

MFCC Yöntemi ve Önerilen Derin Model ile Çevresel Seslerin Otomatik Olarak Sınıflandırılması

Muhammed YILDIRIM^{1*}

^{*1} Bilgisayar Mühendisliği, Mühendislik ve Doğa Bilimleri Fakültesi, Malatya Turgut Özal Üniversitesi, Malatya, Türkiye
^{*1} muhammed.yildirim@ozal.edu.tr

(Geliş/Received: 11/01/2022;

Kabul/Accepted: 03/02/2022)

Öz: Gelişen teknoloji ile birlikte farklı teknolojileri bir araya getirmede Nesnelerin İnterneti (IoT) ön plana çıkmaktadır. Nesnelerin interneti özellikle akıllı şehir uygulamalarında da sıklıkla kullanılmaktadır. Akıllı şehir uygulamaları her geçen gün daha da yaygın bir hale gelmektedir. Yapılan bu çalışmada da akıllı şehir uygulamalarında sıklıkla kullanılacak bir uygulama gerçekleştirilmiştir. Bu çalışmada çevre seslerinden oluşan ve bu konuda literatürdeki en büyük veri setlerinden biri olan UrbanSound8K veri seti kullanılmıştır. Akıllı şehir uygulamalarına katkıda bulunmak amacıyla çevresel sesleri sınıflandırmak için yeni bir derin tek boyutlu (1D-CNN) model önerilmiştir. Geliştirilen modelde ilk olarak Mel-frekans kepstrum katsayısı (MFCC) kullanılarak veri setindeki çevresel seslerin öznelik haritaları elde edilmiştir. Daha sonra elde edilen öznelik haritaları geliştirilen 1D-CNN ağında sınıflandırıldığında yüksek bir doğruluk değeri elde edilmiştir. Elde edilen bu doğruluk değeri önerilen modelin ses verilerini sınıflandırma işleminde kullanılabileceğini göstermektedir.

Anahtar kelimeler: Sınıflandırma, MFCC, Derin Öğrenme, IoT, Akıllı Şehirler

Automatic Classification of Environmental Sounds with the MFCC Method and the Proposed Deep Model

Abstract: With the developing technology, the Internet of Things (IoT) is at the forefront of bringing different technologies together. The Internet of Things is also frequently used, especially in smart city applications. Smart city applications are becoming more common day by day. In this study, an application that will be used frequently in smart city applications has been realized. In this study, the UrbanSound8K dataset, which consists of environmental sounds and is one of the largest datasets in the literature, was used. A new deep one-dimensional (1D-CNN) model is proposed to classify environmental sounds to contribute to smart city applications. In the developed model, firstly, the feature maps of the environmental sounds in the data set were obtained by using the mel-frequency cepstrum (MFCC) method. A high accuracy value was obtained when the feature maps obtained later were classified in the developed 1D-CNN network. This accuracy value obtained shows that the proposed model can be used in the classification process of audio data.

Key words: Classification, MFCC, Deep Learning, IoT, Smart Cities

1. Giriş

Şehirleri verimli, teknolojik olarak daha gelişmiş, daha yeşil ve sosyal olarak daha kapsayıcı hale getirmeye yönelik teknolojiler her geçen gün artmaktadır [1]. Gelişen teknolojiyle birlikte akıllı şehir kavramı her geçen gün daha popüler bir hale gelmeye başlamıştır. Akıllı şehir kavramının ortaya çıkmasında IoT teknolojisinin payı oldukça yüksektir [2]. Nesnelerin interneti, elektronik cihazları internete bağladığından dolayı şehirler üzerinde ciddi anlamda olumlu bir etkiye sahiptir [3]. Bu cihazlardan üretilen ses sinyalleri farklı amaçlar için kullanılabilir. Gerçek zamanlı seslerin IoT cihazları sayesinde toplanıp, sınıflandırılması büyük bir önem arz etmektedir. Ses sinyalleri her geçen gün araştırmacılar tarafından yaygın olarak tartışılan sıcak bir konu olmaya başlamıştır. Çevresel seslerin sınıflandırılması önemli ve zorlu bir problemdir [4, 5].

Yapılan bu çalışmada IoT ağlarının ve akıllı şehir uygulamalarının artmasından dolayı çevresel seslerden oluşan bir veri seti kullanılarak çevresel seslerin sınıflandırılması yapılmıştır. Son dönemlerde sınıflandırma konusunda derin öğrenme mimarileri oldukça yaygın bir şekilde kullanılmaktadır [6]. Derin öğrenme yöntemleri görüntü, metin ve ses sınıflandırma gibi alanlarda yaygın bir şekilde tercih edilmektedir. Derin öğrenme mimarilerinin büyük veri setlerinde başarılı olması ve bu mimarilerin özellik haritalarını otomatik çıkarması gibi sebeplerden dolayı derin öğrenmeye olan ilgi artmıştır. Derin öğrenme mimarileri birbirini takip eden katmanlardan oluşmaktadır. Bu mimariler, bir katmanın çıkışı diğer bir katmanın girişi olacak şekilde

* Sorumlu yazar: muhammed.yildirim@ozal.edu.tr. Yazarın ORCID Numarası: ¹ 0000-0003-1866-4721

tasarlanmaktadır. Derin öğrenme mimarilerinde katman sayısının belirli bir kriteri yoktur. Bundan dolayı literatürde kabul görmüş mimarilerin katman sayıları birbirlerinden farklıdır [7].

