

T.C.  
VAN YÜZÜNCÜ YIL ÜNİVERSİTESİ  
FEN BİLİMLERİ ENSTİTÜSÜ  
ELEKTRİK ELEKTRONİK MÜHENDİSLİĞİ ANABİLİM DALI

**SPEKTROGRAM TEKNİĞİ KULLANILARAK DERİN ÖĞRENME  
YÖNTEMLERİ İLE SES TANIMA**

YÜKSEK LİSANS TEZİ

HAZIRLAYAN: Ferdi KOÇ  
DANIŞMAN: Doç. Dr. Rıdvan SARAÇOĞLU

VAN-2019



T.C.  
VAN YÜZÜNCÜ YIL ÜNİVERSİTESİ  
FEN BİLİMLERİ ENSTİTÜSÜ  
ELEKTRİK ELEKTRONİK MÜHENDİSLİĞİ ANABİLİM DALI

**SPEKTROGRAM TEKNİĞİ KULLANILARAK DERİN ÖĞRENME  
YÖNTEMLERİ İLE SES TANIMA**

YÜKSEK LİSANS TEZİ

HAZIRLAYAN: Ferdi KOÇ

VAN-2019



## KABUL VE ONAY SAYFASI

### KABUL VE ONAY SAYFASI

Elektrik Elektronik Mühendisliği Anabilim Dalı'nda Doç. Dr. Rıdvan SARAÇOĞLU danışmanlığında, Ferdi KOÇ tarafından sunulan "Spektrogram Tekniği Kullanılarak Derin Öğrenme Yöntemleri ile Ses Tanıma" isimli bu çalışma Lisansüstü Eğitim ve Öğretim Yönetmeliği'nin ilgili hükümleri gereğince 05/07/2019 tarihinde aşağıdaki jüri tarafından oy birliği ile başarılı bulunmuş ve Yüksek Lisans tezi olarak kabul edilmiştir.

Başkan: Prof. Dr. M. Emin MERAL

İmza:

Üye: Doç. Dr. Rıdvan SARAÇOĞLU

İmza:

Üye: Dr. Öğr. Üyesi İlker Ali ÖZKAN

İmza:

Fen Bilimleri Enstitüsü Yönetim Kurulu'nun 26/07/2019 tarih ve 2019/40-I sayılı kararı ile onaylanmıştır.





## TEZ BİLDİRİMİ

Tez içindeki bütün bilgilerin etik davranış ve akademik kurallar çerçevesinde elde edilerek sunulduğunu, ayrıca tez yazım kurallarına uygun olarak hazırlanan bu çalışmada bana ait olmayan her türlü ifade ve bilginin kaynağına eksiksiz atıf yapıldığını bildiririm.

(İmza)

(Ferdî KOÇ)







## ÖZET

### SPEKTROGRAM TEKNİĞİ KULLANILARAK DERİN ÖĞRENME YÖNTEMLERİ İLE SES TANIMA

KOÇ, Ferdi

Yüksek Lisans Tezi, Elektrik-Elektronik Mühendisliği Anabilim Dalı

Tez Danışmanı: Doç. Dr. Rıdvan SARAÇOĞLU

Temmuz 2019, 83 sayfa

Ses tanıma insan sesinin bir mikrofon vasıtasıyla bilgisayar tarafından algılanarak tanınması işlemidir. Gelişen teknoloji sürecinde kendine önemli bir yer oluşturmaya başlamıştır. Bu çalışmada, daha önceden kaydedilmiş konuşma komutlarının tanınması için derin öğrenme yöntemleri kullanılmıştır. İki boyutlu derin öğrenme yapılarında işlenmek üzere ses verileri spektrogram yöntemi kullanılarak görüntüye çevrilmiştir. Spektrogram yöntemi hızlı fourier dönüşümü temelli alt örnekleme tekniğidir. Beş farklı evrişimsel sinir ağı modeli oluşturularak spektrogram tekniği ile elde edilen görüntülerin sınıflandırılması gerçekleştirilmiştir. Karşılaştırmalı analiz kapsamında veri seti ayrı kategoriler halinde farklı sınıf sayısına bölünmüştür. Sınıf sayısının derin öğrenme yapıları üzerindeki etkisi araştırılmıştır. Ayrıca, farklı sayıda eğitim verileri kullanılarak ağın duyarlılığı değerlendirilmiştir. Ağların performansını değerlendirmek için Softmax ve Destek Vektör Makinesi olmak üzere iki sınıflandırıcı kullanılmıştır. Elde edilen sonuçlar dikkate alındığında % 95'in üzerinde sınıflama doğruluğuna ulaşıldığı görülmüştür.

**Anahtar kelimeler:** Derin öğrenme yöntemleri, Destek vektör makineleri, Evrişimsel sinir ağları, Sesli komut sınıflandırma, Softmax,



## ABSTRACT

### SOUND RECOGNITION WITH DEEP LEARNING METHODS USING SPECTROGRAM TECHNIQUE

KOÇ, Ferdi

M. Sc.Thesis, Electrical-Electronics Engineering

Supervisor: Doç. Dr. Rıdvan SARAÇOĞLU

July 2019, 83 pages

Voice recognition is the process by which the human voice is detected and recognized by the computer through a microphone. It has started to create an important place for itself in the developing technology process. In this study, deep learning methods were used to recognize previously recorded speech commands. In order to be processed in two-dimensional deep learning structures, audio data were converted to image by using spectrogram method. Spectrogram method is a sub-sampling technique based on fast fourier transform. Five different convolutional neural network models were created by using the spectrogram technique and the trainings were performed with different data numbers. The data set is divided into different number of classes as separate categories. The effect of the number of classes on deep learning structures was investigated. Furthermore, the sensitivity of the network with different training data numbers was evaluated. In order to evaluate the performance of the networks, two classifiers were used as Softmax and Support Vector Machine. Over 95% accuracy was achieved.

**Keywords:** Deep learning methods, Support vector machine, Convolutional neural network, Voice command classification, Softmax,



## ÖN SÖZ

Bu yüksek lisans tez çalışmasının hazırlanmasında, her türlü ilgi ve yardımlarını esirgemeyen beni her konuda yönlendiren destek veren çok değerli hocam ve danışmanım Sayın Doç. Dr. Rıdvan SARAÇOĞLU ile yardımlarını esirgemeyen arkadaşlarıma sonsuz teşekkürlerimi sunarım.

2019  
Ferdî KOÇ



# İÇİNDEKİLER

	<b>Sayfa</b>
ÖZET .....	i
ABSTRACT .....	iii
ÖN SÖZ.....	v
İÇİNDEKİLER.....	vii
ÇİZELGELER LİSTESİ .....	ix
ŞEKİLLER LİSTESİ.....	xi
SİMGELER VE KISALTMALAR .....	xiii
1. GİRİŞ.....	1
2. KAYNAK BİLDİRİŞLERİ .....	3
3. MATERYAL VE YÖNTEM.....	7
3.1. Yapay Sinir Ağları .....	7
3.1.1. Yapay nöron .....	7
3.1.2. Yapay sinir ağları modelleri .....	10
3.1.3. İleri beslemeli yapay sinir ağı .....	11
3.1.4. Geri yayılım.....	12
3.1.5. Tekrarlayan yapay sinir ağları .....	15
3.1.5.1. Hopfield yapay sinir ağları .....	16
3.1.5.2. Elman and Jordan yapay sinir ağları.....	16
3.1.5.3. Uzun kısa süreli bellek .....	17
3.1.5.4. Çift yönlü yapay sinir ağları .....	18
3.1.5.5. Stokastik yapay sinir ağı.....	19
3.2. Konuşma ve Ses Sinyalinin Özellikleri .....	20
3.3. Konuşma ve Sesin Oluşması.....	21
3.4. Öğrenme.....	23
3.4.1. Denetimli öğrenme .....	23
3.4.2. Denetimsiz öğrenme .....	23
3.4.3. Veri çeşitleri ve veri hazırlama yöntemleri .....	24
3.4.4. Veri setlerinin ayrımının gerçekleştirilmesi .....	24
3.4.5. Boyutsallık laneti.....	25
3.4.6. Genelleme ve aşırı uyum .....	25
3.4.7. Boyut indirgeme .....	26

3.5. Derin Öğrenme .....	26
3.5.1. Konvolüsyon .....	28
3.5.2. Evrişimsel sinir ağları ve derin evrişimsel sinir ağları .....	29
3.5.3. Toplu normalizasyon .....	31
3.5.4. Aktivasyon fonksiyonları .....	31
3.5.4.1. Sigmoid.....	32
3.5.4.2. Rectified linear unit (ReLU).....	32
3.5.5. Havuzlama.....	33
3.5.6. Dropout katmanı.....	34
3.5.7. Maliyet fonksiyonu.....	34
3.6. Sınıflandırıcılar .....	35
3.6.1. Softmax .....	35
3.6.2. Destek vektör makineleri.....	36
3.7. Spektogram Analizi .....	40
3.8. Veri Seti .....	43
4. BULGULAR .....	45
4.1. Önerilen Yöntem.....	45
4.2. Deneysel Sonuçlar .....	48
5. TARTIŞMA VE SONUÇ.....	55
KAYNAKLAR.....	57
ÖZ GEÇMİŞ.....	61



## ÇİZELGELER LİSTESİ

Çizelge	Sayfa
Çizelge 3.1. Konuşma verisi sınıfları .....	44
Çizelge 4.1. Önerilen yöntem kapsamındaki ESA yapılarının özellikleri.....	45
Çizelge 4.2. Karşılaştırmalı analiz tablosu. ....	52
Çizelge 4.3. Softmax sınıflandırıcısı için 8 katlı çapraz doğrulama sonuçları. ....	53
Çizelge 4.4. Karşılaştırmalı analiz.....	54





## ŞEKİLLER LİSTESİ

Şekil	Sayfa
Şekil 3.1. McCulloch ve Pitts modeli.....	8
Şekil 3.2. a). Çift Kutuplu Sürekli Fonksiyon, b). Tek Kutuplu Sürekli Fonksiyon (Hatipoğlu. 2016). .....	9
Şekil 3.3. Yapay sinir ağı modeli. ....	10
Şekil 3.4. İleri besleme yapay sinir ağı.....	11
Şekil 3.5. Geri yayılım algoritma diyagramı. ....	13
Şekil 3.6. Gizli birimler ve çıkış birimleri arasındaki duyarlılık ilişkisi.....	15
Şekil 3.7. Tamamen tekrarlayan yapay sinir ağı. ....	16
Şekil 3.8. Elman yapay sinir ağı.....	17
Şekil 3.9. Basit uzun kısa süreli bellek yapay sinir ağı. ....	18
Şekil 3.10. İki yönlü yapay sinir ağları.....	19
Şekil 3.11. İnsan konuşma ve ses oluşum sistemi (Webders Sitesi). ....	22
Şekil 3.12. Konvolüsyon işlemi.....	29
Şekil 3.13. Seyreklik bağlantısı. ....	30
Şekil 3.14. Parametre paylaşımı. ....	31
Şekil 3.15. Sigmoid fonksiyonu. ....	32
Şekil 3.16. ReLU fonksiyonu. ....	33
Şekil 3.17. Maksimum havuzlama. ....	34
Şekil 3.18. Nesnelerin karar düzlemi ile ayrımı. ....	36
Şekil 3.19. Nesnelerin doğrusal olmayan düzlemler ile ayrımı.....	36
Şekil 3.20. Eşleme veya dönüşüm sonucu.....	37
Şekil 3.21. Hiper düzlemler ile elde edilen mesafe .....	38

<b>Şekil</b>	<b>Sayfa</b>
Şekil 3.22 İkili sınıf ayrımı düzlemleri.....	39
Şekil 3.13. Konuşma verisi örnekleri .....	43
Şekil 4.1 Önerilen yöntem.....	48
Şekil 4.2. ESA 1 Modelinin eğitim ve test grafiği.....	49
Şekil 4.3. ESA 2 Modelinin eğitim ve test grafiği.....	49
Şekil 4.4. ESA 3 Modelinin eğitim ve test grafiği.....	50
Şekil 4.5. ESA 4 Modelinin eğitim ve test grafiği.....	50
Şekil 4.6. ESA 5 Modelinin eğitim ve test grafiği.....	51

## SİMGELER VE KISALTMALAR

Bu çalışmada kullanılmış bazı simgeler ve kısaltmalar, açıklamaları ile birlikte aşağıda sunulmuştur.

Simgeler	Açıklama
$b_i$	Bias değeri
$c$	Sınıf Vektör Uzunluğu
$f$	Tahmin Etiket Değeri
$f(ağ)$	Ağ fonksiyonu
$f(nt)$	$f(t)$ Fonksiyonunun Eşit Aralık Numuneleri
$f(t)$	Analog Fonksiyon
$ F(k) $	$k$ . Spektral Nokta Büyüklüğü
$F_i$	Giriş Katmanı Aktivasyon Fonksiyonu
$k_h$	Filtre Yüksekliği
$k_w$	Filtre Ağırlığı
$o$	YSA çıkış değeri
$q_i$	Gizli Katman Ağırlığı
$r_i$	Çıkış Katman Ağırlığı
$t$	Gerçek Etiket Değeri
$w_i$	Giriş Katman Ağırlığı
$w(nt)$	Zaman Penceresi
$y$	Etiket Çıktısı
$x$	Girdi Matrisi
$\lambda$	Ters Sıcaklık Sabiti
$\eta$	Öğrenme Oranı

**Kısaltmalar****Açıklama**

<b>ADD</b>	Ayrık Dalgacık Dönüşümü
<b>AFD</b>	Ayrık Fourier Dönüşümü
<b>ASR</b>	Automatic Speech Recognition
<b>DESA</b>	Derin Evrişimsel Sinir Ağı
<b>DPCM</b>	Differential Pulse Code Modulation
<b>DSK</b>	Digital Signal Kit
<b>DVM</b>	Destek Vektör Makinesi
<b>ESA</b>	Evrişimsel Sinir Ağı
<b>GİB</b>	Grafik İşlem Birimi
<b>SMM</b>	Saklı Markov Modeli
<b>HFD</b>	Hızlı Fourier Dönüşümü
<b>IBM</b>	International Business Machines
<b>MFCC</b>	Mel-Frequency Cepstrum Coefficient
<b>ReLU</b>	Rectified Linear Unit
<b>RMSProp</b>	Root Mean Square Error Propability
<b>SGD</b>	Stochastic Gradient Descent
<b>SUR</b>	Speech Understanding Research
<b>YSA</b>	Yapay sinir Ağı

## 1. GİRİŞ

Konuşma tanıma ve sentezleme alanında yapılan çalışmalar kronolojik olarak incelendiğinde, AT & T Bell Laboratuvarları tarafından 1936 da üretilen ve Voder olarak adlandırılan en eski eserin ilk elektronik konuşma sentezleyicisi olduğu görülmektedir. Konuşma sinyali genellikle sesli ve sesli olmayan olarak iki gruba ayrılır. Sesli sesler periyodik bir yapıya sahiptir, sessiz sesler gürültüye benzerdir. 1970'lerin başında konuşma tanıma, Princeton Üniversitesi'nde Saklı Markov Modelleme (SMM) yaklaşımı, Lenny Baum tarafından keşfedilmiş ve IBM gibi birçok ARPA (Advanced Research Projects Agency) üyesi tarafından desteklenmiştir. DARPA (Defense Advanced Research Projects Agency) tarafından 1971 yılında, sürekli konuşmayı anlayabilen bir bilgisayar sistemi geliştirmek için SUR (Speech Understanding Research) kurulmuştur. 1978 yılında Texas Instruments tarafından popüler bir oyuncak "Speak and Spell" geliştirilmiş. "Speak and Spell" için tasarlanmış konuşma çipi bu alanda büyük adımlar atmıştır. 1984 yılında Speech Works adlı şirket, telefon üzerinden otomatik bir konuşma tanıma sistemi üretti. 1995 yılında ilk defa kelime tabanlı bir dikte yazılımı Dragon Systems tarafından piyasaya sürülmüştür. Daha sonra, IBM ve Kurzwei tarafından benzer bir yazılım yayınlanmıştır. Voice Broker adlı bir konuşma tanıma sistemi 1996 yılında Nuance ve Charles Schwab tarafından geliştirilmiştir (Ocal, 2005).

Son 15-20 yıl, farklı dillerde ve dile özgü formatta çeşitli konuşma tanıma stratejileri uygulamaya çalışılmıştır. Konuşma tanıma aynı şekilde farklı disiplinlerle de ilgilidir. Örneğin işaret işleme, örüntü tanıma, yapay zeka, olasılık teorisi, istatistik, bilgisayar bilimi, psikoloji, dil bilimi, biyoloji, ses bilimi. İlk tanıma adımı olarak, bu alanda karşılaşılan sorunların üstesinden gelmek için ön işleme algoritmaları kullanılmıştır (Karacı, 2006).

Konuşma tanıma gelişen teknoloji sürecinde önemli bir yer edinmeye çalışan bir sistemdir ve insan sesini bilgisayar ve bir mikrofon aracılığıyla tanıma sürecidir. Sayısallaştırılmış konuşmanın dönüştürülmesi ile daha fazla analiz, pencereleme ve filtreleme mümkün olmuştur. İnsan-bilgisayar iletişimindeki bu süreç önemli bir gereklilik haline gelmektedir. Çünkü günümüzde insanlar klavyeyi kullanmadan bir şey yazdırmak veya bilgisayara bir şey yaptırmak istemektedir. Konuşma tanıma disiplini, çok farklı alt disiplinleri içerir. Konuşmacı tanıma, konuşmacıdan bağımsız konuşma

tanıma sistemleri, konuşmacıya bağlı konuşma tanıma sistemleri, sürekli konuşma tanıma ve yalıtılmış sözcük tanıma sistemleri ve sözcük temelli konuşma tanıma adı altında farklı başlıklarda toparlanabilir (Yalçın, 2008).

### 1.1. Tezin Amacı ve Önemi

Birçok araştırmacı tarafından yoğun bir biçimde çalışılmakta olan derin öğrenme, klasik makine öğrenmesinden daha etkili bir öğrenme yöntemidir. Makine öğrenmesinde ham verilerden birçok özellik belirlenmesi ve belirlenen bu özelliklerin öğrenme algoritması tarafından analiz edilmesi gerekmektedir. Öğrenme algoritmalarının başarısıysa çoğunlukla çıkartılan özelliklerin kalitesine ve uygunluğuna bağlıdır yani yanlış özelliklerin kullanılması algoritmaların başarısını düşürür. Derin öğrenme ise özellikleri verilerden otomatik olarak kendisi öğrendiği için bu problemin üstesinden gelebilmektedir.

Özellik çıkarma işlemi oldukça karmaşık ve uzmanlık gerektiren bir süreçtir. Çıkarılan özelliklerin her algoritma için farklı olması da gerekebilir. Derin öğrenme tekniği içerisinde güçlü bir özellik çıkarma süreci barındırmaktadır. Bu sebeple ses verilerinin otomatik olarak sınıflandırılması için bu tez kapsamında derin öğrenme yapısı oluşturulması tercih edilmiştir. Veri setindeki sesler hem eğitim hem de test için ayrılarak oluşturulan ağın performansı grafik halinde görselleştirilecektir. Ayrıca farklı eğitim veri sayıları ile tasarlanan derin öğrenme ağlarının eğitim hassasiyeti değerlendirilecektir. Tez kapsamında önerilen derin öğrenme yapıları ile daha yüksek sınıflama doğruluk oranları elde edilmesi hedeflenmektedir. Oluşturulacak evrimsel sinir ağları (ESA) yapılarının hem eğitim hem de test verilerindeki sınıflama başarıları karşılaştırmalı bir şekilde analiz edilmiştir. Bu karşılaştırmalı analizde eğitim için tasarlanan ağların eğitim için harcadıkları süreler de dikkate alınmıştır. Aynı zamanda sınıflama aşamasında kullanılan sınıflandırıcılar değiştirilmiştir. Tez kapsamında sınıflandırıcı olarak Softmax ve Destek Vektör Makineleri (DVM) kullanılmıştır. Karşılaştırmalı analizde sınıflandırıcıların ağ performansına etkisi de incelenmiştir. Eğitimi tamamlanmış ESA yapılarının test işleminde oldukça hızlı bir şekilde sınıflama yapabilmesi ön görülmektedir.



## 2. KAYNAK BİLDİRİŞLERİ

Ses tanıma ve konuşma tanıma konularında literatürde birçok çalışma bulunmaktadır.

Karakaş (2010), tarafından ses verilerinden Mel frekans Kepstrum Katsayıları elde edilerek dinamik zaman bükme algoritmaları kullanılmıştır. Dört adet sınıf etiketi bulunan veriler konuşmacıya bağımlı olarak %88.5, bağımsız olarak ise %82 başarı oranı ile sınıflandırılmıştır.

Muda ve ark. (2010), beş adet sınıf bulunan ses verilerinde Mel Frekans Kepstrum Katsayıları ve dinamik zaman bükme algoritmalarını uygulamıştır. Ayrıca bu algoritmalar bayan ve erkek kullanıcı ses kayıtları üzerinde çalıştırılmıştır.

Fezari ve ark. (2016), donanımsal olarak ses tanıma gerçekleştirmişlerdir. Bu işlem için RSC 364 kartı ve PIC 16F876 kullanmışlardır. Laboratuvar içinde ve dışında olmak üzere farklı denemeler gerçekleştirilmiştir. Laboratuvarda gürültüsüz ortamda %85 başarı oranı elde edilmiş, gürültülü ortamda ise bu oran %73'e düşmüştür. Laboratuvar dışındaki ortamda ise bu oran gürültüsüz ortamda %78 olup gürültülü ortamda ise %65'dir.

Baygün ve ark. (2009), simülasyon ortamında önerdikleri algoritmaları telefon görüşmeleri üzerinde denemiştir. Elde ettikleri sonuçlarda konuşmacının telaffuzuna göre farklılık gösterdiği görülmüştür.

Bala ve ark. (2010), ses tanıma algoritmaları kullanılarak bir güvenlik sistemi tasarlamışlardır. Ses verilerinde öznitelik çıkarımı ve eşleştirmesi ile %100'e yakın tanıma başarısı elde etmişlerdir.

Hrncar ve ark. (2000), dinamik zaman bükme algoritmalarını geliştirerek konuşma tanıma modeli tasarlamışlardır. Geliştirdikleri yöntemde giriş vektörü ile referans alınan vektör arasındaki mesafeyi maksimum değere göre sınıflandırma işlemi yapılmıştır. Geliştirilen model ile yüksek sınıflama başarıları elde edilmiştir.

Price ve ark. (2006), yapmış oldukları çalışmada karşılaştırmalı analiz gerçekleştirmişlerdir. Analiz kapsamında Mel Frekans Kepstrum Katsayı ile beraber dinamik zaman bükme algoritması ve Saklı Markov Modeli (SMM) kullanılmıştır. Tatmin edici sonuçlar ile algoritmaların başarıları kanıtlanmıştır.

Zhizeng ve ark. (2004), ses komutları ile hareket eden uzaktan kontrollü robot tasarlamışlardır. Algoritma olarak geliştirilmiş zaman bükme algoritması kullanılmıştır. 10 farklı sınıfın bulunduğu test verisinde %88 ortalama başarıya ulaşılmıştır.

Öztürk ve ark. (2007), ses tanıma üzerine yaptıkları çalışmada donanımsal olarak C6713 DSK kartı kullanmışlardır. Üç farklı ses olarak 'a', 'e', 'i' seslerinin tanınmasında %80 oranının üzerinde başarıya ulaşılmıştır.

Demirci (2005), gerçekleştirdiği çalışmada Mel katsayıları ile Vektör mesafe ölçme modellerine birleştirmiştir. Kullanıcı cinsiyetine göre yapılan test işlemlerinde erkeklerde %90, bayanlarda %65 başarıya ulaşmışlardır.

