



**T.C.
YALOVA ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ**

**BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI
BİLGİSAYAR MÜHENDİSLİĞİ BİLİM DALI**

TÜRK MÜZİK TÜRLERİNİN EVRİŞİMLİ SİNİR AĞLARI İLE SINIFLANDIRMASI

YÜKSEK LİSANS TEZİ

**SHAHAD BASSAM HAZIM
218102014**

DANIŞMAN: Dr. Öğr. Üyesi Murat OKKALIOĞLU

**YALOVA
TEMMUZ 2024**



T.C.
YALOVA ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI
BİLGİSAYAR MÜHENDİSLİĞİ BİLİM DALI

TÜRK MÜZİK TÜRLERİNİN EVRİŞİMLİ SINIR AĞLARI İLE SINIFLANDIRMASI

YÜKSEK LİSANS TEZİ

SHAHAD BASSAM HAZIM
218102014

DANIŞMAN: Dr. Öğr. Üyesi Murat OKKALIOĞLU

YALOVA
TEMMUZ 2024

ETİK BEYAN

Yalova Üniversitesi Lisansüstü Eğitim Enstitüsü Tez/Proje Yazım Kuralları'na uygun olarak hazırladığım “Türk Müzik Türlerinin Evrişimli Sinir Ağları İle Sınıflandırması” başlıklı bu tez çalışmasında; tez içinde sunduğum verileri, bilgileri ve dokümanları akademik ve etik kurallar çerçevesinde elde ettiğimi, tüm bilgi, belge, değerlendirme ve sonuçları bilimsel etik ve ahlak kurallarına uygun olarak sunduğumu, tez çalışmasında yararlandığım eserlerin tümüne uygun atıfta bulunarak kaynak gösterdiğimi, kullanılan verilerde herhangi bir değişiklik yapmadığımı, bu tezde sunduğum çalışmanın özgün olduğunu bildirir, aksinin tespiti halinde doğabilecek her türlü hukuki sorumluluğu kabul ettiğimi taahhüt ve beyan ederim.

Shahad Bassam Hazım

İmza



ÖNSÖZ

Yüksek lisans eğitimim boyunca benden yardımlarını esirgemeyen, akademik anlamda bana her zaman destek olan ve doğru yolu gösteren sayın danışman hocam Dr. Öğr. Üyesi Murat OKKALIOĞLU'na teşekkür ederim. Tez jürimde görev yapan Prof. Dr. Murat GÖK ve Doç. Dr. Gürcan YAVUZ'a da katkıları için teşekkür ederim.

Her zaman yanımda oldukları ve benden desteklerini hiçbir zaman esirgemedikleri için kocam ve aileme teşekkür ederim.

Temmuz -2024

Shahad Bassam Hazım





İÇİNDEKİLER

ETİK BEYAN	i
İÇİNDEKİLER.....	v
SİMGELER ve KISALTMALAR LİSTESİ	ix
TABLolar LİSTESİ	xi
ŞEKİLLER LİSTESİ.....	xiii
ÖZET	xiv
ABSTRACT	xvi
1. GİRİŞ	1
1.1. Problem Tanımı	2
1.2. Tezin Katkısı.....	3
1.3. Tez Yapısı.....	3
2. LİTERATÜR TARAMASI	5
2.1. Mevcut Veri Setleri	5
2.2. Mevcut Çalışmalar.....	6
3. ÖN BİLGİLER.....	11
3.1. Ses için Sinyal İşleme.....	11
3.1.1. Fourier dönüşümü	12
3.1.2. Kısa süreli Fourier dönüşümü (Short-Time Fourier Transform, STFT).	12
3.1.3. Mel-spektrogram.....	15
3.2. Derin Öğrenme	17
3.2.1. Evrişimli sinir ağları (CNN)	17
4. YÖNTEM.....	25
4.1. Veri Seti.....	25
4.2. Yöntem	26
4.2.1. Mel-spektrogram tespiti.....	27
4.2.2. CNN modeli.....	29
4.2.3. Yazılım araçları.....	30
4.3. Karşılaştırılan Yöntem.....	32
4.3.1. Akustik özellikler.....	32
4.3.2. Kullanılan makine öğrenimi algoritmaları.....	35
5. SONUÇLAR	37
5.1. Değerlendirme Metrikleri.....	37
5.1.1. Karmaşıklık matrisi (Confusion matrix).....	37
5.1.2. Duyarlılık (Recall)	38



5.1.3. Kesinlik (Precision)	39
5.1.4. F1-skoru (F1-score)	39
5.1.5. Doğruluk (Accuracy)	39
5.1.6. Kayıp (Loss)	40
5.1.7. Makro ortalama (Macro-averaging).....	40
5.1.8. Ağırlıklı ortalama (Weighted average)	41
5.2. Önerilen Yöntemin Sonuçları	41
5.3. Önerilen Yaklaşımın Sonuçlarının Diğer Yöntemle Karşılaştırılması	44
6. SONUÇ VE GELECEK ÇALIŞMALAR.....	49
6.1. Sonuç	49
6.2. Gelecekteki Çalışmalar	50
KAYNAKLAR.....	51
ÖZGEÇMİŞ	57



SİMGELER ve KISALTMALAR LİSTESİ

SİMGELER

p	: Bu parametre normun türünü tanımlar
ω	: Frekansı temsil eder
σ	: Sigma
∞	: Sonsuzluk simgesi
e^x	: Üstel Fonksiyon (Exponential Function)
t	: Zamanı temsil eder

KISALTMALAR

BMD	: Brezilya Müzik Veri Seti
CNN	: Evrişimli Sinir Ağı (Convolutional Neural Network)
CONV	: Evrişimsel Katman (Convolutional Layer)
DFT	: Ayrık Fourier Dönüşümü (Discrete Fourier Transform)
FC	: Tam Bağlantılı Katman (Fully Connected Layer)
FFT	: Hızlı Fourier Dönüşümü (Fast Fourier Transform)
FMA	: Free Music Archive
FN	: Yanlış Negatif (False Negative)
FP	: Yanlış Pozitif (False Positive)
FT	: Fourier dönüşümünün (Fourier Transform)
kHz	: kilohertz
K-NN	: K-en Yakın Komşu (K-Nearest Neighbor)
LR	: Lojistik Regresyon (Logistic Regression)
LSTM	: Uzun-Kısa Vadeli Bellek (Long-Short Term Memory)
MBE	: Müzik Bilgi Erişimi (Music Information Retrieval)
MFCC	: Mel Frekans Cepstral Katsayıları (Mel Frequency Cepstral Coefficients)
MIR	: Müzik Bilgisi Erişiminde (Music Information Retrieval)
MSD	: Million Song Dataset
PPV	: Pozitif Tahmin Değeri (Positive Predicted Value)
ReLU	: Düzeltilmiş Lineer Birim (Rectified Linear Unit)
RMSE	: Ortalama Kare Enerjinin Kökü (Root Mean Square Energy)
RNN	: Tekrarlayan Sinir Ağları (Recurrent Neural Networks)
Sklearn	: Scikit-learn
Sr	: Örnekleme Hızı (Sample Rate)
STFT	: Kısa Süreli Fourier Dönüşümü (Short-Time Fourier Transform)
SVM	: Destek Vektör Makineleri (Support Vector Machines)
SymbTr	: Türk Makam Müziği Sembolik Temsili Veritabanı
tanH	: Hiperbolik tanjant aktivasyon fonksiyonu (Hyperbolic tangent)
TMG	: Türk Müzik Türü (Turkish Music Genre)
TN	: Doğru Negatif (True Negative)
TP	: Doğru Pozitif (True Positive)
ZCR	: Sıfır Geçiş Oranı (Zero_crossing_rate)

TABLolar LİSTESİ

Tablo 4.1. Türk müzik türü	25
Tablo 4.2 Türk müziği türü veri seti	26
Tablo 5.1. F1-skoru Özbacı'nın [10] çalışmasındaki sonucun önerilen yaklaşımla karşılaştırılması	46
Tablo 5.2. Kesinlik(Precision) Özbacı'nın [10] çalışmasındaki sonucun önerilen yaklaşımla karşılaştırılması	46
Tablo 5.3. Duyarlılık (Recall) Özbacı'nın [10] çalışmasındaki sonucun önerilen yaklaşımla karşılaştırılması	47





ŞEKİLLER LİSTESİ

Şekil 3.1 STFT'nin ses sinyalindeki etkisi.....	14
Şekil 3.2. Mel-spektrogramına dönüşüm	16
Şekil 3.3. Evrişimli sinir ağlar mimarisi	19
Şekil 3.4. Havuzlama katmanı (Pooling layer)	20
Şekil 3.5. ReLU aktivasyon fonksiyonu.....	22
Şekil 3.6. Sigmoid fonksiyonu	23
Şekil 3.7. tanH aktivasyon fonksiyonu.....	24
Şekil 4.1 Türk müziği türlerinin sınıflandırılması için önerilen yöntem.....	27
Şekil 4.2 Türk müziğinin Mel-spektrogramı.....	28
Şekil 5.1. Karmaşıklık matrisi (Confusion matrix)	38
Şekil 5.2. Eğitim ve doğrulama için doğruluk fonksiyonu	42
Şekil 5.3. Eğitim ve doğrulama için kayıp fonksiyonu	43
Şekil 5.4. Karmaşıklık matrisi.....	43
Şekil 5.5 Makine öğrenimi algoritmaları için karmaşıklık matrisi(SVM, KNN, Lojistik Regresyon).....	45

TÜRK MÜZİK TÜRLERİNİN EVRİŞİMLİ SİNİR AĞLARI İLE SINIFLANDIRMASI

ÖZET

Dijital müzik veri tabanlarının düzenlenmesinde sıklıkla kullanılan müzik türü sınıflandırması, dijital müzik kayıtlarının hızla artmasıyla birlikte müzik bilgi erişimi alanında zorlu bir görev haline gelmiştir. Tür sınıflandırması, şarkı önerileri geliştirme, benzer şarkıları bulma ve belirli bir müzikten hoşlanabilecek kitleleri keşfetme gibi bazı güncel ilgi çekici zorlukları açıklamakta faydalı olabilir. Bu nedenle, otomatik müzik türü sınıflandırması, müzik analizi, erişim ve öneri sistemleri sunan uygulamalar için önemli bir rol oynamaktadır. Bu süreçte kullanıcıya yardımcı olabilir veya tamamen onun yerini alabilir. Ayrıca, otomatik müzik türü sınıflandırması, müzik sinyallerinin içerik tabanlı analizi için özellikler geliştirme ve değerlendirme konusunda bir temel oluşturur. Bu alandaki önceki araştırmalar ağırlıklı olarak Batı müziğindeki türlerin sınıflandırmasına odaklanmıştır. Ancak, çevrimiçi Türk müziği içeriğini sınıflandırma henüz tam olarak tanımlanmamıştır ve bu durum, Türk müziği bağlamında ses türlerinin otomatik sınıflandırılması için bir engel teşkil etmektedir. Türk müziği türü sınıflandırmasına ilişkin daha önceki çalışmalar, çoğunlukla Klasik tekniklere dayanmaktadır. Bu klasik teknikler, iki temel adıma dayanmaktadır: müzik türünün akustik özelliklerinin çıkarılması ve sınıflandırma için makine öğrenimi algoritmalarının kullanılması. Ancak bu yaklaşım, Türk müziği türü sınıflandırmasının performansını sınırlamaktadır. Bu nedenle, bu çalışmada, Mel-spektrogram analizi ile elde edilen görüntüleri kullanarak evrişimli sinir ağları ile otomatik Türk müziği türü sınıflandırması üzerinde durulmuştur. İlk olarak, Arabesk, Pop, Rap, Rock, Türk Klasik Müziği ve Türk Dini Müziği gibi Türk müziğinin en iyi bilinen altı kategorisini içeren bir veri kümesi oluşturuldu. Ardından, bu çalışma, evrişimli sinir ağları ile derin öğrenme tekniğini önermekte ve şarkıları uygun müzik türlerine sınıflandırmak için Mel-Spektrogram'dan üretilen spektrogram görüntülerini CNN modeline girdi olarak kullanmaktadır. Önerilen yöntemin eğitim için yaklaşık %99,62 ve test için %99,80 gibi yüksek bir doğruluk oranına sahip olması, bu yöntemin başarısını kanıtlamaktadır.

Anahtar Kelimeler: Türk Müziği, Tür Sınıflandırması, Evrişimsel Sinir Ağları, Mel-Spektrogram.



TURKISH MUSIC GENRES CLASSIFICATION USING CONVOLUTIONAL NEURAL NETWORK

ABSTRACT

Music genre classification, frequently used to organize digital music databases, has become a challenging task in the field of music information retrieval due to the exponential growth of digital music recordings. Genre classification can be useful in addressing certain current interesting challenges, such as developing song recommendations, finding similar songs, and discovering audiences who might enjoy that specific music. Therefore, automatic music genre classification plays an important role for applications that offer music analysis, retrieval, and recommendation systems. It can assist or even replace human users in this process. Furthermore, automatic musical genre classification serves as a framework for developing and assessing features for any type of content-based analysis of musical signals. Previous research in this field has primarily focused on exploring the classification of genres in Western music. However, categorizing online Turkish music content remains inadequately defined, posing a challenge for the automatic classification of audio genres in the Turkish music context. The earlier work on Turkish music genre classification relies mostly on classical techniques for genre classification. These classical techniques depend on two basic steps: the first is extracting the acoustic features of music genre and the second is using machine learning algorithms for classification, which limits the performance of Turkish music genre classification. Therefore, in this research, automatic Turkish music genre classification with convolutional neural networks by using images obtained by the Mel-spectrogram analysis is studied. The initial step involves the creation of a dataset comprising of the most six well-known categories of Turkish music, which include: Arabesque, Pop, Rap, Rock, Turkish Classical Music, and Turkish Religious Music. Then, this research proposes a deep learning technique with convolutional neural networks and employs spectrogram images produced from Mel-Spectrogram as the input into a CNN to classify the songs into the appropriate musical genres. The proposed method has an accuracy level of about 99.62% for training and 99.80% for testing, which confirms the success of this method.

Keywords: Turkish Music, Genre Classification, Convolutional Neural Networks, Mel-Spectrogram.



1. GİRİŞ

Müzik, dünya genelinde milyarlarca insanın hayatında önemli bir yer kaplamaktadır. Müzik eserleri ve performansları en karmaşık ve zor kültürel eserlerimiz arasında yer almakta olup, müziğin duygusal gücü bizi beklenmedik ve derin şekillerde etkileyebilir. Müzik, ritim, ses seviyesi ve perde gibi organize bir sesin ortak ön plana çıkaran, geri plana iten veya ihmal eden bir tarz olarak tanımlanabilir. İnsanlar, kültürel veya kişisel sebeplerden ötürü çeşitli müzik türlerine ilgi ve merak duymaktadır.

Teknolojik ve sosyal gelişmelerin etkisiyle 20. yüzyılın başında yeni müzik türleri ortaya çıkmıştır. Caz, blues, rock, pop, funk ve hip-hop gibi yeni türlere ayrılan müzik, müzikal sınıflandırmaları daha da karmaşık hale getirdi. Müzik formatları ve müzik tarzları arasında önemli farklılıklar mevcuttur. Müzik severler, türler arasında ayırım yapmak ve beğendikleri şarkıları dinlemek için teknolojiden faydalanmaktadır. Müzik bilgi erişimi, MBE (music information retrieval, MIR), alanında ses dosyalarını türe göre sınıflandırmak oldukça zorlu bir süreçtir [1].

Son yıllarda, bilgisayarlar müzik tüketiminin hemen hemen her unsurunda yoğun bir şekilde yer almakta ve bugünün müzik üretiminin çoğunda kritik bir rol oynamaktadır. Bu, kişisel koleksiyonları gözden geçirmeyi, yeni sanatçılar keşfetmeyi ve müzik üreticilerinin haklarını yönetmeyi ve korumayı içerir [2]. Günümüzde mobil cihazların artan popülaritesi ve geniş internet erişimi ile dijital müzik, insanlar için artık kolayca erişilebilir durumdadır. Çevrimiçi müzik veri tabanlarının büyümesi ve müzik içeriğinin kolay erişimi ile insanlar dinledikleri şarkıları yönetmekte giderek zorlanmaktadır. Şarkıları kategorize etme ve düzenleme genellikle müzik türlerine dayanmaktadır.

Müzik türü, kültür, sanatçılar ve pazar güçlerinin etkileşimine dayalı olarak müzik tanımlarının kategorize edilmesini ifade eder. Müzik türleri müzisyenler veya besteler arasındaki benzerlikleri göstererek müziği koleksiyonlar halinde sıralamaya yardımcı olurlar [3]. İşitsel müzik verilerinin çok farklı özellikleri nedeniyle, değişik türler arasında ayırım yapmak kolay değildir [4]. Tür sınıflandırması, benzer verileri uyumlu kimlikler halinde kategorize etme yöntemi olarak hizmet eder; onları paylaşılan ritmik kalıplar, enstrümantal kompozisyonlar veya armonik özellikler temelinde tanımlar ve özlerini kapsayan tek bir etiket atar.

Müzik türü, müzik içeriği sınıflandırmasında önemli bir terimdir çünkü müziğin etkili bir şekilde organizasyonunu ve erişimini sağlar. Sonuç olarak, müzikal içeriğe daha iyi erişebilmemiz için doğru ve etkili bir müzik türü sınıflandırma sistemi gereklidir [5]. Bu

nedenle, mzik tr sınıflandırması, dijital mzik platformları olan Tidal, SoundCloud, Spotify ve Apple Music iin hayati nem taşıyan en nemli MBE tekniklerinden biri olarak ortaya ıkmıştır. Otomatik mzik tr sınıflandırma teknikleri, nemli bilgileri doęrudan ses ierięinden ıkarabilir. Bu tr bilgiler, alma listesi oluřturma, mzik nerisi ve arama optimizasyonu gibi birok uygulama sisteminin performansını artırmak iin kullanılır.

Mzik trlerinin sınıflandırılması, gemiřte mzisyenler ve eleřtirmenler tarafından biimsel ve tematik zelliklere dayanarak elle yapılıyordu [2]. Ancak son dnemlerde dijital teknolojinin geliřmesiyle mzik sınıflandırmasında byk bir deęiřim yařandı. Yapay zekâ ve makine ęrenimi algoritmaları kullanılarak mzięi analiz etmek ve sınıflandırmak iin yeni teknikler ortaya ıkmıştır [2]. Bařlangıta, mzik tr sınıflandırması iin geleneksel makine ęrenimi modelleri ve zellik ıkarımı yntemleri kullanılıyordu. Her uygulama, tını, ritim, perde, vuruř gibi farklı zellik trlerini kullanmaktaydı. Hangi zelliklerin ıkarıldıęı kiřiden kiřiye deęiřmekteydi. Daha sonra, byk veri tabanlarını eęitmedeki avantajları nedeniyle derin ęrenme modelleri sınıflandırma problemlerinde kullanılmaya bařlandı [6]. zellikle Evriřimli Sinir Aęları (convolutional neural networks, CNN), ham veya az iřlenmiř verilerden otomatik olarak zellik ęrenme ve ıkarma yetenekleri sayesinde byk bir potansiyel vaat etmektedir. Ses biliminin geliřmesiyle birlikte, mzięin akustik zelliklerini analiz etmek ve spektral desenlere dayalı olarak sınıflandırmak iin spektroskopi kullanılmaya bařlandı. Derin ęrenmede, evriřimli sinir aęları, grsel grnt analizi iin yaygın olarak kullanılan bir derin sinir aęı kategorisidir. Dolayısıyla, CNN’ler, spektral temsillerini kullanarak mzięi farklı trlere ayırmak iin de kullanılabilir.

