

Yapay Zeka Desteği İle Televizyon Kontrolü

Esra ERGİN^{1*}, Umut ÖZKAYA²

¹Elektrik-Elektronik Mühendisliği / Mühendislik ve Doğa Bilimleri Fakültesi, Konya Teknik Üniversitesi, Türkiye
¹(f191202077@ktun.edu.tr)

²Elektrik-Elektronik Mühendisliği / Mühendislik ve Doğa Bilimleri Fakültesi, Konya Teknik Üniversitesi, Türkiye
²(uozkaya@ktun.edu.tr)

Özet – Bu yazıda, televizyon kanal isimlerinin sesli komutlarından oluşan ses sinyalleri toplanarak bir veri seti oluşturuldu. Derin öğrenme yapısı olan, kendi tasarladığımız CNN mimarisi kullanılarak ses sinyallerinin farklı sınıflandırılma işlemlerine tabi tutulması sağlandı. CNN mimarisi kullanılarak yapılan tüm sınıflandırma işlemlerinden önce ses sinyallerinin spektrogram analizi gerçekleştirilerek CNN mimarisinin işlemesine hazır hale getirildi. CNN mimarisi kullanılarak ilk olarak veri setinin yetişkin ve çocuk ses sinyalleri olarak sınıflandırılması işlemi yapılmıştır. Bu işlem sonucunda model %93.54 doğruluk oranında sınıflandırma yapmıştır. Daha sonra yetişkin ses sinyallerinin televizyon kanal isimlerine göre sınıflandırma yapılması için CNN ağı kullanılarak, %81.37 doğruluk oranı elde edilmiştir. Son işlemde CNN ağının kanal isimlerini hem yetişkin hem de çocuk olarak ayrı ayrı sınıflandırma işlemine tabi tutuldu. Sınıflandırma işlemi %74.46 doğruluk oranında tamamladığı gözlemlenmiştir.

Anahtar Kelimeler – Yapay Zeka, Derin Öğrenme, Sinyal İşleme, Konuşma Tanıma, Spektrogram Analizi

I. GİRİŞ

Yapay zeka ile son yıllarda ses işleme teknolojisi kullanılarak yapılan çözümler günlük yaşantımıza uygun ve yaygın şekilde dahil olmaya başlamıştır. Arama motorlarındaki sesli asistanlar, telefonlarımızdaki sesli asistanlar ile doğal şekilde konuşarak ihtiyaçlarımızı karşılamak, aracımızdaki cihazları sesli komutlar ile kontrol edebilmemiz günümüzün kaçınılmaz gerçekleri olmuştur.

NLP (Doğal Dil İşleme) tarihi, incelendiğinde ilk çalışmaların 1600'lü yıllarda başladığı görülmektedir. NLP teknolojisi, sözcükleri diller arasında ilişkilendirebilecek kodlar öneren René Descartes ve Gottfried Wilhelm Leibniz tarafından teoriye dökülmüştür. Buna rağmen NLP'nin ilk uygulama örnekleri yaklaşık üç asırlık bir periyot sonunda ortaya çıkmaya başlamıştır. Ses sinyali işleme ve konuşma tanıma konularında literatürde birçok çalışmanın mevcut olduğu görülmektedir.

Davis ve arkadaşları 1952 yılında yaptıkları çalışmalarında her bir hanenin sesli harflerin sırasında tahmin edilmiş biçimlendirici frekanslarını kullanarak bir konuşmacı için izole olan sayı tanıma sistemi gerçekleştirilmiştir[1].

Forgie ise 1959 yılında MIT Lincoln Laboratuvarında yapılan çalışmada, sesli harfler için spektral rezonanslar ölçümü yapılarak konuşmacıdan bağımsız 10 sesli harf tanıyıcısı geliştirmiştir[2]. 1970'li yıllara gelindiğinde ise ses sinyallerinin işlenmesi ve konuşma tanınmasında dönüm noktası niteliğinde olan birçok gelişme yaşandı. Ayrılmış kelime tanıma çalışmasının uygulanabilir bir teknoloji olarak Rusya'da, Velichko vd. [3]; Japonya'da, Sakoe vd. [4] ve ABD'nde, Itakura [5], gerçekleştirilen temel çalışmalar sonucunda ortaya çıkması idi. Rus bilim insanları ses tanıma içerisindeki örüntü tanıma fikrinin geliştirilmesinde yardımcı olurlarken, Japon bilim insanları ise dinamik programlama yöntemlerinin ses tanıma işlemlerinin geliştirilmesine yardımcı olmuşlardır. ABD' de yapılan araştırmalarda ise Doğrusal Öngörücü Kodlama (LPC)'nin ses ve konuşma tanıma için başarılı bir uygulama olduğu gösterildi. Paliwal ve ark. yaptıkları çalışmalarında ise LPC yöntemini kullanarak, ses sinyallerinden konuşma verilerini filtreleyebilmeyi amaçlamışlardır[6]. 1980'li yıllar daha detaylı incelendiğinde konuşma tanıma adına yapılan en büyük gelişmenin Sovyet bilim

