Teknoloji Fakültesi

MARMARA ÜNİVERSİTESİ

TEKNOLOJİ FAKÜLTESİ

BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ

BİTİRME PROJESİ

Ses Verilerini Kullanarak Şarkı Türü Tahmini Yapılması

PROJE YAZARI

Umut Eren Toraman

170421037

DANIŞMAN

Dr. Öğr. Üyesi Gözde Karataş Baydoğmuş

İL, TEZ YILI

İstanbul, 2025

Teknoloji Fakültesi

MARMARA ÜNİVERSİTESİ

TEKNOLOJİ FAKÜLTESİ

BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ

BİTİRME PROJESİ

Projenin Adı

PROJE YAZARI

Yazar Adı Soyadı

Öğrenci numarası

DANIŞMAN

“Proje Danışmanı Unvanı Adı Soyadı

İL, TEZ YILI

MARMARA ÜNİVERSİTESİ

TEKNOLOJİ FAKÜLTESİ

BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ

Marmara Üniversitesi Teknoloji Fakültesi Bilgisayar Mühendisliği Öğrencisi Umut Eren Toraman nın "Ses Verilerini Kullanarak Şarkı Türü Tahmini Yapılması" başlıklı bitirme projesi çalışması, 19/06/2025 tarihinde sunulmuş ve jüri üyeleri tarafından başarılı bulunmuştur.

Jüri Üyeleri

Prof. Dr. Adı SOYADI (Danışman)

Marmara Üniversitesi ..... (İMZA) .....

Doç. Dr. Adı SOYADI (Üye)

Marmara Üniversitesi ..... (İMZA) .....

Dr. Öğr. Üyesi Adı SOYADI (Üye)

Marmara Üniversitesi ..... (İMZA) .....

İÇİNDEKİLER

Ses Verilerini Kullanarak Şarkı Türü Tahmini Yapılması

Umut Eren Toraman [1], and Gozde Karatas Baydogmus [1,*]

[1] Marmara Üniversitesi, Bilgisayar Mühendisliği, Türkiye

***** İletişim: gkaratas@marmara.edu.tr

Özet: Müziğin bireyler üzerindeki bilişsel, duygusal ve kültürel etkileri, hem akademik hem de teknolojik alanlarda önemli bir araştırma konusu haline gelmiştir. Bu bağlamda, müzik verilerinin analizi ve türlerine göre sınıflandırılması, içerik yönetimi, öneri sistemleri ve dijital medya uygulamalarına önemli katkılar sağlamaktadır. Özellikle, ses sinyallerinden anlamlı temsilcilerin çıkarılması, müziğin yapısal ve işitsel özelliklerinin bilgisayar sistemleri tarafından anlaşılabilir bir biçime dönüştürülmesi açısından kritik bir adımdır.

Bu çalışma, müzik türü sınıflandırmasında farklı ses özelliklerinin ve derin öğrenme mimarilerinin performans etkilerini incelemeyi amaçlamaktadır. Bu doğrultuda, Mel-Frekans Kepstrum Katsayıları (MFCC) ve Chroma özelliklerinin tekil ve birleşik kullanımları değerlendirilmiş; bu özellikler Konvolüsyonel Sinir Ağı (CNN) ve Uzun Kısa Süreli Bellek (LSTM) modelleriyle ayrı ayrı test edilmiştir. Deneysel sonuçlar GTZAN veri seti üzerinde gerçekleştirilmiş ve dört farklı model konfigürasyonu (MFCC-CNN, MFCC+Chroma-CNN, MFCC-LSTM, MFCC+Chroma-LSTM) karşılaştırılmıştır. Sonuçlar, MFCC ve Chroma özelliklerinin birleşik kullanımının LSTM mimarilerinde önemli performans artışları sağladığını gösterirken, CNN mimarisi MFCC özellikleriyle dahi yüksek başarı elde etmektedir. Elde edilen bulgular, özellik seçimi ile model mimarisi arasındaki ilişkinin ses tabanlı sınıflandırma problemlerinde belirleyici bir rol oynadığını ortaya koymaktadır.

Anahtar Kelimeler: Müzik Türü, MFCC, Chroma, GTZAN

---

1. Giriş

Son yıllarda, dijital müzik arşivlerinin hızlı bir şekilde büyümesi, bu tür verilerin otomatik sınıflandırılması ve analizini gerektiren uygulamalara olan ihtiyacı artırmıştır. Bu bağlamda, müzik bilgi erişimi (Music Information Retrieval - MIR) alanı önemli bir araştırma konusu haline gelmiştir. MIR kapsamında gerçekleştirilen temel çalışmalardan biri olan müzik türü sınıflandırması, bir ses kaydının içerdiği müzikal yapıya göre türünün belirlenmesini amaçlamaktadır [1, 2]. Bu süreç, içerik yönetimi, öneri sistemleri ve dijital müzik platformlarında kullanıcı deneyiminin kişiselleştirilmesi gibi alanlara önemli katkılar sağlamaktadır.

Müzik türü sınıflandırmasında başarının belirleyicilerinden biri, ses sinyalinden elde edilen özelliklerin kalitesidir. Mel-Frekans Kepstrum Katsayıları (MFCC), insan işitsel sistemini modelleyerek sesin zaman-frekans yapısını temsil eder ve bu alanda en yaygın

kullanılan özellik türlerinden biridir. Öte yandan, Chroma özellikleri, sesin ton yapısını yansıtarak özellikle armonik yapıların analizinde tamamlayıcı bilgi sağlar. Derin öğrenme tekniklerinin bu özelliklerle birleştirilmesi, sınıflandırma performansını önemli ölçüde artırabilir [1].

Bu çalışmada, MFCC ve Chroma özelliklerinin müzik türü sınıflandırması üzerindeki etkileri, farklı derin öğrenme mimarileri ile kullanılarak incelenmiştir. Sınıflandırma sürecinde Konvolüsyonel Sinir Ağı (CNN) ve Uzun Kısa Süreli Bellek (LSTM) mimarileri tercih edilmiş; özellik-mimari eşleştirmeleri dört farklı model ile yapılandırılmıştır: MFCC-CNN, MFCC+Chroma-CNN, MFCC-LSTM ve MFCC+Chroma-LSTM. Modellerin değerlendirilmesinde, bu alanda yaygın olarak kullanılan GTZAN Genre Collection veri seti tercih edilmiştir.

Bu çalışmanın literatüre katkıları aşağıdaki şekilde özetlenebilir:

- MFCC ve Chroma özelliklerinin tekil ve birleşik kullanımının sınıflandırma başarısı üzerindeki etkileri deneysel olarak karşılaştırılmıştır.
- CNN ve LSTM gibi farklı derin öğrenme mimarilerinin ses özellikleri ile etkileşimi analiz edilmiştir.
- Dört farklı konfigürasyonun doğruluk oranları GTZAN veri seti üzerinde değerlendirilmiştir.
- En yüksek başarıyı sağlayan model-özellik kombinasyonu belirlenmiş ve müzik türü sınıflandırması için öneriler sunulmuştur..

Bulgular, özellik seçimi ile model mimarisi arasındaki ilişkinin sınıflandırma performansında belirleyici bir rol oynadığını göstermektedir.

