

Derin Öğrenme İle Türkçe Müziklerden Müzik Türü Sınıflandırması

Serhat Hızlısoy^{1*}, Zekeriya Tüfekci²

^{1*} Kayseri Üniversitesi, Mühendislik Mimarlık ve Tasarım Fakültesi, Bilgisayar Mühendisliği Bölümü, Kayseri, Türkiye (ORCID: 0000-0001-8440-5539), serhathizlisoy@kayseri.edu.tr

² Çukurova Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği Bölümü, Adana, Türkiye (ORCID: 0000-0001-7835-2741), ztufekci@cu.edu.tr

(2nd International Conference on Access to Recent Advances in Engineering and Digitalization (ARACONF)-10-12 March 2021)

(DOI: 10.31590/ejosat.898588)

ATIF/REFERENCE: Hızlısoy, S., & Tüfekci, Z. (2021). Derin Öğrenme ile Türkçe Müziklerden Müzik Türü Sınıflandırması. *Avrupa Bilim ve Teknoloji Dergisi*, (24), 176-183.

Öz

Bu çalışmada, müzik türü sınıflandırma yapmak için bu alanda daha önce kullanılmamış derin öğrenmeye dayalı Evrişimli Uzun Kısa süreli bellek derin sinir ağı (CLDNN) adı verilen bir mimari kullanılmıştır. Ayrıca çeşitli müzik türlerine ait 200 müzikten oluşan yeni bir Türkçe Müzik Veritabanı oluşturulmuştur. Önerilen mimarinin ve yaygın olarak kullanılan makine öğrenme metodlarının sınıflandırma performansı oluşturulan bu veri tabanı üzerinde değerlendirilmiştir. Ek olarak, kullanılan bu mimarinin ilk kısmını oluşturan Evrişimli Sinir Ağı (CNN) kullanılarak, yeni öznitelikler elde edilmiştir. Bu yeni öznitelikleri elde etmek için Evrişimli Sinir Ağı'na girdi olarak hem Mel Frekansı Kepstrum Katsayıları (MFCC) hem de log mel filtre bankası enerjileri kullanıldı. Bu özniteliklere ek olarak çeşitli araçlar kullanılarak çok sayıda standart öznitelik elde edilmiştir. Bütün metodlar için en başarılı sınıflandırma sonuçlarına standart özniteliklerle yeni öznitelikler bir arada kullanıldığında ulaşılmıştır. Karşılaştırılan sınıflandırıcılar içerisinde en iyi sonuçlara ise %99,5 ile önerilen mimarinin kalan kısmı olan Uzun Kısa Süreli Bellek (LSTM) ile tam bağlantılı katmanlardan oluşan Derin Sinir Ağı (DNN) birleşimi ile ulaşılmıştır.

Anahtar Kelimeler: Müzik, Tür Sınıflandırma, Derin Öğrenme.

Classification Of Music Genres From Turkish Music With Deep Learning

Abstract

In this study, an architecture called Convolutional Long Short-term memory deep neural network (CLDNN) based on deep learning, which has not been used before in this field, is used for music genre classification. In addition, a new Turkish Music Database consisting of 200 music belonging to various music genres has been created. The classification performance of the proposed architecture and commonly used machine learning methods has been evaluated on this database. In addition, new features are obtained by using Convolutional Neural Network (CNN), which is the first part of this architecture. Both Mel Frequency Cepstrum Coefficients (MFCC) and log mel filterbank energies were used as input to the Convolutional Neural Network to obtain these new features. In addition to these features, many standard features have been obtained by using various toolboxes. The most successful classification results for all methods are achieved when standard features are used together with new features. The best results among the compared classifiers were achieved with 99.5% by using the remaining part of the proposed architecture, Long Short Term Memory (LSTM), together with the Deep Neural Network (DNN) consisting of fully connected layers.

Keywords: Music, Genre Classification, Deep Learning.

* Sorumlu Yazar: Kayseri Üniversitesi, Mühendislik Mimarlık ve Tasarım Fakültesi, Bilgisayar Mühendisliği Bölümü, Kayseri, Türkiye, ORCID: 0000-0001-8440-5539, serhathizlisoy@kayseri.edu.tr

1. Giriş

Müzik, kendine özgü dili, yapısı ve anlatımıyla insanların karşılıklı düşüncesi, duygusu ve özlem alışverişinde bulunabilmelerini sağlayan önemli bir iletişim aracıdır. Teknolojinin gelişmesiyle ve yeni iletişim araçlarının hayatımıza girmesiyle müziğe ulaşabilme imkanımızı her geçen gün artmaktadır ve müzik yaşamımızın hemen hemen her anında bizimle beraberdir. Bu durum da müziği en iyi şekilde kullanarak iletişimimizi en üst seviyeye taşıyacak bir bilgi ve beceriye ulaşmayı zorunu hale getirmiştir.

Günümüzde, yapay zekâ teknolojilerinin ve verinin işlenebilirliğinin artmasıyla birlikte müzik endüstrisi dünya ekonomisi içinde önemli paya sahip olmaya başlamıştır. Dünyaca ünlü birçok şirket online müzik dinleme, müzik öneri sistemleri (Spotify) dahil olmak üzere birçok uygulamaya rağbetin artması sonucu müzik endüstrisine yönelmiştir. Bu uygulamalar kullanıcılar tarafından etiketlenen müzikler ile dinlediğimiz müzikleri, içinde barındırdığı sistemler sayesinde analiz ederek, türünü sınıflandırır.

Analiz aşamasında, son zamanlarda yapay zekâ teknolojilerinin sinyal işleme ile birlikte kullanılmasıyla birlikte müzik ile ilgili akademik alanda da birçok çalışma yapıldığı görülmektedir. Bu çalışmaların çoğu; müziğin akustik analizi, makam tanıma, müzik türü sınıflandırma ve müzikten duyu tanma gibi çalışmaları içermektedir (Abidin ve ark., 2017; Çoban, 2017; Er ve Çiğ, 2020).

