

**T.C.
YILDIZ TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

**DERİN ÖĞRENME YÖNTEMLERİ VE UYGULAMALARI HAKKINDA BİR
İNCELEME**

ABDULKADİR ŞEKER

**DOKTORA SEMİNERİ
BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI**

**DANIŞMAN
PROF. DR. HASAN HÜSEYİN BALIK**

İSTANBUL, 2017

İÇİNDEKİLER

Sayfa

İÇİNDEKİLER	ii
BÖLÜM 1	1
GİRİŞ.....	1
1.1 Tanım.....	1
1.2 Seminerin Amacı	2
BÖLÜM 2	3
YAPAY SİNİR AĞLARI.....	3
2.1 Tarihçe.....	3
2.2 Uygulamalar.....	5
BÖLÜM 3	8
DERİN ÖĞRENME	8
3.1 Tanım.....	8
3.2 Tarihçe.....	8
3.3 Derin Öğrenme Mimarileri.....	11
3.3.1 Konvolüsyonel Sinir Ağları	11
3.3.2 Tekrarlayan Sinir Ağı	12
3.3.3 Uzun Kısa Vadeli Memory	14
3.3.4 Kısıtlanmış(Derin) Boltzmann Makineleri	16
3.3.5 Derin İnanç Ağları	16
3.3.6 Derin Oto-kodlayıcılar	17
3.4 Derin Öğrenme Uygulamaları.....	19

3.4.1	Doğal Dil İşleme.....	19
3.4.2	Görüntü/Video İşleme	20
3.4.3	Biyomedikal Görüntü/Bilgi İşleme.....	21
3.4.4	Diğer Uygulamalar	22
3.5	Derin Öğrenme Çalışan Araştırma Ekipleri.....	23
3.6	Derin Öğrenme Kütüphaneleri/Yazılımları.....	25
3.6.1	Theano	25
3.6.2	Caffe	25
3.6.3	TensorFlow	26
3.6.4	Torch.....	26
3.6.5	DeepLearning4j	26
3.6.6	Digits	27
BÖLÜM 4	28
	TARTIŞMA VE POTANSİYEL ÇALIŞMA ALANLARI.....	28
	KAYNAKÇA.....	29
	ÖZGEÇMİŞ	40

KISALTMA LİSTESİ

GPU	Grafik İşlemci Birimi
YSA	Yapay Sinir Ağı
SVM	Destek Vektör Makinesi
CNN	Konvolüsyonel Sinir Ağı
SRN	Basit Tekrarlayan Ağ
RNN	Tekrarlayan Sinir Ağı
LSTM	Kısa Uzun Vadeli Hafıza Ağı
RBM	Kısıtlandırılmış Boltzmann Makinesi
DBN	Derin İnanç Ağı
AE	Oto-kodlayıcı
DAE	Derin/Yığılanmış Oto-kodlayıcı

ŞEKİL LİSTESİ

Şekil 1. Yapay Zekanın Süreçleri [1].....	1
Şekil 2 Sinir hücresi yapısı [12].....	3
Şekil 3 Yapay sinir ağı yapısı.....	4
Şekil 4 Ivankhnenko A. tarafından eğitilen bilinen ilk derin ağ mimarisi [54]	9
Şekil 5 Yıllara göre yapay zeka startup'ların satın alınması [67].....	10
Şekil 6 Maksimum havuzlama için bir örnek [71]	11
Şekil 7 LeNet ağıının mimarisi.....	12
Şekil 8 SRN ağ mimarisi [86]	13
Şekil 9 Basit bir RNN mimarisi [87]	13
Şekil 10 Görüntü tanımlayıcıları oluşturmak için derin görsel-anlamsal hizalama [89].....	14
Şekil 11 RNN'de kullanılan SRN birimi (solda) ve LSTM (sağdaki) ayrıntılı şeması [91].....	15
Şekil 12 Boltzmann Makineleri (solda) ve Kısıtlanmış Boltzmann Makineleri (sağda) [105].....	16
Şekil 13 DBN ve RBM mimarilerinin karşılaştırılması [111].....	17
Şekil 14 Basit AE mimarisi [121]	18
Şekil 15 Genel bir DAE mimarisi [123].....	18

ÇİZELGE LİSTESİ

Tablo 1. Derin Öğrenme çalışan ekipler hakkında bilgiler	25
Tablo 2 Derin Öğrenme Kütüphaneleri.....	25

DERİN ÖĞRENME YÖNTEMLERİ VE UYGULAMALARI HAKKINDA BİR İNCELEME

Abdulkadir ŞEKER

Bilgisayar Mühendisliği Anabilim Dalı
Doktora Semineri

Danışmanı: Prof. Dr. Hasan Hüseyin BALIK

Makine öğrenmesi alanında yapay sinir ağları birçok problemin çözümünde sıklıkla kullanılmaktaydı. Ancak “Yapay Zeka Kışı” (AI winter) olarak da adlandırılan dönemde başta donanımsal kısıtlamalar ve diğer problemler sebebiyle bu alandaki çalışmalar durma noktasına gelmiştir. 2000’lerin başında tekrar gözde bir alan olmaya başlayan yapay sinir ağları, GPU gelişmeleriyle birlikte sığ ağlardan derin ağlara geçiş yapmıştır. Derin öğrenme, her katmanda geri yayılım algoritmasını kullanarak büyük veri kümelerindeki gizli ve ayırt edici özelliklerini keşfetmektedir. Bu yaklaşım görüntü işlemeden, doğal dil işlemeye, medikal uygulamalardan aktivite tanımaya kadar oldukça geniş bir yelpazede başarıyla kullanılmaya başlanmıştır. Bu çalışmada öncelikle yapay sinir ağları ve uygulamaları kısaca anlatılmıştır. Ardından derin öğrenme konusu incelenmiştir. Bu bölümde, derin öğrenmenin tarihçesi, kullanılan yöntemleri ve uygulama alanlarına göre ayrılmış çalışmaları anlatılmıştır. Ayrıca son yıllarda kullanılan kütüphaneler ve derin öğrenme üzerine yoğunlaşan çalışma grupları hakkında da bilgiler verilmiştir.

Anahtar Kelimeler: Derin öğrenme, özellik çıkarma, yapay sinir ağları

A REVIEW ON DEEP LEARNING METHODS AND APPLICATIONS

Abdulkadir ŞEKER

Department of Computer Engineering

PhD Seminar

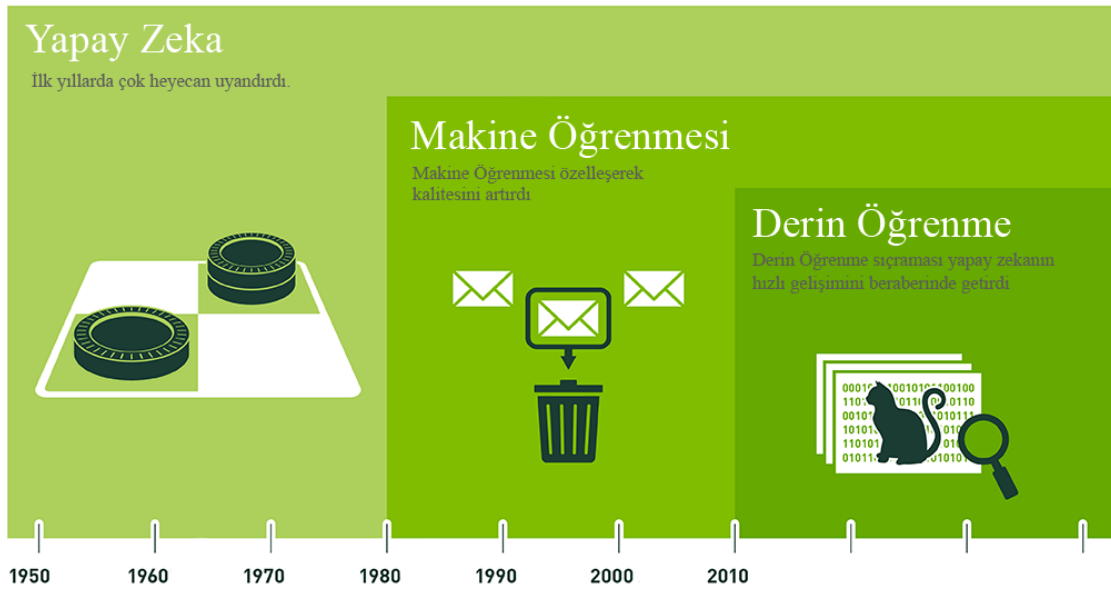
Adviser: Prof. Dr. Hasan Hüseyin BALIK

Artificial neural networks were used in the solution of many problems in the field of machine learning. However, in the period called "AI Winter", studies in this area have come to a halt due to especially hardware limitations and other problem. Artificial neural networks, which started to become a popular area at the beginning of 2000s, have switched from shallow networks to deep networks thanks to GPU developments. Deep learning discovers the hidden and distinctive features of large data sets using the backpropagation algorithm in each layer. This approach has been successfully used in a wide range of fields from image processing to natural language processing, from medical applications to activity identification. In this study, artificial neural networks and applications are briefly explained at first. After that, deep learning is surveyed. It is described history of the deep learning, methods and the implementations separated by the application areas in this section. In addition, information has been given about the libraries used in recent years and working groups focused on deep learning.

Keywords: Deep learning, feature extracting, neural networks

1.1 Tanım

Doğal dil işleme, makine öğrenmesi alanlarında derin öğrenme yaklaşımları sayesinde 2000’li yıllarda büyük sekteye uğrayan yapay zekaya dayalı yöntemler tekrar ön plana çıkmıştır (Şekil 1).



Şekil 1. Yapay Zekanın Süreçleri [1]

Yapay sinir ağlarındaki gizli katman ve düğüm sayılarının artırılmasına karşın donanımsal gelişmelerin yetersiz kalması sebebiyle yapay zeka yöntemleri 2000’li yılların başında artık kullanılmamaktaydı. Ancak GPU ve diğer donanımsal gelişmeler sayesinde çok sayıda gizli katmandan oluşan yapay sinir ağları hesaplama maliyetleri düştüğünden, tekrar kullanılmaya başlanmıştır [2].

Derin öğrenme yaklaşımı çoklu soyutlama yapısı ile verinin temsillerini öğrenmek için bir araya getirilmiş çoklu işleme katmanlarında oluşur [3]. Makine öğrenmesi yaklaşımlarının kalitesi doğru özelliklerin seçimine bağlıdır [4]. Bu nitelikleri bulmak için ön işlemler, boyut indirgeme ve özellik seçme işlemleri yapılmaktadır. Özelliklere olan bu bağımlılıktan kurtulmak gerekmektedir. Sınıflandırıcılar ve diğer tahmin

sistemleri tasarlarlarken, yapay zeka ile veriden faydalı bilgi çıkarmanın daha kolay olabilir [5]. Bu bağlamda derin öğrenme konusu temsili öğrenme ile çok sıkı ilişkilidir. Derin öğrenme algoritmaları özellikle 2005 yıllarından sonra birçok farklı alana uygulanmıştır. Bu çalışmalara ayrıntılarıyla 3. Bölümde anlatılacaktır.

1.2 Seminerin Amacı

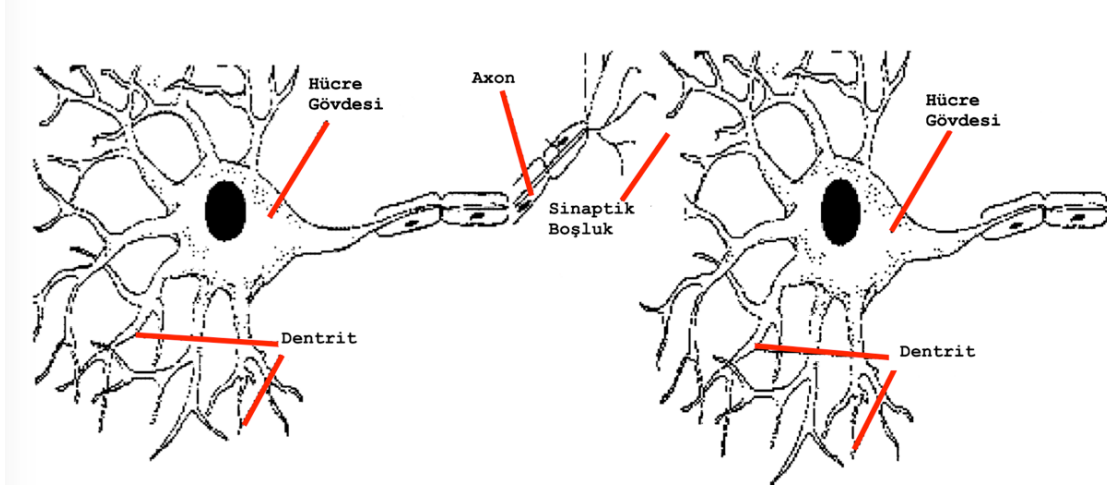
Bu çalışma doktora seminer dersi kapsamında yapılmaktadır. Çalışmanın amacı son yıllarda oldukça popüler bir makine öğrenmesi yaklaşımı olan derin öğrenme kavramını her yönüyle araştırmaktır. 1. Bölümde genel hatlarıyla literatür anlatılacaktır. 2. Bölümde yapay derin öğrenmenin temeli olan yapay sinir ağları konusuna değinilecek, tarihçesi ve uygulama alanları anlatılacaktır. 3. Bölümde yaygın kullanıma sahip derin öğrenme algoritmaları, öne çıkan uygulama alanları, derin öğrenme hakkında araştırmalar yapan ulusal ve uluslararası çalışma grupları, derin öğrenme için kullanılabilecek olan kütüphane ve çerçeveler (framework) anlatılacaktır. 4. Bölümde ise konu ile ilgili genel bir sonuç verilerek, alandaki çalışmaların ne yöne doğru gittiği ve muhtemel çalışma alanları hakkında bilgi sunulacaktır.

Çalışmanın asıl amacı derin öğrenme konusunda bilgi edinmek isteyen veya daha önce başka yöntemlerle ele aldıkları konuları bu alana uygulayacak araştırmacılar için bir vizyon kazandırmaktır. Derin öğrenme algoritmaları çok geniş bir yelpazedeki problemlere uygulanmakta, çoğunda alanında en iyi başarıları elde etmektedir. Bu bağlamda çalışmanın bir başka hedefi de hangi problemlere derin öğrenme yaklaşımlarının uygulanabileceği konusunda fikir vermek olacaktır.

Bu alanda farklı konular üzerine yapılmış tez çalışmaları da bulunmaktadır [6][7][8]. Özellikle derin öğrenmenin doğal dil işleme [9], görüntü ve sinyal işleme konularında başarılı sonuçlar verdiği çalışmalar mevcuttur [10][11].

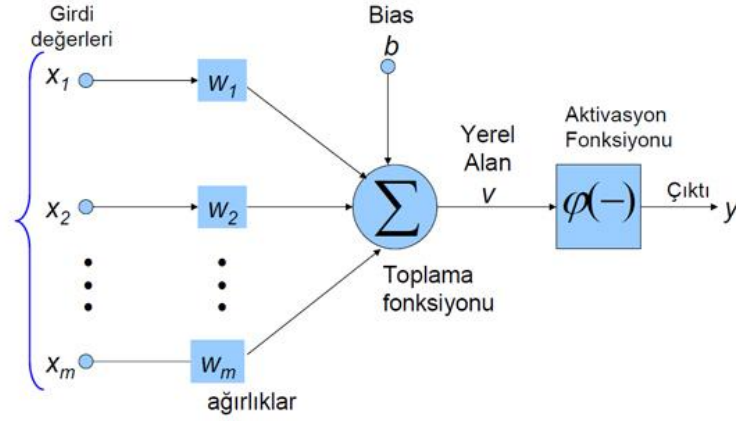
YAPAY SİNİR AĞLARI**2.1 Tarihçe**

Yapay sinir ağı, insan beyninin öğrenme tekniklerinden yola çıkılarak modellenmiş bir sistemdir. Bu yüzden önce kısaca biyolojik sinir hücrelerinden bahsetmek gerekmektedir. Şekil 2’de sinir hücresinin basit bir modeli verilmiştir [12]. Hücreyi inceleyecek olursak; sinyaller hücrelerden gelen sinyalleri çekirdeğe ileten kısma dentrid (dendrite) adı verilir. Gelen sinyalleri toplayan yapıya soma (soma), toplanan bu bilgiyi diğer hücelere dağıtan yapıya akson (axon) denir. Aksonlardan gelen bilgi sinapsis (synaptic) ile ön işlemden geçirilerek diğer hücrelerin dentridlerine iletilir. Bu aşamada sinyaller eşik fonksiyonları kullanılarak iletilir [13]. Öğrenmenin sinapsislerde gerçekleştiği teorisi günümüzde halen geçerliliğini korumaktadır [14].



Şekil 2 Sinir hücresi yapısı [12]

Biyolojik sinirlerden, yapay sinir ağına geçerken elemanlar; nöron-işlem, dentrit-toplama fonksiyonu, hücre gövdesi-aktivasyon fonksiyonu, akson-çıkış, sinapsis-ağırlık şeklinde dönüşüm geçirecektir (Şekil 3).