Phoophuangpairoj (2011), yapmış olduğu çalışmada SMM modeli le ses verilerini tanımaya çalışmışlardır. Sistemi tasarlanan robotun komut tanınması üzerine geliştirmişlerdir. Elde edilen başarı sonuçlarının cinsiyete bağımlı ve bağımsız olmak üzere ortalama %98 oranında olduğu görülmüştür.

Aşlıyan ve ark. (2008), Mel katsayıları ile dinamik zaman bükme algoritmasını kullanarak Türkçe heceleri tanıyabilecek bir sistem tasarımı yapmışlardır. Sistem 2000 adet ses verisi ile test edildiğinde % 95 tanıma başarısı elde edilmiştir.

Baygın ve ark. (2012), ses verilerinde MFCC özelliklerini elde ederek farklı sınıflandırma algoritmaları kullanmışlardır. Sınıflandırıcı olarak Öklid mesafesi, yapay sinir ağları (YSA) ve dinamik zaman bükme algoritmaları denenmiştir. Bu önerilen sınıflandırıcıların başarıları sırasıyla %78, %86 ve %95 olarak elde edilmiştir.

Dede ve ark. (2010), biyometrik tanıma uygulamaları için Mel katsayıları ve YSA'nın bir arada kullanıldığı bir sistem tasarlamışlardır. Matlab programı üzerinde geliştirilen simülasyon uygulamasında 10 farklı kişi için %99 tanıma başarısı elde edilmiştir.

Edizkan (2007), tasarlamış olduğu gezgin araçta Mel katsayıları ve ortak vektör yaklaşımı ile ses komutlarını analiz etmişlerdir. Analizler sonucunda %100 tanıma başarısı elde edilmiştir.

Bolat ve ark. (2004), yapmış oldukları çalışmada istatistiksel YSA kullanmışlardır. Konuşma ve müzik verilerinin değerlendirildiği çalışmada yaklaşık %91 tespit başarısı elde edilmiştir.

Meral (2008), doğrusal öngörülü kodlama ile adaptif bir sistem tasarlamıştır. Ses verilerinin ayrımı için Öklid mesafesi kullanılmış ve %94 tanıma başarısı elde edilmiştir.

Asyalı ve ark. (2011), tasarlamış oldukları yapay el tasarımında ses komutları ile kontrol gerçekleştirmişlerdir. Yapılan tasarım 40 farklı ses komutunu başarılı şekilde yerine getirmiştir.

Babui ve ark. (2012), karşılaştırmalı bir analiz yapmışlardır. Analiz kapsamında doğrusal öngörülü kodlama, vektör nicelemesi ve SMM algoritmaları kullanılmıştır. Eşik değeri kullanılarak ortam gürültüsünün etkisi en aza indirgenmiştir. Zamandan bağımsız olarak 0 ile 9 arasındaki rakamlar tanınmıştır.

Leechor ve ark. (2010), SMM kullanarak ses verilerini analiz edebilecek ara yüz oluşturmuşlardır. Ortam gürültüsü dikkate alınarak yapılan çalışmada gürültüsüz ortamda %98, gürültülü ortamda ise %44 başarı oranı elde edilmiştir.

Abushariah ve ark. (2010), İngilizce rakamları Mel katsayıları ve SMM modeli kullanarak tespit etmeye çalışmışlardır. Matlab programı üzerinden tasarlanan ara yüz ile gürültülü ortamdan %67, gürültüsüz ortamda ise %80 başarı oranı sağlanmıştır.

Jiang ve ark. (2009), 120 adet ses verisi üzerinden 0-9 arasındaki rakamları tanımaya çalışmışlardır. Mel katsayıları ve z dönüşümü algoritması kullanılarak tasarlanan sistemde yaklaşık %86 tanıma başarısı elde edilmiştir.

He ve ark. (2001), Matlab programı üzerinden ActiveX ve SMM algoritmasını koşturarak ses verilerini %92'lik bir oran ile tanımayı başarmışlardır.

Ferrando ve ark. (2009), yapmış oldukları çalışmada su altında hareket eden araç prototipinin ses komutları ile kontrol etmeye çalışmışlardır. Ortama göre sesin ilerleme hızı değişeceğinden dolayı analiz için Ayrık Dalgacık Dönüşümü (ADD) ve Mallat algoritmaları tercih edilmiştir.

Alshuieili ve ark. (2011), tasarlamış oldukları gerçek zamanlı sistemde Zigbee kablosuz haberleşme modülü ve DPCM ses analiz modülü kullanılmıştır. Tasarlanan sistem farklı ortamlarda test edildiğinde %80 başarı gösterdiği tespit edilmiştir.

Öztürk ve ark. (2010), ses komutlarını şablon eşleştirmesi yaparak tanımaya çalışmışlardır. PIC18F452 donanım kullanılarak yapılan test işlemlerinde %67.5 tespit oranı elde edilmiştir.

Özdemircan (2008), Java programı üzerinde Mel ve YSA yapılarını kullanarak ses ile kontrol edilebilen robot tasarımı yapmıştır. Tasarlanan sistem kullanıcıdan bağımlı ve bağımsız olarak değerlendirilmelidir. Kullanıcı bağımlı olarak yaklaşık %95, kullanıcıdan bağımsız olarak %80 başarı oranı elde edilmiştir.

McMahan ve ark. (2017), yapmış oldukları çalışmada 20 sınıflı Google konumu komutları veri setinde derin öğrenme teknikleri kullanmışlardır. Kullandıkları ESA yapılarını ön eğitilmiş ve ölçeklendirilmiş olmak üzere 4 farklı model eğitmişlerdir. Elde ettikleri başarılar %81.32-85.52 arasında değişmektedir.

Andrade ve ark. (2018), yapmış oldukları çalışmada tekrarlayan sinir ağları kullanmışlardır. Önerilen model, Google Konuşma Komutları'nda % 94,1 ve V2'de % 94,1 (20 komut tanıma görevi için) iken, sadece 202 bin eğitilebilir parametre içerir. Sonuçlar, 5 farklı görevde (20 komut tanıma (V1 ve V2), 12 komut tanıma (V1), 35 kelime tanıma (V1) ve sol-sağ (V1)) yapılan önceki çalışmalarla karşılaştırılmıştır.

Warden (2018), yapmış olduğu çalışmada Google Konuşma Komutları'ndaki V1 ve V2 veri setlerinde derin öğrenme modelini denemiştir. Oluşturduğu convnet yapısı ile V1 veri setinde %85.4 ve V2 veri setinde % 88.2 test başarıları elde etmişlerdir.

### 3. MATERYAL VE YÖNTEM

#### 3.1. Yapay Sinir Ağları

Bu başlık altında ilk olarak YSA mimarileri ve çalışma mantığı anlatılmıştır. Sonrasında ise eğitim işleminde kullanılan algoritmalar ve YSA'nın gelişmiş modellerinden bahsedilmiştir. Son olarak ise YSA'nı temel alınarak inşa edilmiş derin öğrenme yapıları sunulmuştur.

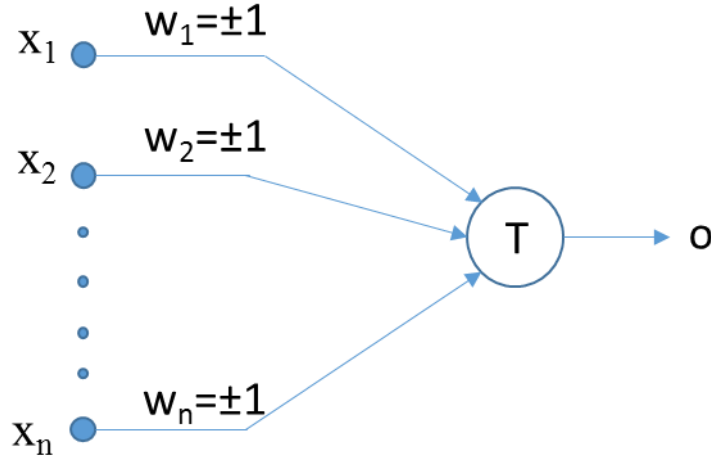
YSA özellikle insan beyninin sinir yapısını matematiksel olarak modellenmesi ile ortaya çıkmıştır. Gerçek sinir sistemi ile yapay sinir ağları tam olarak birbirine benzemese de çalışma prensipleri birbirine oldukça yakındır. YSA uygulama alanları olarak zaman serisi tahmini, sınıflandırma, regresyon analizi, veri işleme görüntü işleme, örüntü tanıma, karar verme ve kümeleme problemlerinde kullanılabilir.

##### 3.1.1. Yapay nöron

YSA'nın ana unsuru gerçek sinir sistemi ile benzerlik gösteren yapay sinirlerden oluşur. Biyolojik nöronlar dendritler sayesinde birbiriyle iletişim sağlarlar. Sonrasında bu sinyaller toplanarak hücre tarafından bir cevap üretilir. Aksonlar aracılığıyla bu cevap hücreler arası transfer edilir.

YSA'nda cevap sinyali ağırlıklar ile işleme girerek iletimi sağlanır. Sonrasında bias değerleriyle bu işlem sonucu toplanır. Ağdaki doğrusallığı azaltmak amacıyla elde edilen toplam transfer fonksiyonundan geçirilir.

YSA'nın ilk matematiksel modeli 1943'de McCulloch ve Pitts tarafından elde edilmiştir (McCulloch ve ark., 1943). Model kapsamında çıktı değerlerine eşikleme uygulanır. Eğer çıktı değeri eşik değerinin altında ise sonuç sıfırdır. Çıktı değeri eşik değerinin üzerinde bir değere sahipse değerini korur. McCulloch ve Pitts'in oluşturdukları model Şekil 3.1'de gösterilmiştir.



Şekil 3.1. McCulloch ve Pitts modeli.

McCulloch ve Pitts modelinin matematiksel ifadesi:

$$o^{k+1} = \begin{cases} 1 & \text{eğer } \sum_{i=1}^n w_i x_i^k \geq T \\ 0 & \text{eğer } \sum_{i=1}^n w_i x_i^k < T \end{cases} \quad (3.1)$$

Model içerisinde pozitif ağırlık değerleri inhibitör bağlantılarını gösterirken negatif ağırlık değerleri eksitator bağlantılarını ifade eder. Model içerisinde normal eşiklemenin yanı sıra transfer fonksiyonu kullanılır. Literatürde çıktı değeri için kullanılan matematiksel ifade de (y, o ve f (ağ)) ifadeleri bulunur.

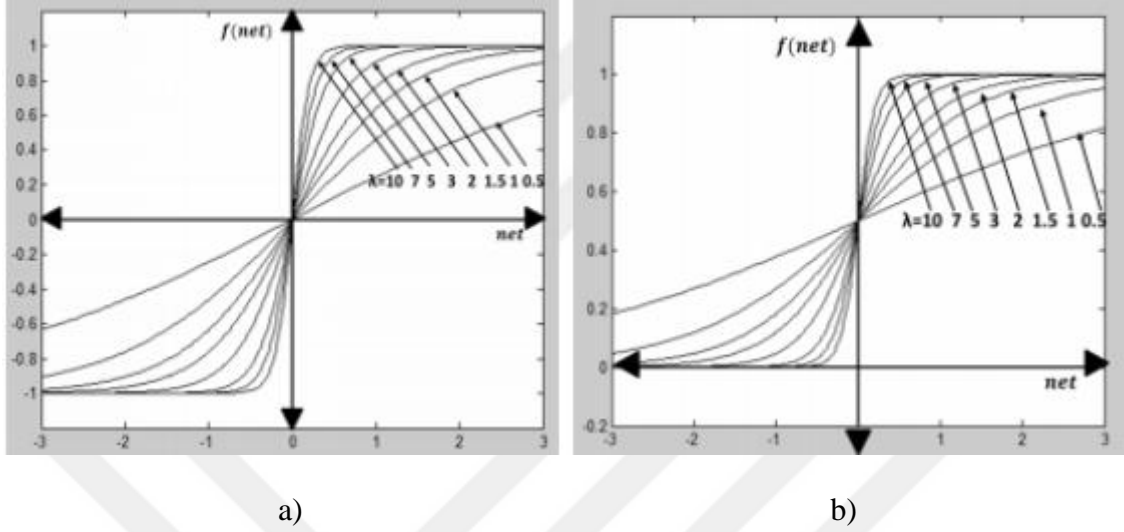
Transfer fonksiyonu ağırlıklandırılmış giriş sinyalini ve bias değerlerini işler. Model için oluşturulan matematiksel ifade Eş. 3.2'deki gibi tanımlanır.

$$y(k) = F\left(\sum_{i=1}^n w_i(k)x_i(k) + b\right) \quad (3.2)$$

Doğrusal fonksiyonların birbirleriyle kombinasyonları doğrusal özellik gösterir. Ağ içerisindeki yapay sinirlerinde doğrusal bir işlem gerçekleşiyorsa elde edilen çıktılarda da doğrusallık oluşacaktır. Bu sebeple ağ yapısındaki doğrusallığı azaltmak için aktivasyon fonksiyonları kullanılır.

Literatürde kullanılan aktivasyon fonksiyonları çift ve tek kutuplu fonksiyonlar olmak üzere ikiye ayrılır. Aktivasyon fonksiyonları arasındaki bu fark üretmiş oldukları cevabın pozitif ve negatif olmasına göre değişir. Çift kutuplu aktivasyon fonksiyonları

negatif ve pozitif cevaplar üretirken tek kutuplu aktivasyon fonksiyonları sadece pozitif cevap üretir.



Şekil 3.2. a). Çift Kutuplu sürekli fonksiyon, b). Tek kutuplu sürekli fonksiyon (Hatipoğlu. 2016).

Çift kutuplu sürekli aktivasyon fonksiyonunun matematiksel ifadesi Eş. 3.3'deki gibi gösterilmiştir.

$$f(ağ) = \frac{2}{1 + \exp(-\lambda ağ)} - 1 \quad (3.3)$$

Tek kutuplu sürekli aktivasyon fonksiyonunun matematiksel ifadesi Eş. 3.4'deki gibi gösterilmiştir.

$$f(ağ) = \frac{1}{1 + \exp(-\lambda ağ)} \quad (3.4)$$

Şekil 3.2'de gösterilen aktivasyon fonksiyonları sigmoid karakteristik gösterir. Çift kutuplu aktivasyon fonksiyonları tek kutuplu aktivasyon fonksiyonlarına göre YSA'nda kullanımı daha yaygındır. Çift kutuplu sürekli aktivasyon fonksiyonuna sahip yapay sinirler sürekli yapay sinirler olarak adlandırılır. Bu çeşit YSA ikili sınıflandırma probleminin çözümünde sıklıkla kullanılır.

Çift kutuplu ikili aktivasyon fonksiyonu Eş. 3.5'te gösterilmiştir.

$$f(ağ) = \text{sgn}(ağ) = \begin{cases} 1 & \text{eğer } ağ > 0 \\ -1 & \text{eğer } ağ < 0 \end{cases} \quad (3.5)$$

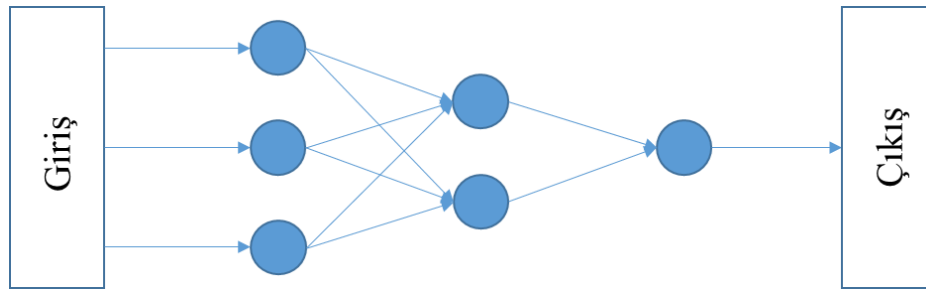
Tek kutuplu aktivasyon fonksiyonu ise

$$f(ağ) = \text{sgn}(ağ) = \begin{cases} 1 & \text{eğer } ağ > 0 \\ 0 & \text{eğer } ağ < 0 \end{cases} \quad (3.6)$$

### 3.1.2. Yapay sinir ağları modelleri

Gerçek yaşamda karşılaşılan problemler tek yapay sinir bulunan modellerle çözülemez. Bu problemlerin çözümü fazla sayıda yapay sinirlerin birleştirilmesi ile gerçekleştirilir. Yapay sinir ara bağlantıları sinir ağı olarak tanımlanır. Sinir ağının giriş ve çıkışları ağırlıklar yoluyla hesaplanır. Gerekli durumlarda gecikme bloğu ara katmanlar arasına yerleştirilir.

YSA'nda bağlantılar rastgele şekilde yapılmaz. Bu bağlantılar bazı topolojiler kullanılarak gerçekleştirilir. Problemlerin çözümü bu topolojiler sayesinde verimli ve kolay şekilde yapılır. Topoloji yapıları ileri veya geri beslemeli olarak iki sınıfa ayrılır. İleri beslemeli yapılarda bilgi girişten çıkışa doğru akar. Bu nedenle ileri beslemeli yapılar tersinir özellik göstermez. Geri beslemeli topolojilerde bilgi hem ileri hem de geri yönde akar. Ayrıca bu yapılar döngüsel bir öğrenmeye sahiptir.



Şekil 3.3. Yapay sinir ağı modeli.

Şekil 3.3'de gösterilen ileri beslemeli YSA, katman katman ifade edilir. Girişten çıkışa doğru gidildiğinde ilk katman giriş katmanı olarak adlandırılır. Sonuncu katman çıkış katmanıdır. Giriş ve çıkış katmanları arasındaki katmanlara ara katman adı verilir. Gizli katmanlar eğitim verilerinin özelliklerinin öğrenildiği yerdir.

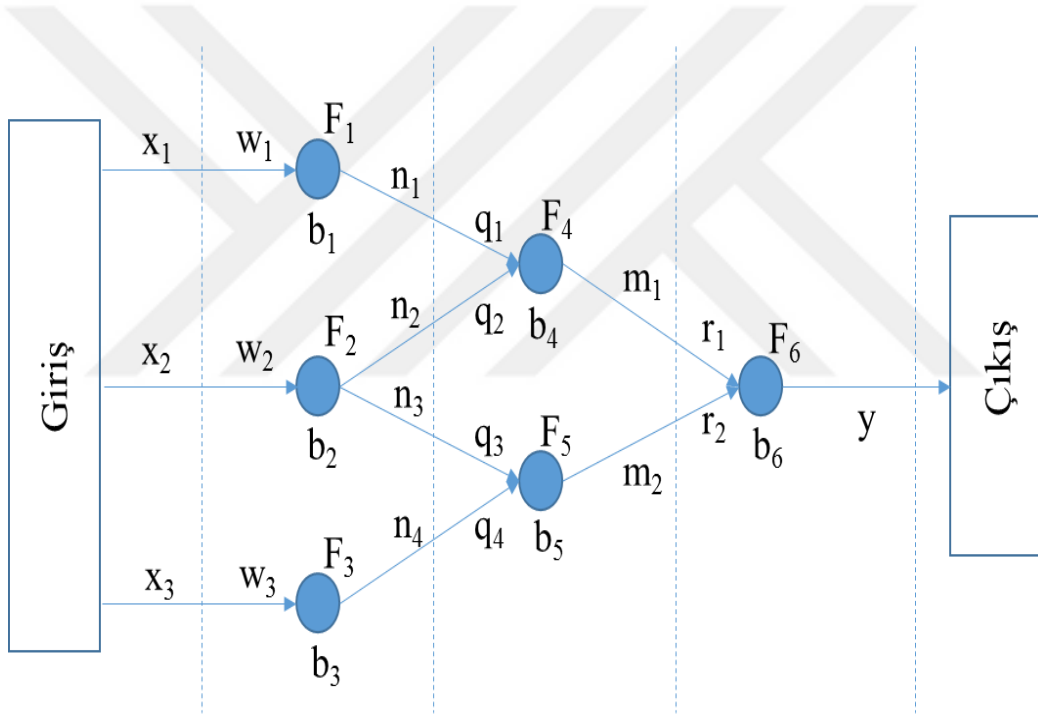


Sonlu sayıdaki tek gizli katmana sahip yapay sinir ağı yakınsama teoremine göre bir problemi çözebilir. Bu çözümü belirli hata oranına göre gerçekleştirir. Bu çözümü tek gizli katmana sahip YSA yapmak yerine çok sayıda gizli katman sayısına sahip YSA ile çözüme ulaşılabilir.

### 3.1.3. İleri beslemeli yapay sinir ağı

İleri beslemeli topolojilerinden biri olan ileri beslemeli YSA’nda bilgi girişten çıkışa doğru akar. Bu yapı oluşturulurken yapıda kullanılan aktivasyon fonksiyonu seçimi, ara katman sayısı ve yapay sinir sayısı oldukça önemlidir.

Bir gizli katmana sahip ileri beslemeli yapay sinir ağı Şekil 3.4’de gösterilmiştir.



Şekil 3.4. İleri besleme yapay sinir ağı.

İleri beslemeli YSA’nın matematiksel olarak hesapları aşağıdaki denklemlerde detaylı olarak verilmiştir. Denklemlerde aktivasyon fonksiyonları  $F_i$  olarak gösterilmektedir.  $F_1$ ,  $F_2$  ve  $F_3$  giriş katmanı aktivasyon fonksiyonlarıdır.  $w_i$ ,  $q_i$  ve  $r_i$  yapıdaki ağırlıkları temsil eden gizli katmanın aktivasyon fonksiyonu  $F_4$  ve  $F_5$ ’dir. Ağda kullanılan bias değerleri  $b_i$  olarak gösterilmiştir. Ağın etiket çıktısı ise  $y$  ile gösterilir.

$$n_1 = F_1(w_1x_1 + b_1) \quad (3.7)$$

$$n_2 = F_2(w_2x_2 + b_2) \quad (3.8)$$

$$n_3 = F_2(w_2x_2 + b_2) \quad (3.9)$$

$$n_4 = F_3(w_3x_3 + b_3) \quad (3.10)$$

$$m_1 = F_4(q_1n_1 + q_2n_2 + b_4) \quad (3.11)$$

$$m_2 = F_5(q_3n_3 + q_4n_4 + b_5) \quad (3.12)$$

$$y = F_6(r_1m_1 + r_2m_2 + b_6) \quad (3.13)$$

Yapay sinir içerisindeki değişkenleri elle hesaplamak oldukça zahmetlidir. Bu nedenle bu değişkenleri optimize etmek amacıyla bilgisayarlar kullanılır.

