

T.C.
YILDIZ TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

AKUSTİK ÖZELLİKLER KULLANAN MÜZİK ÖNERİ
SİSTEMİ

Ahmet ELBİR

DOKTORA TEZİ
Bilgisayar Mühendisliği Anabilim Dalı
Bilgisayar Mühendisliği Programı

Danışman
Prof. Dr. Nizamettin AYDIN

Temmuz, 2020

T.C.
YILDIZ TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

AKUSTİK ÖZELLİKLER KULLANAN MÜZİK ÖNERİ SİSTEMİ

Ahmet ELBİR tarafından hazırlanan tez çalışması 14.07.2020 tarihinde aşağıdaki jüri tarafından Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü Bilgisayar Mühendisliği Anabilim Dalı Bilgisayar Mühendisliği Programı **DOKTORA TEZİ** olarak kabul edilmiştir.

Prof. Dr. Nizamettin AYDIN
Yıldız Teknik Üniversitesi
Danışman

Jüri Üyeleri

Prof. Dr. Nizamettin AYDIN, Danışman
Yıldız Teknik Üniversitesi

Prof. Dr. Ali BULDU, Üye
Marmara Üniversitesi

Prof. Dr. Banu DİRİ, Üye
Yıldız Teknik Üniversitesi

Prof. Dr. Fikret S. GÜRGEN, Üye
Boğaziçi Üniversitesi

Doç. Dr. Mehmet Sıddık AKTAŞ, Üye
Yıldız Teknik Üniversitesi

Danışmanım Prof. Dr. Nizamettin AYDIN sorumluluğunda tarafımca hazırlanan Akustik Özellikler Kullanan Müzik Öneri Sistemi başlıklı çalışmada veri toplama ve veri kullanımında gerekli yasal izinleri aldığımı, diğer kaynaklardan aldığım bilgileri ana metin ve referanslarda eksiksiz gösterdiğimi, araştırma verilerine ve sonuçlarına ilişkin çarpıtma ve/veya sahtecilik yapmadığımı, çalışmam süresince bilimsel araştırma ve etik ilkelerine uygun davrandığımı beyan ederim. Beyanımın aksinin ispatı halinde her türlü yasal sonucu kabul ederim.

Ahmet ELBİR

İmza



Bu çalışma Türkiye Bilimsel ve Teknolojik Araştırma Kurumu (TÜBİTAK TEYDEB-1505 Programı, Proje No:5180069) tarafından desteklenmiştir.

Aileme ithafen



TEŞEKKÜR

Doktora eğitimim süresince değerli yorumları ve katkıları ile bana yön veren danışmanım Prof. Dr. Nizamettin AYDIN hocama teşekkür ederim. Ayrıca, tez çalışmaları sırasında yardımlarıyla katkılar sunan, Tez İzleme Komitesi üyeleri olan hocalarım Prof. Dr. Ali BULDU'ya ve Doç. Dr. Mehmet Sıddık AKTAŞ'a ayrı ayrı teşekkür ederim.

Hayatım boyunca maddeten ve manen desteklerini hep yanımda hissettiğim aileme teşekkürlerimi bir borç bilirim.

Son olarak, tez çalışmamı TÜBİTAK TEYDEP 1505 projesi kapsamında destekleyen Türkiye Bilimsel ve Teknolojik Araştırma Kurumu TÜBİTAK'a katkılarından dolayı teşekkür ederim.

Ahmet ELBİR

İÇİNDEKİLER

SİMGE LİSTESİ	x
KISALTMA LİSTESİ	xi
ŞEKİL LİSTESİ	xiii
TABLO LİSTESİ	xv
ÖZET	xvii
ABSTRACT	xix
1 GİRİŞ	1
1.1 Literatür Özeti	2
1.2 Tezin Amacı	5
1.3 Hipotez	6
2 MÜZİK VE İŞARET İŞLEME KAVRAMLARI	8
2.1 İşaret ve Ses İşareti Kavramları	8
2.1.1 Bir İşaretin Değişkenleri	8
2.2 Müzik Kuramı ve Kavramları	9
2.2.1 Müzik ve İşaret Kavramı	9
2.3 Temel Müzik Kavramları	10
2.4 Sayısal Müzik İşareti İşleme	11
2.4.1 Müzik Analizi ve Fourier Dönüşümü	12
2.4.2 Periyodik İşaretlerin Analizi - Fourier Serileri	13
2.4.3 Periyodik Olmayan İşaretlerin Analizi	14
2.4.4 Kısa Zamanlı Fourier Dönüşümü - STFT	15
2.4.5 Zaman-Frekans Analizi - Spektrogram	17
3 AKUSTİK ÖZELLİK ÇIKARIMI	21
3.1 İstatistiksel Tanımlayıcılar	21
3.1.1 Aritmetik Ortalama	21
3.1.2 Standart Sapma	21

3.1.3	Ortanca - Medyan	21
3.1.4	Tepe Değer - Mod	22
3.1.5	Çarpıklık - Skewness	22
3.1.6	Basıklık - Kurtosis	22
3.1.7	Entropi	22
3.2	Zaman Düzleminde Özellik Çıkarımı	23
3.2.1	Ortalama Enerji - Root Mean Square Energy (RMSE)	23
3.2.2	Sıfır Geçiş Oranı - Zero Crossing Rate	23
3.3	Frekans Düzleminde Özellik Çıkarımı	24
3.3.1	Ortalama Frekans Enerjisi - Root Mean Square Energy (RMSE)	24
3.3.2	Spektral Ağırlık Merkezi - Spectral Centroid	24
3.3.3	Spektral Karşıtlık - Spectral Contrast	25
3.3.4	Spektral Etek - Spectral Roll-off	26
3.3.5	Polinom Özellikleri - Polynomial Features	27
3.3.6	Chroma STFT	27
3.3.7	Chroma Sabit Q Dönüşümü - Chroma CQT	27
3.3.8	Tonnetz	29
3.3.9	Mel-Frekansı Kepstral Katsayıları - MFCC	30
3.4	Dalgacık Dönüşümü İle Özellik Çıkarımı - Wavelet Transform	32
3.5	Evrişimsel Sinir Ağları (ESA) - Convolutional Neural Network (CNN)	
	Tabanlı Özellik Çıkarımı	34
3.5.1	Evrişim Katmanı - Convolution Layer	35
3.5.2	Havuzlama Katmanı - Pooling Layer	36
3.5.3	Düzleştirme Katmanı - Flatten	36
3.5.4	Tam Bağlantılı Öğrenme Katmanı - Fully Connected Layer	37
3.5.5	Bırakma Operasyonu - Dropout	37
3.5.6	Özellik Çıkarımı ve Makine Öğrenmesi	38
3.6	Özellik Son İşleme	38
4	MÜZİK ÖNERİSİ VE MÜZİK TÜRÜ SINIFLANDIRMA	40
4.1	Eğitici Öğrenme Yöntemleri	40
4.1.1	Destek Vektör Makinesi - Support Vector Machines (SVM)	40
4.1.2	Naive Bayes	41
4.1.3	Karar Ağaçları - Decision Trees (DT)	42
4.1.4	Çok Katmanlı Algılayıcı - Multi-Layer Perceptron (MLP)	43
4.1.5	K-En Yakın Komşuluklar - K Nearest Neighbours (K-NN)	45
4.1.6	Rassal Ormanlar - Random Forests (RF)	46
4.1.7	Doğrusal Ayırt Edici Analiz - Linear Discriminant Analysis (LDA)	47
4.2	Eğitici Öğrenme Yöntemleri	48

4.2.1	K-Means Kümeleme	48
4.2.2	Gauss Karışım Modeli (GMM)	48
4.3	Önerilen Derin Öğrenme Mimarileri	49
4.3.1	ESA Model-1 (ESA + Dropout)	49
4.3.2	ESA Model-2 (ESA + Dropout + Auto Encoder)	49
4.4	Müzik Önerisi	49
4.4.1	Akustik Özellikler Kullanan Müzik Önerisi	51
5	DENEYSEL SONUÇLAR	53
5.1	Veri Seti	53
5.2	Müzik Türü Sınıflandırma İçin Başarım Ölçütleri	53
5.2.1	Karmaşıklık Matrisi	53
5.2.2	Doğruluk (Accuracy)	54
5.2.3	Kesinlik (Precision)	55
5.2.4	Duyarlılık (Recall)	55
5.2.5	F Skoru (F Score)	55
5.3	Müzik Önerisinin Değerlendirilmesi	55
5.4	Sayısal İşaret İşleme ile Elde Edilen Özelliklerle Bulunan Sonuçlar . . .	56
5.4.1	K En Yakın Komşuluk Algoritması Sonuçları	56
5.4.2	Destek Vektör Makinesi Sonuçları	57
5.4.3	STFT Parametrelerinin Sınıflandırma Başarısına olan Etkisi . . .	59
5.4.4	Makine Öğrenmesi Algoritmalarının Karşılaştırılması	60
5.4.5	Özelliklerin Doğruluk Değerlerine Etkisi	61
5.4.6	Dalgacık Dönüşümü Özellikleri ile Elde Edilen Sonuçlar	65
5.5	Derin Öğrenme Yöntemleri İle Elde Edilen Sonuçlar	73
5.5.1	ESA Parametre Optimizasyonu Sonuçları	74
5.5.2	ESA Model-1 ile Elde Edilen Sonuçlar	75
5.5.3	ESA Model-2 ile Elde Edilen Sonuçlar	76
5.6	Hibrit Olarak Elde Edilen Sonuçlar	77
5.7	Müzik Türü Sınıflandırma Sonuçları Özeti	78
5.8	Müzik Önerisi Sonuçları	78
5.9	Büyük Veri Teknolojileri İle Özellik Çıkarımı Süre başarımları	82
6	SONUÇ VE ÖNERİLER	83
6.1	Sonuç	83
6.2	Öneriler	85
	KAYNAKÇA	86
A	KAYNAK KODLAR	92

A.1	ESA Model-1 Kaynak Kodu	92
A.2	ESA Model-2 Kaynak Kodu	93
B	MÜZİK TÜRÜ SINIFLANDIRMA VE MÜZİK ÖNERİSİ UYGULAMALARI	94
B.1	Müzik Türü Sınıflandırma Uygulaması	94
B.2	Müzik Öneri Sistemi Uygulaması	104
B.2.1	Derin Öğrenme Özellikleri ile Elde Edilen Sonuçlar	105
B.2.2	İşaret İşleme Yöntemleri ile Belirlenen Özellikler Kullanılarak Elde Edilen Sonuçlar	107
	TEZDEN ÜRETİLMİŞ YAYINLAR	111



SİMGE LİSTESİ

ω	Açısal frekans
$\bar{x}$	Aritmetik Ortalama (Örneklem - sample)
μ	Aritmetik Ortalama (Topluluk - population)
Ψ	Dalgacık Fonksiyonu
e	Euler Sayısı
φ	İşaretin Faz değeri
η	Öğrenme Oranı
σ	Standart Sapma
Σ	Toplam sembolü
Π	Çarpım sembolü

KISALTMA LİSTESİ

ANN	Artificial Neural Network
AUC	Area under the ROC Curve
CDBN	Convolutional Deep Belief Network
CENS	Chroma Energy Normalized Statistics
CNN	Convolutional Neural Network
CRBM	Convolutional Restricted Boltzmann Machines
CQT	Constant Q Transform
DCT	Discrete Cosine Transform
DSP	Digital Signal Processing
DT	Decision Tree
DWCH	Daubechies Wavelet Coefficient Histograms
ESA	Evriřimsel Sinir Aęları
EDİA	Evriřimsel Derin İnanc Aęları
FFT	Fast Fourier Transform
FT	Fourier Transform
GC	Gaussian classifier
GLM	Generalized Linear Model
GMM	Gaussian Mixture Model
HOSVD	High Order Singular Value Decomposition
KNN	K Nearest Neighbours
LBP	Local Binary Pattern
LDA	Linear Discriminant Analysis
LM	Linear Model

LSTM	Long Short Term Memory Network
ML	Machine Learning
MFC	Mel-Frequency Cepstrum
MFCC	Mel-Frequency Cepstral Coefficients
MIDI	Musical Instrument Digital Interface
MIR	Music Information Retrieval
MLP	Multi-Layer Perceptron
MPCA	Multi-linear Principal Component Analysis
NASE	Normalized Audio Spectrum Envelope
NMF	Non-negative Matrix Factorization
NTF	Non-Negative Tensor Factorization
ODL	Online Dictionary Learning
OSC	Octave-Based Spectral Contrast
PCA	Principal Component Analysis
RBF	Radial Basis Function
RELU	Rectified Linear Unit
RF	Random Forests
RMS	Root Mean Square
RMSE	Root Mean Square Energy
SDL	Supervised Dictionary Learning
SPR	Statistical Pattern Recognition
STFT	Short Time/Term Fourier Transform
SVM	Support Vector Machine
VSM	Vector Space Model
WT	Wavelet Transform
YSA	Yapay Sinir Ağları
ZCR	Zero Crossing Rate

ŞEKİL LİSTESİ

Şekil 1.1	Zaman düzlemi örnek gösterimi	6
Şekil 1.2	Zaman-Frekans analizi örnek gösterimi	7
Şekil 2.1	Ses işaretlerinin sınıflandırılması [26]	9
Şekil 2.2	Bir sesin farklı tınılarda görselleştirilmesi örneği [28]	11
Şekil 2.3	Periyodik sinüs işareti örnekleri	13
Şekil 2.4	Bir işaretin bileşenlerine ayrıştırılması	14
Şekil 2.5	Periyodik olmayan işaret örneği	15
Şekil 2.6	İşaretin zaman düzlemi görüntüsü	18
Şekil 2.7	İşaretin tamamının frekans düzlemi görüntüsü	18
Şekil 2.8	İlk tuşa ait frekans bilgileri	19
Şekil 2.9	İşaretin tamamının spektrogram görüntüsü	20
Şekil 2.10	Çalışmanın blok şeması	20
Şekil 3.1	Sıfır geçiş oranı - Zero Crossing Rate [36]	23
Şekil 3.2	Spektral ağırlık merkezi [40]	25
Şekil 3.3	Spektral etek	26
Şekil 3.4	Chroma STFT [31, 45]	27
Şekil 3.5	Chroma CQT [48]	28
Şekil 3.6	Frekans ekseninin doğrusal ve logaritmik olarak ayrılması	28
Şekil 3.7	Örnek Tonnetz	29
Şekil 3.8	MFCC İşlem Adımları	30
Şekil 3.9	MFCC Mel filtre bankası [50, 51]	31
Şekil 3.10	Bazı dalgacık dönüşümü fonksiyonları [56]	32
Şekil 3.11	Ayrık dalgacık dönüşümü	34
Şekil 3.12	Örnek ESA modeli	35
Şekil 3.13	Evrişim işlemi + Aynısı ile doldurma (Same Padding)	36
Şekil 3.14	RELU aktivasyon fonksiyonu	36
Şekil 3.15	Maksimum havuzlama işlemi	36
Şekil 3.16	Düzleştirme katmanı	37
Şekil 3.17	Bırakma operasyonu [60]	38
Şekil 4.1	İki sınıfı birbirinden ayırt eden destek vektörü (Kırmızı) [64]	41
Şekil 4.2	Karar ağacı - Decision tree	42

Şekil 4.3	Örnek çok katmanlı algılayıcı mimarisi	44
Şekil 4.4	K En yakın komşuluk algoritması [72]	46
Şekil 4.5	Rassal ormanlar - Random forests [75]	47
Şekil 4.6	PCA ve LDA karşılaştırması [78]	48
Şekil 5.1	Kullanılan STFT parametreleri ve değerleri [87]	59
Şekil 5.2	Bir MFCC sütununun K-Means ile kümelere ayrılması	63
Şekil 5.3	Bir MFCC sütununun GMM ile kümelere ayrılması	64
Şekil 5.4	MLP için elde edilen Karmaşıklık matrisi (Confusion Matrix)	65
Şekil B.1	Program açılış ekranı - 1	95
Şekil B.2	Program açılış ekranı - 2	96
Şekil B.3	Veri tabanı işlemleri	97
Şekil B.4	Var olan veri tabanının dahil edilmesi	97
Şekil B.5	Özellikler tablosu	98
Şekil B.6	MFCC parametre seçimi	98
Şekil B.7	Yeni şarkılar ekleme	99
Şekil B.8	Eklenen şarkılar listesi	99
Şekil B.9	Özellik çıkarımı ilerleme durum bilgisi	100
Şekil B.10	Sınıflandırma modeli özellikleri ve karmaşıklık matrisi	101
Şekil B.11	Müzik türü sınıflandırma	102
Şekil B.12	Program ana penceresi - 1	103
Şekil B.13	Program ana penceresi - 2	104
Şekil B.14	Müzik öneri sistemi örnek - 1 (Klasik Müzik Türü)	105
Şekil B.15	Müzik öneri sistemi örnek - 2 (Rock Müzik Türü)	106
Şekil B.16	Müzik öneri sistemi örnek - 1 (Klasik Müzik Türü)	107
Şekil B.17	Müzik öneri sistemi örnek - 2 (Chopin-Nocturne C Sharp Minor)	108
Şekil B.18	Müzik öneri sistemi örnek - 3 (Mehter Marşları)	109
Şekil B.19	Müzik öneri sistemi örnek - 4 (Barış Manço Gülpembe)	110

TABLO LİSTESİ

Tablo 5.1	Karmaşıklık Matrisi	54
Tablo 5.2	K-NN Başarım Yüzdeleri (K-Fold = 5, k = Farklı komşuluk değeri) .	57
Tablo 5.3	K-NN Başarım Yüzdeleri (K-Fold = 10, k = Farklı komşuluk değeri)	57
Tablo 5.4	SVM Başarım Yüzdeleri (K-Fold = 5, Kernel Tipi = Lineer)	57
Tablo 5.5	SVM Başarım Yüzdeleri (K-Fold = 10, Kernel Tipi = Lineer)	58
Tablo 5.6	SVM Başarım Yüzdeleri (K-Fold = 5, Kernel Tipi = Poly)	58
Tablo 5.7	SVM Başarım Yüzdeleri (K-Fold = 10, Kernel Tipi = Poly)	58
Tablo 5.8	SVM Başarım Yüzdeleri (K-Fold = 5, Kernel Tipi = Rbf)	59
Tablo 5.9	SVM Başarım Yüzdeleri (K-Fold = 10, Kernel Tipi = Rbf)	59
Tablo 5.10	Pencere boyutu ve örtüşme değerine göre STFT başarım yüzdeleri .	60
Tablo 5.11	Farklı pencere tiplerine göre STFT başarım yüzdeleri	60
Tablo 5.12	Farklı sınıflandırıcıların başarım yüzdeleri (Tüm özellikler)	61
Tablo 5.13	Özellikler tek başına kullanıldığında başarım yüzdeleri	61
Tablo 5.14	Özellikler farklı kombinasyonlarla kullanıldığında başarım yüzdeleri	62
Tablo 5.15	Özellikler sırayla eklendiğinde başarım yüzdeleri	62
Tablo 5.16	Kümeleme algoritmalarının çalışma hızları	63
Tablo 5.17	MFCC'nin farklı küme sayılarında başarım yüzdeleri	64
Tablo 5.18	MFCC'nin türevi ile kullanımında başarım yüzdeleri	65
Tablo 5.19	Haar ailesinin farklı seviyelere göre başarım yüzdeleri	66
Tablo 5.20	Ayrık Meyer ailesinin farklı seviyelere göre başarım yüzdeleri	66
Tablo 5.21	Symlet ailesinin farklı seviyelere göre başarım yüzdeleri	67
Tablo 5.22	Coiflet ailesinin farklı seviyelere göre başarım yüzdeleri	68
Tablo 5.23	Daubechies ailesinin farklı seviyelere göre başarım yüzdeleri	69
Tablo 5.24	Biorthogonal ailesinin farklı seviyelere göre başarım yüzdeleri . . .	70
Tablo 5.25	Ailelerin seviyelere göre ortalama başarım yüzdeleri	70
Tablo 5.26	Seviye 9 ile başarımı arttıran dalgacık fonksiyonları	71
Tablo 5.27	Seviye 6 ile başarımı arttıran dalgacık fonksiyonları	71
Tablo 5.28	Seviye 3 ile başarımı arttıran dalgacık fonksiyonları	72
Tablo 5.29	Başarı ortalaması en yüksek dalgacık fonksiyonları	73
Tablo 5.30	Çift katmanlı yöntem ile başarım yüzdeleri	73
Tablo 5.31	ESA Parametre optimizasyonu başarım yüzdeleri	74

Tablo 5.32 ESA Model-1 başarımları yüzdeleri (Chromagram, 30 saniye)	75
Tablo 5.33 ESA Model-1 başarımları yüzdeleri (Chromagram, 6 saniye)	75
Tablo 5.34 ESA Model-1 başarımları yüzdeleri (Spektrogram, 30 saniye)	75
Tablo 5.35 ESA Model-1 başarımları yüzdeleri (Spektrogram, 6 saniye)	76
Tablo 5.36 ESA Model-2 başarımları yüzdeleri (Chromagram, 30 saniye)	76
Tablo 5.37 ESA Model-2 başarımları yüzdeleri (Chromagram, 6 saniye)	76
Tablo 5.38 ESA Model-2 başarımları yüzdeleri (Spektrogram, 30 saniye)	77
Tablo 5.39 ESA Model-2 başarımları yüzdeleri (Spektrogram, 6 saniye)	77
Tablo 5.40 ESA Model-1 ve (MLP / RF / KNN / SVM) başarımları yüzdeleri	78
Tablo 5.41 Müzik öneri yüzdeleri (ESA Model-1 ve işaret işleme özellikleri) . .	78
Tablo 5.42 N sayıda öneriye müzik öneri yüzdeleri ($N = 5, 10, 20$)	79
Tablo 5.43 Verilen şarkı için öneriler - 1	80
Tablo 5.44 Verilen şarkı için öneriler - 2	80
Tablo 5.45 Önerilen müziklerin farklı türde olma yüzdeleri	81
Tablo 5.46 Apache Spark kullanımı deney sonuçları	82

Akustik Özellikler Kullanan Müzik Öneri Sistemi

Ahmet ELBİR

Bilgisayar Mühendisliği Anabilim Dalı

Doktora Tezi

Danışman: Prof. Dr. Nizamettin AYDIN

İnternetin yaygın kullanımı nedeniyle, her türden veriye erişen cihazlarla internete erişmek çok daha kolay ve ekonomik hale gelmiştir ve bu da çoklu ortam uygulamaları için yeni gelişmelere yol açmıştır. Bu gelişmeler son kullanıcı, eser sahibi ve müzik yayını yapan platformlar için yeni gereksinimler ortaya çıkarmaktadır. Müzik türü sınıflandırma, müzik önerisi ve müzik benzerliği bu gereksinimlerin bazılarıdır. Müzik yayını yapan platformlar ve uygulamalar kullanıcılarına daha fazla müzik satışı yapmak ve platformlarının yaygınlıklarını artırmak için müzik öneri sistemlerine gereksinim duymaktadırlar. Bu gereksinimi karşılayabilmek için; kullanıcının müzik dinleme geçmişi ile birlikte sanatçı adı ve müzik türü gibi bazı üst-veriler kullanılmaktadır. Müzikle kıyaslandığında; sayısal müzik sistemlerinin müziğe nispeten daha yeni hayatımıza girmesi sebebiyle, sayısal müzik sistemlerinin ortaya çıkmasından önce var olan müziklere ait üst veriler yoktur. Hatta, bazı güncel sayısal müzikler için bile üst-veriler mevcut olmayabilir. Bu tür durumda, müzik türü sınıflandırması ve müzik önerisi için müziğin akustik özellikleri kullanılabilir. Ayrıca, bu yaklaşım üst verisi olan müziklerin sınıflandırılmasında ve sonuçların iyileştirilmesi için ek bir özellik olarak da kullanılabilir. Bu tez çalışması kapsamında müzik önerisi, müzik benzerliği ve müzik türü sınıflandırma problemlerinde kullanmak için müzikten akustik özelliklerin çıkarılması gerçekleştirilmiştir. Bu bağlamda sayısal işaret işleme, makine öğrenmesi ve derin öğrenme yöntemleri kullanılmıştır. Sayısal işaret işleme açısından, zaman - frekans analizi yöntemleri ve bu yöntemlerde kullanılan parametrelerin müzik türü sınıflandırma başarısına etkisi araştırılmıştır. Sınıflandırma başarısına anlamlı bir katkının olup olmadığını gözlemlemek için dalgacık dönüşümü teknikleri sayısal işaret işleme başlığı kapsamında ayrıca incelenmiştir. Diğer yandan,

kullanımı gittikçe yaygınlaşan derin öğrenme yöntemleri akustik özellik çıkarımı ve müzik türü sınıflandırma amacıyla araştırılmış ve bu amaçla farklı evrimsel sinir ağları tasarlanmıştır. Aynı türden müziğin tespit edilmesi akustik özellikler kullanan müzik önerisi sistemi için ölçüt olarak kullanılmıştır. Bu ölçüte göre önerilen yöntemlerin başarımları karşılaştırılmıştır. Derin öğrenme yöntemleri ile elde edilen sonuçların geleneksel sayısal işaret işleme yöntemlerine göre daha başarılı olduğu gözlemlenmiştir. Bu çalışmada Turkcell tarafından Fizy uygulamasına eklenecek olan akustik öneri motorunun teorik ve uygulamalı altyapısı geliştirilmiştir. Önerilen model tabanlı sınıflandırma sistemi, etiketsiz (ya da üst verisi bulunmayan) veriler üzerinde kullanılarak girdi olarak verilen bir müziğe benzer müzikleri tavsiye edebilmektedir.

Anahtar Kelimeler: Müzik öneri motoru, akustik müzik özellikleri, müzik işareti işleme, müzik türü sınıflandırma, makine öğrenmesi, derin öğrenme



Music Recommendation System Using Acoustic Features

Ahmet ELBİR

Department of Computer Engineering

Doctor of Philosophy Thesis

Advisor: Prof. Dr. Nizamettin AYDIN

Due to widespread use of the Internet, to access the Internet through devices that access data of any kind has become much easier and more economical, which has lead to new developments for multimedia applications. These developments engender new requirements for end user, author and music streaming platforms. Music genre classification, music recommendation and music similarity are some of these requirements. Music streaming platforms and apps call for music recommendation systems to sell more music to their users and to enhance the prevalence of their platforms. In order to meet these requirements; as well as the user's music history, some meta-data such as artist's name and genre are used. Due to the fact that the digital music systems are relatively new in our lives compared to music, meta-data about music existing before the advent of digital music systems do not exist. Even meta-data for some contemporary digital music may not be available. In this case, acoustic features of music can be used for music genre classification and music recommendation. This approach can also be used as an additional feature to classify music with metadata and to improve the results. Within the scope of this thesis, acoustic features have been extracted from the music in order to use in music recommendation, music similarity and music type classification problems. In this context, digital signal processing, machine learning and deep learning methods have been employed. In terms of digital signal processing, time - frequency analysis techniques have been investigated and the effects of the parameters used in these techniques over the music genre classification performance have been examined. In order to observe whether there is a significant improvement over classification

performance, wavelet transform techniques have been studied separately within the scope of digital signal processing. On the other hand, deep learning methods, which are becoming more and more popular, have been studied for acoustic feature extraction and classification of music, and different convolutional neural networks have been designed for this purpose. Recognition of the same genre of the music has been used as a criterion for the music recommendation system that uses acoustic features. The performances of the proposed methods have been compared according to this criteria. It has been observed that the results obtained by deep learning methods are more successful than conventional digital signal processing methods. In this study, the theoretical and applied infrastructure of the acoustic recommendation engine, which will be added to the Fیزی application by Turkcell, has been developed. By using unlabeled data (or without using meta-data), the proposed model-based classification system can recommend musics that are similar to a music given as input.

Keywords: Music recommendation engine, acoustic music features, music signal processing, music genre classification, machine learning, deep learning

1 GİRİŞ

İnternet kullanımının yaygınlaşması sayesinde, internet erişimi bulunan aygıtlar yoluyla her türden veriye erişmek çok daha kolay ve ekonomik hale gelmiştir. Müziğin insan hayatındaki geçmişi, gelişimi ve önemi göz önünde bulundurulduğunda, müzik insanlar için sürekli artan öneme sahip olmaktadır ve internet sayesinde müzik konusunda çok çeşitli gelişmeler meydana gelmektedir. Geçmişten günümüze üretilen bir çok müzik eserinin sayısal ortamlarda saklanabilmesi, son kullanıcının her türden müziğe zamandan ve mekandan bağımsız olarak erişebilmesi bu gelişmelerden bazılarıdır. Özellikle internet tabanlı müzik işleme ve müzik yayını platformlarının hayatımızın ayrılmaz parçası haline gelmesi ile birlikte bu alanda yoğunlaşılması gereken farklı gereksinimler ve problemler ortaya çıkmaktadır. Müzik yayını yapan platformların kullanıcılarına müzik önerisi yapabilmesi, müzikleri türlerine göre sınıflandırabilmesi bu gereksinimlerden bazılarıdır. Bu iki gereksinim sadece müzik yayını yapan platforma avantaj sağlamakla kalmayıp, kullanıcıya da yeni müzikleri keşfetme, duygu durum tespiti gibi çeşitli avantajlar sağlamaktadır. Müzik türü tanıma ve müzik öneri sistemlerinde, genel olarak müzik verisine ait sanatçı adı, albüm adı, müzik tarzı gibi üst-veriler; kullanıcının müzik dinleme geçmişine ait veriler, ya da işbirlikçi filtreleme (collaborative filtering) kullanarak kullanıcının profiline en çok benzeyen bir başka kullanıcıya ait veriler kullanılmaktadır. Kuramsal açıdan bu veriler söz konusu problemlerde kolaylıkla kullanılmaktadır. Ancak üst verinin veya kullanıcı geçmişinin olmaması gibi nedenlerden dolayı uygulamada bu sorunların çözülmesi gerektiğinde çeşitli ön problemler ortaya çıkmaktadır. Müziğin ayırt edici özelliklerinin çıkarılması olarak da ifade edilebilen bu süreçte sayısal işaret işleme ve yapay zeka tekniklerinin kullanımı ön plana çıkmaktadır. Gelişen yazılımsal ve donanımsal kolaylıklar sayesinde her türden müziğe ait özellik çıkarımı ve müzik türü tanıma kolaylıkla uygulanabilmektedir.

Ülkemizde Turkcell bünyesinde faaliyet gösteren Fizy müzik yayını uygulamasına müzik önerisi ve müzik keşfi motorunun eklenmesi planlanmaktadır. Bu sayede hem uygulamayı kullananlara kolaylık sağlanacak hem de uygulamanın ülkemiz iç

pazarında kullanılabilirliği artacaktır. Bu hedef başta olmak üzere, eser sahiplerinin telif haklarının korunması için müzik benzerliği tespiti ve Türk müziğine ait özellik veri tabanının oluşturulması gibi konularda kullanmak üzere bu tez çalışması gerçekleştirilmiştir.

Akustik özellikler kullanan müzik önerisi ve müzik türü sınıflandırması konusunda hazırlanan bu tezin organizasyonu şu şekildedir. Bölüm 1.1 bu alanda yapılan daha önceki çalışmaları, Bölüm 1.2 tezin amacını ve konusunu, Bölüm 1.3 tezin içeriğine ilişkin öne sürülen hipotezi detaylandırmaktadır. Bölüm 2 müzik teorisine ait temel kavramları, sayısal işaret işleme ve müzik işareti işleme ile ilgili literatürde bulunan yöntemlerin temel bilgilerini ele almaktadır. Bölüm 3 ve Bölüm 4 kapsamında akustik özellik çıkarımı ve müzik türü sınıflandırması yapacak sayısal işaret işleme, makine öğrenmesi ve derin öğrenme yöntemleri açıklanmakta, önerilen ve kullanılan yöntemlerin ve modellerin detayları verilmektedir. Bölüm 5'te, Bölüm 3 ve 4'te önerilen modellerle ve bu modellerin kullandığı parametrelerin incelenmesi ile elde edilen başarımlar sonuçları özetlenmektedir. Son bölüm olan Bölüm 6, tez çalışması sırasında elde edilen sonuçları ve kazanımları, gelecekte yapılması planlanan hedef kazanımları açıklamaktadır.

1.1 Literatür Özeti

Müzik parçalarından bilgi çıkarımı ve müziğin sınıflandırılması özellikle müzik öneri sistemi gibi uygulamalar için sektörde sıkça karşılaşılan problemlerdir. Ünlü müzik yayın platformu Spotify'nın müzik öneri sistemi, işbirlikçi filtreleme (collaborative filtering) ve doğal dil işleme yöntemlerinin yanı sıra bu projenin de odak noktası olan ham ses verisinin modellenmesi yöntemini kullanmaktadır [1]. Literatürde de bu alanda yapılan çalışmalar vardır.

Haggblade, Hong ve Kao, GTZAN veri setindeki müziklerle sadece MFCC (Mel Frequency Cepstral Coefficients) kullanarak, klasik, caz, metal ve pop olmak üzere 4 farklı müzik türü için müzik özellikleri çıkarmış; k-NN (k Nearest Neighbor), k-ortalamalar (k Means), yapay sinir ağları (YSA) ve destek vektör makineleri (Support Vector Machine - SVM) yöntemlerini kullanarak yaklaşık %80 doğrulukla müzik türlerini sınıflandırmıştır [2].

Tzanetakis ve Cook, 10 farklı müzik türü için, tınısal doku (timbral texture), ritmik içerik (rhythmic content) ve ses perdesi içeriği (pitch content) adını verdikleri 3 özellik seti oluşturmuşlardır. Çıkarılan bu özelliklerin performansını, istatistiksel örüntü tanıma sınıflandırıcıları ile hem verinin tamamını hem de gerçek-zamanlı çerçeve-tabanlı sınıflandırma şemaları kullanarak test etmişlerdir [3].

Grzywczak ve Gwardys, m ziklerden  ıkardıkları spektrogramları, resim sınıflandırmak i in e itilmi  bir evri imsel yapay sinir a ına girdi olarak verip,  ıktıları m zik  zelli i olarak kullanmı lar ve SVM kullanarak sınıflandırma yapmı lardır. 10 farklı m zik t r  i in GTZAN veri setinde %78 ba arı oranı elde edilmi tir [4].

Tzanetakis ve arkadaşları, Daubechies dalgacık katsayısı histogramını kullanarak destek vekt r makinesi (SVM) ile %79.5 ba arım elde etmi tir [5].

Nasridinov ve Park yaptıkları  alı mada m zik t r n n sınıflandırılması i in    adımlı bir y ntem  nermi lerdir. İlk adım olan *akor* etiketleme  emasında m zik, ikili sayı sisteminde ifade edilmi tir. İkili sistemi dizi olarak ifade edilen m zikler, ikinci adım olan t r e le tirme a amasında zaman dizisi olarak d   n l rek iki dizi arasında kar ılıklı olarak alt dizi arama i lemleri yapılmı tır. Son adım olan t r belirleme a amasında, ikinci adımda e le en m zik dizisi,  nceden belirlenmi  olan m zik t rlerine g re sınıflandırılmı tır.  alı mada altı farklı t re g re sınıflandırma yapılmı tır: Prehistorik zaman, orta a , r nesans, barok, klasik ve modern [6].

Rosner, Schuller ve Kostek;  alı malarında rock ve metal gibi vurmali  algıların sıklıkla kullanıldı ı m zik t rlerinde sınıflandırma sonu larını iyile tirmi lerdir. M zik i areti davul ve harmonik bile enler olarak ikiye ayrılmı  ve elde edilen belirleyici  zellik sayısı arttırılmı tır [7].

Lidy, Thomas ve arkadaşları m zik t r n n sınıflandırılması i in GTZAN, ISMIRrhythm ve ISMIRgenre veri setlerini kullanmı lardır.  zellik olarak Ritim  zellikleri, istatistiksel spektrum tanımlayıcılar ve ritim histogramı  zellikleri kullanılmı tır. Veri k meleri SVM ile sınıflandırıldıktan sonra, t m  zellik k meleri birle tirilerek en iyi do ruluk %85 olarak bulunmu tur [8].

Li, Tao ve arkadaşları ritim  zellikleri, pitch  zellikleri ve tınısal doku  zelliklerini kullanarak m zik t rlerini sınıflandırmı lardır. Do rusal Diskriminant Analizi (Linear Discriminant Analysis - LDA), SVM, k-NN ve Gauss Karı ım Modelleri (Gaussian Mixture Mode - GMM) gibi farklı makine   renmesi algoritmalarını kar ıla tırmı lardır. SVM ile en iyi do ruluk oranını % 80 olarak elde etmi lerdir [9].

Tao Li ve arkadaşları, GTZAN veri k mesinden elde ettikleri MFCC, Daubechies dalgacık d n   m  katsayılarının histogramlarını (DWCH) ve di er tınısal  zellikleri SVM sınıflandırıcıda kullanarak %75.1 do ruluk elde etmi lerdir [10].

Anders Meng ve arkadaşları kısa s reli  zelliklerin zamansal  zellik entegrasyonunu ara tırmı lardır. LM, GC, GMM ve GLM gibi farklı algoritmalar denemi lerdir. 128 kb/s bit hızlarıyla kodlanan MP3'ler i in en iyi do ruluk %93.3 kaydedilmi tir [11].

Chin-Chia Michael Yeh ve diğerleri MFCC ve Spektrogram ile çıkarılan özellikleri GTZAN veri setinde k-means, çevrimiçi sözlük öğrenmesi (ODL) ve eğitici sözlük öğrenmesi (SDL) algoritmalar kullanarak çalışmışlardır. Spektrogram ile %84.7 doğruluk elde etmişlerdir [12].

Lee ve arkadaşları MFCC, oktav tabanlı spektral karşıtlık (OSC) ve normalize edilmiş ses spektrogramları tekniğini (envelop) (NASE) ile GTZAN veri setinden özellik çıkarmışlardır. LDA kullanarak %90.6 başarıya ulaşmışlardır [13].

Lee ve arkadaşları *CNN derin inanç ağları (deep belief networks)* kullanarak 5 farklı türdeki müziği (klasik, elektrik, jaz, pop, and rock) karşılaştırmış ve %69.75 başarımla elde etmiştir [14].

Costaet ve diğerleri, SVM kullanarak özellik çıkarımı ve sınıflandırılmış türler için LBP (Local Binary Pattern) kullanarak ve 10 müzik türü sınıflandırmada %82.33 başarımla elde etmişlerdir [15].

Ioannis Panagakis ve arkadaşları çoklu ölçekli spektro-temporal modülasyon özellikleri çıkarmışlardır. Negatif olmayan tensör faktörizasyonu (non-negative tensor factorisation - NTF), yüksek dereceli tekil değer ayrıştırma (HOSVD) ile çok doğrusal ana bileşen analizi (Multi-Linear Principle Component Analysis - MPCA) kullanarak GTZAN veri setinde SVM kullanarak %78.2 başarıya ulaşmışlardır [16].

Bo Sahao ve arkadaşları, müziğin başlangıcında ilk bir dakikada sadece enstrüman olacağını deneyimlerinden test ederek 60. ile 90. saniye arasındaki 30 saniyelik bölümü özellik çıkarımı için kullanmışlardır. Özellik çıkarımı sonrasında 80 farklı özelliği içeren özellik vektörü oluşturmuşlardır. Bunlardan 52 tanesi MFCC, 12 tanesi STFT ve kalan 16 tanesi wavelet tabanlı özelliklerdir. Bu çalışmada içerik – akustik tabanlı yaklaşım için bu özellikler kullanmışlardır. Bunlardan başka sanatçı tabanlı, kullanıcı geçmiş tabanlı ve hibrid olmak üzere dört farklı deneme yapılmıştır [17].

Yadin Song ve arkadaşları kullanıcı profili tabanlı, dinleyici profili tabanlı ve akustik tabanlı yöntemleri incelemişlerdir. Akustik özelliklerden tını ve ritmin özellik olarak kullanılabileceğini belirtmişlerdir. Bu özelliklere göre K-Means ve EM Kümeleme ve Euclidean mesafesi kullanarak en benzer müziği bulma işlemi önermişlerdir [18].

Hareesh Bahuleyan, 7 müzik türünü (Pop, rock, hip-hop, tekno, ritim blues, vokal, reggae) sınıflandırmak için VGG 16 CNN kullanmış ve %64 başarımla elde etmiştir [19].

Siddharth Sigtia ve Simon Dixon, CNN kullanarak GTZAN veri kümesinden 10 tür müzik için %83'lük bir başarımla sağlamışlardır [20].

Holzapfel ve Stylianou negatif olmayan matris faktörizasyonu (non-negative matrix factorisation - NMF) yöntemi kullanarak %74 doğruluğa ulaşmışlardır [21].