2. Literatür Taraması

Son yıllarda, ses/konuşma (dil ve konuşmacı) verilerinin işlenmesi ve sınıflandırılması alanında önemli çalışmalar yapılmış ve bu çalışmalar farklı uygulama alanlarında güçlü çözümler sunan birçok yaklaşımın geliştirilmesine öncülük etmiştir. Bu bölümde, literatürde yer alan çeşitli ses çalışmaları kullanılan teknikler, performans kriterleri ve uygulama alanları açısından gruplanarak tartışılmaktadır.

2.1. Müzik Türü Sınıflandırma Yaklaşımları

Çok modlu verilerin kullanımına bağlı olarak doğruluk artışı sağlayan müzik sınıflandırması alanındaki çalışmalardan biri, McKay ve Fujinaga (2008) tarafından jMIR sistemi ile gerçekleştirilen çalışmadır [1]. Bu sistem, ses, sembolik veriler (MIDI) ve kültürel kaynaklar (web tabanlı bilgiler) kullanılarak elde edilen özelliklerin birleştirilmesiyle %96,8'e kadar doğruluk elde etmiştir; ancak yüksek zaman karmaşıklığı ve düşük gürültü bağışıklığı gibi sınırlamalara da sahiptir. Bu yaklaşım, özellikle çok modlu analizlerin önemini vurgulamaktadır.

Mevcut yaklaşımlar arasında, Seo ve ark. (2023), MFCC, STFT ve Mel-spektrogram gibi farklı spektral özellikleri CNN tabanlı "late fusion" yöntemi ile birleştirerek GTZAN

veri setinde %96,8 doğruluk elde etmiştir [2]. Daha sonra, 2025 yılında yayımlanan başka bir çalışmada ise, MFCC, STFT ve DWT özelliklerinin paralel CNN ağlarıyla işlenmesi sonucunda yüksek başarı sağlanmıştır [3, 4]. Bu çalışmalar, spektral çeşitliliğin sınıflandırma performansına katkıda bulunduğunu göstermektedir.

2.2. Konuşmacı Tanıma Yaklaşımları

Konuşmacı tanıma sistemlerinde geleneksel MFCC yerine, daha biyolojik temelli yaklaşımlar önerilmiştir. Gammatone filtreleri ile geliştirilen GFCC tabanlı sistem, GMM modellemesiyle elde edilen MFCC'ye kıyasla daha iyi sonuçlar vermiştir [4]. Bu sistem özellikle gürültülü ortamlardaki kimlik doğrulama uygulamaları için uygundur.

Benzer şekilde, Mohammadi ve Mohammadi (2017) tarafından önerilen başka bir sistemde MFCC, IMFCC ve LPCC gibi özellikler birleştirilmiş ve bu birleşim, tek başına MFCC kullanımına kıyasla daha yüksek doğruluk sağlamıştır [6]. Özellik füzyon stratejisi, özellikle gürültüye dayanıklı güvenlik sistemlerinde etkili sonuçlar vermektedir.

2.3. Biyolojik ve Çevresel Seslerin Sınıflandırılması

Doğal ortamlardan toplanan seslerin sınıflandırılması, gürültü ve sınıf dengesizliği gibi zorlukları içermektedir. Bu bağlamda, Luque ve ark. (2018) kurbağa seslerinin sınıflandırılmasında kare puan serilerini de dikkate alan bir yöntem geliştirmiştir [7]. Karar Ağacı ile %90'ın üzerinde başarı elde edilmiş ve sistemin sınıf dengesizliğine karşı dayanıklı olduğu gösterilmiştir.

Ayrıca, çevresel ses sınıflandırmasında CNN tabanlı modellerin kullanımı artmıştır. 2024 yılında yayımlanan bir çalışmada, MFCC, tempogram ve mel-spektrogram özellikleri birlikte değerlendirilmiş; bu birleşimin özellikle ritim temelli sınıflarda güçlü sonuçlar verdiği gösterilmiştir [8]. Ayrıca, çevresel ses ve müzik aletleri sınıflandırmasına yönelik [9] çalışmada, MFCC özelliklerine veri artırma teknikleri entegre edilmiş ve doğrulukta %10'dan fazla artış sağlanmıştır. Bu yaklaşım, özellikle veri miktarının kısıtlı olduğu senaryolarda faydalı olmuştur.

2.4. Konuşma Tanıma Yaklaşımları

Konuşma tanıma yaklaşımları, ses teknolojilerinin en yaygın uygulama alanlarından biridir. Park ve ark. (2019) tarafından önerilen SpecAugment yöntemi, log-mel spektrogram üzerinde doğrudan zaman ve frekans maskeleri uygulayarak LAS modellerinde Kelime Hata Oranını (WER) %6,8'e düşürmüştür [10]. Bu yaklaşımın en büyük avantajı, ek veri gerektirmeden özellik seviyesinde veri artırımı sağlamasıdır.

2023 yılında yayımlanan başka bir derleme çalışması, MFCC, HMM ve RNN gibi klasik yöntemlerin karşılaştırmalı analizini sunmuş ve hangi özelliklerin hangi senaryolarda daha etkili olduğuna dair kapsamlı bir değerlendirme yapmıştır [11].

2.5. Derin Öğrenme ve Transformer Tabanlı İleri Düzey Modeller

Son yıllarda, ses sınıflandırmasında transformer mimarilerinin artışı dikkat çekmektedir. Causal Audio Transformer (CAT) modeli (2023), çok çözünürlüklü özellikleri zamansal olarak nedensel bir şekilde işleyerek ses sınıflandırmasında yüksek başarı elde etmiştir [12]. Benzer şekilde, Multiscale Audio Spectrogram Transformer (MAST) modeli, daha verimli hesaplama süresi ile ses olaylarının sınıflandırılmasında kullanılabilmektedir [13].

Transformer mimarisine bir alternatif olarak, nöromorfik bilgi işleme sistemleri (2025) enerji verimliliği açısından da öne çıkmaktadır [14]. Spiking Neural Networks kullanılarak geliştirilen bu sistem, mobil ve gömülü cihazlarda düşük güç tüketimi ile ses tanıma imkânı sağlamaktadır.

2.6. Özel Amaçlı Uygulamalar

Konuşan portre sentezi gibi özel uygulamalarda kullanılan modeller arasında, Whisper, Wav2Vec ve HuBERT gibi modern modelleri karşılaştıran 2025 tarihli bir analiz çalışması bulunmaktadır [15]. Bu çalışma, Whisper modelinin gerçek zamanlı senaryolarda daha başarılı olduğunu göstermiştir.

Benzer şekilde, derin öğrenme tabanlı sistemler bebek ağlamalarının paralinguistik analizi için kullanılmış ve MFCC ile perde (pitch) özellikleri kullanılarak sağlık izleme sistemlerinde önemli sonuçlar elde edilmiştir [16]. Son olarak, 2025 yılında AIP konferansında sunulan bir çalışmada, MFCC ve Mel-spektrogram kullanılarak deepfake ses içerikleri tespit edilmiştir [17]. CNN mimarisi ile geliştirilen model, gerçek ve sahte sesleri ayırt etmede yüksek doğruluk sunmaktadır.

Tablo 1'de, ilgili literatürdeki çalışmalar incelenmiş ve karşılaştırmalı olarak sunulmuştur.

