Iğdır Üniversitesi Fen Bilimleri Enstitüsü Dergisi, 12(3): 1250 - 1263, 2022
Journal of the Institute of Science and Technology, 12(3): 1250 - 1263, 2022

ISSN: 2146-0574, eISSN: 2536-4618

Bilgisayar Mühendisliği / *Computer Engineering* **DOI: 10.21597/jist.1124674**

Araştırma Makalesi / *Research Article*

Geliş tarihi / *Received*: 01.06.2022 **Kabul tarihi / *Accepted*: 22.06.2022**

Atıf İçin: Bayat S, Işık G, 2022. Aras Kuş Türlerinin Ses Özellikleri Bakımından Derin Öğrenme Yöntemleriyle Tanınması. Iğdır Üniversitesi Fen Bilimleri Enstitüsü Dergisi, 12(3): 1250 - 1263.

To Cite: Bayat S, Işık G, 2022. Recognition of Aras Bird Species From Their Voices With Deep Learning Methods. Journal of the Institute of Science and Technology, 12(3): 1250 - 1263.

Aras Kuş Türlerinin Ses Özellikleri Bakımından Derin Öğrenme Yöntemleriyle Tanınması

Seda BAYAT[1*], Gültekin IŞIK[2]

ÖZET: Bu çalışmada Iğdır Aras Nehri Kuş Cenneti'nde sıklıkla görülen kuş türlerinin seslerinden tanınması üzerinde durulmuştur. Bu amaçla derin öğrenme yöntemleri kullanılmıştır. Biyolojik çeşitliliğin incelenmesi ve analiz edilmesi için akustik gözetleme çalışmaları yapılmaktadır. Bu iş için pasif dinleyici/kaydedici adındaki aygıtlar kullanılmaktadır. Genel olarak bu kaydedici aygıtlarla toplanan ham ses kayıtlarının üzerinde çeşitli analizler gerçekleştirilir. Bu çalışmada, kuşlardan elde edilen ham ses kayıtları tarafımızca geliştirilen yöntemlerle işlenmiş ve daha sonra derin öğrenme mimarileriyle kuş türleri sınıflandırılmıştır. Sınıflandırma çalışmaları, Aras Kuş Cenneti'nde çokça görülen 22 kuş türü üzerinde yapılmıştır. Ses kayıtları 10 saniyelik klipler haline getirilmiş daha sonra bunlar birer saniyelik log mel spektrogramlara çevrilmiştir. Sınıflandırma yöntemi olarak derin öğrenme mimarilerinden Evrişimsel Sinir Ağları (CNN) ve Uzun Kısa-Dönemli Bellek Sinir Ağları (LSTM) kullanılmıştır. Ayrıca bu iki modelin yanında Öğrenme Aktarımı yöntemi de kullanılmıştır. Öğrenme aktarımı için kullanılan ön-eğitimli evrişimsel sinir ağlarından VGGish ve YAMNet modelleriyle seslerin yüksek seviyeli öznitelik vektörleri çıkarılmıştır. Çıkarılan bu vektörler sınıflandırıcıların giriş katmanlarını oluşturmuştur. Yapılan deneylerle dört farklı mimarinin ses kayıtları üzerindeki doğruluk oranları ve F1 skorları bulunmuştur. Buna göre en yüksek doğruluk oranı (acc) ve F1 skoru sırasıyla %94.2 ve %92.8 ile VGGish modelinin kullanıldığı sınıflandırıcıyla elde edilmiştir.

Anahtar Kelimeler: Ses sınıflandırma, kuş tanıma, öğrenme aktarımı, log mel-spektrogram, vggish, yamnet

Recognition of Aras Bird Species From Their Voices With Deep Learning Methods

ABSTRACT: This study focuses on recognizing bird species from their voices, which are frequently seen in Aras River Bird Sanctuary of Iğdır. For this purpose, deep learning methods were used. Acoustic monitoring is carried out to examine and analyze biological diversity. Passive acoustic listeners/recorders are used for this work. In general, various analyzes are performed on the raw sound recordings collected with these recording devices. In this study, raw sound recordings obtained from birds were processed with the methods developed by us, and then bird species were classified with deep learning architectures. Classifications were carried out on 22 bird species that are frequently seen in Aras Bird Sanctuary. Audio recordings were made into 10-second clips and then converted into one-second log mel spectrograms. Convolutional Neural Networks (CNN) and Long Short-Term Memory Neural Networks (LSTM), which are deep learning architectures, were used as classification methods. In addition to these two models, the Transfer Learning method was also used. High-level feature vectors of sounds were extracted with VGGish and YAMNet models, which are pre-trained convolutional neural networks, used for transfer learning. These extracted vectors formed the input layers of the classifiers. Accuracy rates and F1 scores of four different architectures were found through experiments. Accordingly, the highest accuracy rate (acc) and F1 score were obtained with the classifier using the VGGish model with 94.2% and 92.8%, respectively.

Keywords: Sound classification, bird recognition, transfer learning, log mel-spectrogram, vggish, yamnet

[1] Seda BAYAT (**Orcid ID:** 0000-0002-8427-9971), Iğdır Üniversitesi, Mühendislik Fakültesi, Mekatronik Mühendisliği Bölümü, Iğdır, Türkiye

[2] Gültekin IŞIK (**Orcid ID:** 0000-0003-3037-5586), Iğdır Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği Bölümü, Iğdır, Türkiye

***Sorumlu Yazar/*Corresponding Author*:** Seda BAYAT, e-mail: bayatseda@gmail.com

Bu çalışma Seda BAYAT'ın Yüksek Lisans tezinden üretilmiştir.

GİRİŞ

Kuşlar, biyolojik çeşitliliğin korunmasında önemli bir ekolojik türdür. Kuşlar renkleri, tüyleri, ebat ve şekillerinin yanında sesleriyle de birbirinden ayrılır. Bütün kuş türlerinin kendilerine özgü sesleri vardır. İnsanlar bu sesleri dinleyerek kuş türlerini tanıyabilir. Kuş sesleri konusunda kulak aşinalığı fazla olan bir insan, çok sayıda kuş türünü birbirinden ayırt edebilir. İnsanların bu yeteneği yapay öğrenme çalışmalarına esin kaynağı olmuştur.

Derin öğrenme yapay öğrenme alanının popüler konularından biridir. Bu yöntemler, bilgisayarlara insan-benzeri zeki bilişsel özellikler kazandırmayı amaçlar. Öğrenme, tanıma gibi edinimler bu bilişsel özelliklerdendir. Bu yaklaşımda, öğrenilmesi istenen probleme (göreve) özgü çokça veri toplanır. Bu verileri işleyecek güçlü bilgisayarlar kullanarak veriler arasında bir örüntü bulunmaya çalışılır. Örüntüyü bulmak için son yıllarda yeni algoritmalar geliştirilmiştir.

Derin öğrenme algoritmaları farklı veri türleri ve alanlarda şimdiye kadar başarıyla uygulanmıştır. Görüntü işleme (Akhtar & Mian, 2018), ses işleme (Sprengel vd., 2017), makine çevrimi (Cho vd., 2014), doğal dil işleme (Young vd., 2018) alanları bunlardan bazılarıdır. Derin öğrenme, bunlardan ayrı olarak diğer alanlarda da kullanılmaktadır. Örneğin son yıllarda kanser teşhisi(Pacal vd.,2022; Pacal &Karaboga, 2021), akustik gözetleme (Aide vd., 2013; Nguyen vd., 2017) veya biyolojik çeşitliliğin korunması (Salamon vd., 2016) maksadıyla derin öğrenme algoritmaları kullanılmaktadır.

Bu çalışmada 22 kuş türünün derin öğrenme yöntemleriyle tanınması üzerinde durulmuştur. Derin öğrenme yöntemleri olarak Evrişimsel sinir ağları (Convolutional Neural Network, CNN) ve Uzun kısa-dönemli bellek sinir ağları (Long short-term memory, LSTM) kullanılmıştır. Ayrıca öğrenme aktarımı yöntemiyle yüksek seviyeli öznitelik vektörleri elde etmek için, ön-eğitimli VGGish ve YAMNet modelleri kullanılmıştır.