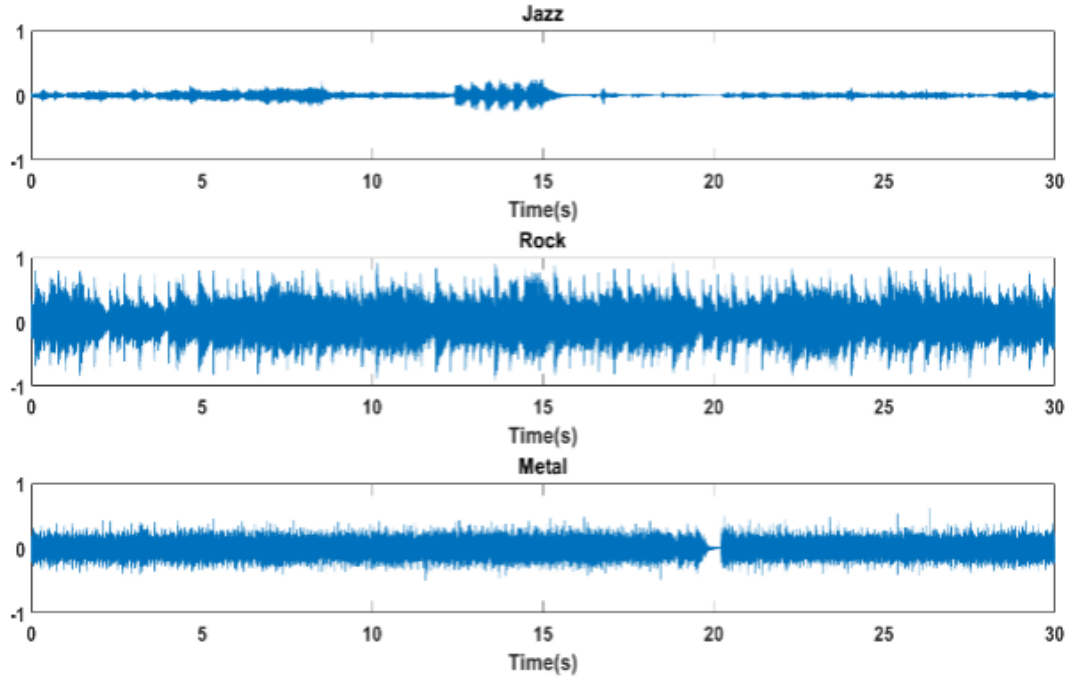
Shin ve arkadaşları önerdikleri *auditory spice code* yöntemiyle 172 farklı özellik elde ederek %85 başarı kaydetmişlerdir [22].

1.2 Tezin Amacı

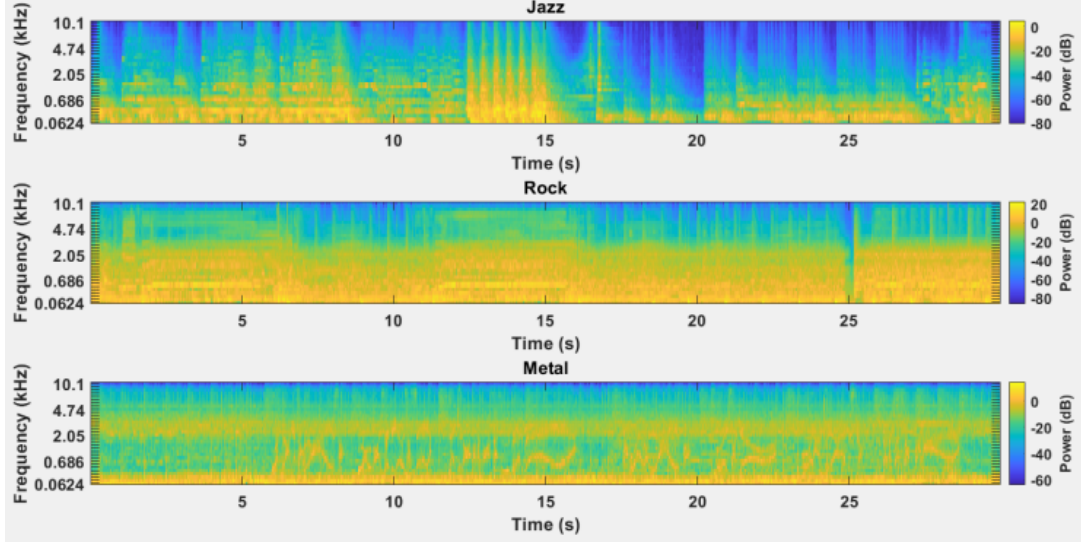
Bu tez çalışmasının amacı, müzik türü tanıma ve müzik önerisi problemlerini müzikten elde edilecek akustik özelliklerle çözecek bir sistem geliştirmektir. Özellikle Türk müziğinin tarihsel zenginliği düşünüldüğünde, bu zenginliğin müzik içeriği bakımından analiz edilmesi bu tez çalışmasıyla birlikte çeşitli yenilikler ortaya çıkaracaktır. Çalışma kapsamında, müzikten akustik özellikler çıkarmak için sayısal işaret işleme yöntemleri öncelikle ele alınmıştır. Bu başlık altında zaman-frekans analizi yöntemleri incelenmiştir. Bu yöntemlerin kullandığı parametrelerin değişiminin müzik türü tanıma sistemi başarımı üzerindeki etkisi çalışmadaki bir başka odak noktası olmuştur. Sayısal işaret işleme yöntemleri kapsamında dalgacık dönüşümü kullanarak zaman-frekans analizlerine ek olarak farklı analizler yapılmıştır. Son zamanlarda her alanda sıklıkla kullanılan makine öğrenmesi ve derin öğrenme yöntemleri, akustik özellik çıkarımı, müzik türü tanıma ve müzik önerisi için ayrıca ele alınmış ve elde edilen sonuçlar diğer yöntemlerle elde edilen sonuçlarla karşılaştırılmıştır. Müzik türü sınıflandırma işlemleri için etiketli veri setine ihtiyaç duyulmuş ve bu alandaki çalışmalarda yaygın olarak kullanılan GTZAN veri tabanı kullanılmıştır. Bu çalışmada GTZAN veri tabanına ek olarak Turkcell'in Fizy uygulamasında kullanılacak olan müzik önerisi işlemleri için Turkcell tarafından sağlanan tür etiketi olmayan müzikler kullanılmıştır. Etiketli veriler üzerinde en yüksek doğruluğu veren sınıflandırıcı model ve akustik özellikler, müzik önerisi işlemlerinde kullanılmıştır. Bu sayede müzik önerilerinin benzer türlerden olması ve öneri sonuçlarının sayısal olarak yorumlanması mümkün olmuştur. GTZAN veri seti üzerinde yapılan çalışmalarda sayısal işaret işleme yöntemleri ile müzik türü sınıflandırması ortalama %70 doğrulukta olurken, derin öğrenme ve makine öğrenmesi yöntemlerinin birlikte kullanımında doğruluk oranı ortalama %95'e ulaşmıştır. Ayrıca, bu çalışma bağlamında kullanılan algoritmaların hesaplama karmaşıklığından dolayı daha hızlı özellik çıkarımı yapabilmek adına büyük veri yöntemleri kullanılmış ve bu sayede geleneksel işaret işleme yöntemlerinden elde edilen akustik özellik veri tabanı daha kısa sürede oluşturulmuştur.

1.3 Hipotez

Tez kapsamında yapılan incelemeler ve karşılaştırmalar sonucu, müzik önerisi, müzik türü sınıflandırma ve müzik telif hakları gibi müzik içeriğine ilişkin konularda akustik özellik çıkarımı ile elde edilen verilerin her müzik için rahatlıkla elde edilebildiği anlaşılmaktadır. Diğer taraftan, müziğe ait üst-veri, kullanıcı hareketleri gibi verilerin elde edilmesinin her zaman mümkün olmayışı nedeniyle bu tür problemlerin çözümü mümkün olmamaktadır. Akustik özellik çıkarımı ile içerik tabanlı müzik benzerliği, müzik türü sınıflandırma söz konusu olduğunda, özellikle müzik türleri arasındaki benzerliğin fazla olduğu durumlar için ayırt edicilik geleneksel işaret işleme yöntemlerinde derin öğrenme yöntemlerine kıyasla daha zor ve başarısız olmaktadır. Şekil 1.1 ve Şekil 1.2 türler arası benzerliğin fazla ve az olduğu durumları zaman-frekans düzleminde görselleştirerek örneklendirmektedir. Burada Jazz - Rock ve Jazz - Metal türlerii arasındaki ayırt edicilik, Rock ve Metal arası ayırt ediciliğine göre daha belirgindir.



Şekil 1.1 Zaman düzlemi örnek gösterimi



Şekil 1.2 Zaman-Frekans analizi örnek gösterimi

Bu tez kapsamında, yukarıda ifade edilen benzerliklerin meydana getirdiği güçlükler sebebiyle, hem özellik çıkarımı yapacak yöntemlerin daha ayırt edici özellikleri çıkarabilmesi hem de sınıflandırıcı modelin daha hassas sınıflandırma yapabilmesi için farklı özellik çıkarımı ve müzik türü sınıflandırma modellerinin kullanılması ve en uygun olanının seçilmesi önerilmektedir. Özellikle, kullanılan ve önerilen yöntemlerin varsayılan parametrelerini doğrudan kullanmak yerine, bu yöntemlerin en uygun parametrelerinin saptanması daha yüksek başarıların elde edilebildiğini deneysel sonuçlarla göstermektedir.

2.1 İşaret ve Ses İşareti Kavramları

İşaret kavramı, bir veya daha fazla değişkene bağlı olarak tanımlanan bir fonksiyon veya günlük hayattaki fiziksel büyüklükler hakkında bilgi bir taşıyan değişken olarak tanımlanabilir. Bahsedilen bu fiziksel büyüklüklerden bir tanesi de zamana bağlı bir değişkenle tanımlanan bir fonksiyon olan sestir.

İşaretin sayısal olarak temsil edilmesi sayısal işaret tanımını ortaya çıkarmaktadır. Sayısal işaretin çeşitli algoritmalar ve yöntemlerle işlenerek yeni işaretler ve bilgiler elde edilmesi süreçleri de sayısal işaret işleme olarak tanımlanabilir [23–25]. Bir kaynaktan gelen ses dalgalarının havadaki basınç değişimi yoluyla algılayıcı özelliği olan kulağa ulaşmasına ses denir. İnsan kulağı 20 Hz ve 20.000 Hz arasındaki frekanslardaki sesleri algılayabilmektedir.

2.1.1 Bir İşaretin Değişkenleri

Zamana bağlı değişen sinüsoidal bir işaretin üç temel değişkeni vardır. Bunlar genlik (Amplitude), frekans ve faz değişkenleridir. Örnek bir sayısal işaretin tüm bileşenleri Eşitlik 2.1, Eşitlik 2.2 ve Eşitlik 2.3 ile gösterilmektedir. Eşitliklerdeki A genlik(Amplitude) değerini, t zaman bilgisini, ω açısal frekansı, f frekansı, φ faz bilgisini ve π ise sabit pi sayısını göstermektedir.

$$X(t) = A \cdot \cos(\omega \cdot t + \varphi) \quad (2.1)$$

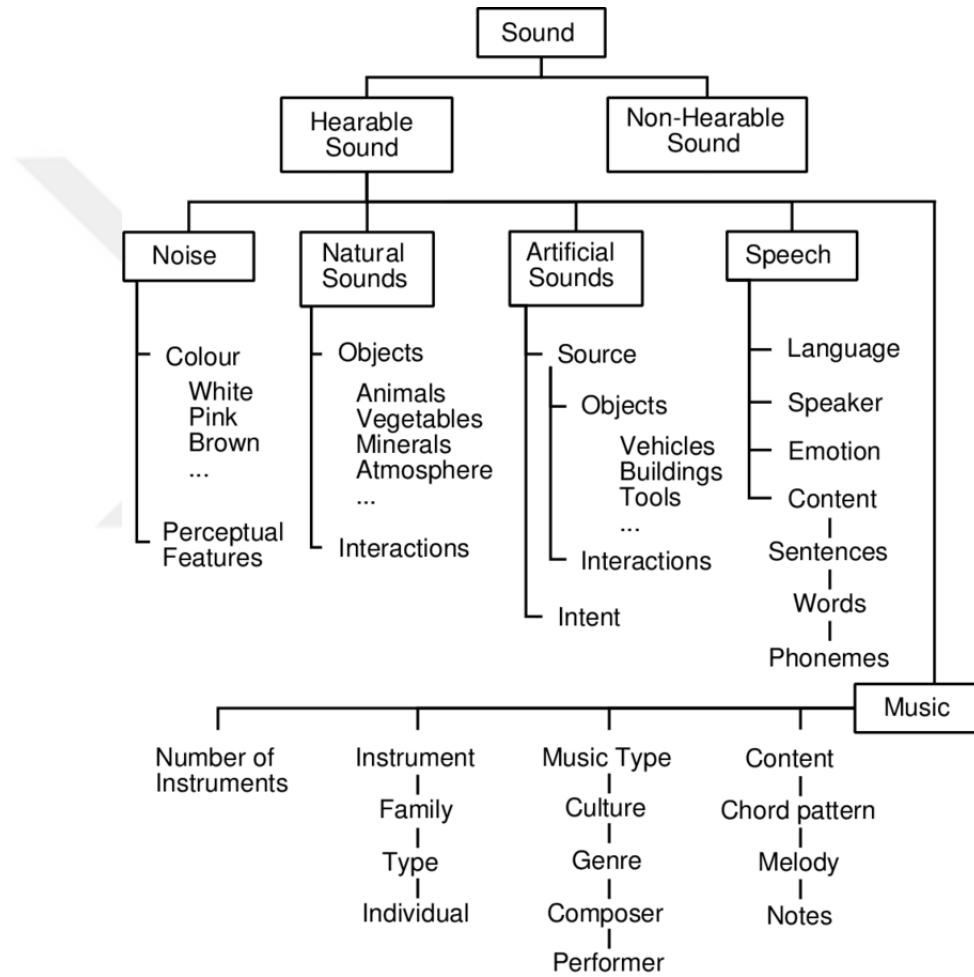
$$\omega = 2 \cdot \pi \cdot f \quad (2.2)$$

$$X(t) = A \cdot \cos(2 \cdot \pi \cdot f \cdot t + \varphi) \quad (2.3)$$

Bu çalışmada; müzik, sayısal bir işaret olarak ele alınmış olup, müzikten anlamlı bilgiler çıkarmayı sağlayacak sayısal işaret işleme yöntemleri incelenmiştir.

2.2 Müzik Kuramı ve Kavramları

Hiyerarşik olarak incelendiğinde müzik bir tür ses işaretidir. David Gerhard'ın ses işaretlerini sınıflandırdığı çalışmasında [26] müziğin işaret işlemedeki yeri ve çeşitleri Şekil 2.1 ile özetlenmiştir. Buna göre müzik, duyulabilir sesler sınıfında yer alırken, farklı enstrüman, içerik ve tür gibi özellikleri bulunmaktadır.



Şekil 2.1 Ses işaretlerinin sınıflandırılması [26]

2.2.1 Müzik ve İşaret Kavramı

Müzik kavramı çok eski zamanlardan beri var olan evrensel bir kavramdır. Müzik kelimesinin bazı dillerdeki karşılıkları: İngilizce'de music, Fransızca'da musique, Latince'de musica, Eski Yunanca'da mousikē şeklindedir. Musiki, musika, muzika, müzik gibi kelimeler Yunanca kökenlidir. Bunun da eski Yunanca'daki Moüsai "ilham perileri" anlamından türediği ifade edilir ve birçok dilde kelime olarak benzer

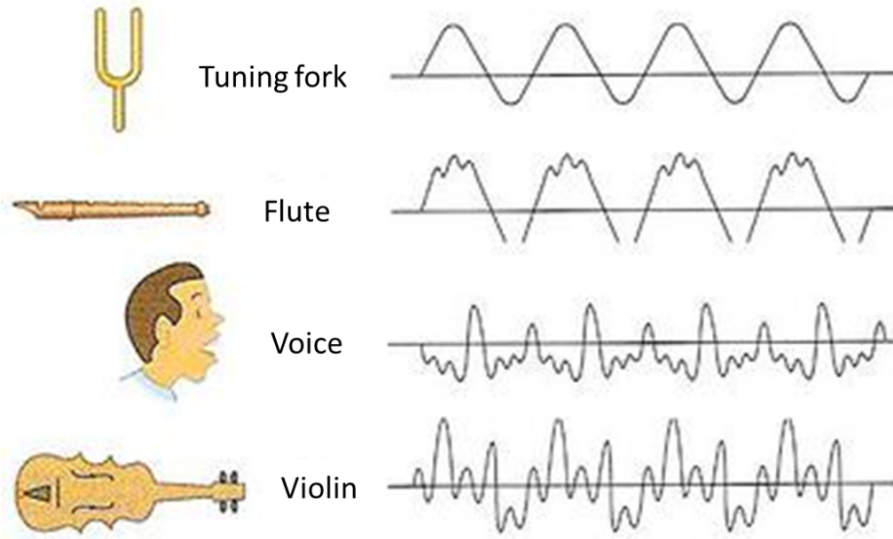
karşılıkları vardır [27]. Bu evrensel özelliği sebebiyle müziğin birçok tanımı vardır. Türk Dil Kurumu'na göre müziğin isim olarak tanımı şöyle yapılmıştır.

- "Birtakım duygu ve düşünceleri belli kurallar çerçevesinde uyumlu seslerle anlatma sanatı, musiki: Örnek: Müzik eğitimi. (TDK)"
- "Bu biçimde düzenlenmiş seslerden oluşan eserlerin okunması veya çalınması: Örnek: Bu akşam güzel bir müzik dinledik. (TDK)"

2.3 Temel Müzik Kavramları

Müzik teorisi altında birçok kavram ve terim bulunmaktadır. Çalışmamızda sayısal işaret işleme yöntemleri ile müzik işaretleri işleneceği için bu kapsamdaki terim ve kavramlara yer verilmiştir.

- **Yükseklik:** Sesin inceliği (tizlik) ve kalınlığı (peslik) ile ilgili bir özelliğidir.
- **Nota:** Müziğin alfabesi olarak tanımlanabilir. Her bir nota farklı frekansların birleşiminden oluşan müzik işaretidir. Do, re, mi, fa, sol, la, si olarak 7 çeşittir. Notaların İngilizce karşılıkları sırasıyla C, D, E, F, G, A ve B'dir. 440 Hz frekansı olan La sesi standart olarak kabul edilmiştir. Kısaca nota, seslerin yükseklikleri ve süreleri hakkında bilgi veren, müziğin icra bileşenidir.
- **Ritim:** Sesin veya müziğin belli bir düzen ve ahenkle sıralanmasıdır.
- **Solfej:** Notaların isimlerinin, bu isimlere karşılık gelen seslerin ve bu seslerin sürelerinin okunması demektir.
- **Oktav:** Aynı isimdeki iki nota arasındaki 8 notadan meydana gelen aralık demektir. Örneğin, do, re, mi, fa, sol, la, si, do diziliminde iki do arası bir oktav olarak adlandırılır.
- **Ton - Tonalite:** Eserin merkez bir notada oluşturulması anlamına gelir. Ses olarak da ifade edilir. Majör ve Minör tonalite olarak ikiye ayrılmıştır. Ton kavramının amacı müziğin belli bir notanın etrafında şekillenmesini ifade etmektir. Örneğin bir müzik do minor notası etrafında şekillenmişse, tonu da do minör olarak ifade edilir.
- **Tını (Timbre):** Belli bir sesin üretilen kaynağa göre ayırt edicilik özelliğidir. Bir notanın insan, keman, flüt tarafından seslendirilmesi aynı nota olmasına rağmen farklı tınlar üretmektedir. Şekil 2.2 aynı notanın farklı kaynaklardan gelen farklı tınlarını göstermektedir [27].



Şekil 2.2 Bir sesin farklı tınılarda görselleştirilmesi örneği [28]

- **Perde - Pitch:** Perde veya Pitch, algılanan frekansla alakalı olan müzik özelliğidir. Örneğin, 440 Hz'lik frekanslı bir notanın frekansında olan küçük değişiklikler algılanan frekansı değiştirmez, bu sayede hangi notaya ait sesin üretildiği algısında bir değişiklik olmaz. Algılama, oktav özelliğindeki ikiye katlanan frekanslarda olduğu gibi logaritmik olarak değişmektedir [25, 29, 30].
- **Zarf - Envelope:** Zarf, müzikte zamanla meydana gelen genlik ve frekans değişimini açıklayan protokole verilen genel bir isimlendirme protokolüdür. En çok kullanılanlardan birisi ADSR (Attack - Yükselme, Decay - İniş, Sustain - Tutma veya Plato, Release - Bırakma) protokolüdür. Bu protokole göre; atak, sesin ilk üretildiği andan en tepe genlik değerine ulaşana kadar geçen süreyi açıklarken; iniş, sesin plato veya durgunluk genlik değerine ulaşana kadar geçen süreyi ifade eder. İşaretin belli bir genlik değerinde değişmeden kalması sürecine tutma, plato veya durgunluk denir. Bırakma, işaretin durgunluk seviyesinden tamamen duyulamaz hale gelen kadar geçen süreyi açıklayan parametredir.

2.4 Sayısal Müzik İşareti İşleme

Bu bölümde Sayısal Müzik işaretinin işlenmesinde kullanılan yöntemlerden genel olarak bahsedilecektir. Çalışmanın konusu olan akustik özellik çıkarımı, müzik önerisi ve müzik türü sınıflandırma konuları Bölüm-3 ve Bölüm-4 içerisinde detaylandırılacaktır.

2.4.1 Müzik Analizi ve Fourier Dönüşümü

Bir müziğin analiz edilmesi ve bu analiz sonrasında ayırt edici özelliklerinin çıkarılması sayısal işaret işlemenin bir alt konusudur. Şekil 2.1 bir müziğin sayısal işaret işlemedeki yerini göstermektedir. Bu görselden yola çıkarak müziğin ayırt edici özelliklerinin çıkarılması için kullanılması gereken yöntemler genelden özele şu şekilde sıralanabilir.

- Sayısal işaret işleme yöntemleri
- Ses işareti işleme yöntemleri
- Müzik işareti işleme yöntemleri

Bu sıralama bir diğer ifadeyle düşük seviye, orta seviye ve yüksek seviye olarak da ifade edilebilir [3, 21, 31, 32].

Bölüm-3 ve Bölüm-4 kapsamında ses işareti işleme ve müzik işareti işlemeye dair bilgiler detaylı olarak verilmiştir. Bu bölümde temel sayısal işaret işleme yöntemlerinin bu çalışmaya yön verecek kullanımı açıklanacaktır.

Eşitlik 2.3’de bir işaretin değişkenleri gösterilmişti. Bu değişkenler, işaretten ayırt edici özellik çıkarımı, işaret analizi ve işaret sentezi gibi konularda sıklıkla kullanılmaktadır.

İşaret işlemede, verilen bir işaretin zaman ve frekans analizinin yapılmasında en sık kullanılan yöntemlerden birisi Fourier dönüşümüdür. Verilen bir $x(t)$ işaretinin içeriğindeki temel sinüs ve kosinüs bileşenlerine ayrıştırmayı sağlayan Fourier dönüşümü Eşitlik 2.4 veya Eşitlik 2.5’de gösterildiği gibi gerçekleştirilir [3, 21, 31, 32].

$$X(\omega) = \int_{-\infty}^{\infty} x(t).e^{-j.2.\omega.t}.dt \quad (2.4)$$

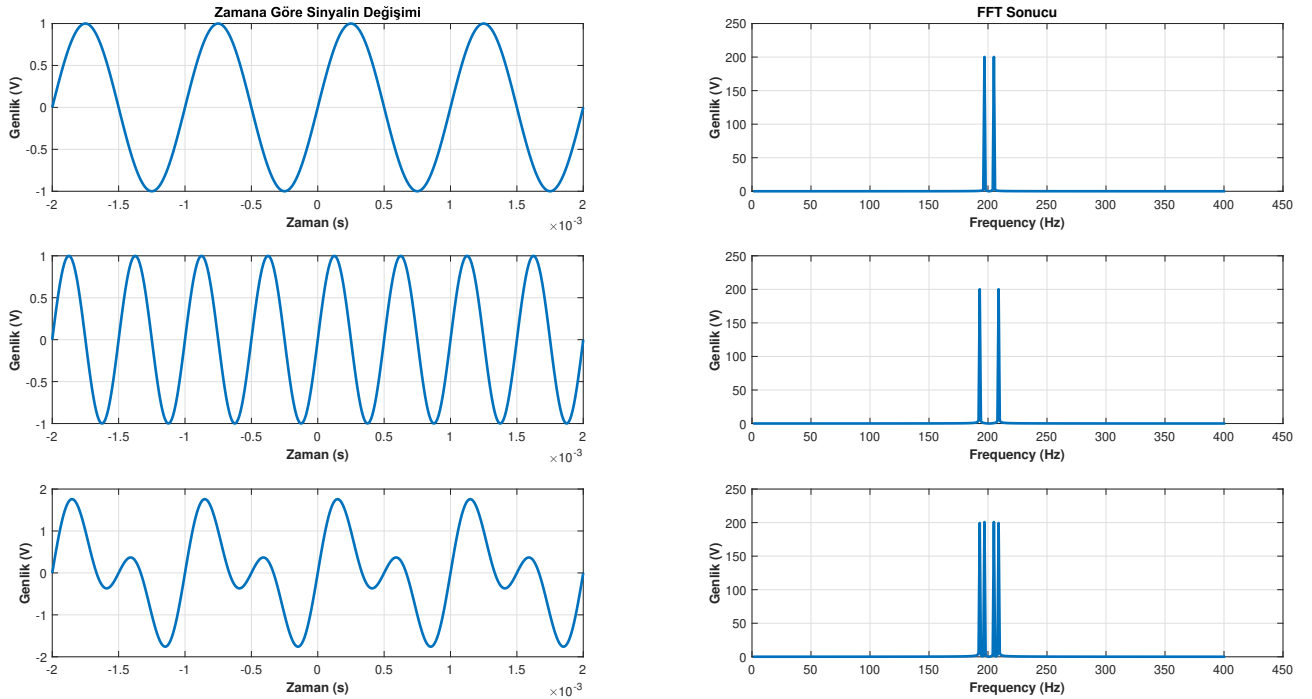
$$X(f) = \int_{-\infty}^{\infty} x(t).e^{-j.2.\pi.f.t}.dt \quad (2.5)$$

Eşitlik 2.4 incelendiğinde zaman düzleminden frekans düzlemine bir geçiş yapıldığı görülmektedir. Fourier dönüşümü, işaretin genlik, frekans ve fazına ait bilgileri elde etmede kullanılır. Fourier dönüşümü işaretin $-\infty$ ile $+\infty$ arasında değiştiğini varsayan bir matematiksel yaklaşımdır, ancak pratikte kullanılan Fourier dönüşümü $-\infty$ ile $+\infty$ arasında değil belli bir zaman aralığında değişmektedir. Sürekli zamanda

tanımlı olan Fouier dönüşümü, ayrık zamanda ayrık zamanlı Fourier dönüşümü olarak uygulanmaktadır. Ayrık zamanlı Fourier dönüşümünün algoritmik karmaşıklığı $\mathcal{O}(n^2)$ büyük boyutlu verilerde sorun oluşturduğu için, algoritma karmaşıklığı $\mathcal{O}(n \log n)$ olan Hızlı Fourier Dönüşümü (Fast Fourier Transform - FFT) algoritması tanımlanmıştır. Hızlı Fourier dönüşümü işaretten zaman bilgisi dışında frekans, faz gibi işaret hakkında fazladan bilgiler edinilmesini sağlayan bir algoritmadır. Ancak hızlı Fourier dönüşümü ile frekans bilgisi elde edilirken zaman bilgisinin kaybedilmesi sorunu söz konusudur. Bu sorunun çözümüne yönelik, hem frekans hem de zaman bilgisini bir arada görmek için kullanılan zaman-frekans analizi ilerleyen bölümlerde açıklanmıştır [3, 21, 31, 32].

2.4.2 Periyodik İşaretlerin Analizi - Fourier Serileri

Belli bir zaman periyodu (T) boyunca kendisini tekrarlayan, bir diğer ifadeyle her bir t değeri için $x(t) = x(t + T)$ şartını sağlayan işaretlere periyodik işaretler denir. Şekil 2.3 bazı periyodik işaretleri göstermektedir.



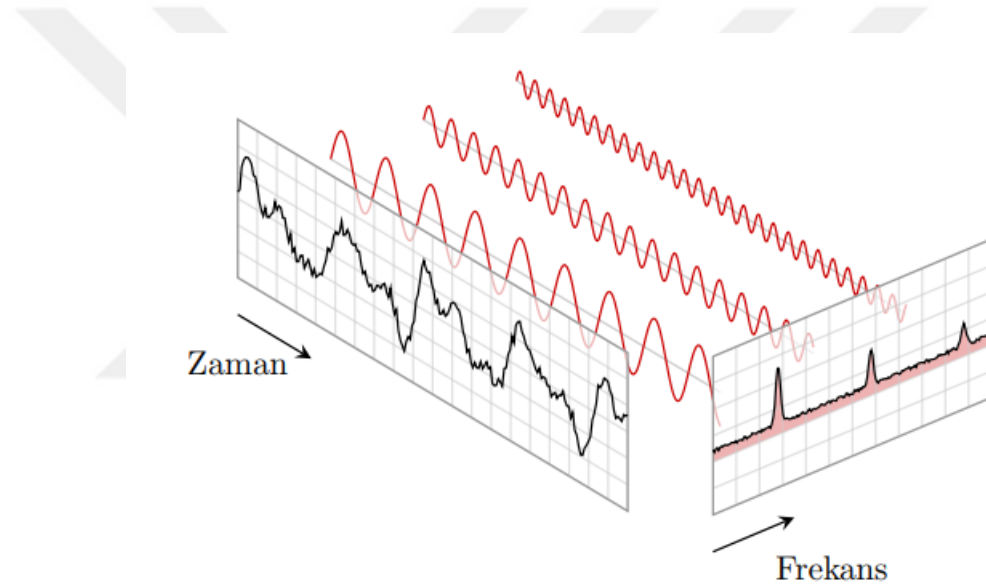
Şekil 2.3 Periyodik sinüs işareti örnekleri

Periyodik işaretler kendisini belli periyotlarla tekrar eder ve dolayısıyla $-\infty$ ve $+\infty$ arasındaki her periyotta işaretin özellikleri aynıdır. Böyle bir sinüsoidal işareti bir periyot boyunca analiz etmek işaret hakkında tüm bilgileri elde etmek için yeterlidir.

Bir periyot boyunca işareti analiz etmeyi sağlayan Fourier serilerinin hesaplanması Eşitlik 2.6 ile gösterilmektedir [3, 21, 31, 32].

$$f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} \{a_n \cos nx + b_n \sin nx\} \quad (2.6)$$

Fourier serileri ile periyodik bir işarettaki sinüs ve kosinüs bileşenlerinin Genlik, Frekans ve Faz değerleri kolaylıkla bulunabilmektedir. İşaretin çok sayıda sinüs ve kosinüs bileşeninden meydana gelmesi işaretin daha karmaşık bir yapıda olduğunu göstermektedir [3, 21, 31, 32]. Şekil 2.4 bir işaretin sinüs ve kosinüs bileşenlerine ayrılmasını ve bu bileşenlerdeki her bir sinüsoidal bileşenin frekans ve genlik bilgisinin elde edilmesini göstermektedir.



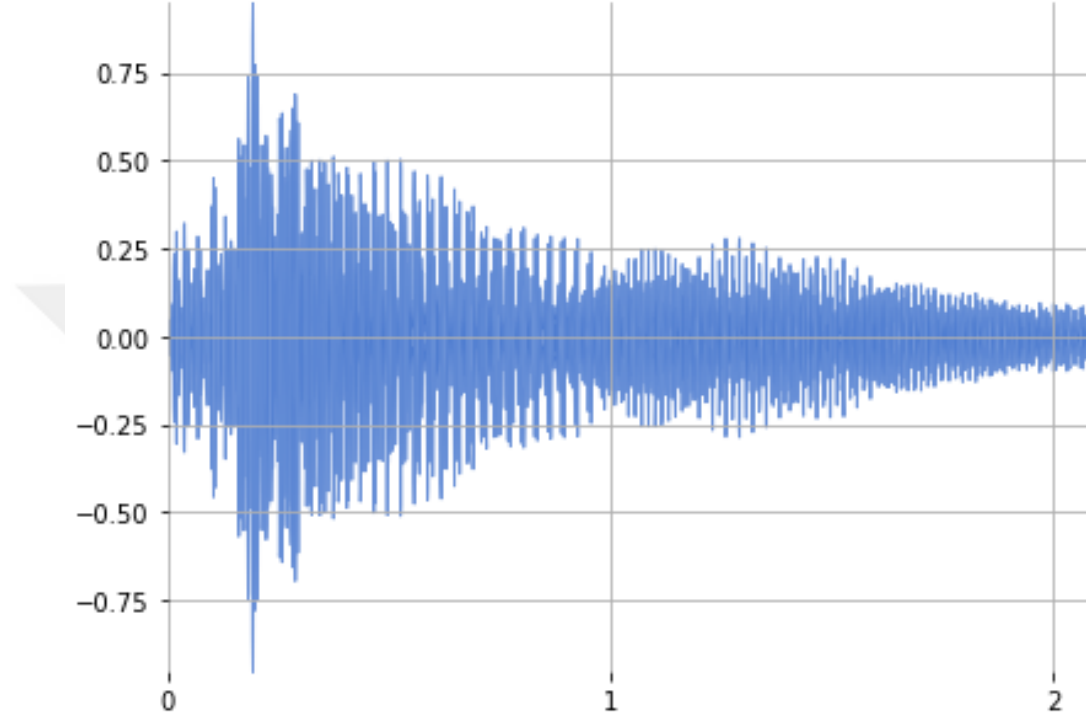
Şekil 2.4 Bir işaretin bileşenlerine ayrıştırılması

Günlük hayatta kullanılan siren, alarm gibi bazı sesler periyodik olduğu için bu yöntemle doğru bir şekilde analiz edilebilir. Ancak bir müzik veya kalp ritmi işareti ele alındığında, bu işaretler periyodik olmadığı için Fourier serileri kullanmak uygun değildir [3, 21, 31, 32].

2.4.3 Periyodik Olmayan İşaretlerin Analizi

Periyodik olmayan işaretler, işaretin kendisini belirli bir zaman periyodunda tekrarlamadığı veya diğer ifadeyle her bir t değeri için $x(t) = x(t + T)$ şartını sağlamayan işaretlerdir. Şekil 2.5 bir bateriden elde edilen ve periyodik olmayan bir işareti göstermektedir.

Bu tür işaretlerin hızlı Fourier dönüşümü ile analiz edilmesi sonucunda frekans bilgileri elde edilir. Ancak, işaret periyodik olmadığı ve sürekli değişken bir yapıda olduğu için frekans bilgisi elde edilirken zaman bilgisi kaybedilir. Bir müziği örnek verecek olursak, müzikteki 5. saniyedeki frekans bileşenleri ile 50. saniyedeki frekans bileşenleri bulunmak istendiğinde, müziğin tüm frekans bilgileri elde edilecek ancak bunların hangi zamanlara ait oldukları bulunamayacaktır [3, 21, 31, 32].



Şekil 2.5 Periyodik olmayan işaret örneği

2.4.4 Kısa Zamanlı Fourier Dönüşümü - STFT

İşarete hızlı Fourier dönüşümü uygulandığında, işaretin hem zaman düzleminde hem de frekans düzleminde değişimini incelemek için önerilen çözümlerden birisi Pencerelemiş Fourier Dönüşümü (Windowed Fourier transform) olarak da bilinen Kısa Zamanlı Fourier dönüşümüdür (Short Time Fourier Transform - STFT). İlk kez 1989 yılında Cohen tarafından önerilmiştir [33]. Bu yöntemle göre, işaret pencere adı verilen daha küçük parçalara ayrılarak analiz edilir. Bu sayede hangi zaman aralığında hangi sinüs ve kosinüs bileşenlerinin olduğu, diğer bir ifadeyle işarete hangi frekans bileşenlerinin olduğu belirlenebilir. Bu dönüşüm yoluyla bir işaretin hem frekans bileşenlerinin hem de genlik değerlerinin zamana bağlı nasıl değiştiği görülebilir. Zaman değişkeni t olan ve $x(t)$ olarak tanımlı bir işaretin $g(t)$ pencere

fonksiyonu kullanılarak STFT dönüşümü aşağıdaki şekilde bulunur.

$$X(f, s) = \int_{-\infty}^{\infty} x(t).g(t-s).e^{-j.2.\pi.f.t} dt \quad (2.7)$$

STFT uygulanırken temel olarak kullanılan üç parametre bulunmaktadır. Bunlar aşağıda kısaca özetlenmiştir [3, 21, 31, 32].

- **Pencere Boyutu:** İşaretin hangi süre boyunca analiz edileceğini gösteren parametredir. Sayısal işaretler için bunun anlamı Fourier dönüşümü için kullanılacak olan örneklerin sayısıdır. Pencere boyutu seçimi zaman-frekans analizinde elde edilecek olan frekans ve zaman çözünürlüğü sonucunu doğrudan etkilemektedir. Örnekleme frekansının f_s ve pencere boyutunun N olarak belirlendiği bir zaman-frekans analizinde, zaman çözünürlüğü Δt ve frekans çözünürlüğü Δf sırasıyla Eşitlik 2.8 ve Eşitlik 2.9 ile gösterildiği gibi hesaplanır.

$$\Delta t = \frac{N}{f_s} \quad (2.8)$$

$$\Delta f = \frac{f_s}{N} \quad (2.9)$$

Eşitliklerden anlaşılacağı üzere, pencere boyutu frekans ve zaman çözünürlüğünü doğrudan etkilemektedir. Yüksek boyutlu bir pencere boyutu seçildiğinde zaman çözünürlüğü yüksek olurken, frekans çözünürlüğü tersi oranda düşük olmaktadır. Örneğin, örnekleme frekansı 1000 Hz olan bir işarete, 10 saniye boyunca örnekleme yapıldığında 10.000 örnek uzunluklu bir sayı dizisi elde edilir. Eğer pencere uzunluğu 100 örnek olacak şekilde seçilirse, 100 milisaniyelik bir pencere uzunluğu elde edilecektir. Böylece zaman frekans analizinin zaman çözünürlüğü 100 milisaniye, frekans çözünürlüğü de 10 Hz olacaktır.

- **Pencere Fonksiyonu:** Normalde $-\infty$ ile $+\infty$ aralığında analiz edilmesi gereken işaret t_1 ve t_2 zaman aralığında analiz edilmektedir. Bu t_1 ve t_2 anlarında yapay olarak oluşturulan süreksizlik noktaları nedeniyle işarete bozulmalar oluşmaktadır. Bu bozulmaları en aza indirmek için yapılan ön işlemlerden birisi de pencere fonksiyonudur. Fourier dönüşümü uygulanmadan önce pencere fonksiyonu ile işaretin analiz edilecek bölgesi karşılıklı olarak çarpılır ve işaret analiz edilmeye hazırlanmış olur. Literatürde Hamming, Hanning vb farklı pencere fonksiyonları bulunmaktadır. Fourier dönüşümü ve derin

öğrenme yöntemlerinde kullanılan bu pencere fonksiyonlarının başarıma etkisi incelenmiş olup ilerleyen bölümlerde detayları verilmiştir. İşaret analizinde pencere boyutu işaretin özelliklerine göre seçilmelidir. Hızlı Fourier dönüşümü uygularken pencere boyutu 2'nin katları olarak (128, 256, 512, ...) seçilmesi gerekir.

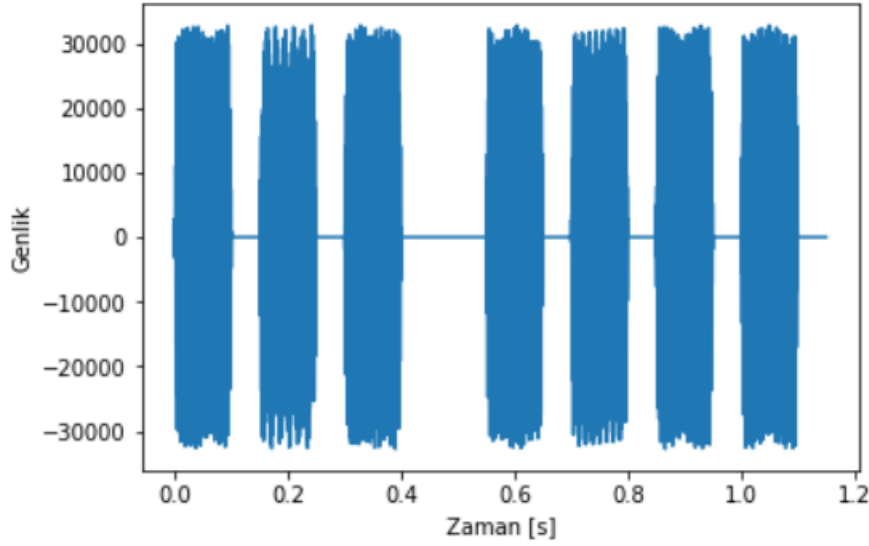
- **Örtüşme (Overlap) Miktarı:** Zaman-frekans analizinde pencere fonksiyonu kullanıldığında, pencereleme sırasında pencere kenarlarına yakın olan kısımlarda var olan işaretlerdeki özellikle kısa süreli bazı bilgilerin kaybolması olasılığı vardır. Pencerelerin ardışık olarak kullanılması sırasında bazı veri kayıpları olabilmektedir. Bu veri kaybını önlemek adına, yeni analiz edilecek pencere bir önceki pencereden bir miktar veriyi de içerecek şekilde analiz gerçekleştirilir. Bu sayede, müzik uygulamalarında notanın başladığı ve bittiği zaman aralığı bu yöntemle daha kolay tespit edilebilmektedir.