insanlarının Dinamik Zaman Atlama (DTW) algoritmasını icat etmeleri olmuştur. 1990'lı yıllarda veri sınıflandırma ve regresyon problemlerinin çözümü amacıyla yeni bir yöntem olan Destek Vektör Makineleri (SVM) yöntemi ortaya atıldı[7] [8] [9]. Bu yöntem 2000'li yıllarda ses tanıma, konuşmacı tanıma ve doğrulama işlemleri için çeşitli araştırmalarda kullanıldı. 2004 yılında Bolat ve ark. yapmış oldukları çalışmada YSA (Yapay Sinir Ağı) kullanımını gerçekleştirerek konuşma ve müzik verilerinin değerlendirildiği çalışmada yaklaşık %91 tespit başarısı elde edilmiştir[10]. Vogt vd. 2007 yılında yaptıkları çalışmada, bir konuşmacıyı tanıma problemi üzerinde Gaussian Karma Modelleri (GMMs) için yeni bir yaklaşım önerdiler[11]. Bala ve ark. 2010 yılındaki çalışmalarında ise ses tanıma algoritmaları kullanılarak bir güvenlik sistemi tasarlamışlardır. Ses verilerinde öznelik çıkarımı ve eşleştirmesi ile %100'e yakın tanıma başarısı elde etmişlerdir[12]. Koç 2019 yılında, çalışmasında ses verilerini spektogram yöntemi kullanılarak görüntüye dönüştürmüştür. Böylelikle iki boyutlu derin öğrenme yapılarında sinyalleri işlemeye hazır hale getirmiş. Beş farklı CNN mimari modeli oluşturmuştur ve sınıflandırma işlemi gerçekleştirmiştir. CNN modellerinin performansını değerlendirmek için Softmax ve SVM olmak üzere iki farklı sınıflandırıcı kullanmış. Sınıflandırma sonucuna göre %95 doğruluk oranında sınıflandırma yapıldığı gözlemlenmiştir[13]. Acar ise çalışmasında, çevre seslerinden oluşan ESC-50 ve ESC-10 veri seti, çeşitli makine öğrenmesi, transfer öğrenme altyapısı ve farklı öznelik çıkarımı yöntemleri kullanarak sınıflandırma çalışmaları yapmıştır. Çalışmasında, K-NN, SVM, Rastgele Orman makine öğrenimi algoritmaları kullanmıştır. Klasik makine öğrenmesi algoritmaları ile elde edilen en iyi doğruluk puanları ESC-50 için %42,15 ve ESC-10 için %77,7'dir. ResNet algoritması ile ESC-50 için %68,95, ESC-10 için ise %87,25 doğruluk oranı elde etmiştir[14].

Bu çalışmada, 10 farklı televizyon kanal isimlerinden oluşan sesli komutların bulunduğu veri seti oluşturulmuştur. Veri setindeki ses sinyallerinin tasarlanan CNN modelinde kullanılabilmesi için spektogram analizi gerçekleştirilerek veriler işlenmek için hazır hale getirildi. Sınıflandırma İşlemi için Softmax ve SVM (Destek Vektör Makinesi) yöntemleri kullanılmıştır. CNN modeli

kullanılarak ilk olarak ses sinyalleri yetişkin ve çocuk olarak sınıflandırma işlemi yapılmıştır. İkinci kısımda CNN modeli televizyon kanal isimlerinin komutlarının yetişkin ses sinyallerine göre sınıflandırılması işlemi yapılmıştır. Son olarak aynı işlemi çocuk ses sinyallerinin de sınıflandırması işlemi için de kullanılarak sonlandırılmıştır.