Müzik türü sınıflandırma, müzik bilgisine erişim (MIR)'in bir alt dalıdır. Müzik Bilgisine Erişim, müzik parçalarından bilgi çıkarma mekanizmaları ile bu görevlerde yer alan disiplinler arası bir araştırma alanıdır. Müzik türü sınıflandırma ile ilgili yapılan literatür taramasında çalışmaların çoğunda batı müziklerinin tercih edildiği (Feng, 2014; Karatana ve Yıldız, 2017; Wong ve ark., 2018) ve Türk müziklerine çok fazla yer verilmemiş gözlendi. Dolayısıyla bu çalışmada Türk müziğinden seçilen örneklerden oluşan yeni bir veri tabanı oluşturulmak istenmiştir.

Tür sınıflandırmasına dayalı çalışmaların çoğunda temelde iki ana unsur vardır. İlk, sınıflandırma başarısını artıracak büyük verilerden onları temsil edecek öznitelikler belirlemek, ikincisi ise, türleri ayırt eden bu öznitelikleri kullanarak iyi bir sınıflandırıcı seçmektir.

Araştırmacılar bu çalışmalarında genellikle melodi, harmoni, ton ve ritim gibi müzik ve türü arasında ilişki kurabilecek standart özellikler kullanırlar. Bu öznitelikleri de tek tek elde etmek yerine yüzlerce hatta binlerce standart öznitelik elde etmeye sağlayan Marsyas (Tzanetakis ve Cook, 2000), Jaudio (McKay, 2005), Opensmile (Eyben ve Schuller, 2015) ve Mirtoolbox (Lartillot ve Toivainen, 2007) gibi çeşitli araçlar kullanırlar. Bu çalışmada da özniteliklerin çıkarılmasında bu araçlardan bazıları kullanılmıştır.

Müzik türlerinin farklılıklarının tespit edilmesini sağlayan en son ve en önemli aşama sınıflandırmadır. Bu alanda yapılan çalışmalara bakıldığından K En Yakın Komşu (k-NN) (Cover ve Hart, 1967), Rastgele Ormanlar (RF) (Breiman, 2001), Destek Vektör Makineleri (SVM) (Platt, 1999) ve Bayes Ağları (Friedman ve ark., 1997) gibi en sık kullanılan makine öğrenme metodları ile birlikte Evrişimli Sinir Ağları (CNN) (Lecun, 2015) ve Uzun Kısa Süreli Bellek Ağları (LSTM) (Hochreiter, 1997) gibi birçok derin öğrenme metodunun kullanıldığı görülmektedir.

Bu çalışmada konuşturma tanıma (Sainath ve ark., 2015), duyu tanımı (Hizlisoy ve ark, 2020) da dahil olmak üzere birçok alanda başarılı olan ancak müzik türü sınıflandırmada daha önce denenmemiş yeni bir yaklaşım olan evrişimli uzun-kısa süreli bellek derin sinir ağları (CLDNN) mimarisini, oluşturulan veritabanı üzerinde kullanılarak sınıflandırma yapılmıştır. Ayrıca bu mimarinin bir parçası olan CNN' e girdi olarak Mel Frekansı Kepstrum Katsayıları (MFCC) ve Log-mel filtre bankası enerjileri kullanılarak 2 farklı öznitelik seti daha elde edilmiş ve bu öznitelikler sınıflandırma aşamasında kullanılmışlardır.

2. Materyal ve Metot

2.1. Veri Tabanı

Bilindiği üzere veritabanı oluşturmak, yoğun emek gerektiren ve zaman alan bir süreçtir. Ayrıca, müziklerin telif hakları ve değerlendirme maliyetleri gibi bazı zorluklarla karşılaşıldığı için de kaliteli veri seti elde etmek ve yayılmak oldukça zordur. Çok sayıda zorlukmasına rağmen son zamanlarda farklı dillerde birçok veritabanı oluşturulmuştur (Sturm, 2012; Bertin-Mahieux ve ark., 2011) ancak kaliteli veri tabanlarının eksikliği ve var olanlara ulaşılamaması, araştırmacıları genellikle kendi veri tabanlarını oluşturmaya zorlamıştır. Bu çalışmalar arasında ulaşılabilir Türk müziği veri tabanlarının sayısı neredeyse yok denecek kadar azdır. Dolayısıyla bizde kendi veri tabanımızı oluşturmaya karar verdik.

Farklı toplumlara ait birçok müzik; ritim, melodi ve kültürel farklılıklara rağmen, dünyanın her bölgesinde sevilerek dinlenir. Fakat her toplum genellikle kendi kültürüne, yoresine ait, kendi dilini ve melodik ezgilerini taşıyan müziğe ayrı bir yer verir. Bu nedenle çalışmada kullanılmak için dört farklı türü içeren müziklerden oluşturulmuş Türkçe Müzik Veri Tabanı oluşturulmuştur.

Bu veritabanı her müzik türünden 50 parça olmak üzere 200 parçanın biraraya getirilmesiyle meydana getirilmiştir. Müzik türleri olarak ise insanlar tarafından için en çok tercih edilen Klasik Türk Müziği, Türk Halk Müziği, Pop ve Arabesk seçilmiştir. Veritabanındaki şarkıların her biri 30 saniye uzunluğundadır. Müziklerin türlerine göre dağılımı Tablo 1'de verilmiştir.

Tablo 1. Çalışmada Kullanılan Müzik Türleri ve Şarkı Sayıları

Tür	Türkçe Veri Tabanı
Klasik Türk Müziği	50
Türk Halk Müziği	50
Pop Müzik	50
Arabesk Müzik	50
<i>Toplam</i>	200

2.2. CLDNN Mimarisi

Bu bölüm, Türk müziklerinin türünü sınıflandırmak için kullanılan CLDNN Mimarısına dayalı olarak yaklaşımı açıklamaktadır. CLDNN (Convolutional Long Short Term Memory Deep Neural Networks), ismini içerisinde barındırdığı mimarilerin ilk harflerinin birleşiminden almaktadır. Bu mimari bir sınıflandırıcı gibi görünmesine rağmen 2 temel bölümden oluşur. İlk bölümde, özellik çıkarma için CNN kullanılır. İkinci bölümde ise sınıflandırıcı olarak LSTM ile DNN birarada kullanılır.



