

**T.C.
HARRAN ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

YÜKSEK LİSANS TEZİ

**PARKİNSON HASTALIĞININ TEŞHİSİNDE SES VERİLERİ ÜZERİNDEN
MAKİNE ÖĞRENMESİ VE DERİN ÖĞRENME YÖNTEMLERİNİN
İNCELENMESİ**

İbrahim Halil GÖKÇİN

BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI

**ŞANLIURFA
2023**

İÇİNDEKİLER

Sayfa No

ÖZET	i
ABSTRACT	ii
TEŞEKKÜR	iii
ŞEKİLLER DİZİNİ	iv
ÇİZELGELER DİZİNİ	v
GİRİŞ	1
1.1.Parkinson Hastalığı	4
1.1.1. Genel bakış	4
1.1.2. Parkinson hastalığı tipleri ve belirtileri	7
1.1.3. Parkinson hastalığının teşhisi	8
1.2. 1.2. Makine Öğrenmesi	9
1.2.1. Öğrenme türleri	10
1.2.2. Sinir ağı (NN)	11
1.2.3.Yapay Sinir Ağları (YSA)	11
1.2.4. Karar ağaçları	15
1.2.5. K-en yakın komşu (EYKA)	18
1.2.6. Destek vektör makineleri (DVM)	21
1.3. 1.3. Performans Metrikleri	38
1.4. 1.3. Parkinson Hastalığı Teşhisi	39
1.3.1. Çizim muayenesi	39
1.3.2. Ses muayenesi	41
1.5. 1.4. Özellik Seçimi	42
1.4.1. Arama stratejisi	42
1.4.2. Arama yönü	43
1.6. 1.5. Rastgele Aşırı Örnekleme	43
2. ÖNCEKİ ÇALIŞMALAR	44
3. MATERYAL ve YÖNTEM	54
1.7. 3.1.Materyal	54
1.8. 3.2.Yöntem	56
4. ARAŞTIRMA BULGULARI ve TARTIŞMA	58
1.9. 4.1.Özellik Seçimi	60
5. SONUÇLAR ve ÖNERİLER	65
KAYNAKLAR	68
ÖZGEÇMİŞ	73

ÖZET

Yüksek Lisans Tezi

PARKİNSON HASTALIĞININ TEŞHİSİNDE SES VERİLERİ ÜZERİNDEN MAKİNE ÖĞRENMESİ VE DERİN ÖĞRENME YÖNTEMLERİNİN İNCELENMESİ

İbrahim Halil GÖKÇİN

Harran Üniversitesi
Fen Bilimleri Enstitüsü
Bilgisayar Mühendisliği Anabilim Dalı

Danışman: Dr.Öğr.Üyesi Nagehan İLHAN
Yıl: 2023, Sayfa: 73

Günümüzde özellikle makine öğrenmesindeki gelişmelerle birlikte ses bozuklukları ve konuşma ölçümleri kapsamında yapılan çalışmalar da artış olmuştur. Özellikle Parkinson gibi konuşma bozukluğu semptomu olan hastalıklar için farklı yöntemler kullanılarak seslerin analizi yöntemiyle sınıflandırma çalışmaları yapılmaktadır. Parkinson hastalığı için yapılan çalışmalara bakıldığında genel olarak çalışmaların makine öğrenmesi metotlarıyla seslerin ve hasta el yazılarının değerlendirilmesi üzerine yoğunlaşmış olduğu görülmektedir. Bu tezin amacı Şakar ve ark. (2019) tarafından oluşturulan “Parkinson's Disease Classification (PDC)” veri seti kullanılarak makine öğrenme ve derin öğrenme yöntemlerinin performansını değerlendirmektir. Çalışmanın diğer çalışmalardan farkı ise daha fazla makine öğrenmesi tekniği kullanarak bu tekniklerin PDC veri seti üzerindeki başarısını test etmektir. Literatürde PD sınıflandırma problemi için algoritma sonuçlarına göre bu çalışmada daha yüksek doğruluk oranlarının rapor edildiği görülmektedir. Çalışma sonucunda en iyi sonuçların Konvolüsyonel Sinir Ağları (KSA) algoritmasından elde edildiği görülmektedir. Gelecekteki çalışmalarda PH sınıflandırma probleminde umut verici sonuçlar gösteren KSA tekniği, PH tele izleme sistemi oluşturmak için PH hastalarının Birleşik Parkinson Hastalığı Derecelendirme Ölçeği puanını tahmin etmek için kullanılabilir.

ANAHTAR KELİMELER: Parkinson hastalığı, makine öğrenmesi, derin öğrenme.

ABSTRACT

MSc Thesis

INVESTIGATION OF MACHINE LEARNING AND DEEP LEARNING METHODS ON VOICE DATA IN DIAGNOSIS OF PARKINSON'S DISEASE

İbrahim Halil GÖKÇİN

**Harran University
Graduate School of Natural and Applied Sciences
Department of Computer Engineering**

**Supervisor: Assoc Prof.Dr. Nagehan İLHAN
Year: 2023, Page: 73**

Today, especially with the developments in machine learning, there has been an increase in the studies carried out within the scope of voice disorders and speech measurements. Especially for diseases with speech impairment symptoms such as Parkinson's, classification studies are carried out by using different methods and analysis of sounds. When the studies for Parkinson's disease are examined, it is seen that the studies are generally focused on the evaluation of sounds and patient handwriting with machine learning methods. The aim of this study was to evaluate the performance of machine learning methods using the "Parkinson's Disease Classification (PDC)" dataset created by Şakar et al. (2019). The difference of the study from other studies is to test the success of these techniques on the PDC dataset by using more machine learning and deep learning techniques. According to the algorithm results for the PD classification problem in the literature, it is seen that higher accuracy rates are reported in this study. As a result of the study, it is seen that the best results are obtained from the Convolutional Neural Networks (KSA) algorithm. The KSA technique, which has shown promising results in the PH classification problem in future studies, can be used to estimate the Unified Parkinson's Disease Rating Scale score of PH patients to create a PH tele-monitoring system.

KEYWORDS: Parkinson's disease, machine learning, deep learning.

TEŞEKKÜR

Tezin konusunun seçiminde, uygulamasında ve çalışmamda yardımlarını esirgemeyen danışmanım sayın Dr. Öğr. Üyesi Nagehan İLHAN'a, yoğun çalışmalarım sırasında sabır gösterdiği ve bana katlandığı için eşim Emine'ye sürekli çalışmama izin verdiği için çocuklarım Selim ve Naz'a, motivasyon desteği için Bölüm Başkanım Doç. Dr. İ. Berkan AYDİLEK'e , bölümümüzün birbirinden kıymetli hocalarına ve çalışmam sırasında küçük veya büyük yardımını esirgemeyen herkes'e teşekkür ederim.



ŞEKİLLER DİZİNİ

Sayfa No

Şekil 1.1. Zaman içindeki engellilik derecesi.....	9
Şekil 1.2. Yapay sinir ağı mimarisi	13
Şekil 1.3. Her elemanın ağırlığı ve ANN sisteminin giriş ve çıkışı.....	15
Şekil 1.4. Bir Karar Ağacı Örneği	18
Şekil 1.5. (a) 1-NN karar kuralı: nokta ? soldaki sınıfa atanmakta; (b) K= 4 ile EYKA karar kuralı: nokta ? soldaki sınıfa da atanmaktadır	20
Şekil 1.6. Bir RBF-SVM sınıflandırıcısının iki sınıf arasında nasıl bir sınır oluşturduğuna dair bir örnek	22
Şekil 1.7. Çeşitli veri kümelerindeki Adaboost sınıflandırıcı.....	24
Şekil 1.8. RF sınıflandırıcısının sınıflandırma akışının bir gösterimi	25
Şekil 1.9. Bir evrişimli sinir ağının (KSA) tipik mimarisi	26
Şekil 1.10. RGB Matrisinin Dizisi	26
Şekil 1.11. Birçok evrişim katmanına sahip bir sinir ağı örneği	27
Şekil 1.12. Görüntü matrisi, çekirdek veya filtre matrisini çarpar	28
Şekil 1.13. Görüntü matrisi, çekirdek veya filtre matrisini çarpar	28
Şekil 1.14. 3 x 3 Çıktı matrisi.....	29
Şekil 1.15. Bazı yaygın filtreler.....	29
Şekil 1.16. ReLU işlemi	30
Şekil 1.17. Maksimum Havuzlama	31
Şekil 1.18. Havuzlama sonrası FC katmanı olarak düzleştirilmiş katman	31
Şekil 1.19. Genel KSA mimarisi	32
Şekil 1.20. RBM mimarisi	33
Şekil 1.21. İki gizli katmana sahip DBM mimarisi	34
Şekil 1.22. ÇKA sınıflandırıcısının sınıflandırma akışının bir gösterimi	36
Şekil 1.23. Test için bir dış döngü ve doğrulama için bir iç döngü ile iç içe Çapraz Doğrulamanın bir resmi	36
Şekil 1.24. 3 girişli tek bir nöron ve ReLU aktivasyon fonksiyonu	38
Şekil 1.25. İleri Beslemeli Çok Katmanlı Algılayıcı.....	39
Şekil 1.26. Spiral şablon üzerinde çizim yapan bir katılımcının resmi	42
Şekil 1.27. Bazı spiral çizimleri	42
Şekil 1.28. Bir dizi dalga formu	43
Şekil 4.1. Uygulanan Algoritmalar ve Doğruluk Değerleri.....	60
Şekil 4.2. Uygulanan Algoritmalar ve Kesinlik Değerleri	61
Şekil 4.3. Uygulanan Algoritmalar ve Duyarlılık Değerleri.....	61

ÇİZELGELER DİZİNİ

Sayfa No

Çizelge 2.1. Literatür taraması.....	51
Çizelge 3.1. Veri gruplarına ilişkin bilgiler.....	55
Çizelge 4.1. Uygulanan algoritma ve sonuçları	58
Çizelge 4.2. Algoritmaların öznitelik seçimi uygulanmadan önce elde edilen sonuçları.....	63
Çizelge 4.3. Öznitelik Seçimi uygulandıktan sonra elde edilen mutual_info_regression sonuçları.....	63
Çizelge 4.4. Öznitelik Seçimi uygulandıktan sonra elde edilen f_classif sonuçları.....	63
Çizelge 4.5. Öznitelik Seçimi uygulandıktan sonra elde edilen mutual_info_classif sonuçları.....	63
Çizelge 4.6. Öznitelik Seçimi uygulandıktan sonra elde edilen chi2 sonuçları.....	64
Çizelge 4.7. Öznitelik Seçimi uygulandıktan sonra elde edilen f_regression sonuçları.....	64
Çizelge 4.8. Öznitelik Seçimi methodlarının birleşimi ile elde edilen sonuçlar	64



1. GİRİŞ

Parkinson hastalığı (PH), dünya nüfusunun 1000’de 1-2’sini etkileyen ve yaşlanmayla birlikte yaygınlığı artan nörolojik bir hastalıktır. Parkinson hastalığının 60 yaş üstü toplam dünya nüfusunun %1’ini etkilediği bilinmektedir. Hastalığın genetik ve çevresel bir takım etkenlere bağlı olduğu düşünülmekle birlikte kesin olarak sebebi bilinmemektedir. Yapılan araştırmalarda %5-10 oranında genetik faktörler saptanırken bazı çevresel faktörlerin de hastalığın oluşumunda katkısı olduğu düşünülmektedir (Tysnes ve Storstein, 2017). Yapılan araştırmalarda bu çevresel faktörlerin arasında CO₂, tarım ilaçları ve ağır metaller gibi zararlı maddelere fazlaca maruz kalmak gibi etkenler bulunmaktadır. Bununla birlikte beyaz ırktan olmak, kırsal kesimde yaşamak ve kuyu suyu içmek gibi bazı durumların da Parkinson hastalığının risk faktörleri arasında olduğu belirtilmektedir (Çevik ve Kilimci, 2020).

PH hastalığı teşhisi konmuş olan bireylerde dopamin seviyesinde gerçekleşen azalmalar sonucu kasların kontrol edilmesinde sorunlar yaşanmakta ve hareket zorlukları meydana gelmektedir. Hastalığın teşhisi konduktan sonra kesin tedavisi olmamasına karşın hekimler tarafından hastalığın ilerlemesine ve semptomların azaltılmasına yönelik tedavi süreci işletilmektedir (Jankoviç, 2008). PH hastalığının klinik görünümü ve dopaminerjik tedaviye diğer nörodejeneratif hastalıklara nazaran vermiş olduğu görünür cevapla ayırt edilebilmesi mümkün olmaktadır. Bu nedenle hastalığın önceden teşhis edilmesi tedavi açısından oldukça önemli olmaktadır. Çünkü PH, ilaç tedavisine en fazla yanıt veren nörodejeneratif hastalıktır (Çakmur, 2011).

Parkinson hastalığının başlıca belirtileri hareket eksikliği ve yavaşlığı, titreme ile eklem bölgelerinde görülen sertliklerdir. Bunlarla birlikte dengede duramama, konuşma ve zihinde bozulmalarla depresif ruh hali yine Parkinson hastalığının belirtileri arasındadır. Bu belirtiler hastaya zamanında tedavi başlatılmadığı sürece zamanla şiddetlenmektedir. Hastalığın ilk belirtilerinin %75 oranında 60 yaşından sonra ortaya çıktığı görülmektedir (Çevik ve Kilimci, 2020). Parkinson hastalığının

konuşma üzerindeki etkileri arasında kısık ses, konuşmaya başlamada ve sürdürmede sıkıntı yaşama, akıcı konuşmada zorluk ve konuşmanın donuklaşması gibi sorunlar bulunmaktadır (Fahn, 2003).

Parkinson hastalığında genellikle hastalığın ilk evrelerinde konuşma bozukluğuna rastlanmamaktadır. Bununla birlikte %50 oranında ilerleyen safhalarda da konuşma bozukluğu yaşamayan hastalar bulunmaktadır. Genel olarak hastalığın karakteristik özelliklerinden dolayı hastaların konuşmalarında sorunlar yaşamaya başladıkları görülmektedir. Parkinson hastalığında meydana gelen konuşma bozukluklarında dopamin eksikliği nedeniyle yüz ve ağız bölgesindeki hareket etme kabiliyetinde azalmalar meydana gelmektedir (Torun, 1991). Aynı zamanda hastaların nefes alma kapasitesinin düşmesi de konuşma bozukluklarının ana nedenlerindendir. Konuşmada yaşanan bozuklukları genel itibariyle başlangıcı konuşma tonundaki yumuşama olmaktadır. Bununla birlikte hastalar konuşmalarında vurgu yapmadan monoton bir şekilde konuşmakta, konuşmaları hem yumuşak hem de hızlı tonda olmaktadır. Kimi durumlarda kelimeler arasındaki boşluklar kaybolmakta ve cümlelerin son kelimelerinde tekrarlar meydana gelmektedir. Yavaş ve zayıf tonda konuşmalar da Parkinson hastalarının konuşma karakteristikleri arasındadır (Özekmekçi ve ark., 2013).

Günümüzde özellikle makine öğrenmesindeki gelişmelerle birlikte ses bozuklukları ve konuşma ölçümleri kapsamında yapılan çalışmalarda artış olmuştur. Özellikle Parkinson gibi konuşma bozukluğu semptomu olan hastalıklar için farklı yöntemler kullanılarak seslerin analizi yöntemiyle sınıflandırma çalışmaları yapılmaktadır. Parkinson hastalığı için yapılan çalışmalara bakıldığında genel olarak çalışmaların makine öğrenmesi metotlarıyla seslerin ve hasta el yazılarının değerlendirilmesi üzerine yoğunlaşmış olduğu görülmektedir. Ses bozukluklarının incelenmesi maksadıyla oluşturulan veri setlerinin bir mikrofon vasıtasıyla ses kaydı alınarak yapıldığı görülmektedir. Daha sonra bu ses kayıtları makine öğrenmesinin farklı metotları kullanılarak analiz edilmekte, bu sayede hasta konuşmaları ve sağlıklı insanlar arasındaki ses ve konuşma sınıflandırmaları yapılabilmektedir. Yapılan

çalışmalarda görülen doğruluk oranları çalışmaların hastalık belirtilerinin analiz edilmesinde faydalı olduğu gözlemlenmiştir (Little ve ark., 2009).

Parkinson hastalarına ait seslerin analiz edilmesinde literatürde oldukça farklı yöntemlerin kullanıldığı görülmektedir. Bu yöntemler arasında ses frekanslarındaki düzensizlik, jitter, shimmer, kısıklık ölçümü, Mel frekansı keştrüm ölçümü, doğrusal tahminleme, işitsel modelleme gibi farklı metotlar bulunmaktadır (Bizal, 2014). Parkinson hastalığının kesin bir tedavisi olmamasına rağmen hastalığın erken teşhisi durumunda semptomların ilerlemesi durdurulabilmekte veya yavaşlatılabilmektedir. Bu nedenle bu tip teşhis yöntemlerinin geliştirilmesi hastalık için oldukça önemlidir.

Literatür incelendiğinde yaygın olarak kullanılan birkaç tane hazır veri setinin olduğu görülmektedir. Bu veri setlerinden biri Little ve ark. (2009) tarafından oluşturulan “Oxford Parkinson's Disease Detection (OPH)” veri seti ve diğeri Şakar ve ark. (2013) tarafından oluşturulan “Parkinson Speech Dataset with Multiple Types of Sound Recordings (PSD)” veri setidir. Bu veri setlerinden PSD’de toplam 40 deneğe, OPH’de ise 31 deneğe ait veriler bulunmaktadır. Denek sayılarına bakıldığında bu veri setlerinin oldukça sınırlı olduğu görülmektedir. Ancak Şakar ve ark. (2019) tarafından oluşturulan “Parkinson's Disease Classification (PDC)” veri seti toplam 252 deneğe ait 756 veriye sahip olması nedeniyle yapılan çalışmalarda son üç senedir yaygın olarak kullanılmaya başlanmıştır. Diğeri veri setleriyle karşılaştırıldığında PDC veri setinin hem örneklem boyutunun büyüklüğü hem de kullanılan sinyal işleme teknikleri nedeniyle Parkinson hastalığına ilişkin daha iyi bir temsil imkânı tanıdığı düşünülmektedir (Badem, 2019).

Bu çalışmanın amacı PDC veri seti kullanılarak makine öğrenme ve derin öğrenme yöntemlerinin performansını değerlendirmektir. Çalışmanın diğeri çalışmalardan farkı ise daha fazla makine öğrenmesi tekniği kullanarak bu tekniklerin PDC veri seti üzerindeki başarısını test etmektir.

Bu çalışmada kullanılan Şakar ve ark. (2019) verileri İstanbul Üniversitesi Cerrahpaşa Tıp Fakültesi Nöroloji Anabilim Dalı'nda yaşları 33 ile 87 (65.1 ± 10.9) arasında değişen 188 Parkinson hastası (107 erkek ve 81 kadın) üzerinden toplanmıştır. Kontrol grubu, yaşları 41 ile 82 arasında değişen (61.1 ± 8.9) 64 sağlıklı bireyden (23 erkek ve 41 kadın) oluşmaktadır. Veri toplama işlemi sırasında mikrofona 44,1 KHz'e ayarlanmış ve doktor muayenesini takiben her denekten üç tekrar ile /a/ ünlüsünün sürekli fonasyonu toplanmıştır. Tüm denekler veri toplama süreci hakkında bilgilendirilmiş, aydınlatılmış onam imzalanmış ve Bahçeşehir Üniversitesi Klinik Araştırmalar Etik Kurulu onayı doğrultusunda gönüllü olarak teste katılmıştır. Bu çalışmada PH hastalarının konuşma kayıtlarına PH değerlendirmesi için klinik olarak yararlı bilgiler çıkarmak için Wavelet Dönüşümü tabanlı temel özellikler, Vokal Kıvrım Özellikleri, Zaman Frekans Özellikleri, Mel Frekans Cepstral Katsayıları (MFCK'ler) özellikleri uygulanmıştır.

Önceki çalışmalar incelendiğinde, Şakar 2019 verisetinin çok fazla kullanılmadığı görülmektedir. Aynı verisetini kullanan çalışmaların değerlendirmeye aldıkları yöntemlerden daha fazla yöntem bu çalışma kapsamında uygulanmıştır. Bir diğer önemli katkı ise aynı verisetini kullanan diğer çalışmalarda parametreler üzerinde değişiklik yapılmadığı görülmektedir. Fakat bu çalışma kapsamında parametreler verilirken overfitting problemini önlemeye yönelik değişiklikler yapılmaktadır. Bahsi geçen özellikler, bu çalışmanın literatüre katkısı olarak belirtilmektedir.

1.1. Parkinson Hastalığı

Bu başlık altında parkinson hastalığına genel bakış, parkinson hastalığı tipleri ve belirtileri, parkinson hastalığının teşhisi detaylı şekilde anlatılmaktadır.

1.1.1. Genel bakış

PH, beyinde vücut hareketinin kontrolünden sorumlu bir alan olan substantia nigra'daki sinir hücrelerinin bozulması veya ölümü nedeniyle ortaya çıkan yaşa bağlı

dejeneratif nörolojik bir hastalıktır. PH, hücreler düzgün çalışmayı bıraktığında veya beynin substantia nigra pars compacta adı verilen bir bölümünde hasar meydana geldiğinde meydana gelen hastalıklardan biridir. Bu bölgedeki hücreler, dopamin adı verilen çok önemli bir kimyasal maddenin üretilmesinden sorumludur. Bu hücreler zamanla kaybolur, bu nedenle beyin, yürüme, yazma ve hatta gülümseme dahil olmak üzere vücuttaki hareket davranışlarını kontrol etmekten sorumlu olan dopamini kaybeder (Vallejo ve ark., 2016).

Genel olarak, Pereira ve ark. (2017), bu hastalığın kronik, ilerleyici bir nörodejeneratif bozukluk olarak tanımlandığını bildirmiştir. PH, özellikle nüfusun ortalama yaşının yüksek olduğu ülkelerde, dünyanın her yerindeki insanları etkiler (Vallejo ve ark., 2016). Parkinson Hastalıkları Vakfı'na (2015) göre, bir milyonu Amerika Birleşik Devletleri'nde olmak üzere dünya çapında yaklaşık 10 milyon insan Parkinson hastasıdır (Parkinson Hastalığı Vakfı, 2015). Parkinson Hastalığı Derneği'nin internet sitesinde, her 500 İngiliz'den birinin bu hastalığa sahip olduğu ve önümüzdeki 50 yılda bu sayının 3 kat artmasının beklendiği belirtilmiştir. Normalde bu hastalık zamanla daha da kötüleşmekte ve çoğunlukla 50-70 yaş arasındaki insanları etkilemektedir. PH ilk olarak 1817'de İngiliz bir doktor olan James Parkinson tarafından tanımlanmıştır ve PH için hala bir tedavi bulunmamaktadır (Lones ve ark., 2014; Pereira ve ark., 2015).

Tedavi edilemez olmakla birlikte, PH ile ilişkili semptomların uzun süreli rahatlama için erken tedavi önemlidir. Evreye ve semptomlara bağlı olarak çeşitli tedaviler alınabilir. Buna ilaçlar ve fiziksel aktiviteler, mesleki terapiler, konuşma-dil terapileri ve cerrahi yöntemler dahildir. Mevcut tedaviler arasında tıbbi tedaviler en yaygın olanıdır ve sinir hücrelerinin dopamin yapmasına yardımcı olmak için kullanılabilir; dopamini taklit etmek; belirli bir enzimin beyindeki dopamini parçalamasını engellemek; PH'nin belirli semptomlarını hafifletmektir (Cleveland Clinic, 2020).

PH'nin en sık görülen semptomları şu şekilde özetlenebilir:

- Bradikinezi: PH tanısı için bir ön koşul olarak hareketin yavaşlığı (Tysnes ve Storstein, 2017).
- “Tremor: Titremeler iki şekilde ortaya çıkar: dinlenme titremeleri ve hareket titremeleri. İstirahat titremeleri vücut hareketsizken ortaya çıkar, ancak kişi harekete geçtiğinde titreme durur. Bu titreme tipik olarak eller veya parmaklarla sınırlıdır. Tersine, hasta gönüllü olarak etkilenen vücut bölümünü hareket ettirdiğinde hareket titremeleri ortaya çıkar” (Cleveland Clinic, 2020).
- Sertlik: Serbest hareket edememe ile sonuçlanan kas sertliği (Raiano ve ark., 2020).
- Postural Değişiklik: Parkinson hastalığının dördüncü ana semptomu (Raiano ve ark., 2020). Genellikle düşmelere (Palakurthi ve Burugupally, 2019) yol açan stabiliteyi koruyamamayı ve PH'nin yaygın bir semptomu olarak çömelme postürünü içerir (Schlenstedt ve ark., 2020).
- Koordinasyon: Hareketleri başlatma ve hareketleri koordine etme zorluğu (Cleveland Clinic, 2020).
- Yürüme bozukluğu: Yürüme hızında azalma, adım uzunluğunda azalma, artan eksenel sertlik veya yürüme hızını koruyamama ile kendini gösteren anormal yürüme paternidir (Mirelman ve ark., 2019).
- Ses bozukluğu: Erken PH hastaları arasında en yaygın semptomlardan biridir. Kısık bir ses, azaltılmış ses seviyesi, kısıtlı perde değişkenliği (monoton), belirsiz artikülasyon (gecikmiş konuşma) ve kararsız konuşma hızı ile karakterize edilir (Gündüz, 2019).

Yukarıda bahsedilen semptomların tümü PH hastalarında yaygın olmakla birlikte, bradikinezi, tremor ve rijidite (sertlik) PH'nin başlıca semptomlarıdır (Cleveland Clinic, 2020) ve dolayısıyla PH tanısında yaygın olarak kullanılan semptomlardır (Tolosa ve ark., 2006). Aslında, Tysnes ve arkadaşlarına (2017) göre bir hasta, ancak hasta tremor veya rijidite ile birlikte bradikinezi belirtileri gösteriyorsa PH olarak teşhis edilebilir.

Bu nedenle, PH'yi teşhis etmek için yaygın bir yaklaşım, doktorun hastaların çizim modellerine bakarak PH'nin üç ana belirtisini arayabilmesi için hastalardan bir spiral veya benzeri çizimlerinin istendiği çizim muayenesidir. Bir çizim muayenesi özellikle bradikinezinin saptanması için uygun olabilir, yavaş hareket veya rahatsız edici derecede küçük el yazısı (mikrografi) ile kendini gösterebilen ve yazma sırasında ilerleyebilen (ilerleyen mikrografi) bir semptomdur (Ma ve ark., 2013).

