MAKİNE ÖĞRENMESİ İLE TÜRK MÜZİĞİNDE DUYGU ANALİZİ

Mehmet Bilal ER
151450101

DOKTORA TEZİ

Bilgisayar Mühendisliği Anabilim Dalı
Danışman: Prof. Dr. Emin Murat ESİN

İstanbul
T.C. Maltepe Üniversitesi
Fen Bilimleri Enstitüsü
Temmuz, 2019

JÜRİ VE ENSTİTÜ ONAYI

MEHMET BİLAL ER'ın "Makine Öğrenmesi ile Türk Müziğinde Duygu Analizi" başlıklı tezi 24.07.2019 tarihinde aşağıdaki jüri tarafından değerlendirilerek "Maltepe Üniversitesi Lisansüstü Eğitim ve Öğretim Yönetmeliği" nin ilgili maddeleri uyarınca Bilgisayar Mühendisliği Anabilim Dalı Yüksek Lisans/Doktora tezi oy birliğiyle/oy çokluğuyla, başarılı/başarısız olarak kabul edilmiştir.

	Unvanı. Adı ve Soyadı	İmza
Üye (Tez Danışmanı)	Prof. Dr. Emin Murat ESİN	.....
Üye	Dr. Öğr. Üyesi Erdal GÜVENOĞLU	.....
Üye	Dr. Öğr. Üyesi Ayşe CİLACI TOMBUŞ	.....
Üye	Doç.Dr. Fazlı YILDIRIM Yeditepe Üniversitesi	.....
Üye	Dr. Öğr.Üyei Can RAZBONYALI Altınbaş Üniversitesi	.....

Prof. Dr. İlter BÜYÜKDIĞAN
Enstitü Müdürü

Revizyon Takip Tablosu

REVİZYON NO	TARİH	AÇIKLAMA
00	01.03.2018	İlk yayın.

ETİK İLKE VE KURALLARA UYUM BEYANI

24../02../20.19

Bu tezin bana ait, özgün bir çalışma olduğunu; çalışmamın hazırlık, veri toplama, analiz ve bilgilerin sunumu olmak üzere tüm aşamalarından bilimsel etik ilke ve kurallara uygun davrandığımı; bu çalışma kapsamında elde edilmeyen tüm veri ve bilgiler için kaynak gösterdiğimi ve bu kaynaklara kaynakçada yer verdiğimi; çalışmanın Maltepe Üniversitesinde kullanılan "bilimsel intihal tespit programı" ile tarandığını ve öngörülen standartları karşıladığını beyan ederim.

Herhangi bir zamanda, çalışmamla ilgili yaptığım bu beyana aykırı bir durumun saptanması durumunda, ortaya çıkacak tüm ahlaki ve hukuki sonuçlara razı olduğumu bildiririm.

(Islak İmza)
Mehmet Bilal Er

TEŞEKKÜR

Doktora öğrenimim ve tez çalışmam süresince deneyim ve bilgisi ile beni yönlendiren değerli danışman hocam Prof. Dr. Emin Murat ESİN'e bilimsel katkılarının yanı sıra sonsuz hoşgörüsü, anlayışı ve desteği için çok teşekkür ederim. Tez izleme jürimde yer alarak araştırmanın biçimlenmesinde yardımcı olan sayın hocalarım Dr. Öğr. Üyesi Ayşe Cilacı TOMBUŞ ve Dr. Öğr. Üyesi Erdal GÜVENOĞLU'na çok müteşekkirim.

Ayrıca akademik hayatta giriştiğim her çalışmanın yoğun ve yorucu döneminde bana hep destek olan, sabrını ve anlayışını benden hiçbir zaman esirgemeyen sevgili eşime ve aileme sonsuz teşekkürlerimi sunarım.

Mehmet Bilal ER

Temmuz 2019

ÖZ

MAKİNE ÖĞRENMESİ İLE TÜRK MÜZİĞİNDE DUYGU ANALİZİ

Mehmet Bilal ER
Doktora Tezi
Bilgisayar Mühendisliği Anabilim Dalı
Danışman: Prof. Dr. Emin Murat ESİN
Maltepe Üniversitesi Fen Bilimler Enstitüsü, 2019

Müziğin duyguları tetikleme veya iletebilme yeteneği sayesinde insan hayatındaki rolü ve önemi büyüktür. Müzik duygularının tanınması bilim, psikoloji, müzikoloji ve sanat gibi birçok disiplinde ele alınan ortak bir çalışma alanı olduğu için son yıllarda yoğun bir şekilde araştırmacıların ilgisini çekerek güncel araştırma konusu olmuştur. Çoğu araştırmacı müzikten akustik özellikler çıkarır ve bu özelliklere karşılık gelen duygu etiketleri arasındaki ilişkileri araştırır. Bu çalışmada müzik kayıtlarından çıkarılan Chroma spektrogramları ile transfer öğrenme kullanılarak müzik duygu tanıma için yeni bir yöntem sunulmuştur. Önceden eğitimli ağ modeli olarak AlexNet mimarisi kullanılmıştır.

AlexNet modelinin Conv5, Fc6, Fc7 ve Fc8 katmanları özellik çıkarıcı katman olarak seçilmiştir ve bu katmanlardan derin görsel özellikler çıkarılmıştır. Çıkarılan derin özellikler Destek Vektör Makineleri ve Softmax sınıflandırıcıları eğitmek ve test etmek için kullanılmıştır. Önerilen yöntemin performansını değerlendirmek için gözetmeli makine öğrenmesi yöntemleri ile de müzik duygu tanıma işlemi yapılmıştır ve elde edilen sonuçlara göre sunulan yöntemin daha iyi performans gösterdiği gözlenmektedir.

Ayrıca bu çalışmada müzik dinleme sırasında hissedilen duyguya göre beyinde oluşan elektriksel sinyallerin yapısı gerçekte algılanan duygunun ifade edilmesi açısından öneme sahip olduğu üzerinde durulmuştur. Bu tür çalışmalarda kullanılan veri setlerindeki duygu etiketleri genelde bir gurup gönüllü dinleyici veya uzmanların görüşü alınarak belirlenir. Fakat hissedilen duygular öznel olabilir ve kültürel farklılıklara göre şekillenebilir. Bu bağlamda, müzik parçaları dinlenirken beyninde oluşan Elektroensefalografi (EEG) sinyalleri analiz edilmiştir ve EEG sinyalleri ile müzik duygu tanıma işlemi yapılmıştır.

Anahtar Sözcükler: Müzik duygu tanıma; Transfer öğrenme; Elektroensefalografi; Makine öğrenmesi; Ses işleme; Sınıflandırma

ABSTRACT

EMOTION ANALYSIS IN TURKISH MUSIC WITH MACHINE LEARNING

Mehmet Bilal ER
PhD Thesis
Computer Engineering Department
Thesis Advisor: Prof. Dr. Emin Murat ESİN
Maltepe University, Institute of Science, 2019

Music has a great role and importance in human life since it has the ability to trigger or convey feelings. As recognizing music emotions is the subject of many studies conducted in many disciplines like science, psychology, musicology and art, it has attracted the attention of researchers as an up-to-date research topic in recent years. Many researchers extract acoustic features from music, and investigate relations between emotional tags corresponding to these features. In the present study, a new method is presented for music emotion recognition by employing transfer learning with Chroma spectrograms extracted from music recordings. The AlexNet Architecture is used as the pre-trained network model.

The Conv5, Fc6, Fc7 and Fc8 layers of the AlexNet model were chosen as the feature extracting layer, and deep visual features were extracted from these layers. The extracted deep features were used to train and test the Support Vector Machines and the SoftMax classifiers. In order to evaluate the performance of the proposed method, music emotion recognition is done by using supervised machine learning techniques as well. Based on the gained results, it is seen that the proposed method has a better performance.

In addition, in this study, we emphasize on explaining the form of electrical signals produced by human brain during the course of listening to music has a deep impact on expressing the perceived real emotion. Emotion labels used in data sets in this kind of studies are determined either getting the idea of a group of volunteer listeners or experts. However, sensed emotions could be subjective and it can change depending on the cultural differences. İn this manner, EEG (Electroencephalogram) signals, that are produced by human brain while listening to the music tracks, are analyzed, and music emotion recognition is done by using EEG signals.

Keywords: Music emotion recognition; Transfer learning; Electroencephalogram; Machine learning; Sound processing; Classification

İÇİNDEKİLER

JÜRİ VE ENSTİTÜ ONAYI iv
ETİK İLKE VE KURALLARA UYUM BEYANI iv
TEŞEKKÜR iv
ÖZ v
ABSTRACT vi
İÇİNDEKİLER vii
TABLOLAR LİSTESİ ix
ŞEKİLLER LİSTESİ x
KISALTMALAR xiii
ÖZGEÇMİŞ xiv
1. BÖLÜM - GİRİŞ 1
1.1. Tezin Amacı ve Önemi 2
1.2. Araştırma Soruları 3
1.3. Literatür Araştırması 3
2. BÖLÜM - ÖNBİLGİ VE TEMEL KAVRAMLAR 8
2.1. Ses Sinyali ve Özellikleri 8
2.2. Duygu 9
2.3. Duygu sınıflandırma Yöntemleri 10
2.3.1. Kategorik Model 10
2.3.2. Boyutsal Model 12
2.4. Öznitelik Çıkarımı İçin Mevcut Platformlar 13
2.4.1. Marsyas 13
2.4.2. jAudio 13
2.4.3. MIRtoolbox 13
2.5. Müzikal Öznitelikler 16
2.5.1. Enerji 16
2.5.2. Ritim 16
2.5.3. Tını 17
2.5.4. Harmoni 17
3. BÖLÜM - Yöntem 18
3.1. Makine Öğrenmesi 18
3.2. Sınıflandırıcılar 18
3.2.1. Destek Vektör Makineleri 18
3.2.1.1. Doğrusal Destek Vektör Makinası 19
3.2.1.2. Doğrusal Olmayan Destek Vektör Makineleri 22
3.2.2. K-NN 24
3.3. Veri Seti 25
3.4. Duygu Sınıflandırmasında Müziksel Özelliklerin Değerlendirilmesi 26
3.4.1. Veri ön işleme 26
3.4.2. Öznitelik Çıkarımı 27
3.4.3. Öznitelik Vektörü 27

3.4.3.1. Kare Ortalamalarının Karekökü (Root Mean Square - RMS) 28
3.4.3.2. Düşük Enerji (Low energy) 28
3.4.3.3. Tempo 29
3.4.3.4. Yüksek Frekans/Tizlik(Brightnes) 30
3.4.3.5. Spektral Özellikler 31
3.4.3.6. Rolloff 31
3.4.3.7. Sıfır Eksenini Geçme Sayısı (Zero-crossing Rate) 32
3.4.3.8. Mel Frekans Kepstrum Katsayıları (MFCC) 32
3.4.3.9. Atak Zamanı (Attack Time) 36
3.4.3.10. Atak Eğimi (Attack Slope) 37
3.4.3.11. Chroma 37
3.4.4. Deneysel Çalışmalar ve Sınıflandırma Sonuçları 38
3.5. Duygu Sınıflandırmasında Transfer Öğrenmenin Değerlendirilmesi 40
3.5.1. Konvolüsyonel Sinir Ağları 40
3.5.2. Alexnet 42
3.5.3. Önerilen Yöntem 44
3.5.3.1. Chroma Spektrogramların Çıkarılması 45
3.5.3.2. Veri Artırma (Data Augmentation) 46
3.5.3.3. Alexnet İle Transfer Öğrenme Ve Derin Özellik Çıkarma İşlemi 47
3.5.4. Ayarlar 49
3.5.5. Deneysel Sonuçlar 49
3.5.6. Müzik Duygu Tanıma Uygulaması 55
4. BÖLÜM – EEG İle Müzik Duygularının Analizi 58
4.1. EEG Dalga Şekli 59
4.2. EEG Elektrotlarının Bağlanış Sekli 63
4.3. Dalgacık Dönüşümü (Wavelet Transform) 64
4.3.1. Sürekli Dalgacık Dönüşümü (SDD) 66
4.3.2. Ayrık Dalgacık Dönüşümü (ADD) 67
4.4. Dalgacık Ailelerinin Farklı Türleri 69
4.5. Uygulama-Araştırma Metodolojisi 71
4.5.1. Uyaran ve Katılımcılar 71
4.5.2. EEG Kayıt 72
4.5.3. EEG Sinyal Ön İşleme 73
4.5.4. Öznitelik Çıkarımı ve Deneysel Sonuçlar 73
5. BÖLÜM - SONUÇ 76
KAYNAKÇA 78

TABLOLAR LİSTESİ

Tablo 2-1: Dokuz Ruh Hali Kümesi (Schubert, 2003) 11

Tablo 2-2: Farklı Kütüphanelerinin Çıkarabileceği Öznitelikler 14

Tablo 3-1: SVM Çekirdek Fonksiyonları 24

Tablo 3-2: Veri setindeki Sınıflar ve Örnek Sayıları 26

Tablo 3-3: Deneyde Kullanılan Akustik Öznitelikler 27

Tablo 3-4: SVM Sınıflandırma Sonuçları 38

Tablo 3-5: K-NN Sınıflandırma Sonuçları 39

Tablo 3-6: Veri Arttırmasından Önce ve Sonra Her Bir Sınıftaki Örnek Sayısı 47

Tablo 3-7: Matlab'da AlexNet'in Modeli (mathworks Documentation, 2019) 48

Tablo 3-8: Conv5 Katmanı İçin Veri Artırma Öncesi ve Sonrası Sınıflandırma Sonuçları 50

Tablo 3-9: Fc6 Katmanı İçin Veri Artırma Öncesi ve Sonrası Sınıflandırma Sonuçları 51

Tablo 3-10: Fc7 Katmanı İçin Veri Artırma Öncesi ve Sonrası Sınıflandırma Sonuçları 52

Tablo 3-11: Fc8 Katmanı İçin Veri Artırma Öncesi ve Sonrası Sınıflandırma Sonuçları 54

Tablo 4-1: Dalgaların Frekans Aralıkları 60

Tablo 4-2: Alt bandlar ve Frekans Aralığı 74

ŞEKİLLER LİSTESİ

Şekil 2-1: Örnek Sinyalin Salınım Hareketi 8
Şekil 2-2: Sinüzoidal Periyodik Sinyal 9
Şekil 2-3: Thayer's Duygu Modeli. 12
Şekil 3-1: İki Sınıfı Ayıran Sonuz Sayıda Doğru 19
Şekil 3-2: İki Sınıfı Ayıran Optimal Hiper Düzlem 21
Şekil 3-3: Doğrusal Olarak Ayrılamayan Veriler 23
Şekil 3-4: Deneysel İşlemin Aşamaları 26
Şekil 3-5: RMS Enerji Eğrisi (Lartillot, 2018) 28
Şekil 3-6: Sinyaldeki Düşük Enerji Miktarının Gösterimi (Lartillot, 2018) 28
Şekil 3-7: Periyodik Sinyalin Pik Noktaları 30
Şekil 3-8: Yüksek frekanslı enerji örneği (Lartillot, 2018) 30
Şekil 3-9: Spektral roll off noktası örneği (Lartillot, 2018) 31
Şekil 3-10: Bir Sinyalin x Eksenini Geçtiği Noktaların Gösterimi (Lartillot, 2018) 32
Şekil 3-11: Frekans- Mel Ölçeği (wikipedia, 2018) 33
Şekil 3-12: MFCC Adımları 33
Şekil 3-13: Sinyal Çerçeveleme 34
Şekil 3-14: Mel Filitre Bankası 36
Şekil 3-15: Sinyal Atak Zamanı (Lartillot, 2018) 36
Şekil 3-16: Sinyal Atak Eğimi (Lartillot, 2018) 37
Şekil 3-17: Chromagram 37
Şekil 3-18: SVM Karışıklık Matrisi 39

Şekil 3-19: KNN Karışıklık Matrisi..... 39
Şekil 3-20 : Farklı Türdeki Özniteliklerden Elde Edilen Sınıflandırma Sonuçları..... 40
Şekil 3-21: Konvolüsyonel Sinir Ağın'ın Genel Yapısı 41
Şekil 3-22: Maksimum Havuzlama 43
Şekil 3-23: Önerilen Yöntemin Çalışma İlkesi..... 45
Şekil 3-24: Ses Sinyali ve Chroma Görüntüsü 46
Şekil 3-25: Veri Artırmanın Gösterimi..... 47
Şekil 3-26: Conv5 Katmanı için - Karşılaştırma Matrisi ve Roc Eğrisi 50
Şekil 3-27: Fc6 Katmanı için - Karşılaştırma Matrisi ve Roc Eğrisi..... 51
Şekil 3-28: Fc7 Katmanı için - Karşılaştırma Matrisi ve Roc Eğrisi..... 53
Şekil 3-29: Fc8 Katmanı için - Karşılaştırma Matrisi ve Roc Eğrisi..... 54
Şekil 3-30: Her katman için en iyi sınıflandırma sonuçları 55
Şekil 3-31: Müzik Duygu Tanıma Sistemi 55
Şekil 3-32: Chroma Özellik Çıkarımı İçin Gerekli Algoritma 56
Şekil 3-33: Chroma Çıkarımı ve Duygu Tahmini 57
Şekil 3-34: Derin Özellik Çıkarma ve Sınıflandırma İçin Gerekli Algoritma..... 57
Şekil 4-1: Delta Dalgası (Lotte, 2009)..... 60
Şekil 4-2: Teta Dalgası (Lotte, 2009) 61
Şekil 4-3: Alfa Dalgası (Lotte, 2009) 61
Şekil 4-4: Beta Dalgası (Lotte, 2009)..... 62
Şekil 4-5: Gama Dalgası (Lotte, 2009)..... 62
Şekil 4-6: Uluslararası 10-20 elektrot yerleştirme sistemi..... 63
Şekil 4-7: Orijinal Zaman Serisi Veri Setine Kıyasla Farklı Dönüşümlerin Zaman ve Frekans Çözünürlükleri Bakımından Karşılaştırılması..... 66
Şekil 4-8: Üç seviyeli dalgacık ayrıştırma ağacı. 68
Şekil 4-9: Yeniden yapılandırılmış üç seviyeli dalgacık ağacı. 69

Şekil 4-10: Örnek dalgacık şekilleri 70
Şekil 4-11: Duygu Tanıma Adımları 71
Şekil 4-12: EMOTIV EPOC 72
Şekil 4-13: EEG Kayıt Alma Süreci 73
Şekil 4-14: KNN için Karışıklık Matrisi 75
Şekil 4-15: SVM için karışıklık Matrisi 75

KISALTMALAR

Kısaltma	İngilizcesi	Türkçesi
MFCC	Mel Frequency Cepstral Coefficient	Mel Frekans Kepstrum Katsayısı
RMS	Root Mean Square	Kök kare ortalama
KNN	K Nearest Neighborhood	K-En Yakın Komşu
SVM	Support Vector Machine	Destek Vektör Makinesi
CNN	Convolutional Neural Network	Konvolüsyonel Sinir Ağı
EEG	Electroencephalogram	Elektroensefalografi
Fc	Fully Connected	Tam Bağlantı
GP	Gaussian Processes	Gauss İşlemleri
RBF	Radial Basis Function	Radyal Tabanlı Fonksiyon
ReLu	Rectified Linear Units	Düzleştirilmiş Doğrusal Birim
FFT	Fast Fourier Transform	Hızlı Fourier Dönüşümü
DFT	Discrete Fourier Transform	Ayrık Fourier Dönüşümü
Conv	Convolutional	Konvolüsyonel
Roc	Receiver Operating Characteristic	Alıcı Çalışma Karakteristiği
μV	Microvolt	Mikrovolt
WT	Wavelet Transform	Dalgacık Dönüşümü
CWT	Continuous Wavelet Transform	Sürekli Dalgacık Dönüşümü
DWT	Discrete Wavelet Transform	Ayrık Dalgacık Dönüşümü
Bpm	Beats Per Minute	Dakika Başı Vuruş Sayısı
Hz	Hertz	Hertz

ÖZGEÇMİŞ

Mehmet Bilal ER

Bilgisayar Mühendisliği Anabilim Dalı

Eğitim

Derece	*Yıl*	*Üniversite, Enstitü, Anabilim/ Anasanat Dalı*
Y.Ls.	2013	Çankaya Üniversitesi, Fen Bilimler Enstitüsü Bilgisayar Mühendisliği Anabilim Dalı
Ls.	2010	Doğu Akdeniz Üniversitesi, Mühendislik Fakültesi Bilgisayar Mühendisliği Bölümü
Lise	2004	Orhan Gazi Lisesi

İş/İstihdam

Yıl	*Görev*
2013 - 2016	Öğretim Görevlisi. İnönü Üniversitesi Arapgir MYO.
2016 - devam	Öğretim Görevlisi Harran Üniversitesi Bilgisayar Mühendisliği

Kişisel Bilgiler

Doğum yeri ve yılı : Şanlıurfa, 1988 Cinsiyet: E

Yabancı diller : İngilizce (iyi)

E-posta bilal.er@harran.edu.tr

1. BÖLÜM - GİRİŞ

Bu bölümde; çalışmanın amacı, önemi, kapsamı ve bu alanla ilgili literatür hakkında bilgiler verilecektir.

Müzik, insanlığın başlangıcından beri hayatımızda var olup, hem sosyal hem de bireysel anlamda büyük bir öneme sahiptir. Spor, eğlence, sağlık gibi çeşitli alanlarda, algıların ve duyguların farklı kültürler ve medeniyetler arasında aktarılmasını sağlar. Ayrıca, müzik duyguların ifade edildiği bir sanat dalıdır. Müzik sesleri ile çeşitli duygular arasındaki ilişkiyi belirli prensipler ile ortaya koymak mümkündür (Hevner, 1936).

Müzik içeriğinin çeşitliliği ve zenginliği nedeniyle farklı alanda bir çok çalışmaya araştırma konusu olmuştur. Müziğin duygusal içeriğinin hesaplamalar ile belirlenmesi yalnızca sinyal işleme ve makine öğrenmeyi değil aynı zamanda işitsel algı, psikoloji, bilişsel bilim ve müzikoloji kavramlarını içeren disiplinler arası bir araştırmadır (Lin & diğerleri., 2016). Müzik duygu tanıma sistemleri, kişisel müzik koleksiyonlarının yönetimi ve müzik öneri sistemlerini geliştirilmesi, müzik terapisi ile duygusal bozuklukların tedavisi gibi çeşitli amaçlar için aktif olarak kullanılmaktadır.

Son zamanlardaki müzik alanındaki gelişmeler ile birlikte bir müzik parçasının duygusal içeriğinin otomatik olarak analiz etme ve tanımayı amaçlayan çalışmalara ilgi artmıştır. Müzikteki duyguyu belirlemek amacıyla önerilen birçok çalışma olmasına rağmen bu alanda araştırılması gereken bir çok problem vardır.

Bu tez çalışmasında müzikteki duygunun tanınmasına yönelik problemler belirlenip, müzik duygu tanımadaki sınıflandırma başarısının artırılmasına yönelik yaklaşımlar geliştirilmiştir. Ayrıca bu çalışmada, müzik parçaları dinlenirken insan duygularının tanınması sorunu da araştırılmıştır.

1.1. Tezin Amacı ve Önemi

Müzik sinyal işleme alanında yapılan çalışmaların çoğu; müzik duygu tanıma, müzik türü sınıflandırma, müzik aletleri belirleme ve nota, akor gibi müzik benzerliklerini ortaya çıkarılması gibi uygulamaları kapsamaktadır. Fakat müzikteki duyguyu belirlemek amacıyla yapılan araştırmalar incelendiğinde bir çoğunda makine öğrenmesi yöntemleri kullanıldığı gözlenmektedir ve araştırmacıların çoğu genelde batı müziğindeki örnekleri seçerek veri setlerini hazırlamışlardır. Literatür araştırmamızda, Türk müziği melodilerinin duygusal yönden sınıflandırılması için analitik bir çalışmaya rastlanmamıştır.

Bu araştırmada Türk müziğindeki veriler kullanılarak sunulan yeni bir yaklaşımla müziğin duygusal içeriğinin sınıflandırılma başarısının artırılması amaçlanmaktadır. Çalışmanın amacı doğrultusunda Türk müziğinden seçilen örneklerden oluşan yeni bir veri seti hazırlanmış olup, özellik tabanlı makine öğrenmesi yöntemlerinden farklı olarak daha önceden eğitilmiş derin ağlara transfer öğrenme yöntemi uygulanarak müzik duygu tanıma işlemi yapılmıştır.

Bu araştırma, müzikoloji, psikoloji, sinyal işleme ve makine öğrenimini içeren çok disiplinli bir çalışma olduğu için bu alanlarda çalışan bir çok araştırmacıya yol gösterici olacağı düşünülmektedir. Literatüre kazandırılan veri seti ve sunulan yöntem sayesinde, Türk müziği örneklerinden oluşan veri setleri ile mühendislik ve müzik alanında daha fazla ortak çalışma oluşmasına katkı sağlanacağı düşünülmektedir.

Bu tür çalışmalar için hazırlanan müzik veri setlerindeki duygu etiketleri genelde ya uzman bir kaynaktan alınır ya da bir gurup dinleyicinin katılım sağladığı deneyler ile belirlenir. Deney sırasında dinleyicilere müzikler dinletilip ve daha sonra hissettikleri duygulara göre müzikleri duygusal olarak etiketlemeleri istenir. Fakat hissedilen duygular kişisel deneyimlere, kültürel farklılıklara veya yaşam tecrübelerine göre şekillenebilir. Müzik dinleme sırasında hissedilen duyguya göre beyinde oluşan elektriksel sinyallerin yapısı, gerçekte algılanan duygunun ifade edilmesi için son derece önemlidir. Bu bağlamda çalışmanın amaçlarından birisi de müzik parçaları dinlerken beyninde oluşan elektriksel dalgaların analizi ve bu dalgaların müzikten hissedilen duyguya göre sınıflandırılmasıdır.

1.2. Araştırma Soruları

Araştırmanın amacı ve hedefleri doğrultusunda aşağıdaki alt problemlere cevap aranacaktır.

- Müzikal verilerden çıkartılan öznitelikler ile işlenebilir anlamlı veriler elde edilebilir mi?
- Matematiksel ve istatistiksel verilerle Türk müziği yorumlanabilir mi?
- Türk müziği eserlerinde geçen duygular daha önceden eğitilmiş derin ağlar kullanılarak sınıflandırılabilir mi?
- Müzikteki her bir nota etrafındaki enerji dağılımını gösteren Chroma spektrogramlardan çıkarılan derin özellikler müzikteki duyguları tanıma için kullanılabilir mi ?
- Transfer öğrenme, müzik duygu tanıma problemi için kullanılabilir mi ve klasik makine öğrenmesi ile elde edilen sınıflandırma başarısı transfer öğrenme kullanılarak artırılabilir mi?
- Müzik dinlerken beyinde nasıl bir elektriksel dalgalanma oluşur ?
- Müzik dinleme sırasında oluşan beyin sinyalleri ile insan duyguları tanınabilir mi?

1.3. Literatür Araştırması

Bu bölümde, müzik duygu tanıma ile ilgili yapılan çalışmalardan bahsedilmektedir.

2003 yılında Feng ve arkadaşları tarafından müzikteki duygunun algılaması için bir yöntem önerilmiştir. Sadece tempo özelliklerini kullanarak duygular tanımlanmıştır. Dört duygudan oluşan kategorik model oluşturulmuş olup, kullanılan kategorik modelde mutluluk, hüzün, öfke ve korku gibi temel duygular yer almaktadır. Daha sonra sınıflandırma için 3 katmalı yapay sinir ağı kullanılarak sınıflandırma sonuçları elde edilmiştir. Elde edilen sonuçların başarısı yüksek olmasına rağmen test veri setinde yeterli miktarda kayıt bulunmamaktadır. Örneğin test veri setinde korku kategorisinde sadece üç adet müzik yer alıyordu (Feng, Zhuang & Pan, 2003).

2005 yılında Korhonen ve arkadaşları tarafından daha sistematik bir yöntem önerilmiştir. Müziğin duygusal içeriğini zamana ve müziksel özelliklere göre modellemek için çalışmalar yapmışlardır. Duygu, zamana göre değişen sürekli bir değişken olarak ölçülmüştür. Bu tür çalışmalarda duygu, değerlik ve uyarılma boyutları kullanılarak ölçülür. Modelden elde edilen sonuçlara göre değerlik ve uyarılma değerlerinin ortalama R2 istatistikleri % 78.4 ve % 21.9' dur (Korhonen, Clausi & Jernigan, 2005).

Li ve Ogihara tarafından iki temel problem araştırılmıştır. Birinci problem müziklerin benzerlikleridir. Belirli bir müzik ses dosyasına benzer müzik ses dosyaları aramak, ikincisi ise müzik seslerinden duygu tespitidir. Bu çalışmalarda çok etiketli bir sınıflandırma sistemi olarak duygu algılama problemini incelenmiş ve müziğin birden fazla duyguya sahip olduğu kabul edilmiştir. Hazırlamış oldukları veri setindeki ses dosyaları iki kişi tarafından duygusal olarak etiketlenmiştir. Daha sonra bu ses dosyalarından çıkarılan akustik özellikler kullanarak sınıflandırma işlemi yapılmıştır (Li & Ogihara, 2006).

Yang ve arkadaşları tarafından yapılan çalışmada insanlar bir şarkıyı dinlediklerinde aynı duyguyu paylaşmaz fikri üzerinde durulmuştur. Bir müziği deterministik bir şekilde sadece bir sınıfa atamak yerine şarkı bölümünün bir duygu sınıfına uygunluğunun ne kadar olası olduğunu belirleyen bir yaklaşım sunmuştur. Ölçümler için bulanık sınıflandırıcı kullanılmıştır (Yang, Liu & Chen, 2006).

2006 yılında Lu ve arkadaşları tarafından akustik müzik verileri ve batı kültüründeki müzikle ilgili psikolojik terimler kullanılarak duygu tanıma için hiyerarşik bir yapı sunulmuştur. Müzik kayıtlarından yoğunluk, tını ve ritim olmak üzere üç akustik özellik çıkarılmıştır. Genel bir müzik parçasında ruh hali sürekli değişebildiği için, müzik parçaları her biri sabit ruh hali içeren bağımsız bölümlere ayrılmıştır. Önerilen yöntem veri setindeki 800 müzik parçası kullanılarak değerlendirilip %86.3'e yakın başarı elde edilmiştir (Lu, Liu & Zhang , 2006).

