

T.C.
FIRAT ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ



**YENİ NESİL MAKİNE ÖĞRENMESİ YÖNTEMLERİ TABANLI
ORTAM SESLERİNDEN DUYGU VE ŞİDDET TESPİTİ**

Arif Metehan YILDIZ

Doktora Tezi

ADLİ BİLİŞİM MÜHENDİSLİĞİ ANABİLİM DALI

HAZİRAN 2025

T.C.
FIRAT ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

Adli Bilişim Mühendisliği Anabilim Dalı

Doktora Tezi

YENİ NESİL MAKİNE ÖĞRENMESİ YÖNTEMLERİ TABANLI
ORTAM SESLERİNDEN DUYGU VE ŞİDDET TESPİTİ

Tez Yazarı

Arif Metehan YILDIZ

Danışman

Doç. Dr. Türker TUNCER

HAZİRAN 2025

ELAZIĞ

T.C.
FIRAT ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

Adli Bilişim Mühendisliği Anabilim Dalı

Doktora Tezi

Başlığı:	Yeni Nesil Makine Öğrenmesi Yöntemleri Tabanlı Ortam Seslerinden Duygu ve Şiddet Tespiti
Yazarı:	Arif Metehan YILDIZ
İlk Teslim Tarihi:	12.05.2025
Savunma Tarihi:	16.06.2025

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:	Doç. Dr. Türker TUNCER Fırat Üniversitesi, Teknoloji Fakültesi	<i>İmza</i> Onayladım
Başkan:	Doç. Dr. Güllü BOZTAŞ Fırat Üniversitesi, Teknoloji Fakültesi	Onayladım
Üye:	Prof. Dr. Şengül DOĞAN Fırat Üniversitesi, Teknoloji Fakültesi	Onayladım
Üye:	Doç. Dr. Mehmet BAYĞIN Erzurum Teknik Üniversitesi, Mühendislik ve Mimarlık Fakültesi	Onayladım
Üye:	Dr. Öğr. Üyesi Fahrettin Burak DEMİR Bandırma Onyedli Eylül Üniversitesi, Mühendislik ve Doğa Bilimleri Fakültesi	Onayladım

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

İmza

Prof. Dr. Burhan ERGEN
Enstitü Müdürü

BEYAN

Fırat Üniversitesi Fen Bilimleri Enstitüsü tez yazım kurallarına uygun olarak hazırladığım “Yeni Nesil Makine Öğrenmesi Yöntemleri Tabanlı Ortam Seslerinden Duygu ve Şiddet Tespiti ” Başlıklı Doktora 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.

16.06.2025

Arif Metehan YILDIZ



ÖNSÖZ

Bu tezin araştırma konusu olan Yeni Nesil Makine Öğrenmesi Tabanlı Ortam Seslerinden Duygu ve Şiddet Tespiti, çevresel ses sınıflandırma, konuşmacı sayma, şiddet yoğunluğu algılama ve topluluk duygu tespiti, kamu güvenliği, duygusal hesaplama ve adli bilişim gibi alanlarda akıllı ve erişilebilir sistemlere duyulan kritik ihtiyacı ele almaktadır. Bu araştırma, dijital sistemlerin duyarlılığını artırabileceği ve sağlık izleme, akıllı gözetim ve davranış analizi gibi uygulamalarda değerli destek sağlayabileceği için büyük önem taşımaktadır. Önemine rağmen, araştırma konusu kayda değer zorluklar ortaya koymuştur. Bu zorluklar arasında gürültülü, belirsiz ve dinamik ses sinyallerini işlemenin doğasında var olan ve özellik çıkarma ve sınıflandırmayı karmaşıklaştıran zorluklar yer alıyordu. Ayrıca, uygun şekilde dengelenmiş ve göreve özgü veri kümelerinin azlığı, birkaç orijinal veri kümesinin titizlikle oluşturulmasını ve açıklanmasını gerektirmiştir. Hesaplama verimliliği sağlarken yüksek sınıflandırma performansı elde etmek, kapsamlı deneyler ve titiz değerlendirme gerektiren bir süreç olan yeni, el yapımı hafif modellerin geliştirilmesini gerektirmiştir.

Araştırma öncelikle gerçek dünyada etki yaratma potansiyeli ile motive edilmiştir. Hem günlük hem de kritik ortamlarda çalışabilen, yorumlanabilir, kaynakları verimli kullanan ve konuşlandırılabilir sistemler tasarlama ihtimali itici bir güç olmuştur. Savunmasız bireylere yardımcı olabilecek veya adli bilişimde dijital kanıt analizini destekleyebilecek teknolojilere katkıda bulunmak, bu çalışmanın altında yatan güçlü bir amaç sağlamıştır.

Uzman rehberliği ve sürekli teşviki için doktora danışmanım Doç. Dr. Türker TUNCER'e, temel akademik desteği ve mentörlüğü için bölüm başkanımız ve yüksek lisans tez danışmanım Prof. Dr. Şengül DOĞAN'a, çalışma hayatımdan dolayı akademik olarak bölümümden uzakta geçirdiğim zamanlarda akademik kariyerimin devamlılığına sonsuz desteklerinden dolayı Doç. Dr. Mehmet BAYĞIN'a derin şükranlarımı sunarım. Adli Bilişim Mühendisliği Bölümü değerli hocalarıma tezin kapsam ve kalitesine önemli ölçüde katkıda bulunan değerli iş arkadaşlarıma teknik yardımları için teşekkür ederim. Bu katkılar, doktora araştırmamın tamamlanmasında etkili olmuş ve bu alandaki akademik yolculuğumu zenginleştirmiştir.

Arif Metehan YILDIZ

ELAZIĞ, 2025

İÇİNDEKİLER

Sayfa

ÖNSÖZ.....	iv
İÇİNDEKİLER	v
ÖZET	ix
ABSTRACT	x
ŞEKİLLER LİSTESİ	xi
TABLOLAR LİSTESİ	xiii
KISALTMALAR	xiv
1. GİRİŞ	1
2. ARKA PLAN VE LİTERATÜR İNCELEMESİ.....	4
2.1. Ortam Seslerine Genel Bakış.....	4
2.1.1. Ortam Seslerinin Türleri.....	4
2.2. Ortam Seslerinin Analizindeki Zorluklar	5
2.2.1. Çakışan Sesler (Overlapping).....	5
2.2.2. Gürültü ve Sinyal-Gürültü Oranı.....	5
2.2.3. Akustik Çeşitlilik.....	6
2.2.4. Zamansal Dinamikler	6
2.2.5. Veri Kümesi Sınırlamaları.....	6
2.2.6. Kültürel ve Bağlamsal Değişimler	7
2.3. Ortam Sesi Analizinin Uygulamaları.....	7
2.3.1. Kamu Güvenliği ve Emniyet.....	7
2.3.2. Sağlık Hizmetleri ve Refah	8
2.3.3. Kentsel Planlama ve Gürültü Yönetimi.....	8
2.3.4. Akıllı Ev ve Nesnelerin İnterneti (IoT)	8
2.3.5. Yaban Hayatı İzleme ve Koruma	9
2.3.6. Eğlence ve Oyun	9
2.3.7. Afet Yönetimi.....	9
2.3.8. Değerlendirme.....	10
2.4. Seslerden Duygu Tanıma.....	10
2.4.1. Konuşma Tabanlı Duygu Tanıma	10
2.4.2. Konuşma Tabanlı Olmayan Duygu Tanıma	13
2.4.3. Kombine Konuşma ve Konuşma Dışı Duygu Tanıma	16
2.4.4. Seslerden Duygu Tanıma Uygulamaları	17
2.4.5. Literatürdeki Boşluklar	22
2.4.6. Literatür Boşlukların Üstesinden Gelmek için Geleceğe Yönelik Öneriler	29
2.4.7. Değerlendirme.....	31
3. SES ANALİZİNİN TEMELLERİ	32
3.1. Seslerin Akustik Özellikleri.....	32
3.1.1. Ses Perdesi	32
3.1.2. Ses Tınısı	33
3.1.3. Ses Yüksekliği.....	34
3.1.4. Sesin Yoğunluğu	36
3.1.5. Değerlendirme.....	37

3.2. Sesten Özellik Çıkarma Teknikleri.....	38
3.2.1. Mel-Frekans Cepstral Katsayıları (MFCC)	39
3.2.2. Spektral Kontrast.....	40
3.2.3. Sıfır Geçiş Oranı (ZCR)	43
3.2.4. Kroma Özellikleri.....	44
3.2.5. Spektral Merkez	46
3.2.6. Spektral Bant Genişliği	47
3.2.7. Spektral Düşüş Eğimi.....	48
3.2.8. Değerlendirme.....	50
3.3. Duygu ve Şiddet Özellikleri	51
3.3.1. Duygu ve Şiddet Özelliklerinin Temel Yönleri.....	51
3.3.2. Konuşma Seslerinde Duygusal Korelasyon	52
3.3.3. Konuşma Dışı Seslerdeki Duygusal Korelasyon.....	55
3.3.4. Duygu ve Şiddet Tespiti için Makine Öğrenimi Yaklaşımları	57
3.3.5. Seslerde Duygu ve Şiddet Analizinin Uygulamaları	58
4. METODOLOJİ	72
4.1. Makine Öğrenmesi	72
4.1.1. Makine Öğrenmesi Türleri	72
4.1.2. Makine Öğrenmesi Bileşenleri	73
4.1.3. Makine Öğrenmesi Uygulamaları	74
4.1.4. Makine Öğrenmesi Çalışma Mantığı.....	75
4.1.5. Makine Öğrenmesinin Zorlukları.....	77
4.1.6. Makine Öğrenmesinin Önemi	78
4.2. Derin Öğrenme	78
4.2.1. Derin Öğrenmenin Türleri.....	79
4.2.2. Derin Öğrenmenin Bileşenleri.....	79
4.2.3. Derin Öğrenmenin Uygulama Alanları	79
4.2.4. Derin Öğrenmenin Çalışma Mantığı	80
4.2.5. Derin Öğrenmenin Zorlukları.....	80
4.2.6. Derin Öğrenmenin Önemi	81
4.3. Makine Öğrenmesi ve Derin Öğrenmede Kullanılan Yöntemler	81
4.3.1. Örüntü Tanıma ve Özellik Çıkarması	81
4.3.2. Optimizasyon ve Kayıp Fonksiyonları	85
4.3.3. Model Değerlendirme ve Özellik Seçimi	86
4.3.4. Sınıflandırma.....	89
5. DİĞER YÖNTEMLER.....	91
5.1. Ses Analizi için Derin Öğrenme	91
5.1.1. Ses Spektrogramlarından Uzamsal Özellik Çıkarması için Evrimsel Sinir Ağları	91
5.1.2. Ses Dizilerindeki Zamansal Bağımlılıkları Modellemek için Tekrarlayan Sinir Ağları	91
5.2. Ses Analizinde Dikkat Mekanizmaları	92
5.2.1. Belirgin Ses Özelliği Tanımlama için Dönüştürücü Mimarilerinin Uygulanması	92
5.3. Ses Analizinde Transfer Öğrenimi	93
5.3.1. Önceden Eğitilmiş Ses Özellik Çıkarma Modellerinin Kullanımı	93
5.4. Gelişmiş Tahmin Sağlamlığı için Topluluk Öğrenme Stratejileri	94
5.4.1. Gelişmiş Performans Optimizasyonu için Model Toplama.....	94
5.5. Değerlendirme	95

6. MATERYAL VE METOT	96
6.1. Yeni Bir Lateral Genikülat Çekirdek Örüntüsü Tabanlı Çevresel Ses Sınıflandırması	97
6.1.1. Literatür Taraması Özeti	97
6.1.2. Mevcut Çalışmalar	98
6.1.3. Mevcut çalışmalardaki eksiklikler.....	98
6.1.4. Çalışmada Kullanılan Veri Seti	98
6.1.5. Çalışmada Önerilen Yöntem	99
6.1.6. Çalışmanın Performans Değerlendirmesi.....	102
6.1.7. Çalışmanın Katkıları	102
6.1.8. Çalışmanın Sonuçları ve Tartışma.....	103
6.2. BP19: Tek Boyutlu İkili Örüntüye Dayalı Doğru Bir Ses Şiddeti Tespit Modeli.....	103
6.2.1. Literatür Taraması Özeti	104
6.2.2. Literatürdeki Mevcut Çalışmalar.....	104
6.2.3. Mevcut çalışmalardaki Eksiklikler	104
6.2.4. Çalışmada Kullanılan Veri Seti	105
6.2.5. Çalışmada Önerilen Yöntem	106
6.2.6. Çalışmanın Performans Değerlendirmesi.....	107
6.2.7. Çalışmanın Katkıları	109
6.2.8. Çalışmanın Sonuçları ve Tartışma.....	109
6.3. James Webb Modeline Dayalı Doğru Bir Otomatik Konuşmacı Sayma Mimarisi	110
6.3.1. Literatür Taraması Özeti	110
6.3.2. Literatürdeki Mevcut Çalışmalar.....	111
6.3.3. Mevcut çalışmalardaki Eksiklikler	111
6.3.4. Çalışmada Kullanılan Veri Seti	112
6.3.5. Çalışmada Önerilen Yöntem	112
6.3.6. Çalışmanın Performans Değerlendirmesi.....	115
6.3.7. Çalışmanın Katkıları	117
6.3.8. Çalışmanın Sonuçları ve Tartışma.....	117
6.4. SKLBP14: Kare Çekirdekli Yerel İkili Örüntüye Dayalı Yeni Bir Dokusal Çevresel Ses Sınıflandırma Modeli	119
6.4.1. Literatür Taraması Özeti	119
6.4.2. Literatürdeki Mevcut Çalışmalar.....	119
6.4.3. Mevcut çalışmalardaki Eksiklikler	120
6.4.4. Çalışmada Kullanılan Veri Seti	120
6.4.5. Çalışmada Önerilen Yöntem	121
6.4.6. Çalışmanın Performans Değerlendirmesi.....	123
6.4.7. Çalışmanın Katkıları	125
6.4.8. Çalışmanın Sonuçları ve Tartışma.....	125
6.5. Ses Sinyallerini Kullanan Yeni Bir Ağaç Örüntüsü Tabanlı Şiddet Tespit Modeli.....	126
6.5.1. Literatür Taraması Özeti	127
6.5.2. Literatürdeki Mevcut Çalışmalar.....	127
6.5.3. Literatürdeki Eksiklikler	127
6.5.4. Çalışmada Kullanılan Veri Seti	128
6.5.5. Çalışmada Önerilen Yöntem	128
6.5.6. Çalışmanın Performans Değerlendirmesi.....	132
6.5.7. Çalışmanın Katkıları	133
6.5.8. Çalışmanın Sonuçları ve Tartışma.....	134

6.6. Yeni Ses Tabanlı Topluluk Duygu Tespiti için FF-BTP Modeli	134
6.6.1. Literatür Taraması Özeti	135
6.6.2. Literatürdeki Mevcut Çalışmalar	135
6.6.3. Literatürdeki Eksiklikler	136
6.6.4. Çalışmada Kullanılan Veri Seti	136
6.6.5. Çalışmada Önerilen Yöntem	137
6.6.6. Çalışmanın Performans Değerlendirmesi	139
6.6.7. Çalışmanın Katkıları	141
6.6.8. Çalışmanın Sonuçları ve Tartışma	142
6.7. FF-Orbital Kaos Tabanlı Özellik Çıkarma Modeli Kullanılarak Toplumun Duygularının Ses Yoluyla Tespiti Üzerine Bir Araştırma	142
6.7.1. Literatür Taraması Özeti	143
6.7.2. Literatürdeki Mevcut Çalışmalar	143
6.7.3. Literatürdeki Eksiklikler	144
6.7.4. Çalışmada Kullanılan Veri Seti	144
6.7.5. Çalışmada Önerilen Yöntem	145
6.7.6. Çalışmanın Performans Değerlendirmesi	148
6.7.7. Çalışmanın Katkıları	148
6.7.8. Çalışmanın Sonuçları ve Tartışma	149
7. BULGULAR VE TARTIŞMA	150
7.1. El Yapımı ve Kaostan Esinlenen Modellere Genel Bakış	150
7.2. Temel Performans Çıktıları	150
7.3. Yeni Veri Seti Katkıları	151
7.4. Modellerin Karşılaştırmalı Avantajları	152
7.5. Pratik Çıkarımlar ve Kullanım Örnekleri	152
7.6. Gelecekteki Araştırma Yönelimleri	153
7.7. Literatürdeki Benzer Çalışmalarla Karşılaştırılması	153
7.7.1. Metodolojik Yaklaşım	153
7.7.2. Veri Seti Yeniliği ve Çeşitliliği	154
7.7.3. Hesaplama Verimliliği	154
7.7.4. Genelleştirme ve Ölçeklenebilirlik	155
7.7.5. Karşılaştırmalı Avantajların Özeti	156
8. SONUÇLAR	157
8.1. Temel Katkıların Özeti	157
8.2. Temel Bulgular	158
8.3. Yaygın Etki ve Uygunluk	158
ÖNERİLER	160
KAYNAKLAR	161
ÖZGEÇMİŞ	

ÖZET

Yeni Nesil Makine Öğrenmesi Yöntemleri Tabanlı Ortam Seslerinden Duygu ve Şiddet Tespiti

Arif Metehan YILDIZ

Doktora Tezi

FIRAT ÜNİVERSİTESİ
Fen Bilimleri Enstitüsü

Adli Bilişim Mühendisliği Anabilim Dalı

Haziran 2025, Sayfa: xv + 178

Bu tez, çevresel ses sınıflandırma, konuşmacı sayma, şiddet yoğunluğu algılama ve topluluk duygu tespiti gibi çeşitli ses tabanlı tanıma görevleri için hafif, yorumlanabilir ve yüksek performanslı makine öğrenimi modellerinin geliştirilmesi için kapsamlı bir çerçeve sunmaktadır. Her biri el yapımı, hafif sıklet, örüntü tanıma veya kaostan ilham alan özellik çıkarma tekniklerini, optimize edilmiş sınıflandırma stratejilerini ve DWT ve TQWT gibi çok seviyeli sinyal dönüşümlerini içeren yedi yeni model önerilmiştir.

Her model, kendi görevleri için özel olarak oluşturulmuş yeni, dengeli ve halka açık veri kümeleri üzerinde eğitilmiş ve değerlendirilmiştir. Tüm modeller GPU hızlandırmasına ihtiyaç duymadan çalışarak düşük hesaplama kaynaklarıyla yüksek doğrulukta ses sınıflandırmasının mümkün olduğunu göstermektedir.

Genel olarak bu tez, gerçek zamanlı, gömülü veya düşük kaynaklı ortamlarda kullanıma uygun açıklanabilir, verimli ve ölçeklenebilir modeller sunarak derin öğrenmeye güçlü bir alternatif sunmaktadır. Bu tez çalışması yeni nesil makine öğrenmesi modelleriyle duygu ve şiddet tespiti alanına önemli bir katkı sağlamaktadır. Önerilen çözümler kamu güvenliği, adli bilişim, akıllı gözetim, sağlık izleme ve duygusal hesaplama alanlarında gelişmiş uygulamaların önünü açmaktadır.

Anahtar Kelimeler: Duygu tanıma, Şiddet tespiti, Çevresel ses sınıflandırması, Makine öğrenmesi, Ses işleme, Ses adli bilişimi.

ABSTRACT

Emotion and Violence Detection from Environmental Sounds Based on Next Generation Machine Learning Methods

Arif Metehan YILDIZ

Ph.D. Thesis

FIRAT UNIVERSITY
Graduate School of Natural and Applied Sciences
Department of Digital Forensic Engineering

June 2025, Pages: xv + 178

This thesis presents a comprehensive framework for the development of lightweight, interpretable, and high-performance machine learning models for various audio-based recognition tasks, such as environmental audio classification, speaker counting, intensity detection, and ensemble emotion recognition. Seven new models are proposed, each incorporating handcrafted, lightweight, pattern recognition or chaos-inspired feature extraction techniques, optimized classification strategies, and multi-level signal transformations such as DWT and TQWT.

Each model is trained and evaluated on novel, balanced and publicly available datasets specifically created for their respective tasks. All models run without the need for GPU acceleration, demonstrating that high accuracy audio classification is possible with low computational resources.

Overall, this thesis presents a powerful alternative to deep learning by providing explainable, efficient and scalable models suitable for use in real-time, embedded or low-resource environments. This thesis makes a significant contribution to the field of emotion and violence detection with next generation machine learning models. The proposed solutions pave the way for advanced applications in public safety, forensics, intelligent surveillance, health monitoring and emotional computing. The proposed solutions pave the way for advanced applications in public safety, forensics, intelligent surveillance, health monitoring and emotional computing.

Keywords: Emotion recognition, Violence detection, Environmental sound classification, Machine learning, Audio processing, Sound forensic.

ŞEKİLLER LİSTESİ

Sayfa

Şekil 6.1.	Önerilen LGNPat-MFMDWT tabanlı ses sınıflandırma modeline genel bakış [128].	100
Şekil 6.2.	Şiddet içeren ve şiddet içermeyen ses segmentlerinin genlik gösterimi [156].	106
Şekil 6.3.	BP özellik çıkarıcının grafiksel açıklaması [156]	107
Şekil 6.4.	BP19 için ablasyon sonuçları ve karışıklık matrisi [156]	108
Şekil 6.5.	JWPat modeline ait akış diyagramı [162]	113
Şekil 6.6.	a) 10 seviyeli dengesiz ağaç yapısı MDWT, 10 düşük geçişli filtre (L) ve 10 yüksek geçişli filtre (H) dalgacık katsayısı üretmek için kullanıldı. b) James Webb uzay teleskopu aynası yapılandırmasına dayalı JWPat. Giriş değeri olarak merkez, SM kullanılır. A, B ve C gruplarındaki altıgen elemanlar, SM etrafında radyal olarak düzenlenir ve saat yönünde ayrı ayrı numaralandırılır [162].	114
Şekil 6.7.	Duyarlılığın sınıf sayısına göre grafiği [162].	116
Şekil 6.8.	Yeni bir veri seti kullanılarak önerilen JWPat tabanlı konuşmacı sayma modeli için elde edilen karışıklık matrisi [162].	118
Şekil 6.9.	Önerilen modelin şematik gösterimi. MDWT: çok seviyeli ayrık dalgacık dönüşümü, L: alçak geçiren filtre dalgacık katsayıları (dalgacık bantları), F: 256 boyutlu özellik vektörleri [184]	121
Şekil 6.10.	Önerilen SKLBP özellik üreticinin genel grafiksel görünümü [184]	122
Şekil 6.11.	Önerilen SKLBP tabanlı TESC modelinin sınıf bazında sınıflandırma performansları [184]	124
Şekil 6.12.	Önerilen SKLBP tabanlı dokusal ESC modelinin sınıfa özgü karışıklık matrisi [184]	124
Şekil 6.13.	Sunulan TreePat23'ün grafiksel açıklaması (a) modele genel bakış, (b) ağaç desenine dayalı özellik çıkarımı [210].	129
Şekil 6.14.	Önerilen ağaç deseninin genel görünümü. Bu çalışmada, b bloğun değerlerini tanımlar ve kenarlardaki sayılar bit çıkarma sırasını gösterir. ($b_2 \leq b_3$) koşulu kullanılarak kırmızı veya mavi alan seçilir [210].	129
Şekil 6.15.	Özellik seçicilerin ve sınıflandırıcıların sınıflandırma doğrulukları. Bu sonuçlar seçilen sekiz özellik vektörü kullanılarak hesaplanmıştır [210].	131
Şekil 6.16.	Önerilen modelin ince ayarlı sınıflandırıcılar kullanılarak oluşturulan karışıklık matrisleri. **1: şiddet var, 2: şiddet yok [210].	132
Şekil 6.17.	FF-BTP tabanlı öznetelik mühendisliği modeli önerildi. * L: Alçak geçiren filtre bantları, FF-BTP: önerilen ileri-ileri ikili üçlü model, f: özellik vektörü, INCA: yinelemeli komşu bileşen analizi özellik seçicisi, SVM: destek vektör makinesi sınıflandırıcısı [125].	137
Şekil 6.18.	İleriye Doğru İkili Üçlü Örüntü (FF-BTP) Özellik Mühendisliği Şeması. *BTP transf.: BTP dönüşümü, S: signum fonksiyonu, LT: alt üçlü, UT: üst üçlü, f: öznetelik vektörü, fv: seçilen öznetelik vektörü, FF: ileri-ileri [125].	138
Şekil 6.19.	Model girişine göre katmanlandırılmış INCA tarafından seçilen özelliklerin dağılımı (ham sinyal veya L1 ile L9 arasındaki dalgacık bantları) [125].	139

- Şekil 6.20.** FF-BTP tabanlı öznitelik mühendisliği modelinin karışıklık matrisi. *1: Olumsuz; 2: Nötr; 3: Pozitif [125]. 140
- Şekil 6.21.** Önerilen FF-Orbital'in grafiksel özellikleri. Burada, Pat: Desen, H: Histogram [236] 146
- Şekil 6.22.** FF-Orbital tabanlı öznitelik mühendisliği modeline ve UTMDWT dönüşümüne genel bakış. Burada: w: dalgacık bandı, f: özellik vektörü, L: düşük geçiren filtre dalgacık bandı, H: yüksek geçiren filtre dalgacık bandı [236]. 147



TABLÖLAR LİSTESİ

	Sayfa
Tablo 3.1. Sesin temel akustik özellikleri	37
Tablo 3.2. Sesten özelliklerin çıkarılması ile ilgili yöntemler.....	50
Tablo 6.1. Fırat ESC-50 veri seti [128].....	99
Tablo 6.2. İstatistiksel özellikler	101
Tablo 6.3. LGNPat modeline ait Performans değerlendirme sonuçları [128].....	102
Tablo 6.4. BP19 üretilirken kullanılan parametreler [156]	108
Tablo 6.5. BP19 modelinde hesaplanan performans değerlendirme sonuçları [156].....	109
Tablo 6.6. Literatürdeki benzer çalışmaların özeti [162]	111
Tablo 6.7. JWPat modeli için toplanan veri setine ait bilgiler [162].....	112
Tablo 6.8. JWPat performans değerlendirme sonuçları [162].....	115
Tablo 6.9. JWPAT modelinde farklı ana dalgacık fonksiyonu kullanılarak elde edilen performanslar [162].....	115
Tablo 6.10. Otomasyon için en son yöntemlerle elde edilen sonuçların karşılaştırılması [162].....	117
Tablo 6.11. Oluşturulan veri setine ait bilgiler.....	120
Tablo 6.12. Önerilen SKLBP tabanlı dokusal ESC modelinin performans sonuçları (%) [184]	123
Tablo 6.13. Önerilen SKLBP modelinin karşılaştırmalı sonuçları.....	125
Tablo 6.14. LORO CV ile kNN birlikte kullanılarak ortaya çıkan sınıflandırma sonuçları (%) [210].	133
Tablo 6.15. LORO CV ile SVM birlikte kullanılarak ortaya çıkan sınıflandırma sonuçları (%) [210]. ...	133
Tablo 6.16. Oluşturulan SCED veri setine ait bilgiler [125].....	136
Tablo 6.17. FF-BTP tabanlı modelin sınıf bazında ve genel sınıflandırma performansı [125].....	140
Tablo 6.18. FF-BTP tabanlı modelin ablasyon çalışmalarında sınıflandırma doğrulukları [125].....	141
Tablo 6.19. Oluşturulan CED veri setine ait bilgiler.....	145
Tablo 6.20. FF-Orbital özellik mühendisliği modelinden özellik çıkarma adımları	146
Tablo 6.21. Önerilen FF-Orbital kaos tabanlı modelin performans sonuçları [236]	148
Tablo 7.1. Tüm modellerde elde edilen sınıflandırma doğrulukları.....	151
Tablo 7.2. Güçlü yönlerin karşılaştırmalı özeti.....	155
Tablo 7.3. Karşılaştırmalı avantajların kısa özeti.....	156

KISALTMALAR

AI	: Artificial Intelligence / Yapay Zekâ
ANN	: Artificial Neural Networks / Yapay Sinir Ağları
AR	: Augmented Reality / Artırılmış Gerçeklik
ASR	: Automatic Speech Recognition / Otomatik Konuşma Tanıma
AVD	: Audio Violence Detection Dataset / Ses Tabanlı Şiddet Tespit Veriseti
BERT	: Bidirectional Encoder Representations from Transformers / Çift Yönlü Kodlayıcı Temsilleri
Bi-RNN	: Bidirectional Recurrent Neural Networks / Çift Yönlü Tekrarlayan Sinir Ağları
BLSTM	: Bidirectional Long Short-Term Memory / Çift Yönlü Uzun Kısa Süreli Bellek
CASA	: Computational Auditory Scene Analysis / Hesaplamalı İşitsel Sahne Analizi
CED	: Community Emotion Detection / Topluluk Duygu Tespit
CCTV	: Closed-Circuit Television / Kapalı Devre Kamera Sistemi
CNN	: Convolutional Neural Networks / Evrimsel Sinir Ağları
dB	: Decibel / Desibel
DCT	: Discrete Cosine Transform / Ayırık Kosinüs Dönüşümü
DNN	: Deep Neural Networks / Derin Sinir Ağları
ESC	: Environmental Sound Classification / Çevresel Ses Sınıflandırması
FFT	: Fast Fourier Transform / Hızlı Fourier Dönüşümü
FinTech	: Financial Technology / Finansal Teknoloji
FPGA	: Field Programmable Gate Array / Saha Programlanabilir Kapı Dizisi
GAN	: Generative Adversarial Networks / Üretken Çekişmeli Ağlar
GBV	: Gender-Based Violence / Cinsiyete Dayalı Şiddet
GMM	: Gaussian Mixture Models / Gauss Karışım Modelleri
GRU	: Gated Recurrent Units / Geçitli Tekrarlayan Üniteler
GPU	: Graphics Processing Unit / Grafik İşlemci Birimi
HMM	: Hidden Markov Model / Gizli Markov Modeli
HNR	: Harmonic-to-Noise Ratio / Harmonik Gürültü Oranı
Hz	: Hertz / Frekans Birimi
IoT	: Internet of Things / Nesnelerin İnterneti
JWST	: James Webb Space Telescope / James Webb Uzay Teleskobu
KHz	: Kilohertz / Frekans Birimi
K-Means	: K-Means Clustering Algorithm / K-Ortalama Kümeleme Algoritması
kNN	: K-Nearest Neighbors / K-En Yakın Komşular Algoritması
LSTM	: Long Short-Term Memory / Uzun Kısa Süreli Bellek
MFCC	: Mel-Frequency Cepstral Coefficients / Mel-Frekans Cepstral Katsayıları
MSE	: Mean Squared Error / Ortalama Kare Hatası
NLP	: Natural Language Processing / Doğal Dil İşleme
OSD	: Overlapped Speech Detection / Örtüşen Konuşma Algılama
PCA	: Principal Component Analysis / Temel Bileşen Analizi
ReLU	: Rectified Linear Unit / Doğrultulmuş Doğrusal Birim
RNN	: Recurrent Neural Networks / Tekrarlayan Sinir Ağları
SAN	: Self-Attention Networks / Öz Dikkat Mekanizmalı Ağlar
SCED	: Sound-Based Community Emotion Detection / Ses Tabanlı Topluluk Duygu Tespiti
SER	: Speech Emotion Recognition / Konuşma Tabanlı Duygu Tanıma
SGD	: Stochastic Gradient Descent / Stokastik Gradyan İnişi

SNR	: Signal-to-Noise Ratio / Sinyal-Gürültü Oranı
STFT	: Short-Time Fourier Transform / Kısa Zamanlı Fourier Dönüşümü
SVM	: Support Vector Machines / Destek Vektör Makineleri
TPU	: Tensor Processing Unit / Tensör İşlem Birimi
VR	: Virtual Reality / Sanal Gerçeklik
YSA	: Yapay Sinir Ağları
YZ	: Yapay Zekâ
ZCR	: Zero Crossing Rate / Sıfır Geçiş Oranı



1. GİRİŞ

Makine öğrenimi alanındaki ortam sesi analizi kritik bir zorluk ortaya çıkmaktadır. Çevresel ses sinyallerinden duyguları ve şiddeti doğru bir şekilde tespit etmek oldukça zordur. Kentsel ortamlar, güvenlik endişeleri, sosyal huzursuzluk ve duygusal refah için giderek daha fazla çözüm gerektirmektedir. İnsan duyguları ve şiddet olayları bağırma, ağlama ve kavga gibi seslere yansırken bu nüanslı örüntüleri gürültülü, üst üste binen ses sinyallerinden sağlam bir şekilde çıkarmak önemli bir engel şeklinde karşımıza çıkmaya devam etmektedir. Bu tez, bu zorluğun üstesinden gelmek için doğru, verimli ve ölçeklenebilir makine öğrenimi yöntemleri geliştirmeye odaklanarak ortam seslerinden gerçek zamanlı duygu ve şiddet tespitini mümkün kılmaktadır [1].

Ortam seslerinden duyguları ve şiddeti doğru bir şekilde tespit etme kapasitesi, kamu güvenliği, ruh sağlığı ve sosyal refah alanlarındaki potansiyel uygulamalarla önemli bir toplumsal değere sahiptir. Okullar, hastaneler ve kentsel alanlar gibi ortamlarda, gerçek zamanlı ses tabanlı sistemler potansiyel şiddet için bir erken uyarı sistemi olarak hizmet edebilir, zamanında müdahaleleri kolaylaştırabilir ve duygusal sıkıntıya ilişkin değerli bilgiler sağlayabilmektedir. Ayrıca bu sistemler, bireyin duygusal durumu hakkında objektif veriler sunarak terapötik ortamlarda ruh sağlığı uzmanlarına yardımcı olabilmektedir.

Bununla birlikte, mevcut yaklaşımlar sınırlı veri kümelerine dayanma, çeşitli akustik ortamlarda zayıf genelleme ve örtüşen seslerin olduğu senaryolarda düşük doğruluk gibi sınırlamalarla karşı karşıyadır. Bu zorlukların üstesinden gelmek yalnızca teknolojik bir ilerleme değil, aynı zamanda topluluklarda güvenliği artırmak ve genel refahı iyileştirmek için yeni araçlar yaratma potansiyeli ile sosyal açıdan etkili bir çaba olarak karşımıza çıkmaktadır.

Bu tezin temel amacı, ortam ses sinyallerindeki duyguların ve şiddetin doğru ve sağlam bir şekilde tanımlanması için yeni makine öğrenimi yöntemleri tasarlamak ve uygulamaktır. Bu araştırma, gerçek dünya senaryolarındaki karmaşıklıkları etkili bir şekilde ele alabilen modeller geliştirmek için derin öğrenme mimarileri, gelişmiş özellik çıkarma yöntemleri ve topluluk öğrenme yaklaşımları dâhil olmak üzere en son tekniklerden yararlanmaya çalışmaktadır. Bu karmaşıklıklar gürültülü ortamları, farklı kültürel bağlamları ve değişken akustik koşulları kapsamaktadır. Ayrıca bu araştırma, geliştirilen bu yöntemlerin kamu gözetim sistemleri, sağlık ortamları ve kentsel güvenlik girişimleri gibi gerçek dünya bağlamlarında uygulanması yoluyla uygulanabilirliğini ve ölçeklenebilirliğini göstermeyi amaçlamaktadır.

Bu araştırma, ortam seslerindeki farklı ve tanımlanabilir akustik özelliklerin belirli duygusal durumlar ve şiddet içeren faaliyetlerle ilişkili olduğu hipotezine dayanmaktadır. Derin sinir ağları, dikkat mekanizmaları ve transfer öğrenme teknikleri de dâhil olmak üzere çağdaş makine öğrenimi paradigmalarının gürültü, üst üste binen sesler ve değişen sinyal yoğunlukları gibi zorlu akustik koşulların varlığında bile bu kalıpları etkili bir şekilde öğrenme kapasitesine sahip olduğu

varsayılmaktadır. Ayrıca araştırma, gelişmiş özellik seçim tekniklerinin kullanılmasının ve hibrit model mimarilerinin geliştirilmesinin, geleneksel yaklaşımlara kıyasla sınıflandırma doğruluğunu önemli ölçüde artıracak ve yanlış pozitiflerin oluşumunu azaltacağını öne sürmektedir.

Bu hipotezleri deneysel olarak doğrulamak için çok aşamalı bir metodolojik yaklaşım kullanılmaktadır. Başlangıçta, kapsamlı ortam sesi veri kümeleri titizlikle toplanır ve gürültü parazitini azaltmak ve veri formatlarını standartlaştırmak için titiz ön işleme tekniklerine tabi tutulur. Mel-frekans cepstral katsayıları (MFCC'ler), spektral kontrast ve sıfır geçiş oranları dâhil olmak üzere ilgili akustik özellikler, duygusal ve şiddetli ses olaylarının ayırt edici özelliklerini yakalamak için titizlikle çıkarılır. Ardından, tahmin modellerini eğitmek ve titizlikle doğrulamak için evrimsel sinir ağları (CNN'ler), tekrarlayan sinir ağları (RNN'ler) ve transformatör tabanlı modelleri kapsayan çeşitli makine öğrenimi algoritmaları kullanılır [2-4]. Bu modeller, doğruluk, kesinlik, geri çağırma ve F1 skoru dâhil olmak üzere kapsamlı bir performans ölçütleri paketi kullanılarak, yerleşik kıyaslama veri kümelerine karşı sistematik olarak değerlendirilmekte ve gerçek dünya senaryolarında detaylı testlere tabi tutulmaktadır. Çalışma ayrıca, genel sağlamlığı ve güvenilirliği artırmak için birden fazla modelin tahminlerini birleştiren topluluk öğrenme tekniklerinin etkinliğini de araştırmaktadır.

Ses sinyallerinden duygu tespiti alanındaki mevcut araştırmalar, öncelikle konuşma analizine odaklanmış ve sözel olmayan ortam seslerinin araştırılmasında önemli bir boşluk bırakmıştır. Destek vektör makineleri (SVM'ler) ve rastgele orman sınıflandırıcıları gibi geleneksel makine öğrenimi algoritmaları, konuşmalardan duygusal durumları tespit etmede etkinlik göstermiş olsa da performansları genellikle gürültü varlığında veya sözlü olmayan ses kaynaklarına uygulandığında kötüleşmektedir [5-7]. Buna karşılık, şiddet tespiti ile ilgili araştırmalar ağırlıklı olarak video tabanlı gözetim sistemlerine odaklanmış ve yalnızca ses metodolojileri sınırlı bir şekilde incelenmiştir. Derin öğrenmedeki son gelişmeler, özellikle de CNN'ler ve RNN'ler, bu sınırlamaları ele alma konusunda umut verici bir potansiyel sergilemiştir. Bununla birlikte, ortam sesi analizi bağlamındaki uygulamaları nispeten keşfedilmemiştir. Bu tez, bu kritik araştırma alanındaki mevcut yetenek sınırlarını genişletmek için transformatör mimarileri ve transfer öğrenme stratejileri gibi en yeni teknikleri dâhil ederek bu temel çalışmaları geliştirmeye çalışmaktadır.

Bu zorluğun ele alınması, çeşitli sektörlerde önemli toplumsal faydalar sunmaktadır. Kamu güvenliği alanında, geliştirilen teknolojiler şiddeti önlemek ve krizleri etkili bir şekilde yönetmek için proaktif önlemleri kolaylaştırabilir, potansiyel olarak can kaybını azaltabilir ve suç oranlarını düşürebilmektedir. Adli bilişim alanında delil niteliğindeki ses kayıtlarının ön analizinin yapılması bütünlüğünün incelenmesi ve doğruluğunun tespit edilmesi gibi durumlarda kritik önem arz etmektedir. Sağlık hizmetleri alanında, bu sistemler duygusal sıkıntıların erken tespit edilmesini sağlayabilir ve böylece ruh sağlığı uzmanlarının hastaları daha etkili bir şekilde teşhis ve tedavi

etmelerine yardımcı olabilmektedir. Eğitim kurumları, zorbalık veya öğrenci stresi vakalarını izlemek ve ele almak için bu teknolojilerden yararlanabilmektedir. Ayrıca, bu gelişmeler akıllı şehir girişimlerinin gerçekleştirilmesine önemli ölçüde katkıda bulunabilir, daha güvenli ve daha duyarlı kentsel ortamları teşvik edebilmektedir. Yenilikçi teknolojik çözümlerin geliştirilmesi yoluyla kritik bir toplumsal ihtiyacı etkin bir şekilde ele alan bu araştırma, akademik sorgulama ile gerçek dünya etkisi arasındaki boşluğu doldurmaktadır.



2. ARKA PLAN VE LİTERATÜR İNCELEMESİ

Bu bölümde ortam sesi analizine genel bakış sağlanarak, ortam seslerinin türleri, zorlukları (örn. üst üste binen sesler, gürültü sesler) ve uygulamalar ile ilgili detaylı bilgiler verilecektir.

2.1. Ortam Seslerine Genel Bakış

Ortam sesi analizi, konuşma ve müziğin ötesinde çevresel seslerin incelenmesine odaklanmaktadır. Bu sesler, belirli ortamların ve olayların akustik özelliklerini yansıtarak insan faaliyetleri ve gerçek dünya çerçevesi hakkında değerli bilgiler sağlamaktadır [8].

Ortam sesi analizinin uygulamaları çok çeşitlidir ve güvenlik izleme, sağlık hizmetleri ve şehir planlama gibi çeşitli alanları kapsamaktadır. Bununla birlikte, ortam seslerini analiz etmek, karmaşık ve genellikle öngörülemeyen doğaları nedeniyle benzersiz zorluklar sunmaktadır [9].

2.1.1. Ortam Seslerinin Türleri

Ortam sesleri çeşitlidir ve genel olarak aşağıdaki türlere ayrılmaktadır:

Çevresel Sesler

Bu sesler doğal veya yapılı çevreden kaynaklanır ve ortam hakkında bağlam sağlamaktadır [10]. Alt kategoriler şunları içermektedir:

- **Doğal sesler:** Meteorolojik olaylar (rüzgâr, yağmur, gök gürültüsü), hidrolojik süreçler (dalgalar) ve biyolojik faaliyetler (kuş sesleri, hışırdayan yapraklar) gibi doğal kaynaklardan kaynaklanan akustik fenomenler olarak tanımlanmaktadır.
- **Kentsel sesler:** Ulaşım gürültüsü (trafik, kornalar), inşaat faaliyetleri ve şehir yaşamıyla ilişkili diğer sesler dâhil olmak üzere kentsel ortamların karakteristik sesleri olarak tanımlanmaktadır.
- **Endüstriyel sesler:** Makinelerin çalışması, fabrika alarmları ve üretim birimlerine özgü sesler gibi endüstriyel süreçler tarafından üretilen sesler olarak tanımlanmaktadır.

İnsan Etkileşimli Sesler

Bu sesler doğrudan insan eylemleri ve etkileşimleriyle ilgilidir [11]. Bunlar ayrıca şu şekilde kategorize edilebilmektedir:

- **Sözel olmayan iletişim:** Ağlama, gülme, alkış ve bağırma gibi dil içermeyen seslendirmeler olarak adlandırılmaktadır.

- **Kalabalık sesleri:** Pazarlardaki konuşmalar, stadyumlardaki tezahüratlar ve kamusal alanlardaki mırıltılar dâhil olmak üzere insan grupları tarafından üretilen sesler olarak tanımlanmaktadır.
- **Sıkıntı veya şiddet gösteren sesler:** Kavga, çığlık ve panik kaynaklı sesler gibi olumsuz olaylarla ilişkili sesler olarak tanımlanmaktadır.

Hayvan Sesleri

Bu sesler, genellikle ormanlar veya hayvanat bahçeleri gibi belirli habitatlarla ilişkilendirilen çağrılar, hırlamalar ve diğer türe özgü sesler dâhil olmak üzere hayvan seslerini kapsamaktadır.

Anormal Sesler

Bu sesler, olağan ortam gürültü profilinden önemli ölçüde sapan, beklenmedik ve atipik yapılarıyla karakterize edilmektedir. Örnekler arasında cam kırılması, silah sesleri veya patlamalar yer alır ve genellikle ilgi çekici olaylara işaret etmektedir.

2.2. Ortam Seslerinin Analizindeki Zorluklar

Ortam seslerinin analizi, doğru ve sağlam ses olayı tanımayı engelleyen birkaç önemli zorluk sunmaktadır. Bu zorluklar, akustik ortamların doğasında var olan karmaşıklık ve değişkenlikten kaynaklanmaktadır. Bu bölümde bu zorluklar ayrıntılı olarak ele alınmıştır [12].

2.2.1. Çakışan Sesler (Overlapping)

Gerçek dünyadaki akustik sahneler sıklıkla birden fazla ses olayının aynı anda meydana gelmesini içermekte ve bu da seslerin üst üste binmesine neden olmaktadır. Bu üst üste binme, tek tek ses kaynaklarının doğru bir şekilde izole edilmesini ve tanımlanmasını zorlaştırmaktadır. Örneğin, kalabalık bir kentsel ortamda, bir bağırış sesi araç gürültüsü, inşaat faaliyetleri ve yaya sohbetlerinin kakofonisi tarafından maskelenebilmektedir [13]. Bu sorunun ele alınması, kör kaynak ayırma veya hesaplamalı işitsel sahne analizi (CASA) gibi gelişmiş sinyal ayırma tekniklerinin, örtüşen ses kaynakları arasında ayırım yapmak için eğitilmiş makine öğrenimi modelleriyle birleştirilmesini gerektirmektedir. Bu modeller genellikle karışık ses sinyallerini ayırmak için spektral ve zamansal ipuçlarından yararlanmaktadır [14].

2.2.2. Gürültü ve Sinyal-Gürültü Oranı

Ortam ses kayıtları genellikle arka plan gürültüsü ile kirlenmekte, bu da sinyal-gürültü oranını (SNR) önemli ölçüde düşürebilmekte ve ilgilenilen sesleri gizleyebilmektedir. Bu durum

özellikle ince veya düşük genlikli sesleri tespit etmeye çalışırken sorun yaratmaktadır [15]. Örneğin, kalabalık bir parkta, kuş sesleri ve oyun oynayan çocuklardan gelen rakip seslerle birlikte zayıf bir ağlama sesini tespit etmek önemli bir zorluk teşkil etmektedir. Gürültünün etkilerini azaltmak için spektral çıkarma, Wiener filtreleme, dalgacık dönüşümü veya derin öğrenme tabanlı yöntemler gibi gürültü azaltma algoritmalarının uygulanması gerekmektedir [16]. Ayrıca, uyarlanabilir filtreleme gibi gelişmiş teknikler, değişen gürültü özelliklerine dinamik olarak uyum sağlayarak elde edilen verilerin netliğini ve kalitesini artırabilmektedir.

2.2.3. Akustik Çeşitlilik

Belirli bir ses olayının akustik özellikleri, çevredeki ortama bağlı olarak önemli ölçüde değişebilmektedir. Oda akustiği (yankılanma, ekolar), atmosferik koşullar ve engellerin varlığı gibi faktörler algılanan sesi önemli ölçüde değiştirebilmektedir. Örneğin bir çığlık, yankılanma ve ses emilimindeki farklılıklar nedeniyle iç mekânda dış mekâna kıyasla büyük ölçüde farklı duyulabilmektedir. Bu zorluğun üstesinden gelmek için modellerin akustik değişkenliğe karşı sağlam olması ve çeşitli akustik koşullarda iyi genelleme yapması gerekmektedir. Farklı akustik ortamları simüle etmek için mevcut ses verilerinin yapay olarak değiştirilmesini içeren veri artırma ve büyük veri kümeleri üzerinde önceden eğitilmiş modellerin belirli hedef ortamlar için ince ayarının yapıldığı transfer öğrenme gibi teknikler, akustik olay algılama sistemlerinin sağlamlığını artırabilmektedir [17].

2.2.4. Zamansal Dinamikler

Ses olayları, süre, ritim ve zamansal kalıplar bakımından farklılık gösteren çok çeşitli zamansal özellikler sergilemektedir. Bir kahkaha patlaması gibi bazı sesler birkaç saniyeye yayılabilirken, silah sesi gibi diğerleri anlıktır. Zamansal dinamiklerdeki bu değişkenlik, özellik çıkarma ve model tasarımı için bir zorluk teşkil etmektedir. RNN'ler, özellikle de Uzun Kısa Süreli Bellek (LSTM) ağları ve Geçitli Tekrarlayan Birimler (GRU'lar), ses sinyallerindeki zamansal bağımlılıkları modellemek için çok uygun bir ortam oluşturmaktadır. Modelin ses girdisinin ilgili kısımlarına odaklanmasını sağlayan dikkat mekanizmaları, göze çarpan zamansal özellikleri yakalama yeteneğini daha da geliştirebilmektedir [3, 18, 19].

2.2.5. Veri Kümesi Sınırlamaları

Ortam sesi analizi için makine öğrenimi modellerinin performansı büyük ölçüde geniş, çeşitli ve temsili veri kümelerinin mevcudiyetine bağlıdır. Ne yazık ki, mevcut birçok ortam sesi veri kümesi boyut, çeşitlilik ve gerçek dünya temsili açısından sınırlamalardan muzdariptir. Bu durum, modelleri gerçek dünya senaryolarında kullanırken aşırı uyum ve zayıf genelleme

performansına yol açabilmektedir. Bu sorunu ele almak, daha kapsamlı ve temsili veri kümeleri oluşturmak için ortak çabalar gerektirmektedir. Kitle kaynaklı veri toplama, veri artırma tekniklerinden yararlanma ve açık kaynaklı, kamuya açık veri setleri oluşturma gibi stratejiler, alanı ilerletmek için oldukça önem arz etmektedir.

2.2.6. Kültürel ve Bağlamsal Değişimler

Belirli seslerin yorumlanması kültürel geçmiş ve bağlamsal faktörlerden etkilenebilmektedir. Örneğin, kahkaha genellikle neşe ile ilişkilendirilirken, kültürel bağlama bağlı olarak alay veya gerginlik de ifade edebilmektedir. Benzer şekilde, bir çığlının anlamı da duruma göre büyük ölçüde değişebilmektedir. Bağlamsal ve kültürel bilginin model eğitime dâhil edilmesi, daha incelikli ve doğru ses olayı tanıma sistemleri geliştirmek için gereklidir. Bu, video veya metin açıklamaları gibi çok modlu bilgilerden yararlanmayı veya bağlamsal ipuçları sağlamak için bilgi tabanlarını ve ontolojileri dâhil etmeyi içerebilmektedir [20].

2.3. Ortam Sesi Analizinin Uygulamaları

Ortam sesi analizi, çevresel seslere gömülü bilgileri kullanarak çeşitli alanlarda zengin bir uygulama yelpazesi sunmaktadır. Bu bölümde daha akademik ve derinlemesine bir bakış açısı sağlanarak bazı belirgin örnekler ayrıntılı olarak açıklanmaktadır [21].

2.3.1. Kamu Güvenliği ve Emniyet

- **Şiddet tespiti:** Ortam seslerinin gerçek zamanlı analizi, silah sesleri, patlamalar veya fiziksel kavgalar gibi şiddet olaylarını tespit etmek için çok önemli olabilmektedir. Otomatik sistemler, bu olayların akustik imzalarını tespit etmek için eğitilebilir ve hızlı müdahale için kolluk kuvvetlerine veya acil servislere uyarıları tetiklemektedir. Bu uygulama, kamu güvenliğini önemli ölçüde artırma ve kritik durumlarda müdahale sürelerini azaltma potansiyeline sahip olmaktadır [22, 23].
- **Gözetim sistemleri:** Ortam sesi analizinin gözetim sistemlerine entegre edilmesi, görsel izlemenin ötesinde yeteneklerini geliştirmektedir. Bu sistemler, kamusal alanlardaki sesleri analiz ederek, potansiyel bir hırsızlığa işaret eden cam kırılma sesi veya gelişmekte olan bir rahatsızlığa işaret eden olağandışı kalabalık gürültüsünün tespiti gibi potansiyel tehditlere karşı erken uyarılar sağlayabilmektedir. Bu işitsel bilgi katmanı görsel verileri tamamlayarak çevrenin daha kapsamlı bir şekilde anlaşılmasını sağlayabilir ve durumsal farkındalığı artırabilmektedir [24].

2.3.2. Sağlık Hizmetleri ve Refah

- **Duygu tanıma:** Ortam sesi analizi, ağlama, kahkaha veya iç çekme gibi seslendirmelere dayalı olarak insanların duygusal durumlarını tespit etmek ve sınıflandırmak için kullanılabilir. Bu bilgi, özellikle sözlü olarak iletişim kurmakta zorluk çeken hastaların duygusal refahını izlemek için sağlık hizmeti ortamlarında değerli olabilmektedir. Ayrıca, ruh sağlığı takibinde sıkıntı veya endişe belirtilerini tespit etmek için kullanılabilir [25].
- **Yaşlı bakımı:** Destekli yaşam tesislerinde veya evde bakım ortamlarında, ortam sesi analizi yaşlı bireylerin güvenliğini ve refahını sağlamada çok önemli bir rol oynayabilmektedir. Sistemler imdat çağrılarını, düşmeleri (çarpmanın karakteristik sesiyle algılanır) veya tıbbi bir acil duruma işaret edebilecek diğer olağandışı sesleri tespit etmek üzere tasarlanabilmektedir. Bu teknoloji zamanında müdahaleyi mümkün kılabilir ve yaşlı nüfus için bakım kalitesini artırabilmektedir [26, 27].

2.3.3. Kentsel Planlama ve Gürültü Yönetimi

- **Gürültü kirliliğinin izlenmesi:** Kentsel gürültü seviyelerinin analiz edilmesi, trafik, inşaat ve endüstriyel faaliyetlerin insan sağlığı ve yaşam kalitesi üzerindeki etkisini değerlendirmek için gerekli olmaktadır. Ortam sesi analizi, kentsel alanların ayrıntılı gürültü haritalarını oluşturmak, gürültü kirliliğinin sıcak noktalarını belirlemek ve kentsel planlama kararlarını bilgilendirmek için kullanılabilir. Bu veriler, gürültü bariyerlerinin inşası veya trafik akışının düzenlenmesi gibi gürültü azaltma stratejilerinin uygulanmasına rehberlik edebilmektedir [28].
- **Ses manzarası tasarımı:** Sadece gürültüyü azaltmanın ötesinde, ortam sesi analizi kentsel ses manzaralarını aktif olarak tasarlamak için kullanılabilir. Şehir planlamacıları, farklı seslerin bir alanın genel işitsel deneyimine nasıl katkıda bulunduğunu anlayarak, yalnızca daha sessiz değil, aynı zamanda daha hoş ve refah için elverişli ortamlar yaratabilmektedir. Bu durum, su sesi gibi doğal seslerin dâhil edilmesini veya gürültüyü en aza indirmek ve daha uyumlu bir akustik ortam yaratmak için trafik akışının stratejik olarak yönetilmesini içerebilmektedir [29].

2.3.4. Akıllı Ev ve Nesnelerin İnterneti (IoT)

- **Sesle etkinleştirilen sistemlerin geliştirilmesi:** Ortam sesi analizinin akıllı ev sistemlerine entegre edilmesi, sesle etkinleştirilen asistanların yeteneklerini artırabilmektedir. Bu sistemler sözlü komutlara yanıt vermenin ötesinde, bebek ağlaması veya su ısıtıcısının kaynama sesi gibi sözel olmayan sesleri yorumlayacak ve uygun eylemleri tetikleyecek

şekilde tasarlanabilmektedir. Bu durum, daha sezgisel ve duyarlı bir akıllı ev ortamı yaratılmasına katkıda bulunmaktadır [30].

- **Güvenlik sistemleri:** Ortam sesi analizi akıllı ev sistemlerinin güvenlik özelliklerini geliştirebilmektedir. Bu sistemler cam kırılması, ayak sesleri veya alarmlar gibi anormal sesleri tespit ederek potansiyel izinsiz girişler konusunda erken uyarılar sağlayabilir ve ev sahiplerini veya güvenlik hizmetlerini uyarabilmektedir. Bu durum, ev ortamına ek bir güvenlik katmanı ekleyerek güvenliği ve iç huzurunu artırabilmektedir [30, 31].

2.3.5. Yaban Hayatı İzleme ve Koruma

- **Hayvan davranış analizi:** Hayvanların doğal yaşam alanlarındaki davranışlarını incelemek kullanılmaktadır. Araştırmacılar, hayvan seslerini analiz ederek iletişim kalıpları, sosyal etkileşimler ve çevresel değişikliklerin hayvan popülasyonları üzerindeki etkisi hakkında bilgi edinebilmektedir. Bu bilgiler, koruma çabaları ve ekosistem dinamiklerini anlamak için çok önemlidir [32].
- **Ekosistem izleme:** Ekosistemlerin sağlığını ve biyolojik çeşitliliğini izlemek için kullanılabilmektedir. Araştırmacılar, belirli bir ortamın kolektif ses manzarasını analiz ederek türlerin varlığındaki değişiklikleri belirleyebilmekte, çevresel stres belirtilerini tespit edebilmekte ve koruma müdahalelerinin etkisini izleyebilmektedir. Bu yaklaşım, ekosistemleri zaman içinde izlemek için invaziv olmayan ve uygun maliyetli bir yol sağlamaktadır [33].

2.3.6. Eğlence ve Oyun

- **Sürükleyici deneyimler:** Sanal gerçeklik (VR) ve artırılmış gerçeklik (AR) uygulamalarında daha sürükleyici ve gerçekçi deneyimler yaratmak için kullanılabilmektedir. Bu uygulamalar, kullanıcının gerçek dünya ortamını analiz ederek, kullanıcının eylemlerine ve çevresine yanıt veren dinamik ses manzaraları oluşturabilir, varlık ve gerçekçilik hissini artırabilmektedir [34].
- **İçerik moderasyonu:** Çevrimiçi platformlarda ve oyun ortamlarında, ortam sesi analizi, şiddet içeren veya küfürlü dil gibi uygunsuz veya zararlı içeriği otomatik olarak tespit etmek için kullanılabilmektedir. Bu durum, kullanıcılar için daha güvenli ve daha olumlu çevrimiçi deneyimler yaratmaya yardımcı olmaktadır [35].

2.3.7. Afet Yönetimi

- **Erken uyarı sistemleri:** Ortam sesi analizi, doğal afetler için erken uyarı sistemlerinin geliştirilmesine katkıda bulunabilmektedir. Örneğin, infrasound izleme, volkanik

patlamalar, nükleer testlerin tespiti veya tsunamilerle ilişkili düşük frekanslı sesleri tespit edebilmekte, tahliyeler ve afet hazırlığı için değerli bir ön hazırlık zamanı sağlayabilmektedir [36-38].

- **Arama ve kurtarma operasyonları:** Afetlerin ardından, arama ve kurtarma operasyonlarına yardımcı olmak için kullanılabilir. Kurtarma ekipleri, afetten etkilenen bölgelerdeki sesleri analiz ederek, yardım çağlıkları veya yaşam belirtileri gibi insan seslerini tespit ederek potansiyel olarak kurbanların yerini belirleyebilmektedir. Bu teknoloji acil durumlarda hayat kurtarmada çok önemli olmaktadır [39-41].

2.3.8. Değerlendirme

Ortam sesi analizi, teknoloji ve toplumsal ihtiyaçlar arasında köprü kuran zorlu ancak dönüştürücü bir alandır. Araştırmacılar, üst üste binen sesleri, gürültüyü ve akustik değişkenliği ele alarak, ortam seslerini etkili bir şekilde yorumlamak için makine öğrenimi modellerinin tüm potansiyelini ortaya çıkarabilmektedir. Bu alanın uygulamaları kamu güvenliği, sağlık hizmetleri, çevre koruma ve ötesini kapsamakta, daha akıllı ve daha güvenli bir dünyayı şekillendirmedeki kritik rolünü ortaya koymaktadır.

2.4. Seslerden Duygu Tanıma

Sesten duygu tanıma, akustik verilerden türetilen duygusal durumların tespiti ve sınıflandırılmasına odaklanan ilgi çekici ancak karmaşık bir araştırma alanı sunmaktadır. Bu alan genel olarak konuşma tabanlı duygu tanıma ve konuşma tabanlı olmayan duygu tanıma şeklinde iki ana araştırma akışına ayrılabilir. Tezin bu bölümü, araştırma alanlarımızın her birindeki ilerlemeyi, zorlukları ve gelişmeleri detaylandırmaktadır [42].

2.4.1. Konuşma Tabanlı Duygu Tanıma

Konuşma tabanlı duygu tanıma (SER), konuşmadaki akustik ipuçları aracılığıyla aktarılan zengin duygusal bilgiler nedeniyle önemli bir araştırma alanı olarak ortaya çıkmıştır. Prozodik, spektral ve ses kalitesi özelliklerini kapsayan bu ipuçları, insanların duygusal durumları hakkında değerli bilgiler sağlamaktadır. Bu alan, geleneksel makine öğrenimi yaklaşımlarından sofistike derin öğrenme mimarilerine doğru evrilerek önemli bir ilerlemeye tanık olmaktadır [43].

Duygu Tanıma için Temel Akustik Özellikler

Prozodik özellikler: Konuşmanın ritmik ve melodik yönleriyle ilgilidir ve duyguların aktarılmasında çok önemli bir rol oynamaktadır [44]. Temel prozodik özellikler şunları içermektedir:

- Perde, tonlama ve melodiyi yansıtan temel frekanstaki değişimler, duygusal durumlarla güçlü bir şekilde ilişkili olmaktadır. Örneğin, daha yüksek perde genellikle heyecan veya öfke ile ilişkilendirilirken, daha düşük perde üzüntü veya sakinliği gösterebilmektedir.
- Ses yüksekliği, konuşmacının enerji seviyesini yansıtan konuşma sinyalinin yoğunluğu da duygusal uyarılmayı gösterebilmektedir. Daha yüksek sesle konuşma genellikle öfke veya heyecanla ilişkilendirilirken, daha yumuşak konuşma üzüntü veya korkuya işaret edebilmektedir.
- Konuşma hızı, birim zamanda hece veya kelime olarak ölçülen konuşma hızı, duygusal duruma göre değişebilmektedir. Daha hızlı konuşma genellikle endişeli veya heyecanlı durumlarda gözlenirken, daha yavaş konuşma üzüntü veya yorgunluğun göstergesi olabilmektedir.
- Ritim, duraklamalar ve hece süresindeki değişimler de dâhil olmak üzere konuşmanın ritmik yapısı duygusal ifadeye katkıda bulunmaktadır.

Spektral Özellikler: Konuşma sinyalinin frekans içeriğini yakalamakta ve ses yolu rezonansları hakkında bilgi sağlamaktadır [45]. Yaygın olarak kullanılan spektral özellikler şunları içermektedir:

- MFCC'ler, insan işitsel sisteminin ses algısına yaklaşma yetenekleri nedeniyle SER'de yaygın olarak kullanılmaktadır. Konuşma sinyalinin spektral zarfını yakalarlar ve gürültüye karşı dayanıklıdır.
- Formantlar, artikulatorlerin şekli ve konfigürasyonundan etkilenen ses yolunun rezonans frekanslarıdır. Formant frekansları, konuşmacının ses yolu özellikleri ve duygusal durumu hakkında bilgi sağlayabilmektedir.
- Spektral merkez, spektrumun ortalama frekansını temsil etmektedir. Ayrıca farklı duygusal durumları ayırt etmek için kullanılabilir.

Ses Kalitesi Özellikleri: Konuşmacının fizyolojik ve duygusal durumuyla ilgili olan sesteki ince varyasyonları yakalamaktadır [46]. Temel ses kalitesi özellikleri şunları içermektedir:

- Titreşim, temel frekanstaki döngüden döngüye değişimleri ölçmekte ve vokal kıvrım titreşiminin stabilitesini yansıtmaktadır. Artan titreşim, genellikle stres veya ses yorgunluğu ile ilişkilendirilmektedir.
- Titreme, konuşma sinyalinin genliğindeki döngüden döngüye değişimi ölçmekte ve aynı zamanda ses kıvrımı kararsızlığını yansıtmaktadır. Literatürde titreşim ve titreme terimleri yerine sıklıkla kullanılan jitter gibi artan shimmer (ses şiddetindeki kısa süreli değişimleri ölçer) de duygusal veya fiziksel stresin göstergesi olarak ilişkilendirilmektedir.
- Harmonik-Gürültü Oranı (HNR), konuşma sinyalindeki harmonik enerjinin gürültü enerjisine oranını ölçmektedir. Daha düşük bir HNR, septe belirli duygusal durumlar veya vokal patolojilerle ilişkilendirilebilecek artan gürültüye işaret etmektedir.

Konuşma Tabanlı Duygu Tanıma için Yöntem ve Modeller

- **Geleneksel makine öğrenimi yaklaşımları:** SER'deki ilk araştırmalar SVM'ler, k-En Yakın Komşular (kNN'ler) ve Karar Ağaçları gibi geleneksel makine öğrenimi algoritmalarına dayanmaktadır [6], [47-48]. Bu yöntemler genellikle konuşma sinyalinin ilgili akustik özelliklerin manuel olarak çıkarılmasını ve ardından bu özellikleri duygusal kategorilerle eşleştirmek için bir sınıflandırıcının eğitilmesini içermektedir. Bu yaklaşımlar bazı başarılar sağlamış olsa da genellikle konuşma verilerindeki karmaşık zamansal dinamikleri ve doğrusal olmayan ilişkileri yakalamakta yetersiz kalmışlardır.
- **Derin öğrenme yaklaşımları:** Derin öğrenmenin ortaya çıkışı, SER alanında devrim niteliğinde bir dönüşüm yaratmış ve doğruluk ile sağlamlık açısından önemli ilerlemeler sağlamıştır. Derin öğrenme modelleri, manuel özellik mühendisliği ihtiyacını ortadan kaldırarak ham konuşma verilerinden karmaşık temsilleri otomatik olarak öğrenebilmektedir. Çeşitli derin öğrenme mimarileri SER'e başarıyla uygulanmıştır. Örneğin CNN'ler, Konuşma spektrogramlarından yerel özelliklerin çıkarılması ve frekans alanındaki ilgili kalıpların tespit edilmesi açısından oldukça etkilidir. Özellikle konuşmadaki spektral bilgilerin modellenmesi sürecinde yaygın olarak kullanılmaktadırlar. RNN'ler, Özellikle LSTM ağları ve GRU'lar, konuşmadaki zamansal bağımlılıkları yakalama sürecinde oldukça uygundur. Konuşmanın sıralı doğasını modelleyerek duygu tanıma için çok önemli olan uzun menzilli bağlamsal bilgileri öğrenebilmektedir. Dönüştürücüler, başlangıçta doğal dil işleme (NLP) için geliştirilen dönüştürücü modeller, son zamanlarda SER alanında umut vaat eden sonuçlar elde etmiştir. Dönüştürücüler uzun vadeli bağımlılıkları etkili bir şekilde modelleyebilme ve konuşma sinyallerindeki küresel bağlamı yakalayarak performansı artırma yeteneğine sahiptir. Buna ek olarak dikkat mekanizmalarından yararlanarak modelin girdi dizisinin en alakalı kısımlarına odaklanmasını sağlamaktadırlar [49-53].

Konuşma Tabanlı Duygu Tanıma için Önemli Veri Kümeleri

Yüksek kaliteli, açıklanmalı veri kümeleri SER modellerinin eğitimi ve değerlendirilmesi için gereklidir [54-56]. Her biri duygusal kategoriler, kayıt koşulları ve konuşmacı demografisi açısından kendine has özelliklere sahip, halka açık çeşitli veri kümeleri geliştirilmiştir. Bazı önemli veri kümeleri şunlardır:

- **IEMOCAP (Interactive Emotional Dyadic Motion Capture):** Bu veri kümesi, duygusal etiketlerle açıklanmış hem senaryolu hem de doğaçlama ikili konuşmaları içermektedir. Zengin duygusal içeriği ve multimodal bilgileri (ses, video, hareket yakalama) nedeniyle SER topluluğunda yaygın olarak kullanılmaktadır [57-58].

- **RAVDESS (Ryerson Audio-Visual Database of Emotional Speech and Song):** Bu veri kümesi, profesyonel aktörler tarafından canlandırılan duygusal konuşma ve şarkı örneklerinden oluşmaktadır. Net duygusal ifadelerle sahip yüksek kaliteli kayıtlar sağlamaktadır [50, 59].
- **CREMA-D (Crowd-Sourced Emotional Multimodal Actors Dataset):** Bu veri kümesi, kitle kaynak yöntemiyle toplanmış ve profesyonel aktörler tarafından seslendirilmiş geniş bir duygusal ifade yelpazesini içermektedir. Veri kümesi, çeşitli duygusal durumları kapsamasının yanı sıra nispeten büyük bir ölçeğe sahiptir [60-61].

Konuşma Tabanlı Duygu Tanıma Alanındaki Zorluklar

SER'de kaydedilen önemli ilerlemeye rağmen, çeşitli zorluklar devam etmektedir.

- **Konuşmacı değişkenliği:** Duygusal ifade, kişilik, kültürel geçmiş ve konuşma tarzındaki farklılıklar nedeniyle bireyler arasında önemli ölçüde değişmektedir. SER modellerinin görünmeyen konuşmacılara iyi genelleme yapabilmesi için bu konuşmacı değişkenliğine karşı sağlam olması gerekmektedir.
- **Kültürel farklılıklar:** Duyguların ifade edilme ve algılanma şekli kültürler arasında farklılık gösterebilmektedir. Bir kültürden gelen veriler üzerinde eğitilen SER modelleri, başka bir kültürden gelen veriler üzerinde iyi performans göstermeyebilmektedir. Bunun ele alınması, kültürel açıdan hassas modellerin geliştirilmesini ve kültürler arası veri kümelerinin bulunmasını gerektirmektedir.
- **Bağlamsal anlama:** Konuşmadaki duygular genellikle ifade edildikleri bağlamdan etkilenmektedir. Örneğin, aynı seslendirme, duruma ve konuşmacılar arasındaki ilişkiye bağlı olarak farklı duygular iletebilmektedir. Çevresel metin veya görsel ipuçları gibi bağlamsal bilgilerin entegrasyonu, SER sistemlerinin doğruluk oranını artırmak açısından kritik bir rol oynamaktadır. Ses verisini metin veya video gibi diğer modalitelerle birleştiren çok modlu analiz yaklaşımı, bu zorluğun aşılmasına yönelik umut vaat eden bir strateji olarak öne çıkmaktadır.

2.4.2. Konuşma Tabanlı Olmayan Duygu Tanıma

Konuşma dilinin ötesinde çok çeşitli seslendirmeleri ve çevresel sesleri kapsayan konuşma dışı sesler, özellikle konuşmanın olmadığı veya kullanılmadığı bağlamlarda duygu tanıma için tamamlayıcı bir yol sunmaktadır. Ağlama, kahkaha, çığlık, iç çekme ve diğer ses patlamalarını içeren bu sesler, duygusal durumlara ilişkin değerli bilgiler sağlayabilmektedir. Bu alandaki araştırmalar, ortam sesi analizinin öneminin ve çeşitli alanlardaki uygulama potansiyelinin giderek daha fazla fark edilmesiyle artan bir ilgi ortamı oluşturmaktadır [62-63].

Konuşma Tabanlı Olmayan Seslerde Temel Akustik Özellikler

- **Zamansal özellikler:** Konuşma dışı seslerin zamansal dinamiklerini yakalayarak duygusal ifadelerin zamanlaması ve süresi hakkında bilgi sağlamaktadır. Temel zamansal özellikler sıklıkla kullanılmaktadır. Örneğin Süre, bir ses olayının uzunluğuyla alakalı bir durumdur. Örneğin bir çığlığın veya kahkahanın süresi, duygunun yoğunluğunun veya doğasının göstergesi olabilmektedir. Örneğin, uzun süreli bir çığlık, kısa bir çığlığa kıyasla daha büyük bir sıkıntıya işaret edebilmektedir. Başlangıç ve Bitiş, bir ses olayının başlangıç ve bitiş zamanlaması da bilgilendirici olabilmektedir. Örneğin başlangıcın aniliği, duygusal bir tepkinin aniliği ile ilgili olabilmektedir.
- **Spektral özellikler:** Bu özellikler, farklı duygularla ilişkili spektral özellikleri yakalayarak konuşma dışı seslerin frekans içeriğini analiz etmektedir. Yaygın olarak kullanılan spektral özelliklerden biri MFCC'dir, konuşma analizinde olduğu gibi MFCC'ler sesin spektral zarfını temsil etme yetenekleri ve gürültüye karşı sağlamlıkları nedeniyle konuşma tabanlı olmayan ses tanımda yaygın olarak kullanılmaktadır. Spektral kontrast ise ses spektrumundaki tepe ve vadiler arasındaki farkı ölçerek sesin netliği ve ayırt ediciliği hakkında bilgi sağlamaktadır. Farklı konuşma tabanlı olmayan ses türlerini ayırt etmek için faydalı olabilmektedir.

Konuşma Tabanlı Olmayan Duygu Tanıma için Yöntemler ve Modeller

- **Geleneksel makine öğrenimi yaklaşımları:** Konuşma dışı ses tanıma yönelik ilk yaklaşımlar SVM'ler, Rastgele Ormanlar ve Gauss Karışım Modelleri (GMM'ler) gibi geleneksel makine öğrenimi algoritmalarını kullanmışlardır. Bu yöntemler genellikle ilgili akustik özelliklerin manuel olarak çıkarılmasını ve ardından bu özellikleri duygusal kategorilerle eşleştirmek için bir sınıflandırıcının eğitilmesini içermektedir. Bu yaklaşımlarla bazı başarılar elde edilmiş olsa da genellikle konuşma tabanlı olmayan seslerin karmaşıklığını ve değişkenliğini yakalamakta yetersiz kalmıştır [5, 64].
- **Derin öğrenme yaklaşımları:** Konuşma tabanlı duygu tanıma benzer şekilde, derin öğrenme de konuşma dışı ses tanıma alanında devrim yaratmıştır. CNN'ler ve RNN'ler dâhil olmak üzere derin sinir ağları (DNN'ler), özellikleri otomatik olarak çıkarmada ve konuşma dışı ses verilerinden duyguları sınıflandırmada üstün performans göstermektedir. Otomatik kodlayıcılar, özellik çıkarma ve boyut azaltma gibi işlemlerde kullanılmak üzere duygusal seslerin sıkıştırılmış temsillerini öğrenebilmektedir. Bunun yanı sıra, eğitim veri kümelerinin çeşitliliğini artırmak ve modellerin sağlamlığını geliştirmek amacıyla sentetik duygusal sesler üreten Üretken Çekişmeli Ağlar (GAN'lar) da detaylı olarak incelenmiştir [65].