#### 3.1.4. Geri yayılım

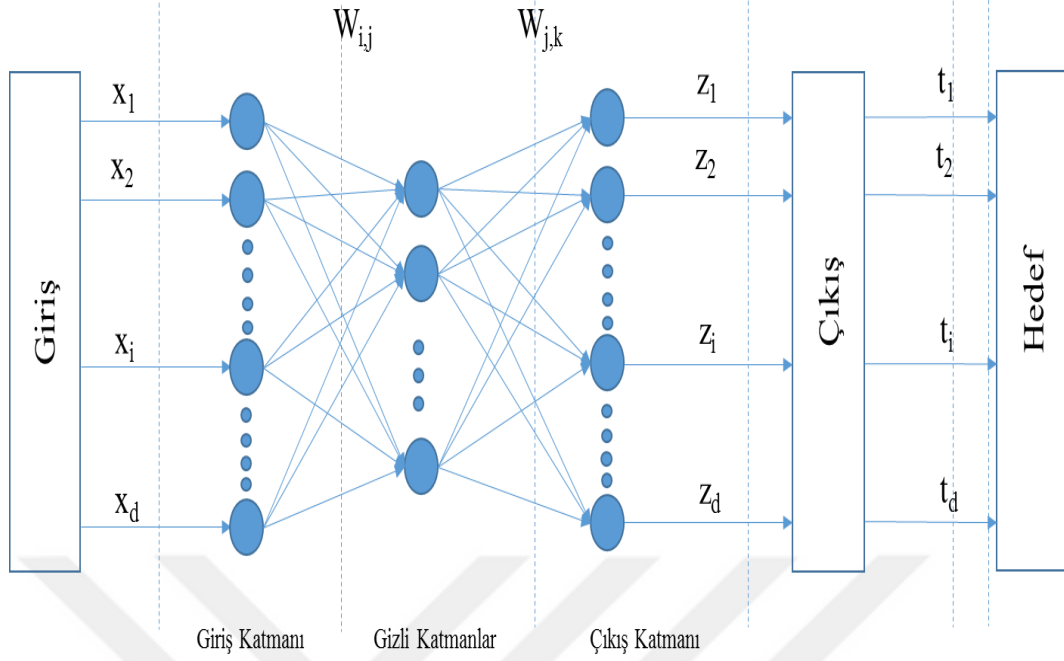
YSA içerisindeki en yaygın optimizasyon yöntemi geri yayılım algoritmasıdır. Geri yayılım algoritmalarından biri olan gradyan azalım yönteminde ortalama karesel hata kullanılır. Geri yayılım algoritması YSA değişkenlerinin optimize edilmesinde güçlü ve oldukça başarılıdır.

YSA'nın başlangıç bias ve ağırlık değerlerini ayarlamak önemli bir problemdir. Eğitim örneklerini kullanarak geri yayılım algoritmasıyla optimum değişkenler elde edilmeye çalışılır. Hedef ve tahmin edilen çıktı değerleri arasındaki farkı kullanarak ortalama karesel hatayı hesaplar (Duda ve ark., 2012).

Eğitim aşamasındaki hata formülize edilirse;

$$j(w) = \frac{1}{2} \sum_{k=1}^c (t_k - z_k)^2 \quad (3.14)$$

Eş. 3.14'te t gerçek etiket değerini ve f tahmin etiket değerini gösterir. w ağırlık değerini, c ise sınıf vektörü uzunluğunu gösterir.



Şekil 3.5. Geri yayılım algoritma diyagramı.

YSA’nda rastgele olarak atanan ağırlık değerleri gradyan azalma yöntemi kullanılarak optimize edilir. Ağırlıklardaki değişim miktarını öğrenme oranı ayarlar. Eş. 3.15’te öğrenme oranı  $\eta$  olarak ifade edilir.

$$\Delta w_{\min} = -\eta \frac{\partial J}{\partial w_{\min}} \quad (3.15)$$

Ağırlıkların bağımlılıkları dikkate alınarak girdi ve gizli katman ağırlıkları güncellenir. Hataya göre güncellemede zincir kuralı kullanılır. Tek gizli katmana sahip YSA’ndaki ağırlıkların güncellenmesi aşağıdaki gibidir. Ortalama karesel hatayı kullanılarak ağırlıkları güncellemek için kısmi türevler kullanılır.

$$\frac{\partial J}{\partial w_{kj}} = \frac{\partial J}{\partial a\check{g}_k} \frac{\partial a\check{g}_k}{\partial w_{kj}} \quad (3.16)$$

Eş. 3.16’da gizli katmanla çıktı katman arasındaki ağırlıklar  $w_{jk}$  ile gösterilir.

$$\zeta_k = -\frac{\partial J}{\partial a\check{g}_k} \quad (3.17)$$

Çıktı katmanının öncesindeki giriş  $a\check{g}_k$  ile ifade edilir. Çıktı değerleri aşağıdaki diferansiyel denklemlere göre hesaplanır.

$$-\frac{\partial J}{\partial a\check{g}_k} = -\frac{\partial J}{\partial z_k} \frac{\partial z_k}{\partial a\check{g}_k} = (t_k - z_k) f'(a\check{g}_k) \quad (3.18)$$

$$\frac{\partial a\check{g}_k}{\partial w_{kj}} = y_i \quad (3.19)$$

Eş. 3.19'daki türev işleminden sonra ağırlık güncellemesi gerçekleştirilir.

$$\Delta w_{kj} = \eta \zeta_k y_i = \eta (t_k - z_k) f'(a\check{g}_k) y_i \quad (3.20)$$

Çıkıştan girişe göre benzer türev işlemleri uygulanır. Aşağıda gösterilen denklemlere göre zincir kuralı uygulanır.

$$\frac{\partial J}{\partial w_{ji}} = \frac{\partial J}{\partial y_i} \frac{\partial y_i}{\partial a\check{g}_j} \frac{\partial a\check{g}_j}{\partial w_{ji}} \quad (3.21)$$

$$\frac{\partial J}{\partial y_i} = \frac{\partial}{\partial y_i} \left[ \frac{1}{2} \sum_{k=1}^c (t_k - z_k)^2 \right] = -\sum_{k=1}^c (t_k - z_k) f'(a\check{g}_k) w_{jk} \quad (3.22)$$

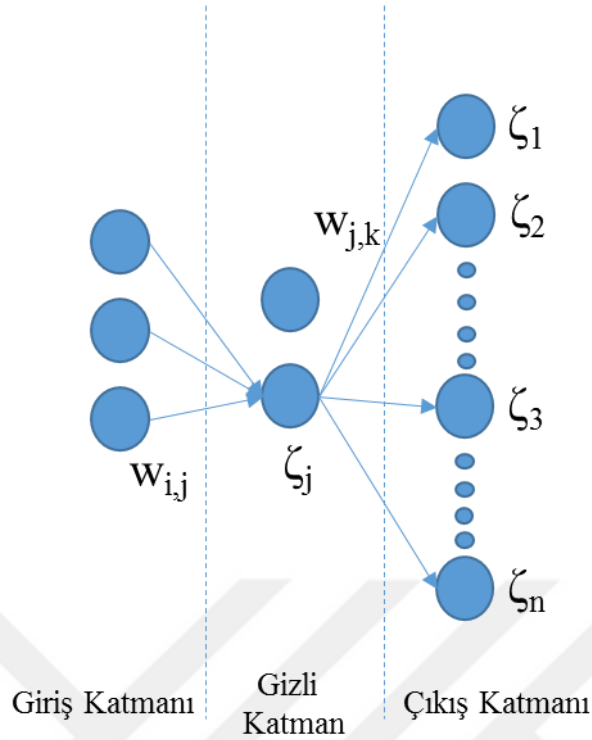
Eş. 3.23'te çıkış değerinin en küçük karesel hata değerine etkisi gösterilmiştir.

$$\zeta_j = f'(a\check{g}_j) \sum_{k=1}^c w_{kj} \zeta_k \quad (3.23)$$

Giriş ve gizli katmanı arasındaki ağırlık güncellemesi aşağıdaki gibi hesaplanır.

$$\Delta w_{ji} = \eta x_i \zeta_j = \eta x_i f'(a\check{g}_j) \sum_{k=1}^c w_{kj} \zeta_k \quad (3.24)$$

Gizli katmandaki duyarlılıklar çıktı katmandaki duyarlılıklara orantılı olarak değişir.



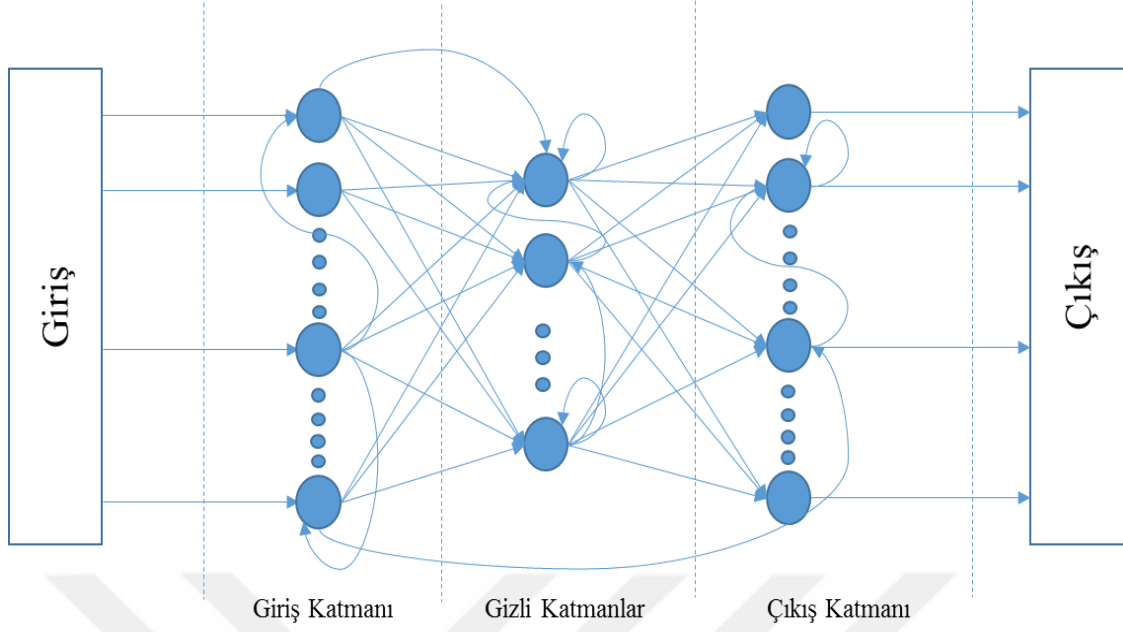
Şekil 3.6. Gizli birimler ve çıkış birimleri arasındaki duyarlılık ilişkisi.

YSA'ndaki en yaygın kullanılan optimizasyon yöntemlerinden biride stokastik eğitimidir. Bu eğitim modelinde eğitim verileri rastgele olarak seçilir. Sonrasında gradyan azalması yöntemi ile ağırlıklar güncellenmeye başlar. Toplu eğitimde bütün eğitim örnekleri kullanılır. Epok sayısı ağır eğitiminde bütün verilerin kullanılmasında geçen iterasyon sayısıdır. Eğitim otomatik olarak tamamlanması için durdurma kriteri eklenir.

### 3.1.5. Tekrarlayan yapay sinir ağları

Tekrarlayan topolojiye sahip olan YSA'nda bilgi akışı ileri veya geri yönde olabilir. Aynı zamanda bilgi aynı seviyedeki yapay sinirler arasında aktarılabilir. Tekrarlayan sinir ağlarında geriye doğru bilgi akışı sayesinde giriş verilerinin incelenmesi sağlanır.

Tekrarlayan YSA'nda bloklar içerisinde ve bloklar arasında her yönde bağlantılar bulunur. Tekrarlayan YSA çeşitleri olarak Hopfield ve Elman tekrarlayan YSA gösterilebilir (Hopfield, 1982; Elman, 1990).



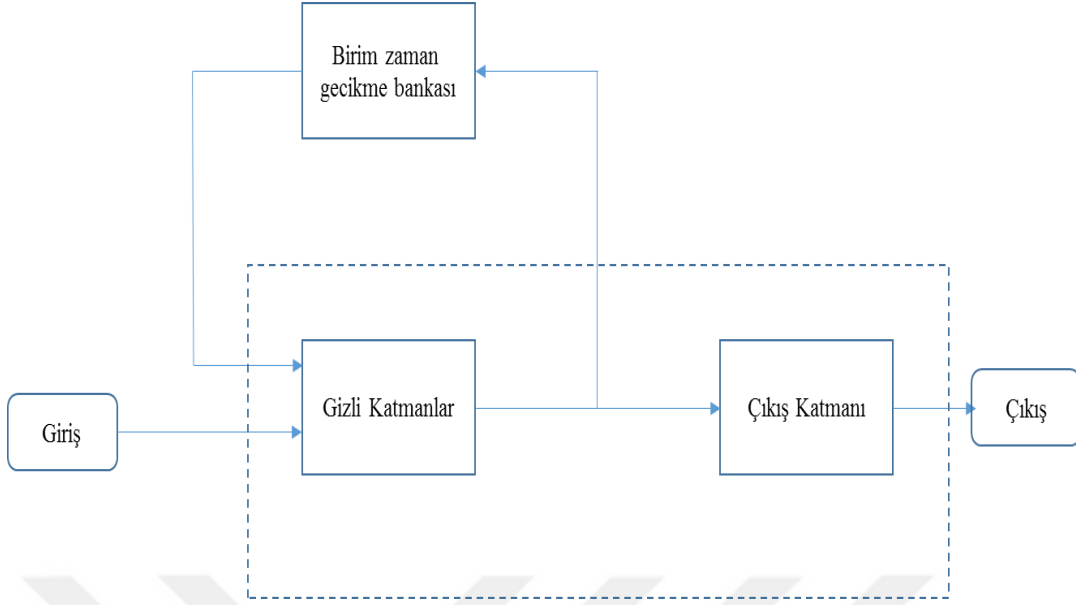
Şekil 3.7. Tamamen tekrarlayan yapay sinir ağı.

### 3.1.5.1. Hopfield yapay sinir ağları

Tekrarlayan YSA'ndan biri olan Hopfield YSA kararlı hedef bilgilerini elde etmek için kullanılır. Hopfield ağları tam bağlı ve eş zamanlı karakteristiğe sahiptir. Hopfield ağları çıkışta -1, +1 veya 0, +1 değerleri üretir. Bu ağlarda iki tip kısıtlama bulunur. İlk olarak ağ içerisindeki yapay sinirlerde kendi içerisinde döngü bulunmaması gerekir. İkincisi ise düğüm bağlantıları arasında simetriklik mevcuttur. Yapı içerisinde bulunan simetriklik ağırlıklar ve aktivasyon fonksiyonları üzerinden enerjinin giderek azalmasını sağlar. Simetrik olmayan yapılar veya periyodik davranışlara sebep olabilir (Krenker ve ark., 2011; Zurada ve ark., 1992.).

### 3.1.5.2. Elman and Jordan yapay sinir ağları

Elman YSA karar ağaçları çalışma prensibine dayanan tekrarlayan YSA modellerindedir. Gizli katman ve giriş katmanı arasında birim zaman gecikmesine sahip geri döngü bulunur. Gizli katmandaki aktivasyon fonksiyonları sigmoid ve doğrusal yapıdadır. Elman YSA verinin desen bilgisini elde etmeye yarayan bir belleğe sahiptir. Bu ağ kullanımı ile uzamsal ve geçici desenler elde edilir (Haykin ve ark., 2004).

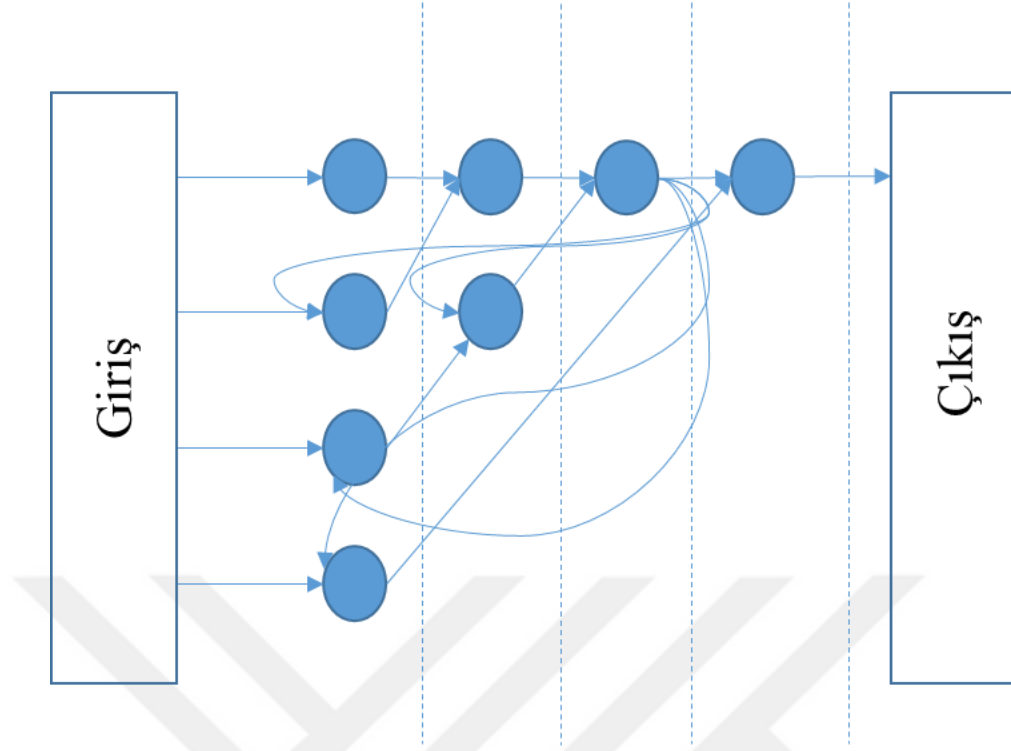


Şekil 3.8. Elman yapay sinir ağı.

Jordan sinir ağını diğer ağlardan ayıran tek fark bağlam biriminin gizli katman yerine çıktı katmanına bağlanmasıdır.

### 3.1.5.3. Uzun kısa süreli bellek

Tekrarlayan YSA topolojilerinden biri olan uzun kısa süreli bellek zaman serilerini hem sınıflandırabilmekte hem de tahmin edebilmektedir. Bu tekrarlayan YSA modeli bilgiyi hafızasında tutma özelliğine sahiptir. Kullanılan aktivasyon fonksiyonu sigmoidtir. Uzun kısa süreli bellek yapısı şekil 3.9'da gösterilmiştir.



Şekil 3.9. Basit uzun kısa süreli bellek yapay sinir ağı.

Şekil 3.10'da gösterilen mimaride giriş katmanı için önem arz eden yapay sinirler üçüncü ve dördüncü yapay sinirdir. Bu yapay sinirdeki bilginin hatırlanma süresini ve bellek verisinin aktarımını bir alt katmandaki yapay sinirler belirler (Krenker ve ark., 2011).

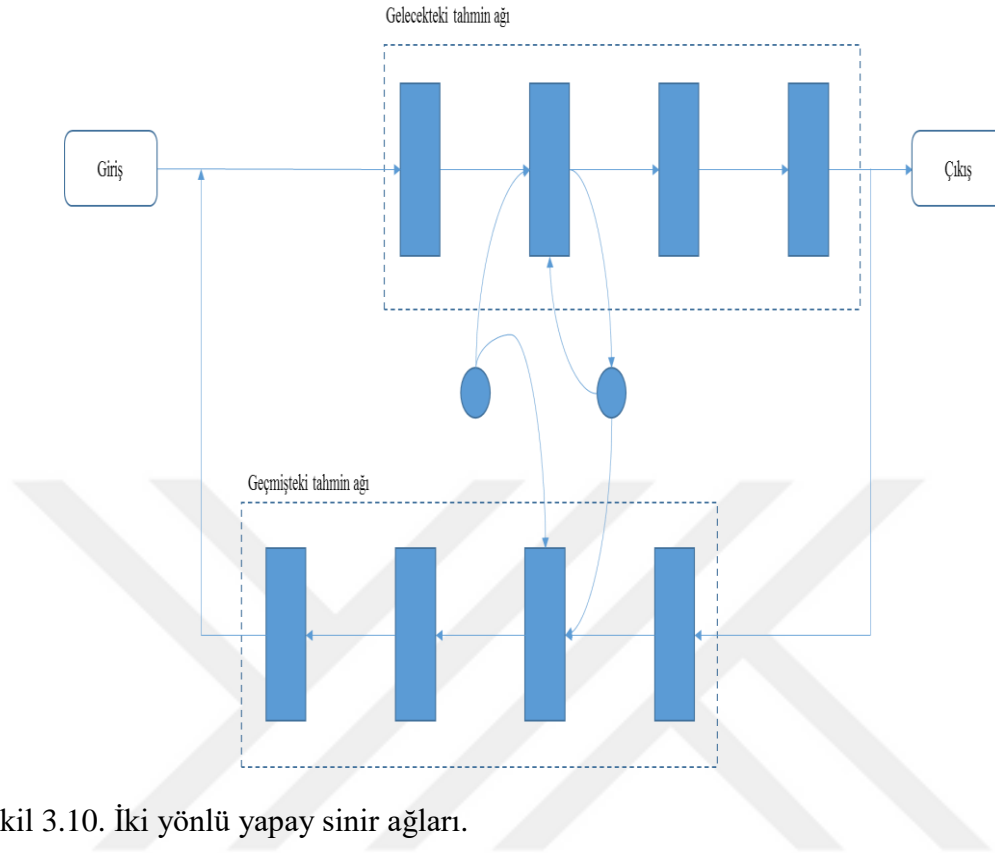
#### 3.1.5.4. Çift yönlü yapay sinir ağları

Çift yönlü YSA tekrarlayan YSA çeşitlerinden olup zaman serilerini tahmin etmek için kullanılır. Uzun kısa süreli bellek yapılarına göre bu yapılar daha iyi sonuçlar elde eder (Wakuya ve ark., 2001).

Çift yönlü YSA ters yönlü YSA olarak da adlandırılır. Çift yönlü YSA durumlarını hatırlayabilme özelliğine sahiptir. Bu yapı gelecek ve geçmiş zaman serilerini tahmin edebilmektedir. İlk aşamada gelecek zaman serilerinin tahmin edebilen bu ağlar sonrasında geçmiş zaman serilerini de tahmin edebilir.



Çift yönlü yapay sinir ağlarının yapısı Şekil 3.10'da gösterilmiştir.



Şekil 3.10. İki yönlü yapay sinir ağları.

### 3.1.5.5. Stokastik yapay sinir ağı

YSA'nın ağırlıklarına stokastik dağılım ile ağı başlangıç değerlerini atama yapar. Bunun yanı sıra stokastik aktivasyon fonksiyonu kullanılmaktadır. Bu dağılım ve aktivasyon fonksiyonlarının kullanımı ile ağı yerel çözümlere takılmasını önler. Bu nedenle stokastik YSA daha hızlı bir yakınsama yapar.

Boltzman makineleri stokastik YSA'nın özel bir versiyonudur. Stokastik hopfield ağlarının karşılığı olarak Boltzman makineleri gösterilebilir (Ackley, 1985).

### 3.2. Konuşma ve Ses Sinyalinin Özellikleri

Dalgalar temelde elektromanyetik dalgalara veya mekanik titreşimlere benzerlik gösterir. Elektromanyetik dalgalar her ortamda yayılabilme özelliğine sahiptir. Ses dalgaları gibi mekanik özelliklere sahip dalgalar ise enerjilerini taşıyabilmek için atomlara ihtiyaç duyarlar. Bu sebeple boşlukta yayılmaları mümkün değildir. Ses dalgaları ortamda basınç dalgaları olarak yayılırken ortamdaki atomlar harekete geçer.

Sesin meydana gelmesi fiziksel bir olay olarak nitelendirilir. Ses dalgalarının oluşmasını sağlayan nesnelere titreşimi ile ortama göre sıkışarak veya genişerek ilerleme sağlar. Aynı zamanda ses bir basıncın oluşmasına neden olur. Sesin özellikleri incelendiğinde yükseklik, tını ve şiddet gibi değişkenlerin olduğu gözlenir.

Sesin yüksekliğini etkileyen en önemli faktör frekanstır. Tiz seslerde yüksek frekans, pes seslerde ise düşük frekans bileşenleri bulunur. Ses dalgaları sinüzoidal dalga biçimindedir. Ses dalgalarının oluşumunda başrol oynayan parçacıkların birim zamanda sıkışıp gevşemesi ile ses dalgaları oluşur. Ses dalgasındaki titreşim aynı zamanda sesin frekansını belirler (Çelebi, 2012). Zaman birimi olarak saniye kullanılır. Saniye başına düşen titreşim sayısını tanımlamak için Hertz terimi kullanılır.