Çalışma toplam dört kısımdan oluşmaktadır. 1.Bölüm giriş kısmı, literatür çalışmalarından bahsedilmiştir. 2. Bölüm materyal ve yöntem kısmıdır. 3.Bölüm çalışmada elde edilen bulgular anlatılmaktadır. 4.Bölüm çalışmadan elde edilen sonuçlar ve tartışma kısmı bulunmaktadır.

II. MATERYAL VE YÖNTEM

A. Veri Setinin Tanımlanması

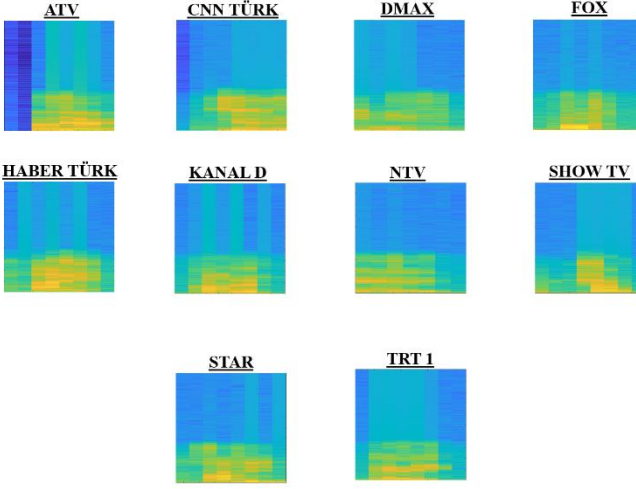
Çalışma kapsamında kullanılan veri seti, 10 farklı televizyon kanal isimlerinin sesli komutlarından oluşan 1253 ses kayıtlarından oluşmaktadır. Veri setindeki ses sinyalleri yetişkin ve çocuk ses kayıtlarından oluşmaktadır.

Tablo 1. Kanal isimlerinin veri dağılımı

Kanal İsimleri	Yetişkin Ses Verisi	Çocuk Ses Verisi
ATV	105	25
CNN TÜRK	105	25
DMAX	105	25
FOX	106	25
HABER TÜRK	106	25
KANAL D	105	25
NTV	105	25
SHOW TV	105	25
STAR	106	25
TRT 1	105	25

Veri setindeki tüm ses kayıtları gürültüsüz ortamda yapılmıştır. Tüm kayıtlar .wav uzantılı dosyalar olarak kaydedilmiştir.

Veri setinin CNN mimarisinde işlenebilmesi için hepsinin spektogram analizi gerçekleştirilerek hazır hale getirilmiştir. Örnek spektogram görüntüleri Şekil 1'de gösterilmiştir.



Şekil 1. Örnek spektrogram görüntüleri

B. Spektrogram Analizi

Spektrogram, zamana göre değişen bir sinyalin frekans spektrumuna göre değişiminin görsel olarak gösterimidir. Genellikle bir ısı haritası, renk veya parlaklığın değiştirilerek gösterilen farklı yoğunluğa sahip bir görüntü olarak düşünülebilir.

Hızlı Fourier Dönüşümü (HFD) algoritması, dijital bilgisayarda spektrumları hesaplamak ve spektral ekranları uygulamak için sıklıkla kullanılan bir algoritmadır. HFD'ye benzer olarak tanımlanan Ayrık Fourier Dönüşümünü (AFD) hesaplamak için kullanılan bir yöntemdir. (1)

$$F(k) = \sum_{n=0}^{N-1} f(nT)e^{-j\frac{2\pi}{N}nk} \quad (1)$$

Eş. 1'de sınırlı bir veri uzunluğundaki bir hesaplama sağlamıştır. Eş. 1'de $f(t)$ 'ye uygulanan bir zaman penceresi, yani $f(t)$ eşit genişliğe sahip NT dikdörtgen bir pencere ile çarpılır. Dikdörtgen pencerenin $w(nT)$ bazı yeni zaman pencereleriyle değiştirilmesiyle spektral şeklin değiştirilmesi mümkündür.

$$F_T(k) = \sum_{n=0}^{N-1} \omega(nT)f(nT + rMT)e^{-j\frac{2\pi}{N}nk} \quad (2)$$

$F_T(k)$ sayı kümesi, $t = rMT$ 'de başlayan ve $t = rMT + (N-1)T$ ile biten analog zaman fonksiyonunun bir bölümünün Ayrık Fourier dönüşümünün bir hesaplamasını temsil eder. Bu sayede $t = rMT + (N-1)T$ zamanında bir filtre bankası çıkışına karşılık gelir. Spektral analizin bir filtre bankası uygulamasında, $w(nT)$ zaman penceresi, filtrelerin her birinin dürtü yanıtının düşük geçişli prototipine karşılık gelir. Bu analizde filtre bankasındaki

filtrelerin her birinin spektral şeklinin yaklaşık olarak aynı olduğu kabul edilir.

