

Hayvan Seslerinin Sınıflandırmasında CNN Yaklaşımı

Fadime COŞKUNER¹, Meryem ŞAHİN², Nurgül AYDIN³

^{1,2,3} Necmettin Erbakan Üniversitesi Seydişehir Ahmet Cengiz Mühendislik Fakültesi Bilgisayar Mühendisliği,
Konya, TÜRKİYE

¹fadimeecoskuner@gmail.com, ²meryemshins@gmail.com, ³aydinnurgul87@gmail.com

ÖZET

Yapay zekâ sistemlerinde etkili olan yapay sinir ağı (YSA)'nın kullanımı gün geçtikçe artmaktadır. YSA modellerinden olan Evrişimsel Sinir Ağı (CNN) yardımıyla, hayvanların seslerinin ayırt edilebilmesi amaçlanmıştır. Sinir ağına yüklenen veri seti ile ilgili hayvanların seslerine göre sınıflandırma yapılmıştır. Çalışmanın sonucunda %98 doğruluk oranı elde edilmiştir. Proje de CNN ile beraber Audio Cats and Dogs veri seti kullanılarak çalışma gerçekleştirilmiştir. Bu yazılımın oluşturulması sürecinde kullanılacak olan yöntem, teknik ve uygulamalar konusunda bilgi verilmiştir.

Anahtar Sözcükler: CNN Evrişimsel sinir ağı, Kedi, Köpek, Sınıflandırma, Yapay sinir ağı, Yapay zekâ.

ABSTRACT

The use of artificial neural network (ANN), which is effective in artificial intelligence systems, is increasing day by day. With the help of Convolutional Neural Network (CNN), which is one of the ANN models, it is aimed to distinguish the sounds of animals. The data set uploaded to the neural network was classified according to the sounds of the related animals. As a result of the study, an accuracy rate of 98% was obtained. In the project, the study was carried out using the Audio Cats and Dogs dataset together with CNN. Information was given on the methods, techniques and applications to be used in the creation of this software.

Keywords: Artificial intelligence, Artificial neural network, Cat, Classification, CNN Convolutional neural network, Dog.

1.GİRİŞ

Günümüzde yaygınlığı günden güne artan yapay zekâ kavramı, bilgisayarda yapılan uzun zamanlı çalışmaların sonucunda bilim adamlarının insan beynini modellemesi ile hayatımıza girmiştir [1]. Yapay zekanın birçok alanda kullanılması ile yapay zekâ yaklaşımları çeşitlendirilmiştir. Bunlar: sürü zekâsı, bulanık mantık, yapay sinir ağları vs. Bu yaklaşımlar içerisinde yapay sinir ağının kullanımı oldukça yaygındır.

Yapay Sinir Ağı (YSA), insan beyninin çalışma prensibinden esinlenerek geliştirilmiş bir çalışma alanıdır. YSA'nın nöron yapısı biyolojik nöronlar model alınarak oluşturulmuştur. Bu teknik, güvenilir sonuçlar vermesi ve doğrusal olmayan sorunların çözümünde aktif olarak kullanılması ile zamanla yaygınlaşmasına neden olmuştur [1]. Kullanım alanlarına göre birçok YSA modeli bulunmaktadır. Bunlar: Algılayıcı (Perceptron), İleri Beslemeli Sinir Ağı (FFNN), Tekrarlayan Sinir Ağı (RNN), Hopfield Ağı, Basamaklı Ağlar (Cascade Network) ve Evrişimsel Sinir Ağları (CNN)'dir.

Literatürde CNN yaklaşımı ile geliştirilmiş çalışma örnekleri bulunmaktadır. Örneğin, [2] bu çalışmada MNIST veri setinin bir varyasyonu olarak el yazısı ile yazılmış Çince sayıların görüntülerini içeren veri kümesi tanıtılmıştır. Makine Öğrenmesi yöntemlerinin güncel ve zorlu araştırma konularından çok katmanlı yapay sinir ağlarının bir türü olan Konvolüsyonel Sinir Ağları (CNN) modeli uygulanıp bir sınıflandırma çalışması yapılmıştır. Sınıflandırma aşamasında %97 doğru sınıflandırma

oranına ulaşılmıştır. Sonuç olarak, önerilen modelin Çin sayı örüntüleri üzerinde yüksek başarımla kullanılabileceği düşünülmektedir. Başka bir çalışmada, [3] yaygın olarak uygulanan dört derin öğrenme yöntemi olan CNN, RNN, LSTM ve GRU kullanılarak geliştirilen modellerin performanslarının bir karşılaştırmasını sağlamayı amaçlamaktadır. Bu modellerin eğitimi ve test edilmesi amacıyla CB513 veri seti kullanılmış, buna ek olarak doğruluk, f1 skoru, doğruluk ve kesinlik gibi performans değerlendirme ölçütleri uygulanmıştır. CNN, RNN, LSTM ve GRU modellerinin doğruluk oranları sırasıyla %82,54, %82,06, %81,1 ve %81,48'dir.

