**T.C.**

**YILDIZ TEKNİK ÜNİVERSİTESİ**

**FEN BİLİMLERİ ENSTİTÜSÜ**

# MAKİNE ÖĞRENMESİ YÖNTEMLERİ İLE KONUŞMA İŞARETLERİNDEN DUYGU ANALİZİ

**Ferhat AYGÜN**

YÜKSEK LİSANS TEZİ

Elektronik ve Haberleşme Mühendisliği Anabilim Dalı

Haberleşme Programı

Danışman

Doç. Dr. Bülent BOLAT

Ocak, 2024

**T.C.**

**YILDIZ TEKNİK ÜNİVERSİTESİ**

**FEN BİLİMLERİ ENSTİTÜSÜ**

# MAKİNE ÖĞRENMESİ YÖNTEMLERİ İLE KONUŞMA İŞARETLERİNDEN DUYGU ANALİZİ

Ferhat AYGÜN tarafından hazırlanan tez çalışması 29.01.2024 tarihinde aşağıdaki jüri tarafından Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü Elektronik ve Haberleşme Mühendisliği Anabilim Dalı, Haberleşme Programı **YÜKSEK LİSANS** olarak kabul edilmiştir.

Doç. Dr. Bülent BOLAT
Yıldız Teknik Üniversitesi
Danışman

**Jüri Üyeleri**

Doç. Dr. Bülent BOLAT, Danışman
Yıldız Teknik Üniversitesi _____

Dr. Öğr. Üyesi Murat TAŞKIRAN, Üye
Yıldız Teknik Üniversitesi _____

Dr. Öğr. Üyesi Nurullah ÇALIK, Üye
İstanbul Medeniyet Üniversitesi _____

Danışmanım Doç. Dr. Bülent BOLAT sorumluluğunda tarafımca hazırlanan "Makine Öğrenmesi Yöntemleri İle Konuşma İşaretlerinden Duygu Analizi" başlıklı çalışmada veri toplama ve veri kullanımında gerekli yasal izinleri aldığımı, diğer kaynaklardan aldığım bilgileri ana metin ve referanslarda eksiksiz gösterdiğimi, araştırma verilerine ve sonuçlarına ilişkin çarpıtma ve/veya sahtecilik yapmadığımı, çalışmam süresince bilimsel araştırma ve etik ilkelerine uygun davrandığımı beyan ederim. Beyanımın aksinin ispatı halinde her türlü yasal sonucu kabul ederim.

Ferhat AYGÜN

İmza

# TEŞEKKÜR

Tez çalışmalarım sırasında ilgi ve yardımlarından dolayı tez danışmanım Doç. Dr. Bülent BOLAT hocama, tez çalışmalarıma verdiği katkılarından dolayı iş arkadaşım Hakan GÜLEÇ'e ve desteklerinden dolayı aileme teşekkür ederim.

Ferhat AYGÜN

# İÇİNDEKİLER

**TEŞEKKÜR** iv
**SİMGE LİSTESİ** vii
**KISALTMA LİSTESİ** viii
**ŞEKİL LİSTESİ** ix
**TABLO LİSTESİ** x
**ÖZET** xi
**ABSTRACT** xiii
**1 GİRİŞ** 1
1.1 Literatür İncelemesi ..... 1
1.1.1 Daha Önce Yapılmış Çalışmalar 1
1.1.2 Yapılan Çalışmalarda Gözden Kaçan Noktalar 3
1.2 Tezin Önemi, Amacı ve İzlenilen Yol ..... 3
1.3 Tezin Organizasyonu ..... 4
**2 YÖNTEM** 6
2.1 Veri Setleri ve Blok Diyagram ..... 6
2.2 Öznitelikler ..... 8
2.2.1 Güç 10
2.2.2 Ses Şiddeti 11
2.2.3 Sıfır Geçiş Oranı 11
2.2.4 Seslilik Olasılığı 11
2.2.5 MFCC (Mel-Frequency Cepstral Coefficients) 11
2.2.6 LPC (Linear Predictive Coding) Katsayıları 12
2.2.7 F0 (Temel Frekans) 13
2.3 Makine Öğrenmesi ..... 15
2.3.1 Destek Vektör Makineleri 15
2.3.2 SMO (Sequential Minimal Optimization) 18
**3 DENEYSEL ÇALIŞMA** 19
3.1 Kurulum ..... 19

3.2 Daha Önce Yapılmış Çalışmaların Tekrarı ..... 19
3.3 Daha Önce Yapılmış Çalışmaların Çapraz Testleri ..... 20
3.4 Duygu Sınıfının Azaltılması ..... 22
3.5 Ses Formatlarının Çağrı Merkezi Formatına Dönüştürülmesi ..... 24
3.6 Ses Kayıtlarına Çağrı Merkezi Arka Plan Gürültüsü Eklenmesi ..... 25
3.7 MFCC Özniteliklerinin Ayrıca İncelenmesi ..... 26
**4 SONUÇ VE ÖNERİLER** **27**
4.1 Tez Çalışması Sonuçlarının Yorumu ..... 27
4.2 Daha Başarılı Sonuçlar İçin Öneriler ..... 27
**KAYNAKÇA** **29**
**TEZDEN ÜRETİLMİŞ YAYINLAR** **32**

# SİMGE LİSTESİ

| | |
|---|---|
| Khz | Kilohertz |

# KISALTMA LİSTESİ

| | |
|---|---|
| AMDF | Average Magnitude Difference Function |
| CNN | Convolutional Neural Network |
| CPU | Central Processing Unit |
| DC | Direct Current |
| DFT | Discrete Fourier Transform |
| EBOB | En Büyük Ortak Bölen |
| F0 | Temel Frekans |
| F1 | Birinci Formant |
| F2 | İkinci Formant |
| F3 | Üçüncü Formant |
| F4 | Dördüncü Formant |
| F5 | Beşinci Formant |
| FFT | Fast Fourier Transform |
| LPC | Linear Predictive Coding |
| LSTM | Long Short Term Memory |
| MFCC | Mel-Frequency Cepstral Coefficients |
| PCM | Pulse Code Modulation |
| SMO | Sequential Minimal Optimization |
| TTS | Text To Speech |

# ŞEKİL LİSTESİ

**Şekil 2.1** Özniteliklerin çıkarılarılması ve model eğitimi **.....6**
**Şekil 3.1** Deneysel çalışma adımları **.....19**

# TABLO LİSTESİ

**Tablo 2.1** Veri setleri ..... **7**
**Tablo 2.2** Duygu sınıfları ..... **8**
**Tablo 3.1** Altı duygulu sınıflandırma başarısı ..... **20**
**Tablo 3.2** Altı duygulu sınıflandırmada çapraz test sonuçları ..... **21**
**Tablo 3.3** Genişletilmiş model için alt duygulu sınıflandırma başarısı ..... **22**
**Tablo 3.4** İkili sınıflandırma başarısı ..... **23**
**Tablo 3.5** Diğer duygular ile ikili sınıflandırma başarısı ..... **23**
**Tablo 3.6** Genişletilmiş model için diğer duygular ile ikili sınıflandırma ..... **24**
**Tablo 3.7** Çağrı merkezi ses formatı ile ikili sınıflandırma ..... **25**
**Tablo 3.8** Arka plan gürültüsü eklenerek ikili sınıflandırma ..... **25**
**Tablo 3.9** MFCC öznitelikleri ile ikili sınıflandırma ..... **26**

# ÖZET

## Makine Öğrenmesi Yöntemleri İle Konuşma İşaretlerinden Duygu Analizi

Ferhat AYGÜN

Elektronik ve Haberleşme Mühendisliği Anabilim Dalı

Haberleşme Programı

Yüksek Lisans Tezi

Danışman: Doç. Dr. Bülent BOLAT

Bu tez çalışmasında, duygusal olarak etiketlenmiş konuşma sesi veri setlerindeki konuşma sesleri üzerinde sinyal işleme teknikleri kullanılarak elde edilen çeşitli özniteliklerle yapay öğrenme yoluyla bir duygu tanıma modeli üretilip üretilemeyeceği, hangi duygular için nasıl bir başarım olabileceği ve daha önce bu yolla yapılmış çalışmaların eksiklikleri ve gözden kaçan noktaları yapılan birçok deneyle tartışılmış ve bu konuda gerçek bir ihtiyaca cevap verebilecek en uygun çözüm önerisi ortaya konmuştur. Model üretimi için kullanılabilecek öznitelikler OpenSMILE isimli açık kaynak kodlu bir uygulama kullanılarak elde edilmiştir. Bu özniteliklerden konuşma tanıma ve konuşmacı tanıma gibi başka konuşma sesi uygulamaları için de yaygın olarak kullanılanların önemli olduğu düşünülerek tanımları yapılmıştır. Öznitelikler çıkarıldıktan sonra elde edilen veri setleri ile yapay öğrenme deneyleri yapılarak model üretilebilmesi için Weka isimli açık kaynak kodlu bir uygulama kullanılmıştır. Yapay öğrenme algoritması olarak destek vektör makineleri seçilmiş ve bütün deneyler bu algoritma kullanılarak yapılmştır. Destek vektör makineleri algoritmasının çalışma prensibi açıklandıktan

sonra bu algoritmanın kullanıldığı birçok deney yapılmış ve sonuçları yorumlanmıştır. Bu deney sonuçları ile özellikle çağrı merkezlerinde bir ihtiyaç olan sinirlilik ve mutluluk duygularının tespit edilerek oluşan çağrıların duygu yönünden analiz edilebilmesine olanak tanıyan bir çözümün mümkün olduğu ortaya konmuş ve bu çözüm önerisinin daha başarılı hale getirilebilmesi için de yapılabilecek yeni çalışmalardan bahsedilerek yeni önerilerde bulunulmuştur.

