

T.C. KIRIKKALE ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI
DOKTORA TEZİ

KARIN AĞRISI İLE ACİL SERVİSE BAŞVURAN HASTALARDA AKUT
APANDİSİT TANISI İÇİN MAKİNE ÖĞRENMESİ YAKLAŞIMLARININ
KULLANIMI

Ömer Faruk AKMEŞE

ŞUBAT 2020

ONAY SAYFASI

Bilgisayar Mühendisliği Anabilim Dalında Ömer Faruk AKMEŞE tarafından hazırlanan KARIN AĞRISI İLE ACİL SERVİSE BAŞVURAN HASTALARDA AKUT APANDİSİT TANISI İÇİN MAKİNE ÖĞRENMESİ YAKLAŞIMLARININ KULLANIMI adlı Doktora Tezinin Anabilim Dalı standartlarına uygun olduğunu onaylarım.

Doç. Dr. Atilla ERGÜZEN
Anabilim Dalı Başkanı

Bu tezi okuduğumu ve tezin **Doktora Tezi** olarak bütün gereklilikleri yerine getirdiğini onaylarım.

Prof. Dr. Hasan ERBAY
Danışman

Jüri Üyeleri

Başkan	: Doç. Dr. Cenker BİÇER	_____
Üye (Danışman)	: Prof. Dr. Hasan ERBAY	_____
Üye	: Dr. Öğr. Üyesi Melih ENGİN	_____
Üye	: Dr. Öğr. Üyesi Bülent Gürsel EMİROĞLU	_____
Üye	: Dr. Öğr. Üyesi Hakan KÖR	_____

21/02/2020

Bu tez ile Kırıkkale Üniversitesi Fen Bilimleri Enstitüsü Yönetim Kurulu Doktora derecesini onaylamıştır.

Prof. Dr. Recep ÇALIN
Fen Bilimleri Enstitüsü Müdürü

ÖZET

KARIN AĞRISI İLE ACİL SERVİSE BAŞVURAN HASTALARDA AKUT APANDİSİT TANISI İÇİN MAKİNE ÖĞRENMESİ YAKLAŞIMLARININ KULLANIMI

AKMEŞE, Ömer Faruk

Kırıkkale Üniversitesi

Fen Bilimleri Enstitüsü

Bilgisayar Mühendisliği Anabilim Dalı, Doktora Tezi

Danışman: Prof. Dr. Hasan ERBAY

Şubat 2020, 82 sayfa

Genel cerrahi kliniklerinde en sık görülen acil hastalıklardan biri akut apandisitir. Özellikle 10 ila 30 yaşları arasında daha yaygındır. Ayrıca, tüm popülasyonun yaklaşık %7'sine yaşamlarının bir döneminde akut apandisit teşhisi konur ve ameliyat gerekir. Çalışma, makine öğrenmesi algoritmaları kullanarak erken akut apandisit tanısı için kolay, hızlı ve doğru bir tahmin yöntemi geliştirmeyi amaçlamaktadır. Retrospektif klinik kayıtlar, tahmin edici veri madenciliği modelleri ile analiz edilmiştir. Çeşitli makine öğrenme algoritmalarıyla elde edilen modellerin öngörü başarıları karşılaştırılmıştır. Çalışmada 348 erkek (%58,49) ve 247 kadın (%41,51) dahil olmak üzere toplam 595 klinik kayıt kullanılmıştır. Gradyan artırılmış ağaçlar algoritmasının %95,31'lik doğru tahmin başarıları ile en iyi başarıya ulaştığı tespit edilmiştir. Bu çalışmada, akut apandisitli bireyleri tanımlamak için makine öğrenmesine dayalı bir tahmin yöntemi geliştirilmiştir. Bu yöntemin özellikle hastanelerde, acil servislerde apandisit belirtileri olan hastalara fayda sağlayacağı düşünülmektedir.

Anahtar kelimeler: Akut Apandisit, Makine Öğrenmesi, Ameliyat Tahmini

ABSTRACT

THE USE OF MACHINE LEARNING APPROACHES FOR THE DIAGNOSIS OF ACUTE APPENDICITIS IN PATIENTS PRESENTING TO THE EMERGENCY DEPARTMENT WITH ABDOMINAL PAIN

AKMEŞE, Ömer Faruk

Kırıkkale University

Graduate School of Natural and Applied Sciences

Department of Computer Engineering, M.Sc. Thesis

Supervisor: Prof. Dr. Hasan ERBAY

February 2020, 82 pages

Acute appendicitis is one of the most common emergency diseases in general surgical clinics. It is more common, especially between the ages of 10 and 30 years. Also, approximately 7% of the entire population is diagnosed with acute appendicitis at some time in a period of their lives and requires surgery. The study aims to develop an easy, fast, and accurate estimation method for early acute appendicitis diagnosis using machine-learning algorithms. Retrospective clinical records were analyzed with predictive data mining models. The predictive success of the models obtained by various machine-learning algorithms was compared. A total of 595 clinical records were used in the study, including 348 males (58.49%) and 247 females (41.51%). It was found that the Gradient boosted tree algorithm achieves the best success with an accurate prediction success of 95.31%. In this study, an estimation method based on machine learning was developed to identify individuals with acute appendicitis. It is thought that this method will benefit patients with signs of appendicitis, especially in emergency departments in hospitals.

