarak çok rahatlıkla dijital medya adli biliminin bu kayıtlar altında yatan gerçekleri bulmaya ve kanıt olarak kullanılmasına olanak sağlar. Bunun yanı sıra aşağıda bulunan sorulara yanıt aramak ilk amacdır.

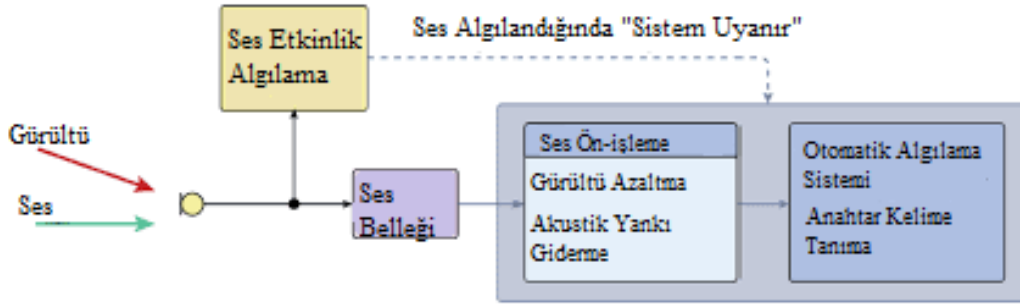
1. İspat niteliğinde bir kayıt "orijinal" mi yoksa birden fazla kaydı birbirine ekleyerek mi oluşturulmuş?
2. Bir delil kayıtlarında varsa sahteciliklerin türleri ve yerleri nelerdir?
3. İddia edildiği gibi, kanıt niteliğindeki kayıt, Y konumunda A edinim cihazı kullanılarak yakalandı mı?
4. İşitsel dosyanın gerçekten de kanıt için, aldatıcı mı yoksa orijinal mi olduğunun kontrolü.

Akustik ortam tanımlama, ses kaydı bütünlüğü kimlik doğrulamasından gerçek zamanlı suç yerelleştirme ve tanımlamaya kadar geniş bir uygulama yelpazesine sahiptir. Örnek verecek olursak, taciz edilen bir kişi tarafından acil çağrı merkezlerine düşen bir senaryo üzerinde düşünmüş olalım. Bu tür suç durumlarında, taciz edilen kişilerin gerçek konumları hakkında herhangi bir ilgili bilgi verememesi çok yaygındır. Ses kaydındaki akustik ipuçları, olay mahallinin akustik alanını (yani araba, cadde, mahalle, oturma odası, banyo odası, yatak odası, mutfak vb.) belirlemek için kullanılabilir.

Bir Ses ve Akustik Olay Tespit Sisteminin Gerçekleştirilmesi ile bir ses dosyası içerisindeki sesin otomatize bir araç ile analiz edilerek etiketlenebilmesi, aynı ses üzerindeki analiz ile ses üzerinden duygu analizinin yapılabilmesi, üst üste binen seslerin (örneğin aynı anda konuşan iki konuşmacının seslerini ayırtmak ya da arka plandaki aynı anda korna sesi ve kuş sesinin ayırtmak) analizine kadar birçok alanda yapılabilecek çalışmalara altyapı sağlanmış olacaktır. Gerçekleştirilmesi planlanan seminer çalışması ışığında akustik sesin analizi ile ses içeriklerinin indekslenebilmesi, ses üzerinden hastalık tespiti, sesin adli bilişim açısından incelenmesinin yapılarak delil niteliğinde verilerin çıkarılabilmesi, ses üzerinden otomatik alarmların üretilmesi, ses bağlamı tanıma, otomatik etiketleme ve ses segmentasyonu gibi birçok araştırma alanında kullanılabilecek bir olay tespit sistemi ortaya konacaktır. Mevcut literatür incelendiğinde konu ile ilgili çalışmaların olduğu görülmektedir. Ancak gerçekleştirilen çalışmaların denetimli öğrenme metodolojilerini kullandığı görülmektedir. Dünya üzerindeki tüm seslerin sınıflandırılmasının mümkün olamayacağı ve öğrenme kümesi için sınıflandırılmış geniş veri setlerinin oluşturulmasının oldukça güç olduğu göz önünde bulundurulduğunda seslerin ayrıştırılabilmesi için denetimsiz öğrenme metodolojilerinin kullanılması gerekliliği ortaya çıkmaktadır [1].

Ses verisi üzerinde denetimli öğrenme modellinden ortaya konulacak hibrit model çalışmanın özgün değerini oluşturmaktadır. Yine aynı şekilde seslerin algılanması ve sınıflandırılması noktasında ortaya koyulacak bir karar destek sistemi oluşturmak, konuşmacı ve arka plan seslerinden anlamlı bilgilerin çıkarılacak bütünlüklü bir sistem oluşturulması çalışmanın bir diğer özgün değeri olarak görülmektedir.

Bu bağlamda tez çalışması ses ve akustik olayların (konuşmacı sesleri ve arka plan sesleri) tespiti ve ayrıştırılarak anlamlandırılması için denetimli ve denetimsiz öğrenme yöntemlerinden oluşan hibrit bir model ortaya koyacaktır. Literatür incelendiğinde tez çalışmasında kullanılan yöntemin özgün olduğu görülmektedir.



Şekil 1.1 Ses Etkinliği Algılama Şeması

Şekil 1.1’de de görüldüğü gibi Ses etkinliği algılama şemasında, Ses belleğinin gürültü ve sesin kaydedildiği ifade edilebilir. Gürültü ve sesin bir arada olduğu ses etkinliğini algılama sürecinde ise algılanan sesler vasıtasıyla sistem uyarıları olarak ses ön işleme noktasında gürültüyü azaltarak akustik yankıyı giderebilmektedir. Otomatik algılama sistemiyle de anahtar kelimeyi tanımlayabilecek bir sürecin oluştuğunu şekil 1.1’den anlayabilmekteyiz.

Şekil 1.1’deki ses kanalına bağlı kalarak ses ve akustik olayların (arka plan seslerin) tespitinin ayrıştırılarak anlamlandırılması süreci tamamlanmış olarak karşımıza çıkabilmektedir.

2. SES VE KONUŞMA

2.1. Ses

Tabiattaki cisimlerinin titreşimi ile oluşan fiziki enerjiye ses denilmektedir. İnsanlarda konuşma veya şarkı söyleme olarak ortaya çıkan enerjidir. Unutulmaması gereken nokta şudur ki: insanlar dünyaya geldiği ilk anda ağlama veya çığlık atmaktadır. Bu enerji sese bir örnektir ve ömür boyu insanlarda konuşma olarak etkisini sürdürmektedir [18].

Sese dair çeşitli tanımlamalar yapılmıştır.[22,24]. Bu kısmı belirtecek olursak:

- Bir kimsenin sesini duymak, onu dinlemektir.
- Kendini ifade biçimi ses olarak adlandırılır.
- Ezgi olarak tanımlanır.
- Konuşan kişinin kim olduğunu bilmemize olanak sağlar.
- Kendini sözlü olarak anlatmaya yarayan fiziksel bir araçtır.
- Kişilerin karakter tanımlanmasında büyük fayda sağlar.

İnsandaki ses ortaya çıkarken farklı şekiller ve kendine özgü bir biçim göstermektedir. Bu süreçte kişiden kişiye değişkenlik gösterir ve çeşitli tonlarda ortaya çıkar. Tüm canlılığın ortak bir paydaşı olan ses çıkarma, kimi canlılarda korunma kimisinde barınma gibi niteliklere göre ortaya çıkmaktadır. İnsanlarda ise konuşma olarak gerçekleşen olaydır. Düşüncelerimizin sözlerle anlatılmasına konuşma denilmektedir [19].

Konuşmaya dair farklı görüşler vardır:

- Düşüncelerin sözlü olarak ifade edilmesi
- Düşüncelerin organize bir yol izleyerek meydana gelmesi olayı
- İnsanın duygusal hareketlerinin fiziksel olarak dışa yansıtılması
- Fikirlerin ne olduğunu anlamaya çalışan bir araç
- İnsanlar arası iletişimi sağlayan bir araç[19].

Bu farklı görüşleri dikkate aldığımızda; insanların çevre ile etkileşimini sağlayan ruhsal, mental, sosyal ve kültürel gelişimlerini etkileyen olaya konuşma diyebiliriz. İnsanlar konuşma yaparak veya şarkı söyleme eyleminde bulunarak meydana bir ses çıkartıyorsa, vücudumuz bu işleme destek verir. Nihai olarak ses ağızdan çıkıyor gibi görünse de vücudumuzun bir çok organı (göğüs, karın, pelvis, boyun, farens) bu olaya yardımcı olmaktadır [21].

İnsanlar arası iletişim ihtiyacı sonucu vücudumuz üzerinden bulunan organların toplu ve düzenli bir hareketi sonucu çevreye çıkan oral iletişimi ses olarak adlandırabiliriz. Bu olay meydana gelirken farklı uyarılar dikkate alınmaktadır. Bu uyarılardan bizi ilgilendiren işitsel kısma bakalım:

2.2. Akustik

Bir ses bilimi olan akustik, sesi inceleyen bilim dalına verilen isim olarak adlandırılır. Katı, sıvı ve gaz halinde olan maddelerde dalga yayılmasının fiziksel özelliklerinin incelenmesini sağlar. Böylelikle katı, sıvı ve gaz halinde bulunan maddeler arasında yayılan dalgalarındaki gürültüye neden olan titreşimin ve gürültünün kontrolünü sağlamaktadır [51].

Birçok bilim insanı akustik ile ilgili olarak insan işitme duyusunu ve ses dalgalarını incelemeye almıştır. Farklı nesnelerin ses üzerinde nasıl bir etki gösterdiği de yine araştırmacıların merak konusu olmuştur. Mühendisler yaptıkları çalışmada insanları farklı ve yüksek seslerden koruma yollarını düşünerek, uygun sesteki iletişim sistemlerini tasarlamaktadırlar. Bilim insanları ve mühendisler tarafından yapılan çalışmalar insanların duymak istemedikleri sesler üzerine kurularak gürültünün kaynağına ulaşma yönündedir. Amaçları gürültünün kaynağını sessiz bir hale getirmektir ve bunun için mühendisler; “Gürültü ile çalışan bir serinletici yanında hiç bulundunuz mu?” sorusunu yönelterek gürültünün kaynağına dikkat çekmişlerdir. Örnek olarak ise; hızı düşük olarak dönen bir pervanenin, hızı yüksek olan bir pervaneye oranla daha az gürültü çıkardığını belirtmişlerdir. Buradan hareketle gürültünün, bir yerden başka bir yere geçmesinin önlenilebileceği konusunda fikir beyan etmişlerdir. Gürültü gelen bir odanın kapısını kapattınız mı gibi sorgulamaları yaparak, perdelerinin ve akustik dalgalanmaların sesi yuttuğuna dikkat çekmektedirler.

Bir ortamda bulunduğunuzda, yansıyarak gelen ses yankı(akis) olarak adlandırılır. Geniş bir odanın içerisinden gelebilecek yankı, uygun bir şekilde odanın yerleştirilmesiyle yankı oluşumu giderilebilir. İnsan kulağı için saniyenin 20’de 1’ine kadar olan bir süre içerisindeki duyulara erişen yankı sesi olmayacaktır. Lakin zaman aralığının uzaması, yankının daha fazla duyulup, rahatsız edici noktasına erişebilmektedir. Duyulan rahatsız edici yankı sesine çoklu yankı denilmektedir. Çoklu yankı olarak bilinen Reverberasyon, yansımış seslerin bir araya gelip daha sonra parça parça sonlanmasıdır. Müzik salonu gibi büyük ve geniş yerlerde reverberasyonların azaltılması için döşeme, perdelerin asılı olması ve duvarların yumuşak olması bu sorunu en aza indirecektir. Böylelikle yapılan yerleştirme ile sesin ortamda emilmesi ya da yutulması daha kolay olacaktır [51].

2.2.1. Akustik Sahne Sınıflandırma Geçmiş

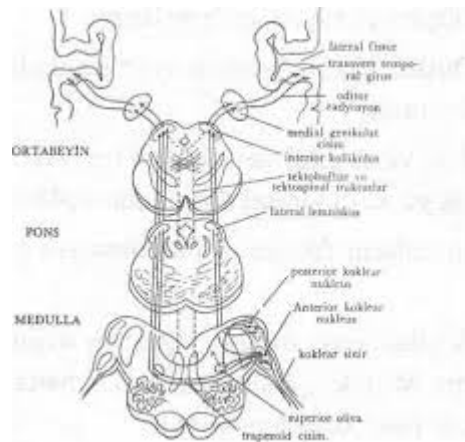
ASC sorununu özel olarak ele almak için literatürde görünen ilk yöntem, MIT MediaLab'ın 1997 yılında yapılan teknik raporunda Sawhney and Maes tarafından önerildiğini söylemiştir. Belirli ses veri kümesi içeren bir dizi kayıtlar yapmışlardır. Bu kayıtlar içerisinde örnek olarak insan, metro, ve trafik sesleri görülmektedir. Bunlar, tekrarlayan sinir ağları ve k-en yakın komşu

kriteri kullanan konuşma analizi ve işitsel araştırmadan ödünç alınan araçları kullanarak ses verilerinden çeşitli özellikleri ayıkladılar. Özellikler ve kategoriler arasında eşleme yaparak % 68 genel sınıflandırma doğruluğunu elde ettiler.

Aynı kurumda görev alan personeller belirli bir süre geçtikten sonra, bisiklet üzerinde seyahat ederken kendilerine takmış oldukları mikrofon ile ses kaydı yapmışlardır. Yapmış oldukları bu ses kayıtlarını daha sonradan kendileri farklı sınıflara tabi tutarak ayırmışlardır. Bazı sesleri ev bazılarını sokak gibi gruplamalar yapmışlardır. Bu tür sınıflandırmaların algısal süreçlere de yardımcı olduğu gözlemlenmiştir. Dubois ve arkadaşları bireylerin semantik kategorilerin kendi taksonomisini nasıl define ettiğini araştırdı. Eronen ve arkadaşları ses sinyallerinin yerel spektral zarfını tanımlamak için Mel-frekans cepstral koefficientlerini (MFCCS) ve istatistiksel dağılımını tanımlamak için Gauss karışım modellerini (GM MS) kullanmıştır. Daha sonra, eğitim sinyalleri kategorileri hakkında bilgi kullanan ayırmacı bir algoritma kullanarak GMMS'nin zamansal evrimini hesaba katmak için HMMS'yi eğittile. Şimdiye kadar belirtilen algoritmalarda, her sinyal bir eğitim kümesi genellikle sıkıştırılmış süre çerçevelere bölünür ve özellik vektörlerinin bir dizi elde etmek için her kareye bir dönüşüm uygulanır. Sonraki süreçte ise karar kriterleri kullanılarak aynı kategoride etiketli olan kayıtlarda atama işlemi yapılmıştır[52].

2.3. İşitsel Uyarılar

İşitsel uyarılar, iç kulak yoluyla başlangıç gösterip beynin temporal lob kısmından, transvers gyrus Heschl merkezine ulaşarak sonlanır [29]. Beynin iki tarafındaki bu uyarılar sol hemisferde bulunan Wernicke alanına transfer edilir. Wernicke alanında ise hafıza, tanıma, anlama ve kayıt gibi bazı işlemlere uğramaktadır. Bu işlemleri tamamlayan işitsel uyarılar cevap için Fasciculus Arcuatus aracılığıyla Beynin frontal lobda Brocca Merkezine iletilir. Burada uyarıya göre motor tarafından cevap düzenlenir (Şekil 2.1.).



Şekil 2.1 İşitmenin Nöroanatomisi

Beynin iki tarafındaki uyarıların sol hemisferde bulunan Wernicke alanına transfer edilme süreci şekil 2.1’de de görüldüğü gibi gösterilmektedir.

2.4. Konuşma Süreci ve Tanımlanması

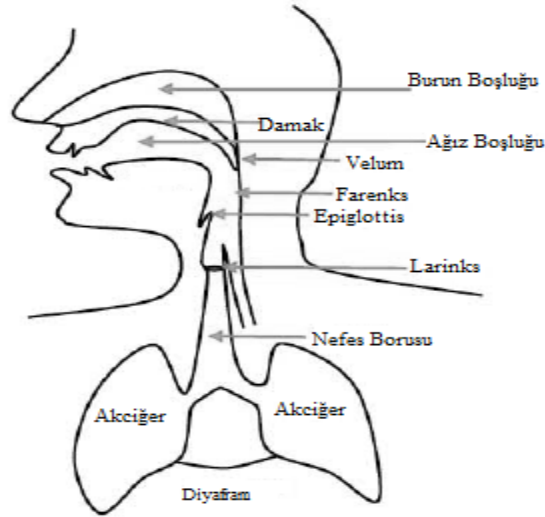
Konuşma süreci ve tanımlamaya yönelik bilgiler literatürden faydalanılarak aşağıdaki gibi tartışılacaktır.

2.4.1. İnsanda Konuşmada Üretim

İnsanda konuşma sisteminden bahsedecek olursak farklı organlardan oluşmaktadır. Bu organlar, nefes borusu, gırtlak, akciğer, boğaz, ağız ve burun boşluğundan oluştuğu gözlemlenmiştir [22]. Bahsedilen organların tek başlarına farklı görevleri vardır. Organlar birbiri ile uyumlu ve senkronize bir halde bir araya gelip iş birliği içinde konuşmayı oluştururlar. Birazdan konuşma için gerekli adımları inceleyeceğiz.

2.4.2. Konuşma Üretimi

Konuşmada üretim denince akla mesaj kısmının geldiği görülmektedir. Konuşmadan önce, konuşmacı mesajı dinleyicilere iletilecek düşünce süreci ile kendi içinde oluşturur. Bu şekilde ileti formül edilmiş olur. Daha sonrasında bu formül dil koduna çevrilir. Kelime oluşturma aşamasına gelinir ve bunun oluşması ise en küçük seslerin birleşmesiyle olur. Bu işlemler yapılırken aynı zamanda sesin yüksekliği, sesin perdesi ve süresi de işlemlerle ortaya çıkarılmış olur. Tüm bu işlemlerin ardından uygun ses dizileri oluşturulur ve konuşmacı tarafından açığa çıkarılır. Sonraki süreçinde akustik dalga olarak açığa çıktığı görülür. Kas kuvvetinin uygulanır, böylece havanın akciğerlerden gırtlak yoluyla çıkar. Hava kesilince yarı olarak bir periyodik dalga oluşumu gerçekleşir. Basınç dürtüsü ağız yolundaki havayı uyarır ve belirli sesler için burun yollarını da uyarır. Çürükler yankılandığında bir ses dalgası (konuşma sinyali) yayılır.

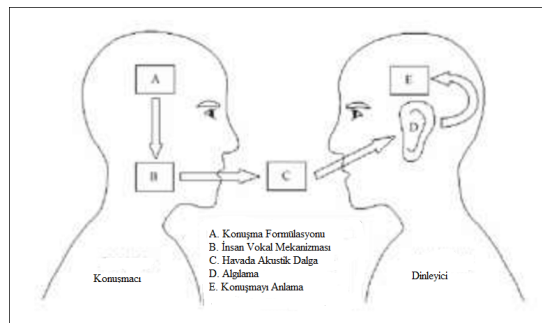


Şekil 2.2 İnsan Konuşma Üretim Sisteminin Diyagramı [23]

İnsandaki konuşma üretim sisteminin diyagramı, şekil 2.2’de gösterilmektedir. Bu diyagrama bağlı kalarak ifade edilebilir ki basınç dürtüsü ağız yolundaki havayı uyarır ve belirli sesler için burun yollarını da uyarır. Çürükler yankılandığında bir ses dalgası (konuşma sinyali) yayılır.

2.4.3. Konuşma Algısı

Konuşma algısı süreci, konuşma oluşturulduktan ve dinleyiciye yayıldıktan sonra başlar. Akustik dalga formunun spektral sinyali, işitsel sinirdeki etkinlik sinyalleriyle aynı dönüştüren baziler membranda analiz edilir. Ses özelliklerinin farklılık gösterdiği görülür. Vücut organlarından beyinde sinir aktivitelerinin dil koduna dönüşümü gerçekleşir. Tüm bu işlem adımlarının ardından mesaj ortaya çıkar ve anlamlı bir hale gelir.



Şekil 2.3 Konuşmada Algı Sürecinin Şematik Diyagramı [24]

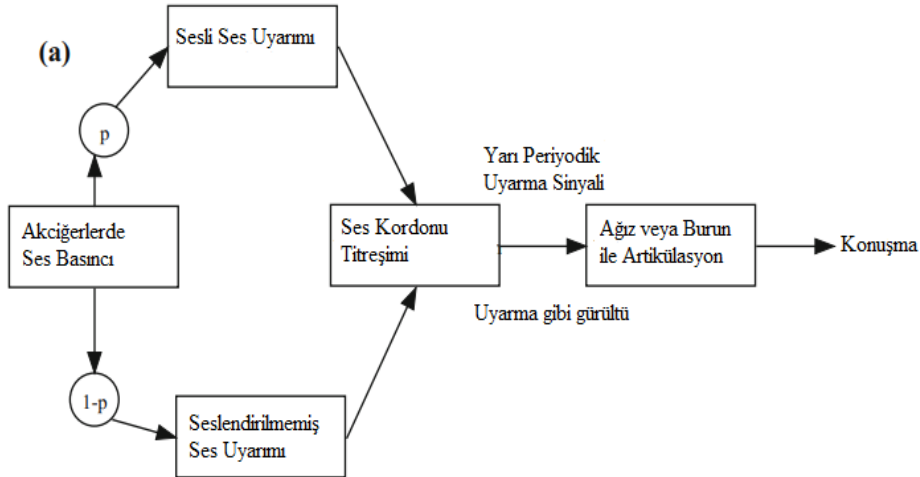
Şekil 2.3'te de görüldüğü gibi Konuşmada Algı Sürecinin Şematik Diyagramı konuşmacı ve dinleyici merkezli olarak vücut organlarından beyinde sinir aktivitelerinin dil koduna dönüşümü gerçekleşir.