Meyers 2007 yılındaki çalışmasında müzikten çıkardığı ses sinyallerinden ve şarkı sözlerinden elde ettiği özellikleri kullanarak istenen bir duyguya dayalı repertuar listeleri oluşturmak için bir araç geliştirmiştir. Ses içeriği bilgisini şarkı sözleri gibi içerikle

uyumlu verilerle birleştirerek, dinleyicinin mevcut durumuna ve ruh hallerine uyacak şekilde otomatik bir çalma listesi oluşturulmuştur (Meyers, 2007).

Huq ve arkadaşları tarafından otomatik içerik tabanlı müzik duygu tanıma sistemi eğitiminin farklı aşamalarının kapsamlı bir değerlendirmesi regresyon yaklaşımıyla sunulmuştur. Ses sinyallerinden 160 öznitelik çıkarılmış ve üç kategorideki farklı regresyonlar ile test edilmiştir. Yöntem olarak Doğrusal Regresyon, Regresyon Ağaçları ve Radyal Tabanlı Fonksiyonlar kullanılmıştır. Ayrıca test işlemi hem özellik seçimi kullanılarak hem de özellik seçimi olmadan yapılmıştır (Huq, Bello & Rowe, 2010).

Schmidt ve Kim özellikleri doğrudan spektrogramdan elde etmek için regresyona dayalı Derin İnanç Ağları kullanmışlardır. Sistem hem belirli müzik duygularını tanıma problemine uygulanırken, hem de herhangi bir regresyona dayalı ses özelliği öğrenme problemine de kolayca uygulanabilir olduğu gösterilmiştir (Schmidt & Kim, 2011).

Song ve arkadaşları tarafından müziksel özelliklerin duygu sınıflandırması üzerindeki etkisi kapsamlı olarak araştırılmıştır. Last.FM web sitesinde "mutlu", "hüzünlü", "kızgın" ve "rahat" dört duygu kelimesi ile etiketlenmiş 2904 şarkıdan oluşan bir veri seti toplamış ve standart algoritmalar kullanılarak çeşitli ses özellikleri çıkartılmıştır. Sınıflandırma için polinom ve radyal temelli fonksiyon çekirdeğine sahip SVM kullanılarak veri seti eğitilmiş ve bunlar 10 kat çapraz doğrulama uygulanarak test edilmiştir. Elde edilen sonuçlara göre spektral özelliklerin daha iyi performans gösterdiği gözlenmektedir (Song, Dixon & Pearce, 2012).

Markov ve Matsui tarafından GP modellerinin müzik duygusu tahmini için uygulanabilirliği araştırılmıştır. SVM, müzik duygu tanımada yaygın kullanılan ve gelişmiş model olmasından dolayı karşılaştırma için kullanılmıştır. Müzik dosyalarından elde edilen ses sinyallileri sinyal işleme teknikleri işlenmiştir ve farklı özellik çıkarılmıştır. Ayrıca çıkarılan özelliklerin çeşitli kombinasyonlarının etkileri de araştırılmıştır. Deney sonuçlarına göre müzik duygularını tanıma için GP, SVM'den daha iyi performans göstermiştir (Markov & Matsui, 2014).

2015 yılında Panda ve arkadaşları tarafından ses sinyallerinden çıkarılan standart ve melodik özellikler birleştirip müzik duygularının tanınması için yeni bir yöntem

önerilmiştir. Yazarlar, müzik duygularının sınıflandırılması için yeni bir ses veri seti hazırladılar. Veri setindeki her bir veri için 253 standart ve 98 melodik öznitelik çıkartılıp, çeşitli sınıflandırma algoritmaları kullanılarak duygu tanıma işlemi yapılmıştır. Ayrıca özellik seçimi kullanılmıştır. Deney sonuçlarına göre, melodik özelliklerin standart özelliklerden daha iyi performans gösterdiği gözlenmektedir. En iyi sonuç, ReliefF özellik seçimi ve SVM ile %64 F ölçütü olarak elde edilmiştir (Panda, Rocha & Paiva, 2015).

Ren ve arkadaşları birkaç farklı veri setlerindeki müziklerin duygularını sınıflandırmak için SVM'yi kullanmışlardır. Akustik frekans ve modülasyon frekans özelliklerini çıkarmak için iki boyutlu bir model önermişlerdir (Ren, Wu & Jang , 2015).

Liu ve arkadaşları tarafından müzik duygu tanıma için hem zaman hem de frekans alanı bilgilerini içeren müzik spektrogramlarını çıkardıktan sonra CNN kullanılmıştır. Öznitelik çıkarmak için ek bir çaba harcanmamıştır. Çünkü öznitelik çıkarma işlemi bir derin öğrenme yöntemi olan CNN modelinin katmanlarında otomatik olarak gerçekleştirilmiştir. Deneyler standart CAL500 ve CAL500exp veri seti kullanılarak gerçekleştirilmiştir. Sonuçlar, her iki veri kümesi için de önerilen yöntemin diğer başarılı yöntemlerden çok daha iyi performans gösterdiğini göstermektedir (Liu & diğerleri., 2017).

2018 yılında Papakostas ve Giannakopoulos tarafından derin özellikler kullanılarak konuşma ve müzik seslerini sınıflandırılması problemi araştırılmıştır. Geleneksel yöntemlerde olduğu gibi ses özelliklerini kullanarak ses içeriğini temsil etmek yerine, öğrenme için ses dosyalarından elde edilen spektrogramlardan derin özelikler çıkarılmıştır. CNN'yi eğitmek için ses spektrogramları giriş görüntüleri olarak kullanılmıştır. Çalışmada önceden eğitilmiş derin mimariler ile transfer öğrenme üzerinde durulmuştur. Deneyler üç farklı veri seti kullanılarak yapılmıştır. Sonuçlara göre CNN'lerin, özellikle transfer öğrenimi uygulandığında, her üç test veri setinde de performans açısından daha iyi sonuç verdiği gözlenmektedir (Papakostas & Giannakopoulos, 2018).

Delbouys ve arkadaşları tarafından ses sinyaline ve bir parçasının sözlerine dayanarak müzik duygu tespiti için yeni bir yaklaşım sunulmuştur. Geleneksel özellik temelli yaklaşımlar kullanılmış ve derin öğrenmeye dayalı yeni bir model önerilmiştir. Her iki yaklaşımın performansı, değerlik ve uyarılma değerleri olan 18.000 ses dosyası içeren bir veri tabanında karşılaştırılmıştır. Sunulan model uyarılma tahmininde daha iyi performans göstermiştir (Delbouys & diğerleri., 2018) .

Liu ve arkadaşları tarafından hem zaman hem de frekans bilgisi içeren spektrogramlar bir CNN ile sınıflandırılmış ve şarkılarda yer alan duyguları tanıma problemi araştırılmıştır. Çalışmalarında kullandıkları veri setinde değerlik ve uyarılma değerleri olan 1000 adet şarkı bulunmaktadır ve her şarkı 45 saniyedir (Liua & diğerleri., 2018).

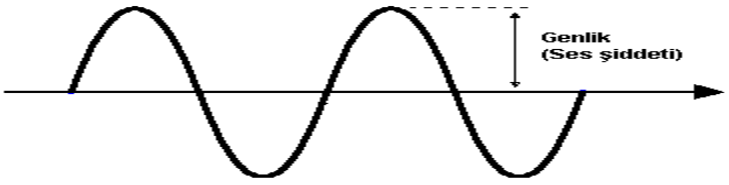
2. BÖLÜM - ÖNBİLGİ VE TEMEL KAVRAMLAR

2.1. Ses Sinyali ve Özellikleri

Ses, titreşen bir nesne tarafından oluşturulan basınç dalgasıdır. Bu titreşimler, çevredeki hava moleküllerini harekete geçirir, böylece ortama enerji iletir. Kulağın algılayabileceği ses bir kaynaktan titreşim oluşturarak doğar ve dalgalar şeklinde etrafa yayılır. Etrafa yayılan bu ses dalgaları hava moleküllerinin ileri geri hareketlerinden oluşur (wikipedia). Her hava molekülü bir önündeki hava molekülünü titreştirir. Bu titreşimler insan kulağına kadar ulaşır ve kulak zarımıza çarpan hava molekülleri kulak zarımızı titretir.

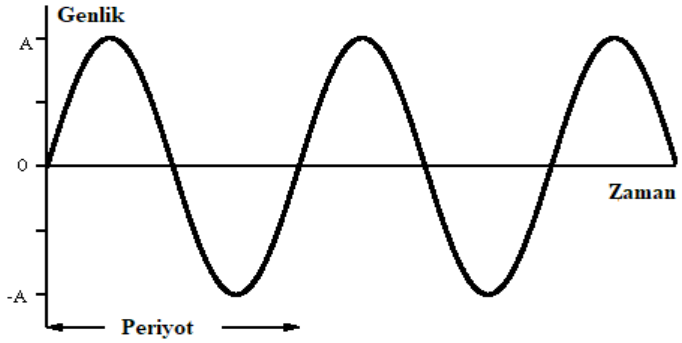
Kulak zarımızın etrafında titreşen hava molekülleri bir enerji oluşturur ve bu enerji beynimize sinyal olarak iletilir. Beynimiz bu sinyalleri anlamlandırdıktan sonra ses olarak algılar (Hearinghaven). Ses sinyalleri, ses dalgalarının elektronik gösterimleridir. Sesin dinlenmesi, ölçülmesi veya saklanması için boşlukta yayılabilen enerji dalgalarının oluşması gerekir. Bir ses sinyali genlik, periyod ve frekans gibi temel özelliklere sahiptir.

Genlik, bir hava molekülünün bozulmamış ya da denge pozisyonu ile yapabileceği maksimum yer değiştirmesi arasındaki farka denir. Bir sinyalin genliği bir salınım hareketinin üst noktası ile en alt noktası arasındaki mesafenin yarısına eşittir. Bir sesin enerjisi ve şiddeti artıkça ses sinyalinin genlik değeri de artar. Buna bağlı olarak yüksek seslerin genlik değerleri yüksek, alçak seslerin genlik değerleri ise düşüktür (Camastra & Vinciarelli, 2007). Şekil 2.1'de örnek bir sinyalin salınım hareketi gösterilmektedir.



Şekil 2-1: Örnek Sinyalin Salınım Hareketi

Periyot, bir hava molekülünün ileri geri hareketini tamamlaması için geçen süredir. Bu ileri geri harekete de döngü denir. Bir sinyalin periyodu kendini tekrar etme süresidir. Periyodun ölçü birimi saniyedir (Camastra & Vinciarelli, 2007). Periyot bir döngünün saniye cinsinden değeri olarak düşünülebilir. Periyot T harfi ile gösterilir. Şekil 2.2'de Örnek sinüzoidal periyodik bir sinyal gösterilmektedir.



Şekil 2-2: Sinüzoidal Periyodik Sinyal

Sesin frekansı, bir saniyedeki yaptığı titreşim sayısını ifade eder. Frekans, bir saniyede meydana gelen dalga döngüsü sayısı olarak ölçülür. Frekans ölçümü birimi Hertz' dir ve Hz olarak gösterilir. Denklem 1'de frekans formülü verilmiştir. Frekans küçük " f " ile gösterilir, "T" ise saniye cinsinden zamanı ifade eder. Seslerin tiz ve pes olması frekans kavramıyla ilgilidir. Tiz sesler yüksek frekans değerlerine sahip, pes sesler ise düşük frekans değerlerine sahiptir. İnsan kulağı 20Hz kadar pes seslerle ile 20000 Hz kadar tiz sesler arasındaki tüm sesleri duyabilir (Doval & Rodet, 1991).

$$f = \frac{1}{T}\,[\mathrm{Hz}] \qquad (1)$$

2.2. Duygu

Duygu terimi üzerinde mutabık kalınan tam bir tanımlama bulunmamaktadır. Duyguların bir başlangıcı ve bir sonu vardır, ancak kesin sürelerinin belirlenmesi zor olabilir. Sonuç olarak duygu, bireyin yaşamında bir bölüm olarak düşünülmelidir (Mulligan & Scherer, 2012). Duygular, boyutsal bir tanımla ile ifade edilebilir. Böyle bir tanım, bir duygu için gerekli yeterli koşulları veya kriterleri ifade eder. Duygu teorisyenleri, duyguların kalıcı olmadığını kabul eder (Moors & Scherer).

Duygu terimi, genellikle öznel duygu, fizyolojik uyarılma, ifade, eylem gibi birkaç alt bileşeni içeren olan oldukça kısa fakat yoğun bir duygusal tepkimeye atıfta bulunmak için kullanılır. Duygular son dakikalardaki yada son birkaç saatteki belirli nesnelere (örneğin mutluluk, üzüntü) odaklanır. Müzikal duygu terimi ise müziğin bir şekilde insanda uyardığı duyguların kesin doğası hakkında herhangi bir ilave yapmadan sadece müziğin neden olduğu duygular olarak tanımlanabilir. (Juslin & Sloboda, 2010). Bu çalışmada müzikal duygular üzerinde durulacaktır.

2.3. Duygu sınıflandırma Yöntemleri

Müzik sinyallerinden duygu tanıma işlemi zordur. Bu zorluğun nedenlerinden biri, duygu algısı özünde özneldir ve insanlar aynı şarkı için farklı duyguları algılayabilirler. Bu öznellik sorunu, bir müzik duygu tanıma sisteminin performans değerlendirmesini temel olarak zorlaştırmaktadır. Bir diğer sebep ise, duyguları evrensel bir şekilde tanımlamak kolay değildir, çünkü aynı duyguyu tanımlamak için kullanılan sıfatlar kişiden kişiye değişebilir. Son olarak, müziğin insanlarda duyguları nasıl uyandırdığı hala tam olarak bilinmemek ile birlikte müziğin içsel öğesi, dinleyici üzerinde nasıl duygusal tepkiler oluşturduğu hala tam olarak anlaşılamamıştır (Yang & diğerleri., 2008) (Yang & diğerleri., 2007).

Müzikteki duyguyu tanımak ve analiz etmek için halen geliştirilmiş standart bir yöntem olmamasına rağmen bu alanda birçok çalışmada psikolojik duygusal modeller yaygın olarak kullanılmaktadır (Han & diğerleri., 2009).

Müzik duygu tanıma için mevcut otomatik sistemler, kategorik veya boyutsal olarak iki ana gurup altında toplanabilir (Panda, Rocha & Paiva, 2015) (Barthet, Fazekas & Sandler, 2012).

2.3.1. Kategorik Model

Ayrık modeller olarak da bilinen kategorik modeller, duyguları tanımlamak için tek kelimeleri veya kelime gruplarını kullanarak sınıflandırır (örneğin, mutlu, üzgün, endişeli, rahatlatıcı). Kategorik model için farklı araştırmacılar farklı duygu kümeleri ortaya çıkarıp birçok duygu taksonomisi kullanmışlardır.

Hevner'in sıfat kontrol listesi, müziğe olan duygusal tepkileri ölçmek için yaygın olarak kullanılmış olup, müzik ve ruh haliyle ilgili araştırmanın temelini oluşturmuştur. Bu kontrol listesi; ciddi, kasvetli, özlem, lirik, canlı, neşeli, huzursuz ve gürültülü (kaba saba) olan sekiz kümeden 67 sıfattan oluşuyor. Bununla birlikte, aynı kümedeki sıfatlar aslında yaklaşık olarak aynı anlama geldiğinden, birini diğerlerinden ayırt etmek çok zordur. Bu belirsizlik temel gerçeği elde etmeyi zorlaştırır (Hevner, 1935).

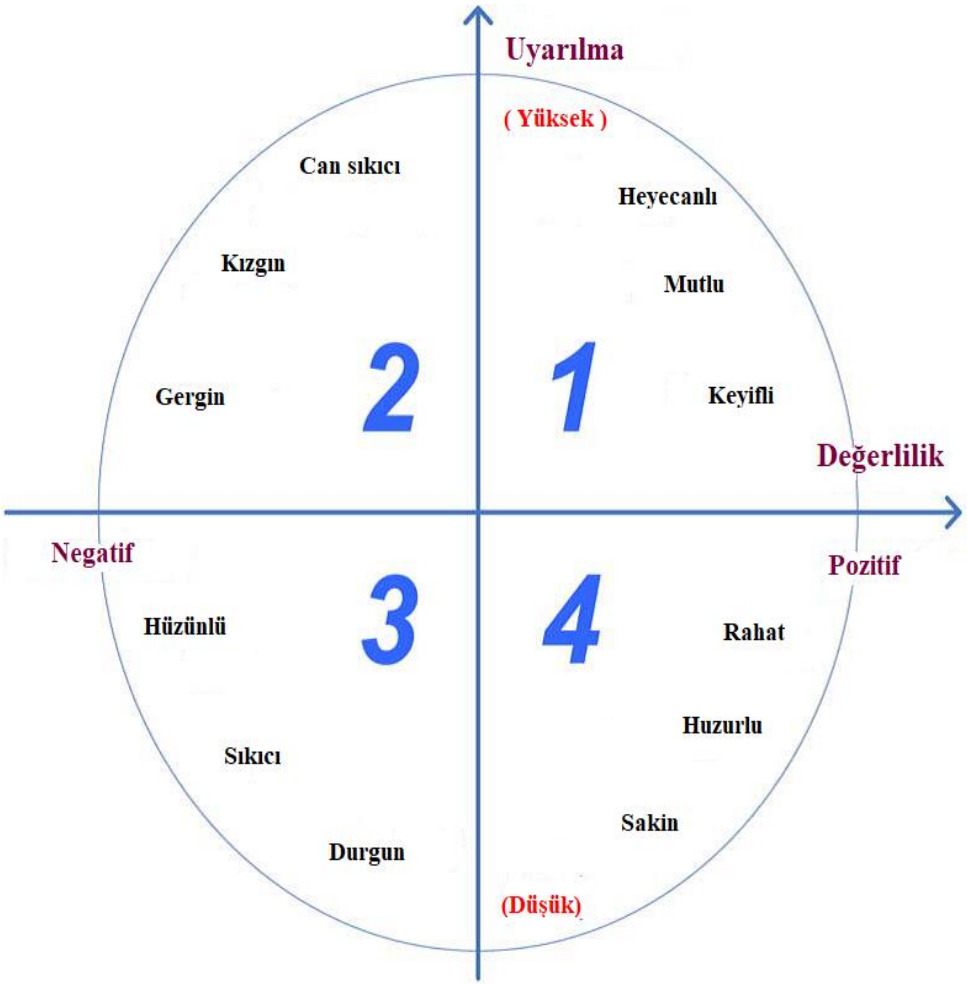
1950'lerde ve 1960'larda kontrol listesinin Farnsworth tarafından revize edilmesinden bu yana liste güncellenmemiştir. Daha sonra müzikal olarak deneyimli 133 kişi listedeki sıfatların müzikal tanıma uygunluğunu araştırmıştır. Schubert 2003 yılında bunları Tablo 2-1'de gösterildiği gibi 46 kelime ve dokuz gurup olacak şekilde güncellemiştir (Schubert, 2003) .

Tablo 2-1: Dokuz Ruh Hali Kümesi (Schubert, 2003)

Kümeler	**Örnekler**
Küme 1	Canlı, neşeli, mutlu, sevinçli
Küme 2	Komik, hafif, lirik, şen, şakacı
Küme 3	Sakin, narin, ağırbaşlı, gürültüsüz, rahat, durgun, dinlendirici, sevecen, huzurlu
Küme 4	Romantik, duygusal
Küme 5	Kasvetli, iç karartıcı, ümitsiz, melankoli, yaslı, hüzünlü, resmi
Küme 6	Ağır, görkemli, kutsal, ciddi, manevi, coşkulu
Küme 7	Trajik, özlem
Küme 8	Tedirgin, kızgın, huzursuz, gergin
Küme 9	Dramatik, heyecan verici, canlı, tutkulu, sansasyonel yükselen, muzaffer

2.3.2. Boyutsal Model

Boyutsal modellerde, farklı duygusal durumlar iki boyutlu alana yerleştirilen duygu durumlarına dönüştürülmüş bir alanda konumlandırır (Panda, Rocha & Paiva, 2015). Thayer'in iki boyutlu duygu modeli boyutsal modele örnektir ve bu model şekil 2-3'de verilmiştir. Bu model duyguları iki boyutlu bir alana yerleştirdikten sonra basit ve oldukça etkili bir model sunar. Modelde, uyarılma ve değerlik miktarı sırasıyla dikey ve yatay eksen boyunca ölçülür (Thayer, 1989). Yaygın olarak kullanılan bu modeller üzerinden müzikler duygusal olarak kategorize edilip bir dizi sınıflara ayrılır ve daha sonra bir sınıflandırıcıyı eğitmek için standart örüntü tanıma prosedürü uygulanır (Li & Ogihara, 2004).



Şekil 2-3: Thayer's Duygu Modeli.

2.4. Öznitelik Çıkarımı İçin Mevcut Platformlar

Ses öznitelik çıkarma kütüphaneleri ses analizi, sentezi ve sesten hesaplamalara dayalı önemli veriler çıkarmak için hazırlanmış olan etkili ve genişletilebilir yazılımlardır. Bu yazılımlar, esas olarak müzik işleme alanındaki geliştiricilerin araştırmalarını desteklemek amacıyla oluşturulmuştur. Ses sinyallerinden öznitelik çıkarımı için mevcut en çok bilinen ve en önemli kütüphaneler MIRtoolbox, Marsyas ve JAudio'dur.

2.4.1. Marsyas

MARSYAS ses işleme ve müzik sinyallerinden öznitelik çıkarmak için çok yaygın kullanılan açık kaynak kodlu bir kütüphanedir ve bu alan için geliştirilen ilk yazılımlardan biridir. MARSYAS, hem akademi hem de endüstride çeşitli projeler için kullanılmıştır. Bu kütüphane, algoritmalarla kolay deney yapılmasına olanak tanıyan ve gerçek zamanlı ses analizi ve sentez araçlarının geliştirilmesinde yararlı olan hızlı performans sağlayan genel, genişletilebilir ve esnek bir mimari sunmaktır (Tzanetakis & Cook , 2000). Bu kütüphanenin temel dezavantajı, görünüşe göre diğer kütüphanelerden daha az özelliklere sahip olmasıdır ve diğerlerine göre daha az öznitelik çıkartmasıdır.

2.4.2. jAudio

jAudio, hem bir kullanıcı ara yüzü ile hem de komut satırı modunda öznitelik çıkarabilen açık kaynak kodlu bir kütüphanedir. Güncellemeler eklenti ara yüzü kullanılarak kolayca kütüphaneye dâhil edilebilir. Java ortamında yazılmıştır (McEnnis & McKay). MARSYAS'a göre daha fazla öznitelik çıkarabilmektedir.

2.4.3. MIRtoolbox

MIRtoolbox, MATLAB'da yazılmış, entegre bir dizi fonksiyona sahiptir. Müzikal verilerden bilgi çıkarma alanındaki hesaplama yaklaşımlarına genel bir bakış sunmak amacıyla hazırlanmış bir yazılımdır. Tasarım, modüler bir yapıya dayanmaktadır: Farklı algoritmalar, formüle edilmiş aşamalara ayrıştırılır (Neeraj, Kumar & Subrata , 2015). Bu araç kutusunun ana avantajları MATLAB ile entegrasyonu (geniş dokümantasyon, topluluk yardımı ve bir çeşit kullanım kolaylığı) ve çok sayıda yüksek seviyeli müzikal öznitelik çıkarma araçları içermesidir. Bu avantajlardan dolayı bu çalışmada

"MIRtoolbox" kütüphanesi kullanılmıştır. Tablo 2-2'de MIRtoolbox, Marsyas ve JAudio kütüphanelerinin çıkarabileceği öznitelikler listelenmiştir.

Tablo 2-2: Farklı Kütüphanelerinin Çıkarabileceği Öznitelikler

Özellik	**Özellik Sınıfı**	**jAudio**	**MARSYAS**	**Toolbox MIR**
Fraction Of Low Energy Frames	Enerji	√		
Less-Than-Average Energy	Enerji	√		√
Root Mean Square Derivative	Enerji	√		
Root Mean Square	Enerji	√		√
Root Mean Square Variability	Enerji	√		√
Pitch (F0)	Nota		√	
Strongest Frequency Via FFT Maximum	Nota	√		
Beat Sum	Ritim	√	√	√
Rhythmic Fluctuation	Ritim			√
Strength Of Strongest Beat	Ritim	√		√
Tempo	Ritim	√		√
Attack Slope	Tını			√
Attack Time	Tını			√
High Frequency Energy (Brightness)	Tını			√

Linear Prediction Reflection Coefficients	Tını	√		
Mel-Frequency Cepstral Coefficients	Tını	√	√	√
Sensory Dissonance	Tını			√
Spectral Centroid	Tını	√	√	√
Spectral Flux	Tını	√	√	√
Spectral Peaks Variability (Irregularity)	Tını	√		√
Spectral Roll Off	Tını	√	√	√
Strongest Frequency Via Spectral Centroid	Tını	√		
Zero Crossing Rate	Tını	√	√	√
Zero Crossing Derivative	Tını	√	√	
Harmonic Change Detection Function	tonality			√
Key (Tonal Center Positions)	Harmoni			√
Modality	Harmoni			√
Tonal Centroid	Harmoni			√

2.5. Müzikal Öznitelikler

Yaygın olarak kabul edilen görüşlerden biri, müziğin duygusal anlamı ortaya çıkan olaylarla ilişkili olduğudur. Müziğin zaman içinde yoğunluğunu ve niteliğini değiştiren farklı müzikal yapıları sayesinde müzikteki duyguyu hesaplamalarla tanımlamak mümkündür (Krumhansl, 2002). Müzikal araştırmalar, müzisyenlerin farklı duyguları ifade etmek için tempo, ritim, tını gibi nitelikleri kullandıklarını göstermektedir. Farklı akustik işaret kalıpları farklı müzik duygularının oluşumunda etkilidir (Juslin, 2000).

Bir müzik parçasının ham ses dalga formundan hesaplanabilecek birçok akustik özellik vardır. Bu özelliklerin düşük seviyeden yüksek seviyeye ve karmaşıklığa göre farklılık gösteren farklı kategorileri vardır. Müzik; notalar, zil sesleri, ritimler, enstrümanlar ve şarkı sözlerinden oluşur (Laar, 2006). Müzikte dinleyicilerin algılayabileceği duygu özellikleri mod, tempo, uyum, ses yüksekliği, ses tonu, ritim vb. gibi faktörlere bağlıdır (Hwang & diğerleri., 2013).

2.5.1. Enerji

Müzik duygu algılamasındaki en önemli özelliklerden biri enerjidir, çünkü uyarılma boyutuyla oldukça ilgilidir. Yüksek enerjinin yoğun olduğu müzik parçaları, coşku ve endişe ile ilgiliyken, düşük enerjinin yoğun olduğu müzik parçaları ise sakin ve depresyon ile ilgilidir (Hampiholi, 2012). Ses tanıma araçları, ses gücü, toplam ses yüksekliği, sinyaldeki düşük enerji miktarı gibi enerji ile ilgili bir dizi özelliği çıkarmak için kullanılabilir.

2.5.2. Ritim

Müzikte kuvvetli ve zayıf melodik ve vuruşların düzenli bir şekilde tekrarlanmasından oluşan notaların desenidir. Tempo ve ritim çeşitliliği duygu ifadesi ile yakından ilgilidir. İnsanın ruh hali tepkisi müzikal bir parçada mevcut olan tempo ve ritim periyodikliği tarafından belirlenir. Örneğin, hızlı tempo genellikle mutluluk, öfke veya korku ifadeleri ile ilişkiliyken, yavaş tempo, hüzün, hassasiyet, ciddiyet ifadeleriyle ilişkilidir ayrıca değişken ritimlerin olduğu bir müzik parçası genellikle sevinç verir (Hwang & diğerleri., 2013) .

Ritim düzenliliği ve netliği başlangıç eğrisine oto korelasyon uygulanarak hesaplanır. Bir müzik segmentinin belirgin ve düzenli bir ritmi varsa, ilgili oto korelasyon eğrisinin zirveleri de bariz ve güçlü olacaktır (Lu, Liu & Zhang , 2006).

2.5.3. Tını

Müzik tınısı, müzik aletlerini ayırt eden ve duygusal tepkiler üzerinde etkili olduğu öne sürülen bir ses niteliğidir. Tını özellikleri konuşma tanıma sistemlerinde de yaygın olarak kullanılmaktadır. Bu özellik setini çıkarmak için ses sinyallerinin ön işlenmesi gerekir (Pouyanfar & Sameti , 2014). Farklı müzik parçaları ayı perdeye ve aynı gürültüye sahip olsalar bile tını belirli bir müzik parçasının sesini diğerinden farklı kılar. Tını, insanın müzik algısında önemli bir rol oynar. Çok daha fazla harmoniye sahip tonlar endişeyle ve coşkuyla ilişkiliyken, az sayıda harmoniye sahip olan tonlar huzur ve depresyon ile ilişkili olabilir (Juslin & Sloboda, 2010).

Birçok araştırma çalışmada MFCC, ses ve müzik sinyallerinin tınısını temsil eden özellik olarak kullanıldıysa spektral özellikler farklı ruh hallerini ayırt etmede önemlidir. Örneğin, coşkulu müziklerde yüksek frekans/tizlik (brightness) oranı genellikle depresif müziklerden daha yüksektir. Coşkulu müzikler genellikle alt bantlarda depresif duygular içeren müziklerden daha büyük bir spektral enerjiye sahiptir (Lu, Liu & Zhang , 2006).

2.5.4. Harmoni

Harmoni, müzikte tonik merkezi bir nota etrafında müzik besteleri düzenleme prensibidir. Mod, Chroma gibi bu kategorideki özellikler müzikteki duygusal ifadeyle ilişkilidir.

Duygular, tempo, tını, ahenk ve ses yüksekliği gibi özelliklerden etkilenebilir ve önceki çalışmaların çoğu akustik özelliklerin geliştirilmesine yöneliktir. Her ne kadar bazı araştırmalar, duygu sınıflandırması için en bilgilendirici özellikleri aramaya odaklanmış olsa da, duyguyu belirlemede hiçbir baskın tek özellik ortaya çıkmamıştır.

3. BÖLÜM - YÖNTEM

3.1. Makine Öğrenmesi

Makine öğrenmesi, mevcut veriler üzerinden modeli öğrenebilen ve bilinmeyene dair tahminler yapabilen algoritmaları araştıran yöntemlerdir. Makine öğrenimi genel olarak üç geniş kategoride sınıflandırılır. Bunlar denetimli öğrenme, denetimsiz öğrenme ve takviyeli öğrenimidir.