Mzik sınıflandırması, basit manuel sınıflandırmalardan yapay zeka teknikleri kullanılarak yapılan karmařık sınıflandırmalara kadar zaman iinde byk geliřmeler gstermiřtir. Bu geliřmeler, mzięi daha iyi anlamamıza yardımcı olmakla kalmamıř, aynı zamanda dijital aęda dinleyicilerin ve mzisyenlerin ihtiyalarını karřılamada da nemli rol oynamıştır. Bir sonraki blmde, sınıflandırma yapmak iin neden Trk mzik trn ve CNN’lerin seildięi aıklanacaktır.

1.1.Problem Tanımı

Otomatik mzik tr sınıflandırmasında kayda deęer bir ilerleme kaydedilmiř olmasına raęmen, bu yaklařımların oęu Batı mzięi trleri iin tasarlanmıştır. Bu trler iin, otomatik sınıflandırma konusunda tr sınıflandırmaları aısından karřılařtırmalı sonular veren zmler ve yaklařımlar bulunmaktadır [7,8]. Batı mzięi arařtırmalarının oęu, Batı mzięinde en iyi

bilinen türleri içeren popüler GTZAN [9] veri setini kullanmıştır. Ne yazık ki, bu durum Türk müziği için geçerli değildir. Türk müziğinin artan popüleritesine rağmen, Türk müziği türü sınıflandırmasıyla ilgili yeterli çalışma bulunmamaktadır. Türk müziği türü literatürü, çeşitli türleri etkili bir şekilde kategorize etmeye yardımcı olacak yöntemler ve veri setlerinden yoksundur. Türk müziği türü sınıflandırmasıyla ilgili az sayıdaki çalışmada kullanılan tüm yöntemler, makine öğrenme algoritmaları ve akustik özellikleri içeren klasik tekniklere odaklanmaktadır[10]. Bu durum, Türk müziği türü sınıflandırmasının performansını sınırladığı gibi, Türk müziği türü veri setlerinin bulunabilirliğini de sınırlamaktadır. Bu da bu müzik veri tabanlarının etkili bir şekilde düzenlenmesinde bir boşluk bırakmaktadır.

1.2.Tezin Katkısı

Bu tez, Türk müziği türü sınıflandırma yöntemlerinde var olan sınırlamaları, CNN'leri ve Mel-spektrogram analizi kullanarak Türk müziği türlerini otomatik olarak sınıflandırmak suretiyle ele almayı amaçlamaktadır. Altı bilinen Türk müziği türünü içeren yeni bir veri seti oluşturulmuştur: Arabesk, Pop, Rap, Rock, Türk Sanat Müziği ve Türk Dini Müziği. Veri setimizde toplam 1167 ses parçası bulunmaktadır. Sunulan yöntem CNN'e girdi olarak Mel-spektrogramlardan üretilen spektrogram görüntüleri kullanarak Türk müzik türü sınıflandırmasında önceki çalışmalara [10] kıyasla sınıflandırma doğruluğunda önemli iyileştirmeler sağlamaktadır. Sunulan yöntem Türk müziğinin otomatik sınıflandırılması için bir çerçeve sunmakta ve müzik sınıflandırma araştırmalarında yeterince keşfedilmemiş bir alanı ele almaktadır.

1.3.Tez Yapısı

Tezin kalan kısmı aşağıdaki bölümlere ayrılacaktır.

Bölüm 2 - Literatür Taraması, bu tezle ilgili konularda yapılan önceki çalışmaları göstermektedir. Bu bölümde, batı müziği ve Türk müziğinde kullanılan önceki veri setleri ile batı ve Türk müziği tür sınıflandırması üzerine yapılan araştırmalar ele alınmaktadır. Ayrıca, sınıflandırma için kullanılan yaklaşımlar açıklanmaktadır.

Bölüm 3 - Ön Bilgiler, bu tezde kullanılacak tüm tekniklerin bir özetini sunar. Ses işleme kısmı, ses sinyalinin kısa zamanlı Fourier dönüşümü kullanılarak Mel spektrograma dönüştürülmesini açıklar. CNN kısmı ise, bu ağların iç yapısını açıklar.

Bölüm 4 - Yöntem , bu bölümde, oluşturulan Türk müziği veri setini ve bu Türk müziğini sınıflandırmak için önerilen yöntemi açıklanmaktadır. Ayrıca, yöntemin başarısını artırmak için önerilen yöntemle karşılaştırmada kullanılan yöntem açıklanmaktadır.

Bölüm 5 – Sonuçlar, bu bölümde Mel spektrogramlarını ve CNN modelini kullanan deneylerden elde edilen sonuçlar sunulmaktadır. Ayrıca, önerilen yaklaşımın sonuçları ile Özbalcı'nın [8] çalışmasından elde edilenler arasında karşılaştırmalı bir analiz sunulmaktadır.

Bölüm 6 - Sonuç ve gelecek çalışmalar, bu bölümde tez kapsamında elde edilen bulguların ve yapılan katkıların bir özeti sunulmaktadır. Ayrıca gelecekteki araştırmaların geliştirilmesine yardımcı olabilecek bazı öneriler de sunulmaktadır.



2. LİTERATÜR TARAMASI

Başlangıçta müzik kategorizasyonu ve tür sınıflandırması insanlar tarafından manuel olarak yapıyordu. Bu alanda çalışan insanlar, önce sesi veya müziği dinler, ardından onu daha önceki benzer müzik formlarına veya tarzlarına göre belirli bir türe göre sınıflandırırlardı. Ancak büyük verinin yükselişinden bu yana, insanların kullanabileceği veri miktarı önemli ölçüde arttı ve bu da manuel olarak çalışmayı pratik bir uygulama olmaktan çıkardı. Bu nedenle çok sayıda araştırmacı çabalarını müzikal parametreleri keşfetmeye ve bunların farklı türlere göre sınıflandırılması için yöntemler geliştirmeye adanmış. Bu bölümde müzik türü sınıflandırmasında daha önce yapılan araştırmalar açıklanacak ve bu alanda kullanılan mevcut veri setleri ve yöntemler açıklanacaktır.

2.1. Mevcut Veri Setleri

Bu bölüm, önceki araştırmalarda müzik türü sınıflandırması için kullanılan en popüler veri setlerini açıklamaktadır. İlk olarak, ünlü GTZAN veri seti [9], müzik türü sınıflandırma için MNIST olmayı hak etmektedir. Bu veri setini kullanan ilk makale [8], modern müzik sınıflandırmasında temel bir çalışma olarak kabul edilmektedir. Bu veri seti, Batı müziğinin 10 türünü içermektedir. Bu türler blues, klasik, country, disko, hip-hop, caz, metal, pop, reggae ve rock'tır. Her türde 100 adet ses kaydı bulunmaktadır. Her kayıt 30 saniye uzunluğundadır.

MagnaTagATune [11], yayınlandığı tarihten itibaren çokça kullanılmıştır. MagnaTagATune, "etiketleme" problemi için tasarlanmış olup, bu ölçekte ve sesle birlikte gelen ilk etiketleme veri setlerinden biridir. MagnaTagATune veri seti, her biri 29 saniye uzunluğunda olan 25.863 müzik içermektedir.

Million Song Dataset (MSD) [12], çığır açıcı bir müzik veri setidir. Boyut, kalite, güvenilirlik ve çeşitli tamamlayıcı özellikler açısından döneminin ilerisinde olan bir veri setidir. Veri setinin temelini, The Echo Nest tarafından sağlanan bir milyon şarkının özellik analizi ve meta verileri oluşturur. Veri seti herhangi bir ses içermemekte, yalnızca türetilmiş özellikleri barındırmaktadır.

Free Music Archive veri seti (FMA) [13], kısa ön izleme klipleri yerine tam parça içeren modern ve büyük ölçekli bir veri setidir. Tam uzunlukta ve yüksek kaliteli ses, önceden hesaplanmış özellikler, parça ve kullanıcı düzeyinde meta veriler, etiketler ve biyografiler gibi serbest biçimli metinler sağlar.

MTG-Jamendo [14], MSD ve FMA'nın bazı avantajlarını paylaşan açık bir veri setidir. Sesleri hazır ve yasal olarak erişilebilir, sesler tam parça ve yüksek kalitede, çeşitli ve gerçekçi etiketler içerir ve uygun şekilde tanımlanmış bölümlerle birlikte gelir.

Görüldüğü gibi, Batı müziğindeki veri setleri geniş çapta bulunmaktadır. Ne yazık ki, Türk müziğindeki veri setleri aynı düzeyde bulunmamaktadır. Bu bölümün devamında Türk müziğinde kullanılan birkaç veri seti açıklanacaktır.

Türk Müziği Duygu Veri Seti [15] , bu veri seti kesikli bir model olarak tasarlanmış olup, veri setinde dört sınıf bulunmaktadır: mutlu, üzgün, öfkeli, rahat. Veri setini hazırlamak için, farklı Türk müziği türlerinden sözlü ve sözsüz müzikler seçilmiştir. Her sınıfta eşit sayıda örnek olması için veri setinde her sınıf için toplam 100 müzik parçası belirlenmiştir. Orijinal veri setinde her biri 30 saniye olan 400 örnek bulunmaktadır.

Türk Makam Müziği Sembolik Temsili Veritabanı (SymbTr) [16], metin, PDF ve MIDI formatlarında müzikal bilgilerden oluşur. Güvenilir kaynaklardan elde edilen ham verilerden oluşur ve 2200 müzik parçasını içerir. Bu iki veri setinde, Türk müzik türlerinin olmadığını gözlemlemektedir. İlk veri setinde Türk müziğini dinlerken insanların hissettiği duyguyu açıklarken, ikincisi sadece Türk müziğinin makamını içerir.

Araştırmalarımızda Türk müziğinin türlerini içeren yalnızca bir veri seti bulunmuştur. Özbacı'nın [10] çalışmasının veri seti, Türk müziği parçalarından derlenmiş olan altı Türk müziği türünü içermektedir. Bunlar arabesk, pop, rock, dini, rap ve klasik müziktir. Veri seti, her biri 100 parça olmak üzere, wav uzantılı 600 parçadan oluşmaktadır. Bu veri seti az sayıda ses örneklerine sahiptir, bu nedenle bu tezde kendi veri setimizi oluşturulmuştur.

2.2. Mevcut Çalışmalar

Birçok araştırmacı, müzikal parametreleri keşfetmeye ve bunları farklı türleri sınıflandırmak için kullanan yöntemler geliştirmeye çalışmıştır. Bu bölüm, alandaki mevcut araştırmaları gözden geçirmeye odaklanacaktır.

Tzanetakis ve Cook [8], makine öğrenimi algoritmalarını kullanarak müzik türü sınıflandırması problemini ele almış ve Gauss Karışım Modeli ile k-en yakın komşu sınıflandırıcıları gibi denetimli öğrenme yaklaşımlarını bu alanda öncü olarak uygulamıştır. Çalışmalarında, tınısal yapı, ritmik içerik ve perde içeriği şeklinde kategorize edilen üç farklı özellik seti sunulmuştur. Tzanetakis ve Cook [8], ayrıca hala geniş çapta bilinen bir ölçüt olarak GTZAN veri setini de tanıtmışlardır .

Changsheng Xu ve diğ. [17], yaptığı çalışmada, destek vektör makineleri (support vector machines, SVM) kullanılarak müzik türü sınıflandırması görevinde denetimli öğrenme yaklaşımlarının etkili bir şekilde kullanılabileceği ortaya konmuştur.

West ve Cox [18], farklı türde ses öğeleri kullanarak karmaşık bir sınıflandırıcı oluşturmuştur. Altı farklı müzik türünü sınıflandırma görevlerinde en iyi sistemleri için yaklaşık %83'lük bir başarı oranı elde etmişlerdir. Özellikle Reggae ve Rock müzik türlerinin ayrımında, kullandıkları özellik çıkarma yönteminin yetersiz kaldığını belirtmişlerdir. Bu iki tür, benzer spektral özelliklere ve armonik/nonharmonik içerik oranlarına sahip olduklarından sınıflandırmayı zorlaştırmıştır.

Scaringella ve diğ. [19], müzik türü sınıflandırması alanında kullanılan hem özellikleri hem de sınıflandırma tekniklerini kapsayan kapsamlı bir genel bakış sunar.

Archit Rathore ve Margaux Dorido [20], GTZAN veri seti için potansiyel çözümler olarak bir dizi farklı kategorizasyon modeli sunmuşlardır. Genellikle, GTZAN veri setinin müzik türü kategorizasyonu için bir kriter olması gerektiği konusunda fikir birliği sağlanmıştır. Yazarlar en iyi doğruluğu polinom çekirdeği ile SVM kullanarak elde etmişlerdir.

J. de Sousa ve diğ. [21], en iyi doğruluk sonucunu SVM'yi temel alarak, sınıflandırmayı hem GTZAN veri seti hem de önerilen Brezilya Müzik Veri Seti (BMD) üzerinde değerlendirmişlerdir. Çalışmada, zaman-bölgesi, spektral, tonal, ses efekti, üst düzey ve ritim olmak üzere altı farklı tanımlayıcı kümesine ait özellikler kullanılmıştır. Buna ek olarak, tonalite, ses şiddeti, uyumsuzluk, keskinlik, tempo, anahtar ve ritim histogramları gibi yaygın özellikler de modele dahil edilmiştir.

P. Fulzele ve diğ. [22], LSTM ve SVM kullanarak müzik türlerini sınıflandırmak için hibrit bir model ortaya koymuştur. Bu modelin uygulanması iki aşamada gerçekleşir. Birinci aşamada, iki farklı sınıflandırıcı ayrı ayrı eğitilir. İkinci aşamada ise, toplam kuralı (iki sınıflandırıcının tahminlerinin olasılıklarını toplayarak) uygulanarak modeller birleştirilir ve nihai sınıflandırma sonucu elde edilir.

Derin öğrenme yöntemleri, son yıllarda müzik türü sınıflandırması alanında kaydedilen ilerlemelerde belirleyici bir rol oynamıştır. T. L. Li ve diğ. [4], CNN kullanarak ses yer alan müzikal örnekleri yoğunlaştırmak için alternatif bir yöntem önermektedir. Deney sonuçları, CNN'in müzikal örneklerin sapmalarından yararlı bileşenleri yakalama konusunda güçlü bir kapasiteye sahip olduğunu göstermiştir. Bu çalışma, müzikal örneklerin önemli özelliklerini otomatik olarak ses müziğinden çıkarmak için bir sistem sunar. Görüntü verisi alanından

uyarlanan CNN ile öznitelik çıkarmak için minimal ön bilgi gerekmektedir. Deneyleri, CNN'in otomatik özellik çıkarımı için uygulanabilir bir seçenek olduğunu ortaya koymuştur. Bu bulgu, müzikal veri çeşitliliğindeki içsel özelliklerin görüntü verileriyle benzer olduğunu savunan hipotezlerini desteklemektedir. Ayrıca, ses müziği tür sınıflandırmasında CNN kullanımı için en uygun parametre setini ve en iyi uygulamaları da sunmuşlardır.

Aaron [23] şarkı hazırlamak için Mel-frekans kepstral katsayıları (Mel-frequency cepstral coefficients, MFCC) spektrogramlarından yararlanmaktadır. Odak noktaları tür tanınmanın ötesine geçerek melodi analizi ve önerilerini kapsamaktadır. Özellikle, MFCC spektrogramları olarak temsil edilen yeniden işlenmiş şarkı alıntıları üzerinde ReLU aktivasyonuna sahip CNN mimarileri kullanmışlardır.

Choi'ye göre [24], derin öğrenme modellerini müzik türü sınıflandırmasına dahil etmek çeşitli avantajlar sunmaktadır. İlk olarak, derin öğrenme yapılandırılmış verilerden türetilen özelliklerin aksine, sinir ağlarının çeşitli hiyerarşik katmanlarından çıkarılan öğrenilmiş özellikler aracılığıyla sınıflandırmayı kolaylaştırır. İkinci olarak, derin öğrenme modellerinin hiyerarşik topoloji özelliklerinin, çeşitli zaman ve frekans aralıklarında müzik analizi için faydalı olduğu kanıtlanmıştır.

Lidy ve Schindler'de [25], aynı görevi başarmak için CNN'ye girdi olarak sabit Q dönüşümü spektrogramı sağlamıştır.

Lonce Wyse'a göre [26], ortak alternatif temsil, hem zaman hem de frekans bilgisini yakalayan bir sinyalin spektrogramıdır. Spektrogramlar görüntü olarak kabul edilir ve CNN'leri eğitmek için kullanılır.

Daha önce Türk müziği üzerine az sayıda araştırma vardır. M. A. Kızrak ve diğ. [27], Türk klasik müziği şarkılarını kullanmışlardır. Bu çalışmada klasik Türk müziği altı makama ayrılmıştır. Özellik olarak MFCC kullanılmıştır. Elde edilen veriler olasılıksal sinir ağları kullanılarak sınıflandırılmıştır. En iyi doğru tanıma oranı 6 sn'lik klip uzunluğu kullanılarak %89,4 olarak elde edilmiştir.

Ö. Çoban ve G. T. Özyer [28], otomatik müzik türü sınıflandırması için Türkçe şarkı sözü veri seti oluşturmuşlardır. Klasik bir metin sınıflandırma problemi olarak kabul edilen müzik türü sınıflandırmasında deneysel sonuçlar destek vektör makineleri üzerinde yapılmış ve özellik modelinin sonuçlara etkisi araştırılmıştır. Özellikler, yapısal ve istatistiksel metin özellikleri, kelime çantası ve NGram olmak üzere üç farklı modelden çıkarılmıştır.

M. B. Er ve E. M. Esin [29], Türk müziğinden farklı örneklerin duygularını tahmin etmek için ses özelliklerinin kullanılmasını önermişlerdir. Veriler [15] destek vektör makineleri, k-en yakın komşu (k-nearest neighbors, KNN) ve yapay sinir ağı kullanılarak sınıflandırılmıştır.

Türk müziğinin tüm türlerini kullanan tek çalışma Özbacı [10] tarafındandır. Türk müzik türlerini makine öğrenme algoritmaları ile akustik özellikleri kullanarak sınıflandırmıştır. Türk müzik türünü sınıflandırmak için kullanılan makine öğrenmesi algoritmaları SVM, KNN ve Lojistik Regresyon algoritmalarıdır. SVM algoritması kullanılarak elde edilen en yüksek doğruluk oranı %78,65'tir. İlgili çalışma, bu tez ile benzer şekilde Türk müziğinin otomatik olarak türlerine ayrılmasını ele almaktadır. Bu tezde Türk müziğinin tüm türlerinin sınıflandırılmasındaki boşluğu doldurmak ve sınıflandırmada daha iyi bir sonuç elde etmek amacıyla Türk müziği türlerini sınıflandırmanın farklı bir yolu önerilmektedir. Tartışılan literatüre uygun olarak Türk müzik içeriğini temsil etmek için Mel-Spektrogramlardan görüntüler üretilmiştir. Daha sonra CNN ile derin öğrenme tekniğini kullanılarak sınıflandırma yapılmıştır. Deneyler ve sonuç bölümünde tartışılacağı gibi deneysel sonuçlar, önerilen çalışmanın Özbacı'nın [10] çalışmasından daha iyi performans gösterdiğini göstermektedir.