Ses dalgalarının genliği aynı zamanda sesin şiddetini belirtir. Sesin genliğinin artmasına paralel olarak sesin yüksekliği de artar (Meral, 2008). Ses dalgalarındaki minimum ve maksimum büyüklük arasındaki fark sesin genliğini belirler. Ortamdaki atomların titreşmesi sonucu ses dalgaları üretilir. Sesin enerjisi genliğin büyüklüğüne göre değişir. Genliği fazla olan ses dalgalarının taşıdığı enerji miktarı da fazladır.

Ses seviyesinin ölçümünde metrik olarak kullanılan ölçüt Bel'dir. Logaritmik bir ölçüt olan belde bir birim artış sesin şiddetinde 10 birim artışa eşittir. Bel'in onda biri olarak desibel kullanılır. Ölçümü yapılması planlanan ses doğrusal bir ölçek içerisinde tanımlanamadığından dolayı desibel ölçeği logaritmik özellik gösterir.

Sesin rengini, diğer bir ifadeyle sesin tınısını analiz etmek için sesin frekansı dikkate alınmalıdır. Aynı tonda aynı notayı farklı iki enstrümanda ayırmasını sağlayan ses tınısı değişkenidir. Sesin tınısı, sesi meydana getiren harmonik bileşenlere göre farklılaşır.

Konuşma sürecinde nefesli ve nefessiz sesler olmak üzere iki farklı uyarı çeşidi vardır. Nefesli seslerin oluşmasında soluk borusunun hava ile zorlandığı görülmektedir. Bu işlem ile ses telleri titreşime uğrar. Elde edilen bu titreşim ile yarı periyodik hava

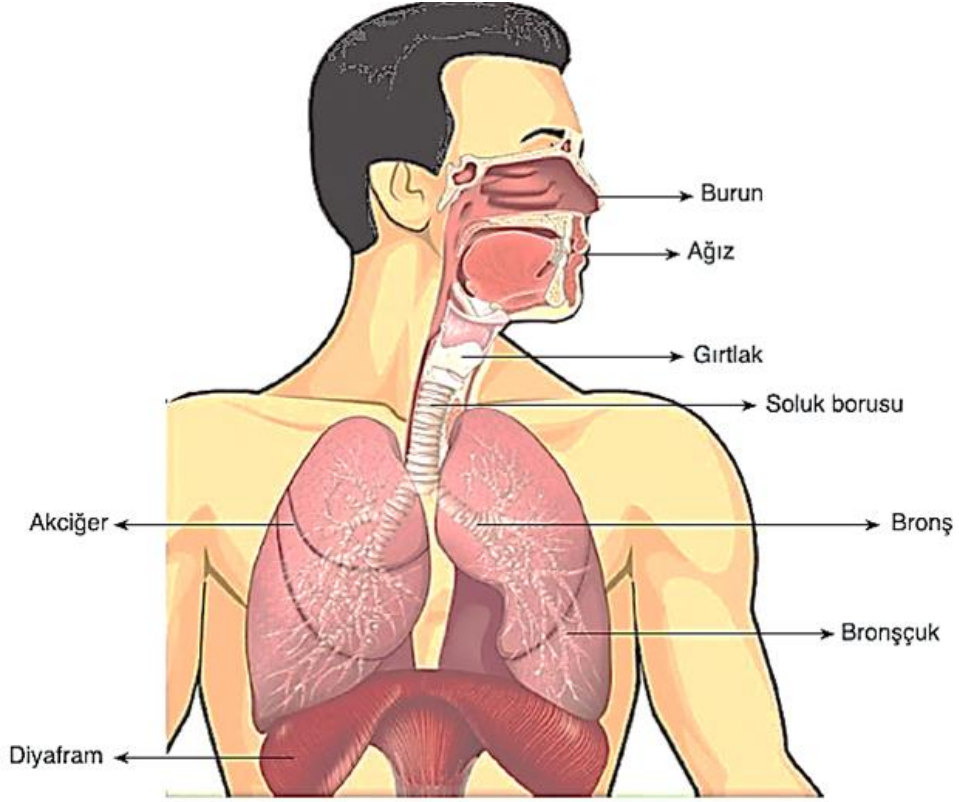
akımı oluşur. Bu sayede gırtlakta periyodik urayimler gerçekleşir. Bu şekilde oluşan seslere nefesli ses adı verilir (Shaneh ve ark., 2009). Nefessiz seslerin oluşumunda ses tellerinde herhangi bir titreşim gerçekleşmez. Nefessiz sesler gürültülü ve düşük frekanslıdır. Konuşma esnasında nefes yolu farklı şekiller olabilir. Meydana gelen rezonans, ses frekansını değiştirerek nefes yolunu karakterize eder.

Gürültü, duyulmak istenilmeyen ve insan sağlığına olumsuz etkisi bulunan ses türü olarak tanımlanır. Ses dalgası üzerinde frekans dağılımı ile gürültü iki şekilde kategorize edilir. Bunlar, sürekli geniş bant ve dar bant gürültüleridir. Eğer gürültü frekans bandında geniş bir spektruma yayılmışsa sürekli geniş bant gürültüsü adı verilir. Farklı frekans bileşenlerine sahip sesler bu kapsamda beyaz gürültü olarak değerlendirilir. Sürekli dar bant gürültülerinde belirli frekanslarda yoğunlaşma görülür. Bu tip gürültüler belirli aralıklarda baskınlık gösterir. Frekans dağılımı incelendiğinde belirli frekanslarda toplanma gözlenir.

### **3.3. Konuşma ve Sesin Oluşması**

Konuşma bireylerin bir biri arasında iletişimi sağlayan en önemli araçtır. Konuşma tanıma ise konuşmanın otomatik veya yarı otomatik sistemler tarafından tanınmasıdır. İşlem yapılması planlanan konuşma verisi kaydedilerek analog veriler dijital verilere dönüştürülür. Makine öğrenmesi algoritmaları tarafından konuşmalar algılanabilir ve anlaşılabilir. Konuşma verisinin işlenmesinde donanımsal olarak bilgisayarlar veya dijital sinyal işleyicileri kullanılır. Gerçek zamanlı konuşma tanıma uygulamalarında genel olarak DSP'ler kullanılır.

Konuşma sonucunda oluşan ses dalgaları basınç dalgaları olarak dinleyicinin kulakları tarafından toplanır. Konuşmanın oluşumunda görev alan başlıca organlar; akciğer, trake, gırtlak, ağız ve burundur. Sesin oluşumu gırtlakta başlayarak dudakta sona erer.



Şekil 3.111. İnsan konuşma ve ses oluşum sistemi (Webders Sitesi).

Ağızda ve gırtlakta bulunan yapılar farklı pozisyonlar ve hareketler ile değişik seslerin oluşmasını sağlarlar. Ses üretiminde hava ve gücü sağlayan yapılar ciğer ve kaslardır. Nefes ile vücuda giren hava akciğerler ve kaslar yoluyla gırtlığa taşınır. Ses tellerinin yer aldığı gırtlakta sesin oluşumu gerçekleşir.

Akciğerlerde depolanan hava kaslar sayesinde gırtlaktan geçerek ses tellerinde titreşime neden olur. Gırtlak yapısındaki farklı pozisyonlara göre elde edilen titreşimlerde değişiklikler olur. Elde edilen bu titreşimler ses organları ile farklı formlara dönüştürülerek işitilebilecek ve tanınabilecek bir yapıya kavuşur. Ses organlarındaki farklılıklar oluşan seslerin birbirinde ayrıt edilmesini sağlar. Ayrıca ses üretim yol uzunluğu cinsiyete göre farklılık gösterir.

İyi ve tanınabilir sesin elde edilebilmesi için ses tellerinin karar yapılarının düzgün ve temas halinde olması gerekir. Bazı yapısal ve işlevsel sorunlardan dolayı ses bozuklukları meydana gelebilir.

### 3.4. Öğrenme

Sınıflandırma ve örüntü tanıma problemlerinde eğitim verilerini kullanarak bilgi edinme yöntemine öğrenme adı verilir. Makine öğrenmesinde denetimli, denetimsiz ve güçlendirilmiş öğrenme olmak üzere üç sınıf öğrenme yöntemi bulunur. Öğrenme türlerinden bahsedilmeden önce veri çeşitlerine açıklama getirmek gerekir. Aynı zamanda öğrenme yöntemleri eğitim setinde bulunan hatalı verileri ayıklamak için kullanılan özel bir tekniktir.

Kullanıcıların uygulamalarına göre veriler, etiketli ve etiketlenmemiş olarak iki gruba ayrılır. Etiketli veriler içerisinde problem ile alakalı özellikler ve sınıf etiketleri bulundurulur. Etiketli verileri kullanarak yeni gözlemler için tahmin yapılabilir. Kümeleme ve karar ağaçları yapılarının kullanıldığı problemlerde daha çok etiketlenmemiş veriler kullanılır. Verilerin etiket bilgisini yansıtan özelliklere makine öğrenmesinde veri özellikleri adı verilir.

#### 3.4.1. Denetimli öğrenme

Denetimli öğrenme kapsamında eğitim ve test verileri etiketleri ile beraber işleme konulur. Bu öğrenme çeşidinde ortalama karesel hata miktarını veya tahmin maliyetlerini azaltmak için çaba sarf edilir. Denetimli öğrenme yöntemleri eğitim verilerini kullanarak eğitim yapılmış bir model geliştirir. Eğitilen bu model test verilerini tahmin etmek veya etiketlemek için kullanılır. Denetimli öğrenmede, test verilerinin etiketleri tahmin ediliyorsa sınıflandırma adı verilir. Eğer sürekli bir sayısal değer elde edilmeye çalışılıyorsa regresyon analizi adı verilir. Sınıflandırma ve regresyon analizinin uygulandığı problemlerde optimizasyon yöntemleri ile tahmin maliyeti en aza indirgenir.

#### 3.4.2. Denetimsiz öğrenme

Denetimsiz öğrenme yönteminde eğitim verilerinin sınıf etiketleri bulunmaz. Denetimsiz öğrenme veriler arasındaki benzerlikler dikkate alınarak eğitim örneklerine gruplama işlemi uygulanır. Gruplama işlemi literatürde aynı zamanda kümeleme işlemi olarak da bilinir.

Kümeleme işleminde asıl amaç veriler arasındaki bağlantıları ve benzerlikleri ortaya koymaktır. Kümeler arasındaki mesafeyi belirleyebilmek önemli bir metriktir.

Diğer önemli bir hususta kümeleri birbirinden ayıran sınırlarının belirlenmesi ve gerekli küme sayısını elde etmektir.

### **3.4.3. Veri çeşitleri ve veri hazırlama yöntemleri**

Öğrenme yöntemlerinde girdi değişkenleri olarak verilen veriler farklılık gösterebilir. Bazı öğrenme yöntemlerinde nicel girdilere ihtiyaç duyulurken bazıları ise nitel verilerin kullanılmasına daha yatkındır. Bu nedenle veri çeşidine göre öğrenme yöntemleri farklılaşabilir. Özellikle nitel veriler için tasarlanan modelleri nicel veriler üzerinde kullanmak için nicel veriler üzerinde kümeleme yapılması gerekir.

Makine öğrenmesinde kullanılan özellik değişkenleri; sıra, nominal, aralık ve oran olmak üzere dört kategoride değerlendirilir. Sınıflama ve kümeleme uygulamalarında nominal ve sıra değişkenlerinin kullanımı daha çok tercih edilir. Nominal değişkenler ile verilerin ayrıtılmasında tanımlayıcı özellikler sağlar. Sıralı değişkenler ise veri içerisinde daha anlamlı bir sıralama gerçekleştirilir.

Sıcaklık ve tarihsel ölçümlerde eşit aralık değişkenleri kullanılabilir. Orantısal hesaplama çerçevesinde aralık değişkenlerinin kullanımı pek tercih edilmez. Bu nedenle oran değişkenlerinin kullanımı daha uygundur. Oran değişkenlerinin anlamsal hesaplamalar içermesi ile daha iyi performansın elde edilmesini sağlar. Veri elde edilmesi sırasında verilerin ölçülmesi veya hatalı ölçümü söz konusu olabilir. Bu faktörler öğrenme süresini negatif yönde etkileyebilir. Eksik ve hatalı verilerde diğer atlama ve interpolasyon teknikleri ile yeniden ayarlama yapılarak öğrenme işleminin performansı artırılabilir. Farklı aralıklara sahip verilerin maliyet fonksiyonuna etkilerini aynı seviyeye getirmek için normalleştirme ve standartlaştırma uygulanır.

### **3.4.4. Veri setlerinin ayrımının gerçekleştirilmesi**

Veri madenciliği ve veri işleme uygulamalarında veriler, rastgele bir şekilde eğitim, doğrulama ve test verisi olmak üzere ayrımı gerçekleştirilir. Denetimli öğrenme yöntemlerinde eğitim verileri etiketli bir şekilde kullanılır. Elde edilen modelin performansını denemek amacıyla eğitim ve doğrulama setlerinden farklı olarak test seti kullanılır.

Kullanılan veri setinin eğitim, doğrulama ve test setlerine ayrılması için genel bir kural bulunmamaktadır. Fakat, eğitim veri sayısının doğrulama ve test veri sayısından fazla olması istenir.

### 3.4.5. Boyutsallık laneti

Makine öğrenmesinde kullanılan girdi verileri yüksek boyutları itibariyle farklı analizler gerektirebilir. Bu durum öğrenme sürecinde bir takım sorunların ortaya çıkmasına ve performans düşüklüğüne sebep olabilir.

Veri azaltma ile verilerin depolanması için gereken bellek sorununun üstesinden gelinirken öğrenme yönteminin hesaplama zamanında da azalma sağlanır. Yeterli miktardaki veri sayısı ile boyutsallık problemi kısmen aşılabilir. Problemin zorluğuna göre fazla sayıda verilere ihtiyaç duyulabilir. Veri sayısının fazla olması elde edilen özellik matrisinin boyutunu artırır. Bu zorluğa boyutsallık laneti adı da verilir.

Boyutsallık laneti temel olarak yüksek boyutlu verilerin işlenmesi için gereken değişken sayısının fazlalığından kaynaklanır. Aynı zamanda bellek sorununa da neden olur. Diğer dezavantajları arasında ise veri seti içinde gürültü oluşturması ve özellikleri gizlemesi ile sınıflandırma probleminin çözümünü zorlaştırır.

### 3.4.6. Genelme ve aşırı uyum

Eğitimi tamamlanmış modelin genelleme yapabilmesi için test verilerinde yüksek performans göstermesi gerekir. Öğrenme yöntemlerinin genelleme yapabilmesi için model hatasının mümkün olduğunca küçük olması gerekir. Öğrenme yöntemlerinin genelleme yeteneği farklı test verileri ile doğrulanmalıdır.

Öğrenme yöntemleri eğitim verileri üzerinde sınıflandırma hatasını azaltmaya çalışır. Diğer dikkat edilmesi gereken husus ise öğrenme yöntemlerinin eğitim ve test verilerinin her ikisini de genellemesinin gerektiğidir. Öğrenme yöntemleri eğitim verileri üzerinde yüksek performans gösterirken doğrulama ve test verilerinde aynı performansı göstermeyebilir. Bu duruma aşırı uyum problemi adı verilir. Yeni test verilerinde düşük performans göstermesi öğrenme yöntemlerinin yapısından kaynaklanır. Makine öğrenmesi uygulamalarında aşırı uyum problemi ile sıkça karşılaşmaktadır.

### 3.4.7. Boyut indirgeme

Makine öğrenmesi uygulamalarında diğer bir sorun ise veri boyutunun yüksek olmasıdır. Hesaplama yükü de bu alanda en önemli sorundur. Hem hesaplama değişkeni sayısını azaltmak hem de önemli özellikleri elde etmek için boyut indirgeme uygulanır. Başarılı bir sınıflandırma yapabilmek için yöntemin genelleme yeteneği yüksek ve boyut indirgemenin uygun bir şekilde yapılması gerekir.

İstatistiksel veya evrimsel işlemler sonucu elde edilen özellikler matrislerinden boyut indirgeme yöntemleri ile en önemlileri seçilir. Boyut indirgeme yöntemlerinden biri olan temel bileşen analizi ile doğrusal boyut indirgeme işlemi yapılabilir. Bu analizde varyans değerleri kullanılarak daha küçük boyutlu özellik haritaları elde edilir.

### 3.5. Derin Öğrenme

Derin öğrenme algoritmaları YSA'nın üzerine inşa edilmiş bir yapıdır. ESA'da oto kodlayıcı ve kısıtlı Boltzman makineleri bulunmaktadır.

Makine öğrenmesinin alt dallarından biri olan derin öğrenme son yıllarda büyük gelişme göstermiştir. Derin öğrenme adını derin ve hiyerarşik yapısından alır. Günümüze kadar sinyal ve görüntüden elde edilen özelliklerin işleminde Gaussian karışım modeli, destek vektör makineleri ve YSA mimarileri klasik makine öğrenmesi algoritmaları olarak kullanılmıştır.

Son zamanlarda kullanılan derin öğrenme algoritmaları gösterdikleri performans ile sinyal ve görüntü işleme uygulamalarına damgasını vurmuştur. Derin öğrenme algoritmalarının makine öğrenme alanındaki problemleri etkin bir şekilde çözülebildiği görülmüştür. Kullanım amacına göre derin öğrenme algoritmaları birçok modele sahiptir. Makine öğrenmesinde gelişen bir alt dal olan derin öğrenme doğrusallık problemini aşmak için birçok karmaşık katmana sahiptir. Derin öğrenmede denetimli ya da denetimsiz özellik seçimi, veri dönüşümü ve sınıflandırma işlemi yapılmaktadır. Daha spesifik ve daha karmaşık özellikler ESA ile elde edilebilir. Derin öğrenmenin daha iyi anlaşılabilmesi için makine öğrenmesinin alt dalı olan YSA'nın öğrenilmesi gerekir.

Derin öğrenmeye olan ilgi her geçen gün giderek artmaktadır. Derin öğrenme algoritmalarının fazlaca kullanılmasının bazı nedenleri bulunmaktadır. İlk olarak Grafik İşlem Biriminde (GİB) yaşanan gelişmeler ile çok sayıdaki veriyi paralel bir şekilde kısa



sürede işlenmektedir. GİB yapısında bulunan fazla sayıdaki işlem çekirdeği ile elde edilen matrisler aynı anda işlenir. GİB'in eğitim sürecini önemli miktarda düşürdüğü gözlemlenmiştir. GİB sayesinde veri miktarındaki artış önemli bir sorun olmayacaktır. Yapay zekâ alanındaki gelişmeler özellik çıkarma yöntemleri ve makine öğrenmesi yöntemlerinin önemi her geçen gün artmaktadır. Bunun yanı sıra hem eğitim hem de test verilerindeki yüksek sınıflama başarısı ile ön plana çıkmaktadır.

Derin öğrenmenin temel avantajı, istatistiksel ve hiyerarşik özellik çıkarma yöntemlerine kıyasla otomatik ve daha etkili bir özellik çıkarım modelinin bulunmasıdır. Derin öğrenme algoritmalarının kullanıldığı uygulamalarda eğitim veri setinden özellikler öğrenir. Derin öğrenme yapısındaki ardışık katmanlarda farklı seviyedeki özellikler öğrenilir. Bu algoritmalar istatistiksel özellikleri öğrenmez. Derin öğrenme üst seviyedeki özellikleri öğrenir ve yararlı özellikleri elde eder. Ancak doğrusal özelliğe sahip olan derin öğrenme modelleri az veri ile eğitildiğinde test verilerinde düşük başarı oranları elde edebilir. Bu durumda üst seviye özellikleri öğrenmek oldukça zordur (Bengio ve ark., 2013).

Derin öğrenme modellerinin birçoğu boyutsallık lanetinin üstesinden gelmiştir. Makine öğrenme problemlerinin çözümünde kullanılan modellerin sığ yapıda olması boyutsallık lanetinin oluşmasında ana unsurdur. Bu problemin çözümünde derin öğrenme algoritmalarında başlangıç değerlerinin uygun bir şekilde atanması gerekir. Ayrıca farklı öğrenme yöntemlerinin kullanılması da boyutsallık laneti probleminin üstesinden gelir.

Derin öğrenme algoritmaları YSA'nın üzerine kurulmuş bir yapıdır. Derin öğrenmedeki katmanlar ile YSA'ndaki katmanlar birbirine benzerlik gösterir. Derin ağlarda farklı özellik eklemesi de yapılabilir. Bu modellere örnek olarak derin Boltzman makineleri gösterilebilir.

Derin ağlarda yapay sinir sayısını ve katman sayısının artırılması aranan çözümü daha optimum şekilde yakınsamasını sağlar. Derin ağlarda katman sayısı ve filtre sayısı artıkça hesaplanması gereken değişken sayısı da artar. Bu durum hesaplama süresini de artırır. Aynı zamanda bellek ve donanımsal kaynaklara da ihtiyaç duyulabilir.

Derin ağlardaki gizli katman sayısının artırılması ağın her zaman iyi performans göstermesini sağlamaz. Gizli katman sayısının artırılması ilk katmanlardaki geri yayılım etkisini azaltır. Böylece gradyan azalma algoritmasının küresel çözümün elde edilmesinde başarısız olur. Makine öğrenme uygulamalarında bu durum yok olma

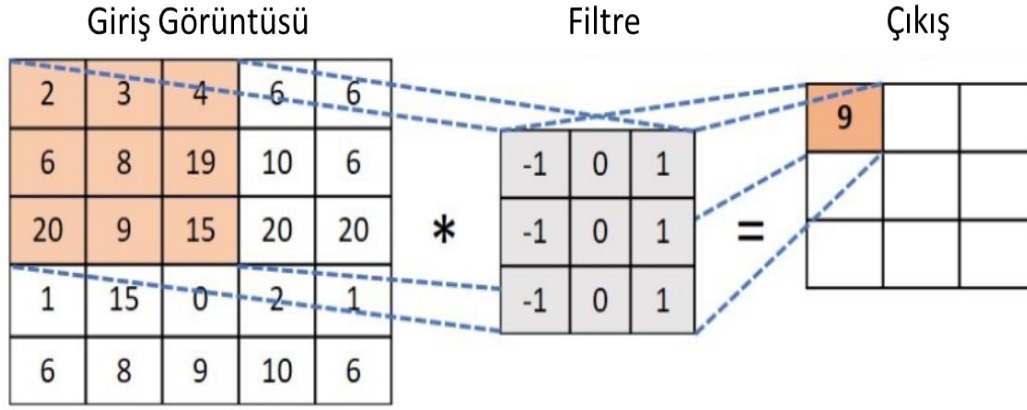
gradyan problemi olarak tanımlanır. Derin ağların diğer bir dezavantajı ise eğitim verilerini ezberlemesidir.

Aşırı uyum ve yok olma gradyanı problemlerinin üstesinden gelmek amacıyla bazı denetimsiz eğitim yöntemleri geliştirilmiştir. Aç gözlü optimizasyon yöntemleri kullanarak eğitim işlemi katmanlar arası ardışıl veya ayrı ayrı yapılmalıdır. Eğitim işlemi ilk katmandan başlayarak son katmana kadar devam eder. Sonrasında ağın tamamına ince ayar işlemi uygulanır.

Derin öğrenme algoritmaları, yapay zekâ, örüntü tanıma, görüntü ve sinyal işleme alanları ile doğrudan ilişkisi bulunur. Bilgisayarlı görme uygulamaları başta olmak üzere karakter tanıma, biyoinformatik, tıbbi alanlar üzerinde de başarılı çalışmalar gerçekleştirilmiştir. Derin öğrenme algoritmalarının endüstri üzerindeki etkileri 2010 yılından bu yana devam etmektedir.