1.1.2. Parkinson hastalığı tipleri ve belirtileri

PH temel olarak motor semptomları (hareket bozukluğu) ve bilişsel işlev bozukluğu gibi motor olmayan semptomları içermektedir (Meireles ve ark., 2012). Motor semptomlar için dört ana belirti ana semptom olarak kabul edilir: dinlenme tremoru, rijidite, bradikinezi ve bazen postural instabilitedir. PH hastalarının yaklaşık %70'inde 3-5 HZ arasında ve asimetrik tremor olarak karakterize edilen istirahat tremoru vardır. PH'nin ikinci belirtisi, eklemlerin hareketleri sırasında bir direnç hissidir ve buna dişli sertliği denir (Samii ve ark., 2004). Başka bir deyişle, yumuşak hareketlerin tersidir (Khan Academy, 2015). Hareketi yavaşlatmak bradikinezi adı verilen üçüncü işarettir ve el yazısı gibi basit hareketlerle genişlemektedir. Dördüncü semptom postural instabilitedir ve bu özellikle genç hastalarda PH'nin erken evresinde olmaz ve denge ile ilgilidir, bu da hastayı ayakları üzerinde dengesiz yapar ve düşmelere neden olabilir (Samii ve ark., 2004; Khan Academy, 2015).

PH'nin hipozmi, hızlı göz hareketi (REM), uyku davranış bozukluğu, kabızlık ve depresyon gibi motor olmayan semptomlarından bazıları, yıllar içinde herhangi bir motor semptomdan önce ortaya çıkabilir (Meireles ve ark., 2012). Birçok hasta aynı zamanda bilişsel işlev bozukluğu da sergiler ve bunlar hafif bilişsel bozukluktan (PD-MCI) PH demansına (PDD) kadar uzanır (Litvan ve ark., 2012). Bazı durumlarda, PD-MCI hastalığın erken evresinde ortaya çıkarken, PDD, PD'ye sahip olduktan 20 yıl sonra ortaya çıkma eğilimindedir. PD-MCI, normal yaşlanma ile beklenenin dışında gerçekleşen, hastanın günlük rutin aktivitelerini gerçekleştirmesine engel olmayan düşünme ve hafıza sorunları olarak tanımlanır.

Ayrıca PD-HBB tanısı YGB'ye geçiş olabileceği için önemlidir (Meireles ve ark., 2012). YGB belirtileri arasında kısa süreli bellekte bozulma, yürütücü işlev bozukluğu, dikkat bozukluğu, görsel-uzaysal eksiklik, psikotik belirtiler (halüsinasyonlar) gibi davranışsal veya nöropsikiyatrik belirtiler, kişilik ve duygudurum değişiklikleri, kaygı ve ilgisizlik sayılabilir (Meireles ve ark., 2012).

1.1.3. Parkinson hastalığının teşhisi

Öncelikle, PH ve parkinsonizm arasında ayrım yapmak önemlidir. Parkinsonizm, titreme, sertlik ve bradikinezi gibi PD'ye atfedilebilecek bir dizi semptomun klinik sunumudur. PH, parkinsonizmin bir biçimi olarak kabul edilir, bu nedenle, parkinsonizm semptomlarını gösteren tüm denekler, vasküler bozukluk veya diğer nörodejeneratif hastalıklar gibi diğer sağlık sorunlarından kaynaklanabileceğinden PD'ye sahip olmayacaktır (Parkinson Hastalığı Vakfı, 2015).

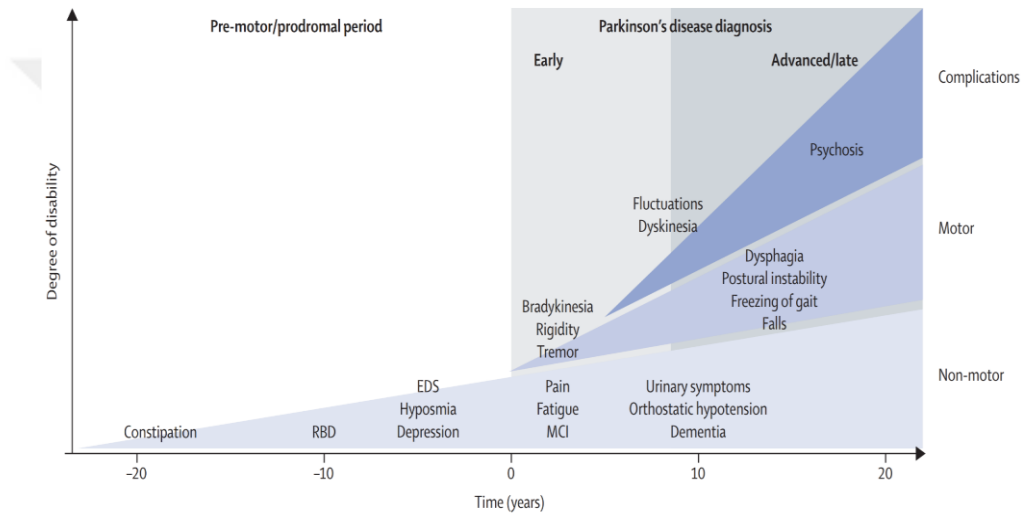
PH hastalarının sadece motor özelliklerini izleyerek PH'yi diğer parkinsonizm türlerinden ayırt etmek güçtür. Bu nedenle, asimetri ve levodopa tedavisine güçlü yanıt gibi klinik tanı doğruluğunu artırmak için başka özelliklerin eklenmesi gerekir (Berg ve ark., 2013). Aşağıdakiler dahil olmak üzere motor özelliklerden tanı kriterlerinin karşı karşıya olduğu çeşitli engeller vardır:

- Bu kriter hareket problemlerine odaklanır, PH normal olarak levodopa tedavisine yanıt vermeyen ve herhangi bir motor semptomdan önce ortaya çıkan motor olmayan özelliklerle de ilişkilidir (Berg ve ark., 2013).
- Kriterlerdeki ana ölçü, diğer parkinsonizm türlerinde de görülen dopamin replasmanına verilen yanıttır (Berg ve ark., 2013).

Birleşik Krallık Parkinson Hastalığı Derneği Beyin Bankası Tanı Kriterleri, daha önceki tanının bir güncellemesi olarak karşımıza çıkmaktadır ve tanının özgüllüğünü artırmak için hem asimetriyi hem de levodopa tedavisine güçlü bir yanıtı içermektedir. Ayrıca, takip araştırmalarının bildirdiği gibi, tanısal doğruluk,

erken çalışmaların önerdiğinden ölümden %76 ile %90 doğruluk olarak yükselmiştir (Berg ve ark., 2013).

Şekil 1.1. zaman içinde engelliliğin derecesini göstermektedir. PH tanısının motor özellikler genişlemeye başladığında (zaman = 0) gerçekleştiği görülebilir, ancak motor dışı semptomlar tanıdan 20 yıl veya daha fazla bir süre önce gelebilir. PH tanısından sonra demansa ulaşana kadar ek motor olmayan semptomlar ortaya çıkmaktadır. Ayrıca postural instabilitenin hastalığın ileri evrelerinde ortaya çıkma eğiliminde olduğu görülebilir.



Şekil 1.1. Zaman içindeki engellilik derecesi (Kalia ve ark., 2015)

Şekil 1.1.'de yer alan EDS=gündüz aşırı uyku hali, HBB=hafif bilişsel bozukluk, RBD=REM uyku davranış bozukluğu terimlerinin kısaltmalarıdır (Kalia ve ark., 2015). UPDRS, Birleşik Parkinson Hastalığı Derecelendirme Ölçeği'ni ifade etmektedir. Bu ölçek her semptom için 5 puanlık bir puan verir (Goetz ve ark., 2008). Parkinson hastalığının tanısının konulması, belirtileri gibi durumların verilmesinden sonra, PH tanısının konulmasında çokça faydalanılan yöntem olan maline öğrenmesinden bahsedilmektedir.

1.2. Makine Öğrenmesi

Makine öğrenmesi (MÖ), bilgisayar bilimlerinde önemli bir alandır. MÖ, bilgisayarların mevcut girdilerden (eğitim verileri) öğrenmesini ve bize uzmanlık

olarak çıktı vermesini amaçlayan bir algoritmalar topluluğudur (Shalev-Shwartz ve ark., 2014). Genel olarak, insanlar tarafından gerçekleştirilen görevlerde insan yeteneklerinin üzerinde olan görevleri yerine getiren çeşitli görevleri otomatikleştirmede makine öğrenimini kullanılmaktadır.

1.2.1. Öğrenme türleri

Shalev-Shwartz ve ark. (2014)'e göre öğrenme, “deneyimi uzmanlık kazanmak için kullanmak” olarak tanımlanabilir. Haykin'e (2009) göre iki ana öğrenme türü vardır:

1. Denetimli öğrenme: Bu tür öğrenmede girdi-çıkıtlı bulunmaktadır. Öğretici her girdi için algoritmaya “optimum” yanıtı (çıkıtlı) verir ama algoritma asıl çıktıyı vermektedir. Hem optimum hem de gerçek çıktı arasındaki farka bağlı olarak, algoritma hatayı azaltmak için parametrelerini değiştirmektedir. Böylece algoritma, öğreticinin sağladığı bilgilerden öğrenir. Bu öğrenme türü normalde bir sınıflandırıcıya modelin girdisi ve çıktısı arasındaki ilişkiyi öğretme süreci olan sınıflandırma görevi ile ilgilidir ve daha sonra bu uzmanlığı daha sonra sonucu görünmeyen girdiler için kullanır (Witten ve ark., 2011).

2. Denetimsiz öğrenme: Bu tür bir öğrenmede öğretmenimiz yoktur. Sonuç olarak, denetimli öğrenmede olduğu gibi öğrenecek bir geri bildirim bulunmamaktadır. Bu tür bir öğrenmede algoritma, bir tahmin işlemi gerçekleştirmek yerine girdiyi nasıl temsil edeceğini öğrenmektedir. Bu tür öğrenme altında, oyun oynamak gibi etkileşim problemlerinde uzmanlaşmış pekiştirmeli öğrenme vardır. Bu tür öğrenmede algoritma, çevre ile etkileşim sırasında yaptığı hatalardan ders çıkarmaktadır (Sutton ve Barto, 1998). Bu öğrenme türü normalde, modelin girdisini çıktının olmadığı durumlarda kümelere ayırma süreci olan Kümeleme ile ilgilenmektedir (Witten ve ark., 2011).

1.2.2. Sinir ağı (NN)

Shalev-Shwartz ve ark. (2014)'e göre bir sinir ağı, birbirine bağlantılar ile bağlanan birkaç düğüm içeren ve düğümün çıktısının ağırlıklı toplamı sonraki düğümlere girdi olarak geçen yönlendirilmiş bir grafik yöntemidir. Bu ağlar insan beyninden ilham almaktadır (Haykin, 2009).

NN, sınıflandırma görevine ait denetimli bir MÖ algoritmasıdır. NN birkaç türe ayrılabilir ancak bu türleri ağ yapısına göre aşağıdaki gibi gruplanabilir.

- Tek Katmanlı İleri Beslemeli Ağlar: NN'nin basit bir şeklidir ve doğrudan bir çıktı katmanına bağlı bir girdi katmanı içermektedir.
- Çok Katmanlı İleri Beslemeli Ağlar: Bu tür, giriş ve çıkış katmanları arasında gizli katmanların (en az bir katman) bulunmasıyla önceki türden farklıdır.
- Tekrarlayan Ağlar: Bu tip, nöronların çıkışından diğer nöronlara bir veya daha fazla geri besleme döngüsünün varlığı ile ileri beslemeli ağlardan farklıdır (Haykin, 2009).

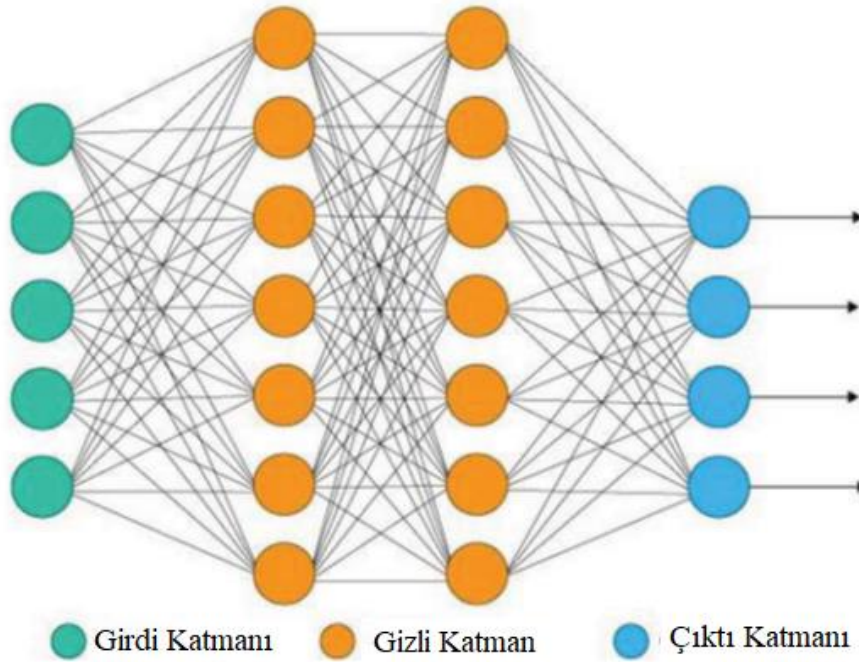
1.2.3. Yapay Sinir Ağları (YSA)

Yapay sinir ağı (YSA) modeli, insan-beyin süreçlerini simüle eden hesaplamaları ve matematiği içerir. Son zamanlarda elde edilen gelişmelerin çoğu, görüntü ve ses tanıma, robotik ve YSA 'lerin kullanımı gibi yapay zeka araştırma alanıyla ilgilidir (Sadiq, 2018).

YSA, fiziksel bir fenomeni veya karar sürecini tanımlayan temsili verilerden örnek olarak öğrenmeye izin veren insan sinir sisteminden esinlenen bir modelleme tekniğidir. YSA'nın benzersiz bir özelliği, bağımsız ve bağımlı değişkenler arasında ampirik ilişkiler kurabilmeleri ve temsili veri kümelerinden ince bilgileri ve karmaşık bilgileri çıkarabilmeleridir. Bağımsız ve bağımlı değişkenler arasındaki ilişkiler, fenomenlerin herhangi bir matematiksel temsili hakkında varsayımlar olmadan kurulabilir. YSA modelleri, gürültülü verilerle başa çıkma kapasitesi de dahil olmak

üzere regresyon tabanlı modellere göre belirli avantajlar sağlar (Balakrishnan ve ark., 2019).

Yapay sinir ağları, insan beyniyle aynı şekilde, nöron düğümleri web benzeri bir şekilde birbirine bağlı olarak tasarlanmıştır. Nöronlar, insan beynini oluşturan milyarlarca hücredir. Her nöron, bilgiyi beyne ve beyinden (girdiler ve çıktılar) getirerek işleyen bir hücre gövdesinden oluşur (Van Gerven ve Bohte 2017). Bu tür ağların ana fikri (bir dereceye kadar) biyolojik sinir sisteminin çalışma şeklinden, bilgiyi öğrenmek ve yaratmak için verileri ve bilgiyi işlemekten esinlenmiştir. Bu fikrin temel unsuru, bilgi işlem sistemi için yeni yapılar oluşturmaktır. Yapay sinir ağı mimarisi Şekil 1.2.'de gösterilmiştir (Bre ve ark., 2018).



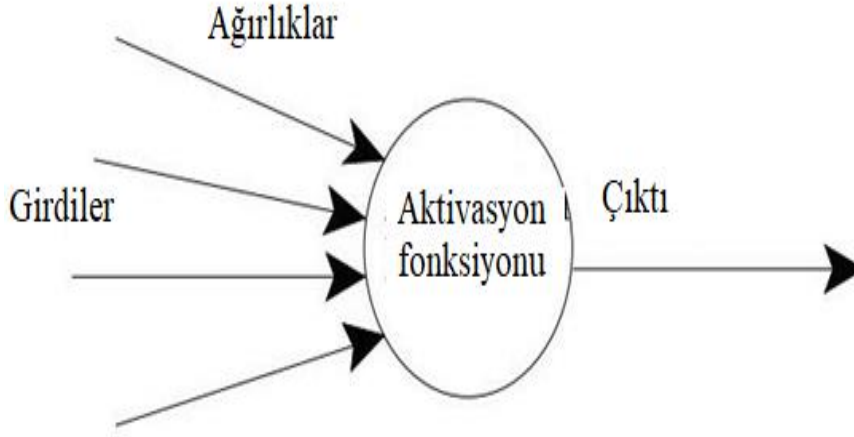
Şekil 1.2. Yapay sinir ağı mimarisi (Bre ve ark., 2018)

YSA modelleri, biyolojik bir sinir sisteminden esinlenen belirli mimari formata sahiptir. İnsan beyninin yapısı gibi, YSA modelleri de karmaşık ve doğrusal olmayan bir formdaki nöronlardan oluşur. Nöronlar birbirlerine ağırlıklı bağlantılarla bağlanır. YSA modellerinde veri toplama ve analizi, ağ yapısı tasarımı, gizli katman sayısı, ağ simülasyonu ve ağırlıklar/önyargı dengesi gibi tüm süreçler öğrenme ve eğitim yöntemleriyle hesaplanır. YSA'lar, bir veya daha fazla gizli düğüm katmanı ile

bağlanan bir giriş düğümleri katmanından ve çıkış düğümleri katmanından oluşur. Giriş katmanı düğümleri, etkinleştirme işlevlerini ateşleyerek gizli katman düğümlerine bilgi aktarır ve gizli katman düğümleri, sunulan kanıtlara bağlı olarak ateşlenir veya uykuda kalır. Gizli katmanlar kanıtlara ağırlıklandırma işlevleri uygular ve gizli katmandaki belirli bir düğümün veya düğüm kümesinin değeri bir eşiğe ulaştığında, çıktı katmanındaki bir veya daha fazla düğüme bir değer geçirilir. YSA'lar çok sayıda vaka (veri) ile eğitilmelidir. YSA'ların uygulanması, verilerin modeli eğitmek için yetersiz olduğu nadir veya aşırı olaylar için mümkün değildir. YSA'lar, nicel kanıtların yerini almak için insan uzmanlığının (uzman görüşü) dahil edilmesine izin vermez (Liu ve ark., 2018).

Sistem, bir problemi çözmek ve sinapslar (elektromanyetik bağlantılar) yoluyla bilgi iletmek için birlikte çalışan nöronlar adı verilen çok sayıda birbirine bağlı işlem elemanından oluşur. Nöronlar birbirine yakından bağlıdır ve katmanlar halinde organize edilir. Giriş katmanı verileri alırken, çıkış katmanı nihai sonucu oluşturur. İki arasında, bir veya daha fazla gizli katman tipik olarak sıkıştırılır. Bu düzenleme, verilerin tam akışını tahmin etmeyi veya bilmeyi zorlaştırır.

Her bağlantının bir bağlantı ağırlığı vardır ve her nöronun bir eşik değeri ve bir aktivasyon fonksiyonu vardır (Balakrishnan ve ark., 2019). Her girdinin, girdinin ağırlığının işaretine bağlı olarak pozitif veya negatif bir ağırlığa sahip olup olmadığı hesaplanır. Ağırlık, bir bağlantıdaki sinyal yoğunluğunu etkiler (Liu ve ark., 2018). Bir sinyalin yalnızca toplam sinyal onu aşarsa iletildiği bir eşik değeri vardır. Etkinleştirme Değeri, toplama biriminin ağırlıklı toplamıdır ve çıktı, bu etkinleştirme değerinden gelen sinyale göre üretilir. Her elemanın ağırlığı ile YSA sisteminin giriş ve çıkışı arasındaki ilişki Şekil 1.3.'te gösterilmiştir.



Şekil 1.3. Her elemanın ağırlığı ve ANN sisteminin giriş ve çıkışı.

Akıllı ağlar, öğrenme ve uyum sağlama yeteneğine sahiptir. Hasarlı hücrelerin telafi edilebileceği ve rejenerasyona katkıda bulunabileceği şekilde çalışırlar. Akıllı sistemlerin en önemli özelliği öğrenme kapasiteleridir. Bir öğrenme sistemi daha esnek ve programlanması daha kolaydır, böylece yeni problemlere ve denklemlere daha iyi yanıt verebilir (Liu ve ark., 2018).

Yapay sinir ağları, insanlar gibi, farklı örnekleri kullanarak öğrenir ve bir öğrenme süreci sırasında kalıpları tanımlamak ve bilgileri kategorize etmek gibi belirli görevleri yerine getirmek için bir sinir ağı kurulur. Örneğin, dokunsal sinir hücrelerine enjekte ederek, hücreler sıcak vücuda gitmemeyi öğrenir ve bu algoritma ile sistem hatasını düzeltmeyi öğrenebilir. Yapay sinir ağları, bilinmeyen veya çok karmaşık iç yapılara sahip sistemlerin kontrolünde veya modellenmesinde giderek daha fazla kullanılmaktadır. Örneğin, bir motorun girişini kontrol etmek için bir sinir ağı kullanılabilir, bu durumda sinir ağının kendisi kontrol işlevini öğrenir. Bu sistemlerde öğrenme uyarlanabilir, yani benzetmeler kullanarak, sinapsların ağırlığı, yeni girdiler verilirse sistemin doğru cevabı üreteceği şekilde değişir (Wu ve Feng 2018). Bir sinir ağına, eğitilirken (eğitim yöntemlerinden birini kullanarak) bir dizi girdi ve bunlara karşılık gelen çıktılar verilir (Van Gerven ve Bohte, 2017).

Gerçek dünyanın problemlerini çözen YSA uygulamaları, finanstan hidrolojiye kadar çok çeşitli bilimsel alanları içerir ve üç kategoriye ayrılır: (i) desen sınıflandırması, (ii) tahmin ve (iii) kontrol ve optimizasyon. YSA modellerinde kümeleme, sınıflandırma ve simülasyon, YSA modelleriyle ilgili farklı yapı türleriyle gerçekleştirilir. YSA modelleri üç gruba ayrılır: (i) statik YSA, (ii) dinamik YSA ve (iii) istatistiksel YSA'dır. Statik YSA modeli, çok katmanlı bir algılayıcı sinir ağı modeli, dokunulmuş gecikme çizgileri ve tekrarlayan sinir ağı modelleri gibi dinamik sinir ağı modelleri ve radyal temel fonksiyon modeli gibi istatistiksel sinir ağı modelleri ve genelleştirilmiş regresyon sinir ağı modeli olarak bilinir (Bre ve ark., 2018). YSA modelini diğer optimizasyon teknikleriyle, örneğin daha iyi tahmin amacıyla uyarlanabilir nöro-bulanık çıkarım sistemi ile birleştirmek de mümkün olabilir. YSA'nın en önemli etkilerinden biri hidrolojik sistemlerdedir ve bu bölüm, YSA modellerinin hidroloji ve su kaynakları yönetiminde kavramı, arka planı, teorisi, sınıflandırılması ve uygulanması hakkında bir fikir vermektedir (Balakrishnan ve ark., 2019).

1.2.4. Karar ağaçları

Karar Ağacı, kolay anlaşılır bir temsil biçimine odaklanan bir sınıflandırma tekniğidir ve en yaygın öğrenme yöntemlerinden biridir. Karar Ağaçları, öznitelik vektörlerinden oluşan veri kümelerini kullanır, bunlar da vektörü tanımlayan bir dizi sınıflandırma özniteliği ve veri girişini belirli bir sınıfa atayan bir sınıf özniteliği içermektedir. Bir Karar Ağacı, belirli bir durdurma kriterine ulaşılan kadar verileri mümkün olduğunca farklı mevcut sınıflara ayıran öznitelik üzerindeki veri kümesini yinelemeli olarak bölerek oluşturulmaktadır. Temsil formu, Karar Ağaçları, insanlar için anlaşılması kolay, ağaç yapılı bir formatta kolayca görselleştirilebildiğinden, kullanıcıların verilere hızlı bir genel bakış elde etmelerini sağlamaktadır (Liu ve Özsu, 2009).

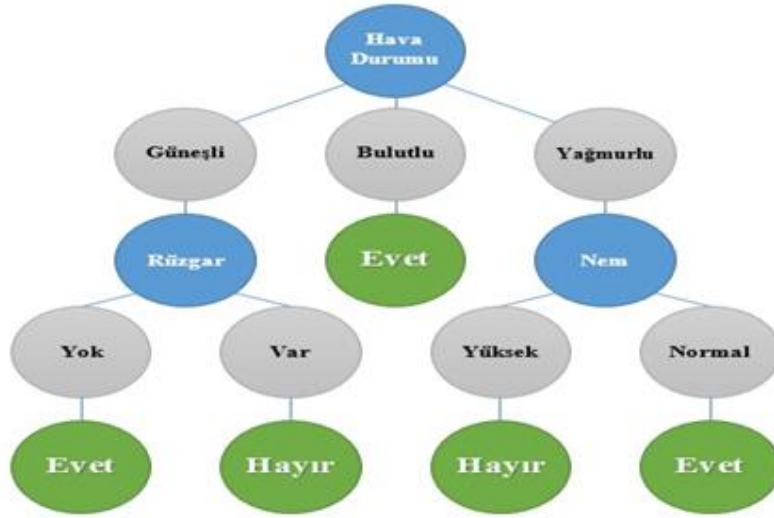
Karar Ağacı eğitimi ile ilgili ilk algoritmalarından biri, 1986 ve 1993'te Ross Quinlan tarafından geliştirilen Yinelemeli Dichotomiser 3 (ID3) ve C4.5 algoritmasıdır (Quinlan, 1986). Bu algoritmalar, daha sonraki birçok gelişmenin

temelini oluşturmuştur. Karar ağaçları, karar destek aracı olarak kullanılan yönlendirilmiş ağaçlardır. Karar kurallarını temsil ederler ve ardışık kararları gösterirler (Liu ve Özsu, 2009).

Karar Ağaçlarında düğümler kök düğüm, iç düğümler ve yaprak olarak da adlandırılan uç düğümler olarak ayrılmaktadır. Kök düğüm, karar destek sürecinin başlangıcını temsil eder ve gelen kenarları yoktur. İç düğümlerin tam olarak bir gelen kenarı vardır ve en az iki giden kenarı vardır (Liu ve Özsu, 2009). Veri setinin bir niteliğine dayalı bir test içermektedir.