**Anahtar Kelimeler:** Duygu tanıma, makine öğrenmesi, çağrı merkezi, destek vektör makineleri, konuşma işleme.

# ABSTRACT

## Emotion Analysis From Speech Signals With Machine Learning Methods

Ferhat AYGÜN

Department of Electronics and Communications Engineering

Master of Science Thesis

Supervisor: Assoc. Prof. Dr. Bülent BOLAT

In this thesis, the feasibility of generating an emotion recognition model through artificial learning using signal processing techniques on speech sounds from emotionally labeled speech datasets is explored. The study discusses the potential success of such a model for different emotions, identifies shortcomings and overlooked aspects of previous related studies, and proposes an optimal solution that addresses a real need in this field. Feature extraction for model creation is performed using an open source application named OpenSMILE. The definitions of features, particularly those commonly used in other speech applications such as speech recognition and speaker recognition, are outlined considering their relevance. After feature extraction, artificial learning experiments are conducted using the obtained datasets and the Weka open source application to create the model. Support vector machines are chosen as the artificial learning algorithm, and all experiments are carried out using this algorithm. After explaining the working principle of support vector machines, numerous experiments are conducted using this algorithm, and the results are interpreted. The study concludes that it is possible to identify emotions such as irritability and happiness, particularly relevant in call

centers, allowing for the emotional analysis of calls. Suggestions for further improvements and new studies to enhance the proposed solution are also discussed.

**Keywords:** Emotion recognition, machine learning, call centers, support vector machines, speech processing.

# 1
# GİRİŞ

## 1.1 Literatür İncelemesi

Yapay öğrenme tekniklerinin gelişmesi, sinyal işleme ve özellikle ses işleme gibi maliyetli CPU işlemlerini hızlı bir şekilde yapabilen donanımların her geçen gün daha iyi özelliklerle üretilebilmesi sonucu, duygusal olarak etiketli ses verisinden yapay öğrenme ile duygu tanıma modellerinin geliştirilmesi konulu çalışmalar da eşzamanlı olarak artmaya başlamıştır. Yapılan çalışmalarda veri seti olarak genellikle bazı topluluklar tarafından oluşturulup internet üzerinde açık olarak paylaşılan veri setlerinden biri veya birkaçı seçilerek kullanılmıştır. Çalışmalarda genellikle ya çıkarılan özniteliklerin seçimi ile ya da farklı bir yapay sinir ağı tasarımı ile daha başarılı tahmin yapan modellerin geliştirilmesi üzerinde durulmuştur.

### 1.1.1 Daha Önce Yapılmış Çalışmalar

Konuşma sesi verisini işleyerek makine öğrenmesi ile duygu tanıma yapabilmek için bir çok çalışma yapılmış, makale ve tezler yazılmıştır. Genellikle duygu tanıma problemini benzer yönleri ile tartışan bu çalışmaların bazıları aşağıdaki gibidir.

Donn Morrison, Ruili Wang ve Liyanage C. De Silva tarafından çağrı merkezlerinde görüşmelerin duygusal olarak sınıflandırılıp analiz edilebilmesi için 2006 yılında iki farklı veri seti ile çalışma yapılmış, veri setlerinin biri ile %77.45, diğeri ile %72.18 başarılı sınıflandırma oranına ulaşabilen iki adet model geliştirilmiştir. Çalışmada genel olarak hangi özniteliklerin seçimi ile doğru sınıflandırma oranının artabileceği üzerinde ve kollektif öğrenme tekniği ile yapay öğrenme kısmındaki doğru sınıflandırma başarısının arttırılabileceği üzerinde tartışılmıştır [1].

Enrique M. Albornoz, Diego H. Milone ve Hugo L. Rufiner tarafından 2011 yılında yaptığı çalışma ile sinirli, şakacı, korkmuş, sıkılmış, iğrenmiş, üzgün ve nötr etiketli 535 konuşma sesi verisini kullanarak geliştirdiği duygu tanıma modelinin doğru sınıflandırma oranını %71.75 olarak ölçmüştür. Çalışmada genel olarak doğru öğrenme algoritması seçiminin ve yaptıkları çalışmada önerdikleri şekilde sınıfları

gruplara ayırıp birden fazla sınıflandırıcı kullanmanın başarımı arttırması üzerinde tartışılmıştır [2].

2014 yılında Chien Shing Ooi, Kah Phooi Seng, Li-Minn Ang ve Li Wern Chew tarafından yapılmış bir başka çalışmada eNTERFACE'05 [3] ve RML [4] isimli veri setleri ile çalışılmış ve özniteliklerin prosodik öznitelikler ve spectral öznitelikler olarak gruplanarak her bir grup için ayrı model üretilmesinin ve sınıflandırmanın da bu iki model ile iki adımda gerçekleştirilmesinin başarımı arttırması üzerinde tartışılmıştır. Bu çalışmada iki ayrı veri seti için doğru sınıflandırma oranı birinde %75.89, diğerinde %68.57 olarak ölçülmüştür [5].

Aslı Demir, Orhan Atila ve Abdulkadir Şengür 2019 yılında yaptıkları çalışmada Uzun Kısa Süreli Bellek (Long Short Term Memory - LSTM) [6] Ağları mimarisi ile tasarlanan bir derin öğrenme ağı ile yapılacak eğitimin sınıflandırma başarısına etkisi üzerinde tartışmışlardır. Bu çalışmada sırasıyla RAVDESS [7], SAVEE [8] ve RML [4] isimli veri setleri kullanılmış ve elde edilen modellerin doğru sınıflandırma başarısı sırasıyla %68.8, %72.13 ve %70.35 olarak ölçülmüştür [9].

2020 yılında yayınlanan ve Dias Issa, M. Fatih Demirci ve Adnan Yazıcı tarafından yapılan bir çalışmada ise IEMOCAP [10] isimli veri seti üzerinde çok katmanlı derin öğrenme ağı mimarisi kullanılarak üretilen model ile yapılan sınıflandırma için doğru sınıflandırma başarısı %64.30 olarak ölçülmüştür [11].

2021 yılında Yeşim Ülgen Sönmez yaptığı çalışmada, özniteliklerin seçimi ile igili deneyler yaparak sınıflandırıcı performansına en çok etki eden öznitelikleri belirleyerek optimum sınıflandırıcıyı tasarlamak için farklı bir öznitelik seçim tekniği önermiştir. Bu şekilde geliştirdiği modelin başarımını ise veri setlerine göre %89.16, %76.67, %74.31 ve %95.16 olarak ölçmüştür [12].

2022 yılında Fatma Gümüş yaptığı çalışmada diğerlerine benzer şekilde ses öznitelikleri çıkartarak veri setini yaptığı deneylerle çeşitli sınıflandırıcılar ve evrişimsel sinir ağları (CNN) [13] kullanarak ayrı ayrı test etmiş ve modeller geliştirmiş, en iyi sonucu CNN [13] kullandığı modelde %95 f-skor ile korku duygusu için ölçmüştür [14].

2022 yılında Özgür Demir, duygu ve konuşmacının cinsiyetini birlikte tespit edilmesi ile ilgili yaptığı çalışmada CNN [13] ve Çok Katmanlı Algılayıcı [15]

kullanmış, ağların duygu tanımada başarılı ancak cinsiyet tanımada başarısız olduğunu gözlemlemiştir [16].

2023 yılında Ülkü Bayraktar, EmoDB [17], RAVDESS [7], TESS [18], CREMA-D [19] ve SAVEE [8] veri setleri ile çalışmış, MFCC ve diğer geleneksel ses işleme öznitelikleri diye ifade ettiği öznitelikleri kullanarak çeşitli makine öğrenmesi algoritmalarıyla sınıflandırıcılar tasarlayarak hem seçilen öznitelikler yönünden hem de seçilen sınıflandırıcı yönünden model başarımını test etmiş, sonuç olarak en iyi özniteliklerin MFCC öznitelikleri, en iyi sınıflandırıcı performansının ise destek vektör makineleri kullanılarak ulaşıldığı sonucuna varmıştır [20].

### 1.1.2 Yapılan Çalışmalarda Gözden Kaçan Noktalar

Konuşma sesinden duygu tanıma probleminin makine öğrenmesi yoluyla çözümü için yapılan çalışmaların birçoğu, yukarıda örnekleri verilen çalışmalardaki gibi öznitelik seçimi ve makine öğrenmesi metodu üzerinde tartışılarak yapılmıştır. Ancak üretilen modeller gerçek dünyadaki problemlerin çözümü için test edildiğinde veya farklı bir veri seti üzerinde test edildiğinde çalışmada ulaşılan başarım oranlarının çok uzağında kalınmaktadır. Bu çelişkinin ana sebepleri genel olarak konuşma sesi verisinden duygu tanıma probleminin aslında zor bir problem olması, duygu etiketli konuşma veri setlerinin nesnel olarak oluşturulma zorluğu, veri setlerindeki farklı konuşmacı sayısının azlığı, doğal bir ses yerine yapay olarak seslendirilmiş veri setlerinin kullanımı, farklı dil, cinsiyet, yaş grubu ve coğrafi bölgeden alınan seslerdeki duygu özelliklerinin de farklılık göstermesi, sesin film, stüdyo kaydı, ortam kaydı, çağrı merkezi görüşmesi gibi kaydedildiği ortamlara göre içerdiği duygunun da farklılık göstermesi, aşırı öğrenme problemi, sınıflandırılmaya çalışılan duygu sayısının çokluğu olarak sıralanabilmektedir.

## 1.2 Tezin Önemi, Amacı ve İzlenilen Yol

Ses verisinden duygu tanıma probleminin çözümü, günümüzde özellikle çağrı merkezi görüşmelerinin duygusal olarak ve özellikle de sinirlilik ve mutluluk duyguları yönünden doğru olarak sınıflandırılıp analiz edilebilmeleri ihtiyacının çözümü için çok önemlidir. Firmaların satış ve pazarlama faaliyetlerindeki stratejilerini daha doğru belirleyebilmeleri ve bu sayede satışlarını arttırabilmeleri, müşteri memnuniyetlerini devam ettirerek müşteri kaçışının önüne hızlı bir şekilde

geçebilmeleri, ileride herhangi bir müşteri ile karşılaşabilme ihtimali olan ve kendilerine zarar verebilecek dava gibi süreçlerin önüne daha önce o müşteri ile yapılan ve müşteri tarafında sinirlilik gözlemlenen çağrının daha önceden analiz edilmesi sonucu zamanında önüne geçilebilmesi, çağrı merkezinde yapılan görüşmelerin tamamının duygusal olarak özellikle de sinirlilik ve mutluluk duyguları yönünden yüksek doğrulukla sınıflandırılabilmeleri sayesinde sağlanabilmektedir.