Şekil 3 Yapay sinir ağı yapısı

Yapay sinir ağıları, insan beyninin öğrenme tekniklerin yola çıkılarak modellenmiş bir sistemdir. İlk olarak, Warren McCulloch ve Walter Pitts (1943), eşik mantığı adı verilen matematik ve algoritmalara dayalı sinir ağıları için hesaplama bir model oluşturulmuştur. Bu model, beyindeki biyolojik süreçlere ve yapay zeka sinir ağlarının uygulanmasına odaklanan bir yaklaşım ortaya atmıştır [15]. Bunun ardından 1940'ların sonlarında psikolog Donald Hebb, Hebbian öğrenme olarak bilinen sinirsel plastisite mekanizmasına dayalı bir öğrenme hipotezi oluşturmuştur. Çalışmalarında sinir hücresini kullanan Hebb iki sinir hücresi arasındaki etkileşimleri araştırmış, ortaya attığı kurallar sinir ağlarının temelini oluşturmuştur. Bu öğrenme, 'tipik' denetimsiz öğrenme (unsupervised learning) kuralı olarak kabul edilmiştir [16]. Bunun ardından ilk çalışma olarak, Farley ve Wesley A. Clark (1954) tarafından MIT'de bir Hebbian ağını taklit etmek için "hesaplayıcılar" olarak adlandırılan hesaplamalı makineleri kullanılmıştır [17]. 1958 yılında Frank Rosenblatt tarafından desen tanıma için, basit toplama/çıkarma işlemlerini kullanarak iki katmanlı bir bilgisayar öğrenme ağına dayalı bir algoritma olan "perceptron" önerilmiştir [18]. Bu alandaki araştırmalar, sinir ağlarının XOR (exclusive-or) problemlerini çözemediği ve büyük sinir ağları için yeterli bilgisayar donanım kapasitesi olmadığını ortaya çıkaran bir çalışmadan sonra donanımsal işleme güçleri artana kadar uzun süre yavaşlamıştır [19]. 1975 yılında Werbos tarafından geliştirilen geri yayılım algoritması XOR problemini etkin bir şekilde çözebilmiştir [20]. 1980'lerde yazılan bir kitap ile sinir ağlarının işlenmesinde paralel dağıtık sistemlerin kullanılmasını tam olarak açıklanabilmiştir [21]. 1982'de Caltech'ten John Hopfield, ulusal Bilimler Akademisine sunduğu bildiride, sadece beyinleri modellemek değil faydalı cihazlar yaratmak üzerine bir yaklaşımı anlatmıştır. Açıklık ve matematik analizi ile yapay sinir ağların nasıl çalıştığını ve neler yapabileceklerini göstermiştir [22]. 1986'da Rumelhart and McClelland tarafından yazılan Paralel Dağıtık İşleme (cilt I ve II) kitabı [23] ile bu alanda tam bir patlama gerçekleşmiştir. 1988'de alandaki ilk dergi olan "Neural Network" dergisi Uluslararası Sinir Ağları Birliği (INNS) tarafından kurulmuştur. Bu yıllardan sonra sinir ağlarının geliştirilmesine odaklanan araştırma nispeten yavaşlamıştır. İşlemcilerin kısıtlamaları nedeniyle sinir ağları öğrenme işlemi haftalar sürmektedir [24]. Bu sebeple 2000 yıllara kadar, donanımsal gelişmeler beklenmiş, sinir ağlarında kış dönemi yaşanmıştır (Şekil 1). Donanımsal gelişmelerin

ardından sığ (shallow) sinir ağlarının yerini derin ağlar almış, tekrar büyük gelişmeler ile uygulanmaya başlanmıştır.

2.2 Uygulamalar

Yapay sinir ağları regresyon [25], sınıflandırma [26], kümeleme [27], robotik [28], kontrol sistemleri [29] gibi bir çok farklı problemin çözümünde kullanılmaktadır.

Uygulama alanları arasında; sistem tanımlama ve kontrol (araç kontrolü, yörünge tahmini, proses kontrolü, doğal kaynak yönetimi) [30]–[33], kuantum kimyası [34], [35], oyun oynama ve karar verme (tavla, satranç, poker), model tanıma (radar sistemleri, yüz tanımlama) [36]–[38], veri madenciliği, görselleştirme ve e-posta spam filtrelemesi [39]–[41] gibi konular bulunmaktadır. Bunların yanında bazı kanser türlerini teşhis etmek için de yapay sinir ağları kullanılmıştır [42]–[45].

Araç kontrolü için yapılan ALVINN (Autonomous Land Vehicle In a Neural Network) isimli bir proje kapsamında gerçek bir sürücünün reaksiyonları ile eğitilen geri yayımlı bir sinir ağı ile, 20 km/s hızla otomatik araç kullanımı başarılmıştır [46]. Bir başka çalışmada, gemilerin gelecekteki davranışlarını (konumu, hızı ve rotayı) tahmin edebilen yapay sinir ağı geliştirilmiştir. Geliştirilen modelin, gemi güzergah planlamasına, yoğun liman tarifelemelerine, hata tespitlerine yardımcı olması amaçlanmıştır [47]. Trafik akışının kontrolü için yapılan bir çalışmada, otoyol trafiğini modellemek ve kontrol etmek için bir yapay sinir ağı tekniği kullanılmıştır. Geliştirilen sistem, şerit boyunca arzulanan trafik akış yoğunluğu dağılımı elde etmek için araçların takip etmesi gereken şeridin her bir bölümü için hız komutları üretirken, bu komutlar sürücülere veya doğrudan araçların kontrol sistemlerine iletilmesi ile sıkışıklığı hafifletmek ve insan kaynaklı hataları azaltmanın mümkün olacağı düşünülmüştür [48].

Yüz tanıma konusunda da yapay sinir ağlarını kullanan çalışmalar vardır. Rowley ve arkadaşlarının yaptığı bir çalışmada otomatik yüz tanıma sistemi sunulmuştur. Gözlerin ve ağzın otomatik tespiti, görüntülerin mekânsal normalizasyonu ile izlenmiş, normalleştirilmiş görüntülerin sınıflandırılması, bir hibrit (denetlenmiş ve denetlenmeyen) sinir ağı tarafından gerçekleştirilmiştir [38]. Başka bir makalede, olasılıksal karar temelli sinir ağlarına (PDBNN) dayanan bir yüz tanıma sistemi önerilmiştir. Model 3 fazda tanımayı gerçekleştirmiş; ilk olarak, görüntüde insan yüzünün konumu tespit edilmiş, ardından iki gözün konumlarından anlamlı özellikler çıkarılmış ve son adımda yüz tanıma işlemi ile tamamlanmıştır. Makale bu alanda bilinen FERET, ORL ve SCR veri kümeleri ile çalıştırılmış ve başarılı sonuçlar üretmiştir [37]. Sinir ağlarının öğrenme yeteneği, görev alanından kısıtlamalar getirerek büyük ölçüde artırılmıştır. Yann LeCun ve arkadaşlarının yaptığı bir çalışmada, bu kısıtlamaların ağı mimarisi yoluyla bir geri yayılım ağına nasıl entegre edilebileceği gösterilmiştir. Bu yaklaşım, ABD Posta Hizmeti tarafından sağlanan el yazısı posta kodu rakamlarının tanınmasına başarıyla uygulanmıştır [49].

E-mail spam tespitinde de yapay sinir ağları kullanılmaktadır. SpamAssassin corpus üzerinde yapılan bir çalışmada, giriş katmanlarında yapay sinir ağı üzerinde

değişiklikler yapılarak, işe yaramadığı düşünülen katmanlar ile daha iyi öğrendiği düşünülen katmanların yer değiştirildiği CLA-ANN adı verilen yaklaşım denenmiştir [39]. Anahtar kelimelerde kelime şaşırtmacaları ile yapılan spam e-mailler için bir algoritma geliştirilmiştir. Bu tür spam mailleri tespit etmek için, aynı sınıfa ait spam varyantlarının istatistiksel özelliklerini yakalamak için gizli Markov modelleri ve yapay sinir ağları kullanılmıştır. TREC 2007 veri kümesi üzerinde test edilmiş, modelin spam içeriğinin %90'ından fazlasını tespit edebildiğini ve yanlış pozitif oranı %13'ün altında olduğunu göstermiştir [40]. Bu konuda yapılan diğer bir çalışmada, Türkçe gibi aglütinatif¹ diller için anti-spam uygulaması geliştirilmiştir. Yapay sinir ağları ve Bayes ağları ile kurulan dinamik sistem ile, kullanıcıya özgü filtreleme yapılmıştır. Morfolojik analiz ile çıkarılan köklerde sınıflandırma yapılmış, 750 mail bulunan bir veri kümesinde %90 başarı elde edilmiştir [41].

Kanser teşhisinde de kullanılan yapay sinir ağları, özellikle meme kanseri konusunda oldukça başarılı sonuçlar vermektedir. Göğüs taraması (mamografi), meme kanserinin erken evrelerinde teşhis etmek için kullanılan bir yöntemdir. Mamogram meme dokusundaki kişinin kendisi veya doktoru tarafından fark edilemeyecek küçüklükteki değişiklikleri tespit edebilir. Yapılan bir çalışmada, meme kanseri tanısına yönelik sinir ağına dayalı yaklaşım tanımlanmıştır. Geliştirilen model, mamografiye başlamadan önce hangi kadının belirli bir tümör türünden muzdarip olma ihtimalini belirleyebilmeyi hedeflemiştir. Meme kanserinin gerçek klinik teşhisinde destek vektör makinelerinin (SVM) polinom çekirdeğinin kullanımını denenmiştir. Test, eğitim ve doğrulama aşamasında hiç kullanılmamış olan yeni veri ile yapılmış, yüksek ihtimalle kanser bulunan hastalarda %100, düşük ihtimalle kanser bulunanlarda ise %64 doğru tanılama gerçekleştirilmiştir. Dolayısıyla bu çalışma SVM yönteminin meme kanseri tanısı için hangi kadınların bir tarama programına girip girmemesi gerektiğini kararlaştırmak için onkologlara yardımcı olabileceğini göstermiştir [44]. İsmail Sarıtaş tarafından yapılan bir çalışmada ise meme kanseri teşhisinde, doktorlara biyopsi alınıp alınması gerektiği konusunda fikir veren yapay sinir ağına dayalı bir model geliştirilmiştir. Hastaların kanser tipine, ve kanser olup olmadıklarına yapay sinir ağı, BI-RADS (Meme Görüntüleme Raporlama ve Veri Sistemleri)² değerlendirmesi yapılarak ve hastanın yaşı, kütle şekli, kitle sınırı ve kütle yoğunluğu temel alınarak belirlenmiştir. Veriler biyopsi ile kesin olarak kanser tanısı alan 800 hastadan elde edilmiştir. Her bir hastaya ait kesin tanı ve YSA model sonuçlarındaki veriler, hata matrisi (confusion matrix) ve ROC analizleri kullanılarak araştırılmıştır. Bu analizlerin sonucunda uygulanan ANN modelinin test verilerinde hastalık tahmini oranı %90.5, sağlıklı hasta oranı %80.9 olarak bulunmuştur [45]. Akciğer kanseri teşhisinde yapılan bir çalışmada, yapay sinir ağları hem klinik öncesi hem de klinik sonrası tanıda sorunlara başarıyla uygulanmıştır. Akciğer kanseri hastalarından gelen demografik verileri (yaş, cinsiyet, vs.) analiz edilmiş, akciğer kanseri teşhisi problemi için, ağdan çıkarılan özlü kurallar, eğitim veri

¹ Kök üzerine eklemeye yapılan, eklerin yoğun kullanıldığı diller.

² Lezyonların tanımlanmasını, şüphe derecesini, hangi lezyona ne tür biyopsi yapılması gerektiğini, hastaya ve klinisyen hekime takip-tedavi yönünden tavsiyeleri içeren rapor.

setinde ve test veri setinde yüksek bir doğruluk oranı elde edilmiştir [42]. Başka bir çalışmada, bir kurumda tedavi gören kolorektal kanser hastaların verileri ile başka bir bölgedeki farklı bir kurumdaki hastalar üzerinde ölüm zamanı ile ilgili tahminler yapılmıştır. Kolorektal kanser nedeniyle tedavi edilen 334 hastadan 5 yıllık takip verileri 9, 12, 15, 18, 21 ve 24 ay içinde ölüm tahmini için tasarlanmış altı sinir ağını eğitmek ve doğrulamak için kullanılmıştır. Ağın tahminleri, aynı veriyi inceleyen iki danışman kolorektal cerrahın öngörülerıyla karşılaştırılmıştır. Sonuçta, 1. kurumdaki hastalar için içinde ölüm tahmini için %80'den fazla, 2. kurumdan alınan verilerde kolorektal cerrahların genel doğruluğu %79 ile karşılaştırıldığında ise %90'lık genel doğruluk elde edilmiştir [43].

Görüldüğü gibi yapay sinir ağı yaklaşımları çok farklı alanlardaki problemlere başarıyla uygulanmaktadır. Ancak veri kümelerinin büyümesi ve hesaplama maliyetlerinin artması bu alandaki çalışmaları yavaşlatmıştır. Son yıllardaki donanımsal gelişmeler ile birlikte, derin öğrenme yaklaşımlarının ortaya çıkmasıyla yapay sinir ağlarının çoklu gizli katmanlı yapıları uygulanmaya başlanmıştır.

DERİN ÖĞRENME

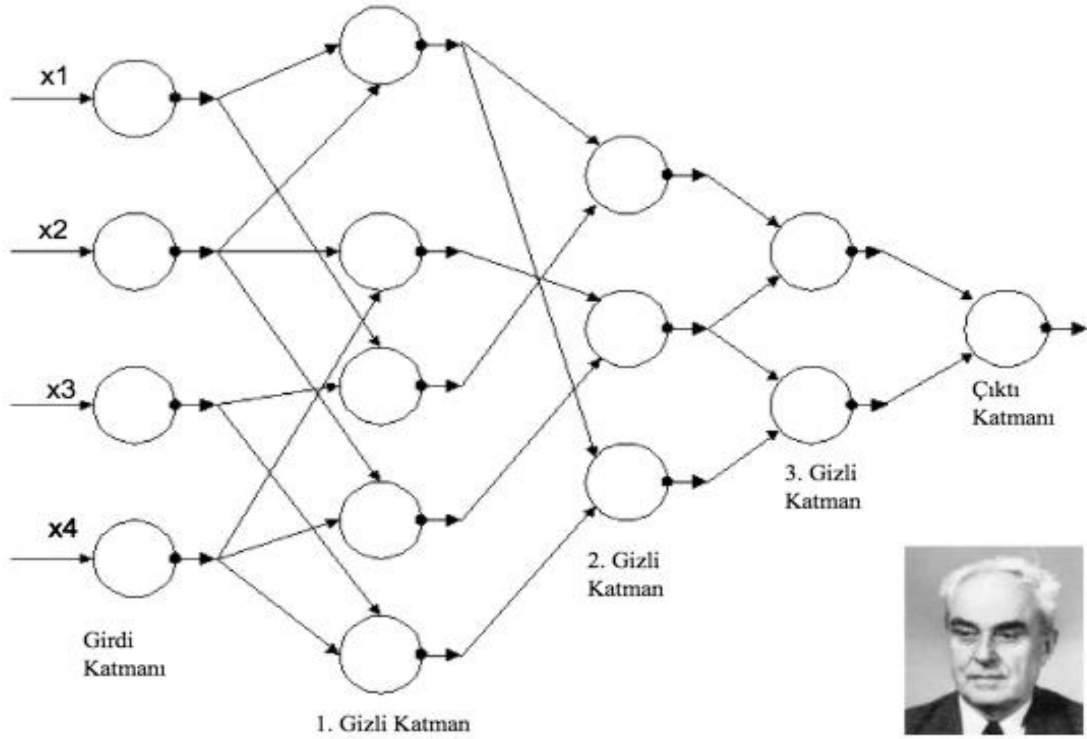
3.1 Tanım

Derin öğrenme bir makine öğrenmesi sınıfıdır. Derin öğrenme, özellik çıkarma ve dönüştürme için birçok doğrusal olmayan işlem birimi katmanını kullanır. Her ardışık katman, önceki katmandaki çıktıyı girdi olarak kullanır [50]. Algoritmalar denetimli (sınıflandırma gibi) veya denetimsiz (desen analizi gibi) olabilir. Derin öğrenmede, verilerin birden fazla özellik seviyesinin veya temsillerinin öğrenilmesine dayanan bir yapı söz konusudur. Üst düzey özellikler, alt düzey özelliklerden türetilerek hiyerarşik bir temsil oluşturur. Bu temsilde, soyutlamanın farklı seviyelerine karşılık gelen birden çok temsil seviyesini öğrenir [51].

Derin öğrenme verinin temsilinden öğrenmeye dayalı bir yöntemdir. Bir görüntü için temsil denildiğinde; piksel başına yoğunluk değerlerinin bir vektörü veya kenar kümeleri, özel şekiller gibi özellikler düşünülebilir. Bazı özellikler veriyi daha iyi temsil etmektedir. Derin öğrenmenin bir diğer avantajı, manuel çıkarılan özellikler (handcrafted features) yerine denetimli/denetimsiz özellik öğrenmesi veya hiyerarşik özellik çıkarımı için etkin algoritmalar kullanmasıdır [52].