Tablo 1. Literatür incelemesi ve makalelerin karşılaştırması

Makale	Yıl	Kullanılan Öznitelikler	Model Yönetimi	Uygulama Alanı	Yayın Türü	Gürültü Direnci	Zaman Karmaşıklığı
SpecAugment (Google) [10]	2019	Log Mel Spectrogram	Listen Attend and Spell (LAS)	Konuşma Tanıma	Arxiv	Yüksek	Düşük
Gammatone Tabanlı GFCC [5]	2012	GFCC (Gammatone + DCT)	GMM	Konuşmacı Tanıma	Konferans	Yüksek	Orta
Multi-Feature CNN Fusion [4]	2022	MFCC, Mel Spectrogram, ZCR	CNN, MobileNet, EfficientNet	Konuşma Tanıma	Arxiv	Orta	Yüksek
Late Fusion CNN - Müzik Türü [2]	2023	MFCC, STFT, Mel-Spectrogram	CNN (Late Fusion)	Müzik Türü Sınıflandırma	Dergi	Orta	Yüksek
jMIR Özellik Birleşimi [1]	2008	Audio + Symbolic + Web	ACE (Meta-learning)	Müzik Türü Sınıflandırma	Konferans	Düşük	Yüksek

Feature Fusion - Gürültülü Ortam [6]	2017	MFCC, IMFCC, LPCC	Özellik Birleşimi (Fusion)	Konuşmacı Tanıma	Konferans	Yüksek	Orta
Score Series & Frame-Classifiers [7]	2018	MFCC + Frame Scores	Sliding Window + DT	Kurbağa Sesi Tanıma	Dergi	Orta	Düşük
Review of Feature Extraction for Speech Recognition [11]	2023	MFCC, HMM, RNN	HMM, RNN	Konuşma Tanıma	Dergi	Orta	Orta
Real-Time Talking Portrait Audio Feature Analysis [GK9]	2025	Whisper, Wav2Vec, HuBERT	Comperative analysis	Konuşan Avatar	Dergi	Yüksek	Yüksek
Parallel CNN for Music Genre Classification [15]	2025	MFCC, STFT, DWT	Paralel CNN Ensemble	Müzik Türü Sınıflandırma	Dergi	Orta	Yüksek
MFCC Augmentation for Instrument Classification [9]	2024	MFCC, Data Augmentation	CNN, RNN	Ses Tanıma	Dergi	Yüksek	Orta
Infant Cry Paralinguistic Classification [16]	2025	MFCC, Pitch, Spectrogram	Deep Learning	Bebek Sesi Analizi	Dergi	Orta	Yüksek
Spectral & Rhythm Features with CNN [8]	2024	Mel Spectrogram, MFCC, Tempogram	CNN	Ses Tanıma	Arxiv	Orta	Yüksek
Causal Audio Transformer (CAT) [12]	2023	Multi-Resolution Features	Causal Transformer	Ses Tanıma	Arxiv	Orta	Yüksek
Multiscale Audio Spectrogram Transformer [13]	2023	Multiscale Spectrogram	Transformer	Ses Tanıma	Arxiv	Orta	Orta
Neuromorphic Audio Classification [14]	2025	Spiking Neural Networks	Neuromorphic Computing	Ses Sınıflandırma	Arxiv	Orta	Düşük
Deepfake Audio Detection with MFCC [17]	2025	MFCC, Mel-Spectrogram	CNN	Deepfake Ses Tanıma	Konferans	Yüksek	Orta

3. Metodoloji

Bu bölümde, çalışmada kullanılan metodolojik yaklaşım sunulmaktadır. Bu kapsamda; veri setinin detaylı açıklaması, ön işleme teknikleri, uygulanan algoritmalar ve değerlendirme metrikleri yer almaktadır. Amaç, verinin sistematik olarak hazırlanması,

anlamlı ses özelliklerinin çıkarılması, uygun derin öğrenme modellerinin uygulanması ve standart sınıflandırma metrikleri kullanılarak performanslarının değerlendirilmesi ile tekrarlanabilir bir müzik türü sınıflandırma süreci oluşturmaktır.

3.1. GTZAN Müzik Türü Veri Seti

George Tzanetakis ve Perry Cook tarafından 2002 yılında tanıtılan GTZAN Müzik Türü Veri Seti, Müzik Bilgi Erişimi (MIR) alanında en yaygın kullanılan kıyaslama veri setlerinden biridir. Otomatik müzik türü sınıflandırma sistemlerinin değerlendirilmesini kolaylaştırmak amacıyla geliştirilmiş ve alanında de facto standart haline gelmiştir. Dosyalar, farklı kayıt koşullarını temsil etmek amacıyla 2000-2001 yıllarında kişisel CD'ler, radyo, mikrofon kayıtları gibi çeşitli kaynaklardan toplanmıştır.

3.1.1. İçerik ve Yapı

Veri seti, her biri 30 saniye uzunluğunda olan 1.000 adet ses parçasından oluşmaktadır. Bu parçalar, 10 farklı müzik türüne eşit şekilde dağıtılmış olup, her tür için 100 örnek bulunmaktadır. Tüm kayıtlar monofonik olup, 22.050 Hz örnekleme frekansında alınmış ve 16 bitlik ses formatında kodlanmıştır; orijinal olarak WAV dosyaları halinde sunulmuştur. Örnek uzunluğu ve formatındaki bu tutarlılık, ön işleme sürecini kolaylaştırmakta ve deneyler arasında karşılaştırılabilirliği sağlamaktadır.

Veri setinde temsil edilen on müzik türü şunlardır: Blues, Klasik, Country, Disco, Hip-hop, Caz, Metal, Pop, Reggae ve Rock. Her tür alt kümesi 100 parçadan oluşmakta olup, dengeli bir sınıf dağılımı sağlayarak denetimli öğrenme görevleri için uygun bir ortam sunmaktadır.

3.1.2. Veri Seti Ön İşlemesi

Önerilen sistemin kritik bir bileşeni, sınıflandırma görevleri için uygun olan ilgili özelliklerin çıkarılması amacıyla ses verilerinin ön işlemden geçirilmesidir. Uygulanan ön işleme süreci, veri seti üzerinde yapılandırılmış bir tarama gerçekleştirir, ses sinyallerini segmentlere ayırır ve ses içeriğinin timbral dokusunu temsil etmede etkinliği geniş çapta kabul gören Mel-Frekans Kepstral Katsayıları (MFCC) çıkarımını yapar.

3.1.2.1. Veri Seti Organizasyonu ve Tarama

Ses veri seti, her alt dizinin farklı bir müzik türüne karşılık geldiği hiyerarşik bir klasör yapısında organize edilmiştir. save_mfcc() fonksiyonu, os.walk() yöntemi ile bu yapıyı sistematik olarak tarayarak dosyaların türlerine göre gruplanmasını ve etiketlenmesini sağlar. Her türe semantik bir etiket atanır ve sınıf karşılığının veri seti boyunca korunması amacıyla mapping listesine eklenir.

3.1.2.2. Ses Segmentasyonu

Eğitim örneklerinin sayısını artırmak ve her ses dosyası içinde yerel çeşitlilik sağlamak amacıyla sistem, sabit segmentasyon stratejisi uygular. Her ses parçası, n eşit segmente (num_segments = 10) bölünür. Her segmentteki örnek sayısı, parçadaki toplam örnek sayısının segment sayısına bölünmesiyle belirlenir. Bu segmentasyon yöntemi, eğitim

verisinin çeşitliliğini artırmakla kalmayıp, aynı zamanda özellik vektörlerinin boyutlarında tutarlılığı sağlayarak derin öğrenme modellerinin eğitimini kolaylaştırır.