Bu tezin amacı sadece bir veri seti üzerinde elde edilen ancak başka veri setleri üzerinde test edildiğinde başarısız kalan sınıflandırma modelleri üretiminin önüne geçerek gerçek dünyada çağrı merkezlerinin ihtiyacı olan görüşme kayıtlarının sinirlilik ve mutluluk duyguları yönünden doğru bir şekilde sınıflandırılıp analiz edilmesine olanak sağlayan bir çözüm üretmeye katkıda bulunmaktır.

Tezin amacını gerçekleştirebilmek için öncelikle daha önce yapılmış çalışmalardaki gibi internette açık olan veri setlerinden bazıları seçilmiş ve seçilen veri setleri için ayrı ayrı eğitimler yapılmış ve benzer başarım oranlarına ulaşıldığı gösterilmiştir. Sonrasında her bir veri seti için üretilen model diğer veri setleri ile test edildiğinde ne kadar büyük bir başarım farkı olduğu ve aslında bu çözümlerin tek başına gerçek problemlerin çözümü için yeterli olamayacağı gösterilmiştir. Sonrasında sırasıyla veri setlerini birleştirerek veri sayısını ve çeşitliliği artırmanın, çağrı merkezi probleminin çözümünde gereksiz duyguları çıkarıp sınıf sayısı azaltılarak sadece sinirlilik ve mutluluk duyguları üzerinde çalışmanın, veri setindeki tüm sesleri standart çağrı merkezi formatı olan 16 bit 8 Khz PCM formatına getirerek tekrar eğitim yapmanın, gürültüsüz olarak kaydedilmiş olan veri setine çağrı merkezi arka plan gürültüsü ekleyip eğitimi tekrarlamanın başarım üzerindeki etkilerini incelemek için ayrı ayrı deneyler yapılmış, sonuçlar gösterilmiş ve yorumlanmıştır. Deney sonuçlarından çıkan verilere göre optimum çözüm ortaya konmuş ve başarımın daha da arttırılabilmesi için yapılabilecek yeni çalışmalar önerilmiştir.

## 1.3 Tezin Organizasyonu

Tezin ikinci bölümünde kullanılan yöntemler, uygulamalar ve veri setleri için bilgi verilmiş, ayrıca sinyal işleme ve konuşma sesi işleme için önemli görülen kavramlar açıklanmıştır. Üçüncü bölümde yapılan deneysel çalışmalar anlatılarak sonuçları yorumlanmış, dördüncü bölümde ise deneysel çalışmalardan çıkan

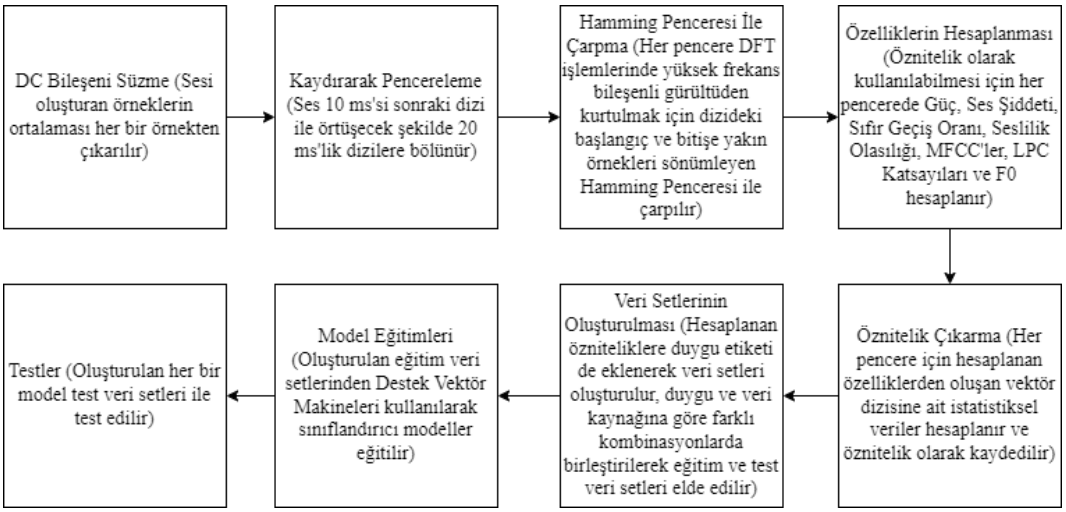
sonuçlara göre çözüm önerisi sunulup başarımın daha da arttırılabilmesi için yapılacak yeni çalışmalar önerilmiştir.

# 2
# YÖNTEM

## 2.1 Veri Setleri ve Blok Diyagram

Bu tez çalışmasında veri seti olarak daha önce yapılmış çalışmalarda da sıklıkla kullanılan veri setlerinden beş tanesi seçilerek kullanılmıştır. Bunlar EmoDB [17], SAVEE [8], RAVDESS [7], CREMA-D [19] ve TESS [18] veri setleridir. Bu veri setlerinde Şekil 2.1'deki blok diyagramda gösterilen işlemler sırayla uygulanarak öznitelikler elde edilip sınıflandırıcı modeller eğitilmiştir. Bu veri setlerinde yer alan toplam veri sayısı ve duygu sınıfları Tablo 2.1'de gösterilmiştir.

Şekil 2.1'deki blok diyagramda gösterildiği gibi önce DC bileşen süzülüp sıfır ortalamalı ses sinyali elde edilmiştir. Sonrasında Öznitelikler başlığı altında daha detaylı incelenen kaydırmalı pencereleme işlemi ile 20 ms'lik pencereler elde edilmiştir. Sonrasında her pencere Hamming Penceresi ile çarpılarak DFT işlemi sonrası yüksek frekanslı gürültü bileşenlerinin oluşması engellenmiştir. Her pencere için kullanılacak özellikler hesaplanmış ve elde edilen bu özellik vektörlerinden öznitelikler hesaplanmıştır. Bu şekilde elde edilen veri setlerinden destek vektör makineleri kullanılarak sınıflandırıcı modeller üretilmiş ve test veri setleri ile test edilmiştir.




**Şekil 2.1** Özniteliklerin çıkarılarılması ve model eğitimi

**Tablo 2.1** Veri setleri

| Veri Seti | Veri Sayısı | Konuşmacı | Dil | Duygu Sınıfları |
|---|---|---|---|---|
| **EmoDB** | **535** | **10** | **Almanca** | **anger, boredom, disgust, fear, happiness, sadness, neutral** |
| **SAVEE** | **480** | **4** | **İngilizce** | **anger, disgust, fear, happiness, sadness, surprise, neutral** |
| **RAVDESS** | **1440** | **24** | **İngilizce** | **anger, disgust, fear, happiness, sadness, surprise, calm, neutral** |
| **CREMA-D** | **7442** | **91** | **Aksanlı İngilizce** | **anger, disgust, fear, happiness, sadness, neutral** |
| **TESS** | **2800** | **2** | **İngilizce** | **anger, disgust, fear, happiness, sadness, neutral, surprise** |

Her bir veri setinin kendi içerisinde verilerin duygu etiketine göre dağılımı yaklaşık olarak eşit şekildedir. Her bir veri setinde ortak olan etiketler anger, disgust, fear, happiness, sadness ve neutral etiketleridir. Diğer etiketler, yani boredom, calm ve surprise etiketleri bazı veri setlerinde bulunup bazılarında bulunmamaktadır. Veri setlerindeki duygu sınıfı etiketlerinin açıklamaları Tablo 2.2'de gösterilmiştir.

**Tablo 2.2** Duygu sınıfları

| Etiket | Açıklama |
|---|---|
| **anger** | **Sinirlilik** |
| **disgust** | **İğrenmiş** |
| **fear** | **Korku** |
| **happiness** | **Mutluluk** |
| **sadness** | **Üzgünlük** |
| **surprise** | **Şaşırmış** |
| **calm** | **Sakin** |
| **boredom** | **Sıkılmış** |
| **neutral** | **Nötr** |

## 2.2 Öznitelikler

İnternet üzerinde bulunan veri setlerinden makine öğrenmesi yoluyla model üretebilmek için öncelikle veri setindeki ses dosyalarından sinyal işleme yoluyla sesteki duygu üzerinde etkisi olabileceği düşünülen özelliklerin çıkarılması gerekmektedir. Yapay öğrenme yoluyla model oluşturulacağından etkisi olup olmadığı bilinmese bile olabildiğince çok özelliğin hesaplanması başarımın artmasına yardımcı olacaktır. Ancak her bir özelliği hesaplamak CPU kaynağı tüketeceğinden bunun maliyeti de hesaba katılmalı ve daha sonra bazı özellikler eksiltilerek yapılan denemelerde o özelliğin başarıma katkısı gözlemlenmiyorsa öznitelik olarak kullanılmamalıdır. Ayrıca aşırı öğrenme problemine yol açabileceğinden özelliğin duyguya etkisi olmadığı bariz olarak düşünülüyorsa da özellik olarak kullanılmamalıdır.

Özniteliklerin çıkarılabilmesi için öncelikle ses verisi zaman domeninde küçük boyutlu pencerelere bölünür. Ses işleme için bu boyut genellikle 20 ms olarak seçilir. 20 ms'lik her bir pencere için ayrı ayrı hem zaman domeninde hem frekans domeninde birçok özellik hesaplanır ve bu özellikler bir vektör olarak ifade edilir. Vektörün boyutu özellik sayısı olacaktır. Sonuç olarak tüm ses kaydı düşünüldüğünde her biri 20 ms'lik pencereye tekabül eden vektörlerin dizisine karşılık bir duygu etiketi olan bir veri seti elde edilmiş olacaktır. Bu vektör dizisinin boyutu ise her bir ses kaydı için ses kaydının uzunluğunun 20 ms'ye bölümü kadar, yani ses kaydındaki 20 ms'lik pencere saysısı kadar olacağından her bir ses kaydı verisi için ses kaydının uzunluğuna göre farklı olabilecektir. Bu durumda öznitelik olarak bu vektörlerdeki her bir özelliğin çeşitli istatistiksel bilgilerini kullanmak doğru bir yaklaşım olacaktır. Bu bilgiler, genellikle maksimum, minimum, maksimum ve minimumun farkı, maksimum ve minimumun yerleri, medyan, ortalama, standart sapma, varyans gibi bilgilerdir. Sonuç olarak elde edilecek öznitelik sayısı yaklaşık olarak vektör boyutu ile hesaplanacak istatistiksel bilgi sayısının çarpımına eşit olacaktır. Yaklaşık denilmesinin sebebi, bazı özellikler için tüm istatistiksel bilgilerin kullanılmasına gerek olmaması, bazı özellikler içinse ekstra istatistiksel bilgi kullanımının faydalı olabilmesidir.

