

**KARADENİZ TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

ELEKTRİK ELEKTRONİK MÜHENDİSLİĞİ ANABİLİM DALI

SES SİNYALİNDEN DUYGU TANIMA

YÜKSEK LİSANS TEZİ

Elek. Elekt. Müh. Onur Erdem KORKMAZ

**OCAK 2016
TRABZON**



KARADENİZ TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

ELEKTRİK ELEKTRONİK MÜHENDİSLİĞİ ANABİLİM DALI

SES SİNYALİNDEN DUYGU TANIMA

Onur Erdem KORKMAZ

Karadeniz Teknik Üniversitesi Fen Bilimleri Enstitüsünce
"ELEKTRONİK YÜKSEK MÜHENDİSİ"
Unvanı Verilmesi İçin Kabul Edilen Tezdir.

Tezin Enstitüye Verildiği Tarih : 07 / 12 / 2015

Tezin Savunma Tarihi : 04 / 01 / 2016

Tez Danışmanı : Doç. Dr. Ayten ATASOY

Trabzon 2016

**KARADENİZ TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

**Elektrik Elektronik Mühendisliği Anabilim Dalında
Onur Erdem KORKMAZ Tarafından Hazırlanan**

SES SİNYALİNDEN DUYGU TANIMA

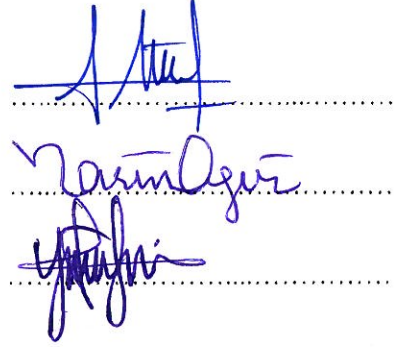
başlıklı bu çalışma, Enstitü Yönetim Kurulunun 15/ 12/ 2015 gün ve 1631 sayılı
kararıyla oluşturulan jüri tarafından yapılan sınavda
YÜKSEK LİSANS TEZİ
olarak kabul edilmiştir.

Jüri Üyeleri

Başkan : Doç. Dr. Ayten ATASOY

Üye : Yrd. Doç. Dr. Yasin OĞUZ

Üye : Yrd. Doç. Dr. Yusuf SEVİM



Prof. Dr. Sadettin KORKMAZ

Enstitü Müdürü

ÖNSÖZ

Tez çalışmam boyunca benden yardımlarını hiçbir zaman esirgemeyen, insani ve ahlaki özelliklerini örnek edindiğim, bilgisinden her daim faydalandığım ve bu süreçte bana göstermiş olduğu sabrından dolayı değerli hocam sayın Doç. Dr. Ayten ATASOY 'a teşekkürlerimi sunarım.

Bu günlere gelmemde en büyük pay sahibi olan aileme, bir harf bile öğrendiğim tüm hocalarıma ve dostlarıma teşekkür ederim.

Onur Erdem Korkmaz

Trabzon 2016

TEZ ETİK BEYANNAMESİ

Yüksek Lisans Tezi olarak sunduğum “ Ses Sinyalinden Duygu Tanıma ” başlıklı bu çalışmayı baştan sona kadar danışmanım Doç. Dr. Ayten Atasoy’ un sorumluluğunda tamamladığımı, sonuçları yaptığım analizlerden elde ettiğimi, başka kaynaklardan aldığım bilgileri metinde ve kaynakçada eksiksiz olarak gösterdiğimi, çalışma süresince bilimsel araştırma ve etik kurallara uygun olarak davrandığımı ve aksinin ortaya çıkması durumunda her türlü yasal sonucu kabul ettiğimi beyan ederim. 04/01/2016

Onur Erdem Korkmaz

İÇİNDEKİLER

Sayfa No

ÖNSÖZ	III
TEZ ETİK BEYANNAMESİ	IV
İÇİNDEKİLER	V
ÖZET.....	VII
SUMMARY	VIII
ŞEKİLLER DİZİNİ.....	IX
TABLolar DİZİNİ	XI
SEMBOLLER DİZİNİ.....	XII
1. GENEL BİLGİLER	1
1.1. Duygusal Konuşma Veri Seti	4
1.2. Tasarım Kriterleri	4
1.3. Mevcut ve Bilinen Konuşmadan Duygu Tanıma Veri Setleri.....	5
1.4. Mevcut Konuşmadan Duygu Tanıma Veri Setlerindeki Problemler.....	6
1.5. Literatürdeki Mevcut Veri Setleri.....	6
1.6. Öznelik Çıkarma.....	8
1.7. Yerel Özellikler ve Global Özelliklerin Karşılaştırması	9
1.8. Konuşma Özellikleri Kategorileri	9
1.8.1. Sürekli Zamanlı Konuşma Özellikleri.....	10
1.8.2. Ses Nitelik Özellikleri	11
1.8.3. Spektral Tabanlı Konuşma Özellikleri	12
1.8.3.1. Mel Frekansı Kepstrum Katsayısı (Mfkk).....	13
1.8.3.1.1. Çerçeveleme	13
1.8.3.1.2. Pencereleme.....	14
1.8.3.1.3. FFT	17
1.8.3.1.4. Mel Frekansına Çevirme	17
1.8.3.1.5. Kepstrum Katsayıları.....	18
1.9. Kullanılan İstatistiksel Hesaplamalar	20
1.10. Sınıflandırıcılar	22
1.10.1. Destek Vektör Makineleri (DVM)	23
1.10.2. k En Yakın Komşu Sınıflandırıcı (k-EK).....	26

1.11.	Temel Bileşen Analizi	28
1.12.	k-Katlı Çapraz Doğrulama Yöntemi ve Başarı Ölçümü.....	28
2.	YAPILAN ÇALIŞMALAR	31
2.1.	Materyal.....	31
2.2.	Yöntem	32
2.3.	Kategorik Değerlendirme ve Öznitelikler	41
3.	BULGULAR	43
3.1.	Destek Vektör Makineleri Sınıflandırıcısına Ait Sonuçlar.....	44
3.2.	k-EK Sınıflandırıcısına Ait Sonuçlar.....	47
3.3.	Farklı Veri Setleri Kullanılarak Elde Edilen Sonuçlar	50
3.4.	Temel Bileşen Analizi Kullanılarak Öznitelik Alt Küme Seçimi	51
4.	SONUÇLAR VE ÖNERİLER	52
5.	KAYNAKLAR	53
	ÖZGEÇMİŞ	60

Yüksek Lisans Tezi
ÖZET
SES SİNYALİNDEN DUYGU TANIMA

Onur Erdem KORKMAZ

Karadeniz Teknik Üniversitesi
Fen Bilimleri Enstitüsü
Elektrik Elektronik Mühendisliği Anabilim Dalı
Danışman: Doç. Dr. Ayten ATASOY
2016, 59 Sayfa

Konuşma sinyalleri insanlar arasındaki hızlı ve en doğal iletişim yöntemlerindendir. Bu durum araştırmacıları, insan ve makine etkileşimini daha hızlı ve verimli hale getirmek için konuşma sinyalinden duygu tanıma alanında çalışmaya sevk etti. Bu tez çalışmasında Kızgın, Nötr, Mutlu ve Üzgün duygu sınıflarına ait ve toplamda 393 veriden oluşan EmoSTAR adlı Türkçe - İngilizce örnekler içeren bir veri tabanı kullanılmıştır. İki farklı dil örneğinin olması duygunun telaffuz ve dilden bağımsız olduğunu göstermek için elverişlidir. Bu veri tabanı kullanılarak, her bir konuşma sinyalinden Mel Frekansı Kepstrum Katsayıları (Mfkk) ve buna ek olarak sıfırıncı Mfkk, enerji ve birinci-ikinci türevleri eklenerek farklı öz nitelikler elde edilerek incelenmiştir. Ayrıca Mfkk çıkarılırken belirlenen çerçeve uzunluğu ve kaydırma süreleri de değiştirilerek sonuçlara olan etkisi incelenmiştir. Bu tez çalışmasında Hu Momentleri ve Doğrusal Öngörü Katsayısı (DÖK) öz nitelikleri de kullanılarak analizler yapılmıştır. Elde edilen öz nitelikler Destek Vektör Makineleri (DVM) ve K En Yakın Komşu (k-EK) sınıflandırıcıları ile çapraz doğrulama yöntemi kullanılarak değerlendirilmiş ve %98,7 başarı oranı elde edilmiştir. Ayrıca bu çalışmada EmoDB veri seti test kümesi olarak kullanılıp, farklı veri setleri arasında doğrulama işlemi gerçekleştirildi. Çalışmanın son aşamasında Temel Bileşenler Analizi ile boyut indirgeme işlemi yapılmış, böylelikle işlem zamanı ve başarı oranı açısından iyi sonuçlar elde edildiği gözlenmiştir.

Anahtar Kelimeler :Ses sinyali, Duygu tanıma, Destek vektör makineleri, K en yakın komşu, Mel frekansı kepsrum katsayısı

Master Thesis

SUMMARY

EMOTION RECOGNITION FROM SPEECH SIGNAL

Onur Erdem KORKMAZ

Karadeniz Technical University
The Graduate School of Natural and Applied Sciences
Department of Electrical and Electronics Engineering
Supervisor: Assoc. Prof. Ayten ATASOY
2016, 59 Pages

Conversation signals are considered as one of the fastest and the natural communication methods among people. This case raised the importance of identifying emotions through conversation signals for researchers in order to make human and machine communication quicker and more efficient. In this thesis study, emotion classes like Angry, Neutral, Happy and Sad also a data base, called as EmoSTAR, which consist of totally 393 data and contains Turkish-English examples, are mentioned. Having two different language examples is sufficient in terms of showing emotions independent from pronunciation and language. Using this data base, it is investigated with different features adding Mel Frequency Cepstral Coefficients (Mfcc), in addition, zeroth Mfcc, energy and first-second derivatives from each speech signal. Furthermore, while Mfcc is extracting, the length of frame and scroll-time were changed in order to study the effect of it on the results. Also, in this thesis study, analyses are made by using Hu Moments and Linear Prediction Coefficient (LPC) features. Obtained features are evaluated using Support Vector Machines (SVM), K Nearest Neighbor (k-NN) classifier and cross-validation method and success rate was obtained as %98,7. Also, in this study EmoDB was used as a test set and verification between different database was performed. The final phase of this study, dimension reduction process has been done by principal componenet analysis and thus it is seen that good results has obtained in terms of processing time and succes rate.

Key Words: Speech signal, Emotion recognition, Support vector machine, K nearest neighbor, Mel frequency cepstral coefficients

ŞEKİLLER DİZİNİ

Sayfa No

Şekil 1.1. Self assessment manikin (SAM) görselleri.....	3
Şekil 1.2. Konuşma özellikleri kategorisi	10
Şekil 1.3. Mfkk katsayılarının elde edilmesi.....	13
Şekil 1.4. Çerçevelere bölünmüş konuşma işareti	14
Şekil 1.5. Pencereleme fonksiyonlarının eğrileri	16
Şekil 1.6. Konuşma çerçevesi ve hamming pencereden geçirilmiş hali	16
Şekil 1.7. Mel Frekans süzgeç öbeği.....	17
Şekil 1.8. Kepstral katsayılarının elde edilişi.....	19
Şekil 1.9. Çerçeveleme ve pencereleme işlemi.....	19
Şekil 1.10. (a) sola çarpık dağılım, (b) normal dağılım, (c) sağa çarpık dağılım	21
Şekil 1.11. Kurtosis için (a) basık dağılım, (b) normal dağılım, (c) sivri dağılım.....	21
Şekil 1.12. (a) $ort < med < mod$, (b) $ort = med = mod$, (c) $mod < med < ort$	21
Şekil 1.13. Duygu tanıma sistemleri yöntem adımları	22
Şekil 1.14. DVM sınıflandırıcı için farklı ayırma düzlemleri.....	23
Şekil 1.15. DVM’de karar sınırı belirleme	24
Şekil 1.16. DVM de iki farklı karar sınırı	25
Şekil 1.17. İki sınıflı problem için doğrusal ayrılamama durumu	26
Şekil 1.18. Test verisi ve en yakın üç komşusu	27
Şekil 1.19. Yeni verinin sınıfının belirlenmesi	27
Şekil 1.20. Veri kümesinin alt parçalara ayrılması	29
Şekil 1.21. (a) Birinci test-eğitim, (b) İkinci test-eğitim, (c) Üçüncü test-eğitim.....	30
Şekil 2.1. Eğitim kümesinde yer alan duygu kategorileri	33
Şekil 2.2. Kızgın sınıfa ait 103 adet ses örneğinin her biri	33
Şekil 2.3. Kızgın sınıfa ait birinci ses örneğinden çıkarılan öz nitelikler	34
Şekil 2.4. Kızgın sınıfa ait birinci ses örneğinden çıkarılan öz nitelikler	35
Şekil 2.5. Karadeniz Teknik Üniversitesi Duygu Tanıma Sistemi ara yüzü.....	36
Şekil 2.6. Diskten ses dosyası yükleme işlemi.....	37
Şekil 2.7. Diskten ses dosyası yükleme işlemi.....	37
Şekil 2.8. Kaydedilen ses dosyasına ait dalga biçimi.....	38
Şekil 2.9. Oluşturulan veri tabanındaki ses örneklerine ait dalga biçimleri.....	38

Şekil 2.10. Ses dosyasına ait pencereleme işlemi	39
Şekil 2.11. Veri tabanı bilgisi.....	39
Şekil 2.12. Ses kaydetme işlemi.....	40
Şekil 2.13. Duygu analizi	40
Şekil 3.1. Eğitim kümesi	43

TABLÖLER DİZİNİ

Sayfa No:

Tablo 1.1. EmoSTAR’da duygu kategorileri ve örnek sayıları	8
Tablo 2.1. Duygu kategorileri	41
Tablo 2.2. Ses sinyallerinden çıkarılan öznitelikler	42
Tablo 3.1. Özniteliklere ait DVM sınıflandırma başarı sonuçları.....	44
Tablo 3.2. Çoklu ve istatistiksel özniteliklere ait DVM sınıflandırma sonuçları.....	45
Tablo 3.3. DVM için farklı çerçeve uzunluğu ve kaydırma süresi sonuçları	46
Tablo 3.4. Özniteliklere ait k-EK sınıflandırma başarı sonuçları	47
Tablo 3.5. Çoklu ve istatistiksel özniteliklere ait k-EK sınıflandırma sonuçları	48
Tablo 3.6. k-EK için farklı çerçeve uzunluğu ve kaydırma süresi sonuçları	49
Tablo 3.7. EmoDB test kümesi olarak kullanıldığında elde edilen sonuçlar.	50

SEMBOLLER DİZİNİ

ABC	Airplane Behaviour Corpus
ASR	Otomatik Konuşma Tanıma
AVIC	Audiovisual Interest Corpus
DVM	Destek Vektör Makineleri
DES	Danish Emotional Speech
GMM	Gaussian Mixture Model
k-EK	K En Yakın Komşu Yöntemi
DÖK	Doğrusal Öngörü Katsayıları
DÖKK	Doğrusal Öngörü Kepstral Katsayıları
Mfkk	Mel Frekanslı Kepstrum Katsayıları
SAL	Sensitive Artificial Listener
SMM	Saklı Markov Model
SUSAS	Speech Under Simulated and Actual Stress
TBA	Temel Bileşenler Analizi
VAM	Vera-Am-Mittag
YSA	Yapay Sinir Ağla

1. GENEL BİLGİLER

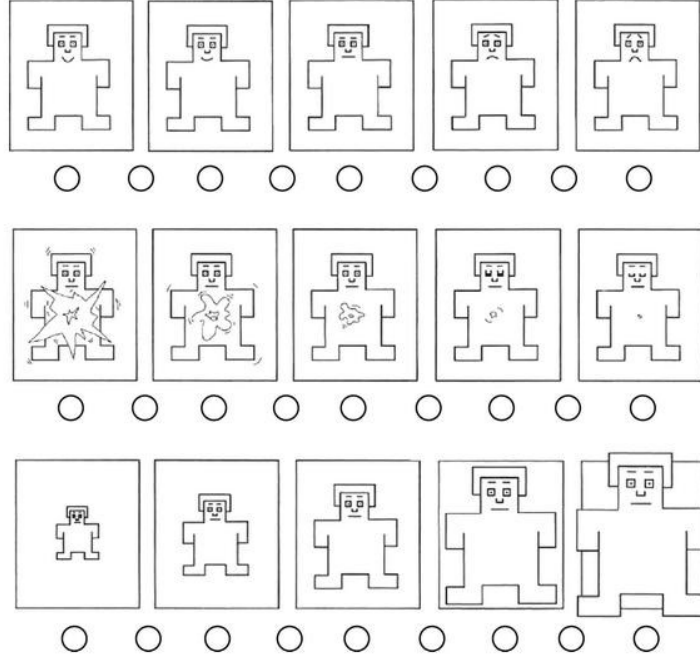
Konuşma sinyalleri insanlar arasındaki hızlı ve en doğal iletişim yöntemlerindendir. Bu durum araştırmacıları insan ve makine etkileşimini daha hızlı ve verimli hale getirmek için konuşma sinyalini kullanmaya yöneltti [1]. Ancak bu durum makinelerin insan sesini tanımada kusursuz çalışması sonucunu doğurdu. Son dönemlerde araştırmacıların büyük bir çoğunluğu konuşma tanıma üzerine çalışmaya başladılar ve insan sesini bir dizi kelimelerden oluşan bir yapı olarak düşündüler. Ancak konuşma tanımada büyük gelişmeler yaşanmasına rağmen, bizler hala makinelerin insan duygusunu tanıyamamasından dolayı insan-makine arasındaki doğal etkileşimden oldukça uzağız. Bu nedenle temeli konuşmacının duygu durumunu ifade eden özellikleri çıkarmaya dayanan ve çoğu araştırmacının son dönemde yöneldiği konuşma sinyalinden duygu tanıma diye isimlendirilen bu alan ortaya çıkmıştır. Konuşmadan duygu tanıma çalışmaları, konuşma sinyalinden anlam bilgisini elde edeceğinden konuşmacı tanıma sistemlerinin performansını da artıracığına inanılmaktadır [2].

Konuşmadan duygu tanıma özellikle insan-bilgisayar arasında doğal etkileşim gerektiren web filmleri ve kullanıcının algılanan duygusuna bağlı olarak tepki veren bilgisayarlı öğretici uygulamalarında kullanılmaktadır [3]. Ayrıca sürücünün zihinsel durumuna göre güvenlik sistemlerini ayarlayan otomobil uygulamaları da bulunmaktadır [3]. Bunlara ilaveten tedavi uzmanları için hastanın duygusunu teşhis eden araçlar tasarlanabilir [4]. Konuşmacının duygu durumunun önemli olduğu otomatik çeviri sistemlerinde de kullanımı faydalı olabilir. Uçak kokpitlerinde konuşmacı tanıma sistemleri bulunur. Stres halinde bu sistemlerin çalışması verimsiz hale gelmektedir. Duygu tanıma sistemleriyle bu istenmeyen durum ortadan kaldırılmaya çalışılmıştır [5]. Konuşmadan duygu tanıma ayrıca çağrı merkezi sistemlerinde ve mobil haberleşme uygulamalarında da kullanılmaktadır [6].

Konuşmadan duygu tanıma aşağıda sıralanan sebeplerden dolayı bir takım zorluklara sahiptir. Öncelikle konuşma özellikleri duygu sınıfları arasında çok etkili bir ayrım gerçekleştirememektedir. Akustik değişkenlik, farklı örneklerin varoluşu, konuşma stili ve konuşmaya eklenen diğer etkenler (şive-ağız vb.) konuşmadan çıkarılan formant ve enerji gibi özellikleri direkt etkilediği için duygu teşhisi zorlaşmaktadır [7]. Ayrıca her bir konuşmanın farklı bölümlerinden farklı duygu sınıfı belirlenebilir ve bu bölümleri

belirlemek de oldukça zordur. Diğer bir sorun ise konuşmacının duygusunu nasıl ifade edeceği onun kültürüne ve çevresel faktörlere bağlı olmasıdır. Çoğu çalışma monolingual (dilli) duygu tanıma üzerine odaklanmıştır ve bu çalışmalarda konuşmacılar arasında kültürel farklılıkların olmadığı varsayılmıştır. Ancak multi-lingual sınıflandırma çalışmalarında kültürel etkenler incelenmiştir [1]. Karşılaşılan diğer bir problem ise konuşma sinyali farklı duygu sınıflarına uğrayabilir olmasıdır. Yani konuşma sinyali içerisinde farklı kısımlarda farklı duygu sınıfına ait örnekler içerebilir. Bundan dolayı duygular geçicidir ve süresizdir. Sonuç olarak otomatik duygu tanıma sistemleri konuşmayı uzun bölümler halinde incelememelidir çünkü bu uzun konuşma farklı duygu sınıflarına sahip olabilir.