Şekil 1. CLDNN Mimarisi

CNN'ler girişlerin özelliklerini çıkarır, konum bilgisini çıkarmada üstündür, LSTM'ler zamansal modellemede iyidir ve DNN'ler özellikleri ayrılabilir bir alana haritalamak için uygundur.

LSTM'in, daha yüksek kaliteli öznitelikler sağlandığında daha iyi performans gösterdiği keşfedildiği için CNN ve DNN arasına eklenir. DNN modülü, çıkışlar ve gizli birimler arasında bir eşleştirme sağlar.

Şekil 1 bu mimarıyı anlatmaktadır. Bu çalışmada kullanılan LSTM katmanı 200 gizli birimden oluşur. LSTM'nin çıkışı 2 tam bağlı (FC) DNN katmanına bağlanır. Bu katmanlar, öznitelikleri hedefleri bulmanın daha kolay olduğu, daha ayırt edici bir alana dönüştürür. Kullanılan her DNN katmanında ise 100 gizli birim vardır.

2.2.1. Öznitelik Çıkarımı için Bir Boyutlu CNN

Derin öğrenmede, sınıfları temsil eden kaliteli öznitelikleri elde etmek için genelde büyük miktarda veri gereklidir. Bu şekilde elde edilen özniteliklerle, tatmin edici sınıflandırma performansları elde edilir ve eğitim verileri ile test verileri arasındaki hata farkı azaltılır.

Bizim veri setimizde olduğu gibi; veri boyutunun sınırlı olduğu durumlarda, bu sorunun üstesinden gelmek için veri büyütme işlemi gereklidir. Müziğe gürültü eklemek veya her parçayı birkaç eşit parçaaya bölmek gibi eldeki veriyi büyütmek için çeşitli teknikler vardır. Bu çalışmada, veri setindeki her örneğin 10 saniyelik 3 eşit parçaaya bölünmesine karar verilmiştir. Böylece, özellik çıkarımı için veri setimizin boyutu tam olarak üç katına çıkarılmıştır.

CNN'ler, yapay sinir ağlarının özel bir türüdür ve en popüler derin öğrenme mimarisidir. Başlangıçta görüntü işleme için geliştirilmiştir. Ancak daha sonra doğal dil işleme ve ses işleme (konuşma tanıma, konuşmadan duygusal tanıma, müzikten duygusal tanıma, müzik türü sınıflandırma) (Liu ve ark., 2017; Thiruvengatanadhan, 2020) gibi birçok farklı alanda da uygulanmaya başlanmıştır.

CNN'ler, bir veya daha fazla evrişimli katmandan oluşur. Her evrişim katmanı, yeni öznitelik haritaları oluşturmak için bir dizifiltreye sahiptir. Evrişim katmanlarından sonra, havuz katmanları eklenir ve genellikle matrisin her alt bölümü için ortalama veya maksimum değerleri alarak parametre sayısını azaltmak için kullanılır. Önerilen mimari Tablo 2'de gösterildiği gibi, dört evrişim katmanı ile birlikte, dört ortaklama katmanı ve özniteliklerin tek bir boyuta indirgenmesi için düzleştirme (flatten) katmanından oluşmaktadır.

CNN'ler farklı giriş boyutlarına uyum sağlayabilirler. CNN'lerin 1, 2 veya 3 boyutları olmasına rağmen temelde aynı şekilde çalışırlar. Farkları ise resimde görüldüğü gibi, giriş verilerinin boyutu ve özellik algılayıcı olarak da adlandırılan filtrelerin veriler boyunca nasıl hareket edeceğini göstermektedir. 1D CNN, sabit uzunluktaki bir süredeki (ses sinyalleri gibi) her türlü sinyal verisinin analizi için kullanılabilir. Bazı çalışmalarında 1D CNN'lerin diğerlerinden daha avantajlı olduğu ve bu nedenle ses sinyalleri ile uğraşırken 2D CNN'lere tercih edildiği görülmüştür.

Mesela, 1D CNN'lerin hesaplama karmaşıklığı, 2D CNN'lere kıyasla çok düşüktür. Çünkü 1D CNN'ler az sayıda gizli katmana ve nörona sahiptir, bu nedenle uygulamak ve eğitmek çok daha kolaydır. Standart bilgisayarlar genellikle 1D CNN'de eğitmek ve uygulamak için yeterlidir, ekstra donanım (çok çekirdekli işlemci, ekran kartı) gerekmeyebilir. Dolayısıyla bu çalışmada, müzik türü sınıflandırma özelliklerinin çıkarılması için 1D CNN'ler kullanılmıştır.

1D-CNN kullanılarak yapılan bazı çalışmalarda doğrudan ham sinyallerden öznitelik çıkarmayı amaçlayan modeller kullanılmıştır. Fakat, CNN tabanlı ham sinyalden elde edilen özniteliklerin sınıflandırma performansı üzerinde çok fazla etkisi olmadığı tespit edilmiştir (Sarkar ve ark., 2020). Bu nedenle, CNN tabanlı öznitelikler elde etmek için bu çalışmada CNN'ye girdi olarak doğrudan ham sinyal kullanmak yerine Log-mel filtre bankası enerjileri ve MFCC'ler kullanıldı. Log-mel filtre

Tablo 2. CNN için evrişimli katmanların (E), ortaklama katmanlarının (O) ve düzleştirme katmanının (F), farklı tip girdiler ve farklı girdi boyutları dikkate alınarak konfigürasyonu.