Evrişimsel sinir ağları insanın görme sistemini taklit eden bir derin öğrenme mimarisidir. Özellikle görüntü tanımanın bütün alanlarında kullanılmaktadır. Ses sinyallerine Fourier dönüşümü uygulanarak elde edilen mel-spektrogramlar resme benzeyen verilerdir. Bundan dolayı mel-spektrogramlar üzerinde de CNN mimarisi başarılı sonuçlar vermektedir. Çoğu ses uygulamasında olduğu gibi hayvan seslerinin tanınmasında ve tespit edilmesinde de sinir ağlarından yararlanılmıştır. Biyoakustik olarak bilinen bu çalışmalar örneğin kuşlar (Sprengel vd., 2017), (Grill & Schluter, 2017; Salamon vd., 2017), (Bayat & Işık, 2020) balıklar (Malfante vd., 2018; Mathur vd., 2020) ağaçkakanlar (Florentin vd., 2020; Vidaña-Vila vd., 2020), yarasalar (Mac Aodha vd., 2018; Nguyen vd., 2017) kurbağalar (LeBien vd., 2020; Xie vd., 2020) üzerinde yoğunlaşmıştır. CNN yapısını kullanan ResNet (He vd.S, 2016) ve Inception (Szegedy vd., 2016) gibi popüler modeller de kullanılmıştır. (Joly vd., 2020) çalışmasına göre evrişimsel sinir ağlarının seslerden elde edilen spektrogramlar üzerindeki başarısının ardından, daha önce çok kullanılan Mel kepstrum katsayıları, destek vektör makinesi gibi yöntemler giderek daha az kullanılmıştır. Örneğin (Sprengel vd., 2017) Evrişimsel sinir ağlarının ilk kullanıldığı çalışmalardan biridir. (Kahl vd., 2017) BirdCLEF 2017 (Joly vd., 2017) kuş tanıma yarışması için sağlanan veri kümesini kullanmışlardır. BirdCLEF Xeno-Canto veri tabanındaki 1500 kuş türünün sınıflandırıldığı bir yarışmadır. 512×256 piksel ebadında 5 saniyelik spektrogramlar çok sayıda ön işlemden geçirilerek CNN mimarisinin giriş katmanına verilmiştir. Bu sayede toplamda 940 bin civarı spektrogram elde edilmiştir. Çalışmada büyük ebatlı spektrogramların daha iyi sonuç verdiği raporlanmıştır.

Uzun kısa-dönemli bellek sinir ağları, yaygın olarak kullanılan temelde yinelemeli sinir ağlarını barındıran diğer bir derin öğrenme mimarisidir. Özellikle konuşma tanıma, ses tanıma gibi zamansal

verilerde başarılı sonuçlar vermektedir. Bunun nedeni, bu tarz verilerde var olan uzamsal bilginin yanında zamansal bilginin de bu mimaride hesaba katılmasıdır. Bu sayede mevcut durumun çıkışı hem mevcut duruma hem de önceki zaman adımlarındaki durumlara bağlı olarak değişir (Işık & Artuner, 2020). (Guo vd., 2019) LSTM modelini ses olayı tespiti için kullanmıştır. Evrişimsel sinir ağının üzerine uzun kısa-dönemli bellek sinir ağları monte edilmiştir. Burada kullanılan model bizim çalışmamıza esin kaynağı olmuştur.

Öğrenme aktarımı temel olarak bir alanda edinilmiş bilginin başka alana (hedef) uygulanmasını sağlamak için kullanılan bir yöntemdir (Dpwe, 2021). Kaynak ve hedef adıyla anılan bu alanlar farklı veya birbiriyle ilişkili olabilir. Öğrenme aktarımında ön-eğitimli modeller kullanılır. VGGish (Hershey vd., 2017) ve YAMNet (Hershey vd., 2017) modelleri büyük ses verileri üzerinde eğitilmiş ve sesle ilgili bir problemde kullanılabilen ön-eğitimli modellerdir. Örneğin (Tolkova vd., 2021) VGGish modelinden elde ettiği yüksek seviyeli öznitelikleri kuş seslerini ayrıştırmak için kullanmıştır.

Hayvanların özelliklerinin incelenmesi, kendi aralarındaki iletişimin araştırılması ve gezinti bölgelerinin belirlenmesine yönelik çalışmalar akustik gözetleme yöntemi ile yapılmaktadır. Ses kaydedici mobil birimler incelenmek istenen doğal ortama bırakılır. Bu kaydedicilerle ortam sürekli dinlenerek veri toplanmaktadır. Bu veriler kaydedici birimler tarafından belirli periyotlarla otomatik olarak uzak bir istasyona gönderilmektedir. Gönderilen veriler üzerinde çeşitli analizler yapılarak ortama ilişkin bilgi edinilmektedir.

Ses kaydedici mobil birimler pillerle çalıştıklarından enerjileri tasarruflu kullanılmalıdır. Bu kaydedici birimler küçük olduğu için depolama üniteleri kısıtlıdır. Bundan dolayı depolama birimleri de verimli kullanılmalıdır. Ses kaydedici mobil birimlerin üzerinde tutulan verilerin ve işlemlerin azaltılması ile pil ve depolama tasarrufu sağlanabilir.

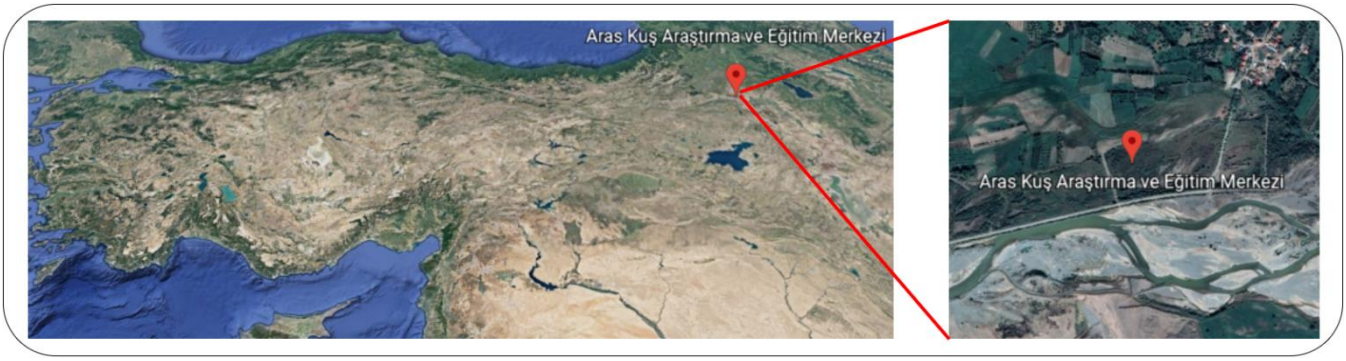
Bu çalışmada, ses kaydedici mobil birimlerin topladığı kuş sesleri ile kuş türlerinin tespiti üzerinde durulmuştur. Dünya üzerinde 10 bine yakın (Barrowclough vd., 2016) kuş türü bulunmaktadır. Bu kadar kuş türünün sınıflandırılması zor olacağından bu türlerin belli sayıda sınırlandırılması gerekir. Bu çalışmada Iğdır Aras Kuş Cenneti'nde (Şekil 1) sıklıkla görülen 22 kuş türü üzerinde sınıflandırma yapılmıştır.

Yapılan çalışmalar genelde kaydediciden alınan ham ses verisi üzerinde olmaktadır. Ancak bu ham verilerin az bir kısmında kuş sesi varken diğer kısımlarda gürültü veya başka sesler bulunur. Bu şekilde ham sesler üzerinde otomatik klipleme (ya da dilimleme, örneğin 10 saniye (s)) yapıldığında kuş ötüşünün bulunmadığı kısımlar da veri kümesine dahil olabilmektedir. Bu şekilde yapılan sınıflandırma işleminin performansı düşmektedir. Bu çalışmada her klipte kuş ötüşünü garantilemek için yeni bir yöntem önerilmektedir. Buna göre, ses dosyalarında kuş ötüşünün bulunmadığı veya belli bir eşik değerinin altında kalan kısımlar silinmektedir. Bu şekilde ham ses verisi işlenmiş yeni ses verisine çevrilmektedir.