Temel olarak denetimli öğrenmede, makineye iyi etiketlenmiş verileri kullanarak öğrettiğimiz veya eğittiğimiz bir öğrenimdir (Kotsiantis, 2007). Bu işlem bazı verilerin zaten doğru cevapla etiketlendiği anlamına gelir. Bundan sonraki adımda, denetlenen öğrenme algoritmasının eğitimi analiz etmesi ve etiketli verilerden doğru bir sonuç çıkarabilmesi için makineye yeni bir örnek veri verilir. Denetimsiz öğrenmede, eğitim veri kümesindeki örneklere karşılık gelen bir çıktıya sahip olmayan girdi verisi kümesi oluştuğunda kullanılır (Kotsiantis & Pintelas, 2004).

Takviye öğrenme tekniği ise bir kazancı en üst düzeye çıkarmak için belirli bir durumda yapılması gereken uygun eylemleri bulma problemi ile ilgilidir. Burada öğrenme algoritması, denetimli öğrenmenin aksine, optimal çıktılara örnekler vermez bunun yerine onları deneme yanılma süreci ile keşfetmelidir. İstenilen işlemin yapılmasını sağlamak için algoritmada ki gerekli adlımlar gerçekleştirdiği eylemler için ödül veya ceza alır. İstenilen sonucu üreten adımlar ödüllendirilir, istenmeyen sonucu üreten adımlar ise cezalandırılır. Amacı toplam ödülü en üst düzeye çıkarmaktır.

3.2. Sınıflandırıcılar

3.2.1. Destek Vektör Makineleri

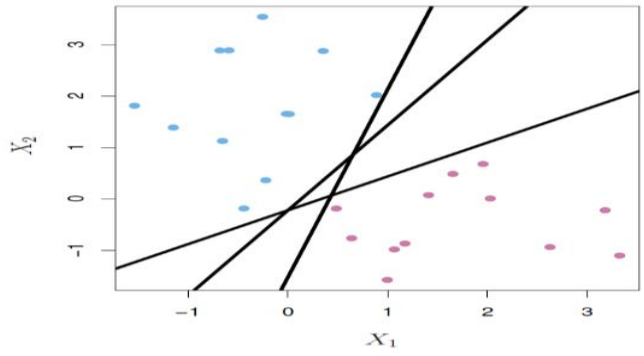
Makine Öğrenmesi, yapay zekanın bir alt alanı olarak kabul edilir ve bilgisayarın öğrenmesini sağlayan teknik ve yöntemlerin geliştirilmesiyle ilgilidir. Sınıflandırma ve regresyon analizi için kullanılan bir makine öğrenme yaklaşımıdır.

Örüntü tanıma problemlerinin çözümü için SVM 1995 de tanıtılmıştır (Vladimir, 1995). Sınıflandırma problemini çözmek için SVM'ler geliştirildi, ancak regresyon problemlerini çözmek için de genişletilmiştir (Vapnik, Golowich & Smola, 1997). SVM, doğrusal ve doğrusal olmayan problemleri çözebilir ve birçok pratik problem için iyi sonuçlar verebilir.

3.2.1.1. Doğrusal Destek Vektör Makinası

Temel amaç, iki sınıfın verileri arasında bir ayırıcı çizgi veya hiper düzlem oluşturmaktır. Hiper düzlem herhangi bir boyuttaki veriyi doğrusal olarak ayırır. Bu hiper düzlem iki boyutlu, veya daha fazla boyutta olabilir. SVM, iki paralel çizgi oluşturarak paralel bölümler oluşturur. Düz ve doğrusal bölümler oluşturmak için alanı tek bir geçişte ayırır. SVM kategoriler arasındaki en geniş aralığı yani en büyük marjini bulmaya çalışır, iki kategoriyi mümkün oldukça geniş bir aralıkta bölünür. Bu bölümleme, hiper düzlem tarafından yapılır.

SVM'de eğer iki sınıf arasında daha geniş bir boşluk varsa sınıflandırma işleminin daha başarılı olunacağı idea edilmektedir (Jakkula). En geniş aralığın ortasından geçen çizgi (marjin) karar sınıfı olur. Şekil 3-1'de görüldüğü gibi "mavi" ve "kırmızı" olmak üzere iki sınıf vardır ve bu iki sınıfı birbirinden ayıran sonuz sayıda doğru çizilebilir. Fakat SVM'deki amaç iki sınıfı ayıran çizginin yani iki sınıf arasındaki boşluğun en fazla olmasıdır.



Şekil 3-1: İki Sınıfı Ayıran Sonuz Sayıda Doğru

Verilen eğitim verileri x_i (1, 2, ……, M), şeklinde olsun ve M örnekten oluşsun. Örneklerin pozitif ve negatif sınıf olmak üzere iki sınıftan oluştukları varsayılmaktadır.

Her sınıf etiketlerle ilişkilendirilmiştir ve y_i=1 pozitif sınıfları, y_i= -1 negatif sınıfları temsi etmektedir. Doğrusal olarak ayrılabilen veri olması durumunda, verilen verileri ayıran f (x) = 0 hiper düzlemini belirlemek mümkündür ve bu düzlem denklem 2'deki gibi yazılabilir.

$$f(x) = \mathrm{w}^{\mathrm{T}}\mathrm{x} + \mathrm{b} = \sum_{\mathrm{j}=1}^{\mathrm{M}} w_j x_j + \mathrm{b} = 0 \tag{2}$$

Denklem 2'de verilen w, M-boyutlu vektördür ve b, bir sabittir. w vektörü ve sabit b, hiper düzlemi ayırma pozisyonunu tanımlamak için kullanılır. Karar fonksiyonu, giriş verilerini hem pozitif sınıfta hem de negatif sınıfta sınıflandıran ayırıcı hiper düzlem oluşturmak için f (x) işaretini kullanır. Sınırları belirleyici bir ayırıcı düzlem aşağıdaki eşitliğe göre yazılabilir:

$$\begin{aligned} f(x_i) &= 1 \;\; eğer \; \mathrm{y}_i = 1, \\ f(x_i) &= -1 \;\; eğer \; \mathrm{y}_i = -1 \end{aligned} \tag{3}$$

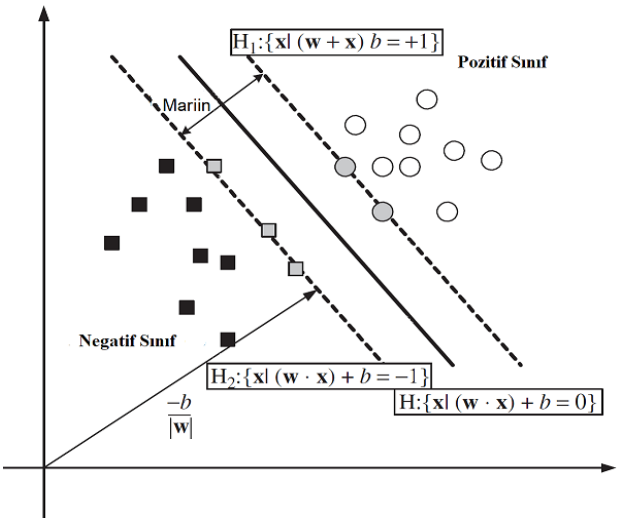
Ya da aşağıdaki gibi tam bir denklemde yazılabilir:

$$y_i f(x_i) = y_i(\mathrm{w}^{\mathrm{T}}\mathrm{x_i} + b) \geq 1, \;\; i = 1,2,\ldots,M \text{ için.} \tag{4}$$

Düzlem ile en yakın veri arasında maksimum mesafeyi oluşturan ayırıcı hiper düzlem, yani maksimum kenar boşluğu, optimum ayırıcı hiper düzlem olarak adlandırılır. Veri setindeki iki sınıfı optimal hiper düzlemde ayıran bir örnek, Şekil 3-2' de sunulmaktadır.

Şekil 3-2'de, iki farklı veri sınıfı için seri veri noktaları yer almaktadır ve negatif sınıftaki veriler siyah karelerle, pozitif sınıftaki veriler ise beyaz daireler ile ifade edilmektedir. SVM, iki farklı sınıfın arasına doğrusal bir sınır koymaya çalışır ve onu noktalı çizgiyle gösterilen kenar boşluğunu maksimize edecek şekilde yönlendirir. Ardından, sınır bu marjın ortasındaki iki nokta arasına yerleştirilir. Marjı tanımlamak için gri daireler ve kareler tarafından temsil edilen en yakın veri noktaları kullanılır ve bu marjı tanımlayan en yakın veri noktalarına ise destek vektörleri denilir. Destek vektörleri seçildiğinde, özellik setinin geri kalanı gerekli değildir, destek vektörleri sınıflandırıcıyı

tanımlamak ihtiyaç duyulan tüm bilgiyi içerebilir. Geometriden geometrik kenar boşluğunun $||w||^{-2}$ olduğu bulunmuştur.



Şekil 3-2: İki Sınıfı Ayıran Optimal Hiper Düzlem

Artık değişkenler (Slack variable) ξi ve hata cezası C alındığında, verileri ayıran en uygun hiper düzlem, aşağıdaki optimizasyon problemine bir çözüm olarak elde edilebilir:

$$\text{minimum } \frac{1}{2}\,||w||^2 + C\sum_{i=1}^{M} \boldsymbol{\xi} i \tag{5}$$

$$yi(w^T x_i + b) \geq 1 - \boldsymbol{\xi} i, \quad i = 1, \ldots\ldots, M \tag{6}$$

$$\boldsymbol{\xi} i \geq 0, \qquad i = 1, \ldots\ldots, M$$

Formülde verilen $\boldsymbol{\xi} i$, kenar boşluğu ile kenar boşluğunun yanlış tarafında duran x_i, örnekleri arasındaki mesafeyi ölçmektedir. Aşağıdaki hesaplama Kuhn-Tucker koşulundaki problemi, eşdeğer Lagrange ikili problemine dönüştürerek basitleştirilebilir;

$$\text{minimum} \quad \mathrm{L(w, b, \alpha)} = \frac{1}{2}\,||w||^2 - \sum_{i=1}^{M} \alpha_i\, y_i(w.x_i + b) + \sum_{i=1}^{M} \alpha_i \tag{7}$$

Formül 7'deki işlem denklemimizi minimize etmektedir. "w" ve "b" ile ilgili olarak, L'den a'ya kadar türevlerinin yok olmasını istiyoruz. Optimal noktada, aşağıdaki eyer noktası denklemlerini elde ediyoruz:

$$\frac{\partial L}{\partial w} = 0, \qquad \frac{\partial L}{\partial b} = 0, \tag{8}$$

Denklemimiz aşağıdaki şekilde dönüştürülebilir:

$$\mathrm{w} = \sum_{\mathrm{i}=1}^{\mathrm{M}} \alpha_i \, y_i x_i , \qquad \sum_{\mathrm{i}=1}^{\mathrm{M}} \alpha_i \, y_i = 0. \tag{9}$$

ve sonuç olarak aşağıdaki denklemler elde edilir:

$$\text{Maksimum} \quad \mathrm{L}(\alpha) = \sum_{\mathrm{i}=1}^{\mathrm{M}} \alpha_i - \frac{1}{2} \sum_{\mathrm{i,j}=1}^{\mathrm{M}} \alpha_i \, \alpha_j y_i y_j x_i . x_j \tag{10}$$

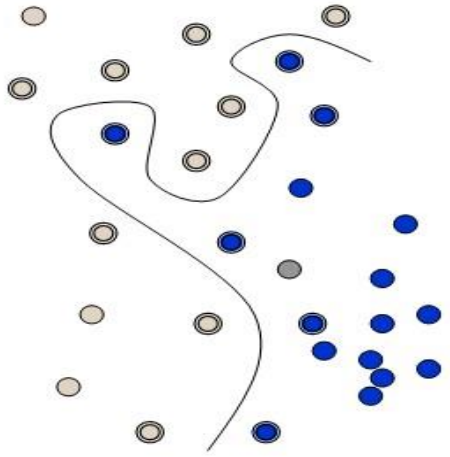
$$\alpha_i \geq 0, \quad i = 1, \ldots\ldots\ldots, M, \quad \sum_{\mathrm{i}=1}^{\mathrm{M}} \alpha_i \, y_i \; = 0 \tag{11}$$

Böylece, ikili optimizasyon problemini çözerek, Denklem 5'i çözmek ve w yi ifade etmek için gereken α_i katsayıları elde edilir.

$$f(\mathrm{x}) = \; \mathrm{sign}(\sum_{\mathrm{i,j}=1}^{\mathrm{M}} \alpha_i \, y_i \left(x_i x_j \right) + \mathrm{b}) \tag{12}$$

3.2.1.2. Doğrusal Olmayan Destek Vektör Makineleri

Doğrusal SVM sınıflandırıcıları verimli olmasına ve çoğu durumda iyi sonuçlar vermesine rağmen, birçok veri kümesi doğrusal olarak ayrılabilir durumda değildir. Bu tür veri kümleri için doğrusal olmayan hiper düzlemlere ihtiyaç vardır. Verileri sınıflandırmak için doğrusal olmayan eğiriler türetmektense çekirdek fonksiyonları kullanılır. SVM, çekirdek fonksiyonlarının uygulanmasıyla doğrusal olmayan sınıflandırma görevlerinde de kullanılabilir. Sınıflandırılacak veriler doğrusal sınıflandırmanın mümkün olduğu yüksek boyutlu bir özellik alanı üzerine eşlenir. Şekil 3-3'de doğrusal olarak ayrılamayan veriler gösterilmektedir.



Şekil 3-3: Doğrusal Olarak Ayrılamayan Veriler

Doğrusal olmayan vektör fonksiyonunu $\Phi(x)=(\Phi_l(x),....,\Phi_l(x))$ kullanarak, *n* boyutlu x girdi vektörünü *l* boyutlu özellik uzayına eşlemek için, ikili formdaki doğrusal karar fonksiyonu aşağıdaki denklemdeki gibidir:

$$f(\mathrm{x}) = \ \mathrm{sign}\left(\sum_{\mathrm{i,j=1}}^{\mathrm{M}} \alpha_i \, y_i \left(\Phi^T(x_i).\Phi(x_j)\right) + \mathrm{b}\right) \tag{13}$$

Yüksek boyutlu özellik alanında çalışmak, karmaşık fonksiyonların meydana gelmesini sağlar, aynı zamanda sorunlarda oluşturur. Çözüm olarak çekirdek fonksiyonları kullanılır. Çekirdek fonksiyonları, belirli eğitim veri vektörleri için doğrudan hesaplanır:

$$K\left(\mathrm{X_i}, \mathrm{X_j}\right) = \ \left(\Phi^T(x_i).\Phi(x_j)\right) \tag{14}$$

Karar fonksiyonu aşağıdaki eşitlikteki gibi ifade edilebilir:

$$f(\mathrm{x}) = \ \mathrm{sign}(\sum_{\mathrm{i,j=1}}^{\mathrm{M}} \alpha_i \, y_i K\left(x_i x_j\right) + \mathrm{b}) \tag{15}$$

Tablo 3-1'de bazı popüler çekirdek fonksiyonları verilmiştir.

Tablo 3-1: SVM Çekirdek Fonksiyonları

Çekirdek Fonksiyonu	**Matematiksel İfadesi**	**Parametre**
Polinom Çekirdeği	$K(x,y) = ((x.y)+1)^d$	Polinom derecesi (d)
Normalleştirilmiş Polinom Çekirdeği	$K(x,y) = \frac{((x.y)+1)^d}{\sqrt{((x.x)+1)^d((y.y)+1)^d}}$	Polinom derecesi (d)
Radyal Tabanlı Fonksiyon Çekirdeği	$K(x,y) = e^{-\gamma\|\|(x-x_i)\|\|^2}$	Çekirdek boyutu (γ)

3.2.2. K-NN

Bu yaklaşım örüntü sınıflandırmada kullanılan en basit ve en eski yöntemlerden biridir. Genellikle verimli performans sağlar ve bazı durumlarda doğruluğu, modern sınıflandırıcılardan daha yüksektir. K-NN sınıflandırıcısı, eğitim setindeki k-en yakın (en benzer) komşularından örneklerin çoğunluğunun etiketini kullanarak etiketsiz bir test örneğini sınıflandırır. K-NN'nin hesaplama karmaşıklığı, eğitim veri setinin büyüklüğü ile artar. Çünkü yeni gelen her örneğin eğitim veri setindeki her örneğe olan uzaklığı tek tek hesaplanır.

K-NN, eğitim veri setini doğrudan kullanarak tahmin yapar. K-NN sınıflandırıcı en yakın komşu sınıfını bularak hedef etiketi tahmin etmektir ve bu tahmini K değerine göre yapmaktadır. Eğitim veri kümesindeki K örneklerinden hangisinin yeni bir girdiye daha çok benzediğini belirlemek için bir mesafe ölçümü kullanılır (Moldagulova & Sulaiman, 2017). Mesafe ölçümü için en popüler olan Öklid ya da, Manhattan yöntemleri kullanılır. Bu çalışmada Öklid uzaklığı kullanılacaktır. K değeri algoritma veri seti üzerinde denenerek bulunabilir. K için birçok farklı değer verilerek en iyi "K" değeri bulunabilir.

K-NN algoritması adımları;

1. “K” parametresini belirlenir.(En yakın komşuların sayısı)
2. Yeni örneği ile tüm eğitim örnekleri arasındaki mesafeyi hesaplanır.
3. Mesafeler sıralanır ve “K” tane en yakın komşu belirlenir.
4. “K” noktadan aynı sınıfta olan örneklerin sayısı toplanır.
5. “K” nokta içinden hangi sınıfta daha çok örnek varsa yeni örneği o sınıfa ata.

3.3. Veri Seti

Bu araştırmada müzik duygu için yeni bir veri seti hazırlanmıştır. Bu alanda yapılan çalışmaları incelediğimizde çoğu araştırmacının ortak bir veri seti kullanmak yerine kendi veri setlerini hazırladıklarını görmekteyiz çünkü hala bu alanla ilgili ortak bir veri setinin var olduğunu tam olarak söylenemez.

Hangi duygu modelinin veya kaç duygu kategorisinin kullanılması gerektiği konusunda hala bir fikir birliği yoktur (Yang & Chen, 2012). Bunun yanında insan algısının öznel doğası ortak bir veri tabanı oluşturulmasını zorlaştırmaktadır. Bu nedenle tarafımızdan bu çalışmaya özel veri seti hazırlanmıştır.

Veri setini hazırlamak için Türk müziğinin farklı türlerinden sözlü ve sözsüz müzikler alınmıştır. Veri seti ayrık model olarak tasarlanmıştır ve veri setinde gergin, hüzünlü, mutlu ve rahatlatıcı olmak üzere dört sınıf mevcuttur. Bu müziklerin duygu etiketlerini belirlemek için 13 kişinin katılım sağladığı bir deneyde katılımcıların seçilen müzikleri gergin, hüzünlü, mutlu ve rahatlatıcı duygu etiketleri ile etiketlemeleri istenilmiştir. Seçilen müziklerden 30 saniyelik bölümler rastgele kesilerek katılımcılara dinletilmiştir ve her katılımcı hissettikleri duygulara göre müzikleri etiketlemiştir.

Daha sonra bu etiketlere göre bir müziğin duygu sınıfı belirlenmiştir. Müzikler en fazla hangi etiketle etiketlendiyse o etiket sınıfa dahil edilmiştir. Örneğin bir müzik 10 rahatlatıcı, 3 de mutlu etiketliyle etiketlendiyse, müzik rahatlatıcı sınıfına dahil edilmiştir. Deney 3 oturum şeklinde yapılmıştır ve her katılımcı toplamda 500 müzik dinlemiştir. Veri tabanında her sınıfta eşit sayıda örnek olsun diye her sınıf için 100 müzik belirlenmiştir. Kalan müzikler ise değerlendirmeye alınmamıştır. Orijinal veri setinde her

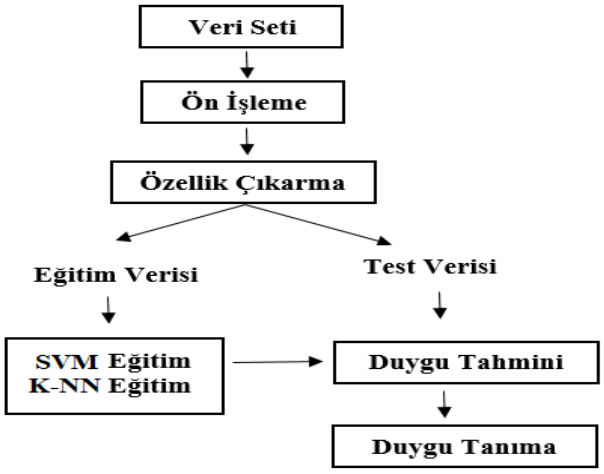
örnekten 30 saniyelik kayıtlar olmak üzere toplam 400 örnek vardır. Veri setindeki her sınıftaki örnek sayısı Tablo 3-2'de verilmiştir.

Tablo 3-2: Veri setindeki Sınıflar ve Örnek Sayıları

Sınıf	Örnek Sayısı
Mutlu	100
Hüzünlü	100
Gergin	100
Rahatlatıcı	100

3.4. Duygu Sınıflandırmasında Müziksel Özelliklerin Değerlendirilmesi

Müzikal duygular ve ses özellikleri arasındaki ilişkiyi daha iyi açıklamak için genel olarak dört kategoride incelenen ses özellikleri karşılaştırılmıştır. Deneysel prosedür; veri ön işleme, özellik çıkarma ve sınıflandırma olarak üç aşamadan oluşmaktadır. Şekil 3-4'de , bu üç aşama gösterilmektedir. Veri setindeki her örnek için farklı kategorilerden öznitelik çıkarılmıştır ve sınıflandırma işlemi için Matlab kullanılmıştır.



Şekil 3-4: Deneysel İşlemin Aşamaları

3.4.1. Veri ön işleme

Veri setindeki müzik kayıtları mp3 formatında 41100 Hz örnekleme frekansına sahiptir. Özellik çıkarılmadan önce her müzik kaydı tek kanala dönüştürülmüştür.

3.4.2. Öznitelik Çıkarımı

Seçilen eserler duygusal yönden kategorize edildikten sonra özniteliklerin çıkarılması için müzik işlemede kullanılan MIRtoolbox araç kutusu kullanılmıştır. MIRtoolbox yardımı ile tempo, ses yüksekliği, ton, ritim vb. gibi farklı gruplardan ses öznitelikleri çıkarabilmektedir. MIRtoolbox kullanılarak toplam 38 özellik değerine sahip özellik vektörü elde edilmiştir.

3.4.3. Öznitelik Vektörü

Deneyde kullanılan öznitelikler tablo 3-3'de verilmiştir.

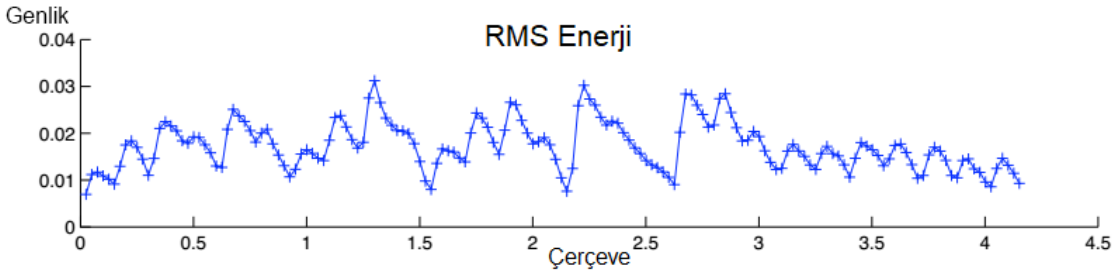
Tablo 3-3: Deneyde Kullanılan Akustik Öznitelikler

Boyut No.	**Özellik Adları**	**Özellik Türü**	**Boyut**
Enerji	Kare Ortalamalarının Karekökü (Root Mean Square Energy)	Ortalama	1
	Düşük Enerji (Low energy)	Ortalama	1
Ritim	Tempo	Ortalama	1
	Ritmik Dalgalanma (Fluctuation)	Ortalama	1
Tını	Yüksek Frekans/Tizlik (Brightness)	Ortalama	1
	Spektral Dağılım (Spectrum centroid)	Ortalama	1
	Spektral Entropi (Spectral entropy)	Entropi	1
	Spektral Çarpıklık (Skewness)	Ortalama	1
	Rolloff	Ortalama	1
	Sıfır Eksenini Geçme Sayısı (Zero Crossing rate)	Ortalama	1
	MFCC	Ortalama	13
	Atak Zamanı (Attacktime)	Ortalama, Standart sapma	2
	Atak Eğimi (Attack Slope)	Ortalama	1
Harmoni	Chromagram	Ortalama	12

3.4.3.1. Kare Ortalamalarının Karekökü (Root Mean Square - RMS)

Ses sistemlerinin güçlerini hesaplanmada kullanılır. Ses sinyali x'in enerjisi, genlik değerlerinin karelerinin toplamının karekök ortalaması alarak hesaplanabilir (Lartillot, 2018). RMS formülü denklem 16'da verilmiştir. Şekil 3-5'de belirli sayıda örneklerle çerçevelere ayrılmış bir sinyalin RMS enerji eğrisi gösterilmektedir.

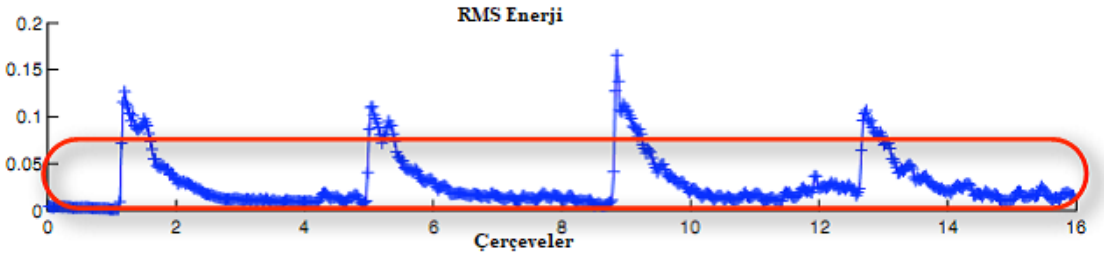
$$x_{rms} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} x^2{}_i} = \sqrt{\frac{x^2{}_1 + x^2{}_2 + \cdots + x^2{}_n}{n}} \qquad (16)$$



Şekil 3-5: RMS Enerji Eğrisi (Lartillot, 2018)

3.4.3.2. Düşük Enerji (Low energy)

Enerji eğrisi, sinyalin zaman boyunca sabit kaldığını veya bazı çerçevelerin diğerlerinden daha farklı olup olmadığını görmek için, geçici enerji dağılımını değerlendirmek için kullanılabilir. Düşük enerji oranını tahmin etmenin bir yolu, ortalamadan daha az enerji gösteren çerçeve yüzdesini hesaplamaktır (Tzanetakis & Cook, 2002). Şekil 3-6'da, bu özelliğin görselleştirilmiş hali sunulmaktadır. Enerji eğrisinin seçilen kısmı, ortalamadan düşük enerji değerini ifade eder.



Şekil 3-6: Sinyaldeki Düşük Enerji Miktarının Gösterimi (Lartillot, 2018)

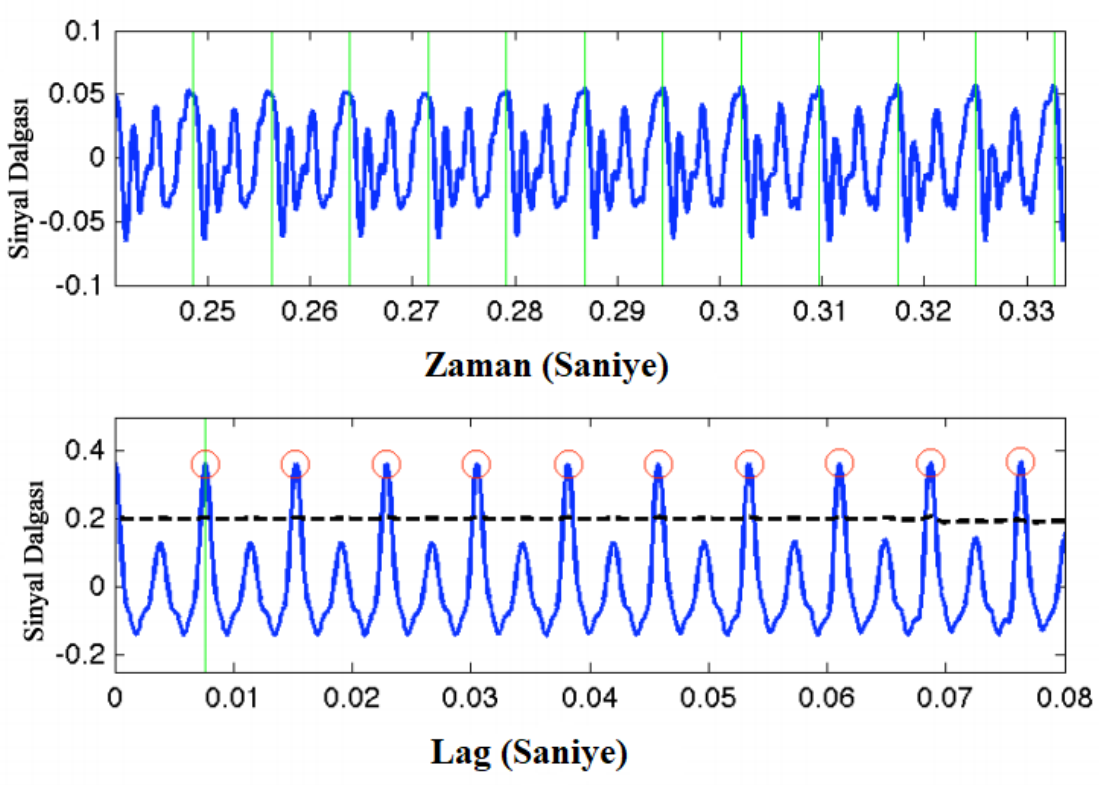
Şekil 3-6’da özellikle yüksek enerji içeren bazı nadir çerçeveler vardır ve bu nedenden dolayı çerçevelerin çoğunun ortalama RMS'nin altında olduğunu görebiliriz.

3.4.3.3. Tempo

Seçilmiş olan birimin bir dakika içeresindeki tekrarı müziğin hızını temposunu belirler. Müzikteki tempoyu ölçmek için kullanılan bir birim olan bpm(dakikadaki vuruş sayısı) dır. Olay algılama eğrisinden periyodiklikleri tespit ederek tempo tahmin edilebilir. Tempo tahmini için klasik paradigma periyodikliklerin tespit edilmesine dayanmaktadır (Davies & Plumbley , 2007).