Tanımlama		Log-mel Filtre Bankası Enerjileri				Mel Frekansı Kepstrum Katsayıları			
		500 x 26				500 x 13			
		Boyut	Filtre	Filtre Boyutu	Atlama (Stride)	Boyut	Filtre	Filtre Boyutu	Atlama (Stride)
Katmanlar	E1	491	64	10	1	491	64	10	1
	O1	122	64	4	4	122	64	4	4
	E2	120	128	3	1	120	128	3	1
	O2	30	128	4	4	30	128	4	4
	E3	28	128	3	1	28	128	3	1
	O3	7	128	4	4	7	128	4	4
	E4	6	128	2	1	6	128	2	1
	O4	2	128	3	3	2	128	3	3
	D	1	256			1	256		

bankası enerjileri ve MFCC'lerin tercih edilmesinin nedeni ise, konuşma tanıma ve duygusal tanıma çalışmaları içerisinde en ilişkili bilgileri taşıdıkları düşünülmeli ve en yaygın kullanılan öznitelikler olmalarıdır.

CNN'ler bağımsız öznitelikleri, bir görüntünün mavi, kırmızı ve yeşil bileşenleri gibi ayrı kanallar olarak paralel şekilde kullanabilir. Dolayısıyla bu öznitelikler, bir görüntünün kanalları gibi birbirlerinden bağımsız olmaları nedeniyle 1D CNN ile kullanılmaya da uygundur. 1D-CNN için gerekli girdinin yapısı, her örnek için zaman adımlarının sayısına ve bu zamanlarda elde edilen öznitelik sayısı sayısına ihtiyaç duyar. Her müzik parçası daha önce belirtildiği gibi 10 saniyelik üç parçaya ayrıldı ve ardından her 20 ms'de 30 ms'lık Hamming penceresi kullanılarak 13 MFCC ve 26 log-mel filtre bankası enerjisi elde edildi. Her seferde pencere 20 ms kaydırdıktan, 10 saniye içinde zaman adımlarının sayısına 500 olacaktır. Sonuç olarak, her 10 saniyelik müzik için CNN'e girdi olarak kullanabileceğimiz 500 * 13 boyutunda MFCC ve 500 * 26 boyutunda log-mel filtre bankası enerjisi oldu. Bu öznitelikler, her 10 saniyelik müzik için 256 CNN tabanlı özellik elde etmek için CNN'e beslenir. Tüm bunların sonucunda her 30 saniyelik müzik için 768 adet CNN tabanlı MFCC özniteligi ve 768 CNN tabanlı log-mel filtre bankası enerjisi özniteligi elde edilir.

Log-mel filtre bankası enerjilerinin ve MFCC'ların hesaplanması için adımlar gereklidir. İlk adımda, müzikler hamming penceresi kullanılarak üst üste binen çerçevelere bölünür. İkinci adımda, her çerçeve için ayrık zamanlı Fourier dönüşümü (DTFT) hesaplanır. Daha sonra spektrumun büyüğünün karesi hesaplanır. Dördüncü adım, her bir mel-ölçeklenmiş üçgen filtre öibeğindeki tüm enerjilerin toplanmasıyla hesaplanan filtre bankası enerjilerini verir. Beşinci adımda çıktıları, mel filtre bankası enerjilerinin logaritması alınarak hesaplanan log-mel filtre bankası enerjileridir. Son adımda ise, MFCC'ler log-mel filtre bankası enerjilerinin ayrık kosinus dönüşümü (DCT) alınarak hesaplanır.

2.2.2. Standart Öznitelikler

Günümüzde, anlamlı öznitelikler elde etmenin karmaşıklığı nedeniyle standart öznitelikleri çıkarmak için genellikle belirli araçlar kullanılır. Bu araçlar, sağlam analiz ve senteze dayalı önemli öznitelikleri çıkarmak için hazırlanmış etkili ve genişletilebilir yazılımlardır. Bu çalışmada, bu standart öznitelikleri müzik alıntılarından çıkarmak için üç araç kullanıldı.

Opensmile aracı, ilk olarak konuşma işleme için tasarlanmış olsada, duygusal tanıma gibi farklı alanlarda da iyi performans göstermiştir. Bu öznitelik seti, 6553 özellik içerir. Özellik sayısının yanı sıra, Emo-large özellik seti, bugüne kadar bilinen en büyük özellik kümesidir. Mirtoolbox, müzik verilerinden bilgi çıkarma alanındaki hesaplama yaklaşımlarına genel bir bakış sağlamak için hazırlanmış bir yazılımdır. Bu araç kutusunun önemli avantajları, MATLAB ile entegrasyonu ve çok sayıda müzikal özelliği çıkarma yeteneğidir. Bu araç kullanılarak 348 özellik elde edilmiştir.

Jaudio hem genişletilebilirdir hem de kullanımı kolaydır. Çok sayıda özelliğin çıkarılmasına olanak sağlar. Standart sapma, türev ve ortalama dahil olmak üzere yeni özellikler oluşturmak için herhangi bir özelliğe uygulanabilecek meta özellikleri kullanır. Bu araç kutusu ile 468 özellik elde edilmiştir. Bu 3 araç kutusundan elde edilen özellikler daha sonra birleştirilerek bir özellik havuzu oluşturulmuştur. Bu araç kutuları kullanılarak 200 müzikten oluşan veritabanı içindeki her bir müzik parçası için toplam 7369 standart özellik çıkarılmıştır.

2.3. Öznitelik Seçimi

Veri setinden elde edilen özelliklerin kalitesi, sınıflandırmadaki başarı oranını ölçüde etkiler. Öznitelik seçme işlemi bittikten elde edilen binlerce öznitelikten en faydalı öznitelikleri seçerek gereksiz özniteliklerden kurtulma işlemine öznitelik seçimi denir. Böylelikle o veritabanını temsil edecek en iyi öznitelik alt kümesi belirlenerek veri boyutu azaltılır. Aynı zamanda bu işlem sayesinde eğitim sırasında modele çok fazla öznitelik verilmeyen ve modelin çalışmasını

kolaylaşır. Dolayısıyla eğitim süresi ve iş yükü azalır. Özellikle seçim işlemi yapan yöntemler, tahmin başarısını en yükseğe çıkarma amacıyla sahiptir. Bu çalışmada, öznitelik seçim yöntemi olarak Korelasyona dayalı Özellik Seçimi algoritması (Correlation-based Feature Selection, CFS) (Hall ve Smith, 1998) uygulanmıştır. Bu yöntem, birbiriyile bağlantılı olmayan ancak sınıfla ilişkisi yüksek olan özniteliklerin alt kümelerini, bulmayı amaçlayan bir algoritmadır.

