ANKARA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

YÜKSEK LİSANS TEZİ

TEK MİKROFON İLE SES KAYNAĞI UZAKLIĞI TAHMİNİ

Hulusi EROĞLU

ELEKTRİK-ELEKTRONİK MÜHENDİSLİĞİ ANABİLİM DALI

ANKARA
2018

Her hakkı saklıdır

TEZ ONAYI

Hulusi EROĞLU tarafından hazırlanan "**Tek Mikrofon ile Ses Kaynağı Uzaklığı Tahmini**" adlı tez çalışması 10/07/2018 tarihinde aşağıdaki jüri tarafından oy birliği ile Ankara Üniversitesi Fen Bilimleri Enstitüsü Elektrik-Elektronik Mühendisliği Anabilim Dalı'nda **YÜKSEK LİSANS TEZİ** olarak kabul edilmiştir.

Danışman: Prof. Dr. Hakkı Gökhan İLK
Ankara Üniversitesi Elektrik-Elektronik Mühendisliği Anabilim Dalı

Jüri Üyeleri:

Başkan : Prof. Dr. Özgül SALOR DURNA
Gazi Üniversitesi Elektrik-Elektronik Mühendisliği Anabilim Dalı

Üye : Prof. Dr. Hakkı Gökhan İLK
Ankara Üniversitesi Elektrik-Elektronik Mühendisliği Anabilim Dalı

Üye : Dr. Öğretim Üyesi Aykut KALAYCIOĞLU
Ankara Üniversitesi Elektrik-Elektronik Mühendisliği Anabilim Dalı

Yukarıdaki sonucu onaylarım.

Prof. Dr. Atila YETİŞEMİYEN
Enstitü Müdürü

ETİK

Ankara Üniversitesi Fen Bilimleri Enstitüsü tez yazım kurallarına uygun olarak hazırladığım bu tez içindeki bütün bilgilerin doğru ve tam olduğunu, bilgilerin üretilmesi aşamasında bilimsel etiğe uygun davrandığımı, yararlandığım bütün kaynakları atıf yaparak belirttiğimi beyan ederim.

10/07/2018

Hulusi EROĞLU

TEŞEKKÜR

Bu çalışmam boyunca yardımlarını ve desteklerini esirgemeyen çok kıymetli danışman hocam Prof. Dr. Hakkı Gökhan İLK'e (Ankara Üniversitesi Elektrik-Elektronik Mühendisliği Anabilim Dalı) teşekkür ederim.

Ayrıca eğitim hayatım boyunca maddi ve manevi destekleriyle hep yanımda olan aileme teşekkürlerimi iletiyorum.

Hulusi EROĞLU
Ankara, Temmuz 2018

ÖZET

Yüksek Lisans Tezi

TEK MİKROFON İLE SES KAYNAĞI UZAKLIĞI TAHMİNİ

Hulusi EROĞLU

Ankara Üniversitesi
Fen Bilimleri Enstitüsü
Elektrik-Elektronik Mühendisliği Anabilim Dalı

Danışman: Prof. Dr. Hakkı Gökhan İLK

Bu tez çalışmasında, geleneksel yöntemler (öznitelik çıkarımı ve sınıflandırma) ve derin ağlar ile öğrenme olmak üzere iki ayrı yöntem, tek mikrofon ile ses kaynağı mesafesi tahmini problemine çözüm olarak sunulmuştur.

Veri seti olarak kullanılmak üzere öncelikle 1 metre, 2 metre, 3 metre ve 4 metrelik uzaklıklardan ses kayıtları toplanmıştır. Problem ilk olarak geleneksel yöntemler ile üç adımda çözülmüştür. İlk adımda ses aktivatörü kullanılarak sadece konuşma içeren ses sinyali alınmıştır ve Hanning pencereleme uygulanmıştır. İkinci adımda ise bu sinyalden öznitelikler çıkartılmıştır. Son adımda ise çıkartılan bu öznitelikler k-nn (k-nearest neighbour, k-en yakın komşuluk) sınıflandırıcı ile sınıflandırılıp mesafeye karar verilmiştir. İkinci yöntem olarak ise derin sinir ağları kullanılarak mesafeye karar verilmiştir. Derin sinir ağ yapısı, 1 giriş, 7 konvolüsyönel katman ve 1 çıkış katmanından oluşturulmuştur. Geleneksel yönteme göre derin sinir ağları ile elde edilen başarımda % 14'lük bir artış gözlenmiştir.

Temmuz 2018, 36 sayfa

Anahtar Kelimeler: Öznitelik Çıkarımı, Sınıflandırma, Derin Sinir Ağları

ABSTRACT

Master Thesis

AUDIO SOURCE DISTANCE ESTIMATION VIA SINGLE MICROPHONE

Hulusi EROĞLU

Ankara University
Graduate School of Natural and Applied Sciences
Department of Electric and Electrical Engineering

Supervisor: Prof. Dr. Hakkı Gökhan İLK

In this thesis, traditional methods (feature extraction and classification) and deep neural networks, are presented as a solution to the sound source distance estimation problem via single microphone.

As dataset, 1 meter, 2 meter, 3 meter and 4 meter recordings were used. The problem is solved in three steps by conventional methods. In the first step, a VAD (voice activity detector) and Hanning windowing are applied to speech signal. In the second step, features are extracted from this signal. In the last step, these extracted features are classified with k-nn (k-nearest neighborhood) classifier. As a second method, it has been decided to use deep neural networks. The deep network structure is composed of 1 input, 7 convolutional layers and 1 output layer. Compared with the conventional method, deep networks increased the performance of the overall system by 14 %.

July 2018, 36 pages

Key Words: Feature Extraction, Classification, Deep Neural Networks

İÇİNDEKİLER

TEZ ONAY SAYFASI
ETİK……i
TEŞEKKÜR……ii
ÖZET ……iii
ABSTRACT ……iv
KISALTMALAR DİZİNİ……vii
ŞEKİLLER DİZİNİ……viii
ÇİZELGELER DİZİNİ……ix
1. GİRİŞ……1
2. KURAMSAL TEMELLER……3
2.1 Sesin Oluşumu……3
2.2 Doğrusal Öngörüm Katsayıları……3
2.2.1 LPC yönteminin uygulama adımları……4
2.3 İstatistiksel Kavramlar……5
2.4 Sınıflandırma……8
2.4.1 K-en yakın komşuluk algoritması……9
2.4.2 Derin sinir ağları……10
2.4.2.1 Konvolüsyonel ve genişletilmiş konvolüsyonel sinir ağları ……12
2.4.2.2 Hiperparametreler……13
2.4.2.3 Aktivasyon fonksiyonları……13
2.4.2.4 Eğitim……16
2.4.2.5 İnce ayar(Fine tuning)……17
2.4.2.6 Test……17
3. MATERYAL VE YÖNTEM……18
3.1 Öznitelik Çıkarımı ve Sınıflandırma……19
3.1.1 Öznitelikler……20
3.2 Derin Sinir Ağlarının Uygulanması……21
3.2.1 Model oluşturma……21
3.2.2 Model eğitimi……22
3.2.3 Model testi……22

4. ARAŞTIRMA BULGULARI……...…….....23
4.1 Öznitelik Çıkarımı ve Sınıflandırma Yöntemi Bulguları……23
4.1.1 Öznitelik seçiminin etkisi…….…23
4.2 Derin Sinir Ağları Yöntemi Bulguları……...26
4.2.1 Katman türü seçiminin etkisi……..26
4.2.2 Epoch sayısının etkisi……...26
4.2.3 Aktivasyon fonksiyonu etkisi……...28
4.2.4 GPU kullanımı etkisi……28
4.2.5 Veri miktarının etkisi……...29
5. SONUÇ……...……30
5.1 Değerlendirme……..30
5.2 Öneriler…….33
KAYNAKLAR……...34
ÖZGEÇMİŞ……...36

KISALTMALAR DİZİNİ

VAD	Ses Aktivasyon Detektörü (Voice Activity Detector)
LPC	Doğrusal Öngörüm Katsayıları (Linear Predictive Coefficients)
PDF	Olasılık Yoğunluk Fonksiyonu (Probabilty Density Function)
DNN	Derin Sinir Ağları (Deep Neural Network)
KNN	K-en Yakın Komşuluk (K-Nearest Neighbour)
CNN	Konvolüsyonel Sinir Ağları (Convolutional Neural Network)
RNN	Yinelemeli Sinir Ağları (Recurrent Neural Network)
CPU	Merkezi İşlem Birimi (Central Processing Unit)
GPU	Grafiksel İşlem Birimi (Graphical Processing Unit)
RELU	Düzeltilmiş Doğrusal Birim (Rectified Linear Unit)
SGD	Stokastik Gradyan Düşümü (Sthocastic Gradient Descent)

ŞEKİLLER DİZİNİ

Şekil 2.1 LPC uygulama adımları…….4

Şekil 2.2 Negatif çarpıklık değerine göre dağılım eğrisi……..6

Şekil 2.3 Pozitif çarpıklık değerine göre dağılım eğrisi……...7

Şekil 2.4 Negatif basıklık değerine göre dağılım eğrisi……..…….…...7

Şekil 2.5 Pozitif basıklık değerine göre dağılım eğrisi……..…….....…..8

Şekil 2.6 k = 4 için komşuluk uzaklıkları gösterimi……..…….……...10

Şekil 2.7 Sigmoid fonksiyonu…….……14

Şekil 2.8 Tanh fonksiyonu…….…….…14

Şekil 2.9 ReLU fonksiyonu…….……...15

Şekil 2.10 Leakly ReLU fonksiyonu…….……16

Şekil 3.1 Kayıt ortamından alınan 1. görüntü.....18

Şekil 3.2 Kayıt ortamından alınan 2. görüntü…….…….…….……19

Şekil 3.3 Ses kaynağı mesafe tespiti sistemi blok diyagramı…….…….……...19

Şekil 4.1 1 metre, 2 metre, 3 metre ve 4 metre için hata oranı özniteliği grafiği…..…..23

Şekil 4.2 1 metre, 2 metre, 3 metre ve 4 metre için savrukluk oranı özniteliği grafiği..24

Şekil 4.3 1 metre, 2 metre, 3 metre ve 4 metre için basıklık oranı özniteliği grafiği.....24

Şekil 4.4 Epoch Etkisinin Grafiksel Gösterimi……...28

ÇİZELGELER DİZİNİ

Çizelge 4.1 Hata Oranı Özniteliği ve k-nn sınıflandırıcı karışıklık matrisi…….....25

Çizelge 4.2 Basıklık Özniteliği ve k-nn sınıflandırıcı karışıklık matrisi……25

Çizelge 4.3 Savrukluk Özniteliği ve k-nn sınıflandırıcı karışıklık matrisi…….....25

Çizelge 4.4 Üç Öznitelik ve k-nn sınıflandırma sonucu karışıklık matrisi…..….....26

Çizelge 4.5 10 Epoch ile eğitim sonucu karışıklık matrisi……..27

Çizelge 4.6 100 Epoch ile eğitim sonucu karışıklık matrisi……27

Çizelge 4.7 500 Epoch ile eğitim sonucu karışıklık matrisi……27

Çizelge 4.8 1000 Epoch ile eğitim sonucu karışıklık matrisi……..……27

1. GİRİŞ

Ses sinyali işleme, günümüzde yaygın olarak kullanılan ses kaydetme sistemleri ve güçlü donanıma sahip işlem gücü yüksek bilgisayarlar sayesinde çalışılan güncel bir konudur. Ses sinyali işleme alanında başta konuşmacı tanıma, duygu analizi, cinsiyet, yaş tahmini ve ses kaynağı uzaklığı tespiti olmak üzere birçok problem çözülmeye çalışılmaktadır.