2.4.4. Sesli ve Sessiz Konuşma

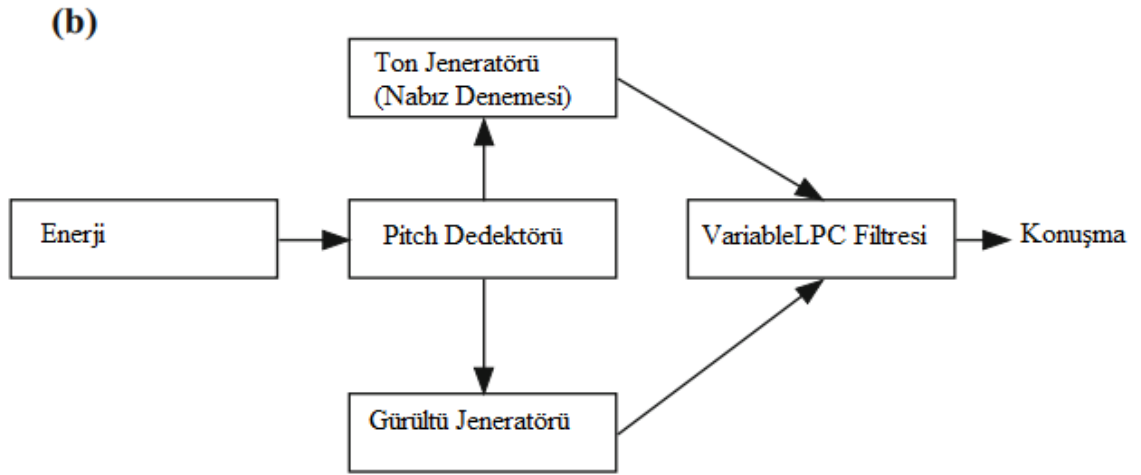
İki tür konuşma vardır: sesli ve sessiz. Ses yolu ve ses tellerinin kullanımını birbirinden ayırır. Vokal işlemlerin konuşma esnasında sesli durumlarda kullanımı söz konusudur. Konuşmanın oluşabilmesi için farklı seslerin üretimi söz konusudur. Ünlü seslerin sesli olduğunu herkes tarafından bilinmektedir. Sessiz olan ünlülere ise örnek verecek olursak p, s, h gibidir. Konuşmada bir sesin sesli veya sessiz olduğunun anlaşılabilmesi için farklı yöntemlerin kullanıldığı görülmektedir.

2.4.5. İnsan Konuşma Modeli

Hem burun hem de ağız boşluğundan oluşan supralaringeal vokal sistem, diğer bazı frekanslarda geçişine izin verirken belirli frekanslarda sesleri bastıran zaman değişken bir düşük geçiş filtresi (LPF) görevi görür. Bu LPF için transfer fonksiyonu, akustik rezonans sistemi olarak hizmet veren tüm ses yolu tarafından belirlenebilir. Karar mekanizması olarak konuşmanın duyulmadığına karar verilir. [33-34]



Şekil 2.4 a-) İnsanda Konuşma Sistemi



Şekil 2.5 b-) Elektrik Sisteme Eşdeğer Model

Şekil 2.4 a-) 'da gösterildiği gibi insandaki konuşma sisteminin aşamaları tablolaştırılmıştır. Ayrıca şekil 2.5 b-) 'de ise Elektrik sisteme Eş değer model konuşma tablosuyla bütünleştirilerek ifade edilmektedir.

3. ÖNİŞLEMLER

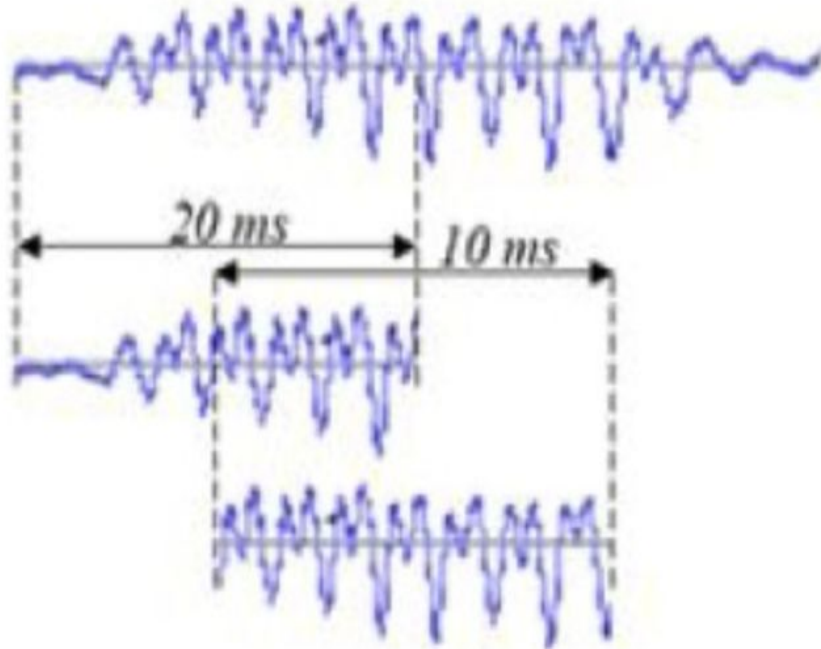
3.1. Ön Vurgu Filtresi

Sinyallerin yüksek frekans değişimine ilişkin olarak ön vurgu filtresi enerjiyi arttırmak amacına hizmet etmektedir. Vurgu filtresine dair transfer fonksiyonu aşağıda formül ile gösterilmiştir [53].

$$y(n) = x(n) - 0.95x(n - 1)$$

3.2. Çerçeveleme

Konuşma sinyalinin x miktarda konuşma örneğini taşıyarak parçalara bölünmüş olması işlemine çerçeveleme denilmektedir. Çerçeveleme işlemi devam ederken ilk olan çerçeve a miktarında konuşma içeriyorsa bir sonraki gelen çerçeve ise b miktarı sonra başlatılır. Bu işlem yapılırken dikkat edilmesi gereken bazı hususlar vardır. Bunlardan en önemlisi b miktarının a miktarından daha küçük bir değer olarak seçilmesidir. Böyle seçim işlemi olduktan sonra her çerçeve kendinden sonra gelecek olan çerçeveyi örtmüş olacaktır. Bu sayede çerçeveler arasında keskinlik ortadan kalkacaktır [54]. Bu çalışmada kullanılan çerçevelerin uzunluğu 25 ms olup örtüşme uzunluğu 10 ms şeklindedir.



Şekil 3.1 Ses Sinyalinde Çerçeveleme

Şekil 3.1’de gösterildiği gibi ses sinyalinde çerçeveleme kısmı ses dalgalarıyla ifade edilmektedir. Yapılan bu araştırma da kullanılan çerçevelerin uzunluğu 25 ms olup örtüşme uzunluğu 10 ms şeklinde şekil 3.1 aracılığıyla gösterilmektedir.

3.3. Pencereleme

N adet örnek bulunan bir sinyal de matematiksel bir ifade ile belirtilen pencere ifadesiyle çarpılması; pencereleme işlemini tanımlamaktadır. Bu işlem için X kadar örnek çerçeve adı verilen sinyal bölümünden oluşmaktadır. Bu ses sinyalinden kendine özgü nitelikler ortaya çıktığı gibi, sinyalin çerçevelere bölünmesiyle de nitelikler ortaya çıkarılabilir. Hamming Hann(Hanning), Gauss, Welc, Blacman, Triangular ve Bartlett gibi kullanılan yöntemler bazı pencereleme yöntemleri olarak literatürde yer almaktadır. Bu yöntemlerin çoğu zaman sinyali sıfıra doğru azaltabildiği ve çoğu zaman da sinyali azaltmadığı belirtilmektedir [54].

3.4. Sıfır Geçiş Oranı

Sinyallerin zamanlama açısından ayırık olması ve sinyal içerisindeki ardışık örneklerin farklı matematiksel işaretlere sahip olması oranına, sıfır geçiş oranı denir. Bir sinyalin frekans içeriğindeki basit bir ölçüm olarak adlandırılmasına sıfır geçiş oranı denilmektedir. Ses sinyallerinde sıfır geçiş oranı ise bir çerçevedeki sinyalin genlik değerine göre hesaplanır. Genlik değeri kaç defa geçmiş ise ölçüm böyle yapılmaktadır. Ses sinyalleri geniş bantlı sinyaller olduğundan ortama sıfır geçiş oranının yorumlanması daha az kesinlik ile sonuçlanabilmekte iken ses sinyalinin spektral özellikleri kısa zamanlı ortalama sıfır geçiş oranı (short-time average zero crossing rate) ile daha kolay hesaplanabilmektedir [54].

3.5. Enerji

Kısa zamanlı ses sinyalleri üzerinde gelen değişimler enerji olarak adlandırılmaktadır. Büyük ölçüde çeşitlilik gösterebilen ses sinyallerinin enerji değerleri sesli – sessiz bölge ayrımında kullanılabilir. Örneğin bir konuşma sinyalinde, sesli bölgelerde temel frekansın (fundamental frequency) ölçülebilir boyutta değiştiği gözlemlenebilmektedir. Bu yüzden zaman domeninde yapılan işlemler sonucunda sinyalin yoğunluk, perde ve formant frekansı gibi vokal yol özellikleri elde edilebilmektedir [54].

3.6. Hızlı Fourier Dönüşümü

Fourier dönüşümü olurken bazı hesaplamalar yapılır. Bu hesaplama da her bir A_x değeri için N adet karmaşık çarpma ve $N-1$ tane toplama işlemi kullanılır. Bundan dolayı N adet Fourier değeri bulurken N^2 çarpma ve $N(N-1)$ toplama işlemi yapılmalıdır. Yine farklı toplama işlemleri yapıldıktan sonra Fourier dönüşümü hesaplanması yapılmış olur. Nihai olarak elimizde bir dizinin olduğunu bilelim. Bu dizinin uzunluğu N olarak bilinen değerin 1000'in üzerinde olması neticesinde Fourier bulunması için çok fazla işlemin yapılması gerekecektir. Buradan N sayısı ile doğru orantılı sonucunu çıkarabilir ve N sayısı arttıkça işlemin de hızlı arttığı görülmektedir. Bu nokta da karşımıza çıkan sonuçlar Hızlı Fourier Dönüşümü algoritmaları diğer algoritmalara göre hesaplamadaki hız ve ekonomiklik açısından daha başarılı bir yöntemdir [53].

3.7. Mel Ölçeği Filtre Bankası

Konuşma tanıma sistemlerinde sıklıkla Mel frekansı kepsral katsayıları kullanılmaktadır. Kepsral katsayılar doğrusal ölçekleri temsil etmektedir. İnsan kulağı ise 1kHz'in altındaki ve üstündeki frekansları doğrusal ve logaritmik şeklinde ayırtılabilmektedir. 1kHz'in altındaki frekanslar insan işitme sisteminde doğrusal ölçekli sınıfa dâhil olurken, 1 kHz'in üstündeki frekanslar logaritmik ölçekli olarak duyulmaktadır. MFCC'nin kullanılmasındaki amaç ise kepsral katsayılarının insan işitme sistemine uyumlu olmasıyla alakalıdır [53].

3.8. Öznitelik Çıkarma Yöntemleri

Konuşma sinyali konuşmacının aklında tasarladığı fikir hakkında bilgiler içermektedir. Konuşma sinyalleri içerisinde konuşmacıya ait akustik özelliklerin bulunabileceği gibi, mekânda bulunan ortam sesleri, donanımsal gürültüler de yer alabilmektedir. MFCC, LPC ve PLP gibi yöntemlerle konuşmacıya özgü akustik model çıkarılarak konuşmacı tanıma işlemleri rahatlıkla sağlanabilmektedir. MFCC'nin ortaya çıkış sebebinde insan kulağından ilham alınmıştır. En yaygın olarak kullanılan öznitelik çıkarma yöntemlerinden biri Mel ölçeği yöntemidir. MFCC'nin en büyük özelliklerinden birisi de zaman döneminde işlem yaparak net sonuçlar vermesidir. Şimdi bir konuşma da var olan sinyallere uygulanan MFCC adımlarını izleyelim.

- Ses sinyali öncelikle çerçevelere ayrılmalı ve her çerçevedeki sayıları aynı olmalı.
- Ayrılan bu çerçevelerin her birisine pencereleme yöntemlerinden birisi uygulanmalı.
- Daha sonrasında ses sinyalleri üzerinde bulunan frekansların değerlerini çıkarabilmek için çerçevelerin her birisine FFT uygulanmalı.
- FFT uygulanmış çerçevelerin her bir çerçevesine yüksek frekans içeren mel ölçekli filtreler uygulanmalıdır.

- Mel ölçeđi uygulanan bu çerçevelerin her birisine daha sonrasında DCT uygulama işlemi yapılır.

Bu işlemler yapıldıktan sonra elde bir MFCC seti edilir. Bu sete artık ses içerisinde bulunan fonetikleri içeren bir akustik veri seti de diyebiliriz [54].



4. DENETİMLİ ÖĞRENME YÖNTEMLERİ

Sınıflandırma denetimli öğrenme altına girer. Eğitim veri kümeleri bir öğrenme süreci ister. Bu geçen sürecin en önemli ölçütüne denetimli öğrenme diyebiliriz. Eğitim veri kümelerini gözlemlene de ve bunları sınıflandırma karar mekanizması olarak çalıştığı gözlemlenir. Sınıflandırma denince aklımıza ilgili verilerin aynı kümeye atanması gelir. Bu işlem yapılırken belirli özellikler baz alınarak yapılır. Sınıflandırma yapılırken öncelikle model oluşturulur. Daha sonrasında oluşturulan bu model kullanılır. Model oluşturulurken bir öğrenme sürecinden geçer. Bu aşamalardan geçerken sınıflandırma etiketleri kullanılır. Etiketleme işlemi sırasında eğitim kümesinde yer alan veriler analiz edilir ve bunlara karşılık gelen etiket sınıfına dahil edilir. Bu veri seti oluşturulduktan sonra test verileri oluşturulur. Test verilerinin oluşturulma amacı doğruluk tahmin etmek içindir. Sınıflandırmanın kabul edilebilmesi için gerekli ölçüt, önceden belirlenen eşiğin üzerinde olması gereklidir. Doğruluk kabul edilebilir ise sınıflandırma kuralları yeni veri kümesi için geçerli olabilir. Bu bölümde bazı son teknoloji sınıflandırma tekniklerini ve bunların otomatik konuşma tanıma sisteminde uygulanmasını özetlemektedir. Birazdan inceleyeceğimiz bölümde denetimli öğrenme yöntemlerinden bahsedeceğiz. Ayrıca bahsettiğimiz yöntemle dair uygulama alanlarına da değineceğiz.

4.1. Denetimli Öğrenme Algoritmaları

Denetimli öğrenme yönetimi algoritmalarını aşağıdaki alt başlıklarda incelenecektir.

4.1.1. K-En Yakın Komşular Yöntemi (k-NN)

En basit ve temel sınıflandırma tekniği olarak K-en yakın komşu yöntemi karşımıza çıkmaktadır [35]. Verilerin dağıtımı ile ilgili bilginin zayıf veya az olduğu sistemler için geçerlidir. Bu metodoloji, olasılık yoğunluklarının parametrik tahmininin mevcut olmadığı veya tahmin edilmesi zor olduğunda verilerin ayrımcı analizini gerçekleştirmek için geliştirilmiştir. k-NN, sınıflandırma için kullanılan parametrik olmayan, tembel ve örnek tabanlı denetlenen makine öğrenimi algoritmasıdır. Parametrik olmayan terim, k-NN'de, dikkate alınan verilerle ilgili hiçbir varsayım yapılmadığı anlamına gelir. K-NN en büyük özelliği var olan verileri sınıflandırma amacı gütmektir. Tahmin işlemi için ham eğitim verileri veya örneği kullanıldığından, örnek tabanlı algoritma olarak da adlandırılır. k-NN prosedürünün önceki eğitime ihtiyacı yoktur. k-NN, bir genelleme yapmak için önceki veri kümesini öğrenmek yerine tüm veri kümesini saklar. k-NN'de hiçbir öğrenme süreci olmadığından, gerekli tüm işler tahmin anında yapılır. k-NN'ye tembel bir süreç olarak nitelendirilmesinin en büyük sebebi budur. Veriler özellik veya metrik uzay içerisinde dahil bulunmuş olabilir. Skaler veri veya çok boyutlu veri olabilir. Hesaplamanın bir özellik

uzayında yapılması gerektiği için mesafe ölçümü kavramı gerekliliği vardır. Eğitim veri kümesi vektör kümesini ve ilgili sınıf etiketlerini içerir. k-NN, eğitim veri kümesini doğrudan tahmin eder.

Ağırlıklı KNN algoritması kullanarak SROL veritabanı için konuşma duygu tanıma için yeni bir metodoloji Feraru ve Zbancioc tarafından tanıtıldı [25]. Wk-NN yöntemini dayanıklılık seviyesine çıkarmada kullanmak için önerdiler. K-NN bazı değerlerinin değiştirilerek güncellenmiş haline denilmektedir. Rizwan ve Anderson tarafından k-NN ve hoparlör sıralaması kullanılarak fonem tahmini için son teknoloji bir metodoloji önerilmiştir [26]. k-NN, verilen test hoparlörünün örnek alanı ile benzerlik analizi yaparak hoparlör sıralamasını hesaplamak için kullanılmıştır. En yakın komşu ve k-NN çoğunluk oylama yaklaşımı ile deneysel sonuçlarının karşılaştırmalı bir çalışma yaptılar. Önerilen şema gelişmiş tahmin doğruluğu göstermiştir. Bu yaklaşım, belirli bir konuşmacı için konuşma tanıyıcının anında özelleştirme ve benzerlik ölçüsüne bağlı olarak eğitim verileri özelleştirme konusunda ek bir avantaja sahiptir.

4.1.2. Naïve Bayes (NB) Sınıflandırıcı

k-NN kullanmanın tipik bir dezavantajı, sınıflandırma mesafe ölçüsüne dayandığından önemli nitelikleri tanımlayamamasıdır. Bazen, aynı mesafe ölçüsüne sahip daha az önemli olan niteliklerin önemli nitelikler olarak kabul edildiği ortaya çıkabilir. k-NN'nin en büyük dezavantajı, eksik verileri işleyemez. KNN, yönteminin dezavantajı olarak görülmesinin en büyük sebebi, büyük veriler üzerinde yavaş hareket ediyor olmasıdır. Naïve bayes bu konuda KNN den farklı olarak büyük veriler üzerinde uygulanmaya müsait bir yöntemdir. Bu yöntemin başka avantajlarından birisi ise eksik veri olduğu takdirde etkin bir şekilde işlemesidir. Bu avantajları kendisini KNN yönteminden daha hızlı kılar. KNN yönteminden daha hızlı olan bir öğrenme sınıfı olan Naïve Bayes yönteminde max olasılık hipotezi yer alır. Naïve Bayes yönteminin bir diğer önemli özelliği ise öğrenme şeklinin zamanla olmasıdır.

Naïve Bayes yöntemi ile ilgili geniş araştırmalar literatürde karşımıza sıkça çıkmaktadır. Naïve Bayes sınıflandırma yöntemi olarak kullanılsa da çok güçlü bir yöntem olmakla beraber kendi kategorisinde yer alan sınıflandırma tekniklerinden daha iyi performans gösterdiği yıllar içinde görülmüştür. 1960'lı yıllardan sonra ise popüler olmaya başlamıştır [26].

Özelliklerin sonuçlar üzerinde etkisi bulunmaktadır ve Naïve Bayes eşit bir bağımsız etkiye sahip bir yöntemdir. Bayes teoremi, olasılık üzerine çalışarak daha önce gerçekleşmiş olan başka bir olayın olasılığı içerisinde verilen bir olayın, olasılığını bulur. Son yıllarda Naïve Bayes sınıflandırıcı müzik veya konuşma tanıma ve sınıflandırma yaygın olarak uygulanır. 2010 yılında Naïve Bayes sınıflandırıcı Zhouyu Fu et al. müzik sınıflandırma ve geri alma için keşfedildi [27]. Naïve Bayes yönteminde tek düzey özellik yerine tüm ses özelliklerinin kullanılması baz alınmıştır.

Naïve Bayes yöntemine ait olan sınıflandırma çeşitlerinden iki tanesini incelediler. Naïve Bayes sınıflandırıcısını 2012'de Sanchis ve arkadaşları öneri olarak sundular. Çalışmalarında konuşma tanımda güven tahmini yapmaktaydılar [28]. Güven tahmini, ayıklanan özellik kümesine göre doğru olup olmadığı varsayımlı bir sözcüğü sınıflandırmak için kullanılmıştır. Bundan dolayı bu kısım da Naïve Bayes yöntemi öneri olarak sunuldu. Birleştirilmiş özellik çıkarma tekniği ve konuşma tanıma için Naïve Bayes sınıflandırıcısı Sonia Sunny ve diğerleri tarafından 2013'te önerilmiştir. Malayalam'daki rakamları tanımlamak için bir konuşma tanıma sistemi geliştirdiler.

Burada, özelliklerin çıkarılması için ayrık dalgacık dönüşümleri (DWT) ve dalgacık paket ayrışması (WPD) gibi dalgacık yöntemleri kullanılmıştır. Özellikler çıkarıldıktan sonra, sınıflandırma amacı için Naïve Bayes sınıflandırıcısı uygulanmıştır. 2016 yılında, Naïve Bayes sınıflandırıcısını uygulayarak ses sinyaline dayanan duyguları tanımak için yeni bir metodoloji Bhakre ve Bang tarafından önerildi [29]. Ses sinyalini Naïve Bayes sınıflandırıcısını kullanarak dört temel duygusal duruma ayırdılar. Sesin ön işlemlerinden önemli işlem adımları olan enerji, mel-frekansı, sıfır geçiş oranı, frekans gibi özellikler değerlendirilmiştir.

4.1.3. Karar Ağacı ve Konuşma Sınıflandırılması

Karar ağacı, iş analitiği ve kritik karar verme süreçlerinde çeşitli uygulamalara sahip tahmin tabanlı bir analitik araçtır. Ön koşullara dayalı veri kümeleri vardır. Bu veri kümelerini parçalara ayırmak ve ayrılan bu parçaları ise belirli niteliklere göre ayırmak için kullanılan önemli bir yöntemdir. Denetimli öğrenme yöntemi olarak karşımıza çıkan karar ağaçları, sınıflandırmanın yanı sıra regresyon görevi olarak da yer almaktadır. Bir tahmin modeli ortaya konulur. Konulan bu tahmin modeli belirli karar kuralları oluşturması için veri özelliklerinden çıkarılır. Çözümsele sonuçlar için grafiksel olarak temsil etmede karar ağaçları kullanılır. Öznitelikler üzerinde test veya farklı koşullar uygulamak için kullanılan ağaç benzeri yapılara iç düğümler olarak isimlendirme verilir. Test sonucunu temsil eden kısma dal ismi verilir. Sınıf etiketleri veya çıkan karar sonuçlarını tutan kısma ise yaprak düğüm denir. Bahsedilen bu süreç tüm alt ağaçlarda geçerlidir. Belirli bir tuple T için sınıf etiketi bilinmiyorsa, aynı karar ağacından elde edilen karar kuralları tarafından tahmin edilir. Karar ağacı ile ilgili çeşitli tanımlamalar yapılmıştır. Bunları sırasıyla inceleyecek olursak,

- Kök Düğüm: Örnek veri kümesinin temsil edilmesidir. Homojen yapıdadır.
- Yaprak Düğümü: Sınıf etiketleri yaprak düğümlerinde bulunur.
- Bölme: Kök düğüm veya yaprak olmayan düğümler, bazı koşullara bağlı olarak alt düğümlere ayrılabilir ve sınıf etiketlerine doğru farklı yollar oluşturabilir.
- Dallar veya alt ağaçlar: Dallar veya alt ağaçlar bölme işlemi ile oluşturulur.
- Budama: Budama, istenmeyen dalları karar ağacından çıkarma işlemidir.

