T.C.
SAKARYA UYGULAMALI BİLİMLER ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

DERİN ÖĞRENME ALGORİTMALARI KULLANARAK BİR KONUŞMA TANIMA UYGULAMASI

YÜKSEK LİSANS TEZİ

Harun KUTUCU

Enstitü Anabilim Dalı	**:**	**ELEKTRİK-ELEKTRONİK MÜHENDİSLİĞİ BÖLÜMÜ**
Tez Danışmanı	**:**	**Prof. Dr. Abdullah FERİKOĞLU**

Mayıs 2020

T.C.
SAKARYA UYGULAMALI BİLİMLER ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

DERİN ÖĞRENME ALGORİTMALARI KULLANARAK BİR KONUŞMA TANIMA UYGULAMASI

YÜKSEK LİSANS TEZİ

Harun KUTUCU

Enstitü Anabilim Dalı : ELEKTRİK-ELEKTRONİK MÜHENDİSLİĞİ BÖLÜMÜ

Bu tez 28/05/2020 tarihinde aşağıdaki jüri tarafından oybirliği ile kabul edilmiştir.

JÜRİ	BAŞARI DURUMU
Jüri Başkanı: Prof. Dr. Abdullah FERİKOĞLU	BAŞARILI
Üye: Doç. Dr. Ahmet ZENGİN	BAŞARILI
Üye: Doç. Dr. Metin VARAN	BAŞARILI

BEYAN

Tez içindeki tüm verilerin akademik kurallar çerçevesinde tarafımdan elde edildiğini, görsel ve yazılı tüm bilgi ve sonuçların akademik ve etik kurallara uygun şekilde sunulduğunu, kullanılan verilerde herhangi bir tahrifat yapılmadığını, başkalarının eserlerinden yararlanılması durumunda bilimsel normlara uygun olarak atıfta bulunulduğunu, tezde yer alan verilerin bu üniversite veya başka bir üniversitede herhangi bir tez çalışmasında kullanılmadığını beyan ederim

Adı Soyadı
28/05/2020

ÖNSÖZ

İlk olarak bu çalışmanın gerçekleştirilmesinde,değerli bilgilerini benimle paylaşan, kendisine ne zaman danışsam benim için kıymetli zamanını ayırıp sabırla ve büyük bir ilgiyle bana yol gösteren danışmanım sayın Prof. Dr. Abdullah FERİKOĞLU hocama teşekkürlerimi sunarım.

Teşekkürlerin az kalacağı diğer üniversite hocalarımında bana kazandırdıkları herşey için ve beni gelecekte söz sahibi yapacak bilgileriyle donattıkları için hepsine teker teker teşekkürlerimi sunuyorum.

Ayrıca bu zorlu süreç boyunca arkamda duran ve destekleyen başta ailem olmak üzere arkadaşlarıma ve sevdiklerime teşekkür ederim.

Harun KUTUCU

Mayıs 2020

İÇİNDEKİLER

KISALTMALAR

AI	: Artificial Intelligence (Yapay Zeka)
CIFAR100	: 60000 adet 32x32 renk görüntüsü içeren veriseti
CNN	: Convolutional Neural Network (Evrişimsel Sinir Ağı)
CPU	: Central Processing Unit (Merkezi İşlemci Birimi)
CVA	: Common Vector Analysis (Ortak Vektör Analizi)
ÇKA	: Çok Katmanlı Algılayıcı
DVM	: Destek Vektör Makinesi
DZB	: Dinamik Zaman Bükmesi
GANs	: Generative Adversial Autoencoders
GPU	: Graphics Processing Unit (Grafik İşlemci Birimi)
HD-CNN	: Hierarchical Deep Convolutional Neural Network
HMM	: Hidden Markov Model (Saklı Markov Modeli)
Hz	: Hertz (Frekans)
IEMOCAP	: The Interactive Emotional Dyadic Motion Capture (Veriseti)
ILSVRC	: ImageNet Large Scale Visual Recognition Competition
LSTM	: Long Short-Term Memory (Erişimli Uzun Kısa Süreli Bellek)
LSVRC	: Large Scale Visual Recognition Competition
MCDNN	: Multi Column Deep Neural Network
MFCC	: Mel Frequency Cepstrum Coefficient (Mel Frekans Kepstral Katsayıları)
MJO	: Madden-Julian Oscillation (Madden-Julian Salınımı
NLP	: Natural Language Processing (Doğal Dil İşleme)
NORB	: Nesne tanımlama veriseti
RNN	: Recurrent Neural Network (Yinelenen Sinir Ağı)
SER	: Speech Emotion Recognition
SMM	: Saklı Markov Modeli
SOM	: Self Organizing Map (Öz Düzenleyici Harita)

SVM	: Support Vector Machine(Destek Vektör Makinesi)
TV	: Television (Televizyon)
VAD	: Volume Activation Detector(Sesli Etkinlik Algılayıcısı)
VAEs	: Variational Autoencoders

TABLOLAR LİSTESİ

ŞEKİLLER LİSTESİ

DERİN ÖĞRENME ALGORİTMALARI KULLANARAK KONUŞMA TANIMA UYGULAMASI

ÖZET

Bu araştırmada, İngilizce konuşma tanıma konusu tartışılmıştır. Konuşma tanıma, sesi metne dönüştürmek isteyen yapay zeka bilimlerinde ana alanlarından biridir. Yapay zeka artık her alanda liderlik yapmaya başlamıştır. Konuşma tanımada da makine öğrenme ve derin öğrenme konuları gittikçe başarılı yöntemler sunmuştur.

Konuşma tanıma, son yıllarda daha fazla dikkat çekmesine rağmen, şimdiye kadar, derin öğrenme algoritmalarının kullanımı fazla ilgi görmemiştir. Bu çalışmada, ilk önce konuşma tanıma ve derin öğrenme algoritmalarından bahsedilmiştir. Konuşma tanıma modelini oluşturmak için kullanılan Konvolüsyonel(Evrişimsel) Sinir Ağları algoritması daha sonra açıklanmıştır.

Uygulamanın sonuçlarına dayanarak, kabul edilebilir bir performans sağlayabilen bir konuşma tanıma modeli elde edilmiştir. Konuşma tanıtma modeli için elde edilen en iyi doğruluk oranı %83 olarak tespit edilmiştir. Bu çalışmada kullanılan veri, 65.000 ses bulunan İngilizce ses veri kümesidir. Araştırmaya göre, derin öğrenme algoritmaları ile konuşma tanıma sorununu çözebileceği sonucuna varılmıştır.

Anahtar sözcükler: Konuşma Tanıma - Makine Öğrenme Algoritması - Derin Öğrenme Algoritması – Sinir Ağ

SPEECH RECOGNITION APPLICATION USING DEEP LEARNING ALGORITHMS

SUMMARY

In this research, the topic of speech recognition in english voice is discussed. Speech recognition is one of the main areas of study in artificial intelligence sciences that seek to convert sound to text. Artificial intelligence is now leading in every field. In speech recognition, machine learning and deep learning have been more and more successful ways of doing things.

Speech recognition has received more attention in recent years, but so far, the use of deep learning algorithms has not received much attention. In this study, we first deal with speech recognition and deep learning algorithms. Then , we explained the Convolutional Neural Networks (CNN) algorithm used to create the speech recognition model. In the implementation part of the CNN deep learning algorithm, a speech recognition model was used, which, based on the results of the deep learning algorithm was able to provide acceptable performance. The best accuracy obtained for the speech recognition model was reported at 83%, obtained by CNN algorithm in epoch 48. The dataset used in this study is the English Voice dataset containing 65,000 voices. Based on the research, it can be concluded that deep learning algorithms can solve the problem of speech recognition.

Keywords: Speech Recognition - Machine Learning Algorithm - Deep Learning Algorithm - Neural Network

1. GİRİŞ

Konuşma tanıma, bir makinenin veya programın, konuşulan dilde sözcükleri ve ifadeleri tanımlama ve bunları makinede okunabilir bir biçime dönüştürme yeteneğidir. Başka bir tanımla, insan sesinin bilgisayar tarafından algılanması olarak bilinmektedir. Konuşma tanıma konusunun tarihçesine bakıldığında ilk yıllarda sadece sayılar ve rakamlar üzerine çalışılmıştır. 1952 yılında 'Audrey' adlı uygulama, sadece rakamları anlayabilen Bell Laboratuvarları tarafından icat edilmiştir. Daha sonra bu uygulamalar daha da genişletilmiş ve daha başarılı uygulamalar sunulmuştur.

Konuşma tanıma çok fazla alanda kullanılmaktadır. Şuan birçok işletme içinde en yaygın şekilde kullanılıyor ve bu alanlardan arama yönlendirme, konuşmadan metne işleme, sesli konuşma ve sesli aramayı söyleyebiliriz. Konuşma tanıma uygulaması geliştirmek için genellikle makine öğrenme algoritmaları kullanılmaktadır.

Konuşma tanıma uygulaması geliştirmek için derin öğrenme algoritmaları kullanılacaktır. Derin öğrenme algoritmaları son yıllarda makine öğrenme yöntemleri içerisinde parlamayı başarmıştır. Bu yüzden konuşma tanıması için bu çalışmada derin öğrenme yöntemi kullanılacaktır. Derin öğrenme kullanıldığından beri elde ettiği iyi sonuçlar nedeniyle bilim dünyasında birçok alanda kullanılmaktadır. Derin öğrenme; bilgisayarların, deneyimlerden öğrenmelerini ve dünyayı kavramların hiyerarşisi açısından anlamalarını sağlayan bir makine öğrenimi olarak tanımlanmıştır. Derin öğrenme algoritmaları, insan beyninin prensibinden ve görsel korteksin insan beyninde nasıl çalıştığından ilham alırlar. Derin öğrenme algoritmaları son yıllarda çok başarılı uygulamalar ortaya koymuşlar ve bu yüzden bu çalışmada bu algoritmalar kullanılacaktır.

Makine öğrenimi uzun yıllardır, doğal insan ve makine etkileşimini mümkün kılacak algoritmalar sunarak dünyayı açıkça değiştiren genişleyen bir alan olmuştur. Bu algoritmaların tasarımı genellikle yapay sinir ağları gibi doğadan ilham almaktadirlar. Her biri birkaç nöron içeren birçok katmandan oluşan yapay bir sinir ağı, çok boyutlu ve doğrusal olmayan verileri analiz etme yeteneklerinden dolayı sınıflandırma amacıyla kullanılır. Bu sinir ağlarının kullanımı, makine çevirisi (Arnold, et al., 1994), konuşma tanıma (Juang & Rabiner, 1993), el yazısı oluşturma (Boser, et al., 1998), karakter metni üretimi (Alpaydin, 1989), görüntü tanıma (King-Sun, 1976) gibi birçok alanda başarılı olduğu bilinmektedir.

Son zamanlarda, derin öğrenme olarak bilinen çok ilginç ve gelecek vaat eden bir teknik yapay sinir ağlarından ilham almıştır. Bu tekniğin derin olduğu söylenir, çünkü ağdaki katmanların sayısı, gizli bir katmana sahip sığ ağlara kıyasla büyüktür. Daha derin ağlar makine öğreniminde performans iyileştirme sağlayabilir. Girişler ve çıkışlar arasındaki doğrusal olmayan ilişkileri öğrendikleri için başarılı oldukları bilinmektedir. Günümüzde daha derin ağları eğitmek için iyi organize edilmiş mekanizmalar var ve bu yüzden bu tekniği kullanarak konuşma tanıma uygulaması yapmak iyi bir hamle olacaktır.

Makine öğrenimi uzun yıllardır Türkiye'de de özellikle doğal dil işleme ve konuşma tanıma alanlarında kullanılmaktadır. Çakır 2017'de gerçek zamanlı yüksek kalitede ses tanıma uygulaması yapmıştır. Bu çalışmada, konuşmacı ve dil bağımsız gerçek zamanlı ve kaliteli bir konuşma tanıma sistemi önerilmiştir. Çalışma, her bir konuşmanın sisteme metni ile etiketlenmesiyle geliştirilir. Her bir konuşma özellik çıkarımı ve sınıflandırma aşamalarından geçerek tanınır. Bu aşamaların seçimi esnasında geçmişte yapılan çalışmalardan yola çıkarak en verimli teknikler belirlenerek sistem önerimi yapılmıştır. Özellik çıkarımı aşamasında tekniğin daha iyi sonuçlar vermesi için bazı karşılaştırmalar yapılmıştır. Sınıflandırma aşamasında teknik içerisinden bazı algoritmalar ile sistemin eğitilmesi ve testi gerçekleştirilmiştir (Çakır, 2017).

Derin öğrenmeye dayalı imza tanıma uygulaması Çalık ve arkadaşları tarafından geliştirilmiştir(Çalık, et al., 2017). Konuşma sinyali tanımı için Durak ve arkadaşları makine öğrenme algoritmaları kullanmışlar(Durak & Seke & Özkan, 2015). Canlı internet yayınları için otomatik konuşma tanıma tekniği kullanılarak alt yazı oluşturulmuş (Koruyan, 2015). Otomatik ses tanıma: türkçe için genel dağarcıklı akustik model

oluşturulması ve test edilmesi çalışması yapılmış(Özbey & Bayar, 2017). Duygu tanımada konuşma tanıma yöntemi kullanılmıştır ve yeni bir yöntem sunulmuştur (Özseven, 2019).

Özkan ve arkadaşları konuşma tanıma için yeni bir yaklaşım sunduklarını iddia etmişler. 4 seviye 8 yapılandırılmış arka plan gürültü kaydı eklenmiş 6 kişi tarafından konuşulan 30 cümle üzerinde yapılan testler, önerilen CVA(Common Vector Analysis) yönteminin diğer 5 yönteme göre üstün olduğu sonucuna varılmıştır. CVA tanıma uygulamalarında kullanılan bir alt metottur. CVA'da, her bir konuyu temsil eden eğitim verileri kendi sınıfını oluşturmak için kullanılır. Bir konuşma tanıma uygulamasında, ortam gürültüsü, yaş etkeni ve konuşanların cinsiyetleri sınıf içinde farklılıklara neden olmaktadır. CVA, temelde sınıftaki bu farklılıkları kaldırarak, bunların ortak bileşenine dayanır. Bu çalışmada CVA yöntemi kullanılarak bu yöntemim başka algoritmalardan daha başarılı olduğu sonucuna varılmştır (Özkan & Seke & Işık, 2016).

Yakar ve Aşlıyan tarafından saklı markov modeli kullanılarak konuşma tanıma için yeni bir yöntem kullanılmıştır. Bu çalışmada, hece ve sözcük tabanlı Türkçe konuşma tanıma sistemleri geliştirilerek karşılaştırılmıştır. Yapılan bu uygulamalar, orta ölçekli, ayrık ve kişiye bağımlı sistemlerdir. Bu sistemlerde, Dinamik Zaman Bükmesi (DZB), Destek Vektör Makinesi (DVM), Çok Katmanlı Algılayıcı (ÇKA) ve Saklı Markov Modeli (SMM) metotları kullanılarak eğitim ve test işlemleri yapılmıştır. SMM, ÇKA ve DVM metotlarıyla her hece ve sözcük için hece ve sözcük modelleri oluşturulmuştur. Bu modellere göre tanıma işlemi gerçekleştirilmiştir. Sistemler genel olarak önişleme, öznitelik çıkarma, hece ve sözcük eğitim ve tanıma safhalarından oluşmaktadır. Hece tabanlı sistemlerde artişleme işleminde uygulanmıştır. Önişleme safhasında ses sinyalleri düzleştirilir ve pencereleme işlemi yapılır. Sonra, sözcük ve hece sınırları belirlenir. Öznitelik çıkarma aşamasında, her bir sözcük ve hece için MFCC öznitelik vektörleri oluşturulur. Vektör olarak temsil edilen bu hece ve sözcükler SMM, ÇKA ve DVM metotlarıyla eğitildikten sonra tanıma işlemi yapılır. Hece tabanlı sistemlerde, artişleme yapılarak sistemlerin başarısı önemli ölçüde artırılmıştır. 200 Türkçe sözcükle yapılan test işleminde, hece tabanlı sistemlerdeki en iyi doğru tanıma oranları DZB için %94,2; ÇKA için %88; SMM için %82,6; DVM için ise %90,8 olmuştur. Sözcük tabanlı sistemlerde ise DZB için %96; ÇKA için %82,6; SMM için %89,4; DVM için ise %90,7 oranında doğru tanıma gerçekleştirildi (Yakar & Aşlıyan, 2016).

Konuşma tanıma teknolojisi kullanılarak devre tasarım ve analizi yapılmıştır. Bu çalışmada, öğrencinin sesli komutlar ile istediği deneyi seçebildiği; deney parametrelerini değiştirilebildiği; deney sonuçlarını gözlemleyerek analiz edebildiği ve tüm bunları da ellerini kullanmadan sadece konuşarak gerçekleştirebildiği bir uygulama geliştirilmiştir. Bu çalışma ile Elektrik-Elektronik Mühendisliği, Mekatronik Mühendisliği, Kontrol ve Otomasyon Mühendisliği lisans programları ile Meslek Yüksekokullarında Elektronik ve Otomasyon Bölümü önlisans programları müfredatında yer alan Devre Analizi ve Elektronik Devreler derslerine katkı sağlanması amaçlanmıştır. Web tabanlı olarak geliştirilen bu uygulamada, laboratuvar çalışmalarına yeni bir özellik ekleyerek, deneylerin gerçekleştirilmesinin tüm aşamalarında konuşma verisinin kullanılması sağlanmıştır (Yayla, 2018).

Yurtcan ve Kılıç tarafından gürültülü ortamlar için konuşma tanıma yöntemi sunulmuştur. Bu çalışmada gürültülü ortamlarda küçük bir mikrofon dizisi ile konuşma örnekleri kaydedilmiş ve geliştirilen gerçek zamanlı gürültü temizleme algoritmasıyla işlenerek bir veri seti oluşturulmuştur. Oluşturulan veri seti üzerinde Google bulut sistemi kullanılarak konuşma tanıma performansı test edilmiştir. Yapılan test sonucunda gürültü seviyesine göre bulut sistemlerinin konuşma tanıma başarısı gözlemlenmiştir. Sonuçlar göstermiştir ki mobil cihazlarda bulut sistemleri kullanarak konuşma tanıma yapılabilmesi için gürültü seviyesinin az olması veya gürültünün gerçek zamanlı olarak temizlenmesi gereklidir (Yurtcan & Kılıç, 2018).

Literatüre kattığı veya katacağı özgün yönler açısından bu çalışamda derin öğrenmenin konuşma tanıma da ne kadar etkili olabilir sorusuna cevap aranmıştır. Derin çğrenme algoritmaları oldukça ilgi göörmektedir ve konuşma tanıma için bu algoritmaları kullnmak iyi sonuçlar verebilir düşüncesi ile bu araştırma başlatılmıştır.

Tez bu şekilde yapılandırılmıştır:

Bölüm 1, Konuşmayı tanıma sistemi hakkında ayrıntılı bir genel bakış sunmaktadır.

2. Bölüm, literatür taraması yapılmıştır. konuşma tanıma, sinir ağları ve derin sinir ağları konusunda detaylı bilgi verilmiştir. Derin öğrenme tanıtılmıştır ve derin öğrenme uygulamaları hakkında bilgi verilmiştir.

Bölüm 3 ön işleme, özellik çıkarımı, veri bölümü ve ağ eğitim algoritması açıklamalarını içeren uygulamalı metodolojiyi önermektedir.

Bölüm 4, bu tez çalışmasının uygulanmasında kullanılan geliştirme araçlarını sunmaktadır ve deneyleri ve uygulama detaylarını ortaya koymaktadır. Bu bölüm, kullanılan verilerin tanımını ve derin sinir ağlarının uygulanmasını ve ayrıca özel uygulama parametrelerinin ağ performansları üzerindeki etkisini içermektedir.

Bölüm 5 sonuçları ve tartışmaları sunmaktadır. Son olarak, gelecekteki araştırma alanları için sonuç ve öneriler bölümünde açıklanmıştır.

2. LİTERATÜR

2.1.Konuşma Tanıma

Bu bölümde konuşma tanıma ve derin öğrenme konusundaki geçmişin kısa bir özeti verilmektedir. Çalışmaya konuşma tanıma ve ortak özelliklerinin gözden geçirilmesiyle başlanmıştır. Buna ek olarak, yapay sinir ağı kısaca tarif edilmiş; türleri ve mimarileri sunulmuştur. Son olarak, derin öğrenme ve özellikleri gözden geçirilmiştir.

Konuşma tanıma, bir makinenin veya programın, konuşulan dilde sözcükleri ve ifadeleri tanımlama ve bunları makinede okunabilir bir biçime dönüştürme yeteneğidir (Cakir & Sirin, 2018). Bu bölümde konuşma tanıma veya otomatik konuşma tanıma olarak da bilinen kavram hakkında bilgi verilecektir.

2.1.1. Konuşma Tanıma Nedir?

Konuşma tanıma; bir program veya sistemin, insan konuşmalarını işlemesini sağlayan bir teknik veya yetenektir. Aynı zamanda ses tanıma veya metinden konuşma özellikleri de bu kavrama dâhil edilir(Aktürk, 2015) (Büyük, 2018).

Techopedia'ya göre konuşma tanıma; "(…)İnsan sesini tanımlamak ve işlemek için bilgisayar donanımının ve yazılım tabanlı tekniklerin kullanılmasıdır. Öncelikle, konuşulan kelimeleri bilgisayar metnine dönüştürmek için kullanılır. Ek olarak, otomatik konuşma tanıma, kullanıcıların seslerini doğrulamak ve insan tarafından tanımlanan talimatlara dayanarak bir eylem yapmak için kullanılır." (Techopedia, 2012).

Günümüzdeki yaygın uygulamalar arasında ahizesiz kullanım cihazları, dikte yazılımı, Siri ve Alexa gibi sanal asistanlar bulunmaktadır. Birçok işletme daha verimli bir çağrı yönetimi gerçekleştirmek için, sesle etkinleşen çağrı merkezi hizmetleri sunar. Ayrıca konuşma tanıma, sesle etkinleşen navigasyon sistemleri ve araç radyoları için arama yetenekleri sağlayarak sürüşü daha güvenli hâle getirmeye yardımcı olur. Cihazlarla ve

uygulamalarla sesli iletişim kurmak, gün geçtikçe popülerlik kazanmaktadır. Konuşma tanıma pazarının 2023 yılına kadar 18 milyar dolar olması beklenmektedir (Techopedia, 2012).

2.1.2.Konuşma Tanıma Evrimi

Konuşma tanıma ilk kez 1950'lerde, sesle çalışan "Audrey" adlı bir makine ile sahneye çıkmıştır (Pinola, 2017). Bell Labs tarafından yaratılan bu makine, 0-9 arasındaki konuşulan sayıları anlayabilir ve tahminleri yüzde 90 doğruluk oranına sahiptir. 1962'de IBM, zamanının en gelişmiş konuşma tanıma makinesi olan ve konuşulan 16 kelimeyi anlayabilen "Shoebox"u piyasaya sürmüştür. Shoebox, 1971'de "Harpy" adlı bir sistem tarafından takip edilmiştir. Carnegie Mellon Üniversitesi'nde geliştirilen teknoloji, 1000'den fazla kelimeyi tanımıştır(Juang & Rabiner, 1993).

Kalkınma 1980 ve 1990'larda hızlanmıştır. Bilgi işlem gücü arttıkça, sistemlerin anlayabileceği terimlerin sayısı da artmıştır. 1996 yılında IBM, VoiceType Simply Speaking yazılımını piyasaya sürmüştür. Uygulamada 42.000 kelimelik bir sözcük hazinesi bulunmaktadır. Ayrıca bu yazılım İngilizce ve İspanyolca dillerini de desteklemekte ve 100.000 kelimelik heceleme sözlüğü içermektedir. Günümüzde bilişsel ve hesaplamalı yeniliklerin de desteğiyle, konuşma tanıma programları neredeyse sınırsız sayıda konuşulan kelimeyi tanıyabilme yeteneğine sahiptir(Juang & Rabiner, 1993).

2.1.3. İnsan Konuşmasının Karmaşıklıkları

İnsan konuşmalarının değişmezleri, gelişme sorunları yaratmıştır. Dilbiliminin matematiği ve istatistikleri içeren bilgisayar bilimi kolunun, en karmaşık alanlardan biri olduğunu düşünülmektedir(Yu & Deng, 2016).

