

**BAŐKENT ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

SES SİNYALLERİNDE DUYGU TANIMA VE GERİ ERİŐİMİ

E. SONAT ERDEM

YÜKSEK LİSANS TEZİ

2014

SES SİNYALLERİNDE DUYGU TANIMA VE GERİ ERİŞİMİ

EMOTION RECOGNITION AND RETRIEVAL IN AUDIO SIGNALS

E. SONAT ERDEM

Başkent Üniversitesi
Lisansüstü Eğitim Öğretim ve Sınav Yönetmeliğinin
BİLGİSAYAR Mühendisliği Anabilim Dalı İçin Öngördüğü
YÜKSEK LİSANS TEZİ
olarak hazırlanmıştır.

2014

“Ses Sinyallerinde Duygu Tanıma ve Geri Erişimi” başlıklı bu çalışma, jürimiz tarafından 12/09/2014 tarihinde, **BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI** 'nda **YÜKSEK LİSANS TEZİ** olarak kabul edilmiştir.

Başkan :
(Prof. Dr. Adnan YAZICI)

Üye (Danışman) :
(Yrd. Doç. Dr. Mustafa SERT)

Üye :
(Doç. Dr. Hasan OĞUL)

ONAY

...../...../.....

Prof. Dr. Emin AKATA
Fen Bilimleri Enstitüsü Müdürü

TEŐEKKÜR

Tez alıőmamda ufkumu aan sevgili hocam Yrd. Do. Dr. Mustafa Sert'e ve daima beni destekleyen biricik eőim Ümran Iőık Erdem'e gönülden teőekkür ederim.

Tez jürime katıldıkları için sayın Prof. Dr. Adnan Yazıcı ve Do. Dr. Hasan Ođul hocalarıma içten teőekkür ederim.

ÖZ

SES SİNYALLERİNDE DUYGU TANIMA VE GERİ ERİŞİMİ

E. Sonat ERDEM

Başkent Üniversitesi Fen Bilimleri Enstitüsü

Bilgisayar Mühendisliği Anabilim Dalı

Ses sinyalinde duygu tanıma özellikle, görsel bilginin kısıtlı ya da hiç olmadığı durumlarda önem kazanmaktadır. Bu çalışmada, tam ve genişletilebilir bir ses tabanlı duygu tanıma ve geri erişim çatısı önerilmiştir. Makine öğrenme yöntemi olarak Destek Vektör Makineleri (DVM) kullanılmış ve performansını artırmak amacıyla parametre optimizasyonu gerçekleştirilmiştir. Ses içerik analizlerinde, uygun pencere ve atlama sürelerine karar verebilmek için ampirik analizler gerçekleştirilmiştir. Çalışmada, gürbüz öznitelikler bulmak amacıyla, 20 ses özniteliği üzerinde, DVM kullanılarak kapsamlı analizler yapılmış ve sonuçlar değerlendirilmiştir. Ayrıca, ses sinyallerinin duygu-tabanlı geri erişimi için, nokta, aralık ve en yakın komşuluk olarak adlandırılan sorgu türleri geliştirilmiş ve geri erişim başarımları değerlendirilmiştir. Deneysel sonuçlara göre, sınıflandırıcı parametre optimizasyonu ve önerilen ses analiz yöntemleri, dayanak tanıma başarımlarını arttırmaktadır.

ANAHTAR SÖZCÜKLER: Ses duygu tanıma, ses duygu geri erişimi, Destek Vektör Makineleri (DVM), MFCC, MPEG-7.

Danışman: Yrd.Doç.Dr. Mustafa SERT, Başkent Üniversitesi, Bilgisayar Mühendisliği Bölümü.

ABSTRACT

EMOTION RECOGNITION AND RETRIEVAL IN AUDIO SIGNALS

E. Sonat ERDEM

Baskent University Institute of Science

Department of Computer Engineering

Emotion recognition from audio signals become more of significance especially when visual information is limited or absent. In this study, a complete and extensible audio-based emotion recognition and retrieval framework is proposed. Support Vector Machine (SVM) is employed as the machine learning scheme and parameter optimization methods are carried out to improve the performance of the learner. In audio content analysis, empirical analyses are performed to decide the proper window and hop sizes. In the study, extensive analyses are conducted using 20 audio features with SVM classifier to determine robust audio features and to evaluate the results. In addition, flexible querying abilities, namely point, range, and nearest neighbor are developed and retrieval performance is evaluated for emotion-based retrieval of audio signals. Based on the experiments, parameter optimization of the classifier along with the proposed audio analysis methods improve the baseline recognition accuracy.

KEYWORDS: Audio emotion recognition, audio emotion retrieval, Support Vector Machines (SVM), MFCC, MPEG-7.

Supervisor: Asst. Prof. Dr. Mustafa SERT, Baskent University, Department of Computer Engineering.

İçindekiler Listesi

ÖZ	i
ABSTRACT	ii
İçindekiler Listesi	iii
Çizelgeler Listesi	vi
Şekiller Listesi	viii
Kısaltmalar	ix
1. GİRİŞ.....	1
1.1 İlgili Çalışmalar	2
1.2 Tezin Organizasyonu	8
2. TEMEL TANIM, KAVRAM VE YÖNTEMLER	9
2.1 Duygu ve Ses Sinyali	9
2.1.1 Duygu	9
2.1.2 Duygu sınıflandırma	9
2.1.3 Ses sinyali.....	12
2.2 Makine Öğrenme	13
2.2.1 Öğrenme çeşitleri.....	15
2.2.1.1 Güdümlü öğrenme	15
2.2.1.2 Güdümsüz öğrenme	15
2.3 Duygu Tanıma Problemi	16
2.4 Destek Vektör Makineleri	17
2.4.1 Doğrusal olarak ayrılabilme durumuna göre sınıflandırma.....	18
2.4.2 Doğrusal olarak ayrılamama durumuna göre sınıflandırma	22
2.4.3 Doğrusal olmayan sınıflandırma	23
2.4.3.1 Doğrusal çekirdek	24
2.4.3.2 Polinom çekirdek.....	25
2.4.3.3 Radyal tabanlı çekirdek.....	25
2.4.3.4 Sigmoid çekirdek.....	26
2.4.4 Çok sınıflı sınıflandırma	26
2.4.4.1 Bire karşı hepsi (BKH)	27
2.4.4.2 Bire karşı bir (BKB)	27
2.5 En Yakın N Komşu Sorgusu	28
3. DUYGU TANIMA.....	30

3.1. Veri Kümeleri	30
3.1.1 Emo-DB Veri Kümesi [26]	31
3.2 Yöntemler	33
3.2.1 Ön işlem ve zamansal bölütleme	33
3.2.1.1 Pencere boyutu.....	33
3.2.1.2 Atlama süresi	34
3.2.1.3 Zamansal bölütleme.....	35
3.2.2 Öznitelik çıkarım	36
3.2.2.1 Öznitelik vektörleri.....	37
3.2.2.1.1. Mel Frequency Cepstral Coefficients (MFCC).....	40
3.2.3 Sınıflandırma yöntemi	43
3.2.4 Çapraz doğrulama	44
3.2.5 Recognition accuracy (tanıma doğruluğu) ve hata matrisi	45
3.3 Araçlar	46
3.3.1 İşletim sistemi	46
3.3.2 Matlab	47
3.3.3 libSVM	47
4. DUYGU GERİ ERİŞİMİ	49
4.1 Veri Kümesi	49
4.2 Sorgular	49
4.2.1 Anahtar kelime sorgusu	49
4.2.2 İçerik tabanlı sorgu.....	50
4.2.2.1 Öznitelik tabanlı	50
4.2.2.2 Sınıflandırma tabanlı	51
4.2.3 Zamansal sorgular	51
5. DENEYSEL ÇALIŞMALAR VE DEĞERLENDİRME.....	53
5.1 Duygu Sınıflandırma Deneyleri	53
5.1.1 DVM parametre optimizasyonu.....	53
5.1.2 Pencere ve atlama süresi analizi	58
5.1.3 Gürbüz öznitelik analizi	60
5.2 Duygu Geri Erişim Deneyleri.....	68
5.2.1 Veri kümesi	68
5.2.2 Anahtar kelime sorgusu	69
5.2.3 İçerik tabanlı sorgu.....	70

5.2.3.1 Öznitelik tabanlı	70
5.2.3.2 Sınıflandırma tabanlı	72
5.2.4 Zamansal sorgu	72
6. SONUÇ	74
7. REFERANSLAR.....	76
Ekler	81
Ek-1: Öznitelik performansları çalışması için diğer özniteliklerin hata matrisi ...	81

Çizelgeler Listesi

Çizelge 2.1 Ekman duygu sınıflandırması.	10
Çizelge 3.1 Emo-DB veri kümesinin 535 adet duygu ses sinyalinin dağılımları ...	31
Çizelge 3.2 Emo-DB veri kümesinin konuşmacı bilgileri	32
Çizelge 3.3 Emo-DB metin içerikleri	32
Çizelge 3.4 Deneysel çalışmada kullanılan pencere ve atlama boyutları	36
Çizelge 3.5 Kullanılan ses sinyali öznelik vektörü çıkarım yöntemleri.....	38
Çizelge 3.6 İki sınıflı veri için hata matrisi	45
Çizelge 4.1 TV dizileri veri kümesinin 409 adet duygu ses sinyalinin dağılımları .	49
Çizelge 5.1 DVM varsayılan parametreleri	54
Çizelge 5.2 DVM varsayılan parametreler için hata matrisi	54
Çizelge 5.3 DVM varsayılan parametreler için performans değerleri	55
Çizelge 5.4 DVM parametre değişikliklerinde tanıma doğruluk oranları.....	56
Çizelge 5.5 DVM için maksimum doğruluk veren parametre değerleri	56
Çizelge 5.6 DVM için maksimum doğruluk hata matrisi	57
Çizelge 5.7 DVM için maksimum doğruluk performans değerleri.....	57
Çizelge 5.8 DVM için maksimum doğruluk veren pencere boyutu ve atlama süresi için hata matrisi.....	59
Çizelge 5.9 DVM için maksimum doğruluk veren pencere boyutu ve atlama süresi için performans değerleri	59
Çizelge 5.10 MFCC+ASS+ASC bileşik özneliği hata matrisi.....	62
Çizelge 5.11 MFCC+ASS+ASC bileşik özneliği performans değerleri	62
Çizelge 5.12 MFCC özneliği hata matrisi	63
Çizelge 5.13 MFCC özneliği performans değerleri.....	63
Çizelge 5.14 MFCC+ASF bileşik özneliği hata matrisi	64
Çizelge 5.15 MFCC+ASF bileşik özneliği performans değerleri.....	64
Çizelge 5.16 ASF+ASS+ASC bileşik özneliği hata matrisi.....	65
Çizelge 5.17 ASF+ASS+ASC bileşik özneliği performans değerleri	65
Çizelge 5.18 ASF özneliği hata matrisi	66
Çizelge 5.19 ASF özneliği performans değerleri	66
Çizelge 5.20 TV dizilerinden oluşturulmuş veri kümesi için hata matrisi	68
Çizelge 5.21 TV dizilerinden oluşturulmuş veri kümesi için performans değerleri	69

Çizelge 5.22 Öznitelik tabanlı sorgulama için performans değerleri	71
Çizelge Ek1.1 MP+ASS+ASC bileşik özniteliği hata matrisi	81
Çizelge Ek1.2 ZCR özniteliği hata matrisi.....	81
Çizelge Ek1.3 MP özniteliği hata matrisi.....	82
Çizelge Ek1.4 LPC+ASC+ASS bileşik özniteliği hata matrisi	82
Çizelge Ek1.5 ASE özniteliği hata matrisi	83
Çizelge Ek1.6 LPC özniteliği hata matrisi	83
Çizelge Ek1.7 ASP özniteliği hata matrisi	84
Çizelge Ek1.8 HSC özniteliği hata matrisi	84
Çizelge Ek1.9 HSD özniteliği hata matrisi	85
Çizelge Ek1.10 SC özniteliği hata matrisi	85
Çizelge Ek1.11 F0 özniteliği hata matrisi	86
Çizelge Ek1.12 HSS özniteliği hata matrisi.....	86
Çizelge Ek1.13 ASC özniteliği hata matrisi.....	87
Çizelge Ek1.14 ASS özniteliği hata matrisi	87
Çizelge Ek1.15 HSV özniteliği hata matrisi.....	88

Şekiller Listesi

Şekil 2.1	Plutchik'in duygular tekerleği.....	11
Şekil 2.2	Russel'in Circumplex Modeli	12
Şekil 2.3	Analog ve dijital ses sinyali	13
Şekil 2.4	Örnek bir duygu tanıma sistemi	17
Şekil 2.5	Doğrusal olarak ayrılabilen veriler, aşırı düzlemler	18
Şekil 2.6	Optimal aşırı düzlem	19
Şekil 2.7	Verilerin doğrusal olarak ayrılabilmesi	22
Şekil 2.8	Doğrusal ayrılabilmeyen verilerin yüksek boyutlu uzaya haritalanması	24
Şekil 2.9	Gaussian çekirdek parametre değişikliği ve veri ayrışımı.....	26
Şekil 2.10	En yakın N(K) komşu sorgusu [63].....	29
Şekil 3.1	Sistem blok diyagramı.....	33
Şekil 3.2	Duygu barındıran ses sinyalinde pencere ve atlama boyutu.....	35
Şekil 3.3	MFCC özniteliklerinin çıkarılma blok diyagramı	40
Şekil 3.4	Ses sinyaline hamming penceresi uygulaması	41
Şekil 3.5	Ses sinyali için Mel Frekansı Kepstrumu.....	42
Şekil 3.6	Farklı C ve γ parametreleri için sınıflandırma durumları [52].....	44
Şekil 3.7	N kere çapraz doğrulama	45
Şekil 3.8	Matlab uygulaması ekran görüntüsü	47
Şekil 4.1	Allen'in aralık algoritması	52
Şekil 5.1	Pencere boyutu ve atlama sürelerinin ortalama doğruluk oranları	58
Şekil 5.2	Emo-DB üzerinde özniteliklerin ortalama doğrulukları.....	60
Şekil 5.3	Anahtar kelimeye göre sorgu	70
Şekil 5.4	İçerik tabanlı sorgulama - öznitelik karşılaştırması.....	71
Şekil 5.5	İçerik tabanlı sorgulama – sınıflandırma	72
Şekil 5.6	Zamansal sorgulama.....	73

Kısaltmalar

DVM	Destek Vektör Makineleri
MFCC	Mel-Frequency Cepstral Coefficients
YSA	Yapay Sinir Ağları
SA-DVM	Sezgisel Arama- Destek Vektör Makineleri
SMM	Saklı Markov Modeli
ÇYUKSH	Çift Yönlü Uzun Kısa Dönem Hafıza
ÇTF	Çapsal Tabanlı Fonksiyonlar
ÇTF-DVM	Çapsal Tabanlı Fonksiyonlar – Destek Vektör Makineleri
ÇEBA	Çekirdek Entropi Bileşen Analizi
BAG	Beyin Akım Grafiği
EKG	Elektrokardiyografi
GDC	Galvanik Deri Cevabı
TBA	Temel Bileşen Analizi
DVR	Destek Vektör Regresyon
ÇDR	Çoklu Doğrusal Regresyon
GKM	Gaussian Karıştırılmış Modeli
BKH	Bire Karşı Hepsi
BKB	Bire Karşı Bir
ODÖ	Ortalama Duyarlık Ölçütü

1. GİRİŞ

İnsan ve makine etkileşiminde en doğal yollar olarak kabul edilen işitsel ve görsel öğeler, son yıllarda yapılan duygu tanımlama çalışmalarında önemli bir veri kaynağı olarak kullanılmaktadır. Duyguları tanımlayabilmek psikoloji, bilişsel bilişim, bilgisayar bilimi, sinir bilimi gibi birçok disiplini de içine alan geniş bir kapsama dağılmıştır. Duygu tanımlama da nabız, kan basıncı, yüz mimikleri, el kol hareketleri, ses, beyin dalgaları gibi birçok fizyolojik parametreden faydalanılmaktadır [5]. Günümüzde, duygusal parametrelerin hepsi ölçülüp, anlamlandırılabilir. İnsan davranışındaki duygusal desenlerin, yüksek güvenilirlikte tam otomatik olarak algılanması ve bunun değerlendirilmesi, insan-makine etkileşimindeki en büyük hedeflerden biri olmaya devam etmektedir.

İnsan duygu çıkarımında var olan yaklaşımları, genel olarak tekil ve çoklu kip olmak üzere iki kip altında toplayabiliriz. Tekil kip yaklaşımında, bilgisayar tarafından işlenen veri, ses veya görüntü sinyallerinden alınır. Çoklu kip yaklaşımında ise bu iki kaynaktan gelen tamamlayıcı bilgiler birleştirilerek, duygunun tanınma doğruluğu artırılır. Geçmiş çalışmaların birçoğu, görüntü sinyalinden duygu çıkarımı konusuna ağırlık vermiş, pek çok araştırmacı bu konuda çalışmalar yapmıştır. Son yıllarda, ses sinyali ile ilgili çalışmaların sayısı artmaktadır [1]-[5], [22]-[24]. Yapılan çalışmalar veri kümeleri, öznitelikler, yöntemler konusunda sürekli artan bir şekilde katkı sağlamaktadır. Ses sinyalinde farklı doğruluk yüzdeleri ile mutluluk, üzüntü, kızgınlık, korku, sıkıntı, iğrenme, nötr gibi duygular çıkarılabilmektedir. Ses sinyalinde duygu tanımlayabilmenin, insan davranışını anlamlandırma, karar destek mekanizmalarını destekleme, çoklu ortam verilerini anlamlandırma ve insan-robot etkileşimi gibi birçok uygulama alanı bulunmaktadır. Ses sinyalinde duygu tanımlama özellikle, görsel bilginin kısıtlı ya da hiç olmadığı durumlarda önem kazanmaktadır. Örneğin, insan-robot etkileşiminde; görsel kanalların yetersiz kaldığı ya da hiç olmadığı durumlarda, ses bileşeni, duygu tanımada önemli bir kanaldır.

Ses sinyali, konuşanın fizyolojisi, ruh hali, yaşı, cinsiyeti, lehçesi gibi birçok bilgiyi aynı anda barındırabilen karmaşık bir işarettir. Ses sinyalini işlemek, görsel sinyale

göre, görece daha zordur ve ses sinyalinde duygu tanımda önemli zorluklar karşımıza çıkmaktadır. İlki, duygu tanımlama ile ilgili, güçlü akustik öz nitelikler tanımlayabilmektir. İkincisi ise, sınıflandırıcının öğrenme algoritmaları için, en uygun parametreleri bulabilmektir. Üçüncü ve önemli bir zorluk ise, sesin içindeki duygunun alınabileceği, pencere boyutu (window size) ve atlama boyutu'nun (hop size) belirlenmesidir.

Literatürde farklı yöntemleri kullanarak ses sinyallerinde duygu tanımanın, otomatik olarak tanınması üzerine birçok çalışma yer almaktadır. Bu çalışmalardan bazıları aşağıda incelenmiştir.

1.1 İlgili Çalışmalar

Literatürde ses sinyalinde duygu tanımlamak için farklı yöntemler olmakla birlikte, ağırlıklı yönelim olarak, makine öğrenme problemi olan sınıflandırma ile çözüme ulaşılmaya çalışılmaktadır [1][2][3][4][5].

Ses sinyalinden çıkarılabilecek öznitelik sayısı görece fazladır. Sesin akustik, spektral gibi özelliklerinin yanı sıra içeriğin de duygu tanımlama da rolü olduğuna dair çalışmalar yapılmaktadır. Tawari ve Trivedi [1] yaptıkları çalışmada, perde ve yoğunluk kontürü için kepstrum analizi kullanmıştır. İki farklı veri kümesi üzerinde içerik üzerine çalışılmıştır. Duygu tanımda, konuşmacının, metnin ve cinsiyetin önemli olduğunu göstermişlerdir. Ses sinyali için, Destek Vektör Makinesi'ni (DVM) ardışıl en küçük en iyileme algoritması ile eğitip %84 e varan doğruluk oranlarına ulaşmış, spektral ve akustik özniteliklerin duygu tanımda başarılı oldukları anlaşılmıştır. Çalışmalarında ses sinyali ve farklı öznitelikler için, farklı pencere boyutu ve atlama süreleri kullanmışlardır. Cinsiyet bilgisinin de içinde bulunduğu duygu tanıma yöntemlerinde daha yüksek başarı alındığı görülmüştür. Konuşmacının cinsiyetini algılayabilen sistemler tasarlayarak, cinsiyet bilgisinin de içinde bulunduğu duygu tanıma yöntemleri üzerinde çalışmalarına devam edeceklerdir.

Paleari ve Chellali [22] yaptıkları çoklu-model çalışmada, ses ve video sinyalinden gerçek zamanlı ve kişi bağımsız duygu tanımla yapmak için, en ideal özniteliği aramışlardır. Öznitelik seçmek için ses sinyali için perde, enerji, titreşim, doğrusal öngörü katsayıları ve Mel-frequency cepstral coefficients(MFCC) özniteliklerini

kullanmışlardır. Her bir duygu için farklı olmak üzere, öfke ve kızgınlık duygusunun en başarılı tanımlanan duygular olduklarını ve doğrusal öngörü katsayıları ile MFCC özniteliklerinin bu iki duygu için %90'lar seviyesinde doğruluk verdiklerini raporlamışlardır. Öğrenme yöntemi için kullanılan Yapay Sinir Ağları (YSA) için, genel sonuçlarda %60 oranında doğruluk yakalamışlardır. Kullanılan veri kümesinin çözünürlüğünün düşük olması ve aktör performanslarının da, başarıyı olumsuz etkilediğini yorumlamışlardır. Video çalışmalarında sesli ve sessiz çerçeve öğelerini, iki farklı sınıfa ayırıp, ayrı ayrı süreçlerden geçirip incelemek üstüne çalışacaklardır.

Yang ve Chen [24], müzik parçalarındaki duyguyu tanımlayabilmek için oylama yöntemi yerine derece skoru yöntemini kullanmışlardır. Tanınma doğruluğunu arttırmak için, nesnel ve öznel bir etiketleme yöntemi çalışılmıştır. Bu yöntemde, deneklere önce müzik parçalarının duyguları etiketlenmiştir. Eğitim kümesi bu şekilde oluşturulduktan sonra, duygusu bilinmeyen müzik parçalarının, derece skorları üstünden duygularına karar verilmiştir. Boyutsal gösterim için, derece skorları iki boyutlu duygu uzayına yerleştirilmiştir. En yüksek dereceye sahip olan şarkılar, maksimum değerlik ya da uyarılma olarak atanmış, diğerleri ise doğrusal olarak işaretlenmiştir. Sonuçların yüksek öznel başarısı, duygu tabanlı görselleştirmenin başarılı sonuçlar verdiğini göstermektedir.

Li, Liu ve Wang [23] çalışmalarında DVM öğrenme yönteminde kullanılacak parametre optimizasyonu için Sezgisel Arama- Destek Vektör Makineleri (SA-DVM), sezgisel arama algoritması önermiş ve parametre optimizasyonu için gerekli süreyi, ızgara arama yöntemine kıyasla azalttıklarını bildirmişlerdir. Ancak kullandıkları parametre sayısını c , γ , ϵ ile sınırlı tutmuşlardır.

Metallinou, Wollmer ve diğerleri [28] çalışmalarında, Saklı Markov Modeli (SMM) ve Çift Yönlü Uzun Kısa Dönem Hafıza (ÇYUKSH) yöntemlerini füzyonlayarak kullanmışlardır. Her konuşma aralığı için, duyguya özel SMM modelleri tarafından üretilen log olasılıkları toplanılmış, bunlarla N boyutlu, log olasılığının konuşma seviyesi öznitelikleri yaratılmıştır. Bu öznitelikler yüksek seviyeli ÇYUKSH'nin girdilerini oluşturmuş, ÇYUKSH eğitim verilerindeki bu verilere göre eğitilmiştir. Eğitim kümesinde değerlik, uyarılma ve kategorik (öfke, mutluluk...) duygu olarak, kendi kullanıcılarına etiketlenmiştir. Yapay sinir ağlarının, özellikle de ÇYUKSH ağlarının, konuşma seviyesinde öznitelikler kullanılarak eğitilmesinin, SMM

sınıflandırıcılarından daha yüksek performans verdiğini göstermişlerdir. Ancak performans kişiden kişiye değişiklik gösterebilmektedir. Buna karşın SMM ve SMM/ÇYUKSH melez çatıları, düşük performans göstermelerine karşın kişiden kişiye daha kararlı sınıflandırma sonuçları göstermiştir. Gelecekte, gerçek zamanlı veriler ile çalışma yapmayı planlamaktadırlar.

Lampropoulos ve Tsihrintzis [17], Çapsal Tabanlı Fonksiyonlar – Destek Vektör Makineleri (ÇTF-DVM) sınıflandırıcıları kullanmışlardır. İkili sınıflandırma yerine çoklu sınıfı tercih etmişler ve topluluk tekniğini kullanmışlardır. Öncelikle ikili DVM sınıflandırıcısı ile sistemi eğitmişlerdir. Daha sonra sınıflandırmayı yapabilmek için, ilk adımdaki sınıflandırma sonuçlarını dinamik olarak bir araya getirmişlerdir. MPEG-7 tanımlayıcılarının üst seviye algı özelliklerinde, özellik seçim metotlarında, sınıflandırıcıların topluluk olarak kullanılmasında, daha karışık yaklaşımlara göre karşılaştırılabilir performans sergiledikleri anlaşılmıştır. MPEG-7 tanımlayıcıları ile MFCC ve tonlama özelliklerinin yüksek başarımla sağladığı göstermişlerdir. Potansiyel olarak daha güçlü olabilecek yapay bağımsızlık sistemleri tabanlı sınıflandırıcılar ile çalışmalarına devam edeceklerdir.