3.2 Tarihçe

Denetimli derin beslemeli çok katmanlı perceptronlar için ilk genel, öğrenme algoritması Ivakhnenko ve Lapa tarafından 1965 yılında yayınlanmıştır [53]. Bu çalışmada, her katmanda, en iyi özellikleri istatistiksel yöntemlerle seçilip bir sonraki katmana iletilmektedir. Ağlarını uçtan uca eğitmek için geri yayılımı (backpropagation) kullanılmamış, önceki katmanlardan sonraki katmanlara en küçük kareler yöntemi kullanılmıştır



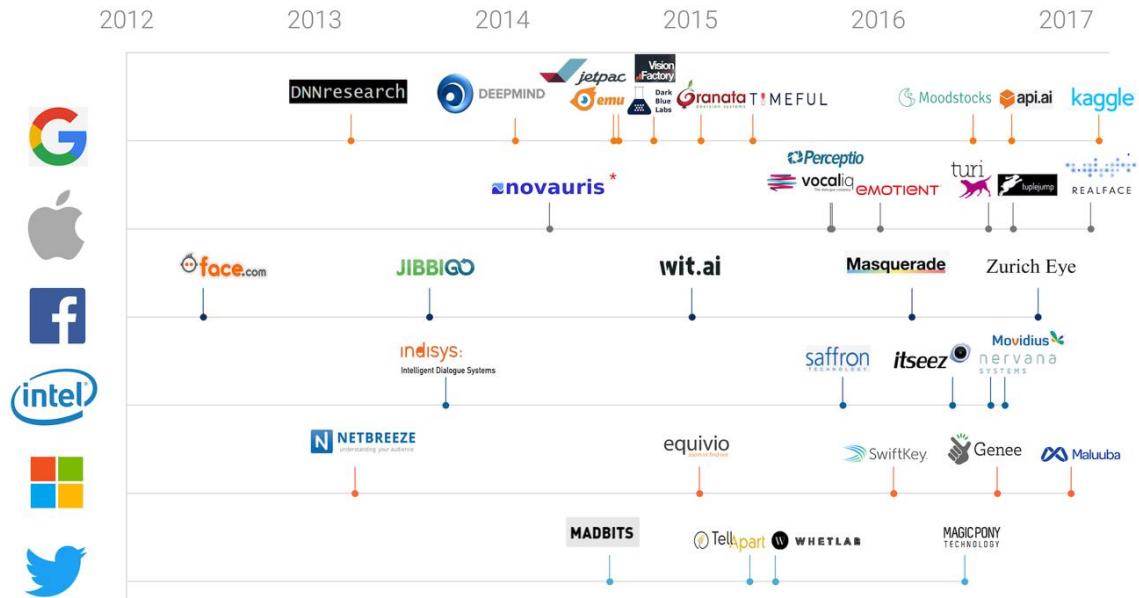
Şekil 4 Ivankhnenko A. tarafından eğitilen bilinen ilk derin ağ mimarisi [54]

Ivankhnenko'dan sonra ilk derin öğrenme mimarisi "Neokognitron", 1979 yılında Fukushima tarafından önerilmiştir (Şekil 4). Omurgalı canlıların görsel sinir sistemlerinden ortaya atılan yapıda, "öğretmensiz öğrenme" ile kendi kendini organize eden bir ağ geliştirilmiştir. Fukushima'nın ağları modern ağlara benzer çoklu bükülme ve havuz katmanlarını içermiştir [55].

Derin mimarilerdeki öğrenmenin eksikliği çoklu katmanlardaki hataların geri yayılımındaydı. Geri yayılım algoritmaları önceki yıllarda ortaya atılmış olsa da ilk başarılı derin sinir ağı uygulamasını Yann LeCun ve arkadaşları posta kutusu yazıları üzerinde geliştirmişlerdir [56]. Her ne kadar ağ başarılı çalışsa da, eğitiminin yaklaşık olarak 3 gün sürdüğü için pratik için uygun olmadığı görülmüştür. Bu çalışmadan sonra yine Yann LeCun "LeNet" ağını kullanarak el yazısı rakamlarını (MNIST) sınıflandırmak için kıvrımlı ağlarla geri yayılımı birlikte uygulamıştır [49]. 1995 yılında Brendan Frey, Peter Dayan ve Geoffrey Hinton ile ortaklaşa geliştirilen uyanık-uyku (wake-sleep) algoritmasını kullanarak altı adet tamamen bağlı katmanlar ve yüzlerce gizli katman içeren bir ağı, her ne kadar eğitim iki gün sürse de, eğitilmesi mümkün olduğunu göstermişlerdir [57]. Hochreiter ve Schmidhuber'in tekrarlayan sinir ağları (recurrent neural networks) için 1997'de uzun kısa vadeli bellek (long short-term memory) gibi bazı önemli gelişmeler yapılmıştır [58]. YSA'ların hesaplama maliyetinden dolayı, avantajlarına rağmen, destek vektör makineleri [59] (support vector machine) gibi probleme özel çalışan, elle hazırlanmış özellikleri kullanan daha basit modeller, 1990'lı ve 2000'li yıllarda popüler bir tercih olmuştur.

Bilgisayarların daha hızlı çalışmasını ve daha sonra grafik işlemci birimlerinin (GPU'lar) kurulmasıyla büyük gelişmeler gerçekleşmiştir. Daha hızlı bilgisayarların ve GPU'lar ile, hesaplama hızını, 10 yıllık bir sürede 1000 kat artmıştır. Bu dönemde, sinir ağı yavaş yavaş destek vektör makinelerine rakip olmaya başlamıştır [2]. YSA bağlamında "Derin Öğrenme" ifadesi 2000 yılında Igor Aizenberg ve arkadaşları tarafından tanıtılmıştır [60]. 2006'da Geoffrey Hinton yayınladığı bir makalede, çoklu katmanlı ileri beslemeli sinir ağının her seferinde bir katmanı etkili bir şekilde nasıl eğitebildiğini (Her bir katmanı denetimsiz olarak kısıtlanmış bir Boltzmann makinesi ile eğitmiştir.), ardından denetimli bir geri yayılım yöntemi ile ince ayar yapabileceklerini göstermiştir [61]. GPU hızlarının artmasıyla birlikte derin ağların önceden eğitmeden de eğitilebilmesi mümkün hale gelmiştir. Ciresan ve arkadaşlarına trafik işaretleri, medikal görüntüleme ve karakter tanıma yarışmalarını kazandıran derin ağlarında bu yaklaşımı kullanmışlardır [62]–[64]. Krizhevsky, Sutskever ve Hinton 2012'de benzer bir mimari kullanmışlardır. Eğitim hızını artırmak için GPU kullanılmış ve tam-bağlı (fully-connected) katmanlardaki ezberlemeyi (overfitting) azaltmak için, son yıllarda geliştirilen, “dropout”³ adı verilen normalleştirme yöntemini [65] kullanarak etkinliğini kanıtlamışlardır [66]. Bu yaklaşımları onlara ILSVRC-2012 ImageNet yarışmasında olağanüstü sonuçlar getirmiştir.

Bu gelişmelerden sonra Google, Facebook ve Microsoft gibi teknoloji firmaları bu trendi fark edip derin öğrenme alanına yatırım yapmaya başladılar. Şekil 5'te görüldüğü gibi, 2012-2017 yıllarında çalışma grupları ve startup şirketleri satın alınmasıyla başlayan süreçte derin öğrenme alanındaki gelişmeler bir hayli hızlanmıştır.



Şekil 5 Yıllara göre yapay zeka startup'ların satın alınması [67]

³ Eğitimin her iterasyonunda, gizli katmanlardaki düğümlerin her biri 0,5 olasılıkla ağdan atılır, bu sayede mevcut gizli bir birim diğer gizli birimleri bağlamaz ve ezberleme durumu azaltılır.

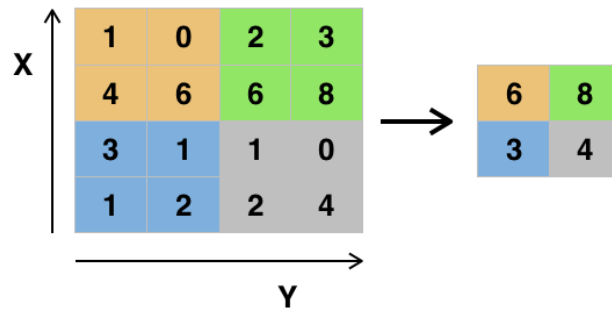
3.3 Derin Öğrenme Mimarileri

Yapay sinir ağlarındaki katman sayılarının artırılmasıyla kurulan çok farklı türde derin öğrenme mimarileri bulunmaktadır. Bu bölümde sık kullanılan mimarilerden kısaca bahsedilecektir.

3.3.1 Konvolüsyonel Sinir Ağları

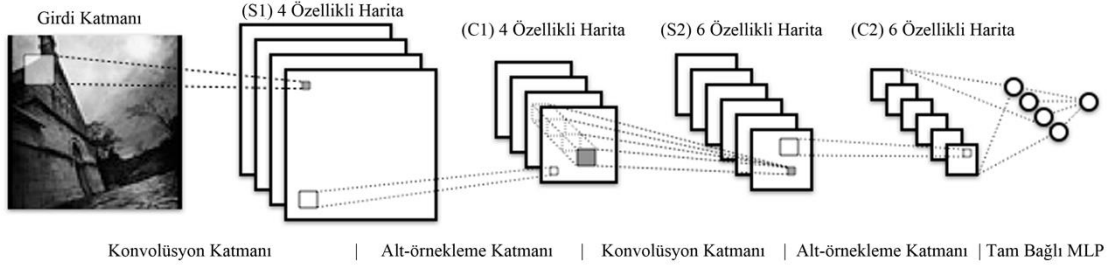
Konvolüsyonel Sinir Ağları (CNN veya ConvNet) çok katmanlı algılayıcıların (MLP) bir türüdür. Görme merkezindeki hücreler tüm görseli kapsayacak şekilde alt bölgelere ayrılmıştır. Basit hücreler, kenar benzeri özelliklere yoğunlaşırken, karmaşık hücreler ise daha geniş alıcılarla, tüm görsele yoğunlaşır. İleri yönlü bir sinir ağı olan CNN, hayvanların görme merkezinden esinlenilerek ortaya atılmıştır. Buradaki matematiksel konvolüsyon işlemi, bir nöronun kendi uyarı alanından uyarılara verdiği cevap olarak düşünülebilir [55], [68], [69].

CNN, bir veya daha fazla konvolüsyonel katman, altörnekleme (subsampling) katmanı ve bunun ardından standart çok katmanlı bir sinir ağı gibi bir veya daha fazla tamamen bağlı katmanlardan oluşur [3]. CNN'lerin bir yararı, aynı sayıda gizli birimle tamamen bağlı ağlardan daha az sayıda eğitime ve daha az parametreye sahip olmalarıdır. Bir konvolüsyonel katmanındaki girdi, $m \times m \times r$ 'lik bir resimdir. Burada m değerleri sırasıyla görüntünün yüksekliği ve genişliği, r ise kanalların sayısıdır (örneğin, bir RGB görüntüde $r = 3$ 'tür.). Konvolüsyonel katmanı ise $n \times n \times q$ boyutlarında k adet filtreden oluşur. Filtre için görüntünün boyutundan daha küçük olan n , ve r (kanal) değeriyle genellikle aynı seçilen q değeri seçilir. Bu filtreler ile, her biri $m-n+1 \times m-n+1$ boyutunda k adet birbirine yerel olarak bağlı özellik haritaları üretir [64]. Daha sonra her harita için, Şekil 6'de verilen örnek gibi, maksimum veya ortalama havuzlama (pooling) ile $p \times p$ 'lik altörnekleme işlemi (p değeri, görüntü büyüklüğüne göre genellikle 2-5 arasında) uygulanır [70].



Şekil 6 Maksimum havuzlama için bir örnek [71]

Şekil 7'de verilen ilk CNN ağı 1988 yılında Yann LeCun tarafından ortaya atılan, 1998'lere kadar iyileştirmeleri devam eden LeNet isimli mimaridir [72].



Şekil 7 LeNet ağının mimarisi

LeNet ağında, alt katmanlar art arda yerleştirilmiş konvolüsyon ve maksimum havuzlama katmanlarından oluşur. Sonraki üst katmanlar ise tamamen bağlantılı geleneksel MLP'ye karşılık gelmektedir.

CNN algoritmaları görüntü ve ses işleme alanı başta olmak üzere doğal dil işleme, biyomedikal gibi bir çok farklı alanda uygulanmaktadır. Özellikle görüntü işleme alanında state-of-art sonuçlar elde edilmiştir. MNIST veri kümesi üzerinde, Cireşan yaptığı çalışmada, CNN ile hata oranı %2'lere kadar düşürmeyi başarmışlardır [73]. MNIST ve NORB veri kümeleri üzerinde denenen başka bir çalışmada öğrenme sürecinin çok hızlı gerçekleştiği ve yayınlandığı zamana kadar olan yöntemler içinde en başarılı olduğu belirtilmiştir [74]. 2014 yılında, ImageNet Büyük Ölçekli Görsel Tanıma Yarışması'nda, milyonlarca resim ve yüzlerce nesne sınıfı ile nesne sınıflandırması ve algılaması kriterlerinde en başarılı dereceleri alan ekiplerin hepsi temelde CNN algoritmalarını kullanmıştır [75]. 2015'te çok katmanlı bir CNN, ters yüzler de dahil olmak üzere geniş açı aralıklarındaki yüzleri yakalayabilme yeteneğini göstermiştir. Bu ağ, çeşitli açılar ve yönlerde yüzleri içeren 200.000 görüntü ve yüzleri olmayan 20 milyon görüntü daha içeren bir veri tabanı üzerinde eğitilmiştir [76].

CNN modelleri çeşitli NLP problemleri için etkili oldukları gösterilmiştir. Anlamsal ayrıştırmada [77], arama sorgusu elde etmede [78], cümle modelleme [79], sınıflandırma [80], tahmin problemlerinde [81] mükemmel sonuçlar elde edilmiştir.

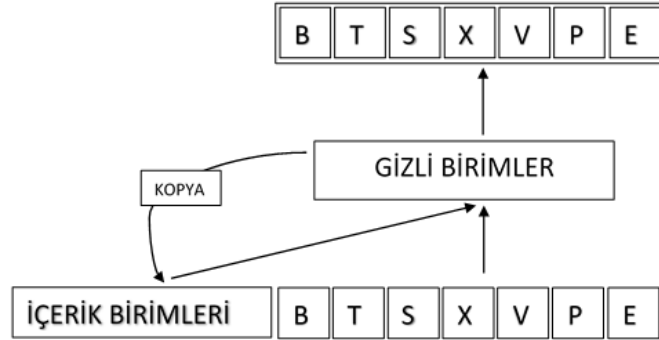
CNN algoritmaları ilaç keşfinde de kullanılmıştır. 2015 yılında Atomwise şirketinin geliştirdiği AtomNet, ilaç tasarımı için geliştirilen ilk derin sinir ağı olmuştur [82]. Kimyasal tepkimelerin 3 boyutlu temsilleriyle eğitilen sistem, Ebola ve skleroz gibi hastalıklarda yeni biyomolekülleri keşfedebilmek için kullanılmıştır [83].

Go oyunu için de yine CNN kullanılmış, geleneksel yöntemlerle geliştirilen GNU Go algoritmasını, önceden eğitilmiş 12-katmanlı bir CNN algoritması oyunların %97'sinde yenmiştir [84]. Google DeepMind tarafından geliştirilen CNN tabanlı AlphaGo ise ilk kez bir profesyonel insanı yenebilen program olmuştur [85].

3.3.2 Tekrarlayan Sinir Ağı

Basit Tekrarlayan Ağ (SRN), Jeff Elman tarafından tasarlanmıştır (Şekil 8). Elman'ın cümle yapısı simülasyonunda kullanılan her bir kelime için gizli kalıpların üzerinde ortalama örüntü kümeleme sonucunda isim ve fiil kategorileri temiz şekilde ayrılmıştır.

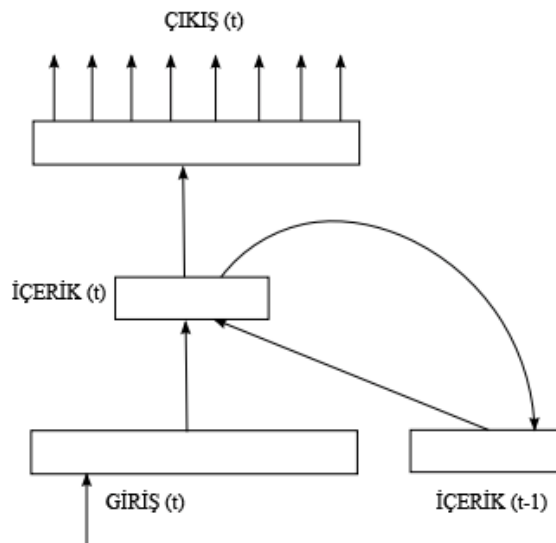
Bunun yanında isimler arasında canlı-cansız ayrımları, hatta insan-hayvan, hayvanlar arasında avcı-yırtıcı gibi kümeler de ayrılmıştır [86].



Şekil 8 SRN ağ mimarisi [86]

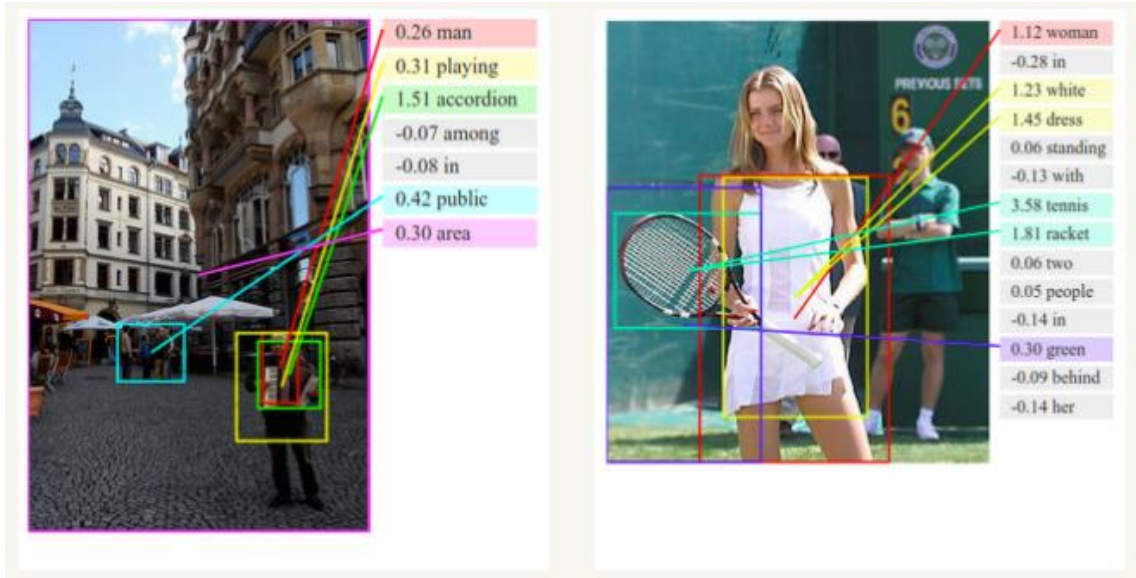
RNN, birimler arasındaki bağlantıların yönlendirilmiş bir döngü oluşturduğu yapay sinir ağı sınıfıdır. Bu döngü ile, dinamik zamansal davranış sergilemesine olanak tanıyan bir ağ iç durumu oluşturulmuştur. İleri beslemeli sinir ağların aksine, RNN'ler kendi giriş belleğini girdilerin rastgele dizilerini işlemek için kullanabilmektedirler (Şekil 9) [87].