3. ÖN BİLGİLER

3.1. Ses için Sinyal İşleme

Sinyal işleme, verileri doğrudan gözlem yoluyla mümkün olmayan bir şekilde görmemizi sağlayacak şekilde dönüştürmeyi veya değiştirmeyi içerir. Bu süreç, mühendislerin ve bilim insanlarının bilimsel verileri, ses akışları, görüntüler ve videolar dahil olmak üzere sinyalleri analiz etmelerine, optimize etmelerine ve düzeltmelerine olanak tanır.

Ses sinyali işleme, ses sinyallerinin elektronik manipülasyonu ile ilgilenen bir sinyal işleme alt alanıdır [30]. Ses sinyalleri, hava yoluyla yayılan ve sıkışma ve genişlemelerden oluşan boyuna dalgalar olan ses dalgalarının elektronik temsilleridir. Ses sinyallerinde veya ses gücü seviyesinde bulunan enerji genellikle desibel cinsinden ölçülür. Ses sinyalleri dijital veya analog formatta temsil edilebildiği için işleme her iki alanda da gerçekleştirilebilir. Analog işlemciler doğrudan elektrik sinyali üzerinde çalışırken, dijital işlemciler matematiksel olarak sinyallerin dijital temsili üzerinde çalışırlar.

Ses dalgaları, hava veya su gibi bir ortamda yayılan hava moleküllerinin titreşimleridir. Bu dalgalar, müzik aletleri, hoparlörler veya insan sesi gibi ses kaynakları tarafından oluşturulur ve insan kulağı tarafından ses olarak algılanır. Ses dalgalarını dijitalleştirmek, analog ses dalgalarını bir makinenin depolayabileceği ve işleyebileceği dijital sinyallere dönüştürmek anlamına gelir. Bu süreç birkaç adımı içerir:

1. Örneklem (Sampling): Analog ses dalgası düzenli aralıklarla ölçülür ve her ölçüme sayısal bir değer atanır. Bu ölçümlerin alındığı hız, örneklem hızı olarak adlandırılır ve genellikle kilohertz (kHz) cinsinden ölçülür. Örneklem hızı ne kadar yüksek olursa, dijital sinyalin orijinal analog dalgayı o kadar doğru bir şekilde temsil etmesi sağlanır. Yaygın örneklem hızları, CD kalitesi için 44.1 kHz ve profesyonel ses kayıtları için 48 kHz veya daha yüksek olabilir.

2. Niceleme (Quantization): Örneklem yoluyla elde edilen sayısal değerler en yakın tam sayıya yuvarlanır. Niceleme işlemi, sinyalin her bir örneğinin belirli bir bit derinliğinde ifade edilmesiyle gerçekleştirilir. Kullanılan bit sayısı, dijital sinyalin dinamik aralığını, yani temsil edilebilecek en sessiz ve en yüksek sesler arasındaki farkı belirler. Bit sayısı ne kadar yüksek olursa, dinamik aralık da o kadar geniş olur ve daha hassas bir ses temsil edilir.

3. Kodlama (Encoding): Nicelendirilmiş değerler, bir bilgisayarda veya diğer dijital depolama ortamlarında saklanabilecek dijital bir formata, örneğin WAV veya MP3 dosyasına,

dönüştürülür. Bu süreç, ses verilerinin sıkıştırılmasını ve daha verimli bir şekilde saklanmasını sağlar. Kodlama formatı, dosya boyutunu ve ses kalitesini doğrudan etkileyebilir; WAV dosyaları genellikle daha yüksek kaliteli ve kayıpsızdır, MP3 dosyaları ise daha küçük boyutlu ve kayıplıdır.

Ses sinyali, sesin bir temsilidir. Notalar ve sembolik temsillerin aksine, ses temsili, bir müzik parçasının akustik bir gerçekleştirmesini yeniden üretmek için gereken tüm bilgileri kodlar. Ses olaylarının akustik özellikleri, spektrogram gibi bir zaman-frekans görüntüsünde görselleştirilebilir. Spektrogram oluşturma, Fourier dönüşümünü (Fourier Transform, FT) zamanla değişen sinyal analizine uyarlanması esasına dayanır [31].

3.1.1. Fourier dönüşümü

Fourier dönüşümü, ses sinyali işlemede en önemli matematiksel araçtır. Fourier dönüşümü, zamana bağlı bir sinyali frekansa bağlı bir fonksiyona dönüştürür. Ters işlem ise, bir sinyali bağımsız temel fonksiyonların ağırlıklı bir üst üste binmesi olarak temsil eden Fourier temsili ile gerçekleştirilir [2]. Bu işlem, sinyali zaman alanından frekans alanına dönüştürür ve ortaya çıkan sonuç spektrum olarak adlandırılır.

Fourier dönüşüm spektrumları kısa aralıklarla üst üste binen pencereleri olan sektörler için hesaplanır. Benzer ses olarak algılanan gerçek dünya sinyallerinin zaman alanı dalga biçimleri, aslında frekans bileşenleri arasındaki değişen faz ilişkilerine bağlı olarak çok fazla değişiklik gösterir. Kısa süreli faz spektrumu, karşılık gelen büyüklük veya güç spektrumu kadar algısal olarak anlamlı değildir ve sinyal gösteriminde ihmal edilmiştir [32]. Çalışan hacim spektrumlarından, sinyalin zaman-frekans içeriğinin veya spektrogramın grafiksel bir gösterimi üretilir.

Her bir sinyal uzayının kendi Fourier dönüşümüne sahip olduğunu ve ilgili Fourier dönüşümünün varlığını ve özelliklerini kanıtlamak için gereken matematiksel kavramların çeşitlilik gösterdiğini belirtmek önemlidir. Bu tezde ses sinyalinin Mel-spektrogramını elde etmek için kısa süreli Fourier dönüşümüne odaklanılmıştır.

3.1.2. Kısa süreli Fourier dönüşümü (Short-Time Fourier Transform, STFT)

Kısa zamanlı Fourier dönüşümü (Short-Time Fourier Transform, STFT), zaman içinde değişen bir sinyalin yerel parçalarının sinüzoidal frekansını ve faz içeriğini belirlemek için kullanılan bir Fourier dönüşümüdür [33]. Bu dönüşüm, zaman ve frekans temelli gösterimler arasında bir denge sağlar, böylece bir sinyalin yerel parçalarının sinüzoidal frekansını ve faz içeriğini

belirlerken zamansal deęişimleri de dikkate alır. Bu nedenle, STFT yalnızca sinyalin hangi frekansları içerdiğini deęil, aynı zamanda bu frekansların hangi zaman noktalarında veya zaman aralıklarında ortaya çıktığını da belirtir [2].

Pratikte, STFT'leri hesaplamanın prosedürü, daha uzun bir zaman sinyalini eşit uzunluktaki daha kısa segmentlere bölmek ve ardından her kısa segmentte ayrı ayrı Fourier dönüşümünü hesaplamaktır. Bu, her kısa segmentteki Fourier spektrumunu ortaya çıkarır. Genellikle, zamanın bir fonksiyonu olarak deęişen spektrumları, spektrogram olarak adlandırılır. Sürekli alandaki STFT'nin matematiksel Denklem (3.1)'de verilmiştir:

$$STFTx(t) = X(T, w) = \int_{-\infty}^{\infty} x(t)w(t - T)e^{-jwt}dt \quad (3.1)$$

Burada $x(t)$ zaman alanında orijinal sinyaldir. $w(t - T)$ pencere fonksiyonudur. Pencere fonksiyonu, sinyalin belirli bir zaman aralığında analiz edilmesini sağlar. $w(t - T)$, T zaman kaydırması ile $w(t)$ pencere fonksiyonunu kaydırır. Pencere fonksiyonu, genellikle belirli bir genişlikte (örneğin Hamming, Hanning, veya Gauss penceresi) seçilir. T zaman kaymasıdır. T , sinyalin analiz edildięi zaman anını belirtir. Farklı T deęerleri, sinyalin farklı zaman dilimlerinde analiz edilmesine olanak tanır. e^{-jwt} karmaşık üstel fonksiyondur ve Fourier dönüşümünde kullanılır. Burada w açısal frekansıdır ve sinyalin frekans içeriğini analiz etmek için kullanılır. STFT, zaman ve frekans arasında bir ilişki kurarak sinyalin frekans bileşenlerini zamanın belirli noktalarında analiz eder.

Ayrıca, ayrık alan Denklem (3.2)'de verilmiştir:

$$X(n, w) = \sum_{m=-\infty}^{\infty} x(m)w(n - m)e^{-jwm} \quad (3.2)$$

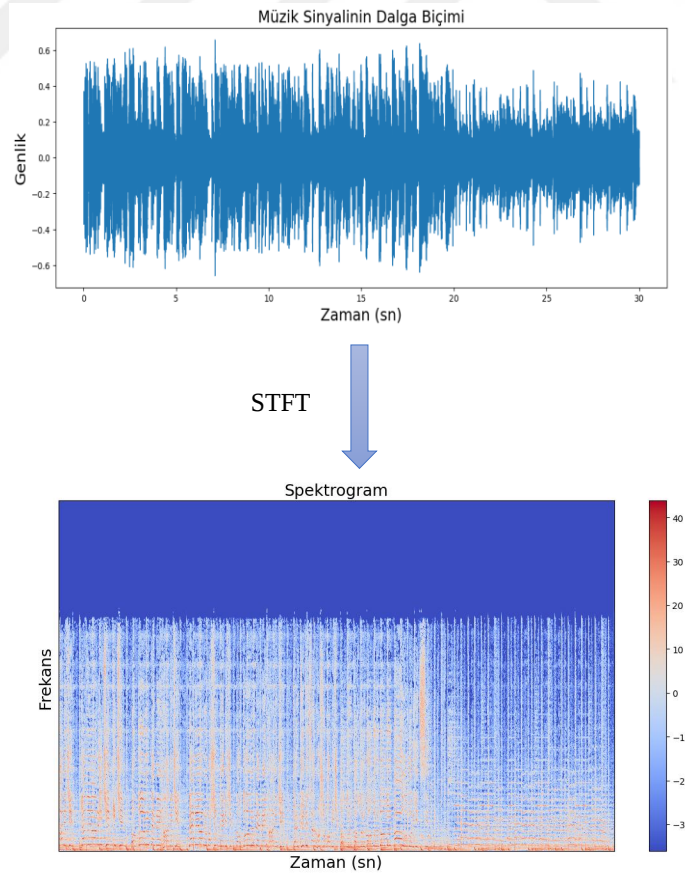
Burada $x(m)$ zaman alanında ayrık bir sinyaldir. Sinyal örneklerinin indekslendięi zaman deęişkeni m ile gösterilir. $w(n - m)$ pencere fonksiyonudur. $w(n - m)$ pencere fonksiyonunun n anında m ile kaydırılmış halidir. Bu pencere fonksiyonu, sinyalin belirli bir zaman dilimindeki analizini sağlar. $w(n - m)$ genellikle belirli bir genişlikte seçilen bir fonksiyondur ve sinyalin belirli bir kısmını izole etmek için kullanılır. n zaman kaydırma indeksidir. n sinyalin hangi zaman anında analiz edileceğini belirler. Bu deęişken, sinyalin zaman-frekans analizi sırasında kullanılır. e^{-jwm} karmaşık üstel fonksiyondur ve Fourier dönüşümünün temel bileşenidir. Bu terim, sinyalin frekans bileşenlerinin analiz edilmesine olanak tanır. ω frekans deęişkenidir. Her sabit n için, STFT o belirli zaman indeksindeki bir spektrum (frekans içerięi) sağlar.

Ayrık STFT'nin filtreleme görünümünü Denklem (3.3)'te ifade edilmiştir:

$$X(n, w_0) = \sum_{m=-\infty}^{\infty} (x(m)e^{-jw_0m})w(n-m) = (x(n)e^{-jw_0n}) * w(n) \quad (3.3)$$

Burada $X(n, w_0)$ zaman ve frekansa bağlı bir dönüşümü veya analizi temsil eder. Burada, n zamandaki konumu ve w_0 belirli bir frekansı gösterir. Bu dönüşüm, sinyalin hem zaman hem de frekans alanındaki özelliklerinin analiz edilmesine olanak tanır. $x(m)$ süreksiz zaman sinyali. m zaman indeksidir ve sinyalin o andaki değerini gösterir. $w(n-m)$ pencere fonksiyonudur. Sinyalin belirli bir bölümünü seçmek ve o bölümü analiz etmek için kullanılır. e^{-jw_0m} kompleks üstel bir ifade olup, sinyalin frekans kaydırmasını temsil eder. Bu çarpan, sinyalin belirli bir frekans bileşenine odaklanmak için kullanılır. Burada j sanal birimdir ve w_0 açısal frekansın bir ölçüsüdür.

Şekil 3.1 ses sinyaline STFT uygulandıktan sonra elde edilen spektrogram görüntüsünü göstermektedir.



Şekil 3.1 STFT'nin ses sinyalindeki etkisi

3.1.3. Mel-spektrogram

Mel spektrogramı, Mel-frekans spektrogramının kısaltmasıdır ve bir ses sinyalindeki frekansların spektrumunun görsel bir temsidir [34]. Bu spektrogram, frekansları Mel ölçeğine dönüştürür. Akustik bir sinyalin spektral analizi, Fourier dönüşümü ile elde edilir ve bu dönüşüm, frekansın genlik (veya büyüklük) spektrumu ve faz spektrumu olarak adlandırılan iki gerçek değerli fonksiyonunu üretir. Sinyalin zamanla değişen özelliklerini izlemek için, bu dönüşüm kare büyüklüğün iki boyutlu bir temsidir [2]. Spektrogram matematiksel Denklem (3.4)'te verilmiştir:

$$Spec(t, \omega) = |\tilde{f}_g(t, \omega)|^2 = |\tilde{f}^g(t, \omega)|^2 \quad (3.4)$$

Burada $Spec(t, \omega)$ sinyalin spektrogramını temsil eder. Sinyalin frekans içeriğinin zamanla nasıl değiştiğini gösteren bir zaman-frekans temsidir. $\tilde{f}_g(t, \omega)$ veya $\tilde{f}^g(t, \omega)$ sinyalin frekans içeriğinin zamanla nasıl değiştiğini tanımlayan karmaşık bir fonksiyondur. Genellikle bunun, orijinal sinyalin $f(t)$ dönüşmüş bir versiyonu olduğunu gösterir. Alt veya üst simge olarak kullanılan g , belirli bir pencere fonksiyonunu veya işleme tabi tutulmuş sinyalin genelleştirilmiş bir formunu gösterebilir. t zaman ve ω frekans temsil eder.

Mel ölçeği, insan kulağının farklı frekanslara verdiği tepkiyi taklit eden bir perde ölçeğidir. İnsanların perdeyi algılama şeklini daha iyi temsil eden doğrusal olmayan bir ölçektir. Bu logaritmik ölçek, dinleyiciler tarafından birbirinden eşit uzaklıkta olarak algılanan perdeleri temsil eder, her perde sınıfı bir öncekinin iki katı frekansa sahiptir. Mel ölçeği, belirli formüller kullanılarak Hertz'e ve tersine Denklem (3.5), (3.6)'de dönüştürülebilir [35].

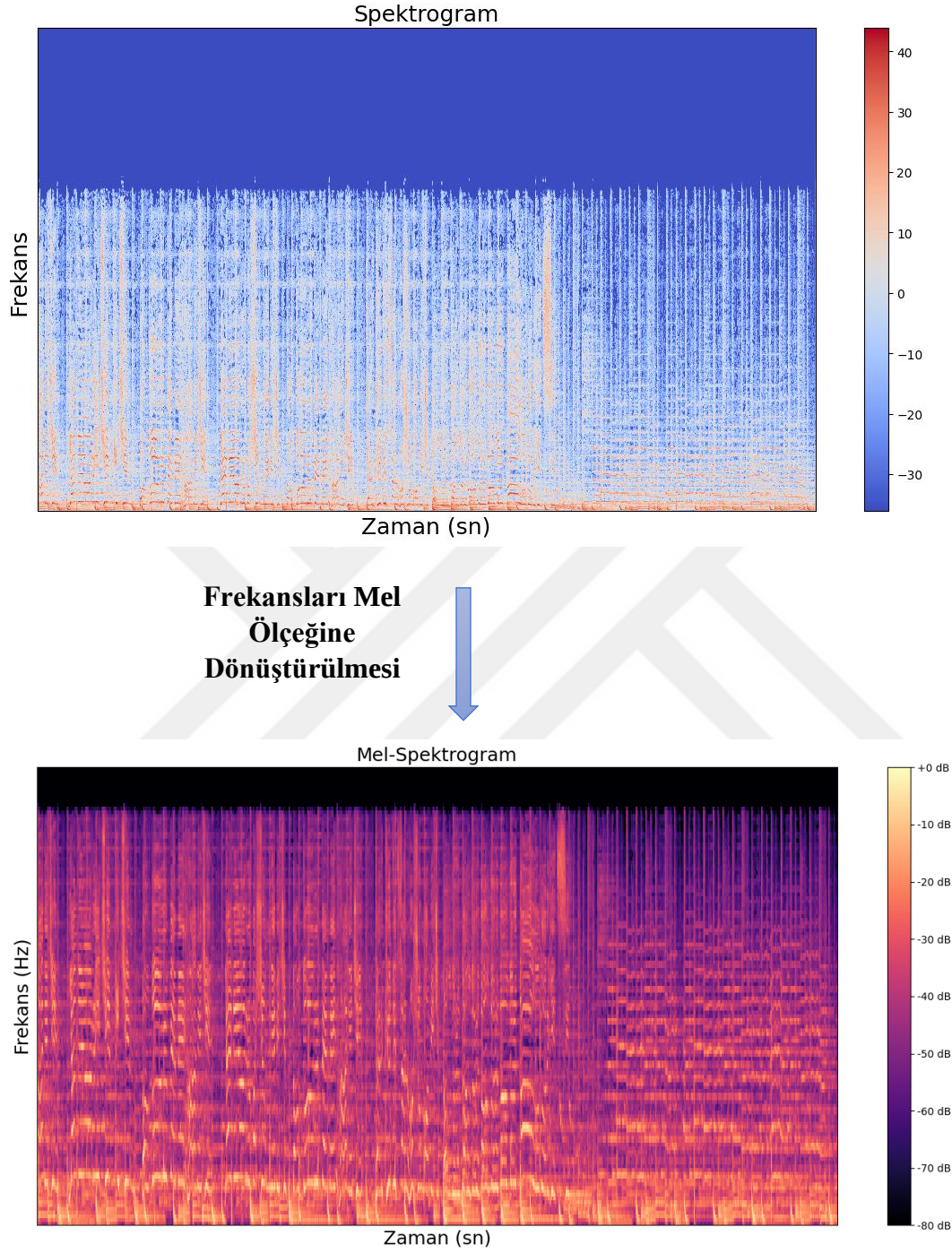
$$m = 2595 \cdot \log \left(1 + \frac{f}{500} \right) \quad (3.5)$$

$$f = 700 \left(10^{\frac{m}{2595}} - 1 \right) \quad (3.6)$$

Burada m , Mel'i ve f , Hertz'i temsil eder.