2.4. Sınıflandırıcılar

Öznitelik çıkarma işlemi sonucunda elde edilen özellikler kullanılarak sınıfı belli olmayan verilerin, yüksek doğruluk oranları ile sınıflara ayrılması işlemine sınıflandırma denir. Sınıflandırmanın amacı, hangi sınıf'a ait oldukları bilinen örneklerin özelliklerini kullanarak bir model oluşturmak ve bu modeli, özellikleri bilinmeyen yeni örneklerin sınıflandırılmasında kullanmaktır.

Bu çalışmada sınıflandırma yapabilmek için içerisinde çok sayıda sınıflandırma ve kümeleme algoritmaları bulunduran WEKA (Hall ve ark., 2009) aracı kullanılmıştır. Yapılan müzik türü sınıflandırma çalışmasında kullanmak için oldukça popüler olan Naive Bayes, K- en yakın komşu, rastgele orman seçilmiştir ve bu sınıflandırıcılar Weka'da varsayılan parametreleri ile kullanılmıştır.

2.4.1. Naïve Bayes

Olasılık modeli Bayes teoremine dayanan Naïve Bayes sınıflandırıcısı, yüksek boyutlu verilere uygulanabilir ve basit, kolay ve iyi performans gösteren bir algoritmadır. Temel mantık, sınıfı tahmin etmek için gözlemlenen bir grup örnek için en yüksek olasılık çıktısını hesaplamaktır. Bu sınıflandırıcı, olasılık ilkelerine göre tanımlanan bir dizi hesaplama ile test için sisteme gönderilen veri sınıfını belirlemeyi hedefler.

Naïve Bayes yöntemi, bir sınıfın her özelliğinin eşit ve diğer özelliklerden bağımsız olduğunu varsayar. Yani, tüm özellikler bağımsız olarak sonuç olasılığını etkiler ve herhangi bir özelliğin varlığı veya yokluğu diğer özelliklerini etkilemez. Avantajları, veri sınıflandırımda hızlı ve kolay uygulanabilmesi ve çok sınıflı tahminlerde iyi performans göstermesidir.

2.4.2. K En Yakın Komşu

K en yakın komşu (k-NN), hem sınıflandırma hem de regresyon problemleri için kullanıldığı için en popüler makine öğrenme algoritmalarından biridir, basittir. Kolay çıktı yorumlama ve tahmin gücü ile genellikle verimli performans sağlar ve bazı durumlarda doğruluğu modern sınıflandırıcılarından daha yüksektir.

K-NN algoritmasının üç temel öğesi vardır: sınıfı bilinen örnekler, örnekler arasındaki mesafeyi hesaplamak için uzaklık ölçüsü ve en yakın komşuların sayısı için k değeri. Hesaplama maliyetinin yüksek olması ve diğer yöntemlere göre yavaş çalışması dezavantajı olarak görülebilir.

Bu yöntemde, sınıflandırılacak örnekler ile mevcut sınıflandırılmış örnek arasındaki mesafelerin hesaplanarak sınıflandırma işlemi yapılır. (Euclidean, Minkowski ve Manhattan) Sınıflandırılacakorneğe en yakın k komşuları belirlenir ve sıralanır, daha sonra bu k komşularının sınıf etiketlerine bakılarak sınıflandırma yapılır.

2.4.3. Rastgele Ormanlar

Rastgele Ormanlar, en çok kullanılan makine öğrenme algoritmalarından biridir. Çünkü kullanımı kolay, esnek ve başarılı bir yöntemdir. Avantajı hem sınıflandırma hem de regresyon problemleri için kullanılabilmesi ve eğitim süresinin diğer sınıflandırıcılara göre daha az olmasıdır.

Bu algoritma, düğümde kullanılacak değişken sayısını ve geliştirilecek ağaç sayısını olmak üzere sırasıyla 2 farklı parametreye dayanmaktadır. RF, eğitim aşamasında çoklu karar ağaçları oluşturarak çalışan bir yöntemdir. Bu yönteme göre rasgele birden fazla karar ağacı oluşturulur. Sınıflandırma sonucu veya karar, bir yaprak düğümünde taşınır.

Birbirinden farklı kurulan karar ağaçları, bizi sonuca götürecek karar ormanı topluluğunu oluşturur. Karar ormanın oluşturulması sırasında elde edilen sonuçlar birleştirilerek nihai tahmin yapılır. Ağaçların coğunluğunun kararı, rastgele orman tarafından nihai karar olarak seçilir.

2.4.4. LSTM + DNN

Yinelenen Sinir Ağrı (RNN)'nın, doğal dil işleme, konuşma tanıma, görüntü sınıflandırma gibi farklı alanlarda yapılan birçok çalışmada başarılı olduğu, fakat öğrenme yeteneğinin bazı durumlarda sınırlı olduğu görülmüştür. Bunun nedeni olarak da RNN'nin sadece en son bilgilere bakması ve bağamlar arası boşluk arttığında geçmişten gelen bilgiyi kullanmadan zorlanması gösterilmiştir. LSTM katman yapıları işte bu yaşanan problemleri aşmak amacıyla geliştirilmişlerdir.

LSTM, birçok görev için standart RNN'den çok daha iyi çalışan bir RNN türüdür. Yapısı gereği, hafıza mekanizması ile uzun süreli bağımlılıkları öğrenebildiğinden, konuşma tanıma, doğal dil işleme, müzik üretimi ve duyu tanıma gibi zamansal modelleme çalışmalarında yaygın olarak kullanılmaktadır. Uzun süre bilgiyi hatırlama davranışını sergilerler. Çünkü LSTM katman yapısında katmanlarda bulunan nöronların kendi hafızaları bulunmaktadır. Bu hafızalarda geçmiş zaman verileri depolanır ve model geliştirilirken bu verilerden de yararlamlar.