C. Derin Öğrenme

Büyük miktarlarda etiketlenmiş eğitim verilerinden özellik saptama yapabilen sistemler oluşturmak için ileri teknoloji, çok seviyeli "derin" sinir ağların kullanılması olan Derin Öğrenme, önemli derecede yatırım ve araştırmanın yapıldığı bir alandır. Günümüzde derin öğrenme kullanım alanları gün geçtikçe artmaktadır. Akıllı telefonlar, sosyal medya, arama motoru eşleştirme ve savunma sanayisi gibi birçok farklı alanda kullanımına örnek olarak verilebilir. Farklı türlerde yapay zeka sinir ağları vardır, Evrişimli Sinir Ağları (CNN) bunlardan biridir.

D. Evrişimli Sinir Ağları

Evrişimli sinir ağları (CNN), yaygın olarak görüntülerin analiz edilmesinde kullanılan bir derin öğrenme yapısıdır. İki boyutlu görüntü verileriyle çalışmak için tasarlanmış ancak tek boyutlu ve üç boyutlu verilerin analizi için de kullanılmaktadır.

Evrişimli sinir ağı (CNN) kendi içinde temel olarak iki bölümden oluşmaktadır. İlk bölüm bir veya daha fazla sayıda konvolüsyonel, havuzlama ve ReLu katmanlarından oluşan özellik çıkarımının yapıldığı kısımdır. İkinci bölüm ise tam bağlı katmanlardan oluşan sınıflandırma işleminin yapıldığı kısımdır.

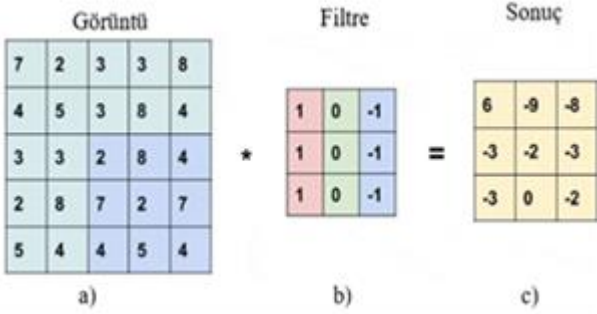
Görüntü sınıflandırma problemlerinde zor olan görüntü piksellerinin içerisinde özellik çıkarımıdır. CNN, zor olan bu süreci otomatik olarak ilk bölümde işleterek özellik haritalarını elde eder. Sonrasında ise elde edilen özellikler tam bağlı katmanlara iletilerek çok katmanlı algılayıcılar sayesinde belirli sayı değerleri elde edilerek görüntünün hangi sınıfa ait olduğu ile alakalı tahminler yapılır.



Şekil 2. Evrişimli Sinir Ağlarının (CNN) Genel Yapısı

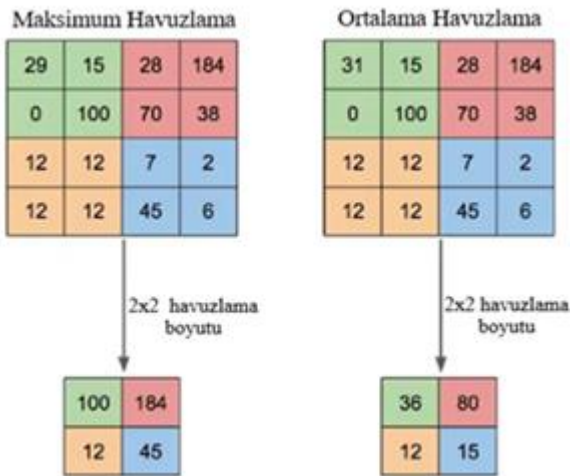
Konvolüsyon katmanı, girdi verilerinin filtre edildiği katmandır. Bu katmanda verinin orijinal boyutlarından daha küçük bir filtre gezerek verinin özellik haritası çıkarılır. CNN algoritmalarında öğrenilen parametreler filtrelerdeki değerlerdir.