Ses sinyali, sadece kelime ve bu kelimelerin anlamlarını içermenin yanı sıra içerdiği istatistiksel bilgiler ile farklı özellikler de barındırmaktadır. Bu istatistiksel özellikler de bize konuşmacıdan bağımsız olarak ses kaynağı hakkında çeşitli bilgiler sunmaktadır.

Ses kaynağı uzaklığı tespiti ise tek mikrofon veya mikrofon dizileri kullanılarak çözülmeye çalışılan oldukça zor bir problemdir. Ses kaynağı uzaklığı tespiti yöntemlerinin başta konuşma tanıma (Omologo vd. 1998), akıllı işitme cihazları (Hamacher vd. 2005), iyileştirilmiş gerçeklik sesi (Härmä vd. 2004) ve eller serbest haberleşme sistemleri (Härmä 2009) olmak üzere geniş bir uygulama alanı vardır. Savunma sanayisinde akıllı silah sistemleriyle entegre olarak uygulanmaktadır.

Literatürde, ses kaynağı uzaklığı tespiti yaygın olarak mikrofon dizileri ile çözülmüştür. Mikrofon dizileri ile farklı gecikmelerde kaydedilen ses işlenilerek konum ve mesafe bilgisi elde edilmektedir (Fan vd. 2010).

Mikrofon dizileri ile problem irdelendiğinde, mikrofonlara gelen ses sinyallerindeki gecikmeler, faz farkları gibi özellikler kullanılarak çözüm yapılmaktadır (Kwok vd. 2005). Tek mikrofon ile böyle bir imkan bulunmadığı için bu çözüm mikrofon dizilerine nispeten daha zordur. Çözüm için sinyalin istatistiksel özelliklerinden öznitelik çıkarmalı (Georganti vd. 2011) ya da daha güncel bir yaklaşım olan derin sinir ağlarından faydalanılmalıdır.

Bu tez kapsamında, ses kaynağı uzaklığı tespiti probleminin çözümü tek mikrofon ile ele alınacağından ve literatürde tek mikrofon ile çözümü çok az olduğu için bu çalışma son derece önemlidir. Yapılan bu çalışmada problemin çözümüne tek mikrofonla

yaklaşıldığı için mikrofon dizilerinde olduğu gibi mikrofonlar arasındaki mesafeden dolayı oluşan faz farkı gibi bir özellik kullanılamayacaktır. Ses sinyalinin mesafesi hakkında bilgi verecek özelliklerini kullanmak gerekmektedir. Geleneksel yöntem diye tanımladığımız öznitelik çıkarma işlemi, ses sinyalinin istatistiksel özelliklerini kullanarak mesafe bilgisi veren özellikleri çıkarma işlemidir. Ses sinyalindeki bu istatistiksel özellikler kullanılarak elde edilen öznitelikler, herhangi bir sınıflandırıcı kullanılarak sınıflandırılır ve mesafe tahmini yapılmış olur. Bu çalışma kapsamında ilk olarak geleneksel yöntem uygulanmış Bölüm 4'te bulgular sunulacaktır. Geleneksel yöntem ile elde edilen başarı değeri nasıl artırılır diye düşünülmüş ve sinyal işleme alanında son yıllarda yaygın olarak kullanılan derin sinir ağları ile de problemin çözümü gerçekleştirilmiştir. Derin sinir ağları ile başarı değeri daha da artırılmış ve Bölüm 4'te bulgular sunulacaktır.

Tezin bölümleri şu şekildedir: Bölüm 2'de LPC (doğrusal öngörüm katsayıları) uygulaması, bazı istatistiksel terimlerin, kullanılan sınıflandırıcının ve derin sinir ağlarının kuramsal tanımlamaları anlatılacaktır. Bölüm 3'te ise problemin ilk olarak geleneksel yöntemler (öznitelik çıkarımı + sınıflandırma) ile çözümü yani ses sinyalinin bazı istatistiksel özelliklerinin kullanılarak öznitelik çıkarımını ve çıkarılan bu özniteliklerin k-en yakın komşuluk algoritması kullanılarak sınıflandırılması ve mesafe tahmini yapılması detaylandırılacaktır. İkinci yöntem olan derin sinir ağları kullanılarak ses kaynağı uzaklığı tespiti probleminin çözümü de Bölüm 3'te detaylandırılacaktır. Bölüm 4'te ise geleneksel yöntem ve derin sinir ağları ile çözülen problemin bulguları paylaşılacaktır. Son bölüm olan Bölüm 5'te ise sonuçlar ve öneriler sunulacaktır.

2. KURAMSAL TEMELLER

2.1 Sesin Oluşumu

Ses dalgası, akciğerlerden gelen havanın ses tellerini titreştirmesiyle oluşan akustik bir basınç dalgasıdır. Ses yolu ise boğaz ve ağız boşluğundan oluşmaktadır. Dolayısıyla ses yolu, gırtlak çıkışından başlayıp, dudaklarda son bulmaktadır (Selen 1979).

Ses fizyolojisinde bazı önemli tanımlamalar aşağıda açıklanmaktadır.

Ses Dalgası: Genliği sabittir. İçerisinde farklı frekanslar bulundurabilen sinüzoidal formda bir dalgadır.

Şiddet: Ses dalgalarının genliğidir. Ses şiddeti desibel (dB) cinsinden ölçülür.

Frekans: Saniyedeki titreşim sayısıdır. İnsan kulağı 20-20 kHz arasındaki sesleri algılayabilir. Günlük insan konuşma seslerinin frekansı ise 500-2000 Hz aralığındadır.

2.2 Doğrusal Öngörüm Katsayıları

Doğrusal öngörüm katsayıları (LPC), n = 1,2,3,4…N olmak üzere, N adet örnekten oluşan konuşma sinyalinin, kendisinden önceki p adet örnekle ifade edilmesidir. LPC'nin analiz denklemi, aşağıdaki eşitlik 2.1'de sunulmuştur.

$$s\ n\ =\ \ \sum_{i=1}^{p} a_i s\ n-i \qquad (2.1)$$

Eşitlik 2.1'de görüldüğü üzere, $s\ n$ lpc katsayılarıyla öngörülen konuşma sinyali, a_i katsayıları ise LPC katsayıları olarak tanımlanmaktadır.

Orijinal konuşma sinyali ile LPC katsayıları ile öngörülen sinyal arasındaki fark hata sinyalidir. Bu fark öngörüm hatasıdır ve LPC tortusu olarak da tanımlanmaktadır. Eşitlik 2.2'de hata sinyali matematiksel ifadesi sunulmuştur.

$$e\ n\ = s\ n\ - s\ n\ = s\ n\ - \ \sum_{i=1}^{p} a_i s\ n-i \tag{2.2}$$

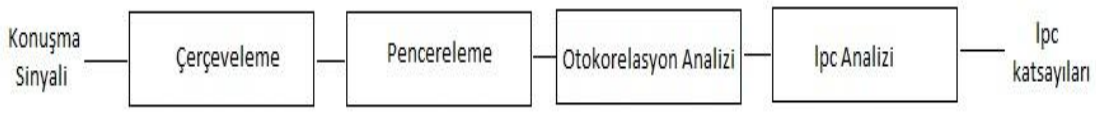
LPC katsayıları ile sentezlenmiş konuşma sinyali ise eşitlik 2.3'te verilmiştir.

$$s\ n\ = \ \sum_{i=1}^{p} a_i s\ n-i\ + e\ n \tag{2.3}$$

2.2.1 LPC yönteminin uygulama adımları

LPC yöntemi dört adımda uygulanmaktadır. İlk adımda konuşma sinyalinin belirli bir uzunlukta parçası alınır (çerçeveleme). İkinci adımda konuşma sinyaline pencereleme uygulanır. Üçüncü adımda konuşma sinyalinin otokorelasyon analizi yapılır. Son adımda ise LPC analizi yapılarak LPC katsayıları elde edilir.

Şekil 2.1'de LPC yönteminin uygulanmasının blok diyagramı sunulmuştur.



Şekil 2.1 LPC uygulama adımları

Çerçeveleme: Konuşma sinyali 20ms-30ms aralığında durağan olduğu kabul edildiğinden, çerçeve uzunluğu bu aralıkta olmalıdır.

Pencereleme: Konuşma sinyaline bu aşamada Hanning, Gauss vb. pencere gibi fonksiyonlar uygulanmaktadır.

Otokorelasyon analizi: LPC analizine geçişte en yaygın olarak kullanılan yöntemdir. eşitlik 2.4'te formülasyonu sunulmuştur.

$S\ n$ konuşma sinyali olmak üzere,

p, LPC analizinin derecesi olmak üzere,

$$r\ m\ = \ \sum_{n=0}^{N-1-m} S\ n\ S\ n+m \qquad\qquad m = 0,1 \dots \dots p \tag{2.4}$$

LPC analizi: Otokorelasyon analizinden sonra LPC analizine geçiş için Durbin metodu kullanılabilir. Durbin metodu, LPC katsayıları ile ilgili hata farkını yinelemeli olarak bulmaktadır. Bulunan bu katsayılar LPC katsayıları olarak adlandırılmaktadır (Rabiner vd. 1993).

2.3 İstatistiksel Kavramlar

Ortalama değer: Bir veride bulunan tüm değerleri tek bir değer ile ifade etmek için kullanılabilir. Aritmetik, geometrik ve harmonik ortalama gibi farklı kullanım türleri vardır.