Bu çalışmada her bir 20 ms'lik pencere için oluşturulacak vektörde hesaplanacak özellikler güç, ses şiddeti, sıfır geçiş oranı, seslilik/sessizlik olasılığı, MFCC (1-12), LPC ve F0 (Temel Frekans) özellikleri olacaktır. Bu çalışmada özelliklerin hesaplanmasıyla oluşan vektör dizisinden elde edilen öznitelik sayısı 988'dir. Bu öznitelikler 2.4.1'de anlatılan OpenSMILE programı ile çıkarılmış olup bu programın kurumu ile birlikte gelen ve duygu öznitelikleri çıkarmak için örnek olarak sunulan emobase.conf dosyasında belirtilen konfigürasyon ile elde edilmiştir.

Çıkarılacak öznitelikler için kullanılacak özellikler açıklanmadan önce 20 ms'lik pencereler üzerinde yapılması gereken bazı işlemler bulunmaktadır ve bunların açıklanması gerekmektedir. Bu işlemler sesin DC bileşeninin süzülmesi, DFT işleminde yüksek frekans bölgesinde oluşacak gürültü bileşenlerinin oluşmasını minimuma indirmek için bir zaman domeninde bir pencere ile çarpılması ve bu işlemden dolayı oluşan bilgi kaybını telafi etmek için pencerelerin birinin bitip

diğerinin başlayacağı 20 ms'lik pencereler olarak değil de, kaydırmalı şekilde yani pencerelerin bir kısmı üst üste çakışacak şekilde seçilmesi işlemleridir.

DC bileşen süzme işlemi, tüm örneklerden pencerenin ortalaması çıkarılarak (2.1) denkleminde gösterildiği gibi yapılabilir.

$$y[n] = x[n] - \frac{(\sum_{k=1}^{N} x[k])}{N} \tag{2.1}$$

Her bir pencere DFT işlemine direkt olarak alındığında DFT işleminin matematiksel ifadesinin sonucu olarak son örneğin ardından ilk örneğin tekrar geldiği şeklinde bir hesaplama yapıldığından, oluşan bu ani değişim DFT sonucunda yüksek frekans bileşenleri eklenmesine neden olmaktadır. Bunun önlenmesi amacıyla 20 ms'lik bu pencerenin, zaman domeninde (2.2) denklemi ile verilen ve Hann penceresi olarak adlandırılan bir pencere ile çarpılması gerekmektedir. Bu şekilde pencerenin ilk ve son örnekleri sönümlenerek birbirine yaklaştırılacağından eklenecek yüksek frekans bileşenli gürültü azaltılmış olacaktır.

$$w[n] = a_0 - (1 - a_0)\cos\left(\frac{2\pi n}{N}\right) \tag{2.2}$$

Bu çalışmada, (2.2) denkleminde verilen Hann penceresi denkleminde $a_0$ değerinin yaklaşık 0.54 olarak seçilmesi durumunda oluşan ve Hamming penceresi olarak isimlendirilen pencere kullanılmıştır.

Pencerelerin Hamming penceresinden geçirilmesi sebebiyle zaman domeninde bilgi kaybı oluşacağından, pencereler, bir sonraki pencere bir önceki pencerenin ortasından başlayacak şekilde, yani her bir pencerenin bir kısmı ardışık pencere ile örtüşecek şekilde oluşturulur. Bu şekilde bir pencerede zaman domeninde kaybedilmiş olan bilgi sonraki pencerede örtüşen bölümde var olacağından sesin tamamı düşünüldüğünde bilgi kaybı olmamış olur. Bu çalışmada, her pencere ardışık pencere ile 10 ms boyunca örtüşecek şekilde oluşturulmuştur.

### 2.2.1 Güç

Her bir 20 ms'lik pencere için hesaplanacak vektördeki parametrelerden ilki güçtür. Bu parametre zaman domeninde ve (2.3) denklemini hesaplayan bir algoritma kullanılarak kolaylıkla bulunabilir. Her bir örneğin karesi alınıp toplanarak elde

edilen toplam örnek sayısına bölünür ve çıkan sonucun karekökü ilgili 20 ms'lik pencerenin gücünü vermektedir.

$$rms = \sqrt{\frac{1}{N}\sum_{n}|x(n)|^2} \tag{2.3}$$

### 2.2.2 Ses Şiddeti

İngilizce loudness olarak da ifade edilen ses şiddeti her bir 20 ms'lik pencere için hesaplanan gücün pencereye ait örneklerin alabileceği maksimum değere bölünerek çıkan değerin 10 tabanında logaritmasının 20 ile çarpımı o pencere için sesin şiddetini verecektir. Burada örneklerin signed 16 bit olmasından dolayı alabileceği maksimum değer 32768 olduğundan ses şiddeti (2.4) denklemi ile hesaplanabilir.

$$dBFS = 20Log\left(\frac{rms}{32768}\right) \tag{2.4}$$

### 2.2.3 Sıfır Geçiş Oranı

Her bir 20 ms'lik pencere için art arda gelen örneklerin kaç tanesinde pozitiften negatife veya negatiften pozitife geçtiğini gösteren özelliktir ve (2.5) denklemi ile hesaplanabilir.

$$ZCR = \sum_{n}|\text{sign}(x[n]) - \text{sign}(x[n-1])| \tag{2.5}$$

### 2.2.4 Seslilik Olasılığı

Her bir 20 ms'lik pencerenin konuşma sesi parçası içerip içermediği olasılığını verir. 1 veya 0 değerini alır. Güç, ses şiddeti ve sıfır geçiş oranı hesaplandıktan sonra kural tabanlı olarak belirlenir. Güç, ses şiddeti ve sıfır geçiş oranının belirlenen eşik değeri geçmesi durumunda 1, aksi takdirde 0 olarak hesaplanır.

### 2.2.5 MFCC (Mel-Frequency Cepstral Coefficients)

Ses işlemede önemli özniteliklerden biri de MFCC'dir. Özellikle konuşma işleme uygulamalarından konuşma tanımada öznitelik olarak kullanılır. MFCC, sesin frekans içeriğini temsil etmek üzere kepstral analiz ile mel ölçeğinde örneklenmiş frekansları kullanmaktadır [21].

Frekans domeninde hesaplanan bir özelliktir. Her bir 20 ms'lik pencere için öncelikle DFT yani ayrık Fourier dönüşümü hesaplanır. Ayrık Fourier

dönüşümünün ifadesi (2.6) denklemi ile verilmiştir. Örnek sayısının ikinin katları olması durumunda FFT, yani hızlı Fourier dönüşümü algoritması kullanılarak da daha az CPU kaynağı ile DFT hesaplanabilir. Bu sebeple DFT hesaplanması gereken durumlarda örnek sayısının ikinin kuvveti olacak şekilde seçilmesine ve DFT'nin FFT algoritması ile hesaplanmasına dikkat edilir.

$$X_k = \sum_{n=0}^{N-1} x_n e^{-i2\pi kn/N} \ , k = 0,1, \dots, N-1 \tag{2.6}$$

DFT ile elde edilen k boyutlu vektör, Mel Filtre Bankası adı verilen ve bu tez çalışmasında 26 adet olan filtreden geçirilerek her bir filtre çıkışı ayrı ayrı hesaplanır. Bu filtreler insan kulağını taklit etmek için tasarlanmış olup genel olarak düşük frekanslar için dar yüksek frekanslar için geniş olarak seçilmiş, frekans domeninde üçgen biçimli ve birbiri ile kısmen örtüşen ve ardışık sıralanan filtrelerdir. 26 adet filtrenin her birinin çıkışındaki enerjiler toplanarak, her bir filtre çıkışının toplam enerjisi ayrı ayrı (2.7) denklemindeki gibi hesaplanır.

$$S_i = \sum_k X(k) M_i(k) \ , i = 1,2, \dots ,26 \tag{2.7}$$

Elde edilen Mel filtre çıkışlarının toplam enerjilerinin logaritması alınarak ve buna ayrık kosinüs dönüşümü uygulanarak MFFC'ler hesaplanır. Bu çalışmada 12 adet MFCC (2.8) denklemi ile hesaplanır.

$$MFCC_i = \sum_{n=1}^{N} \log(S_n) \cos\left[ i \left( n - \frac{1}{2} \right) \frac{\pi}{N} \right] \ , i = 1,2, \dots ,12 \ , n = 1,2, \dots ,26 \tag{2.8}$$

### 2.2.6 LPC (Linear Predictive Coding) Katsayıları

LPC Katsayıları, zaman domeninde önceki örneklerin değerleri ile sonraki örnekleri tahmin etmeyi sağlayan bir model oluşturabilmek için kullanılan katsayılardır. Bu katsayılar tahmin yaparken hatayı minimuma düşürecek şekilde deneme yanılma ile döngüsel olarak çalıştırılan bir algoritma ile zaman domeninde hesaplanmaktadır. Minimum hata ile çalışacak modelin katsayıları LPC katsayıları olarak isimlendirilmekte ve modelin karmaşıklığına göre farklı boyutlarda bir vektör olarak ifade edilebilmektedir. Bu tez çalışmasında 8 adet katsayı

kullanılacak olup LPC katsayıları, her bir 20 ms'lik pencere için 8 boyutlu bir vektör olarak hesaplanacaktır.