Medium.com'daki Clark Boyd'a göre; "Çok daha fazla standardizasyon seviyesine sahip metnin aksine; konuşulan kelime bölgesel lehçelere, hıza, vurguya, hatta sosyal sınıfa ve cinsiyete göre değişmektedir. Bu nedenle, herhangi bir konuşma tanıma sistemini ölçeklemek her zaman önemli bir engel oldu… Aslında, ortalama bir insanı birkaç yıl alan bir yolculuğu tamamlamak için yüzlerce yıl öğretim makineleri harcadık." (3)

Bilim yazarı Chris Woodford, bilgisayarların insan konuşmasını anlamak için kullandıkları dört yaklaşımı ana hatlarıyla şöyle açıklamaktadır: (4)

- Basit Örüntü Eşleme: Her sözlü kelime, analiz gerekmeksizin bütünüyle tanınır.
- Desen ve Özellik Analizi: Bir kelime bitlere bölünür ve sesli harf gibi özellikler tarafından tanınır.
- Dil Modelleme ve İstatistiksel Analiz: Dilbilgisi ve birbirini takip eden belirli kelimeler, olasılığı tanımak ve doğruluğu artırmak için kullanılır.
- Yapay Sinir Ağları: Beyin benzeri bilgisayar modelleri, yoğun bir eğitimden sonra desenleri tanımak için kullanılır.

Yapay sinir ağları alanında doğal dil işleme ve bilişsel hesaplama, insanların nasıl düşündüğünü ve konuştuğunu daha yakından tanımak ve örnekleri çoğaltmak için kullanılmaktadır. İnsan paritesine ulaşmak, yani konuşan iki insan ile aynı derecede bir hata oranı, uzun zamandır amaç olmuştur. Böylece konuşma tanıma ilerlemekte ve giderek insan konuşmasına yaklaşmaktadır(Aktürk, 2015).

2.1.4.Konuşma Tanıma Önemi

Konuşma tanıma neredeyse sonsuz fırsatlar sunmaktadır. Ancak bu; erişim kolaylığını artırmaktan, daha az dikkat dağıtıcı olan bir aracı sürmeye değin mümkündür. Ses tabanlı kimlik doğrulama da birçok sistem için uygun bir güvenlik düzeyi oluşturmaktadır(Arora, & Reetz, 2017). Buna örnek olarak gösterilebilecek durumlar ise şunlardır:

- Bir çağrı merkezinin, ortak çağrı modellerini ve sorunlarını tanımlamak için müşteriler ve aracılar arasında kayıtlı binlerce konuşmayı kopyalaması gerekir.
- Bir tıbbi servis, doktorların hasta teşhis ve tedavi notlarını yakalamak ve günlüğe kaydetmek için bir dikte uygulaması oluşturmak istemektedir.
- Bir perakende satış şirketi, müşterileri ile olan satış sözleşmesini bir online konuşma uygulaması yoluyla genişletmek istemektedir.

İyi bir konuşma tanıma sistemi bu tür görevleri destekleyerek hız, verimlilik ve maliyet tasarrufu gibi faydalar sağlayabilir. Bu aynı zamanda insanları daha karmaşık işler için serbest bırakmaya yardımcı olur(Bennett & Babu & Morkhandikar, Gururaj, 2015).

Ses tanıma; tüketiciler için kolaylık, erişilebilirlik ve hatta güvenlik sunar. Pazar, sadece sanal asistanlar için katlanarak artmaktadır. INSEAD Kıdemli Ortak Strateji Profesörü Annet Aris'e göre; “Ses ekosistemi o kadar hızlı gelişiyor ki, ABD'deki hanelerin yüzde

75'inin şaşırtıcı bir şekilde önümüzdeki iki yıl içinde sesle çalışan bir akıllı konuşmacıya sahip olacağı tahmin ediliyor.” (5)

IBM blogunda Audioburst CTO ve Kurucu Ortağı Gal Klein şöyle açıklıyor: “Ses; çalışırken, seyahat ederken, araç kullanırken, yemek yerken ve ekrana bakmanın zor veya tehlikeli olduğu diğer birçok senaryoda kolayca tüketilebildiği için gözsüz içerik, internetteki en hızlı büyüyen içerik türlerinden biridir.” Şirket, sesle aranabilecek bir ses kütüphanesi oluşturmak için IBM Watson, şirket içi doğal dil işleme ve segmentasyon algoritmalarını kullanıyor. Klein ses kütüphanesi ile ilgili düşüncelerini; “Aylık milyonlarca dakika ses içeriğini yazıyoruz. Bu tür içerik konuşma, konuşmacı değişikliği, müzik, kahkaha, alkış, sessizlik ve bir ses programından veya podcast'ten bekleyebileceğiniz diğer her şeyi içerebilir.” şeklinde ifade ediyor. “Ses patlaması, bölümleme algoritmalarımızın konu başlığının ne zaman başladığını ve ne zaman bittiğini tam olarak anlamalarına yardımcı olmak için tüm bu ses ipuçlarını algılayabilir. Ses verileri daha sonra konuya göre düzenlenir ve arama için havuzumuzda saklanır. Bu, içeriğin genellikle yayınlanmasından birkaç saniye sonra bulunabileceği anlamına gelir.” Dolayısıyla ses odaklı sesli arama, araştırmacı ve geliştiricilerin konuşma tanıma erişimini ve potansiyelini genişletme yollarından biridir.

2.1.5.Konuşma Tanımanın Temel Özellikleri

Birçok konuşma tanıma uygulaması ve cihazı mevcuttur. Daha gelişmiş çözümler AI ve makine öğrenmeyi kullanır. Bunlar insan konuşmasını anlamak ve işlemek için dilbilgisi, sözdizimi, yapı, ses ve ses sinyallerinin kompozisyonunu birleştirirler. İdeal olarak devamlı öğrenme halindedirler ve her etkileşimde yanıtları geliştirirler(Chiu, et al., 2018).

En iyi sistem türü; kuruluşların konuşma dili ve konuşma farklarından marka bilinirliğine kadar, teknolojiyi kendi özel gereksinimlerine göre uyarlamalarına olanak tanır(Alpaydin, 1989). Örneğin:

- Dil Ağırlıklandırma: Sıkça konuşulan belirli sözcükleri (ürün adları veya endüstri jargonu gibi) halihazırda temel kelime haznesinde bulunan terimlerin ötesinde ağırlıklandırarak hassasiyeti artırmak demektir.
- Konuşmacı Etiketleme: Her konuşmacının çok katılımlı bir sohbete olan katkılarını gösteren veya etiketleyen bir transkript yazdırmak anlamına gelir.

- Akustik Eğitimi: Burada amaç, işletmenin akustik yönüne katılmaktır. Sistemi bir akustik ortama (bir çağrı merkezindeki ortam gürültüsü gibi) ve hoparlör stillerine (ses perdesi, ses ve tempo gibi) adapte olacak şekilde eğiterek gerçekleştirilir.
- Küfür Filtreleme: Belirli kelimeleri veya cümleleri tanımlamak ve konuşma çıktısını sterilize etmek için filtreler kullanılabilir.

Bu esnada konuşma tanıma ilerlemeye devam etmektedir. IBM gibi şirketler, insan ve makine etkileşimini geliştirmek için birçok alanda yol kat etmektedir(Amodei, et al., 2016).

IBM, programların bir sohbetteki kişileri daha iyi ayırt etmesine yardımcı olmak için Watson ve ishalleri (konuşmayı konuşmacı kimliğine göre tanımlayan ve bölümleyen algoritmalar) kullanır. Watson canlıları blogunda Michael Picheny; "Canlı bir sohbette, konuşmacılar arasında hızlı bir şekilde ileri geri kayma oluşuyor. Bu da sistemin başka bir kişi konuşmadan önce belirli bir konuşmacıda bir konuşma modeli geliştirmesini zorlaştırıyor." der. Picheny sözlerine; "Konuşma sırasında her konuşmacının spektral frekans içeriğindeki farklılıkları tanıyacak bir sistem geliştirdik. Bir soda şişesinin üstüne üflerseniz duyabileceğiniz nüansları düşünün - ses frekanslarımız aynı şekilde çalışır… Watson'un yapabildiği şey, bu profillerin her birini anında oluşturmak ve çıktı metnini belirli hoparlörlere atamaktır." şeklinde devam eder(Chorowski, et al., 2015).

Araştırmacılar onlarca yıl boyunca, yüzde 4'lük bir kelime hata oranı olduğu tahmin edilen insan konuşma doğruluğuna ulaşmak için çalışmıştır(Diğken & İbrikçi, 2015). Son zamanlarda ise IBM, yüzde 5,5'lik yeni bir endüstri rekoru kırmıştır. Bu konu hakkında, IBM Watson blogunda George Saon; "Çok zor bir konuşma tanıma görevi ölçüldü. İnsanlar arasında araba satın almak gibi günlük konuları tartışan sohbetler kaydedildi." şeklinde bir açıklama yapmıştır. Saon sözlerini şöyle devam ettirir; "IBM araştırmacıları derin öğrenme teknolojileri uygulamamızı genişletmeye odaklandı. Uzun Kısa Süreli Bellek ve WaveNet dil modellerini üç güçlü akustik modelle birleştirdik… Son model ile ilgili benzersiz olan şey, yalnızca olumlu örneklerden öğrenmenin yanı sıra olumsuz örneklerden de yararlanmasıdır; benzer konuşma kalıplarının tekrarlandığı yer. "

2.1.6. Konuşma Tanıma Teknolojisinin Uygulamaları

Konuşma tanıma teknolojisi ve dijital asistanların kullanımı, cep telefonlarımız vasıtasıyla hızlı bir şekilde evlerimize taşınmıştır(Fu, 2019). Ayrıca bu teknolojinin; işletme, bankacılık, pazarlama ve sağlık gibi sektörlerdeki uygulamaları da hızla ortaya çıkmaktadır(Arora, & Reetz, 2017).

1. İş Yerinde

İş yerinde konuşma tanıma teknolojisi; geleneksel görevleri yerine getirmesi beklenen kişilerin görevlerinin ötesinde, verimliliği artırmak için daha basit görevlerin bir araya gelmesiyle oluşmuştur(Yu & Deng, 2016). Ofis görevlerine örnek olarak dijital asistanların yapabilecekleri:

- Bilgisayardaki raporları veya belgeleri aramak
- Verileri kullanarak bir grafik veya tablo oluşturmak
- Bir belgeye dâhil etmek istenilen bilgileri dikte etmek
- İstek üzerine belge yazdırmak
- Video konferansları başlatmak
- Toplantıları planlamak
- Tutanak dakikalarını gözetmek
- Seyahat düzenlemeleri yapmak

2. Bankacılıkta

Bankacılık ve finans sektörünün amacı, müşterinin bekleme süresini azaltmak ve kolaylık sağlamak için konuşmanın tanınmasıdır. Sesle aktive olan bankacılık, insan müşteri hizmetlerine duyulan ihtiyacı büyük ölçüde azaltabilir ve çalışan maliyetlerini düşürebilir(Ko & Peddinti & Povey, Khudanpur, 2015). Ayrıca kişiselleştirilmiş bir bankacılık asistanı müşteri memnuniyetini ve sadakatini artırabilir. Konuşma tanımanın bankacılığı iyileştirmesine örnek olarak şunlar gösterilebilir:

- Cep telefonunu açmak zorunda kalmadan bakiyeyi öğrenmek, işlemleri gerçekleştirmek ve harcama alışkanlıkları hakkında bilgi talebinde bulunmak
- Ödeme yapmak
- İşlem geçmişi hakkında bilgi almak

3. Pazarlamada

Sesli arama, pazarlamacıların tüketicilerine ulaşmalarına yeni bir boyut ekleme potansiyeline sahiptir(Han, et al., 2017). İnsanların cihazlarıyla nasıl etkileşime gireceği konusundaki değişiklikle pazarlamacılar, kullanıcı verileri ve davranışları konusunda trendler geliştirmeye çalışırlar.

- Veri: Konuşma tanıma ile, pazarlamacıların analiz etmesi gereken yeni bir veri türü meydana gelecektir. İnsanların aksanları, konuşma kalıpları ve kelime bilgisi; tüketicinin yeri, yaşı ve kültürel ilişkileri gibi demografik özellikleri ile ilgili diğer bilgileri yorumlamak için kullanılabilir.
- Davranış: Pazarlamacıların ve optimize edicilerin, bu eğilimlerin ötesinde kalabilmesi için uzun kuyruklu anahtar kelimeler sohbet içeriği üretmeleri gerekebilir. Bu tür bir hızlı arama, kullanıcıların karakterlerinde sabırsızlaşmaya yol açabilir ve interneti ana bilgi kaynağı olarak kullanmaya daha bağımlı hale getirebilir. (16) Bu nedenle kullanıcıların bir ekrana bakmak için harcadıkları zaman azalabilir. Pazarlamacılar bunun görsel içerik için ne anlama gelebileceğini düşünmelidir çünkü işitsel ve bilgi ağırlıklı içeriğe odaklanmaya doğru bir kayma olabilir.

4. Sağlıkta

Saniyelerin hayatî önem taşıdığı ve steril çalışma koşullarının öncelikli olduğu bir ortamda, “eller serbest” özelliği vasıtasıyla bilgiye anında erişim, hastanın güvenliği ve tıbbi verimlilik üzerinde önemli derecede olumlu bir etkiye sahip olabilir. Bu bağlamdaki avantajlar şunları içerir:

- Tıbbi kayıtlardan hızla bilgi bulunabilir.
- Hemşirelere süreçler hatırlatılabilir veya özel talimatlar verilebilir.
- Hemşireler, kattaki hasta sayısı ve mevcut birim sayısı gibi idari bilgiler isteyebilirler.
- Evde ebeveynler, doktora ne zaman gitmeleri ve hasta bir çocuğa nasıl bakmaları gerektiği konusunda ortak hastalık belirtileri isteyebilir.
- Evrak işlerinin azalması sağlanabilir.
- Veri girişi için daha az zamana ihtiyaç duyulabilir.
- İş akışları daha da geliştirilebilir.

Sağlık hizmetlerinde konuşma tanımayı kullanmanın en önemli kaygısı, dijital asistanın erişebildiği içeriktir. Bu alanda uygulanabilir bir seçenek olması için içeriğin tanınmış sağlık kurumları tarafından tedarik edilmesi ve doğrulanması gerektiği kabul edilmiştir (Wilcox & Bush, 2018).

5. Nesnelerin İnternetinde

Nesnelerin İnterneti[1], bir zamanlar olduğu gibi fütüristik bir olasılık değil, çevremizdeki nesnelerle ilgili bir gelişmedir(Han, et al., 2017).

Şu anda nesnelerin, internette konuşma tanıma kapsamında en belirgin uygulamalarından biri otomobillerde bulunmaktadır. Her beş otomobilden birinin 2020'ye kadar konuşma tanıma uygulaması olduğu tahmin ediliyor(Wilcox & Bush, 2018). Bunun avantajı, sürücü dikkatini daraltmak amacıyla, araçlarla etkileşime girme biçimini değiştirmesidir. Otomobillerde dijital asistan uygulamasına şunlar örnek verilebilir:

- Mesajları ahizesiz dinlemek
- Radyoyu kontrol edebilmek
- Rehberlik ve navigasyon ile yolculuk esnasında yardım almak
- Ses komutlarına cevap almak

6. Dil öğreniminde

Konuşma tanıma teknolojisinin en dönüştürücü uygulamalarından biri, insan bakış açısı vasıtasıyla, sosyal yaşamdaki dil engellerini ve kültürel sınırları ortadan kaldırma kabiliyetidir(Arnold, et al., 1994). Dil engeli olmayan bir dünya, çeşitli ülkeler ve kültürler arasında birtakım iş birliklerinin gelişme olasılıklarını artırır. Belki de artan çeşitliliğin bir sonucu olarak daha hızlı bir inovasyon oranına katkıda bulunur(Chao & Bourguet, 2017).

Konuşma tanımanın gelecek uygulamaları da çok fazla olacaktır. Bu teknoloji hâlâ büyük ölçüde bebeklik dönemindedir. Ancak tüketicilerin geçmişte olduğundan daha hızlı bir şekilde yeni teknolojileri benimseme eğiliminde oldukları “hiper benimseme” teorisi ile, bu teknolojinin hızla büyüme ve gelişme ihtimali oldukça yüksektir. Teknolojinin yaşam döngüsünün bu aşamasında, mevcut potansiyelinin ve yakın gelecekte günlük yaşamımıza ortak olma olasılığının bilincinde olmak çok önemlidir. İşletmeler, konuşma

[1] Internet of Things

tanıma teknolojisini dijital pazarlama stratejilerine ve bütçelerine dâhil etme yaklaşımlarında proaktif olmalı; bireyler ise konuşma tanımanın faydalarını araştırmaya devam etmelidir. Doğruluk oranları geliştikçe ve tüketici katılımları arttıkça, endüstrilerin daha önceki yıllarda göründüğünden daha fazla konuşma merkezli olması güçlü bir ihtimaldir. Endüstriler konuşma tanımanın sonuç aşamasında, daha fazla insan özelliği kazanmak için adapte olma ihtiyacını karşılayabilirler(Wilcox & Bush, 2018).

2.1.7. Konuşma Tanıma Yapısı

Konuşma tanıma dışında bu kavrama benzer kaç kavram vardır. Otomatik konuşma tanıma, konuşmacı tanıma[2] ve dil tanıma[3]; konuşma sinyali işleme[4] kullanımları olarak bilinmektedir. Bu kavramların hepsi, bir özellik çıkarım aşaması gibi sistemin başlangıcında bazı ortak özellikleri paylaşırlar ancak hedefleri farklıdır(Chiu, et al., 2018).

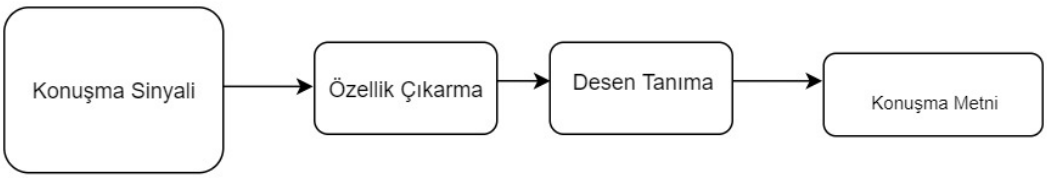
Konuşmacının tanınmasında amaç, konuşmacının kimliğini tanımaktır. Dil tanıma, konuşma dilinde hangi dilin kullanıldığını belirlemeyi hedefler. Otomatik konuşma tanıma ise, ses sinyalinin bir bölümündeki mesajın ne olduğunu, bu sinyalde ne söylendiğini belirleme sorunuyla ilgilenir. Konuşma tanıma çıkışı genellikle giriş konuşmasının bir metin versiyonudur.

Daha önce belirtildiği gibi, tüm bu söz konusu görevlerin ortak özellikleri vardır. Ancak bu tez konuşma tanıma problemiyle ilgili olacağı için derinlemesine açıklanacak olan konu konuşma tanımadır(Wilcox & Bush, 2018).

Otomatik konuşma tanıma, dil bilgisini konuşma sinyallerinden çıkarma işlemi olarak tanımlanabilir. Bu dil bilgisine "fonetik bilgi" denir. Dolayısıyla konuşma tanıma, konuşma seslerinin (telefonların) fiziksel özellikleriyle ilgilenmektedir. Bunlar; anatomik üretimi, akustik özellikleri, işitsel ayırt etme ve nörofizyolojik konum olarak sıralanabilir. Telefon; tam bir sesin kelimelerin anlamı için önemli olup olmadığını hesaba katmadan dikkate alınarak, ayrı bir konuşma sesi olarak tanımlanabilir.

[2] Speaker Recognition
[3] Language Recognition
[4] Speech Signal Processing

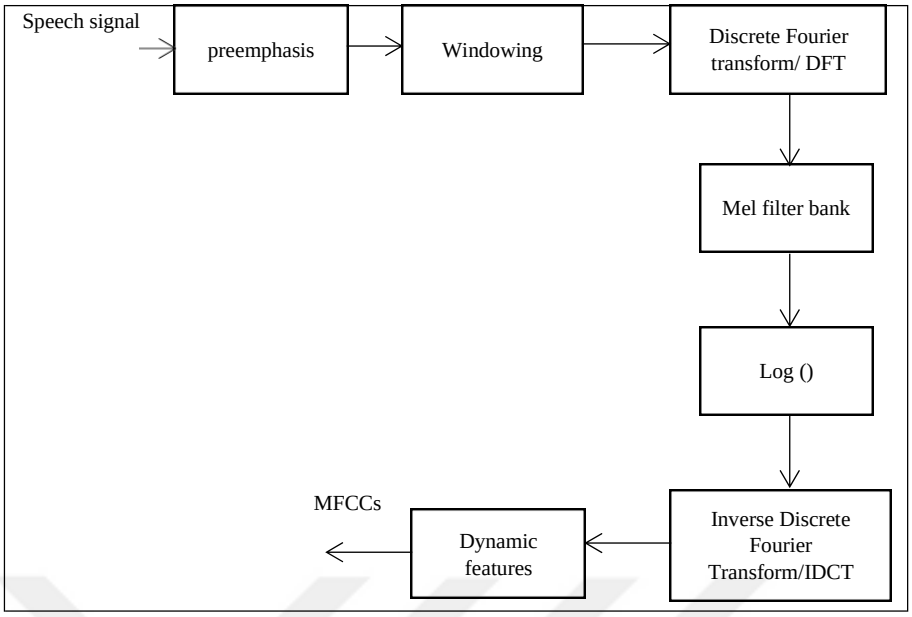
Bunun aksine eğer bir fonem, dilde alternatif bir fonemle değiştirildiyse kelimenin anlamını değiştiren bir konuşma sesi olarak tanımlanır. Yukarıdaki tanımlara göre, fonemlerin mutlak olduğunu ima eden herhangi bir dile özgü olmadığı tespit edilebilir. Öte yandan fonemler sadece belirli dillere atıfta bulunarak konuşulabilir. Türk dili, ses birimini temsil eden 8 sesli 23 sessiz olmak üzere 31 adet foneme sahiptir. Bu fonemlerin kombinasyonu, fonetik olarak bilinen akustik bir gösterimle sonuçlanır. Şekil 2.1'de basit bir konuşma tanıma sistemi verilmiştir.



Şekil 2.1: Basit Bir Konuşma Tanıma Sistemi

2.1.7.1. Özellik Çıkarma

Konuşma tanımada konuşma, özellik olarak bilinen matematiksel bir gösterime dönüştürülür. Özellikler kolay analiz ve işleme sağlar. Bilindiği gibi, konuşma sinyalleri ham bir dalga biçimindedir ve bu onları anında değiştirir. Bu değişikliklerden kaçınmanın tek yolu, konuşma sinyalini zaman alanı yerine başka bir alanda temsil etmektir. Genellikle Fourier alanı tercih edilir çünkü konuşma, frekans alanında en iyi şekilde temsil edilir. Ancak bir insanın bir konuşma sinyalini tanıması için, zaman ve frekans çözünürlüğünün bir kombinasyonu gerekir, sadece bir Fourier alanı kullanarak gerekli olmaz, çünkü konuşma sinyali içinde bulunan zamanlama bilgisini ortadan kaldıracaktır. Bu nedenle bölümlendirme tekniğinin, yani 5 ila 30 milisaniye uzunluktaki kareler olarak bilinen birçok kısa bölümün kullanılmasının nedeni budur. Özellik çıkarımı için kullanılan tüm teknikler arasında en yaygın teknik; hesaplama basitliği, düşük boyutlu kodlama ve tanıma aşamasındaki başarısı nedeniyle Mel-Frekans Cepstral Katsayıları'dır.



Şekil 2.2: Mel-Frekans Cepstral Katsayıları blok şeması

Bu elde edilen özellikler konuşma tanımada kullanılmak üzere aşağıdaki yetkinliklere sahip olmalıdır:

- Farklı türleri ayırt etmek için yeterli bilgiye sahip olunmalıdır.
- Telefonların yüksek kaliteli zaman çözünürlüğü -10ms; yüksek kaliteli frekans çözünürlüğü 20-40 kanal olmalıdır.
- Hoparlör varyasyonuna karşı esnek olunmalıdır.
- Kanal bozulmalarına veya gürültüye karşı dayanıklı olunmalıdır.
- Yüksek standart örüntü tanıma özelliği mevcut olmalıdır.

MFCC tarafından birinci mertebeden ve ikinci mertebeden hesaplandıktan sonra; MFCC zamansal farklılıkları, örüntü tanıma sistemine giriş olan son bir vektör elde etmek için birleştirilir(Wilcox & Bush, 2018).

2.1.7.2. Desen Tanıma

Örüntü tanıma görevleri farklı düzeylerde gerçekleştirilebilir. Bunlara örnek olarak telefonlar, üç telefonlar veya kelimeler verilebilir. Bu tezde telefon tanıma üzerine çalışılmaktadır.