CNN; temel olarak görüntüleri sınıflandırmak, benzerliğe göre kümelemek, fotoğraf arama, ses analizi vb. için kullanılan derin YSA modelidir. Giriş, evrişim ve tam bağlantı katmanı olmak üzere üç katmandan oluşmaktadır. Evrişim katmanında, evrişim (convolution), piksel ekleme(padding), kaydırma (stride), ortaklama (pooling) işlemleri gerçekleştirilir. Bu bildiride, Audio Cats and Dogs veri seti üzerinde CNN yaklaşımı kullanılacaktır. Kedi ve köpek seslerinin analizi CNN ile gerçekleştirilecektir.

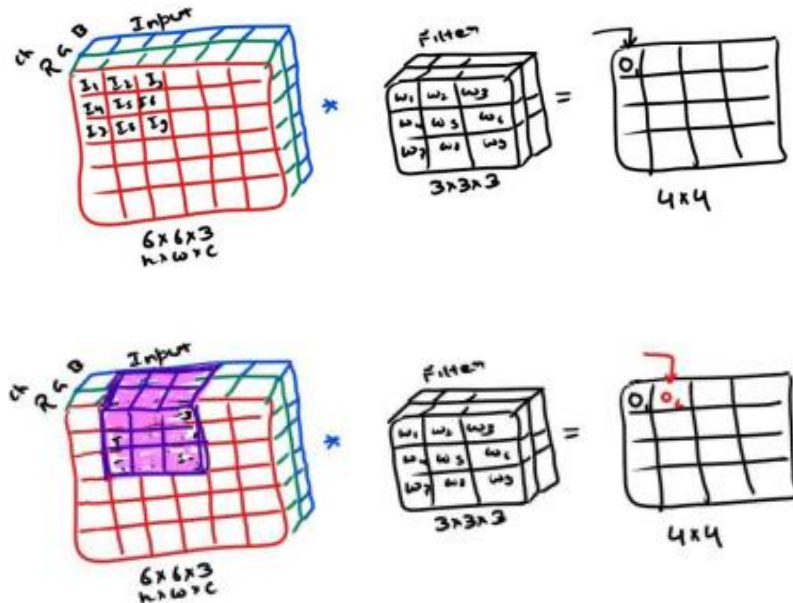
2. EVİRİŞİMSSEL SİNİR AĞLARI (CNN-CONVOLUTIONAL NEURAL NETWORKS)

CNN modelinde giriş katmanı, verilerin girildiği katmandır. Evrişim katmanı, girdi görüntüsünden özellikleri çıkaran katmandır. Bu katman, girdi verilerinin piksellerini kullanarak görüntü özelliklerini öğrenir. Bu katmanda yapılan işlemler:

2.1 Evrişim Katmanı

2.1.1 Evrişim (Convolution)

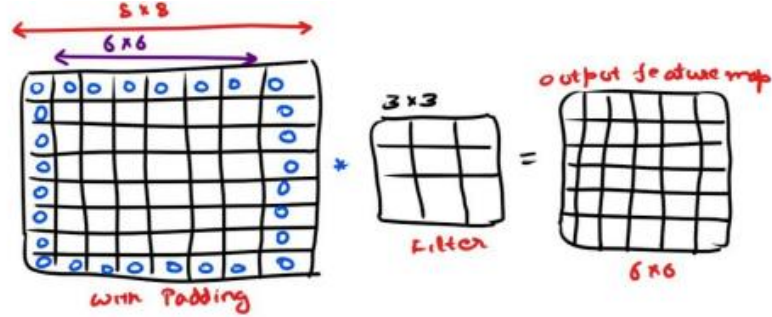
Girdi ve filtre matrisleri arasındaki matematiksel bir işlemdir. Çıktısı da bir matristir. Filtre sol üst köşeden başlayarak her seferinde girdi görüntüsünün üzerinde bir piksel kayar. Görüntünün üzerinde kayarken kendi değerleriyle örtüşen değerleri çarpar ve hepsini toplayarak her örtüşme için tek değer verir. Elde edilen bu çıktı matrisi evrişim işleminin sonuç matrisidir.