Duygular yaygın olarak kabul edilen teoriksel bir tanıma sahip değildirler [8]. Ancak insanlar duyguları hissettikleri gibi bilirler. Bu nedenle araştırmacılar duyguları farklı görüşlerde tanımladılar. Yaygın bir düşünceye göre duygular iki grupta kategorize edilebilir: aktivasyon ve değerlilik [9]. Aktivasyonda belli olan duyguyu ifade etmek için enerji miktarından bahseder. Diğer bir çalışmada ise [10] duygular üç grupta kategorize edilir. Bunlar aktivasyon, değerlilik ve baskınlıktır. Duyguları üç boyutlu duygu uzayında gösterirken yine Bradley ve Lang'ın (1994) geliştirdiği Self Assessment Manikin (SAM, Öz Değerlendirme Modeli) görselleri (Şekil 1.1.) kullanılabilir. Fizyoloji alanında çalışan Williams ve Stevens'a göre [11] duygu üretme mekanizmalarından sempatik sinirler Sevinç, Öfke ve Korku duygularını tetikler-uyarırlar. Bu uyarımlar kalp atışlarını hızlandırırlar, kan basıncını yükseltirler, bilinçaltı ve solunum hareketliliğini değiştirirler, gırtlak altı basıncını artırır, ağızda kuruluğa sebebiyet verirler ve bazen de kaslarda titremelere sebep olurlar. Bütün bunların sonucunda konuşma, yüksek sesle hızlı ve geniş formant aralığına sahip, güçlü ve yüksek frekans enerjisiyle söylenen bir hale dönüşür. Diğer yandan parasempatik sinirlerin uyarılması ile üzüntü duygu türü tetiklenir, kalp atışları ve kan basıncı azalır, tükürük salgısı azalır ve üretilen ses yavaş, alçak formantlı ve düşük frekans enerjili olur. Tüm bu sebeplerden dolayı formant, zaman, ses kalitesi ve telaffuz gibi sese ait akustik özellikler sese ait duyguyla yüksek oranda ilişkilidir [12]. Ancak duygu sadece aktivasyon kullanılarak ayrılamaz. Örneğin hem öfke hem de mutlu duygular yüksek aktivasyonla ilişkilidir. Fakat farklı etki taşırlar yani duygu sınıfları farklıdır. Bu farklılık değerlilik boyutuyla karakterize edilir. Ancak yüksek aktivasyon ve düşük aktivasyon arasında sınıflandırma yapılırken yüksek başarı oranı elde ediliyorken, aynı aktivasyona sahip farklı sınıfları ayırt etmek hala daha zorlu olmaktadır.



Şekil 1.1. Self assessment manikin (SAM) görselleri [13].

Duygu tanıma sistemlerinde en önemli problemlerden biri otomatik sınıflandırıcının birçok duygu sınıfı içerisinde ihtiyaç duyulanı tespit etmesidir. Dil bilimciler yaptıkları araştırmalar sonucunda hayatımıza giren çok fazla duygu sınıfının varlığını tanımladılar. Tipik çalışmalardan bir tanesi Schubiger ve O'Connor tarafından belirlenen ve 300 duygu sınıfı içeren çalışmadır. Ancak bu kadar duygu sınıfı içerisinde sınıflandırma yapmak oldukça zordur. Birçok araştırmacının kabul ettiği 'palette teori' diye isimlendirilen ve birçok renk tonunun olmasına rağmen ana renkler diye adlandırılan renklerin varlığı gibi duygu sınıflarını da öncelikli duygu sınıflarının varlığını belirleyen bu teori ortaya çıkmıştır [14]. Bu teoriye göre öncelikli duygu sınıfları Kızgın, Nefret, Korku, Mutlu, Üzgün ve Şaşkın olmak üzere altı grup altında toparlanmıştır. Bu duygu sınıfları hayatımızdaki en açık ve farklı duygulardır. Bunlar 'archetypal duyguları' olarak isimlendirilirler [14].

Sesten duygu tanıma çalışmaları temelde üç adımdan oluşmaktadır [1]. Birinci adımda çalışmalarda çok önemli bir yeri olan veri tabanından bahsedilecektir. Mevcut veri tabanlarını kullanmak hem zaman açısından hem de yapılan çalışmanın diğer çalışmalarla karşılaştırılmasına olanak sağlamak için önemlidir. İkinci adımda ise konuşmadan öznitelik çıkarma yöntemlerinden detaylı bir şekilde bahsedilecektir. Üçüncü adımda ise duygu tanıma sistemlerinin son aşaması olan sınıflandırma teknikleri incelenecektir.

1.1. Duygusal Konuşma Veri Seti

Konuşmadan duygu tanıyıcı değerlendirilmesinde hesaba katılması gereken en önemli ölçütlerinden bir tanesi de doğallık derecesi yüksek olan bir veri seti belirlemektir. Böylelikle çalışmanın performansı artırılmış olur. Düşük kalitede veri setinin kullanılması yanlış sonuçların oluşmasına sebebiyet verebilir. Hatta veri seti tasarlamak sınıflandırıcı başarısında kritik öneme sahiptir.

Bu bölüm üç alt bölüme ayrılacaktır. Birinci alt başlıkta duyguların seslendirilmesi, söylemlerin benzetilmesi, dengeli-dengesiz söylem, duygu sınıflarının eşit dağıtıldığı söylemler gibi farklı kıstasların kullanılmasının duygusal konuşma veri tabanı değerlendirmesine olumlu etkilerinden bahsedilecektir. İkinci alt başlıkta ise literatürde bulunan bazı veri setlerine kısa bir genel bakış sunulacak. Son olarak da duygusal konuşma veri setlerindeki bazı sınırlamalardan üçüncü alt bölümde bahsedilecektir

1.2. Tasarım Kriterleri

Duygusal konuşma veri setinin daha gerçekçi ve iyi olduğuna hüküm verebilmek için bazı ölçütlerin olması gerekir. Bazı çalışmalara göre [15, 16] duygu setinin belirlenmesinde aşağıda sıralanan faktörler göz önüne alınmalıdır.

Gerçek duygular ya da aktörlerin seslendirdikleri: En gerçekçi veri setleri gerçek hayattan toplanmış örnekler içerenlerdir. En tanınmış örnekler ise içerisinde büyük kaza ya da olayların geçtiği radyo-televizyon haber yayınlarından alınanlardır. Bu kayıtlar doğal duyguları taşıyan örnekler içerir. Ama ne yazık ki bu tür kayıtlar bazı yasal ve ahlaki sıkıntılar içerdiğinden dolayı araştırmalar için yasaklanmıştır. Bu duruma karşın alternatif olarak şuan mevcut birçok veri tabanında olduğu gibi stüdyolarda kaydedilen duygu örnekleri oluşturulabilir. Ancak aktörlerin seslendirdiği bu örnekler gerçek hayattakilere tam manasıyla benzemediği için araştırmacılar tarafından sıkça eleştirilirler. Williams ve Stevens [11] gerçek örneklerine nazaran daha fazla abartılı olmasına rağmen aktörlerin oluşturduğu yapay duygu örneklerini oluşturdular. Yine de yapay duygu örneklerinin akustik bağı ile gerçek duygu örneklerinin akustik bağı arasında çok fazla çelişki olmadığını gözlemlediler.

Duyguların seslendirilmesi: Birçok duygu setinde, gerekli duygular için önceden belirlenmiş örnekleri seslendirmeleri için profesyonel aktörler davet edilir. Ancak bazı veri setlerinde örneğin Danish Emotional Speech (DES) veri setinde [17] gerçeğe daha yakın ve

abartılı örneklerden kaçınmak için yarı profesyonel aktörlerin seslendirilmesi tercih edilmiştir.

* Söylemlerin benzetilmesi: Birçok duygu veri setinde kaydedilen söylemler kaynak açısından eksik değildir. Ancak bu söylemler yani kaydedilen örnekler gerçek hayatta farklı durumlara cevaben oluşturulan duyguların doğallığı açısından eksik olabilmektedir. Genellikle duygu söylemlerinin ortaya çıkarılmasında iki yaklaşım mevcuttur. İlk yaklaşım özellikli duygu sınıfları için örneğin mutlu, kızgın, üzgün vb. aktörlerin tecrübelerinden faydalanmaktır. Son dönemdeki çalışmalarda ise doğal duygu örneği çıkarmak için bilgisayar oyunları kullanılmaktadır [18]. Ses örnekleri oyuncunun kazanması ya da kaybetmesi gibi oyunun farklı durumları içerir ve böylelikle gerçeğe daha yakın keyifli ya da keyifsiz ses örnekleri ortaya çıkarılmış olur.

* Dengeli söylem veya dengesiz söylem: Dengeli söylemleri kullanmak bilimsel analizlerin ve deneylerin kontrolünde faydalıdır ancak dataların doğruluğunu azaltabilir. Alternatif olarak dengesiz söylemleri kullanmak doğruluk açısından fayda sağlayabilir.

* Duygu sınıflarının eşit dağıtıldığı söylemler: Bazı veri seti oluşturucular Berlin veri setinde olduğu gibi [19], sınıflandırma hesaplamasının uygun olması için neredeyse tüm duygu sınıflarının örnek sayısının eşit dağıtılmasını tercih ederler. Diğer yandan birçok başka araştırmacı ise dünyada duygu sınıflarının görülme sıklıklarıyla orantılı olarak veri setinde duyguların dağıtılmasını tercih ederler. Bundan dolayı dünyada en sık rastlanan duygu sınıfı nötr olduğu için veri setinde bu duygu türünün ağırlıklı olmasını tercih ederler.

1.3. Mevcut ve Bilinen Konuşmadan Duygu Tanıma Veri Setleri

Geliştirilen birçok duygu tanıma veri setleri genel kullanım için ulaşılabilir değildir. Bu yüzden araştırmacılar için çok az veri tabanı paylaşılmıştır. Bu gizlemenin yani bazı veri setlerinin paylaşılmamasının diğer bir sonucu ise araştırmacılar arasındaki koordinasyon eksikliğine sebebiyet vermektedir. Farklı veri setleri oluşturulurken benzer hataların yapılması araştırma sonuçlarının verimini etkilemektedir.

1.4. Mevcut Konuşmadan Duygu Tanıma Veri Setlerindeki Problemler

Mevcut duygu tanıma veri setlerinin çoğunda duygu tanıma değerlendirme performanslarında bir takım sınırlamalar mevcuttur. Aşağıda kısaca bu sınırlamalardan bahsedilmektedir.

(1) Birçok veri setinde yeteri kadar açık ve doğal duygu benzetimi yapılamamaktadır. Buna kanıt olarak nispeten düşük başarı oranına sahip (%65) çalışmalar gösterilebilir [20].

(2) Bazı veri setlerinde örneğin KISMET, kayıt kalitesi yeteri kadar iyi değildir. Ayrıca örnekleme frekansı da biraz düşüktür (8 kHz).

(3) Fonetik bilgi duygu tanıma açısından önemlidir. Fonetik kayıt olmayan bazı veri setleri vardır. BabyEars veri seti buna örnek verilebilir [21]. Bu veri setinde dilbilimsel içerik eksikliğinden dolayı başarı oranı düşüktür.

1.5. Literatürdeki Mevcut Veri Setleri

Duygusal konuşma kayıtları oluşturmaya yönelik yapılan çalışmalar da son yıllarda büyük artış gözlemlenmektedir. Mevcut veri tabanlarının kapsamlı değerlendirmelerine [22] [23, 24] ve HUMAINE ekibinin çalışmalarında [25] yer verilmiştir. Bu veri tabanlarının bir kısmı mutlu, kızgın, şaşkın, üzgün, korku, vb. duygu sınıflarına ait önceden hazırlanmış metinleri seslendiren profesyonel aktörler tarafından oluşturulmuştur [26]. Bu tip veri tabanlarına yani profesyonel aktörler tarafından oluşturulan veri setlerine örnek olarak Berlin Duygusal Konuşma Veri Tabanı [19] ve Danimarka’ca Duygusal Konuşma Veri Tabanı [27] verilebilir. Berlin Veri Tabanı konuşma dili Almanca olan ve profesyonel aktörlerin seslendirdiği Kızgın, Neşeli, Üzgün Korku, Nefret, Sıkın ve Nötr olmak üzere 7 duygu sınıfına ait toplamda 800 örnek içeren bir veri setidir. Danimarka Veri Seti ise profesyonel olmayan aktörlerin seslendirdiği konuşma dili Danimarka’ca olan Kızgın, Neşeli, Üzgün, Şaşırılmış ve Nötr olmak üzere 5 duygu sınıfına ait toplam 260 örnek içeren bir veri setidir.

Veri tabanlarında bulunan örneklerin doğallığı ve gerçekliği kabul edilebilir bir düzeyde olması önemlidir. Son yıllarda yapılan veri seti hazırlama çalışmalarında daha doğal örnekler oluşturabilmek için hazırlanan bir senaryo çerçevesinde ve istenilen duyguyu tetikleyecek ortam hazırlanarak kayıtlar oluşturulmaya çalışılmaktadır [26]. Bu tür veri tabanlarına örnek olarak Sensitive Artificial Listener (SAL) [28], Airplane Behaviour Corpus

(ABC) [29], Speech Under Simulated and Actual Stress (SUSAS) [30], Audiovisual Interest Corpus (AVIC) [31], SmartKom [32] ve FAU AIBO [33] verilebilir. Örneğin FAU AIBO [33] veri tabanı 51 çocuğun Sony AIBO robotla WoZ yöntemiyle Şaşkın, Tedirgin, Neşeli, Çaresiz, Öfkeli, Nötr ve Sıkkin duygu başlıklarıyla kaydedilmesiyle oluşturulmuş veri tabanıdır [26]. SUSAS veri tabanı ise dili İngilizce olan ve gerçek duyguya benzetmek için ayarlanmış ortamlarda 13 bayan 19 erkek aktörün seslendirdiği toplamda 16000 örnek içeren bir veri tabanıdır.

Veri tabanı oluşturmada kullanılan diğer yöntemlerden bir tanesi de acil merkez arama kayıtları, tedavi seans kayıtları ve müşteri hizmetleri arama kayıtları gibi doğal konuşma kayıtlarının kullanılmasıdır [26, 34, 35], Capital Bank Service and Stock Exchange çağrı merkezi ses kayıtları [36] ve Geneva Airport Lost Luggage çağrı merkezi ses kayıtları [37] gibi veri tabanlarını bu türe örnek olarak verebilir.

Diğer bir veri tabanı oluşturma yöntemi ise haber, film-dizi ve TV programlarından alınmış ses kayıtlarıdır. Alman TV talk Show programı Vera-Am-Mittag (VAM) [38] kayıtları kullanılarak oluşturulan veri tabanı bu türe örnek olarak gösterilebilir. Bu veri tabanı TV programına katılmış 47 konuktan kaydedilmiş toplamda 946 örnekten oluşmaktadır. Bu veri tabanında konuşma sinyalleri 3-boyutlu duygu düzleminde Self-Assessment Manikin (SAM) [10] kullanılarak değerlendirmeye tabi tutulmuştur. Belfast Naturalistic Veri Tabanı [39] TV sohbet programları ve röportajlardan oluşturulmuş geniş duygu içeriğine sahip bir veri tabanı olarak bu türe girmektedir. Yine bu tür veri tabanına, orijinal dili Türkçe olan dizi ve filmlerden alınmış profesyonel aktörler tarafından seslendirildiği için yarı-doğal/simule bir veri tabanı olan Türkçe Duygusal Konuşma Veri Tabanı (TURES) [26], kategorik olarak Mutlu, Üzgün, Kızgın, Nötr, Korku, Şaşkın ve Diğer olmak üzere toplam 7 farklı duygu kategorisinin yanı sıra üst seviye duygu kategorilerinden Pozitif, Negatif ve Nötr kategorisine sahip veri tabanı örnek verilebilir.

Bu tez çalışmasında EmoSTAR adlı yeni bir veri tabanı kullanılmıştır. EmoSTAR TV ve internetteki kaynaklardan derlenerek hazırlanmış bir veri tabanıdır [40].

EmoSTAR toplamda 393 olmak üzere Kızgın, Nötr, Mutlu ve Üzgün duygu sınıflarına ait Türkçe ve İngilizce örnekler içermektedir. Nötr örnekler haber spikeri konuşmalarından, kızgın örnekler dizi-filmlerden, üzgün örnekler internetteki videolardan, mutlu örnekler ise Oscar, Golden Globe gibi ödül törenlerinde ödül alan sanatçıların konuşmasından alınmıştır [41]. Nötr ve mutlu örnekler doğal konuşmalardır. Üzgün örneklerin de çok azı hariç büyük çoğunluğu doğal konuşmalardır. Kızgın örnekler ise yapay duygulardır. Örneklerin hepsi

farklı cümleler içermektedir ve uzunlukları 2,2 ile 14,5 saniye arasında değişmektedir. Konuşmacı başına örnek sayısı ise 1 ile 29 arasındadır. Kategorik sınıflandırmaya sahip veri tabanında etiketleme yazar tarafından sesli ve görsel değerlendirmeye videoların sahibi tarafından etiketlenen üzgün ve kızgın gibi etiketlerde göz önüne alınarak yapılmıştır [41]. Veri setinin örnek sayısı ve etiketleri aşağıda Tablo 1.1.'de verilmiştir.

Tablo 1.1. EmoSTAR'da duygu kategorileri ve örnek sayıları (T: türkçe, İ:ingilizce)

	Kızgın	Nötr	Mutlu	Üzgün
Erkek	33 İ – 30 T	35 İ – 34 T	45 İ	12 İ
Kadın	40 İ	37 İ – 20 T	37 İ	51 İ – 19 T
Toplam	103	126	82	82

1.6. Öznitelik Çıkarma

Duygu sınıflarını ayırt edebilmek için ses sinyalinin çıkarılan özelliklere öznitelik denilmektedir. Duygu tanıma sistemi oluştururken karşılaşılan en önemli problem verimli bir şekilde farklı duygu sınıflarını karakterize edebilen uygun özniteliklerin çıkarılmasıdır. Uygun öznitelik seçiminin sınıflandırma performansını ciddi miktarda artırdığı gözlemlenmiştir.

Öznitelik çıkarırken dört problem göz önünde bulundurulmaktadır. Bunlardan ilki özellik çıkarmada kullanılan bölge analizidir. Bazı araştırmacılar sıradan çalışmalarında konuşma sinyalini çerçeve (frame) diye adlandırılan küçük aralıklara bölmüşler ve her bir çerçeveden yerel özellikler çıkarmışlardır. Diğer araştırmacılar ise konuşma sinyalini bir bütün halinde düşünüp bundan global özellikler çıkarmışlardır [42]. Diğer bir önemli soru ise yapılan çalışma için hangi özellik türünün en iyisi olduğudur. Üçüncü soru ise filtreleme sonrası veya gürültü giderme gibi sıradan ses işleme yöntemlerinin sınıflandırma performansı üzerine etkisinin ne olacağıdır? Son olarak duygu modellemede akustik özelliklerin yeterli olup olmayacağı veya eğer gerekli ise dilbilim, söylem bilgisi veya yüz ifadeleri gibi özelliklerin hangisiyle birleştirilmesi gerektiğidir [1].

Bu konu iki alt başlıkta daha detaylı bir şekilde incelenecektir. Bölüm 2.1.'de yerel özellikler ile global özelliklerin karşılaştırılması verilmiştir. Bölüm 2.2.'de konuşmadan duygu tanımda kullanılan farklı türdeki özniteliklerinden bahsedilecektir. Yine bu bölümde duygu tanımda kullanılması gereken öznitelikler hakkında tavsiyelere yer verilmiştir.

1.7. Yerel Özellikler ve Global Özelliklerin Karşılaştırması

Konuşma sinyalleri sabit olmayan geniş örnekler olduğu için ses işleme alanında yaygın olarak kullanılan çerçeve (frame) olarak adlandırılan sesi küçük parçalara bölerek işlem yapılır. Böylelikle her çerçevenin sürekli olması sağlanır. [43]. Prosodic (aruz) özellikler örneğin formant ve enerji gibi, her bir çerçeveden çıkarılır ve yerel özellikler olarak adlandırılır. Diğer yandan global özellikler ses sinyalinin tamamından istatistiksel olarak hesaplanır. Konuşmadan duygu tanımada hangi yöntemin (yerel veya global) daha uygun olduğu hala tartışma konusudur. Birçok araştırmacının ortak kanaati sınıflandırma başarısı ve sınıflandırma zamanı açısından global özelliklerin yerel özelliklerden üstün olduğudur [44, 45, 46, 47]. Global özelliklerin yerel özelliklere göre diğer bir üstünlüğü ise özellik sayısının daha az olmasıdır. Bundan dolayı çapraz doğrulama ve özellik seçme uygulamalarında yerel özelliklere göre daha hızlı sonuç vermiştir.