Karar ağacı oluşturmak için kullanılan en önemli algoritma olarak karşımıza çıkan ID3 algoritması Quinlan tarafından literatüre kazandırılmıştır[28]. ID3 algoritmasında kullanılacak özelliklerin ne olacağına karar verilirken entropi ve bilgi kazanç hesaplamaları dikkate alınır. Entropi bir sistemin belirsizlik ölçüsüdür. Entropi kavramı tarafından bilgisayar bilimine uyarlanmıştır ve sıklıkla kullanılmaktadır. Bilgisayar biliminde entropi verinin karmaşıklık ölçümüdür. Entropi değerinin yüksek olması verinin karmaşık ve örüntülerinin az olduğunu ifade eder. Entropi (1) formülüne göre hesaplanır. Entropi değerini temel alarak elde edilen bir diğer değer bilgi kazanç değeridir ve (2) formülüne göre hesaplanır. ID3 algoritması ile kazanç değeri hesaplanır ve bu değer en yüksek olan kök düğüm olarak adlandırılır. Bu olay döngü içerisinde devam ederek ağacın yapısı oluşturulur. ID3 algoritması entropi ve bilgi kazanç değerlerini baz almasından ötürü entropi temelli karar ağacı olarak da adlandırılmaktadır [30].

$$H(S) = \sum_{i=1}^n -p(c) \log_2 p(c) \quad (1) = H(S) - \sum p(t)H(T) \quad (2)$$

4.1.4. Destek Vektör Makinesi (SVM)

SVM olarak karşımıza çıkan destek vektör makinesi sınıflandırma için kullanacağımız denetlenebilen bir öğrenme algoritmasıdır. Geniş bir yelpazeye sahip olmakla beraber çok çeşitli alanlarda kullanılması mümkündür. Bu alanlar için örnek verecek olursak, metin analizinden tutun da desen tanımaya kadar örnek olarak verilebilir. Bu makine öğrenmesi yönteminin temel ilkesi, tüm verilerin aynı sınıfa ait veri noktalarını toplamak için sınır tespit eder ve bu sınır içinde olup olmadığını bilmek için tekrar ürettiği bit veri noktası ile karşılaştırmasını yapar [43-44]. SVM yöntemi, bir sınır sabitlendikten sonra eğitim verilerinin çoğunun gereksiz hale gelmesi anlamında genel bir yöntemdir. Sınırı kurmaya yardımcı olan veri noktalarına destek vektörleri denir. Destek vektörü olarak adlandırılma sebepleri, bu noktaların sınırı destekliyor olmalarıdır. Vektörler, veri noktası farklı öznelik ve özellik değerleri içermektedir. Genel olarak verilen ismi ise hiper düzlemdir. Bu veri noktalarını sınıflandırmak için hiper düzlem kullanılır ve iki boyutlu boşluklarla bu sınıfları ayırır. Ayırım yapan bu hiper düzlemler ve veri noktaları arasında ne kadar fazla uzaklık varsa doğru şekilde sınıflandırma olasılıkları da o derece yüksek çıkar. Hangi sınıfa ait olduğunu bulmak için ise hiper düzlemin hangi tarafa indiğine bakmak yeterlidir.

Sınıflandırma için önemli bir karar olan hiper düzlem seçiminde veri noktaları arasında birçok defa bulunabilir. Hiper düzlemlerin uygunsuz bir şekilde seçilmesi sonucu ise yanlış sınıflandırmalar olma ihtimali çok yüksektir.

SVM makine öğrenme yöntemi ve algoritması karşımıza çıkan son teknoloji bir sınıflandırma yöntemidir. SVM maksimum marj çözümüne dayandığından, gürültü varlığında doğrusal olmayan sınıflandırıcıların çoğuna kıyasla daha iyi sonuç verir. ASR sistemi gürültüye

duyarlı olduğundan, son yıllarda konuşma tanıma ve sınıflandırma alanında SVM uygulamasına ilgi giderek artmaktadır [45-46]. 2011 yılında, Shen ve diğerleri tarafından otomatik konuşma duygu tanıma için yeni bir metodoloji önerildi [31]. Shen ve arkadaşları önerdikleri bu metotta beş sınıftan oluşan bir duygu sınıfı oluşturdular. Bu sınıf içerisinde; hüznün, sevin, iğrenme, sıkıntı ve nötr olmak üzere bir sınıf oluşturuldu. Bilinen ses özellikler çıkarma yöntemlerinde, enerji, doğrusal tahmin katsayısı ve mel frekans katsayısı gibi özellikler çıkarılmıştır. Buradan ibaretle farklı duygusal durumları sınıflandırmak için destek vektör makinesi yöntemi uygulandığı görülmektedir. Davood ve arkadaşları aynı yıl içerisinde insanın konuşma özelliklerine dayalı yaş tahmin etmek için bir metodoloji [32] önerdi. Otomatik bir yaş tanıma sistemi için, farklı altı yaş grubuna dahil farsça konuşmacıların seslerini içeren bir veri seti oluşturdular. Bu veri setini sınıflandırmayı hedeflediler. Ses özellik çıkarma yöntemlerini kullanmak için destek vektör makinesini kullandılar. 2013 yılında, Matoušek ve Tihelka tarafından yanlış açıklamalı kelimeleri tespit etmek için bir metodoloji önerildi [33]. Yanlış açıklamalı sözcükleri tespit etmek için SVM'den yararlandılar. 2016 yılında Kamil Aida-zade vd. MFCC ve LPC özelliklerine dayalı Azerbaycan DataSet için ASR sisteminin akustik modeline destek vektör makineleri uyguladı [34]. Çeşitli sonuçlar elde ettiler ve bu elde ettikleri sonuçları analiz ettiler. Shi ve arkadaşları tarafından SVM algoritmasına dayanan bir konuşma sınıflandırma yöntemini önerdiler [35]. Destek vektör makine yönteminin daha gelişmiş bir versiyonunu geliştirdiler. Farklı seslerden oluşan dört çeşit müzik seçerek bir veri seti oluşturdular. Bu veri setinde türk, pop, rock vs gibi farklı sesleri içeren sesler vardır. Konuşma özellikleri yöntemler kullanılarak çıkarıldı ve veri setinde eğitim için geliştirdikleri yeni algoritmayı uyguladılar. Bunun sonucunda uygun bir sınıflandırıcı modelini literatüre kazandırmış oldular.

4.1.5. Random Forest Algoritması

Günümüzde RO Algoritması, sınıflandırmada çok iyi performans sergilediği için toplu öğrenme yöntemlerinde sıklıkla tercih edildiği görülmektedir. Son yıllarda geliştirilen RO sınıflandırıcısı, hem hızlı hem de yüksek doğruluk sağlaması yönünden toplu öğrenmede çok iyi iki yöntem olarak bilinen Hızlandırma ve Torbalama yöntemlerine göre daha fazla avantaj sağladığı görülür.

Öğrenme yöntemleriyle karşılaştırıldığında RO sınıflandırıcısı, özellikle hızlandırma yöntemine göre, eğitim aşamasında çok daha hızlı olduğu görülmektedir. Yeterliliği ve doğruluğu ile çok kullanışlı bir sınıflandırıcıdır. Hızlandırma yönteminin hesap yükü fazla olduğu için Torbalama yönteminden daha yavaştır. Fakat çoğu durumda bu yöntemden daha doğru sonuçlar verildiği görülmüştür. Hızlandırma yönteminin, çok yavaş ve gürültüye karşı duyarlı olması,

tekrarlı eğitimin olabilmesi gibi dezavantajlarına karşın RO, hesapsal olarak hızlandırma sınıflandırıcısından çok daha basittir ve gürültüye karşı duyarlı değildir.

RO, ağaç tipi sınıflandırıcılar topluluğudur. Daha gelişmiş bir sınıflandırma yöntemidir. Breiman ve Cutler RO'yu, şuan ki algoritmalar arasında doğruluğu eşsiz olan bir sınıflandırıcı olarak tanımlamıştır. Ayrıca hızlı ve belirli bir kalıbı olmayan bir yöntem olduğunu söylemektedir. Bu yöntemde ne kadar istenirse o kadar ağaçla çalışılacağını söylemişlerdir.

Breiman'ın RO yönteminde, Torbalama, rastgele özellik seçimi ile birlikte ele alınır. RO'da Torbalama yönteminin tercih edilmesinin iki önemli nedeni vardır; birincisi, Torbalama "işleminde rastgele özellik kullanıldığından doğruluğun artması; ikincisi, genelleştirilmiş hataların hesaplanmasıdır. Random Forest algoritması uygulamak için gerçek veri seti olarak bilinen veri setinde birbiri ile yer değişen yeni bir veri seti oluşturulmuştur. Oluşturulan bu veri setine random forest uygulanarak yeni bir ağaç geliştirilmiştir. Herhangi bir budama işlemi yapılmaz bunun sebebi ise performansı etkiliyor olmasıdır. Budama olmaması random forest algoritmasını diğer karar ağacı yöntemlerinden avantajlı kılmaktadır. Kullanıcılar tarafından tanımlanan iki parametre random forest algoritmasına gönderilir ve yeni bir ağaç üretilir. Bu gönderilen iki parametreden birisi en iyi bölümlmeyi belirlemek için düğümlerde kullanılan değişken sayısı iken diğer parametre ise geliştirilecek ağaç sayısını vermektedir.

Kullanıcı tarafından başlangıç m değeri rastgele seçilir sonraki m 'ler genelleştirilmiş hatalara göre artırılır ya da azaltılır. m azalınca korelasyon ve güç azalır, m artınca korelasyon ve güç artar. Bu şekilde en uygun m bulunur ve sınıflandırma duyarlılığı artar, hata azalır. Breiman'a göre, m değişken değeri seçilirken, m değerinin M (toplam değişken sayısı) değişkeninin kareköküne eşit olarak alınması genellikle uygun değer sonuçlar verir. Tüm girdi değişkenleri random forest algoritmasında kullanılmamaktadır. Random Forest hesap karmaşıklığını azaltır ve bunun yanı sıra ağaçlar arasındaki korelasyon katsayısını azaltır. Bunun için bir m değeri seçimi yapar. Seçilen bu değer, sınıflandırmada önemli bir role sahiptir. Önem ölçümlerine göre değişkenler belirlenmektedir. Random forest yönteminde elde edilen üç parametreye sahiptir. Bunlar yakınlık analizi, genel hata ve değişken önemi olarak karşımıza çıkmaktadır. Genel hata olarak bilinen parametre, sınıflandırma da doğruluğun anlaşılmasına yardımcı olmaktadır. Eğitim veri setinin T olduğunu düşünelim. Birbiri ile yer değişen yeni veri setine ise T_k ifadesini verelim. Bu yeni veri setinden $h(x, T_k)$ sınıflandırıcı elde edilir. Sınıflandırıcı ile çantaya atılmış tahminlerden oylama yapılır. Eğitim verisi olarak adlandırılan sette bulunan her x ve y değerleri için oylama yapılır ve bu sınıflandırmaya tabi tutulur.

RO algoritması, genelleştirilmiş hata verisindeki verilerin yerleri değiştirildiğinde tahmin hatasının ne kadar olduğunu inceleyerek değişkenlerin önemini, etkilerini (variable importance) hesaplar (Liaw ve Wiener, 2002). Değişkenlerin yerlerinin değiştirmesindeki amaç, kullanılan değişkenlerin ne kadar önemli olduğunu anlamak içindir. Bu değişimler sonucunda herhangi bir

hata oluşuyor ise o değişkenin yapılan işlemdeki önem derecesini ortaya çıkarmaktadır. 4 çeşit değişken önem yöntemi karşımıza çıkmaktadır. Sırasıyla bunlar; hata artışı, ortalama hata payı artışı, hata payı artışının diferansiyeli ve gini düşüştür. Furlanello 'nun çalışmasında 2. ve 4. yöntem diğerlerinden daha tutarlı sonuçlar vermiştir [46].

Random forest budalama olmadan da işlem yapabilmektedir. Bu işlemi de yapabilmesi için CART algoritmasını kullandığı görülmektedir. CART algoritmasında, bir düğümde belirli bir ölçüt uygulanarak bölünme işlemi gerçekleştirilir. Bunun için önce tüm niteliklerin var olduğu değerler göz önüne alınır ve tüm eşleşmelerden sonra iki bölünme elde edilir. Daha sonraki adımda ise seçme işlemi uygulanmaktadır. Bu işlemler yapılırken homojen sınıflar olması tercih edilir. Düğüm homojenliğinin ölçümünde; Gini İndeks, Entropy, Yanlış Sınıflama Hatası, Gain Oranı Kriteri gibi ölçütler kullanılmaktadır. Random forest yöntemi içerisinde gini indeksi kullanılmaktadır. RO algoritmasının sonuç olarak avantajlarını söyleyecek olursak;

- RO ile uyum problemi ortadan kaldırılmaktadır.
- Sınıflandırmanın yanı sıra regresyon problemlerinde de RO algoritması kullanılır.
- Mevcut olan bir veri setinden en uyumlu olanı tanımlamak için kullanılacak bir yöntemdir.

4.1.6. SMO

SMO, matris depolaması gerekmeden, sayısal QP optimizasyon adımlarına ihtiyaç duymadan, SVM QP problemini hızlıca çözebilen bir algoritma türüdür. SMO algoritması her adımda en küçük optimizasyon problemini çözmeyi seçmektedir. Standart bir SVM QP probleminde en küçük optimizasyon probleminde iki Lagrange çarpanı bulunur. Bu Lagrange çarpanlarının ise lineer bir eşitlik kısıtlamasına uyması gerekmektedir. SMO, ortaklaşa optimize için her adımda iki lagrange çarpanını seçerek, en uygun değeri bulur ve SVM'yi en uygun değerleri yansıtabilecek şekilde yeniden gerçekleştirir. SMO'nun bu süreçteki avantajı iki Lagrange çarpanının çözümlenmesindeki analitik yapısıdır. Bu analitik yapı sayesinde sayısal QP optimizasyonu tamamen önlenmiş olabilmektedir.

Her ne kadar algoritma esnasında optimizasyon alt problemleri ortadan kalkmış olsa da her bir alt problem barındırdıkları hız seviyesiyle beraber, genel OP problemini çabucak çözerek ortadan kaldırabilmektedir. Ek matris depolama gerektirmeyen SMO sayesinde büyük SVM eğitim sorunları, kişisel bir bilgisayar veya iş istasyonunun hafızasına sığabilmektedir. Ayrıca SMO'da matris algoritması kullanılmadığından sayısal hassasiyet problemlerine karşı daha az duyarlı olduğu ifade edilebilir [47].

Sıralı Minimal Optimizasyon (SMO) hızlı bir şekilde herhangi bir ekstra matris depolama olmadan ve sayısal QP optimizasyonu tüm stepsat kullanmadan SVMQP sorunu çözebilen basit bir algoritma olarak adlandırılır. SMO barındırdığı yapısıyla en küçük optimizasyon sorununa karşılık her adımda çözüm aramaktadır. SMO, yakınsamayı sağlamak için osuna teoremi kullanılarak genel

QP sorununun QP alt problemlerine ayrılmasını sağlar. Diğer yöntemlerin aksine SMO her adımda olabildiğince en küçük problemi seçmeyi hedefler. Bu problem içerisinde iki adet Lagrange çarpanı bulunmaktadır. Bulunan bu çarpanlar lineer eşitlik kısıtına uygun olarak seçilmelidir. Böylece sayısal QP optimizasyonu tamamen önlenir. Algoritmanın iç döngüsü, bütün bir QP kitaplığı rutin olarak çağırmak yerine kısa bir C kodu cinsinden ifade edilebilir. Algoritma sürecinde daha fazla optimizasyon alt problem çözülsede, her alt problemin çok hızlı gerçekleşmesiyle genel QP problemi hızla çözülür. SMO algoritması matris algoritması kullanmadığından dolayı sayısal tahmin problemleri için duyarlılık seviyesi daha azdır [48].

4.1.7. J48 Algoritması

J48 algoritması, quinalan tarafından geliştirilen bir algoritma türüdür. Bu algoritma bir karar ağacı algoritması olmakla beraber doğrusal olmayan küçük boyutlu verileri sınıflandırmada kullanılır. Karar ağacı yaklaşımı sınıflandırma problemlerini çözmede önemlidir. Bu yöntemle sınıflandırma sürecini modellemek için bir ağaç oluşturulur. Ağaç oluşturulduktan sonra veri tabanındaki her veri grubuna uygulanarak sınıflandırma işlemi gerçekleşir. Ağaç oluşturulurken eksik değerleri göz ardı edilir. Böylece kalan diğer veriler kullanılarak tahmin işlemi gerçekleştirilir. J48 yönteminde temel fikir karar ağaçları tarafından üretilen kurallar kullanılarak sınıflandırma işlemi yapmaktır [49].

4.1.8. IB1 Algoritması

IB1 algoritması öklid mesafesini kullanır. Birden fazla örnek test örneğine sahipse bu set içerisinde en küçük mesafeye sahip olan değerlere bakılır. Bu değerler arasında ilk bulunan değer kullanılır. Oluşturulan bu küme içerisindeki örnekler y_i ile ifade edilir. X_i olarak ifade edilen kısım ise sınıflandırılmak istenen değerdir. Bu iki değer öklid hesabı ile değer bulunur. Formül hali aşağıda gösterildiği gibidir [50].

$$Similarity(x, y) = - \sqrt{\sum_{i=1}^n f(X_i, Y_i)}$$

4.1.9. Çok Katmanlı Yapay Sinir Ağları

Yapay Sinir Ağları(YSA) insan sinir sistemini taklit ederek öğrenmeyi hedefleyen denetimli bir makine öğrenmesi yöntemidir. Yapay sinir ağı modeli, bir katmanda bulunan

nöronların takip eden katmandaki nöronlara bağlanması ile oluşturulur. En sık kullanılan yapay sinir ağı modeli olan çok katmanlı algılayıcı sinir ağı (MLP) modeli; girdi katmanı, gizli katman ve çıktı katmanı olmak üzere üç farklı katmandan meydana gelmektedir. Verilerin okunduğu katman, girdi katmanı olarak kabul edilmektedir. Her bir nöron bir özelliği temsil ettiği için özellik sayısı kadar nöron içermektedir. Çıktı katmanı ise sınıfların belirlendiği katmandır.

Bu katman oluşturulan modele göre tek bir nöron içerebileceği gibi sınıf çeşidi sayısı kadar nöron da içerebilmektedir. Gizli katman ise girdi katmanı ile çıktı katmanı arasında yer alan verilerin ara işleme maruz kaldığı katmandır. Gizli katman sayısı ve bir gizli katmandaki nöron sayısı tam olarak belli olmamakla birlikte eğitimin kalitesini etkileyen önemli iki faktördür. MLP modelinde öğrenme bir önceki katmandan takip eden katmana doğru yapıldığı için ileri beslemeli YSA olarak da bilinir. Kullanılan eğitim algoritması hatanın karesini en aza indirecek şekilde ağırlıkları güncellemeyi hedefler [50].

4.2. Konuşma Sınıflandırmasında Sinir Ağları

Bu bölümde, zaman içinde yapay sinir ağı uygulayarak konuşma sınıflandırması üzerine yapılan öncü çalışmalar ele alınmaktadır. Farklı konuşma aksanlarının sınıflandırılması için bir model Chan ve ark tarafından önerildi [36]. Aksanları tanımak ve tanımlamak için çeşitli ANN modelleri burada önerilmiştir. ASR sisteminin birden fazla İngiliz aksanının bulunduğu bir ortamda aksanları tanımlamasına yardımcı olmak için uygulanabilir bir yaklaşım önerildi. Buradan hareketle üç modelden bahsedebilmek mümkündür. Uygulanan bu üç model rekabetçi öğrenme [37], geri yayılma [38] ve karşı yayılma [39] şeklinde literatürde yer alabilmektedir. Uygulama esnasında iki konuşmacı seçilerek iki kategori oluşturuldu. Ve aynı görev için üç sinir ağı tercih edildi. Bu sinir ağı uygulanırken giriş olarak yedi adet özellik kullanıldı. Bu yedi özelliğin ilk üç özelliğin ortalaması alınarak formant frekansı, ortalama zift frekansı, İngilizce konuşma süresinin yüzdesi ve İngilizce ilk tanıtıldığı yaş ve İngilizce konuşma yıllarının sayısı olarak belirlendi. Üç tür sinir ağı arasında, geri yayılım ANN üstün performansı sergiledi. Anadili ve anadili olmayan İngilizce konuşmacıları tanımak için ANN kullanımını gösterdiler. 1996 yılında, stres altında konuşmanın sinirsel ağ tabanlı bir model Hansen ve Womack tarafından önerildi [40]. Stres altında vokal yolun değişimi, kesit alanları, akustik tüpler ve spektral özellikler kullanılarak düşünülür. Bu varyasyonlardan hareketle, 11 tip dikkate alınarak bir sinir ağı tabanlı sınıflandırma modeli inşa edilmiştir. Bu sınıflandırmaları Kızgın, net, cond50, cond70, Lombard, hızlı, normal, yüksek sesli, soru, yavaş ve yumuşak olarak ifade edebilmek mümkündür. Sınıflandırıcı olarak tek bir telefon sınıf bölümü yani mono bölüm özellikleri ile eğitilmiş bir sinir ağı tercih edilmiştir. Sinir ağının iki ara katmanı aracılığıyla bireysel konuşma özelliği yayıldığı ifade edilebilir. Çıktı katmanında ise gerilme olasılık skoru tahmin edilmiştir. Stres altındaki Mevcut konuşma (SUSAs) veritabanı özellik kümesinin seçimi, stres sınıflandırma algoritma performansı için değerlendirmeye

sınıflandırma algoritması dahil edilmiştir. 2001 yılında, Polur ve ark tarafından sinir ağı tabanlı izole kelime konuşma tanıma sistemi önerildi [41].