Yaprak düğümler, çoğunlukla bir sınıf tahmini ile temsil edilen karar problemlerine bir cevaptan oluşmaktadır. Örnek olarak, bir karar problemi, sınıf tahminlerinin evet ve hayır olduğu bir çevrimiçi mağazadaki bir müşterinin bir satın alma yapip yapmayacağı sorusu olabilir. Yaprak düğümlerin giden ve tam olarak bir gelen kenarı yoktur. Kenarlar, önceki düğümden alınan kararı temsil eder (Liu ve Özsu, 2009).

Bir n düğümü verildiğinde, n 'ye tam olarak bir kenarla ayrılan tüm takip eden düğümler n 'nin çocukları olarak adlandırılırken, n 'ye tüm alt düğümlerinin ebeveyni denir. Şekil 1.4.'te bir Karar Ağacı örneği gösterilmektedir. Örneğin, soğuk özneliliklere sahip bir veri kaydı; “Kutup Ayısı”, sıcaklık özelliği soğuk olduğu için sol alt ağaca ve ardından ilgili etiketle sınıflandırılan “Kuzey Kutbu” yaprağına aktarılacaktır (Liu ve Özsu, 2009).



Şekil 1.4. Bir Karar Ağacı Örneği (Liu ve Özsu, 2009)

Karar Ağacı Eğitimi, çoğunlukla sınıflandırma amacıyla kullanılan yaygın bir veri madenciliği yöntemidir. Amacı, bir dizi girdi özniteliğine dayalı olarak bir hedef özniteliğin değerini tahmin etmektir. Denetlenen bir senaryoda bir Karar Ağacı eğitimi, veriler içindeki kalıpları bulmak ve Karar Ağacını oluşturmak için bir eğitim seti kullanılarak yapılır. Daha sonra, hedef özniteliklerinin değerini tahmin etmek için daha önce görülmemiş bir dizi örnek kullanılabilir. Eğitim seti şu şekilde veri kayıtlarını içermektedir (Liu ve Özsu, 2009):

$$(\vec{x}, Y) = (x_1, x_2, x_3, \dots, x_n, Y) \quad (1.1)$$

Denklem 1.1’de Y, hedef öznitelik değeridir ve $\vec{x}$, n girdi değeri içeren bir vektördür; burada n, veri kümesindeki özniteliklerin sayısıdır (Liu ve Özsu, 2009).

Bir Karar Ağacını eğitmek ve böylece bir sınıflandırıcı oluşturmak için, bir hedef öznitelik, girdi öznitelikleri, bir bölme ölçütü ve bir durdurma ölçütü içeren bir eğitim kümesine ihtiyaç vardır. Belirli bir düğümde, bölme kriteri tüm nitelikler için bir değer hesaplar. Bu değer, bu özniteliği kullanarak düğümü bölerek elde edilen bilgi miktarının bir ölçüsünü temsil eder. Daha sonra, tüm özniteliklerden en iyi değer alınır ve düğüm, ilgili özniteliğin farklı sonuçlarına bölünür. Bu noktada, öznitelikler arasında en iyi bölünmeyi bulma işlemi, bir durdurma kriterine ulaşılan

kadar oluşturulan tüm alt ağaçlara özyinelemeli olarak uygulanır (Liu ve Özsu, 2009).

Ortak durdurma kriterleri şunlardır:

- Ağacın maksimum yüksekliğine ulaşıldığında
- Düğümdeki kayıt sayısı izin verilen minimumdan daha az olduğunda
- En iyi bölme kriteri, kazanılan bilgi açısından belirli bir eşiği aşmadığında.

Bölme özniteliği sayısal türdeyse, kayıtları özniteliğin tüm sonuçlarına bölme imkanı yoktur. Bu, ID3'e kıyasla C4.5 Karar Ağacının ana yükseltmelerinden biridir. C4.5 ayrıca sayısal öznitelikler için de en iyi bölme noktalarını hesaplayabilir ve bunları büyük veya eşit ve operatörlerden küçük kullanarak bölebilir (Liu ve Özsu, 2009).

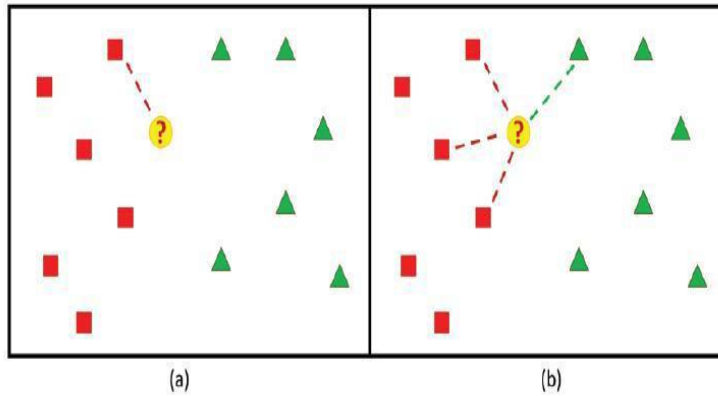
Bir Karar Ağacını bu otomatik süreçle eğitmek, sınıflandırma açısından çok az güce sahip bölümlere sahip büyük Karar Ağaçları ile sonuçlanabilir. Ek olarak, ağaçlar fazla yakınsama eğilimindedir, bu da eğitim örneklerine çok yakınsak oldukları anlamına gelir. Bu ağaçlar görünmeyen verilere uygulandığında kötü performanslara neden olur. Bu nedenle budama adı verilen bir teknik geliştirilmiştir. Amacı, gürültülü veya hatalı verilere dayalı parçalar veya fazla takılan parçalar gibi daha az veya üretken olmayan parçaları Karar Ağacından çıkarmaktır. Bu genellikle doğruluk açısından daha fazla iyileştirme ile sonuçlanır ve ağaç boyutunu küçültür. Bu süreç, her gerçek dünya veri kümesinin hatalı veya gürültülü veriler içermesi nedeniyle özellikle önemlidir (Liu ve Özsu, 2009).

1.2.5. K-en yakın komşu (EYKA)

Örüntü tanımada, EYKA algoritması, öznitelik alanındaki en yakın eğitim örneklerine dayalı olarak nesneleri sınıflandırmak için kullanılan bir yöntemdir. EYKA, bir tür örnek tabanlı öğrenme veya işlevin yalnızca yerel olarak yaklaşık olduğu ve tüm hesaplamaların sınıflandırmaya kadar ertelendiği düşük öğrenmedir (Devroye, 1981; Cover ve Hart, 1967). EYKA, verilerin dağılımı hakkında çok az

veya hiç ön bilgi olmadığında kullanılan temel ve en basit sınıflandırma tekniğidir (Devroye ve ark., 1994). Bu kural, öğrenme sırasında tüm eğitim setini korumakta ve her sorguya eğitim setindeki k -en yakın komşularının çoğunluk etiketi ile temsil edilen bir sınıf atamaktadır. En Yakın Komşu kuralı (NN), $K = 1$ olduğunda EYKA'nın en basit şekli olmaktadır. Bu yöntemde her örneğin, çevresindeki örneklerle benzer şekilde sınıflandırılması gerekmektedir. Bu nedenle, bir örneğin sınıflandırması bilinmiyorsa, en yakın komşu örneklerinin sınıflandırması dikkate alınarak tahmin yürütülebilmektedir. Bilinmeyen bir örnek ve bir eğitim seti verildiğinde, bilinmeyen örnek ile eğitim setindeki tüm örnekler arasındaki tüm mesafeler hesaplanabilmektedir.

En küçük değere sahip mesafe, bilinmeyen örneğe en yakın eğitim kümesindeki örneğe karşılık gelmektedir. Bu nedenle, bilinmeyen örnek, bu en yakın komşunun (Domeniconi ve ark., 2002) sınıflandırmasına göre sınıflandırılabilir. Şekil 1.3., 2 sınıfa bölünmüş bir dizi örnek için $K=1$ ve $K=4$ için EYKA karar kuralını göstermektedir. Şekil 1.5. (a)'da, bilinmeyen bir numune, yalnızca bir bilinen numune kullanılarak sınıflandırılmakta; Şekil 1.5. (b)'de bilinen birden fazla örnek kullanılmıştır. Son durumda, K parametresi 4'e ayarlanmaktadır, böylece bilinmeyi sınıflandırmak için en yakın dört örnek dikkate alınmaktadır. Bunlardan üçü aynı sınıfa aitken, sadece biri diğer sınıfa aittir. Her iki durumda da bilinmeyen örnek soldaki sınıfa ait olarak sınıflandırılmaktadır (Devroye ve ark., 1994).



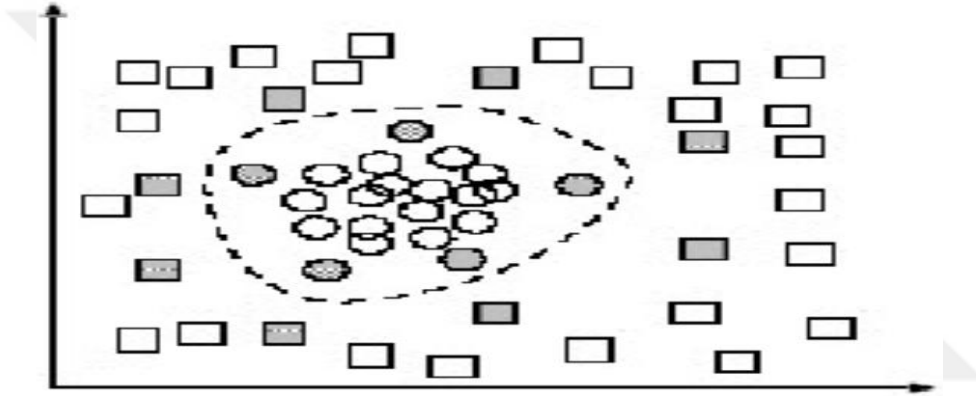
Şekil 1.5. (a) 1-NN karar kuralı: nokta ? soldaki sınıfa atanmakta; (b) $K=4$ ile EYKA karar kuralı: nokta ? soldaki sınıfa da atanmaktadır (Guo ve ark., 2006)

Bir EYKA sınıflandırıcının performansı, öncelikle K seçiminin yanı sıra uygulanan mesafe metriği ile belirlenmektedir (Guo ve ark., 2006). Tahmin, komşuluk büyüklüğü K seçiminin duyarlılığından etkilenmektedir, çünkü yerel bölgenin yarıçapı, sorguya en yakın Kth komşusunun mesafesi ile belirlenmektedir ve farklı K, farklı koşullu sınıf olasılıkları vermektedir. K çok küçükse, veri seyrekliği ve gürültülü, belirsiz veya yanlış etiketlenmiş noktalar nedeniyle yerel tahmin çok zayıf olma eğiliminde olmaktadır. Tahmini daha da yumuşatmak için K artırılabilen ve sorgunun etrafındaki geniş bir bölge hesaba katılabilmektedir. Büyük bir K değeri, diğer sınıflardan aykırı değerlerin eklenmesiyle sınıflandırma performansını düşürmektedir. Problemlerle başa çıkmak için, EYKA'nın sınıflandırma performansını iyileştirmek için ilgili araştırma çalışmaları yapılmıştır. Uygun bir komşuluk büyüklüğünün nasıl seçileceği K, EYKA'nın sınıflandırma performansını büyük ölçüde etkileyen kilit bir konudur (Domeniconi ve ark., 2002).

Küçük eğitim örnekleme büyüklüğü, optimal komşuluk büyüklüğü K'nin seçimini büyük ölçüde etkileyebilmektedir ve EYKA'nın sınıflandırma performansının bozulması, K seçiminin duyarlılığı ile doğrudan alakalıdır. Genel olarak değerlendirildiğinde de sınıflandırma sonuçlarının çok hassas olduğu görülmektedir. Teorik bir bakış açısından, EYKA'nın sınıflandırma performansı, sorguya en yakın Kth komşusunun mesafesi ile belirlenen veri uzayının yerel bir bölgesindeki sorgunun koşullu sınıf olasılıklarının tahmini ile belirlenmektedir. Bu nedenle, sınıflandırma performansı seçilen K değerine çok duyarlı olmaktadır. K vektör büyüklüğünün farklı seçimlerinin hassasiyet sorununu ele almak amacıyla, EYKA için bazı ağırlıklı oylama yöntemleri geliştirilmiştir. Noktaların düzgün dağılmadığında gösterilmektedir ve K değerini önceden belirlemek zorlaşmaktadır. Genel olarak, daha büyük K değerleri sunulan gürültüye karşı daha bağıştır ve sınıflar arasındaki sınırları daha pürüzsüz hale getirmektedir. Sonuç olarak, aynı (optimal) K'yi seçmek, farklı uygulamalar için neredeyse imkansız hale gelmektedir (Domeniconi ve ark., 2002).

1.2.6. Destek vektör makineleri (DVM)

Destek Vektör Makineleri (DVM), iki sınıf arasında bir sınır belirlemek için matematiksel bir formül kullanan ve böylece ikili sınıflandırma problemini çözen bir öğrenme yöntemidir. Sıradan bir DVM ile karşılaştırıldığında, Radius Basis Function (RBF)-DVM, lineer olarak ayrılamaz olan iki sınıfın ayrılmasını sağlamak için Gauss çekirdeği kullanan bir DVM'dir (Nie ve ark., 2020). Şekil 1.6.'da bir RBF-DVM sınıflandırıcı tarafından bir sınıflandırmanın nasıl yapılabileceğinin bir örneğini göstermektedir.



Şekil 1.6. Bir RBF-SVM sınıflandırıcısının iki sınıf arasında nasıl bir sınır oluşturduğuna dair bir örnek (Kamarulzaini ve ark., 2018)

Şekil 1.6.'daki örnekler sınıfa göre şekillenmektedir. Gri olanlar sınıra yakın olanlardır. SVM, farklı kernellerle kullanılabilmektedir. Yüksek boyutlu bir uzayda, doğrusal bir ayırıcı bazen verileri iyi bir şekilde sınıflandıramazken, doğrusal olmayan kernel dönüşümleri, farklı sınıflardan veri noktalarını ayırmak için doğrusal olmayan karar sınırları sağlayabilmektedir. Python'un scikit-learn modülünde seçilebilecek birçok farklı kernel bulunmaktadır. Örneğin, RBF kerneli (Gauss radyal tabanlı fonksiyon kerneli) şu şekilde tanımlanmaktadır (Nie ve ark., 2020):

$$K(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right) \quad (1.2)$$

Denklem 2.2'de $\|x_i - x_j\|^2$, T gözlemleri i ve j iki boyutlu öznitelik vektörü

$x_i = (x_{i1}, x_{i2}, \dots, x_{ip})^T$ ve $x_j = (x_{j1}, x_{j2}, \dots, x_{jp})^T$ arasındaki kare Öklid mesafesi olarak tanımlanabilir.

Yaygın olarak kullanılan diğer kerneller şunları içermektedir (Kamarulzaini ve ark., 2018):

- Doğrusal kernel:

$$K(x_i, x_j) = x_i^T x_j \quad (1.3)$$

- Polinom kernel:

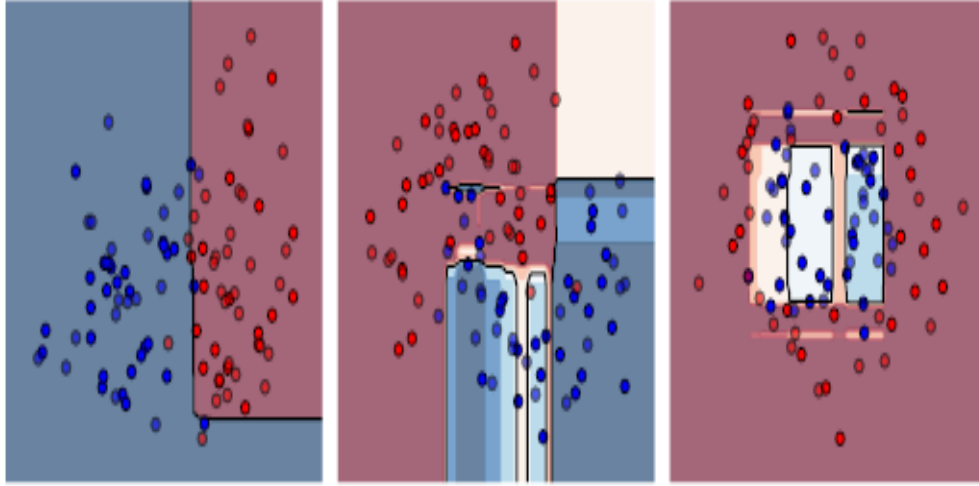
$$K(x_i, x_j) = (\gamma x_i^T x_j + r)^d, \gamma > 0 \quad (1.4)$$

- Sigmoid kernel:

$$K(x_i, x_j) = \tanh(\gamma x_i^T x_j + r) \quad (1.5)$$

1.2.6.1. Adaboost karar ağaçları

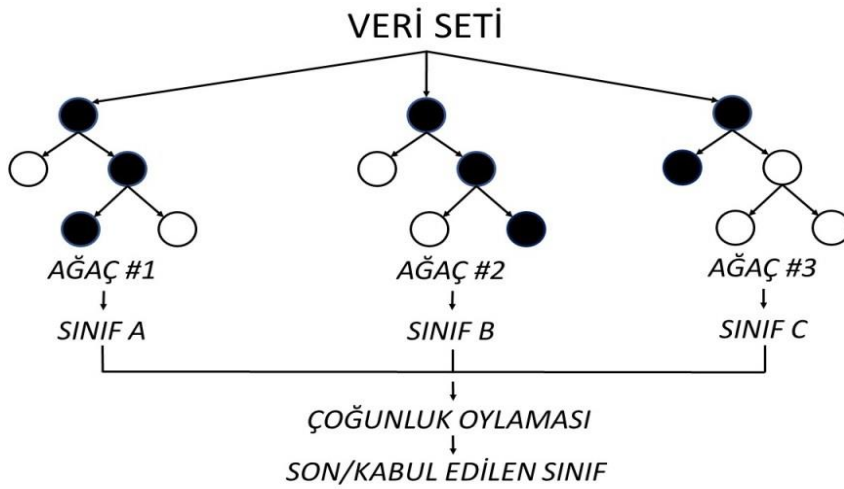
AdaBoost (Freund ve Schapire, 1995), uyarlamalı güçlendirmenin kısaltması, sürekli olarak değiştirilen veriler üzerinde bir dizi başka sınıflandırıcıyı birleştiren bir sınıflandırıcıdır. Eğitim setine bir sınıflandırıcı yerleştirerek başlamaktadır. Yanlış sınıflandırılan örnekler daha sonra ağırlaşmaktadır (artırılmaktadır), bu da sıradaki bir sonraki sınıflandırıcının onları doğru şekilde sınıflandırmak için "daha çok uğraşmasına" neden olmaktadır. Yeni nihai sınıflandırma, sınıflandırıcı serisinin çoğunluğu veya ağırlığıdır. Aykırı değerler sürekli olarak yanlış sınıflandırıldığından, bu örnekler giderek daha fazla etki sahibi olmaktadır ve sınıflandırıcı aşırıya kaçmaya eğilimlidir (Freund ve Schapire, 1995). AdaBoost sınıflandırıcısının karar sınırları Şekil 1.7.'de gösterilmiştir.



Şekil 1.7. Çeşitli veri kümelerindeki Adaboost sınıflandırıcı (Freund ve Schapire, 1995)

1.2.6.2. Rastgele orman

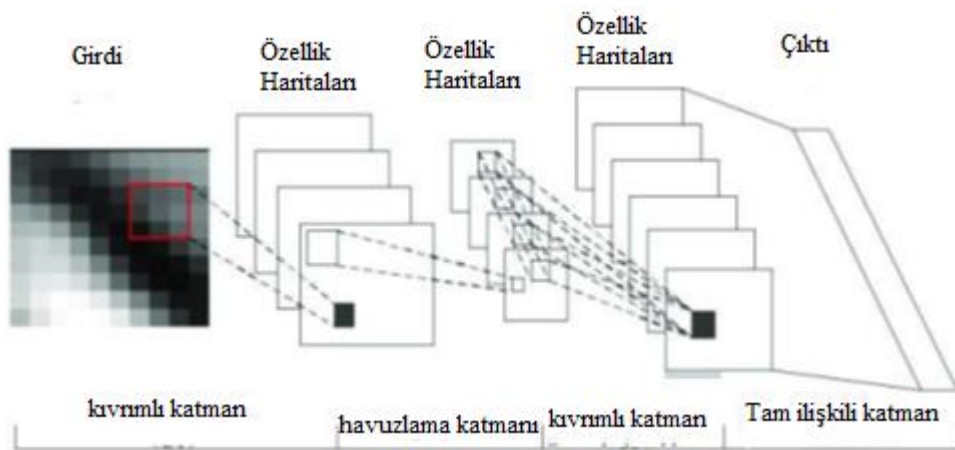
Rastgele Orman (RF), her bir ağaç için, bazı özelliklerin rastgele seçilmesi ve ardından ağaçtaki düğümlerin ayarlanması için bu özelliklerin kullanılarak birden fazla karar ağacının oluşturulduğu bir topluluk öğrenme yöntemidir. Dolayısıyla bir karar ağacı, eğitim verilerindeki özellik değerlerinin öğrenilmesiyle oluşturulan bir kurallar hiyerarşisidir. Hiyerarşiyi kullanarak, daha sonra her bir karar ağacı tarafından bir karar verilebilmektedir. Çoğunluk oylaması gerçekleştirerek, daha sonra bir RF sınıflandırıcısı tarafından temel alınan karar ağaçları kullanılarak bir sınıflandırma yapılabilir (Myles ve ark., 2004) (Şekil 1.8.).



Şekil 1.8. RF sınıflandırıcısının sınıflandırma akışının bir gösterimi (Golze ve ark., 2020)

1.2.6.3. Konvolüsyonel sinir ağları (KSA)

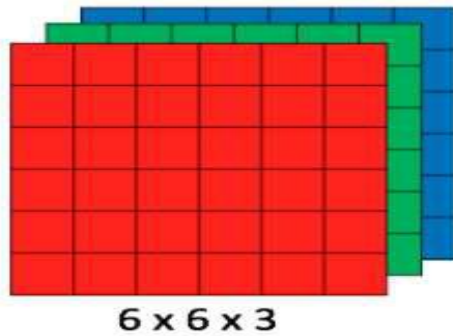
Derin öğrenme, bir beyninin derin yapısından ilham alan makine öğrenme yöntemidir (LeCun ve ark., 2006). Bu yöntem, birden çok gizli katmanın kullanıldığı, farklı özellik düzeylerinin soyutlanmasına izin veren derin bir mimari ile karakterize edilmektedir. 2006 yılında Hinton ve Salakhutdinov açgözlü tabakalı eğitim olarak adlandırdıkları nöron katmanlarının bu derin mimarisini eğitmek için yeni bir algoritma geliştirmişlerdir (Hinton ve Salakhutdinov, 2006). Sonrasında, bu yöntem daha etkili hale gelmiştir ve daha sonra önerilen derin ağlar üzerinde birçok eğitim için kullanılmaya başlanmıştır. En güçlü derin ağlardan biri, girdi verilerinin düşük ve yüksek seviyelerdeki özelliklerini çıkarmak için alt örnekleme gerçekleştiren birçok gizli katmandan oluşan derin bir ağ olan Evrişimli sinir ağıdır (Helwan ve ark., 2018). Bu ağ, uygulandığı farklı alanlarda, yani bilgisayarla görme (Krizhevsky ve ark., 2012), biyolojik hesaplama (Rios ve Kavuluru, 2015), tıbbi görüntülerin sınıflandırılması (Helwan ve ark., 2018) vb. gibi farklı alanlarda büyük bir verimlilik göstermiştir. Temel olarak, bu tür ağlar üç ana katmandan oluşmaktadır: evrişim katmanları, alt örnekleme veya havuzlama katmanları ve tam bağlantı katmanları. Her katman türü aşağıdaki paragrafta kısaca açıklanmıştır (Myles ve ark., 2004). Şekil 1.9.'da bir evrişimli sinir ağının (KSA) tipik bir mimarisini göstermektedir.



Şekil 1.9. Bir evrişimli sinir ağının (KSA) tipik mimarisi (Pereira ve ark., 2016)

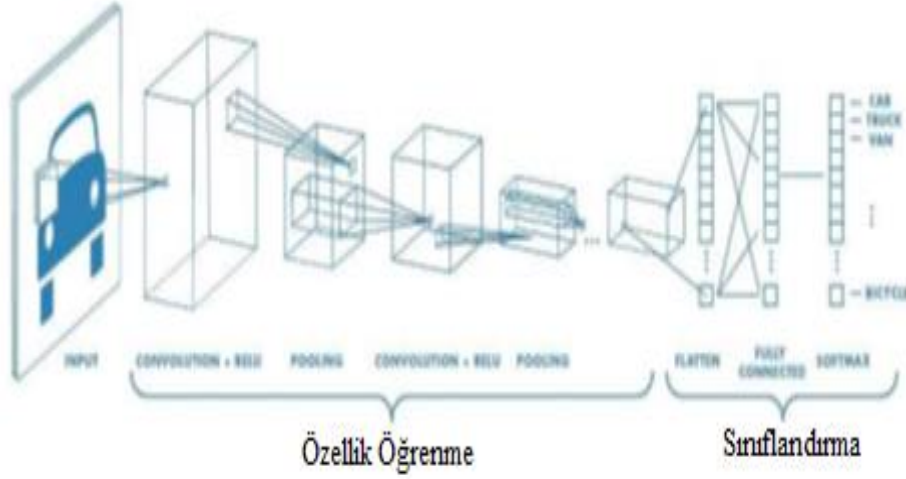
Evrişimli sinir ağları (ConvNet'ler veya KSA'lar), görüntü tanıma ve sınıflandırma, nesne algılama, yüz tanıma ve diğer uygulamalarda kullanılan sinir ağlarının önemli bir kategorisidir. KSA'lar, aşağıdakiler de dahil olmak üzere çeşitli alanlarda sıklıkla uygulanmaktadır (Pereira ve ark., 2016).

KSA görüntü sınıflandırmaları, daha sonra analiz edilen ve Köpek, Kedi, Kaplan veya Aslan gibi belirli gruplara ayrılan bir görüntünün girişini içerir. Görüntü bilgisayarlar tarafından bir piksel dizisi olarak algılanır ve görüntünün çözünürlüğü bu sınıflandırma işleminde çok önemli bir rol oynar. Görüntü çözünürlüğüne bağlı olarak, $h \times w \times d$ (h = Yükseklik, w = Genişlik, d = Boyut) değerini görecektir. Örneğin, $6 \times 6 \times 3$ dizi RGB matrisi (3 RGB değerlerini ifade eder) ve $4 \times 4 \times 1$ matris gri tonlamalı görüntü matrisi görüntüsüdür (Pereira ve ark., 2016).