Şekil 1: Evrişim İşlemi [4]

2.1.2 Piksel Ekleme (Padding)

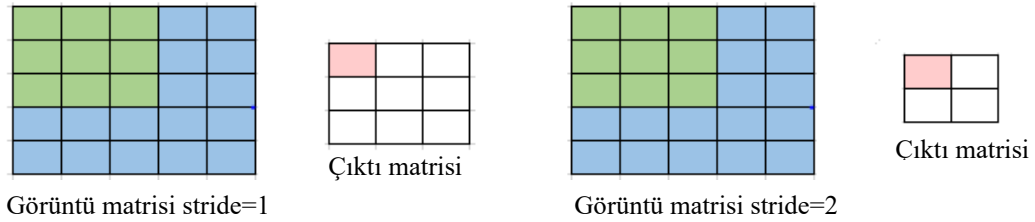
Padding, girdi matrisine eklenen piksel miktarıdır. Evrişim işleminden sonra girdi ve çıktı matrisi arasındaki boyut farkının kontrol edilebilmesini sağlar. Filtre matrisinin görüntüyü kaplaması için daha fazla alan sağlamak amacıyla görüntünün çerçevesine piksel eklenir. Bu işlem görüntülerin daha doğru analizine olanak tanır.



Şekil 2: Piksel Ekleme [4]

2.1.3 Kaydırma (Stride)

Kaydırma adımı, girdi matrisi üzerinde filtre matrisinin kaç adım ile taşınacağını belirtir. Kaydırma sayısı bir olması durumunda filtreler tek seferde bir piksel kaydırılır. Adım değeri ne kadar küçükse çıktı o kadar büyüktür.



Şekil 3: Kaydırma İşlemi

2.1.4 Ortaklama (Pooling)

Ortaklama işlemi ağdaki parametre ve hesaplama miktarını azaltmak için gösterimin uzamsal boyutunu aşamalı olarak azaltmaktadır. Bu işlem aynı zamanda baskın özelliklerin çıkarılması için de yararlı olup modülün etkin bir şekilde eğitilmesini sağlamaktadır. Ortaklama yöntemi için maksimum ortaklama veya ortalama ortaklama kullanılır.

Max (Maksimum) Pooling

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



100	184
12	45

Average (Ortalama) Pooling

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

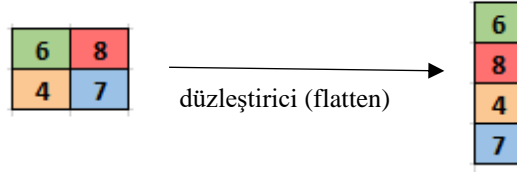


36	80
12	15

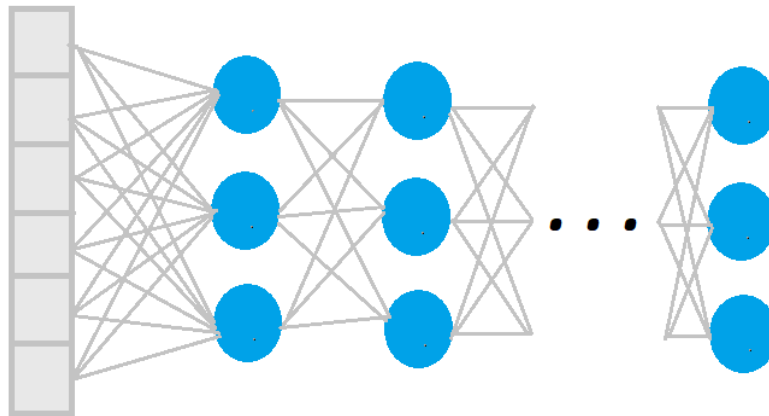
Şekil 4: Ortaklama İşlemi

2.2 Tam Bağlantı Katmanı

Ortaklama işlemi sonucu elde edilen çıktı bu katmanda bir boyutlu vektöre dönüştürülür. Bunun nedeni tek boyutlu vektöre dönüşmüş olan girdinin yapay sinir ağına eklenecek olmasıdır. Bir boyuta dönüştürme işlemini tam bağlantılı katmanda bulunan düzleştirici (flatten) gerçekleştirir.



Şekil 5: Flatten İşlemi



Şekil 6: Tam Bağlantı Katmanı

3. CNN İLE SINIFLANDIRMA

3.1 Veri Seti

Projede veri seti olarak kedi ve köpek seslerinden oluşan Audio Cats and Dogs veri seti kullanılmıştır. Veri seti içerisinde eğitim ve test verisi bulunmaktadır. Kedi ve köpek için bulunan eğitim verisinin yaklaşık %25'i test verisi olarak ayrılmıştır.