Varyans: Veride bulunan değerlerin ortalama değer etrafında dağılımına bakar. Verideki değerlerin her birinin ortalama değerden olan farkının karelerinin aritmetik ortalaması olarak tanımlanır. Eşitlik 2.5'te formülasyonu sunulmuştur.

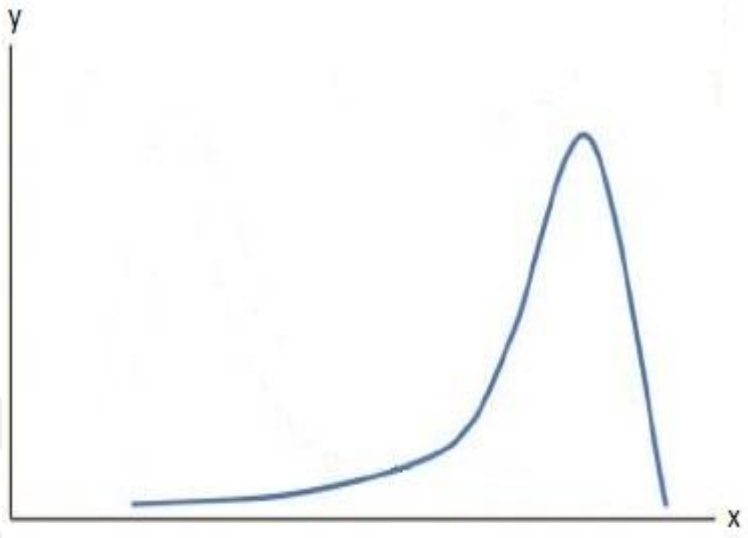
$$\sigma^2 = \frac{1}{N} \sum_{i=1}^{N} (x_i - X_A)^2 \qquad X_A\text{: ortalama değer} \tag{2.5}$$

Standart Sapma: Ortalama değer etrafındaki dağılmanın sayısal ölçütüdür. Varyansın karekökünе eşittir.

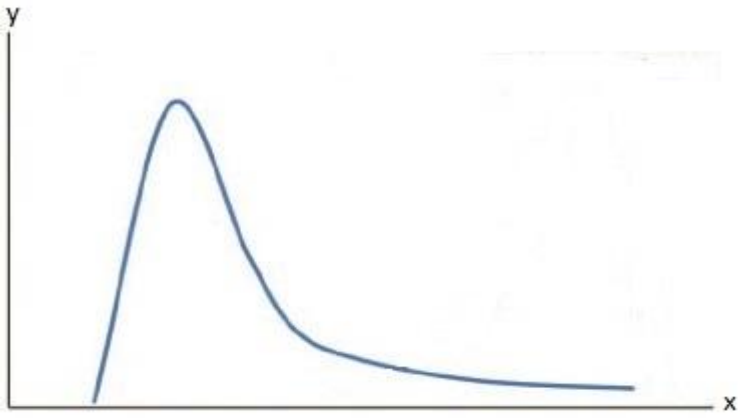
Çarpıklık: Normal dağılım grafiğinde, dağılım eğrisinin simetrisinin bozulma değerine çarpıklık denilmektedir. Çarpıklık katsayısı değeri negatif ise, dağılım eğrisi sağ tarafa, pozitif ise dağılım eğrisi sol tarafa yığılma gösterecektir. Şekil 2.2'de negatif çarpıklık değerine göre dağılım eğrisi, şekil 2.3'te ise pozitif çarpıklık değerine göre dağılım eğrisi sunulmuştur.

x rasgele değişken,

y olasılık yoğunluk fonksiyonu (pdf) olmak üzere,



Şekil 2.2 Negatif çarpıklık değerine göre dağılım eğrisi

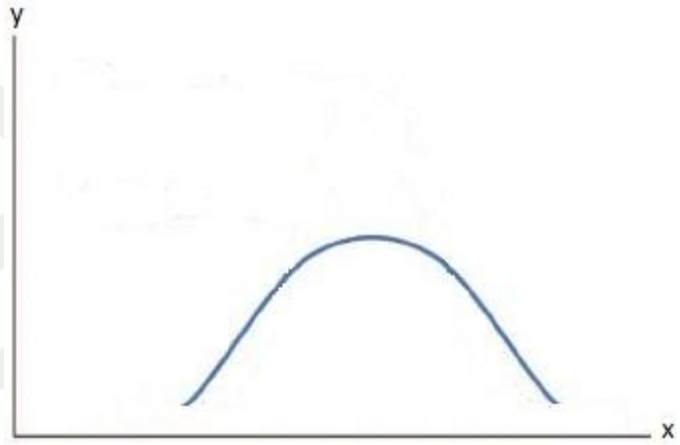


Şekil 2.3 Pozitif çarpıklık değerine göre dağılım eğrisi

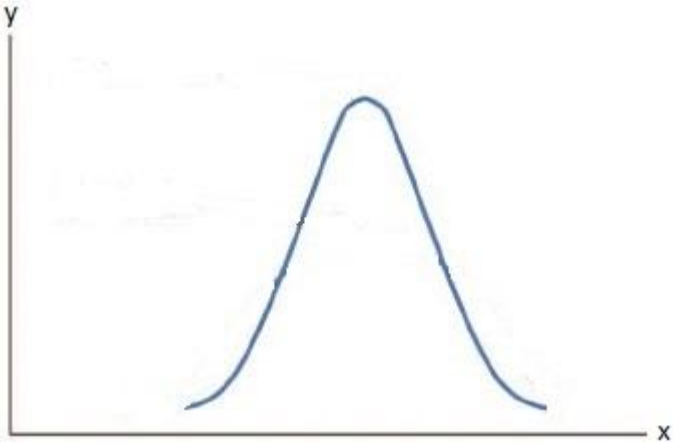
Basıklık: Normal dağılım grafiğinde, dağılım eğrisinin sivrilik ya da basıklık değerine basıklık denilmektedir. Basıklık katsayısı değeri negatif ise, dağılım eğrisi basık, pozitif ise dağılım eğrisi sivri bir şekil alacaktır. Şekil 2.4'te negatif basıklık değerine göre dağılım eğrisi, şekil 2.5'te ise pozitif basıklık değerine göre dağılım eğrisi sunulmuştur.

x rasgele değişken,

y olasılık yoğunluk fonksiyonu (pdf) olmak üzere,



Şekil 2.4 Negatif basıklık değerine göre dağılım eğrisi



Şekil 2.5 Pozitif basıklık değerine göre dağılım eğrisi

2.4 Sınıflandırma

Sınıflandırma problemi için literatürde eğiticili ve eğiticisiz olmak üzere farklı çözümler sunulmuştur. Bu çözümlerden klasik olan ve en çok kullanılanları k-nn (k-en yakın komşuluk) algoritması ve destek vektör makineleridir (SVM-support vector machine).

K-nn algoritması eğiticili bir algoritmadır.

K-nn sınıfını belirleyeceğimiz veri elemanının, kendimizin belirlediği *k* parametresine göre *k* adet komşuya en yakın olması şartının sağlanması prensibine göre çalışmaktadır (Duda vd. 2000). Bu çalışma kapsamında k-nn algoritması kullanıldığından dolayı k-nn algoritması bölüm 2.4.1'de detaylandırılmıştır.

SVM algoritması ise hiper düzlem ile veriyi ikiye ayırma prensibine göre çalışmaktadır. Hiper düzleme en yakın öğrenme verileri destek vektörleri olarak adlandırılmaktadır (Duda vd. 2000). Bu çalışma kapsamında kullanılmayacağı için detaylandırma yapılmamıştır.

Sınıflandırma problemine diğer bir yaklaşım ise yapay sinir ağlarıdır. Yapay sinir ağları ise 1940'lardan bu yana çalışılan bir konudur. Yapay sinir ağları, insan beyninin modellenmesini esas alan bir sistemdir.

Günümüzde artan veri miktarı ve güçlü işlem kapasitesine sahip bilgisayarlar ve grafik ekran kartlarında işlem yapabilme teknolojisi sayesinde derin öğrenme ile yapay sinir ağları tekrar önem kazanmıştır. 2012 yılında (Krizhevsky vd. 2012) görüntü sınıflandırma alanında ImageNet veri seti ile oldukça yüksek bir başarı oranı elde etmiştir. ImageNet veri seti yaklaşık 1,4 milyon görüntü içermektedir. Derin ağlar günümüzde sinyal işleme alanında oldukça yaygın olarak kullanılmaktadır.

2.4.1 K-en yakın komşuluk algoritması

K-en yakın komşuluk algoritması, örnek kümedeki gözlemlerin her birinin, sonradan belirlenen bir gözlem değerine olan uzaklıkların hesaplanması ve en küçük uzaklığa sahip k sayıda gözlemin seçilmesi esasına dayanmaktadır (Duda vd. 2000).

Uzaklıkların hesaplanmasında genellikle Öklid uzaklık yöntemi seçilir.

a_x, x eksenindeki değer,

a_y, y eksenindeki değer olmak üzere;

p verideki eleman sayısı olmak üzere;

D, Öklid uzaklık değeri Eşitlik 2.6'da sunulmuştur.

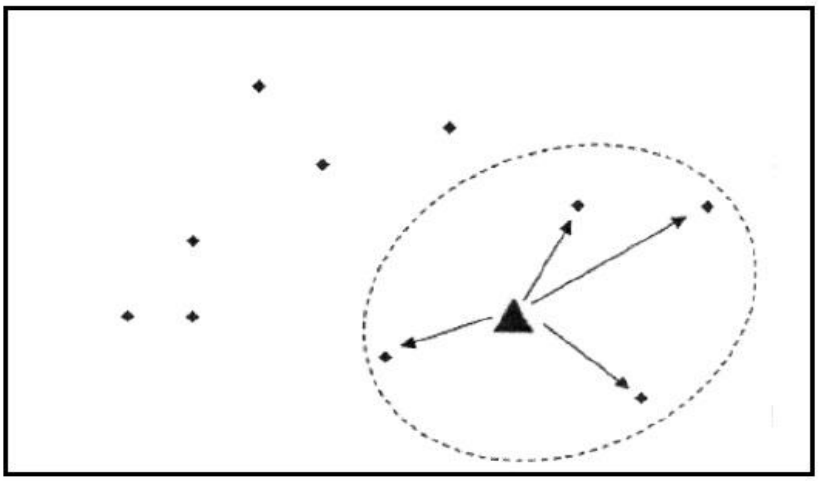
k=1,2,3….p olmak üzere,

$$D\ x,y\ = \sqrt{\sum_{k=1}^{p} a_{xk} - a_{yk}^{\ 2}} \tag{2.6}$$

K-en yakın komşuluk algoritmasının çalışma esası şu şekildedir,

1) K parametresi belirlenir. Bu parametre verilen noktaya en yakın k adet komşunun sayısıdır (Duda vd. 2000).
2) Söz konusu nokta ile diğer tüm noktalar arasındaki uzaklıklar tek tek hesaplanır (Duda vd. 2000).
3) Hesaplanan uzaklıklara göre satırlar sıralanır ve bunların arasından en küçük k tanesi seçilir (Duda vd. 2000).
4) Seçilen satırların hangi kategoriye ait oldukları belirlenir ve en çok tekrar eden kategori değeri seçilir (Duda vd. 2000).
5) Seçilen kategori, tahmin edilmesi beklenen gözlem değerinin kategorisi olarak kabul edilir (Duda vd. 2000).