Derin öğrenme, biyomedikal görüntülerden hastalık teşhisi ve derecelendirilmesi [8], sürücüsüz arabalar [9], nesnelerin interneti [10], video işleme ve görüntü sınıflandırma gibi birçok alanda yaygın olarak kullanılmaktadır [11]. Ayrıca son dönemlerde derin öğrenme mimarileri ses sinyallerinin sınıflandırılması sürecinde de kullanılmaya başlanmıştır. Bu konuda yapılmış farklı alanlarda çalışmalar mevcuttur.

1.1. Literatür Araştırması

Son yıllarda yapay zekâ yöntemleri kullanılarak ses sınıflandırması sıklıkla yapılmaktadır. Çevresel seslerin sınıflandırılmasında da yapılmış çalışmalar mevcuttur.

Song ve arkadaşları yapmış oldukları bu çalışmada çevre seslerini sınıflandırmak için Urbansound8K veri setini kullanmışlardır. Çevre seslerini sınıflandırmak için evrişimsel tekrarlayan sinir ağlarını (CRNN) kullanmışlardır. Araştırmacılar yaptıkları bu çalışmada 4 farklı model önermişlerdir. Bu modeller Keras ve Tensorflow ortamlarında yürütülmüştür. Araştırmacılar önerdikleri modellerde sırasıyla %67.41, %73.92, %77.06 ve %68.07 oranlarında doğruluk değerleri elde etmişlerdir [12].

Chen ve arkadaşları UrbanSound8K veri setini kullanarak çevresel sesleri sınıflandırmayı amaçlamışlardır. Araştırmacılar veri setindeki çevresel sesleri sınıflandırmak için CNN tabanlı bir model önermişlerdir. Ayrıca araştırmacılar yaptıkları bu çalışmada katman sayısının başarımlar üzerindeki etkisini de incelediklerini belirtmişlerdir. Önerilen bu modelde %78 oranında bir doğruluk değeri elde edilmiştir [13].

Demir ve arkadaşları, yaptıkları çalışmada UrbanSound8K veri setindeki çevresel sesleri sınıflandırmak için Sort Time Fourier Dönüşümünü (STFT) kullanarak ses dosyalarını görüntüye dönüştürmüşlerdir. Araştırmacılar, ses görüntülerini oluşturduktan sonra önceden eğitilmiş Evrişimsel sinir ağları (Convolutional Neural Network-CNN) mimarilerini kullanmışlardır. Önerilen modelde VGG16, VGG19 ve Densenet201 mimarileri özellik çıkarımı için kullanılmıştır. Elde edilen özellik haritaları SVM sınıflandırıcısında sınıflandırılmıştır. Yapılan bu çalışmada önerilen modelde %78.14 oranında bir doğruluk değeri elde edilmiştir [14].

Davis ve arkadaşları yaptıkları bu çalışmada UrbanSound8K veri setindeki çevresel sesleri sınıflandırmak için CNN tabanlı bir model önermişlerdir. Araştırmacılar önerdikleri modelin başarımlarını artırmak için veri artırma tekniklerini kullanmışlardır. Ayrıca yapılan çalışmada çevresel ses analizi için en iyi büyütme tekniğinin hangisi olduğunun belirlenmesi için farklı büyütme yöntemleri kullanılmıştır. Araştırmacılar veri büyütme yöntemleri içerisinde en başarılı yöntemin LPCC olduğunu belirtmişlerdir. Yapılan bu çalışmada orijinal veri setinde %67.8 oranında bir doğruluk değeri elde edilmiştir [15].

Piczak ve arkadaşları yapmış oldukları bu çalışmada çevresel sesleri sınıflandırmak için MFCC yöntemini kullanarak özellik çıkarımı yapmışlardır. Yapılan bu çalışmada UrbanSound8K veri seti kullanılmıştır. Araştırmacılar maksimum ortaklama katmanı ile birlikte 2 evrişim katmanı ve 2 tam bağlantılı katmandan oluşan bir model önermişlerdir. Önerilen bu modelde %72.7 oranında bir doğruluk değeri elde edilmiştir [16].

Salamon ve arkadaşları yapmış oldukları çalışmada çevresel seslerin sınıflandırılması için UrbanSound8K veri setini kullanmışlardır. Araştırmacılar yaptıkları çalışmada çevresel seslerin az olmasından dolayı veri çoğullaması yaptıklarını belirtmişlerdir. Önerilen CNN tabanlı modelde çevresel seslerin sınıflandırılmasında %73 oranında bir doğruluk değeri elde edilmiştir [17].

1.2. Yenilik ve Katkıları

Bu çalışmada UrbanSound8K veri seti kullanılarak çevresel seslerin sınıflandırılması amaçlanmıştır. Nesnelerin interneti ağının ve bulut bilişim teknolojilerin gelişmesi beraberinde akıllı şehir kavramını da getirmiştir. Yapılan bu çalışmada akıllı şehirlerde çevresel seslerin sınıflandırılması için bu konuda en büyük veri setlerinden biri olan UrbanSound8K veri seti kullanılmıştır. Bu veri setindeki ses sinyallerinin öznelik haritalarının elde edilmesinde literatürde yaygın olarak kullanılan MFCC yöntemi kullanılmıştır. Daha sonra elde edilen öznelik haritaları geliştirdiğimiz model ile sınıflandırılmıştır. Önerilen modelimizde elde ettiğimiz performans metrikleri diğer çalışmalarla karşılaştırıldığında yüksek bir performansın elde edildiği görülmektedir. Modelimizin çevresel seslerin sınıflandırılmasında elde ettiği doğruluk değeri %84.02'dir.