Tekrarlayan sinir ağındaki (RNN) temel düşünce sıralı bilgileri kullanmaktır. Geleneksel bir sinir ağında tüm girdilerin (ve çıktıların) birbirinden bağımsız olduğunu varsayılmaktadır. Fakat NLP gibi birçok alan için bu çok kötü bir fikirdir. Örneğin, bir cümle içinde bir sonraki kelimeyi tahmin etmek için, hangi sözcüklerin o anki kelimedenden önce geldiğini bilmek gerekmektedir. RNN mimarisinin yinelenen (recurrent) olarak adlandırılmasının sebebi, bir dizinin her ögesi için aynı görevi önceki çıktılarına bağlı olarak yerine getirmesidir.



Şekil 9 Basit bir RNN mimarisi [87]

Alex Graves'in çalışmasında, ses verilerinin fonetik sunumuna gerek kalmadan doğrudan metine çeviren bir RNN tabanlı konuşma tanıma sistemi sunulmuştur [88]. Başka bir çalışmada, RNN'ler CNN ile birlikte, etiketlenmemiş görüntüler için tanımlayıcı üreten bir modelin parçası olarak kullanılmıştır. Birleştirilmiş model, görüntüdeki nesneleri tanımlamanın yanında, tanımlayıcıların görüntülerdeki konumlarını bile bulmayı başarmıştır (Şekil 10) [89].



Şekil 10 Görüntü tanımlayıcıları oluşturmak için derin görsel-anlamsal hizalama [89]

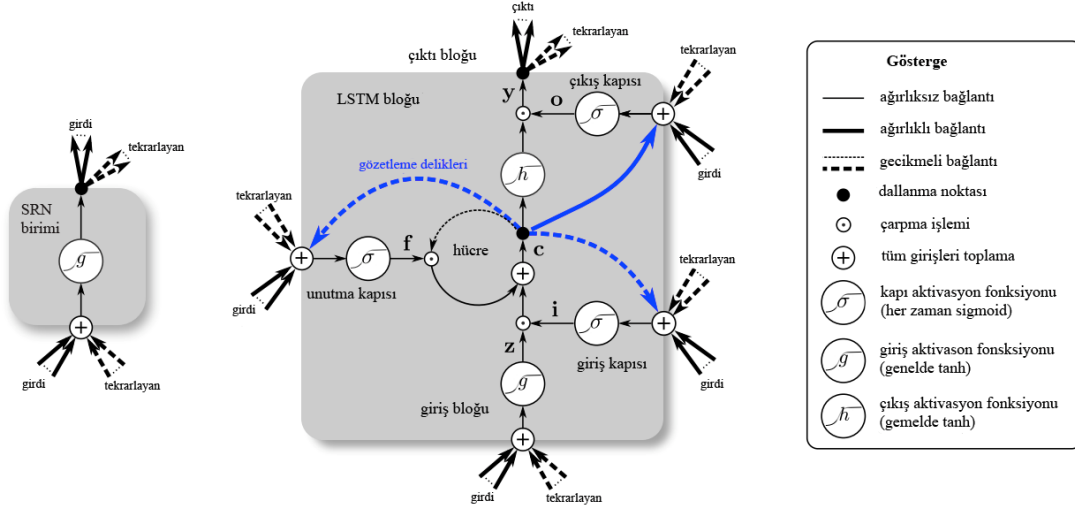
3.3.3 Uzun Kısa Vadeli Memory

RNN mimarilerinde önceki bilgi kullanımına dayalı bir yaklaşım vardır. Örneğin “Ağaç toprakta yetişir” cümlesinde “toprak” kelimesini tahmin etmek kolaydır. Fakat bağlamlar arası boşluk arttığında RNN geçmişten gelen bilgiyi kullanması oldukça zordur. Örneğin,

“İngiltere’de büyüdüm Akıcı bir şekilde İngilizce konuşurum.”

gibi bir metinde “İngilizce” kelimesini tahmin ederken, içinde bulunduğu cümleden yola çıkarak bir dil adı olacağını tahmin edilebilir, ancak doğru kelimenin “İngilizce” olduğunu tahmin etmek için, metnin başındaki cümleyi hafızada tutmak gerekmektedir. Teoride mümkün olan “uzun-vadeli bağımlılıklar”, pratikte büyük problemlere yol açtığı ortaya çıkmıştır [90].

Bu problemi çözmek için, uzun vadeli bağımlılıkları öğrenebilen özel bir RNN türü olan Uzun Kısa Vadeli Hafıza Ağları (LSTM) Hochreiter ve Schmidhuber tarafından 1997 yılında tanıtılmıştır [58].



Şekil 11 RNN’de kullanılan SRN birimi (solda) ve LSTM (sağdaki) ayrıntılı şeması [91]

Şekil 11’de de görüldüğü gibi LSTM mimarisinde giriş, unutma ve çıkış olmak üzere 3 üç kapı, blok girişi, tek bir hücre (Sabit Hata Döngüsü), bir çıkış aktivasyon fonksiyonu ve gözetleme (peephole) bağlantılarına sahiptir. Bloğun çıktısı tekrar tekrar bloğun girişine ve tüm kapılarına bağlanır. Gözetleme bağlantıları ve unutma kapısı ilk geliştirilen mimaride bulunmamaktadır. LSTM’in kendi durumunu sıfırlamak için unutma kapısı [92], kesin zamanlamaları öğrenmeyi kolaylaştırmak için ise gözetleme bağlantıları eklenmiştir [93].

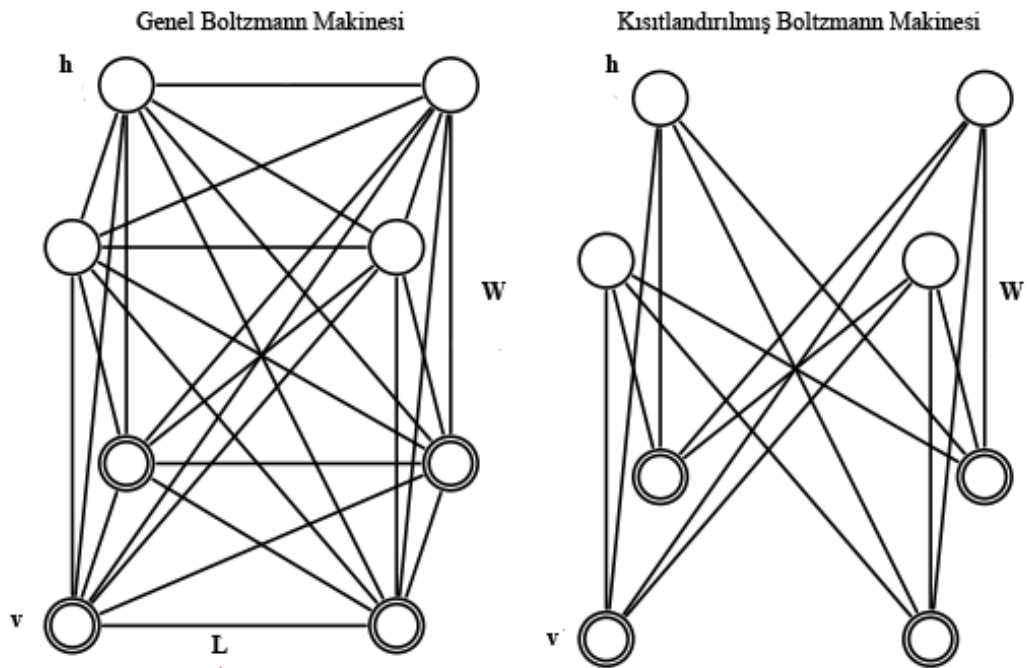
LSTM mimarileri konuşma/metin işleme konularında oldukça iyi sonuçlar vermektedir. Çerçeve tabanlı ses sınıflandırma çalışmasında farklı lehçelerdeki zengin içerikler barındıran TIMIT veri kümesi kullanılmış, %70 başarı oranı elde edilmiştir [94]. Uygun ayarlamalar ile uçtan uca eğitimden geçen başka bir LSTM yaklaşımı ile aynı veri kümesi üzerinde %83 başarı oranına ulaşılmıştır [95]. Başka bir çalışmada, anahtar kelime tespitinde geniş kapsamlı konuşma verileri barındıran Verbmobil veri kümesi üzerinde %84.5’lik bir doğruluk oranı verilmiştir [96]. Başka bir çalışmada, insan aktivitelerini önceden bilgi kullanmadan sınıflandırmayı öğrenen tamamen otomatikleştirilmiş CNN ve LSTM algoritmalarını kullanan bir derin model başarıyla uygulanmıştır [97].

İlginç bir başka uygulama ise LSTM ile müzik besteleme çalışmasıdır. RNN ile nota tahmini yapılabilir de beste yapmak zordur. Bu çalışmada, ise uzun vadeli öğrenme yeteneği ile LSTM’nin bir blues türünde müzikleri başarılı bir şekilde öğrendiğini ve bu tarzda yeni besteler oluşturabileceğini gösteren model tasarlanmıştır [98].

Hızalamadan bağımsız olarak protein homolojisinin algılanması [99], robotik kalp cerrahisinde düğümleri bağlamayı öğrenen bir sistem tasarlanması [100], düzensiz dillerde öğrenme [101], çevrimdışı elyazı tanınması [102] gibi çok farklı konularda da LSTM mimarileri kullanılmıştır.

3.3.4 Kısıtlanmış(Derin) Boltzmann Makineleri

Kısıtlanmış Boltzmann makinesi (RBM), Harmonium [103] adıyla 1986'da ilk olarak ortaya atılmış, ancak çok uzun zaman sonra 2006'da Geoffrey Hinton ve arkadaşları tarafından hızlı bir öğrenme algoritması olarak ön plana çıkarılmıştır [104]. RBM girdi seti üzerinde olasılık dağılımı öğrenebilen üretken bir rastgele yapay sinir ağıdır. RBM, Boltzmann Makinalarının bir türüdür. Görünür ve gizli olmak üzere aralarında simetrik bağlantı bulunan iki parçalı graflardan oluşurlar. Bir graf içindeki düğümlerde kendi aralarında bağlantı bulunmaz (Şekil 12) [105]. Buna karşın, Kısıtlamasız Boltzmann makineleri gizli birimler arasında da bağlantı vardır. Bu kısıtlama, Boltzmann makinelerinin genel sınıfı için mevcut olanlardan daha etkili eğitim algoritmalarına imkan tanır [106].

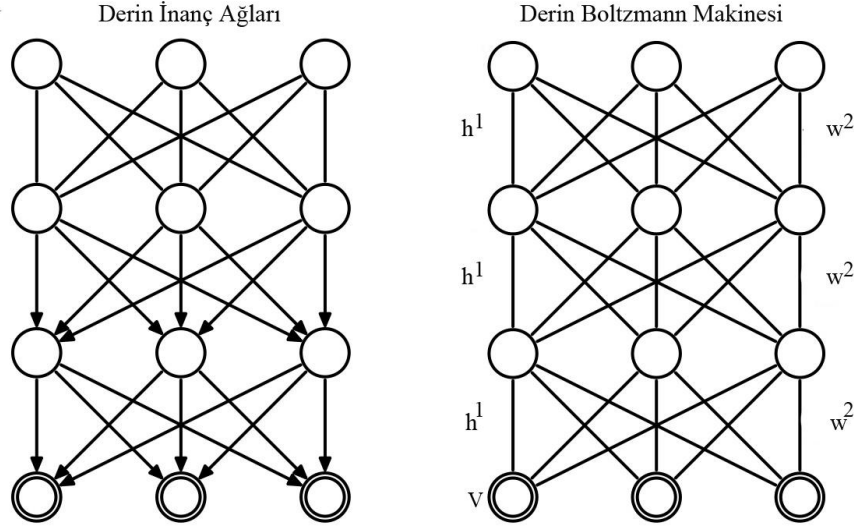


Şekil 12 Boltzmann Makineleri (solda) ve Kısıtlanmış Boltzmann Makineleri (sağda) [105]

RBM'ler, boyut indirgeme [104], sınıflandırma [107], işbirlikçi filtreleme [108], özellik öğrenimi [109] ve konu modelleme [110] gibi farklı konular için kullanışlı bir algoritmadır.

3.3.5 Derin İnanç Ağları

Geoffrey Hinton tarafından tanıtılan, Derin inanç ağları (DBN), önceki bölümde verilen RBM'lerin yığını olarak tanımlanmaktadır. Her RBM katmanı hem önceki hem de sonraki katmanlarla bağlıdır. Ancak herhangi bir katmanın düğümleri birbirleriyle yatayda iletişimi bulunmamaktadır (Şekil 13). En son katman olarak bir softmax katmanı ile sınıflandırma veya denetimsiz bir öğrenme için kümeleme yeteneğine sahiptir [104].



Şekil 13 DBN ve RBM mimarilerinin karşılaştırılması [111]

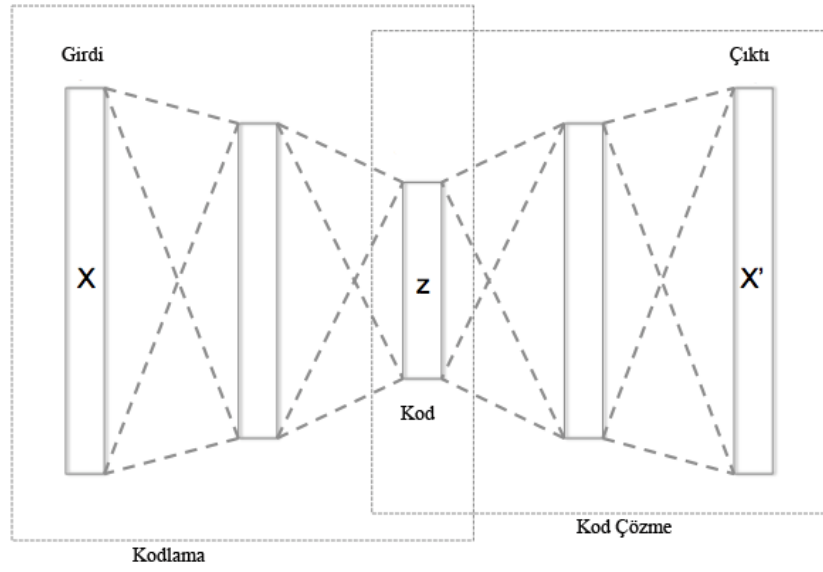
X bir dizi özellik vektörü olan girdi verisi olarak kabul edildiğinde, DBN bir ağda öğrenme işlemi aşağıdaki adımlar ile gerçekleştirilir [112];

- Ağırlık matrisini elde etmek için X üzerinde bir RBM eğitilir ve W ağırlıkları elde edilir. Bu (W) ağın alttaki iki katmanı arasındaki ağırlık matrisi olarak kullanılır.
- Örnekleme veya gizli düğümlerin ortalama aktivasyonunu hesaplayarak, RBM ile X bir sonraki katmanın için girdi olan yeni veri X' dizisine dönüştürülür.
- Ağın en üstteki iki katına ulaşıncaya kadar, $X \leftarrow X'$ dönüşümü tekrarlanır.
- Bu derin mimarının tüm parametreleri için, DBN olasılığı için bir vekile (proxy) göre ince ayar yapılır.

DBN mimarileri görüntü tanıma ve üretme konularında uygulanmıştır [112]–[114]. Başka bir çalışmada, bütün dokümana TF-IDF uygulamak yerine DBN kullanan bir sistem ile büyük bir belge veri kümesinde kelime-sayı vektörleri elde edilmiştir [115]. İnsan hareketi tanımlamak için Taylor ve arkadaşları çeşitli hareket dizilerini sentezleyerek ve hareket yakalama sırasında kaybolan verilerin çevrimiçi olarak doldurularak gösterilmesini sağlamışlardır [116].