Şekil 2.6'da belirli bir noktaya yakın k = 4 komşuluk için grafiksel gösterim yapılmıştır.



Şekil 2.6 k = 4 için komşuluk uzaklıkları gösterimi

2.4.2 Derin sinir ağları

Derin öğrenme, öznitelik çıkarımı ve dönüştürme işlemleri için doğrusal olmayan birçok katmandan oluşur. Katman sayısına göre ağın derinliği oluşmaktadır. Katmanlar sıralı bir şekilde oluşturulur ve her bir katmanın çıkışı bir sonraki katmanın girişi olmaktadır. Derin öğrenmede birden fazla öznitelik ve varsa verinin izlediği doku öğrenilir. Derin öğrenme 2012 yılında oluşturulan AlexNet (Krizhevsky vd. 2012) ağ modeli ile görüntü sınıflandırma alanında oldukça yüksek başarı elde etmiştir. Krizhevsky vd. AlexNet'i 8 katmandan oluşturmuştur. 2014 yılında ise VGG (Simonyan and Zisserman 2014) (16 ve 19 katman) ve GoogleNet (Szegedy vd. 2014) (22 katman) daha derin bir yapıyla oluşturulmuşlardır. GoogleNet AlexNet'e göre daha yüksek bir başarım oranı elde etmiştir. 2015 yılında ise daha derin bir yapıyla ResNet (He vd. 2015) (152 katman) ortaya çıkmıştır. ResNet diğerlerine göre oldukça yüksek başarı elde etmiştir.

Derin sinir ağları, klasik sinir ağlarından nöron ağırlık katsayıları, hiperparametreler ve her bir katmanda filtre öğrenmesi yönleriyle farklılığa ve zenginliğe sahiptir (Chollet 2018).

Derin sinir ağları veri miktarına oldukça bağımlıdır. Gerçek bir öğrenme sisteminde modelin kapasitesinin yüksek, eğitim verisinin de yeterli miktarda olması gerekmektedir. Model kapasitesinin yüksek eğitim verisinin az olduğu durum ezberleme (overfitting) problemini ortaya çıkarmaktadır. Bu durumu ortadan kaldırmak için ya veri miktarını artırmak gerek ya da "dropout (yok sayma)" denilen yöntemi uygulamak gereklidir.

Dropout: Katmanda bulunan bazı nöronların yok sayılmasıyla ezberlenmenin engellenmesi sağlanmaktadır (Chollet 2018). Katmanda bulunan nöronların kaçta kaçının yok sayılacağı dropout fonksiyonuna parametre olarak verilebilmektedir.

Model kapasitesinin düşük veya derinliği yeterli olmayan modellerde ise öğrenememe (underfitting) problemi ortaya çıkmaktadır. Çözümü için modelin hiperparametreleri veya derinliği değiştirilmelidir. Hiperparametreler bir sonraki bölümde açıklanacaktır.

Derin sinir ağlarının gerçekleştirilmesinde ise farklı kütüphaneler ve programlama dilleri kullanılmaktadır. Bu kütüphanelerden alt yapı olarak en çok kullanılanları tensorflow, theano ve caffe'dir (Gulli ve Pal 2017). Kodlama tarafında ise keras, tflearn kütüphaneleri yaygın olarak tercih edilmektedir. Programlama dillerinde ise python ve Matlab tercih edilmektedir. Oluşturulacak derin öğrenme sistemi için tensorflow alt yapısında keras kütüphanesi ile python dilinde kolaylıkla kodlanabilir.

Günümüzde hızlı işlem yapma kapasitesine ulaşmış ekran kartı teknolojisi bu alanda da yaygın olarak kullanılmaktadır. Özellikler NVIDIA bu alanda kütüphaneler oluşturmuş ve kullanıma sunmuştur.

Oluşturulan derin öğrenme sisteminde GPU teknolojisini destekleyen tensorflow sürümü kurulduğunda ve sistemsel özellikleri yeterli bir NVIDIA ekran kartı var ise eğitim ve test zamanında ciddi miktarda bir kazanç sağlanmaktadır.

Derin sinir ağlarının en çok kullanılan modelleri konvolüsyonel sinir ağları (CNN), yinelemeli sinir ağları (RNN) ve genişletilmiş konvolüsyonel sinir ağlarıdır (Dilated CNN) (Chollet 2018). Bu çalışma kapsamında yinelemeli sinir ağları

kullanılmadığından detaylandırılmayacak, konvolüsyonel ve genişletilmiş konvolüsyonel sinir ağları beraber kullanıldığından Bölüm 2.4.2.1'de detaylandırılacaktır.

2.4.2.1 Konvolüsyonel ve genişletilmiş konvolüsyonel sinir ağları

Konvolüsyonel sinir ağları, derin sinir ağları modellerinden en yaygın kullanılan ve en popüler olan derin öğrenme mimarisidir. Görüntü tanıma ve doğal dil işleme alanlarında oldukça başarılı sonuçlar vermektedir. Giriş verisindeki kıymetli özellikleri kendisi tespit edebilmektedir. Konvolüsyonel bir sinir ağı tasarladığımızda, giriş katmanının boyutu, konvolüsyonel katmanda kaç adet filtre öğrenmesi gerektiği ve bu filtrenin boyutları katman parametreleri ile belirlenmektedir. Uygulanan konvolüsyon işlemi, ilgili nöronun kendi uyarı bölgesindeki uyaranlara verdiği tepki olarak düşünülebilir (Fukushima 1980). Konvolüsyonel sinir ağı modeli, konvolüsyonel katman, genişletilmiş konvolüsyonel katman ve örnek azaltma (pooling) gibi katmanlardan oluşabilir. Konvolüsyonel sinir ağlarında, bir katmanın çıkışı diğer bir katmanın girişi olmaktadır ve sıralı bir sistem olarak çalışmaktadır.

Genişletilmiş konvolüsyonel sinir ağlarında ise sıra esaslı sistem daha farklı uygulanmaktadır. Genişletilmiş konvolüsyonel katmana bir önceki katmanın çıkışının yanı sıra belirlenen parametre ile kendisinden önceki katmanlar da giriş olarak verilmektedir (Gulli ve Pal 2017). Bu sayede genişletilmiş konvolüsyon katmanı gelen veriye daha geniş bir pencereden bakmış, daha zengin bilgiyle beslenmiş olmaktadır. Kullanılan hiperparametreler bir sonraki bölümde detaylandırılacaktır.

2.4.2.2 Hiperparametreler

Kaydırma (Stride): Konvolüsyon filtresinin veriye kaç adımda bir uygulanacağını gösteren parametredir (Chollet 2018). Adım aralığı arttıkça doğal olarak öğrendiği öznitelikler azalacaktır.

Ekleme (Padding): Oluşturulan sinir ağının herhangi bir katmanındaki giriş için küçük boyutta verileri giriş katmanı boyutuna çıkarmak için kullanılan yöntemdir (Chollet 2018).

Örnek Azaltma (Pooling): Oluşturulan sinir ağının herhangi bir katmanındaki giriş için büyük boyutta verileri giriş katmanı boyutuna indirmek için kullanılan yöntemdir(Chollet 2018).

MaxPooling en yaygın kullanılan türüdür. MaxPooling çalışma yöntemi şu şekildedir; örneğin, 2x2 boyutunda bir verinin hangi elemanı en büyükse o tutulur diğerleri atılır böylece 1x1'lik bir veri oluşturulmuş olur.

Genişleme Oranı (Dilated_rate): Oluşturulan sinir ağının sadece bir önceki katmandan değil daha önceki katmanlardan da giriş almasını sağlamaktadır (Chollet 2018). Bu sayede katmanın algı alanı artırılmış olup veriye daha büyük çerçeveden bakılmış olur.

2.4.2.3 Aktivasyon fonksiyonları

Nöron çıkışı için farklı çeşitte aktivasyon fonksiyonları vardır. Çalışılan alana göre, aktivasyon fonksiyonun da çalışma aralığı dikkate alınarak hangi aktivasyon fonksiyonun kullanılacağına karar verilmelidir.

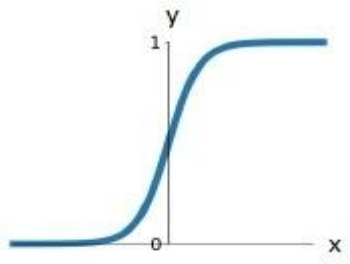
Sigmoid: Çalışma aralığı (0,1)'dir. Grafiksel gösterimi şekil 2.7'de, matematiksel ifadesi ise eşitlik 2.7'de sunulmuştur.

x = 0,1,2…N olmak üzere,

$$\sigma\ x\ = \frac{1}{1+ e^{-x}} \tag{2.7}$$

x giriş değişkeni,

y çıkış fonksiyonunun aldığı değer olmak üzere,



Şekil 2.7 Sigmoid fonksiyonu

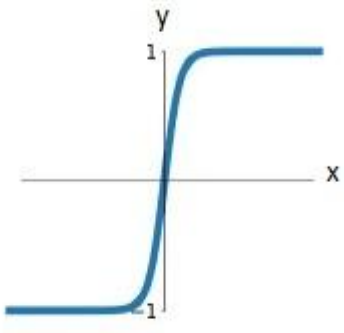
Tanh: Çalışma aralığı (-1,1)'dir. Matematiksel ifadesi tanh(x)'dir. Şekil 2.8'de grafiksel gösterimi sunulmuştur.