Konuşma Tabanlı Olmayan Ses Tanıma Alanındaki Önemli Çalışmalar ve Uygulamalar

Konuşma dışı ses tanıma alanındaki araştırmalar, aşağıdakiler de dâhil olmak üzere çeşitli uygulamalara odaklanmıştır.

- **Ağlama ve kahkaha algılama:** Ağlamanın otomatik tespiti, bebek izleme ve yaşlı bakımı gibi uygulamalarda değerli olabilirken, kahkaha tespiti sosyal etkileşimleri analiz etmek veya medyanın duygusal içeriğini değerlendirmek için kullanılabilir [66].
- **Çığlık algılama:** Çığlıkları otomatik olarak tespit etme yeteneği, kamu gözetimi ve acil müdahale sistemleri gibi güvenlik açısından kritik uygulamalarda çok önemli bir yer tutmaktadır. Çığlıkların erken tespiti, potansiyel tehditlerin veya sıkıntılı durumların belirlenmesine yardımcı olabilmesinin yanı sıra zamanında müdahaleyi mümkün kılmaktadır [67].

Konuşma Tabanlı Olmayan Sesler için Veri Kümeleri

Büyük, açıklamalı veri kümelerinin mevcudiyeti, konuşma dışı ses tanıma modellerinin eğitimi ve değerlendirilmesi için oldukça önemlidir. Bazı önemli veri kümeleri şunlardır:

- **ESC-50:** Bu veri kümesi, insan ve hayvan sesleri de dâhil olmak üzere çok çeşitli çevresel sesler içermektedir. Ayrıca çeşitli ses tanıma görevleri için modellerin eğitilmesi ve değerlendirilmesi için değerli bir kaynak sağlamaktadır [68].
- **UrbanSound8K:** Öncelikle kentsel seslere odaklanmış olsa da bu veri kümesi duygu tanıma araştırmaları için kullanılabilecek ağlama ve çığlık gibi bazı konuşma tabanlı olmayan insan seslerini içermektedir [69].
- **AudioSet:** Bu büyük ölçekli veri kümesi, duygular, müzik ve çevresel sesler gibi çok çeşitli sesleri içeren milyonlarca etiketli ses segmenti içermektedir. Çeşitli ses analizi görevleri için modelleri eğitmek ve değerlendirmek için zengin bir kaynak sağlamaktadır [70].

Konuşma Tabanlı Olmayan Duygu Tanıma Alanındaki Zorluklar

Konuşma dışı ses tanıma kaydedilen ilerlemeye rağmen, bazı zorluklar devam etmektedir.

- **Çakışan sesler:** Konuşma dışı sesler genellikle diğer çevresel seslerle birlikte ortaya çıkmaktadır. Bu da duygusal ipuçlarını izole etmeyi ve analiz etmeyi zorlaştırmaktadır. Örtüşen sesleri ayırmak ve analiz etmek için sağlam yöntemler geliştirmek, konuşma dışı duygu tanıma sistemlerinin doğruluğunu artırmak için oldukça önem arz etmektedir.
- **İnce duygusal ifadeler:** Memnuniyet veya hafif kızgınlık gibi bazı duygular, konuşma dışı seslerde ince bir şekilde ifade edilebilmektedir. Bu da doğru bir şekilde tanınmalarını zorlaştırmaktadır. Bu ince ipuçlarına duyarlı modeller geliştirmek, önemli bir teknik ve metodolojik zorluk olarak karşımıza çıkmaktadır.

- **Veri kıtlığı:** Konuşma veri kümeleriyle karşılaştırıldığında, konuşma dışı duygu tanıma için yüksek kaliteli, açıklamalı veri kümeleri nispeten azdır. Bu veri kıtlığı, modellerin performansını ve genelleştirilebilirliğini sınırlayabilmektedir. Daha büyük ve daha çeşitli veri kümeleri oluşturma çabaları, alanı ilerletmek için oldukça önem arz etmektedir.

2.4.3. Kombine Konuşma ve Konuşma Dışı Duygu Tanıma

Duyguların tanınması, duygusal ifadenin nadiren tek bir kanalla sınırlı olduğunu kabul ederek, çoklu modaliteleri entegre eden bütünsel bir yaklaşımdan önemli ölçüde fayda sağlamaktadır. Sonuç olarak, araştırma çabaları giderek daha kapsamlı ve doğru duygu tanıma sistemleri için konuşma ve konuşma dışı ses ipuçlarını birleştirmeye ve hatta diğer modaliteleri dâhil etmeye odaklanmaktadır [71].

Çok Modelli Duygu Tanıma

İnsan duygusal ifadesi, sözlü ve sözsüz ipuçlarının bir araya gelmesiyle oluşan karmaşık bir olgudur. Konuşma, duygusal durum hakkında önemli bilgiler sunarken, genellikle yüz ifadeleri, vücut dili ve fizyolojik sinyaller gibi diğer modalitelerle tamamlanmakta ve hatta bazen bu modalitelerle çelişebilmektedir. Çok modlu duygu tanıma, duygu tanıma sistemlerinin doğruluğunu ve sağlamlığını artırmak için bu farklı modaliteler tarafından sağlanan tamamlayıcı bilgilerden yararlanmayı amaçlamaktadır [72].

- **Konuşma ve konuşma dışı ses entegrasyonu:** Konuşmanın ötesinde, kahkaha, iç çekme, nefes alma ve ses patlamaları gibi konuşma dışı ses ipuçları da zengin duygusal bilgiler aktarmaktadır. Bu konuşma dışı sesleri konuşma analizi ile entegre etmek, duygusal durumun daha eksiksiz bir resmini sağlayabilmektedir. Örneğin, bir iç çekiş rahatlama veya boyun eğme anlamına gelebilirken, bir soluk alma şaşkınlık veya korku anlamına gelebilmektedir. Bu ipuçlarının konuşmadan elde edilen bilgilerle birleştirilmesi, daha doğru ve incelikli duygu tanımaya yol açabilmektedir.
- **Görsel ve fizyolojik modalitelerle füzyon:** İşitsel ipuçlarına ek olarak, yüz ifadeleri ve beden dili gibi görsel bilgiler de duygusal iletişimde hayati bir rol oynamaktadır. Özellikle yüz ifadeleri genellikle duygusal durumlarla yüksek oranda ilişkilendirilmekte ve konuşmaya değerli tamamlayıcı bilgiler sağlamaktadır. Ayrıca, kalp atış hızı değişkenliği, deri iletkenliği ve beyin aktivitesi gibi fizyolojik sinyaller, duygularla ilişkili altta yatan fizyolojik değişikliklere dair içgörü sağlayabilmektedir. Bu modalitelerin erken füzyon (özelliklerin erken aşamada birleştirilmesi) veya geç füzyon (ayrı modalite sınıflandırıcılarından alınan kararların birleştirilmesi) gibi tekniklerle entegre edilmesi, duygu tanıma doğruluğunu önemli ölçüde artırabilmektedir. Örneğin, bir sistem daha sağlam ve doğru bir duygu sınıflandırması elde etmek için konuşma ve konuşma dışı

seslerden elde edilen akustik özellikleri, yüz ifadelerinden elde edilen görsel özellikler ve kalp atış hızından elde edilen fizyolojik özelliklerle birleştirebilmektedir.

Geliştirilmiş Duygu Yorumlaması için Ses Bağlamı Analizi

İster konuşma ister konuşma dışı seslerde olsun, duygusal ifadelerin yorumlanması genellikle içinde bulundukları bağlama bağlıdır. Aynı seslendirme veya yüz ifadesi, durumun özelliklerine, ilgili bireyler arasındaki ilişki dinamiklerine ve çevresel faktörlere bağlı olarak farklı duygular iletebilmektedir. Dolayısıyla, bağlamsal bilgilerin sisteme dâhil edilmesi, doğru duygu tanıma için kritik bir rol oynamaktadır.

- **Dilsel ve konuşma bağlamı:** Konuşma söz konusu olduğunda, dilsel içeriği ve konuşma bağlamını anlamak, duyguları doğru bir şekilde yorumlamak için çok önemli olabilmektedir. Örneğin, alaycı bir ses tonu, dilsel bağlam dikkate alınmazsa gerçek bir öfke olarak yanlış yorumlanabilmektedir. Çevredeki kelimeleri ve cümleleri analiz etmek, konuşmacının gerçek duygusal durumu hakkında değerli ipuçları sağlayabilmektedir. Ayrıca konuşma konusu ve konuşmacılar arasındaki etkileşim geçmişi gibi konuşma bağlamı da duyguların yorumlanmasını etkileyebilmektedir.
- **Akustik ve çevresel bağlam:** Bir sesin oluştuğu akustik ortam da bağlamsal bilgi sağlayabilmektedir. Örneğin, sessiz bir parkta duyulan bir çığlık, kalabalık bir caddede duyulan bir çığlıktan farklı yorumlanabilmektedir. Arka plandaki seslerin ve genel akustik sahnenin analiz edilmesi duygusal ifadelerin ayırt edilmesine yardımcı olabilmektedir. Benzer şekilde, günün konumu ve saati gibi çevresel bağlam da duyguların yorumlanmasını etkileyebilmektedir. Örneğin, bir silah sesi bir yerleşim bölgesinde atış poligonuna göre daha endişe verici olabilmektedir. Bu bağlamsal bilgilerin, muhtemelen çevresel ses sınıflandırması (ESC) veya konum tabanlı hizmetler aracılığıyla entegre edilmesi, duygu tanıma sistemlerinin doğruluğunu artırabilmektedir.

2.4.4. Seslerden Duygu Tanıma Uygulamaları

Hem konuşma hem de konuşma dışı seslendirmeleri kapsayan seslerden duyguları otomatik olarak tanıma yeteneği, çeşitli alanlardaki çeşitli uygulamalar için önemli bir potansiyel oluşturmaktadır. Bu uygulamalar, insan-bilgisayar etkileşimini geliştirmek, refahı artırmak ve çeşitli hizmetleri optimize etmek için akustik ipuçları aracılığıyla aktarılan zengin duygusal bilgilerden yararlanmaktadır [73].

Ruh Sağlığı İzleme ve Terapi

Sesten duygu tanıma, ruh sağlığı izleme ve terapötik müdahalelerde çok önemli bir rol oynamaktadır. Konuşma ve konuşma dışı seslendirmelerin otomatik analizi, bireyin duygusal

durumu hakkında değerli bilgiler sağlayarak duygusal sıkıntıların veya ruh halindeki değişikliklerin erken tespit edilmesini sağlayabilmektedir. Bu bilgiler aşağıdakiler için kullanılmaktadır:

- **Duygusal durumun objektif değerlendirmesi:** Geleneksel ruh sağlığı değerlendirmeleri genellikle önyargıya veya yanlılıklara eğilimli olabilen öznel öz-raporlamaya dayanmaktadır. Sesten otomatik duygu tanıma, duygusal durumun daha objektif ve sürekli bir şekilde ölçülmesini sağlayarak klinisyenlere tanı koyma ve tedavi planlaması süreçlerinde önemli veriler sunmaktadır.
- **Kişiselleştirilmiş terapötik müdahaleler:** Bireyin duygusal durumunu sürekli olarak izleyerek, en uygun zamanlarda kişiselleştirilmiş terapötik müdahaleler sunulabilmektedir. Örneğin, sistem anksiyete veya sıkıntı belirtileri tespit ederse, rahatlama egzersizlerini veya farkındalık tekniklerini tetikleyebilmektedir.
- **Uzaktan izleme ve tele sağlık:** Sesten duygu tanıma, tele-sağlık platformlarına entegre edilerek hastaların duygusal refahının uzaktan izlenmesini sağlayabilmektedir. Bu durum, özellikle yüz yüze ruh sağlığı hizmetlerine erişimi kısıtlı olan bireyler açısından oldukça değerli bir avantaj sunmaktadır.

Kamu Güvenliği ve Emniyeti

Kamuya açık alanlarda çılgınlık veya yardım çılgınlıkları gibi tehlike seslerinin tespit edilebilmesi, kamu güvenliği ve emniyeti açısından önemli sonuçlar doğurmaktadır. Otomatik sistemler şu amaçlarla kullanılabilir:

- **Gerçek zamanlı tehdit tespiti:** Bu sistemler parklar, alışveriş merkezleri veya ulaşım merkezleri gibi kamusal alanlardaki ortam seslerini analiz ederek potansiyel tehditleri veya acil durumları belirleyebilmektedir. Çılgınlıklar veya diğer tehlike seslerinin tespiti, kolluk kuvvetlerine ve acil servislere otomatik uyarı sistemleri aracılığıyla hızlı müdahale imkânı sağlamakta ve potansiyel olarak zararlı olayların genişlemesini önleyebilmektedir.
- **Geliştirilmiş gözetim sistemleri:** Sesten duygu tanımayı gözetim sistemlerine entegre etmek, bu sistemlerin yeteneklerini görsel izlemenin ötesinde geliştirebilmektedir. İşitsel bilgi katmanı, önemli bir bağlam sunabilmekte ve durumsal farkındalığı artırarak güvenlik personelinin potansiyel tehditleri daha etkili bir şekilde belirlemesine ve bu tehditlere müdahale etmesine imkân sağlayabilmektedir.

Müşteri Hizmetleri ve Deneyimi

Sesten duygu tanıma, çağrı merkezi temsilcileri veya sohbet robotları ile etkileşimler sırasında müşteri duygularını analiz etmek için müşteri hizmetleri ortamlarında kullanılabilir. Bu bilgiler şu amaçlarla kullanılmaktadır:

- **Hizmet kalitesini iyileştirmek:** Çağrı merkezi temsilcileri, müşteri duygularını anlayarak iletişim tarzlarını ve yaklaşımlarını müşteri ihtiyaçlarını ve endişelerini daha iyi karşılayacak şekilde uyarlayabilmektedir. Sistem ayrıca temsilcilere gerçek zamanlı geri bildirim sağlayarak gergin durumları yatıştırmak veya empatik destek sağlamak için stratejiler önerebilmektedir.
- **Kişiselleştirilmiş müşteri deneyimleri:** Duygu tanıma, müşterinin duygusal durumuna göre yanıtları ve teklifleri uyarlayarak müşteri deneyimlerini kişiselleştirmek için kullanılabilir. Örneğin, hayal kırıklığı yaşayan bir müşteriye indirim imkânı sağlanabilmekte veya kişiye özel bir özür mesajı sunulabilmektedir.
- **Otomatik duygu analizi:** Müşteri duygularının analiz edilmesi, bir şirketin ürün veya hizmetlerine yönelik genel müşteri duyarlılığı hakkında önemli bilgiler sunmaktadır. Bu bilgiler, iyileştirme gerektiren alanların tespit edilmesinde kullanılmakta ve müşteri memnuniyetini artırmaya yönelik etkili stratejilerin oluşturulmasına imkân tanımaktadır.

Eğlence ve Oyun

Sesten duygu tanıma, daha dinamik ve kişiselleştirilmiş etkileşimler yaratarak eğlence ve oyun deneyimini geliştirebilmektedir.

- **Uyarlanabilir oyun dinamikleri:** Öncelikle oyunlar, oyuncunun duygusal durumuna bağlı olarak zorluk derecelerini veya hikâye akışını dinamik bir şekilde uyarlayabilecek biçimde tasarlanabilmektedir. Örneğin, oyuncu hayal kırıklığı yaşadığında, oyun ek ipuçları sunabilmekte veya zorluk seviyesini düşürebilmektedir. Öte yandan, oyuncu sıkılmış hissediyorsa, oyun yeni zorluklar veya ödüller ekleyerek deneyimi yeniden ilgi çekici hale getirebilmektedir.
- **Kişiselleştirilmiş medya içeriği:** Akış hizmetleri veya medya oynatıcıları, kullanıcının mevcut ruh haline uygun içerik önermek için duygu tanıma özelliğini kullanabilmektedir. Örneğin, kullanıcı üzgün hissediyorsa, sistem canlandırıcı müzik veya filmler önerebilmektedir.
- **Sürükleyici deneyimler:** VR veya AR uygulamalarında, duygu tanıma daha sürükleyici ve kişiselleştirilmiş deneyimler oluşturmak için kullanılabilir. Örneğin, sanal ortam kullanıcının duygusal durumuna yanıt vererek daha gerçekçi ve ilgi çekici bir deneyim yaratabilmektedir.

Akıllı Cihazlar ve Sanal Asistanlar

Sesten duygu tanımayı akıllı cihazlara ve sanal asistanlara entegre etmek, bu cihazların yeteneklerini önemli ölçüde artırabilmektedir. Ayrıca onları daha sezgisel ve duyarlı hale getirebilmektedir.

- **Bağlama duyarlı etkileşimler:** Akıllı cihazlar, kullanıcı duygularına daha incelikli ve bağlama duyarlı bir şekilde yanıt verebilecek şekilde tasarlanabilmektedir. Örneğin, bir akıllı hoparlör, kullanıcının stresli olduğunu algıladığında rahatlatıcı mesajlar iletebilir veya rahatlatıcı bir müzik listesi çalabilmektedir.
- **Proaktif yardım:** Sanal asistanlar, kullanıcının duygusal durumuna göre proaktif olarak yardım sunmak üzere eğitilebilmektedir. Örneğin, sistem kullanıcının bunalmış hissettiğini tespit ettiğinde, bir mola vermeyi önerebilmekte veya rehberli meditasyon egzersizleri sunabilmektedir.
- **Kişiselleştirilmiş öneriler:** Akıllı cihazlar, kullanıcının mevcut ruh haline göre müzik, film, haber makaleleri veya diğer içerik önerilerini kişiselleştirmek için duygu tanıma özelliğini kullanabilmektedir.

Değerlendirme

Seslerden duygu tanıma alanı, son yıllarda hem konuşma tabanlı hem de konuşma tabanlı olmayan yaklaşımları kapsayan önemli bir ilerlemeye tanık olmuştur. Konuşma tabanlı duygu tanıma (SER) göreceli bir olgunluk seviyesine ulaşmış olsa da ortam akustik ortamlarını analiz etmenin artan önemi nedeniyle konuşma dışı seslerden duygu tanıma araştırmaları ivme kazanmaktadır. Bu alan, konuşmanın olmadığı veya kullanılmadığı bağlamlarda duygusal ifadeyi anlamak için önemli bir potansiyel oluşturmaktadır. Derin öğrenme metodolojilerindeki önemli ilerlemelerle birlikte çeşitli ve açıklamalı veri kümelerinin artan kullanılabilirliği, daha sağlam ve doğru duygu tanıma sistemlerinin geliştirilmesinde etkili olmuştur. Derin öğrenme modelleri, ham ses verilerinden karmaşık temsilleri otomatik olarak öğrenme yetenekleriyle, elle hazırlanmış özelliklere dayanan geleneksel makine öğrenimi yaklaşımlarından daha iyi performans göstermektedir. Konuşma için IEMOCAP, RAVDESS, CREMA-D ve konuşma dışı sesler için ESC-50, UrbanSound8K ve AudioSet gibi daha büyük ve daha çeşitli veri kümelerinin kullanılabilirliği, daha genelleştirilebilir modellerin eğitilmesini sağlamaktadır. Bu ilerlemelere rağmen, bazı zorluklar devam etmektedir. Temel zorluklardan biri, özellikle birden fazla ses olayının aynı anda gerçekleştiği konuşma dışı ses tanıma sürecinde ortaya çıkan üst üste binen sesler sorunudur.

Bu örtüşen seslerin ayrıştırılması ve hedeflenen duygusal ipuçlarının izole edilmesi, ileri düzey sinyal işleme ve makine öğrenimi tekniklerini zorunlu kılmaktadır. Konuşmadaki konuşmacı değişkenliği ve konuşma dışı ses ifadesindeki bireysel farklılıklar da bir zorluk teşkil etmektedir. Modellerin görünmeyen bireylere iyi genelleme yapabilmesi için bu varyasyonlara karşı dayanıklı olması gerekmektedir. Ayrıca hem konuşma prozodisinde hem de konuşma tabanlı olmayan seslendirmelerde duygusal ifadelerdeki kültürel farklılıklar, kapsamlı bir analiz ve değerlendirme sürecini gerektirmektedir. Bir kültüre özgü verilerle eğitilen modeller, farklı kültürel

kökenlere sahip veriler üzerinde yetersiz performans sergileyebilmektedir. Bu kültürel nüansların göz önünde bulundurulması, kültürel açıdan duyarlı modellerin tasarlanması ve kültürler arası veri kümelerinin geliştirilmesini zorunlu hale getirmektedir. Ses temelli duygu tanıma alanında gelecekteki araştırma eğilimleri, birkaç temel alana yoğunlaşmalıdır:

- **Çok modlu veri entegrasyonu:** İnsan duygusal ifadesi doğası gereği çok modludur ve sözlü, sözsüz, görsel ve fizyolojik ipuçlarının bir kombinasyonunu içermektedir. Konuşma, yüz ifadeleri, vücut dili ve fizyolojik sinyaller gibi çoklu modalitelerden gelen bilgilerin entegre edilmesi, duygu tanıma sistemlerinin doğruluğunu ve sağlamlığını önemli ölçüde artırabilmektedir. Bu alandaki araştırmalar, zamansal hizalama ve özellik ağırlıklandırma gibi konuları ele alarak farklı modalitelerden gelen bilgileri birleştirmek için etkili yöntemleri keşfetmektedir.
- **Farklı bağlamlar arasında genelleme:** Duygu tanıma modelleri, farklı akustik ortamlar, kayıt koşulları ve kültürel ortamlar dâhil olmak üzere çeşitli bağlamlar arasında iyi genelleme yapabilmektedir. Bu durum, belirli faktörlerdeki değişikliklere karşı dayanıklı modellerin geliştirilmesini gerektirmektedir. Genelleme performansını artırmak için veri artırımı, etki alanı uyarlaması ve transfer öğrenimi gibi teknikler kullanılabilir.
- **Gerçek dünyadaki zorluklar:** Seslerden duygu tanımanın gerçek dünya uygulamaları genellikle gürültülü ve öngörülemez ortamları içermektedir. Gelecekteki araştırmalar gürültüye, yankılanmaya ve diğer akustik bozulmalara karşı dayanıklı modeller geliştirmeye odaklanmalıdır. Ayrıca, bu sistemlerin gerçek dünya uygulamalarında kullanılmasını sağlamak için hesaplama verimliliği ve ölçeklenebilirlikle ilgili konuların ele alınması gerekmektedir. Bu, model sıkıştırma tekniklerini keşfetmeyi, özellik çıkarma yöntemlerini optimize etmeyi ve verimli çıkarım algoritmaları geliştirmeyi içermektedir.
- **Etik hususlar:** Duygu tanıma teknolojisi ilerledikçe, kullanımının etik sonuçlarını göz önünde bulundurmak çok önemlidir. Gizlilik, önyargı ve teknolojinin potansiyel kötüye kullanımı gibi konuların dikkatle ele alınması gerekmektedir. Araştırmalar, duygu tanıma sistemlerinin geliştirilmesi ve yaygınlaştırılması için sorumlu ve etik yönergeler geliştirmeye odaklanmaktadır.

Bu zorlukların aşılması ve belirtilen araştırma yönergelerinin izlenmesiyle, ses temelli duygu tanıma alanı, gerçek dünya uygulamalarında etkili bir şekilde kullanılacak daha doğru, güvenilir ve ölçeklenebilir sistemlerin geliştirilmesine yönelik önemli adımlar atabilmektedir. Bu gelişmeler, ruh sağlığı izleme, kamu güvenliği, müşteri hizmetleri, eğlence ve insan-bilgisayar etkileşimi gibi alanlarda önemli ilerlemelere yol açabilmektedir.

2.4.5. Literatürdeki Boşluklar

Ses sinyallerine dayalı sağlam ve genelleştirilebilir şiddet tespit sistemlerinin ilerlemesinin önündeki temel engel, mevcut veri kümelerinin taşıdığı sınırlılıklardır. Bu sınırlamalar aşağıdaki kategoriler altında incelenebilmektedir:

Sınırlı ve Dengesiz Veri Kümeleri

Bu tür sistemlerin performansı ve güvenilirliği kritik ölçüde eğitim verilerinin miktarına, çeşitliliğine ve dengesine bağlıdır. Mevcut veri kümeleri sıklıkla bu önemli hususlarda eksiklikler göstermekte ve etkili modellerin geliştirilmesini engellemektedir.

- **Sınırlı örneklem büyüklüğü:** Şiddet tespiti araştırmalarında kullanılan birçok veri kümesi, genellikle yalnızca birkaç yüz veya birkaç bin ses örneğinden oluşan nispeten küçük örnek boyutlarından oluşmaktadır. Bu veri azlığı, karmaşık örüntüleri öğrenmek ve aşırı uyumdan kaçınmak için önemli miktarda veri gerektiren karmaşık makine öğrenimi modellerinin, özellikle de derin öğrenme mimarilerinin eğitimi için önemli bir zorluk teşkil etmektedir. Yetersiz veri, görülmeyen verilere iyi genelleme yapamayan modellere yol açarak gerçek dünya senaryolarında uygulanabilirliklerini sınırlayabilmektedir.
- **Çeşitlilik eksikliği:** Sağlam şiddet tespit sistemleri geliştirilmesinde temel bir gereklilik, eğitim veri kümelerine çeşitli ses olaylarının eklenmesidir. Bununla birlikte, mevcut veri kümelerinin önemli bir bölümü, çığlık, silah sesi veya cam kırılması gibi şiddetle bağlantılı seslerin yalnızca dar bir alt kümesini kapsamaktadır. Bu dar bakış açısı, fiziksel kavga sesleri, patlamalar veya bağlamsal olarak şiddet göstergesi olabilecek diğer ses olayları gibi oldukça geniş olan akustik ipuçları yelpazesini dikkate almamaktadır. Eğitim verilerinde çeşitli ses olaylarının olmaması, aşırı uzmanlaşmış ve daha karmaşık veya nüanslı akustik ortamlarda şiddeti tespit edemeyen modellerle sonuçlanabilmektedir. Ayrıca, eğitim veri kümelerinde çeşitli arka plan seslerinin ve çevresel varyasyonların olmaması, bir modelin çeşitli gerçek dünya koşullarında doğru performans gösterme yeteneğini daha da sınırlayabilmektedir.
- **Dengesiz sınıflar:** Ses olaylarının gerçek dünya kayıtları, genellikle şiddet içermeyen seslerin, şiddet içeren seslere göre önemli ölçüde daha fazla olduğu bir sınıf dengesizliği göstermektedir. Bu dengesizlik, şiddet içeren olayların, şiddet içermeyen olaylara kıyasla daha az sıklıkta meydana gelmesi gerçeğinden kaynaklanmaktadır. Makine öğrenimi modellerini dengesiz veri kümeleri üzerinde eğitirken, modeller çoğunluk sınıfına (şiddet içermeyen sesler) doğru önyargılı olma eğilimindedir ve bu da azınlık sınıfı (şiddet içeren sesler) için zayıf geri çağırma ile sonuçlanmaktadır. Şiddet tespiti bağlamında, bu önyargı ciddi sonuçlar doğurabilmektedir. Çünkü gerçek şiddet olaylarının gözden kaçmasına neden olabilmekte ve bu da önemli güvenlik etkileri ortaya çıkarmaktadır.

Örneğin, veri seti sınırlamalarına dair yapılan bazı çalışmalar, ESC-50 ve AudioSet gibi kamuya açık veri setlerini veya özel olarak geliştirilmiş veri setlerini kullanmaktadır. Bununla birlikte, bu kaynaklar genellikle yukarıda belirtilen sınırlamaları karşılamakta yetersiz kalmaktadır. ESC-50, çeşitli çevresel sesler içermesine karşın, şiddetle ilişkili gerçek dünya akustik sahnelerinin karmaşıklığını yeterince yansıtamamaktadır. AudioSet, büyük ölçekli bir veri kümesi olmakla birlikte, özellikle şiddetle ilgili sesler için yeterli açıklamalara sahip olmayabilir ve mevcut açıklamalar gürültülü olma eğilimindedir. Özel olarak seçilmiş veri kümeleri, belirli uygulamalara uyarlanabilme potansiyeli taşısa da genellikle küçük örneklem boyutları ve sınırlı çeşitlilik gibi sorunlarla karşılaşmaktadır. Bu sınırlamalar, ses sinyallerine dayalı etkili şiddet tespit sistemlerinin geliştirilmesini desteklemek için daha büyük, çeşitli ve dengeli veri kümelerine duyulan kritik ihtiyacı açıkça göstermektedir.

Farklı Ortamlar Arasında Zayıf Genelleme

Ses analizine dayalı şiddet tespit sistemlerinin karşılaştığı önemli zorluklardan biri, farklı ortamlar arasında genelleme yapma yeteneklerinin sınırlı olmasıdır. Bu genelleme açığı birkaç temel faktörden kaynaklanmaktadır. Bunlarla ilgili bilgiler bu bölümde anlatılmaktadır.

- **Bağlama özel eğitim:** Mevcut modellerin çoğu kontrollü veya laboratuvar benzeri ortamlarda toplanan veri kümeleri üzerinde eğitilmiştir. Bu ortamlar gerçek dünyadaki akustik senaryoların karmaşıklığını ve çeşitliliğini tam olarak yansıtmamaktadır. Sonuç olarak, bu tür veri kümeleri üzerinde eğitilen modeller, değişen akustik koşullar, arka plan gürültü seviyeleri ve çevresel yankılanmalar ile gerçek dünya ortamlarında kullanıldığında düşük performans sergileyebilmektedir.
- **Veri kümeleri arası değerlendirme eksikliği:** Birçok çalışmada yaygın bir uygulama, modelleri yalnızca eğitim için kullanılan veri kümesi veya tek, sınırlı bir değerlendirme veri kümesi üzerinde değerlendirmektir. Bu yaklaşım, bir modelin gerçek genelleme yeteneklerinin eksik bir resmini sunmaktadır. Modelleri farklı ortamlarda ve farklı koşullar altında toplanan çok sayıda ve çeşitli veri kümeleri üzerinde değerlendirmek, sağlamlıklarını değerlendirmek ve olası sınırlamaları belirlemek için çok önemlidir.
- **Ses özelliklerinde değişkenlik:** Bağırma veya silah sesi gibi belirli bir ses olayının akustik özellikleri, çevredeki ortama bağlı olarak önemli ölçüde değişebilmektedir. Bu faktörler şu şekilde açıklanmaktadır:
 - **Konum:** Bir sesin akustik özellikleri, olayın iç veya dış mekânda, kapalı bir alanda veya açık bir alanda gerçekleşmesine bağlı olarak büyük ölçüde farklılık gösterebilmektedir.

- **Mikrofondan uzaklık:** Ses kaynağı ile kayıt cihazı arasındaki mesafe, alınan sinyali önemli ölçüde etkileyerek yoğunluğunu, frekans içeriğini ve genel özelliklerini etkilemektedir.
- **Ortaman akustik özellikleri:** Yankılanma, yankılar ve arka plan gürültü seviyeleri gibi faktörler algılanan sesi önemli ölçüde değiştirebilmekte ve modellerin şiddet olaylarını doğru bir şekilde tanımlamasını ve sınıflandırmasını zorlaştırabilmektedir.

Örneğin, sessiz bir laboratuvar ortamında toplanan ses kayıtlarından oluşan bir veri kümesi üzerinde eğitilen bir model, bu belirli ortamda iyi performans gösterebilmektedir. Ancak, kalabalık bir tren istasyonu veya hareketli bir gece kulübü gibi gürültülü bir ortamda kullanıldığında, model arka plan gürültüsünün varlığı ve hedef seslerin maskelenmesi nedeniyle şiddet olaylarını doğru bir şekilde tespit etmekte zorlanabilmektedir. Bu durum, modellerin akustik ortamlardaki değişikliklere karşı dirençli olması ve çeşitli ve zorlu gerçek dünya senaryolarında başarılı bir şekilde genelleme yapabilmesi için kritik bir gereksinimi vurgulamaktadır.

Bu zorluğun üstesinden gelmek, gerçek dünya ortamlarının akustik çeşitliliğini yakalayan veri kümelerinin geliştirilmesini ve bu farklı koşullardan etkili bir şekilde öğrenebilen ve genelleştirebilen sağlam makine öğrenimi modellerinin geliştirilmesini gerektirmektedir.

Gürültü ve Çakışan Seslerin Yetersiz Ele Alınması

Güvenilir şiddet tespit sistemlerinin geliştirilmesindeki en önemli engellerden biri, gerçek dünya ortamlarındaki gürültü ve üst üste binen seslerin yetersiz şekilde ele alınmasıdır. Bu sistemler, arka plan gürültüsü, üst üste binen konuşma ve birden fazla eş zamanlı ses olayının yaygın olduğu karmaşık akustik senaryolarda etkili bir şekilde çalışmaktadır. Ancak mevcut birçok model bu zorlukları yeterince ele alamamaktadır [74].

- **Yetersiz gürültü ön işleme:** Birçok çalışma, basit filtreleme yöntemleri gibi temel gürültü azaltma tekniklerine dayanmaktadır. Bu teknikler ses sinyallerinden karmaşık arka plan gürültüsünü etkili bir şekilde kaldırmak için yeterli olmamaktadır. Spektral çıkarma, Wiener filtreleme veya derin öğrenme tabanlı yöntemler gibi gelişmiş gürültü azaltma teknikleri, gürültünün etkisini azaltmak ve SNR'yi iyileştirmek için oldukça önemlidir. Bu teknikler, çığlıklar veya silah sesleri gibi hedef seslerin netliğini artırmaya yardımcı olarak model tarafından daha kolay tespit edilebilir hale getirebilmektedir [16].
- **Ses olaylarının üst üste binmesi:** Gerçek dünyadaki şiddet olayları genellikle birden fazla eşzamanlı ses kaynağı içermektedir. Örneğin, fiziksel bir kavga bağırma, boğuşma ve fırlatılan veya kırılan nesnelerin çarpma seslerini içerebilmektedir. Seslerin bu şekilde üst üste binmesi, şiddetle ilişkili belirli akustik ipuçlarının izole edilmesini ve tanımlanmasını zorlaştırmaktadır. Örtüşen ses kaynaklarını ayırmak ve ilgili sesleri izole etmek için kör kaynak ayırma ve CASA gibi sofistike sinyal işleme teknikleri gerekli olmaktadır.

- **Dinamik arka plan koşulları:** Şehir sokakları, alışveriş merkezleri veya stadyumlar gibi gerçek dünya ortamları dinamik ve öngörülemeyen arka plan gürültü seviyeleri ile karakterize edilmektedir. Bu ortamlarda, bir aracın geçmesi, inşaat çalışmalarının başlaması veya ani insan akını gibi ani ve önemli gürültü seviyeleri değişiklikleri yaşanabilmektedir. Arka plan gürültüsündeki bu dalgalanmalar şiddet algılama sistemlerinin performansını önemli ölçüde düşürebilir, yanlış alarmlara veya kaçırılan tespitlere yol açabilmektedir. Modellerin arka plan gürültü koşullarındaki bu dinamik değişikliklere karşı dayanıklı olması ve değişen gürültü girişim seviyelerine uyum sağlaması gerekmektedir.

Örneğin, arka plan gürültüsünden arındırılmış temiz ses örneklerinden oluşan bir veri kümesi üzerinde eğitilen bir makine öğrenimi modeli, kontrollü laboratuvar ortamlarında iyi performans gösterebilmektedir. Ancak, araba kornaları, sirenler veya şehir trafiğinin sürekli uğultusunun olduğu bir ortamda kullanıldığında, modelin performansı önemli ölçüde düşebilmektedir. Arka plan gürültüsü şiddet seslerini maskeleyerek tespitlerin gözden kaçmasına veya yanlış alarmlara yol açabilmektedir. Bu zorluğun üstesinden gelmek için modellerin gerçekçi arka plan gürültüsü içeren veri kümeleri üzerinde eğitilmesi ve sağlam gürültü azaltma ve sinyal ayırma tekniklerini içermesi gerekmektedir.

Bu zorlukların aşılması ve gelişmiş gürültü işleme tekniklerinin sisteme entegre edilmesiyle, şiddet algılama sistemleri gerçek dünya ortamlarında daha sağlam ve güvenilir bir hale gelebilmekte ve şiddet olaylarını tespit etme ile bunlara müdahale etme konusundaki performanslarını önemli ölçüde artırabilmektedir.

Sınırlı Özellik Mühendisliği ve Temsil Öğrenimi

Mevcut şiddet tespit sistemlerindeki önemli bir sınırlama, yetersiz özellik mühendisliği ve temsil öğrenmesi eksikliğinden kaynaklanmaktadır. Birçok çalışma, basit veya yetersiz özellik çıkarma yöntemlerine bağlı kalmakta ve bu durum, şiddet olaylarıyla ilişkili ses sinyallerinin doğasında bulunan karmaşık örüntüleri etkili bir şekilde yakalama kapasitelerini önemli ölçüde kısıtlamaktadır [75].

- **Klasik özelliklere güven:** Yaygın bir sınırlama, MFCC'ler, spektral kontrast veya sıfır geçiş oranları gibi geleneksel, elle hazırlanmış özelliklere aşırı güvenmek olarak bilinmektedir. Bu özelliklerin konuşma tanımda yararlı olduğu kanıtlanmış olsa da bir çığlığın aniden başlaması, bir nesnenin çarpması veya bir mücadeleyle ilişkili enerjideki hızlı dalgalanmalar gibi şiddetle ilgili seslerin nüanslı özelliklerini yeterince yakalayamamaktadır. Bu özellikler, karmaşık akustik ortamlarda şiddet içeren ve içermeyen sesleri doğru bir şekilde ayırt etmek için yeterince ayırt edici olmamaktadır.

- **Derin temsil öğrenimi eksikliği:** Birçok çalışma hala SVM'ler veya Rastgele Ormanlar gibi elle hazırlanmış özelliklere dayanan sığ makine öğrenimi modellerini kullanmaktadır. Bu modeller, ses verilerinin hiyerarşik temsillerini otomatik olarak öğrenme yeteneğinden yoksundur ve akustik özellikler ile şiddetin varlığı arasındaki karmaşık ve doğrusal olmayan ilişkileri yakalama yeteneklerini sınırlamaktadır. CNN'ler, RNN'ler ve Dönüştürücüler gibi derin öğrenme modelleri, ayırt edici özellikleri doğrudan ham ses verilerinden öğrenmede üstün performans göstermektedir. Ancak, şiddet tespiti alanında yapılan birçok çalışma bu derin öğrenme mimarilerinin gücünden tam olarak yararlanamamıştır.

Zamansal ve Bağlamsal Bilgilerin Yetersiz Kullanımı

Şiddetle ilgili sesler genellikle karmaşık zamansal dinamikler sergilemektedir. Örneğin bir şiddet olayı, ani bir bağırış ve ardından bir nesnenin çarpması ve ardından bir mücadele dönemi gibi bir dizi ses verisi içerebilmektedir. Mevcut birçok model, öncelikle kısa vadeli spektral veya zamansal özelliklere odaklanarak bu zamansal bağımlılıkları etkili bir şekilde yakalayamamaktadır. Bu sınırlama, zaman içinde ortaya çıkan olayları doğru bir şekilde tespit etme ve sınıflandırma yeteneklerini engelleyebilmektedir. Ayrıca birçok model, çevredeki akustik ortam veya ses sahnesinin zamansal gelişimi gibi tek tek seslerin önemini doğru bir şekilde yorumlamak için çok önemli olabilecek bağlamsal bilgileri dâhil etmeyi ihmal etmektedir. Örneğin yalnızca spektrogram tabanlı özelliklere sahip CNN'lerin kullanıldığı bir çalışma, bir sesin spektral özelliklerini etkili bir şekilde yakalayabilir ancak olayın zamansal gelişimini yakalamakta başarısız olabilmektedir. Örneğin, daha ciddi veya uzun süreli bir şiddet olayına işaret edebilecek tek bir izole çığlık ile hızlı bir şekilde art arda gelen bir dizi çığlık arasında yeterli ayrım yapamamaktadır [76].

Bu sınırlamaların ele alınması, daha gelişmiş özellik mühendisliği ve temsil öğrenme tekniklerine geçilmesini gerektirmektedir. Bu durum, CNN'ler, RNN'ler ve Dönüştürücüler gibi ses verilerinin hiyerarşik temsillerini otomatik olarak öğrenebilen derin öğrenme mimarilerinin kullanımını araştırmayı içermektedir. Ayrıca, bu modellere dikkat mekanizmalarının dâhil edilmesi, ses sinyali içindeki en ilgili zamansal bölümlere ve bağlamsal bilgilere odaklanmalarını sağlayabilmektedir. Bu gelişmiş tekniklerden yararlanarak, şiddet tespit sistemleri ses verilerindeki şiddetle ilişkili karmaşık örüntülerin daha doğru ve sağlam bir şekilde anlaşılmasını sağlayabilmektedir.

Gerçek Zamanlı ve Düşük Gecikmeli Sistemlerin Eksikliği

Mevcut şiddet tespit modellerinin çoğu öncelikle çevrimdışı işleme için tasarlanmıştır ve gözetim sistemleri, acil durum müdahalesi ve canlı yayınlar gibi önemli gerçek zamanlı

senaryolarda uygulanabilirliklerini engellemektedir. Bu sınırlama birkaç temel faktörden kaynaklanmaktadır:

- **Yüksek hesaplama maliyeti:** Derin öğrenme modelleri, özellikle de transformatörler ve RNN'ler gibi karmaşık mimariler kullananlar, genellikle yüksek hesaplama talepleri sergilemektedir. Bu da önemli miktarda işlem gücü gerektirdiğinden bu modellerin uç cihazlar, cep telefonları veya gömülü sistemler gibi kaynak kısıtlaması olan cihazlara yerleştirilmesini zorlaştırmaktadır. Bu sınırlama, anında tespit ve müdahalenin kritik olduğu gerçek zamanlı uygulamalarda kullanımlarını kısıtlamaktadır.
- **Gecikmeli karar verme:** Yüksek doğruluk elde etmek için bazı modeller, tipik olarak birkaç saniyeye (örn. 5-10 saniye) yayılan nispeten büyük ses parçalarını işlemektedir. Bu yaklaşım doğruluğu artırırken, tespit sürecinde önemli bir gecikmeye neden olmaktadır. Bu gecikme, hızlı müdahalenin çok önemli olduğu aktif tetikçi durumları veya diğer acil durumlar gibi zaman açısından kritik senaryolarda zararlı olabilmektedir.
- **Dağıtım için sınırlı optimizasyon:** Gerçek zamanlı dağıtım için şiddet tespit modellerinin optimizasyonunda önemli bir boşluk bulunmaktadır. Çok az sayıda çalışma, düşük güçlü cihazlar için modellerin optimize edilmesine, gecikmenin en aza indirilmesine ve verimli kaynak kullanımının sağlanmasına öncelik vermektedir. Sonuç olarak, mevcut birçok model, kaynakların kısıtlı olduğu ortamlarda veya gerçek zamanlı performansın kritik olduğu uygulamalarda kolayca kullanılamamaktadır.

Örneğin, Aktif tetikçi durumları gibi kolluk kuvvetleri uygulamalarında hızlı tespit ve müdahale çok önemli bir yer tutmaktadır. Bir uyarı vermeden önce ses verilerini işlemek için birkaç saniyeye ihtiyaç duyan bir sistem, gerçek zamanlı veya gerçek zamana yakın algılama sağlayabilen bir sistemden önemli ölçüde daha az etkili olma potansiyeline sahiptir.

Bu sınırlamaların ele alınması, araştırma odağının hafif ve hesaplama açısından verimli modeller geliştirmeye, gerçek zamanlı çıkarım tekniklerini keşfetmeye ve uç cihazlarda dağıtım için modelleri optimize etmeye doğru kaydırılmasını gerektirmektedir. Bu durum, gerçek dünya senaryolarında minimum gecikme ve ek hesaplama yükü ile kullanılabilir pratik ve etkili şiddet tespit sistemlerinin geliştirilmesini sağlayacaktır.

Etik ve Gizlilik Kaygıları

Ses analizi kullanan şiddet tespit sistemlerinin geliştirilmesi ve uygulanması, dikkatle değerlendirilmesi gereken önemli etik ve mahremiyet kaygılarını gündeme getirmektedir:

- **Kötüye kullanım ve gözetim riskleri:** Bu teknolojilerin kötüye kullanım potansiyeli büyük bir endişe kaynağıdır. Sesli gözetim sistemleri kitlesel gözetim için kullanılabilir, bireylerin izinsiz izlenmesini sağlayabilir ve mahremiyet haklarını ihlal edebilmektedir.

Hükümetlerin veya diğer kuruluşların bu sistemleri sosyal kontrol veya siyasi baskı için kullanma olasılığı göz ardı edilemez. Bu sistemler belirli bireyleri veya grupları haksız yere hedef almak için kullanılabilir, bu da potansiyel ayrımcılığa ve sosyal adaletsizliğe yol açabilmektedir.

- **Veri kümeleri ve modellerde önyargı:** Bu modelleri eğitmek için kullanılan veri kümeleri belirli kültürel veya sosyal bağlamları yansıtacak şekilde önyargılıysa, ortaya çıkan modeller yanlış ve adil olmayan tahminlere yol açabilecek eğilimler sergilemektedir. Örneğin, öncelikle kentsel ortamlardan gelen verilerle eğitilen bir model, kırsal ortamlarda iyi performans göstermeyebilir ve potansiyel olarak gözden kaçan tespitlere veya yanlış alarmlara yol açabilmektedir. Bu önyargıların ele alınması, adalet ve eşitliği sağlamak için veri toplama, iyileştirme ve model geliştirmenin dikkatli bir şekilde değerlendirilmesini gerektirmektedir.

- **Yanlış pozitifler ve yanlış negatifler:** Sistemin şiddet içermeyen sesleri yanlışlıkla şiddet olayları olarak tanımladığı yanlış pozitifler gereksiz müdahalelere yol açabilmekte, kaynakları boşa harcayabilmekte ve potansiyel olarak bireylerde sıkıntı veya rahatsızlığa neden olabilmektedir.

Tersine, sistemin gerçek şiddet olaylarını tespit edemediği yanlış negatifler ciddi sonuçlar doğurabilir, potansiyel olarak hayatları tehlikeye atabilir ve zamanında müdahale çabalarını engelleyebilmektedir. Bu hataların en aza indirilmesi için dikkatli model geliştirme, titiz testler ve sistemin sürekli olarak izlenmesi ve iyileştirilmesi gerekmektedir. Örneğin, Alışveriş merkezi gibi kamusal bir alanda kullanılan bir şiddet tespit sistemi, gerçek şiddet olayları (örn. çığlıklar, silah sesleri vb.) ile yüksek sesli konuşmalar, oyun oynayan çocuklar veya inşaat sesleri gibi yanlışlıkla şiddet olarak yorumlanabilecek günlük sesleri doğru bir şekilde ayırt edebilmelidir.

Bu etik ve mahremiyet kaygılarının ele alınması, aşağıdakileri de içeren çok yönlü bir yaklaşım gerektirmektedir:

- **Şeffaflık ve hesap verilebilirlik:** Veri toplama, kullanma ve depolamaya ilişkin açık kılavuzlar ve düzenlemeler de dâhil olmak üzere bu sistemlerin geliştirilmesi, dağıtımı ve kullanımında şeffaflığın sağlanması gerekmektedir
- **Sağlam etik çerçeveler:** Bireysel gizliliğe, adalete ve bu teknolojilerin sorumlu kullanımına öncelik veren etik kurallar ve ilkeler geliştirmek ve bunlara bağlı kalmak gerekmektedir.
- **Sürekli izleme ve değerlendirme:** Bu sistemlerin performansının düzenli olarak değerlendirilmesi ve ortaya çıkabilecek önyargıların veya yanlışlıkların ele alınması gerekmektedir.

- **Halkın katılımı ve diyalog:** Bu teknolojilerin geliştirilmesi ve yaygınlaştırılmasının toplumsal değerler ve kaygılarla uyumlu olmasını sağlamak için açık diyalog ve kamu katılımının teşvik edilmesi gerekmektedir.

Bu etik ve gizlilik kaygıları dikkatle değerlendirilerek ve uygun tedbirler uygulanarak, şiddet tespit sistemlerinin geliştirilmesi ve yaygınlaştırılmasının sorumlu ve etik bir şekilde yürütülmesi sağlanabilmekte, potansiyel riskler en aza indirilirken faydaları en üst düzeye çıkarabilmektedir.

2.4.6. Literatür Boşlukların Üstesinden Gelmek için Geleceğe Yönelik Öneriler

Bu bölümde boşlukları gidermek için gelecekteki araştırmalar ile ilgili bilgiler verilmektedir.

Veri Kaynaklarının Geliştirilmesi

- **Daha büyük ve daha çeşitli veri setleri:** Daha büyük ve daha çeşitli veri setlerinin geliştirilmesi çok önemlidir. Bu veri setleri, kentsel, banliyö ve kırsal ortamların yanı sıra çeşitli kültürel ve sosyo-ekonomik bağlamlar da dâhil olmak üzere daha geniş bir akustik ortam yelpazesini kapsamaktadır.
- **Arka plan gürültüsü ve gerçek dünya değişkenliğinin dâhil edilmesi:** Veri setleri, gerçek dünya koşullarını daha doğru bir şekilde simüle etmek için trafik, inşaat ve çevresel sesler gibi gerçekçi arka plan gürültüsünü içermektedir.
- **Sınıf dengesizliğinin ele alınması:** Sağlam model performansı sağlamak için azınlık sınıflarını fazla örnekleme, çoğunluk sınıflarını az örnekleme veya maliyete duyarlı öğrenme algoritmaları kullanma gibi sınıf dengesizliğini ele alma stratejileri araştırılmaktadır.

Genelleme ve Sağlamlığın Geliştirilmesi

- **Transfer öğrenimi ve etki alanı adaptasyonu:** Büyük, genel amaçlı ses veri kümelerinden elde edilen bilgilerden yararlanmak ve bunları belirli şiddet tespit görevlerine uyarlamak için transfer öğrenimi gibi tekniklerden yararlanılmaktadır. Etki alanı uyarlama teknikleri, modellerin yeni ve görülmemiş ortamlara daha iyi genelleştirilmesine yardımcı olabilmektedir.
- **Veri büyütme:** Mevcut veri kümelerinin gürültü enjeksiyonu, zaman uzatma ve perde kaydırma gibi tekniklerle zenginleştirilmesi, akustik koşullardaki değişikliklere karşı modelin sağlamlığını artırabilmektedir.

Gelişmiş Özellik Temsili ve Öğrenme

- **Derin öğrenme mimarileri:** Dönüştürücüler, kendi kendine dikkat mekanizmaları ve CNN'ler ile RNN'leri birleştiren hibrit modeller gibi gelişmiş derin öğrenme mimarilerini keşfetmek, ilgili özelliklerin çıkarılmasını iyileştirebilir ve ses sinyallerindeki karmaşık zamansal bağımlılıkları yakalayabilmektedir [77].
- **Kendi kendine denetimli öğrenme:** Kontrastlı öğrenme veya oto kodlayıcılar gibi kendi kendine denetimli öğrenme tekniklerinin kullanılması, büyük miktarda etiketlenmemiş ses verisinden sağlam temsiller öğrenebilir ve daha sonra belirli şiddet tespit görevleri için ince ayar yapılabilir.

Gerçek Zamanlı Performans ve Optimizasyon

- **Hafif model geliştirme:** MobileNet, EfficientNet veya ses işleme için özel mimariler gibi hafif ve hesaplama açısından verimli modeller geliştirmek, uç cihazlarda gerçek zamanlı dağıtım için çok önemlidir.
- **Model sıkıştırma ve niceleme:** Model budama, niceleme ve bilgi damıtma gibi teknikler, performans düşüşünü en aza indirirken model boyutunu ve hesaplama karmaşıklığını azaltmak için kullanılabilir.
- **Donanım Hızlandırma:** Özel donanım hızlandırıcıları (örn. Tensör işleme birimleri (TPU'lar)) veya sahada programlanabilir kapı dizileri (FPGA'lar) gibi donanım hızlandırma tekniklerini keşfetmek, çıkarım hızını ve verimliliğini önemli ölçüde artırabilmektedir [78].

Etik Hususlar ve Sorumlu Dağıtım

- **Gizliliği koruyan teknikler:** Kullanıcı verilerini korumak ve gizlilik risklerini en aza indirmek için federe öğrenme ve diferansiyel gizlilik gibi gizliliği koruyan tekniklerin uygulanması gerekmektedir.
- **Önyargı azaltma:** Veri kümeleri ve modellerdeki önyargıları azaltmak için stratejiler geliştirmek ve uygulamak, farklı popülasyonlar ve demografik özellikler arasında adil ve eşit performans sağlamak gerekmektedir.
- **Şeffaflık ve açıklanabilirlik:** Güveni artırmak ve sorumlu kullanımı kolaylaştırmak için şiddet tespit modellerinin şeffaflığını ve açıklanabilirliğini artırmak gerekmektedir.
- **Etik yönergeler ve düzenlemeler:** Bu teknolojilerin geliştirilmesi, dağıtımı ve kullanımını düzenleyen etik yönergelere ve yönetmeliklere uyulması, sorumlu ve etik kullanımlarının sağlanması gerekmektedir.

Bu zorluklar ele alınarak ve bu araştırma yönergeleri takip edilerek, potansiyel riskler azaltılırken ve sorumlu kullanım sağlarken gerçek dünyadaki ihtiyaçlar etkili bir şekilde karşılanabilecek daha sağlam, güvenilir ve etik şiddet tespit sistemleri geliştirilebilecektir.

2.4.7. Değerlendirme

Makine öğrenimi tabanlı ses ve şiddet tespiti önemli ilerleme kaydetmiş olsa da sınırlı veri kümeleri, zayıf gürültü işleme, yetersiz genelleme ve gerçek zamanlı kısıtlamalar dâhil olmak üzere bazı kritik boşluklar devam etmektedir. Bu eksikliklerin giderilmesi, daha doğru, adil ve pratik şiddet tespit sistemleri geliştirmek için gelişmiş derin öğrenme teknikleri, sağlam veri toplama stratejileri ve etik hususların bir kombinasyonunu gerektirecektir.



3. SES ANALİZİNİN TEMELLERİ

Bu bölümde ses analizinin temelleri ile ilgili detaylı açıklamalar yapılmaktadır.

3.1. Seslerin Akustik Özellikleri

Mekanik enerjinin bir formu olan ses, hava, su ve katı cisimler gibi çeşitli ortamlarda dalga olarak yayılmaktadır. Bu dalga benzeri hareket, titreşimlerin ve basınç değişimlerinin iletilmesini içermektedir. Bu ses sinyallerinin analizi, konuşulan dili anlamak için konuşmanın akustik özelliklerinin deşifre edildiği konuşma tanıma dâhil olmak üzere birçok alanda çok önemlidir. Duygu tespitinde, perde ve tonlama gibi ses ipuçlarının analizi duygusal durumların ayırt edilmesine yardımcı olmaktadır [79].

Çevresel ses sınıflandırması, belirli bir ortamdaki farklı seslerin benzersiz akustik imzalarını analiz etmeye dayanmaktadır. Ayrıca, tıbbi teşhis, iç organları ve vücut fonksiyonlarını görselleştirmek ve değerlendirmek için ultrason ve ekokardiyografi gibi ses analiz tekniklerini kullanmaktadır. Etkili ses sinyali işleme ve analizi için temel akustik özelliklerin anlaşılması çok önemlidir. Bu özellikler şunları içermektedir:

3.1.1. Ses Perdesi

Perde, insanların sesleri alçaktan yükseğe doğru bir ölçekte sıralamasına olanak tanıyan bir sesin öznel algısal özelliği olarak bilinmektedir. Öncelikle ses kaynağının titreşim oranını temsil eden ses dalgasının temel frekansı (F_0) tarafından belirlenir. F_0 , saniyedeki titreşim döngüsü sayısını temsil eden Hertz (Hz) cinsinden ölçülmektedir.

Perdenin Temel Yönleri

- **Frekans ve perde algısı:** Frekans ve perde arasında doğrudan bir ilişki vardır. Daha yüksek frekanslar genellikle daha tiz seslere karşılık gelirken, daha düşük frekanslar daha alçak seslere neden olmaktadır. Örneğin, bir düdüğün tiz sesi yüksek bir frekansla ilişkilendirilirken, bir bas davulun alçak gürültüsü düşük bir frekansa karşılık gelmektedir.
- **İnsan işitsel sınırları:** İnsan işitme sistemi, tipik olarak 20 Hz ile 20.000 Hz arasında algılayabileceği sınırlı bir frekans aralığına sahiptir. Bu aralık yaşa ve bireysel farklılıklara göre değişmektedir.
- **Doğrusal olmayan algı:** Perde algısı doğrusal değildir. Frekanstaki eşit değişiklikler, perdede eşit değişiklikler olarak algılanmamaktadır. Düşük frekanslardaki frekans değişiklikleri, yüksek frekanslardaki değişikliklere göre algısal olarak daha belirgin olmaktadır. Örneğin, 100 Hz'den 200 Hz'e olan bir değişiklik, 10.000 Hz'den 10.100 Hz'e

olan bir deęişiklikten daha belirgin olmaktadır. Bu olgu, frekans ve algılanan perde arasındaki doğrusal olmayan ilişkiyi modellemeye çalışan Mel ölçeęi olarak bilinmektedir.

Örnek Uygulamalar

- **Konuşma işleme:** Perde, konuşma işlemede çok önemli bir rol oynamaktadır. Temel frekansı etkileyen ses yolu uzunluklarındaki doğal farklılıklar nedeniyle erkek ve kadın konuşmacılar arasında ayırım yapılmasına yardımcı olmaktadır.
- **Müzik analizi:** Müzikte perde temeldir. Perde algılama algoritmaları, aşağıdakiler de dâhil olmak üzere çeşitli müzik uygulamalarında kullanılmaktadır:
 - **Melodi çıkarma,** bir müzik parçasındaki melodik çizgiyi tanımlama olarak bilinmektedir.
 - **Müzikal akortlama,** notalarının frekanslarını analiz ederek bir müzik aletinin akortlu olup olmadığını belirlemek için kullanılmaktadır.
 - **Otomatik transkripsiyon,** ses kayıtlarından müzik notalarının yazıya dökülmesini sağlamaktadır.
- **Duygu tanıma:** Perde deęişimleri duygusal durumun önemli bir göstergesi olarak kabul edilmektedir. Örneęin, heyecanlı veya mutlu konuşmalar genellikle daha yüksek bir perde sergilerken, üzgün veya depresif konuşmalar daha düşük bir perdeye sahip olma eğilimindedir. Perde konturlarını analiz etmek, konuşmanın duygusal içerięi hakkında önemli bilgiler sağlayabilmektedir.

Bu detaylı açıklama, perde kavramının ve bunun akustik ile sinyal işlemenin farklı alanlarındaki öneminin daha kapsamlı bir şekilde kavranmasına olanak tanımaktadır.

3.1.2. Ses Tınısı

Ton olarak da bilinen tını, aynı perde ve ses yüksekliğine sahip sesleri birbirinden ayıran algısal bir nitelik olarak bilinmektedir. Trompet ve klarnet gibi farklı ses kaynaklarını, aynı notayı aynı ses seviyesinde çalıyor olsalar bile ayırt etmemizi sağlamaktadır. Tını, öncelikle ses dalgasının harmonik içerięi ve genel şekli tarafından belirlenmektedir [80].

Ses Tınısının Temel Yönleri

- **Harmonik içerik:** İnsan sesi ve çoęu müzik aleti saf tonlar (tek frekanslar) olmayan sesler üretmektedir. Bunun yerine, temel frekans ve temel frekansın tam sayı katları olan bir dizi üst ton veya harmonik de dâhil olmak üzere karmaşık bir frekans karışımı üretmektedirler. Bu harmoniklerin göreceli genlikleri ve frekansları algılanan tınıyı önemli ölçüde etkilemektedir. Örneęin, aynı notayı çalan bir keman ve bir flüt, farklı harmonik kombinasyonları ve yoğunlukları ürettikleri için farklı tınılara sahip olmaktadır.

- **Dalga formu şekli:** Bir osiloskopta görselleştirilebilen ses dalgasının şekli de tınıyı belirlemede önemli bir rol oynamaktadır. Sinüs dalgaları gibi basit dalga biçimleri, tek bir frekansa sahip saf tonlar üretmektedir. Müzik aletleri veya insan sesleri tarafından üretilenler gibi daha karmaşık dalga biçimleri, zaman içinde genlik ve frekansta değişiklikler sergileyerek daha zengin ve daha karmaşık tınılara neden olmaktadır.

Örnek Uygulamalar

Özet olarak tını, farklı sesleri ayırt etme ve çevremizdeki akustik dünyayı anlama becerimizde önemli bir rol oynayan karmaşık bir algısal özellik olarak göze çarpmaktadır.

- **Müzik işleme:** Tını analizi, müzik işleme ve analizinde temel faktördür.
 - **Enstrüman tanımlama:** Algoritmalar, benzersiz tını özelliklerine dayalı olarak farklı müzik aletlerini tanımlamak için geliştirilebilmektedir.
 - **Müzik türü sınıflandırması:** Tını, her türün genellikle karakteristik tınısal niteliklere sahip olması nedeniyle müziği farklı türlere sınıflandırmak için kullanılabilir.
 - **Ses sentezi:** Sentezleyiciler ve diğer elektronik müzik enstrümanları, akustik enstrümanların tınısal özelliklerini taklit ederek çok çeşitli sesler oluşturmak için tını sentezleme tekniklerini kullanmaktadır.
- **Konuşmacı tanımlama:** Konuşmacı tanımlama ve tespitinde tını oldukça önemli bir detaydır.
 - **Tını:** Konuşmacı tanımlamada çok önemli bir faktördür. Her bireyin ses yolu şekli, ses teli özellikleri ve konuşma tarzı gibi faktörlerden etkilenen benzersiz bir ses tınısı olduğu bilinmektedir.
 - **Tını analizi:** Ses tanıma sistemlerinde bireyleri benzersiz vokal özelliklerine göre tanımlamak ve kimliklerini doğrulamak için kullanılmaktadır.
 - **Ses kaynağı lokalizasyonu:** Karmaşık akustik ortamlarda, tını farklı ses kaynaklarını ayırt etmeye yardımcı olabilmektedir. Örneğin, kalabalık bir odada, farklı tını özelliklerine dayanarak insan konuşması, çalan telefonun sesi ve arka plan müziği ayırt edilebilmektedir.

3.1.3. Ses Yüksekliği

Yükseklik, bir sesin insan kulağına ne kadar güçlü veya yoğun algılandığını belirleyen öznel bir algısal özellik olarak ifade edilmektedir. Bu özellik, ses kaynağının neden olduğu havadaki basınç değişimlerinin büyüklüğünü yansıtan ses dalgasının genliği ile doğrudan ilişkilidir [81].

Ses Yüksekliğinin Temel Yönleri

- **Genlik ve ses yüksekliği:** Daha yüksek genlikli bir ses dalgası genellikle daha yüksek algılanan bir sesle sonuçlanmaktadır. Ancak, genlik ve algılanan ses yüksekliği arasındaki ilişki doğrusal değildir.
- **Logaritmik ölçek:** İnsanların ses yüksekliği algısı logaritmik bir ölçek izlemektedir. Bu durum, ses yoğunluğundaki on katlık bir artışın (metrekare başına Watt olarak ölçülür.) ses yüksekliğinin iki katına çıkması olarak algılandığı anlamına gelmektedir. Bu ilişki desibel (dB) ölçeği kullanılarak ifade edilir; burada 10 dB'lik bir artış ses yoğunluğunda on katlık bir artışa karşılık gelmektedir.
- **Frekans bağımlılığı:** İnsan kulağının sese karşı hassasiyeti frekansa göre değişmektedir. Eşit ses yüksekliği konturları olarak da bilinen Fletcher-Munson eğrileri, kulağın en çok orta frekans aralığındaki (yaklaşık 2-5 kHz) seslere duyarlı olduğunu göstermektedir. Çok düşük frekanslardaki (infrasound) ve çok yüksek frekanslardaki (ultrason) seslerin orta frekans aralığındaki seslerle eşit derecede yüksek olarak algılanması için daha yüksek yoğunluklara ihtiyaç duyulmaktadır.
- **İşitme eşiği ve hasar:** İnsan kulağı tarafından algılanabilen minimum ses yoğunluğu olan işitme eşiği tipik olarak 0 dB civarında olmaktadır. Uzun süre 85 dB üzerindeki seslere maruz kalmak işitme hasarına neden olabilmektedir.

Örnek Uygulamalar

- **Gürültü azaltma:** Gürültü azaltma algoritmalarında ses yüksekliği analizi oldukça önemlidir. Bu algoritmalar, yüksek arka plan gürültüsünü tanımlayarak ve azaltarak konuşmanın ve diğer istenen seslerin anlaşılabilirliğini artırmaktadır.
- **İşitme cihazı teknolojisi:** İşitme cihazları, sesleri seçici olarak yükseltmek için ses yüksekliği analizini kullanmaktadır. Rahatsızlığı önlemek ve kullanıcının işitme duyusunu korumak için daha yüksek seslerin amplifikasyonunu sınırlandırırken daha yumuşak sesleri yükseltebilmektedirler.
- **Konuşma duygusu tanıma:** Ses yüksekliği varyasyonları, duygu tanıma için değerli ipuçları sağlamaktadır. Örneğin, öfke genellikle artan ses yoğunluğu ile ilişkilendirilir ve bu da daha yüksek sesle konuşmaya neden olmaktadır. Tersine, üzüntü veya depresyon daha sessiz konuşmayla ilişkilendirilebilmektedir. Algoritmalar, ses yüksekliğindeki değişimleri analiz ederek konuşmacının duygusal durumunu çıkarabilmektedir.
- **Ses montajı ve iyileştirme:** Müzik prodüksiyonunda ses yüksekliği analizi, ses kayıtlarının genel ses seviyesinin optimize edilmesi amacıyla kullanılmakta olup, farklı cihazlar ve platformlar arasında tutarlı bir dinleme deneyimi sunmaktadır. Bu ayrıntılı

açıklama, ses yüksekliği, ses yoğunluğu ile ilişkisi ve çeşitli uygulamalardaki önemi hakkında kapsamlı bir genel bakış sağlamaktadır.

3.1.4. Sesin Yoğunluğu

Ses yoğunluğu, sesin yayılma yönüne dik bir birim alandan geçen ses gücü miktarını tanımlayan fiziksel bir niceliktir. Ses dalgası tarafından taşınan enerjinin bir ölçüsüdür ve tipik olarak metrekare başına Watt (W/m^2) birimleriyle ifade edilmektedir [82].

Ses Yoğunluğunun Temel Yönleri

- **Genlik ile ilişki:** Yoğunluk, ses dalgasının genliğinin karesi ile doğru orantılıdır. Bu durum, genliğin iki katına çıkmasının ses yoğunluğunda dört kat artışa neden olduğu anlamına gelmektedir.
- **Nesnel ölçüm:** Öznel bir algısal nicelik olan ses yüksekliğinin aksine, yoğunluk ses dalgasının nesnel, ölçülebilir fiziksel bir özelliği olarak bilinmektedir. Ses seviyesi ölçer gibi özel aletler kullanılarak ölçülebilmektedir.
- **Yayılma ve mesafe:** Ses yoğunluğu bir ortamda yayıldıkça azalmaktadır. Bu zayıflamanın başlıca nedeni, ses dalgası kaynaktan uzaklaştıkça ses enerjisinin daha geniş bir alana yayılması olarak bilinmektedir. Bu ilişki, bir ses dalgasının yoğunluğunun kaynaktan uzaklığın karesi ile orantılı olarak azaldığını belirten ters kare yasası ile tanımlanmaktadır. Örneğin bir ses kaynağına olan mesafe iki katına çıkarılırsa, ses yoğunluğu dört kat azalmaktadır.

Örnek Uygulamalar

- **Çevresel gürültü değerlendirmesi:** Yoğunluk ölçümleri, trafik gürültüsü, uçak gürültüsü ve endüstriyel gürültü gibi çevresel gürültü seviyelerini değerlendirmek için oldukça önemlidir. Bu bilgiler şehir planlaması, gürültü kontrol önlemleri ve çevresel etki değerlendirmeleri için kullanılmaktadır.
- **Tıbbi teşhis:** Yoğunluk ölçümleri, aşağıdakiler de dâhil olmak üzere çeşitli tıbbi uygulamalarda kullanılmaktadır:
 - **Ultrason görüntüleme:** İç organların görüntülerini oluşturmak için dokulardan yansıyan ultrason dalgalarının yoğunluğunun analiz edilmesi gerçekleştirilmektedir.
 - **Akustik emisyon testi:** Bütünlüklerini değerlendirmek için malzemeler içinde üretilen akustik emisyonların (yüksek frekanslı stres dalgaları) tespit ve analiz edilmesi gerçekleştirilmektedir.

- **Akciğer ve kalp sesi analizi:** Anormalliklerin belirlenmesi ve kalp üfürümleri veya akciğer tıkanıklığı gibi durumların tanınması için akciğer ve kalp seslerinin yoğunluğunun düzenli olarak izlenmesi büyük önem taşımaktadır.
- **Gözetim sistemleri:** Ses yoğunluğu değişimlerinin analiz edilmesi gözetim sistemlerinde kullanılabilmektedir. Örneğin, silah sesleri veya patlamalarla ilişkili olanlar gibi ses yoğunluğundaki ani artışlar, güvenlik sistemlerinde uyarıları tetikleyebilmektedir. Bu ayrıntılı açıklama, ses yoğunluğu, temel özellikleri ve akustik ve sinyal işlemenin çeşitli alanlarındaki önemi hakkında kapsamlı bir genel bakış sağlamaktadır.

3.1.5. Değerlendirme

Sesin temel akustik özelliklerini (perde, ton, ses yüksekliği ve yoğunluk) anlamak, konuşma işleme, müzik analizi, çevresel izleme ve makine öğrenimi tabanlı ses sınıflandırması gibi çeşitli uygulamalar için gerekli bir durum oluşturmaktadır. Sesin temel akustik özelliklerinin yer aldığı bilgiler Tablo 3.1’de sunulmuştur

Tablo 3.1. Sesin temel akustik özellikleri

Özellik	Tanımlama	Ölçü Birimi	Anahtar Etki	Örnek Uygulamalar
Perde	Bir sesin algılanan frekansı	Hertz (Hz)	Ses Dalgası Frekansı	Konuşma analizi, müzik ayarlama, duygu algılama
Tını	Bir sesin kalitesi veya rengi	Standart birimi yok	Harmonik içerik ve dalga biçimi	Enstrüman sınıflandırma, konuşmacı tanıma
Ses Yüksekliği	Algılanan ses gücü/hacmi	dB	Ses dalgası genliği	Gürültü azaltma, konuşma geliştirme, duygu tanıma
Ses Yoğunluğu	Ses dalgasının fiziksel gücü	Metrekare başına Watt (W/m ²)	Genliğin karesi ve kaynaktan uzaklık	Çevresel gürültü analizi, güvenlik sistemleri

Bu özelliklerin her biri, insanların ve makinelerin ses sinyallerini nasıl algıladığı ve işlediği konusunda farklı bir rol oynamaktadır. Derin öğrenme ve sinyal işleme alanındaki gelecekteki gelişmeler, ses verilerini analiz etme, sınıflandırma ve bu verilerden anlamlı bilgiler çıkarma becerisini geliştirmeye devam edecektir.

Bu akustik özelliklerin etkileşimi, sesin doğasını ve çevremizdeki etkilerini kavramamızın temelini oluşturmaktadır. CNN’ler, RNN’ler ve dönüştürücü modeller gibi derin öğrenme ve sinyal işleme tekniklerindeki devam eden gelişmeler, karmaşık ses sinyallerini analiz etme ve yorumlama yeteneğimizde devrimsel bir dönüşüm yaratma potansiyeli taşımaktadır. Bu ilerlemeler, konuşma tanıma, duygu tanıma, şiddet tespiti, müzik bilgisi çıkarımı, çevresel izleme ve tıbbi teşhis gibi çeşitli alanlarda daha doğru ve etkili çözümler sunmaktadır.

3.2. Sesten Özellik Çıkarma Teknikleri

Özellik çıkarma, ses sinyali analizinde temel bir adımı oluşturmaktadır. Esasen zamanla değişen basınç dalgaları olan ham ses sinyallerinin bir dizi anlamlı sayısal temsile dönüştürülmesini içermektedir. Çıkarılan bu özellikler, ses sinyalinin ayırt edici özelliklerini içermekte ve makine öğrenimi modellerinin farklı ses modellerini etkili bir şekilde anlaması ile ayırt etmesine olanak tanımaktadır.

Bu süreç, insanların sesleri nasıl algıladığı ve yorumladığına benzemektedir. Ham dalga formu doğrudan işlenememektedir. Bunun yerine işitsel sistemden perde, tını, ses yüksekliği ve belirli seslerin varlığı gibi ilgili bilgiler çıkarılmaktadır [83-84].