Bir ASR'nin model tanıma aşaması üç ana bileşenden oluşur. Bunlar; dilin seslerinin özelliklerini veren “akustik bir model”, telaffuzlarıyla birlikte kullanılan kelimeleri veren

fonetik bir “sözlük”; konuşulabilecek kelime dizileri hakkında bilgi veren “dil modeli”dir. Bu üç bileşenin elde edilmesi, fazla miktarda ses verisine ve metinlere istatistiksel işlemlerin uygulanmasını gerektirir(Wilcox & Bush, 2018).

2.1.8. Konuşma Tanıma Neden Zordur?

Dinlemek, düşünülenden daha zor ve daha karmaşıktır. Bu yüzden konuşma tanıma işlemi de oldukça zordur. Bu eyleme yönelik olarak makinenin ve kullanıcının yapması gerekenler şunlardır:

1- Bir makine, analog sinyali (akustik dalga) dijital gösterime dönüştürmek zorundadır. Dolayısıyla fizyoloji ve anatomi, bir akustik dalgayı kabul etmektedir.

2- Örneğin restoranda birisi konuştuğunda, onun sözlerini - ki bu sözlere sinyal denilmektedir- tüm arka plan gürültüsünden ayırmak zorunludur. Gürültü, telefon görüşmeleri, oda akustiği, diğer kişilerin konuşmaları, trafik gibi farklı biçimlerde olabilir. Bir makine, sesi bu gürültülerden ayırmak zorundadır.

3- İnsanlar zaman zaman çok hızlı veya yavaş konuşabilirler. Yeni bir cümleye başlamadan, bir önceki cümlenin sonunda duraklamamaktadırlar. Cümleler makineye sürekli ve uzun bir sözcük dizisi gibi gelmektedir Bu tür durumlarda cümle yapısını sadece ses tonundan "duymak zor" dur. Ayrıca bir kelimenin ne zaman bitip diğerinin başlayacağı belirsiz içerisindedir. Makine ise bu son noktaları baz alır.

4- Herkes, birbirinden farklı görünümlere sahiptir. Yaş, cinsiyet, aksan, stil, kişilik, bağlam, niyet gibi durumların hepsi ses ve konuşmayı etkiler. Sesler her seferinde aynı şekilde düzenli olarak değişir. Ayrıca bir makine konuşma tanırken yaş, cinsiyet, aksan gibi farklılıkları hesaba katmalıdır.

5- Bir varsayım olarak; hepsi farklı lehçeler konuşan ya da birbirinden farklı aksanları olan, 9 yaşından 90 yaşına kadar zengin bir yaş yelpazesine sahip bir grubun var olduğu düşünülsün. Bu insanlar konuştuğu zaman, temel konuşma akışına ve herkesin ne dediğine önem vermek gerekir. Örneğin "kedi" kelimesini kim söylerse söylesin aynı anlama geldiğini algılamak gereklidir. Dolayısıyla bir makine, "kedi" ve "cat" örneğinde olduğu gibi, sesler farklı söylense bile bunu ayırt etmek zorundadır.

6- “To", "de", "iki" gibi eş sesli birçok sözcük vardır. Ancak birbirinden çok farklı anlamlar ifade edebilirler. Bunlara homofon denilir. Konuşmacının hangi kelimeyi hangi anlamda kullandığını bilmek gerekir. Dolayısıyla bir makine, homofonları ayırt etmek zorundadır.

7- Konuşma esnasında kullanılan "um", "hmm" ve buna benzer birçok dolgu bulunmaktadır. İnsan, içgüdüsel olarak bunları nasıl filtreleyeceğinin bilincindedir. Bunlar, konuşmacının sözlerinin yanlış yorumlanmasına neden olmaz. Ancak bir makinenin bu dolguları filtrelemesi gerekmektedir.

8- Konuşma esnasında sıklıkla yanlış anlaşılmalar yaşanmaktadır. Yanlış anlaşılmalar, cümlelerin yanlış duyulması sonucu meydana gelir. Örneğin bir adres sorulduğu zaman, karşı taraf yanlış anlaşılma sonucu birçok farklı sokak adı tahmin edebilir ancak bunların tamamı yanlış çıkabilir. Dolayısıyla bir makine de bu tür yanlış anlamaları yönetmek zorundadır ve bu görevde insan zihninden çok daha iyi olmaları gerekmektedir.

9- Son olarak, içeriğe dikkat edildiği kadar, kullanılan dilin söz dizimine ve semantiğine de hâkim olmak gerekmektedir.

Bunların hepsini günlük yaşamda insanın kolaylıkla gerçekleştirmesi oldukça dikkat çekicidir. Bu açıdan, insan beyninin inanılmaz bir düşünme ve konuşma yetisinin olduğunu söylemek mümkündür. Makineler ise yalnızca onun bu yetilerinin bir kısmını gerçekleştirmek için büyük bir çaba sarf etmektedir. Yine de konuşma tanıma bugüne kadar uzun bir yol kat etmiştir ve bu sadece başlangıç da olabilir. Konuşma, insan için en doğal iletişim şeklidir. Günümüzde ise artık makineler insan konuşmasını tanımaya başlamakta ve onunla iletişim kurmada gittikçe daha iyi hale gel mektedir.

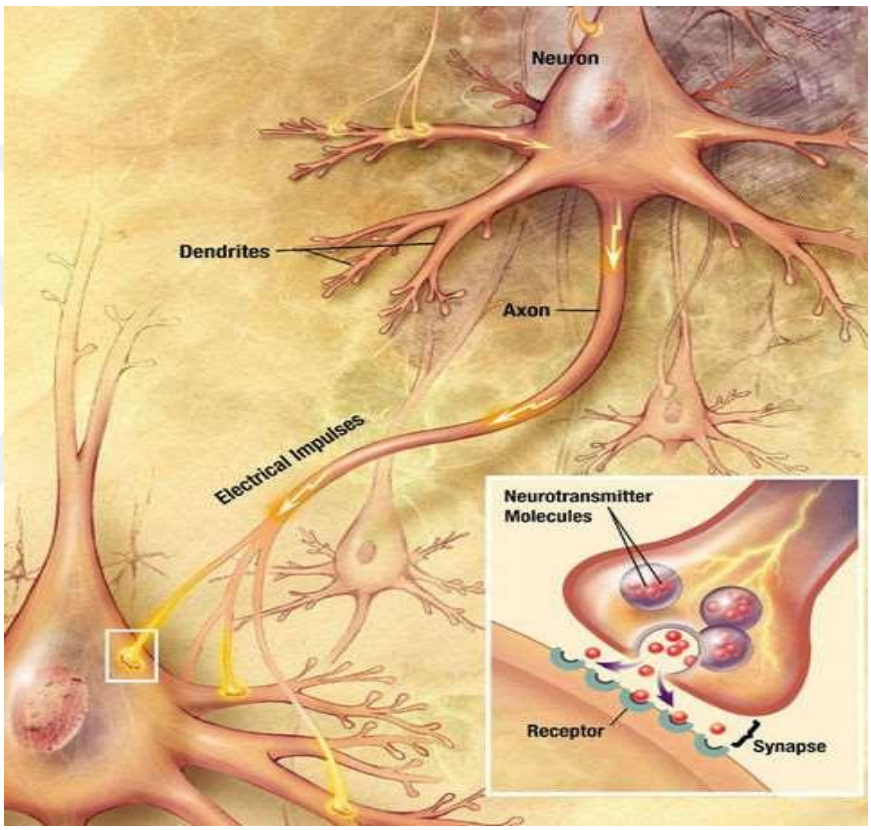
Amazon Alexa ve Google Home gibi mevcut ses asistanları ve aygıtlar her ay giderek daha popüler hale gelmektedir. Kişinin alışveriş yapma şeklini, arama şeklini, cihazlarla ve bizzat insanların birbirleriyle nasıl iletişim kurduğunu değiştirmektedirler.

2.2.Derin Öğrenme Tanımı

Derin öğrenmeyi anlamak için Sinir ağları konusunu iyi bilmek gerekir, bu yüzden bu bölümde Sinir ağları hakkında bilgi verilecektir.

2.2.1. Sinir Ağları

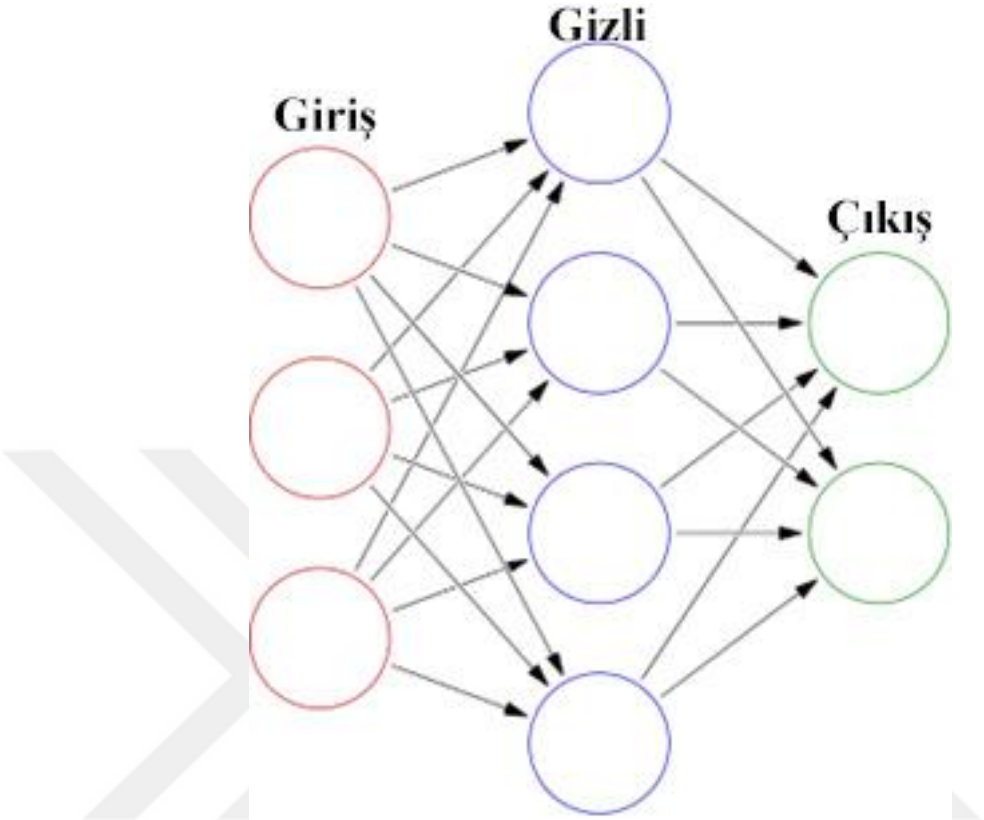
Yapay sinir ağları, görüntü işleme ve otomatik konuşma tanıma gibi örüntü tanıma görevleri için birçok alanda kullanılan çok güçlü sınıflandırıcılardır. Sinir ağı beynimizin yapısına dayanarak kurulur, bir insanın beyni yaklaşık 10 milyar nöron ve nöronları bağlayan 60 trilyon bağlantıdan (sinaps) oluşur. Beyin sürecini taklit eden bu matematiksel modelin çok büyük tanıma yetenekleri vardır. Şekilde biyolojik bir nöron temsili olarak gösterilmektedir (Deng, & Yu, 2014).



Şekil 2.3: Biyolojik bir nöronun temsili

Sinir ağı, insan nörobiyolojik sisteminden esinlenilen bir bilgi işlem şablonudur ve yeni tür sinir ağı, bilgi işlemi için yeni bir yapı kullanır. Yapay sinir ağları çeşitli alanlarda kullanılmaktadır. Üç katmanlı sinir ağlarında ilk katman giriş katmanıdır, girdi verileri daha sonra sinir ağına girilir, ikinci katman ise ilk katmanın kenarlarıyla birbirine bağlanan gizli katmandır. Bu kenarların her biri, bu ağırlıkların eğitim aşamasında değişen öngörücü ağırlığa sahiptir. Üçüncü katman, kenarlarla gizli katmana bağlanan çıkış katmanıdır. İkinci ve üçüncü katmanlar arasındaki kenarlar aynı zamanda egzersiz sürecinde değişen önceden ayarlanmış değerlere sahiptir, aşağıdaki şekil katmanları

göstermektedir. Şekil 1. Bir sinir ağı ve katmanları arasındaki ilişkileri göstermektedir(LeCun, Bengio & Hinton, 2015).



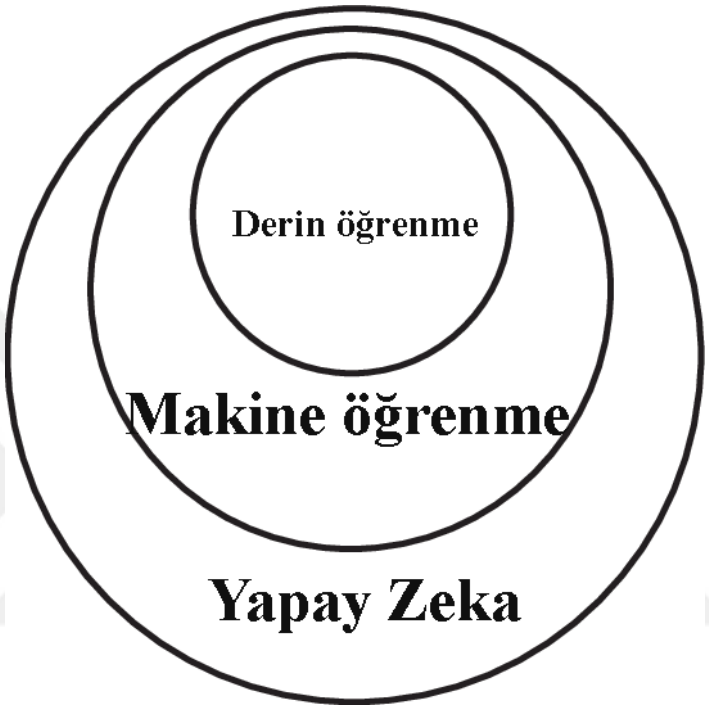
Şekil 2.4: Bir sinir ağı ve katmanları arasındaki ilişkiler

Ağ, giriş bilgisine ve çıkış verilerine göre öğrenir. Eğitim yapılırken, gizli katman kenarlarının ve çıktı katmanının ağırlıkları, gerçek çıktı verileriyle beklenen çıktı farkına göre belirlenir. Eğitim yapıldığı zaman yeni girdi verileri, kenarların ağırlığına göre sınıflandırılır(Najafabadi, et al., 2015).

Sinir ağları, özellikle giriş verisi sayısı yüksek olduğunda ve veriler gürültü içerdiğinde çok yararlıdır. Sinir ağının güçlü yönlerinden biri, gürültü verilerini analiz etme yeteneğidir. Yapay sinir ağları, bu avantajların yanı sıra problemleri de vardır, bu yöntemlerin kusurlarından biri ağ yapısının anlaşılmazlığıdır. Diğer bir dezavantaj, makine öğrenim algoritmasında izlenen diğer öğrenme yöntemlerinden daha uzun öğrenme süresidir(LeCun, Bengio & Hinton, 2015).

2.2.2. Derin öğrenme

Derin öğrenme, makine öğrenmesi araştırmalarında yeni bir alandır. Bu tür öğrenmeler aynı zamanda derin yapılandırılmış öğrenme veya hiyerarşik öğrenme olarak da adlandırılır, ancak genellikle derin öğrenme bu alanı adlandırmak için kullanılır(Deng, 2014). Bu bölüm bu algoritmayı tartışacaktır.



Şekil 2.5: Yapay zekâda derin öğrenme algoritmasının yeri

Makine öğrenmesi algoritmaları, veri madenciliği ve yapay zekâ alanlarında yaygın olarak kullanılmaktadır. Makine öğrenmesi algoritmaları yapay zekânın bir alt kümesi olarak tanımlanabilirler. Bununla birlikte, derin öğrenme algoritmaları makine öğrenme algoritmalarının alt kümesidirler(Deng, & Yu, 2014)..

Son yıllarda derin öğrenme, görüntü ve bilgi işleme araştırmacıları tarafından büyük ölçüde kabul görmüştür. Böylece, Derin öğrenmenin birçok tanımından bahsedilmiştir: Bu tanımlardan biri:

Derin öğrenme; bilgisayarların, deneyimlerden öğrenmelerini ve dünyayı kavramların hiyerarşisi açısından anlamalarını sağlayan bir makine öğrenimi olarak tanımlanmıştır (Lecun & Bengio & Hinton, 2015).

Denetimli veya denetimsiz özellik çıkarma, dönüştürme, desen analizi ve sınıflandırma için birçok doğrusal olmayan gizli katmandan yararlanan bir makine öğrenme teknikleri sınıfı olarak tanımlanmıştır.

İnsan beyninin son derece karmaşık problemler için gözlemleme, analiz etme, öğrenme ve karar verme yeteneğini taklit etmeyi amaçlayan, büyük miktarda denetimsiz veri kullanan bir makine öğrenmesi olarak tanımlanmıştır(Gu & Zhang & Zhang, Kim, 2016).

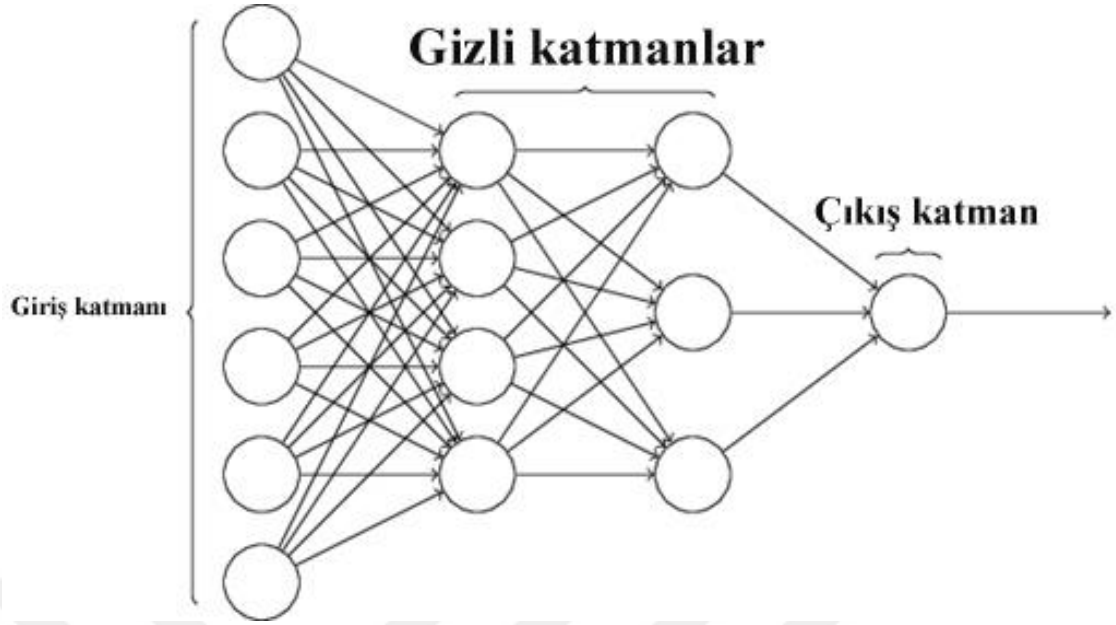
Şekil 2.5'de Yapay Zekâda, Makine Öğrenimi ve Derin Öğrenme ilişkisi farklı bir bakışla anlatılmaktadır (Süzen & Kayaalp, 2018).

Yapay zeka[5], ilk olarak yirminci yüzyılın ortalarında açığa çıkmıştır. Yapay zeka makinaların insanlar kadar becerili bir şekilde belirli işlemleri yapabilmesi olarak tanımlanabilir. Bazı yapay zeka makinaların sadece programladığınızı yerine getirirken başka yapay zeka makinaların algoritmik hesaplarda bulunarak programladığınızı iyileştirebilen, hatalardan öğrenebilen sistemlerdir.

Makine öğrenimi[6], 1980'lerde ortaya çıkmıştır ve yıllar sonra veri madenciliğin popüler hale gelmesi ile beraber kullanılması artmıştır. Sunmuş olduğunuz veriler ve parametreler ile benzetimler yaparak, sizden daha iyi tespitlerde bulunan, programlamadıklarınızı da açığa çıkarabilen, kendi kendini eğitebilen sistemlerdir.

Derin öğrenme[7], çok yeni bir kavramdır ve yapay zeka ve makine öğrenimi kadar tarihçesi yoktur. Derin öğrenme büyük veri denizi ile tek bir katmanda değil, birçok katmanda makine öğreniminde kullanılan hesapları tek bir seferde yapan, makine öğreniminde tanımlamanız gereken parametreleri bile kendisi keşfeden, belki de daha iyi parametreler ile değerlendirmelerde bulunabilen bir sistemdir(Deng, & Yu, 2014).

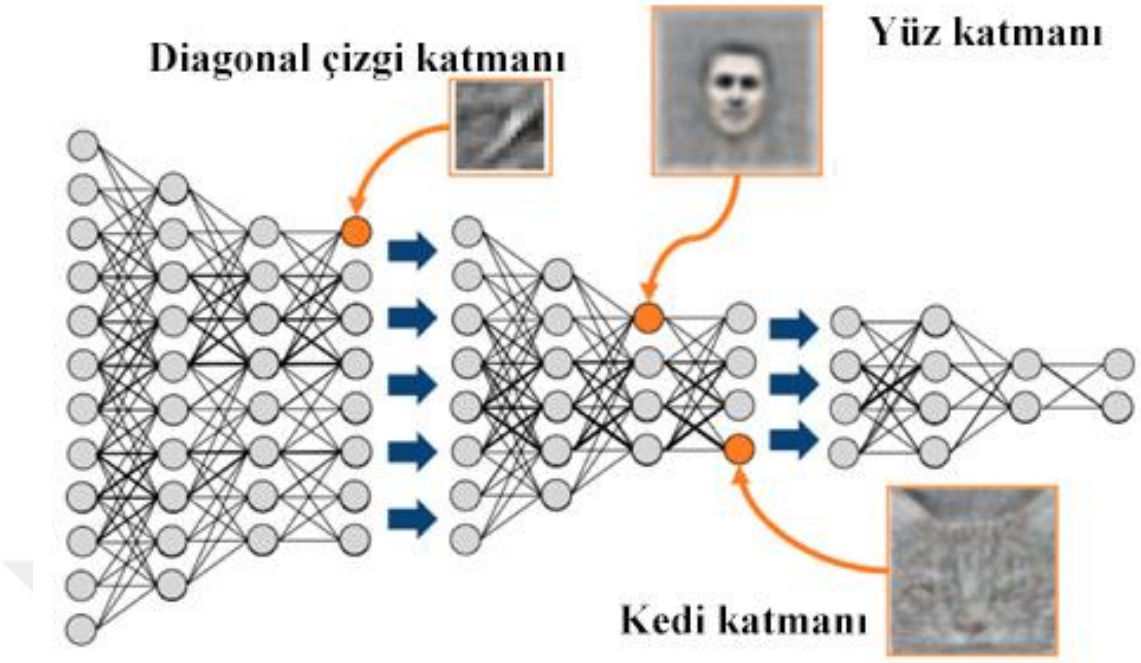
[5] Artificial Intelligence
[6] Machine Learning
[7] Deep Learning



Şekil 2.6: Derin öğrenme algoritmasının yapısı

Derin öğrenme algoritmaları, insan beyninin prensibinden ve görsel korteksin insan beyninde nasıl çalıştığından ilham alırlar. İnsan beyninde, beynin görsel korteksindeki birincil hiyerarşideki nöronlar, kenarlara ve kütlelere duyarlı olarak aldıkları bilgiyi alırlar ve çıktıları, nöronlar daha karmaşık yapılara dönüşene kadar sonraki bir hiyerarşide devam eder. Bu iş parçacığında, beyin fonksiyonunun simüle edildiği sinir ağı algoritmasının bir alt kümesi vardır(Najafabadi, et al., 2015).

Şekil 2.6'de, bir kedinin görüntüsünü tanımlamak için derin bir öğrenme algoritması kullanılmaktadır.



Şekil 2.7: Bir kedinin görüntüsünü tanımlamak için kullanılan derin öğrenme algoritmasının yapısı

Şekil 2.7'de görebileceğiniz gibi, bu görüntü sinir ağlarını kullanarak derin bir öğrenmedir. Bu sinir ağı, kedinin kafası için bir katmanın yüz çizgileri için bir katmanı ve tüm kedi görüntüsü için bir katmanı olarak görev yapar. Bir kedi ya da erkeğin görüntüsünün bu şekilde analiz edilmesini kolaylaştıran farklı katmanlar halinde sunulur(Deng, & Yu, 2014).