x = 0,1,2…N olmak üzere,

$$y = \tanh(x) \tag{2.8}$$

x giriş değişkeni,

y çıkış fonksiyonunun aldığı değer olmak üzere,



Şekil 2.8 Tanh fonksiyonu

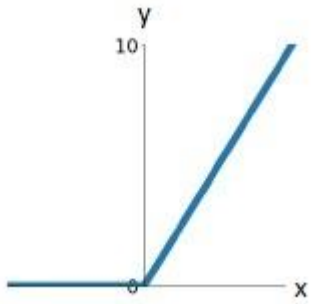
ReLU(Rectified Linear Unit): Çalışma aralığı (0,∞)'dur. Matematiksel ifadesi max(0,x)'dir. Şekil 2.9'da grafiksel gösterimi sunulmuştur.

x = 0,1,2….N olmak üzere

$$y = \max(0,x) \tag{2.9}$$

x giriş değişkeni,

y çıkış fonksiyonunun aldığı değer olmak üzere,



Şekil 2.9 ReLU fonksiyonu

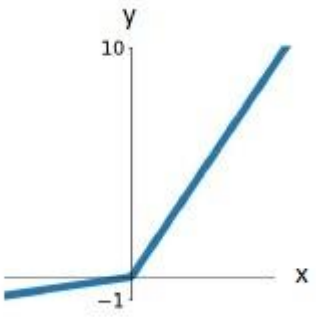
Leakly ReLU: Çalışma aralığı(-∞,∞)'dur. Matematiksel ifadesi max(0.1x,x)'dir. Şekil 2.10'da grafiksel gösterimi sunulmuştur.

x = 0,1,2….N olmak üzere

$$y = \max(0.1x,x) \tag{2.10}$$

x giriş değişkeni,

y çıkış fonksiyonunun aldığı değer olmak üzere,



Şekil 2.10 Leakly ReLU fonksiyonu

2.4.2.4 Eğitim

Derin sinir ağlarında model kapasitesinin, ağ derinliğinin ve kullanılan veri miktarının yeterliliğinin önemini daha önceki bölümlerde vurgulanmıştır. Eğitim aşamasını ilgilendiren bazı terimler aşağıda açıklanmıştır.

Veri Miktarı (Batch_size): Eğitim veya test aşamasında işlenecek verinin ne kadar büyüklükte alınmasını istediğimiz parametredir.

Öğrenme Oranı (Learning_rate): Öğrenmede ulaşılmak istenilen hedef noktadır.

Yineleme (Epoch): Eğitim esnasında hedef öğrenme noktasına ulaşılana kadar verinin işleneceği yineleme (iterasyon) sayısıdır.

Softmax: Çıkış katmanına konularak sınıflandırıcı görevi yapmaktadır (Chollet 2018). En büyük avantajı çıkış katmanında olasılıksal değerler üretmesidir.

Eğitim aşamasında veri miktarından kaynaklanan zaman problemi ise GPU destekli bir altyapı kullanarak çözülmektedir.

2.4.2.5 İnce ayar(Fine tuning)

Farklı verilerle eğitilmiş bir modelin, belirli bir amaç için daha özel verilerle tekrar eğitilmesi ince ayar olarak tanımlanmaktadır. Bu yöntem sayesinde özel verilerin az olduğu sistemlerde başarılı öğrenme sonuçlarına ulaşılmaktadır.

2.4.2.6 Test

Eğitim ve ince ayar yapılmış modeller kullanıma hazırdır. Eğitim aşamasında olduğu gibi veri miktarından kaynaklanan zaman problemi var ise GPU destekli bir altyapı kullanılarak bu problem çözülmektedir.

3. MATERYAL VE YÖNTEM

Bu çalışmada ses kaynağı uzaklığı tespiti probleminin çözümü iki farklı uygulamayla yapılmıştır. İlk çözüm olarak Matlab programı kullanılarak, ses sinyalinin bazı öznitelikleri çıkartılmış ve bu öznitelikler k-en yakın komşuluk algoritması kullanılarak sınıflandırılmıştır. Sınıflandırma sonucunda da mesafenin uzaklığı tahmin edilmeye çalışılmıştır. İkinci çözüm olarak ise tensorflow altyapısında keras kütüphanesi kullanılarak python programlama dili ile derin sinir ağ modeli oluşturulmuş ve bu model kullanılarak ses kaynağı uzaklığı tahmini yapılmıştır. Ayrıca ikinci yöntem GPU ve CPU'da ayrı ayrı gerçeklenmiştir. GPU olarak NVIDIA 4GB GTX-965M kullanılmıştır. GPU olmadan ise INTEL i5 işlemci üzerinde 4GB ram 4 Core özelliklere sahip bilgisayar kullanılmıştır.

Veri seti olarak, kendi oluşturduğumuz set kullanılmıştır. Kayıtlar 1 metre, 2 metre, 3 metre, 4 metre olmak üzere 4 farklı uzaklıktan yapılmıştır. Şekil 3.1-3.2'de kayıt ortamından alınan görüntü sunulmuştur.



Şekil 3.1 Kayıt ortamından alınan 1. görüntü



Şekil 3.2 Kayıt ortamından alınan 2. görüntü

Kullanılan ses kaynağı erkek ve bayan sesleri içermektedir. Konuşma dili İngilizcedir. Örnekleme frekansı 8 kHz'dir. Kendi kayıt ortamımızın oluşturulduğu sistemin şekil 3.2'de de görüldüğü gibi, ses seviyesi ölçer cihaz ile ortamın ses dB seviyesi de ölçülmüştür.

3.1 Öznitelik Çıkarımı ve Sınıflandırma

Ses kaynağı uzaklığı tespiti için ses sinyalinin bazı istatistiksel özellikleri mesafeye bağlı olarak değişmektedir (Georganti vd. 2011). Bu çalışma kapsamında da öznitelikler çıkartılırken ses sinyalinin istatistiksel özelliklerinden faydalanılmıştır. Çözüm üç adımda yapılmıştır. Sistemin genel blok diyagramı şekil 3.3'te sunulmuştur.



Şekil 3.3 Ses kaynağı mesafe tespiti sistemi blok diyagramı

İlk adımda ses sinyali 20ms'lik çerçeveler halinde alınmış ve çerçevenin enerjisine bakılarak ses içerip içermediğine karar verilmiştir. Eğer ses içeriyorsa Hanning pencereleme yapılmıştır. İkinci adımda ise pencereleme yapılan bu sinyalden hata tortusu, 0,90'lık oran (percentile), savrukluk ve basıklık gibi öznitelikler çıkarılmıştır. Son olarak ise k-en yakın komşuluk algoritması kullanılarak mesafenin uzaklığı tahmin edilmiştir.

3.1.1 Öznitelikler

Gerçek ses sinyalinden, LPC katsayıları ile oluşturulmuş öngörüm sinyalinin çıkarılmasıyla elde edilen hata (tortu) sinyalinin kare ortalama hataları hesaplanmıştır. $e(n)$ hata sinyali olmak üzere, eşitlik 3.1'de matematiksel ifadesi sunulmuştur. Daha sonra hata sinyalinde genliği % 90'ın üzerinde olan değerler elenerek tekrar kare ortalama hatası elde edilmiştir. Bu percentile olarak adlandırılmış ve kare ortalama hatanın percentile'a olan oranı ilk öznitelik olarak elde edilmiştir (Georganti vd. 2011).

$$r = \overline{\frac{1}{\tau} \sum_{n=1}^{\tau} e^2(n)} \tag{3.1}$$

Hata sinyalinin basıklığı (kurtosis) hesaplanmış ve ikinci öznitelik olarak kullanılmıştır. e ortalama değer olmak üzere, basıklığın matematiksel ifadesi eşitlik 3.2'de sunulmuştur (Georganti vd. 2011).

$$k = \frac{\tau \sum_{n=1}^{\tau} (e\ n - e)^4}{(\sum_{n=1}^{\tau} (e\ n - e)^2)^2} \tag{3.2}$$

Hata sinyalinin hızlı Fourier dönüşümü sonucu elde edilen kepstral katsayıların savrukluğu (skewness) hesaplanmış ve üçüncü öznitelik olarak kullanılmıştır. $C\ z$ kepstral katsayılar, C kepstral katsayıların ortalaması olmak üzere, savrukluğun matematiksel ifadesi eşitlik 3.3'de sunulmuştur (Georganti vd. 2011).

$$g = \frac{\overline{\tau} \sum_{n=1}^{\tau} (C\ z - C)^3}{(\sum_{n=1}^{\tau} (C\ z - C)^2)^{3\ 2}} \tag{3.3}$$

Matematiksel ifadeleri verilen bu üç öznitelik ses sinyalinin 2s'lik bloklar halinde tüm ses içeren çerçevelere uygulanmış ve en son olarak nxm boyutlu öznitelik matrisi elde edilmiştir.

Çıkarılan bu öznitelikler k-en yakın komşuluk algoritması kullanılarak sınıflandırılmış ve test sonuçları Araştırma Bulguları bölümünde sunulmuştur.

3.2 Derin Sinir Ağlarının Uygulanması

Ses kaynağı uzaklığı tespiti problemi, geleneksel yöntem olan öznitelik çıkarımı ve sınıflandırma işleminden sonra günümüzde birçok sinyal işleme probleminin çözümünde kullanılan derin sinir ağları ile çözümü gerçekleştirilmiştir.

Derin sinir ağlarında sinyal işleme problemine çözüm olarak konvolüsyonel sinir ağları yaygın olarak kullanılmaktadır. Genişletilmiş konvolüsyonel sinir ağları ise başarıyı daha da artırmaktadır. Bunun sebeplerinden birisi genişletilmiş konvolüsyonel sinir ağlarındaki katman, dilation_rate parametresine bağlı olarak gerisinde bulunan katmanlardan da beslenebilmektedir. Bu sayede işlenecek sinyale daha geniş bir pencereden bakılır ve dolayısıyla daha çok bilgi içeren sinyal ile besleme yapılmaktadır. Bu çalışma kapsamında da hem konvolüsyonel sinir ağları, hem de genişletilmiş konvolüsyonel sinir ağları kullanılarak iki farklı model oluşturulmuş ve bu iki modele ait test sonuçları Araştırma Bulguları bölümünde sunulmuştur.