Şekil 1.10. RGB Matrisinin Dizisi (LeCun ve ark., 2006)

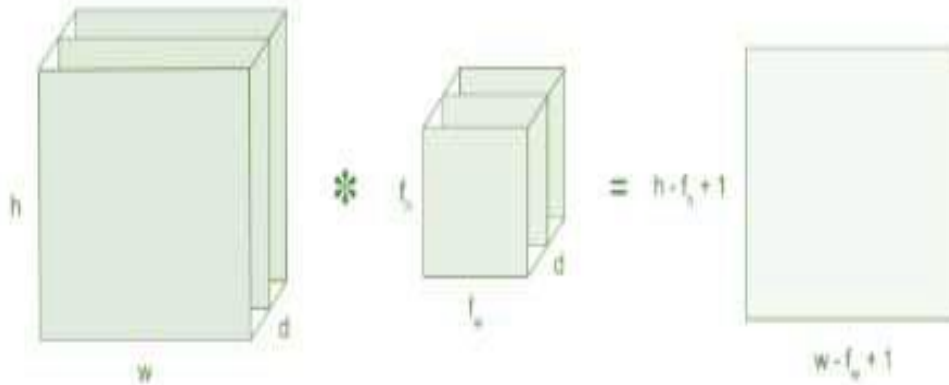
Derin öğrenme KSA modellerini eğitme ve test etme süreci, giriş görüntülerinin filtreleri (Kernels olarak bilinir), havuzlamayı, tamamen bağlı katmanları (FC) ve Softmax işlevini kapsayan bir dizi evrişim katmanı aracılığıyla çalıştırılmasını içerir. Bu teknik, bir nesnenin 0 ile 1 arasında bir olasılık değeri atayarak sınıflandırılmasını sağlar. Aşağıdaki şekil, bir girdi görüntüsünü işlemek ve nesneleri değerlere göre sınıflandırmak için tam bir KSA akışını göstermektedir (LeCun ve ark., 2006).



Şekil 1.11. Birçok evrişim katmanına sahip bir sinir ağı örneği

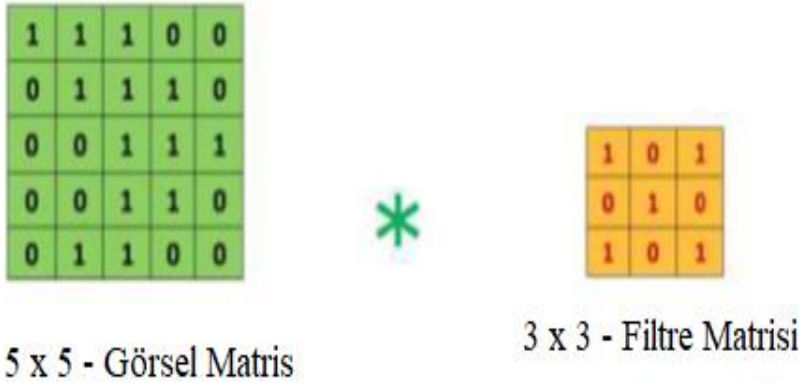
Evrişim Katmanı

Bir girdi görüntüsünden özellik çıkarmanın ilk aşamasına Evrişim denir. Girdi verilerinin küçük karelerini işleyerek Convolution, pikseller arasındaki korelasyonu etkili bir şekilde korur ve böylece görüntünün ayırt edici özelliklerini öğrenir. Görüntü matrisi ve bir filtre veya çekirdek gibi iki girdi alan matematiksel bir işlemdir (Hinton ve Salakhutdinov, 2006).



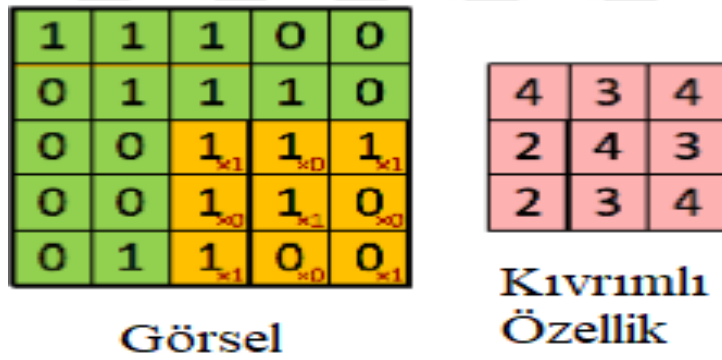
Şekil 1.12. Görüntü matrisi, çekirdek veya filtre matrisini çarpar (Hinton ve Salakhutdinov, 2006).

Görüntü piksel değerleri 0, 1 ve filtre matrisi 3 x 3 olan bir 5 x 5 ele alınır aşağıda gösterildiği gibi olmaktadır.



Şekil 1.13. Görüntü matrisi, çekirdek veya filtre matrisini çarpar

5 x 5 görüntü matrisinin evrişimi, aşağıda gösterildiği gibi “Özellik Haritası” olarak adlandırılan 3 x 3 filtre matrisi ile çarpılmaktadır (Helwan ve ark., 2018).



Şekil 1.14. 3 x 3 Çıktı matrisi (Krizhevsky ve ark., 2012)

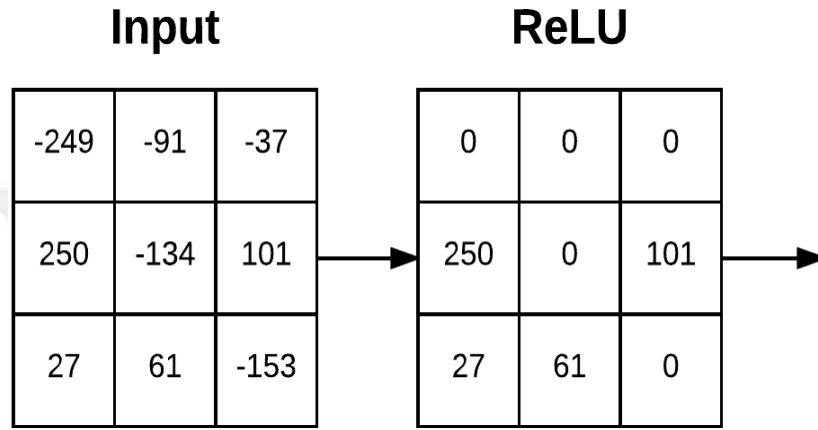
Bir görüntü, evrişim sırasında farklı filtreler uygulanarak kenar algılama, bulanıklaştırma ve keskinleştirme gibi çeşitli işlemlerden geçirilebilir. Örneğin, farklı filtreler uygulandığında, aşağıdaki örnekte gösterildiği gibi, evrişim ekranı farklı efektler ortaya çıkarır (Krizhevsky ve ark., 2012).

Operasyon	Filtre	kıvrımlı görüntü
Tanımlama	$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$	
Kenar algılama	$\begin{bmatrix} 1 & 0 & -1 \\ 0 & 0 & 0 \\ -1 & 0 & 1 \end{bmatrix}$	
	$\begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix}$	
	$\begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix}$	

Şekil 1.15. Bazı yaygın filtreler (Krizhevsky ve ark., 2012)

Rektifiye Doğrusal Birim (ReLU):

ReLU, doğrusal olmayan bir işlem için Düzeltilmiş Doğrusal Birim anlamına gelmektedir. Çıktı $f(x) = \max(0, x)$ 'dir. ReLU'nun amacı, ConvNet'imize doğrusal olmayanlığı tanıtmaktır. Çünkü gerçek dünya verileri ConvNet'inin negatif olmayan doğrusal değerler olacağını öğrenmesini istemektedir (Krizhevsky ve ark., 2012).



Şekil 1.16. ReLU işlemi (Helwan ve ark., 2018).

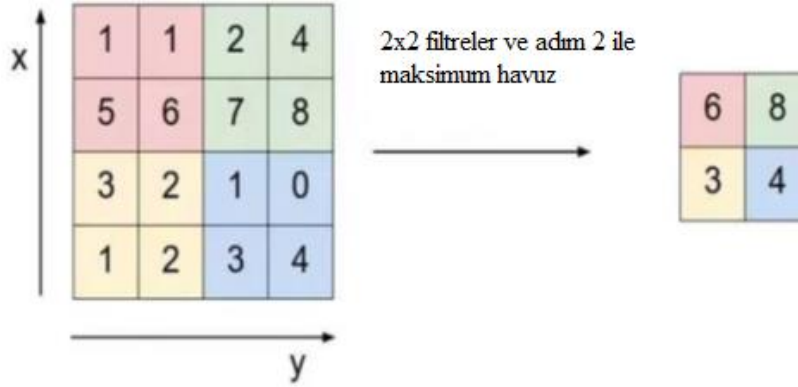
ReLU yerine tanh veya sigmoid gibi doğrusal olmayan başka fonksiyonlar da kullanılabilir. Veri bilimcilerin çoğu, performans açısından ReLU diğer ikisinden daha iyi olduğu için ReLU kullanmaktadır (Helwan ve ark., 2018)

Havuzlama katmanı:

Havuzlama katmanları bölümü, görüntüler çok büyük olduğunda parametre sayısını azaltmaktadır. Alt örnekleme olarak da adlandırılan mekansal havuzlama, her haritanın boyutsallığını azaltmaktadır ancak önemli bilgileri korumaktadır (Rios ve Kavuluru, 2015). Mekansal havuzlama farklı tiplerde olabilmektedir:

- Maksimum Havuzlama
- Ortalama Havuzlama

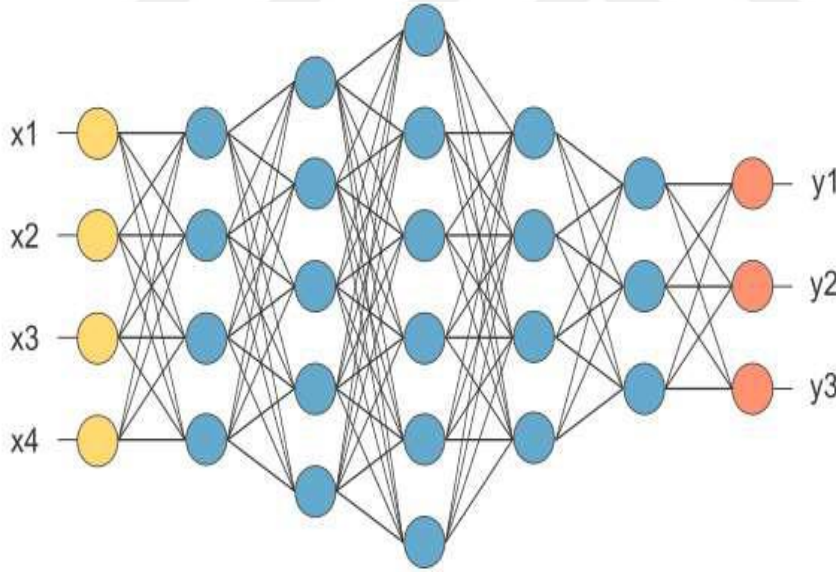
Maksimum havuzlama, düzeltilmiş özellik haritasından en büyük öğeyi almaktadır. En büyük elemanı almak aynı zamanda ortalama havuzlamayı da alabilmektedir (Krizhevsky ve ark., 2012).



Şekil 1.17. Maksimum Havuzlama (Krizhevsky ve ark., 2012)

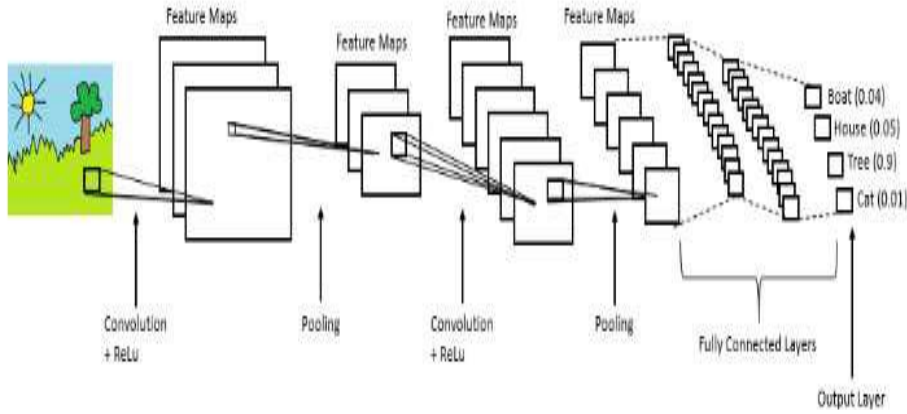
Tam bağlantılı katman (FC):

Katman FC katmanı olarak adlandırılmaktadır; matris vektöre düzleştirilmiştir ve onu sinir ağı gibi tamamen bağlantılı bir katman olacak şekilde beslemiştir (Helwan ve ark., 2018).



Şekil 1.18. Havuzlama sonrası FC katmanı olarak düzleştirilmiş katman (Helwan ve ark., 2018)

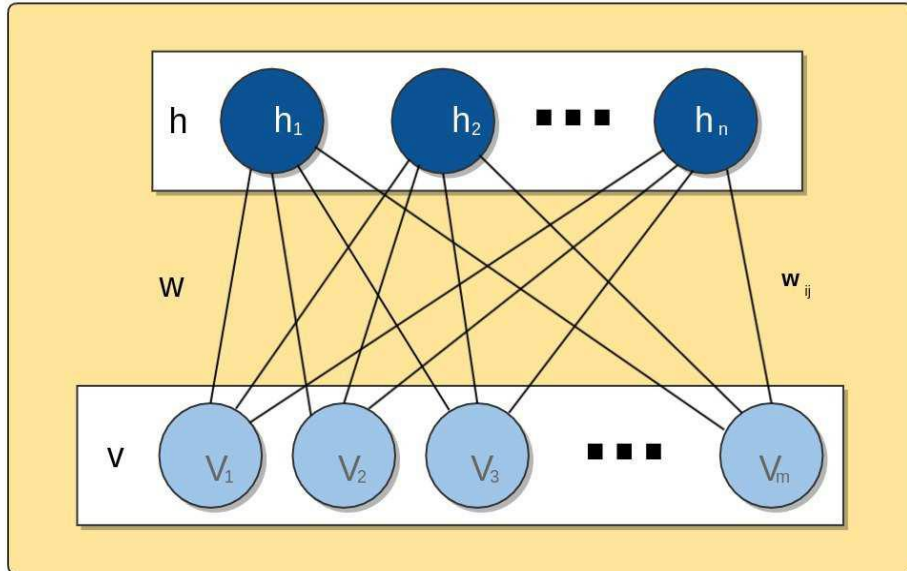
Şekil 1.16.'daki diyagramda öznitelik haritası matrisi vektör ($x_1, x_2, x_3, \dots$) olarak dönüştürülecektir. Tamamen bağlantılı katmanlarla, bu özellikler bir model oluşturmak için birleştirilmektedir. Son olarak çıktıları köpek, kedi, kamyon, araba vb. olarak sınıflandırmak için softmax veya sigmoid gibi bir aktivasyon fonksiyonları bulunmaktadır (Helwan ve ark., 2018).



Şekil 1.19. Genel KSA mimarisi (Helwan ve ark., 2018)

1.2.6.4. Derin boltzmann makinesi (DBN)

Kısıtlı Boltzmann Makineleri, öğrenme prosedürü bir enerji fonksiyonunun minimizasyonuna dayanan, görünür ve gizli bir nöron katmanından oluşan stokastik modellerdir. m birimli görünür bir katman v ve n birimli gizli bir katman h içeren bir Kısıtlı Boltzmann Makinesinin vanilya mimarisi Şekil 1.20.'de gösterilmektedir. Ayrıca, $W_{m \times n}$, her iki katman arasındaki ağırlıkları modelleyen gerçek değerli bir matrisi temsil etmektedir ve w_{ij} , görünür birim v_i ile gizli birim h_j arasındaki ağırlığı temsil etmektedir (Rios ve Kavuluru, 2015).



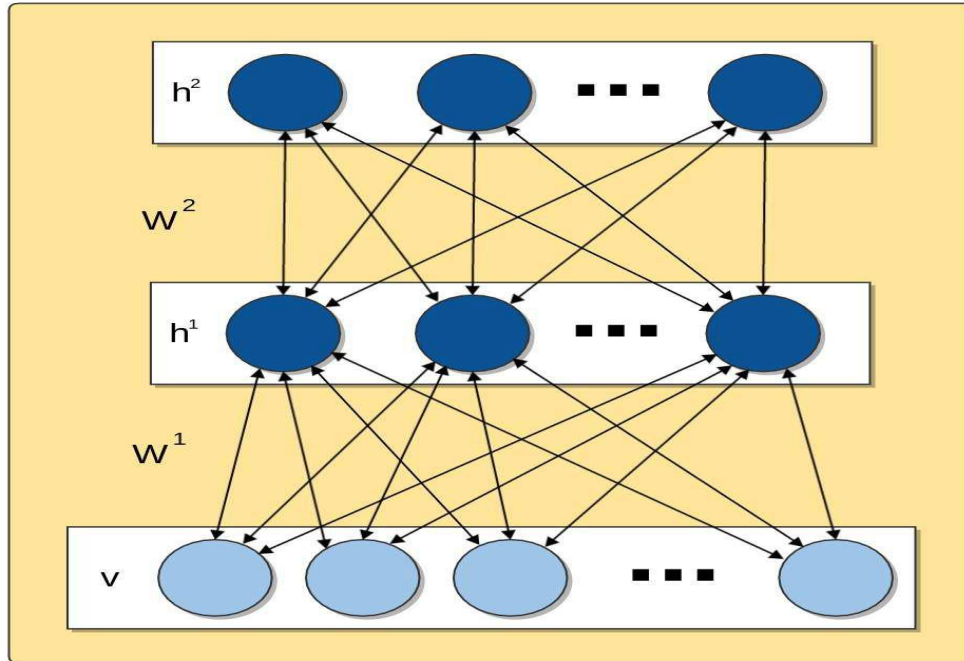
Şekil 1.20. RBM mimarisi (Krizhevsky ve ark., 2012)

DBN'ler, öğrenme algoritması ve Persistent Contrastive Divergence (PCD) kullanılarak açgözlü bir şekilde eğitilmiş bir dizi yığılmış RBM'den oluşan derin mimarilerdir. RBM, modeli belirli bir katmanda eğitirken i katmanındaki gizli birimlerin $i+1$ katmanına giriş birimleri olması dışında diğer katmanların birimlerinin durumlarını dikkate almamaktadır (Rios ve Kavuluru, 2015).

DBN'lerin yanı sıra DBM'ler, yığılmış RBM'leri kullanan girdi verilerinin daha karmaşık ve içsel temsillerini öğrenebilen derin mimarilerdir. Şekil 1.21.'de formülasyonun DBN 2'den hafif farklılıkları olan standart bir DBM'nin mimarisini göstermektedir.

Birinci ve ikinci katmanlarda h^1 gizli birimleri ve h^2 görünür birimleri temsil eden iki katmanlı bir DBM'nin enerjisi aşağıdaki gibi hesaplanabilmektedir:

$$E(v, h^1, h^2) = -\sum_{i=1}^{m^1} \sum_{j=1}^{n^1} v_i h_j^1 w_{ij}^1 - \sum_{i=1}^{m^2} \sum_{j=1}^{n^2} h_j^1 h_i^2 w_{ij}^2 \quad (1.6)$$



Şekil 1.21. İki gizli katmana sahip DBM mimarisi (Krizhevsky ve ark., 2012)

Şekil 1.21.'deki çift taraflı okların, DBM yaklaşık çıkarım prosedüründe kullanılan olağan aşağıdan yukarıya geçiş ek olarak yukarıdan aşağıya geri bildirimi temsil ettiğine dikkat edilmelidir.

Burada m^1 ve m^2 sırasıyla birinci ve ikinci katmanlardaki görünür birimlerin sayısını ve n^1 ve n^2 sırasıyla birinci ve ikinci katmanlardaki gizli birimlerin sayısını temsil etmektedir. Ayrıca, $W_{m^1 \times n^1}^1$ ve $W_{m^2 \times n^2}^2$ ağırlık matrisleri, sırasıyla v ve h^1 vektörleri ile h^1 ve h^2 vektörleri arasındaki bağlantıların ağırlıklarını kodlamaktadır. Önyargı terimleri, sadeleştirme amacıyla çıkarılmıştır. Karmaşıklığı nedeniyle, RBM tabanlı modellerin türevlerini hesaplamak imkansız bir iş haline gelmektedir. Böyle bir kısıtlamayla başa çıkmak için, Kontrast Diverjans algoritması kullanılabilmektedir ve görünür ve gizli birimlerin tahmini bir durumu örneklenebilmektedir. Böylece, görünen ve iki gizli birim üzerindeki koşullu olasılıklar şu şekilde verilmektedir (Salakhutdinov ve Hinton, 2006):

$$P(v_i = (1|h^1)) = \phi(\sum_{j=1}^{n^1} h_j^1 w_{ij}^1) \quad (1.7)$$

$$P(h_z^2 = (1|h^1)) = \phi(\sum_{i=1}^{m^2} h_i^1 w_{iz}^2) \quad (1.8)$$

ve

$$P(h_j^1 = (1|v, h^2)) = \phi(\sum_{i=1}^{m^1} v_i w_{ij}^1 + \sum_{z=1}^{n^2} h_z^2 w_{jz}^2) \quad (1.9)$$

Son olarak, üretici model aşağıdaki gibi yazılabilmektedir (Salakhutdinov ve Hinton, 2006):

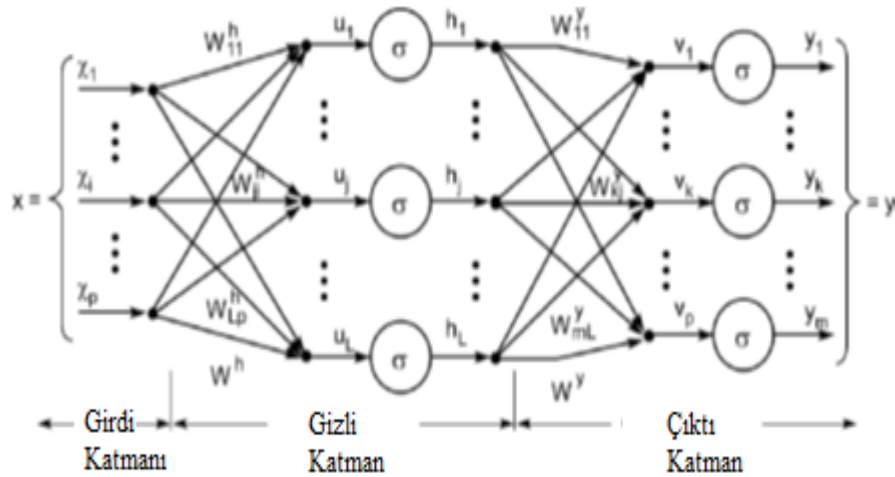
$$P(v) = P(h^1) P(1|v, h^2) \quad (1.10)$$

Burada $P(h^1) = \sum_v P(h^1, v)$ daha sonra $P(h^1) = \sum_{h^2} P(h^1, h^2)$ denkleminin yerini alacak olan ikinci RBM'nin öğrenme sürecine geçilmektedir. Böyle bir prosedür kullanılarak, denklemler 1.7-1.9 ve Kontrast Diverjans tarafından verilen koşullu olasılıklar kullanılarak, DBM parametreleri bir seferde bir katman

olarak öğrenilebilmektedir (Salakhutdinov ve Hinton, 2006). Daha sonra, daha doğru bir model elde etmek için ortalama alan tabanlı öğrenme uygulanabilir.

1.2.6.5. Çok katmanlı algılayıcı (ÇKA)

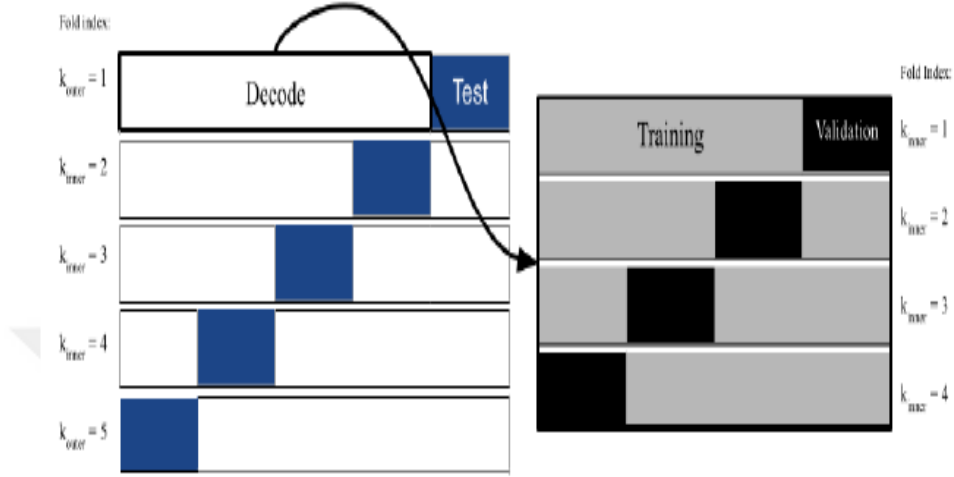
Çok Katmanlı Algılayıcı (ÇKA) sınıflandırıcısı, aralarında en az bir gizli katman bulunan bir girdi katmanı ve bir çıktı katmanından oluşan ileri beslemeli bir yapay sinir ağıdır. Giriş katmanında olmayan her bir düğüm değerini, önceki katmandaki düğüm değerlerinin ağırlıklı toplamı olarak hesaplayarak, girdi katmanı, çıktı katmanında tahmin edilen etiketlerle sonuçlanmak üzere ileriye doğru yayılacak değerleri kabul etmek için kullanılmaktadır. Konsept, bir ileri ve geri geçiş döngüsü kullanarak bir sinir ağını eğitmektir. İleri geçişte, numunelerin etiketlerini tahmin etmek için değerler giriş katmanından çıkış katmanına yayılmaktadır. Ortaya çıkan etiketleri gerçek etiketlerle karşılaştırarak, tipik olarak belirli bir hata fonksiyonunun kısmi türevlerini kullanarak ağırlıkları ayarlayarak ayarlamalar yapmak için geriye doğru bir geçiş kullanılabilir. Ağ ayarlandığında, yeni tahminler yapmak için yeni bir ileri geçiş tetiklenebilmektedir (Pedro ve ark., 2017). Şekil 1.22.'de bir ÇKA sınıflandırıcı örneğini göstermektedir.