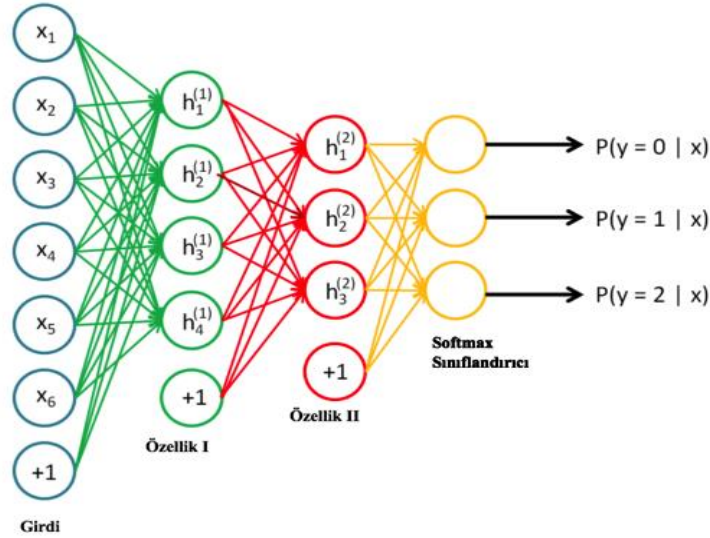
3.3.6 Derin Oto-kodlayıcılar

Diabolo ağı olarak adlandırılan Oto-kodlayıcılar (AE) denetimsiz öğrenme için kullanılan bir özel yapay sinir ağıdır [51], [117]. AE, bir veri kümesi için boyut indirgeme amacıyla bir temsil (kodlama) öğrenmeyi hedefler. Son yıllarda, AE konsepti, verinin üretken modellerini öğrenmek için daha yaygın olarak kullanılmaya başlamıştır [118], [119]. AE, kabaca girdi verisinin sıkıştırılmış gösteriminden en iyi özelliklerin öğrenilmesini hedefleyen bir ileri beslemeli sinir ağıdır [120]. Şekil 14’de görüldüğü gibi, AE, girdi verisi şifreleme-şifre çözme işleminden sonra çıktı olarak yine aynı girdiyi görene kadar ağırlıklar değiştirilir. Hedefe ulaşıldığında gizli katmandaki düğüm sayısı ile girdi verisi temsil edilmiş olmaktadır.



Şekil 14 Basit AE mimarisi [121]

Derin veya yığınlı oto-kodlayıcılar ise (DAE), her bir katmandaki çıktıların ardışık katmanın girişlerine bağlandığı AE'lerin çok katmanlarından oluşan sinir ağıdır [122].



Şekil 15 Genel bir DAE mimarisi [123]

Şekil 16'da ardışık olarak yerleştirilmiş iki AE ağı bulunmaktadır (Şifre çözme kısımları gösterilmemiştir.). Girdi dizisi (Input) ilk gizli katmanda şifrelenmiş Özellik I (Features I) dizisi elde edilmiştir. Sonraki gizli katmana elde edilen özellik dizisi girdi olarak verilmiş ve şifreleme sonucunda özellik II (features II) dizisi elde edilmiştir. Özellik II dizisi ise özellik vektörü olarak son katmanda softmax sınıflandırıcıya girdi olarak verilmiş, sınıflandırma işlemi bu vektörle gerçekleştirilmiştir.

DAE mimarisi kumaş hatası tespiti için uygulanmış, 2 ardışık AE ile oluşturulan derin ağ ile oldukça başarılı özellikler elde edilmiştir [122]. Mail spam yakalama için

kullanılan bir başka çalışmada ise sonuçlar Naive Bayes, Destek Vektör Makineleri, Karar Ağaçları gibi diğer yöntemler ile kıyaslanmış hem doğruluk hem de fl ölçümlerinde daha iyi performans gösterdiği görülmüştür [124]. Konuşma spektrogram kodlaması üzerindeki bir çalışmada ise DAE tabanlı bir mimari uygulanmış, konuşma verisi ile, alınan sonuçların, daha önce görüntü yamalarını kodlamak için elde edilen başarı ile yaklaşık olarak aynı olduğunu gözlemlenmiştir [125].

3.4 Derin Öğrenme Uygulamaları

Bir önceki bölümde mimariler üzerinden bir inceleme yapılmıştır. Burada ise konu bazında bir ayrıştırma yapılarak, özellikler son yıllarda hangi alana hangi yöntemler uygulanıyor konusu irdelenecektir.

3.4.1 Doğal Dil İşleme

Doğal dil işleme alanında oldukça başarılı olan derin öğrenme algoritmalarından, özellikle DAE, LSTM, CNN mimarileri etkileyici sonuçlar vermiştir. Az miktarda etiketlenmiş veri ile duygu analizi yapmak bilinen bir problemdir. Shusen Zhou, Aktif derin ağ (ADN) olarak adlandırılan bir yarı-eğitici öğrenme algoritması bu problem üzerinde çalışmıştır. RBM tabanlı kurulan derin ağ, 5 duygu analizi veri kümesi üzerinde uygulanmıştır [126]. Yine duygu analizi ile ilgili bir başka çalışmada, bulanık DBN adı verilen iki aşamalı bir yarı-eğitici öğrenme metodu sunulmuştur. İlk olarak, DBN ile yarı-eğitici öğrenme tarafından oluşturulan veri seti ile eğitilmiştir. Ardından, derin mimari tarafından öğrenilmiş görüşleri baz alan bir bulanık üyelik fonksiyonu devreye girmiş, kıyaslanan önceki çalışmalara göre etkili sonuçlar alınmıştır [127]. Duygu analizinden esinlenilerek yapılan diğer bir çalışmada, hapishanelerdeki kısa mesajların otomatik güvenlik denetimi için kısa mesajları (güvenli ve güvensiz) sınıflandıran yeni derin öğrenme tabanlı bir ağ olan RNN modeli kullanılmıştır. Kısa mesajlardan word2vec [128] vasıtasıyla tipik özellikler çıkarılmış ve RNN ile sınıflandırılmıştır. DeneySEL sonuçlar, RNN modelinin SVM'ye göre daha yüksek olan ortalama %92.7 doğruluk elde ettiğini göstermiştir [129]. Cümle sınıflandırma doğal dil işlemede önemli bir problemdir. CNN tabanlı bir yaklaşım ile önceden eğitilmiş kelime vektörleri cümle sınıflandırma probleminde kullanılmış ve 4'ü çok değerli 7 konu üzerinde (duygu analizi ve soru sınıflandırma dahil) yüksek başarılar göstermiştir [80]. Ayrıca çoklu konu (multi-tasking) öğrenme için de yine CNN yöntemi kullanılmıştır [81]. Ses tanıma üzerine yapılan çalışmalarda da RNN ağları kullanılmaktadır [95]. Metin özetleme ile ilgili bir çalışmada, DAE ile, Terim Frekans (tf) girdisinden bir özellik alanı hesaplamak için tek belge özetleme yöntemlerini sunulmuştur. ENAE adı verilen, tf değerlerine rastgele gürültüler eklenmiş topluluk (ensemble) tabanlı öğrenme yöntemi denenmiştir. İki farklı konudaki e-mail veri kümesi üzerinde çalışılmış ortalama %11,2 hatırlama (recall) değerlerine ulaşılmıştır [130]. Metinlerdeki anahtar kelimelerin tespiti için LSTM tabanlı bir yaklaşım, bu konuda başarılı olan HMM ile kıyaslanmış ve daha başarılı olduğu görülmüştür [96].

Yazar cinsiyeti belirlemek amaçlı bir çalışmada CNN ve LSTM birlikte kullanılmıştır. Rusça metinler üzerinde yapılan çalışma, geleneksel yöntemlerle kıyaslanmış %86 doğruluk ile state-of-art çalışmalara yakın başarı göstermiştir [131]. Yazarlık doğrulama konusunda yapılan bir çalışmada ise DBN kullanımı araştırılmıştır. Modelden sözcüksel, sözdizimsel ve uygulamaya özgü özellikler elde edilmiştir. Twitter ve Enron e-mail veri kümesi üzerinde test edilen model %8-16 arası hata oranlarıyla umut vadetmiştir [132].

Web üzerinden elde edilen metinlerde veya sorgularda düşük boyutlu anlamsal temsiller elde etmek, veri işleme için önemli bir problemdir. CLSM (Convolutional Latent Semantic Model) adı verilen model kullanılarak web sayfalarından tıklama verisi ile eğitilen büyük çaplı bir gerçek dünya verisi üzerinde çalışılmıştır. Döküman sıralama (document ranking) değerlendirmesi yapılmış, belgelerdeki göz çarpan semantik bilgilerin elde edildiği, önceki yöntemlerden açıkça üstün olduğu görülmüştür [133]. Jianfeng Gao ve arkadaşlarının çalışmasında, okuduğu kaynak belgeye dayanarak bir kullanıcıya ilgi duyacağı belgeleri tavsiye etmek için CNN tabanlı derin bir anlamsal benzerlik modeli (DSSM) sunulmuştur. DSSM, milyonlarca Web geçişi üzerinde eğitim alıp ve kaynak-hedef belge çiftlerini özellik vektörlerine eşleştirmiştir. Bu sayede kaynak belgeler ile ilgi duyulabilecek belgeler arasındaki mesafeler azaltılmış olmuştur. Model sadece ilgililik (interestingness) alanında başarılı olmamış, ayrıca konu modelleme konusunda da etkili sonuçlar üretmiştir [134]. Anlamsal ayrıştırma konusunda buna benzer başka yaklaşımlar da uygulanmıştır [77], [78]. Cümlelerin doğru ifade edilmesi dilin anlaşılması için önemlidir. CNN yöntemiyle yapılan bir çalışmada cümlelerin semantik modellemesi üzerinde çalışılmış ve %25 hata oranlarıyla ciddi bir başarı göstermiştir [79]. Doğal dil işleme alanında kullanıcının sözleri konsept/alana göre farklı etiketlenmelidir. Yapılan bir çalışmada veriler benzer olsa da, etiket veri setleri farklı olduğunda, mevcut ek açıklamalardan (annotation) yararlanmak için bir yaklaşım önerilmiştir. Bu yaklaşım CCA (canonical correlation analysis)'dan gelen etiket yerleştirmelerine dayanır. Sonuçlar incelendiğinde “takvim-hatırlatıcı” (Calendar-reminder) gibi önemli yakınlıklar ortaya çıkmıştır [135]. Başka bir çalışmada görüntü verilerindeki doğru bölgelere uygun tanımlayıcılar eklemek için görüntü üzerinde CNN, cümleler üzerinde çift yönlü RNN algoritmalarını kullanarak geliştirilen model, Flickr8K, Flickr30K ve MSCOCO veri kümeleri üzerinde uygulanmıştır. Bölge seviyesi ve tam görsel üzerinde üretilen tanımlayıcıların oldukça doğru olduğu görülmüştür [89]. Yine başka veri kümeleri üzerinde aynı problem benzer modeller ile çözülmüştür [136]. Mail spam tespiti de son yıllarda derin öğrenme alanında uygulama alanlarına girmiştir. DAE metoduyla yapılan bir çalışmada model, geleneksel yöntemlere göre çok daha iyi sonuçlar vermiştir [124].

Konuşma tanıma konusunda da DBN modeliyle yapılan çalışmalar ses kayıt verileri üzerinde etkileyici sonuçlar vermiştir [137],[138].

3.4.2 Görüntü/Video İşleme

Görüntü işleme alanında en popüler problemlerden biri görüntü sınıflandırma problemidir. Bu görev için ortak bir değerlendirme seti olarak genellikle MNIST

kullanılır. MNIST eğitim için 60000 ve test için 10000 el yazısı rakamlarından oluşmaktadır. Yapay sinir ağlarından, istatistiksel yöntemlere kadar farklı türde çok sayıda yöntem bu veri kümesi üzerinde test edilmiştir [139]. Bugüne kadarki en başarılı sonuç 0.23 hata ile Ciresan'ın 2012'de yaptığı çalışmadır. CNN modeli kullanılan bu çalışmada ön işlem olarak genişlik normalizasyonu uygulanmıştır [73]. Yine 2011'de CNN yaklaşımı, görsel desen tanıma yarışmasında ilk defa süper insan performansı elde etmiştir [63]. Ayrıca 2011 yılında ICDAR Çin el yazısı yarışması ve Mayıs 2012'de ISBI resim bölümlleme yarışması kazanılmıştır [62]. Ekim 2012'de, Alex Krizhevsky, Geoffrey Hinton ve ekibi tarafından oluşturulan benzer bir sistem, büyük ölçekli ImageNet yarışmasını, sıg makine öğrenme yöntemleri üzerinde belirgin bir farkla kazanmıştır [66]. Benzer bir model aynı zamanda kanser tespiti için büyük medikal görüntülerin analizi konusundaki ICPR ve MICCAI yarışmalarını kazanmıştır [140]. 2014'te derin öğrenmeyi kullanan ImageNet yarışmasındaki hata oranı, benzer derin öğrenme modelleri kullanılarak daha da azaltılmıştır [66], [141], [142]. Bunların yanında derin öğrenme tabanlı denetimsiz öğrenme algoritmaları da etiketsiz verilerden sınıfa özgü özellikler çıkarılması için kullanılmaktadır. Yapılan bir çalışmada havuzlama ve normalizasyon ile birlikte 9-katmanlı bir DAE ile internetten indirilmiş 10 milyon görüntü bulunan bir veri kümesi eğitilmiştir. 1000 makine (16.000 çekirdek) ile oluşturulan bir paralel küme ile eğitim 3 günde yapılmıştır. Bu model sayesinde yüz etiketlemesi yapmadan yüz tanıma sistemi, hatta kedi yüzü ve insan vücudu parçası da tanımlamak mümkün olmuştur [143].

Nesne tanıma problemleri için de derin öğrenme algoritmaları oldukça başarılıdır. Derin kalıntı ağları (deep residual network) kullanılan bir çalışma ILSVRC 2015, ImageNet algılama, ImageNet yerelleştirme, COCO segmentasyon yarışmalarında 1. olmuştur [144]. Bir başka çalışmada görüntüdeki nesneleri tanımlamak ve anlamsal segmentasyon yapmak için bölgeler ve CNN birleştirildiği R-CNN adı verilen algoritmayla zengin özellikler elde edilmiştir [145]. Sahne etiketleme problemi, bir görüntüdeki ham piksellerin ait olduğu nesne kategorisiyle etiketlenmesiyle çözülebilir. Clement Farabet ve ekibinin yaptığı çalışmada çoklu ölçekli (multiscale) CNN modeli önerilmiştir. SIFT Flow, Barcelona, Stanford arka plan veri kümeleri üzerinde diğer yaklaşımlara göre çok daha hızlı ve başarılı sonuçlar üretilmiştir [146]. Yayaların tespiti önemli bir sorundur. Bu konuda seyrek (sparse) CNN modeliyle geliştirilen denetimsiz öğrenme yöntemi rekabetçi sonuçlar vermiştir [147].

3.4.3 Biyomedikal Görüntü/Bilgi İşleme

Derin öğrenme, görüntü işlemenin bir alt dalı sayılabileceğinden, tıbbi görüntü analizi problemlerin çözümünde de oldukça etkilidir [148]. Beyin tümörü araştırmalarında, görüntü segmentasyonuna ihtiyaç duyulmaktadır. Bu görev için de son yıllarda derin öğrenme yöntemleriyle başarılı bir şekilde otomatik segmentasyon yapılmaktadır. Derin öğrenme modelleri, özellikle büyük miktarda MRI verileri ile nesnel yorumlar yapılmasına imkan sağlamıştır [149], [150]. Beyin tümörü için otomatik MRI Segmentasyonu konusunda özellikle CNN tabanlı algoritmalar, ayırt edici özellikler çıkararak başarılı sonuçlar üretmiştir [149], [151]. Segmentasyon problemi diğer

hastalıklar için çekilen MRI görüntülerinde de uygulanmıştır. Yapılan bir çalışmada düşük alan diz MRI taramalarında tibial kırıkdağın segmentasyonuna uygulanmış ve 114 görünmeyen taramada test edilmiştir [152]. Meme kanseri klinik bilgisayar teşhisinde önemli bir parametre olan SWE (share wave elastography) verisinden elde edilen görüntülerden iyi-kötü huylu tümörlerin tespiti için otomatik özellik çıkarımı amacıyla bir derin öğrenme modeli kurulmuştur. RBM ve PGBM (point-wise gated Boltzmann Machine) ile 2 katmanlı oluşturulan model 227 SWE verisiyle eğitilmiş, %93.4 doğruluk göstermiştir [153]. Başka bir çalışmada akciğer kanseri tespitinde otomatik çıkarılan özellikler kullanan derin öğrenme ile, geleneksel yöntemlerle (elle) çıkarılan özellikleri kullanan CADx sistemi karşılaştırılmıştır. Derin öğrenme ağı CNN, DBN ve DAE olmak üzere 3 kanallı oluşturulmuştur. Sonuçlar AUC (area under curve) değerine göre incelendiğinde CADx karşısında, CNN belirgin bir şekilde üstün çıkmış, DBN yakın ama üstün bir değer vermiş, DAE ise daha kötü sonuçlar göstermiştir [153].

İlaç endüstri için de aday ilaçların olumsuz sonuçlardan dolayı piyasaya sürülememesi büyük bir problemdir. Kimyasal bileşiklerdeki bu başarısızlıklar, yetersiz etkinlik göstermesi, hedefte olmayan bileşikler ile hatalı tepkimeye girmesi veya beklenmeyen toksik etkilerden kaynaklanmaktadır [154]. 2012 yılında çoklu görev derin sinir ağı ile George Dahl ve ekibi bileşiğin biyomoleküler hedef tahmini konusundaki “Merck Moleküler Aktivite Sorunu” [155] isimli bir yarışmayı kazanmışlardır [156]. 2014 yılında ise Sepp Hochreiter ve ekibi yine derin öğrenme kullanarak ilaçlardaki hedef dışı ve toksik etkileri tespit etmek için açılan bir yarışmayı kazanmışlardır [157]. Bu ve benzeri gelişmeler derin öğrenmenin sanal tarama konusunda üstün olabileceğini göstermiştir [158], [159]. 2015 yılında yapı temelli ilaç tasarımı için derin öğrenme kullanılan AtomNet, Atomwise şirketi tarafından piyasaya sürülmüş ve ardından ebola ve skleroz gibi hastalıklar için yeni aday biyomoleküllerin keşfinde kullanılmıştır [82] .