Xie ve Guan [16], kurguladıkları sistem ile sesin spektral (MFCC gibi), ve prosodik (vurgu, perde, tonlama, duraksama, vb. gibi özellikler) özniteliklerini kullanmışlardır. Duygu tanıma için, öznitelik seviyesinde birleştirilmiş, Çekirdek Entropi Bileşen Analizi (ÇEBA) temelli yeni bir çözüm önermişlerdir. Kullanılan öznitelik vektörü sayısı görece fazla olduğu için, ÇEBA yöntemi ile orijinal uzayı dönüştürülmüş uzaya taşıyarak, orijinal özniteliklerin karakteristiklerini taşıyan alt veri kümesi oluşturmuşlardır. Bu sayede, boyutluluk sorunu denilen ve verilerin boyutları arttıkça veri uzayının hacminin üstel bir hızla artması sorununun önüne geçmişlerdir. Sistemi SMM ile eğitip sonuçlar incelendiğinde, sonuçlarda gürültü verisi olmasına karşın, yararlı verinin büyük oranda korunduğu ve tutarlı bir doğruluk oranının yakalandığı gösterilmiştir.

Soleymani ve arkadaşları [9], yaptıkları çalışmada, duygu tanıma için kullanılacak yeni bir veri kümesi oluşturmuşlardır. Seçilen film sahnelerinden oluşturulan video kliplerini 30 katılımcıya etiketlendirmişlerdir (değerlik, uyarılma, duygu grubu, baskınlık, tahmin edilebilirlik). Aynı zamanda 6 kamera ile katılımcıların 6 farklı pozisyondan görüntüsünü ve oda mikrofonu ve yaka mikrofonu

ile katılımcıların doğal tepkilerinin sesini kaydetmişlerdir. Ancak veri yetersiz olduğu için katılımcıların seslerini kullanmamışlardır. Göz takip yazılımı ile göz hareketlerini, gözbebeğinin genişliği gibi parametreleri ölçmüşlerdir. Fizyolojik sinyalleri(Elektrokardiyografi (EKG), Galvanik Deri Cevabı (GDC), nefes alma genliği, cilt sıcaklığı) ve Beyin Akım Grafiği (BAG)'yi ölçmüşlerdir. Bu ölçümlerin hepsini eş zamanlı yapmışlardır. Kelime tabanlı geri dönüşlerden sınıfları oluşturup, ölçülen değerlerden çıkarılan öznitelikler ile DVM sınıflandırıcı ve Çapsal Tabanlı Fonksiyonlar (ÇTF) çekirdek ile 3 kipe göre (fizyolojik sinyaller, BAG, göz takip verisi) sınıflandırmaları yapmışlardır. Katılımcıların yüz ifadelerinin işlenmesinde SMM algoritması kullanmışlardır. Duygu tanımlamada kullanılabilir, 5 kipi de barındıran ilk veri kümesini oluşturmuşlardır.

Sezgin, Günsel ve Kurt [15], oluşturulan yeni algısal öznitelik kümesi ile veri normalizasyonu yapmadan, duygusal farklılıkların istatistiksel karakteristiklerini öğrenme üstüne çalışmışlardır. Algısal model, ayırık duyguların referans duygusundan farkını çıkarmak üstüne kurulmuştur. Yöntemin avantajı olarak normalizasyon yapılmamıştır. Sinyal 42ms'lik çerçevelere bölünmüş ve dış ile orta kulağın frekans cevabı spektral katsayılarla uygulanmıştır. Veri kümesinin, hem eğitim hem de test için referans kümesinden duygusal farklılıklarını hesaplamışlardır. 1 saniyelik aralıklarla ses sinyalinde N algısal fark vektörü çıkarmışlardır. Çıkarılan vektörleri sınıflandırıcıya vererek, çoğunluk oylaması (majority voting) yöntemi ile sınıflandırma yapılmıştır. Karşılaştırmalar kendi öznitelikleri ile openEAR yazılımının benzer öznitelikleri arasında yapılmıştır. Kendi öznitelik tanımlama sonuçlarının, geleneksel özniteliklere göre ortalama %6 civarında daha fazla tanıma oranı ile daha iyi sonuç verdiğini gözlemlemişlerdir. Hesaplama karmaşıklığını azaltmak için, referans kümesinin boyutunu azaltmak üzere çalışmalarına devam edeceklerdir.

Yang [31] ve arkadaşları, regresyon yaklaşımı ile müzik verisindeki duygu içeriğini tanıma üzerine çalışmışlardır. Müzik verisindeki duygu tanımayı, her bir müzik örneği için, regresyon problemi olarak, değerlik ve uyarılma uzayında formülize etmişlerdir. Kategorik sınıflandırma olmadığı için, geleneksel kategorize etme yaklaşımlarındaki doğal belirsizliklerden de kurtulmuşlardır. Performansı arttırmak, değerlik ve uyarılma arasındaki ilişkiyi azaltmak için Temel Bileşen Analizi (TBA)

kullanmışlardır. Özniteliklerin değerleri cinsinden birbirine yakın olan örneklerin sınıflarının aynı ya da farklı olmasına bakarak öznitelik ağırlıklarını güncelleme fikrine dayanan RReliefF yöntemini kullanarak, önemli özniteliklerin seçimini yapmışlardır. DVR, ÇDR ve AdaBoost.RT regresyon algoritmaları ile sistemi eğitip, test ederek sonuçlara ulaşmışlardır. RReliefF yöntemi ile öznitelik seçimi yapmanın tanımlama oranını önemli şekilde arttırdığı sonucuna ulaşmışlardır. İleriki çalışmalarında, şarkının sözleri ve söylenmesi üzerine öznitelik çözümlenmeleri yapmayı ve büyük boyutlu veri kümeleri ile çalışmayı istemektedirler.

Jia ve arkadaşları [32], 2 boyutlu değerlik, uyarılma uzayına, memnuniyet (haz) boyutunu da ekleyerek 3 boyuta çıkararak bir çalışma yapmışlardır. Ayrıca, yüz tanıma sentez modeli olarak da tanımlanabilecek, Arttırılmış- Gaussian Karıştırılmış Modeli (Arttırılmış-GKM) kavramını önermektedirler. Önerdikleri model, duygusal ses ve görüntü sentezi yaklaşımına dayanmaktadır. Çeşitli metinlerin duygu durumunu işaretledikten sonra, metinden-konuşmaya çevrim yapabilen yazılımlar ile, metni doğal konuşmalara çevirmişlerdir. Arttırılmış-GKM ile doğal konuşmaları, duygusal konuşma ve 3 boyutlu yüz ifadelerine çevirmişlerdir. Duygusal konuşmanın akustik özniteliklerini (enerji, süre, perde gibi) kullanarak, ses ve görüntü sentezindeki yüz ifadelerini yumuşatmışlardır. 3 nesnel, 5 öznel deney yaparak her bir modelin ve toplam yaklaşımın performansını ölçmüşlerdir. Sonuçlar önerilen yaklaşımın, etkin ve verimli bir şekilde doğal ve anlamlı ses ve video sentezi yapabildiğini göstermiştir.

Sesten duygu tanıma çalışmalarının birçoğunda, insan işitsel sistemini ifade edebilen MFCC ve MPEG-7 gibi spektral öznitelikler kullanılmaktadır [1][4][13][17][22]. Ancak, öznitelik çıkarımı ve temsilinde standart yöntemler uygulanmaktadır ve öznitelik çıkarımında analiz edilecek pencere boyutu ve atlama süresi gibi parametreler özniteliklerin gürbüzlüğünde etkili olmaktadır. Benzer olarak, sınıflandırıcı tasarımında, öğrenici parametrelerin en iyilenmesi de başarımda önemli bir rol oynamaktadır [55][56].

Ses sinyallerinde duygu geri erişimi konusunda, literatür taraması sonucu yapılmış bir çalışmaya rastlanılmamıştır. Konuya benzer olabilecek çalışmalar müzik duygu geri erişimi ve ses sinyallerinde bilgi geri erişimi olarak özetlenebilir [64]-[69]. Zhou ve arkadaşları [63], müzik bilgi geri erişiminde duygu tanımayı kullanmışlardır.

Çalışmalarında, müzik içerisinde saklı bulunan duygu durumunu, makine öğrenme duygu sınıflandırma algoritmaları ile tanımlayabilmeye ve farklı öznitelik kümeleri ile performans karşılaştırmaları yapmaya çalışmışlardır. Sınıflandırma sonuçları, klasik yöntemlere göre başarılı gözlenmiştir. Doğan ve arkadaşları [65], karışık ses sinyallerinden, esnek ve ölçeklenebilir ses bilgi geri erişimi üzerine çalışmışlardır. Çalışmaları, kullanıcıya ses sinyali üzerinde 4 farklı şekilde (karışık tipli ses sınıflandırma sorgusu, alan temelli bulanık sorgulama, zamansal bilgi ve yakınlık sorgulaması, örnek ile sorgulama) anlamsal sorgulama yapabilme imkanı vermektedir. Geri erişimin süresini azaltmak için, özet-tabanlı indeksleme tekniği önermişlerdir. Yaptıkları iki deneysel çalışmanın ilkinde, ASF özniteliğinin, müzik ses örneklerinde, diğer ses örneklerine göre daha başarılı sonuç verdiğini gözlemlemişlerdir. İkinci deneysel çalışmada ise, tasarlanan geri erişim sisteminin çeşitli koşullar altında başarılı olduğunu gözlemlemişlerdir. Sert ve Baykal [66], ses verisinde, içerik tabanlı ve anlamsal sorguları web tabanlı bir uygulama geliştirerek uygulamışlardır. Geliştirdikleri uygulama ile, örnek ile sorgulama ve metin tabanlı sorgulama kullanmışlardır. Örnek ile sorgulamada MPEG-7 tanımlayıcı ailesinden öznitelikler kullanmışlar, metin tabanlı anlamsal sorgular için zamansal ve konseptel ilişkileri nesne ve olay konseptinde kullanmışlardır. Kullanıcı sorgulamalarında, kullanıcıya daha fazla adapte olabilecek bir arayüz tanıtmışlardır. Ses verisinden anlamsal bilgi geri erişimi ve düşük seviye öznitelik çıkarımı için kullanılacak yeni bir etketleme aracı tanıtmışlardır. Yoo ve arkadaşları [69], müzik verisinde, duygu tanıma çalışmaları yapmışlardır. Çalışmaları kapsamında, 55 farklı özniteliği, öznitelik seçim yöntemleri ile 27 özniteliğe indirerek kullanmışlardır. Öznitelik vektörlerini, müzik verisindeki duyguyu, 2 boyutlu değerlik, uyarılma uzayı üzerinde etiketlemek için kullanmışlardır. 2 boyutlu duygu uzayından duyguları tanımlamak için, renkler (açık, koyu, siyah, mor, kırmızı...) ve duygu yoğunluğuna bağlı olarak çeşitli eğriler kullanmışlar, duyguların tanımını ve şiddetini ifade eden grafiksel simgelere dönüştürmüşlerdir.

Bu tez çalışmasında, öncelikle ses sinyalinde 7 adet ayırık duygu (Kızgınlık, Korku, Sıkıntı, İğrenme, Mutluluk, Üzüntü, Nötr) durumunu tanımlayabilen, bir uygulama çatısı yaratılmaya çalışılmış, tasarlanan deneysel çalışma ile de öğrenme yönteminin başarımını arttıracak etkin öznitelik bulma, parametre optimizasyonu, sınıflandırıcı tasarımının başarıma etkisi incelenmiştir. Sonuçların uygulamasını

görmek amacı ile Matlab uygulamasında örnek bir uygulama geliştirilmiştir. Örüntü tanıma sistemlerindeki görece yüksek başarımlar ve yaygın kullanımından dolayı öğrenme yöntemi olarak DVM kullanılmıştır. 7 farklı duygu durumunu tanımak için 19 tekil ve bileşik ses özniteliği incelenmiş ve sonuçları değerlendirilmiştir.

Çalışmanın ikinci bölümünde, duygu geri erişimi için örnek bir veri kümesi yaratılmıştır. Yaratılan veri kümesi üzerinde, esnek sorgulamalar (nokta, aralık, en yakın komşuluk) ile duygu geri erişimi incelenmiş ve değerlendirilmiştir.

1.2 Tezin Organizasyonu

Bölüm 2'de duygu ve ses sinyali hakkında temel tanım ve kavramlar verilmiş, duygu tanımının tarihi ve gelişme süreci anlatılmıştır. Ayrıca, ses sinyalinde duygu tanıma amacıyla kullanılan yöntemler incelenmiştir. Makine öğrenme ve algoritmaları, öğrenme çeşitleri, duygu tanımadaki ana problem ve tez kapsamındaki deneysel çalışmalarda kullanılan DVM öğrenme algoritması, özellikleri ve yöntemleri anlatılmıştır. Bu tez kapsamında önerilen duygu tanıma, duygu geri erişimi yöntemleri ve kullanılan araçlar 3. ve 4. Bölüm'de ayrıntılı olarak sunulmuştur. 5. Bölümde deneysel çalışmaların sonuçları ve performans değerlendirmeleri yapılmıştır. Son bölümde ise sonuçlar ve gelecek planları aktarılmıştır.

2. TEMEL TANIM, KAVRAM VE YÖNTEMLER

2.1 Duygu ve Ses Sinyali

2.1.1 Duygu

Duygunun kısa bir tanımını yapmak zordur. Birçok kişi farklı tanımlamalarda bulunmuştur. Cevizci, “Duyduğumuz, duyumsadığımız her şey; özellikle tüm tutkularımızın, hafif veya ortalama şiddetteki heyecanlarımızın, aşk, sevgi gibi genel hallerimizin, genel ve içgüdüsel eğilimlerimizin genel adıdır” şeklinde tanımlamıştır [34]. Akarsu, “Anlık ve istencin yanında, duyma yetisi, haz ve acı duyabilme olarak beliren üçüncü temel yetiye verilen ad” olarak tanımlamıştır [35]. Goleman ise duyguyu, “kişinin öğrenme potansiyelini uyararak öğrenmesini sağlayan, soru sormaya yönlendirerek merak uyandıran, bilinmeyeni aramaya yönlendiren, kapasitesini geliştirip öğrendiklerini pratiğe dökmesini ve böylece davranış sergilemesini ifade eden özellikler bütünü” olarak tanımlamaktadır [36]. Duygular davranışlarımızı motive eder ve diğer insanlarla ve çevremizle etkileşim kurma şeklimizi yönlendirir.

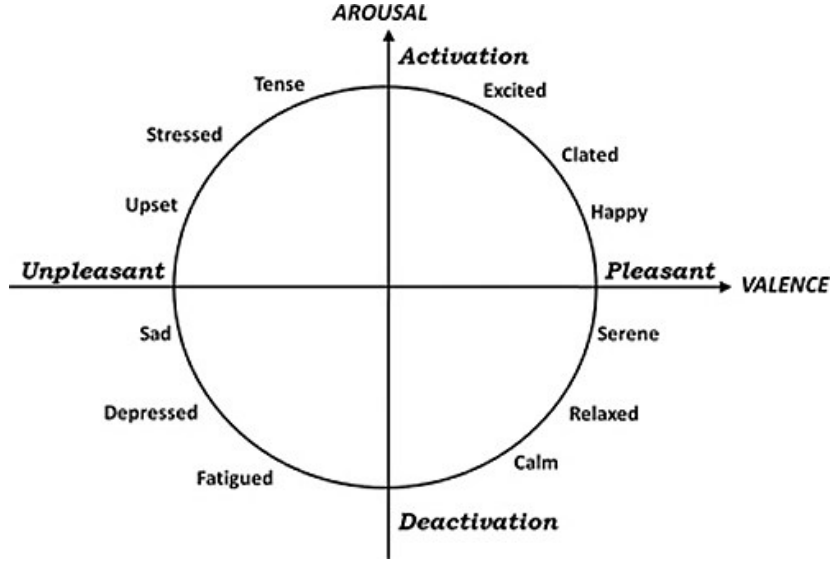
2.1.2 Duygu sınıflandırma

Charles Darwin (1872), duyguların biyolojik olarak belirlendiğini ve tüm insanlık kültürü için genel olduğunu savunmuştur [37]. Ancak Ekman’dan önce, bilim dünyası çoğunlukla, yüz ifadelerinin ve duyguların kültür tarafından belirlendiğini savunmaktaydılar. Ekman (1974), dış kirlenmeye çok az maruz kalmış izole bir kültüre sahip olan Papua Yeni Gine yerlileri ile çalışmış ve tüm insan halklarında, her bireyde aynı olan temel duygu ifadelerini 6 ayrı ayrı kategoride sınıflandırmıştır [38]. Ekman, 1999 yılında duygu durumlarını, sadece yüz mimiklerinden çıkarılamayan olumlu ve olumsuz duyguları da ekleyerek güncellemiştir [39]. Çizelge 2-1 de Ekman tarafından tanımlanan duygu durumları incelenebilir.

Çizelge 2.1 Ekman duygu sınıflandırması.

Temel duygular (1974)	Temel duygu adayları (1999)
Öfke	Eğlenme
İğrenme	Aşağılama (hor görme)
Korku	Hoşnutluk
Mutluluk	Mahcubiyet
Üzüntü	Heyecanlanma
Şaşkınlık	Suçluluk duygusu
	Gurur
	Rahatlama
	Memnuniyet
	Zevk
	Utanç

Plutchik (1980), duygunun yeni tarifini sundu [40]. Bu tarifi duygular tekerleği olarak adlandırdı. Tarife göre, farklı duygular birbiri içinde eritilerek yeni duygular çıkmaktadır. Tarifte duygular, yoğunluk ve şiddet açısından kendi içlerinde derecelendirilirler; coşkunluk, neşe, huzur gibi. Komşu olan duygular birbirlerine, diğerlerine göre daha çok benzemektedirler; coşkunluk, hayranlık gibi. Karşılıklı yer alan duygular, birbirlerinin tersi duygulardır; nefret, hayranlık gibi. Komşu duygular birleşerek geniş çapta yeni duygular elde edilir; korku + güven = itaat gibi. Şekil 2-1 de Plutchik'in duygular tekerleği incelenebilir.



Şekil 2.2 Russel'in Circumplex Modeli

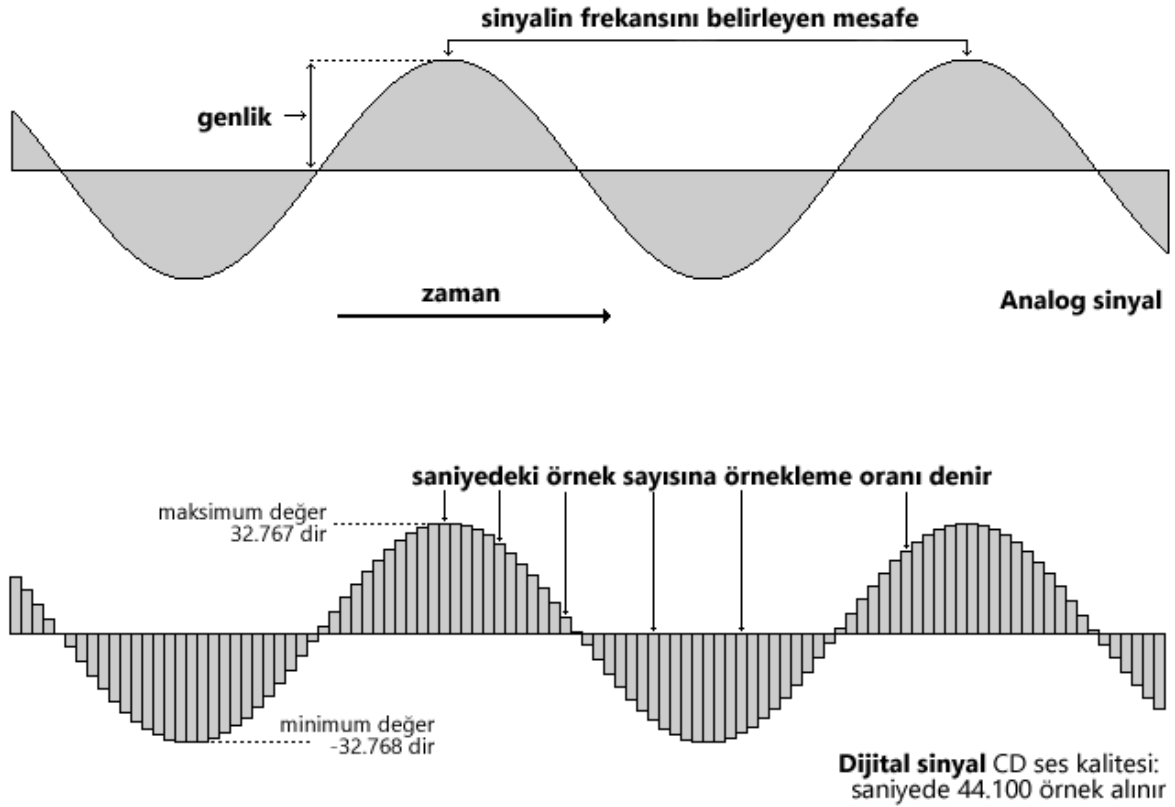
Tez çalışması kapsamında, 7 ayrıık duygu olan, mutluluk, üzüntü, kızgınlık, korku, sıkıntı, iğrenme ve nötr duyguları alınmıştır.

2.1.3 Ses sinyali

Ses, havadaki titreşimin insan kulağında yarattığı fiziksel baskının, elektrik sinyallerine çevrilerek, sinir hücrelerine aktarılması sonucu anlamlandırılır. Bilgisayar sistemlerine aktarılması ve ses sinyali haline gelebilmesi için, mikrofon gibi çeviriciler vasıtası ile kulakta olduğu gibi, titreşimin sinyallere çevrimi şeklinde oluşur. Ses sinyali sesin elektriksel formudur. İnsan kulağı tarafından doğrudan anlamlandırılmaz. Analog ve dijital olmak üzere iki farklı şekildedir. Analog ses sinyali, sesin birebir kopyalanmış biçimidir. Dijital ses sinyali ise analog ses sinyalinde birim zamanda, belli örnekler alınması ve örneklerin sadece 1 ve 0'lardan oluşan dijital sisteme dönüştürülmesiyle elde edilmiş sayısal biçimidir. Sesin bir frekansı, boyu, periyodu ve hızı bulunmaktadır.

Frekans: Sesin birim zamandaki titreşim sayısına frekans denir. Birimi ise Hertz (Hz)'dir. Normal insan kulağının duyabileceği ses sinyalin en alt ve en üst frekans sınırı 20 Hz ile 20kHz arasında rapor edilir [57].

Periyot: Frekansın tersi olarak bir tam titreşim için geçen zaman denilir.



Şekil 2.3 Analog ve dijital ses sinyali

2.2 Makine Öğrenme

1940lı yıllarda, bilim insanlarının, nöronların elektriksel çalışmalarına dair yaptıkları keşiflerle, insan karar verme mekanizmasının çok basit bir şekilde topla ve ateşle nöron modeli ile açıklanabildiği gösterilmiştir. Bu gelişme ile birlikte yapay zeka çalışmaları 1950li yıllarda başlamıştır.

İnsan için öğrenme, hayatı boyunca kazandığı tecrübelerden çıkarım yapabilme, davranışlarını bu çıkarımlara göre değiştirebilme yetisidir. Benzer olarak makine öğrenme için de, makinenin performansı, yaptığı işi tecrübe ettikçe artıyorsa, makine öğrenmesi gerçekleşmiştir denilebilir. Makine öğrenimi kendisine verilen örnekleri ve varsa sonuçları inceler ve bu işlemi nasıl yeniden yapacağını öğrenir ve yeni durumlar hakkında genellemeler yapar. Makine öğreniminin, var olan bilgileri kullanarak yeni bilgiler çıkarma ve bu bilgileri makinelerin davranışlarına aktararak, yeni bilgileri işlemede ve sonuç çıkarmada tahminlerde bulunmaya dayandığını

söyleyebiliriz. Makine öğrenimi veri madenciliği, yapay zeka, örüntü tanıma, istatistik gibi bir çok çalışma alanı ile bağlantılı ve iç içe geçmiş durumdadır.

Makine öğrenme yöntemlerini aşağıdaki gibi maddeleyebiliriz:

- Sınıflandırma: Sınıflama belli özniteliklere bakarak veriyi önceden belli olan sınıflardan birisine vermektir.
- Kümeleme: Geçmiş bilgilerinin sınıfları bilinmediği veya verilmediği durumlarda verilerden birbirine benzerlerin yer aldığı kümelerin bulunması işlemidir.
- Regresyon: Verinin bazı özelliklerini kullanarak diğer özelliklerini tahmin etmek ya da veriyi kullanarak belli sonuçlar çıkarmak işlemidir.
- Öznitelik seçimi/çıkarımı: Bazı durumlarda veriye ait birçok özellik bilirse de verinin kümesini, sınıfını, değerini belirleyen özniteliklerin hangileri olduğu bilinmeyebilir. Bu durumlarda tüm öznitelik kümesinin bir alt kümesi seçilmesi (öznitelik seçimi) ya da bu özniteliklerin birleşimlerinden yeni özelliklerin elde edilmesi işlemidir.
- İlişki belirleme: Veri yığınları arasındaki ilişkileri ya da ilişki bağıntılarını belirleme işlemidir. Hangi ürünlerin çoğunlukla birlikte satıldığı, bir ürünün alıcısının başka hangi ürünü satın alabileceği gibi sorulara ilişki belirleme yöntemi ile cevap bulunabilir.

Literatürde duygu tanıma alanında sıklıkla kullanılan makine öğrenme algoritmaları şu şekildedir [1][4][6][8][9].