Sinir ağının ve derin sinir sisteminin nüanslarını açıklamak gerekir. Geleneksel sinir ağlarının gücü, paralel olarak işlem birimi sayısına göre hesaplanır, aslında, işlem birimlerinin paralel bağlanması ne kadar olursa olsun, sinir ağının gücü daha yoğundur, o yıllarda, sinir ağlarının paralel işlenmesi en büyük avantajıydı olarak bilinmektedir.

Yapay sinir ağları, uzun yıllar boyunca çeşitli alanlarda karmaşık sorunları başarıyla çözmüştür, ancak zamanla ve daha karmaşık sorunların ortaya çıkmasıyla araştırmacılar, derin öğrenme ağlarının tasarımına yönelmiştir.

Derin öğrenme ağlarına geçmenin temel nedeni, sıradan sinir ağlarının güçlü modeller için sağlayamadığı karmaşık doğrusal olmayan sorunların ortaya çıkmasıydı. Bir olayda, derin öğrenme ağında çözülen sorun, üniteleri gösterebilmeyi gerektiriyor . İki çözülebilir bir sinir ağında daha küçük problemleri çözebilen tipik bir sinir ağı, derin bir öğrenme

ağı vardır, bu yüzden karmaşık problemlerdeki derin öğrenme ağının geleneksel sinir ağından çok daha fazla yeteneği olduğu ilkesine dayanır(Deng, & Yu, 2014).

İlk denemede (2006 yılında yapılan bir araştırmaya göre), derinlemesine oturmuş bir öğrenme ağı el yazısı sayılarını büyük bir veri setinde sınıflandırmak için geleneksel sinir ağından çok daha iyi bir performans göstermiştir(Deng, & Yu, 2014).

Aslında, derin öğrenme ağları ve sinir ağları arasında iki temel fark vardır:

1- Ara katmanların sayısı

Geleneksel sinir ağlarında, bazı durumlarda ara katmanların sayısı verimliliği azaltabilir, aslında, katman sayısındaki artışla, sinir ağının işlevi de azalır, ancak derin öğrenme ağlarında ara katların sayısı sınırlı değildir ve bu katmanların sayısı Modelin performansını olumlu yönde etkiler. Yapay sinir ağlarının geleneksel ağlarında katman sayısı arttıkça her çalışmadaki her katman ağırlığının güncellenmesi büyük bir problemdir.

2. Doğrusal olmayan aktivatör fonksiyonları

Yukarıda belirtildiği gibi, sinir ağları üzerindeki en büyük kısıtlamalardan biri doğrusal olmayan problemlerin karmaşıklığıdır, bu nedenle derin öğrenme ağlarında, doğrusal olmayan aktivatör fonksiyonlarını kullanarak karmaşık problemlerin modellenmesi mümkündür.

Derin kelime, derin öğrenme ağlarında kullanılır. Aslında, soru şu ki, bir katmanı olan bir öğrenme ağının derin olabileceği, derin bir öğrenme ağının iki katmanı olan bir ağ mı?

Derin öğrenme ağlarında, derin kavramın belirsizlikleri vardır. Bu bölümde, bir ağın derin öğrenme ağı katmanlarının sayısı ile nasıl adlandırıldığını açıkça ortaya koyuyoruz.

Tablo 2.1: Derin Öğrenme Ağındaki Derinlik Kavramı

İsim	Katman sayısı
Derin	+3
Çok Derin	+16
Aşırı Derin	+1000

Tablo 2.1'de derin öğrenme ağındaki derinlik kavramı açıklanmaktadır. Orta katmanlar bir ağda iki kattan fazlaysa, derin olarak adlandırılabilir.

Aşağıda, bu algoritmanın bazı olumlu ve olumsuz yanları açıklanacaktır.

Derin öğrenme sinir ağı algoritmasının avantajlarından biri bu algoritmaların kullanım kolaylığıdır. Katmanların yapısına ve her katmanın ağırlığına hiçbir şekilde dahil olmadan kolayca kullanılabilir. Bu avantaj, derin öğrenme sinir ağı algoritmasını araştırmacılar tarafından daha saygın hale getirmiştir. Bu algoritmanın bir diğer önemli avantajı, tüm fonksiyonları kabaca tahmin edebilmesidir. Bu algoritmanın diğer avantajları aşağıda anlatılmaktadır(Suzen & Kayaalp, 2018).

Pozitif:

Sinir ağı algoritmasının kolay kullanımı konularında derin bir öğrenmedir.
Tüm fonksiyonları kabaca tahmin edebilir.
Görüntü işleme gibi karmaşık ve karmaşık sorunlara karşı çok iyi çalışır.
Düzenli yapı katmanlar arasındadır.
Çok sayıda eğitim veri seti varsa, çok başarılıdır.

Bu algoritmanın zayıf yönleri, sonuçların yorumlanmamasıyla açıklanabilir.

Negatif:

Diğer algoritmalara kıyasla, regresyon algoritması basit problemlere göre daha az başarılıdır.
Sonuçlar ciddi bir yorum sağlayabilir.

2.2.3. Derin Öğrenme Uygulamaları

Derin Öğrenme, teknolojilere bakış açımızı değiştiriyor. Yapay zeka konularında, şu anda makine öğrenimi ve derin öğrenme gibi yöntemler fazlasıyla kullanılmaktadır.

Yakın gelecekte birçok derin öğrenme uygulamasının yaşamınızı etkileyeceği tahmin edilmektedir. Önümüzdeki beş ila 10 yıl içinde, derin öğrenme geliştirme araçları, kütüphaneler ve diller, her yazılım geliştirme araç setinin standart bileşenleri olacaktır.

1. Otomatik metin üretimi

Bir metin otomatik olarak derin öğrenme algoritmaları kullanılarak üretilir.

Derin öğrenme algoritmaları, hecelemeyi, noktalamayı, cümleleri oluşturmayı ve hatta metnin stilini yakalamayı öğrenebilir. Büyük tekrarlayan sinir ağları, giriş dizgileri dizisindeki öğeler arasındaki ilişkiyi öğrenmek ve ardından metin üretmek için kullanılır.

2. Sağlık

Meme veya deri kanseri teşhisi artık derin öğrenme algoritmaları kullanılarak yapılmaktadır. Biobank verilerine dayanarak tahmin ve kişiselleştirilmiş tıp artık hayal değil. Yapay zeka, yaşam bilimlerini, ilaçları ve sağlık hizmetlerini bir endüstri olarak tamamen yeniden şekillendiriyor. Yapay zeka'daki yenilikler, hassas tıp ve nüfus sağlığı yönetiminin geleceğini inanılmaz şekillerde ilerletiyor. Bilgisayar destekli algılama, kantitatif görüntüleme, karar destek araçları ve bilgisayar destekli tanı önümüzdeki yıllarda büyük rol oynayacaktır(Mamoshina, Vieira, Putin, & Zhavoronkov, 2016).

3. Sesli arama ve sesle aktifleşen asistanlar

Derin öğrenmenin en popüler kullanım alanlarından biri sesle arama ve sesi aktive eden akıllı asistanlardır. Büyük teknoloji devleri zaten bu alanda önemli yatırımlar yapmış, hemen hemen her akıllı telefonda sesle çalışan asistanlar bulunabilir. Bu alan bizim yapacağımız konuşma tanıma çalışmamızada çok ilgilidir. Apple'ın Siri'i Ekim 2011'den bu yana piyasada. Android için sesle etkinleşen yardımcısı Google Now, Siri'den bir yıldan daha az bir süre sonra piyasaya sürüldü. Sesle çalışan akıllı asistanların en yenisi Microsoft Cortana olarak bilinmektedir(Wang, 2015).

4. Sessiz filmlere otomatik olarak ses eklemek

Bu görevde, sessiz bir videoyu eşleştirmek için sistem sesleri sentezlemelidir. Sistem, farklı yüzeylere çarpan ve farklı sesler yaratan bir baget sesine sahip 1000 video örneği kullanılarak eğitilmiştir. Derin bir öğrenme modeli, sahnede olup bitene en uygun olanı seçmek için video karelerini önceden kaydedilmiş seslerin bir veritabanı ile ilişkilendirir.

Sistem daha sonra insanların hangi videonun gerçek veya sahte sesleri olduğunu belirlemek zorunda kaldığı bir kurulum gibi bir turing testi kullanılarak değerlendirebilir(Wang, 2015).

5. Otomatik makine çevirisi

Otomatik makine çevirisi, bir dile verilen sözcüklerin, cümlenin veya cümlenin otomatik olarak başka bir dile çevrilmesidir. Otomatik makine çevirisinde gelişme, uzun süredir devam etmektedir. Bu alan iki bölüme ayrılır:

- Metinlerin otomatik çevirisi
- Resimlerin otomatik tercümesi

Otomatik makine çevirisi derin öğrenme algoritmaları kullanrark daha başarılı sonuçlar sunmuştur(Ker, Wang, Rao, & Lim, 2017).

6. Sürücüsüz Arabalar

Bu tür sürücü yardımı hizmetlerinin yanı sıra, Google'lar gibi tam gelişmiş sürücüsüz otomobilleri inşa eden şirketler, bir bilgisayara duyuların yerine dijital sensör sistemlerini kullanarak araba sürmenin kilit kısımlarını (veya tamamını) nasıl devralmaları gerektiğini öğretmelidir. Bunu yapmak için şirketler genellikle büyük miktarda veri kullanarak eğitim algoritmalarıyla işe başlarlar. Derin öğrenme algoritmaları sürücüsüz arabalardan çok kullanılmaktadır(Pouyanfar, et al., 2019).

7. Görüntü Tanıma

Derin öğrenmeyle ilgili bir diğer popüler alan ise görüntü tanımadır. Görüntüdeki insanları ve nesneleri tanımayı, içeriğini ve bağlamını anlamayı amaçlar. Görüntü tanıma zaten oyun, sosyal medya, perakende satış, turizm vb. birçok sektörde kullanılmaktadır(Pouyanfar, et al., 2019).

8. Konuşma Tanıma

Derin öğrenme algoritmaları konuşma tanımada da çok başarılı olduğu bilinmektedir. Konuşma tanıma, bir makinenin veya programın, konuşulan dilde sözcükleri ve ifadeleri tanımlama ve bunları makinede okunabilir bir biçime dönüştürme yeteneğidir. Derin öğrenme algoritmaları kullanılarak başarılı modeller geliştirilmiştir(Ker, Wang, Rao, & Lim, 2017).

9. Otomatik Görüntü Altyazısı Oluşturma

Otomatik resim yazısı, bir resim verildiğinde, sistemin resmin içeriğini tanımlayan bir resim yazısı oluşturması gereken alandır.

2014 yılında, bu sorunla ilgili çok etkileyici sonuçlar elde eden derin öğrenme algoritmaları patlaması olmuş ve fotoğraflarda nesne sınıflandırma ve nesne tespiti için en iyi modellerden yapılan çalışmalardan yararlanılmıştır.

Fotoğraflardaki nesneleri algıladıktan ve bu nesneler için etiketler oluşturduktan sonra, bir sonraki adımın bu etiketleri tutarlı bir cümle açıklamasına dönüştürmek olduğunu görebilirsiniz(Ker, Wang, Rao, & Lim, 2017).

10. Otomatik Renklendirme

Görüntü renklendirme, siyah beyaz fotoğraflara renk ekleme çalışmasıdır. Görüntüyü renklendirmek için fotoğraftaki nesneleri ve içeriklerini kullanmak için derin öğrenme kullanılabilir, tıpkı bir insan operatörünün soruna yaklaşması gibi düşünülebilir. Bu yetenek, ImageNet için eğitilmiş ve görüntü renklendirme sorunu için ortak çalışılan yüksek kaliteli ve çok büyük evri sinir ağlarından yararlanır. Genel olarak, yaklaşım, çok büyük evrişimli sinir ağlarının ve görüntüyü renk ekleyerek yeniden yaratan denetimli katmanların kullanılmasını içermektedir(Pouyanfar, et al., 2019).

11. Reklamcılık

Reklam sektörü, derin öğrenmeyi kullanan başka bir alandır. Derin öğrenme algoritmaları, hem yayıncılar hem de reklamverenler tarafından, reklamlarının alaka

düzeyini artırmak ve reklam kampanyalarının yatırım getirisini artırmak için kullanılmıştır. Örneğin, derin öğrenme, reklam ağları ve yayıncıların veriye dayalı tahminli reklamcılık, reklamları için gerçek zamanlı teklif, tam olarak hedefli görüntülü reklamcılık ve daha fazlasını oluşturmak için içeriklerini kullanmalarını mümkün kılmaktadır(Pouyanfar, et al., 2019).

12. Depremlerin Tahmini

Derin öğrenme algoritmaları depremlerin tahmininde başarılı sonuçlar sunmuştur. Harvard bilim adamları, viskoelastik hesaplamaları yapmak için bir bilgisayarı öğretmek için Derin Öğrenme'yi kullandılar, bunlar depremlerin tahminlerinde kullanılan hesaplamalardır(Pouyanfar, et al., 2019).

13. Beyin kanseri tespiti

Fransız araştırmacılardan oluşan bir ekip, ameliyathanede aydınlatmanın etkileri nedeniyle, ameliyat sırasında invaziv beyin kanseri hücrelerinin tespit edilmesinin zor olduğunu belirtmiştir. Nöral ağların operasyonlar sırasında Raman spektroskopisi ile birlikte kullanılmasının, kanserli hücreleri daha kolay tespit etmelerini ve ameliyat sonrası rezidüel kanseri azaltmalarını sağladığını bulmuşlardır. Aslında bu parça, son birkaç haftanın üzerinde, çeşitli kanser ve tarama cihazları ile gelişmiş görüntü tanıma ve sınıflandırma ile eşleşen konulardır(Ker, Wang, Rao, & Lim, 2017).

14. Finans

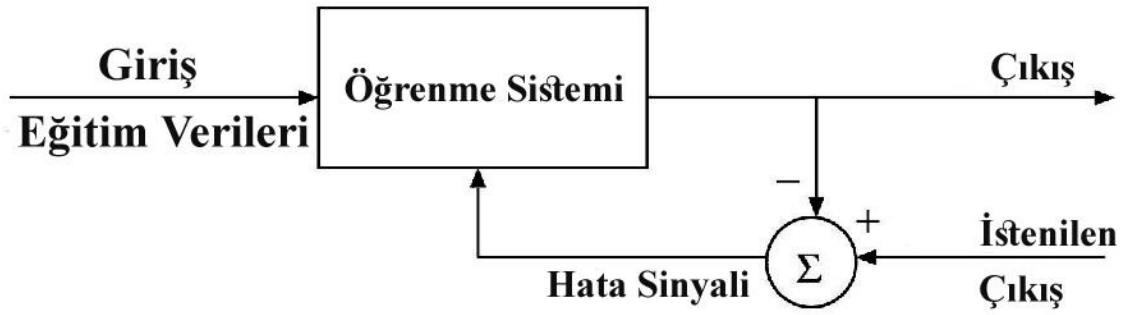
Derin öğrenme algoritmaları vadeli işlem piyasalarında, son dört yılda hem gelişmiş hem de gelişmekte olan ülkelerdeki kuruluşlarından bu yana olağanüstü başarılar elde etmiştir. Bu başarı, vadeli işlemlerin piyasa katılımcılarına sağladığı büyük kaldıraç oranına bağlanabilir (Ker, Wang, Rao, & Lim, 2017).

15. Enerji piyasası fiyat tahmini

Araştırmacılar, fiyat ve kullanım dalgalanmalarını tahmin etmek için enerji şebekesine derin öğrenme algoritmaları uyguladılar. Bölge için günlük ve günlük pazarlar, ertesi gün satış ve elektrik alım işlemlerinin yapıldığı günlük bir oturumda ve günlük uygulanabilirlik zamanlamasını takip eden saatlerde ortaya çıkabilecek enerji arz ve talebini dikkate alan altı gün içi oturumda düzenlenir. günlük seanstan sonra. Kısacası, tüketim ve kullanılabilirlik modellerine dayanarak yeterli tahminler yapabilmek, çok daha yüksek verimlilik ve maliyet tasarrufu sağlaması tespit edilmiştir(Mamoshina, Vieira, Putin, & Zhavoronkov, 2016).

3. YÖNTEM

Bu çalışmada denetimli bir öğrenme süreci kullanılmıştır. Konvolüsyonel bir derin öğrenme algoritması kullanılmıştır.



Şekil 3.1: Denetimli bir öğrenme süreci örneği

Şekil 3.1'de görüldüğü gibi, denetlenen yöntemlerin genel eğilimi tanımlanabilir. Bu yöntemlerde, ilk adım eğitim veri seti adı verilen etiketli veri kümesiyle başlar, daha sonra sınıflandırma modelinin seçilen algoritmasına dayanır. Bu işlem, sınıflandırma modeli makul ve kabul edilebilir bir doğruluk elde edene kadar devam etmektedir (Ghosh-Dastidar & Adeli, 2009).

Konuşma tanıma alanında araştırmacıları yönlendiren gözlemsel yöntemlerin güçlü yanları, önemli ve büyük etiketli bir veri seti varsa, sınıflandırmada en iyi performansı sağlamalarıdır. Denetlenen yöntemler modellerini eğitmek için etiketli veri setine erişimleri varsa, modelin kabul edilebilir performansı sağlanabilir. Örneğin, etiketli veri kümesine erişimle ilgili bir sorun varsa, denetlenen yöntemler kullanılarak güçlü modeller oluşturulabilir, ancak etiketli veri kümesi mevcut değilse, bu yöntemler zorlaşacaktır (Kashima, Kato, Yamanishi, Sugiyama, & Tsuda, 2009).

Denetimli yöntemler için ifade edilen güçlü yönlere ek olarak, bu yöntemlerin zayıflıkları da bu yöntemlerin kullanılmasının zorluğuna bağlanabilir. Bu yöntemlerin dezavantajlarından biri, model oluşturmak için etiketli veri kümesine ihtiyaç duymalarıdır. Etiketli veri kümesinde bir sorun yoksa, kullanım kısıtlanacaktır. Bu araştırmada, derin öğrenme algoritmaları kullanarak modeller oluşturmak için kullanılabilecek binlerce ses dosyası içeren çok iyi bir veri seti bulunmaktadır(Kashima, Kato, Yamanishi, Sugiyama, & Tsuda, 2009).

Denetimli yöntemleri kullanma yaklaşımı, konuşma tanıma konusundaki en eski ve en ünlü yaklaşımdır. Bu yaklaşımda konuşma tanıma, yapay zekanın bir alt kümesi olarak tanıtılır. Dolayısıyla, veri sınıflandırma problemlerinin çözümü için önerilen aynı mantık bu konu için de geçerlidir. Veri sınıflandırma, belirtilen gruplara sunulan verileri sınıflandırmak için bir sınıflandırma modeli oluşturmayı amaçlamaktadır. Örneğin, haberlerin sınıflandırılmasında, haberleri farklı sporlara, ekonomik, eğitimsel, kültürel ve diğer gruplara sınıflandırmak için bir model sunmaya çalışan durumlar vardır. Etiketli veri kümesi de modeli oluşturmak için kullanılır. İlk olarak, bir etiketli veri kümesi öğrenme makinesi algoritmalarından birine verilir ve daha sonra bu etiketli veri setine dayanarak, öğrenme makinesi algoritması bir sınıflandırma modeli oluşturmaya çalışır. Aşağıdakiler yeni veri setinin etiketli veri setine dayanarak oluşturulan sınıflandırma modeline göre sınıflandırılır. Bu örnek veri madenciliğinde klasik bir sınıflandırmayı göstermektedir. Dolayısıyla bu fikir gözlem yaklaşımında ana tema olarak kullanılmaktadır (Xu, Zeng, & Zhong, 2013). Derin öğrenme algoritmasının konvolüsyonel algoritması da denetlenen algoritmalardan biridir.

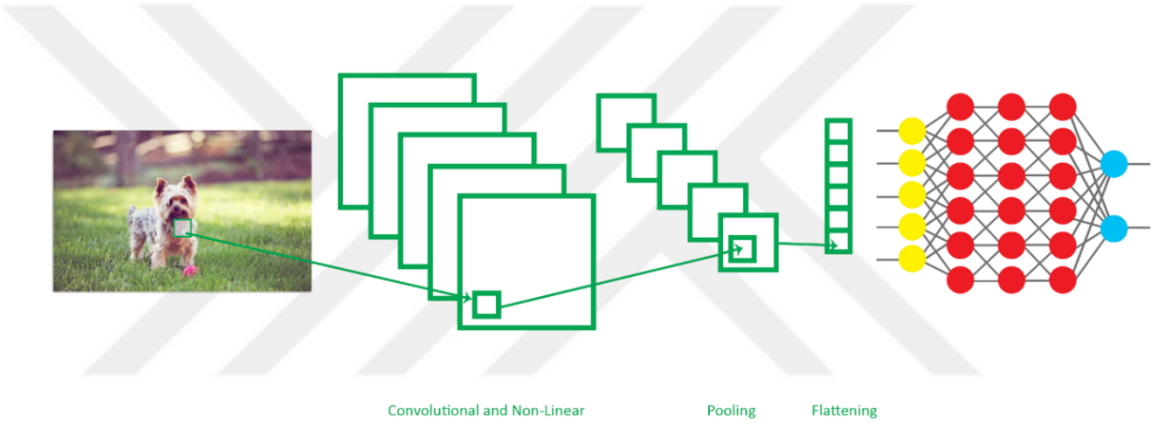
3.1. Konvolüsyonel(Evrişimsel) Sinir Ağları Yapısı ve Kullanılan Sistemin Matematiksel Modeli

3.1.1. Evrişimsel Sinir Ağları

Bu çalışmada evrişimsel sinir ağları konuşma tanıma modeli için kullanılmıştır. Bu yüzden ilk olarak evrişimsel sinir ağlarının yapısını anlatmak lazım(Kalchbrenner, Grefenstette, & Blunsom, 2014).

Konuşma tanıma modeli elde etmek için, evrişimsel sinir ağları çeşitli katmanlarla kullanır. Bu katmanları daha detaylı ve ayrıntılı olarak incelenecektir, ancak bu katmanlara ve amaçlarına genel bir bakış atmak gerekirse, evrişimsel sinir ağlarının katmanları şu şekildedir:

- Convolutional Layer — Özellikleri saptamak için kullanılır
- Non-Linearity Layer — Sisteme doğrusal olmayanlığın (non-linearity) tanıtılması
- Pooling (Downsampling) Layer — Ağırlık sayısını azaltır ve uygunluğu kontrol eder
- Flattening Layer — Klasik Sinir Ağı için verileri hazırlar
- Fully-Connected Layer — Sınıflamada kullanılan standart sinir ağı



Şekil 3.2: ConvNet şeması

Şekil 3.1'de ConvNet şeması gösterilmiştir. Temel olarak, evrişimsel sinir ağları, sınıflandırma konularında kullanılır ve bu yöntem sınıflandırma sorununu çözmek için standart sinir ağları kullanır. Evrişimsel sinir ağlarıda temel olarak kullandığı mantık sınır ağları mantığı ama farklı yönleride var ve bilgileri belirlemek ve bazı özellikleri tespit etmek için daha fazla katman kullanır.

3.1.2. Evrişimsel Katman

Bu katman evrişimsel sinir ağlarının ana katmanıdır. Verinin özelliklerini algılamaktan sorumludur. Bu veri resim, ses ve metin olabilir. Sinir ağlarının ilk kullanım alanı resim dünyası olmuştur. O yüzden evrişimsel sinir ağlarının tanımı içinde bu çerçeve kullanılmıştır. Bu katman, verideki düşük ve yüksek seviyeli özellikleri çıkarmak için resme bazı filterler uygulamaktadır. Örneğin, bu filtre kenarları algılayacak bir filtre

olabilir. Bu filtreler genellikle çok boyutludur ve piksel değerleri içermektedirler. (5x5x3) 5 matrisin yükseklik ve genişliğini, 3 matrisin derinliğini temsil eder.