Model (2.9) denklemini sağlayarak tahmin yapmaktadır. Bu durumda hata fonksiyonu (2.10) denklemindeki gibi olacaktır. Sonuç olarak LPC, (2.11) denklemindeki ortalama karesel hatayı minimum yapan katsayıların bulunması, yani bir optimizasyon problemidir.

$$\hat{x}[n] = \sum_{k=1}^{N} a_k x[n-k] \tag{2.9}$$

$$e[n] = x[n] - \hat{x}[n] \tag{2.10}$$

$$E\{e^2[n]\} = E\left\{\left(x[n] - \sum_{k=1}^{N} a_k x[n-k]\right)^2\right\} \quad , \quad a_k : LPC\ katsayıları \tag{2.11}$$

Çok sayıda örnekten oluşan bir penceredeki örneklerin tamamının yalnızca bir miktar katsayının hesaplanıp bir modelle tahmin edilebilmesi sebebiyle LPC, ses sıkıştırma uygulamalarında sıklıkla kullanılmaktadır. G729 [22] ve GSM [23] gibi özellikle sesli iletişim kanallarında kullanılan, konuşma sesi sıkıştırılması için geliştirilmiş kodekler, konuşma sesini LPC katsayıları ile kodlayan kodeklerdir. LPC katsayıları hesaplanarak bellekte ses örneklerinin tamamını tutmak yerine sadece bu katsayıları tutmak yeterli olacağından çok az bir kayıpla sesin sıkıştırılması mümkün olmaktadır. Ses kodeklerinin çoğunda LPC kullanılmakta olup, LPC katsayıları, makine öğrenmesi içeren ses uygulamalarında sıklıkla öznitelik olarak da kullanılmaktadır.

### 2.2.7 F0 (Temel Frekans)

Konuşma sesleri için kullanılan F0, yani temel frekans kavramı, ses tellerinin titreşim frekansı için kullanılan bir ifadedir ve pitch olarak da ifade edilmektedir. Örnekleme frekansına göre DFT işlemi sonucu elde edilen frekans spektrumunun çözünürlüğü için kullanılan ve yine $f_0$ olarak ifade edilen kavramdan farklıdır. F0, akciğerlerden gelen havanın ses tellerinde meydana getirdiği titreşimin frekansıdır. Dolayısıyla konuşma sesinin sadece ses tellerinin titreştiği sesli harf içeren bölümlerinde bulunmaktadır. Konuşma seslerinde F0 değerlerinin alabileceği aralık, konuşan kişiye ait vokal tract ya da ses yolu da denilen, akciğerlerden gelen

havanın ses tellerine gelmeden önce geçtiği gırtlaktaki bir bölgenin boyu ile ilişkilidir. Ses yolunun boyu arrtıkça F0 azalmakta, ses yolu boyu azaldıkça F0 artmaktadır. Buna bağlı olarak da yine F0 ne kadar düşükse ses o kadar kalın, ne kadar yüksekse ses de o kadar ince olacaktır. Kadınların, erkeklerin ve çocukların seslerinin birbirlerinden ayırt edilebilmelerinin sebebi, ses yolu boylarının farklı olmalarından dolayı çıkardıkları seslerdeki F0 değerlerinin de farklı aralıklarda olmasıdır.

F0 konuşma boyunca sabit olmayıp konuşmacının ses tellerinin ve ses yolunun özelliklerine göre belirli bir aralıkta, belirli bir olasılık dağılım fonksiyonu ile değişim göstermektedir. Dolayısıyla F0, konuşmacı için de ayırt edici bir özellik olarak özellikle konuşmacı tanıma uygulamalarında ve TTS denilen metinden sese dönüştürme uygulamalarında önemli bir öznitelik olarak kullanılmaktadır. TTS uygulamalarında F0'ın sabit olması sesin doğal bir insan sesi gibi değil de robotik bir ses gibi olmasına sebep olmaktadır. Bu sebeplerle konuşma boyunca F0'ın değişimi, sesin içerdiği duygu bakımından da önemli miktarda bilgi taşımakta ve makine öğrenmesi yoluyla duygu tanıma yapılması için çok önemli bir öznitelik olarak kullanılmaktadır.

Temel frekans hem zaman domeninde hem frekans domeninde geliştirilmiş bir çok algoritma ile elde edilebilir. Her bir 20 ms'lik ses sinyali için (2.6) denkleminde hesaplanan DFT'deki frekans bileşenleri enerji seviyelerine göre sırasıyla F1 veya birinci formant, sonrakiler de yine F2, F3, F4, F5 … şeklinde adlandırılır. Buradaki F1, F2, F3, F4 ve F5 bileşenlerinin her biri F0'ın katlarıdır. Bu sebeple F0, yani temel frekans, harmonik frekans olarak da adlandırılmaktadır. Konuşma sesi işleme için ilk 4 veya 5 formantın seçilerek EBOB'unun alınması yaklaşık olarak F0 değerini hesaplayabilmek için yeterli olmaktadır ancak birbirlerine göre kesinlik, performans ve maliyet açılarından artı ve eksileri olan, zaman veya frekans domeninde hesaplama yapan YIN [24], AMDF [25] veya Mcleod [26] gibi algoritmalar da bulunmaktadır. Yapılacak uygulamada DFT işlemi varsa işlem maliyet ve hızından tasarruf için EBOB alınarak hesaplanabileceği gibi daha kesin bir sonuç bulmak için bu algoritmalar da kullanılabilmektedir.

## 2.3 Makine Öğrenmesi

Bu tez çalışmasında elde edilen veri setleri ile sınıflandırma problemini çözebilmek için makine öğrenmesi kullanılmıştır. Makine öğrenmesi, basitçe ifade etmek gerekirse problemin regresyon mu yoksa sınıflandırma mı olduğuna göre değişebilen, farklı problemler için farklı başarımlara sahip, temelde döngüsel olarak deneme yanılma yapılarak minimum hatayı yakalamaya çalışan, çeşitli algoritmalar yoluyla var olan bir veri setinin analiz edilerek yeni gelecek veriler için tahmin yapabilen bir modelin üretimi yöntemidir.

Makine öğrenmesi konusu çok geniş bir konu olduğundan ve bu tez çalışmasının doğrudan bir konusu olmayıp sadece çalışmada araç olarak kullanılmasından dolayı tez çalışması içeriğinde detaylarına girilmemiştir.

Bu çalışmada algoritma olarak destek vektör makineleri seçilmiştir.

### 2.3.1 Destek Vektör Makineleri

Destek vektör makineleri algoritması, matematiksel olarak düzlem üzerinde bulunan noktaların bir doğru ile ayrıştırılması prensibine dayanmaktadır. Bu tezin konusu yapay öğrenme algoritmalarını tartışmak olmadığından bütün öğrenme algoritmalarının detaylarına girilmemiştir. Farklı öğrenme algoritmalarının kullanılması durumunda da yine birbirine yakın sonuçlar çıkmakla beraber bu tezin öncesinde yapılan testlerde en başarılı sonucu destek vektör makinelerinin vermesi sebebiyle bu çalışmada da yine destek vektör makineleri sınıflandırıcı algoritma olarak seçilmiştir. Bu sebeple bu çalışmada sadece destek vektör makinelerinin çalışma prensibi detaylarını girilmiş, diğer algoritmalar anlatılmamıştır.

Destek vektör makineleri, daha çok örüntü tanıma problemleri için sınıflandırıcı olarak kullanılsa da regresyon kestirimi için de kullanılırlar. Sınıflandırıcıları genel olarak doğrusal sınıflandırıcılar ve doğrusal olmayan sınıflandırıcılar olarak ikiye ayırabiliriz [27]. Doğrusal destek vektör makineleri ayrılabilen veya ayrılamayan veri setleri üzerinde kullanılabilirler [28].

Destek vektör makineleri, el yazısıyla yazılmış karakter tanımada [29], nesne tanımada [30], konuşmacı tanımada [31], yüz tanımada [32] ve metin sınıflandırmada [33] sınıflandırıcı olarak kullanılmaktadır.

Genel olarak, verilen bir öğrenme görevi ve sınırlı sayıda veri için, en iyi performans, eğer eğitim verisi ve makinenin kapasitesi arasındaki denge sağlanabilirse gösterilebilmektedir [28].

Doğrusal destek vektör makinelerinin ayrık veri üzerinde çalışması durumu için veri (2.12) denklemindeki gibi etiketlenir.

$$\{\boldsymbol{x}_i, y_i\}, i = 1, \dots, l, y_i \in \{-1,1\}, \boldsymbol{x}_i \in \boldsymbol{R}^d \tag{2.12}$$

Aynı zamanda, pozitif etiketlenmiş verileri negatif etiketlenmiş verilerden ayıran, adına ayrıştırıcı hiperdüzlem denilen bir düzlemin olduğu varsayılır [27] [28]. Bu düzlem üzerindeki noktalar düzleme dik olan bir $\boldsymbol{w}$ vektörü için, (2.13) denklemini sağlarlar.

$$\boldsymbol{w}.\boldsymbol{x} + b = 0 \tag{2.13}$$

Doğrusal ayrılabilen durum için, destek vektör algoritması adı verilen eğitim verisi noktalarının bu dözleme olan en büyük marjinini elde etmeyi sağlayan hiperdüzlemi bulmaya çalışır.

Verilen eğitim verilerinin (2.14) ve (2.15) eşitsizliklerini sağladıkları varsayılarak bu eşitsizlikler Lagrange formülasyonuna dönüştürülerek (2.16) denklemi elde edilir.

$$\boldsymbol{x}_i.\boldsymbol{w} + b \geq +1 \; for \; y_i = +1 \tag{2.14}$$

$$\boldsymbol{x}_i.\boldsymbol{w} + b \geq +1 \; for \; y_i = +1 \tag{2.15}$$

$$L_p = \frac{1}{2}\|\boldsymbol{w}\|^2 - \sum_{i=1}^{l} \alpha_i y_i (\boldsymbol{x}_i.\boldsymbol{w} + b) + \sum_{i=1}^{l} \alpha_i \tag{2.16}$$

Bahsedilen hiperdüzlemi elde edebilmek için $L_P$'nin $\boldsymbol{w}$ ve $b$'ye göre minimize edilmesi gerekmektedir. Bu gereklilik beraberinde $L_P$'nin $\alpha_i$'lere göre türevlerinin sıfırlanması anlamına da gelir. Bu da (2.17) ve (2.18) koşullarının sağlanmasını gerektirir.

$$\boldsymbol{w} = \sum_i \alpha_i y_i \boldsymbol{x}_i \tag{2.17}$$

$$\sum_i \alpha_i y_i = 0 \tag{2.18}$$

Bu koşullar (2.16) denkleminde yerine konulursa (2.19) denklemi elde edilir.

$$L_D = \sum_i \alpha_i - \frac{1}{2}\sum_{i,j} \alpha_i \alpha_j y_i y_j \boldsymbol{x}_i . \boldsymbol{x}_j \tag{2.19}$$

Burada $L_D$, $L_P$'nin aksine maksimize edilmesi gereken bir değerdir. Destek vektör makinelerinin yaptığı da bu değeri $\alpha_i$'lere göre maksimize etmektir [28].