2.4.5 Zaman-Frekans Analizi - Spektrogram

Spektrogram, işaretin kıza zamanlı fourier dönüşümünden görselleştirilerek elde edilen bir zaman-frekans analizi yöntemidir [23]. Kısa zamanlı fourier dönüşümünde kullanılan pencere boyutuna göre oluşacak zaman aralıklarının her birinde genlik ve frekans bilgilerinin elde edilmesinde kullanılır. Spektrogram bir eksenini zaman, diğer eksenini frekans olarak elde edildiğinde iki boyutlu bir matris bir diğer ifadeyle görüntü elde edilir. Bu görüntünün her bir pikseli bir zaman aralığındaki tespit edilen frekansların enerji bilgisidir. Diğer taraftan, spektrogram üç boyutlu olarak da gösterilebilir. Bu boyutlar zaman, frekans ve genlik olarak şelale diyagramı (waterfall plot) ile gösterilir. Çalışmamızda spektrogram görüntüleri iki boyutlu olarak verilecektir. Spektrogram, zaman-frekans analizi yöntemlerinden olan STFT'yi görselleştirme yöntemidir. Müzik ve ses işareti uygulamalarında, farklı zaman-frekans analizi yöntemleri kullanılarak elde edilen farklı görselleştirme teknikleri de kullanılmaktadır. Mel-spektrogram, Chromagram bunlardan bazılarıdır.

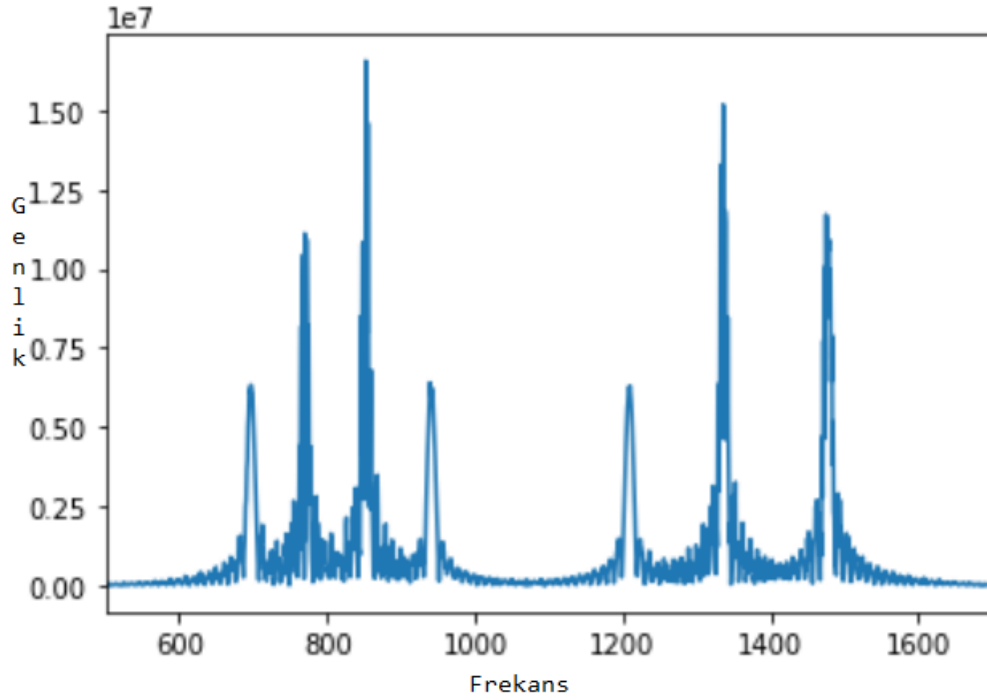
İşareti belli zaman aralıklarında analiz edebilmek için pencere boyutu, pencere fonksiyonu ve örtüşme miktarı belirlenir. Aşağıda 7 haneli bir telefon tuşu sesinin spektrogramının bulunması işlemi adım adım açıklanacaktır.

- **İşaretin Zaman Düzleminde İncelenmesi:** Bu aşamada işaretin zamana bağlı fonksiyonu aşağıdaki şekilde görüldüğü gibi çizilir. İşaret incelendiğinde, işaretin yedi farklı zamanda başlayıp ve bir süre sonra biten bilgiler içerdiği görülmektedir.



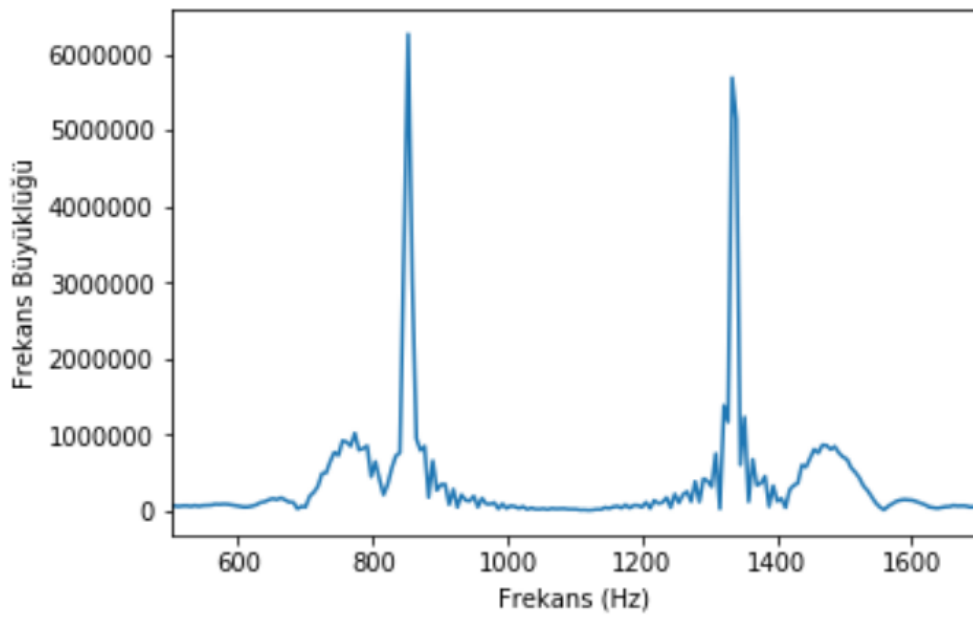
Şekil 2.6 İşaretin zaman düzlemi görüntüsü

- **İşaretin Frekans Düzleminde İncelenmesi:** Zaman düzleminde verilen işaretin hızlı fourier dönüşümü ile frekans düzlemine geçildiğinde işaretin frekans bileşenleri aşağıdaki şekilde gösterilmektedir. Bu şekilden de yedi farklı işarete ait frekans değerleri görülmektedir. Ancak frekans bilgilerinin hangi zamana ait olduğu veya bir diğer ifadeyle her bir tuşa ait frekans bilgisi elde edilememektedir.



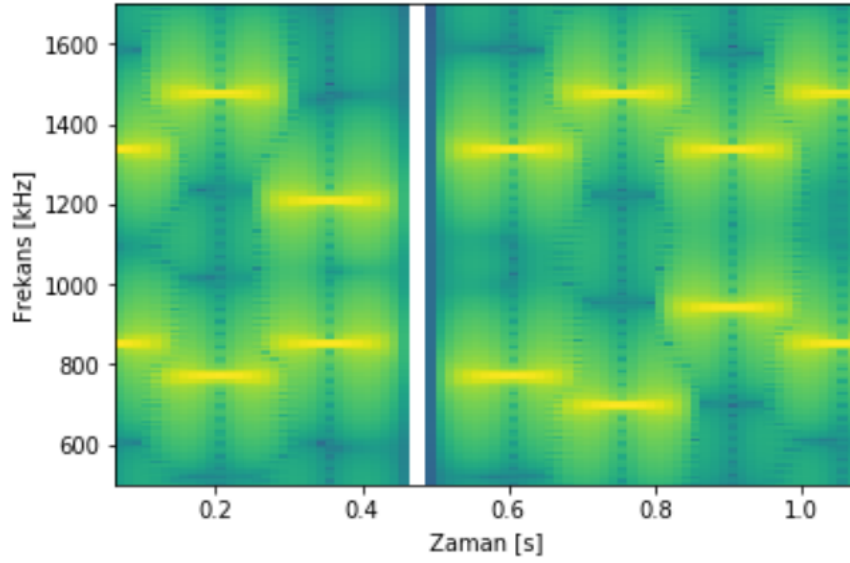
Şekil 2.7 İşaretin tamamının frekans düzlemi görüntüsü

- **İşaretin Zaman-Frekans Düzleminde İncelenmesi:** Önceki maddede verilen işaretin zaman düzleminde frekans düzlemine geçişiyle sadece frekans bilgisi görüldüğü için işaretin bölümlere ayrılarak frekans bilgisinin analiz edilmesi daha yerinde olacaktır. Bu nedenle işaret her bir tuş sesinin süresine göre parçalara ayrıldıktan sonra her bir parçanın Fourier dönüşümü alınır ve her bir tuşa ait frekans bilgisi elde edilir. Telefon tuş sesleri iki sinüsoidal bileşenden oluşacak şekilde standart olarak kodlanmıştır. Bu nedenle aşağıda zaman düzlemindeki ilk tuş olan 8 tuşu için 852 Hz ve 1366 Hz frekanslarında değerler görülmektedir. Bu yapılan işlem tüm tuşlar için tekrarlanarak yapılır ve tuş seslerinin hangi tuşa ait olduğu elde edilir [34].



Şekil 2.8 İlk tuşa ait frekans bilgileri

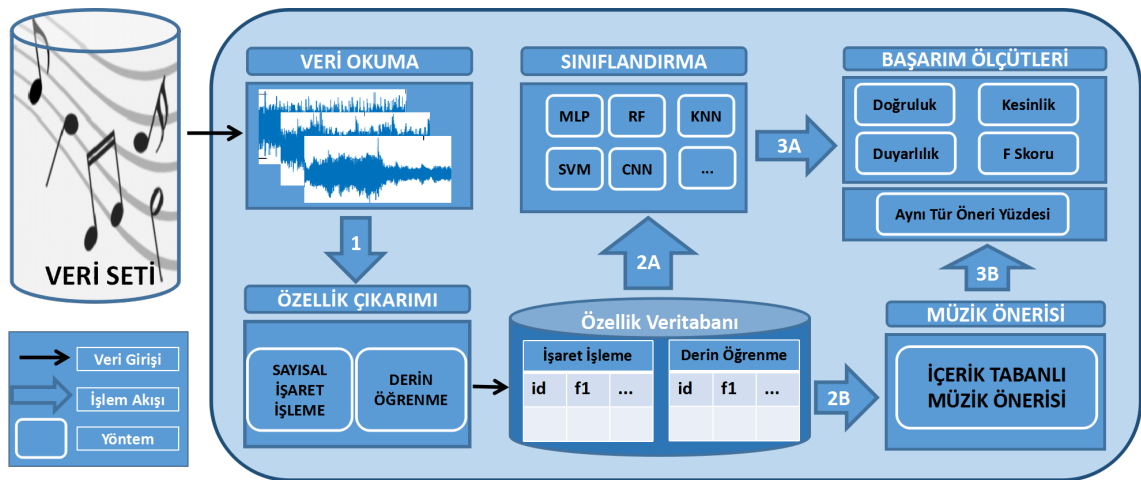
Yukarıda belirtilen prosedürün görselleştirilmiş hali olan spektrogram bu tür işaretlerin analizinde çok yaygın olarak kullanılan bir araçtır. Aşağıdaki spektrogram aynı işaretin analizini göstermektedir.



Şekil 2.9 İşaretin tamamının spektrogram görüntüsü

Önceki bölümlerde açıklanan zaman frekans analizi için kısa zamanlı Fourier dönüşümü ve spektrogram bu çalışmada müziklerin analiz elde edilmesinde kullanılan temel bir yöntemdir. Ses ve müzik işaretlerinden akustik özellik çıkarımı yaparken kullanılan yöntemlerin parametreleri ve girdileri spektrogramdan türetilmiştir. İlerleyen bölümlerde bu yöntemler detaylı olarak açıklanacaktır.

Tez çalışması boyunca bu bölümde açıklanan yöntemlerden yola çıkarak, müzik veri setinden işaret işleme yöntemleri ve derin öğrenme yöntemleri ile özellik çıkarımı yapılacaktır. Bu özelliklerle müzik türünün sınıflandırılması ve müzik önerisi gerçekleştirilecek, elde edilen sonuçlar farklı ölçütlerle karşılaştırılacak ve değerlendirilecektir. Çalışmanın blok şeması Şekil 2.10 ile gösterilmiştir.



Şekil 2.10 Çalışmanın blok şeması

3

AKUSTİK ÖZELLİK ÇIKARIMI

Bu bölümde tez süresince kullanılan özellik çıkarımı yöntemleri açıklanmaktadır. Temel olarak sayısal işaret işleme yöntemleri zaman düzlemi ve frekans düzlemi olarak ikiye ayrılmıştır.

3.1 İstatistiksel Tanımlayıcılar

Bu bölümde istatistiksel özellik çıkarımı yöntemlerinden bahsedilecektir.

3.1.1 Aritmetik Ortalama

Aritmetik ortalama ($\bar{x}$) bir veri dizisindeki değerlerin toplamının, dizinin eleman sayısına (N) bölümüyle hesaplanır. Verilerin dağılımının normal dağılıma uygun olduğu durumlarda kullanılması uygundur.

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i \quad (3.1)$$

3.1.2 Standart Sapma

Standart sapma (σ) bir veri dizisindeki değerlerin ortalamadan ne kadar saptığını belirlemede kullanılan bir tanımlayıcıdır.

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2} \quad (3.2)$$

3.1.3 Ortanca - Medyan

Sıralı bir veri dizisindeki en ortadaki elemanın değerini bulmaya yarayan ölçüttür. Gürültü veya aykırı değer bulduğu verilerde dağılım hakkında aritmetik

ortalamaya göre daha kesin bilgi verir. Verilerin dağılımının normal dağılıma uygun olmadığı durumlarda kullanılması aritmetik ortalamaya göre daha uygundur.

3.1.4 Tepe Değer - Mod

Bir veri dizisindeki en fazla görülen sayı olarak ifade edilir

3.1.5 Çarpıklık - Skewness

Bir veri dizisindeki verilerin ortalamaya göre dağılımının yönünü belirler. Veri dizisinde aritmetik ortalama > medyan > mod durumu varsa, bu veri dizisi sağdan çarpık veya pozitif çarpıklığa sahiptir; diğer taraftan aritmetik ortalama < medyan < mod durumu varsa, bu verisi dizisi soldan çarpık veya negatif çarpıklığa sahiptir.

$$\text{Çarpıklık} = \sqrt{N} \times \frac{\sum_{i=1}^N (X_i - \bar{X})^3}{(\sum_{i=1}^N (X_i - \bar{X})^2)^{3/2}} \quad (3.3)$$

3.1.6 Basıklık - Kurtosis

Bir veri dizisindeki verilerin dağılımının basıklığı veya sivriliği hakkında bilgi veren istatistiksel tanımlayıcıdır.

$$\text{Basıklık} = N \times \frac{\sum_{i=1}^N (X_i - \bar{X})^4}{(\sum_{i=1}^N (X_i - \bar{X})^2)^2} \quad (3.4)$$

3.1.7 Entropi

Bir veri dizisindeki verilerin karmaşıklığı ve saflığı hakkında bilgi veren ölçüttür. Veri dağılımının tekdüze (uniform) olup olmadığını belirlemede sıklıkla kullanılır. Aşağıdaki eşitlikle hesaplanır. Buradaki p_i her bir verinin veri dizisindeki bulunma olasılığını ifade eder.

$$E(S) = \sum_{i=1}^c -p_i \log_2 p_i \quad (3.5)$$

3.2 Zaman Düzleminde Özellik Çıkarımı

3.2.1 Ortalama Enerji - Root Mean Square Energy (RMSE)

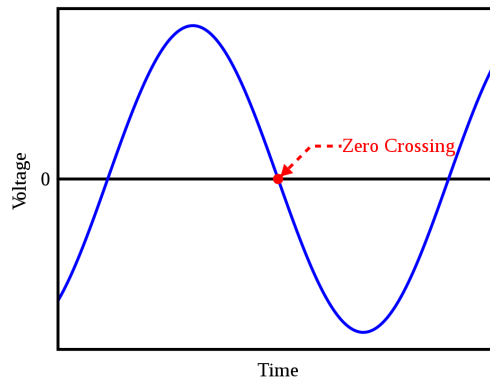
Ortalama enerji sesin belli bir zaman aralığındaki genliklerinin karelerinin toplamının karekökü ile hesaplanır. Bu yöntem farklı şekillerde frekans düzleminde de kullanılabilir. Özellikle iki ses arası karşılaştırma yaparken gürültüye göre bir ayırım yapmakta kullanılan yöntemdir. Eşitlik 3.6 ile hesaplanır.

$$RMS = \sqrt{\frac{1}{N} \sum_t^N x(t)^2} \quad (3.6)$$

3.2.2 Sıfır Geçiş Oranı - Zero Crossing Rate

Sıfır geçiş oranı, bir işaretin toplam sıfır geçişlerinin T pencere boyutuna oranıdır. Bu oran bize işaretin frekans bilgisini verir. Frekans, işaret değişikliği ile doğru orantılıdır. Daha fazla değişiklik frekansın yüksek olduğunu ifade ederken; daha az değişiklik frekansın düşük olduğunu ifade eder. Bu yöntem işaret ve ses işlemede değişim bilgisinin yorumlanmasında kullanılan basit ve etkili bir yöntemdir [35]. Eşitlik 3.7 ile gösterildiği gibi hesaplanır. Eşitlikteki S değişkeni, T uzunluklu bir işareti temsil etmektedir.

$$ZCR = \frac{1}{T-1} \sum_{t=1}^{T-1} f(s_t s_{t-1} < 0) \quad (3.7)$$



Şekil 3.1 Sıfır geçiş oranı - Zero Crossing Rate [36]

3.3 Frekans Düzleminde Özellik Çıkarımı

Bu bölümde müzik verilerinden frekans düzlemi kullanılarak özellik çıkarımı için kullanılan yöntemler açıklanmıştır.

3.3.1 Ortalama Frekans Enerjisi - Root Mean Square Energy (RMSE)

Bir işaretin enerjisi, işaretin genliklerinin karelerinin toplamı ile, aşağıdaki formüldeki gösterildiği gibi hesaplanır [31, 37].

$$\sum_f^N |x(f)|^2 \quad (3.8)$$

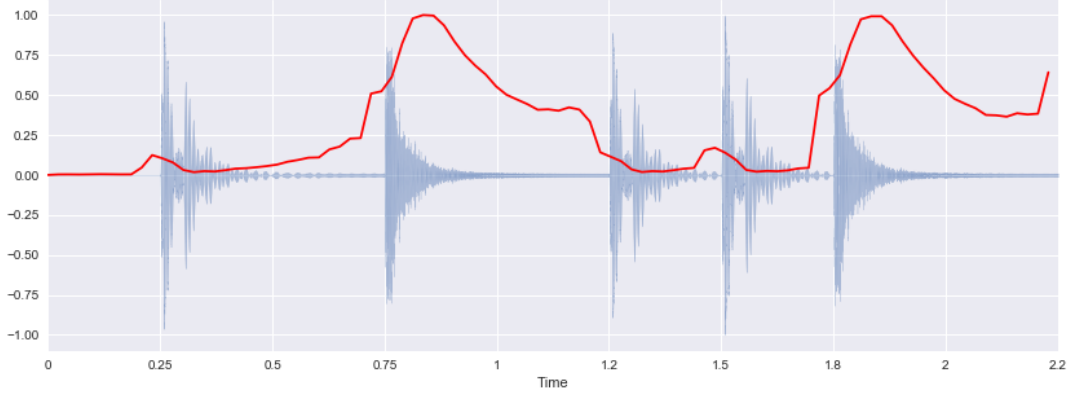
RMSE ise, bir işaretin enerjisinin ortalamasının kareköküdür ve aşağıdaki şekilde tanımlanır [31, 37].

$$\sqrt{\frac{1}{N} \sum_f^N |x(f)|^2} \quad (3.9)$$

3.3.2 Spektral Ağırlık Merkezi - Spectral Centroid

Spektral ağırlık merkezi, genellikle bir sesin belirginliği veya keskinliği (brightness) ile ilişkilidir [38]. İşaretin frekansı hakkında bilgi verir. Tını üzerinde çalışan araştırmacılara göre, sesler arasındaki algısal olarak en güçlü ayrımlardan biri sesin keskinliğidir [39]. Keskinlik, bir sesteki yüksek frekanslı içerik miktarı ile ilgilidir. Spektral ağırlık merkezi spektrumun ağırlık merkezini gösterir ve Fourier dönüşümünün frekans ve genlik değerleri kullanılarak elde edilir. Bu, frekansların ağırlıklı ortalaması olarak hesaplanır. Spektral ağırlık merkezi, bir spektrumun enerjisinin hangi frekansta ortalandığını ifade eder. Spektral ağırlık merkezi f_c ; k numaralı frekans kutusundaki spektral genlik $S(k)$ ve frekans $f(k)$ olmak üzere Eşitlik 3.10'deki gibi hesaplanır [3].

$$f_c = \frac{\sum_k S(k)f(k)}{\sum_k S(k)} \quad (3.10)$$



Şekil 3.2 Spektral ağırlık merkezi [40]

Sıfır Geçme Oranı yöntemine benzer olarak işaretin başlarında spektral ağırlık merkezi değerinin olmaması gerektiği şekilde yüksek bantlarda olduğu gibi görülmektedir. Bu durumun sebebi olarak işaretin başlarındaki düşük genlik değerine sahip bileşenler gösterilebilir. Sabit küçük bir değer eklenilerek bu sorunun üstesinden gelinebilir [40].

3.3.2.1 Spektral Düzlük - Spectral Flatness

Spektral düzlük, işaretin beyaz gürültüye olan yakınlığının ölçüsüdür. Beyaz gürültünün güç spektrumu düz bir şekildedir. Spektral düzlük işaretin güç spektrumunun geometrik ortalamasının aritmetik ortalamasına bölünmesi ile elde edilir [41].

$$\text{Spektral Düzlük} = \frac{\sqrt[N]{\prod_{n=1}^{N-1} x(n)}}{\frac{\sum_{n=1}^{N-1} x(n)}{N}} \quad (3.11)$$

Spektral düzlük değeri 0 ile 1 arasında değer alabilmektedir. Bunun nedeni geometrik ortalamanın aritmetik ortalamadan büyük olmamasıdır. Spektral düzlük değerinin 1'e yakın olması, işaretin beyaz gürültüye daha yakın olması demektir.

3.3.3 Spektral Karşıtlık - Spectral Contrast

Spektral karşıtlık, işaretin bağıl spektral özelliklerini temsil eder ve spektrumdaki zirveleri, vadileri ve aralarındaki farkı dikkate alır. Buna desibel farkı da denilmektedir. Çoğu müzik için güçlü spektral zirveler harmonik bileşenler olarak nitelendirilirken, vadiler genellikle gürültü ile nitelendirilir. Bazı çalışmalar, spektral

karşıtlığın müzik türlerini ayırmak açısından iyi sonuçlar verdiğini göstermiştir [42].

$$Zirve_k = \log \left\{ \frac{1}{aN} \sum_{i=1}^{aN} x_{k,i} \right\} \quad (3.12)$$

$$Vadi_k = \log \left\{ \frac{1}{aN} \sum_{i=1}^{aN} x_{k,N-i+1} \right\} \quad (3.13)$$

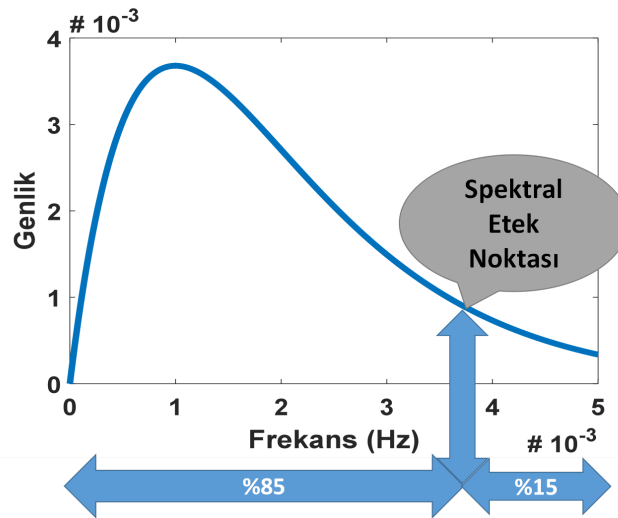
$$Karşıtlık_k = Zirve_k - Vadi_k \quad (3.14)$$

Eşitlikler 3.12, 3.12 ve 3.12'deki k FFT vektörünün alt bant bilgisini temsil ederken $\{x_{k,1}, x_{k,2}, \dots, x_{k,N}\}$; N ise k . banttaki toplam sayıyı göstermektedir. α sabit sayı değeridir.

3.3.4 Spektral Etek - Spectral Roll-off

Spektral etek, toplam spektral enerjinin belirli bir kısmının (örneğin %85) altında bulunduğu frekansı ifade eder. Sözsüz müziklerde, enerjinin büyük bir kısmı spektrumun yüksek frekanslarında bulunur. Böylelikle sözlü ve sözsüz müzikleri birbirinden ayırt etmek için bu özellik kullanılabilir [3, 43].

$$\sum_{n=1}^{R_i} M_i[n] = 0.85 \sum_{n=1}^N M_i[n] \quad (3.15)$$



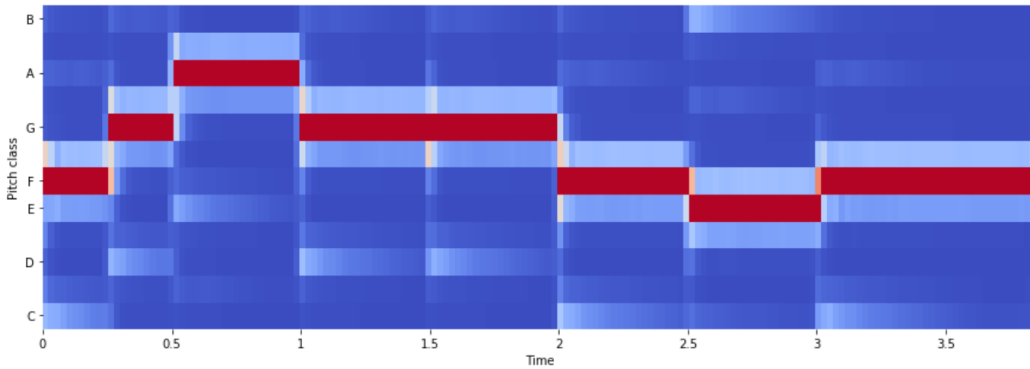
Şekil 3.3 Spektral etek

3.3.5 Polinom Özellikleri - Polynomial Features

Bu yöntemle işaretin spektrumu için istenilen dereceden bir polinom üretilir ve bu polinomun katsayıları geri döndürülür [44]. Polinomu üretmek için, istenilen dereceye göre en uygun enterpolasyon yöntemi kullanılır. Elde edilen polinom katsayıları özellik vektörü olarak kullanılabilir.

3.3.6 Chroma STFT

Chroma özellikleri tüm spektrumu, her biri oktavdaki bir yarım tona denk gelen 12 farklı kutuya bölerek müziği etkili bir biçimde temsil edebilen bir özellik çıkarma ve zaman-frekans analizi yöntemidir. İşareti herhangi bir sayısal işaret olarak görmekten daha çok, işaretin bir alt kümesi olan müzik işareti üzerinde analiz gerçekleştirir. Müzikte aralarında tam 1 oktav fark olan iki ton, insanlar tarafından çok benzer olarak algılandıklarından; müziğin chroma dağılımını bilmek, müzik hakkında işe yarar bilgiler verebilir [31, 45]. Bu vektörü elde etmek için, Chroma STFT’de öncelikle işareten elde edilen frekans spektrumunda frekans eksenini parçalara bölünerek her bir parçanın bir ses sınıfını temsil etmesi sağlanır. Bu işlem yapıldıktan sonra frekans eksenini logaritmik ayrımlı olarak MIDI ses sınıflarını temsil edecek şekilde doğrusal olarak etiketlenir. Etiketlendirme işlemi yapıldıktan sonra işaretin her bir penceresi (frame) için aynı ses sınıfına ait katsayılar birbirleriyle toplanarak işaretin her bir ses sınıfı için ne kadar enerjiye sahip olduğu bulunur. Chroma STFT, müziğin melodik özelliklerinin elde edilmesinde sıklıkla kullanılmaktadır. Şekil 3.4 örnek bir chromagram görüntüsünü göstermektedir. [31, 46].

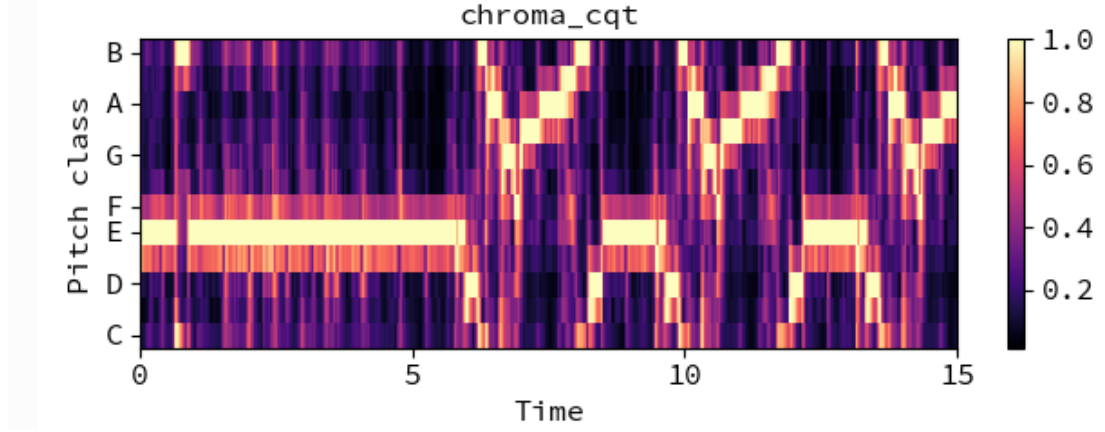


Şekil 3.4 Chroma STFT [31, 45]

3.3.7 Chroma Sabit Q Dönüşümü - Chroma CQT

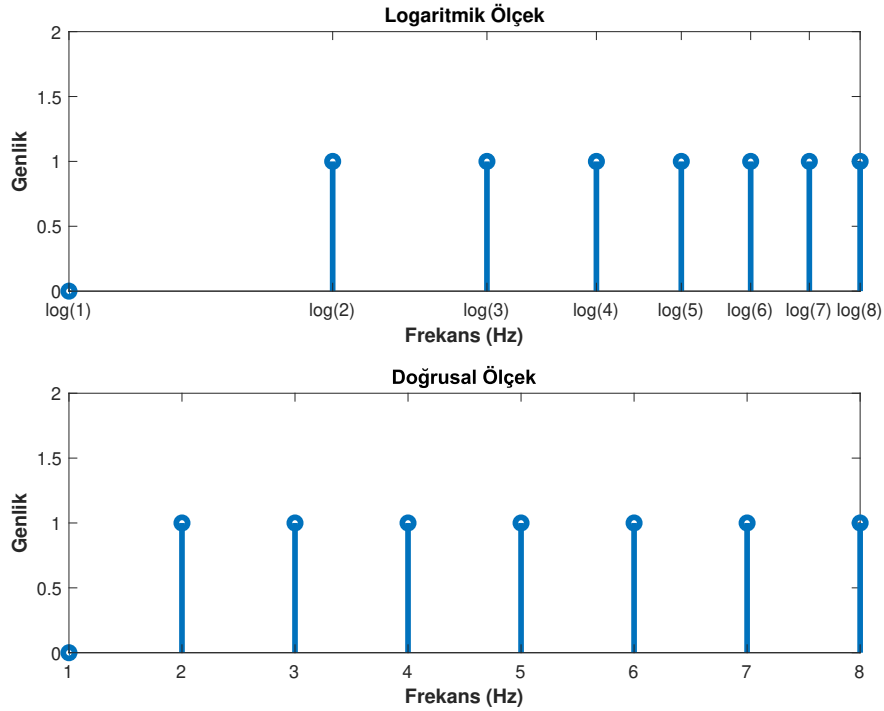
Chroma vektörü, 12 farklı perde(pitch) sınıfının enerji seviyelerini tutan vektördür. Bu vektörün elde edilmesinde farklı yöntemler kullanılabilir. Bu yöntemlerden birisi de CQT’dir. CQT yönteminde STFT’ye benzer şekilde zaman düzleminde

frekans düzlemine geçişi sağlanır. Ancak CQT’de frekans eksenini logaritmik olarak ölçeklenir. Bu da insan kulağının duyma şekline çok daha yakındır. Ses birçok harmonik frekansın bir araya gelmesiyle oluşur. Logaritmik ölçeklemede bu frekansların konumları farklı olsa da birbirlerine göre konumları sabittir. Bu da enstrümanların ayırt edilmesi için önemli bir özelliktir [47]. Chroma CQT yöntemiyle elde edilen bir chromagram Şekil 3.5 ile gösterilmiştir.



Şekil 3.5 Chroma CQT [48]

Şekil 3.6 frekans ekseninin doğrusal ve logaritmik olarak bölümlenmesini örnek olarak göstermektedir.



Şekil 3.6 Frekans ekseninin doğrusal ve logaritmik olarak ayrılması

3.3.8 Tonnetz

Tonnetz ilk olarak Alman matematikçi Euler tarafından önerilmiştir. Daha sonra müzik teorikiler bu yöntemi geliştirmişlerdir. Tonnetz müzikte harmonik ilişkiyi göstermek için kullanılır. Yöntemin kullanılabilmesi için öncelikle Chroma vektörünün elde edilmiş olması gerekmektedir. Tonnetz yönteminin çıktısı, tonal ağırlık merkezleridir. Bir tonal ağırlık merkezi Eşitlik 3.16, 3.17 ve 3.18'deki gibi hesaplanır [49].

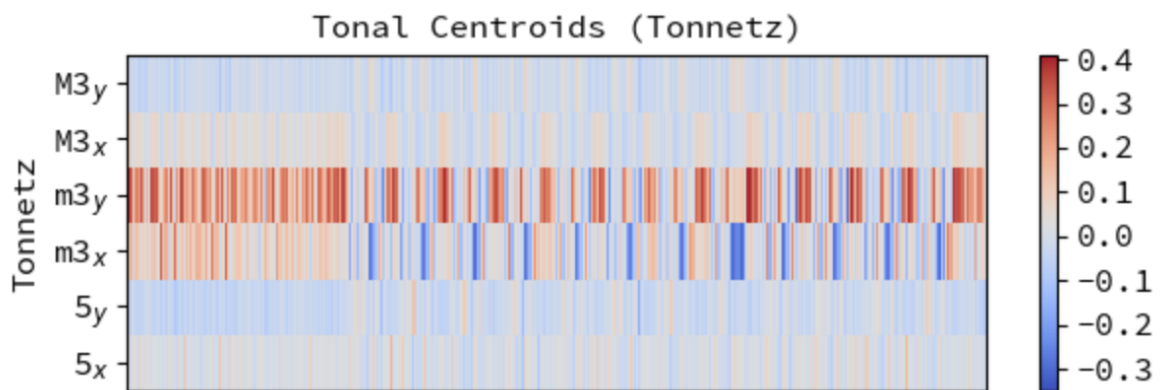
$$T(d) = \frac{1}{\sum_{k=0}^L |C(k)|} \times \sum_{k=0}^L \Phi(d, k) \times C(k) \quad (3.16)$$

$$\Phi = [o_0, o_1, o_2, \dots, o_{L-1}] \quad (3.17)$$

$$o_k = \begin{bmatrix} r1 \sin(k \frac{7\pi}{6}) \\ r1 \cos(k \frac{7\pi}{6}) \\ r2 \sin(k \frac{3\pi}{2}) \\ r2 \cos(k \frac{3\pi}{2}) \\ r3 \sin(k \frac{2\pi}{3}) \\ r3 \cos(k \frac{2\pi}{3}) \end{bmatrix} \quad (3.18)$$

$1 \leq d \leq 6$ ve $0 \leq k < L$

Şekil 3.7 ile örnek bir tonnetz ortalama vektörü görüntülenmektedir.



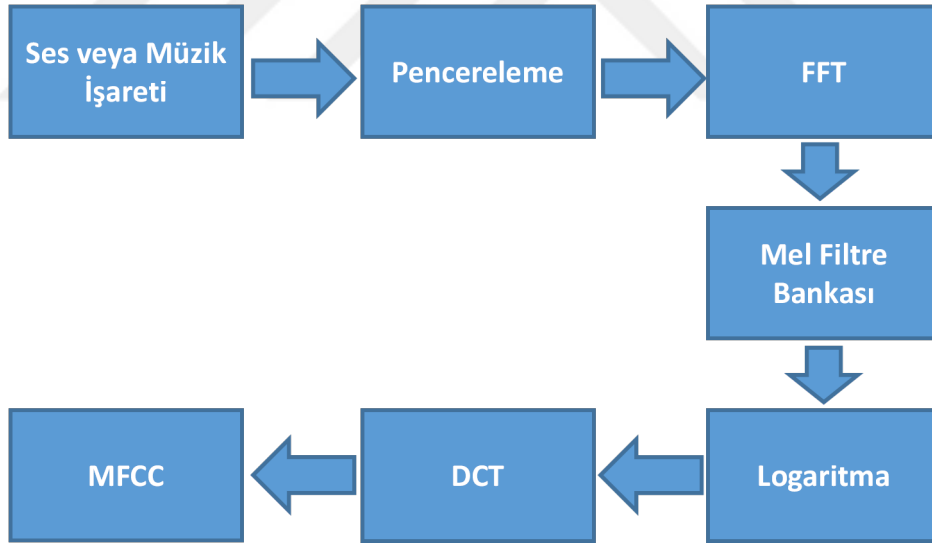
Şekil 3.7 Örnek Tonnetz

3.3.9 Mel-Frekanslı Kepstral Katsayıları - MFCC

Ses işlemede; Mel-frekanslı kepstrumu (MFC), sesin kısa zamanlı güç spektrumunu temsil eder. Mel-frekanslı kepstral katsayıları (MFCC) kollektif olarak bir MFC meydana getirirler. Kepstrum ve Mel-frekanslı kepstrumu arasındaki fark, MFC’de frekans bantları Mel ölçeğine göre eşit aralıklardadır. Mel ölçeği, insan ses algısını doğrusal ölçeğe kıyasla çok daha iyi modellemektedir [2].

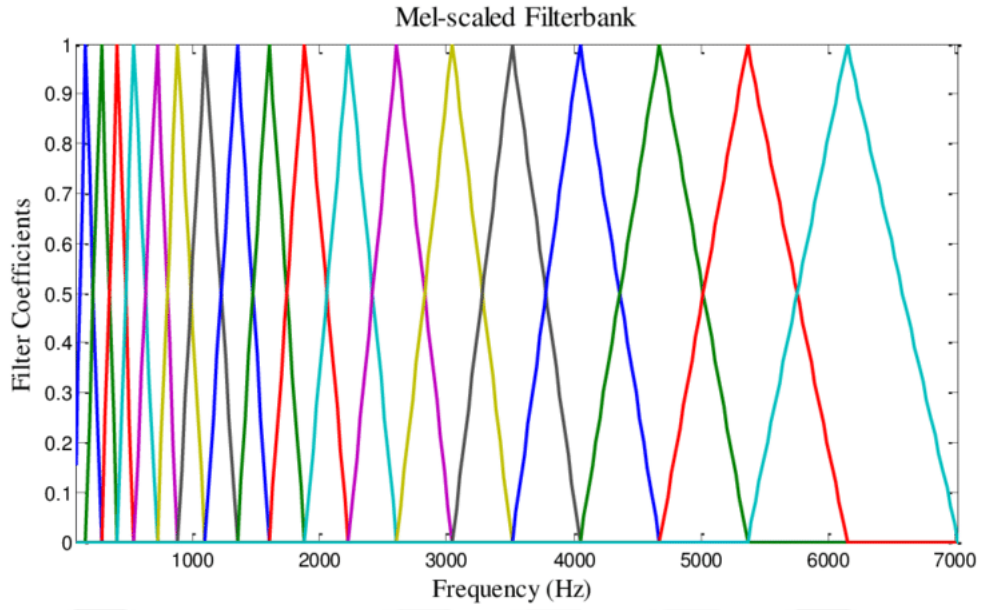
MFCC şu adımların uygulanmasıyla elde edilir:

- İşaretin Fourier dönüşümü alınır.
- Elde edilen spektrumun katsayıları Mel ölçeğine eşleştirilir.
- Her bir Mel frekansındaki güç katsayılarının logaritması alınır.
- Elde edilen katsayılar listesine, bir işaretmiş gibi ayrık kosinüs dönüşümü (DCT) uygulanır.
- Sonuçta ortaya çıkan spektrumun katsayıları MFCC’yi verir.



Şekil 3.8 MFCC İşlem Adımları

Şekil 3.9, MFCC algoritmasında kullanılan Mel Filtre bankasını görsel olarak örneklendirmektedir.