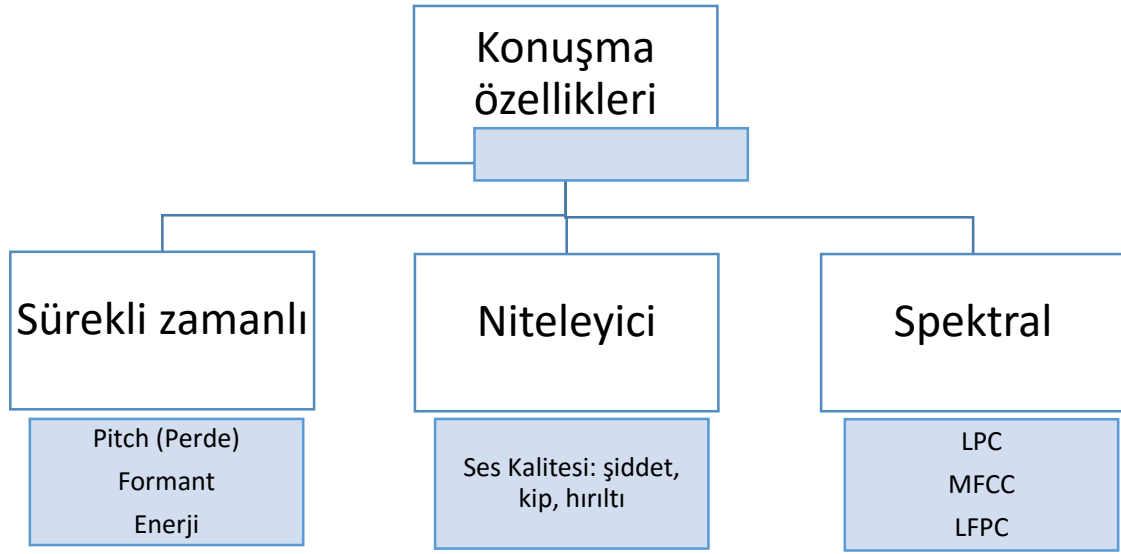
Ancak yapılan araştırmalar sonucunda global özelliklerin sadece yüksek uyarımlı duygular örneğin kızgın, korku ve neşeli, ile düşük uyarımlı duyguları örneğin üzgün, ayırabildiği gözlemlenmiştir [20]. Global özelliklerin aynı uyarımlı sınıftaki duyguları örneğin kızgın ve neşeli, ayırmadaki başarısının kötü olduğu görülmüştür. Global özelliklerin diğer bir dezavantajı ise mevcut sinyaldeki geçici bilginin tamamen kaybolmasıdır. Ayrıca Saklı Markov Model (SMM) ve Destek Vektör Makineleri (DVM) gibi kompleks sınıflandırıcılar ile global özelliklerin kullanılması, model parametresinin oluşturulmasında yetersiz sayıda eğitim vektör elemanın bulunmasından dolayı güvensiz ve verimsizdir. Yani global özelliklerden elde edilen eğitim vektörünün eleman sayısının az olması Destek Vektör Makineleri (DVM) ve Saklı Markov Model (SMM) gibi karmaşık sınıflandırıcıların sistem modellemesini olumsuz etkilediğinden sınıflandırma başarıları yeterli olmayacaktır. Bu tez çalışmasında hem çok duygu sınıflı veri tabanı kullanıldığından (benzer uyarımlı duygu sınıfları içeren) hem de Destek Vektör Makineleri sınıflandırıcısı kullanıldığından dolayı yerel özellikler kullanılmıştır.

1.8. Konuşma Özellikleri Kategorileri

Konuşmadan duygu tanıma sistemlerinde en önemli problem içerik ve konuşmacıya bağlı olmadan verimli bir şekilde farklı duygu sınıflarını kategorize edebilecek uygun

özelliklerin seçilmesidir. Duygu tanımda birçok özellik bulunmasına rağmen hangisinin en ayırt edici olduğu kesinleşmiş bir durum değildir.

Konuşma özellikleri üç grupta kategorize edilmiştir. Bunlar; sürekli zamanlı özellikler, niteleyici özellikler ve spektral özelliklerdir. Şekil 1.2.'de her bir gruba ait özellik örnekleri görülmektedir. Bu bölümün ana amacı her bir kategorinin artılarını ve eksilerini açıklamaktır. Ancak konuşma işleme veya duygu tanıma çalışmalarında farklı kategorilerdeki özellikleri birleştirerek bir özellik vektörü çıkarmak oldukça yaygın bir davranıştır.



Şekil 1.2. Konuşma özellikleri kategorisi

1.8.1. Sürekli Zamanlı Konuşma Özellikleri

Birçok araştırmacının bulgularına göre pitch ve enerji gibi sürekli zamanlı özelliklerin ölçüsü konuşmadaki farklı duygu içeriklerini taşımaktadır [48, 49, 14]. Williams'ın çalışmasına göre konuşmacının uyarım durumu (yüksek uyarımlı veya düşük uyarımlı) ortalama enerji, frekans spektrumundaki çapraz enerji bozunumu ve konuşma sinyalinin durma sürelerini etkiler [11]. Son çalışmalar bunu desteklemektedir [50, 51].

Sürekli zamanlı konuşma özellikleri duygu tanıma çalışmalarında sıklıkla kullanılmaktadır. Temel frekans (F0), enerji, söyleyiş netlik oranı ve sesli-sessizlik bölümlerinin spektral bilgileri bu alanda sıklıkla kullanılan özelliklerdendir. Yapılan çalışmalara göre [52, 15, 14], akustik özellikler aşağıdaki gibi gruplandırılabilir.

- Pitch ile ilişkili özellikler
- Formant özellikleri
- Enerji ile ilgili özellikler
- Zaman özellikleri
- Telaffuz tabanlı özellikler

Temel Frekans (F0): ortalama, medyan, standart sapma, maksimum, minimum, oran(maksimum, minimum), lineer regresyon katsayıları, birinci türevin ortalaması, jilter gibi istatistiki özellikler çıkarılır.

Enerji: ortalama, medyan, standart sapma, maksimum, minimum, oran(maksimum, minimum), lineer regresyon katsayıları, shimmer. gibi istatistiki özellikler çıkarılır.

Süre: konuşma oranı, sesli ve sessiz bölgelerin sürelerinin oranı ve en uzun konuşmanın süresi gibi özellikleri içerir.

Formants: birinci ve ikinci formantlar, ve bunların bant genişliklerini içerir.

1.8.2. Ses Nitelik Özellikleri

Yapılan çalışmalarda [53, 54, 14] söylemlerdeki duygusal içeriğin sesin niteliğine güçlü bir şekilde bağlı olduğu gözlemlenmiştir. Yapılan deneysel çalışmalarda da [55] dinlenen insan sesinden algılanan duygunun ses niteliğiyle ilişkili olduğu ispatlanmıştır. Birçok araştırmacı da işitsel etki ile duygu arasındaki ilişkiyi tanımlamaya çalışmaktadır [14, 56, 57]. Ses niteliği doğrudan duyguyu tanımlar. Örneğin duygular güçlü bir şekilde insanların davranışlarını yönlendirirler [14]. Bununla birlikte temelde yatan duygu karşılıklı olarak pozitif veya negatif olarak insanların hareketlerini etkiler ve düşünceler kontrolümüz dışında gerçekleşir [14]. Ayrıca fonetik çeşitlilik de ses niteliğindeki öznel ifadeye katkıda bulunmaktadır [52]. Bu alanda çalışma yapan [14]'e göre akustik bağ, aşağıda sıralanan dört kategori altında ses niteliğiyle ilişkilidir.

- Ses seviyesi: Sinyal genliği ve enerji
- Ses perdesi (pitch)
- İfade, ses birimi (fonem), kelime ve özellik sınırları
- Geçici yapılar

Bu yaklaşımda ses niteliği, direk ses sinyalinin hesaplanarak nümerik olarak temsil edilir. [58]'in çalışmasında ses niteliği jitter ve shimmer ile temsil edilmiştir [59]. Bu çalışmada sınıflandırıcı olarak Saklı Markov Model (SMM) kullanılmış ve veri seti olarak da SUSAS kullanılmıştır. Sınıflandırma konuşmacıdan bağımsız bir şekilde gerçekleştirilmiştir. Sınıflandırma başarısı ise sadece Mel Frekansı Kepstrum Katsayıları (Mfkk) kullanılarak %65.5, Mfkk ye jitter özelliği eklenerek %68.1, Mfkk ye shimmer özelliği eklenerek %68.5 ve Mfkk ye her ikisi eklenerek %69.1 elde edilmiştir.

Özetle ses niteliği parametreleri kabaca, ses sinyalinin elde edilen pitch (perde), ilk dört formant frekansı ve bunların bant genişliklerinden hesaplanır [60, 61, 62].

1.8.3. Spektral Tabanlı Konuşma Özellikleri

Pitch (perde) ve enerji gibi zamana bağlı akustik özelliklerin yanı sıra ses sinyalleri için kısa zaman aralıklarında temsil edilen spektral özelliklerde konuşmadan duygu tanıma çalışmalarında özellik olarak sıklıkla kullanılırlar. Konuşmadan tanımlanan duygu içeriği spektral enerjiyi ve dolayısıyla konuşmanın frekans tepkisini etkiler [20]. Örneğin mutlu bir ses örneği yüksek enerji ve yüksek frekans tepkisine sahip iken üzgün ses örneği düşük enerji ve frekans tepkisine sahiptir [7, 63].

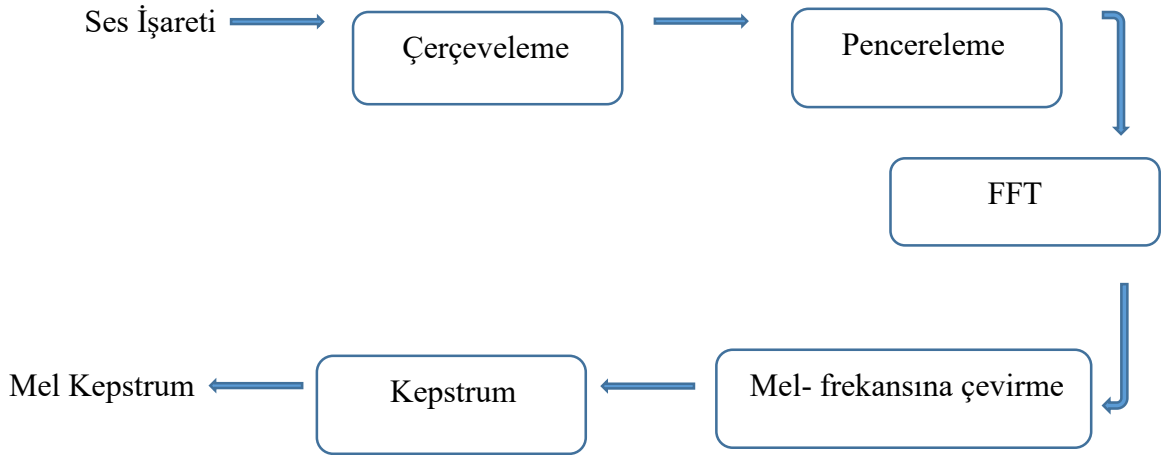
Spektral özelliklere örnek olarak doğrusal öngörü katsayıları (DÖK) [43], tek taraflı oto korelasyon doğrusal öngörü katsayıları (TOKDÖK) [64] ve kısa zaman uyumlu metot (SMC) [65] gösterilebilir. Ancak spektral dağılımdan daha fazla yararlanmak için bant geçiren filtre kullanılmalıdır. Spektral özellikler bu filtrelemeden sonra elde edilir.

Spektral özellikleri doğrusal ve kepsral tabanlı olarak ayırabiliriz. Doğrusal tabanlı yöntemlere örnek olarak DÖK ve TOKDÖK verilebilir. Kepsral tabanlı yöntemlere ise doğrusal öngörü kepsral katsayıları (DÖKK) [66] ve TOKDÖKK [67] ve mel frekansı kepsral katsayıları (Mfkk) örnek olarak verilebilir. Kepsral tabanlı veya doğrusal tabanlı yöntemlerden hangisinin daha iyi olduğuna dair çelişkili sonuçlar vardır. [67]'in çalışmasında kepsral tabanlı yöntemlerin (DÖKK, Mfkk), doğrusal tabanlı yöntemlere (DÖK, TOKDÖK) göre duygu sınıfını ayırt etmede daha iyi sonuçlar verdiği görülmüştür. Ancak Nwe ve arkadaşlarının çalışmasında ise doğrusal tabanlı logaritmik frekans güç katsayıları (LFGK) olarak adlandırılan özellik ile kepsral tabanlı DÖKK ve Mfkk özelliklerinin karşılaştırması yapılmıştır [20]. Çalışma sonucunda LFPC'nin sınıflandırma

sonucu %71.1 iken DÖKK ve Mfkk'nin sınıflandırma sonuçları sırasıyla %56.1 ve %59.0 çıkmıştır.

1.8.3.1. Mel Frekanslı Kepstrum Katsayısı (Mfkk)

Konuşmadan duygu tanıma çalışmalarında sıklıkla kullanılan özneliliklerden birisi Mel Frekanslı Kepstrum Katsayılarıdır. Bunun nedeni Mfkk insan kulağının frekans seçiciliğini taklit ederek iyi bir değerlendirme yapmasıdır. Ayrıca Mfkk değişimlerden ve ses dalga yapısından çok daha az etkilenir. Reynolds'un önerdiği Mfkk elde etme yöntem basamakları Şekil 1.3.'de gösterilmektedir [68].



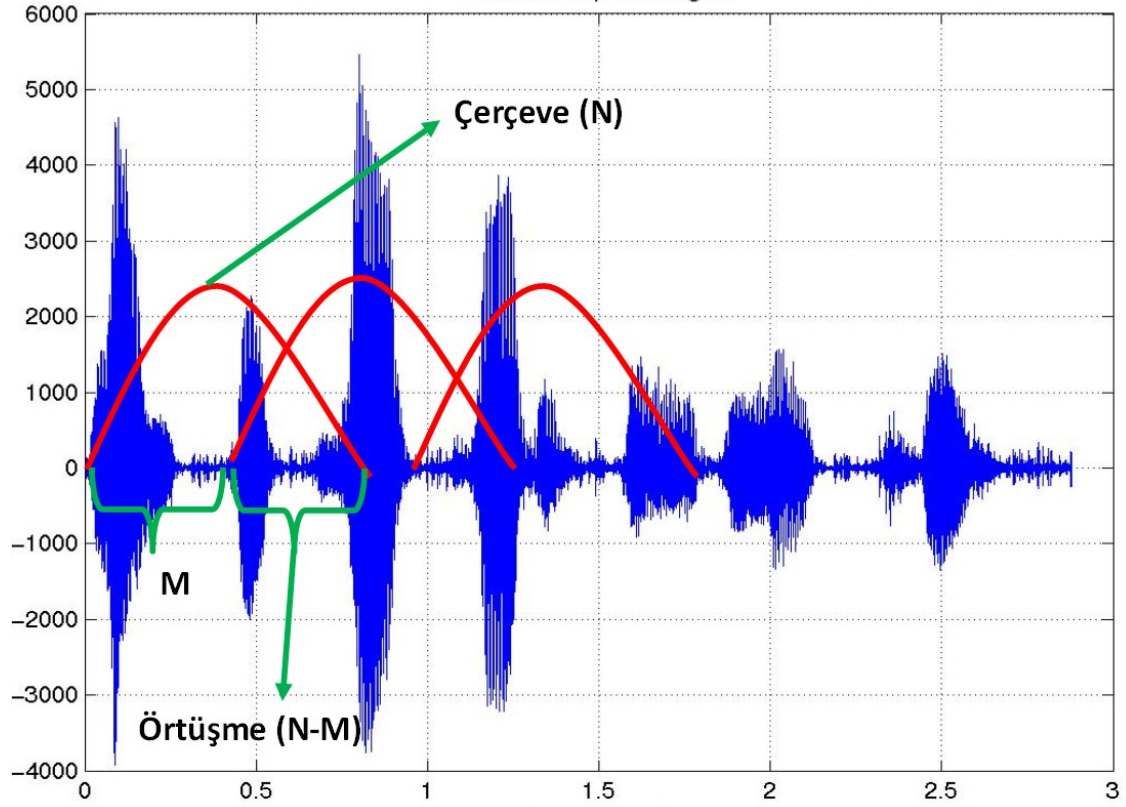
Şekil 1.3. Mfkk katsayılarının elde edilmesi

1.8.3.1.1. Çerçeveleme

Konuşma sinyalleri küçük zaman aralıklarında kararlıdır. Bu nedenle bu sinyaller küçük zaman aralıklarında incelenmelidirler [69]. Konuşma işareti, parametrelerin sabit kaldığı çerçeve diye adlandırılan küçük zaman dilimlerindeki parçalara ayrılmalıdır. Yapılan çalışmalar sonucunda en etkili zaman aralığının 20-30 ms olduğu görülmüştür [70]. Tüm işaret boyunca FFT hesaplanırsa sinyalin farklı spektral bilgisinde kayıplar oluşur. Bunun için tüm işaretin FFT'sini almak yerine her bir çerçevenin FFT'si alınmalıdır [71].

Her çerçeve N tane konuşma örneği ve bir önceki komşu çerçevenin M tane örneğini içerir ($M < N$). Her çerçeve kendinden önceki çerçevenin belirli bir kısmını örtmüştür.

Yani ilk çerçeve N örnekten oluşurken bir sonraki çerçeve ilk çerçeveden M örnek sonra başlar böylelikle $N-M$ örnek örtüşmüş olur (Şekil 1.4.). Bu örtüştürme işleminin amacı bir çerçeveden diğerine geçişin yumuşak olmasını sağlamak ve çerçeve sonundaki bilgi kaybını önlemektir.



Şekil 1.4. Çerçevelere bölünmüş konuşma işareti

1.8.3.1.2. Pencereleme

Mel Frekanslı kepstrum katsayılarını elde etmede ikinci adım pencerelemedir. Pencerelemenin amacı çerçevelerin başında ve sonundaki süreksizliği gidermek ve spektral etkilerini azaltmaktır [72]. Böylelikle ses sinyalinin orta bölgeleri güçlendirilirken kenar bölgeleri zayıflatılmaktadır. Yaygın olarak kullanılan pencereleme fonksiyonları (Şekil 1.5.);

- Hamming
- Hanning
- Blackman

- Gauss
- Dikdörtgen
- Üçgen

En yaygın kullanılan pencere fonksiyonu Hamming dir. Hamming ve diğer pencere fonksiyonlarına ait matematiksel ifadeler aşağıdaki gibidir.

Hamming:

$$w[k+1] = 0.54 - 0.46 \cos\left(2\pi \frac{k}{N-1}\right) \quad k = 0, \dots, N-1 \quad (1)$$

Hanning:

$$w[k+1] = 0.5 \left(1 - \cos\left(2\pi \frac{k}{N-1}\right)\right) \quad k = 0, \dots, N-1 \quad (2)$$

Blackman:

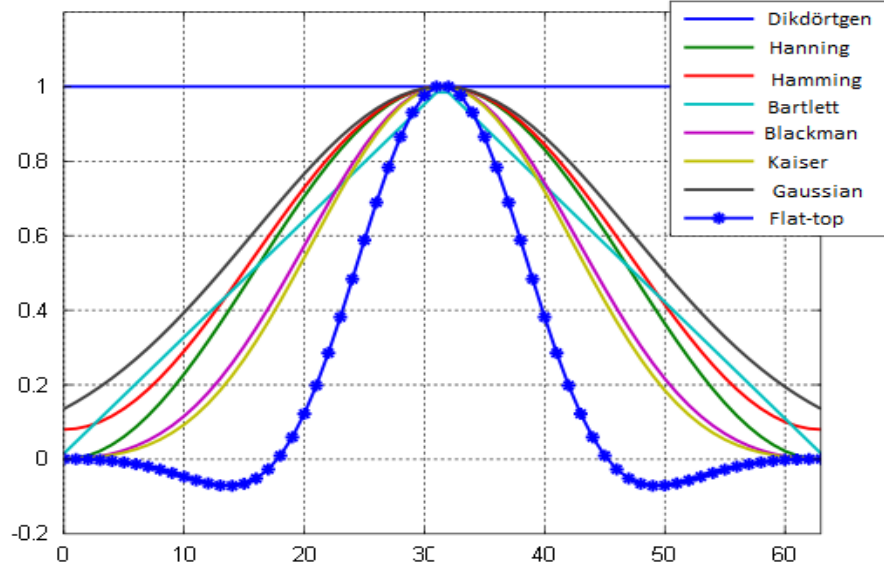
$$w[k+1] = 0.42 - 0.5 \cos\left(2\pi \frac{k}{N-1}\right) + 0.08 \cos\left(4\pi \frac{k}{N-1}\right), \quad k = 0, \dots, N \quad (3)$$

Gauss:

$$w[k+1] = e^{-\frac{1}{2}\left(\frac{k-\frac{N}{2}}{\frac{N}{2}}\right)^2}, \quad k = 0 \leq k \leq N \quad (4)$$

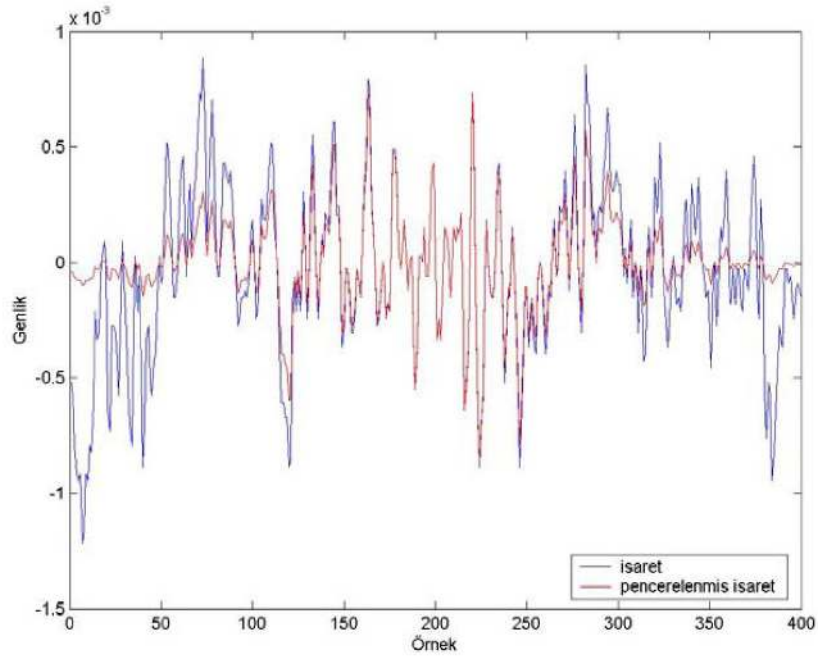
Dikdörtgen:

$$w[k+1] = 1, \quad k = 0, \dots, N-1 \quad (5)$$



Şekil 1.5. Pencereleme fonksiyonlarının eğrileri [71].