Model sürekli olarak bu değerleri günceller ve özellikleri daha iyi tespit etmeye başlar. Genellikle, birden çok özelliği tespit etmek için birden fazla filtre kullanılır. Yani bir CNN ağında birden fazla konvolüsyon katmanı bulunur. Konvolüsyon katmanından elde edilen özelliklerin seçilmesi için havuzlama katmanına geçilir.



Şekil 3. a) 5x5 görüntü matrisi b) 3x3 filtre matrisi c) 3x3 sonuç matrisi

Havuzlama katmanının asıl amacı görüntü katmanlarını ve boyutu azaltmaktır. CNN yapılarında genellikle kullanılan 2 tip havuzlama tekniği vardır. Bunlardan biri maksimum havuzlama diğeri ise ortalama havuzlama tekniğidir. Havuzlama katmanında da bir filtre görsel üzerinde gezer. Seçilen havuzlama tekniğini uygular. Maksimum havuzlama tekniğinde filtrenin kapsadığı alandaki en büyük değeri, ortalama havuzlama tekniğinde ise filtrenin kapsadığı alandaki değerlerin ortalamasını alır. Bu sayede boyut azalır ve önemli özellikler çıkarılmış olur.

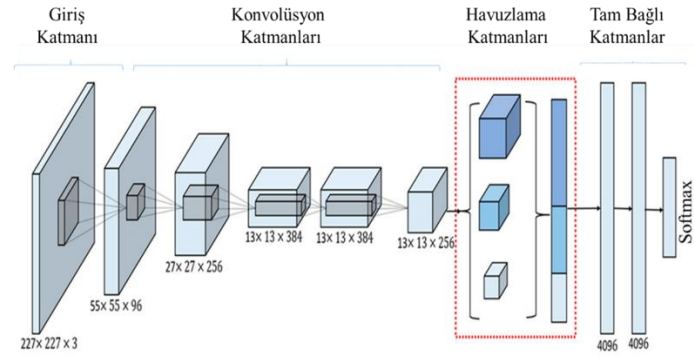


Şekil 4. Maksimum Havuzlama ve Ortalama Havuzlama Örneği

Aktivasyon fonksiyonları, doğrusal olmayan fonksiyonlardan oluşur. Bu fonksiyonlar özellik

haritalarına uygulanarak doğrusallık sorunu giderilmiş olur[15]. Kaçınma katmanı, çıkarılan özelliklerin gauss dağılımı incelenerek elde edilen önemli özellikleri bir sonraki katmana aktarılır.

Tam bağlı katmanlar içerisinde ağırlık ve bias değerleri ile klasik bir yapay sinir ağına benzetilmektedir. Sınıflama işlemi için genel olarak Softmax yapısı kullanılır. Softmax, olasılıksal sınıf tahminleyicisi olarak işlem yapar[15].



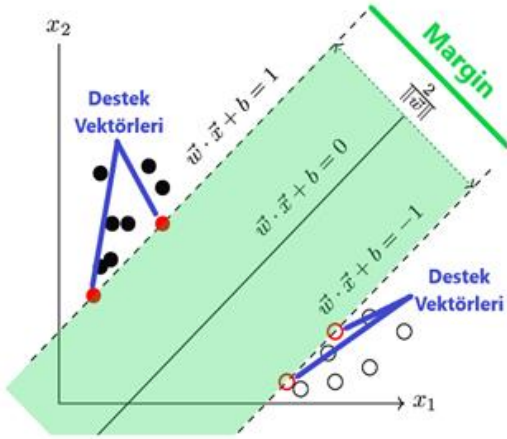
Şekil 5. Örnek CNN Modeli

E. Destek Vektör Makineleri

Destek Vektör Makinesi (SVM), sınıflandırma veya regresyon problemleri için kullanılabilen denetimli bir makine öğrenmesi algoritmasıdır. Şekil 6. incelenirse, siyahlar ve beyazlar olmak üzere iki sınıf vardır. Sınıflandırma problemlerindeki amaç, sıradaki verinin hangi sınıfta yer alacağını tayin etmektir. Sınıflandırmayı yapabilmek için iki sınıfı ayıran bir doğru çizilir. Bu doğrunun ± 1 değeri arasında kalan bölgeye Margin adı verilir. Margin bölgesi ne kadar geniş ise iki veya daha fazla sınıfın iyi bir sonuçla ayrıştırılması yapılmış olur.

$$\hat{y} = \begin{cases} 0, & w^T \cdot x + b < 0, \\ 1, & w^T \cdot x + b \geq 0, \end{cases} \quad (3)$$

Eş. 3'te w değeri ağırlık vektörü, x değeri girdi vektörü, b değeri sapmayı temsil etmektedir. Eğer çıkan sonuç sıfırdan küçükse beyaz noktalara yakın olacaktır. Çıkan sonuç sıfıra eşit veya büyükse bu durumda siyah noktalara yakın olacaktır.