Şekil 3.9 MFCC Mel filtre bankası [50, 51]

MFCC'yi elde edebilmek için öncelikle ses işareti üzerinde çerçeveleme işlemi ile bölütlere ayrılır. Ardından her bir çerçevenin başlangıç ve sonu arasında süreksizliği engellemek için pencereleme işlemi gerçekleştirilir. En çok kullanılan pencereleme yöntemlerinden birisi Hamming pencerelemesidir. [52].

$$W_n(m) = 0.54 - 0.46 \times \cos(2\pi m / (N_m - 1)) \quad (3.19)$$

Pencereleme işleminden sonra FFT uygulanır. Daha sonra işaretin frekansları mel frekansına dönüştürülür. Bu işlem aşağıdaki Eşitlik 3.20 [52].

$$mel(f) = 2595 \times \log_{10}\left(1 + \frac{f}{700}\right) \quad (3.20)$$

1 kHz e kadar olan frekanslar için mel frekansları doğrusal olarak ölçeklenir. Ancak 1 kHz den yüksek frekanslarda logaritmik aralıklar kullanılır [52].

En son olarak ayrık kosinüs dönüşümü (DCT) uygulanarak mel frekansı keprum katsayıları hesaplanmış olur [52].

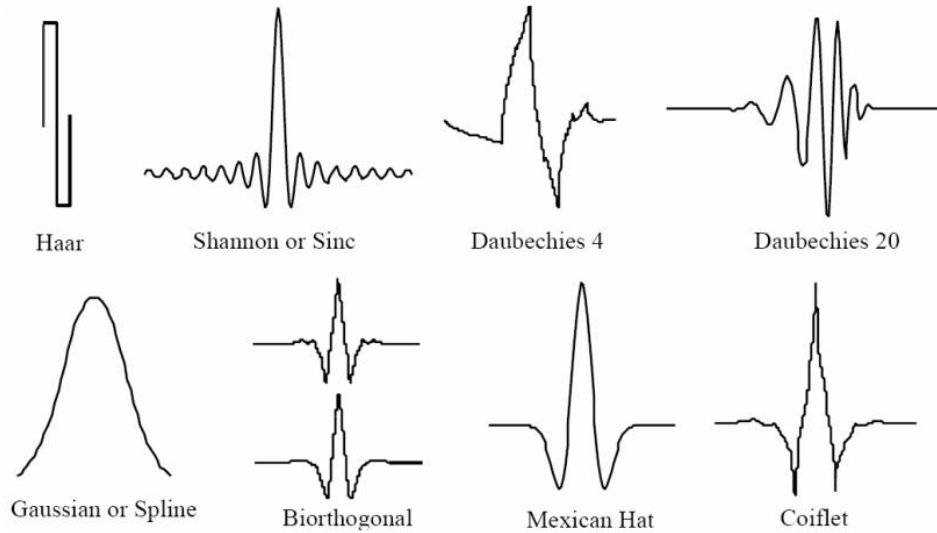
$$C_n = \sum_{k=1}^k (\log D_k) \cos \left[mel\left(k - \frac{1}{2}\right) \frac{\pi}{k} \right] \quad (3.21)$$

3.4 Dalgacık Dönüşümü İle Özellik Çıkarımı - Wavelet Transform

İşaretin belirli bir zamanında ani bir değişiklik olduğu durumda, hızlı Fourier dönüşümü kullanmak işaret analizinde yetersiz kalmaktadır. Çünkü, değişiklik ne kadar ani olursa, bu değişikliğin zaman-frekans analizi için kosinüs ve sinüs bileşenleri o kadar çok kullanılır. Ani değişimler içermeyen sinüs ve kosinüs fonksiyonları kullanmak yerine, bu ani değişimleri içerisinde bulunduran fonksiyonlarla gerçekleştirilen Dalgacık dönüşümünü kullanmak, işaretin analizinde daha başarılı sonuçlar vermektedir. Dalgacık dönüşümünün genel ifadesi Eşitlik 3.22 ile gösterilmiştir [5, 53–55].

$$C(a, b) = \left(\frac{1}{\sqrt{a}}\right) \int_{-\infty}^{+\infty} X(t) \cdot \Psi\left(\frac{x-b}{a}\right) dt \quad (3.22)$$

Yukarıdaki Eşitlik 3.22 b öteleme miktarını, a ölçekleme katsayısını, Ψ ise ani değişimleri içerisinde barındıran dalgacık fonksiyonunu göstermektedir. Dalgacık fonksiyonu, hızlıca sönümlenen ve ortalaması sıfır olan fonksiyon demektir. Sinüsoidal fonksiyonlardan farklı olarak fonksiyonun her noktasında sürekli değildir. Şekil 3.10 bazı dalgacık fonksiyonlarını göstermektedir.



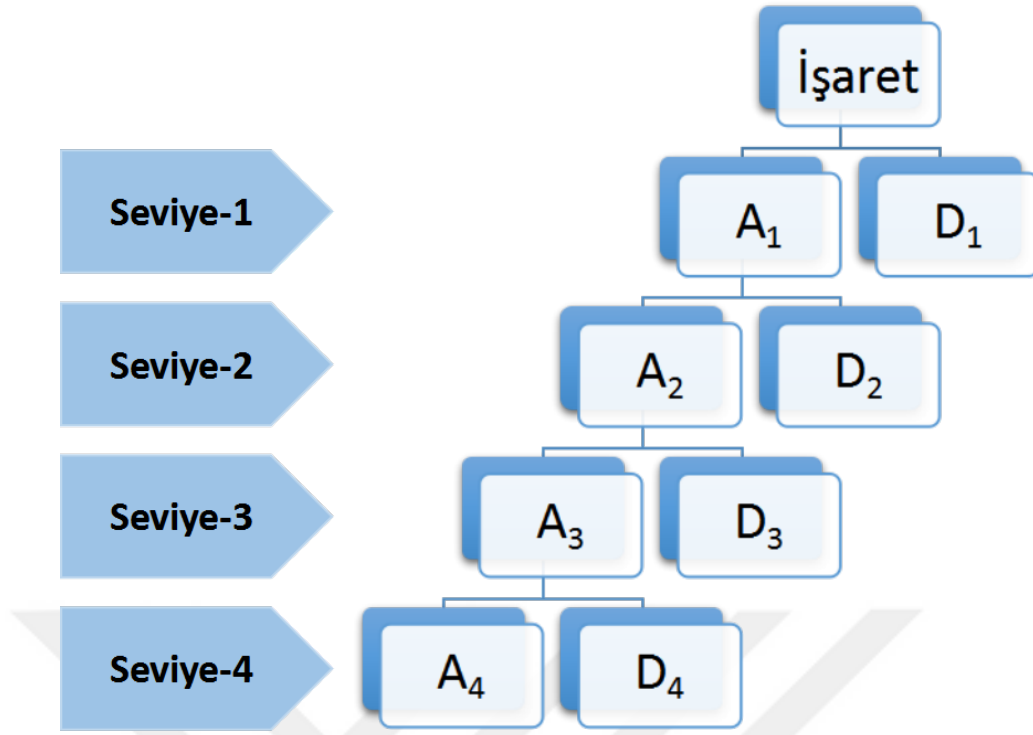
Şekil 3.10 Bazı dalgacık dönüşümü fonksiyonları [56]

Dalgacık dönüşümlerindeki ölçekleme işlemi ise, işareti genişletmek veya kapatmak anlamına gelen bir operasyondur. Örneğin, frekansı 10 Hz olan bir sinüs işareti 2 ile ölçeklendirilirse oluşan işaretin frekansı 5 Hz olur. Dalgacık dönüşümünde kullanılan diğer bir değişken olan kaydırma veya öteleme miktarı, dalgacık dönüşümünü gerçekleştirirken kullanılan dalgacık fonksiyonunu eksen boyunca belirlenen miktara

göre ötelemede kullanılır.

Dalgacık fonksiyonları sürekli ve ayrık olarak ikiye ayrılabilir. Sürekli dalgacık fonksiyonları çekirdek fonksiyonu olarak kullanıldığında sürekli dalgacık dönüşümü gerçekleştirirken, ayrık dalgacık fonksiyonları ayrık dalgacık dönüşümünü gerçekleştirir. Ayrık dalgacık dönüşümü, dalgacık dönüşümünü sürekli zamanda uygulamanın meydana getireceği algoritma karmaşıklığından kaçınmak için geliştirilmiş bir yöntemdir ve işareti daha hızlı bir şekilde analiz etmeyi mümkün kılar. Bu işlem yapılırken işaret belirlenen kriterlere göre parçalara ayrılır. En çok kullanılan ve bilinen yöntem, işareti biri frekans düzleminde diğeri zaman düzleminde olmak üzere iki bölüme ayırmaktır. Örnekleme frekansı f_s ise, işarete bulunan frekans bileşeni 0 ile $f_s/2$ arasındadır. Örnek olarak, 800 örnekleme frekansı ile bir işaret alınırsa, bu işarete bilgi içeren kısım 0 ile 400 arasındadır [5, 29, 53].

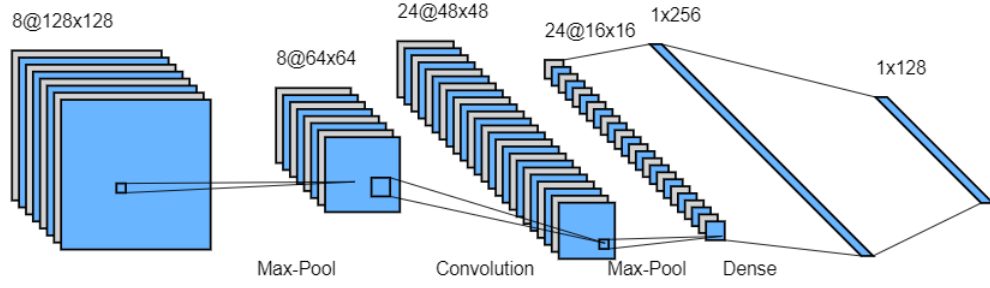
Bu çalışmada ayrık dalgacık dönüşümleri kullanılmıştır. Ayrık dalgacık dönüşümü yapılırken ölçeklendirme için 2 ve 2'nin kuvvetleri kullanılmış, kaydırma yapılırken ise 2'nin kuvvetlerinin katları kullanılmıştır. Bu sayede gereksiz bilgilerin tekrar hesaplanmasının önüne geçilmiştir. Ayrık dalgacık dönüşümü yapılırken ilk adım olarak; işaret, biri düşük geçiren (wavelet function, dalgacık fonksiyonu), diğeri ise yüksek geçiren (scaling function, ölçeklendirme fonksiyonu) iki filtreden geçirilir [57]. Filtreleme sonrası her iki çıktıya da aşağı örnekleme (down sampling) uygulanarak örnek sayısı yarıya düşürülür. Alçak geçiren filtrenin sonucu yaklaşım bileşeni (Approximation), yüksek geçiren filtrenin sonucu ayrıntı bileşeni (Detail) olarak adlandırılır. Daha sonra yaklaşım bileşeni alınır ve bu önceki iki adım istenilen seviye kadar tekrarlanır. Bu adımların ne kadar tekrarlanacağı seviye parametresi ile seçilir. Bu seviye parametresi 4 seçilerek yapılırsa Şekil 3.11'e benzer bir durum ortaya çıkar [5, 53–55]. Aşağı örnekleme işleminin bir sonucu olarak, dalgacık katsayıları elde edilir. Bu katsayılar hem zaman hem de frekans bilgisinin analizi sonucunda elde edilir. Elde edilen bu dalgacık katsayıları aritmetik ortalama, standart sapma gibi istatistiksel tanımlayıcılarla indirgenerek müzik türü sınıflandırma aşamasında özellik olarak kullanılır. Dalgacık katsayıları, sürekli aynı şekilde devam etmeyen ani değişikliklere sahip şarkılar için iyi bir özellik olarak kullanılabilir. Çalışmanın deneysel sonuçlar bölümünde farklı seviyelerde, farklı dalgacık fonksiyonu ile elde edilen özelliklerle yapılan sınıflandırma başarımları karşılaştırılacaktır.



Şekil 3.11 Ayrık dalgacık dönüşümü

3.5 Evrişimsel Sinir Ağları (ESA) - Convolutional Neural Network (CNN) Tabanlı Özellik Çıkarımı

Çalışmanın bu bölümünde, müzikten akustik özellik çıkarımı ve müzik türü sınıflandırması yapmak için kullanılan ESA modellerinden bahsedilecektir. Genellikle görüntü işleme problemlerinde kullanılan ESA, bu çalışma kapsamında spektrogram, mel-spektrogram ve chromagram gibi zaman-frekans analizi ile elde edilen iki boyutlu işaretlerle eğitilecektir. ESA, farklı işlevlere sahip farklı katmanlara sahiptir. Tüm ESA modelleri Python programlama dili ve Keras kütüphanesi ile oluşturulmuştur. Modellerin temelini oluşturan bazı kısımlar ve işlemler şunlardır: evrişim (convolution) katmanı, aktivasyon fonksiyonu, havuzlama (pooling) katmanı, düzleştirme (flatten) katmanı ve bırakma (dropout) operasyonu. Şekil 3.12’de bir ESA modeli örneği gösterilmektedir.



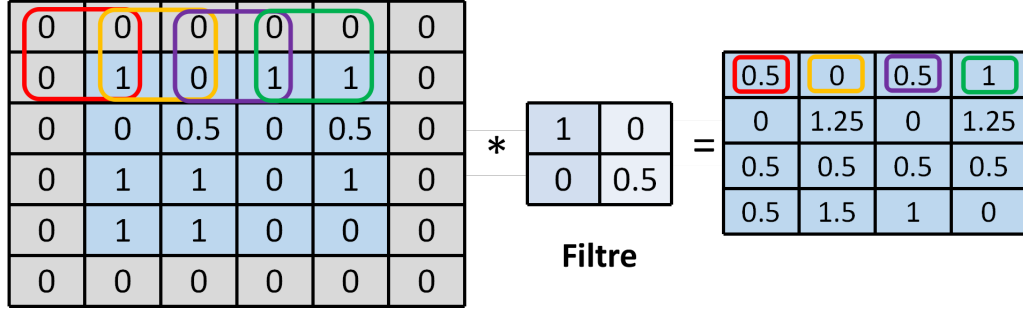
Şekil 3.12 Örnek ESA modeli

3.5.1 Evrişim Katmanı - Convolution Layer

Bu katman özellik tespit edici olarak da adlandırılabilir. Bir görüntüdeki kenar veya köşe gibi ayırt edici olabilen bölgeleri tespit etmeyi sağlar. İşaret işlemedeki evrişim (konvolüsyon) işlemi ile aynıdır. Derin öğrenmede işaret işlemeden farklı olarak çok fazla sayıda evrişim işlemi ile özellik çıkarımı yapılır. Kullanılan evrişim elemanlarına çekirdek (kernel) fonksiyonu veya filtre adı verilir. Bu filtrelerin tipi ve boyutu bu katmanda belirlenmesi gereken parametrelerdendir [58]. Evrişim sırasında işaretteki ilk ve son sınır elemanlarının kaybını önlemek için doldurma (padding) operasyonu gerçekleştirilir. Bu çalışmada bu sınır değerlerin aynısını kullanarak doldurma (same padding) yapılmıştır. Bu katmanda ayrıca klasik yapay sinir ağlarında olduğu gibi aktivasyon fonksiyonları da kullanılır. En çok kullanılan RELU aktivasyon fonksiyonu bu çalışma kapsamında da kullanılmıştır. $[0, t]$ sonlu aralığında tanımlanan f ve g fonksiyonları için evrişim işlemi denklem 3.23'deki gibidir.

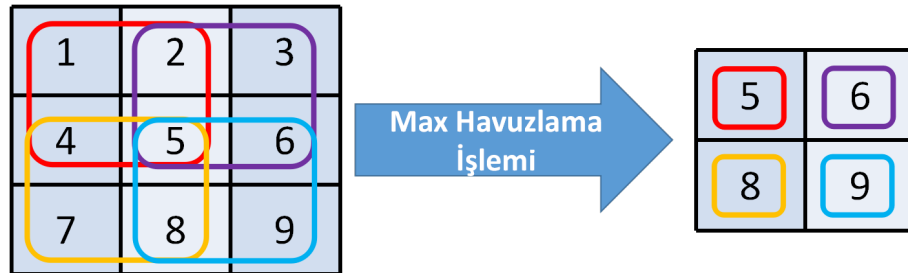
$$(f * g)(t) = \int_{-\infty}^{+\infty} f(\tau)g(t - \tau)d\tau \quad (3.23)$$

Evrişim işlemi ve aynısı ile doldurma (same padding) işlemi Şekil 3.13 ile gösterilmektedir. RELU transfer/aktivasyon fonksiyonunun çalışması ise, Şekil 3.14 ile gösterilmiştir.



3.5.2 Havuzlama Katmanı - Pooling Layer

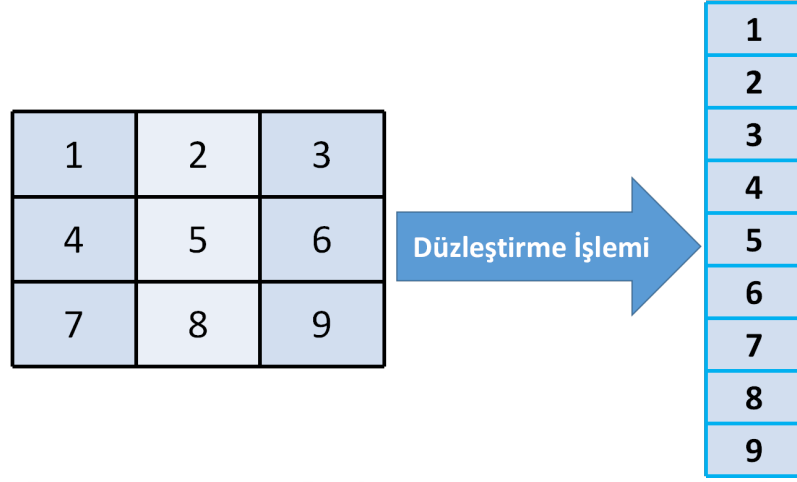
Bu katman, aşağı örnekleme veya alt örnekleme yapar (down sampling), parametre sayısını azaltır, ölçek veya yönlendirme değişikliklerine karşı olmayan özelliklerin algılanmasını sağlar ve aşırı öğrenmeyi (ezberlemeyi) kontrol etmeye yardımcı olur. Bu katmanda ortalama, minimum, medyan gibi başka alternatif yöntemler de vardır, ancak çalışmamız için literatürde sıklıkla kullanılan maksimum havuzlama yöntemi seçilmiştir. Havuzlama işlemi Şekil- 3.15 ile gösterilmektedir.



3.5.3 Düzleştirme Katmanı - Flatten

Düzleştirme, önceki katmanlardan gelen matrisleri bir sonraki katmanda yapılacak olan çok katmanlı yapay sinir ağları işlemlerinde kullanıma uygun hale getirmek için

vektöre dönüştürme operasyonudur. Düzleştirme işlemi Şekil-3.16 ile görsel olarak özetlenmektedir.



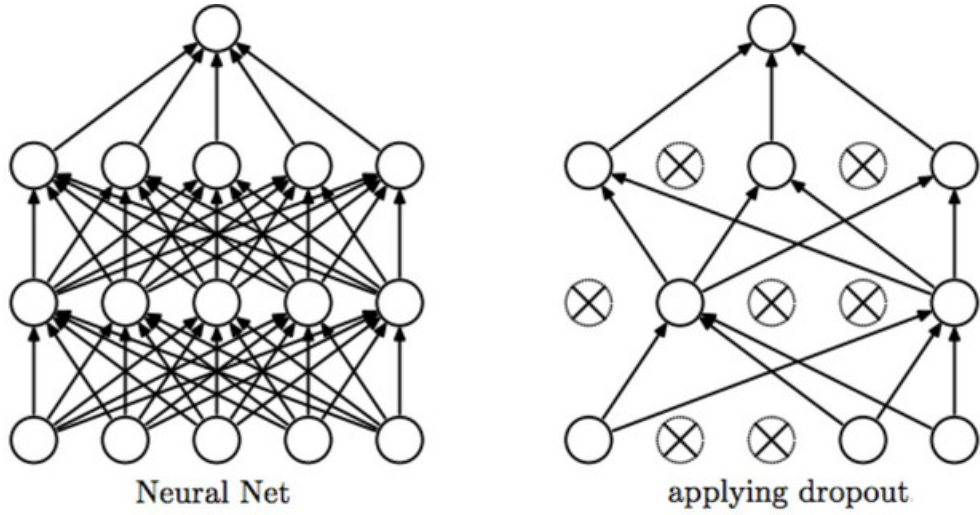
Şekil 3.16 Düzleştirme katmanı

3.5.4 Tam Bağlantılı Öğrenme Katmanı - Fully Connected Layer

Bu katman derin öğrenme ve ESA'da eğitim, ağırlık güncelleme gibi yapay sinir ağları yöntemlerinin bulunduğu kısımdır. Tam bağlantılı olarak, nöronların önceki katmandaki tüm aktivasyonlarla bağlantıları vardır. Çalışmamızda Keras kütüphanesi ile modellere farklı sayıda yoğun katmanlar eklenmiş, düğüm sayısı ve aktivasyon fonksiyonları gibi katman parametreleri belirlenmiştir.

3.5.5 Bırakma Operasyonu - Dropout

Tamamen bağlanmış katmanlarda belirli eşik değerinin altındaki düğümlerin seyreltilmesinin performansı arttırdığı gözlenmiştir. Başka bir ifadeyle, ezberlemenin önüne geçmenin ve zayıf bilgiyi unutmanın test işlemleri sırasında öğrenmeyi arttırdığı gözlemlenmiştir. Bu yüzden bırakma operasyonu eğitim aşamasında tercih edilmiştir. Bu aşamada parametre olarak bir katmandaki işlem elemanlarının ne kadarının işlevsiz kalacağı belirlenmesi gerekmektedir. Örneğin, bırakma olasılığı değeri 0.5 olarak belirlendiğinde o katmandaki işlem elemanlarının yarısının işlevsiz hale geleceği belirlenmiş olur. Bu işlemler sayesinde, ezberlemenin önüne geçilebilir ve test sırasında yapılan denemelerde doğruluk oranı artabilir. Ayrıca, daha az sayıda işlem elemanı kullanılacağı için derin ağın mimarisindeki karmaşıklık da azaltılır. [59]. Bırakma operasyonunun görsel hali Şekil-3.17'de bulunmaktadır.



Şekil 3.17 Bırakma operasyonu [60]

3.5.6 Özellik Çıkarımı ve Makine Öğrenmesi

Derin öğrenme eğitimi tamamlandıktan sonra, elde edilen özellik vektörleri başka makine öğrenmesi sınıflandırıcılarla eğitilebilir. Bu işlemin amacı ESA modelindeki *softmax* sınıflandırıcının diğer sınıflandırma algoritmalarına nispeten daha basit bir yapıda olmasıdır. Derin öğrenme aşamasında, en iyi modelin son tam bağlantılı katmanı eğitim tamamlandıktan sonra kaydedilmektedir. Böylece, katmanın düğüm sayısı kadar özelliğe sahip yeni bir akustik özellik veri kümesi elde edilir. Bundan sonra, bu yeni veri kümesine istenilen makine öğrenmesi yöntemi ve sınıflandırma algoritması uygulanabilir. Örneğin, destek vektör makinesi (SVM), sınıflandırmada kullanılan başarılı ve etkili yöntemlerden biridir. Sınıflandırma için, iki veri grubu arasındaki sınır noktalarını en etkili şekilde tespit ederek bir ayırım yüzeyi meydana getirir. Bu sayede derin öğrenme ile geleneksel işaret işleme yöntemlerine göre daha ayırt edici özellikler elde edilir. Derin öğrenme ile elde edilen özelliklerin daha karmaşık sınıflandırıcılar kullanarak başarıyı daha yüksek müzik türü sınıflandırma sonuçları elde edilmektedir. Deneysel sonuçlarda, mfcc, chroma gibi özelliklerin makine öğrenmesi sonuçları derin öğrenme sonuçlarıyla karşılaştırıldığında kayda değer bir fark olduğu görülmüştür. Deneysel sonuçlar Bölüm-5 kapsamında açıklanmıştır.

3.6 Özellik Son İşleme

Sayısal işaret işleme yöntemleri ham müzikten bir veya çok boyutlu özellik vektörleri elde edilebilmektedir. Örneğin, spektral ağırlık merkezi kısa zamanlı Fourier dönüşümü tabanlı bir işlem olduğundan bu dönüşümdeki her bir pencere için bir

ağırlık merkezi elde edilir. İşaretin M sayıda pencereye bölündüğü varsayılırsa, M uzunluklu bir ağırlık merkezleri vektörü elde edilir. Aynı işlemler MFCC yönteminde her bir pencere için 20 farklı katsayı döndürecek ve dolayısıyla M sayıda pencereden oluşan müzik için $M \times 20$ boyutlarında bir matris elde edilecektir. İşaretlerin ve müziklerin boyutları çok farklı olduğu için elde edilen bu özelliklerin makine öğrenmesi yöntemlerinde doğrudan kullanımı mümkün değildir. Bu nedenle bazı istatistiksel tanımlayıcılar kullanılarak sabit boyutta özellik vektörleri elde edilmiştir. Çalışmamız kapsamında özellik vektörü veya matrisinin elde edildiği tüm durumlarda aritmetik ortalama, standart sapma, medyan, mod, basıklık, çarpıklık gibi istatistiksel tanımlayıcılar kullanılmıştır.

Bu yöntemlere ek olarak özellik vektörlerinin veya matrislerinin olduğu yöntemlerde elde edilen bu verilerin zamana göre değişimi de bir bilgi içerdiğinden, elde edilen bu verilerin birinci mertebeden türevleri alınmış ve yeni özellik vektörleri elde edilmiştir. N elemanlı ve x 'e bağımlı olarak değişen bir f vektörünün birinci mertebeden türevi Eşitlik 3.24 ile gösterildiği gibi hesaplanabilir.

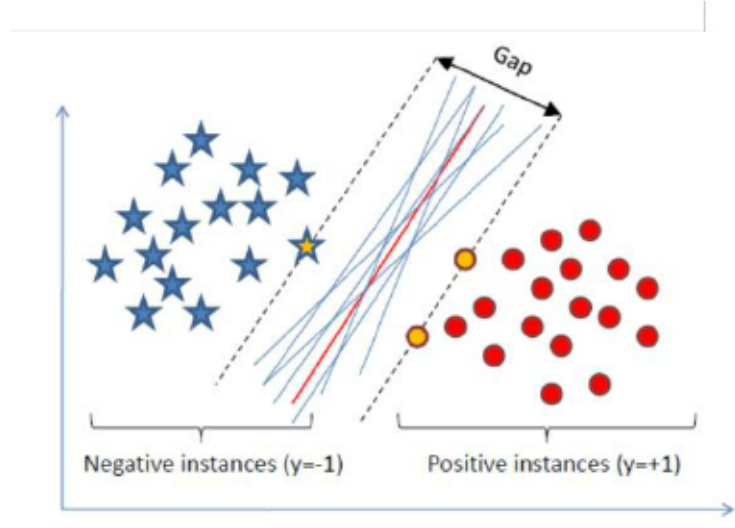
$$f'(x) = f(x) - f(x - 1), x : 1, 2, \dots, N \quad (3.24)$$

Bu bölümde, müzik önerisi ve müzik türü sınıflandırma için tez çalışmasında kullanılan sınıflandırma, kümeleme, derin öğrenme ve müzik benzerliği yöntemlerinden bahsedilmiştir.

4.1 Eğitici Öğrenme Yöntemleri

4.1.1 Destek Vektör Makinesi - Support Vector Machines (SVM)

Destek Vektör Makineleri, veri kümesindeki iki sınıfı bir karar sınırı (hiperdüzlem) ile ayıracak destek vektörünü bulmaya yarayan, eğitici bir makine öğrenmesi yöntemidir. SVM iki sınıf arasındaki sınır noktaların (marjın) tespitini yapar ve bu sınır noktalardan faydalanarak sınıflar arası ayrımı yapacak olan karar sınırlarının oluşturulması mantığıyla eğitim gerçekleştirir [61, 62]. Veri kümeleri ayırım yapılabilirlik bakımından, doğrusal olarak ayrılabilen ve doğrusal olarak ayrılamayan olmak üzere iki şekilde sınıflandırılabilir. Doğrusal ayrılabilen veri kümelerinde, SVM iki kümeyi birbirinden en az hata miktarı ile ayıran doğruyu üretmeye çalışır ve eğitim sonunda bulunan bu doğrunun denklemini üretir. Doğrusal ayırım yapmanın mümkün olmadığı durumlarda ise doğrusal fonksiyonlardan farklı olan çeşitli çekirdek (kernel) fonksiyonları kullanılır. Çekirdek fonksiyonları çok fazla sayıda olmakla birlikte, en yaygın kullanılanları; doğrusal, sigmoid, polinom veya radyal temelli fonksiyonlardır. Polinom çekirdek fonksiyonu, en iyi ayrımı yapabilecek derecede veriyi üst boyutlara taşır ve bu sayede veri doğrusal olarak ayrılabilir hale getirilir. Radyal temelli çekirdek fonksiyonu (Radial Basis Function), sınıfları ayırt etmek için, kapalı karar sınırları kullanır [63]. Doğru bir şekilde veya bir diğer ifadeyle en az miktarda hata ile sınıflandırma yapmak için iki sınıf arasında çizilecek ayırım yüzeyi her iki grubun üyelerinden mümkün oldukça en uzak olmalıdır. Şekil 4.1 bu duruma örnek olarak verilebilir.



Şekil 4.1 İki sınıfı birbirinden ayırt eden destek vektörü (Kırmızı) [64]

4.1.2 Naive Bayes

Naive Bayes algoritması, özellikle girdilerin yüksek boyutlu olduğu durumlarda Bayes teoremine dayanan bir sınıflandırma tekniğidir. Basitliğine rağmen, bazı problemlerde karmaşık sınıflandırma yöntemlerinden daha iyi performans gösterebilir. En basit haliyle, bir örneğin tahmin edilecek olan tüm kategorilere ait olma olasılıklarını hesaplar ve en yüksek olasılığın hesaplandığı kategoriyi sınıf bilgisi olarak geri döndürür. Bu yöntemi kullanmada en önemli kriterlerden birisi, özelliklerin birbirinden bağımsız olmasıdır. Daha fazla özellik birbirlerini etkilediğinden, olasılığı hesaplamak daha zordur. Sonuç olasılığı, sonucu etkileyen tüm özelliklerin olasılıklarının çarpımı yoluyla hesaplanır [65].

Naive Bayes algoritması temel olarak koşullu olasılık kavramından geliştirilmiştir. Koşullu olasılık, önceki olaylara bakarak bir sonraki olayın gerçekleşme olasılığı hakkında çıkarımlar yapmaktadır. Bir müziğin bir kategoride olma olasılığının çarpımının, veri setindeki niteliklerin var olma olasılığı ile bölünmesiyle bulunur [65].

$$P(A|B) = \frac{P(B|A) \times P(A)}{P(B)} \quad (4.1)$$

Formüldeki $P(A|B)$ koşullu olasılık biçiminde, B olayının olması durumunda A olayının olması olasılığı ifade edilmektedir ve bu olasılık sonraki olasılık değeri olarak bilinir. $P(B|A)$ koşullu olasılık biçiminde ise, A olayı meydana geldiğinde B olayının olma olasılığıdır. $P(A)$ ve $P(B)$, sırasıyla A ve B olaylarının önceki olasılıklarıdır [65].

Naive Bayes sınıflandırıcının Gaussian, Multinomial ve Bernoulli gibi farklı türleri

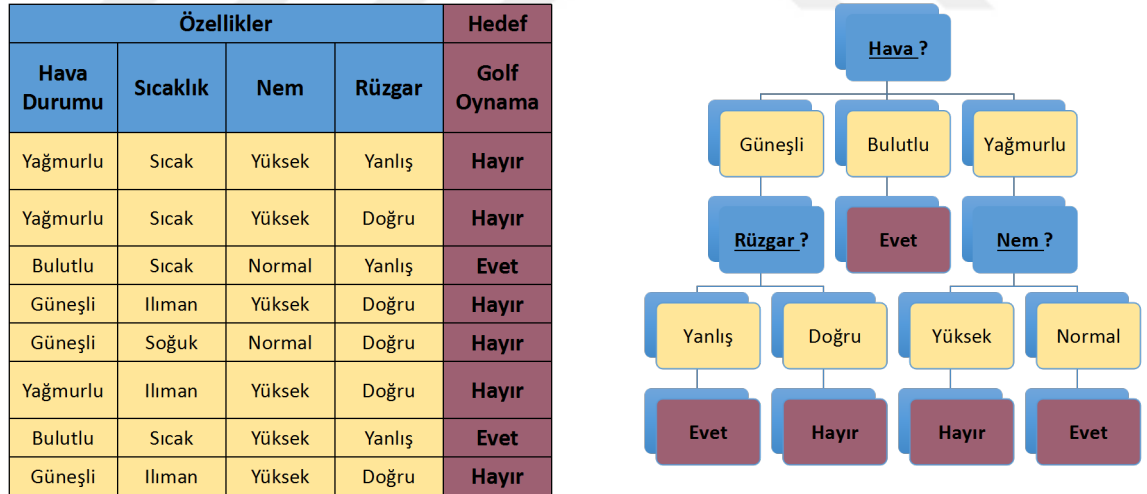
bulunmaktadır. Bu çalışmada kullanılan Gaussian Naive Bayes sınıflandırıcı için, sınıflara ait özelliklerin Gauss dağılımına uygun şekilde dağıldıkları varsayılmıştır [66].

$P(B|A)$ değeri hesaplanırken, bu çalışmada kullanılan akustik özellikler, sayısal sürekli veriler olduğundan Gauss normal dağılımından (Gauss normal distribution) faydalanılır. Gauss normal dağılımı aşağıdaki formülde gösterilmiştir. Formüldeki, e Euler sayısını, σ değişkeni verilerin standart sapmasını verirken, μ değişkeni verilerin aritmetik ortalamasını temsil etmektedir [65].

$$P(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{\sigma^2}} \quad (4.2)$$

4.1.3 Karar Ağaçları - Decision Trees (DT)

Karar ağaçları algoritması kullanılarak ağaç veri yapısında kural tabanlı sınıflandırma modelleri oluşturulmaktadır. Veri seti küçük alt gruplara ayrılırken aynı zamanda artımlı bir şekilde karar ağacı modeli oluşturulmaktadır [67]. Ağacın üstteki düğümlerinin alttaki düğümlere göre daha ayırt edici özelliklere ait olması gerektiği esasıyla eğitim gerçekleştirilir.



Şekil 4.2 Karar ağacı - Decision tree

Karar ağaçları oluşturulurken kök düğümün (root node) hangi özellik olacağına entropi hesapları ile karar verilir. Entropi bir veri seti için Eşitlik 4.3 ile bulunur [67].

$$E(S) = \sum_{i=1}^c -p_i \log_2 p_i \quad (4.3)$$

Karar ağacına yeni bir düğüm (özellik) eklenirken öncelikle ata düğümün ilgili özellik değerine sahip veri setindeki tüm veriler için ata düğümler dışındaki düğümlerin bilgi kazancı (information gain) hesaplanır. Bilgi kazancının en yüksek olduğu özellik (en ayırt edici özellik), karar ağacında ata düğüm olarak belirlenir ve daha az etkili olan diğer özellikler bu ata düğümün alt seviyelerinde yerlerini alır.

$$infoGain = entropy(parent) - [avgWeights] * entropy(children) \quad (4.4)$$

4.1.4 Çok Katmanlı Algılayıcı - Multi-Layer Perceptron (MLP)

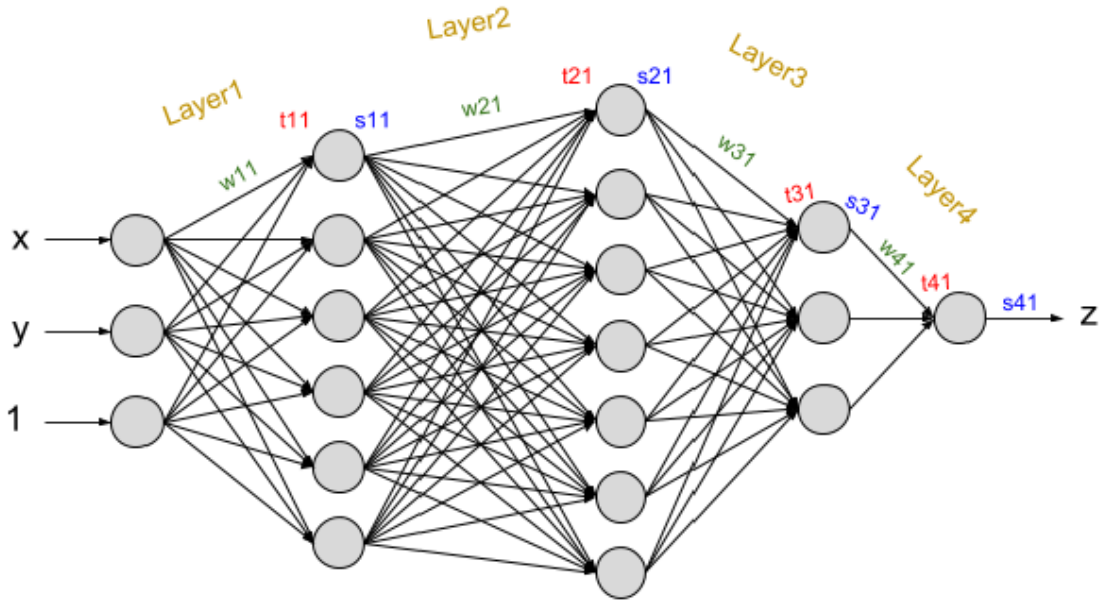
Çok Katmanlı Algılayıcı çok sıklıkla kullanılan yapay sinir ağı mimarilerinden biridir. Makinelerin düşünebileceği ve karar verebileceği düşüncesini temel alarak yapay sinir ağları önerilmiş ve bir çok problemin çözümünde kullanılmıştır. Özellikle doğrusal olarak ayırt edilemeyen sistemlerin modellenmesinde doğrusal olmayan ayırt ediciler önererek bu alanda yapılan çalışmalara önemli gelişmeler kazandırmıştır. Günümüzün en güncel yöntemlerinden olan derin öğrenme algoritmalarının da temelini oluşturması, bu yöntemlerin önemini göstermektedir. Yapay sinir ağı ve çok katmanlı algılayıcı kendilerine verilen örneklerden bilgi toplayarak, genelleme yaparak ve kendilerine gösterilmeyen örneklerle kıyasla daha önce öğrenilen bilgileri toplayarak sınıflandırıcı modeller oluşturan ve bu modelleri kullanarak sisteme sunulan yeni örneklerin sınıf bilgisinin tespit edilmesini sağlamaktadır. Bu nedenle, günümüzde, karmaşık sorunları çözme yeteneği nedeniyle, birçok alanda geniş kullanımı vardır [68, 69].

Basit bir yapay sinir ağının yapısı; biyolojik sinir sistemi açısından nöronlar, dendritler, hücre gövdeleri, aksonlar ve sinapslar; yapay sinir sistemi mimarisi açısından, sırasıyla işlem elemanı, toplama fonksiyonu, aktivasyon fonksiyonu, eleman çıkışı ve ağırlıklardan oluşur. İşlem elemanı kendisine verilen girdileri alır, ağırlıklandırır ve toplama fonksiyonuna ekler. Toplama işlevinden gelen değer aktivasyon fonksiyonuna eklenir ve bir sonraki işlem elemanına aktarılır. Son olarak çıktı değerleri bulunur ve varsa bu çıktı değeri sonraki katmanlara girdi olarak verilir. Birden fazla katmana sahip yapay sinir ağları çok katmanlı yapay sinir ağları olarak adlandırılır [68, 69].

Çok katmanlı yapay sinir ağları, bir giriş katmanı, ara veya gizli katmanlar (hidden layers) ve bir çıkış katmanı olan ağ mimarileridir. Giriş katmanı dış sisteme verilecek olan verinin özelliklerini alır, bu özellikleri bir müziğin akustik özelliği olabileceği gibi bir görüntünün piksel değerleri gibi çok farklı şekillerde olabilir. Ara katmanlar, giriş katmanından gelen bilgileri işler ve her bir işlem elemanına ait ağırlıkların güncellenmesine yardımcı olur. Ara katmanların karmaşıklığı çözülmesi gereken

problemin karmaşıklığına göre belirlenir. Çıkış katmanı ara katmanlardan gelen ağırlık ve çıktı bilgilerini toplayarak çıkışta anlamlı bir sonuç üretir. Bu sonuç, örneğin bir müziğin tür bilgisi olabileceği gibi bir görüntünün içeriğinin bilgisi gibi çok farklı türlerde olabilir [68, 70].

Çok katmanlı algılayıcı, ileri beslemeli bir yapay sinir ağı mimarisine sahiptir. Bu algoritma bir lojistik regresyon gibi çalışır fakat girdiler lojistik regresyona verilmek yerine, gizli katman olarak adlandırılan bir ara katmana verilir. Bu ara katmanlarda doğrusal olmayan bir aktivasyon fonksiyonu (genellikle tanh veya sigmoid) kullanılır [69]. Örnek bir çok katmanlı algılayıcı mimarisi Şekil 4.3 ile gösterilmektedir.