Genetik bilgilerin açıklanması biyoenformatik için önemli bir problemdir. Bu sorunun çözümü için DAE tabanlı bir çözüm önerilmiş ve standart makine öğrenmesi yöntemlerine göre daha iyi sonuçlar üretmiştir [160]. Başka bir çalışmada, LSTM, bir proteinin bir SCOP süper ailesine ait olarak sınıflandırılması gereken uzak protein homolojisi tespiti için iyi bilinen bir kritere uygulanmıştır. Model sınıflandırma için benzer sınıflandırma performansına sahip diğer yaklaşımlara göre doğruluk açısından yakın değerler verse de, onlara göre oldukça hızlıdır [99]. DNA ve RNA bağlayıcı proteinlerin dizilim özelliklerini bilmek, biyolojik sistemlerde düzenleyici süreç modelleri geliştirmek ve nedensel hastalık değişkenlerini tanımlamak için önem arz etmektedir. Yapılan bir çalışmada, dizi özelliklerinin desen keşfi için ölçeklenebilir, esnek ve birleşik hesaplama yaklaşımı sunan bir derin öğrenme modeli ile başarıyla kullanılmıştır [161].

3.4.4 Diğer Uygulamalar

Yukarıdaki bölümde en sık uygulama alanları verilmiştir. Bunların dışında yine farklı alanlarda derin öğrenme yöntemlerinin uygulandığı görülmüştür.

Uyku kalitesi tahmini için yapılan bir çalışmada bazı derin öğrenme yöntemleri (CNN, RNN, LSTM, vs.) ile geleneksel lojistik regresyonun öngörücülerin performansını karşılaştırılmıştır. CNN algoritması en iyi sonucu vermiştir [162]. Ayrıca insan aktivite tanıma ve insan hareket tespiti gibi konularda da derin öğrenme algoritmaları uygulanmıştır [97], [116].

CRM (Müşteri İlişkileri Yönetimi) otomasyonu konusunda da derin takviye öğrenme (deep reinforcement learning) yöntemi denenmiştir. RFM (güncellik-frekans-tutar) değişkenleri açısından tanımlanan müşteri durumu alanlarındaki muhtemel doğrudan pazarlama eylemlerinin değerini yaklaşık olarak hesaplamak için sinir ağı kullanılmıştır [163].

DARPA, doğal dil işlemenin gücünü kullanmak için Derin Arastırma ve Metin Filtreleme (DEFT) programını yaratmıştır. Bu yapay zeka, savunma analistlerinin büyük boyutlu belgeleri verimli bir şekilde araştırmalarını sağlayacak potansiyele sahip olduğu ve bu belgeler içinde örtülü olarak ifade edilen, aktivite içeren bilgileri keşfedebileceği (anormallik tespiti) açıklanmıştır [164].

Araç otonom sistemleri, multimedya yönetimi gibi konularda da derin öğrenme uygulamaları denenmekte, önümüzdeki yıllarda bu yönde gelişmeler beklenmektedir. NVIDIA tarafından üretilen derin öğrenme kartları ve süper bilgisayarlar bu hedefler için çalıştırılmaktadır [165].

Akıllı şehirlerin kurulmasında akıllı trafik planlaması önemli bir etkidir. Ulaşım ağındaki doğrusal olmayan ve rastgele özelliklerin üstesinden gelmek önemli bir zorluktur. Yapılan bir çalışmada DeepSense adı verilen derin öğrenme modeli ile taksi GPS verileriyle eğitilerek trafik akış tahmini yapılmıştır. Önerilen model ile mevcut yöntemlere göre %5 iyileştirme sağlanmıştır [166]. Yine başka bir çalışmada RBM kullanılarak trafik sıklığı modeli üzerinde çalışılmıştır [167].

Kapalı alan yönlendirmesi önemli bir problemdir. Yapılan bir çalışmada bir quadcopter bağımsız bir şekilde iç mekanlarda dolaştığı ve tek bir kamera kullanarak belli bir hedefi (örn, bir çantayı) bulması düşünülen sistem önerilmiştir. CNN algoritması ile uzman bir pilotun eylem seçimini taklit edilmiş, sistem performansını, çeşitli kapalı mekanlardaki gerçek zamanlı deneylerle gösterilmiştir .

3.5 Derin Öğrenme Çalışan Araştırma Ekipleri

Derin öğrenme konusunda birçok üniversite ve büyük çaplı şirketler çalışmalar yapmaktadır. Bunun yanında yine bu konulara yönelen, iyi işler çıkaran start-up şirketler de büyük firmalar tarafından satın alınmaktadır [67]. Başlıca çalışma grupları ve ekiplerin önde gelen isimleri aşağıda Tablo 1’de verilmiştir.

Araştırma Grupları ve Lab.	Ekip	Platformlar ve Kütüphaneler	Çalışma Alanları
Toronto Üniversitesi - Makine Öğrenmesi	Geofrey Hinton Ruslan Salakhutdinov	Torch	Doğal Dil İşleme Görüntü İşleme

Grubu	Roger Grosse		
Montréal Üniversitesi - MILA Lab.	Yoshua Bengio Pascal Vincent Christopher Pal	Theano Pylearn2 Blocks	Doğal Dil İşleme Sinyal İşleme Görüntü İşleme
New York Üniversitesi - CILVR Lab.	Yann Lecun Rob Fergus Leon Bottou	Theano C++ (CUDA)	Bilgisayar Algısı Doğal Dil İşleme Sağlık ve Robotik
Stanford Üniversitesi – SAIL ve SVL	Andrew Ng Christopher Manning Fei-fei Li	TensorFlow Java	Bilgisayarlı Görü Doğal Dil İşleme Robotik
Oxford Üniversitesi - Derin Öğrenme Grubu	Nando de Freitas Phil Blunsom Andrew Zisserman	-	Doğal Dil İşleme Görüntü İşleme
Kaliforniya Üniversitesi - BAIR	Pieter Abbeel Trevor Darrell Peter Bartlett	Caffe	Bilgisayarlı Görü Doğal Dil İşleme Robotik
Koç Üniversitesi - AI Lab.	Deniz Yuret	KNET	Doğal Dil İşleme Görüntü İşleme
Google - DeepMind	Demis Hassabis Mustafa Suleyman Shane Legg	Python	Sağlık Hizmetleri Doğal Dil İşleme
Google Research	Jeff Dean Corinna Cortes Úlfar Erlingsson	TensorFlow	Görüntü İşleme Bilgi Getirimi
Facebook - FAIR	Yann Lecun Larry Zitnick Antoine Bordes	Caffe2	Doğal Dil İşleme Bilgisayar-İnsan Etkileşimi
Twitter - Cortex	Hugo Larochelle Ryan Adams Clement Farabet	Torch	Twitter, Vine ve Periscope için teknoloji üretimi
Microsoft - DLTC	Xiaodong He Li Deng Jianfeng Gao	Microsoft Cognitive Toolkit Caffe	Doğal Dil İşleme Görüntü İşleme
İsviçre Uygulamalı Bilimler Üniversitesi	Jurgen Schmidhuber Luca M. Gambardella	-	Görüntü İşleme Robotik

(SUPSI) - IDSIA	Alessandro Antonucci		
Baidu - Derin Öğrenme Enstitüsü (IDL)	Lin Yuanqing Wu Zhongqin Wei Xu	-	Doğal Dil İşleme Görüntü İşleme

Tablo 1. Derin Öğrenme çalışan ekipler hakkında bilgiler

3.6 Derin Öğrenme Kütüphaneleri/Yazılımları

Derin öğrenme için geliştirilen çok farklı çerçeve ve kütüphaneler bulunmaktadır [168]. Bunlardan bazıları bu bölümde kısaca tanıtılacaktır.

Kütüphane	Yazıldığı Dil	Geliştirici
Theano	Python	MILA Lab
Caffe	Python	Berkeley Vision and Learning Center (BVLC)
Torch	Lua	Ronan Collobert, Clement Farabet, ve diğerleri
Digits	C++	NVIDIA
TensorFlow	Python	Google
DeepLearning4j	Java	Adam Gibson
KNET	Julia	Deniz Yuret

Tablo 2 Derin Öğrenme Kütüphaneleri

3.6.1 Theano

Theano, Montreal Üniversitesi'ndeki Yoshua Bengio tarafından en son teknoloji derin öğrenme algoritmalarının araştırılması ve geliştirilmesi için işletilen LISA grubu (şimdi MILA) tarafından geliştirilen bir Python . Theano, GPU'larda veya CPU'larda verimli çalışabilen bir sistemdir [169].

Derin bir öğrenme kütüphanesi değil, bir araştırma platformudur. İstenilen modelleri oluşturmak için kendi kodlarınızı yazmanız gerekmektedir [170]. Örneğin, sinir ağı sınıfları yoktur. Bununla birlikte, derin öğrenme için sınıflar ve metotlar oluşturmayı gösteren mükemmel bir derin öğrenme öğreticileri (tutorial) vardır (CNN, DBN, DAE, vs) [171].

Theano üzerinde Keras, Lazanya, Blocks, Pylearn2 gibi API'lerde bulunmaktadır.

3.6.2 Caffe

Caffe, Yangqing Jia tarafından denetimli bilgisayar görme problemleri için Berkeley Görme ve Öğrenme Merkezi'nde geliştirilen bir Python derin öğrenme kütüphanesidir[172]. Kütüphanenin en büyük yararı, Caffe Model Zoo üzerinden

indirilebilecek ve hemen kullanılacak önceden eğitilmiş ağların sayısıdır. Buna, standart bilgisayar görme veri kümeleri üzerinde dünya standartlarında sonuçlar verebilecek en modern modeller dahildir. Örneğin, dünya standartlarındaki modeller için bazı öğreticiler şunlardır:

Caffe ile Alex'in CIFAR-10 öğreticisi

MNIST'de Caffe ile LeNet Eğitimi

Caffe ile ImageNet

3.6.3 TensorFlow

TensorFlow, veri akış grafikleri kullanarak sayısal hesaplama için kullanılan açık kaynaklı bir yazılım kütüphanesidir. Esnek mimari, hesaplamayı tek bir API ile bir masaüstü, sunucu veya mobil cihazdaki bir veya daha fazla CPU'ya veya GPU'ya dağıtmanıza olanak tanır. TensorFlow başlangıçta, makine öğrenimi ve derin sinir ağları araştırması yürütmek amacıyla Google'ın Makine İstihbarat araştırma organizasyonu içindeki Google Beyin Ekibi üzerinde çalışan araştırmacılar ve mühendisler tarafından geliştirildi, ancak sistem, diğer alanlarda geniş bir yelpazede kullanılmaya başlanmıştır [173].

3.6.4 Torch

Torch, Ronan Collobert, Clement Farabet ve Koray Kavukcuoğlu tarafından derin öğrenme algoritmalarına yönelik araştırma ve geliştirme için geliştirilmiş bir Lua derin öğrenme çerçevedir [174]. CILVR Laboratuvarı tarafından kullanılmıştır. Ayrıca Torch Facebook AI laboratuvarı, Google DeepMind, Twitter ve birçoğu grup tarafından kullanılarak daha da geliştirilmiştir. Torch, GPU için CUDA'nın yanı sıra C / C ++ kitaplıklarını kullanmaktadır. Kullanıcı dostu bir arayüz sağlamak için C dostu dil Lua'yı benimseyen hızlı bir hedef vardır. Torch'un amacı, işlemi son derece basit hale getirirken bilimsel algoritmalarınızı oluşturma konusunda maksimum esnekliğe ve hıza sahip olmaktır [170].

Torch'un popüler uygulamaları, görüntü işleme için CNN, ve daha karmaşık alanlarda ise derin takviyeli öğrenmedir (deep reinforcement learning-DRL).

3.6.5 DeepLearning4j

DeepLearning4J (kısaca DL4J), ticari derin öğrenme projeleri için Adam Gibson tarafından Java dilinde geliştirilmiş derin öğrenme çerçevesidir [175]. DL4J, JVM tabanlı, endüstriye odaklı, ticari olarak desteklenen, dağıtılan derin öğrenme çerçevesidir ve büyük miktarda veriyi makul bir süre içinde çözen problemleri çözmeye yöneliktir. DBN, DAE, RNN, LSTM gibi birçok model için çözümler sunmaktadır [176]. JVM'deki diğer diller (ör. Scala) ve büyük veriler için platformlar (Hadoop ve Spark) de dahil olmak üzere yazılım geliştirmesinde baskın bir platform olan tüm Java ekosistemiyle çalışmanın büyük yararları vardır [170].

3.6.6 Digits

NVIDIA Derin Öğrenme GPU Eğitim Sistemi, DIGITS, görüntü sınıflandırma, segmentasyon ve nesne algılama görevleri için son derece hassas derin sinir ağını (DNN) hızlı bir şekilde eğitmek için kullanılabilir. DIGITS, veri yönetimi, çoklu GPU sistemleri üzerinde sinir ağları tasarımı ve eğitimi, gelişmiş görselleştirmelerle performansı gerçek zamanlı olarak izleme ve dağıtım için sonuç tarayıcıdan en iyi performans gösteren modeli seçme gibi ortak derin öğrenme görevlerini basitleştirir. DIGITS tamamen etkileşimli, böylece veri bilimcileri programlama ve hata ayıklama yerine ağ tasarımı ve eğitimine odaklanabilmektedir [177].

TARTIŞMA VE POTANSİYEL ÇALIŞMA ALANLARI

Önceki bölümlerde de anlatıldığı gibi derin öğrenme algoritmaları birçok alanda ve problemde kullanılmaktadır. Artan veri miktarıyla birlikte görüntü işleme alanında özellikle görüntü sınıflandırma, nesne tanıma gibi problemler yer almaktadır. Konu ile ilgili ILSVRC gibi yarışmalarda artık büyük veri ile başa çıkabilecek bir yöntem olarak kendisini ispatlayan derin öğrenme metotları sıklıkla kullanılmakta ve en başarılı sonuçları vermektedir. Bunun yanında otonom araç teknolojilerinde gerçek zamanlı veri işleme yeteneği ile GPU tabanlı derin öğrenme algoritmaları kullanılmaktadır. Araçlardaki 4-5 kamera ve birçok sensörden gelen verilerle başa çıkabilmek için derin öğrenme gelecekte de kullanılacak bir yol olarak görünmektedir.

Büyük verinin en önemli problemlerinden bir tanesi etiketleme maliyetidir. Bunu aşmak için yine gelecekte AE gibi denetimsiz derin öğrenme yöntemleri ön plana çıkacaktır.

Bir diğer dikkat çekici alan ise doğal dil işleme ve ses işleme olarak göze çarpmaktadır. Derin öğrenme algoritmaları, görüntü işleme ile doğal dil işleme konusunu bir araya getirmiş, özellikle görüntü etiketleme ve segmentleme, görüntülere başlık ekleme gibi konularda uygulanmıştır. Derin öğrenme üzerinde araştırma yapan ekiplerin neredeyse hepsi bu alanda geliştirmeler yapmaktadır (Tablo 1). Ancak yapılan çoğu çalışma İngilizce dili için yapıldığı görülmüştür. Özellikle Türkçe için bu alanlarda yapılan çalışmalar büyük önem arz etmektedir.

Doğal dil işleme konusunda metin özetleme, sınıflandırma gibi problemlerin çözümünde derin öğrenme algoritmalarının (özellikle RNN, LSTM, CNN gibi) başarı bir şekilde kullanılacağı düşünülmektedir. Bu sebeple çalışma konusu olarak, bu alana yönelmek, Türkçe için çözümler sunmak değerli olacaktır.

Bu amaç doğrultusunda metinlerden konu çıkarımı, anahtar kelimelerin belirlenmesi, soru-cevap sistemleri geliştirilmesi gibi problemler çözmek amacıyla bir çalışma yapılacaktır. Çalışmada kullanılacak olan veri kümesi bir çağrı merkezine ait operatör-müşteri görüşmelerinin metne dönüştürülmüş hallerinden oluşmaktadır. Bu problemlerin çözümünde derin öğrenme yöntemleri ve istatistiksel tabanlı yöntemler bir arada kullanılacaktır.