Şekil 6. Destek Vektör Makinesi (SVM) Sınıflaması

F. Önerilen Yöntem

Veri setinin ilk olarak yetişkin ve çocuk ses sinyalleri olarak sınıflandırılması işlemi, yetişkin ses sinyallerini kullanarak televizyon kanal isimlerinin sınıflandırılması ve son olarak çocuk ses sinyalleri de eklenerek televizyon kanal isimlerinin hem yetişkin hem de çocukların ses sinyallerine göre sınıflandırılma işlemi için kendi tasarladığımız bir CNN mimarisini kullandık. Sınıflandırma işlemi için tasarlanan CNN ağındaki Softmax katmanı dışında Destek Vektör Makinesi algoritması kullanılarak daha doğru sınıflandırma işlemi yapmak amaçlanmaktadır.

III. BULGULAR

Çalışma kapsamında toplamda 1253 ses sinyalinden oluşan veri seti yetişkin-çocuk, yetişkin kanal ve yetişkin-çocuk kanal isimleri sınıflandırması işleminin gerçekleştirilmesi için oluşturulan CNN mimarisinin katmanları Çizelge 1'de görülmektedir.

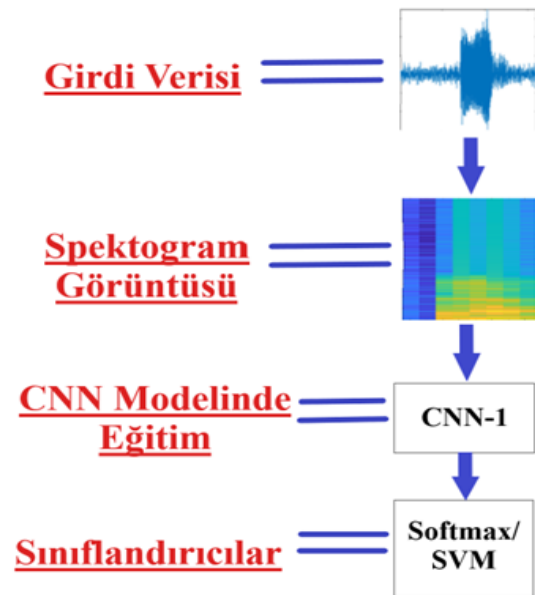
Çizelge 1. Önerilen CNN Modelinin Katman Yapıları

CNN Modeli	CNN-1
Konvolüsyon Katmanı	3x3x12
	3x3x24
	3x3x48
	3x3x48
	3x3x48
Havuzlama Katmanı	3x2
	13x13
ReLU Katmanı	5 Adet
Dropout Katmanı	0.2

CNN-1 modelinin katman yapıları incelendiğinde birden fazla konvolüsyon katmanı içerdiği görülmektedir. Modelde bulunan konvolüsyon katmanları özdeş olup 3x3 boyutu filtreler ile sırasıyla 12, 24, 48 adet filtrelerden oluşan 5 katman mevcuttur. 0.2 indirgeme oranına sahip bir dropout katmanı da modelde bulunmaktadır. Havuzlama katmanlarında ise iki farklı havuzlama işlemi yapılmaktadır. İlk havuzlama katmanı 3x2 boyutunda, ikinci havuzlama katmanı ise 13x13 boyutunda maksimum havuzlama işlemi gerçekleştirmektedir. CNN-1 modelinde aktivasyon fonksiyonu olarak ReLu aktivasyon fonksiyonu 5 katmanda kullanıldı. Ayrıca tüm CNN mimarilerinde mevcut olan toplu normalleştirme katmanı ve tam bağlı katmanlar da CNN-1 modelinde kullanılmıştır.