Şekil 1.6.'da bir konuşma sinyali ve hamming pencereleme işlemi gerçekleştirildikten sonraki hali görülmektedir. Şekilden de görüleceği gibi sinyalin baş ve son kısımları sıfıra yaklaştırılarak başında ve sonundaki ani değişimler ortadan kaldırılmış olur [73].



Şekil 1.6. Konuşma çerçevesi ve hamming pencereden geçirilmiş hali [71].

1.8.3.1.3. FFT

Bu aşamada ise N örnekten oluşan zaman uzayındaki her bir çerçeveye FFT uygulanarak frekans uzayına çevrilir. FFT ayrık fourier dönüşümü (DCT) uygulamak için hızlı bir yoldur. N elemanlı her bir çerçeve için ayrık fourier dönüşümü denklem 6'daki gibi tanımlanır;

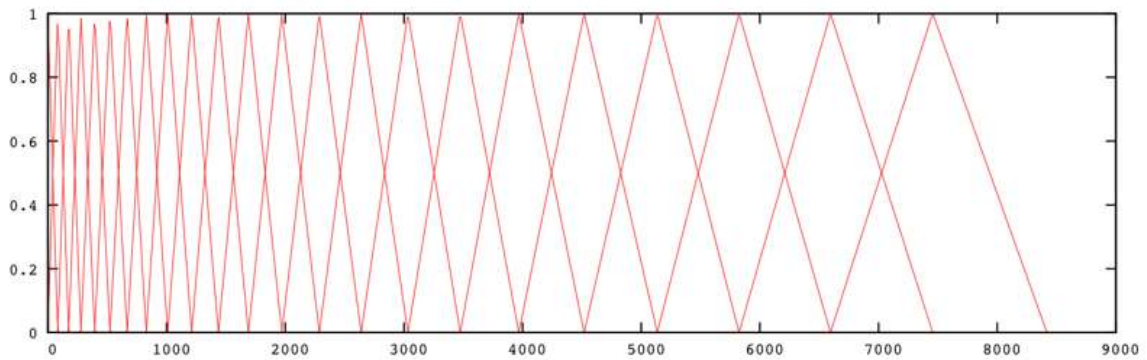
$$X_n = \sum_{k=0}^{N-1} (x_k e^{-2\pi jkn/N}) \quad n = 0,1,2 \dots \dots \dots N - 1 \quad (6)$$

1.8.3.1.4. Mel Frekansına Çevirme

Mel birimi insan kulağının yapısını taklit eden bir yapıdır. Yapılan araştırmalar göstermiştir ki, ölçüler 1kHz'e kadar doğrusal daha yüksek değerlerde ise logaritmiktir. Mel ölçeği ile frekans ölçeği arasındaki dönüşüm aşağıdaki formülle gerçekleştirilmektedir.

$$mel(f) = 2595 * \log\left(1 + \frac{f}{700}\right) \quad (7)$$

Bu algısal spektrumu uygulamanın bir yolu mel skalasına düzgün dağılımlı yerleştirilmiş süzgeç kullanmaktır. Bu süzgeç %50 oranında çakıştırılan, üçgen bant geçiren ve bant genişliği, aralığı mel frekansına bağlı bir süzgeçtir (Şekil 1.7.). Genellikle 20 süzgeç kullanılır [74].



Şekil 1.7. Mel Frekans süzgeç öbeği

1.8.3.1.5. Kepstrum Katsayıları

İnsanların gırtlak yapısından dolayı sesler arasında farklılıklar ve benzerlikler ortaya çıkmaktadır. Kepstral katsayılar bu farklılıkları ve benzerlikleri ortaya koyan yöntemlerdendir.

Kepstrum, homomorfik sinyal işleme teknikleri içinde yer alır [74]. Homomorfik sistemler doğrusal olmayan sistemlerin bir alt dalıdır. Homomorfik sistemler kaynağı ($u(n)$), ses yolunun şeklinden ayırt edebildiklerinden ses sinyali uygulamalarında sıklıkla kullanılırlar. Ayırma işlemi katlama olarak adlandırılır ve aşağıdaki gibi ifade edilir.

$$s(n) = u(n) * h(n) \quad (8)$$

Burda $s(n)$ konuşma işareti, $u(n)$ kaynağı ve $h(n)$ süzgeci tanımlar.

Frekans bölgesindeki gösterim ise;

$$S(w) = U(w).H(w) \quad (9)$$

olarak tanımlanır.

Eşitliğin her iki tarafının da logaritması alınır

$$\begin{aligned} \text{Log } S(w) &= \text{Log}(U(w).H(w)) \\ \text{Log } S(w) &= \text{Log } U(w) + \text{Log } H(w) \end{aligned} \quad (10)$$

böylece çarpımsal ifade toplamsal bir ifadeye dönüşmüş olur. Bu tip dönüşümlere homomorfik dönüşüm denir. Kepstrum, homomorfik bir dönüşümdür ve ayırık zamanda şu şekilde ifade edilir;

$$c(m) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \log |S(w)e^{jwm}d(w)| \quad (11)$$

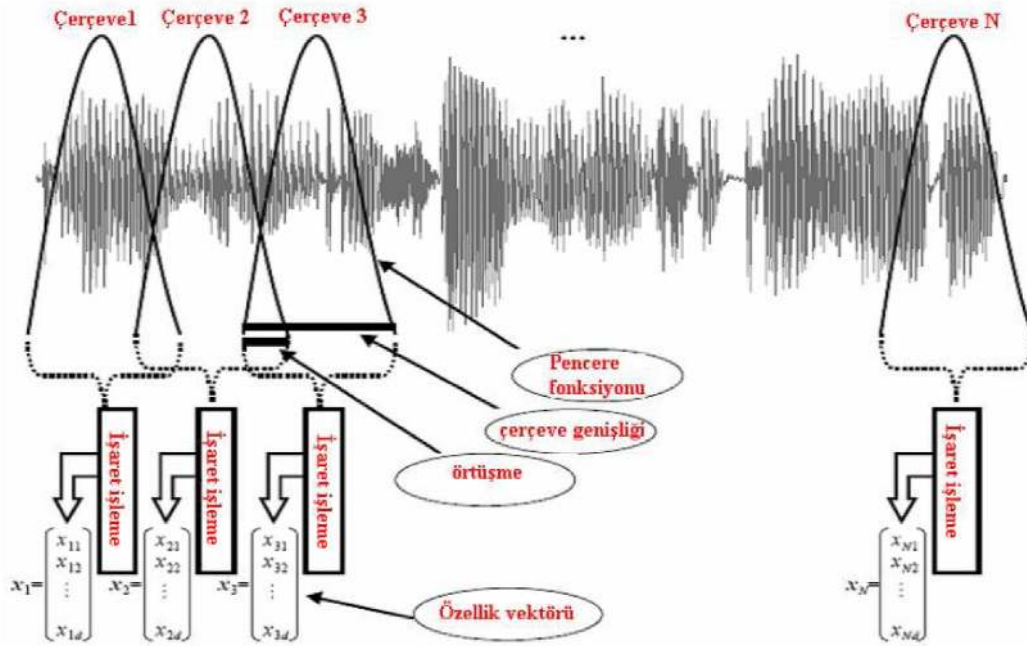
Bu işleme ait blok diyagramı aşağıda verilmiştir (Şekil 1.8.)



Şekil 1.8. Kepstral katsayılarının elde edilişi

Öncelikle ses sinyali pencerelenir. Daha sonra her bir pencerenin Ayırık Fourier Dönüşümü alınır. Alınan ses örneğinin frekans bileşen genliklerinin logaritması alınır. En sonunda ise elde edilen değerlerin Ters Ayırık Fourier Dönüşümü alınarak ses sinyaline ait kepsral değerler elde edilmiş olur.

Aşağıdaki Şekil 1.9.'da ise ses sinyaline ait bilgilerin çerçeveleme ve pencereleme işlemlerini kullanarak çıkarılması işleminin özet hali gösterilmiştir.



Şekil 1.9. Çerçeveleme ve pencereleme işlemi

1.9. Kullanılan İstatistiksel Hesaplamalar

Aritmetik Ortalama: Bir veri setindeki değerlerin toplamının, o veri setindeki terim sayısına bölerek elde edilen merkezsiz konum değeridir.

$$\mu = \frac{1}{n} \sum_{i=1}^n x_i \quad (12)$$

Medyan: Bir sayısal veri seti sıralandığında ortada kalan sayıdır.

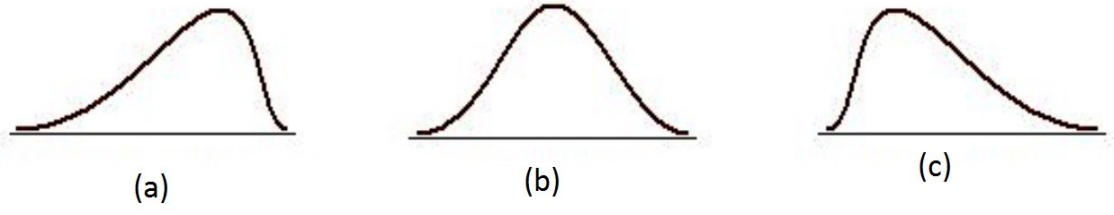
Standart Sapma: Bir veri setindeki verilerin ne kadarının ortalamaya yakın olduğunu bulmaya yarayan yöntemdir.

$$s = \sqrt{\frac{\sum_{n=1}^{\infty} (x_i - \mu)^2}{n}} \quad (13)$$

Global maksimum ve global minimum: Bir veri setinin minimum veya maksimum değerinin bulunmasıyla, veri kümesi optimize (en iyileme) edilmesine yardımcı olur.

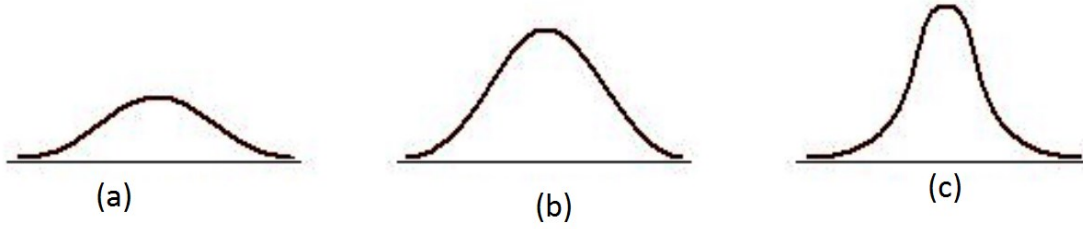
Skewness - Kurtosis: Verilerin normal dağılım gösterip göstermediği grafiği çizilip yorumlanarak yapılabilir. Normal dağılımı test etmenin bir diğer yolu da; ortalama, mod ve medyan değerlerine bakmaktır. Normal dağılımda bu değerler çakışıktır. Verilerin dağılımının normal dağılım gösterip göstermediğini anlamının bir diğer yolu da skewness (çarpıklık) ve kurtosis (basıklık) katsayılarına bakmaktır. Çarpıklık katsayısı normal dağılımda sıfırdır. Negatif çarpıklık katsayısı sağa çarpık dağılımda, pozitif çarpıklık katsayısı sola çarpık dağılıma işaret eder. Basıklık katsayısı da normal dağılımda sıfırdır. Pozitif basıklık katsayısı sivri dağılıma, negatif basıklık katsayısı ise basık bir dağılıma işaret eder. Dağılımın normal dağılımdan manidar düzeyde farklılaşmıyor olması için bu değerlerin (-1, +1) aralığında kalması beklenir.

$$skewnes = \frac{\sum_{i=1}^N (Y_i - \mu)^3 / N}{s^3} \quad (14)$$



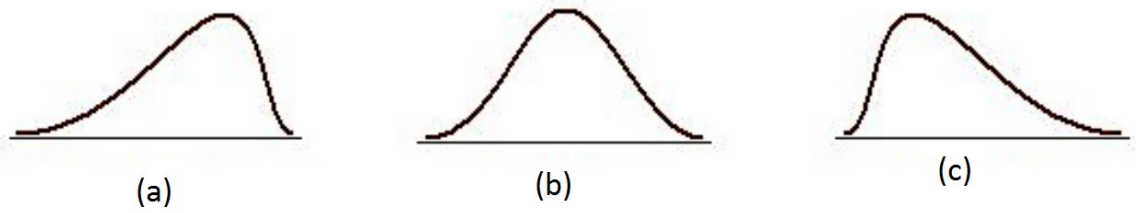
Şekil 1.10. (a) sola çarpık dağılım, (b) normal dağılım, (c) sağa çarpık dağılım

$$kurtosis = \frac{\sum_{i=1}^N (Y_i - \mu)^4 / N}{s^4} \quad (15)$$



Şekil 1.11. Kurtosis için (a) basık dağılım, (b) normal dağılım, (c) sivri dağılım

Ortalama, mod ve medyan değerleri için dağılım grafikleri ise Şekil 2.33. de görülmektedir.



Şekil 1.12. (a) $ort < med < mod$, (b) $ort = med = mod$, (c) $mod < med < ort$

1.10. Sınıflandırıcılar

Bundan önceki bölümlerde sırasıyla veri tabanı seçimi ve sinyal işlemenin önemli adımlarından birisi olan öznitelik çıkarma işlemlerinden bahsedildi. Bu bölümde ise çıkarılan özniteliklere bağlı olarak sınıflandırma algoritmaları hazırlayan yöntemlerden bahsedilecektir. Sesten duygu tanıma sistemleri için ana başlıklar Şekil 1.13.'de görülmektedir.



Şekil 1.13. Duygu tanıma sistemleri yöntem adımları

Aslında konuşmadan duygu tanıma alanında çalışan birçok araştırmacının odaklandığı nokta sınıflandırma aşamasıdır. Problem ve sınıflandırıcı arasındaki ara yüzü bu adım oluşturmaktadır. Sistemin son ve en önemli adımlarındandır.

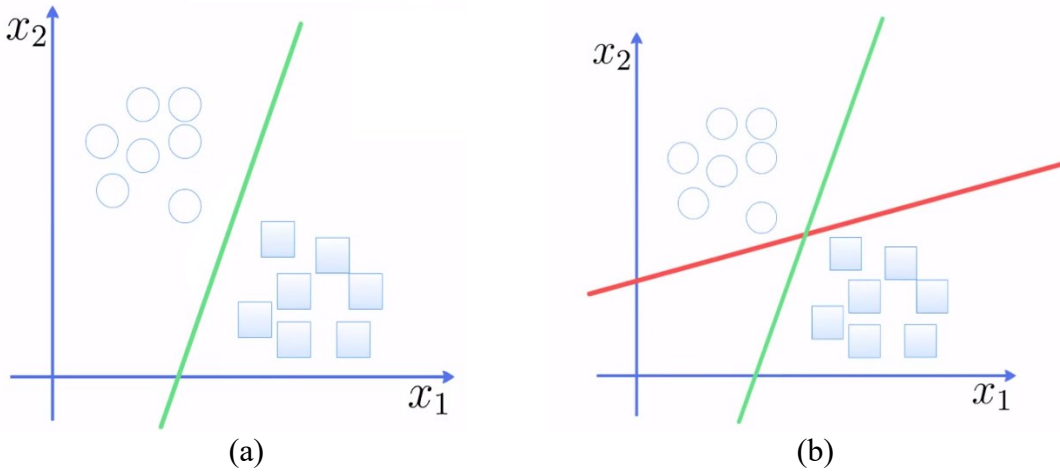
Konuşmadan duygu tanımada çok sayıda farklı tipte sınıflandırıcı kullanılmaktadır. Bunlar; SMM, DVM, yapay sinir ağları (YSA), k en yakın komşu (k-EK) ve diğer birçok sınıflandırma yöntemleridir. Aslında hangi tip sınıflandırıcının duygu sınıflandırmasında daha uygun olduğu kararlaştırılamamıştır. Görünen o ki her sınıflandırıcının kendine has avantajları ve limitleri vardır. Bu alanda kullanılan esas sınıflandırıcılar bir araya getirilerek gruplandırılabilir [62, 75].

Bu alanda yapılan bazı araştırmalara göre [76, 20, 77, 78, 79, 19] SMM, GMM ve DVM konuşma uygulamalarında geniş bir alana sahip olduklarından dolayı, duygu tanıma çalışmalarında da en çok kullanılan sınıflandırıcılardır. Bu bölümün de özellikle tez çalışmasında kullanılan DVM sınıflandırıcısının özellikleri ve limitlerinden bahsedilmektedir.

Sınıflandırmada kullanılan istatistiksel temelli yaklaşımlarda, eğitim datasındaki her bir sınıf olasılık dağılımı tabanlı modellenir. İstatistiksel sınıflandırıcılar birçok ses uygulamalarında kullanılmaktadır. SMM daha çok otomatik konuşma tanıma (ASR) alanında yaygın olarak kullanılmaktadır [80].

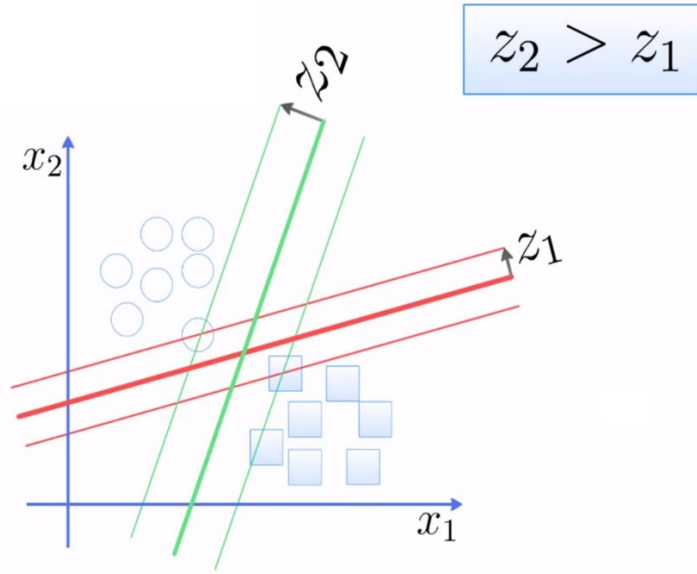
1.10.1. Destek Vektör Makineleri (DVM)

Destek vektör makineleri kendisine verilen eğitim kümesindeki verileri sınıflandırmak için $x_i, r_i, i \in 1, 2, \dots, N$ $r_i \in \{-1, +1\}$ kullanarak bir model geliştirmektedir (-1 ve +1 sınıfları temsil etmektedir). Buradaki r_i değeri x_i ile ilişkili sınıf bilgisini temsil etmektedir. DVM'ler sınıflandırma modellerini ayırma düzlemi (karar sınırı) ile bu düzlemin ayırdığı sınıf örnekleri arasındaki mesafenin maksimum olması için düzlemin pozisyonunu optimize etmektedir. DVM sınıflara ait örnek noktalar arasındaki karar yüzeyini (marjin) oluştururken, yüzeyin sınıflara olan uzaklığını en yüksek yapmaya çalışmaktadır.



Şekil 1.14. DVM sınıflandırıcı için farklı ayırma düzlemleri

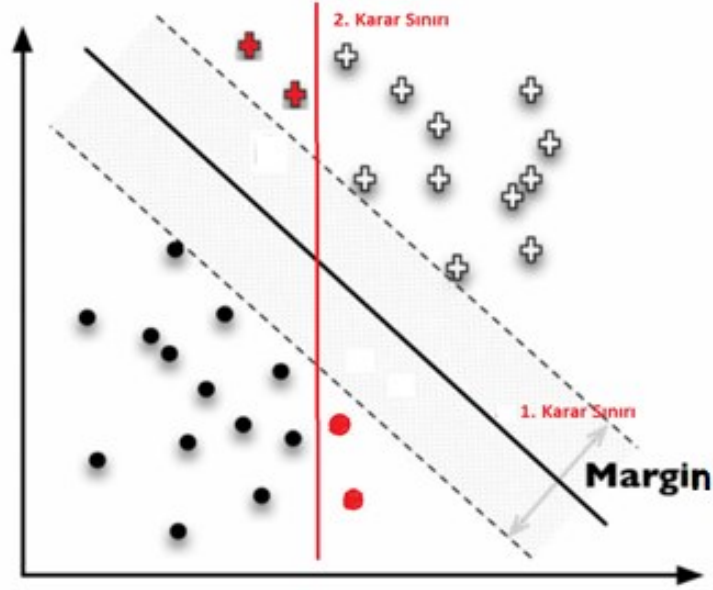
Örneğin yukarıda Şekil 1.14.'de görüldüğü gibi yeşil ve kırmızı renklerle gösterilen 2 farklı karar sınırı bulunmaktadır. DVM'ler bunun gibi birçok karar sınırları oluşturup hangisinin sınıflardaki örneklere en uzak mesafeli olduğunu belirleyerek en sonunda optimum karar sınırını belirlemeye çalışırlar.