Şekil 4.3 Örnek çok katmanlı algılayıcı mimarisi

Şekil 4.3 üzerinde gösterilen X , Y , 1 olarak verilen değerler giriş değerlerini; w_{11} , w_{21} , w_{31} , w_{41} değerleri ağırlıkları; t_{11} , t_{21} , t_{31} ve t_{41} toplam fonksiyon değerlerini; s_{11} , s_{21} , s_{31} ve s_{41} aktivasyon fonksiyonu sonuç değerlerini ve z çıktı değerini temsil etmektedir.

MLP sınıflandırıcının eğitim aşaması iki ana bölüme ayrılabilir: ileri besleme ve geri yayılım. İleri beslemede, nöronların değerleri denklemde gösterildiği gibi hesaplanır (Eşitlik 4.5). Hesaplanan değer Eşitlik 4.6 ile gösterilen sigmoid fonksiyonundan geçirilir, bir sonraki katman veya işlem elemanına bu değer aktarılır.

$$net = bias + \sum_{i=1}^n x_i w_i \quad (4.5)$$

$$f(net) = sig(net) = \frac{1}{1 + e^{-net}} \quad (4.6)$$

Geri yayılımda, hata değeri E bulunan sonuç ile gerçek sonuç arasındaki fark hesaplanarak bulunur (Eşitlik 4.7). Elde edilen hata değeri kullanılarak son katmandan ilk katmana doğru tüm ağırlıklar güncellenir ve eğitim bu şekilde devam eder. Hatanın geriye doğru güncellenmesi sırasında farklı iyileştirme yöntemleri bulunmaktadır. Eşitlik 4.8 ile gösterilen gradyan düşümü (gradient descent) bu yöntemlerden bir tanesidir.

$$E = \frac{1}{2} \sum_j e_j^2 = \frac{1}{2} \sum_j (\bar{y}_j - y_j)^2 \quad (4.7)$$

$$\Delta w_{i,j} = -\frac{\partial E}{\partial w_{i,j}} = -\frac{\partial E}{\partial y_j} \frac{\partial y_j}{\partial net_i} \frac{\partial net_i}{\partial w_{i,j}} \quad (4.8)$$

Eğer aktivasyon fonksiyonu olarak *sigmoid* kullanılırsa, ağırlıkların güncellenmesi Eşitlik 4.9 ve Eşitlik 4.10'da gösterildiği gibi yapılır.

$$\Delta w_{i,j} = e_j(1 - y_j)y_jx_i \quad (4.9)$$

$$\Delta w_{i,j} = w_{i,j} + \eta \Delta w_{i,j} \quad (4.10)$$

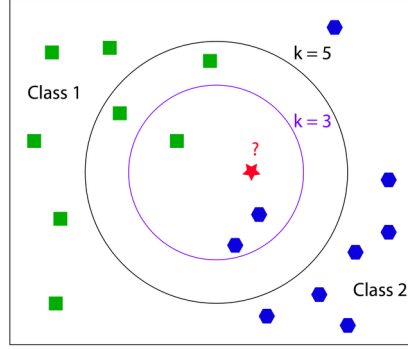
Eşitlikteki η değeri öğrenme oranını göstermektedir. Bu değer, gradyan düşümü ile güncelleme sırasında hata değerini iyileştirirken yerel minimum bölgelerinde takılmayı engeller, daha hassas bir ağırlık ayarlaması yapar. Ancak bu hassas arama işlemi bir gecikme de meydana getirir.

4.1.5 K-En Yakın Komşuluklar - K Nearest Neighbours (K-NN)

K-NN; parametrik olmayan, uzaklık temelli, sınıflandırma ve regresyonda kullanılan bir makine öğrenmesi algoritmasıdır [71]. Eğitim örnekleri, test örnekleriyle aralarındaki mesafeye göre küçükten büyüğe sıralanarak ilk k örnekte fazla çıkan sınıf etiketi test örneğinin sınıf etiketi olarak belirlenmektedir.

K-NN algoritması iki farklı şekilde kullanılmaktadır: Çoğunluk oylaması (Majority vote) ve mesafe ağırlıklandırılmış skor (Distance-weighted score). Çoğunluk

oylamasında basit olarak test ve eğitim örnekleri arasındaki mesafe sadece eğitim verilerini mesafeye göre küçükten büyüğe sıralama için kullanılmaktadır. Mesafe ağırlıklandırılmış skorda ise mesafeye göre küçükten büyüğe sıralanmış eğitim örneklerinin sınıf etiketleri, test örneğine olan mesafeye göre ağırlıklandırılmaktadır.

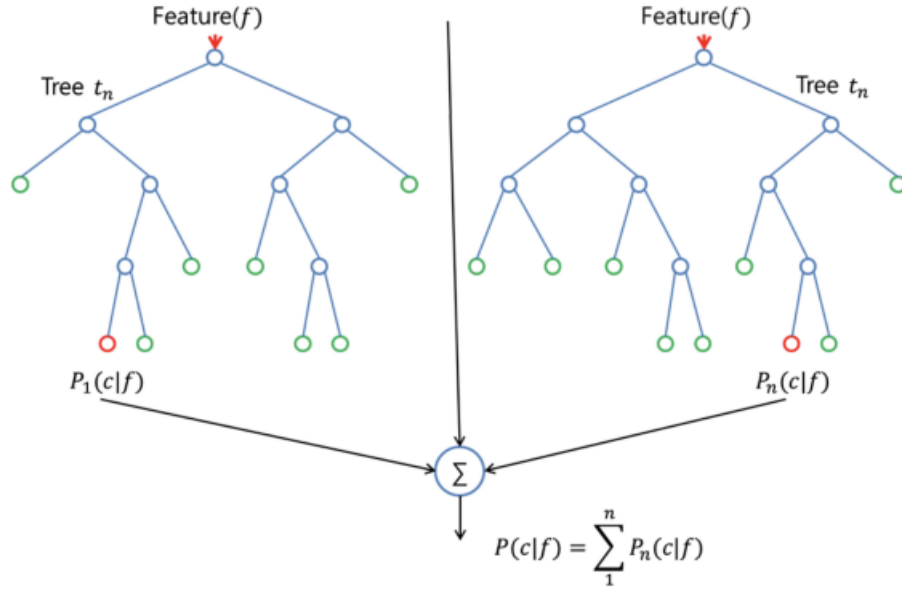


Şekil 4.4 K En yakın komşuluk algoritması [72]

4.1.6 Rassal Ormanlar - Random Forests (RF)

Kollektif öğrenme, makine öğrenmesi konusunda başarıyı artırmak için geliştirilen stratejilerden birisidir. Çok karmaşık yapıda olan öğrenme algoritmalarını kullanmak yerine çok sayıda zayıf öğrenme algoritmasını kollektif olarak kullanmayı esas alır. Bu strateji temel olarak örneklerin seçimi (bagging) ve özelliklerin seçimi (boosting) yöntemleriyle gerçekleştirilir. Rassal Ormanlar, çok sayıda karar ağacından oluşan kollektif bir sınıflandırma algoritmasıdır. Her bir karar ağacı, verilen girdiler doğrultusunda bir sınıf üzerine karar verir. Bu kararlar arasında en fazla oya sahip olan sınıf, verilen girdinin sınıfı olarak belirlenir. Karar ağaçlarının eğitimi, her bir ağaç için eğitim kümesinden rastgele seçilen bir örnek kümesiyle yapılır [73, 74].

Bu algoritmanın en önemli avantajlarından biri, her bir karar ağacının birbirinden bağımsız olarak karar vermesiyle sonuca oybirliğiyle ulaşmasıdır. Ağaçların bir kısmı yanlış sonuca ulaşsa bile, çoğunluğun doğru sonuca ulaşması karar verme aşamasında yanlış sonuca gitmeyi engelleyecektir [73, 74].



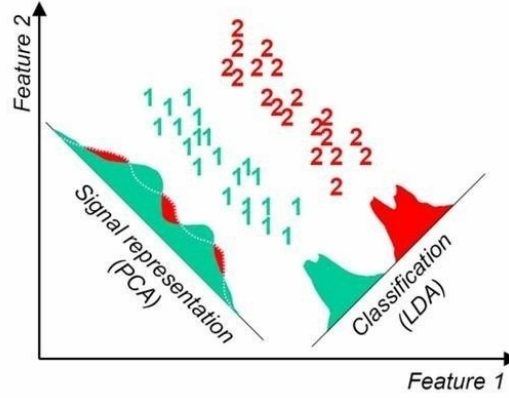
Şekil 4.5 Rassal ormanlar - Random forests [75]

Şekil 4.5, Rassal Ormanlar algoritmasının sınıflandırıcı olarak N sayıda karar ağacı ürettiğini ve genel sonucun bu N farklı ağacın toplamına göre hesaplanacağını göstermektedir.

4.1.7 Doğrusal Ayırt Edici Analiz - Linear Discriminant Analysis (LDA)

Doğrusal Ayırt Edici Analiz, makine öğrenimi uygulamalarının ön işleme aşamalarında boyut indirgeme amaçlı ve sınıflandırma aşamasında yaygın olarak kullanılan bir algoritmadır. Amaç, ezberlemeyi önlemek ve aynı zamanda hesaplama maliyetlerini azaltmak için, iyi bir sınıf ayrılabilirliği olan veri kümesi sağlamaktır [76, 77].

Normalde, orijinal doğrusal ayırt edici iki sınıflı problemler için tanımlanır, ancak daha sonra çok sınıflı ayırt edici analizi olarak genelleştirilebilir. LDA genel olarak Temel Bileşen Analizine (Principal Component Analysis - PCA) çok benzemektedir. Ancak LDA, PCA'daki verilerin varyansını en üst düzeye çıkaran bileşenler bulmaya ek olarak, çoklu sınıflar arasındaki ayrımı en üst düzeye çıkaran eksenleri bulmakla ilgilenir. LDA'nın amacı, özellik alanını daha küçük bir alt alana indirgemek ve sınıf ayrımcılığı bilgisini korumaktır. PCA denetimsiz bir algoritma olmasına rağmen, LDA eğitici bir algoritma olarak kullanılabilir.



Şekil 4.6 PCA ve LDA karşılaştırması [78]

4.2 Eğiticişiz Öğrenme Yöntemleri

Eğiticişiz öğrenme veya Kümeleme yöntemleri, sayısal işaret işleme yöntemleriyle elde edilen özellik vektörünü nicemlemek (vector quantization), diğer ifadeyle özellik boyutunu indirgemek için kullanılmıştır. Özellik vektörünün belirli sayıda kümelere ayrıldıktan sonra vektöre istatistiksel analiz fonksiyonlarının kullanılmasının başarıya etkisinin gözlemlenmesi hedeflenmiştir. Kullanılan kümeleme yöntemleri K-Ortalamlar (K-Means) ve Gauss Karışım Modeli (Gaussian Mixture Model - GMM) olarak seçilmiştir.

4.2.1 K-Means Kümeleme

K-Means; verilen bir veri setini K adet kümeye, küme içi benzerliklerin maksimum ve kümeler arası benzerliklerin minimum olacak şekilde bölünmesini sağlayan bir kümeleme algoritmasıdır.

Algoritma öncelikle, rastgele K merkez noktası belirler. Daha sonra tüm noktalar, kendilerine en yakın olan merkez noktasıyla gruplandırılır. Elde edilen grupların, ortalamaları alınarak orta noktaları bulunur. Bulunan bu orta noktalar yeni merkez noktaları olarak kabul edilerek, tekrar gruplandırma yapılır. Bu işlem, merkez noktalar değişmeyene kadar (veya ihmal edilebilecek kadar az değişene kadar) tekrarlanır. Son durumda elde edilen gruplar, başlangıçta verilen veri setinin kümelere bölünmüş hali olur [79].

4.2.2 Gauss Karışım Modeli (GMM)

GMM, veri kümesinin belirli sayıda Gauss dağılımlarından oluştuğunu kabul eder. Her bir Gauss dağılımı, dağılımın merkezini ifade eden bir ortalama (μ) ve dağılımın

genişliğini ifade eden bir standart sapma (σ) değişkeninden oluşmaktadır.

Veri kümesindeki Gauss dağılımlarını en iyi ifade eden μ ve σ değerleri, Beklenti Maksimizasyonu (Expectation Maximization) algoritması ile bulunur. Böylece veri seti Gauss dağılımlarıyla ifade edilen kümelere bölünmüş olur [80].

4.3 Önerilen Derin Öğrenme Mimarileri

Bu bölüm, farklı derin öğrenme ve makine öğrenimi modellerinin başarılarını, doğruluk, kesinlik ve hassasiyet gibi ölçütleri içerir. Tüm modeller, k -katlı çapraz doğrulama yöntemiyle farklı k değerleri için eğitilmiştir. Deneysel sonuçlarda kullanılan k değerleri ayrıca belirtilmiştir [81].

4.3.1 ESA Model-1 (ESA + Dropout)

ESA Model-1, sistem tasarımında açıklandığı gibi basit bir ESA modelidir. İçerisinde evrişim katmanı, havuzlama ve düzleştirme işlemleri ile birlikte bırakma (dropout) operasyonu yapılmaktadır. Modelin kaynak kodu Ek-A'da verilmiştir.

4.3.2 ESA Model-2 (ESA + Dropout + Auto Encoder)

Bu modele ESA Model-1'e ek olarak eğitici-siz öğrenme yöntemi olan Autoencoder eklenmiştir. Bu şekilde, giriş ve çıkış verilerinin boyutları korunarak ESA modelinde daha az kayıpla özellik vektörleri oluşturulur. Modelin kaynak kodu Ek-A'da verilmiştir.

4.4 Müzik Önerisi

Müzik önerisi öneri sistemleri altında çalışılan bir konudur. Öneri sistemleri, genel olarak, beğenilen bir nesneyle birlikte kişiye aynı veya benzer türde özellikleri olan yeni nesnelerin önerilmesini sağlayan sistemler olarak tanımlanabilir. Burada önerilen nesne internette alışveriş esnasında tercih edilen bir kitaba göre farklı kitapların önerilmesi, film izleme ortamlarında izlenen filmin beraberinde farklı film veya filmlerin önerilmesi gibi çok çeşitli şekilde olabilir. Öneri sistemi hem kullanıcının yeni şeyler keşfetmesini sağlayacaktır, hem de ticari kuruluşların daha fazla kazanç elde etmesini sağlayacaktır. Öneri sistemleri temel olarak aşağıda özetlenen yöntemlerle çalışmaktadır [82, 83].

- **İşbirlikçi Filtreleme Öneri Sistemleri:** Bu yöntemde göre, kişiler arasındaki işlem geçmişini veri kümesindeki ortaklıkların ve farklılıkların tespit edilerek farklılıklar üzerinden kişilere yeni öneriler sunmak amaçlanmaktadır. A ve B kişilerinin dinledikleri 100 müzikten 70 tanesi aynıyken, 30 tanesi farklı ise, bu farklı olan müzikleri olmayan kişiye önerilmesi örnek olarak verilebilir. Bu yöntemde en önemli sorun kişilerin belli bir işlem geçmişinin olması gerektiğidir. Hiç bir işlemi olmayan birisi için bu işlem mümkün olmamaktadır. Kişiler arasındaki benzerliklerin tespit edilmesi işlemlerine genel olarak işbirlikçi filtreleme (collaborative filtering) adı verilir. Bu yöntem nesnelerin içeriği ile doğrudan ilgilenmediği için, nesnelerin içeriği ile ilgili metin madenciliği, işaret işleme, görüntü işleme ve benzeri içerik temelli konular yoluyla nesneler arasında benzerlik kurmaz. Bu da bu tür yöntemlerin kullanılabilirliğini kolaylaştırmaktadır. Özellikle popüler kültür açısından kullanıcılar arasında içinde bulunulan döneme uygun nesnelerin yaygınlığı göz önüne alındığında, bu yöntemle güncel olan nesnelere erişmek daha mümkün olmaktadır [82].
- **İçerik Tabanlı Öneri Sistemleri:** İçerik tabanlı öneri sistemleri, ilgilenilen nesnelerin içeriğinden elde edilecek özellikler yardımıyla benzer nesnelerin önerilmesine imkan sağlayan sistemlerdir. Bu yöntemle nesnelerin içeriğine ilişkin özelliklerin elde edilmesi gerektiğinden, nesnenin türüne uygun özellik çıkarımı yöntemleri kullanmak zorunludur. Örnek olarak bu tez çalışmasında müzikler üzerinde öneri sistemi gerçekleştirileceği için müziklerin akustik özellikleri sayısal işaret işleme yöntemleri ve derin öğrenme ile elde edilmektedir. Bu sayede kullanıcı geçmişini olmasa bile, müzik önerisi yapabilmek mümkün hale gelmektedir [82].
- **Bilgi tabanlı Öneri Sistemleri:** Bu yöntemle öneri yapılırken kullanıcının gereksinimlerine uygun olarak kullanıcıya kısıtlar sunulur. Bu kısıtlar altında sorgulanan nesnelerden kullanıcının seçim yapması beklenir. Özellikle tekil kullanıcıların çok olduğu elektronik alışveriş sistemlerinde gereksinim duyulan nesnelerin kişiye özgün nesneler olması bu tarz öneri sistemlerini gerekli kılmıştır. Örnek olarak, kullanıcı bilgisayar almak istediğinde işlemci hızı, bellek kapasitesi, kullanım alanı gibi farklı kısıtların kullanıcıdan alınması yapılacak önerilerin daha çok amacına uygun olmasını sağlayacaktır [82].
- **Sepet Analizi Tabanlı Öneri Sistemleri:** İşbirlikçi öneri sistemlerine benzer olarak çalışmaktadır. Ancak burada kullanıcılar arasında bir ilişki olmasına gerek yoktur. Bu yöntemde şu ana kadar işlem gören nesneler arasında birliktelik ilişkisinden yola çıkılarak öneri gerçekleştirilir. Bir çevrimiçi alışveriş sitesinde şimdiye kadar kurşun kalem, pilot kalem satın alanların %90'ı çay, kahve gibi ürünler de satın almışsa bu ürün gruplarından birisini tercih edene

diğerininin indirimli veya indirimsiz bir fiyatla önerilmesi bu sistemlere örnek olarak verilebilir [82].

- **Hibrit Yaklaşımlar:** Bu yaklaşım önceki maddelerde açıklanan yöntemlerin iki veya daha fazlasının bir arada kullanılmasıyla yapılan öneri sistemleridir.

4.4.1 Akustik Özellikler Kullanan Müzik Önerisi

Çalışmanın esas konusu olan müzik önerisi için kullanıcı geçmişinin olmadığı, işbirlikçi öneri ve filtreleme yöntemlerinin çalışmadığı durumlar göz önüne alınarak içerik temelli yöntem kullanılmaktadır.

Müzik içeriğine ilişkin ayırt edici özelliklerin elde edilmesi bu çalışmanın özellik çıkarımı başlığı altında açıklanmıştır. Elde edilen özelliklerden akustik özellik veri tabanı oluşturulduktan sonra kullanıcının seçtiği bir müziğe en çok benzer içerikte olan N sayıda müziğin önerilmesi benzerlik yöntemleriyle gerçekleştirilmiştir. Bu bağlamda ifade edilen yöntem en yakın komşuluk ilişkisinin bulunması olarak da ifade edilebilir. En benzer müziklerin bulunması için literatürde en çok kullanılan yöntemlerden olan Euclidean uzaklık ölçüsü, Pearson korelasyon katsayısı ve Kosinüs benzerlik fonksiyonu kullanılmıştır. Özellik vektörü sırasıyla X ve Y olan n boyutlu iki müzik için, Euclidean uzaklık ölçümü Eşitlik 4.11 ile verilmiştir.

$$d(X, Y) = \sqrt{\sum_{i=1}^n (X_i - Y_i)^2} \quad (4.11)$$

Özellik vektörü sırasıyla X ve Y olan n boyutlu iki müzik için, Pearson Korelasyon Katsayısına ait eşitlikler aşağıda verilmiştir. Burada $\bar{X}$ ve $\bar{Y}$, X ve Y vektörlerinin aritmetik ortalamalarını; S_{XX} ve S_{YY} ise X ve Y vektörlerinin varyans değerlerini göstermektedir. X ve Y vektörü arasındaki kovaryans ve Pearson Korelasyon Katsayısı değerleri ise sırasıyla S_{XY} ve r_p ile gösterilmiştir.

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i, S_{XX} = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 \quad (4.12)$$

$$\bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i, S_{YY} = \frac{1}{n-1} \sum_{i=1}^n (Y_i - \bar{Y})^2 \quad (4.13)$$

$$S_{XY} = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y}) \quad (4.14)$$

$$r_p = \frac{S_{XY}}{\sqrt{S_{XX}S_{YY}}} \quad (4.15)$$

Özellik vektörü sırasıyla X ve Y olan n boyutlu iki müzik için, Kosinüs benzerlik değeri Eşitlik 4.16 ile hesaplanabilir. $\|X\|$ ve $\|Y\|$, X ve Y vektörlerinin gücünü temsil etmektedir.

$$\cos(X, Y) = \frac{XY}{\|X\|\|Y\|} = \frac{\sum_{i=1}^n X_i Y_i}{\sqrt{\sum_{i=1}^n (X_i)^2} \sqrt{\sum_{i=1}^n (Y_i)^2}} \quad (4.16)$$



5

DENEYSEL SONUÇLAR

5.1 Veri Seti

Deneysel çalışmalarda akustik özelliklerin elde edilmesi, müzik türü sınıflandırma ve müzik önerisi işlemlerinin yapılması için GTZAN veri seti kullanılmıştır. Bu veri setinin tercih edilmesindeki amaç, 1000 farklı örnek ve 10 farklı kategoride 30’ar saniyelik tür etiketli müzikler içermesidir [84, 85]. Bu sayede hem her türden eşit sayıda müzik olduğu için düzensiz veri dağılımı problemi meydana gelmemiş hem de müzikler 30 saniye olduğu için algoritma karmaşıklığı yüksek yöntemlerden aşırı yavaş sonuçlar alınmasına engel olunmuştur.

Diğer yandan, tür etiketi bulunan GTZAN veri seti ile eğitilen modeller kullanılarak Türkcell tarafından sağlanan 17.000 müziğin akustik özellikleri çıkarılmış ve elde edilen bu özelliklerle müzik önerisi gerçekleştirilmiştir. Gelecek bölümlerde bu verilerle elde edilen deneysel sonuçlara detaylı olarak yer verilecektir.

5.2 Müzik Türü Sınıflandırma İçin Başarım Ölçütleri

5.2.1 Karmaşıklık Matrisi

Karmaşıklık matrisi bir sınıflandırma algoritmasından tahmin edilen test sonuçlarının gerçekte olan veya beklenen sonuçlarla karşılaştırıldığı tablodur. İki sınıflı problemlerde kullanıldığında sınıflandırma sonuçları pozitif ve negatif işaretli olarak düşünülür ve böylece doğru ve yanlış yapılan sınıflandırma sonuçlarının sayısı hesaplanır [65, 86]. İki sınıflı bir sınıflandırma problemi için örnek bir karmaşıklık matrisi Tablo 5.1 ile gösterilmiştir.

Tablo 5.1 Karmaşıklık Matrisi

KARMAŞIKLIK MATRİSİ	Tahmin Sonucu Pozitif	Tahmin Sonucu Negatif
Gerçekte Pozitif	Doğru Pozitif Sayısı (TP)	Yanlış Negatif Sayısı (FN)
Gerçekte Negatif	Yanlış Pozitif Sayısı (FP)	Doğru Negatif Sayısı (TN)

Karmaşıklık matrisine ilişkin temel kavramlar aşağıda kısaca özetlenmiştir [65, 86].

- **Doğru Pozitif - True Positive (TP):** TP; gerçekte pozitif olan ve pozitif etiketli olarak doğru tahmin edilen örneklerin sayısıdır.
- **Doğru Negatif - True Negative (TN):** TN; gerçekte negatif olan ve negatif etiketli olarak doğru tahmin edilen örneklerin sayısıdır.
- **Yanlış Pozitif - False Positive (FP):** FP; gerçekte negatif olan ancak pozitif etiketli olarak yanlış tahmin edilen örneklerin sayısıdır.
- **Yanlış Negatif - False Negative (FN):** FN; gerçekte pozitif olan ancak negatif etiketli olarak yanlış tahmin edilen örneklerin sayısıdır.

Devam eden bölümlerde karmaşıklık matrisinden türetilen bazı başarımlar ölçütleri detaylı olarak açıklanacaktır.

5.2.2 Doğruluk (Accuracy)

Doğruluk en sık kullanılan ölçütlerden birisidir. Gerçekte pozitif olup pozitif olarak tahmin edilen örneklerin sayısı (TP) ile gerçekte negatif olup negatif olarak tahmin edilen örneklerin (TN) toplam sayısının (TP+TN), tüm tahminlerin sayısına (TP + FP + FN + TN) oranıdır. Doğruluk hesabı Eşitlik 5.1 ile gösterildiği gibi yapılır [65, 86].

$$\text{Doğruluk} = \frac{TP + TN}{TP + FP + FN + TN} \quad (5.1)$$

5.2.3 Kesinlik (Precision)

Kesinlik, gerçekte pozitif olan örneklerin sayısının (TP), pozitif olarak tahmin edilen örneklerin sayısına (TP+FP) oranıdır. Kesinlik oranı ne kadar yüksek olursa FP örneklerin sayısı o kadar az olur. Kesinlik hesabı Eşitlik 5.2 ile gösterildiği gibi yapılır [65, 86].

$$\text{Kesinlik} = \frac{TP}{TP + FP} \quad (5.2)$$

5.2.4 Duyarlılık (Recall)

Duyarlılık, pozitif olarak tahmin edilen örneklerin sayısının (TP), gerçekte pozitif olarak bilinen bir diğer ifadeyle pozitif olarak tahmin edilmesi gereken örneklerin sayısına (TP+FN) oranıdır. Duyarlılık hesabı Eşitlik 5.3 ile gösterildiği gibi yapılır [65, 86].

$$\text{Duyarlılık} = \frac{TP}{TP + FN} \quad (5.3)$$

5.2.5 F Skoru (F Score)

Duyarlılık ve kesinlik değerlerinden türetilen F skoru başarımlı ölçütlerinde kullanılan önemli bir ölçüttür. Hem FP hem de FN değerleri dikkate alındığı için kesinlik ve hassasiyetin ikisine ilişkin bir bilgi vermektedir [65, 86].

$$\text{F Skoru} = \frac{2 * \text{Duyarlılık} * \text{Kesinlik}}{\text{Duyarlılık} + \text{Kesinlik}} \quad (5.4)$$

5.3 Müzik Önerisinin Değerlendirilmesi

Öneri sistemlerinde değerlendirme yapma işlemi, öneri yapılacak nesneye göre ve öneri yapılan yönteme göre değişiklik göstermektedir. Diğer yandan, öneri sisteminin değerlendirilmesi için öznel (sübjektif) ve nesnel (objektif) değerlendirme olarak iki tür değerlendirme ölçütü genel bir çerçeve olarak belirlenmiştir [82]. Önerilen nesnenin belli bir oran veya yüzde üzerinden oylanarak değerlendirilmesi öznel değerlendirmeye örnek olarak verilebilir.

Müzik önerisinde ve gerçekleştirdiğimiz bu çalışmada öznel değerlendirmeyi, Fizy uygulamasında aktif hale gelecek öneri motoru sayesinde önerilen müziklerin dinlenme süresi veya tıklanma oranı gibi metriklerle yapmak mümkün olacaktır. Diğer taraftan, öneri işlemlerinin nesnel ve bilimsel ölçütlerle yapılması önem arz etmektedir. Nesnel müzik önerisi değerlendirmesi için aynı türde veya aynı döneme ait müzik önerisinin kontrolü, aynı sanatçının müziklerinin öneri listesinde bulunma oranı gibi çeşitli ölçütler kullanılabilir. Bu çalışma kapsamında akustik özellikleri çıkarılan müziklere yapılan önerilerin değerlendirilmesi için etiketli veri setinde yapılan denemelerde aynı türden müziğin önerilme yüzdeleri kullanılmıştır. Seçilen bir müzik için akustik özellikler kullanarak önerilen farklı N sayıda müziğin bu seçilen müzikle aynı türden olma yüzdeleri ölçüt olarak kullanılmıştır. Bu sayede hem müzik önerisinin başarımları elde edilmiş hem de birbirine çok fazla veya çok benzeyen müzik türleri analiz edilebilmiştir. Elde edilen başarımlarla ilgili detaylara deneysel sonuçlar bölümünde yer verilmiştir.

5.4 Sayısal İşaret İşleme ile Elde Edilen Özelliklerle Bulunan Sonuçlar

Bu kısımda müzik türlerini belirlemek için sayısal işaret işleme ile çıkartılan bütün özellikler kullanılmıştır. Özellik vektörleri boyut sayısını indirmek için iki farklı şekilde oluşturulmuştur. Bunlardan birincisi, her bir özellik çıkarma yönteminden elde edilen sonuçların aritmetik ortalaması ve standart sapması kullanılarak oluşturulmuştur. İkinci olarak, işaret işleme yöntemlerinden elde edilen sonuçların aritmetik ortalama, standart sapma, mod, medyan, basıklık, çarpıklık ve entropi değerlerini içerecek şekilde bir özellik veri seti gerçekleştirilmiştir.

5.4.1 K En Yakın Komşuluk Algoritması Sonuçları

KNN algoritmasında k (komşu sayısı) değeri müziğin türünün doğru tahmin edilmesinde bir etkiye sahiptir. Aşağıdaki tablolarda farklı k değerleri ile elde edilen sonuçlar özetlenmiştir. Ortalama doğruluk değerinde %2'lik bir artış olduğu gözlemlenmiştir.

Tablo 5.2 K-NN Başarım Yüzdeleri (K-Fold = 5, k = Farklı komşuluk değeri)

k	Kesinlik	Duyarlılık	F Skoru	Min Acc	Avg Acc	Max Acc
1	63.71	63.11	63.40	58.50	62.30	64.50
3	65.07	64.81	64.93	62.50	64.30	65.50
5	65.15	65.29	65.21	63.50	64.80	65.50
7	65.16	65.29	65.22	61.50	64.30	67.00
9	65.57	66.91	66.23	61.00	64.80	67.00

Tablo 5.3 K-NN Başarım Yüzdeleri (K-Fold = 10, k = Farklı komşuluk değeri)

k	Kesinlik	Duyarlılık	F Skoru	Min Acc	Avg Acc	Max Acc
1	65.51	64.24	65.87	59.00	64.70	69.00
3	66.23	65.47	65.84	62.00	65.60	70.00
5	65.52	65.92	65.71	62.00	66.00	72.00
7	66.67	66.44	66.55	63.00	66.60	73.00
9	66.85	66.91	66.88	62.00	66.80	73.00

5.4.2 Destek Vektör Makinesi Sonuçları

SVM makine öğrenmesi algoritmasında algoritmanın başarımını belirleyen önemli etkenlerden biri sınıfları ayırt etmek için kullanılan kernel tipi parametresidir. Bu parametre *lineer*, *poly* ve *rbf* olabilir. Bu kernel tipleri içerisinde ise başarı oranını etkileyecek bazı parametreler vardır. Bunlardan bir tanesi de *c* parametresidir. *c* parametresi sınıfları ayırt etmek için kullanılan hiper düzleme olan uzaklıkla ilgilidir. Yani *c* büyük olursa daha küçük bir uzaklıkta hiper düzlem oluşturulacaktır. Bu da çok daha iyi bir ayırım sağlayacaktır. *c* parametresi üç kernel tipi için de kullanılmıştır.

Tablo 5.4 SVM Başarım Yüzdeleri (K-Fold = 5, Kernel Tipi = Lineer)

c	Kesinlik	Duyarlılık	F Skoru	Min Acc	Average Acc	Max Acc
0.1	71.95	72.44	72.19	69.00	72.20	74.00
1	72.47	72.60	72.53	71.50	72.90	76.50
5	70.65	71.15	70.90	68.00	71.50	76.50

Tablo 5.5 SVM Başarım Yüzdeleri (K-Fold = 10, Kernel Tipi = Linear)

c	Kesinlik	Duyarlılık	F Skoru	Min Acc	Average Acc	Max Acc
0.1	71.61	71.65	71.63	66.00	72.10	79.00
1	71.54	71.65	71.59	67.00	72.10	79.00
5	70.76	70.27	71.51	67.00	71.40	81.00

Görüldüğü üzere lineer SVM ile elde edilen başarı oranı c değeri değişimine göre büyük bir değişiklik göstermemektedir. Kernel tipi olarak *poly* veya *rbf* seçildiğinde c parametresi yanında gamma parametresi de gerekmektedir. Gamma parametresi ise iki nokta arasındaki uzaklığı ifade etmektedir. Eğer gamma parametresi büyük olursa gamma fonksiyonunun varyansı da büyük olur. Bu eğer iki nokta arasındaki mesafe büyükse kullanılması düşünülebilir.

Tablo 5.6 SVM Başarım Yüzdeleri (K-Fold = 5, Kernel Tipi = Poly)

c	gamma	Kesinlik	Duyarlılık	F Skoru	Min Acc	Average Acc	Max Acc
0.1	1/25	64.81	65.46	65.13	57.00	64.50	70.00
	1/5	73.96	74.86	74.41	73.00	74.70	76.50
1	1/25	74.78	75.27	75.02	70.50	74.90	78.00
	1/5	73.78	74.64	74.20	72.25	74.60	77.00
5	1/25	73.92	74.74	74.33	72.50	74.70	77.50
	1/5	73.78	78.64	76.13	72.50	74.60	77.00

Tablo 5.7 SVM Başarım Yüzdeleri (K-Fold = 10, Kernel Tipi = Poly)

c	gamma	Kesinlik	Duyarlılık	F Skoru	Min Acc	Average Acc	Max Acc
0.1	1/25	64.00	64.41	64.20	57.00	64.40	69.00
	1/5	74.35	74.16	74.25	71.00	75.00	82.00
1	1/25	75.21	75.21	72.21	70.00	76.20	84.00
	1/5	74.59	74.57	74.58	71.00	75.50	81.00
5	1/25	74.98	75.04	75.01	71.00	76.00	81.00
	1/5	74.59	74.57	74.57	71.00	75.50	81.00

Tablo 5.8 SVM Başarım Yüzdeleri (K-Fold = 5, Kernel Tipi = Rbf)

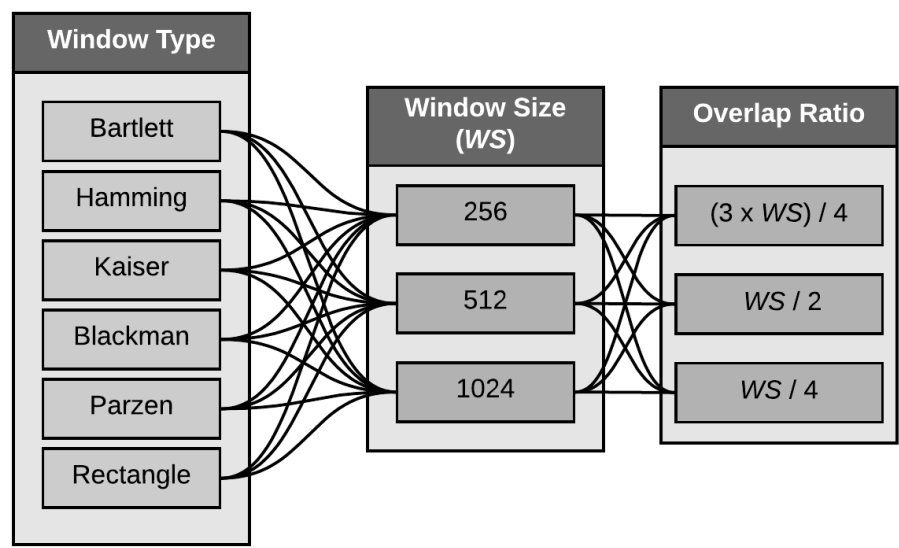
c	gamma	Kesinlik	Duyarlılık	F Skoru	Min Acc	Average Acc	Max Acc
0.1	1/25	42.10	45.96	43.95	38.50	42.00	48.00
	1/5	15.27	17.29	16.22	7.00	11.70	15.00
1	1/25	72.17	72.22	72.19	68.50	71.90	75.00
	1/5	71.47	71.80	71.63	66.50	71.10	74.00
5	1/25	75.88	76.62	76.25	73.50	76.30	80.00
	1/5	71.09	71.73	71.41	68.50	71.10	72.50

Tablo 5.9 SVM Başarım Yüzdeleri (K-Fold = 10, Kernel Tipi = Rbf)

c	gamma	Kesinlik	Duyarlılık	F Skoru	Min Acc	Average Acc	Max Acc
0.1	1/25	49.71	48.12	48.90	39.00	45.90	60.00
	1/5	16.83	23.71	19.68	5.00	16.80	27.00
1	1/25	72.55	72.32	72.73	68.00	73.30	77.00
	1/5	72.09	71.19	71.63	67.00	72.40	77.00
5	1/25	76.48	76.47	76.47	72.00	76.90	84.00
	1/5	73.78	73.62	73.70	71.00	74.50	79.00

5.4.3 STFT Parametrelerinin Sınıflandırma Başarısına olan Etkisi

Bu kısımda Kısa Zamanlı Fourier Dönüşümü parametrelerinin sınıflandırma başarısına olan etkisi incelenmiştir. Sınıflandırma için Rassal Ormanlar algoritması ve SVM algoritması lineer, rbf ve polynomial kernel ile kullanılmıştır. Kullanılan parametrelerin türleri ve değerleri, Şekil 5.1 ile özet olarak gösterilmiştir.

**Şekil 5.1** Kullanılan STFT parametreleri ve değerleri [87]

Tablo 5.10 farklı pencere boyutu ve farklı örtüşme değerleri kullanılarak elde edilen başarımların sonuçlarını göstermektedir.

Tablo 5.10 Pencere boyutu ve örtüşme değerine göre STFT başarımlar yüzdeleri

Pencere Boyutu	Örtüşme Değeri	SVM(Poly)	SVM(RBF)	SVM(Linear)	Rassal Ormanlar
256	64	68.40	60.10	64.80	63.50
	128	68.80	60.50	65.30	63.80
	196	68.60	59.90	65.10	63.40
512	128	69.20	62.30	66.10	63.50
	256	69.50	62.50	66.20	64.70
	384	69.20	62.20	66.10	63.90
1024	256	69.10	58.30	65.50	63.60
	512	69.40	59.30	66.40	64.40
	768	68.50	58.70	65.70	63.50

Tablo 5.11 Farklı pencere tiplerine göre STFT başarımlar yüzdeleri

Pencere Tipi	SVM(Poly)	SVM(RBF)	SVM(Linear)	Rassal Ormanlar
Bartlett	70.00	62.80	67.20	65.70
Hamming	70.40	63.7	67.20	66.30
Kaiser	70.00	62.30	66.00	64.70
Blackman	69.90	63.10	67.10	64.60
Rectangle	69.30	62.10	66.00	64.30
Parzen	71.30	63.30	67.50	65.70

5.4.4 Makine Öğrenmesi Algoritmalarının Karşılaştırılması

Bu kısımda bütün özellikler kullanıldığında kullanılan makine öğrenmesi algoritmalarının başarı oranları kıyaslanacaktır. Tablo 5.12 akustik özellikler ve dalgacık dönüşümü özellikleri ortalama, standart sapma, kurtosis, skewness, medyan ve mod istatistiksel tanımlayıcıları ile elde edilen en fazla sayıda özellik içeren veri tabanının sınıflandırılması ile kaydedilen sınıflandırma sonuçlarını göstermektedir. Sınıflandırma için k -fold değeri 5 ve 10 olarak seçilmiştir. En yüksek ortalama doğruluk sonucu LDA algoritması %83.90 doğrulukla, SVM algoritmasının RBF kernel, $c = 100$ ve $\gamma = 0.01$ parametreleri ile %83.50 doğrulukla elde edilmiştir.

Tablo 5.12 Farklı sınıflandırıcıların başarımlar yüzdeleri (Tüm özellikler)

Algoritma	k-fold = 5	k-fold = 10
MLP	73.60	75.60
Rassal Ormanlar	75.10	76.50
LDA	83.20	83.90
KNN (k=1)	73.10	75.70
SVM (Poly)	82.70	83.30
Gaussian Naive Bayes	68.00	68.80
SVM (Lineer)	79.80	80.80
SVM (Rbf)	82.90	83.50

5.4.5 Özelliklerin Doğruluk Değerlerine Etkisi

Özelliklerin başarı oranına etkisini bulmak için SVM algoritması kernel tipi olarak lineer ve $k - fold$ değeri 10 olarak alınmıştır.