Doğrusal olmayan destek vektör makineleri içinse öncelikle (2.20) denkleminde gösterildiği gibi verinin bir başka Öklid uzayına eşlemesi ile başlanır.

$$\Phi = \boldsymbol{R}^d \rightarrow H \tag{2.20}$$

Doğrusal sınıflandırmadaki gibi doğrusal olmayan sınıflandırmada da eğitim verisi eşitliklerde skaler çarpım halinde bulunur. Bu durumda eşlenilen uzaya eğitim verisi kernel fonksiyonları ile taşınır. Eğitim algoritmasında ise (2.21) denklemindeki gibi $K$ fonksiyonu kullanılır ve bu fonksiyona bir örnek olarak (2.22) verilebilir.

$$K(\boldsymbol{x}_i, \boldsymbol{x}_j) = \Phi(\boldsymbol{x}_i)\Phi(\boldsymbol{x}_j) \tag{2.21}$$

$$K(\boldsymbol{x}_i, \boldsymbol{x}_j) = e^{-\|x_i - x_j\|^2 / 2\sigma^2} \tag{2.22}$$

Bu sayede, farklı bir uzayda doğrusal sınıflandırma durumundaki işlemler uygulanmış olur. Bu durumda destek vektör makineleri, (2.23) denklemindeki optimizasyon problemini, işaretini tespit ederek çözer.

$$f(\boldsymbol{x}) = \sum_{i=1}^{N_S} \alpha_i y_i \Phi(\boldsymbol{s}_i).\Phi(\boldsymbol{x}) + b = \sum_{i=1}^{N_S} \alpha_i y_i K(\boldsymbol{s}_i, \boldsymbol{x}) + b \tag{2.23}$$

Bu eşitlikte, $\boldsymbol{s}_i$'ler destek vektörlerini ifade eder. Bu sayede $\Phi(\boldsymbol{x})$'leri hesaplamak yerine (2.24) denklemi kullanılır.

$$K(\boldsymbol{s}_i, \boldsymbol{x}) = \Phi(\boldsymbol{s}_i).\Phi(\boldsymbol{x}) \tag{2.24}$$

### 2.3.2 SMO (Sequential Minimal Optimization)

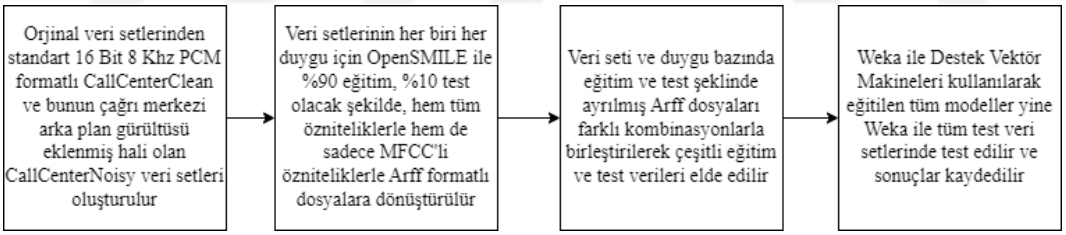
SMO, özellikle destek vektör makineleri için geliştirilmiş bir optimizasyon algoritmasıdır. SMO, destek vektör makinelerinin çekirdek metot ve Lagrange çarpanları kullanılarak gerçekleştirilen öğrenme sürecini hızlandırmakta ve daha etkili hale getirmektedir. Algoritmanın temel amacı, destek vektör makinelerini daha hızlı eğitmek ve daha iyi performans elde etmektir. [34]

Bu tez çalışmasında yapay öğrenme metot olarak Weka uygulaması üzerinden destek vektör makinelerini kullanan SMO fonksiyonu seçilmiştir.

# 3
# DENEYSEL ÇALIŞMA

## 3.1 Kurulum

Bu tez çalışmasında konuşma sesi verisinden öznitelik çıkarmak için açık kaynak kodlu OpenSMILE [35] uygulaması, etiketli veri setlerini yapay öğrenme algoritması seçerek eğitmek ve oluşan modelleri test etmek için ise yine açık kaynak kodlu bir uygulama olan Weka [36] uygulaması kullanılmıştır. Şekil 3.1'de gösterilen blok diyagramdaki deneysel çalışmalarda çok fazla sayıda veri seti oluşturulup deney yapıldığından pratik olması açısından tüm işlemler manuel yapılmak yerine script dili kullanılarak programatik olarak yapılmıştır. Script dili olarak JavaScript dili seçilmiş ve nodejs ortamında çalışılmıştır. Yazılan scriptler ile OpenSMILE ve Weka programları komut satırından çağrılmış ve döngüsel olarak çalıştırılarak otomatik olarak toplam 786 adet eğitim ve test verisi, birbirinden farklı toplam 102 model oluşturulmuş ve bu modellerle toplam 1026 adet deney yapılmıştır.




**Şekil 3.1** Deneysel çalışma adımları

## 3.2 Daha Önce Yapılmış Çalışmaların Tekrarı

Tüm veri setleri ile ayrı ayrı olarak, hepsinde bulunan ortak duygu etiketleri anger, fear, sadness, happiness, disgust ve neutral için sınıflandırma yapabilecek modeller üretilmiştir. Model üretimi için eğitim seti olarak her veri setinin verilerinin %90'ı rastgele seçilmiş olup ortaya çıkan model, veri setinin kalan %10'u ile test edilmiştir. Rastgele seçimde her etiketten eşit miktarda alınmasına dikkat edilmiştir.

Her bir veri seti için oluşturulan modelin bu altı duygulu sınıflandırma için test seti üzerinde doğru sınıflandırma oranı Tablo 3.1'de gösterilmiştir.

**Tablo 3.1** Altı duygulu sınıflandırma başarısı

| Veri Seti | Doğru Sınıflandırma Oranı |
|---|---|
| **EmoDB** | **%91.66** |
| **SAVEE** | **%79.16** |
| **RAVDESS** | **%72.72** |
| **CREMA-D** | **%58.34** |
| **TESS** | **%100** |

Tablo 3.1 tek başına yorumlandığında yapılan çalışmanın sonuçları EmoDB veri seti ve TESS veri seti için oldukça başarılı, SAVEE ve RAVDESS veri seti için kabul edilebilir seviyede başarılı, CREMA-D veri seti için ise ortalama seviyede ancak yetersiz olduğu sonucuna varabiliriz. Bu güne kadar yapılan çalışmaların temel hatası ise sadece bu sonuçları yorumlayıp yapay öğrenme yöntemi ve öznitelik seçimlerinde yapılacak değişikliklerle daha başarılı modellerin oluşturulabileceği üzerinde tartışmalarıdır.

Ancak her bir veri seti için üretilen model diğer veri setlerinin test verisi ile test edildiğinde Tablo 3.1'den çok uzak bir başarım oranı gözlemlenmektedir.

## 3.3 Daha Önce Yapılmış Çalışmaların Çapraz Testleri

Tüm veri setleri ile ayrı ayrı elde edilen modeller diğer veri setleri ile de test edildiğinde Tablo 3.2'de sonuçlar elde edilmiştir. Kolonlar veri setlerini (V), satırlar ise modellerin (M) sonuçlarını göstermektedir.

**Tablo 3.2** Altı duygulu sınıflandırmada çapraz test sonuçları

| Veri Setileri | EmoDB (V) | SAVEE (V) | RAVDESS (V) | CREMA-D (V) | TESS (V) |
|---|---|---|---|---|---|
| **EmoDB (M)** | **%91.66** | **%20.83** | **%16.36** | **%19.09** | **%28.86** |
| **SAVEE (M)** | **%20.83** | **%79.16** | **%20** | **%25.63** | **%16.26** |
| **RAVDESS (M)** | **%14.58** | **%10.41** | **%72.72** | **%18.29** | **%26.82** |
| **CREMA-D (M)** | **%45.83** | **%35.41** | **%28.18** | **%58.34** | **%36.99** |
| **TESS (M)** | **%25** | **%22.91** | **%18.18** | **%21.36** | **%100** |

Tablo 3.2'deki sonuçlar incelendiğinde her modelin yalnızca kendi test veri setinde başarılı olabildiği görülmektedir. Bu sonuçlara göre, makine öğrenmesi yoluyla duygusal anlamda tahmin edebilen modellerin geliştirilmesinin mümkün olduğu, ancak bu modelin gerçek dünyadaki problemleri başarılı bir şekilde çözebilmesi için yapılan çalışmaların yetersiz olduğu sonucuna varılmaktadır.

Tüm veri setlerinin birleştirilerek hem veri setindeki eğitim verisi sayısının arttırılması, hem de dil, kanuşmacı sayısı, kayıt ortamı, konuşulan metin gibi önemli parametrelerin çeşitlendirilmesinin sağlanması ile yeniden eğitim yapıldığında Tablo 3.2.1'de görülen problemin çözülebileceği görülmektedir. Bu şekilde yapılan deney sonuçları Tablo 3.2.2'de gösterilmiştir.

**Tablo 3.3** Genişletilmiş model için alt duygulu sınıflandırma başarısı

| Veri Seti | Doğru Sınıflandırma Oranı |
|---|---|
| **EmoDB** | **%70.83** |
| **SAVEE** | **%60.41** |
| **RAVDESS** | **%59.09** |
| **CREMA-D** | **%54.20** |
| **TESS** | **%95.93** |
| **TÜMÜ** | **%64.11** |

Tablo 3.2.2'ye bakıldığında altı duygulu sınıflandırmanın aslında gerçek hayatta Tablo 3.1'deki gibi olmayacağı anlaşılmaktadır. Genişletilmiş veri seti ile eğitilen modelin diğer modellere göre daha güvenilir bir başarım oranına sahip olduğu görülmektedir ancak gerçek hayata uygulandığında veya burada kullanılmayan başka bir veri setine uygulandığında başarımın diğer modeller kadar olmasa da yine düşeceği açıktır.