Şekil 1.15. DVM’de karar sınırı belirleme

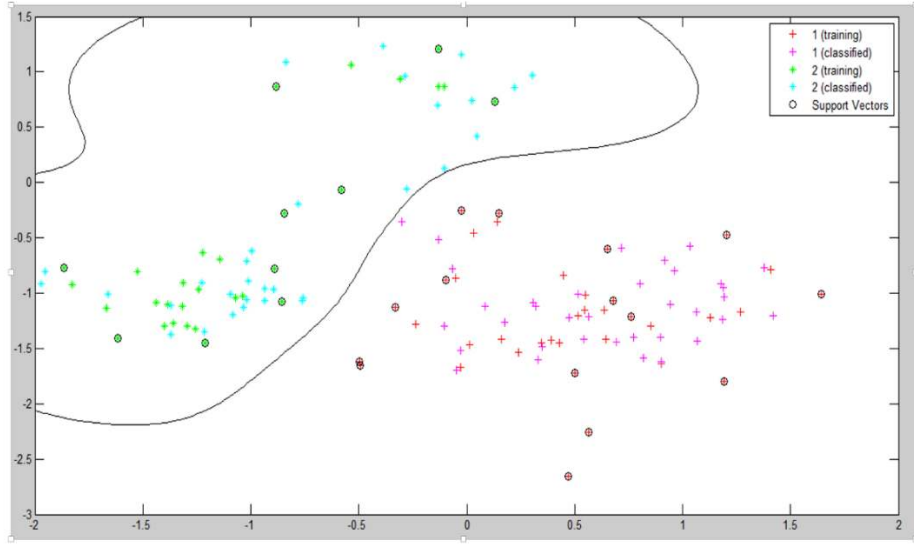
Şekil 1.15.’de görüleceği gibi yeşil renkli karar sınırının yuvarlak ve kareyle gösterilen sınıflara olan uzaklığı, kırmızı renkle gösterilen karar sınırınınkinden daha büyük olduğundan buradaki en uygun ayırma düzlemi yeşil renkli olandır.

Karar sınırının sınıflara en uzak mesafede belirlenememesi sınıflandırıcı performansını düşürecektir. Şekil 1.16.’da görüleceği gibi iki sınıf arasındaki en uzak mesafeli (marjini en büyük) olan 1.karar sınırıdır. Burada sınıflandırıcı yüzeyi olarak 1.karar sınırı seçilmelidir. Ancak marjini en büyük olan seçilmediği durumda ise yani 2.karar sınırı seçildiği takdirde sınıflandırıcının yanlış karar verdiği Şekil 1.16.’da görülmektedir. Burada kırmızıyla gösterilen (+) sınıfına ait örneklerin 2.karar sınırının belirlenmesiyle (o) sınıfa dâhil edileceği böylelikle yanlış sınıflandırılmış olacağı görülmektedir. Aynı şekilde kırmızı renkli (o) sınıf örneklerinin de (+) sınıf olarak yanlış sınıflandırılacağı görülmektedir.



Şekil 1.16. DVM de iki farklı karar sınırı

Destek Vektör Makineleri eğitim verisinde öğrenme yaparak (model geliştirme) yeni veri örneği üzerinde doğru tahmin yapmaya çalışan makine öğrenmesidir. Eğitim kümesi yukarda bahsedilen örnekler gibi doğrusal olarak ayrılabiliriyorsa verileri ayırabilecek sonsuz sayıda doğru (karar sınırı) içerisinde marjini en yüksek yapacak doğruyu bulmayı hedeflemektedir. Ancak veriler her zaman doğrusal olarak ayrılamamaktadır. Veriler doğrusal olarak ayrılabilir değilse, bu durumda doğrusal olmayan haritalama kullanılmaktadır. Burada amaç en geniş marjin ile en düşük toplam ceza arasında bir denge kurmaktır. Şekil 1.17.'de doğrusal olarak ayrılamayan iki sınıflı bir problemin DVM kullanarak ayrıştırma işlemine örnek görülmektedir.



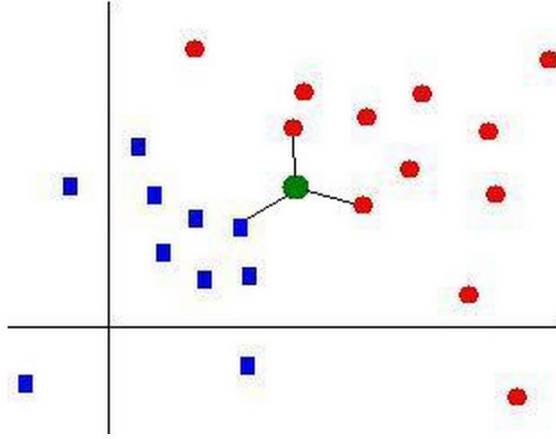
Şekil 1.17. İki sınıflı problem için doğrusal ayrılamama durumu

1.10.2. k En Yakın Komşu Sınıflandırıcı (k-EK)

K En Yakın Komşu Sınıflandırıcısı (k-EK), öznelilik uzayında bir birine en yakın eğitim örneklerine bakarak sınıflandırma yapan en temel örüntü tanıma yöntemlerinden birisidir [81]. Bu sınıflandırıcı verilen k değeri (Örneğin 3 veya 5) kadar en yakın komşusuna bakarak sınıflandırma yapmaktadır. K en yakın komşu sınıflandırıcısında bir vektörün sınıflandırılması sınıfı bilinen vektörler kullanılarak yapılmaktadır [82]. Yani örnek tabanlı yöntemlerdendir. Test edilecek örnek eğitim kümesindeki her bir örnekle teker teker karşılaştırılır. Test edilen örneğin sınıfını belirlemek için eğitim kümesinde o örneğe en yakın k tane değer seçilir. Seçilen bu örneklerden hangisinin sayısı fazla ise test edilen örnek bu sınıfa aittir denilir. Örnekler arası mesafe Öklit (Euclidean) uzaklığı ile bulunur. Literatürde en uygun k değerinin deneme yanılma yoluyla ve genelde 3, 5 veya 10 olarak alındığı belirtilmektedir [83].

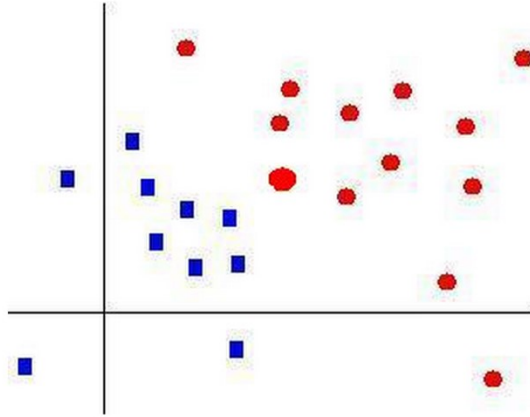
Örneği $k=3$ için yeni bir örneğin sınıflandırılmasında, eski sınıflandırılmış örneklerden (eğitim kümesi) bu yeni örneğe en yakın 3 nokta belirlenir. Belirlenen bu noktalardan en fazla hangi kümenin elemanı varsa yeni verinin sınıfı da bu olmaktadır.

k-EK yöntemiyle yeşil renkli yeni verinin sınıfının belirlenmesi işlemi. (Şekil 1.18.)



Şekil 1.18. Test verisi ve en yakın üç komşusu

En yakın 3 komşunun 2 tanesi kırmızı yuvarlak sınıf olduğu için yeni verinin sınıfı da kırmızı yuvarlak olarak belirlenir (Şekil 1.19.).



Şekil 1.19. Yeni verinin sınıfının belirlenmesi

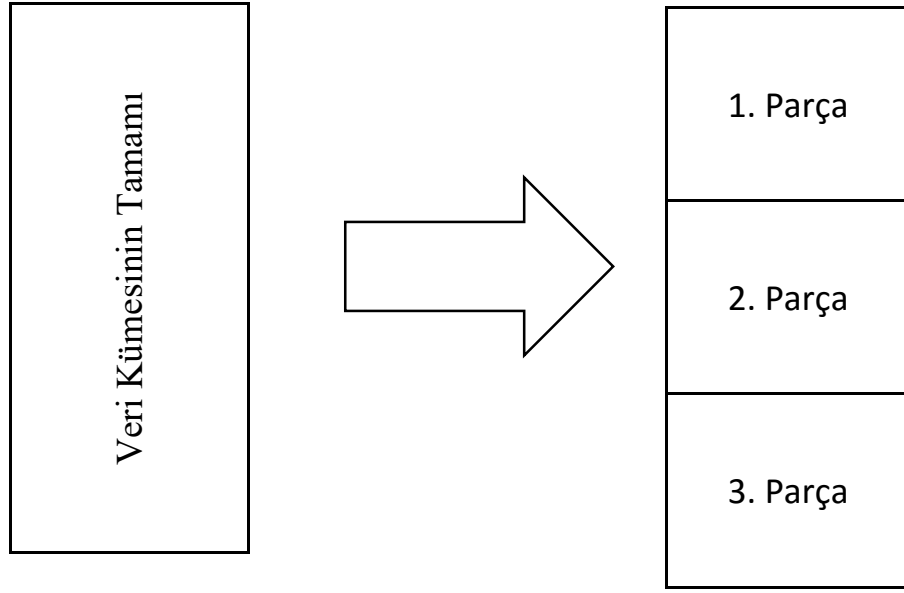
1.11. Temel Bileşen Analizi

Temel Bileşenler Analizi (TBA) bağımlılık yapısını ortadan kaldırma ve boyut azaltma amacıyla kullanılmaktadır. Tanıma, boyut azaltma ve sınıflandırma işlemi yapan çok değişkenli bir istatistik yöntemidir. Bu yöntem veri içindeki en önemli örüntüyü bulmaya çalışır. Bu yüzden örüntü tanıma tekniği olarak da kullanılır. Çoğunlukla verinin sahip olduğu özellik, tüm veri takımından seçilen küçük bir boyut setiyle yakalanabilir. Verideki gürültü, örüntüden daha güçsüz olduklarından, boyut azaltma sonucunda bu gürültüler temizlenebilir. Temel Bileşenler Analizinin; verilerin boyutunu azaltma, tahminleme yapma ve veri setini bazı analizler için görüntüleme gibi temel üç görevi vardır. Temel Bileşenler analizi uygulandığında p boyutlu uzayın gerçek boyutu belirlenir. Bu gerçek boyuta temel bileşenler adı verilir [87].

Temel Bileşenler Analiziyle çok boyutlu verilere doğru açıdan bakarak genellikle verideki ilişkiler açıklanabilir. TBA boyut azaltmada çok faydalı bir yöntemdir ve çok boyutlu verileri yaklaşık olarak ve daha az boyutlu veriyle temsil eder.

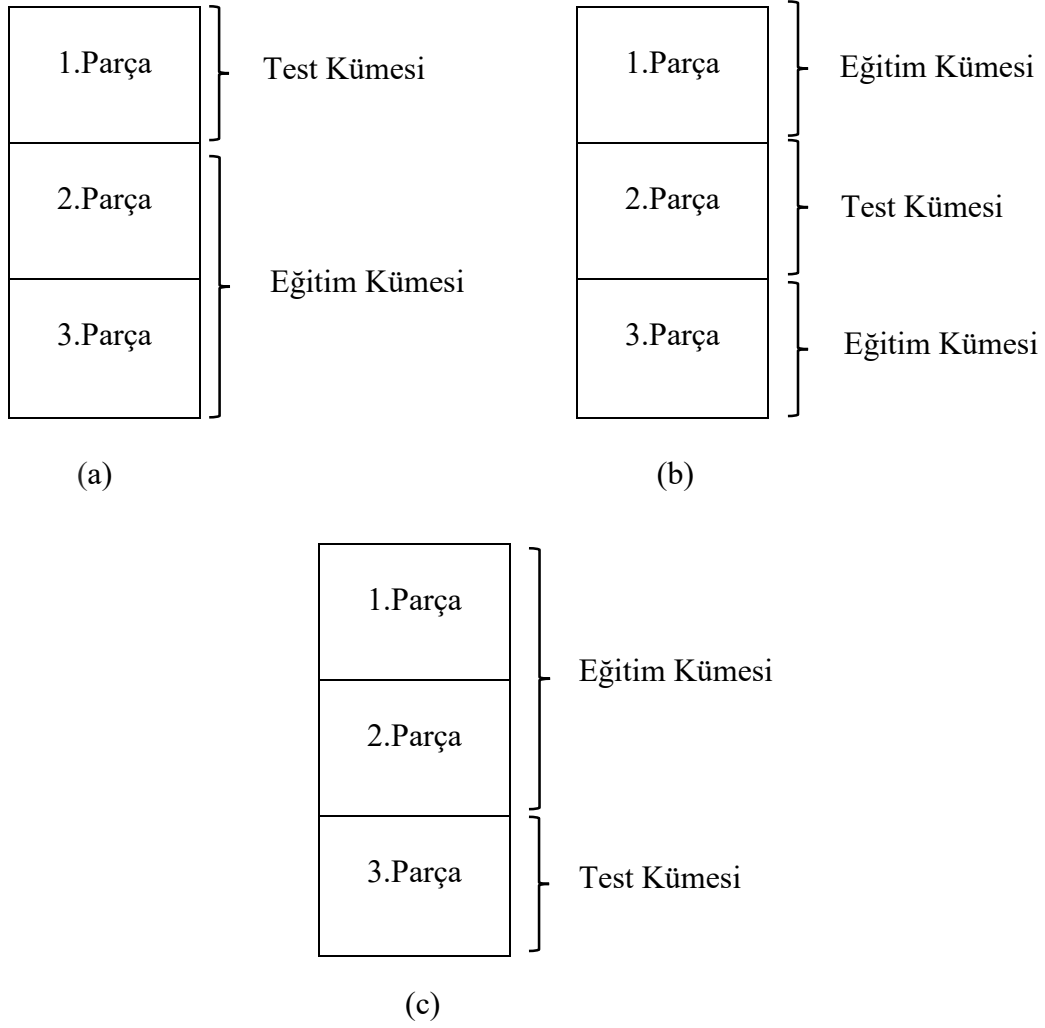
1.12. k-Katlı Çapraz Doğrulama Yöntemi ve Başarı Ölçümü

Her hangi bir amaç doğrultusunda oluşturulan sistemden elde edilen verilerin doğruluklarının ölçülmesiyle sistemin başarı durumu belirlenebilir. Bu amaçla kullanılan en yaygın yöntemlerden bir tanesi ise k-katlı çapraz doğrulama sistemidir. Bu yöntemde D ile gösterilen bir veri kümesi $D_1, D_2, D_3, \dots, D_k$ olmak üzere ve her birisi yaklaşık eşit sayıda olan k sayıda alt veri kümesine ayrılır (Şekil 1.20.).



Şekil 1.20. Veri kümesinin alt parçalara ayrılması

Yöntem veri üzerinde modelleme yaparken her seferinde bir küme ‘test kümesi’ ve kalan kümeler ‘eğitim kümesi’ olacak şekilde ayarlamalar yapar (Şekil 1.21.). Böylelikle veri üzerindeki her küme ‘test kümesi’ ve ‘eğitim kümesi’ olarak ele alındığından sistemin başarısı daha sağlıklı değerlendirilmiş olur.



Şekil 1.21. (a) Birinci test-eğitim, (b) İkinci test-eğitim, (c) Üçüncü test-eğitim

Yapılan deneysel çalışmalarda k için optimum değerin 10 olduğu belirlenmiştir [85]. Ancak kullanılan verinin çokluğuna göre bu sayı 2 veya 5 olarak da kullanılabilir [86]. Bu çalışmada DVM ve k -EK için 10 katlı çapraz doğrulama yöntemi kullanılmıştır.

K tane eğitim ve test kümesiyle modellenen ve test edilen sistemin k tane başarı oranı elde edilmiş olur. Bu nedenle sistemin genel başarısını değerlendirmek için bu değerlerin ortalaması alınır. Sistemin sınıflandırma başarı oranı (SBO), doğru sınıflandırılan veri sayısının (t) toplam veri sayısına (n) oranlanmasıyla elde edilir ve aşağıda verildiği gibi ifade edilir.

$$SBO = \frac{t}{n} , \quad ortalama SBO = \frac{\sum_{i=1}^t SBO_i}{t} \quad (16)$$

2. YAPILAN ÇALIŞMALAR

2.1. Materyal

Bu çalışmada EmoSTAR adlı yeni bir veri seti kullanılmıştır. EmoSTAR TV ve internetten faydalanılarak oluşturulmuş Türkçe ve İngilizce örneklerin bulunduğu bir veri setidir. İki farklı dil örneğinin olması duygunun telaffuz ve dilden bağımsız olduğunu göstermek için elverişlidir.

EmoSTAR toplam 393 örneğe sahip Kızgın, Mutlu, Nötr ve Üzgün duygu kategorilerine ayrılmış bir veri setidir. Kızgın örnekler sinema ve dizilerden, Mutlu örnekler ödül törenlerinde konuşma yapan aktörlerden, Nötr örnekler haber spiker konuşmalarından ve Üzgün örnekler ise internette bulunan videolardan alınarak oluşturulmuştur. Nötr ve Mutlu örnekler doğal konuşmalardır üzgün örneklerin ise çok azı hariç çoğu doğal konuşmadır. Kızgın örnekler ise yapay konuşmalardır. Duygu etiketlemesi ise yazar tarafından video isim ve örnekleri de göz önüne alınarak sesli ve görsel değerlendirme sonucunda kategorik olarak yapılmıştır. EmoSTAR veri setine ait bilgiler Tablo 1.1. de gösterilmiştir.

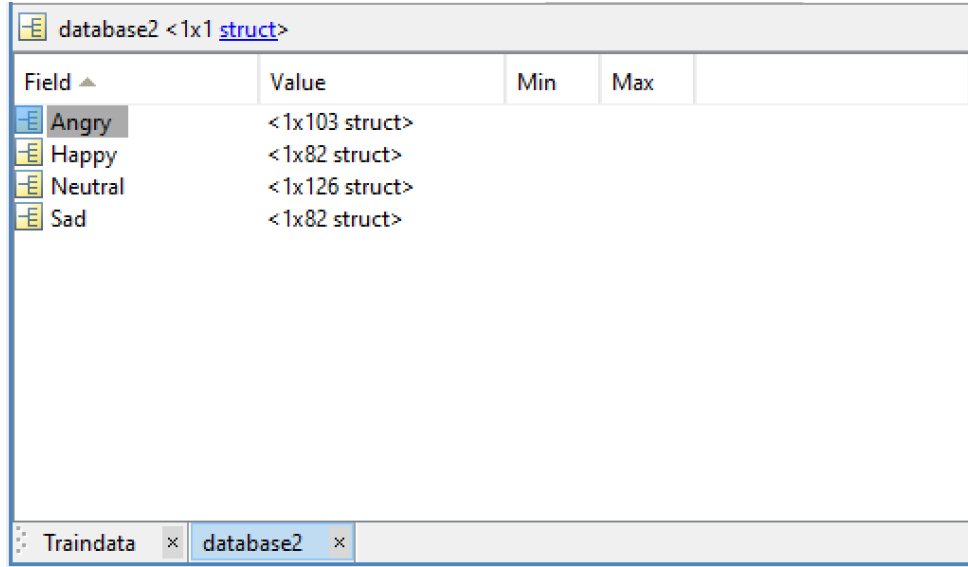
Ses sinyalinden öznitelik çıkarmada ve çıkarılan bu özniteliklere göre sınıflandırma işlemi MATLAB programı kullanılarak gerçekleştirildi. MATLAB, dünya çapında bilim adamları ve mühendisler tarafından benimsenmiş, yüksek düzeyde hesaplama ve grafikleme işlemi yapabilen bir programdır. Karmaşık matematiksel ve istatistiksel işlemler için hazır kütüphaneler barındırmakla beraber, matrise dayalı programlama yapısı ile kolay ve hızlı bir şekilde işlem yapabilmektedir. Öznitelik çıkarma aşamasında literatürde kullanılan OpenSMILE programından da faydalanılabilir [84].

MATLAB ortamında geliştirilen fonksiyonları oluştururken, ses sinyallerini işlemede birçok genel fonksiyon içeren VOICEBOX kütüphanesinden faydalanıldı. Bu kütüphane İngiltere Imperial Koleji Elektrik Elektronik Mühendisliği bölümünden Mike Brookes tarafından hazırlanmıştır.

2.2. Yöntem

Literatürde konuşma sinyallerinin duygusal sınıfının belirlenmesinde kullanılan en yaygın ve klasik olanı, konuşma sinyallerindeki duygu içeriğinin kategorik olarak (Mutlu, Kızgın, Neşeli vb.) sınıflandırılmasıdır. Bu tez çalışması kapsamında EmoSTAR veri setinde dört farklı duygu (Kızgın, Mutlu, Nötr, Üzgün) kategorisinde bulunan ses sinyalleri kullanılmıştır. Veri setinde bulunan 393 ses örneğine ait öznitelikler çıkarılarak database isminde bir kümeye kaydedilmiştir. Böylelikle herhangi bir duygu sınıfına ait herhangi bir öznitelik veya öznitelikler istenildiği zaman kayıtlı olan veri setinden elde edilebilir hale getirilmiştir. Bu yöntemdeki amaç hem işlem yükünü azaltmak hem de özniteliklerin hepsini değil de bir kısmının kullanılması istendiğinde veya duygu sınıflarından bir kısmını kullanmak gerektiğinde daha hızlı işlem gerçekleştirebilmektir. Veri seti ilk başta gelişmiş güzel olarak Eğitim (283 adet) ve Test (110 adet) kümelerine ayrıldı. Bu durumda sabit veriler üzerinde modelleme gerçekleştirilip yine sabit test verileri üzerinde sınıflandırma işlemi gerçekleştirildi. Bu yöntemin en büyük dezavantajı sürekli aynı veriler üzerinde model geliştirilip test edildiği için sağlıklı sonuçlar vermemektedir. Bu nedenle çapraz doğrulama yöntemi kullanılarak tüm veri setinin modellenmesi ve test edilmesi işlemleri gerçekleştirildi.