Bir örnekle evrişimsel katmanın filterinin nasıl uygulandığı araştırılacaktır;

1	1	1	0	0
0	1	1	1	0
0	0	1	1	1
0	0	1	1	0
0	1	1	0	0

Şekil 3.3: Evrişimsel katmanın filtresini göstermek için örnek resim

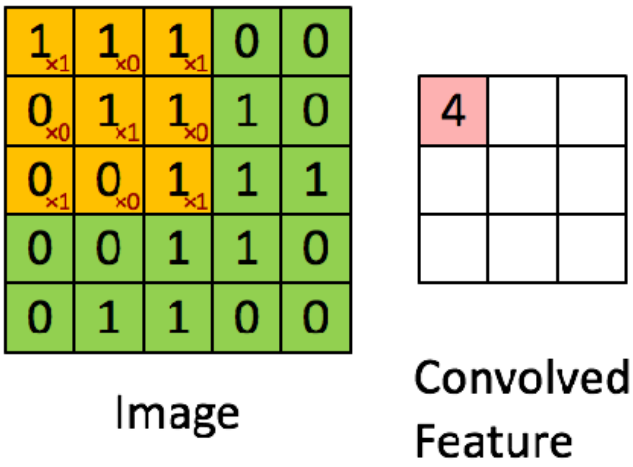
Evrişimsel katmanın filtresini göstermek için örnek resim şekil 3.2'de gösterilmiştir.

Örneğin anlaşılır olması için burada sadece bir kanal işlenecektir.

Resmin 5×5 boyutunda ve "1" ve "0" 'lardan oluşan bir resim olduğunu varsayalım. Filtremizi 3×3 boyutunda oluşturulduğunu varsayalım.

1	0	1
0	1	0
1	0	1

Evrişimsel katmanın filtresini göstermek için örnek filtre şekil 3.3'de gösterilmiştir.



Şekil 3.4: Evrişimsel katmanın filtresinin uygulanması

Evrişimsel sinir ağlarında filtrenin nasıl uygulandığını gösteren resim şekil 3.4'de gösterilmiştir.

Şimdi evrişimsel sinir ağlarında filtrenin nasıl uygulanışı hakkında bilgi verilecektir. Öncelikle, filtre görüntünün sol üst köşesine konumlandırılır. Burada, iki matris arasında (resim ve filtre) indisler birbirisi ile çarpılır ve tüm sonuçlar toplanır, daha sonra sonucu çıktı matrisine depolanır. Ardından, bu filtreyi sağa 1 piksel ("basamak" olarak da bilinir) kadar hareket ettirip işlemi tekrarlanır. 1. Satır bittikten sonra 2 satıra geçilir ve işlemler tekrarlanır. Tüm işlemler bittikten sonra çıktı matrisi oluşturulur. Burada çıktı matrisinin 3×3 olmasının nedeni 5×5 matrisinde 3×3 filtresi yatayda ve dikeyde 3 kez hareket etmesinden kaynaklanır.

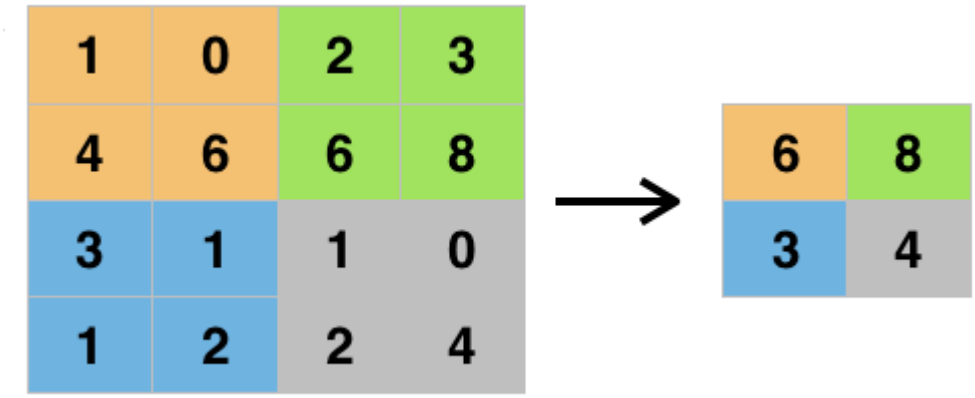
Eğer resim 6×4 ve filtre 3×3 boyutunda olsaydı çıkış matrisi 4×2 boyutunda olurdu.

Peki çıktı matrisi bize ne anlatıyor? Bu matrise genellikle Feature Map denir. Filtre tarafından temsil edilen özellikte görüntünün bulunduğu yeri gösterir. Kısacası, filtreyi görüntü üzerinden hareket ettirerek ve basit matris çarpımını kullanarak, özelliklerimizi tespit ediyoruz.

Genellikle, birden çok özelliği tespit etmek için birden fazla filtre kullanlır, yani bir CNN ağında birden fazla konvolüsyonel (Convolutional) katman bulunur(Kalchbrenner, Grefenstette, & Blunsom, 2014).

3.1.3. Pooling Layer

Bu katman, CovNet'teki ardışık convolutional katmanları arasına sıklıkla eklenen bir katmandır. Bu katmanın görevi, gösterimin kayma boyutunu ve ağ içindeki parametreleri ve hesaplama sayısını azaltmak içindir. Bu sayede ağdaki uyumsuzluk kontrol edilmiş olur. Birçok Pooling işlemleri vardır, fakat en popüleri max pooling'dir. Yine aynı prensipte çalışan average pooling, ve L2-norm pooling algoritmalarıda vardır. Bu işlemi şekiller üzerinden açıklayarak gidelim. Öncelikle 2×2 boyutunda bir filtre oluşturalım. Bu filtreyi aşağıdaki (4×4) resim üzerinde görebilirsiniz. Resimde gördüğünüz gibi, filtre, kapsadığı alandaki en büyük sayıyı alır. Bu sayede, sinir ağının doğru karar vermesi için yeterli bilgiyi içeren daha küçük çıktıları kullanmış olur.

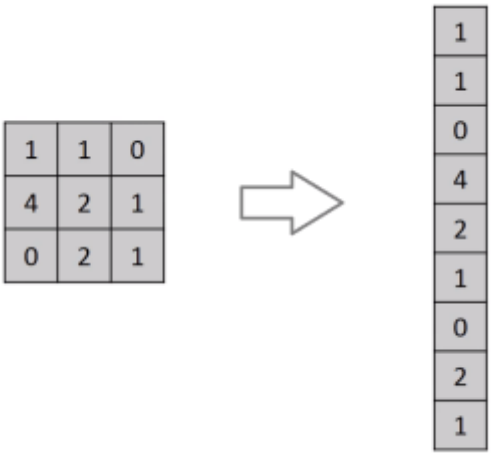


Şekil 3.5: Maxpooling işlemi

Bununla birlikte birçok kişi bu katmanı kullanmayı tercih etmez. Bunun yerine convolutional katmanında daha büyük Stride (Filtreyi kaydırma işlemi) tercih edilir. Ayrıca variational autoencoders (VAEs) or generative adversarial networks (GANs) gibi daha üretken modellerde pooling katmanını tamen çıkartırlar.

3.1.4. Flattening Layer

Bu katmanın görevi net söylemek gerekirse sadece başka katmana veri hazırlamaktır. En son ve en önemli katman olan Fully Connected Layer'ın girişindeki verileri bu katman tarafından hazırlanmaktır. Bu sinir ağındaki veriler ise Convolutional ve Pooling katmanından gelen matrixlerin tek boyutlu diziye çevrilmiş halidir. Bu bir genel kuraldır, sinir ağlarında, giriş verilerini tek boyutlu bir diziden almaktadır.



Şekil 3.6: Flatting işlemi

3.1.5. Fully-Connected Layer

Bu katman ConvNet'in son ve en önemli katmanıdır çünkü öğrenme işlemini gerçekleştirir. Aslında verileri Flattening işleminden alır ve daha sonra sinir ağı yoluyla öğrenme işlemini geçekleştirir. Bu katmanı detaylı anlatmak başlı başına bir derin konu olduğundan dolayı daha ayrıntılı incelenmeyecktir.

3.1.6. Sistemin Matematiksel Modeli

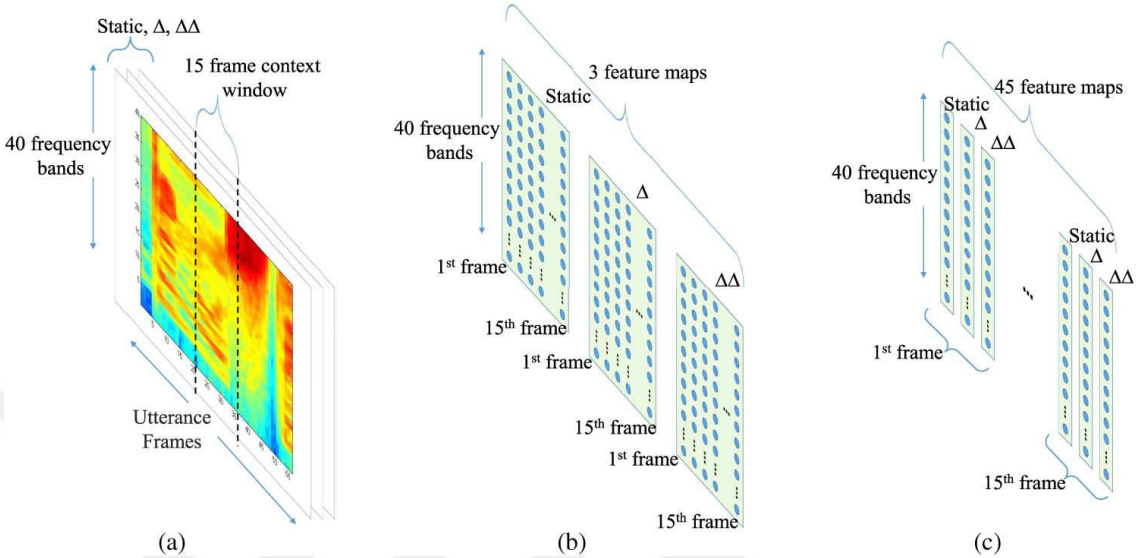
Örüntü tanıma için CNN'ler kullanılırken,ilk olarak kullanılan kelimelerin yapısına göre bir özellik haritası çıkarmak gerekir.Daha sonra bu özellik haritasındaki veriler CNN'in giriş verilerini oluştururlar.

CNN'ler giriş verisi üzerinde küçük bir pencerede aynı zamanda hem eğitim,hem test yaparak ağın ağırlıklandırma parametrelerini belirlerler.Bu ağırlıklandırma parametrelerini ise daha önceden çıkartılan özellik haritalarına göre ayarlarlar.

Biz bu çalışmada MFCC katsayılarını kullanarak,işlemleri gerçekleştirdik.Bu bize her konuşmanın içindeki statik, delta(1.türev) ve delta-delta (2.türev) arası noktaların tayinini yapmamızı sağladı.

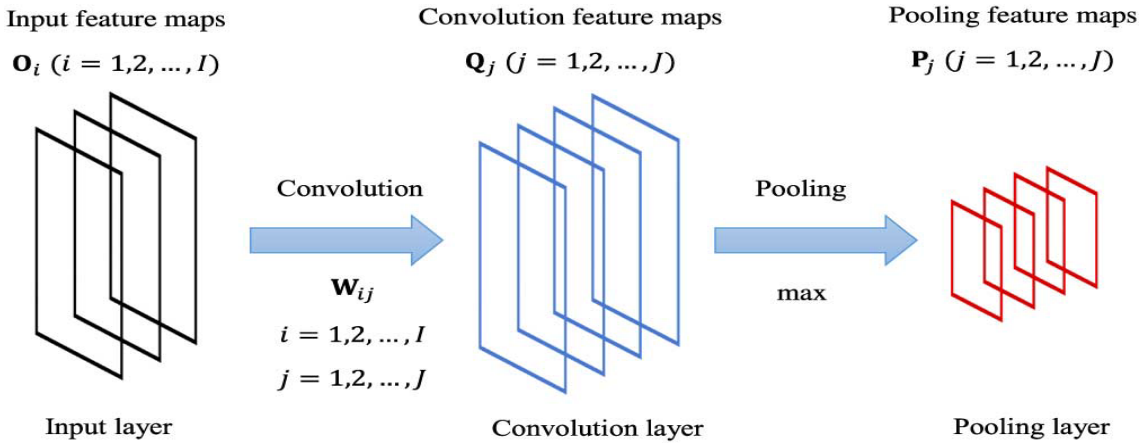
Şekil (a)'da ham ses sinyalinin çerçevesini görmekteyiz. Şekil (b)'de ise bu konuşma sinyalinin 1-15 arasında çerçevelere bölünmüş 2 boyutlu halini görmekteyiz.Burada frekans ve zamansal degisim aynı anda yapılmaktadır. Şekil (c)'de ise sadece frekans

alanında normalleştirme yapılmıştır. Burada MFCC özellikleri bir boyutlu olarak düzenlenir. Burada çerçeve sayısı ve filtre bankalarını kullanarak tek boyutlu özellik haritaları üzerinden yani sadece frekans alanında inceleme yapıldı.



Sekil 3.7: Çerçeveleme ve filtre bankaları kullanarak yapılan matris çıkarma işlemi

CNN girişi için özellik haritaları oluşturulduktan sonra,evrişim ve pooling katmanı oluşturulduktan sonra,sırayla bu birimlerin aktivasyonları Şekil-2'deki gibi gerçekleştirilir.Giriş katmanına benzer şekilde konvolüsyonel katman ve Pooling katmanı haritalar içinde düzenlenebilir.CNN terminolojisi,bir çift kıvrım ve havuz katmanları arka arkaya bir "CNN" olarak adlandırılır. Dolayısıyla derin bir CNN oluşturmak için bu katmanlardan ard arda kullanmak gerekir.



Şekil 3.8: CNN modeli

Her giriş özelliği haritası (toplam sayısı I olan),bir çok özellik haritasına (toplam sayısı J olan) bağlıdır. Q_j(j= 1,…..,J), konvolüsyon katındaki yerel ağırlık matrislerinin sayısı (toplam IxJ) şekil-2'de gösterilmiştir.

Giriş özellik haritalarının hepsi tek boyutlu olduğuna göre,evrişim katındaki her bir özellik haritası şu şekilde hesaplanabilir:

$$q_{j,m} = \sigma\left(\sum_{i=1}^{I}\sum_{n=1}^{F} o_{i,n+m-1} w_{i,j,n} + w_{0,j}\right) \ (j = 1, \ldots, J) \qquad (3.1)$$

Burada F filtre boyutu olarak kullanılır. Her bir giriş verisi özelliğinin eşleşmesindeki frekans bandı sayısı evrişim katındaki girdi olarak alınır.MFCC özellikleri seçiminden kaynaklanan yerellik yüzünden,bu özellik haritaları konuşma sinyalinin sınırlı bir frekans aralığıyla sınırlıdır.

Konvolüsyon katındaki özellik haritaları,konvolusyon katında kullanılan matrislerinin yerel ağırlıklandırılmalarında belirleyici etkendir.Uygulama yaparken aynı katsayıya sahip birçok matrisi sınırlayacağız.

CNN normal ağdan biraz daha farklıdır.Bunu şöyle açıklayabiliriz;

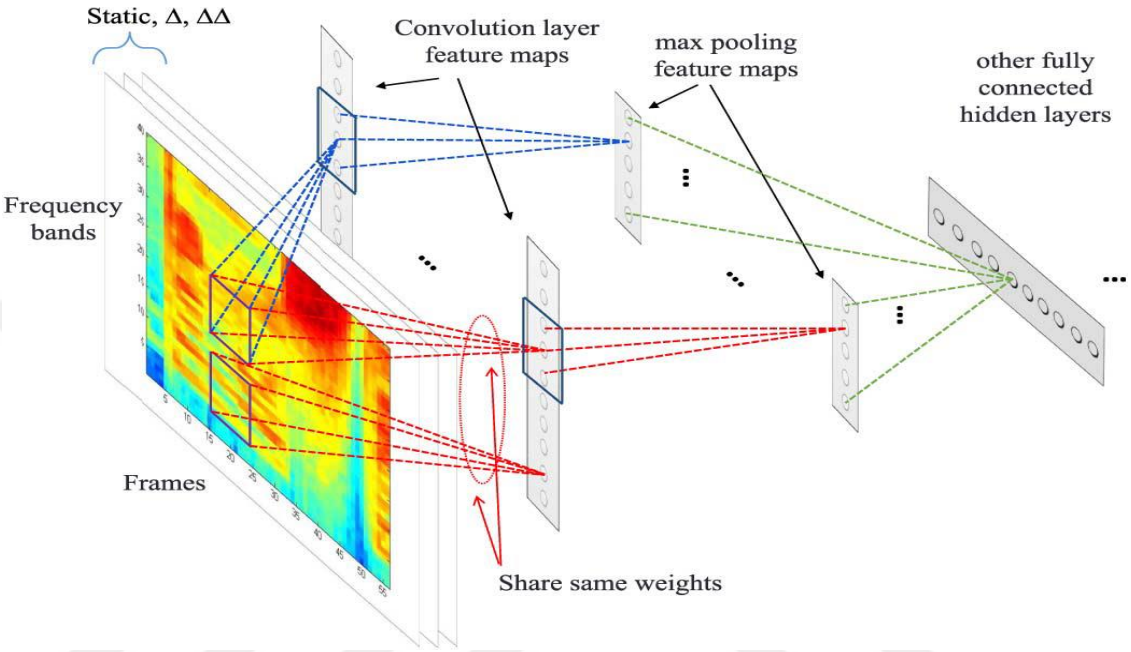
İki önemli özelliği olan tamamen bağlı katmanı(Fully Connected Layer) içerir.Bu katmanın ilk önemli özelliği,her evrişimsel birim girişi, yalnızca girişin yerel bir alanından alır.Bu her birimin yerel bir bölgenin bazı özelliklerini temsil ettiği anlamına gelir.İkincisi, evrişim kat birimleri kendileri bir dizi özellik haritasına yerleştirilebilir;aynı özellik haritasındaki tüm birimler aynı ağırlıkları paylaşır,ancak verileri alt katmanın farklı yerlerinden alırlar.

Pooling katmanı, evrişimsel katmandaki özellik haritasına eşit sayıda harita içerir,ancak bu haritalar daha küçük matrisler biçimindedir.Bu katmandaki amaç özellik haritalarının çözünürlüğünü azaltmaktır.Bunun sayesinde belirli bir frekanstaki özellikler daha net ortaya çıkarılmaktadırlar.Bu çalışmada, max-pooling uygulaması gerçekleştirildi. Aşağıdaki şekilde;

kayma boyutu "s" , pooling boyutu "G" ile ifade edilmektedir.

$$p_{i,m} = \max_{n=1}^{G} q_{i,(m-1)\times s+n} \tag{3.2}$$

Aşağıdaki şekilde bir boyutlu matrisler üzerinden,frekans boyunca max-pooling uygulanan CNN ağının modellemesi gösterilmiştir.



Şekil 3.9: Bir boyutlu matrisler üzerinden frekans boyunca max-pooling uygulanan CNN modeli

Evrişim katındaki tüm ağırlıklandırma parametreleri geri yayılım algoritması kullanılarak öğrenilir,ayrıca buna ek olarak bazı özel modifikasyonlar,ağ arasındaki zayıf bağlantılar ve ağırlık paylaşımı da kullanılır.

Basit olarak W matrisi aşağıda oluşturulmuştur.

$$\mathbf{W} = \begin{bmatrix} w_{1,1,1} & w_{1,2,1} & \cdots & w_{1,J,1} \\ \vdots & \vdots & \ddots & \vdots \\ w_{I,1,1} & w_{I,2,1} & \cdots & w_{I,J,1} \\ \vdots & \vdots & \ddots & \vdots \\ w_{I,1,2} & w_{I,2,2} & \cdots & w_{I,J,2} \\ \vdots & \vdots & \ddots & \vdots \\ w_{I,1,F} & w_{I,2,F} & \cdots & w_{I,J,F} \end{bmatrix}_{I\cdot F\times J} \tag{3.3}$$

Burada W matrisi I*F satırlarının çarpımından oluşturulmuştur, burada F filtre boyutunu belirtmektedir.Her frekans bandı I giriş özellik haritası içeren I satırını içermektedir ve W aynı zamanda evrişim katındaki özellik haritasına göre ağırlıklandırmaları içeren J sütununa sahiptir.Bu matris sayesinde ağdaki ağırlık parametrelerinin belirlenmesi sağlanır.

3.2. Konvolüsyonel(Evrişimsel) Sinir Ağlarının Uygulamaları

Bu çalışmada, "Konvolüsyonel (Evrişimsel)" sinir ağlarının günümüzde nasıl kullanıldığı anlatılmıştır. Ayrıca çalışma esnasında konvolüsyonel sinir ağları için "CNN" kısaltması da kullanılmıştır. Şirketlerde CNN'ler, genellikle insanların yaşam kalitelerini zenginleştirmeye yönelik uygulamalarda kullanılır(Sainath, Mohamed, Kingsbury & Ramabhadran, 2013).

Gündelik hayatta görülebilecek CNN'lerin basit uygulamaları; yüz tanıma yazılımı, görüntü sınıflandırma, konuşma tanıma programları gibi çok geniş bir alanı kapsamaktadır. Özellikle Instagram gibi görüntü odaklı sosyal medya uygulamaları evrişimsel sinir ağlarından faydalanmaktadır. CNN'nin kilit uygulamalarından bazıları ise şöyle listelenmiştir:

3.2.1. Kod Çözme/Yüz Tanıma

Yüz tanıma, evrişimli bir sinir ağı tarafından aşağıdaki ana bileşenlere bölünür:

- Resimdeki her yüzü belirlemek
- Işık, açı, poz gibi harici faktörlere rağmen her yüze odaklanmak
- Eşsiz özellikleri belirlemek
- Toplanan tüm verileri, veri tabanında önceden var olan verilerle bir ad altında eşleştirmek için karşılaştırma yapmak
- Benzer bir işlemin sahne etiketlemesi için bunu takip etmek

Bir görüntünün yapısını tanımak için CNN'ler görevlendirilir. Normalde, bir görüntü ağa bir sayı ızgarası olarak beslenir. Ancak, bunu yapmanın daha iyi bir yolu, görüntüyü küçük bir sinir ağına gönderilen kesişen görüntü döşemelerine bölmektir.

Yüzler; yerel resim örneklemesini, öz-düzenleyici bir harita sinir ağını ve bir CNN'yi birleştiren hibrit bir sinir ağı tarafından görüntülenen karmaşık, çok boyutlu görsel uyaranlardır. Karhunen-Loe`ve Transform, iyi performans gösteren öz-düzenleyici harita (% 3.8'e karşı %5.3 hata) ve kötü performans gösteren çok katmanlı bir algılayıcı (%3.8'e karşı %40 hata) yerine sonuçları göstermek için kullanılmıştır (Lawrence vd., 1997).

3.2.2. Belgeleri Analiz Etme

Konvolüsyonel sinir ağları da doküman analizi için kullanılabilir. Bu sadece el yazısı analizi için yararlı değildir, aynı zamanda tanıyıcılar da büyük bir paya sahiptir. Bir makinenin, bir bireyin yazısını tarayabilmesi ve sahip olduğu geniş veri tabanıyla karşılaştırması için dakikada yaklaşık bir milyon komut vermesi gerekmektedir. CNN'lerin, daha yeni modellerin ve algoritmaların kullanılmasıyla, hata testinin karakter seviyelerinin minimum % 0,4'e düşürüldüğü söylenmektedir(Li, Lin, Shen, Brandt & Hua, 2015).

Belgeler ve cümleler matrisler halinde gösterilir ve NLP görevleri kullanılarak ele alınır. Her simge, matristeki, bir sözcük veya karakter olabilen bir satırı temsil eder. Bu nedenle, her satır, temelde bir belirteç olan bir vektördür. Bu vektörler, kelime gömmeleri olarak adlandırılan düşük boyutlu temsiller/gösterimlerdir. Kelime gömme, kelime dağarcığındaki kelimelerin düşük boyutlu bir alanda gerçek sayıların vektörlerine eşlendiği NLP'de, bir dil modelleme ve özellik öğrenme tekniği grubudur (Mikolov vd., 2013). Kelime gömme yöntemlerine; Mikolov ve arkadaşları tarafından 2013 yılında önerilen Word2vec, ve Pennington ve arkadaşları tarafından 2014 yılında önerilen GloVe örnek verilebilir. Word2vec, Google News üzerinden erişimi herkese açık olan 100 milyar kelime kullanarak eğitilmiştir. Buna ek olarak, sabit veya değişken bir filtre boyutu kullanılarak, kıvrım hesaplanır ve özellik haritası üretilir. Her özellik haritası için Pooling (Ortaklama) uygulanır. Sınıflandırma gibi gerekli görevleri tamamlamak için son bir özellikler vektörü üretilir ve bir son katmandan geçirilir. NLP'de CNN'lerin temel

özellikleri olan konumda değişmezlik ve yerel bütünlük, bilgisayar görme uygulamalarında olduğu gibi geçerli değildir. Bir kelimenin cümle içindeki yeri son derece önemlidir. Piksel durumunda, birbirine yakın olanlar aynı nesneye ait olabilir ve bağlı olabilirler. Ancak, birbirine yakın kelimelerin mutlaka aynı anlama gelmesi gerekmediğinden, aynı durum cümlelerde geçerli değildir ve sonuçta, bağlı olmayabilirler. Bundan dolayı, CNN'ler yalnızca sınıflandırma görevlerini, konu kategorizasyonu veya duyarlılık analizi gibi görevleri yerine getirmek için uygulanır. Bu görevlerde dizinin önemli olmasından ve, evrişim ve pooling (ortaklama) süreçlerinin kelimelerin sırasını takip etmemesinden dolayı, Klasik CNN'ler için PoS etiketleme veya giriş çıkarma gibi görevleri yerine getirmek zordur.