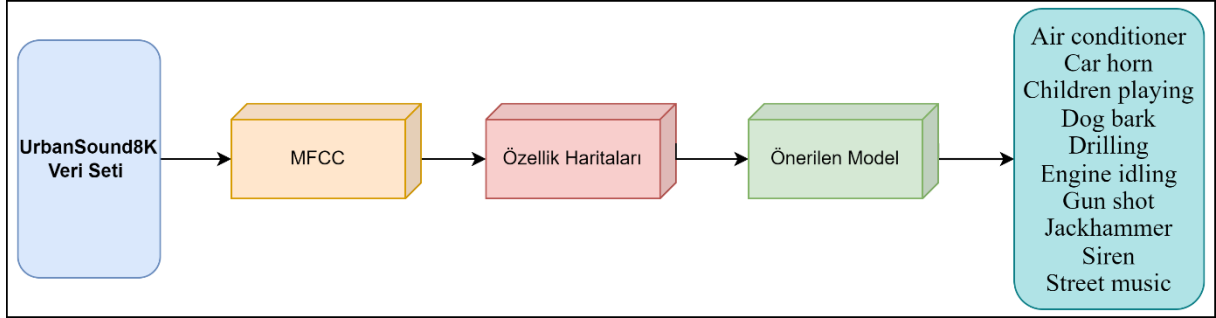
1.3. Makalenin Organizasyonu

Çalışmanın ilk aşamasında genel bilgi sağlamaya ve literatür taramasına odaklanılmıştır. Ayrıca birinci bölümde çalışmanın amacına ve çalışmadaki yeniliklere değinilmiştir. İkinci bölümde kullanılan veri seti,

geliştirilen model ve MFCC yöntemi incelenmiştir. Üçüncü bölümde, çalışmanın uygulama sonuçları, performans ölçütleri ve elde edilen sonuçlara yer verilmiştir. Dördüncü bölüm tartışma ve son bölüm ise sonuç bölümüdür.

2. Materyal ve Yöntem

Bu çalışmada çevresel sesleri sınıflandırmak için UrbanSound8K veri seti kullanılmıştır. Önerilen model ile veri setindeki bu çevresel sesler sınıflandırılmıştır. Şekil 1, önerilen modelin blok diyagramını göstermektedir.



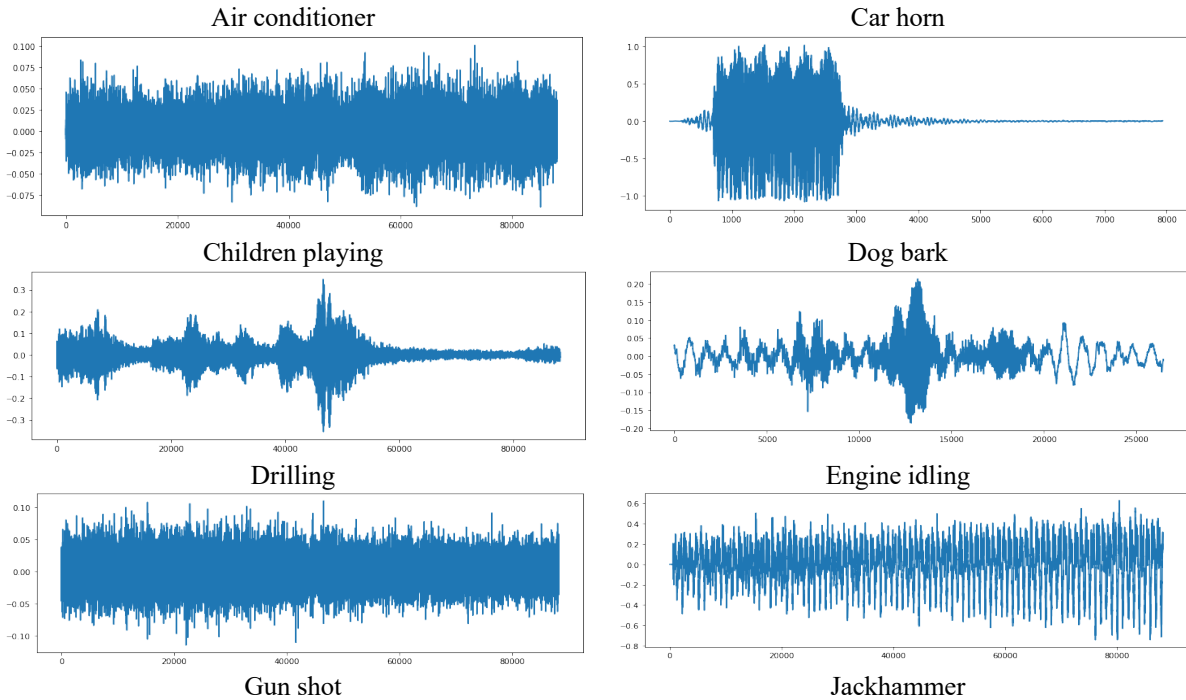
Şekil 1. Önerilen Modelin Kabaca Gösterimi