3.2.1 Model oluşturma

Derin sinir ağlarında model, giriş katmanı, saklı katman ve çıkış katmanından oluşmaktadır. Giriş katmanı boyutu, işlenecek sinyalin boyutuyla aynı olmalıdır. Ses sinyali 20-30 ms aralığında durağan olduğu için çerçeve uzunluğu bu aralıkta olmalıdır. Bu çalışma kapsamında 20ms çerçeve uzunluğu seçilmiştir. Ses sinyalimizin örnekleme frekansı 8 kHz olduğu için işlenecek çerçeveler 160 örnek içermektedir. Modelimizin giriş katmanının boyutu da 160 olacak şekilde ayarlanmıştır. Ara katmanlarda konvolüsyonel ve genişletilmiş konvolüsyonel ağlar kullanılmıştır. Aktivasyon fonksiyonu olarak "tanh" kullanılmıştır. Katman çıkışlarında boyut azaltma işlemi

yapmak için çalışma prensibi 2. Bölümde açıklanan “maxpooling” yöntemi kullanılmıştır. Çıkış katmanında ise sınıflandırıcı görevi yapan “softmax” fonksiyonu kullanılmıştır.

3.2.2 Model eğitimi

Oluşturulan model epoch sayısı 10, 100, 500 ve 1000 olmak üzere farklı varyasyonlarla eğitilmiş ve model doğruluk değerleri Araştırma Bulguları bölümünde sunulmuştur. Optimizer parametresi olarak “adam” yöntemi seçilmiştir. Ayrıca GPU ve CPU’da olmak üzere iki ayrı platform kullanılmış ve eğitim süresindeki kazanç Araştırma Bulguları bölümünde sunulmuştur.

3.2.3 Model testi

Eğitimi tamamlanmış olan model, eğitim verisi içermeyen test verileri ile test edilmiş ve karışıklık matrisi Araştırma Bulguları bölümünde sunulmuştur.

4. ARAŞTIRMA BULGULARI

Yapılan bu çalışmada ses kaynağı uzaklığı tespiti için ilk olarak geleneksel yöntem olan öznitelik çıkarımı ve sınıflandırma yöntemiyle sonuçlar elde edilmiştir. Elde edilen bu başarıyı artırmak için günümüzde yaygın olarak kullanılan derin sinir ağları yöntemi seçilmiş ve başarımda % 10 civarında bir artış sağlanmıştır.

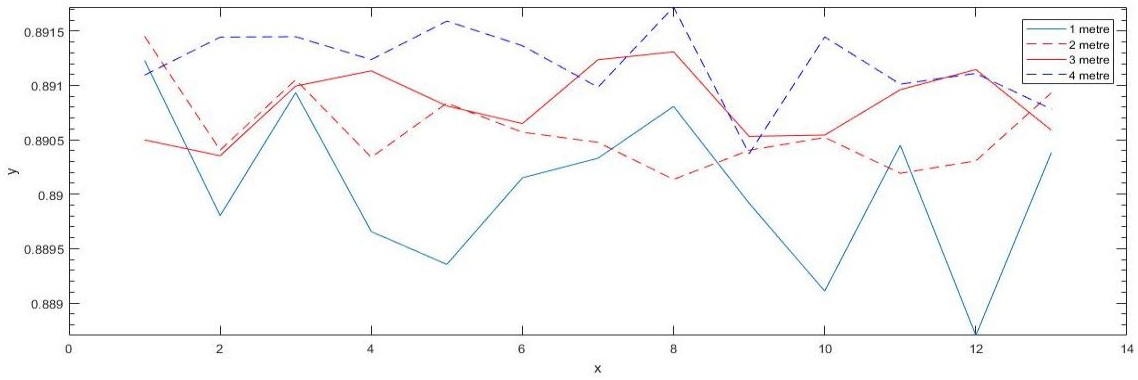
4.1 Öznitelik Çıkarımı ve Sınıflandırma Yöntemi Bulguları

4.1.1 Öznitelik seçiminin etkisi

Bu yöntemde, sinyalin Bölüm 3.1.1 Öznitelikler kısmında açıklanan, LPC katsayıları ile oluşturulmuş öngörüm sinyalinin çıkarılmasıyla elde edilen hata (tortu) sinyalinin kare ortalama hatalarının, percentile'a yani hata sinyalinde genliği % 90'ın üzerinde olan değerler elenerek tekrar kare ortalama hatası hesaplanan değere olan oranı yani hata oranı, savrukluk ve basıklık gibi istatistiksel özellikleri kullanılarak öznitelikler çıkartılmıştır.

İlk öznitelik olan Bölüm 3.1.1 Öznitelikler başlığı altında detaylı açıklaması yapılan hata oranını 1 metre, 2 metre, 3 metre ve 4 metre uzaklıkları için hesaplanmış değerleri şekil 4.1'de sunulmuştur. Şekil 4.1'de görüldüğü üzere 1 metre ve 4 metre için tam bir ayrım sağlayamamıştır.

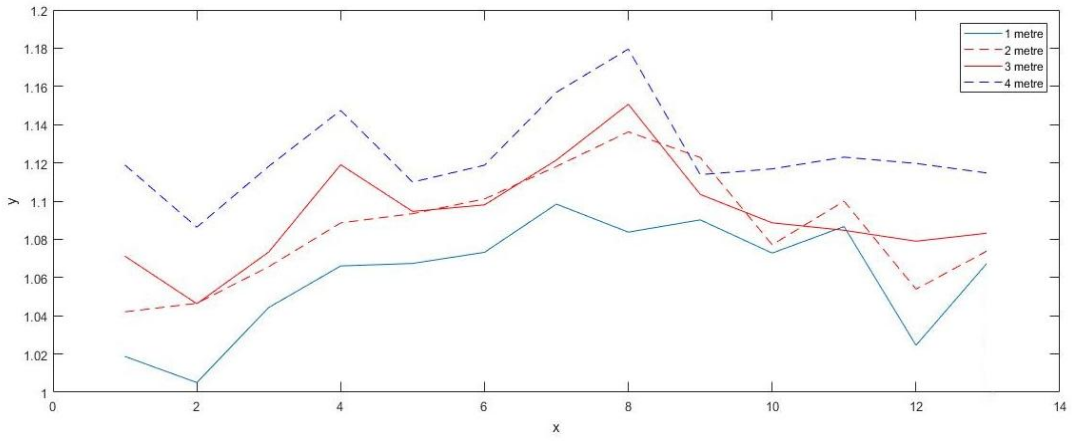
x sayı, y genlik değerleri olmak üzere,



Şekil 4.1 1 metre, 2 metre, 3 metre ve 4 metre için hata oranı özniteliği grafiği

İkinci öznitelik olan savrukluk oranını 1 metre, 2 metre, 3 metre ve 4 metre uzaklıkları için hesaplanmış değerleri şekil 4.2'de sunulmuştur. Şekil 4.2'de görüldüğü üzere 1 metre, 2 metre, 3 metre ve 4 metre için tam bir ayrım sağlayamamıştır.

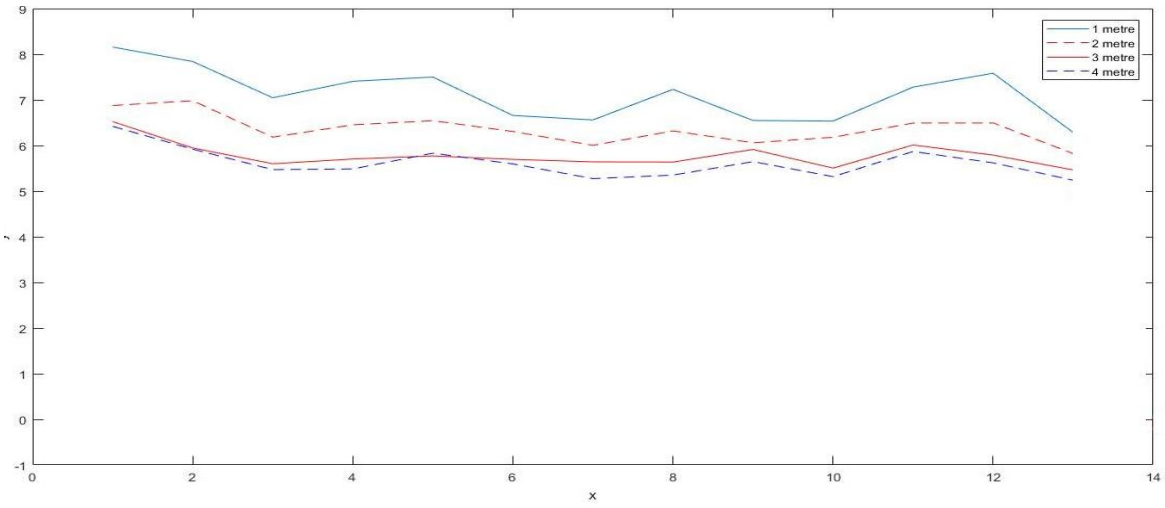
x sayı, y genlik değerleri olmak üzere,



Şekil 4.2 1 metre, 2 metre, 3 metre ve 4 metre için savrukluk oranı özniteliği grafiği

Üçüncü öznitelik olan basıklık oranını 1 metre, 2 metre, 3 metre ve 4 metre uzaklıkları için hesaplanmış değerleri şekil 4.3'te sunulmuştur. Şekil 4.3'te görüldüğü üzere 1 metre, 2 metre, 3 metre için tam bir ayrım sağlanmıştır. 3 ve 4 metre için kısmi bir ayrım sağlanmıştır.

x sayı, y genlik değerleri olmak üzere,



Şekil 4.3 1 metre ve 4 metre için basıklık oranı özniteliği grafiği

Elde edilen sonuçlar, karışıklık matrisi (confisuion matrix) dediğimiz tablo yapısında sunulmuştur. Karışıklık matrisinde tablonun diagonal değerlerinin % 100, diğer sütünların 0 çıkması beklenmelidir. Yani elde edilen başarı, verilen tablonun diagonal değerleri olarak elde edilmektedir.

Sadece hata oranı özniteliği kullanıldığında, elde edilen karışıklık matrisi çizelge 4.1'de sunulmuştur.

Çizelge 4.1 Hata oranı özniteliği ve k-nn sınıflandırıcı karışıklık matrisi

Mesafe	1 metre	2 metre	3 metre	4 metre
1 metre	**%41.7**	%30.4	%26.7	%1.2
2 metre	%30	**%35.5**	%32	%2.5
3 metre	%2.3	%30.8	**%33.2**	%33.7
4 metre	%21	%23	%29	**%27**

Sadece savrukluk özniteliği kullanıldığında elde edilen karışıklık matrisi çizelge 4.2'de sunulmuştur.

Çizelge 4.2 Savrukluk Özniteliği ve k-nn sınıflandırıcı karışıklık matrisi

Mesafe	1 metre	2 metre	3 metre	4 metre
1 metre	**%44.9**	%30.8	%24.3	0
2 metre	%12	**%41**	%38	%9
3 metre	%2	%32.2	**%35**	%30.8
4 metre	%14	%25.5	%30.5	**%30**

Sadece basıklık özniteliği kullanıldığında elde edilen karışıklık matrisi çizelge 4.3'te sunulmuştur.