İşlenmiş ses dosyalarının kullanılması üç açıdan yarar sağlar: 1) Yukarıda bahsedildiği gibi veri kümelerinin tamamen amaca yönelik seslerden oluşması garanti edilmektedir. Bu şekilde oluşturulmuş veri kümelerinin kullanılmasıyla sınıflandırma (tanıma) başarısı yükselmektedir. 2) İşlenmiş ses kayıtlarının boyutu daha düşük olduğu için hem zamandan hem de işlem gücünden tasarruf edilmektedir. Bunun sonucunda hesaplama kapasitesi nispeten daha düşük Raspberry Pi gibi bilgisayarlarla çalışmak kolaylaşmaktadır. 3) Kaydedici birimlerde enerji ve bellek tasarrufu yapılmaktadır.

Bu çalışmada, ham ve işlenmiş ses örnekleri üzerinde deneyler yapılmıştır. Bunun amacı, işlenmiş verilerle gerçekleştirilen modelin sınıflandırma başarısının ham verilere göre yapılandan iyi olup olmadığını görmektir. Her iki deney türünde de 1) ses kayıtları otomatik olarak 10 saniyelik kliplere

ayrılmış ve 2) daha sonra bu kliplerin logaritmik mel-spektrogramları çıkarılmıştır. 3) Bu mel-spektrogramlar kullanılan modellerin giriş katmanına verilmiştir. 4) Sinir ağının, girişte verilen mel-spektrogram için ürettiği tahmin skoru elde edilmiştir. 5) Nihayetinde modelin performans ölçümleri hesaplanmıştır.



Şekil 1. Aras Kuş Cenneti'nin coğrafi konumu

MATERYAL ve METOT

Verilerin Elde Edilmesi

Iğdır Aras Kuş Cenneti'nde 300'den fazla kuş türü görülmüştür (*Kuzeydoğa Derneği,* 2020). Görülen bu kuş türleri Xeno-Canto web portalında (xeno-canto, 2020) incelenmiştir. İnceleme sonucunda, Aras yöresinde görülen kuşlardan 22 tür için fazla sayıda ses kaydının olduğu belirlenmiştir. Bunun yanında bazı ses kayıtlarının gürültülü oldukları tespit edilmiştir. Bu sonuçlardan hareketle: 1) Yeterli ve dengeli sayıda ses kaydı bulunan türler seçilmiştir. 2) Nispeten gürültüsü az ve Xeno-Canto tarafından A-kalite olarak adlandırılan kategorideki (44 kHz) sesler alınmıştır. Bu iki kıstasa göre 22 kuş türü belirlenmiştir. Bu türler Iğdır-Aras konumunda oldukça yaygın olarak görülmektedir.

Bu aşamadan sonra Xeno-Canto web portalından kuş şarkıları kategorisindeki sesler indirilmiş ve tür bazında klasörlere ayrılmıştır. Alarm ve çağrı seslerinin seçilmemesinin sebebi ise kuş türlerinin frekans zaman grafiği ile ayrımı yapıldığı için, bir kuşun şarkı söylerken çıkardığı sese ait frekansı, başka kuşun alarm sesiyle çakışması riskidir. Benzerlikler karışıklıklara yani kimliklendirme hatalarına sebep olacağı için tüm türlerde sadece şarkı sesleri tercih edilmiştir. Çalışmada oluşturulan veri setindeki 22 kuş türünün isimleri ve ham ses kayıt süreleri Çizelge 1'de verilmiştir.

Seslerdeki Ötüşsüz Kısımların Temizlenmesi

SpeechRate betiği (de Jong & Wempe, 2009) Praat (Boersma & Weenink, 2018) programı için sesteki enerjiye bağlı olarak hecelerin bulunmasını sağlamak üzere yazılmıştır. (de Jong & Wempe, 2009) çalışmasında bahsedilen eşik değerler, enerji ve perdenin bulunması için bu çalışmada değiştirilmeden kullanıldı. Bu eşik değerler ve tarafımızca yazılan ayrı bir Praat betiği kullanılarak ses dosyasında ötüşsüzlük olarak belirlenen kısımlar silinmiştir. Bu şekilde ham ses kayıtlarının boyutu yaklaşık olarak yarıya düşürülerek işlenmiş ses kayıtları elde edilmiştir.

Tarafımızca yazılan betiklerle ham seslerin içindeki ötüşlülük oranları bulunmuştur. Örneğin Karatavuk (Tu Me) için ham kaydın içindeki ötüşlü kısım oranı %34.51'dir. Bu oranın küçük olması ham ses kaydının içindeki ötüş miktarının az olması demektir. Kayıt sürelerinin içindeki ötüş süresinin az olması akustik gözetlemede beklenen bir şeydir.

Mel-Spektrogram Oluşturma

İnsanın işitme sistemi logaritmik çalışmaktadır. Yani sesler insan kulağında 1 kHz'e kadar doğrusal olarak duyulurken 1 kHz'den sonra logaritmik bir hal almaktadır. Bu işitme davranışını

modellemek amacıyla frekans mel-ölçekli frekansa dönüştürülmektedir. Ses kayıtlarının (Hershey vd., 2017) çalışmasında belirtildiği gibi log mel-spektrogramları elde edildi:

Çizelge 1. 22 türe ait veri kümesinin özellikleri

Türün Türkçe Adı	Türün Latince Adı	Kısaltması	Kayıt Süresi (dakika)
Ağaç İncirkuşu	Anthus Trivialis	An Tr	57
Ağaç Kamışçını	Locustella Fluviatilis	La Fl	55
Akgerdanlı Ötleğen	Sylvia Communis	Sy Co	71
Alakarga	Garrulus Glandarius	Ga Gl	59
Benekli Bülbül	Luscinia Luscinia	Lu Lu	72
Boyunçeviren	Jynx Torquilla	Jy To	56
Boz Ötleğen	Sylvia Borin	Sy Bo	59
Bülbül	Luscinia Megarhynchos	Lu Me	63
Büyük Baştankara	Parus Major	Pa Ma	73
Büyük Kamışçın	Acrocephalus Arundinaceus	Ac Ar	83
Çalı Kamışçını	Acrocephalus Palustris	Ac Pa	58
Çayır Taşkuşu	Saxicola Rubetra	Sa Ru	79
Çekirge Kamışçını	Locustella Naevia	Lo Na	66
Bayağı Çıvgın	Phylloscopus Collybita	Ph Co	51
Bayağı Çitkuşu	Troglodytes Troglodytes	Tr Tr	62
Bayağı Çütre	Carpodacus Erythrinus	Ca Er	51
Bayağı Guguk	Cuculus Canorus	Cu Ca	56
İbibik	Upupa Epops	Up Ep	58
Karatavuk	Turdus Merula	Tu Me	62
Bayağı Kızılkuyruk	Phoenicurus Phoenicurus	Ph Ph	65
Orman Toygarı	Lullula Arborea	Lu Ar	61
Bayağı Sumru	Sterna Hirundo	St Hi	45
		Toplam Süre:	1362

1. Tüm sesler 16 kHz ve tek kanallı (mono) olacak şekilde yeniden örneklenmiştir.

2. Kısa-Zamanlı Fourier dönüşümü için 25 ms pencere genişliği, 15 ms örtüşme genişliği kullanılarak spektrogram bulunmuştur.