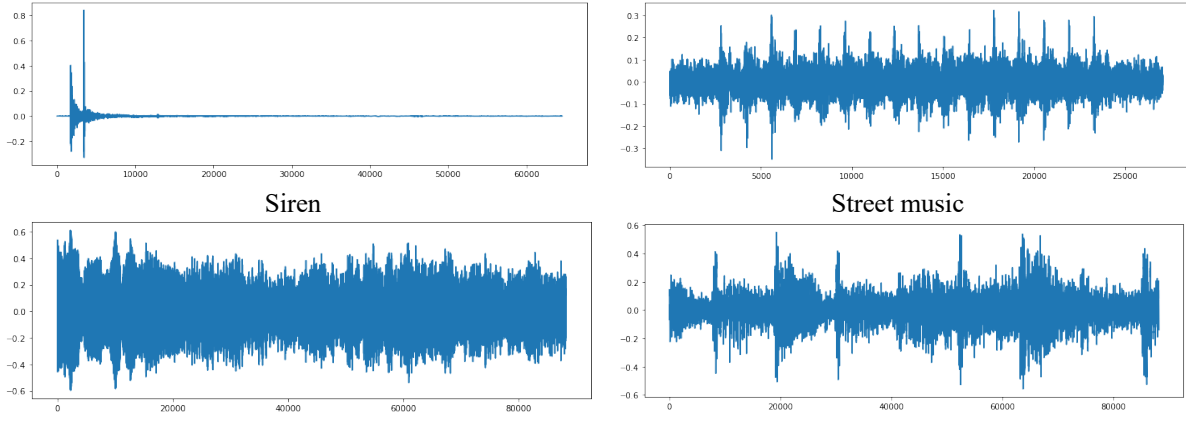
Bu bölümde çalışmada kullanılan veri seti, ses verilerinden öznelik çıkarmak için kullanılan MFCC yöntemi ve ses verilerini sınıflandırmak için geliştirilen model detaylı olarak incelenmiştir.

2.1. Veri Seti

Yapılan bu çalışmada çevresel sesleri sınıflandırmak için UrbanSound8K veri seti kullanılmıştır [18]. UrbanSound8K veri seti çevresel seslerden oluşan literatürdeki en büyük veri setlerinden biridir. Bu veri seti toplamda 9,7 saat olmak üzere 4 saniyeye kadar süreli 8732 etiketli çevresel ses verisinden oluşmaktadır [12]. UrbanSound8K veri setindeki çevresel sesler dış ortamdan elde edildiğinden dolayı gürültü oranı yüksektir. Bundan dolayı UrbanSound8K veri setindeki çevresel sesleri sınıflandırmak oldukça zorlu bir süreçtir.

UrbanSound8K veri seti 10 farklı sınıftan oluşmaktadır. Çalışmada kullanılan ses dosyalarının sinyal olarak gösterimi Şekil 2'de verilmiştir.

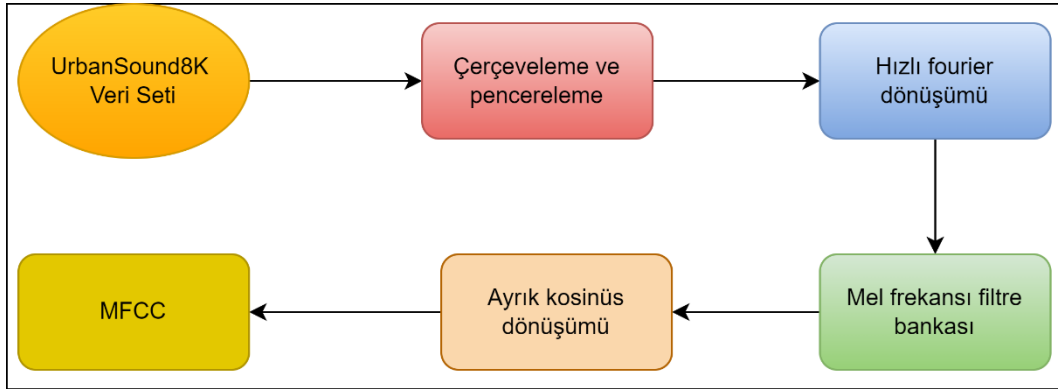




Şekil 2. Çevresel seslerin sinyal olarak gösterimi

2.2. MFCC Özellik Çıkarım Yöntemi

Bu çalışmada UrbanSound8K veri setindeki çevresel seslerin özelliklerini çıkarmak için çok önemli bir Python paketi olan Librosa Kütüphanesi kullanılmıştır. Bu kütüphane, ses sinyali kayıtlarını değerlendirmek için kullanılır. Bu çalışmada, çevresel seslerin özellik haritalarını elde etmek için Mel-frekans kepstrum katsayısı (MFCC) yöntemi kullanılmıştır. MFCC, ses sinyallerinden öznitelikleri çıkarmak için literatürde yaygın olarak kullanılmaktadır. Bu çalışmada, çevresel seslerden oluşan veri kümemizdeki seslerin öznitelikleri çıkarmak için MFCCs yöntemi kullanıldı. MFCCs yöntemi ilk olarak 1980'lerde Davis ve Mermelstein tarafından kullanılmıştır [19]. MFCC algoritması, Hamming penceresini kullanarak bir ses akışını daha küçük pencerelelere yeniden şekillendirerek çerçevelere bölmektedir. Spektrum, Fast Fourier Dönüşümü kullanılarak her çerçeve için üretilir ve her spektrum filtre bankası kullanılarak ağırlıklandırılır. Son olarak, Logaritma ve Ayrık Kosinüs Dönüşümü kullanılarak MFCC vektörü hesaplanır. MFCC'lerin özellik işleme süreci Şekil 3'te gösterilmektedir.