Mel-spektrogram elde etmenin iki ana adımı SFFT'yi çıkarmak ve frekansları Mel ölçeğine dönüştürmektir. Pratik olarak Mel-spektrogramın tanımı, zamanın x-ekseninde ve Mel frekansının y-ekseninde bulunduğu, ses sinyalinin iki boyutlu bir temsili olarak kabul edilebilir. Belirli bir zaman aralığında belirli bir frekansın büyüklüğünü (veya gücünü) görsel olarak

iletmek için bir renk haritası kullanır. Şekil 3.2. Mel-spektrogramına dönüşümünü açıklamaktadır.



Şekil 3.2. Mel-spektrogramına dönüşüm

Mel-spektrogramları, insan algısıyla uyumlu bir şekilde sinyalin frekans içeriğini yakalamak için önemli olduğu ses işleme görevlerinde sıkça kullanılır; bu görevler arasında konuşma tanıma ve müzik analizi yer alır. Ses sinyalinin spektrum özellikleri hakkında bilgilendirici bilgiler sunarlar.

3.2. Derin Öğrenme

Derin öğrenme, temsili öğrenmeye dayalı sinir ağları ile gerçekleştirilen makine öğrenmesi yaklaşımlarının bir alt kümesidir. Derin öğrenme yöntemleri, birkaç temsil katmanına sahip temsili öğrenme yöntemleridir ve basit ancak doğrusal olmayan modüller inşa edilerek, bir seviyedeki temsili ardışık olarak daha yüksek ve biraz daha soyut bir seviyedeki temsile dönüştürülmesiyle oluşturulur [36]. Yeterli sayıda dönüşüm birleştirildiğinde, çok karmaşık fonksiyonların öğrenilmesi mümkün olur. Temsilin daha yüksek katmanları, sınıflandırma görevleri için kritik olan girdi özelliklerini vurgular ve sınıflandırma için önemsiz olan varyasyonları bastırır. Bu katmanlar, birlikte çalışarak girdilerden öğrenen ve onları işleyen düğümler (nöronlar) ile birbirine bağlıdır.

Derin öğrenme, yapay zekâ alanının yıllardır çözmekte zorlandığı sorunları çözmede önemli ilerlemeler kaydetmektedir [37]. Yüksek boyutlu verilerdeki karmaşık yapıları algılamada son derece başarılı olduğunu kanıtlamıştır ve bu nedenle birçok araştırma ve iş sektörüne uygulanabilir.

Derin öğrenme mimarileri, derin sinir ağları, derin inanç ağları, tekrarlayan sinir ağları, evrişimli sinir ağları ve transformatörler gibi, bilgisayarla görme, konuşma tanıma, doğal dil işleme, makine çevirisi, biyoinformatik, ilaç tasarımı, görüntü analizi, iklim bilimi, malzeme denetimi ve masa oyunu programları gibi alanlarda uygulanmıştır ve bazı durumlarda insan uzmanlarıyla karşılaştırılabilir, hatta onları aşan sonuçlar üretmiştir [38, 39].

Derin öğrenmede en yaygın kullanılan mimariler ileri beslemeli sinir ağları, evrişimli sinir ağları (CNN'ler) ve tekrarlayan sinir ağlarıdır (RNN'ler). Bu tez, Türk müzik türlerini sınıflandırmak için evrişimli sinir ağlarını kullanmaya odaklanmaktadır.

3.2.1. Evrişimli sinir ağları (CNN)

Evrişimli sinir ağları (CNN), en çok görsel görüntüleri analiz etmek için kullanılan bir derin sinir ağı türüdür. CNN, özellikle görüntü ve video tanıma görevleri için çok boyutlu diziler biçimindeki girdileri yorumlamak amacıyla tasarlanmıştır [40]. CNN, özellikleri manuel olarak çıkarmaya ihtiyaç duymaz. Görüntülerden otomatik olarak özellikler öğrenebilir, bu da onları görüntü sınıflandırma, nesne tanımlama ve segmentasyon gibi görevler için ideal hale getirir. Örneğin, bir renkli görüntü, üç renk kanalındaki piksel yoğunluklarını içeren üç boyutlu diziden oluşur. Birçok veri biçimi, farklı boyutlarda diziler şeklindedir: sinyaller ve diziler için tek boyutlu; görüntüler veya ses spektrogramları için iki boyutlu; video veya hacimsel görüntüler için üç boyutlu. CNN'nin, doğal sinyallerin özelliklerinden yararlanmak için dört

ana fikri vardır: yerel bağlantılar, paylaşılan ağırlıklar, havuzlama ve birçok katmanın kullanımı.

CNN mimarisi, geleneksel sinir ağları gibi yalnızca matris çarpımlarına dayanmak yerine, özel bir teknik olan Evrişim'i (konvolüsyon) kullanır. CNN'nin mimarisi, görsel algıdan esinlenmiştir [41]. Biyolojik bir nöron, yapay bir nörona karşılık gelir ve CNN çekirdekleri çeşitli özelliklere yanıt veren reseptörler gibi davranır. Aktivasyon fonksiyonları, yalnızca belirli bir eşiği aşan sinir sinyallerinin bir sonraki nörona iletildiği biyolojik süreci simüle eder. Kayıp fonksiyonları ve optimizasyon algoritmaları ise, tüm CNN sistemini istenen sonuçları öğrenmesi için eğitmek amacıyla tasarlanmış yapay yapılardır.

CNN'nin birçok avantajı vardır:

- Yerel bağlantılar: Her bir nöronu, önceki katmandaki tüm nöronlara bağlamak yerine daha küçük bir nöron grubuna bağlayarak parametreleri azaltır ve yakınsamayı hızlandırır.
- Ağırlık paylaşımı: Birden çok bağlantının aynı ağırlıkları paylaşmasını sağlar, bu da parametrelerin sayısını azaltır.
- Alt örnekleme ile boyut azaltma: Bir havuzlama katmanı, bir görüntüyü alt örnekleme için görüntü yerel korelasyon ilkesinden yararlanır. Bu işlem, görüntünün en önemli özelliklerini korurken daha küçük boyutlu bir temsil elde etmeyi sağlar. Bu üç özellik, CNN'i derin öğrenme alanındaki en temsili algoritmalarından biri haline getiriyor.

CNN'ler üç temel katman türünden oluşur: evrişimli katmanlar, havuzlama katmanları ve tam bağlantılı (fully connected, FC) katmanları. Bu katmanları üst üste koyarak, Şekil 3.3'te gösterildiği gibi bir CNN mimarisi ortaya çıkar. Bu üç katmana ek olarak iki önemli parametre daha vardır: nöron seyreltme katmanı (dropout) ve aktivasyon fonksiyonu. Bu katmanlar aşağıda ayrıntılı olarak açıklanmaktadır. Şekil 3.3 CNN mimarisi gösterilmiştir.

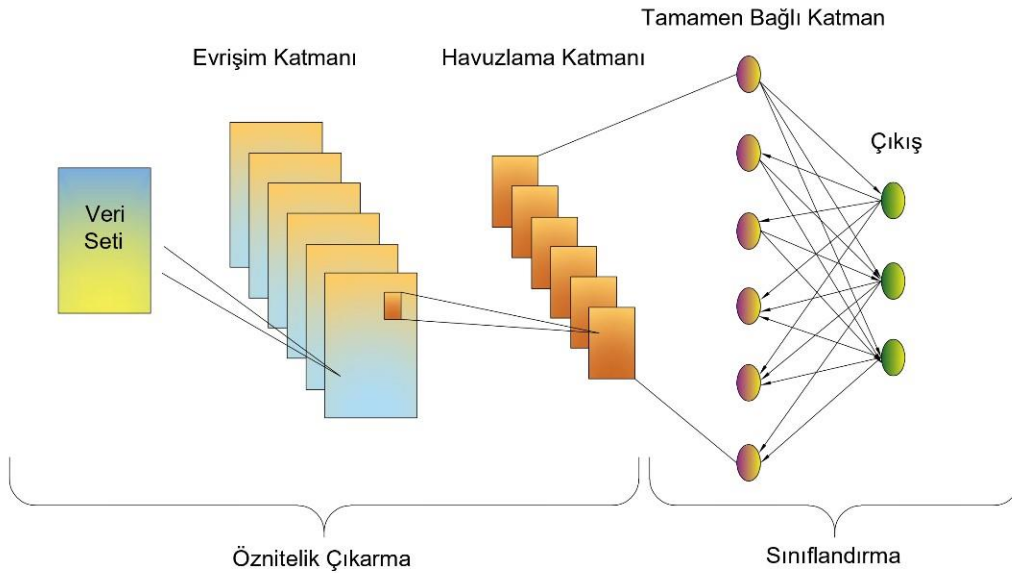
1. Evrişimsel Katman (Convolutional Layer, CONV):

Evrişim katmanı, CNN yapısının ilk adımıdır ve girdi olan görüntülerden çeşitli özellikler çıkarır. Çekirdek/Filtre, bu katmanda evrişimin matematiksel işlemini gerçekleştiren bileşendir. Evrişim işlemi, giriş görüntüsü ile belirli bir $M \times M$ boyutundaki filtre arasında gerçekleştirilir. Filtreyi giriş görüntüsü üzerinde kaydırarak, filtre boyutuna ($M \times M$) göre filtre ile giriş görüntüsünün parçaları arasında nokta çarpım alınır. Çekirdek bir görüntüden daha küçük boyuttadır ancak daha fazla derinliğe

sahiptir. Bu, görüntünün üç (RGB) kanalı varsa, çekirdek yüksekliği ve genişliğinin daha küçük olacağı, ancak derinliğin üçünü de kapsayacağı anlamına gelir.

Çıktı, bize görüntü hakkında köşeler ve kenarlar gibi bilgiler veren öznitelik haritası olarak adlandırılır. Daha sonra bu özellik haritası, giriş görüntüsünün diğer bazı özelliklerini öğrenmek için diğer katmanlara beslenir.

Girişe evrişim işlemi uygulandıktan sonra CNN'nin evrişim katmanı, çıkışı bir sonraki katmana aktarır. CNN'deki evrişimli katmanları oldukça yararlıdır çünkü pikseller arasındaki uzamsal ilişkiyi korurlar.



Şekil 3.3. Evrişimli sinir ağlar mimarisi

2. Havuzlama Katmanı (Pooling Layer):

Havuzlama katmanı genellikle bir evrişimsel katmanın ardından eklenir. Bu katmanın temel amacı, evrişimsel sonucu oluşan öznitelik haritasının boyutunu azaltarak hesaplama maliyetlerini düşürmektir [42]. Bu, katmanlar arasındaki bağlantıları azaltarak ve her öznitelik haritası üzerinde bağımsız olarak çalışarak gerçekleştirilir. Havuzlama katmanındaki öznitelik haritasının boyutu, çekirdeklerin hareket ettiği adımın belirlediği şekilde belirlenir. Tipik Havuzlama işlemleri iki türe ayrılabilir: maksimum havuzlama ve ortalama havuzlama. Şekil 3.4'te havuzlama katmanı gösterilmiştir.

Maksimum havuzlama

29	15	28	184
0	100	70	38
12	12	7	2
12	12	45	6

2×2

Havuz boyutu

100	184
12	45

Ortalama havuzlama

31	15	28	184
0	100	70	38
12	12	7	2
12	12	45	6

2×2

Havuz boyutu

36	80
12	15

Şekil 3.4. Havuzlama katmanı (Pooling layer)

Max havuzlamada, özellik haritasından en büyük öge seçilir. Ortalama havuzlama, önceden belirlenmiş bir görüntü segmentindeki bileşenlerin ortalamasını hesaplar. Toplam havuzlama, önceden belirlenmiş bölümdeki öğelerin toplamını hesaplar. Havuzlama katmanı genellikle evrişim katmanı ile tam bağlantılı katman arasında bir köprü görevi görür. Birkaç evrişim ve havuzlama katmanı bir araya getirilerek girişlerin yüksek seviye özellikleri çıkarılabilir.

3. Tam Bağlantılı Katman (Fully Connected Layer, FC):

Tam bağlantılı katman, düzleştirilmiş bir girişle çalışır, bu da her girişin her bir nörona bağlı olduğu anlamına gelir [43]. Bundan sonra, düzleştirilmiş vektör matematiksel işlemlerin sıklıkla gerçekleştirildiği birkaç ek tam bağlantılı katmana iletilir. Bu noktada sınıflandırma süreci başlar. Tam bağlantılı katmanlar genellikle çıkış katmanından önce yer alır ve bir CNN mimarisinin son katmanlarını oluşturur.

4. Nöron seyreltme (Dropout):

Nöron seyreltme katmanı, bazı nöronların sonraki katmana katkılarını ortadan kaldırırken diğerlerini değiştirmeden bırakan bir maskedir [44]. Bu katman, giriş vektörüne uygulanabilir ve bazı özelliklerini iptal edebilir veya gizli bir katmana uygulanabilir ve belirli gizli nöronları devre dışı bırakabilir. Nöron seyreltme katmanları

CNN eğitiminde çok önemlidir çünkü eğitim verilerinin aşırı öğrenmeyi önlerler. Bunlar olmadan, başlangıçtaki eğitim verileri grubu, öğrenmeyi orantısız bir şekilde etkileyebilir ve modelin daha sonraki örneklerde veya gruplarda görünen özellikleri öğrenmesini engelleyebilir.

5. Aktivasyon Fonksiyonu (Activation Function):

Aktivasyon fonksiyonu CNN modelindeki en önemli parametrelerden biridir. Ağ değişken türleri arasındaki karmaşık ve sürekli ilişkilerin öğrenilmesinde ve yaklaştırılmasında kullanılırlar [45]. Basit bir ifadeyle, hangi bilgilerin etkinleştirilmesi ve ağ boyunca ileriye doğru yayılması gerektiğini, hangisinin ağın sonunda olmaması gerektiğini belirlerler.

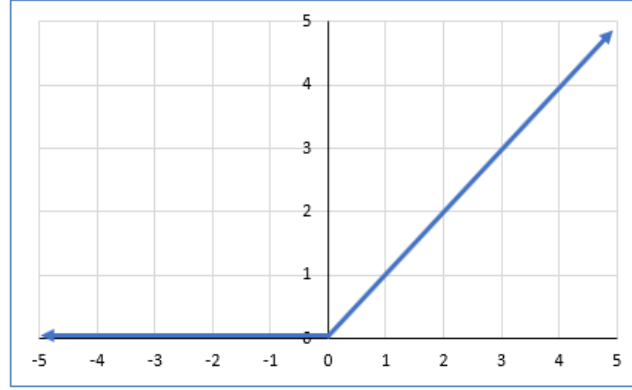
Aktivasyon fonksiyonu, ağa doğrusal olmayan bir yapı kazandırır. ReLU, Softmax, tanH ve Sigmoid fonksiyonları gibi yaygın olarak kullanılan birkaç aktivasyon fonksiyonu vardır. Her bir fonksiyonun belirli bir amacı vardır. İkili sınıflandırma CNN modellerinde sigmoid ve softmax fonksiyonları tercih edilirken, çoklu sınıflandırmada genellikle softmax kullanılır. Bir CNN modelinde, aktivasyon fonksiyonları bir nöronun aktive edilip edilmeyeceğine karar verir. Girdi verisinin iş için önemini tahmin açısından belirlemek için matematiksel işlemler kullanır.

Düzeltilmiş Lineer Birim (Rectified Linear Unit, ReLU):

Düzeltilmiş lineer birim aktivasyon fonksiyonu, yapay sinir ağlarında yaygın olarak kullanılan matematiksel bir fonksiyondur [45]. Bu fonksiyon, bir nöronun çıktısına uygulanarak ağın doğrusal olmayan bir yapıya sahip olmasını sağlar ve girişteki karmaşık desenleri ve ilişkileri anlamasına yardımcı olur. ReLU fonksiyonu, pozitif olan giriş değerini geri döndürür ve negatifse sıfır döndürür. Ayrıca, argümanının pozitif kısmı olarak tanımlanır. Şekil 3.5 ReLU aktivasyon fonksiyonu gösterilmiştir. Matematiksel olarak, ReLU fonksiyonu Denklem (3.7) ifade edilmiştir:

$$f(x) = x^+ = \max(0, x) = \frac{x + |x|}{2} = \begin{cases} x & \text{if } x > 0, \\ 0 & \text{otherwise,} \end{cases} \quad (3.7)$$

Burada x bir nöronun girişidir. $\max(0, x)$, fonksiyonun iki değerden (0 veya x) büyük olanı döndüreceği anlamına gelir. Yani, x pozitifse fonksiyon x değerini döndürür; x negatif veya sıfır ise 0 değerini döndürür.



Şekil 3.5. ReLU aktivasyon fonksiyonu

Softmax fonksiyonu:

Softmax fonksiyonu, sınıflandırma görevleri için bir sinir ağı modelinin son katmanında sıkça kullanılır ve ham çıktı skorlarını — logitler olarak da bilinir — olasılıklara dönüştürür [46]. Bu dönüşüm, her bir çıktının üstelini hesaplayarak ve bu değerleri tüm üstel değerlerin toplamına bölerek gerçekleştirilir. Bu teknik, çıktı değerlerinin 0 ile 1 arasında olmasını ve toplamalarının 1 olmasını sağlar, böylece olasılık olarak yorumlanabilirler. Softmax fonksiyonu, çok sınıflı sınıflandırma için uygulanan daha genel bir lojistik aktivasyon fonksiyonudur. Matematiksel olarak, softmax fonksiyonu Denklem (3.8) ifade edilmiştir:

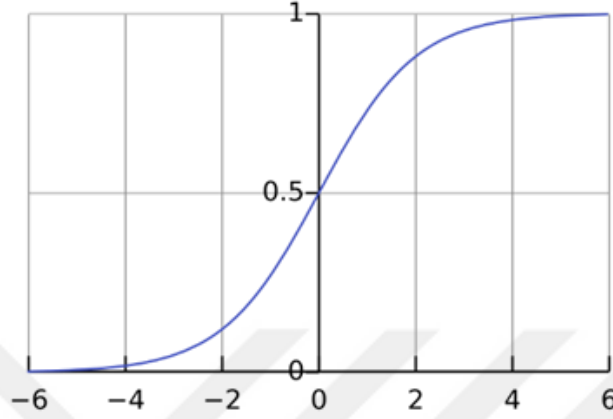
$$\sigma(z)_i = \frac{e^{z_i}}{\sum_{j=1}^K e^{z_j}} \quad (3.8)$$

Burada standart (birim) softmax fonksiyonu $\sigma: R^K \rightarrow (0,1)^K$, neresi $K \geq 1$ olmak üzere, bir vektör alır $z = [z_1, z_2, \dots, z_K] \in R^K$ ve vektörün her bileşenini hesaplar, $\sigma(z) \in (0,1)^K$.