### **3.5.1. Konvolüsyon**

Matematiksel bir işlem olan konvolüsyon iki boyutlu görüntülerde pikselin rengi, komşu piksellerin özelliklerini dikkate alarak yeni bir değer üretir. Konvolüsyon penceresi içerisindeki ağırlık katsayıları görüntü üzerinde işleme girer ve özellik haritaları oluşturur. Konvolüsyon penceresindeki ağırlıklara çekirdek veya filtre adı verilir. Filtre içerisindeki ağırlık değerlerinin her birinin etki değeri vardır. Şekil 3.12.'de gösterilen konvolüsyon işleminde görüntü ve filtre katsayıları işleme girerek toplam filtre değeri özellik haritasına atanır. Piksel değerleri Eş. 3.25'e göre hesaplanır. Denklemde filtre (i, j) değerleri satır ve sütun filtre değerini gösterir. Piksel (x,y) x. Satır y. sütundaki piksel değerini temsil eder. kw ve kh filtrelerin ağırlık ve yüksekliğini gösterir. Eş. 3.26'da x girdi matrisini gösterir. Elde edilen sonuç matrisi özellik haritası olarak da isimlendirilir (Bengio ve ark., 2016).



Şekil 3.112. Konvolüsyon işlemi.

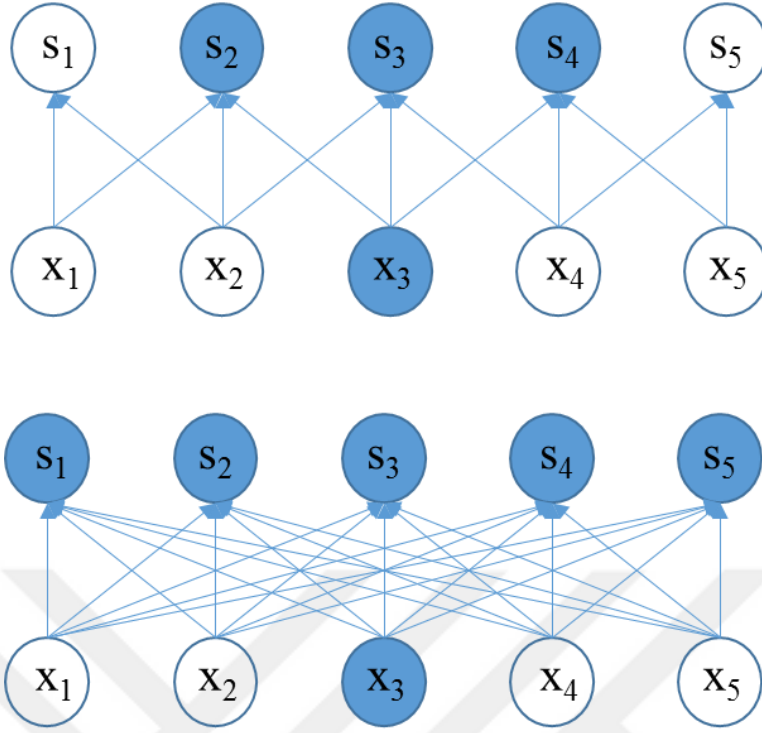
$$piksel(x, y) = \sum_{j=0}^{kh} \sum_{i=0}^{kw} Görüntü(x+i - \frac{kw-1}{2}, y+j - \frac{kh-1}{2}) filtre(i, j) \quad (3.25)$$

$$s(t) = (x * w)(t) \quad (3.26)$$

### 3.5.2. Evrişimsel sinir ağları ve derin evrişimsel sinir ağları

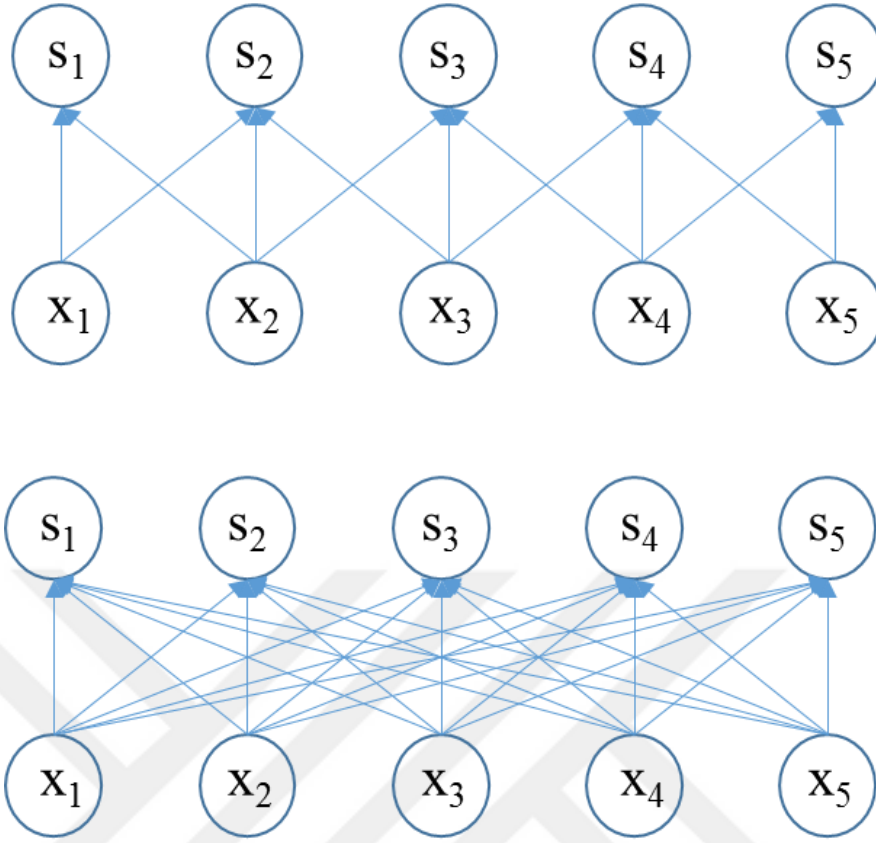
ESA ve Derin Evrişimsel Sinir Ağları (DESA) diğer derin öğrenme yapılarından biraz daha farklıdır. Bu yapılar daha çok görüntü işleme ve sinyal işleme uygulamalarında kullanılır. Bu ağlar konvolüsyon işlemi kaydırarak bütün görüntüyü tarayarak özellik haritası çıkarır. Son kısmında YSA'na benzer tam bağlı katmanlar ile sınıflandırıcılarla görüntü tanıma işlemi yapılır.

Makine öğrenmesinde bulunan ağlarda giriş birimi ile çıkış birimi arasındaki etkileşim çeşitli değişkenler ile açıklanmaktadır. ESA'da seyrek gösterim matrisleri elde edilir. Bu işlemde konvolüsyon filtreleri kullanılarak giriş görüntüsü boyutu düşürülür. Yapı içerisindeki konvolüsyon filtreleri giriş görüntüsünden seyrek gösterim matrisi elde eder. Elde edilen seyrek gösterim matrisleri bellekte daha az yer kaplar. Şekil 3.13.'de giriş ve çıkış arasındaki bağlantı gösterimi gerçekleştirilmiştir. Şekilde gösterilen yapı incelendiğinde üç çıkış sadece x girişinden etkilenir. Diğer ağ yapısında ise x girişi bütün çıkışları etkilemektedir.



Şekil 3.113. Seyreklik bağlantısı.

Model içerisinde aynı değişkenlerin birden fazla kullanılması değişken paylaşımı olarak nitelendirilir. YSA'nda ağırlık katsayılarının her biri çıktı hesabı içinde bir kez kullanılır. Bu yapılar tam bağlı katmanlar olarak adlandırılır. ESA'da filtre katsayılarının oluşumunda giriş elemanlarının etkisi vardır. Her görüntüde ayrı ayrı değişken öğrenmek yerine veri seti üzerinden tek değişken öğrenilir. Şekil 3.14'de kullanılan oklar iki farklı model için belirli değişken bağlantılarını gösterir. Şekil 3.14'de bulunan ilk modelde çekirdek değişkenlerinin üç farklı çekirdeğe etkisi görülmektedir. Parametre paylaşımı ile tüm giriş konumlarında bir parametre kullanılır. İkinci modelde ise tam bağlı bir yapı gösterilmiştir. Parametre bir kez kullanılır ve herhangi bir paylaşım yoktur.



Şekil 3.114. Parametre paylaşımı.

### 3.5.3. Toplu normalizasyon

Toplu normalizasyon işlemi katmanlar arasında gerçekleştirilir. Eğitim aşamasından mini batch yapıları içerisinde normalizasyon işlemi yapılır. Bu sayede ağıın eğitim hızı artar. RMSProp ve SGD gradyan azalma optimizasyon yöntemleri DESA'nın eğitimi için kullanılır.

### 3.5.4. Aktivasyon fonksiyonları

YSA'nın doğruluk oranlarını etkileyen faktörlerden biri de aktivasyon fonksiyonu seçimidir. Aktivasyon fonksiyonu seçimi ile ağıın işlevselliği ve öğrenme yöntemi belirlenir. Klasik geri yayılım algoritması iki veya üç katmana sahip ağılar üzerinde yüksek performans sağlar. Fakat derinliğin artması ile farklı problemler oluşur.

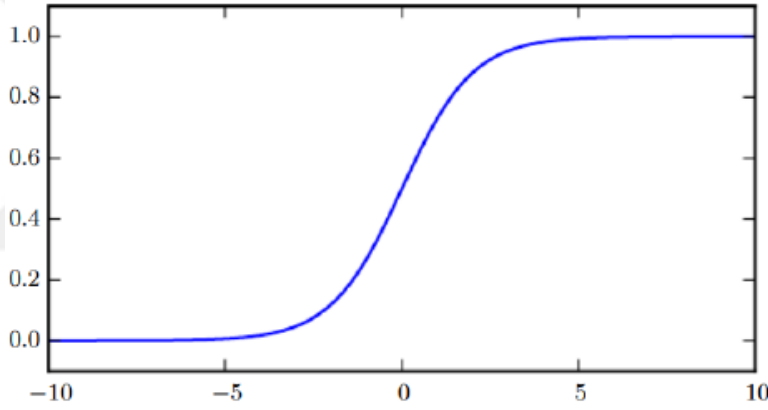
Hatanın etkisi bütün katmanlara yayıldığı için mevcut sonuç aktivasyon fonksiyonunun türevi ile işleme girer. Sigmoid fonksiyonu türevinin algılama alanı kısıtlı

olduğundan dolayı birkaç katman sonrası hata sifira yaklaşır. Aktivasyon fonksiyonunun tersinin türevi sınırsız olması durumunda ise geri yayılım sonrasında hatada artışlar meydana gelir. Bu durum ağda kararsızlıklara sebep olur.

### 3.5.4.1. Sigmoid

Sigmoid aktivasyon fonksiyonu rastgele bir giriş verisi alır ve çıkışta sıfır ile bir arasında gerçek sayılar üretir.  $-\infty$ 'a doğru gidildikçe değer sıfıra yaklaşırken  $+\infty$ 'a doğru değer bir'e yaklaşır. Şekil 3.15'de sigmoid aktivasyon fonksiyonu kullanılmıştır. Sigmoid fonksiyonunun sonucu Eş. 3.27'e göre hesaplanır.

$$f(s) = \frac{1}{1 + e^{-s}} \quad (3.27)$$



Şekil 3.115. Sigmoid fonksiyonu.

### 3.5.4.2. Rectified linear unit (ReLU)

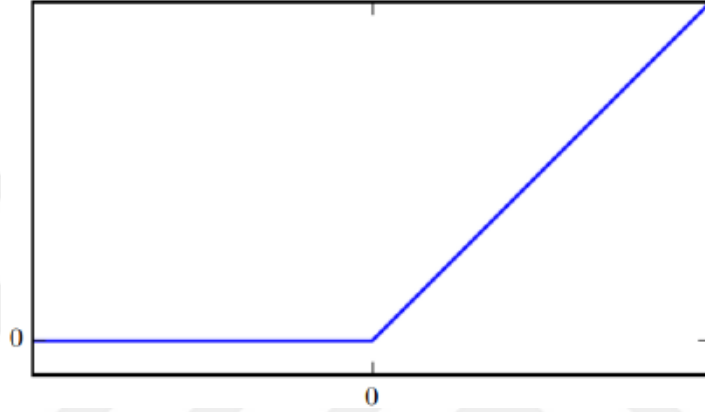
Yeteri kadar katman barındıran YSA'nda aktivasyon fonksiyonları doğrusallık özellik göstermiyorsa kompleks fonksiyonlara yakınsama yapabilir. ReLU fonksiyonu doğrusal olmayan özellik gösterir (Nielsen, 2015). Fakat artan veya azalan gradyanlarla birlikte çeşitli problemlere yol açar. Aktivasyon fonksiyonu olarak ReLU'nun seçimi ağı yakınsama hızını ve oranını önemli ölçüde etkiler.

Sinir ağlarının, içlerinde yeterli katman varsa ve aktivasyon fonksiyonu lineer değilse, rasgele karmaşık bir fonksiyona yaklaşma kabiliyetleri olduğu bilinmektedir. Sigmoid gibi aktivasyon fonksiyonları doğrusal değildir, ancak azalan veya artan

gradyanlarla birlikte problemlere yol açar. ReLU kullanılması, sigmoid ile karşılaştırıldığında stokastik gradyan inişinin yakınsama oranını önemli ölçüde iyileştirir (Nair ve ark., 2010).

ReLU fonksiyonu Şekil 3.16'da gösterilmiştir ve Eş. 3.28 ile hesaplanabilir.

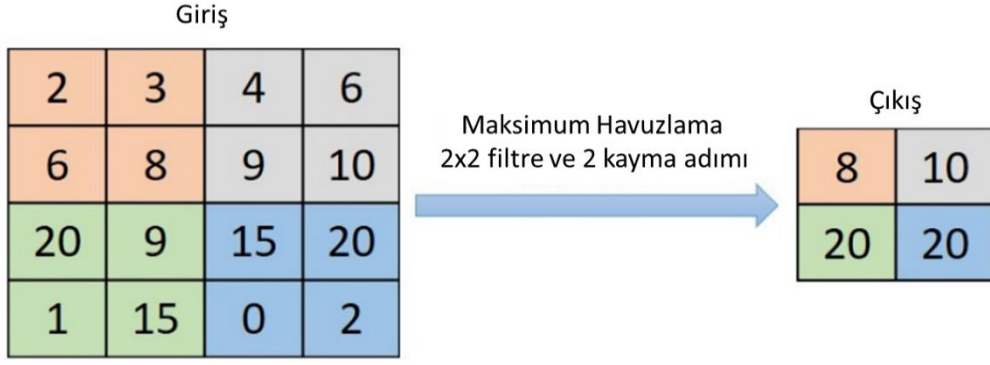
$$f(s) = \max(0, s) \quad (3.28)$$



Şekil 3.116. ReLU fonksiyonu.

### 3.5.5. Havuzlama

ESA'da konvolüsyon katmanlarının yanı sıra havuzlama katmanları da bulunur. Havuzlama katmanı ile özellik seçimi yapıp boyut indirgenmesi gerçekleştirilir. ESA modellerinde konvolüsyon katmanlarında alt örnekleme için havuzlama katmanı yaygın bir şekilde kullanılır. Ağdaki parametre sayısı ve hesap yükünü azaltmak için elde edilen özellik haritalarının boyutunun azaltılması gerekir. Havuzlama katmanı ağın derinliğinden etkilenmez. Maksimum veya ortalama fonksiyonları kullanarak ölçekleme işlemi gerçekleştirilir. Şekil 3.17'de  $2 \times 2$  boyutunda 2 kayma adımına sahip havuzlama katmanı kullanılmıştır. Bu özellikteki havuzlama katmanı ile görüntünün genişlik ve yüksekliğinde iki birim azalma olur. Maksimum havuzlama işleminde dört değer içerisinde maksimum değer seçilir ve özellik haritasına eklenir.



Şekil 3.117. Maksimum havuzlama.

### 3.5.6. Dropout katmanı

ESA aşırı uyum durumunda belirli eğitim verilerinde yüksek sınıflama başarısı gösterir. Başka bir ifadeyle model sadece eğitim setini öğrenir. DESA modelleri tam bağlı katmanlarından farklı değişkenler içerir. Yeterince öğrenmenin olmadığı durumlardan aşırı uyum problemi kendini gösterebilir. Eğitim aşamasında doğrulama veya test verileri ile ağın performansı belirli döngü sayılarından test edilebilir. Böylece eğitim verileri aşamasında doğrulama veya test verileri ile ağın performansı belirli döngü sayılarında test edilmelidir. Böylece eğitim verileri üzerinde bulunan gürültüler giderilebilir. Bu işlemler için dropout katmanı önerilir. Bu işlem aşırı uyum problemi için uygulanacak en basit yöntemdir.

### 3.5.7. Maliyet fonksiyonu

Veri setinde ESA modellerinin performansını belirlemek amacıyla maliyet fonksiyonlarına ihtiyaç duyulur. Çapraz entropi maliyet fonksiyonu ESA modellerinde en çok kullanılan maliyet fonksiyonudur. Eş. 3.29'da çapraz entropi maliyet fonksiyonu gösterilmektedir.

$$L(w) = -\frac{1}{N} \sum_{i=0}^N \sum_{j=0}^J y_{ij} \cdot \log(y_{ij}^*) + (1 - y_{ij}) \cdot \log(1 - y_{ij}^*) \quad (3.29)$$

Eş. 3.22'de, J sınıf sayısını, N örnek sayısını belirtir,  $y^*$  modelin tahminini ifade eder. Çapraz entropi fonksiyonu tahmin edilen çıktı değerleri ile gerçek çıktı değerleri



arasında farkı olasılıksal olarak değerlendirir. Entropi kaybını azaltmak için sürekli olarak parametre güncellemesi yapılır.

### 3.6. Sınıflandırıcılar

#### 3.6.1. Softmax

Literatürde en çok kullanılan ve kabul gören Softmax fonksiyonunun matematiksel alt yapısı aşağıdaki gibi ifade edilir.  $\sigma$  ile beraber softmax fonksiyonu Eş. 3.30'da verilmiştir.

$$\sigma(z) = \frac{1}{\sum_{j=1}^n \exp(\lambda z_j)} \begin{bmatrix} \exp(\lambda z_1) \\ \vdots \\ \exp(\lambda z_n) \end{bmatrix}, \quad \lambda > 0 \quad (3.30)$$

$\lambda$  ifadesi fonksiyon içerisinde ters sıcaklık sabiti olarak adlandırılır. Softmax fonksiyonu literatürde farklı bileşen içerir.

$$\sigma_i(z) = \frac{\exp(\lambda z_i)}{\sum_{j=1}^n \exp(\lambda z_j)}, \quad 1 \leq i \leq n \quad (3.31)$$

$\lambda=1$  olduğu durumda Eş. 3.31 standart Softmax adını alır.  $\lambda=0$  ditiğinde  $\sigma$  fonksiyonu çıkışı düzgün olasılıksal dağılıma göre yakınsama yapar. Bu softmax fonksiyonu lojistik fonksiyon formuna indirgendiğinde;

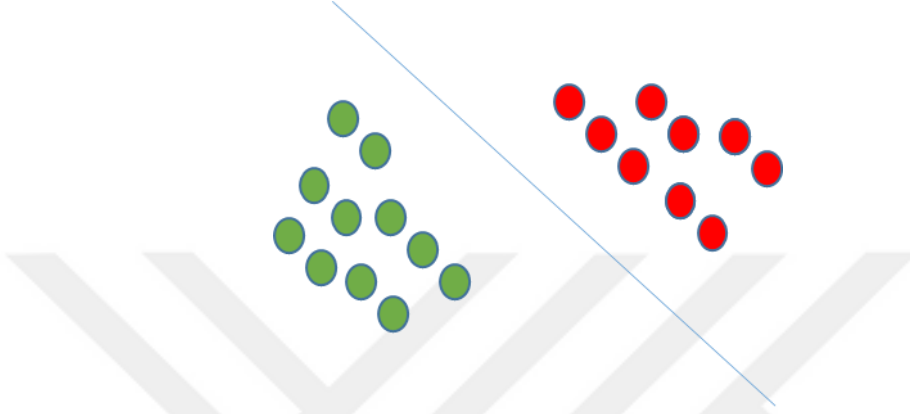
$$\sigma_i(z) = \frac{\exp(\lambda z_i)}{1 + \exp(-\lambda(z_i - z_j))}, \quad j \neq i \quad (3.32)$$

$$\sigma_i(z) = \exp(\lambda z_i - \log(\sum_{j=1}^n \exp(\lambda z_j))) \quad (3.33)$$

Eş. 3.33 nadiren softmax fonksiyonunun bir temsili olarak kullanılmasına rağmen ESA teorisinin kopyalayıcı dinamiğinin çözümü olduğunu belirtmiştir.

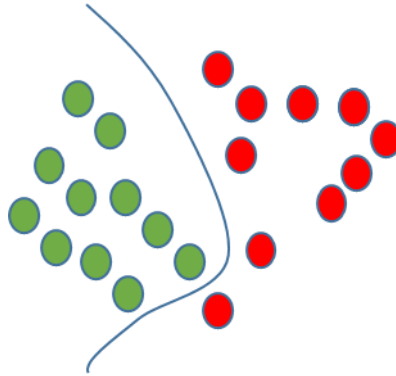
### 3.6.2. Destek vektör makineleri

Destek Vektör Makinesi (DVM) düzlemler kullanarak karar verebilme yeteneğine sahip bir algoritmadır. Kullanılan düzlemler farklı özelliklere sahip sınıfları birbirinden ayırır. Şekil 3.18’de gösterildiği gibi farklı iki renge sahip nesnelere bulunmaktadır. Ortalarında bulunan düzlem bu sınıfların birbirinden ayrımını sağlar.



Şekil 3.118. Nesnelere karar düzlemi ile ayrımı.

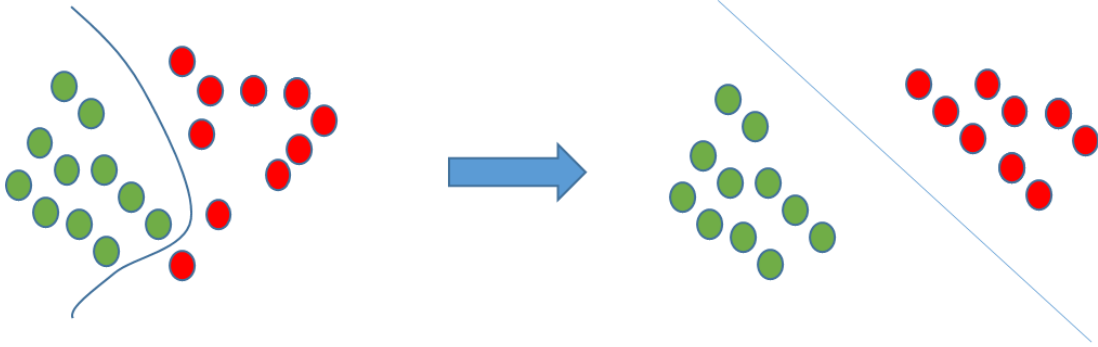
Şekil 3.19’de farklı renge sahip nesnelere birbirinden ayırmak için doğrusal düzlem yerine parabolik düzlem gerekir. Farklı özellikler barındıran sınıfları birbirinden ayırmak için hiper düzlemler kullanılabilir. Düzlemsel ayırım noktasında DVM oldukça uygun bir yapıya sahiptir.



Şekil 3.119. Nesnelere doğrusal olmayan düzlemler ile ayrımı.

Şekil 3.20’de DVM’nin çalışma prensibi gösterilmektedir. Şekildeki nesnelere DVM içerisinde bulunan çekirdekler sayesinde düzlem değiştirirler. Bu işleme eşleme ve dönüşüm adı da verilir. Dönüşümü gerçekleştirilmiş nesnelere doğrusal düzlemler ile

ayrımı yapılabilir. Bu sayede karmaşık eğriler kullanılmadan daha doğru bir şekilde sınıflandırma gerçekleştirilmiştir.