- Destek Vektör Makineleri
- Yapay Sinir Ağları
- Karar Ağaçları ve Karar Ormanları
- Bayes Öğrenme ve Bayes Ağları
- Kümeleme
- Gizli Markov Modelleri
- Genetik Algoritmalar

2.2.1 Öğrenme çeşitleri

Öğrenme, sistemin bilenen veya bilinmeyen veriye göre eğitilmesine ya da kendi kendine veri içinden sonuç çıkarmasına göre güdümlü ve güdümsüz olmak üzere ikiye ayrılır.

2.2.1.1 Güdümlü öğrenme

Güdümlü öğrenme, önceden sonuçları (sınıfları) bilinen verileri kullanarak bu verileri ve sonuçlarını kapsayan bir model oluşturmayı amaçlayan makine öğrenimi metodudur.

Önceden bilinen eğitim verileri, etiketleri (sınıfları) ile birlikte sisteme verilerek, makinenin kendi başına öğrenmesi, öğrenme sonunda ortaya çıkan model sayesinde gelecekteki bilinmeyen veriler için de doğru sonuçları veren işlemler yapabilmesi üzerine kurulmuştur. Öğrenmeyi gerçeklemek için makine, olması gereken çıktıya en yakın sonucu elde edecek optimum fonksiyonu bulmayı amaçlar.

Eğitim aşamasından sonra, makinenin ürettiği model test verileri ile teste tabi tutulur. Test sonuçlarından, doğru olarak tanımlanmış sonuçların, toplam test sayısına bölünmesiyle tanıma ya da doğruluk oranı çıkarılır. Doğruluk, ne kadar yüksek ise oluşturulan model o derece başarılıdır.

Güdümlü öğrenmede kural, ezberlemek değil, öğrenmedir. Sınıflandırma, güdümlü öğrenme yöntemidir. Güdümsüz öğrenmeye göre, öğrenme süresi daha hızlıdır. Sebebi ise hatanın, sisteme öğrenme olarak giriyor olmasıdır. Sistem, yaptığı hatalardan ders çıkarır ve bu dersleri sistemi öğrenmek için kullanır.

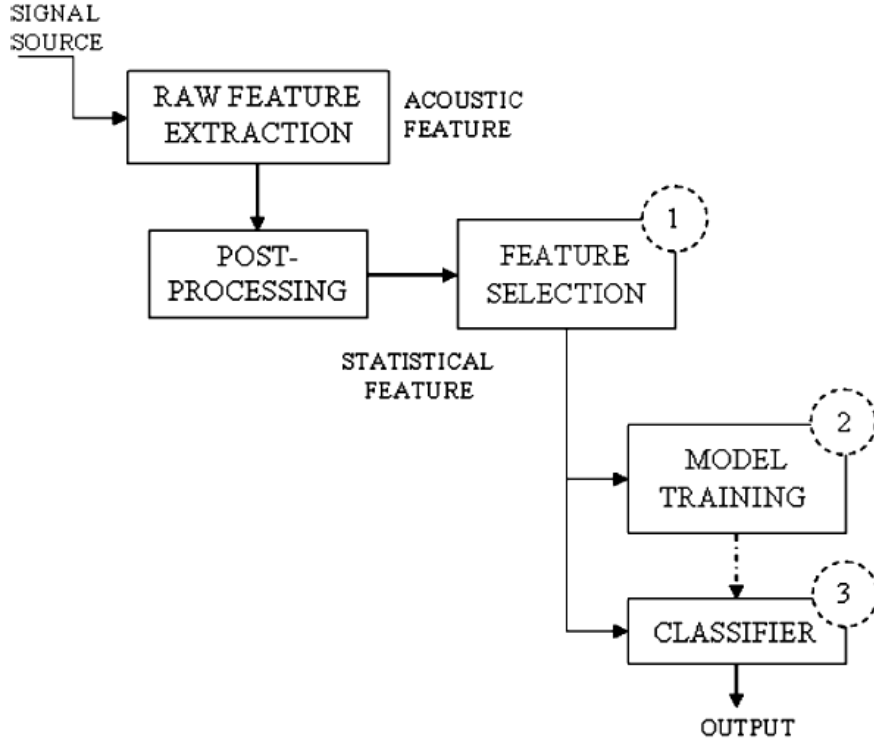
2.2.1.2 Güdümsüz öğrenme

Güdümsüz öğrenmede, sisteme hedef değeri ve etiketi bilinmeyen veriler verilir. Etiketlenmemiş verideki, varsa bilinmeyen yapıyı bulma işlemidir. Veriler arasında var olan ama gözle görülmeyen bağıntının açığa çıkarılması işlemidir. Veriler arasında böyle bir ilişki olmayabilir. Dolayısı ile sonuçların kesin doğru olması beklentisi yoktur.

Verileri benzer deęerler etrafında toplamayı amaçlayan kümeleme yöntemleri güdümsüz öğrenme yöntemidir.

2.3 Duygu Tanıma Problemi

Literatürde ses sinyalinde duygu tanımak için farklı yöntemler kullanılmakla birlikte, ağırlıklı yönelim olarak, makine öğrenme problemi olan sınıflandırma ile çözüme ulaşılmaya çalışılmaktadır [1][2][3][4][5]. Problem makine öğrenme problemi olarak ele alındığında izlenen genel akış şöyledir. İçerdiği duygular bilinen ses sinyalleri, eğitim ve test verisi olarak ikiye ayrılır. Eğitim verisi olarak adlandırılan eğitim kümesinden, konuşmacıya özgü bilgiler çeşitli öznitelik çıkarım yöntemleri ile çıkarılıp, içerdiği duygular etiketlenir. Öğrenme algoritmasına verilen eğitim verileri ile eğitilen sistemden, eğitim kümesi için örnek bir model alınır. Eğitim kümesi gibi içindeki duygular etiketlenen test kümesi üzerinde, aynı model çalıştırılarak, eğitim kümesi üzerinde oluşturulan modelin başarımı çıkarılır. Başarım yeterli bulunmazsa, öznitelikler ya da öğrenme algoritmaları değiştirilebilir. Yeterli bulunan başarım seviyesinde, duygusu bilinmeyen ses sinyalleri sisteme verilerek, eğitilmiş modelden, duygunun doğru olarak tanımlanması beklenir. Tanımlamanın doğruluğunu etkileyen birçok etken mevcuttur. Eğitim ve test kümesinin, tanımlanmak istenen veri ile benzerlikler göstermesi, problem uzayını temsil edebilmesi önemlidir. Aksi takdirde, model daha önce deneyimlemediği ses sinyalini doğru tanımlayamaz. Bu durumda, bilinmeyen veriden, eğitim kümesine eklemeler yapılarak, eğitim kümesi ve model geliştirilmelidir. Eğitim kümesinin yeterince veri içermemesi durumu da aynı sorunu doğuracaktır. Bir diğer sorun ise, boyutluluk sorunu olarak da adlandırılan, özniteliklerden vektörlerinin sayısının artması ile verilerin boyutları arttıkça veri uzayının hacminin üstel bir hızla artması sorunu ortaya çıkmasıdır. Boyutluluk sorunu yaşanması durumunda, orijinal uzayı dönüştürülmüş uzaya taşıyarak, orijinal özniteliklerin karakteristiklerini taşıyan alt veri kümesi oluşturulabilir [16]. Bu şekilde veri kümesi azaltılabilir ve hem veri içerisindeki gürültü denilen, alakalı olmayan veri ayrıştırılabilir hem de öğrenme algoritmasının işlem süresi kısaltılabilir. Şekil 2-4'de örnek bir duygu tanıma sisteminin blok şeması verilmiştir [1].



Şekil 2.4 Örnek bir duygu tanıma sistemi

2.4 Destek Vektör Makineleri

DVM; değişkenler arasındaki bağlantıların bilinmediği veri kümelerinde sınıflama ve regresyon için kullanılan bir öğrenme algoritmasıdır. Destek vektörleri yöntemi 1960'ların sonunda Vapnik tarafından geliştirilmiş ve 1995 yılında sınıflandırma konusunda Vapnik tarafından önerilmiştir [21]. DVM sınıflaması, güdümlü öğrenme yöntemidir. Eğitim veri kümesindeki bilgilerden model çıkarılarak, sınıfı bilinmeyen yeni verilerin sınıfını bulmaya çalışır. DVM istatistiksel öğrenme teorisi üzerine kurulmuştur. İki sınıflı veri sınıflandırma ve uyumlama metodu olup son yıllarda literatürde sıklıkla kullanılır olmuştur [1-5][9][12][24].

DVM, verilerin doğrusal olarak ayrılabilmediği durumda verileri ayırabilecek sonsuz sayıda doğru içerisinden aradaki mesafeyi en yüksek yapacak olan doğruyu seçmeyi hedeflemektedir. Veri doğrusal olarak ayrılamadığında doğrusal olmayan örnek uzayını, örneklerin doğrusal olarak ayrılabilmediği bir yüksek boyuta haritalayarak, farklı örnekler arasındaki maksimum sınırın bulunması esasına dayanır. Algoritma yeni örnekleri en uygun yüksek düzlem ile sınıflandırmaktadır.

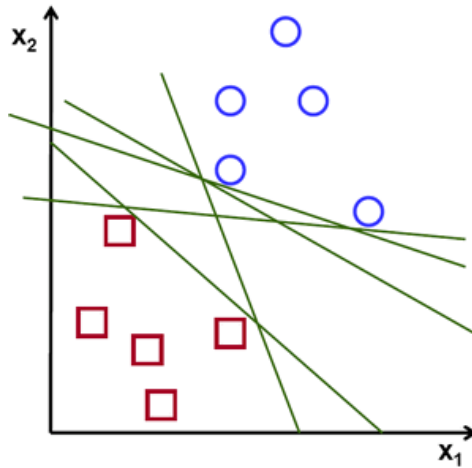
DVM, doğrusal sınıflayıcı olarak iki sınıflı problemler için düşünülmüş, daha sonra doğrusal olmayan sınıflayıcı olarak da kullanılmıştır. Çeşitli yöntemler ile de ikiden çok sınıf içeren problemlerin çözümünde de kullanılmaya başlanılmıştır [42]. DVM'i, doğrusal olarak ayrılabilme durumuna göre, doğrusal olarak ayrılamama durumuna göre, doğrusal olmaması ve çok sınıflı sınıflandırma altında dört ana önemli başlık altında inceleyebiliriz.

2.4.1 Doğrusal olarak ayrılabilme durumuna göre sınıflandırma

Eğitim verimizin formunun

$$(x_1, y_1), \dots, \dots, (x_m, y_m) \in R^n \times \{-1, 1\} \quad (2.1)$$

olduğunu kabul edelim. Bu formda, x_i , n boyutlu özellikler vektörleri, y_i de etiketleri olsun. $\{-1, +1\}$ kümesi gözlemlerin çıktı etiketlerini temsil etmektedir. Girdide +1 ile etiketlenmiş olan vektör, +1 sınıfına dahil olduğunu -1 ile etiketlenmiş olan vektör, -1 sınıfına dahil olduğunu gösterir. Verimizin doğrusal olarak ayrılabilme durumu için amacımız, veri kümesini verilen etiketlere göre bir düzlemlerle ayırıp, aynı sınıfa ait bütün veri noktalarını düzlemin aynı tarafında bırakmaktır. Şekil 2.5 de doğrusal olarak ayrılabilen veriler için, çeşitli noktalardan geçen düzlemler görülmektedir.



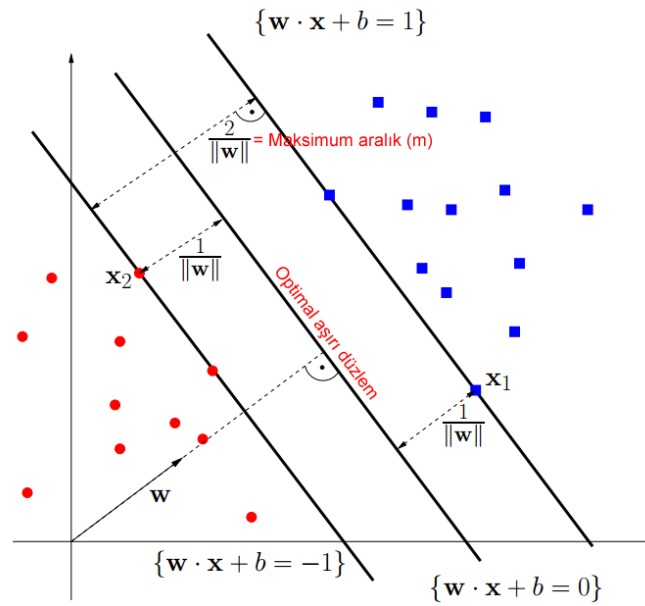
Şekil 2.5 Doğrusal olarak ayrılabilen veriler, aşırı düzlemler

Veriler iki sınıfa belirgin şekilde ayrılmış olmasına karşın, sınıfı ayıran en ideal düzlemi bulmak gerekmektedir. DVM, iki sınıf arasındaki aşırı düzlemi seçmek için birbirinden en uzak olan iki aşırı düzlemi seçer.

Doğrusal olarak ayrılabilir durumu için, aşırı düzlem (hyper plane) şu şekilde tanımlanabilir.

$$f(x) = w^T \cdot x + b = 0 \quad (2.2)$$

Burada, w aşırı düzlemin normalini ve ağırlık vektörünü ifade eder. b doğrusal değeri ise sapma değeridir ve aşırı düzlemin konumunu belirler. x ise $w^T \cdot x + b = 0$ aşırı düzlemi üzerinde olan herhangi bir noktadır. DVM'nin amacı bu aşırı düzlemin iki gruba da eşit uzaklıkta olmasını sağlamaktır.



Şekil 2.6 Optimal aşırı düzlem

Bu durumda, aralığı en büyük tutacak iki düzlemin fonksiyonları

$$y_1 = w \cdot x_1 + b \quad (2.4)$$

ve

$$y_2 = w \cdot x_2 + b \quad (2.5)$$

olacaktır. Eşitlik (2.4) ve (2.5), y_i değeri için $\{-1, +1\}$ sonuç sınıf çıktı kümesine eşit uzaklıkta seçilmelidir. Dolayısıyla aşırı düzlemin üstünde kalan verilerin +1 sınıfına, altında kalanların ise -1 sınıfına dahil olduğunu kabul eder isek, bu durumda denklemler

$$w \cdot x_1 + b = 1 \quad (2.6)$$

ve

$$w \cdot x_2 + b = -1 \quad (2.7)$$

şeklinde tanımlanabilir. (2.6) yı (2.7) den çıkarır isek,

$$w \cdot x_1 + b - w \cdot x_2 - b = 2 \quad \text{ve}$$

$w(x_1 - x_2) = 2$ eşitliği elde edilir. Eşitliği w nin öklid formu olan $\|w\|$ ye böler isek,

$\frac{w(x_1 - x_2)}{\|w\|} = \frac{2}{\|w\|}$ elde ederiz. $\frac{w}{\|w\|}$ uzunluğu 1 olan birim vektör olduğu için denklemin son hali olan ve maksimum aralık (margin) olarak adlandırılan mesafe (m)

$$m = \frac{2}{\|w\|} \quad (2.8)$$

bulmuş olur. DVM in, aralığı maksimum yapma davranışını bildiğimiz için eşitlik (2.8)'in maksimum değerini bulmamız gereklidir. Maksimum aralık değerine sahip olan bu düzlemler üzerinde yer olan veri noktalarına “Destek Vektörleri” denilir. (2.8)'in maksimum olması için $\|w\|$ değerinin minimum olması gereken yeni bir problem ile karşılaşırız. Öncelikle $\|w\|$ nin öklid form dönüşümünü yaparsak

$$\frac{2}{\|w\|} = \frac{2}{w^T w} \quad \text{olur,} \quad \frac{2}{w^T w} \quad \text{nin}$$

maksimum olması, pay ve paydayı yer değiştirdiğimizde minimum olması gereken

$$\frac{1}{2} w^T w \quad \text{ifadesine dönüşür.}$$

Maksimum aralığın bulunması için

$$\text{Minimizasyon} \quad \frac{1}{2} w^T w \quad (2.9)$$

$$\text{Kısıtlama} \quad y_i(w^T \times x_i + b) \geq +1, \quad i = 1, 2, \dots, n \quad (2.10)$$

Eşitliğe uygun olarak minimumlaştıran w ve w_0 hesaplanır. Eşitlik (2.9) çözülecek problem, (2.10) problem çözümünde kullanılacak kısıttır(koşuldur). Bu aşamada çözüm için ikinci derece, optimizasyon tekniklerine geçmemiz gereklidir. Çözüm için Lagrange yöntemi kullanılır. Lagrange yöntemini kullanmanın yararları ise, problemin Lagrange çarpanlarının hesaplanmasının daha kolay ve problemin doğrusal olarak ayrılmasını içinde genelleştirilmesinin daha uygun olmasıdır. Problemin Lagrange formülü,

$$L_p(w, b, \alpha) = \frac{1}{2} w^T w - \sum_{i=1}^l \alpha_i (y_i(w^T \times x_i + b) - 1) \quad (2.11)$$

$$L_p(w, b, \alpha) = \frac{1}{2} w^T w - \sum_{i=1}^l \alpha_i y_i w^T \times x_i - \sum_{i=1}^l \alpha_i y_i b + \sum_{i=1}^l \alpha_i \quad (2.12)$$

(2.12) de ki , $\alpha = \alpha_1, \alpha_2, \dots, \alpha_n$ değerleri pozitif Lagrange çarpanları olarak adlandırılır. Bu denklemin bu çarpanlara göre kısmi türevlerinin alınıp sıfıra eşitlenmesi sonucunda elde edilirler. (2.12) nin çözümünün bulunması için Karush-Kuhn-Tucker (KKT) koşulları kullanılırsa problem dual probleme dönüştürülür. KKT koşulları ise şu şekilde tanımlanır.

$$\frac{\partial L_p}{\partial w} = 0, \quad w = \sum_{i=1}^l \alpha_i y_i x_i \quad (2.13)$$

$$\frac{\partial L_p}{\partial b} = 0, \quad w = \sum_{i=1}^l \alpha_i y_i \quad (2.14)$$

Koşullar (2.12) de ki Lagrange formülü konulurlarsa, sonuç olarak oluşan dual problem şu şekilde olur.

$$L_D = \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i,j=1}^l \alpha_i \alpha_j y_i y_j x_i^T \cdot x_j \quad (2.15)$$

$$\sum_{i=1}^l \alpha_i y_i = 0 \quad (2.16)$$

$$\alpha_i \geq 0 \quad i = 1, 2, \dots, l \quad (2.17)$$

Eşitlik (2.15) probleminin, (2.16) kısıtları(koşulları) altında çözümü, ikinci dereceden optimizasyon problemidir.

(2.15) in çözümü ile bulunacak Lagrange çarpanlarının çoğunluğu 0 olacaktır. Geriye kalan $\alpha_i > 0$ değerli örnekler destek vektörleridir ve iki aşırı düzlem üzerinde

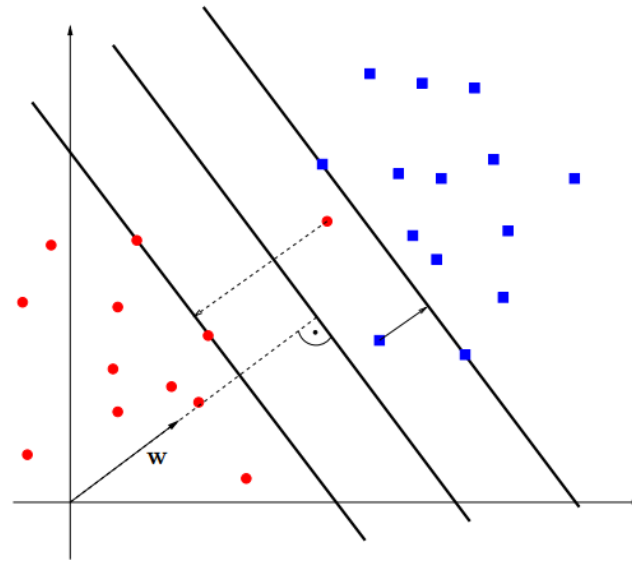
yer alırlar. $\alpha_i = 0$ olan değerlerse, iki aşırı düzelmelerin arkalarında kalırlar [43]. b değeri ise;

$$b = \frac{1}{n_{sv}} \left(\sum_{s=1}^{n_{sv}} \left(\frac{1}{y_s} - x_s^T w \right) \right) \quad s = 1, 2, \dots, n_{sv} \quad (2.18)$$

(2.18) de sv destek vektörleri, n_{sv} ise destek vektörlerinin sayısını ifade etmektedir.

2.4.2 Doğrusal olarak ayrılamama durumuna göre sınıflandırma

Bazı durumlarda, örneğin verilerde gürültü bulunması durumunda, veriler tam olarak doğrusal ayrılamayabilir. Şekil 2-7 de olduğu gibi, veri noktalarının kısıtlı bir kısmı iki aşırı düzlemin içerisinde kalabilir. Önceki alt bölümde anlatılan, ayırıcı fonksiyon tarafından ayırım düzgün yapılamaz. Sınıf verileri olabildiğince, aşırı düzlemin uygun tarafında bırakılmaya çalışılır. Bu durumda sınıf ayrımları, minimum hata ile yine doğrusal olarak yapılabilir. Kısıtlamaları zayıflatıp, sisteme esneklik kazandırmak gereklidir ve pozitif zayıflık (ya da gevşeklik) değişkenleri ξ kullanılmalıdır [45].



Şekil 2.7 Verilerin doğrusal olarak ayrılamaması

$$\text{Minimizasyon} \quad \frac{1}{2} w^T w + C \sum_1^n \xi_i \quad (2.19)$$

$$\text{Kısıtlamalar} \quad y_i = +1 \Rightarrow w^T \times x_i + b \geq +1 - \xi_i, \quad \xi_i \geq 0 \quad (2.20)$$

$$y_i = -1 \Rightarrow w^T \times x_i + b \leq -1 + \xi_i, \quad \xi_i \geq 0 \quad (2.21)$$

(2.19) ve (2.20) için $\xi_i=0$ olması durumunda, örnek doğru sınıflandırılmıştır. $\xi_i \geq 1$ ise yanlış sınıflandırılmıştır. 0 ve 1 arasında olma durumunda ise doğru sınıflandırılmış olmasına karşın, aşırı düzlemlerin arasında kalmıştır. C katsayısı, Lagrange çarpanının alabileceği üst sınır değerini gösteren ceza parametresini ifade etmekte aralık içinse düzeltme faktörü olarak değerlendirilmektedir. Lagrange formülü ise;

$$L_p = \frac{1}{2} w^T w + C \sum_{i=1}^l \xi_i - \sum_{i=1}^l \alpha_i (y_i (w^T \times x_i + b) - 1 + \xi_i) - \sum_{i=1}^l \mu_i \xi_i \quad (2.22)$$

formülüne dönüşür. Karush-Kuhn-Tucker (KKT) koşulları kullanılarak daha önceki bölümdeki gibi problem dual probleme dönüştürülür.

$$L_D = \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i,j=1}^l \alpha_i \alpha_j y_i y_j x_i^T \cdot x_j \quad (2.23)$$

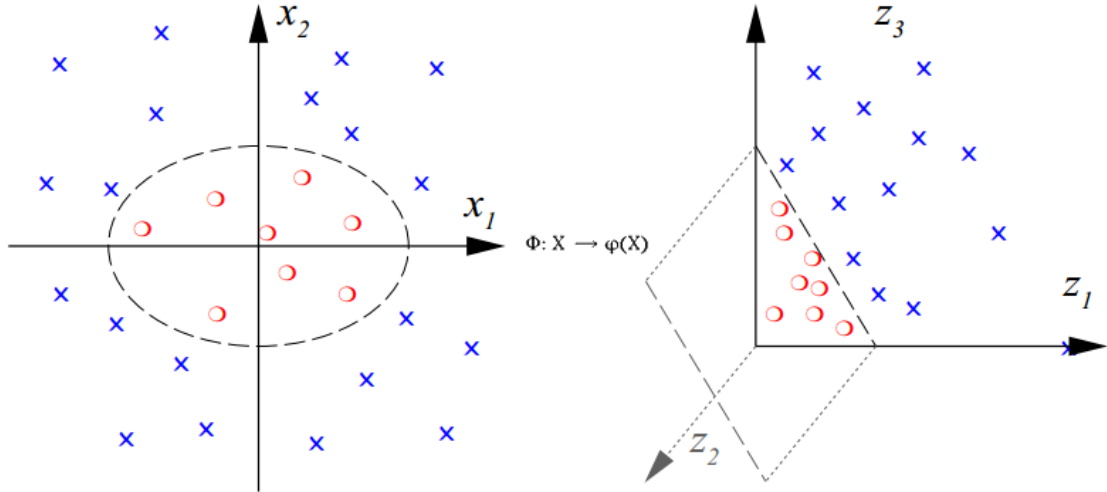
$$\sum_{i=1}^l \alpha_i y_i = 0 \quad (2.24)$$

$$0 \leq \alpha_i \leq C \quad i = 1, 2, \dots, l \quad (2.25)$$

b, (2.18) le aynı şekilde hesaplanır. $0 \leq \alpha_i \leq C$ aralığında yer alan Lagrange çarpanlarına karşılık gelen x_i vektörleri destek vektörleridir. Doğrusal ayrılamama optimizasyon problemi, doğrusal ayrılabilen durumdaki gibi, Lagrange denklemi ve KKT koşulları ile çözümlenmiştir.

2.4.3 Doğrusal olmayan sınıflandırma

Doğrusal olarak ya da bir ceza parametresi ile ayrılamayan veriler, yüksek boyutlu bir uzaya haritalanarak doğrusal ayrılabilir hale getirilebilir. x_i verisi, girdi uzayından, bir dönüşüm ile $\Phi(x_i)$ özellik uzayına aktarılır. Özellik uzayında işlem yapmanın çeşitli avantajları vardır. Özellik uzayındaki doğrusal işlemler, girdi uzayındaki doğrusal işlemlere karşılık gelir. Düzgün bir dönüşüm ile sınıflandırma daha kolay yapılabilir. Ancak özellik uzayı, çok boyutlu olduğu için işlem yapmak maliyetlidir. Şekil 2-8'de yüksek boyutlu uzaya haritalama gösterilmektedir.