5.4.5.1 Her Özellik Tek Başına Kullanıldığında

Tablo 5.13 Özellikler tek başına kullanıldığında başarımlar yüzdeleri

Özellik	Keskinlik	Duyarlılık	Min Acc	Avg Acc	Max Acc
Sıfır Geçiş Oranı	26.60	27.26	21.00	25.40	37.00
Spektral Ağ. Merkezi	30.70	34.70	29.00	34.40	40.00
Spektral Karşıtlık	20.17	31.79	26.20	32.10	41.00
Spektral Bant Gen.	25.46	32.34	24.00	31.80	38.00
Spektral Etek	26.62	35.06	30.00	34.40	42.00
MFCC	70.75	70.13	60.00	70.80	77.00
RMSE	24.88	30.88	23.00	28.90	37.00
Chroma CQT	31.86	32.26	28.00	32.50	40.00
Tonnetz	30.96	30.39	22.00	30.00	35.00
Spektral Düzlük	10.05	17.89	11.00	14.96	19.00
Polinom Özelliği	27.85	30.30	19.00	28.70	39.00
Polinom (3.Derece)	23.70	26.78	19.00	25.60	37.00

Tablo 5.14 Özellikler farklı kombinasyonlarla kullanıldığında başarımlar yüzdeleri

Özellik Grubu	Polinom Derece	Kesinlik	Duyarlılık	Min Acc	Avg Acc	Max Acc
Sadece Spektral Özellikler	-	48.83	47.70	40.00	47.30	52.00
Sadece Chroma Cqt ve Tonnetz	-	35.35	36.15	29.00	36.30	43.00
Chroma Cqt ve Tonnetz Hariç Hepsi	1	75.98	75.33	71.00	76.30	80.00
Chroma Cqt ve Tonnetz Hariç Hepsi	3	75.52	74.58	69.00	75.70	82.00
Bütün Özellikler	1	71.54	71.65	67.00	72.10	79.00
Bütün Özellikler	3	71.53	71.04	66.00	72.10	80.00

Tablo 5.15 Özellikler sırayla eklendiğinde başarımlar yüzdeleri

Özellikler	Kesinlik	Duyarlılık	Min Acc	Avg Acc	Max Acc
Zero Crosssing	26.6	27.26	21	25.4	37
+Spektral Ağ. Merkezi	35.76	38.46	32.00	38.40	46.00
+Spektral Karşıtlık	41.68	46.65	34.00	46.50	55.00
+Spektral Bant Genişliği	45.91	49.38	41.00	49.00	56.00
+Spektral Etek	46.54	50.02	42.00	49.80	57.00
+RMSE	54.40	54.55	48.00	54.50	62.00
+Chroma CQT	56.64	56.72	53.00	57.50	63.00
+Tonnetz	58.21	59.04	56.00	59.80	64.00
+Spektral Düzlük	58.29	59.23	56.00	59.90	64.00
+MFCC	71.48	70.76	66.00	71.90	81.00
+Polinom 1. Derece	71.54	70.65	67.00	72.10	79.00
+Polinom 3. Derece	71.53	71.04	66.00	72.10	80.00

5.4.5.2 MFCC Yönteminin Analizi İle Elde Edilen Sonuçlar

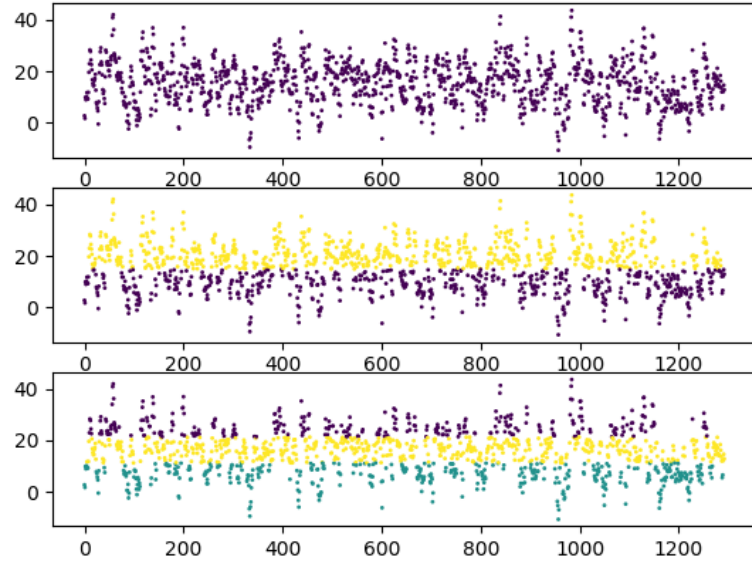
Bu bölümde MFCC özelliğinin farklı şekillerde işlenmesinin ve farklı yöntemlerle sınıflandırılmasının başarıyı nasıl etkilediği incelenmiştir.

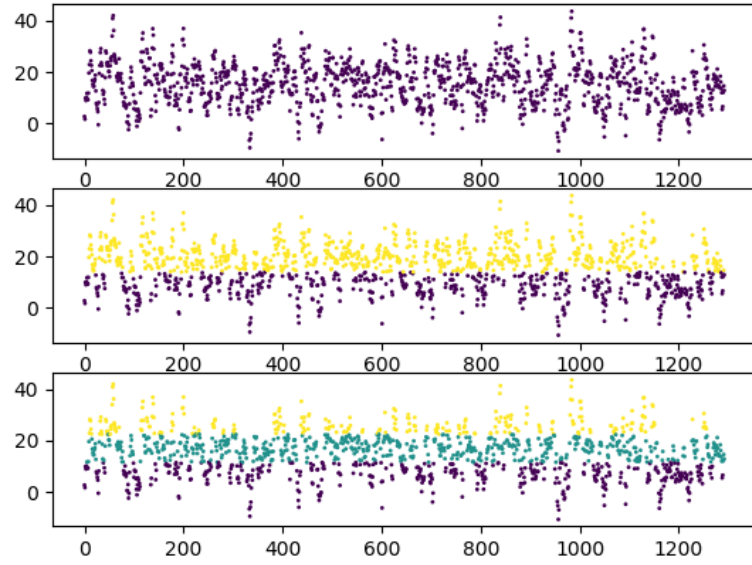
13 katsayılı MFCC vektörünün öncelikle sütun bazında ortalamaları ve standart sapmaları alınmıştır. Elde edilen 26 özellik sınıflandırma aşamasında kullanılmak üzere saklanmıştır. Daha sonra aynı MFCC vektörü, sütunları ayrı ayrı ve tüm sütunlar bütün olmak üzere iki farklı şekilde; *K*-means ve GMM kümeleme algoritmaları aracılığıyla 2'şer ve 3'er kümeye bölünmüştür. Elde edilen 56 özellikli ve 78 özellikli vektörlerin farklı sınıflandırıcılarla başarıları değerlendirilmiştir. MFCC vektörünün farklı yöntemlerle kümelenebilmesinin ortalama gecikme süreleri Tablo 5.16'de verilmiştir.

Tablo 5.16 Kümeleme algoritmalarının çalışma hızları

	K-Means (k = 2)	K-Means (k = 3)	GMM (k = 2)	GMM (k = 3)
MFCC kaysayılarını kümeleme	384 ms	575 ms	150 ms	204 ms
MFCC tamamen kümeleme	56 ms	90 ms	29 ms	51 ms

Şekil 5.2 ve Şekil 5.3’de MFCC vektörünün bir sütununun farklı k (küme sayısı) değerleri için K-Means ve GMM ile kümelmesi gösterilmiştir. Tablo 5.17, MFCC özellik çıkarımı sonrasında elde edilen verilerin hiç kümeleme olmadan, K-Means yöntemiyle iki ve üç kümeli, GMM yöntemiyle iki ve üç kümeli olarak ortalama ve standart sapmaları bulunduğu zaman elde edilen başarımlarını göstermektedir.

**Şekil 5.2** Bir MFCC sütununun K-Means ile kümelere ayrılması



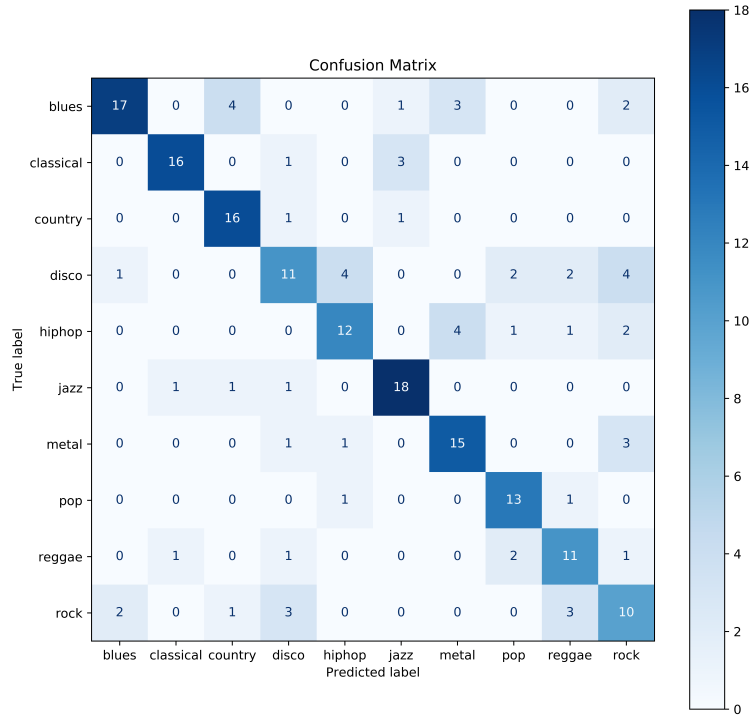
Şekil 5.3 Bir MFCC sütununun GMM ile kümelere ayrılması

Tablo 5.17 MFCC'nin farklı küme sayılarında başarıım yüzdeleri

Algoritma	Tek Küme	2 küme (K-Means)		3 küme (K-Means)		2 küme (GMM)	
		Sütuna Göre	Tüm Mfcc	Sütuna Göre	Tüm Mfcc	Sütuna Göre	Tüm Mfcc
SVM	68.60	59.90	63.20	55.70	61.70	53.90	63.90
Rassal Ormanlar	62.80	57.30	61.10	51.20	61.60	54.30	62.70
Gaussian NB	49.10	49.50	49.00	52.00	50.00	47.70	48.30
MLP	68.70	59.50	64.50	59.20	61.20	55.90	64.20

Yukarıda verilen başarı sonuçlarına bakıldığında, MFCC özelliğinin kümelenmesinin başarıyı olumsuz etkilediği görülmektedir. Ayrıca MFCC özelliği için; sütunların bir bütün olarak kümelenmesine kıyasla, ayrı ayrı kümelenmesi başarı oranını düşürmektedir. Tablo 5.17 ile verilen grafiklere bakıldığında en başarılı sınıflandırıcılar, %68.70 başarı oranıyla MLP ve %68.60 başarı oranıyla SVM olmuştur.

Şekil 5.4'de MLP sınıflandırıcı için elde edilmiş karışıklık matrisi (confusion matrix) verilmiştir.



Şekil 5.4 MLP için elde edilen Karmaşıklık matrisi (Confusion Matrix)

Ayrıca MFCC yönteminde tek küme kullanıldığında, özellik vektörlerinin birinci mertebeden türevinin de kullanılması durumu incelenmiş olup, başarımlar sonuçları Tablo 5.18 ile özetlenmiştir. Sonuçlardan görüleceği üzere, MFCC yöntemi ile birlikte birinci mertebeden türev kullanımı başarıyı yaklaşık %3 artırmıştır.

Tablo 5.18 MFCC'nin türevi ile kullanımında başarımlar yüzdeleri

MFCC Kullanım Durumu	Doğruluk Yüzdesi
MFCC	73.40
MFCC 1. Türevi	72.80
MFCC + MFCC 1. Türevi	76.20

5.4.6 Dalgacık Dönüşümü Özellikleri ile Elde Edilen Sonuçlar

Bu bölümde farklı dalgacık dönüşümleri ile elde edilen akustik özelliklerden deneysel olarak elde edilen sonuçlar takip eden tablolarda verilecektir.

5.19, 5.20, 5.21, 5.22, 5.23 ve 5.24 tablolarında farklı fonksiyon ve her bir fonksiyon için 4 farklı seviye değeri için doğruluk yüzdeleri verilmektedir. Deneyler yapılırken spektral ağırlık merkezi, sıfır geçiş oranı gibi sadece klasik özelliklerin kullanıldığı durumda başarımlar %67.83 olmuştur.

Tablo 5.19 Haar ailesinin farklı seviyelere göre başarımlar yüzdeleri

Fonksiyon	Seviye 3	Seviye 6	Seviye 9	Seviye 12	Ortalama
haar	68	68	67	67	67
Ortalama	68	68	67	67	67

Tablo 5.19 incelendiğinde Haar ailesinde seviye 3 ve seviye 6, seviye 9 ve seviye 12'ye göre daha başarılı olmuştur. Dalgacık dönüşümü olmaksızın elde edilen başarımlar dikkate alındığında kayda değer bir başarımlar artışı gözlenmemiştir.

Tablo 5.20 ile Ayrik Meyer ailesi kullanılarak elde edilen sonuçlar gösterilmiştir.

Tablo 5.20 Ayrik Meyer ailesinin farklı seviyelere göre başarımlar yüzdeleri

Fonksiyon	Seviye 3	Seviye 6	Seviye 9	Seviye 12	Ortalama
dmey	67	67	66	65	66
Ortalama	67	67	66	65	66

Tablo 5.21 Symlet ailesi ile elde edilen başarımlar sonuçlarını göstermektedir. Sonuçlara göre başarımların en fazla %69 olduğu görülmektedir.

Tablo 5.21 Symlet ailesinin farklı seviyelere göre başarıım yüzdeleri

Fonksiyon	Seviye 3	Seviye 6	Seviye 9	Seviye 12	Ortalama
sym2	69	68	67	65	67
sym3	68	69	67	66	68
sym4	69	69	67	67	68
sym5	68	68	66	64	67
sym6	69	69	67	65	68
sym7	68	68	66	66	67
sym8	68	68	66	66	67
sym9	68	67	66	64	66
sym10	68	68	66	65	67
sym11	68	68	66	65	67
sym12	68	68	67	65	67
sym13	67	68	66	66	67
sym14	68	68	66	65	67
sym15	69	68	67	67	67
sym16	68	68	67	65	67
sym17	68	67	66	64	66
sym18	68	68	66	66	67
sym19	68	67	66	65	67
sym20	68	68	66	65	67
Ortalama	68	68	66	65	67

Tablo 5.22 Coiflet ailesi ile elde edilen başarıım sonuçlarını göstermektedir. Sonuçlara göre başarıımın en fazla %69 olduğu görülmektedir.

Her iki tabloda da görüldüğü üzere, seviye artışının başarıyı düşürdüğü ve hatta olumsuz etkisi olduğu gözlemlenmektedir.

Tablo 5.22 Coiflet ailesinin farklı seviyelere göre başarımlar yüzdeleri

Fonksiyon	Seviye 3	Seviye 6	Seviye 9	Seviye 12	Ortalama
coif1	69	69	67	67	68
coif2	69	68	67	65	67
coif3	68	68	66	64	67
coif4	68	68	66	65	67
coif5	68	68	66	65	67
coif6	68	68	66	66	67
coif7	68	67	65	65	66
coif8	68	67	66	65	67
coif9	68	68	66	66	67
coif10	67	67	66	65	66
coif11	68	68	65	66	67
coif12	68	67	65	65	66
coif13	68	68	65	65	67
coif14	68	67	65	65	66
coif15	68	67	65	64	66
coif16	68	67	65	64	66
coif17	68	67	65	65	66
Ortalama	68	68	66	65	67

Dalgacık dönüşümünde sıklıkla kullanılan Daubechies ailesine ait başarımlar sonuçları Tablo 5.23 ile özetlenmiştir. Sonuçlara göre Daubechies ailesinin kullanımı, kayda değer bir başarımlar artışı sağlamamıştır.

Tablo 5.23 Daubechies ailesinin farklı seviyelere göre başarıım yüzdeleri

Fonksiyon	Seviye 3	Seviye 6	Seviye 9	Seviye 12	Ortalama
db1	67	67	66	66	67
db2	68	67	66	65	67
db3	68	68	67	65	67
db4	68	68	66	65	67
db5	67	67	65	65	66
db6	68	68	66	65	67
db7	68	68	66	64	66
db8	68	67	65	64	66
db9	68	67	65	64	66
db10	67	67	65	66	66
db11	68	67	66	65	67
db12	67	67	66	66	66
db13	67	67	65	64	66
db14	66	67	65	64	66
db15	68	67	65	66	66
db16	67	67	65	65	66
db17	67	68	65	65	66
db18	68	67	64	64	66
db19	67	67	64	63	65
db20	67	67	65	64	66
db21	67	67	65	64	66
db22	67	66	63	64	65
db23	67	67	64	64	66
db24	67	67	65	64	66
db25	67	66	65	65	66
db26	68	67	64	65	66
db27	67	67	65	65	66
db28	67	66	65	65	66
db29	67	67	65	66	66
db30	67	67	65	67	67
db31	68	67	65	66	67
db32	68	68	65	65	66
db33	68	67	65	66	66
db34	67	67	65	66	66
db35	68	68	65	67	67
db36	67	67	64	66	66
db37	68	67	65	65	66
db38	68	68	65	67	67
Ortalama	68	67	65	65	66

Tablo 5.24 Biorthogonal ailesinin farklı seviyelere göre başarımları yüzdeleri

Fonksiyon	Seviye 3	Seviye 6	Seviye 9	Seviye 12	Ortalama
bior1.1	68	68	67	67	67
bior1.3	68	67	66	67	67
bior1.5	68	68	66	67	67
bior2.2	69	69	68	67	68
bior2.4	68	68	67	66	68
bior2.6	68	68	67	65	67
bior2.8	68	68	67	66	67
bior3.1	68	67	65	62	66
bior3.3	69	67	67	65	67
bior3.5	68	67	67	64	67
bior3.7	68	67	66	64	66
bior3.9	68	67	66	65	66
bior4.4	68	69	67	66	68
bior5.5	68	68	66	66	67
bior6.8	68	68	67	65	67
Ortalama	68	68	67	65	67

Tablo 5.24'e göre, başarımları %62 - %69 arasında değişmekte, ortalama başarımları %68 olmaktadır. Tablo 5.25, daha önceki tablolardaki her bir ailenin ortalama başarımlarını özetlemektedir. Genel ortalama %67 olarak hesaplanmıştır. Ayrıca seviye artışının başarımlara olumsuz etkisinin olduğu gözlemlenmektedir.

Tablo 5.25 Ailelerin seviyelere göre ortalama başarımları yüzdeleri

Aile	Seviye 3	Seviye 6	Seviye 9	Seviye 12	Ortalama
haar	68	68	67	67	67
db	68	67	65	65	67
sym	68	68	66	65	67
coif	68	68	66	65	67
bior	68	68	67	65	67
rbio	68	68	67	66	67
dmey	67	67	66	65	66
Ortalama	68	68	66	66	67

Tablo 5.26, Tablo 5.27 ve Tablo 5.28 seviyelere göre başarımlarda artış olan dalgacık fonksiyonlarını göstermektedir.

Tablo 5.26 Seviye 9 ile başarımları arttıran dalgacık fonksiyonları

Fonksiyon	Seviye 9
bior2.2	68
rbio3.1	68

Tablo 5.27 Seviye 6 ile başarımları arttıran dalgacık fonksiyonları

Fonksiyon	Seviye 6	Fonksiyon	Seviye 6
db2	68	coif2	68
db3	69	coif6	68
db4	68	bior1.5	68
db6	68	bior2.2	69
db7	68	bior2.4	68
db9	68	bior2.6	68
db17	68	bior2.8	68
sym2	68	bior4.4	69
sym3	69	bior5.5	68
sym4	69	bior6.8	68
sym6	69	rbio1.3	68
sym7	68	rbio2.2	69
sym8	68	rbio2.4	69
sym11	68	rbio2.6	69
sym12	68	rbio3.1	68
sym15	68	rbio3.5	68
sym16	68	rbio3.7	68
sym18	68	rbio4.4	68
sym20	68	rbio5.5	69
coif1	69	rbio6.8	68

Tablo 5.28 Seviye 3 ile başarıımı arttıran dalgacık fonksiyonları

Fonksiyon	Seviye 3	Fonksiyon	Seviye 3
db2	68	coif6	68
db3	69	coif7	68
db4	68	coif8	69
db5	68	coif9	68
db6	68	coif12	68
db7	68	coif13	68
db8	68	coif14	68
db9	68	coif15	68
db10	68	coif16	68
db11	68	coif17	68
db15	68	bior1.1	68
db16	68	bior1.3	68
db18	68	bior1.5	68
db24	68	bior2.2	69
db25	68	bior2.4	68
db26	68	bior2.6	69
db32	68	bior2.8	68
db37	68	bior3.1	68
sym2	69	bior3.3	69
sym3	68	bior3.7	68
sym4	69	bior3.9	68
sym6	69	bior4.4	68
sym7	68	bior5.5	68
sym8	68	bior6.8	68
sym10	68	rbio1.3	68
sym11	68	rbio1.5	68
sym12	68	rbio2.2	68
sym14	68	rbio2.4	69
sym15	69	rbio2.6	68
sym16	68	rbio2.8	68
sym17	68	rbio3.1	68
sym18	68	rbio3.3	68
sym19	68	rbio3.5	68
sym20	68	rbio3.7	68
coif1	69	rbio3.9	68
coif2	69	rbio4.4	68
coif4	68	rbio5.5	68

Tablo 5.29 başarımlı ortalaması en yüksek olan dalgacık fonksiyonlarını listelemektedir.

Tablo 5.29 Başarı ortalaması en yüksek dalgacık fonksiyonları

Fonksiyon	Seviye 3	Seviye 6	Seviye 9	Seviye 12	Ortalama
rbio3.1	68	68	68	67	68
bior2.2	69	69	68	67	68
coif1	69	69	67	67	68
sym4	69	69	67	67	68
Ortalama	69	69	68	67	68

Tablo 5.30, çift katmanlı dalgacık dönüşümü uygulanarak elde edilen sonuçları göstermektedir. Sonuçlar incelendiğinde %72 oranında ortalama başarıma ulaşılmış olup, başarımda yaklaşık %5'lik bir artış olduğu görülmektedir.

Tablo 5.30 Çift katmanlı yöntem ile başarımlı yüzdeleri

Fonksiyon	Seviye	148 Pencere	596 Pencere	Ortalama
db3	3	73	71	72
sym2	3	73	71	72
sym4	3	73	72	72
rbio2.4	3	72	71	72
sym3	6	72	70	71
sym6	6	72	71	71
coif1	6	72	70	71
bior2.2	6	71	71	71
bior4.4	6	72	70	71
bior2.2	9	70	70	70
rbio3.1	9	72	70	71
Ortalama		72	71	71

5.5 Derin Öğrenme Yöntemleri İle Elde Edilen Sonuçlar

Bu bölümde derin öğrenme ESA-1 ve ESA-2 modelleri ile elde edilen sonuçlar özetlenmiştir. ESA-1 modeli ile parametre iyileştirmesi için farklı parametreler denenmiş ve elde edilen sonuçlar özetlenmiştir. Diğer yandan, kullanılan veri setinde Chromagram ve Spektrogram verileri elde edilmiş ve derin öğrenme yöntemleri

uygulanmıştır. Bu veriler elde edilirken veri setindeki 30 saniyelik her bir müzik 6 saniyelik parçalara ayrılmış ve böylece 5000 farklı müzik verisi elde edilmiştir. Veri setinde 10 farklı türde müzik olduğu için kesinlik ve hassasiyet değerleri bu tabloda verilmemiş, bu metrikler müzik önerisi kısmında tür benzerliği yoluyla önerilen yöntemle test edilmiş ve özetlenmiştir.

5.5.1 ESA Parametre Optimizasyonu Sonuçları

Derin öğrenme modellerinin çok sayıda meta parametresi olduğundan ve bu meta parametrelerdeki değişiklikler modelin doğruluğunu etkilediğinden, en yüksek başarımla sahip olmak için parametrelerin optimum değerlerinin bulunması önemlidir. ESA Model-1 için bu parametrelerin farklı değerlerindeki doğruluk yüzdeleri ayrıntılı olarak Tablo 5.31 ile özetlenmiştir.

Tablo 5.31 ESA Parametre optimizasyonu başarımlı yüzdeleri

Parameter Type	Parametre	Avg Acc.
Giriş Boyutu	128x128	72.33
	64x64	75.70
	32x32	67.4
Batch Boyutu	64	74.01
	32	75.70
	16	73.90
Filtre Boyutu	7x7	73.70
	5x5	75.70
	3x3	73.90
Filtre Sayısı	64	74.90
	32	75.70
	16	74.70
Hata Fonksiyonu	Mean Squared Error	74.90
	Cross Entropy	74.60
	Hinge	68.70
Transfer Fonksiyonları	Tanh	74.40
	Relu	75.70
	Sigmoid	52.70
İyileştirme Fonksiyonu	SGD()	63.01
	Adadelat()	75.70
	RMSprop()	70.70
Son Katman Düğüm Sayısı	512	74.80
	128	75.70
	32	71.80

5.5.2 ESA Model-1 ile Elde Edilen Sonuçlar

Chromagram ile elde edilen sonuçlar, Tablo-5.32 ve Tablo-5.33 ile özetlenmiştir.

Tablo 5.32 ESA Model-1 başarımlar yüzdeleri (Chromagram, 30 saniye)

Deney No	Doğruluk
Deney-1	16.66
Deney-2	26.67
Deney-3	24.44
Ortalama	22.59

Tablo 5.33 ESA Model-1 başarımlar yüzdeleri (Chromagram, 6 saniye)

Deney No	Doğruluk
Deney-1	32.22
Deney-2	43.11
Deney-3	32.88
Ortalama	36.07

Chroma CQT özellik çıkarma yöntemi müzik ton özelliklerine yoğunlaşsa da tınısal özellikleri kaybettiği için müzik türü tespitinde başarısız olmuştur. Şarkıların farklı tınılardan meydana gelen seslerin harmonisinden oluşması ve tını ile ilgili özelliklerin bu yöntemde göz ardı edilmesi nedeniyle bu yöntemin başarımlarını daha düşük olmuştur. Müzikler aynı notalardan oluşsa bile tını ile ilgili özelliklerinden dolayı farklı türlerde olabilir.

Spektrogram ile elde edilen sonuçlar, Tablo-5.34 ve Tablo-5.35 ile özetlenmiştir.

Tablo 5.34 ESA Model-1 başarımlar yüzdeleri (Spektrogram, 30 saniye)

Deney No	Doğruluk
Deney-1	57.56
Deney-2	62.22
Deney-3	60.44
Ortalama	60.07

Tablo 5.35 ESA Model-1 başarımlı yüzdeleri (Spektrogram, 6 saniye)

Deney No	Doğruluk
Deney-1	78.11
Deney-2	79.20
Deney-3	76.89
Ortalama	78.06

Tablo-5.34 ve Tablo-5.35 incelendiğinde, Chromagram ile elde edilen sonuçlara göre başarımda dikkate değer artış olmuştur. Bunun en önemli sebebi, spektrogram'ın müziğin karakteristik özelliklerine ait daha çok bilgiyi içermesidir.

5.5.3 ESA Model-2 ile Elde Edilen Sonuçlar

Chromagram ile elde edilen sonuçlar, Tablo-5.36 ve Tablo-5.37 ile özetlenmiştir.

Tablo 5.36 ESA Model-2 başarımlı yüzdeleri (Chromagram, 30 saniye)

Deney No	Doğruluk
Deney-1	31.11
Deney-2	28.89
Deney-3	34.44
Ortalama	31.48

Tablo 5.37 ESA Model-2 başarımlı yüzdeleri (Chromagram, 6 saniye)

Deney No	Doğruluk
Deney-1	43.55
Deney-2	42.44
Deney-3	48.44
Ortalama	44.81

Tablo-5.36 ve Tablo-5.37 incelendiğinde, Chromagram ile elde edilen sonuçlarda ESA Model-1'e göre yaklaşık %10'luk bir artış olmuştur. ESA Model-2'deki boyut indirgemeyi sağlayan kodlayıcı sisteminin daha çok ayırt edici özellikler çıkarması bunun göstergesidir.

Spektrogram ile elde edilen sonuçlar, Tablo-5.38 ve Tablo-5.39 ile özetlenmiştir.

Tablo 5.38 ESA Model-2 başarımlı yüzdeleri (Spektrogram, 30 saniye)

Deney No	Doğruluk
Deney-1	82.67
Deney-2	76.44
Deney-3	81.77
Ortalama	80.29

Tablo 5.39 ESA Model-2 başarımlı yüzdeleri (Spektrogram, 6 saniye)

Deney No	Doğruluk
Deney-1	93.02
Deney-2	85.96
Deney-3	90.13
Ortalama	89.70

Tablo-5.38 ve Tablo-5.39 incelendiğinde, ESA Model-2'nin 30 saniyelik müziklerde %20 , 6 saniyelik müziklerde %10'luk bir artışa katkı sağladığı gözlemlenmiştir.

Tüm tablolar dikkate alındığında derin öğrenme modelinin sınıflandırma kabiliyetinin artması için eklenen bırakma (dropout), Auto-Encoder gibi yapıların kullanılması kaçınılmazdır. Diğer yandan, derin öğrenmede başarı için gerekli olan veri sayısının fazla olmasının önemi her iki modelden bağımsız olarak iki tür veri için de gözlemlenmektedir. Bu sayede daha ayırt edici özellikler elde edebilmek mümkün olacak ve böylece daha başarılı sınıflandırma yapılabilecektir.

5.6 Hibrit Olarak Elde Edilen Sonuçlar

Bu bölümde önerilen ESA modellerinden birisi olan ESA Model-1'in son yoğun katmanından elde edilen özelliklerin, daha karmaşık yapıda olan farklı sınıflandırma algoritmalarıyla sınıflandırılması sonucunda elde edilen başarımlı sonuçları özetlenmiştir. Elde edilen sonuçlar Tablo 5.40 ile özetlenmiştir. ESA model ile elde edilen ve ayırt ediciliği geleneksel yöntemlere kıyasla daha fazla olan akustik özelliklerin daha karmaşık yapıda olan sınıflandırma algoritmaları ile sınıflandırılmasının başarımlı artırdığı sonuçlarda görülmektedir.

Tablo 5.40 ESA Model-1 ve (MLP / RF / KNN / SVM) başarımları yüzdeleri

Algoritma	Ortalama Doğruluk	
	5 Fold	10 Fold
MLP	91.20	92.20
RF	87.70	88.70
KNN	88.60	87.60
SVM Poly=3	97.60	97.90

5.7 Müzik Türü Sınıflandırma Sonuçları Özeti

Önceki bölümlerde elde edilen başarımların performansları, etiketli bir veri seti olan GTZAN veri setinde yapılan testler sonucunda kaydedilmiştir. Kullanılan yöntemler ve algoritmalar, sadece müzik önerisi yapma amacıyla Turkcell'e ait verilerde de kullanılacaktır.

5.8 Müzik Önerisi Sonuçları

Bu bölümde akustik özellikleri elde edilen müzik önerisi sonuçlarına yer verilecektir. Bu kapsamda elde edilen sonuçlar, aynı türden müziğin önerilmesi esasına dayanmaktadır. Aşağıda bu kapsamda elde edilen sonuçlara ait tablolar yer almaktadır.

Tablo 5.41 Müzik öneri yüzdeleri (ESA Model-1 ve işaret işleme özellikleri)

Tür	Sayısal İşaret İşleme		ESA Model-1	
	İlk 5 Öneri	İlk 10 Öneri	İlk 5 Öneri	İlk 10 Öneri
Blues	60.00	48.00	63.00	49.70
Classical	88.00	90.00	90.80	87.70
Country	40.00	50.00	56.80	49.40
Disco	48.00	40.00	53.60	45.90
Hip-hop	72.00	60.00	64.20	57.30
Jazz	52.00	42.00	65.00	52.70
Metal	84.00	76.00	79.40	75.10
Pop	60.00	50.00	78.40	73.80
Reggae	24.00	26.00	57.20	48.90
Rock	60.00	50.00	49.80	42.20

Tablo 5.42’de çalışmamızda önerilen ESA Model-1 ile farklı bir öneri sistemi başarımı ölçülmüştür. Buna göre veri setindeki her bir müzik için N sayıda müzik önerisi yapılmıştır. Bu N değerine göre gelen önerilerin seçilen müzikle aynı türde olan müziklerin yüzdesi hesaplanmıştır. Sonrasında bütün türlere ait müziklere yapılan bu önerilerden gelen başarımların ortalaması alınmıştır. Elde edilen öneri sonuçları farklı N değerinde (5, 10, 20) öneriye göre tabloda özetlenmiştir [88] .

Tablo 5.42 N sayıda öneriye müzik öneri yüzdeleri ($N = 5, 10, 20$)

Müzik Türü	Öneri Sayısı		
	5	10	20
Pop	85.30	84.60	80.70
Metal	90.10	86.90	85.10
Disco	70.40	66.90	60.00
Blues	82.90	71.40	56.40
Reggea	72.80	67.50	61.40
Classical	98.00	96.80	94.70
Rock	58.20	51.70	42.80
Hiphop	77.90	73.50	65.40
Country	82.70	77.10	70.55
Jazz	85.80	81.00	72.30

Tablo 5.43 ve Tablo 5.44’de Turkcell’in veri tabanındaki 17000 müzikten bu projede kullanılan özellik çıkarma yöntemleri yoluyla elde edilmiş akustik özellikler kullanılarak K en yakın komşu yöntemiyle oluşturulan öneri sisteminden elde edilmiş çıktılar sunulmuştur. Tablodaki ilk satır seçilen şarkıyı, diğer satırlar önerilen şarkıları göstermektedir.

Tablo 5.43 Verilen şarkı için öneriler - 1

Id	Şarkıcı İsmi	Şarkı Adı
TRARG1900021	Merve Özbey	Yaramızda Kalsın
TRARG1900017	Merve Özbey	Kalır mı Günah
TRARG1900018	Merve Özbey	Kül
TR2811600046	Tuğçe Haşimoğlu	Ayrılık Şarkısı
TR0070700062	Sibel Can	Benim Yerime de Sev
TRARG1900016	Merve Özbey	Eyvallah
TRA160900011	Sıla	Yara Bende
TR0011900047	Bengü	Yazık
TR9491200117	Burcu Güneş	Hasret Kaldım
TR0011100089	Ömür Gedik, Halil Sezai	Paramparça

Tablo 5.44 Verilen şarkı için öneriler - 2

Id	Şarkıcı İsmi	Şarkı Adı
TR0681700362	Sagopa Kajmer	366.Gün
TR0681700371	Sagopa Kajmer	366.Gün - 2017 Edit
TR0441722101	Hadise	Aşk Dediğin
FR96X1706667	Ezman Sterk	E Kalima Bıkın
TR0061100109	Aydilge	Sorma
QM4DW1665224	No.1	Ateş ve Barut
TR0531705867	Berkay	Açık Ara
TR0531913072	Ufuk Çakır	Senin Yüzünden
FRX201860502	Rauf	Hevesim Kaçtı
LBA131401470	Yara	Ma Baaref

Öneri testleri sırasında, her bir şarkının kendi türü hariç en çok tavsiye edilen türü Tablo 5.45 içinde verilmiştir. Eğer aynı yüzdede birden fazla türde öneri yapılmışsa her türün bilgisi verilmektedir.

Tablo 5.45 Önerilen müziklerin farklı türde olma yüzdeleri

Müzik Türü	Test 1	Test 2	Test 3	Test 4	Test 5
Blues	Rock & Jazz (20%)	Jazz (20%)	Jazz & Country (20%)	Rock & Country (10%)	Metal (30%)
Classical	Jazz (10%)	Country & Jazz (10%)	Country & Disco & Jazz & Rock (20%)	- (0%)	Country (30%)
Country	Rock (40%)	Jazz (40%)	Blues (20%)	Jazz (30%)	Rock (20%)
Disco	Rock & Country (20%)	Hiphop & Metal & Pop & Rock (10%)	Blues (20%)	Metal & Rock (10%)	Rock & Blues (20%)
Hiphop	Disco (10%)	Reggae (20%)	Reggae (40%)	Metal & Disco (10%)	Reggae (60%)
Jazz	Classical (30%)	Country (30%)	Classical (20%)	Classical (20%)	Classical (50%)
Metal	Hiphop (10%)	Hiphop (10%)	Rock & Disco (10%)	Rock & Hiphop (20%)	Hiphop (30%)
Pop	Disco (50%)	Disco (20%)	Disco (40%)	Reggae (30%)	Hiphop (60%)
Reggae	Disco & Metal & Hiphop (20%)	Hiphop (10%)	Country (30%)	Hiphop (20%)	Pop (50%)
Rock	Country (20%)	Metal (30%)	Disco (40%)	Blues & Reggae & Disco & Metal & Jazz (10%)	Disco (30%)

5.9 Büyük Veri Teknolojileri İle Özellik Çıkarımı Süre başarımları

Bu bölümde, özellik çıkarma yöntemleri büyük veri teknolojilerinden Apache Spark ile gerçekleştirilmiştir. Farklı sayıda işlem elemanı ile alınan akustik özellikler zaman açısından karşılaştırılmıştır. Apache Spark kütüphanesi kullanırken temel yaklaşımımız, veri tabanındaki müziklerin tamamını farklı sayıda alt kümelere bölerek her alt kümeye farklı bir işlem elemanı ile özellik çıkarımı yapmak olmuştur. Farklı işlem elemanlarına göre spektral ağırlık merkezi yönteminin ve tüm işaret işleme yöntemlerinin sonuç verme süresi Tablo 5.46 ile gösterilmiştir.

Tablo 5.46 Apache Spark kullanımı deney sonuçları

Tek İşlemcili ve Çok Düğümlü Başarım Karşılaştırma Süreleri (saniye)			
DENEY-1		DENEY-2	
Ortam	Spektral Ağırlık Merkezi	Ortam	Tüm İşaret İşleme Yöntemleri
Tek İşlemci	569	Tek İşlemci	1197
Spark - 4	70	Spark - 1	1171
Spark - 6	67	Spark - 2	673
Spark - 8	67	Spark - 3	542
Spark - 10	72	Spark - 4	535
Spark - 16	65	Spark - 5	516
Spark - 18	68	Spark - 7	552
Spark - 24	67	Spark - 8	546

6

SONUÇ VE ÖNERİLER

6.1 Sonuç

Tez çalışması kapsamında, müzik veri setlerinden akustik özellikler çıkarımı, bu akustik özellikleri kullanarak müzik önerisi ve müzik türü sınıflandırması yapılmıştır. Bu işlemler yapılırken kullanılan yöntemlerle birden fazla açıdan karşılaştırma yapılmıştır. Çıkarılan özelliklerin her birinin sınıflandırma sonucu üzerindeki etkisinin incelenmesi, kullanılan sınıflandırıcıların meta-parametrelerinin sonuçlarının başarımlarına olan etkisinin karşılaştırılması, sınıflandırıcı algoritmalarının doğruluk, kesinlik gibi ölçütler açısından karşılaştırılması, kullanılan GTZAN veri setinde sınıflandırması en kolay ve en zor müzik türlerinin tespiti, özellik çıkarımı için büyük veri teknolojilerinin kullanımının gerekliliği yapılan karşılaştırmaların özetidir.

Akustik özellikler açısından yapılan karşılaştırmalar sonucunda; MFCC özelliğinin akustik özellikler olarak en fazla etkiye sahip olduğu gözlenmiştir. MFCC özelliğinin türevinin kullanımının müzik türü sınıflandırması üzerinde yaklaşık %3 kadar başarımlar artışı sağladığı gözlenmiştir. Gruplandırılmış özellikler arasında en başarılı sonuçlar MFCC ve Spektral özellikler kullanılarak elde edilmiştir. Sonuç olarak, tüm spektral özellikler birlikte kullanılarak en yüksek başarımlar oranına ulaşılmıştır. Sayısal işaret işleme yöntemlerinden elde edilen verilerde MFCC özellik türüne kümeleme yapmadan özellik çıkarımı yapmanın kümeleme yaparak özellik çıkarımı yapmaya göre daha başarılı sonuçlar verdiği gözlemlenmiştir. Spektrogram üzerinde farklı pencere fonksiyonu, pencere boyutu ve örtüşme miktarı kullanmanın yapılan deneylerde %2-%3 oranında bir iyileştirme sağladığı görülmüştür. Müziğe ait ton özelliklerinin elde edilmesinde kullanılan Chroma STFT, Chroma CQT ve Tonnetz gibi özelliklerin müziğin tınısal özelliklerini göz ardı ettikleri için müzik türü başarısında Spektrograma veya Mel-Spektrograma göre daha başarısız oldukları gözlenmiştir. Dalgacık dönüşümü ile elde edilen özelliklere gelince, dalgacık fonksiyonunun ve seviye değişkeninin değişimlerinin başarımlara bir etkileri olmamıştır. Diğer yandan, önerilen farklı mimarilerdeki derin öğrenme yöntemleriyle elde edilen akustik özelliklerin sayısal işaret işleme yöntemleriyle elde edilenlere göre başarıyı önemli

miktarda arttırdığı gözlemlenmiştir. Özellikle Auto-Encoder özelliğine sahip olan ESA Model-2'nin Auto-Encoder içermeyen ESA Model-1'e göre daha ayırt edici akustik özellikler verebildiği gözlemlenmiştir.

Sınıflandırma algoritmalarının meta-parametreleri açısından yapılan karşılaştırmalar sonucunda; bu parametreler farklı değerlerle test edilmiş ve başarımlar sonuçları birbirleri ile karşılaştırılmıştır. Karşılaştırma sonucunda, MLP yönteminde kullanılan aktivasyon fonksiyonunun başarı oranı üzerindeki etkisi çok fazla olmamıştır. K-NN yönteminde k değerleri birbiriyle karşılaştırılmış ve k değeri 7 olarak verildiğinde, algoritmanın en yüksek başarımlarına sahip olduğu görülmüştür. SVM yönteminde, kullanılan çekirdek fonksiyonunun iyileştirilen parametreler sonucunda başarıya katkı sağladığı görülmüştür. Diğer yandan, derin öğrenme ile elde edilen özellik veri setinin sınıflandırılmasında optimize edilmiş c ve γ değerlerinin test başarımlarını kayda değer miktarda arttırdığı görülmüştür.