Audio Cats and Dogs veri seti ile 50 eğitim gerçekleştirilmiştir. Burada eğitim sayısı ve veri kümesindeki veri miktarının artırılmasıyla tahmin oranında arttığı görülmüştür. Örnek olarak 1., 10., 20., 30., 40., ve 50. eğitimlerin sonuçları görülmektedir. 1. eğitimdeki doğruluk oranı %65 civarında iken 50. eğitim sonunda %98 civarında doğruluk oranı elde edilmiştir.

Epoch 1/50

14/14 [=====] - 4s 128ms/step - loss: 1.1633 - **accuracy: 0.6619** - val_loss: 1.6172 - val_accuracy: 0.8060

Epoch 10/50

14/14 [=====] - 1s 57ms/step - loss: 0.4205 - **accuracy: 0.8857** - val_loss: 0.3560 - val_accuracy: 0.9104

Epoch 20/50

14/14 [=====] - 1s 59ms/step - loss: 0.2830 - **accuracy: 0.9190** - val_loss: 0.3256 - val_accuracy: 0.8955

Epoch 30/50

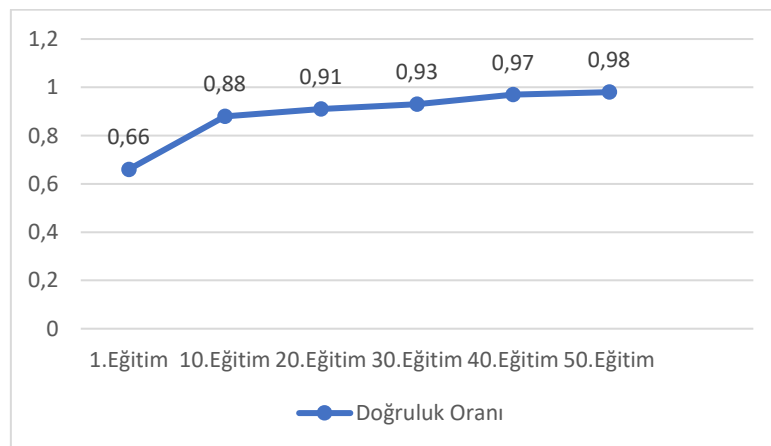
14/14 [=====] - 1s 49ms/step - loss: 0.2824 - **accuracy: 0.9333** - val_loss: 0.2662 - val_accuracy: 0.9552

Epoch 40/50

14/14 [=====] - 1s 52ms/step - loss: 0.1500 - **accuracy: 0.9714** - val_loss: 0.2349 - val_accuracy: 0.9104

Epoch 50/50

14/14 [=====] - 1s 51ms/step - loss: 0.1332 - **accuracy: 0.9810** - val_loss: 0.1704 - val_accuracy: 0.9403



Şekil 7: Eğitim Doğruluk Oranı

Her eğitimin sonunda doğru tahmin oranı artmaktadır. Bu süreç tepe değere geldiğinde doğruluk oranında kayda değer bir değişiklik göstermemektedir. Doğruluk oranının artması noktasında daha farklı ve çok sayıda resim veri tabanına ihtiyaç duyulmaktadır.

3.2 Program Kodları

Program için oluşturulan kod kümesi incelendiğinde öncelikle gerekli kütüphaneler tanımlanmıştır. Ardından veriler listelerde toplanmıştır. Girdi ve doğrulama verileri için listelerde toplanan veriler üzerinde birtakım işlemler yapılmıştır. Bu işlemlerin açıklamaları program içerisinde yorum satırı olarak belirtilmiştir. Veriler CNN formatına uyarlandıktan sonra modeldeki katmanlarda sınıflandırılmak üzere kullanılmıştır. Katmanlar üzerinde ara işlemler uygulanmıştır. Daha sonra modeli compile ederek eğitim işlemi gerçekleştirilmiş olur. Aşağıda proje için oluşturulmuş CNN ağı gösterilmiştir. CNN üç katmandan ve ara işlemlerden oluşmaktadır.

```
model = Sequential()
model.add(Conv1D(8, (3), activation='relu', padding='same', input_shape=(129,1283)))
model.add(MaxPooling1D((2)))
model.add(BatchNormalization())
model.add(Flatten())
model.add(Dropout(rate=0.7))
model.add(Dense(8, kernel_regularizer=l2(l=0.01)))
model.add(Dense(1, activation='sigmoid'))
model.summary()
model.compile(optimizer='adam',
              loss='binary_crossentropy',
              metrics=['accuracy'])

model.fit(X_train, Y_train, epochs=50, batch_size=16,
          validation_data=(X_test, Y_test))
```