3. Spektrogramlar 64 adet mel filtresinden geçirilerek mel-spektrogram elde edilmiştir.

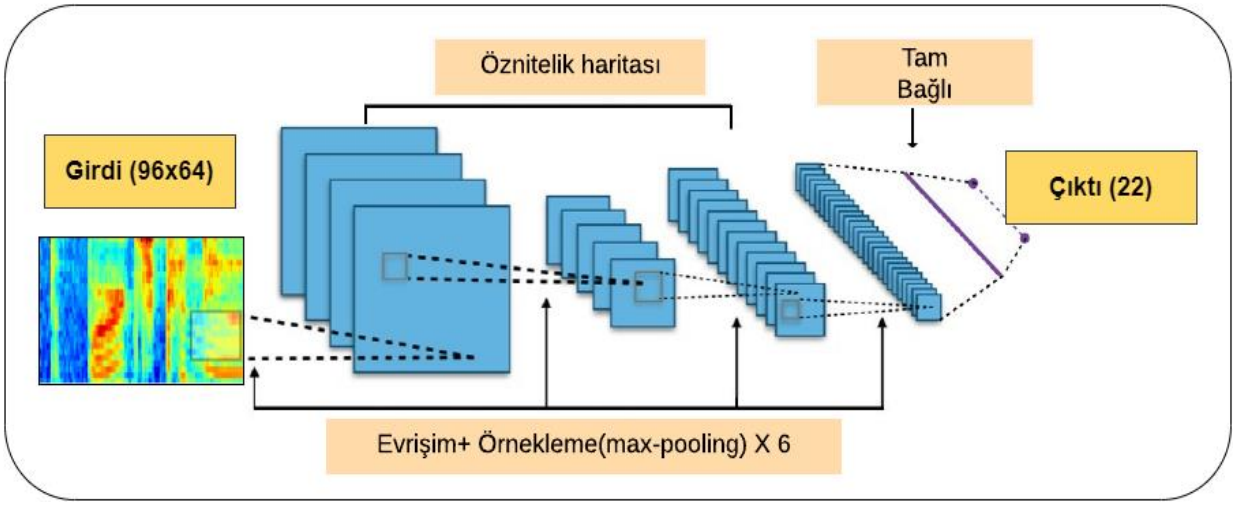
4. Mel spektrogramın logaritması alınmıştır.

5. Bu özelliklerle, örtüşmesiz olarak alınan her örneğin 96 çerçeve ve 64 mel bandından (96×64) oluşacak şekilde 1 saniyelik örnekler elde edilmiştir. Bu işlemler için Librosa (McFee vd., 2015) kütüphanesi kullanılmıştır.

Log mel-spektrogramlardan ham sesler için ayrı, işlenmiş sesler için ayrı veri kümeleri oluşturulmuştur. Bu veri kümeleri sırasıyla %70, %15, %15 oranlarında eğitim, doğrulama ve sınama kümelerine ayrılmıştır.

Kullanılan Evrişimsel Sinir Ağı (CNN) Mimarisi

Bu çalışmada kullanılan derin öğrenme modellerinden biri, evrişimsel sinir ağı modelidir. İnsanlar bir görsele baktıklarında nesneleri, nesnelerin renklerini, şekil ve boyutlarını, nesne sayısını, duruş şekilleri gibi özelliklerini rahatlıkla söyleyebilirler. Aynı görseli bilgisayarlar, sayısal değerlere dönüştürerek bir sayı matrisi olarak algılar. Evrişimsel sinir ağları, görüntüleri birden fazla gizli katmandan geçirerek içindeki tüm nesneleri ve bu nesnelerin özniteliklerini çıkarmaktadır. Günümüzde görüntü sınıflandırma, nesne tanıma, görüntü bölütleme gibi işler evrişimsel sinir ağları ile kolaylıkla ve başarılı bir şekilde gerçekleştirilmektedir (Yamashita vd, 2018). Bu çalışmada evrişimsel sinir ağının tanımlanması ve eğitilmesi işlemleri için Tensorflow (Abadi vd., 2016) arka ucunda olmak üzere Keras (Chollet, 2015) kütüphanesi kullanılmıştır. Keras açık kaynak kodlu bir kütüphanedir ve Python dilinde yazılmıştır.



Şekil 2. Evrişimli sinir ağının adımları

Evrişimsel sinir ağının gizli katmanlarında doğrultulmuş lineer birim (rectified linear unit, ReLU) aktivasyon fonksiyonu, çıkış katmanında ise Softmax fonksiyonu kullanılmıştır (Şekil 2). Gizli katmanlarda 6 evrişim ve örnekleme işlemi yapılarak mel-spektrogramların öznitelik haritaları çıkartılmıştır. Daha sonra bu öznitelik haritaları düzleştirilerek tam bağlı katmandan, sonrasında ise çıkış katmanından geçirilmiştir. En sonda softmax fonksiyonu ile sınıflara ait olasılık değerleri hesaplanmıştır. Softmax fonksiyonu bütün sınıflar için [0, 1] aralığında ve toplamda 1 olacak şekilde olasılık değerleri hesaplar. Bunun sonucunda en yüksek olasılık değerini alan sınıf, model tarafından seçilir.

Kullanılan CNN mimarisinde 6 adet evrişim (sırasıyla 32, 32, 64, 64, 128, 128 öznitelik haritası ve 3×3 kernel) ve örnekleme (2×2 kernel) katmanı bulunmaktadır. Sonrasında sırayla 100 ve 10 adet birimden (nöron) oluşan iki tam bağlı katman gelmektedir. Amaç fonksiyonu olarak Keras'ta bulunan kategorik çapraz entropi kullanılmıştır.

CNN-LSTM Modeli

Bu çalışmada kullanılan diğer model Evrişimsel uzun kısa-dönemli bellek (CNN-LSTM) sinir ağı modelidir. 10 saniyelik kliplerden elde edilen birer saniyelik örnekler bu modele girdi olarak verilmiştir. Bu modelde iki evrişimsel sinir ağı ve bir uzun kısa-dönemli bellek (LSTM) bulunmaktadır.

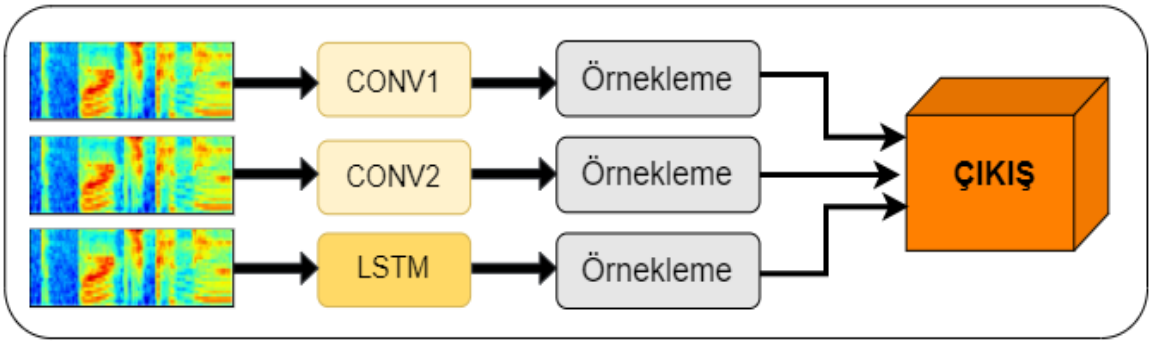
İlk evrişimsel sinir ağında Log Mel-spektrogramlar *Relu* aktivasyon fonksiyonun olduğu gizli katmanlardan geçirilerek öznitelikler çıkarılmıştır. Örnekleme işlemlerinden sonra tam bağlı katman gelmektedir. Birinci evrişim işlemi (sırasıyla 16, 32, 64, 128, 128 öznitelik haritası, 4×4 kernel) bu şekilde tamamlanmıştır.

İkinci evrişimsel sinir ağında *tanh* fonksiyonlu gizli katmanlar vardır. 7×7 bir kernel ile filtrelenen mel-spektrogramların öznitelikleri çıkarılmıştır. Daha sonra *Relu* fonksiyonunun kullanıldığı gizli katman ve tam bağlı katman bulunmaktadır. Örneklemeler 2×2 kernel ile evrişim ise (sırasıyla 16, 32, 64, 64, 128, 128 öznitelik haritası, sırasıyla 7×7, 5×5, 3×3, 3×3, 3×3 kernel) hiper parametreleriyle tamamlanmıştır.

LSTM modelinde tek bir LSTM katmanı kullanılmıştır. Verilen birer saniyelik log mel-spektrogramlar 10 zaman adımı süresince bu katmandan geçirilir. En son zaman adımından sonra örnekleme işlemi yapılır. 22 türe ait ses verileri Şekil 3'te gösterilen CNN-LSTM modelini kullanmıştır.

Öğrenme Aktarımı

Bir ağı ön-eğitimli ağırlık parametreleri ile eğitmek zaman ve performans bakımından yarar sağlar. Öğrenme aktarımı temel olarak bir alanda (kaynak) edinilmiş bilginin başka alana (hedef) uyarlanmasını sağlayan bir yöntemdir. Kaynak ve hedef olarak adlandırılan bu alanlar farklı veya birbiriyle ilişkili olabilir. Bu sayede, çok ve büyük veriyle eğitilmiş kaynak modelinin bilgisi hedef modele transfer edilmektedir.