Yapılan bu deneylerden iki sonuç çıkmaktadır. Başarılı bir duygu tanıma modeli oluşturmak, eğitim verisinin mümkün olduğunca çeşitli ve yüksek sayıda olması ile ve modelin uygulanacağı dil ve kayıt ortamının eğitim verisinde kullanılan dil ve kayıt ortamı ile aynı olması ile yakından ilişkilidir. Bu iki duruma dikkat edilmesi durumunda makine öğrenmesi ile başarılı modellerin oluşturulabilmesi mümkün görünmektedir.

## 3.4 Duygu Sınıfının Azaltılması

Bu çalışmanın hedefi çağrı merkezleri için mutluluk ve sinirlilik duygularının tespit ve analizi ihtiyacının karşılanabilmesi olduğundan altılı sınıflandırma yerine, biri happiness ve neutral etiketli veriden, diğeri ise anger ve neutral etiketli veriden

üretilen iki ayrı model üretilmiştir. Modellerin başarım oranları Tablo 3.3.1'de gösterilmiştir.

**Tablo 3.4** İkili sınıflandırma başarısı

| Model | Doğru Sınıflandırma Oranı |
|---|---|
| **Happiness – Neutral** | **%89.09** |
| **Anger – Neutral** | **%93.84** |

Tablo 3.3.1'deki sonuçlara göre duygu sınıfının azaltılarak ikili sınıflandırma yapılması başarımı oldukça arttırmaktadır.

Ancak bu modellerin başarısının her iki duygu ile de etiketli olmayan diğer duygular ile test edildiğinde düştüğü görülmektedir. Bu durumda diğer duyguları neutral olarak tanıma başarısının çok düşük olduğu görülmektedir. Tablo 3.3.2, bu modellerin diğer duygular ile etiketli veriyi doğru bir şekilde neutral olarak sınıflandırıp sınıflandıramadığının başarısını göstermektedir.

**Tablo 3.5** Diğer duygular ile ikili sınıflandırma başarısı

| Model | Doğru Sınıflandırma Oranı |
|---|---|
| **Happiness – Neutral** | **%42.49** |
| **Anger – Neutral** | **%52.42** |

Tablo 3.3.2 incelendiğinde diğer duyguları içeren kayıtların bu modellerle başarılı olarak sınıflandırılamadıkları gözlemlenmektedir.

Bu sorunun çözümü için eğitim verisi iki model için de tekrar oluşturulmuş, bu sefer neutral etiketli verinin %30'u eğitim ve test setlerinden kaldırılarak yerine diğer duygular ile etiketli verilerden rastgale ve eşit olarak seçilerek yine neutral etiketi ile eklenmiştir. Yeni veri setleri ile elde edilen modeller ile yapılan testlerin sonucu Tablo 3.3.3'te gösterilmiştir.

Bu sonuçlara göre başarımın arttığı görülmektedir. Sinirlilik ve mutluluk duygularını nötr duygudan ayırt etmek için ilk modeller daha başarılı olsa da gerçek hayatta diğer duygular da olacağından bu duyguların da başarılı bir şekilde nötr olarak tanınması gerekmektedir. Bu sebeple yeni üretilen Anger-NotAnger ve Happiness-NotHappiness modelleri gerçek hayatta daha güvenilir ve daha başarılı olacaklardır.

**Tablo 3.6** Genişletilmiş model için diğer duygular ile ikili sınıflandırma

| Model | Happiness ve Anger | Diğer Duygular |
|---|---|---|
| **Happiness – NotHappiness** | **%88.17** | **%57.83** |
| **Anger – NotAnger** | **%91.87** | **%75.44** |

## 3.5 Ses Formatlarının Çağrı Merkezi Formatına Dönüştürülmesi

Bu çalışmanın amacı çağrı merkezi ihtiyaçlarını karşılamak olduğundan tüm veri setlerindeki ses kayıtları standart çağrı merkezi ses kaydı formatı olan 16 bit 8 Khz PCM formatına dönüştürülerek tüm deneyler tekrarlanmış ve sonuçlara etkisi incelenmiştir.

Tüm sonuçlar tek tek gösterilmeyecek olup Tablo 3.3.3'te gösterilen deney sonuçları için yeni veri seti oluşturulup deney tekrarlandığında Tablo 3.4'deki sonuçlar elde edilmiştir.

**Tablo 3.7** Çağrı merkezi ses formatı ile ikili sınıflandırma

| Model | Happiness ve Anger | Diğer Duygular |
|---|---|---|
| **Happiness – NotHappiness** | **%83.03** | **%54.32** |
| **Anger – NotAnger** | **%93.65** | **%74.22** |

Sonuçlar incelendiğinde sadece Anger-NotAnger modelinin sinirlilik ve nötr sınıflandırmasında başarımın arttığı diğer durumlarda azaldığı gözlemlenmiştir. Ancak çağrı merkezi uygulamaları için bu modelin kullanılması daha güvenilir olacaktır.

## 3.6 Ses Kayıtlarına Çağrı Merkezi Arka Plan Gürültüsü Eklenmesi

Çağrı merkezi ortamını daha doğru taklit edebilmek için tüm veri setlerindeki ses kayıtlarına çağrı merkezi arka plan gürültüsü de eklenerek tüm deneyler tekrarlanmıştır. Tüm sonuçlar tek tek verilmeyecek olup Tablo 3.3.3'te gösterilen deney sonuçları için yeni veri seti oluşturulup deney tekrarlandığında Tablo 3.5'deki sonuçlar elde edilmiştir.

**Tablo 3.7** Arka plan gürültüsü eklenerek ikili sınıflandırma

| Model | Happiness ve Anger | Diğer Duygular |
|---|---|---|
| **Happiness – NotHappiness** | **%84.06** | **%56.29** |
| **Anger – NotAnger** | **%91.62** | **%75.66** |

Sonuçlar incelendiğinde sadece Anger-NotAnger modelinin sinirlilik ve nötr sınıflandırmasında hafif bir düşüş, diğerlerinde ise hafif bir artış görülmektedir.

Ancak gerçek durumu daha iyi taklit ettiğinden dolayı çağrı merkezi uygulamalarında bu modelin tercih edilmesi daha güvenilir olacaktır.

## 3.7 MFCC Özniteliklerinin Ayrıca İncelenmesi

Duygu tanıma modelinin konuşma tanıma ile birlikte kullanılması durumunda öznitelik çıkarma maaliyetinden tasarruf etmek için sadece MFCC öznitelikleri kullanılarak da model oluşturulmuş ve yapılan sınıflandırma başarısı test edilmiştir. MFCC öznitelikleri zaten konuşma tanıma için çıkarılan öznitelikler olduğundan konuşma tanıma yapılan durumlar için tekrardan öznitelik çıkarma maaliyeti oluşmadan aynı özniteliklerle duygu tanıma da yapılabilir.

Bu durumu test etmek için oluşturulan veri setlerinden, sadece MFCC öznitelikler seçilerek yeni veri setleri oluşturulmuş ve yapılan deneyler tekrarlanmıştır. Sadece MFCC öznitelikleri ile elde edilen modeller test veri setleri ile test edildiğinde Tablo 3.6'daki sonuçlar gözlemlenmiştir. Tablo 3.6 incelendiğinde model başarımında her ne kadar bir miktar düşüş olsa da yerine göre CPU kaynağından tasarruf etmek için zaten konuşma tanıma yapılan uygulamalarda sadece konuşma tanıma için çıkarılan MFCC öznitelikleri kullanılarak da tekrardan başka öznitelik çıkarılmasına gerek olmadan duygu tanıma yapılması mümkün olduğu söylenebilmektedir.

**Tablo 3.8** MFCC öznitelikleri ile ikili sınıflandırma

| **Model** | **Doğru Sınıflandırma Başarısı** |
|---|---|
| **Happiness – Neutral** | **%81.03** |
| **Anger – Neutral** | **%89.23** |

# 4
# SONUÇ VE ÖNERİLER

## 4.1 Tez Çalışması Sonuçlarının Yorumu

Yapılan çalışmanın en önemli sonucu, konuşma sesinde duygu tanıma probleminin gerçek anlamıyla başarılı olarak çözümü için en önemli gereksinimin kullanılan veri setinin mümkün olduğunca büyük, çeşitli ve uygulanacak alanla uyumlu olmasıdır. Öznitelik seçimleri, sınıflandırma için seçilen makine öğrenmesi yöntemi ile başarım bir miktar daha arttırılabilmektedir. İkinci en önemli sonuç ise başarımın daha çok artabilmesi için sınıflandırılacak duygu sayısının mümkün olduğunca düşürülmesinin gerekliliğinin anlaşılmasıdır. Gerçek hayattaki problemlerde çoğunlukla tespit edilmesi istenen duygu sinirlilik olmaktadır. Bunun yanında bazen mutluluk duygusunun tespiti de önem taşıyabilmektedir. Ancak diğer duyguların tespitine nadiren ihtiyaç duyulmaktadır. Bu çalışmada gerçek bir problem olan çağrı merkezi ihtiyacına odaklanıldığından sinirlilik ve bunun yanında mutluluk duygularının tespiti üzerinde durulmuştur. Duygu sayısının azaltılması ile yapılan sınıflandırma başarısının arttığı gözlemlenmiştir. Çalışmadan çıkarılacak üçüncü en önemli sonuç ise bir çok ses işleme uygulamasında kullanılan MFCC özniteliklerinin tek başına kullanılmasıyla da başırılı modellerin üretilebilmesinin mümkün olduğunun gösterilmesidir. Bu durum, duygu tanımanın, konuşma tanıma gibi uygulamalarla birlikte kullanımında tekrar yeni özniteliklerin çıkarılması gerekmeden zaten konuşma tanıma için çıkarılmış olan MFCC öznitelikleri ile duygu tanımanın da yapılabilmesine olanak sağlamış olması açısından ve böylelikle CPU kaynağı kullanım maliyeti oluşturmaması açısından önemlidir.