4.SONUÇ

Çalışmanın sonucunda kullanımı giderek artan CNN uygulamalarının temel mantığı hakkında bilgi sahibi olunmuştur. Bu proje sınıflandırma tekniğinin kullanıldığı alanlarda etkin bir şekilde kullanılabilir. Örnek olarak veri setindeki hayvan çeşitliliği arttırılarak okul öncesi eğitim de çocukların hayvan seslerini ayırt edebilmesi sağlanabilir. Hayvanlar üzerinde yapılan tür araştırmalarında veri setine ekleme veya değiştirme yapılarak ilgili bir hayvanın tür analizi yapılabilir. Kullanım alanları çeşitlendirilebilmektedir.

Projenin geliştirilmesinde önemli etkenlerden biri veri kümesidir. Veri kümesindeki verilerin sayısı ve çeşitliliği arttırılarak doğruluk oranı tahmini daha iyi bir sonuç verecektir. Örneğin ilgili bir hayvana ait farklı ortamlarda kaydedilen sesler ile bu hayvanın farklı duygu durumlarına verdiği tepkilerin bulunduğu sesler veri setine dahil edilerek projenin çalışma alanı genişletilebilir. Diğer bir etken eğitim sayıdır. Eğitim sayısının veri sayısı ile orantılı bir şekilde arttırılması tahminin doğruluk oranını da arttıracaktır.

Projenin sonucunda kedi ve köpek seslerinin sınıflandırılması için ses analizi yapılmıştır. Ses analizi yapılırken CNN kullanılmıştır. Projedeki veri sayısına uygun olacak şekilde 50 eğitim gerçekleştirilmiştir ve eğitim sonucunda %98 doğruluk oranı elde edilmiştir.

KAYNAKÇA

- 1- Zontul, M. ve Yangın, A., (2017), Yapay Sinir Ağı Teknikleri Kullanarak Eğitim Yayıncılığı Sektöründe Veri Madenciliği, 1(2), 1-15.
- 2- Kayalı, N., Z., ve Omurca S., (2021), Konvolüsyonel Sinir Ağları (CNN) ile Çin Sayı Örüntülerinin Sınıflandırılması, 2548(1304), 184-191.
- 3- Çakmak, E., ve Selvi, İ., H., (2022), Derin Öğrenme Kullanarak (CNN, RNN, LSTM, GRU) Kullanarak Protein İkincil Yapı Tahmini, 6(1), 43-52.
- 4- Yılmaz, T., (2020), Yüz Tanıma ile Duygu Durumu Belirleme, Kastamonu Üniversitesi Mühendislik ve Mimarlık Fakültesi Bilgisayar Mühendisliği Bölümü, Bitirme Projesi
- 5- Sönmez, Zontul, Bülbül (2015), Mevduat Bankalarının Karlılığının Yapay Sinir Ağları ile Tahmini: Bir Yazılım Modeli Tasarımı, BDDK Bankacılık ve Finansal Piyasalar Cilt: 9, Sayı: 1.
- 6- Yazıcıoğlu, N. Yapay Zeka İle Talep Tahmini, Uludağ Üniversitesi, Fen Bilimleri Enstitüsü, Endüstri Mühendisliği Bölümü, Yüksek Lisans Tezi, 2010.
- 7- Deng, L., & Yu, D. (2014), Deep learning: methods and applications. Foundations and trends in signal processing, 7(3–4), 197-387.
- 8- Tüfekçi, M., & Karpat. (2019), Derin Öğrenme Mimarilerinden Konvolüsyonel Sinir Ağları (CNN) Üzerinde Görüntü İşleme-Sınıflandırma Kabiliyetinin Arttırılmasına Yönelik Yapılan Çalışmaların İncelenmesi, HORA2019, Ürgüp, Nevşehir, Türkiye.
- 9- Işcan, Y., & Albora, A., M., (2008), Determination of structure boundaries in Archaeological ruins by using Cellular Neural Networks (CNN), 2165(0608). 2008 IEEE 16th Signal Processing, Communication and Applications Conference
- 10- Utku, A., & Can, Ü., (2022), Covid-19'un Yayılım Tahminine Yönelik Makine Öğrenmesi ve Derin Öğrenme Tabanlı Karşılaştırmalı Bir Analiz: Türkiye İçin Örnek Bir Çalışma, 34(2), 709-717.