4.3. Konuşmada Derin Sinir Ağları Tanıma ve Sınıflandırma

Sinir ağı iki katmandan fazla katman barındırarak, derin bir sinir ağı olarak bilinmektedir. Konuşmada yinelenen sinir ağı veya derin sinir ağı doğal dinamik özellikleri ASR sisteminde uygun bir alternatif olarak kabul edilebilir. Bu bölümde, DNN uygulamayla ilgili araştırma eğilimleri tartışılmaktadır.

Derin sinir ağı tarafından akustik modelleme için yeni bir metodoloji Mohamed ve diğerleri tarafından önerildi. 2012 yılında [42]. Gauss karışım modellerini derin sinir ağlarıyla değiştirdiler ve DNN'nin birkaç özellik katmanı ve çok sayıda parametre içerdiğinden daha iyi bir fonetik tanıma elde edilebileceğini gösterdiler. Bu ağlarda yapılan ön eğitim, bir ayırmacı bilgiyi kullanmadan, spektral özellik vektörlerinin bir penceresinin çok katmanlı bir jeneratif modelini kapsamaktadır.

Özellikleri tasarladıktan sonra, geri yayılım kullanılarak özelliklerin hafif ayarlanması yapıldı; bu nedenle, monofonik HMM durumları üzerinde olasılık dağılımının daha iyi tahmin edilebildi. 2013 yılında, hem konuşmacı hem de dil tanıma için derin sinir ağı uygulaması Richardson ve diğ. [43]. Bu araştırmanın ana amacı ve literatüre olan katkısı hem konuşmacı hem de dil tanıma açılarından ASR için eğitilmiş tek bir DNN'nin uygulanması analiz edebilmektir. 2015 yılında Rajanna ve ark tarafından derin öğrenme yöntemleri kullanılarak müzik türü sınıflandırması üzerine bir vaka çalışması yapıldı. [44]. Müzik türü için ise iki katmanlı bir sinir ağının olduğu keşfedildi. Onlar vb iyi bilinen geleneksel sınıflandırma teknikleri, örneğin SVM, regresyon ile DNN doğruluk oranı üzerinde karşılaştırmalı bir çalışma yaptık Derin sinir ağı kullanarak stresli konuşma için doğaçlama konuşma tanıma sistemi Dumpala ve Kopparapu [45] tarafından 2017 yılında önerildi. Hoparlör tanıma eğitim aşamasında nefes sesinin dahil edilmesinin etkisi analiz edilmiş ve son teknoloji Gauss karışım model-evrensel arka plan modeli (GMM-UBM) ve DNN tabanlı sistemler kullanılarak karşılaştırılmıştır. Ayrıca DNN tabanlı sistemin, GMM –UBM tabanlı sisteme göre öğrenme kabiliyeti açısından daha üstün olduğu gözlemlenmiştir.

4.4. Denetimli Öğrenme Yöntemleri Kullanan Ses Çalışmalarının İncelenmesi

Önerilen çalışma ile ilgili literatür araştırmaları aşağıda özetlenmiştir: Matthew ve diğerleri tarafından yapılan çalışmada temiz bir konuşma kaynağından büyük miktarlarda gürültülü konuşma verilerinin oluşturulmuştur. Oluşturulan bu sinir modelleri bir kez eğitildikten sonra, canlı bir ortama nasıl yerleştirilebilir sorusuna cevap aranmıştır. Bu modellerin

kapalı alan test ortamlarında iyi performans göstermesine rağmen, canlı dağıtım senaryosunun ilgili zorluklar sunduğu görülmüştür[2].

Muhammad Salman Ishaq yaptığı çalışmayı 3 kısıma ayırmıştır. Bu çalışmalarında, ilk olarak Fin akustik modeli incelenmiş ve eğitim verileri farklı depo ortamlarına kaydedilerek gerçek dünyadaki konuşmaları doğrulamak için kullanmışlar. İkincisi, sesin algılanması VAD kararları uygulanarak sınıflandırılmış ve konuşmaya dahil olup olmadığını belirlemişlerdir. Son olarak, OOV kelime dağarcığı için bir çöp modeli geliştirmişlerdir. Bu model, dilbilgisinin dışındaki kelimeleri tanır ve işaretler. Bu çalışma üzerinde bu üç görevin Fin ses veri tabanı ile başarısını değerlendirmişler ve kelime hata oranındaki genel gelişmeyi rapor etmişlerdir[3].

Simon ve diğerlerinin yaptığı çalışmada farklı konuşma özelliklerini hedefleyen birçok yerleşik VAD özelliğine ilişkin yapılandırılmış bir genel bakış sunulmuştur. Güç, uyumluluk veya modülasyon gibi sömürülen özelliklere göre özellikleri sınıflandırılması ve bazı özel özelliklerin performansını değerlendirilmesi yapılmıştır. Zamansal bağlamın önemi, farklı uygulamaların getirdiği gecikme kısıtlamaları ile ilişkili olarak tartışılmıştır. Bu çalışma ile , gelecek vaat eden VAD özelliklerini seçmek ve performans ile karmaşıklık arasında makul bir denge bulmak hedeflenmiştir [1].

Marco ve diğerlerinin yapmış olduğu çalışmada, derin sinir ağları olan konuşmacı tanıma problemlerini çözmek için günümüzün en etkili istatistiksel modellerinden birinin kullanılması hedeflenmiştir. Dolayısıyla, bu çalışmanın amacı, konuşmacıları belirlemek için derin sinir ağı modellerini uygulamak, bu yaklaşımın umut verici olup olmadığını göstermek ve etkinliğini kanıtlamaktır. Sonuçların diğer tekniklerle karşılaştırılması yapılmıştır[4].

Zhang çalışmasında parkinson hastalarını baz almıştır. Parkinson hastalarının mPower uygulamasıyla toplanan ses verileri üç gruba ayırarak, İlaçların alındığı ilk esnada (en iyi durumda), ilaçlar alınmadan önce (en kötü durumda) ve ilaç dozları arasında bir yerde(ne en iyi, ne de en kötü durumda) gruplandırmayı çalışmasına dâhil etmektedir. Amacı sesi doğru bir sınıflandırıcı olarak doğrulamaktır. Yöntem olarak: Veri ön işleme, lojistik regresyon, destek vektör makineleri (SVM) sonrası, model için karar ağaçları, Gaussian Naïve Bayes ve Multi-layer Perceptron (MLP) uygulamıştır. Sonuç olarak ses ile Parkinson Hastalığı arasında bağlantı olduğunu göstermiştir [5].

Sangwook ve diğerleri üç açıdan çeşitli yöntem araştırmışlardır. Bunlar: özellik çıkarma, özellik dönüşümü ve nihai karar için puan füzyonu. Özellik çıkarma bölümünde, etkili akustik sahne sınıflandırması için çeşitli özellikler incelemişlerdir. Sorunu, çözmek için ise, sınıflandırma uygulamışlardır. Sınıflandırma için farklı özellik setlerine dayalı çeşitli sistemler geliştirmişlerdir. Nihai sonuç, bireysel sistemleri eriterek belirlenmiş. Yöntem gösterilerek doğrulanması yapılmıştır [6].

Gustavo ve diğerlerinin yaptığı çalışmada DCASE 2016 meydan okumasının veritabanını kullanarak otomatik bir sınıflandırma sisteminin performansı incelenmiş. Bu 15 farklı kategori göz

önünde bulundurularak ses örneğinin ham halinin kaydedildiği –ki bu orijinal halini gösterir- ortamlar tanınmıştır. Numunenin ses sinyali bir spektrograma dönüştürülmüş ve sonra doku tanımlayıcıları kullanılarak özellikler çıkarılmış. Stereo olarak kaydedilen dijital sesin sağ ve sol kanallarından gelen sinyaller arasındaki tamamlayıcı etken araştırma konusu olmuştur. Akustik ve görsel özelliklerin kullanıldığı kombinasyonlar da test edilmiş ve sınıflandırma hızında iyileşme sağlanmış. Sonuçta, DCASE 2016 meydan okumasının kendisinin sağladığı orijinal taban çizgisini aşarak, %77,2'lik bir doğruluğa ulaşmışlardır. Daha sonra, görsel alanda elde edilen özelliklerin bu uygulamada verimli bir sınıflandırma sisteminin geliştirilmesini destekleyebileceği sonucuna varmışlardır[7].

Zixing ve diğerlerine göre otomatik horlama sesi sınıflandırmasının (ASSC) geliştirilmesini ciddi şekilde engelleyen sınır konularından biri, yeterli denetimli eğitim verilerinin eksikliğiyle ilişkilendirilmiş.. Bu sorunla birlikte, rastgele bir gürültü alanından orijinal veri dağıtımına otomatik olarak haritalama stratejisini öğrenmeyi amaçlayan yarı denetimli koşullu Generative Ad-versarial Networks (scGAN) dayalı yeni bir veri artırma yaklaşımı önerilmiş.. Önerilen yaklaşım, 'gerçekçi' yüksek boyutlu verileri iyi sentezleme yeteneğine sahip, ek açıklama işlemi gerektirirken. GAN'ların mod çöküşü sorununu ele almak için, oluşturulan verilerin çeşitliliğini artırmak için bir topluluk stratejisi daha da tanıtılmış. Yaygın olarak kullanılan bir Munich-Passau horlama korpus üzerinde yapılan sistematik deneyler, scGAN tabanlı sistemlerin diğer klasik veri büyütme sistemlerinden daha iyi performans gösterdiğini ve aynı zamanda son zamanlarda bildirilen diğer sistemlere rekabetçi olduğunu göstermişlerdir [8].

Esmailpour ve diğerlerine göre küresel K-Means++ algoritması aracılığıyla denetimsiz özellik öğrenimini içeren yeni bir çevresel ses sınıflandırma yaklaşımı ve üst düzey veri büyütme için yeni bir mimari önerilmiştir. Ses sinyali, ayrı bir dalgacık dönüşümü (DWT) kullanılarak 2D gösterime dönüştürülmüş. DWT spektrogramları daha sonra döngü tutarlı üretken rakip ağı için yeni bir mimari ile artırılmış.. Bu üst düzey büyütme önyüklemesi, yapısal özellikleri numuneden numuneye çevirerek hem içi hem de sınıflar arası davranışlarda spektrogramlar üretmiş. Bir kod kitabı, DWT spektrogramlarını hızlandırılmış güçlü özellik dedektörü ve K-Means++ algoritması ile kodlayarak oluşturulmuşlardır[9].

Jamal ve diğerleri bu çalışmalarında ses etkinliği algılama (VAD), support vektör makineleri (SVM) kullanarak iki sınıf sınıflandırma problemi olarak formüle etmişler.. Önerilen yöntem, konuşma/konuşma dışı sınıflandırma için farklı arka zemin seslerinde eğitilmiş SVM modelleri ile birlikte gürültülü sağlam konuşma işleme özelliği ayıklama işlemini birleştirmedir. Çok sınıflı bir SVM, VAD için SVM modelini tercih etmek üzere arka plan seslerini sınıflandırmak için de kullanıldığı söylenebilir. Önerilen VAD'ler, diferantaditif gürültü türleri tarafından yapısal olarak bozulmuş TIMIT verileriyle test edilmiş ve state-of-the-art VAD'lerle karşılaştırılmıştır. Deneysel sonuçlar, önerilen VAD'nin zayıf SNR koşulları altında konuşma aktivitesini

ayıklayabileceğini ve aynı zamanda değişken gürültü seviyelerine karşı duyarsız olduğunu göstermişlerdir [10].

Zheng ve diğerleri bu makalede, güçlü ses etkinliği algılama (rVAD) için denetimsiz segment tabanlı bir yöntem sunmuşlardır. Yöntem, ses etkinliği algılama (VAD) aşamasının ardından iki ifade geçişinden oluşur. Bu ifade geçişlerinin ilkinde bir konuşma sinyalindeki yüksek enerjili segmentler posteriori sinyal-gürültü oranı (SNR) ağırlıklı enerji farkı kullanılarak tespit edilir ve bir segment içinde perde tespit edilmezse, segment yüksek enerjili bir gürültü segmenti olarak kabul edilir ve sıfıra uyarlanmaktadır. İkinci geçişte ise, konuşma sinyali, birkaç yöntemin araştırıldığı bir konuşma geliştirme yöntemi ile ifade edilir. Daha sonra, perdeye sahip komşu çerçeveler, zift segmentleri oluşturmak üzere ayrıştırılarak gruplandırılır ve konuşma istatistikleri dikkate alınarak hem sesli hem de sessiz sesleri ve muhtemelen de konuşma dışı bölümlerde katmak için perde segmentleri her iki uçtan daha da genişletilir. Sonucunda ise ses aktivitesinin tespiti için ifade edilen konuşma sinyalinin uzatılmış perde segmentlerine posteriori SNR ağırlıklı enerji farkı uygulanır. Önerilen yöntemin VAD performansını, çok çeşitli gürültü koşulları içeren iki veri tabanını, RATS ve Aurora-2 kullanarak değerlendirmişlerdir.. RVAD yöntemi, hoparlör doğrulama performansı açısından RedDots 2016 meydan okuma veritabanında ve gürültü bozuk sürümlerinde daha fazla değerlendirilmiş. Yapılan deney sonuçlarına göre rVAD'nin var olan mevcut yöntemle pozitif bir şekilde karşılaştırıldığını göstermektedir. İlaveten hesaplama açısından yoğun perde ayrıştırmasıyla verimli bir spektral düzlük hesaplamasıyla değiştirilen, rVAD'ın yeni değiştirilmiş versiyonunu sunmuşlardır. Bu değiştirilen sürüm, hesaplama karışıklığı rizki orta seviyelerdeki VAD performansı pahasına büyük oranda azaltmaktadır. Bu işlenen büyük veri miktarı düşük kaynak cihazlarındaki çalışma esnasında sunduğu bir avantajdır. RVAD'nin kaynak kodu herkese açık hale getirilmiştir [11].

Amir ve diğerleri yaptıkları çalışmada, gerçek hayat senaryolarında sıklıkla meydana gelen geçici ve sabit seslerin akustik ortamlarında ses etkinliği tespitini ele almışlardır. Temel geometrik yapılarını bağımsız olarak öğrenerek, konuşma ve konuşma dışı ses çerçevelerinin benzersiz mekânsal kalıplarından yararlanmışlar. Bu uygulama derin kodlayıcı kod çözücü sinir ağı yapısı kullanılarak yapılır. Bu yapı, spektral özellikleri zamansal bilgilerle eşleyen ve difüzyon haritaları yöntemi uygulanarak oluşturulan düşük boyutlu gösterimlerine eşleyen bir kodlayıcı içerir.

Kodlayıcı, sistemdeki gömülü verileri yüksek kapasiteli uzaya geri eşleme mantığıyla bir kod çözücü besler. Konuşma ve konuşma dışı ayrımı yapabilen derin bir sinir ağı, difüzyon ağlarına benzeyen yapıyla kod çözücü kodlayıcıya birleştirilerek elde edilir. Deney sonuçları, bir başka alanda ses etkinliği gösteren rakip ses etkinliği algılama sistemlerine kıyasla daha gelişmiş bir performans göstermektedir. Performans sonucu iyileştirme, doğruluk, sağlamlık ve genelleme kabiliyetlerine göre elde edilir. Modellerin ise gerçek zamanlı çalışarak ses tabanlı iletişim

sistemlerine uyarlanabilir hale getirilmiştir. Çevrimdışı uygulamalar için ise doğruluğu daha yüksek elde edilen toplu bir algoritma oluşturulmuştur [12].

Vishala ve diğerleri yapmış oldukları çalışmada, yeni bir yöntem açıklayarak, tek frekans filtreleme(SFF) analizi ve sinir ağı modeline dayanan konuşma etkinliği algılama (SAD) yönteminden bahsetmişlerdir. Her örnekleme anında konuşma sinyalinin an içerisindeki spektrumunu SFF, analizini vermektedir. Bu spektrum frekans arasındaki çözünürlük değeri, SFF analizinde kullanılan frekansların adetlerine göre karşılaştırılmaktadır. Bu karşılaştırma ise frekans aralığını kapsamaktadır. Örneğin 10 Hz'lik bir frekans aralığıyla 8 kHz'lik örnekleme frekansı kullanılarak, 0 – 4 kHz kapsayan 401 boyutlu bir spektrum her örnekleme anında elde edilerek sonuçlandırılır. Bu özellik ise ANN yani yapay sinir ağı modelini (gürültülü) Konuşma ve konuşma yapmamayı ayırt etmek için kullanılmaktadır. Bu kullanımın ise daha çok gürültüyü ayırt etmede etkin olduğu ifade edilebilir. Belirli bir test konuşması için eğitilmiş ANN modelinin çıktısı, her örnekleme anında konuşma/konuşma dışı karar verir. Kararın işlenmesi sonrası SAD için kullanılır. Sistem oluşturulan SAD, SAD görevi için Apollo korpus üzerinde tespit maliyet fonksiyonu (DCF) açısından değerlendirmiştir. Geliştirme ve değerlendirme veri kümeleri üzerinde önerilen sistemin DCF değerleri sırasıyla%3.1 ve%4,6, rapor edilen temel sistemin DCF değerleri sırasıyla%8.6 ve%11.7 olduğu görülmektedir [13].

Carlos ve diğerleri ilgili çalışmalarında uçtan uca işitsel ses tespiti üzerine çalışma yapmışlardır. Konuşma etkinliğinin algılama sistemleri (SAD) otomatik konuşmayı tanıma da (ASR) ve mevcut konuşma işleme sistemlerinde büyük bir rol üstlenmektedir. Konuşma etkinliğinin algılama sistemleri olarak bilinen SAD özellikle akustik gürültülü ortamlarda zor bir şekilde görev üstlenmektedir. Bu zorluğa pratik bir çözüm olarak ise, görsel bilgilerinin birleştirilmesiyle sunulmuştur. Böylelikle SAD yaklaşımının sağlamlığı arttırılabilmektedir. Görsel-ışitsel bir sistem, farklı konuşma modlarına (örn. fısıltı konuşması) veya arka plan gürültüsüne karşı sağlam olma avantajına sahiptir.

Derin öğrenmeyi kullanarak görsel-ışitsel konuşma işlemedeki son gelişmeler, akustik ve görsel özellikler arasındaki zamansal ilişkileri ilkel bir şekilde yakalamak için fırsatlar yaratmıştır. Bu çalışma, SAD için bimodal tekrarlayan sinir ağı (BRNN) çerçevesini dikkate alarak, literatüre katkı sağlamak üzere araştırılmaktadır. Yaklaşım, sıralı görsel-ışitsel verilerin zamansal dinamiğini modelleyerek önerilen SAD sisteminin doğruluğunu ve sağlamlığını iyileştirir. Yapılan çalışma daha çok, el yapımı özellikleri tahmin etmek yerine, akustik ve görsel özelliklerin eğitimi sırasında ham bilgilerden doğrudan öğrenilen uçtan uca bir eğitim yaklaşımını incelemeyi baz almaktadır. Bu deneysel değerlendirme sürecinde 105 hoparlörden toplanan 60.8 saatten fazla kayıt barındıran bir görsel işitsel süreç dikkate alınmıştır. Sonuçlar, önerilen çerçevenin yalnızca derin sinir ağı (DNN) ile uygulanan ses kullanılarak VAD taban çizgisi üzerinde pratik senaryolar altında%1,2'ye varan mutlak iyileştirmelere yol açtığını göstermiştir. Bu yaklaşım, gürültülü akustik ortam altında

taşınabilir bir tablettan sensörler kullanılarak değerlendirildiğinde, ideal koşullar altında elde edilen performanstan sadece% 1.0 daha düşük olan (örneğin, yüksek çözünürlüklü bir kamera ve yakın konuşan bir mikrofon ile elde edilen temiz konuşma) %92,7 F1 puanı elde etmiş ve sonuca ulaşmıştır [14].

Alexey ve diğerleri, konuşma ve konuşma dışı çerçevelerin kayda özgü Gauss karışım modellemesine (GMM) dayanan basit bir konuşma etkinliği dedektörü (SAD) önermişlerdir. Yarı denetimli öğrenmeyi kullanarak GMM eğitimi için geleneksel beklenti-maksimizasyon (EM) algoritmasını genişletmişlerdir. Etiksiz verileri SAD eğitim sürecine dahil etmek için bir metodoloji sağlar ve veri dağıtımının yapısından yararlanarak daha doğru istatistiksel modellere yol açar. İnsan yardımı gerektirebilecek çevrimdışı uygulamalara ya da metinlerden bağımsız olarak yapılan hoparlör doğrulması konuşmacının veya ses gözetimi gibi büyük orandaki ses verisinin işlenmesini içeren sistem uygulamalarında kullanılarak entegre edilir. Önerilen SAD, denetlenen SAD'lerin yaptığı gibi çevrimdışı eğitim verisi gerektirmediği görülmüştür. Daha ziyade, belirli bir ses kaydının küçük bir parçasından üretilen ilk etiketleri, daha basit bir SAD (veya bir insan operatörü) yardımıyla kullanır. Önerilen SAD sistemleri, farklı kovaryans türlerinde, konuşma ve konuşma dışı sınıf için başlatma yöntemleri tercih edilerek başlatma eylemini gerçekleştirecek etiketleme yapılmış veri miktarı ve konuşma özellikleri çözümlenmiştir. Bağımsız bir SAD sistemi ile yapılan deneylerde, son NIST OpenSAD değerlendirmesinden gelen zorlu veri kümesinde artan doğruluk Alexey ve diğerleri tarafından gözlenlenmiştir. NIST 2010 hoparlör tanıma değerlendirme (SRE) verileri ile metinden bağımsız deneyler ve RSR2015 ve RedDots corpora ile metin bağımlı deneyler de dahil olmak üzere kapsamlı otomatik hoparlör doğrulama (ASV) deneylerinde, sabit olmayan içeren uzun konuşma segmentleri için yeni yaklaşımın faydalarını göstermişlerdir. Metne bağımlı deney sonuçlarında daha kısa veri koşulları için basitlik düzeyi en altta olan denetimsiz SAD'ler daha iyi performans gösterecektir. Ayrıca bunoktada SAD izlemesi ve yanlış alarm noktasında NIST 2010 SRE verileri üzerindeki ASV performansına etkisi araştırılmıştır. İki SAD hatası sonucunda ampirik maliyet fonksiyonu elde edilerek, SAD yöntemi ve sinyal – gürültü oranı (SNR) hangi ölçüde olursa olsun ASV hata oranının aynı SAD çalışma noktası etrafında en az seviyeye ulaştığı gözlemlenmiştir. Optimum ASV performansı, yanlış bir şekilde dahil edilen konuşmamanın cevapsız konuşmalardan 4-5 kat daha pahalı olduğu SAD işletim bölgesinde yaklaşık olarak gerçekleşir. Önerilen yarı denetimli SAD, diğer zıt SAD yöntemlerine karşılık olarak SAD karar noktasına göre bağımlılığının daha az olduğu sonucu elde edilmiştir[15]. Inyoung ve diğerleri bu çalışmalarında, istatistiksel modele dayalı ses etkinliği algılama (VAD) için akustik ortam sınıflandırma (AEC) tekniği kullanarak derin sinir ağlarının (DNN) topluluğunu araştırmışlardır. İstatistiksel model tabanlı VAD araştırmasından, geleneksel karar kuralının, sıfır veya bir gizli katman ile sığ bir model olan olasılık oranının geometrik ortalamasına veya destek vektör makinesine (SVM) dayandığı bilinmektedir.