KAYNAKÇA

- [1] M. Copeland, “The Difference Between AI, Machine Learning, and Deep Learning?,” 2016. [Online]. Available: <https://blogs.nvidia.com/blog/2016/07/29/whats-difference-artificial-intelligence-machine-learning-deep-learning-ai/>. [Accessed: 17-Apr-2017].
- [2] J. Schmidhuber, “Deep learning in neural networks: An overview,” *Neural Networks*, vol. 61, pp. 85–117, 2015.
- [3] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, pp. 436–444, 2015.
- [4] A. L. Blum and P. Langley, “Selection of relevant features and examples in machine learning,” *Artif. Intell.*, vol. 97, no. 1–2, pp. 245–271, Dec. 1997.
- [5] Y. Bengio, A. Courville, and P. Vincent, “Representation Learning: A Review and New Perspectives,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 1798–1828, 2013.
- [6] E. Karabulut, “Investigation of deep learning approaches for biomedical data classification,” ÇUKUROVA UNIVERSITY, 2016.
- [7] O. Vinyals, “Beyond Deep Learning: Scalable Methods and Models for Learning,” University of California Berkeley, 2013.
- [8] A. Prasoon, “Deep Feature Learning and Cascaded Classifier for Large Scale Data,” University of Copenhagen , 2014.
- [9] R. Socher, “Recursive deep learning for natural language processing and computer vision,” Stanford University, 2014.
- [10] H. Özcan, “DEEP LEARNING APPLICATIONS FOR VERY LOW RESOLUTION FACE IMAGES,” DENİZ BİLİMLERİ VE MÜHENDİSLİĞİ ENSTİTÜSÜ, 2014.
- [11] A. V. Singh, “Content-Based Image Retrieval using Deep Learning,” Rochester Institute of Technology, 2015.
- [12] I. B. Levitan and L. K. Kaczmarek, *The Neuron: Cell and Molecular Biology*, Fourth Edi. New York: Oxford University Press, 2015.
- [13] E. Keleş, S. Çepni, A. Gör, K. Teknik Üniversitesi, F. Eğitim Fakültesi, and İ. Bölümü, “Beyin ve Öğrenme,” *Türk Fen Eğitimi Derg.*, vol. 3, no. 2, pp. 66–82, 2006.
- [14] I. J. Arreguín-González and A. De Revisión Resumen, “Sinapsis y memoria procedimental,” vol. 148, no. 3, 2013.
- [15] W. S. McCulloch and W. Pitts, “A logical calculus of the ideas immanent in

- nervous activity,” *Bull. Math. Biophys.*, vol. 5, no. 4, pp. 115–133, Dec. 1943.
- [16] D. Hebb, *The Organization of Behavior*. New York: Wiley & Sons, 1949.
 - [17] B. Farley and W. Clark, “Simulation of self-organizing systems by digital computer,” *Trans. IRE Prof. Gr. Inf. Theory*, vol. 4, no. 4, pp. 76–84, Sep. 1954.
 - [18] F. Rosenblatt and F., “The perceptron: A probabilistic model for information storage and organization in the brain,” *Psychol. Rev.*, vol. 65, no. 6, pp. 386–408, 1958.
 - [19] M. Minsky and S. Papert, *Perceptrons: an introduction to computational geometry*. MIT Press, 1988.
 - [20] P. J. Werbos, *Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences*. Harvard University, 1975.
 - [21] D. E. Rumelhart and J. L. McClelland, *Parallel distributed processing: explorations in the microstructure of cognition*. MIT Press, 1986.
 - [22] J. J. Hopfield, “Neural networks and physical systems with emergent collective computational abilities,” in *Proc. Natl. Acad. Sci.*, 1982, vol. 79, pp. 2554–2558.
 - [23] J. McClelland, D. Rumelhart, and P. R. Group, *Parallel distributed processing*. The MIT Press, 1987.
 - [24] N. Yadav, A. Yadav, and M. Kumar, “History of Neural Networks,” Springer Netherlands, 2015, pp. 13–15.
 - [25] T. Hill, L. Marquez, M. O’Connor, and W. Remus, “Artificial neural network models for forecasting and decision making,” *Int. J. Forecast.*, vol. 10, no. 1, pp. 5–15, Jun. 1994.
 - [26] G. P. Zhang, “Neural networks for classification: a survey,” *IEEE Trans. Syst. Man Cybern. Part C (Applications Rev.)*, vol. 30, no. 4, pp. 451–462, 2000.
 - [27] Mu-Chun Su, N. DeClaris, and Ta-Kang Liu, “Application of neural networks in cluster analysis,” in *1997 IEEE International Conference on Systems, Man, and Cybernetics. Computational Cybernetics and Simulation*, vol. 1, pp. 1–6.
 - [28] A. M. S. (Ali M. S. . Zalzal and A. S. Morris, *Neural networks for robotic control: theory and applications*. Ellis Horwood, 1996.
 - [29] W. Miller, P. Werbos, and R. Sutton, *Neural networks for control*. London: The MIT Press, 1995.
 - [30] J. C. Patra and A. C. Kot, “Nonlinear dynamic system identification using Chebyshev functional link artificial neural networks,” *IEEE Trans. Syst. Man Cybern. Part B*, vol. 32, no. 4, pp. 505–511, Aug. 2002.
 - [31] S. A. Kalogirou, “Applications of artificial neural-networks for energy systems,” *Appl. Energy*, vol. 67, no. 1, pp. 17–35, 2000.
 - [32] C. H. Dagli, “Intelligent manufacturing systems,” in *Artificial Neural Networks for Intelligent Manufacturing*, Dordrecht: Springer Netherlands, 1994, pp. 3–16.
 - [33] S. I.-J. Chien, Y. Ding, and C. Wei, “Dynamic Bus Arrival Time Prediction with Artificial Neural Networks,” *J. Transp. Eng.*, vol. 128, no. 5, pp. 429–438, Sep. 2002.
 - [34] R. M. Balabin and E. I. Lomakina, “Neural network approach to quantum-

- chemistry data: Accurate prediction of density functional theory energies,” *J. Chem. Phys.*, vol. 131, no. 7, p. 74104, Aug. 2009.
- [35] I. E. Lagaris, A. Likas, and D. I. Fotiadis, “Artificial neural network methods in quantum mechanics,” *Comput. Phys. Commun.*, vol. 104, no. 1–3, pp. 1–14, Aug. 1997.
 - [36] N. K. Ibrahim, R. S. A. R. Abdullah, and M. I. Saripan, “Artificial neural network approach in radar target classification,” *J. Comput. Sci.*, vol. 5, no. 1, pp. 23–32, 2009.
 - [37] Shang-Hung Lin, Sun-Yuan Kung, and Long-Ji Lin, “Face recognition/detection by probabilistic decision-based neural network,” *IEEE Trans. Neural Networks*, vol. 8, no. 1, pp. 114–132, 1997.
 - [38] H. A. Rowley, S. Baluja, and T. Kanade, “Neural network-based face detection,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 1, pp. 23–38, 1998.
 - [39] A. Sabri, A. H. Mohammad, B. Al-Shargabi, and M. A. Hamdeh, “Developing new Continuous Learning Approach for Spam Detection using Artificial Neural Network,” *Eur. J. Sci. Res.*, vol. 42, no. 3, 2010.
 - [40] D. Ndumiyana and L. Sakala, “Hidden Markov Models And Artificial Neural Networks For Spam Detection,” *Int. J. Eng. Res. Technol.*, vol. 2, no. 4, pp. 2069–2084, 2013.
 - [41] L. Özgür, T. Güngör, and F. Gürgen, “Adaptive anti-spam filtering for agglutinative languages: a special case for Turkish,” *Pattern Recognit. Lett.*, vol. 25, pp. 1819–1831, 2004.
 - [42] D. N. Ganesan, D. K. Venkatesh, D. M. A. Rama, and A. M. Palani, “Application of Neural Networks in Diagnosing Cancer Disease using Demographic Data,” *Int. J. Comput. Appl.*, vol. 1, no. 26, pp. 81–97, Feb. 2010.
 - [43] L. Bottaci *et al.*, “Artificial neural networks applied to outcome prediction for colorectal cancer patients in separate institutions,” *Lancet*, vol. 350, no. 9076, pp. 469–472, Aug. 1997.
 - [44] L. Álvarez Menéndez, F. J. de Cos Juez, F. Sánchez Lasheras, and J. A. Álvarez Riesgo, “Artificial neural networks applied to cancer detection in a breast screening programme,” *Math. Comput. Model.*, vol. 52, no. 7, pp. 983–991, 2010.
 - [45] I. Saritas, “Prediction of Breast Cancer Using Artificial Neural Networks,” *J. Med. Syst.*, vol. 36, no. 5, pp. 2901–2907, Oct. 2012.
 - [46] D. A. Pomerleau, “Efficient Training of Artificial Neural Networks for Autonomous Navigation,” *Neural Comput.*, vol. 3, no. 1, pp. 88–97, Feb. 1991.
 - [47] D. Zissis, E. K. Xidias, and D. Lekkas, “A cloud based architecture capable of perceiving and predicting multiple vessel behaviour,” *Appl. Soft Comput.*, vol. 35, pp. 652–661, 2015.
 - [48] Fu-Sheng Ho and P. Ioannou, “Traffic flow modeling and control using artificial neural networks,” *IEEE Control Syst. Mag.*, vol. 16, no. 5, pp. 16–26, 1996.
 - [49] LeCun Yann, Boser B, and Denker J S, “Handwritten Digit Recognition with a Back-Propagation Network,” 1989.
 - [50] L. Deng and D. Yu, “Deep Learning: Methods and Applications,” *Found.*

- Trends® Signal Process.*, vol. 7, no. 3–4, pp. 197–387, 2014.
- [51] Y. Bengio, “Learning Deep Architectures for AI,” *Found. trends® Mach. Learn.*, vol. 2, no. 1, pp. 1–127, 2009.
 - [52] H. A. Song and S.-Y. Lee, “Hierarchical Representation Using NMF,” in *International Conference on Neural Information Processing.*, 2013, pp. 466–473.
 - [53] A. G. Ivakhnenko and V. G. Lapa, “Cybernetic Predicting Devices,” 1966.
 - [54] Tim Dettmers, “Deep Learning in a Nutshell: History and Training Parallel Forall,” 2015. [Online]. Available: <https://devblogs.nvidia.com/parallelforall/deep-learning-nutshell-history-training/>. [Accessed: 20-Mar-2017].
 - [55] K. . N. Fukushima, “A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position.,” *Biol. Cybern.*, vol. 36, no. 4, pp. 193–202, 1980.
 - [56] Y. LeCun *et al.*, “Backpropagation Applied to Handwritten Zip Code Recognition,” *Neural Comput.*, vol. 1, no. 4, pp. 541–551, Dec. 1989.
 - [57] G. E. Hinton, P. Dayan, B. J. Frey, and R. M. Neal, “The wake-sleep algorithm for unsupervised neural networks,” *Science*, vol. 268, no. 5214, pp. 1158–61, May 1995.
 - [58] S. Hochreiter and J. Schmidhuber, “Long Short-Term Memory,” *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997.
 - [59] C. Cortes and V. Vapnik, “Support-Vector Networks,” *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995.
 - [60] I. N. Aizenberg, N. N. Aizenberg, and J. Vandewalle, “Multiple-Valued Threshold Logic and Multi-Valued Neurons,” in *Multi-Valued and Universal Binary Neurons*, Boston, MA: Springer US, 2000, pp. 25–80.
 - [61] G. E. Hinton, “Learning multiple layers of representation,” *Trends Cogn. Sci.*, vol. 11, no. 10, pp. 428–434, Oct. 2007.
 - [62] D. Cireşan, A. Giusti, L. M. Gambardella, and J. Schmidhuber, “Deep Neural Networks Segment Neuronal Membranes in Electron Microscopy Images,” in *Advances in neural information processing systems*, 2012, pp. 2843–2851.
 - [63] D. Cireşan, U. Meier, J. Masci, and J. Schmidhuber, “Multi-column deep neural network for traffic sign classification,” *Neural Networks*, vol. 32, pp. 333–338, 2012.
 - [64] D. C. Cireşan, U. Meier, L. M. Gambardella, and J. Schmidhuber, “Convolutional Neural Network Committees for Handwritten Character Classification,” in *2011 International Conference on Document Analysis and Recognition*, 2011, pp. 1135–1139.
 - [65] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov, “Improving Neural Networks by Preventing Co-adaptation of Feature Detectors,” *Neural Evol. Comput.*, Jul. 2012.
 - [66] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks,” in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

- [67] “The Race For AI: Google, Twitter, Intel, Apple In A Rush To Grab Artificial Intelligence Startups,” *cbinsights*, 2017. [Online]. Available: <https://www.cbinsights.com/blog/top-acquirers-ai-startups-ma-timeline/>. [Accessed: 21-Apr-2017].
- [68] D. H. Hubel and T. N. Wiesel, “Receptive fields and functional architecture of monkey striate cortex,” *J. Physiol.*, vol. 195, no. 1, pp. 215–243, Mar. 1968.
- [69] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition,” *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.
- [70] D. Scherer, A. Müller, and S. Behnke, “Evaluation of Pooling Operations in Convolutional Architectures for Object Recognition,” Springer, Berlin, Heidelberg, 2010, pp. 92–101.
- [71] “File:Max pooling.png - Wikimedia Commons.” [Online]. Available: https://commons.wikimedia.org/wiki/File:Max_pooling.png. [Accessed: 26-Apr-2017].
- [72] Y. Le Cun *et al.*, “Handwritten digit recognition: applications of neural network chips and automatic learning,” *IEEE Commun. Mag.*, vol. 27, no. 11, pp. 41–46, Nov. 1989.
- [73] D. Cireşan, U. Meier, and J. Schmidhuber, “Multi-column Deep Neural Networks for Image Classification,” Feb. 2012.
- [74] D. C. Cirean, U. Meier, J. Masci, and L. M. Gambardella, “Flexible, High Performance Convolutional Neural Networks for Image Classification,” in *Proceedings of the Twenty-Second international joint conference on Artificial Intelligence*, 2012, pp. 1237–1242.
- [75] “Results of ILSVRC2014,” 11-Nov-2014. [Online]. Available: <http://arxiv.org/abs/1311.2524>. [Accessed: 26-Apr-2017].
- [76] S. S. Farfade, M. Saberian, and L.-J. Li, “Multi-view Face Detection Using Deep Convolutional Neural Networks,” in *Proceedings of the 5th ACM on International Conference on Multimedia Retrieval.*, 2015, pp. 643–650.
- [77] E. Grefenstette, P. Blunsom, N. de Freitas, and K. M. Hermann, “A Deep Architecture for Semantic Parsing,” Apr. 2014.
- [78] Y. Shen, X. He, J. Gao, L. Deng, and G. Mesnil, “Learning semantic representations using convolutional neural networks for web search,” in *Proceedings of the 23rd International Conference on World Wide Web - WWW '14 Companion*, 2014, pp. 373–374.
- [79] N. Kalchbrenner, E. Grefenstette, and P. Blunsom, “A Convolutional Neural Network for Modelling Sentences,” Apr. 2014.
- [80] Y. Kim, “Convolutional Neural Networks for Sentence Classification,” Aug. 2014.
- [81] R. Collobert and J. Weston, “A unified architecture for natural language processing: Deep neural networks with multitask learning,” in *Proceedings of the 25th international conference on Machine learning - ICML '08*, 2008, vol. 20, no. 1, pp. 160–167.
- [82] I. Wallach, M. Dzamba, and A. Heifets, “AtomNet: A Deep Convolutional

- Neural Network for Bioactivity Prediction in Structure-based Drug Discovery,” Oct. 2015.
- [83] J. Yosinski, J. Clune, A. Nguyen, T. Fuchs, and H. Lipson, “Understanding Neural Networks Through Deep Visualization,” Jun. 2015.
 - [84] C. J. Maddison, A. Huang, I. Sutskever, and D. Silver, “Move Evaluation in Go Using Deep Convolutional Neural Networks,” Dec. 2014.
 - [85] C.-S. Lee *et al.*, “Human vs. Computer Go: Review and Prospect [Discussion Forum],” *IEEE Comput. Intell. Mag.*, vol. 11, no. 3, pp. 67–72, Aug. 2016.
 - [86] J. L. Elman, “Finding Structure in Time,” *Cogn. Sci.*, vol. 14, no. 2, pp. 179–211, Mar. 1990.
 - [87] T. Mikolov, “Recurrent neural network based language model,” in *Interspeech*, 2010.
 - [88] A. Graves and N. Jaitly, “Towards End-To-End Speech Recognition with Recurrent Neural Networks,” in *ICML*, 2014, pp. 1764–1772.
 - [89] A. Karpathy and L. Fei-Fei, “Deep Visual-Semantic Alignments for Generating Image Descriptions,” in *CVPR*, 2015, pp. 3128–3137.
 - [90] Y. Bengio, P. Simard, and P. Frasconi, “Learning long-term dependencies with gradient descent is difficult,” *IEEE Trans. Neural Networks*, vol. 5, no. 2, pp. 157–166, Mar. 1994.
 - [91] K. Greff, R. K. Srivastava, J. Koutník, B. R. Steunebrink, and J. Schmidhuber, “LSTM: A Search Space Odyssey,” Mar. 2015.
 - [92] F. A. Gers, J. Schmidhuber, and F. Cummins, “Learning to forget: continual prediction with LSTM,” *Neural Comput.*, vol. 12, no. 10, pp. 2451–71, Oct. 2000.
 - [93] F. A. Gers and J. Schmidhuber, “Recurrent nets that time and count,” in *Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks. IJCNN 2000. Neural Computing: New Challenges and Perspectives for the New Millennium*, 2000, pp. 189–194 vol.3.
 - [94] A. Graves and J. Schmidhuber, “Framewise phoneme classification with bidirectional LSTM and other neural network architectures,” *Neural Networks*, vol. 18, no. 5–6, pp. 602–610, Jul. 2005.
 - [95] A. Graves, A. Mohamed, and G. Hinton, “Speech recognition with deep recurrent neural networks,” in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, 2013, pp. 6645–6649.
 - [96] S. Fernández, A. Graves, and J. Schmidhuber, “An Application of Recurrent Neural Networks to Discriminative Keyword Spotting,” in *International Conference on Artificial Neural Networks*, 2007, pp. 220–229.
 - [97] M. Baccouche, F. Mamalet, C. Wolf, C. Garcia, and A. Baskurt, “Sequential Deep Learning for Human Action Recognition,” Springer, Berlin, Heidelberg, 2011, pp. 29–39.
 - [98] D. Eck and J. Schmidhuber, “Learning the Long-Term Structure of the Blues,” Springer, Berlin, Heidelberg, 2002, pp. 284–289.
 - [99] S. Hochreiter, M. Heusel, and K. Obermayer, “Fast model-based protein