## 4.2 Daha Başarılı Sonuçlar İçin Öneriler

Bu çalışmada Tablo 3.5'de sonuçları gösterilen modellerle gerçek çağrı merkezi verileri analiz edilerek sınıflandırılması durumunda başarım burada gösterilen başarımdan daha düşük olacaktır. Ancak bu modeller yine de kullanışlı bir model olarak işe yarayabilecektir. Modelin başarımının arttırılması için ise model tarafından sınıflandırılan veriler gerçek insanlar tarafından kontrol edilerek yapılan

hatalı sınıflandırmalar düzeltilmeli ve bu düzeltme ile elde edilen verilerden aynı yöntemle yeni modeller oluşturulmalıdır. Bu durumda oluşturulan bu yeni modellerin başarısının, bu çalışmada her bir veri setinin ayrı ayrı eğitilip kendi içinde test edildiğindeki gibi yüksek oranlara ulaştığı görülecektir. Sonraki aşamada ise CPU kaynağından tasarruf için sadece MFCC öznitelikleri ile model üretilmesi önerilmektedir. Son olarak ise destek vektör makineleri dışında farklı sınıflandırma algoritmaları ile de deneyler yapılarak başarımın arttırılıp arttırılamayacağının test edilmesi önerilmektedir.

Konuşma seslerinde temel frekans, yani F0, sinirlilik duygusu için önemli bilgi taşıdığından, sadece sinirlilik analizi için, veri setleri sadece F0 hesaplanarak elde edilen istatistiksel özelliklerle oluşturulan öznitelikler kullanılarak yeniden oluşturulup bir ikili sınıflandırma modeli üretilirse, hem hızlı hem de düşük maliyetli sinirlilik analizi yapabilen bir sistemin geliştirilmesinin mümkün olup olmadığı üzerinde yeni çalışmalar yapılması önerilmektedir.

# KAYNAKÇA


[1] D. Morrison, R. Wang, and L. C. De Silva, “Ensemble methods for spoken emotion recognition in call-centres,” Speech Communication, vol. 49, no. 2, pp. 98–112, 2007, doi:10.1016/j.specom.2006.11.004.

[2] E. M. Albornoz, D. H. Milone, and H. L. Rufiner, “Spoken emotion recognition using hierarchical classifiers,” Computer Speech & Language, vol. 25, no. 3, pp. 556–570, 2011, doi:10.1016/j.csl.2010.10.001.

[3] Y. Wang and L. Guan, “Recognizing human emotional state from audiovisual signals,” IEEE Transactions on Multimedia, vol. 10, no. 5, pp. 936–946, 2008, doi:10.1109/tmm.2008.927665.

[4] O. Martin, I. Kotsia, B. Macq, and I. Pitas, “The Enterface’05 Audio-Visual Emotion Database,” 22nd International Conference on Data Engineering Workshops (ICDEW’06), 2006, doi:10.1109/icdew.2006.145.

[5] C. S. Ooi, K. P. Seng, L.-M. Ang, and L. W. Chew, “A new approach of audio emotion recognition,” Expert Systems with Applications, vol. 41, no. 13, pp. 5858–5869, 2014, doi:10.1016/j.eswa.2014.03.026.

[6] B. Lindemann, T. Müller, H. Vietz, N. Jazdi, M. Weyrich, "A survey on long short-term memory networks for time series prediction", Procedia CIRP, 99, 650-655, 2021, 10.1016/j.procir.2021.03.088.

[7] S. R. Livingstone and F. A. Russo, “The Ryerson Audio-Visual Database of emotional speech and Song (RAVDESS),” Zenodo, https://zenodo.org/records/1188976, 15.12.2023.

[8] SAVEE, Surrey Audio-visual expressed emotion (SAVEE) database, http://kahlan.eps.surrey.ac.uk/savee, 15.12.2023.

[9] A. Demir, O. Atila, and A. Sengur, “Deep learning and audio based emotion recognition,” 2019 International Artificial Intelligence and Data Processing Symposium (IDAP), 2019, doi:10.1109/idap.2019.8875932.

[10] C. Busso, M. Bulut, C.-C. Lee, A. Kazamzadeh, E. Mower, S. Kim, J.N. Chang, S. Lee, S.S. Narayanan, Iemocap: interactive emotional dyadic motion capture database, Lang. Resour. Eval. 42 (2008) 335.

[11] D. Issa, M. Fatih Demirci, and A. Yazici, “Speech emotion recognition with deep convolutional neural networks,” Biomedical Signal Processing and Control, vol. 59, p. 101894, 2020, doi:10.1016/j.bspc.2020.101894.

[12] Y.Ü. Sönmez, A. Varol, “Konuşma seslerinden duyguları tanımak için yerel ikili ve üçlü örüntülere dayalı yeni bir modelin geliştirilmesi (Development of a new model based on local binary and ternary patterns to recognize emotions from speech sounds)", Doktora Tezi, Fen Bilimleri Enstitüsü, Fırat Üniversitesi, Elazığ, 2021.

[13] R. Yamashita, M. Nishio, R. K. Do, and K. Togashi, “Convolutional Neural Networks: An overview and application in Radiology,” Insights into Imaging, vol. 9, no. 4, pp. 611–629, 2018, doi:10.1007/s13244-018-0639-9.

[14] F. Gümüş, T. Çakar, “Comparing Audio Features For Speech Emotion Recognition Using Machine Learning Algorithms", Yüksek Lisans Tezi, MEF University, İstanbul, 2022.

[15] F. Murtagh, “Multilayer perceptrons for classification and regression,” Neurocomputing, vol. 2, no. 5–6, pp. 183–197, 1991, doi:10.1016/0925-2312(91)90023-5.

[16] Ö. Özdemir, G. Bilgin, “Emotion and Gender Identification on Speech Signals Using Deep Learning Methods”, Yüksek Lisans Tezi, Fen Bilimleri Enstitüsü, Yıldız Teknik Üniversitesi, İstanbul, 2022.

[17] Berlin database of emotional speech, Emo, http://emodb.bilderbar.info/start.html. 15.12.2023.

[18] TESS, Toronto Emotional Speech Set, https://tspace.library.utoronto.ca/handle/1807/24487, 15.12.2023.

[19] H. Cao et al., “Crema-d: Crowd-sourced emotional Multimodal Actors Dataset,” IEEE Transactions on Affective Computing, vol. 5, no. 4, pp. 377–390, 2014, doi:10.1109/taffc.2014.2336244.

[20] Ü. Bayraktar, Z. H. Kilimci, "Makine Öğrenimi Tekniklerini Kullanarak Konuşma Duygularını Tanıma", Yüksek Lisans Tezi, Fen Bilimleri Enstitüsü, Kocaeli Üniversitesi, Kocaeli, 2023.

[21] S. Davis and P. Mermelstein, "Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences," in IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. 28, no. 4, pp. 357-366, 1980, doi: 10.1109/TASSP.1980.1163420.

[22] R. Salami, C. Laflamme, B. Bessette and J. -P. Adoul, "Description of ITU-T Recommendation G.729 Annex A: reduced complexity 8 kbit/s CS-ACELP codec," 1997 IEEE International Conference on Acoustics, Speech, and Signal Processing, Munich, Germany, 1997, pp. 775-778 vol.2, doi: 10.1109/ICASSP.1997.596039.

[23] K. Jarvinen et al., "GSM enhanced full rate speech codec," 1997 IEEE International Conference on Acoustics, Speech, and Signal Processing, Munich, Germany, 1997, pp. 771-774 vol.2, doi: 10.1109/ICASSP.1997.596038.

[24] A. de Cheveigné, H. Kawahara, "YIN, A fundamental frequency estimator for speech and music", The Journal of the Acoustical Society of America, 111 4, 1917-30, 2002, doi:10.1121/1.1458024.

[25] Yu-Min Zeng, Zhen-Yang Wu, Hai-Bin Liu and Lin Zhou, "Modified AMDF pitch detection algorithm," Proceedings of the 2003 International Conference on Machine Learning and Cybernetics (IEEE Cat. No.03EX693), Xi'an, 2003, pp. 470-473 Vol.1, doi: 10.1109/ICMLC.2003.1264523.

[26] P. McLeod, G. Wyvill, "A smarter way to find pitch", In ICMC, Vol. 5, pp. 138-141, 2005.

[27] O. Eray, S. İplikçi, "Destek Vektör Makineleri İle Ses Tanıma Uygulaması", Yüksek Lisans Tezi, Fen Bilimleri Enstitüsü, Pamukkale Üniversitesi, Denizli, 2008.

[28] C. J. C. Burges, "A Tutorial on Support Vector Machines for Pattern Recognition", Data Mining and Knowledge Discovery, vol. 2, no. 2, pp. 121–167, 1998, doi:10.1023/a:1009715923555.

[29] C. Cortes, V. Vapnik, "Support-Vector Networks", Machine Learning 20, 273–297, 1995, doi: 10.1023/A:1022627411411.

[30] V. Blanz, B. Schölkopf, H. Bülthoff, C. Burges, V. Vapnik, T. Vetter, "Comparison of view-based object recognition algorithms using realistic 3D models", Artificial Neural Networks, ICANN 1996 - 1996 International Conference, Proceedings (pp. 251-256), 1996, doi: https://doi.org/10.1007/3-540-61510-5_45.

[31] M. Schmidt, “Identifying speaker with support vector networks”, In Interface ’96 Proceedings, Sydney, 1996.

[32] E. Osuna, R. Freund and F. Girosi, "An improved training algorithm for support vector machines", Neural Networks for Signal Processing VII. Proceedings of the 1997 IEEE Signal Processing Society Workshop, Amelia Island, FL, USA, pp. 276-285, 1997, doi: 10.1109/NNSP.1997.622408.

[33] T. Joachims, "Text categorization with support vector machines: Learning with many relevant features", European conference on machine learning pp. 137-142, Berlin, 1998,

[34] J. Platt, “Sequential minimal optimization: A fast algorithm for Training Support Vector Machines,” Microsoft Research, https://www.microsoft.com/en-us/research/publication/sequential-minimal-optimization-a-fast-algorithm-for-training-support-vector-machines, 23.11.2023.

[35] OpenSMILE, openSMILE Documentation, https://audeering.github.io/opensmile, 02.01.2024.

[36] Weka, Weka 3 - Data Mining with Open Source Machine Learning Software in Java, https://waikato.github.io/weka-site/index.html, 02.01.2024.

# TEZDEN ÜRETİLMİŞ YAYINLAR

**Konferans Bildirileri**

**1**. F. AYGÜN And B. BOLAT, "Makine Öğrenmesi İle Seste Sinirlilik Analizi," 1. BİLSEL ULUSLARARASI TURABDİN BİLİMSEL ARAŞTIRMALAR VE İNOVASYON KONGRESİ, vol.1, Mardin, Turkey, pp.615-624, 2023.