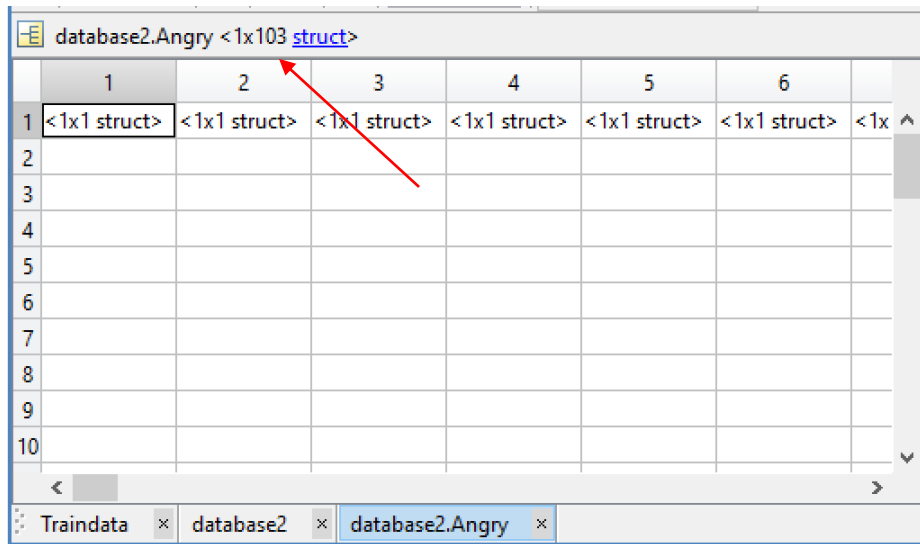
Oluşturulan bu veri setinden her bir duygu kategorisine ait ses sinyalleri ve bu sinyallerden çıkarılan öznitelikler bulunmaktadır. Şekil 2.1.'de Eğitim kümesi içerisinde yer alan farklı duygu kategorileri ve Şekil 2.2. ve Şekil 2.3., Şekil 2.4.'de ise sırasıyla kızgın sınıfa ait ses örnekleri ve bu ses örneklerine ait öznitelikler bulunmaktadır.



Field	Value	Min	Max
Angry	<1x103 struct>		
Happy	<1x82 struct>		
Neutral	<1x126 struct>		
Sad	<1x82 struct>		

Şekil 2.1. Eğitim kümesinde yer alan duygu kategorileri

Şekil 2.1.'de görüleceği gibi database (Toplam 393 örnek) içerisinde yer alan Kızgın (103 örnek), Mutlu (82 örnek), Nötr (126 örnek) ve Üzgün (82 örnek) duygu sınıflarına ait veri gurupları bulunmaktadır.



	1	2	3	4	5	6
1	<1x1 struct>	<1x1 struct>	<1x1 struct>	<1x1 struct>	<1x1 struct>	<1x1 struct>
2						
3						
4						
5						
6						
7						
8						
9						
10						

Şekil 2.2. Kızgın sınıfa ait 103 adet ses örneğinin her biri

Field	Value	Min	Max
wav	<122773x1 double>	-0.4846	0.5430
mfcc	<958x12 double>	-10.77...	6.9728
mfcc_mean	<1x12 double>	-6.3751	1.0743
mfcc_med	<1x12 double>	-6.5800	1.0807
mfcc_std	<1x12 double>	0.6379	2.1477
mfcc_skv	<1x12 double>	-0.1195	0.5517
mfcc_max	<1x12 double>	0.1932	6.9728
mfcc_min	<1x12 double>	-10.77...	-0.7331
mfcc_kurt	<1x12 double>	2.6630	3.7625
mfcc_range	<1x12 double>	3.9863	11.9651
mfcc_E_mean	<1x13 double>	-6.3751	4.4126
mfcc_E_med	<1x13 double>	-6.5800	5.0301
mfcc_E_std	<1x13 double>	0.6379	2.1477
mfcc_E_skv	<1x13 double>	-1.1412	0.5517
mfcc_E_max	<1x13 double>	0.1932	6.9728
mfcc_E_min	<1x13 double>	-10.77...	-0.7331
mfcc_E_kurt	<1x13 double>	2.6630	3.7875
mfcc_E_range	<1x13 double>	3.9863	11.9651
mfcc_E0_mean	<1x14 double>	-6.3751	4.4126
mfcc_E0_med	<1x14 double>	-6.5800	5.0301
mfcc_E0_std	<1x14 double>	0.6379	3.5816
mfcc_E0_skv	<1x14 double>	-1.1412	0.5517
mfcc_E0_max	<1x14 double>	0.1932	6.9728
mfcc_E0_min	<1x14 double>	-14.80...	-0.7331
mfcc_E0_kurt	<1x14 double>	2.6630	4.0297
mfcc_E0_range	<1x14 double>	3.9863	21.0075
mfcc_Ed_mean	<1x26 double>	-6.3751	4.4126
mfcc_Ed_med	<1x26 double>	-6.5800	5.0301
mfcc_Ed_std	<1x26 double>	0.0987	2.1477
mfcc_Ed_skv	<1x26 double>	-1.1889	1.3991

Şekil 2.3. Kızgın sınıfa ait birinci ses örneğinden çıkarılan öznitelikler

Şekil 2.3. deki wav matrisi birinci ses sinyalinin matris halini, Mfkk ise birinci sinyalinden elde edilen mel frekansı kepsrum katsayılarını, Mfkk_mean ile gösterilen özellik 958x12 lik Mfkk özneliğinin sütun bazında ortalaması (mean) alınarak elde edilmiş başka bir özneliği temsil etmektedir. Bunu gibi Mfkk_med, Mfkk özneliğinin sütun bazında medyanını, Mfkk_std ise standart sapmasını, Mfkk_skv skewnessını, Mfkk_max maksimumunu, Mfkk_min minimumunu, Mfkk_kurt kurtosisini ve Mfkk_range ise range istatistiksel özelliğini temsil etmektedir. Yine şekil 2.20.'de Mfkk_E mel frekansı kepsrum katsayılarına Enerji özelliği eklenerek elde edilmiş 13 özellikli matrisi, Mfkk_E0 ise Mfkk'ye Enerji ve sıfırıncı kepsrum katsayısı eklenerek elde edilmiş 14 özellikli matrisi temsil etmektedir. Şekilde gösterilen Mfkk_Ed, Mfkk'ye Enerji eklenerek (13 öznelik) ve

bu 13 boyutlu matrisin birinci türevi (d) eklenerek (13 öznitelik) elde edilmiş 26 özellikli matrisi temsil etmektedir.

database2.Angry(1,1) <1x1 struct>				
Field ▲	Value	Min	Max	
mfcc_E0_kurt	<1x14 double>	2.6630	4.0297	
mfcc_E0_range	<1x14 double>	3.9863	21.0075	
mfcc_Ed_mean	<1x26 double>	-6.3751	4.4126	
mfcc_Ed_med	<1x26 double>	-6.5800	5.0301	
mfcc_Ed_std	<1x26 double>	0.0987	2.1477	
mfcc_Ed_skv	<1x26 double>	-1.1889	1.3991	
mfcc_Ed_max	<1x26 double>	0.1932	6.9728	
mfcc_Ed_min	<1x26 double>	-10.77...	-0.2674	
mfcc_Ed_kurt	<1x26 double>	2.6630	7.5086	
mfcc_Ed_range	<1x26 double>	0.5528	11.9651	
mfcc_E0d_mean	<1x28 double>	-6.3751	4.4126	
mfcc_E0d_med	<1x28 double>	-6.5800	5.0301	
mfcc_E0d_std	<1x28 double>	0.0987	3.5816	
mfcc_E0d_skv	<1x28 double>	-1.1889	1.3991	
mfcc_E0d_max	<1x28 double>	0.1932	6.9728	
mfcc_E0d_min	<1x28 double>	-14.80...	-0.2674	
mfcc_E0d_kurt	<1x28 double>	2.6630	7.5086	
mfcc_E0d_range	<1x28 double>	0.5528	21.0075	
mfcc_E0dD_mean	<1x42 double>	-6.3751	4.4126	
mfcc_E0dD_med	<1x42 double>	-6.5800	5.0301	
mfcc_E0dD_std	<1x42 double>	0.0433	3.5816	
mfcc_E0dD_skv	<1x42 double>	-1.1889	1.3991	
mfcc_E0dD_max	<1x42 double>	0.1440	6.9728	
mfcc_E0dD_min	<1x42 double>	-14.80...	-0.1379	
mfcc_E0dD_kurt	<1x42 double>	2.6630	8.1722	
mfcc_E0dD_range	<1x42 double>	0.2818	21.0075	
lpc	[1 -2.7958 4.5132 -5.2...	-5.2395	4.7019	
Pitch	[208.9096 200 49.7723...	1.3480	372.09...	
formant	[628.3605 1.1437e+03...	628.36...	1.9617...	
Hu_moment	<1x28 double>	-0.0141	0.0058	

Şekil 2.4. Kızgın sınıfa ait birinci ses örneğinden çıkarılan öznitelikler

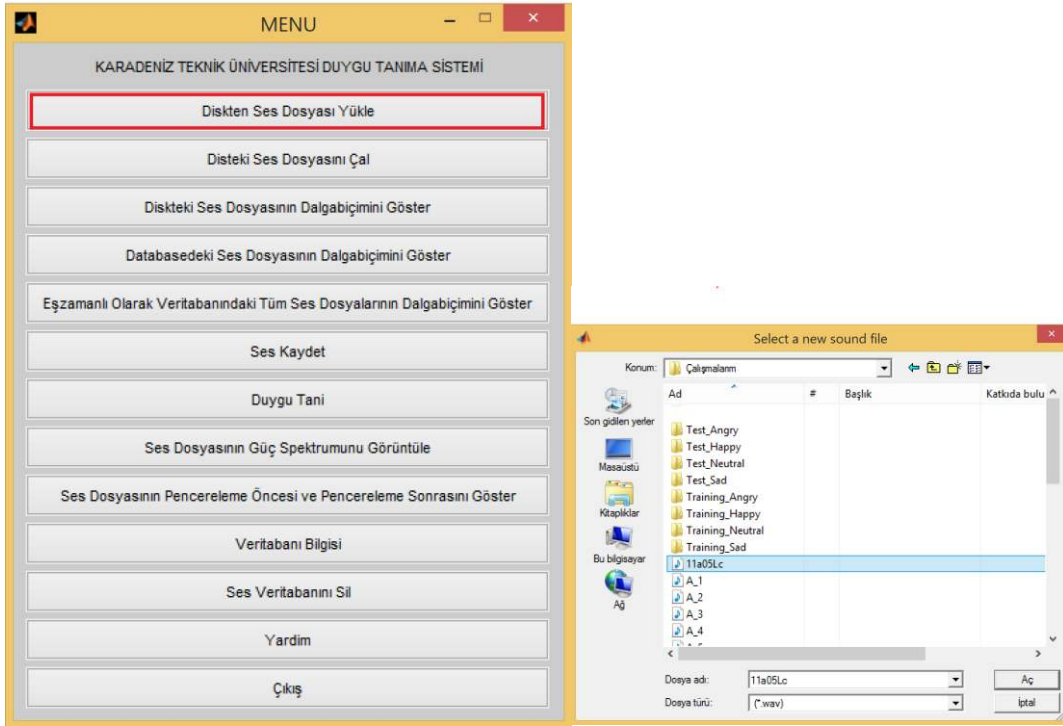
Şekil 2.4.'de Mfkk_E0d, Mfkk'ye enerji, sıfırcı Mfkk eklenerek (14 öznitelik) birinci türevi (d) alındıktan sonra (14 öznitelik) elde edilen 28 özellikli matrisi, Mfkk_E0dD ise Mfkk_E0d ye ek olarak ikinci türevin (D) eklenmesiyle (14 öznitelik) elde edilmiş 42 özellikli matrisi ifade etmektedir. Şekil 2.21.'de DÖK ile 8 özellikli doğrusal öngörü katsayıları, Pitch ile 8 özellikli ses perdesi özelliği, formant ile 3 özellikli ses formantı, Hu_moment ise 28 özellikli Hu Momentleri temsil edilmektedir.

Bu tez çalışmasında EmoSTAR veri setine ait örnekler kullanılmış ve sonuçlar elde edilmiştir. Bunun yanında oluşturulan bu sisteme dışardan bir ses dosyası yüklenerek veya ses kaydı gerçekleştirilerek duygu analizi yapabilmek için gerçek zamanlı KARADENİZ TEKNİK ÜNİVERSİTESİ DUYGU TANIMA SİSTEMİ oluşturuldu. Bu sisteme ait ara yüz Şekil 2.5.'de görülmektedir.

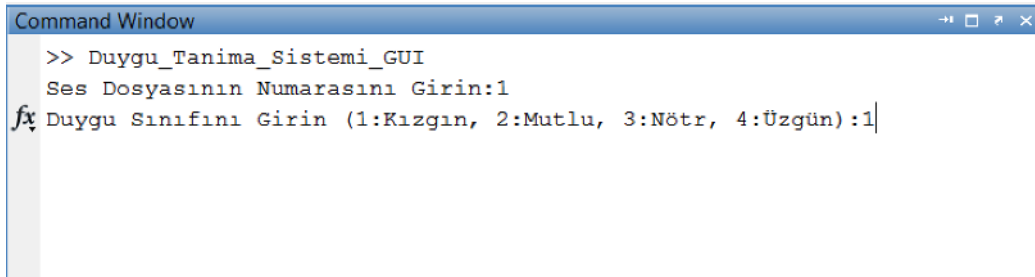


Şekil 2.5. Karadeniz Teknik Üniversitesi Duygu Tanıma Sistemi ara yüzü

Oluşturulan bu sistem sayesinde dışardan ses dosyaları girilerek küçük bir veri seti oluşturulabilir ve bu veri setindeki örneklerin duygu analizi yapılabilir, bu ses dosyalarına ait dalga biçimi, güç spektrumu, pencereleme öncesi ve sonrası gibi çeşitli grafikler de görülebilir. Yine oluşturulan bu ara yüz sayesinde gerçek zamanlı ses kaydı ve bu kaydın duygu analizi gerçekleştirilebilir.

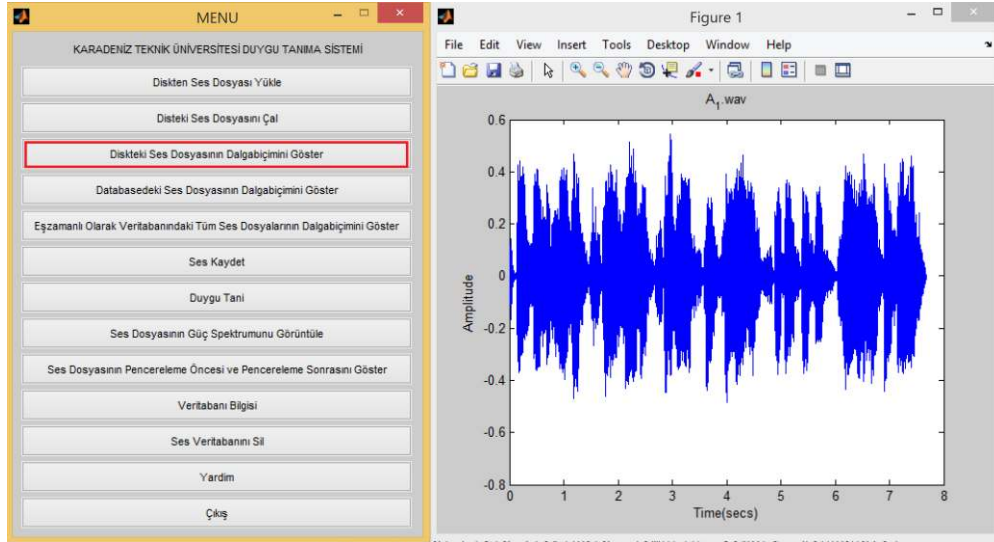


Şekil 2.6. Diskten ses dosyası yükleme işlemi



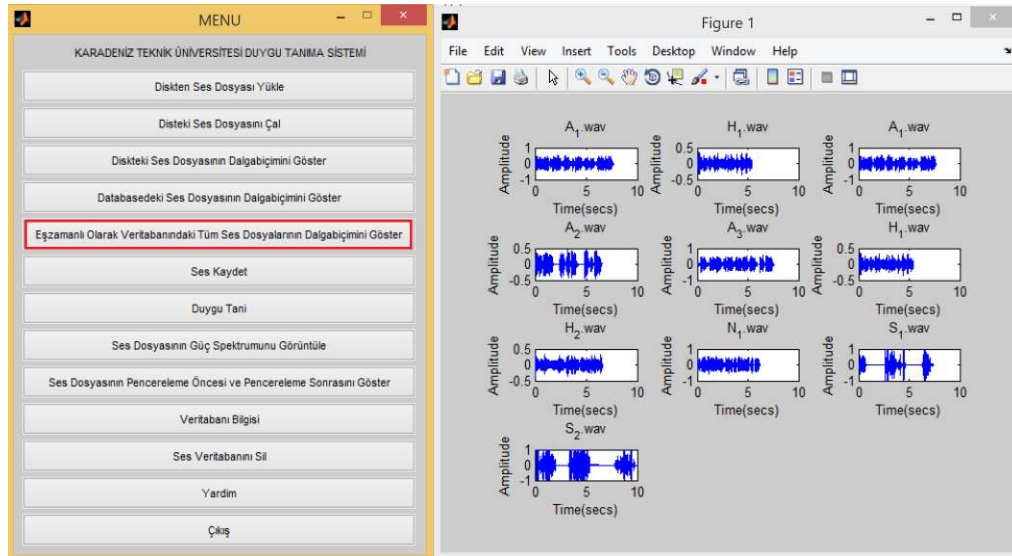
Şekil 2.7. Diskten ses dosyası yükleme işlemi

Diskten duygu analizi yapılmak istenen ses dosyası-dosyaları kaydedilerek bu örneklerle ait duygu analizi ve başka diğer işlemler yapılabilir.



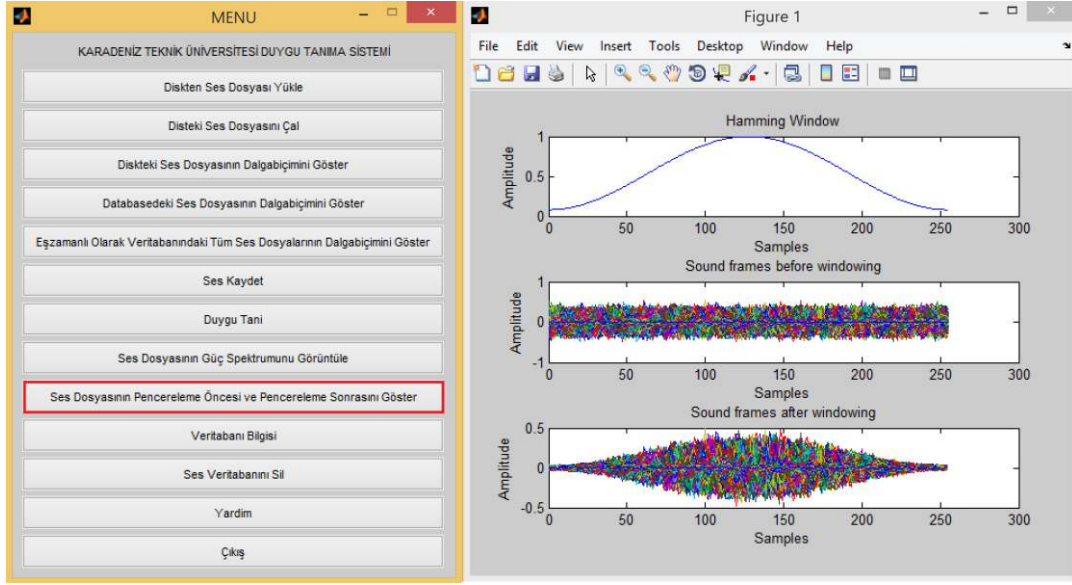
Şekil 2.8. Kaydedilen ses dosyasına ait dalga biçimi

Şekil 2.8.'de kaydedilen ses dosyasına ait dalga biçimi görülmektedir. Ses analizi yaparken dalga biçimi sıklıkla görüntülenmektedir. Bu ara yüz sayesinde bu işlem hızlı ve kolay bir şekilde gerçekleştirilmiş olur.



Şekil 2.9. Oluşturulan veri tabanındaki ses örneklerine ait dalga biçimleri

Şekil 2.9.'da ise kaydedilen ses dosyalarına ait dalga biçimlerinin toplu olarak gösterildiği ara yüz ve grafik yer almaktadır.