Şekil 2.8 Doğrusal ayrılamayan verilerin yüksek boyutlu uzaya haritalanması

$K(x_i, x_j) = \phi(x_i^T) \cdot \phi(x_j)$ çekirdek fonksiyonu olarak tanımlanır ve Lagrange formülü, çekirdek fonksiyonu kullanılarak dönüştürülür.

$$L_\alpha = \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i,j=1}^l \alpha_i \alpha_j y_i y_j \phi(x_i^T) \cdot \phi(x_j) + b \quad (2.26)$$

$$x \rightarrow \phi(x)$$

(2.26) da doğrusal aşırı düzlem söz konusu olmadığından b terimi ihmal edilir. b terimi çekirdek fonksiyonu içinde kapalı biçimde yer almaktadır.

$$L_\alpha = \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i,j=1}^l \alpha_i \alpha_j y_i y_j K(x_i, x_j) \quad (2.27)$$

olarak gösterilir. Yeni tanımlanmış yüksek boyutlu uzayda karar fonksiyonu ise

$$\text{sgn}(f(x)) = \text{sgn}(w \cdot \phi(x) + b) \quad (2.28)$$

Literatürde kullanılan birçok çekirdek fonksiyonu mevcuttur. En yaygın olanlar ise şu şekilde sıralanabilir.

2.4.3.1 Doğrusal çekirdek

$$K(x_i, x_j) = x_i^T x_j + c \quad (2.29)$$

Doğrusal çekirdek en basit çekirdek fonksiyonudur. Nokta çarpıma c sabitinin eklenmesi ile çıkarılır. Özellik sayısının çok fazla olduğu durumlarda daha iyi sonuç vermektedir.

2.4.3.2 Polinom çekirdek

$$K(x_i, x_j) = (1 + x_i \cdot x_j)^d \quad (2.30)$$

Polinom derecesi d kullanıcı tanımlı olmak üzere 2,3,... derecelerini alabilir. Polinom çekirdek fonksiyonu yönlü bir fonksiyondur. Çıktı, çekirdekteki nokta çarpımı nedeniyle düşük boyutlu uzaydaki iki vektöre bağlıdır. Aynı yöndeki tüm vektörler, çekirdekten yüksek çıktı verirler. Çıktının genliği, vektör x_j nin genliğine bağlıdır.

2.4.3.3 Radyal tabanlı çekirdek

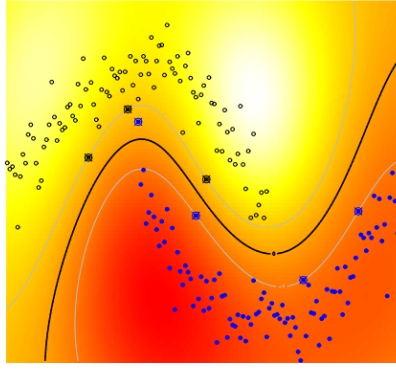
Sıklıkla kullanılan radyal tabanlı çekirdek, gaussian çekirdektir. Formülü:

$$K(x_i, x_j) = e^{-\frac{\|x_i - x_j\|^2}{2\sigma^2}} \quad (2.31)$$

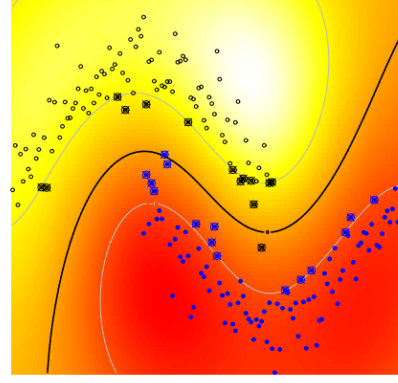
Parametre $\gamma = -\frac{1}{2\sigma^2}$ olarak tanımlanır ve denklem

$$K(x_i, x_j) = e^{\gamma(\|x_i - x_j\|^2)} \quad (2.32)$$

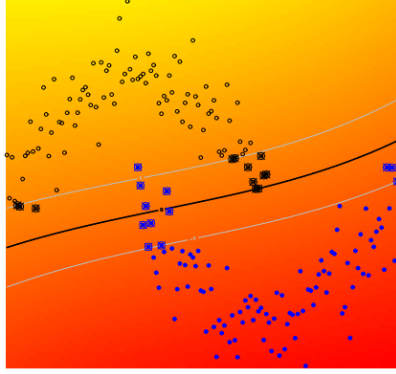
olarak tanımlanabilir. Çekirdek çıktısı, Öklid mesafesi olan x_i den x_j ye olan mesafeye bağlıdır. Bu iki değişkenden biri destek vektörü, diğeri ise test veri noktasıdır. Destek vektörü radyal tabanlı fonksiyon çekirdeğin merkezinde yer almaktadır. Büyük σ değeri, destek vektörleri sayısını azaltır. Her destek vektörü geniş bir uzayı tarif edebileceği için, daha az sayıda destek vektörü tercih edilir.



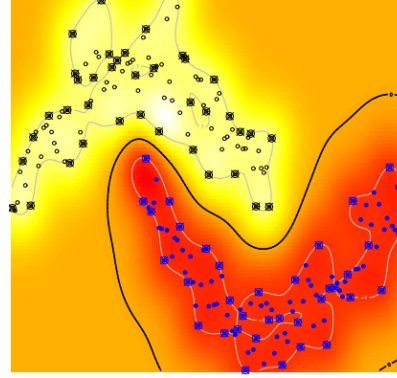
$\gamma = 0.5, C = 50$



$\gamma = 0.5, C = 1$



$\gamma = 0.02, C = 50$



$\gamma = 10, C = 50$

Şekil 2.9 Gaussian çekirdek parametre değişikliği ve veri ayrışımı

2.4.3.4 Sigmoid çekirdek

$$K(x, y) = \tanh(ax^T y + c) \quad (2.33)$$

Sigmoid çekirdek, çok katmanlı geriye yayılma algoritması çekirdeğidir. Yapay sinir ağları alanından gelmektedir. Çift kutuplu sigmoid çekirdeği çoklukla, yapay sinirlerdeki aktivasyon fonksiyonu olarak kullanılır.

2.4.4 Çok sınıflı sınıflandırma

DVM, temelinde ikili sınıflandırma yapan bir öğrenme algoritmasıdır. Ancak çeşitli yöntemler ile birden çok sınıflı sınıflandıracak ya da verilen birden çok sınıf arasında istenen sınıflı bulacak hale getirilebilmektedir. DVM için çok sınıflı sınıflandırmada sıklıkla iki yöntem kullanılmaktadır. Bunlar, bire karşı hepsi ve bire karşı bir yöntemleridir.

2.4.4.1 Bire karşı hepsi (BKH)

Bire karşı hepsi stratejisinde, sınıf başına bir DVM yaratılır. DVM'lerin her biri, bir sınıfın örneklerini tanımak ve diğer sınıflardan ayıt edebilmek amacıyla eğitilir. Çıkışların hepsi, olasılıklarına göre ayrı ayrı haritalandırır. Bunun için ek bir sigmoid fonksiyonu kullanır.

$$\hat{P}(w_i|f_i(x)) = \frac{1}{1+e^{(A_i f_i(x)+B_j)}} \quad (2.34)$$

$f_i(x)$, w_i sınıfını diğer sınıflardan ayıran eğitilmiş DVM çıktısıdır. Her sigmoidin A_i ve B_j parametreleri, (2.35) deki logaritmik formüle uygun hale getirilir.

$$- \sum_{k=1}^n t_k \log(p_k) + (1 - t_k) \log(1 - p_k) \quad (2.35)$$

p_k , sigmoidin çıktısı olmak üzere, t_k olasılık hedefidir. Bu optimizasyon problemini çözmek için Levenberg-Marquardt algoritmasına dayanan (2.36) kullanılır.

$$\sum_{j=1}^c \hat{P}(w_i|f_i(x)) = 1 \quad (2.36)$$

Olasılığın normalizasyonu ise

$$\hat{P}(w_j|x) = \frac{\hat{P}(w_i|f_i(x))}{\sum_{j=1}^c \hat{P}(w_i|f_i(x))} \quad (2.37)$$

şeklini alır [47].

n sınıflandırıcı, her sınıflandırıcı diğerlerinden bir sınıf ayıracak şekilde oluşturulmuştur. En kolay çok-sınıflı sınıflandırma yapılarından biri olmakla birlikte, çoğu uygulamalarda, bu yaklaşım diğerlerine göre alt seviyede kalmaktadır [46].

2.4.4.2 Bire karşı bir (BKB)

Bire karşı bir stratejisinin birçok ismi daha vardır; "ikili birleştirme", "tüm çiftler", "dairesele denetim". Her bir sınıf çifti için bir DVM yaratılması esasına dayanır. N sınıftan oluşan bir sınıflandırma problemini bu yöntemle çözmek için, $\frac{n(n-1)}{2}$ adet DVM eğitilmesi gerekir. Her bir DVM'in bir sınıfa oy verdiği maksimum oylama yöntemi kullanılır.

Çıktıların hepsi, olasılıklarına göre ayrı ayrı haritalandırır. Bunun için ek bir sigmoid fonksiyonu kullanır. Amaç, genel olasılık olan $\hat{P}(w_j|x)$ yi, yerel olasılık olan $\hat{P}(w_i|f_i(x))$ cinsinden ifade etmektir. Literatürde çeşitli yöntemler kullanılmakta olup, en çok kullanılanlardan birini inceleyebiliriz [48].

Tüm w_j sınıfları için

$$\sum_{j'=1, j' \neq j}^c \hat{P}(w_{j,j'}|x) - (c-2)P(w_j|x) = 1 \quad (2.38)$$

$w_{j,j'}$, w_j ve $w_{j'}$ sınıflarının bileşimidir. Devam edecek olursak;

$$\hat{P}(w_j|f_{j,j'}x) \approx \frac{\hat{P}(w_j|x)}{\hat{P}(w_{j,j'}|x)} \quad (2.39)$$

(2.39)'dan takip eden formülü çıkarabiliriz.

$$\hat{P}(w_j|x) = \frac{1}{\sum_{j'=1, j' \neq j}^c \frac{1}{\hat{P}(w_{j,j'}|x)}^{-(c-2)}} \quad (2.40)$$

Bire karşı hepsine benzer şekilde, bire karşı bir de genelleştirme hatalarında belirlenen sınırlara sahip değildir.

2.5 En Yakın N Komşu Sorgusu

Veri kümesindeki verilerin her birisinin, bir değerine olan uzaklıkların hesaplanması ve en küçük uzaklığa sahip n sayıda verinin seçilmesi esasına dayanmaktadır. Veri noktalarının birbirlerine olan uzaklıkların hesaplanmasında kullanılan çeşitli mesafe ölçütleri vardır. Sıklıkla kullanılan Euclidean metriği;

Euclidean mesafesi: x_s ve y_t vektörleri arasındaki mesafe

$$d_{st}^2 = (x_s - y_t)(x_s - y_t)' \quad (2.41)$$

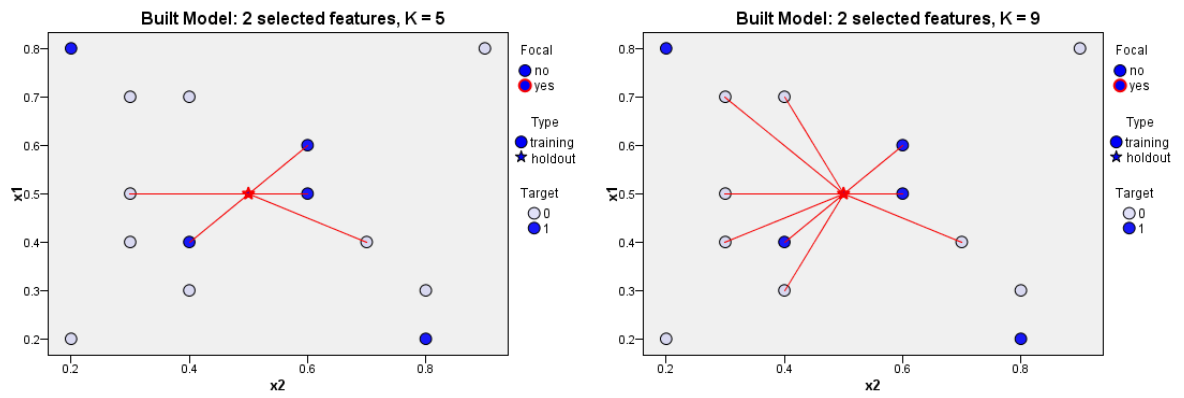
ile verilir. Euclidean metriği, Minkowski metriğinin özel bir halidir (p=2 durumu).

Diğer mesafe ölçütleri [62];

- Standardized Euclidean
- Mahalanobis
- City block
- Minkowski

- Chebychev
- Cosine
- Correlation
- Hamming
- Jaccard
- Spearman

olarak sıralanabilir.



Şekil 2.10 En yakın N(K) komşu sorgusu [63]

3. DUYGU TANIMA

Bu bölümde, öncelikle kullanılan veri kümesi tanıtılmıştır. Ses sinyalinde öznitelik vektörlerini çıkarmak için kullanılan pencere boyutu, atlama süresi, zamansal bölümlenme ve öznitelik çıkarım yöntemi aktarılmıştır. Son olarak da, DVM ile çıkartılan model ile yapılan sınıflandırma, çapraz doğrulama ve başarımların performansı ile hata matrisi anlatılmış ve kullanılan araçlar ile bölüm sonlandırılmıştır.

3.1. Veri Kümeleri

Ses sinyalinde duygu tanımlamada ilk aşama, kullanılacak olan veri kümesinin yaratılması ya da seçilmesidir. Literatürde birçok hazır veri kümesi bulunmaktadır. Bunlardan bazıları şunlardır.

eNTERFACE'05 Veri Kümesi [16]

42 katılımcıya 6 hikâyeye okutularak kaydedilmiş ses ve video görüntülerinden oluşmaktadır. Katılımcılar 14 farklı ülkeden olup, hepsi okumaları İngilizce seslendirmişlerdir. Katılımcılardan 6 adet duygu vermeleri istenmiştir. Bunlar: mutluluk, üzüntü, şaşkınlık, öfke, tiksinti ve korku. Örnekler, mini-DV kamera ile ve yüksek kaliteli mikrofona alınmıştır.

HUMAINE Veri Kümesi [59]

Tetiklenmiş ve doğal olarak oluşmuş duygusal ses, görüntü ve hareket verilerinden oluşmaktadır. Çeşitli veri tabanlarından ve konuşmalardan yakalanmış her biri 5 saniyeden 3 dakika ya kadar olabilen 52 klip içermektedir. Kategorik ve zaman düzleminde duygu tanımlayıcıları ile etiketlemeleri yapılmıştır.

RML Veri Kümesi [16]

Sessiz ve temiz bir ortamda, dijital video kamera kullanılarak çekilen 720 adet ses ve video kaydından oluşmaktadır. 6 adet duygudan oluşmaktadır: Öfke, tiksinti, korku, mutluluk, üzüntü, sürpriz. 8 farklı gönüllü, 6 farklı dil ile duyguları canlandırmıştır. Örnekler, 22050 Hz örnekleme oranı, 16 bit çözünürlükte, 30 kare/saniye olarak kayda alınmıştır.

Veri Kümesi [15]

Alman talk showu olan "Vera am Mittag" dan kaydedilmiş 12 saatlik ses kaydından oluşmaktadır. Almanca konuşmaları içermekte olup, doğal konuşmalardır. Görüntü, ses ve yüzden duygu analizi için ayrı modülleri mevcuttur. VAM-Audio: Bağımsız değerlendirmeciler tarafından etiketlenmiş toplam 1018 konuşmadan oluşmaktadır. 16 kHz örnekleme frekansında, 16 bit çözünürlükteki ses dosyalarıdır.

3.1.1 Emo-DB Veri Kümesi [26]

Tez çalışmasının deneysel kısmında da kullanılmış olan Emo-DB, ana dilleri Almanca olan 5 erkek 5 bayan oyuncudan oluşan toplam 10 kişinin, Almanca 10 cümleyi farklı duygular ile seslendirmesinden oluşturulmuştur. Ses dosyaları, 16 kHz, 16 bit, mono olarak kayıt edilmiştir. Ses dosyalarının her biri, biri duyguyu canlandırmaktadır. 20 kişi tarafından duyguların değerlendirilmesi ve etiketlenmesi yapılmıştır.

Çizelge 3.1 Emo-DB veri kümesinin 535 adet duygu ses sinyalinin dağılımları

Duygu	Kızgınlık	Korku	Sıkıntı	İğrenme	Mutluluk	Nötr	Üzüntü
Toplam	127	69	81	46	71	79	62

Çizelge 3.2 Emo-DB veri kümesinin konuşmacı bilgileri

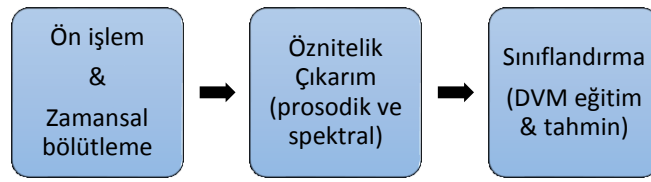
Konuşmacı numarası	Cinsiyet	Yaş
03	Erkek	31
08	Kadın	34
09	Kadın	21
10	Erkek	32
11	Erkek	26
12	Erkek	30
13	Kadın	32
14	Kadın	35
15	Erkek	25
16	Kadın	31

Çizelge 3.3 Emo-DB metin içerikleri

Metin kodu	Metin (Almanca)
a01	Der Lappen liegt auf dem Eisschrank.
a02	Das will sie am Mittwoch abgeben.
a04	Heute abend könnte ich es ihm sagen.
a05	Das schwarze Stück Papier befindet sich da oben neben dem Holzstück.
a07	In sieben Stunden wird es soweit sein.
b01	Was sind denn das für Tüten, die da unter dem Tisch stehen?
b02	Sie haben es gerade hochgetragen und jetzt gehen sie wieder runter.
b03	An den Wochenenden bin ich jetzt immer nach Hause gefahren und habe Agnes besucht.
b09	Ich will das eben wegbringen und dann mit Karl was trinken gehen.
b10	Die wird auf dem Platz sein, wo wir sie immer hinlegen.

3.2 Yöntemler

Ses sinyalinde duygu tanıma işlemleri için tez kapsamında kullanılan yöntemlerin blok şeması Şekil 4-1 de gösterilmiştir. Emo-DB veri kümesinden alınan ses dosyaları, duygularına göre gruplanmış, zamansal olarak bölütlenmiş, öznitelik vektörleri çıkarma yöntemleri ile öznitelik vektörleri çıkarılmış ve bu öznitelik vektörleri kullanılarak sınıflandırma yapılmıştır. Sonraki aşamalarda sonuçlar ve başarımlar değerlendirilmiştir.



Şekil 3.1 Sistem blok diyagramı

3.2.1 Ön işlem ve zamansal bölütleme

Bölüm 4.1.1 de tanıtılmış olan Emo-DB veri kümesi içerisindeki ses dosyaları, dosya sisteminde saklanma şekli olarak, XXXXXXXX.wav (örneğin: 03a01Fa.wav) formatında olup, ilk iki hane konuşmacının numarasını (03), 3,4,5 inci haneler konuşulan metni (a01: Der Lappen liegt auf dem Eisschrank), 6. hane seslendirilen duygunun almaca karşılığının ilk harfini (L Langeweile yani Sıkıntı), 7. hane de o seslendirmenin birden çok versiyonu varsa ona karşılık gelen ve a'dan başlayan alfabe harfine karşılık gelmektedir (a,b,c,...). Duygu dosyaları, duygu durumlarına göre ayrı dizinlere aktarılmış, dosya sisteminde kategorilendirilmiştir. Matlab'da yazılan kodlama ile her bir ses dosyası matlab uygulamasına alınmış, duygu adı, ses verisi, örnekleme oranı, toplam örnek sayısı, toplam süresi, saniyedeki bit sayısı, kanal sayısı gibi özellikleri bir yapı içinde saklanmıştır.

3.2.1.1 Pencere boyutu

Pencere boyutu, doğru bir ses analizi yapılabilmesi için önemli bir parametredir. Pencere, içindeki ses ile ilgili parametrelerin sabit kaldığının varsayıldığı belirli uzunluktaki zamansal ses bölütleridir. Tüm bir ses sinyali ile işlem yapmak yerine,

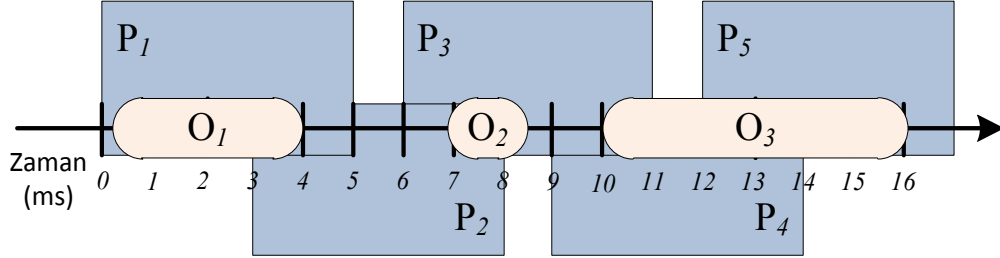
pencerelerle işlem yapmak, sinyaldeki veriyi anlamlı küçük parçalara ayırmamıza ve sinyaldeki değişiklikleri anlayabilmemize yardımcı olur. Pencere boyutu, istenilen analize göre değişiklik gösterebilmek ile birlikte, sesteki duygu tanımlama, sesteki kullanıcı tanımlama gibi bazı sinyal analizi çalışmalarında [13][15][24][28][30][31], 10 ila 200 ms aralığında kullanılmaktadır. Bu aralıkta, özellikle konuşma sinyali, değişkenliği az olan karakteristik yapı göstermektedir.

3.2.1.2 Atlama süresi

Analiz edilen ses sinyalinde, bir sonraki analiz edilecek pencerenin nereden başlayacağını ifade eder. Başka bir ifade ile yeni analiz edilecek pencerenin, bir önceki pencerenin başladığı yerden ne kadar atlama yapılarak başlayacağı şeklinde de anlatılabilir. Atlama süresi pencere boyutu kadar olursa, pencereler arasında üst üste binme olmaz. Pencere boyutundan daha büyük atlama süresinde ise pencereler arasında analiz edilmeyen örnekler kalacaktır. Atlama kullanılmasının ana amacı hesaplanan özniteliklerdeki değişimlerin daha yumuşak olmasının sağlanmasıdır.

Pencere boyutu ve atlama süresi ile ilgili Şekil 3.2'de ki örneği inceleyebiliriz. Bir ses sinyalinde, gülme ve silah atışı olaylarının olduğu bir ses sinyali olduğunu varsayalım. Bu iki olayın süreleri belirgin şekilde farklıdır. Gülme olayı, silah atışına göre oldukça uzun bir olaydır. Dolayısıyla ses sinyalinde gülbüz(güçlü) öznitelik vektörleri çıkarmak için pencere boyu ve atlama süresi önemli etkenler olacaktır. Şekil 3.2, pencere boyu ve atlama süresi arasındaki ilişkiyi açıklamaktadır. Şeklin x eksenindeki her bir aralığın 1 ms olduğunu kabul edelim ve pencere boyunu 5 ms, atlama süresini de 3 ms olarak tanımladığımızı varsayalım. O_1 tanımlamak istediğimiz ses olayı ve P_1 de analiz edilecek pencere olsun. Burada, 5 ms'lik pencere boyutu, O_1 ve O_2 ses olaylarını yakalayabiliyor ancak, seçilen pencere boyutu O_3 ün karakteristiğini yakalamak için yeterli olmuyor. Atlama süresini 3 ms olarak belirlediğimizde, O_2 ses olayının süresinin belirlenen pencere boyutunun içinde kalmadığı için, O_3 ün karakteristiğini tam olarak yakalayamamış oluyoruz. Tez çalışmada, bu sorunun üstesinden gelebilmek için, doğru boyutları ve parametreleri bulmak adına deneysel analizler yapılmıştır. Bu amaçla ses sinyalinde, pencere boyunca alınan veri, belirlenen atlama süreleri ile ötelenerek

tüm ses sinyali taranmıştır. Ses sinyalinde aranan bilginin zamansal boyutu ile değerlendirme yapılan pencere boyutu olabildiğince birbirine yakın olmalıdır [60].



Şekil 3.2 Duygu barındıran ses sinyalinde pencere ve atlama boyutu

3.2.1.3 Zamansal bölütleme

Öznitelik çıkarımında, pencere ve atlama süreleri önemli bir etken olduğu için, içinde duygunun karakteristiğini barındıran küçük ses bölümlenmeleri seçilmesi önem kazanmıştır. Farklı pencere ve atlama sürelerine göre, çıkarılan öznitelik vektörleri ile sınıflandırma yapılmış ve duygu tanıma başarımlarına bakılarak, en uygun pencere ve atlama sürelerine karar verilmiştir. Çizelge 3.4 de, test edilen pencere ve atlama süreleri gösterilmiştir.

Çizelge 3.4 Deneysel çalışmada kullanılan pencere ve atlama boyutları

Pencere Boyutu (ms)	Atlama Boyutu (ms)
10	5
10	10
20	10
20	20
30	10
30	15
30	30
50	10
50	25
50	50
100	10
100	50
100	100

3.2.2 Öznitelik çıkarım

Genel uygulama olarak, ses sinyalinde duygu çıkarımında, öncelikle gürbüz özniteliklerin belirlenmesi gerekir. Belirlenen özniteliklerden çıkarılacak vektörlerin, sayıca görece az, ancak ses sinyalinde aranan karakteristik özellikleri yansıtmaları istenir. İlerleyen bölümde, literatürde de yaygın olarak kullanılan öznitelik vektörleri çıkarım yöntemleri anlatılmış, MFCC üzerinde ayrıntılı durulmuştur [55].