Şekil 1.22. ÇKA sınıflandırıcısının sınıflandırma akışının bir gösterimi (Faghfour ve Frish, 2011)

Şekil 1.23.'deki beyaz, mavi, gri ve siyah kutular sırasıyla kod çözme, test, eğitim ve doğrulama verileridir. Her bir dış katlamada, kod çözme verileri, kod

çözme verileri kullanılarak bir iç çapraz doğrulamanın yapılabilmesi için 4 iç kata bölünmüştür. Çizim, Varoquaux ve ark. (2017) tarafından sağlanan açıklamaya göre yapılmıştır.



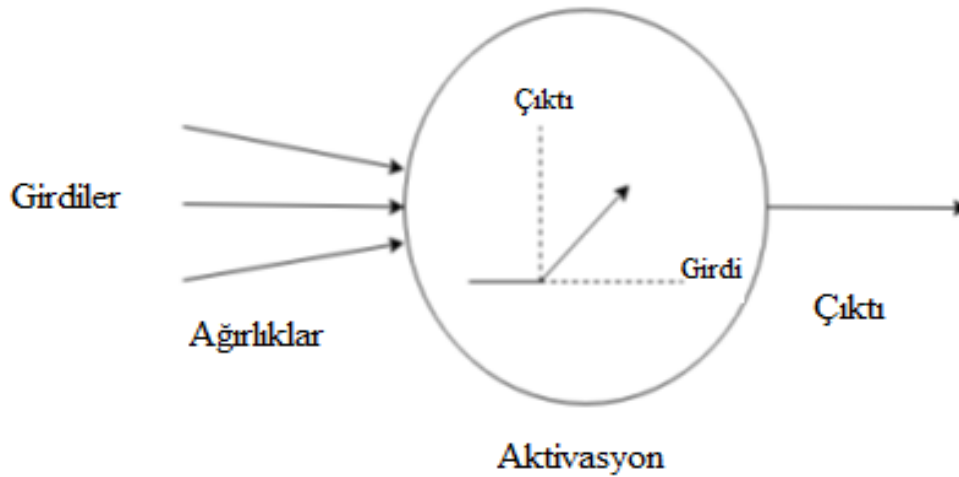
Şekil 1.23. Test için bir dış döngü ve doğrulama için bir iç döngü ile iç içe Çapraz Doğrulamanın bir resmi (Varoquaux ve ark., 2017)

Bir ÇKA sinir ağının, sıradan doğrusal regresyonla birçok benzerliği bulunmaktadır. Lineer regresyonda (Montgomery ve ark., 2021), veri noktalarına (x,y) en iyi uyan $f(x) = w \cdot x + b$ model fonksiyonunu bulmak için w ağırlığı ve b önyargısı ayarlanmaya çalışılmaktadır. Yani, $f(x)$ ve y arasındaki toplam hatayı, kayıp fonksiyonunu en aza indiren $f(x)$ fonksiyonunun bulunması gerekmektedir.

Bir ÇKA ağındaki en küçük birime nöron denmektedir. Her nöron, çıktıyı oluşturan doğrusal olmayan bir aktivasyon fonksiyonuna geçirilen ağırlıklı girdi toplamından tek bir çıktı oluşturan bir bilgi işlem hücresidir. Doğrusal regresyona benzer şekilde, çıktıyı istenen değere yaklaştırmak için girdi üzerindeki ağırlıklar ayarlanmaya çalışılmaktadır. Her nöronun girişinde ayrıca bir önyargı terimi bulunmaktadır. Önyargı genellikle 1 sayısına ayarlanmaktadır ve diğer girdilerle birlikte ağırlıklandırılmaktadır. Aktivasyon işlevi, bir nörondan çıktının belirli bir eşikte tutulmasına yardımcı olmaktadır. Aksi takdirde, çıktı çok büyüyebilmekte ve optimal ağırlıkları bulmak için problemler yaratabilmektedir. Aktivasyon işlevi

ayrıca doğrusal olmayanlığı enjekte etmek için kullanılmaktadır ve ağıın verilerdeki doğrusal olmayan modelleri bulmasını sağlamaktadır (Noriega, 2005).

Bir ÇKA ağı, bir girdi katmanı, bir dizi gizli katman ve bir çıktı katmanından oluşmaktadır (Gardner ve Dorling, 1998). Her katman birkaç düğüm tarafından oluşturulmaktadır. Giriş katmanındaki düğümler, giriş vektörü x 'in x_i girişlerinden oluşmaktadır. Gizli katmandaki düğümler, yukarıda açıklanan nöronlardan oluşmaktadır. Ayrıca çıktı katmanı, ağıın sayısal çıktısına sahip düğümlerden oluşmaktadır ve herhangi bir boyutta olabilmektedir. Her düğüm, önceki katmandaki düğümlere, her bağlantı için bir ağırlık (w) ile bağlanmaktadır. Ayrıca her düğüm için bir önyargı terimi, b eklenmektedir. Böylece h katmanındaki j düğümü aşağıdaki girdiyi almaktadır.



Şekil 1.24. 3 girişli tek bir nöron ve ReLU aktivasyon fonksiyonu (Montgomery ve ark., 2021)

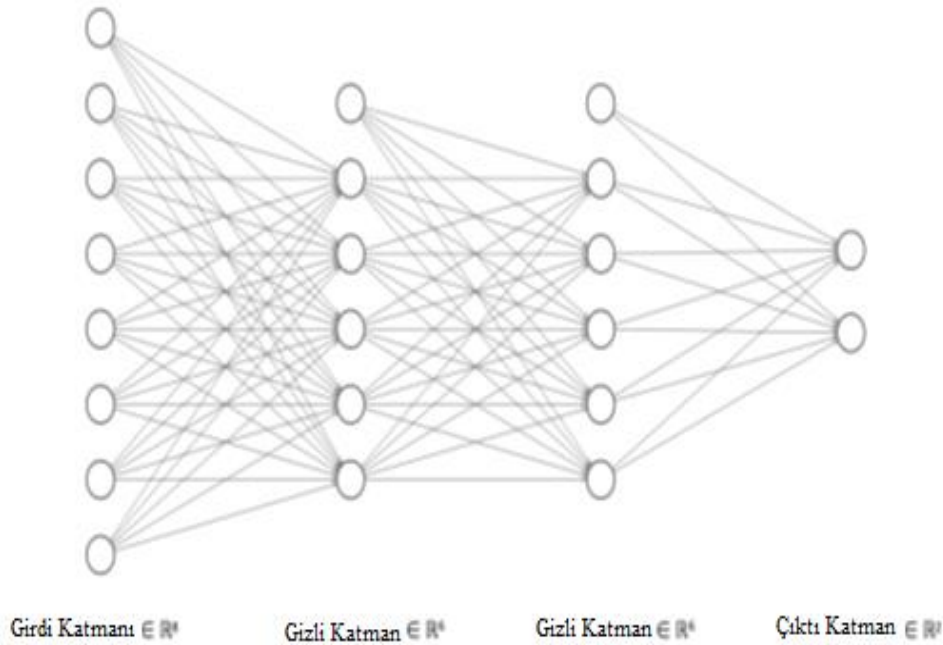
$$x_j^h = \sum_i^N y_i^{h-1} w_{ij}^{h-1} + b_i^h \quad (1.11)$$

N , önceki katmandaki düğüm sayısıdır ($h - 1$). y_i^{h-1} , önceki katmanın her bir düğümünün çıktısıdır ve $w_{i,j}^{h-1}$, bağlantı ağırlığıdır. Ayrıca, b_j^h önyargı terimidir. İlk gizli katman için, önceki katman y_i^{h-1} 'den gelen çıktılar, girdi verileridir. Nöronların aktivasyon fonksiyonu, diğer katmanlardan gelen çıktıyı belirlemektedir. Geçmişte bir sigmoid fonksiyonunun aktivasyon fonksiyonu olarak kullanılması popülerdi,

ancak bugün ReLU fonksiyonu en yaygın olarak olarak kullanılmaktadır (Sharma, 2017).

$$H(x) = \max(0, x) \quad (1.12)$$

Böylece bir nöronun çıktısı $y_i^h = \max(0, x_i^h)$ olmaktadır. ReLU aktivasyon fonksiyonu ve 3 girişi olan bir nöronun yapısı Şekil 1.22.'de gösterilmiştir. Şekil 1.23.'te bir ÇKA'nın yapısı verilmiştir. Bu özel yapı 7 boyutlu girdi, her birinde 5 düğüm bulunan 2 gizli katman ve 2 çıktı boyutundan oluşmaktadır. Uyarı okuyucusu, çıktı dışında her katman için bu boyutların şekilde gösterilen sayılarla kapalı olduğunu fark edebilmektedir. Bu yapı, örneğin, önceki 7 günün ortalama sıcaklığından (giriş) yarının güneşli mi yoksa bulutlu mu olacağını (çıkıtı) tahmin eden model fonksiyonunu bulmak için kullanılabilir. Bu fonksiyon modelini bulmak için son 2 yılın ortalama sıcaklığına ilişkin veriler bulunmaktadır. Dolayısıyla bu 2 yılın girdisi ve karşılık gelen çıktısı bilinmektedir. En iyi model fonksiyonunu bulmak için, tıpkı lineer regresyonda yapıldığı gibi yapılmaktadır, ağırlıkları ve önyargıları, model fonksiyonu çıktısı $f(x)$ ile bilinen çıktı y 'ye en iyi uyacak şekilde ayarlanmaktadır (Noriega, 2005).



Şekil 1.25. İleri Beslemeli Çok Katmanlı Algılayıcı (Noriega, 2005)

Ağırlıkları ve önyargıları ayarlama yöntemi, gradyan inişidir (Ruder, 2016), Gradyan iniş, bir amaç fonksiyonunun minimumuna ulaşmak için kullanılan bir yöntemdir, bu durumdaki kayıp fonksiyonu, modelin parametreleri tarafından parametrelendirilmiştir. Örneğin, kayıp fonksiyonu $E \propto (f(x) - y)$ olarak tanımlanırsa ve burada $f(x)$ 'in ağırlık çıktısını tanımladığı ve bu çıktının önceki katmanın çıktılarının her birinden etkilendiği varsayılırsa; daha sonra, ağırlıklara ve önyargılara göre gradyanını alırken bu hatanın geriye doğru yayılmasına izin verilmektedir. Ardından, öğrenme oranı olarak adlandırılan bir faktör olan η ile çarpılarak gradyanın negatif yönündeki ağırlıkları ve önyargıları güncellenmektedir. Öğrenme oranı, ağırlıkların ne kadar güncellenmesi gerektiğini kontrol etmektedir (Gardner ve Dorling, 1998).

1.3. Performans Metrikleri

Makine öğrenimindeki önemli bir konu, bir bilgisayar programının hangi sonuçlarının uygun olduğunu ve hangilerinin hatalar içerdiğini nasıl fark ettiği sorunudur. Sınıflandırma görevlerinin değerlendirilmesi normalde veri kümesini bir eğitim veri kümesine ve bir test veri kümesine bölerek yapılır. Makine öğrenimi algoritması daha sonra birincisi üzerinde eğitilirken, test veri seti algoritmanın kalitesini değerlendirmek için performans göstergelerini hesaplamak için kullanılır. Makine öğrenimi algoritmaları için yaygın bir problem, sınırlı test ve eğitim verilerine erişimde yatmaktadır. Bu nedenle, bu programları değerlendirirken aşırı uyum ciddi bir sorun olabilir (Yasen ve Du, 2016). Makine öğrenimi algoritmalarının performanslarının değerlendirilmesi için performans metrikleri bulunmaktadır. Bu performans metrikleri duyarlılık, doğruluk, kesinlik ve F1 puanıdır.

Duyarlılık (veya Geri Çağırma), Gerçek Pozitif örneklerin sayısının Gerçek Pozitif ve Yanlış Negatif örneklerin sayısının toplamına bölünmesiyle elde edilen orandır. Bu nedenle, hassasiyet ne kadar yüksekse, doğru bir şekilde 1 olarak tahmin edilen sınıf etiketi miktarı da daha yüksektir. Bu performans ölçüsü, herhangi bir pozitif sınıf etiketinin yanlış sınıflandırılmasını göze alamayacağımız uygulamalarda önemlidir (Bahl, 2017).

Doğruluk, Gerçek Negatif örneklerin sayısının, Gerçek Negatif ve Yanlış Pozitif örneklerin sayısının toplamına bölünmesiyle elde edilen orandır. Bu nedenle, doğruluk daha yüksek, doğru olarak 0 olarak tahmin edilen sınıf etiketi miktarı daha yüksektir. Bu performans ölçüsü, 1 [Duyarlılık ve Doğruluk Vikipedi Sayfası] olarak sınıflandırılmış herhangi bir negatif sınıf etiketine sahip olmayı göze alamayacağımız uygulamalarda önemlidir. Örneğin, bilgi alımı için, bir arama motorunda, yanlış bir pozitifte sahip olmak, arama motorunun işlevselliği ve işi için maliyetli olabilir (Bahl, 2017).

Kesinlik (veya Pozitif Tahmine Dayalı değer), Gerçek Pozitif örneklerin sayısının sayısının, Gerçek pozitif ve Yanlış pozitif örneklerin sayısının toplamına bölünmesiyle elde edilen orandır. Burada, hassasiyet ne kadar yüksekse, tahmin edilen toplam pozitif sınıf etiketlerinden tahmin edilen pozitif sınıf etiketlerinin fraksiyon miktarı daha yüksektir (Bahl, 2017).

F1 puanı, sınıflandırma testinin doğruluğunun bir ölçüsüdür. Hem Hassasiyeti (Geri Çağırma) hem de Hassasiyeti dikkate alır. F1 skoru, hassasiyet ve hatırlamanın (hassasiyet) ağırlıklı harmonik ortalamasıdır (Bahl, 2017).

1.4. Parkinson Hastalığı Teşhisi

PH'nin otomatik teşhisi, tıp uzmanlarının iş yükünü azaltırken PH'nin daha doğru ve verimli bir şekilde teşhis edilmesini sağlamanın bir yolu olarak birçok kişinin ilgisini çekmiştir.

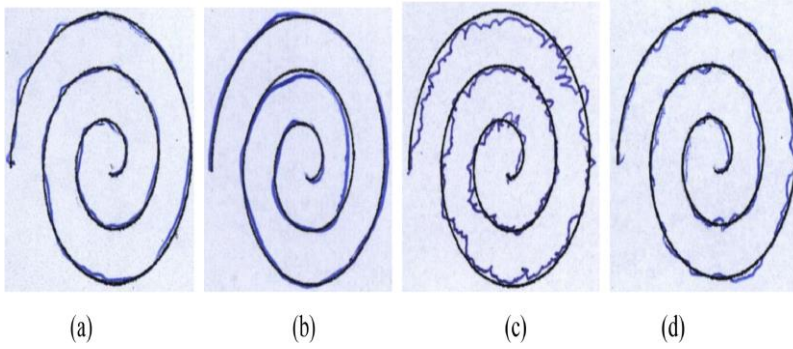
1.4.1. Çizim muayenesi

Daha önce belirtildiği gibi, çizim muayenesi, PH'nin ana semptomlarını yakalama yeteneği sayesinde, genellikle PH'nin klinik teşhisinin bir parçasıdır (Gupta ve ark., 2018). Bu nedenle, bu prosedürü otomatikleştirmek için birkaç model önerilmiştir. Çoğu zaman, otomasyon, hasta tarafından yapılan çizimlerden ilgili özelliklerin çıkarılması ve daha sonra bu özelliklere dayalı bir tahmin yapmak için önceden eğitilmiş bir modelin kullanılmasıyla bir teşhisin yapılacağını ifade

etmektedir (Poon ve ark., 2019). Şekil 1.24.'de Memedi ve ark. (2015), katılımcılardan dijital bir cihaz üzerinde spiral bir şablon üzerinde çizim yapmalarının istendiği çalışmasında kullanılan cihazı göstermektedir. Şekil 1.25.'de taranan çizimler kullanılarak PH'yi otomatik olarak teşhis etmek için bir algoritmanın önerildiği Gupta ve ark. (2018) çalışmasından bazı çizimleri göstermektedir.



Şekil 1.26. Spiral şablon üzerinde çizim yapan bir katılımcının resmi (Memedi ve ark., 2015)

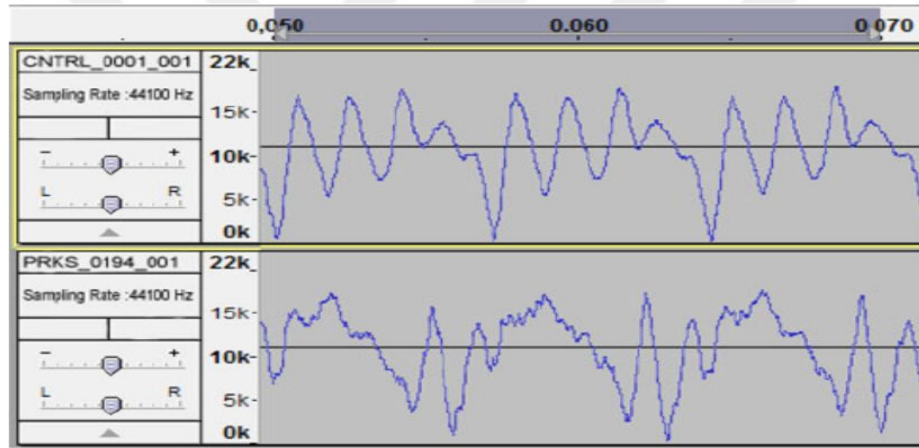


Şekil 1.27. Bazı spiral çizimleri (Gupta ve ark., 2018)

Şekil 1.27.'deki çizimler Gupta ve ark. (2018)'in çalışmasında yer alan ve her biri farklı katılımcılara ait olan çizimlerdir. (a) 58 yaşında sağlıklı katılımcı ve (b) 28 yaşında sağlıklı katılımcı ve (c) 56 yaşında PH hastası ve (d) 65 yaşında yaşlı PH hastasıdır.

1.4.2. Ses muayenesi

Pratikte çizim muayeneleri kadar yaygın olarak kullanılsa da, ses muayeneleri, PH'nin erken tespitini sağlamanın bir yolu olarak birçok makine öğrenimi uzmanının dikkatini çekmiştir. Bunun nedeni, hem erken PH hastaları arasında ses bozukluğunun en yaygın semptomlar arasında yer alması (Liu ve Özsü, 2009) hem de erken PH hastalarının, ses anormallikleri insanlar tarafından algılanamayacak kadar belirsiz olabilen kişiler olma eğiliminde olmasıdır. Başka bir deyişle, ses muayeneleri yoluyla PH'nin otomatik teşhisinin, PH'yi daha erken tespit etmesine yardımcı olabileceği ve bu nedenle araştırmaya değer olduğu varsayılmaktadır (Cleophas ve Zwinderman, 2013).



Şekil 1.28. Bir dizi dalga formu (Şakar ve ark., 2013)

Şekil 1.28.'de yer alan dalga formunda, üst dalga formu sağlıklı bir bireyin sesine aittir. Alt dalga formu bir PH hastasına aittir. Y eksenini sinyalin genliğini gösterirken, x eksenini zaman çizelgesini gösterir (Şakar ve ark., 2013). Çizim muayenelerinden farklı olarak, bir ses muayenesi, kişinin PH'yi ses bozukluğu yoluyla teşhis etmesine izin verir. Tipik olarak bu, hastanın sesini bir dalga biçimine dönüştürerek ve ardından o dalga biçiminden çıkarılan özelliklere dayalı bir tahmin yapmak için önceden eğitilmiş bir model kullanarak yapılır (Naranjo ve ark., 2017). Bunu göstermek için, Şekil 1.26., Şakar ve ark. (2013) sağlıklı bir kişiye ait dalga formunun bir PH hastasına ait dalga formuyla karşılaştırıldığı çalışmasından bir grafiği göstermektedir.

1.5. Özellik Seçimi

Tipik olarak, PH'nin otomatik teşhisi, PH'nin göstergeleri olarak kullanılan özelliklerin çıkarılmasını içermektedir. Örneğin, titremenin bir göstergesi olarak hastanın çizim hızındaki standart sapmayı veya rijidite ve bradikinezinin bir göstergesi olarak hastanın çizim hızının ortalama değerini ölçmek istenebilir (Poon ve ark., 2019). Listeler uzayıp gidebilir, bu da teşhisin öngörücü gücü için hangi özelliklerin gerekli olduğunu belirlemeyi zorlaştırır. Gerçekten de, özellik tabanlı bir makine öğrenme algoritmasının hem verimliliği hem de tahmin gücü için en uygun özellik kümesini bulmak esastır (Kumar ve Minz, 2014). Aynı zamanda, bilginin işlenmesi zaman ve kaynak gerektirdiğinden, Büyük Veriyi işlemek için tasarlanmış bir algoritma için yüksek boyutluluğu kritik bir sorun haline getirirken, çok yüksek boyutluluğa sahip olmak, modelin boyut başına çok az numuneye sahip olmasına yol açabilir (Liu ve Gillies, 2016). Sonuç olarak, model, daha az ilgili veya gürültülü verilerden kolayca etkilenecek ve böylece fazla uyum ve yüksek varyansa yol açacaktır (Lever ve ark., 2016). Bu özellik seçimine, verinin gürültülü ve alakasız kısmını kaldırarak bir makine öğrenme algoritmasının performansını iyileştirme potansiyeli verir. Bir özellik seçimi yapmak için arama stratejisine, arama yönüne karar verilmelidir (Venkatesh ve Anuradha, 2019).

1.5.1. Arama stratejisi

Kumar ve ark. (2014) tarafından belirtildiği gibi, bir özellik seçim metodolojisinde kullanılabilecek ana arama stratejileri kategorileri vardır:

- Sıralı arama: Bu, açgözlü arama algoritması gibi öğeleri sırayla ekleyerek veya kaldırarak özellikleri seçenlere atıfta bulunur. Basit olmakla birlikte, bu, yerel bir optimumda takılıp kalma eğiliminde olma dezavantajına sahiptir.
- Üstel arama: Bu, ya kapsamlı arama yoluyla ya da bir buluşsal yöntem yoluyla farklı özellik alt kümelerini araştıranlara atıfta bulunur. Bu, kapsamlı bilgi işlem kaynağına ve uzun bilgi işlem süresine olan ihtiyacı nedeniyle genellikle önlenir.

- **Rastgele arama:** Bu, rastgele seçilen özelliklerle başlayan ve rastgele öğelerle yeniden oluşturma ve seçimlerle ilerleyen genetik algorithma olduğu gibi, aramada rastgele öğelere sahip olanları ifade eder. Bu, tüm olasılıkları denemek zorunda kalmadan yerel bir optimumda takılıp kalmamak için bir yaklaşım olarak birçok araştırmacının dikkatini çekmiştir.

1.5.2. Arama yönü

Arama stratejisine karar verdikten sonra arama yönüne de karar verilmelidir. Genel olarak, arama yönü aşağıdakilerden biri olabilir (Kumar ve Minz, 2014):

- **İleri arama:** Bu, önceki seçenekleri değiştirmeden kademeli olarak yeni özellikler eklemek için boş bir listeyle başlayanlara atıfta bulunur.
- **Geriye arama:** Bu, önceki seçenekleri değiştirmeden özellikleri sırayla kaldırmak için tam verilerle başlayanlara atıfta bulunur.
- **Bileşik arama:** Bu, herhangi bir rastgele öğe eklemekten ileri seçimi geriye doğru eleme ile birleştirenlere atıfta bulunur.
- **Rastgele arama:** Bu, bir öğeyi ekleyip çıkarmamaya rastgele karar vermek için bir algoritma kullananlara atıfta bulunur.

1.6. Rastgele Aşırı Örnekleme

Katmanlı k-katlı Çapraz Doğrulama, önyargılı değerlendirmeyi azaltmak için kullanılabilirken, eğitim veri setinde azınlık sınıfına ait örnekleri rastgele çoğaltmak için bir teknik olarak rastgele aşırı örnekleme, makine öğrenme algoritmalarını çoğunluk sınıfına karşı önyargılı olmaktan alıkoymak için kullanılabilecek bir tekniktir. Aşırı uyumlanmaya yol açma riskiyle basit olmakla birlikte, rastgele aşırı örnekleme ampirik çalışmalarda tatmin edici bir performans göstermiştir. Bu nedenle, rastgele aşırı örnekleme en popüler yeniden örnekleme stratejilerinden biri olmuştur (Cheng ve ark., 2019).

2. ÖNCEKİ ÇALIŞMALAR

Bu başlık altında geleneksel makine öğrenimi algoritmalarıyla parkinson hastalığı, zaman serisi veri setini kullanan KSA ile parkinson hastalığı, görüntüleme veri setini kullanan KSA ile parkinson hastalığı, parkinson hastalığı ile ilgili diğer çalışmalar incelenmektedir.

Yıldırım (2020), Uzmanlar, Parkinson hastalığını sesli sinyallerle teşhis etmek için hibrit makine öğrenimi tabanlı modelleri ve cinsiyete özgü yaklaşımları araştırıyor. Teşhisi daha doğru hale getirmek için karar ağaçları ve Destek Vektör Makineleri kullanılıyor. Kullanılan veri seti, yüksek boyutlu özelliklere ve örnekleme bilgilerine sahip olan PDC'dir. Oluşturulan modeller, hastalığı teşhis etme potansiyellerini gösteren umut verici başarı oranları göstermiştir. Ayrıca, cinsiyete dayalı veri setlerinin dahil edilmesinin teşhis sürecini geliştirdiği gözlemlenmiştir.

Esmer ve ark. (2020), Bu projenin amacı, PH'den mustarip olabilecek bireylerin ses kayıtlarından oluşturulan bir veri setini kullanarak doktorların tıbbi kararlar vermelerine yardımcı olmaktır. Veriler bir dengeleme sürecinden geçti ve sistematik olarak örneklendi. Bir özellik seçme algoritması kullanılarak, her bir özelliğin etiket üzerindeki etkisi hesaplandı ve veri kümeleri oluşturuldu. Kullanılan sınıflandırma algoritmaları Karar Ağacı, Destek Vektör Makineleri ve K En Yakın Komşu Algoritması idi ve performans Doğruluk, Hassasiyet, Spesifiklik, F-Ölçü, Kappa ve Auc kullanılarak değerlendirildi. En iyi performans gösteren veri grubu ve sınıflandırma algoritmasına dayalı olarak bir model oluşturulmuştur.