Veri setindeki tüm ses sinyallerinin spektrogram analizi gerçekleştirilerek bir data haline getirilip tasarlanan CNN mimarisinde işlenmesi için hazır hale getirildi.

Bütün eğitim işlemleri MATLAB 2021a programında 8GB RAM ile 2.40GHz'de Intel Core i5-10200 H işlemci üzerinde gerçekleştirilmiştir. Analiz kapsamında TFI verilerinin %75'i eğitim aşamasında kullanılmak için, %25'i ise test aşamasında kullanılmak üzere ayrılmıştır.



Şekil 7. Önerilen Yöntem

Tablo 2’de yetişkin-çocuk, yetişkin kanal ve yetişkin-çocuk kanal isimlerinin sınıflandırılması işlemlerinin sonuçları gözlemlenmektedir.

Tablo 2. Sınıflandırma Sonuçları

Sınıflandırma	Doğruluk Oranı
Yetişkin-Çocuk Sınıflandırması	%93.54
Yetişkin Kanal İsmi Sınıflandırması	%81.37
Yetişkin-Çocuk Kanal İsmi Sınıflandırması	%74.46

Tablo 2 incelendiğinde “Yetişkin-Çocuk Sınıflandırma” işlemini CNN-1 modeli, %93.54 doğruluk oranında gerçekleştirmiştir. “Yetişkin Kanal İsimleri Sınıflandırma” işlemini CNN-1 modeli, %81.37 doğruluk oranında gerçekleştirmiştir. “Yetişkin-Çocuk Kanal İsmi Sınıflandırması” işleminde ise CNN-1 modeli %74.46 doğruluk oranında tamamlamıştır.

“Yetişkin Kanal İsmi Sınıflandırma” işleminde “TRT-1” kanalı %96.2 doğruluk oranında en yüksek doğru sınıflandırma oranına sahiptir. En düşük doğrulukta sınıflandırma oranı ise %70.4 ile “STAR” kanal ismi olmuştur.

“Yetişkin-Çocuk Kanal İsmi Sınıflandırma” işleminde yetişkinlerde en yüksek doğru sınıflandırma oranı %96.2 doğruluk oranında “TRT 1” kanal ismi olurken, çocuklarda ise %83.3 doğruluk oranında “SHOW TV” kanal ismi olmuştur.

IV. SONUÇLAR VE TARTIŞMA

Bu çalışmada veri setindeki 10 farklı televizyon kanal isimlerinin yetişkin ve çocuklara ait sesli komutlarından oluşan ses sinyallerinin spektrogram analizi yapılarak tasarlanan CNN-1 modelinin işlemlerine hazır hale getirdik. Kendi tasarladığımız CNN-1 modelini kullanarak veri setindeki sınıf ve ses sinyallerini arttırarak modelin sınıflandırma başarısı test edilerek asıl amacımız olan kanal isimleri sınıflandırma işlemini gerçekleştirdik.

Gelecekte gerçekleştirilmesi planlanan çalışmalarda, veri setindeki ses sinyallerinin

arttırılması ve özellikle çocuk ses sinyalleri toplanması hususunda altı yaş ve üzeri çocukların kullanılarak telaffuz probleminden kaynaklanan problemlerin giderilmesi sonucunda CNN modelinin daha yüksek doğrulukta sınıflandırma işlemi gerçekleştirilmesi ile proje geliştirilebilir.

TEŞEKKÜR

Bu çalışmada değerli bilgileri ve faydalı tartışmaları için Dr. Öğr. Üyesi Umut ÖZKAYA’ya çok teşekkür ederim.