Öznitelik vektörleri çıkarımı için, zamansal bölütlenmiş (ya da öznitelik çıkarma yöntemi tarafından bölümlenecek olan) her bir ses dosyası, öznitelik vektörü çıkarma yöntemi için yazılmış Matlab kodu tarafından işlenmiş ve öznitelik vektörleri olarak matrislerde saklanmıştır. Çıkarımı yapılan öznitelik vektörleri, doğrusal ya da vektörel olmalarına göre işlenmiş ve aynı düzleme indirgenmişlerdir. Bu durumu açıklamak için vektörel sonuç veren MFCC üzerinden örnek verebiliriz.

MFCC öznitelik çıkarma yöntemi, istenen uzunluktaki pencere verisinde, ses sinyalindeki her bir örnek için 20 adet katsayı (katsayı sayısı parametrik olup,

değiştirilebilir) çıkarır. Örnek vermek gerekir ise, 20 ms pencere boyutu ve 10 ms atlama süresi olan analiz edilen ses sinyalindeki bir penceresinde, 186 adet ses örneği vardır. Dolayısı ile bu pencere de 186x20 adet katsayı çıkar. Tüm ses dosyasındaki duygu durumlarındaki katsayıya erişmek için, örneklerin aynı katsayı için alınmış verilerinin aritmetik ortalaması alınır. Her bir örnek katsayısı (toplam 20 adet) matrisin bir satırına karşılık gelir. Böylece satırlarda katsayıların olduğu $n \times m$ bir matris oluşur. (3.1) de matrisin gösterimi mevcuttur.

$$F = \begin{bmatrix} F_{1,1} & \dots & F_{1,m} \\ \vdots & \dots & \vdots \\ F_{n,1} & \dots & F_{n,m} \end{bmatrix}_{n \times m} \quad (3.1)$$

Her bir sütunun aritmetik ortalaması alınması ile (3.2), her bir katsayı için ortalama \bar{F} değerleri bulunur ve her bir ses dosyası için toplam 20 adet katsayı elde edilmiş olur.

$$\bar{F}_i = \frac{\sum_{j=1}^n F_{j,i}}{n} \quad (3.2)$$

Doğrusal sonuç veren öznitelikler için, her bir ses dosyasının her bir penceresi için sonuçlar alınmış, her ses için çıkan satır matrisinin ortalaması alınarak tek değere indirgenmiştir.

3.2.2.1 Öznitelik vektörleri

Öznitelik vektörü, analiz edilen sinyalin taşıdığı bilgiyi temsil eden ve özgün sinyalin uygun miktarda bölünmesi ile oluşturulmuş sayısal veridir. Ses sinyalinde çıkarılabilecek öznitelik sayısı görece fazladır. Tez kapsamında 15 adet tekil ve 5 adet bileşik olmak üzere toplam 20 adet öznitelik vektörü çıkarma yöntemi kullanılmıştır. Deneysel çalışmada kullanılan 20 öznitelik arasında MFCC özniteliğinin duygu tanımadaki başarımı yüksek olduğu bulunduğu için, çalışmanın sonraki kısımlarında MFCC ile çalışılmaya devam edilmiştir. Öznitelik çıkarım yöntemlerinde MFCC üzerinde ayrıntılı olarak durulmuştur. Kullanılan öznitelikler, literatürde ses sinyalinde duygu analizi için sıklıkla kullanılan, MPEG-7 ses öznitelikleri [17], insan kulağı algısına yakın bir temsili olan Mel Frequency Cepstral Coefficients (MFCC) [4], Linear Prediction Coding (LPC), Zero Crossing Rate

(ZCR), Fundamental Frequency (F0), Matching Pursuit (MP)'dir. Ek olarak 4 farklı bileşke öznelik kullanılmıştır. Çizelge 3.5'de ayrıntılı olarak verilmiştir.

Çizelge 3.5 Kullanılan ses sinyali öznelik vektörü çıkarım yöntemleri

Ses öznelik vektörü çıkarım yöntemi	Sunum
Audio Spectrum Centroid (ASC)	Doğrusal
Audio Spectrum Spread (ASS)	Doğrusal
Audio Spectrum Flatness (ASF)	Vektörel
Audio Spectrum Envelope (ASE)	Vektörel
Audio Spectrum Projection (ASP)	Vektörel
Harmonic Spectral Centroid (HSC)	Doğrusal
Harmonic Spectral Deviation (HSD)	Doğrusal
Harmonic Spectral Spread (HSS)	Doğrusal
Harmonic Spectral Variation (HSV)	Doğrusal
Spectrum Centroid (SC)	Doğrusal
Zero Crossing Rate (ZCR)	Doğrusal
Fundamental Frequency (F0)	Doğrusal
Linear Predictive Coding (LPC)	Vektörel
Mel Frequency Cepstral Coef. (MFCC)	Vektörel
Matching Pursuit (MP)	Vektörel
MFCC + ASS + ASC	Vektörel
ASF + ASS + ASC	Vektörel
LPC + ASS + ASC	Vektörel
MP + ASS + ASC	Vektörel
MFCC + ASF	Vektörel

MPEG-7 Audio Spectrum Centroid (ASC): Ses sinyalinin log frekans güç spektrumundaki ağırlık merkezini tanımlamaktadır. Sinyaldeki baskın olan yüksek ve düşük frekans bileşenlerini gösterir.

MPEG-7 Audio Spectrum Spread (ASS): Ses sinyalinin ağırlık merkezi etrafındaki spektrum dağılımını ortaya çıkarmaktadır.

MPEG-7 Audio Spectrum Flatness (ASF): Ses sinyalinin güç spektrumundaki düzlük özelliğini yansıtmaktadır. Her alt bant ve her pencere aralığı için spektrum gücünün geometrik ortalamasının aritmetik ortalamasına oranıdır.

MPEG-7 Audio Spectrum Envelope (ASE): Sinyalin her bir çerçevesindeki her bir bant için, kısa zamanlı güç spektrumunu tanımlar.

MPEG-7 Audio Spectrum Projection (ASP): ASE öznitelik vektörü çıkarım yönteminin normalize edilmiş halidir.

MPEG-7 Harmonic Spectral Centroid (HSC): Genlik ağırlıklı ortalama bir sinyalin süresi boyunca, harmonik tepe noktalarının ortalaması olarak tanımlanmaktadır.

MPEG-7 Harmonic Spectral Deviation (HSD): Harmonik tepe noktalarının, yerel spektrumdan kapsamından sapmasını ölçer.

MPEG-7 Harmonic Spectral Spread (HSS): HSC ilişkisinde, ortalama spektrum ayrışmasını ölçer.

MPEG-7 Harmonic Spectral Variation (HSV): Yakın çerçeveler arasındaki spektral değişiklikleri yansıtır.

MPEG-7 Spectrum Centroid (SC): Log-frekans spektrum gücünün çekim merkezini verir.

Zero Crossing Rate (ZCR): Ses sinyalinin 0'dan geçiş sayısıdır. Ses sinyalinin bulunmadığı yerlerde, yüksek gürültüden dolayı değeri artmakta, ses sinyalinin olduğu yerlerde düşmektedir.

Fundamental Frequency (F0): Ses üretim mekanizmasında ses üretimi için ses tellerinde yarı-periyodik titreşimler oluşur. Ses tellerinde oluşan bu titreşimin frekansına temel frekans denir. Perde frekansı olarak da adlandırılır.

Linear Predictive Coding (LPC): Şu andaki örnek sesli ifade ile doğrusal olarak tahmin edilen sesli ifade arasındaki farkların karelerinin toplamı en aza indirilmeye çalışılarak, sesli ifadenin tahminini sağlayacak bir dizi birim katsayı bulunabilir. Bu katsayılara tahmin edici katsayılar (Linear Predictor Coefficients) denir ve tahmin

edilen sesli ifadenin doğrusal olarak birleştirilmesi sırasında kullanılan ağırlıklandırma katsayıları olarak da tanımlanabilirler.

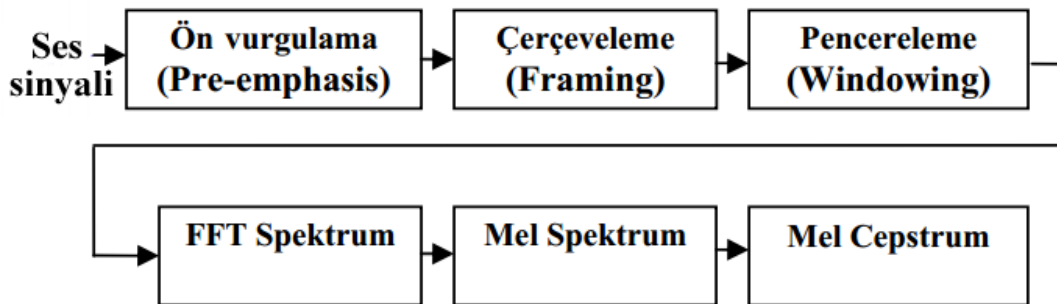
Matching Pursuit (MP): Seyrek yakınsama (sparse approximation) yöntemlerinden biridir. Öncelikle ölçülen sinyalin parametrik modelini kullanarak, olabilecek her parametre kombinasyonu için bir model sözlük oluşturulur. Sonra, en küçük kareler hatasını en aza indiren sözlük öge kombinasyonu bulunur. Yönteme seyrek denilmesinin sebebi, ölçülen sinyalin çok az sayıda ögenin doğrusal kombinasyonu olarak temsil edilmesidir [49].

3.2.2.1.1. Mel Frequency Cepstral Coefficients (MFCC)

İnsan kulaklarının algılamasını taklit eden ve hızlı Fourier transformu tabanlı olarak hesaplanan bir sayısal teknik analizdir. MFCC, değişimlerden, ses dalga yapısından çok daha az etkilenir. Ses sinyalinin düşük frekans bileşenleri, daha fazla fonem bilgisi taşır. Doğrusal olmayan mel ölçek bankası, yüksek olanlar üzerinden, düşük frekans bileşenlerini güçlendirir. Konuşma çerçevesinin kısa dönem güç spektrumu, doğrusal olmayan mel frekans ölçeğinde, log güç spektrumu üzerinde doğrusal kosinüs dönüşümü kullanır. Normal frekans f den mel frekansı olan m ye dönüşüm için

$$m = 2595 \log_{10}\left(\frac{f}{700} + 1\right) \quad (3.3)$$

eşitliği kullanılır [50].



Şekil 3.3 MFCC özneliklerinin çıkarılma blok diyagramı

Bir ses sinyalinde, MFCC katsayılarını almak için izlenen yol Şekil 3.3'de gösterilmiştir.

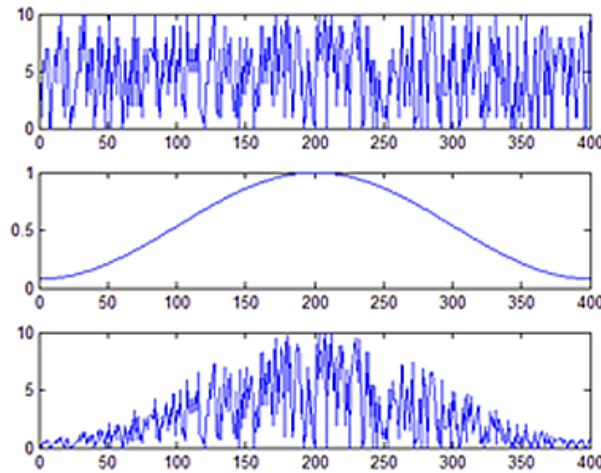
Ön vurgulama: Ses sinyali birinci dereceden, sonlu dürtü yanıtı (FIR) bir filtreden geçirilerek spektral olarak düzlenir. Yüksek frekanslı bölgeleri ise güçlendirir, spektrumu dengeler.

Çerçeveleme: Ses sinyali karakteristikleri yeteri kadar küçük bir zaman aralığında kararlı kalmaktadır. Bu nedenle ses sinyalleri kısa zaman aralıklarında işlenmektedir. Sinyaller, genellikle 10 ila 200 milisaniye arasında değişen uzunluklarda çerçevelere bölünürler. Her çerçeve kendinden önceki çerçevenin bir bölümünü içerisinde barındırır, örter. Örtüşmeli çerçevelerin kullanılmasının nedeni, çerçeveden çerçeveye geçişlerde oluşacak süreksizlik etkilerini en aza indirmektir, bir çerçeveden diğerine geçişin yumuşak olmasını sağlamaktır.

Pencereleme: İkinci aşama olarak çerçeveler, pencerelenir. Analiz pencerelerinin başındaki ve sonundaki süreksizliği azaltmak için bir pencere fonksiyonu uygulanır. Ses uygulamalarında sıklıkla (3.4) eşitliği ile verilen hamming penceresi tercih edilir.

$$w(n) = 0.54 - 0.46 \times \cos\left(\frac{2\pi n}{N-1}\right) \quad (3.4)$$

Şekil 3.4'de bir ses sinyaline hamming penceresi uygulanarak başlangıç ve bitiş yumuşatması gösterilmektedir.



Şekil 3.4 Ses sinyaline hamming penceresi uygulaması

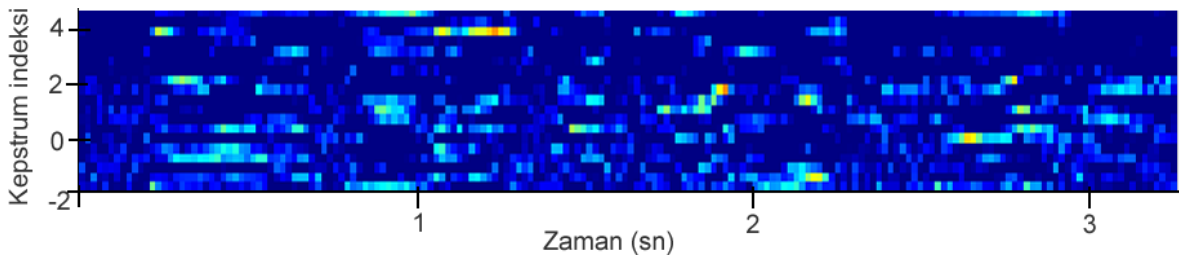
FFT Spektrum: Pencereler üzerinde hızlı fourier dönüşümü (Fast Fourier Transform) uygulanır. Bu dönüşüm ses sinyalini zaman uzayından frekans uzayına çevirir. Bu dönüşüm ile hangi frekanslarda ne şiddette bir titreşim olduğunu gösteren spektogram elde edilir.

Mel Spektrum: İnsan kulağı, ses frekanslarını doğrusal olmayan şekilde algılar. 1 kHz e kadar normal, 1 kHz den sonra ise logaritmik olarak algılar. Bu ölçüye ise mel ölçüsü denilir. Mel spektrumuna geçiş için, bant genişliği mel ölçeğine göre değişen üçgen filtre kullanılır. Bir frekans mel frekansına çevirmek için (3.3)'deki formül uygulanır.

Kepstrum: Sinyalin genlik spektrumu mel ölçeğinde eşit aralıklarla dağılmış ve birbiriyle %50 oranında kesişen N adet band geçiren üçgen filtreye çarpılır. Çarpım sonucunda her bir filtre içinde kalan enerjinin logaritması hesaplanır. Mel spektrum katsayılarının logaritması gerçek sayılar olduğu için (3.5) bağıntısıyla verilen ayrık kosinüs dönüşümü kullanılarak tekrar zaman bölgesine geçilir. Sonuçta elde edilen katsayılar mel frekanslı kepstrum katsayısı (MFCC) olarak isimlendirilir [51].

$$C_n = \sum_{k=1}^K (\log S_k) \cos \left[n \left(k - \frac{1}{2} \right) \frac{\pi}{K} \right] \quad (3.5)$$

Bu eşitlikte elde edilen $S_k, k = 1, 2, \dots, K$ mel spektrum katsayılarıdır.



Şekil 3.5 Ses sinyali için Mel Frekanslı Kepstrumu

3.2.3 Sınıflandırma yöntemi

Sınıflandırma için kullanılan öğrenme yöntemi olarak literatürdeki konu ile ilgili yoğun kullanımı nedeni ile DVM seçilmiştir [1]-[5]. Uygulama olarak sayısal hesaplama yapabilen Matlab uygulaması kullanılmıştır. DVM'nin, Matlab uygulamasında yüksek başarımından dolayı libSVM, DVM kütüphanesi kullanılmıştır [25]. Verileri çok boyutlu öznitelik uzayına eşlemler için Bölüm 2.5.3.3'de açıklanmış olan radyal tabanlı çekirdek fonksiyonu (RBF) kullanılmıştır. Verilerde normalizasyon yapmak için z-score algoritması kullanılmıştır. Normalizasyon, makine öğrenme ve örüntü tanıma sisteminin performansını artırmak amacıyla, veri kümesinin bulunduğu uzaydan başka bir uzaya taşınması işlemidir. Algoritmada var olan x özniteliği, (3.6) ile x' özniteliğine dönüştürülerek normalizasyon gerçekleştirilir.

$$x' = \frac{x - x_{ort}}{x_{std}} \quad (3.6)$$

x_{ort} , x özniteliğinin ortalaması, x_{std} ise x özniteliğinin standart sapmasıdır. DVM, ikili sınıflandırma yaptığı için, Bölüm 3.3.4'de anlatılan BKH yöntemi kullanılarak, DVM çoklu sınıflama yapabilecek duruma getirilmiştir.

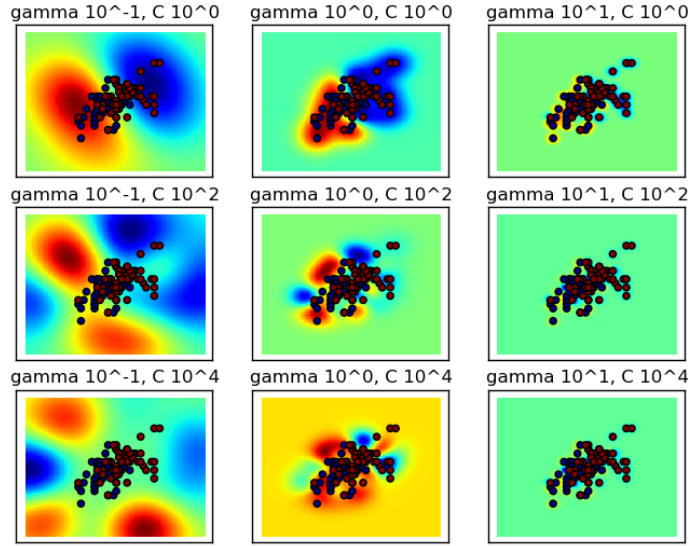
Eğitim ve test veri kümeleri için n kere çapraz doğrulama yöntemi kullanılmış olup $n=10$ alınmıştır. BKH yöntemi ile kazanan hepsini alır strateji kullanılır. Her bir veri operasyonu 10 kez çalıştırılmış, her çalıştırma için BKH yaklaşımı ile bir model oluşturulmuştur. Oluşturulan model üzerinden test verisi için sınıflandırma yapılmış, genel başarımlar için tüm test verilerinin ortalamaları alınarak, doğru tanıma başarımlar oranı ve sonuçlar değerlendirilmiştir.

Parametre optimizasyonu için yapılan çalışmada, DVM öğrenme algoritmasının 4 farklı parametresinin optimizasyonu ile çalışılmıştır. Parametrelerin çeşitli değerleri tekil ve kombinasyonlu şekilde, çapraz doğrulama yapılarak sınıflandırma için kullanılmış ve başarımlar ve sonuçlar karşılaştırılarak yorumlanmıştır. Kullanılan DVM parametreleri şunlardır.

Ceza Sabiti C: C sabiti bölüm 3.3.2'de de formülüne edildiği gibi, eğitim hataları ile karar uzayının basit kalması arasındaki ödünleşimdir. İkisinin de olumlu olamayacağı, birini düzeltirken diğerinin kötüleştiği bir etkisi vardır. Düşük C

değerleri, karar uzayının yumuşak, yüksek C değerleri ise eğitim verilerinin doğru sınıflandırmasını hedefler.

Gama (γ): γ parametresi, belli bir boyutu olan bir veri alt uzayına karşılık gelir. Bu veri alt uzayı da aşırı düzlemi ayıran optimal karmaşıklığın derecesini kısıtlar.



Şekil 3.6 Farklı C ve γ parametreleri için sınıflandırma durumları [52]

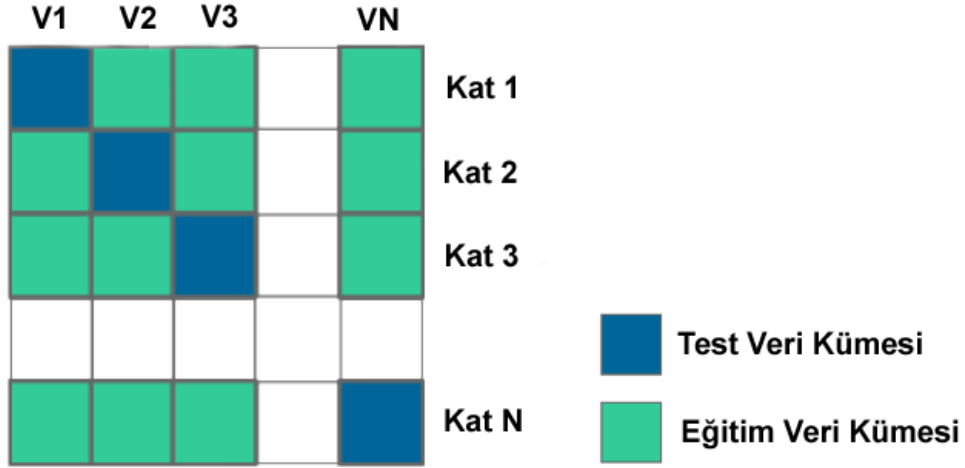
Sonlandırma Kriterinin Toleransı (ϵ): ϵ , öğrenme döngüsünün eşik değeridir. Öğrenme algoritmasının, öğrenmeye nerede son vereceği, bu parametre ile belirlenir.

Katsayı (r): r , çekirdeğin katsayı operasyonudur.

3.2.4 Çapraz doğrulama

Çapraz doğrulama, öğrenme algoritmasının sonuçlarını değerlendirmek ve karşılaştırmak için kullanılan istatistiksel bir yöntemdir. Veri kümesini eğitim ve doğrulama kümesi olmak üzere ikiye ayırır. Veri kümesi rastgele n gruba ayrılır. 1. grup test için ayrılırken geriye kalan gruplarla model kurulur. Kurulan model doğrulama için ayrılan veriler üzerinden test edilir ve doğruluk oranı hesaplanır. Süreç n defa tekrar eder ve modelin doğruluk oranı, n tane doğruluk oranının

aritmetik ortalaması kadar olur. $N=10$ seçilirse, çapraz doğrulamada, 10 kümeden 9'u eğitim verisi, 1'si test verisi olarak modelleme gerçekleştirilir. Bir testte toplam 10 çapraz doğrulama gerçekleştirilir. Böylece her bir küme, hem eğitim hem de test verisi olarak test sürecine dahil edilir. Elde edilen sonuçlar 10 test üzerinden gerçekleştirilmiş olur.



Şekil 3.7 N kere çapraz doğrulama

3.2.5 Recognition accuracy (tanıma doğruluğu) ve hata matrisi

Eğitim sonucu oluşturulan modelin başarımı değerlendirilirken çeşitli kavramlar kullanılır. Sıklıkla kullanılanları; hata oranı, duyarlılık (precision), anma (recall), doğruluk (accuracy) ve f-ölçütü (f-ratio) olarak sıralanabilir. Test sonucunda ulaşılan sonuçların başarım bilgileri hata matrisi ile ifade edilebilir (Çizelge 4.6). Hata matrisinde satırlar test kümesindeki örneklere ait gerçek sayıları, kolonlar ise modelin tahminlemesini ifade eder. Hata matrisi, yalnızca modelin nasıl tahmin edildiğini değil, aynı zamanda sorunlu işlemlerin nerede olduğunu detaylı biçimde gösterir.

Çizelge 3.6 İki sınıflı veri için hata matrisi

Gerçek Sınıf	Tahmin Edilen Sınıf	
	Sınıf 1 (Pozitif)	Sınıf2 (Negatif)
Sınıf 1 (Pozitif)	TP	FN
Sınıf 2 (Negatif)	FP	TN

Çizelge 3.6'daki kısaltmaların açılımlarını yapacak olur isek;

True Positive (TP): Doğru Negatif; kendisi ve tahmin edilen sınıfı aynı olan örnek sayısı

False Negative (FN): Yanlış Negatif; kendisi başka bir sınıf olarak tahmin edilen örnek sayısı

False Positive (FP): Doğru Pozitif; tahmini başka bir sınıf olarak sınıflandırılan örnek sayısı

True Negative (TN): Yanlış Pozitif; doğru olarak sınıflandırılmış, diğer sınıfa ait örnek sayısı

dırlar.

Başarım kriterleri ise;

$$Duyarlılık = \frac{TP}{TP+FP} \quad (3.7)$$

$$Anma = \frac{TP}{TP+FN} \quad (3.8)$$

$$Doğruluk = \frac{TP+TN}{TP+TN+FP+FN} \quad (3.9)$$

$$F - ölçütü = \frac{2 \times Duyarlılık \times Anma}{Duyarlılık + Anma} \quad (3.10)$$

$$Hata Oranı = \frac{FP+FN}{TP+TN+FP+FN} \quad (3.11)$$

olarak tanımlanır.