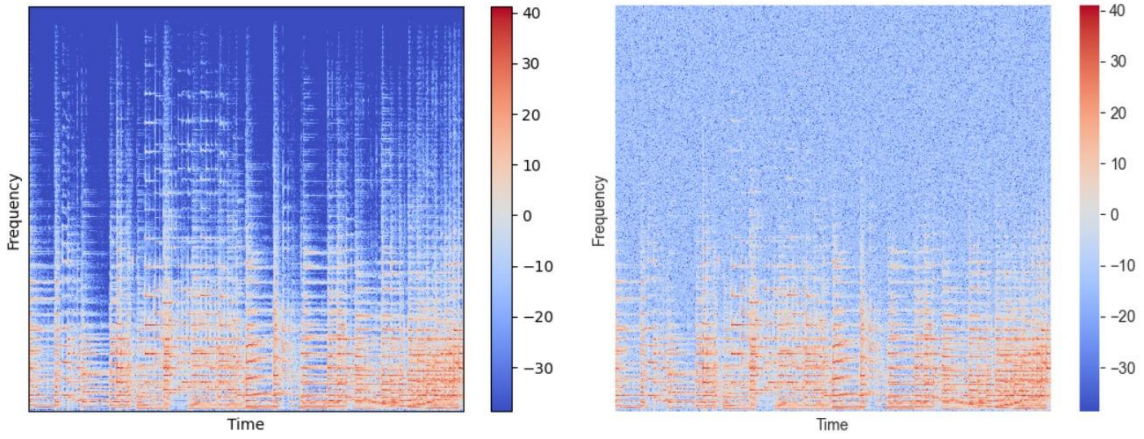
3.1.2.3. Veri Artırma

Orijinal veri seti kullanılarak yapılan ilk eğitim denemeleri, sınırlı veri çeşitliliği nedeniyle model performansının yetersiz olduğunu ve genelleme yeteneğinin düşük, aşırı öğrenme (overfitting) riskinin yüksek olduğunu göstermiştir. Buna karşılık, eğitim setinin boyutunu ve çeşitliliğini yapay olarak artırmak amacıyla veri artırma (data augmentation) teknikleri uygulanmıştır. Veri artırma, modeli daha geniş bir giriş varyasyonuna maruz bırakarak dayanıklılığını artırmakta ve böylece modelin görülmemiş verilere genelleme yapabilme yeteneğini geliştirmektedir.

Eğitim setindeki her bir ses örneğine aşağıdaki veri artırma yöntemleri uygulanmıştır:

- **Beyaz Gürültü Ekleme:** Gerçek dünya ses ortamlarını simüle etmek ve gürültüye karşı dayanıklılığı artırmak amacıyla orijinal sinyale rastgele Gaussian gürültüsü eklenir. Gürültünün yoğunluğu, belirli bir çarpan ile ölçeklendirilir.
- **Zaman Germe (Time Stretching):** Sesin perdesini değiştirmeden hızını değiştiren bu teknik, sinyalin zamansal dinamiklerini etkili bir şekilde değiştirir. Modelin tempo değişimlerine karşı değişmezlik öğrenmesini sağlamak için özellikle faydalıdır.
- **Perde Kaydırma (Pitch Shifting):** Sesin süresi değiştirilmeden belirli sayıda yarım ton kadar perde kaydırılır. Bu teknik, ton farklılıklarını çeşitlendirerek modelin kayıtlar arasındaki perde ile ilgili varyasyonlara karşı daha dayanıklı olmasını sağlar.

Bu dönüşümler, librosa ses işleme kütüphanesi kullanılarak uygulanmış ve artırılmış veriler, model eğitimi için Mel-Frekans Kepstral Katsayıları (MFCC) çıkarımı açısından orijinal verilerle aynı şekilde işlenmiştir. Bu artırılmış versiyonların eğitim veri setine dahil edilmesiyle, model her sınıfın daha zengin ve çeşitli bir temsilini görmüş ve bu durum sınıflandırma performansının artmasına yol açmıştır.



Şekil 1. WAV dosyasının gürültü eklenmeden ve eklendikten sonraki spektrogramları

Şekil 1, WAV dosyasının gürültü eklenmeden ve eklendikten sonraki spektrogramlarını göstermektedir [20].

3.1.2.4 MFCC Öznitelik Çıkarımı

Her segment için MFCC'ler, Librosa kütüphanesi kullanılarak hesaplanmaktadır. MFCC'ler, sesin algısal ve spektral özelliklerini yakalama yetenekleri nedeniyle sınıflandırmada kullanılan temel öznitelikler olarak görev yapmaktadır. Öznitelik çıkarım parametreleri şu şekilde yapılandırılmıştır: MFCC sayısı: n_mfcc = 13, FFT pencere boyutu: n_fft = 2048, Hop uzunluğu: hop_length = 512.

Sinyalin her segmenti, kısa zamanlı Fourier dönüşümüne (STFT) tabi tutulduktan sonra mel ölçekli filtreleme ve logaritmik sıkıştırma işlemlerinden geçer. Elde edilen sonuç, zaman içinde algısal frekans bantlarındaki enerji dağılımını temsil eden bir MFCC matrisidir. Bu matris, zaman adımlarının birinci eksen boyunca hizalanması için transpoze edilir; böylece, konvolüsyonel veya tekrarlayan sinir ağları gibi sıralı giriş modelleri ile uyumlu hale gelir.

Veri bütünlüğünü sağlamak amacıyla, yalnızca beklenen zaman çerçevesi sayısını (segment uzunluğu ve hop uzunluğu ile belirlenen) karşılayan MFCC segmentleri saklanmaktadır. Bu adım, farklı ses uzunlukları veya eksik segmentler nedeniyle ortaya çıkabilecek giriş boyutu tutarsızlıklarının önüne geçmektedir.

3.1.2.5 Etiketleme ve Serileştirme

Her geçerli MFCC segmenti, bir Python sözlüğündeki "mfcc" listesine eklenirken, ilgili sınıf indeksi (yani tür etiketi) "labels" listesinde saklanır. Tür isimleri ise, tahmin edilen etiketlerin yorumlanmasını kolaylaştırmak amacıyla "mapping" listesine kaydedilir.

Son olarak, MFCC özellikleri, etiketleri ve sınıf eşlemelerini içeren tüm veri seti serileştirilerek JSON formatında saklanır. Bu yapılandırılmış çıktı, sınıflandırma modelinin eğitim ve değerlendirme aşamalarında verilerin verimli bir şekilde yüklenip ayrıştırılmasını sağlar.

Oluşan JSON dosyası aşağıdaki yapıya uygun olarak düzenlenmiştir:

```
{
   "mapping": ["classical", "rock", "jazz", ...],
   "labels": [0, 1, 0, ...],
   "mfcc": [[[...], [...], ...], ...]
}
```

Bu yaklaşım, yüksek kaliteli özellik çıkarımı ve tutarlı veri temsilini garanti eden standartlaştırılmış ve ölçeklenebilir bir ön işleme hattı sunar; böylece müzik türü sınıflandırmasında denetimli öğrenme modelleri için güvenilir bir temel oluşturur.