3.2.3. Görüntü Sınıflandırması

CNN'lerin ortak özellikleri ve sınıflandırıcı öğrenme yeteneği bulunduğundan, büyük ölçekli veri kümelerinde çalıştırıldıklarında diğer yöntemlere kıyasla daha yüksek sınıflandırma doğruluğu üretirler (Gu vd., 2018). AlexNet'i geliştiren Krizhevsky vd. (2012), ILSVRC 2012'de en iyi performansı elde etmiştir. Bu başarının ardından, birkaç başka çalışmada, filtre boyutu en aza indirgenerek (Strigl vd., 2010) veya ağın derinliği artırılarak sınıflandırma doğruluğunda önemli iyileştirmeler elde edilmiştir (Simonyan ve Zisserman, 2014; Szegedy vd., 2015). CNN'nin hızlı, tamamen parametrelendirilebilir GPU kullanımı; %2.53, %19.51 blunder oranları ile nesne tespiti (NORB, CIFAR10) için karşılaştırmalı değerlendirme sonuçlarını vermiştir. Lowe (1999), resim sınıflandırmasında bir GPU kodunun CPU muadilinden iki kat daha hızlı olduğunu göstermiştir. Strigl vd. (2010) ve, Uetz ve Behnke (2009), çok sütunlu derin sinir ağlarının (MCDNN'ler) görüntü sınıflandırması için her geçmiş stratejiyi yenebileceğini belirtmiş ve hata oranını %30-40 azaltırken (arada bir küçük veri kümeleri için yararlı olsa da) ön hazırlığa gerek olmadığını göstermiştir (Cireşan vd., 2012). Doyuma ulaşmayan nöronlar ve konvolüsyon işleminin etkin GPU uygulaması; ImageNet LSVRC-2010'daki 1.2 milyon yüksek çözünürlüklü görüntünün 1000 benzersiz sınıfa ayrılması için ILSVRC-2012 yarışmasının en iyi ikinci girişinin elde ettiği %26,2 ile karşılaştırıldığında, %15,3'lük başarı en iyi 5 test hata oranını beraberinde getirmiştir (Krizhevsky vd., 2012).

Görüntü sınıflandırmadaki bazı sınıfların diğerlerinden daha belirsiz olmasına dayalı olarak, Hiyerarşik Derin Konvolüsyonel Sinir Ağı (HD-CNN) geliştirilmiştir. N-way sınıflandırıcılar olup, kaba-ince sınıflandırma stratejisi ve tasarım modülünü takip eden geleneksel CNN'lere dayanmaktadır. CIFAR100-NIN yapı taşına sahip HD-CNN, %65.33'lük bir test doğruluğu gösterir. Bu, CIFAR100 veri kümesindeki diğer standart derin modellerin ve HD-CNN modellerinin doğruluğundan daha yüksek bir orandır (Yan vd., 2015). İnce-taneli görüntülerle ilgilenen görüntü sınıflandırma sistemleri, benzersiz özellikleri ayırt etmek için ön plan nesnelerini tanıma kavramına dayanır. İnce-taneli sınıflandırmalara dikkat uygulamak, sadece, sınıf etiketinin verildiği en az izleme ayarları ile yapılabilir. Bu, bir sınıflandırma görevi ile eğitilmiş CNN'den alınan dikkat kullanılarak yapılabilir. Eğitmek veya test etmek için bir nesne sınırlama kutusuna veya parça işaretine ihtiyaç duyan diğer tekniklerin tersidir. Bu yöntem; CUB200-2011 veri kümesi için, en kötü denetim ayarı altında, en yüksek doğruluğu üretmektedir (Xiao vd., 2015).

3.2.4. İklimi Anlamak

CNN'ler, iklim değişikliğine karşı mücadelede oldukça önem arz etmektedir. Çünkü ciddi iklim değişikliklerinin gerçekleşmesinin nedenlerini gözler önüne sermektedir. Ayrıca söz konusu etkiyi azaltmak için neler yapılması gerektiğinin anlaşılabilmesinde de oldukça etkilidir. Bu tür doğal tarih koleksiyonlarındaki verilerin de daha fazla sosyal ve bilimsel görüş sağlayabildiği söylenmektedir. Ancak bunun, bu tür depoları fiziksel olarak ziyaret edebilecek araştırmacılar gibi, yetenekli insan kaynakları gerektireceği düşünülmektedir. Bu alanda daha derin araştırmalar yapmak için daha fazla insan gücüne ihtiyaç duyulmaktadır(Kalchbrenner, Grefenstette, & Blunsom, 2014).

İklim verileri analizi için bir dizi gelişmiş metodolojiye ihtiyaç duyulur. Sinir ağı tabanlı makine öğrenimi yaklaşımı, generatif bir analiz tekniği olarak son yıllarda büyük ilgi görmüş olup, çeşitli iklim sorunlarının çözülmesinde faydalanılmıştır. Chattopadhyay vd. [Chattopadhyay ve ark., 2013], Madden-Julian salınımının (MJO) yapısal evrimini incelemek için, Öz-düzenleyici Harita'ya (SOM) dayalı doğrusal olmayan bir kümeleme yöntemi geliştirmiştir. Geliştirmiş oldukları yöntem, diğer yöntemlerdeki gibi önemli

modların seçilmesini veya zaman ve mekanda mevsim-içi bant geçiş filtresini gerektirmemektedir. Sonuçlar, SOM tabanlı yöntemin yalnızca MJO yapısındaki ve gelişimindeki brüt özelliği yakalamakla kalmayıp, aynı zamanda MJO'da ortaya çıkan uzun dalga radyasyonunun ve diyabetik ısıtmanın dipol ve tripol yapısı gibi, diğer yöntemlerin ortaya çıkaramadığı kavrayışları açığa çıkardığını göstermektedir. Gorricha ve Costa [Gorricha ve ark, 2013], İspanya'daki bir ada üzerindeki aşırı yağış modellerini kategorize etmek ve görselleştirmek için üç boyutlu bir Öz-düzenleyici Harita kullanmışlardır. Geleneksel yağış endeksi yaklaşımının ortaya çıkaramadığı mekansal yağış modellerini bulmuş ve, üç boyutlu Öz-düzenleyici Harita'nın keşifsel uzamsal model analizi için oldukça yararlı bir araç olduğu sonucuna varmışlardır. Yakın zamanda, Shi ve arkadaşları ; anlık yağış tahmini için, yeni geliştirilen, erişimli uzun kısa süreli bellek (LSTM) derin sinir ağını kullanmışlardır[Xingjian ve ark., 2015]. İki boyutlu radar harita zaman serileri üzerinde eğitilmiş olan sistemleri, çeşitli değerlendirme metriklerinde, mevcut en gelişmiş anlık yağış tahmini sisteminden daha iyi performans gösterebilir. Iglesias vs., tarihsel zaman serileri verisi üzerinde eğitilmiş, ısı dalgaları tahmini üzerine çok görevli, tam bağlantılı bir sinir ağı geliştirmişlerdir [Iglesias ve ark., 2015]. Sinir ağı yaklaşımının, doğrusal ve lojistik regresyondan önemli ölçüde daha iyi olduğunu ve potansiyel olarak, aşırı ısı dalgalarını tahmin etme performansını artırabileceğini ortaya koymuşlardır. Bu çalışmalar, sinir ağının generatif bir yöntem olduğunu ve çeşitli iklim sorunlarına yönelik uygulanabileceğini göstermektedir. Çalışma kapsamında, iklim örüntüsü algılama sorununu çözme konusuna ilişkin olarak, derin Evrişimli Sinir Ağı araştırılmaktadır.

3.2.5. Gri Alanlar

CNN'lerde tanıtılan gri alanlar,, gerçek dünyanın çok daha gerçekçi bir resmini oluşturmak için hazırlanmıştır. Şu anda CNN'ler büyük ölçüde bir makine gibi işlev görmekte ve her soru için doğru ve yanlış bir değer taşımaktadır. Ancak insanlar gerçek dünyanın, grinin birçok farklı tonunda ortaya çıktığının bilincindedir. Makinenin belirsiz mantığı anlama ve işlemesine izin vermek, onun, insanların yaşadığı gri alanı anlamasına ve bunun için çaba harcamasına yardımcı olacaktır. Böylece CNN'ler, insanların

gördükleri hakkında daha bütünsel bir görüş elde edecektir (Kalchbrenner, Grefenstette, & Blunsom, 2014).

3.2.6. Reklam

CNN'ler, programatik satın alma ve veri odaklı kişiselleştirilmiş reklamcılığın getirilmesiyle, reklamcılıkta bir fark yaratmıştır(Li, Lin, Shen, Brandt & Hua, 2015). Li ve ark. yaptığı bir çalışmada TV yayınlarında CNN tabanlı bir ticari algılama araştırması yapılmıştır. Bu çalışmada, TV yayıncılığı için otomatik bir ticari tespit sistemi tasarlanmış ve uygulanmıştır. Bu sistem çekim düzeyinde çalışır ve TV yayını ve çevrimiçi videolar da dahil olmak üzere akış videolarındaki reklamları sınıflandır ve algılar. Atış sınır algılama modülü ve atış sınıflandırma modülü olmak üzere iki modülden oluşturulmuştur. Bu çalışmda, evrişimli sinir ağı ve geleneksel makine öğrenimi teknikleri kullanılarak, çeşitli program türlerini, reklamları ve normal TV programlarını 93% dikkat oranıyla sınıflandırlımıştır (Li ve ark., 2017). Başka bir çalışmada, Vo ve ark. çevrimiçi reklamın gösterimini tanımlamak için uygulanan görüntü sınıflandırma modeli sunmuşlardır. Sınıflandırma modeli için bir CNN modeli sunulmuştur. Evrişimli sinir ağı kullanarak reklam görüntü sınıflandırması modelinin başarısı 86% olarak bildirilmiştir (Vo ve ark., 2017).

3.2.7. Diğer İlginç Alanlar

Sürücüsüz otomobillere, insan davranışlarını taklit edebilen robotlara, insan genom haritalama projelerine yardımcı olan; depremleri ve doğal afetleri öngören ve hatta kendi kendine tıbbi sorunların teşhislerini yapan CNN'ler, geleceği aydınlatmaya hazırdır. CNN'ler sayesinde; hapşırma atağının veya yüksek ateşin nadir görülen bir hastalığın belirtisi değil, sadece basit bir grip semptomu olduğundan emin olmak için bir kliniğe gitme veya doktordan randevu alma gereksinimleri ortadan kalkar. CNN'lerin bu yönünü geliştirmek için çalışan araştırmacıların son yıllarda üzerinde durdukları konulardan biri de, CNN'lerle beyin kanseri tespit etmektir(Liang & Hu, 2015). Beyin kanserinin erken

teşhisi, bu hastalıktan etkilenen insanların kurtarılma oranlarının artmasında büyük bir adım olabilir(Sainath, Mohamed, Kingsbury & Ramabhadran, 2013).

4.UYGULAMA

Önceki bölümlerde, konuşma tanıma konusundaki kavramların teorik temelleri derin öğrenme algoritmaları kullanılarak tartışılmıştır. Bu bölüm ise, bir önceki bölümde önerilen yöntemi uygulamaya yöneliktir. Konuşma tanıma bölümü, konuşma tanıma veri kümesinde uygulanacaktır. Bu çalışmada, konuşma tanıma için farklı derin öğrenme algoritmalarının kullanımı tartışılacak ve bu problemi çözmek için önerilen yöntem uygulanacaktır. Elde edilen sonuçlar, sonuç bölümünün içerisinde yorumlanmıştır. Ayrıca sonuç bölümünde, bu çalışmada kullanılan veri setinin detayları da sunulacaktır.

4.1. Veri Seti

Bu araştırmada kullanılan veri seti, 65.000 adet, uzunluğu bir saniye ve 30 kısa kelime olarak yayınlanmıştır. Bu veri seti kaggle web sitesinde[8] yayınlanmaktadır. Söz konusu web sitesi, farklı veri kümelerini paylaşan güvenilir web sitelerinden biridir. Kaggle web sitesi dünya genelinde 194 ülkeden veri bilimiyle ilgilenen 500.000'in üzerinde aktif üyeye sahiptir. Aynı zamanda burada veri bilimi meraklıları için veri setleri ve başka faydalı şeyler de yayınlanmaktadır. Bu web sitesinin popülaritesinin sebeplerinden biri, veri bilimi meraklıları arasında yaydığı rekabettir. En iyi rakiplere nakit ödüller verilir. Bu araştırmada, mevcut itibarı nedeniyle, Kegel web sitesinde yayınlanan veri seti sunulmuştur. Dünyadaki birçok insan veri teminatı için bu veri setini kullanmıştır. Bu veri tabanı, binlerce insanın otuz adet İngilizce kelime telaffuz ettiği 64728 ses dosyası içermektedir.

Konuşma tanıma ve veri madenciliğinde en önemli ve en çok kullanılan programlama dillerinden biri olan Python, önerilen yöntemi uygulamak ve geliştirmek için kullanılacaktır.

[8] https://www.kaggle.com/

4.2. Uygulanan Adımların Özeti

Bu bölümde uygulanan adımların özeti anlatılacaktır. Ön işlemden sonra, makine öğrenme algoritmaları için özellik seçim aşaması gerçekleştirilir. Ancak derin öğrenme algoritmalarında, özellik seçimi adımları yoktur. Bu çalışmada, derin öğrenme algoritmaları kullanmanın dil çerçevesinde konuşma tanıma modelleri geliştirmek için en iyi seçim olduğu iddiası kanıtlanmaya çalışılmaktadır. Daha sonra CNN derin öğrenme algoritması bir konuşma tanıma modeli oluşturmak için kullanılır.

Bu bölümde, önceki bölümlerde söylenenlere dayanarak, konuşma tanıma veri kümesi için konuşma tanıma modelinin uygulanmasından elde edilen sonuçlar gösterilecektir.

Bu çalışma, konuşma tanıma konusunda dikkatli bir yaklaşım kullanmaktadır. "Gözlem Yöntemlerini Kullanmaya Yaklaşım", makine öğrenmede en eski ve en popüler yaklaşımdır. Bu yöntemde konuşma tanıma bir veri madenciliği alt kümesi olarak tanıtılır. Dolayısıyla, sınıflandırma problemlerini çözmek için sunulan aynı gerekçe de bu tartışmada mevcuttur. Veri sınıflandırma, verileri belirtilen gruplara göre tasnif etmek için bir sınıflandırma modeli oluşturmayı amaçlamaktadır. Örneğin, haberlerin sınıflandırılmasında haberleri; ekonomik, eğitimsel, kültürel ve diğer gruplara sınıflandırmak için bir model sunmayı amaçlayan durumlar vardır. Etiketli veri kümesi de modeli oluşturmak için kullanılır. İlk olarak, bir etiketli veri kümesi, makine öğrenme algoritmalarından birine atanır. Daha sonra makine öğrenme algoritması, bu etiketli veri setine dayanarak bir sınıflandırma modeli oluşturmaya çalışır. Aşağıdakiler, yeni veri setinin etiketli veri setine dayanarak oluşturulan sınıflandırma modeline göre tasnif edilir. Bu örnek, veri madenciliğinde klasik bir sınıflandırmayı göstermektedir. Dolayısıyla, denetlenen yaklaşımda bu fikir ana tema olarak kullanılmaktadır.

Veri kümesi, 65.000 ses içermektedir. Bu veri setinde otuz kelime içeren 65.000 ses bulunmaktadır. Önerilen yönteme dayanarak CNN algoritması, konuşma tanıma veri setine uygulanır ve sonuçlar tablolarda sunulur.

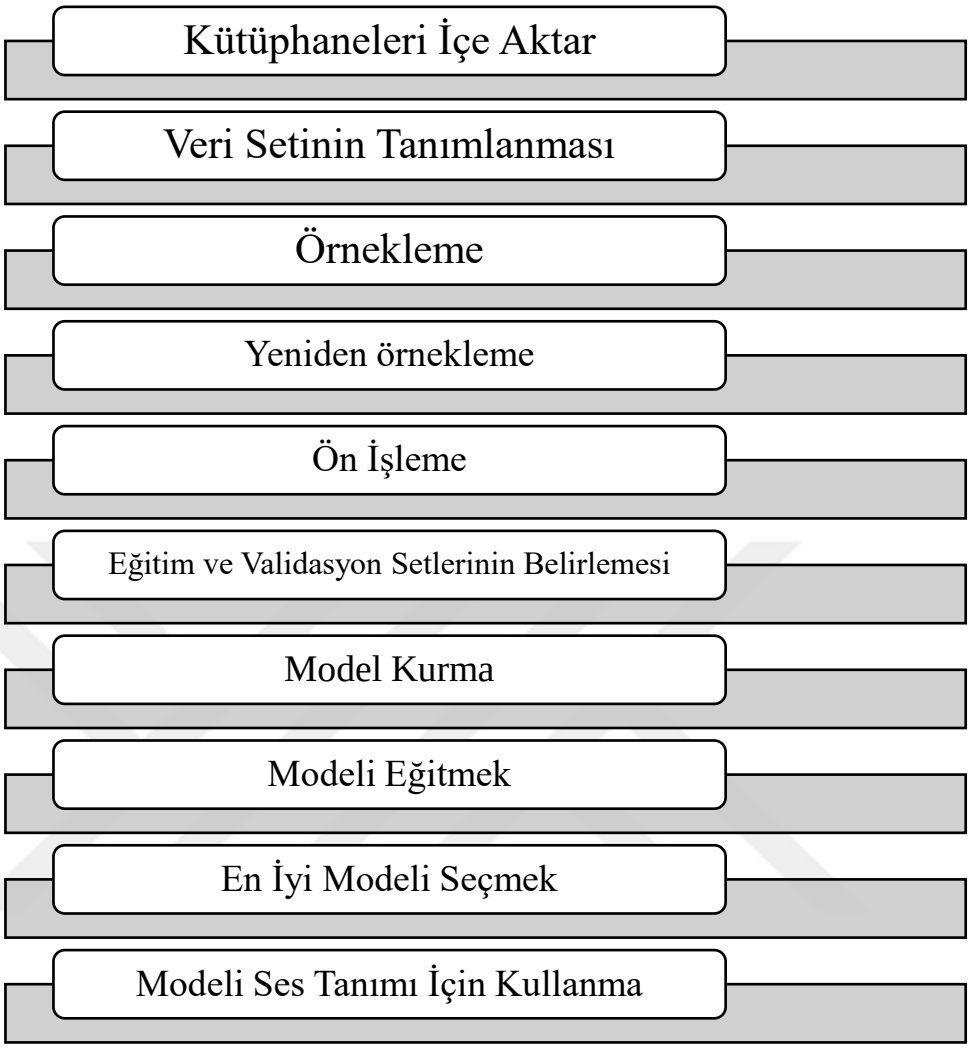
Derin öğrenme, makine öğrenmesi araştırmalarında yeni bir alandır. Bu tür öğrenme, aynı zamanda derin yapılandırılmış öğrenme veya hiyerarşik öğrenme olarak da adlandırılır. Derin öğrenme, genellikle yukarıda sözü edilen bu alanı adlandırmak için kullanılır. Bu bölümde bu algoritmadan bahsedilecektir. Giriş bölümünde yapay zekâ; veri madenciliği ve makine öğrenmesi algoritmaları arasındaki ilişki hakkında derin öğrenme algoritmaları

tartışılmıştır. Bunun yanı sıra derin öğrenme algoritmaları, makine öğrenme algoritmalarının bir alt kümesidir. Bu yüzden bu algoritma türü, makine öğrenme algoritmaları ve yaklaşımlarında detaylı olarak açıklanmıştır.

Ayrıca derin öğrenme algoritmaları kullanan yöntemlerde de sabit adımlar vardır. Her modelin oluşturulması bir "ön işleme" ile başlamaktadır. Ön işleme, yapılandırılmamış verileri yapılandırılmış verilere dönüştürmeye ve metinleri en üst düzeye getirmeye çalışır. Ön işleme adımından sonra, yukarıda açıklanan kelime çıkarma ve seçme işlemi gerçekleştirilmektedir. Derin öğrenme algoritmalarında, doğrusal olmayan özellikler iç tabakalar tarafından elde edilir. Birtakım özellikler belirlenerek istenen kümeleme modeline ulaşılabilir ve asıl hali çıkarılan özellik modele gönderilir. Sınıflandırma modeli ile, daha önceden elde edilen özelliğin derin öğrenme algoritmaları kullanılarak istenen model oluşturulur. Kategori modelini oluşturduktan sonra değerlendirme aşaması gerçekleştirilir. Yapılacak araştırma, modeller oluşturmak için derin öğrenme algoritmaları kullanmaktadır. Bu özellik de bu şekilde üretilmektedir. Dolayısıyla doğru öğrenme tanıma modelini oluşturmak için derin öğrenme algoritmalarının kullanıldığı söylenebilir.

Derin öğrenme algoritmalarının çalışma yolu, temel olarak insan beyninden alınan fikir ve görsel korteksin insan beyninde nasıl çalıştığı sorusunun yanıtıdır. Yani insan beyninde bulunan görsel kortekseteki birincil hiyerarşinin nöronlarının aldıkları bilgilerin, kenarlara ve kütlelere karşı duyarlı olmalarıdır. Nöronlar cevap verene kadar çıktısı bir sonraki hiyerarşide devam etmektedir. Örneğin yüzler gibi daha sofistike yapılar böylece okunabilir. Beyin fonksiyonunu simüle eden ortak bir sinir ağı algoritması vardır.

Bu çalışmada, CNN derin öğrenme algoritmaları bu fikri uygulamak için kullanılmıştır. Aşağıda kullanılan unsur, Python kodlarının bir açıklamasıdır.



Şekil 4.1: Uygulanan Adımların Özeti

Şekil 4.1'de uygulanan adımların özeti gösterilmiştir. Şimdi bir bir bu adımlar anlatılacaktır. Bu adımlar, kütüphaneleri içe aktar, veri setinin tanımlanması, örnekleme, yeniden örnekleme, ön işleme, eğitim ve validasyon setlerinin belirlemesi, model kurma, modeli eğitmek, en iyi modeli seçmek ve modeli ses tanımı için kullanma olarak belirlenmiştir.

4.2.1 Kütüphaneleri İçe Aktar

Konuşma tanımanın derin öğrenme algoritmalar ile yapılması için ilk adım kütüphaneleri içe aktarmaktır. Bu bölümde de bu görev yapılmıştır.

os, librosa ,IPython.display as ipd, matplotlib.pyplot as plt, numpy as np, from scipy.io wavfile ve warnings gibi kütüphaneler bu kod için kullanılmıştır.

```
import os
import librosa
import IPython.display as ipd
import matplotlib.pyplot as plt
import numpy as np
from scipy.io import wavfile
import warnings
from sklearn.preprocessing import LabelEncoder
from keras.utils import np_utils
from sklearn.model_selection import train_test_split
from keras.layers import Dense, Dropout, Flatten, Conv1D, Input,
MaxPooling1D
from keras.models import Model
from keras.callbacks import EarlyStopping, ModelCheckpoint
from keras import backend as K
from matplotlib import pyplot
from keras.models import load_model
import random
```

Şekil 4.2: Kodlamada kullanılan kütüphaneler

Yukarıda kullanılan tüm python kütüphaneler listelenmiştir. Konuşma tanıma modeli oluşturmak için bu python dilindeki kütüphaneler kullanılmıştır. Python popüler bir programlama dilidir ve Guido van Rossum tarafından yaratılmıştır ve 1991 yılında piyasaya sürülmüştür. Python programlama dili son yıllarda özellikle makine öğrenme ve

derin öğrenme çalışmalarında sıkça kullanılmaya başlamıştır ve bu konuda çok iyi kütüphanelere sahiptir.

4.2.2. Veri Setinin Tanımlanması

Veri kümesi, 65.000 ses içermektedir. Bu veri setinde otuz kelime içeren 65.000 ses bulunmaktadır. Tablo 4.1'de veri setini oluşturan kelimeler verilmiştir. Bu kelimeler farklı kişiler seslerini kaydederek oluşturulmuştur. Ortalama her kelime için 2200 ses kaydı bulunmaktadır.