Şekil 3. MFCC özellik işleme süreci

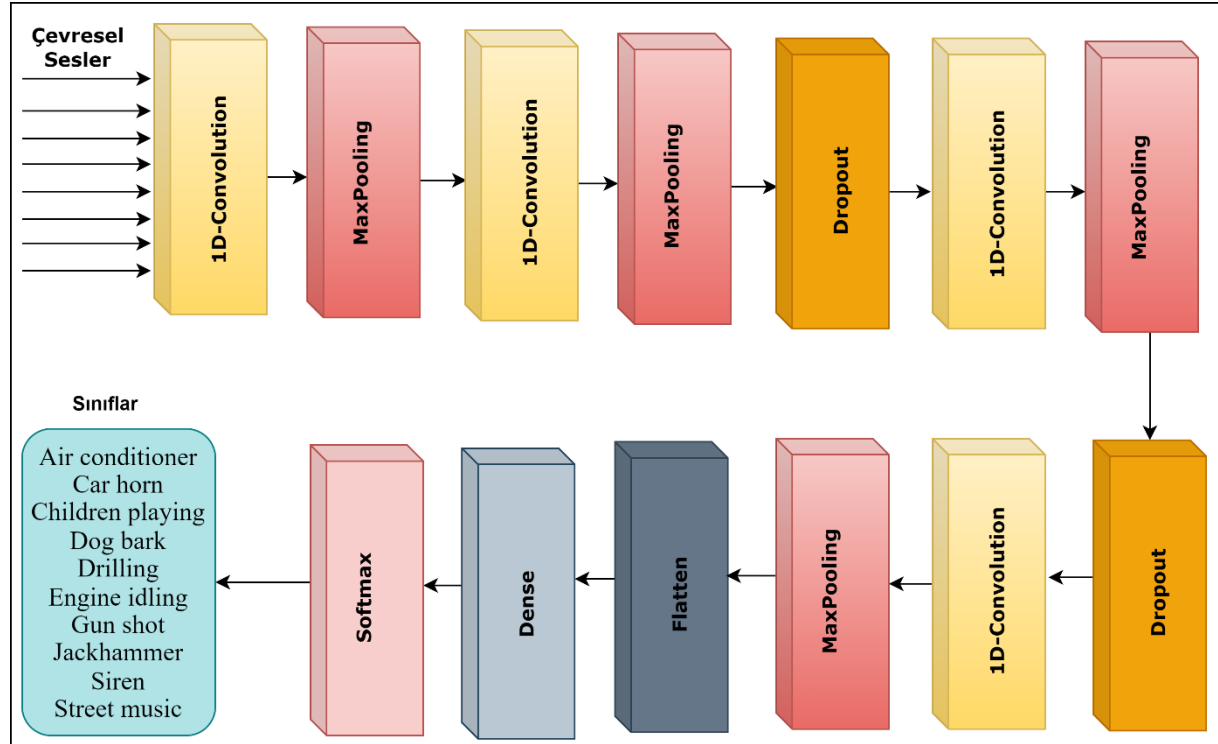
Bu çalışmada her bir ses sinyalinden 40 özellik elde edilmiştir. Daha sonra elde edilen bu özellikler derin öğrenme tabanlı modellerde kullanılmak üzere düzenlenmiştir. Sonuç olarak elde edilen özellik haritaları önerdiğimiz derin model ile sınıflandırılmıştır.

2.3. Önerilen Derin Model

Çalışmada ilk olarak, çevresel seslerin sınıflandırılması için MFCC yöntemi kullanılarak ses sinyallerinin öznitelik haritaları elde edilmiştir. UrbanSound8K veri setinden elde edilen özellik haritaları daha sonra önerdiğimiz derin model ile sınıflandırılmıştır. Son yıllarda CNN'ler görüntü işleme, metin işleme ve ses işlemede sıklıkla kullanılmaya başlanmıştır. CNN'lerin bu özellikleri boyut sayısından bağımsızdır. Görüntü

sınıflandırmasında 2D-CNN ve 3D-CNN kullanılırken, özellikle sinyal analizinde tek boyutlu CNN'ler kullanılmaktadır.

UrbanSound8K veri setindeki çevresel sesleri sınıflandırmak için önerdiğimiz derin ağ modelinde 1D-CNN kullanılmıştır. Geliştirilen model 1D-CNN, 1D-MaxPooling, Dropout, Flatten ve Dense (softmax) katmanlarına sahiptir. Elde edilen öznetelik haritalarının aktivasyonu için Relu, hesaplama karmaşıklığını azaltmak için Pooling katmanı kullanılan bir diğer katmandır. Ağ eğitilirken, ezberleme olayının önüne geçmek için Dropout katmanı kullanılarak, ağdaki bazı düğümleri devre dışı bırakmak amaçlanmıştır. Flatten katmanı, verileri matris biçiminde düzleştirmek için kullanılan bir diğer katmandır. Sınıflandırma işlemi için Softmax tercih edilmiştir. Bu katman, aldığı değerlere göre sınıflandırma işlemi için olasılık değerleri üretir. Bu değerler hangi sınıfa daha yakınsa ilgili sınıfa yerleştirilir. Modelin kabaca gösterimi Şekil 4'te gösterilmiştir.