Sıg modeller, özelliklerin alan dağılımının çeşitliliğinden yararlanamadığından, eğitim adımıda temel olarak, istatistiksel model tabanlı VAD algoritmasının parametrelerini kullanarak farklı gürültü türlerine göre birden fazla DNN'yi oluşturmuştur. Buradan hareketle, ayrı DNN, her gürültü için en iyi DNN'yi tercih etmek için AEC algoritması için tasarlandığı belirtilmiştir.

Çevrimiçi gürültüye duyarlı VAD adımıda, AEC ilk olarak ayrı DNN kullanılarak kare bazında gerçekleştirilir, böylece gürültüyü tanımlama posteriori olasılıkları elde edilir. Her gürültü için olasılıklar elde edildikten sonra, çevresel bilgi, bireysel gürültü için eğitilmiş DNN'lerin topluluğundan elde edilen konuşma varlığı olasılıklarının birleştirilmesine izin verir. VAD yaklaşımlarında objektif önlemler noktasında değerlendirmeler yapılarak, konvansiyonel değerlendirme ve konvansiyonel algoritmanın verilerine göre önemli bir ilerleme gösterdiği sonucuna ulaşılmıştır [16].

Moattar ve HomayounpourSes ilgili çalışmalarında ses etkinlik algılamasından bahsetmişlerdir. Aktivite Algılama (VAD), tüm Konuşma ve Ses işleme uygulamalarında önem arz etmektedir. Çoğu konuşma/ses işleme yöntemlerinin tümü olmasa da performansı, Sesli Etkinlik Algılamanın performansına çok bağlıdır. İdeal bir ses etkinliği dedektörünün uygulama alanı ve gürültü durumundan bağımsız olması ve gerçek uygulamalarda en az parametre ayarına sahip olması gerekir. Bu çalışmalarında, daha önceki bazı yöntemlerle kıyaslayarak hem uygulanması kolay hem de gürültüye dayanıklı neredeyse ideal bir VAD algoritması önerdiklerini görmekteyiz. Önerdikleri yöntem,, Spektral Düzlük (SF) ve Kısa Vadeli Enerji gibi kısa vadeli özellikleri kullanır. Kullanılan bu yöntemin faydası ise çevrimiçi işleme görevleri için yardımcı olmasıdır. Önerilen yöntem, katkı maddesi gürültülü birkaç konuşma corpora üzerinde değerlendirmişler ve en son önerilen algoritmalarından bazılarıyla karşılaştırmışlardır. Ex- periments, çeşitli gürültü koşullarında tatmin edici performans gösterdiği görülmüştür [17].

Ahmet Ali ve arkadaşları yaptıkları çalışmada LSTM algoritmasını kullanarak vücut kitle indeksi hesaplamasını yapmışlardır. Obezite vücut yağ miktarının artması ile çeşitli sorunlara yol açan ve tedavi edilebilen bir hastalıktır. Artan vücut ağırlığı, kas iskelet sistemi üzerinde bozukluklara yol açabilmektedir. Ayak, vücut ağırlığının yere aktarılmasında görevli bir yapıdır. Obezitenin ayağın anatomik yapısını, taban temas yüzey alanını ve basınç değişimlerini etkilediği görülmüştür. Bu çalışmalarında Uzun Kısa-Vadeli Hafıza (LSTM) tabanlı derin sinir ağı (DSA) ile bireylerin ölçülen ayak tabanı basınç değerleri ile vücut kitle indeksi (VKİ) durumları sınıflandıran bir model geliştirmişlerdir. Öncelikle bireyin ayak tabanı ait 12 farklı bölgeye ait basınç değerleri baropodometre ile ölçülmüşt. Ölçüm sonucunda sinir ağının veri kümesinde kullanılacak 13 farklı giriş parametresi elde edilmiş. Bunun yanında her bireyin VKİ verisi hesaplanarak, bu verilere ait 4 gruptan birine yerleştirilmiş. Böylelikle modelin öğrenmesi için gerekli veri kümesi oluşturulmuş. Derin sinir ağında veri kümesi %80 eğitim %20 test olarak bölümlenmiş. Ağın hiper parametrelerin belirlenmesinden sonra eğitim ve test işlemleri gerçekleştirilmiş. Eğitilen modelin test sonucunda

%93.2 doğruluk elde edilmiş. Önerilen LSTM model, diğer sınıflandırma algoritmaları ile kıyaslanmış. Bunun sonucunda en yüksek performans elde etmiş. Sonuç olarak önerilen sınıflandırma modeli ile bireyin VKİ'sine göre belirlenmiş obezite durumunu, ayak basınç verilerinin sınıflandırılması ile doğrulanmış olduğunu göstermişlerdir. Ayrıca LSTM tabanlı derin sinir ağlarının, sınıflandırma işlemlerinde yüksek başarı performansı verdiğini gözlemlemişlerdir [55].

Sehgal ve Kehtarnavaz yaptıkları bu çalışmada, evrimsel sinir ağına dayalı gerçek zamanlı ses etkinliği algılama gerçekleştiren bir akıllı telefon uygulaması sunmuşlardır. Gerçek zamanlı olarak değerlendirilebilecek uygulama konuları, evrimsel sinir ağını ilişkisel olarak dikkate alarak yavaş çıkarım süresinin nasıl ele alındığını göstermektedir. Geliştirilen akıllı telefon uygulaması, işitme cihazlarının sinyal işleme boru hatlarında gürültüyü azaltmak için bir anahtar görevi gördüğü ifade edilebilir. Bu noktada yapılan uygulama gürültü tahmini ya da sınıflandırma noktasında gürültülü konuşma sinyallerinin sadece bahsedilen bölümlerinde yapılmasını sağlamaktadır. Geliştirilen akıllı telefon uygulaması, daha önce geliştirilmiş bir ses etkinliği algılama uygulaması ve iki yüksek sesli etkinlik algılama algoritması ile karşılaştırılmıştır. Deneysel sonuçlar, evrimsel sinir ağı kullanan geliştirilmiş uygulamanın daha önce geliştirilmiş akıllı telefon uygulamasından daha iyi performans gösterdiğini göstermektedir [56].

Ferroni ve arkadaşları derin sinir ağları yaklaşımı ile ses aktivite algılama işlemi yapmışlardır. Çok odalı yerli senaryolar için bir Ses Aktivite Dedektörü (VAD) gösterilmektedir. Çok odalı bir VAD (mVAD), bir konuşma segmentinin zaman sınırlarını eşzamanlı olarak algılar ve oluşturulduğu odayı belirler. Önerdikleri yaklaşım tamamen veri odaklı olup ve Derin İnanç Ağı (DBN) olarak önceden eğitilmiş ve standart bir hata geri yayılım yöntemi ile ince ayarlanmış bir Derin Sinir Ağına (DNN) dayandırılmaktadır. Sınıflandırmayı gerçekleştirmek için altı farklı özellik setleri çıkarmışlar ve birden fazla mikrofon sinyalinden birleştirmişlerdir. Bir evde kaydedilen büyük birçoklu mikrofon veri kümesi, ağ boyutu ve giriş mikrofonları seçimleri ile dönüşümlü olarak çoklu özellik seçim aşamalarını içeren çok aşamalı bir analiz stratejisi aracılığıyla performansı değerlendirmek için kullanmışlardır. Önermiş oldukları bu yaklaşım, ilk özellik seçim aşamasında ve ağ seçiminde alternatif algoritmalarından daha iyi performans gösterdiğini görmüşlerdir. Hassas geri çağırma eğrisi (AUC) altındaki alan açısından BLST-mVAD'ye mutlak artış oranı%5,55'ken, MLP-mVAD ile oranı%2,65'tir. Bu nedenle, yalnızca önerilen yaklaşım kalan seçim aşamalarından geçtiğini görmüşler. Özellikle, DBN-mVAD önemli gelişmeler elde eder: AUC ve F-ölçümü açısından mutlak artışlar DBN-mVAD'nin ilk aşamasına göre % 10.41 ve% 8.56'ya eşit olduğu görülmüştür [57].

Bu alıřmada, mevcut literatür alıřmalardan farklı olarak belirli ortamlarda ses kayıtları yapılması ve yeni gelen bir sesin algılanarak hangi ortama ait olduėunun sınıflandırılması hedeflenmektedir.



5. MATERYAL VE METOT

Bu bölümde, deneysel çalışmalar için kullanılan materyaller ve metotlar açıklanmıştır.

5.1. Materyal

Bu bölümde deneysel çalışmalar için kullanılan materyaller açıklanmaktadır. Kullanılan ses veri tabanı hakkında bilgilerden, sınıflandırmada kullanılan eğitim ses verisini oluşturmak için gerçekleştirilen işlemlerden ve sınıflandırma işleminin parametrelerinden bahsedilmektedir.

Bu çalışmada önerilen modelin sınıflandırma başarısını ölçmek için Google'nin sunmuş olduğu AudioSet kullanılmıştır. Google tarafından ses datası için belirli bir format hazırlanmıştır. Hazırlanan format csv dosyası halinde kullanıcılara sunulmuştur. Hazırlanan dosya youtube üzerinden bulunan videolara ait özellikleri içermektedir. Youtube kimliği, videoda sesin başlangıç zamanı, videoda sesin bitiş zamanı, sesin ait olduğu etiket id'si şeklinde csv dosyasında yer almaktadır. Seslerin ait olduğu etiket için ise ayrı bir csv dosyası yer almaktadır. Etiket başlığının bulunması için, ses setine ait olan dosyada bulunan id alanı ile etiket dosyası içindeki id alanları eşleştirilerek sesin ait olduğu etiket bulunmuştur.

Her csv dosyası, her satırı “#” ile başlayan ve ilk iki satır oluşturulması zamanını ve genel istatistikleri gösteren üç satırlık bir başlığa sahiptir.

```
# Segments csv oluşturuldu
```

```
# num_id = 20371, num_segments = 20371, num_unique_label = 527, num_positive_label = 51804
```

Sonraki her satırda üçüncü başlık satırıyla tanımlanan sütunlar vardır:

```
# YTID, başlangıç_saniye, bitiş_saniye, pozitif_etiketler
```

Örneğin:

```
-0RWZT-aXcV 320.000, 330.000, "/ m / 03v3yw, / m / 0k4j"
```

Başta bulunan -0RWZT-aXcV you tube üzerinde bulunan videonun id değeridir. 320.000 değeri video içerisinde bulunan sesin başlangıç zamanı, 330.000 ise sesin bitiş zamanıdır. /m/03v3yw (Anahtar) ve /m/0k4j (Araba)değerleri ise bu sesin ait olduğu etiket sınıfının id alanlarıdır. Bu data setinin anlamlı hale gelmesi için gerekli işlemlerden geçirilmiştir. Shell komutları kullanılarak terminal kodu yazılmıştır. Bu shell kodu setin bulunduğu csv dosyasını okumuştur. Okuduktan sonra you tube üzerinde bulunan id'sine göre youtube_dll kütüphanesi kullanılarak video indirme işlemi yapılmıştır.

Video local sunucuya indirildikten sonra ffmpeg toolunun local sunucuya kurulumu sağlanmıştır. Setin dahil olduğu csv üzerinde indirilen video üzerinde bağlı bulunan sesin başlangıç

ve bitiş zamanları shell kodu tarafından okunarak ffmpeg toolu yardımı ile video kesme işlemi yapılmıştır. Belirtilen zaman aralığında bir video haline gelmiş olan bu dosyayı çıktısı ses olabilmesi adına .wav şeklinde kaydı yapılarak işlem başarı ile gerçekleştirilmiştir. Tüm bu işlemler yapılırken etiket dosyası da okunarak ses grubunun ait olduğu etiket ismi şeklinde klasörler oluşturulmuştur. Shell kodları işlem sonucunda bu ses dosyasının hangi etikete ait olduğunu bularak ilgili etiketli klasör içerisine id'si ile .wav şeklinde kaydını yaparak işlemi sonlandırmıştır.

Kullanılan bu set dosyasında, aynı kriterlerle seçilen farklı videolardan 22.716 video ve bunların ait olduğu 59 etiket sınıfı bulunmaktadır. Bu videoların bazılarının telif hakkı ve youtube üzerinden kaldırıldığı için ise işlemleri olmamıştır. Audio Set ontolojik olarak incelendiğinde bir hiyerarşi içerisinde düzenlenmiş ses etkinliklerinin oluşturduğu koleksiyon olarak tanımlanabilir. Ontoloji ise, insan ve hayvan seslerinden, doğal ve çevredeki etkin seslere aynı zamanda müzik ve diğer çeşitli seslere kadar çeşitli günlük seslerden oluşmaktadır. Aşağıda sunulan AudioSet grubuna ait örnek bir ontolojisi şeması gösterilmiştir.

İnsan sesleri

- İnsan sesi
- ıslık
- Solunum sesleri
- İnsan hareketi
- Sindirim
- Eller
- Kalp sesleri, kalp atışı
- Otoakustik emisyon
- İnsan grubu eylemleri

Kaynak belirsiz sesler

- Genel etki sesleri
- Yüzey teması
- Deforme edilebilir kabuk
- yansıma
- Sessizlik
- Diğer kaynaksız

Şekil 5.1 AudioSet Ontolojisi

Yukarıda şekil 5.1’de de görüldüğü gibi Audio Set ontolojik olarak incelendiğinde bir hiyerarşi içerisinde düzenlendiği görülmektedir. Yukarıda sunulan AudioSet grubuna ait örnek bir ontolojisi şeması gösterilmiştir.

Bu sette bulunan bütün seslerin örnekleme frekansı (sampling frequency) 22.050’dir. Literatürde bu değer 44.100 dür. Örnekleme frekansınca sesin çözünürlüğü diyebiliriz. Bu değer sıklığı ne kadar az olursa sesin kulağa hitap etme netliği de o kadar artar. Örnek verecek olursak, 10 saniyelik bir ses kaydının 1 saniyesinde 22.050 tane değer vardır. Fakat literatürde bu 44.100 dür. Bu değer arttığı takdirde sesin kalitesi de düşmektedir.

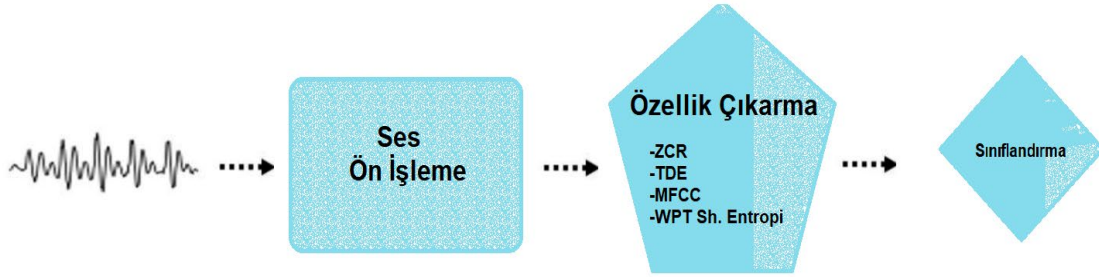
Bu çalışmada 2 farklı data set oluşturulup sonuçlar alınmıştır. 1. data seti hibrit bir data seti olarak hazırlanmıştır. Bu data setinde, hava, alarm, ambulans, bebek, zil, bot, helikopter, motor, müzik ve konuşma ses grupları olmak üzere 10 adetten oluşan bir data set hazırlanmıştır. Yukarıda bahsedilen Google AudioSet dosyaları indirilerek ses grupları oluşturulmuştur. Bu ses grupları içinde 10 farklı grup seçilmiştir. Seçilen bu ses gruplarında hand craft (el işi) işlemi yapılmıştır. Bu işlem yapılırken PRAAT programı tarafından hazırlanmıştır. PRAAT programına bir ses grubuna ait tüm ses dosyaları .wav halinde dahil edilmiştir. Bu ses dosyaları tek tek açılarak düzenlenmiştir. İlgili ses grubuna ait sesin geçtiği aralıklar belirlenmiş ve o aralık kesilerek farklı bir .wav dosyası olarak kayıt yapılmıştır. O ses grubuna ait işlemler tekrarlanarak birden fazla .wav dosyası halinde sesler oluşmuştur. PRAAT programı kullanılarak oluşturulan bu sesler birleştirme işlemine tabi tutulmuştur. Birleştirme işleminden sonra 50 sn ile 100 sn arasında değişen ses dosyaları oluşturulmuştur.

Aynı işlemler 2. data seti içinde yapılmıştır. 2. data setinde, alarm, blender, kedi, bulaşık, köpek, tıraş makinası, kızartma, akan su, konuşma ve elektrik süpürgesinden olmak üzere 10 farklı ses grubu yer almaktadır. Bu ses grubunun 1. data setinden farkı, farklı sesler olması ve 50 sn olarak olmasıdır. Bu data setinde yer alan tüm sesler PRAAT programı sonucunda 50 sn olarak sabit bir uzunlukta birleştirilmiştir. Bu data seti DCASE 2018’ de Task 4 grubundan yer alan çalışmadan ilham alınarak seçilmiştir. Bu görevde önerilen sesler bu ses grubunda yer almaktadır.

İlk data setinin ikinci data setinden farkları, ilk data setinde 50 sn. ile 100 sn. arasında değişken olan zaman aralıklarına sahip seslerden oluşurken, ikinci ses grubunda sabit 50 saniyeden oluşan ses grupları yer almaktadır. İlk ve ikinci data setleri içinde yer alan ses grupları birbirinden bağımsızdır ve ikinci data seti DCASE 2018’de önerilen seslerden ilham alınarak seçim işlemi yapılmıştır. İlk data seti bu durumda dengesiz (unbalanced) iken ikinci data seti dengeli (balanced) bir ses grubudur. Konuşma işleminin tüm algoritmaları MATLAB 2021a ortamında oluşmaktadır. Sınıflandırma işlemlerinde “MATLAB Classification Learner” aracı olarak tercih edilmektedir. Bu uygulamayı kullanarak, çeşitli sınıflandırıcılarında dikkate alınmasıyla denetimli makine öğrenimindeki verilere ulaşılabilmektedir. Aynı zamanda özellikler seçilebilir, doğrulama şeması belirlenebilir, oluşturulan modeller eğitilerek çıkan sonuçlar değerlendirilmeye alınabilmektedir.

5.2. Metot

Bu çalışmada, sese özellik çıkarımlarında bulunmak için farklı işlemlerden geçmiştir. Öncelikle ses alınarak ön işlemlere tabi tutuldu. Ön işlemlerin ardından ses üzerinde özellik çıkarma yöntemleri uygulanarak sınıflandırmaya tabi tutuldu.



Şekil 5.2 Önerilen Denetimli Öğrenme Yöntemi Kullanılarak Gerçekleştirilen Ses Tespit Sistemi

Yukarı da Şekil 5.2’de görüldüğü gibi ses üzerinde ön işlemler yapıldıktan sonra aşağıda bahsedilen özellik çıkarma aşamaları uygulanmıştır.

5.2.1. Özellik Çıkarma Yöntemleri

5.2.2. MFCC

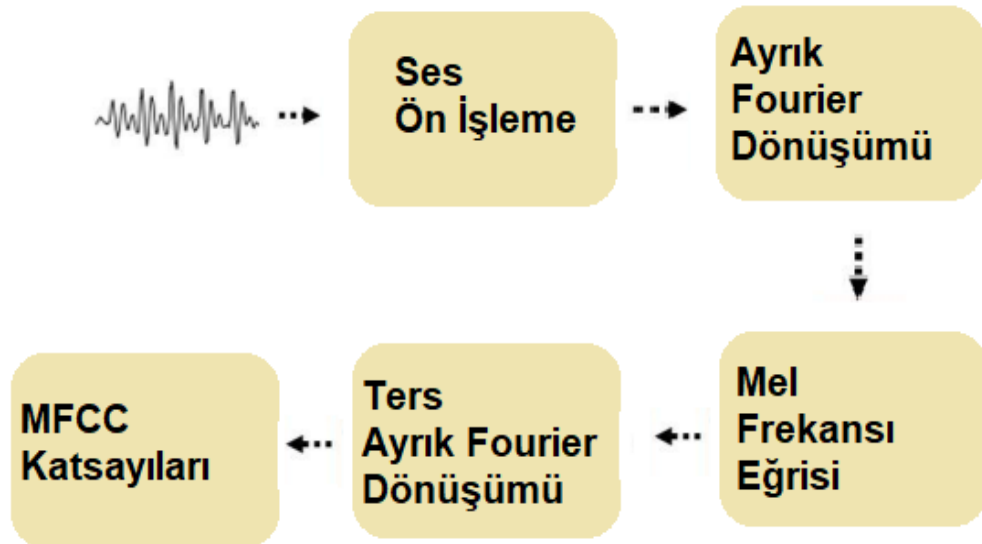
Akustik özellik olarak konuşma işleme alanında en çok kullanılan özellik çıkarma yöntemlerinden biri olan Mel Frekansı Kepstrum Katsayılarından faydalanılmıştır. Doğrusal olmayan sistemlerden olarak kabul edilen başka bir sınıfa homomorfik sistemler denilmektedir. Bu sistem doğrusal sistemlerin özel bir durumuna dahildir. Homomorfik sistemi bir formüle ile ifade edecek olursak, $a(f) = b(f).c(f)$ olarak ifade edebiliriz. Burada $a(f)$ olan kısım sesli ifadeyi, $b(f)$ olan kısım gırtlak, yani sesin gittiği yol ve $c(f)$ olan kısım ise sesin sinyalini kısaca sesi temsil etmektedir. Cepstrum değerlerinin hesaplanması için aşağıdaki eşitlik kullanılır.

$$c(n) = \frac{1}{N_s} \sum_{k=0}^{N_s-1} \log_{10} |S(k)| e^{(2n|N_s)kn} \quad 0 \leq n \leq N_s - 1$$

Formüle bakıldığı zaman $c(n)$ cepstrum olarak bilinmektedir. Formülü incelemeye devam edecek olursak $S(k)$ ise cepstrum değerinin ait olduğu frekans aralığında fourier dönüşümünü belirtir. Çerçevenin boyutunu ise N_s ile ifade edildiği görülmektedir. Görüldüğü üzere öncelikle spektrumun logaritması alınıp ters fourier dönüşümü yapılır. Bunun sebebi ise seslerin gürültüden

ayırt edilmesi içindir. Bu işlemler sonucunda kepstrum değerleri fourier dönüşümünden türetilmiş olurlar. Bu işlemler sonucunda kepstrum değerleri hesaplanmış olur. Hesaplanan bu değerler sesleri tanımda kullanılır. Mel frekansı özellikleri için var olan pencerelere bazı işlemler uygulanır. Bu işlem adımları şöyledir.