Gerger (2019), Parkinson hastalığını teşhis etmek için araştırmacılar, sarmalları analiz etmek için k-en yakın komşu ve karar ağacı örüntü tanıma yöntemlerini kullandılar. İstanbul Üniversitesi Cerrahpaşa Tıp Fakültesi Nöroloji Anabilim Dalı bu çalışmada kullanılan verileri sağladı ve sonuçlar çapraz doğrulama yoluyla doğrulandı. Deneyler, karar ağacı sınıflandırmasında 1.00 doğruluk değeri üreten

genetik algoritma ile özellik seçme yöntemi ile önerilen yöntemin oldukça etkili olduğunu ortaya koymuştur.

Torun ve ark. (2015), Bu çalışmanın amacı, Parkinson hastalığının ses sinyallerine dayalı olarak sınıflandırılmasında beş farklı makine öğrenme tekniğinin (KYK, ROS, SVM, NB ve KA) etkinliğini değerlendirmektir. Bunu başarmak için, yüksek boyutlu nitelikler ve örnekleme içeren yeni tanıtılan PDC veri seti kullanıldı. Deneysel analiz oldukça doğru sonuçlar verdi ve yöntemler istatistiksel karşılaştırmaya tabi tutuldu. Ayrıca, TBA ve DDA boyut küçültme tekniklerinin performans üzerindeki etkisi, yukarıda belirtilen bulgularla birlikte değerlendirilmiştir.

Bizal (2014), Makine öğrenmesi teknikleri kullanılarak Parkinson Hastalığı verileri üzerinde incelemeler yapılmış ve hastalığın belirlenmesi amaçlanmıştır. Sınıflandırma tekniği olarak Naive Bayes, K-En Yakın Komşuluk, Destek Vektör Makineleri, Lojistik Regresyon Analizi, Diskriminant Analiz yöntemleri kullanılmıştır. Bu çalışmada Little ve ark. tarafından oluşturulan (23 Parkinson hastası ve 8 sağlıklı birey) Oxford Parkinson's Disease Detection (OPD) veri seti kullanılmıştır. Bu veri seti üzerinde Knn diğer sınıflandırma tekniklerine göre en başarılı sonucu elde etmiştir.

Pereira ve ark. (2015), PH teşhisinde derin olmayan öğrenme algoritmaları kullanmıştır. Araştırmaları, 1-“HandPD” adlı halka açık bir veri seti oluşturmayı, 2-görüntülerden bazı özellikleri çıkarmak için görüntü işleme için bir boru hattı oluşturmayı temel almıştır. İki ana görevi vurgulamışlardır. Bunlar; görüntü işleme ve özellik çıkarmadır. Görüntü işleme görevi, görüntü deneğin izlemesi istenen şablonu içerdiğinden, spiral şablon ST'yi ve el yazısı izi HT'yi çıkarmak için birkaç aşamadan oluşan bir boru hattı içermektedir. Özellik çıkarımı, ST ve HT arasındaki farkları, aralarındaki dokuz istatistiksel değeri hesaplayarak değerlendirmeyi amaçlamıştır. Deneylerinde üç sınıflandırıcı model kullandılar: - Naive Bayes (NB), Optimum-Path Forest (OPF) ve SVM kullanılmıştır.

55 kişiden “HandPD” adlı bir veri seti oluşturulmuştur. Bunlar; 37 (25 erkek ve 12 kadın) hasta ve 18 (6 erkek ve 12 kadın) sağlıklı kontrol şeklinde belirlenmiştir. Bu veri seti, spiral el yazısı test görüntülerini içermektedir. Her muayenede birey başına yaklaşık dört spirale sahip oldukları için toplamda 373 örnek kullanılmıştır. Bu örnekleri %90 eğitim ve %10 test seti olarak ayırmışlardır.

Çalışmalarını hatırlama, kesinlik, F-skoru ve sınıflandırma doğruluğuna göre değerlendirmişlerdir. Sonuçlar, NB'nin en yüksek doğruluğu (%78.9) elde ettiğini, ardından OPF'nin (%77.1) ve SVM'nin (%75.8) geldiğini göstermiştir. Çalışmalarını tamamladıktan sonra, sınıflandırıcıların hatırlama oranında NB dışında hasta grubu ile karşılaştırıldığında kontrol grubunu ayırt etmede iyi bir performans elde edemediklerini fark etmişlerdir.

Badem (2019)'in yapmış olduğu bir çalışmada, Parkinson hastalığının ses sinyalleri temelli sınıflandırılmasında Yüksek boyutlu öznelik içeren PDC veri seti kullanılmıştır. Veri setinde bulunan hastalığı sınıflandırılmak için örneklem tabanlı algoritmalarından KYK, ensemble ve karar ağacı tabanlı algoritmalar'dan ROS ve KA, olasılık ve margin tabanlı algoritmalarından ise NB ve DVM makine öğrenmesi algoritmaları kullanılmıştır. Yapılan deneyler sonucunda, önemli derecede yüksek doğruluk oranları tespit edilmiştir. Ek olarak, uygulanan algoritmaların performansları istatistiksel metotlar kullanılarak karşılaştırılmıştır. Ayrıca, TBA ve DDA gibi boyut azaltma tekniklerinin sonuçlar üzerindeki etkisi de incelenmiştir.

Gerger (2019), Parkinson hastalığı teşhis spiralleri, örüntü tanıma teknikleri kullanılarak analiz edildi. Hastaların spiral çizimleri karakter olarak kabul edildi ve bir veri setine çevrildiğinde yüksek bir plow n sorununa neden oldu. Bu konuyu ele almak için çalışma, genetik algoritmaların özellik seçimi üzerindeki etkisini araştırdı. Sınıflandırma, hem k-en yakın komşu hem de karar ağaçları yöntemlerini içeriyordu ve doğrulama işlevi gören çapraz bırakma doğrulama yöntemi. Çalışmada kullanılan veriler İstanbul Üniversitesi Cerrahpaşa Tıp Fakültesi Nöroloji Anabilim Dalı tarafından sağlanmıştır. Deneysel sonuçlar, önerilen yöntemin etkinliğini, özellik

seçim yönteminin genetik algoritma tarafından 1.00 doğruluk değeri üreten karar ağacı sınıflandırmasıyla göstermiştir.

Esmer ve ark. (2020), Hekimin teşhisini doğrulamak için, 252 katılımcının ses verilerini kullanarak PH teşhisi için bir çalışma yürütüldü. Veriler, sonuçların doğruluğunu artırmak için ön işleme tabi tutuldu. İşlenecek verileri seçmek için sistematik bir örnekleme yöntemi kullanıldı ve bir dengeleme süreci uygulandı. Özelliklerin etiket üzerindeki etkisini belirlemek için özellik seçim algoritması gerçekleştirilmiş ve performanslarına göre veri grupları oluşturulmuştur. Kullanılan üç sınıflandırma algoritması Karar Ağacı, K En Yakın Komşu Algoritması ve Destek Vektör Makineleridir. Bu algoritmaların performansı, Doğruluk, Hassasiyet, Spesifiklik, F-Measure, Kappa ve Auc gibi çeşitli kriterlere göre değerlendirildi. Elde edilen sonuçlara göre performans değeri en yüksek veri grubu belirlenmiş ve model oluşturmak için Destek Vektör Makineleri algoritması kullanılmıştır. Model, en ilgili olandan en az ilgili olana doğru sıralanan veri setinin %45'i kullanılarak geliştirilmiştir. Sonuçlar, performans kriterlerinde %85 doğruluk oranı ve diğer kriterlerde olumlu sonuçlar gösterdi. Bu nedenle, PB semptomları olan bireylerin ses kayıtlarından oluşturulan veri setinin hekimlere tıbbi karar desteği sağlamak için kullanılabileceği sonucuna varılmıştır.

Pereira ve ark. (2016), PH hastalarının motor özelliklerini analiz etmek için KSA'lerin kullanımını araştırmıştır. Başka bir deyişle, zaman serisi sinyallerinden elde edilen görüntülerde KSA'yi kullanmışlardır. Üç farklı KSA mimarisi kullanılmıştır. Bunlar; ImageNet, CIFAR-10 ve LeNet ve ayrıca temel yaklaşım olarak OPF şeklinde sıralanabilir.

Orijinal “HandPD” veri kümesini bir smartpen'den alınan sinyallerle genişletmişlerdir. Bu yeni veri setini, menderes ve spiral çizimlerden sinyaller çıkararak oluşturmuşlardır. Bu veri seti 35 birey, 14 hasta (10 erkek ve 4 kadın) ve 21 kontrol (11 erkek ve 10 kadın) içermektedir. Bireylerden bu etkinliklerden sinyalleri alabilmeleri için smartpen kullanarak bir form doldurmaları istenmiştir. Ardından hareket sinyallerini görüntüye dönüştürmüşlerdir.

Deney sonuçlarını Wilcoxon işaretli sıra testi kullanarak analiz ettiler ve çalışmalarını iki görüntü çözünürlüğü (64x64 ve 128x128 piksel) üzerinde değerlendirdiler. Farklı eğitim-test seti bölümleri, %75 eğitim ve %25 test, ardından %50 eğitim ve %50 test setleri kullandılar. Menderes veri setinde, imageNet yaklaşımı her iki çözünürlükte de tüm deneylerde en iyi sonuçları (sınıf başına genel doğruluk) elde etmişlerdir. %84 ile %87 arasında değişen, ardından genel doğrulukta %76 ile %84 arasında değişen OPF şeklinde sıralanabilir. En kötü doğruluk ve tanıma (sınıf başına doğruluk), bu deneyde daha sığ ağ olduğu için LeNet tarafından elde edilmiştir. Spiral veri setinde, en iyi doğruluklar %77 ile %83 arasında değişen OPF tarafından elde edilmiş, ardından %77 ile %80 arasında değişen ImageNet ve menderes veri seti gibi daha kötü sonuçlar LeNet'ten alınmıştır. CIFAR-10'un %50 eğitim-%50 test bölümünde ve 128x128 çözünürlükte PH hastalarını ayırt etmede %98.53 doğruluk elde ettiğini fark etmişlerdir.

Araştırmalarını tamamladıktan sonra, PH hastası ve kontrol arasındaki tanıma (sınıf başına doğruluk) sonuçlarının spiral veri setinde kıvrımlı veri setinden daha iyi olduğunu fark etmişlerdir, çünkü spirali çizmek düz bir çizgiden daha zordur. Tüm zaman serisi verilerinin KSA tarafından ayırt edilmesi için bazen 64x64 görüntü çözünürlüğünün yeterli olmadığını belirtmişlerdir.

Pereira ve ark. (2017), menderes ve spiral görüntüler kullanarak PH teşhisine yardımcı olmak için KSA'lerin kullanımını araştırmıştır. KSA uygulamasını aşağıdakiler de dahil olmak üzere çeşitli şekillerde karşılaştırdılar. Bunlar; temel olarak standart KSA (optimizasyon olmadan); Bat Algoritması BA ile KSA (optimizasyon ile); Ateşböceği Algoritması (FA) ile KSA (optimizasyon ile); Parçacık Sürü Optimizasyonu (PSO) ile KSA ve rastgele arama (RA) ile KSA şeklinde sıralanabilir.

Genişletilmiş “HandPD” veri setini kullanmışlardır. Bu veri setini spiral ve menderes grupları olmak üzere iki gruba ayırmışlardır. Görüntü çözünürlüğü 256x256 pikseldir. Her bir veri seti eğitim seti olarak %30, doğrulama seti olarak %20 ve test seti olarak %50 olarak ayrılmıştır.

Wilcoxon işaretli sıra testine dayanarak, menderes veri seti için elde ettikleri sonuçlar, diğer teknikler arasında genel doğruluk olarak BA, PSO ve standart KSA'nin sırasıyla en iyi sonuçları (%79.62, %75.76, %78.18) aldığını göstermiştir. BA, PH sınıfını sınıflandırmada en iyi doğruluğu elde etmiştir. Benzer şekilde, aynı önceki küme, spiral veri kümesi için sırasıyla (%87.20, %88.33, %89.55) en yüksek doğruluğu elde etmiştir. Önceki deney gibi, BA, PH sınıfını sınıflandırmada en iyi doğruluğu elde etmiştir.

Önceki çalışmada olduğu gibi, spiral testinin hastalar için menderes testinden daha zor olduğunu fark etmişlerdir. Çünkü bu test, yani kıvrımlı olan, sadece düz bir çizgiyi takip etmektir. Bu nedenle spiral üzerinde yapılan deneyler, hastalardan sağlıklı olanı menderes testinden daha anlamlı bir şekilde tanımasını sağlamıştır.

Parkinson hastalığı teşhisi konan denekleri sınıflandırmak amacıyla ses ve sınıflandırma algoritmalarının kullanma alanında çok sayıda çalışma yapılmıştır (Upadhyaya ve ark., 2018; Şakar ve ark., 2013; Vaiciukynas ve ark., 2017; Şakar ve ark., 2019). Bu araştırmalardaki algoritmalar genellikle yüksek bir doğruluğa ulaşmaktadır, genellikle %80'in üzerinde ve bazıları %90'ın üzerindedir. Bu etki alanında kullanılan en yaygın sınıflandırma algoritmalarından bazıları, Destek Vektör Makinesi (SVM), Çok Katmanlı Algılayıcı (ÇKA), k-En Yakın Komşu (EYKA) ve Rastgele Orman (RF) şeklinde sıralanabilir.

Bu alandaki araştırmaların çoğu, çalışmalarını halka açık iki veri kümesinden biri veya her ikisi üzerinde yapmaktadır. Bu veri kümeleri, Little ve ark. tarafından (Little ve ark., 2008) ve Şakar ve ark. tarafından (Şakar ve ark., 2013), sırasıyla 2008 ve 2013 yıllarında elde edilmiştir. Little'dan alınan veri seti, 23'ü PH'li ve 8 kontrol deneği olan 31 kişiden kaydedilmiştir. Şakar veri seti, 20'si PH'li ve 20'si kontrol deneği olmak üzere 40 kişiden kaydedilmiştir. Little veri seti için, her denek sesli harflerin sürekli fonasyonlarının 6 kaydını yapmıştır. SAKAR2019 veri setinde, her bir öznenin, üçü sesli harflerin sürekli fonasyonları olan 26 kayıt bulunmaktadır. "a" sesli harfinin sürekli fonasyonunun kullanılması, araştırmalarda Parkinson hastalığı sınıflandırması (Tsanas ve ark., 2012) ve diğer tıbbi uygulamalarda (Sidorova ve

Anisimova, 2020) uygun bilgiler içerdiğini göstermiştir. Bu veri setlerinin her ikisinde de veriler, kayıtlardan çıkarılan öznitelikler olarak verilmiştir.

Konularının küçük olması nedeniyle yukarıda bahsedilen iki veri setini kullanmanın sonuçlarını genelleştirmeye yönelik eleştirilerle Şakar ve ark., makalede (Şakar ve ark., 2019) yeni bir veri seti tanıtmışlardır. Bu veri seti, her kayıt için 188 Parkinson Hastası ve 64 kontrol denekten kaydedilmektedir. Bu veri setinden bu tez boyunca SAKAR2019 olarak bahsedilecektir. Orijinal makalede, bu veri seti sunulduğunda, yazarlar bir ÇKA dahil olmak üzere farklı sınıflandırma yöntemlerini karşılaştırmaktadırlar. Modellerini toplam 754 özniteliğin farklı alt kümeleri üzerinde uygulamaktadırlar ve toplam 754 öznitelikten yalnızca 50'sinin kullanıldığı öznitelik seçim yöntemlerini kullanmaktadırlar. Kullandıkları alt kümelerden ikisi yalnızca temel özellikleri ve yalnızca Mel-Frekans Kepstral Katsayıları (MFCK) özelliklerini kullanmaktadırlar.

Gabriel Solana-Lavalle ve ark., SAKAR2019 veri setini kullanmışlardır ve öznitelik seçme yöntemlerini yapmışlardır ve bir ÇKA dahil olmak üzere Parkinson sınıflandırması için farklı yöntemleri değerlendirmişlerdir. Literatürün gözden geçirilmesi, PH tanısında araştırmacıların karşılaştığı bazı ortak sorunlar ve zorluklar olduğunu göstermektedir.

Pereira ve ark. (2015), hem sağlıklı hem de erken evre hastaların elle çizilmiş çizimlerinin PH'nin semptomlarını keşfetmek çok zordur. Pereira ve ark. (2016) ve Pereira ve ark. (2017) gibi spiral olan çizimler daha basit olduğu için bu noktada özellikle menderes çizimi ile yapılan testlere genişletilmesi ve daha doğru sonuçlar alınması sağlanmaya çalışılması gerekmektedir. Veri kümeleri küçük, taraflı ve nadirdir. Bu, bir önceki nokta ile sadece elle çizilmiş görüntülerden teşhiste yer alan zorlukların nedenleridir. Pereira ve ark. (2016) araştırmasında yaptığı gibi PH tanısında bize yardımcı olacak motor özellikler ve Smartpen gibi daha fazla özellik çıkarılması faydalı olacaktır.

Evde kullanılan aletlerin pahalı olması ve her zaman bulunamaması nedeniyle hastayı evde izlemek zordur. Lones ve ark. (2017), ilaçla tedavi sürecinde uzmanlara yardımcı olmak için diskineziyi ölçmek için bir izleme cihazı geliştirmiştir.

Çizelge 2.1. Literatür taraması

Yazar	Yayın tarihi	Amaç	Veri seti ve kullanılan öz nitelikler	Başarı oranı ve sınıflandırma sonuçları
Little ve ark.	2009	Ses sinyalleri ile parkinson hastalığının teşhis edilmesi	Sesin dalgalanma, titreşim, ses seviyesi ve uyum özellikleri üzerinde bir ayırım yapmışlar 23 Parkinson hastası ve 8 sağlıklı bireyden oluşturdukları Oxford Parkinson's Disease Detection (OPD) veri seti kullanılmış.	91,4%
Tsanas ve ark	2010	Ses sinyalleri ile parkinson hastalığının teşhis edilmesi	Kaba küme yöntemi kullanılmıştır. Araştırmacılar çalışmalarında verileri farklı algoritmalarla filtreleyerek sınıflandırma yapmıştır. 23 Parkinson hastası ve 8 sağlıklı bireyden oluşturdukları Oxford Parkinson's Disease Detection (OPD) veri seti kullanılmış.	90%
Das	2010	Ses sinyalleri ile parkinson hastalığının teşhis edilmesi	Sinir ağları yöntemi kullanılmıştır. 23 Parkinson hastası ve 8 sağlıklı bireyden oluşturdukları Oxford Parkinson's Disease Detection (OPD) veri seti kullanılmış.	92,9%
Şakar ve ark.	2013	Parkinson Hastalığı Sınıflandırması	K-En Yakın Komşuluk ve Destek Vektör Makineleri yöntemlerini kullanarak sınıflandırma yapmışlardır. Bununla birlikte araştırmacılar hem Özetlenmiş Birini Dışarıda Bırakma (ö-BDB) ve Bir Bireyi Dışarıda Bırakma (BBDB) yöntemlerinin de sınıflandırma yöntemi olarak etkisini incelemiştir. 20 Parkinson hastası ve 20 sağlıklı bireyden oluşturdukları Parkinson Speech Dataset with Multiple Types of Sound Recordings (PSD) veri seti kullanılmış.	85%
Pereira ve ark.	2015	1-“HandPD” adlı halka açık bir veri seti oluşturmayı, 2-Görüntülerden bazı özellikleri çıkarmak için görüntü işleme için bir boru hattı oluşturmayı temel almıştır. Ana görev görüntü işleme ve özellik çıkarmadır.	Görüntü işleme görevi, görüntü deneğin izlemesi istenen şablonu içerdiğinden, spiral şablon ST'yi ve el yazısı izi HT'yi çıkarmak için birkaç aşamadan oluşan bir boru hattı içermektedir. Özellik çıkarımı, ST ve HT arasındaki farkları, aralarındaki dokuz istatistiksel değeri hesaplayarak değerlendirmeyi amaçlamıştır. Deneylerinde üç sınıflandırıcı model kullandılar: - Naive Bayes (NB), Optimum-Path Forest (OPF) ve SVM kullanılmıştır. HandPD adlı 37 Hasta ve 18 Sağlıklı bireyden oluşturdukları veri seti kullanılmışlardır. Bu veri seti, spiral el yazısı test görüntülerini içermektedir. Her muayenede birey başına yaklaşık dört spirale sahip oldukları için toplamda 373 örnek kullanılmıştır. Bu örnekleri %90 eğitim ve %10 test seti olarak ayırmışlardır	DOĞRULUK ORANLARI NB : % 78,9 OPF : % 77,1 SVM : % 75,8
Torun ve ark.	2015	Parkinson hastalığının ses sinyalleri üzerinden sınıflandırılması la hastalığın tespit edilmesi amaçlanmıştır	Sınıflandırma sonuçları filtreli ve filtresiz olarak değerlendirilmiş olup sınıflandırıcı olarak KNN (K-En Yakın Komşu) ve MLP (Çok Katmanlı Algılayıcı) fonksiyonu kullanılmış sonuçların geliştirilmesi için Adaboost ve Bagging algoritmaları kullanılmıştır. Bu amaç için Max Little ve ark. (2007) tarafından kaydedilen 147 Parkinson hastası ve 48 sağlıklı olan toplam 195 bireyin ses kayıtlarının olduğu veri seti kullanılmıştır.	Elde edilen sonuçlara bakıldığında filtre kullanılmadan KNN, topluluk algoritması olarak Adaboost'un kullanıldığı sınıflandırma işleminde başarı % 96,4103 olarak elde edilmiştir.

Çizelge 2.1. (Devam)

Yazar	Yayın tarihi	Amaç	Veri seti ve kullanılan öz nitelikler Sınıflandırma sonuçları	Başarı oranı ve sınıflandırma sonuçları
Esmer ve ark.	2020	PH olma ihtimali olan bireyin ses kayıtlarından oluşturulan veri seti ve uygulanan model yardımı ile doktora tıbbi karar destek sağlanması amaçlanmıştır.	Verilerde dengeleme işlemi yapılmış ve sistematik örnekleme metodu ile işleme alınacak veriler belirlenmiştir. Özellik seçme algoritması ile niteliklerin etiket üzerindeki etki gücü hesaplanıp bazı veri grupları oluşturulmuştur. Sınıflandırma algoritmalarından Karar ağacı, Destek Vektör Makineleri ve K En Yakın Komşu Algoritması kullanılıp, performans değerlendirme kriterleri - bunlar; Doğruluk Oranı, Duyarlılık, Özgünlük, F-Ölçümü, Kappa, Auc - değerlendirilmiştir. En yüksek performans değerine sahip veri grubu ve kullanılan sınıflandırma algoritması belirlenip model oluşturulmuştur.	85%
Bizal	2014	Makine öğrenmesi teknikleri kullanılarak Parkinson Hastalığı verileri üzerinde incelemeler yapılmış ve hastalığın belirlenmesi amaçlanmıştır	Sınıflandırma tekniği olarak Naive Bayes, K-En Yakın Komşuluk, Destek Vektör Makineleri, Lojistik Regresyon Analizi, Diskriminant Analiz yöntemleri kullanılmıştır. Bu çalışmada Little ve ark. tarafından oluşturulan (23 Parkinson hastası ve 8 sağlıklı birey) Oxford Parkinson's Disease Detection (OPD) veri seti kullanılmıştır	Bu veri seti üzerinde Knn diğer sınıflandırma tekniklerine göre en başarılı sonucu elde etmiştir.
Şakar ve ark.	2019	Parkinson Hastalığı Sınıflandırması	NaiveBayes, K-En Yakın Komşu (KNN), Rastgele Orman, DVM-RBF, DVM-Çok Katmanlı ve DVM-Doğrusal yöntemleriyle sınıflandırma yapmışlardır. Araştırmacılar bu çalışmalarında toplamda 6 ana öz nitelik setiyle toplam da 753 öz nitelik belirlemişlerdir. 188 Parkinson hastası ve 64 sağlıklı bireyden oluşturdukları Parkinson's Disease Classification (PDC) veri seti kullanılmıştır.	
			Temel öz niteliklerde	KNN
			MFCC öz niteliğinde	DVM (RBF)
			Dalga öz niteliklerinde	MLP
			Bant genişliği öz niteliğinde	KNN
			Yoğunluk öz niteliğinde	KNN ve Rastgele Orman Algoritması
			Vokal Kıvrım öz niteliğinde	Rastgele Orman Algoritması
Badem	2019	Parkinson Hastalığı Sınıflandırması	Karar Ağaçları, NaiveBayes, K-En Yakın Komşu, Rastgele Orman ve Destek Vektör Makineleri algoritmaları kullanılmıştır. 188 Parkinson hastası ve 64 sağlıklı bireyden oluşturdukları Parkinson's Disease Classification (PDC) veri seti kullanılmıştır.	KNN
Gerger	2019	Parkinson hastalığının tanısında çizilen spirallerin örüntü tanıma yöntemleriyle analiz edilmesi	Sınıflandırma sürecinde k-EK (k - en yakın komşu) ve karar ağaçları yöntemleri kullanılmıştır. Sonuçlar, birini dışarda bırak çapraz doğrulama yöntemiyle doğrulanmıştır. Bu çalışmada, İstanbul Üniversitesi Cerrahpaşa Tıp Fakültesi Nöroloji Anabilim Dalı tarafından hazırlanan veriler kullanılmıştır. Elde edilen deneysel sonuçlar önerilen yöntemin başarılı sonuçlar verdiğini göstermektedir.	Doğruluk değeri, genetik algoritma ile öz nitelik seçme yöntemi karar ağacı sınıflandırmasında 1.00 olarak elde edilmiştir.