Şekil 3.20. Eşleme veya dönüşüm sonucu.

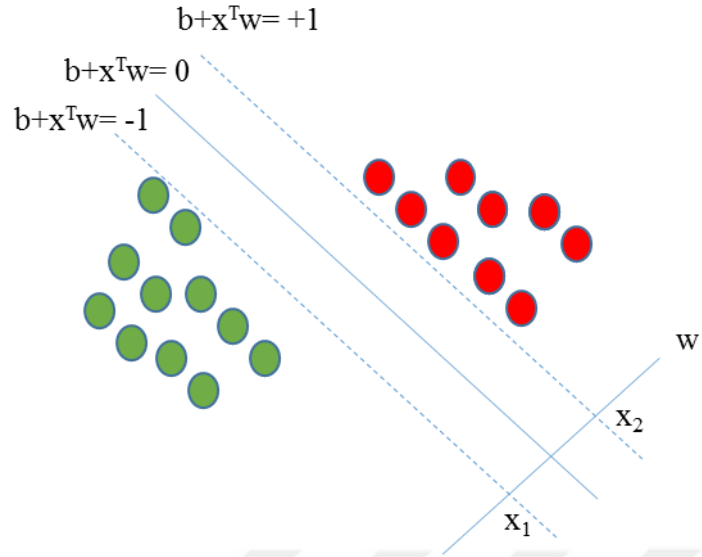
DVM özelliklerin ayrımında kullanacağı düzlemleri eğitebilecek bir yapıya sahiptir. Bu eğitim işlemi öğrenme algoritmaları kullanılarak gerçekleştirilir. Bu kapsamda Vapnik tarafından geliştirilen öğrenme algoritması geniş bir uygulama alanına sahiptir. DVM yapısı sınıfları birbirinden ayırabilmek için düzlemler arasında mesafenin artırılması ve çekirdek fonksiyonları ile uzay dönüşümü kullanılır (Masud ve ark. 2011).

Denetimli öğrenme yöntemlerinden biri olan DVM algoritması, verileri eğitim, doğrulama ve test olmak üzere üç kategoriye ayırır. DVM doğrusal ve doğrusal olmayan sınıflama işlemi yapabilir.

DVM problemlerdeki genelleme yeteneği ve boyutsallık sorununun üstesinden gelmesinden dolayı kullanımı oldukça yaygındır. DVM içerisinde arama uzayında küresel çözüme yaklaşarak hatayı azaltır. DVM algoritmasının avantajlarından biri de çok hızlı olmalarıdır. DVM ile gerçekleştirilen sınıflandırma işlemi özellik matrisinin boyutundan ve sayısından etkilenmez. Bu nedenler çok fazla örüntü kümesi öğrenebilir. Aynı zamanda anlık olarak örüntüleri güncelleyebilir.

DVM’de doğrusal düzlemler kullanarak sınıflandırma işlemi gerçekleştirebildiği gibi doğrusal olmayan düzlemler ile de bu işlem yapılabilir. Doğrusal olmayan sınıflandırmalarda çekirdek fonksiyonundan yararlanır. Kullanılan çekirdeklerin radyal, polinom, sigma ve iki katmanlı olarak üzere birçok çeşidi bulunur.

Şekil 3.21’de görüldüğü üzere  $w$  vektörü ayrıca hiper düzlem olarak kullanılmış ve diğer hiper düzlemler ile en büyük mesafeyi elde etmiştir. Elde edilen bu mesafe  $2/\|w\|_2$  olarak elde edilir (Watt ve ark., 2016).



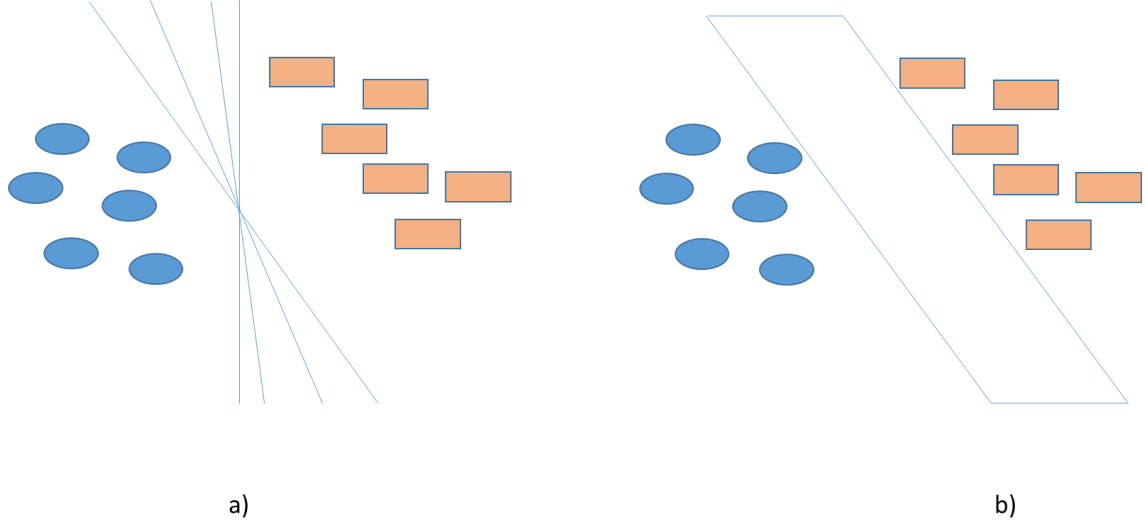
Şekil 3.21. Hiper düzlemler ile elde edilen mesafe.

İkili sınıflandırma problemlerinde DVM Eş. 3.34'daki gibi matematiksel olarak ifade edilebilir.

$$\min_{(w,b)} \frac{1}{2} w^T w \quad (3.34)$$

$$f(x) = \text{sign}(w \cdot x - b) \quad (3.35)$$

Eş. 3.35 kullanılarak ikili sınıflama problemi için doğrusal hiper düzlemler ile sınıflar arası mesafe artırılmıştır. Bu mesafenin büyüklüğü ve sınırları Öklid mesafesine göre ayarlanmaktadır. Şekil 3.22 a).’da çok sayıda ayırıcı bulunmaktadır. Fakat b).’de ise bir düzlem yardımıyla sınıflar arası mesafe artırılmıştır.



Şekil 3.22 İkili sınıf ayırma düzlemleri.

DVM’de kullanılan optimizasyon tekniği iki boyuta sahiptir. Eş. 3.36’de optimizasyon için gerekli olan fonksiyonu gösterir.

$$\max(L(w, b, \alpha)) = \frac{1}{2} w^T w - \sum_{i=1}^N \alpha_i y_i (w \cdot x_i - b) + \sum_{i=1}^N \alpha_i \quad (3.36)$$

$\alpha_i$ ’nin çözümü ile beraber  $w$  ifadesi farklı bir forma dönüşür. Eş. 3.37 kullanılarak sınıflara işlemini gerçekleştirilir.

$$f(x) = \text{sign}(wx - b) = \text{sign}\left(\sum_{i=1}^N \alpha_i y_i (x \cdot x_i - b)\right) \quad (3.37)$$

Denklemler dikkate alındığında skaler çarpıma göre işlem yapılmaktadır. Her eğitim verisi için noktasal  $\alpha_i$  bulunmaktadır. Maksimum mesafe elde edildiğinde aradaki hiper düzleme yakın noktalar için  $\alpha_i > 0$  durumunda olacaktır. Diğer noktalarda ise  $\alpha_i = 0$ ’dır. Hiper düzlem çerçevesindeki noktalar sınıflandırıcı olarak nitelendirilir. Aynı zamanda bu değerler DVM’nin sınıflama güvenilirliğini test etmek içinde kullanılabilir.

DVM’de beklenen hata en aza indirmeye çalışılır. DVM algoritması içerisinde ikilik teorisine dayalı etkin bir matematiksel alt yapı bulunmaktadır. DVM hata faktörü adı verilen bir değişkeni kullanıcıya sunabilmektedir. Aynı zamanda tasarımcıların örnek seçimi ve sınır belirlenmesinde konularda müdahale hakkı tanımaktadır. DVM’de işleme

girmemiş değişkenlere öznitelik adı verilir. Eğer öznitelikler çekirdekler ile işleme girerse özellik adını alırlar. Özelliklerin uygunluğunu tespit etmek için seçim algoritmaları bulunmaktadır (Jaiganesh ve ark., 2013).

DVM avantajları olarak;

- Mesafe artırımı ile etkin bir sınıflandırma yapabilir.
- Boyutsallıktan etkilenmezler
- Fazla sayıdaki uzaylarda yüksek performans gösterebilir.
- Donanımsal Bellek tasarrufu sağlar.

Dezavantajları;

- Veri kümelerinde gürültü olması durumunda düşük performans gösterebilir.
- Olasılıksal bir tahmin gerçekleştiremeyen bir yapıya sahiptir.

### 3.7. Spektrogram Analizi

Konuşma üretimi için basitleştirilmiş, ancak çoğu zaman kullanışlı bir model olan spektrogram, sesler için yarı periyodik bir uyarma işlevine sahip ve doğrusal olmayan sesler sırasında uyarma gibi bir gürültüye sahip doğrusal bir sistemden oluşur. Doğrusal sistem ses boşluğunu temsil eder. Ünlü sesler sırasında yarı-periyodik uyarma fonksiyonu, ses tellerinden geçen hava akışına karşılık gelir. Sürekli bir sesli harfin üretimi sırasında, ses yolu konfigürasyonunun sabit olduğu varsayılabilir ve üretilen ses, bir rezonans boşluğunun yanıtına karşılık gelir. Bir sabit durumlu sesli harfin uyarılma işlevi olup, buradaki sabitleme durumlu bir sesli harfin üretimini temsil eder. Konuşmada bulunan sesli harf periyodik olduğundan, spektrumu, perde periyodunun karşılığına göre frekansla aralıklı harmoniklere sahip bir çizgi spektrumudur. Ses boşluğunun rezonans frekansları içerisinde uyarılmaya karşılık gelen çizgi spektrumunu, vokal boşluğun frekans tepkisine karşılık gelen sabit durum sesli harfinin spektrumuna karşılık gelen bileşik çizgi spektrumunu temsil eder. Bu gibi sürtünme seslerinin üretimi sırasında, uyarma, dudaklarda ve dişlerde türbülans tarafından üretilen dalga biçiminde olmayan bir düzgün dalga biçimidir. Böylece, sürtünmeli sesler için, çıkış gürültüsüzdür ve çizgi spektrumu yoktur.

Sürekli konuşmaların üretimi sırasında, ses boşluğunun şekli sabit değildir ve rezonanslar farklı sesler üretmek üzere değişmektedir. Bir uyarma periyodu zamana bağlı bir değişkendir. Değişim çok hızlı değilse, sistemi kısa süreli ve sabit olarak görmek mantıklıdır, böylece konuşma dalga biçiminin kısa süreli bir spektral analizi, ses kanalının rezonanslarına karşılık gelen zirveler gösterecektir. Bunun yanı sıra uyarıma karşılık gelen harmonik bir yapı elde edilecektir. Zaman penceresi uzunluğu arttıkça, frekans çözünürlüğü artar ve harmonik yapı daha belirgin hale gelir, ancak spektral analiz hızlı değişiklikleri takip etme yeteneğini kaybeder.

Daha kısa bir zaman penceresine ve sonuç olarak daha geniş bir frekans penceresine sahip bir spektral analiz, spektral çözünürlük açısından daha iyi bir zaman çözünürlüğü sağlayacaktır. Spektrumdaki bireysel perde harmoniklerini çözme eğilimi göstermeyecek ancak hızlı değişiklikleri daha iyi izleyebilecektir. Zaman ve frekans çözünürlüğü arasındaki bu değişimler nedeniyle, iyi zaman çözünürlüğüne ve düşük frekans çözünürlüğüne karşılık gelen dar bant analizini kullanmak, konuşmanın spektral analizinde yaygındır. Dar bantlı bir analizde, filtre bant genişlikleri tipik olarak 45 Hz'dir. Geniş bant analizler için 300 Hz'dir. Dar bantlı analizde bireysel perde harmoniklerinin sıklıkta çözüldüğü açıktır, geniş bantlı analizde ise artık belirgin değildir. Bununla birlikte, geniş bantlı spektrogramda, bireysel adım periyodlarına karşılık gelen dikey çizgiler görülebilir. Bu özellik, geniş bantlı spektrogramın iyi zaman çözünürlüğü sağlar. Dar bant analizinde bu çizgiler belirgin değildir. Hem geniş hem de dar bant spektrogramlarında da belirgin olanı, ses yolu rezonanslarını belirleyen yüksek spektral genliğe karşılık gelen frekans bölgeleridir (Lynn, T. J. & Shaameri, A.Z., 2007).

Hızlı Fourier Dönüşümü (HFD) algoritması, dijital bilgisayarda spektrumları hesaplamak ve spektral ekranları uygulamak için bir araç olarak çok büyük bir öneme sahiptir. Özellikle, bir spektrograf makinesi tarafından uygulanan bir konuşma analizinin dijital bir bilgisayarda uygulanması için HFD algoritmasını kullanarak analizi gerçekleştirmek bir filtre bankası uygulamaktan çok daha etkilidir. HFD'ye benzer olarak tanımlanmış Ayrık Fourier Dönüşümünü (AFD) hesaplamak için kullanılan bir algoritmadır.

$$F(k) = \sum_{n=0}^{N-1} f(nT) e^{-j \frac{2\pi}{N} nk} \quad (3.38)$$

$f(nT)$ , analog  $f(t)$  zaman fonksiyonunun eşit aralıklı numunelerine karşılık gelir. Örnekleme Nyquist oranına eşit veya daha yüksek bir hızda yapıldığı varsayılarak,  $k$ . spektral noktasının büyüklüğünün  $|F(k)|$  Eş. 3.38'de bulunmaktadır.  $f(t)$  analog fonksiyonunun örnekleri,  $H(\omega)$  tarafından verilen bir frekans tepkisi ile analog bir filtreden oynatıldığında,  $t = (N-1)T$  zamanında elde edilebilecek büyüklüğe karşılık gelir (Gueguen, L., Sayrac, B., & Depierre, D, 2009).

$$H(\omega) = \frac{\sin \frac{NT}{2} (\omega - \frac{2\pi k}{NT})}{(\omega - \frac{2\pi k}{NT})} \quad (3.39)$$

$K$ , 0 ile  $N-1$ 'e eşit iken  $F(k)$  sayıları kümesi, daha sonra her filtre bir spektral şekle sahip, bir filtre bankasından elde edilen merkez frekansına ( $\omega = 2\pi k/NT$ ) sahip çıktı kümesine karşılık gelir.

Kısa süreli bir spektral analiz elde etmek için,  $t = (N-1)T$  hesaplamayı zamanın art arda kalıcılarında ve ayrıca filtre şeklini değiştirebilmek için yapılır. Örneğin, filtre özelliğindeki kenar loblarının azaltılması istenebilir. Ayrıca, geniş banttan dar bant analizine geçtikçe, filtredeki merkezi lobun genişliğini değiştirmek gerekir. Çalışan bir spektrum belirlemek ve filtre karakteristiği açısından esneklik sağlamak için Eş. 3.38'deki ifade şu şekilde değiştirilebilir:

$$F_T(k) = \sum_{n=0}^{N-1} w(nT) f(nT + rMT) e^{-j \frac{2\pi}{N} nk} \quad (3.40)$$

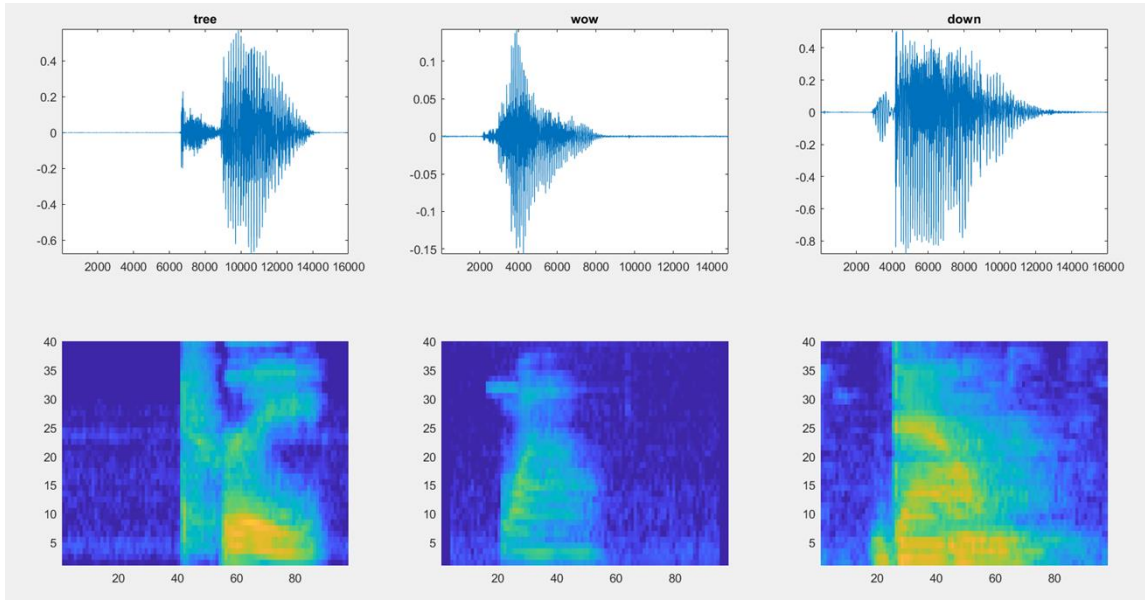
Eş. 3.39'da iki değişiklik yapılmıştır. Birincisi, daha iyi bir spektral özellik sağlamak için bir pencere  $w(nT)$  eklemektir. Bu sayede Ayrık Fourier Dönüşümü (AFD) hesaplamasına da fayda sağlar. Eş. 3.38'de sınırlı bir veri uzunluğundaki bir hesaplama sağlamıştır. Eş. 3.38'de  $f(t)$  'ye uygulanan bir zaman penceresi, yani  $f(t)$  eşit genişliğe sahip  $NT$  dikdörtgen bir pencere ile çarpılır. Dikdörtgen pencerenin  $w(nT)$  bazı yeni zaman pencereleriyle değiştirilmesiyle spektral şeklin değiştirilmesi mümkündür. Eş. 3.40'de dalga formunun ardışık bölümlerinin spektral analizinin uygulanmasına karşılık gelir. Başka bir deyişle,  $F_T(k)$  sayı kümesi,  $t = rMT$ 'de başlayan ve  $t = rMT + (N-1)T$  ile biten analog zaman fonksiyonunun bir bölümünün ayrık Fourier dönüşümünün bir hesaplamasını temsil eder. Bu sayede  $t = rMT + (N-1)T$  zamanında bir filtre bankası çıkışına karşılık gelir. Spektral analizin bir filtre bankası uygulamasında,  $w(nT)$  zaman



penceresi, filtrelerin her birinin dürtü yanıtının düşük geçişli prototipine karşılık gelir. Bu analizde filtre bankasındaki filtrelerin her birinin spektral şeklinin yaklaşık olarak aynı olduğu kabul edilir.

### 3.8. Veri Seti

Derin öğrenme yöntemlerinde yüksek performans sağlayabilmek için, veri setindeki ses sinyallerinin sayısı çok fazla olmalıdır. Bu çalışmada 64718 konuşma sinyali kullanılmıştır. Bazı konuşma verileri Şekil 3.23'te gösterilmektedir.



Şekil 3.23. Konuşma verisi örnekleri.

Herhangi bir kullanıcıdan gelen ses ihtiyaca göre saklanmalı ve kullanılmalıdır. Bu çalışmada ses sinyalleri veri seti olarak etiketlenmiştir. Tez kapsamında Google konuşma komutları veri seti kullanılmıştır (Andrade, 2018). Veri seti İngilizce komutlardan oluşmaktadır. Çizelge 3.1.'deki gibi verilerde otuz sınıf vardır.

Çizelge 3.1. Konuşma verisi sınıfları

Bed	Five	Marvin	Right	Tree
Bird	Four	Nine	Seven	Two
Cat	Go	No	Sheila	Up
Dog	Happy	Off	Six	Wow
Down	House	On	Stop	Yes
Eight	Left	One	Three	Zero



## 4. BULGULAR

### 4.1. Önerilen Yöntem

Tez kapsamında önerilen yöntemde beş farklı ESA modeli kullanılmıştır. ESA modelleri incelendiğinde birbirileri arasında derinlik ve katman sayısı farklılığı bulunur. Önerilen yöntem kapsamında kullanılan ESA yapılarının özellikleri Çizelge 4-1’de verilmektedir.