Şekil 4. Önerilen derin model

Önerilen model incelendiğinde, çalışmada 4 tane 1D-Convolution, 4 tane Maxpooling, 1 tane Dropout, 1 tane Flatten, 1 tane Dense ve 1 tane Softmax katmanının kullanıldığı görülmektedir. Tablo 1, önerilen modelde kullanılan parametreler ve katmanlar hakkında ayrıntılı bilgi vermektedir.

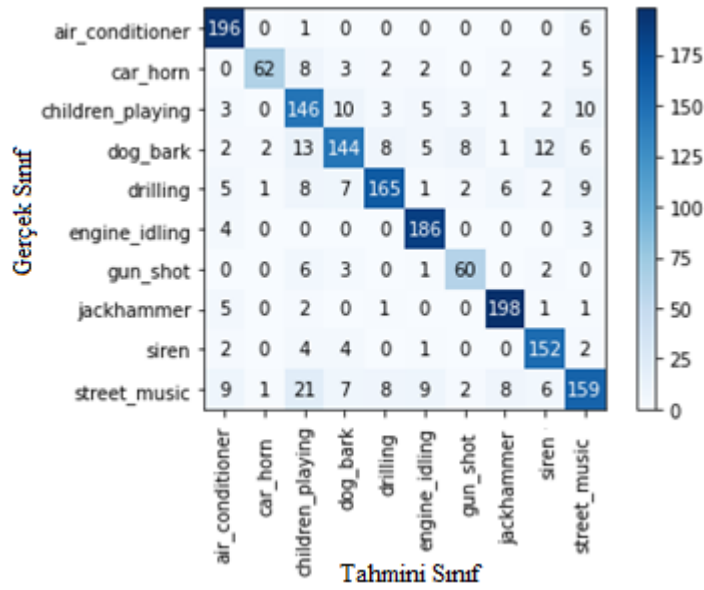
Tablo 1. Önerilen Derin Modeldeki Katman ve Parametreler

Katmanlar	Çıktı Biçimi	Aktivasyon Fonksiyonu	Parametre
Conv1d	(None, 20, 32)	Relu	96
Max_pooling1d	(None, 10, 32)	-	0
Conv1d	(None, 5, 64)	Relu	4160
Max_pooling1d	(None, 2, 64)	-	0
Dropout	(None, 2, 64)	-	0
Conv1d	(None, 1, 64)	Relu	8256
Max_pooling1d	(None, 1, 64)	-	0
Dropout	(None, 1, 64)	-	0
Conv1d	(None, 1, 32)	Relu	4128
Max_pooling1d	(None, 1, 32)	-	0
Flatten	(None, 32)	-	0
Dense	(None, 10)	-	330

3. Deneysel Sonuçlar

Bu çalışma, Intel(R) Core masaüstü bir bilgisayarda gerçekleştirilmiştir. Kullanılan bilgisayar, 16 GB RAM bellek ve 8 GB ekran kartına sahiptir. Çalışmada UrbanSound8K veri seti kullanılarak önerilen derin model ile çevresel seslerin sınıflandırılması amaçlanmıştır. Önerilen modelin performansını ölçmek için farklı değerlendirme parametreleri kullanılmıştır. Doğruluk, Hassaslık, Duyarlılık ve F1-ölçümü bu çalışmada önerilen modelin performansını ölçmek için kullanılan değerlendirme ölçütleridir. Bu performans ölçme metriklerinin çoğu karışıklık matrisi kullanılarak hesaplanmaktadır [20].

UrbanSound8K veri setindeki çevresel sesleri sınıflandırmak için öncelikle MFCC yöntemi kullanılarak özellik haritaları elde edilmiştir. Elde edilen bu özellik haritaları önerilen derin model ile sınıflandırılmıştır. Önerilen model eğitilirken veri setindeki verilerin %80'i modelin eğitimi için kullanılırken, geriye kalan %20'si ise modelin test edilmesi için kullanılmıştır. Bu değerler rastgele seçilmiştir. Önerilen modelde elde edilen karışıklık matrisi Şekil 5'te sunulmuştur.