Tablo 4.1: Veri setini oluşturan kelimeler

bed	house	sheila
bird	left	six
cat	marvin	stop
dog	nine	three
down	no	tree
eight	off	two
five	on	up
four	one	wow
go	right	yes
happy	seven	zero

bed , bird, cat, dog, down, eight, five, four, go, happy, house, left, marvin, nine, no, off, on, one, right, seven, sheila, six, stop, three, tree, two, up, wow, yes, zero kelimelerinin ses kayıtları bu veri setinde mevcuttur. Bu veri seti hakkında detaylı bilgi bu[9] bağlantıda verilmiştir.

4.2.3. Örnekleme ve Yeniden Örnekleme

Ses frekansı, konuşmanın iletimi için kullanılan frekanslardan biridir. Gennelikle kullanılabilir ses frekans bandı yaklaşık 300 ila 3400 Hz arasında değişir. Çalışmada veri setindeki ses kayıtlarının frekansları 16000 hz olduğunu biliyoruz. Ama Pythonda kullanacağımız kütüphaneler 8000 hz frekanlarına göre tasarlanmıştır. Bu yüzden veri

[9] https://www.kaggle.com/c/tensorflow-speech-recognition-challenge

setinde olan ses kayıtlarının frekanslarının librosa.resample kütüphanesini kullanrak değiştirmemiz lazım. Veri setinden olan ses kayıtlarının örneklemesi 8000 hz'e göre yeniden yapılmıştır.

samples = librosa.resample(samples, sample_rate, 8000) ipd.Audio(samples, rate=8000)

4.2.4. Ön İşleme

Ön işlemede yapılandırılmamış ses verilerini yapılandırılmış ses verilerine dönüştürmek istiyoruz.

Daha önceki veri araştırma bölümünde, birkaç kaydın süresinin 1 saniyeden az olduğunu ve örnekleme oranının çok yüksek olduğunu görülmüştür. Ses dalgalarını okumak ve ses kayıtları ile başa çıkmak için ön işleme adımlarını kullanılmıştır.

1 saniyeden daha kısa komutlar kaldırılmıştır ve ön işleme adımlarını aşağıdaki kod pasajında tanımlanmıştır.

```
#ses kayıtlarının süresi
records_duration =[]
for label in labels:
    waves = [f for f in os.listdir(train_audio_path + '/'+ label) if f.endswith('.wav')]
    for wav in waves:
        sample_rate, samples = wavfile.read(train_audio_path + '/' + label + '/' + wav)
        records_duration.append(float(len(samples)/sample_rate))
plt.hist(np.array(records_duration))
```

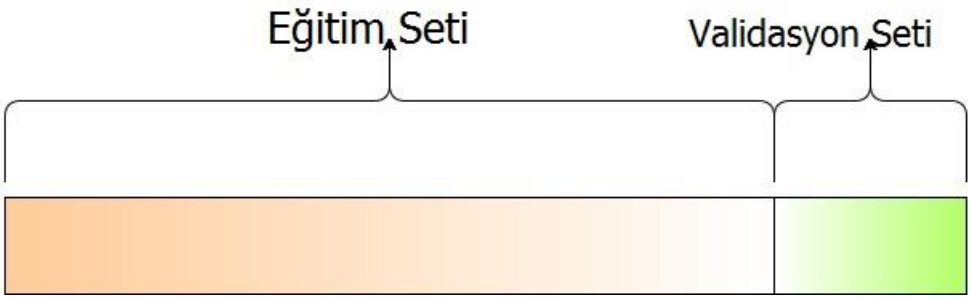
Şekil 4.3: Kodlamada ses kaydı süreleri tanımlanması

4.2.5. Eğitim ve Validasyon Setlerinin Belirlemesi

Bu adımda, eğitim ve validasyon setlerinin belirlemesi işlemi yapılmıştır. Eğitim ve validasyon setlerinin ne olduğu açıklanacaktır.

Eğitim Veri Kümesi: Modele uyması için kullanılan veri örneğidir.

Modeli eğitmek için kullandığımız gerçek veri kümesi (Sinir Ağı durumunda ağırlık ve önyargı). Model bu verileri görür ve öğrenir ve bu verilere göre modeli oluşturur.



Şekil 4.4: Ağı eğitmede kullanılacak veri oranları

Validasyon veri kümesi: Model hiperparametrelerini ayarlarken eğitim veri setine uygun bir modelin tarafsız bir değerlendirmesini sağlamak için kullanılan veri örneğidir. Doğrulama veri setindeki beceri, model konfigürasyonuna dahil edildiğinden değerlendirme daha önyargılı hale gelmektedir.

Doğrulama seti belirli bir modeli değerlendirmek için kullanılır. Makine öğrenimi mühendisleri olarak biz bu verileri model hiperparametrelerinde ince ayar yapmak için kullanıyoruz. Bu nedenle model zaman zaman bu verileri görür, ancak bundan asla "öğrenme" yapmaz. Doğrulama seti sonuçlarını kullanırız ve üst seviye hiperparametreleri güncelleriz. Dolayısıyla, bir şekilde belirlenmiş doğrulama bir modeli etkiler, ancak dolaylı olarak.

Bu çalışmada eğitim ve test veri kümelerini belirlemek için train_test_split kütüphanesi kullanılmıştır. Bu kütüphane kullanılarak veri kümesi iki alt gruba ayrılıyor. Eğitim ve test veri kümelerini ne oranda olduğunu manuel olarak belirleye imaknı olduğu için bu kısımda verilerin 80%'ni eğitim için ver 20%'si test için kullanılmıştır. Bu oran belirleme kodumuzun alt kısmında belirlenmiştir.

```
x_tr, x_val, y_tr, y_val = train_test_split (np.array(all_wave),np.array(y),stratify=y,test_size = 0.2,random_state=777,shuffle=True)
```

Şekil 4.5: Kodlamada veri seti ayarlamaları

4.2.6. Model Kurma

Derin öğrenme modellerini geliştirmeye ve değerlendirmeye yönelik en güçlü ve kullanımı kolay Python kitaplıklarından biri de Keras'tır. Keras fonksiyonel API kullanarak model kurulmuştur. Conv1d kullanarak ses, metne çevrilmiştir ve model

Conv1d kullanılarak çevirilmiştir. Conv1d, evrişimi yalnızca bir boyutta gerçekleştiren evrişçi bir sinir ağıdır.

```
inputs = Input(shape=(8000,1))
#Birinci Conv1D katman
conv = Conv1D(8,13, padding='valid', activation='relu', strides=1)(inputs)
conv = MaxPooling1D(3)(conv)
conv = Dropout(0.3)(conv)
#İkinci Conv1D katman
conv = Conv1D(16, 11, padding='valid', activation='relu', strides=1)(conv)
conv = MaxPooling1D(3)(conv)
conv = Dropout(0.3)(conv)
#Üçüncü Conv1D katman
conv = Conv1D(32, 9, padding='valid', activation='relu', strides=1)(conv)
conv = MaxPooling1D(3)(conv)
conv = Dropout(0.3)(conv)
#Dordunuc Conv1D katman
conv = Conv1D(64, 7, padding='valid', activation='relu', strides=1)(conv)
conv = MaxPooling1D(3)(conv)
conv = Dropout(0.3)(conv)
#Flatten katman
conv = Flatten()(conv)
#Dense Katman 1
conv = Dense(256, activation='relu')(conv)
conv = Dropout(0.3)(conv)
#Dense Katman 2
conv = Dense(128, activation='relu')(conv)
conv = Dropout(0.3)(conv)
outputs = Dense(len(labels), activation='softmax')(conv)
model = Model(inputs, outputs)
```

Şekil 4.6: Model kurma ayarları

Kullanılan modelin özeti aşağıda gösterilmiştir.

Tablo 4.2: Parametre ayarları ve çıktılar

Katman (Türü)	Çıktı Şekli	Parametre
input_1 (Input Layer)	(None, 8000, 1)	0
conv1d_1 (Conv1D)	(None, 7988, 8)	112
max_pooling1d_1 (MaxPooling1	(None, 2662, 8)	0
dropout_1 (Dropout)	(None, 2662, 8)	0
conv1d_2 (Conv1D)	(None, 2652, 16)	1424
max_pooling1d_2 (MaxPooling1	(None, 884, 16)	0
dropout_2 (Dropout)	(None, 884, 16)	0
conv1d_3 (Conv1D)	(None, 876, 32)	4640
max_pooling1d_3 (MaxPooling1	(None, 292, 32)	0
dropout_3 (Dropout)	(None, 292, 32)	0
conv1d_4 (Conv1D)	(None, 286, 64)	14400
max_pooling1d_4 (MaxPooling1	(None, 95, 64)	0
dropout_4 (Dropout)	(None, 95, 64)	0
flatten_1 (Flatten)	(None, 6080)	0
dense_1 (Dense)	(None, 256)	1556736
dropout_5 (Dropout)	(None, 256)	0
dense_2 (Dense)	(None, 128)	32896
dropout_6 (Dropout)	(None, 128)	0
dense_3 (Dense)	(None, 10)	1290

Bu çalışmada evrişimsel(Konvolüsyonel) sinir ağları konuşma tanıma modeli için kullanılmıştır. Kullanılan ve kurulan konvolüsyonel sınır ağı detayları yukarıda verilmiştir. Kurulan model konuşma tanıma uygulamasını yapmak için kullanılmıştır.

4.2.7. Modeli Eğitmek

Makine öğreniminde, temel olarak test verilerini tahmin etmek için bir model oluşturmaya çalışıyoruz. Bu nedenle, eğitim verilerini modele sığdırmak ve verileri test etmek için test etmek için kullanıyoruz. Oluşturulan modeller, test seti olarak adlandırılan bilinmeyen sonuçları tahmin etmektir. Belirttiğiniz gibi, veri seti doğrulukları, hassasiyetleri üzerinde çalışarak ve test ederek kontrol etmek için eğitim ve test setine bölünmüştür.

Bir makine öğrenme modeli, gerçek dünyadaki sürecin matematiksel bir gösterimi olabilmektedir. Bir makine öğrenme modeli oluşturmak için, bir makine öğrenme algoritmasına eğitim verisi sağlamanız gerekmektedir. Bu çalışmada derin öğrenme kullanılarak model eğitip oluşturmaya çalışımıştır.

Makine Öğrenmesi algoritması, eğitimin gerçek dünya verileriyle başlamadan önce başında alınan hipotezdir. Derin öğrenme algoritması derken, derin öğrenme tarafından tanımlandığı gibi benzer özellikleri tanımlayan bir fonksiyonlar kümesidir ve bu fonksiyonlar grubundan eğitim verisine en uygun olanı seçmektedir.

Makine öğrenimi için eğitim alırken, eğitim verilerini içeren bir algoritma iletirsiniz. Bu çalışmada bu algoritma, derin öğrenme algorştmasıdır. Öğrenme algoritması, eğitim verilerinde girdi parametrelerinin hedefe karşılık geleceği şekilde kalıplar bulur. Eğitim sürecinin çıktısı, tahminlerde bulunmak için kullanabileceğiniz bir makine öğrenme modelidir. Bu süreç aynı zamanda “öğrenme” veya “eğitmek”olarak da adlandırılır.

```
kt_model=model.fit(x_train, y_train ,epochs=50, callbacks=[es,mc], batch_size=32,
validation_data=(x_validation,y_validation))
```

Şekil 4.7: Model eğitim ve doğrulama kodu

Bir sinir ağının fitlemesi, girdilerin çıkışlara iyi bir şekilde eşlenmesi için model ağırlıklarını güncellemek için bir eğitim veri setinin kullanılmasını içerir. Bu eğitim süreci, sinir ağları modeli için olası değerler alanı araştırarak, egzersiz veri setinde iyi bir performansa yol açan bir dizi ağırlık için bir optimizasyon algoritması kullanılarak çözülmektedir.

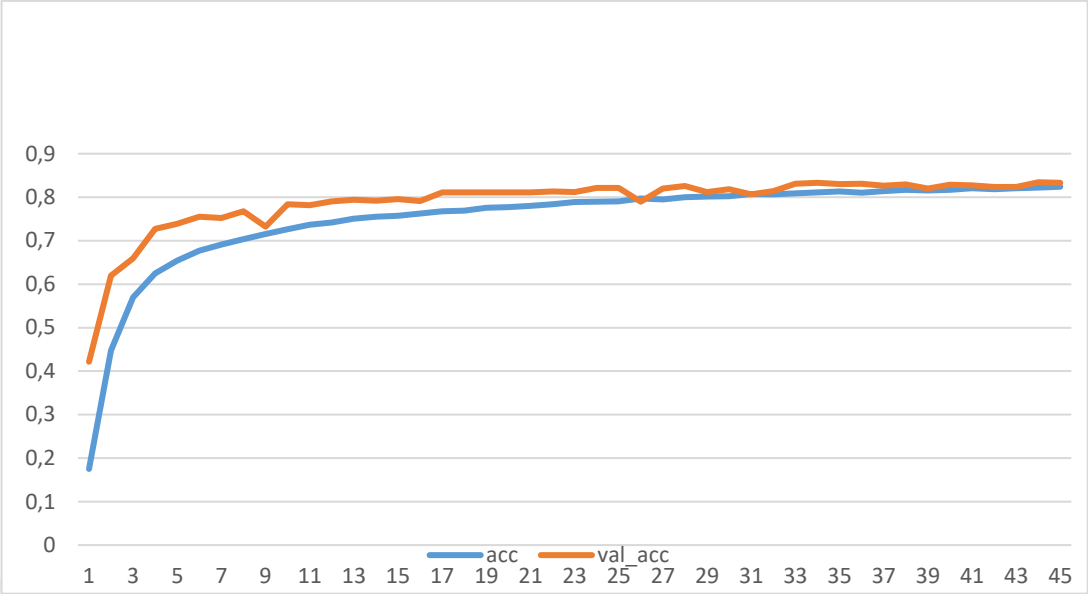
4.2.8. En İyi Modeli Seçmek

Konuşma modeli oluşturmak için en iyi ve başarılı modeli seçmek lazım. Konuşma tanıma için hangi model en başarılıdır? Bu sorunun cevabını bulmak için modeli 48 kere eğitik ve en başarılı model 83% oranlar seçilmiş oldu.

```
Epoch 00038: val_acc improved from 0.82233 to 0.82783, saving model to kt_model.hdf5
Epoch 39/50
46601/46601 [==============================] - 36s 783us/step - loss: 0.6346 - acc: 0.8059 - val_loss: 0.6251 - val_acc: 0.8167

Epoch 00039: val_acc did not improve from 0.82783
Epoch 40/50
46601/46601 [==============================] - 37s 805us/step - loss: 0.6348 - acc: 0.8058 - val_loss: 0.6893 - val_acc: 0.7896

Epoch 00040: val_acc did not improve from 0.82783
Epoch 41/50
46601/46601 [==============================] - 37s 789us/step - loss: 0.6420 - acc: 0.8065 - val_loss: 0.6396 - val_acc: 0.8087

Epoch 00041: val_acc did not improve from 0.82783
Epoch 42/50
46601/46601 [==============================] - 36s 774us/step - loss: 0.6224 - acc: 0.8102 - val_loss: 0.6891 - val_acc: 0.7965

Epoch 00042: val_acc did not improve from 0.82783
Epoch 43/50
46601/46601 [==============================] - 36s 777us/step - loss: 0.6217 - acc: 0.8104 - val_loss: 0.7290 - val_acc: 0.7817

Epoch 00043: val_acc did not improve from 0.82783
Epoch 44/50
46601/46601 [==============================] - 36s 772us/step - loss: 0.6239 - acc: 0.8086 - val_loss: 0.6345 - val_acc: 0.8108

Epoch 00044: val_acc did not improve from 0.82783
Epoch 45/50
46601/46601 [==============================] - 36s 774us/step - loss: 0.6105 - acc: 0.8133 - val_loss: 0.6156 - val_acc: 0.8188

Epoch 00045: val_acc did not improve from 0.82783
Epoch 46/50
46601/46601 [==============================] - 36s 776us/step - loss: 0.6092 - acc: 0.8130 - val_loss: 0.6491 - val_acc: 0.8103

Epoch 00046: val_acc did not improve from 0.82783
Epoch 47/50
46601/46601 [==============================] - 36s 773us/step - loss: 0.6058 - acc: 0.8140 - val_loss: 0.6033 - val_acc: 0.8233

Epoch 00047: val_acc did not improve from 0.82783
Epoch 48/50
46601/46601 [==============================] - 36s 778us/step - loss: 0.6039 - acc: 0.8164 - val_loss: 0.6056 - val_acc: 0.8235

Epoch 00048: val_acc did not improve from 0.82783
Epoch 00048: early stopping
```

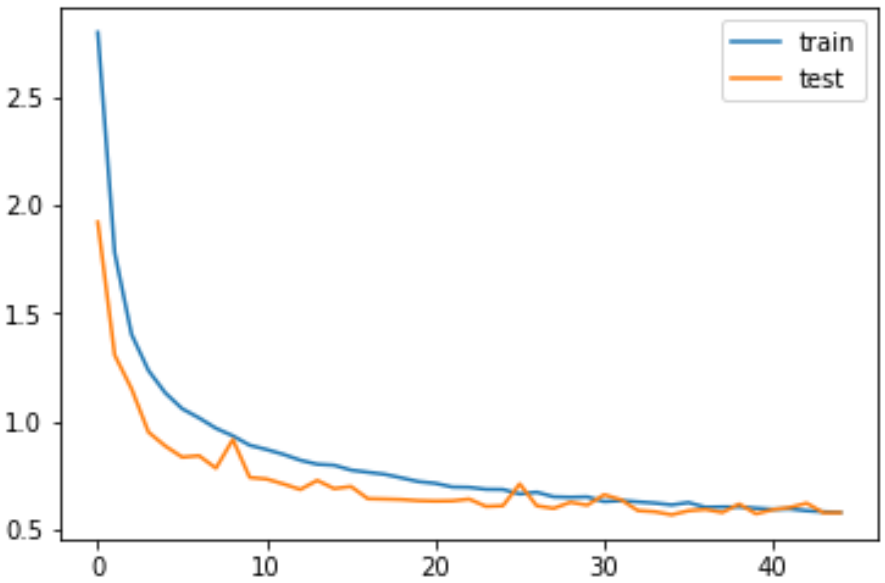
Şekil 4.8: Gerçekleştirilen modelin iterasyon sayısına göre doğruluk oranları

Model kendini eğitmeyi tekrarladığı zaman daha iyi sonuçlar vermektedir. İlk etaplarda modelin dikkatı çok düşük ama ilerledikçe daha iyi noktalara gelimiştir. Şekil 4.9'da bu konu gösterilmiştir.



Şekil 4.9: Modeli eğitmede dikkat sonuçları

Şekil 4.9'da görüldüğü gibi 45 inci epoch en iyi sonuç elde edilmiştir. Genel olarak sinir ağlarında, bir epoch tam eğitim setinden geçen tek bir geçiştir. Eğitim setini yalnızca bir kez çalıştırmazsınız, geri yayılım algoritmanızın ağırlıkların birleşiminde kabul edilebilir bir doğruluk düzeyiyle birleşmesi için binlerce epoch alabilir. Ama bu çalışmada 45'inci epochda en iyi sonuç elde edilmiştir.



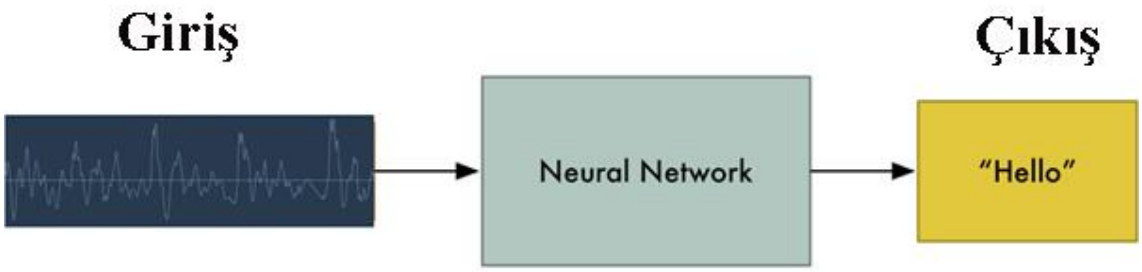
Şekil 4.10: Modeli eğitmede hata sonuçları

Şekil 4.10'da görüldüğü gibi 45 inci epoch en az hata elde edilmiştir ve en iyi model elde edilmektedir.

Otomatik konuşma tanıma sisteminin ana amacı, insan konuşmalarını simüle edebilecek bir sistem kurmaktır. Bu sistem ne kadar dikkatli olursa o kadar başarılıdır. Hata oranı ne kadar az olursa otomatik konuşma tanıma sisteminin daha başarılı olduğu ortaya çıkıyor. Burada az hata demek az yanlış kelime tahmin etmek demektir. Sesler ne kadar doğru metne çevirilirse o kadar başarı oranı yüksektir. Şimdi elde edilen sistem kullanılarak ses tanıma yapılacaktır.

4.2.9. Modeli Ses Tanımı İçin Kullanma

En iyi sonuç elde edildikten sonra, model ses tanıma için kullanılacaktır.



Şekil 4.11: Modeli Ses Tanımı İçin Kullanma Örneği

Şekil 4.11'de modeli ses tanımı için kullanma örneği gösterilmiştir. Otomatik konuşma tanıma sistemi bir donanım ve yazılım sistemidir; burada giriş sesin (konuşmanın) sesidir ve çıktı da bu konuşulan sözcüklerin tanımlanması şeklindedir. Ortadaki kutu, genellikle mikrofon tarafından alınan konuşmayı analiz edebilen ve bir makine tarafından kullanılabilen bir metin biçiminde kopyalayan tüm bir sistemdir. Bu sistem bizim çalışmada derim öğrenme algoritması kullanılarak yapılmıştır.

```
model=load_model(kt_model')

def predict(audio):

   prob=model.predict(audio.reshape(1,8000,1))

   index=np.argmax(prob[0])

   return classes[index]

index=random.randint(0,len(x_val)-1)

samples=x_val[index].ravel()

print("Audio:",classes[np.argmax(y_val[index])])

ipd.Audio(samples, rate=8000)
```

Şekil 4.12: Oluşturulan modeli kullanma

Derin öğrenme algoritması kullanılarak model oluşturulmuştur ve daha sonra model, ses tanımı için kullanılmıştır. Yukarıdaki kodda model kullanılarak kaç kelime doğru şekilde tahmin edilmiştir.

Şekil 4.13: Ses Tanıma İçin Kullanım

Şekil 4.13'de nasıl model ses tanıma için kullanılmış görülmektedir. "Three" ve "stop" kelimeleri model tarafından doğru şekilde metne dönüştürülmüştür.

4.3. Tartışma

Yapay zeka, bir insan gibi düşünerek ve hareket ederek tasarlanıp programlanmış makinelerdir. Yapay zeka günlük hayatımızın önemli bir parçası haline gelmektedir. Hayatımız yapay zeka tarafından değiştirildi çünkü bu teknoloji günden güne çok çeşitli hizmetlerde kullanılmaktadır. Konuşma tanıma bu hizmetlerden biridir.

İş dünyası liderleri ve yenilikçiler rekabet avantajının yanı sıra maliyet ve zaman tasarrufu da sağlamak için yapay zeka vaadine ulaşmaya çalışırken, teknoloji sanayileri yeni ürünler, işlemler ve becerilerle finans sektöründen imalata geçiriyor. 150 yaşındaki Heineken gibi şirketler yapay zeka, topladıkları büyük miktarda veri ve Pazarlama kararları ve girişimlerini yönlendiren şeyler İnterneti, operasyonları ve müşteri hizmetlerini geliştiriyor. Küresel tedarik zincirlerini yönetmekten teslimat rotalarını optimize etmeye kadar, yapay zeka her büyüklükteki şirketlere ve tüm sektörlerdeki şirketlere, kaynak malzemesinden satış ve muhasebe işlemlerine kadar iş yaşamının her aşamasında üretkenliği ve karlılığı arttırmalarında yardımcı olmaktadır. Şirketlerin, ürünleri ve hizmetleri her zamankinden daha iyi tasarlaması, üretmesi ve sunmasına olanak vermektedir. Konuşma tanıma artık her şirket için kullanılması zorunlu hale gelmiştir. Özellikle teknoloji şirketleri konuşma tanımayı çok fazla kullanıyorlar. Konuşma tanıma sistemleri o yüzden çok başarılı olmak zorundadır.