Şekil 3. CNN-LSTM modelinin özet hali

Öğrenme aktarımında ön-eğitimli ağın parametreleri dondurulur yani güncellenmez. Ağ parametrelerinin bu şekilde dondurulmasının iki nedeni vardır. Birincisi zaten az olan verilerle çok katmanlı bir ağın eğitilmesi aşırı uyum sorunlarına neden olmaktadır. İkincisi ise ön-eğitimli modellerin genel ve büyük ama hedef alanla ilişkili veri kümesi üzerinde eğitilmiş olmasıdır. Kaynak modelin transfer edilmesiyle hedef ağ, büyük veri kümesinden edinilen temel bilgileri alır. Hedef ağın eğitimi sırasında elde edilen ağırlık parametrelerinin geriye doğru yayılması (back-propagation) gerekir. İşte bu, parametrelerin kaybedilmesi anlamına gelir ki istenen bir şey değildir. Bu durumda, transfer edilen ağ, kendisine verilen giriş üzerinde sadece ileriye doğru (forward propagation) işlemler yapmaktadır.

VGGish Ön Eğitimli Modeli

Öznitelik çıkartımı için derin ön eğitimli VGGish modeli (Hershey vd., 2017) kullanılmıştır. VGGish, VGG modelinin (Simonyan & Zisserman, 2015) değiştirilip eğitilerek elde edilen ve ses vektörü üreten bir derin evrişimli sinir ağı modelidir. VGGish modeli AudioSet veri kümesi (Gemmeke vd., 2017) üzerinde eğitilmiştir. AudioSet insanlar tarafından etiketlenmiş 2 milyonun üzerinde 10 saniyelik Youtube videolarının ses kayıtlarından oluşan bir veri kümesidir. Bu veri kümesinde, içinde hayvan ve kuş seslerinin de olduğu 632 ses sınıfı mevcuttur. VGGish modeli, kendisine giriş olarak verilen log mel-spektrogram özniteliklerini 128 boyutlu (128-B) yüksek seviyeli vektörlere dönüştürür. Bu vektörler daha sonra bir sınıflandırma modelinin girişinde kullanılabilir.

Ön-eğitimli VGGish evrişimli sinir ağı modelinin Tensorflow'daki (Abadi vd., 2019) parametreleri (Dpwe, 2021) adresinden indirildi. VGGish modeliyle aynı yapıda bir model inşa edildi. VGGish modelinden alınan ağırlık parametreleri kendi içindeki son tam bağlı katmandan alındığı için tekrar aktivasyon fonksiyonundan geçirilmesi gerekir. Bunun için VGGish modelinin üzerine *Relu* aktivasyon fonksiyonunu kullanan bir adet tam bağlı katman ve Softmax katmanları eklendi. Softmax katmanındaki düğümler 22 sınıfı ayıracak şekilde oluşturuldu. Öğrenme aktarımın gereği olarak ön-eğitimli VGGish ağının parametreleri dondurulduu. Bu şekilde ağdaki parametrelerin tekrar güncellenmesinin önüne geçilerek sadece yeni eklenen katmanların eğitilmesi sağlanmaktadır. Bu süreç: 1) Verilen ses kayıtlarının VGGish modelinden geçirilerek özniteliklerinin çıkartılması, 2) Çıkartılan bu özniteliklerin yeni eklenen katmanlarca sınıflandırılmasının sağlanması adımlarıyla özetlenebilir.

YAMNet Ön Eğitimli Modeli

YAMNet (Hershey vd., 2017), VGGish gibi AudioSet veri kümesi üzerinde 521 ses sınıfını tanıyan ön-eğitimli bir derin öğrenme ağıdır. MobileNetV1 (Howard vd., 2017) modelinin evrişim mimarisi kullanılarak eğitilmiştir. YAMNet kendi başına bir sınıflandırıcı olarak kullanılabilir ancak yaptığımız bu çalışmada YAMNet bir öznitelik dönüştürücü olarak kullanılmıştır. YAMNet modeli, kendisine giriş olarak verilen ses klibini 3 adet çıktıya çevirir: 1) Her 1 saniyelik örneğin (çerçevenin) sınıf skoru, 2) Her çerçevenin yüksek seviyeli öznitelik vektörü, 3) Her çerçevenin log mel spektrogramı. Buradaki 1 saniyelik çerçeveler 0.5 saniyede bir alınmaktadır.

Sınıf skorları 1 saniyelik çerçeve düzeyinde alındığı için bu skorların ortalaması alınarak klip düzeyinde skorlar bulunur. Elde edilen yüksek seviyeli öznitelik vektörü her çerçeve için 1024-boyutludur (1024-B). Öznitelik vektörleri daha üst katmanlar için bir giriş verisi olarak kullanılabilir. Bu çalışmada YAMNet modeli de VGGish modeli gibi çok fazla etiketli örneğe ve uçtan uca eğitime gerek duymadan öğrenme aktarımı amacıyla kullanılmıştır. Öğrenme aktarımı için kullanılan ön eğitimli modellerin akış şeması Şekil 4'te görülmektedir.

10 saniyelik klipler VGGish modeli için 1 saniyelik log mel spektrogramlara çevrildi ve bu örneklerle model eğitildi. YAMNet modeline kliplerin bir boyutlu (1-B) sinyal dizisi [-1.0, 1.0] aralığında normalize edilerek verildi. Daha sonra test aşamasında bu klipler eğitilen modellerden geçirilerek çıkan sonuçların ortalaması alındı. Bu ortalamaların en yükseği bulunarak test için verilen klibin sınıfı belirlendi.

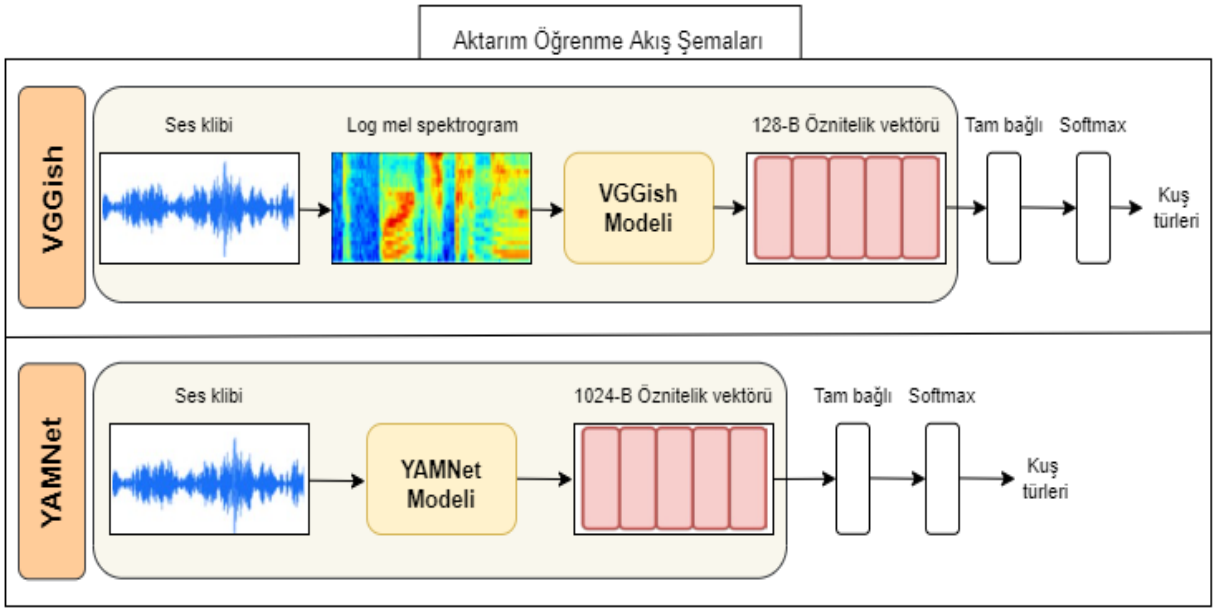
Adam optimizasyon algoritması (Kingma & Ba, 2015) 0.01 öğrenme katsayısı ile başlatıldı. Daha sonra, doğrulama kümesinde üst üste 10 döngüde doğruluk oranı artmazsa öğrenme katsayısı yarılanarak güncellendi. Amaç fonksiyonu olarak çapraz entropi kullanıldı.