Sigmoid fonksiyonu:

Sigmoid fonksiyonu, grafiği belirgin bir S-şekilli veya sigmoid eğriye sahip olan matematiksel bir fonksiyondur [47]. Sigmoid fonksiyonu, tüm reel giriş değerleri için tanımlı, türevlenebilir, bir fonksiyondur ve her noktada sıfırdan büyük bir türevelere sahiptir. Tam olarak bir tane bükülme noktası vardır Şekil 3.6 Sigmoid fonksiyonu gösterilmiştir. Matematiksel olarak, sigmoid fonksiyonu Denklem (3.9) ifade edilmiştir:

$$\sigma(x) = \frac{1}{1 + e^{-x}} = \frac{e^x}{1 + e^x} = 1 - \sigma(-x) \quad (3.9)$$

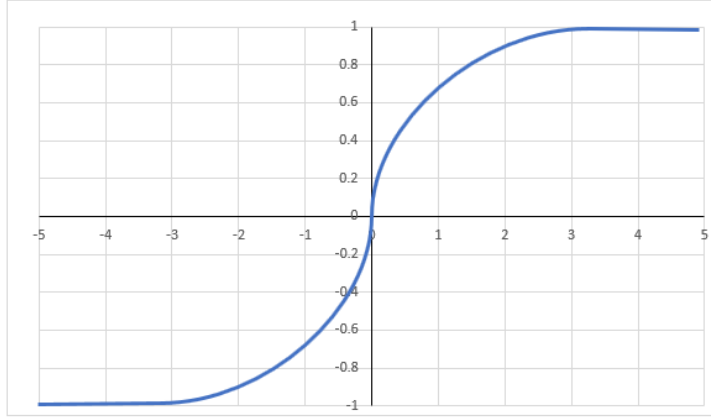


Şekil 3.6. Sigmoid fonksiyonu

Hiperbolik tanjant aktivasyon fonksiyonu (tanH):

Hiperbolik tanjant aktivasyon fonksiyonu, genellikle yapay sinir ağlarının gizli katmanlarında kullanılan matematiksel bir fonksiyondur [48]. Giriş sayılarını -1 ile 1 arasında çıkış değerlerine dönüştürür. Bu fonksiyon, giriş değerinin üssü ile onun negatifinin üssü arasındaki farkın, bu üslerin toplamına bölünmesiyle hesaplanır. tanh fonksiyonu, genellikle iki sınıf arasında sınıflandırma yapmak için kullanılır. Şekil 3.7 tanH aktivasyon fonksiyonu gösterilmiştir. Tanh aktivasyon fonksiyonunun matematiksel Denklem (3.10) ifade edilmiştir

$$\tanh(x) = \frac{(e^x - e^{-x})}{(e^x + e^{-x})} \quad (3.10)$$



Şekil 3.7. tanH aktivasyon fonksiyonu

4. YÖNTEM

Bu bölümde, oluşturulan Türk müziği veri setini ve bu Türk müziğini sınıflandırmak için bu tezde önerilen yöntem açıklanmaktadır. Ayrıca, önerilen yöntemin başarısını kanıtlamak için karşılaştırma yapılan Özbacı'nın çalışması da [10] açıklanmıştır.

4.1. Veri Seti

Türk müziğinin sınıflandırılması için yapılan çalışmaların yetersizliği, sınıflandırma amaçları için yeterli veri setinin bulunmamasına neden olmuştur. Bu nedenle, her tür için bir müzik listesi çıkarmak amacıyla Spotify API'sini kullanarak bir veri seti oluşturulmuştur. Ardından, Türk müziği ses klipleri YouTube'dan indirilmiştir. Veri seti, altı farklı tanınmış müzik türünden oluşmaktadır. Bu türler, Arabesk, Pop, Rap, Rock, Türk Sanat Müziği ve Türk Dini Müziği olarak Tablo 4.1'de sunulmuştur. Türk Müzik Türü (TMG) veri seti, her biri 30 saniye uzunluğunda olan 1167 ses parçasından oluşmakta ve MP3 dosyası olarak saklanmaktadır.

Tablo 4.1. Türk müzik türü

Müzik Türleri	Müzik Sayısı
Arabesk	210
Dini	150
Pop	150
Rap	150
Rock	213
Sanat	294
Toplam	1167

Türk müziği türünde müzik ses parçaları içeren dosyayı oluşturduktan sonra, her ses parçası 3 saniyelik segmentlere bölünmüştür. Ardından, her bir segmenti Mel-Spektrogram görüntüsüne dönüştürülmüştür. Toplam spektral görüntü sayısı 11670 olarak Tablo 4.2'de gösterilmiştir.

Tablo 4.2 Türk müziği türü veri seti

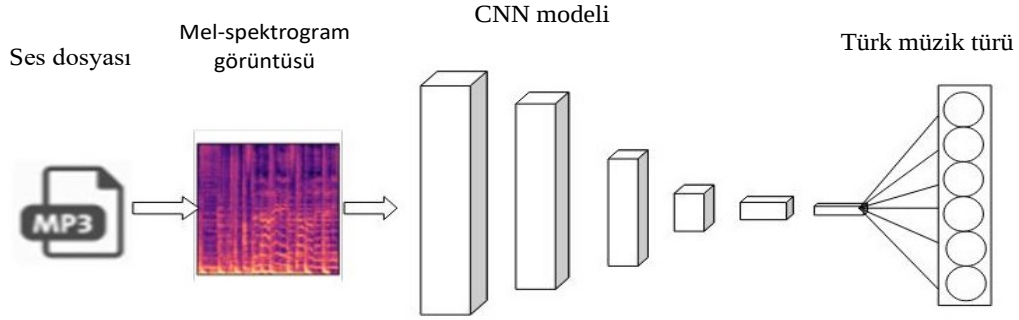
Müzik Türleri	Müzik Sayısı	Görüntü Segmentasyon Sayısı
Arabesk	210	2100
Dini	150	1500
Pop	150	1500
Rap	150	1500
Rock	213	2130
Sanat	294	2940
Toplam Sayısı	1167	11670

Veri seti aşağıdaki klasörleri içermektedir:

- Türk müzik türü orijinal — 6 türden oluşan ve her biri 30 saniye uzunluğunda toplam 1167 ses dosyası içeren bir koleksiyon.
- Spektrogram görüntüleri orijinal — Her ses dosyası için Mel-Spektrogramın görsel temsili. Tüm müzik türleri için toplamda 11670 görüntü içermektedir.

4.2. Yöntem

Önerilen yöntemin iki önemli adım bulunmaktadır. Birincisi, ses sinyalini Mel-spektrogram görüntüsüne dönüştürmek ve ikincisi, Türk müziği türlerini sınıflandırmak için CNN kullanmaktır. Önerilen Türk müziği türü sınıflandırma yöntemi Şekil 4.1’de açıklanmıştır.



Şekil 4.1 Türk müziği türlerinin sınıflandırılması için önerilen yöntem

4.2.1. Mel-spektrogram tespiti

Veri seti bölümünde açıklandığı gibi, altı türden oluşan ve 1167 adet ses dosyasından oluşan bir veri seti oluşturuldu. Her ses dosyası 30 saniye uzunluğunda olup, bunları daha sonra 3 saniyelik 10 parçaya bölündü. Her ses örneği Mel-spektrogramına dönüştürüldü (MEL frekans bantlarıyla y ekseninde). Güç spektrogramını oluşturmak için, ses sinyalinin kısa süreli Fourier dönüşümü (STFT), örnekleme frekansı (sr) = 22050, kareler arasındaki zaman ilerlemesi (atlama boyutu, hop length) = 512 ve çerçeve/pencere boyutu (n -fft) = 2048 parametreleriyle hesaplandı. Bu parametreyi aşağıda açıklanmıştır:

- **Örnekleme Frekansı (Sample Rate, sr)**

Örnekleme frekansı veya örnekleme hızı, sürekli bir sinyalden sayısal veya dijital bir sinyal oluşturmak için her saniyede toplanan örnek sayısını belirler ve hertz (Hz) cinsinden ölçülür. Nyquist-Shannon örnekleme teoremine göre, kusursuz sinyal yeniden oluşturma, örnekleme frekansının sinyalin maksimum frekansının iki katından fazla olması durumunda mümkündür [49, 50]. Örneğin, bir ses akışının üst sınırı 20.000 Hz ise (insan işitme aralığının en üst sınırı), 40.000 Hz'den (40 kHz) daha büyük bir örnekleme frekansı örtüşmeyi önleyecek ve potansiyel olarak hatasız yeniden yapılandırmaya olanak sağlayacaktır.

Önerilen yöntemde 22050 örnekleme hızını kullandık çünkü 22.05 kHz örnekleme hızı, birçok konuşma işleme uygulaması ve ultra yüksek frekansların kritik olmadığı bazı müzik türleri için yeterli olan 11.025 kHz'e kadar olan frekansları yakalayabilir.

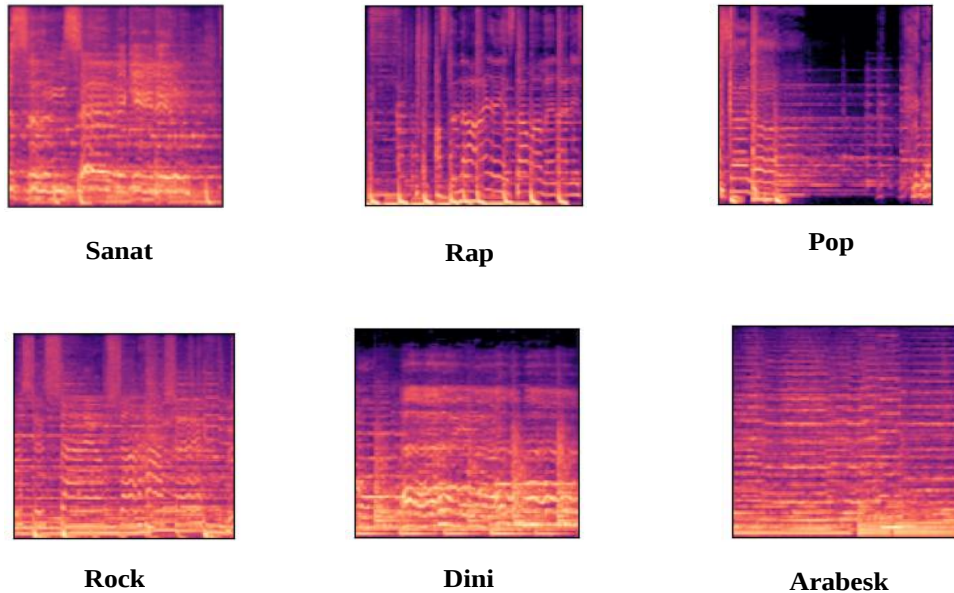
- **Atlama Boyutu (hop size -hop length)**

Atlama boyutu, kısa süreli Fourier dönüşümü (STFT) gibi işlemler sırasında iki ardışık kısa süreli çerçeve arasındaki mesafeyi (örnek cinsinden) belirtir. Bir sinyal için STFT hesaplanırken aynı sinyal iki kez işlenir; hop uzunluğu ne kadar kısa ise, ses sinyalinin belirli bir bölümü STFT’de daha fazla temsil edilir.

- **Çerçeve/Pencere boyutu (Frame/Window size (n-fft))**

Çerçeve boyutu, temel frekans, yoğunluk ve sinyaldeki değişikliklere göre belirlenir. Çerçeve boyutu, Fourier serisinin prensipleri sonucunda belirlenir: analiz penceresinin frekans çözünürlüğünü ayarlamak için kaç frekans bandına bölüneceğini ifade eder. Örneğin, FFT(Hızlı Fourier Dönüşümü) çerçeve boyutunuz (n_fft) 2048 örnektir ve hop uzunluğunuz 512 örnektir. Bu durumda, her 512 örnekte bir yeni bir FFT penceresi başlar. Bu, her biri 2048 örnek kaplayan ancak 512 ile kaydırılmış örtüşen pencereler oluşturur.

Şekil 4.2’da Türk müziğinin her türü için bir görsel gösterilmektedir. Spektrogramlar PNG görüntüleri olarak kaydedilir. Her görselin boyutu 640×480’dir. Bu görüntüler CNN modeline girdi olarak kullanılır.



Şekil 4.2 Türk müziğinin Mel-spektrogramı

4.2.2. CNN modeli

Türk müzik türüne ait ses sinyalinin Mel-spektrograma dönüştürdüğümüzde bu işlemde elde edilen spektrogram görüntülerinin toplam sayısı 11670'dir. Bu görüntüler CNN modeline girdi olarak kullanılmaktadır. Daha sonra bu görüntülerin %70'ini eğitim için, %30'unu doğrulama için kullanıldı. Eğitim 20 epoch ile gerçekleştirildi ve kullanılan parti büyüklüğü (batch_size) 15'e eşittir. Tekrarlanabilirlik açısından rastgele durum sifıra eşittir. Yöntemimizdeki CNN modeli aşağıdaki katmanları içerir:

- Girdi Katmanı:
 - Girdi şekli: 224x224 piksel, üç renk kanalı (standart RGB görüntüsü).
- Birinci Evrişimli Katman:
 - Filtre sayısı: 64
 - Çekirdek boyutu: 3x3
 - Aktivasyon fonksiyonu: ReLU
 - Takip eden Max-Pooling Katmanı:
 - Havuzlama boyutu: 2x2
 - Nöron Seyreltme Katmanı:
 - Aşırı uyumu engellemek için: 0.5 olasılık
- İkinci Evrişimli Katman:
 - Filtre sayısı: 128
 - Çekirdek boyutu: 3x3
 - Aktivasyon fonksiyonu: ReLU
 - Takip eden Maksimum Havuzlama Katmanı:
 - Havuzlama boyutu: 2x2
 - Nöron Seyreltme Katmanı:
 - Aşırı uyumu engellemek için: 0.8 olasılık
- Üçüncü Evrişimli Katman:
 - Filtre sayısı: 128

- Çekirdek boyutu: 3x3
- Aktivasyon fonksiyonu: ReLU
- Takip eden Maksimum Havuzlama Katmanı:
 - Havuzlama boyutu: 2x2
- Nöron Seyreltme Katmanı:
 - Aşırı uyumu engellemek için: 0.4 olasılık
- Tekrarlanan Üçüncü Katman:
 - Aynı özelliklerle bir kez daha uygulanır.
- Düzleştirme Katman:
 - Girdi verilerini düzleştirir.
- Tam Bağlantılı Katmanlar:
 - Birinci Yoğun Katman:
 - Birim sayısı: 1024
 - Aktivasyon fonksiyonu: ReLU
 - İkinci Yoğun Katman:
 - Birim sayısı: 6
 - Aktivasyon fonksiyonu: Softmax

Dikkatlice seçilmiş bir öğrenme oranına sahip (0,0001'e eşdeğer olan $1e-4$ 'e ayarlanmış) Adam optimize edici, eğitim sırasında modelin parametrelerini yinelemeli olarak güncellemek, etkili yakınsama sağlamak ve CNN'nin eğitim verilerinden kalıpları öğrenmesini ve genelleştirmesini sağlamak için kullanılır. CNN'nin eğitimi değerlendirmek için metrik olarak doğruluk ve kayıp fonksiyonu olarak ise categorical_crossentropy kullanıldı.

4.2.3. Yazılım araçları

Bu bölümde kodu yazmak için kullanılan Yazılım araçlarını açıklanmaktadır.

GoogleColab notebook:

Colab, kullanmak için herhangi bir kurulum gerektirmeyen ve GPU'lar ve TPU'lar dahil bilgi işlem kaynaklarına ücretsiz erişim sağlayan, barındırılan bir Jupyter Notebook hizmetidir. Colab özellikle makine öğrenimi, veri bilimi ve eğitime çok uygundur [51].

TensorFlow:

TensorFlow, makine öğrenimi ve yapay zekaya yönelik ücretsiz ve açık kaynaklı bir yazılım kütüphanesidir [52]. Çeşitli görevlerde kullanılabilir ancak derin sinir ağlarının eğitimi ve çıkarımına özel olarak odaklanır. TensorFlow, aralarında Python, JavaScript, C++ ve Java'nın da bulunduğu çok çeşitli programlama dillerinde kullanılabilir ve birçok sektördeki çeşitli uygulamalarda kullanımını kolaylaştırır. Esnek mimarisi, bilgi işlemin çeşitli platformlarda (CPU'lar, GPU'lar, TPU'lar) ve masaüstü bilgisayarlardan sunucu kümelerine, mobil ve uç cihazlara kadar kolayca dağıtılmasına olanak tanır.

Python 3:

Python, birçok uygulamaya sahip, güçlü ve esnek bir genel amaçlı dildir [53]. Python 3, dilin en son sürümüdür ve hem yeni hem de deneyimli geliştiriciler için uygundur. Dünyadaki en popüler programlama dillerinden biridir.

Keras:

Keras, sinir ağlarını uygulamak için Google tarafından geliştirilen üst düzey, derin bir öğrenme API'sidir [54]. Python'da yazılmıştır ve sinir ağlarının uygulanmasını kolaylaştırmak için kullanılır. Aynı zamanda birden fazla arka uç sinir ağı hesaplamasını da destekler.

Librosa:

Librosa, müzik ve ses analizi için bir Python paketidir [55]. Müzik bilgi erişim sistemleri oluşturmak için gerekli yapı taşlarını sağlar.

NumPy:

NumPy, Python programlama diline destek ekleyen bir kütüphanedir [56]. Bu kütüphane, büyük, çok boyutlu diziler ve matrisler için destek sağlar ve bu diziler üzerinde çalışmak için geniş bir yüksek seviye matematiksel işlev koleksiyonu sunar.

Sklearn:

Scikit-learn (Sklearn), Python programlama diline yönelik ücretsiz ve açık kaynaklı bir makine öğrenimi kütüphanesidir [57]. SVM, rastgele ormanlar, gradyan artırma, k-ortalama ve DBSCAN gibi çeşitli sınıflandırma, regresyon ve kümeleme algoritmalarını içerir ve NumPy

ve SciPy gibi Python sayısal ve bilimsel kütüphaneleriyle etkileşimli olarak tasarlanmıştır. Scikit-learn, NumFOCUS tarafından mali olarak desteklenen bir projedir.

4.3. Karşılaştırılan Yöntem

Karşılaştırma yapmak ve önerilen yöntemin başarısını onaylamak için Özbacı'nın [10] çalışmasının yöntemini veri setimizde uyguladık. Bu yöntem, çalışmasının yöntemini bu çalışmada toplanan veri setine uyguladık. Bu yöntem, Türk müziğinin ses dosyasının akustik özelliklerinin çıkartılmasına dayanmaktadır. Bu akustik özellikler chroma_stft, rmse, spectral_centroid, spectral_bandwidth, Spectral Roll-off, Zero_crossing_rate ve mfcc'dir. Daha sonra müzik türünü sınıflandırmak için Lojistik Regresyon (LR), KNN ve SVM gibi çeşitli makine öğrenme algoritmaları uygulanmaktadır.

4.3.1. Akustik özellikler

Bu bölümde Özbacı'nın [10] çalışmasını akustik özellikleri ayrıntılı olarak açıklanacaktır:

Chroma_STFT:

Chroma, bir ses sinyalinin on iki farklı perde sınıfının yoğunluğunu temsil eder ve müzik analizinde kullanılır. Bu özellikler, ses sinyalleri arasındaki perde sınıfı profillerini ayırt etmek için kullanılabilir. Chroma STFT Chroma özelliklerini hesaplamak için kısa süreli Fourier dönüşümünü (STFT) kullanır. STFT'ler, perde sınıflandırması ve sinyal yapısı hakkında bilgi sağlar. Grafikte renk çubuğu ile gösterilen yüksek değerlerle (spike) ve düşük değerlerle (koyu bölgeler) ilgili bilgi sunar.