Hafızalarda hangi verilerin depolanacağı eğitim sürecinde belirlenir. Böylelikle girdilerin veya girdilerin dizilimlerinin bile yeni girdilerle etkileşime girebileceği ve çıktı üzerinde etkilerinin olacağı bir ağ yapısı oluşturulmuş olur. Bununla birlikte, RNN'den farklı olarak, temel LSTM mimarisinde, bir veya daha fazla bellek hücresına sahip tek bir kapı yerine, giriş kapısı, çıkış kapısı ve unutma kapısı gibi 3 geçit vardır.

CNN ve LSTM'e ek olarak, DNN de ses sinyallerine uygulanabilir. DNN'de, bir katmandaki her nöron, önceki ve sonraki katmanlardaki tüm nöronlara bağlıdır. Genelde sınıflandırma problemlerinde gizli birimler ile çıktı arasındaki eşleşmeyi daha da derinleştirerek tahminlerini iyileştirmek için son adımda kullanılır. Sınıflandırmada bir başka sinir ağları ile beraberde kullanılabilir. Ek olarak, başka bir mimarının bir parçasını da oluşturabilir.

CNN, LSTM ve DNN modülleri modelleme yeteneklerinde tamamlayıcı olduğundan ve sadece LSTM modellerinin kullanıldığı durumlardan daha iyi performans gösterdiği için CLDNN daha önce birçok alanda geniş çapta kabul görmüştür. Bundan dolayı, müzik türü sınıflandırma için CNN'ler, LSTM'ler ve DNN'ler birleşik bir çerçevede birleştirilerek kullanıldı.

Uzun kısa süreli bellek (LSTM) katmanından ve ardından iki tam bağlı sinir katmanından (FC) oluşan bir LDNN (LSTM + DNN) modeli, önerilen modellerin her biri için ortak alt ağ mimarisi işlevi görür. Bu nedenle, her modeli sınıflandırmak için LDNN kullanılır ve ardından sistemi tamamlamak için bir çıktı softmax katmanı kullanılır. Böylece CLDNN mimarisi tamamlanmış olur.

3. Araştırma Sonuçları ve Tartışma

3.1. Gerçekleştirme

Bu çalışma için üç ayrı öznitelik seti oluşturulmuştur. İlk öznitelik setinde daha önce çeşitli çalışmalarında kullanılan melodi, harmoni, ton, ritim gibi müzik ve türü arasında ilişki kurabilecek standart özellikler kullanılmıştır. Standart özniteliklere ek olarak, öznitelik çıkarmak için log-mel filtre bankası enerjileri ve MFCC'ler CNN'e girdi olarak beslenmesiyle 768'er adet öznitelikten oluşan iki ayrı öznitelik seti daha elde edilmiştir. CNN yapısı içerisinde her biri sırasıyla 64, 128, 128 ve 128 filtreden oluşan dört evrişimli katman kullanılmıştır. Temel modüller; her evrişim katmanın ardından Doğrultulmuş Lineer Birim (Rectified Linear Unit) uygulanmıştır. Ortaklama katmanlarında (pooling), boyutu azaltmak için 4 ve 3 maksimum ortaklama katmanı (max-pooling) kullanılmıştır. Ayrıca aşırı öğrenmeyi azaltmak için 0,05 bırakma oranı (dropout rate) seçilmiştir. 0,0001 öğrenme oranına (learning rate) sahip Uyaranabilir Moment Tahmin Optimize Edici (ADAM) (Kingma ve Ba, 2015) en iyi özelliklerini bulmak için seçilmiştir. CNN'yi eğitmek için 10 numunelik (batch size) bir boyut seçilerek ve 100 tur (epoch) kadar

eğitilmiştir. Ayrıca, sinir ağlarını optimize etmek için kategorik çapraz entropi (categorical cross entropy) bir kayıp fonksiyon (loss function) olarak tercih edilmiştir. Daha sonra bütün öznitelikler hem ayrı ayrı hem de birarada kullanılıp karşılaştırılarak 4 türde göre müzik türü sınıflandırması yapılmak istenmiştir.

Bütün deneysel çalışmalar WEKA programı ile yapılmıştır. Çapraz doğrulama, öğrenme algoritmasının sonuçlarını değerlendirmek ve karşılaştırmak için kullanılan istatistiksel bir yöntemdir. Weka, bu işlemi yapmamıza olanak sağlar. N katlı çapraz doğrulamada, veri kümesi rastgele n eşit parçaya bölünür. Bundan sonra veriler, tüm eğitim setleri için yaklaşık olarak aynı sınıf dağılımını elde edecek şekilde katmanlara ayrılır. Daha sonra n-1 kısımları eğitim verisi olarak kullanılır, geri kalan bölüm ise bir hata oranını hesaplamak için test için kullanılır. Bu işlem, her bir alt küme hem eğitim verisi hem de test verisi olarak kullanılanın kadar n kez tekrarlanır. Modelin hata oranı, ayrı ayrı hesaplanan n hata oranının aritmetik ortalaması alınarak bulunur. Çalışmamızda 10 kat çapraz doğrulama kullanılmıştır.