Tempo tahmininde son derece doğru sonuç veren yöntem oto korelasyon dur. Oto korelasyon verilen bir sinyalin farklı zamanlardaki benzerliklerini belirler. Referans olarak seçilen bir sinyal penceresinin diğer pencereler ile korelasyonunu gerçekleştirir. Sıfırıncı zaman periyodunda da korelasyon, maksimum olacaktır çünkü referans penceresi kendi kopyası ile karşılaştırmış olacaktır.

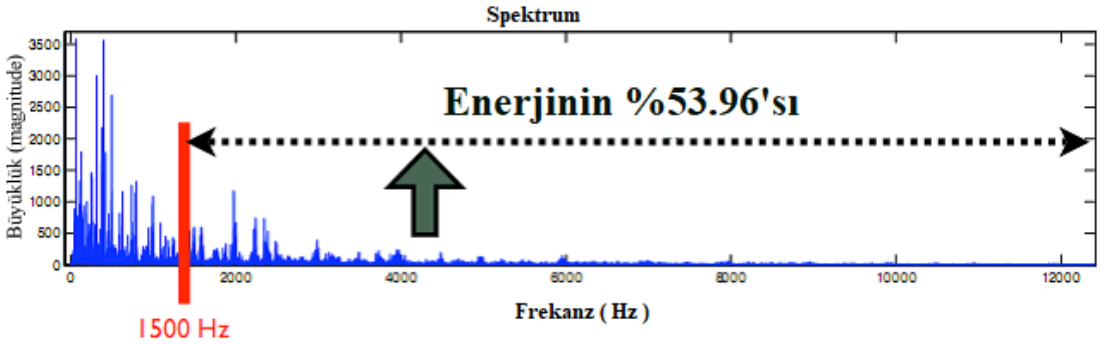
Zaman periyodu ilerledikçe korelasyon mutlaka azalacaktır, fakat sinyalin periyodik olması durumunda, bir noktada tekrar artmaya başlayacak, sonra yerel bir maksimum seviyeye ulaşacaktır. Başlangıç ile ilk pik noktası arasındaki mesafe bize temponun periyodunu verecektir. Örneğin bu mesafe 0.5 saniyelik bir periyod olursa tempo (dakikadaki vuruş sayısı, beat per minute) 120 bpm olacaktır. Şekil 3-7’de oto korelasyon uygulanmış periyodik bir sinyalin pik noktalarının gösterimi yer almaktadır.



Şekil 3-7: Periyodik Sinyalin Pik Noktaları

3.4.3.4. Yüksek Frekans/Tizlik(Brightnes)

Bu özellik, belirli bir kesme frekansını sabitleyerek ve bu frekansın üzerindeki enerji miktarını ölçerek elde edilir (Juslin, 2000). Şekil 3-8'de bir örnek (1500 Hz) gösterilmektedir.



Şekil 3-8: Yüksek frekanslı enerji örneği (Lartillot, 2018)

3.4.3.5. Spektral Özellikler

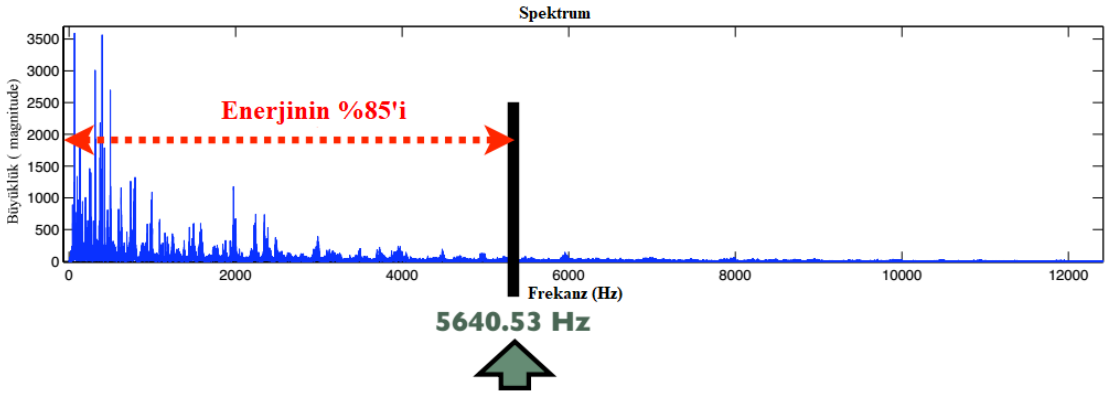
Spektral Dağılım (Spectrum centroid) : Spektral dağılımın istatistiksel ortalaması, isteğe bağlı olarak ayrı frekans alt bantlarından hesaplanan centroidler üzerinden elde edilen ortalama olarak ifade edilebilir. Centroid, bir spektral şeklin ölçüsüdür ve daha yüksek centroid değerleri daha yüksek frekanslara sahip seslerle ilgilidir. Spektral centroid kullanıcı deneyleri tarafından enstrüman tınılarının karakterizasyonunda önemli bir algısal özellik olduğu gösterilmiştir. (Tzanetakis & Cook, 2002)

Spektral Entropi (Spectral Entropy) : Spektral dağılımın entropisi, isteğe bağlı olarak ayrı frekans alt-bantlarında hesaplanan spektral entropiler üzerinden alınan ortalama ile elde edilir (Toh, Togneri & Nordholm, 2005).

Spektral Çarpıklık (Skewness) : İsteğe bağlı olarak ayrı frekans alt-bantlarında hesaplanan çarpıklık ölçümleri üzerinden elde edilen ortalama spektral dağılımın çarpıklık katsayısıdır (Lidy & Rauber, 2006).

3.4.3.6. Rolloff

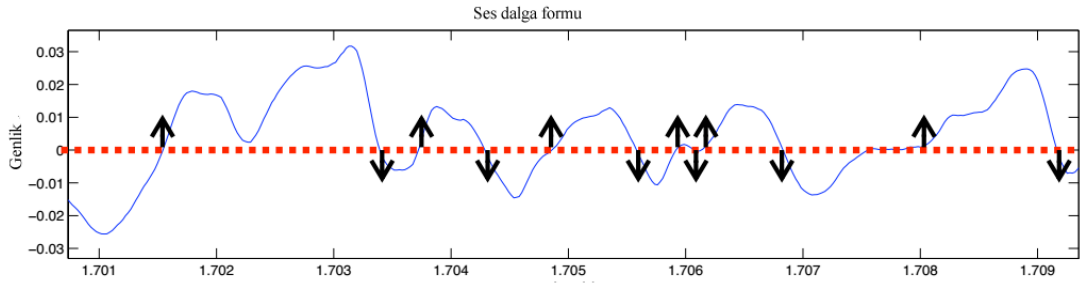
Sinyaldeki yüksek frekansın miktarını tespit etmenin bir yolu, toplam enerjinin belirli bir parçasının o frekansın altında olduğu frekansı bulmayı içerir. Bu oran aksi belirtilmediği sürece %85'e sabitlenmiştir (Tzanetakis & Cook, 2002). Yani öyle bir frekans değeri bulunmalıdır ki sitemdeki frekansın %85'i bu frekans değerinin altında yer alsın. Buna bir örnek Şekil 3-9'da verilmiştir.



Şekil 3-9: Spektral roll off noktası örneği (Lartillot, 2018)

3.4.3.7. Sıfır Eksenini Geçme Sayısı (Zero-crossing Rate)

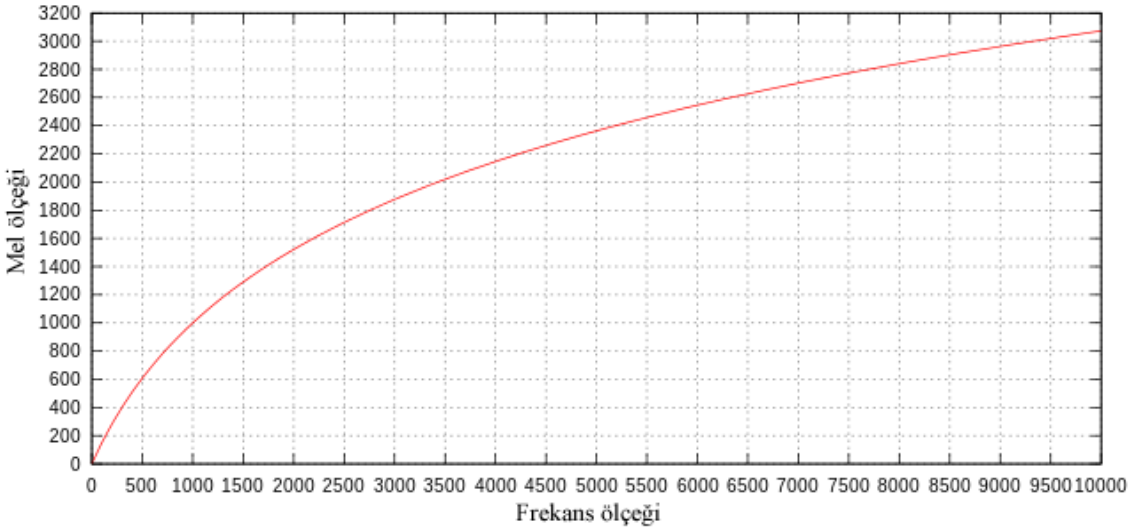
Dalga formunun bir pencerede yer değiştirilme sayısını belirtir. Sinyalin X ekseni kaç kez geçtiğini gösterir, frekansın yanı sıra gürültü belirtisi olarak da kullanılabilir (Lartillot, 2018). Ses de bölünmeler olup olmadığını incelemek için kullanılır. Şekil 3-10'de örnek bir sinyalin x eksenini geçtiği noktaların gösterimi yer almaktadır.



Şekil 3-10: Bir Sinyalin x Eksenini Geçtiği Noktaların Gösterimi (Lartillot, 2018)

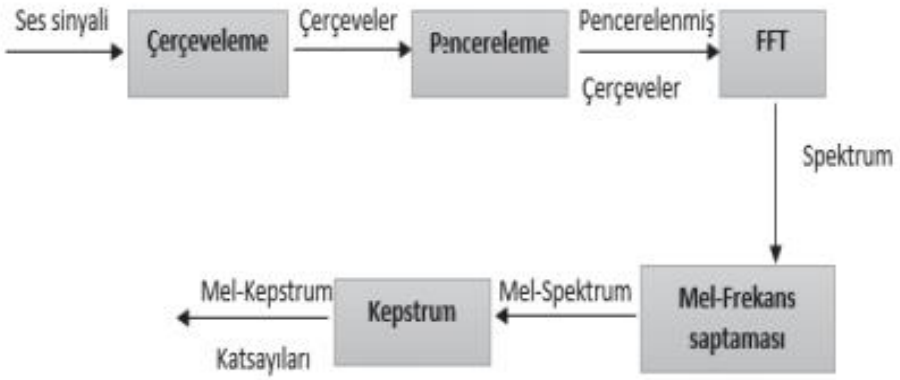
3.4.3.8. Mel Frekans Kepstrum Katsayıları (MFCC)

Mel, bir ses tonun insan kulağı tarafından algılanan frekansının ölçüsüdür. Bu ölçülen değer ses tonun fiziksel frekansıyla lineer olarak uyuşmaz çünkü bir insan kulağı ses frekanslarını lineer olmayan bir şekilde algılar (Chauhan & Desai, 2014). Yapılan çalışmalar sonucunda ölçüler, 1 kHz e kadar doğrusal, daha yüksek değerlerde ise logaritmiktir bir artış olduğu gözlenmiştir. Mel ölçeği, insanların müzik tonunu nasıl duyduklarını yansıtır. Aşağıdaki şema da fiziksel frekans (Hz) ve algılanan frekans (mel) ölçeği arasındaki ilişki yer almaktadır. Şekil 3-11'de frekans- mel ölçeği verilmiştir.



Şekil 3-11: Frekans- Mel Ölçeği (wikipedia, 2018)

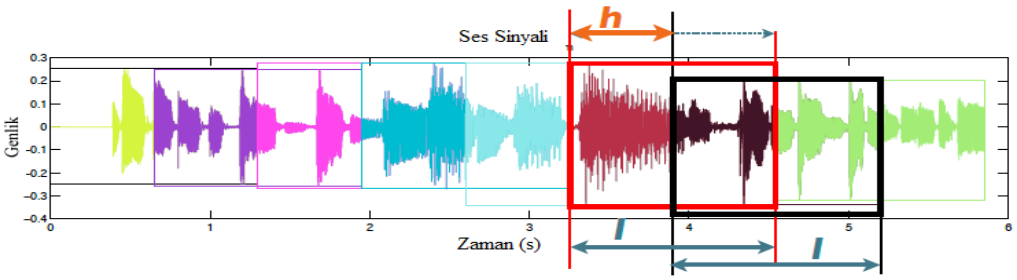
MFCC ses tanımada yaygın olarak kullanılmaktadır. MFCC insan kulağının frekans seçiciliğini taklit ederek iyi bir şekilde konuşmacıları ayırt edici değerler elde edilmesidir (On & diğerleri., 2006). Mel frekans kepstrum katsayılılarının çıkarılabilmesi için uygulanması gereken adımlar şekil 3-12'de verilmiştir.



Şekil 3-12: MFCC Adımları

Çerçeveleme: Ses sinyalindeki işaretler zamana bağlı olarak sürekli değişmektedir. Bu yüzden ses sinyallerinin karakteristik özelliklerini belirleyen parametreler çok küçük zaman diliminde kararlı kalmaktadır. Yapılan araştırmalara göre bu zaman dilimi de 20 ve 100 ms arasındadır. Bu aralıklarda ses sinyalleri sabit karakteristik özellik gösterir. Bu çalışmada ses sinyalleri 25 ms uzunluklarda çerçevelere bölünmüştür. Ayrıca her çerçeve

kendinden bir önceki çerçevenin bir kısmını örtecek şekildedir. Çerçevelerinin birbirini örtmesinin amacı çerçeveler arsı geçişin yumuşak olmasını sağlamaktır (ATAL, 1976). Eğer bu örtüşme olmasaydı birbirini izleyen iki çerçeveyi izlediğimiz zaman veya bu çerçevedeki sinyalleri işlenirken çerçeveler arasındaki özellik değişimi kesintiye veya sıçramaya neden olabilecekti. Ayrıca komşu çerçevelerin parametre değerlerinin farkı da daha yüksek olacaktı. Bu nedenlerden dolayı çerçeveler birbirini örtecek şekildedir. Bu çalışmada çerçevelerin örtüşme oranı bir çerçevenin %50 si kadar olarak seçilmiştir. Şekil 3-13'de örnek bir sinyal üzerinde çerçeveleme işlemi gösterilmektedir.



Şekil 3-13: Sinyal Çerçeveleme

Pencereleme: Bir sinyale çerçeveleme uygulandıktan sonra bu işlem yapılır. Amaç her bir çerçevenin başında sonunda oluşan süreksizliği önlemektir. Her çerçevedeki ses sinyalinin orta bölgesi güçlendirilir uç bölgeler ise zayıflatılır. Bu çalışmada yaygın olarak kullanılan "Hamming penceresi" kullanılmıştır. "Hamming Pencereleme" sayesinde sinyalin merkez bölgesi daha belirgin hale gelir. Uç kısımlarda sıfıra yaklaştırılır. Hamming Penceresi sinyalin uç bölgelerdeki istenmeyen radyasyonları en aza indirir (Gupta & diğerleri., 2013). Böylece sesin kalitesini veya harmonilerini sinyalin sinüs dalga şeklinden uzaklaşması iyileştirir.

Ayrıca sinyali Fourier dönüşümüne uygun hale getiren fonksiyondur. Hamming fonksiyonun çalışma mantığı hamming penceresi ile çerçevelenmiş ses sinyali çarpılır ve sonunda pencerelenmiş sinyal elde edilir. Hamming penceresinin uzunluğu ile çerçevenin uzunluğu eşittir. Hamming pencere formülü denklem 17'de verilmiştir.

$$w[n] = 0.54 - 0.46 * \cos\left(2 * \pi * \frac{n}{N}\right), \quad N: Pencere\ uzunluğu \tag{17}$$

$$n: 0, 1, 2 \dots \dots N - 1$$

FFT: Sinyal işlemede dönüşüm bir sinyalin başka parametrelerle ifade edilmesi anlamına gelir. Fourier dönüşümü bir sinyalin zaman domaininden frekans domainine dönüşümünü sağlar. Foruier dönüşümü sinyal işlemede sinyalin içindeki bilgileri elde etmek için çok önemli bir yöntemdir. Sinyali oluşturan bileşenleri (frekans, genlik, faz) kosinüs ve sinüs fonksiyonlarının toplamı olarak ifade etmeye yarar. Bu çalışmada Hızlı Fourier Dönüşümü (FFT) kullanılmıştır. FFT, Ayrık Zamanlı Foruier Dönüşümü ile aynı sonucu verecektir. Fakat yapılması gereken aritmetiksel işlem sayısı az olduğu için daha hızlı sonuç verecektir.

Bu adımda N örnekten oluşan her çerçeve Hızlı Fourier dönüşümü uygulanarak zaman domainden frekans domainine geçirilmiştir. N örnekli bir küme denklem 18'deki gibi tanımlanır.

$$X_n = \sum_{k=0}^{N-1} x_k \, e^{-2\pi jkn \,/\, N}, \qquad \mathrm{n} = 0, 1, 2, \ldots\ldots, \mathrm{N} - 1 \tag{18}$$

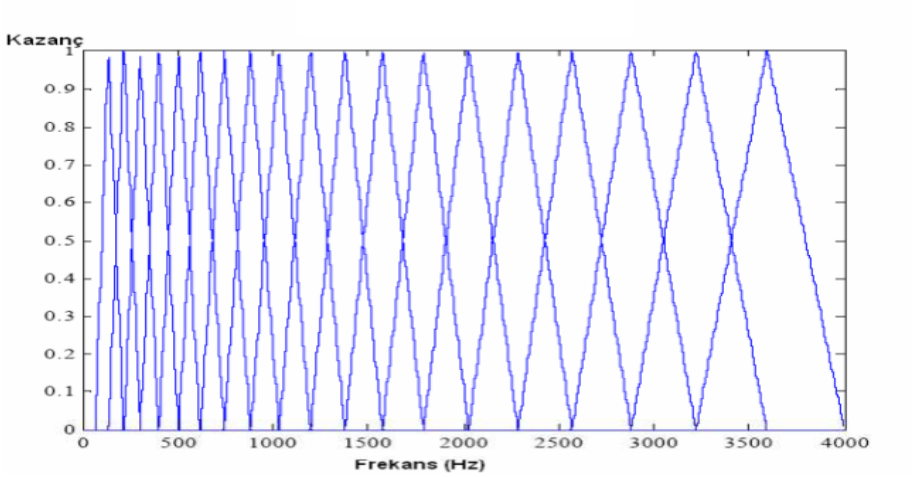
Mel -Frekansına Çevirme: Mel-ölçeği, düşük frekanslarda daha ayrımcı ve daha yüksek frekanslarda daha az ayrımcı olarak, sesin doğrusal olmayan insan kulağı algısını taklit etmeyi amaçlamaktadır. Üst bölümde de belirtildiği gibi insan kulağı sesleri 1khz'e kadar doğrusal, 1kz'den yüksek değerler için logaritmik olarak algılamaktadır.

Denklem 18 ve 19 kullanarak Mel ölçeği(M) ile frekans ölçeği(Hz) arasındaki dönüşüm sırasıyla yapılabilir.

$$m = 2595 log_{10}(1 + \frac{f}{700}) \tag{18}$$

$$f = 700(10^{\frac{m}{2595}} - 1) \tag{19}$$

Bu algısal yapıyı ifade etmenin bir yolu mel filitre bankası kullanmaktır. Bu filtre bankasının yapısı üçgen bant geçiren şekildedir. Bant genişliği sabit mel frekansı aralıklarıyla belirlenir. Filtre bankaları %50 oranında birbiri örtecek şekilde üst üste gelmiştir (Öztürk, 2010). Kulağın duyma hassasiyetini modellemek için bu yapı kullanılmıştır. Mel filitre bankası şekil 3-14'de verilmiştir.



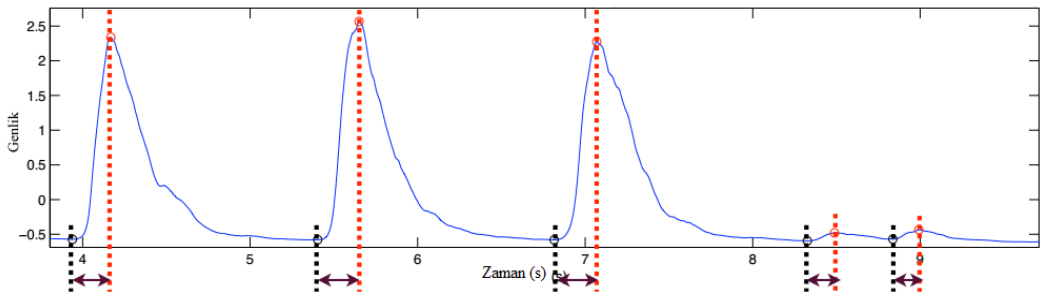
Şekil 3-14: Mel Filitre Bankası

Kepstrum: Son adım, kepstrum katsayılarının elde edilmesine yönelik bir çalışmadır. Tüm filtre bankasısın enerjilerinin logaritmasını alınır. Ayrık kosinüs dönüşümü uygulanarak Mel Frekans Kepstrum katsayıları elde edilir (Öztürk, 2010). MFCC aşağıdaki formülü denklem 20'de verilmiştir.

$$MFCC_i = \sum_{k=1}^{20} X_k . cos\left[i.\left(\frac{k-1}{2}\right).\frac{\pi}{20}\right] \quad i = 1,2,\dots,M \tag{20}$$

3.4.3.9. Atak Zamanı (Attack Time)

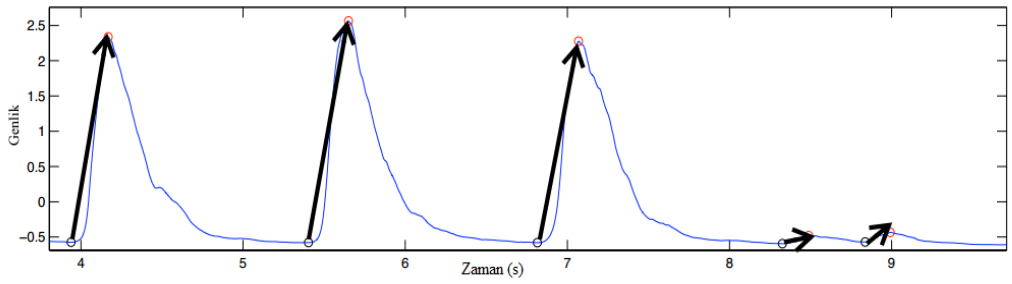
Atak zamanı, bir sinyalin zirvesine yükselmesi için geçen zamansal sürenin tahminidir. Bu özelliği tanımlamanın ve hesaplamanın basit bir yolu sinyalin genliğinin yükselişe geçtiği fazın aralığının zamansal süresini tahmin etmektir (Lartillot, 2018). Sinyal atak zamanı örneği şekil 3-15'de verilmiştir.



Şekil 3-15: Sinyal Atak Zamanı (Lartillot, 2018)

3.4.3.10. Atak Eğimi (Attack Slope)

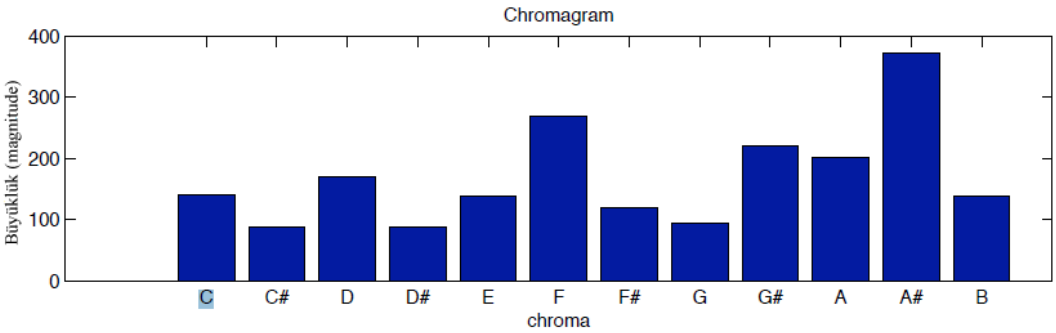
Bu özellik, atak periyodunun bitiş ve başlangıç değerlerini farkının, bu periyodun oluşması için gereken zaman farkına oranı olarak hesaplanır. Bu özelliğin örneği şekil 3-16'da verilmiştir.



Şekil 3-16: Sinyal Atak Eğimi (Lartillot, 2018)

3.4.3.11. Chroma

Her nota etrafındaki enerji dağılımını gösterir. Her notanın belirli bir frekans aralığı vardır. Bu notaların frekans aralıklarındaki enerji yoğunluğunu hesaplar (Lartillot, 2018). Bu özelliğin örneği şekil 3-17'de verilmiştir. Toplamda 7 nota vardır ve batı müziğinde mi ve fa notası hariç diğer her iki nota arası iki eşit parçaya bölündüğü için aradaki seslerde alınarak 12 tane öz nitelik ve Chroma spektrogramı elde edilebilmektedir.



Şekil 3-17: Chromagram

3.4.4. Deneysel Çalışmalar ve Sınıflandırma Sonuçları

Bu deneyde, Matlab kullanarak, sınıflandırıcı olarak SVM ve K-NN seçilmiştir ve sınıflandırma işlemi için 10- kat çapraz doğrulama uygulanmıştır. SVM'ler polinom ve radyal temelli fonksiyon çekirdekleri kullanılarak eğitilmiştir. Dört müzik boyutundaki özellikleri daha iyi anlamak ve karşılaştırmak için deney iki farklı şekilde yapılmıştır.

Deney1'de SVM'nin ve K-NN'nin performansı çeşitli özellikler kullanılarak karşılaştırılmıştır. Deney 1'de, iki farklı çekirdekle eğitilmiş SVM'ler ve KNN karşılaştırılmıştır. Her bir modelin sınıflandırma süreleri farklıdır. SVM Polinom çekirdeği ile en yüksek sınıflandırma başarısı %66 iken, RBF çekirdeği için en yüksek sınıflandırma başarısı %52'dir. Veri Seti'nin K-NN ile sınıflandırılması sonucu en yüksek %53.25 oranında başarı elde edilmiştir. (K sayısı 5 olarak seçilmiştir.) SVM, polinom çekirdeğinin burada kullanılan parametre değerleri için duygu sınıflandırma işleminde RBF çekirdeğinden daha iyi performans gösterdiği gözlenmektedir. SVM ve K-NN için sınıflandırma sonuçları sırasıyla tablo 3-4 ve 3-5'de verilmiştir.

Tablo 3-4: SVM Sınıflandırma Sonuçları

Özellik Sınıfı	**Polynomial kernel**		**RBF kernel**		**Boyut**
	Doğruluk %	**Zaman**	**Doğruluk %**	**Zaman**	
Enerji	31.75	0.06	29.5	0.11	2
Ritim	39.25	0.09	34.75	0.12	2
Tını	63	0.11	46	0.13	22
Harmoni	42.75	0.11	40.25	0.13	12
Enerji+Ritim+Tını+Harmoni	**66**	0.13	**52**	0.14	38

Tablo 3-5: K-NN Sınıflandırma Sonuçları

Özellik Sınıfı	K=5 **Doğruluk %**	**Boyut**
Enerji	34.25	2
Ritim	44.25	2
Tını	52	22
Harmoni	45	12
Enerji+Ritim+Tını+Harmoni	**53.25**	38

Şekil 3-18 ve 3-19'de sırasıyla SVM ve K-NN den elde edilen en iyi sınıflandırma sonuçlarına ait karışıklık matrisleri verilmiştir.

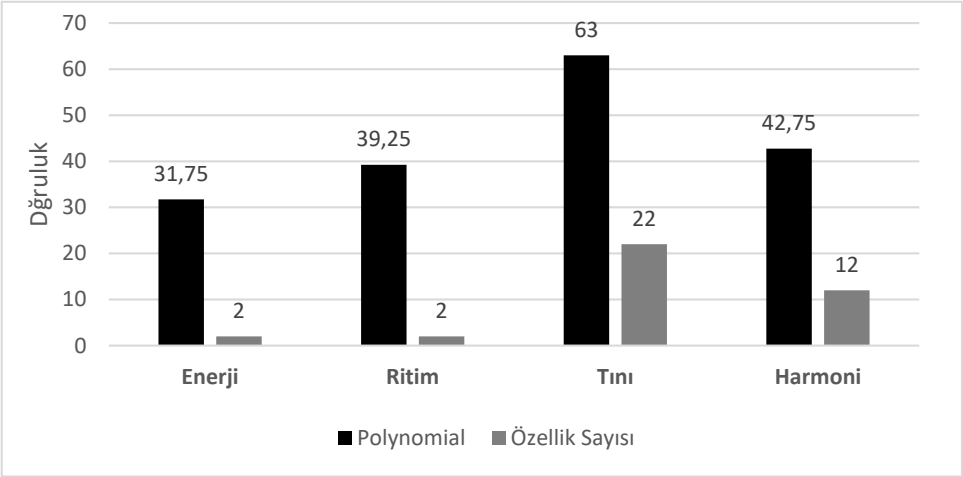
Karışıklık Matrisi		TAHMİN				
		MUTLU	**HÜZÜNLÜ**	**GERGİN**	**RAHATLATICI**	**TOPLAM**
GERÇEK	**MUTLU**	**81**	10	9	0	**100**
	HÜZÜNLÜ	16	**45**	14	25	**100**
	GERGİN	10	12	**71**	7	**100**
	RAHATLATICI	6	17	10	**67**	**100**
	TOPLAM	**113**	**84**	**104**	**99**	

Şekil 3-18: SVM Karışıklık Matrisi

Karışıklık Matrisi		TAHMİN				
		MUTLU	**HÜZÜNLÜ**	**GERGİN**	**RAHATLATICI**	**TOPLAM**
GERÇEK	**MUTLU**	**75**	5	18	2	**100**
	HÜZÜNLÜ	16	**40**	33	11	**100**
	GERGİN	17	4	**78**	1	**100**
	RAHATLATICI	15	28	37	**20**	**100**
	TOPLAM	**123**	**77**	**166**	**34**	

Şekil 3-19: KNN Karışıklık Matrisi

Deney 2'de farklı sınıftaki müzikal özellikler ayrı ayrı test edilip baskın bir sınıf bulmak için sonuçlar karşılaştırılmıştır. Sonuçlar Şekil 3-20'de verilmiştir. Harmoni ve ritim özellik gurubu yaklaşık olarak benzer sonuçlar vermiştir, fakat en iyi sınıflandırma sonucu tını özellik gurubu ile % 63 olarak edilmiştir.