3.2. Sınırlar ve Eleştiriler

Yaygın olarak kullanılmasına rağmen, GTZAN veri seti akademik literatürde çeşitli eleştirilere konu olmuştur:

> *Veri Tekrarı:* Bazı ses örnekleri aynı tür içinde veya farklı türler arasında tekrar etmekte olup, bu durum önyargı oluşturabilir ve performans ölçümlerini şişirebilir.
> *Etiket Hataları:* Veri setinin bir bölümünde yanlış etiketlenmiş ses parçaları

bulunmakta ve bu durum sınıflandırma sonuçlarının güvenilirliğini olumsuz etkileyebilir.
Müzik Dışı Artifaktlar: Bazı parçalar konuşma, gürültü veya sessizlik içermekte olup, müzikal içeriği temsil etmeyebilir.
Telif Hakkı Sorunları: Veri seti telif hakkıyla korunan materyaller içermekte ve bu durum, ticari bağlamda yeniden dağıtım ve kullanım açısından hukuki ve etik kaygılar doğurmaktadır.

Araştırmacıların, bu sınırlamaları ele almak için sağlam doğrulama protokolleri uygulamaları, alternatif veri setlerini değerlendirmeleri veya mevcutsa temizlenmiş ve açıklamalı versiyonları kullanmaları teşvik edilmektedir.

3.3. Algoritmalar

Bu çalışmada, çıkarılan ses özelliklerine dayalı müzik türü sınıflandırması yapmak için iki derin öğrenme mimarisi—Uzun Kısa Süreli Bellek (LSTM) ağları ve Konvolüsyonel Sinir Ağları (CNN)—kullanılmıştır.

Uzun Kısa Süreli Bellek (LSTM) ağları, ardışık verilerde uzun süreli zamansal bağımlılıkları yakalamak için tasarlanmış özel bir tekrarlayan sinir ağı (RNN) türüdür. Kapılı yapıları sayesinde, LSTM'ler, geleneksel RNN'lerde sıkça karşılaşılan kaybolan gradyan problemini azaltarak zaman içindeki örüntüleri etkin bir şekilde öğrenebilirler [18]. Ses sınıflandırma bağlamında, LSTM'ler, frekans ve tonal içeriğin zamanla evrimini temsil eden MFCC ve Chroma özellikleri gibi zaman serisi verilerindeki zamansal değişimleri modellemek için oldukça uygundur.

Convolutional Neural Networks (CNN'ler) öncelikle görüntü işleme alanındaki başarılarıyla tanınsa da, ses sınıflandırma görevlerinde de etkin bir şekilde kullanılmıştır. CNN katmanları, girişin mekansal boyutları (yükseklik ve genişlik) ile derinlik olmak üzere üç boyutta organize edilmiş nöronlardan oluşur. Buradaki derinlik, yapay sinir ağındaki toplam katman sayısını değil, bir aktivasyon hacminin üçüncü boyutunu ifade eder. Standart yapay sinir ağlarının aksine, herhangi bir katmandaki nöronlar yalnızca kendilerinden önceki katmanın küçük bir bölgesiyle bağlantılıdır [19]. Ses sinyallerinin iki boyutlu temsillerine (örneğin, MFCC spektrogramları) uygulandığında, CNN'ler yerel örüntüleri kullanarak özelliklerin mekansal hiyerarşisini çıkarabilir. Bu çeviriye duyarlı olmayan (translation-invariant) özellikleri yakalama yetenekleri, ritmik veya harmonik dokular gibi müzik sinyallerindeki karakteristik yapıları tanımada onları etkili kılar.

Her iki model de adil bir karşılaştırma sağlamak amacıyla aynı özellik setleri kullanılarak eğitildi ve sınıflandırma sürecinde MFCC ile Chroma özelliklerinin birleştirilmesinin etkisini belirlemek için performansları değerlendirildi.

3.4. Performans Değerlendirmesi

Bu çalışmada kullanılan sınıflandırma modellerinin etkinliğini değerlendirmek için birkaç standart değerlendirme metriği kullanılmıştır. Bunlar arasında Doğruluk (Accuracy), Duyarlılık (Recall) ve F1-Skoru, model performansına çeşitli sınıflandırma görevlerinde niceliksel içgörü sağlama yetenekleri nedeniyle yaygın olarak kabul görmüştür.

3.4.1. Doğruluk (Accuracy)

Doğruluk (Accuracy), değerlendirilen toplam örnek sayısı içindeki doğru sınıflandırılmış örneklerin oranını gösteren temel bir metriktir. Modelin genel tahmin yeteneğine dair bir ölçüm sağlar ancak sınıf dengesizliği durumunda yanıltıcı olabilir.

Matematiksel olarak, TP gerçek pozitif, TN gerçek negatif, FP yanlış pozitif ve FN yanlış negatif olmak üzere, Denklem (1) ile tanımlanır:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

3.4.2. Recall

Duyarlılık (Recall) ya da gerçek pozitif oranı olarak da bilinen bu metrik, modelin pozitif sınıfa ait tüm ilgili örnekleri doğru şekilde tanımlama yeteneğini ölçer. Pozitif vakaların tespit edilememesinin yüksek maliyete yol açtığı uygulamalarda özellikle önemlidir.

$$Recall = \frac{TP}{TP + FN} \tag{2}$$

Burada, Denklem (2)'de, gerçek pozitif modelin pozitif sınıfı doğru şekilde tahmin etmesi anlamına gelirken, yanlış negatif modelin pozitif sınıfı negatif olarak yanlış tahmin etmesi anlamına gelir.

3.4.3. Kesinlik (Precision)

Makine öğrenmesi ve istatistikte, kesinlik (precision), gerçek pozitif tahminlerin (gerçekte pozitif olanların) tüm pozitif tahminlere oranıdır. Yani, "Tahmin ettiğim pozitif sonuçların kaçı gerçekten pozitiftir?" sorusuna yanıt verir. Kesinlik formülü Denklem (3)'te verilmiştir.

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

3.4.4. F1-Skoru

F1-Skoru, Kesinlik (Precision) ve Duyarlılık'ın (Recall) harmonik ortalamasıdır. Hem yanlış pozitifleri hem de yanlış negatifleri tek bir metrikte birleştirerek dengeli bir değerlendirme sunar. F1-Skoru, sınıf dağılımının dengesiz olduğu ve Kesinlik ile Duyarlılık arasında bir denge gerektiği durumlarda özellikle faydalıdır. F1-Skoru formülü Denklem (4)'te verilmiştir.

$$F1\_score = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4}$$

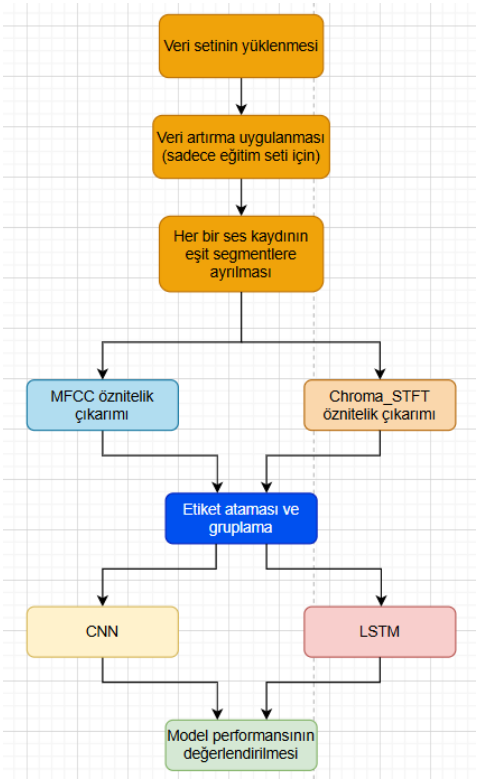
3.5. Önerilen Model

3.5.1. Genel Bakış ve Amaç

Bu çalışmanın temel amacı, farklı özellik çıkarma tekniklerinin ses türü sınıflandırmasının doğruluğu üzerindeki etkisini incelemektir. İki tür özellik kullanılmıştır: (i) Sesin tınısal özelliklerini temsil etmek için yaygın olarak kullanılan Mel-Frekans Kepstral Katsayıları (MFCC) ve (ii) armonik içeriği ekleyen MFCC ile kromanın (chroma short-time Fourier transform - chroma_stft) kombinasyonu. Bu özelliklerin sınıflandırma performansı, iki farklı derin öğrenme mimarisi kullanılarak araştırılmıştır: Bir Konvolüsyonel Sinir Ağı (CNN) ve Uzun Kısa Süreli Bellek (LSTM) ağı. En iyi performans gösteren model-özellik çiftini belirlemek için performans doğruluk (accuracy), duyarlılık (recall) ve F1-skoru temelinde değerlendirilmiştir.