- homology detection without alignment,” *Bioinformatics*, vol. 23, no. 14, pp. 1728–1736, Jul. 2007.
- [100] H. Mayer, F. Gomez, D. Wierstra, I. Nagy, A. Knoll, and J. Schmidhuber, “A System for Robotic Heart Surgery that Learns to Tie Knots Using Recurrent Neural Networks,” in *2006 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2006, pp. 543–548.
 - [101] J. Schmidhuber, F. Gers, and D. Eck, “Learning Nonregular Languages: A Comparison of Simple Recurrent Networks and LSTM,” *Neural Comput.*, vol. 14, no. 9, pp. 2039–2041, Sep. 2002.
 - [102] A. Graves, “Offline Handwriting Recognition with Multidimensional Recurrent Neural Networks.”
 - [103] P. Smolensky, “Information Processing in Dynamical Systems: Foundations of Harmony Theory,” 1986.
 - [104] G. E. Hinton and R. R. Salakhutdinov, “Reducing the Dimensionality of Data with Neural Networks,” *Science (80-.)*, vol. 313, no. 5786, pp. 504–507, Jul. 2006.
 - [105] R. Salakhutdinov and G. Hinton, “Deep Boltzmann Machines,” in *International Conference on Artificial Intelligence and Statistics*, 2009, pp. 3–11.
 - [106] M. A. Carreira-Perpiñán and G. E. Hinton, “On Contrastive Divergence Learning,” *Artif. Intell. Stat.*, vol. 10, 2005.
 - [107] H. Larochelle and Y. Bengio, “Classification using discriminative restricted Boltzmann machines,” in *Proceedings of the 25th international conference on Machine learning - ICML '08*, 2008, pp. 536–543.
 - [108] R. Salakhutdinov, A. Mnih, and G. Hinton, “Restricted Boltzmann machines for collaborative filtering,” in *Proceedings of the 24th international conference on Machine learning - ICML '07*, 2007, pp. 791–798.
 - [109] A. Coates, A. Ng, and H. Lee, “An Analysis of Single-Layer Networks in Unsupervised Feature Learning,” in *PMLR*, 2011, pp. 215–223.
 - [110] G. E. Hinton and R. R. Salakhutdinov, “Replicated Softmax: an Undirected Topic Model,” in *Advances in Neural Information Processing Systems 22*, 2009, pp. 1607–1614.
 - [111] H. Sarkar, “Deep Belief Network,” 2013. [Online]. Available: <https://agollp.wordpress.com/2013/12/29/deep-belief-network/>. [Accessed: 27-Apr-2017].
 - [112] G. E. Hinton, S. Osindero, and Y.-W. Teh, “A Fast Learning Algorithm for Deep Belief Nets,” *Neural Comput.*, vol. 18, no. 7, pp. 1527–1554, Jul. 2006.
 - [113] M. Ranzato, F. J. Huang, Y.-L. Boureau, and Y. LeCun, “Unsupervised Learning of Invariant Feature Hierarchies with Applications to Object Recognition,” in *2007 IEEE Conference on Computer Vision and Pattern Recognition*, 2007, pp. 1–8.
 - [114] Y. Bengio, P. Lamblin, D. Popovici, and H. Larochelle, “Greedy layer-wise training of deep networks,” *Proceedings of the 19th International Conference on Neural Information Processing Systems*. MIT Press, pp. 153–160, 2006.

- [115] R. Salakhutdinov and G. Hinton, "Semantic hashing," *Int. J. Approx. Reason.*, vol. 50, pp. 969–978, 2009.
- [116] G. W. Taylor, G. E. Hinton, and S. T. Roweis, "Modeling Human Motion Using Binary Latent Variables," in *Advances in Neural Information Processing Systems 19*, 2006, pp. 1345–1352.
- [117] C.-Y. Liou, W.-C. Cheng, J.-W. Liou, and D.-R. Liou, "Autoencoder for words," *Neurocomputing*, vol. 139, pp. 84–96, 2014.
- [118] A. B. L. Larsen and S. K. Sønderby, "Generating Faces with Torch," 13-Nov-2015. [Online]. Available: <http://arxiv.org/abs/1506.05751>. [Accessed: 27-Apr-2017].
- [119] D. P. Kingma and M. Welling, "Auto-Encoding Variational Bayes," Dec. 2013.
- [120] A. Krizhevsky and G. E. Hinton, "Using Very Deep Autoencoders for Content Based Image Retrieval," in *European Symposium on Artificial Neural Networks*, 2011, pp. 489–494.
- [121] "Autoencoder structure," 2015. [Online]. Available: https://commons.wikimedia.org/wiki/File:Autoencoder_structure.png. [Accessed: 27-Apr-2017].
- [122] A. Seker and A. Gürkan YUKSEK, "Stacked Autoencoder Method for Fabric Defect Detection," *Sci. Sci. J.*, vol. 38, no. 2, 2017.
- [123] "Stacked Autoencoders," *Stanford University*, 2013. [Online]. Available: http://ufldl.stanford.edu/wiki/index.php/Stacked_Autoencoders. [Accessed: 27-Apr-2017].
- [124] G. Mi, Y. Gao, and Y. Tan, "Apply Stacked Auto-Encoder to Spam Detection," *Adv. Swarm Comput. Intell.*, vol. 9141, pp. 3–15, 2015.
- [125] L. Deng, M. Seltzer, D. Yu, A. Acero, A. Mohamed, and G. Hinton, "Binary Coding of Speech Spectrograms Using a Deep Auto-encoder," in *Interspeech 2010*, 2010, p. 1692–1695.
- [126] S. Zhou, Q. Chen, and X. Wang, "Active deep learning method for semi-supervised sentiment classification," *Neurocomputing*, vol. 120, pp. 536–546, 2013.
- [127] S. Zhou, Q. Chen, and X. Wang, "Fuzzy deep belief networks for semi-supervised sentiment classification," *Neurocomputing*, vol. 131, pp. 312–322, 2014.
- [128] Y. Goldberg and O. Levy, "word2vec Explained: deriving Mikolov et al.'s negative-sampling word-embedding method," Feb. 2014.
- [129] L. You, Y. Li, Y. Wang, J. Zhang, and Y. Yang, "A deep learning-based RNNs model for automatic security audit of short messages," in *2016 16th International Symposium on Communications and Information Technologies (ISCIT)*, 2016, pp. 225–229.
- [130] M. Yousefi-Azar and L. Hamey, "Text Summarization Using Unsupervised Deep Learning," *Expert Syst. Appl.*, vol. 68, pp. 93–105, Feb. 2017.
- [131] A. Sboev, T. Litvinova, D. Gudovskikh, R. Rybka, and I. Moloshnikov, "Machine Learning Models of Text Categorization by Author Gender Using

- Topic-independent Features,” *Procedia Comput. Sci.*, vol. 101, pp. 135–142, 2016.
- [132] M. L. Brocardo, I. Traore, I. Woungang, and M. S. Obaidat, “Authorship verification using deep belief network systems,” *Int. J. Commun. Syst.*, p. e3259, 2017.
 - [133] Y. Shen, X. He, J. Gao, L. Deng, and G. Mesnil, “A Latent Semantic Model with Convolutional-Pooling Structure for Information Retrieval,” in *CIKM*, 2014.
 - [134] J. Gao, L. Deng, M. Gamon, and X. He, “Modeling interestingness with deep neural networks,” 14/304,863, 2014.
 - [135] Y.-B. Kim, K. Stratos, R. Sarikaya, and M. Jeong, “New Transfer Learning Techniques For Disparate Label Sets,” in *Association for Computational Linguistics*, 2015.
 - [136] O. Vinyals, A. Toshev, S. Bengio, and D. Erhan, “Show and Tell: A Neural Image Caption Generator,” *arXiv*, vol. 32, no. 1, pp. 1–10, Nov. 2014.
 - [137] A.-R. Mohamed, T. N. Sainath, G. Dahl, B. Ramabhadran, G. E. Hinton, and M. A. Picheny, “DEEP BELIEF NETWORKS USING DISCRIMINATIVE FEATURES FOR PHONE RECOGNITION,” in *Acoustics, Speech and Signal Processing*, 2011, p. 5060–5063.
 - [138] G. Hinton *et al.*, “Deep Neural Networks for Acoustic Modeling in Speech Recognition: The Shared Views of Four Research Groups,” *IEEE Signal Process. Mag.*, vol. 29, no. 6, pp. 82–97, Nov. 2012.
 - [139] Y. LeCun, C. Cortes, and C. Burges, “MNIST handwritten digit database.” [Online]. Available: <http://yann.lecun.com/exdb/mnist/>. [Accessed: 30-Apr-2017].
 - [140] D. C. Cireřan, A. Giusti, L. M. Gambardella, and J. Schmidhuber, “Mitosis Detection in Breast Cancer Histology Images with Deep Neural Networks,” in *Medical Image Computing and Computer-Assisted Intervention*, 2013, pp. 411–418.
 - [141] “ImageNet,” *Stanford Vision Lab*, 2016. [Online]. Available: <http://image-net.org/>. [Accessed: 30-Apr-2017].
 - [142] O. Russakovsky *et al.*, “ImageNet Large Scale Visual Recognition Challenge,” *Int. J. Comput. Vis.*, vol. 115, no. 3, pp. 211–252, Dec. 2015.
 - [143] Q. V. Le *et al.*, “Building high-level features using large scale unsupervised learning,” *arXiv Prepr. arXiv1112.6209*, Dec. 2011.
 - [144] K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition,” in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.
 - [145] R. Girshick, J. Donahue, T. Darrell, and J. Malik, “Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation,” in *The IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 580–587.
 - [146] C. Farabet, C. Couprie, L. Najman, and Y. LeCun, “Learning Hierarchical Features for Scene Labeling,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 1915–1929, Aug. 2013.

- [147] P. Sermanet, K. Kavukcuoglu, S. Chintala, and Y. Lecun, "Pedestrian Detection with Unsupervised Multi-stage Feature Learning," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 3626–3633.
- [148] S. K. Zhou, H. Greenspan, and D. Shen, *Deep learning for medical image analysis*. .
- [149] M. Havaei *et al.*, "Brain tumor segmentation with Deep Neural Networks," *Med. Image Anal.*, vol. 35, pp. 18–31, Jan. 2017.
- [150] A. Işın, C. Direkoğlu, and M. Şah, "Review of MRI-based Brain Tumor Image Segmentation Using Deep Learning Methods," *Procedia Comput. Sci.*, vol. 102, pp. 317–324, 2016.
- [151] G. Urban, M. Bendszus, F. A. Hamprecht, and J. Kleesiek, "Multi-modal Brain Tumor Segmentation using Deep Convolutional Neural Networks," in *MICCAI BraTS (Brain Tumor Segmentation) Challenge. Proceedings, winning contribution*, 2014, pp. 31–35.
- [152] A. Prasoon, K. Petersen, C. Igel, F. Lauze, E. Dam, and M. Nielsen, "Deep Feature Learning for Knee Cartilage Segmentation Using a Triplanar Convolutional Neural Network," in *MICCAI*, 2013, pp. 246–253.
- [153] Q. Zhang *et al.*, "Deep learning based classification of breast tumors with shear-wave elastography," *Ultrasonics*, vol. 72, pp. 150–157, 2016.
- [154] J. Arrowsmith and P. Miller, "Trial Watch: Phase II and Phase III attrition rates 2011–2012," *Nat. Rev. Drug Discov.*, vol. 12, no. 8, pp. 569–569, Aug. 2013.
- [155] "Merck Molecular Activity Challenge," *Kaggle*. [Online]. Available: <https://www.kaggle.com/c/MerckActivity/details/winners>. [Accessed: 30-Apr-2017].
- [156] G. E. Dahl, N. Jaitly, and R. Salakhutdinov, "Multi-task Neural Networks for QSAR Predictions," Jun. 2014.
- [157] "Tox21 Data Challenge Final Subchallenge Leaderboard," 2014. [Online]. Available: <https://tripod.nih.gov/tox21/challenge/leaderboard.jsp>. [Accessed: 01-May-2017].
- [158] T. Unterthiner, A. Mayr, M. J. At Unter Klambauer, K. J. At, S. Hochreiter, and H. J. At, "Toxicity Prediction using Deep Learning."
- [159] T. Unterthiner, A. Mayr, M. Steijaert OpenAnalytics, J. K. Wegner Johnson, H. Ceulemans, and S. Hochreiter, "Deep Learning as an Opportunity in Virtual Screening," in *Workshop on Deep Learning and Representation Learning*, 2014.
- [160] D. Chicco, P. Sadowski, and P. Baldi, "Deep autoencoder neural networks for gene ontology annotation predictions," in *Proceedings of the 5th ACM Conference on Bioinformatics, Computational Biology, and Health Informatics - BCB '14*, 2014, pp. 533–540.
- [161] B. Alipanahi, A. Delong, M. Weirauch, and B. Frey, "Predicting the sequence specificities of DNA-and RNA-binding proteins by deep learning," *Nat. Biotechnol.*, 2015.
- [162] A. Sathyanarayana *et al.*, "Sleep Quality Prediction From Wearable Data Using Deep Learning," *JMIR mHealth uHealth*, vol. 4, no. 4, p. e125, Nov. 2016.

- [163] Y. Tkachenko, “Autonomous CRM Control via CLV Approximation with Deep Reinforcement Learning in Discrete and Continuous Action Space,” *ArXiv*, Apr. 2015.
- [164] B. Onyshkevych, “Deep Exploration and Filtering of Text (DEFT).” [Online]. Available: <http://www.darpa.mil/program/deep-exploration-and-filtering-of-text>. [Accessed: 01-May-2017].
- [165] “The NVIDIA DGX-1 Deep Learning System,” *NVIDIA*. [Online]. Available: <http://www.nvidia.com/object/deep-learning-system.html>. [Accessed: 01-May-2017].
- [166] X. Niu, Y. Zhu, and X. Zhang, “DeepSense: A novel learning mechanism for traffic prediction with taxi GPS traces,” in *2014 IEEE Global Communications Conference*, 2014, pp. 2745–2750.
- [167] X. Ma, H. Yu, Y. Wang, Y. Wang, and M. González, “Large-Scale Transportation Network Congestion Evolution Prediction Using Deep Learning Theory,” *PLoS One*, vol. 10, no. 3, p. e0119044, Mar. 2015.
- [168] “Comparison of deep learning software.” [Online]. Available: https://en.wikipedia.org/wiki/Comparison_of_deep_learning_software. [Accessed: 01-May-2017].
- [169] Theano Development Team, “Theano: A {Python} framework for fast computation of mathematical expressions,” *arXiv e-prints*, vol. abs/1605.02688.
- [170] Jason Brownlee, “Popular Deep Learning Libraries,” 2016. [Online]. Available: <http://machinelearningmastery.com/popular-deep-learning-libraries/>. [Accessed: 01-May-2017].
- [171] “DeepLearning 0.1 documentation.” [Online]. Available: <http://deeplearning.net/tutorial/contents.html>. [Accessed: 01-May-2017].
- [172] Y. Jia *et al.*, “Caffe: Convolutional Architecture for Fast Feature Embedding,” Jun. 2014.
- [173] “TensorFlow.” [Online]. Available: <https://www.tensorflow.org/>. [Accessed: 01-May-2017].
- [174] R. Collobert, C. Farabet, and K. Kavukcuoglu, “Torch | Scientific computing for LuaJIT,” *NIPS Workshop on Machine Learning Open Source Software*, 2008. [Online]. Available: <http://torch.ch/>. [Accessed: 01-May-2017].
- [175] “Deeplearning4j: Open-source, Distributed Deep Learning for the JVM.” [Online]. Available: <https://deeplearning4j.org/>. [Accessed: 01-May-2017].
- [176] “Skymind - Deep learning for Enterprise on Hadoop and Spark.” [Online]. Available: <https://skymind.ai/>. [Accessed: 01-May-2017].
- [177] “NVIDIA DIGITS.” [Online]. Available: <https://developer.nvidia.com/digits>. [Accessed: 01-May-2017].

ÖZGEÇMİŞ

KİŞİSEL BİLGİLER

Adı Soyadı :Abdulkadir ŞEKER
Doğum Tarihi ve Yeri :21.11.1988 Gülşehir
Yabancı Dili :İngilizce
E-posta :aseker@cumhuriyet.edu.tr

ÖĞRENİM DURUMU

Derece	Alan	Okul/Üniversite	Mezuniyet Yılı
Y. Lisans	Elektrik ve Bilgisayar Mühendisliği	Erciyes Üniversitesi	2016
Lisans	Bilgisayar Mühendisliği	Karadeniz Teknik Üniv..	2012
Lise	Sayısal	Prof.Dr.Necati Erşen AÖL	2007

İŞ TECRÜBESİ

Yıl	Firma/Kurum	Görevi
2013	Cumhuriyet Üniversitesi	Araştırma Görevlisi
2012	Netsim Yazılım	Yazılım Geliştirici

YAYINLARI

Bildiri

1. 2016 ŞEKER A. ,AMASYALI M.F., Semi supervised Learning Based on Ensemble Algorithm. International Conference on Computer Science and Engineering, 393-401.
2. 2016 TAKCI H.,ŞEKER A., Educational Data Mining For Pre Detection Of Qualified Personnel. International Conference on Computer Science and Engineering, 398-404
3. 2016 DELİBAŞ E.,ŞEKER A., Üniversite Öğrencilerinin Başarıları Arasındaki İlişkinin Birliktelik Analizi İle İncelenmesi. 5th International Vocational Schools Symposium, 2, 1045-1050

Proje

1. 2015-2016 Derin Öğrenme Yaklaşımı ile Otomatik Kumaş Hata Tespiti Sistemi İçin Performansı ve Verimi Yüksek Algoritma Geliştirme / BAP Destekli Proje / Araştırmacı
- 2.2014-2015 Robot Laboratuvarlarının Kurulması ve Robot Eğitimi verilmesi / Avrupa Birliği Destekli Proje, Araştırmacı