KAYNAKLAR

- [1] Davis, K., Biddulph, R., & Balashek, S. (1952). Automatic Recognition of Spoken Digits. *Journal of the Acoustic Society of America*, 24 (6):637-642.
- [2] Forgie, J., & Forgie, C. (1959). Results Obtained From a Vowel Recognition Computer Program. *Journal of the Acoustic Society of America*, 31 (11):1480-1489.
- [3] Velichko, V., & Zagoruyko, N. (1970). Automatic Recognition of 200 Words. *Intenational Journal of Man-Machine Studies*, 2:223.
- [4] Sakoe, H., & Chiba, S. (1978). Dynamic Programming Algorithm Optimization for Spoken Word Recognition. *IEEE Transactions on Acoustic, Speech and Signal Processing ASSP-26* (s. (1):43-49). IEEE.
- [5] Itakura, F. (1975). Minimum Prediction Applied to Speech Recognition. *IEEE Transactions on Acoustics, Speech, Signal Processing, ASSP-23*, (1):67-72.
- [6] Paliwal, K., Agarwal, A., & Sinha, S. (1982). A Modification over Sakoe and Chiba's Dynamic Time Warping Algorithm for Isolated Word Recognition. *Signal Processing*, 4 (4):329-333.
- [7] Vapnik, V. (1995). The Nature of Statistical Learning Theory. *Springer-Verlag Publishing*, (s. 187). New York.
- [8] Vapnik, V. (1998a). Statistical Learning Theory. *John Wiley Publishing*, (s. 768). New York.
- [9] Vapnik, V. (1998b). The Support Vector Method of Function Estimation, In *Nonlinear Modelling Advanced Black Box Techinques. Kluwer Academic Publishers*, (s. 55-85). Boston.
- [10] Bolat, B., & Yıldırım, T. (2004). Aktif Öğrenme ile Konuşma Müzik Sınıflandırma. *Akıllı Sistemlerde Yenilikler ve Uygulamaları Sempozyumu, ASYU-INISTA*.
- [11] Vorgt, R., & Sridharan, S. (2008). Explicit Modelling of Session Variability for Speaker Verification. *Computer Speech and Language*, 22 (1):17-38.
- [12] Bala, A., Kumar, A., & Birla, N. (2010). Voice Command Recognition System Based on MFCC and DTW. *Intenational Journal of Engineering Science and Technology*, (12):7335-7342.
- [13] Koç, F. (2019). *Spektrogram Tekniği Kullanılarak Derin Öğrenme Yöntemleri ile Ses Tanıma*. Van: Van Yüzüncü Yıl Üniversitesi Fen Bilimleri Enstitüsü.

- [14] Acar, M. (2021). *Makine Öğrenmesi ve Transfer Öğrenimi Kullanılarak Ses Tanıma*. İstanbul: Kadir Has Üniversitesi Lisansüstü Eğitim Yüksekokulu, Yüksek Lisans Tezi .
- [15] U. Özkaya ve L. Seyfi, "Yere Nüfuz Eden Radar B Tarama Görüntülerinin Az Parametreye Sahip Konvolüsyonel Sinir Ağı İle Değerlendirilmesi", *Geomatik*, cilt 6, hayır. 2, s. 84-92, Ağustos 2021.
- [16] Aydemir, E. (2021). *Uygulamalar İle PYTHON ve Yapay Zeka*. Ankara: NOBEL Akademik Yayıncılık.
- [17] Eray, O. (2018). *Destek Vektör Makineleri ile Ses Tanıma Uygulaması* . Denizli: Pamukkale Üniversitesi Fen Bilimleri Enstitüsü .
- [18] Korçuklu, B. (2020). *Derin Öğrenme Tabanlı Konuşma Sistem Tasarımı*. Bursa: Bursa Uludağ Üniversitesi Fen Bilimleri Enstitüsü.
- [19] Öztürk, B., & Çakar, T. (2007). *Gerçek Zamanlı Ses Tanıma*. İstanbul: İstanbul Üniversitesi Mühendislik Fakültesi, Lisans Tezi.
- [20] Price, J., & Eydgahi, A. (23-28 July 2006). Design of Matlab Based Automatic Speaker Recognition Systems. *9th International Conference on Engineering Education*.
- [21] Reddy, D. (1967). Computer Recognition of Connected Speech . *Journal of the Acoustic Society of America*, 42:329-347.
- [22] Vintsyuk, T. (1968). Speech Discrimination by Dynamic Programming. *Kibernetika*, 4 (2):81-88.
- [23] Fezari, M., & Bousbia-Salah, M. (2006). A Voice Command System for Autonomous Robots Guidance . *9th International Workshop on Advanced Motion Control* (s. 261-265). İstanbul: IEEE.
- [24] Kımanuke, U. (2019). *Derin Sinir Ağlarına Dayalı Türkçe Ses Tanıma*. Kocaeli: Kocaeli Üniversitesi Doğa ve Uygulamalı Bilimler Yüksekokulu.
- [25] Krüger, S., Schafföner, M., Katz, M., Andelic, E., & Wendemuth, A. (2006). Mixture of Support Vector Machines for HMM Based Speech Recognition . *The 18th International Conference on Pattern Recognition*, (s. 326-329). Magdeburg.