Ortalama Kare Enerjinin Kökü (Root Mean Square Energy ,RMSE):

RMSE, ses sinyalinin zaman alanı analizinde önemli bir özelliktir [58]. Sayısal sinyal işlemede RMSE, bir sesin küresel enerjisidir [59]. Bu, çerçevedeki tüm genlik değerlerinin karelerinin ortalamasının karekökü alınarak hesaplanır. Bir sinyaldeki enerji Denklem (4.1)'de matematiksel olarak ifade edilmiştir:

$$\sum_{n=1}^N |x(n)|^2 \quad (4.1)$$

Burada ifadesi, $n = 1$ ile $n = N$ arasında $x(n)$ dizisinin karelerinin toplamını temsil eder.bağlam, $x(n)$ sinyalin n -inci örnekteki genliğini temsil eder ve bu değerlerin karesi alınıp toplandığında, N örnek boyunca sinyalin toplam enerjisi elde edilir.

Kök ortalama kare değeri hesaplamak için Denklem (4.2)'de görüldüğü gibi matematiksel olarak hesaplanabilir:

$$\sqrt{\frac{1}{N} \sum_{n=1}^N |x(n)|^2} \quad (4.2)$$

Spectral_centroid:

Bu, en popüler ses özelliğidir. Spektrumun ağırlık merkezinin nerede olduğunu gösterir, yani enerjinin yoğunlukla hangi frekans aralığında toplandığını belirtir. Yüksek frekans içeriği ne kadar fazlaysa, spektral merkez frekansı da o kadar yüksek olur. Spectral_centroid Denklem (4.3)'te matematiksel olarak ifade edilmiştir:

$$Centroid = \frac{\sum_{n=1}^{N-1} f(n)x(n)}{\sum_{n=0}^{N-1} x(n)} \quad (4.3)$$

Burada $x(n)$, n bölme numarasının ağırlıklı frekans değerini veya büyüklüğünü temsil eder ve $f(n)$ bu bölmenin merkez frekansını temsil eder.

Spektral merkez frekansı, bir sesin parlaklık algısıyla önemli bir ilişkiye sahiptir çünkü bir sinyalin enerji yoğunluğunu daha yüksek frekanslara kaydırmak onun daha parlak duyulmasını sağlar [60].

Spectral_bandwidth:

Bant genişliği, sürekli bir frekans bandında üst ve alt frekanslar arasındaki fark olarak tanımlanır. Bildiği gibi sinyaller bir nokta etrafında salınır, dolayısıyla eğer nokta sinyalin ağırlık merkezi ise, noktanın her iki tarafındaki sinyalin en yüksek sapmasının toplamı, sinyalin andaki bant genişliği olarak kabul edilebilir. Spectral_bandwidth Denklem (4.4)'de matematiksel olarak ifade edilmiştir:

$$\left[\sum_k (S(k)f(k) - f_c)^p \right]^{\frac{1}{p}} \quad (4.4)$$

Burada $S(k)$ genellikle k 'inci frekans bölgesindeki bir ağırlığı veya sinyal genliğini temsil eder. $f(k)$, k 'inci kutuya karşılık gelen frekansı temsil eder. f_c referans frekansı, örneğin ağırlık merkezi (centroid) veya merkez frekansı temsil eder. p , hesaplanan momentin veya metrik türünü belirler.

Spectral roll-off:

Bu özellik, spektrumdaki toplam enerjinin %85'inin (kullanıcı bu eşiği seçebilir) altındaki frekans değerini ifade eder [58].

Zero_crossing_rate:

Sıfır geçiş hızı (ZCR), bir sinyalin pozitiften sıfıra, negatife veya negatiften sıfıra ve pozitifte değişme hızıdır [61]. Değeri, vurmali sesleri sınıflandırmak için önemli bir özellik olarak hem konuşma tanımda hem de müzik bilgisi alımında yaygın olarak kullanılmaktadır [62]. ZCR Denklem (4.5)'te matematiksel olarak ifade edilmiştir:

$$zcr = \frac{1}{T-1} \sum_{t=1}^{T-1} 1_{R<0}(s_t s_{t-1}) \quad (4.5)$$

Burada T , sinyaldeki toplam örnek sayısıdır. s bir sinyal uzunluğu. T ve $1_{R<0}$ bir gösterge fonksiyonudur. s_t , t -inci zamandaki sinyalin değeridir. s_{t-1} bir önceki zaman dilimindeki, yani $t-1$ anındaki sinyalin değeridir.

Belirli uygulamalar için ZCR, düşük hesaplama maliyetiyle spektrum bilgisi sağlar. Sinüsoid gibi dar bantlı sinyaller için ZCR doğrudan temel frekansla ilişkilidir. Daha karmaşık sinyaller için ZCR, birincil enerji konsantrasyonunun ortalama frekansı ile iyi bir korelasyon gösterir. Konuşma sinyallerinde kısa süreli ZCR, farklı spektral enerji konsantrasyonlarından dolayı sesli ve sessiz bölümler arasında hızla dalgalanır. Tersine, müzik sinyallerinde ZCR daha uzun süreler boyunca daha istikrarlı kalma eğilimindedir.

MFCC:

MFCC, kısa süreli spektrum zarfının kompakt bir temsili vererek algısal olarak motive edilmiş özelliklerdir. Uzun zamandır ses tanımlamada ve son zamanlarda müzikte kullanılmaktadır [63]. MFCC'yi hesaplamak için, pencereleme yapılmış ses veri çerçevesi önce Ayrık Fourier Dönüşümü (Discrete Fourier transform, DFT) kullanılarak dönüştürülür. Daha sonra, frekans

alanında Mel ölçeği filtre bankası uygulanır ve her alt bant içindeki güç, bantlar içindeki spektral büyüklüklerin karelerinin toplanmasıyla hesaplanır. Mel-frekans ölçeği, kritik bant ölçeğine benzer algısal bir ölçektir ve 1 kHz'in altında doğrusal, bu frekansın üstünde ise logaritmiktir. Son olarak, bant bazındaki güç değerlerinin logaritması alınır ve Ayırık Kosinüs Dönüşüm uygulanarak cepstral katsayılar elde edilir. Logaritmik dönüşüm, spektrumun kaynak ve filtre aktarım fonksiyonu gibi çarpan bileşenlerini ayrıştırmaya yardımcı olur. Bu işlem sayesinde, enerjinin çoğunun birkaç cepstral katsayıda yoğunlaşır. Örneğin, 16 kHz örneklemeli konuşmada, fonemler arasındaki spektral zarfı temsil etmek için 13 düşük dereceli MFCC yeterlidir.

4.3.2 Kullanılan makine öğrenimi algoritmaları

Veri setindeki Türk müziğini Özbacı'nın [10] çalışmalarına dayanarak sınıflandırmak için çeşitli makine öğrenmesi algoritmaları uyguladık. Bu algoritmaları aşağıda açıklanmaktadır:

Lojistik Regresyon (LR):

Lojistik regresyon, genellikle ikili sınıflandırma görevleri için kullanılan bir gözetimli makine öğrenimi algoritmasıdır. Girdi özelliklerinin doğrusal bir kombinasyonunu 0 ile 1 arasında bir olasılık değerine dönüştürmek için lojistik (veya sigmoid) fonksiyonunu kullanır ve bu, belirli bir girdinin iki önceden tanımlanmış kategoriden birine ait olma olasılığını gösterir. Lojistik regresyonun temel mekanizması, lojistik fonksiyonun ikili sonuçların olasılığını doğru bir şekilde modelleme yeteneğine dayanır. Lojistik fonksiyonun karakteristik S şeklindeki eğrisi, herhangi bir reel sayıyı 0 ile 1 arasında bir değere etkili bir şekilde dönüştürür. Lojistik regresyon ayrıca çok sınıflı sınıflandırma için de uygulanabilir. Scikit-learn, çok sınıflı sınıflandırma görevlerini destekleyen verimli bir lojistik regresyon uygulaması sunar.

K-En Yakın Komşu (KNN):

KNN, hem regresyon hem de sınıflandırma görevleri için kullanılan bir gözetimli öğrenme tekniğidir [64]. Bu algoritma, büyük ölçekli bir veri kümesinde bir sorgu örneğine en çok benzeyen veri alt kümesini bulmayı amaçlar. KNN, boyut indirgeme, model sınıflandırma ve görüntü alımı gibi çeşitli uygulamalarda yaygın olarak kullanılır. Örnek tabanlı bir öğrenme algoritmasıdır ve veri kümesindeki örneklerin genellikle benzer özelliklere sahip diğer örneklerin yakınında bulunduğu prensibine dayanır. Yeni bir veri noktasını sınıflandırmak için algoritma, yeni noktaya en yakın k eğitim noktasını belirler. Sınıflandırma, bu komşuların çoğunluk oyuna göre belirlenir; bir veri noktası, en yakın komşuları arasında en sık temsil edilen sınıfa atanır ve bu, bir mesafe fonksiyonu ile ölçülür [65].

Destek Vektör Makinleri (SVM):

SVM, Vapnik'in istatistiksel öğrenme teorisinden türetilen doğrusal ve doğrusal olmayan çizgilerle ayırt edilebilen en iyi makine öğrenimi tekniklerinden biridir [66]. SVM, hem sınıflandırma hem de regresyon görevleri için yaygın olarak kullanılan bir gözetimli makine öğrenimi yöntemidir ve özellikle küçük örnek boyutlarını sınıflandırmada etkilidir [67]. SVM'de, girdi verileri daha yüksek boyutlu bir alana haritalanır ve algoritma iki sınıf arasındaki optimal ayırma hiper düzlemini oluşturur. SVM'nin arkasındaki ilke, iki sınıf arasındaki daha geniş bir marjinin daha başarılı bir sınıflandırmaya yol açmasıdır. İkili sınıflandırmanın yanı sıra, SVM çok sınıflı veri kümelerine de uygulanabilir.



5. SONUÇLAR

Bu bölümde Mel spektrogramlarını ve CNN modelini kullanan önerilen metot ile elde edilen sonuçlar sunulmaktadır. Ayrıca, yaklaşımın sonuçları ile Özbacı'nın [10] çalışmasından elde edilenler arasında karşılaştırmalı bir analiz sunmaktadır.

5.1. Değerlendirme Metrikleri

Algoritmanın karşılaştırmalı niteliklerini anlamak için bunları değerlendirmek önemlidir. Değerlendirme için ana yöntem deneyseldir. Deneysel değerlendirme, algoritmanın pratik performansını gözlemlemek için öğrenme görevlerine uygulanmasını içerir. Bazı modeller değerlendirme sırasında tatmin edici sonuçlar verebilirken, diğerleri kötü performans gösterebilir. Bu, sınıflandırma modelinin performansını değerlendirmek için kullanılan çeşitli metrikleri kapsar, bu çalışmada karmaşıklık matrisi(confusion matrix), kesinlik (precision), duyarlılık (recall) ve F-skoru (F-score) değerlendirme metrikleri olarak kullanılmıştır.

5.1.1. Karmaşıklık matrisi (Confusion matrix)

Karmaşıklık matrisi, hata matrisi olarak da bilinir, bir sınıflandırma algoritmasının sonuçlarını değerlendirmek için kullanılan bir görselleştirme tekniğidir. Bu tablo, her sınıfın gerçek örneklerinin sayısını, tahmin edilen örneklerle karşılaştırır. Karmaşıklık matrisleri, sınıflandırma modelinin performansını ölçmek için kullanılan çeşitli değerlendirme metriklerinden biridir ve hem ikili hem de çok sınıflı sınıflandırma problemlerine uygulanabilir. Karmaşıklık matrisinin amacı, kesinlik (precision), duyarlılık (recall), F-skoru (F-score) ve hata oranını değerlendirerek performansı karakterize etmektir.

Matris, doğru pozitifler (True Positive, TP), doğru negatifler (True Negative, TN), yanlış pozitifler (False Positive, FP) ve yanlış negatifler (False Negative, FN) kavramlarına dayanmaktadır. Bir modelin farklı sınıflardaki performansına ilişkin ayrıntılı bir görünüm sağlar. Şekil 5.1'de örnek bir karmaşıklık matrisi verilmektedir.

Doğru Pozitifler (TP): Modelin pozitif bir sınıfı doğru bir şekilde pozitif olarak tahmin ettiği örneklerdir. Örneğin, bir kanser teşhis modeli düşünüldüğünde, doğru pozitif, modelin kanserli bir hastayı doğru bir şekilde hastalığa sahip olarak tanımladığı durumdur. TP, modelin pozitif örnekleri doğru bir şekilde tanıma yeteneğinin önemli bir ölçüsüdür.

Doğru Negatifler (TN): Modelin negatif bir sınıfı doğru bir şekilde negatif olarak tahmin ettiği örneklerdir. Tıbbi benzetmeye devam edilecek olursa, doğru negatif, modelin sağlıklı bir

hastayı doğru bir şekilde hastalığa sahip olmadığını tanımladığı durumdur. TN, modelin negatif örnekleri tanıma becerisini yansıtır.

	Pozitif Tahmin	Negatif Tahmin
Gerçek Pozitif	Doğru Pozitif (TP)	Yanlış Negatif (FN)
Gerçek Negatif	Yanlış Pozitif (FP)	Doğru Negatif (TN)

Şekil 5.1. Karmaşıklık matrisi (Confusion matrix)

Yanlış Pozitifler (FP): Modelin negatif olması gereken bir durumda yanlış bir şekilde pozitif sınıfı tahmin ettiği örneklerdir. Tıbbi senaryoda, yanlış pozitif, modelin sağlıklı bir hastaya yanlış bir şekilde hastalığı olduğunu belirtmesi anlamına gelir. FP, modelin pozitif sonuçları tahmin ederken aşırı güven sergilediği durumları gösterir.

Yanlış Negatifler (FN): Modelin pozitif olması gereken bir durumda yanlış bir şekilde negatif sınıfı tahmin ettiği örneklerdir. Tıbbi bağlamda, yanlış negatif, modelin aslında hastalığı olan bir hastada hastalığı tespit edememesi durumudur. FN, modelin gerçek pozitif örnekleri yakalamakta başarısız olduğu durumları vurgular.

5.1.2. Duyarlılık (Recall)

Hassasiyet veya gerçek pozitif oran olarak da bilinir. Duyarlılık metriği, bir modelin bir veri kümesindeki tüm uygun vakaları keşfetme yeteneğini gösterir. Bu metrik, doğru pozitif tahminlerin gerçek pozitif örneklerin toplamına oranıdır. Bir veri kümesindeki tüm ilgi çekici veri noktalarını bulmak bir modelin becerisi olarak düşünülebilir. Duyarlılığın matematiksel formülü Denklem (5.1)'de gösterilmiştir:

$$Duyarlılık = \frac{TP}{TP + FN} \quad (5.1)$$

Olumlu örnekleri kaçırmamanın maliyetli olduğu durumlarda bu ölçüm önemlidir.

5.1.3. Kesinlik (Precision)

Kesinlik metriği, pozitif sınıfın doğruluğunu gösterir. Ayrıca, pozitif sınıf tahmininin doğru olma olasılığını değerlendirir. Kesinlik, doğru pozitiflerin sayısının, doğru pozitifler ve yanlış pozitiflerin toplamına bölünmesiyle tanımlanır. Yanlış pozitifler, modelin yanlış bir şekilde pozitif olarak tanımladığı ancak aslında negatif olan vakalardır. Bu metrik, yanlış pozitiflerin maliyetinin yüksek olduğu durumlarda çok önemlidir. Kesinlik, pozitif tahmin değeri (positive predicted value, PPV) olarak da adlandırılabilir. Kesinliğin matematiksel formülü Denklem (5.2)'de gösterilmiştir:

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

5.1.4. F1-skoru (F1-score)

Bazı durumlarda, duyarlılık veya kesinlik ölçütlerinin kullanımını diğer ölçütler pahasına optimize etmek dikkate değer olabilir. F1-skoru, kesinlik ve duyarlılığın harmonik ortalamasıdır; F1-skoru için en iyi değer 1 ve en kötü değer 0'dır. Bu metrik, duyarlılık ve kesinliğin ağırlıklı ortalamasını sağlar. Kesinlik ve duyarlılığı birleştiren birkaç metrik vardır, ancak F1-skoru en yaygın olarak kullanılanıdır. Duyarlılık ve kesinlik arasında ideal dengeye sahip dengeli bir sınıflandırma modeli geliştirirken, F1-skorunun kullanımı mükemmel bir tercihtir. F1-skorunun matematiksel formülü Denklem (5.3)'te gösterilmiştir:

$$F1 = 2 * \frac{Kesinlik * Duyarlılık}{Kesinlik + Duyarlılık} \quad (5.3)$$

5.1.5. Doğruluk (Accuracy)

Doğruluk, modelin tahminlerinin genel doğruluğunu ölçen temel bir metriktir. Bu, doğru pozitifler ve doğru negatiflerin toplamının, toplam örnek sayısına bölünmesiyle hesaplanır. Sınıflar dengeli olduğunda ve yanlış pozitifler ile yanlış negatiflerin maliyetleri karşılaştırılabilir olduğunda uygundur. Doğruluğun matematiksel formülü Denklem (5.4)'de gösterilmiştir:

$$Doğruluk = \frac{TP + TN}{TP + TN + FP + FN} \quad (5.4)$$

5.1.6. Kayıp (Loss)

Kayıp fonksiyonu, maliyet fonksiyonu olarak da bilinir ve bir tahminin gerçek değerden ne kadar uzaklaştığına bağlı olarak olasılığı veya belirsizliği dikkate alır. Bu, modelin ne kadar iyi çalıştığına dair daha ayrıntılı bir anlayış sağlar. Basit bir şekilde ifade edersek, kayıp, eğitim veya doğrulama setlerindeki her örnek için hataların toplamıdır. Kayıp fonksiyonu, genellikle eğitim sürecinde modelin ağırlıkları gibi optimal parametre değerlerini belirlemek için kullanılır. Eğitim sırasında ana hedef, bu değeri minimize etmektir.

5.1.7. Makro ortalama (Macro-averaging)

Her sınıf için performans ölçütü (kesinlik veya duyarlılık gibi), makro ortalama ile belirlenir ve bu daha sonra tüm sınıfların aritmetik ortalamasını hesaplar. Sonuç olarak, makro ortalama her sınıfa, örnek sayısından bağımsız olarak aynı ağırlığı verir. Makro ortalama istatistikleri, çok sınıflı sınıflandırma problemleri için eğitilmiş modelleri değerlendirir. Kesinlik, duyarlılık ve F1-skoru için makro ortalamayı açıklıyoruz.