Çizelge 2.1. (Devam)

Yazar	Yayın tarihi	Amaç	Veri seti ve kullanılan öznitelikler	Başarı oranı ve sınıflandırma sonuçları
Yıldırım	2020	Parkinson hastalığının ses sinyalleri üzerinden teşhisinde hibrit makine öğrenmesi temelli modellerle cinsiyete bağlı olarak hastalığın teşhisinde, Karar ağaçları ve Destek vektör makineleri kullanılarak hibrit makine öğrenmesi temelli yeni yaklaşımlar amaçlanmıştır.	Bu amaç için literatüre sunulan yüksek boyutlu öznitelik ve örnekleme sahip 188 Parkinson hastası ve 64 sağlıklı bireyden oluşturulan Parkinson's Disease Classification (PDC) veri seti kullanılmıştır.	Elde edilen başarı oranları oluşturulan modellerin hastalığın teşhisi için kullanılabileceğini göstermiştir. Ayrıca cinsiyet bazlı veri setlerinin kullanılması hastalığın teşhis edilmesi sürecinde performansı artırdığı gözlemlenmiştir
GÖKÇİN	2023	Parkinson hastalığının teşhisinde ses verileri üzerinden makine öğrenmesi ve derin öğrenme yöntemlerinin incelenmesi	Destek Vektör Makinelerinin (SVM) doğrusal, polinomsal, radial tabanlı çekirdek fonksiyonları, K-En Yakın Komşu (EYKA), Karar Ağaçları, Adaboostla Karar Ağaçları ve Rastgele Orman, algoritmaları kullanılmıştır. Bununla birlikte Konvolüsyonel Sinir Ağları, Çok Katmanlı Algılayıcılar, Derin Boltzmann Makinesi ayrı ayrı denenmiştir. 188 Parkinson hastası ve 64 sağlıklı bireyden oluşan Parkinson's Disease Classification (PDC) veri seti kullanılmıştır.	Doğruluk : % 89 (CNN) Kesinlik : % 92 (CNN) Duyarlılık : % 99 (MLP) F1 Skor : % 93 (CNN)

3. MATERYAL ve YÖNTEM

3.1.Materyal

Bu çalışmada SAKAR2019 veri seti kullanılmıştır. Bu veri setinde Cerrahpaşa Tıp Fakültesi Nöroloji Anabilim dalında tedavi görmekte olan yaşları 33-87 arasında değişen, 107 erkek ve 81 kadından oluşan toplam 188 Parkinson hastasından toplanan veriler kullanılmıştır. Çalışmada bu verilerle kontrol grubu olan 23 erkek ve 41 kadından oluşan toplam 64 sağlıklı bireyden toplanan veriler birlikte analiz edilmiştir. Çalışmada kullanılan veri ise 44.1 KHz'e ayarlanan bir mikrofon vasıtasıyla deneklerin üç defa "a" ünlüsünü söyledikleri ses kayıtlarıdır.

PH, orta beyindeki substantia nigra pars compacta'da nöromelanin içeren nöronların kaybı sonucu ortaya çıkan ve striatal dopamin azalmasına yol açan nörodejeneratif bir hastalıktır. Literatürde (Yunusova ve ark., 2008; Falk ve ark., 2012), PH'nin konuşmayı erken bir aşamada bile etkilediği gösterilmiştir ve bu nedenle konuşma özellikleri, PH'yi değerlendirmek, cerrahi veya farmakolojik tedaviden sonra gelişimini izlemek için başarıyla kullanılmıştır. Jitter, shimmer, temel frekans parametreleri, harmoniklik parametreleri, Tekrarlama Periyodu Yoğunluk Entropisi (RPDE), Eğilimsiz Dalgalanma Analizi (DFA) ve Perde Periyodu Entropisi (PPE), PD çalışmalarında kullanılan en popüler konuşma özellikleridir (Peker ve ark., 2015; Tsanas ve ark., 2012). Çalışmamızda bu özellikler "temel özellikler" olarak adlandırılmakta ve bu çalışmada incelenen diğer öznelik çıkarma yöntemlerinin performansını karşılaştırmak için kullanılmaktadır. RPDE, DFA ve PPE dışında, kullanılan temel özelliklerin ayrıntılı ifadeleri Çizelge 3.1.'de verilmiştir. Spektrogramlardan konuşma yoğunluğu, formant frekansları ve bant genişliğine dayalı özellikler de çıkarılır. Bu veri setindeki veriler bazı öznelik gruplarına bölünmüştür. Bu gruplar temel olarak Temel Özellikler (Baseline Features), Zaman Frekansı Özellikleri (Time Frequency Features), Mel Frekans Cepstral Katsayıları (MFCK), Dalgacık Özellikleri (Wavelet features) ve Vokal Kıvrım Özelliklerinden (VocalFold Features) oluşmaktadır. Oluşturulan veri gruplarına ilişkin bilgiler Çizelge 3.1.'de gösterilmiştir.

Çizelge 3.1. Veri gruplarına ilişkin bilgiler

Veri Özellik Grupları			Öznitelik Sayısı
Temel	Titreşim varyantları	Jitter varyantları, vokal kordların salınımlı modelinde meydana gelen kararsızlıkları yakalamak için kullanılır ve bu özellik alt kümesi, temel frekanstaki döngüden döngüye değişiklikleri nicelendirir.	5
	Parıltı varyantları	Parıltı varyantları da vokal kordların salınım modelinin kararsızlıklarını yakalamak için kullanılır, ancak bu sefer bu özellik alt kümesi, genlikteki döngüden döngüye değişiklikleri nicelleştirir.	6
	Temel frekans parametreleri	Vokal kord titreşim frekansı. Ortalama, ortanca, standart sapma, minimum ve maksimum değerler kullanılır.	5
	Uyum parametreleri	Ses tellerinin tam kapanmaması nedeniyle konuşma patolojilerinde artan gürültü bileşenleri oluşur. Özellik olarak, sinyal bilgisinin gürültüye oranını ölçen Harmonik Gürültü Oranı ve Gürültü Harmonik Oranı parametreleri kullanılır.	2
	Tekrarlama Periyodu Yoğunluk Entropisi (RPDE)	RPDE, vokal kordların stabil vokal kord salınımlarını sürdürme yeteneği hakkında bilgi verir ve F0'dan sapmaları nicelendirir.	1
	Eğilimsiz Dalgalanma Analizi (DFA)	DFA, türbülanslı gürültünün stokastik kendi kendine benzerliğini ölçer.	1
	Perde Periyodu Entropisi (PPE)	PPE, logaritmik ölçek kullanarak temel frekans F0'ın bozulmuş kontrolünü ölçer.	1
Mel Frekans Cepstral Coefficients MFCK	MFCK'ler	MFCK'ler, ses yolundaki PD etkilerini vokal kordlardan ayrı olarak yakalamak için kullanılır.	84
Zaman Frekansı	Yoğunluk Parametreleri	Yoğunluk, dB cinsinden konuşma sinyalinin gücü ile ilgilidir. Ortalama, minimum ve maksimum yoğunluk değerleri kullanılır.	3
	Formant Frekansları	Ses yolu tarafından yükseltile frekanslar, ilk dört formant özellik olarak kullanılır.	4
	Bant genişliği	Formant frekansları arasındaki frekans aralığı, ilk dört bant genişliği öznitelikler olarak kullanılmıştır.	4
Vokal Kıvrım	Glottis Bölümü (GQ)	GQ, glottisin açılma ve kapanma süreleri hakkında bilgi verir. Glottis hareketlerindeki periyodikliğin bir ölçüsüdür.	3
	Glottal - Gürültü Uyarma (GNE)	GNE, konuşma sinyalindeki eksik vokal kord kapanmasının neden olduğu türbülanslı gürültünün boyutunu ölçer.	6
	Vokal Kıvrım Uyarma Oranı (VFER)	VFER, doğrusal olmayan enerji ve entropi kavramlarını kullanarak patolojik ses teli titreşimi nedeniyle üretilen gürültü miktarını nicelleştirir.	7
	Ampirik Mod Ayırıştırma (EMD)	EMD, uyarlanabilir temel fonksiyonları kullanarak bir konuşma sinyalini temel sinyal bileşenlerine ayırıştırır ve bu bileşenlerden elde edilen enerji/entropi değerleri gürültüyü ölçmek için kullanılır.	6
Dalga	F0 ile ilgili dalgacık dönüşümü (WT) özellikleri	WT özellikleri, F0'daki sapmaları ölçer.	182

3.2.Yöntem

Bu çalışmada sınıflandırma yöntemi olarak Destek Vektör Makinelerinin (SVM) doğrusal, polinomsal, radial tabanlı çekirdek fonksiyonları, K-En Yakın Komşu (EYKA), Karar Ağaçları, Adaboostla Karar Ağaçları ve Rastgele Orman, algoritmaları kullanılmıştır. Bununla birlikte Konvolüsyonel Sinir Ağları, Çok Katmanlı Algılayıcılar, Derin Boltzmann Makinesi ayrı ayrı denenmiştir.

İlk olarak veri import edilmektedir. Verinin özellikleri Çizelge 3.1.'de yer almaktadır. Daha sonra ilk yapılan verinin içerisinde eksik veya anlamsız (missing or NAN) bir bilgi olup olmadığı kontrol edilmektedir. Çünkü eğer öyle bir şey olması durumunda, sıkıntılı satırları çıkarmak veya data amputation yapmak gerekmektedir. Fakat mevcut verilerde öyle bir durum gözlenmemiştir. Bu sebeple bir sonraki aşamaya geçilmektedir. Bir sonraki aşama verinin sütun tiplerini anlamlandırmaktır. Bir diğer ifadeyle sütundaki verilerin kategorik, binary veya sayı şeklinde linear olma durumu incelenmektedir. Ardından verinin test ve eğitim şeklinde ayrılmaktadır. Sonrasında sayısal olan sütunlar 0 ile 1 arasında normalize edilmektedir. Bir diğer ifadeyle her bir sayı 0 ile 1 arasında değere sahip hale gelmektedir. Normalizasyon yapılmasındaki amaç, bazı modellerde eğer sütunlar arasında çok fazla fark varsa anlamsız sonuçlar çıkabilmektedir. Bu durumu engellemek için normalizasyon yapılmaktadır. Literatürde bu durumun önüne geçmek için normalizasyon haricinde farklı yöntemlerde bulunmaktadır. Örneğin, mean=0, deviation=1 olacak şekilde normal dağılımına da fit ettirilebilir. Normalize edildiği sürece yöntemler arasında ciddi farklılıklar bulunmamaktadır. Bu işlem sonrasında sınıflandırma yapılacak olan modeller oluşturulmakta ve veriye fit ettirilerek sonuçları karşılaştırılmaktadır. Boltzmann Machines algoritması implementasyonları sürecinde GitHub'da yell isimli kullanıcının paylaştığı olduğu Boltzmann Machines in TensorFlow with examples temel alınıp, doğruluk ve duyarlılık üzerinde iyileştirmeler yapılmıştır.

Öznitelik çıkarma adımını takiben, elde edilen öznitelik vektörleri, her bir öznitelik normalize edilmektedir. Daha sonra, sağlıklı denekleri PH hastalarından ayırt etmek için özellik alt kümeleri çoklu sınıflandırıcılara beslenmektedir. Amaç,

disfoni tabanlı PH telediagnoz sistemlerinde her bir özellik çıkarma yönteminin performansını değerlendirmektir. Sınıflandırma modellerinin genelleme yeteneğini doğrulamak için bir konuyu dışarıda bırakma (LOSO) çapraz doğrulama tekniğini kullanılmaktadır. LOSO yöntemi, bir bireyin ses örneğinin sanki görünmeyen bir bireymiş gibi doğrulama için kullanılmaması ve diğer deneklerin ses örneklerinin eğitim için kullanılması esasına dayanmaktadır. Her denek üç kayıt örneği sağladığından, bir konuda nihai tahmini yapmak için bu örneklerin tahminleri üzerinde çoğunluk oylaması yapılmaktadır.

Standartlaştırılmış özellikler, Karar Ağaçları, K-En Yakın Komşu, Destek Vektör Makineleri (Doğrusal Kernel), Destek Vektör Makineleri (Polinomsal Kernel), Destek Vektör Makineleri (RBF Kernel), Adaboost Karar Ağaçları ve Rastgele Orman, Konvolüsyonel Sinir Ağları, Çok Katmanlı Algılayıcılar, Derin Boltzmann Makinesi yöntemlerini beslemektedir. Çeşitli özellik alt kümelerinin ve sınıflandırıcıların sonuçlarını karşılaştırmak için genel doğruluk, kesinlik ve duyarlılık metrikleri kullanılmaktadır.

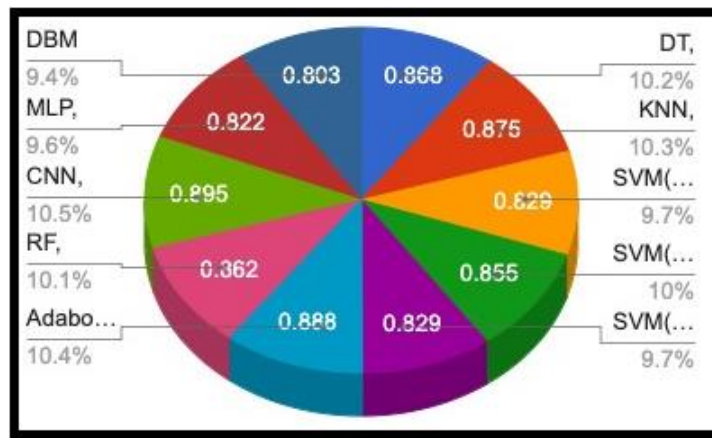
4. ARASTIRMA BULGULARI ve TARTISMA

Çalışmada kullanılan Karar Ağaçları, K-En Yakın Komşu, Destek Vektör Makineleri (Doğrusal Kernel), Destek Vektör Makineleri (Polinomsal Kernel), Destek Vektör Makineleri (RBF Kernel), Adaboost Karar Ağaçları, Rastgele Orman, Konvolüsyonel Sinir Ağları, Çok Katmanlı Algılayıcılar ve Derin Boltzmann Makinesi yöntemlerinden uygulama sonucunda elde edilen doğruluk, kesinlik ve duyarlılık metriklerine ilişkin bulguları Çizelge 4.1.'de gösterilmektedir.

Çizelge 4.1. Uygulanan algoritma ve sonuçları

	Doğruluk	Kesinlik	Duyarlılık
Karar Ağaçları	0.868	0.907	0.907
K-En Yakın Komşu	0.875	0.873	0.963
Destek Vektör Makineleri (Doğrusal Kernel)	0.829	0.858	0.907
Destek Vektör Makineleri (Polinomsal Kernel)	0.855	0.846	0.972
Destek Vektör Makineleri (RBF Kernel)	0.829	0.819	0.972
Adaboost Karar Ağaçları	0.888	0.902	0.944
Rastgele Orman	0.862	0.858	0.963
Konvolüsyonel Sinir Ağları (KSA)	0.895	0.925	0.941
Çok Katmanlı Algılayıcılar (ÇKA)	0.822	0.818	0.992
Derin Boltzmann Makinesi	0.803	0.806	0.983

Çizelge 4.1.'de verilen değerlerden doğruluk değerleri doğru bir sınıflandırma yapılmış olan veri sayısı ile toplam veri sayısının oranıdır. Dolayısıyla Parkinson ve sağlıklı bireylerin hangi oranda doğru tahmin edilebildiğini göstermektedir.



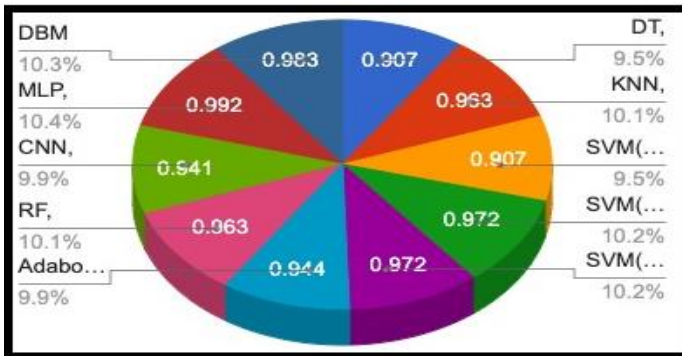
Şekil 4.1. Uygulanan Algoritmalar ve Doğruluk Değerleri

Çalışma sonucunda elde edilen doğruluk oranlarına bakıldığında bütün algoritmalarda %80 üzeri bir başarı elde edilmiş olduğu, en yüksek başarı oranının ise %89.5 oranla Konvolüsyonel Sinir Ağları (KSA) algoritmasıyla elde edilmiş olduğu görülmektedir. Kesinlik oranı ise Parkinson hasta örneklerinin oranlamasını göstermektedir. Bir diğer ifadeyle Parkinson hastası olarak tahmin edilmiş olan örneklerden ne kadarı gerçek Parkinson hastası olduğunu gösteren orandır.



Şekil 4.2. Uygulanan Algoritmalar ve Kesinlik Değerleri

Çalışma sonucunda tüm algoritmalarda yine %80 üzerinde bir başarı oranı olduğu, en yüksek oran %92.5 olarak Konvolüsyonel Sinir Ağları (KSA) ile elde edilmiştir. Duyarlılık oranı ise sağlıklı bireyler için elde edilen bir orandır. Toplam veri kümesi içerisinde sağlıklı olarak tahmin edilen hastaların oranını göstermektedir.



Şekil 4.3. Uygulanan Algoritmalar ve Duyarlılık Değerleri

Çalışmada tüm algoritmalarda %90 üzerinde bir oran elde edilmiş olup en yüksek %99.2 oranla Çok Katmanlı Algılayıcılar (ÇKA) algoritmasında elde edilmiştir.

4.1.Özellik Seçimi

Özellik Seçimi, veri seti içerisinde, eğitilme sonucu en doğru ve performanslı olan öznitelikleri bulma ve seçme, model için en doğru ve iyileştirici olmayan özniteliklerin elenmesi işlemidir. Bu işlem Makine Öğrenmesi modelinin sonuçlarını doğrudan ve önemli ölçülerde etkilemektedir. Model sonuçları neticesinde başarı göstermeyen öznitelikleri eleyip en doğru ve optimize olan verilerle en doğru sonuca ulaşmamızı sağlar. Özellik Seçimi işlemi, modelimizin eğitim süresini arttırabilmektedir.

Özniteliklerin elenmesi aşağıdaki sebeplerle önem taşıyabilir;

- Modelin öğrenme sürecini hızlandırmak adına gereksiz özniteliklerin kaldırılması önemli bir adımdır. Bu durum, algoritmanın performansını artırabilir.
- Komplekslik ve anlaşılabilirlik arasında bir denge sağlamak gereklidir. Fazlaca öznitelik, modelin anlaşılabilirliğini ve yorumlanabilirliğini zorlaştırabilir.
- Aşırı uyum (overfitting) riskini azaltmak için özniteliklerin seçimi kritiktir. Eğer model, eğitim verilerine aşırı uyum sağlarsa, test verilerinde düşük performans gösterebilir. Özellikle test verisetinde gelecek olan girdiler eğitim verisetindeki girdilere benzerlik göstermiyorsa, modelin hata oranı artacaktır.

Bu projede sütun sayısının azaltılarak makine öğrenme algoritmalarının uygulanması için öznitelik seçim (Özellik Seçimi) yöntemi uygulanmıştır.

Öznitelik seçim yöntemleri temel olarak bir makine öğrenmesi algoritması içinde açıklanmaya çalışılan değişken ile açıklayıcı olarak kullanılan değişkenler arasında

- Açıklanan değişken ile ilişkisi olmayan veya açıklanan değişkeni açıklama yeteneği düşük olan açıklayıcı değişkenlerin elenmesi.
- Oluşturulan modelin yorumlanmasını kolaylaştırmak amacıyla değişken sayısını azaltmak

- Makine öğrenmesi algoritmalarında eğitim aşaması için harcanan süreyi önemli ölçüde kısaltmak

Amacıyla kullanılmaktadır. Öznitelik seçim yöntemlerinin birçok gruplandırması olmakla birlikte temel olarak supervised (denetimli) ve unsupervised (denetimsiz) olarak ikiye ayrılmaktadır. Supervised yöntemlerde açıklanan değişken göz önünde tutulurken unsupervised yöntemlerde sadece açıklayıcı değişkenler arasındaki ilişki göz (korelasyon ve benzeri) önünde bulundurulur.

755 sütunu bulunan Parkinson Veri Setinde daha az değişken ile algoritmaların çalıştırılması için öznitelik seçim algoritmalarından K-en yüksek skor (SelectKBest) uygulanmıştır. K-en yüksek skor algoritması da denetimli algoritmalar arasında yer almaktadır. Bu yöntemde açıklayıcı değişkenlerin açıklanan değişken üzerindeki açıklama yeteneğini ölçmek için bir skor hesaplanır ve bu skoru en yüksek olan değişkenden başlanarak en düşük olana doğru k tanesi nihayi modele eklenir.

Bu çalışmada değişkenlerin açıklama yeteneğinin ölçülmesinde Python Scikit kütüphanesinde yer alan SelectKBest fonksiyonu kullanılmıştır. Bu yöntem içinde mutual_info_regression, f_classif, mutual_info_classif, chi2 ve f_regression metodları kullanılmıştır.

- mutual_info_regression: İki rastgele değişken arasındaki negatif olmayan bir değerdir ve değişkenler arasındaki bağımlılığı ölçer. Bu metot k en yakın komşuluk uzaklıkları ile elde edilen entropi tahminlerine dayalı olarak değişkenler arasındaki bağımlılığı hesaplamaktadır.
- f_classif: Yöntem, değişkenler arasındaki ANOVA analiz sonuçlarına dayanmaktadır. ANOVA analizinden çıkan F-değerleri üzerinden karşılaştırma yapmaya imkân sağlar.
- mutual_info_classif: mutual_info_regresyon yöntemine benzer olarak bir karşılaştırma yapılmaktadır. Mutual_info_regresyon yönteminden farklı olarak regresyon temeline değil de ikili karşılaştırma temeline dayalı olarak skorlar vermektedir.

- chi2: Bu yöntemde değişkenler arasındaki chi-kare istatistiği hesaplanmaktadır. Chi-kare istatistiği beklenen ve gözlenen değerler arasındaki farkın frekanslar üzerinden ağırlıklandırılarak ortalamasının alınmasına dayanmaktadır.
- f_regression: Bu yöntem tahmin edilen ve gerçek değerler arasındaki bağımlılığı hesaplamak için regresyon modeli kurmaktadır. Oluşturulan doğrusal regresyon modelinin F-istatistiği ve p-değerleri üzerinden karşılaştırma yapma imkânı sunmaktadır.

Öznitelik seçim algoritması bu çalışmada aşağıdaki şekilde uygulanmıştır:

- Seçilen makine öğrenmesi algoritmasını (A) bütün sütunlar modele dahil iken uygula.
- A algoritması sonucunda elde edilen doğruluk değerini *en iyi doğruluk* olarak al.
- $k = \text{MIN_FEATURE}$ al ve $k < \text{MAX_FEATURE}$ olduğu müddetçe aşağıdaki adımları tekrarla
- k parametresi için yukarıda belirtilen öznitelik algoritması ile bir değişkenler kümesi bul ve bunu D kümesine ata.
- D kümesindeki değişkenler ile A algoritmasını çalıştır.
- Algoritmanın sonundaki doğruluk değeri *en iyi doğruluktan* daha büyük ise *en iyi doğruluk* değerini algoritma sonu ile değiştir.
- $k = k + 1$ al

Özellik Seçimi algoritması uygulanırken seçilebilecek sütun sayılarına ilişkin alt ve üst sınır girdi olarak alınmıştır. Bu değerler arasındaki K-en yüksek skora sahip olan sütunlar ile yeniden Çizelge 4.2.'de yer alan algoritmalar çalıştırılmış ve doğruluk (accuracy) değerleri hesaplanmıştır. Bu doğruluk değerleri en yüksek olan sütunlar Özellik Seçimi algoritması ile seçildiği belirlenmiş ve bu sütunlar üzerinden Çizelge 4.2.'de ki algoritmaların sonuçları Çizelge 4.3.'de, sütun bilgileri ise Çizelge 4.4.'de verilmiştir.

Çizelge 4.2. Algoritmaların öznitelik seçimi uygulanmadan önce elde edilen sonuçları

	Doğruluk	Kesinlik	Duyarlılık
Decision Tree	0.77	0.843	0.851
KNN	0.855	0.859	0.965
SVM (Linear Kernel)	0.849	0.842	0.982
SVM (Polynomial Kernel)	0.849	0.858	0.956
SVM (RBF Kernel)	0.842	0.836	0.982
Adaboost with Decision Trees	0.868	0.873	0.965
Random Forest	0.849	0.847	0.974
Multilayer Preceptron Model (MLP)	0.822	0.818	0.982
Convolutional Neural Network (CNN)	0.875	0.936	0.895
Deep Boltman Machine (DBM)	0.75	0.75	1

Çizelge 4.3. Öznitelik Seçimi uygulandıktan sonra elde edilen mutual_info_regression sonuçları

mutual_info_regression	Doğruluk	Kesinlik	Duyarlılık
Decision Tree	0.868	0.905	0.921
KNN	0.862	0.866	0.965
SVM (Linear Kernel)	0.84	0.849	0.982
SVM (Polynomial Kernel)	0.855	0.848	0.982
SVM (RBF Kernel)	0.855	0.848	0.982
Adaboost with Decision Trees	0.868	0.905	0.921
Random Forest	0.888	0.888	0.974
Multilayer Preceptron Model (MLP)	0.868	0.879	0.956
Convolutional Neural Network (CNN)	0.849	0.847	0.974
Deep Boltman Machine (DBM)	0.75	0.75	1.00

Çizelge 4.4. Öznitelik Seçimi uygulandıktan sonra elde edilen f_classif sonuçları

f_classif	Doğruluk	Kesinlik	Duyarlılık
Decision Tree	0.789	0.831	0.904
KNN	0.816	0.826	0.956
SVM (Linear Kernel)	0.809	0.810	0.974
SVM (Polynomial Kernel)	0.822	0.822	0.974
SVM (RBF Kernel)	0.822	0.822	0.974
Adaboost with Decision Trees	0.822	0.827	0.965
Random Forest	0.816	0.826	0.956
Multilayer Preceptron Model (MLP)	0.822	0.832	0.956
Convolutional Neural Network (CNN)	0.816	0.826	0.956
Deep Boltman Machine (DBM)	0.783	0.800	0.947

Çizelge 4.5. Öznitelik Seçimi uygulandıktan sonra elde edilen mutual_info_classif sonuçları

mutual_info_classif	Doğruluk	Kesinlik	Duyarlılık
Decision Tree	0.862	0.897	0.921
KNN	0.862	0.878	0.947
SVM (Linear Kernel)	0.849	0.842	0.982
SVM (Polynomial Kernel)	0.855	0.848	0.982
SVM (RBF Kernel)	0.855	0.848	0.982
Adaboost with Decision Trees	0.868	0.905	0.921
Random Forest	0.895	0.895	0.974
Multilayer Preceptron Model (MLP)	0.875	0.874	0.974
Convolutional Neural Network (CNN)	0.862	0.860	0.974
Deep Boltman Machine (DBM)	0.75	0.75	1.00

Çizelge 4.6. Öznitelik Seçimi uygulandıktan sonra elde edilen chi2 sonuçları

chi2	Doğruluk	Kesinlik	Duyarlılık
Decision Tree	0.77	0.862	0.825
KNN	0.829	0.838	0.957
SVM (Linear Kernel)	0.803	0.804	0.974
SVM (Polynomial Kernel)	0.803	0.809	0.965
SVM (RBF Kernel)	0.809	0.815	0.965
Adaboost with Decision Trees	0.796	0.812	0.947
Random Forest	0.822	0.832	0.956
Multilayer Preceptron Model (MLP)	0.822	0.832	0.956
Convolutional Neural Network (CNN)	0.803	0.809	0.965
Deep Boltman Machine (DBM)	0.816	0.821	0.965

Çizelge 4.7. Öznitelik Seçimi uygulandıktan sonra elde edilen f_regression sonuçları

f_regression	Doğruluk	Kesinlik	Duyarlılık
Decision Tree	0.789	0.825	0.912
KNN	0.816	0.826	0.956
SVM (Linear Kernel)	0.809	0.810	0.974
SVM (Polynomial Kernel)	0.822	0.822	0.974
SVM (RBF Kernel)	0.822	0.822	0.974
Adaboost with Decision Trees	0.822	0.827	0.965
Random Forest	0.816	0.826	0.956
Multilayer Preceptron Model (MLP)	0.829	0.844	0.947
Convolutional Neural Network (CNN)	0.813	0.833	0.895
Deep Boltman Machine (DBM)	0.783	0.800	0.947

Çizelge 4.8. Öznitelik Seçimi methodlarının birleşimi ile elde edilen sonuçlar

	Doğruluk	Kesinlik	Duyarlılık
Decision Tree	0.796	0.849	0.886
KNN	0.842	0.846	0.965
SVM (Linear Kernel)	0.855	0.859	0.965
SVM (Polynomial Kernel)	0.842	0.846	0.965
SVM (RBF Kernel)	0.859	0.866	0.965
Adaboost with Decision Trees	0.868	0.880	0.965
Random Forest	0.862	0.855	0.982
Multilayer Preceptron Model (MLP)	0.875	0.875	0.974
Convolutional Neural Network (CNN)	0.869	0.871	0.974
Deep Boltman Machine (DBM)	0.789	0.811	0.939

Özellik Seçimi ile seçilen sütunlar üzerinden uygulanan algoritma sonuçlarına göre bütün algoritmaların Doğruluk değerlerinde iyileşme olduğu görülmüştür. Kesinlik ve Duyarlılık değerlerinde ise bazı algoritmalarda iyileşme olduğu bazılarında ise iyileşme olmadığı görülmüştür.