		Tahmin Edilen	
		Doğu Pozitif (DP)	Yanlış Pozitif (YP)
Gerçek	Yanlış Negatif (YN)	Yanlış Negatif (YN)	Doğu Negatif (DN)

Şekil 2. Hata Matrisi

Oluşturulan bu modellerin başarısını değerlendirmek için kullanılan en popüler yöntem modele ait doğruluk oranıdır. Doğru sınıflandırılmış verilerin (DN+DP) toplam veri sayısına (DN+DP+YP+YN) bölünmesi ile bulunur (1).

$$\text{Doğruluk} = \frac{DP + DN}{DP + YN + YP + DN} \quad (1)$$

3.2. Araştırma Sonuçları ve Tartışma

Bu çalışmada farklı yollar izlenerek elde edilen 3 farklı öznitelik seti, 4 farklı sınıflandırıcıda test edilerek sonuçlar alınmıştır ve bu sonuçlar Tablo 3, Tablo 4 ve Tablo 5 de gösterilmiştir. Bu 3 tablodaki sonuçlara bakıldığında en başarılı sonuçların öznitelik seçim işlemi öncesinde MFCC öznitelikleri ile öznitelik seçim sonrasında ise standart öznitelikle elde edildiği görülmektedir. Kullanılan öznitelik sayısı göz önüne

Tablo 3. Log mel Filtre Bankası Enerjileri Kullanılarak Elde Edilen Özniteliklerle Sınıflandırma Sonuçları

Sınıflandırıcı	Log mel Filtre Bankası Enerjileri	
	Öznitelik Seçimi Öncesi Doğruluk	Öznitelik Seçimi Sonrası Doğruluk
Naive Bayes	79,5	86
K En Yakın Komşu	75	84
Rastgele Ormanlar	84,5	89
LSTM + DNN	85	89,5

alındığında 10 kat daha az öznitelik içermesine rağmen MFCC öznitelikleri kullanılarak elde edilen sonuçların standart öznitelikler kullanılarak elde edilen sonuçlara hemen hemen yakın sonuçlar göstermesi yeni yöntemler kullanılarak oluşturulan öznitelik setlerinin oldukça başarılı olduğunu göstermektedir. Log mel filtre bankası enerjileri kullanarak elde edilen başarı diğer iki öznitelik setine oranla daha düşük olmasına rağmen yaklaşık %90 sınıflandırma başarısına ulaşmıştır. Yine bu tablolara bakıldığında, hepsinde sınıflandırma metotları arasında en iyi doğruluğu veren sonuçların hem öznitelik seçimi öncesinde hem de öznitelik seçimi sonrasında LSTM + DNN ile elde edildiği görülmektedir.

Tablo 6 da ise, bütün öznitelik setleri biraraya getirildiğinde elde edilen sonuçları göstermektedir. Bu sonuçlara bakıldığında standart özniteliklere yeni öznitelikler eklendiğinde öznitelik seçim öncesinde 3,5 ile 6,5 puan arasında, öznitelik seçim sonrasında ise 0,5 ile 4 puan arasında iyileşmeler olduğu görülmektedir.

Sınıflandırıcı başarısı açısından bakıldığında ise LSTM + DNN mimarisini kullanılarak yapılan sınıflandırma başarının %99,5 ulaşarak bu çalışmada elde edildiği en yüksek müzik türü sınıflandırma başarısı olduğu gözlenmektedir. LSTM + DNN'ı hemen hemen bütün tablolarda sırasıyla Rastgele Ormanlar, Naive Bayes ve K En Yakın Komşu metotlarının takip ettiği görülmektedir. Ayrıca yine bütün tablolarda öznitelik seçim işleminin sınıflandırmadaki başarı oranını iyileştirmede ne kadar önemli olduğu ortaya koyulmuştur.

4. Sonuç

Bu çalışmada, müzik türü sınıflandırma için derin öğrenmeye dayalı yöntemler kullanarak hem yeni öznitelikler elde etmek hemde sınıflandırıcı olarak kullanarak performansı artırmak amaçlandı. Ham sinyalden elde edilen MFCC ve log-mel filtre bankası enerjileri CNN'e girdi olarak beslenerek her biri için 768'er öznitelik elde edildi. Bu öznitelikler müzik türü tanıma için kullanılan LSTM + DNN sınıflandırıcısı ile CFS öznitelik seçim işlemi sonrasında MFCC ile %96, log mel filtre bankası enerjiler ile %89,5 sınıflandırma doğruluğu elde edildi.

Tablo 4. Mel Frekansi Kepstrum Katsayıları Kullanılarak Elde Edilen Özniteliklerle Sınıflandırma Sonuçları

Sınıflandırıcı	Mel Frekansi Kepstrum Katsayıları	
	Öznitelik Seçimi Öncesi Doğruluk	Öznitelik Seçimi Sonrası Doğruluk
<i>Naive Bayes</i>	84	91,5
<i>K En Yakın Komşu</i>	82,5	90
<i>Rastgele Ormanlar</i>	86,5	93,5
<i>LSTM + DNN</i>	90	96

Tablo 5. Standart Özniteliklerle Sınıflandırma Sonuçları

Sınıflandırıcı	Standart Öznitelikler	
	Öznitelik Seçimi Öncesi Doğruluk	Öznitelik Seçimi Sonrası Doğruluk
<i>Naive Bayes</i>	82,5	93,5
<i>K En Yakın Komşu</i>	80	91,5
<i>Rastgele Ormanlar</i>	86	96,5
<i>LSTM + DNN</i>	87	97

Tablo 6. Büütün Öznitelikler Birlikte Kullanıldığından Elde Edilen Sınıflandırma Sonuçları

Sınıflandırıcı	Özniteliklerin Tamamı ile	
	Öznitelik Seçimi Öncesi Doğruluk	Öznitelik Seçimi Sonrası Doğruluk
<i>Naive Bayes</i>	86,5	97,5
<i>K En Yakın Komşu</i>	86,5	95,5
<i>Rastgele Ormanlar</i>	90	97
<i>LSTM + DNN</i>	90,5	99,5

Yine ham sinyalden çeşitli araçlar kullanılarak standart öznitelikler elde edilerek kullanılan sınıflandırıcı ile %97 sınıflandırma başarısı elde edilerek diğer bütün popüler methodlardan daha iyi sonuçlar alındı. Son olarak Bütün öznitelikler birleştirildi ve ardından öznitelik seçim işlemi uygulanarak önerilen mimari ile %99,5 başarı ile müzik türleri sınıflandırıldı. CFS öznitelik seçim işlemi uygulandıktan sonra toplam öznitelik sayısı 8905'den 201'e düşmüştür. Bu özniteliklerin 35'i MFCC kullanılarak elde edilen özniteliklerden, 21'i log mel filtre bankası enerjileri kullanılarak elde edilen özniteliklerden, 144' i ise standart özniteliklerden oluşmaktadır. Yine bu sonuçlardan da yeni özniteliklerin sınıflandırma başarısının iyileşmesinde katkısı olduğu gözlenmektedir.