Kesinlik için makro ortalama, farklı sınıflar için tüm kesinlik değerlerinin aritmetik ortalamasıdır. Kesinlik için makro ortalama Denklem (5.5)'te görüldüğü gibi matematiksel formülü ile hesaplanabilir:

$$\text{Kesinlik makro ortalama} = \frac{(\text{Kesinlik}_1 + \text{Kesinlik}_2 + \dots + \text{Kesinlik}_n)}{n} \quad (5.5)$$

Duyarlılığın makro ortalaması, farklı sınıflara ait tüm duyarlılık puanlarının aritmetik ortalamasıdır. Duyarlılık için makro ortalama Denklem (5.6)'daki matematiksel formül ile hesaplanabilir:

$$\text{Duyarlılık makro ortalama} = \frac{(\text{Duya}_1 + \text{Duya}_2 + \dots + \text{Duya}_n)}{n} \quad (5.6)$$

Makro ortalama F1-skoru, sınıf bazında elde edilen F1-skorlarının basit bir ortalamasıdır. Makro ortalama F1-skoru, yalnızca veri kümesinde her bir sınıfta aynı sayıda veri noktası kullanıldığında yararlıdır. Matematiksel olarak Denklem (5.7) ile ifade edilmiştir:

$$\text{F1 - skoru makro ortalama} = \frac{\sum_{i=1}^n \text{F1skoru}_i}{n} \quad (5.7)$$

5.1.8. Ağırlıklı ortalama (Weighted average)

Ağırlıklı ortalama, göreceli önemlerine göre bireysel veri noktalarına değişen önem düzeyleri atayan istatistiksel bir ölçümdür. Bu yöntem, genel veri kümesinin daha kesin bir temsilini sağlar. Bunu hesaplamak için her veri noktası kendi ağırlığıyla çarpılır, elde edilen ürünler toplanır ve daha sonra bu toplam ağırlıkların toplamına bölünür. Ağırlıklı ortalama, sınıf başına tüm puanların ortalamasını hesaplarken her sınıfın desteğini dikkate alır.

Destek, sağlanan veri kümesindeki sınıf örneklerinin sayısıdır. Eğitim verilerindeki dengesiz destek, sınıflandırıcının rapor edilen puanlarında temel sorunlara işaret edebilir ve tabakalı örnekleme veya yeniden dengeleme gerekliliğini gösterebilir. Destek, modeller arasında değişmez, aksine değerlendirme sürecini teşhis eder.

Kesinlik için Ağırlıklı ortalama Matematiksel olarak Denklem (5.8) ile ifade edilmiştir:

$$\text{Kesinlik Ağırlıklı ortalama} = \frac{Kesi_A \times Destek_A + Kesi_B \times Destek_B + \dots}{Destek_A + Destek_B + \dots} \quad (5.8)$$

Duyarlılığın Ağırlıklı ortalaması Matematiksel olarak Denklem (5.9) ile ifade edilmiştir:

$$\text{Duyarlılık Ağırlıklı ortalama} = \frac{Duya_A \times Destek_A + Duya_B \times Destek_B + \dots}{Destek_A + Destek_B + \dots} \quad (5.9)$$

F1-skoru için Ağırlıklı ortalama Matematiksel olarak Denklem (5.10) ile ifade edilmiştir:

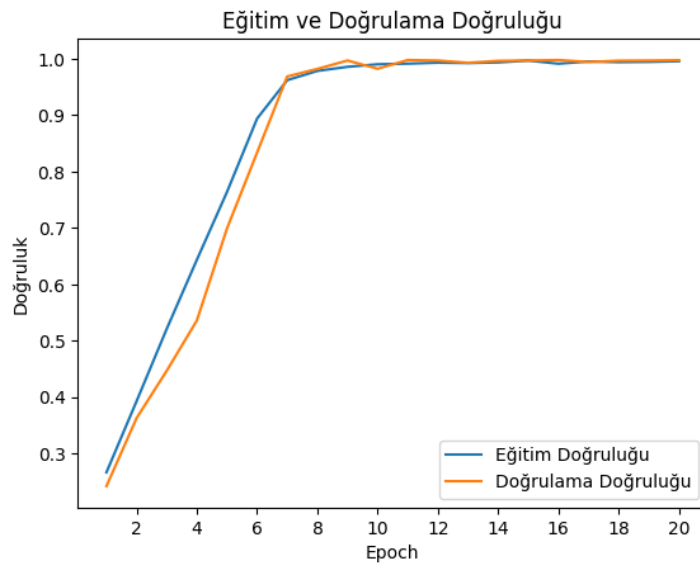
$$\text{F1 Ağırlıklı ortalama} = \frac{F1_A \times Destek_A + F1_B \times Destek_B + \dots}{Destek_A + Destek_B + \dots} \quad (5.10)$$

Burada denklemlerde, $Destek_A$, A sınıfının gerçek örneklerinin sayısını, $Destek_B$, B sınıfının gerçek örneklerinin sayısını belirtir.

5.2. Önerilen Yöntemin Sonuçları

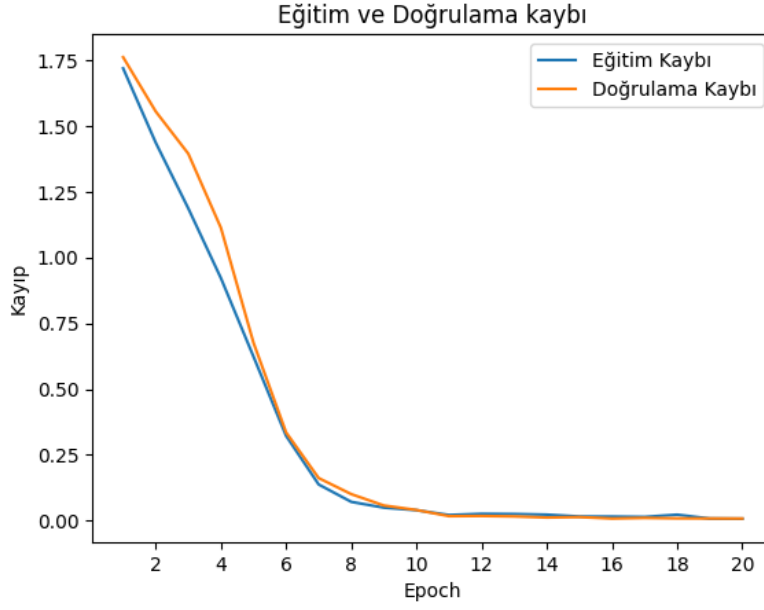
Önceki bölümde açıklandığı gibi, altı tür ve 1167 ses parçasından oluşan bir veri kümesi oluşturuldu. Her ses parçası 30 saniye uzunluğunda olup, her biri 3 saniye uzunluğunda olan 10 segmente ayrılmıştır. Bu ses kayıtları, Python'un Librosa modülü kullanılarak Mel-spektrogram görüntülerine dönüştürülmüştür. Bu spektrogram görüntülerinin toplam sayısı 11670'tir. Daha sonra, bu görüntülerin %70'i eğitim için ve %30'u doğrulama için kullanılmıştır. Eğitim, 20 epoch boyunca gerçekleştirilmiş ve tekrarlanabilirlik için random state sıfır olarak ayarlanmıştır. Deneyler, Keras kütüphanesi ve Python 3 kullanılarak bir GPU sunucusunda (Google Colab notebook) gerçekleştirilmiştir.

Bu bölümde, önerilen yöntemle elde edilen sonuçlar açıklanmaktadır. İlk adımda, deneyin değerlendirme kriteri olarak doğruluk ve kayıp fonksiyonunu kullanılmıştır. Şekil 5.2, modelin farklı epoch'lar boyunca eğitim ve doğrulama doğruluk fonksiyonu eğrisini göstermektedir. Doğruluk eğrileri, modelin eğitim sürecinde eğitim verileri üzerindeki doğru tahminler yapma yeteneğini gösterir. Şekilde, eğitim ve doğrulama eğrilerinin başlangıçta arttığını, ardından 6. epoch'tan sonra plato yapmaya başladığını göstermektedir. Bu, modelin toplam örnek sayısından doğru bir şekilde sınıflandırıldığını kanıtlar. Bu yaklaşımda, eğitim sırasında elde edilen en yüksek doğruluk %99.62 ve doğrulama doğruluğu %99.80 olarak elde edilmiştir. Bu, modelin doğru tahminler yapma yeteneğini kanıtlamaktadır.



Şekil 5.2. Eğitim ve doğrulama için doğruluk fonksiyonu

Kayıp fonksiyonu, modelin tahminlerinin eğitim sırasında gerçek hedef değerlerle ne kadar iyi örtüştüğünü değerlendirir. Kayıp fonksiyonunun minimize edilmiş değeri, modelin tahminlerinin gerçek değerlere daha da yaklaştığını gösterir. Şekil 5.3'de gösterildiği gibi, eğitim ve doğrulama eğrileri başlangıçta epoch'lar boyunca azalmakta ve 6. epoch'tan sonra plato yapmaya başlamaktadır ve eğitim kaybı 0.0123'e, doğrulama kaybı ise 0.0045'e düşmüştür. Kayıp eğrilerinin azalması, modelin hatasının öğrenme süreci boyunca nasıl azaldığını ve bu durumun performansında bir iyileşme olduğunu gösterir.



Şekil 5.3. Eğitim ve doğrulama için kayıp fonksiyonu

Şekil 5.2 ve 5.3'te gösterildiği gibi, eğitim ve doğrulama kaybı ve doğruluk eğrileri birlikte hizalanır ve bu da eğitim modeli sırasında herhangi bir aşırı uyumun meydana gelmediğini gösterir. Bu, müzik türlerini sınıflandırmaya yönelik çoğu yöntemin yaşadığı aşırı uyum sorununun bu yaklaşımla çözüldüğü anlamına gelir.

Ayrıca sınıflandırma modelindeki hataların nerede yapıldığını değerlendirmek için Karmaşıklık matrisi de incelenebilir. Şekil 5.4 sınıflandırma modelin Karmaşıklık matrisini göstermektedir.

Gerçek Etiket	arabesk	627	0	0	0	0	3
	dini	0	450	0	0	0	0
	rap	0	0	450	0	0	0
	rock	0	0	0	639	0	0
	sanat	0	0	0	0	450	0
	pop	6	0	0	0	0	876
Tahmin Etiket		arabesk	dini	rap	rock	sanat	pop

Şekil 5.4. Karmaşıklık matrisi

Satır indisleri Türk müzik türlerinin gerçek sınıf etiketlerine, sütun indisleri ise Türk müzik türlerinin tahmin edilen sınıf etiketlerine karşılık gelir. Gözlemlendiği gibi, sanat, rock, rap ve dini türlerinin Türk müzik veri setindeki sınıf etiketleri sıfıra eşittir. Bu, değerlerin doğru bir şekilde tahmin edildiği anlamına gelir. Arabesk kategorisinde üç müzik parçası pop müzik olarak sınıflandırılmış, ayrıca altı pop müzik parçası da arabesk müzik olarak sınıflandırılmıştır. Bu, bir hata veya yanlış tahmin olduğunu gösterir. Yaklaşımın Karmaşıklık matrisi, Türk müzik türlerini sınıflandırmak için yaklaşımın yeteneğini doğrulamıştır, yalnızca test verilerinden yedi müzik parçasının yanlış tahmin edildiği ortaya çıkmıştır.

5.3. Önerilen Yaklaşımın Sonuçlarının Diğer Yöntemle Karşılaştırılması

Bu bölüm, önerilen yaklaşımın Özbacı'nın [10] çalışmasında kullanılan diğer bir yaklaşım ile karşılaştırma sonuçlarını açıklar. Bölüm 4'te açıklandığı gibi, ilk adımda veri setinden (TMG) akustik özellikler çıkarılmıştır. Bu özellikler chroma_stft, rmse, spectral_centroid, spectral_bandwidth, rolloff, zero_crossing_rate ve mfcc'dir. Daha sonra SVM, KNN ve Lojistik Regresyon olmak üzere üç farklı makine öğrenimi algoritması kullanılmıştır. Veri kümesini eğitim için %70 ve doğrulama için %30 olarak böldülmüştür ve random state değeri sıfır olarak ayarlandı, sınıflandırmada kullanılan değerler önerilen yaklaşımla aynıdır.

Bu iki yaklaşım arasındaki karşılaştırma sonuçlarını değerlendirmek amacıyla ilk aşamada, her bir algoritma için sınıflandırmadaki hataların nerede olduğunu değerlendirmek amacıyla karmaşıklık matrisini uygulanmıştır. SVM, KNN ve Lojistik Regresyon için karmaşıklık matrisi Şekil 5.5'te gösterilmektedir.

Bu algoritmaların karmaşıklık matrisini (Şekil 5.5) önerilen yöntemin karmaşıklık matrisi (Şekil 5.4) ile karşılaştırıldığında, SVM, KNN ve Lojistik Regresyon ile sınıflandırma sürecinde birçok hata veya yanlış tahmin olduğunu görülmektedir. Ayrıca her sınıf için elde edilen kesinlik, hatırlama ve f1-skor değerlerini de sunulmuştur. Ayrıca her algoritmadaki kesinlik, duyarlılık ve F1-skoru için makro ortalama ve ağırlıklı ortalama Tablo 5.1, 5.2, ve 5.3'te gösterilmiştir.

	arabesk	dini	rap	rock	sanat	pop
Gerçek Etiket	arabesk	dini	rap	rock	sanat	pop
	41	1	1	0	2	8
	3	30	0	0	0	5
	3	1	27	7	9	1
	2	1	5	39	4	0
	7	3	6	3	44	1
	14	5	2	0	4	72
	Tahmin Etiket					

(a) SVM karmaşıklık matrisi

	arabesk	dini	rap	rock	sanat	pop
Gerçek Etiket	arabesk	dini	rap	rock	sanat	pop
	44	1	2	0	3	3
	5	26	0	0	1	6
	5	1	28	3	11	0
	7	0	5	32	7	0
	5	0	7	4	46	2
	23	2	1	0	5	66
	Tahmin Etiket					

(b) KNN karmaşıklık matrisi

	arabesk	dini	rap	rock	sanat	pop
Gerçek Etiket	arabesk	dini	rap	rock	sanat	pop
	40	1	1	0	3	8
	3	29	0	1	1	4
	3	1	27	6	10	1
	2	0	7	38	4	0
	3	1	5	5	45	5
	14	5	2	0	3	73
	Tahmin Etiket					

(c) Lojistik Regresyon karmaşıklık matrisi

Şekil 5.5 Makine öğrenimi algoritmaları için karmaşıklık matrisi(SVM, KNN, Lojistik Regresyon)

Önerilen yöntem Özbalcı'nın sunduğu yöntemi [10] önemli ölçüde geride bıraktığını doğrulanmaktadır. Özellikle, önerilen yöntem yaklaşık %99.80'lik önemli bir sınıflandırma doğruluğuna ulaşırken, Özbalcı'nın kullandığı SVM, KNN ve LR makine öğrenimi algoritmaları sırasıyla %72.07, %68.94 ve %71.79 doğruluk oranlarına ulaşmaktadır. Özellikle, SVM algoritması tarafından elde edilen en yüksek doğruluk oranı %72.07 iken, önerilen yaklaşım %99.80 gibi dikkate değer bir doğruluk oranına ulaşmaktadır. Bu büyük doğruluk farkı, önerilen yöntemin Özbalcı'nın yaklaşımına kıyasla etkinliğini ve başarısını vurgulamaktadır.

Tablo 5.1. F1-skoru Özbacı'nın [10] çalışmasındaki sonucun önerilen yaklaşımla karşılaştırılması

TMG	SVM	LR	KNN	Önerilen Yaklaşım
Arabesk	0.67	0.68	0.62	0.99
Dini	0.76	0.77	0.76	1.00
Pop	0.61	0.60	0.62	1.00
Rap	0.78	0.75	0.71	1.00
Rock	0.69	0.69	0.67	1.00
Sanat	0.78	0.78	0.76	1.00
Makro ort.	0.71	0.71	0.69	1.00
Ağırlıklı ort.	0.72	0.72	0.70	1.00

Tablo 5.2. Kesinlik(Precision) Özbacı'nın [10] çalışmasındaki sonucun önerilen yaklaşımla karşılaştırılması

TMG	SVM	LR	KNN	Önerilen Yaklaşım
Arabesk	0.59	0.62	0.49	1.00
Dini	0.73	0.78	0.87	1.00
Pop	0.66	0.64	0.65	1.00
Rap	0.80	0.76	0.82	1.00
Rock	0.70	0.68	0.63	1.00
Sanat	0.83	0.80	0.86	1.00
Makro ort.	0.72	0.71	0.72	1.00
Ağırlıklı ort.	0.73	0.72	0.73	1.00

Tablo 5.3. Duyarlılık (Recall) Özbalcı'nın [10] çalışmasındaki sonucun önerilen yaklaşımla karşılaştırılması

TMG	SVM	LR	KNN	Önerilen Yaklaşım
Arabesk	0.77	0.75	0.83	0.99
Dini	0.79	0.76	0.68	1.00
Pop	0.56	0.56	0.58	1.00
Rap	0.76	0.75	0.63	1.00
Rock	0.69	0.70	0.72	1.00
Sanat	0.74	0.75	0.68	1.00
Makro ort.	0.72	0.71	0.69	1.00
Ağırlıklı ort.	0.72	0.72	0.69	1.00



6. SONUÇ VE GELECEK ÇALIŞMALAR

6.1. Sonuç

Bu tez, Mel-spektrogram görüntüleri ve Evrişimli sinir ağı (CNN) modelleri kullanarak Türkçe şarkıları türlerine göre sınıflandırmayı amaçlamaktadır. Çalışma, ses verilerini Mel-spektrogram görüntülerine dönüştürerek, CNN modellerinin görüntülerden önemli özellikleri tanıma ve çıkarma konusundaki yüksek yeteneğinden yararlanmıştır.

Bu tezin önemli noktaları, ilk olarak, Türk müzik türlerinin sınıflandırılması üzerine yapılan araştırmaların eksikliğini ele almak ve bu konuda katkı sağlamaktır. Türk müziğinin en bilinen türlerini içeren, toplam 1167 ses parçasından oluşan bir Türk müziği veri kümesi oluşturulmuştur. Bu türler Arabesk, Pop, Rap, Rock, Türk Sanat Müziği ve Türk Din Müziği'dir. İkinci sırada ise Mel-spektrogram görüntüleri ile CNN modelinin üstün performansı yer alıyor. Sonuçlar, Mel-spektrogram görüntülerinin CNN modellerine girdi olarak kullanılmasının, Türkçe şarkıların sınıflandırılmasında çok iyi performans sağladığını göstermiştir. Model, geleneksel yöntemlere göre açık bir üstünlük göstererek yaklaşık %99,80'lik bir sınıflandırma doğruluğuna ulaşmayı başarmıştır. Geleneksel yöntemlerle karşılaştırıldığında, elde edilen sonuçlar, Özbalcı [10] tarafından kullanılan geleneksel yöntemle kıyaslandığında, bu yöntemin SVM, KNN, ve LR için sırasıyla %72.07, %68.94 ve %71.79 gibi bir sınıflandırma başarısı gösterdiği tespit edilmiştir. Performanstaki bu büyük boşluk, Mel-spektrogram görüntülerinin CNN modelleri ile kullanılmasının büyük etkinliğini vurgulamaktadır. Ayrıca bu süreç derin öğrenme ve genelleme yeteneğini de kanıtlar niteliktedir. CNN'ler, karmaşık ve çeşitli verilerden öğrenme yeteneklerini ve en ayırt edici özellikleri otomatik olarak çıkarma yeteneklerini gösterir, bu da modelin genelleme yapabilme ve yeni, görülmemiş veri kümelerine uygulama yeteneğini artırır. Bu tezin etkisi sadece yüksek sınıflandırma doğruluğu elde etmekle sınırlı kalmayıp, aynı zamanda konuşma tanıma, müzik analizi, sanat ve sanatta yapay zeka uygulamaları gibi diğer alanlardaki ses modellerinin tanınması gibi çeşitli alanlardaki diğer uygulamalara da yeni ufuklar açmaktadır.