Şekil 5. Önerilen modelde elde edilen karışıklık matrisi

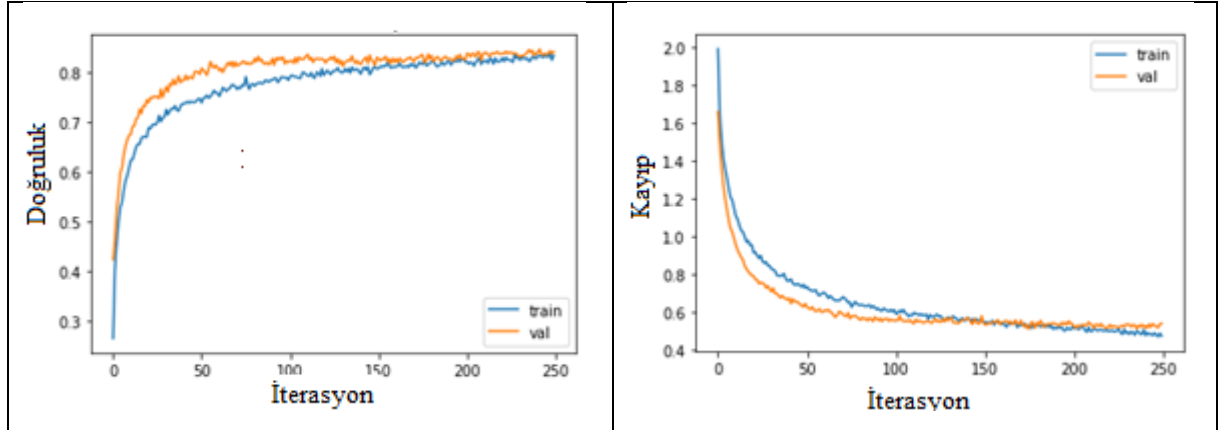
Şekil 5 incelendiğinde sınıflandırma başarımı en yüksek olan sınıf Air conditioner sınıfıdır. Önerilen model bu sınıftaki 203 test sesinin 196 tanesini doğru sınıflandırırken 7 tanesini de yanlış sınıflandırmıştır. Sınıflandırma başarımı en düşük olan sınıf ise Street music sınıfıdır. Önerilen model bu sınıftaki 230 test sesinin 159 tanesini doğru sınıflandırırken 71 tanesini ise yanlış sınıflandırmıştır.

Önerilen modelin başarımleri Tablo 2'de verilmiştir.

Tablo 2. Önerilen Modelin Başarım Metrikleri

Sınıflar	Doğruluk(%)	Duyarlılık(%)	F1-Ölçümü(%)
Air conditioner	96.55	87	91
Car horn	72.09	94	82
Children playing	79.78	70	74
Dog bark	71.64	81	76
Drilling	80.09	88	84
Engine idling	96.37	89	92
Gun shot	83.33	80	82
Jackhammer	95.19	92	93
Siren	92.12	85	88
Street music	69.13	79	74

Tablo 2 incelendiğinde en başarılı sınıfın Air conditioner sınıfı olduğu görülmektedir. Bu sınıfta elde edilen doğruluk oranı %96.55'tir. En başarısız sınıf ise Street music sınıfıdır. Bu sınıfta elde edilen doğruluk değeri %69.13'tür. Önerilen modelin UrbanSound8K veri setindeki çevresel sesleri sınıflandırma doğruluğu %84.02'dir. Önerilen modelin doğruluk ve kayıp eğrileri Şekil 6'da verilmiştir.

**Şekil 6.** Önerilen modelin doğruluk ve kayıp eğrileri

4. Tartışma

Derin öğrenme yöntemleri son yıllarda özellikle görüntü, metin, sinyal ve ses verilerini sınıflandırma da yaygın bir şekilde kullanılmaktadır. Yapılan bu çalışmada UrbanSound8K veri seti kullanılarak çevresel sesler sınıflandırılmıştır. Bu sınıflandırma işlemi için derin öğrenme tabanlı bir model önerilmiştir. Bu çalışma gelişen IoT teknolojisiyle birlikte akıllı şehir uygulamalarının popüler olmaya başlamasından dolayı büyük bir önem arz etmektedir. Literatürde akıllı şehirler ile ilgili yapılmış birçok çalışma mevcuttur [21, 22]. UrbanSound8K veri setindeki çevresel sesleri sınıflandırmak için önerdiğimiz model literatürdeki benzer çalışmalar ile karşılaştırıldığında, önerilen modelin yüksek bir başarı oranı elde ettiği görülmektedir. Çevresel seslerin sınıflandırılması için literatürde yapılan benzer çalışmalar Tablo 3'te verilmiştir.

Tablo 3. Çevresel Seslerin Sınıflandırılması ile İlgili Yapılan Çalışmalar

Referans	Yıl	Yöntem	Veri Seti	Doğruluk(%)
Sang ve arkadaşları[12]	2018	CRNN	UrbanSound8K	%67.41, %73.92, %77.06, %68.07
Chen ve arkadaşları[13]	2019	CNN	UrbanSound8K	%78
Demir ve arkadaşları[14]	2020	STFT, VGG16, VGG19 ve Densenet201	UrbanSound8K	%78.14
Davis ve arkadaşları[15]	2018	CNN	UrbanSound8K	%67.8
Piczak ve arkadaşları[16]	2015	MFCC, CNN	UrbanSound8K	%72.7
Salamon ve arkadaşları[17]	2017	CNN	UrbanSound8K	%73
Önerilen Model	2022	MFCC, CNN	UrbanSound8K	%84.02

Tablo 3 incelendiğinde önerilen modelin %84.02 gibi yüksek bir doğruluk değerine sahip olduğu görülmektedir. Bu çalışmanın temel avantajlarından biri, UrbanSound8K veri setindeki çevresel sesleri 10 farklı kategoride sınıflandırmaktadır. Bu çevresel seslerin sınıflandırılması özellikle akıllı şehir uygulamalarında büyük bir önem arz etmektedir. Uygulamada kullanılan veri seti 8732 tane ses dosyasından oluşmaktadır. Bu sesler dış ortamda kayıt edildiğinden dolayı gürültü oranı yüksektir. Buna rağmen önermiş olduğumuz model literatürdeki benzer çalışmalar ile karşılaştırıldığında, önerdiğimiz modelin yüksek bir doğruluk değeri elde ettiği görülmektedir.