5. SONUÇLAR ve ÖNERİLER

Günümüzde, Parkinson hastalığının semptomlarından olan dil ve yazı yeteneğindeki aksaklıklar, makine öğrenmesi metotlarıyla analiz edilebilmektedir.

Bu çalışma boyunca “temel özellikler” olarak adlandırılan bu alanda en sık kullanılan öznitelikler kümesi de ayrı bir grup olarak dahil edilmektedir. Bu çalışma kapsamında 252 deneğin (188 Parkinson hastası ve 64 sağlıklı kontrol) ses kayıtlarını toplanan veri seti kullanılmaktadır. Ses kayıtlarından çeşitli özellik alt kümeleri çıkarılmakta ve her bir özellik alt kümesinin etkinliğini ve ayrıca bir dizi sınıflandırıcı kullanarak bunların kombinasyonunu değerlendirilmektedir.

Bu çalışmada konuşma bozukluklarının değerlendirilmesine yönelik kullanılan makine öğrenmesi algoritmalarının karşılaştırılmasının yapılması amaçlanmıştır. Ayrıca çalışmada daha önce yapılmış çalışmalardan daha fazla algoritma kullanılarak literatüre katkı sağlanması hedeflenmiş, bu kapsamda makine öğrenmesi algoritmalarından Karar Ağaçları, K-En Yakın Komşu, Destek Vektör Makineleri (Doğrusal Kernel), Destek Vektör Makineleri (PolinomsalKernel), Destek Vektör Makineleri (RBF Kernel), Adaboost Karar Ağaçları, Rastgele Orman, Konvolüsyonel Sinir Ağları, Çok Katmanlı Algılayıcılar (ÇKA) ve Derin Boltzmann Makinesi kullanılmaktadır.

Parkinson gibi ses ve görüntü özellikli belirtileri olan hastalıkların teşhisinde makine öğrenmesi tekniklerinin oldukça önemli olduğu ve yaygın olarak kullanıldığı bilinmektedir. Ancak yapılacak çalışmalar için esas teşkil edecek veri setlerinin ve örneklem sayılarının artırılmasının çalışma sayılarının artmasında fayda sağlayacağı değerlendirilmektedir. Bununla birlikte hem ses hem de el yazısı veri setlerinin birlikte ele alınarak incelemeler yapılmasının karşılaştırmalı analizlere imkan tanınması açısından fayda sağlayacağı değerlendirilmektedir. Parkinson hastalığının konuşma bozukluğu üzerinden makine öğrenmesi yöntemleriyle incelendiği çalışmalara bakıldığında;

Little ve ark. (2009), yaptıkları çalışmada 23 Parkinson hastası ve 8 sağlıklı bireyden oluşturdukları OPH örneklemelerinde sesin dalgalanma, titreşim, ses seviyesi ve uyum özellikleri üzerinde bir ayırım yapmışlar ve çalışmalarında %91,4 oranında bir başarı elde etmişlerdir. Tsanas ve ark. (2010) çalışmalarında OPH veri setini kullanmış ve metot olarak kaba küme yöntemini tercih etmişlerdir. Araştırmacılar çalışmalarında verileri farklı algoritmalarla filtreleyerek sınıflandırma yapmış, çalışma sonucunda %90 oranında bir başarı oranı elde etmişlerdir. Aynı şekilde OPH veri setini kullanan Das (2010) da çalışmasında sinir ağları yöntemini kullanarak %92,9 başarı oranı yakalamıştır. Bizim çalışmamızda bu çalışmadaki başarı oranından çok daha yüksek olan %99.2 oranla Çok Katmanlı Algılayıcılar (ÇKA) algoritmasında elde edilmektedir.

Şakar ve ark. (2013), kendi oluşturdukları Cerrahpaşa Tıp Fakültesinde Nöroloji departmanında tedavi gören 20 Parkinson hastası ve 20 sağlıklı bireyden oluşan toplam 40 denekli veri setleriyle yapmış oldukları çalışmada K-En Yakın Komşuluk ve Destek Vektör Makineleri yöntemlerini kullanarak sınıflandırma yapmışlardır. Bununla birlikte araştırmacılar hem Özetlenmiş Birini Dışarıda Bırakma (ö-BDB) ve Bir Bireyi Dışarıda Bırakma (BBDB) yöntemlerinin de sınıflandırma yöntemi olarak etkisini incelemiş, çalışma sonucunda ö-BDB yönteminde %85 oranında başarı elde etmişlerdir. Bizim çalışmamızda en yüksek başarı oranının ise %89.5 oranla Konvolüsyonel Sinir Ağları (KSA) algoritmasıyla elde edilmektedir.

Şakar ve ark. (2019), yılında veri setlerini geliştirerek bir çalışma daha yapmışlar, bu çalışmalarında toplam 252 denekten oluşan 756 veriyle NaiveBayes, K-En Yakın Komşu (EYKA), Rastgele Orman, DVM-RBF, DVM-Çok Katmanlı ve DVM-Doğrusal yöntemleriyle sınıflandırma yapmışlardır. Araştırmacılar bu çalışmalarında toplamda 6 ana öznitelik setiyle toplam da 753 öznitelik belirlemişlerdir. Yazarlar, temel özniteliklerde K- En Yakın Komşu algoritmasında, MFCK özniteliğinde Destek Vektör Makinelerinde (RBF Kernel), dalga özniteliklerinde Çok Katmanlı Algılayıcı algoritmasında, bant genişliği özniteliğinde K- En Yakın Komşu algoritmasında, yoğunluk özniteliğinde K- En Yakın Komşu ve

Rastgele Orman Algoritmalarında ve son olarak Vokal Kıvrım özneliğinde Rastgele Orman algoritmasında en yüksek doğruluk oranına ulaşmışlardır. Bizim çalışmamızda diğer çalışmalarda yer alan algoritmaların verdikleri sonuçlara göre KSA algoritması en yüksek kesinlik oranını (%92.5) vermektedir.

Badem (2019)'in bu çalışmada kullanılan aynı veri setiyle Karar Ağaçları, NaiveBayes, K-En Yakın Komşu, Rastgele Orman ve Destek Vektör Makineleri algoritmalarını kullanarak yapmış olduğu çalışmasında en iyi performansı K- En Yakın Komşu algoritmasından elde etmiştir. Bizim çalışmamızda her ne kadar K-En Yakın Komşu algoritmasından da iyi sonuçlar elde edilmiş olsa da Rastgele Orman algoritmasının diğerlerine nazaran daha iyi bir performansa sahip olduğu görülmektedir.

Literatürde PD sınıflandırma problemi için algoritma sonuçlarına göre bu çalışmada daha yüksek doğruluk oranlarının rapor edildiği görülmektedir. Çalışma sonucunda en iyi sonuçların Konvolüsyonel Sinir Ağları (KSA) algoritmasından elde edildiği görülmektedir. Bununla birlikte, bu çalışmalarda kullanılan konuşma veri kümeleri denek başına birden fazla konuşma kaydı içermesine rağmen, bu çalışmaların çoğu, eğitimdeki bir bireyin bazı örneklerini ve bazılarını eğitim için ayırarak önyargılı tahmin modelleriyle sonuçlanan bir dışarıda bırakma çapraz doğrulama tekniğini kullanmaktadır. Bu süreç, eğitim ve test setleri arasında yapay bir örtüşme oluşturmaktadır.

Gelecekteki çalışmalarda PH sınıflandırma probleminde umut verici sonuçlar gösteren KSA tekniği, PH tele izleme sistemi oluşturmak için PH hastalarının Birleşik Parkinson Hastalığı Derecelendirme Ölçeği puanını tahmin etmek için kullanılabilir.

KAYNAKLAR

- AYHAN, S. ve ERDOĞMUŞ, Ş., 2014. Destek vektör makineleriyle sınıflandırma problemlerinin çözümü için çekirdek fonksiyonu seçimi. Eskişehir Osmangazi Üniversitesi İktisadi ve İdari Bilimler Dergisi, 9(1), 175-201.
- BADEM, H., 2019. Parkinson Hastalığının Ses Sinyalleri Üzerinden Makine Öğrenmesi Teknikleri ile Tanımlanması. Niğde Ömer Halisdemir Üniversitesi Mühendislik Bilimleri Dergisi, 8(2), 630-637.
- BAHL, N. 2017. Optimizing Performance Measures in Classification Using Ensemble Learning Methods, Master of Science, Arizona State University, USA.
- BALAKRİSHNAN, H., NELLİPALLİL, A., KATHPALIA, S.S., and NİTHİN, N., 2019. "ChaosNet: A chaos based artificial neural network architecture for classification." Review of. Chaos: An Interdisciplinary Journal of Nonlinear Science 29 (11):113125.
- BAŞER, B., YANGIN, M. ve SARIDAŞ, E.S., 2021. Makine Öğrenmesi Teknikleriyle Diyabet Hastalığının Sınıflandırılması. Süleyman Demirel Üniversitesi Fen Bilimleri Enstitüsü Dergisi, 25(1), 112-120.
- BİZAL, Ö., 2014. Makine Öğrenmesi Teknikleriyle Parkinson Hastalığının Belirlenmesi.
- BRE, F., JUAN, M., GİMENEZ, and VÍCTOR, D.F., 2018. "Prediction of wind pressure coefficients on building surfaces using artificial neural networks." Review of. Energy and Buildings 158:1429-41.
- CHENG, K., ZHANG, C., YU, H., YANG, X., ZOU, H. and GAO, S., 2019. Grouped smote with noise filtering mechanism for classifying imbalanced data, IEEE Access, vol. 7, pp. 170 668–170 681.
- CLEOPHAS T. J. and ZWINDERMAN, A.H.EDS., 2013. Machine Learning in Medicine. Springer Netherlands.
- CLEVELAND C. and "PARKINSON'S DİSEASE," J., 2020. <https://my.clevelandclinic.org/health/diseases/8525-parkinsons-disease-an-overview>
- ÇAKMUR, R., 2011. Parkinson hastalığı ve medikal tedavisi. Klinik Gelişim, 53, 58.
- ÇEVİK, F. ve KİLİMCİ, Z.H., 2020. Derin Öğrenme Yöntemleri ve Kelime Yerleştirme Modelleri Kullanılarak Parkinson Hastalığının Duygu Analiziyle Değerlendirilmesi. Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi, 27(2), 151-161.
- DAS, R., 2010. A comparison of multiple classification methods for diagnosis of Parkinson disease. Expert Systems with Applications, 37(2), 1568-1572.
- DOĞAN, F. ve TÜRKOĞLU, İ., 2019. Derin Öğrenme Modelleri ve Uygulama Alanlarına İlişkin Bir Derleme. Dicle Üniversitesi Mühendislik Fakültesi Mühendislik Dergisi, 10(2), 409-445.
- ESMER, S., UÇAR, M. K. , ÇİL, İ. ve BOZKURT, M.R., 2020. Parkinson Hastalığı Teşhisi İçin Makine Öğrenmesi Tabanlı Yeni Bir Yöntem. Düzce Üniversitesi Bilim ve Teknoloji Dergisi , 8 (3) , 1877-1893.

- FAGHFOURİ, A.E. and FRİSH, M.B., "Robust discrimination of human footsteps using seismic signals," in *Unattended Ground, Sea, and Air Sensor Technologies and Applications XIII*, vol. 8046. International Society for Optics and Photonics, 2011, p. 80460D.
- FAHN, S., 2003. Description of Parkinson's Disease as a Clinical Syndrome. *Annals of the New York Academy of Sciences*, 991(1), 1-14.
- FALK, T., CHAN, W. and SHEİN, F., 2012. Characterization of atypical vocal source excitation, temporal dynamics and prosody for objective measurement of dysarthric word intelligibility, *Speech Commun.* 54 (5) 622–631.
- GERGER, 2019. Makine Öğrenme Yöntemlerini Kullanarak Spiral Test İle Parkinson Hastalığının Tanı Tahmini, Yüksek Lisans Tezi, Harran Üniversitesi, Şanlıurfa.
- GOLZE, J., ZOURLİDOU, S. and SESTER, M., 2020. Traffic regulator detection using gps trajectories, *KN-Journal of Cartography and Geographic Information*, vol. 70, no. 3, pp. 95–105.
- GUPTA, D., SUNDARAM, S., KHANNA, A., HASSANIEN, A.E. and DE ALBUQUERQUE, V. H. C., 2018. Improved diagnosis of parkinson's disease using optimized crow search algorithm, *Computers&Electrical Engineering*, vol. 68, pp. 412–424.
- GÜNDÜZ, H., 2019. Deep learning-based parkinson's disease classification using vocal feature sets, *IEEE Access*, vol. 7, pp. 115 540–115 551.
- JANKOVIĆ, J., 2008. Parkinson's disease: clinical features and diagnosis. *Journal of neurology, neurosurgery & psychiatry*, 79(4), 368-376.
- KAMARULZAINİ, K. A.A., ISMAİL, N., RAHİMAN, M.H. F., TAİB, M.N., ALİ, N. A. M. and TAJUDDİN, S.N., 2018. Evaluation of rbf and mlp in svm kernel tuned parameters for agarwood oil quality classification, in *2018 IEEE 14th International Colloquium on Signal Processing & Its Applications (CSPA)*. IEEE, pp. 250–254.
- KUMAR, V. and MİNZ, S., 2014. Özellik Seçimi: A literature review, *Smart Computing Review*, vol. 4.
- LECUN, Y., CHOPRA, S. and HADSELL, R., 2006. A tutorial on energy-based learning. *Predicting Structured Data*, MIT Press
- LEVER, J., KRZYWIŃSKI, M. and ALTMAN, N., 2016. Points of significance: model selection and overfitting.
- LİTTLE, M.A., MC SHARRY, P.E., HUNTER, E.J., SPIELMAN, J., and RAMİG, L.O., 2009. Suitability of Dysphonia Measurements for Telemonitoring of Parkinson's Disease. *IEEE Transactions on Biomedical Engineering*, 56(4), 1015-1022.
- LİU, L. and ÖZSU, M.T., 2009. eds., *Encyclopedia of Database Systems*. Springer US.
- LİU, JİNJİN,, YONGCHUN, CHEN,, Lİ LAN, BOLİ, LİN., WEIJIAN, CHEN, MEİHAO WANG, RUİ Lİ, YUNJUN YANG, BİNG ZHAO, and ZİLİONG HU., 2018. "Prediction of rupture risk in anterior communicating artery aneurysms with a feed-forward artificial neural network." *Review of. European radiology* 28 (8):3268-75.
- LİU, R. and GİLLİES, D.F., 2016. Overfitting in linear feature extraction for classification of high-dimensional image data, *Pattern Recognition*, vol. 53, pp. 73–86.

- MA, H.I., HWANG, W.J., CHANG, S.H. and WANG, T.Y., 2013. Progressive micrographia shown in horizontal, but not vertical, writing in parkinson's disease, *Behavioural neurology*, vol. 27, no. 2, pp. 169–174.
- MEMEDİ, M., SADIKOV, A., GROZNIK, V., ŽABKAR, J., MOŽINA, M., BERGQUIST, F., JOHANSSON, A., HAUBENBERGER, D. and NYHOLM, D., 2015. Automatic spiral analysis for objective assessment of motor symptoms in parkinson's disease, *Sensors*, vol. 15, no. 9, 23 727–23 744.
- MIRELMAN, A., BONATO, P., CAMICIOLİ, R., ELLIS, T.D., GILADI, N., HAMILTON, J.L., HASS, C.J., HAUSDORFF, J.M., PELOSIN, E. and ALMEIDA, Q. J., 2019. Gait impairments in parkinson's disease, *The Lancet Neurology*, vol. 18, no. 7, pp. 697–708.
- MYLES, A.J., FEUDALE, R. N., LIU, Y., WOODY, N.A. and BROWN, S.D., 2004. An introduction to decision tree modeling, *Journal of Chemometrics: A Journal of the Chemometrics Society*, vol. 18, no. 6, pp. 275–285.
- NARANJO, L., PEREZ, C.J., MARTÍN, J. and CAMPOS-ROCA, Y., 2017. A two-stage variable selection and classification approach for parkinson's disease detection by using voice recording replications, *Computer Methods and Programs in Biomedicine*, vol. 142, pp. 147–156.
- NİE, F., ZHU, W., and Lİ, X., 2020. Decision tree svm: An extension of linear svm for non-linear classification, *Neurocomputing*, vol. 401, pp. 153– 159.
- NİZAM, H. and AKIN, S.S., 2014. Sosyal Medyada Makine Öğrenmesi İle Duygu Analizinde Dengeli ve Dengesiz Veri Setlerinin Performanslarının Karşılaştırılması. XIX. Türkiye'de İnternet Konferansı, 1(6).
- ORHAN, U., HEKİM, M. ve ÖZER, M., 2010. EEG İşaretlerinin Çok-Katmanlı Algılayıcı Yapay Sinir Ağı Modeli ile Sınıflandırılmasında Ayırıklaştırma Yaklaşımı. In *BiomedicalEngineering Meeting (BIYOMUT)* 1, 4.
- ÖZEKMEKÇİ, S., APAYDIN, H., OĞUZ, S. ve ZİLELİ, İ., 2013. Parkinson Hastalığı Hasta ve Yakınları İçin El Kitabı. İstanbul, Turkey: Bayçınar Tıbbi Yayıncılık ve Reklam Hiz. Tic. Ltd. Şti, 98.
- PALAKURTHİ, B. and BURUGUPALLY, S.P., 2019. Postural instability in parkinson's disease: A review, *Brain sciences*, 9(9), 239.
- PEDRO, H.T., INMAN, R.H. and COIMBRA, C.F., 2017. 4 - mathematical methods for optimized solar forecasting,” in *Renewable Energy Forecasting*, ser. Woodhead Publishing Series in Energy, G. Kariniotakis, Ed. Woodhead Publishing, 111–152. ISBN 978-0-08-100504-0
- PEKER, M., ŞEN, B. And DELEN, D., 2015. Computer-aided diagnosis of Parkinson's disease using complex-valued neural networks and mRMR Özellik Seçimi algorithm, *J. Healthcare Eng.* 6 (3), 281–302.
- POON, C., GORJİ, N. LATT, M., TSOİ, K. CHOİ, B. LOY, C. and POON, S., 2019. Derivation and analysis of dynamic handwriting features as clinical markers of parkinson's disease, in *Proceedings of the 52nd Hawaii International Conference on System Sciences*, 3721–3730.
- QUINLAN, J.R., 1986. Induction of decision trees, *Machine Learning*, 1(1), 81–106.
- RAIANO, L., Dİ PİNO, G., Dİ BİASE, L., TOMBİNİ, M., TAGLIAMONTE, N. L. and FORMICA, D., 2020. Pdmeter: A wrist wearable device for an at-home assessment of the parkinson's disease rigidity, *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 28(6):1325– 1333.

- SADİQ, R., 2018. Reference Module in Earth Systems and Environmental Sciences, Empirical Models to Predict Disinfection By-Products in Drinking Water: An Updated Review ☆. , (), –.
- SCHLENSTEDT, C., BOBE, K., GAVRİLİUC, O., WOLKE, R., GRANERT, O., DEUSCHL, G. and MARGRAF, N.G., 2020. Quantitative assessment of posture in healthy controls and patients with parkinson's disease, *Parkinsonism & related disorders*, 76: 85–90.
- ŞAKAR C., SERBES, G. , GÜNDÜZ, A. , TUNÇ, H. , NİZAM, H., ŞAKAR, B., TÜTÜNCÜ M., AYDIN T., İSENKUL M. and APAYDIN, H., 2019. A Comparative Analysis Of Speech Signal Processing Algorithms For Parkinson's Disease Classification And The Use Of The Tunable Q-Factor Wavelet Transform, *Applied Soft Computing*, 74, 255-263.
- ŞAKAR, B.E. İSENKUL, M. E. ŞAKAR, C. O. SERTBAS, A. GURGEN, F. DELİL, S. APAYDİN, H. and KURSUN, O., 2013. Collection and analysis of a parkinson speech dataset with multiple types of sound recordings, *IEEE Journal of Biomedical and Health Informatics*, vol. 17(4): 828–834.
- ŞAKAR, B. E.,İSENKUL, M.E., ŞAKAR, C. O., SERTBAS, A., GURGEN, F., DELİL, S. and KURSUN, O., 2013. Collection and analysis of a Parkinson speech dataset with multiple types of sound recordings. *IEEE Journal of Biomedical and Health Informatics*, 17(4), 828-834.
- ŞAKAR, C.O., SERBES, G., GUNDUZ, A., TUNC, H.C., NİZAM, H., ŞAKAR, B.E., TUTUNCU, M., AYDİN, T., İSENKUL, M.E. and APAYDİN, H., 2018. A comparative analysis of speech signal processing algorithms for Parkinsonâ's disease classification and the use of the tunable Q-factor wavelet transform. *Applied Soft Computing*.
- ŞAKAR, C.O., SERBES, G., GUNDUZ, A., TUNC, H.C., NİZAM, H., ŞAKAR, B.E., TÜTÜNCÜ, M., AYDIN, T., İSENKUL, M.E. and APAYDİN, H., 2018. A Comparative Analysis of Speech Signal Processing Algorithms for Parkinsonâ, Disease Classification and The Use of The Tunable Q-Factor Wavelet Transform. *Applied Soft Computing*.
- TEKELİ, K. ve AŞLIYAN, R., 2016. Çok Katmanlı Algılayıcı, K-NN ve C4.5 Metotlarıyla İstenmeyen E-postaların Tespiti. Adnan Menderes Üniversitesi.
- TOLOSA, E., WENNİNG, G. and POEWE, W., 2006. The diagnosis of parkinson's disease, *The Lancet Neurology*, 5: 75–86.
- TORUN, Ş., 1991. Parkinson'lularda konuşma fonksiyonunun subjektif ve objektif (elektrolaringografik) yöntemlerle incelenmesi, *Uzmanlık Tezi*, Anadolu Üniversitesi Tıp Fakültesi, Nöroloji Anabilim Dalı, Eskişehir.
- TSANAS, A., LITTLE, M., MCSHARRY, P., SPIELMAN, J. and RAMİG, L., 2012. Novel speech signal processing algorithms for high-accuracy classification of Parkinson's disease, *IEEE Trans. Biomed. Eng.* 59 (5): 1264–1271.
- TSANAS, A., LITTLE, M.A., MCSHARRY, P.E. and RAMİG, L.O., 2010. Accurate telemonitoring of Parkinson's disease progression by noninvasive speech tests. *IEEE transactions on biomedicalengineering*, 57(4), 884-893.
- TYSNES, O.B. and STORSTEİN, A., 2017. Epidemiology of parkinson's disease, *Journal of Neural Transmission*, 124(8): 901–905.
- TYSNES, O. B. and STORSTEİN, A., 2017. Epidemiology of Parkinson's Disease. *Journal of Neural Transmission*, 124(8), 901-905.

- VAN, G.M. and SANDER B., 2017. "Artificial neural networks as models of neural information processing." Review of. *Frontiers in Computational Neuroscience* 11:114.
- VAROQUAUX, G., RAAMANA, P.R., ENGEMANN, D.A., HOYOS-IDROBO, A., SCHWARTZ, Y., and THIRION, B., 2017. Assessing and tuning brain decoders: Cross-validation, caveats, and guidelines, *NeuroImage*, vol. 145, pp. 166–179.
- VENKATESH, B. and ANURADHA, J., 2019. A review of Özellik Seçimi and its methods, *Cybernetics and Information Technologies*, 19 (1): 3–26.
- VERANYURT, Ü., DEVECİ, A., ESEN, M. F. ve VERANYURT, O., 2020. Makine Öğrenmesi Teknikleriyle Hastalık Sınıflandırması: Random Forest, K-Nearest Neighbour ve Adaboost Algoritmaları Uygulaması. *Uluslararası Sağlık Yönetimi ve Stratejileri Araştırma Dergisi*, 6(2), 275-286.
- WU, YU-CHEN, and JUN-WEN FENG., 2018. "Development and application of artificial neural network." Review of. *Wireless Personal Communications* 102 (2):1645-56.
- YASEN, J. and DU, P., 2016. Performance measures in evaluating machine learning based bioinformatics predictors for classifications. *Quantitative Biology*. 4.
- YUNUSOVA, Y. WEISMER, G. WESTBURY, J. and LINDSTROM, M. 2008. Articulatory movements during vowels in speakers with dysarthria and healthy controls, *J. Speech Lang. Hear. Res.* 51 (3): 596.