Özellik çıkarma, ses analizinde birkaç önemli görevi kolaylaştırmaktadır:

- **Boyutsallık azaltma:** Ham ses sinyalleri genellikle yüksek boyutludur ve gereksiz bilgiler içermektedir. Özellik çıkarma, yalnızca en alakalı bilgileri çıkararak verilerin boyutluluğunu azaltmakta, sonraki işlemleri basitleştirmekte ve hesaplama verimliliğini artırmaktadır.
- **Geliştirilmiş model performansı:** Makine öğrenimi modellerine ses sinyalinin özlü ve bilgilendirici bir temsili sağlayarak, özellik çıkarma, aşağıdaki gibi görevlerde performanslarını önemli ölçüde artırır:
 - **Konuşma tanıma:** Konuşulan kelimeleri ve cümleleri doğru bir şekilde tanıma ile ilgili işlemler gerçekleştirilmektedir.
 - **Duygu algılama:** Vokal ipuçlarından duygusal durumları tanımlama ile ilgili işlemler gerçekleştirilmektedir.
 - **Müzik bilgisi tespiti:** Müzik türlerinin sınıflandırılması, enstrümanların tanımlanması ve müzikal özelliklerin çıkarılması ile ilgili işlemler gerçekleştirilmektedir.
 - **Çevresel ses sınıflandırması:** Kuş sesleri, trafik gürültüsü ve doğal sesler gibi çevreden gelen sesleri tanıma ile ilgili işlemler gerçekleştirilmektedir.
- **Geliştirilmiş yorumlanabilirlik:** Çıkarılan özellikler, ses sinyalinin altında yatan özelliklere ilişkin içgörüler sağlayarak sesin yorumlanmasına ve anlaşılmasına yardımcı olabilmektedir. Özellik çıkarma teknikleri genel olarak aşağıdaki gibi kategorize edilebilmektedir:
 - **Zaman alanı özellikleri:** Bu özellikler doğrudan ham ses dalga biçiminden çıkarılmakta ve genlik değişimleri, sıfır geçiş oranları ve enerji gibi zamansal özellikleri yakalamaktadır.
 - **Frekans alanı özellikleri:** Bu özellikler, Fourier Dönüşümü gibi teknikler kullanılarak ses sinyalinin zaman alanından frekans alanına dönüştürülmesiyle elde edilmektedir. Bu, seslerin spektral özelliklerini anlamak için çok önemli olan sinyalin frekans içeriğinin analiz edilmesini sağlamaktadır.

Araştırmacılar ve mühendisler, ilgili özellikleri etkili bir şekilde çıkararak ses verilerinin potansiyelini ortaya çıkarabilmektedir. Ayrıca yapay zekâ, insan-bilgisayar etkileşimi ve multimedya gibi alanlarda geniş bir uygulama yelpazesi sağlayabilmektedir.

3.2.1. Mel-Frekans Cepstral Katsayıları (MFCC)

MFCC'ler, özellikle konuşma ve ses tanıma görevlerinde ses sinyalleri için yaygın olarak kullanılan bir özellik çıkarma tekniği olarak bilinmektedir. Bir sesin spektral özelliklerini insan işitsel algısına yakın bir şekilde etkili bir şekilde yakalamaktadır [85].

MFCC'lerin Hesaplanması

- **Ön vurgu:** Daha yüksek frekansları yükseltmek için ses sinyaline yüksek geçişli bir filtre uygulanmaktadır. Bu durum, farklı sesleri ayırt etmek için genellikle çok önemli olan yüksek frekans bölgelerindeki enerjiyi artırmaya yardımcı olmaktadır.
- **Çerçeveleme:** Ses sinyali kısa, üst üste binen karelere bölünmektedir (tipik olarak 20-25 milisaniye süreli). Bu yaklaşım, ses sinyalinin zamanla değişen özelliklerinin analiz edilmesini sağlamaktadır.
- **Pencereleme:** Her kareye Hamming penceresi gibi bir pencereleme işlevi uygulanmaktadır. Bu, bir karenin ani başlangıcı ve sonu sinyale yapay yüksek frekanslı bileşenler eklediğinde ortaya çıkan spektral sızıntıyı azaltmaktadır.
- **Hızlı fourier dönüşümü (FFT):** Fourier dönüşümü, zaman alanı sinyalini frekans alanına dönüştürmek için her kareye uygulanmaktadır. Bu durum, sinyalin frekans bileşenlerinin bir temsilini sağlamaktadır.
- **Mel-frekans filtreleme:** FFT'den elde edilen güç spektrumu, Mel ölçeğine göre aralıklandırılmış bir üçgen filtre bankası kullanılarak filtrelenmektedir. Mel ölçeği, insan işitme sisteminin doğrusal olmayan frekans tepkisini taklit eder ve algısal olarak daha alakalı olan frekansları vurgulamaktadır.
- **Logaritma:** Her bir Mel filtre bankasındaki enerji logaritmik bir ölçeğe dönüştürülür. Bu sıkıştırma, sinyalin dinamik aralığını normalleştirmeye ve insan işitme sisteminin ses yüksekliğine verdiği tepkiyi daha iyi temsil etmeye yardımcı olmaktadır.
- **Ayrık kosinüs dönüşümü (DCT):** Ayrık kosinüs dönüşümü log-Mel filtre bankası enerjilerine uygulanmaktadır. Bu işlem, filtre bankası enerjilerinin ilişkisini bozarak fazlalığı azaltmakta ve en önemli spektral özellikleri yakalamaktadır. Elde edilen katsayılar Mel-Frekans Cepstral Katsayıları olarak bilinmektedir.

MFCC'lerin Kullanımı

- **İnsan benzeri algı:** MFCC hesaplama sürecinde uygulanan Mel frekansı çarpıtma ve logaritmik sıkıştırma, insan işitme sisteminin sese verdiği tepkiyi büyük ölçüde taklit edebilmekte ve bu durum, onları insan algısıyla bağlantılı ses özelliklerini temsil etmede son derece etkili kılmaktadır.
- **Boyut azaltma:** MFCC'ler, en önemli spektral bilgileri korurken orijinal ses sinyalinin boyutluluğunu önemli ölçüde azaltmaktadır. Bu boyutsallık azaltımı, sonraki işlemleri basitleştirerek hesaplama verimliliğini artırmaktadır.
- **Dayanıklılık:** MFCC'lerin çeşitli gürültü ve bozulma türlerine karşı dayanıklı olduğu kanıtlanmıştır. Bu da onları ses sinyallerinin arka plan gürültüsü veya diğer parazitlerle bozulabileceği gerçek dünya uygulamaları için uygun hale getirmektedir.

MFCC'nin Uygulamaları

- **Konuşma tanıma:** MFCC'ler, otomatik konuşma tanıma (ASR) sistemlerinde birincil özellikler olarak yaygın şekilde kullanılmaktadır. Ayrıca konuşma dilinin doğru bir şekilde yazıya dökülmesini sağlamaktadır.
- **Konuşmacı tanımlama:** MFCC'ler, ses yolu şekli ve konuşma tarzı gibi seslerinin benzersiz özelliklerine dayalı olarak bireysel konuşmacıları tanımlamak için kullanılabilir.
- **Müzik bilgisi alma:** MFCC'ler müzik türü sınıflandırması, enstrüman tanımlama ve diğer müzik bilgisi alma görevlerinde kullanılmaktadır.
- **Çevresel ses sınıflandırması:** MFCC'ler kentsel sesler (trafik, sirenler), doğal sesler (kuş sesleri, hışırdayan yapraklar) ve endüstriyel sesler gibi çevresel seslerin sınıflandırılmasında etkinlik göstermektedir.

Özetle, MFCC'ler ses sinyallerinin güçlü ve sağlam bir temsilini sağlayarak çeşitli uygulamalarda etkili analiz ve sınıflandırmaya olanak tanımaktadır. Sesin spektral özelliklerini insan işitsel algısıyla uyumlu bir şekilde yakalama yetenekleri, onları birçok ses işleme sisteminin temel taşı haline getirmektedir.

3.2.2. Spektral Kontrast

Spektral kontrast, ses spektrumundaki belirgin frekans bantlarının (tepe noktaları) enerjileri ile komşu vadileri (düşük enerji bölgeleri) arasındaki farkı ölçen bir ses özelliğidir. Öncelikle genel spektral şekle odaklanan MFCC'lerin aksine, spektral kontrast farklı frekans bantları içindeki göreceli enerji dağılımını vurgulamaktadır [86].

Spektral Kontrastın Temel Yönleri

- **Tepe-Vadi analizi:** Spektral kontrast, belirli frekans bantları içindeki spektral tepe ve vadilerin enerji seviyeleri karşılaştırılarak hesaplanmaktadır.
- **Frekans bandı bölme:** Ses spektrumu tipik olarak, insan işitsel algısını daha iyi yansıtmak için genellikle logaritmik aralıklarla bir dizi frekans bandına bölünmektedir.
- **Kontrast hesaplama:** Her bir frekans bandı için tepe frekansının enerjisi ve komşu vadi frekanslarının enerjisi hesaplanmaktadır. Bu enerjiler arasındaki fark (genellikle desibel cinsinden ifade edilir) o bant için spektral kontrastı temsil etmektedir.

Spektral Kontrast Kullanımı

- **Harmonik ve vurmali seslerin ayırt edilmesi:** Spektral kontrast, özellikle aşağıdaki gibi farklı spektral özelliklere sahip sesleri ayırt etmede etkili olmaktadır:
 - **Harmonik sesler:** Müzik aletleri (örn. piyano, keman vb.) veya insan konuşması gibi güçlü harmonik içeriğe sahip sesler, iyi tanımlanmış tepe ve vadilere sahip nispeten pürüzsüz spektrumlar sergileme eğilimindedir.
 - **Vurmali sesler:** Davul, zil veya patlama gibi geçici başlangıçlara sahip sesler, genellikle keskin tepe noktaları ve enerji dağılımındaki hızlı değişikliklerle daha düzensiz ve dinamik spektrumlar sergilemektedir. Spektral kontrast, enerjideki bu hızlı dalgalanmaları etkili bir şekilde yakalamaktadır.
- **Spektral dinamikleri yakalama:** Öncelikle ortalama spektral zarfa odaklanan MFCC'lerin aksine, spektral kontrast keskin tepe ve vadilerin varlığı gibi spektrumun dinamiklerini yakalamaktadır. Bu bilgi aşağıdaki gibi görevler için oldukça önemli olabilmektedir:
 - **Müzik türü sınıflandırması:** Klasik müzik (armonikler açısından zengin) ve rock müzik (vurmali unsurlarla karakterize edilen) gibi farklı tınsal özelliklere sahip türler arasında ayırım yapmak için kullanılabilir.
 - **Çevresel ses sınıflandırması:** Kuş sesleri, trafik gürültüsü ve endüstriyel sesler gibi farklı spektral özelliklere sahip sesler arasında ayırım yapmak için kullanılabilir.
- **Ses kaynağı ayırma:** Spektral kontrast, bir karışımdaki farklı ses kaynaklarını ayırmaya yardımcı olmak için kullanılabilir. Örneğin, konuşmacının sesiyle ilişkili spektral tepe ve vadileri tanımlayarak konuşmayı arka plan gürültüsünden ayırt etmek için kullanılabilir.

Sonuç olarak spektral kontrast, ses sinyallerinin dinamik ve incelikli spektral özelliklerini yakalayıp MFCC'ler gibi özelliklere değerli bir tamamlayıcı sağlamaktadır. Farklı frekans bantlarındaki göreceli enerji dağılımına odaklanarak, seslerin tınısı ve karakteri hakkında içgörüler sunarak daha doğru ve sağlam ses analizi ve sınıflandırması sağlamaktadır.

Spektral kontrast uygulamaları

Spektral kontrast, özellikle müzik ve ses işleme alanlarında, çeşitli uygulamalarda önemli bir rol oynamaktadır. İşte bu bölümde spektral kontrastın kullanıldığı başlıca alanlardan bahsedilmektedir.

- **Müzik sınıflandırma:** Burada müzik sınıflandırma ile ilgili uygulamalar açıklanmaktadır.
 - **Janr ayrımı:** Farklı müzik türlerinin belirgin spektral karakteristikleri vardır. Örneğin, klasik müzik genellikle zengin harmonik içeriğe sahipken, rock müzik perküsyif seslerle karakterize edilmektedir. Spektral kontrast, bu farklılıkları belirleyerek müzik türlerinin sınıflandırılmasında kullanılmaktadır.
 - **Enstrüman tanıma:** Farklı müzik aletlerinin ürettiği seslerin spektral kontrastları farklıdır. Örneğin, bir piyano ve bir gitarın aynı notayı çalması durumunda bile, farklı harmonik içerikleri sayesinde birbirlerinden ayırt edilebilmektedirler. Spektral kontrast, bu farklılıkları analiz ederek enstrüman tanıma uygulamalarında kullanılmaktadır.
- **Ses kaynak ayrımı:** Bu bölümde ses ve kaynağıyla ilgili uygulamalar açıklanmaktadır.
 - **Konuşma ve gürültü ayrımı:** Karışık bir ses ortamında konuşma sesini arka plan gürültüsünden ayırmak için spektral kontrast kullanılabilir. Konuşma genellikle daha düzenli bir spektral yapıya sahipken, arka plan gürültüsü daha rastgele ve değişken bir spektral yapıya sahiptir. Spektral kontrast, bu farklılıkları belirleyerek konuşma sinyallerinin daha net bir şekilde elde edilmesini sağlamaktadır.
 - **Çoklu konuşmacı ayrımı:** Birden fazla kişinin konuştuğu bir ortamda, her bir konuşmacının sesini diğerlerinden ayırmak için spektral kontrast kullanılabilir. Farklı konuşmacıların seslerinin farklı spektral karakteristikleri olduğu için spektral kontrast bu farklılıkları kullanarak konuşmacıları ayırt etmeye yardımcı olmaktadır.
- **Diğer uygulamalar:** Literatürde kullanılan diğer uygulamalar burada açıklanmaktadır.
 - **Ses etkisi tanıma ile ilgili uygulamalar:** Patlama, çarpma gibi ani ses olaylarının tespiti için spektral kontrast kullanılabilir.
 - **Ortam ses sınıflandırması ile ilgili uygulamalar:** Şehir gürültüsü, doğa sesleri gibi farklı çevresel seslerin sınıflandırılmasında spektral kontrast önemli bir rol oynamaktadır.
 - **Biyolojik ses analizi ile ilgili uygulamalar:** Hayvan sesleri, kalp sesleri gibi biyolojik seslerin analizinde spektral kontrast kullanılarak farklı türler veya hastalıklar tespit edilebilmektedir.

Özetle, spektral kontrast, ses sinyallerinin spektral yapısındaki farklılıkları analiz ederek, çeşitli ses işleme uygulamalarında önemli bir rol oynamaktadır. Özellikle harmonik ve perküsyif

seslerin ayrımı, farklı ses kaynaklarının ayrılması ve ses sınıflandırma gibi görevlerde etkin bir şekilde kullanılmaktadır.

3.2.3. Sıfır Geçiş Oranı (ZCR)

Belirli bir zaman dilimi içinde bir ses sinyalinin Sıfır Geçiş Oranı (ZCR), sinyalin o zaman aralığında sıfır voltaj seviyesini kaç kez geçtiğini (pozitiften negatife veya negatiften pozitif) belirtmektedir [87].

ZCR Temel Yönleri

- **Basit ve verimlilik:** ZCR, hesaplama açısından ucuz bir özelliktir ve gerçek zamanlı uygulamalar için uygundur.
- **Ayırt edici güç:** ZCR, ses sinyalinin özellikleri hakkında değerli bilgiler sağlamaktadır:
 - **Sesli ve sessiz sesler:** Sesli harfler gibi sesli sesler, dalga formlarının periyodik yapısı nedeniyle tipik olarak daha düşük ZCR değerleri sergilemektedir. Sürtünmeli sesler (örn. “s”, “f”) ve pesifler (örn. “p”, “t”) gibi sessiz sesler, daha düzensiz ve geçici yapıları nedeniyle daha yüksek ZCR değerlerine sahip olma eğilimindedir.
 - **Vurmalı sesler:** Davul ve zil gibi vurmalı sesler, ani başlangıçlar ve genlikteki hızlı değişimlerle karakterize edilmektedir. Bu ani değişiklikler genellikle yüksek ZCR değerlerine neden olmaktadır.

ZCR Kullanımı

ZCR kullanımı ikiye ayrılmaktadır. Bunlar sesli/sessiz ayrımı ve vurmalı ses algılama olarak ikiye ayrılmaktadır.

- **Sesli / sessiz ses ayrımı:** Her dilin kendine özgü sesli ve sessiz harfleri vardır. Bunların ayırt edilmesi problemi de bazı araştırmaların çözümünde yardımcı olmuştur.
 - **Sesli sesler:** Sesli harfler gibi sesli sesler, ses tellerinin titreşimiyle üretilmektedir. Bu periyodik titreşim, daha düşük sıfır geçiş oranına sahip nispeten pürüzsüz ve tutarlı bir dalga formu ile sonuçlanmaktadır.
 - **Sessiz sesler:** Sürtünmeli (/s/, /f/, /th/) ve patlamalı (/p/, /t/, /k/) sesler gibi tonsuz sesler, ses yolundaki daralmalar yoluyla türbülanslı hava akışı tarafından üretilmektedir. Bu türbülanslı hava akışı, daha fazla sayıda sıfır geçişi ile düzensiz ve hızla değişen dalga formları oluşturmaktadır.

Algoritmalar ZCR'yi analiz ederek, konuşma tanıma ve diğer konuşma işleme görevleri için çok önemli olan sesli ve sessiz sesleri etkili bir şekilde ayırt edebilmektedir.

- **Vurmalı ses algılama:**

- **Davul vuruşları:** Zil sesleri veya silah sesleri gibi vurmalı sesler, ani başlangıçlar ve genlikteki hızlı değişikliklerle karakterize edilmektedir. Dalga biçimindeki bu ani değişiklikler çok sayıda sıfır geçişine neden olmaktadır.
- **Algoritmalar:** ZCR'yi analiz ederek vurmalı sesleri etkili bir şekilde tespit edip sınıflandırabilmekte ve bunları müzik notaları veya sürekli konuşma gibi sürekli seslerden ayırt edebilmektedir. Bu bilgi, müzik bilgisi alma (örn. ritim algılama, tür sınıflandırma), ses olayı algılama (örn. patlamaları veya cam kırılmasını algılama) ve ses gözetimi gibi uygulamalarda değerlidir.

Uygulamalar

- **Konuşma işleme:**

- **Sesli / sessiz ses ayrımı:** ZCR, sesli ve sessiz sesleri ayırt etmek için konuşma işlemede çok önemli bir özelliktir. Bu bilgi konuşma tanıma, ses sentezi ve konuşmayla ilgili diğer görevler için gereklidir.
- **Konuşmacı tanımlama:** ZCR, vokal kıvrım titreşimlerinin hızı gibi vokal özelliklerine göre konuşmacılar arasında ayırım yapmak için kullanılabilir.

- **Müzik bilgisi alma:**

- **Vurmalı ses algılama:** ZCR, müzikteki davul vuruşları ve zil çarpmaları gibi vurmalı sesleri tespit etmede oldukça etkili olmaktadır. Bu bilgiler vuruş takibi, tür sınıflandırması ve müzik bilgisi alma gibi çeşitli müzik analizi görevleri için kullanılabilir.

- **Duygu algılama:** ZCR, konuşmacının duygusal durumu hakkında içgörü sağlayabilmektedir. Örneğin, heyecanlı veya tedirgin konuşma, artan vokal yoğunluğu ve genlikteki hızlı değişiklikler nedeniyle daha yüksek ZCR değerleri sergileyebilmektedir.

- **Çevresel ses sınıflandırması:** ZCR, yaprak hışırtısı (düşük ZCR) ve trafik sesi (yüksek ZCR) gibi farklı çevresel ses türlerini ayırt etmek için kullanılabilir.

3.2.4. Kroma Özellikleri

Kroma özellikleri, bir ses sinyalinin harmonik içeriğinin güçlü bir temsili olarak göze çarpmaktadır. Her bir boyutun Batı müziğindeki 12 perde sınıfından birine (C, C#, D, D#, E, F, F#, G, G#, A, A#, B) karşılık geldiği 12 boyutlu bir vektör sağlamaktadırlar. Esasen kroma özellikleri, mutlak frekanslarına bakılmaksızın, bu 12 perde sınıfı boyunca enerjinin dağılımını temsil etmektedir [88].

Kroma Özelliklerinin Hesaplanması

- **Frekans alanı gösterimi:** Ses sinyali ilk olarak FFT gibi teknikler kullanılarak frekans alanına dönüştürülmektedir.
- **Perde sınıfı eşleme:** Frekans spektrumu daha sonra 12 perde sınıfıyla eşleştirilmektedir. Örneğin, C notasına (ve oktavlarına) karşılık gelen tüm frekanslar, C perde sınıfındaki enerjiyi temsil etmek için toplanmaktadır.
- **Normalleştirme:** Elde edilen 12 boyutlu vektör, genel sinyal enerjisindeki varyasyonları hesaba katmak için genellikle normalleştirilmektedir (örneğin ögeler 1'e toplanarak yapılmaktadır).

Kroma Özelliklerinin Kullanımı

- **Tını değişimlerine karşı dayanıklılık:** Kroma özellikleri, tınıdaki (bir ses kaynağının benzersiz kalitesi) değişikliklere karşı nispeten sağlamdır. Aynı notayı çalan iki enstrüman farklı tınlara sahip olmaktadır, ancak her ikisi de aynı perde sınıfını paylaştığı için kroma vektörleri benzer olabilmektedir. Bu da MFCC gibi tınıya karşı daha hassas olan özelliklere kıyasla enstrüman türü veya çalma stilindeki değişikliklere karşı daha dayanıklı olmalarını sağlamaktadır.
- **Transpozisyona karşı değişmezlik:** Kroma özellikleri transpozisyona karşı değişmezdir, yani bir melodiyi bir oktav yukarı veya aşağı kaydırmak kroma vektörünü değiştirmemektedir. Bu özellik, anahtar algılama ve melodi çıkarma gibi birçok müzik bilgisi alma görevi için oldukça önemlidir.
- **Verimli temsil:** Kroma özellikleri, bir sinyalin harmonik içeriğinin kompakt ve verimli bir temsilini sağlamaktadır. Bu da onları gerçek zamanlı uygulamalar ve hesaplama açısından zorlu görevler için uygun hale getirmektedir.

Kroma'nın Uygulamaları

- **Müzik türü sınıflandırması:** Kroma özellikleri müzik türlerini sınıflandırmada oldukça etkilidir. Farklı türler genellikle farklı harmonik modeller sergilemektedir. Örneğin, rock müzik tipik olarak belirli perde sınıfları üzerinde güçlü bir vurguya sahipken, klasik müzik perde sınıfları arasında daha eşit dağılmış bir enerjiye sahip olabilmektedir.
- **Ses parmak izi:** Kroma özellikleri, birçok ses parmak izi algoritmasının temel bir bileşenidir. Bu algoritmalar, kroma içeriklerine dayalı olarak ses kayıtları için benzersiz "parmak izleri" oluşturmaktadır. Bu parmak izleri daha sonra büyük veri tabanlarındaki şarkıları hızlı ve doğru bir şekilde tanımlamak ve eşleştirmek için kullanılabilir.

- **Akor tanıma:** Enerjinin 12 perde sınıfındaki dağılımını analiz ederek, kroma özellikleri bir müzik parçasında bulunan akorları tanımlamak için kullanılabilir. Bu bilgi müzik analizi, transkripsiyon ve çeşitli müzik bilgisi alma görevleri için değerlidir.
- **Melodi çıkarma:** Kroma özellikleri, bir müzik parçasından melodik çizgiyi çıkarmak için kullanılabilir. En güçlü kroma bileşenlerinin zaman içindeki gelişimini izleyerek, melodiyi oluşturan notaları tanımlamak mümkündür.

Özetle, kroma özellikleri ses sinyallerinin harmonik içeriğinin güçlü ve sağlam bir temsiliyi sağlamaktadır. Transpozisyona karşı değişmezlikleri ve tını değişimlerine karşı dayanıklılıkları, onları çok çeşitli müzik bilgisi alma görevleri için değerli bir araç haline getirmektedir.

3.2.5. Spektral Merkez

Spektral merkez, bir ses sinyalinin frekans spektrumundaki enerji dağılımının “kütle merkezini” temsil eden değerli bir özelliktir. Sinyalin enerjisinin yoğunluğunun nerede yoğunlaştığını göstermektedir [89].

Spektral Merkezin Hesaplanması

Literatürde spektral centroid olarak da bilinen spektral merkez, sinyalde bulunan frekansların ağırlıklı ortalaması olarak hesaplanmaktadır; burada ağırlıklar, karşılık gelen frekans bileşenlerinin büyüklükleridir.

Yüksek bir spektral merkez değeri, sinyalin enerjisinin daha yüksek frekanslarda yoğunlaştığını belirtmekte ve bu durum, “daha parlak” veya “daha keskin” olarak algılanan bir ses üretmektedir. Düşük bir spektral merkez değeri ise sinyalin enerjisinin daha düşük frekanslarda yoğunlaştığını göstermekte ve bu durum da “daha koyu” veya “daha derin” olarak nitelendirilen bir sesle sonuçlanmaktadır.

Spektral Merkez Kullanımı

- **Tını analizi:** Spektral merkez, farklı enstrümanların ve seslerin tınısını (bir sesin benzersiz kalitesini) tanımlamak için kritik bir özelliktir. Yüksek spektral merkez değerine sahip enstrümanlar (örn. ziller ve trompetler), daha parlak ve keskin bir ses üretme eğilimi göstermektedir. Düşük spektral merkez değerine sahip enstrümanlar (örn. bas davullar ve cellolar) ise daha koyu ve derin bir ses üretme eğilimindedir.
- **Perküsif ve tonal sesler:** Spektral merkez, vurmali ve tonal sesler arasında ayırım yapılmasına yardımcı olmaktadır. Davul ve zil gibi vurmali sesler, geçici yapıları ve yüksek frekanslı bileşenlerin varlığı nedeniyle genellikle daha yüksek spektral merkezlere sahiptir. Enstrümanlardan veya insan konuşmasından gelen sürekli notalar gibi tonal sesler, düşük

frekanslı harmoniklerin baskınlığı nedeniyle genellikle daha düşük spektral merkezlere sahiptir.

Spektral Merkez Uygulamaları

- **Müzik tını sınıflandırması:** Spektral merkez, farklı enstrümanları sınıflandırmak, müzikteki vurmali unsurların varlığını belirlemek ve farklı müzik türlerini ayırt etmek için önemli bir özelliktir.
- **Konuşma analizi:**
 - Konuşma işlemede, spektral merkez, konuşmacıların seslerinin parlaklığı veya koyuluğu gibi vokal özelliklerini analiz etmek için kullanılabilir.
 - Konuşma sinyalinin spektral dağılımını etkileyebileceğinden, vokal çaba veya duygusal durumdaki değişiklikleri tespit etmek için de kullanılabilir.
- **Çevresel ses sınıflandırması:** Spektral merkez, kuş seslerinin yüksek frekanslı bileşenleri ile trafik gürültüsünün düşük frekanslı bileşenleri gibi farklı çevresel ses türlerini ayırt etmeye yardımcı olabilir.

Özetle spektral merkez, bir ses sinyalinin spektral dağılımının önemli bir ölçüsünü temsil etmektedir. Sinyalin genel parlaklığını yakalayarak, tını analizi, ses kaynağı tanımlama ve diğer ses işleme uygulamalarına önemli bir katkıda bulunmaktadır.

3.2.6. Spektral Bant Genişliği

Spektral bant genişliği, bir ses sinyalinin spektral merkezi etrafındaki frekansların yayılımının bir ölçüsüdür. Sinyalin enerjisinin frekans spektrumunda ne kadar geniş bir alana dağıldığını ölçmektedir [90].

- **Yüksek spektral bant genişliği:** Yüksek spektral bant genişliği, sinyalin enerjisinin geniş bir frekans aralığına dağıldığını göstermektedir. Bu durum, geniş bantlı gürültü, vurmali sesler veya zengin harmonik içeriğe sahip sesler gibi geniş bir spektruma sahip seslerin karakteristiğidir.
- **Düşük spektral bant genişliği:** Düşük spektral bant genişliği, sinyalin enerjisinin dar bir frekans aralığında yoğunlaştığını göstermektedir. Bu durum, saf tonlar (örn. sinüs dalgaları) veya sınırlı harmonik içeriğe sahip sesler gibi dar bantlı seslerin karakteristik özelliğidir.

Bant Genişliğinin Hesaplanması

Spektral bant genişliği genellikle spektral merkez etrafındaki frekans dağılımının standart sapması olarak hesaplanmaktadır.

Spektral Bant Genişliği Kullanımı

- **Geniş bantlı ve dar bantlı sesleri ayırt etme:** Spektral bant genişliği, geniş bantlı sesler (örn. beyaz gürültü, trafik gürültüsü) ile dar bantlı sesleri (örn. saf tonlar, tek frekanslı sinyaller) birbirinden ayırabilmek için oldukça önemli bir özelliktir.
- **Tını karakterizasyonu:** Seslerin tınısı hakkında değerli bilgiler sağlamaktadır. Yüksek spektral bant genişliğine sahip sesler daha parlak, daha karmaşık bir tınıya sahip olma eğilimindeyken, düşük spektral bant genişliğine sahip sesler daha odaklanmış, daha az karmaşık bir tınıya sahip olma eğilimindedir.
- **Gürültü özelliklerini analiz etme:** Çevresel gürültü analizinde, spektral bant genişliği, gürültü kirliliğinin doğasını tanımlamak amacıyla kullanılmaktadır. Trafik ve endüstriyel makineler gibi geniş bantlı gürültü kaynakları, tek bir makinenin uğultusu gibi dar bantlı gürültü kaynaklarına göre daha yüksek spektral bant genişliği değerleri göstermektedir.

Spektral Bant Genişliği Uygulamaları

- **Enstrüman sınıflandırması:** Farklı tınısal özelliklere sahip enstrümanlar arasında ayırım yapma olarak bilinmektedir. Örneğin, yüksek spektral bant genişliğine sahip enstrümanlar (örn. ziller, trampetler), daha düşük spektral bant genişliğine sahip enstrümanlara (örn. flütler, fagotlar) kıyasla daha geniş bir frekans aralığına sahip sesler üretmektedir.
- **Ses kaynağı tanımlama:** Endüstriyel uygulamalarda gürültü kaynağının belirlenmesi olarak bilinmektedir. Örneğin, yüksek spektral bant genişliği dönen parçalara sahip makinelerin varlığına işaret edebilirken, düşük bant genişliği tek ve baskın bir gürültü kaynağının varlığına işaret edebilmektedir.
- **Çevresel gürültü izleme:** Farklı gürültü kaynaklarının çevre üzerindeki etkisinin değerlendirilmesi gerçekleştirilebilmektedir. Geniş bantlı gürültü kirliliği kaynaklarının belirlenmesi ve azaltılması uygulamaları gerçekleştirilebilmektedir.

Özetle, spektral bant genişliği bir ses sinyalinin frekans spektrumundaki enerji yayılımı hakkında değerli bilgiler sağlamaktadır. Bu özellik, seslerin tınısını karakterize etmek, farklı ses türleri arasında ayırım yapmak ve çeşitli akustik ortamların özelliklerinin analizi, farklı disiplinlerdeki uygulamalar için de temel teşkil etmektedir.

3.2.7. Spektral Düşüş Eğimi

Spektral rolloff olarak da bilinmektedir. Bir ses sinyalinin toplam spektral enerjisinin önceden tanımlanmış bir yüzdesinin (genel olarak %85) altında yoğunlaştığı frekans eşikini belirleyen bir akustik parametredir. Bu parametre, frekans spektrumunda sinyal enerjisinin baskın olduğu bölgeyi nicel olarak ifade etmektedir [91].

Spektral Düşüş Eğiminin Temel Yönleri

Spektral düşüş eğimi, enerjinin frekans spektrumundaki dağılımı hakkında değerli bilgiler sağlamaktadır.

- **Tonal ve gürültülü sesler:** Tonal sesler, temel frekans ve harmonik yapısal bütünlükleri nedeniyle, spektral enerjilerinin büyük bir kısmını dar bir frekans bandında yoğunlaştırır, bu da düşük spektral düşüş eğimi değerlerine yol açar. Buna karşılık, geniş bantlı gürültülü sesler, spektral enerjilerini geniş bir frekans aralığına yayarak yüksek spektral düşüş eğimi değerleri sergilemektedirler.

Spektral Düşüş Eğimi Kullanımı

- **Tonal ve gürültülü sesleri ayırt etme:** Spektral rolloff, tonal ve gürültülü sesleri ayırt etmede oldukça etkilidir. Konuşmayı tanımlamak ve arka plan gürültüsünden ayırmak, müzik aletlerini izole etmek ve çevresel kayıtlarda gürültünün varlığını tespit etmek için kullanılabilir.
- **Tını karakterizasyonu:** Spektral düşüş eğimi, ses sinyallerinin spektral içeriklerinin frekans alanındaki dağılımını nicel olarak ifade eden bir parametredir. Bu parametre, sesin algılanan tınısal nitelikleri ile doğrudan ilişkilidir. Yüksek spektral düşüş eğimi değerleri, spektral enerjinin yüksek frekans bileşenlerinde yoğunlaştığını ve dolayısıyla sesin algılanan tınısının “parlak” veya “gürültülü” olduğunu göstermektedir. Buna karşılık, düşük spektral düşüş eğimi değerleri, spektral enerjinin düşük frekans bileşenlerinde yoğunlaştığını ve dolayısıyla sesin algılanan tınısının “odaklanmış” veya “tonal” olduğunu göstermektedir.

Spektral Düşüş Eğimi Uygulamaları

- **Konuşma iyileştirme:** Konuşma sinyallerindeki arka plan gürültüsünü tanımlamak ve filtrelemek için kullanılabilir. Algoritmalar, gelen sesin spektral düşüş eğimini analiz ederek konuşma ve arka plan gürültüsü arasında etkili bir ayırım yapabilmektedir.
- **Müzik türü sınıflandırması:** Müzik türlerinin spektral enerji dağılımlarındaki farklılıkları nicel olarak ifade ederek, türler arasında ayırım yapılmasına olanak tanımaktadır. Rock müzik gibi yüksek frekans bileşenlerinin baskın olduğu türler, yüksek spektral düşüş eğimi değerleri sergilerken, klasik müzik gibi daha dengeli bir spektral yapıya sahip türler, düşük spektral düşüş eğimi değerleri sergilemektedir. Bu parametre, müzik sınıflandırma ve bilgi erişimi gibi uygulamalarda kullanılmaktadır.
- **Enstrüman sınıflandırması:** Enstrümanların spektral enerji dağılımlarındaki farklılıkları nicel olarak ifade ederek, enstrüman sınıflandırma ve tanıma gibi uygulamalarda kullanılmaktadır. Geniş spektrumlu enstrümanlar, yüksek spektral düşüş eğimi değerleri

sergilerken; dar spektrumlu enstrümanlar, düşük spektral düşüş eğimi değerleri sergilemektedir. Bu parametre, enstrümanların tınsal karakteristiklerini analiz ederek enstrümanlar arasında ayırım yapılmasına olanak tanımaktadır.

Özetle spektral düşüş eğimi, ses sinyallerinin frekans spektrumundaki enerji dağılımını nicel olarak ifade eden bir akustik parametredir. Bu parametre, sinyal enerjisinin baskın olduğu frekans aralığını belirleyerek, sesin tınsal nitelikleri hakkında değerli bilgiler sunmaktadır. Spektral düşüş eğimi, çeşitli ses analizi ve sınıflandırma uygulamalarında, özellikle müzik bilgisi erişimi, konuşma tanıma ve çevresel ses analizi gibi alanlarda kullanılmaktadır.

3.2.8. Değerlendirme

Özellik çıkarma, ses analizinde çok yönlü ve güçlü bir araç olarak kabul edilmektedir. Uygun özelliklerin seçilip çıkarılmasıyla, makine öğrenimi modellerinin ses verilerinden değerli bilgiler edinmesi ve çeşitli uygulamalarda başarılı olması sağlanmaktadır. Sesten özelliklerin çıkarılması ile ilgili yöntemler Tablo 3.2’de özetlenmiştir.

Tablo 3.2. Sesten özelliklerin çıkarılması ile ilgili yöntemler

Özellik	Tanım	Temel Kullanım Örnekleri
MFCC’ler	Spektral özellikleri yakalayarak insan işitme algısını modellemektedir.	Konuşma tanıma, müzik sınıflandırması, çevresel ses analizinde kullanılmaktadır.
Spektral Kontrast	Spektral tepeler ve vadiler arasındaki farkı ölçmektedir.	Müzik sınıflandırması, ses ayırımında kullanılmaktadır.
Sıfır Geçiş Oranı (ZCR)	Bir sinyaldeki işaret değişim oranını ölçmektedir.	Konuşma-müzik ayırımı, duygu tanıma, vurmali ses algılama ayırımında kullanılmaktadır.
Kroma Özellikleri	Müzikal perde sınıflarına dayalı armonik içeriği temsil etmektedir.	Müzik analizi, akor tanıma, şarkı tanımlamada kullanılmaktadır.
Spektral Ağırlık Merkezi	Frekans spektrumunun “kütle merkezini” ölçmektedir.	Tını sınıflandırması, konuşma analizinde kullanılmaktadır.
Spektral Bant Genişliği	Spektral merkez etrafındaki frekans yayılımını ölçmektedir.	Çevresel gürültü analizi, makine sesi sınıflandırmasında kullanılmaktadır.
Spektral Düşüş Eğimi	Spektral enerjinin çoğunun hangi frekansın altında bulunduğunu ölçmektedir.	Konuşma filtreleme, müzik ve gürültü ayırımında kullanılmaktadır.

Özellik çıkarma tekniklerinde devam eden ilerlemeler, derin öğrenmenin gücüyle birleştiğinde, ses analizinde inovasyonu teşvik etmeye devam edecek ve çeşitli alanlarda çok çeşitli heyecan verici uygulamalara olanak sağlayacaktır.

3.3. Duygu ve Şiddet Özellikleri

Ses yalnızca bir iletişim aracı değildir; aynı zamanda duyguları ifade etmek için de güçlü bir araçtır. Bir sesin nüansları dikkatlice analiz edilerek konuşmacının duygusal durumu hakkında içgörü kazanılabilmektedir. Sesin duygusal içeriğini anlamamıza yardımcı olan akustik özellikleri daha derinlemesine incelemek gerekmektedir [92].

3.3.1. Duygu ve Şiddet Özelliklerinin Temel Yönleri

- **Perde ve duygusal korelasyon:** Genel olarak, daha yüksek perde heyecan, öfke veya şaşkınlık gibi duygularla ilişkilendirilirken, daha düşük perde üzüntü veya sakinlik gibi duygularla bağlantılı olmaktadır. Ancak bu ilişki her zaman açık değildir ve kültürler arasında farklılık gösterebilmektedir.
- **Ses yüksekliği ve duygusal korelasyon:** Daha yüksek ses yüksekliği genellikle heyecan, öfke veya korku gibi güçlü duygularla ilişkilendirilirken, daha düşük ses yüksekliği üzüntü veya sakinlik gibi duygularla ilişkilendirilmektedir. Ancak, ses yüksekliği ve duygu arasındaki ilişki her zaman doğrusal değildir. Örneğin, fısıltıyla söylenen bir tehdit, bağırarak söylenen bir teşekkürden daha korkutucu olabilmektedir.
- **Konuşma hızı ve duygusal korelasyon:** Daha yüksek konuşma hızı genellikle heyecan, gerginlik veya endişe ile ilişkilendirilirken, daha yavaş konuşma hızı üzüntü, yorgunluk veya düşüncelilik belirtisi olabilmektedir.
- **Ses tınısı ve duygusal korelasyon:** Farklı duygular farklı tınlar üretebilmektedir. Örneğin, mutluluk genellikle daha parlak, daha açık bir tını ile ilişkilendirilirken, üzüntü daha koyu, daha boğuk bir tını ile ilişkilendirilebilmektedir.
- **Spektral enerji ve duygusal korelasyon:** Spektral enerjinin dağılımı farklı duygusal durumlara göre değişebilmektedir. Örneğin, heyecanlı konuşma daha yüksek frekans aralığında daha fazla enerjiye sahip olabilirken, üzgün konuşma daha düşük frekans aralığında daha fazla enerjiye sahip olabilmektedir.
- **Titreşim ve pırıltı:** Bunlar vokal kıvrım titreşimlerinin stabilitesini ölçer ve duygusal durum hakkında ipuçları sağlayabilmektedir.
- **Formant frekansları:** Bunlar ses yolunun rezonans frekanslarıdır ve farklı sesli ve sessiz harfleri ayırt etmeye yardımcı olmanın yanı sıra ses yolu boyutu ve şekli hakkında bilgi sağlayabilmektedir.

Uygulamalar

- **Sesli asistanlar:** Daha kişiselleştirilmiş yanıtlar sağlamak için kullanıcı duygularını anlamak.

- **Oyun:** Daha sürükleyici ve duygusal olarak ilgi çekici oyun deneyimleri yaratmak.
- **Pazarlama:** Reklamlara yanıt olarak müşteri duygularını analiz etmek.
- **Akıl sağlığı:** Ruh sağlığı bozukluklarının erken teşhisi için duygusal durumların izlenmesini sağlamak.

Gelecekteki Yönelimler

- **Multimodel analiz:** Daha doğru duygu tanıma için akustik özelliklerin görsel ve fizyolojik verilerle birleştirilmesi sağlanabilecektir.
- **Derin öğrenme:** Ses sinyallerinden daha karmaşık ve soyut özellikler çıkarmak için derin sinir ağlarından yararlanmak.
- **Kültürel hususlar:** Duygusal ifadedeki kültürel farklılıklara duyarlı modeller geliştirmek. Akustik özellikler ve duygular arasındaki ilişkiyi anlayarak, insan-bilgisayar etkileşimi ve duygusal zekâ için daha sofistike ve etkili araçlar geliştirilebilecektir.

3.3.2. Konuşma Seslerinde Duygusal Korelasyon

Bu bölümde konuşma seslerindeki duygusal korelasyon açıklanacaktır [93].

Öfke durumu

- **Perde:** Vokal frekansta önemli bir artışa neden olmaktadır. Bu artış, genellikle titreme eşiğine yaklaşan yüksek frekans değerleriyle kendini göstermektedir. Duygunun yoğun ve uçucu doğasını yansıtan, perdede hızlı ve öngörülemeyen değişimler gözlenmektedir.
- **Ses şiddeti:** Tipik olarak yüksektir ve vokal çıktının artan enerjisini ve kuvvetini yansıtmaktadır.
- **Konuşma hızı:** Değişebilir, ancak genellikle öfkeyle ilişkili aciliyeti veya dürtüsellliği yansıtan ani başlangıçlar, duraklamalar ve hızlanmalar ile karakterize edilmektedir.
- **Tını:** Genellikle sert ve gergin bir hal almaktadır. Artan vokal çabası, spektrumda “törpülenmiş” veya “raspy” olarak tanımlanan bir kaliteye neden olmaktadır.
- **Spektral özellikler:** Yüksek frekans bantlarında yoğunlaşmaktadır. Bu durum, artan vokal yoğunluğunu ve ses tellerindeki potansiyel gerginliği yansıtmaktadır.

Örneğin Bağırma, saldırgan ton, aniden ses tonunun yükselmesi vb. gibi akustik değişiklikler, öfke duygusunun vokal ifadesinin karakteristik özellikleridir ve otomatik duygu tanıma sistemlerinde kullanılmaktadır.

Hüzün durumu

- **Perde:** vokal frekansta (perde) belirgin bir düşüşe neden olmaktadır. Perde değişimleri monotonlaşır, duygunun içe dönük ve durağan doğasını yansıtmaktadır.

- **Ses şiddeti:** Vokal genlik (ses şiddeti) azalır, enerjinin düşüşünü ve duygusal çekilmeyi ifade etmektedir.
- **Konuşma hızı:** Konuşma hızı yavaşlar, uzun duraklamalar ve hezeyanlar gözlemlenmektedir. Bu durum, hüznün getirdiği düşünce yavaşlamasını ve motivasyon eksikliğini yansıtmaktadır.
- **Tını:** Vokal tını, donuk ve boğuk bir hal almaktadır. Azalan vokal çabası, spektrumda “soluk” veya “mat” olarak tanımlanan bir kaliteye neden olmaktadır.
- **Spektral özellikler:** Düşük frekans bantlarında artan enerji, vokal enerjisindeki azalmayı ve rezonans değişikliklerini yansıtmaktadır.

Örneğin yavaş, titreşim ses, iç çekmeler vb. gibi akustik değişiklikler, hüznün duygu durumunun vokal ifadesinin karakteristik özellikleridir ve otomatik duygu tanıma sistemlerinde kullanılmaktadır.

Mutluluk ve Heyecan durumları

- **Perde:** Nötr konuşmaya kıyasla belirgin bir artış gözlemlenmektedir. Perde değişimleri, artan uyarılma ve enerjiyi yansıtan daha sık ve geniş aralıklı dalgalanmalar göstermektedir.
- **Ses şiddeti:** Heyecanın yoğunluğuna bağlı olarak değişkenlik göstermektedir. Genlik, artan vokal enerjisini ifade etmekle birlikte, duygusal nüansları yansıtmak amacıyla değişkenlik gösterebilmektedir.
- **Konuşma hızı:** Nötr konuşmaya kıyasla belirgin bir artış göstermektedir. Vokal akıcılık artar, duraklamalar azalır ve konuşma ritmi hızlanmaktadır.
- **Tını:** Parlak ve rezonanslı bir karaktere bürünmektedir. Ses, net ve iyi artiküle edilmiş bir kalite sergilemektedir.
- **Spektral özellikler:** Vokal çabayı ve duygusal yoğunluğu yansıtan, geniş bir frekans aralığında enerji artışı gözlemlenmektedir.

Örneğin gülme, neşeli konuşma, abartılı tonlama vb. gibi akustik değişiklikler, mutluluk ve heyecan duygu durumlarının vokal ifadesinin karakteristik özellikleridir ve otomatik duygu tanıma sistemlerinde kullanılmaktadır.

Korku ve Kaygı durumları

- **Perde:** genellikle yüksek ve dengesiz bir seyir izlemektedir. Hızlı dalgalanmalar ve ani artışlar, duygunun değişken ve kontrolsüz doğasını yansıtmaktadır.
- **Ses şiddeti:** Geniş bir aralıkta değişkenlik göstermektedir. Ani yükselmeler (çığlıklar) ve düşüşler (fısıltılar), duygusal yoğunluğun ani değişimlerini ifade etmektedir.
- **Konuşma hızı:** Hızlı ve düzensiz bir ritim sergilemektedir. Sık duraklamalar ve kesintiler, zihinsel dağınıklığı ve kaygıyı yansıtmaktadır.

- **Tını:** Gergin ve nefes nefese bir karaktere bürünmektedir. Ses tellerindeki artan gerginlik, vokal titremesine (jitter ve shimmer) neden olabilmektedir.
- **Spektral özellikler:** Hızlı ve sık nefes alma nedeniyle seste artan gürültü sergileyebilmektedir.

Örneğin çılgın atma, kısa soluklu sözler, soluk almalar vb. gibi akustik değişiklikler, korku ve kaygı duygu durumlarının vokal ifadesinin karakteristik özellikleridir ve otomatik duygu tanıma sistemlerinde kullanılmaktadır.

Sakinlik ve Nötr durumlar

- **Perde:** Nispeten stabil ve tutarlı bir seyir izlemektedir. Düzgün ve kademeli perde değişimleri, duygunun dengeli ve durağan doğasını yansıtmaktadır.
- **Ses şiddeti:** Önemli dalgalanmalar olmaksızın orta düzeyde ve tutarlı bir seyir izlemektedir. Bu durum, vokal enerjinin dengeli dağılımını ifade etmektedir.
- **Konuşma hızı:** Sabit ve eşit bir ritim sergilemektedir. İyi tanımlanmış duraklamalar ve kelimeler arasında yumuşak geçişler, konuşmanın düzenli ve kontrollü olduğunu göstermektedir.
- **Tını:** Rahat ve doğal bir karaktere bürünmektedir. Ses, net ve iyi artiküle edilmiş bir kalite sergilemektedir.
- **Spektral özellikler:** Frekans spektrumu boyunca dengeli bir enerji dağılımı gözlemlenmektedir. Bu durum, vokal sinyalin spektral stabilitesini ifade etmektedir.

Örneğin rahat konuşma, eşit tempo, dengeli tonlama vb. gibi akustik değişiklikler, sakinlik ve nötr duygu durumlarının vokal ifadesinin karakteristik özellikleridir ve otomatik duygu tanıma sistemlerinde kullanılmaktadır.

Konuşma Seslerinde Duygusal Korelasyon ile İlgili Değerlendirme

- **Bireysel değişkenlik:** Bunlar genel gözlemlerdir ve duyguların ses yoluyla bireysel ifadesi önemli ölçüde değişebilmektedir.
- **Kültürel farklılıklar:** Kültürel normlar ve sosyal beklentiler duyguların sesli olarak nasıl ifade edildiğini etkileyebilmektedir.
- **Bağlamsal faktörler:** Konuşma konusu ve sosyal durum gibi çevredeki bağlam, konuşmanın akustik özelliklerini önemli ölçüde etkileyebilmektedir.

Bu analiz, farklı duyguların konuşmanın akustik özelliklerinde nasıl ortaya çıktığına dair daha derin bir anlayış sağlamaktadır. Bunların genel eğilimler olduğunu ve bireysel varyasyonların önemli olabileceğini unutmamak çok önemlidir.

3.3.3. Konuşma Dışı Seslerdeki Duygusal Korelasyon

Konuşmanın ötesinde, sözel olmayan seslendirmeler ve çevresel sesler de duygu ve şiddetle ilgili bilgi taşımaktadır. Bu bölümdeki analiz, bu konuşma dışı seslerin akustik özelliklerinin ve farklı duygusal durumlarla ilişkilerinin daha derinlemesine anlaşılmasını sağlamaktadır [94-95].

Kahkaha

Duygusal çağrışım olarak bakıldığında öncelikle mutluluk, eğlence ve neşe ile ilişkilendirilmektedir. Ancak kahkaha, bağlama bağlı olarak rahatlama, sinirlilik veya alaycılık gibi diğer duyguları da ifade edebilmektedir.

Akustik özellikler:

- **Ritmik patlamalar:** Kahkaha tipik olarak bir dizi kısa, ritmik ses patlamasından oluşmaktadır.
- **Dalgalanan perde:** Perde, her patlamada önemli ölçüde ve hızla değişerek karakteristik bir “müzikal” kalite yaratmaktadır.
- **Enerji dağılımı:** Enerji tipik olarak orta ila yüksek frekanslara yayılır ve kahkahanın parlak ve havadar kalitesine katkıda bulunmaktadır.
- **Spektral kontrast:** Her bir patlama içinde genlik ve frekanstaki hızlı değişimler nedeniyle yüksek spektral kontrast oluşmaktadır.

Çığlık Atma

Duygusal çağrışım olarak bakıldığında öncelikle korku, öfke ve acı ile ilişkilendirilmektedir. Panik veya terör vakalarında olduğu gibi aşırı sıkıntıya da işaret edebilmektedir.

Akustik özellikler:

- **Keskin perde yükselişi:** Genellikle perdede ani ve dramatik bir artış ile karakterize edilmektedir.
- **Kararsız frekans kalıpları:** Duygusal yoğunluğu ve potansiyel sıkıntıyı yansıtan hızlı ve öngörülemeyen frekans değişimleri gerçekleşmektedir.
- **Yüksek ses şiddeti ve spektral enerji:** Geniş bir frekans aralığında yüksek yoğunluk, havanın güçlü bir şekilde dışarı atılmasını yansıtmaktadır.
- **Yüksek sıfır geçiş oranı (ZCR):** Ani başlangıç ve genlikteki hızlı dalgalanmalar nedeniyle oluşmaktadır.

Ağlama

Duygusal çağrışım olarak bakıldığında öncelikle üzüntü, acı, sıkıntı ve keder ile ilişkilendirilmektedir. Bebeklerde korku veya hayal kırıklığı ile de ilişkilendirilebilmektedir.

Akustik Özellikler:

- **Düşük frekanslı bileşenler:** Düşük frekanslı bileşenler baskındır, genellikle nazal bir kaliteye sahip olmaktadır.
- **Düzensiz yoğunluk modülasyonları:** Genlikte ani artış ve azalışlarla birlikte yoğunlukta düzensiz değişimlerle karakterize edilmektedir.
- **Harmonik olmayan yapı:** Net harmonik içeriğin olmaması, daha “gürültülü” ve daha az yapılandırılmış bir sesle sonuçlanmaktadır.
- **Yavaş atak ve çürüme:** Ağlama seslerinin başlangıcı ve bozulması genellikle kademeli ve uzun bir formda karşımıza çıkmaktadır.

Ağır Solunum

Duygusal çağrışım olarak bakıldığında öncelikle endişe, gerginlik, korku ve efor ile ilişkilendirilir. Panik atak yaşayan bireylerde de görülebilmektedir.

Akustik Özellikler:

- **Yüksek titreşim ve pırıltı:** Hızlı ve düzensiz solunum kalıpları nedeniyle perde ve genlikte önemli değişimler oluşturmaktadır.
- **Düzensiz yoğunluk dalgalanmaları:** Solunumun çaba gerektiren doğasını yansıtan yoğunlukta hızlı artış ve azalışlar ile ortaya çıkmaktadır.
- **Kısa ve hızlı inhalasyonlar:** Genellikle kısa, hızlı nefes alışverişi birbirine eşlik ederek karakteristik bir nefes sesi modeli oluşturmaktadır.

Patlama ve Silah Sesi

Duygusal çağrışım olarak bakıldığında öncelikle tehdit, tehlike ve şiddet ile ilişkilendirilmektedir. Korku, endişe ve irkilme tepkilerini tetikleyebilmektedir.

Akustik Özellikler:

- **Darbe sesi:** Ani, yüksek yoğunluklu bir tepe noktası ve ardından hızlı bir azalma ile karakterize edilmektedir.
- **Geniş bant enerji:** Enerji geniş bir frekans aralığına dağılmıştır ve enerjinin ani salınımını yansıtmaktadır.
- **Yüksek sıfır geçiş oranı:** Sesin hızlı başlaması ve bozulması nedeniyle ortaya çıkmaktadır.

3.3.4. Duygu ve Şiddet Tespiti için Makine Öğrenimi Yaklaşımları

Otomatik ses tabanlı duygu tespiti için çeşitli makine öğrenmesi ve derin öğrenme teknikleri kullanılmaktadır [96-97].

Duygu Analizi için Kullanılan Özellik Çıkarma Teknikleri

- **MFCC'ler:** İnsan algısıyla daha ilgili frekansları vurgulayarak insan işitme sistemini taklit eden yaygın olarak kullanılan bir tekniktir. Ses sinyalinin spektral zarfını yakalayarak spektrumun genel şekli hakkında bilgi sağlamaktadırlar. Duygu uygunluğuna bakıldığında MFCC'ler, duygusal ifadeyle yakından bağlantılı olan ses yolu şekli ve ses kıvrımı titreşimlerdeki ince varyasyonları yakalamak için çok önemlidir. Örneğin, formant frekanslarındaki (ses yolunun rezonans frekansları) farklılıklar MFCC'ler tarafından etkili bir şekilde yakalanabilmekte ve duygu tanıma için değerli ipuçları sağlamaktadır.
- **Spektral kontrast:** Öne çıkan frekans bantlarının (tepe noktaları) enerjileri ile komşu vadileri arasındaki farkı ölçmektedir. Duygu uygunluğuna bakıldığında farklı duygusal durumların göstergesi olabilecek farklı tınısal özelliklere sahip sesler arasında ayırım yapılmasına yardımcı olmaktadır. Örneğin, öfke, artan vokal efor nedeniyle daha yüksek spektral kontrast ile karakterize edilen daha “törpülenmiş” bir sesle ilişkilendirilebilmektedir.
- **ZCR:** Ses sinyalinin belirli bir zaman dilimi içinde sıfır voltaj seviyesini kaç kez geçtiğini ölçer. Duygu uygunluğuna bakıldığında sinyalin zamansal dinamikleri hakkında bilgi sağlamaktadır. Daha yüksek ZCR değerleri artan ses yoğunluğunu ve heyecanı gösterebilirken, daha düşük ZCR değerleri üzüntüde olduğu gibi daha yavaş, daha bastırılmış konuşma ile ilişkilendirilebilmektedir.
- **Kroma özellikleri:** Batı müziğindeki 12 perde sınıfı boyunca enerji dağılımını temsil etmektedir. Duygu uygunluğuna bakıldığında kroma özellikleri, duygu tanıma için önemli ipuçları olan perde ve tonlamadaki ince varyasyonları yakalayabilmektedir. Örneğin, belirli perde sınıflarında artan enerji belirli duygularla ilişkilendirilebilmektedir.
- **Prozodik özellikler:** Bu özellikler konuşmanın zamansal ve ritmik yönlerini yakalamaktadır. Bunlar perde, yoğunluk, konuşma hızı ile duraklamalar ve süreler olarak bilinmektedir. Duygu uygunluğuna bakıldığında prozodik özellikler duygu tanıma için oldukça bilgilendiricidir. Örneğin, artan perde ve ses yüksekliği genellikle heyecanla ilişkilendirilirken, azalan perde ve ses yüksekliği üzüntüye işaret edebilmektedir.

Duygu Analizi için Kullanılan Makine Öğrenmesi Teknikleri

Özelliklerin ve makine öğrenimi modellerinin seçimi, özel uygulamaya ve mevcut verilere bağlıdır. Araştırmacılar, etkili özellik çıkarma tekniklerini güçlü makine öğrenimi modelleriyle

birleştirerek, otomatik ses tabanlı duygu tanıma için doğru ve sağlam sistemler geliştirme konusunda önemli ilerlemeler kaydetmişlerdir.

- **SVM'ler:** Özellik vektörlerinden karar sınırlarını etkili bir şekilde öğrenebilen güçlü sınıflandırma algoritmalarıdır. Ses sinyallerinden üretilenler gibi yüksek boyutlu özellik uzaylarını işlemek için çok uygundur [98-99].
- **Gizli markov modelleri (HMM'ler):** HMM'ler özellikle konuşma sinyalleri gibi sıralı verilerin modellenmesi için oldukça uygundur [100]. Zaman içinde farklı duygusal durumlar arasındaki geçişleri modelleyerek duyguların zamansal dinamiklerini yakalayabilmektedirler.
- **CNN'ler:** Verilerden uzamsal özelliklerin çıkarılmasında etkilidir. Ses sinyallerinin spektrogramlarına uygulandığında, CNN'ler farklı duygularla ilişkili kalıpları ve özellikleri tanımlamayı öğrenebilmektedir [101-102].
- **RNN'ler ve LSTM'ler:** Bu yöntemler zamansal bağımlılıkları ve uzun vadeli bağlamı yakalayıp ses sinyalleri gibi sıralı verileri işlemek için çok uygundur. LSTM'ler özellikle kaybolan gradyan sorununun üstesinden gelmede etkilidir ve ses verilerindeki uzun menzilli zamansal bağımlılıkları öğrenmelerine olanak tanımaktadır [18, 103].
- **Transformatör tabanlı modeller:** Doğal dil işlemede kullanılanlar gibi dönüştürücü modeller, son zamanlarda ses analizinde umut verici sonuçlar göstermişlerdir. Ses sinyalindeki uzun menzilli bağımlılıkları ve bağlamsal bilgileri etkili bir şekilde yakalayıp duygu tanıma görevlerinde daha iyi performans elde edilmesini sağlayabilmektedirler [104-105].

3.3.5. Seslerde Duygu ve Şiddet Analizinin Uygulamaları

Seslerdeki duygu ve şiddetin doğru analizinin, güvenlik, acil durum müdahalesi, sağlık hizmetleri ve akıllı gözetim sistemleri dâhil olmak üzere çeşitli alanlarda derin etkilere sahip olduğu bilinmektedir. Makine öğrenimi modelleri tehlike sinyallerini, agresif konuşmaları ve şiddet içeren sesleri etkili bir şekilde tespit ederek güvenliğini artırabilmekte, acil durum müdahalelerini otomatikleştirebilmekte ve genel refahı iyileştirebilmektedir.

Bu bölümde Duygu ve şiddet analizi için gerçekleştirilen ya da gerçekleştirilebilecek olan uygulamalar açıklanmaktadır.

Ses Analizi Yoluyla Kamu Güvenliği ve Şiddet Tespiti

Ses analizi, şiddet içeren veya saldırgan olayları tespit etmek için güçlü bir yol sunarak çeşitli ortamlarda çok önemli bir güvenlik katmanı sağlamaktadır. Bu bölümde ilgili temel alanlar ve teknolojiler açıklanmaktadır.

- **Agresif konuşma ve bağırımların tespit edilmesi:** Kızgın tonlar, bağırma veya tehdit edici dil gibi sözlü saldırganlığı tespit etmek, potansiyel şiddet durumlarına önleyici müdahale için gereklidir. Bu durum ile ilgili uygulamalar şu şekilde gerçekleştirilebilmektedir:
 - **Kolluk kuvvetleri:** Agresif etkileşimleri belirlemek ve kanıt sağlamak için vücut kameralarından veya gözetim sistemlerinden gelen sesin analiz edilmesiyle sağlanabilmektedir.
 - **Kamusal alan izleme:** Yıkıcı veya tehdit edici davranışları tespit etmek ve bunlara müdahale etmek için tren istasyonları, havaalanları, okullar ve halka açık meydanlar gibi yerlerdeki sesin izlenmesi gerekmektedir. Bu durum, güvenlik personelinin hızlı bir şekilde müdahale etmesine ve durumları yatıştırmasına yardımcı olabilmektedir.
- **Teknik yaklaşım:** Bu durumların tespit edilmesi için kullanılan yöntemler açıklanmaktadır.
 - **Özellik çıkarma:** İlk adım, agresif konuşmayı karakterize eden ses sinyalinin ilgili özelliklerin çıkarılmasını içermektedir. Bu özellikler konuşma hızı ve ritmi, frekans, yoğunluk ve formantlar olarak bilinmektedir.
 - **Makine öğrenimi:** Özellikler çıkarıldıktan sonra, sesi agresif veya agresif olmayan olarak sınıflandırmak için makine öğrenimi algoritmaları kullanılmaktadır.
 - **Derin öğrenme:** Sinir ağları, özellikle CNN'ler ve RNN'ler, ses verilerini analiz etmek için güçlü araçlar olarak nitelendirilmektedir. CNN'ler frekans alanındaki karmaşık kalıpları öğrenebilirken, RNN'ler konuşmadaki zamansal bağımlılıkları yakalamak için oldukça uygundur. Derin öğrenme modelleri, genellikle geleneksel özellik çıkarma yöntemlerinden daha iyi performans göstererek ilgili özellikleri otomatik olarak öğrenebilmektedir.
- **Silah atışı ve patlamaların tespit edilmesi:** Silah seslerini ve patlamaları hızlı ve doğru bir şekilde tespit etmek, anında müdahale ve bu tür olayların etkisini en aza indirmek için kritik öneme sahiptir. Bu durum ile ilgili uygulamalar şu şekilde gerçekleştirilebilmektedir:
 - **Akıllı şehir gözetimi** alanında Akustik silah atışı tespitinin şehir genelindeki gözetim sistemlerine entegre edilmesi, kolluk kuvvetlerinin olay yerine hızlı bir şekilde sevk edilmesini sağlayabilmektedir.
 - **Suç önleme** alanında gerçek zamanlı silah atışı tespiti, suç için caydırıcı olabilir ve genel kamu güvenliğini artırabilir.
 - **Askeri uygulamalar** alanında akustik sensörler düşman ateşini tespit etmek ve yerini belirlemek için kullanılabilir.

- **Teknik Yaklaşım:** Bu durumların tespit edilmesi için kullanılan yöntemler açıklanmaktadır.
 - **Akustik parmak izi:** Silah sesleri ve patlamaların benzersiz akustik imzaları vardır. Akustik parmak izi teknikleri bu imzaları analiz ederek havai fişek veya araba geri tepmesi gibi diğer yüksek seslerden ayırır. Bu, bilinen silah ve patlama seslerinden oluşan bir veri tabanı oluşturmayı ve gelen sesi bu parmak izleriyle karşılaştırmayı içermektedir.
 - **Spektral özellik çıkarımı:** Agresif konuşma tespitine benzer şekilde, spektral özellikler çok önemlidir. MFCC'ler ve Spektral kontrast oldukça sık kullanılmaktadır.
 - **Sınıflandırma:** Agresif konuşma tespiti için kullanılanlara benzer makine öğrenimi algoritmaları daha sonra çıkarılan özellikleri silah sesi, patlama veya diğer sesler olarak sınıflandırmak için kullanılmaktadır.
- **Kalabalık rahatsızlığı ve isyan tespiti:** Büyük toplantılar bazen yoğun bağırışlar, agresif kalabalık gürültüsü ve şiddetli çatışmalar içeren tehlikeli durumlara dönüşebilmektedir. Bu rahatsızlıkların erken tespiti, daha fazla tırmanmayı önlemek ve kamu güvenliğini sağlamak için çok önemlidir. Bu durum ile ilgili uygulamalar şu şekilde gerçekleştirilebilmektedir:
 - **Kolluk kuvvetleri:** Artan huzursuzluk belirtilerinin tespit edilmesi ve kaynakların etkili bir şekilde dağıtılması için halka açık etkinlikleri, protestoları ve gösterileri izlemek için kullanılabilir.
 - **Güvenlik güçleri:** Ayaklanmalar veya kitlesel panik durumlarını tespit etmek ve bunlara müdahale etmek için gözetim sistemlerinde ses analizinden yararlanmak için kullanılabilir.
 - **Kamusal etkinlik yönetimi:** Kalabalığın güvenliğinin sağlanması ve olası sorunları belirlemek için konserlerde, spor etkinliklerinde ve diğer büyük toplantılarda seslerin izlenmesi gerekebilir.
- **Teknik yaklaşım:** Gerçek ortamda yaşanan durumların tespit edilmesi için kullanılan yöntemler açıklanmaktadır.
 - **Ses yoğunluğu ve dinamikleri:** Ses yüksekliğindeki hızlı artışlar, sesin genel enerjisindeki değişiklikler ve ani, keskin seslerin varlığı tırmanan gerilime işaret edebilir.
 - **Seslendirmeler:** Kalabalık içindeki bağırma, çağırma ve çığlık atma kalıplarının tespit edilmesi. Bu seslendirmelerin sıklığı ve yoğunluğu kalabalığın duygusal durumunun göstergesi olabilir.

- **Spektral özellikler:** Saldırganlık veya panik ile ilişkili kalıpları belirlemek için kalabalık gürültüsünün frekans dağılımının analiz edilmesi sağlanmaktadır. Örneğin, yüksek frekanslı bileşenlerin varlığı duygusal uyarılmanın arttığını gösterebilmektedir.
- **Yapay zekâ tabanlı modeller:** Yapay zekâ (AI) ve makine öğrenimi modelleri, çıkarılan akustik özellikleri sınıflandırmak ve kalabalık rahatsızlıklarını veya isyanları tanımlamak için kullanılmaktadır. Bu modeller, barışçıl toplantılar, protestolar ve isyanlar dâhil olmak üzere çeşitli kalabalık senaryolarından alınan ses kayıtlarının büyük veri kümeleri üzerinde eğitilmektedir. Modeller, farklı kalabalık davranış türlerinin akustik imzalarını tanımayı öğrenmektedir.
- **Sınıflandırma:** Yapay zekâ modelleri, gelen ses verilerini gerçek zamanlı olarak sınıflandırarak bir kargaşa veya isyanın varlığını tespit edebilmektedir. Bu bilgi daha sonra yetkilileri uyarmak ve uygun bir müdahaleyi tetiklemek için kullanılabilir.

Akıllı Gözetim ve Acil Durum Sistemleri

Ses analizinin akıllı CCTV ve IoT cihazlarına entegre edilmesi, şiddet olaylarını, duygusal sıkıntıları ve diğer acil durumları tespit etmek için güçlü bir sistem oluşturmaktadır. Bu durum, geleneksel video gözetimine işitsel bir boyut katarak yeteneklerini önemli ölçüde artırmaktadır [106, 107].

- **Otomatik tehdit tespiti:** Bu durum, kapsamlı bir görsel-işitsel tehdit algılama sistemi oluşturmak için ses analizinin CCTV kameraları ve diğer gözetim sistemleri ile entegre edilmesini içermektedir.
- **Ses olayı tespiti:** Sistem, şiddet veya acil durumlarla ilişkili belirli sesleri tanımak üzere eğitilmiştir:
 - **Çığlıklar ve bağırsıklar:** Sıkıntı, korku veya çatışma göstergesi olarak tanımlanmaktadır.
 - **Cam kırılması:** Zorla girme veya başka bir saldırı olayı düşündürmektedir.
 - **Silah sesleri:** Potansiyel tehlikenin açık bir göstergesi olarak kabul edilmektedir.
 - **Fiziksel münakaşalar:** Kavgaya, vurma veya diğer fiziksel şiddet sesleri tehdit ortamı oluşturmaktadır.
- **CCTV ile entegrasyon:** Tehdit edici bir ses algılandığında, sistem CCTV kamerasını otomatik olarak tetikleyerek sesin geldiği yere odaklanabilir, olayı kaydedebilir ve güvenlik personeli uyarabilir. Bu durum, değerli bir görsel bağlam ve kanıt sağlamaktadır.
- **Akıllı gözetim ve acil durum sistemlerinde uygulamalar:**

- **Akıllı şehir güvenlik sistemleri:** Ses analizinin şehir genelindeki gözetim ağlarına entegre edilmesi, şiddet olaylarına hızlı müdahale edilmesini sağlayarak kamu güvenliğini artırabilmektedir.
- **Bina güvenliği:** Ofis binaları, okullar ve diğer tesisler güvenlik tehditlerini tespit etmek ve bunlara yanıt vermek için bu sistemleri kullanabilmektedir.
- **Kritik altyapı izleme:** Enerji santralleri veya ulaşım merkezleri gibi kritik altyapıların korunması, ses tabanlı tehdit algılama ile geliştirilebilmektedir.

Panik ve Tehlike Sinyali Tanıma

Yapay zekâ sistemleri, seslerdeki panik ve sıkıntının akustik özelliklerini tanımak için eğitilebilmektedir.

- **Ses analizi:** Sistem, panik, korku veya duygusal sıkıntı belirtilerini tanımlamak için tonu, perdeyi ve diğer ses özelliklerini analiz etmektedir. Hızlı konuşma, ses yoğunluğundaki değişiklikler ve belirli seslendirmeler gösterge olabilmektedir.
- **Acil durum müdahale ekiplerinin uyarılması:** Bir tehlike sinyali tespit edildiğinde, sistem acil durum hizmetlerini otomatik olarak uyarabilir, onlara tehlikedeki bireyin yerini ve hatta potansiyel olarak ses kaydını sağlayabilmektedir.
- **Panik ve tehlike sinyali tanıma ile ilgili uygulamalar:**
 - **Acil çağrı merkezleri:** Yapay zekâ destekli sistemler, operatörlerine tehlike altındaki bireylerden gelen çağrıları hızlı bir şekilde tanımlama ve önceliklendirme konusunda yardımcı olabilmektedir.
 - **Akıllı ev güvenlik sistemleri:** Ev güvenlik sistemleri, tehlike sinyallerini dinleyen mikrofonlarla donatılabilmektedir. Bir panik sesi algılanırsa, sistem otomatik olarak yardım çağırabilmektedir.
 - **Yapay zekâ destekli ev asistanları:** Akıllı hoparlörler ve diğer ev asistanları, seslerdeki sıkıntıyı tanıyacak ve acil servislerle veya belirlenen bir irtibat kişisiyle iletişime geçmek gibi uygun eylemleri gerçekleştirecek şekilde tasarlanabilmektedir.

Cinsiyete Dayalı Şiddet (GBV) Tespiti

Potansiyel aile içi şiddeti veya diğer cinsiyete dayalı şiddet türlerini tespit etmek için ses analizinin kullanılması oldukça önem arz etmektedir [108, 109].

- **Örüntü tanıma:** Makine öğrenimi modelleri, GBV ile ilişkili ses kalıplarını tanımak için eğitilebilmektedir. Örneğin, seslerde yüksek ton ve ses seviyesi; hakaret, tehdit ve bağırma dâhil olmak üzere sözlü saldırganlık; fiziksel kavgaların göstergesi olan belirli ses kalıpları vb.

- **Gizlilikle ilgili hususlar:** GBV tespiti, mahremiyet konularının dikkatle değerlendirilmesini gerektirmektedir. Veriler etik bir şekilde toplanmalı ve analiz edilmeli, bireylerin mahremiyetini korumak için uygun önlemler alınması gerekmektedir. Potansiyel kullanım alanları şu şekilde açıklanmaktadır:
 - **Akıllı ev güvenliği:** Ev güvenlik sistemleri aile içi şiddet belirtilerini tespit edecek ve yetkilileri veya belirlenmiş destek kuruluşlarını uyaracak şekilde tasarlanabilmektedir.
 - **Kadın güvenliği uygulamaları:** Sesi analiz etmek ve potansiyel olarak tehlikeli durumlarda kadınlara yardım sağlamak için mobil uygulamalar geliştirilebilmektedir. Bu, ses kanıtlarının kaydedilmesi veya acil servislerle hızlı bir şekilde iletişime geçilmesi gibi özellikleri içerebilmektedir.

Sese dayalı gözetimin temel avantajları şu şekilde açıklanmaktadır:

- **Erken tespit:** Ses genellikle yaklaşan bir olayla ilgili erken uyarı işaretleri sağlayarak daha hızlı müdahaleye olanak tanıyabilmektedir.
- **Video ile tamamlayıcıdır:** Ses analizi, geleneksel video gözetimine çok önemli bir bilgi katmanı ekleyerek durumun daha eksiksiz bir resmini sunmaktadır.
- **Gizlilikle ilgili hususlar:** Gizlilik, özellikle ev içi sistemlerde bir endişe kaynağı olsa da tehlikede olanlara yardım etme potansiyeli önemlidir. Etik geliştirme ve dağıtım oldukça önemlidir.

Akıllı gözetim ve acil durum sistemlerine sağlam analizler eklenerek daha güvenli ortamlar yaratılabilir ve çeşitli kritik durumlara yanıt verme becerisi geliştirilebilir.

Ses Analizi ile Sağlık ve Ruh Sağlığı İzleme

Ses sinyallerini analiz etmek, bir kişinin duygusal ve zihinsel durumu hakkında değerli bilgiler sunarak sağlık hizmetleri ve terapi için yeni olanaklar sağlamaktadır. İşte bu bölüm uygulamalara ayrıntılı bir bakış sunmaktadır.