Kısaca bu tez, Mel-spektrogram görüntüleri ile kullanıldığında CNN modellerinin açık üstünlüğünü vurgulayarak, Türkçe şarkı sınıflandırmasında derin öğrenme tekniklerinin kullanımının muazzam potansiyelini ortaya koymaktadır. Elde edilen sonuçlar, bu yaklaşımların etkinliğini güçlü bir şekilde desteklemekte ve bunların birçok alanda gelecekteki uygulamalarına yönelik büyük potansiyeli vurgulamaktadır. Bu gelişmiş modelden

yararlanılarak ses düzeni sınıflandırmasının doğruluğu ve etkinliği önemli ölçüde artırılarak teknolojinin gelişmesine ve günlük yaşamdaki uygulamalarına katkıda bulunulabilir.

6.2. Gelecekteki Çalışmalar

Bu bölüm, özellikle Türk müzik türlerinin sınıflandırılması ve genel olarak müzik türlerinin sınıflandırılması için yapılabilecek gelecekteki çalışmalar için bazı öneriler sunmaktadır. Türk müzik türlerinin sınıflandırılması açısından, veri setini genişletebilir ve daha fazla sayıda Türk müzik parçasını içeren bir veri seti oluşturabilir. Ayrıca, Türk müzik türlerinin sınıflandırılmasını duygusal bağlamla birleştirmek de mümkün olabilir. Örneğin, model yalnızca türü (örneğin, Rap veya Pop) değil, aynı zamanda ilgili duyguyu (örneğin, mutlu, üzgün, romantik) sınıflandıracak şekilde eğitilebilir. Bu yaklaşım, "mutlu Rap" veya "romantik Pop" gibi daha ayrıntılı ve bağlam farkındalığına sahip sınıflandırmalar yapılmasını sağlar.

Başka bir yöntem olarak, modelin uygulanabilirliği kültürler arası analiz için alan adaptasyon teknikleri keşfedilerek geliştirilebilir. Örneğin, CNN modelini Arapça (yazarın ana dili) müzik türlerini sınıflandırmak için uyarlamak, Türk ve Arap müzik gelenekleri arasındaki nüansları ve ortak özellikleri araştıran ilgi çekici bir yol sunar. Bu ilerlemeler, yalnızca Türk müziği sınıflandırmasını geliştirmekle kalmaz, aynı zamanda farklı kültürel müzik türlerinde de daha geniş uygulamalar sunar.

KAYNAKLAR

- [1] N. Choudhury, D. Deka, Satyajit Sarmah, and P. Sarma, "Music Genre Classification Using Convolutional Neural Network," Mar. 2023, doi: <https://doi.org/10.1109/i3cs58314.2023.10127554>.
- [2] M. Müller, *Fundamentals of Music Processing*. 2021. doi: 10.1007/978-3-030-69808-9.
- [3] A. Kumar, A. Rajpal, and D. Rathore, "Genre Classification using Feature Extraction and Deep Learning Techniques," *Genre Classification Using Feature Extraction and Deep Learning Techniques*, Nov. 2018, doi: 10.1109/kse.2018.8573325.
- [4] T. L. H. Li, A. B. Chan, and A. H. W. Chun, "Automatic musical pattern feature extraction using convolutional neural network," *International MultiConference of Engineers and Computer Scientists*, pp. 546–550, Jan. 2010, [Online]. Available: https://www.researchgate.net/profile/Antoni_Chan2/publication/44260643_Automatic_Musical_Pattern_Feature_Extraction_Using_Convolutional_Neural_Network/links/02e7e523dac6bb86b0000000.pdf
- [5] Y. Yu, S. Luo, S. Liu, H. Qiao, Y. Liu, and L. Feng, "Deep attention based music genre classification," *Neurocomputing*, vol. 372, pp. 84–91, Jan. 2020, doi: <https://doi.org/10.1016/j.neucom.2019.09.054>.
- [6] S. Vishnupriya and K. Meenakshi, "Automatic Music Genre Classification using Convolution Neural Network," *International Conference on Computer Communication and Informatics (ICCCI)*, Jan. 2018, doi: <https://doi.org/10.1109/iccci.2018.8441340>.
- [7] U. Bagci and E. Erzin, "Automatic classification of musical genres using Inter-Genre similarity," *IEEE Signal Processing Letters*, vol. 14, no. 8, pp. 521–524, Aug. 2007, doi: 10.1109/lsp.2006.891320.
- [8] G. Tzanetakis and P. Cook, "Musical genre classification of audio signals," *IEEE Transactions on Speech and Audio Processing*, vol. 10, no. 5, pp. 293–302, Jul. 2002, doi: 10.1109/tsa.2002.800560.
- [9] G. Tzanetakis and P. Cook, "Gtzan genre collection," *web resource*. Available: <http://marsyas.info/downloads/datasets.html>, 2001.
- [10] M. C. Özbalcı, *Olasılıksal modeller ile Türkçe müzik türlerinin sınıflandırılması, Master's thesis*, 2022.
- [11] E. Law, K. West, M. Mandel, B. Mert, and J. S. Downie, "Evaluation of algorithms using games: The case of music tagging," *Int. Society for Music Information Retrieval Conf. (ISMIR)*, Kobe, Japan, 2009. [Online]. Available: <https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=8a1384e041cc6ea2735b01c734aeef666dc92884>
- [12] T. Bertin-Mahieux, D. P. W. Ellis, B. Whitman, and P. Lamere, "THE MILLION SONG DATASET," *International Symposium/Conference on Music Information Retrieval*, pp. 591–596, Jan. 2011, doi: 10.7916/d8nz8j07.

- [13] M. Defferrard, K. Benzi, P. Vandergheynst, and X. Bresson, "FMA: a dataset for music Analysis.," *arXiv (Cornell University)*, pp. 316–323, Jan. 2017, [Online]. Available: <https://arxiv.org/abs/1612.01840>
- [14] D. Bogdanov, M. Won, P. Tovstogan, A. Porter, and X. Serra, "The MTG-Jamendo dataset for automatic music tagging," *International Conference on Machine Learning*, Jan. 2019, [Online]. Available: <http://mtg.upf.edu/node/3957>
- [15] M. B. Er and I. B. Aydılek, "Music emotion recognition by using Chroma Spectrogram and deep visual features," *International Journal of Computational Intelligence Systems*, vol. 12, no. 2, p. 1622, Jan. 2019, doi: 10.2991/ijcis.d.191216.001.
- [16] M. K. Karaosmanoğlu, "A Turkish makam music symbolic database for music information retrieval: SymbTr," in *Proceedings of 13th International Society for Music Information Retrieval Conference*, Oct. 8-12, 2012, Porto, Portugal, pp. 223–228, 2012.
- [17] C. Xu, N. C. Maddage, X. Shao, F. Cao, and Q. Tian, "Musical genre classification using support vector machines," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'03)*, vol. 5, pp. V-429, Apr. 2003.
- [18] K. West and S. Cox, "Features and classifiers for the automatic classification of musical audio signals," *International Symposium/Conference on Music Information Retrieval*, Jan. 2004, [Online]. Available: <http://www2.cmp.uea.ac.uk/~sjc/ismir-2004.pdf>
- [19] N. Scaringella, G. Zoia, and D. Mlynek, "Automatic genre classification of music content: a survey," *IEEE Signal Processing Magazine*, vol. 23, no. 2, pp. 133-141, 2006.
- [20] A. Rathore and M. Dorido, "Music genre classification," *CS365: Artificial Intelligence*, 2015.
- [21] J. M. De Sousa, E. T. Pereira, and L. R. Veloso, "A robust music genre classification approach for global and regional music datasets evaluation," in *IEEE International Conference on Digital Signal Processing (DSP)*, Oct. 2016, pp. 109-113.
- [22] P. Fulzele, R. Singh, N. Kaushik, and K. Pandey, "A hybrid model for music genre classification using LSTM and SVM," in *Eleventh International Conference on Contemporary Computing (IC3)*, Aug. 2018, pp. 1-3.
- [23] A. Van den Oord, S. Dieleman, and B. Schrauwen, "Deep content-based music recommendation," in *Advances in Neural Information Processing Systems*, vol. 26, 2013.
- [24] K. Choi, G. Fazekas, and M. B. Sandler, "Automatic tagging using deep convolutional neural networks.," *arXiv (Cornell University)*, pp. 805–811, Aug. 2016, doi: 10.5072/zenodo.242702.
- [25] T. Lidy and A. Schindler, "Parallel Convolutional Neural Networks for Music Genre and Mood Classification." Available: <https://www.ifs.tuwien.ac.at/~schindler/pubs/MIREX2016.pdf>
- [26] L. Wyse, "Audio Spectrogram Representations for Processing with Convolutional Neural Networks," *Audio Spectrogram Representations for Processing With Convolutional*

- Neural Networks*, pp. 37–41, May 2017, [Online]. Available: <http://dblp.uni-trier.de/db/journals/corr/corr1706.html#Wyse17>.
- [27] M. A. Kizrak, K. S. Bayram, and B. Bolat, “Classification of Classic Turkish Music Makams,” *Classification of Classic Turkish Music Makams*, Jun. 2014, doi: 10.1109/inista.2014.6873650.
 - [28] Ö. Çoban and G. T. Özyer, “Music genre Classification from Turkish lyrics,” *Music Genre Classification From Turkish Lyrics*, May 2016, doi: 10.1109/siu.2016.7495686.
 - [29] M. B. Er and E. M. EsiN, “Music Emotion Recognition with Machine Learning Based on Audio Features,” *Computer Science*, Jul. 2021, doi: 10.53070/bbd.945894.
 - [30] A. Spanias, T. Painter, and V. Atti, *Audio signal processing and coding*. 2005. doi: 10.1002/0470041978.
 - [31] B. Prasad and S. R. M. Prasanna, *Speech, Audio, Image and Biomedical Signal Processing using Neural Networks*. 2008. doi: 10.1007/978-3-540-75398-8.
 - [32] A. V. Oppenheim and J. S. Lim, “The importance of phase in signals,” *Proceedings of the IEEE*, vol. 69, no. 5, pp. 529–541, Jan. 1981, doi: 10.1109/proc.1981.12022.
 - [33] E. Sejdić, I. Djurović, and J. Jiang, “Time–frequency feature representation using energy concentration: An overview of recent advances,” *Digital Signal Processing*, vol. 19, no. 1, pp. 153–183, 2009.
 - [34] B. Z. J. L. S. Thornton, “Audio recognition using mel spectrograms and convolution neural networks,” 2019.
 - [35] D. O’Shaughnessy, *Speech communication: human and machine*. 1987. [Online]. Available: <http://ci.nii.ac.jp/ncid/BA03401045>
 - [36] P. P. Shinde and S. Shah, “A Review of Machine Learning and Deep Learning Applications,” *IEEEExplore*, Aug. 01, 2018. <https://ieeexplore.ieee.org/document/8697857>
 - [37] L. Zhang, S. Wang, and B. Liu, “Deep learning for sentiment analysis: A survey,” *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 8, no. 4, Mar. 2018, doi: <https://doi.org/10.1002/widm.1253>
 - [38] D. Cireşan, U. Meier, and J. Schmidhuber, “Multi-column deep neural networks for image classification,” *IEEE Conference on Computer Vision and Pattern Recognition*, Jun. 2012, doi: <https://doi.org/10.1109/cvpr.2012.6248110>
 - [39] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks,” *Neural Information Processing Systems*, vol. 25, pp. 1097–1105, Dec. 2012, [Online]. Available: http://books.nips.cc/papers/files/nips25/NIPS2012_0534.pdf
 - [40] R. Zhu, R. Zhang, and D. Xue, “Lesion detection of endoscopy images based on convolutional neural network features,” *Int. Congr. Image Signal Process. (CISP)*, Chengdu, China, Oct. 2015, pp. 372–376.

- [41] D. H. Hubel and T. N. Wiesel, "Receptive fields, binocular interaction and functional architecture in the cat's visual cortex," *The Journal of Physiology*, vol. 160, no. 1, pp. 106–154, Jan. 1962, doi: 10.1113/jphysiol.1962.sp006837.
- [42] H. C. Ceylan, N. Hardalaç, A. C. Kara, and H. Firat, "Automatic Music Genre Classification and Its Relation with Music Education," *World Journal of Education*, vol. 11, no. 2, pp. 36–45, 2021, Available: <https://eric.ed.gov/?id=EJ1302228>
- [43] T. Guo, J. Dong, H. Li, and Y. Gao, "Simple convolutional neural network on image classification," *IEEE 2nd International Conference on Big Data Analysis (ICBDA)*, Mar. 2017, doi: <https://doi.org/10.1109/icbda.2017.8078730>
- [44] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1929–1958, Jan. 2014, [Online]. Available: <https://jmlr.csail.mit.edu/papers/volume15/srivastava14a/srivastava14a.pdf>
- [45] A. D. Rasamoelina, F. Adjailia, and P. Sincak, "A Review of Activation Function for Artificial Neural Network," *IEEE 18th World Symposium on Applied Machine Intelligence and Informatics (SAMI)*, Jan. 2020, doi:<https://doi.org/10.1109/sami48414.2020.91087>
- [46] C. E. Nwankpa, W. Ijomah, A. Gachagan, and S. Marshall, "Activation functions: comparison of trends in practice and research for deep learning," *arXiv (Cornell University)*, Jan. 2021, [Online]. Available: <https://arxiv.org/pdf/1811.03378.pdf>
- [47] J. Han and C. Moraga, "The influence of the sigmoid function parameters on the speed of backpropagation learning," in *Lecture notes in computer science*, 1995, pp. 195–201. doi: 10.1007/3-540-59497-3_175.
- [48] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015, doi: 10.1038/nature14539.
- [49] M. H. Weik, *Communications Standard Dictionary*. 1996. doi: 10.1007/978-1-4613-0429-6.
- [50] T. J. Moir, *Rudiments of signal processing and systems*. 2022. doi: 10.1007/978-3-030-76947-5.
- [51] J. Praveen Gujjar, R. Prasanna Kumar H, and N. N. Chiplunkar, "Image Classification and Prediction using Transfer Learning in Colab Notebook," *Global Transitions Proceedings*, Aug. 2021, doi: <https://doi.org/10.1016/j.gltp.2021.08.068>
- [52] B. Pang, E. Nijkamp, and Y. N. Wu, "Deep Learning With TensorFlow: A Review," *Journal of Educational and Behavioral Statistics*, vol. 45, no. 2, pp. 227–248, Sep. 2019, doi: <https://doi.org/10.3102/1076998619872761>
- [53] D. K. India, *Programming in Python 3: A complete introduction to the Python language*. 2009. [Online]. Available: <https://www.amazon.com/Programming-Python-Complete-Introduction-Language/dp/8131727017>

- [54] “Deep Learning with Keras,” Google Books, 2017. <https://books.google.com.tr/books?hl=tr&lr=&id=20EwDwAAQBAJ&oi=fnd&pg=PP1&dq=%5B54%5D%09A.+Gulli+and+S.+Pal>
- [55] B. McFee *et al.*, “librosa: Audio and Music Signal Analysis in Python,” *Proceedings of the Python in Science Conferences*, Jan. 2015, doi: 10.25080/majora-7b98e3ed-003.
- [56] S. van der Walt, S. C. Colbert, and G. Varoquaux, “The NumPy Array: A Structure for Efficient Numerical Computation,” *Computing in Science & Engineering*, vol. 13, no. 2, pp. 22–30, Mar. 2011, doi: <https://doi.org/10.1109/mcse.2011.37>.
- [57] P. F., “Scikit-learn: Machine learning in python Fabian,” *CiNii Research*, 2011, [Online]. Available: <https://cir.nii.ac.jp/crid/1370298337102937743>
- [58] H. Bahuleyan, “Music Genre Classification using Machine Learning Techniques,” *arXiv.org*, Apr. 03, 2018. <https://arxiv.org/abs/1804.01149>
- [59] O. Lartillot, “MIRtoolbox 1.7.1 User’ s Manual,” 2018. Accessed: Aug. 12, 2024. [Online]. Available: <https://www.jyu.fi/hytk/fi/laitokset/mutku/en/research/materials/mirtoolbox/manual1-7-1.pdf>
- [60] J. J. Burred and A. Lerch, “Hierarchical Automatic Audio signal classification,” *Journal of the Audio Engineering Society*, vol. 52, no. 7, pp. 724–739, Jul. 2004, [Online]. Available: <https://dialnet.unirioja.es/servlet/articulo?codigo=971067>
- [61] C. H. Chen, *Signal processing handbook*. 1988. [Online]. Available: <http://ci.nii.ac.jp/ncid/BA04249467>
- [62] F. Gouyon, F. Pachet, and O. Delerue, “On the use of zero-crossing rate for an application of classification of percussive sounds, ” *Proceedings of the COST G-6 Conference on Digital Audio Effects (DAFX-00)*, Verona, Italy, Dec. 2000, vol. 5, p. 16.
- [63] B. Logan, “Mel frequency cepstral coefficients for music modeling,” *International Symposium/Conference on Music Information Retrieval*, Jan. 2000, [Online]. Available: http://ismir2000.ismir.net/papers/logan_abs.pdf
- [64] “A Double Weighted KNN Algorithm and Its Application in the Music Genre Classification,” *IEEE Conference Publication | IEEE Xplore*, Jan. 01, 2020. <https://ieeexplore.ieee.org/abstract/document/9045778>
- [65] H. Jégou, M. Douze, and C. Schmid, “Product quantization for nearest neighbor search,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 1, pp. 117–128, Jan. 2011, doi: 10.1109/tpami.2010.57.
- [66] M. Y. Kao, *Encyclopedia of Algorithms*. 2008. doi: 10.1007/978-0-387-30162-4.
- [67] A. Widodo and B.-S. Yang, “Support vector machine in machine condition monitoring and fault diagnosis,” *Mechanical Systems and Signal Processing*, vol. 21, no. 6, pp. 2560–2574, Aug. 2007, doi: 10.1016/j.ymssp.2006.12.007.



ÖZGEÇMİŞ

İlk ve orta öğrenimini Musul/Irak'ta tamamladı. 2013 yılında Musul/Irak'taki Northern Teknik Üniversitesi Bilgisayar Teknolojisi Mühendisliği Bölümü'nden mezun oldu. 2021 yılından bu yana Yalova Üniversitesi Yüksek Lisans Enstitüsü Bilgisayar Mühendisliği Bölümü öğrencisi.

TEZDEN TÜRETİLEN YAYIN VE ESERLER

Hazim, S. B., & Okkalioglu, M. (2024, May). Turkish Music Genre Classification Using Convolutional Neural Network. In Proceedings of the Cognitive Models and Artificial Intelligence Conference (pp. 219-223).