3.5.2. Sistem Mimarisi

Tüm sistem akışı Şekil 2'de gösterilmiştir. Altı ana aşamadan oluşmaktadır: veri yükleme, ön işleme, özellik çıkarma, veri seti hazırlama, model eğitimi ve performans değerlendirme. Aşağıda, proje betiğinde uygulanan işlemleri yansıtarak her aşamanın detaylı açıklaması yer almaktadır.



Şekil 2. Önerilen algoritmanın akış şeması

3.5.3. Veri Yüklenmesi ve Öznitelik Çıkarımı

Ses özellikleri önceden çıkarılarak JSON dosyalarında saklanmıştır (eğitim/doğrulama için oye_augmented.json ve test için oye.json). load_data() fonksiyonu bu dosyaları ayrıştırarak MFCC veya MFCC+chroma_stft özelliklerini (X) ve bunlara karşılık gelen

etiketleri (y) yükler. Özellik matrisleri, modele doğrudan girdi sağlamak amacıyla NumPy dizileri olarak saklanmaktadır.

3.5.4. Veri Seti Hazırlığı ve Etiket Kodlama

"prepare_dataset()" fonksiyonu, artırılmış eğitim verisini train_test_split kullanarak eğitim ve doğrulama alt kümelerine böler. Ayrıca, string tabanlı tür etiketlerini model eğitimi için gereken tamsayı kodlu değerlere dönüştürmek amacıyla "LabelEncoder" uygular. Etiket kodlama, eğitim, doğrulama ve test veri setleri arasında tutarlılığı sağlar.

3.5.5. Modelin Derlenmesi ve Eğitilmesi

İki tür model eğitilmiştir: CNN ve LSTM. Her iki model de düşük öğrenme oranına (0.0001) sahip Adam optimizatörü ve tamsayı etiketli çok sınıflı problemler için uygun olan sparse_categorical_crossentropy kayıp fonksiyonu ile derlenmiştir. Eğitim, 30 epoch boyunca 32'lik batch boyutuyla gerçekleştirilmiş ve doğrulama seti eş zamanlı olarak izlenmiştir. Eğitim sonrası doğruluk ve kayıp eğilimlerinin değerlendirilmesi için eğitim geçmişi plot_history() fonksiyonu ile kaydedilmiştir.

3.5.6. Model Mimarileri

Özellik kombinasyonlarını test etmek için iki farklı derin öğrenme modeli kullanılmıştır. Aşağıda bu modellerin mimarilerinin detayları yer almaktadır:

3.5.6.1. LSTM Mimarisi

LSTM modeli, zaman dizisi olarak yeniden şekillendirilen MFCC veya birleşik özellikler kullanılarak sıralı modelleme için oluşturulmuştur. Mimari şu bileşenlerden oluşmaktadır:

- İki katmanlı yığılmış (stacked) LSTM katmanı:
    - İlk LSTM katmanı 128 birimden oluşur ve return_sequences=True parametresi ile tam diziyi bir sonraki katmana aktarır.
    - İkinci LSTM katmanı 64 birimden oluşur ve boyut indirgeme işlemi yapar.
- 64 nöronlu ve ReLU aktivasyon fonksiyonlu yoğun (dense) katman.
- Aşırı öğrenmeyi önlemek için %10 (0.1) oranında dropout katmanı.
- 10 birimli (müzik türü sınıf sayısı) softmax çıkış katmanı.

Bu model, ses dizilerindeki zamansal bağımlılıkları yakalamak için oldukça uygundur.

3.5.6.2. CNN Mimarisi

CNN modeli, 2B matrislerden (örneğin MFCC spektrogramları) mekansal özellik çıkarımı için tasarlanmıştır. Mimari şu bileşenlerden oluşmaktadır:

Üç adet konvolüsyon bloğu:

- Conv2D → MaxPooling2D → BatchNormalization
- Filtre sayıları: 32, 64 ve 128; Kernel boyutları: (3×3) veya (2×2)

Tam bağlantılı (fully connected) katmanlar:

- Flatten → Dense(128) → Dropout(0.3) → Dense(64) → Dropout(0.3)

Çıkış katmanı:

- Sınıflandırma için softmax aktivasyonlu Dense(10) katmanı

Bu mimari, frekans değişimleri ve zamana bağlı olmayan yerel desenleri öğrenmede üstün performans gösterir.

4. Deneysel Sonuçlar

Bu projenin gerçekleştirildiği geliştirme ortamı Tablo 2'de gösterilmiştir:

Tablo 2. Geliştirme Ortamı

Donanım	**Özellikler**
CPU	11th Gen Intel® Core(Tm) i7-11390H @3.40GHz
İşletim Sistemi	64 bit, Windows 11 Pro
Grafik Kartı	GTX 1650
L1/L2/L3 Önbellek	96 KB / 1.25 MB / 12 MB
RAM	16.00 GB
Python Version	3.10.11 64-bit

Bu bölümde, modellerin performansı doğruluk (accuracy), F1-skoru ve duyarlılık (recall) metrikleri kullanılarak değerlendirilmiştir. Çalışmanın temel amacı, MFCC ve Chroma özelliklerinin birleştirilmesinin müzik türü sınıflandırmasında performansı artırıp artırmadığını incelemektir. Bu amaçla dört model konfigürasyonu test edilmiştir: MFCC-CNN, MFCC+Chroma-CNN, MFCC-LSTM ve MFCC+Chroma-LSTM. Her model, on sınıflı bir veri seti üzerinde eğitilmiş ve genel performans yeteneğini değerlendirmek için test setinde performans metrikleri hesaplanmıştır.