Sınıflandırma algoritmalarının performanslarının karşılaştırılması için farklı testler yapılmıştır. Sonuçlardan elde edilen verilere göre; en başarılı sonuçların birbirine çok yakın olduğu ve bu başarılı sonuçların SVM, LDA, ve MLP yöntemlerinden elde edildiği görülmüştür. Özellikle SVM için rbf çekirdek fonksiyonu ve bu fonksiyonun parametrelerinin iyileştirilmesi ile %83.50 başarımlar elde edilebilmiştir. Bu algoritmalar daha karmaşık sınıflandırıcılar oldukları için, Naive bayes, k-NN gibi daha basit olan sınıflandırma algoritmalarına göre nispeten daha başarılı sonuçlar vermiştir. Derin öğrenme ile elde edilen sonuçlar incelendiğinde Auto-Encoder özelliği bulunan ESA-2 modelinin özellik çıkarımında, sınıflandırmada ve müzik önerisinde çok daha başarılı sonuçlar verdiği gözlemlenmiştir. Özellikle hazır olarak kullanılan ResNet ve benzeri derin ağların karmaşıklığı ile kıyaslandığında; daha basit bir mimariye sahip olan ESA Model-1 derin sinir ağı, müzik türü sınıflandırma ve özellik çıkarımı problemi için SVM ile birlikte kullanıldığında en yüksek başarımları vermiştir.

Müzik sınıflandırması için kullanılan GTZAN veri setindeki müziklerin türleri açısından başarımlar incelendiğinde; 10 farklı müzik türünden (blues, klasik, ülke, disko, hiphop, caz, metal, pop, reggae, rock), rock-country ve reggae-hiphop birbirine en çok karıştırılan müzikler olarak gözlemlenmiştir. Diğer yandan klasik müzik en az karıştırılan müzik türü olarak gözlemlenmiştir.

Özellik çıkarımı aşamasında yapılan çalışmalar sonucunda, özellik çıkarımı yöntemlerinin algoritmik karmaşıklığının fazla olması ve kullanılan veri setlerinin çok fazla sayıda örnekten oluşması nedeniyle büyük veri teknolojilerinden faydalanmanın eş zamanlı işlem yapma kabiliyetini artırdığı ve dolayısıyla yanıt süresinin kayda değer miktarda azaldığı görülmüştür. Özellikle yüksek boyutlu veri tabanlarında bulunan ve

milyonlarla ifade edilen sayıdaki müziklerin özelliklerinin çıkarılması için büyük veri teknolojilerinin kullanılması kaçınılmaz olacaktır.

Sonuç olarak, sayısal işaret işleme ve derin öğrenme yöntemleri kullanılarak, müziklerin akustik özellikleri çıkarılmış ve bu özellikler müzik türlerini sınıflandırmak ve müzik önerisi yapmak için kullanılmıştır. Veri setinde yer alan müziklerin akustik özelliklerine göre benzerlikleri hesaplanmış ve en benzer N sayıda müzik önerilmiştir.

6.2 Öneriler

Bu çalışmada üst-verisi bulunmayan müzikler üzerinde akustik özellik çıkarımı ve müzik önerisi yapılmıştır. Gelecek çalışmalarda üst-verisi bulunan, müzik dinleme listesinde dinleme geçmişini olan kullanıcılardan alınan verilerle yapılan müzik öneri sistemleri akustik özellikler kullanan müzik önerisi sistemleri ile birlikte hibrit olarak kullanılabilir. Bu sayede akustik özellikler kullanan müzik öneri sistemlerinin diğer öneri sistemlerine olan üstünlükleri ve zayıflıkları analiz edilebilecektir.

Bu çalışmadan elde edilen bilgi birikimi sayesinde, müzik veri setlerinden akustik özellikler elde edilerek, farklı türde problemlere içerik tabanlı çözümler üretilebilir. Özellikle müzik telif hakları söz konusu olduğunda, bu yöntemlerle elde edilecek akustik özellikler sayesinde eser sahiplerinin haklarını korumak amacıyla müzik benzerlik analizleri yapılabilir. Ayrıca kültürel, yöresel veya dönemsel olarak sınıflandırılan müziklerden karakteristik özelliklerin çıkarılması ve müzik analizi bu başlık altında ele alınabilecek çalışmalardan bir diğeridir. Sayısal işaret işleme ve derin öğrenme konularının yeniliklere ve gelişimlere açık olması, bu konularda meydana gelecek gelişmelerin var olan problemlere daha iyi çözümler üretilebilmesini kazanan bilgi birikimi sayesinde mümkün olacaktır.

- [1] G. Giacaglia. (2019). Spotify's recommendation engine, [Online]. Available: <https://medium.com/datadriveninvestor/behind-spotify-recommendation-engine-a9b5a27a935> (visited on 11/02/2019).
- [2] M. Hagglade, Y. Hong, K. Kao, "Music genre classification," *Department of Computer Science, Stanford University*, 2011.
- [3] G. Tzanetakis P. Cook, "Musical genre classification of audio signals," *IEEE Transactions on speech and audio processing*, vol. 10, no. 5, pp. 293–302, 2002.
- [4] D. Grzywczak G. Gwardys, "Deep image features in music information retrieval," *Intl Journal of Electronics and Telecommunications*, vol. 60, no. 4, pp. 321–326, 2014.
- [5] T. Li G. Tzanetakis, "Factors in automatic musical genre classification of audio signals," in *2003 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (IEEE Cat. No. 03TH8684)*, IEEE, 2003, pp. 143–146.
- [6] A. Nasridinov Y.-H. Park, "A study on music genre recognition and classification techniques," *International Journal of Multimedia and Ubiquitous Engineering*, vol. 9, no. 4, pp. 31–42, 2014.
- [7] A. Rosner, B. Schuller, B. Kostek, "Classification of music genres based on music separation into harmonic and drum components," *Archives of Acoustics*, vol. 39, no. 4, pp. 629–638, 2014.
- [8] T. Lidy A. Rauber, "Evaluation of feature extractors and psycho-acoustic transformations for music genre classification," in *ISMIR*, 2005, pp. 34–41.
- [9] T. Li, M. Ogihara, Q. Li, "A comparative study on content-based music genre classification," in *Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval*, 2003, pp. 282–289.
- [10] T. Li M. Ogihara, "Music genre classification with taxonomy," in *Proceedings. (ICASSP'05). IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005.*, IEEE, vol. 5, 2005, pp. v–197.
- [11] A. Meng, P. Ahrendt, J. Larsen, L. K. Hansen, "Temporal feature integration for music genre classification," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 15, no. 5, pp. 1654–1664, 2007.
- [12] C.-C. M. Yeh Y.-H. Yang, "Supervised dictionary learning for music genre classification," in *Proceedings of the 2nd ACM International Conference on Multimedia Retrieval*, 2012, pp. 1–8.
- [13] C. Lee, J. Shih, K. Yu, H. Lin, "Automatic music genre classification based on modulation spectral analysis of spectral and cepstral features," *IEEE Transactions on Multimedia*, vol. 11, no. 4, pp. 670–682, 2009.

- [14] H. Lee, P. Pham, Y. Largman, A. Y. Ng, "Unsupervised feature learning for audio classification using convolutional deep belief networks," in *Advances in neural information processing systems*, 2009, pp. 1096–1104.
- [15] Y. Costa, L. Oliveira, A. Koerich, F. Gouyon, J. Martins, "Music genre classification using lbp textural features," *Signal Processing*, vol. 92, no. 11, pp. 2723–2737, 2012, ISSN: 0165-1684. DOI: <https://doi.org/10.1016/j.sigpro.2012.04.023>. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0165168412001478>.
- [16] I. Panagakis, E. Benetos, C. Kotropoulos, "Music genre classification: A multilinear approach," in *ISMIR*, 2008, pp. 583–588.
- [17] B. Shao, D. Wang, T. Li, M. Ogihara, "Music recommendation based on acoustic features and user access patterns," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 17, no. 8, pp. 1602–1611, 2009.
- [18] Y. Song, S. Dixon, M. Pearce, "A survey of music recommendation systems and future perspectives," in *9th International Symposium on Computer Music Modeling and Retrieval*, vol. 4, 2012, pp. 395–410.
- [19] H. Bahuleyan, *Music genre classification using machine learning techniques*, 2018. arXiv: 1804.01149 [cs.SD].
- [20] S. Sigtia S. Dixon, "Improved music feature learning with deep neural networks," in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2014, pp. 6959–6963.
- [21] A. Holzapfel Y. Stylianou, "Musical genre classification using nonnegative matrix factorization-based features," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 16, no. 2, pp. 424–434, 2008.
- [22] S.-H. Shin, H.-W. Yun, W.-J. Jang, H. Park, "Extraction of acoustic features based on auditory spike code and its application to music genre classification," *IET Signal Processing*, vol. 13, no. 2, pp. 230–234, 2018.
- [23] B. Boashash, *Time-frequency signal analysis and processing: a comprehensive reference*. Academic Press, 2015.
- [24] J. H. McClellan, R. Schafer, M. Yoder, *Dsp first*. Prentice Hall Press, 2015.
- [25] A. Lerch, *An introduction to audio content analysis: Applications in signal processing and music informatics*. Wiley-IEEE Press, 2012.
- [26] D. Gerhard, *Audio signal classification: History and current techniques*. Citeseer, 2003.
- [27] C. Weihs, D. Jannach, I. Vatulkin, G. Rudolph, *Music data analysis: Foundations and applications*. Chapman and Hall/CRC, 2016.
- [28] *What is timbre*. [Online]. Available: <http://www.simplifyingtheory.com/timbre/>.
- [29] E. Gopi, *Algorithm collections for digital signal processing applications using Matlab*. Springer Science & Business Media, 2007.
- [30] *Notes on music information retrieval*. [Online]. Available: <https://musicinformationretrieval.com/>.

- [31] M. Müller, *Fundamentals of music processing: Audio, analysis, algorithms, applications*. Springer, 2015.
- [32] D. W. Kammler, *A first course in Fourier analysis*. Cambridge University Press, 2007.
- [33] L. Cohen, "Time-frequency distributions-a review," *Proceedings of the IEEE*, vol. 77, no. 7, pp. 941–981, 1989.
- [34] J. Stroobach, *Digital dtmf tone detector*, US Patent 5,119,322, 1992.
- [35] R. Bachu, S. Kopparthi, B. Adapa, B. D. Barkana, "Voiced/unvoiced decision for speech signals based on zero-crossing rate and energy," in *Advanced Techniques in Computing Sciences and Software Engineering*, Springer, 2010, pp. 279–282.
- [36] W. user: Ktims. (2006). A simple diagram showing a zero crossing point of a simple signal, [Online]. Available: https://upload.wikimedia.org/wikipedia/commons/thumb/0/03/Zero_crossing.svg/1200px-Zero_crossing.svg.png.
- [37] B. M. Meinard Müller. (). Energy and rmse, [Online]. Available: <https://musicinformationretrieval.com/energy.html>.
- [38] U. Nam. (2001). Spectral centroid, [Online]. Available: <https://ccrma.stanford.edu/~unjung/AIR/areaExam.pdf>.
- [39] D. L. Wessel, "Timbre space as a musical control structure," *Computer music journal*, pp. 45–52, 1979.
- [40] (2018). Spectral features, [Online]. Available: https://musicinformationretrieval.com/spectral_features.html (visited on 12/03/2018).
- [41] J. D. Cook. (2016). Spectral flatness, [Online]. Available: <https://www.johndcook.com/blog/2016/05/03/spectral-flatness/>.
- [42] D.-N. Jiang, L. Lu, H.-J. Zhang, J.-H. Tao, L.-H. Cai, "Music type classification by spectral contrast feature," in *Multimedia and Expo, 2002. ICME'02. Proceedings. 2002 IEEE International Conference on*, IEEE, vol. 1, 2002, pp. 113–116.
- [43] (2015). A diagram for spectral rolloff, [Online]. Available: <http://www.pvsm.ru/images/2015/02/19/obzor-algoritmov-audioanalitiki-30.jpg>.
- [44] (). Poly features, [Online]. Available: https://librosa.github.io/librosa/generated/librosa.feature.poly_features.html.
- [45] P. D. M. Müller. (2015). Short-time fourier transform and chroma features, [Online]. Available: https://www.audiolabs-erlangen.de/content/05-fau/professor/00-mueller/02-teaching/2016s_ap1/LabCourse_STFT.pdf.
- [46] (2018). Short-time fourier transform and chroma features, [Online]. Available: https://www.audiolabs-erlangen.de/content/05-fau/professor/00-mueller/02-teaching/2016s_ap1/LabCourse_STFT.pdf (visited on 12/05/2018).
- [47] J. C. Brown, "Calculation of a constant q spectral transform," *The Journal of the Acoustical Society of America*, vol. 89, no. 1, pp. 425–434, 1991.

- [48] (), [Online]. Available: http://librosa.github.io/librosa/generated/librosa.feature.chroma_cqt.html.
- [49] C. Harte, M. Sandler, M. Gasser, "Detecting harmonic change in musical audio," in *Proceedings of the 1st ACM workshop on Audio and music computing multimedia*, ACM, 2006, pp. 21–26.
- [50] (2019). Filter bank for mfcc piecewise function question, [Online]. Available: <https://it.mathworks.com/matlabcentral/answers/367893-filter-bank-for-mfcc-piecewise-function-question> (visited on 01/09/2019).
- [51] M. Yusnita, M. Paulraj, S. Yaacob, R. Yusuf, A. Shahriman, "Analysis of accent-sensitive words in multi-resolution mel-frequency cepstral coefficients for classification of accents in malaysian english," *International Journal of Automotive and Mechanical Engineering*, vol. 7, p. 1053, 2013.
- [52] S. Gupta, J. Jaafar, W. W. Ahmad, A. Bansal, "Feature extraction using mfcc," *Signal & Image Processing: An International Journal (SIPIJ)*, vol. 4, no. 4, p. 101, 2013.
- [53] R. Polikar et al., *The wavelet tutorial*, 1996.
- [54] I. Daubechies, *Ten lectures on wavelets*. Siam, 1992, vol. 61.
- [55] N. Aydin, F. Marvasti, H. S. Markus, "Embolic doppler ultrasound signal detection using discrete wavelet transform," *IEEE Transactions on Information Technology in Biomedicine*, vol. 8, no. 2, pp. 182–190, 2004.
- [56] G. Dallas. (2014). Wavelets 4 dummies: Signal processing, fourier transforms and heisenberg, [Online]. Available: <https://georgemdallas.wordpress.com/2014/05/14/wavelets-4-dummies-signal-processing-fourier-transforms-and-heisenberg/>.
- [57] A. Teolis J. J. Benedetto, *Computational signal processing with wavelets*. Springer, 1998, vol. 182.
- [58] Tensorflow. (2019). F.keras.layers.conv2d, [Online]. Available: https://www.tensorflow.org/api_docs/python/tf/keras/layers/Conv2D (visited on 02/07/2019).
- [59] (2018). Dropout, [Online]. Available: https://www.tensorflow.org/api_docs/python/tf/keras/layers/Dropout (visited on 10/05/2018).
- [60] DATAI, *Convolutional neural network (cnn) tutorial*, [Online; accessed 21-April-2020], 2018. [Online]. Available: <https://www.kaggle.com/kanncaa1/convolutional-neural-network-cnn-tutorial>.
- [61] (2018). Destek vektör makinesi, [Online]. Available: https://tr.wikipedia.org/wiki/Destek_vekt%C3%B6r_makinesi (visited on 12/06/2018).
- [62] T. Joachims, "Text categorization with support vector machines: Learning with many relevant features," in *European conference on machine learning*, Springer, 1998, pp. 137–142.
- [63] J. A. Suykens J. Vandewalle, "Least squares support vector machine classifiers," *Neural processing letters*, vol. 9, no. 3, pp. 293–300, 1999.

- [64] S. Upadhyaya R. Ramsankaran, "Support vector machine (svm) based rain area detection from kalpana-1 satellite data," 2014.
- [65] P-N. Tan, M. Steinbach, V. Kumar, *Introduction to data mining*. Pearson Education India, 2016.
- [66] G. H. John P. Langley, "Estimating continuous distributions in bayesian classifiers," in *Proceedings of the Eleventh conference on Uncertainty in artificial intelligence*, Morgan Kaufmann Publishers Inc., 1995, pp. 338–345.
- [67] (2018). Decision tree - classification, [Online]. Available: https://www.saedsayad.com/decision_tree.htm (visited on 12/06/2018).
- [68] Ç. Elmas, *Yapay Sinir Ağları*. Seçkin Yayıncılık, 2003.
- [69] T. Hastie, R. Tibshirani, J. Friedman, *The elements of statistical learning: data mining, inference, and prediction*. Springer Science & Business Media, 2009.
- [70] E. Öztemel, *Yapay Sinir Ağları*. Papatya Bilim, 2003.
- [71] (2018). K nearest neighbors algorithm, [Online]. Available: https://en.wikipedia.org/wiki/K-nearest_neighbors_algorithm (visited on 12/06/2018).
- [72] A. Bronshtein. (2017). A quick introduction to k-nearest neighbors algorithm, [Online]. Available: <https://medium.com/@adi.bronshtein/a-quick-introduction-to-k-nearest-neighbors-algorithm-62214cea29c7>.
- [73] N. Donges. (2018). The random forest algorithm, [Online]. Available: <https://towardsdatascience.com/the-random-forest-algorithm-d457d499ffcd>.
- [74] L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.
- [75] (2018). Random forest simple explanation, [Online]. Available: <https://medium.com/@williamkoehrsen/random-forest-simple-explanation-377895a60d2d> (visited on 12/06/2018).
- [76] S. Raschka. (2014). Linear discriminant analysis, [Online]. Available: http://sebastianraschka.com/Articles/2014_python_lda.html.
- [77] M. I. Kamel, M. J. Alhaddad, H. M. Malibary, K. Thabit, F. Dahlwi, E. A. Alsaggaf, A. A. Hadi, "Eeg based autism diagnosis using regularized fisher linear discriminant analysis," *International Journal of Image, Graphics and Signal Processing*, vol. 4, no. 3, p. 35, 2012.
- [78] B. Guo. (2017). Pca (4) : Lda (linear discriminant analysis), [Online]. Available: <https://algorithmsdatascience.quora.com/PCA-4-LDA-Linear-Discriminant-Analysis>.
- [79] J. A. Hartigan, *Clustering algorithms*, ser. Wiley series in probability and mathematical statistics. New York, NY: Wiley, 1975. [Online]. Available: <http://cds.cern.ch/record/105051>.
- [80] J. D. Banfield A. E. Raftery, "Model-based gaussian and non-gaussian clustering," *Biometrics*, pp. 803–821, 1993.
- [81] D. Pedamonti, "Comparison of non-linear activation functions for deep neural networks on mnist classification task," *arXiv preprint arXiv:1804.02763*, 2018.

- [82] D. Jannach, M. Zanker, A. Felfernig, G. Friedrich, *Recommender systems: an introduction*. Cambridge University Press, 2010.
- [83] N. Orio, *Music retrieval: A tutorial and review*. now publishers Inc, 2006.
- [84] J. Leben. (2015). Gtzan genre collection, [Online]. Available: [http : / / marsyas . info / downloads / datasets . html](http://marsyas.info/downloads/datasets.html).
- [85] (). Gtzan license, [Online]. Available: [http : / / marsyas . info / about / license . html](http://marsyas.info/about/license.html) (visited on 11/03/2019).
- [86] (). Simple guide to confusion matrix terminology, [Online]. Available: [http : / / www . dataschool . io / simple - guide - to - confusion - matrix - terminology /](http://www.dataschool.io/simple-guide-to-confusion-matrix-terminology/) (visited on 2018).
- [87] A. Elbir, H. O. İlhan, G. Serbes, N. Aydın, “Short time fourier transform based music genre classification,” in *2018 Electric Electronics, Computer Science, Biomedical Engineerings’ Meeting (EBBT)*, 2018, pp. 1–4.
- [88] A. Elbir N. Aydın, “Music genre classification and music recommendation by using deep learning,” *Electronics Letters*, vol. 56, no. 12, pp. 627–629, 2020.

A.1 ESA Model-1 Kaynak Kodu

ESA Model-1'e ait kaynak kodu aşağıda verilmiştir.

```
def createModel(X_train):
    model = Sequential()
    #
    model.add(Conv2D(filters = 16, kernel_size = (3,3),padding = 'Same',
                     activation = 'relu', input_shape =
                     ↪ (X_train.shape[1],X_train.shape[2],X_train.shape[3])))
    model.add(MaxPool2D(pool_size=(2,2)))
    model.add(Dropout(0.25))
    #
    model.add(Conv2D(filters = 32, kernel_size = (3,3),padding = 'Same',
                     activation = 'relu'))
    model.add(MaxPool2D(pool_size=(2,2)))
    model.add(Dropout(0.25))
    #
    model.add(Flatten())
    model.add(Dense(256, activation = "relu"))
    model.add(Dropout(0.5))
    model.add(Dense(10, activation = "softmax"))

    # Define the optimizer
    #optimizer = tf.compat.v1.train.AdamOptimizer(1e-3, epsilon=1e-4)
    optimizer = tf.keras.optimizers.Adam(lr=0.001, beta_1=0.9,
    ↪ beta_2=0.999)

    model.compile(optimizer = optimizer , loss =
    ↪ "categorical_crossentropy", metrics=["accuracy"])

    return model
```

A.2 ESA Model-2 Kaynak Kodu

ESA Model-2'e ait kaynak kodu aşağıda verilmiştir.

```
def createModel(X_train):
    model = Sequential()
    #
    model.add(Conv2D(filters = 16, kernel_size = (3,3),padding = 'Same',
        activation = 'relu', input_shape =
        ↪ (X_train.shape[1],X_train.shape[2],X_train.shape[3])))
    model.add(MaxPool2D(pool_size=(2,2)))
    model.add(Dropout(0.25))
    #
    model.add(Conv2D(filters = 32, kernel_size = (3,3),padding = 'Same',
        activation = 'relu'))
    model.add(MaxPool2D(pool_size=(2,2)))
    model.add(Dropout(0.25))
    ##decode
    model.add(Conv2D(filters = 16, strides=(2,2), kernel_size =
    ↪ (3,3),padding = 'Same',
        activation = 'relu'))
    #model.add(UpSampling2D((2,2)))
    model.add(Conv2D(filters = 16, kernel_size = (3,3),padding = 'Same',
        activation = 'relu'))
    model.add(UpSampling2D((2,2)))
    ##
    # fully connected
    model.add(Flatten())
    model.add(Dense(256, activation = "relu"))
    model.add(Dropout(0.5))
    model.add(Dense(10, activation = "softmax"))
    #%
    # Define the optimizer
    optimizer = tf.keras.optimizers.Adam(lr=0.001, beta_1=0.9,
    ↪ beta_2=0.999)
    #%
    model.compile(optimizer = optimizer , loss =
    ↪ "categorical_crossentropy", metrics=["accuracy"])
    model.summary()
    return model
```

B

MÜZİK TÜRÜ SINIFLANDIRMA VE MÜZİK ÖNERİSİ UYGULAMALARI

B.1 Müzik Türü Sınıflandırma Uygulaması

Bu bölümde, sistemin ana özelliklerinin uygulandığı grafik arabirim tanıtılmaktadır. Grafik arabirimi, PyQt5 kütüphanesinde bulunan Spyder IDE üzerindeki Qt Designer5 uygulaması kullanılarak Python programlama dilinde tasarlanmıştır. Program şu bölümlerden oluşmaktadır. Veritabanı işlemlerinin yapıldığı menü bölümü, özellik çıkarma için MFCC parametrelerinin seçildiği bölüm, yeni müziğin içe aktarıldığı ve özelliklerinin veritabanına eklendiği bölüm, sınıflandırma ve sınıflandırma için müzik özelliklerinin dikkate alınacağı bölüm. Programda müzik türü sınıflandırma için uygulanacak algoritma seçilir ve model oluşturulur. Bu aşamada eğitim verileri için karmaşıklık matrisi gösterilir. Ayrıca, modelin doğruluk oranının test verileri üzerinde hesaplandığı bölüm, seçilen şarkının türü ve son olarak özellik tablosu bölümünde tüm şarkıların özellikleri görülebilir. Bu çalışmada sayısal işaretleme yöntemleri ile elde edilen özelliklerle eğitim ve test yapılacağı için derin öğrenme eğitimi ile ilgili sınıflandırma arayüzü hazırlanmamıştır. Derin öğrenme modelini kullanan öneri sistemi bir sonraki başlıkta açıklanmıştır.

Music Genre Classification Program

Operations

Classification Part 1 Classification Part 2

CHOOSE MFCC PARAMETERS

Window Type: Hamming

Window Size: 1024

Hop Length: 256

Coefficient Number: 20

CHANGE APPLY

IMPORT NEW SONGS

SELECT PROPERTIES OF MODEL

Features

- ☐ Zero Crossing Rate
- ☐ Spectral Centroid
- ☐ Spectral Bandwidth
- ☐ Spectral Contrast
- ☐ Spectral Rollof
- ☐ RMSE
- ☐ MFCC
- ☐ MFCC Derivative
- ☐ Chroma STFT
- ☐ Wavelet

Classifier

- ☐ Neural Network (MLP)
- ☐ Logistic Regression
- ☐ Random Forest
- ☐ LDA
- ☐ K-NN
- ☐ SVM
- ☐ Naive Bayes

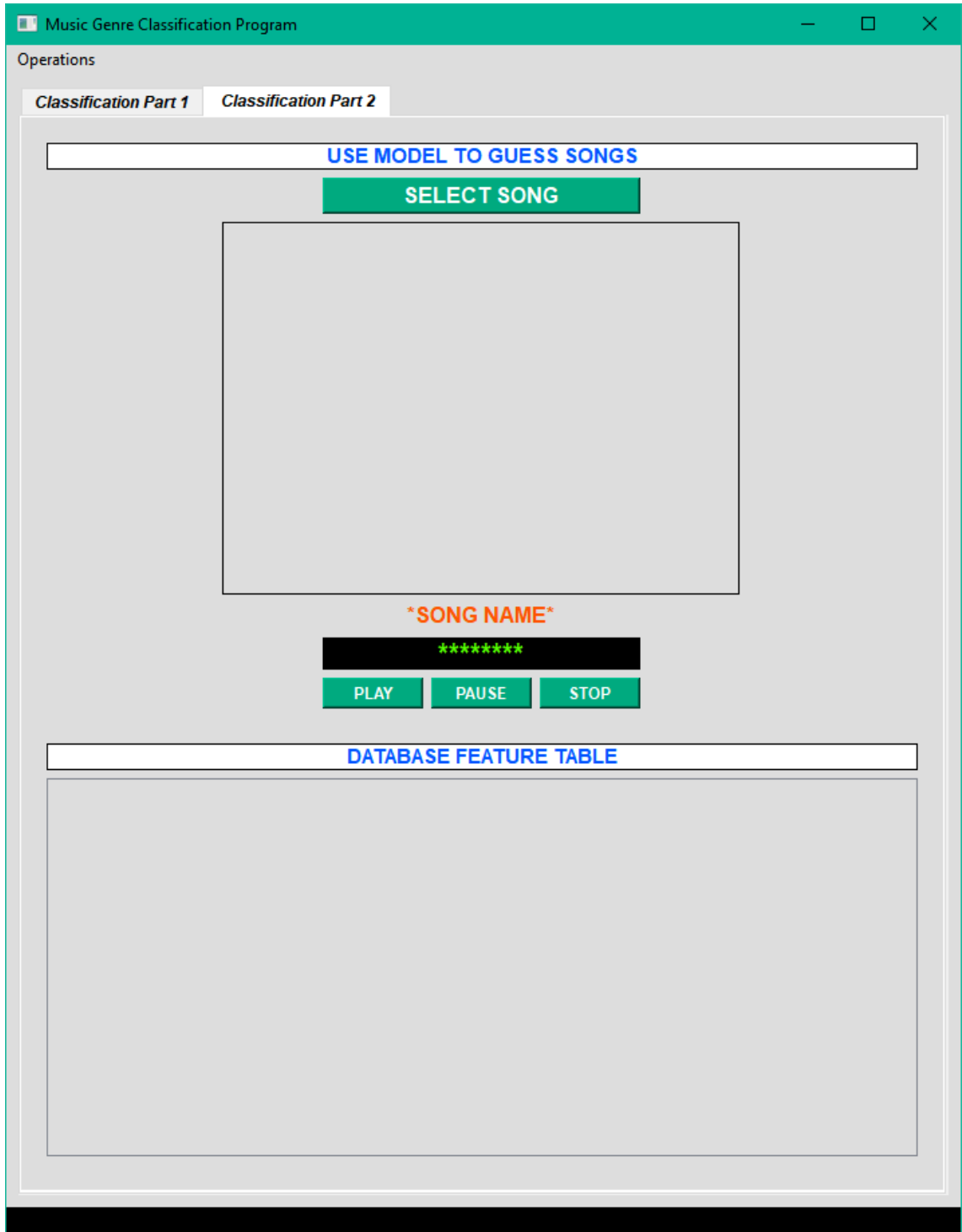
☐ Use the same sets

START CLASSIFICATION

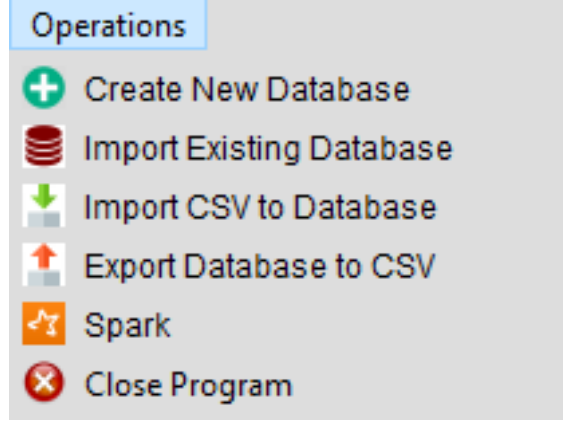
Average Accuracy Maximum Accuracy

AVG ACCURACY

Şekil B.1 Program açılış ekranı - 1

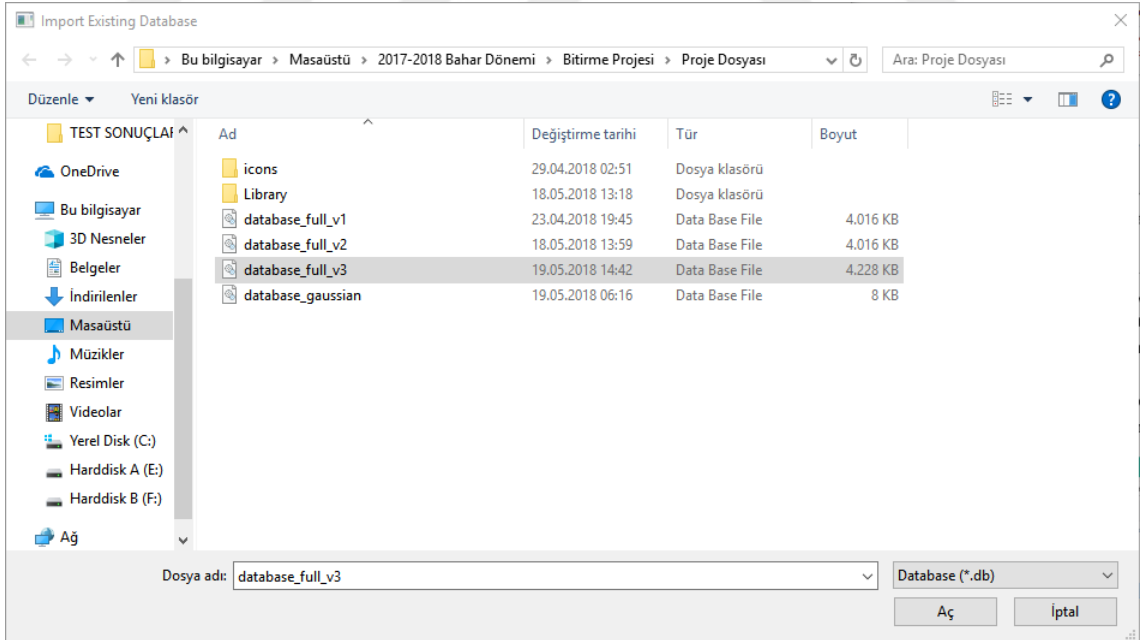


Şekil B.2 Program açılış ekranı - 2



Şekil B.3 Veri tabanı işlemleri

Veritabanı işlemleri menü işlemlerinde gerçekleştirilir. Bu bölüm programın ilk adımıdır. Hazır veritabanı yoksa, "Create New Database" menüsünden yeni bir veritabanı oluşturulur. Bir veritabanı zaten varsa, bu veritabanı "Import Existing Database" bölümünden seçilir ve işlemler bu veritabanı üzerinden gerçekleştirilir. Oluşturulan veritabanının CSV dosyası olarak dışa aktarmak için "Export Database to CSV" seçeneği seçilir ve CSV dosyasından veritabanına veri aktarmak için "Import CSV to Database" komutu kullanılır. Apache Spark ile özellik çıkarma işlemini gerçekleştirmek için "Spark" seçeneği seçilir. Programdan çıkmak için "Close Program" komutu seçilmelidir.



Şekil B.4 Var olan veri tabanının dahil edilmesi

Veritabanı oluşturulduktan veya içe aktarıldıktan sonra, müziklere ait akustik özelliklerin tablosu Şekil B.5 ile gösterildiği gibidir.

DATABASE FEATURE TABLE						
	Name	mZCR	sZCR	mCentroid	sCentroid	mBandwidth
1	C:/Users/Bilal/...	0.0830448206...	0.0276937836...	1784.1658495...	360.24167516...	2002.4490601...
2	C:/Users/Bilal/...	0.0560398087...	0.0380463071...	1530.1766787...	613.06612502...	2039.0365160...
3	C:/Users/Bilal/...	0.0762912074...	0.0317305700...	1552.8118647...	395.55991122...	1747.7023116...
4	C:/Users/Bilal/...	0.0333088626...	0.0205611037...	1070.1066149...	429.36690885...	1596.4128721...
5	C:/Users/Bilal/...	0.1014605387...	0.0442054827...	1835.0042655...	586.00336114...	1748.1721158...
6	C:/Users/Bilal/...	0.0940415337...	0.0789505897...	1831.9939396...	1015.1267751...	1729.6532869...
7	C:/Users/Bilal/...	0.0730522374...	0.0436940575...	1459.3664716...	661.70945138...	1389.0091311...
8	C:/Users/Bilal/...	0.0614422461...	0.0429978635...	1451.6670658...	670.49844329...	1577.2709414...

Şekil B.5 Özellikler tablosu

Sayısal işaret işleme kullanarak özellik çıkarma yöntemlerinden biri olan MFCC'nin parametre seçimi şu şekilde gerçekleştirilir. Burada "Window Type", işarete STFT gerçekleştirmek kullanılan pencere tipini belirtirken, "Window Size" bu pencerenin boyutunu belirtir. "Hop Length", pencereleme yapılırken alınan alınan arasındaki örtüşme değerini (overlap ratio) ifade eder. "Coefficient Number", MFCC işlemi sonrasında istenen MFCC katsayılarının sayısını belirtir. Kullanıcı "Apply" seçeneğiyle girilen bu değerleri varsayılan değerlerle değiştirebilir ve bir sonraki adımı geçebilir. "Change" düğmesine tıklandığında ise ayarlanan değerlerin tekrar değiştirilmesi mümkün olmaktadır.

CHOOSE MFCC PARAMETERS

Window Type

Hamming

Window Size

1024

Hop Length

256

Coefficient Number

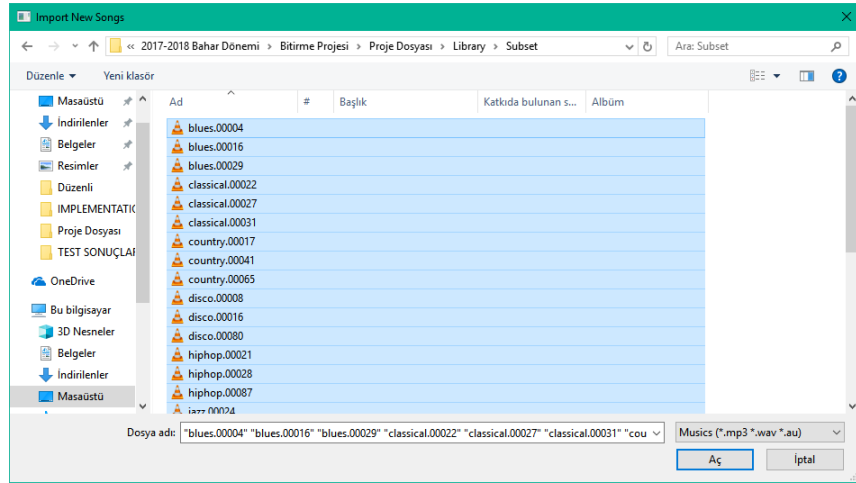
20

CHANGE

APPLY

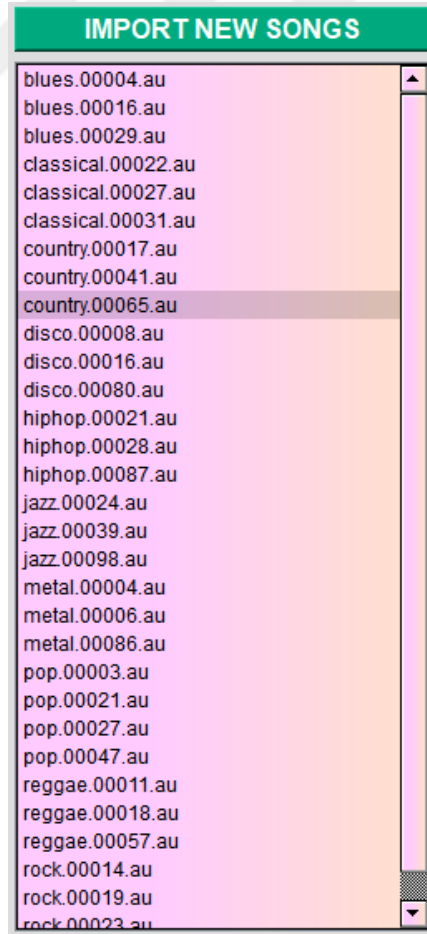
Şekil B.6 MFCC parametre seçimi

Parametre seçimini yaptıktan sonra, veritabanına şarkı eklemek için "İmport New Songs" düğmesi tıklanır.



Şekil B.7 Yeni şarkılar ekleme

Şarkılar eklendikten sonra, şarkı listesine eklenen şarkılar aşağıdaki gibi görüntülenebilir.