Çizelge 4.3 Basıklık Özniteliği ve k-nn sınıflandırıcı karışıklık matrisi

Mesafe	1 metre	2 metre	3 metre	4 metre
1 metre	**%60**	%25	%15	0
2 metre	%7.2	**%51.5**	%35.5	%5.8
3 metre	%1.2	%23	**%45**	%30.8
4 metre	%6	%20	%35	**%39**

En yüksek başarım oranı ise bu üç istatiksel özelliğin bir arada kullanılmasıyla sağlanmıştır. Bu üç öznitelik ile yapılan test sonucu oluşturulmuş olan karışıklık matrisi çizelge 4.4'de sunulmuştur.

Çizelge 4.4 Öznitelik çıkarma ve k-nn sınıflandırıcı karışıklık matrisi

Mesafe	1 metre	2 metre	3 metre	4 metre
1 metre	**%71.9**	%20.4	%7.69	0
2 metre	%7.69	**%50.83**	%33.76	%7.69
3 metre	0	%23.07	**%46.83**	%30.07
4 metre	0	%25.69	%36.83	**%37.45**

4.2 Derin Sinir Ağları Yöntemi Bulguları

Oluşturulan derin sinir ağı modeli, 1 giriş, 7 saklı (konvolüsyonel ve genişletilmiş konvolüsyonel), 1 de çıkış katmanı olmak üzere 9 katmandan oluşturulmuştur. Katman çıkışlarında aktivasyon fonksiyonu olarak "tanh" fonksiyonu kullanılmıştır. Çıkış katmanında ise sınıflandırıcı görevi gören ve olasılıksal değerler üreten "softmax" fonksiyonu kullanılmıştır. Oluşturulan ağın başarı oranına katman(konvolüsyonel, genişletilmiş konvolüsyonel) türü seçimi, epoch sayısı ve veri miktarının etki ettiği gözlenmiştir.

4.2.1 Katman türü seçiminin etkisi

Bu çalışma kapsamında ikinci bir yöntem olarak derin sinir ağları kullanılmıştır. Model oluşumunda konvolüsyonel sinir ağı ve genişletilmiş konvolüsyonel sinir ağları denenmiştir. Genişletilmiş konvolüsyonel sinir ağı kullanılan model de başarımın % 7 civarında arttığı görülmüştür.

4.2.2 Epoch sayısının etkisi

Epoch sayısı, modelin hedef öğrenme noktasına ulaşması için eğitim aşamasındaki yineleme (iterasyon) olarak tanımlanmaktadır. Bu çalışma kapsamında 10, 100, 500 ve

1000 epoch için modelin eğitimi yapılmış ve test sonuçları çizelge 4.5-4.8'de sunulmuştur.

Epoch parametresinin oluşturulan derin sinir ağı modelinin başarısına etkisi doğrudan gözlenmiştir. Grafiksel gösterimi şekil 4.4'te sunulmuştur.

Çizelge 4.5 10 Epoch ile eğitim sonucu karışıklık matrisi

Mesafe	1 metre	2 metre	3 metre	4 metre
1 metre	**%48,07**	%22,78	%18,44	%10,67
2 metre	%30,07	**%34,08**	%22,62	%13,21
3 metre	%24,58	%28,50	**%30,80**	%16,10
4 metre	%24,66	%28,93	%28,35	**%18,04**

Çizelge 4.6 100 Epoch ile eğitim sonucu karışıklık matrisi

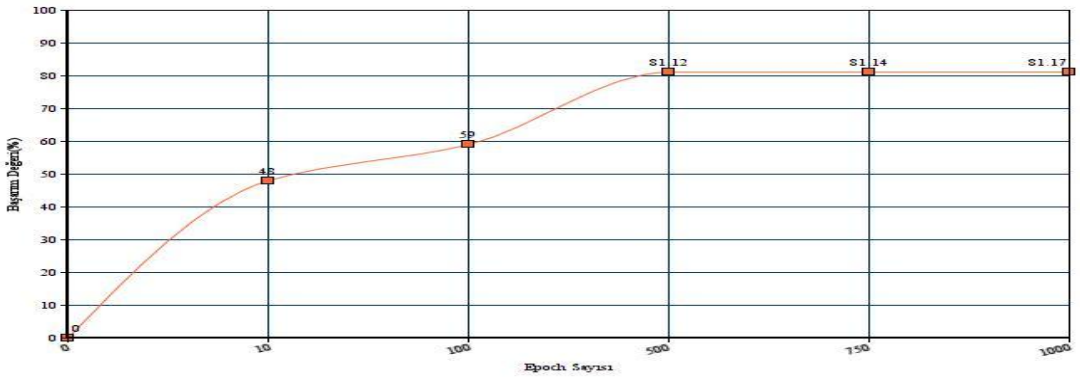
Mesafe	1 metre	2 metre	3 metre	4 metre
1 metre	**%59,48**	%19,86	%12,60	%8,04
2 metre	%31,79	**%40,30**	%17,5	%10,32
3 metre	%27,67	%24,94	**%32,59**	%14,78
4 metre	%29,94	%23,82	%23,79	**%22,43**

Çizelge 4.7 500 Epoch ile eğitim sonucu karışıklık matrisi

Mesafe	1 metre	2 metre	3 metre	4 metre
1 metre	**%81,12**	%9,64	%6,77	%2,46
2 metre	%10,09	**%78,28**	%7,94	%3,67
3 metre	%11,66	%12,86	**%70,33**	%5,14
4 metre	%16,11	%22,84	%22,32	**%38,71**

Çizelge 4.8 1000 Epoch ile eğitim sonucu karışıklık matrisi

Mesafe	1 metre	2 metre	3 metre	4 metre
1 metre	**%81,17**	%9,61	%6,76	%2,45
2 metre	%10,08	**%78,38**	%7,91	%3,64
3 metre	%11.5	%12.5	**%71**	%5
4 metre	%18	%21	%23	**%38**



Şekil 4.4 Epoch Etkisinin Grafiksel Gösterimi

4.2.3 Aktivasyon fonksiyonu etkisi

Aktivasyon fonksiyonu seçimi çözülmesi gereken probleme göre belirlenmelidir. Çünkü aktivasyon fonksiyonlarının çalışma aralığı tanımlıdır. Kullanılan veri setinin veri değer aralığı, aktivasyon fonksiyonunun çalışma aralığına düşmez ise veri kullanılamayacak dolayısıyla veri azlığı nedeniyle bir öğrenme gerçekleştirilemeyecektir.

Bu çalışma kapsamında da ses sinyali verisi negatif ve pozitif değerler içerdiği için "tanh" fonksiyonunun çalışma aralığı da negatif ve pozitif değerlerde olduğu için bu aktivasyon fonksiyonu seçilmiştir.

4.2.4 GPU kullanımı etkisi

Günümüzde GPU kartlarının yüksek işlem kapasitesi ve yeteneği sayesinde bu donanım hesaplama amacıyla yaygın olarak kullanılmaktadır. Derin sinir ağlarında da kullanılan tensorflow kütüphanesi sürümü ile NVIDIA GPU kartlarında hızlı işlem yapma imkanı vardır. Bu çalışma kapsamında da model hem yalnız CPU hem de yalnız GPU da denenmiştir. Eğitim süresi yalnız CPU'da 500 epoch için 60 dakika sürmüş iken yalnız GPU'da (kendi bilgisayarımızda bulunan NVIDIA 4GB GTX-965M GPU kartı ile) 30 dakika sürmüştür.

4.2.5 Veri miktarının etkisi

Bu çalışma kapsamında mevcut verinin sadece % 40'ı ile eğitim yapıldığında öğrenme gerçekleştirilememiş ve derin sinir ağlarının başarımının veri miktarına doğrudan bağlı olduğu gözlemlenmiştir. Tüm eğitim seti ve 500 epoch ile eğitim sonucu oluşturulan modelin test sonuçları çizelge 4.7'de sunulmuştur. Bu çalışma kapsamında 9 ayrı konuşmacı tarafından, 55 dakikalık, 8 kHz'de örneklenmiş ses kayıtları kullanılmıştır.

5. SONUÇ

5.1 Değerlendirme

Bu tez kapsamında tek mikrofon ile ses kaynağı uzaklığı tespiti probleminin çözümü iki ayrı yaklaşım kullanılarak çözülmüştür. İlk çözüm olarak geleneksel yöntem olan öznitelik çıkarımı ve çıkarılan bu özniteliklerin sınıflandırıcı yardımı ile sınıflandırılarak ses kaynağı uzaklığı tahmini yapılmıştır. Öznitelikler, ses sinyalinin istatistiksel özellikleri olan, Bölüm 3.1.1'de detaylı anlatımı yapılmış olan hata oranı, savrukluk ve basıklık olarak belirlenmiştir. Çıkarılan bu öznitelikler Bölüm 2.4.1'de detaylı anlatımı yapılmış olan k-nn sınıflandırıcı ile sınıflandırılarak ses kaynağı uzaklığının tahmini yapılmıştır. İkinci çözüm ise, son yıllarda sinyal işleme problemlerinde oldukça yaygın kullanılan ve başarılı sonuçlar veren derin sinir ağları yöntemi kullanılarak ses kaynağı uzaklığı tespiti problemi çözülmüştür.

Geleneksel yöntemin gerçeklenmesinde Matlab programı kullanılarak öznitelikler çıkarılmış ve k-nn sınıflandırıcı ile mesafe tahmin edilmiş. İkinci yöntemde ise Tensorflow+Keras kütüphaneleri kullanılmış ve python programı kullanılarak derin sinir ağları ile mesafe tahmini yapılmıştır. Sistemlerin birbirlerine kıyasla başarı oranları karşılaştırılmış ve bu başarıyı etkileyen parametreler incelenmiştir.

Öznitelik çıkarımı ve sınıflandırma yönteminde, öznitelik seçiminin sistemin başarısına doğrudan etki ettiği gözlenmiştir.

İlk öznitelik olan şekil 4.1'de grafiksel gösterimi verilen hata oranı, tek başına kullanıldığında mesafe tahmininde yeterli başarı oranını vermemiştir. Şekil 4.1'de görüldüğü üzere 1 metre ile 2 metre, 2 metre ile de 3 metre ve 4 metre ses sinyalleri karışmaktadır.