Tablo 3. Test setindeki farklı model ve özellik kombinasyonlarının performans metrikleri

Model	**Doğruluk (%)**	**F1-Skoru (%, ort)**	**Recall (%, ort)**
MFCC-CNN	98.38	98.38	98.38
MFCC+Chroma-CNN	98.02	98.03	98.02
MFCC-LSTM	89.32	88.64	88.38
MFCC+Chroma-LSTM	95.94	95.93	95.94

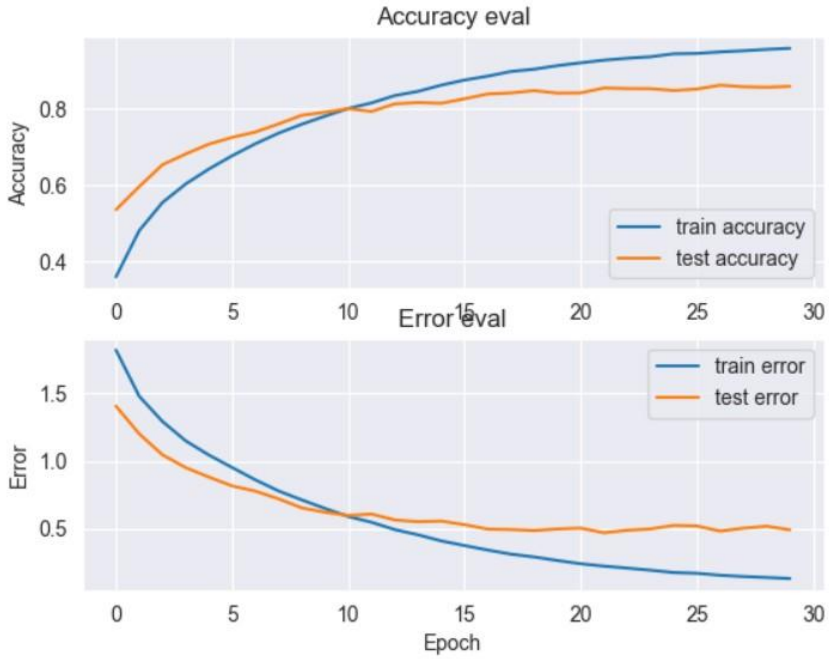
Tablo 3'te gösterilen sonuçlar, MFCC ve Chroma özelliklerinin birleşiminin LSTM tabanlı modellerde belirgin performans artışlarına yol açtığını göstermektedir. Özellikle, MFCC-LSTM modeli %89,32 doğruluk elde ederken, MFCC+Chroma-LSTM modeli

%95,94 doğrulukla onu önemli ölçüde geride bırakmıştır. Chroma özelliklerinin eklenmesi, neredeyse tüm sınıflarda F1-skorları ve duyarlılık (recall) değerlerinde iyileşme sağlamıştır; özellikle MFCC-only modelinin zorlandığı sınıflarda (örneğin, sınıf 9'da F1-skorda %75,84'ten %90,48'e, recall'da ise %69,20'den %84,10'a yükselme) bu artış dikkat çekicidir. Bu durum, Chroma'nın sağladığı armonik özelliklerin, MFCC'nin tınısal temsilini tamamlayıcı önemli bir rol oynadığını ve LSTM gibi sıralı modeller için kritik olan zamansal desenlerin yakalanmasında etkili olduğunu ortaya koymaktadır.

Ayrıca, CNN tabanlı modeller Chroma özellikleri olmadan da çok yüksek performans göstermiştir. MFCC-CNN modeli genel olarak en yüksek doğruluk olan %98,38'i kaydederken, MFCC+Chroma-CNN modeli biraz daha düşük olan %98,02'ye ulaşmıştır. Sınıf bazında F1-skorları ve recall değerleri her iki CNN modelinde de tutarlı şekilde yüksek olup, güçlü genelleme ve özellik çıkarma yeteneklerini göstermektedir. Ancak, Chroma özelliklerinin eklenmesi CNN performansında anlamlı bir artış sağlamamış; bu da konvolüsyonel mimarilerin yalnızca MFCC özelliklerinden yeterli ayırt edici bilgiyi çıkarabildiğini düşündürmektedir. Tüm skorlar Tablo 4'te gösterilmiştir.

Tablo 4. Tüm modeller için sınıf bazında F1-skoru ve Duyarlılık (Recall) değerleri

Sınıf	MFCC-CNN (%)		MFCC+Chroma-CNN (%)		MFCC-LSTM (%)		MFCC+Chroma-LSTM (%)	
	F1	Recall	F1	Recall	F1	Recall	F1	Recall
0	98.37	98.37	98.36	98.37	93.72	93.78	98.33	98.35
1	98.28	98.30	98.31	98.33	87.38	86.89	95.88	95.83
2	98.35	98.39	98.10	98.13	90.45	89.78	96.14	96.07
3	98.38	98.36	98.34	98.33	89.17	89.47	96.03	95.99
4	98.39	98.37	98.37	98.37	88.47	87.89	95.88	95.73
5	98.37	98.38	98.31	98.30	89.41	89.56	96.23	96.32
6	98.40	98.41	98.37	98.38	93.56	93.78	94.59	94.22
7	98.37	98.37	98.34	98.35	89.64	89.66	96.18	95.85
8	98.41	98.41	98.35	98.34	89.30	88.86	95.91	95.24
9	98.38	98.38	98.04	98.03	75.84	69.20	90.48	84.10



Şekil 3. En iyi model konfigürasyonunun performansı

Şekil 3, en iyi model konfigürasyonunun performansını göstermektedir; üst panelde eğitim ve test doğruluğu, alt panelde ise hata değerleri 30 epoch boyunca sunulmuştur. Hem eğitim hem de test doğruluğu artarken, hata oranları azalmaktadır. Test doğruluğu yaklaşık 0.82-0.84 aralığında (epoch 15-20 civarı) zirve yapıp ardından durağanlaşmakta, test hatası da benzer şekilde stabilize olmaktadır. Yaklaşık epoch 10-15 sonrasında eğitim ve test eğrileri arasındaki artan fark, modelin eğitim verisine aşırı uyum sağlamaya başladığını (overfitting) göstermektedir. Buna rağmen, bu konfigürasyon, belirgin overfitting etkileri ortaya çıkmadan önce en yüksek test seti performansını yakalaması nedeniyle en iyi model olarak seçilmiştir.

Özetle, deneysel sonuçlar MFCC ve Chroma özelliklerinin entegrasyonunun özellikle zamansal bağımlılıkları kullanan LSTM mimarileri için oldukça faydalı olduğunu doğrulamaktadır. CNN modelleri için ise yalnızca MFCC özellikleri yüksek etkinlik gösterirken, Chroma'nın sağladığı sınırlı performans artışı ek hesaplama maliyetini haklı çıkaramayabilir. Bu nedenle, MFCC ve Chroma özelliklerinin birleştirilmesinin etkinliği modele bağlıdır; zamansal modellerde önemli kazanımlar sağlarken, konvolüsyonel modellerde sınırlı iyileştirmeler sunar.

5. Tartışma

Deneysel bulgular, özellik seçimi ve model mimarisinin müzik türü sınıflandırmasındaki rolüne dair önemli bilgiler sunmaktadır. Çalışmanın temel amaçlarından biri, MFCC ve Chroma özelliklerinin birleştirilmesinin sınıflandırma performansını artırıp artırmadığını incelemekti. Sonuçlar, bu birleşimin özellikle ses

sinyallerindeki zamansal bağımlılıkları kullanan LSTM gibi sıralı modeller için oldukça faydalı olduğunu açıkça göstermektedir. Özellikle, MFCC+Chroma-LSTM modeli doğruluk, F1-skoru ve duyarlılık (recall) açısından MFCC-LSTM konfigürasyonunu önemli ölçüde geride bırakmış; bu durum, önceden düşük performans gösteren sınıflarda daha belirgindir. Bu da, perde ve armonik içeriği yakalayan Chroma özelliklerinin, MFCC'lerin temsil ettiği tını bilgisi ile tamamlayıcı bir rol oynayarak, zamansal modeller için daha zengin ve ayırt edici girdiler sağladığını ortaya koymaktadır.