Çizelge 4.1. Önerilen yöntem kapsamındaki ESA yapılarının özellikleri

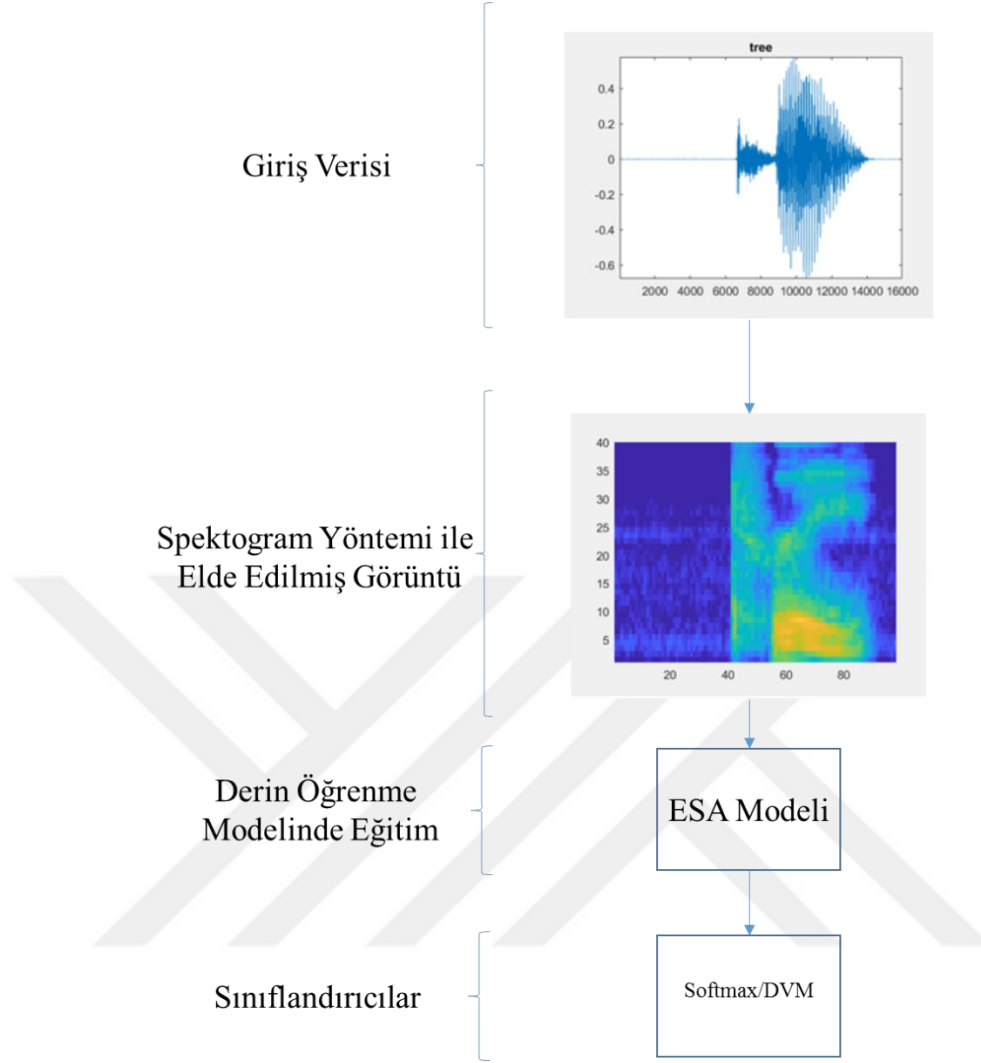
ESA Modelleri	Konvolüsyon Katmanı	Havuzlama Katmanı	Relu Katmanı	Dropout Katmanı	Toplam Katman Sayısı
ESA 1	3x3x12				10
	3x3x12	13x13	2 Adet	-	
ESA 2	3x3x12				16
	3x3x24	3x3 13x13	4 Adet	0.2	
	3x3x36				
ESA 3	3x3x12				20
	3x3x24	3x3 13x13	4 Adet	0.2	
	3x3x36				
	3x3x48				
ESA 4	3x3x24				20
	3x3x48	3x3			
	3x3x48	3x3 13x13	4 Adet	0.2	
	3x3x48				
ESA 5	3x3x24				24
	3x3x48				
	3x3x48	3x3 3x3 3x3	5 Adet	0.2	
	3x3x48	13x13			
	3x3x48				

Çizelge 4.1.'de bulunan ESA yapıları incelendiğinde farklı özelliklere sahip konvolüsyon katmanları kullanılmıştır. Bununla birlikte ESA yapılarının hepsi farklı derinlik ve uzunluğa sahiptir. ESA 1 yapısı içerisinde herhangi bir Dropout katmanı olmayıp 2 adet farklı konvolüsyon katmanı vardır. Bu konvolüsyon katmanları birbirleri ile özdeş olup 3x3 boyutu filtreler ile 12 adet filtre bulundurmaktadır. Havuzlama katmanında ise 13x13 boyutunda havuzlama işlemi yapılmıştır. Kullanılan ReLU aktivasyon fonksiyonunun sayısı da 2 adettir. ESA 2 modelinde 3x3 konvolüsyon filtre boyutuna sahip 3 adet konvolüsyon katmanı bulunmaktadır. Bu katmanlarda sırasıyla 12, 24 ve 36 adet filtre bulunmaktadır. 0.2 indirgeme oranına sahip 1 adet dropout katmanı bulunmaktadır. Bu modelde 2 adet havuzlama katmanı kullanılarak 3x3 ve 13x13 pencere boyutları bulunmaktadır. 4 adet ReLU katmanı olup ESA 2 modeli toplamda 16 katmana sahiptir. ESA 3 modelinde 3x3 konvolüsyon filtre boyutuna sahip 4 adet konvolüsyon katmanı bulunur. Filtre sayıları katman sırasına göre 12, 24, 36, 48'dir. 3x3 ve 13x13 havuzlama pencere boyutuna sahip 2 adet havuzlama katmanı bulunmaktadır. 4 adet ReLU aktivasyon fonksiyonu bulunup 1 adet 0.2 indirgeme oranına sahip 1 adet dropout katmanı modelde mevcuttur. ESA 3 modeli içerisinde toplamda 20 adet katman bulundurur. ESA 4 Modelinde 4 adet konvolüsyon katmanı olup 3x3 konvolüsyon pencere boyutuna sahiptir. Katmanlardaki filtre sayısı dikkate alındığında 24, 48, 48 ve 48 adet filtre bulunur. 3x3, 3x3 ve 13x13 pencere boyutuna sahip 3 adet havuzlama katmanı kullanılmıştır. Modeldeki dropout katmanında 0.2 indirgeme oranı olup 4 adet ReLU aktivasyon fonksiyonu bulunmaktadır. ESA 4 modelinde toplam katman sayısı 20'dir. ESA 5 modeli en derin ve uzun model olup toplamda 3x3 konvolüsyon filtre boyutuna sahip 5 adet konvolüsyon katmanı içinde barındırır. Filtre sayıları sırasıyla 24, 48, 48, 48 ve 48 adettir. Havuzlama katmanlarında 3x3, 3x3, 3x3 ve 13x13 pencere boyutları kullanılır. 5 adet ReLU aktivasyon fonksiyonu ESA 5 modelinde olup dropout katmanında 0.2 indirgeme oranı kullanılmıştır. ESA 5 modelinde toplam 24 adet katman bulunur.

Dünya çapında milyonlarca mühendis ve bilim adamı, tasarladıkları ürünleri analiz etmek ve modellemek için MATLAB'ı kullanmaktadır. Makine öğrenmesi, sinyal işleme, görüntü işleme, bilgisayarlı görme, iletişim, hesaplamalı finans, kontrol tasarımı, robotik ve çok daha fazlası alanda da kullanılmaktadır.

MATLAB’da gerçekleştirilen analizler büyük veri kümeleri ile çalışabilmektedir. MATLAB kodu, web, işletme ve üretim sistemleri gibi alanlarda algoritmalar ve uygulamalar ile bütünleştirilebilir. Önerilen yöntem kapsamında MATLAB programındaki derin öğrenme kütüphanesinden faydalanılmıştır. Önerilen ESA ağlarının tasarlanmasında kütüphanede bulunan bazı hazır katmanlardan yararlanılmıştır.

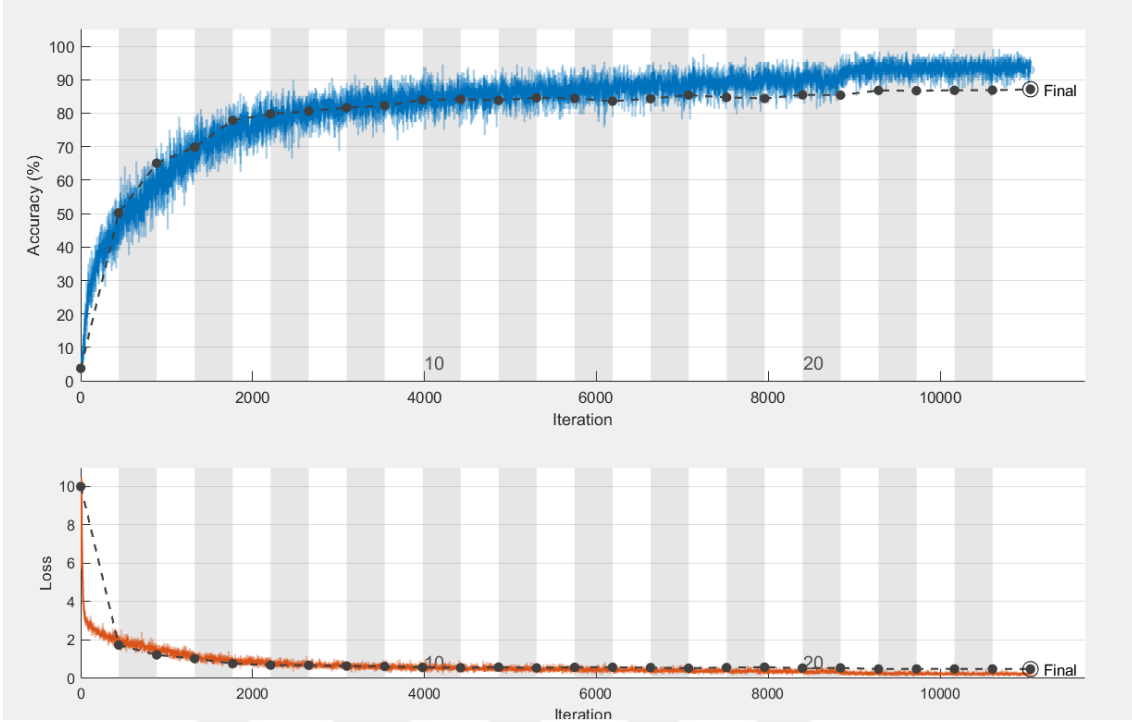
Önerilen yöntemin çalışmasının prensibi Şekil 4.1’deki gibi olup, daha önceden kaydedilmiş ses kayıtları ile çalışılmıştır. Bu ses kayıtlarının ait oldukları sınıflara göre veri seti oluşturulmuş ve oluşturulan bu veri setinin bir kısmı ağın eğitilmesinde bir kısmı ise test edilmesinde kullanılmıştır. Derin öğrenme ağına ses sinyallerinin girdi olarak verilmeden önce spektogram işlemi ile ses sinyalleri 2 boyutlu görüntülere çevrilir. Oluşturulan farklı modellerdeki ağlar eğitim işlemi için hazırlanacaktır. Sonrasında ağın eğitim işlemi gerçekleştirilir. Bu eğitim süreci ağın eğitileceği bilgisayarın performansına bağlı olarak günler veya haftalar sürebilmektedir fakat eğitilmiş olan ağın çalışması oldukça hızlıdır yani sesli komutlara çok hızlı bir şekilde tepki verebilmektedir. Bununla birlikte karşılaştırmalı analiz çerçevesinde Softmax ve (DVM) kullanılarak sınıflama başarısı artırılmaya çalışılmıştır. Yapılan analizler doğruluk ve zaman açısından değerlendirilmiştir. Elde edilen sonuçlar kapsamında test işlemi spektogram dönüşümü haricinde fazladan bir ön işlemeye ihtiyaç duymadan hızlı bir sonuç vermektedir.



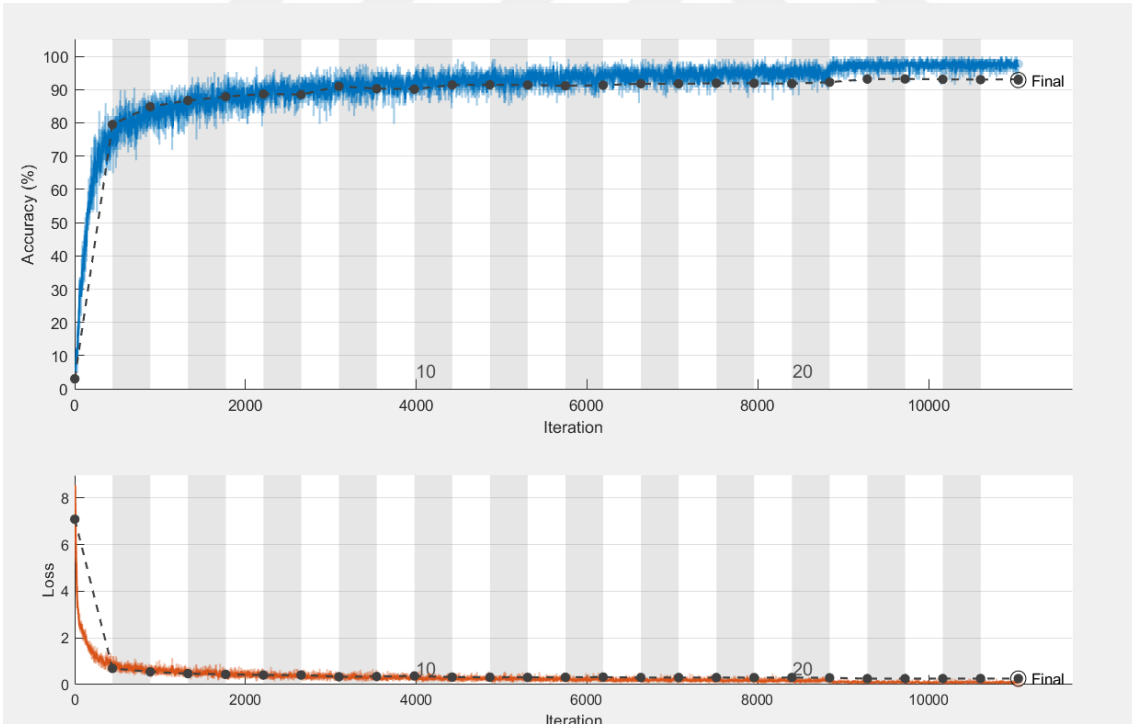
Şekil 4.1 Önerilen Yöntem.

## 4.2. Deneysel Sonuçlar

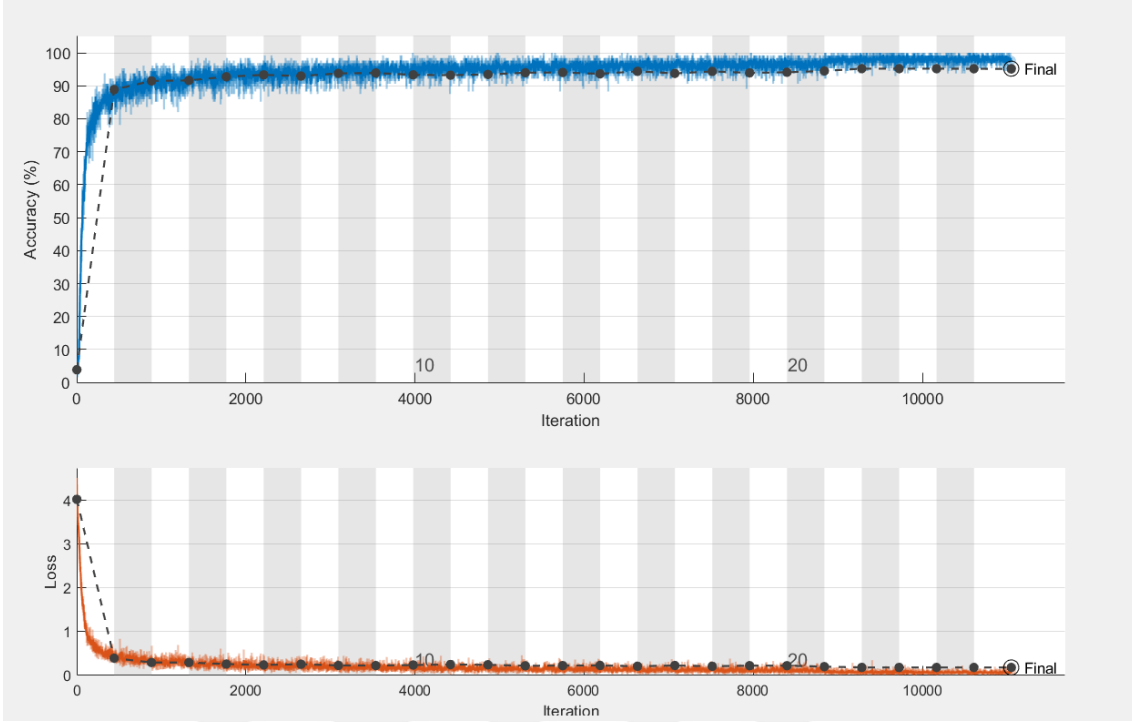
Bu bölümde önerilen yönteme ait 5 farklı ESA modellerinin eğitim ve test analizleri karşılaştırmalı olarak sunulmuştur. Gerçekleştirilen karşılaştırmalı analiz incelendiğinde eğitim süreleri, sınıflandırıcı etkisi, eğitim/ test veri sayılarının etkisi, sınıf sayısı ve ESA modellerine göre farklı sınıflama doğrulukları bulunmaktadır. Bununla birlikte 30 sınıf bulunan veri setinde 8 katlı çapraz doğrulama işlemi yapılarak eğitilen ESA ağlarının genelleme yeteneği analiz edilmiştir. En yüksek doğruluk oranlarının sağlandığı 30 sınıf için ESA 1/2/3/4/5 modellerinin %87.5 eğitim-toplam veri oranına göre eğitim ve test grafikleri Şekil 5.2-5.6 arasında bulunmaktadır.



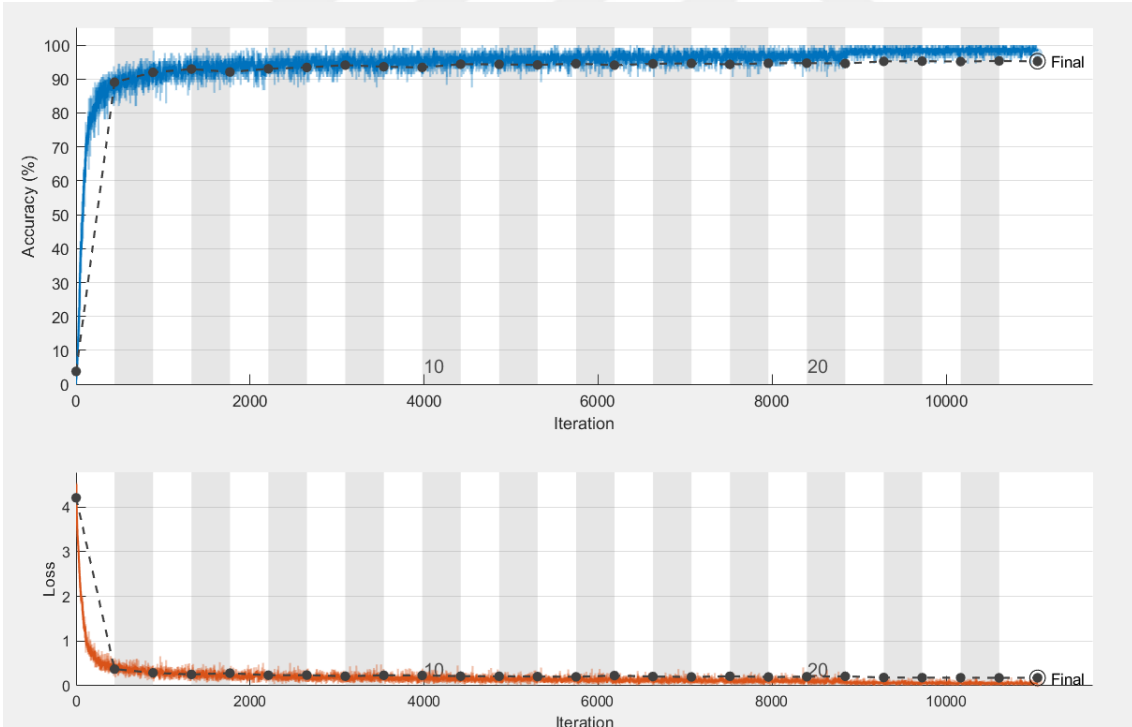
Şekil 4.2. ESA 1 Modelinin eğitim ve test grafiği.



Şekil 4.3. ESA 2 Modelinin eğitim ve test grafiği.

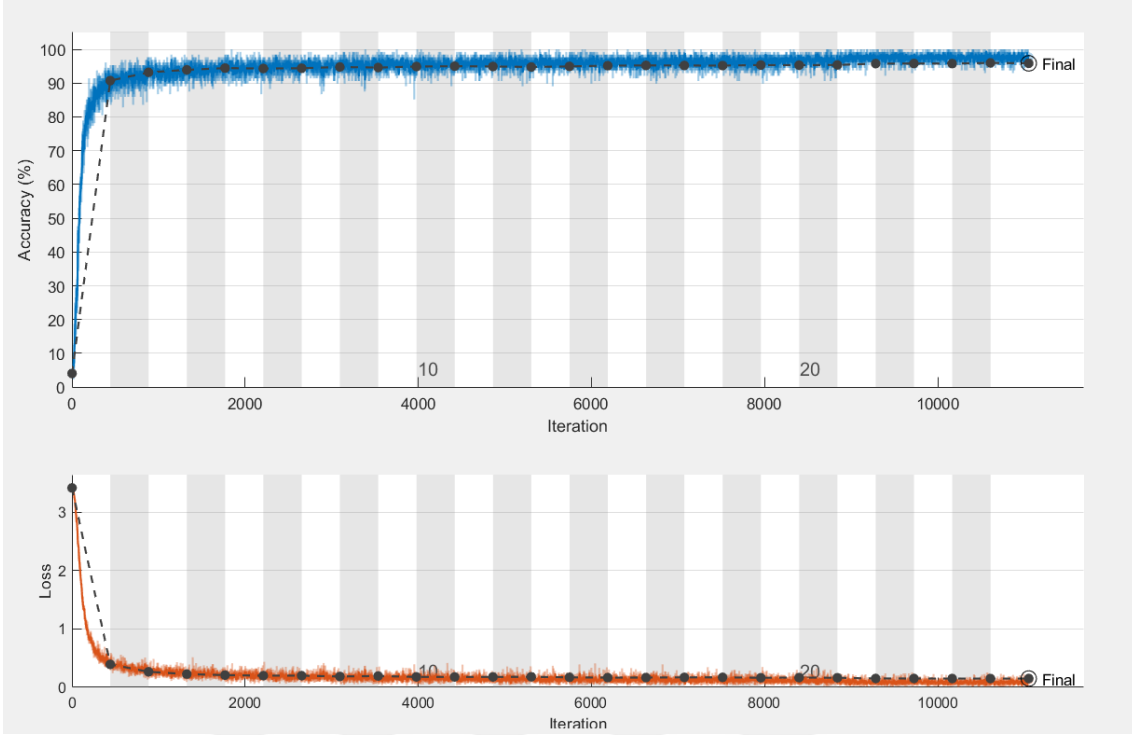


Şekil 4.4. ESA 3 Modelinin eğitim ve test grafiği.



Şekil 4.5. ESA 4 Modelinin eğitim ve test grafiği.





Şekil 4.6. ESA 5 Modelinin eğitim ve test grafiği.

Eğitim grafiğinde eğitim verileri için mavi eğriler sınıflama yüzdesini göstermektedir. Siyah çizgiler ise belirli iterasyonlarda ağı test edilmesiyle elde edilmiştir. Hata grafiğindeki kırmızı çizgiler eğitim verisinin hatasını gösterirken siyah çizgiler test verisinin hatasını göstermektedir. ESA 1'deki aşırı uyum problemini ortadan kaldırmak için diğer ESA modellerinde dropout katmanı kullanılmıştır. Ayrıca ESA modellerinde derinlik ve katman sayısı arttıkça modellerin daha başarılı olduğu görülmektedir. Filtre sayısı ve filtre boyutları ESA ağının başarısını etkileyen ana unsurdur.