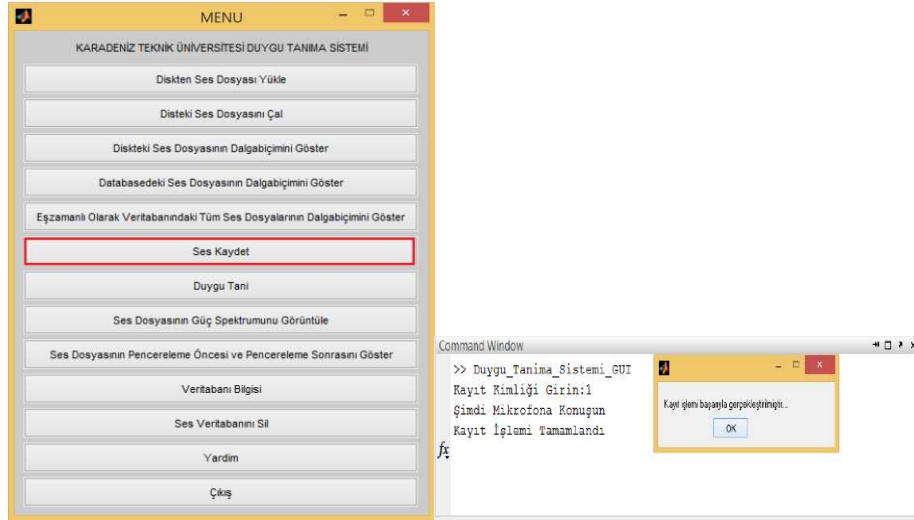
Şekil 2.10. Ses dosyasına ait pencereleme işlemi

Şekil 2.10.'da Mfkk özelliğinin çıkarımında kullanılan pencereleme işlemi öncesi ve sonrasını gösteren grafik bulunmaktadır. Pencerelemenin amacı çerçevelerin başında ve sonundaki süreksizliği gidermektir.

```
Command Window
Dosya:A_1.wav
Konum:C:\Users\onurerdem\Desktop\Çalışmalarım\
Duygu:Kızgın
Ses Numarası:1
-
Dosya:A_2.wav
Konum:C:\Users\onurerdem\Desktop\Çalışmalarım\
Duygu:Kızgın
Ses Numarası:2
-
Dosya:A_3.wav
Konum:C:\Users\onurerdem\Desktop\Çalışmalarım\
Duygu:Kızgın
Ses Numarası:3
-
Dosya:H_1.wav
Konum:C:\Users\onurerdem\Desktop\Çalışmalarım\
Duygu:Mutlu
Ses Numarası:4
-
Dosya:H_2.wav
Konum:C:\Users\onurerdem\Desktop\Çalışmalarım\
Duygu:Mutlu
Ses Numarası:5
-
Dosya:N_1.wav
Konum:C:\Users\onurerdem\Desktop\Çalışmalarım\
Duygu:Nötr
```

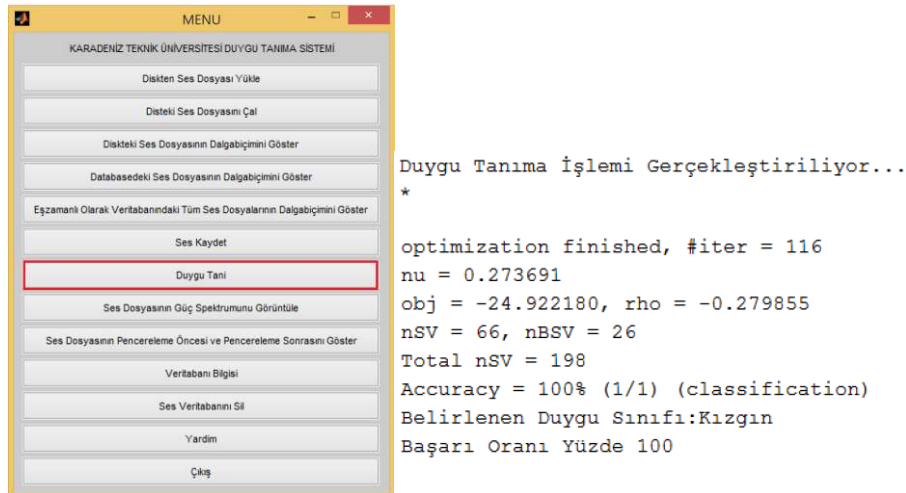
Şekil 2.11. Veri tabanı bilgisi

Oluşturulan veri tabanına ait özellikler Şekil 2.11.'de görülmektedir. Böylece oluşturulan veri tabanında kaç örneğin bulunduğu ve bu örneklerin duygu sınıfı gibi bilgilere ulaşmak mümkün olmaktadır.



Şekil 2.12. Ses kaydetme işlemi

Bu ara yüzün oluşturulmasındaki amaçlardan bir tanesi de gerçek zamanlı konuşmadan duygu tanıma işlemi gerçekleştirmektir. Bu sayede EmoSTAR veri setinden geliştirilen modelle bilgisayar başındaki konuşmacının duygu analizi gerçekleştirilmiş olacaktır. Duygu tanıma ara yüzü ise Şekil 2.13.'de görülmektedir.



Şekil 2.13. Duygu analizi

2.3. Kategorik Değerlendirme ve Öznitelikler

Bu tez çalışmasında kategorik duygu bazında Kızgın, Mutlu, Nötr ve Üzgün olmak üzere dört farklı duygu üzerinde çalışılmıştır (Tablo 2.1.).

Tablo 2.1. Duygu kategorileri

Duygu Sınıfı	Bayan		Erkek		TOPLAM
	Türkçe	İngilizce	Türkçe	İngilizce	
Kızgın	0	40	30	33	103
Mutlu	0	37	0	45	82
Nötr	20	37	34	35	126
Üzgün	19	51	0	12	82
Toplam	39	165	64	125	393
TOPLAM	204		189		

Yukardaki duygu örneklerine ait ses sinyallerinden çıkarılan öznitelikler ise Tablo 2.2.'de gösterilmiştir. Burada duygu tanıma sistemlerinde sıklıkla kullanılan Mel Frekans Kepstrum Katsayılarının ilk 12 tanesi ve bunlara ait istatistiksel hesaplamalar, Mel Frekans Kepstrum Katsayılarına ilaveten 0.kepstral katsayısı, enerji, birinci türev ve ikinci türevden oluşan toplam 46 özniteliğe sahip mel frekansı tabanlı öznitelikler bulunmaktadır. Ayrıca duygu analizi alanında kullanılmaya yeni başlanılan Hu Momentleri özniteliği de bulunmaktadır [88]. Yine bu alanda sıklıkla kullanılan Doğrusal Öngörü Katsayıları ve Temel Frekans öznitelikleri de Tablo 2.2.'de yer almaktadır.

Tablo 2.2. Ses sinyallerinden çıkarılan öznitelikler

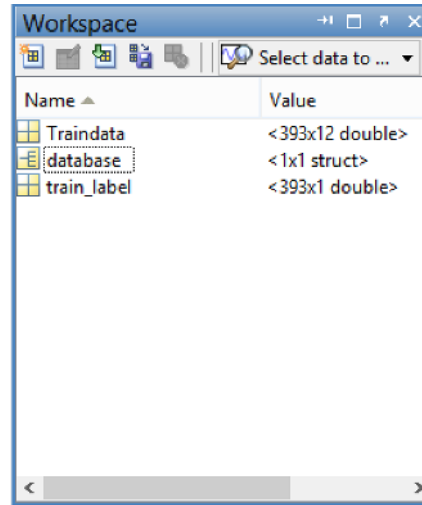
ÖZNİTELİK İSMİ	Kısaltması	İSTATİSTİKSEL ÖZELLİKLER	Top
Mel Frekans Kepstrum Katsayıları(12)	Mfkk	ortalama(12),medyan(12),standartsapma(12) skewness(12), kurtosis(12), minimum(12), maksimum(12), range(12)	96
Mel Frekans Kepstrum Katsayıları(12) + Enerji (1)	Mfkk_E	ortalama(13),medyan(13),standartsapma(13) skewness(13), kurtosis(13), minimum(13), maksimum(13), range(13)	104
Mel Frekans Kepstrum Katsayıları(12) + Enerji (1) + 0.Mel katsayısı (1) + 1.Türev (14) + 2.Türev (14)	Mfkk_E0dD	ortalama(42),medyan(42),standartsapma(42) skewness(42), kurtosis(42), minimum(42), maksimum(42), range(42)	336
Doğrusal Öngörü Katsayıları (8)	DÖK		8
Hu Momentleri	Hu		28

Ses sinyallerinden çıkarılan bu özelliklerin çeşitli istatistiksel hesaplamaları (ortalama, medyan, standart sapma, skewness, kurtosis, minimum, maksimum, range) da eklenerek öznitelikler çeşitlendirilmiştir. Bu özniteliklerin sınıflandırılmasında ise k-En Yakın Komşu (k-EK) ve Destek Vektör Makineleri (DVM) sınıflandırıcıları kullanılmış ve bu iki sınıflandırıcının başarı oranları 10 kat çapraz doğrulama yöntemi kullanılarak karşılaştırılmıştır.

3. BULGULAR

Bu çalışmada, 393 örnekten oluşan veri seti, rastgele alınan ‘eğitim kümesi’ (283 tane örnek) ve ‘test kümesi’ (110 tane örnek) olarak iki alt kümeye ayrılır. Ancak bu durumda sürekli sabit veriler modellenip test edildiği için sistemin genel başarısı hakkında pek sağlıklı sonuçlar vermemektedir. Bunun için bu çalışmada veri setinin tamamı kullanılarak öznitelikler çıkarılmış ve bunlar 10 kat çapraz doğrulama yöntemiyle test edilmiştir. Örneğin Şekil 3.1.’de EmoSTAR veri setinde bulunan tüm veriler (393 adet) kullanılarak çıkarılan Mfkk_mean (12) özneliğine ait veri kümesi görülmektedir.

Verilen test kümesindeki ses işaretlerinin Kızgın, Mutlu, Nötr ve Üzgün olan duygu sınıflarından hangisine ait olduğunun sınıflandırılması için DVM ve k-EK yöntemleri kullanılarak sonuçlar elde edilmiştir.



The screenshot shows a MATLAB Workspace window with the following variables:

Name	Value
Traindata	<393x12 double>
database	<1x1 struct>
train_label	<393x1 double>

Şekil 3.1. Eğitim kümesi

3.1. Destek Vektör Makineleri Sınıflandırıcısına Ait Sonuçlar

Bu bölümdeki sonuçlar Matlab ortamı için LIBSVM kütüphanesinden faydalanılarak [89] hazırlanan Destek Vektör Makineleri sınıflandırıcısına aittir. Daha önceki bölümlerde de bahsedildiği gibi sınıflandırıcı başarısının daha güvenilir olması için 10-katlı çapraz doğrulama yöntemi kullanılmıştır. Bu yöntem 25 kez çalıştırılarak en iyi sonuç ve elde edilen 25 sonucun ortalaması alınarak iki sonuç elde edilmiştir.

Tablo 3.1. Özniteliklere ait DVM sınıflandırma başarı sonuçları

Öznitelik	Özellik sayısı	En iyi sonuç (%)	Ortalama sonuç (%)
Mel Frekansı Kepstrum Katsayısı (Mfkk_mean)	12	99.5	98.7
Mfkk_E0dD (mean)	42	94	92.68
Mfkk_E	13	94	92.28
Hu Momentleri	28	79	77.38
Doğrusal Öngörü Katsayısı (DÖK)	8	73.5	70.5

Tablo 3.1.'deki öznitelikler içerisinde en başarı sonucu mel frekansı kepstrum katsayıları vermektedir. Diğer göze çarpan sonuç ise bu alanda yeni yeni kullanılmaya başlanan hu momentleri özniteliğinin kayda değer bir sonuç vermesidir.

Tablo 3.2.'de ise elde edilen öznitelikler bir birlerine eklenerek elde edilmiş çok boyutlu öznitelik tablosu görülmektedir. Yine bu tablodan mel frekansı kepstrum katsayıları özniteliği ve çeşitli istatistiki hesaplamaları eklenerek oluşturulmuş yeni özniteliklerin başarı oranının yüksek olduğu görülmektedir. Ayrıca Mfkk özniteliğine doğrusal öngörü katsayıları ve hu momentleri eklenerek özellik vektörü zenginleştirilmiş ve başarılı sonuçlar elde edilmiştir.

Tablo 3.2. Çoklu ve istatistiksel özniteliklere ait DVM sınıflandırma sonuçları

Öznitelik	Özellik sayısı	En iyi sonuç (%)	Ortalama sonuç (%)
Mfkk_mean(12), Mfkk_median(12), Mfkk_std(12), Mfkk_skewness(12), Mfkk_kurtosis(12), Mfkk_max(12), Mfkk_min(12), Mfkk_range(12),	96	97.5	96.9
MfkkE_mean(13), MfkkE_median(13), MfkkE_std(13), MfkkE_skewness(13), MfkkE_kurtosis(13), MfkkE_max(13), MfkkE_min(13), MfkkE_range(13), DÖK(8)	112	97	96.6
Mfkk_mean(12), DÖK(8)	20	96.5	95.9
Hu momentleri(28), Mfkk_mean(12)	40	96	94.5
Mfkk_mean(12), Mfkk_median(12), Mfkk_std(12), Mfkk_skewness(12), Mfkk_kurtosis(12), Mfkk_max(12), Mfkk_min(12), Mfkk_range(12), DÖK(8)	104	97.5	92.1

Yine Tablo 3.2.'den görüleceği gibi duygu tanıma çalışmalarında kullanılan en iyi öznitelik mel frekansı kepsrum katsayılarıdır. Diğer dikkat çeken bir durum ise ses sinyalleri çalışmalarında yeni yeni kullanılmaya başlanan Hu Momentleri yönteminin, bu alanda sıklıkla kullanılan doğrusal öngörü katsayıları ve temel frekans gibi özniteliklere göre daha iyi sonuç verdiğidir. Bu yöntem daha da geliştirilerek daha iyi sonuçlar elde edilebilir.

Tablo 3.3. DVM için farklı çerçeve uzunluğu ve kaydırma süresi sonuçları

Özellik İsmi	Özellik sayısı	En iyi sonuç (%)	Ortalama Sonuç (%)
Mfkk_20_10	12	98.5	98.3
Mfkk_30_10	12	99.5	98.7
Mfkk_40_10	12	97.5	96.3
Mfkk_50_30	12	99	97.7
Mfkk_30_20	12	99	97.8

Tablo 3.3.'de ise farklı çerçeve uzunluğu ve kaydırma süresine bağlı olarak Destek Vektör Makineleri kullanılarak hesaplanan başarı sonuçları görülmektedir. Buradaki birinci sayı değeri çerçeve uzunluğunu ikinci sayı değeri ise kaydırma süresini göstermektedir. Yani Mfkk_20_10 özelliğinde çerçeve uzunluğu 20ms ve kaydırma miktarı ise 10ms dir. Öznitelik sayısı ise Mel Frekansı Kepstrum Katsayısı adedini göstermektedir. Tablodan da görüleceği gibi en ideal değerlerin çerçeve süresi için 30ms, kaydırma süresi içinse 10ms olduğu görülmektedir.

3.2. k-EK Sınıflandırıcısına Ait Sonuçlar

Bu tez çalışmasında kullanılan sınıflandırıcılardan bir tanesi de k En Yakın Komşu yöntemidir ve k değeri literatürdeki çalışmalar dikkate alınarak 5 olarak seçilmiştir. Yine bu yöntemde 10 katlı çapraz doğrulama yöntemi kullanılmış ve 25 kez çalıştırılarak en iyi sonuç ve elde edilen 25 sonucun ortalaması alınarak iki sonuç elde edilmiştir.

Tablo 3.4. Özniteliklere ait k-EK sınıflandırma başarı sonuçları

Öznitelik	Özellik sayısı	En iyi sonuç (%)	Ortalama sonuç (%)
Mel Frekansı Kepstrum Katsayısı (Mfkk_mean)	12	95.1	94.1
Mfkk_E0dD	42	81.9	80.9
Mfkk_E	13	87.2	85.9
Hu Momentleri	28	75.1	72.5
Doğrusal Öngörü Katsayısı (DÖK)	8	67.4	65.4

Tablo 3.4. ve Tablo 3.5. den görüleceği gibi k En Yakın Komşu yönteminde de en iyi sınıflandırma başarısı Mel Frekansı Kepstrum Katsayıları özniteliği ile olmaktadır. Hu Momentleri özniteliği sınıflandırmada önem açısından yine ikinci sırada yer almaktadır. Eklenen yeni özniteliklerin işlem süresini uzatmasının yanında, bazı durumlarda ortalama başarı oranına da katkısı olabilmektedir. Sadece mel frekansı kepsrum katsayılarının ortalaması alınarak elde edilen özellik vektörü en iyi başarı oranını vermektedir.

Tablo 3.5. Çoklu ve istatistiksel özniteliklere ait k-EK sınıflandırma sonuçları

Öznitelik	Özellik sayısı	En iyi sonuç (%)	Ortalama sonuç (%)
Mfkk_mean(12), Mfkk_median(12), Mfkk_std(12), Mfkk_skewness(12), Mfkk_kurtosis(12), Mfkk_max(12), Mfkk_min(12), Mfkk_range(12),	96	93.3	92.3
MfkkE_mean(13), MfkkE_median(13), MfkkE_std(13), MfkkE_skewness(13), MfkkE_kurtosis(13), MfkkE_max(13), MfkkE_min(13), MfkkE_range(13), DÖK(8)	112	89.3	88.4
Mfkk_mean(12), DÖK(8)	20	95.1	94.5
Hu momentleri(28), Mfkk_mean(12)	40	95.1	94.1
Mfkk_mean(12), Mfkk_median(12), Mfkk_std(12), Mfkk_skewness(12), Mfkk_kurtosis(12), Mfkk_max(12), Mfkk_min(12), Mfkk_range(12), DÖK(8)	104	94.4	93.4

Tablo 3.1.-3.2. ve Tablo 3.4.-3.5. kıyaslandığı zaman Destek Vektör Makineleri sınıflandırıcısının başarı oranının, k En Yakın Komşu yöntemine göre daha başarılı olduğu görülmektedir. Ancak işlem zamanı açısından k-EK sınıflandırıcısının DVM sınıflandırıcısına göre üstünlüğü bulunmaktadır.

Tablo 3.6. k-EK için farklı çerçeve uzunluğu ve kaydırma süresi sonuçları

Özellik İsmi	Özellik sayısı	En iyi sonuç (%)	Ortalama Sonuç (%)
Mfkk_20_10	12	91.1	90.1
Mfkk_30_10	12	91.6	90.6
Mfkk_40_10	12	92.8	92.2
Mfkk_50_30	12	91.3	90.4
Mfkk_30_20	12	90.8	90

Tablo 3.6.'da ise farklı çerçeve uzunluğu ve kaydırma süresine bağlı olarak k En Yakın Komşu sınıflandırıcısı kullanılarak hesaplanan başarı sonuçları görülmektedir. Tablodan da görüleceği gibi en iyi sonucu çerçeve süresinin 40ms ve kaydırma süresinin 10ms olduğu değerler vermektedir.

3.3. Farklı Veri Setleri Kullanılarak Elde Edilen Sonuçlar

Bu tez çalışmasında ayrıca farklı iki veri seti kullanılarak farklı eğitim ve test kümeleri oluşturulmuştur. Daha önceki bölümlerde sadece EmoSTAR veri seti üzerinde eğitim işlemi yapılmış ve aynı veri seti örnekleri test edilmişti. Bu bölümde ise EmoSTAR veri seti üzerinde sistem öğrenmesi yapılmış, farklı bir veri seti olan EmoDB ile de bu sistem öğrenmesi test edilmiştir. Yani EmoSTAR ile eğitim kümesi oluşturulmuş, EmoDB ile de test kümesi oluşturularak sonuçlar elde edilmiştir. Literatür incelendiği zaman bu tarz çalışmaların başarı oranının çok yüksek olmadığı görülmektedir [41]. Bunun sebebi ise veri setlerindeki örneklerin farklı ortam ve şartlarda alınmasından kaynaklanmasıdır.

Yapılan çalışmada EmoDB test verilerinin duygu sınıflandırma işlemi için, öznitelik olarak mel frekansı kepsrum katsayısı kullanılmıştır. Birinci adımda sadece mel frekansı kepsrum katsayılarının ortalaması (Mfkk_mean (12)), ikinci adımda ise mel frekansına ait sekiz istatistiksel hesap kullanılarak (Mfkk_mean (12), Mfkk_med (12), Mfkk_std (12), Mfkk_skv (12), Mfkk_max (12), Mfkk_min (12), Mfkk_kurt (12), Mfkk_range (12)) toplam doksan altı öznitelik kullanılmış ve sonuçlar elde edilmiştir. Bu çalışmaya ait sonuçlar Tablo 3.7.'de görülmektedir.

Tablo 3.7. EmoDB test kümesi olarak kullanıldığında elde edilen sonuçlar.