Şekil 3-20 : Farklı Türdeki Özniteliklerden Elde Edilen Sınıflandırma Sonuçları

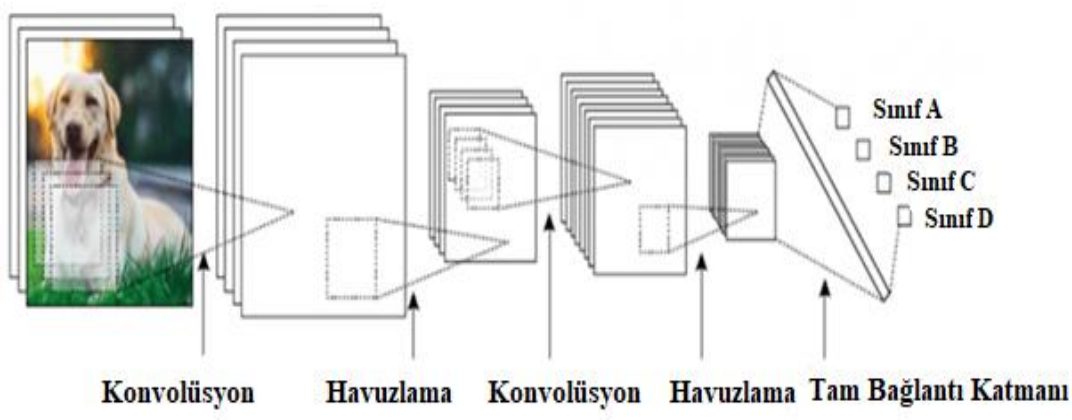
3.5. Duygu Sınıflandırmasında Transfer Öğrenmenin Değerlendirilmesi

Çalışmanın bu bölümünde, veri setindeki müziklerden elde edilen Chroma Spektrogram görüntülerinden derin özellikler çıkarılıp, önceden eğitilmiş ağlar ile transfer öğrenme uygulanarak müzik duygu tanıma için yeni bir yöntem önerilmiştir. Önerilen yöntemden elde edilen sonuçlar gözetimli makine öğrenmesi yöntemlerinden elde edilen sonuçlar ile karşılaştırılmıştır. Ayrıca önerilen yöntemdeki teknikler kullanılarak MATLAB ortamında müzik duygu tanıma için yazılım geliştirilmiştir. Müzik duygu tanıma sisteminde mp3 formatında kaydedilmiş müzikten elde edilen Chorma özelliği ile daha önceden veri tabanına kaydedilmiş müziklerden elde edilen Chorma özellikleri arasında eşleştirme yapılarak ve müzikte geçen duygunun ne olduğu tespit edilmektedir.

3.5.1. Konvolüsyonel Sinir Ağları

CNN birçok farklı disiplinde görüntü analizine başarıyla uygulanan derin ve ileri beslemeli yapay sinir ağları sınıfına aittir. Biyolojik olarak hayvan görsel kortekslerinin çalışma biçiminden ilham almıştır ve çok katmanlı algılayıcıların türevi olan en güçlü

görsel işleme sistemidir (Fu & Aldrich, 2009). Bu tür ağlar geniş ve derin yapıda olmalarından dolayı derin öğrenmenin bir türüdür. Konvolüsyonel ağlar, resim tabanlı nesnelerin sınıflandırılmasıyla ilgili problemlerde elde ettikleri büyük başarılar sebebiyle günümüzdeki en popüler modelleme tekniğidir. Şekil 3-21'de bir Konvolüsyonel ağın genel mimari yapısı verilmiştir. Klasik bir CNN genel olarak, konvolüsyonel katmanları, havuzlama katmanları, tamamen bağlı katmanlar ve sınıflandırma katmanı dâhil olmak üzere alternatif katman türlerine sahiptir (Cirean & diğerleri., 2011). CNN'lerde tamamen bağlanmış katmanlar, yapay sinir ağlarındakine benzer olarak nöronlar önceki katman ile tam olarak bağlanır.



Şekil 3-21: Konvolüsyonel Sinir Ağın'ın Genel Yapısı

Derin sinir ağları olarak, CNN'ler görüntü verilerinden karmaşık derin özellikler çıkararak, görüntü tanıma problemleri için oldukça verimli çözümler üretmektedir ayrıca bu ağlar, geleneksel yaklaşımları büyük bir farkla geride bırakmıştır (Simonyan & Zisserman, 2015). Fakat derin sinir ağlarının tasarımı ve bu ağlarının etkili bir şekilde eğitilmesi, büyük miktarda veri ve üst düzey hesaplama kaynakları gerektirir. Bir derin öğrenme modelini eğitmek için gerekli olan verileri ve kaynakları göz önünde bulundurduğumuzda, transfer öğrenme derin öğrenmeden son derece avantajlıdır.

Transfer öğrenme, bir problem türünde eğitim sırasında edinilen bilgi kazanımının diğer problem tiplerini eğitmek için kullanan bir makine öğrenme tekniğidir. Transfer öğrenme, yeteri miktarda veri olmadığı durumda derin sinir ağını eğitmek için çok uygun ve etkili bir yöntemdir. Önceden tanımlanmış ağdaki tüm parametrelerin başlangıç olarak

kullanılması, büyük görüntülerden öğrenilen özelliklerden yararlanması sınıflandırma başarısını artırmada etkili olmaktadır (Lu, Lu & Zhang, 2019).

Bu çalışmada, transfer öğrenme için önceden eğitilmiş ağ modeli olarak AlexNet mimarisi kullanılmıştır. Transfer öğrenme işlemini uygulamak için, önceden eğitilmiş olan AlexNet mimarisi özellik çıkarıcı olarak kullanılmıştır ve çıkarılan özellikler softmax ve SVM sınıflandırıcıları eğitmek için kullanılmıştır.

3.5.2. Alexnet

AlexNet, yıllık olarak yapılan ImageNet büyük ölçekli Görsel tanıma yarışmasını 2012 yılında kazanan evrimsel sinir ağıdır (Krizhevsky, Sutskever & Hinton, 2012). Bu mimari, 1000 ortak nesnenin yaklaşık 1.2 milyon görüntüden oluşan ImageNet veri tabanı üzerinde önceden eğitilmiştir ve 1000 adet görüntüyü ayırt edebilen bir ağ yapısına sahiptir (Russakovsky & diğerleri., 2015). Orijinal çıkış katmanının eldeki sınıflandırma problemiyle ilgili uygun bir katman olarak belirlenmesi ile bu ağın bir özellik çıkarıcı olarak kullanılması mümkündür. Önceden eğitilmiş AlexNet mimarisi temel olarak 5 konvolüsyonel katmanı ve 3 tamamen bağlı (fully connected) katmandan oluşur.

Konvolüsyon katmanında belirli bir yükseklikte ve genişlikte olan filtreler giriş görüntüsü üzerinde belirlenen kaydırma adımları (Stride) ile soldan sağa doğru dolaştırılır. Konvolüsyon, model görüntülerinden özellik öğrenilmesini sağlar ve model karmaşıklığını azaltır. (i, j) boyutunda bir R görüntüsü verildiğinde, konvolüsyon Denklemi 21'deki gibi tanımlanır.

$$f(i,j) = (R * w)(i,j) = \sum_{x}\sum_{y} R(i-x, j-y)w(x,y) \tag{21}$$

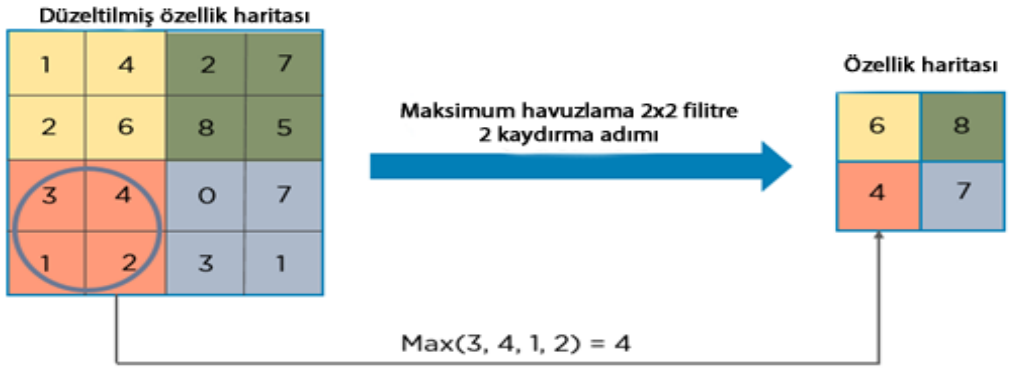
Formülde verilen F, özellik haritasını, w ise (x,y) boyutunda konvolüsyon çekirdeğini ifade etmektedir. Birinci katmana 227 piksel genişlikte 227 piksel yükseklikte 3 renk kanalından oluşan giriş görüntüsü verilir. Bu giriş görüntüsü üzerine genişliği 11 piksel yüksekliği 11 piksel olan 3 kanallı filtre uygulanır. Birinci katmanda toplam 96 adet filtre uygulanmaktadır. Daha sonraki adım aktivasyon fonksiyonudur. Aktivasyon için Düzleştirilmiş Doğrusal Birim (Rectified Linear Units) kullanılır. Bu işlem her konvolüsyon katmanından sonra uygulanır. Konvolüsyon sonunda giriş verisinde oluşan

negatif değerleri sıfıra çeker. ReLu'nun amacı derin ağı doğrusal olamayan bir yapıya getirmektir. ReLu fonksiyonu Denklem 22'de verilmiştir.

$$\mathrm{Re}\,LU(x) = \max(x,0) \tag{22}$$

Birinci katmana benzer şekilde ikinci katmanda 5x5 ebatlarında 256 filtre, üçüncü katmanda 3x3 ebatlarında 384 filtre, dördüncü katmanda 3x3 ebatlarında 384 ve beşinci katmanda 3x3 ebatlarında 256 filtre uygulanır. Ayrıca her konvolüsyon ve ReLu işleminden sonra havuzlama (pooling) uygulanır. Havuzlama işleminin temel amacı, özellik azaltmadır. Bir sonraki konvolüsyon katmanına gönderilecek girdi görüntüsünün boyutunu genişlik ve yükseklik bakımdan azaltılır.

Özellik haritasındaki komşu piksel gurupları temsil eden bir değer oluşturur. Şekil 3-22'de 'de maksimum havuzlama (maxpooling)gösterilmektedir. Özellik haritası 4×4' tür, maksimum havuzlama özelliği her 2 × 2 bloktaki maksimum değeri oluşturur, bu da özellik boyutunu önemli ölçüde azaltır.



Şekil 3-22: Maksimum Havuzlama

Son üç katman tam bağlantı katmandır. Tamamen bağlanmış olan 6. ve 7. katmanlarda bütün nöronların birbirine bağlandığı 4096 nöron bulunur. Son katman ise sınıflandırma işleminin yapıldığı katmandır. Konvolüsyon, ReLu ve havuzlama işlemlerinden sonra tam bağlı katman gelir. Bu katmadaki her nöron kendisinden bir önceki katmanın tüm nöronlarına tam bağlı şekildedir. AlexNet'in FC6 ve FC7 katmanının her birinde 4096 tane nöron bulunur ve bu nöronlar birbirine tam bağlıdır.

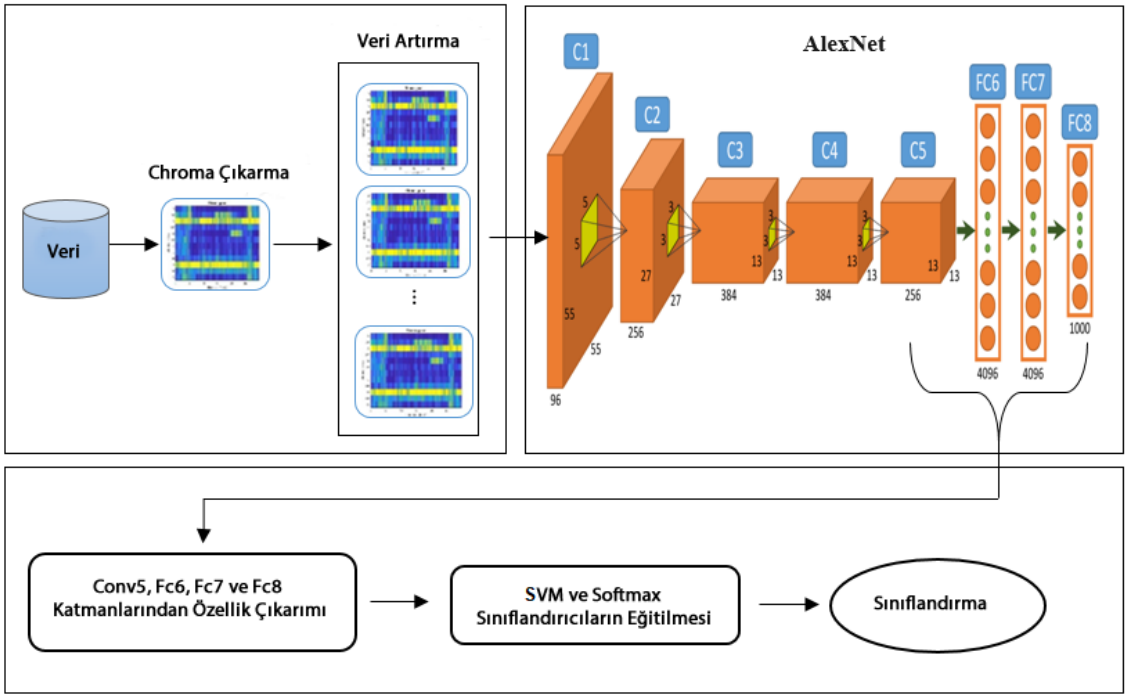
Son katman ise sınıflandırma katmanıdır. Bu katmandaki nöron sayısı veri setindeki sınıf sayısı kadardır. Genellikle Softmax sınıflandırıcı kullanılırken farklı sınıflandırıcılarda kullanılabilir. Herhangi bir test verisinin her bir sınıfa ait olma olasılığını hesaplar. Softmax Denklemi 23'de verilmiştir. Softmax, nöronların çıkış değerlerini (0,1) aralığında olmasını sağlar.

$$softmax(x)_j = \frac{e^{xi}}{\sum_{n=1}^{N} e^{xn}} \ , \ j = 1 \ldots\ldots N \tag{23}$$

3.5.3. Önerilen Yöntem

Bu çalışmada derin ağlara transfer öğrenme uygulanarak müzik duygu tanıma için yeni bir yöntem önerilmiştir. Müzik duygu tanıma için önerilen yöntem 5 adımdan oluşmaktadır.

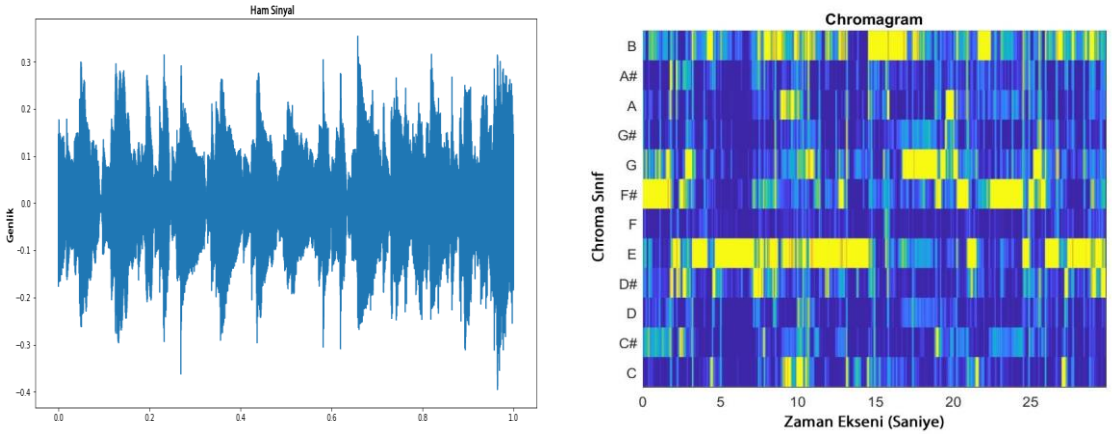
Birinci adımda veri setindeki müzik kayıtlarından Chroma spektrogramlar çıkarılmıştır. İkinci adımda veri setini büyütmek için veri artırma işlemi yapılmıştır. Veri artırma için veri setindeki her müzik kaydına üç farklı deformasyon uygulanıp veri seti 4 katına çıkarılmıştır. Üçüncü adımda Chroma spektrogramlar daha önceden yaklaşık olarak bir milyon görüntüyle eğitilmiş olan AlexNet modeline girdi olarak verilerek transfer öğrenme uygulanmıştır. Dördüncü adımda derin öznitelik çıkarma katmanları belirlenmiştir. Ağın daha iyi öğrenmesini sağlamak ve daha derin özellikler çıkarmak için "Conv5", "Fc6", "Fc7" ve "Fc8" katmanları öznitelik çıkarıcı katman olarak seçilmiştir. Beşinci adımda ise çıkarılan derin görsel özellikler kullanılarak sınıflandırma işlemi yapılmıştır. Sınıflandırıcı olarak SVM (Polinom Çekirdek) ve Softmax kullanılıp, farklı oranlarda alınan eğitim ve test verileri ile bu sınıflandırıcılar eğitilmiştir. Önerilen yöntemin çalışma ilkesi şekil 3-23'deki gibidir.



Şekil 3-23: Önerilen Yöntemin Çalışma İlkesi

3.5.3.1. Chroma Spektrogramların Çıkarılması

Her nota belirli bir frekans aralığına sahiptir ve bu notların müzik içindeki enerji dağılımını gösteren Chroma özellikleri spektrogram haline getirilebilir. MIRtoolbox yardımıyla ham ses verilerinden Chroma spektrogramlar elde edilebilir. Örnek bir ses sinyali ve Chroma görüntüsü Şekil 3-24'de verilmiştir. Şekil 3-24'de verilen Chroma spektrogramındaki C, D, E, F, G, A, B harfleri sırasıyla do, re, mi, fa, sol, la ve si notalarını ifade eder. "#" işaretli harfler ise iki nota arasındaki koma sesleri ifade eder.

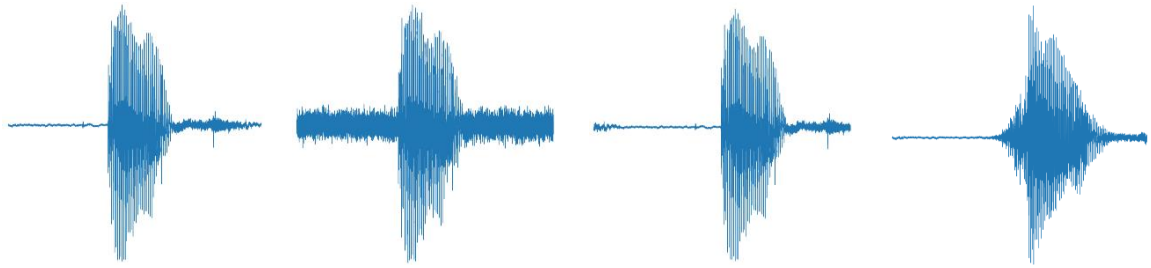


Şekil 3-24: Ses Sinyali ve Chroma Görüntüsü

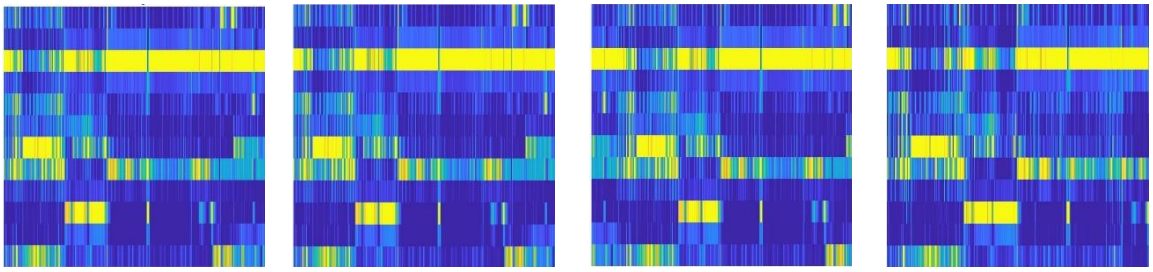
3.5.3.2. Veri Artırma (Data Augmentation)

Derin öğrenme de, tatmin edici sınıflandırma performans oranlarına ulaşmak ve eğitim verisiyle test verisi arasındaki hata farkını azaltmak için çoğu durumlarda büyük miktarda eğitim verisi gerekir (Wang & Perez, 2017). Orijinal veri boyutunun sınırlı olduğu durumlarda, bu veri azlığı probleminin üstesinden gelmek için veri artırma işlemi gereklidir. Veri artırma, eğitim veri setindeki verilere bir dizi deformasyon uygulayarak ek eğitim veri örnekleri oluşturulmasıdır (McFee, Humphrey & Bello, 2015). Veri artırmadaki temel prensip, etiketli verilere uygulanan deformasyonlar sayesinde oluşan yeni verilerin etiketlerinin değiştirilmemesidir (Salamon & Bello, 2017).

Veri artırma için, görüntünün farklı açılarda döndürülmesi, yatay dikey çevrilmesi, görüntüye gürültü ve renk manipülasyonu ekleme gibi birçok teknik mevcuttur. Bu çalışmada, veri setindeki her bir örneğinin Chroma spektrogramını çıkarmadan önce sinyale 3 farklı işlem uygulandı. Birincisi gürültü (noise) eklenmesi, ikincisi sinyalin kaydırılması (shifting) üçüncüsü ise sinyalin uzatılması (stretch) işlemidir. Her işlem sonunda elde edilen ses örnekleri orijinal ses örneğinin sınıfına yeni veri olarak ekledi. Veri setindeki orijinal örnek de dâhil edildiğinde, veri setimizin boyutu tam dört katına çıkarılmış oldu. Veri artırma öncesi ve sonrası ses sinyallerinin gösterimi ve Chroma özellikleri şekil 3-25'de verilmiştir. Orijinal veri seti ve veri artırma sonrasında elde edilen kayıt sayısı Tablo 3-6'de verilmiştir.



Orijinal Sinyal Gürültülü Sinyal Kaydırılmış Sinyal Gerilmiş Sinyal



Orijinal Chroma Gürültülü Chroma Kaydırılmış Chroma Gerilmiş Chroma.

Şekil 3-25: Veri Artırmanın Gösterimi

Tablo 3-6: Veri Arttırmasından Önce ve Sonra Her Bir Sınıftaki Örnek Sayısı

Sınıf	**Veri Arttırma Öncesi Kayıt Sayısı**	**Veri Arttırma Sonrası Kayıt Sayısı**
Mutlu	100	400
Hüzünlü	100	400
Gergin	100	400
Rahatlatıcı	100	400

3.5.3.3. Alexnet İle Transfer Öğrenme Ve Derin Özellik Çıkarma İşlemi

Bu çalışmada yeni nesne sınıfını sınıflandırmak için önceden eğitilmiş derin öğrenme modellerinin uygulaması olan transfer öğrenme yaklaşımı kullanılmıştır ve bu

bağlamda AlexNet modelli incelenmiştir. Kendi veri setlerimizi önceden eğitilmiş bu mimariler ile sınıflandırmak istediğimiz zaman farklı modellerde transfer öğrenme işlemi uygulanabilir. Daha önceden eğitilmiş ağların sonundaki bazı katmanları değiştirip kendi veri setimiz ile bu ağları yeniden eğitebilir veya bu ağları kullanarak kendi veri setimizden derin öznitelikler çıkarıp bir sınıflandırıcı eğitebiliriz.

Bu araştırmada AlexNet' in katmanlarından derin özellikler çıkarılarak transfer öğrenme işlemi yapılmış ve daha sonra bu verilerin bir kısmı sınıflandırıcıları eğitmek bir kısım da sınıflandırıcıları test etmek için kullanılmıştır. AlexNet' in temel olarak 5 konvolüsyonel ve 3 tam bağlı katmanı vardır. Bu katmanlar transfer öğrenme için özellik çıkarıcı katman olarak kullanılabilir. Bu çalışmada öznitelik çıkarıcı katman olarak conv5, Fc6, Fc7 ve Fc8 katmanları seçilmiştir. Matlab kütüphanesinde bulunan AlexNet modeli uygulama için kullanılmıştır. AlexNet modelinin yapısı Tablo 3-7'de verilmiştir.

Tablo 3-7: Matlab'da AlexNet'in Modeli (mathworks Documentation, 2019)

No	Katman İsmi	Katman Tipi	Açıklama
1	'data'	Image Input	227x227x3 images with 'zerocenter' normalization
2	'conv1'	Convolution	96 11x11x3 convolutions with stride [4 4], and padding [0 0 0 0]
3	'relu1'	ReLU	ReLU
4	'norm1'	Cross Channel Normalization	cross channel normalization with 5 channels per element
5	'pool1'	Max Pooling	3x3 max pooling with stride [2 2] and padding [0 0 0 0]
6	'conv2'	Convolution	256 5x5x48 convolutions with stride [1 1], and padding [2 2 2 2]
7	'relu2'	ReLU	ReLU
8	'norm2'	Cross Channel Normalization	cross channel normalization with 5 channels per element
9	'pool2'	Max Pooling	3x3 max pooling with stride [2 2] and padding [0 0 0 0]
10	'conv3'	Convolution	384 3x3x256 convolutions with stride [1 1], and padding [1 1 1 1]
11	'relu3'	ReLU	ReLU
12	'conv4'	Convolution	384 3x3x192 convolutions with stride [1 1], and padding [1 1 1 1]
13	'relu4'	ReLU	ReLU
14	'conv5'	Convolution	256 3x3x192 convolutions with stride [1 1], and padding [1 1 1 1]
15	'relu5'	ReLU	ReLU
16	'pool5'	Max Pooling	3x3 max pooling with stride [2 2] and padding [0 0 0 0]

17	'fc6'	Fully Connected	4096 fully connected layer
18	'relu6'	ReLU	ReLU
19	'drop6'	Dropout	50% dropout
20	'fc7'	Fully Connected	4096 fully connected layer
21	'relu7'	ReLU	ReLU
22	'drop7'	Dropout	50% dropout
23	'fc8'	Fully Connected	1000 fully connected layer
24	'prob'	Softmax	softmax
25	'output'	Classification Output	crossentropyex with 'tench' and 999 other classes

3.5.4. Ayarlar

Önerilen yöntemin uygulaması, i7 2.50GHz işlemci, 12GB hafıza ve NVIDIA 940M GPU donanım özelliklerine sahip makine üzerinde yapılmıştır. Uygulama için gerekli yazılım kodları Matlab yazılımı matlab2018a versiyonu kullanılarak hazırlanmıştır. Veri büyütme işlemi için Python programlama dili kullanılmıştır.

3.5.5. Deneysel Sonuçlar

Sınıflandırma doğruluğunu hesaplamak için, AlexNet'in farklı katmanlarından çıkarılan derin öznitelikler kullanılarak artırılmış veri seti ve orijinal veri seti kullanılarak deneyler yapıldı. Veri setinden rastgele alınan verilerin bir kısmı eğitim, diğer kısmı da test için kullanılmıştır. Eğitim ve test için ayrılan veri boyutunun ağın performansı üzerindeki etkisini belirlemek için veri 3 farklı şekilde bölünmüştür. İlk deneyde verinin %50'si eğitim %50'si test, ikinci deneyde verinin %60'ı eğitim %40'i test üçüncü deneyde ise verinin %70'i eğitim %30'u test için kullanılmıştır. AlexNet + Transfer Öğrenme işleminin uygulanmasından sonra "Conv5", "Fc6", " Fc7" ve "Fc8 " gibi farklı katmanlardan elde edilen derin özniteliklerle eğitilen SoftMax ve SVM sınıflandırma sonuçları sırasıyla Tablo 3-8, Tablo 3-9,Tablo 3-10 ve Tablo3-11'de verilmiştir. Tablo 3-8'da "Conv5" katmanı için sınıflandırma sonuçları verilmiştir.

Tablo3-8'deki verilerde görüldüğü gibi veri artırma işleminden önce en yüksek sınıflandırma başarısı %58.3 olarak SVM ile elde edilmiştir. Veri artırma sonrasında ise en yüksek sınıflandırma başarısı yine SVM ile elde edilmiştir ve bu değer %72.1 dir. Şekil 3-26'da " Conv5" katmanı için veri artırma öncesi ve sonrası elde edilen en iyi sınıflandırma sonuçlarına ait karışıklık matrisleri ve Roc eğrileri verilmiştir.