- Ses ön işleme alınarak başlanır.
- Ayırık Fourier Dönüşümü uygulanır.
- Ayırık fourier dönüşümü ile mel ölçeği uygulanarak çıkan genlik frekans sonucuna göre işlem yapılır
- Logaritmik enerji hesaplama işlemi yapılır.
- MFCC katsayılarının bulunma işlemi yapılır.



Şekil 5.3 MFCC özellik çıkarımının akış diyagramı

Şekil 5.3’de görüldüğü gibi Mel frekansı özellikleri için var olan pencerelere bazı işlemler uygulandığı şema olarak gösterilmektedir.

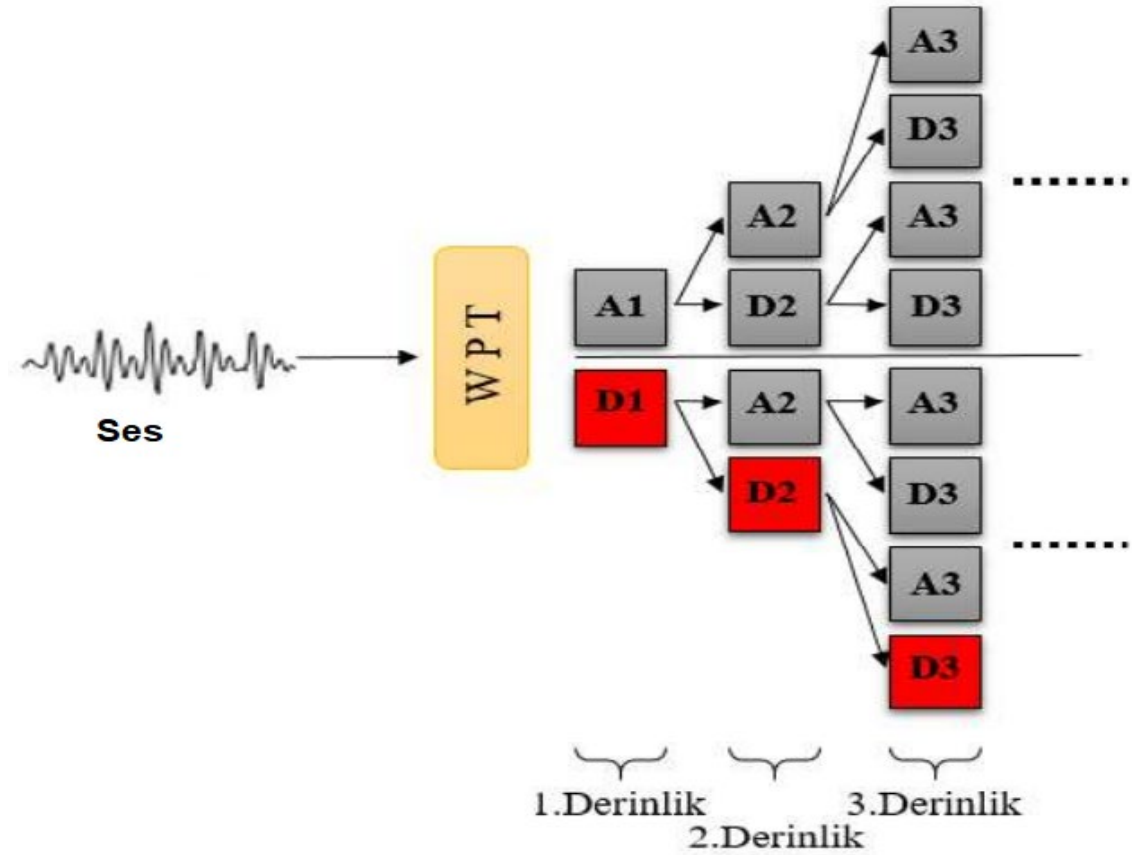
Mel Cepstrum değerlerini hesaplamak için Mel filtre bankası yöntemi aşağıdaki formül ile uygulanır. Mel Filtre bankasının amacı duyma mekanizmasının kritik bant filtrelerini simülasyonunu yapmaktır. Filtreler Mel Skalasında düzenli bir şekilde yerleşen üçgensel filtrelerdir. Üçgensel filtre çıkışları $Y(i)$; $i = 1; \dots; M$ logaritma kullanarak sıkıştırılmış ve ayırık kosinüs dönüşümü uygulanmıştır.

$$melc(n) = \sum_{i=0}^M \log Y(i) \cos \left[\frac{\pi n}{M} \left(i - \frac{1}{2} \right) \right]$$

Bu fonksiyonun sonucunda elde edilen katsayılar ile sesimizin artık yapay sinir ağına girebilecek şekle yani işlenebilecek şekle getirildiğini söyleyebiliriz [58].

5.2.3. Wavelet Packet Transform Shannon Entropi

Özellik çıkarma adımının ikinci kısmında sinyalin ayrışimsal özelliği olarak Dalgacı Paket Dönüşümü Shannon Entropi (Wavelet Packet Transform-WPT Shannon Entropy) kullanılmıştır. Dalgacık dönüşümü sinyali düşük ve yüksek frekanslı alt sinyallere ayıran bir 33 matematiksel dönüşüm olarak bilinmektedir. WPT, sinyali zaman ve ölçek bağlamında değerlendirdiğinden konvansiyonel Ayırık fourier dönüşümü, hızlı fourier dönüşümü ve kısa zamanlı fourier dönüşümü gibi dönüşümlere göre daha çok tercih edilmektedir. Bu çalışmada, ayrışimsal özellik çıkarımı için sinyali art arda yaklaşım (approximation) ve detay (detail) alt sinyal parçalarına bölen “daubechies” filtreli 5.derinlikte WPT kullanılmıştır



Şekil 5.4 WPT'nin ayrışimsal yapısı.[59]

Şekil 5.4'de görüldüğü gibi kullanılan WPT'ye ait ayrışimsal yapı yukarıdaki şekilde gösterilmiştir.

5.2.4. Zero Crossing Rate

Sinyalin sifıra yakınlama oranına denir. Bir sinyalin sıfır geişlerinin yüksek veya düşük olması 32 o sinyalin frekans domeninde bir özelliğini tanımlamaktadır. Sıfır geiş oranını biraz daha açacak olursak, ses sinyalinin sıfır değeri üzerinden kaç defa getiğı ile hesaplanmaktadır. Sıfır geişin meydana geldiğı hız, bir sinyalin frekans içeriğinin basit bir ölçüsü olarak bilinir.

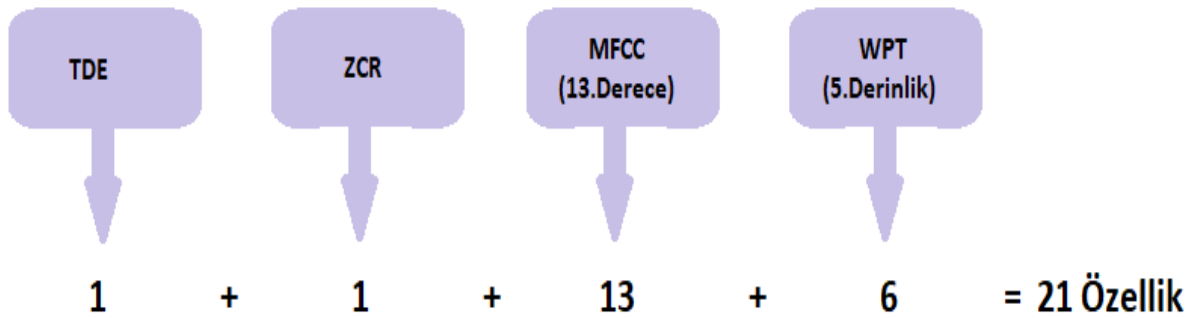
$$ZCR = \frac{\text{Sinyalin sıfırı geme miktarı}}{\text{Çerçeveedeki örnek sayısı}}$$

5.2.5. Time Domain Energy

Ses sinyalinin kısa zamanlı enerji (Time Domain Energy-TDE) değeri sinyal genliğindeki değışimleri yansıtmaktadır. Ses sinyalinin enerji değeri sinyalin konuşma, gürültü ve sessizlik sınıflarından hangisine ait olduğunu tespit etme noktasında önemli ipuçları içermektedir.

$$E = \sum_{k=1}^n |x(k)|^2$$

Bu çalışmada özellik çıkarma aşamasından sonra incelenen her ses verisine ait 1'i TDE, 1'i ZCR, 13'ü 13.dereceden MFCC ve 6'sı 5.derinlikten WPT çıkarılmış olmak üzere toplam 1x21'lik özellik vektörü elde edilmiştir. Çerçeveleme kısmı 20 ms olarak alınmıştır. 10 ms örtüşme ile ses üzerinde gezilmiştir. 50 saniyelik data setleri üzerinde bu işlem yapılmıştır. Aşağıdaki şekil 5.5 aracılığıyla ses datasına ilişkin özellik gruplarının bilgisi gösterilmektedir.



Şekil 5.5 Ses Datasına Ait Özellik Grupları

6. BULGULAR VE TARTIŞMA

Bu bölümde önerilen denetimli öğrenme yöntemiyle bir ses olay tespit metodunun sonuçları verilmektedir. Konuşma aktivite tespit sisteminin ses ön işleme, özellik çıkarma ve sınıflandırma aşamalarının tamamı MATLAB 2021a yüklü Intel Core i7 (2.6 GHz, 8 çekirdek) işlemcisi barındıran ve 8 GB DDR3 RAM'e sahip bir laptopta gerçekleştirilmiştir. Özellik çıkarma aşamasından sonra elde edilen sonuçlar verilmiştir. Birinci data set ve ikinci data set için uygulanan algoritmalar ve bulguları yer almaktadır. Birinci data setine uygulanan algoritmalar ve bulguları aşağıda yer almaktadır.

6 farklı SVM yöntemi kullanılarak yapılan sınıflandırma sonuçları aşağıda yer almaktadır. En yüksek başarımların Quadratic SVM'de olduğu görülmektedir. 21 tane özellik baz alınarak yapılan işlemlerde farklı ses aralıklarına sahip birinci data setinin değerleri tablo 6.1'de görüldüğü gibidir.

SVM Türü	Özellik	Sonuç
Linear SVM	21/21	66.7%
Quadratic SVM	21/21	92.9%
Cubic SVM	21/21	90.8%
Fine Gaussian SVM	21/21	85.1%
Medium Gaussian SVM	21/21	92.8%
Coarse Gaussian SVM	21/21	73.7%

Tablo 6.1 SVM Sonuçları (Data Set 1)

SVM kullanılarak çıkan sonuçların matris değeri aşağıda yer almaktadır. 1. data sete ait olan 10 farklı ses grubunun sonuçları yer almaktadır. Ses grupları sırasıyla;

- Hava (5349), Alarm (7798), Ambulans (5873), Bebek (5079), Zil (5572), Bot (6067), Helikopter (5303), Motor (6043), Müzik (5928), Konuşma (5506)

Yukarıdaki sıra ile ses grubu ve içerisinde bulunan değer sayısı yer almaktadır. Şimdi şeklimiz üzerinden çıkan sonuçları yorumlayalım.

Model 1.9										
Tahmini Sınıf	1	2	3	4	5	6	7	8	9	10
1	4932	18	1	13		87	169	32	85	12
2	13	7557	7	82	25	4	4	11	25	70
3	2	5	5787	30	8	16		5	14	6
4	35	104	53	4720	70	8	5	13	7	64
5	3	17	3	34	5392	12	9	14	48	40
6	99		15	7	11	5497	195	59	148	36
7	87	4		1	3	125	4844	74	112	53
8	19	5	3	2	5	149	105	5576	148	31
9	107	6	11	14	12	212	181	132	5181	72
10	72	70	14	50	39	130	81	41	161	4848
Gerçek Sınıf	1	2	3	4	5	6	7	8	9	10

Şekil 6.1 SVM Sınıflandırıcısına Ait Karışıklık Matrisi (Data Set 1)

SVM (Quadratic) ile yapılan sınıflandırma da çıkan sonuçlar şekil 6.1 de görülmektedir. Boş olan kutucuklar sınıflandırma da dahil olmayanlar olarak görülmektedir. Mavi olarak doğrusal giden değerler ait olduğu sınıf içerisinde bulunduğu miktar sayısıdır. N*M şeklinde bulunan matriste ses sınıflarının birbiri içerisindeki değerleri görülmektedir. Örnek verecek olursak, 1*1 matrisi incelendiğinde, 1. Sınıfın hava ses grubuna ait olduğu ses grubudur. Verilen hava sınıfı içerisinde hava olarak algıladığı değer sayısı 4932 olarak çıkmıştır.

1*2 matrisi incelendiğinde, 1. Sınıfın hava, 2.sınıfın alarm olduğu bilinmektedir. Hava sınıfı içerisinde 18 tane değeri alarm sesi olarak algılamıştır.

Şeklinde değerler gitmektedir. Genel bir sonuç çıkarımında bulunacak olursak;

N*M matrisinde ses grubunun kendi ses grubuna aitlik değeri ve diğer ses gruplarının kendi içerisinde bulunduğu değerleri görülmektedir. SVM kullanılarak çıkarılan özellik grupları ve bunlara ait değerler yukarıda görülmekteydi. Farklılıkların çıkmasının sebebi data set üzerinde bulunan ses gruplarının saniye farklılıkları ve bir saniye başına düşen seslerin yoğunluk oranından kaynaklanmaktadır. Şimdi Neural Network sınıflandırması kullanılarak çıkan sonuçları inceleyeceğiz. 5 farklı NN yöntemi kullanılarak yapılan sınıflandırma sonuçları aşağıda yer almaktadır. En yüksek başarı oranının Wide NN olduğu görülmektedir. 21 tane özellik baz alınarak yapılan işlemlerde farklı ses aralıklarına sahip birinci data setinin değerleri tabloda görüldüğü gibidir.

Neural Network Türü	Özellik	Sonuç
Narrow NN	21/21	74.0%
Medium NN	21/21	89.0%
Wide NN	21/21	94.3%
Bilayered NN	21/21	78.2%
Trilayered NN	21/21	79.6%

Tablo 6.2 Neural Network Sonuçları (Data Set 1)

Tablo 6.2'deki 1. data seti kullanılarak Neural Network sınıflandırmasına ait çıkan sonuçların matris değeri aşağıda yer almaktadır.

		Model 1.27									
Tahmini Sınıf	1	5056	8	1	18	8	34	90	25	84	25
	2	6	7642	17	47	18	3	5	4	13	43
	3		15	5750	55	10	13		1	16	13
	4	22	58	42	4817	40	14	7	11	15	53
	5	9	15	8	36	5404	17	5	13	24	41
	6	55	2	13	11	19	5627	103	53	141	43
	7	70	2		3	5	96	4941	34	91	61
	8	12	11	2	5	17	37	33	5768	126	32
	9	110	16	17	24	34	173	94	126	5200	134
	10	41	62	14	56	41	55	69	41	156	4971
		1	2	3	4	5	6	7	8	9	10
		Gerçek Sınıf									

Şekil 6.2 NN Sınıflandırıcısına Ait Karışıklık Matrisi (Data Set 1)

Şekil 6.2'de Neural Network sınıflandırmasına ait çıkan sonuçların matris değerlerine ait data set 1 fonksiyonları yüzdelik olarak gösterimi yapılmıştır.

Model 1.27										
Tahmini Sınıf	1	2	3	4	5	6	7	8	9	10
	94.5%	0.1%	0.0%	0.3%	0.1%	0.6%	1.7%	0.5%	1.6%	0.5%
	0.1%	98.0%	0.2%	0.6%	0.2%	0.0%	0.1%	0.1%	0.2%	0.6%
		0.3%	97.9%	0.9%	0.2%	0.2%		0.0%	0.3%	0.2%
	0.4%	1.1%	0.8%	94.8%	0.8%	0.3%	0.1%	0.2%	0.3%	1.0%
	0.2%	0.3%	0.1%	0.6%	97.0%	0.3%	0.1%	0.2%	0.4%	0.7%
	0.9%	0.0%	0.2%	0.2%	0.3%	92.7%	1.7%	0.9%	2.3%	0.7%
	1.3%	0.0%		0.1%	0.1%	1.8%	93.2%	0.6%	1.7%	1.2%
	0.2%	0.2%	0.0%	0.1%	0.3%	0.6%	0.5%	95.4%	2.1%	0.5%
	1.9%	0.3%	0.3%	0.4%	0.6%	2.9%	1.6%	2.1%	87.7%	2.3%
	0.7%	1.1%	0.3%	1.0%	0.7%	1.0%	1.3%	0.7%	2.8%	90.3%
Gerçek Sınıf										
										TPR FNR
										94.5% 5.5%
										98.0% 2.0%
										97.9% 2.1%
										94.8% 5.2%
										97.0% 3.0%
										92.7% 7.3%
										93.2% 6.8%
										95.4% 4.6%
										87.7% 12.3%
										90.3% 9.7%

Şekil 6.3 NN Sınıflandırıcısına Ait Karışıklık Matrisinin Yüzdelik Gösterimi (Data Set 1)

Şekil 6.3’de Neural Network sınıflandırmasına ait çıkan sonuçların matris değerlerine ait yüzdelik olarak gösterimi yapılmıştır. TPR (True Pozitif Rate) olarak adlandırılan kısımda, bulunduğu sınıf baz alınıp gerçekten yine o sınıfta olduğunun değeridir. Başarılı sınıflandırma sonucudur. FNR (False Negatif Rate) olarak adlandırılan kısım ise, bulunduğu sınıf baz alınır fakat bu defa o sınıfta olmayan seslerin sınıflandırılmasıdır. Yukarıdaki tablo baz alındığında en yüksek sonucun 2. Ses grubunda çıktığı görülmektedir. 2. Ses grubumuzu alarm sesi oluşturmaktadır. 98.0% TPR değeri ile alarm olan ses grubunda gerçekten de %98’i alarm sesini oluşturmaktaymış. Geri kalan %2’lik kısmı ise alarm sesi olmayan ses sınıfları oluşturmaktadır.

İlk data set grubunun diğer bir sınıflandırma işlemi KNN ile yapılmıştır. KNN sınıflandırması kullanılarak çıkan sonuçları inceleyeceğiz. 6 farklı KNN yöntemi kullanılarak yapılan sınıflandırma sonuçları aşağıda yer almaktadır. En yüksek başarımların Weighted KNN olduğu görülmektedir. 21 tane özellik baz alınarak yapılan işlemlerde farklı ses aralıklarına sahip birinci data setinin değerleri tabloda görüldüğü gibidir.

KNN Türü	Özellik	Sonuç
Fine KNN	21/21	90.1%
Medium KNN	21/21	90.1%
Coarse KNN	21/21	84.2%
Cosine KNN	21/21	88.6%
Cubic KNN	21/21	88.7%
Weighted KNN	21/21	91.1%

Tablo 6.3 KNN Sonuçları (Data Set 1)

Tablo 6.3'deki 1. data seti kullanılarak KNN sınıflandırmasına ait çıkan sonuçların karışıklık matrisi ile gösterimi aşağıda yer almaktadır. Yukarıda bahsedilen N*M matris değerlendirmesi geçerlidir.

Model 1.14										
Tahmini Sınıf	1	2	3	4	5	6	7	8	9	10
1	4754	28	2	24	16	165	202	67	66	25
2	13	7717	6	13	2	11	7	9	11	9
3	2	1	5776	39	12	27		6	4	6
4	112	27	64	4570	40	67	77	59	42	21
5	27	4	7	12	5464	23	12	6	11	6
6	272		16	35	28	5094	295	132	170	25
7	201	1		8	1	158	4702	112	98	22
8	67	7	4	13	5	127	123	5567	110	20
9	229	5	8	27	24	299	365	273	4609	89
10	192	24	19	56	30	155	218	103	246	4463
Gerçek Sınıf	1	2	3	4	5	6	7	8	9	10

Şekil 6.4 KNN Sınıflandırıcısına Ait Karışıklık Matrisi (Data Set 1)

Şekil 6.4'de farklı ses aralıklarına sahip birinci data setinin değerleri ve karışıklık matrisleri yer almaktadır.