Şekil B.8 Eklenen şarkılar listesi

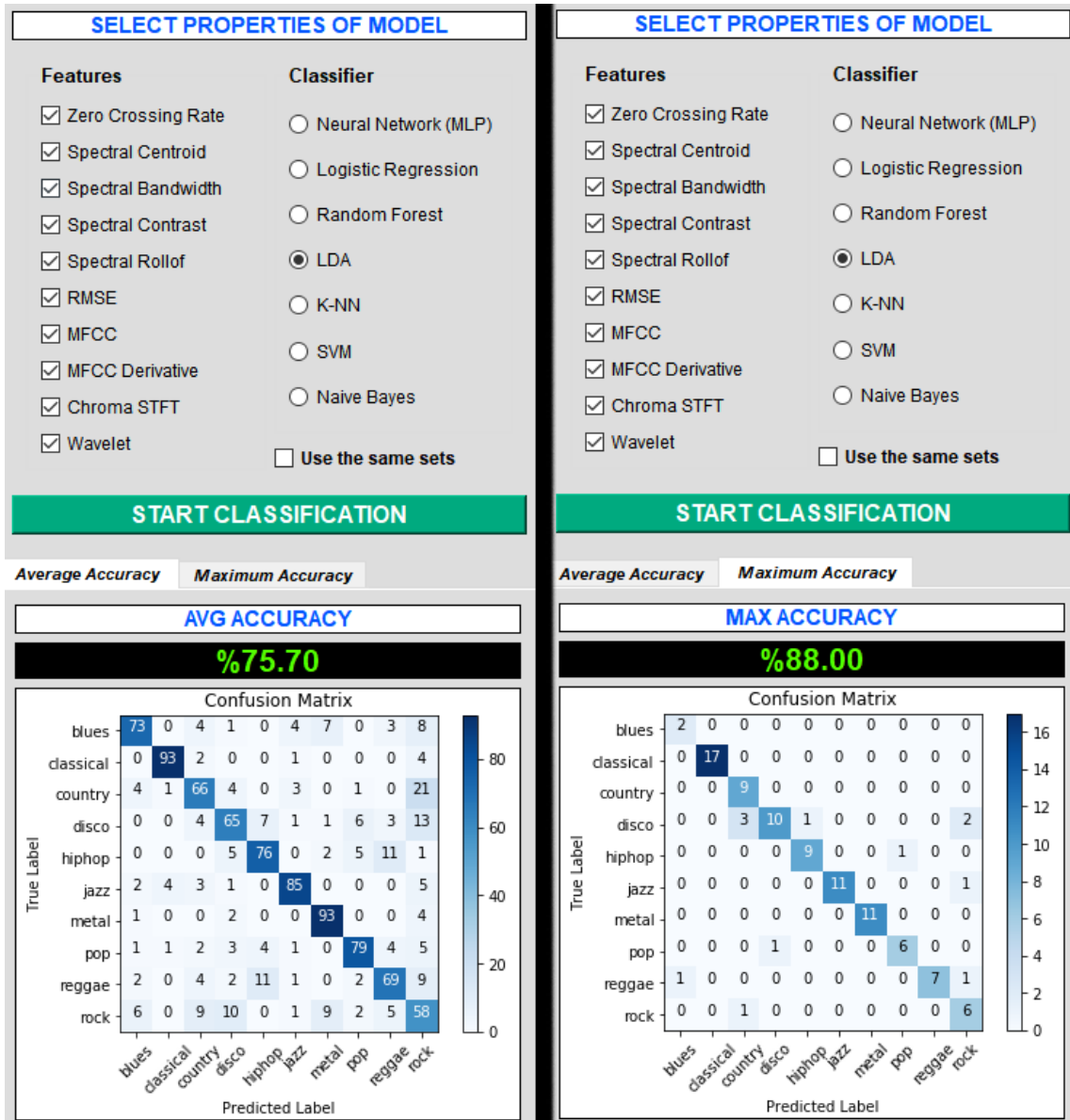
Şarkılar eklendikten sonra, özellik çıkarımı için gerçekleştirilen her bir aşama, ilerleme ekranında aşağıdaki gibi gösterilir.

```
blues.00094.au features are being extracted... Completed!  
blues.00095.au features are being extracted... Completed!  
blues.00096.au features are being extracted... Completed!  
classical.00009.au features are being extracted... Completed!  
classical.00010.au features are being extracted... Completed!  
classical.00011.au features are being extracted... Completed!  
country.00052.au features are being extracted... Completed!  
country.00053.au features are being extracted... Completed!  
country.00054.au features are being extracted... Completed!  
disco.00025.au features are being extracted... Completed!  
disco.00028.au features are being extracted... Completed!  
hiphop.00010.au features are being extracted... Completed!  
hiphop.00043.au features are being extracted... Completed!  
jazz.00002.au features are being extracted... Completed!  
jazz.00049.au features are being extracted... Completed!  
metal.00021.au features are being extracted... Completed!  
metal.00042.au features are being extracted... Completed!  
metal.00047.au features are being extracted... Completed!  
pop.00009.au features are being extracted... Completed!  
pop.00058.au features are being extracted... Completed!  
reggae.00010.au features are being extracted... Completed!  
reggae.00014.au features are being extracted... Completed!  
reggae.00017.au features are being extracted... Completed!  
reggae.00019.au features are being extracted... Completed!  
rock.00009.au features are being extracted... Completed!  
rock.00012.au features are being extracted... Completed!
```

Şekil B.9 Özellik çıkarımı ilerleme durum bilgisi

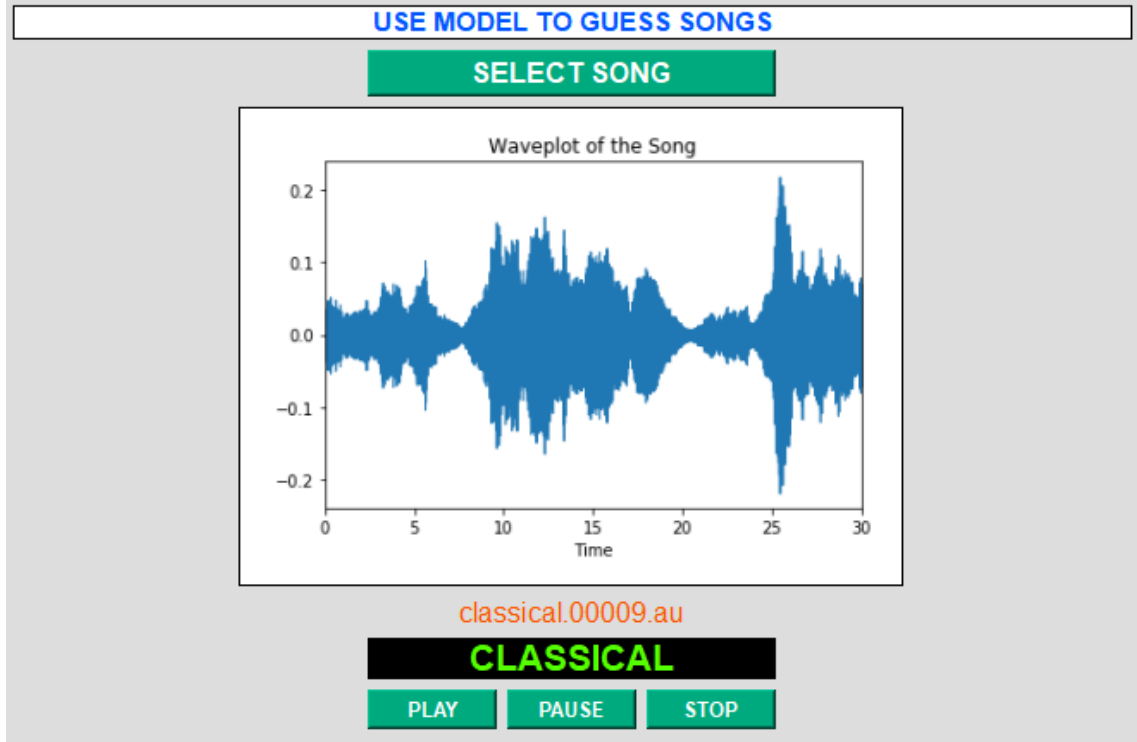
Veritabanı işlemleri, MFCC parametre işlemleri ve şarkı ekleme işlemleri tamamlandıktan sonra veriler programda işlenmeye hazır hale gelir. Bu adımlardan sonra, kullanıcının girdiği şarkılar artık eğitim ve test verisi olmak üzere ikiye ayrılır. Eğitim seti ile makine öğrenmesi yöntemleri kullanılarak bir model oluşturur ve test veri seti ile oluşturulan bu model test edilir. Model oluşturulmadan önce veritabanından çıkarılacak özellikler, onay kutuları ile seçilerek kaldırılabilir. Bu sayede kullanıcı, istenen özellikleri seçerek özelliklerin başarıma etkisini detaylı olarak gözlemleyebilir.

Özellikler seçildikten sonra, modeli oluşturacak sınıflandırma algoritması seçilir. Burada kullanıcı uygun bir sınıflandırma algoritmasını seçebilir ve "Start Classification" düğmesiyle sınıflandırma işlemini başlatabilir. Sonuç olarak, program test veri seti ile modelin doğruluğunu ölçer. Bu ölçüm, hesaplama işleminden sonra "Accuracy" bölümünde görülebilir. Aynı zamanda, modelin eğitim veri seti üzerinde hesaplanan karmaşıklık matrisi Şekil B.10 ile gösterildiği gibi görülebilir. Burada, satırların toplamı müzik türü sayısıdır ve sütunların toplamı tahmini müzik türü sayısıdır. Matristeki köşegen üzerindeki değerler ne kadar fazla olursa, modelin doğruluğu da o kadar doğru fazla olur. Çünkü bu rakamlar şarkının doğru tahmin edildiğini göstermektedir.



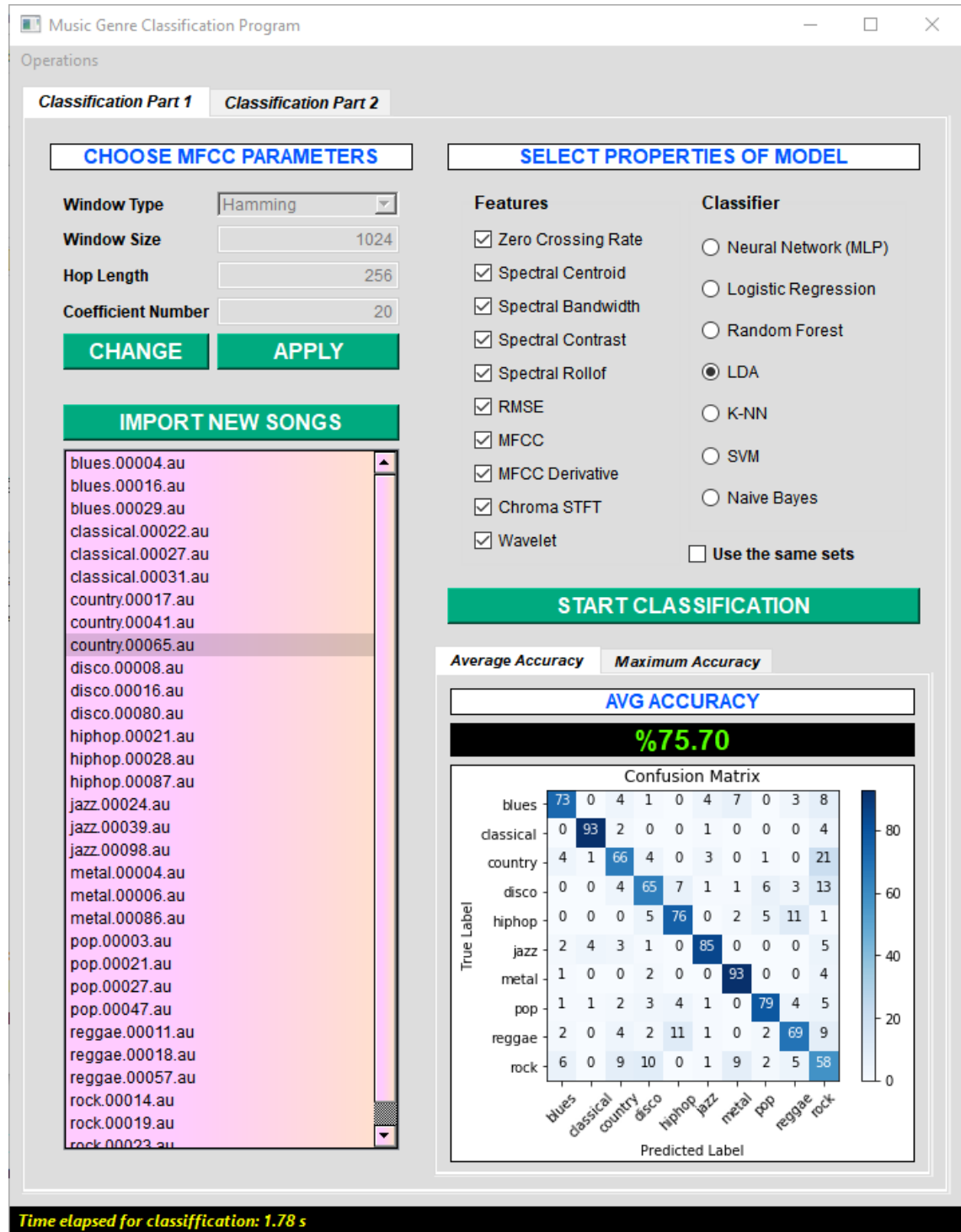
Şekil B.10 Sınıflandırma modeli özellikleri ve karmaşıklık matrisi

Model oluřturulduktan sonra kullanıcı, "Select Song" düğmesini tıklayarak istediğı bir řarkıyı seřebilir. Bu seřimden sonra, program bu řarkının ait olduğı müziğın türünü daha önce eğıtilen sınıflandırma modelini kullanarak tahmin eder ve grafik arayüzde tahmin sonucu görülebilir. Kullanıcı bu müziğı dinlemek istediğinde, "Play", "Pause" ve "Stop" düğmelerini tıklatarak dinlemeyi başlatabilir veya devam ettirebilir, durdurabilir ya da sonlandırabilir. Görselleřtirme ve fikir vermesi açısından, seřilen řarkıya ait işaretin zaman düzlemindeki grafiğı de verilir.

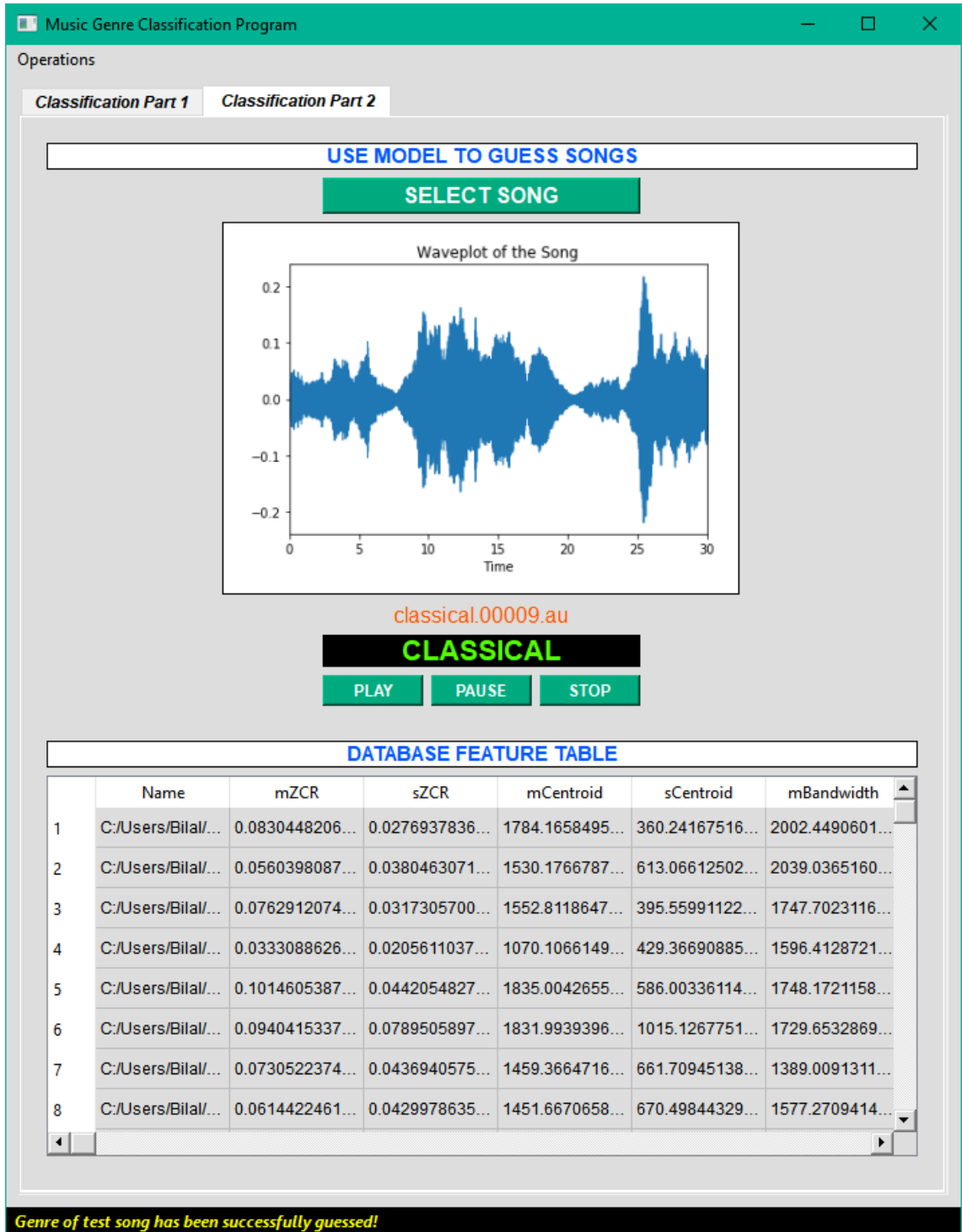


Şekil B.11 Müzik türü sınıflandırma

Program çalıştırıldıktan sonraki Şekil B.12 ve Şekil B.13 ile gösterildiği gibidir.



Şekil B.12 Program ana penceresi - 1



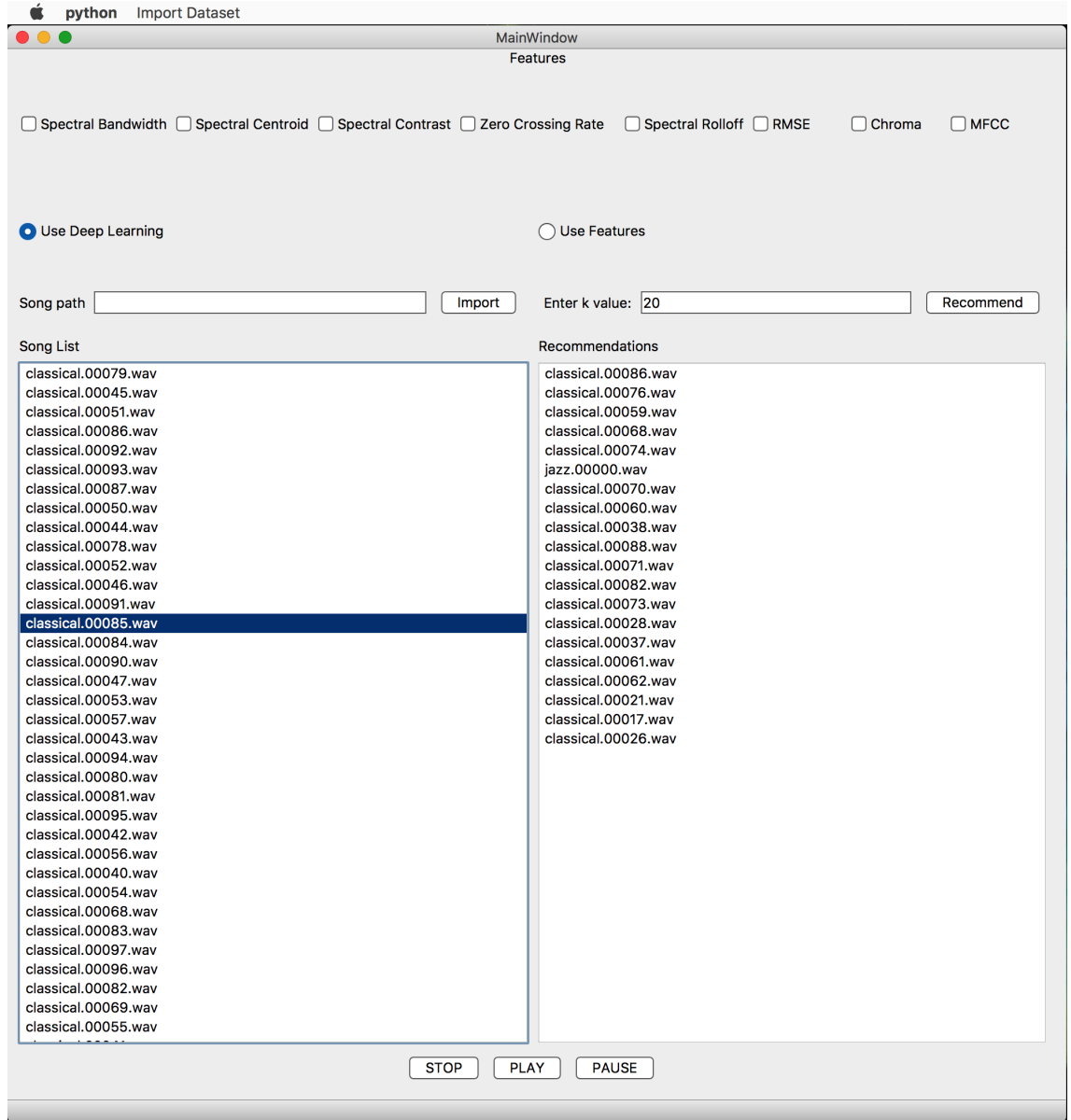
Şekil B.13 Program ana penceresi - 2

B.2 Müzik Öneri Sistemi Uygulaması

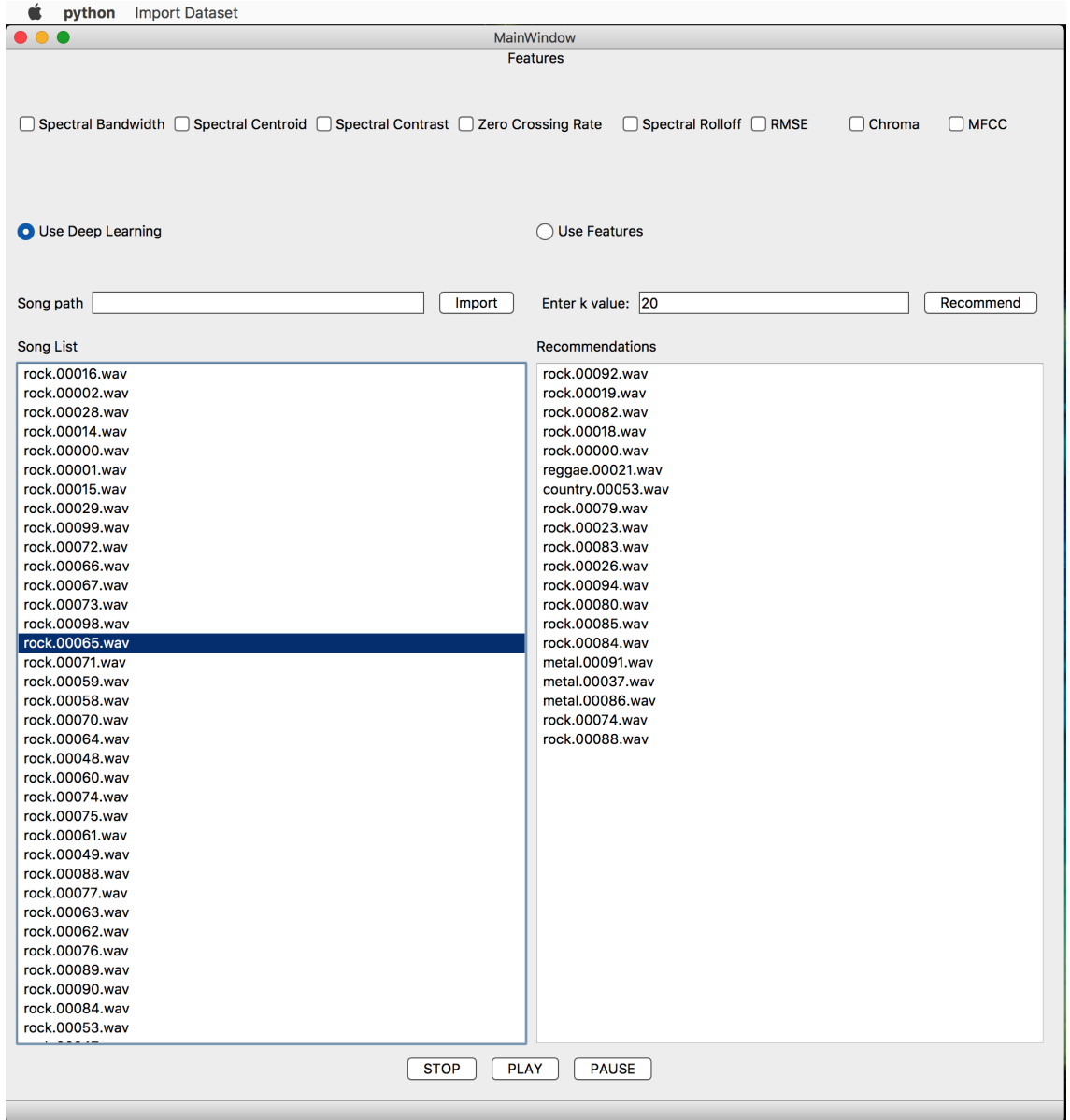
Bu bölümde akustik özellikler kullanarak müzik önerisi yapan uygulamaya ait ekran görüntüleri verilmiştir. Bu çalışma kapsamında sayısal işaret işleme ve derin öğrenme yöntemleri ile özellik çıkarımı gerçekleştirildiği için, müzik önerileri iki farklı şekilde yapılmaktadır. Uygulamaya işaret işleme ile elde edilen özellik veri tabanı ve derin öğrenme modeli dahil edildikten sonra aşağıdaki şekilde öneriler ekranda gösterilir.

Bu aşamada hem GTZAN veri seti hem de Turkcell tarafından sağlanan müzikler deneme amaçlı kullanılmıştır. Uygulamada "Song List" bölümünden istenilen bir şarkı seçildikten sonra, şarkı önerisi sayısı k belirlenir ve "Recommend" düğmesi kullanılarak seçilen bir şarkıya k sayıda şarkı önerisi yapılır. İki farklı modelle elde edilen sonuçlar devam eden bölümlerde gösterilmektedir.

B.2.1 Derin Öğrenme Özellikleri ile Elde Edilen Sonuçlar

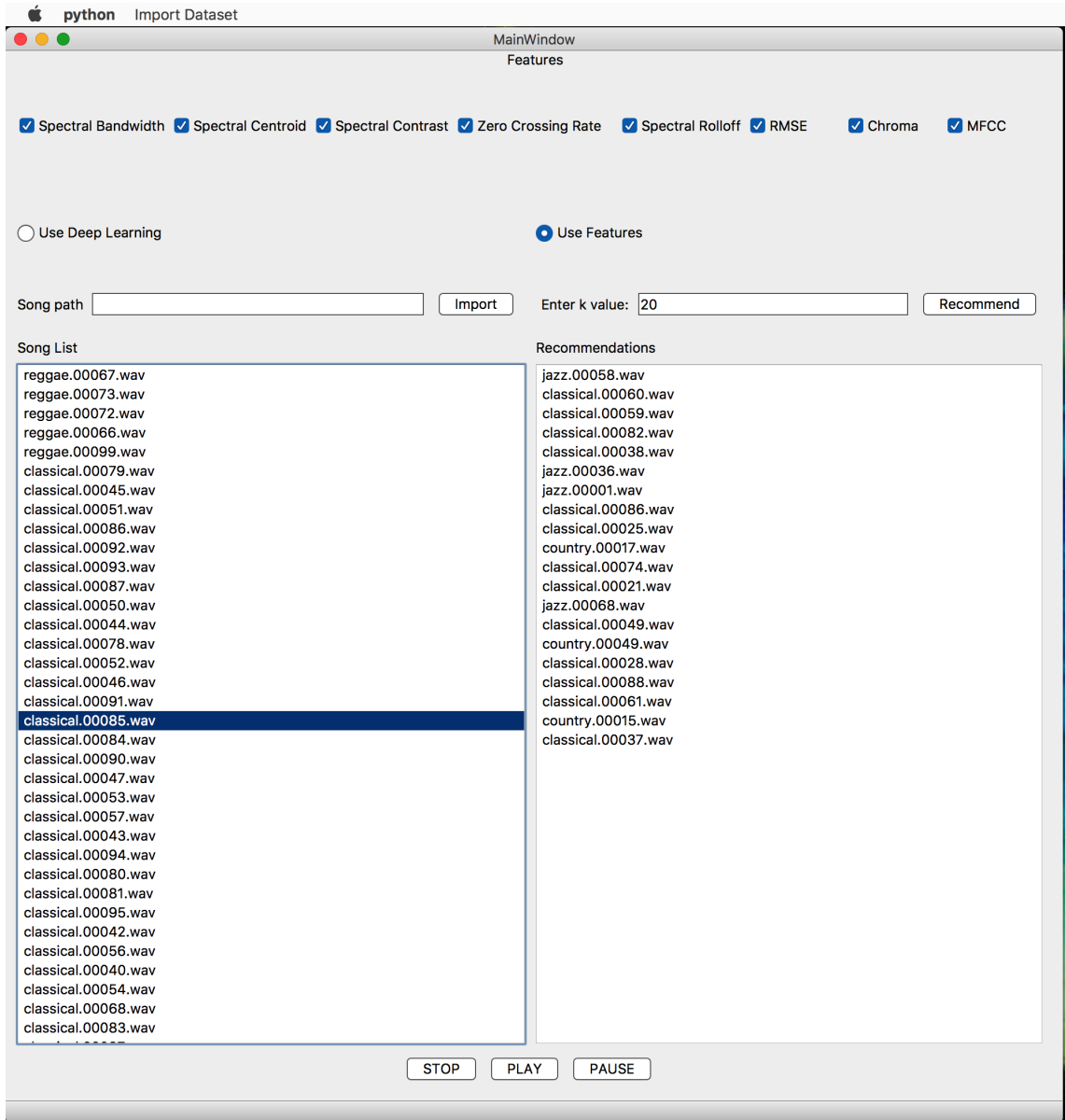


Şekil B.14 Müzik öneri sistemi örnek - 1 (Klasik Müzik Türü)

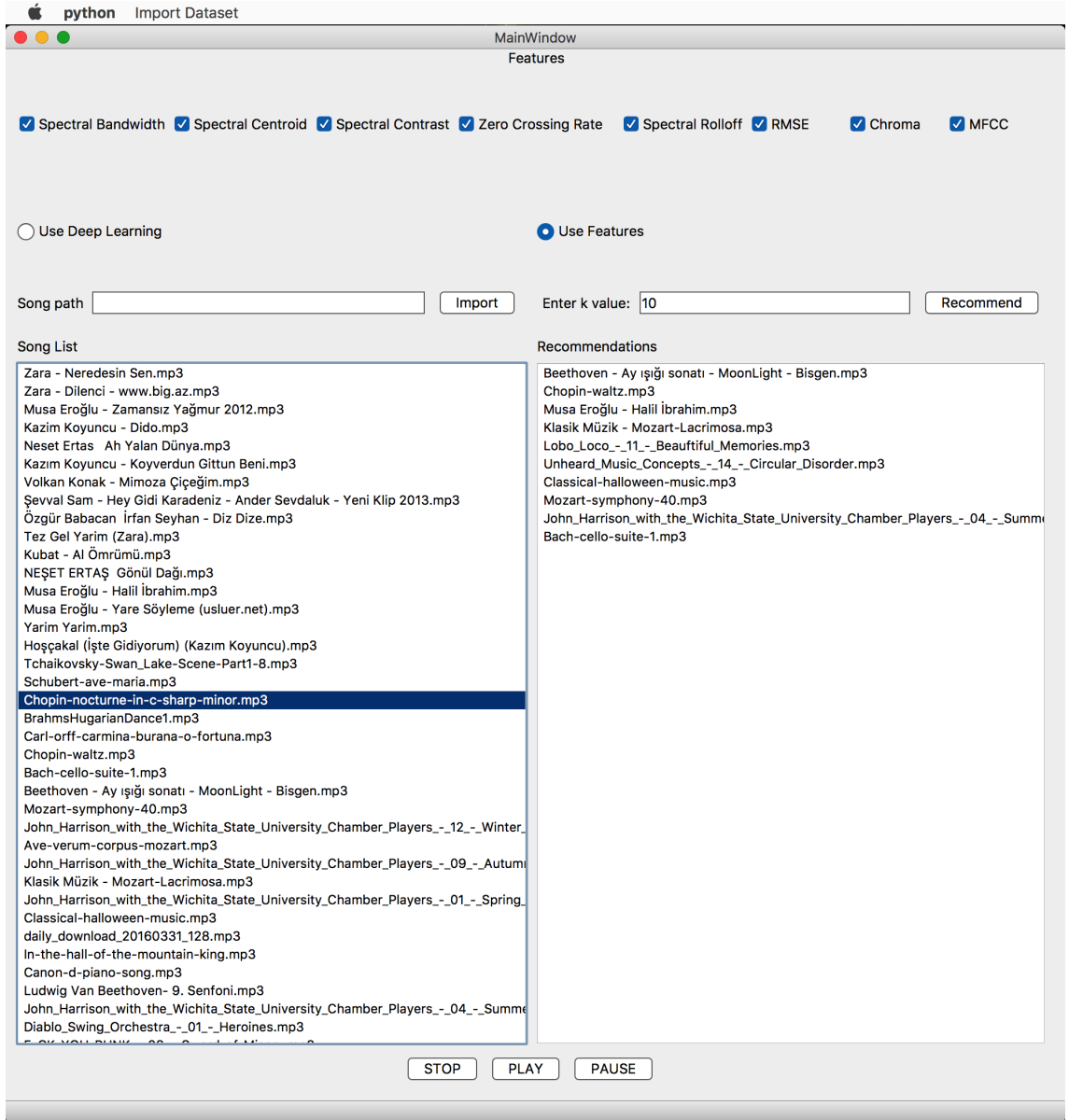


Şekil B.15 Müzik öneri sistemi örnek - 2 (Rock Müzik Türü)

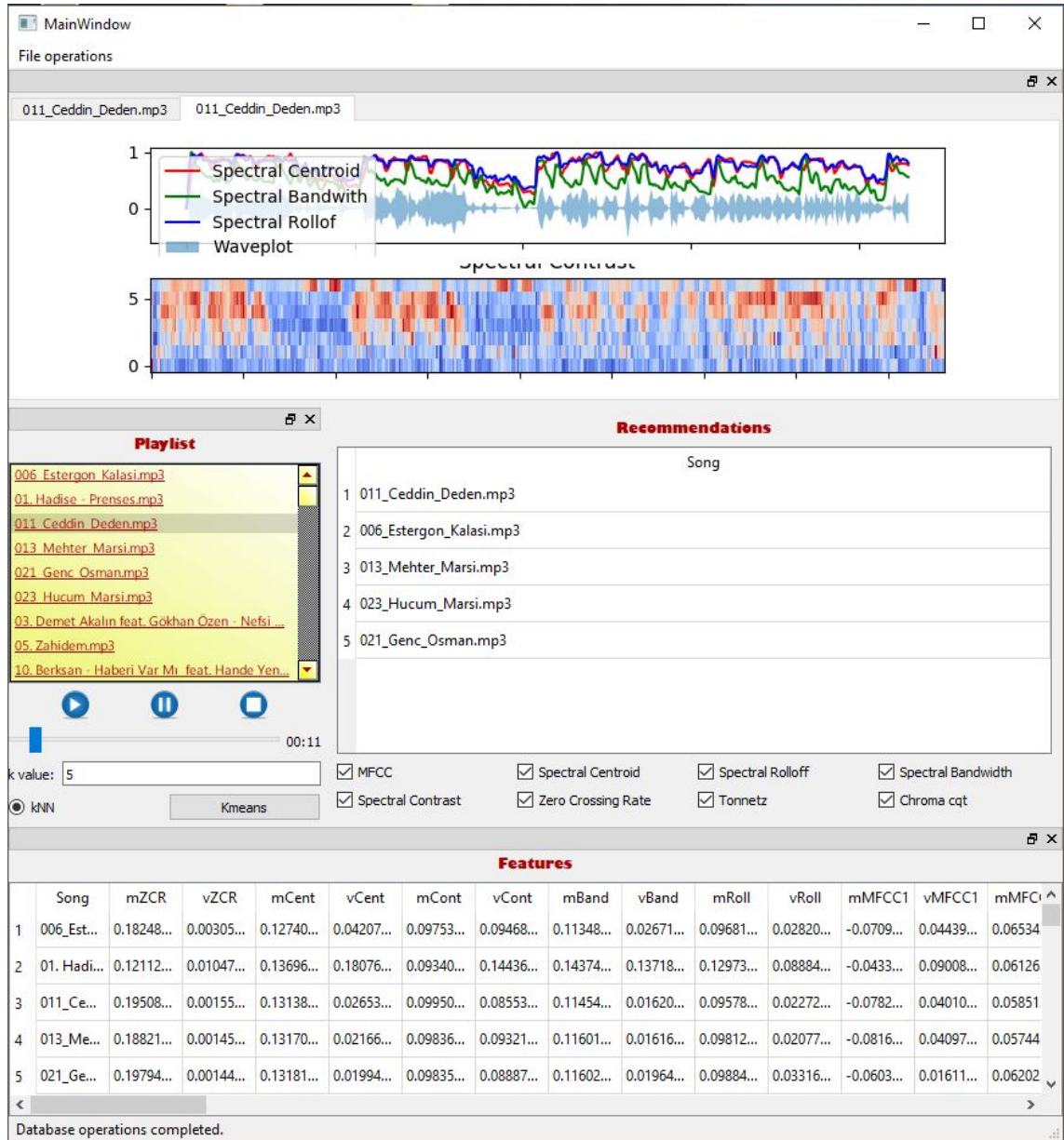
B.2.2 İşaret İşleme Yöntemleri ile Belirlenen Özellikler Kullanılarak Elde Edilen Sonuçlar



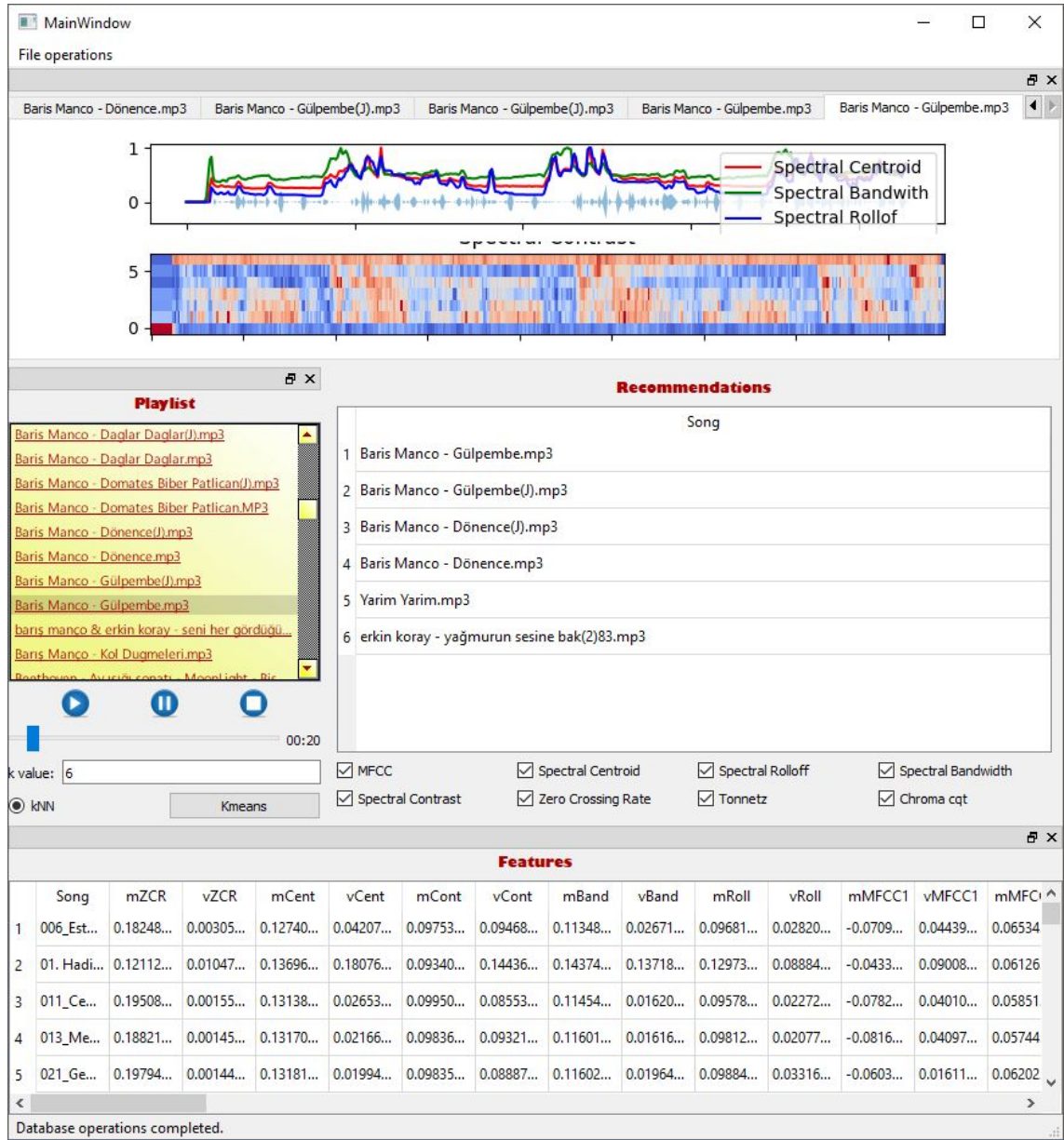
Şekil B.16 Müzik öneri sistemi örnek - 1 (Klasik Müzik Türü)



Şekil B.17 Müzik öneri sistemi örnek - 2 (Chopin-Nocturne C Sharp Minor)



Şekil B.18 Müzik öneri sistemi örnek - 3 (Mehter Marşları)



Şekil B.19 Müzik öneri sistemi örnek - 4 (Barış Manço Gülpembe)

TEZDEN ÜRETİLMİŞ YAYINLAR

İletişim Bilgileri: aelbir@gmail.com

Makale

1. A. Elbir and N. Aydın, "Music genre classification and music recommendation by using deep learning," in Electronics Letters, vol. 56, no. 12, pp. 627-629, 11 6 2020, doi: 10.1049/el.2019.4202.

Konferans Bildirisi

1. A. Elbir, H. Bilal Çam, M. Emre Iyican, B. Öztürk and N. Aydın, "Music Genre Classification and Recommendation by Using Machine Learning Techniques," 2018 Innovations in Intelligent Systems and Applications Conference (ASYU), Adana, 2018, pp. 1-5, doi: 10.1109/ASYU.2018.8554016.
2. A. Elbir, H. O. İlhan, G. Serbes and N. Aydın, "Short Time Fourier Transform based music genre classification," 2018 Electric Electronics, Computer Science, Biomedical Engineerings' Meeting (EBBT), Istanbul, 2018, pp. 1-4, doi: 10.1109/EBBT.2018.8391437.

Proje

1. 2019-2021, "Akustik Özellikler Kullanan Müzik Tavsiye Motoru", "5180069" nolu Tubitak TEYDEP 1505 projesi, Araştırmacı