- **Hastalarda duygusal sıkıntıyı tespit etme:** Stres, anksiyete veya depresyon gibi duygusal sıkıntıların belirlenmesi ve ölçülmesi zor olabilmektedir. Geleneksel yöntemler genellikle öznel ve güvenilir olmayan öz raporlamaya dayanmaktadır. Ses analizi, duygusal durumu değerlendirmek için objektif ve müdahaleci olmayan bir yol sağlamaktadır.
- **Uygulamalar:**
 - **Ruh sağlığı terapisi:** Terapi seansları sırasında konuşma kalıplarının analiz edilmesi, terapistlere hastanın duygusal durumu hakkında değerli bilgiler sağlayarak yaklaşımlarını buna göre uyarlamalarına olanak tanıyabilmektedir.
 - **İntiharı önleme yardım hatları:** Yapay zekâ destekli sistemler, yüksek risk altındaki bireyleri belirlemek için intiharı önleme yardım hatlarını arayanların seslerini analiz

edebilmektedir. Bu, operatörlerin çağrılara öncelik vermesine ve zamanında müdahale sağlamasına yardımcı olabilmektedir.

- **Uzaktan hasta izleme:** Ruh sağlığı sorunları olan hastalar için ses analizi, duygusal durumlarını uzaktan izlemek ve potansiyel nüksleri tespit etmek için kullanılabilmektedir.
- **Teknik yaklaşım:** Konuşmadan duygusal sıkıntının göstergesi olan akustik özelliklerin çıkarılması işlemine özellik çıkarma denir ve bu özellikler şunları içermektedir:
 - **Perde değişiklikleri:** Stres ve kaygı genellikle bir kişinin sesinin perdesinde ve değişkenliğinde değişikliklere yol açabilmektedir.
 - **Vokal yoğunluğu:** Ses yüksekliği ve ses seviyesindeki değişimler duygusal uyarılmayı yansıtabilmektedir.
 - **Konuşma hızı ve ritmi:** Konuşma hızı ve ritmindeki değişiklikler de duygusal sıkıntının göstergesi olabilmektedir.
 - **Ses kalitesi:** Titreme ve parıltı gibi özellikler, duygusal durumdan etkilenebilen ses teli gerginliği ve kontrolündeki değişiklikleri yansıtabilmektedir.
 - **Makine öğrenimi:** Makine öğrenimi modelleri, farklı duygusal durumlarla ilişkili olan bu akustik özelliklerdeki kalıpları tanımak için eğitilmektedir. Örneğin yapay zekâ destekli sohbet robotları, bireylerle etkileşim kurmak ve depresyon veya diğer ruh sağlığı sorunlarının belirtilerini tespit etmek için konuşma kalıplarını analiz etmek için kullanılabilmektedir. Bu durum, erken müdahale ve destek için değerli bir araç sağlayabilmektedir.
- **TSSB ve travma tepkisi izleme:** Travma Sonrası Stres Bozukluğu (TSSB) olan bireyler önemli duygusal sıkıntılar yaşayabilmekte ve tetikleyicilere yanıt olarak belirli ses değişiklikleri sergileyebilmektedir. Bu teknoloji özellikle askeri gaziler ve travma mağdurları için uygundur.
- **Uygulamalar:**
 - **Travma tepkilerinin izlenmesi:** Ses analizi, TSSB ataklarıyla ilişkili ses değişikliklerini izlemek için kullanılabilmekte ve bireyin duygusal durumu hakkında değerli bilgiler sağlamaktadır.
 - **Terapi desteği:** Bu bilgiler terapi seanslarını desteklemek için kullanılabilir ve terapistlerin tetikleyicileri daha etkili bir şekilde tanımlamasına ve ele almasına yardımcı olmaktadır.
- **Teknik yaklaşım:**
 - **Perde dalgalanmaları:** Artan perde değişkenliği kaygı ve stres belirtisi olabilmektedir.

- **Titreşim ve pırlıltı:** Bu ses kalitesi ölçümleri duygusal uyarılmadan etkilenebilmektedir.
- **Nefeslilik:** Nefes kontrolündeki değişiklikler de duygusal sıkıntının göstergesi olabilmektedir.
- **Yaşlılar ve hasta bakımı:** Yaşlı bireyler ve hareketlilik sorunları olan hastalar düşme ve diğer kazalar açısından risk altındadır. Ses analizi bu olayları tespit etmek ve zamanında yardım sağlamak için kullanılabilir.
- **Uygulamalar:**
 - **Düşme tespiti:** Ani darbe sesleri düşmeleri tespit etmek ve bakıcıları veya acil servisleri otomatik olarak uyarmak için kullanılabilir.
 - **Yardım çağrıları:** Yapay zekâ sistemleri, bir yardım çağrısına işaret eden belirli sesleri veya konuşma kalıplarını tanımak üzere eğitilebilir.
- **Kullanım örnekleri:**
 - **Yaşlı izleme sistemleri:** Bu sistemler, yaşlı sakinlerin güvenliğini izlemek için evlere veya destekli yaşam tesislerine kurulabilir.
 - **Hastane hasta izleme:** Ses analizi, düşmeleri veya diğer acil durumları tespit etmek için hastane hasta izleme sistemlerine entegre edilebilir.
- **Sağlık hizmetlerinde ses analizinin temel faydaları:**
 - **Objektif ölçüm:** Duygusal durum ve sıkıntının objektif bir ölçümünü sağlayarak kendi kendine raporlamaya olan güveni azaltmaktadır.
 - **Müdahaleci olmayan:** Ses analizi, bireyin aktif katılımını gerektirmeden pasif olarak gerçekleştirilebilir.
 - **Uzaktan izleme:** Hastaların uzaktan izlenmesini sağlayarak potansiyel sorunların erken tespit edilmesine olanak tanımaktadır.
 - **Erken müdahale:** Ruh sağlığı sorunları olan veya düşme riski taşıyan bireyler için erken müdahaleyi ve desteği kolaylaştırmaktadır.

Ses analizinin gücünden yararlanılarak sağlık hizmeti sonuçları iyileştirilebilir, ruh sağlığı desteği geliştirebilir ve savunmasız nüfuslar için daha güvenli ortamlar yaratılabilir.

İnsan-Bilgisayar Etkileşimi ve YZ Asistanları

Ses analizinin, özellikle de duygu ve şiddet tespitinin YZ sistemlerine entegre edilmesi, daha uyarlanabilir, akıllı ve empatik etkileşimlere olanak tanımaktadır. Bu durum, teknolojiyle etkileşim şeklini başka bir sonuca dönüştürmektedir [110-112].

- **Yapay zekâ destekli müşteri hizmetleri:** Müşteri hizmetleri etkileşimleri, özellikle de telefon üzerinden hem müşteri hem de temsilci için stresli olabilmektedir. Müşterinin hayal

kırıklığını veya saldırganlığını erken tespit etmek, durumları yatıştırmak ve müşteri memnuniyetini artırmak için oldukça önemlidir.

- **Duygu algılama:** Yapay zekâ sistemleri, hayal kırıklığı, öfke veya diğer olumsuz duyguların belirtilerini tespit etmek için bir telefon görüşmesi sırasında müşterinin sesini analiz edebilmektedir. Bu, ses perdesi, ton, konuşma hızı ve enerji gibi ses özelliklerinin analiz edilmesini içermektedir.
- **Saldırganlık tespiti:** Sistem ayrıca agresif dili veya sözlü tehditleri de tespit edebilmektedir.
- **Otomatik yanıt:** Algılanan duygulara dayanarak, yapay zekâ sistemi çeşitli eylemlerde bulunabilmektedir:
 - **İnsan temsilcilere eskalasyon:** Müşteri çok tedirgin veya agresifse, çağrı otomatik olarak zor durumlarla başa çıkmak için eğitilmiş bir insan temsilciye yükseltilebilmektedir.
 - **Kişiselleştirilmiş yanıtlar:** Yapay zekâ, yanıtlarını müşterinin duygusal durumuna göre uyarlayabilmektedir. Örneğin, müşteri hayal kırıklığına uğramışsa, yapay zekâ özür dileyebilir ve sorunu hızlı bir şekilde çözmeye çalışabilmektedir.
 - **Duygu analizi:** Basit duygu tespitinin ötesinde, YZ konuşmanın genel tonunu anlamak ve etkileşimi buna göre uyarlamak için duygu analizi yapabilmektedir.

Örneğin otomatik çağrı merkezleri, müşteri seslerini gerçek zamanlı olarak analiz etmek için yapay zekâyı kullanabilmektedir. Bir müşterinin aşırı sinirli olduğu tespit edilirse, sistem çağrıyı otomatik olarak gerginliği azaltma teknikleri konusunda uzman bir müşteri hizmetleri temsilcisine aktarabilmektedir.

- **Akıllı Asistanlar (Alexa, Google Asistan, Siri):** Mevcut akıllı asistanlar öncelikle doğrudan komutlara yanıt vermektedir. Ancak kullanıcının duygusal durumunu anlayarak çok daha yardımcı ve sezgisel olabilmektedirler.
 - **Duygu tanıma:** Akıllı asistanlar, sesli etkileşimlerde korku, öfke, sıkıntı, üzüntü ve neşe gibi bir dizi duyguyu tanıyacak şekilde tasarlanabilmektedir.
 - **Bağlamsal anlama:** Duygu tanımının doğal dil işleme ile birleştirilmesi, asistanın kullanıcının ihtiyaçlarını ve niyetlerini daha iyi anlamasını sağlamaktadır.
 - **Proaktif yardım:** Algılanan duygulara dayanarak, akıllı asistan proaktif adımlar atabilmektedir:
 - **Acil durum yardımı:** Asistan, kullanıcının sesinde sıkıntı veya korku tespit ederse, otomatik olarak acil servisleri veya belirlenen bir irtibat kişisini aramayı teklif edebilmektedir.

- **Duygusal destek:** Kullanıcının sesi üzgün veya sinirli geliyorsa, asistan rahatlatıcı sözler söyleyebilir, rahatlatıcı bir müzik çalabilir veya başka yardımcı kaynaklar önerebilmektedir.
- **Kişiselleştirilmiş öneriler:** Asistan, müzik, film veya diğer içerik önerilerini kullanıcının duygusal durumuna göre uyarlayabilmektedir.

Örneğin bir kullanıcının akıllı hoparlörüne “Kendimi çok bunalmış hissediyorum” dediği bir senaryo düşünün. Kullanıcının sıkıntılı tonunu tanıyan akıllı asistan şöyle yanıt verebilmektedir: “Bunu duyduğuma üzıldüm. Sakinleştirici bir müzik çalmamı ya da rahatlama teknikleri bulmamı ister misiniz?” Ya da daha kritik bir durumda, bir kullanıcı acı içinde bağırırsa, asistan hemen tıbbi yardıma ihtiyacı olup olmadığını sorabilmekte ve 112'yi aramayı teklif edebilmektedir.

Duygusal olarak akıllı yapay zekâ asistanlarının temel faydaları:

- **Geliştirilmiş kullanıcı deneyimi:** Yapay zekâ asistanları duyguları anlayarak ve bunlara yanıt vererek daha kişiselleştirilmiş ve empatik bir kullanıcı deneyimi yaratabilmektedir.
- **Gelişmiş güvenlik:** Sıkıntı veya acil durumları tespit etme yeteneği, akıllı asistanları kullanıcı güvenliğini sağlamak için değerli bir araç haline getirebilmektedir.
- **Artan yardımseverlik:** Duygusal olarak akıllı asistanlar kullanıcı ihtiyaçlarını daha iyi tahmin edebilir ve daha ilgili ve yararlı yardım sağlayabilmektedir.
- **Daha güçlü insan-bilgisayar bağı:** Duygulara yanıt vererek, yapay zekâ insanlar ve teknoloji arasında daha güçlü bir bağlantı ve güven duygusu yaratabilmektedir.

YZ teknolojisi gelişmeye devam ettikçe, cihazlarımızda daha da sofistike ve duygusal olarak akıllı etkileşimler görülmesi beklenmektedir. Ses analizi, bunu gerçeğe dönüştürmede çok önemli bir rol oynayacaktır.

Kolluk Kuvvetleri ve Adli Uygulamalar

Ses analizi, modern polislik ve adli bilimlerde çok önemli bir rol oynamaktadır. Özellikle suç tespiti, soruşturma ve kanıt toplama için değerli araçlar sunmaktadır.

- **112 çağrılarında şiddet tespiti:** Acil çağrı merkezleri çok sayıda çağrı alır ve şiddet ya da acil tehlike içeren çağrılar hızla bir şekilde tespit edilmesi, kaynakların etkin bir şekilde sevk edilmesi açısından kritik önem taşımaktadır.
- **Duygusal sıkıntı analizi:** Yapay zekâ sistemleri, korku, panik veya ajitasyon gibi duygusal sıkıntı belirtilerini tespit etmek için arayanın sesini analiz etmektedir. Bu durum, ses perdesi, ton, konuşma hızı ve enerji gibi ses özelliklerinin incelenmesini içermektedir.
- **Tehdit edici konuşma örüntüsü tanıma:** Sistem, şiddet tehdidine işaret eden belirli dil veya konuşma kalıplarını tanımlayabilmektedir.

- **Arka plan şiddet sesi tespiti:** Yapay zekâ algoritmaları, çağrının arka planındaki bağırma, vurma veya cam kırma gibi fiziksel kavgaların seslerini tanımak için eğitilebilmektedir.
- **Gerçek zamanlı uyarı:** Sistem şiddet belirtileri tespit ettiğinde, 112 operatörünü anında uyararak çağrıyı önceliklendirmesini ve polis memurlarını mümkün olan en kısa sürede olay yerine göndermesini sağlayabilmektedir.
- **112 çağrılarında şiddet tespitinin faydaları:**
 - **Yüksek riskli çağrılara öncelik verme:** Yakın tehlike içeren çağrılar hızlı bir şekilde belirleyin ve yanıtlayın.
 - **Müdahale sürelerini iyileştirme:** Polis memurlarını olay yerine daha hızlı göndererek potansiyel olarak daha fazla zararı önlemektedir.
 - **Kanıt toplama:** Ses kayıtları cezai soruşturmalarda değerli kanıtlar olarak kullanılabilir.
- **Adli ses analizi:** Suç mahallerinden, gözetleme sistemlerinden veya diğer kaynaklardan alınan ses kayıtları değerli bilgiler içerebilmektedir. Ancak bu bilgileri açığa çıkarmak zor olabilmektedir.
 - **Ses iyileştirme:** Adli ses uzmanları ses kayıtlarını geliştirmek, arka plan gürültüsünü azaltmak, konuşmayı netleştirmek ve kaydın genel kalitesini artırmak için özel yazılımlar kullanmaktadır.
 - **Ses tanımlama:** Bir şüphelinin sesini suç mahallinde kaydedilen bir sesle karşılaştırmak için ses özelliklerini analiz etmek gerekmektedir. Bu durum, perde, formant frekansları ve konuşma kalıpları gibi özelliklerin incelenmesini içermektedir.
 - **Konuşmacı diyarizasyonu:** Bir kayıttaki farklı konuşmacıların seslerini tanımlama ve ayırma olarak bilinmektedir.
 - **Transkripsiyon:** Konuşulan kelimelerin metne dönüştürülmesi işlemi içermektedir.
 - **Akustik olay analizi:** Bir kayıttaki silah sesi, patlama veya cam kırılması gibi belirli sesleri tanımlama ve analiz etme durumunu açıklamaktadır.
- **Uygulamalar:**
 - **Suç soruşturmaları:** Kanıt toplamak, şüphelileri belirlemek ve olayları yeniden yapılandırmak için ses kayıtlarını analiz etme adımlarını içermektedir.
 - **Mahkeme kanıtları:** Geliştirilmiş ses kayıtları ve uzman analizi mahkemede kanıt olarak sunulabilmektedir.
- **Sahte tehlike çağrılarını tespit etme:** Sahte acil durum çağrılarını veya “swatting”, değerli kaynakları boşa harcayan ve potansiyel olarak insanları tehlikeye atan ciddi bir sorun olabilmektedir.

- **Aldatıcı konuşma örüntüsü analizi:** Yapay zekâ sistemleri, aldatma göstergesi olan konuşma kalıplarını tanımak üzere eğitilebilmektedir. Bu durum, arayanın hikayesindeki tutarsızlıkları, olağandışı duraklamaları veya diğer dilsel ipuçlarını içerebilmektedir.
- **Duygusal uyumsuzluk tespiti:** Sistem, arayan kişinin sesini analiz ederek duygusal durumunun bildirdiği durumla tutarlı olup olmadığını görebilmektedir. Örneğin, şiddet içeren bir suçu bildiren ve sesi alışılmadık derecede sakin çıkan bir arayan şüphe uyandırabilmektedir.
- **Bağlamsal analiz:** Aramanın güvenilirliğini değerlendirmek için ses analizinin arayanın konumu ve telefon numarası gibi diğer bilgilerle birleştirilmesi işlemini gerçekleştirmektedir.
- **Sahte tehlike çağrılarını tespit etme faydaları:**
 - **Kaynak yönetimi:** Kolluk kuvvetlerinin yanlış alarmlar için zaman ve kaynak israfını önlemesine yardımcı olmaktadır.
 - **Kamu güvenliği:** “Swatting” olaylarından kaynaklanan tehlikeli durum riskini azaltmaktadır.
- **Kolluk kuvvetlerinde ses analizinin temel avantajları:**
 - **Nesnel kanıt:** Ses kayıtları, soruşturmalarda ve mahkeme işlemlerinde kullanılabilecek objektif kanıtlar sağlamaktadır.
 - **Gelişmiş durumsal farkındalık:** Ses analizi, video gözetimi tarafından yakalanamayabilecek olaylar hakkında değerli bilgiler sağlayabilmektedir.
 - **İyileştirilmiş müdahale süreleri:** 112 çağrılarının gerçek zamanlı analizi, acil durum müdahale ekiplerinin yüksek riskli durumlara öncelik vermesine ve kaynakları daha etkili bir şekilde sevk etmesine yardımcı olabilmektedir.

Kolluk kuvvetleri, ses analizinin gücünden yararlanarak suçları önleme ve çözme, kamu güvenliğini artırma ve adaleti sağlama becerilerini geliştirebilmektedir.

Ses Aracılığıyla Eğlence ve Medya Analizi

Duygu ve şiddet tespiti de dâhil olmak üzere ses analizi, eğlence ve medya alanında giderek artan bir kullanım alanı bulmakta ve içerik moderasyonunu, kullanıcı deneyimini ve daha sürükleyici ve kişiselleştirilmiş eğlencenin yaratılmasını etkilemektedir [113, 114].

- **Otomatik içerik derecelendirme ve moderasyon:** Günümüzde çok sayıda kullanıcı tarafından oluşturulan içerik ve profesyonel olarak üretilen medya mevcutken, içeriği yaşa uygunluk veya politika ihlalleri açısından manuel olarak derecelendirmek ve denetlemek göz korkutucu bir görev olarak öne çıkmaktadır.

- **Şiddet içeren konuşma tespiti:** Yapay zekâ sistemleri, tehditler, hakaretler ve nefret söylemi dâhil olmak üzere şiddet içeren veya saldırgan konuşma örneklerini tespit etmek için filmlerin, TV şovlarının, video oyunlarının ve diğer medyanın ses kaydını analiz edebilmektedir.
- **Agresif ses algılama:** Sistem, konuşmanın ötesinde, silah sesleri, patlamalar, çığlıklar veya fiziksel kavga sesleri gibi diğer saldırgan sesleri de tespit edebilmektedir.
- **Duygusal ton analizi:** Yapay zekâ, sesin genel duygusal tonunu değerlendirerek aşırı rahatsız edici veya korkutucu içeriği belirleyebilmektedir.
- **Otomatik etiketleme ve kategorizasyon:** Ses analizine dayanarak, sistem içeriği otomatik olarak etiketleyebilir ve kategorize edebilir, böylece filtrelemeyi ve yönetmeyi kolaylaştırmaktadır.
- **Ses yoluyla eğlence ve medya analizinin avantajları:**
 - **Ölçeklenebilirlik:** İçerik moderasyonunun otomatikleştirilmesi, yayın platformları ve medya sağlayıcıları için geniş ölçekli içerik kütlelerinin etkin bir şekilde yönetilmesine olanak tanımaktadır.
 - **Tutarlılık:** Yapay zekâ tabanlı sistemler, içerik derecelendirme ve moderasyon politikalarını insan denetçilere göre daha tutarlı ve tarafsız bir şekilde uygulayabilmektedir.
 - **İnsan denetimi yükünün hafifletilmesi:** Otomasyon, insan denetçilerin daha karmaşık ve nüanslı durumlara odaklanabilmesi için gereken zamanı ve kaynakları serbest bırakmaktadır.
 - **Yaşa uygun içerik sunumu:** Akış platformları, ses analizi teknolojilerini kullanarak kullanıcıların yalnızca yaşlarına uygun içeriklerle etkileşime girmesini garanti altına almaktadır.
- **Uyarlanabilir oyun ve sanal gerçeklik:** Geleneksel video oyunları ve sanal gerçeklik deneyimleri, sabit bir zorluk ve daldırma seviyesi sunmaktadır. Ses analizi, oyuncunun duygusal durumuna yanıt veren daha dinamik ve uyarlanabilir deneyimler oluşturmak için kullanılabilir.
- **Oyuncu duygu tanıma:** Yapay zekâ sistemleri, korku, heyecan, hayal kırıklığı veya can sıkıntısı gibi duygusal durumlarını tespit etmek için oyun sırasında oyuncunun sesini analiz edebilmektedir. Bu durum, ses perdesi, hızı ve yoğunluğu gibi ses ipuçlarını analiz ederek yapılabilir.
- **Gerçek zamanlı oyun ayarlama:** Oyuncunun duygularına bağlı olarak oyun, oyunun çeşitli yönlerini dinamik olarak ayarlayabilmektedir:

- **Zorluk:** Eğer oyuncu oyunda sürekli olarak hızlı ilerliyorsa, zorluk derecesi artırılabilir. Tersine, oyuncu mücadele ediyor ve hayal kırıklığına uğruyorsa, oyun ipuçları sunabilir veya zorluğu azaltılabilir.
- **Hikâye:** Hikâye, oyuncunun duygusal durumuna yanıt vermek için ince bir şekilde değiştirilebilir. Örneğin, bir korku oyununda, atlama korkuları, yüksek korku anlarına denk gelecek şekilde zamanlanabilir.
- **Ses tasarımı:** Oyunun ses ortamı oyuncunun duygularını yansıtacak şekilde uyarlanabilir. Örneğin, müzik gerilim anlarında daha yoğun hale gelebilir veya rahatlama dönemlerinde daha sakinleştirici olabilir.

Mesela oyuncunun gerçek zamanlı korku seviyelerini tespit etmek için ses analizi kullanan bir korku oyunu düşünüldüğünde eğer oyuncu gerçekten korkmuşsa, oyun korkutucu sesleri yükseltebilir, Tersine, oyuncu sıkılmış veya etkilenmemiş görünüyorsa, oyun yoğunluğu azaltılabilir veya yeni zorluklar getirebilir.

- **Uyarlanabilir oyun ve sanal gerçekliğin faydaları:**
 - **Geliştirilmiş sürükleyicilik:** Uyarlanabilir oyun, oyuncu için daha sürükleyici ve ilgi çekici bir deneyim yaratabilir.
 - **Kişiselleştirilmiş eğlence:** Oyun, deneyimi bireysel oyuncunun tercihlerine ve duygusal durumuna göre uyarlayabilir.
 - **Artan ilgi:** Uyarlanabilir sistemler, oyunun zorluk derecesini ve diğer yönlerini dinamik olarak ayarlayarak oyuncuların ilgisini çekmeye ve eğlenmelerine yardımcı olabilir.
- **Uyarlanabilir oyun ve sanal gerçeklikte gelecekteki yönelimler:** Ses analizi teknolojisi gelişmeye devam ettikçe, eğlence ve medya endüstrisinde daha da yenilikçi uygulamalar görülmesi beklenebilir. Bu uygulamalar şunları içerebilir:
 - **Kişiselleştirilmiş müzik önerileri:** Bir kullanıcının sesini analiz ederek o anki ruh haline uygun müzikler öneren yapay zekâ sistemleri geliştirilebilir.
 - **İnteraktif hikâye anlatımı:** Hikâyenin kullanıcının duygusal tepkilerine göre gerçek zamanlı olarak uyarlandığı deneyimler gerçekleştirilebilir.
 - **Geliştirilmiş sanal gerçeklik:** Kullanıcının duygularına ve fiziksel tepkilerine yanıt veren gerçekten sürükleyici VR deneyimleri yaratılabilir.

Değerlendirme

Seslerde hareket ve şiddet analizi, güvenlik, sağlık, yapay zekâ ve eğlence alanlarında geniş uygulamalara sahiptir. Makine öğrenimi, sistemlerin saldırganlığı, sıkıntıyı ve tehditleri tespit etmesini sağlayarak kamu güvenliğini, acil durum müdahalesini ve insan-bilgisayar etkileşimlerini iyileştirmektedir.

4. METODOLOJİ

Bu bölümde, tez çalışmasının bilimsel temelini oluşturan araştırma metodolojisi ve uygulanan yöntemler ve gerçekleştirilen çalışmalara özgü bilgiler ayrıntılı olarak sunulmaktadır.

4.1. Makine Öğrenmesi

Makine öğrenimi sistemleri temelde bilgi edinimi için birincil kaynak olarak verilere dayanır, altta yatan kalıpları belirlemek ve değerli içgörüler oluşturmak için verileri kullanmaktadır. Bu veriler, geliştirilmiş modellerin eğitimi için temel taşı görevi görmektedir. Açık talimatların titizlikle tanımlandığı geleneksel programlama paradigmasının aksine, makine öğrenimi algoritmaları, yeni verilere sürekli maruz kalarak performanslarını otonom bir şekilde öğrenme ve geliştirme yeteneğine sahiptir. Bu süreç, giderek daha doğru tahminler ve karar verme becerileriyle sonuçlanır. Makine öğrenimini benzersiz kılan özellik, sistemlerin insan tarafından belirlenmiş açık kurallara bağlı kalmak yerine, verilerin içindeki doğal kalıpları ve ilişkileri tanımlayarak çözümler üretebilmesidir. Bu yaklaşım, makine öğrenimini geleneksel yöntemlerden ayıran temel bir unsurdur.

4.1.1. Makine Öğrenmesi Türleri

Makine öğrenmesi yöntemleri 4'e ayrılır.

Denetimli Öğrenme

Denetimli öğrenme, etiketli veri kümeleri üzerinde eğitilen makine öğrenimi algoritmalarını kapsar; burada her bir girdi örneği, karşılık gelen bir çıktı (etiket) ile açıkça ilişkilendirilmektedir. Bu paradigmadaki temel amaç, görünmeyen girdi verileri için çıktıyı doğru bir şekilde tahmin eden sağlam bir eşleme işlevi öğrenmektir. Denetimli öğrenme görevlerinin açıklayıcı örnekleri arasında, çeşitli özelliklere dayalı olarak konut fiyatlarının tahmin edilmesi gibi regresyon analizi ve bir e-posta mesajının spam mı yoksa yasal yazışma mı olduğunun belirlenmesi gibi sınıflandırma sorunları yer almaktadır [115].

Denetimsiz Öğrenme

Denetimsiz öğrenme, açık bir çıktı veya hedef değişkenin sağlanmadığı etiketsiz veri kümeleri üzerinde çalışan bir makine öğrenimi algoritmaları sınıfını kapsamaktadır. Bu paradigmadaki birincil amaç, verilerin kendi içindeki yapıları, örüntüleri veya ilişkileri tanımlamaktır. Denetimsiz öğrenme görevlerinin açıklayıcı örnekleri arasında, müşterileri satın

alma davranışlarına göre gruplandırmak gibi kümeleme ve en önemli bilgileri korurken bir veri kümesindeki özelliklerin sayısını azaltmayı amaçlayan boyut azaltma teknikleri yer almaktadır.

Yarı Denetimli Öğrenme

Yarı denetimli öğrenme hem denetimli hem de denetimsiz öğrenme paradigmalarının unsurlarından yararlanan hibrit bir yaklaşımı kapsamaktadır. Bu metodoloji, sınırlı miktarda etiketli veriyi çok daha büyük hacimde etiketsiz veriyle birlikte etkin bir şekilde kullanır. Bu yaklaşım, veri etiketleme işleminin çok pahalı ya da aşırı zaman alıcı olduğu senaryolarda özellikle değerlidir.

Takviyeli Öğrenme

Takviyeli öğrenme, bir ajanın bir çevre ile etkileşime girme ve ödülleri veya cezalar şeklinde geri bildirim alma gibi yinelemeli bir süreç yoluyla karar vermeyi öğrendiği makine öğrenimi içinde bir paradigmadır. Bu geri bildirim, ajanın öğrenme sürecine rehberlik ederek zaman içinde karar verme stratejilerini iyileştirmesini sağlamaktadır.

4.1.2. Makine Öğrenmesi Bileşenleri

Veri, modeli besleyen hammaddestir. Özellikler, verilerin anlamlı temsidir. Model, Verilerdeki ilişkileri temsil eden matematiksel yapıdır. Algoritma, modeli oluşturan ve eğiten prosedürdür. Değerlendirme, Modelin performansını ölçmek için kullanılan metriklerdir. Bu beş kavram makine öğreniminin yapı taşlarını oluşturmaktadır. Birlikte, verilerden anlamlı bilgiler çıkarabilen ve gelecekteki olaylar hakkında tahminler yapabilen güçlü sistemler oluşturmamızı sağlamaktadırlar [116].

Veri: Makine öğreniminin temelidir. Yüksek kaliteli ve ilgili veriler daha iyi model performansı sağlamaktadır. Modelin öğreneceği bilgileri içermektedir.

Özellikler: Ham verilerden türetilen nitelikler veya özellikler ortaya çıkarılmaktadır. Makine öğrenimi modeli için girdi olarak kullanılmaktadır. Verilerin anlamlı bir temsili sağlamaktadırlar.

Model: Verilerdeki ilişkilerin matematiksel bir temsil etmektedir. Verilerden öğrenilen kalıpları genelleştirmektedir. Örnekler arasında doğrusal regresyon, karar ağaçları ve sinir ağları yer almaktadır.

Algoritma: Verilerdeki örüntüleri öğrenerek model oluşturmaya yönelik bir prosedür veya formül olarak adlandırılmaktadır. Örnekler arasında Gradient Descent, k-Ortalamalar (k-Means), SVM yer almaktadır. Modelin nasıl eğitildiğini belirlemektedir.

Değerlendirme metrikleri: Modelin performansını ölçmek için kullanılan metriklerdir. Doğruluk (accuracy), kesinlik (precision), F1-skor (F1-score), duyarlılık (sensitivity / recall), özgüllük (specificity) ve geometrik ortalama (G-mean) gibi metrikler kullanılmaktadır. Bu metrikler,

modelin ne kadar iyi genelleme yaptığı hakkında bilgi vermektedir. Söz konusu metrikler, aşağıdaki temel sınıflandırma parametrelerine dayanır:

- **TP (True Positive):** Gerçek pozitif sınıfın doğru şekilde tahmin edilmesi olarak bilinmektedir.
- **TN (True Negative):** Gerçek negatif sınıfın doğru şekilde tahmin edilmesi olarak bilinmektedir.
- **FP (False Positive):** Gerçekte negatif olan bir örneğin yanlış şekilde pozitif olarak tahmin edilmesi olarak bilinmektedir.
- **FN (False Negative):** Gerçekte pozitif olan bir örneğin yanlış şekilde negatif olarak tahmin edilmesi olarak bilinmektedir.

Doğruluk, kesinlik, F1-skor, duyarlılık, özgüllük ve geometrik ortalama ile ilgili bilgiler bölüm 6'da detaylı bir şekilde açıklanmaktadır.

4.1.3. Makine Öğrenmesi Uygulamaları

Makine öğreniminin endüstrileri ve günlük yaşamımızı dönüştürme potansiyeline dair birçok önemli örnek bulunmaktadır. Makine öğrenimi gelişmeye devam ettikçe, önümüzdeki yıllarda daha da yenilikçi ve etkili uygulamalar görülmesi beklenmektedir. Bu bölümde bunlarla ilgili örneklerle yer verilmektedir [117].

Sağlık

Hastalık teşhisi için tıbbi görüntülerden, hasta verilerinden ve genetik bilgilerden kanser, diyabet ve kalp hastalığı gibi hastalıkları belirlemektedir. İlaç keşfi, çok miktarda kimyasal ve biyolojik veriyi analiz ederek yeni ilaçlar bulma sürecini hızlandırmaktadır. Kişiselleştirilmiş tedavi planları, Tedavi planlarını genetik yapılarına, tıbbi geçmişlerine ve yaşam tarzlarına göre bireysel hastalara göre uyarlamaktadır.

Finans

Dolandırıcılık tespiti, müşterileri ve finansal kurumları korumak için hileli işlemleri gerçek zamanlı olarak tespit etmektedir. Borsa tahmini, yatırım kararlarını bilgilendirmek için hisse senedi fiyatlarını ve piyasa eğilimlerini tahmin etmektedir. Risk değerlendirmesi, kredibilitenin değerlendirilmesi ve yatırım risklerinin değerlendirilme işlemi yapılabilmektedir.

Perakende

Tavsiye sistemleri, müşterilere geçmiş davranışlarına ve tercihlerine göre ürün ve hizmet önermektedir. Talep tahmini, envanteri ve fiyatlandırmayı optimize etmek için ürünlere yönelik

müşteri talebini tahmin etmektedir. Kişiselleştirilmiş pazarlama, pazarlama kampanyalarını belirli müşteri segmentlerine hedeflemektedir.

Otonom araçlar

Nesne tespitinde yayalar, araçlar ve trafik işaretleri gibi nesneleri belirleme ve izleme işlemlerinde kullanılmaktadır. Navigasyon, karmaşık ortamlarda güvenli ve verimli rotalar planlamak ve yürütmek için kullanılmaktadır. Karar verme, dinamik sürüş durumlarında gerçek zamanlı kararlar almaktadır.

Doğal dil işleme

Sohbet robotları, insanlarla doğal dilde etkileşime girebilen konuşma araçları oluşturmaktadır. Duygu analizi, metin verilerinde ifade edilen duygusal tonu ve hissiyatı belirlemektedir. Makine çevirisi, metni bir dilden diğerine doğru ve verimli bir şekilde çevirebilmektedir [118].

4.1.4. Makine Öğrenmesi Çalışma Mantığı

Veri toplama, ön işleme, özellik mühendisliği, model seçimi, eğitim, test ve dağıtımdan oluşan yinelemeli bir süreç vardır. Bunlar makine öğrenimi iş akışının özünü oluşturmaktadır. Uygulayıcılar, her bir adımı dikkatlice uygulayarak çok çeşitli gerçek dünya zorluklarını ele alan etkili ve sağlam makine öğrenimi modelleri geliştirebilmektedirler.

Veri Toplama

Bu önemli ilk aşama, sorun alanını doğru bir şekilde yansıtan ilgili ve yüksek kaliteli verilerin toplanmasını içermektedir. Veri kaynakları, veri tabanlarından gelen yapılandırılmış veriler, metin ve görüntüler gibi yapılandırılmamış veriler ve çeşitli kaynaklardan gelen sensör verilerini kapsayacak şekilde geniş bir çeşitlilik gösterebilmektedir. Toplanan verilerin kalitesi, sonraki makine öğrenimi modelinin performansını önemli ölçüde etkilemektedir.

Veri İşleme

Ham veriler genellikle tutarsızlıklar, gürültü ve ilgisiz bilgiler içermektedir. Veri ön işleme, verileri analize hazırlamak için bir dizi dönüşüm gerekmektedir. Temel adımlar şunları içermektedir:

- Veri temizleme, eksik değerlerin işlenmesi, aykırı değerlerin giderilmesi ve tutarsızlıkların düzeltilmesi gerekmektedir.
- Veri dönüşümü, ölçeklendirme, normalleştirme ve kategorik değişkenleri sayısal gösterimlere kodlama şeklinde gerçekleştirilmektedir.

- Veri indirgeme, boyutsallık azaltma gibi teknikler, temel bilgileri korurken özellik sayısını azaltmak için kullanılmaktadır.

Özellik Mühendisliği

Bu kritik adım, ham verilerden ilgili özelliklerin tanımlanmasını ve çıkarılmasını içermektedir. Özellikler, makine öğrenimi modeli için girdi olarak kullanılan verilerin belirli nitelikleri veya özellikleridir. Etkili özellik mühendisliği, modelin kalıpları öğrenme ve doğru tahminler yapma becerisini geliştirerek model performansını önemli ölçüde artırabilmektedir. Tez dönemi boyunca üretilen makalelerde en çok fayda sağlanan durumlardan biri olarak öne çıkmaktadır.

Model Seçimi

Uygun makine öğrenimi algoritmasının veya modelinin seçilmesi, belirli sorunun türüne, verilerin niteliğine ve istenen sonuca bağlıdır. Yaygın modeller şunları içermektedir.

- **Denetimli öğrenme:** Doğrusal regresyon, lojistik regresyon, destek vektör makineleri, karar ağaçları, rastgele ormanlar, sinir ağları vb. algoritmaları içermektedir. Etiketli veri ile çalışmaktadır.
- **Denetimsiz öğrenme:** K-ortalamlar kümelemesi, hiyerarşik kümeleme, temel bileşen analizi (PCA) vb. yöntemlerle verideki gizli bağlantılar keşfedilmektedir.
- **Yarı denetimli öğrenme:** Öz eğitim, çift eğitim, grafik tabanlı öğrenme, üretici modeller vb. gibi yaklaşımlar çok miktarda etiketsiz ve az miktarda etiketli verilerle birlikte değerlendirilmektedir.
- **Takviyeli öğrenme:** Q-öğrenme, derin Q-ağları vb. gibi yöntemlerle, ajanların ödül maksimizasyonu odaklı öğrenmesi sağlanmaktadır.

Model Eğitimi

Eğitim aşaması, önceden işlenmiş verilerin seçilen modele beslenmesini içerir. Model, yinelemeli bir süreç aracılığıyla verilerin altında yatan kalıpları ve ilişkileri öğrenmektedir. Amaç, modelin tahminleri ile eğitim verilerindeki gerçek değerler arasındaki farkı en aza indirmektir.

Modelin Test Edilmesi

Eğitimden sonra, modelin performansı test seti adı verilen ayrı bir veri kümesi üzerinde değerlendirilmektedir. Bu adım, modelin görülmeyen verilere genelleme yeteneğini değerlendirir ve performansının tarafsız bir tahminini sağlamaktadır. Temel değerlendirme ölçütleri arasında doğruluk, kesinlik, F1 skoru, duyarlılık (geri çağırma), geometrik ortalama, ortalama karesel hata (MSE) ve diğerleri yer almaktadır [119].

Model Dağıtımı

Tatmin edici bir performans seviyesine ulaşıldığında, eğitilen model gerçek dünya ortamına dağıtılmaktadır. Bu durum, modelin bir yazılım uygulamasına entegre edilmesini, bir sunucuya yerleştirilmesini veya bir cihaza gömülmesini içerebilmektedir. Konuşlandırılan model daha sonra yeni girdi verilerine dayalı tahminler veya kararlar almak için kullanılabilir.

4.1.5. Makine Öğrenmesinin Zorlukları

Bu zorluklar, makine öğrenimi sistemlerinin sorumlu ve etkili bir şekilde geliştirilmesini ve uygulanmasını sağlamak için dikkatli bir değerlendirme ve yenilikçi çözümler gerektirmektedir.

Veri Kalitesi ve Miktarı

Veri kıtlığı, yetersiz veri, bir modelin sağlam kalıpları öğrenme ve görülmeyen verilere etkili bir şekilde genelleme yapma becerisini engelleyebilmektedir. Sınırlı veri, modelin eğitim verilerinde iyi performans gösterdiği ancak yeni, görünmeyen örneklerde kötü performans gösterdiği aşırı uyuma yol açabilmektedir.

Veri kalitesi sorunları, gürültülü veriler, eksik değerler ve tutarsızlıklar model performansını önemli ölçüde düşürebilmektedir. Yanlış veya güvenilir olmayan veriler önyargılara yol açabilmekte ve öğrenme sürecini yanlış yönlendirebilmektedir.

Veri yanlılığı, veriler genellikle gerçek dünyadaki mevcut önyargıları yansıtır ve bunlar makine öğrenimi modeline istemeden dâhil edilebilmektedir. Bu durum, özellikle hassas uygulamalarda ayrımcı veya adil olmayan sonuçlara yol açabilmektedir.

Aşırı uyum ve yetersiz uyum

Aşırı uyumda bir model aşırı karmaşık olduğunda ve eğitim verilerini gürültü ve alakasız detaylar da dâhil olmak üzere çok iyi öğrendiğinde ortaya çıkmaktadır. Model, eğitim örneklerini esasen “ezberlediği (overfitting)” için bu durum yeni verilere zayıf genelleme ile sonuçlanmaktadır. Yetersiz uyumda, bir model verilerin altında yatan örüntüleri yakalayamayacak kadar basit olduğunda ortaya çıkmaktadır. Bu durum hem eğitim hem de test verilerinde yüksek yanlılığa ve düşük performansa yol açmaktadır.

Model karmaşıklığını dengeleme, model karmaşıklığı ve genelleme yeteneği arasında doğru dengeyi bulmak çok önemlidir. Düzenli hale getirme (normalizasyon), çapraz doğrulama (cross validation, örn. 5 CV, 10 CV.) ve erken durdurma (early stopping) gibi teknikler aşırı uyumu azaltmak için kullanılmaktadır.

Önyargı ve Tarafsızlık

Algoritmik önyargı, makine öğrenimi modellerinin eğitim verilerinde mevcut olan önyargıları farkında olmadan devam ettirebilmesi durumunu ifade etmektedir. Bu durum, kredi başvuruları, işe alım süreçleri ve ceza adaleti gibi alanlarda ayrımcı sonuçlara neden olabilmektedir. Adalet ve eşitlik ilkeleri, makine öğrenimi sistemlerinin tüm bireyler ve gruplar için adil ve hakkaniyetli olmasını sağlamak açısından temel bir etik zorunluluktur. Önyargının azaltılması için adil temsil, önyargı tespiti ve önyargı giderme algoritmaları gibi teknikler geliştirilmekte ve uygulanmaktadır.

Ölçeklenebilirlik

Büyük veri kümelerinin işlenmesi, birçok gerçek dünya uygulaması devasa veri kümeleri içermektedir. Makine öğrenimi modellerinin bu kadar büyük veri kümeleri üzerinde eğitilmesi ve kullanılması hesaplama açısından pahalı ve zaman alıcı olabilmektedir.

Verimli algoritmalar, büyük hacimli verileri verimli bir şekilde işleyebilen ve bunlardan öğrenebilen ölçeklenebilir algoritmalar geliştirmek birçok uygulama için oldukça önemlidir. Dağıtık hesaplama, dağıtık hesaplama ve paralel işleme gibi teknikler, hesaplama iş yükünü birden fazla makineye dağıtmak için kullanılır ve modellerin büyük veri kümeleri üzerinde eğitilmesini sağlamaktadır.

4.1.6. Makine Öğrenmesinin Önemi

Makine öğrenimi, bilgisayarları dinamik ve karmaşık ortamlardaki karmaşık sorunları açık, kural tabanlı programlamaya ihtiyaç duymadan çözme konusunda güçlendirerek dünyamızda devrim yaratmaktadır. Bu paradigma değişimi, makinelerin verilerden öğrenmesine, kalıpları tanımlamasına ve akıllı kararlar almasına, yeni durumlara uyum sağlamasına ve zaman içinde performanslarını artırmalarına olanak tanımaktadır.

4.2. Derin Öğrenme

Makine öğreniminin bir alt alanı olan derin öğrenme, karmaşık veri işleme ve bilgi edinmeyi kolaylaştırmak için çok katmanlı yapay sinir ağlarından yararlanmaktadır. Bu metodoloji, karmaşık kalıpları ayırt etmek, bilinçli kararlar almak ve deneyimsel öğrenme yoluyla performansı yinelemeli olarak artırmak için hesaplama sistemlerini güçlendirir. Derin öğrenmenin ayırt edici bir özelliği, geleneksel makine öğrenimi paradigmalarının aksine, ham verilerden otomatik özellik çıkarma kapasitesidir, böylece manuel özellik mühendisliğine olan bağımlılığı azaltmaktadır [120].

4.2.1. Derin Öğrenmenin Türleri

- **Denetimli derin öğrenme:** Bu paradigma, modelleri eğitmek için etiketli veri kümelerini kullanır; örnek olarak görüntü sınıflandırma görevlerinde CNN'ler gösterilebilir; burada model, girdi görüntülerini önceden tanımlanmış kategorik etiketlerle ilişkilendirmeyi öğrenmektedir.
- **Denetimsiz derin öğrenme:** Bu yaklaşım, genellikle kümeleme ve boyut azaltma gibi görevler için kullanılan ve gizli temsillerin keşfedilmesini sağlayan etiketsiz verilerdeki doğal yapıların ve kalıpların tanımlanmasına odaklanmaktadır.
- **Takviyeli derin öğrenme:** Bu paradigma, karar verme politikalarını optimize etmek için ödül ve ceza sinyallerini kullanarak bir ortamla yinelemeli bir etkileşim süreci yoluyla modelleri eğitmektedir. Uygulamalar, ajanların deneme-yanılma yoluyla en uygun eylemleri öğrendiği robotik ve stratejik oyun gibi alanlarda yaygındır.

4.2.2. Derin Öğrenmenin Bileşenleri

Yapay sinir ağları (YSA'lar): Bunlar, verileri ağ boyunca dönüştürmek ve yaymak için tasarlanmış birbirine bağlı yapay nöron katmanlarından oluşan çekirdek işlem birimlerini oluşturmaktadır [121].

Aktivasyon fonksiyonları: Rectified Linear Unit (ReLU), Sigmoid ve Softmax gibi bu doğrusal olmayan fonksiyonlar, ağın gerçek dünya verilerinde bulunan karmaşık, doğrusal olmayan ilişkileri modellemesini sağlamanın ayrılmaz bir parçasıdır.

Kayıp fonksiyonları: Bu fonksiyonlar, modelin tahmin edilen çıktıları ile gerçek hedef değerler arasındaki tutarsızlığı ölçerek model performansını değerlendirmek ve optimizasyon sürecine rehberlik etmek için bir metrik sağlamaktadır.

Optimizasyon algoritmaları: Stokastik Gradyan İnişi (SGD) ve Adam optimizasyonu dâhil olmak üzere algoritmalar, kayıp fonksiyonunu en aza indirmeyi ve model performansını optimize etmeyi amaçlayarak modelin ağırlıklarını ve önyargılarını iteratif olarak ayarlamak için kullanılır.

Eğitim veri kümeleri: Büyük ölçekli, temsili veri kümeleri, derin öğrenme modeli tarafından anlamlı örüntülerin ve sağlam genellemelerin öğrenilmesini kolaylaştırmak için vazgeçilmezdir.

4.2.3. Derin Öğrenmenin Uygulama Alanları

Derin öğrenmenin endüstriyel alanların yanı sıra günlük yaşam pratiklerinde kullanılmasına dair birçok örnek mevcuttur.

Bilgisayarla Görme

Bu alan, görsel veri analizi için derin öğrenmeden yararlanarak görüntü sınıflandırma, nesne algılama ve yüz tanıma gibi görevleri kapsamaktadır.

Doğal Dil İşleme (NLP)

Derin öğrenme, konuşma araçları (chatbotlar), konuşmadan metne dönüştürme ve makine çevirisi gibi uygulamalardaki gelişmeleri destekleyerek gelişmiş dil anlama ve üretme yetenekleri sağlamaktadır [118].

Sağlık Hizmetleri

Derin öğrenme tıbbi teşhis, ilaç keşfi ve kişiselleştirilmiş tedavi önerilerinin geliştirilmesinde önemli bir rol oynayarak tıp bilimindeki ilerlemelere katkıda bulunmaktadır.

Otonom Sistemler

Derin öğrenme, otonom araçların, robotik sistemlerin ve drone navigasyonunun geliştirilmesinin temelini oluşturarak dinamik ortamlarda akıllı karar vermeyi sağlamaktadır.

Finansal Teknoloji (FinTech)

Bu sektördeki uygulamalar arasında dolandırıcılık tespiti, finansal tahmin ve algoritmik ticaret yer almakta ve veri odaklı finansal analiz ve karar verme için derin öğrenmeden yararlanılmaktadır.

4.2.4. Derin Öğrenmenin Çalışma Mantığı

Derin öğrenme modelleri, girdi verilerini bir dizi birbirine bağlı yapay nöron katmanı aracılığıyla işleyerek çalışmaktadır. Bu hiyerarşik işleme, ilk katmanların ilkel kalıpları ayırt ettiği ve sonraki, daha derin katmanların giderek daha karmaşık temsiller oluşturduğu aşamalı olarak soyut özelliklerin çıkarılmasını sağlamaktadır. Öğrenme süreci, geri yayılım ve optimizasyon algoritmaları aracılığıyla elde edilen dâhili model parametrelerinin yinelemeli olarak ayarlanmasıyla kolaylaştırılmıştır. Başarılı bir eğitimin ardından, bu modeller daha önce görülmemiş veriler üzerinde doğru tahminler üretme kapasitesini sergilemektedir.

4.2.5. Derin Öğrenmenin Zorlukları

Hesaplama yoğunluğu: Büyük ölçekli derin öğrenme modellerinin eğitimi, genellikle Grafik İşleme Birimleri (GPU'lar) veya TPU'lar gibi özel donanımlar gerektiren önemli hesaplama kaynakları gerektirmektedir.

Veri bağımlılığı: Etkili model eğitimi, kapsamlı, etiketli veri kümelerinin mevcudiyetine bağlıdır ve bu da verilerin az olduğu alanlarda sınırlayıcı bir faktör olabilmektedir.

Aşırı uyum olgusu: Yaygın bir zorluk, modellerin eğitim verilerini ezberleyerek yeni veriler üzerinde zayıf genelleme performansına yol açtığı aşırı uyumdur.

Sınırlı model yorumlanabilirliği: Derin öğrenme modelleri genellikle “kara kutu” doğaları ile karakterize edilir ve karar verme süreçleri için şeffaf ve yorumlanabilir açıklamalar sağlama yeteneğini engellemektedir.

Eğitim kaynak talepleri: Karmaşık derin öğrenme mimarilerinin eğitimi önemli ölçüde zaman ve hesaplama kaynağı gerektirmekte, bu da geliştirme döngüleri üzerinde pratik bir sınırlama oluşturmaktadır.

4.2.6. Derin Öğrenmenin Önemi

Otomasyonun artırılması: Derin öğrenme metodolojileri, özellikle görüntü ve konuşma tanıma gibi alanlarda karmaşık görevlerin otomasyonunu kolaylaştırarak insan müdahalesi gerekliliğini en aza indirmektedir.

Üstün tahmin doğruluğu: Büyük ölçekli, karmaşık problem alanlarında, derin öğrenme modelleri geleneksel makine öğrenimi paradigmalarına kıyasla genellikle üstün tahmin doğruluğu sergilemektedir.

Alanlar arası ölçeklenebilirlik: Derin öğrenmenin uyarlanabilirliği, sağlık hizmetleri, otonom sistemler ve finansal teknoloji dâhil olmak üzere geniş bir sektör yelpazesinde uygulanmasına olanak tanıyarak sağlam ölçeklenebilirliğini göstermektedir.

Teknolojik inovasyon için katalizör: Derin öğrenme, yapay zekâ alanındaki ilerlemelerin kritik bir itici gücü olarak çeşitli bilimsel ve teknolojik alanlarda atılımları teşvik etmektedir.

4.3. Makine Öğrenmesi ve Derin Öğrenmede Kullanılan Yöntemler

Bu bölümde tez ilerleme sürecinde gerçekleştirilen çalışmalarda kullanılan teknikler, yöntemler ve özellikler açıklanmıştır.

4.3.1. Örüntü Tanıma ve Özellik Çıkarımı

- **BP:** İkili eşiklemeye dayalı görüntü/sinyal tanımlayıcıdır. Yerel bilgilerin, önceden tanımlanmış bir eşik veya merkezi bir piksel değerine göre karşılaştırma işlemi yoluyla oluşturulan ikili bir gösterime kodlanmasıyla ilgilidir [122]. Örneğin merkez pikselin değeri I_c ve komşu piksellerin değeri de I_p olarak verilmiştir. Matematiksel ifadesi Denklem 4.1 ve Denklem 4.2’de gösterilmiştir.

$$BP_p = \begin{cases} 1 & \text{eğer } I_p \geq I_c \\ 0 & \text{eğer } I_p < I_c \end{cases} \quad (4.1)$$

Burada tüm ikili değerler tek bir ikili sayıya birleştirilmektedir.

$$BP = \sum_{p=0}^{P-1} BP_p \cdot 2^p \quad (4.2)$$

- **LBP:** Her bir komşu pikseli merkezi pikselle karşılaştırarak ve sonucu ikili bir değere kodlayarak çalışan ikili örüntüler alanında daha yapılandırılmış bir yaklaşımı temsil etmektedir [123]. Yerel ikili örüntü olarak bilinmektedir. Matematiksel ifadesi Denklem 4.3 ve Denklem 4.4’de gösterilmektedir.

$$s(x) = \begin{cases} 1 & \text{eğer } x \geq 0 \\ 0 & \text{eğer } x < 0 \end{cases} \quad (4.3)$$

Burada merkez pikselin (I_c) etrafında 3x3’lük bir bölge verilmektedir.

$$LBP(x, y) = \sum_{p=0}^{P-1} s(i_p - i_c) \cdot 2^p \quad (4.4)$$

Burada yerel örüntüyü tanımlayan bir P-bit ikili sayı oluşturulmaktadır.

- **BTP:** LBP’nin iki durumunun aksine, temsilde üç durumun kullanılmasıyla elde edilen daha büyük bir yerel varyasyon derecesinin kodlanmasını sağlayarak LBP’yi geliştirir [124]. Örneğin merkez pikselin değeri I_c ve diğer iki eşik değeri de T_{low} ve T_{high} olarak verilmiştir. Matematiksel ifadesi Denklem 4.5’de gösterilmektedir.

$$BTP_p = \begin{cases} 1 & \text{eğer } I_p \geq I_c + T_{high} \\ 0 & \text{eğer } |I_p - I_c| < T_{low} \\ -1 & \text{eğer } I_p < I_c - T_{high} \end{cases} \quad (4.5)$$

Burada genellikle iki ikili dizeye bölünen üçlü bir dize (-1, 0, 1 değerleriyle) elde edilmektedir. Pozitif geçişler için UT (Üst Üçlü) kullanılırken Negatif geçişler için LT (Alt Üçlü) kullanılmaktadır.

- **FF-BTP:** BTP’nin İleri-İleri (FF) öğrenme paradigmasını (ilk olarak Geoffrey Hinton tarafından 2022’de önerilmiştir) entegre eden geliştirilmiş bir versiyonunu temsil etmektedir [125]. Bu yöntem, üçlü örüntü kodlamasını yalnızca ileri adımlarla karakterize edilen bir nöral eğitim süreciyle birleştirmektedir. Matematiksel ifade Denklem 4.6’da gösterilmektedir.

$$z = \sum_{i=1}^n w_i \cdot x_i + b \quad \{a = f(z)\} \quad (4.6)$$

Burada f genellikle ReLU ya da Sigmoid olarak bilinmektedir.

- **LT:** Dalgacık dönüşümü uygulandıktan sonra elde edilen düşük frekans (L) katsayıları olarak tanımlanmaktadır. Eşikleme işlemleri Denklem 4.7’de gösterilmiştir.

$$LT_p = \begin{cases} 1 & \text{eğer } BTP_p = -1 \\ 0 & \text{diğer durumda} \end{cases} \quad (4.7)$$

- **UT:** Dalgacık dönüşümü uygulandıktan sonra elde edilen yüksek frekans (H) katsayıları olarak tanımlanmaktadır. Eşikleme işlemleri Denklem 4.8’de gösterilmiştir.

$$UT_p = \begin{cases} 1 & \text{eğer } BTP_p = +1 \\ 0 & \text{diğer durumda} \end{cases} \quad (4.8)$$

Bunlar daha sonra özellik çıkarma veya model eğitiminde kullanılmak üzere ikili değerler olarak kodlanır.

- **MFCC:** Ses verisinden elde edilen spektral özellikler olarak tanımlanmaktadır. Log-filtre bankası enerjilerinin DCT’si kullanılarak türetilmektedir. Önce sinyal üst üste binen pencerelere bölünür. Daha sonra hızlı fourier dönüşümü (FFT) ve Mel filtre bankası uygulanmaktadır. Bu adımdan sonra çıktıların log enerjisi (M_m) alınarak DCT uygulanmaktadır. Bu işleme ait matematiksel ifade Denklem 4.9’da verilmiştir.

$$MFFC_n = \sum_{m=1}^M \log(M_m) \cdot \cos\left(\frac{\pi n(m-0.5)}{M}\right) \quad (4.9)$$

- **DWT:** Dalgacıkları kullanarak bir sinyali yaklaşık (düşük frekans) ve ayrıntı (yüksek frekans) bileşenlerine ayırmaktadır [126]. Düşük frekansa ait matematiksel ifade Denklem 4.10, yüksek frekansa ait matematiksel ifade Denklem 4.11’de gösterilmektedir.

$$A[n] = \sum_k x[k] \cdot \varphi_{j,k}(n) \quad (4.10)$$

$$D[n] = \sum_k x[k] \cdot \psi_{j,k}(n) \quad (4.11)$$

Burada φ ve ψ , j ölçeğindeki ölçekleme (baba) ve dalgacık (ana) fonksiyonlarıdır.

- **MDWT:** Bu, dönüşümü birden fazla seviye için yaklaşım katsayılarına yinelemeli olarak uygulayarak DWT’yi genişletmektedir [122, 127]. Bu yöntemin matematiksel ifadesi Denklem 4.12’de gösterilmektedir.

$$x[n] \xrightarrow{DWT} A_1[n], D_1[n] \quad A_1[n] \xrightarrow{DWT} A_2[n], D_2[n] \dots A_L[n], D_1[n], D_2[n], \dots, D_L[n] \quad (4.12)$$

- **MFMDWT:** Farklı seviyelerde birden fazla filtre bankasına sahip DWT uygulaması olarak tanımlanmaktadır. MDWT’nin bu varyasyonu, çeşitli zaman-frekans özelliklerini

yakalamak için her ayrıştırma seviyesinde birden fazla dalgacık filtresi uygulamaktadır [128]. Bu yöntemle ait matematiksel ifade Denklem 4.13’de gösterilmektedir.

$$x[n] \xrightarrow{\psi_1} A_1^{(1)}, D_1^{(1)} \quad x[n] \xrightarrow{\psi_2} A_1^{(2)}, D_1^{(2)} \quad (4.13)$$

Daha sonra, farklı filtrelerdeki tüm özellikler daha zengin gösterimler için birleştirilmektedir.

- **UTMDWT:** Standart DWT'nin (dengeli bir ikili ağaç oluşturan) aksine, UTMDWT DWT'yi seçici olarak uygular, örneğin, yalnızca entropi veya enerjiye bağlı olarak ayrıştırma veya yaklaşıklık katsayıları üzerinde gerçekleştirilmektedir [129]. Bu işlemin matematiksel ifadesi Denklem 4.14’de gösterilmektedir.

$$\textbf{Level 1: } x[n] \xrightarrow{DWT} A_1, D_1 \quad \textbf{Level 2: } D_1 \xrightarrow{DWT} A_2, D_2 \quad (A_1' \text{ de özyineleme yok.}) \quad (4.14)$$

Yüksek frekanslı veya düşük frekanslı bileşenleri uyarlamalı olarak vurgulamak için özellik çıkarmada kullanılır.

- **TQWT:** Titreşim davranışını ayarlayan ayarlanabilir bir Q faktörüne sahip yedekli bir dönüşümdür [130]. Değişen rezonanslı sinyalleri analiz etmek için kullanışlıdır (örn. EKG, solunum sesleri). Sinyal, ayarlanabilir parametrelere dayalı yinelemeli filtreleme kullanılarak alt bantlara ayrıştırılır. $H_0(w)$ düşük geçiş, $H_1(w)$ yüksek geçiş olarak adlandırılmaktadır. Bu durumun matematiksel ifadeleri Denklem 4.15, Denklem 4.16, Denklem 4.17 ve Denklem 4.18’de gösterilmektedir.

$$\text{Yüksek geçiş ölçeği: } \beta = \frac{2}{r+1} \quad (4.15)$$

$$\text{Düşük geçiş ölçeği: } \alpha = 1 - \frac{\beta}{Q} \quad (4.16)$$

$$J \text{ ölçeğinde dalgacık: } \psi_j(w) = H_1(\alpha^{j-1}w) \prod_{k=0}^{j-2} H_0(\alpha^k w) \quad (4.17)$$

$$\text{Ölçekleme fonksiyonu: } \phi_j(w) = \prod_{k=0}^{j-1} H_0(\alpha^k w) \quad (4.18)$$

Burada görüldüğü gibi Q: kalite faktörünü ifade etmekle birlikte hem geçici sinyaller (düşük Q) hem de salınımlı sinyaller (yüksek Q) için uygundur. r: yedeklilik ifade eder ve gürültü sağlamlılığına yardımcı olmaktadır. J: seviye sayısı olmak üzere, FFT tabanlı filtre bankaları kullanılmaktadır. TQWT ortogonal değildir ve üst üste binen dalgacık paketleri içermektedir.

4.3.3. Optimizasyon ve Kayıp Fonksiyonları

- **Normalizasyon:** Normalleştirme, verileri genellikle (0, 1) olmak üzere sabit bir aralığa dönüştürür. Matematiksel ifadesi Denklem 4.19'da verilmiştir.

$$x_{norm} = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (4.19)$$

Burada x orijinal özellik değeri, x_{max} ve x_{min} özelliğin minimum ve maksimum değerleridir.

- **Stokastik gradyan inişi (SGD):** Her eğitim örneğine göre kayıp fonksiyonunun eğimini kullanarak model parametrelerini güncelleyen basit bir optimizasyon yöntemidir [131]. Matematiksel ifadesi Denklem 4.20'de verilmiştir.

$$\theta_{t+1} = \theta_t - n \nabla_{\theta} L(\theta_t; x_i, y_i) \quad (4.20)$$

- **Adam optimizasyonu:** Adam hem momentum hem de uyarlanabilir öğrenme oranlarını kullanmaktadır. Önce eğimler hesaplanarak ön yargılı birinci ve ikinci moment tahminleri güncellenir. Sonra ön yargı düzeltilir ve parametreler güncellenir [132].
- **Bayes optimizasyonu:** Bayes Optimizasyonu, genellikle pahalı kara kutu fonksiyonlarını optimize etmek için kullanılan olasılıksal model tabanlı küresel bir optimizasyon yöntemidir [133]. Matematiksel olarak Denklem 4.21'de gösterilmektedir.

$$EI(x) = E[\max(0, f(x) - f(x^+))] = (\mu(x) - f(x^+))\Phi(z) + \sigma(x)\phi(z) \quad (4.21)$$

Burada $\mu(x)$ ve $\sigma(x)$: Gauss (GP) işleminden elde edilen standart sapma; $f(x^+)$: gözlenen en iyi değer; Φ ve ϕ : standart normal dağılımın kümülatif dağılım fonksiyonu (CDF) ve olasılık yoğunluk fonksiyonu (PDF) olarak tanımlanmaktadır.

- **Sigmoid fonksiyonu:** Sigmoid (veya lojistik) fonksiyonu, gerçek değerli herhangi bir sayıyı (0, 1) aralığına eşlemek için yaygın olarak kullanılır ve bu da onu ikili sınıflandırma için kullanışlı hale getirmektedir [134]. Matematiksel ifadesi Denklem 4.22'de gösterilmektedir.

$$\sigma(z) = \frac{1}{1 + e^{-z}} \quad (4.22)$$

Burada z giriş değeri, e ise Euler sayısıdır (~ 2.718).

- **Signum fonksiyonu:** Signum (veya Sign) fonksiyonu gerçek bir sayının işaretini belirlemek için kullanılmaktadır [135]. Optimizasyonda veya seyrek sinir ağlarında kullanılmaktadır.

$$Sign(x) = \begin{cases} 1 & \text{eğer } x > 0 \\ 0 & \text{eğer } x = 0 \\ -1 & \text{eğer } x < 0 \end{cases} \quad (4.23)$$

Genellikle büyüklüğü değil sadece yönü dikkate almak için gradyan güncellemelerinde kullanılmaktadır.

- **Kayıp fonksiyonu:** Tahmin edilen değer ile gerçek değer (etiket) arasındaki farkı ölçmektedir. Sigmoid çıktısı $y' = \sigma(z)$ olan ikili sınıflandırma için, ikili çapraz entropi kaybı (log kaybı) ya da MSE buna örnek olarak gösterilebilmektedir [136]. Matematiksel olarak Denklem 4.24'deki gibi ifade edilmektedir.

$$L(y, y') = [y \log(y') + (1 - y) \log(1 - y')] \quad (4.24)$$

4.3.4. Model Değerlendirme ve Özellik Seçimi

- **LORO CV:** Bu yöntem, Leave-One-Out Cross-Validation (LOOCV) yönteminin bir çeşididir ancak bireysel örnekler yerine çalışmalara/oturumlara uygulanmaktadır [23, 137]. R çalışması olan bir veri kümesi verildiğinde, her yinelemede test için bir tam çalışma kullanılır ve kalan $R-1$ çalışması eğitim için kullanılmaktadır. Açık bir formülü yoktur ancak matematiksel olarak Denklem 4.25'deki gibi ifade edilebilmektedir.

$$Accuracy_{LORO} = \frac{1}{R} \sum_{i=1}^R Accuracy_{test_run} \quad (4.25)$$

- **10-Fold CV:** Veriler 10 eşit (veya neredeyse eşit) kata bölünmektedir. Her yinelemede $i \in \{1, \dots, 10\}$ i katını test kümesi olarak kullanılmaktadır. Kalan 9 katı eğitim kümesi olarak kullanılmaktadır [138]. Bu işlemin matematiksel ifadesi Denklem 4.26'da gösterilmektedir.

$$CV_{10} = \frac{1}{10} \sum_{i=1}^{10} Accuracy_i \quad (4.26)$$

- **Hold-Out validation:** Bu yöntemde veriler eğitim ve test olmak üzere iki bölüme ayrılmaktadır [139]. Örneğin V veri seti olarak kabul edildiğinde matematiksel ifadesi Denklem 4.27'de gösterilmiştir.

$$V = V_{train} \cup V_{test} \quad (4.27)$$

Kendi başına bir öğrenme formülü yoktur, ancak eğitim V_{train} üzerinde ve değerlendirme ise V_{test} üzerinde gerçekleştirilmektedir.

- **ReliefF:** Bu yöntem özellikleri sınıflar arasında ne kadar iyi ayırım yaptıklarına göre sıralayan filtre tabanlı bir özellik seçimi algoritmasıdır [140]. Örneğin f özellik vektörü için matematiksel ifadesi Denklem 4.28’de gösterilmektedir.

$$W[f] = W[f] - \frac{(x_f - nearestHit_f)^2}{m} + \frac{(x_f - nearestMiss_f)^2}{m} \quad (4.28)$$

Burada m : belirlenen örnek sayısını, x_f : x örneği için f özelliğinin değerini ifade etmektedir.

Çok sınıflı sorunları ve gürültülü verileri dikkate alan ve birden fazla komşu üzerinde ortalama alan geliştirilmiş bir sürümü de RReliefF olarak adlandırılmaktadır.

- **Ki-Kare:** Bu yöntem, özellikler ve sınıf etiketleri arasındaki bağımsızlığı test etmektedir [141]. Ki-kare (Chi^2), matematiksel olarak Denklem 4.29’daki gibi ifade edilmektedir.

$$Chi^2 = \sum_{i=1}^r \sum_{j=1}^c \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \quad (4.29)$$

Burada O_{ij} : gözlenen frekansı, E_{ij} : bağımsızlık altında beklenen frekansı, r : satır sayısını (özellikleri), c : sütun sayısını (sınıfları) ifade etmektedir.

- **Yinelemeli Ki-Kare:** Ki-Kare’nin iteratif bir çeşidi olarak bilinmektedir. Chi^2 testi veri setine bir çok kez uygulanarak gerçekleştirilmektedir. Her iterasyonda düşük puanlı özellikler kaldırılarak özellik seti güncellenmektedir. Özellikle yüksek boyutlu veri setlerinde sağlamlığı artırmaktadır. Kesin bir formülü yoktur ancak her adımda sıralama ve eşiklemeye dayalı bir döngüyle çalışmaktadır [142].
- **PCA:** Temel Bileşen Analizi (PCA), verilerdeki varyansı mümkün olduğunca korumak amacıyla boyut azaltma için kullanılan istatistiksel bir tekniktir. Orijinal değişkenleri, temel bileşenler olarak adlandırılan ve verilerde açıkladıkları varyans miktarına göre sıralanan yeni bir ortogonal (ilişkisiz) değişkenler kümesine dönüştürmektedir [143]. Örneğin en büyük varyans ilk ana bileşende, ikinci en büyük varyans ikincide yer alır. Sınıflandırılacak olan bir veri setinde $X \in R^{n \times f}$ olmak üzere n : örnek sayısı, f : özellik sayısı olarak ifade edilmektedir. İlk olarak verilerin Denklem 4.30’da gösterildiği gibi ortalaması alınır.

$$X_c = X - \mu, \mu = \frac{1}{n} \sum_{i=1}^n X_i \quad (4.30)$$

Bu adımdan sonra kovaryans matris hesaplanarak özdeğerler ve özvektörler ortaya çıkarılmaktadır. Bu işlemler Denklem 4.31 ve Denklem 4.32’de gösterilmektedir.

$$C = \frac{1}{n-1} X_c^T X_c \quad (4.31)$$

$$Cv_i = \lambda_i v_i, \quad i = 1, 2, \dots, d \quad (4.32)$$

Burada v_i ana bileşenleri, λ_i her bileşen tarafından açıklanan varyansı ifade etmektedir. PCA denetimsiz öğrenmenin bir parçasıdır ve sınıf etiketi kullanılmaz. Sınıfları ayırmak yerine varyansı en üst düzeye çıkarma eğilimindedir.

- **NCA:** Sınıflandırma performansını iyileştirmeyi amaçlayan bir mesafe metriği öğrenme yöntemidir. Genellikle kNN ile birlikte ve özellik seçim aşamasında kullanılmaktadır. Amacı her bir veri örneğini d boyuttan m boyuta indirgeyen bir lineer dönüşüm matrisi $A \in R^{m \times d}$ öğrenmektir [144].

Bir A verildiğinde, i noktasının j 'yi komşu olarak seçme olasılığı Denklem 4.33'da gösterilmektedir.

$$p_{ij} = \frac{\exp(-\|A_{xi} - A_{xj}\|^2)}{\sum_{k \neq i} \exp(-\|A_{xi} - A_{xk}\|^2)} \quad (4.33)$$

Bu çıktıdan beklenen bir tanesini dışarıda bırakma sınıflandırma doğruluğu (LOO CV) Denklem 4.34'de ifade edilmektedir.

$$f(A) = \sum_i \sum_{j \in \text{same class}} p_{ij} \quad (4.34)$$

Burada NCA, A 'yi $f(A)$ 'yi maksimize etmek için optimize etmektedir.

- **INCA:** Yinelemeli Komşuluk Bileşen Analizi (INCA), NCA yönteminin geliştirilmiş ve yinelemeli bir versiyonu olarak sunulmaktadır [145]. INCA, özellikle yüksek boyutlu veri kümelerinde en bilgilendirici özelliklerin seçilmesi ve sonuç olarak sınıflandırma performansının iyileştirilmesi amacıyla kullanılmaktadır. Yinelemeler arasında diğer özellik seçimi yöntemlerini (Chi², ReliefF gibi) entegre edebilmesi de muhtemeldir. Kesin bir formül yoktur ancak Denklem 4.35'deki gibi ifade edilebilmektedir.

$$A^{(t+1)} = \text{Update}(A^{(t)}, \text{gradient of } f(A)) \quad (4.35)$$

Burada $A^{(t)}$ t yinelemesindeki dönüşümdür. Bu yöntem, yinelemeli bir süreç aracılığıyla daha uygun özellik alt kümelerini belirlemeye çalışırken, aynı zamanda NCA tarafından sunulan doğal avantajları korumayı amaçlamaktadır.

IMV: Özellikle çok modlu veya çok kaynaklı veriler için tipik olarak sınıflandırma görevlerinde kullanılan topluluk tabanlı bir karar füzyon yöntemidir. Kararlı bir çoğunluk konsensüsüne ulaşılan kadar birden fazla sınıflandırıcı veya karar sisteminin çıktılarını değerlendirip birleştirerek nihai kararı yinelemeli olarak iyileştirmektedir [146]. Algoritma, yakınsama veya bir durdurma kriterine kadar çoğunluk oylamasını kullanarak

kararı yinelemeli olarak güncellemektedir. Her örnek x_i ve bir sınıflandırıcı kümesi için $C = \{c_1, c_2, \dots, c_K\}$ çoğunluk oyu matematiksel olarak Denklem 4.36'da gösterilmektedir.

$$y'_i = \text{mode} (\{c_1(x_i), c_2(x_i), \dots, c_K(x_i)\}) \quad (4.36)$$

Burada $c_k(x_i)$: k -inci sınıflandırıcının x_i örneği üzerindeki tahmini, $\text{mode}(\cdot)$: tahminler arasında en sık rastlanan sınıftır. Ancak IMV'de, bu çoğunluk oyu bir kez hesaplanmaz, bunun yerine Denklem 4.37'deki gibi iterasyon t olarak verilirse $y_i^{(t)}$, i örneğinin tahmin edilen etiketi olsun, Her bir c_k sınıflandırıcısı için $y_i^{(t)}$ ile uyuma göre ağırlıklar ve güvenilirlik ayarlanarak bir sonraki iterasyonda $(t + 1)$ tahminler yeniden değerlendirilmektedir.

$$y_i^{(t+1)} = \text{mode} \left(\sum_{k=1}^K w_k^{(t)} \cdot \delta(c_k(x_i) = y) \right) \quad (4.37)$$

4.3.5. Sınıflandırma

- **SVM:** Farklı sınıfları ayıran en uygun hiper düzlemi bulan bir gözetimli öğrenme modelidir. Matematiksel gösterimi Denklem 4.38'de verilmiştir.

$$\min_{w,b} \frac{1}{2} \|w\|^2 \quad \text{eğer} \quad y_i(w^T x_i + b) \geq 1 \quad (4.38)$$

Burada w : Ağırlık vektörü, b : Sapma terimi, x_i : Giriş özellik vektörü, $y_i \in \{-1, +1\}$: Sınıf etiketi olarak tanımlanmaktadır.