Çizelge 4.2. Karşılaştırmalı analiz tablosu

ESA Modelleri	Eğitim/ Toplam Veri (%)	20 Sınıf için Test Doğruluğu (%)		25 Sınıf için Test Doğruluğu (%)		30 Sınıf için Test Doğruluğu (%)		Ortalama Eğitim Süresi
		Softmax	DVM	Softmax	DVM	Softmax	DVM	
ESA 1	50	85.89	83.07	84.65	82.34	85.08	82.55	7 dak. 24 sn
	62.5	86.92	84.94	85.78	83.80	85.14	82.70	
	75	87.40	84.82	86.07	83.61	85.32	82.84	
	87.5	<b>89.11</b>	86.27	87.96	86.22	87.17	84.41	
ESA 2	50	89.18	89.75	91.59	91.64	89.79	90.52	11 dak. 41 sn
	62.5	89.33	89.90	92.08	90.71	91.87	91.00	
	75	91.74	91.26	92.70	90.84	92.06	91.33	
	87.5	92.14	91.93	93.04	92.39	92.87	92.22	
ESA 3	50	93.46	94.25	92.54	94.18	92.81	93.95	11 dak. 50 sn
	62.5	93.86	93.81	93.47	93.60	93.66	93.56	
	75	93.98	93.77	93.62	93.91	93.67	93.64	
	87.5	94.32	94.36	94.18	94.39	95.20	94.96	
ESA 4	50	94.50	94.21	93.55	93.98	92.10	94.46	11 dak.56 sn
	62.5	94.54	94.24	93.66	93.91	93.58	93.82	
	75	94.22	94.42	94.78	94.36	93.79	94.06	
	87.5	94.67	94.76	<b>95.37</b>	94.76	95.25	94.83	
ESA 5	50	95.39	95.56	95.65	95.55	95.25	95.08	12 dak. 10 sn
	62.5	95.42	95.16	95.36	95.23	95.45	95.28	
	75	95.73	95.36	95.22	95.15	95.09	94.91	
	87.5	<b>95.79</b>	95.28	<b>95.80</b>	95.64	<b>95.86</b>	95.76	
<b>Toplam Veri</b>		44241		55436		64718		

Çizelge 4.2. incelendiğinde, en yüksek test doğruluğuna sahip algoritma % 95.86 ile 30 sınıf için ESA 5 + Softmax yapısıdır. ESA 5 yapısı, 24 katmanlı bir sayıyla diğer ESA yapılarından daha derin bir yapıya sahiptir. Eğitim süresi göz önüne alındığında, ESA 5 modeli yaklaşık 12 dakikalık en uzun eğitim süresine sahiptir. Eğitim süreleri, katman sayısına ve iterasyon sayısına bağlı olarak değişmiştir. Aynı zamanda, ESA 5 yapısında dropout katmanları kullanılarak aşırı uyum problemi önlenmiştir. Test doğruluğu açısından, ESA 1 + DVM modeli 25 sınıf için % 82,34 oran ile en düşük doğruluk oranına sahiptir. ESA 1'in toplam 10 katmanı bulunmaktadır. Eğitim süresi 7 dakika 24 sn sürmektedir. Eğitim süresi incelendiğinde, ESA 1 yapısı en düşük eğitim süresine sahip olduğu görülmektedir. 20 sınıf için ESA 1 + Softmax yapısı, % 89.11 sınıflandırma oranıyla kendi içinde en yüksek doğruluk oranına sahiptir. ESA 2-4 yapıları incelendiğinde, yapı derinleştikçe daha fazla eğitim verisine ihtiyaç duyulmaktadır. Bu nedenle, 30 sınıf veri setinde daha fazla veri içerdiğinden dolayı sınıflama doğruluk oranlarının daha yüksek olduğu görülmüştür. ESA 2 + Softmax yapısı 25 sınıf için % 93.04 oranı ile en yüksek doğruluk oranını elde etmiştir. ESA 3+ Softmax yapısı 30 sınıf için % 95,20 oran ile en yüksek doğruluk oranına ulaşmıştır. ESA 4 + Softmax yapısı % 87,5 eğitim veri oranlarını kullanarak 30 sınıf için % 95,25 oran ile en yüksek doğruluk oranına sahiptir. ESA yapıları derinleştikçe daha fazla eğitim verisi olan 30 sınıfa sahip veri setinde daha yüksek sınıflandırma başarısı elde edilmiştir. Daha derin ESA yapılarında hesaplanan parametre sayısını artıracığından, eğitim süresi artar. Genel olarak, Softmax sınıflandırıcısı DVM sınıflandırıcısından daha başarılı olduğu görülmektedir.

Çizelge 4.3. Softmax sınıflandırıcısı için 8 katlı çapraz doğrulama sonuçları

Test Veri Tipi	ESA 1	ESA 2	ESA 3	ESA 4	ESA 5
Test 1	%88.10	%91.82	%95.33	%93.73	%94.33
Test 2	<b>%84.05</b>	%91.90	%93.04	%93.51	%95.03
Test 3	%86.00	%93.36	%93.93	%95.30	%95.29
Test 4	%88.38	%94.73	%97.25	%96.31	<b>%97.68</b>
Test 5	%88.28	%94.39	%95.40	%95.78	%97.09
Test 6	%86.24	%93.43	%94.65	%94.94	%96.32
Test 7	%86.83	%92.84	%94.09	%94.25	%95.17
Test 8	%86.50	%94.15	%95.36	%95.99	%97.38
<b>Ortalama Doğruluk</b>	<b>%86.79</b>	%93.32	%94.72	%94.97	<b>%96.03</b>

Çizelge 4.3.'de 30 sınıflı veri setine 8 kat çapraz doğrulama uygulanmış ve elde edilen ortalama doğrulukları gösterilmiştir. Bu sonuçlar değerlendirildiğinde, Çizelge 4.1'deki test sonuçları ile oldukça yakındır. En iyi ortalama doğruluk % 96.03 ile ESA 5 + Softmax'a modeline aittir. ESA 1 + Softmax modelinde % 86,79 oranı ile en düşük ortalama doğruluk elde edilmiştir. Test 4 verilerinde ESA 5 + Softmax modelinde %96.03 oran ile en yüksek doğruluğa ulaşılmıştır. Test 2 verileri, ESA 1 + Softmax modelinde % 84.05 ile en düşük doğruluk oranına sahiptir.

Çizelge 4.4. Karşılaştırmalı analiz

Model	Doğruluk Yüzdesi	Sınıf Sayısı
DenseNet-121 ön eğitimsiz ve ölçeklenmemiş (McMahan., 2017)	81.32	20
DenseNet-121 UrbanSound8K very seti üzerinde ön eğitilmiş, ölçeklenmemiş (McMahan., 2017)	82.48	20
DenseNet-121 ön eğitimsiz ve ölçekli (McMahan., 2017)	82.22	20
DenseNet- UrbanSound8K very setinde ön eğitilmiş, ve ölçekli (McMahan., 2017)	85.52	20
ConvNet (Warden. 2018)	85.4	20
Attention RNN (Andrade, 2018)	94.1	20
Attention RNN-V2 (Andrade, 2018)	94.5	20
<b>ESA 4</b>	94.67	20
<b>ESA 5</b>	95.79	20

Çizelge 4.4. incelendiğinde DenseNet modelleri yaklaşık % 85.52 ile % 81.32 sınıflama başarısı elde etmiştir. ConvNet'in gösterdiği performans ise %85.4'tir. Attention RNN ve Attention RNN-V2 modelleri ise %94.5'lik başarıya ulaşılmıştır. Önerdiğimiz ESA 4 ve ESA 5 modelleri ise bu oran %94.67 ve 95.79 olarak en üstün performansa sahiptir. Aynı zamanda 8 katlı çapraz doğrulama sonuçlarına bakıldığında önerilen yöntem ESA modelleri bütün veri setini genelleyebilecek kabiliyettedir. ESA 4 ve ESA 5 modellerinin ortalama 8 katlı çapraz doğrulama performansları dikkate alındığında %94.97 ve %96.04'dür. Bu sonuçlar ile mevcut çalışmalardan daha yüksek performans gösteren ESA modelleri tezde yer almaktadır.

## 5. TARTIŞMA VE SONUÇ

Konuşulan dil anlayışı, konuşulan konuşma dili arayüzleri için hayal ettiğimiz performansı elde etmek için karmaşık sinyal işleme, transkripsiyon ve daha üst düzey akıl yürütme sistemleri gerektirir. Gerçekleştirilen çalışmalarda gürültü azaltma, konuşma transkripsiyonu ve duyarlılık analizi gibi görevler hakkında el yapımı varsayımlara sahip karmaşık bireysel sistemler mevcuttur. Bu görevlerin her biri için, belki de yanlış modelleme varsayımlarını ortadan kaldırmak ve performansı iyileştirmek için derin öğrenme yöntemleri uygulanmaya çalışılmıştır.

Bu tez çalışmasında, konuşma tanıma için, genel performansı iyileştiren ESA'ya dayalı bir derin öğrenme yöntemi uygulanmıştır. Bunun için yeterli eğitim verisi mevcut olup gittikçe daha karmaşık fonksiyonlar öğrenilmiştir. Derin öğrenme, büyük miktarda eğitim verisi içeren karmaşık sistemlerin oluşturulması için mükemmel bir test ortamı sağladı. Mobil ve her yerde kullanıma açık bilgi işlem, konuşma arabirimlerinin hızlı bir şekilde benimsenmesi, karmaşık konuşma dili anlama görevleri için sistemlerin eğitimi aşamasında daha fazla veri üretilmesini sağlamaktadır. Bu verileri kullanarak yüksek performansa sahip ESA modellerinin oluşturulması mümkündür. Ayrıca transfer öğrenmesi yöntemi kullanılarak yüksek sınıflama başarısı elde etmek mümkündür. Farklı ses verileri ile ESA eğitilerek problemde asıl olarak kullanılması planlanan verilerde kısa zamanda daha yüksek performans elde edilebilir. Transfer öğrenmesinde ESA modelinin yapısı ve parametreleri problemin çözümü için kullanılır. Bunun yanı sıra daha derin ve fazla parametre içeren ESA yapılarının kullanımı için yüksek özelliklere sahip bilgisayarlar gerekir. Tasarımcılar bu sorunun üzerinden gelmek amacıyla Google firması tarafından sunulan Colab platformunu kullanabilmektedirler.

Önerilen yöntemde:

- ❖ Klasik yöntemlere kıyasla belirli ortamlar için bir gürültü azaltma yaklaşımı yerine, mevcut verileri öğrenerek gürültü sorununu ortadan kaldırmaktadır.
- ❖ ESA, eğitim verilerinde ne tür gürültü ve bozulma varsa, kişiselleştirilmiş bir gürültü azaltma işlevi oluşturur.
- ❖ ESA mimarisinin, konuşma verisi üzerinde klasik SMM fonksiyonu tasarımını bir makine öğrenme yaklaşımıyla değiştirerek kompleks bir model ortaya koymaktadır.

- ❖ Kelime vektörleri, bir dizi akustik özellik ile dil anlayışı için gerekli kelime kimlikleri arasında olası bir köprü sunar. Tüm bir kelimenin akustiğini bir kelime vektör uzayına yansıtan evrimsel bir ağ mimarisine tanıtılmıştır. Yapılan çalışma konuşma alanındaki çok sayıdaki veri üzerinde ve kelime düzeyinde derin öğrenme yaklaşımlarının potansiyelini göstermiştir.
- ❖ Genel olarak, Softmax sınıflandırıcısı DVM sınıflandırıcısından daha başarılı olduğu görülmüştür.
- ❖ ESA yapıları konuşma ve örüntü tanıma problemleri için oldukça yararlı algoritmalarıdır.

Bu tez çalışması, ESA'nın konuşmayı tanımada oldukça yararlı olduğunu ve yüksek tanıma kabiliyetine sahip olduğunu göstermektedir. Derin öğrenme yöntemlerine dayanan önerilen ESA algoritmalarından ESA-5 modeli, önerilen beş farklı modeli içinde en iyi konuşma tanıma oranını vermiştir. ESA-5 modeli diğer modellerden daha derin yapıya sahip olması ve filtre sayısının çokluğu ile daha fazla özellik öğrenebildiği için diğer ESA yapılarına göre performansı daha yüksektir. Aynı zamanda konuşma tanıma birçok kullanım ve uygulama alanına sahiptir ve birçok araştırmacının dikkatini çekmektedir. Bu tez çalışma ile sınıflandırıcıların konuşma tanıma üzerindeki etkilerinin görülebileceği ve Otomatik Konuşma Tanıma üzerine çalışan araştırma gruplarından yararlanabileceği bir doküman olması beklenmektedir.

Daha karmaşık derin öğrenme teknikleri ile daha başarılı sistemler gerçekleştirilebilirler. Ancak bu karmaşık sistemlerin etkili bir şekilde uygulanması için daha güçlü donanım özellikli bilgisayar sistemlerine ihtiyaç vardır.

## KAYNAKLAR

- Abushariah A.A.M., Gunawan T.S., Khalifa O.O. 2010. English digits speech based on Hidden Markov models. *International Conference on Computer and Communication Engineering (ICCCE 2010)*, 11-12 Mayıs 2010.
- Ackley D. H., Hinton G. E., and Sejnowski T. J., 1985. A learning algorithm for boltzmann machines. *Cognitive science*, 9(1):147–169.
- Alshuieili H., Gupta G.S., Mukhopadhyay S., 2011. Voice Recognition Based Wireless Home Automation System, 2011. **4th International conference on mechatronics (ICOM)**.
- 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.
- Asyalı M. H., Yılmaz M., Tokmakçı M., Sedef K., Aksebzeci B.H., Mittal R., 2011. Design and Implementation of a Voice Controlled Prosthetic Hand. *Turk J. Elec. Eng. and Comp. Sci.*, doi: 10.3906/elk-0912-315.
- Aşlıyan R., 2008. *Dinamik Zaman Bükmesi Yöntemiyle Hece Tabanlı Konuşma Tanıma Sistemi*. Yüksek Lisans Tezi, Çanakkale On sekiz Mart Üniversitesi Fen Bilimleri Enstitüsü, Çanakkale.
- Babui G., Kumar H., Vanathi P.T. 2012. Performance analysis of hybrid robust Automatic Speech Recognition System. *IEEE International Conference on Signal Processing, Computing and Control*, 15-17 Mart 2012.
- Bala A., Kumar A. A., Birla N., 2010. Voice command recognition system based on MFCC and DTW. *International Journal Of Engineering Science and Technology*, (12):7335-7342.
- Baygın M., Karaköse M., 2012. *Gerçek Zamanlı Ses Tanıma Tabanlı Akıllı Ev Uygulaması*. IEEE 978-4673-0056.
- Baygün M.K., Yaldır A.K., 2009. *Linear Predictive Coding ve Dynamic Time Warping-Teknikleri Kullanılarak Ses Tanıma Sistemi Geliştirilmesi*. Pamukkale Üniversitesi, 2009.
- Bengio I., Courville A., 2016. *Deep Learning*, The MIT Press.
- Bengio Y., Courville A., Vincent P., 2013. Representation learning: A review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35 (8):1798–1828.
- Bolat B., Küçük Ü., Yıldırım T. 2004, Aktif Öğrenen PNN ile Konuşma/Müzik Sınıflandırma. *Akıllı Sistemlerde Yenilikler ve Uygulamaları Sempozyumu, ASYU-INISTA*, 2004.
- Çelebi M., 2012. *Ses Komut Tanıma ile Gezgin Araç Kontrolü*. Yüksek Lisans Tezi, Marmara Üniversitesi Fen Bilimleri Enstitüsü.
- Dede G., Sazlı M.H., 2010. *Biyometrik Sistemlerin Örüntü Tanıma Perspektifinden İncelenmesi ve Ses Tanıma Modülü Simülasyonu*. Savunma Bilimleri Enstitüsü.
- Deniz Demirci M., 2005. *Bilgisayar Destekli Ses Tanıma Sistemi Tasarımı*. Yüksek Lisans Tezi, İstanbul Üniversitesi, İstanbul.
- Duda R. O., Hart P. E., and Stork D. G., 2012. *Pattern Classification*. John Wiley & Sons. 2012.
- Edizkan R., 2007. *Ses Komut Tanıma ile Gezgin Araç Kontrolü*, Yüksek Lisans Tezi, Dumlupınar Üniversitesi, Kütahya.

- Elman J. L., 1990. Finding structure in time. *Cognitive science*, **14**(2):179–211.
- Ferrando F., Nouveau G., Philip B., Pradeilles P., Soulenq V., Stean G.V. 2009, Courmontagne P., 2009. *A Voice Recognition System for a Submarine Piloting*. IEEE 1-4244-2523-5.
- Fezari M., Salah M.B. 2016, A Voice command system for autonomous robots guidance. **9th IEEE International Workshop on Advanced Motion Control**, 27-29 Mart 2016.
- Hatipoğlu P. U., 2016. *Time Series Classification Using Deep Learning*. Yüksek Lisans Tezi, ODTÜ, Fen Bilimleri Enstitüsü. Ankara
- Haykin S. and Network N., 2004. A comprehensive foundation. *Neural Networks*, 2.
- He Q., Zhang Y. 2001. A Speech Recognition and Speech Corpus System Based on Matlab. *International Symposium on Intelligent Multimedia, Video and Speech Processing*, 6-7 May 2001.
- Hopfield J. J., 1982. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences*, **79**(8):2554–2558,
- Hrncar M., 2000. *Voice Command Control for Mobile Robots*. Department of Control and Information Systems Faculty of Electrical Engineering University of Zilina. [http://download.tensorflow.org/data/speech\\_commands\\_v0.01.tar.gz](http://download.tensorflow.org/data/speech_commands_v0.01.tar.gz)
- Jaiganesh V., Mangayarkarasi, S., Sumathi, P., 2013. Intrusion Detection Systems: A Survey and Analysis of Classification Techniques. *International Journal of Advanced Research in Computer and Communication Engineering*.
- Jiang Z., Huang H., Yang S., Lu S., Hao Z., 2009. Acoustic feature comparison of MFCC and CZT based cepstrum for speech recognition. *Fifth International Conference on Natural Computation*, 13-14 March 2009.
- Karacı A., 2006. Bilgisayar Ortamında Sesli İfadeleri Tanıma. *Kastamonu Eğitim Dergisi*. **15**(1):45-62.
- Karakaş M., 2010. *Computer Based Control Using Voice Input*. Yüksek Lisans Tezi, Dokuz Eylül Üniversitesi Fen Bilimleri Enstitüsü, İzmir.
- Krenker A., Kos A., and Bešter J., 2011. *Introduction to the artificial neural networks*. INTECH Open Access Publisher
- Leechor P., Pornpanomchai C., Sukklay P., 2010. Operation of a Radio Controlled Car by Voice Commands. **2nd International Conference on Mechanical and Electronics Engineering**, 1-3 August 2010.
- Masud M., Khan, L., Thuraisingham, B., 2011. *Data Mining Tools for Malware Detection*, CRC Press. Taylor & Francis, London.
- McCulloch W. S. and Pitts W., 1943. A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, **5**(4):115–133.
- McMahan B., Rao D., 2017. *Listening to The World Improves Speech Command recognition*. CoRR abs/1710.08377.
- Meral O., 2008. *Doğrusal Öngörülü Kodlama ve Adaptif Algoritma Tabanlı Konuşmacı Tanıma*. Yüksek Lisans Tezi, İstanbul Üniversitesi Fen Bilimleri Enstitüsü, İstanbul.
- Muda L., Begam M., Elamvazuthi I., 2010. Voice recognition using mel frequency cepstral coefficient and dynamic time warping techniques. *Journal of Computing*, **2**, ISSN 2151-9617.
- Nair V. and Hinton G.E. 2010, Rectified linear units improve restricted boltzmann machines. **Proceedings of the 27th International Conference on International**



- Conference on Machine Learning**, ICML'10, Omnipress, USA, 807–814, <http://dl.acm.org/citation.cfm?id=3104322:3104425>.
- Nielsen, M. A., 2015. **Neural networks and deep learning 25**. USA: Determination press.
- Ocal K., 2005. *Application of Automatic Speech Recognition Algorithms*. Ms. Thesis, Özdemircan M. Z., 2008. *Robot Control With Voice Command*. Bitirme Projesi, Yıldız Teknik Üniversitesi Bilgisayar Mühendisliği Bölümü.
- Öztürk B., Çakar T., 2007. *Gerçek Zamanlı Ses Tanıma*. Bitirme Projesi, İstanbul Üniversitesi Mühendislik Fakültesi Elektrik/Elektronik Mühendisliği Bölümü, İstanbul.
- Öztürk N., Ünözkan U., 2010. *Microprocessor Based Voice Recognition System Realization*. IEEE, 978-1-4244-6904-8/10.
- Phoophuangpairaj R. 2011. Using Multiple HMM Recognizers and the Maximum Accuracy Method to Improve Voice Controlled Robots. **International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS)**, 7-9 December 2011.
- Price J., Eydgahi A., 2006. Design of Matlab Based Automatic Speaker Recognition Systems. **9th International Conference on Engineering Education**, 23-28 July 2006.
- Shaneh M., Taheri A., 2009. Voice Command Recognition System Based on MFCC and VQ Algorithms, **World Academy of Science, Engineering and Technology**, 2009.
- Wakuya H. and Shida K. 2001, Bi-directionalization of neural computing architecture for time series prediction. iii. application to laser intensity time record “data set a. In Neural Networks, 2001. Proceedings. **IJCNN'01. International Joint Conference on**, 3, 2098–2103. IEEE, 2001.
- Warden P.. 2018. *Speech commands: A dataset for Limited-Vocabulary Speech Recognition*. CoRR abs/1804.03209.
- Watt J., Borhani, R., Katsaggelos, A.K., 2016. *Machine Learning Refined: Foundations, Algorithms, And Applications*. Cambridge University Press, New York.
- Webders Sitesi <http://webders.net/solunum-sistemleri-ders-20-368p2.html>. Erişim tarihi: 26.05.2019.
- Yalçın N., 2008. *Speech Recognition Theory and Techniques*. **Kastamonu Eğitim Dergisi**, 16, (1), 249-266.
- Zhizeng L., Jinghing Z., 2004. Speech Recognition and Its Application in Voice-based Robot Control System. *Proceedings Of International Conference On Intelligent Mechatronics*.
- Zurada J. M.. 1992. *Introduction to artificial neural systems*, 8. West St. Paul.



## ÖZ GEÇMİŞ

Ferdi KOÇ 1984 yılı, Hakkâri ili merkez ilçesi doğumlu olup. İlk, Orta ve Liseyi yine aynı merkez ilçede tamamladıktan sonra, Afyon Kocatepe Üniversitesi, Teknik Eğitim Fakültesi Elektrik Öğretmenliği Bölümüne kaydolmuştur. Bu bölümden 2009 yılında mezun olduktan sonra, askerlik görevini, 2010 yılında Erzurum'da tamamlamıştır. 2015 yılından itibaren, Erzincan Binali Yıldırım Üniversitesi'nde Öğretim görevlisi görevini sürdürmektedir. 2015 yılında da, Van Yüzüncü Yıl Üniversitesi, Elektrik Elektronik Anabilim Dalında yüksek lisans eğitimine başlamıştır.



T.C  
VAN YÜZÜNCÜ YIL ÜNİVERSİTESİ  
FEN BİLİMLERİ ENSTİTÜSÜ  
LİSANSÜSTÜ TEZ ORJİNALLİK RAPORU

Tarih: 22/07/2019

Tez Başlığı / Konusu: Spektrogram Tekniği Kullanılarak Derin Öğrenme Yöntemleri İle Ses Tanıma

Yukarıda başlığı/konusu belirlenen tez çalışmamın Kapak sayfası, Giriş, Ana bölümler ve Sonuç bölümlerinden oluşan toplam 14 sayfalık kısmına ilişkin, 22/07/2019 tarihinde şahsım/tez danışmanım tarafından Turnitin intihal tespit programından aşağıda belirtilen filtreleme uygulanarak alınmış olan orijinallik raporuna göre, tezimin benzerlik oranı % 3 (üç) dür.

Uygulanan filtreler aşağıda verilmiştir:

- Kabul ve onay sayfası hariç,
- Teşekkür hariç,
- İçindekiler hariç,
- Simge ve kısaltmalar hariç,
- Gereç ve yöntemler hariç,
- Kaynakça hariç,
- Alıntılar hariç,
- Tezden çıkan yayınlar hariç,
- 7 kelimededen daha az örtüşme içeren metin kısımları hariç (Limit inatch size to 7 words)

Van Yüzüncü Yıl Üniversitesi Lisansüstü Tez Orijinallik Raporu Alınması ve Kullanılmasına İlişkin Yönergeyi inceledim ve bu yönergede belirtilen azami benzerlik oranlarına göre tez çalışmamın herhangi bir intihal içermediğini; aksinin tespit edileceği muhtemel durumda doğabilecek her türlü hukuki sorumluluğu kabul ettiğimi ve yukarıda vermiş olduğum bilgilerin doğru olduğunu beyan ederim.

Gereğini bilgilerinize arz ederim.

  
22/07/2019

Adı Soyadı: Ferdi KOÇ

Öğrenci No: 159101016

Anabilim Dalı: Elektrik Elektronik Mühendisliği ABD

Programı: .....

Statüsü: Y. Lisans  Doktora

DANIŞMAN ONAYI  
UYGUNDUR

  
Doç. Dr. Ridvan SARAÇOĞLU

ENSTİTÜ ONAYI  
UYGUNDUR

(Unvan, Ad Soyad, İmza)

  
Prof. Dr. Suat GENÇY  
Enstitü Müdürü