Değerlendirme Kriterleri

Modellerin sınıflandırma performansları doğruluk (*Acc*) ve F_1 skorlarına göre hesaplanmıştır. Bu kriterlerin hesaplanması için kullanılan formüller aşağıda verilmiştir. Burada *DP*: *Doğru Pozitif*, *DN*: *Doğru Negatif*, *YP*: *Yanlış Pozitif*, *YN*: *Yanlış Negatif*, *P*: *Pozitif*, *N*: *Negatif* anlamlarına gelmektedir.

$$Acc = \frac{DP + DN}{P + N} \tag{1}$$

$$F_1 = \frac{2 \times DP}{2 \times DP + YP + YN} \tag{2}$$



Şekil 4. Öğrenme aktarımı için uygulanan modellerin akış şeması.

BULGULAR VE TARTIŞMA

Çalışmada kullanılan CNN, CNN-LSTM ve VGGish modellerinin girişine uygulanan 1 saniyelik mel-spektrogramlar 10 saniyelik kliplerden üretilmiştir. YAMNet modelinin girişine ise Şekil 4'te görüldüğü gibi 1 saniyelik klipler dalga formunda verilmiştir. YAMNet modelinin çıktılarından birisi zaten log mel spektrogramlardır. Modeller eğitildikten sonra 10 ve 30 saniyelik ses klipleri üzerinde sınamalar yapılmıştır. Modeller her 1 sn için skor ürettiğinden, bu skorların 10 ve 30 saniye için ortalamaları alınmıştır. Örneğin 10 saniyelik bir sınama verisi için 10 adet olasılık skorunun ortalaması alınarak bu sınama verisinin sınıfı bulunmuştur.

Bu şekilde sınama kümesinde yapılan sınıflandırma işlemlerinin bütün modeller için sonuçları Çizelge 2'de verilmiştir. Ham ve işlenmiş ses verileri üzerinde sınamalar ayrı ayrı yapılmıştır. Bunun nedeni daha önce belirtildiği gibi tarafımızca yazılan Praat betikleriyle yapılan temizleme işlemlerinin sınıflandırma başarımına etkisinin görülmek istenmesiydi. Bu açıdan bakıldığında bütün modellerin işlenmiş veriler üzerinde doğruluk oranlarının ve F_1-skorlarının arttığı görülmektedir. Bu sayede, sahada toplanmış olan ham verilerin içindeki sessizliğin ve diğer gürültülerin temizlenmesi gerektiği ortaya konulmaktadır.

Ham veriler içindeki sessizlik bölgelerinin ve gürültünün temizlenmesi sonucunda ses kaydının boyutu düşürülmektedir. Bunun sonucunda elde edilen yeni ses kaydının işlenmesi için zamandan ve bilgisayar işlem gücünden tasarruf edileceği açıktır. Ayrıca sahada bulunan kaydedicilerin enerjisinin ve belleğinin daha uzun sürelerle kullanılabileceği söylenebilir. Bunların yanısıra sınıflandırma başarımının artması bir diğer avantaj olarak karşımıza çıkmaktadır.

Modeller içinde VGGish modelinin hem ham veriler hem de işlenmiş veriler üzerinde en yüksek doğruluk oranına (0.942) ve F_1 skoruna (0.928) ulaştığı görülmektedir. Bütün sınama sürelerinde de ayrıca diğer modelleri geçmiştir. VGGish modeli öğrenme aktarımı yaklaşımıyla yüksek seviyeli öznitelik çıkarıcı olarak kullanılmıştır. Bu açıdan bakıldığında, çalışmada kullanılan modelin (Şekil 4) eğitim süresinin düşük olduğu söylenebilir. Hem yüksek başarım hem de eğitim süresinin düşük olması VGGish modelinin avantajıdır. VGGish modelinin CNN ve CNN-LSTM modellerini geride bırakması, üzerinde ön-eğitildiği AudioSet veri kümesinin kuş ve hayvan seslerini barındırmasına bağlanabilir.

Sınama süreleri açısından bakıldığında bütün modellerin 30 saniyelik veriler üzerinde daha iyi performans sağladıkları görülmektedir. Sınama süresi olarak bu çalışmada en fazla 30 sn kullanılmıştır. Daha uzun süreli sınama verileri de kullanılabilir. Ancak sınama süresinin artmasının, sonucun elde edilmesi için geçen süreyi de arttıracağı hesaba katılmalıdır. Dolayısıyla bu iki süre arasında bir denge kurulması gerektiği açıktır.

Çizelge 2. Ham ve işlenmiş ses verilerinden 10 ve 30 saniyelik klipler üzerinde yapılan sınamalar sonucunda elde edilen doğruluk (*Acc*) ve F_1 skorları [0, 1] aralığında verilmiştir.

	Ham Ses Verileri (Sınama Kümesi)				**İşlenmiş Ses Verileri (Sınama Kümesi)**			
	10 sn		**30 sn**		**10 sn**		**30 sn**	
Modeller	***Acc***	F_1	***Acc***	F_1	***Acc***	F_1	***Acc***	F_1
CNN	0.771	0.755	0.883	0.872	0.815	0.801	0.921	0.892
CNN-LSTM	0.672	0.662	0.721	0.715	0.718	0.712	0.773	0.754
VGGish	0.775	0.764	**0.901**	**0.890**	0.820	0.816	**0.942**	**0.928**
YAMNet	0.764	0.762	0.885	0.860	0.815	0.812	0.926	0.914

Her bir kuş türünün hangi türle daha çok karıştırıldığını görmek için karışıklık matrisi kullanılmıştır. Çizelge 3'te en iyi performansı sağlayan VGGish modelinin karışıklık matrisi görülmektedir. Karışıklık matrisinde hücrelerdeki değerler yüzdelik değerlerdir ve diagonalde bulunan değerler doğru sınıflandırma yüzdesini vermektedir. Buna göre Boyunçeviren kuşu (*Jy To*) için verilen bütün örnekler doğru tahmin edilmiştir. Bu türün diğerlerinden iyi ayrıştığı söylenebilir. Alakarga (*Ga Gl*) ve Karatavuk (*Tu Me*) kuşları için ise % 69 ile en düşük doğru tahmin oranı elde edilmiştir.

Çizelge 3. En yüksek doğruluk oranını sağlayan VGGish modelinin (işlenmiş 30 sn) karışıklık matrisi.

	An Tr	La Fl	Sy Co	Ga Gl	Lu Lu	Jy To	Sy Bo	Lu Me	Pa Ma	Ac Ar	Ac Pa	Sa Ru	Lo Na	Ph Co	Tr Tr	Ca Er	Cu Ca	Up Ep	Tu Me	Ph Ph	Lu Ar	St Hi
An Tr	**97.8**	0	0	1.1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1.1	0	0	0
La Fl	0	**89.7**	0	0	0	0	0	0	0	0	0	0	6.9	3.4	0	0	0	0	0	0	0	0
Sy Co	0	0	**76.2**	0	0	0	0	0	0	0	0	19.0	0	4.8	0	0	0	0	0	0	0	0
Ga Gl	3.4	0	3.4	**69.0**	0	0	3.4	0	0	0	0	0	0	0	13.8	0	0	0	3.4	0	0	3.4
Lu Lu	0	0	0	0	**86.7**	0	6.7	3.3	0	0	0	0	0	0	0	0	0	0	3.3	0	0	0
Jy To	0	0	0	0	0	**100**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Sy Bo	1.3	0	0	0	0	0	**96.1**	0	0	0	0	0	0	1.3	0	0	0	0	0	0	0	1.3
Lu Me	0	3.2	0	0	0	3.2	0	**74.2**	3.2	0	3.2	0	0	0	3.2	0	3.2	0	0	0	0	6.5
Pa Ma	0	0	3.3	0	0	0	0	6.7	**83.3**	0	0	3.3	0	0	0	0	0	0	0	3.3	0	0
Ac Ar	3.3	3.3	0	0	0	0	0	0	0	**76.7**	0	3.3	0	0	3.3	0	0	0	0	6.7	3.3	0
Ac Pa	0	0	0	0	1.7	0	0	0	0	0	**97.3**	0	0	0	0	0	0	0	0	0	1.0	0
Sa Ru	0	0	0	0	0	0	0	0	0	3.6	0	**85.7**	0	0	0	3.6	3.6	0	0	0	0	3.6
Lo Na	0	6.7	0	0	0	0	0	0	0	0	0	3.3	**86.7**	0	0	0	0	0	3.3	0	0	0
Ph Co	3.3	0	0	0	0	0	0	0	0	6.7	0	0	0	**83.3**	6.7	0	0	0	0	0	0	0
Tr Tr	0	0	0	3.4	0	0	0	0	0	3.4	0	0	0	3.4	**86.2**	0	0	0	3.4	0	0	0
Ca Er	1.2	0	2.1	0	0	0	0	0	0	0	0	1.2	0	0	0	**92.9**	0	0	0	1.3	0	1.2
Cu Ca	0	0	3.4	0	0	0	0	0	0	0	0	0	0	0	0	0	**89.7**	6.9	0	0	0	0
Up Ep	3.4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3.4	**82.8**	0	3.4	0	6.9
Tu Me	0	0	0	0	6.9	0	0	6.9	0	3.4	0	0	0	0	0	0	6.9	0	**69.0**	0	0	6.9
Ph Ph	0	2.4	0	0	0	0	0	0	0	0	0	5.8	0	0	0	2.4	0	0	0	**89.3**	0	0
Lu Ar	0	3.3	6.7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**86.7**	3.3
St Hi	2.0	1.4	0	0	0	0	0	0	0	0	0	0	0	0	0	1.4	0	0	0	0	0	**95.1**