- **kNN:** “k” en yakın eğitim örneğini analiz ederek sınıflandırma ve regresyon için kullanılan parametrik olmayan bir yöntemdir. Temel amacı bir veri noktasının sınıfı veya değeri, ona en yakın k komşusunun çoğunluğuna veya ortalamasına göre belirlenir. Matematiksel gösteri Denklem 4.39'da verilmiştir.

$$\text{Test noktası} = x_{test} \quad (4.39)$$

Burada sınıflandırma aşamasında x_{test} 'in k en yakın komşusu bulunarak komşuların sınıfları sayılmaktadır. İndikatör fonksiyonu (I) ile komşu sınıfı c ise 1, değilse 0 olarak belirlenmektedir. Bu adımların matematiksel ifadeleri Denklem 4.40 ve Denklem 4.41'de gösterilmektedir.

$$\text{Öklid uzaklığı} = d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (4.40)$$

$$\text{Sınıf}(x) = \text{argmax}_c \sum_{i=1}^k I(y_i = c) \quad (4.41)$$

Regresyon aşamasında x_{test} 'in k en yakın komşusu bulunarak komşuların çıktı değerlerinin ortalaması alınmaktadır. Bu adımın matematiksel ifadeleri Denklem 4.41 ve Denklem 4.43'de gösterilmektedir.

$$Değer(x) = \frac{1}{k} \sum_{i=1}^k y_i \quad (4.41)$$

$$Eğitim kümesi = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\} \quad (4.43)$$

Tüm eğitim noktaları ile x_{test} arasındaki uzaklık hesaplanarak küçükten büyüğe doğru sıralanmaktadır. En küçük k tanesi seçilerek sınıflandırma için komşuların modu, regresyon için ise ortalaması alınmaktadır. Burada k değeri küçük olduğunda sınırlar daha karmaşık ve aşırı öğrenme riski (overfitting) oluşabilmektedir. Ancak k değeri büyük olduğunda sınırlar daha yumuşak olup yetersiz öğrenme (underfitting) riski oluşabilmektedir.

5. DİĞER YÖNTEMLER

Bu bölümde literatürde kullanılan diğer yöntemlerden bahsedilmektedir.

5.1. Ses Analizi için Derin Öğrenme

Derin öğrenme metodolojileri, ham veya önceden işlenmiş ses sinyallerinden bilgilendirici temsiller türetmek için sinir ağı mimarilerini kullanarak ses analizini önemli ölçüde dönüştürmüştür. Bu alanda, CNN'ler ve RNN'ler öne çıkan mimarilerdir ve her biri özellik çıkarma ve sıralı veri modellemede farklı işlevleri yerine getirmektedir.

5.1.1. Ses Spektrogramlarından Uzamsal Özellik Çıkarımı için Evrimsel Sinir Ağları

CNN'ler, görüntüler ve ses spektrogramları gibi yapılandırılmış veri temsillerinden uzamsal desenleri çıkarma konusunda mükemmeldir. Ses analizinde, ham dalga formu verileri genellikle Kısa Zamanlı Fourier Dönüşümü (STFT) veya MFCC'ler gibi tekniklerle spektrogramlara dönüştürülmektedir. Daha sonra, CNN'ler bu temsiller içindeki yerleştirilmiş desenleri ayırt etmek için evrimsel filtreler uygulamaktadır.

Ses Analizinde CNN'lerin Temel Özellikleri

- **Hiyerarşik özellik temsili:** Alt evrimsel katmanlar harmonik bileşenler gibi temel kalıpları yakalarken, daha derin katmanlar tınısal özellikler ve fonemik unsurlar dâhil olmak üzere daha karmaşık yapıları ayırt edebilmektedir.
- **Çeviri değişmezliği:** CNN'ler, spektrogram içindeki uzamsal konumlarına bakılmaksızın kalıpları tanıyarak perde ve süredeki değişikliklere karşı sağlamlık sergilemektedir.
- **Hesaplama verimliliği:** Konvolüsyonel katmanların doğasında bulunan ağırlık paylaşımı, tam bağlantılı sinir ağlarına kıyasla daha az hesaplama karmaşıklığı sağlamaktadır.

CNN ile Gerçekleştirilen Uygulamalar

CNN'ler, farklı frekans ve zamansal modellerin tanımlanmasının çok önemli olduğu müzik türü sınıflandırması, konuşma duygusu tanıma, çevresel ses sınıflandırması, müzik kaynağı ayırımı, çözünürlüğü artırma ve gürültü giderme gibi uygulamalarda yaygın olarak kullanılmaktadır.

5.1.2. Ses Dizilerindeki Zamansal Bağımlılıkları Modellemek için Tekrarlayan Sinir Ağları

CNN'ler uzamsal örüntüleri etkili bir şekilde çıkarırken, sıralı verilerdeki uzun menzilli zamansal bağımlılıkları yakalamada sınırlamalar sergilemektedirler. RNN'ler, özellikle de LSTM

ve GRU'lar, sıralı veri modelleme mekanizmalarını dâhil ederek bu zorluğun üstesinden gelmektedir.

Ses Analizinde RNN'lerin Temel Özellikleri

- **Hafıza tutma mekanizmaları:** İleri beslemeli sinir ağlarının aksine, RNN'ler gizli durum temsillerini koruyarak önceki zaman adımlarından gelen bilgileri koruyarak sıralı verilerin işlenmesini sağlamaktadır.
- **Zamansal değişkenlik modellemesi:** RNN'ler, zamansal alanları kapsayan bağlamsal bilgilerin kritik olduğu konuşma tanıma, akustik olay algılama ve müzik transkripsiyonu gibi uygulamalarda etkilidir.
- **Çift yönlü bağlamsal işleme:** Çift yönlü RNN'ler (Bi-RNN'ler) ve LSTM'ler, ağıın hem geçmiş hem de gelecek bağlamsal bilgileri dikkate almasını sağlayarak ses sınıflandırma görevlerinde doğruluğu artırmaktadır.

RNN ile Gerçekleştirilen Uygulamalar

RNN tabanlı bir model, konuşmadan metne dönüştürme sistemlerinde kullanılabilir; burada doğru fonem tanıma, hassas transkripsiyonu kolaylaştırmak için önceki ve sonraki fonemlerin bağlamsal olarak anlaşılmasını gerektirmektedir. Müzik türü sınıflandırması, konuşmadan duygu tanıma, ses olayı algılama, ses etkinliği algılama, şiddet tespiti ve konuşmacı sayma vb. gibi uygulamalarda kullanılabilmektedir.

5.2. Ses Analizinde Dikkat Mekanizmaları

Geleneksel RNN'ler etkili olmakla birlikte, kaybolan gradyanlar ve bellek darboğazları olguları nedeniyle uzun dizileri işlemede zorluklarla karşılaşmaktadır. Dikkat mekanizmaları, bir ses dizisi içinde göze çarpan bölümlerin seçici olarak önceliklendirilmesini sağlayarak bu sınırlamaları hafifletmektedir.

5.2.1. Belirgin Ses Özelliği Tanımlama için Dönüştürücü Mimarilerinin Uygulanması

Transformatör mimarileri, özellikle Öz Dikkat Mekanizmalı Ağlar (SAN'lar) ve Transformatörlerden Çift Yönlü Kodlayıcı Temsilleri (BERT), ses işleme uygulamalarında giderek daha fazla kullanılmaktadır [147].

Ses Analizinde Dikkat Mekanizmalarının Özellikleri

- **Küresel bağlam farkındalığı:** RNN'lerin sıralı işleminin aksine, dönüştürücüler tüm giriş dizisinin eşzamanlı analizini kolaylaştırır ve böylece küresel karşılıklı bağımlılıkları yakalamaktadır.

- **Geliştirilmiş duygusal ve yoğunluk özellik çıkarımı:** Konuşma duygusu tanıma görevlerinde, dikkat mekanizmaları, önemli duygusal bilgiler ileten belirli hece veya perde modülasyon modellerine hedefli odaklanmayı sağlamaktadır.
- **Ölçeklenebilirlik ve transfer öğrenme etkinliği:** Transformatör modelleri kapsamlı veri kümeleri üzerinde önceden eğitilebilir ve daha sonra özel alana özgü görevler için ince ayar yapılabilir, bu da yüksek ölçeklenebilirlik göstermektedir.

Transformerler ile Gerçekleştirilen Uygulamalar

SER bağlamında, dönüştürücü tabanlı bir model, stres veya tonlama varyasyonları gibi duygusal ipuçlarıyla karakterize edilen belirli ses çerçevelerine SAN ile farklı ağırlıklar atayabilir ve böylece sınıflandırma doğruluğunu artırabilmektedir. HuBERT, Wav2Vec 2.0 gibi modeller, konuşma analizinde tonlama, vurgu ve duraklamaları modelleyerek duygu tanımda %90'lara varan doğruluk sağlamaktadır. Gerçek dünya senaryosunda ise çağrı merkezlerinde müşteri öfkelerini tespit etme, terapide depresyon analizi gibi çalışmalar yapılmaktadır [148].

5.3. Ses Analizinde Transfer Öğrenimi

Derin öğrenme modelleri, güçlü olmakla birlikte, karakteristik olarak veri yoğundur ve optimum eğitim için önemli veri kümeleri gerektirir. Bununla birlikte, ses verilerinin elde edilmesi ve ek açıklama eklenmesi, doğal masrafları ve zamansal talepleri nedeniyle önemli zorluklar ortaya çıkarmaktadır. Transfer öğrenimi, önceden eğitilmiş modellerden yararlanarak bu kısıtlamaya pragmatik bir çözüm sunmaktadır.

5.3.1. Önceden Eğitilmiş Ses Özellik Çıkarma Modellerinin Kullanımı

VGG mimarisinden türetilen VGGish ve MobileNet mimarisine dayanan YAMNet gibi modeller, kapsamlı ses veri setleri üzerinde eğitilerek ses temsillerini öğrenmektedir. Bu modeller, sınırlı veri kümeleri kullanan sonraki görevler için transfer öğrenme yoluyla sağlam özellik çıkarıcılar olarak hizmet etmekte ve hedef görevlerin performansını arttırmaktadır.

Ses Analizinde Transfer Öğrenmenin Avantajları

- **Sınırlı açıklamalı veri ile geliştirilmiş model performansı:** Önceden eğitilmiş modellere ince ayar yapmak, önemli ölçüde azaltılmış etiketli veri hacimleriyle üstün doğruluk elde edilmesini sağlamaktadır.
- **Alanlar arası adaptasyon:** Önceden eğitilmiş modeller genelleştirilmiş ses özelliği temsilleri edinirken, sağlık hizmetlerinde tanısal ses analizi ve vahşi yaşamın

korunmasında akustik izleme dâhil olmak üzere özel uygulamalar için etkili bir şekilde uyarlanabilirler.

- **Hesaplamalı kaynak optimizasyonu:** Önceden eğitilmiş modellerden elde edilen önceden hesaplanmış özellik yerleştirmelerinin kullanılması, eğitim süresini ve hesaplama kaynağı tüketimini önemli ölçüde azaltmaktadır.

Transfer Öğrenme ile İlgili Uygulamalar

Örneğin sağlık odaklı bir ses sınıflandırma sistemi VGGish'i kullanabilir. Bu modelin AudioSet veri kümesi üzerinde öğrendiği genel akustik özellikleri transfer ederek akciğer ses kayıtlarındaki solunum anomalilerini yüksek doğrulukta tespit edebilmektedir. Böylece kapsamlı manuel açıklama ihtiyacını azaltır ve klinik teşhis araçlarının geliştirilmesini hızlandırmaktadır [8, 149].

5.4. Gelişmiş Tahmin Sağlamlığı için Topluluk Öğrenme Stratejileri

Toplu öğrenme (Ensemble learning) metodolojileri, birden fazla kurucu modelden elde edilen çıktıları bir araya getirerek tahmin performansını artırmayı ve böylece güvenilirliği artırmayı amaçlamaktadır. Bu yaklaşım, farklı modellerin ses verilerinin doğasında bulunan farklı, tamamlayıcı özellikleri yakaladığı ses analizinde özellikle etkilidir.

5.4.1. Gelişmiş Performans Optimizasyonu için Model Toplama

Yaygın topluluk teknikleri şunları içermektedir:

- **Torbalama (örn. Rastgele ormanlar):** Bu teknik, birden fazla zayıf sınıflandırıcıdan gelen tahminlerin toplanmasını içermektedir. Varyansı etkili bir şekilde azaltmakta ve model kararlılığını artırmaktadır.
- **Boosting (örn. Gradient boosting machines-GBM):** Bu sıralı yaklaşım, önceki tahmin hatalarının düzeltilmesine odaklanarak zayıf modelleri yinelemeli olarak iyileştirir ve genel doğruluğun artmasını sağlamaktadır [150].
- **Hibrit model mimarileri (CNN + RNN):** Bu mimariler, uzamsal özellik çıkarımı için CNN'ler zamansal bağımlılıkların modellenmesi için RNN'ler ile birleştirilerek her bir bileşenin güçlü yönlerinden yararlanmaktadır.

Ses İşlemede Toplu Öğrenmenin Avantajları

- **Artırılmış tahminsel doğruluk:** Birden fazla modelin kombinasyonu, aşırı uyum eğilimini azaltır ve modelin görülmeyen verilere genelleme yeteneğini geliştirmektedir.

- **Geliştirilmiş gürültü sağlamlığı:** Farklı modellerin sinerjik etkileşimi, bireysel model zayıflıklarını telafi ederek gürültülü ortam koşullarında daha iyi performans sağlamaktadır.
- **Uygulama çok yönlülüğü:** Topluluk öğrenimi, otomatik müzik etiketleme ve silah sesi tespiti gibi gözetim sistemlerinde akustik olay tespiti de dâhil olmak üzere geniş bir uygulama yelpazesinde fayda sağlamaktadır.

Toplu Öğrenme ile İlgili Uygulamalar

CNN, RNN ve Transformer tabanlı mimarilerden oluşan bir topluluk modeli, kentsel akustik sahne sınıflandırması için kullanılabilir; burada her bir bileşen modeli, araç trafiği, insan sesleri ve alarm sinyalleri gibi farklı işitsel olayların tanınmasında uzmanlaşmıştır.

5.5. Değerlendirme

Sonuç olarak, derin öğrenme metodolojileri, uzamsal özellik çıkarımı için CNN'lerin, sıralı örüntü tanıma için RNN'lerin ve verimli, dikkat tabanlı modelleme için Transformatör mimarilerinin stratejik uygulamasıyla kolaylaştırılan ses analizinde bir teknik değişikliğine yol açmaktadır. Transfer öğrenme teknikleri, araştırmacıların büyük ölçekli önceden eğitilmiş modellerin içerdiği bilgiden yararlanmasını sağlarken, topluluk öğrenme stratejileri gerçek dünya dağıtım senaryolarında model sağlamlığını artırmaktadır. Toplu olarak, bu geliştirilmiş teknikler konuşma işleme, duygu tanıma, şiddet tanıma, çevresel akustik sahne sınıflandırması ve tıbbi teşhis uygulamaları dâhil olmak üzere çeşitli alanlarda önemli ilerlemeler sağlamıştır.

6. MATERYAL VE METOT

Bu bölümde tez aşamasında yayınlanan ve literatüre katkı sağlanan makaleler ile ilgili detaylı bilgiler verilmektedir. Bu bölümde kullanılacak olan performans değerlendirme ölçütlerinin matematiksel ifadeleri aşağıda verilmiştir [151].

Doğruluk: Tüm tahminlerin ne kadarının doğru olduğunu ölçmektedir. Dengeli veri setlerinde genel performansı değerlendirmek için sıklıkla kullanılmaktadır. Denklem 6.1 kullanılarak hesaplanmaktadır.

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

Kesinlik: Pozitif olarak tahmin edilenlerin ne kadarının gerçekten pozitif olduğunu ölçmektedir. Yanlış pozitiflerin maliyetli olduğu durumlarda kullanımı tercih edilmektedir (örn. spam tespiti vb.). Denklem 6.2 kullanılarak hesaplanmaktadır.

$$Kesinlik = \frac{TP}{TP+FP} \quad (6.2)$$

F1 Skor: Kesinlik ve duyarlılığın harmonik ortalaması olarak hesaplanır. Dengesiz veri setlerinde performansı dengeli değerlendirmek için kullanılmaktadır. Denklem 6.3 kullanılarak hesaplanmaktadır.

$$F1 \text{ skor} = \frac{2TP}{2TP+FP+FN} \quad (6.3)$$

Duyarlılık: Gerçek pozitiflerin ne kadarının doğru tespit edildiğini ölçmektedir. Yanlış negatiflerin kritik olduğu durumlarda (örneğin, kanser teşhisi vb.) kullanımı tercih edilmektedir. Denklem 6.4 kullanılarak hesaplanmaktadır.

$$Duyarlılık = \frac{TP}{TP+FN} \quad (6.4)$$

Özgüllük: Gerçek negatiflerin ne kadarının doğru tespit edildiğini ölçmektedir. Yanlış pozitiflerin istenmediği durumlarda (örn. sağlıklı bireylerin yanlış teşhisi vb.) kullanımı tercih edilebilmektedir. Denklem 6.5 kullanılarak hesaplanmaktadır.

$$Özgüllük = \frac{TN}{TN+FP} \quad (6.5)$$

Geometrik Ortalama: Duyarlılık ve özgüllük arasındaki dengeyi ölçmektedir. Dengesiz veri setlerinde modelin genel yeteneğini ölçmek için kullanımı tercih edilebilmektedir. Matematiksel ifadesi Denklem 6.6'da gösterilmektedir.

$$Geometrik Ortalama = \sqrt{\frac{TP}{TP+FN} \times \frac{TN}{TN+FP}} \quad (6.6)$$

MSE: Gerçek ve tahmini değerler arasındaki farkın karesinin ortalaması olarak tanımlanır. Tahminlerin veya öngörülerin gerçek değerlere ne kadar yakın olduğunu kontrol etmek için kullanılır. MSE ne kadar düşükse, tahmin gerçek değere o kadar yakındır. Bu, regresyon modelleri için bir model değerlendirme ölçüsü olarak kullanılır ve daha düşük değer daha iyi bir uyumu göstermektedir. Matematiksel olarak denklem 6.7’de ifade edilmektedir.

$$MSE = \frac{1}{N} \sum_{i=1}^N (y_i - y'_i)^2 \quad (6.7)$$

6.1. Yeni Bir Lateral Genikülat Çekirdek Örüntüsü Tabanlı Çevresel Ses Sınıflandırması

İnsan İşitsel Sistemine (HAS) benzer veya daha üstün performansa sahip ses sınıflandırma sistemleri arayışı önemli bir araştırma zorunluluğu olmaya devam etmektedir [152-153]. Bu çalışma yeni, kamuya açık bir Çevresel Ses Sınıflandırma (ESC) veri kümesi ve Lateral Genikülat Çekirdek Örüntüsü (LGNPat) olarak adlandırılan biyolojik olarak esinlenmiş bir özellik çıkarma işlevi sunarak bu hedefe katkıda bulunmaktadır [154]. Bu unsurlar yeni geliştirilen, yüksek doğruluk oranına sahip ve elle tasarlanmış bir ses sınıflandırma mimarisinin çekirdeğini oluşturmaktadır.

Metodolojik olarak değerlendirme platformu şeklinde hizmet etmek üzere 50 sınıfta 5000 sestten oluşan dengeli bir ESC veri seti toplanmıştır. Önerilen sınıflandırma çerçevesi şunları entegre etmektedir: (1) istatistiksel momentler ve DWT ile birlikte yenilikçi LGNPat kullanılarak özellik çıkarımı; (2) sağlam özellik seçimi için INCA; ve (3) 10 kat çapraz doğrulama protokolü ile doğrulanan kNN algoritması tarafından gerçekleştirilen sınıflandırma işlemidir. Bu mimarinin önemli bir özelliği, tek boyutlu sinyal sınıflandırmasına yönelik gelecekteki elle tasarlanmış yaklaşımlar için bir paradigma sunan doğal genişletilebilir olma özelliğinde olmasıdır.

LGNPat tabanlı modelin bu veri kümesi üzerindeki deneysel değerlendirmesi %93,34'lük bir sınıflandırma doğruluğu göstermiştir. Bu sonuç, önerilen modelin etkinliğinin altını çizmektedir. ESC veri setinin kamuya açık hale getirilmesinin daha fazla ilerlemeyi kolaylaştırması ve bu alandaki sonraki araştırmalar için bir ölçüt sağlaması beklenmektedir.

6.1.1. Literatür Taraması Özeti

Çevresel ses sınıflandırmasının öneminden bahsedilmiştir. Çünkü ESC İşitme engelliler için güvenlik sistemleri, akıllı ev uygulamaları ve adli bilişim gibi alanlarda kritik bir rol

oynamıştır. Genellikle seslerin non-stasyoneryapısı ve arka plan gürültüsüne duyarlılık ile HAS seviyesinde doğruluk gereksinimi karşılaşılan zorluklar olarak ortaya çıkmıştır.

6.1.2. Mevcut Çalışmalar

ESC-10, ESC-50 ve UrbanSound8K gibi halka açık veri setleri kullanılarak derin öğrenme (CNN, LSTM) ve öznitelik mühendisliği (MFCC, LBP) tabanlı yöntemler sunulmuştur. Mevcut veri setleri ESC-10, ESC-50 gibi veri setleri 10 ya da 50 sınıflı küçük ölçekli veri setleridir. UrbanSound8K ise kent seslerine odaklanır ancak sınırlı çeşitliliğe sahiptir.

6.1.3. Mevcut çalışmalardaki eksiklikler

Bu bölümde mevcut çalışmalardaki eksikliklerden bahsedilmiştir.

Sınırlı veri kümesi boyutu ve sınıf çeşitliliği: Derin öğrenme modellerinin etkin bir şekilde eğitilmesi, genellikle büyük ve çeşitli veri kümelerine dayanır. Ancak, bazı alanlarda veri toplama ve etiketleme süreçlerinin maliyeti ve zaman alıcı olması nedeniyle veri kümesi boyutları sınırlı kalabilmektedir. Ayrıca, veri kümesindeki sınıf çeşitliliğinin azlığı, modelin genelleme yeteneğini olumsuz etkileyebilmektedir.

Yüksek hesaplama maliyeti: Derin öğrenme modelleri göz önüne alındığında büyük ve karmaşık modellerin eğitimi yüksek hesaplama maliyeti gerektirmektedir. Yüksek performanslı ekran kartları GPU'lar veya TPU'lar gibi donanımların kullanımı bu maliyeti artırabilmektedir.

Negatif ses örneklerini işleyemeyen öznitelik çıkarıcılar: Ses analizi uygulamalarında, öznitelik çıkarıcıların negatif ses örneklerini etkili bir şekilde işleyememesi modelin performansını düşürebilmektedir. Örneğin, arka plan gürültüsü veya sessizlik buna örnek olarak verilmektedir. Bu durum özellikle gürültülü ortamlarda çalışan sistemlerde önemli bir sorun teşkil etmektedir.

6.1.4. Çalışmada Kullanılan Veri Seti

Kullanılan veri seti 50 sınıftan oluşmaktadır. Bu veri setinde her sınıfta 200 örnek olmak üzere toplam 10.000 ses örneği bulunmaktadır. Ses uzunluğu 2 saniye ve 44.1 kHz olarak belirlenmiştir. Dosya formatları WAV ve MP3 formatında olup tüm sınıflar eşit sayıda örnek içermektedir. Bu çalışmada homojen bir veri seti kullanılmıştır. Bu sesler kategorilere göre numaralandırıldı ve toplanan 50 kategori Tablo 6.1'de sunulmuştur. Veri kümesine şu bağlantı adresinden erişilebilir: <https://www.kaggle.com/datasets/buraktaci/firat-esc50> [155].

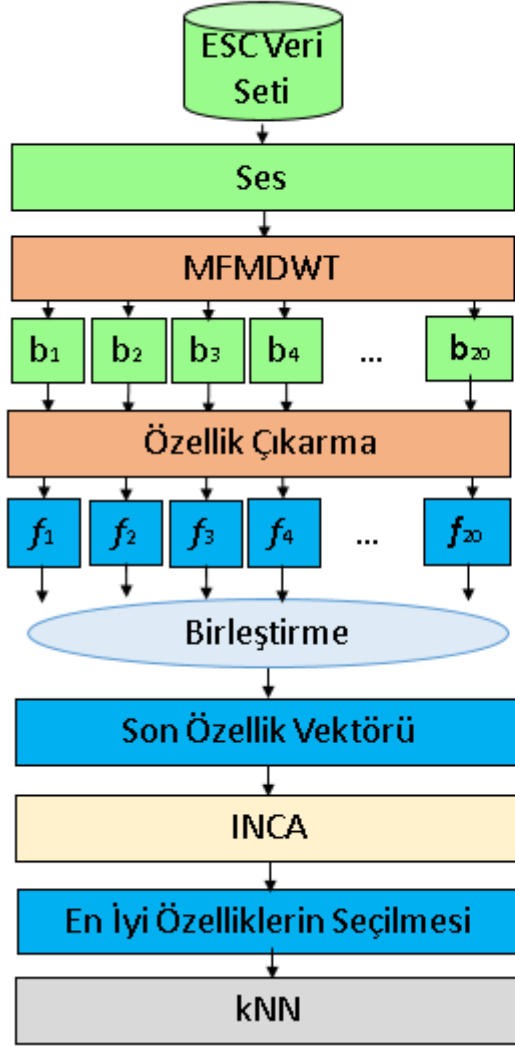
Tablo 6.1. Fırat ESC-50 veri seti [128]

No	Kategori	No	Kategori	No	Kategori	No	Kategori
1	Erkek konuşması	14	Tabak tıktırması	27	Telefon tuş sesi	40	Telefon çalması
2	Tuvalet sifonu	15	Telefon zil sesi	28	Kız konuşması	41	Su sıçraması
3	Sıvı dökme	16	Mikrodalga	29	Araba kornası	42	Bilgisayar konuşması
4	Diş fırçalama	17	Köpek havlaması	30	Yazı yazma	43	Çalar saat
5	Kadın konuşması	18	Sert yüzeyde yürüme	31	Bilgisayar başlama sesi	44	Topuklu ayakkabıyla yürüme
6	Araba hızlanması	19	Yol trafiği	32	Arka plan konuşması	45	Elektrik süpürgesi
7	Isırma ve çiğneme	20	Fermuar sesi	33	Kuş şarkısı	46	Rüzgâr
8	Gülme	21	Cep telefonu titreşimi	34	Su dökme	47	Erkek çocuk konuşması
9	Klavye yazma	22	Su damlaması	35	Pop şarkı	48	Sandalye yuvarlanması
10	Araba motoru çalışması	23	Kaşım	36	Su kaynaması	49	Rock şarkı
11	Akan su	24	Araba camları	37	Gitar	50	Kapı tıklatma
12	Nefes alma	25	Telefon çalması	38	Öksürme		
13	Anahtar şıkırtısı	26	Yiyecek doğrama	39	Kâğıt buruşturma		

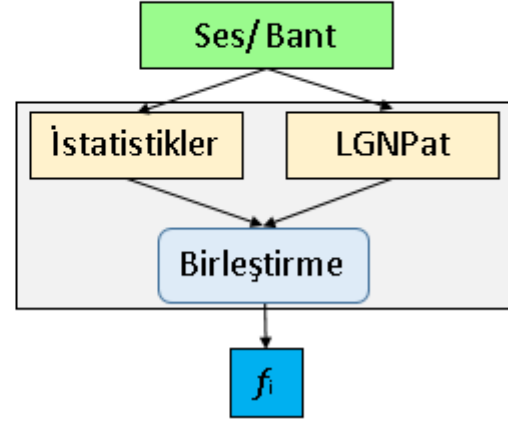
6.1.5. Çalışmada Önerilen Yöntem

Bu çalışmada, el yapımı öznitelik mühendisliği yaklaşımına dayanan bir ses sınıflandırma modeli önerilmiştir. Model, üç temel bileşenden oluşmaktadır. Bunlar özellik çıkarımı, özellik seçimi ve sınıflandırma adımlarıdır. Bu makalede önerilen modelin akış diyagramı Şekil 6.1’de sunulmuştur.

LGNPat-MFMDWT tabanlı ses sınıflandırma



Özellik çıkarma



Açıklamalar:

MFMDWT: Çoklu Filtre Tabanlı ÇokSeviyeli Ayırık Dalgacık Dönüşümü

b: alt bantlar

f: özellik vektörü

INCA: Yinelemeli Komşuluk Bileşen Analizi

kNN: k en yakın komşular



Şekil 6.1. Önerilen LGNPAT-MFMDWT tabanlı ses sınıflandırma modeline genel bakış [128].

MFMDWT kullanılarak 20 alt bant oluşturulmuştur. Özellik oluşturma aşamasında, iki özellik çıkarma fonksiyonu kullanılmıştır: (1) istatistiksel özellik oluşturucu ve (2) önerdiğimiz LGNPAT dokusal özellik çıkarıcı. Bu fonksiyonlar (istatistiksel ve LGNPAT) kullanılarak istatistiksel ve dokusal özellikler üretilmiştir. Burada, iki filtre ile beş seviyeli DWT kullanılmıştır. Bu nedenle 20 alt bant oluşturulmuştur. Hesaplanan 21 özellik vektörü (20 tanesi alt bantlar kullanılarak ve bir tanesi de ham ses sinyali kullanılarak oluşturulmuştur). Bu 21 özellik vektörü nihai özellik vektörünü elde etmek için birleştirilmiştir. En bilgilendirici özellikler INCA özellik seçici kullanılarak seçilmiştir ve bu özellikler kNN sınıflandırıcıya aktarılmıştır.

Özellik Çıkarımı

- **Lateral Geniculate Nucleus Pattern (LGNPAT):** İnsan görsel sisteminden esinlenen grafik tabanlı bir özellik çıkarma yöntemi kullanılmıştır.

- **Çok Seviyeli Ayrık Dalgacık Dönüşümü (MFMDWT):** Ses sinyallerinin farklı frekans bileşenlerini ayırmak için kullanılmıştır.
- **İstatistiksel Momentler:** Ses sinyallerinden 20 farklı istatistiksel özellik çıkarılmıştır. Bu özelliklere ait formüller Tablo 6.2’de gösterilmiştir.

Tablo 6.2. İstatistiksel özellikler

No	İstatistik	Formüller
1	Tsallis entropisi	$S_q = \frac{1-\sum p_i^q}{q-1}$
2	Shannon entropisi	$H_j = -\sum_n p_j(n) \log p_j(n)$
3	Sure entropisi	Sinyal gürültüsünü azaltmak için; Stein’in Tarafsız Risk Tahmincisi’ne dayalıdır.
4	Dalgacık entropisi	Dalgacık katsayılarından elde edilmektedir.
5	Log enerji entropisi	$E = \sum_i \log p_i^2$
6	Renyi entropisi	$H_\alpha = \frac{1}{1-\alpha} \sum \log p_i^\alpha$
7	Çeyrek 3 (Quartile 3)	$(n+1) \times \frac{3}{4}$
8	Ortalama (Mean)	$\mu = \frac{1}{n} \sum_{i=1}^n x_i$
9	Medyan	n tek ise: $x_{\left[\frac{n+1}{2}\right]}$ n çift ise: $\frac{x_{\left[\frac{n}{2}\right]} + x_{\left[\frac{n}{2}+1\right]}}{2}$
10	Mod	$1 + h \left(\frac{f_m - f_1}{2f_m - f_1 - f_2} \right)$
11	Aralık (Range)	$\max(x) - \min(x)$
12	Moment	$\mu_k = \frac{1}{n} \sum (x_i - \mu)^k$
13	Higuchi	$D = \frac{\log(L(k))}{\log(k)}$
14	Lyapunov üssü	$\lambda = \lim_{n \rightarrow \infty} \frac{1}{n} \sum \ln \left \frac{df}{dx} \right $
15	Enerji	$E = \sum x_i^2$
16	Çeyrek 1 (Quartile 1)	$(n+1) \times \frac{1}{4}$
17	Çeyrek aralık (Q. range)	$(n+1) \times \frac{3}{4} - (n+1) \times \frac{1}{4}$
18	Varyans	$\sigma^2 = \frac{1}{n} \sum (x_i - \mu)^2$
19	Çarpıklık (Skewness)	$Skew = \frac{1}{n} \sum \left(\frac{x_i - \mu}{\sigma} \right)^3$
20	Basıklık (Kurtosis)	$Kurt = \frac{1}{n} \sum \left(\frac{x_i - \mu}{\sigma} \right)^4$

Dalgacık dönüşümü ve LGNPat yöntemleri, düşük, orta ve yüksek frekans bantlarındaki ses sinyallerinden özellik çıkarılmasını sağlamıştır.

Özellik Seçimi

- **INCA:** Özellik vektöründen en bilgilendirici özellikleri seçmek için kullanılmıştır.

Başlangıçta 6216 boyutunda olan özellik vektörü INCA yöntemi ile 92 boyuta indirilmiştir.

Sınıflandırma

- **k-En Yakın Komşu (kNN) Sınıflandırıcısı:** Sınıflandırma görevi için kNN sınıflandırıcı kullanılmış ve modelin performansı 10 kat çapraz doğrulama yöntemi ile değerlendirilmiştir.

Modelin performansı doğruluk, kesinlik, geri çağırma ve F1 skoru gibi metrikler kullanılarak hesaplanmıştır.

6.1.6. Çalışmanın Performans Değerlendirmesi

Genel olarak bakıldığında literatürde sıklıkla kullanılan performans değerlendirme ölçütleri bilinmektedir. Bu değerlendirme ölçütlerine (Denklem 6.1-Denklem 6.6) göre ortaya çıkan sonuçlar Tablo 6.3’de gösterilmektedir.

Tablo 6.3. LGNPat modeline ait Performans değerlendirme sonuçları [128]

Performans Metrikler	Sonuçlar (%)
Doğruluk	85.57
Duyarlılık	85.57
Kesinlik	85.72
F1-skoru	85.65

Genel sınıflandırma doğruluğu: %85,57 olarak ölçülmüştür. Duyarlılık %85,57, Kesinlik %85,72 ve F1 skoru %85,65 olarak hesaplanmıştır.

Önerilen modelin avantajları

- **Düşük hesaplama maliyeti:** GPU gerektirmeyen el yapımı özellik mühendisliği önerilmiştir.
- **Yüksek doğruluk:** Önceki derin öğrenme modelleriyle karşılaştırılabilir performans sağlamıştır.
- **Yeni veri seti:** ESC-50’ye alternatif olarak FIRAT-ESC50 oluşturulmuştur.

6.1.7. Çalışmanın Katkıları

FIRAT-ESC-50 adında 50 sınıflı 10.000 ses örneği içeren yeni bir veri seti önerilmiştir. Biyolojik esinli Lateral Genikulat Nükleus yapısı temelli LGNPat adında yeni bir öznetelik çıkarıcı

önerilmiştir. Derin öğrenme yerine ise öznelik mühendisliği ve kNN tabanlı hafif bir model kullanılmıştır.

6.1.8. Çalışmanın Sonuçları ve Tartışma

Bu araştırma yeni bir veri seti ve biyolojik esinli hafif ses sınıflandırma modeli önermiştir. Toplanan veri seti 10.000 ses içermekte dengeli olarak nitelendirilmektedir. Bu veri setinden düşük zaman karmaşıklığıyla sonuçlar elde etmek için biyolojik esinli yerel bir özellik üretici önerildi ve bu özellik üretici LGNPat olarak adlandırıldı. Önerimiz bir özellik mühendisliği modelidir. Kullanılan seslerin özellikleri, önerilen özellik çıkarma yöntemimizden üretildi. Özellik çıkarma yöntemimizde, özellikler oluşturmak için yeni bir DWT modeli kullanıldı ve bu MFMDWT olarak adlandırıldı. İstatistiksel bir özellik üretici, önerilen LGNPat kullandığımızdan ve bu özellik oluşturma işlevleri elle hazırlanmış özellik üreticileri olduğundan düşük, orta ve yüksek seviyelerde özellikler oluşturmak için MFMDWT kullanıldı. Her ikisi de (istatistiksel özellik üretici ve LGNPat) tek başlarına kullanıldıklarında düşük seviyede özellik üretmektedir. Değişken frekanslı alt bantlar, değişken seviyelerle orta ve yüksek seviyelerde özellikler oluşturmak için kullanılmaktadır. Bu çalışmada, MFMDWT ve ham ses sinyalleri kullanılarak özellikler hem frekans hem de uzay alanlarından çıkarılmıştır. Önerilen LGNPat-MFMDWT modeli, el yapımı özellik çıkarma yöntemi ile **%85,57** doğruluk oranına ulaşmıştır. ESC-50 ve diğer çevresel ses veri kümeleri ile karşılaştırıldığında benzer veya daha iyi doğruluk sağlamıştır. Modelin hafif yapısı ve düşük hesaplama maliyeti ile yoğun hesaplama gerektiren derin öğrenme modellerine bir alternatif sunduğu belirtilmiştir.

6.2. BP19: Tek Boyutlu İkili Örüntüye Dayalı Doğru Bir Ses Şiddeti Tespit Modeli

Ses Şiddeti Tespiti (AVD), ses adli bilimlerinde yeni ortaya çıkan ve kritik bir araştırma alanıdır ancak mevcut çalışmaların sınırlı bir çerçevesiyle ile karakterize edilmektedir. Bu araştırma, yeni bir ses veri kümesine katkıda bulunarak ve AVD için sağlam bir ikili örüntü tabanlı sınıflandırma algoritması önererek alanı ilerletmeyi amaçlamaktadır. Metodolojik olarak “şiddet içeren” veya “şiddet içermeyen” olarak sınıflandırılan 301 ses örneğini kapsayan yeni bir AVD veri kümesi oluşturulmuş ve bu veri kümesi değerlendirme ölçütü olarak kullanılmıştır [156].

Önerilen özellik mühendisliği hattı, tek boyutlu İkili Örüntülerin (BP) çıkarılmasını ve hem frekans hem de ilgili dönüşüm alanlarından özellikler türetmek için TQWT uygulanmasını içermektedir. Özellik seçimi daha sonra INCA kullanılarak gerçekleştirilmiş ve ardından seçilen özellikler optimize edilmiş bir SVM ile sınıflandırılmıştır. Model, toplanan veri kümesi üzerinde **%97,01**'lik bir sınıflandırma doğruluğu elde etmiş ve 10 kat çapraz doğrulama ile doğrulanmıştır.

Bu sonuçlar, önerilen özellik mühendisliği yaklaşımının ses tabanlı şiddet tespiti için oldukça etkili bir çözüm sunduğunu göstermektedir.

6.2.1. Literatür Taraması Özeti

Şiddet ve duygu tanıma alanında önemli zararlara yol açabilecek şiddet içeren davranışların belirlenmesi, agresif seslerin ve ortam gürültüsünün tespit edilmesini gerektirdiğinden bahsedilmiştir. Önceki araştırmalar, konuşma tanıma, görüntü analizi ve NLP'yi kapsayan metodolojiler aracılığıyla duygusal durumların ve agresif eğilimlerin sınıflandırılmasını araştırmıştır. “Facebook, Twitter ve YouTube” dâhil olmak üzere sosyal medya platformları, duygusal içeriği ayırt etmek için etkileşimli alışverişlerin analizi için derlem olarak kullanılmıştır. Duygu tanıma alanındaki zorluklar, bireysel deneyimsel bağlamlardan etkilenen duygusal durumların doğasında var olan öznellik, doğru tanıma için önemli bir zorluk teşkil etmektedir.

NLP, nefret söylemi tespiti ve siber zorbalık analizinde etkinliğini kanıtlamış olsa da gerçek dünyadaki konuşmaların analizi daha fazla karmaşıklık ortaya çıkarmıştır.

Adli bilimlerde akustik analizin önemine göre adli soruşturmalar, kanıt bütünlüğünü sağlamak için gerçek ve değiştirilmemiş ses verilerinin elde edilmesini zorunlu kılmıştır. Konuşma içeriğini, arka plan gürültüsünü ve çevresel akustik özellikleri entegre eden kapsamlı bir analitik yaklaşımın etkili adli ses analizi için gerekli olduğu ortaya çıkmıştır.

Hesaplama açısından verimli bir model gereksinimi düşünüldüğünde mevcut modellerin önemli bir kısmı, kaynağı kısıtlı ortamlarda uygulanabilirliklerini sınırlayabilecek şekilde derin öğrenme mimarilerine dayanan yüksek hesaplama karmaşıklığı sergilemiştir. Önerilen model, görsel-işitsel şiddet tespiti (AVD) için hesaplama açısından hafif, verimli ve yüksek hassasiyetli bir çözüm sağlamayı amaçlamaktadır.

6.2.2. Literatürdeki Mevcut Çalışmalar

Genelde ses tabanlı şiddet algılama alanı, bilimsel araştırma için cazip bir yol sunmaktadır. Bununla birlikte, mevcut literatürün eleştirel bir analizi, mevcut metodolojilerin ya büyük ölçüde hesaplama açısından yoğun derin öğrenme mimarilerine dayandığı ya da gerçek zamanlı işleme etkinliğine ulaşmada sınırlamalar sergilediği bir ikilemi ortaya koymuştur. Sonuç olarak, bu çalışma, tahmin doğruluğundan ödün vermeden hesaplama verimliliği elde etmek için tasarlanmış yeni bir ikili örüntü tabanlı özellik çıkarma modeli önererek bu boşluğu gidermeye çalışmaktadır.

6.2.3. Mevcut Çalışmalardaki Eksiklikler

Bu bölümde mevcut çalışmalardaki eksikliklerden bahsedilmiştir.

Akustik modalitelerin yeterince keşfedilmemesi: Görsel veriler (video ve görüntüler) kullanılarak şiddetin tespit edilmesine yönelik kapsamlı araştırmalara rağmen ses tabanlı yaklaşımların potansiyeli önemli ölçüde keşfedilmemiştir. Akustik sinyaller değerli bağlamsal bilgiler sunar ve genellikle görsel verilerden daha az hesaplama gerektirir; ancak şiddet olaylarını tanımlamadaki faydalarından araştırma topluluğu içinde tam olarak yararlanılmamıştır.

Uzmanlaşmış ve kamuya açık veri Setlerinin azlığı: AVD'de ilerlemenin önündeki en büyük engellerden biri uygun veri setlerinin olmamasıdır. Mevcut araştırmalar sıklıkla genel amaçlı ses derlemelerine veya gerçek dünya karmaşıklığından ve ekolojik geçerlilikten yoksun simüle edilmiş senaryolara dayanmaktadır. Şiddet olaylarının gerçek dünyadaki ses kayıtlarını, özellikle de adli analiz için özel olarak açıklanmış olanları içeren yüksek kaliteli, dengeli, kamuya açık veri kümelerinde kritik bir eksiklik vardır. Bu eksiklik, farklı çalışmalar arasında sağlam model kıyaslamasını ve karşılaştırmalı performans değerlendirmelerini engellemektedir.

Akustik duygu ve bağlam tanımadaki doğal zorluklar: Şiddet tespiti için sesin yorumlanması, duygusal durumların ve bağlamsal bilgilerin tanınmasındaki doğal zorluklar nedeniyle karmaşıktır. Konuşma yoluyla duygu ifadesi öznel ve bireysel özelliklerden, kültürel normlardan ve arka plan gürültüsü gibi akustik parazitlerden kaynaklanan değişikliklere açıktır. Ayrıca şiddet olaylarının gerçek kayıtları genellikle üst üste binen ses kaynakları, belirsiz seslendirmeler ve çeşitli çevresel özellikler içeren karmaşık akustik sahneler içerir ve geleneksel analitik modeller için önemli zorluklar ortaya çıkarır.

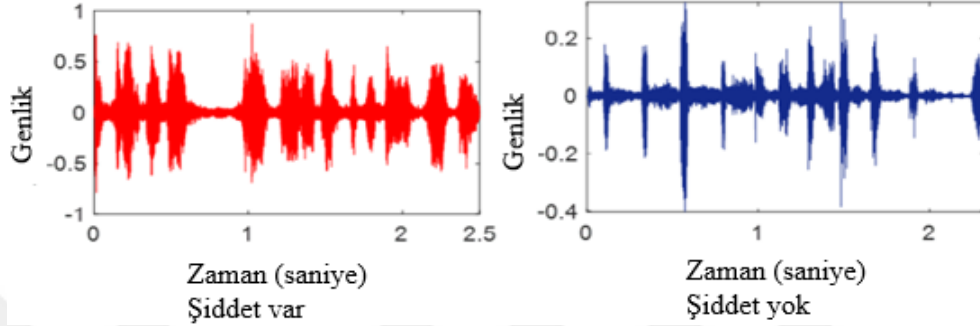
Baskın derin öğrenme modellerinin hesaplama kısıtlamaları: Derin öğrenme mimarileri (örn. CNN'ler), LSTM ağları AVD görevlerinde yüksek performans göstermiş olsa da pratik dağıtımları genellikle önemli hesaplama gereksinimleri ile sınırlıdır. Bu modeller tipik olarak büyük hacimlerde etiketli eğitim verisi, güçlü GPU kaynakları ve uzun eğitim süreleri gerektirmektedir. Bu da onları gerçek zamanlı uygulamalar veya sınırlı hesaplama altyapısına sahip adli laboratuvarlar gibi kaynak kısıtlı ortamlarda dağıtım için daha az uygulanabilir hale getirmektedir.

Model doğrulamada metodolojik zayıflıklar: Mevcut bazı AVD sistemlerinin rapor edilen etkinliği, optimal olmayan doğrulama tekniklerinin uygulanması nedeniyle sorgulanabilir. Özellikle, basit rastgele eğitim-test bölünmelerinin yaygın kullanımı, potansiyel olarak modelin aşırı uyumuna yol açabilir ve sonuç olarak gerçek dünya genelleme yeteneklerini doğru bir şekilde yansıtmayan şişirilmiş performans ölçümleri sağlayabilir. AVD modellerinin güvenilirliğini ve genelleştirilebilirliğini artırmak için sağlam çapraz doğrulama protokolleri de dâhil olmak üzere daha titiz doğrulama stratejilerinin benimsenmesine acil ihtiyaç vardır.

6.2.4. Çalışmada Kullanılan Veri Seti

“YouTube” platformundan yeni bir görsel-ışitsel şiddet algılama (AVD) veri kümesi seçilmiştir. Bu veri kümesi, iki farklı kategoriye ayrılmıştır. Bunlar 154 adet Şiddet içeren örnek

ve 147 adet Şiddet içermeyen örnek olmak üzere toplam 301 ses örneğinden oluşmaktadır. Her ses segmenti yaklaşık 2 ila 3 saniyelik bir süreye ve 44.1 kHz frekansa sahiptir. Bu veri kümesi, önerilen sınıflandırma modelinin değerlendirilmesi için bir test ortamı görevi görmektedir. Bu sınıflara ait ses sinyallerinin genlik grafiği Şekil 6.2’de gösterilmektedir.



Şekil 6.2. Şiddet içeren ve şiddet içermeyen ses segmentlerinin genlik gösterimi [156].

6.2.5. Çalışmada Önerilen Yöntem

Bu çalışmada önerilen modelin temel amacı, basit ve hafif bir özellik mühendisliği modeliyle yüksek sınıflandırma performansı elde etmektir. Modelimiz, özellikleri elde etmek için TQWT ve tek boyutlu BP'yi kullanır [130, 157]. Bu özelliklerin en üstleri INCA seçicisini dağıtarak seçilmiştir ve tahmin vektörünü elde etmek için optimize edilmiş SVM kullanılmıştır [122]. Önerilen BP19 modeli, hesaplama verimliliğini korurken yüksek sınıflandırma etkinliği elde etmek için tasarlanmıştır. Metodolojik çerçeve 4 aşamadan oluşmaktadır:

TQWT kullanarak Özellik Çıkarma

TQWT ses sinyallerine uygulanarak 18 dalgacık alt bandına ayrıştırılır. TQWT'de kullanılan parametreler şu şekildedir.

- **Q faktörü (q):** 2
- **Artıklık değeri (r):** 3
- **Ayrıştırma seviyesi sayısı (L):** 17

Tek Boyutlu İkili Örüntü (BP) Analizi ile Özellik Mühendisliği

BP analizi ham ses sinyaline ve 18 dalgacık alt bandına uygulanmıştır. BP çıkarma işlemi şunları içermektedir:

- Sinyalin, her biri 9 örnek uzunluğunda olan örtüşen bloklara bölünmüştür.

- Her bloktaki merkezi değere ve komşu değerlere dayalı ikili özellik temsilleri oluşturulmuştur.
- Bir özellik haritası sinyali oluşturmak için ikili özellik temsillerinin ondalık değerlere dönüştürülmesi işlemi gerçekleştirilmiştir.
- Özellik haritası sinyalinden histogramların çıkarılması ve sinyal başına 256 boyutlu bir özellik vektörü elde edilmesi sağlanmıştır.

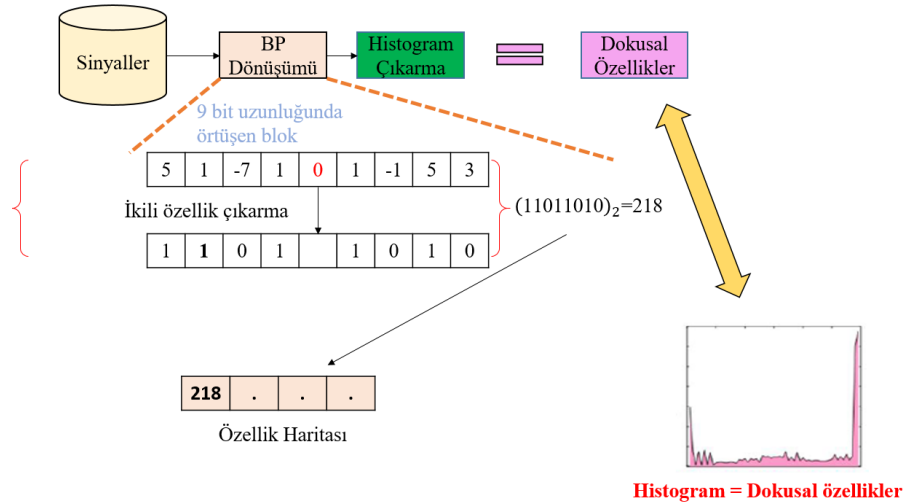
Tüm alt bantları içeren birleştirilmiş özellik vektörü, 4864 ($=256 \times 19$ sinyal) nihai özellik vektörü boyutunu vermiştir.

INCA Kullanılarak Özellik Seçimi

INCA, çıkarılan 4864 boyutlu özellik vektöründen en önemli 966 özelliği seçmek için kullanılmıştır. Bu adım, hesaplama karmaşıklığını azaltmaya ve sınıflandırma performansını artırmaya hizmet etmiştir.

Optimize Edilmiş SVM Kullanılarak Sınıflandırma

Sınıflandırma için Gauss çekirdeğine sahip bir SVM kullanılmıştır. Model, 10 kat çapraz doğrulama tekniği kullanılarak eğitilmiştir. SVM parametrelerine ince ayar yapmak için Bayesian optimizasyonu kullanılarak sınıflandırma doğruluğu en üst düzeye çıkarılmıştır. Bu modelde kullanılan BP özellik çıkarıcının grafiksel açıklaması ise Şekil 6.3'de gösterilmektedir.



Şekil 6.3. BP özellik çıkarıcının grafiksel açıklaması [156]

6.2.6. Çalışmanın Performans Değerlendirmesi

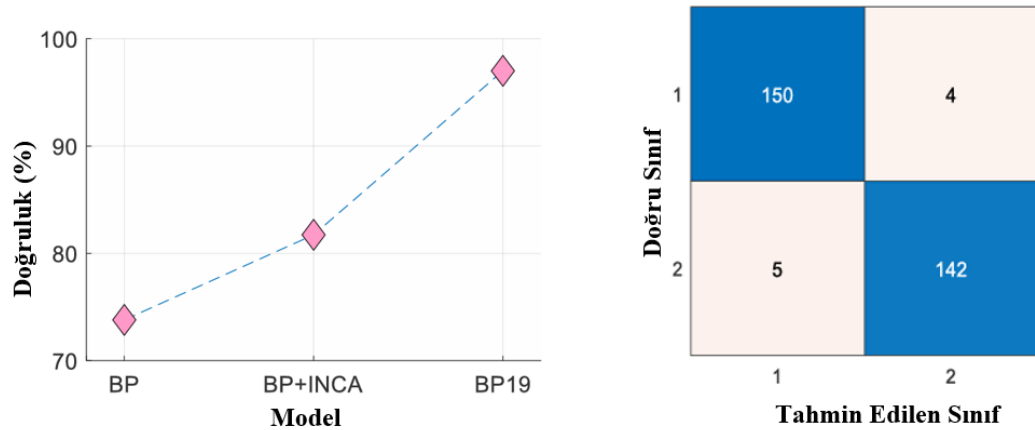
Bu araştırmada, TQWT ve BP tabanlı bir model kullanılmıştır. BP, 19 sinyalden özellikler üretir. Bu nedenle modele VGG19 veya DarkNet19 gibi BP19 adı verilmiştir. Ancak, bu model

hafif bir modeldir. Bu modeli uygulamak için MATLAB (2022a) programlama ortamı kullanılmıştır. Sınıflandırma sonuçlarını elde etmek için herhangi bir grafiksel işlem birimi veya paralel programlama tekniği kullanılmamıştır. Ayrıca BP19 parametrik bir sınıflandırma modelidir. BP19'u oluşturmak için kullanılan parametreler Tablo 6.4'de listelenmiştir.

Tablo 6.4. BP19 üretilirken kullanılan parametreler [156]

Metot	Parametreler	Çıkış
TQWT [157]	q:2, r:3, J:17	18 dalgacık bandı
BP [130]	Örtüşen bloğun uzunluğu: 9, Çekirdek: signum	256 özellik
Önerilen özellik çıkarma metodu	19 giriş (18 dalgacık bandı + 1 ham ses)	4864 özellik
INCA [158]	Döngü aralığı: [100-1000], Sınıflandırıcı: kNN	966 özellik
SVM [159]	Çekirdek: Gauss Çekirdek ölçeği: 86,94, Kutu kısıtlaması: 17,83, Doğrulama: 10 katlı çapraz doğrulama	Sonuçlar
Bayes optimizasyonu [160]	Uygunluk fonksiyonu: Doğruluğu en üst düzeye çıkarma, Yineleme sayısı: 1000	Optimize edilmiş SVM parametresi

Önerilen BP19, Tablo 6.4'de listelenen bu parametreleri dağıtarak oluşturulmuştur. Bu problem ikili bir sınıflandırma problemidir [161]. Bu nedenle, doğruluk, duyarlılık, özgüllük ve geometrik ortalama performans ölçütleri kullanılmıştır. Bu performans ölçütlerini hesaplamak için, Şekil 6.4'de gösterilen ablasyon sonuçları ve karışıklık matrisi kullanılmıştır.



Şekil 6.4. BP19 için ablasyon sonuçları ve karışıklık matrisi [156]

Önerilen BP19 modeli, seçilmiş görsel-işitsel şiddet tespiti (AVD) veri kümesinde %97,01'lik bir sınıflandırma doğruluğu göstermiştir. Modelin etkinliğini gösteren diğer performans ölçümleri Tablo 6.5'de gösterilmiştir.

Tablo 6.5. BP19 modelinde hesaplanan performans değerlendirme sonuçları [156]

Performans metrikler	Sonuçlar (%)
Doğruluk	97.01
Duyarlılık	97.40
Özgüllük	96.60
Geometrik Ortalama	97.00

Özellikle, BP19 modeli hesaplama verimliliği sergileyerek Grafik İşlem Birimi (GPU) hızlandırma ihtiyacını ortadan kaldırmaktadır. Alternatif metodolojilerle yapılan karşılaştırmalı analiz, BP19 modelinin şiddet içeren işitsel olayların tespitinde üstün verimlilik sağladığını ortaya koymuştur.

6.2.7. Çalışmanın Katkıları

Bu araştırma yüksek duyarlılıklı, hesaplama açısından verimli ve kaynak açısından hafif bir sınıflandırma modeli sunarak ses adli bilişimi ve şiddet tespiti alanlarına önemli bir ilerleme sağlamıştır. Yeni bir veri kümesinin oluşturulması, geliştirilmiş performans için ikili örüntü tabanlı bir özellik çıkarma metodolojisinin uygulanması, optimize edilmiş bir SVM sınıflandırıcısının kullanılması, ses tabanlı şiddet tespiti alanındaki boşluğu gidermesi bakımından güvenlik uygulamaları ve adli bilimler alanında gelecekteki araştırmalara değerli katkılar sağlayacaktır.

6.2.8. Çalışmanın Sonuçları ve Tartışma

BP19 modeli, şiddet içeren ve şiddet içermeyen ses örneklerini etkili bir şekilde ayırt etme kabiliyetini göstermiş ve düşük hesaplama yükünü korurken yüksek sınıflandırma doğruluğuna ulaşmıştır. Önerilen BP19, 19 sinyal oluşturmakta ve BP işlevini kullanarak bu sinyallerden özellikler üretmektedir. INCA ile birlikte 10 katlı çapraz doğrulama ile optimize edilmiş bir SVM kullanılarak sınıflandırılan en bilgilendirici 966 özellik seçilmiştir. Modelimiz %97,01 sınıflandırma doğruluğu ve %97 geometrik ortalama elde etmiştir. Ayrıca üretilen sinyallere göre özellikler analiz edilmiştir ve önerilen modelin üstünlüğünü göstermek için ablasyon sonuçları da karışıklık matrisiyle birlikte verilmiştir. Bu çalışmada bir özellik mühendisliği modelinin sınıflandırma yeteneği araştırılmıştır. Sonuçlar ve bulgular, şiddet sesinin basit bir modeliyle yüksek sınıflandırma performansı elde edildiğini göstermektedir. Bu model, gerçek dünyadaki adli uygulamalara ve çeşitli ses sınıflandırma zorluklarına sorunsuz entegrasyon potansiyeline sahiptir.

Gelecekteki araştırma yönleri arasında veri kümesinin genişletilmesi ve otomatik görsel-işitsel şiddet algılama (AVD) sistemleri için kendi kendini organize eden özellik çıkarma modellerinin geliştirilmesi yer almaktadır.

6.3. James Webb Modeline Dayalı Doğru Bir Otomatik Konuşmacı Sayma Mimarisi

Ses adli bilişimi alanı, konuşmacı sayımını büyük önem taşıyan bir araştırma alanı olarak kabul etmektedir. Ancak kamuya açık veri kümelerinin azlığı ve bunların toplanmasındaki doğal zorlukların bir sonucu olarak ilerleme kısıtlanmıştır. Bu araştırma, özellikle konuşmacı sayma metodolojilerini geliştirmek için örtüşen konuşma sinyallerinden oluşan yeni, büyük ölçekli bir veri kümesi oluşturup yeni bir özellik mühendisliği modeli sunarak bu sınırlamaları aşmayı amaçlamaktadır. Bu model, James Webb Uzay Teleskobunun ikonik ayna bölümlerinden esinlenerek James Webb Örüntüsü (JWPat) olarak adlandırılan benzersiz bir dokusal özellik çıkarma tekniği içermektedir [162].

Yeni bir araya getirilen veri kümesi, sistematik olarak 32 sınıfa ayrılmış 3.121 konuşma kaydından oluşuyor ve her sınıf etiketi mevcut konuşmacıların kesin sayısını göstermektedir. Derin öğrenme modellerinin özelliklerini taklit eden yeni bir sınıflandırma çerçevesi önerilmiştir. Bu çerçeve, çok seviyeli özellik çıkarımı (düşük ve yüksek seviyeli özellikler) için çeşitli ana dalgacık fonksiyonlarından yararlanarak özellik mühendisliğine yönelik kendi kendini organize eden bir yaklaşım kullanmaktadır.

Ampirik değerlendirme, symlet4 ana dalgacık fonksiyonu kullanıldığında %86,74'lük bir maksimum sınıflandırma doğruluğu sağlamıştır. Genel olarak sekiz sınıflandırma deneyi %75,94 ile %86,74 arasında değişen doğruluklar üretmiştir. Yüzde 10 puanın üzerindeki bu önemli varyasyon, ana dalgacık seçiminin model performansındaki kritik rolünü vurgulamaktadır. Ayrıca, bu bulgular JWST'den esinlenen yaklaşımın etkili özellik çıkarma yeteneklerini göstermektedir [163]. Bu kapsamlı veri kümesinde bildirilen %86,74'lük doğruluk, önerilen modelin etkinliğini kanıtlamaktadır.

6.3.1. Literatür Taraması Özeti

Bu makale, ses adli bilişimi ve konuşma işleme alanlarında kritik bir zorluk olan konuşmacı sayma problemini ele almaktadır. Çalışma, gelişmiş konuşma teknolojisi uygulamaları için örtüşen konuşma algılama (OSD) ve konuşmacı saymanın öneminin altını çizmektedir; ancak bu alanlar kapsamlı, titizlikle açıklanmış veri kümelerinin azlığı nedeniyle nispeten az keşfedilmiştir [164-166].

6.3.2. Literatürdeki Mevcut Çalışmalar

Mevcut literatür incelendiğinde, çok sayıda çalışmanın derin öğrenme metodolojilerinin uygulanması yoluyla konuşmacı sayma sorununu çözmeye çalıştığı görülmektedir. Bu makalede sunulan Tablo 6.6, kullanılan yöntemleri ve ilgili sınırlamalarını detaylandırarak önceki araştırmaların kapsamlı bir özetini sunmaktadır. CNN'ler ve Çift Yönlü Uzun Kısa Süreli Bellek ağları (BLSTM'ler) dâhil olmak üzere derin öğrenme mimarileri, OSD ve konuşmacı sayma görevleri için kapsamlı bir şekilde kullanılmıştır [167-169].

Önceki araştırmalarda kullanılan veri kümelerinin çoğu, konuşmacı sayma uygulamaları için özel olarak düzenlenmiş olmaktan ziyade, öncelikle OSD için tasarlanmıştır. Önceki çalışmaların kapsamı, kısa zamansal pencerelerin (200 ms-500 ms) kullanılması ve sınırlı sayıda konuşmacıyla (en fazla 5) sınırlandırılmasıyla kısıtlanmıştır [4], [170-177].

Derin öğrenme modelleri, büyük ölçekli veri kümeleri ve yüksek performanslı GPU kaynakları gerektiren önemli hesaplama talepleri sergilemektedir.

Tablo 6.6. Literatürdeki benzer çalışmaların özeti [162]

Yıl	Araştırmacılar	Model Mimarisi	Doğruluk Oranı	Konuşmacı Sayısı Sınırı	Dezavantajlar
2018	Stöter ve ark. [178]	CNN (CountNet)	%98,7	5'e kadar	Uzun işlem süreleri
2020	Peng ve ark. [174]	BLSTM (Spektral + Uzamsal Özellikler)	%83,7	5'e kadar	Düşük doğruluk oranı
2021	Wang ve Wang [174]	CNN (OSD Modeli)	%97	3'e kadar	Maksimum 3 konuşmacı ile sınırlılık

6.3.3. Mevcut Çalışmalardaki Eksiklikler

Bu çalışma, konuşmacı sayma araştırmalarına ilişkin mevcut literatürdeki bazı kritik boşlukları tespit etmiştir:

- **Özel konuşmacı sayma veri kümelerinin eksikliği:** Konuşmacı sayma görevleri için özel olarak düzenlenmiş, kamuya açık veri kümeleri şeklinde dikkate değer bir eksiklik mevcuttur. Birden fazla konuşmacının sayılması için optimize edilmemiş olan OSD veri kümelerine yaygın olarak güvenilmesi önemli bir sınırlama oluşturmaktadır [179].
- **Düşük sınıflandırma doğruluğu:** Mevcut hesaplama modelleri, konuşmacı sayısı beşi aştığında düşük performans sergilemekte ve konuşmacı sayısı arttıkça doğrulukta belirgin bir düşüş görülmektedir.
- **Derin öğrenme modellerinin yüksek hesaplama karmaşıklığı:** kapsamlı veri kümeleri ve yüksek performanslı GPU kaynakları gerektiren derin öğrenme modellerinin doğasında

bulunan önemli hesaplama talepleri, bu modelleri gerçek zamanlı uygulama senaryoları için kullanışsız hale getirmektedir.

- **Özellik mühendisliği metodolojilerinin eksikliği:** Literatürde, el yapımı özellik çıkarma yaklaşımları yerine, derin öğrenme mimarilerine ağırlıklı olarak odaklanıldığı görülmektedir. Ancak el yapımı özellik çıkarma gelişmiş hesaplama verimliliği ve kolay model yorumlanabilirliğini de sunabilmektedir.

6.3.4. Çalışmada Kullanılan Veri Seti

Konuşmacı sayma görevleri için özel olarak hazırlanmış kamuya açık veri kümelerinin tespit edilen eksikliğini gidermek için yazarlarla birlikte özellikle bu çalışma için yeni bir veri kümesi oluşturuldu. Bu veri kümesine ait bilgiler Tablo 6.7’de özetlenmiştir.

Tablo 6.7. JWPat modeli için toplanan veri setine ait bilgiler [162]

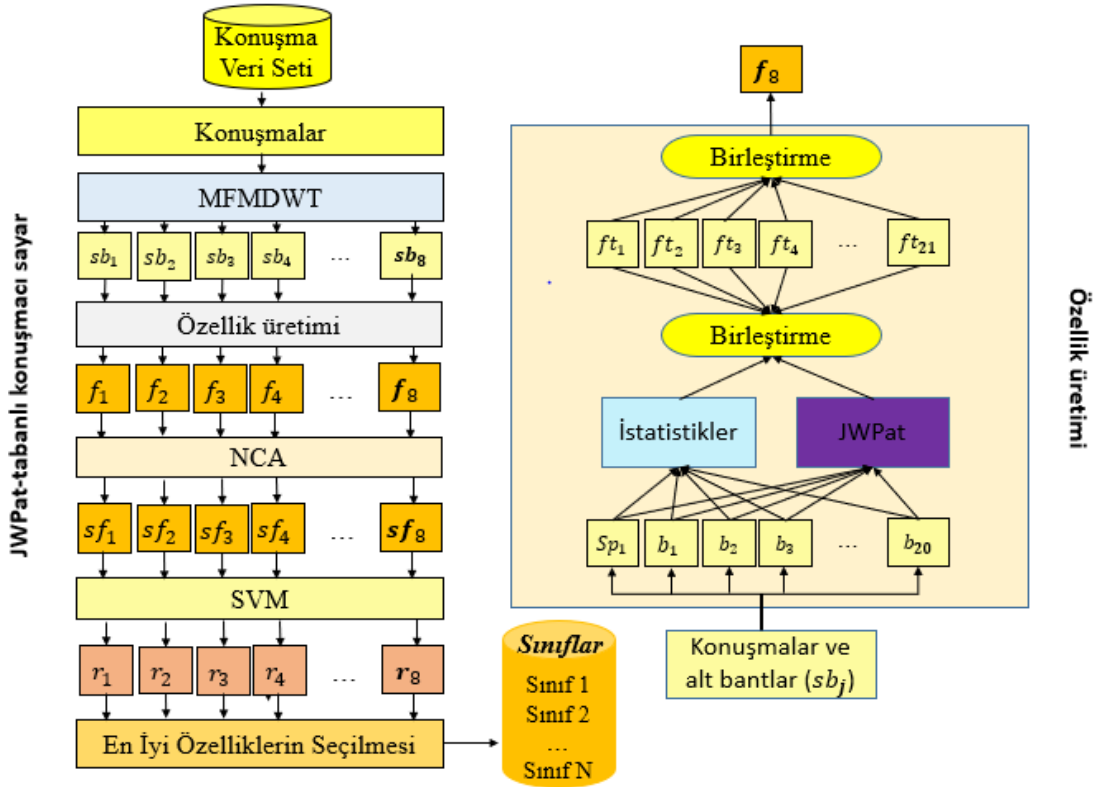
Özellik	Tanım
Toplam Ses Örnekleri	3.121
Sınıf Sayısı	32 (1’den 32’ye kadar konuşmacı sayısını temsil eder)
Sınıf Temsili	Sınıflar 1-14: Doğal Konuşma Kayıtları 15-32. Sınıflar: Sentetik Olarak Üretilen Çakışan Konuşma Bölümleri
Ses Süresi	Klip başına 2–3 saniye
Ses Dosyası Formatı	M4A
Örnekleme Oranı	44,1 kHz
Kaynak	YouTube

Doğal ve sentetik konuşma kompozisyonu: Konuşmacı sayısı sınıfları 1-14, doğal olarak meydana gelen konuşmaların kayıtlarından oluşmaktadır. Konuşmacı sayısı sınıfları 15-32, veri kümesinin sağlamlığını sağlamak için doğal konuşma dosyalarının üst üste bindirilmesiyle sentetik olarak oluşturulmuştur.

Veri kümesinin önemi ve katkıları: Bu veri kümesi, mevcut literatürde bilinen konuşmacı sayma uygulamaları için en kapsamlı külliyyatı temsil etmektedir. Bu alanda gelecekteki araştırma çabalarını kolaylaştırmak ve teşvik etmek için kamuya açık erişim sağlanmıştır. Veri kümesi, 1 ila 32 arasında değişen konuşmacı sayılarının dengeli bir temsili sunmaktadır. Veri setine http://web.firat.edu.tr/turkertuncer/df_speaker_counting_v1.rar adresinden erişilebilmektedir [180].

6.3.5. Çalışmada Önerilen Yöntem

Bu çalışma, derin öğrenme ilkelerinden esinlenen, hesaplama açısından verimli bir metodoloji olan James Webb Örüntüsü (JWPat) özellik çıkarma modelini tanıtmaktadır. Bu modele ait akış diyagramı Şekil 6.5’de gösterilmektedir.



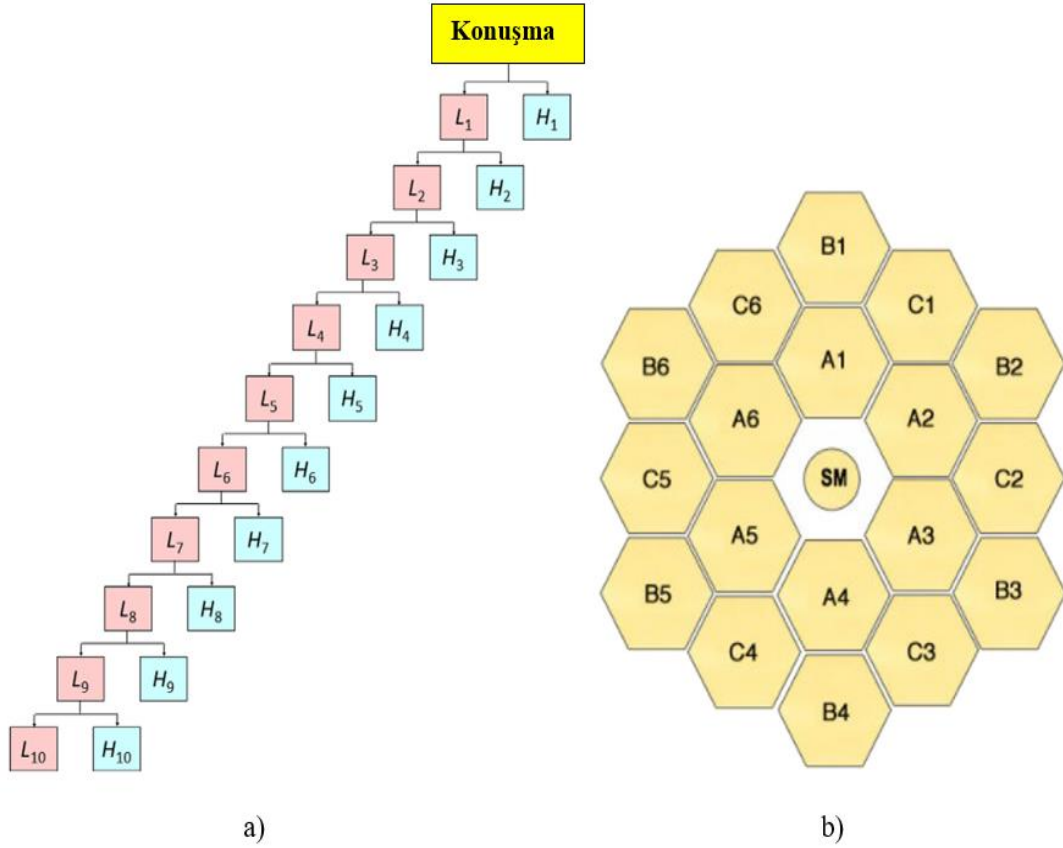
Şekil 6.5. JWPat modeline ait akış diyagramı [162]

Akış diyagramından da anlaşılacağı üzere model üç farklı aşama olarak yapılandırılmıştır:

Özellik Çıkarma

- James Webb Uzay Teleskobu'nun altıgen ayna dizisinden ilham alan James Webb Deseni (JWPat) algoritması, dokusal özelliklerin çıkarılması için kullanılmaktadır.
- İstatistiksel özellikler, ortalama, varyans ve entropi dâhil olmak üzere 20 istatistiksel moment kullanılarak sinyal başına toplam 40 istatistiksel özellik türetilmiştir.
- Çoklu Filtre Tabanlı Çok Seviyeli Ayrık Dalgacık Dönüşümü (MFMDWT), ses sinyallerini yüksek ve düşük frekanslı bileşenlere ayırmak için 8 ana dalgacık fonksiyonundan oluşan bir paket ile uygulanmıştır. Symlet 4 (sym4), Symlet 5 (sym5), Symlet 6 (sym6), Daubechies 4 (db4), Daubechies 5 (db5), Daubechies 6 (db6), Haar ve Biorthogonal 3.5 (bior 3.5) [181]. Bu yaklaşım, farklı frekans bantlarında hem üst düzey hem de alt düzey özelliklerin çıkarılmasını kolaylaştırmaktadır [127], [154], [182]. JWPat

tabanlı dokusal özellik çıkarımı için dört gruba yapılandırılabilen James Webb uzay teleskop aynası ve 10 seviyeli yüksek ve düşük geçiş filtresinin oluşturulması adımları Şekil 6.6’da sunulmuştur.