Daha önce müzik türü sınıflandırma da kullanılmamış bir mimari, klasik makine öğrenmesi yöntemleri ile karşılaştırılarak; farklı türdeki müziklerden oluşan yeni bir veri tabanı uygulanmıştır. Sonuç olarak deneysel sonuçlar oluşturulan veri tabanının kalitesini, yeni özelliklerin performansı artırıcı etkisini ve önerilen sınıflandırıcının diğer sınıflandırma yöntemlerine göre daha iyi sonuçlar verdiği göstermiştir.

Kaynakça

- Abidin, D., Öztörk, Ö., & Öztörk, T.Ö., (2017). Using data mining for makam recognition in Turkish traditional art music, *J. Fac. Eng. Archit. Gazi Univ.* 32 1221–1232. <https://doi.org/10.17341/gazimmd.369557>.
- Bertin-Mahieux, T., Ellis, D. P. W., Whitman, B., & Lamere, P., (2011). The million song dataset, in: Proc. 12th Int. Soc. Music Inf. Retr. Conf. ISMIR 2011.
- Breiman, L., (2001). Random forests, *Machine. Learning*.
- Cover, T. M., & Hart, P. E., (1967). Nearest Neighbor Pattern Classification, *IEEE Trans. Inf. Theory*.
- Çoban, Ö., (2017). Turkish Music Genre Classification using Audio and Lyrics Features, *Şüleyman Demirel Üniversitesi Fen Bilim. Enstitüsü Derg.* 21 322. <https://doi.org/10.19113/sdufbed.88303>.
- Er, M. B., & Çığ, H., (2020). Türk Müziği Uyarlanları Kullanılarak İnsan Duygularının Makine Öğrenmesi Yöntemi İle Tanınması, *Gazi Üniversitesi Fen Bilim. Derg. Part C Tasarım ve Teknol.* 8 458–474. <https://doi.org/10.29109/gujsc.687199>.
- Eyben, F., & Schuller, B., (2015). OpenSMILE – The Munich Versatile and Fast Open-Source Audio Feature Extractor Florian, *ACM SIGMultimedia Rec.*
- Feng, T., (2014). Deep learning for music genre classification, *Tech. Rep. Univ. Illinois*.
- Friedman, N., Geiger, D., & Goldszmidt, M., (1997). Bayesian Network Classifiers, *Mach. Learn.* <https://doi.org/10.1023/a:1007465528199>.
- Hall, M., & Smith, L., (1998). Feature subset selection: a correlation based filter approach, in: Proc. Int. Conf. Neural Inf. Process. Intell. Inf. Syst.
- Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H., (2009). The WEKA data mining software: An update, *ACM SIGKDD Explor. Newsl.*
- Hizlisoy, S., Yıldırım, S., & Tufekci, Z., (2020). Music emotion recognition using convolutional long short term memory deep neural networks, *Eng. Sci. Technol. an Int. J.* <https://doi.org/10.1016/j.jestch.2020.10.009>.
- Hochreiter, S., & Schmidhuber, J., (1997). Long Short-Term Memory, *Neural Comput.* <https://doi.org/10.1162/neco.1997.9.8.1735>.
- McKay, C., (2005). JAudio: Towards a standardized extensible audio music feature extraction system, *Course Pap. McGill Univ. Canada*.
- Lartillot, O., & Toiviainen, P., (2007). Mir in matlab (II): A toolbox for musical feature extraction from audio, in: *Proc. 8th Int. Conf. Music Inf. Retrieval, ISMIR*.
- LeCun, Y., Hinton, G., & Bengio, Y., (2015). Deep learning *Nature.* 28;521(7553):436-44. doi: 10.1038/nature14539. PMID: 26017442.
- Liu, X., Chen, Q., Wu, X., Liu, Y., & Liu, Y., (2017). CNN based music emotion classification, *ArXiv*.
- Karatana, A., & Yıldız, O., (2017). Music Genre Classification using Machine Learning Techniques, *Signal Process. Commun. Appl. Conf. (SIU), 2017 25th.* (2017) 1–4. <https://doi.org/10.1109/siu.2017.7960694>.
- Kingma, D. P., & Ba, J. L., (2015). Adam: A method for stochastic optimization, in: *3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc.*
- Platt, J., (1999). Fast Training of Support Vector Machines using Sequential Minimal Optimization, in: *Adv. Kernel Methods --- Support Vector Learn.*
- Sainath, T. N., Weiss, R. J., Senior, A., Wilson, K.W., & Vinyals, O., (2015). Learning the speech front-end with raw waveform CLDNNS, in: *Proc. Annu. Conf. Int. Speech Commun. Assoc. INTERSPEECH*.
- Sarkar, R., Choudhury, S., Dutta, S., Roy, A., & Saha, S. K., (2020). Recognition of emotion in music based on deep convolutional neural network, *Multimed. Tools Appl.* <https://doi.org/10.1007/s11042-019-08192-x>.
- Sturm, B. L., (2012). An analysis of the GTZAN music genre dataset, in: *MIRUM 2012 - Proc. 2nd Int. ACM Work. Co-Located with ACM Multimed.* <https://doi.org/10.1145/2390848.2390851>.
- Tzanetakis, G., & Cook, P., (2000). MARSyas: A framework for audio analysis, *Organised Sound.* <https://doi.org/10.1017/S1355771800003071>.
- Thiruvengatanadhan, R., (2020). Musical Genre Classification using Convolutional Neural Networks, *Int. Journal of Innovative Technology and Exploring Engineering,* <https://doi.org/10.35940/ijitee.a8172.1110120>.
- Wong, K.H., Tang, C. P., Chui, K. L., Yu, Y. K., & Zeng, Z., (2018). Music genre classification using a hierarchical long short term memory (LSTM) model, 7. <https://doi.org/10.1117/12.2501763>.