Örneğin Alakarga için verilen örneklerin % 13.8'i Bayağı Çitkuşu (*Tr Tr*) olarak yanlış sınıflandırılmıştır. Bir başka ilginç örnek ise, Akgerdanlı Ötleğen (*Sy Co*) için verilen örneklerin % 19'u Çayır Taşkuşu (*Sa Ru*) olarak yanlış sınıflandırılmıştır. Bunların dışında Ağaç İncirkuşu (*An Tr*) ve Çalı Kamışçını (*Ac Pa*) sırasıyla % 97.8 ve % 97.3 oranlarıyla doğru tahmin edilmiştir.

SONUÇ

Bu çalışmada Iğdır'da bulunan Aras Kuş Cenneti'nde sıklıkla görülen 22 kuş türünün ses özellikleri bakımından sınıflandırılması ele alınmıştır. Sınıflandırma işlemini gerçekleştirmek için derin öğrenme mimari ve teknikleri kullanılmıştır. Evrişimsel sinir ağları özellikle görüntü, resim gibi uzamsal veriler üzerinde son yıllarda etkili olmuştur. Log mel-spektrogramların uzamsal özelliğinden dolayı evrişimsel sinir ağlarında çokça kullanılmaktadır. Bundan dolayı bu çalışmada da kuş sesleri kısa zamanlı Fourier dönüşümü yardımıyla zaman alanından zaman-frekans alanına çevrilmiş, daha sonra mel süzgeçlerinin logaritması alınarak log mel-spektrogramlara dönüştürülmüştür. Elde edilen bu öznitelikler derin öğrenme modellerinin giriş verisini oluşturmuştur.

Uzun kısa-dönemli bellek sinir ağları zamansal veriler üzerinde etkindir ve çok fazla çalışmaya konu olmuştur. Seslerin zamansal doğasının hesaba katılması için LSTM ağlarının kullanılması da yine son yılların popüler yaklaşımlarındandır. Bunun için bu çalışmada hem uzamsal hem de zamansal bilgiyi işleyen CNN-LSTM ortak mimarisi kullanılmıştır.

Öğrenme aktarımı temel olarak bir alanda (kaynak) edinilen bilginin başka bir alanda (hedef) kullanılmasını sağlayan bir tekniktir. Az verinin olduğu problemlerde, çok veriyle eğitilmiş kaynak modelinin bilgisi hedef modele transfer edilmektedir. Bu sayede az veriden kaynaklanan genelleştirme sorununun üstesinden gelmeye çalışılır. Kaynak ve hedef olarak adlandırılan alanlar farklı veya birbiriyle ilişkili olabilir. Bu çalışmada kaynak bir alanda önceden eğitilmiş VGGish ve YAMNet evrişimli modelleri kullanılmıştır. Ön eğitimli modellerden elde edilen yüksek seviyeli öznitelik vektörleriyle yeni bir model inşa edilmiştir. Bu şekilde oluşturulan modeller CNN ve CNN-LSTM modellerini doğruluk değeri (Acc) ve F_1 skoruna göre geride bırakmayı başarmıştır.

Ses kaydedicilerle toplanan ses kayıtları tarafımızca yazılan betiklerle temizlenmiştir. Bu sayede uzun sessizlik bölgelerinden ve gürültüden arındırılarak işlenmiş ses kayıtları elde edilmiştir. Bu sayede ses kayıtlarının boyutu düşürülerek ses kaydedicinin belleği etkin kullanılırken bu kayıtları işleyen bilgisayarın da işlem gücünden tasarruf edilmektedir. Bu avantajlarının yanında bu çalışmada işlenmiş ses kayıtlarının sınıflandırma başarımına etkisi incelenmiştir. Buna göre işlenmiş ses verileriyle eğitilip sınanan derin öğrenme modelleri ham ses verilerine göre daha yüksek performans göstermiştir.

Bu deneysel çalışmanın sonucu tatmin edicidir. Sonraki çalışmalarda, az örneğin bulunduğu durumlarda kullanılan diğer derin öğrenme tekniklerinin araştırılması planlanmaktadır. Bu sayede veri azlığı probleminin üstesinden gelinmesi arzu edilmektedir.

Çıkar Çatışması

Makale yazarları aralarında herhangi bir çıkar çatışması olmadığını beyan ederler.

Yazar Katkısı

Yazarlar makaleye eşit oranda katkı sağlamış olduklarını beyan eder.

KAYNAKLAR

Abadi, M, Agarwal A, Barham P, Brevdo E, Chen Z, Citro C, Corrado G. S, Davis A, Dean J, & Devin M. (2016). Tensorflow: Large-scale machine learning on heterogeneous distributed systems. arXiv 2016. *arXiv preprint arXiv:1603.04467*.

Aide T. M, Corrada-Bravo C, Campos-Cerqueira M, Milan C, Vega G, & Alvarez R. (2013). Real-time bioacoustics monitoring and automated species identification. *PeerJ, 2013*(1).

Akhtar N, & Mian A. (2018). Threat of Adversarial Attacks on Deep Learning in Computer Vision: A Survey. Içinde *IEEE Access* (C. 6, ss. 14410–14430). Institute of Electrical and Electronics Engineers Inc.

Bardeli R, Wolff D, Kurth F, Koch M, Tauchert K. H, & Frommolt K. H. (2010). Detecting bird sounds in a complex acoustic environment and application to bioacoustic monitoring. *Pattern Recognition Letters*, *31*(12), 1524–1534.

Barrowclough G. F, Cracraft J, Klicka J, & Zink R. M. (2016). How Many Kinds of Birds Are There and Why Does It Matter? *PLOS ONE*, *11*(11), 1–15.

Bayat S, & Işık G. (2020). Identification of Aras Birds with Convolutional Neural Networks. *4th International Symposium on Multidisciplinary Studies and Innovative Technologies, ISMSIT 2020 - Proceedings*.

Boersma P, & Weenink D. (2018). *Praat: doing phonetics by computer [Computer program]*. *Version 6.0.43*. retrieved 8 September 2018.

Chalmers C, Fergus P, Wich S, & Longmore S. (2021). *Modelling Animal Biodiversity Using Acoustic Monitoring and Deep Learning*.

Cho K, van Merriënboer B, Bahdanau D, & Bengio Y. (2014). On the properties of neural machine translation: Encoder–decoder approaches. *Proceedings of SSST 2014 - 8th Workshop on Syntax, Semantics and Structure in Statistical Translation*.

Chollet F. (2015). Keras: The Python Deep Learning library. *Keras.Io*.

de Jong N. H, & Wempe T. (2009). Praat script to detect syllable nuclei and measure speech rate automatically. *Behavior Research Methods*, *41*(2), 385–390.