3.3 Araçlar

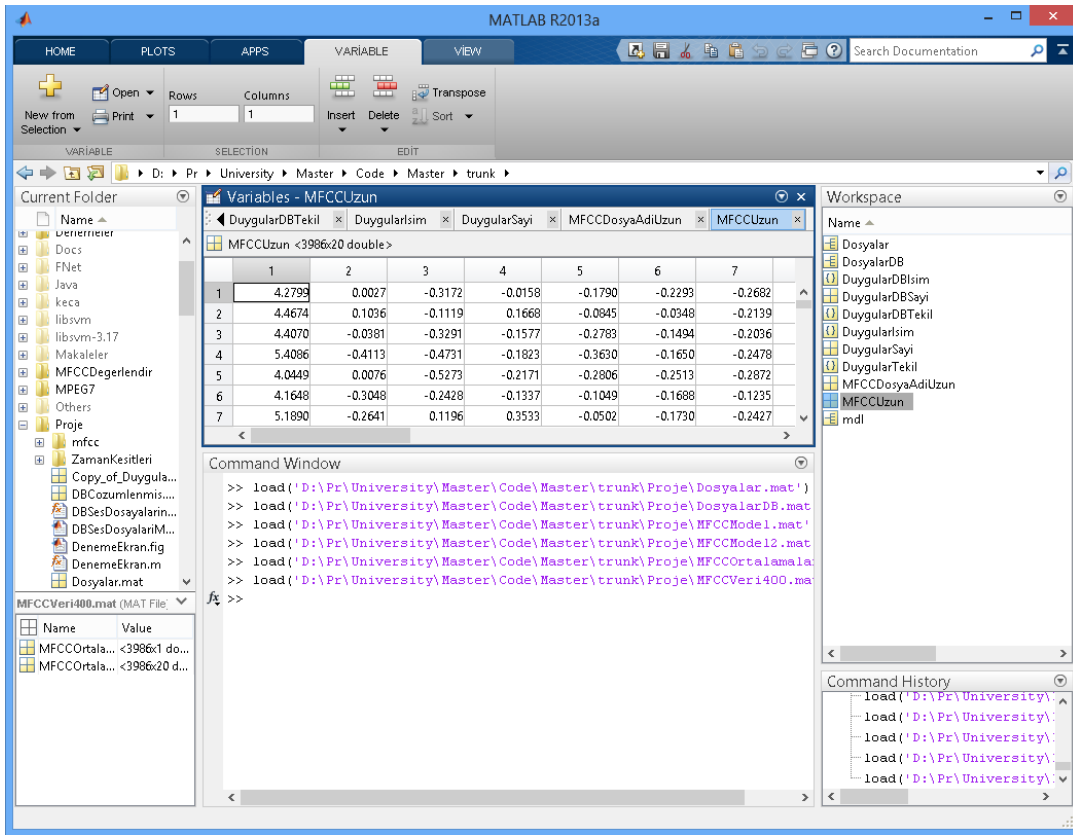
Bu bölümde, ses sinyalinde duygu tanımlama için kullanılan araçlar tanıtılmıştır.

3.3.1 İşletim sistemi

İşletim sistemi olarak, masaüstünde yoğun kullanımı ve yazarın genel tecrübesi olduğu için Microsoft Windows8 seçilmiştir. Kullanılan diğer araçlar da (Matlab ve libSVM) Windows8 üzerinde kararlı ve sorunsuz çalıştıkları gözlemlenmiştir.

3.3.2 Matlab

Matlab, Matworks firması tarafından geliştirilmiş, temel olarak nümerik hesaplama, grafiksel veri gösterimi ve programlamayı içeren teknik ve bilimsel hesaplamalar için yazılmış yüksek performansa sahip bir yazılımdır [54]. Yazılım çalışmasında araç olarak kullanılmıştır.



Şekil 3.8 Matlab uygulaması ekran görüntüsü

3.3.3 libSVM

LibSVM, DVM öğrenme algoritmasını uygulayan bir yazılım kütüphanesidir. National Taiwan University de geliştirilmiş olup, C ve C++ ile yazılmış ve Java ile güncellenmiştir. Linux altında çalışan paket program versiyonu da mevcuttur. Birçok DVM formülünü destekler. Çoklu DVM sınıflandırma için araçlar sunar. Çeşitli çekirdekler ile çalışabilmektedir. Matlab, Python, R, Perl, Ruby, WEKA, Common LISP, CLISP, Haskel, PHP, C# ve CUDA için genişleme paketleri mevcuttur.

Tez çalışmasında, Matlab uygulaması içerisinde, libSVM kütüphaneleri ile DVM eğitimi ve testleri gerçekleştirilmiştir.

4. DUYGU GERİ ERIŞİMİ

4.1 Veri Kümesi

Deneysel çalışmada kullanılmak üzere, içlerinde çeşitli duygular barındıran Türk TV dizilerinden ortalama 20'şer saniye uzunluğunda, 20 adet ses klipi yakalanmıştır. Ses dosyaları, 16 kHz, 16 bit, mono olarak kayıt edilmiştir. Her bir klip, 1'er saniyelik alt kliplere bölünmüştür. Toplam 20 klip ve 409 alt klip oluşturulmuştur. Klipler ve alt klipler 2 kişi tarafından, içerdiği ayrık duygu durumuna göre etiketlenmiştir. Etiketleme sonucu ses dosyalarının MFCC öznitelikleri çıkarılmış, DVM ile eğitilmiş, 10 kere çapraz doğrulama yöntemi ile test edilmiştir.

Veri kümesi için duygu durumları Çizelge 4-1 de gösterilmiştir. Veri kümesi için eğitim sonucu oluşturulan ve ortalama başarıma yakın model, çalışmanın sınıflandırma kısımlarında kullanılmak üzere sisteme kayıt edilmiştir.

Çizelge 4.1 TV dizileri veri kümesinin 409 adet duygu ses sinyalinin dağılımları

Duygu	Kızgınlık	Korku	Sıkıntı	İğrenme	Mutluluk	Nötr	Üzüntü
Toplam	71	5	49	14	121	116	53

4.2 Sorgular

Sert ve Baykal [66] çalışmalarında, ses verilerinin içerik tabanlı ve anlamsal sorgulanması için 3 farklı sorgu yaklaşımını incelemektedir. Bu yaklaşımların çalışmamızdaki kullanım biçimleri izleyen bölümlerde açıklanmaktadır.

4.2.1 Anahtar kelime sorgusu

Bu sorgu türünde, daha önceden hangi duygu durumunda olduğu etiketlenmiş ses bölütlerinin, kelime-tabanlı olarak aranması hedeflenmektedir. Etiketleme işlemi, el ile ya da sınıflandırıcı kararı olabilmektedir. Eşleşme için, aramada verilen anahtar kelime ile duygu etiketlerinin tam eşleşmesi söz konusudur.

4.2.2 İçerik tabanlı sorgu

Sisteme örnek olarak verilen bir ses dosyası üzerinden, bu sese ya da barındırdığı duyguya en yakın ses sinyalini, veri kümesi üzerinde sorgulama yöntemidir. Öznitelik ve otomatik sınıflandırma olmak üzere iki farklı şekilde incelenmiştir.

4.2.2.1 Öznitelik tabanlı

Bu sorgu türünde, öncelikle sisteme örnek olarak bir ses klipi verilir. Verilen bu ses klibinin MFCC öznitelikleri çevrimiçi olarak çıkarılır ve veritabanında daha önceden öznitelikleri çıkarılmış ses klipleri ile karşılaştırılır. Belirli bir ölçüte göre yapılan karşılaştırma sonucu kullanıcıya geri döndürülür. Bu aşamada, 3 farklı sorgu türüne göre sonuç üretilebilir. Bu sorgu türleri;

- Nokta sorgu,
- Aralık sorgusu,
- En yakın N komşu sorgusu

olarak tanımlanmıştır. Seçime göre en yakın sonuçlar gösterilir. En yakın komşu aramasında mesafe ölçütü olarak;

- Euclidean Mesafesi,
- Standardized Euclidean Mesafesi,
- City Block Mesafesi,
- Chebychev Mesafesi,
- Minkowski Mesafesi,
- Mahalanobis Mesafesi,
- Cosine,
- İlinti,
- Spearman,
- Hamming Mesafesi,
- Jaccard,

yöntemleri seçilebilir olarak uygulanmıştır.

Sonuçların başarımlarını ölçmek için Ortalama Duyarlık Ölçütü (ODÖ – Mean Average Precision – MAP) kullanılmıştır. ODÖ, birden çok sorgu barındıran sistemler için sistemin duyarlık değerlerinin ortalamasıdır ve (4-1) ile ifade edilir.

$$MAP = \frac{1}{N} \sum_{j=1}^N \frac{1}{Q_j} \sum_{i=1}^{Q_j} P_i \quad (4-1)$$

(4-1) eşitliğinde;

P_i : i. ilgili eleman için duyarlık

N : sorgu sayısı

Q_j : j sorgusu için ilgili eleman sayısıdır.

Veri kümesinde, her bir duygu durumu için 5'er adet test klipi sisteme verilmiş, her bir klip bir sorgu olarak ele alınmıştır. Her bir sorgunun ortalama duyarlılığı hesaplanmış, sonuçta ise her bir duygu durumu için ODÖ hesaplanmıştır.

4.2.2.2 Sınıflandırma tabanlı

Bu sorgu türünde, duygusu bilinmeyen örnek bir ses klipi sisteme verilir. Bu örnek, makine öğrenme algoritması ile daha önceden oluşturulan model ile test edilir ve duygu durumu öğrenilir. Son aşamada, öğrenilen duygu durumu için veritabanında arama gerçekleştirilir. Bu aşamada gerçekleştirilen sorgu “kelime-tabanlı” sorgu olarak değerlendirilir ve sonuçları listelenir.

4.2.3 Zamansal sorgular

Zamansal sorgularda, iki duygu durumunun birbirine olan zamansal durumlarına göre sorgulanması amaçlanmaktadır. Bu sorgu türüne, “Üzüntü'den sonra mutluluk gelen ses kliplerini getir” sorgusu örnek olarak verilebilir. Zamansal sorgulama için Allen'in aralık algoritması [61] kullanılmıştır. Allen'in aralık algoritmasında, 13 farklı operatöre göre zamansal sorgu yapılabilmektedir. Bu sorgu operatörleri a ve b olarak adlandırılan ses klipleri ile örneklenirse;

- Precedes: önünde olmak. A klipi b klipinin önündedir ve o başlamadan biter.
- Meets : buluşmak. A klipi biter bitmez b klipi başlar.

- Overlaps: üst üste binmek. B klibi a klibi başladıktan sonra ve bitmeden başlar. O bittikten sonra da devam eder.
- Finished by: ile bitmek. A klibi, b klibi başlamadan başlar ve b klibi ile aynı anda biterler.
- Contains: kapsamak. A klibi b klibinden önce başlar ve sonra biter.
- Starts: başlamak. A klibi başladığında b klibi de başlar.
- Equals: eşittir. A klibi, b klibine eşittir.
- Started by: ile başlar . B'yi kapsayan a klibi başladığında, b klibi de başlar.
- During: süresince. A klibi, b klibi süresi içinde devam eder.
- Finishes: bitmek. A klibi ile b klibi aynı anda biter.
- Overlapped by: üst üste binmek. A klibi b klibi başladıktan sonra ve bitmeden önce başlar ve o bittikten sonra da devam eder.
- Met by: buluşmak. B klibi biter bitmez a klibi başlar.
- Preceded by: önünde olmak. B klibi a klibinden önce başlar ve a klibi başlamadan biter.

Allen aralık algoritmasının grafiksel gösterimi Şekil 4.1 de verilmiştir.

precedes	meets	overlaps	finished by	contains	starts	equals	started by	during	finishes	overlapped by	met by	preceded by
p	m	o	F	D	s	e	S	d	f	O	M	P

Şekil 4.1 Allen'in aralık algoritması

Allen'in aralık algoritması, önceden etiketlenmiş ses dosyalarının duygu durumları üzerinde metin tabanlı arama yapmaktadır. Şekil 4.1 de gösterildiği üzere, mutluluk duygusu precedes (önünde olma) operatörü ile üzüntü duygusu aranmaktadır.

5. DENEYSEL ÇALIŞMALAR VE DEĞERLENDİRME

Tez çalışmasında yapılan çalışmaları, duygu sınıflandırma ve duygu geri erişim çalışmaları olarak iki ayrı başlık olarak inceleyebiliriz.

Duygu sınıflandırma başlığında; gerçekleştirilen uygulama altyapısını test edebilmek amacı ile, 3 deneysel çalışma gerçekleştirilmiştir. İlk çalışmada, MFCC öznitelik vektörleri ile, 10-kere çapraz doğrulama yöntemi kullanılarak, DVM parametre optimizasyonu gerçekleştirilmiştir. İkinci çalışmada, öğrenme algoritmasının başarımını arttırmak için farklı pencere boyutları ve atlama süreleri ile ilgili testler gerçekleştirilmiştir. Üçüncü olarak, 20 öznitelik kümesinin her bir elemanın duygu tanımlama güçlerini ölçmek için DVM ile değerlendirme yapılmıştır.

Duygu geri erişim başlığında; yeni bir veri kümesi oluşturulmuştur. Veri kümesi, yerli TV dizilerinden 20'şer saniyelik klipler şeklinde alınmış, 1'er saniyelik alt kliplere bölünerek duyguları etiketlenmiştir. Veri kümesi üzerinde anahtar kelimeye göre, içerik tabanlı ve zamansal sorgular oluşturulmuş ve sonuçları değerlendirilmiştir.

5.1 Duygu Sınıflandırma Deneyleri

5.1.1 DVM parametre optimizasyonu

DVM öğrenme algoritmasının parametre optimizasyonu için, literatürde duygu tanıma da sıklıkla kullanılan MFCC öznitelikleri kullanılmıştır [4]. Emo-DB veri kümesinin 535 adet ses sinyalinin hepsi çalışmada kullanılmıştır. Hazırlanan veri kümesinin 10 da 9'u eğitim ve 10 da 1'i test için kullanılmıştır (10-kere çapraz doğrulama).

MFCC özniteliği için varsayılan pencere boyutu 30 ms ve varsayılan atlama süresi 10 ms kullanılmıştır. DVM'nin varsayılan parametreleri olan Çizelge 5.1'de gösterilmiş değerlerle analiz çalışmasına başlanılmıştır.

Çizelge 5.1 DVM varsayılan parametreleri

C	r	ε	Y
1	0	0,001	1/7

Varsayılan DVM parametreleri ile yapılan çalışma sonucu, ortalama başarımları oranı %72,14 olarak çıkmıştır. Ortalama değere yakın (%72,22) hata matrisi Çizelge 5.2 de gösterilmiştir.

Çizelge 5.2 DVM varsayılan parametreler için hata matrisi

Referans Duygu	Tanınan Duygu						
	Öfke	Endişe / Korku	Sıkıntı	İğrenme	Mutluluk	Nötr	Üzüntü
Öfke	11	1	0	1	0	0	0
Endişe/Korku	1	4	0	1	0	1	0
Sıkıntı	0	1	2	0	0	5	0
İğrenme	0	0	0	4	0	0	0
Mutluluk	2	0	0	0	5	0	0
Nötr	0	0	0	0	0	8	0
Üzüntü	0	0	1	0	0	1	5

Çizelge 5.3 DVM varsayılan parametreler için performans değerleri

	Duyarlık	Anma	F-ölçütü	Doğruluk	Hata Oranı
Öfke	0,7857	0,8462	0,8148	0,7222	0,2778
Endişe/Korku	0,6667	0,5714	0,6154		
Sıkıntı	0,6667	0,2500	0,3636		
İğrenme	0,6667	1,0000	0,8000		
Mutluluk	1,0000	0,7143	0,8333		
Nötr	0,5333	1,0000	0,6957		
Üzüntü	0,7143	1,0000	0,8333		

Hata matrisini incelediğimizde, sıkıntı duygusunun nötr duygusu ile karıştırıldığını görmekteyiz. Farklı parametreler ve çapraz doğrulama ile yapılan çalışmalar sonucu Çizelge 5.4'de gösterilmiştir. Maksimum doğruluk oranı veren parametreler ise Çizelge 5.5'deki değerler için ulaşılmıştır.

Çizelge 5.4 DVM parametre değışikliklerinde tanıma doğruluk oranları

Parametre	Ortalama Doğruluk (%)
$\gamma = 0.2$	80,7722
$\gamma = 0.5$	74,3816
$\gamma = 1.0$	50,4752
$c=1$	80,0070
$c=5$	82,0510
$c=10$	82,4074
$c=20$	82,4284
$\acute{e} = 0.02, r=0$	82,4284
$\acute{e} = 0.03, r=0$	82,2607
$\acute{e} = 0.04, r=0$	81,9043
$\acute{e} = 0.05, r=0$	82,2642
$\acute{e} = 0.10, r=0$	82,9769
$\acute{e} = 0.02, r=1$	82,9595
$\acute{e} = 0.03, r=1$	83,0084
$\acute{e} = 0.04, r=1$	83,1481
$\acute{e} = 0.05, r=1$	82,6205
$\acute{e} = 0.10, r=1$	82,2432

Çizelge 5.5 DVM için maksimum doğruluk veren parametre değeri

C	r	\acute{e}	γ
20	1	0,04	0,2

Çizelge 5.6 DVM için maksimum doğruluk hata matrisi

Referans Duygu	Tanınan Duygu						
	Öfke	Endişe / Korku	Sıkıntı	İğrenme	Mutluluk	Nötr	Üzüntü
Öfke	13	0	0	0	0	0	0
Endişe/Korku	0	6	0	1	0	0	0
Sıkıntı	0	0	7	0	0	1	0
İğrenme	1	0	0	3	0	0	0
Mutluluk	2	1	0	1	3	0	0
Nötr	0	0	2	0	0	6	0
Üzüntü	0	0	0	0	0	0	7

Çizelge 5.7 DVM için maksimum doğruluk performans değerleri

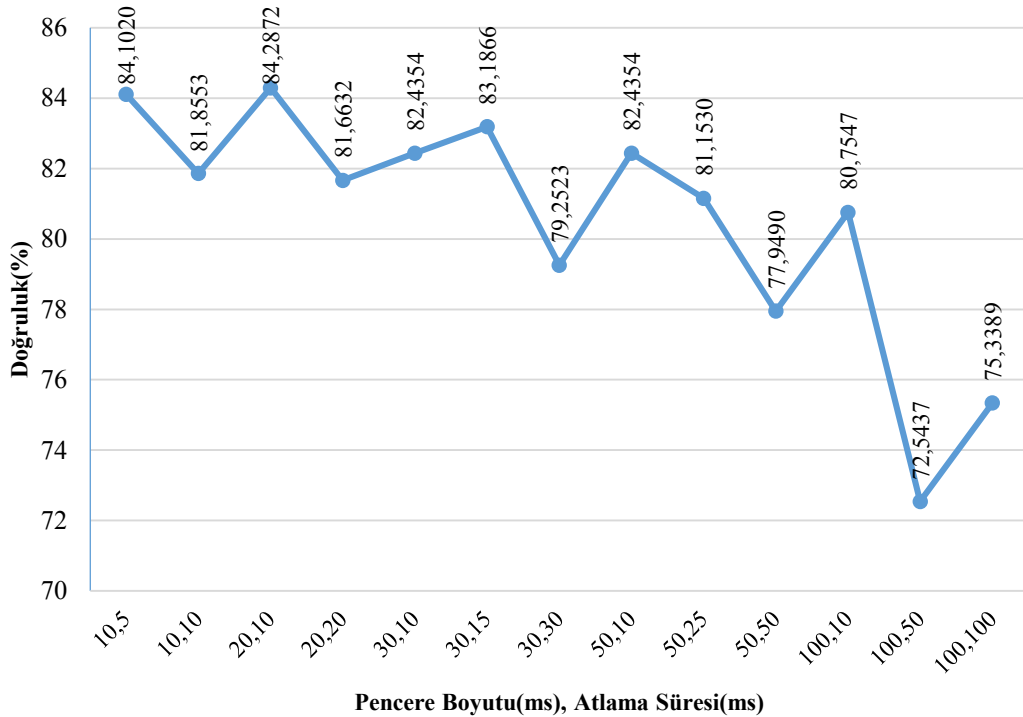
	Duyarlık	Anma	F-ölçütü	Doğruluk	Hata Oranı
Öfke	0,8667	1,0000	0,9286	0,8333	0,1667
Endişe/Korku	0,8571	0,8571	0,8571		
Sıkıntı	0,7778	0,8750	0,8235		
İğrenme	0,6000	0,7500	0,6667		
Mutluluk	1,0000	0,4286	0,6000		
Nötr	0,8571	0,7500	0,8000		
Üzüntü	1,0000	1,0000	1,0000		

DVM parametre optimizasyonu çalışmasında, varsayılan parametrelerde % 72,14 olan ortalama doğruluk oranı, optimize edilmiş parametreler ile %83,36'ya çıkmıştır. Sadece DVM parametre optimizasyonu yapılarak % 11,22'lik bir tanıma başarıımı artışı sağlanmış olması deneysel çalışmanın başarılı olduğunu göstermektedir.

5.1.2 Pencere ve atlama süresi analizi

DVM öğrenme algoritmasının pencere ve atlama süreleri ile ilişkisini deneyimlemek ve başarımı arttırabilmek için MFCC öznitelikleri ile çalışılmıştır. Emo-DB veri kümesinin 535 adet ses sinyalinin hepsi çalışmada kullanılmıştır. Hazırlanan veri kümesinin 10 da 9'u eğitim ve 10 da 1'i test için kullanılmıştır (10-kere çapraz doğrulama). Bölüm 5.1 çalışmasında elde edilmiş DVM parametre optimizasyonu sonuçları, bu çalışmanın girdisi olarak kullanılmış, DVM parametreleri olarak alınmıştır.

Çalışmaya 10 ms pencere boyutu, 5 ms atlama süresi ile başlanılmıştır. Bu kurulum, %84,10'luk bir ortalama doğruluk oranı vermiştir. Atlama süresini 10 ms'ye çıkarıp pencere boyutunu sabit tuttuğumuz da, doğruluk oranının azalarak %81,86'ya indiğini görülmüştür. Bu sonuçtan, üst üste binme olamayan pencerelerde, doğruluk yani duygu tanıma başarımın düştüğünü anlıyoruz. Kurulumu, pencere boyutunun 20 ms, atlama süresinin 10 ms olduğu duruma getirip ortalama doğruluk oranına baktığımızda, bu deneysel çalışmadaki en yüksek başarımlı değer olan %84,29'a çıktığını görmekteyiz.



Şekil 5.1 Pencere boyutu ve atlama sürelerinin ortalama doğruluk oranları

Açıkça görülmüştür ki, pencere boyutu ve atlama süresini eş zamanlı arttırmak, ortalama doğruluk oranı yani duygu tanıma başarımını sürekli bir şekilde azaltmaktadır (Şekil 5.1). En yüksek ortalama doğruluk oranına yakın hata matrisi (%84,91) Çizelge 5-8’de gösterilmektedir.

Çizelge 5.8 DVM için maksimum doğruluk veren pencere boyutu ve atlama süresi için hata matrisi

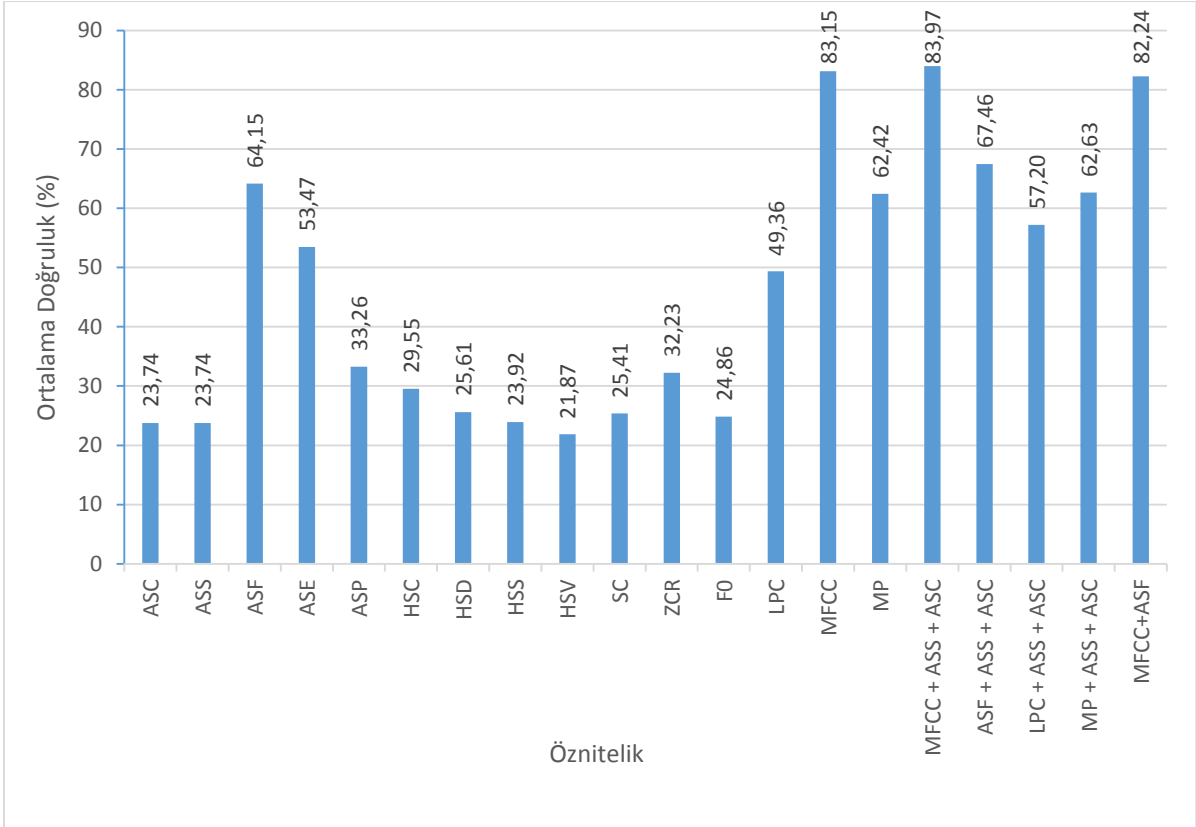
Referans Duygu	Tanınan Duygu						
	Öfke	Endişe / Korku	Sıkıntı	İğrenme	Mutluluk	Nötr	Üzüntü
Öfke	12	0	0	0	1	0	0
Endişe/Korku	0	6	0	0	1	0	0
Sıkıntı	0	0	6	0	0	2	0
İğrenme	0	0	0	3	1	0	0
Mutluluk	0	0	0	0	7	0	0
Nötr	0	1	2	0	0	5	0
Üzüntü	0	0	0	0	0	0	6

Çizelge 5.9 DVM için maksimum doğruluk veren pencere boyutu ve atlama süresi için performans değerleri

	Duyarlık	Anma	F-ölçütü	Doğruluk	Hata Oranı
Öfke	1,0000	0,9231	0,9600	0,8491	0,1509
Endişe/Korku	0,8571	0,8571	0,8571		
Sıkıntı	0,7500	0,7500	0,7500		
İğrenme	1,0000	0,7500	0,8571		
Mutluluk	0,7000	1,0000	0,8235		
Nötr	0,7143	0,6250	0,6667		
Üzüntü	1,0000	1,0000	1,0000		

5.1.3 Gürbüz öznitelik analizi

Çalışma 1 ve Çalışma 2 (Bölüm 5.1.1 ve 5.1.2)'de, sınıflandırıcı parametre ve pencere ile atlama sürelerinin optimizasyonu gerçekleştirilmiştir. Çalışmanın bu bölümünde 15 tekil, 5 bileşik öznitelik ile çalışılmıştır. Emo-DB üzerinde, özniteliklerin sonuç performans sonuçları Şekil 5-2'de gösterilmiştir.