Tablo 3-8: Conv5 Katmanı İçin Veri Artırma Öncesi ve Sonrası Sınıflandırma Sonuçları

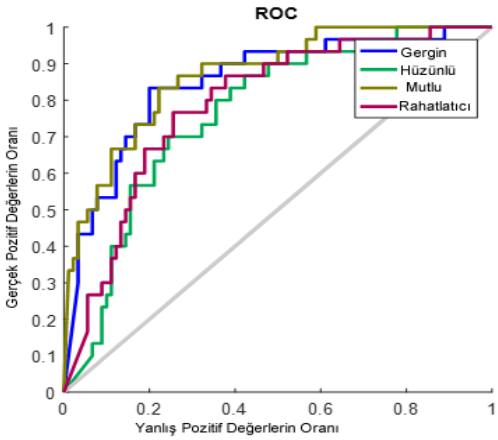
Katman	Sınıflandırıcı	Eğitim-Test Verilerini Ayırma	Veri Arttırmadan önceki Doğruluk (%)	Veri Arttırdıktan Sonraki Doğruluk (%)
Conv5	SVM	%50 - %50	53.5	65.8
Conv5	SVM	%60 - %40	56.9	69.5
Conv5	SVM	%70 - %30	58.3	72.1
Conv5	Softmax	%50 - %50	53.5	66.6
Conv5	Softmax	%60 - %40	54.4	68.4
Conv5	Softmax	%70 - %30	57.5	69.2

Karışıklık Matrisi		Tahmin			
		Gergin	Hüzünlü	Mutlu	Rahatlatıcı
Gerçek	Gergin	**19**	3	7	1
	Hüzünlü	4	**15**	3	8
	Mutlu	5	2	**20**	3
	Rahatlatıcı	1	10	3	**16**

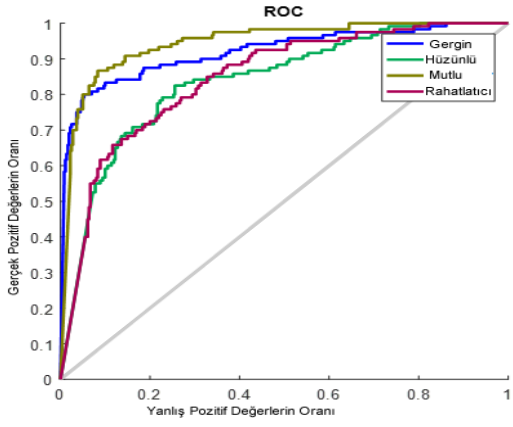
(a) Veri büyütmeden önce karışıklık matrisi

Karışıklık Matrisi		Tahmin			
		Gergin	Hüzünlü	Mutlu	Rahatlatıcı
Gerçek	Gergin	**88**	12	13	7
	Hüzünlü	2	**83**	5	30
	Mutlu	11	7	**97**	5
	Rahatlatıcı	4	34	4	**78**

(b) Veri arttırmadan sonra karışıklık matrisi



(c) Veri arttırmadan önce Roc eğrisi



(d) Veri arttırma sonrası Roc eğrisi

Şekil 3-26: Conv5 Katmanı için - Karşılaştırma Matrisi ve Roc Eğrisi

Tablo 3-9'da "Fc6" katmanı için sınıflandırma sonuçları verilmiştir. Tablo 3-9' daki verilerde görüldüğü gibi veri artırma işleminden önce veri setinin eğitim ve test için

sırasıyla (%70 - %30) oranında bölünmesiyle en yüksek sınıflandırma başarısı %74.2 olarak Softmax ile elde edilmiştir. Veri artırma sonrasında ise veri setinin eğitim ve test için sırasıyla (%70 - %30) oranında bölünmesiyle en yüksek sınıflandırma başarısı SVM ile elde edilmiştir ve bu değer %91' dir. Şekil 3-27'de " Fc6" katmanı için veri artırma öncesi ve sonrası elde edilen en iyi sınıflandırma sonuçlarına ait karışıklık matrisleri ve Roc eğrileri verilmiştir.

Tablo 3-9: Fc6 Katmanı İçin Veri Artırma Öncesi ve Sonrası Sınıflandırma Sonuçları

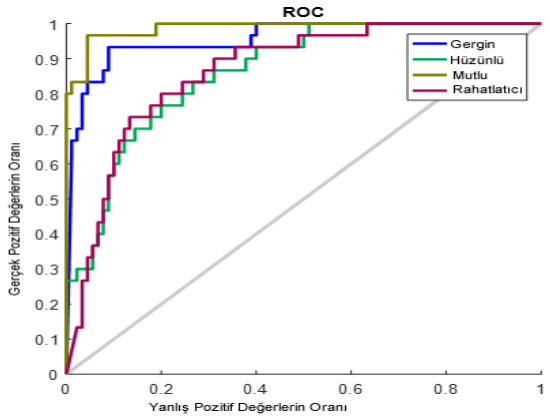
Katman	Sınıflandırıcı	Eğitim-Test Verilerini Ayırma	Veri Arttırmadan önceki Doğruluk (%)	Veri Arttırdıktan Sonraki Doğruluk (%)
Fc6	SVM	%50 - %50	69.5	88.6
Fc6	SVM	%60 - %40	73.1	89.8
Fc6	SVM	%70 - %30	74.0	91.0
Fc6	SoftMax	%50 - %50	67.5	86.9
Fc6	SoftMax	%60 - %40	71.3	88.9
Fc6	SoftMax	%70 - %30	74.2	90.6

Karışıklık Matrisi		Tahmin			
		Gergin	Hüzünlü	Mutlu	Rahatlatıcı
Gerçek	Gergin	**26**	1	2	1
	Hüzünlü	3	**17**	0	10
	Mutlu	2	2	**25**	1
	Rahatlatıcı	3	6	0	**21**

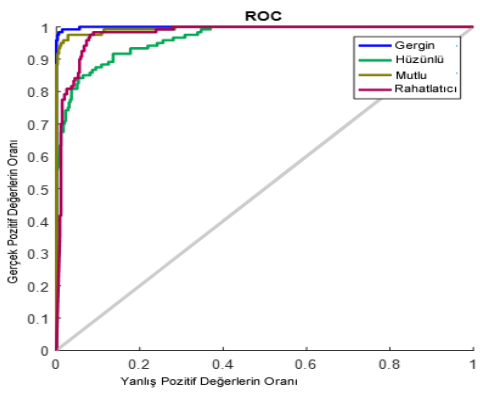
(a) Veri büyütmeden önce karışıklık matrisi

Karışıklık Matrisi		Tahmin			
		Gergin	Hüzünlü	Mutlu	Rahatlatıcı
Gerçek	Gergin	**119**	0	1	0
	Hüzünlü	2	**97**	1	20
	Mutlu	2	3	**114**	1
	Rahatlatıcı	0	12	1	**107**

(b) Veri arttırmadan sonra karışıklık matrisi



(c) Veri arttırmadan önce Roc eğrisi



(d) Veri arttırma sonrası Roc eğrisi

Şekil 3-27: Fc6 Katmanı için - Karşılaştırma Matrisi ve Roc Eğrisi

Tablo 3-10'da "Fc7" katmanı için sınıflandırma sonuçları verilmiştir. Tablo 3-10' daki verilerde görüldüğü gibi veri artırma işleminden önce veri setinin eğitim ve test için sırasıyla (%70 - %30) oranında bölünmesiyle en yüksek sınıflandırma başarısı %73.3 olarak Softmax ile elde edilmiştir. Veri artırma sonrasında ise veri setinin eğitim ve test için sırasıyla (%70 - %30) oranında bölünmesiyle en yüksek sınıflandırma başarısı SVM ile elde edilmiştir ve bu değer %92.7 dir. Şekil 3-28' de " Fc7" katmanı için veri artırma öncesi ve sonrası elde edilen en iyi sınıflandırma sonuçlarına ait karışıklık matrisleri ve Roc eğrileri verilmiştir.

Tablo 3-10: Fc7 Katmanı İçin Veri Artırma Öncesi ve Sonrası Sınıflandırma Sonuçları

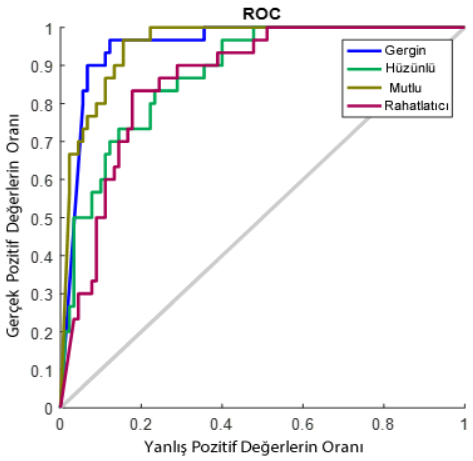
Katman	**Sınıflandırıcı**	**Eğitim-Test Verilerini Ayırma**	**Veri Arttırmadan önceki Doğruluk (%)**	**Veri Arttırdıktan Sonraki Doğruluk (%)**
Fc7	SVM	%50 - %50	70.0	87.3
Fc7	SVM	%60 - %40	71.3	89.7
Fc7	SVM	%70 - %30	72.5	92.7
Fc7	SoftMax	%50 - %50	69.0	86.0
Fc7	SoftMax	%60 - %40	70.6	87.8
Fc7	SoftMax	%70 - %30	73.3	91.0

Karışıklık Matrisi		Tahmin			
		Gergin	**Hüzünlü**	**Mutlu**	**Rahatlatıcı**
Gerçek	**Gergin**	**27**	0	1	2
	Hüzünlü	0	**17**	3	10
	Mutlu	5	1	**23**	1
	Rahatlatıcı	1	6	2	**21**

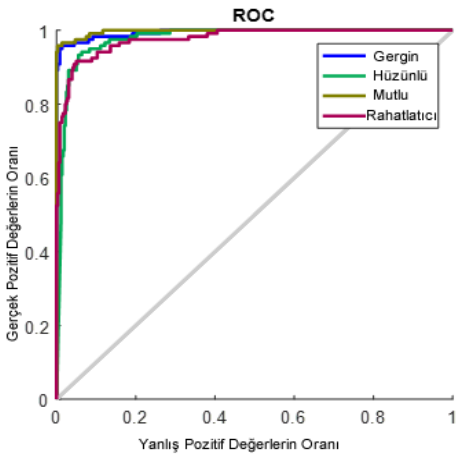
(a) Veri büyütmeden önce karışıklık matrisi

Karışıklık Matrisi		Tahmin			
		Gergin	**Hüzünlü**	**Mutlu**	**Rahatlatıcı**
Gerçek	**Gergin**	**114**	2	1	3
	Hüzünlü	0	**107**	4	9
	Mutlu	1	2	**116**	1
	Rahatlatıcı	1	9	2	**108**

(b) Veri arttırmadan sonra karışıklık matrisi



(c) Veri arttırmadan önce Roc eğrisi



(d) veri arttırma sonrası Roc eğrisi

Şekil 3-28: Fc7 Katmanı için - Karşılaştırma Matrisi ve Roc Eğrisi

Tablo 3-11'de "Fc8" katmanı için sınıflandırma sonuçları verilmiştir. Tablo 3-11' deki verilerde görüldüğü gibi veri artırma işleminden önce veri setinin eğitim ve test için sırasıyla (%70 - %30) oranında bölünmesiyle en yüksek sınıflandırma başarısı %70.8 olarak Softmax ile elde edilmiştir. Veri artırma sonrasında ise veri setinin eğitim ve test için sırasıyla (%70 - %30) oranında bölünmesiyle en yüksek sınıflandırma başarısı SVM ile elde edilmiştir ve bu değer %87.3 dür. Şekil 3-29' da " Fc8" katmanı için veri artırma öncesi ve sonrası elde edilen en iyi sınıflandırma sonuçlarına ait karışıklık matrisleri ve Roc eğrileri verilmiştir. Ayrıca her katman için elde edilen en iyi sınıflandırma sonuçları da Şekil 3-30'da verilmiştir.

Tablo 3-11: Fc8 Katmanı İçin Veri Artırma Öncesi ve Sonrası Sınıflandırma Sonuçları

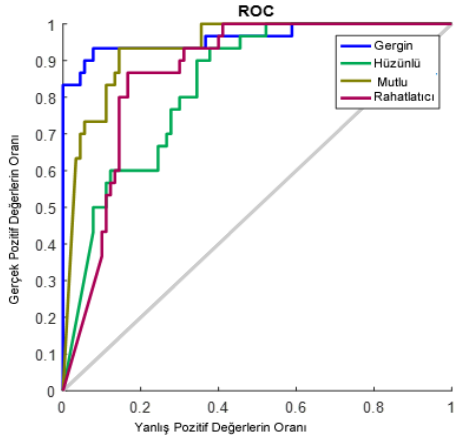
Katman	Sınıflandırıcı	Eğitim-Test Verilerini Ayırma	Veri Arttırmadan önceki Doğruluk (%)	Veri Arttırdıktan Sonraki Doğruluk (%)
Fc8	SVM	%50 - %50	63.5	84.4
Fc8	SVM	%60 - %40	63.7	85.6
Fc8	SVM	%70 - %30	68.8	87.3
Fc8	SoftMax	%50 - %50	63.1	82.0
Fc8	SoftMax	%60 - %40	66.9	84.7
Fc8	SoftMax	%70 - %30	70.8	85.2

Karışıklık Matrisi		Tahmin			
		Gergin	Hüzünlü	Mutlu	Rahatlatıcı
Gerçek	Gergin	**25**	1	4	0
	Hüzünlü	1	**18**	1	10
	Mutlu	1	4	**22**	3
	Rahatlatıcı	1	9	0	**20**

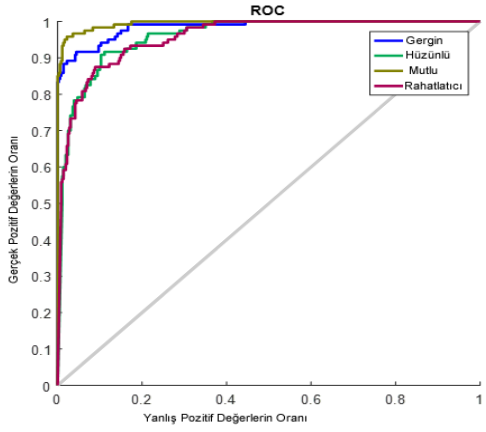
(a) Veri büyütmeden önce karışıklık matrisi

Karışıklık Matrisi		Tahmin			
		Gergin	Hüzünlü	Mutlu	Rahatlatıcı
Gerçek	Gergin	**108**	6	3	3
	Hüzünlü	4	**100**	2	14
	Mutlu	1	2	**115**	2
	Rahatlatıcı	4	18	2	**96**

(b) Veri arttırmadan sonra karışıklık matrisi

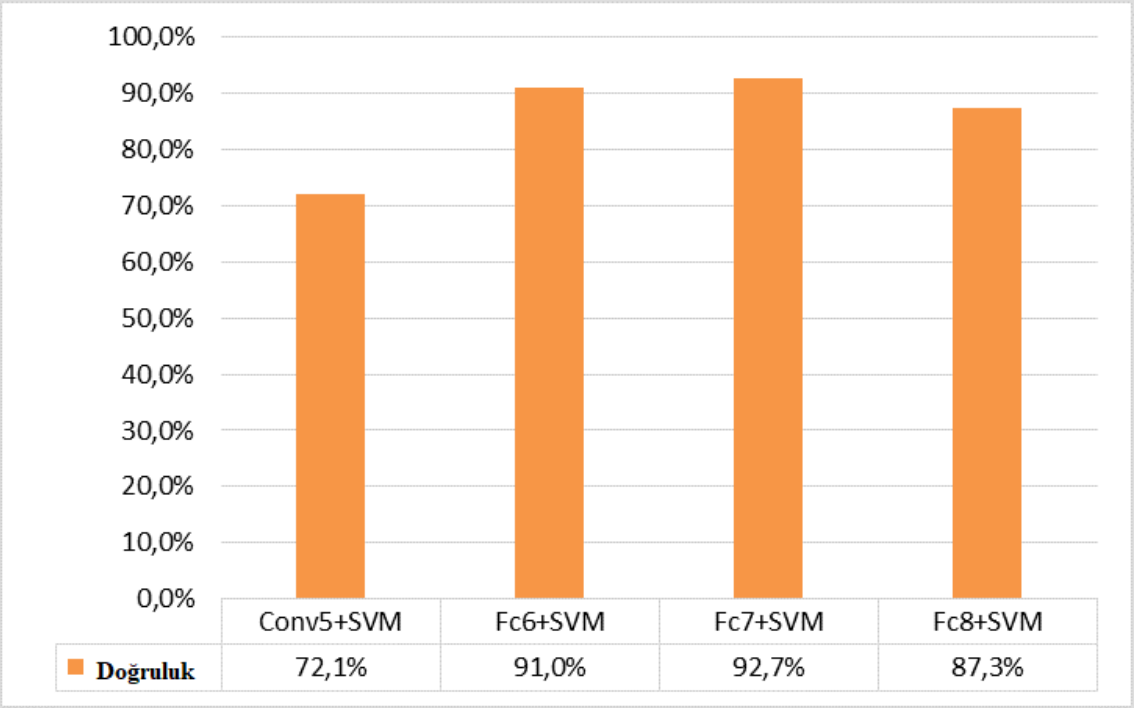


(c) Veri arttırmadan önce Roc eğrisi



(d)) veri arttırma sonrası Roc eğrisi

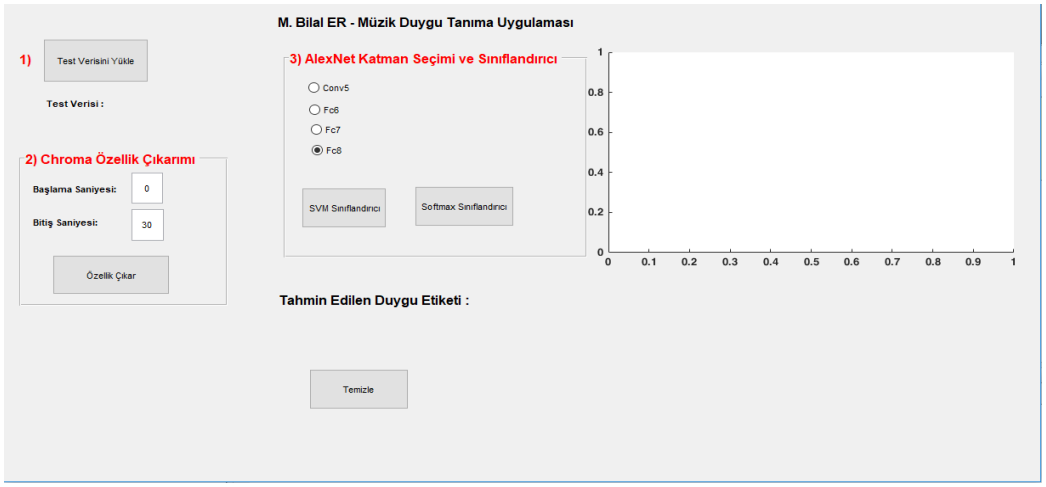
Şekil 3-29: Fc8 Katmanı için - Karşılaştırma Matrisi ve Roc Eğrisi



Şekil 3-30: Her katman için en iyi sınıflandırma sonuçları

3.5.6. Müzik Duygu Tanıma Uygulaması

Müzik duygu tanıma uygulamasının çalışma görüntüsü Şekil 3-31'de gösterilmektedir. Uygulamada bulunan butonlar ile mp3 müzik dosyasını yükleme, Chroma özellik çıkarımı, AlexNet katman seçimi ve sınıflandırma işlemleri yapılabilmektedir. Ara yüzde yapılan her işlem adımı aşama verilmiştir.



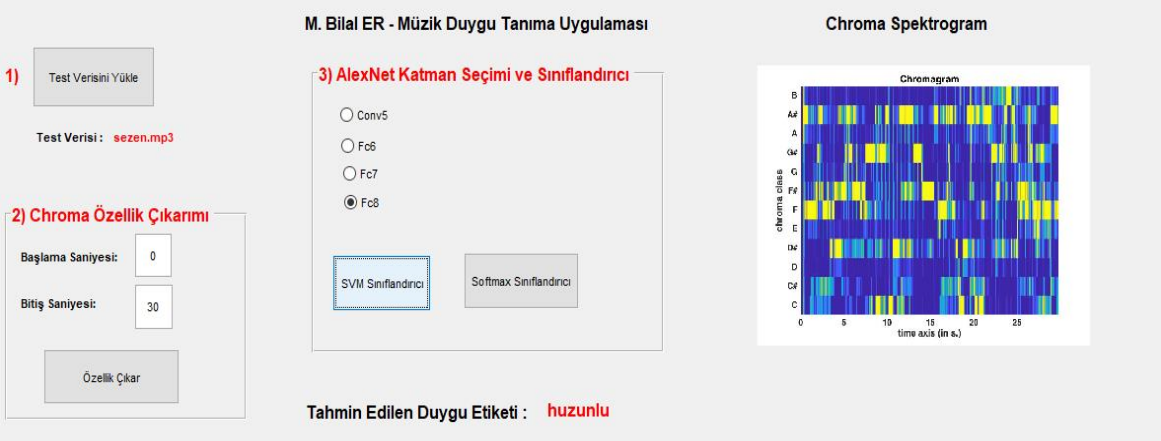
Şekil 3-31: Müzik Duygu Tanıma Sistemi

Uygulamanın çalışması için Alexnet ve Mirtoolbox kütüphanelerine ihtiyaç vardır. “Test Verisini Yükle” butonu kullanılarak dosyadan .mp3 formatındaki müzik verisi alınmaktadır. İkinci adımda yüklenen mp3 dosyasının herhangi bir bölümünden 30 saniyelik kesit alınabilmektedir ve daha sonra “özellik çıkar” butonu ile test verisinin kesilen 30 saniyelik bölümü için Chroma spektrogramı çıkarılmaktadır. Çıkarılan Chroma spektrogramı test klasörüne “jpg” formatında kaydedilmektedir. Chroma çıkarımı için gerekli algoritma Şekil 3-32’de verilmiştir. Ayrıca Şekil 3-33’de örnek test verisinden elde edilen Chorma Spektrogramın uygulma üzerinde gösterimi verilmiştir.

```
filename, foldername] = uigetfile('*.mp3',);
fullname = fullfile(foldername, filename);
startpoint= handles.edit1;
endpoint= handles.edit2;
audio = miraudio(fullname,'Extract', startpoint, endpoint);
k=mirchromagram(audio,'frame')
adress = ['test\'];
 img=[adres filename(1:end-4) '.jpg'];
saveas(gcf,img);
```

Şekil 3-32: Chroma Özellik Çıkarımı İçin Gerekli Algoritma

Üçüncü adım, AlexNet katman seçimi ve sınıflandırma işlemidir. Test verisinden elde edilen Chroma görüntüsünden derin görsel özellikler çıkarmak için AlexNet’in farklı katmanlarından derin özellik çıkarma seçeneği uygulamaya eklenmiştir. Sınıflandırma işlemi uygulamada bulunan SVM ve Softmax Sınıflandırıcı butonları kullanılarak yapılmaktadır. Test verisinin Chroma görüntüsünden çıkarılan derin görsel özellikler ile eğitim veri setindeki müziklerden elde edilen Chroma görüntülerinin derin özellikleri arasında karşılaştırma yapılarak müzik duygu tanıma işlemi uygulama sayesinde kolayca yapılabilmektedir. AlexNet’ten derin özellik çıkarma ve sınıflandırma için gerekli algoritma şekil 3-34’de verilmiştir.



Şekil 3-33: Chroma Çıkarımı ve Duygu Tahmini

```
allImages = imageDatastore('.\chroma', 'IncludeSubfolders', true,...
 'LabelSource', 'foldernames','ReadFcn',@readFunctionTrain);

validateImages = imageDatastore('.\test', 'IncludeSubfolders', true,...
'LabelSource', 'foldernames','ReadFcn',@readFunctionTrain);

Selected_layer=get(handles.uibuttongroup2,'SelectedObject')

layer = Selected_layer;
myNet=alexnet;

trainingFeatures = activations(myNet,allImages,layer);
validateFeatures = activations(myNet,validateImages,layer);

trainingLabels = allImages.Labels;
validateLabels = validateImages.Labels;

classifier = fitcecoc(trainingFeatures,trainingLabels); % Binary SVM
[predictedLabels,score] = predict(classifier,validateFeatures);
emotion=string(predictedLabels(1));
set(handles.text9,'String', emotion);
```

Şekil 3-34: Derin Özellik Çıkarma ve Sınıflandırma İçin Gerekli Algoritma

4. BÖLÜM – EEG İLE MÜZİK DUYGULARININ ANALİZİ

Müzikte geçen duygunun tahmini için hazırlanan veri setlerindeki duygu etiketleri genellikle bir gurup uzmanın kişisel kanaatlerine göre ya da bir gurup gönlü katılımcı üzerinde yapılan deneyler sonucu elde edilen veriler ile hazırlanmaktadır. Bu şekilde hazırlanan veri setleri üzerinde yapılan çalışmalar her ne kadar olumlu sonuçlar veresede müzikten hissedilen duygu özneldir ve kişinin yaşına, kültürüne, mesleğine ve diğer nedenlere bağlı olarak farklılık gösterebilir yani bir şarkıdan herkes aynı duyguyu hissetmeyebilir. Ayrıca bir insanın dinlediği müziğe hüzünlü yada mutlu duygu içeriyor demesi gerçekte hissettiği duyguyu ortaya koymayabilir. Fakat müzik dinleme anında hissedilen duyguya göre beyin içinde meydana gelen elektriksel dalgalanmalar, algılanan gerçek duygunun yapısını daha doğru bir şekilde ortaya koyabilmektedir. Bu nedenlerden dolayı çalışmanın bu bölümünde ses müziği parçalarını dinlerken insan duygularının tanınması sorunu araştırılmıştır. Müzik dinlerken beyninde oluşan elektriksel dalgalar analiz edilerek duygusal (mutlu, hüzünlü, rahatlatıcı, gergin) olarak sınıflandırılmıştır.

İnsan beyni, milyarlarca sinir hücresinin birbirine bağlandığı karmaşık bir sistemdir. Beyin hücreleri elektriksel impulslarla birbirleriyle iletişim kurarlar. EEG(elektroansefalogram), beyin tarafından üretilen elektriksel aktiviteyi kafa derisi yüzeyine yerleştirilen elektrotlar aracılığıyla kaydetmek için kullanılan fizyolojik yöntemdir. 1929'da Hans Berger, ilk insan beyni EEG' sini kaydetmiştir (Vaid, Preeti & Kaur, 2015). EEG, doktorlar ve bilim adamları tarafından beyin fonksiyonlarını incelemek ve nörolojik hastalıkları teşhis etmek için yaygın olarak kullanılmaktadır.

Bir EEG, beyin dalgası paternlerini izler ve kaydeder. EEG testi sırasında, elektrot olarak adlandırılan küçük düz metal diskler teller ile kafa derisine sabitlenir ve daha sonra her elektrot bir EEG kayıt makinesine bağlanır. Elektrotlar beyindeki elektriksel impulsları analiz eder ve sonuçları kaydeden bir bilgisayara sinyaller

gönderir. Kullanımlarına bağlı olarak, paralel olarak veri kaydedebilen çok kanallı EEG kayıt cihazları 256 ya kadar elektrot içerebilir. Her kanal bir EEG kaydı sırasında bir sinyal üretir (Siuly, 2012).

Bir EEG kaydındaki elektriksel dürtüler, tepe noktaları ve çukur noktaları olan dalgalı çizgiler gibi görünür (Farnsworth, 2018). Bu çizgiler, doktorların anormal kalıpların olup olmadığını hızlı bir şekilde değerlendirmesini sağlar. Herhangi bir düzensizlik nöbet veya diğer beyin bozukluklarının bir belirtisi olabilir. İşaret işleme tekniklerini kullanarak normal ve anormal bir kişinin beyin aktivitesi kolayca ayırt edilebilir.

Normal bir EG sinyalinin frekans aralığı 1 Hz-100 Hz'dir, ancak 100 Hz çok nadir görülür. Sinyalin genliği ise genel olarak 10 μV - 100 μV arasındadır. EEG'nin karakteristik özellikleri, bireyin kendisi, yaşı ve öznenin zihinsel durumu gibi birçok faktöre bağlıdır (Subha & diğerleri., 2010).

4.1. EEG Dalga Şekli

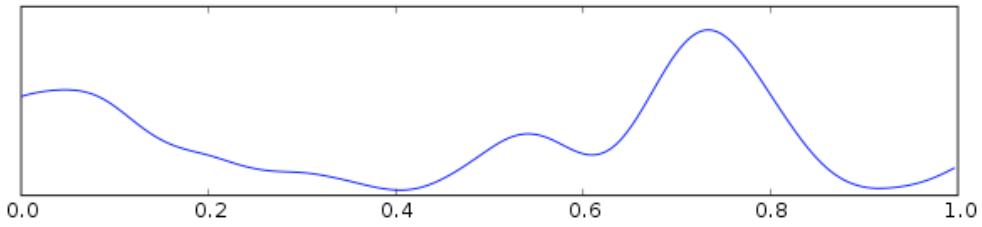
Beyin yüzeyinde meydana gelen elektriksel aktiviteler, EEG makinesinin ekranında, voltaj olarak ölçülen değişen frekans ve genlikteki dalga formları olarak görünür. EEG aktivitesi mikrovolt (mV) cinsinden ölçülen sinyallerdir. Yaşadığımız andaki duruma göre beynimizde farklı şekillerde elektriksel dalgalanmalar olur ve durumlara göre farklı frekans kalıpları oluşur. EEG sinyalleri frekans genlik ve şekilleri bakımından sınıflandırılır.

Beyin dalgası hızı Hertz (saniyedeki devir) cinsinden ölçülür ve yavaş, orta ve hızlı dalgaları tanımlayan frekans bantlara ayrılır. Frekans aralıklarına bağlı olarak, beş tip dalga tanımlanabilir. Sırasıyla düşük frekanstan yüksek frekansa doğru; Alfa (α), teta (θ), beta, (β), delta (δ) ve gamma (γ) şeklindedir. Bu frekans bantları tablo 4-1'deki gibidir.

Tablo 4-1: Dalgaların Frekans Aralıkları

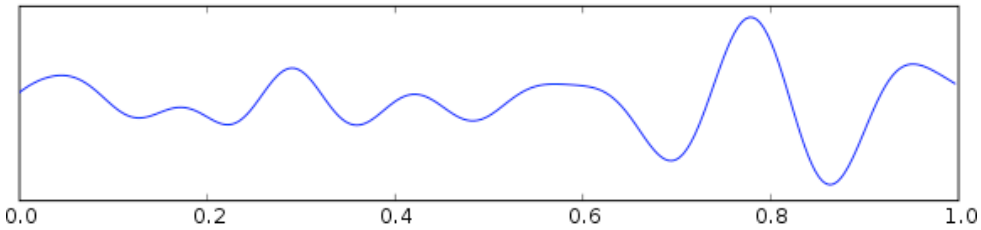
Dalga	Frekans
Delta	0-4 HZ
Teta	4-8 HZ
Alfa	8-13 HZ
Beta	13-30 HZ
Gama	>30 HZ

Delta: Delta dalgalar 0 - 4 Hz frekans aralığındadır. Şekilsel bakımından en yüksek genliğe fakat yavaş dalga formuna sahiptir. Bu dalgalarla ilişkili zihinsel durumlar derin uyku, koma, hipnoz ve bazen de uyanıklık halindedir (Tauseef, 2012). Çok yüksek genlikte ortaya çıkmaları halinde ciddi beyin bozukluklarıyla ilişkilendirilebilir. Bu dalgalar yaşa göre azalır ve normalde sağlıklı insanlarda uyanık halde bulunurlar. Delta dalgası örneği şekil 4-1'de verilmiştir.