Tablo 6.4'deki data set olarak kullanılan 1. Data set grubuna son uygulanan sınıflandırma ensembledir. Ensemble sınıflandırması kullanılarak çıkan sonuçları inceleyeceğiz. 5 farklı Ensemble yöntemi kullanılarak yapılan sınıflandırma sonuçları aşağıda yer almaktadır. En yüksek başarımların Bagged Trees olduğu görülmektedir. 21 tane özellik baz alınarak yapılan

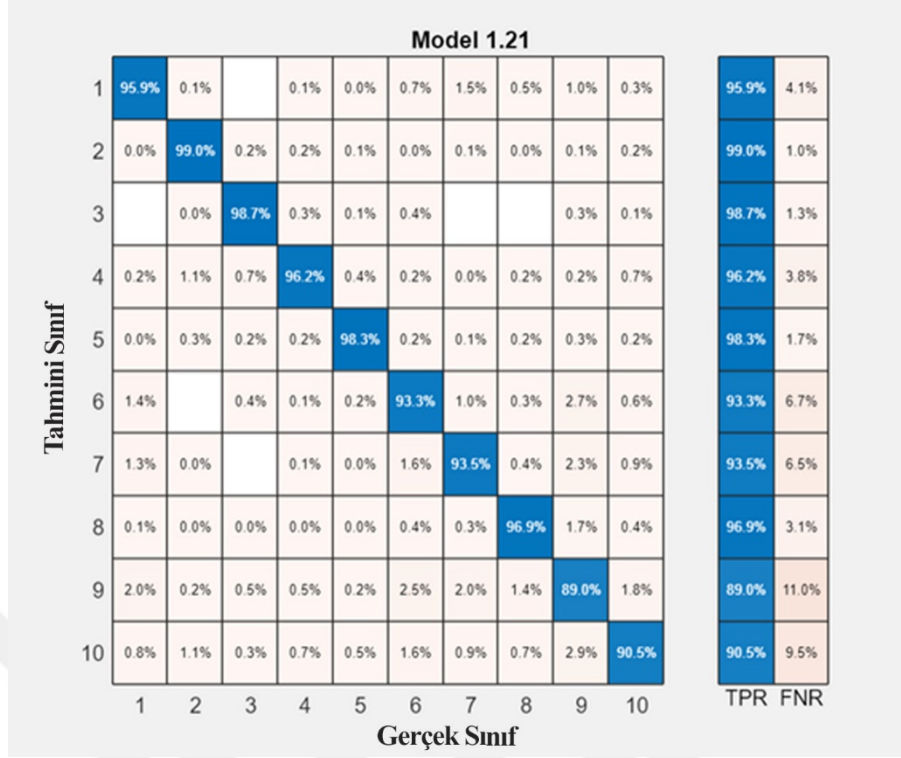
işlemlerde farklı ses aralıklarına sahip birinci data setinin değerleri ve karışıklık matrisleri aşağıda yer almaktadır.

Ensemble Türü	Özellik	Sonuç
Boosted Trees	21/21	57.2%
Bagged Trees	21/21	95.3%
Subspace Discriminant	21/21	44.3%
Subspace KNN	21/21	94.7%
RUSBoosted Tress	21/21	54.1%

Tablo 6.4 Ensemble Sonuçları (Data Set 1)

Model 1.21										
Tahmini Sınıf	1	2	3	4	5	6	7	8	9	10
1	5131	4		3	1	35	78	27	54	16
2	3	7721	16	16	7	3	5	1	10	16
3		2	5798	20	5	23			19	6
4	8	56	38	4887	22	8	1	10	12	37
5	1	14	11	12	5479	9	4	12	19	11
6	82		23	6	15	5663	61	16	162	39
7	67	1		3	1	86	4958	20	121	46
8	8	3	1	1	2	26	18	5853	105	26
9	117	11	29	31	9	147	121	83	5275	105
10	42	61	18	39	29	88	50	38	159	4982
Gerçek Sınıf	1	2	3	4	5	6	7	8	9	10

Şekil 6.5 Ensemble Sınıflandırıcısına Ait Karışıklık Matrisi (Data Set 1)



Şekil 6.6 Ensemble Sınıflandırıcısına Ait Karışıklık Matrisinin Yüzdelik Gösterimi (Data Set 1)

Şekil 6.5 ve 6.6’da Ensemble sınıflandırmasına ait çıkan sonuçların matris değerlerine ait yüzdelik olarak gösterimi yapılmıştır. TPR (True Pozitif Rate) olarak adlandırılan kısımda, bulunduğu sınıf baz alınıp gerçekten yine o sınıfta olduğunun değeridir. Başarılı sınıflandırma sonucudur. FNR (False Negatif Rate) olarak adlandırılan kısım ise, bulunduğu sınıf baz alınır fakat bu defa o sınıfta olmayan seslerin sınıflandırılmasıdır. Yukarıdaki tablo baz alındığında en yüksek sonucun 2. Ses grubunda çıktığı görülmektedir. 2. Ses grubumuzu alarm sesi oluşturmaktadır. 99.0% TPR değeri ile alarm olan ses grubunda gerçekten de %99’u alarm sesini oluşturmaktaymış. Geri kalan %1’lik kısmı ise alarm sesi olmayan ses sınıfları oluşturmaktadır. Böylelikle 1.data setine uygulanan 4 farklı sınıflandırma ve bu sınıflandırma da kullanılan farklı yöntemlerin işlem adımlarını incelemiş olduk. Çıkan sonuçların karışıklık matris ile incelemesini yaptık.

Yukarıdaki kısımda 1. data setine ait incelemelerde bulunduk. 1. Data setinde ses grupları arasında zaman farklılıkları, bir saniye üzerine düşen ses yoğunlukları ve birbirine benzeyen ses grupları olarak görülmesi bize 2. Hazırladığımız data seti için bir ilham kaynağı olmuştur. 2. Data setinde tüm ses grupları 50 sn olarak sabit tutulup, ses yoğunlukları artırılmıştır. Birbirinden bağımsız ses grupları da seçilerek başarı oranının artmış olduğu gözlemlenmiştir. Aynı sınıflandırma ve aynı yöntemler 2. Data setine de uygulanarak sonuçları aşağıda verilmiştir. Bu datasetine ait olan sınıf grupları sıra ve sayılarıyla şöyledir;

- Alarm (5100), Blender (5023), Kedi (5107), Bulaşık (5074), Köpek (5098), Traş Makinası (5045), Kızartma (5038), Su (5049), Konuşma (5001), Süpürge Makinası (5073)

Bu data setine ilk sette olduğu gibi 4 farklı sınıflandırma ve farklı yöntemleri kullanılarak bulgular elde edilmiştir. Sırasıyla bu sınıflandırma ve yöntemleri bulgularıyla inceleyelim. İlk olarak uygulanan SVM sınıflandırmasından bahsedelim.

SVM Türü	Özellik	Sonuç
Linear SVM	21/21	86.8%
Quadratic SVM	21/21	96.5%
Cubic SVM	21/21	97.4%
Fine Gaussian SVM	21/21	83.9%
Medium Gaussian SVM	21/21	96.3%
Coarse Gaussian SVM	21/21	88.0%

Tablo 6.5 SVM Sonuçları (Data Set 2)

6 farklı SVM yöntemi kullanılarak yapılan sınıflandırma sonuçları tablo 6.5’de görülmektedir. En yüksek başarımların Cubic SVM’de olduğu görülmektedir. 21 tane özellik baz alınarak yapılan işlemlerde farklı ses aralıklarına sahip ikinci data setinin değerleri tabloda görüldüğü gibidir. SVM kullanılarak çıkan sonuçların matris değeri aşağıda yer almaktadır. Karışıklık matrisi tablosunda ses sınıflarının kendi ses sınıfları içerisindeki değer sayıları ve kendisi dışında bulunan 9 tane ses sınıfındaki ses değerleri ile karşılaştırılmış değerleri görülmektedir. (Şekil 6.7.)

Model 1.10

1	4957		28	34	10	3		57	6	5
2		4999	1			1	3	9		10
3	14	2	4889	44	42	3	2	78	12	21
4	17	5	19	4831	28	23	7	107	17	20
5			19	49	4975	12		26	16	1
6		4	3	15	12	4990	17	4		
7				7	2	25	5001			3
8	18	15	15	76	6			4812	13	94
9	1		13	32	19			24	4912	
10		11	12	10		5		107	1	4927
	1	2	3	4	5	6	7	8	9	10

Tahmini Sınıf

Gerçek Sınıf

Şekil 6.7 SVM Sınıflandırıcısına Ait Karışıklık Matrisi (Data Set 2)

Model 1.10

1	97.2%		0.5%	0.7%	0.2%	0.1%		1.1%	0.1%	0.1%
2		99.5%	0.0%			0.0%	0.1%	0.2%		0.2%
3	0.3%	0.0%	95.7%	0.9%	0.8%	0.1%	0.0%	1.5%	0.2%	0.4%
4	0.3%	0.1%	0.4%	95.2%	0.6%	0.5%	0.1%	2.1%	0.3%	0.4%
5			0.4%	1.0%	97.6%	0.2%		0.5%	0.3%	0.0%
6		0.1%	0.1%	0.3%	0.2%	98.9%	0.3%	0.1%		
7				0.1%	0.0%	0.5%	99.3%			0.1%
8	0.4%	0.3%	0.3%	1.5%	0.1%			95.3%	0.3%	1.9%
9	0.0%		0.3%	0.6%	0.4%			0.5%	98.2%	
10		0.2%	0.2%	0.2%		0.1%		2.1%	0.0%	97.1%
	1	2	3	4	5	6	7	8	9	10

Tahmini Sınıf

Gerçek Sınıf

97.2%	2.8%
99.5%	0.5%
95.7%	4.3%
95.2%	4.8%
97.6%	2.4%
98.9%	1.1%
99.3%	0.7%
95.3%	4.7%
98.2%	1.8%
97.1%	2.9%

TPR FNR

Şekil 6.8 SVM Sınıflandırıcısına Ait Karışıklık Matrisinin Yüzdelik Gösterimi (Data Set 2)

Şekil 6.8’ de SVM sınıflandırmasına ait çıkan sonuçların matris değerlerine ait yüzdelik olarak gösterimi yapılmıştır. TPR (True Pozitif Rate) olarak adlandırılan kısımda, bulunduğu sınıf baz alınıp gerçekten yine o sınıfta olduğunun değeridir. Başarılı sınıflandırma sonucudur. FNR (False Negatif Rate) olarak adlandırılan kısım ise, bulunduğu sınıf baz alınır fakat bu defa o sınıfta olmayan seslerin sınıflandırılmasıdır. Yukarıdaki tablo baz alındığında en yüksek sonucun 2. Ses grubunda çıktığı görülmektedir. 2. Ses grubumuzu blender sesi oluşturmaktadır. 99.5% TPR değeri ile blender olan ses grubunda gerçekten de %99.5’i blender sesini oluşturmaktaymış. Geri kalan %0.5’lik kısmı ise blender sesi olmayan ses sınıfları oluşturmaktadır. Tablonun genel itibari ile boş olan sütunları ilk data setine oranla daha fazla olduğu görülmektedir. Bu da başarı oranının yüksek olduğunu ve hesaplanan ses sınıfının başka ses sınıfı olarak algılanmadığını göstermektedir. Şimdi Neural Network sınıflandırması kullanılarak çıkan sonuçları inceleyeceğiz. 5 farklı NN yöntemi kullanılarak yapılan sınıflandırma sonuçları aşağıda yer almaktadır (Tablo 6.6.). En yüksek başarı oranının Wide NN olduğu görülmektedir. 21 tane özellik baz alınarak yapılan işlemlerde farklı ses aralıklarına sahip ikinci data setinin değerleri tabloda görüldüğü gibidir.

Neural Network Türü	Özellik	Sonuç
Narrow NN	21/21	90.6%
Medium NN	21/21	96.0%
Wide NN	21/21	96.6%
Bilayered NN	21/21	91.2%
Trilayered NN	21/21	92.0%

Tablo 6.6 Neural Network Sonuçları (Data Set 2)

2. data seti kullanılarak Neural Network sınıflandırmasına ait çıkan sonuçların matris değeri aşağıda yer almaktadır. (Şekil 6.9.)

Model 1.19

1	4883	4	25	49	9	4	19	82	7	18
2	1	4990	1			2	11	4		14
3	17	11	4681	78	33	13	54	110	27	83
4	24	13	13	4502	14	76	49	289	8	86
5	1	2	51	92	4819	24	3	51	45	10
6	1	7	2	19	4	4944	40	21		7
7		1	1	2	1	45	4985			3
8	19	20	10	78		10	58	4617	9	228
9	1	1	8	107	18	2	2	90	4765	7
10	5	8	1	8		6	31	222		4792
	1	2	3	4	5	6	7	8	9	10

Gerçek Sınıf

Şekil 6.9 NN Sınıflandırıcısına Ait Karışıklık Matrisi (Data Set 2)

Model 1.27

1	97.0%	0.0%	1.0%	0.8%	0.2%	0.1%	0.0%	0.6%	0.1%	0.1%
2	0.0%	99.4%	0.1%	0.0%		0.1%	0.0%	0.1%		0.2%
3	1.1%	0.1%	94.6%	0.6%	0.6%	0.1%	0.0%	1.2%	0.9%	0.8%
4	0.7%	0.2%	0.8%	93.4%	1.3%	0.9%	0.1%	1.9%	0.6%	0.3%
5	0.3%		0.6%	1.2%	97.0%	0.4%	0.0%	0.2%	0.4%	
6	0.0%	0.2%	0.1%	0.8%	0.2%	97.9%	0.7%	0.1%		0.0%
7				0.2%	0.1%	0.7%	99.0%			0.0%
8	0.4%	0.2%	1.1%	1.6%	0.2%	0.1%		93.9%	0.5%	2.0%
9	0.3%		0.6%	0.7%	0.4%			0.5%	97.4%	0.0%
10	0.1%	0.2%	0.7%	0.2%	0.0%	0.0%	0.1%	2.1%	0.1%	96.6%
	1	2	3	4	5	6	7	8	9	10

Gerçek Sınıf

1	97.0%	3.0%
2	99.4%	0.6%
3	94.6%	5.4%
4	93.4%	6.6%
5	97.0%	3.0%
6	97.9%	2.1%
7	99.0%	1.0%
8	93.9%	6.1%
9	97.4%	2.6%
10	96.6%	3.4%
	TPR	FNR

Şekil 6.10 NN Sınıflandırıcısına Ait Karışıklık Matrisinin Yüzdelik Gösterimi (Data Set 2)

Yukarıda şekil 6.10’da Neural Network sınıflandırmasına ait çıkan sonuçların matris değerlerine ait yüzdelik olarak gösterimi yapılmıştır.(Şekil 6.10.) TPR (True Pozitif Rate) olarak adlandırılan kısımda, bulunduğu sınıf baz alınıp gerçekten yine o sınıfta olduğunun değeridir. Başarılı sınıflandırma sonucudur. FNR (False Negatif Rate) olarak adlandırılan kısım ise, bulunduğu sınıf baz alınır fakat bu defa o sınıfta olmayan seslerin sınıflandırılmasıdır. Yukarıdaki tablo baz alındığında en yüksek sonucun 2. Ses grubunda çıktığı görülmektedir. 2. Ses grubumuzu blender sesi oluşturmaktadır. 99.4% TPR değeri ile blender olan ses grubunda gerçekten de %99.4’ü blender sesini oluşturmaktaymış. Geri kalan %0.6’lık kısmı ise blender sesi olmayan ses sınıfları oluşturmaktadır.

İkinci data set grubunun diğer bir sınıflandırma işlemi KNN ile yapılmıştır. KNN sınıflandırması kullanılarak çıkan sonuçları inceleyeceğiz. 6 farklı KNN yöntemi kullanılarak yapılan sınıflandırma sonuçları aşağıda yer almaktadır. En yüksek başarımların Weighted KNN olduğu görülmektedir.

KNN Türü	Özellik	Sonuç
Fine KNN	21/21	93.9%
Medium KNN	21/21	94.4%
Coarse KNN	21/21	89.5%
Cosine KNN	21/21	93.8%
Cubic KNN	21/21	93.3%
Weighted KNN	21/21	94.8%

Tablo 6.7 KNN Sonuçları (Data Set 2)

21 tane özellik baz alınarak yapılan işlemlerde farklı ses aralıklarına sahip ikinci data setinin değerleri tablo 6.7’de görüldüğü gibidir.

2. data seti kullanılarak KNN sınıflandırmasına ait çıkan sonuçların karışıklık matris ile gösterimi aşağıda yer almaktadır. (Şekil 6.11.)

		Model 1.19									
Tahmini Sınıf	1	4883	4	25	49	9	4	19	82	7	18
	2	1	4990	1			2	11	4		14
	3	17	11	4681	78	33	13	54	110	27	83
	4	24	13	13	4502	14	76	49	289	8	86
	5	1	2	51	92	4819	24	3	51	45	10
	6	1	7	2	19	4	4944	40	21		7
	7		1	1	2	1	45	4985			3
	8	19	20	10	78		10	58	4617	9	228
	9	1	1	8	107	18	2	2	90	4765	7
	10	5	8	1	8		6	31	222		4792
		1	2	3	4	5	6	7	8	9	10
		Gerçek Sınıf									

Şekil 6.11 KNN Sınıflandırıcısına Ait Karışıklık Matrisi (Data Set 2)

		Model 1.19									
Tahmini Sınıf	1	95.7%	0.1%	0.5%	1.0%	0.2%	0.1%	0.4%	1.6%	0.1%	0.4%
	2	0.0%	99.3%	0.0%			0.0%	0.2%	0.1%		0.3%
	3	0.3%	0.2%	91.7%	1.5%	0.6%	0.3%	1.1%	2.2%	0.5%	1.6%
	4	0.5%	0.3%	0.3%	88.7%	0.3%	1.5%	1.0%	5.7%	0.2%	1.7%
	5	0.0%	0.0%	1.0%	1.8%	94.5%	0.5%	0.1%	1.0%	0.9%	0.2%
	6	0.0%	0.1%	0.0%	0.4%	0.1%	98.0%	0.8%	0.4%		0.1%
	7		0.0%	0.0%	0.0%	0.0%	0.9%	98.9%			0.1%
	8	0.4%	0.4%	0.2%	1.5%		0.2%	1.1%	91.4%	0.2%	4.5%
	9	0.0%	0.0%	0.2%	2.1%	0.4%	0.0%	0.0%	1.8%	95.3%	0.1%
	10	0.1%	0.2%	0.0%	0.2%		0.1%	0.6%	4.4%		94.5%
		1	2	3	4	5	6	7	8	9	10
		Gerçek Sınıf									

TPR	FNR
95.7%	4.3%
99.3%	0.7%
91.7%	8.3%
88.7%	11.3%
94.5%	5.5%
98.0%	2.0%
98.9%	1.1%
91.4%	8.6%
95.3%	4.7%
94.5%	5.5%

Şekil 6.12 KNN Sınıflandırıcısına Ait Karışıklık Matrisinin Yüzdelik Gösterimi (Data Set 2)

Şekil 6.12’de KNN sınıflandırmasına ait çıkan sonuçların matris değerlerine ait yüzdelik olarak gösterimi yapılmıştır. TPR olarak adlandırılan kısımda, bulunduğu sınıf baz alınıp gerçekten yine o sınıfta olduğunun değeridir. FNR (olarak adlandırılan kısım ise, bulunduğu sınıf baz alınır fakat bu defa o sınıfta olmayan seslerin sınıflandırılmasıdır. Yukarıdaki tablo baz alındığında en yüksek sonucun 2. Ses grubunda çıktığı görülmektedir. 2. Ses grubumuzu blender sesi oluşturmaktadır. 99.3% TPR değeri ile blender olan ses grubunda gerçekten de %99.7’si blender sesini oluşturmaktadır. Geri kalan %0.7’lik kısmı ise blender sesi olmayan ses sınıfları oluşturmaktadır.

Data set olarak kullanılan 2. Data set grubuna son uygulanan sınıflandırma ensembledir. 5 farklı Ensemble yöntemi kullanılarak yapılan sınıflandırma ve matris sonuçları tablo 6.8’de yer almaktadır.

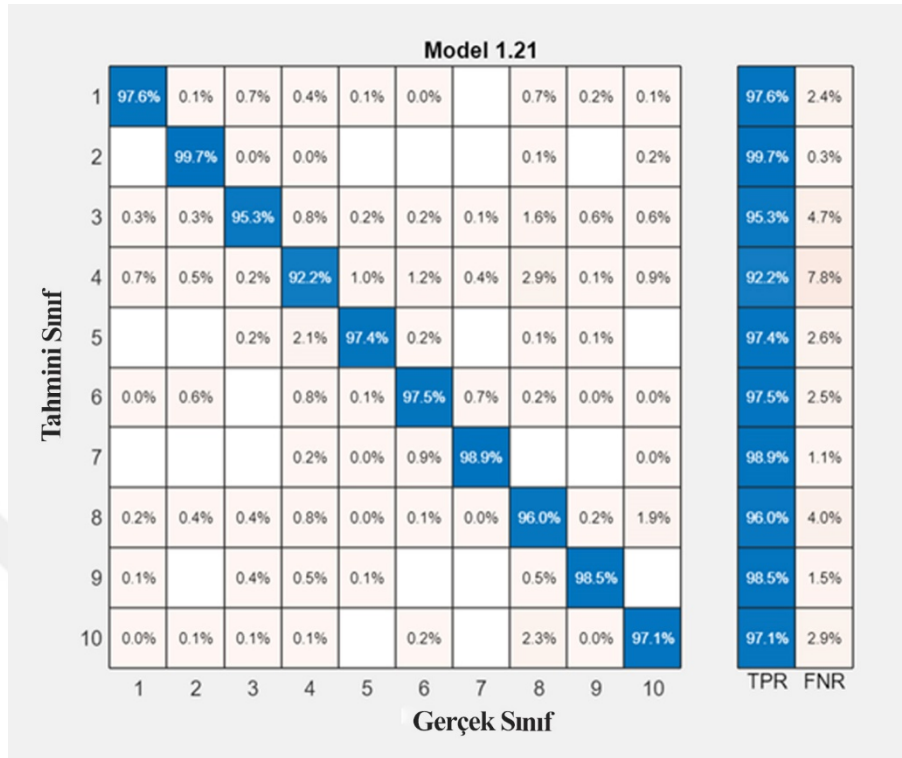
Ensemble Türü	Özellik	Sonuç
Boosted Trees	21/21	72.2%
Bagged Trees	21/21	97.0%
Subspace Discriminant	21/21	68.8%
Subspace KNN	21/21	96.9%
RUSBoosted Tress	21/21	64.7%

Tablo 6.8 Ensemble Sonuçları (Data Set 2)

Model 1.21										
Tahmini Sınıf	1	2	3	4	5	6	7	8	9	10
1	4976	5	36	21	7	1		38	9	7
2		5007	1	1				4		10
3	16	14	4869	43	8	11	3	81	32	30
4	34	24	11	4680	51	60	20	145	4	45
5			9	105	4966	9		6	3	
6	2	29		41	7	4921	34	9	1	1
7				10	2	44	4981			1
8	11	18	19	42	1	4	2	4846	11	95
9	3		19	23	4			27	4925	
10	1	6	6	5		8		119	1	4927
Gerçek Sınıf	1	2	3	4	5	6	7	8	9	10

Şekil 6.13 Ensemble Sınıflandırıcısına Ait Karışıklık Matrisi (Data Set 2)

Ensemble sınıflandırmasına ait çıkan sonuçların matris değerlerine ait yüzdelik olarak gösterimi şekil 6.13’de görüldüğü gibidir.