5. Sonuçlar

Ses verilerinin sınıflandırılması gelişen teknoloji ile beraber önemli konulardan biri haline gelmiştir. Çevresel seslerinin sınıflandırılması akıllı şehir uygulamaları başta olmak üzere farklı alanlarda da kullanılmaktadır. Bu çalışmada, 8732 ses dosyası ve 10 sınıftan oluşan UrbanSound8K veri setindeki çevresel sesler önerilen CNN tabanlı bir model ile sınıflandırılmıştır. Önerilen model, 10 farklı çevresel sesi sınıflandırırken %84.02 gibi yüksek bir doğruluk değeri elde etmiştir. Bu doğruluk değeri, önerilen modelin çevresel sesleri sınıflandırmak için kullanılabilirliğini göstermektedir.

Teşekkür

Araştırmacılara veri setlerini paylaştıkları için teşekkür ederim.

Kaynaklar

- [1] Ghazal, T.M., et al., IoT for smart cities: Machine learning approaches in smart healthcare—A review. *Future Internet*, 2021. 13(8): p. 218.
- [2] Teng, H., et al., A low-cost physical location discovery scheme for large-scale Internet of things in smart city through joint use of vehicles and UAVs. *Future Generation Computer Systems*, 2021. 118: p. 310-326.
- [3] Sarkar, N.I. and S. Gul, Green computing and internet of things for smart cities: technologies, challenges, and implementation, in *Green Computing in Smart Cities: Simulation and Techniques*. 2021, Springer. p. 35-50.
- [4] Mandalapu, H., et al., Audio-visual biometric recognition and presentation attack detection: A comprehensive survey. *IEEE Access*, 2021. 9: p. 37431-37455.
- [5] Luz, J.S., et al., Ensemble of handcrafted and deep features for urban sound classification. *Applied Acoustics*, 2021. 175: p. 107819.
- [6] Eroglu, Y., et al., Diagnosis and grading of vesicoureteral reflux on voiding cystourethrography images in children using a deep hybrid model. *Computer Methods and Programs in Biomedicine*, 2021. 210: p. 106369.
- [7] Cengil, E., A. Çınar, and M. Yıldırım. A Case Study: Cat-Dog Face Detector Based on YOLOv5. in *2021 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies (3ICT)*. 2021. IEEE.

- [8] BİNGÖL, H. and B. ALATAS, Classification of Brain Tumor Images using Deep Learning Methods. *Turkish Journal of Science and Technology*, 2021. 16(1): p. 137-143.
- [9] Karmakar, G., et al., Assessing Trust Level of a Driverless Car Using Deep Learning. *IEEE Transactions on Intelligent Transportation Systems*, 2021.
- [10] Ullah, I. and Q.H. Mahmoud, Design and development of a deep learning-based model for anomaly detection in IoT networks. *IEEE Access*, 2021. 9: p. 103906-103926.
- [11] Sarma, M.S., et al., Traditional Bangladeshi Sports Video Classification Using Deep Learning Method. *Applied Sciences*, 2021. 11(5): p. 2149.
- [12] Sang, J., S. Park, and J. Lee. Convolutional recurrent neural networks for urban sound classification using raw waveforms. in *2018 26th European Signal Processing Conference (EUSIPCO)*. 2018. IEEE.
- [13] Chen, Y., et al., Environmental sound classification with dilated convolutions. *Applied Acoustics*, 2019. 148: p. 123-132.
- [14] Demir, F., et al., A new pyramidal concatenated CNN approach for environmental sound classification. *Applied Acoustics*, 2020. 170: p. 107520.
- [15] Davis, N. and K. Suresh. Environmental sound classification using deep convolutional neural networks and data augmentation. in *2018 IEEE Recent Advances in Intelligent Computational Systems (RAICS)*. 2018. IEEE.
- [16] Piczak, K.J. Environmental sound classification with convolutional neural networks. in *2015 IEEE 25th international workshop on machine learning for signal processing (MLSP)*. 2015. IEEE.
- [17] Salamon, J. and J.P. Bello, Deep convolutional neural networks and data augmentation for environmental sound classification. *IEEE Signal processing letters*, 2017. 24(3): p. 279-283.
- [18] Salamon, J., C. Jacoby, and J.P. Bello. A dataset and taxonomy for urban sound research. in *Proceedings of the 22nd ACM international conference on Multimedia*. 2014.
- [19] Davis, S. and P. Mermelstein, Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. *IEEE transactions on acoustics, speech, and signal processing*, 1980. 28(4): p. 357-366.
- [20] Eser, S., A deep learning based approach for the detection of diseases in pepper and potato leaves. *Anadolu Tarım Bilimleri Dergisi*, 2021. 36(2): p. 167-178.
- [21] Chang, V., An ethical framework for big data and smart cities. *Technological Forecasting and Social Change*, 2021. 165: p. 120559.
- [22] Chen, D., P. Wawrzynski, and Z. Lv, Cyber security in smart cities: a review of deep learning-based applications and case studies. *Sustainable Cities and Society*, 2021. 66: p. 102655.