Şekil 4-1: Delta Dalgası (Lotte, 2009)

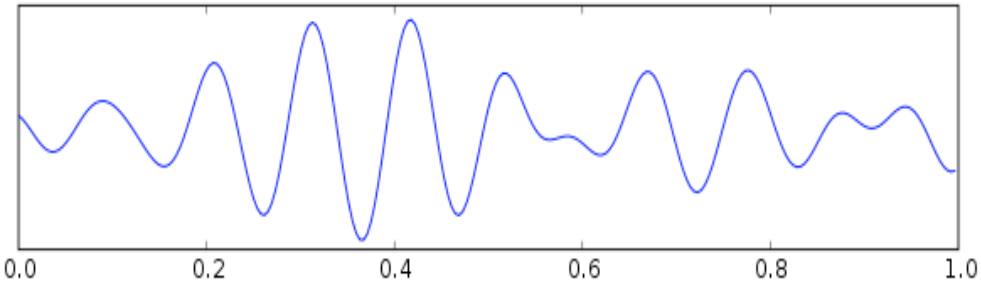
Teta: Bu dalgalar genellikle 20 μ V'den büyük bir genliğe sahip 4 ila 8 Hz frekans bandı aralığındadır. Bu dalgalarla ilişkili zihinsel durumlar, duygusal stres, hayal kırıklığı ve derin meditasyon şeklindedir (Siuly, 2012). Bu dalgalar normalde sağlıklı insanlarda derin uykudayken bulunur. Bu dalgalar beynin temporal ve parietal bölgelerinde görünür. Teta dalgası örneği şekil 4-2'de verilmiştir.



Şekil 4-2: Teta Dalgası (Lotte, 2009)

Alfa: Alfa dalgası aktivitesi, uyanık halde beynin rahat bir durumda olduğu anlamına gelir. Bu dalgalar, gözler kapalıyken veya gevşeme durumundayken, esas olarak başın arka bölgelerinde (oksipital lob) görünen 30-50m μ V genlik ile 8 ila 13 Hz frekans aralığını içerir (Tauseef, 2012). Bu dalgalar normalde sakin, uyanık halde olan insanlarda sık görünür. Gözleri kapatırken ve gevşetirken belirir, gözler açılırken veya herhangi bir mekanizma ile uyarırken kaybolur (düşünme, hesaplama). Uyku halinde, beyin genellikle alfa dalgaları üretmez. Ancak bazı durumlarda, uygun olmayan alfa etkinliği uyku bozukluklarına neden olabilir. Alfa dalgası örneği şekil 4-3'de verilmiştir.

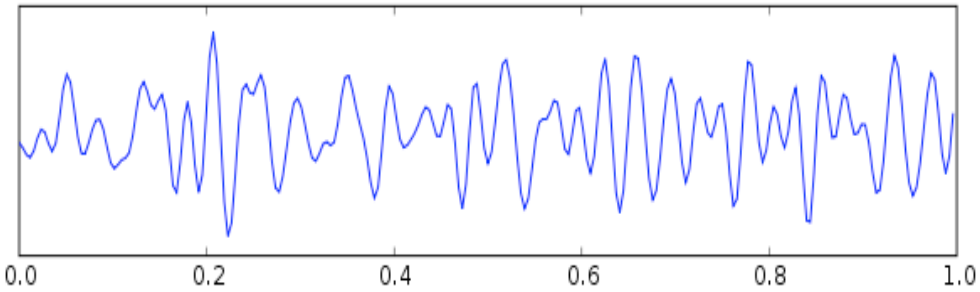
Normal yetişkinlerde görülen ana ritimdir. Özellikle on üçüncü yıldan sonra yaşamın çoğunda bulunur. EEG cihazları ile ölçülen sinyaller genelde alfa dalgası olarak adlandırılır. Artan alfa aktivitesinin yaratıcılığı arttırdığı ve hatta duyumları görmezden gelme yeteneğinizi artırarak depresif semptomları ve kronik ağrıları azaltabildiği bulunmuştur.



Şekil 4-3: Alfa Dalgası (Lotte, 2009)

Beta: 13 Hz-30 Hz frekans aralığındadır. Frontal alanda beynin her iki tarafta simetrik olarak düşük genlikli ve değişken frekanslarda görülür. Diğerlerine göre daha büyük bir frekans bandına sahiptir. Beyin uyandırıldığında ve zihinsel faaliyetlerde aktif olarak yer aldığında, beta dalgaları oluşturur (Lotte, 2009). Bir çok yaş gurubunda görünür. Beta

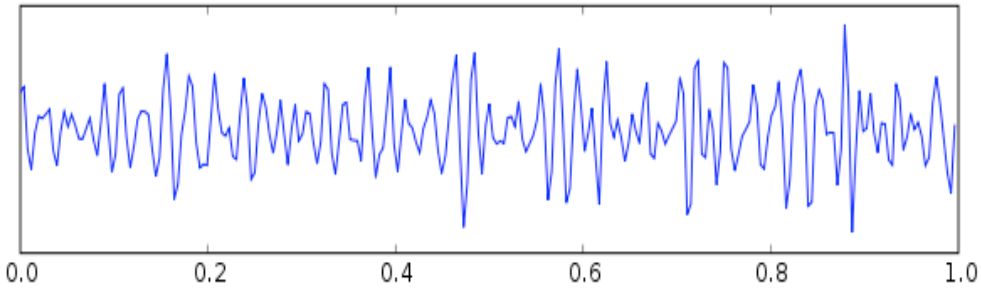
dalgaları, güçlü bir şekilde çalışan bir aklın özellikleridir. Beta, genellikle yoğunlaşma, derin düşünme, dış dünyaya odaklanma, karar verme veya somut problemleri çözme ile ilişkili beyin dalgasıdır. Beta dalgası örneği şekil 4-4'de verilmiştir.



Şekil 4-4: Beta Dalgası (Lotte, 2009)

Gama: Gama dalgaları, sinirbilimi alanında daha yeni bir keşiftir, bu nedenle nasıl işlev gördükleri anlayışı sürekli gelişmektedir. Bugüne kadar, sağlıklı bilişsel fonksiyona ek olarak, Gamma dalgalarının daha karmaşık görevlerin işlenmesinde rol oynadığı bilinmektedir. Bu beyin dalgaları en hızlı beyin dalgalarıdır. Gama dalgası örneği şekil 4-5'de verilmiştir.

Gama dalgaları 30 Hz frekans aralığındadır. Bu dalgaların çok düşük genliklerinden oluşmalarından dolayı fark edilmesi zordur. Bu dalgalar beynin her kısmında ortaya çıkabilirler. Gama dalgaları, öğrenme, yeni bilgileri zihne işleme, duyularımız ve algılarımızla ilişkilidir. Zihinsel engelli ve öğrenme güçlüğü çeken bireylerin ortalamadan daha düşük gama aktivitesine sahip olma eğiliminde oldukları bulunmuştur.



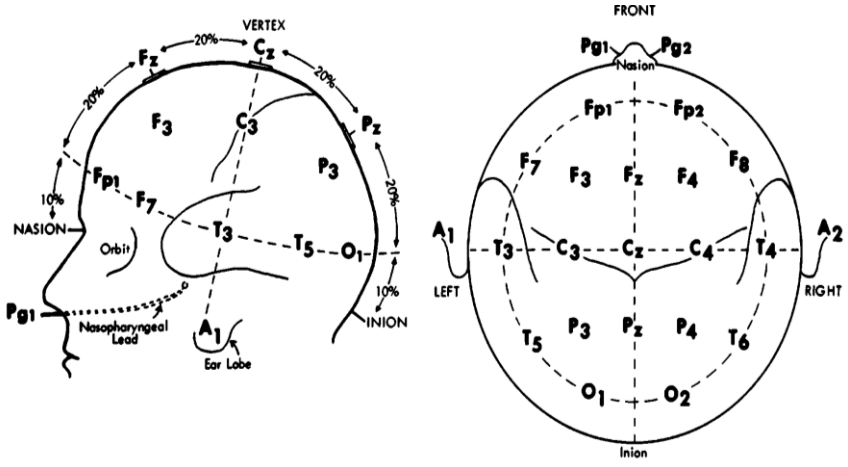
Şekil 4-5: Gama Dalgası (Lotte, 2009)

4.2. EEG Elektrotlarının Bağlanış Sekli

EEG Sinyalleri kafanın neresinden alındığına bağlı olarak iki türe ayrılabilir vardır. Bunlar, kafa derisinden alınan EEG sinyalleri ve kafa içinden alınan EEG sinyalleridir. Kafa derisinden alınan EEG için, kafa derisinin üzerine iyi mekanik ve elektriksel teması olan küçük elektrotlar yerleştirilir. kafa içinden alınan EEG sinyali almak için ameliyat sırasında beyine implant edilen özel elektrotlar kullanılır.

Beynin mimarisi üniform olmaması ve korteksin fonksiyonel yapıya sahip olmasından dolayı kaydedilen EEG sinyalleri elektrotlarının konumuna bağlı olarak değişebilir. EEG sinyalleri kayıt edilmeden önce elektrotların yerleştirilmesine dikkat edilmelidir. Elektrotların kafa derisinin üzerine nasıl yerleştirileceği önemli bir süreçtir, çünkü farklı türdeki aktivitelerin işlenmesinden farklı beyinsel korteks lobları sorumludur. Kafa derisi elektrot lokalizasyonu için yaygın olarak kullanılan elektrot yerleştirme yöntemlerinden biri, Amerikan Elektroensefalografik Topluluğu tarafından standart hale getirilmiş olan 10-20 elektrot sistemidir (Jasper, 1958).

Bu sistemi kullanarak, kafa derisi üzerine toplam 21 elektrot yerleştirilir. Şekil 4-6'da, uluslararası 10-20 sistemine göre beyindeki elektrot pozisyonunu göstermektedir. Her bölge lobu tanımlamak için bir harf ve yarım küre konumunu tanımlamak için bir sayı kullanır. F, T, C, P ve O harfleri sırasıyla Frontal, Temporal, Central, Parietal ve Oksipital anlamına gelir. "Z" orta hatta yerleştirilen bir elektrodu ifade eder. Çift rakamlar sağ yarım küre üzerindeki elektrot pozisyonlarını belirtirken, tek rakamlar sol yarım küre üzerindekileri gösterir.



Şekil 4-6: Uluslararası 10-20 elektrot yerleştirme sistemi

4.3. Dalgacık Dönüşümü (Wavelet Transform)

EEG sinyalleri farklı frekans bantlarına, genliğe sahip periyodik olmayan sinyallerdir. Ham EEG verilerini işlemek, analiz etmek ve öz nitelik çıkarmak için Fourier Analizi, Bilgisayar Destekli Beyin Elektriksel Aktivitesinin Haritalanması ve dalgacık dönüşümü gibi çeşitli sinyal işleme teknikleri kullanılır. Öznitelik çıkarımı için bu çalışmada dalgacık dönüşümü kullanılırmıştır.

Dalgacık dönüşümü sinyalin zaman ve frekans bilgisinin aynı anda gösterimini sağlayan bir dönüşümdür. Spektral bileşenlerin zaman lokalizasyonu gerektiğinde, sinyalin zaman-frekans bileşenlerini temsil eden bir dönüşüm gereklidir. Bu bileşenleri temsil eden dönüşüm de dalgacık dönüşümü ile sağlanır. Bazı durumlarda, belirli spektral bileşenlerin meydana geldiği zaman aralıklarını bilmek çok faydalı olabilir (Adeli, Zhou & Dadmehr, 2003).

Bir sinyalin frekansını hesaplamak için Foruier dönüşümünden faydalanılır. Zaman alanındaki bir sinyalin Foruier dönüşümü alınırsa, o sinyalin frekans-genlik gösterimini elde ederiz ve bir ekseni frekans, diğer ekseni ise genlik olan bir gösterime sahip oluruz. Bu gösterim bize sinyalimizde her frekansın ne kadar olduğunu gösterir fakat hangi zamanda hangi frekans bileşenlerinin olduğu bilgisini vermez.

EEG sinyali sabit olmayan bir sinyaldir (Adeli, Zhou & Dadmehr, 2003). Foruier dönüşümü, sinyalin spektral içeriğini verir, ancak bu spektral bileşenlerin nerede ortaya çıktığı konusunda hiçbir bilgi vermez. Bu nedenle, Foruier dönüşümü istisnalar dışında durağan olmayan sinyal için uygun bir teknik değildir (Vetterl, 2001).

Dalgacık dönüşümü dalgacık adı verilen fonksiyonlar yardımıyla yapılır. Ana dalgacık ismi verilen bir fonksiyon zamanda ileriye doğru hareket eden bir pencere görevi görür. Ana dalgacık fonksiyonu zaman eksini boyunca ötelenir ve ölçeklendirilir. Ötelemek dalgacık fonksiyonun kaydırılmasına, ölçeklemek ise dalgacık fonksiyonun daraltılıp genişletilmesi anlamına gelir (Arı, Özen & Çolak , 2008).

Dalgacık dönüşümü, hem frekans hem de zaman alanında yüksek bir çözünürlüğe sahiptir. İlk önce sinyale büyük bir ölçek/pencere ile bakar ve büyük özellikleri analiz eder, sonra daha küçük özellikleri analiz etmek için sinyale daha küçük ölçeklerle bakar.

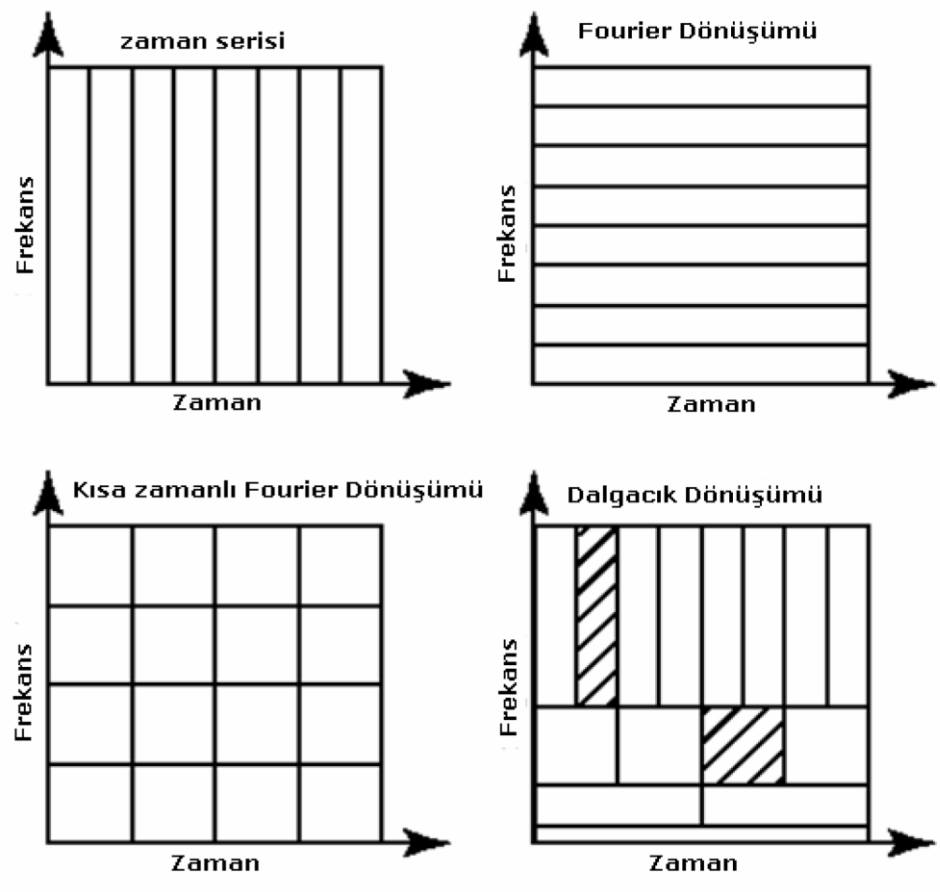
Kısa zamanlı Foruier dönüşümlerinin aksine, ana dalgacık sürekli değişen pencere boyutlarıyla ölçeklenir.

Şekil 4-7'de farklı dönüşümlerin zaman ve frekans çözünürlüklerini görebiliriz. Orijinal zaman serileri, zaman alanında yüksek çözünürlüğe ve frekans alanında sıfır çözünürlüğe sahiptir. Bu, zaman alanındaki çok küçük özellikleri ayırt edebileceğimiz ve frekans alanındaki hiçbir özelliği ayırt edemeyeceğimiz anlamına gelir. Buna zıt olarak, frekans alanında yüksek çözünürlüğe ve zaman alanında sıfır çözünürlüğe sahip olan Fourier Dönüşümleri frekans bilgilerini gösterir fakat frekansların ne zaman oluştuğu göstermezler. Kısa Zamanlı Fourier Dönüşümünde, hem frekans hem de zaman alanında orta ölçekli çözünürlüğe sahibizdir. Pencere zaman ekseni boyunca bütün frekanslar için sabit boyuttadır.

Dalgacık Dönüşümü de ise :

- Küçük frekans değerleri için, frekans alanında yüksek çözünürlük, zaman alanında düşük çözünürlük sağlanır.
- Büyük frekans değerleri için, frekans alanında düşük bir çözünürlük, zaman alanında yüksek çözünürlük sağlanır.

Dalgacık Dönüşümleri bir takas yapar; zamana bağlı özelliklerin olduğu ölçeklerde, zaman alanında yüksek bir çözünürlüğe ve frekansa bağlı özelliklerin olduğu ölçeklerde, frekans alanında yüksek bir çözünürlüğe sahiptir.



Şekil 4-7: Orijinal Zaman Serisi Veri Setine Kıyasla Farklı Dönüşümlerin Zaman ve Frekans Çözünürlükleri Bakımından Karşılaştırılması

4.3.1. Sürekli Dalgacık Dönüşümü (SDD)

Sürekli dalgacık dönüşümü ve ayrık dalgacık dönüşümü olmak üzere İki tür dalgacık analizi vardır. Sürekli dalgacık dönüşümü ile analiz edilecek sinyal, sürekli zaman ve frekans artışlarında dalgacık temeli fonksiyon ile eşleşir. Sürekli dalgacık dönüşümünde veriler dijitalleştirilmelidir. Sürekli zaman ve frekans artışları, sayısallaştırılmış her noktadaki veya artımdaki verilerin kullanıldığını gösterir. Sonuç olarak, orijinal sinyal sürekli olarak dalgacık fonksiyonunun ağırlıklı bir integrali olarak ifade edilir.

Matematiksel olarak bir Sürekli dalgacık dönüşümü formülü denklem 24'de verilmiştir.

$$X_{WT}(\tau, s) = \frac{1}{\sqrt{|s|}} \int_{-\infty}^{\infty} x(t) C_{\psi}{}^{*} \left(\frac{t-\tau}{s}\right) dt \quad (24)$$

Dönüştürülmüş sinyal $X_{WT}(\tau, s)$, τ kaydırma parametresinin ve s ölçek parametresinin bir fonksiyonudur. Ana dalgacık ψ ile gösterilir, $*$ ise kompleks eşleniği temsil eder.

Dalgacık katsayıları $1/\sqrt{|s|}$ 'ye bölünerek sinyal enerjisi her ölçekte normalleştirilir. Bu işlem**,** dalgacıkların her ölçekte aynı enerjiye sahip olmasını sağlar (Addison, 2017).

Ana dalgacık, ölçek parametreleri değiştirilerek daraltılır. "s" ölçeğindeki değişiklik sadece dalgacıktaki merkezi frekansı f_c'yi değil, aynı zamanda pencere uzunluğunu da değiştirir. Bu nedenle s ölçeği, dalgacık analizinin sonuçlarını temsil etmek için frekans yerine kullanılır. τ kaydırma parametresi zaman içinde dalgacık konumunu belirler, τ değiştirilerek dalgacık sinyal üzerinden kaydırılabilir. $X_{WT}(\tau, s)'$deki elemanlara dalgacık katsayıları denir, her dalgacık katsayısı bir ölçek (frekans) ve zaman alanındaki bir nokta ile ilişkilendirilir.

Dalgacık dönüşümü, Fourier dönüşümlerinde olduğu gibi ters bir dönüşüme de sahiptir. Ters sürekli dalgacık dönüşümü denklem 25'deki gibi tanımlanır.

$$x(t) = \frac{1}{C_{\psi}^{2}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} X_{\mathrm{WT}}(T, S) \frac{1}{S^{2}} \psi \left(\frac{t-T}{s}\right) d_T d_s \quad (25)$$

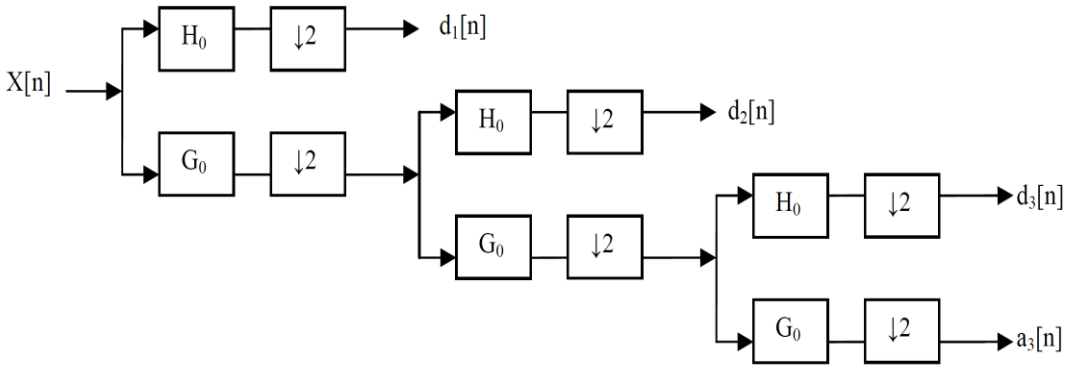
C_{ψ} dalgacık koşulunun kabul edilebilirlik sabitidir.

Sürekli dalgacık dönüşümündeki ölçekleme ve kaydırma parametrelerinin sürekli değişmesi nedeniyle sürekli dalgacık dönüşümünün pratik uygulamalarda kullanımı çok fazla değildir (Übeyli & Güler, 2004).

4.3.2. Ayrık Dalgacık Dönüşümü (ADD)

Ayrık dalgacık dönüşümünde, dalgacık fonksiyonlarının gerilip ve sıkıştırılmasıyla çoklu çözünürlük analizi gerçekleştirir. Zaman-frekans düzleminde çoklu-çözünürlük sağlamak için filtre bankaları kullanır. Dijital filtreleme teknikleri kullanılarak dijital sinyalin zaman ölçekli bir gösterimi elde edilir. Analiz edilecek sinyal, farklı ölçeklerde farklı kesim frekanslarına sahip filtrelerden geçirilir.

Sinyaldeki ayrıntılı bilgi miktarının bir ölçüsü olan sinyalin çözünürlüğü, filtreleme işlemleri tarafından belirlenirken, ölçek ise yukarı örnekleme ve alt örnekleme işlemleriyle belirlenir. Ayrık dalgacık dönüşümü, Şekil 4-8' de gösterildiği gibi ayrık sinyalinin ardışık alçak bant geçiren filtre ve yüksek bant geçiren filtrelerden geçirilmesiyle hesaplanır. Sinyal x[n] dizisi ile gösterilir, burada n bir tam sayıdır. Düşük geçişli filtre G_0 ile gösterilirken, yüksek geçişli filtresi H_0 ile gösterilir. Her seviyede, yüksek geçişli filtre sinyal ile ilgili ayrıntılı kat sayıları verir ve d[n] ile gösterilir. Ölçekleme fonksiyonuyla ilişkili düşük geçişli filtresi ise yaklaşık katsayılar üretir ve a[n] ile gösterilir.



Şekil 4-8: Üç seviyeli dalgacık ayrıştırma ağacı.

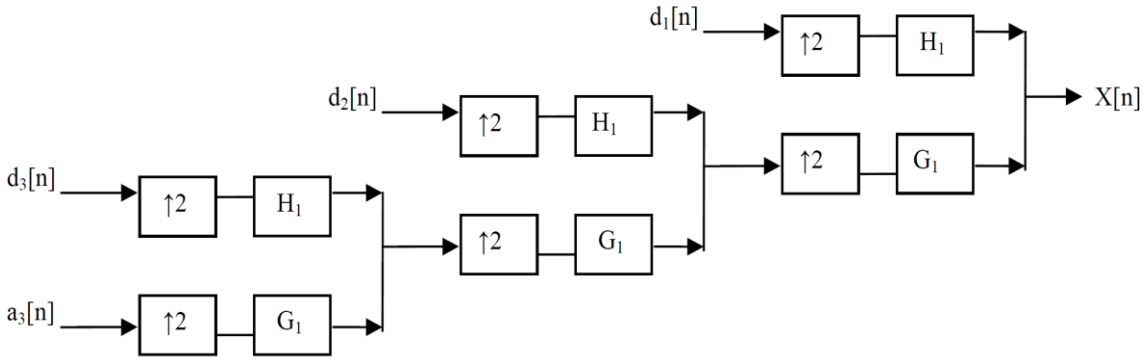
Ayrık dalgacık dönüşümünü bir sinyale uygulamak için en küçük ölçekle başlarız. Küçük ölçekler yüksek frekanslara karşılık gelir ve buda öncelikle yüksek frekansları analiz ettiğimiz anlamına gelir. İkinci aşamada, ölçek iki kat artar (frekans iki kat azalır) ve maksimum frekansın yarısı kadarı analiz edilir. Üçüncü aşamada, ölçek faktörü dörttür ve maksimum frekansın dörtte biri civarında frekans davranışını analiz edilir. Bu işlememler maksimum ayrışma seviyesine ulaşana kadar devam eder.

Düşük frekans değerlerinde, Nyquist oranını karşılamak için daha az örnek sayısına ihtiyaç vardır ve daha yüksek sayıda örneği sinyalde tutmaya gerek yoktur. Alt örneklemeden dolayı, işlemin bir aşamasında sinyalimizdeki örnek sayısı dalgacık filtresinin uzunluğundan daha küçük olacak ve maksimum ayrışma seviyesine ulaşmış olunacaktır.

Örneğin 1000 Hz'e kadar frekanslara sahip bir sinyal ilk aşamada, düşük frekanslı bir parçaya ve yüksek frekanslı bir parçaya, yani 0-500 Hz ve 500-1000 Hz'e ayrılır. İkinci aşamada, düşük frekanslı parça alınır ve tekrar iki parçaya bölünür: 0-250 Hz ve 250-500 Hz. Üçüncü aşamada 0-250 Hz parçasını 0-125 Hz parçasına ve 125-250 Hz parçasına ayrılır. Bu, ihtiyacımız olan inceltme seviyesine ulaşana veya örnekler bitene kadar devam eder.

Orijinal sinyali tekrar elde etmek için, son ayrışma seviyesinden başlayarak, tüm [a] ve d[n] katsayılarının bir araya getirilir.

Şekil 4-9'da dalgacık katsayılarından orijinal sinyalin yeniden yapılandırılmasını göstermektedir. Temel olarak, yeniden yapılanma, ayrıştırmanın ters işlemidir. Her seviyedeki yaklaşık ve detay katsayıları iki katına çıkarılır, düşük geçişli ve yüksek geçişli sentez filtrelerinden geçirilir ve sonra eklenir. Bu işlem, orijinal sinyali elde etmek için ayrıştırma işlemindeki gibi aynı seviyelerde devam eder.



Şekil 4-9: Yeniden yapılandırılmış üç seviyeli dalgacık ağacı.

4.4. Dalgacık Ailelerinin Farklı Türleri

Fourier Dönüşümü ve Dalgacık Dönüşümü arasındaki diğer bir fark, dalgacıkların birçok farklı familyasının (türünün) olmasıdır. Dalgacık aileleri birbirlerinden farklıdır. Bu durumda sinyalimizde aradığımız özelliklere en iyi şekilde uyan belirli bir dalgacık ailesini seçebileceğimiz seçebiliriz.

Dalgacık analizi, söz konusu sinyal ile bir dalgacık fonksiyonu ψ(t) arasındaki korelasyonu hesaplar. Sinyal ve dalga fonksiyonu arasındaki benzerlik, farklı zaman aralıkları için ayrı olarak hesaplanır.

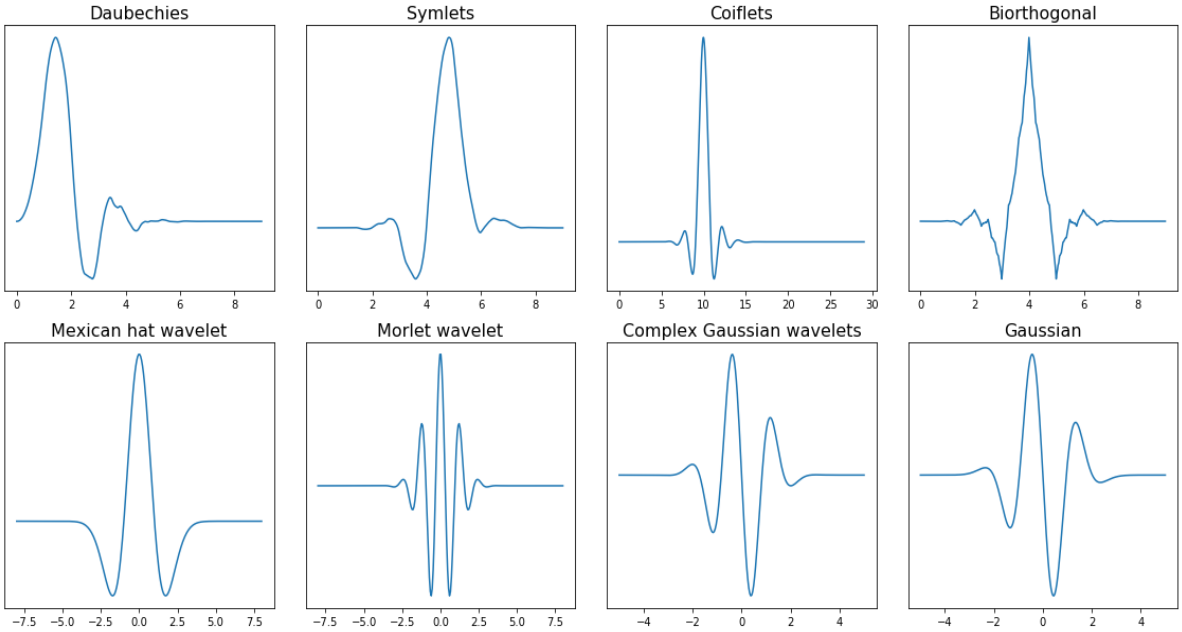
Bir analiz fonksiyonu ψ(t), aşağıdaki matematiksel kriteri karşılarsa dalgacık olarak sınıflandırılır:

Bir dalgacık sonlu enerjiye sahip olmalıdır ve gerekli formül denklem 26'da verilmiştir.

$$E = \int_{-\infty}^{\infty} |\psi(t)|^2 \, dt < \infty \tag{26}$$

Enerji E, dalgacık fonksiyonunun kare büyüklüğünün integraline eşittir ve sonsuzdan daha az olmalıdır.