Şekil 6.6. a) 10 seviyeli dengesiz ağaç yapısı MDWT, 10 düşük geçişli filtre (L) ve 10 yüksek geçişli filtre (H) dalgacık katsayısı üretmek için kullanıldı. b) James Webb uzay teleskopu aynası yapılandırmasına dayalı JWPat. Giriş değeri olarak merkez, SM kullanılır. A, B ve C gruplarındaki altıgen elemanlar, SM etrafında radyal olarak düzenlenir ve saat yönünde ayrı ayrı numaralandırılır [162].

- Nihai Özellik vektörünün oluşturulması, 21 sinyalin her biri için (orijinal sinyal ve 20 dalgacık alt bandı) 424 özellik çıkarılır (384 JWPat özelliği + 40 istatistiksel özellik). Sonuç olarak, özellik seçiminden önceki ilk özellik vektörü 8.904 boyuttan (= 424 özellik $\times$ 21 sinyal) oluşmaktadır.

Özellik Seçimi

NCA, başlangıçtaki 8.904 boyutlu özellik uzayından en ayırt edici 848 özelliği seçmek için kullanılmıştır.

Sınıflandırma

Sınıflandırma için 10 kat çapraz doğrulama paradigması ile birlikte kübik bir çekirdek kullanan bir SVM sınıflandırıcısı kullanılmıştır. Sınıflandırma için özellik setini optimize etmek amacıyla, SVM'nin performansı, her biri MFMDWT aşamasında farklı ana dalgacık filtrelerinin uygulanmasından elde edilen 8 farklı özellik setinde değerlendirilmiştir. Çapraz doğrulama yoluyla en yüksek sınıflandırma doğruluğunu sağlayan özellik seti daha sonra nihai model için seçilmiştir. Bütün bu işlemler MATLAB kullanılarak, Intel Core i7 işlemcili ve 16 GB Rastgele Erişim Belleğine (RAM) sahip standart bir merkezi işlem birimi (CPU) sistemi üzerinde kolayca uygulanmıştır. Önerilen JWPat tabanlı model hesaplama açısından verimlidir.

Kendi Kendini Düzenleyen Model Tasarımı

Model, optimum ana dalgacık fonksiyonunun otomatik seçimi için kendi kendini organize eden bir mekanizma içermekte ve böylece sınıflandırma doğruluğunu en üst düzeye çıkarmaktadır. Bu metodoloji, derin öğrenme mimarilerine olan gereksinimi ortadan kaldırarak onu kaynak kısıtlı ortamlarda kullanıma uygun hale getirmektedir.

6.3.6. Çalışmanın Performans Değerlendirmesi

Önerilen James Webb Örüntüsü (JWPat) tabanlı model, 3.121 konuşma örneğinden oluşan yeni düzenlenmiş veri kümesi kullanılarak değerlendirilmiştir. Temel performans ölçümleri Tablo 6.8'de sunulmuştur.

Tablo 6.8. JWPat performans değerlendirme sonuçları [162]

Model	Doğruluk (%)	Duyarlılık (%)	Kesinlik (%)	F1-Skor (%)
JWPat + SVM (en iyi)	86.74	86.62	86.99	86.80
JWPat + SVM (en kötü)	75.94	75.83	76.93	76.38

Ayrıca her ana dalgacık fonksiyonu için önerilen modelin (JWPat tabanlı model) genel sınıflandırma sonuçları (MFMDWT'de sekiz filtre kullanıldı) performans üzerindeki etkisini daha iyi anlamak için Tablo 6.9'da gösterilmektedir.

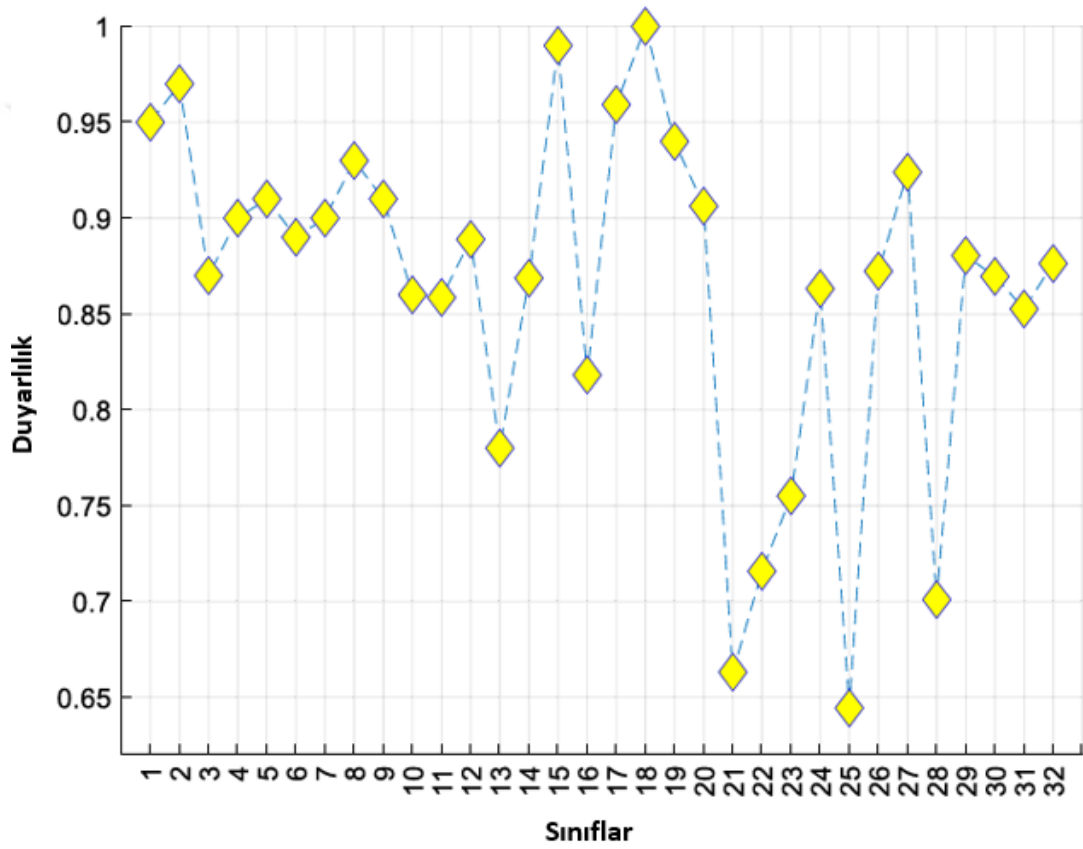
Tablo 6.9. JWPAT modelinde farklı ana dalgacık fonksiyonu kullanılarak elde edilen performanslar [162]

Filtre	Doğruluk (%)	Duyarlılık (%)	Kesinlik (%)	F1-Skor (%)
Sym4	86.74	86.62	86.99	86.80
Sym5	83.76	83.67	84.36	84.02
Sym6	83.37	83.30	83.83	83.57
Db4	82.31	82.23	83.20	82.72
Db5	83.05	82.94	83.49	83.22

Db6	84.56	84.49	85.19	84.84
Haar	75.94	75.83	76.93	76.38
Bior 3.5	81.80	81.70	82.56	82.13

Tablo 6.9, sym4 ana dalgacık fonksiyonu kullanılarak elde edilen en iyi performansı göstermektedir.

En düşük performans Haar ana dalgacık fonksiyonu kullanılarak elde edilmiştir. En iyi sonucun sınıf bazında doğrulukları Şekil 6.7'de gösterilmiştir.



Şekil 6.7. Duyarlılığın sınıf sayısına göre grafiği [162]

JWPat modeli, 32 konuşmacıya kadar işleme kapasitesine rağmen üstün sınıflandırma doğruluğu (%86,74) göstermiş ve maksimum 3-5 konuşmacı ile kısıtlanan önceki modellerin performansını aşmıştır. Sym4 ana dalgacık fonksiyonu en iyi performansı verirken, Haar dalgacık fonksiyonu optimal olmayan sonuçlar sergilemiştir.

6.3.7. Çalışmanın Katkıları

Bu araştırma çalışması, Fırat Üniversitesi Bilimsel Araştırma Projesi tarafından sağlanan TEKF.22.11 proje fonu tarafından desteklenmiştir. Ayrıca Konuşmacı sayma ve adli bilişim alanına çeşitli yeni katkılar sağlamaktadır:

- **Yeni bir veri kümesinin tanıtımı:** Konuşmacı sayımı için halka açık ilk veri kümesidir. 32 sınıfta 3.121 konuşma örneği içerir, bu da onu alandaki en büyük veri kümesi yapmaktadır.
- **Yeni bir özellik çıkarma modelinin geliştirilmesi:** James Webb Uzay Teleskobu'nun ayna tasarımıyla esinlenerek JWPat tabanlı el yapımı özellik çıkarımı önerildi. Yüksek ve düşük seviyeli özellikler oluşturmak için kendi kendini organize eden bir Çoklu Filtre Tabanlı Çok Seviyeli Ayrık Dalgacık Dönüşümü (MFMDWT) tanıtıldı.
- **Hesaplama açısından verimli model:** Derin öğrenmenin aksine GPU gerektirmez. Özellik mühendisliği ve SVM kullanır, bu da onu hafif ve yorumlanabilir hale getirmektedir. Bu önerilen modelin literatürdeki diğer çalışmalarla karşılaştırmalı sonuçlarının bir özeti de Tablo 6.10'da gösterilmektedir.

Tablo 6.10. Otomasyon için en son yöntemlerle elde edilen sonuçların karşılaştırılması [162]

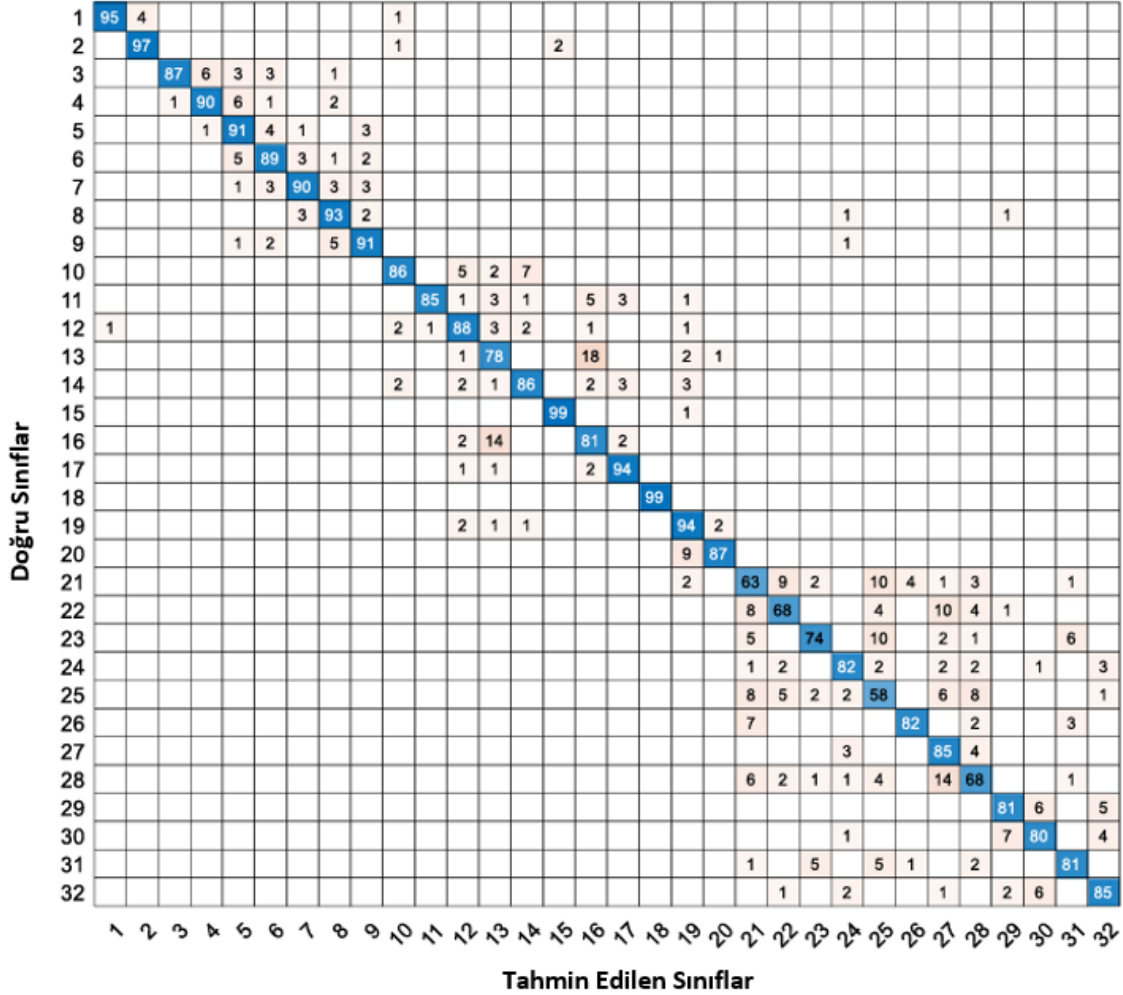
Çalışmalar	Metot	Maksimum Konuşmacı Sayısı	Doğruluk (%)
Yousefi & Hansen (2021) [175]	Spektrogram çıkarma + CNN	3	76.15
Andrei ve ark. (2017) [171]	Tek boyutlu Evrişimsel Sinir Ağı (1D CNN)	3	80.2
Grumiaux ve ark. (2021) [176]	Evrişimsel Tekrarlayan Sinir Ağları	5	51.43
Zhang ve ark. (2019) [183]	Özelleştirilmiş CNN	4	76.66
Önerilen JWPat modeli	Özellik mühendisliği + SVM	32	86.74

- **Literatürdeki en yüksek doğruluk:** CNN tabanlı yöntemlerden daha iyi performans göstererek %86,74 doğruluk elde etmiştir. Mevcut yöntemler 3-5 konuşmacı ile sınırlıyken önerilen yöntemimizde 32 konuşmacıya kadar işlenebilmektedir.

6.3.8. Çalışmanın Sonuçları ve Tartışma

Bu araştırmanın bulguları, mevcut derin öğrenme merkezli metodolojilere kıyasla üstün performans gösteren bir özellik mühendisliği yaklaşımının konuşmacı sayma alanındaki etkinliğinin altını çizmektedir. Özellikle, Sym4 ana dalgacık fonksiyonunun kullanımı en olumlu sonuçları vermiş ve gözlemlenen en yüksek sınıflandırma doğruluğu ile sonuçlanmıştır. Önerilen

JWPat modeli derin öğrenme metodolojilerine ihtiyaç duymadan adli ses analizi, konuşmacı günlüğü oluşturma ve akıllı ev sistemleri bağlamı dâhil olmak üzere bir dizi gerçek dünya senaryosunda umut verici bir uygulanabilirlik sergilemektedir. Bu önerilen modeldeki çoklu sınıflandırma problemi ilgili karışıklık matrisi Şekil 6.8’da gösterilmektedir.



Şekil 6.8. Yeni bir veri seti kullanılarak önerilen JWPat tabanlı konuşmacı sayma modeli için elde edilen karışıklık matrisi [162].

Gelecekteki araştırma çalışmaları, sınıflandırma çerçevesinin sürekli optimizasyonunun yanı sıra modelin sağlamlığını ve genelleştirilebilirliğini daha da artırmak için veri kümesinin genişletilmesine odaklanacaktır.

6.4. SKLBP14: Kare Çekirdekli Yerel İkili Örüntüye Dayalı Yeni Bir Dokusal Çevresel Ses Sınıflandırma Modeli

Bu araştırma, yeni bir dokusal ESC veri kümesi sunarak ve Kare Çekirdekli Yerel İkili Örüntü (SKLBP) tabanlı yeni bir özellik mühendisliği modeli önererek ESC metodolojisine katkıda bulunmaktadır. SKLBP, son zamanlarda popüler olan İleri-İleri (FF) algoritmasında öne çıkan kare tabanlı aktivasyon fonksiyonundan esinlenen yeni bir LBP çekirdeğidir.

Seçilmiş ESC veri kümesi, sınıf başına 100 ses içeren 15 sınıf içerecek şekilde dengelenmiştir ve önerilen modelin sınıflandırma yeteneklerini değerlendirmek için kullanılmıştır. SKLBP tabanlı model mimarisi, DWT aracılığıyla çok seviyeli bir özellik çıkarma yaklaşımı kullanarak derin öğrenme yapılarını taklit etmektedir. Yinelemeli bir özellik seçici, çıkarılan özellikleri iyileştirir ve bir kNN sınıflandırıcısı nihai sınıflandırmayı gerçekleştirir. Bu yaklaşım %94'lük bir sınıflandırma doğruluğu sağlayarak SKLBP tabanlı özellik mühendisliğinin ESC görevleri için potansiyelini ortaya koymuştur [184].

6.4.1. Literatür Taraması Özeti

Bu çalışma, farklı çevresel bağlamlardan kaynaklanan akustik sinyallerin sınıflandırılması için makine öğrenimi tekniklerinden yararlanan bir alan olan ESC alanını incelemektedir. Araştırma, sınıflandırma doğruluğu ve hesaplama kaynağı harcaması arasında makul bir denge sağlayan verimli özellik çıkarma metodolojilerinin geliştirilmesi için kritik gereksinimi vurgulamaktadır [185-189].

6.4.2. Literatürdeki Mevcut Çalışmalar

ESC alanındaki önceki araştırmalar ağırlıklı olarak CNN'ler ve RNN'ler dâhil olmak üzere derin öğrenme modellerinin uygulanmasına odaklanmıştır. Mevcut literatürden kayda değer katkılar şunlardır:

Salamon ve Bello (2017) bir çalışmada, çevresel seslerin sınıflandırılması için veri artırma stratejileri ile birlikte derin evrimsel sinir ağları kullanılmıştır [187]. Abdoli ve arkadaşları (2019) bir çalışmada ESC için tek boyutlu bir Evrimsel Sinir Ağı (1D CNN) modeli önermiş ve yüksek sınıflandırma doğruluğu göstermiştir, ancak ilgili hesaplama maliyetleri önemli boyutlardadır [189].

Demir ve diğerleri (2020): Bu araştırma, çevresel ses sınıflandırması bağlamında etkili ve verimli özellik mühendisliği tekniklerinin zorunluluğuna özel bir vurgu yaparak derin bir CNN modeli uygulamıştır [185].

6.4.3. Mevcut Çalışmalardaki Eksiklikler

Bu çalışma, ESC araştırmalarına ilişkin mevcut literatürdeki bazı kritik boşlukları tespit etmektedir:

- **Hafif özellik mühendisliği yaklaşımlarının sınırlı keşfi:** Literatürde, kapsamlı veri kümeleri ve yüksek performanslı GPU kaynakları gerektiren, hesaplama açısından yoğun derin öğrenme modellerine ağırlıklı olarak odaklanıldığı görülmektedir.
- **Özel dokusal çevresel ses veri kümelerinin eksikliği:** Dokusal çevresel seslerin analizi için özel olarak düzenlenmiş halka açık veri kümeleri şeklinde dikkate değer bir eksiklik mevcuttur.
- **Geleneksel özellik çıkarma metodolojilerinin yetersiz incelenmesi:** LBP gibi el yapımı özellik çıkarma tekniklerinin potansiyeli, ESC bağlamında yeterince araştırılmamıştır.

6.4.4. Çalışmada Kullanılan Veri Seti

Bu çalışmada özellikle dokusal çevresel seslere odaklanan kamuya açık veri kümelerinde tespit edilen eksikliği gidermek için TESC (Dokusal Çevresel Ses Sınıflandırması) olarak adlandırılan yeni bir veri kümesi oluşturulmuştur. Bu veri kümesine ait bilgiler Tablo 6.11’de gösterilmektedir.

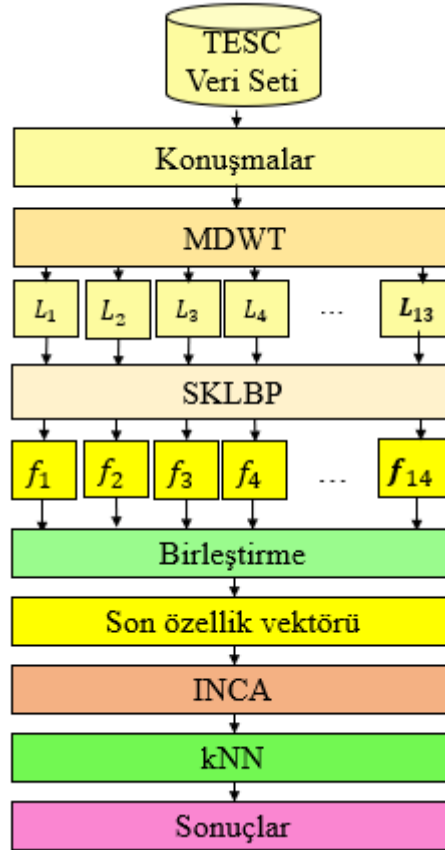
Tablo 6.11. Oluşturulan veri setine ait bilgiler

Özellik	Tanım
Veri Kümesi Adı	Dokusal Çevresel Ses Sınıflandırması (TESC) Veri Seti
Sınıf Sayısı	15
Sınıf Başına Örnekler	100
Toplam Ses Klipleri	1.500
Klip Başına Süre	Yaklaşık 2 saniye
Örnekleme Oranı	44,1 kHz
Veri Kaynakları	Freesound.org, YouTube
Sınıflar	Alkışlar, köpüren su, ateş, kurbağalar, kazlar, helikopter, böcekler, matkap, yağmur, hışırdayan kâğıt, serçeler, akarsu, tren, dalgalar, rüzgâr.

- **Veri kümesinin önemi ve katkıları:** Bu veri seti, dokusal çevresel ses sınıflandırması için özel olarak düzenlenmiş, halka açık ilk kaynağı temsil etmektedir. Veri kümesi, sınıf başına 100 örnekten oluşan dengeli bir dağılım sergilemektedir. Veri setine Kaggle platformu üzerinden herkese açık olarak erişilebilir ve gelecekteki araştırma çabalarını kolaylaştıracaktır (<https://www.kaggle.com/datasets/senguldogan/texturalesc>) [190].

6.4.5. Çalışmada Önerilen Yöntem

Bu çalışma, İleri-İleri (FF) algoritmasından esinlenerek ESC için hesaplamalı olarak verimli bir el yapımı metodoloji olan Kare Çekirdekli Yerel İkili Örüntü özellik çıkarma modelini tanıtmaktadır. Bu çalışmada önerilen modelin genel akış diyagramı Şekil 6.9’da gösterilmektedir.



Şekil 6.9. Önerilen modelin şematik gösterimi, MDWT: çok seviyeli ayrık dalgacık dönüşümü, L: alçak geçiren filtre dalgacık katsayıları (dalgacık bantları), F: 256 boyutlu özellik vektörleri [184]

Yukarıda şematik olarak açıklanan modelimiz dört temel aşamadan oluşmaktadır:

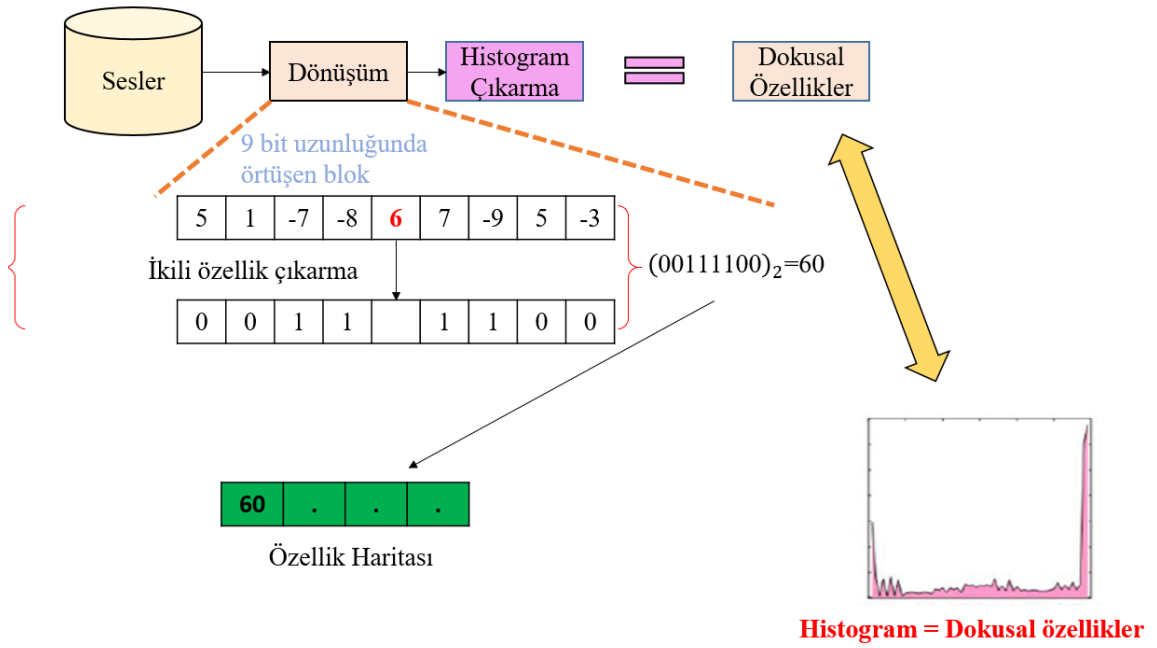
Özellik Çıkarma

Özellik çıkarma işlemi, tek boyutlu ses sinyalleri için tasarlanmış LBP’nin değiştirilmiş bir çeşidi olan Kare Çekirdekli Yerel İkili Örüntü (SKLBP) algoritması kullanılarak gerçekleştirilmiştir. SKLBP algoritması, dokusal özellikleri yakalamak için kare tabanlı bir aktivasyon fonksiyonu kullanır. SKLBP çıkarma işlemi şunları içermektedir:

- Ses sinyalinin, her biri 9 örnekten oluşan tanımlanmış bir boyuta sahip örtüşen bloklara bölünmüştür.

- Bir kare çekirdek aktivasyon fonksiyonunun uygulanması yoluyla ikili özellik temsillerinin oluşturulması sağlanmıştır.
- Bir özellik haritası sinyali üretmek için ikili özellik temsillerinin ondalık değerlere dönüştürülmesi gerçekleştirilmiştir.
- Özellik haritası sinyalinden bir histogram çıkarılarak 256 boyutlu bir özellik vektörü elde edilmiştir.

Özellik çıkarma işlemine ait akış diyagramı Şekil 6.10'da gösterilmektedir.



Şekil 6.10. Önerilen SKLBP özellik üreticinin genel grafiksel görünümü [184]

Özellik mühendisliği

- Çok Seviyeli Ayrık Dalgacık Dönüşümü (MDWT), 13 ayrıştırma seviyesinde Symlet 4 ana dalgacığı kullanılarak ses sinyallerine uygulanmıştır [122], [126].
- Bu işlem, ham sinyali ve 13 dalgacık alt bandını kapsayan toplam 14 özellik vektörü vermiştir.
- Nihai birleştirilmiş özellik vektörü boyutu 3.584'tür ($=256 \text{ özellik} \times 14 \text{ sinyal}$).

Özellik Seçimi

Başlangıçtaki 3.584 boyutlu özellik uzayından en belirgin 449 özelliği seçmek için INCA uygulanmıştır. Bu adım, sınıflandırma doğruluğunu korurken hesaplama karmaşıklığını azaltmayı amaçlamıştır.

Sınıflandırma

Sınıflandırma görevi, k parametresi 1 olarak ayarlanmış bir kNN sınıflandırıcısı ve Manhattan mesafe metriği kullanılarak gerçekleştirilmiştir. Model performansı 10 kat çapraz doğrulama metodolojisi kullanılarak değerlendirilmiştir.

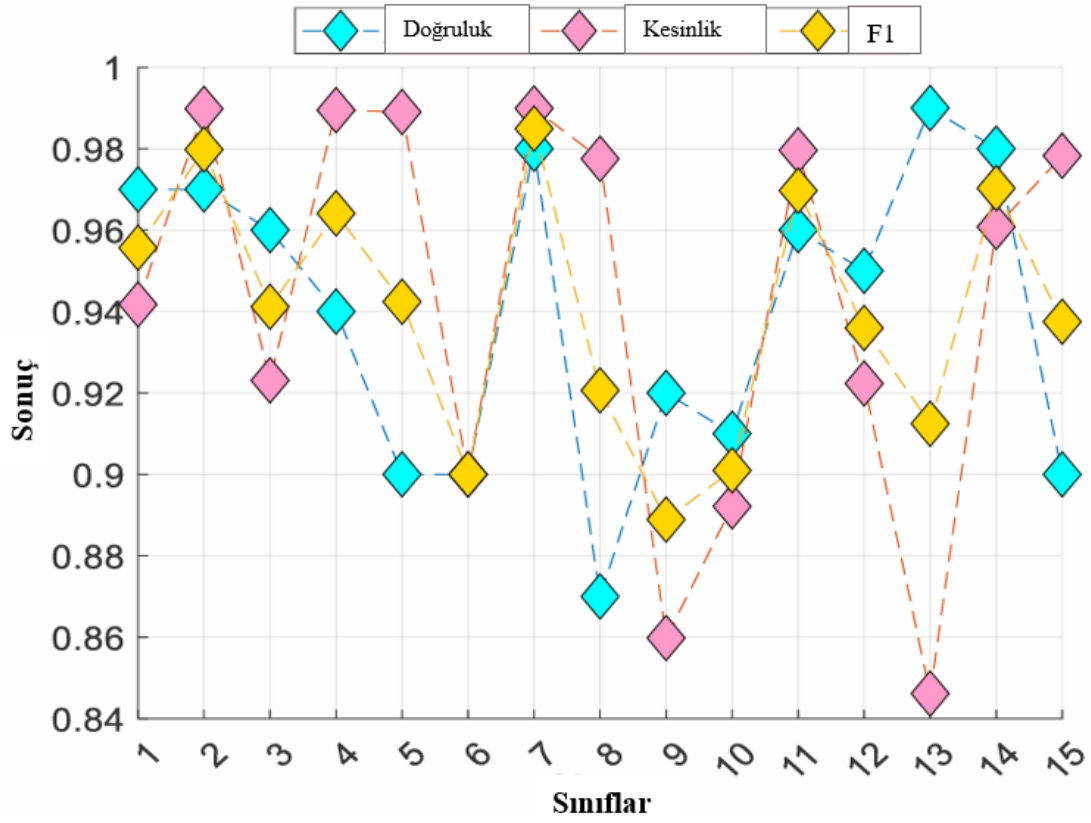
6.4.6. Çalışmanın Performans Değerlendirmesi

Önerilen SKLBP tabanlı model, belirtilen parametreler kullanılarak uygulanmıştır. Modelin 15 sınıflı veri kümesi üzerindeki performansını kapsamlı bir şekilde değerlendirmek için çok yönlü bir değerlendirme stratejisi kullanılmıştır. Genel sınıflandırma performansı doğruluk, ağırlıklandırılmamış ortalama kesinlik (UAP) ve genel F1-skoru kullanılarak ölçülmüştür; bu toplam ölçütler, problemin çok sınıflı yapısı göz önüne alındığında oldukça uygundur. Bu sonuçlar Tablo 6.12’de gösterilmektedir.

Tablo 6.12. Önerilen SKLBP tabanlı dokusal ESC modelinin performans sonuçları (%) [184]

Metrik	Sonuç (%)
Doğruluk	94.00
Ağırlıklandırılmamış Ortalama Kesinlik (UAP)	94.26
F1-Skor	94.03

Ayrıca modelin bireysel kategorilerdeki sınıflandırma etkinliğinin ayrıntılı bir şekilde anlaşılmasını sağlamak için sınıfa özgü performans ölçümleri de hesaplanmıştır. Bu ölçümler Şekil 6.11 ve Şekil 6.12’de karışıklık matrisi ile birlikte sunulmuştur. Bu değerlendirme ölçütlerinin temelini oluşturan matematiksel tanımlar referanslarda ayrıntılı olarak verilmiştir [191, 192].



Şekil 6.11. Önerilen SKLBP tabanlı TESC modelinin sınıf bazında sınıflandırma performansları [184]

1	97									2				1	
2		97								1		1	1		
3			1	96								2	1		
4				1	94					5					
5	3				1	90	1			2	1		2		
6	2					1	90			1			4	1	1
7							98	1						1	
8						2		87	5	1		1	4		
9						3			92			1	2	1	1
10	1		1						3	91	2	2			
11							1	1	1	1	96				
12				5								95			
13				1									99		
14													2	98	
15						4			3			1	2		90
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15

Tahmin Edilen Sınıflar

Şekil 6.12. Önerilen SKLBP tabanlı dokusal ESC modelinin sınıfa özgü karışıklık matrisi [184]

Bu sonuçlara göre en iyi doğruluğa sahip olan sınıfın 13. sınıf olduğunu (sınıf bazında %99 doğruluk oranına ulaşmıştır) ve bu sınıfın tren sesini içerdiğini ve en kötü sınıfın 8. sınıf olduğunu (sınıf bazında %87 doğruluk elde edilmiştir) ve bu sınıfın bir matkap olduğunu göstermiştir. Dahası, bütün sınıflar göz önüne alındığında sınıf bazında sonuçlar %84'ten yüksektir. Önerilen modelin benzer çalışmalarla karşılaştırılması Tablo 6.13’de sunulmuştur.

Tablo 6.13. Önerilen SKLBP modelinin karşılaştırmalı sonuçları

Çalışma	Yöntem	Doğruluk (%)
Bu çalışma (SKLBP) [184]	Özellik Mühendisliği+kNN	94.00
Salomon ve Bello (2017) [187]	Deep CNN	87.50
Abdoli ve ark. (2019) [189]	1D CNN	91.20

6.4.7. Çalışmanın Katkıları

Bu çalışma, ESC alanına birkaç önemli katkı sunmaktadır.

Yeni Bir Veri Setinin Tanıtımı: Dokusal çevresel ses analizi için özel olarak düzenlenmiş ilk halka açık kaynağı temsil eden TESC veri setinin oluşturulması ve kamuya açıklanması. Veri seti, 15 farklı çevresel ses sınıfı arasında dengeli bir dağılım sergilemektedir.

SKLBP14 özellik çıkarma modelinin geliştirilmesi: Ses sinyallerinin analizi için özel olarak tasarlanmış kare çekirdekli yerel ikili örüntü (SKLBP) algoritmasını içeren ve SKLBP14 olarak adlandırılan yeni bir el yapımı özellik çıkarma modelinin tanıtımı gerçekleştirilmiştir. Önerilen model, derin öğrenme tabanlı yaklaşımların aksine, Grafik İşlem Birimi (GPU) hızlandırma gerekliliğini ortadan kaldıran hesaplama verimliliği ile karakterize edilmiştir.

Çok seviyeli özellik çıkarma stratejisinin uygulanması: 13 ayrıştırma seviyesinde MDWT kullanan çok seviyeli bir özellik çıkarma yaklaşımının uygulanması gerçekleştirilmiştir. Bu strateji, ses sinyallerinin hem zamansal hem de frekans alanlarından bilgilendirici özelliklerin çıkarılmasını sağlamaktadır [122].

Hesaplama yükü ile yüksek sınıflandırma doğruluğunun elde edilmesi: SKLBP14 modeli, mevcut derin öğrenme modellerine kıyasla üstün performans göstererek %94'lük bir sınıflandırma doğruluğu elde edilmiştir. Derin öğrenme mimarileri tarafından tipik olarak kullanılan binlerce özelliğin aksine yalnızca 449 özelliğin kullanılması, eğitim süresinde ve hesaplama maliyetinde önemli bir azalma sağlamıştır.

6.4.8. Çalışmanın Sonuçları ve Tartışma

Deneysel değerlendirme, önerilen SKLBP14 modelinin “Tren” ses sınıfı için en yüksek sınıflandırma doğruluğunu elde ettiğini ve %99'luk bir hassasiyete ulaştığını ortaya koymuştur.

Buna karşılık, “Matkap” ses sınıfı %87’lik bir doğrulukla en düşük sınıflandırma performansını sergilemiştir. Ayrıca SKLBP14 modeli hesaplama verimliliği ve yorumlanabilirliği ile karakterize edilerek gerçek zamanlı ESC uygulamalarında kullanılmaya çok uygun hale getirilmiştir. Gelecekteki araştırma çalışmaları, TESC veri setinin temsil gücünü artırmak için genişletilmesini ve karşılaştırmalı analiz ve potansiyel performans iyileştirmeleri için tek boyutlu Yerel İkili Örüntüleri (1D-LBP) içeren bir derin öğrenme modelinin geliştirilmesini kapsayacaktır.

Özetle bu araştırma, SKLBP14 özellik çıkarma modelinin tanıtılmasıyla ESC alanında önemli bir ilerleme sunmuştur. Bu metodoloji, sınıflandırma doğruluğunda belirgin bir iyileşme göstererek %94’lük bir performans seviyesine ulaşırken aynı zamanda kayda değer bir hesaplama verimliliği sergilemiştir. Ayrıca, yeni oluşturulan TESC veri seti, çevresel ses analizi alanında gelecekteki araştırma çabalarını kolaylaştırmaya ve katalize etmeye hazır olan değerli ve kamuya açık bir kaynak sunmuştur [190].

6.5. Ses Sinyallerini Kullanan Yeni Bir Ağaç Örüntüsü Tabanlı Şiddet Tespit Modeli

Multimedya verilerini kullanan fiziksel şiddet tespiti, kamu güvenliği ve emniyetinin artırılması için kritik bir araştırma alanını temsil etmektedir. Bu alan bilgi güvenliği ve adli bilişim alanlarında büyük önem taşımaktadır [193-195]. Video tabanlı şiddet tespitine (VVD) odaklanan araştırmalar, video gözetim sistemlerinin dünya çapında yaygınlaşmasının da etkisiyle son yıllarda önemli bir büyüme kaydetmiştir. Buna karşın sözlü saldırganlığı tespit etmeye yönelik teknolojiler, büyük ölçüde videolarda şiddet tespiti için araştırma topluluğu içinde bilgisayarla görme modellerine odaklanması nedeniyle nispeten sınırlı kalmaktadır [196-198]. Bu eşitsizliği ele alan bu çalışma, yeni bir otomatik sesten şiddet algılama (AVD) modelini tanıtmaktadır [23]. Önerilen AVD modeli, çeşitli işleme aşamalarını kapsayan el yapımı bir yaklaşıma dayanmaktadır. Bu çalışma için YouTube’den sözlü saldırganlık örneklerini içeren özel bir ses veri kümesi derlenmiştir. Çok seviyeli özellik çıkarma, özellik seçimi, sınıflandırma ve çoğunluk oylaması aşamalarını içeren yeni bir el yapımı model mimarisi geliştirilmiştir.

Ses sinyallerinden özellik çıkarımı için ikili ağaç konseptine dayanan yeni bir yerel özellik fonksiyonu geliştirilmiştir. Bu fonksiyon ağaç deseni-23 (TreePat23) olarak adlandırılmıştır; burada 23 rakamı ses sinyallerinden türetilen giriş dalgacık bantlarının sayısını ifade etmektedir. Bu dalgacık bantları, özellik üretimi için TreePat23 fonksiyonuna girilmeden TQWT kullanılarak oluşturulmuştur. Özellik seçimi, INCA ve Chi2 istatistiksel testi kullanılarak gerçekleştirilmiştir. Seçilen özellikler daha sonra kNN ve SVM algoritmaları kullanılarak sınıflandırılmıştır [145], [199]. Bu sınıflandırıcılardan elde edilen çıktılar daha sonra yinelemeli çoğunluk oylama (IMV) yöntemi kullanılarak birleştirilmiştir [200-202]. Optimum tahmin edilen vektör, açgözlü bir algoritma uygulanarak belirlenmiştir. Model doğrulaması, bir kaydı dışarıda bırak (LORO) çapraz

doğrulama (CV) olarak bilinen yeni bir teknik kullanılarak gerçekleştirilmiştir. Önerilen TreePat23 modeli, kNN ve SVM sınıflandırıcıları ile sırasıyla %89,68 ve %89,75'lik kayda değer sınıflandırma doğrulukları elde etmiştir. Geliştirilen sistem, her sınıflandırıcı için birden fazla sonuç (bu çalışmada 14) üretme ve değerlendirme yeteneğine sahiptir ve en uygun sonucu otomatik olarak seçmektedir. Bu doğal yetenek, LORO CV stratejisi altında kullanılan her iki sınıflandırıcı için %89'u aşan sınıflandırma doğruluğu sağlayan bir ses sınıflandırma sistemi olarak modelin kendi kendini organize eden doğasının altını çizmektedir [203].

6.5.1. Literatür Taraması Özeti

Bu çalışma özellikle çocuklar ve ergenler üzerindeki potansiyel etkisini vurgulayarak, multimedya içeriğindeki şiddet tespitinin kritik önemini ele almaktadır. Bilgisayarla görme metodolojileri kullanılarak video içeriklerinde şiddet tespiti üzerine önemli sayıda araştırma yapılmış olsa da AVD alanı nispeten daha az incelenmiştir [204]. AVD'ye yönelik mevcut yaklaşımlar ağırlıklı olarak geleneksel makine öğrenimi paradigmalarına dayanmaktadır. Sonuç olarak ses sinyallerinde şiddet sınıflandırmasının doğruluğunu artırmak için daha sağlam özellik çıkarma tekniklerinin geliştirilmesine yönelik önemli bir ihtiyaç vardır.

6.5.2. Literatürdeki Mevcut Çalışmalar

Şiddet tespiti alanı, birkaç önemli çalışmanın da gösterdiği gibi çeşitli metodolojik yaklaşımlarla incelenmiştir:

- **Video tabanlı yaklaşımlar:** Sahay ve diğerleri (2022) ile Khalil ve diğerleri (2021) tarafından yapılanlar da dâhil olmak üzere araştırma çabaları, şiddet içeren davranışları tespit etmek amacıyla video gözetim sistemleri bağlamında derin öğrenme modellerini kullanmıştır [205-206].
- **Ses tabanlı yaklaşımlar:** Giannakopoulos ve diğerleri (2006) ile Lin ve diğerleri (2010) gibi araştırmacılar tarafından yapılan incelemeler hem zaman hem de frekans alanlarından çıkarılan ses özelliklerinin analizi yoluyla şiddet olaylarının sınıflandırılmasına odaklanmıştır [207-208].
- **Hibrit yöntemler:** Bazı çalışmalar, şiddet tespiti için hem işitsel hem de görsel ipuçlarını entegre eden hibrit yaklaşımlar benimsemiştir. Bu yöntemler gelişmiş performans sergilemiş olsa da genellikle daha yüksek bir hesaplama yükü gerektirirler.

6.5.3. Literatürdeki Eksiklikler

Bu çalışma, aşağıdaki temel araştırma eksikliklerini tespit etmiştir:

- **AVD üzerine sınırlı araştırma:** Önceki çalışmaların çoğu video tabanlı şiddet tespitine odaklanırken, ses tabanlı yöntemler yeterince araştırılmamıştır.
- **Özellik çıkarma sorunları:** Birçok çalışmanın, sese şiddet tespiti için uygun olmayabilecek geleneksel özellik çıkarma tekniklerine dayandığı anlatılmıştır.
- **Doğrulama zorlukları:** Rastgele ayırma tabanlı doğrulama yöntemleri yanıltıcı sınıflandırma sonuçlarına yol açabilir, bu da daha güvenilir doğrulama stratejileri gerektirmektedir.

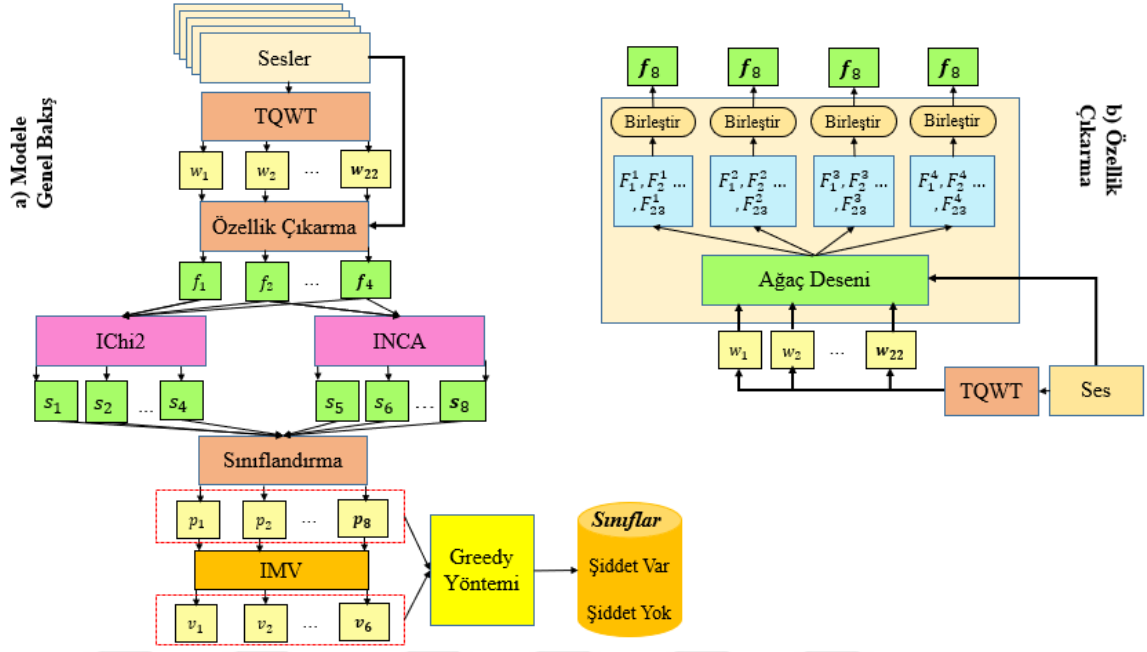
6.5.4. Çalışmada Kullanılan Veri Seti

Bu çalışmada, kamuya açık YouTube videolarından ve film kaynaklarından, ses tabanlı şiddet tespiti üzerine bu çalışma için özel olarak tasarlanmış yeni bir ses veri kümesi oluşturulmuştur.

Bu veri kümesi 46'sı şiddet içeren ve 76'sı şiddet içermeyen olarak etiketlenmiş toplam 122 ham ses dosyası içermektedir. Ses dosyaları mp4 ve m4a dâhil olmak üzere çeşitli formatlarda sağlanmıştır. Her bir orijinal ses dosyası daha sonra 2 saniyelik parçalara bölünerek 746 adet şiddet içeren ve 698 adet şiddet içermeyen örnek olarak kategorize edilmiş toplam 1444 adet ses örneği elde edilmiştir. Çalışmada kullanılan veri setine kamuya açık bir şekilde <https://www.kaggle.com/datasets/senguldogan/violencedetection> adresinden ulaşılabilir [209].

6.5.5. Çalışmada Önerilen Yöntem

Bu çalışma, ses tabanlı şiddet tespiti için yeni bir ağaç örüntüsü tabanlı özellik çıkarma modeli sunmaktadır. Önerilen yöntemin akış diyagramı Şekil 6.13'de gösterilmektedir [23].

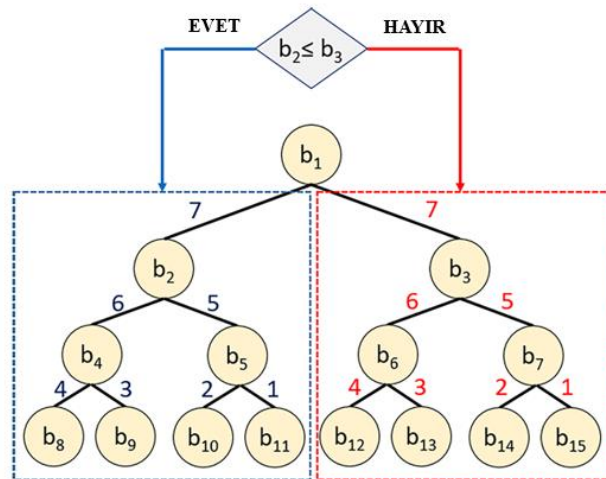


Şekil 6.13. Sunulan TreePat23'ün grafiksel açıklaması (a) modele genel bakış, (b) ağaç desenine dayalı özellik çıkarımı [210].

Önerilen metodoloji dört temel aşamayı kapsamaktadır:

Özellik Çıkarma

Özellik çıkarma işlemi, ham ses sinyallerine uygulanan ikili ağaç tabanlı bir fonksiyon kullanılarak gerçekleştirilmiştir [141], [211]. İkili ağaç kullanımıyla ilgili detaylar Şekil 6.14'de gösterilmektedir.



Şekil 6.14. Önerilen ağaç deseninin genel görünümü. Bu çalışmada, b bloğun değerlerini tanımlar ve kenarlardaki sayılar bit çıkarma sırasını gösterir. ($b_2 \leq b_3$) koşulu kullanılarak kırmızı veya mavi alan seçilir [210].

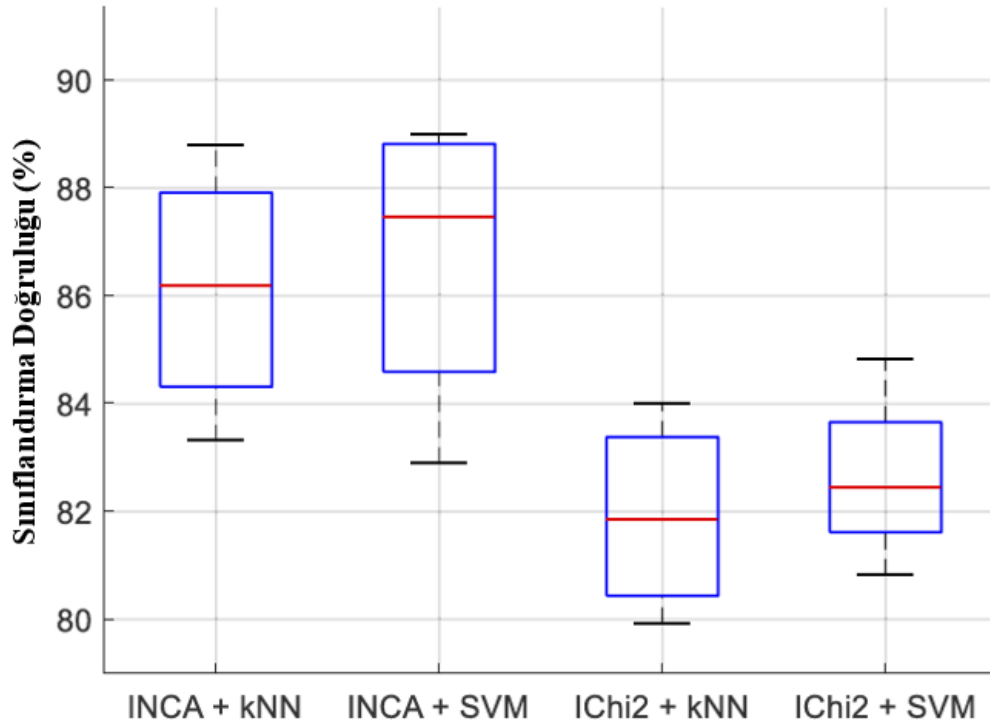
Şekil 6.13’de sunulduğu gibi önerilen TreePat23 mimarisinin ilk aşaması, giriş ses sinyallerinden kapsamlı bir özellik kümesinin çıkarılmasını içerir. Her 2 saniyelik ses segmenti, TQWT aracılığıyla ayrıştırılarak 22 farklı alt dalgacık bandı oluşturulur. Bu 22 alt bant, orijinal ses sinyaliyle birlikte, sonraki özellik çıkarımı için 23 giriş sinyalinden oluşan bir set oluşturmaktadır. Bu 23 sinyalin her biri daha sonra yeni bir ağaç deseni özellik çıkarıcısı tarafından işlenmektedir. Bu özellik çıkarıcı dört farklı çekirdek fonksiyonu kullanılarak uygulanmaktadır: (1) Signum, (2) Üst Üçlü, (3) Alt Üçlü ve (4) Mutlak Üst Üçlü. Her bir çekirdeğin 23 giriş sinyaline uygulanmasıyla her biri 2944 boyutluluğa sahip dört ayrı özellik vektörü ortaya çıkarmaktadır (23 sinyal boyunca sinyal başına çıkarılan 128 özellikten türetilmiştir).

Özellik Seçimi

En göze çarpan özellikleri belirlemek için iki farklı özellik seçim tekniği kullanılmaktadır:

- **INCA:** k-En Yakın Komşu (kNN) sınıflandırma kaybının minimizasyonuna dayanan iteratif bir seçim yöntemidir.
- **Yinelemeli Ki-Kare (IChi2):** Chi-squared istatistiği tarafından belirlenen özelliklerin istatistiksel uygunluğuna dayanan iteratif bir seçim yöntemidir [199], [212].

Özellik çıkarma aşamasında üretilen 2944 boyutlu dört özellik vektörünün her biri bağımsız olarak hem INCA hem de IChi2 özellik seçim süreçlerine tabi tutulmaktadır [213]. Bunun sonucunda, her biri orijinal özellik uzayının azaltılmış ve potansiyel olarak daha bilgilendirici bir alt kümesini temsil eden toplam 8 seçilmiş özellik vektörüyle elde edilen grafiksel sonuçlar Şekil 6.15’de gösterilmektedir.



Şekil 6.15. Özellik seçicilerin ve sınıflandırıcıların sınıflandırma doğrulukları. Bu sonuçlar seçilen sekiz özellik vektörü kullanılarak hesaplanmıştır [210].

Sınıflandırma ve Doğrulama

Özellik seçimi aşamasının ardından, seçilen sekiz özellik vektörünün her biri bağımsız olarak model değerlendirmesi için iki sınıflandırma algoritması kullanılmaktadır:

- **kNN:** k: 1; mesafe: L1-norm (Manhattan mesafesi); ağırlık: yok; doğrulama: LORO CV.
- **SVM:** Çekirdek: 3. derece polinom sırası; çekirdek ölçeği: otomatik; kutu kısıtlaması: 1, doğrulama: LORO CV.

Bu sınıflandırma işlemi, sınıflandırıcı başına toplam sekiz tahmin edilen çıktı vektörü üretir ve her ses segmenti için 8 adet kNN ve SVM'den olmak üzere toplam 16 aday tahmin vektöründen oluşan birleşik bir özellik vektörü oluşturulmuştur.

Geleneksel doğrulama metodolojilerinin doğasında bulunan sınırlamaları ele almak için yeni bir çapraz doğrulama tekniği olan Bir Kaydı Dışarıda Bırakarak Çapraz Doğrulama (LORO CV) önerilmiştir. Bu teknik makine öğrenmesinde özellikle küçük veri setleri veya dengesiz dağılımlı veriler üzerinde model performansını değerlendirmek için kullanılan bir yöntemdir.

Yinelemeli Çoğunluk Oylaması (IMV)

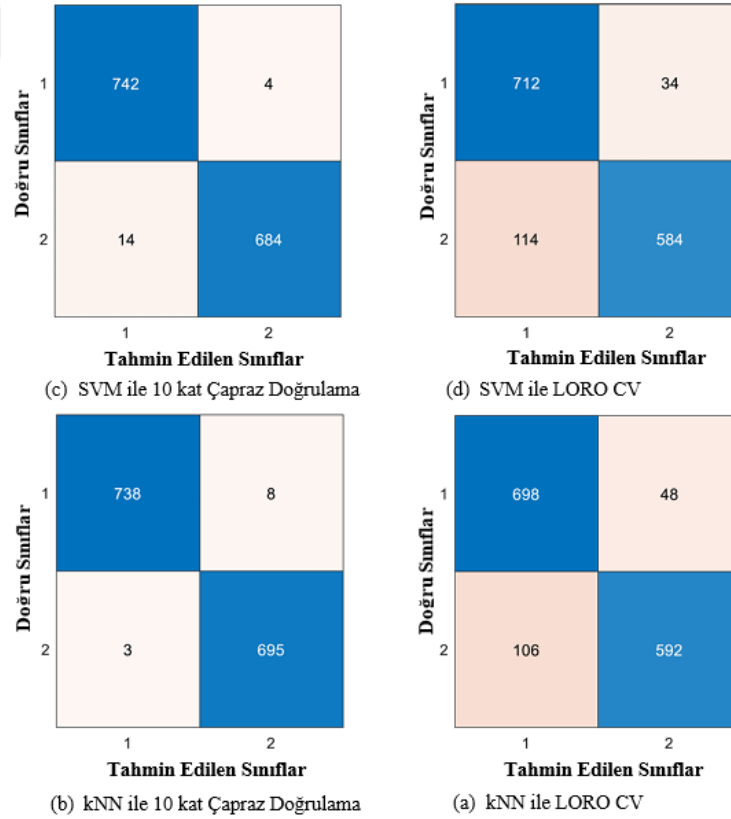
Sınıflandırma sisteminin sağlamlığını ve genel performansını artırmak için, bireysel tahminleri toplamak üzere bir Yinelemeli Çoğunluk Oylaması (IMV) şeması kullanılmıştır [202]. On altı aday tahmin vektöründen, yinelemeli olarak altı ek oylanmış tahmin vektörü oluşturmak

için açgözlü bir seçim algoritması uygulanmıştır. Bu yinelemeli süreç, her yinelemede alt kümenin boyutu artacak şekilde aday tahminlerin alt kümelerini değerlendirir ve her alt kümedeki çoğunluk sınıfı oylanan tahmin olarak belirlenmiştir. Nihai sınıflandırma kararı daha sonra bir doğrulama kümesi üzerinde veya çapraz doğrulama yoluyla değerlendirildiği üzere en yüksek sınıflandırma doğruluğunu sağlayan aday veya oylanmış tahmin vektörü seçilerek verilmiştir.

6.5.6. Çalışmanın Performans Değerlendirmesi

Önerilen ağaç deseni tabanlı modelin performansı, yerleşik performans ölçütleri kullanılarak titizlikle değerlendirilmiştir. Önerilen TreePat23 mimarisinin sınıflandırma performansı LORO CV kullanılarak değerlendirilmiştir.

- **LORO CV kullanımının önemi:** Aynı ses kaydından kaynaklanan segmentleri yanlışlıkla hem eğitim hem de test setlerine bölebilen 10 katlı CV gibi geleneksel k katlı çapraz doğrulama tekniklerinin aksine, LORO CV daha katı bir değerlendirme protokolü kullanmaktadır. Bu çalışmada gerçekleştirilen karşılaştırmalı sonuçlara ait karışıklık matrisi Şekil 6.16'da gösterilmektedir.



Şekil 6.16. Önerilen modelin ince ayarlı sınıflandırıcılar kullanılarak oluşturulan karışıklık matrisleri. **1: şiddet var, 2: şiddet yok [210].

Özellikle, LORO CV her bir tam ses kaydının her iterasyonda özel test seti olarak tutulmasını sağlamaktadır. Böylece eğitim ve test aşamaları arasında veri sızıntısını titizlikle önlemektedir. Bu doğrulama stratejisi, modelin genelleme kabiliyetinin daha gerçekçi ve güvenilir bir şekilde tahmin edilmesini sağlamaktadır, bu da özellikle ses adli bilişimi gibi bireysel kayıtların bütünlüğünün korunması gereken alanlardaki uygulamalar için oldukça önemlidir. Modelin LORO CV ile kNN birlikte kullanılarak ortaya çıkan sınıflandırma sonuçları Tablo 6.14’de, LORO CV ile SVM birlikte kullanılarak ortaya çıkan sınıflandırma sonuçları ise Tablo 6.15’de sunulmuştur.

Tablo 6.14. LORO CV ile kNN birlikte kullanılarak ortaya çıkan sınıflandırma sonuçları (%) [210].

Metrik	Sonuç (%)
Doğruluk	89.68
F1-Skor	90.11
Kesinlik	90.17
Duyarlılık	89.81
Özgüllük	89.54

Tablo 6.15. LORO CV ile SVM birlikte kullanılarak ortaya çıkan sınıflandırma sonuçları (%) [210].

Metrik	Sonuç (%)
Doğruluk	89.75
F1-Skor	90.28
Kesinlik	92.09
Duyarlılık	88.53
Özgüllük	87.25

Temel bulgular şunlardır:

- **Sınıflandırma doğruluğu:** Elde edilen en yüksek sınıflandırma doğruluğu SVM sınıflandırıcısı kullanılarak %89,75 ve kNN sınıflandırıcısı ile %89,68 olmuştur.
- **Karşılaştırmalı metrik analizi:** Model, hassasiyet, geri çağırma, F1-skoru ve özgüllük gibi bir dizi ek değerlendirme ölçütünde üstün performans göstererek genel etkinliğini vurgulamıştır.
- **Doğrulama stratejisi karşılaştırması:** Çalışmanın bulguları, önerilen Bir Kaydı Dışarıda Bırak Çapraz Doğrulama (LORO CV) tekniğinin rastgele veri bölümlmesine dayalı doğrulama stratejilerine kıyasla daha sağlam ve güvenilir sonuçlar verdiğini göstermiştir.

6.5.7. Çalışmanın Katkıları

Bu çalışma, ses tabanlı şiddet tespiti alanına birkaç önemli katkı sunmuştur.

- **Yeni bir özellik çıkarma modelinin tanıtımı:** Ses sinyalleri için özel olarak tasarlanmış yeni bir ağaç deseni tabanlı özellik çıkarma işlevi sunan TreePat23 modelinin geliştirilmesi ve uygulanması sunulmuştur.
- **Yeni bir veri kümesinin sağlanması:** Şiddet tespiti için özel bir ses veri setinin küratörlüğü ve halka açık olarak yayınlanması, böylece gelecekteki araştırma çabaları için değerli bir kaynak sunulmuştur.
- **Doğrulama metodolojisinde ilerleme:** Geleneksel doğrulama prosedürlerinin doğasında bulunan sınırlamaları ele alan Bir Kaydı Dışarıda Bırakarak Çapraz Doğrulama (LORO CV) tekniğinin önerilmesi ve uygulanması gerçekleştirilmiştir.
- **Gösterilen üstün sınıflandırma performansı:** Önerilen model, mevcut metodolojilere kıyasla daha yüksek bir sınıflandırma doğruluğu elde etmekte ve böylece ses verilerindeki şiddetin tespitindeki etkinliğini doğrulamıştır.

6.5.8. Çalışmanın Sonuçları ve Tartışma

Bu araştırmanın bulguları önceki araştırmalarla karşılaştırılmış ve böylece önerilen metodolojinin üstün performansı gösterilmiştir. Ayrıca, ampirik analiz, üst üçlü çekirdeğin değerlendirilen dört çekirdek türü arasında en uygun sınıflandırma doğruluğunu sağladığını ortaya koymuştur.

İnce ayarlı SVM ve kNN sınıflandırıcılarının hiperparametre optimizasyonunu takiben uygulanması, sınıflandırma doğruluğunun artmasıyla sonuçlanmıştır. Sonuç olarak, TreePat23 modeli, AVD için ağaç örüntüsü tabanlı özellik çıkarma tekniklerinin etkinliğini başarıyla doğrulamaktadır.

6.6. Yeni Ses Tabanlı Topluluk Duygu Tespiti için FF-BTP Modeli

Çağdaş duygu sınıflandırma şemaları ağırlıklı olarak bireysel düzeydeki sinyallere odaklanmakta, kalabalıklar içindeki duygusal durumların tanınmasını büyük ölçüde ihmal etmektedir. Bu boşluğu doldurmak için bu araştırma, makine öğrenimi alanında yeni bir meydan okuma olarak Ses Tabanlı Topluluk Duygu Tespitini (SCED) tanıtmaktadır. Bu amaç doğrultusunda, FF-BTP olarak adlandırılan, derin öğrenme ilkelerinden ilham alan, özellikle kalabalık duygularının sesten ayırt edilmesi için uyarlanmış, el yapımı bir özellik mühendisliği modeli geliştirildi [125].

Benzersiz veri kümemiz 187 YouTube videosundan derlenmiş ve her biri 3 saniye süren (44,1 kHz'de örneklenmiş) 2733 segment elde edilmiştir. Üst üste binen konuşma ve ortam sesleri de dâhil olmak üzere karmaşık akustik ortamları yakalayan bu segmentler, kalabalık düzeyinde negatif, nötr ve pozitif duygusal kategorilere titizlikle eklenmiştir.

Önerilen mimari tasarım, dokusal bir özellik çıkarıcı olan BTP'yi Hinton'un Fast Forward algoritmasından esinlenen yenilikçi bir el yapımı özellik seçim mekanizmasıyla bütünleştirmektedir. Bu kombinasyon, hesaplanan ortalama kare hatasına dayalı olarak en belirgin özellik vektörünün tanımlanmasını kolaylaştırmaktadır. Diğer metodolojik geliştirmeler arasında uzamsal ve frekans alanlarındaki özellikleri çıkarmak için çok düzeyli bir ayrık dalgacık dönüşümünün yanı sıra rafine özellik alt kümesi seçimi için sofistike bir yinelemeli komşuluk bileşeni analizinin dâhil edilmesi yer almaktadır. Sınıflandırma daha sonra bir SVM kullanılarak gerçekleştirilmiştir.

Ampirik değerlendirme, FF-BTP modelimizin, geliştirilen veri kümesini kullanarak tanımlanan üç SCED kategorisinde %97,22'lik bir sınıflandırma doğruluğuna ulaştığını göstermektedir. Bu el yapımı yaklaşım, derin öğrenmenin karakteristik derin özellik analizi yetenekleri tarafından bilgilendirilirken, önemli ölçüde daha düşük hesaplama kaynağı gereksinimleri sergilemekte, ancak olağanüstü performans sunmaktadır. Sonuçlar, ses tabanlı topluluktan duygu tanıma üzerine odaklanan gelecekteki uygulamalar için önemli bir vaat olduğunu vurgulamaktadır.

6.6.1. Literatür Taraması Özeti

Bu çalışma, makine öğrenimi alanında yeni ve gelişmekte olan bir zorluk olan SCED araştırmaktadır. Öncelikle tek kişilik konuşma veya fizyolojik sinyalleri analiz eden geleneksel bireysel duygu tanıma modellerinden farklı olarak SCED, ortamdaki işitsel ipuçlarından yararlanarak kalabalıklar veya topluluk ortamlarındaki kolektif duygu durumlarını sınıflandırmaya çalışmaktadır [214-224].

6.6.2. Literatürdeki Mevcut Çalışmalar

Duygu algılama alanındaki önceki araştırmaların hâkim gövdesi, ağırlıklı olarak bireysel düzeydeki modalitelere, özellikle de konuşma ve metne odaklanmıştır. İncelenen literatür genel olarak aşağıdaki gibi kategorize edilebilir:

- **Konuşma tabanlı duygu tanıma:** Çok sayıda araştırma, perde, tonlama ve konuşma hızı dâhil olmak üzere konuşma özelliklerinin analizi yoluyla duygusal durumların sınıflandırılmasına odaklanmıştır. Hâkim metodolojik yaklaşımlarda tipik olarak MFCC'ler özellik temsilleri olarak kullanmakta ve derin öğrenme tabanlı mimari çerçevelerle birleştirmektedir [225-226].
- **Metin tabanlı duygu tanıma:** Sosyal medya gönderilerinden ve karşılıklı konuşmalardan elde edilen duygusal içeriğin analizi, köklü bir araştırma alanı oluşturmaktadır.

Dönüştürücü tabanlı mimariler kullanan duygu analizi modelleri, metinsel duygu sınıflandırması için yaygın olarak kullanılmaktadır [227-231].

- **Topluluk düzeyinde duygu tespiti:** Topluluk çapında duygu tespiti için ortam sesinin analizine ilişkin araştırmalar nispeten sınırlı kalmaktadır. Bazı çalışmalar sosyal medya verilerinin incelenmesi yoluyla kalabalık duygu analizini araştırmış olsa da çevresel işitsel sinyallerin bu amaçla doğrudan kullanımı daha az yaygındır.

6.6.3. Literatürdeki Eksiklikler

Bu çalışmada, SCED ile ilgili mevcut araştırmalarda birkaç kritik boşluk tespit edilmiştir:

- **Çevresel ses tabanlı duygu tanıma modellerinin azlığı:** Mevcut duygu algılama çalışmalarının çoğu, öncelikle bireysel konuşma analizini vurgulamakta, topluluk ortamlarındaki ortam gürültüsü ve örtüşen konuşma sinyallerinin analizine nispeten sınırlı ilgi gösterilmektedir.
- **Derin öğrenme yaklaşımlarının hesaplama karmaşıklığı:** Mevcut derin öğrenme modelleri yüksek sınıflandırma performansı gösterirken, genellikle önemli hesaplama kaynağı gereksinimleri ile karakterize edilmektedirler.
- **Özel bir SCED veri setinin bulunmaması:** Ses tabanlı topluluk duygu tespiti için özel olarak düzenlenmiş, kamuya açık bir veri kümesi şeklinde kayda değer bir boşluk mevcuttur, bu da bu alandaki araştırma ilerlemesini engellemektedir.

6.6.4. Çalışmada Kullanılan Veri Seti

SCED veri setinin eksikliğini gidermek için yazarlar, kalabalık ve topluluk tabanlı sesler içeren YouTube videolarından yeni bir veri seti topladılar. Bu veri seti ile ilgili bilgiler Tablo 6.16'da özetlenmiştir.

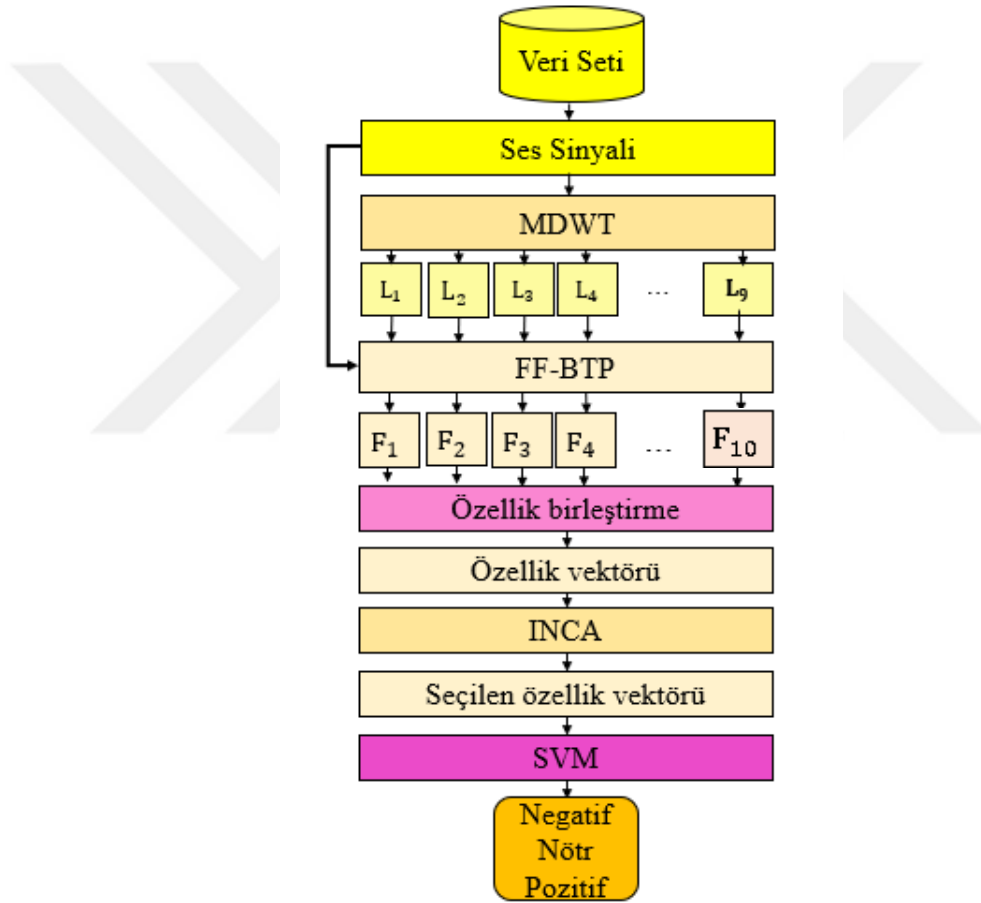
Tablo 6.16. Oluşturulan SCED veri setine ait bilgiler [125]

Özellik	Tanım
Veri Kümesi Adı	Ses Tabanlı Topluluk Duygu Tespiti (SCED) Veri Seti
Kaynak	Kalabalık ve çevresel seslerin yer aldığı YouTube videoları
Ses Dosyalarının Sayısı	2.733 (.wav formatı)
Segment Uzunluğu	Dosya başına 3 saniye
Örnekleme Oranı	44,1 kHz
Duygu Sınıfları	Negatif (919 dosya), Nötr (905 dosya), Pozitif (909 dosya)
Veri Filtreleme	Ön plandaki sesler, şarkılar ve açık konuşmalar kaldırıldı; ortam, üst üste binen ve dolaylı sesler eklendi
Doğrulama	Klip başına 3 kişilik doğrulama ile manuel etiketleme

Tabloda özellikleri açıklanan veri setine şu bağlantıdan <https://www.kaggle.com/datasets/arifmetehanyldz/sced-v1> ulaşılabilir [232].

6.6.5. Çalışmada Önerilen Yöntem

Bu çalışma, derin öğrenme kavramlarından ilham alarak SCED için hesaplama açısından verimli bir yaklaşım olan İleri-İleri İkili Üçlü Örüntü (FF-BTP) özellik mühendisliği modelini tanıtmaktadır. Önerilen modelin genel akış şeması Şekil 6.17’de sunulmuştur.



Şekil 6.17. FF-BTP tabanlı öznelik mühendisliği modeli önerildi. * L: Alçak geçiren filtre bantları, FF-BTP: önerilen ileri-ileri ikili üçlü model, f: özellik vektörü, INCA: yinelemeli komşu bileşen analizi özellik seçicisi, SVM: destek vektör makinesi sınıflandırıcısı [125].