Şekil 5.2 Emo-DB üzerinde özniteliklerin ortalama doğrulukları

Şekil 5.2, tekil öznitelikler ile bazı özniteliklerin kombinasyonlarının (bileşik öznitelikler) duygu tanıma sonuçlarını göstermektedir. Sonuçlardan anlaşıldığı üzere, tekil ve doğrusal öznitelikler (örneğin: ASC, ASS, HSC, HSD, HSS, HSV, SC, ZCR, ve F0), duygu tanıma için yeterli katkıyı yapmamaktadırlar. Vektörel olmasına karşın, ASP özniteliği de tekil ve doğrusal öznitelikler gibi zayıf performans vermektedir.

MFCC öznitelik kümesi, %83,15 ile en yüksek ikinci ortalama tanıma doğruluğunu vermiştir. Bu sonuç beklenen bir sonuçtur. MFCC özniteliği, insan algısına en yakın ve sesin kısa dönem güç spektrumunun yakalanmasında gürbüz sonuçlar vermektedir. Diğer tekil özniteliklerden MPEG-7 ASF, MP, MPEG-7 ASE, ve LPC de MFCC özniteliği ile karşılaştırıldığını iyi performans vermişlerdir. Bu tekil özniteliklerin prosodik öznitelikler ile (ASS ve ASC) kombine edilmesi ile en MFCC+ASS+ASC bileşik özniteliği, %83,97 ile en yüksek ortalama tanıma doğruluğuna ulaşmıştır. Buna karşın, bu bileşik öznitelik ile tekil MFCC özniteliğinin başarımı arasındaki fark %0,82 gibi düşük bir değerdir. Dolayısı ile bileşik özniteliğin getirdiği karmaşıklığa girmemek için, MFCC tekil özniteliğinin tek başına kullanılması tavsiye edilebilir.

En yüksek 5 ortalama doğruluk oranına sahip MFCC+ASS+ASC, MFCC, MFCC+ASF, ASF+ASS+ASC, ASF özniteliklerinin, ortalama doğruluk oranlarına en yakın hata matrisleri Çizelge 5-10,5-12,5-14,5-16'da verilmiştir. Diğer hata matrisleri Ek-1'de sunulmuştur.

Çizelge 5.10 MFCC+ASS+ASC bileşik özneliği hata matrisi

Referans Duygu	Tanınan Duygu						
	Öfke	Endişe / Korku	Sıkıntı	İğrenme	Mutluluk	Nötr	Üzüntü
Öfke	12	0	0	0	0	0	0
Endişe/Korku	1	6	0	0	0	0	0
Sıkıntı	0	0	6	0	0	2	0
İğrenme	0	1	0	4	0	0	0
Mutluluk	2	0	0	0	5	0	0
Nötr	0	0	1	0	1	6	0
Üzüntü	0	0	0	0	0	0	6

Çizelge 5.11 MFCC+ASS+ASC bileşik özneliği performans değerleri

	Duyarlık	Anma	F-ölçütü	Doğruluk	Hata Oranı
Öfke	0,8000	1,0000	0,9231	0,8491	0,1509
Endişe/Korku	0,8571	0,8571	0,8571		
Sıkıntı	0,8571	0,7500	0,8000		
İğrenme	1,0000	0,8000	0,8889		
Mutluluk	0,8333	0,7143	0,7692		
Nötr	0,7500	0,7500	0,7500		
Üzüntü	1,0000	1,0000	1,0000		

Çizelge 5.12 MFCC özniteliği hata matrisi

Referans Duygu	Tanınan Duygu						
	Öfke	Endişe / Korku	Sıkıntı	İğrenme	Mutluluk	Nötr	Üzüntü
Öfke	13	0	0	0	0	0	0
Endişe/Korku	0	6	0	0	1	0	0
Sıkıntı	0	0	5	1	0	2	0
İğrenme	0	0	0	4	0	0	0
Mutluluk	2	0	0	0	5	0	0
Nötr	0	0	3	0	0	5	0
Üzüntü	0	0	0	0	0	0	6

Çizelge 5.13 MFCC özniteliği performans değerleri

	Duyarlık	Anma	F-ölçütü	Doğruluk	Hata Oranı
Öfke	0,8667	1,0000	0,9286	0,8302	0,1698
Endişe/Korku	1,0000	0,8571	0,9231		
Sıkıntı	0,6250	0,6250	0,6250		
İğrenme	0,8000	1,0000	0,8889		
Mutluluk	0,8333	0,7143	0,7692		
Nötr	0,7143	0,6250	0,6667		
Üzüntü	1,0000	1,0000	1,0000		

Çizelge 5.14 MFCC+ASF bileşik özneliği hata matrisi

Referans Duygu	Tanınan Duygu						
	Öfke	Endişe / Korku	Sıkıntı	İğrenme	Mutluluk	Nötr	Üzüntü
Öfke	9	0	1	0	2	0	0
Endişe/Korku	2	5	0	0	0	0	0
Sıkıntı	0	0	7	0	0	1	0
İğrenme	0	0	0	5	0	0	0
Mutluluk	0	0	0	0	7	0	0
Nötr	0	1	0	0	0	7	0
Üzüntü	0	0	1	0	0	2	3

Çizelge 5.15 MFCC+ASF bileşik özneliği performans değerleri

	Duyarlık	Anma	F-ölçütü	Doğruluk	Hata Oranı
Öfke	0,8182	0,7500	0,7826	0,8113	0,1887
Endişe/Korku	0,8333	0,7143	0,7692		
Sıkıntı	0,7778	0,8750	0,8235		
İğrenme	1,0000	1,0000	1,0000		
Mutluluk	0,7778	0,7143	0,7447		
Nötr	0,7500	1,0000	0,8571		
Üzüntü	1,0000	0,5000	0,6667		

Çizelge 5.16 ASF+ASS+ASC bileşik özneliği hata matrisi

Referans Duygu	Tanınan Duygu						
	Öfke	Endişe / Korku	Sıkıntı	İğrenme	Mutluluk	Nötr	Üzüntü
Öfke	11	1	0	0	1	0	0
Endişe/Korku	1	5	0	1	0	0	0
Sıkıntı	0	0	6	0	0	2	0
İğrenme	0	0	0	4	0	0	0
Mutluluk	4	1	0	0	2	1	0
Nötr	0	0	4	0	0	3	0
Üzüntü	0	0	2	0	0	0	5

Çizelge 5.17 ASF+ASS+ASC bileşik özneliği performans değerleri

	Duyarlık	Anma	F-ölçütü	Doğruluk	Hata Oranı
Öfke	0,6875	0,8462	0,7586	0,6667	0,3333
Endişe/Korku	0,7143	0,7143	0,7143		
Sıkıntı	0,5000	0,7500	0,6000		
İğrenme	0,8000	1,0000	0,8889		
Mutluluk	0,6667	0,2500	0,0976		
Nötr	0,5000	0,4286	0,4615		
Üzüntü	1,0000	0,7143	0,8333		

Çizelge 5.18 ASF özniteliği hata matrisi

Referans Duygu	Tanınan Duygu						
	Öfke	Endişe / Korku	Sıkıntı	İğrenme	Mutluluk	Nötr	Üzüntü
Öfke	11	0	2	0	0	0	0
Endişe/Korku	1	1	1	0	3	0	1
Sıkıntı	0	0	7	0	0	1	0
İğrenme	0	0	0	4	0	0	0
Mutluluk	2	1	0	0	4	0	0
Nötr	3	1	1	0	0	3	0
Üzüntü	0	0	2	0	0	0	5

Çizelge 5.19 ASF özniteliği performans değerleri

	Duyarlık	Anma	F-ölçütü	Doğruluk	Hata Oranı
Öfke	0,6471	0,8462	0,7333	0,6481	0,3519
Endişe/Korku	0,3333	0,1429	0,2000		
Sıkıntı	0,5385	0,8750	0,6667		
İğrenme	1,0000	1,0000	1,0000		
Mutluluk	0,5714	0,5714	0,5714		
Nötr	0,7500	0,3750	0,5000		
Üzüntü	0,8333	0,7143	0,7692		

MFCC özniteliği için Çizelge 5.2 ile 5.8 i karşılaştırdığımızda, nötr duygusunun tanınma oranının düşmesine karşın, diğer 5 duygunun tanınma oranında, ciddi bir artış gözlemlenmiştir. Nötr duygusunun tanınmasındaki düşüş, nötrün sıkıntı ile karıştırılmasından kaynaklanmaktadır. Nötrde düşüş, sıkıntıda artışla aynı orandadır. Birbirlerinin tanınma oranlarını etkilemektedirler. Genel olarak nötr duygusu ile sıkıntı duygusu benzer duygular oldukları için karıştırılmaktadırlar. Karakteristik olarak benzerlikler gösteren diğer iki duygu olan mutluluk ve öfkenin de, en çok karıştırılan duygular olduğu anlaşılmaktadır.

En yüksek ortalama doğruluk oranına sahip MFCC+ASS+ASC bileşik özniteliğini Çizelge 5.10'da inceleyebiliriz. Çizelge verisinden MFCC özniteliği ile çok benzer sonuçlar verdiği, doğruluk oranındaki küçük farktan da anlaşılmaktadır. Bu küçük fark, sıkıntı ve nötr duygularındaki 1'er fazla doğru tanımadan kaynaklanmaktadır. Ancak bileşik özniteliğin, hem öğrenme algoritmasına hem de fiziksel bilgisayar donanımına fazladan yük getirdiği dikkate alınmalıdır. Başarımdaki bu ufak artışa karşın, bileşik öznitelik kullanımının getirdiği yükü üstlenmeye gerek olmadığı düşünülmektedir.

İki yüksek başarımlı ve benzer öznitelikler olan MFCC ve ASF özniteliğinin bileşik özniteliğini olan MFCC+ASF bileşke özniteliği, Çizelge 5.14'de inceleyebiliriz. Öfkenin mutluluk ile, korkunun endişe ile ve üzüntünün nötr ile karıştırıldığı görülmektedir. Bileşik öznitelik olmasına karşın MFCC tekil özniteliğinden başarımlı daha düşüktür. Benzer öznitelikler oldukları için, sonucun MFCC sonuçlarına yakın çıkması beklenilmekte idi.

Çizelge 5.16'da ise üçüncü en yüksek tanıma oranına sahip ASF+ASS+ASC bileşik özniteliği bulunmaktadır. Önceki iki çizelge incelememiz ile aynı duygusal tanımlamaları yapabilmekle birlikte, daha önce karıştırılan duygular olduğunu anladığımız sıkıntı-nötr ile mutluluk-öfke duygularındaki başarımlı düşüğünü görmekteyiz.

ASF özniteliğinde, endişe/korku ile mutluluğun, nötr ile de öfkenin karıştırıldığını görmekteyiz (Çizelge 5.18). Özniteliğin bu duygu durumlarını ayırt etmekte zayıf kaldığını söyleyebiliriz.

5.2 Duygu Geri Erişim Deneyleri

5.2.1 Veri kümesi

Duygu bilgisi çıkarımı için oluşturulan veri kümesinin, tanıma doğruluğunu belirlemek amacıyla veri kümesi 1 erkek 1 kadın, 2 kişi tarafından etiketlenmiştir. Klipler, ana ve alt olmak üzere ayrılmıştır. Ana klipler, ortalama 20 saniye uzunluğunda yerli tv dizilerinden alınmış ses klipleri olup, her bir ana ses klipi 1'er saniyelik alt kliplere ayrılmış ve ayrı ayrı etiketlenmiştir. Her bir alt klipin, 20 ms pencere boyutu, 10 ms atlama süresi ile MFCC öznitelikleri çıkarılmıştır. Çıkarılan öznitelikler, DVM öğrenme algoritması ile eğitilmiş ve öğrenme modelleri oluşturulmuştur. 10 kere çapraz doğrulama yöntemi ile sistem test edilmiş, tanıma doğruluğu, duyarlık, anma, f-ölçütü ve hata oranı çıkarılmıştır.

Çizelge 5.20 TV dizilerinden oluşturulmuş veri kümesi için hata matrisi

Referans Duygu	Tanınan Duygu						
	Öfke	Endişe / Korku	Sıkıntı	İğrenme	Mutluluk	Nötr	Üzüntü
Öfke	7	0	0	0	0	0	0
Endişe/Korku	0	0	0	0	1	0	0
Sıkıntı	1	0	1	0	1	1	0
İğrenme	0	0	0	0	1	1	0
Mutluluk	1	0	0	0	6	2	2
Nötr	0	0	0	0	0	11	0
Üzüntü	0	0	0	0	1	0	4

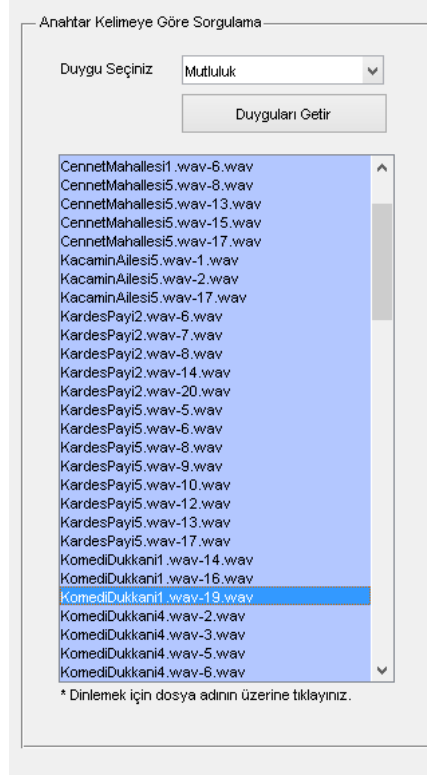
Çizelge 5.21 TV dizilerinden oluşturulmuş veri kümesi için performans değerleri

	Duyarlık	Anma	F-ölçütü	Doğruluk	Hata Oranı
Öfke	0,7778	1,0000	0,8750	0,7073	0,2927
Endişe/Korku	-	0,0000	-		
Sıkıntı	1,0000	0,2500	0,4000		
İğrenme	-	0,0000	-		
Mutluluk	0,6000	0,5455	0,5714		
Nötr	0,7333	1,0000	0,8462		
Üzüntü	0,6667	0,8000	0,7273		

Oluşturulan veri kümesinin % 70,73 lük bir ortalama tanıma doğruluğu verdiği görülmüştür. Çizelge 5.20 ve Çizelge 5.21’de hata matrisi ve performans durumu incelenebilmektedir. Sonuçları incelediğimizde, mutluluk duygusunun, nötr ve üzüntü ile, sıkıntının, öfke, mutluluk ve nötr, üzüntünün ise mutluluk ile yüksek olmayan oranlarda karıştırıldığını görmekteyiz. Dizi kayıtlarının genellikle, temiz bir stüdyo ortamında değil de açık stüdyolarda (örn: şehir içindeki bir herkese açık bir üniversitenin girişi) çekiliyor olması arka plan gürültüsünü arttırmaktadır. Ayrıca klip seçimlerinde olmamasına dikkat edilmesine karşın arka plan müziği gürültüyü arttırmaktadır. Başarımının düşmesindeki önemli bir etken gürültü olabilmektedir. Buna karşın genel tanıma doğruluğunu Emo-DB genel tanıma doğruluğuna %10 civarında bir fark ile görece yakın sayılabilir.

5.2.2 Anahtar kelime sorgusu

Veri kümesi üzerinde etiketlenmiş olarak sisteme kaydedilmiş ses kliplerinin duyguları, seçilen anahtar kelimeye göre metin tabanlı olarak sorgulanmaktadır. Sorgulama arayüzü, Şekil 5.3’de verilmiştir.



Şekil 5.3 Anahtar kelimeye göre sorgu

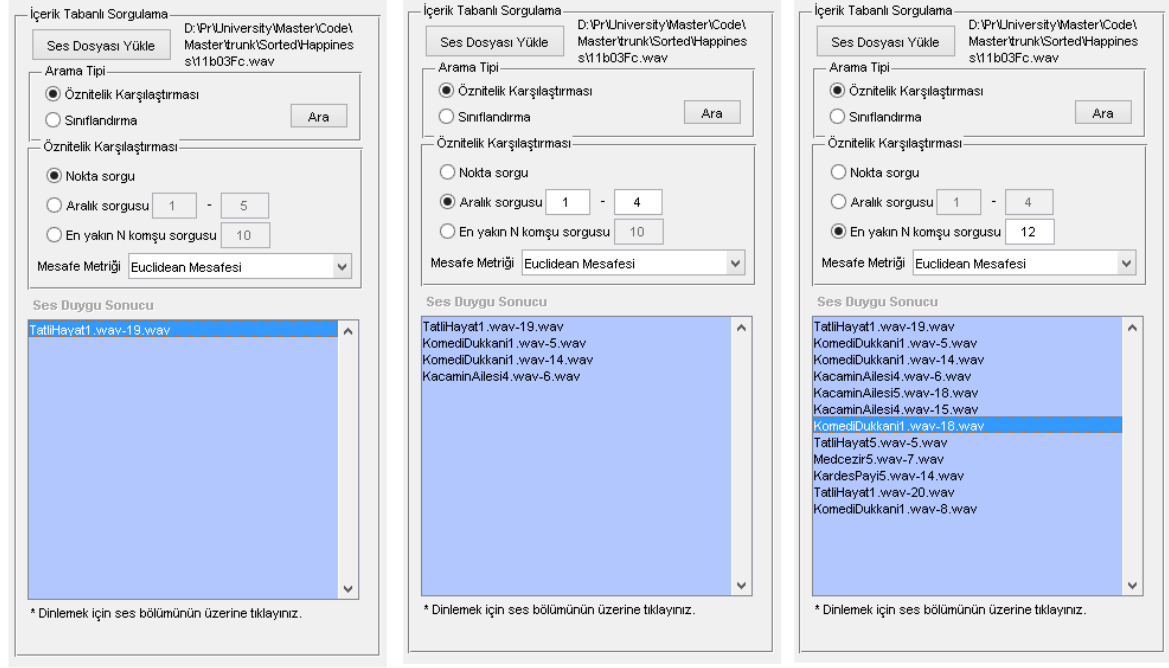
Sorgulama, anahtar kelimelerin tam eşleşmesi ile gerçekleştirildiğinden, geri erişim performansı ses kliplerinin doğru etiketlenme başarımı ile sınırlıdır.

5.2.3 İçerik tabanlı sorgu

Sisteme verilen duygusu bilinmeyen ses klibinden, öncelikle 20 ms pencere, 10 ms öteleme süreleri için MFCC öznitelikleri çıkarılmıştır. Çıkarılan 20 adet öznitelik 2 farklı sorgu türünde iki farklı şekilde kullanılmaktadır.

5.2.3.1 Öznitelik tabanlı

MFCC öznitelikleri çıkarılan ses klibi, var olan veri kümesi üzerindeki kliplerin MFCC öznitelikleri ile kullanıcının seçtiği mesafe metriği kullanılarak hesaplanan mesafeye göre karşılaştırılır. Sorgu türüne göre (nokta, aralık, en yakın n komşu) seçilen sonuçlar, sıralanmış yakınlık listesinden listelenir. Şekil 5.4'de sorgulama arayüzü incelenebilmektedir.



Şekil 5.4 İçerik tabanlı sorgulama - öznitelik karşılaştırması

Bu geri erişim yöntemi için yapılan testlerin sonucu Çizelge 5.22 de verilmiştir.

Çizelge 5.22 Öznitelik tabanlı sorgulama için performans değerleri

	Sorgu1	Sorgu2	Sorgu3	Sorgu4	Sorgu5	ODÖ
Öfke	0,95	0,79	0,83	0,86	0,78	0,84
Endişe/Korku	1,00	1,00	1,00	1,00	0,70	0,94
Sıkıntı	1,00	0,93	0,75	0,83	0,75	0,85
İğrenme	1,00	1,00	1,00	0,80	1,00	0,96
Mutluluk	0,65	1,00	1,00	0,86	0,81	0,86
Nötr	0,99	0,91	0,86	0,75	1,00	0,90
Üzüntü	0,72	0,61	0,65	0,65	0,97	0,72

5.2.3.2 Sınıflandırma tabanlı

İçerik Tabanlı Sorgulama

Ses Dosyası Yükle D:\Pr\University\Master\Code\Master\trunk\Sorted\Happines s\11b03Fc.wav

Arama Tipi

Öznitelik Karşılaştırması

Sınıflandırma Ara

Öznitelik Karşılaştırması

Nokta sorgu

Aralık sorgusu 1 - 4

En yakın N komşu sorgusu 12

Mesafe Metriği Euclidean Mesafesi

Ses Duygu Sonucu M Mutluluk (%50)

AltıManti1 .wav-10 .wav	
AltıManti3 .wav-2 .wav	
AltıManti3 .wav-3 .wav	
AltıManti3 .wav-4 .wav	
AltıManti3 .wav-6 .wav	
CennetMahallesi1 .wav-2 .wav	
CennetMahallesi1 .wav-6 .wav	
CennetMahallesi5 .wav-6 .wav	
CennetMahallesi5 .wav-13 .wav	
CennetMahallesi5 .wav-15 .wav	
CennetMahallesi5 .wav-17 .wav	
KacaminAilesi5 .wav-1 .wav	
KacaminAilesi5 .wav-2 .wav	
KacaminAilesi5 .wav-17 .wav	
KardesPayi2 .wav-6 .wav	
KardesPayi2 .wav-7 .wav	
KardesPavi2 .wav-8 .wav	

* Dinlemek için ses bölümünün üzerine tıklayınız.

Şekil 5.5 İçerik tabanlı sorgulama – sınıflandırma

Bu geri erişim yönteminde, sınıflandırma sonuçları kullanıldığından, geri erişim başarımı Bölüm 4’de sunulan sınıflandırma başarımı ile aynıdır.

5.2.4 Zamansal sorgu

Zamansal sorgulama yapılırken precedes, meets, equals, met by, preceded by operatörleri ile arama yapıldığında, her bir ana klibin içindeki alt klipler kendi aralarında aramaya tabi tutulmaktadır. Starts, during ve finishes operatörleri ise, alt klip ve ana klip arasında uygulanmaktadır. Overlaps, finished by, contains, started by ve overlaped by operatörleri, verimizin ayrık duygular barındırmasından dolayı işlevsizdirler. Şekil 5.6’da zamansal sorgulama arayüzü incelenebilir.

Zamansal Sorgulama

1. Duyguyu Seçiniz Mutluluk

Operatör Seçiniz Precedes G

2. Duyguyu Seçiniz Üzüntü

Ara

Sorgu Sonucu Bulunan Klip Sayısı: 129

AtiManti3.wav-2.wav :2-7sn
AtiManti3.wav-2.wav :2-9sn
AtiManti3.wav-2.wav :2-12sn
AtiManti3.wav-2.wav :2-15sn
AtiManti3.wav-3.wav :3-7sn
AtiManti3.wav-3.wav :3-9sn
AtiManti3.wav-3.wav :3-12sn
AtiManti3.wav-3.wav :3-15sn
AtiManti3.wav-4.wav :4-7sn
AtiManti3.wav-4.wav :4-9sn
AtiManti3.wav-4.wav :4-12sn
AtiManti3.wav-4.wav :4-15sn
AtiManti3.wav-6.wav :6-7sn
AtiManti3.wav-6.wav :6-9sn
AtiManti3.wav-6.wav :6-12sn
AtiManti3.wav-6.wav :6-15sn
CennetMahallesi1.wav-2.wav :2-12sn
CennetMahallesi1.wav-2.wav :2-15sn
CennetMahallesi1.wav-6.wav :6-12sn
CennetMahallesi1.wav-6.wav :6-15sn
CennetMahallesi5.wav-8.wav :8-16sn
CennetMahallesi5.wav-8.wav :8-19sn
CennetMahallesi5.wav-8.wav :8-21sn
CennetMahallesi5.wav-13.wav :13-16sn
CennetMahallesi5.wav-13.wav :13-19sn
CennetMahallesi5.wav-13.wav :13-21sn
CennetMahallesi5.wav-15.wav :15-16sn

* Dinlemek için dosya adının üzerine tıklayınız.