İkinci öznitelik olan şekil 4.2'de grafiksel gösterimi verilen savrukluk oranı da tek başına kullanıldığında mesafe tahmininde yeterli başarı oranını vermemiştir. Şekil 4.2'de görüldüğü üzere 1 metre diğerlerine karışmazken, 2 metre, 3 metre ve 4 metre ses sinyali birbirine karışmaktadır.

Üçüncü öznitelik olan şekil 4.3'te ise grafiksel gösterimi verilen basıklık oranı özniteliği ise mesafe tahmini için ayırt edici bir öznitelik olduğu gözlenmiştir. Şekil 4.3'te görüldüğü üzere 1 metre, 2 metre, 3 metre ve 4 metre ses sinyalleri birbirine karışmamış, kısmi olarak 3 metre ve 4 metre ses sinyalleri karışmıştır. Fakat bu üçüncü özniteliğin de tek başına kullanımı yeterli başarı oranını vermemektedir.

Çizelge 4.1'de görüldüğü üzere sadece hata oranı özniteliği kullanıldığında başarılı sonuçlar vermemiştir. Benzer olarak çizelge 4.2'de de sadece savrukluk özniteliği kullanıldığında elde edilen başarım değerleri sunulmuştur ve başarılı sonuçlar vermediği gözlemlenmiştir. Çizelge 4.3'te ise sadece basıklık özniteliği kullanılmış ve diğer 2 özniteliğe kıyasla daha başarılı sonuçlar elde edilmiştir.

Bu üç özniteliğin birlikte kullanımı ile en yüksek başarı oranı elde edilmiş ve çizelge 4.4'te sunulmuştur. Çizelge 4.4'te görüldüğü üzere, 1 metre ses sinyali, 4 metre ses sinyaline hiç karışmamış, sadece 2 metre ve 2 metreden daha düşük oranda 3 metre ses sinyaline karışmıştır. Benzer şekilde 2 metre ses sinyali bir sonraki adım olan 3 metre ses sinyaline karışmış, 3 metre ses sinyalinden daha düşük oranda 1 metre ve 4 metre ses sinyallerine karışmıştır.

Çizelge 4.4'te görüldüğü üzere, ses kaynağından uzaklaştıkça başarı oranı düşmektedir. Bunun sebebi olarak, ses sinyalinin zayıflayarak istatistiksel özelliklerinin tam çıkarılamaması olduğu düşünülmektedir.

Ayrıca ses kaynağından uzaklaştıkça, başarı oranını düşüren diğer bir faktörün de kayıt alınan odanın ses yalıtımlı olmadığı ve yankı etkisinin de ses sinyalinin istatistiksel özelliklerini bozduğu düşünülmektedir.

Geleneksel yöntemle elde edilen bu başarı oranı nasıl artırılabilir diye düşünülmüş ve son yıllarda sinyal işleme problemlerine yaygın uygulanan, oldukça başarılı sonuçlar veren derin sinir ağları ile ses kaynağı uzaklığı tespiti problemi çözümüne karar verilmiştir.

Derin sinir ağları ile oluşturulan modelin başarı oranı, veri miktarı, eğitim epoch sayısı ve aktivasyon fonksiyonu gibi parametrelere doğrudan bağlıdır. Veri miktarının az olduğu durumlarda, modelin öğrenemediği gözlenmiştir. Ayrıca epoch sayısının düşük tutulduğu durumlarda, model hedef öğrenme noktasına ulaşamamış ve dolayısıyla öğrenme başarı oranı düşük kalmıştır.

Bu çalışmada epoch sayısının etkisi incelenmiş ve model 10, 100, 500 ve 1000 epoch sayısı olmak üzere eğitilmiştir. Çizelge 4.5'de görüldüğü üzere 10 epoch ile eğitilen modelde, 1 metre ses sinyali için % 48,07 başarı oranı elde edilirken, epoch sayısı 100 yapıldığında çizelge 4.6'de görüldüğü üzere 1 metre ses sinyali için başarı oranı % 24 civarında bir artış sağlamıştır. Çizelge 4.7'te ise epoch sayısı 500 olarak eğitilmiş modelin sonuçları sunulmuştur. Burada ise başarı oranı daha da artmış ve 10 epoch sayısı ile eğitilmiş modele göre % 33'lük bir artışla, % 48 civarında olan başarı oranı % 81 civarlarına çıkartılmıştır. Çizelge 4.7'te görüldüğü üzere, en iyi öğrenme başarı oranı, epoch sayısının 500, veri miktarının çok olduğu durumda elde edilmiştir. Ayrıca epoch sayısı 1000 olarak alındığında başarım değerleri 500 epoch sayısına kıyasla değişmemiştir. Başarım değerleri çizelge 4.8'de sunulmuştur.

Ayrıca derin sinir ağlarında dikkat edilmesi gereken önemli bir nokta da aktivasyon fonksiyonu seçimidir. Çünkü verilerimizin değer aralığı aktivasyon fonksiyonunun çalışma aralığına düşmez ise veri kaybı olacak ve dolayısıyla veri miktarındaki azlıktan dolayı eğitim aşamasında öğrenme işlemi gerçekleşemeyecektir. Eğitim esnasında zamandan kazanç sağlama durumu ise GPU ile model eğitim esnasında gerçekleşmiştir.

5.2 Öneriler

Ses kaynağı mesafe tespiti problemi geleneksel yöntem olan öznitelik çıkarımı ve sınıflandırma ile çözülmek isteniyorsa, ses sinyalinin istatistiksel özelliklerinden faydalanılmalıdır. Bu istatistiksel özellikleri bozan ortamın yankılanma oranı mümkün mertebe minimize edilmelidir. Bu işlem için ses sinyali bir ön işleme bloğuna alınmalı ve ses sinyalindeki yankılanma minimize edilmeli bu işlemden sonra ses sinyalinin istatistiksel özellikleri kullanılarak öznitelik çıkarımı yapılmalıdır.

Ses kaynağı mesafe tespiti problemi derin sinir ağları ile çözülmek isteniyorsa, veri miktarına, kullanılacak olan Bölüm 2.4.2 ve 2.4.2.1 kısımlarında detaylandırılan katman tipine ve seçilecek olan Bölüm 2.4.2.3'te detaylandırılan aktivasyon fonksiyonuna dikkat edilmelidir. Ayrıca veri miktarı fazla ise eğitim esnasında zamandan tasarruf etmek için GPU destekli kütüphaneler ve yaklaşımlar kullanılmalıdır.

KAYNAKLAR

Chollet, F., 2018. Deep Learning with Python. Manning Publications Co. ISBN: 978-1-61729-443-3.

Duda, R.O., Peter, E., Hart, P.E. and Stork, D.G. 2000. Pattern classification. Wiley-Interscience, ISBN:0471056693, USA.

Fan, J., Luo, Q., Ma, D., 2010. Symposium on Localization Estimation of Sound Source by Microphones Array. Vol. 7, pp. 312-317 China.

Fukushima, K.N., 1980. A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. Biol. Cybern., vol. 36, pp. 193–202.

Georganti, E., May, T., Van, P.S., Harma, A. and Mourjopoulos, J., 2011. Speaker Distance Detection Using a Single Microphone. IEEE Transactions on Audio, Speech and Language Processing. Audio, Speech and Language Processing, 19(7), 1949-1961. DOI: 10.1109/TASL.2011.2104953.

Gulli, A., Pal, S., 2017. DeepLearning with Keras. Packt Publishing Ltd. ISBN: 978-1-78712-842-2.

Härmä, A., Jakka, J., Tikander, M., Karjalainen, M., Lokki, T., Hiipakka, J. and Lorho, J., 2004. Augmented reality audio for mobile and wearable appliances, J. Audio Eng. Soc., vol. 52, pp. 618–639, Jun.

Härmä, A., 2009. Ambient human-to-human communication, in Handbook of Ambient Intelligence and Smart Environments. New York: Springer, pp. 795–823.

He K., Zhang, X., Ren, S., Sun, J., 2015. Deep Residual Learning for Image Recognition. Computer Vision and Pattern Recognition (CVPR). Vol. 90, pp. 770-778.

Hamacher, V., Chalupper, J., Eggers, J., Fischer, E., Kornagel, U.,Puder, H. and Rass U., 2005. Signal processing in high-end hearing aids: State of the art, challenges, and future trends, EURASIP J. Appl. Signal Process., vol. 2005, pp. 2915–2929.

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. ImageNet Classification with Deep Convolutional Neural Networks. Neural Processing Systems Conference.

Kwok, N.M., Buchholz, J., Fang, G., Gal, J., 2005. Sound Source Localization: Microphone Array Design and Evolutionary Estimation, DOI: 10.1109/ICIT.2005.1600650, 14-17 Dec. 2005, Hong Kong, China.

Omologo, M., Svaizer, P., and Matassoni, M., 1998. Environmental conditions and acoustic transduction in hands-free speech recognition, Speech Commun., vol. 25, no. 1-3, pp. 75–95, Aug.

Rabiner, L.R. and Juand, B.H. 1993. Fundamentals of Speech Recognition. Prentice-Hall, Englewood Cliffs, N.J., ISBN: 0-13-015157-2.

Schmidhuber, J., 2015. Deep learning in neural networks: An overview, Neural Networks vol. 61, pp. 85–117, 2015.

Selen, N. 1979. Söyleyis sesbilimi, akustik sesbilim ve Türkiye türkçesi, Türk Dil Kurum Yayınları.

Simonyan, K., Zisserman, A., 2015. Very Deep Convolutional Networks for Large-Scale Image Recognition. ICLR Conference.

Szegedy, C., Liu, W., Jia, Y., Sermanent, P., Reed, S., Angualev, D., Erhan, D., Vanhoucke, V. and Robinevich, A., 2014. Going Deeper with Convolutional. ILSVCR2014.

ÖZGEÇMİŞ

Adı Soyadı : Hulusi EROĞLU

Doğum Yeri : Bayat

Doğum Tarihi : 12/04/1989

Medeni Hali : Bekar

Yabancı Dili : İngilizce

Eğitim Durumu (Kurum ve Yıl)

Lise : Keçiören Lisesi (2007)

Lisans : Ankara Üniversitesi Mühendislik Fakültesi Elektrik-Elektronik Mühendisliği (2014)

Yüksek Lisans : Ankara Üniversitesi Fen Bilimleri Enstitüsü Elektrik-Elektronik Anabilim Dalı (Şubat 2015 – Temmuz 2018)

<u>Ulusal Kongre Sunum</u>

Eroğlu H., Gökçe C. O., İlk H. G., Uzamsal ve Zamansal Yöntemler ile İnsan Hareketi Algılama, SIU-2014,Trabzon