Önerilen metodoloji üç ana aşamadan oluşmaktadır:

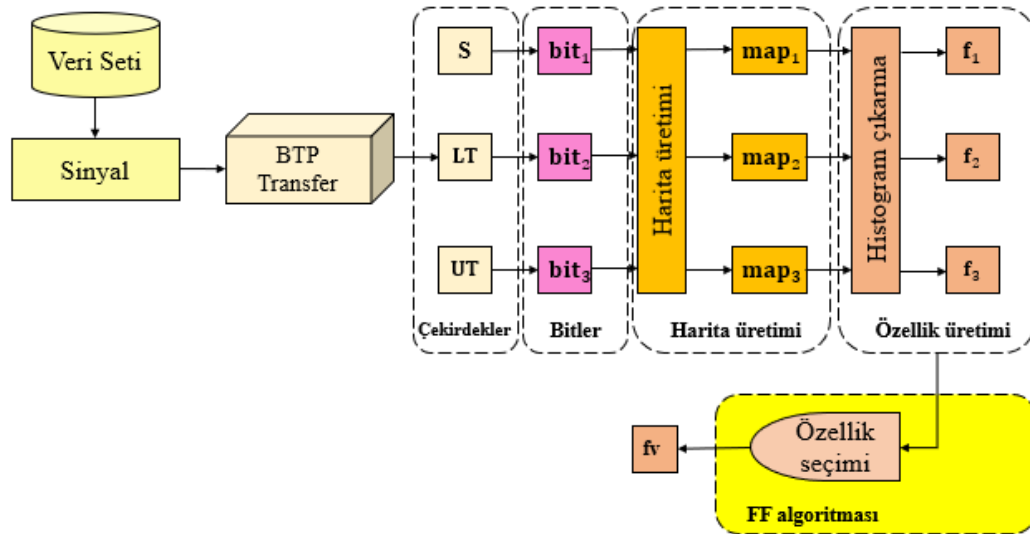
Özellik Çıkarımı

- **İleri-ileri ikili üçlü örüntü (BTP) özellik çıkarımı:** İkili Üçlü Örüntü (BTP) algoritması, işaretlenmiş, üst üçlü ve alt üçlü çekirdekleri kullanarak ses sinyallerinden dokusal

özellikler çıkarmak için kullanılmıştır. Çıkarılan her bir özellik vektörü 256 özellikten oluşmaktadır. Daha sonra, Hinton'un İleri-İleri algoritmasından esinlenen bir seçim mekanizması, on sinyalin her biri için en uygun 256 boyutlu vektörü otonom olarak seçer. Bu seçim, birincil duygu sınıfına ait sinyallerin özellik temsilleriyle karşılaştırıldığında MSE'nin en üst düzeye çıkarılmasına dayanır ve örüntü tanıma için insan benzeri bir ayırt edici süreci etkili bir şekilde taklit etmektedir.

- **MDWT:** Çok Seviyeli Ayrık Dalgacık Dönüşümü, frekans ve uzamsal alan özelliklerini yakalamak için dokuz ayrıştırma seviyesinde Symlet 4 ana dalgacığını kullanarak uygulanır. Bu işlem ses sinyali başına toplam 10 özellik vektörü vermiştir.
- **Son özellik vektörü oluşturma:** Tam özellik vektörü, çıkarılan tüm özelliklerin birleştirilmesiyle oluşturulur ve 2560 boyutlu bir gösterim elde edilmiştir ($=256 \text{ özellik} \times 10 \text{ sinyal}$).

Özellik çıkarma adımı kullanılan FF-BTP özellik mühendisliği işlevinin şeması Şekil 6.18'de sunulmuştur.

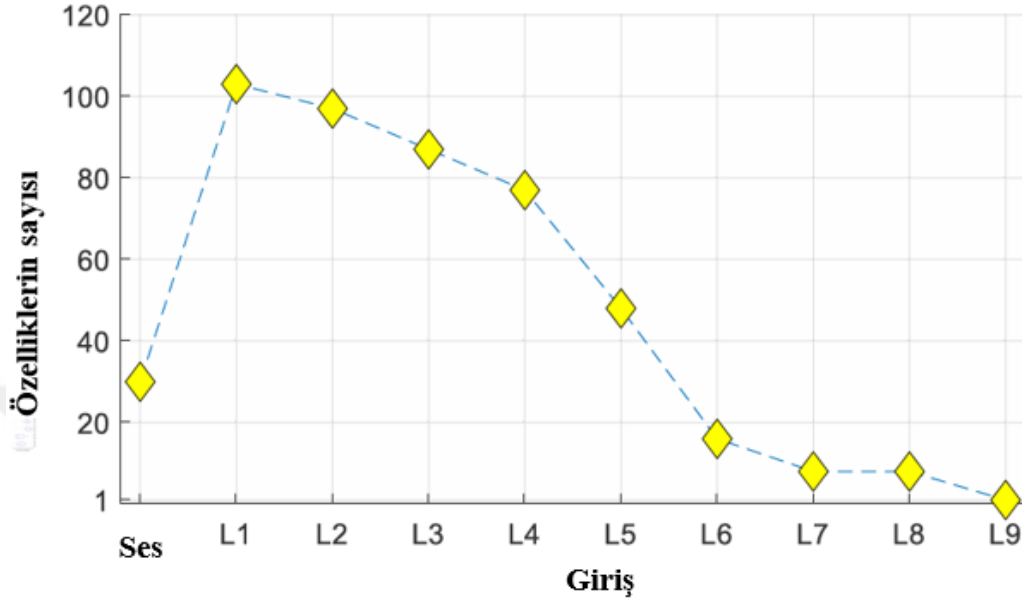


Şekil 6.18. İleriye Doğru İkili Üçlü Örüntü (FF-BTP) Özellik Mühendisliği Şeması. *BTP transf.: BTP dönüşümü, S: signum fonksiyonu, LT: alt üçlü, UT: üst üçlü, f: öznelik vektörü, fv: seçilen öznelik vektörü, FF: ileri-ileri [125].

Özellik Seçimi

Başlangıçtaki ham ses sinyalinden ve 9 adet MDWT ile ayrıştırılmış dalgacık bandından üretilen 2560 boyutlu özellik uzayından en ayırt edici 475 özelliği seçmek için Yinelemeli Komşuluk Bileşen Analizi (INCA) kullanılmıştır. Bu adım, sınıflandırma doğruluğunu korurken hesaplama karmaşıklığını azaltmayı amaçlamıştır. Şekil 6.19'dan da anlaşılacağı üzere 10 olası

giriş sinyalinin tümü arasında, L1 ve L9 dalgacık bantları, seçilen özelliklerin en yüksek (475'in 103'ü) ve en düşük (475'in 1'i) oranlarına katkıda bulunmuştur.



Şekil 6.19. Model girişine göre katmanlandırılmış INCA tarafından seçilen özelliklerin dağılımı (ham sinyal veya L1 ile L9 arasındaki dalgacık bantları) [125].

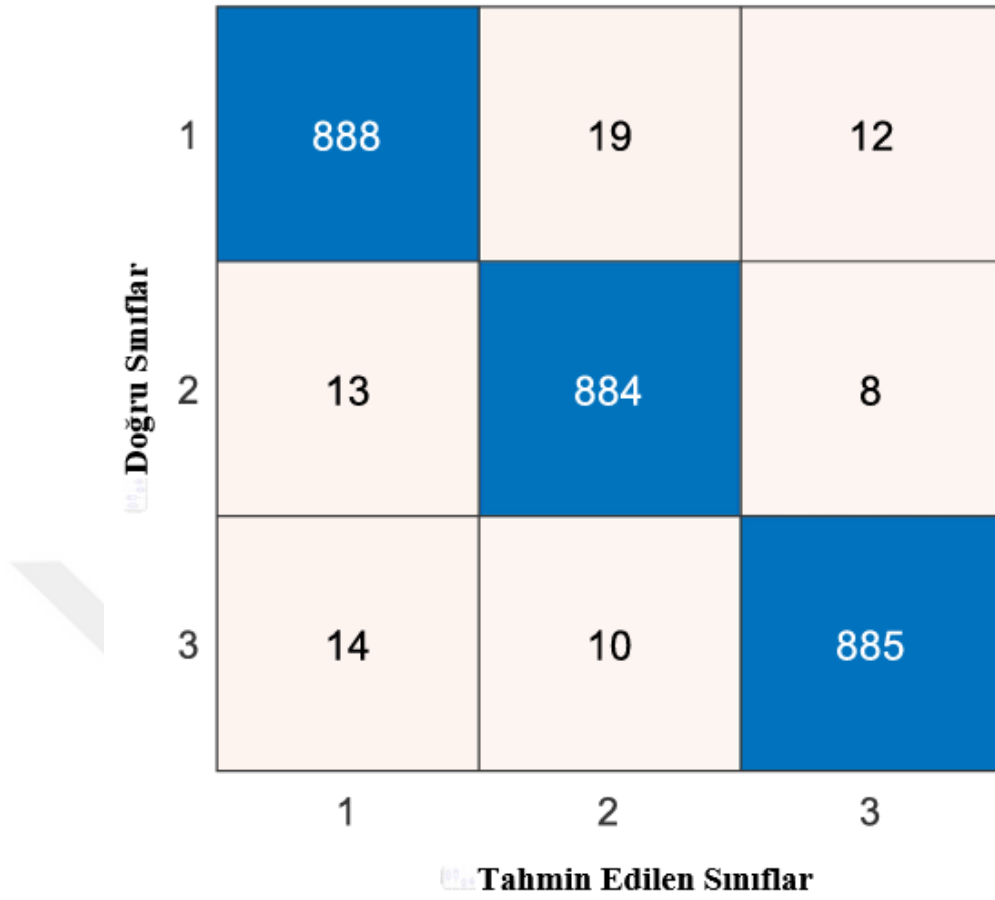
Bu seçim süreci, yinelenmeli denemeler ve bir SVM sınıflandırıcısından gelen geri bildirimler kullanılarak yanlış sınıflandırma oranını en aza indirme ilkesiyle yönlendirilmiştir. Özellik seçiminden önce tutarlı ölçeklendirme sağlamak için özellik vektörlerine min-max normalizasyonu uygulanmıştır.

Sınıflandırma

Sınıflandırma görevi için bir SVM sınıflandırıcısı kullanılmıştır. Model performansı 10 kat çapraz doğrulama paradigması kullanılarak titizlikle değerlendirilmiştir. SVM hiperparametrelerinin ince ayarı için Bayes optimizasyonu uygulanmıştır.

6.6.6. Çalışmanın Performans Değerlendirmesi

Sonuçlar, modelin topluluk duygu tespiti alanında oldukça yüksek bir doğruluğa ulaştığını göstermektedir. Önerilen modelin sonuçlarına ait karışıklık matrisi Şekil 6.20'de sunulmuştur.



Şekil 6.20. FF-BTP tabanlı öznelik mühendisliği modelinin karışıklık matrisi. *1: Olumsuz; 2: Nötr; 3: Pozitif [125].

Topluluk duygularının üç farklı sınıfa (Negatif, Nötr, Pozitif) ayrılmasını içeren sınıflandırma görevi, Tablo 6.17’deki sınıflandırma bazında standart performans ölçütleri kullanılarak değerlendirilmiştir.

Tablo 6.17. FF-BTP tabanlı modelin sınıf bazında ve genel sınıflandırma performansı [125].

Metrik	Negatif (%)	Nötr (%)	Pozitif (%)	Genel Toplam (%)
Doğruluk	-	-	-	97.22
Kesinlik	97.05	96.82	97.79	97.22
Duyarlılık	96.63	97.68	97.36	97.22
F1-Skor	96.84	97.25	97.57	97.22

Parametrik FF-BTP özellik mühendisliği modelinin bireysel bileşenlerinin katkılarını incelemek için, Tablo 6.18’de gösterildiği gibi ablasyon çalışmaları gerçekleştirilmiştir.

Tablo 6.18. FF-BTP tabanlı modelin ablasyon çalışmalarında sınıflandırma doğrulukları [125].

Durum	Tanım	Doğruluk (%)
1	Yalnızca alt üçlü çekirdekli özellikler kullanılarak; diğer model bileşenleri sabit tutuldu.	94.10
2	Yalnızca signum çekirdeği + LBP kullanılarak; diğer model bileşenleri sabit tutuldu.	96.05
3	Yalnızca üst üçlü çekirdek özellikleri kullanılarak; diğer model bileşenleri sabit tutuldu	95.80
4	Tam FF-BTP Modeli (tüm bileşenler)	97.22

Durum 4'deki sonuçlardan da görüldüğü üzere en iyi sınıflandırma performansına ulaşıldı ve %96,05 sınıflandırma doğruluğuyla bir sonraki en iyi vaka olan Vaka 2, %1,17 oranında geride bırakıldı. Performans değerlendirmesi aşamasındaki ablasyon çalışmalarını kısaca özetleyecek olursak:

- Tam İleri-İleri İkili Üçlü Örüntü (FF-BTP) modeli, %97,22'lik bir değere ulaşarak en yüksek sınıflandırma doğruluğunu göstermiştir.
- Signum çekirdeğinin kullanımı %96,05 ile ikinci en iyi sınıflandırma doğruluğunu sağlamıştır.
- Ablasyon çalışmaları, üç çekirdek türünün (signum, üst üçlü ve alt üçlü) entegrasyonunun sınıflandırma performansında istatistiksel olarak anlamlı bir artışa yol açtığını doğrulamıştır.

Bu bulgular, ses tabanlı topluluk duygu tespiti için modelin ayırt edici kapasitesini en üst düzeye çıkarmak üzere çeşitli özellik çıkarma bileşenlerini entegre etmenin önemini vurgulamaktadır.

6.6.7. Çalışmanın Katkıları

Bu çalışma, Ses Tabanlı Toplumsal Duygu Tespiti (SCED) alanına birkaç önemli katkı sunmaktadır:

- **Yeni Bir Veri Kümesinin Tanıtımı:** Ses tabanlı topluluktan duygu tanıma için özel olarak düzenlenmiş, kamuya açık ilk veri setinin oluşturulması ve kamuya sunulması. Veri kümesi üç farklı duygusal kategoride dengeli bir dağılım sergilemektedir. Sonraki araştırma çabalarını kolaylaştırmak ve teşvik etmek için Kaggle platformu aracılığıyla kamuya açık erişim sağlanmıştır.
- **İleri-İleri İkili Üçlü Örüntü (FF-BTP) Özellik Çıkarma Modelinin Geliştirilmesi:** FF-BTP olarak adlandırılan ve hesaplama ek yükünü azaltırken derin öğrenme mimarilerinin performansını taklit etmek için tasarlanmış yeni bir el yapımı özellik çıkarma modelinin tanıtımı yapılmıştır. FF-BTP modeli şunları entegre etmiştir:

- Dokusal özelliklerin çıkarılması için İkili Üçlü Örüntü (BTP) algoritması kullanılmıştır [233].
- Frekans alanı özelliklerinin çıkarılması için MDWT uygulanmıştır [234].
- Özellik seçimi için INCA kullanılmıştır [144].

Bu çalışma, Hinton'ın İleri-İleri (FF) algoritmasından esinlenen ilk el yapımı özellik çıkarma işlevini temsil etmektedir [235].

Azaltılmış Hesaplama Gereksinimleri ile Yüksek Sınıflandırma Doğruluğunun Elde Edilmesi

- Önerilen FF-BTP modeliyle, mevcut derin öğrenme modellerine kıyasla üstün performans göstererek %97,22'lik bir sınıflandırma doğruluğu elde edilmiştir.
- Modelin hesaplama verimliliği, GPU'ların kullanımını gerektirmediğinden, onu gerçek zamanlı uygulamalarda dağıtım için çok uygun hale getirmiştir.

6.6.8. Çalışmanın Sonuçları ve Tartışma

Önerilen İleri-İleri İkili Üçlü Örüntü (FF-BTP) modeli, SCED için özel olarak geliştirilen ilk el yapımı özellik mühendisliği metodolojisini temsil etmektedir. Çekirdeğe özgü performansın analizi, alt üçlü çekirdeğin nötr duygusal durumların sınıflandırılması için en sık seçimi sergilediğini ortaya koymuştur. Dahası, L1 ve L9 dalgacık alt bantları sırasıyla en yüksek ve en düşük özellik katkısını göstermiştir. Ablasyon çalışmaları, tüm özellik çıkarma bileşenlerinin sinerjik entegrasyonunun optimum sınıflandırma doğruluğunu sağladığını doğrulamıştır. Gelecekteki araştırma çabaları birkaç temel alana yönelecektir.

SCED veri setinin daha geniş bir duygusal kategori yelpazesini kapsayacak şekilde genişletilmesi gerekecektir. FF-BTP modeli gibi özellik mühendisliği tekniklerini derin öğrenme mimarileriyle bütünleştiren hibrit modellerin keşfi gerçekleşmiştir. FF-BTP çerçevesinin fizyolojik ve biyometrik verilerin analizi için uyarlanabilirliğinin araştırılması gerekmektedir.

6.7. FF-Orbital Kaos Tabanlı Özellik Çıkarma Modeli Kullanılarak Toplumun Duygularının Ses Yoluyla Tespiti Üzerine Bir Araştırma

Topluluk Duygu Tespiti (CED), kamu güvenliği ve adli bilişim açısından büyük önem taşımakta ve halka açık toplantılar, protestolar ve acil durumlar gibi dinamik ortamlardaki kolektif duygusal durumların analizini kolaylaştırmaktadır. Bu araştırma, CED için özel olarak tasarlanmış yeni, uygun maliyetli, otomatik ve kendi kendini organize eden bir sınıflandırma modelini tanıtmaktadır. Birincil amaç, CED araştırma alanında gelişmiş sinyal işleme tekniklerinin etkinliğini göstermektir.

Bu çalışma için bilinen en büyük CED ses veri seti oluşturulmuştur. Üç saniyelik üst üste binen 5.051 ses örneğinden oluşan veri seti, topluluk düzeyinde negatif, nötr ve pozitif duygusal

durumlar olarak kategorize edilmiştir. Bu kapsamlı veri kümesi, modelin çeşitli akustik ve duygusal senaryolar arasında genellenebilirliğini artırmakta ve önerilen otomatik CED modelinin sınıflandırma yetenekleri için sağlam bir test ortamı olarak hizmet vermektedir [236].

Önemli bir yenilik, İleri Yönlü Örüntü Tabanlı Özellik Oluşturucu (FF-Orbital) olarak adlandırılan kendi kendini organize eden bir özellik çıkarma işlevinin tanıtılmasıdır. Bu işlev, önceden tanımlanmış altı şablondan oluşan bir setten en uygun kalıbı otonom olarak seçer ve böylece manuel özellik mühendisliği gerekliliğini ortadan kaldırır. Ayrıca hem uzamsal hem de frekans alanı özelliklerini yakalayan özellikleri çıkarmak için frekans bantları üreten UTMDWT entegrasyonu yoluyla çok seviyeli bir özellik çıkarma yaklaşımı sağlanmıştır. Özellik seçimi, kendi kendini organize eden özellik seçimi yetenekleriyle tanınan ve en bilgilendirici özellik alt kümesini tanımlayan bir yöntem olan INCA kullanılarak gerçekleştirilmiştir. Nihai sınıflandırma Bayes optimizasyonlu bir SVM kullanılarak gerçekleştirilmiştir. Modelin performansı 10 kat çapraz doğrulama kullanılarak titizlikle değerlendirilmiş ve %98,81'lik kayda değer bir sınıflandırma doğruluğu elde edilmiştir. Bu deneysel sonuçlar, önerilen CED modelinin adli bilişim, kamu güvenliği izleme ve büyük ölçekli topluluk düzeyinde duygu analizi gibi pratik uygulamalarda kullanılabilirliğini ve etkinliğini vurgulamaktadır.

Bu çalışma, hesaplama açısından yoğun derin öğrenme modellerine özellik mühendisliği tabanlı bir alternatif sunarak CED araştırma ortamına önemli bir katkı sağlamaktadır. Elde edilen yüksek sınıflandırma doğruluğu, önerilen sinyal işleme yaklaşımının hem CED hem de ses adli bilişimi uygulamaları için önemli bir değer olduğunu göstermektedir.

6.7.1. Literatür Taraması Özeti

Bu makale, ses sinyallerini kullanarak topluluk duygu tespiti (CED) alanını araştırmaktadır. Kamu güvenliği, ruh sağlığı izleme ve insan-bilgisayar etkileşimi dâhil olmak üzere bir dizi uygulama genelinde duygu tanımanın kritik önemini vurgulamaktadır [237-238]. Geleneksel metodolojiler duygu analizi için ağırlıklı olarak metinsel ve konuşma tabanlı verilere odaklanmış olsa da bu çalışma daha büyük kolektif varlıklar içindeki duygusal durumları ayırt etmenin bir yolu olarak ses tabanlı analizin önemini vurgulamaktadır [239-244].

6.7.2. Literatürdeki Mevcut Çalışmalar

Önceki araştırmalar, konuşma ve metinsel verilerin analizi yoluyla bireysel duygu tanımayı kapsamlı bir şekilde araştırmıştır. Bu çalışmalar, sınıflandırma doğruluğunu artırma amacıyla CNN'ler, LSTM'ler ve dikkat mekanizmalarını kapsayan çeşitli makine öğrenimi modellerini birleştirmiştir. Özellikle, birkaç önemli çalışma, ses verilerini yüz tanıma ve fizyolojik sinyallerle birleştiren çok modlu yaklaşımları entegre ederek yüksek performans elde etmiştir. Ancak, bu

yaklaşımların büyük ölçekli topluluk düzeyinde duygu tespitine doğrudan uygulanabilirliği belirli sınırlamalar sunmaktadır [125], [245-250].

6.7.3. Literatürdeki Eksiklikler

Bu makale, mevcut araştırma gövdesindeki birkaç kritik boşluğu vurgulamaktadır:

- **Topluluk düzeyinde akustik duygu analizi için tasarlanmış veri kümelerinin azlığı:** Önceki çalışmalarda kullanılan veri kümelerinin çoğunluğu, özellikle çevresel ve kalabalık sesi verilerini kapsayan veri kümelerinin dikkate değer bir eksikliğiyle, öncelikle bireysel konuşma tabanlı duygu tanıma üzerine yoğunlaşmıştır.
- **Derin öğrenme yaklaşımlarının hesaplama kısıtlamaları:** Derin öğrenme modelleri yüksek etkinlik göstermiş olsa da önemli hesaplama kaynağı talepleri onları gerçek zamanlı uygulama senaryoları için genellikle pratik olmaktan çıkarmaktadır.
- **Etkili el yapımı özellik çıkarma tekniklerinin sınırlı keşfi:** Hesaplama verimliliğini korurken rekabetçi sınıflandırma doğruluğu elde edebilen el yapımı özellik çıkarma metodolojilerinin geliştirilmesi ve doğrulanmasında fark edilebilir bir araştırma eksikliği mevcuttur.
- **SCED için sistematik bir çerçevenin yokluğu:** Çevresel ses tabanlı topluluk duygu tespiti için özel olarak tasarlanmış kapsamlı ve sistematik bir yaklaşımla ilgili literatürde önemli bir boşluk bulunmaktadır.

6.7.4. Çalışmada Kullanılan Veri Seti

SCED için özel olarak yeni bir veri kümesi oluşturulmuştur. Bu veri kümesi YouTube platformundaki 189 halka açık videodan elde edilmiştir. Elde edilen veri kümesi, her biri 3 saniye süreli ve 44,1 kHz standart örnekleme hızına sahip toplam 5.051 ayrı ses örneğinden oluşmaktadır [236]. Bu veri kümesi için küratörlük süreci, ön plandaki konuşma ve müzik içeriğinin hariç tutulmasına öncelik verildi. Bu kasıtlı yaklaşım, yakalanan duygusal ipuçlarının bireysel sözlü ifadelerden ziyade ağırlıklı olarak ortam seslerinden ve örtüşen kalabalık seslendirmelerinden türetilmesini sağlamak için üstlenildi. Bu metodolojinin amacı, topluluk çapında duygu tespitinin araştırılması için daha genelleştirilmiş ve pratik olarak uygulanabilir bir veri kümesi oluşturmaktır. Veri kümesi üç farklı duygu sınıfına ayrılmıştır: Negatif, Nötr ve Pozitif. Ses örneklerinin içeriği, protestolar ve halka açık etkinlikler gibi gerçek hayattaki topluluk durumlarının karakteristik özelliği olan örtüşen konuşma ve ortam gürültüsüne odaklanmaktadır. Veri kümesindeki baskın dil Türkçe olmakla birlikte, örneklerin bir alt kümesi dilsel çeşitliliği tanıtmak için İngilizce ve İspanyolca içerik içermektedir. Duygu etiketlerinin güvenilirliğini ve geçerliliğini sağlamak için, bağımsız uzmanlar tarafından manuel ek açıklama ve ardından beş bağımsız uzmandan oluşan bir

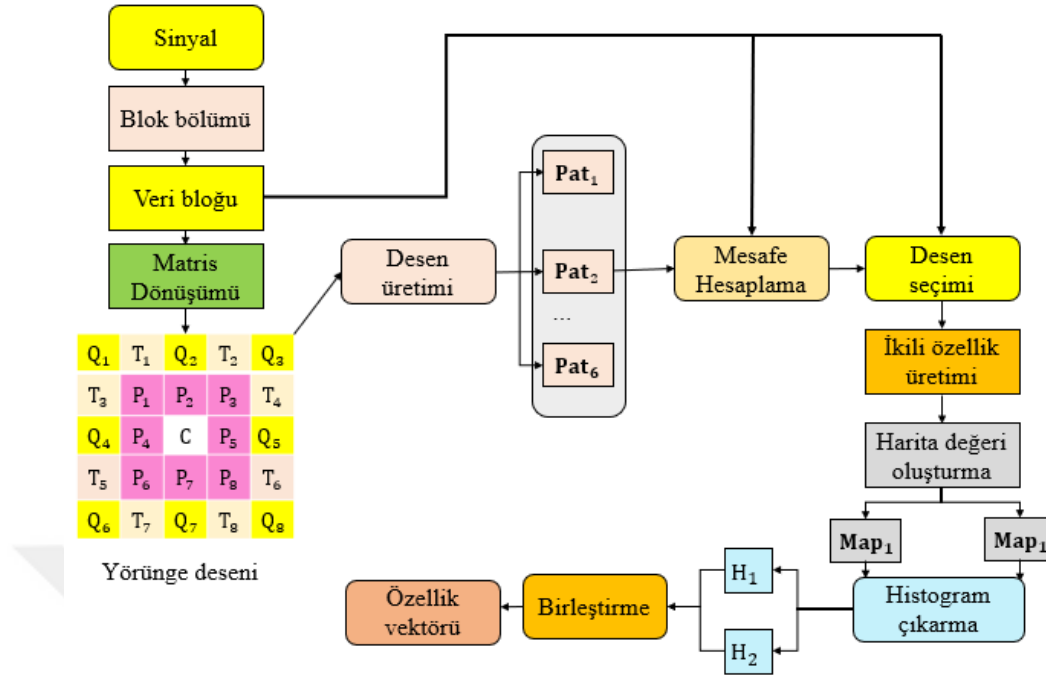
panel tarafından doğrulamayı içeren titiz bir açıklama ve doğrulama süreci uygulanmıştır. Etiketleme sürecinin tutarlılığını ölçmek için değerlendiriciler arası uyum da analiz edilmiştir. Veri seti ile ilgili bilgiler Tablo 6.19’da sunulmuştur.

Tablo 6.19. Oluşturulan CED veri setine ait bilgiler

Özellik	Tanım
Kaynak	189 adet herkese açık YouTube videosu kullanılmıştır.
Örnek Sayımı	5.051 ses örneği vardır
Örnek Süresi	Örnek başına 3 saniye
Örnekleme Oranı	44,1 kHz
Sınıflar	Negatif, Nötr, Pozitif
İçerik Odaklı	Gerçek yaşam topluluk durumlarından (örn. protestolar, etkinlikler vb.) kaynaklanan örtüşen konuşma ve ortam gürültüsü içermektedir.
Dil Çeşitliliği	Çoğunlukla Türkçe, bir miktar İngilizce ve İspanyolca içerik
Doğrulama	Bağımsız uzmanlar tarafından manuel açıklama, beş bağımsız uzmandan oluşan bir panel tarafından doğrulandı; değerlendiriciler arası uyum analizi gerçekleştirildi.

6.7.5. Çalışmada Önerilen Yöntem

Bu çalışma, CED için tasarlanmış İleri-İleri Yörünge Kaos Tabanlı (FF-Orbital) olarak adlandırılan yeni bir özellik çıkarma modeli sunmaktadır. Önerilen modelin akış diyagramı Şekil 6.21’de gösterilmektedir.



Şekil 6.21. Önerilen FF-Orbital'in grafiksel özellikleri. Burada, Pat: Desen, H: Histogram [236]

Önerilen metodoloji üç ayrı aşamaya yapılandırılmıştır:

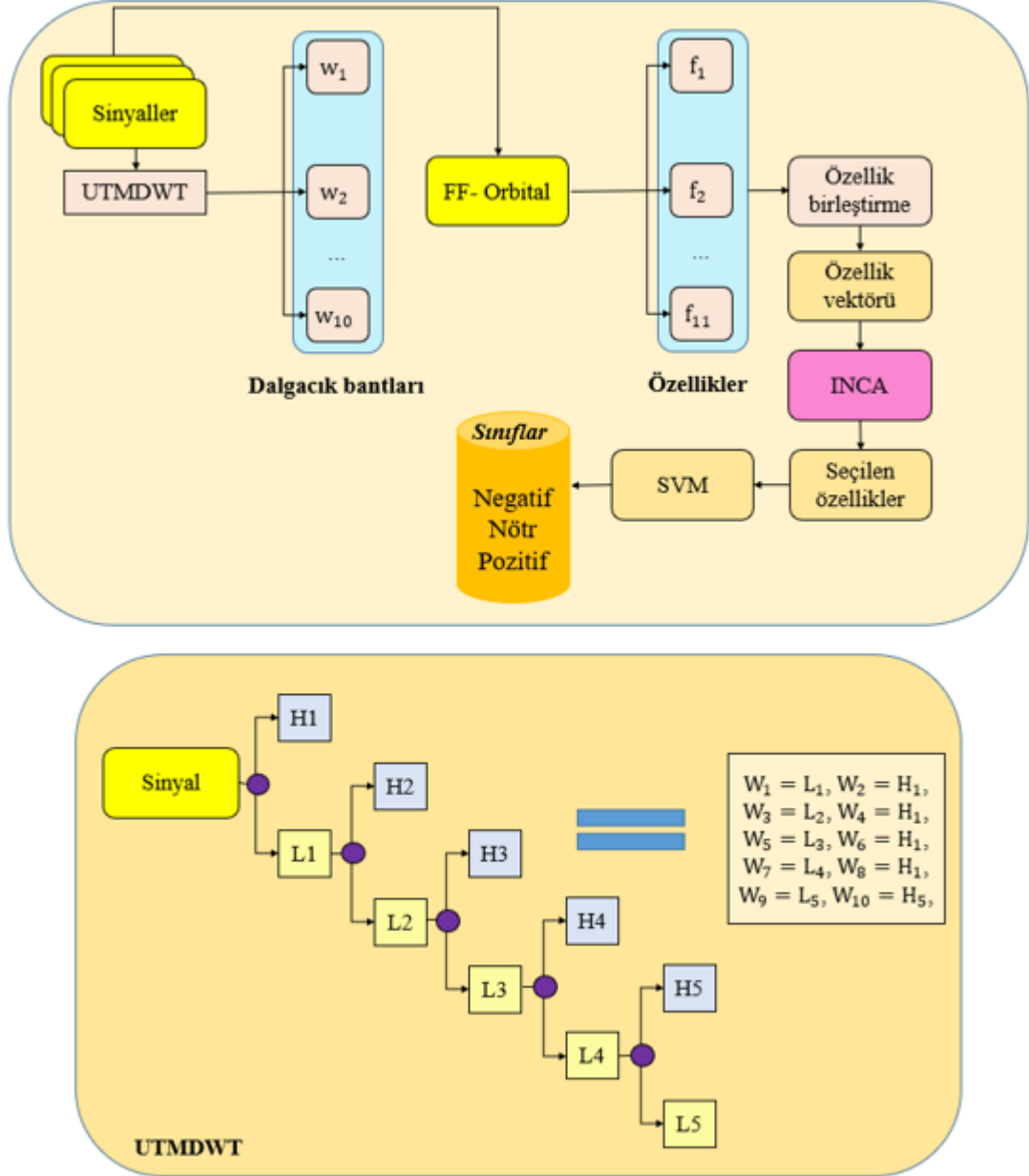
Özellik Çıkarma

- **Dengesiz ağaçlı çok seviyeli ayrık dalgacık dönüşümü (UTMDWT):** Ses sinyallerinden Symlet 6 filtresi kullanılarak her sinyalden 10 dalgacık bandı çıkarmaktadır.
- **FF-Orbital özellik mühendisliği:** Hinton'un ileri-ileri algoritmasından esinlenen bu yeni bileşen, ses verilerindeki belirgin desenleri otonom olarak seçer ve böylece hesaplama karmaşıklığını azaltmayı amaçlamıştır [235], [251]. Kendi kendini organize eden bu özellik çıkarıcı Tablo 6.20'deki adımlardan oluşmaktadır.

Tablo 6.20. FF-Orbital özellik mühendisliği modelinden özellik çıkarma adımları

Adımlar	Durumlar
1	Sinyaller örtüşen 25 noktalı bloklara bölünmüştür.
2	Her blok 5x5 matrisler halinde yeniden şekillendirilmiştir.
3	6 farklı uzamsal desen (örn. P-C, Q-C, vb.) üretilmiştir.
4	Genel sinyalden ortalama uzaklığa göre en iyi örüntü seçilmiştir.
5	Signum ve absolute signum fonksiyonlarını kullanarak ikili özelliklerin çıkarılması gerekmektedir.
6	İkili eşlemelerden histogramlar oluşturulmuştur.
7	Sinyal başına 512 uzunlukta bir özellik vektörü üretilmiştir.

Özellik çıkarma aşamasında gerçekleştirilen adımların şematik gösterimi Şekil 6.22’de verilmiştir.



Şekil 6.22. FF-Orbital tabanlı öznelik mühendisliği modeline ve UTMDWT dönüşümüne genel bakış. Burada: w: dalgacık bandı, f: özellik vektörü, L: düşük geçiren filtre dalgacık bandı, H: yüksek geçiren filtre dalgacık bandı [236].

FF-Orbital ile ham sinyal ve 10 dalgacık bandı olmak üzere buradaki 11 ses örneği başına 512 özellik çıkarılarak toplamda 5632 (=512 özellik x 11 sinyal) boyutlu bir özellik vektörü oluşturulmuştur.

Özellik Seçim

En ayırt edici özellikleri seçmek için INCA kullanılmış ve yalnızca en alakalı alt küme korunmuştur. Ortaya çıkarılan 5632 özellikten en anlamlı 320 özellik seçilmiştir.

Sınıflandırma

Bayes optimizasyonu ile optimize edilmiş bir SVM sınıflandırıcısı, ses örneklerinin negatif, nötr ve pozitif olarak üç duygu kategorisine nihai sınıflandırılması için kullanılmıştır. Aşırı uyumu önlemek için 10 kat çapraz doğrulama tekniği kullanılmıştır.

6.7.6. Çalışmanın Performans Değerlendirmesi

Önerilen FF-Orbital modelin etkinliği, 10 katlı çapraz doğrulama prosedürüyle titizlikle doğrulanmıştır. Sınıflandırma performansı Tablo 6.21’de sunulmuştur.

Tablo 6.21. Önerilen FF-Orbital kaos tabanlı modelin performans sonuçları [236]

Metrik	Doğruluk (%)
Doğruluk	98.81
Duyarlılık	98.80
Kesinlik	98.83
F1-Skoru	98.82

Önerilen modelin genel sınıflandırma performansı, ses tabanlı topluluk duygu tespiti görevinde yüksek derecede doğruluk, hassasiyet, geri çağırma ve F1-skoru göstermektedir. FF-Orbital tabanlı model, **%98,81**'lik bir sınıflandırma doğruluğu göstermiştir ve böylece mevcut derin öğrenme modellerine kıyasla üstün performans sergilerken aynı zamanda daha düşük bir hesaplama maliyetini korumuştur.

6.7.7. Çalışmanın Katkıları

Bu çalışma, Ses Tabanlı Topluluk Duygu Tespiti alanına birkaç önemli katkı sunmaktadır:

- **Yeni bir özellik çıkarma modeli geliştirme:** SCED için özel olarak tasarlanmış, FF-Orbital olarak adlandırılan yeni bir el yapımı özellik çıkarma modelinin tanıtımı gerçekleştirilmiştir.
- **En büyük özel CED veri setinin oluşturulması:** Şu anda mevcut olan en kapsamlı CED veri setinin oluşturulması ve kamuoyuna açıklanması, özellikle çevresel ses tabanlı duygu tanıma görevi için özel olarak düzenlenmiştir.

- **En son teknoloji performansının elde edilmesi:** Önerilen model, mevcut metodolojilere kıyasla önemli ölçüde azaltılmış hesaplama gereksinimlerini gösterirken, %98,81'lik en son teknoloji sınıflandırma doğruluğuna ulaşmaktadır.
- **Gerçek zamanlı duygu izleme için pratik bir çözüm sunulması:** Geliştirilen model, kamusal alanlarda gerçek zamanlı duygu izleme için uygulanabilir ve etkili bir çözüm sunmuştur.

Bu araştırma, Çin'in Zhejiang Üniversitesi'ndeki Endüstriyel Kontrol Teknolojisi Devlet Ana Laboratuvarı'nın Açık Araştırma Projesi tarafından desteklenmiştir, Hibe No: ICT2024B42.

6.7.8. Çalışmanın Sonuçları ve Tartışma

Bu araştırmanın sonuçları FF-Orbital özellik çıkarma tekniği, çevresel akustik sinyallere gömülü belirgin duygusal ipuçlarını yakalamada yüksek derecede etkili olmuştur. Önerilen sistem, mevcut derin öğrenme metodolojileriyle karşılaştırıldığında daha yüksek sınıflandırma doğruluğu ve gelişmiş hesaplama verimliliği ile kanıtlandığı üzere üstün performans elde etmiştir. Bu çalışma, ses tabanlı duygu tanıma ve adli ses analizi alanlarında sonraki araştırma çabaları için sağlam ve önemli bir temel oluşturmaktadır.

7. BULGULAR VE TARTIŞMA

Bu tezde sunulan yedi çalışma, çevresel ses sınıflandırması, şiddet tespiti, konuşmacı sayma ve topluluk duygu tanıma konularına odaklanarak ses tabanlı makine öğrenimi alanını toplu olarak geliştirmektedir. Her bir çalışma yeni metodolojiler, veri kümeleri ve hibrit yaklaşımlar sunarak doğruluk, hesaplama verimliliği ve genellenebilirlik konularında önemli gelişmeler ortaya koymaktadır. Bu yönleriyle “Yeni Nesil Makine Öğrenmesi Yöntemleri Tabanlı Ortam Seslerinden Duygu ve Şiddet Tespiti” isimli tezimize katkı sağlamaktadır. Aşağıda temel bulguların ve bunların çıkarımlarının bir sentezi yer almaktadır.

7.1. El Yapımı ve Kaostan Esinlenen Modellere Genel Bakış

Bu çalışmada yedi farklı makine öğrenimi modeli geliştirilmiş ve incelenmiştir. Çoğunlukla el yapımı olan ve kaotik sistemlerden esinlenen bu modeller şunlardır:

- **LGNPat:** ESC için çok seviyeli dalgacık ayrıştırması ve yerel istatistiksel özellikler kullanan biyolojik esinli bir doku tanımlayıcısı olarak tanımlanmaktadır.
- **BP19:** Hesaplama açısından verimli ses şiddeti tespiti için optimize edilmiş ikili örüntü tabanlı 18 dalgacık ve 1 ham ses sinyalinde oluşan bir modelidir.
- **JWPAT:** James Webb Uzay Teleskobu'nun altıgen ayna yapısından ilham alan bir konuşmacı sayma modelidir.
- **SKLBP14:** Dokusal çevresel ses sınıflandırması için özel olarak uyarlanmış bir kare çekirdekli Yerel İkili Örüntü modelidir.
- **TreePAT23:** Şiddet tespiti için tasarlanmış ağaç yapılı bir kaotik özellik çıkarma mekanizması kullanan bir modeldir.
- **FF-BTP:** Topluluk duygularının tespiti için ileri beslemeli bir özellik seçimi metodolojisi ile güçlendirilmiş üç çekirdekli bir örüntü modelidir.
- **FF-Orbital:** Sağlam ses duygusu tanıma için çok bantlı dalgacık ayrıştırma ve özellik füzyonundan yararlanan kaostan ilham alan bir yörünge modelidir.

Bu modeller toplu olarak, özellikle çok katmanlı dönüşümler DWT veya TQWT gibi ve INCA veya ileri beslemeli optimizasyon gibi akıllı seçim stratejileri ile geliştirildiğinde, el yapımı özellik mühendisliğinin etkinliğinin altını çizmektedir.

7.2. Temel Performans Çıktıları

Geliştirilen modeller, değerlendirilen ses sınıflandırma görevlerinde rekabetçi ve birkaç örnekte üstün performans göstermiştir. Kritik olarak, bu GPU'lar gibi özel donanımlar

gerektirmeden veya derin öğrenme mimarilerine dayanmadan başarılmıştır. Her model tarafından elde edilen sınıflandırma doğruluklarının bir özeti Tablo 7.1’de sunulmuştur.

Tablo 7.1. Tüm modellerde elde edilen sınıflandırma doğrulukları

Örnek	Görev	Doğruluk (%)
LGNPat	Çevresel Ses Sınıflandırması	85.57
BP19	Şiddet Tespiti	97.01
JWPat	Konuşmacı Sayımı (1–32 sınıf)	86,74
SKLBP14	Dokusal ESC (15 sınıf)	94.00
TreePAT23	İkili Şiddet Tespiti	89,75
FF-BTP	Topluluk Duygu Sınıflandırması	97.22
FF-Orbital	Kaos Tabanlı Duygu Algılama	98.81

Doğruluğun ötesinde, modeller F1-skoru, duyarlılık ve özgüllük dâhil olmak üzere diğer temel performans göstergeleri için sürekli olarak yüksek değerler sergilemiştir. Çeşitli ölçütler arasında tutarlı bir şekilde güçlü olan bu performans, önerilen modellerin çeşitli ses sınıflandırma görevlerinde sağlamlığını ve etkinliğini daha da doğrulamaktadır.

7.3. Yeni Veri Seti Katkıları

Bu tez, beş yeni, kamuya açık veri setinin tanıtılması yoluyla araştırmanın yeniden üretilebilirliğini ve ölçeklenebilirliğini artırmaya önemli bir katkı sağlamaktadır. Bu kaynaklar aşağıda detaylandırılmıştır:

FIRAT-ESC50: 10.000 örnekten oluşan 50 sınıflı bir çevresel ses sınıflandırma veri kümesidir [155].

TESC: 1.500 örnek içeren 15 sınıflı doku tabanlı çevresel ses sınıflandırma veri kümesidir [190].

AVD Veri Kümesi: Özellikle ses şiddeti tespiti için 301 örnekten oluşan bir koleksiyondur [156].

Konuşmacı Sayma Veri Kümesi: 32 konuşmacı sınıfına dağıtılmış 3.121 örnekten oluşur [180].

SCED Veri Kümesi: Pozitif, nötr veya negatif duygularla etiketlenmiş 2733 örnekten oluşan veri kümesidir [232].

Topluluk Duygu Veri Kümesi (CED): Pozitif, nötr veya negatif duygularla etiketlenmiş 5.051 örnekten oluşmaktadır [236].

Her veri kümesi, ilgili alanlardaki hem çok sınıflı hem de ikili sınıflandırma problemlerinin güvenilir bir şekilde değerlendirilmesini kolaylaştırmak için titizlikle seçilmiş ve dengelenmiştir.

7.4. Modellerin Karşılaştırmalı Avantajları

Önerilen modellerin değerlendirilmesi, karşılaştırmalı avantajlarına ilişkin bazı temel bulguları ortaya çıkarmıştır:

Verimlilik: Tüm modeller MATLAB ortamında geliştirilmiş ve grafik işlem birimi (GPU) hızlandırması gerektirmeden standart merkezi işlem birimlerinde (CPU) etkin bir şekilde çalışmıştır.

Yorumlanabilirlik: Kaos tabanlı ve el yapımı tanımlayıcılara olan güven, adli bilişim, sağlık ve güvenlik gibi hassas alanlardaki uygulamalar için özellikle değerli olan bir şeffaflık derecesi sağlar.

Ölçeklenebilirlik: Çok sayıda sınıf içeren görevler için tasarlanan 32 sınıflı konuşmacı sayma modeli (JWPat) gibi modeller bile sınırlı hesaplama kaynakları kullanırken yüksek performans elde etmiştir.

Sağlamlık: BP19 ve FF-Orbital gibi modeller, gürültülü veya duygusal açıdan belirsiz ses segmentlerine karşı dayanıklılık göstermiş ve önemli bir bozulma olmadan performans seviyelerini korumuştur.

Özellikle, kaostan ilham alan özellik üreteçleri (TreePAT23, FF-BTP, FF-Orbital) LBP veya MFCC'lere dayalı olanlar gibi geleneksel yöntemlerden daha iyi performans gösterme konusunda benzersiz bir kapasite sergilemiştir. Bu üstün performans, gerçek dünyadaki ses sinyallerinin doğasında bulunan doğrusal olmayan ve düzensiz davranışları etkili bir şekilde modelleme yeteneklerine atfedilir, böylece ses verilerinin karmaşıklığına iyi uyum sağlamaktadır.

7.5. Pratik Çıkarımlar ve Kullanım Örnekleri

Bu tezde önerilen modeller önemli pratik sonuçlara sahiptir ve özellikle birkaç kilit alanda kullanılmak için çok uygundur:

Gömülü sistemler ve nesnelerin interneti (IoT): Mütevazı hesaplama gereksinimleri ve özel donanımdan bağımsız olmaları nedeniyle bu modeller, kaynak kısıtlı gömülü sistemler ve IoT cihazları için son derece uygundur.

Güvenlik ve gözetim: BP19 ve TreePAT23 gibi modeller özellikle güvenlik ve gözetim uygulamaları için uygundur ve şiddetli akustik olayların tespit edilmesini sağlar.

Akıllı evler ve sağlık hizmetleri: Konuşmacı sayımı için JWPat modeli ve duygu tespiti modelleri (FF-BTP ve FF-Orbital), konuşmacı takibi için akıllı ev ortamlarında ve ruh hali algılama için sağlık hizmetlerinde uygulanabilir.

Akademik ve endüstriyel araştırma: Kamuya açık olarak yayınlanan veri kümeleri ve geliştirilen model çerçeveleri, sonraki araştırmalar için değerli temeller sunar ve ses tabanlı makine öğrenimi alanında gelecekteki yenilikleri teşvik etmeye hizmet etmektedir.

7.6. Gelecekteki Araştırma Yönelimleri

Bu tezde sunulan çalışmaya dayanarak, gelecekteki araştırmalar için potansiyel yollar şunlardır:

Hibrit model entegrasyonu: Geliştirilen el yapımı özelliklerin geleneksel yöntemlerle derin öğrenme modellerini birleştiren hibrit mimarilere entegrasyonunun araştırılmasına katkıda bulunabilir.

Veri kümesi granülerliği: Duygular veya belirli şiddet türleri için daha ince taneli etiketleri dâhil etmek üzere mevcut veri kümelerinin kapsamını genişletmek veya yeni veri kümeleri geliştirmek için kullanılabilir [252].

Düşük etiketli öğrenme: Sınırlı etiketli verilerle karakterize edilen senaryoları ele almak için denetimsiz veya kendi kendine denetimli öğrenme tekniklerinin uygulanmasının araştırılmasına katkıda bulunabilir.

Gerçek zamanlı uç dağıtım: Bu modellerin mobil veya uç yapay zekâ platformlarında gerçek zamanlı dağıtımının uygulanması ve değerlendirilmesi ile ilgilidir.

7.7. Literatürdeki Benzer Çalışmalarla Karşılaştırılması

Bu tezde tanıtılan modellerin ve veri kümelerinin önemine ilişkin bağlam sağlamak için literatürdeki mevcut çalışmalarla karşılaştırmalı bir analiz yapılmıştır. Bu analiz, metodoloji, veri kümesi özellikleri, hesaplama maliyeti ve sınıflandırma performansı gibi temel hususlara odaklanmaktadır.

7.7.1. Metodolojik Yaklaşım

Özellikle duygu tanıma, çevresel ses sınıflandırması ve şiddet tespiti ile ilgili ses sınıflandırması alanındaki son araştırmaların büyük çoğunluğu, CNN'ler, RNN'ler ve bunların çeşitli mimari uzantıları gibi derin öğrenme modellerine yönelmiştir. Bu derin mimariler, özellikle kapsamlı veri kümeleri üzerinde eğitildiklerinde güçlü performans gösterebilirler de genellikle çeşitli sınırlamalarla ilişkilendirilirler:

- Sıklıkla GPU'lar gibi özel donanımlar ve uzun eğitim süreleri gerektiren yüksek hesaplama karmaşıklığı vardır.
- Öğrenilen özelliklerin sınırlı yorumlanabilirliği bazı dezavantajlar oluşturmaktadır.
- Küçük veya dengesiz veri kümelerine uygulandığında genelleştirilebilirliğin azalması olasıdır.

Buna karşılık, bu tezde sunulan modeller el yapımı, kaostan ilham alan ve yapısal olarak tasarlanmış tanımlayıcılar (örn. TreePAT23, JWPat, FF-Orbital) kullanmaktadır. Bu yaklaşımlar farklı avantajlar sunmaktadır:

- GPU'lara ihtiyaç duymadan verimli bir şekilde çalışırlar, bu da onları uç cihazlarda dağıtım için uygun hale getirmektedir.
- Yorumlanabilir özellik uzayları üreterek sınıflandırma sürecinin şeffaflığını artırır.
- Orta ölçekli ve dengeli veri kümeleri kullanıldığında bile güçlü genelleme yetenekleri sergilerler.
- Çeşitli deneysel ortamlarda derin öğrenme modelleriyle karşılaştırılabilir veya onları aşan performans gösterirler.

Örnek olarak, SKLBP14 modeli, k-En Yakın Komşular (kNN) sınıflandırıcısı kullanarak TESC veri kümesinde %94,00'lük bir doğruluk elde etmiştir. Bu performans, Salamon & Bello (2017) tarafından derin CNN modeli (%87,50) için ve Abdoli ve arkadaşları tarafından 1D-CNN (%91,20) için karşılaştırılabilir ESC görevlerinde gerçekleşen performansı aşmaktadır [187], [189].

7.7.2. Veri Seti Yeniliği ve Çeşitliliği

Ses sınıflandırmasında önceki çalışmalarda yaygın olarak kullanılan çok sayıda kıyaslama veri kümesi (örn. ESC-10, ESC-50, UrbanSound8K, RAVDESS) [21], [69], [101], [219], [253-256] aşağıdaki sınırlamalardan bir veya daha fazlasını sergilemektedir:

- Dar bir alan odağı sergilemektedir (örn. yalnızca kentsel sesler veya harekete geçirilmiş duygular).
- Sınırlı sınıf çeşitliliğine sahiptir (tipik olarak 10 ila 20 sınıf arasında değişir).
- Dengesiz örnek dağılımlarından muzdariptir.

Buna karşılık, bu tezde çeşitli avantajlar sunan beş özel veri seti tanıtılmaktadır:

- Daha geniş ve daha temsili kullanım durumlarını kapsamaktadırlar (örn. kalabalık duygu analizi, şiddet içeren olay tespiti vb.).
- FIRAT-ESC50'deki 50 sınıf gibi önemli ölçüde daha fazla sayıda sınıf içerirler.
- Titizlikle dengelenmişlerdir ve kamuya açıktırlar, böylece araştırmanın tekrarlanabilirliğini artırır ve daha sağlam modellerin geliştirilmesini teşvik ederler.

Örnek olarak, burada tanıtılan Konuşmacı Sayma veri kümesi 32 farklı konuşmacı sayma sınıfını kapsamaktadır. Bu durum, tipik olarak yalnızca 3-5 konuşmacı içeren senaryolara odaklanan önceki çalışmaların (örn. CountNet, BLSTM tabanlı yöntemler) çoğunluğuyla tezat oluşturmaktadır [119].

7.7.3. Hesaplama Verimliliği

Bu tezin temel ilkesi, hesaplama açısından verimli çözümlerin önceliklendirilmesidir. Tablo 7.2 önerilen modellerin literatürden seçilmiş çalışmalarla karşılaştırılmasını sunarak gereksinimlerini ve performanslarını vurgulamaktadır:

Tablo 7.2. Güçlü yönlerin karşılaştırmalı özeti

Çalışma / Model	GPU gerektirir mi?	Özellik Türü	Sınıflandırıcı	Keskinlik (%)	Notlar
BP19 (Bu Tez)[156]	Hayır	İkili örüntü + TQWT	SVM	97.01	Düşük hesaplama ile yüksek hassasiyet
CountNet [178]	Evet	CNN tabanlı Spektrogram	CNN	98,7	Sadece 5 konuşmacıya kadar değerlendirildi.
SKLBP14(Bu Tez)[184]	Hayır	Kare LBP + DWT	kNN	94.00	Güçlü ESC performansı
Abdoli ve ark. [189]	Evet	Ham dalga formu + CNN	1D-CNN	91.20	Yüksek kaynak talebi
FF-Orbital (Bu Tez) [236]	Hayır	Kaos tabanlı yörünge	SVM	98.81	En iyi genel performansı elde etti.
Wang ve Wang [174]	Evet	CNN tabanlı OSD	CNN	97.00	En fazla 3 konuşmacı ile sınırlıdır.
JWPat (Bu Tez) [162]	Hayır	JW'den esinlenen desen	SVM	86,74	32 konuşmacıya kadar uygulanabilir.

Bu karşılaştırmalar, bu tezde geliştirilen el yapımı modellerin, tipik olarak derin öğrenme yaklaşımlarıyla ilişkili önemli ek hesaplama yükünü azaltırken, en son yöntemlerle karşılaştırılabilir veya bunları aşan performans seviyelerine ulaşma kapasitesinin göstermektedir.

7.7.4. Genelleştirme ve Ölçeklenebilirlik

Derin öğrenme modelleri genellikle etkili eğitim için önemli miktarda veri gerektirir ve alan koşullarındaki değişimlerle karşılaşıldığında genellikle kapsamlı yeniden eğitim gerektirir. Buna karşılık, bu tezde sunulan el yapımı modeller, genelleme ve ölçeklenebilirlik açısından dikkate değer özellikler göstermektedir:

- **Esneklik ve uyarlanabilirlik:** Modelin yeniden ayarlanması için minimum gereksinim ile farklı alanlar arasında esneklik ve uyarlanabilirlik sergilemektedir.
- **Onaylanmış genellenabilirlik:** Genelleştirilebilirlik, titiz çapraz doğrulama protokolleri, özellikle 10-kat ve Bir-Kayıt-Dışı Bırak (LORO) şemaları aracılığıyla deneysel olarak doğrulanmıştır.
- **Büyük Sınıf Büyüklükleri için Ölçeklenebilirlik:** JWPat'ın 32 konuşmacı sayısı sınıfını ele alması ve FIRAT-ESC50'nin 50 sınıfına uygulanması ile örneklendiği gibi çok sayıda sınıfı barındırmak için ölçeklenebilirlik gösterilmiştir.
- **Sağlam Performans:** Gürültülü veya duygusal olarak belirsiz verileri işlerken bile doğruluğu koruyarak çeşitli görevler arasında tutarlı performans göstermektedir.

7.7.5. Karşılaştırmalı Avantajların Özeti

Bu tezde önerilen modellerin geleneksel derin öğrenme yaklaşımlarına göre sunduğu karşılaştırmalı avantajların kısa bir özeti Tablo 7.3’de sunulmaktadır:

Tablo 7.3. Karşılaştırmalı avantajların kısa özeti

Kriter	Bu Tez	Geleneksel Derin Öğrenme Modelleri
Doğruluk	Karşılaştırılabilir veya Aşılabilir	Yüksek (büyük veri kümelerinin kullanılabilirliğine bağlı)
Veri Seti Çeşitliliği	Yüksek (5 yeni kamuya açık veri seti)	Düşük ila orta
Hesaplama Gereksinimleri	Düşük (GPU hızlandırması olmayan CPU'larda çalışır)	Yüksek (genellikle eğitim için GPU kaynaklarına ihtiyaç duyar)
Yorumlanabilirlik	Yüksek (el yapımı özellik tasarımı nedeniyle)	Düşük (modellerin kara kutu yapısına atfedilir)
Küçük Verilere Uyum Sağlama	Yüksek	Düşük
Pratik Dağıtım	Yüksek (gerçek zamanlı ve uç dağıtım için oldukça uygundur)	Düşük ila orta

Bu tezde sunulan çalışma, ses sinyali işleme ve sınıflandırmada yeni bir paradigma sunarak, iyi tasarlanmış el yapımı ve kaostan ilham alan modellerin derin öğrenme yöntemlerinin performansıyla eşleşebileceğini veya aşabileceğini göstermektedir. Bu modeller, özellikle güvenlik, adli bilişim, duygu tanıma ve çevresel ses analizi gibi verimlilik, yorumlanabilirlik ve uygulanabilirliğin çok önemli olduğu alanlarda yüksek kaynak yaklaşımlarının bıraktığı kritik boşlukları doldurmaktadır.

8. SONUÇLAR

Bu tez, çevresel ses sınıflandırması, şiddet tespiti, konuşmacı sayma ve duygu tanıma dâhil olmak üzere bir dizi ses tabanlı sınıflandırma problemini ele almak için tasarlanmış kapsamlı bir yeni, hesaplama açısından verimli ve yorumlanabilir makine öğrenimi mimarileri paketi sunmaktadır. Yedi farklı model sistematik olarak önerildi ve yeni düzenlenmiş, dengeli ve halka açık veri kümeleri kullanılarak deneysel olarak doğrulandı. Bu modeller öncelikle el yapımı özelliklerden, kaostan ilham alan modellerden ve biyolojik veya yapısal tasarımlardan türetilen ilkelerden yararlanmaktadır. Böylece geleneksel derin öğrenme paradigmalarına oldukça rekabetçi bir alternatif sunmaktadır. Yapılan araştırmaya dayanarak bazı sonuçlar çıkarılmıştır.

8.1. Temel Katkıların Özeti

Bu tez, ses tabanlı makine öğrenimi alanına birçok önemli katkı sağlamaktadır. Model inovasyonu olarak yedi yeni ve farklı model geliştirilmiştir:

- ESC için LGNPAT.
- Şiddet yoğunluğunun tespiti için BP19.
- Konuşmacı sayımı için 32 konuşmacıya kadar ölçeklenebilir JWPAT
- Dokusal ses sınıflandırması için SKLBP14.
- Ses tabanlı şiddet ve duygu tespiti için kaostan ilham alan ve ileri beslemeli özellik seçim mekanizmalarını içeren TreePAT23, FF-BTP ve FF-Orbital.

Veri kümesi oluşturma: Mevcut kıyaslama kaynaklarında yaygın olan temel sınırlamaları ele alan, halka açık ve dengeli beş veri kümesi tanıtıldı. Bunlar şunları içermektedir:

- FIRAT-ESC50
- TESC
- AVD Veri Seti
- JW Konuşmacı Sayma Veri Kümesi
- Topluluk Duygu Veri Seti (CED)

Hesaplama verimliliği: Önerilen tüm modeller CPU tabanlı sistemler üzerinde tasarlanmış ve deneysel olarak değerlendirilmiş, GPU hızlandırma ihtiyacı ortadan kaldırılmıştır. Bununla birlikte, derin öğrenme metodolojileri ile elde edilenlerle rekabet edebilecek ve bazı durumlarda onlardan daha üstün sınıflandırma doğrulukları elde etmişlerdir.

Performans mükemmelliği: BP19 (%97,01), FF-BTP (%97,22) ve FF-Orbital (%98,81) dâhil olmak üzere belirli modeller, kendi görevlerinde olağanüstü performans göstermiş ve literatürde bildirilen derin CNN veya LSTM tabanlı alternatiflerin yeteneklerini sıklıkla aşmıştır.

8.2. Temel Bulgular

Bu araştırmadan elde edilen başlıca bulgular aşağıdaki gibidir:

- Ampirik kanıtlar, Ayırık Dalgacık Dönüşümü DWT veya TQWT gibi çok seviyeli dönüşümler kullanılarak çıkarılan ve daha sonra akıllı özellik seçim teknikleri (örn. INCA, FF) ile geliştirilen el yapımı ve kaostan ilham alan özelliklerin yüksek ayırt edici güç, gelişmiş yorumlanabilirlik ve gürültüye karşı gelişmiş sağlamlık sağladığını göstermektedir.
- Önerilen JWPat modeli, %86,74'lük bir doğrulukla 32'ye kadar konuşmacıyı işleme kapasitesini göstererek konuşmacı sayma literatürüne öncü bir katkıyı temsil eder ve böylece tipik olarak 3-5 konuşmacı senaryolarıyla sınırlı olan mevcut son teknoloji yöntemlerin performansını aşar.
- TreePAT23 ve FF-Orbital modelleri, geleneksel tanımlayıcılar tarafından genellikle göz ardı edilen doğrusal olmayan ve dinamik özellikleri etkili bir şekilde yakalayarak ses sinyali modellemesinde kaotik modellerin önemli uygulanabilirliğinin altını çizmektedir.

Bulgular, gömülü cihazlar, akıllı gözetim sistemleri ve mobil uygulamalar da dâhil olmak üzere gerçek zamanlı veya kaynak kısıtlı ortamlarda kullanıma uygun düşük maliyetli, yüksek doğruluklu sistemler geliştirmenin fizibilitesini teyit etmektedir.

8.3. Yaygın Etki ve Uygunluk

Bu tezde sunulan araştırma, önemli ölçüde daha geniş bir etki ve pratik uygunluk taşımaktadır:

- Bu tez, kamusal alanlarda şiddet tespiti, kalabalık duyarlılığının izlenmesi, akıllı ev ses analizi ve adli ses incelemesi dâhil olmak üzere gerçek dünyadaki zorlukları ele almak için ölçeklenebilir ve pratik çözümler sunmaktadır.
- Açık erişimli veri kümelerinin sağlanması, araştırmanın tekrarlanabilirliğini kolaylaştırmakta ve daha geniş araştırma topluluğu içinde gelecekteki kıyaslama çabalarını teşvik etmektedir.
- Önerilen metodolojiler, derin öğrenme yaklaşımlarının uygulanmasının hesaplama maliyeti, gecikme kısıtlamaları veya yetersiz mevcut veri gibi faktörler nedeniyle genellikle pratik olmadığı kaynak kısıtlı ortamlarda makine öğrenimi sistemlerinin konuşlandırılması için giriş engelini etkili bir şekilde azaltmaktadır.

Son olarak bu tezde sunulan bulgular, çağdaş makine öğrenimi paradigmalarının yalnızca derin mimari tasarımlara dayanmadığı önermesini güçlü bir şekilde desteklemektedir. Mantıklı tasarım ilkeleri sayesinde, el yapımı özellikler ve kaos tabanlı modeller, minimum hesaplama kaynakları kullanırken en son teknolojiye orantılı performans seviyelerine ulaşabilir ve böylece

ses anlama için daha erişilebilir, şeffaf ve sürdürülebilir yapay zekâ çözümlerinin geliştirilmesinin önünü açabilir.



ÖNERİLER

Bu tezin çalışmasındaki umut verici sonuçlara dayanarak, gelecekteki araştırma çalışmaları için aşağıdaki yönergeler önerilmektedir:

- **Hibrit sistem geliştirme:** El yapımı özelliklerin doğal yorumlanabilirliğini derin öğrenme modellerinin uyarlanabilir yetenekleriyle etkili bir şekilde birleştiren hibrit sistemlerin geliştirilmesi beklenmektedir.
- **Düşük kaynaklı öğrenme paradigmaları:** Özellikle sınırlı etiketli veriye sahip senaryolarda duygu ve şiddet tespiti gibi görevler için sıfır atımlı, az atımlı veya kendi kendine denetimli öğrenme paradigmalarının uygulanabilirliğinin araştırılması önerilmektedir.
- **Edge ve IoT dağıtımı:** Uç cihazlarda veya IoT sistemlerinde gerçek zamanlı dağıtım için bu modellerin uygulanması ve değerlendirilmesi gerekmektedir.
- **Veri kümesinin genişletilmesi ve çeşitliliği:** Çok dilli veya gürültüyle değişen ses kaynakları da dâhil olmak üzere daha çeşitli ses kaynaklarının dâhil edilmesi yoluyla önerilen veri kümelerinin daha da geliştirilmesi ve genişletilmesi beklenmektedir.

KAYNAKLAR

- [1] A. Bakhshi, J. García-Gómez, R. Gil-Pita, and S. Chalup, "Violence Detection in Real-Life Audio Signals Using Lightweight Deep Neural Networks," *Procedia Comput. Sci.*, vol. 222, pp. 244–251, 2023, doi: <https://doi.org/10.1016/j.procs.2023.08.162>.
- [2] M. Jangid and K. Nagpal, "Sound classification using residual convolutional network BT - Data Engineering for Smart Systems," 2022, pp. 245–254.
- [3] B. Bahmei, E. Birmingham, and S. Arzanpour, "CNN-RNN and data augmentation using deep convolutional generative adversarial network for environmental sound classification," *IEEE Signal Process. Lett.*, vol. 29, pp. 682–686, 2022.
- [4] W. Wang, F. Seraj, N. Meratnia, and P. J. Havinga, "Speaker counting model based on transfer learning from SincNet bottleneck layer," in *IEEE PerCom 2020*, 2020, pp. 1–8.
- [5] M. Sewak, S. K. Sahay, and H. Rathore, "Comparison of deep learning and the classical machine learning algorithm for the malware detection," *arXiv*, 2018.
- [6] V. Vapnik, "Three Remarks on the Support Vector Method of Function Estimation," 1998. doi: <https://doi.org/10.7551/mitpress/1130.003.0006>.
- [7] V. Vapnik, *The Nature of Statistical Learning Theory*. Springer, 2013.
- [8] M. Mahyub, L. S. Souza, B. Batalo, and K. Fukui, "Signal latent subspace: A new representation for environmental sound classification," *Appl. Acoust.*, vol. 225, p. 110181, 2024, doi: <https://doi.org/10.1016/j.apacoust.2024.110181>.
- [9] K. J. Piczak, "Environmental sound classification with convolutional neural networks," in *2015 IEEE 25th International Workshop on Machine Learning for Signal Processing (MLSP)*, 2015.
- [10] K. J. Piczak, "ESC: Dataset for environmental sound classification," *MM 2015 - Proc. 2015 ACM Multimed. Conf.*, pp. 1015–1018, 2015, doi: <https://doi.org/10.1145/2733373.2806390>.
- [11] B. Chen and J. Peng, "Environmental sound classification using two-stream deep neural network with interactive time-frequency attention," *Appl. Acoust.*, vol. 238, p. 110794, 2025, doi: <https://doi.org/10.1016/j.apacoust.2025.110794>.
- [12] K. A. Moges, N. Dalila, P. Plaskota, and S. Pyo, "Evaluation methods, testing standards, and simulation techniques of sound absorption capabilities of cementitious materials: A review," *J. Build. Eng.*, vol. 96, p. 110468, 2024, doi: <https://doi.org/10.1016/j.jobte.2024.110468>.
- [13] Z. Sun, M. Gao, M. Zhang, M. Lv, and G. Wang, "Research on recognition method of broiler overlapping sounds based on random forest and confidence interval," *Comput. Electron. Agric.*, vol. 209, p. 107801, 2023, doi: <https://doi.org/10.1016/j.compag.2023.107801>.
- [14] M. Cooke, "Auditory Scene Analysis: Computational Models," J. D. B. T.-I. E. of the S. & B. S. (Second E. Wright, Ed., Oxford: Elsevier, 2015, pp. 223–228. doi: <https://doi.org/10.1016/B978-0-08-097086-8.43010-2>.
- [15] Z.-R. Pan, B. Qiu, A.-L. Luo, Q. Li, Z.-J. Liu, and F.-J. Ren, "CoaddNet: Enhancing signal-to-noise ratio in single-shot images using convolutional neural networks with coadded image effect," *Appl. Soft Comput.*, vol. 167, p. 112395, 2024, doi: <https://doi.org/10.1016/j.asoc.2024.112395>.
- [16] Z. Li, Y. Li, W. Han, K. Chen, Y. Guo, and C. Rao, "End-to-end Wiener-CNN optimization for lensless extended depth of field imaging with phase mask," *Opt. Commun.*, vol. 546, p. 129828,

2023, doi: <https://doi.org/10.1016/j.optcom.2023.129828>.

- [17] D. Sriveni and D. L. R., “An active learning driven deep spatio-textural acoustic feature ensemble assisted learning environment for violence detection in surveillance videos,” *Eng. Sci. Technol. an Int. J.*, vol. 66, p. 102050, 2025, doi: <https://doi.org/10.1016/j.jestch.2025.102050>.
- [18] A. F. M. Agap, “A neural network architecture combining gated recurrent unit (GRU) and support vector machine (SVM) for intrusion detection in network traffic data,” *ACM Int. Conf. Proceeding Ser.*, no. 1, pp. 26–30, 2018, doi: <https://doi.org/10.1145/3195106.3195117>.
- [19] R. Akter, M. R. Islam, S. K. Debnath, P. K. Sarker, and M. K. Uddin, “A hybrid CNN-LSTM model for environmental sound classification: Leveraging feature engineering and transfer learning,” *Digit. Signal Process.*, vol. 163, p. 105234, 2025, doi: <https://doi.org/10.1016/j.dsp.2025.105234>.
- [20] A. Pudasaini, M. Al-Hawawreh, M. R. Bouadjenek, H. Hacid, and S. Aryal, “A comprehensive study of audio profiling: Methods, applications, challenges, and future directions,” *Neurocomputing*, vol. 640, p. 130334, 2025, doi: <https://doi.org/10.1016/j.neucom.2025.130334>.
- [21] H. Xu, Z. Fan, and X. Liu, “Application of personalized federated learning methods to environmental sound classification: A comparative study,” *Eng. Appl. Artif. Intell.*, vol. 135, p. 108760, 2024, doi: <https://doi.org/10.1016/j.engappai.2024.108760>.
- [22] J. Ye, T. Kobayashi, and M. Murakawa, “Urban sound event classification based on local and global features aggregation,” *Appl. Acoust.*, vol. 117, pp. 246–256, 2017, doi: <https://doi.org/10.1016/j.apacoust.2016.08.002>.
- [23] A. M. Yildiz, P. D. Barua, S. Dogan, M. Baygin, T. Tuncer, C. P. Ooi, H. Fujita ve U. R. Acharya, “A novel tree pattern-based violence detection model using audio signals,” *Expert Systems with Applications*, cilt 224, s. 120031, 2023. doi: <https://doi.org/10.1016/j.eswa.2023.120031>
- [24] P. P. Jesudhas and P. V. Ranjan, “A novel approach to build a low complexity smart sound recognition system for domestic environment,” *Appl. Acoust.*, vol. 221, p. 110028, 2024, doi: <https://doi.org/10.1016/j.apacoust.2024.110028>.
- [25] X. Han, F. Chen, and J. Ban, “A GAI-based multi-scale convolution and attention mechanism model for music emotion recognition and recommendation from physiological data,” *Appl. Soft Comput.*, vol. 164, p. 112034, 2024, doi: <https://doi.org/10.1016/j.asoc.2024.112034>.
- [26] R. D. Bullard, “Environmental justice in the 21st century: Race, class, and gender ,” 2020. doi: <https://doi.org/10.1111/josi.12389> .
- [27] R. A. P. Rodrigues, L. M. Silva, F. L. Seredynskyj, J. R. S. Fhon, A. F. Bolina ve K. Miyamura, “Analysis of violence against older adults in police reports,” *Archives of Psychiatric Nursing*, cilt 33, sayı 4, ss. 407–413, 2019.
- [28] R. Viveros-Muñoz, M. A. López-García, J. J. Valentin, A. Fernández-Alegría ve C. Hurtado-Pascal, “The SPASS dataset: A new synthetic polyphonic dataset with spatiotemporal labels of sound sources,” *Applied Acoustics*, cilt 214, s. 109665, 2023.
- [29] K. Sun, B. De Coensel, K. Filipan, F. Aletta, J. Kang, T. Brown, M. Sanchez ve L. Vu, “Classification of soundscapes of urban public open spaces,” *Landscape and Urban Planning*, cilt 189, ss. 139–155, 2019.
- [30] A. Bansal and N. K. Garg, “Environmental Sound Classification using Hybrid Ensemble Model,” *Procedia Comput. Sci.*, vol. 218, pp. 418–428, 2023, doi: <https://doi.org/10.1016/j.procs.2023.01.024>.