Şekil 6.14 Ensemble Sınıflandırıcısına Ait Karışıklık Matrisinin Yüzdelik Gösterimi (Data Set 2)

Şekil 6.14’de Ensemble sınıflandırmasına ait çıkan sonuçların matris değerlerine ait yüzdelik olarak gösterimi yapılmıştır. TPR (True Pozitif Rate) olarak adlandırılan kısımda, bulunduğu sınıf baz alınıp gerçekten yine o sınıfta olduğunun değeridir. Başarılı sınıflandırma sonucudur. FNR (False Negatif Rate) olarak adlandırılan kısım ise, bulunduğu sınıf baz alınır fakat bu defa o sınıfta olmayan seslerin sınıflandırılmasıdır. Yukarıdaki tablo baz alındığında en yüksek sonucun 2. Ses grubunda çıktığı görülmektedir. 2. Ses grubumuzu blender sesi oluşturmaktadır. 99.7% TPR değeri ile blender olan ses grubunda gerçekten de %99.7’si blender sesini oluşturmaktaymış. Geri kalan %0.3’lük kısmı ise blender sesi olmayan ses sınıfları oluşturmaktadır.

Böylelikle 2.data setine uygulanan 4 farklı sınıflandırma ve bu sınıflandırma da kullanılan farklı yöntemlerin işlem adımlarını incelemiş olduk. Çıkan sonuçların karışıklık matris ile incelemesini yaptık.

7. SONUÇLAR

Bu çalışmada, denetimli öğrenme yöntemleri kullanılarak ses olay tespit sisteminin gerçekleştirilmesi hedeflenmiştir. Önerilen model, ön işleme, özellik çıkarımı ve sınıflandırma olmak üzere 3 temel bölümden oluşmaktadır. Ön işleme aşamasında, bir yandan sınıflandırmada kullanılacak eğitim veri setinin hazırlanmış, diğer yandan ses kayıtlarına çerçeveleme, pencereleme ve ön vurgu filtresi gibi konuşma işlemede sıklıkla kullanılan yöntemler uygulanmıştır. Akustik özellik çıkarma aşamasında sıfır geçiş oranı (ZCR), zaman uzayında enerji (TDE), ve MFCC, ayrışimsal özellik çıkarma aşamasında ise dalgacık paket dönüşümü yöntemlerinden faydalanılmaktadır. Farklı iki data seti üzerinde yapılan çalışmalar da önerilen yapının birçok sınıflandırıcı ile yüksek sınıflandırma başarısına (%97.4) ulaştığı gözlemlenmiştir.

Farklı ses gruplarından oluşan 2 data set kullanılmıştır. Kullanılan ilk data setinde sabit bir zaman aralığı bulunmamaktadır. 50 saniye ile 100 saniye arasında değişkenlik gösteren 10 adet ses grubu bulunmaktaydı. Bu ses gruplarının yoğunluk seviyesi normal düzeyde olmaktaydı. İkinci data setinde ise sabit 50 saniyeden oluşan ve ses yoğunluğunun fazla olduğu 10 grup ses datası bulunmaktadır. Bu durumdan kaynaklı olarak ikinci data setinde alınan başarımlar oranları daha yüksek çıktığı görülmektedir. Her iki data setine, SVM, Neural Network, KNN ve Ensemble olmak üzere 4 sınıflandırma ve farklı yöntemleri kullanılmıştır. Toplamda MFCC, ZCR, TDE, ve WPT olmak üzere 21 adet özellik kullanılmıştır. Uygulanan bu yöntemler sonucunda, 1. Data setinde en yüksek başarımlar oranının Ensemble (Bagged) sınıflandırması (95.3%) olduğu görülmektedir. 2. Data setinde ise en yüksek başarımlar oranının SVM (Cubic) sınıflandırması (97.4) olduğu görülmektedir.

Gerçekleştirilen çalışma ile ses ve akustik konusunda gerekli araştırmalar yapılmıştır. İlgili literatür çalışmaları taranmıştır. Başka araştırmacılar tarafından gerçekleştirilen çalışmalar irdelenerek ilham alınmıştır. Ayrıca bu ses ve akustik veriler üzerinde gerçekleştirilen yapay zekâ, makine öğrenmesi araştırmaları yapılmıştır. Kullanılan ve kullanılması muhtemel olan denetimli öğrenme yöntemleri konusunda literatür çalışması yapılarak incelemelerde bulunulmuştur. Bu çalışmanın ortaya koyduğu en büyük avantaj, makine öğrenmesi yöntemleri kullanılarak akustik olay tespitinde kullanılacak olan akustik veri seti ve yöntem olarak da denetimli öğrenim yöntemleri konusunda yapılacak çalışmalara ilham olunmuştur.

KAYNAKLAR

- [1] Graf, S., Herbig , T., Buck, M. and Schmidt, G. (2015) Features for voice activity detection: a comparative analysis, *EURASIP Journal on Advances in Signal Processing*
- [2] McEachern, M. (2018) Neural Voice Activity Detection and its Practical Use, *Massachusetts Institute of Technology*
- [3] Muhammed, I. (2017) Voice Activity Detection and Garbage Modelling for a Mobile Automatic Speech Recognition Application
- [4] Do, B.N, (2015) Neural Networks for Automatic Speaker, Language and Sex Identification
- [5] Zhang, Q., (2018) Mpower Voice Activity Motiroring And Classification For Parkinson's Diagnosis
- [6] Analysis Acoustic Features for Acoustic Scene Classification and Score fusion of multi-classification systems applied to DCASE 2016 challenge(2016)
- [7] Felipe,G.Z., Costa,Y.M.G.,Helal, L.G., (2018), Acoustic Scene Classification Using Spectrograms
- [8] Zhang,Z., Han, J., Qian, K., Janott, C., Guo, Y., Schuller,B., (2019),Snore-GANs: Improving Automatic Snore Sound Classification with Synthesized Data
- [9] Esmaeilpour,M.,Cardinal, P., Koerich, A.L.,(2019),Unsupervised Feature Learning for Environmental Sound Classification Using Weighted Cycle-Consistent Generative Adversarial Network
- [10] Saeedi,J.S.,Ahadi, S.M.,Faez,K.,(2013), Robust voice activity detection directed by noise classification
- [11] Tan, Z.H., Sarkar,A., Dehak,N.,(2019), rVAD: An unsupervised segment-based robust voice activitydetection method
- [12] Ivry,A., Berdugo, B. and Cohen,I.,(2019), Voice Activity Detection for Transient Noisy Environment Based on Diffusion Nets
- [13] Pannala, V. And Yegnanarayana, B.,(2020), A neural network approach for speech activity detection for Apollo corpus
- [14] Tao, F., Busso, C.,(2018), End-to-end audiovisual speech activity detection with bimodal recurrent neural models
- [15] Sholokhov,A.,Sahidullah, M.,Kinnunen,T.,(2017), Semi-supervised speech activity detection with an application to automatic speaker verification
- [16] Hwang,I., MinPark, H., HyukChang, J., (2015), Ensemble of deep neural networks using acoustic environment classification for statistical model-based voice activity detection
- [17] Kovacevic, B., Milosavljevic, M.M., Veinović, M., Marković, (2017), Robust Digital Processing of Speech Signals
- [18] Gerçeker,M., Yorulmaz,İ.,Ural, A.,(2000) Ses ve Konuşma (+)
- [19] Fugoin, C.,(1992) Fonksiyonel disfoni. Ses ve Ses Bozuklukları Konferanslar Dizisi, Haseki Hastanesi, İstanbul.
- [20] Erkan, M., Aslan, T., (1992), Konuşma ve ses bozuklukları. Erciyes Tıp Dergisi Ek 1: 297-313.
- [21] Burad, E., Çağlar, Y.,(1972), Sesimiz DSİ Matbaası, ANKARA.

- [22] Mahendru,H.C.(2014), Quick review of human speech production mechanism. *International Journal of Engineering Research and Development*, 9(10), 48–54.
- [23] Goswami,S., Deka, P., Bardoloi, B., Dutta, D., Sarma, D., (2013), ZCR Based Identification of Voiced, Unvoiced and Silent Parts of Speech Signal in Presence of Background Noise
- [24] Vyas,J., Suthar, A., (2017), Study of Speech Recognition Technology and its Significance in Human-Machine Interface
- [25] Feraru, M., & Zbancioc, M. (2013). Speech emotion recognition for SROL database using weighted KNN algorithm. *In 2013 International Conference on Electronics, Computers and Artificial Intelligence (ECAI)* (pp. 1–4). IEEE
- [26] Russell, S., & Norvig, P. (2003). *Artificial intelligence: A modern approach* (2nd ed.). *Prentice Hall*. ISBN 978-0137903955. [1995].
- [27] Fu, Z., Lu, G., Ting, K. M., & Zhang, D. (2010). Learning Naïve Bayes classifiers for music classification and retrieval. *In 2010 20th International Conference on Pattern Recognition (ICPR)* (pp. 4589–4592). IEEE.
- [28] Sanchis, A., Juan, A., & Vidal, E. (2012). A word-based Naïve Bayes classifier for confidence estimation in speech recognition. *IEEE Transactions on Audio, Speech, and Language Processing*, 20(2), 565–574.
- [29] Bhakre, S. K., & Bang, A. (2016). Emotion recognition on the basis of audio signal using Naïve Bayes classifier. *In 2016 International Conference on Advances in Computing, Communications and Informatics (ICACCI)* (pp. 2363–2367). IEEE.
- [30] Güldal, H., (2018). Karar Ağacı Algoritmalarının Eğitsel Veriler Üzerindeki Performanslarının İncelenmesi. (syf 6-7)
- [31] Shen, P., Changjun, Z., & Chen, X. (2011). Automatic speech emotion recognition using support vector machine. *In 2011 International Conference on Electronic and Mechanical Engineering and Information Technology (EMEIT)* (Vol. 2, pp. 621–625). IEEE.
- [32] Mahmoodi, D., Marvi, H., Taghizadeh, M., Soleimani, A., Razzazi, F., & Mahmoodi, M. (2011). Age estimation based on speech features and support vector machine. *In 2011 3rd Computer Science and Electronic Engineering Conference (CEEC)*, (pp. 60–64). IEEE.
- [33] Matoušek, J., & Tihelka, D. (2013). SVM-based detection of misannotated words in read speech corpora. *In International Conference on Text, Speech and Dialogue* (pp. 457–464). Springer, Heidelberg.
- [34] Aida-zade, K., Xocayev, A., & Rustamov, S. (2016). Speech recognition using support vector machines. *In 2016 IEEE 10th International Conference on Application of Information and Communication Technologies (AICT)*, (pp. 1–4). IEEE.
- [35] Shi, W., & Fan, X. (2017). Speech classification based on cuckoo algorithm and support vector machines. *In 2017 2nd IEEE International Conference on Computational Intelligence and Applications (ICCIA)* (pp. 98–102). IEEE.

- [36] Chan, M. V., Feng, X., Heinen, J. A., & Niederjohn, R. J. (1994). Classification of speech accents with neural networks. *In 1994 IEEE International Conference on Neural Networks, IEEE World Congress on Computational Intelligence* (Vol. 7, pp. 4483–4486). IEEE
- [37] Kohonen, T. (2012). Self-organization and associative memory (Vol. 8). *Springer Science & Business Media, New York*.
- [38] Rumelhart, D. E., & McClelland, J. L. (1986). Parallel distributed processing: explorations in the microstructure of cognition. volume 1. foundations.
- [39] Hecht-Nielsen, R. (1990). Neurocomputing. Boston: Addison-Wesley.
- [40] Hansen, J. H., & Womack, B. D. (1996). Feature analysis and neural network-based classification of speech under stress. *IEEE Transactions on Speech and Audio Processing*, 4(4), 307–313.
- [41] Polur, P. D., Zhou, R., Yang, J., Adnani, F., & Hobson, R. S. (2001). Isolated speech recognition using artificial neural networks. *Virginia Commonwealth Univ Richmond School of Engineering*
- [42] Mohamed, A., Dahl, G.E., & Hinton, G. (2012). Acoustic modeling using deep belief networks. *IEEE Transactions on Audio, Speech, and Language Processing*, 20(1), 14–22.
- [43] Richardson, F., Reynolds, D., & Dehak, N. (2015). Deep neural network approaches to speaker and language recognition. *IEEE Signal Processing Letters*, 22(10), 1671–1675.
- [44] Rajanna, A. R., Aryafar, K., Shokoufandeh, A., & Ptucha, R. (2015). Deep neural networks: A case study for music genre classification. *In 2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA)* (pp. 655–660). IEEE.
- [45] Rajanna, A. R., Aryafar, K., Shokoufandeh, A., & Ptucha, R. (2015). Deep neural networks: A case study for music genre classification. *In 2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA)* (pp. 655–660). IEEE.
- [46] Ok,A., Akar,Ö., Gungor,O., (2011).. Rastgele Orman Sınıflandırma Yöntemi Yardımıyla Tarım Alanındaki Ürün Çeşitliliğinin Sınıflandırılması
- [47] İşletme Bilim Dergisi 2018 Cilt:6 Sayı:3 (Syf 15-16)
- [48] Platt, J., (1998). Sequential minimal optimization: A fast algorithm for training support vector machines.
- [49] Eraldemir, S.G., Arslan, M.T., Yıldırım, E., (2017). Comparison Of Random Forest And j48 Decision Tree Classifiers Using HHT Based Features In EEG. *International Advanced Researches & Engineering Congress-2017*
- [50] Nizam, H., Akın, S.S. Sosyal Medyada Makine Öğrenmesi ile Duygu Analizinde Dengeli ve Dengesiz Veri Setlerinin Performanslarının Karşılaştırılması
- [51] <https://tr.wikipedia.org/wiki/Akustik>.
- [52] Barchiesi, D., Dimitrios, D.G., Plumbley, M.D., Stowell, D., (2015). Acoustic Scene Classification: Classifying Environments From The Sounds They Produce
- [53] Yılmaz. Ö., Kartal, Y., Özer, Y.B., Ses Etiketleme
- [54] Korkmaz, Y., (2018). Türkçe’deki Ünlü Harflerin Formant Frekans Değerlerine Dayalı Adli Aksan Analizi Gerçekleştirimi
- [55] Süzen, A.A., Yıldız, S., Yılmaz, T., (2019). LSTM Tabanlı Derin Sinir Ağı ile Ayak Taban Basınç Verilerinden VKİ Durumlarının Sınıflandırılması. *BEÜ Fen Bilimleri Dergisi* 8 (4), 1392-1398.

- [56] Sehgal, A., Kehtarnavaz, N., (2018). A Convolutional Neural Network (CNN) Smartphone App for Real-Time Voice Activity Detection
- [57] Ferroni, G., Bonfigli, R., Principi, E., Squartini, S. and Piazza, F., (2015). A Deep Neural Network Approach for Voice Activity Detection in Multi-Room Domestic Scenarios
- [58] Yüzkollar, C. (2007). *Yapay sinir ağları kullanarak parmak izi ve konuşmacı tanıma* (Master's thesis, Sakarya Üniversitesi)
- [59] Korkmaz, Y., Boyacı, A.(2020). Unsupervised and supervised VAD systems using combination of time and frequency domain features. *Biomedical Signal Processing And Control*, 61, 102044.



EKLER

EK- 1: CIEA'2021 KONFERANSI KATILIM SERTİFİKASI



ÖZGEÇMİŞ

Hakan GÜLER

Category	Percentage	Percentage
Very good	100%	100%
Good	100%	100%
Fair	100%	100%
Poor	100%	100%
Very poor	100%	100%
Don't know	100%	100%
No answer	100%	100%
Refused to answer	100%	100%

[REDACTED]

- [REDACTED]
- [REDACTED]
- [REDACTED]

[REDACTED]

	(a)	(b)
1. Name of the reporting company:	[REDACTED]	[REDACTED]
2. Date of filing:	[REDACTED]	[REDACTED]
3. Title of the report:	[REDACTED]	[REDACTED]
4. Author(s) name:	[REDACTED]	[REDACTED]
5. Institution:	[REDACTED]	[REDACTED]
6. Country:	[REDACTED]	[REDACTED]
7. City:	[REDACTED]	[REDACTED]
8. State:	[REDACTED]	[REDACTED]
9. Zip code:	[REDACTED]	[REDACTED]
10. Telephone number:	[REDACTED]	[REDACTED]
11. E-mail address:	[REDACTED]	[REDACTED]
12. Fax number:	[REDACTED]	[REDACTED]
13. Web site:	[REDACTED]	[REDACTED]
14. Other contact information:	[REDACTED]	[REDACTED]
15. Summary of findings:	[REDACTED]	[REDACTED]
16. Conclusions:	[REDACTED]	[REDACTED]
17. Recommendations:	[REDACTED]	[REDACTED]
18. Acknowledgments:	[REDACTED]	[REDACTED]
19. References:	[REDACTED]	[REDACTED]
20. Appendix:	[REDACTED]	[REDACTED]
21. Notes:	[REDACTED]	[REDACTED]
22. Glossary:	[REDACTED]	[REDACTED]
23. Bibliography:	[REDACTED]	[REDACTED]
24. Index:	[REDACTED]	[REDACTED]
25. Table of contents:	[REDACTED]	[REDACTED]
26. Executive summary:	[REDACTED]	[REDACTED]
27. Introduction:	[REDACTED]	[REDACTED]
28. Methodology:	[REDACTED]	[REDACTED]
29. Results:	[REDACTED]	[REDACTED]
30. Discussion:	[REDACTED]	[REDACTED]
31. Conclusion:	[REDACTED]	[REDACTED]
32. References:	[REDACTED]	[REDACTED]
33. Appendix:	[REDACTED]	[REDACTED]
34. Notes:	[REDACTED]	[REDACTED]
35. Glossary:	[REDACTED]	[REDACTED]
36. Bibliography:	[REDACTED]	[REDACTED]
37. Index:	[REDACTED]	[REDACTED]
38. Table of contents:	[REDACTED]	[REDACTED]
39. Executive summary:	[REDACTED]	[REDACTED]
40. Introduction:	[REDACTED]	[REDACTED]
41. Methodology:	[REDACTED]	[REDACTED]
42. Results:	[REDACTED]	[REDACTED]
43. Discussion:	[REDACTED]	[REDACTED]
44. Conclusion:	[REDACTED]	[REDACTED]
45. References:	[REDACTED]	[REDACTED]
46. Appendix:	[REDACTED]	[REDACTED]
47. Notes:	[REDACTED]	[REDACTED]
48. Glossary:	[REDACTED]	[REDACTED]
49. Bibliography:	[REDACTED]	[REDACTED]
50. Index:	[REDACTED]	[REDACTED]
51. Table of contents:	[REDACTED]	[REDACTED]
52. Executive summary:	[REDACTED]	[REDACTED]
53. Introduction:	[REDACTED]	[REDACTED]
54. Methodology:	[REDACTED]	[REDACTED]
55. Results:	[REDACTED]	[REDACTED]
56. Discussion:	[REDACTED]	[REDACTED]
57. Conclusion:	[REDACTED]	[REDACTED]
58. References:	[REDACTED]	[REDACTED]
59. Appendix:	[REDACTED]	[REDACTED]
60. Notes:	[REDACTED]	[REDACTED]
61. Glossary:	[REDACTED]	[REDACTED]
62. Bibliography:	[REDACTED]	[REDACTED]
63. Index:	[REDACTED]	[REDACTED]
64. Table of contents:	[REDACTED]	[REDACTED]
65. Executive summary:	[REDACTED]	[REDACTED]
66. Introduction:	[REDACTED]	[REDACTED]
67. Methodology:	[REDACTED]	[REDACTED]
68. Results:	[REDACTED]	[REDACTED]
69. Discussion:	[REDACTED]	[REDACTED]
70. Conclusion:	[REDACTED]	[REDACTED]
71. References:	[REDACTED]	[REDACTED]
72. Appendix:	[REDACTED]	[REDACTED]
73. Notes:	[REDACTED]	[REDACTED]
74. Glossary:	[REDACTED]	[REDACTED]
75. Bibliography:	[REDACTED]	[REDACTED]
76. Index:	[REDACTED]	[REDACTED]
77. Table of contents:	[REDACTED]	[REDACTED]
78. Executive summary:	[REDACTED]	[REDACTED]
79. Introduction:	[REDACTED]	[REDACTED]
80. Methodology:	[REDACTED]	[REDACTED]
81. Results:	[REDACTED]	[REDACTED]
82. Discussion:	[REDACTED]	[REDACTED]
83. Conclusion:	[REDACTED]	[REDACTED]
84. References:	[REDACTED]	[REDACTED]
85. Appendix:	[REDACTED]	[REDACTED]
86. Notes:	[REDACTED]	[REDACTED]
87. Glossary:	[REDACTED]	[REDACTED]
88. Bibliography:	[REDACTED]	[REDACTED]
89. Index:	[REDACTED]	[REDACTED]
90. Table of contents:	[REDACTED]	[REDACTED]
91. Executive summary:	[REDACTED]	[REDACTED]
92. Introduction:	[REDACTED]	[REDACTED]
93. Methodology:	[REDACTED]	[REDACTED]
94. Results:	[REDACTED]	[REDACTED]
95. Discussion:	[REDACTED]	[REDACTED]
96. Conclusion:	[REDACTED]	[REDACTED]
97. References:	[REDACTED]	[REDACTED]
98. Appendix:	[REDACTED]	[REDACTED]
99. Notes:	[REDACTED]	[REDACTED]
100. Glossary:	[REDACTED]	[REDACTED]
101. Bibliography:	[REDACTED]	[REDACTED]
102. Index:	[REDACTED]	[REDACTED]
103. Table of contents:	[REDACTED]	[REDACTED]
104. Executive summary:	[REDACTED]	[REDACTED]
105. Introduction:	[REDACTED]	[REDACTED]
106. Methodology:	[REDACTED]	[REDACTED]
107. Results:	[REDACTED]	[REDACTED]
108. Discussion:	[REDACTED]	[REDACTED]
109. Conclusion:	[REDACTED]	[REDACTED]
110. References:	[REDACTED]	[REDACTED]
111. Appendix:	[REDACTED]	[REDACTED]
112. Notes:	[REDACTED]	[REDACTED]
113. Glossary:	[REDACTED]	[REDACTED]
114. Bibliography:	[REDACTED]	[REDACTED]
115. Index:	[REDACTED]	[REDACTED]
116. Table of contents:	[REDACTED]	[REDACTED]
117. Executive summary:	[REDACTED]	[REDACTED]
118. Introduction:	[REDACTED]	[REDACTED]
119. Methodology:	[REDACTED]	[REDACTED]
120. Results:	[REDACTED]	[REDACTED]
121. Discussion:		

[illegible]