Her dalgacık tipi farklı bir şekle, yapıya sahiptir ve farklı bir amaç için kullanışlıdır. Şekil 4-10'da dalgacık ailesinde yer alan bazı dalgacık fonksiyonları verilmiştir.



Şekil 4-10: Örnek dalgacık şekilleri

4.5. Uygulama-Araştırma Metodolojisi

Ses müziği parçalarını dinlerken insan duygularının tanınması sistemi dört adımdan oluşmaktadır. Bu adımlar şekil 4-11'de verildiği gibi, veri toplama, sinyal önişleme, özellik çıkarma, ve sınıflandırma şeklidedir. Müzik parçalarının dinlenmesi sırasında beyin üzerinde oluşan elektriksel aktiviteler EEG kulaklık kullanarak insan kafa derisinden kaydedilir. Kaydedilen EEG sinyalleri daha sonra dış ortamdan gelen gürültüleri gidermek ve ham EEG verisini işlenebilir hale getirmek için ön işlemeden geçirilir. Ön işleme tabi tutulduktan sonra dalgacık dönüşümü kullanılarak sinyallerden öznitelik çıkarılır. Son olarak, çıkarılan öznitelikler sınıflandırıcıları eğitmek ve insan duygularını tanımak için kullanılır. Her adımın detayı aşağıdaki alt bölümlerde açıklanmıştır.



Şekil 4-11: Duygu Tanıma Adımları

4.5.1. Uyaran ve Katılımcılar

Otuz ses müzik parçası, EEG tabanlı duygu tanıma deneyi için harici bir uyarıcı olarak kullanılmıştır. Tüm müzik parçaları MP3 formatındadır ve örnekleme hızı 320 kbps'dir. Bu çalışmada kullanılan müzikler, Türk Sanat Müziği, Türk Halk Müziği, Türk Pop ve Türk Caz müziği olmak üzere dört türden oluşuyordu. Seçilen türler fark edilebilir duygular üretmektedir.

Farklı kültürlere sahip toplam dokuz katılımcı gönüllü olarak çalışmaya katılmıştır. Dokuz sağlıklı denekten hepsi erkektir ve katılımcılar 20 ila 25 yaş arasındadır. Tüm katılımcıların herhangi bir beyin hastalığı rahatsızlığı yoktur ve normal işitme gücüne sahiptir. Deneyden önce, katılımcılara çalışmanın kapsamı ve prosedürü hakkında bilgi verilmiştir.

4.5.2. EEG Kayıt

Bu çalışmada EEG sinyallerini alabilmek için "EMOTIV EPOC" kullanılmıştır. EMOTIV EPOC insan beyni araştırması için tasarlanmıştır ve hızlı ve kullanımı kolay bir tasarımla profesyonel yüksek kalitede beyin verilerine erişim sağlar. Bu cihazda toplam 14 kanal vardır ve 5 dakikaya kadar kayıt alabilmektedir. Cihaz 128 örnekleme frekansıyla veri alabilmektedir. Bluetooth bağlantısı ile Matlab üzerinde hazırlanan yazılım kullanılarak ham EEG verileri kaydedilmiştir. Tüm EEG kayıtları izole ve gürültüsüz bir odada yapılmıştır. Şekil 4-12'de "EMOTIV EPOC" cihazının fotoğrafı yer almaktadır.

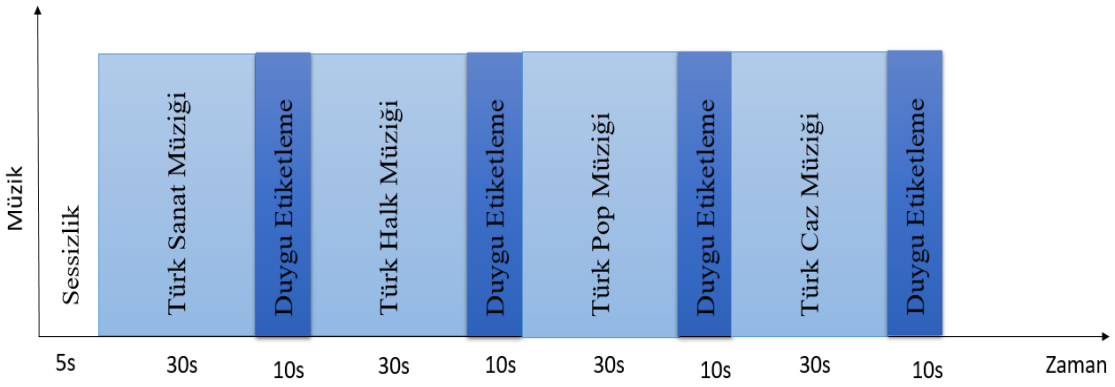


Şekil 4-12: EMOTIV EPOC

Cihazdaki elektrodlar uluslararası 10-20 sistemine göre beyinin çeşitli yerlerine konumlandırılarak EEG kayıtları alınmıştır. Mutlu, hüzünlü, rahatlatıcı ve gergin gibi 4 farklı duyguyu tanımlamak için 9 katılımcının katılım sağladığı bir deney hazırlanmıştır.

Her katılımcıya 30 saniyelik bölümlerden oluşan toplam 16 adet müzik dinletilmiştir. Her katılımcıya her türden birer birer dört şarkı dinletilmiştir ve her şarkıdan sonra 10 saniye sessizlik sağlanmıştır. Bu müziklere katılımcıya dinletilirken dinleyicinin EEG kayıtları alınmıştır ve her şarkı sonunda hissettiği duygu sorularak alınan EEG kaydı hissedilen duyguya göre etiketlenmiştir. Duygu tanıma ve sınıflandırıcının eğitimi işlemi için manuel olarak etiketlenen duygu etiketleri kullanılmıştır. Şekil 4-13'de, katılımcılar için EEG veri toplama aşamasındaki genel

yapı ifade edilmiştir. Toplamda 144 kayıt alınmıştır. Fakat kayıt sırasında bazı verilerin etkin olarak alınamadığından dolayı kaydedilen 15 veri değerlendirme dışına alınarak toplamda 129 EEG kaydı içeren veri seti üzerinden duygu tanıma işlemi yapılmıştır.



Şekil 4-13: EEG Kayıt Alma Süreci

4.5.3. EEG Sinyal Ön İşleme

İnsan kafa derisinden kaydedilen EEG sinyalleri beyin dalgalarının, seslerin ve farklı dış gürültülerin birleşiminden oluşur. Özellik çıkarımı ve sınıflandırma için kaliteli, gürültüsüz sinyallere sahip olmak gerekir. Gürültü azaltmak için, 1.0 Hz - 50 Hz arasında band geçişine sahip bir filtre kullanılarak sinyal analiz için uygun hale getirilmiştir.

4.5.4. Öznitelik Çıkarımı ve Deneysel Sonuçlar

Yukarıdaki bölümde bahsedildiği gibi EEG sinyalleri farklı frekans bantlarına, genliğe sahip periyodik olmayan sinyallerdir. Ham EEG verilerini işlemek, analiz etmek ve öz nitelik çıkarmak için Fourier Analizi, Bilgisayar Destekli Beyin Elektriksel Aktivitesinin Haritalanması ve dalgacık dönüşümü gibi çeşitli sinyal işleme teknikleri kullanılır. Öznitelik çıkarımı için bu çalışmada Dalgacık dönüşümü kullanılırmıştır.

Sinyal toplamda 14 kanaldan alınmıştır fakat 4. 8. 13. ve 17 kanaldaki sinyaller değerlendirmeye alınmıştır. Çünkü beynin işitsel uyarıları algıladığı bölge tempolar bölgedir. Temporal bölge kafatasının yan taraflarında yer aldığı için sağ bölgeden iki kanal ve sol bölgeden iki kanal seçilerek toplamda 4 kanal üzerinden öznitelik çıkarılmıştır. Daha sonra her kanal için 4 seviye dalgacık dönüşümü uygulanarak dalgacık

katsayıları elde edilmiştir ve bu dalgacık kat sayıları öznitelik olarak kullanılmıştır. Ön işlemden geçirilen EEG sinyaline ayrık dalgacık dönüşümü uygulandıktan sonra Delta (1-4 Hz), Teta (4-8 Hz), Alfa (8-13 Hz) ve Beta (13-30 Hz) alt bantlarına indirgenmiştir. Tablo 4-2'de EEG sinyallerinin ayrıldığı band aralıkları verilmiştir. Bu dalgacık katsayılarının hesaplanmasında 2. dereceden Daubechies dalgacık (db2) kullanılmıştır.

Her 30 saniyelik EEG sinyalinde toplam 3840 adet veri vardır. Bu veri 256'lık kesitlere ayrılarak her 256'lık kesit üzerinden D4, D3,D2, D1 ve A4 yaklaşık dalgacık katsayıları (18+18+34+66+129 olmak üzere toplam 265 adet dalgacık) elde edilmiştir. Çıkarılan dalgacık kat sayıları üzerinden istatistiksel hesaplamalar yapılmıştır.

Her katsayının;

- Standard Sapması
- Ortalaması
- Kare Ortalamalarının Karekökü hesaplanmıştır.

Her bir EEG verisi 256'lık kesitlere ayrılarak D4, D3,D2, D1 ve A4 her alt bant için Standard sapma, ortalama ve kareler toplamının karekökü hesaplanmıştır. Her bir kesit için 15 tane öznitelik elde edilmiştir. 30 saniyelik bir sinyalde 15 tane kesit olduğu için her kesitten elde edilen 15 tane öz nitelik birleştirerek bir kanaldan 225 adet öznitelik çıkarılmıştır. Ayrıca yukarda belirtildiği gibi her bir kayıt 4 kanallı EEG verisinden oluşmaktadır. Son olarak her 4 kanaldaki öz nitelik vektörleri birleştirilerek sonuç olarak bir EEG kaydından 900 adet öz nitelik çıkarılmıştır.

Tablo 4-2: Alt bandlar ve Frekans Aralığı

Alt Bandar	Frekans Aralığı (Hz)
D1	32-64
D2	16-32
D3	8-16
D4	4-8
A4	0-4

Veri Seti'nin KNN(k sayısı =7) ile sınıflandırılması sonucunda %53,4, SVM ile sınıflandırılması sonucunda ise %66.67 oranında başarı elde edilmiştir. SVM ve KNN' den elde sınıflandırma sonuçlarına ait karışıklık matrisleri sırasıyla Şekil 4-14 ve 4-15'de verilmiştir.

Karışıklı Matrisi		TAHMİN				
		MUTLU	HÜZÜNLÜ	GERGİN	RAHATLATICI	TOPLAM
GERÇEK	MUTLU	**11**	12	6	5	**34**
	HÜZÜNLÜ	1	**24**	2	4	**31**
	GERGİN	0	6	**20**	4	**30**
	RAHATLATICI	7	7	6	**14**	**34**
	TOPLAM	**19**	**49**	**34**	**27**	

Şekil 4-14: KNN için Karışıklık Matrisi

Karışıklık Matrisi		TAHMİN				
		MUTLU	HÜZÜNLÜ	GERGİN	RAHATLATICI	TOPLAM
GERÇEK	MUTLU	**25**	1	2	6	**34**
	HÜZÜNLÜ	2	**24**	1	4	**31**
	GERGİN	4	2	**19**	5	**30**
	RAHATLATICI	6	6	4	**18**	**34**
	TOPLAM	**37**	**33**	**26**	**33**	

Şekil 4-15: SVM için karışıklık Matrisi

5. BÖLÜM - SONUÇ

Bu çalışmada, derin öğrenme modellerinin uygulaması olan transfer öğrenme kullanılarak müzik duygu tanıma için yeni bir yöntem sunulmuştur. Müzikal verilerden elde edilen veri artırımlı Chroma spectrogramlar kullanılarak önceden eğitilmiş ağa transfer öğrenme uygulanmıştır ve Chroma spektrogramlarından çıkarılan derin görsel öznitelikler kullanılarak müzik duygu tanımanın sınıflandırma başarı oranı arttırılmıştır. Bu amaçla önceden eğitimli ağ modeli AlexNet mimarisi kullanılmıştır.

Bu çalışma için hazırlanan veri seti ile AlexNet'in Conv5, Fc6, Fc7 ve Fc8 katmanlarından derin görsel öznitelikler çıkarılmıştır ve transfer öğrenme süreçleri yapılmıştır. Bu derin görsel öznitelikler SVM ve Softmax sınıflandırıcıları eğitmek ve test etmek için kullanılmıştır. Veri artırma öncesi en iyi sınıflarıma başarısı Softmax sınıflandırıcı ile "Fc6" katmanından elde edilmiştir ve bu başarım % 74.2' dir. Veri artırma sonrası ise en iyi sınıflarıma başarısı SVM sınıflandırıcı ile Fc7 katmanından elde edilmiştir ve bu başarım % 92.7' dir. Beklendiği gibi veri artırma işleminin gerçekleştirilmesinden sonra doğruluk oranı artmıştır. Önerilen yöntemin performansını ölçmek için elde edilen sonuçlar makine öğrenmesi tekniklerinden elde edilen sonuçlarla karşılaştırılmıştır. Deney sonuçlara göre, önerilen yöntem makine öğrenmesi tekniğiyle yapılan müzik duygu tanıma çalışmalarından daha iyi performans göstermiştir.

Elde edilen sonuçlara göre transfer öğrenmenin müzik duygu tanıma problemi için kullanılabileceği ayıca orijinal eğitim verileri sınırlı olsa bile transfer öğrenmenin çok etkili bir yöntem olduğu ortaya konulmuştur.

Bu çalışmanın diğer bölümünde müzik dinlerken beyninde oluşan elektrik dalgaların analizi amaçlanmıştır. Bu amaç doğrultusunda sesiz ve karanlık ortamda 9 kişinin katılım sağladığı bir deney yapılmıştır. Farklı müzik türlerinden bazı şarkılar seçilerek deneye katılan dinleyicilere dinletilmiştir ve dinleme sırasında bu kişilerin EEG sinyalleri alınmıştır. Dinletilen her şarkının sonunda katılımcının hissettiği duyguya göre

şarkıyı gergin, hüzünlü, mutlu veya rahatlatıcı duygu etiketlerinden bir tanesiyle etiketlemesi istenilmiştir. Elde edilen bu EEG sinyalleri filtreden geçirilip dalgacık dönüşümü uygulanarak öznitelik çıkarılmıştır. Son olarak K-NN ve SVM algoritmaları kullanılarak EEG sinyalleri sınıflandırılmaya çalışılmıştır. En yüksek sınıflandırma başarısı ise %66.67 olarak SVM ile elde edilmiştir

İleriki çalışmalarda daha önceden eğitilmiş farklı derin ağ modellerinin mimarileri kullanılarak, mevcut veri setleri ile bu ağların yeniden eğitilmesi ve müzik duygu tanıma işlemlerinin yapılması önerilmektedir. Aynı zamanda derin öğrenme ve transfer öğrenme kullanılarak EEG ile müzik duygu tanıma işleminin yapılması önerilmektedir.

KAYNAKÇA

Addison, Paul S. The Illustrated Wavelet Transform Handbook. 2. CRC Press , 2017

Adeli, Hojjat, Ziqin Zhou & Nahid Dadmehr. "Analysis of EEG records in an epileptic patient using wavelet transform," Journal of Neuroscience Methods. 123, 1: 69-87, 2003.

Adeli, Hojjat, Ziqin Zhou & Nahid Dadmehr. "Analysis of EEG records in an epileptic patient using wavelet transform," Journal of Neuroscience Methods. 123, 1: 69-87, 2003.

Arı, Niyazi, Şükrü Özen & Ömer H. Çolak. Dalgacık Teorisi. Ankara: Palme Yayıncılık, 2008

ATAL, BISHNU S. "Automatic recognition of speakers from their voices," Proceedings of the IEEE. 64, 4: 460-475, 1976.

Barthet, Mathieu, György Fazekas & Mark Sandler. "ultidisciplinary Perspectives on MusicEmotion Recognition: Implications for Contentand Context-Based Models," *9th International Symposium on Computer Music Modelling and Retrieval (CMMR 2012)* : 492-507. London: 2012.

Camastra, Francesco & Alessandro Vinciarelli. Machine Learning for Audio,Image and Video Analysis. Springer, 2007

Chauhan, Paresh M. & Nikita P. Desai. ""Mel Frequency Cepstral Coefficients (MFCC) based speaker identification in noisy environment using wiener filter," *2014 International Conference on Green Computing Communication and Electrical Engineering (ICGCCEE)*: 1-5. Coimbatore: 2014.

Cirean, D.C., U. Meier, J. Masci, L.M. Gambardella & J. Schmidhuber. "Flexible, high performance convolutional neural networks for image classification," *Twenty-Second International Joint Conference on Artificial Intelligence (IJCAI'11)*: 1237–1242. Barcelona: 2011.

Davies, Matthew E. P. & Mark D. Plumbley. "Context-Dependent Beat Tracking of Musical Audio," IEEE Transactions on Audio, Speech, and Language Processing. 15, 3: 1009-1020, 2007.

Delbouys, Rémi, Romain Hennequin, Francesco Piccoli, Jimena Royo-Letelier & Manuel Moussallam. "Music Mood Detection Based On Audio And Lyrics With Deep Neural Net," *ISMIR 2018*: 2018.

Doval, Boris & Xavier Rodet. "Estimation of fundamental frequency of musical sound signals," *International Conference on Acoustics, Speech, and Signal Processing*: 3657-3660. Toronto: 1991.

Farnsworth, Bryn. "What is EEG (Electroencephalography) and How Does it Work?," imotions. *https://imotions.com/blog/what-is-eeg/*. 2018. (8 Ocak 2019).

Feng, Yazhong, Yueting Zhuang & Yunhe Pan. "Popular music retrieval by detecting mood," *Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval*: 376-376. Toronto: 2003.

Fu, Y. & C. Aldrich. "Flotation froth image recognition with convolutional neural networks," Minerals Engineering. 132: 183-190, 2009.

Gupta, Shikha, Jafreezal Jaafar, Wan Fatimah , wan Ahmad & Arpit Bansal. "FEATURE EXTRACTION USINGMFCC," Signal & Image Processing : An International Journal (SIPIJ). 4, 4: 101-108, 2013.

Hampiholi, Vallabha Vasant. "A method for Music Classification based on Perceived Mood Detection for Indian Bollywood Music," *World Academy of Science, Engineering and Technology 72*: 507-514. 2012.

Han, Byeong-jun, Rho Seungmin , Roger B. Dannenberg & Eenjun Hwang. "SMERS: Music Emotion Recognition Using Support Vector Regression.," *10th International Society for Music Information Retrieval Conference*: 651-656. 2009.

Hearinghaven. "How Do We Hear?," *http://www.hearinghaven.com/how-do-we-hear*. [t.y.]. (08 Şubat 2018).

Hevner, K. "Expression in music: a discussion of experimental studies and theories," Psychological Review. ss. 186-204, 1935.

Hevner, Kate. "Experimental Studies of the Elements of Expression in Music," The American Journal of Psychology. 48, 2: 246-268, 1936.

Huq, Arefin, Juan Pablo Bello & Robert Rowe. "Automated Music Emotion Recognition: A Systematic Evaluation," Journal of New Music Research. 39, 3: 227-244, 2010.

Hwang, Fang-Chen, JeenShing Wang, Pau-Choo Chung & Ching-Fang Yang. "Detecting emotional expression of music with feature selection approach," *2013 1st International Conference on Orange Technologies (ICOT)*: 282-286. Tainan: 2013.

Jakkula, Vikramaditya. *"Tutorial on Support Vector Machine (SVM)"*. School of EECS,Washington State University, [t.y.].

Jasper, H.H. "The Ten-Twenty Electrode System of the International Federation," Electroencephalography and Clinical Neurophysiology. 10: 371-375, 1958.

Juslin, Patrik N. "ue Utilization in Communication of Emotion in Music Performance: Relating Performance to Perception ," Journal of Experimental Psychology: Copyright 2000 by the American Psychological Association, Inc. Human Perception and Performance. 26, 6: 1797-1813, 2000.

Juslin, Patrik N. & John Sloboda Sloboda. Handbook of Music and Emotion: Theory, Research, Applications. New York : Oxford University Press, 2010

Korhonen, Mark D., David A. Clausi & M. Ed Jernigan. "Modeling emotional content of music using system identification," IEEE Transactions on Systems, Man, and Cybernetics. 36, 3: 588-59, 2005.

Kotsiantis, Sotiris B. "Supervised machine learning: A review of classification techniques," *Proceedings of the 2007 conference on Emerging Artificial Intelligence Applications in Computer Engineering: Real Word AI Systems with Applications in eHealth, HCI, Information Retrieval and Pervasive Technologies* : 3-24. Amsterdam: 2007.

Kotsiantis, S. & P. E Pintelas. "Recent Advances in Clustering: A Brief Survey," WSEAS Transactions on Information Science and Applications. 1, 1: 73-81, 2004.

Krizhevsky, Alex, Ilya Sutskever & Geoffrey E. Hinton. "ImageNet Classification with Deep Convolutional Neural Networks," *NIPS*: 1097-1105. 2012.

Krumhansl, Carol L. "Music: A Link Between Cognitionand Emotion," Current Directions in Psychological Science. 11, 2: 45-50, 2002.

Laar, Bram van de. "Emotion detection in music, a survey," *4th Twente Student Conference on IT*: Enschede : 2006.

Lartillot, Olivier. *"MIRtoolbox 1.7.1 User's Manual"*. Jyväskylä, Finland.: 2018.

Lidy, Thomas & Andreas Rauber. "Computing statistical spectrum descriptors for audio music similarity and retrieval," *MIREX 2006 - Music Information Retrieval Evaluation*: 2006.

Lin, Chingshun, Mingyu Liu, Weiwei Hsiung & Jhihsiang Jhang. "MUSIC EMOTION RECOGNITION BASED ON TWO-LEVEL SUPPORTVECTOR CLASSIFICATION," *2016 International Conference on Machine Learning and Cybernetics (ICMLC)*: 375-386. Jeju: 2016.

Li, Tao & Mitsunori Ogihara. "Content-based music similarity search and emotion detection," *2004 IEEE International Conference on Acoustics, Speech, and Signal Processing*: Montreal: 2004.

Li, Tao & Mitsunori Ogihara. "Content-based music similarity search and emotion detection," *2004 IEEE International Conference on Acoustics, Speech, and Signal Processing*: Montreal: 2006.

Liua, Tong, Li Hanb, Liangkai Mac & Dongwei Guo. "Audio-based deep music emotion recognition," *AIP Conference Proceedings*: 2018.

Liu, Xin, Chen Qingcai, Xiangping Wu, Yan Liu & Yang Liu. *CNN based music emotion classification*. arXiv preprint arXiv, 2017.

Lu, Lie, Dan Liu & Hong-Jiang Zhang. "Automatic mood detection and tracking of music audio signals," in IEEE Transactions on Audio, Speech, and Language Processing. 14, 1: 5-18, 2006.

Lu, Siyuan, Zhihai Lu & Yu-Dong Zhang. "Pathological brain detection based on AlexNet and transfer learning," Journal of Computational Science. 30: 41-47, 2019.

Markov, Konstantin & Tomoko Matsui. "Music Genre and Emotion Recognition Using Gaussian Processes," IEEE Access. 2: 688-697, 2014.

mathworks Documentation. "deeplearning: Alexnet," *https://www.mathworks.com/help/deeplearning/ref/alexnet.html*. 2019. (2019).

McEnnis, Daniel & Cory McKay. "Audio: jAudio 2," *http://jmir.sourceforge.net/jAudio.html*. [t.y.]. (8 Haziran 2018).

McFee, B., E. Humphrey & J. Bello. "A software framework for musical data augmentation," *16th Int. Soc. Music Inf. Retrieval Conf.*: 248-254. Malaga: 2015.

Meyers, Owen Craigie. *A mood-based music classification and exploration system*. Master's Thesis. Massachusetts Institute of Technology. 2007.

Moldagulova, A. & R. B. Sulaiman. "Using KNN algorithm for classification of textual documents," *8th International Conference on Information Technology (ICIT)*: 665-671. Amman: 2017.

Moors, Agnes & Klaus R. Scherer. "The Role of Appraisal in Emotion," Handbook of Cognition and Emotion: 135-155. [t.y.].

Mulligan, Kevin & Klaus R. Scherer. "Toward a Working Definition of Emotion," Emotion Review. 4, 7: 345-347, 2012.

Neeraj, Kumar, Raubin Kumar & Bhattacharya Subrata. "Testing reliability of Mirtoolbox," *Testing reliability of Mirtoolbox," 2015 2nd International Conference on Electronics and Communication Systems (ICECS)*: 710-717. Coimbatore: 2015.

On, Chin Kim, Paulraj M. Pandiyan, Sazali Yaacob & Azali Saudi. "Mel-frequency cepstral coefficient analysis in speech recognition," *2006 International Conference on Computing & Informatic*: 1-5. Kuala Lumpur: 2006.

Öztürk, Merve. "Ses Tanıma," mcu turkey. *https://www.mcu-turkey.com/ses-tanima-3/*. 2010. (2018).

Panda, Renato, Bruno Rocha & Rui Pedro Paiva. "Music Emotion Recognition with Standard and Melodic Audio Features," Applied Artificial Intelligence. 29, 4: 313-334, 2015.

Papakostas, Michalis & Theodoros Giannakopoulos. "Speech-music discrimination using deep visual feature extractors," Expert Systems with Applications. 114: 334-344, 2018.

Pouyanfar, Samira & Hossein Sameti. "Music emotion recognition using two level classification," *2014 Iranian Conference on Intelligent Systems (ICIS)*: 1-6. Bam: 2014.

Ren, Jia-Min, Ming-Ju Wu & Jyh-Shing Roger Jang. "Automatic Music Mood Classification Based on Timbre and Modulation Features," IEEE Transactions on Affective Computing. 6, 3: 236-246, 2015.

Russakovsky, O., Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg & Li Fei-Fei. "ImageNet Large Scale Visual Recognition Challenge," International Journal of Computer Vision. 115, 3: 211-252, 2015.

Salamon, J. & J. P. Bello. "Deep Convolutional Neural Networks and Data Augmentation for Environmental Sound Classification," IEEE Signal Processing Letters. 24, 3: 279-283, 2017.

Schmidt, Erik M. & Youngmoo E. Kim. "Learning emotion-based acoustic features with deep belief networks," *2011 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*: 65-68. New Paltz: 2011.

Schubert, Emery. "Update of the Hevner Adjective Checklist ," Perceptzlaland Motor Skills. 96, 3: 1117-1122, 2003.

Simonyan, K. & A. Zisserman. "Very deep convolutional networksfor large-scale image recognition," *ICLR*: 2015.

Siuly, Siuly. *ANALYSIS AND CLASSIFICATION OF EEG SIGNALS*. Doktora Tezi. UNIVERSITY OF SOUTHERN QUEENSLAND, 2012.

Song, Yading, Simon Dixon & Marcus Pearce Pearce. "Evaluation of musical features for emotion classification," *Proc. ISMIR*: 523-528. 2012.

Subha, D. Puthankattil, Paul K. Joseph, Rajendra Acharya U & Choo Min Lim. "EEG Signal Analysis: A Survey," Journal of Medical Systems. 34, 2: 195-212, 2010.

Thayer, Robert E. The Biopsychology of Mood and Arousal. Oxford University Press, 1989

Toh, A. M., R Togneri & S Nordholm. "pectral entropy as speech features for speech," *In Proceedings of PEECS*: 22-25. 2005.

Tzanetakis, George & Perry Cook. "MARSYAS: a framework for audio analysis," Organised Sound. ss. 169-175, 2000.

Tzanetakis, George & Perry R. Cook. "Musical genre classification of audio signals," IEEE Transactions on Speech and Audio Processing. 10, 5: 293-302, 2002.

Tzanetakis, George & Perry Cook. "Musical Genre Classification of Audio Signals," IEEE TRANSACTIONS ON SPEECH AND AUDIO PROCESSING. 10, 5: 293-302, 2002.

Übeyli, Elif Derya & İnan Güler. "Dalgacık Dönüşümü İle EEG İşaretlerinden Çıkarılan Öznitelik Vektörleri Üzerinde İstatistiksel İşlemlerin Gerçekleştirilmesi," *Elektrik - Elektronik ve Bilgisayar Mühendisliği Sempozyum*: Bursa: 2004.

Vaid, Swati, Singh Preeti & Chamandeep Kaur. "EEG Signal Analysis for BCI Interface: A Review," *2015 Fifth International Conference on Advanced Computing & Communication Technologies*: 143-147. Haryana: 2015.

Vapnik, Vladimir, Steven E. Golowich & Alex Smola. "Support Vector Method for Function Approximation, Regression Estimation, and Signal Processing," *Advances in Neural Information Processing Systems 9 --- Proceedings of the 1996 Neural Information Processing Systems Conference (NIPS 1996)*: 281-287. Cambridge: 1997.

Vetterl, Martin. "Wavelets, approximation, and compression," IEEE Signal Processing Magazine. 18, 5: 59-73, 2001.

Vladimir, Vapnik. The Nature of Statistical Learning Theory. 1995

Wang, Jason & Luis Perez. *"The Effectiveness of Data Augmentation in Image Classification using Deep Learning"*. 2017.

wikipedia. "Acoustic transmission," *https://en.wikipedia.org/wiki/Acoustic_transmission*. [t.y.]. (5 Kasım 2018).

Yang, Y.-H. & H.-H. Chen. "Machine recognition of music emotion: A review," ACM Trans. Intell. Syst. Technol. 3, 4, 2012.

Yang, Yi-Hsuan, Yu-Ching Lin, Ya-Fan Su & Homer H. Chen. "A Regression Approach to MusicEmotion Recognition," IEEE Transactions on Audio, Speech, and Language Processing. 16, 2: 448-457, 2008.

Yang, Yi-Hsuan, Chia-Chu Liu & Homer H. Chen. "Music emotion classification: a fuzzy approach," *Proceedings of the 14th ACM international conference on Multimedia*: 81-84. Santa Barbara: 2006.

Yang, Yi-Hsuan, Yi-Hsuan Yang, Yu-ching Lin & Homer H. Chen. "Music emotion recognition: the role of individuality," *Proceedings of the international workshop on Human-centered multimedia* : 13-22. 2007.