- [31] B. Omarov, S. Narynov, Z. Zhumanov, A. Gumar, and M. Khassanova, "State-of-the-art violence detection techniques in video surveillance security systems: A systematic review," *PeerJ Comput. Sci.*, vol. 8, pp. 1–41, 2022, doi: <https://doi.org/10.7717/PEERJ-CS.920>.
- [32] E. Kucukkulahli and A. T. Kabakus, "Towards Understanding Cat Vocalizations: A Novel Cat Sound Classification Model Based on Vision Transformers," *Appl. Acoust.*, vol. 226, p. 110218, 2024, doi: <https://doi.org/10.1016/j.apacoust.2024.110218>.
- [33] S. I. Salem, S. Shirayama, S. Shimazaki, and K. Oki, "Ensemble deep learning and anomaly detection framework for automatic audio classification: Insights into deer vocalizations," *Ecol. Inform.*, vol. 84, p. 102883, 2024, doi: <https://doi.org/10.1016/j.ecoinf.2024.102883>.
- [34] M. A. Aslam, L. Zhang, X. Liu, M. Irfan, Y. Xu, N. Li, P. Zhang, Z. Jiangbin, and L. Yaan, "Underwater sound classification using learning based methods: A review," *Expert Systems with Applications*, cilt 255, s. 124498, 2024.
- [35] F. Rahman and J. Siddiqi, "Semantic annotation of digital music," *J. Comput. Syst. Sci.*, vol. 78, no. 4, pp. 1219–1231, 2012, doi: <https://doi.org/10.1016/j.jcss.2011.10.010>.
- [36] V. K. Singh, K. Sharma, and S. N. Sur, "A survey on preprocessing and classification techniques for acoustic scene," *Expert Syst. Appl.*, vol. 229, p. 120520, 2023, doi: <https://doi.org/10.1016/j.eswa.2023.120520>.
- [37] M. M. Islam, M. Hasan, M. S. Mia, A. Al Masud, and A. R. M. T. Islam, "Early Warning Systems in Climate Risk Management: Roles and Implementations in Eradicating Barriers and Overcoming Challenges," *Nat. Hazards Res.*, 2025, doi: <https://doi.org/10.1016/j.nhres.2025.01.007>.
- [38] Z. Sa'adi, M. W. A. Ramli, W. A. N. W. A. Tajuddin, N. Z. Arman, C. H. C. Hassan, M. A. Ramzan, Z. Yusop, and N. E. Alias, "Evaluating flood early warning system and public preparedness and knowledge in urban and semi-urban areas of Johor, Malaysia: Challenges and opportunities," *Int. J. Disaster Risk Reduct.*, vol. 113, p. 104870, 2024, doi: <https://doi.org/10.1016/j.ijdr.2024.104870>.
- [39] S. A. Argyroudis, S. A. Mitoulis, L. Hofer, M. A. Zanini, E. Tubaldi, and D. M. Frangopol, "Resilience assessment framework for critical infrastructure in a multi-hazard environment: Case study on transport assets," *Sci. Total Environ.*, vol. 714, p. 136854, 2020, doi: <https://doi.org/10.1016/j.scitotenv.2020.136854>.
- [40] M. M. Hasan, M. A. Rahman, S. Arya, U. K. Ana, A. T. Asyhari, H. Tao, and S. Abu Bakar, "Search and rescue operation in flooded areas: A survey on emerging sensor networking-enabled IoT-oriented technologies and applications," *Cogn. Syst. Res.*, vol. 67, pp. 104–123, 2021, doi: <https://doi.org/10.1016/j.cogsys.2020.12.008>.
- [41] I. Abi-Zeid and J. R. Frost, "SARPlan: A decision support system for Canadian Search and Rescue Operations," *Eur. J. Oper. Res.*, vol. 162, no. 3, pp. 630–653, 2005, doi: <https://doi.org/10.1016/j.ejor.2003.10.029>.
- [42] P. Rajesh Kanna and V. Kumararaja, "Enhancing speech emotion detection with Windowed Long-Term Average Spectrum and Logistic-Rectified Linear Unit," *Eng. Appl. Artif. Intell.*, vol. 137, p. 109103, 2024, doi: <https://doi.org/10.1016/j.engappai.2024.109103>.
- [43] G. Balachandran, S. Ranjith, G. C. Jagan, and T. R. Chenthil, "Advanced speech emotion recognition utilizing optimized equivariant quantum convolutional neural network for accurate emotional state classification," *Knowledge-Based Syst.*, vol. 316, p. 113414, 2025.
- [44] N. Naderi and B. Nasersharif, "Cross Corpus Speech Emotion Recognition using transfer learning

- and attention-based fusion of Wav2Vec2 and prosody features,” *Knowledge-Based Syst.*, vol. 277, p. 110814, 2023, doi: <https://doi.org/10.1016/j.knosys.2023.110814>.
- [45] F. Daneshfar, S. J. Kabudian, and A. Neekabadi, “Speech emotion recognition using hybrid spectral-prosodic features of speech signal/glottal waveform, metaheuristic-based dimensionality reduction, and Gaussian elliptical basis function network classifier,” *Appl. Acoust.*, vol. 166, p. 107360, 2020, doi: <https://doi.org/10.1016/j.apacoust.2020.107360>.
- [46] L. Sun, B. Zou, S. Fu, J. Chen, and F. Wang, “Speech emotion recognition based on DNN-decision tree SVM model,” *Speech Commun.*, vol. 115, pp. 29–37, 2019, doi: <https://doi.org/10.1016/j.specom.2019.10.004>.
- [47] L. E. Peterson, “K-nearest neighbor,” *Scholarpedia*, vol. 4, p. 1883, 2009.
- [48] M. Galar, A. Fernández, E. Barrenechea, H. Bustince, and F. Herrera, “An overview of ensemble methods for binary classifiers in multi-class problems: Experimental study on one-vs-one and one-vs-all schemes,” *Pattern Recognit.*, vol. 44, no. 8, pp. 1761–1776, 2011, doi: <https://doi.org/10.1016/j.patcog.2011.01.017>.
- [49] C. Deepika and S. Kuchibhotla, “Deep-CNN based knowledge learning with Beluga Whale optimization using chaogram transformation using intelligent sensors for speech emotion recognition,” *Meas. Sensors*, vol. 32, p. 101030, 2024, doi: <https://doi.org/10.1016/j.measen.2024.101030>.
- [50] V. Sareen and S. K.R., “Speech Emotion Recognition using Mel Spectrogram and Convolutional Neural Networks (CNN),” *Procedia Comput. Sci.*, vol. 258, pp. 3693–3702, 2025, doi: <https://doi.org/10.1016/j.procs.2025.04.624>.
- [51] E. Boitel, A. Mohasseb, and E. Haig, “MIST: Multimodal emotion recognition using DeBERTa for text, Semi-CNN for speech, ResNet-50 for facial, and 3D-CNN for motion analysis,” *Expert Syst. Appl.*, vol. 270, p. 126236, 2025, doi: <https://doi.org/10.1016/j.eswa.2024.126236>.
- [52] S. Zhang, Y. Yang, C. Chen, X. Zhang, Q. Leng, and X. Zhao, “Deep learning-based multimodal emotion recognition from audio, visual, and text modalities: A systematic review of recent advancements and future prospects,” *Expert Syst. Appl.*, vol. 237, p. 121692, 2024, doi: <https://doi.org/10.1016/j.eswa.2023.121692>.
- [53] S. Rajesh and N. J. Nalini, “Instrument Emotion Recognition from Polyphonic Instrumental Music using MFCC and CENS Features with Deep Neural Networks,” *Procedia Comput. Sci.*, vol. 235, pp. 2548–2556, 2024, doi: <https://doi.org/10.1016/j.procs.2024.04.240>.
- [54] S. P. Mishra, P. Warule, and S. Deb, “Speech emotion recognition using multi resolution Hilbert transform based spectral and entropy features,” *Appl. Acoust.*, vol. 229, p. 110403, 2025, doi: <https://doi.org/10.1016/j.apacoust.2024.110403>.
- [55] B. Vijayan and M. V. Judy, “EmoFusionNet: A unified approach for robust speech emotion recognition,” *Digit. Signal Process.*, vol. 162, p. 105173, 2025, doi: <https://doi.org/10.1016/j.dsp.2025.105173>.
- [56] M. Z. Akhtar, R. Jahangir, Q. U. Ain, M. A. Nauman, M. Uddin, and S. S. Ullah, “UrduSER: A comprehensive dataset for speech emotion recognition in Urdu language,” *Data Br.*, p. 111627, 2025, doi: <https://doi.org/10.1016/j.dib.2025.111627>.
- [57] P. S.S., V. Menon, and S. Gopalan, “Hybrid CNN-BiLSTM architecture with multiple attention mechanisms to enhance speech emotion recognition,” *Biomed. Signal Process. Control*, vol. 100, p.

106967, 2025, doi: <https://doi.org/10.1016/j.bspc.2024.106967>.

- [58] N. Saleem, H. Elmannai, S. Bourouis, and A. Trigui, "Squeeze-and-excitation 3D convolutional attention recurrent network for end-to-end speech emotion recognition," *Appl. Soft Comput.*, vol. 161, p. 111735, 2024, doi: <https://doi.org/10.1016/j.asoc.2024.111735>.
- [59] V. V. N. Raju, R. Saravanakumar, N. Yusuf, R. Pradhan, H. Hamdi, K. A. Saravanan, V. S. Rao, and M. A. Askar, "Enhancing emotion prediction using deep learning and distributed federated systems with SMOTE oversampling technique," *Alexandria Eng. J.*, vol. 108, pp. 498–508, 2024, doi: <https://doi.org/10.1016/j.aej.2024.07.081>.
- [60] Z. Sun, H. Liu, H. Li, Y. Li, and W. Zhang, "AVERFormer: End-to-end audio-visual emotion recognition transformer framework with balanced modal contributions," *Digit. Signal Process.*, vol. 161, p. 105081, 2025, doi: <https://doi.org/10.1016/j.dsp.2025.105081>.
- [61] M. Rayhan Ahmed, S. Islam, A. K. M. Muzahidul Islam, and S. Shatabda, "An ensemble 1D-CNN-LSTM-GRU model with data augmentation for speech emotion recognition," *Expert Syst. Appl.*, vol. 218, p. 119633, 2023, doi: <https://doi.org/10.1016/j.eswa.2023.119633>.
- [62] M. A. Almulla, "A multimodal emotion recognition system using deep convolution neural networks," *J. Eng. Res.*, 2024, doi: <https://doi.org/10.1016/j.jer.2024.03.021>.
- [63] A. Hashem, M. Arif, and M. Alghamdi, "Speech emotion recognition approaches: A systematic review," *Speech Commun.*, vol. 154, p. 102974, 2023, doi: <https://doi.org/10.1016/j.specom.2023.102974>.
- [64] A. R. Bagasta, Z. Rustam, J. Pandelaki, and W. A. Nugroho, "Comparison of Cubic SVM with Gaussian SVM: Classification of Infarction for Detecting Ischemic Stroke," in *IOP Conference Series*, 2019, p. 52016.
- [65] D. Kulkarni and V. V. Dixit, "Hybrid classification model for emotion detection using electroencephalogram signal with improved feature set," *Biomed. Signal Process. Control*, vol. 100, p. 106893, 2025, doi: <https://doi.org/10.1016/j.bspc.2024.106893>.
- [66] K. P. Truong and D. A. van Leeuwen, "Automatic discrimination between laughter and speech," *Speech Commun.*, vol. 49, no. 2, pp. 144–158, 2007, doi: <https://doi.org/10.1016/j.specom.2007.01.001>.
- [67] R. O'Donovan, E. Sezgin, S. Bambach, E. Butter, and S. Lin, "Detecting Screams From Home Audio Recordings to Identify Tantrums: Exploratory Study Using Transfer Machine Learning," *JMIR Form. Res.*, vol. 4, no. 6, 2020, doi: <https://doi.org/10.2196/18279>.
- [68] T. She, H. Shao, X. Deng, and Y. Jiang, "Design and analysis of a novel quantum-classical hybrid neural network for environmental sound classification," *Appl. Acoust.*, vol. 231, p. 110527, 2025, doi: <https://doi.org/10.1016/j.apacoust.2024.110527>.
- [69] H. Xu, Y. Tian, H. Ren, and X. Liu, "A Lightweight Channel and Time Attention Enhanced 1D CNN Model for Environmental Sound Classification," *Expert Syst. Appl.*, vol. 249, p. 123768, 2024, doi: <https://doi.org/10.1016/j.eswa.2024.123768>.
- [70] F. Zhu-Zhou, D. Tejera-Berengué, R. Gil-Pita, M. Utrilla-Manso, and M. Rosa-Zurera, "Computationally constrained audio-based violence detection through transfer learning and data augmentation techniques," *Appl. Acoust.*, vol. 213, p. 109638, 2023, doi: <https://doi.org/10.1016/j.apacoust.2023.109638>.
- [71] D. Griol, J. M. Molina, and Z. Callejas, "Combining speech-based and linguistic classifiers to

- recognize emotion in user spoken utterances,” *Neurocomputing*, vol. 326–327, pp. 132–140, 2019, doi: <https://doi.org/10.1016/j.neucom.2017.01.120>.
- [72] A. Kumar, K. Sharma, and A. Sharma, “MEMoR: A Multimodal Emotion Recognition using affective biomarkers for smart prediction of emotional health for people analytics in smart industries,” *Image Vis. Comput.*, vol. 123, p. 104483, 2022, doi: <https://doi.org/10.1016/j.imavis.2022.104483>.
- [73] L. Tang, S. Hu, C. Yang, R. Deng, A. Chen, and G. Zhou, “JL-TFMSFNet: A domestic cat sound emotion recognition method based on jointly learning the time–frequency domain and multi-scale features,” *Expert Syst. Appl.*, vol. 255, p. 124620, 2024, doi: <https://doi.org/10.1016/j.eswa.2024.124620>.
- [74] R. Jahangir, M. A. Nauman, R. Alroobaea, J. Almotiri, M. M. Malik, and S. M. Alzahrani, “Deep Learning-based Environmental Sound Classification Using Feature Fusion and Data Enhancement,” *Comput. Mater. Contin.*, vol. 74, no. 1, pp. 1069–1091, 2022, doi: <https://doi.org/10.32604/cmc.2023.032719>.
- [75] H. Xu, H. Chen, X. Liu, and H. Ren, “Improving one-dimensional-based environmental sound classification models with progressive projection knowledge distillation from two-dimensional-based models,” *Eng. Appl. Artif. Intell.*, vol. 151, p. 110622, 2025, doi: <https://doi.org/10.1016/j.engappai.2025.110622>.
- [76] S. Slade, L. Zhang, H. Asadi, C. P. Lim, Y. Yu, D. Zhao, A. Panesar, P. F. Wu, and R. Gao, “Cluster search optimisation of deep neural networks for audio emotion classification,” *Knowledge-Based Syst.*, vol. 314, p. 113223, 2025, doi: <https://doi.org/10.1016/j.knosys.2025.113223>.
- [77] V. S. Nallanthighal, Z. Mostaani, A. Härmä, H. Strik, and M. Magimai-Doss, “Deep learning architectures for estimating breathing signal and respiratory parameters from speech recordings,” *Neural Networks*, vol. 141, pp. 211–224, 2021, doi: <https://doi.org/10.1016/j.neunet.2021.03.029>.
- [78] M. S. Azzaz, R. Kaibou, and B. Madani, “Co-design based FPGA implementation of an efficient new speech hyperchaotic cryptosystem in the transform domain,” *Integration*, vol. 97, p. 102197, 2024, doi: <https://doi.org/10.1016/j.vlsi.2024.102197>.
- [79] T. Özseven, “Investigation of the effectiveness of time-frequency domain images and acoustic features in urban sound classification,” *Appl. Acoust.*, vol. 211, p. 109564, 2023, doi: <https://doi.org/10.1016/j.apacoust.2023.109564>.
- [80] M. Velankar, “Study paper for Timbre identification in sound,” <http://www.ijert.org/view.php?id=5942&title=study-paper-for-timbre-identification-in-sound>, vol. 2, Oct. 2013, doi: <https://doi.org/10.13140/2.1.4565.0566>.
- [81] G. Berthomieu, V. Koehl, and M. Paquier, “Does Loudness Relate to the Strength of the Sound Produced by the Source or Received by the Ears? A Review of How Focus Affects Loudness,” *Front. Psychol.*, vol. 12, p. 583690, Jan. 2021, doi: <https://doi.org/10.3389/fpsyg.2021.583690>.
- [82] M. Crocker and J. Arenas, “Fundamentals of the direct measurement of sound intensity and practical applications,” *Acoust. Phys.*, vol. 49, pp. 163–175, Jan. 2003, doi: <https://doi.org/10.1134/1.1560378>.
- [83] M. Chen, D. Tang, Y. Xiang, L. Shi, T. Tuncer, F. Ozyurt, and S. Dogan, “Instrument sound classification using a music-based feature extraction model inspired by Mozart's Turkish March pattern,” *Alexandria Eng. J.*, vol. 118, pp. 354–370, 2025, doi: <https://doi.org/10.1016/j.aej.2025.354-370>.

<https://doi.org/10.1016/j.aej.2025.01.059>.

- [84] H. Hall, M. Bencsik, N. Capela, J. P. Sousa, and D. C. de Graaf, "Remote and automated detection of Asian hornets (*Vespa velutina nigrithorax*) at an apiary, using spectral features of their hovering flight sounds," *Comput. Electron. Agric.*, vol. 235, p. 110307, 2025, doi: <https://doi.org/10.1016/j.compag.2025.110307>.
- [85] S. Ali, S. Tanweer, S. Khalid, and N. Rao, *Mel Frequency Cepstral Coefficient: A Review*. 2021. doi: <https://doi.org/10.4108/eai.27-2-2020.2303173>.
- [86] J. Yang, F.-L. Luo, and A. Nehorai, "Spectral contrast enhancement: Algorithms and comparisons," *Speech Commun.*, vol. 39, no. 1, pp. 33–46, 2003, doi: [https://doi.org/10.1016/S0167-6393\(02\)00057-2](https://doi.org/10.1016/S0167-6393(02)00057-2).
- [87] D. O'Shaughnessy, "Review of analysis methods for speech applications," *Speech Commun.*, vol. 151, no. May, pp. 64–75, 2023, doi: <https://doi.org/10.1016/j.specom.2023.05.008>.
- [88] A. Shah, M. Kattel, A. Nepal, and D. Shrestha, *Chroma Feature Extraction*. 2019.
- [89] M. Massar, M. Fickus, E. Bryan, D. Petkie, and A. Terzuoli, "Fast computation of spectral centroids," *Adv. Comput. Math.*, vol. 35, pp. 83–97, Jul. 2011, doi: <https://doi.org/10.1007/s10444-010-9167-y>.
- [90] M. Feldman, "Hilbert transform in vibration analysis," *Mech. Syst. Signal Process.*, vol. 25, no. 3, pp. 735–802, 2011, doi: <https://doi.org/10.1016/j.ymssp.2010.07.018>.
- [91] Y. Yu, Z. Yang, Y. You, G. Liu, P. Li, Z. Yang, and W. Shan, "Spectral Roll-off Points: Estimating Useful Information Under the Basis of Low-frequency Data Representations," *CoRR*, vol. abs/2102.00369, 2021, [Çevrimiçi]. URL: <https://arxiv.org/abs/2102.00369>.
- [92] P. Spector, S. Fox, and T. Domagalski, *Emotions, Violence, and Counterproductive Work Behavior*. 2006. doi: <https://doi.org/10.4135/9781412976947.n3>.
- [93] B. Myers-Schulz, M. Pujara, R. C. Wolf, and M. Koenigs, "Inherent emotional quality of human speech sounds," *Cogn. Emot.*, vol. 27, no. 6, pp. 1105–1113, 2013, doi: <https://doi.org/10.1080/02699931.2012.754739>.
- [94] M. Tawdrous, K. D'Onofrio, R. Gifford, and E. Picou, "Emotional Responses to Non-Speech Sounds for Hearing-aid and Bimodal Cochlear-Implant Listeners," *Trends Hear.*, vol. 26, p. 233121652210830, Apr. 2022, doi: <https://doi.org/10.1177/23312165221083091>.
- [95] S. Morgan, E. Picou, E. D. Young, and S. Gustafson, "Relationship Between Auditory Distraction and Emotional Dimensionality for Non-Speech Sounds," *Ear Hear.*, Feb. 2025, doi: <https://doi.org/10.1097/AUD.0000000000001643>.
- [96] T. Ensari and B. Ensari, "Violence Detection with Machine Learning: A Sociodemographic Approach," *Eur. J. Sci. Technol.*, Jan. 2023, doi: <https://doi.org/10.31590/ejosat.1225896>.
- [97] T. Chutia and N. Baruah, *A review on emotion detection by using deep learning techniques*, vol. 57, no. 8. Springer Netherlands, 2024. doi: <https://doi.org/10.1007/s10462-024-10831-1>.
- [98] Y. Potharaju, M. Kamsali, and C. R. Kesavari, "Classification of ontological violence content detection through audio features and supervised learning," *Int. J. Intell. Eng. Syst.*, vol. 12, no. 3, pp. 20–30, 2019, doi: <https://doi.org/10.22266/IJIES2019.0630.03>.
- [99] M. Awad and R. Khanna, "Support Vector Machines for Classification," 2015, pp. 39–66. doi:

https://doi.org/10.1007/978-1-4302-5990-9_3.

- [100] M. Franzese and A. Iuliano, "Hidden Markov Models (HMM)," S. Ranganathan, M. Cannataro, and A. M. B. T.-E. of B. and C. B. (Second E. Khan, Eds., Oxford: Elsevier, 2025, pp. 324–335. doi: <https://doi.org/10.1016/B978-0-323-95502-7.00196-2>.
- [101] B. Mocanu, R. Tapu, and T. Zaharia, "Multimodal emotion recognition using cross modal audio-video fusion with attention and deep metric learning," *Image Vis. Comput.*, vol. 133, p. 104676, 2023.
- [102] K. Presannakumar and A. Mohamed, "An enhanced model for environmental sound classification using bio-inspired multi-kernel optimization algorithm," *Appl. Acoust.*, vol. 231, no. November 2024, p. 110463, 2025, doi: <https://doi.org/10.1016/j.apacoust.2024.110463>.
- [103] H. Zhao, B. Wei, P. Zhang, P. Guo, Z. Shao, S. Xu, L. Jiang, H. Hu, Y. Zeng, and P. Xiang, "Safety analysis of high-speed trains on bridges under earthquakes using a LSTM-RNN-based surrogate model," *Comput. Struct.*, vol. 294, p. 107274, 2024, doi: <https://doi.org/10.1016/j.compstruc.2024.107274>.
- [104] N. Ahmed, Z. Al Aghbari, and S. Girija, "A systematic survey on multimodal emotion recognition using learning algorithms," *Intell. Syst. with Appl.*, vol. 17, no. December 2022, p. 200171, 2023, doi: <https://doi.org/10.1016/j.iswa.2022.200171>.
- [105] J. Wagner, A. Triantafyllopoulos, H. Wierstorf, M. Schmitt, F. Burkhardt, F. Eyben, and B. W. Schuller, "Dawn of the Transformer Era in Speech Emotion Recognition: Closing the Valence Gap," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 9, pp. 10745–10759, 2023, doi: <https://doi.org/10.1109/TPAMI.2023.3263585>.
- [106] M. Bogen, S. Rilling, S. Guder, and R. Ammon, *Emergency Management and System Surveillance based on Smart Monitors as a U-CEP-based Exocortex Application*. 2013.
- [107] S. Ifkat, A. Mandal, R. Mandal, and P. Kumar, "IOT Based Smart Surveillance System," *Int. J. Res. Publ. Rev.*, vol. 4, pp. 4530–4543, Apr. 2023, doi: <https://doi.org/10.55248/gengpi.4.423.38048>.
- [108] K. R. van Daalen, S. S. Kallesøe, F. Davey, S. Dada, L. Jung, L. Singh, R. Issa, C. A. Emilian, I. Kuhn, I. Keygnaert, and M. Nilsson, "Extreme events and gender-based violence: a mixed-methods systematic review," *Lancet Planet. Heal.*, vol. 6, no. 6, pp. e504–e523, 2022, doi: [https://doi.org/10.1016/S2542-5196\(22\)00088-2](https://doi.org/10.1016/S2542-5196(22)00088-2).
- [109] M. Silva, U. Anaba, N. Jani Tulsani, P. Sripad, J. Walker, and A. Aisiri, "Gender-Based Violence Narratives in Internet-Based Conversations in Nigeria: Social Listening Study," *J. Med. Internet Res.*, vol. 25, 2023, doi: <https://doi.org/10.2196/46814>.
- [110] A. B. T.-R. M. in S. S. Saavedra, "Voice Assistants: Harmonizing Human-Computer Interactions," Elsevier, 2024. doi: <https://doi.org/10.1016/B978-0-443-13701-3.00365-0>.
- [111] J. Alcaide-Marzal and J. A. Diego-Mas, "Computers as co-creative assistants. A comparative study on the use of text-to-image AI models for computer aided conceptual design," *Comput. Ind.*, vol. 164, p. 104168, 2025, doi: <https://doi.org/10.1016/j.compind.2024.104168>.
- [112] B. Paneru, B. Paneru, and K. N. Poudyal, "Advancing human-computer interaction: AI-driven translation of American Sign Language to Nepali using convolutional neural networks and text-to-speech conversion application," *Syst. Soft Comput.*, vol. 6, p. 200165, 2024, doi: <https://doi.org/10.1016/j.sasc.2024.200165>.
- [113] A. Lerch, *Audio Content Analysis*. 2020. doi: <https://doi.org/10.48550/arXiv.2101.00132>.

- [114] M. Nasrullah, "Music: entertainment media with millions of benefits for physical performance," *Indones. J. Soc. Sci.*, vol. 12, p. 12, Jun. 2020, doi: <https://doi.org/10.20473/ijss.v12i1.21155>.
- [115] S. Sah, *Machine Learning: A Review of Learning Types*. 2020. doi: <https://doi.org/10.20944/preprints202007.0230.v1>.
- [116] F. Maleki, K. Ovens, K. Najafian, B. Forghani, C. Md, and R. Forghani, "Overview of Machine Learning Part 1," *Neuroimaging Clin. N. Am.*, vol. 30, pp. e17–e32, Nov. 2020, doi: <https://doi.org/10.1016/j.nic.2020.08.007>.
- [117] Y. Chen, W. You, L. Ou, and H. Tang, "A Review of Machine Learning Techniques For Urban Resilience Research: The Application And Progress Of Different Machine Learning Techniques In Assessing And Enhancing Urban Resilience," *Syst. Soft Comput.*, p. 200269, 2025, doi: <https://doi.org/10.1016/j.sasc.2025.200269>.
- [118] N. M. Mansour and T. Li, "540 Utilizing Chatgpt For The Development Of A Natural Language Processing (Nlp) Algorithm For The Automated Calculation Of Colonoscopy Quality Metrics," *Gastroenterology*, vol. 166, no. 5, Supplement, p. S-122-S-123, 2024, doi: [https://doi.org/10.1016/S0016-5085\(24\)00773-X](https://doi.org/10.1016/S0016-5085(24)00773-X).
- [119] Z. Wang, T. Zhang, Y. Shao, and B. Ding, "LSTM-convolutional-BLSTM encoder-decoder network for minimum mean-square error approach to speech enhancement," *Appl. Acoust.*, vol. 172, p. 107647, 2021, doi: <https://doi.org/10.1016/j.apacoust.2020.107647>.
- [120] I. Sarker, "Deep Learning: A Comprehensive Overview on Techniques, Taxonomy, Applications and Research Directions," *SN Comput. Sci.*, vol. 2, Aug. 2021, doi: <https://doi.org/10.1007/s42979-021-00815-1>.
- [121] A. Saxena and A. Saad, "Evolving an artificial neural network classifier for condition monitoring of rotating mechanical systems," *Appl. Soft Comput. J.*, vol. 7, no. 1, pp. 441–454, 2007, doi: <https://doi.org/10.1016/j.asoc.2005.10.001>.
- [122] B. Taşcı, "Multilevel hybrid handcrafted feature extraction based depression recognition method using speech," *J. Affect. Disord.*, vol. 364, no. August, pp. 9–19, 2024, doi: <https://doi.org/10.1016/j.jad.2024.08.002>.
- [123] Z. Sedaghatjoo, H. Hosseinzadeh, and B. S. Bigham, "Local Binary Pattern(LBP) Optimization for Feature Extraction," pp. 1–21, 2024, [Çevrimiçi]. URL: <http://arxiv.org/abs/2407.18665>
- [124] T. Tuncer and S. Dogan, "Novel dynamic center based binary and ternary pattern network using M4 pooling for real world voice recognition," *Appl. Acoust.*, vol. 156, pp. 176–185, 2019, doi: <https://doi.org/10.1016/j.apacoust.2019.06.029>.
- [125] A. M. Yildiz, M. Tanabe, M. Kobayashi, I. Tuncer, P. D. Barua, S. Dogan, T. Tuncer, R. S. Tan, and U. R. Acharya, "FF-BTP Model for Novel Sound-Based Community Emotion Detection," *IEEE Access*, vol. 11, pp. 108705–108715, 2023, doi: <https://doi.org/10.1109/ACCESS.2023.3318751>.
- [126] W. Na, S. Kovvuri, and J. Kim, "Multi-Variate Discrete Wavelet Technique for Advanced State-of-Charge Estimation of a Lithium-Ion Cell," *Energy Procedia*, vol. 105, pp. 4525–4530, 2017, doi: <https://doi.org/10.1016/j.egypro.2017.03.971>.
- [127] P. D. Barua, T. Keles, S. Dogan, M. Baygin, T. Tuncer, C. F. Demir, H. Fujita, R. S. Tan, C. P. Ooi, and U. R. Acharya, "Automated EEG sentence classification using novel dynamic-sized binary pattern and multilevel discrete wavelet transform techniques with TSEEG database," *Biomed. Signal Process. Control*, vol. 79, p. 104055, 2023, doi: <https://doi.org/10.1016/j.bspc.2022.104055>.

- [128] B. Taşcı, M. R. Acharya, P. D. Barua, A. M. Yildiz, M. V. Gun, T. Keles, S. Dogan, and T. Tuncer, "A new lateral geniculate nucleus pattern-based environmental sound classification using a new large sound dataset," *Appl. Acoust.*, vol. 196, p. 108897, 2022, doi: <https://doi.org/10.1016/j.apacoust.2022.108897>.
- [129] G. Boztas, T. Tuncer, O. Aydogmus, and M. Yildirim, "A DCSLBP based intelligent machine malfunction detection model using sound signals for industrial automation systems," *Comput. Electr. Eng.*, vol. 119, p. 109541, 2024, doi: <https://doi.org/10.1016/j.compeleceng.2024.109541>.
- [130] W. Zeng, J. Yuan, C. Yuan, Q. Wang, F. Liu, and Y. Wang, "A new approach for the detection of abnormal heart sound signals using TQWT, VMD and neural networks," *Artif. Intell. Rev.*, vol. 54, no. 3, pp. 1613–1647, 2021.
- [131] I. Dagal, K. Tanrıöven, A. Nayir, and B. Akin, "Adaptive Stochastic Gradient Descent (SGD) for erratic datasets," *Futur. Gener. Comput. Syst.*, vol. 166, p. 107682, 2025, doi: <https://doi.org/10.1016/j.future.2024.107682>.
- [132] H. Kohli, J. Agarwal, and M. Kumar, "An improved method for text detection using Adam optimization algorithm," *Glob. Transitions Proc.*, vol. 3, no. 1, pp. 230–234, 2022, doi: <https://doi.org/10.1016/j.gltp.2022.03.028>.
- [133] Z. Su, Y. Wang, B. Tan, Q. Cheng, X. Duan, D. Xu, L. Tian, and T. Qi, "Performance prediction of disc and doughnut extraction columns using bayes optimization algorithm-based machine learning models," *Chem. Eng. Process. - Process Intensif.*, vol. 183, p. 109248, 2023, doi: <https://doi.org/10.1016/j.cep.2022.109248>.
- [134] K. Zhang and Y. Shen, "Interface capturing schemes based on sigmoid functions," *Comput. Fluids*, vol. 280, p. 106352, 2024, doi: <https://doi.org/10.1016/j.compfluid.2024.106352>.
- [135] B. E. Tatlıcıoğlu, "A simple chaotic system using signum function," *Math. Comput. Simul.*, vol. 225, pp. 1072–1088, 2024, doi: <https://doi.org/10.1016/j.matcom.2023.11.006>.
- [136] J. Terven, D. M. Cordova-Esparza, A. Ramirez-Pedraza, E. A. Chavez-Urbiola, and J. A. Romero-Gonzalez, "Loss Functions and Metrics in Deep Learning," pp. 1–53, 2023, [Çevrimiçi]. URL: <http://arxiv.org/abs/2307.02694>
- [137] A. Mahmoudi, S. Takerkart, F. Regragui, D. Boussaoud, and A. Brovelli, "Review Article Multivoxel Pattern Analysis for fMRI Data: A Review," *Comput. Math. Methods Med.*, vol. 14, Dec. 2012, doi: <https://doi.org/10.1155/2012/961257>.
- [138] S. Widodo, H. Brawijaya, and S. Samudi, "Stratified K-fold cross validation optimization on machine learning for prediction," *Sinkron*, vol. 7, pp. 2407–2414, Oct. 2022, doi: <https://doi.org/10.33395/sinkron.v7i4.11792>.
- [139] L. C. Lee, C.-Y. Liong, and A. A. Jemain, "Validity of the best practice in splitting data for hold-out validation strategy as performed on the ink strokes in the context of forensic science," *Microchem. J.*, vol. 139, pp. 125–133, 2018, doi: <https://doi.org/10.1016/j.microc.2018.02.009>.
- [140] R. J. Urbanowicz, M. Meeker, W. La Cava, R. S. Olson, and J. H. Moore, "Relief-based feature selection: Introduction and review," Sep. 01, 2018, *Academic Press Inc.* doi: <https://doi.org/10.1016/j.jbi.2018.07.014>.
- [141] T. Ahonen, A. Hadid, and M. Pietikäinen, "Face Recognition with Local Binary Patterns BT - Computer Vision - ECCV 2004," T. Pajdla and J. Matas, Eds., Berlin, Heidelberg: Springer Berlin Heidelberg, 2004, pp. 469–481.

- [142] S. Key, H. Kurum, O. Esmez, A. H. Baig, R. Hajiyeva, S. Dogan, and T. Tuncer, "Automated hip dysplasia detection using novel FlexiLBPHOG model with ultrasound images," *Ain Shams Eng. J.*, vol. 16, no. 1, p. 103235, 2025, doi: <https://doi.org/10.1016/j.asej.2024.103235>.
- [143] K. Pothuganti, "Overview On Principal Component Analysis Algorithm In Machine Learning," *Int. Res. J. Sci. Technol.*, pp. 2582–5208, Oct. 2020.
- [144] T. Tuncer, S. Dogan, F. Özyurt, S. Brahim Belhaouari, and H. Bensmail, "Novel Multi Center and Threshold Ternary Pattern Based Method for Disease Detection Method Using Voice," *IEEE Access*, vol. PP, p. 1, Oct. 2020, doi: <https://doi.org/10.1109/ACCESS.2020.2992641>.
- [145] T. Tuncer, S. Dogan, F. Özyurt, S. Brahim Belhaouari, and H. Bensmail, "Novel Multi Center and Threshold Ternary Pattern Based Method for Disease Detection Method Using Voice," *IEEE Access*, vol. PP, p. 1, 2020, doi: <https://doi.org/10.1109/ACCESS.2020.2992641>.
- [146] J. Hansot, W. Wongsaroj, T. Sangsuwan, and N. Thong-un, "A low-cost autonomous portable poultry egg freshness machine using majority voting-based ensemble machine learning classifiers," *Smart Agric. Technol.*, vol. 10, p. 100768, 2025, doi: <https://doi.org/10.1016/j.atech.2025.100768>.
- [147] X. Qin, Z. Wu, T. Zhang, Y. Li, J. Luan, B. Wang, L. Wang, and J. Cui, "BERT-ERC: fine-tuning BERT is enough for emotion recognition in conversation," in *Proc. AAAI Conf. Artif. Intell.*, 2023, vol. 37, pp. 13492–13500, doi: <https://doi.org/10.1609/aaai.v37i11.26582>.
- [148] A. Chakhtouna, S. Sekkate, And A. Adıb, "Unveiling embedded features in Wav2vec2 and HuBERT msodels for Speech Emotion Recognition," *Procedia Comput. Sci.*, vol. 232, pp. 2560–2569, 2024, doi: <https://doi.org/10.1016/j.procs.2024.02.074>.
- [149] I. Mohino-Herranz, J. García-Gómez, S. Alonso-Diaz, J. G. Gallegos, F. J. Perez-Sanz, M. Aguilar-Ortega, and R. Gil-Pita, "Implementing transfer learning for sound event classification using the realised audio database," *Meas. Sensors*, vol. 38, p. 101711, 2025, doi: <https://doi.org/10.1016/j.measen.2024.101711>.
- [150] A. Natekin and A. Knoll, "Gradient Boosting Machines, A Tutorial," *Front. Neurorobot.*, vol. 7, p. 21, Dec. 2013, doi: <https://doi.org/10.3389/fnbot.2013.00021>.
- [151] D. Chicco and G. Jurman, "The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation," *BMC Genomics*, vol. 21, no. 1, pp. 1–13, 2020, doi: <https://doi.org/10.1186/s12864-019-6413-7>.
- [152] T. K. Chan and C. S. Chin, "A comprehensive review of polyphonic sound event detection," *IEEE Access*, vol. 8, pp. 103339–103373, 2020.
- [153] I. McLoughlin, H. Zhang, Z. Xie, Y. Song, and W. Xiao, "Robust sound event classification using deep neural networks," *IEEE Trans. Audio, Speech Lang. Process.*, vol. 23, no. 3, pp. 540–552, 2015, doi: <https://doi.org/10.1109/TASLP.2015.2389618>.
- [154] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971–987, 2002, doi: <https://doi.org/10.1109/TPAMI.2002.1017623>.
- [155] B. Taşçı, "Firat-ESC50: Environmental Sound Classification Dataset." [Çevrimiçi]. URL: <https://www.kaggle.com/datasets/buraktaci/firat-esc50>
- [156] A. M. Yıldız, T. Keleş, K. Yıldırım, S. Dogan, And T. Tuncer, "BP19: An Accurate Audio Violence Detection Model Based On One-Dimensional Binary Pattern," *Turkish J. Sci. Technol.*, vol. 18, no. 1, pp. 215–222, 2023, doi: <https://doi.org/10.55525/tjst.1244759>.

- [157] A. Subasi, T. Tuncer, S. Dogan, and D. Tanko, “Local binary pattern based feature extraction and machine learning for epileptic seizure prediction and detection,” 2020. doi: <https://doi.org/10.1088/978-0-7503-3411-2ch6>.
- [158] T. Tuncer and F. Ertam, “Neighborhood component analysis and reliefF based survival recognition methods for Hepatocellular carcinoma,” *Phys. A Stat. Mech. its Appl.*, vol. 540, p. 123143, Oct. 2019, doi: <https://doi.org/10.1016/j.physa.2019.123143>.
- [159] S. S. Keerthi, S. K. Shevade, C. Bhattacharyya, and K. R. K. Murthy, “A fast iterative nearest point algorithm for support vector machine classifier design,” *IEEE Trans. Neural Networks*, vol. 11, no. 1, pp. 124–136, 2000, doi: <https://doi.org/10.1109/72.822516>.
- [160] J. Wu, X.-Y. Chen, H. Zhang, L.-D. Xiong, H. Lei, and S.-H. Deng, “Hyperparameter Optimization for Machine Learning Models Based on Bayesian Optimizationb,” *J. Electron. Sci. Technol.*, vol. 17, no. 1, pp. 26–40, 2019, doi: <https://doi.org/10.11989/JEST.1674-862X.80904120>.
- [161] P. Primus, V. Haunschmid, P. Praher, and G. Widmer, “Anomalous Sound Detection as a Simple Binary Classification Problem with Careful Selection of Proxy Outlier Examples,” 2020, *arXiv*. doi: <https://doi.org/10.48550/ARXIV.2011.02949>.
- [162] P. D. Barua, A. M. Yildiz, N. Canpolat, T. Keles, S. Dogan, M. Baygin, I. Tuncer, T. Tuncer, R. S. Tan, H. Fujita, and U. R. Acharya, “An accurate automated speaker counting architecture based on James Webb Pattern,” *Eng. Appl. Artif. Intell.*, vol. 119, p. 105821, 2023, doi: <https://doi.org/10.1016/j.engappai.2023.105821>.
- [163] J. Daniel, T. Hull, and J. B. Barentine, “JWST: Tinsley achievements on the largest beryllium polishing project, modern technologies in space- and ground-based telescopes and instrumentation II,” in *SPIE Optical Engineering + Applications*, 2012, p. 845021.
- [164] S. Cornell, M. Omologo, S. Squartini, and E. Vincent, “Overlapped speech detection and speaker counting using distant microphone arrays,” *Comput. Speech Lang.*, vol. 72, p. 101306, 2022.
- [165] P. D. Barua, S. Dogan, T. Tuncer, M. Baygin, and U. R. Acharya, “Novel automated PD detection system using aspirin pattern with EEG signals,” *Comput. Biol. Med.*, vol. 137, p. 104841, 2021.
- [166] Z. Du, K. Liu, X. Wan, and H. Zhou, “Joint speech activity and overlap detection with multi-exit architecture,” 2022. [Çevrimiçi]. URL: <https://arxiv.org/abs/2209.11906>
- [167] B. Jiang, S. Chen, B. Wang, and B. Luo, “MGLNN: Semi-supervised learning via multiple graph cooperative learning neural networks,” *Neural Networks*, 2022.
- [168] W. Khan, K. Raj, T. Kumar, A. M. Roy, and B. Luo, “Introducing Urdu digits dataset with demonstration of an efficient and robust noisy decoder-based pseudo example generator,” *Symmetry (Basel)*, vol. 14, p. 1976, 2022.
- [169] C. Peng, X. Wu, and T. Qu, “Competing speaker count estimation on the fusion of the spectral and spatial embedding space,” 2020, pp. 3077–3081.
- [170] V. Andrei, H. Cucu, and C. Burileanu, “Overlapped Speech Detection and Competing Speaker Counting - Humans Versus Deep Learning,” *IEEE J. Sel. Top. Signal Process.*, vol. 13, no. 4, pp. 850–862, 2019, doi: <https://doi.org/10.1109/JSTSP.2019.2910759>.
- [171] V. Andrei, H. Cucu, and C. Burileanu, “Detecting overlapped speech on short timeframes using deep learning,” in *Interspeech*, 2017, pp. 1198–1202.
- [172] F.-R. Stöter, S. Chakrabarty, B. Edler, and E. A. P. Habets, “Classification vs. regression in

- supervised learning for single channel speaker count estimation,” in *ICASSP 2018*, 2018, pp. 436–440.
- [173] W. Zhang, M. Sun, L. Wang, and Y. Qian, “End-To-End Overlapped Speech Detection and Speaker Counting with Raw Waveform,” *2019 IEEE Autom. Speech Recognit. Underst. Work. ASRU 2019 - Proc.*, pp. 660–666, 2019, doi: <https://doi.org/10.1109/ASRU46091.2019.9003962>.
 - [174] Z.-Q. Wang and D. Wang, “Count and separate: Incorporating speaker counting for continuous speaker separation,” in *ICASSP 2021*, 2021, pp. 11–15.
 - [175] M. Yousefi and J. H. Hansen, “Real-time speaker counting in a cocktail party scenario using attention-guided convolutional neural network,” 2021. [Çevrimiçi]. URL: <https://arxiv.org/abs/2111.00316>
 - [176] P.-A. Grumiaux, S. Kitic, L. Girin, and A. Guérin, “Multichannel CRNN for speaker counting: An analysis of performance,” 2021. [Çevrimiçi]. URL: <https://arxiv.org/abs/2101.01977>
 - [177] T. T.-H. Duong, P.-L. Nguyen, H.-S. Nguyen, D.-C. Nguyen, H. Phan, and N. Q. Duong, “Speaker count: A new building block for speaker diarization,” in *APSIPA ASC 2021*, 2021, pp. 1149–1155.
 - [178] F.-R. Stöter, S. Chakrabarty, B. Edler, and E. A. Habets, “CountNet: Estimating the number of concurrent speakers using supervised learning,” *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 27, pp. 268–282, 2018.
 - [179] M. Kunešová, M. Hruš, Z. Zajíc, and V. Radová, “Detection of overlapping speech for the purposes of speaker diarization,” in *International Conference on Speech and Computer*, 2019, pp. 247–257.
 - [180] T. Tuncer, “df_speaker_counting_v1: Speaker Counting Dataset or Tool.” [Çevrimiçi]. URL: http://web.firat.edu.tr/turkertuncer/df_speaker_counting_v1.rar
 - [181] A. Wei, X. Li, L. Yan, Z. Wang, and X. Yu, “Machine learning models combined with wavelet transform and phase space reconstruction for groundwater level forecasting,” *Comput. Geosci.*, vol. 177, p. 105386, 2023, doi: <https://doi.org/10.1016/j.cageo.2023.105386>.
 - [182] Y. Zhang, H. Cao, H. Jiang, and B. Li, “Memory-efficient high-speed VLSI implementation of multi-level discrete wavelet transform,” *J. Vis. Commun. Image Represent.*, vol. 38, pp. 297–306, 2016, doi: <https://doi.org/10.1016/j.jvcir.2016.03.014>.
 - [183] C. Han, R. Xue, R. Zhang, and X. Wang, “A new audio steganalysis method based on linear prediction,” *Multimed. Tools Appl.*, vol. 77, no. 12, pp. 15431–15455, Jun. 2018, doi: <https://doi.org/10.1007/s11042-017-5123-x>.
 - [184] A. M. Yıldız, M. V. Gun, K. Yıldırım, T. Keles, S. Dogan, T. Tuncer, and U. R. Acharya, “SKLBP14: A new textural environmental sound classification model based on a square-kernelled local binary pattern,” *Firat Univ. J. Exp. Comput. Eng.*, vol. 2, no. 2, pp. 46–54, 2023, doi: <https://doi.org/10.5505/fujece.2023.03521>.
 - [185] F. Demir, K. Demir, and B. Arı, “Detection of brain tumor with a pre-trained deep learning model based on feature selection using MR images TT - MR görüntüleri kullanılarak öznelik seçimine dayalı ön-eğitilmiş bir derin öğrenme modeliyle beyin tümörünün tespiti,” *Firat Univ. J. Exp. Comput. Eng.*, vol. 2, no. 1, pp. 23–31, 2023, doi: <https://doi.org/10.5505/fujece.2023.36844>.
 - [186] A. Sevinç and F. Özyurt, “Classification of recyclable waste using deep learning architectures,” *Firat Univ. J. Exp. Comput. Eng.*, vol. 1, no. 3, pp. 122–128, 2022.
 - [187] J. Salamon and J. P. Bello, “Deep convolutional neural networks and data augmentation for

- environmental sound classification,” *IEEE Signal Process. Lett.*, vol. 24, no. 3, pp. 279–283, 2017.
- [188] S. Matwin, A. Kouznetsov, D. Inkpen, O. Frunza, and P. O’Blenis, “A new algorithm for reducing the workload of experts in performing systematic reviews,” *J. Am. Med. Informatics Assoc.*, vol. 17, no. 4, pp. 446–453, 2010.
 - [189] S. Abdoli, P. Cardinal, and A. L. Koerich, “End-to-end environmental sound classification using a 1D convolutional neural network,” *Expert Syst. Appl.*, vol. 136, pp. 252–263, 2019.
 - [190] S. Dogan, “TexturalESC: A Texture Dataset for Classification,” 2023. [Çevrimiçi]. URL: <https://www.kaggle.com/datasets/senguldogan/texturalesc>
 - [191] A. M. Carrington, D. G. Manuel, P. W. Fieguth, T. Ramsay, V. Osmani, B. Wernly, C. Bennett, S. Hawken, O. Magwood, Y. Sheikh, M. McInnes, and A. Holzinger, “Deep ROC Analysis and AUC as Balanced Average Accuracy, for Improved Classifier Selection, Audit and Explanation,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 1, pp. 329–341, 2023, doi: <https://doi.org/10.1109/TPAMI.2022.3145392>.
 - [192] D. M. Powers, “Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation,” 2020. [Çevrimiçi]. URL: <https://arxiv.org/abs/2010.16061>
 - [193] L. Baker, T. H. Wagner, S. Singer, and M. K. Bundorf, “Use of the Internet and E-mail for Health Care Information Results From a National Survey,” *JAMA*, vol. 289, no. 18, pp. 2400–2406, May 2003, doi: <https://doi.org/10.1001/jama.289.18.2400>.
 - [194] C. L. Hsu, H. P. Lu, and H. H. Hsu, “Adoption of the mobile Internet: An empirical study of multimedia message service (MMS),” *Omega*, vol. 35, no. 6, pp. 715–726, 2007, doi: <https://doi.org/10.1016/j.omega.2006.03.005>.
 - [195] D. U. Patton, J. S. Hong, M. Ranney, S. Patel, C. Kelley, R. Eschmann, and T. Washington, “Social media as a vector for youth violence: A review of the literature,” *Comput. Human Behav.*, vol. 35, pp. 548–553, 2014, doi: <https://doi.org/10.1016/j.chb.2014.02.043>.
 - [196] S. M. Mohtavipour, M. Saeidi, and A. Arabsorkhi, “A multi-stream CNN for deep violence detection in video sequences using handcrafted features,” *Vis. Comput.*, vol. 38, no. 6, pp. 2057–2072, 2022, doi: <https://doi.org/10.1007/s00371-021-02266-4>.
 - [197] B. Omarov, S. Narynov, Z. Zhumanov, A. Gumar, and M. Khassanova, “State-of-the-art violence detection techniques in video surveillance security systems: a systematic review,” *PeerJ Comput. Sci.*, vol. 8, p. e920, 2022, doi: <https://doi.org/10.7717/peerj-cs.920>.
 - [198] X. Hu, Z. Fan, L. Jiang, J. Xu, G. Li, W. Chen, X. Zeng, G. Yang, and D. Zhang, “TOP-ALCM: A novel video analysis method for violence detection in crowded scenes,” *Inf. Sci.*, vol. 606, pp. 313–327, 2022, doi: <https://doi.org/10.1016/j.ins.2022.05.045>.
 - [199] T. Tuncer, S. Dogan, and A. Subasi, “A new fractal pattern feature generation function based emotion recognition method using EEG,” *Chaos, Solitons and Fractals*, vol. 144, p. 110671, 2021, doi: <https://doi.org/10.1016/j.chaos.2021.110671>.
 - [200] F. Güneş Eriş and E. Akbal, “Enhancing speech emotion recognition through deep learning and handcrafted feature fusion,” *Appl. Acoust.*, vol. 222, p. 110070, 2024, doi: <https://doi.org/10.1016/j.apacoust.2024.110070>.
 - [201] T. Keles, A. M. Yildiz, P. D. Barua, S. Dogan, M. Baygin, T. Tuncer, C. F. Demir, E. J. Ciaccio, and U. R. Acharya, “A new one-dimensional testosterone pattern-based EEG sentence classification method,” *Eng. Appl. Artif. Intell.*, vol. 119, p. 105722, 2023, doi: <https://doi.org/10.1016/j.engappai.2023.105722>.

<https://doi.org/10.1016/j.engappai.2022.105722>.

- [202] L. Xue and T. Zhu, "Hybrid resampling and weighted majority voting for multi-class anomaly detection on imbalanced malware and network traffic data," *Eng. Appl. Artif. Intell.*, vol. 128, p. 107568, 2024, doi: <https://doi.org/10.1016/j.engappai.2023.107568>.
- [203] H. Cheng, D. J. Garrick, and R. L. Fernando, "Efficient strategies for leave-one-out cross validation for genomic best linear unbiased prediction," *J. Anim. Sci. Biotechnol.*, vol. 8, no. 1, pp. 1–5, 2017, doi: <https://doi.org/10.1186/s40104-017-0164-6>.
- [204] H. W. Loh, C. P. Ooi, S. Seoni, P. D. Barua, F. Molinari, and U. R. Acharya, "Application of explainable artificial intelligence for healthcare: A systematic review of the last decade (2011--2022)," *Comput. Methods Programs Biomed.*, p. 107161, 2022.
- [205] K. B. Sahay, B. Balachander, B. Jagadeesh, G. Anand Kumar, R. Kumar, and L. Rama Parvathy, "A real time crime scene intelligent video surveillance systems in violence detection framework using deep learning techniques," *Comput. Electr. Eng.*, vol. 103, no. January, p. 108319, 2022, doi: <https://doi.org/10.1016/j.compeleceng.2022.108319>.
- [206] T. Khalil, J. I. Bangash, A. W. Khan, S. A. Lashari, A. Khan, and D. A. Ramli, "Detection of violence in cartoon videos using visual features," *Procedia Comput. Sci.*, vol. 192, pp. 4962–4971, 2021, doi: <https://doi.org/10.1016/j.procs.2021.09.274>.
- [207] T. Giannakopoulos, D. Kosmopoulos, A. Aristidou, and S. Theodoridis, "Violence Content Classification Using Audio Features BT - Advances in Artificial Intelligence," G. Antoniou, G. Potamias, C. Spyropoulos, and D. Plexousakis, Eds., Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, pp. 502–507.
- [208] J. Lin, Y. Sun, and W. Wang, "Violence detection in movies with auditory and visual cues," *Proc. - 2010 Int. Conf. Comput. Intell. Secur. CIS 2010*, pp. 561–565, 2010, doi: <https://doi.org/10.1109/CIS.2010.128>.
- [209] S. Dogan, "ViolenceDetection Dataset," 2023. [Çevrimiçi]. URL: <https://www.kaggle.com/datasets/senguldogan/violencedetection>
- [210] A. M. Yildiz, P. D. Barua, S. Dogan, M. Baygin, T. Tuncer, C. P. Ooi, H. Fujita, and U. R. Acharya, "A novel tree pattern-based violence detection model using audio signals," *Expert Syst. Appl.*, vol. 224, p. 120031, 2023, doi: <https://doi.org/10.1016/j.eswa.2023.120031>.
- [211] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, pp. 971–987, 2002.
- [212] T. Tuncer, S. Dogan, and A. Subasi, "LEDPatNet19: Automated Emotion Recognition Model based on Nonlinear LED Pattern Feature Extraction Function using EEG Signals," *Cogn. Neurodyn.*, vol. 16, 2022, doi: <https://doi.org/10.1007/s11571-021-09748-0>.
- [213] H. Liu and R. Setiono, "Chi2: feature selection and discretization of numeric attributes," *Proc. Int. Conf. Tools with Artif. Intell.*, pp. 388–391, 1995, doi: <https://doi.org/10.1109/tai.1995.479783>.
- [214] C. Regenkoben, D. A. Schneider, R. E. Gur, F. Schneider, U. Habel, and T. Kellermann, "Multimodal human communication—targeting facial expressions, speech content and prosody," *Neuroimage*, vol. 60, no. 4, pp. 2346–2356, 2012.
- [215] S. Koolagudi and K. Rao, "Emotion recognition from speech: A review," *Int. J. Speech Technol.*, vol. 15, 2012, doi: <https://doi.org/10.1007/s10772-011-9125-1>.

- [216] K. Ezzameli and H. Mahersia, "Emotion recognition from unimodal to multimodal analysis: A review," *Inf. Fusion*, vol. 99, p. 101847, 2023, doi: <https://doi.org/10.1016/j.inffus.2023.101847>.
- [217] T. Mcelroy, J. Salapska-Gelleri, K. Schuller, and M. Bourgeois, "Thinking About Decisions: How Human Variability Influences Decision-Making," 2023, pp. 487–510. doi: https://doi.org/10.1007/978-3-031-15959-6_23.
- [218] S. Alanazi, M. Shabbir, N. Alshammari, M. Alruwaili, I. Hussain, and F. Ahmad, "Prediction of Emotional Empathy in Intelligent Agents to Facilitate Precise Social Interaction," *Appl. Sci.*, vol. 13, p. 1163, 2023, doi: <https://doi.org/10.3390/app13021163>.
- [219] A. Moin, F. Aadil, Z. Ali, and D. Kang, "Emotion recognition framework using multiple modalities for an effective human–computer interaction," *J. Supercomput.*, vol. 79, 2023, doi: <https://doi.org/10.1007/s11227-022-05026-w>.
- [220] W. Lin and C. Li, "Review of Studies on Emotion Recognition and Judgment Based on Physiological Signals," *Appl. Sci.*, vol. 13, p. 2573, 2023, doi: <https://doi.org/10.3390/app13042573>.
- [221] C. Filippini and A. Merla, "Systematic Review of Affective Computing Techniques for Infant Robot Interaction," *Int. J. Soc. Robot.*, vol. 15, pp. 1–17, 2023, doi: <https://doi.org/10.1007/s12369-023-00985-3>.
- [222] F. Daneshfar and M. B. Jamshidi, "An octonion-based nonlinear echo state network for speech emotion recognition in Metaverse," *Neural Networks*, vol. 163, 2023, doi: <https://doi.org/10.1016/j.neunet.2023.03.026>.
- [223] J. Wen, D. Jiang, G. Tu, C. Liu, and E. Cambria, "Dynamic interactive multiview memory network for emotion recognition in conversation," *Inf. Fusion*, vol. 91, 2022, doi: <https://doi.org/10.1016/j.inffus.2022.10.009>.
- [224] J. Li, X. Wang, G. Lv, and Z. Zeng, "GA2MIF: Graph and Attention Based Two-Stage Multi-Source Information Fusion for Conversational Emotion Detection," *IEEE Trans. Affect. Comput.*, vol. PP, pp. 1–14, 2023, doi: <https://doi.org/10.1109/TAFFC.2023.3261279>.
- [225] B. Wang, G. Dong, Y. Zhao, R. Li, Q. Cao, K. Hu, and D. Jiang, "Hierarchically stacked graph convolution for emotion recognition in conversation," *Knowledge-Based Syst.*, vol. 263, p. 110285, 2023, doi: <https://doi.org/10.1016/j.knosys.2023.110285>.
- [226] X. X. Qin, Z. Wu, J. Cui, T. Zhang, Y. Li, J. Luan, B. Wang, and L. Wang, "BERT-ERC: Fine-tuning BERT is Enough for Emotion Recognition in Conversation," 2023. [Çevrimiçi]. URL: <https://arxiv.org/abs/2301.06745>.
- [227] M. Ali, M. Hassan, K. Kifayat, J. Y. Kim, S. Hakak, and M. K. Khan, "Social media content classification and community detection using deep learning and graph analytics," *Technol. Forecast. Soc. Change*, vol. 188, no. C, Oct. 2023, doi: <https://doi.org/10.1016/j.techfore.2022.1>.
- [228] Y. Xiaoxu, S. Chen, and G. Yu, "Effects of members' response styles in an online depression community based on text mining and empirical analysis," *Inf. Process. Manag.*, vol. 60, p. 103198, 2023, doi: <https://doi.org/10.1016/j.ipm.2022.103198>.
- [229] D. Gu, M. Li, X. Yang, Y. Gu, Y. Zhao, C. Liang, and H. Liu, "An analysis of cognitive change in online mental health communities: A textual data analysis based on post replies of support seekers," *Inf. Process. Manag.*, vol. 60, no. 2, p. 103192, 2023, doi: <https://doi.org/10.1016/j.ipm.2022.103192>.
- [230] A. Utku, U. Can, and S. Aslan, "Detection of hateful twitter users with graph convolutional network

- model,” *Earth Sci. Informatics*, vol. 16, Oct. 2023, doi: <https://doi.org/10.1007/s12145-023-00940-w>.
- [231] M. Arbane, R. Benlamri, Y. Brik, and A. Alahmar, “Social media-based COVID-19 sentiment classification model using Bi-LSTM,” *Expert Syst. Appl.*, vol. 212, p. 118710, Oct. 2022, doi: <https://doi.org/10.1016/j.eswa.2022.118710>.
 - [232] A. M. # K. adından tahmini: A. M. Y. Yıldız, “SCED-V1: A Dataset for Sound Classification/Emotion Detection or Signal Analysis.” [Çevrimiçi]. URL: <https://www.kaggle.com/datasets/arifmetehanyldz/sced-v1>
 - [233] E. Akbal and T. Tuncer, “A learning model for automated construction site monitoring using ambient sounds,” *Autom. Constr.*, vol. 134, p. 104094, 2022, doi: <https://doi.org/10.1016/j.autcon.2021.104094>.
 - [234] S.-H. Fang, W.-H. Chang, Y. Tsao, H. Shih, and C. Wang, “Channel State Reconstruction Using Multilevel Discrete Wavelet Transform for Improved Fingerprinting-Based Indoor Localization,” *IEEE Sens. J.*, vol. 16, pp. 7784–7791, 2016, doi: <https://doi.org/10.1109/JSEN.2016.2602840>.
 - [235] G. Hinton, “The Forward-Forward Algorithm: Some Preliminary Investigations,” 2022.
 - [236] L. Xu, A. M. Yildiz, I. Tuncer, F. Ozyurt, S. Dogan, and T. Tuncer, “Detection of community emotions through Sound: An Investigation using the FF-Orbital Chaos-Based feature extraction model,” *Ain Shams Eng. J.*, vol. 16, no. 2, 2025, doi: <https://doi.org/10.1016/j.asej.2024.103248>.
 - [237] A. Mehrish, N. Majumder, R. Bharadwaj, R. Mihalcea, and S. Poria, “A Review of Deep Learning Techniques for Speech Processing,” *Inf. Fusion*, p. 101869, 2023.
 - [238] S. E. Sahin, G. Gulhan, P. D. Barua, T. Tuncer, S. Dogan, and O. Faust, “PrismPatNet: Novel Prism Pattern Network for Accurate Fault Classification Using Engine Sound Signals,” *Expert Syst.*, 2023.
 - [239] S. Lewis, S. L. Thomas, R. W. Blood, D. J. Castle, J. Hyde, and P. A. Komesaroff, “How do obese individuals perceive and respond to the different types of obesity stigma that they encounter in their daily lives? A qualitative study,” *Soc. Sci. Med.*, vol. 73, pp. 1349–1356, 2011.
 - [240] M. Bakardjieva, “Virtual togetherness: an everyday-life perspective,” *Media, Cult. Soc.*, vol. 25, pp. 291–313, 2003.
 - [241] R. F. Baumeister and B. J. Bushman, “Emotions and aggressiveness,” *Int. Handb. violence Res.*, pp. 479–493, 2003.
 - [242] M. Soleymani, J. Lichtenauer, T. Pun, and M. Pantic, “A Multimodal Database for Affect Recognition and Implicit Tagging,” *IEEE Trans. Affect. Comput.*, vol. 3, pp. 42–55, 2011.
 - [243] R. Aminzade and D. McAdam, “Emotions and Contentious Politics BT - Silence and Voice in the Study of Contentious Politics,” 2001, pp. 14–50.
 - [244] A. S. Imran, S. M. Daudpota, Z. Kastrati, and R. Batra, “Cross-Cultural Polarity and Emotion Detection Using Sentiment Analysis and Deep Learning on COVID-19 Related Tweets,” *IEEE Access*, vol. 8, pp. 181074–181090, 2020.
 - [245] Y. B. Ozçelik, A. Altan, and C. Kaya, “Machine Learning Approach for Early Diagnosis of Alzheimer’s Disease Using rs-fMRI and Metaheuristic Optimization with Functional Connectivity Matrices,” 2024.
 - [246] S. G. Ozkaya, N. Baygin, P. D. Barua, A. R. Singh, M. Bajaj, and M. Baygin, “Most complicated lock pattern-based seismological signal framework for automated earthquake detection,” *Int. J.*

Appl. Earth Obs. Geoinf., vol. 118, p. 103297, 2023.

- [247] N. Baygin, E. Aydemir, P. D. Barua, M. Baygin, S. Dogan, and T. Tuncer, “Automated mental arithmetic performance detection using quantum pattern-and triangle pooling techniques with EEG signals,” *Expert Syst. Appl.*, vol. 227, p. 120306, 2023.
- [248] C. Barış, C. Yanarates, and A. Altan, “A robust chaos-inspired artificial intelligence model for dealing with nonlinear dynamics in wind speed forecasting,” *PeerJ Comput. Sci.*, vol. 10, 2024.
- [249] T. Hussain, K. Dashtipour, Y. Tsao, and A. Hussain, “Audio-Visual Speech Enhancement in Noisy Environments via Emotion-Based Contextual Cues,” 2024. [Çevrimiçi]. URL: <https://arxiv.org/abs/2402.16394>
- [250] S. Ghosh, A. Priyankar, A. Ekbal, and P. Bhattacharyya, “Multitasking of sentiment detection and emotion recognition in code-mixed Hinglish data,” *Knowledge-Based Syst.*, vol. 260, p. 110182, 2023, doi: <https://doi.org/10.1016/j.knosys.2022.110182>.
- [251] J. Goldberger, G. E. Hinton, S. Roweis, and R. R. Salakhutdinov, “Neighbourhood components analysis,” *Adv. Neural Inf. Process. Syst.*, vol. 17, pp. 513–520, 2004.
- [252] J. W. Chang, H. S. Ma, and Z. Y. Hu, “Multi-Level Transfer Learning using Incremental Granularities for environmental sound classification and detection,” *Appl. Soft Comput.*, vol. 169, no. November 2024, p. 112619, 2025, doi: <https://doi.org/10.1016/j.asoc.2024.112619>.
- [253] Y. Zhang, J. Wang, Y. Liu, L. Rong, Q. Zheng, D. Song, P. Tiwari, and J. Qin, “A Multitask learning model for multimodal sarcasm, sentiment and emotion recognition in conversations,” *Inf. Fusion*, vol. 93, pp. 282–301, 2023, doi: <https://doi.org/10.1016/j.inffus.2023.01.005>.
- [254] J. Singh, L. B. Saheer, and O. Faust, “Speech Emotion Recognition Using Attention Model,” *Int. J. Environ. Res. Public Health*, vol. 20, no. 6, 2023, doi: <https://doi.org/10.3390/ijerph20065140>.
- [255] S. Dong, Z. Xia, X. Pan, and T. Yu, “Environmental sound classification based on improved compact bilinear attention network,” *Digit. Signal Process.*, vol. 141, p. 104170, 2023, doi: <https://doi.org/10.1016/j.dsp.2023.104170>.
- [256] L. Yu, F. Xu, Y. Qu, and K. Zhou, “Speech emotion recognition based on multi-dimensional feature extraction and multi-scale feature fusion,” *Appl. Acoust.*, vol. 216, p. 109752, 2024, doi: <https://doi.org/10.1016/j.apacoust.2023.109752>.

ÖZGEÇMİŞ

Arif Metehan YILDIZ

KİŞİSEL BİLGİLER

Doğum Yeri :
Doğum Yılı :
Uyruğu :
Adres :
E-posta :
Yabancı Diller :

ARAŞTIRMACI BİLGİLERİ

Öğrenci Orcid ID : 0000-0003-0451-8600
Danışman Orcid ID : 0000-0002-5126-6445

EĞİTİM BİLGİLERİ

Yüksek Lisans : “Makine Öğrenmesi Tabanlı Kötücül Yazılımların Tespiti”
Fırat Üniversitesi, Fen Bilimleri Enstitüsü, Adli Bilişim Mühendisliği Anabilim Dalı, 2020
Danışman: Prof. Dr. Şengül DOĞAN
Lisans : Fırat Üniversitesi, Teknoloji Fakültesi, Adli Bilişim Mühendisliği Bölümü, 2018
Lise : Mehmet Koloğlu Anadolu Lisesi, Elazığ, 2013

ARAŞTIRMA DENEYİMİ

- ✓ Pc-3000 Express, Tableau TD3, Ultra Dock
- ✓ C++, Java, HTML, MATLAB, Python
- ✓ Forensics Explorer, EnCase, Magnet Axiom, FTK, X-Ways, Autopsy, MobilEdit Forensics

İŞ DENEYİMİ

2016 – 2016 : FORDEFENCE Adli Bilişim Laboratuvarı
2017 – 2017 : TRT WORLD IT
2018 – 2018 : MyDISK Adli Bilişim Ve Veri Kurtarma
2020 – 2021 : InitAck Adli Bilişim
2021 – ∞ : Ardahan Üniversitesi (Öğretim Görevlisi)

AKADEMİK FAALİYETLER

Makaleler:

1. T. Keleş, A. M. Yıldız, P. D. Barua, S. Doğan, M. Baygın, T. Tuncer, C. F. Demir, E. J. Ciaccio, and U. R. Acharya, “A new one-dimensional testosterone pattern-based EEG sentence classification method,” Eng. Appl. Artif. Intell., 2023
2. B. Taşcı, M. R. Acharya, P. D. Barua, A. M. Yıldız, M. V. Gün, T. Keleş, S. Doğan, and T. Tuncer, “A new lateral geniculate nucleus pattern-based environmental sound classification using a new large sound dataset,” Eng. Appl. Artif. Intell., 2023.
3. S. Kirik, S. Doğan, M. Baygın, P. D. Barua, C. F. Demir, T. Keleş, A. M. Yıldız, N. Baygın, İ. Tuncer, and T. Tuncer, “FGPat18: Feynman graph pattern-based language detection model using EEG signals,” Biomed. Signal Process. Control, vol. 85, p. 104927, 2023.

4. I. Topaloglu, P. D. Barua, A. M. Yildiz, T. Keles, S. Dogan, M. Baygin, H. F. Gul, T. Tuncer, R. S. Tan, and U. R. Acharya, "Explainable attention ResNet18-based model for asthma detection using stethoscope lung sounds," *Eng. Appl. Artif. Intell.*, vol. 126, p. 106887, 2023, doi: <https://doi.org/10.1016/j.engappai.2023.106887>.
5. M. Kilic, P. D. Barua, T. Keles, A. M. Yildiz, I. Tuncer, S. Dogan, M. Baygin, T. Tuncer, M. Kuluozturk, R. S. Tan, and U. R. Acharya, "GCLP: An automated asthma detection model based on global chaotic logistic pattern using cough sounds," *Eng. Appl. Artif. Intell.*, vol. 127, p. 107184, 2024, doi: <https://doi.org/10.1016/j.engappai.2023.107184>.
6. S. Kirik, S. Dogan, M. Baygin, P. D. Barua, C. F. Demir, T. Keles, A. M. Yildiz, N. Baygin, I. Tuncer, T. Tuncer, R. S. Tan, and U. R. Acharya, "FGPat18: Feynman graph pattern-based language detection model using EEG signals," *Biomed. Signal Process. Control*, vol. 85, p. 104927, 2023, doi: <https://doi.org/10.1016/j.bspc.2023.104927>.
7. P. D. Barua, A. M. Yıldız, N. Canpolat, T. Keleş, S. Doğan, M. Baygin, İ. Tuncer, T. Tuncer, R.-S. Tan, H. Fujita, and U. R. Acharya, "An accurate automated speaker counting architecture based on James Webb Pattern," *Eng. Appl. Artif. Intell.*, vol. 119, Mar. 2023, Art. no. 105821.
8. A. M. Yildiz, P. D. Barua, S. Doğan, M. Baygin, T. Tuncer, C. P. Ooi, H. Fujita, and U. R. Acharya, "A novel tree pattern-based violence detection model using audio signals," *Expert Syst. Appl.*, vol. 224, Aug. 2023, Art. no. 120031.
9. A. M. Yildiz, M. Tanabe, M. Kobayashi, I. Tuncer, P. D. Barua, S. Dogan, T. Tuncer, R. S. Tan, and U. R. Acharya, "FF-BTP Model for Novel Sound-Based Community Emotion Detection," *IEEE Access*, vol. 11, pp. 108705–108715, 2023, doi: <https://doi.org/10.1109/ACCESS.2023.3318751>.
10. S. Kirik, I. Tasci, P. D. Barua, A. M. Yildiz, T. Keles, M. Baygin, I. Tuncer, S. Dogan, T. Tuncer, A. Devi, R. S. Tan, and U. R. Acharya, "DSWIN: Automated hunger detection model based on hand-crafted decomposed shifted windows architecture using EEG signals," *Knowledge-Based Syst.*, vol. 300, p. 112150, 2024, doi: <https://doi.org/10.1016/j.knosys.2024.112150>.
11. L. Xu, A. M. Yıldız, İ. Tuncer, F. Özyurt, S. Doğan, and T. Tuncer, "Detection of community emotions through sound: An investigation using the FF-Orbital Chaos-Based feature extraction model," *Ain Shams Eng. J.*, vol. 16, Feb. 2025, Art. no. 103248.
12. A. M. Yıldız, M. V. Gun, K. Yıldırım, T. Keles, Ş. Doğan, T. Tuncer, and U. R. Acharya, "SKLBP14: A new textural environmental sound classification model based on a square-kernelled local binary pattern," *Firat University Journal of Experimental and Computational Engineering*, vol. 2, no. 2, pp. 46–54, Jun. 2023.
13. A. M. Yıldız, T. Keleş, K. Yıldırım, Ş. Doğan, and T. Tuncer, "BP19: An Accurate Audio Violence Detection Model Based On One-Dimensional Binary Pattern," *Turkish Journal of Science and Technology*, vol. 18, no. 1, pp. 215–222, Jan. 2023.
14. K. Yıldırım, M. E. Demir, T. Keleş, A. M. Yıldız, Ş. Doğan, and T. Tuncer, "A YARA-based approach for detecting cyber security attack types," *Firat University Journal of Experimental and Computational Engineering*, vol. 2, no. 2, pp. 55–68, 2023.
15. T. Keleş, Y. E. Çolak, N. İlhan, K. Yıldırım, A. M. Yıldız, T. Tuncer, and Ş. Doğan, "Investigation of Microsoft Teams and Skype as distance education tools in terms of digital forensics," *Firat University Journal of Experimental and Computational Engineering*, vol. 3, no. 2, pp. 204–215, Dec. 2024.

Bildiriler:

1. A. M. Yıldız, T. Tuncer, Ş. Doğan, and E. Akbal, “Malware detection method based on exemplar and machine learning,” in Proc. Int. Eng. Nat. Sci. Conf. (IENSC), Diyarbakır, Turkey, Nov. 2019. p. 474-489.
2. A. M. Yıldız and T. Tuncer, “A novel local cross T pattern (LCTP) for facial image recognition,” in Proc. Int. Conf. Cyber Security Digit. Forensics (ICONSEC), Yalova, jun. 2021, p. 33-44

Projeler:

1. EEG ve EKG Sinyalleri Kullanılarak Makine Öğrenmesi Tabanlı Otomatik Duygu Tespiti ve Sınıflandırma, BAP, Araştırmacı, 2021-2024, Bütçe: 14365 TL
2. Yoğun Bilişsel Aktivite Sırasında EEG Sinyalleri Kullanılarak Mental Stres Ölçümü, BAP, Araştırmacı, 2021-2024, Bütçe: 14361 TL
3. EEG Sinyallerinin Tercüme Edilmesi İçin Kimyasal Yapıların Çizgeleri Tabanlı Yeni Nesil Öğrenme Modellerinin Oluşturulması, TÜBİTAK 1002, 2021-2023.
4. Makine öğrenmesi ve Derin Öğrenme Tabanlı Kötücül Yazılım Tespiti, TÜBİTAK 2242 Üniversite Öğrencileri Proje Yarışması, Erzurum Bölge BİRİNCİLİĞİ, 2020.
5. KONUSMACISAYAR, TEKNOFEST, İnsanlık Yararına Teknoloji Yarışması, Sosyal İnovasyon, FİNALİST,2022.