Şekil 5.6 Zamansal sorgulama

Bu yöntemin geri erişim performansı, kelime ve sınıflandırma tabanlı yöntemlere benzer olarak ses kliplerinin doğru etiketleme başarımı ile sınırlıdır.

6. SONUÇ

Bu tez çalışmasında, ses duygu tanıma ve geri erişimi için bir uygulama çatısı geliştirilmiştir. Bu kapsamda, gürbüz öznitelik tespiti, öğrenme algoritması parameter optimizasyonu ve ses içerik analizi konularında ampirik yöntemler önerilmiştir.

Çalışmada yedi (7) ayrıık duygu durumu (mutluluk, üzüntü, şaşkınlık, öfke, tiksinti ve korku) ele alınmıştır.. Uygulama altyapısı, verilen bir ses klibinin, çalışmada tanıtılan düşük seviye öznitelikler ile tahminlenmesine dayanmaktadır. Deneysel çalışmalarda, Emo-DB veri kümesi üzerinde, 7 ayrıık duygu durumu için, 20 ses özniteliği (15 tekil, 5 bileşik) üzerinden, DVM sınıflandırma algoritması kullanılarak duygu durumları sezimlenmiştir. Deneysel çalışmamız sonucunda, ses sinyalinde duygu tanıma, öznitelik çıkarımı ve sınıflandırıcı optimizasyonu sonucunda, %11'in üzerinde ciddi bir başarıım artışı gözlenmiştir.

Optimizasyon çalışmalarının çıktıları ile gerçekleştirilen DVM test sonuçlarını incelediğimizde MFCC, MPEG-7 ASF, LPC ve MP özniteliklerinin, birbirinin yerine kullanılabilecekleri ve ses sinyalinde duygu tanıma başarılı oldukları görülmüştür. Bu 4 öznitelik vektörel öznitelikler olup, vektörel özniteliklerin, doğrusal özniteliklere göre başarımları daha yüksektir. Diğer ses öznitelikleri, tek başına kullanıldığında, incelenen veri kümesi üzerinde düşük başarımlar vermiştir. Bu öznitelikler, uygulama gereksinimlerine bağlı olarak, güçlü spektral özniteliklerle (MFCC, ASF) birlikte kullanılabilir.

Duygu geri erişimi çalışmasında, yerli TV dizilerinden yeni bir veri kümesi yaratılmıştır. Veri kümesinin test sonuçlarını incelediğimizde, %70,73'lük görece yüksek bir tanıma başarımsı elde edilmiştir. Yeni yaratılmış veri kümesi, Emo-DB'de olduğu gibi, aktörlerin duyguları canlandırmaları şeklinde gerçekleştirilmediği, dizi filmlerden alındığı için, ayrıık duyguların yakalanabilmesi görece daha zordur. Ayrıca dizi filmlerdeki geri plan gürültüsü, diyalogların üst üste binebilmesi ve sessizlikler de dezavantaj olabilmektedir. Bunlara karşın Emo-DB ile aralarındaki %10 civarındaki başarıım farkı, başarılı bir sonuç olarak kabul edilebilir. Üç (3) farklı sorgu türüne (anahtar kelime, içerik tabanlı ve zamansal) göre ve iki farklı yaklaşıma göre

(sınıflandırma ve en yakın komşu) yapılan sorgulamalar ile deneysel örnek bir çalışma gerçekleştirilmiştir.

Gelecek çalışmalarda, süreklilik içeren ses sinyalindeki duygu durumlarını tanıma üzerine çalışılması düşünülmektedir. Ayrıca, bulanık duygu durumları ile çalışılması hedeflenmektedir. Bazı duygu durumları öznedir ve ses kliplerinden anlaşılması insanođlu için bile zordur. Bu nedenle, yaratılan yeni veri kümesinin genişletilmesi ve daha kalabalık bir grup tarafından etiketlenmesi geri erişim performansını artıracaktır.

7. REFERANSLAR

- [1] Tawari A., Trivedi Mohan., Speech Emotion Analysis: Exploring the Role of Context, *IEEE Transactions On Multimedia*, 12(16):502-509, 2010.
- [2] Mion L., De Poli G., Score-Independent Audio Features for Description of Music Expression, *IEEE Transactions On Audio, Speech, And Language Processing*, 16(2):458-466, 2008.
- [3] Kim J., Andre E., Emotion Recognition Based on Physiological Changes in Music Listening, *IEEE Transactions On Pattern Analysis And Machine Intelligence*, 30(12):2067-2083, 2008.
- [4] Muhammad G., Alotaibi Y., Alsulaiman, M., Environment Recognition Using Selected MPEG-7 Audio Features and Mel-Frequency Cepstral Coefficients, 2010 Fifth International Conference on Digital Telecommunications, 11-16, 2010.
- [5] Zeng Z., Pantic M., Roisman G., Huang T.S., A Survey of Affect Recognition Methods: Audio, Visual, and Spontaneous Expressions, *IEEE Transactions On Pattern Analysis And Machine Intelligence*, 31(1):39-58, 2009.
- [6] Mower E., Mataric, M., Naryanan S., A Framework for Automatic Human Emotion Classification Using Emotion Profiles, *IEEE Transactions On Audio, Speech, And Language Processing*, 19(5):1057-1070, 2011.
- [7] Wang Y., Guan L., Venetsanopoulos A., Kernel Cross-Modal Factor Analysis for Information Fusion With Application to Bimodal Emotion Recognition, *IEEE Transactions On Multimedia*, 14(3):597-607, 2012.
- [8] Jia J., Zhang S., Meng F., Chai L. , Emotional Audio-Visual Speech Synthesis Based on PAD *IEEE Transactions On Audio, Speech, And Language Processing*, 19(3):570-582, 2011.
- [9] Soleymani M., Lichtenauer J., Pun,, T., Pantic, M., A Multimodal Database for Affect Recognition and Implicit Tagging, *IEEE Transactions On Affective Computing*, Vol. 3, No. 1:42-55, 2012.
- [10] Mower, E., Mataric, M. and Naryanan, S., Human Perception of Audio-Visual Synthetic Character Emotion Expression in the Presence of Ambiguous and Conflicting Information, *IEEE Transactions On Multimedia.*, 11(5):843-855, 2009.
- [11] Wang, Y., Guan, L., Recognizing Human Emotional State From Audiovisual Signals, *IEEE Transactions On Multimedia*, 10(4):659-668, 2008.
- [12] Sezgin C., Günsel B., Hacıoğlu C. , Audio Emotion Recognition By Perceptual Features, *Signal Processing and Communications Applications Conference (SIU)*, pp.1,4, 18-20, 2012.
- [13] Malandrakis N., Potamianos A., Evangelopoulos G., Zlatintsi A., A Supervised Approach To Movie Emotion Tracking , *Acoustics Speech and Signal Processing (ICASSP)*, pp.2376,2379, 2011.

- [14] Schuller B., Hantke S., Weninger F., Han W., Zhang Z., Narayanan S., Automatic Recognition Of Emotion Evoked By General Sound Events, Acoustics Speech and Signal Processing (ICASSP), pp.341,344 2012.
- [15] Sezgin C., Günsel B., Kurt G., A Novel Perceptual Feature Set for Audio Emotion Recognition, Automatic Face & Gesture Recognition and Workshops, pp.780,785,2012
- [16] Xie Z., Guan L., Multimodal Information Fusion of Audio Emotion Recognition Based on Kernel Entropy Component Analysis, Multimedia (ISM), pp.1,8,2012.
- [17] Lampropoulos A., Tsihrintzis G., Evaluation of MPEG-7 Descriptors for Speech Emotional Recognition, 2012 Eighth International Conference on Intelligent Information Hiding and Multimedia Signal Processing, pp.98,101,2012.
- [18] C. J. C. Burges, A tutorial on support vector machines for pattern recognition, Data Mining Knowledge Discovery, vol. 2, no. 2, pp. 1-47, 1998.
- [19] K. Mineichi, S. Petr, Comparison of classifier-specific feature selection algorithms, Springer-verlag Berlin Heidelberg, 2000:677-686.
- [20] Hsu C., Chang C. , Lin C., A Practical Guide to Support Vector Classification, National Taiwan University, 2010.
- [21] Vapnik V., The Nature of Statistical Learning Theory, Springer-Verlag, New York, 1995.
- [22] Paleari, M., Chellali, R., Huet, B., Features for multimodal emotion recognition: An extensive study, Cybernetics and Intelligent Systems (CIS), 2010 IEEE Conference, pp.90,95, 28-30, 2010.
- [23] Li M., Liu Y., Wang J., A new parameter optimization algorithm of SVM, Advanced Intelligence and Awareness Internet (AIAI 2011), 2011 International Conference, pp.174,178, 28-30, 2011.
- [24] Yang, Y., Chen, H, Ranking-Based Emotion Recognition for Music Organization and Retrieval, IEEE Transactions On Audio, Speech, And Language Processing, Vol.19, No:4, 2011.
- [25] Chang C.C., Lin C.J., LIBSVM: A library for support vector machines. ACM Transactions on Intelligent Systems and Technology, 2:27:1—27:27, 2011.
- [26] Burkhardt F., Paeschke A., Rolfes M., Sendlmeier W., Weiss ., A Database of German Emotional Speech (EMO-DB), Proc. Interspeech, 2005.
- [27] Gunes H., Pantic M., Automatic, Dimensional and Continuous Emotion Recognition, Int'l Journal of Synthetic Emotion, 1(1): pp. 68 - 99, 2010.
- [28] Metallinou A., Wollmer M., Katsamanis A., Eyben F., Schuller B., Narayanan S., Context-Sensitive Learning for Enhanced Audiovisual Emotion Classification, IEEE Transactions On Affective Computing, Vol. 3, No. 2, 2012.

- [29] Bellegarda J., A Data-Driven Affective Analysis Framework Toward Naturally Expressive Speech Synthesis, *IEEE Transactions On Audio, Speech, And Language Processing*, Vol. 19, No. 5, 2011.
- [30] Yang Y.H., Lin Y.C., Su Y., Chen H., A Regression Approach to Music Emotion Recognition, *IEEE Transactions On Audio, Speech, And Language Processing*, Vol. 16, No. 2, 2008.
- [31] Jia J, Zhang S., Meng F., Wang Y., Cia L., Emotional Audio-Visual Speech Synthesis Based on PAD, *IEEE Transactions On Audio, Speech, And Language Processing*, Vol. 19, No. 3, 2011.
- [32] DeGangi GA, A Terapist's Guide to Assessment and Treatment, 2-31,120-159. London: Academic Press, 2000.
- [33] Malatesta CZ., Culver C., Tesman JR., Shepard B., The development of emotion expression during the first two years of life, *Monographs of the Society for Research in Child Development*, 1989.
- [34] Cevizci A., *Paradigma Felsefe Sözlüğü*, Paradigma Yayınları, 3.Basım, p.269, 1999.
- [35] Akarsu B., *Felsefe Terimleri Sözlüğü*, İnkılap Kitabevi, 9. Baskı, s.60, 1998.
- [36] Yelkikalan N., 21. Yüzyılda Girişimcinin Yeni Özelliği: Duygusal Zeka, *Girişimcilik ve Kalkınma Dergisi*, C. 1, S. 2, 2006.
- [37] Darwin C., *The Expression of the Emotions in Man and Animals*,1872.
- [38] Ekman P. Friesen, W.V., Constants across cultures in the face and emotion, *Journal of Personality and Social Psychology* 17: 124–129, 1971.
- [39] Ekman P., *Basic Emotions*, *Handbook of Cognition and Emotion*, Sussex, UK: John Wiley & Sons, 1999.
- [40] Plutchik R., *Emotion: Theory, research, and experience: Vol. 1. Theories of emotion*, New York: Academic, 1980.
- [41] Bozoklu Ç., Alkibay S., Nöropazarlama kapsamında tv reklamlarına yönelik duygulanım tepkilerinin değerlendirilmesi: yüz kasları analizi ile anket yönteminin karşılaştırılması, *19.Ulusal Pazarlama Kongresi Bildiri Kitabı*, S 143-144, 2014.
- [42] Takahashi K., Remarks on Computational Emotion Recognition from Vital Information, *Proceedings of the 6th International Symposium on Image and Signal Processing and Analysis*, 299-304, 2009.
- [43] Vapnik, V., *The Nature of Statistical Learning Theory*, Second Edition, Springer Science, 2000.
- [44] Akkoca, B. S., *Durgun Görüntülerden Yüz İfadelerinin Tanınması*, İstanbul Teknik Üniversitesi, Yüksek Lisans Tezi, 2014.
- [45] Cortes, C., Vapnik, V. N., Support Vector Networks, *Machine Learning*, 20(3):273-297, 1995.

- [46] Yıldırım, S., Arıza Teşhisinde Destek Vektör Makinelerinin Kullanımı, Yüksek Lisans Tezi, Fırat Üniversitesi, s113, 2006.
- [47] Milgram J., Cheriet M., Sabourin R., Speeding Up the Decision Making of Support Vector Classifiers, Proceedings of the 9th Int'l Workshop on Frontiers in Handwriting Recognition, 2004.
- [48] Price D., Knerr S., Personnaz L., Dreyfus G., Pairwise Neural Network Classifiers with Probabilistic Outputs, Neural Information Processing Systems, MIT Press, s1109-1116, 1995.
- [49] Baghaee S., Gürbüz S.Z., Uysal-Bıyıkoğlu E., Orthogonal Matching Pursuit For Ferromagnetic Target Localization And Identification, Signal Processing and Communications Applications Conference (SIU), 2012.
- [50] Rao K.S., Koolagudi S.G., Robust Emotion Recognition using Pitch Synchronous and Sub-syllabic Spectral Features, s17-45, 2013.
- [51] Nabiyev V.V., Yücesoy E., VQ Yöntemiyle Konuşmacı Cinsiyetinin Belirlenmesi, Turkish Journal of Computer and Mathematics Education, Vol.1 No.1, s 35-47, 2009.
- [52] Internet: http://scikit-learn.org/stable/auto_examples/svm/plot_rbf_parameters.html.
- [53] Li C.H., Lin C., Ku B., Chu H., An Automatic Method For Selecting The Parameter Of The RBF Kernel Function To Support Vector Machines, Geoscience and Remote Sensing Symposium (IGARSS), s 836-839, 2010.
- [54] Ateş Y., Ders Notları, Yıldız Üniversitesi, Enformatik Bölümü, 2004.
- [55] Okuyucu C., Sert M., Yazıcı A., Audio Feature and Classifier Analysis for Efficient Recognition of Environmental Sounds, In Proceedings of IEEE International Symposium on Multimedia (ISM'2013), Anaheim, California, USA, pp. 125-132, 2013.
- [56] Sert M., Baykal B., Yazıcı A., Structural and Semantic Modeling of Audio for Content-based Querying and Browsing, Lecture Notes in Computer Science (Springer-Verlag), LNCS 4027:319-330, 2006.
- [57] Kyriakakis C., Fundamental and Technological Limitations of Immersive Audio Systems, Proceedings Of The IEEE, Vol. 86, No. 5, 1998.
- [59] Douglas-Cowie E., Cowie R., Sneddon I., Cox C., Lowry O., McRorie M., Martin J. C., Devillers L., Abrilian S., Batliner A., Amir N., Karpouzis K., The HUMAINE database: addressing the collection and annotation of naturalistic and induced emotional data, Affective Computing and Intelligent Interaction, Vol. 4738/2007, pp. 488–500, Springer, 2007.
- [60] Behbahan N. S., Azari S., Bahadori H. An Introduction To The Features Extracted From The Audio Signal, International Journal of Advanced Research in IT and Engineering, Vol. 2, No. 8, 2013.
- [61] Allen J.F., Maintaining Knowledge About Temporal Intervals, Communications of the ACM, ACM Press. pp. 832–843, 1983.

- [62] Internet <http://www.mathworks.com/help/stats/classification-using-nearest-neighbors.html#bsehyk>.
- [63] Internet http://www-01.ibm.com/support/knowledgecenter/SSLVMB_20.0.0/com.ibm.spss.statistics.help/syn_knn_overview.htm.
- [64] Zhou L., Lin H., Liu W., Enriching Music Information Retrieval Using Emotion Detection, SIGIR 2011 Workshop on Enriching Information Retrieval, 2011.
- [65] Doğan E., Sert M., Yazıcı A., A Flexible and Scalable Audio Information Retrieval System for Mixed-Type Audio Signals, International Journal Of Intelligent Systems, Vol. 26, 952–970 (2011).
- [66] Sert M., Baykal B., Web-based Query Engine for Content-based and Semantic Retrieval of Audio, 2004 IEEE International Symposium on Consumer Electronics, s. 485-490, 2004.
- [67] G. Lu, "Indexing and retrieval of audio: A survey," Multimedia Tools and Applications, vol. 15, pp. 269-290, 2001.
- [68] J. Foote, "An overview of audio information retrieval," ACM Multimedia Systems, vol. 7, pp. 2-10, 1999.
- [69] Yoo M-J., Lee I-K., Affecticon: Emotion-Based Icons for Music Retrieval, IEEE Computer Graphics and Applications, Vol. 31, s. 89-95, 2011.

Ekler

Ek-1: Öznitelik performansları çalışması için diğer özniteliklerin hata matrisi

Çizelge Ek1.1 MP+ASS+ASC bileşik özniteliği hata matrisi

Referans Duygu	Tanınan Duygu						
	Öfke	Endişe / Korku	Sıkıntı	İğrenme	Mutluluk	Nötr	Üzüntü
Öfke	12	1	0	0	0	0	0
Endişe/Korku	1	4	0	1	0	1	0
Sıkıntı	0	0	6	1	0	0	1
İğrenme	0	1	2	0	0	0	1
Mutluluk	2	0	0	2	3	1	0
Nötr	0	2	2	0	0	3	0
Üzüntü	0	0	0	0	0	1	6

Çizelge Ek1.2 ZCR özniteliği hata matrisi

Referans Duygu	Tanınan Duygu						
	Öfke	Endişe / Korku	Sıkıntı	İğrenme	Mutluluk	Nötr	Üzüntü
Öfke	13	0	0	0	0	0	0
Endişe/Korku	7	0	0	0	0	0	0
Sıkıntı	7	0	0	0	0	1	0
İğrenme	5	0	0	0	0	0	0
Mutluluk	6	0	0	0	0	1	0
Nötr	4	0	0	0	0	4	0
Üzüntü	6	0	0	0	0	0	0

Çizelge Ek1.3 MP özniteliği hata matrisi

Referans Duygu	Tanınan Duygu						
	Öfke	Endişe / Korku	Sıkıntı	İğrenme	Mutluluk	Nötr	Üzüntü
Öfke	13	0	0	0	0	0	0
Endişe/Korku	0	5	0	1	1	0	0
Sıkıntı	0	0	3	0	1	4	0
İğrenme	0	1	1	2	0	0	0
Mutluluk	2	2	0	0	2	1	0
Nötr	0	0	3	0	0	5	0
Üzüntü	0	0	3	0	0	1	3

Çizelge Ek1.4 LPC+ASC+ASS bileşik özniteliği hata matrisi

Referans Duygu	Tanınan Duygu						
	Öfke	Endişe / Korku	Sıkıntı	İğrenme	Mutluluk	Nötr	Üzüntü
Öfke	12	0	0	0	1	0	0
Endişe/Korku	0	2	0	0	0	5	0
Sıkıntı	1	1	4	0	0	2	0
İğrenme	1	1	0	1	1	0	0
Mutluluk	4	0	0	0	2	1	0
Nötr	0	0	1	1	1	5	0
Üzüntü	0	0	1	0	0	1	5

Çizelge Ek1.5 ASE öz niteliği hata matrisi

Referans Duygu	Tanınan Duygu						
	Öfke	Endişe / Korku	Sıkıntı	İğrenme	Mutluluk	Nötr	Üzüntü
Öfke	12	0	0	0	1	0	0
Endişe/Korku	1	1	0	1	2	0	1
Sıkıntı	0	1	1	0	1	4	2
İğrenme	1	1	0	1	0	1	0
Mutluluk	4	0	0	1	2	0	0
Nötr	0	1	0	0	0	7	0
Üzüntü	0	1	0	0	0	0	5

Çizelge Ek1.6 LPC öz niteliği hata matrisi

Referans Duygu	Tanınan Duygu						
	Öfke	Endişe / Korku	Sıkıntı	İğrenme	Mutluluk	Nötr	Üzüntü
Öfke	11	0	1	0	0	0	0
Endişe/Korku	2	2	2	0	0	0	1
Sıkıntı	3	0	4	0	0	0	1
İğrenme	0	0	2	2	0	1	0
Mutluluk	5	0	2	0	0	0	0
Nötr	1	0	2	0	0	4	1
Üzüntü	0	0	2	0	1	0	3

Çizelge Ek1.7 ASP öz niteliği hata matrisi

Referans Duygu	Tanınan Duygu						
	Öfke	Endişe / Korku	Sıkıntı	İğrenme	Mutluluk	Nötr	Üzüntü
Öfke	10	0	0	0	1	0	1
Endişe/Korku	3	0	1	0	0	2	1
Sıkıntı	1	0	2	0	3	2	0
İğrenme	5	0	0	0	0	0	0
Mutluluk	5	0	0	0	2	0	0
Nötr	0	0	2	0	3	3	0
Üzüntü	1	0	2	0	1	1	1

Çizelge Ek1.8 HSC öz niteliği hata matrisi

Referans Duygu	Tanınan Duygu						
	Öfke	Endişe / Korku	Sıkıntı	İğrenme	Mutluluk	Nötr	Üzüntü
Öfke	13	0	0	0	0	0	0
Endişe/Korku	7	0	0	0	0	0	0
Sıkıntı	5	0	3	0	0	0	0
İğrenme	4	0	0	0	0	0	0
Mutluluk	7	0	0	0	0	0	0
Nötr	5	0	3	0	0	0	0
Üzüntü	6	0	1	0	0	0	0

Çizelge Ek1.9 HSD özneliği hata matrisi

Referans Duygu	Tanınan Duygu						
	Öfke	Endişe / Korku	Sıkıntı	İğrenme	Mutluluk	Nötr	Üzüntü
Öfke	13	0	0	0	0	0	0
Endişe/Korku	7	0	0	0	0	0	0
Sıkıntı	7	0	1	0	0	0	0
İğrenme	5	0	0	0	0	0	0
Mutluluk	7	0	0	0	0	0	0
Nötr	5	0	3	0	0	0	0
Üzüntü	6	0	0	0	0	0	0

Çizelge Ek1.10 SC özneliği hata matrisi

Referans Duygu	Tanınan Duygu						
	Öfke	Endişe / Korku	Sıkıntı	İğrenme	Mutluluk	Nötr	Üzüntü
Öfke	13	0	0	0	0	0	0
Endişe/Korku	6	0	1	0	0	0	0
Sıkıntı	5	0	3	0	0	0	0
İğrenme	5	0	0	0	0	0	0
Mutluluk	7	0	0	0	0	0	0
Nötr	5	0	3	0	0	0	0
Üzüntü	6	0	0	0	0	0	0

Çizelge Ek1.11 F0 özneliği hata matrisi

Referans Duygu	Tanınan Duygu						
	Öfke	Endişe / Korku	Sıkıntı	İğrenme	Mutluluk	Nötr	Üzüntü
Öfke	12	0	0	0	0	0	0
Endişe/Korku	6	0	0	0	0	1	0
Sıkıntı	8	0	0	0	0	0	0
İğrenme	4	0	0	0	0	1	0
Mutluluk	7	0	0	0	0	0	0
Nötr	7	0	0	0	0	1	0
Üzüntü	5	0	0	0	0	1	0

Çizelge Ek1.12 HSS özneliği hata matrisi

Referans Duygu	Tanınan Duygu						
	Öfke	Endişe / Korku	Sıkıntı	İğrenme	Mutluluk	Nötr	Üzüntü
Öfke	13	0	0	0	0	0	0
Endişe/Korku	6	0	1	0	0	0	0
Sıkıntı	5	0	3	0	0	0	0
İğrenme	5	0	0	0	0	0	0
Mutluluk	7	0	0	0	0	0	0
Nötr	5	0	3	0	0	0	0
Üzüntü	6	0	0	0	0	0	0

Çizelge Ek1.13 ASC öz niteliği hata matrisi

Referans Duygu	Tanınan Duygu						
	Öfke	Endişe / Korku	Sıkıntı	İğrenme	Mutluluk	Nötr	Üzüntü
Öfke	11	0	0	0	0	0	1
Endişe/Korku	7	0	0	0	0	0	0
Sıkıntı	6	0	0	0	0	0	2
İğrenme	5	0	0	0	0	0	0
Mutluluk	6	0	0	0	0	0	1
Nötr	7	0	1	0	0	0	0
Üzüntü	6	0	0	0	0	0	0

Çizelge Ek1.14 ASS öz niteliği hata matrisi

Referans Duygu	Tanınan Duygu						
	Öfke	Endişe / Korku	Sıkıntı	İğrenme	Mutluluk	Nötr	Üzüntü
Öfke	12	0	0	0	0	0	0
Endişe/Korku	7	0	0	0	0	0	0
Sıkıntı	8	0	0	0	0	0	0
İğrenme	5	0	0	0	0	0	0
Mutluluk	7	0	0	0	0	0	0
Nötr	8	0	0	0	0	0	0
Üzüntü	6	0	0	0	0	0	0

Çizelge Ek1.15 HSV özniteliği hata matrisi

Referans Duygu	Tanınan Duygu						
	Öfke	Endişe / Korku	Sıkıntı	İğrenme	Mutluluk	Nötr	Üzüntü
Öfke	7	0	4	0	0	0	2
Endişe/Korku	5	0	1	0	0	0	0
Sıkıntı	7	0	1	0	0	0	1
İğrenme	2	0	1	0	0	1	0
Mutluluk	5	0	2	0	0	0	0
Nötr	1	0	3	0	0	2	2
Üzüntü	5	0	0	0	0	0	1