Öznitelikler	Başarı Sonucu (%)
Mfkk_mean (12)	41.1
Mfkk_mean(12), Mfkk_median(12), Mfkk_std(12), Mfkk_skewness(12), Mfkk_kurtosis(12), Mfkk_max(12), Mfkk_min(12), Mfkk_range(12),	39.9

3.4. Temel Bileşen Analizi Kullanılarak Öznitelik Alt Küme Seçimi

Yapılan çalışmalarda çok yüksek boyutlu öznitelik vektörünün olması hem sınıflandırıcının başarısını, hem de işlem zamanını olumsuz yönde etkileyebilmektedir. Bunun için öznitelik vektörü içerisinde, sınıflandırmada en etkili olan öznitelik alt kümeleri seçilerek işlem zamanı azaltılmış ve başarı oranı artırılmış olur. Bu amaçla çeşitli öznitelik seçme yöntemleri bulunmaktadır. Bu tez çalışmasında Weka programı kullanılarak, Temel Bileşen Analizi yöntemi ile öznitelik seçme işlemi gerçekleştirilmiştir. Programa girdi olarak tüm özniteliklerin bulunduğu 336 boyutlu (393x336, 393:örnek sayısı, 336 tüm özniteliklerin sayısı) matris verilmiş ve program 1,2,3,4,5,6,7,8,9,10,212 ve 213. özniteliklerin en belirleyici olduğunu göstermiştir. Buradaki ilk on özellik mel frekansı keppstrum katsayılarının ortalamasının ilk on örneğini (Mfkk_mean (10)) ve 212. ve 213. öznitelikler ise mel frekansı keppstrum katsayılarına sıfırıncı katsayı, enerji, birinci türev ve ikinci türev eklenerek (Mfkk_E0dD) elde edilmiş matrisin iki elemanını temsil etmektedir. Elde edilen seçilmiş bu matrisle sonuçlar tekrarlandığında Mfkk_mean özneliğiyle elde edilen sonuca yakın bir sonuç elde edilmektedir. Bunun nedeni ise öznitelik seçmeyle elde edilen yeni matrisin elemanlarının çoğunun zaten Mfkk_mean'in elemanları olmasıdır. Elde edilen bu matrisle yapılan çapraz doğrulamalı sınıflandırma işleminde en yüksek başarı oranı %99.6 ve ortalama başarı oranı % 98,7'dir. Bu sonuçta yaklaşık olarak Mfkk_mean özneliğiyle elde edilen sonuca yakındır.

4. SONUÇLAR VE ÖNERİLER

Bu tez çalışmasında ses sinyalinin duygu tanıma için çeşitli öznitelikler kullanıp, Destek Vektör Makineleri ve k En Yakın Komşu sınıflandırıcılarıyla çapraz doğrulama yapılarak sonuçlar elde edilmiştir. Elde edilen sonuçlara göre Mel Frekansı Kepstrum Katsayıları özniteliğinin ortalaması alınarak elde edilen yeni öznitelikle, Destek Vektör Makineleri sınıflandırıcısı kullanılarak çapraz doğrulama işlemi sonunda %98.7 lik bir başarı oranı elde edilmiştir. Ayrıca Mfkk elde edilirken kullanılan çerçeve ve kaydırma süreleri değiştirilerek başarı oranı artırılmaya çalışılmıştır. Çıkarılan yüksek boyutlu öznitelik vektörüne Temel Bileşenler Analizi uygulanarak boyut indirgeme işlemi yapılmış, böylelikle işlem zamanı ve başarı oranında iyi sonuçlar elde edilmiştir.

Farklı sınıflandırıcılar ve öznitelikler kullanılarak başarı oranı artırılabilir. Diğer dikkat çeken bir durum ise ses sinyalleri çalışmalarında yeni yeni kullanılmaya başlanan Hu Momentleri yönteminin, bu alanda sıklıkla kullanılan doğrusal öngörü katsayıları ve temel frekans gibi özniteliklere göre daha iyi sonuç verdiğidir. Bu yöntem daha da geliştirilerek daha iyi sonuçlar elde edilebilir.

5. KAYNAKLAR

1. El Ayadi, M., S. Kamel, M. ve Karray, F., Survey on speech emotion recognition: Features, classification schemes, and databases, Pattern Recognition, 44, 3 (2011) 572-587.
2. Nicholson, J., Takahashi, K. ve Nakatsu, R., "Emotion recognition in speech using neural networks, Neural Computing & Applications, 9 (2000) 290-296.
3. Schuller, B., Rigoll, G. ve Lang, M., Speech emotion recognition combining acoustic features and linguistic information in a hybrid support vector machine - Belief Network architecture, 2004 Ieee International Conference on Acoustics, Speech, and Signal Processing, 1 (2004) 577-580.
4. France, D. J., Shiavi, R. G., Silverman, S., Silverman, M. ve Wilkes, D. M., Acoustical properties of speech as indicators of depression and suicidal risk, Ieee Transactions on Biomedical Engineering, 47, 7 (2000) 829-837.
5. Hansen, J. H. L. ve Cairns, D. A., Icarus - Source Generator Based Real-Time Recognition of Speech in Noisy Stressful and Lombard Effect Environments, Speech Communication, 16, 6 (1995) 391-422.
6. Sun, D. Q., Pan, J., Cao, Q. Y., Li, T. ve Yang, F., Ubiquitous computing service model based on SPKI/SDSI, Dynamics of Continuous Discrete and Impulsive Systems-Series B-Applications & Algorithms, 13, 12 (2006) 2218-2223.
7. Banse, R. ve Scherer, K. R., Acoustic profiles in vocal emotion expression, Journal of Personality and Social Psychology, 70, 3 (1996) 614-636.
8. Kleinginna, P. R. ve Kleinginna, A. M., A categorized list of emotional definitions with suggestions for a consensual definition, Motivation Emotion, 4 (1981) 345-379.
9. Fernandez, R., A computational model for a automatic recognition of affect in speech, Ph.D Thesis, Massachusetts Institute of Technology, 2004.
10. Bradley, M. M. ve Lang, P. J., Measuring Emotion - the Self-Assessment Mannequin and the Semantic Differential, Journal of Behavior Therapy and Experimental Psychiatry, 25, 3 (1994) 49-59.
11. Williams, C. E. ve Stevens, K. N., Emotions and Speech - Some Acoustical Correlates, Journal of the Acoustical Society of America, 52 (1972) 1238-&.
12. Chan, J., The generation of affect in synthesized speech, J. Am. Voice Input/Output Soc., 8 (1990) 1-19.

13. Morris, J. D., Observations: SAM: the Self-Assessment Manikin; an efficient cross-cultural measurement of emotional response, Journal of advertising research, 35 (1995) 63-68.
14. Cowie, R., Douglas-Cowie, E., Tsapatsoulis, N., Votsis, G., Kollias, S. ve Fellenz, W., Emotion recognition in human-computer interaction, Ieee Signal Processing Magazine, 18, 1 (2001) 32-80.
15. Lee, C. M. ve Narayanan, S. S., Toward detecting emotions in spoken dialogs, Ieee Transactions on Speech and Audio Processing, 13 , 3 (2005) 293-303.
16. Campbell, W., Database of emotional speech, Proceedings of the ISCA ITRW o speech and emotion, (2000) 34-38.
17. Engberg, I. ve Hansen, A., Documentation of the Danish emotional speech database, <http://cpk.auc.dk/tb/speech/Emotion/>, 03.06.2015
18. Johnstone, T., Van Reekum, C. M., Hird, K. Kirsner, K., ve Scherer, K. R., Affective speech elicited with a computer game, Emotion, 5 , 12 (2005) 513-518.
19. Burkhardt, F., Paeschke, A., Rolfes, M., Sendlmeier, W. ve Weiss, B., A database of German Emotional Speech, Interspeech, (2005) 1517-1520.
20. Nwe, T. L., Foo, S. W. ve De Silva, L. C., Speech emotion recognition using hidden Markov models, Speech Communication, 41 , 11 (2003) 603-623.
21. Slaney, M. ve McRoberts, BabyEars, G., A recognition system for affective vocalizations, Speech Communication, 39 , 2 (2003) 367-384.
22. Douglas-Cowie, E., Campbell, N., Cowie, R. ve Roach, P., Emotional speech: Towards a new generation of databases, Speech Communication, 40 , 4(2003) 33-60.
23. Ververidis, D. ve Kotropoulos, C., A review of emotional speech database, PCI 2003, 9th Panhellenic Conference on Informatics, 2003, 560-574.
24. Schuller, B., Vlasenko, F., Eyben, B., Rigol, G. ve Wendemuth, A., Acoustic Emotion Recognition: A Benchmark Comparison of Performances, Automatic Speech Recognition and Understanding Workshop (ASRU 2009), 2009 552-557.
25. Humaine. <http://emotion-research.net/wiki/Databases>, 12.07.2015.
26. Oflazoğlu, Ç., Türkçe Konuşma Sinyalinden Duygu Tanıma, Yüksek Lisans Tezi, Mustafa Kemal Üniversitesi Fen Bilimleri Enstitüsü, 2011.
27. Engbert, I. S., ve Hansen, V. A., Documentetion of the danish emotional speech database (DES). Center for Person Kommunikation, Tech. Rep., Denmark, 2007.

28. Douglas-Cowie, E., Cowie, R. Sneddon, I., Cox, C., Lowry, O. ve McRorie, M., The HUMAINE database: Addressing the Collection and Annotation of Naturalistic and Induced Emotional Data. , Affective Computing and Intelligent Interaction, 2007.
29. Schuller, B., Wimmer, M., Arsic, D., Rigol, G. ve Radig, B., Audiovisual behavior modeling by combined feature spaces, ICASSP, (2007) 733-736.
30. Hansen, J., Getting started with susas: A speech under simulated and actual stress database, EUROSPEECH-97, 4 (1997) 1743-1746.
31. Schuller, B., Müller, R., Eyben, F., Gast, J. Börnler, H.ve Wöllmer M., Being Bored? Recognising Natural Interest by Extensive Audiovisual Integration for Real-Life Application, Image and vision Computing Jornal (IMAVIS), (2009).
32. Steininger, S., Schiel, F., Dioubina, O. ve Raubold, S., Development of user state conventions for the multimodal corpus in smartkom., Workshop on Multimodal Resources and Multimodal System Evaluation, (2002) 33-37.
33. Batliner, A., Steidl, S., Schuller, B., Seppi, D., Laskowski, K. ve Vogt, T., Combining Efforts for Improving Automatic Classification of Emotional User States., IS-LTC, (2006) 240-245.
34. Vidrascu, L. ve Devillers, L., Detection of real-life emotions in call center., Eurospeech, (2005) 1841-1844.
35. France, D. J., Shiavi, R. G., Silverman, M. ve Wilkes, D. M., Acoustical properties of speech as indicators of depression and suicidal risk., IEEE Transactions on Biomedical Engineering, 47 , (2003) 829-837.
36. Devillers, L. ve Vasilescu, I., Reliabilty of Lexical and Prosodic Cues in two Real-life Spoken Dialog Corpora, LREC, (2004).
37. Scherer, K. R. ve Ceschi, G., Studiying affective communication in the airport: The case of lost baggage claims, Personality and Social Psychological Bulletin, 26 (2000) 327-339.
38. Grimm, M., Kroschel, K. ve Narayanan, S., The Vera Am Mittag German Audio-Visual Emotional Speech Database, 2008 Ieee International Conference on Multimedia and Expo, 2008, Almanya, 1-4: 865-868.
39. Douglas-Cowie, E., Cowie, R. ve Schroeder, M., A new emotion database: considerations, sources and scope, 2000 , ISCA ITRW on speech and Emotion: 39-44.
40. Parlak, C., Diri, B. ve Gürgen, F., A Cross-Corpus Experiment in Speech Emotion Recognition, International Workshop on Speech, Language and Audio in Multimedia (SLAM 2014), (2014) 58-61.

41. Parlak, C. ve Diri, B., Farklı Veri Setleri Arasında Duygu Tanıma Çalışması, DEÜ Mühendislik Fakültesi Mühendislik Bilimleri Dergisi, 16 (2014) 21-29.
42. Sato, N. ve Obuchi, Y., Emotion Recognition using Mel-Frequency Cepstral Coefficients, Journal of Natural Language Processing, 1 (2007) 83-96.
43. Rabiner, L. ve Schafer, R., Digital processing of speech signals, firsted, Pearson Education, (1978).
44. Ververidis, D. ve Kotropoulos, C., Emotional speech classification using Gaussian mixture models and the sequential floating forward selection algorithm, 2005 IEEE International Conference on Multimedia and Expo (ICME), 1501-1504.
45. Shami, M. T. ve Kamel, M. S., Segment-based approach to the recognition of emotions in speech, 2005 IEEE International Conference on Multimedia and Expo (ICME), 366-369.
46. Picard, R. W., Vyzas, E. ve Healey, J., Toward machine emotional intelligence: Analysis of affective physiological state, Ieee Transactions on Pattern Analysis and Machine Intelligence, 23 , 10 (2001) 1175-1191.
47. Hu, H., Xu, M. X. ve Wu, W., Fusion of Global Statistical and Segmental Spectral Features for Speech Emotion Recognition, Interspeech 2007: 8th Annual Conference of the International Speech Communication Association, 1013-1016.
48. Ten Bosch, L., Emotions speech and the ASR framework, Speech Communication, 40, 4 (2003) 213-225.
49. Busso, C., Lee, S. ve Narayanan, S., Analysis of Emotionally Salient Aspects of Fundamental Frequency for Emotion Detection, Ieee Transactions on Audio Speech and Language Processing, 17 , 5 (2009) 582-596.
50. Cowie, R. ve Cornelius, R. R., Describing the emotional states that are expressed in speech, Speech Communication, 40 , 4 (2003) 5-32.
51. Johnstone, T. ve Scherer, K., Vocal Communication of Emotion, second ed., Guilford, (2000) 226-235.
52. Murray, I. ve Arnott, J., Toward a simulation of emotions in synthetic speech: A review of the literature on human vocal emotion, J. Acoust. Soc. Am., 93 (1993)1097-1108.
53. Scherer, K. R., Vocal Affect Expression - a Review and a Model for Future-Research, Psychological Bulletin, 99 , 3 (1986) 143-165.
54. Davitz, J. R., The Communication of Emotional Meaning, McGraw-Hill, (1964).
55. Gobl, C. ve Ni, A., Chasaide, The role of voice quality in communicating emotion, mood and attitude, Speech Communication, 40 , 4 (2003) 189-212.

56. Cowie, R. ve DouglasCowie, E., Automatic statistical analysis of the signal and prosodic signs of emotion in speech, Icslp 96 - Fourth International Conference on Spoken Language Processing, Proceedings, 1989-1992.
57. Schlosberg, H., Three dimensions of emotion, Psychological rev., 61 (1954) 81-88.
58. Li, X., Tao, J., Johnson, M. T., Soltis, J., Savage, A. ve Leong, K. M., Stress and emotion classification using jitter and shimmer features, 2007 IEEE International Conference on Acoustics, Speech, and Signal Processing, 1081-1084.
59. Gelfer, M. P. ve Fendel, D. M., Comparisons of Jitter, Shimmer, and Signal-to-Noise Ratio from Directly Digitized Versus Taped Voice Samples, Journal of Voice, 9 , 12 (1995) 378-382.
60. Lugger, M. ve Yang, B., The relevance of voice quality features in speaker independent emotion recognition, 2007 IEEE International Conference on Acoustics, Speech, and Signal Processing, 17-20.
61. Lugger, M. ve Yang, B., Cascaded emotion classification via psychological emotion dimensions using a large set of voice quality parameters, 2008 Ieee International Conference on Acoustics, Speech and Signal Processing, 4945-4948.
62. Lugger, M. ve Yang, B., On the relevance of high-level features for speaker independent emotion recognition of spontaneous speech, Interspeech 2009: 10th Annual Conference of the International Speech Communication Association 2009, 1959-1962.
63. Kaiser, L., Communication of affects by single vowels, Synthese, 14 (1962) 300-319.
64. Hernando, J. ve Nadeu, C., Linear prediction of the one-sided autocorrelation sequence for noisy speech recognition, Ieee Transactions on Speech and Audio Processing, 5 , 1 (1997) 80-84.
65. Bouquin, R., Enhancement of noisy speech signals: application to mobile radio communication, Speech Communication, 18 (1996) 3-19.
66. Atal, B. S., Effectiveness of Linear Prediction Characteristics of Speech Wave for Automatic Speaker Identification and Verification, Journal of the Acoustical Society of America, 55 (1974) 1304-1312.
67. Bou-Ghazale, S. E. ve Hansen, J. H. L., A comparative study of traditional and newly proposed features for recognition of speech under stress, Ieee Transactions on Speech and Audio Processing, 8 , 7 (2000) 429-442.
68. Reynolds, D. A., A Gaussian Mixture Modeling Approach to Text Independent Speaker Idendification, Ph.D. thesis, Georgia Ins. of Technology, 1992.

69. Schafer, R. W. ve Rabiner, L. R., Digital Representations of Speech Signals, Proceedings of the Ieee, 63 (1975) 662-677.
70. Atal, B. S., Automatic Recognition of Speakers from Their Voices, Proceedings of the Ieee, 64 (1976) 460-475.
71. Eskidere, Ö. ve Ertaş, F., Mel Frekansı Kepstrum Katsayılarındaki Değişimlerin Konuşmacı Tanımaya Etkisi, Uludağ Üniversitesi Mühendislik-Mimarlık Fakültesi Dergisi, 14 (2009) 92-110.
72. Rabiner, L. R. ve Juang, B. H., Fundamentals of speech recognition, Prentice Hall, Englewood Cliffs, (1993).
73. Karpov, E., Real Time Speaker identification, Master Thesis, University of Joensuu Department of Computer Science, 2003.
74. Karasartova, S., Metinden bağımsız konuşmacı tanıma sistemlerinin incelenmesi ve gerçekleştirilmesi, Yüksek Lisans Tezi, Ankara Üniversitesi Elektronik Mühendisliği Anabilim Dalı, Ankara, 2011.
75. Schuller, B., Lang, M. ve Rigoll, G., Robust acoustic speech emotion recognition by ensembles of classifiers, Proceedings of the DAGA'05, (2005) 329-330.
76. Cairns, D. ve Hansen, J., Nonlinear analysis and detection of speech under stressed conditions, J. Acoust. Soc. Am., (1994) 3392-3400.
77. Lee, C. ve Pieraccini, R., Combining acoustic and language information for emotion recognition, Proceedings of the ICSLP 2002, (2002) 873-876.
78. Schuller, B., Rigoll, G. ve Lang, M., Hidden Markov model-based speech emotion recognition, Ieee International Conference on Acoustics, Speech, and Signal Processing, 2003, 1-4.
79. Schuller, B., Towards intuitive speech interaction by the integration of emotional aspects, IEEE International Conference on Systems, Man and Cybernetics, 2002, 6.
80. Reynolds, D. A., Quatieri, T. F. ve Dunn, R. B., Speaker verification using adapted Gaussian mixture models, Digital Signal Processing, 10, 6 (2000) 19-41.
81. Alpaydin, E., Introduction to Machine Learning, MIT Press, 2004.
82. Küçük, H., Tepe, C. ve Eminoğlu, İ., K-En Yakın Komşu Algoritması ve Destek Vektör Makinesi Yöntemleri ile EMG İşaretlerinin Sınıflandırılması, IEEE, (2013).
83. Jansen, B. J., Making sense of data: a practical guide to exploratory data analysis and data mining, Information Processing & Management, 44 , 3 (2008) 978-979.
84. Wöllmer, M. ve Schuller, B., The Munich Versatile and Fast Open-Source Audio Feature Extractor, ACM Multimedia (MM), (2010) 25-29.

85. Breiman, L., Friedman, J. H., Olsen, R. A. ve Stone, C. J., Classification and regression trees, Wadsworth & Brooks/Cole Advanced Books & Software, (1984).
86. Erpolat, S. ve Öz, E., Kanser verilerinin sınıflandırılmasında yapı sinir ağları ile destek vektör makinalarının karşılaştırılması, 71-83.
87. Cömert, Z., Temel Bileşenler Analizine Genel Bir Bakış <http://www.zafercomert.com>, 10.10.2015.
88. Yaxin, S., Guihua, W. ve Jiabing, W., Weighted spectral features based on local Hu moments for speechemotion recognition, Biomedical Signal Processing and Control, (2014) 80-90.
89. Chih-Chung, C. ve Chih-Jen. L., LIBSVM : a library for support vector machines. ACM Transactions on Intelligent Systems and Technology 2:27:1--27:27, 2011. Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>, 10.10.2015.

ÖZGEÇMİŞ

Onur Erdem Korkmaz 1990 yılı Erzurum doğumludur. 2008 yılında Erzurum Mecidiye Anadolu Lisesinde mezun olmuş, aynı yıl Atatürk Üniversitesi Elektrik Elektronik Mühendisliğini kazanmıştır. Buradaki mühendislik eğitimini 2012 yılında tamamlamıştır. 2013 yılında Karadeniz Teknik Üniversitesi Elektrik Elektronik Mühendisliği Elektronik Bilim Dalında yüksek lisans eğitimine başlamıştır. 2013 yılında Atatürk Üniversitesi İspir Hamza Polat Meslek Yüksekokulunda Araştırma Görevlisi olarak göreve başlamış ve halen görevini sürdürmektedir. Gerçekleştirdi bu çalışma ile uluslararası ‘ELECO 2015’ konferansına ‘ Emotion Recognition from Speech Signal Using Mel-Frequency Cepstral Coefficient ’ isimli bildiri ile katılan Onur Erdem, İngilizce bilmektedir.