Ferdiana R, Dicka W. F. & Boediman A. (2021). Cat sounds classification with convolutional neural network. *International Journal on Electrical Engineering and Informatics*.

Florentin J, Dutoit T, & Verlinden O. (2020). Detection and identification of European woodpeckers with deep convolutional neural networks. *Ecological Informatics*.

Gemmeke J. F, Ellis D. P. W, Freedman D, Jansen A, Lawrence W, Moore R. C, Plakal M, & Ritter M. (2017). Audio Set: An ontology and human-labeled dataset for audio events. *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*.

Grill T, & Schluter J. (2017). Two convolutional neural networks for bird detection in audio signals. *25th European Signal Processing Conference, EUSIPCO 2017*, *2017-Janua*, 1764–1768.

Guo Y, Xu M, Wu Z, Wu J, & Su B. (2019). Multi-Scale Convolutional Recurrent Neural Network with Ensemble Method for Weakly Labeled Sound Event Detection. *2019 8th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos, ACIIW 2019*.

He K, Zhang X, Ren S, & Sun J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*.

Hershey S, Chaudhuri S, Ellis D. P. W, Gemmeke J. F, Jansen A, Moore R. C, Plakal M, Platt D, Saurous R. A, Seybold B, Slaney M, Weiss R. J, & Wilson K. (2017). CNN architectures for large-scale audio classification. *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*.

Hershley S, et al.: Models for audioset: a large scale dataset of audio events (2016). https://github.com/tensorflow/models/tree/master/research/audioset/vggish

Howard A. G, Zhu M, Chen B, Kalenichenko D, Wang W, Weyand T, Andreetto M, & Adam H. (2017). MobileNets. *arXiv preprint arXiv:1704.04861*.

Işık G, & Artuner H. (2020). Turkish Dialect Recognition Using Acoustic and Phonotactic Features in Deep Learning Architectures. *Bilişim Teknolojileri Dergisi*, *13*, 207–216.

Jalal A, Salman A, Mian A, Shortis M, & Shafait F. (2020). Fish detection and species classification in underwater environments using deep learning with temporal information. *Ecological Informatics*.

Joly A, Goëau H, Glotin H, Spampinato C, Bonnet P, Vellinga W. P, Lombardo J. C, Planqué R, Palazzo S, & Müller H. (2017). LifeCLEF 2017 lab overview: Multimedia Species identification challenges. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*.

Joly A, Goëau H, Kahl S, Deneu B, Servajean M, Cole E, Picek L, Ruiz de Castañeda R, Bolon I, Durso A, Lorieul T, Botella C, Glotin H, Champ J, Eggel I, Vellinga W. P, Bonnet P, & Müller H. (2020). Overview of LifeCLEF 2020: A System-Oriented Evaluation of Automated Species Identification and Species Distribution Prediction. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*.

Jung D. H, Kim N. Y, Moon S. H, Kim H. S, Lee T. S, Yang J. S, Lee J. Y, Han X, & Park S. H. (2021). Classification of Vocalization Recordings of Laying Hens and Cattle Using Convolutional Neural Network Models. *Journal of Biosystems Engineering*.

Kahl S, Wilhelm-Stein T, Hussein H, Klinck H, Kowerko D, Ritter M, & Eibl M. (2017). Large-scale bird sound classification using convolutional neural networks. *CEUR Workshop Proceedings*.

Kingma D. P, & Ba J. L. (2015). Adam: A method for stochastic optimization. *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*.

Kuzeydoğa Derneği. 10 Ekim 2020 tarihinde, https://kuzeydoga.net/ adresinden erişildi.

LeBien J, Zhong M, Campos-Cerqueira M, Velev J. P, Dodhia R, Ferres J. L, & Aide T. M. (2020). A pipeline for identification of bird and frog species in tropical soundscape recordings using a convolutional neural network. *Ecological Informatics*, *59*.

Lezhenin I, Bogach N, & Pyshkin E. (2019). *Urban Sound Classification using Long Short-Term Memory Neural Network*.

M. Lasseck, "Acoustic bird detection with deep convolutional neuralnetworks," DCASE2018 Challenge, Tech. Rep., September 2018.

Mac Aodha O, Gibb R, Barlow K. E, Browning E, Firman M, Freeman R, Harder B, Kinsey L, Mead G. R, Newson S. E, Pandourski I, Parsons S, Russ J, Szodoray-Paradi A, Szodoray-Paradi F, Tilova E, Girolami M, Brostow G, & Jones K. E. (2018). Bat detective—Deep learning tools for bat acoustic signal detection. *PLOS Computational Biology*, *14*(3), e1005995.

Malfante M, Mars J. I, Dalla Mura M, & Gervaise C. (2018). Automatic fish sounds classification. *The Journal of the Acoustical Society of America*, *143*(5), 2834–2846.

Mathur M, Vasudev D, Sahoo S, Jain D, & Goel N. (2020). Crosspooled FishNet: transfer learning based fish species classification model. *Multimedia Tools and Applications*.

McFee B, Raffel C, Liang D, Ellis D, Mcvicar M, Battenberg E, & Nieto O. (2015). *librosa: Audio and Music Signal Analysis in Python*.

Nguyen H, Maclagan S. J, Nguyen T. D, Nguyen T, Flemons P, Andrews K, Ritchie E. G, & Phung D. (2017). Animal recognition and identification with deep convolutional neural networks for automated wildlife monitoring. *Proceedings - 2017 International Conference on Data Science and Advanced Analytics, DSAA 2017*, *2018-Janua*, 40–49.

Pacal I, & Karaboga D. (2021) A robust real-time deep learning based automatic polyp detection system, *Computers in Biology and Medicine*, Volume 134, 104519, ISSN 0010-4825

Pacal I, Karaman A, Karaboga D, Akay B, Basturk A, Nalbantoglu U, & Coskun S. (2022) An efficient real-time colonic polyp detection with YOLO algorithms trained by using negative samples and large datasets, *Computers in Biology and Medicine*, Volume 141, 105031, ISSN 0010-4825

Salamon J, Bello J. P, Farnsworth A, & Kelling S. (2017). Fusing shallow and deep learning for bioacoustic bird species classification. *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, 141–145.

Salamon J, Bello J. P, Farnsworth A, Robbins M, Keen S, Klinck H, & Kelling S. (2016). Towards the automatic classification of avian flight calls for bioacoustic monitoring. *PLoS ONE*, *11*(11).

Simonyan K, & Zisserman A. (2015). Very deep convolutional networks for large-scale image recognition. *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*.

Sprengel E, Jaggi M, Kilcher Y, & Hofmann T. (2016). *Audio Based Bird Species Identification Using Deep Learning Techniques*. In CEUR Workshop Proceedings (Vol. 1609, pp. 547–559). CEUR-WS.

Stowell D, Wood M, Stylianou Y, & Glotin H. (2016). Bird detection in audio: A survey and a challenge. *IEEE International Workshop on Machine Learning for Signal Processing, MLSP*.

Szegedy C, Vanhoucke V, Ioffe S, Shlens J, & Wojna Z. (2016). Rethinking the Inception Architecture for Computer Vision. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*.

Tolkova I, Chu B, Hedman M, Kahl S, & Klinck H. (2021). Parsing Birdsong with Deep Audio Embeddings. *CoRR, abs/2108.0*. https://arxiv.org/abs/2108.09203

Vidaña-Vila E, Navarro J, Alsina-Pagès R. M, & Ramírez Á. (2020). A two-stage approach to automatically detect and classify woodpecker (Fam. Picidae) sounds. *Applied Acoustics*, *166*.

xeno-canto. 10 Ekim 2020 tarihinde, *https://www.xeno-canto.org/* adresinden erişildi.

Xie J, Hu K, Zhu M, & Guo Y. (2020). Bioacoustic signal classification in continuous recordings: Syllable-segmentation vs sliding-window. *Expert Systems with Applications*, *152*.

Yamashita, R, Nishio, M, Do, RKG. et al. (2018) Convolutional neural networks: an overview and application in radiology. Insights Imaging 9, 611–629.

Young T, Hazarika D, Poria S, & Cambria E. (2018). Recent trends in deep learning based natural language processing [Review Article]. Içinde *IEEE Computational Intelligence Magazine* (C. 13, Sayı 3, ss. 55–75). Institute of Electrical and Electronics Engineers Inc.