Buna karşılık, CNN tabanlı modeller Chroma özellikleri eklendiğinde anlamlı bir iyileşme göstermemiştir. MFCC-CNN modeli zaten en yüksek genel performansı elde etmiş olup, Chroma özelliklerinin eklenmesi doğrulukta yalnızca küçük bir düşüşe yol açmıştır. Bunun sebebi, konvolüsyonel mimarilerin yalnızca MFCC'lerden yerel spektral desenleri çıkarmada son derece etkili olması ve ek armonik özelliklerin bu sürece kayda değer bir katkı sağlamamasıdır. Bazı durumlarda, artan giriş boyutu fazlalık veya gürültüye sebep olarak performansta küçük düşüşlere neden olabilir.

Bir diğer dikkat çekici gözlem ise farklı sınıflar arasındaki performans değişkenliğidir. CNN tabanlı modeller tüm türlerde tutarlı performans gösterirken, LSTM modeli özellikle sadece MFCC kullanıldığında sınıf 9'da belirgin sınıf bazlı zayıflıklar sergilemiştir. Chroma özelliklerinin eklenmesi bu zayıflıkları azaltmaya yardımcı olmuş ve bazı müzik türlerinin armonik temsilden diğerlerine göre daha fazla fayda sağlayabileceğini göstermiştir.

Genel olarak, bu bulgular, öznitelik kombinasyonlarının etkililiğinin büyük ölçüde model mimarisine bağlı olduğunu göstermektedir. MFCC'ler tek başına CNN tabanlı sınıflandırma için yeterli olabilirken, MFCC ve Chroma özniteliklerinin birleştirilmesi, zamansal bağlama dayalı modellerin performansını önemli ölçüde artırmaktadır. Bu durum, öznitelik çıkarım stratejilerinin seçilen model mimarisinin özellikleriyle uyumlu hale getirilmesinin önemini vurgulamaktadır.

6. Sonuç ve Gelecek Çalışmalar

Bu çalışmada, ses sinyallerinden elde edilen MFCC ve Chroma öznitelikleri, CNN ve LSTM mimarileri ile birlikte kullanılarak müzik türü sınıflandırma performansı değerlendirilmiştir. Yapılan deneyler sonucunda, LSTM tabanlı modellere Chroma özniteliğinin eklenmesiyle sınıflandırma performansında belirgin bir artış gözlemlenmiştir. Diğer yandan, CNN modelleri sadece MFCC özniteliği ile dahi yüksek doğruluk oranlarına ulaşırken, Chroma katkısı sınırlı kalmıştır. Bu sonuçlar, özniteliklerin modele özgü etkileşimlerinin performans açısından kritik olduğunu göstermektedir.

Gelecekteki çalışmalar kapsamında, GTZAN veri setinin bilinen sınırlılıkları (etiket hataları, tekrarlar vb.) nedeniyle deneylerin daha dengeli ve temiz veri setleriyle tekrarlanması, müzik türü sınıflandırmasında Transformer tabanlı mimarilerin potansiyelinin araştırılması, gerçek zamanlı ve kaynak kısıtlı ortamlarda çalışabilecek hafif model yapılarının geliştirilmesi ve MFCC ile Chroma'ya ek olarak tempo, ritim ve spektral

kontrast gibi ek akustik özelliklerin entegre edilerek daha zengin bir öznitelik seti oluşturulması hedeflenmektedir.

Bu bağlamda çalışma, hem metodolojik yaklaşımıyla hem de deneysel bulgularıyla müzik bilgi erişimi alanındaki ileri düzey çalışmalara katkıda bulunmayı amaçlamaktadır.

Referanslar

[1] McKay, C., & Fujinaga, I. (2008, September). Combining Features Extracted from Audio, Symbolic and Cultural Sources. In ISMIR (pp. 597-602).

[2] Seo, W., Cho, S. H., Teisseyre, P., & Lee, J. (2023). A short survey and comparison of cnn-based music genre classification using multiple spectral features. IEEE Access, 12, 245-257.

[3] Zhang, Y., & Li, T. (2025). Music genre classification with parallel convolutional neural networks and capuchin search algorithm. Scientific Reports, 15(1), 9580.

[4] Turab, M., Kumar, T., Bendechache, M., & Saber, T. (2022). Investigating multi-feature selection and ensembling for audio classification. arXiv preprint arXiv:2206.07511.

[5] Xu, H., Lin, L., Sun, X., & Jin, H. (2012, May). A new algorithm for auditory feature extraction. In 2012 international conference on communication systems and network technologies (pp. 229-232). IEEE.

[6] S. Al-Kaltakchi, M. T., Woo, W. L., Dlay, S., & Chambers, J. A. (2017). Evaluation of a speaker identification system with and without fusion using three databases in the presence of noise and handset effects. EURASIP Journal on Advances in Signal Processing, 2017, 1-17.

[7] Luque, A., Romero-Lemos, J., Carrasco, A., & Barbancho, J. (2018). Improving classification algorithms by considering score series in wireless acoustic sensor networks. Sensors, 18(8), 2465.

[8] Wolf-Monheim, F. (2024). Spectral and Rhythm Features for Audio Classification with Deep Convolutional Neural Networks. arXiv preprint arXiv:2410.06927.

[9] Rezaul, K. M., Jewel, M., Islam, M. S., Siddiquee, K. N. E. A., Barua, N., Rahman, M. A., ... & Asha, U. F. T. (2024). Enhancing Audio Classification Through MFCC Feature Extraction and Data Augmentation with CNN and RNN Models. International Journal of Advanced Computer Science and Applications, 15(7), 37-53.

[10] Park, D. S., Chan, W., Zhang, Y., Chiu, C. C., Zoph, B., Cubuk, E. D., & Le, Q. V. (2019). Specaugment: A simple data augmentation method for automatic speech recognition. arXiv preprint arXiv:1904.08779.

[11] Yadav, S., Kumar, A., Yaduvanshi, A., & Meena, P. (2023). A Review of Feature Extraction and Classification Techniques in Speech Recognition. SN Computer Science, 4(6), 777.

[12] Liu, X., Lu, H., Yuan, J., & Li, X. (2023, June). Cat: Causal audio transformer for audio classification. In ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 1-5). IEEE.

[13] Zhu, W., & Omar, M. (2023, June). Multiscale audio spectrogram transformer for efficient audio classification. In ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 1-5). IEEE.

[14] Basu, A., Chaudhari, P., & Di Caterina, G. (2025). Fundamental Survey on Neuromorphic Based Audio Classification. arXiv preprint arXiv:2502.15056.

[15] Salehi, P., Sheshkal, S. A., Thambawita, V., Gautam, S., Sabet, S. S., Johansen, D., ... & Halvorsen, P. (2025). Comparative analysis of audio feature extraction for real-time talking portrait synthesis. Big Data and Cognitive Computing, 9(3), 59.

[16] Owino, G., & Shibwabo, B. (2025). Advances in Infant Cry Paralinguistic Classification—Methods, Implementation, and Applications: Systematic Review. JMIR Rehabilitation and Assistive Technologies, 12(1), e69457.

[17] Jbara, W. A., & Soud, J. H. (2025, March). Deepfake audio detection via MFCC features and mel-spectrogram using deep learning. In AIP Conference Proceedings (Vol. 3264, No. 1, p. 030027). AIP Publishing LLC.

[18] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural computation, 9(8), 1735-1780.

[19] O'shea, K., & Nash, R. (2015). An introduction to convolutional neural networks. arXiv preprint arXiv:1511.08458.

[20] Schlüter, J., & Grill, T. (2015, October). Exploring data augmentation for improved singing voice detection with neural networks. In ISMIR (pp. 121-126).