Bu araştırmada, İngilizce konuşma tanıma konusu ele alınmıştır ve yapay zeka yöntemleri kullanılarak bir konuşma tanıma uygulaması sunulmuştur. Konuşma tanıma uygulaması içinde makine öğrenme ve derin öğrenme algoritması kullanılmıştır. Konuşma tanıma modelini oluşturmak için kullanılan Konvolüsyonel(Evrişimsel) Sinir Ağları algoritması kullanılarak doğruluk oranı %83 olan bir konuşma tanıma uygulaması sunulmuştur. Uygulamanın sonuçlarına dayanarak, kabul edilebilir bir performans sağlayabilen bir konuşma tanıma modeli ve uygulaması elde edilmiştir.

Net bir şekilde sonuçların literatüre katkısını anlatmak gerekirse, derin öğrenme algoritmalarının konuşma tanıma uygulamalarında, başarılı sonuçlar verebileceği olarak tespit edilmiştir.

Erdoğan ve arkadaşları LVSCR yöntemi kullanarak, konuşma tanıma modeli oluşturmaya çalışmışlar. Büyük kelime dağarcığını belirlemek ve büyük kelime sürekli konuşma

tanıma için dil modelleri geliştirmek için yeni yöntemler sunmuşlardır. Elde ettikleri dikkat oranı %47-%53 arasıdır (Erdogan,Buyuk & Oflazer,2005).

Mirsamadi ve arkadaşları, konuşma duygularını tanımada özellik öğrenme için farklı RNN mimarileri kullanmışlar. Derin RNN'leri kullanarak, hem çerçeve düzeyinde karakterizasyonu hem de daha uzun zaman aralıklarında zamansal toplamayı öğrenebildiğimiz gösterilmiştir. Dahası, basit bir dikkat mekanizması kullanarak, ağın bir ifadenin duygusal olarak dikkat çekici kısımlarına odaklanmasını sağlayan yeni bir ağırlıklı zaman havuzu oluşturma stratejisi önerilmiştir. IEMOCAP verileri üzerinde yapılan deneyler, öğrenilen özelliklerin, sabit tasarlanmış özellikler kullanan geleneksel SVM tabanlı SER ile karşılaştırıldığında daha iyi sınıflandırma doğruluğu sağladığını göstermektedir. Bu çalışmada da %53-%64 arası dikkat oranı raporlanmıştır (Mirsamadi,Barsoum & Zhang, 2017).

Büyük, 2018 senesinde,konuşma tanıma için model geliştirmiştir. Yapılan çalışmada mobil platformlar için yüksek başarım ile çalışan bir konuşma tanıma sisteminin gerçekleştirilmesi amaçlanmıştır. Yapılan çalışmada akıllı telefonlardan alınmış kayıtlardan oluşan yeni bir ses veri tabanı oluşturulmuştur. Sistemin performansı üç farklı konuşma tanıma uygulaması kullanılarak ölçülmüştür. i) Televizyon kumanda uygulaması, ii) Sesli mesaj uygulaması, iii) Genel metin yazdırma uygulaması. Yaptığımız testlerde tanıma performansının televizyon kumanda uygulaması için %95'in üzerinde olduğu görülmüştür. Sesli mesaj ve genel metin yazdırma uygulamalarında yaklaşık %40 ve %60 başarım oranları elde edilmiştir (Büyük, 2018).

Yapılan çalışmalara bakıldığında bu tezde elde edilen sonuçların kabul edilebilir olduğu görülmektedir. Bu arada tam bir karşılaştırma yapılması için verilerin aynı olması gerekmektedir. Andrade ve arkadaşları tarafından sunulan modelde de derin öğrenme algoritmaları kullanılmıştır. Elde edilen sonuç %94 olarak belirlenmiştir ama bu çalışmada sadece iki kategori sınıflandırmak içim kullanılmıştır. Bizim çalışmada kategori sayısı oldukça fazladır. Bu yüzden yapılan çalışma ve elde edilen sonuçların kabul edilebilir düzeyde olduğu söylenebilir (Andrade, Leo, Viana & Bernkopf, 2018).

5. SONUÇ

Konuşma tanımanın kökleri, 1980'lerin başında Bell Laboratuvarları'nda yapılan araştırmalara dayanmaktadır. Bu dönemdeki erken sistemler, tek konuşmacılar ve belirli kelimelerle sınırlı kalmıştır. Modern konuşma tanıma sistemleri, eski hâllerinden günümüzdeki seviyelerine ulaşmak için uzun bir yol kat etmiştir. Bu sistemlerin, birden fazla konuşmacıdan gelen konuşmaları tanıyabilme ve birden çok dilde geniş bir kelime bilgisine sahip olma yetenekleri mevcuttur. Konuşma tanımanın ilk bileşeni kesinlikle "konuşma" kavramıdır. Konuşma; fiziksel sesten mikrofona, sonra elektriksel bir sinyale ve ardından "analog-dijital dönüştürücü" vasıtasıyla dijital verilere dönüştürülmelidir. Konuşma dijitalleştirildikten sonra, sesi metne dönüştürme amacına yönelik birkaç model mevcuttur.

Modern konuşma tanıma sistemlerinin çoğu, Gizli Markov Modeli (HMM) adlı kavrama dayanmaktadır. Bu yaklaşım, konuşma sinyalinin yeterince küçük zaman ölçeklerinde (örneğin, on milisaniyede) görüntülendiğinde, "durağan bir işlem" olarak kabul edilebileceği düşüncesinden kaynaklanır. Anlamlandırma süreci, istatistiksel özelliklerin zaman içinde değişmediği bir süreçtir.

Ortak bir Markov gizli zincirinde, konuşma sinyali on milisaniyelik parçalara bölünmüştür. Sinyal gücü, aslında frekansın bir fonksiyonudur. Sinyal gücünün bir parçası olan her bölümün güç spektrumu, "Cepstral Katsayıları" olarak bilinen gerçek sayılara eşlenir. Bu vektörün boyutları genellikle küçük ve bazen 2 değeri kadar düşüktür. Ancak daha kesin sistemlerde 2 veya daha fazla boyutta da olabilirler. Markov gizli zincirinin (HMM) son çıktısı bu vektörlerin bir dizisidir.

Metnin konuşmasını "çözmek" için bir vektör dizisi, bir veya daha fazla fonemle örtüşür. Fonem, konuşmanın temel birimidir. Bu hesaplamalar eğitim gerektirmektedir. Bir ses tonunun sesi, konuşmacıdan konuşmacıya değişir. Daha sonra, verilen fonem sırasını üreten en muhtemel kelimeyi veya kelimeleri belirlemek için özel bir algoritma uygulanır.

Kişi böyle bir işlemin maddiyat açısından pahalı olduğunu düşünebilir. Birçok modern konuşma tanıma sisteminde yapay sinir ağları, Markov'un gizli zincir tespitinden önce "Özellik Dönüşümü" ve "Boyut Düşüşü" için teknikler kullanarak konuşma tanıma sinyallerini basitleştirmiştir. Ayrıca Sesli Etkinlik Dedektörleri (VAD), tek bir ses sinyalini yalnızca konuşma içermesi muhtemel alanlara indirgemek için kullanılır.

Bu araştırmada, ingilizce konuşma tanıma konusu tartışılmıştır. Konuşma tanıma, sesi metne dönüştürmek isteyen yapay zeka bilimlerinde ana alanlarından biridir. Yapay zeka artık her alanda liderlik yapmaya başlamıştır. Konuşma tanımada da makine öğrenme ve derin öğrenme konuları gittikçe başarılı yöntemler sunmuştur.

Konuşma tanıma, son yıllarda daha fazla dikkat çekti, fakat şimdiye kadar, derin öğrenme algoritmalarının kullanımı fazla ilgi görmemiştir. Bu çalışmada, ilk önce konuşma tanıma ve derin öğrenme algoritmalarından bahsedilmiştir. Konuşma tanıma modelini oluşturmak için kullanılan Konvolüsyonel(Evrişimsel) Sinir Ağları algoritması daha sonra açıklanmıştır.

Uygulamanın sonuçlarına dayanarak, kabul edilebilir bir performans sağlayabilen bir konuşma tanıma modeli elde edilmiştir. Konuşma tanıtma modeli için elde edilen en iyi doğruluk oranı %83 olarak tespit edilmiştir. Bu çalışmada kullanılan veri, 65.000 ses bulunan İngilizce ses veri kümesidir. Araştırmaya göre, derin öğrenme algoritmalarının konuşma tanıma sorununu çözebileceği sonucuna varılmıştır.

Önerilen yöntemin başarısına ek olarak, bu yöntem ve araştırmalar genişletilebilir ve verimliliği artırmak için başka yöntemler kullanılabilir. Türkçe veri setleri üzerinde araştırma yapmak, daha doğru konuşma tanıma modellerinin elde edilmesinde faydalı olabilir. Ne yazık ki, Türkçe dili için derin öğrenme algoritması için uygun veri seti yoktur.

KAYNAKÇA

Abdel-Hamid, O., Mohamed, A. R., Jiang, H., Deng, L., Penn, G., & Yu, D. (2014). Convolutional neural networks for speech recognition. *IEEE/ACM Transactions on audio, speech, and language processing*, *22*(10), 1533-1545.

Aktürk, F. (2015). *Örneklem Tabanlı Gürbüz Konuşma Tanıma* (Doctoral dissertation, Fen Bilimleri Enstitüsü).

Alpaydin E., Optical Character Recognition Using Artificial Neural Networks, 1989 First IEEE International Conference on Artificial Neural Networks, London, United Kingdom, 16-18 October 1989.

Amodei, D., Ananthanarayanan, S., Anubhai, R., Bai, J., Battenberg, E., Case, C., ... & Chen, J. (2016, June). Deep speech 2: End-to-end speech recognition in english and mandarin. In *International conference on machine learning* (pp. 173-182).

Andrade, D. C., Leo, S., Viana, M. L. D. S., & Bernkopf, C. (2018). A neural attention model for speech command recognition. arXiv preprint arXiv:1808.08929.

Arnold D., Balkan L., Meijer S., Humphreys R., Sadler L., Machine Translation: An Introductory Guide, UK: NEC Blackwell, Manchester, 1994

Arora, V., & Reetz, H. (2017). Automatic speech recognition: What phonology can offer. *The Speech Processing Lexicon: Neurocognitive and Behavioural Approaches*, *22*, 211.

Bennett, I. M., Babu, B. R., Morkhandikar, K., & Gururaj, P. (2015). *U.S. Patent No. 9,076,448*. Washington, DC: U.S. Patent and Trademark Office.

Bennett, I., Babu, B. R., Morkhandikar, K., & Gururaj, P. (2015). *U.S. Patent No. 9,190,063*. Washington, DC: U.S. Patent and Trademark Office.

Boser B., Denker J. S., Henderson D., Howard R. E., Hubbard W., Jackel L. D., Lecun Y., Backpropagation Applied to Handwritten Zip Code Recognition, *Neural Computation*, 1989, 1(4), 541–551.

Büyük, O. (2018). Mobil araçlarda Türkçe konuşma tanıma için yeni bir veri tabanı ve bu veri tabanı ile elde edilen ilk konuşma tanıma sonuçları. *Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi*, *24*(2), 180-184.

Büyük, O. (2018). Mobil araçlarda Türkçe konuşma tanıma için yeni bir veri tabanı ve bu veri tabanı ile elde edilen ilk konuşma tanıma sonuçları. Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi, 24(2), 180-184.

Cakir, M. Y., & Sirin, Y. (2018). Comparison of windowing techniques for speech recognition system [Pencereleme tekniklerinin konuşma tanima sistemi için karşilaştirilmasi].

Chao, Y., & Bourguet, M. L. (2017, September). What Speech Recognition Accuracy is Needed for Video Transcripts to be a Useful Search Interface?. In *International Conference on Speech and Computer* (pp. 820-828). Springer, Cham.

Chattopadhyay, R., Vintzileos, A., & Zhang, C. (2013). A description of the Madden–Julian oscillation based on a self-organizing map. *Journal of climate*, *26*(5), 1716-1732.

Chiu, C. C., Sainath, T. N., Wu, Y., Prabhavalkar, R., Nguyen, P., Chen, Z., ... & Jaitly, N. (2018, April). State-of-the-art speech recognition with sequence-to-sequence models. In *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 4774-4778). IEEE.

Chorowski, J. K., Bahdanau, D., Serdyuk, D., Cho, K., & Bengio, Y. (2015). Attention-based models for speech recognition. In *Advances in neural information processing systems* (pp. 577-585).

Cireşan, D., Meier, U. & Schmidhuber, J. (2012). Multi-column deep neural networks for image classification. arXiv preprint arXiv:1202.2745.

Çakır, M. Y. (2017). *Gerçek zamanlı yüksek kalitede ses tanıma* (Master's thesis, İstanbul Sabahattin Zaim Üniversitesi, Fen Bilimleri Enstitüsü, Bilgisayar Mühendisliği Anabilim Dalı).

Çalik, N., Kurban, O. C., Yilmaz, A. R., Ata, L. D., & Yildirim, T. (2017, May). Signature recognition application based on deep learning. In *2017 25th Signal Processing and Communications Applications Conference (SIU)* (pp. 1-4). IEEE.

Deng, L., & Yu, D. (2014). Deep learning: methods and applications. Foundations and Trends® in Signal Processing, 7(3–4), 197-387.

Deng, L., Yu, D. (2014). Deep Learning: Methods and Applications. Foundations and Trends® in Signal Processing, 7(3–4), 197–387. Https://doi.org/10.1561/2000000039

Diğken, G., & İbrikçi, T. (2015, May). Recognition of non-speech sounds using Mel-frequency cepstrum coefficients and dynamic time warping method. In *2015 23nd Signal Processing and Communications Applications Conference (SIU)* (pp. 144-147). IEEE.

Durak, M. H., Seke, E., & Özkan, K. (2015, May). Denoising speech signal using common vector approach. In *2015 23nd Signal Processing and Communications Applications Conference (SIU)* (pp. 1961-1964). IEEE.

Erdogan, H., Buyuk, O., & Oflazer, K. (2005). Incorporating language constraints in sub-word based speech recognition. In IEEE Workshop on Automatic Speech Recognition and Understanding, 2005. (pp. 98-103). IEEE.

Fu, K. S. (2019). *Applications of pattern recognition*. CRC press.

Ghosh-Dastidar, S., & Adeli, H. (2009). A new supervised learning algorithm for multiple spiking neural networks with application in epilepsy and seizure detection. *Neural networks*, *22*(10), 1419-1431.

Gorricha, J., Lobo, V., & Costa, A. C. (2013). A framework for exploratory analysis of extreme weather events using geostatistical procedures and 3D self-organizing maps. *International Journal on Advances in Intelligent Systems*, *5*(NA), 16-26.

Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., Liu, T., Wang, X., Wang, G., Cai, J. & Chen, T. (2018). Recent advances in convolutional neural networks. Pattern Recognition, 77, 354-377.

Gu, X., Zhang, H., Zhang, D., Kim, S. (2016, November). Deep API learning. In Proceedings of the 2016 24th ACM SIGSOFT International Symposium on Foundations of Software Engineering (pp. 631-642). ACM.

Han, S., Kang, J., Mao, H., Hu, Y., Li, X., Li, Y., ... & Yang, H. (2017, February). Ese: Efficient speech recognition engine with sparse lstm on fpga. In *Proceedings of the 2017 ACM/SIGDA International Symposium on Field-Programmable Gate Arrays* (pp. 75-84). ACM.

Iglesias, G., Kale, D. C., & Liu, Y. (2015). An examination of deep learning for extreme climate pattern analysis. In *The 5th International Workshop on Climate Informatics*.

Juang B. H., Rabiner L., Fundamentals of Speech Recognition, Signal Processing Series, PTR Prentice Hall, New Jersey, 1993.

Kalchbrenner, N., Grefenstette, E., & Blunsom, P. (2014). A convolutional neural network for modelling sentences. arXiv preprint arXiv:1404.2188.

Kashima, H., Kato, T., Yamanishi, Y., Sugiyama, M., & Tsuda, K. (2009, April). Link propagation: A fast semi-supervised learning algorithm for link prediction. In *Proceedings of the 2009 SIAM international conference on data mining* (pp. 1100-1111). Society for Industrial and Applied Mathematics.

Ker, J., Wang, L., Rao, J., & Lim, T. (2017). Deep learning applications in medical image analysis. Ieee Access, 6, 9375-9389.

King-Sun F., Azriel R., Pattern Recognition and Image Processing, *IEEE Transactions on Computers*, 1976, c-25(12), 1336-1346.

Ko, T., Peddinti, V., Povey, D., & Khudanpur, S. (2015). Audio augmentation for speech recognition. In *Sixteenth Annual Conference of the International Speech Communication Association*.

Koruyan, K. (2015). Canlı İnternet Yayınları İçin Otomatik Konuşma Tanıma Tekniği Kullanılarak Alt Yazı Oluşturulması. *International Journal of Informatics Technologies*, *8*(2), 111.

Krizhevsky, A., Sutskever, I. & Hinton, G.E. (2012). Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems, 1097-1105.

Lawrence, S., Giles, C.L., Tsoi, A.C. & Back, A.D. (1997). Face recognition: A convolutional neural-network approach. IEEE transactions on neural networks, 8(1), 98-113.

Lecun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. nature, 521(7553), 436.

Li, H., Lin, Z., Shen, X., Brandt, J., & Hua, G. (2015). A convolutional neural network cascade for face detection. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 5325-5334).

Li, M., Guo, Y., & Chen, Y. (2017, December). Cnn-based commercial detection in tv broadcasting. In *Proceedings of the 2017 VI International Conference on Network, Communication and Computing* (pp. 48-53).

Liang, M., & Hu, X. (2015). Recurrent convolutional neural network for object recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 3367-3375).

Lowe, D.G. (1999). Object recognition from local scale-invariant features. In iccv, 99(2), 1150-1157.

Mairesse, F., Raccuglia, P. F., & Vitaladevuni, S. N. P. (2016). *U.S. Patent No. 9,484,021*. Washington, DC: U.S. Patent and Trademark Office.

Mamoshina, P., Vieira, A., Putin, E., & Zhavoronkov, A. (2016). Applications of deep learning in biomedicine. Molecular pharmaceutics, 13(5), 1445-1454.

Mikolov, T., Chen, K., Corrado, G. & Dean, J. (2013). Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781.

Mirsamadi, S., Barsoum, E., & Zhang, C. (2017). Automatic speech emotion recognition using recurrent neural networks with local attention. In 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 2227-2231). IEEE.

Najafabadi, M. M., Villanustre, F., Khoshgoftaar, T. M., Seliya, N., Wald, R., Muharemagic, E. (2015). Deep learning applications and challenges in big data analytics. Journal of Big Data, 2(1), 1–21. Https://doi.org/10.1186/s40537-014-0007-7

Özbey, C., & Bayar, S. (2017). Otomatik Ses Tanıma: Türkçe için Genel Dağarcıklı Akustik Model Oluşturulması ve Test Edilmesi. *Akademik Bilişim*, 8-10.

Özkan, K., Seke, E., & Işık, Ş. (2016, May). A new approach for speech denoising. In *2016 24th Signal Processing and Communication Application Conference (SIU)* (pp. 2109-2112). IEEE.

Özseven, T. (2019). Konuşma Tabanlı Duygu Tanımada Ön İşleme ve Öznitelik Seçim Yöntemlerinin Etkisi. *DÜMF Mühendislik Dergisi*, *10*(1), 99-112.

Parthasarathi, S. H. K., Hoffmeister, B., King, B., & Maas, R. (2017). *U.S. Patent Application No. 15/196,228*.

Pecorari, John. "Methods for training a speech recognition system." U.S. Patent Application No. 14/619,093.

Pennington, J., Socher, R. & Manning, C. (2014). Glove: Global vectors for word representation. In Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP), 1532-1543.

Pinola, M. (2017). History of voice recognition: from Audrey to Siri.

Pouyanfar, S., Sadiq, S., Yan, Y., Tian, H., Tao, Y., Reyes, M. P., ... & Iyengar, S. S. (2019). A survey on deep learning: Algorithms, techniques, and applications. ACM Computing Surveys (CSUR), 51(5), 92.

Sainath, T. N., Mohamed, A. R., Kingsbury, B., & Ramabhadran, B. (2013, May). Deep convolutional neural networks for LVCSR. In 2013 IEEE international conference on acoustics, speech and signal processing (pp. 8614-8618). IEEE.

Shostak, R. E. (2015). *U.S. Patent No. 8,977,548*. Washington, DC: U.S. Patent and Trademark Office.

Simonyan, K. & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.

Strigl, D., Kofler, K. & Podlipnig, S. (2010). Performance and scalability of GPU-based convolutional neural networks. In 2010 18th Euromicro Conference on Parallel, Distributed and Network-based Processing, 317-324.

Suzen. A.L, Kayaalp.K, (2018), Derin Öğrenme ve Türkiye'deki Uygulamaları, IKSAD INTERNATIONAL PUBLISHING HOUSE ISBN: 978-605-7510-53-2

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V. & Rabinovich, A. (2015). Going deeper with convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition, 1-9.

Techopedia, Voice Recognition. (2012, November 20). Techopedia. https://www.techopedia.com/definition/9961/voice-recognition

Uetz, R. & Behnke, S. (2009). Large-scale object recognition with CUDA-accelerated hierarchical neural networks. In 2009 IEEE international conference on intelligent computing and intelligent systems, 1, 536-541.

Vo, A. T., Tran, H. S., & Le, T. H. (2017, October). Advertisement image classification using convolutional neural network. In *2017 9th International Conference on Knowledge and Systems Engineering (KSE)* (pp. 197-202). IEEE.

Wang, Z. (2015). The applications of deep learning on traffic identification. BlackHat USA, 24.

Wilcox, L. D., & Bush, M. A. (2018). 15.4 Speech Recognition. *The Electrical Engineering Handbook-Six Volume Set*.

Xiao, T., Xu, Y., Yang, K., Zhang, J., Peng, Y. & Zhang, Z. (2015). The application of two-level attention models in deep convolutional neural network for fine-grained image classification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 842-850.

Xingjian, S. H. I., Chen, Z., Wang, H., Yeung, D. Y., Wong, W. K., & Woo, W. C. (2015). Convolutional LSTM network: A machine learning approach for precipitation nowcasting. In *Advances in neural information processing systems* (pp. 802-810).

Xu, Y., Zeng, X., & Zhong, S. (2013). A new supervised learning algorithm for spiking neurons. *Neural computation*, *25*(6), 1472-1511.

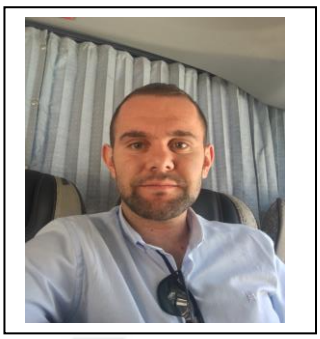
Yakar, Ö., & Aşlıyan, R. (2016). Saklı Markov Modeli Kullanarak Türkçe Konuşma Tanıma. *Akademik Bilişim*, *16*, 1-7.

Yayla, A. (2018). Konuşma tanıma teknolojisi kullanılarak devre tasarım ve analizi.

Yu, D., & Deng, L. (2016). *AUTOMATIC SPEECH RECOGNITION*. Springer london limited.

Yurtcan, Y., & Kılıç, B. G. (2018, May). Speech recognition on mobile devices in noisy environments. In *2018 26th Signal Processing and Communications Applications Conference (SIU)* (pp. 1-4). IEEE.

ÖZGEÇMİŞ



Ad-Soyad **: Harun KUTUCU**

Doğum Tarihi ve Yeri **: 1994 - Muğla**

E-posta : **harun.kutucu@gmail.com**

ÖĞRENİM DURUMU:

- **Lisans** **:** 2016, Sakarya Üniversitesi, Elektrik-Elektronik Mühendisliği
- **Yüksek lisans** **:** 2020, Sakarya Uygulamalı Bilimler Üniversitesi , Elektrik-Elektronik Mühendisliği ABD, Elektrik-Elektronik Mühendisliği

MESLEKİ DENEYİM VE ÖDÜLLER:

- 2016 yılında OTOKAR firmasında Ar-Ge departmanın 5 ay süreyle staj yaptı.
- 2019 yılında Optimum STU firmasında 8 ay boyunca Ar-Ge Mühendisi olarak görev yaptı.

YÜKSEK LİSANS TEZİNDEN TÜRETİLEN YAYINLAR, SUNUMLAR VE PATENTLER:

Kutucu H., Ferikoğlu A., 2020. Speech Recognition Using Deep Learning Algorithm, Sakarya University Journal of Computer and Information Sciences