Keywords: Acute appendicitis, Machine learning, Surgery prediction

TEŞEKKÜR

Yüksek lisans ve Doktora eğitimim süresince eğitimime her türlü desteği veren, araştırmalarımda bana rehberlik eden, birikim ve deneyimlerini paylaşan, mesleğine yaklaşımıyla bizlere örnek olan, tezimin hazırlanmasında yardımlarını esirgemeyen ve her zaman hocam olmaya devam edecek olan Sayın Prof. Dr. Hasan Erbay’a, büyük fedakârlıklarla desteklerini esirgemeyen ve yol gösteren Dr. Öğr. Üyesi Emre Demir ve Dr. Öğr. Üyesi Hakan Kör’e, her zaman varlıklarıyla bana güç veren aileme, akademik kariyerim süresince her zaman desteğiyle yanımda olan ve akademik çalışmalarımı bitirmemde çok büyük emeği bulunan sevgili eşim Bediha Akmeşe’ye ve hayatımı güzelleştiren canım kızlarım Zümra Betül ve Zülal’e sonsuz sevgi ve teşekkürlerimi sunarım.

İÇİNDEKİLER DİZİNİ

Sayfa

ÖZET	iii
ABSTRACT	iii
TEŞEKKÜR	iv
İÇİNDEKİLER DİZİNİ	v
ŞEKİLLER DİZİNİ	vi
ÇİZELGELER DİZİNİ	x
KISALTMALAR DİZİNİ	x
1. GİRİŞ	1
2. AKUT APANDİSİT İLE İLGİLİ YAYINLARIN BİBLİYOMETRİK ANALİZİ	4
3. AKUT APANDİSİT	7
3.1. Tanım	7
3.2. İnsidans (Sıklık)	8
3.3. Tanı	9
3.3.1. Fiziki Bulgular	10
3.3.2. Laboratuvar Bulguları	10
3.3.3. Radyolojik Bulgular	11
3.3.4. Skorlama Sistemleri	14
3.4. Cerrahi Tedavi	16
4. VERİ MADENCİLİĞİ VE MAKİNE ÖĞRENMESİ	17
4.1. Veri Madenciliği Tanımı	18
4.2. Veri Tabanından Bilgi Keşfi	20
4.3. Çapraz-Endüstri Standardı Süreci	20
4.4. Veri Madenciliği Modelleri	24

4.5. Makine Öğrenmesi	25
4.6. Makine Öğrenmesi Türleri	26
4.6.2. Denetimli Öğrenme	26
4.6.3. Denetimsiz Öğrenme	27
4.6.4. Takviyeli Öğrenme	27
5. MATERYAL VE METOT	29
5.1. Veri Toplama	29
5.2. Özellik Seçimi ve Ağırlıklandırma	31
5.3. Veri Ön İşleme	32
5.4. Makine Öğrenmesi Modelinin Oluşturulması	33
5.5. Model Performans Değerlendirmesi	36
5.6. Makine Öğrenmesi Algoritmaları	38
5.6.1. Lojistik Regresyon	39
5.6.2. Yapay Sinir Ağları	41
5.6.2.1. Derin Öğrenme	42
5.6.3. K-NN (K En Yakın Komşu)	44
5.6.4. Destek Vektör Makineleri	46
5.6.5. Sınıflandırma ve Regresyon Ağaçları (CART)	47
5.6.6. Rastgele Orman	48
5.6.7. Gradyan Artırılmış Ağaçlar	50
6. ARAŞTIRMA BULGULARI	53
7. SONUÇLAR VE TARTIŞMA	62
EKLER	64
KAYNAKLAR	65

ŞEKİLLER DİZİNİ

<u>ŞEKİL</u>	<u>Sayfa</u>
2.1. Çalışmaların doküman tiplerine göre dağılımı.....	4
2.2. Yayın sayısına göre ilk on ülke.....	5
2.3. Çalışmaların dahil oldukları ilk on alan.....	5
2.4. Yayınların yıllara göre dağılımı.....	6
3.1. Appendix vermiformis	7
3.2. Apandisit ile ilgili demografik bilgiler, risk faktörleri ve gelir etkisi.....	8
3.3. 2012 yılında küresel çapta milyon kişi başına apandisit ölümü.....	9
3.4. Apandisit iltihabını gösteren 3 boyutlu grafik	9
3.5. Akut apandisit ultrason görüntüsü	12
3.6. Akut apandisit gösteren bir BT taraması.....	13
3.7. Akut apandisit gösteren bir MRG taraması.....	14
3.8. Alvarado skorlama sistemi.....	15
3.9. Apandisit şüphesi olan hastalarda yaklaşım.....	15
4.1. Veri bilimi ile ilgili kavramlar	18
4.2. Veri tabanından bilgi keşfi süreci	20
4.3. Veri madenciliği yöntemlerinin kullanım dağılımları.....	21
4.4. CRISP-DM sürecinin aşamaları.....	21
4.5. Genel veri madenciliği süreci.....	22
4.6. Veri Madenciliği Modelleri.....	24
4.7. Denetimli öğrenme.....	26
4.8. Denetimsiz öğrenme.....	27
4.9. Takviyeli öğrenme	28
5.1. Ön işlemeden sonra cinsiyete göre ameliyat geçirenlerin grafiği	33
5.2. Önerilen sistemin mimarisi	34
5.3. Önerilen sistemin mimarisi (çapraz doğrulama).....	35
5.4. Çapraz doğrulama	36
5.5. Karışıklık Matrisi	37
5.6. Makine öğrenmesi haritası	39
5.7. Lojistik regresyon.....	40

5.8. Basit yapay sinir ağı yapısı	42
5.9. Yapay sinir ağı ve derin öğrenme	42
5.10. Veri miktarına göre derin öğrenme ve klasik öğrenme.....	43
5.11. Makine öğrenmesi ve derin öğrenmede özellik çıkarımı	44
5.12. K En Yakın Komşu sınıflandırması	45
5.13. Doğrusal ve doğrusal olmayan iki ayrı veri seti.....	46
5.14. Torbalama ve Artırma yöntemleri arasındaki fark.....	49
5.15. Rastgele orman diyagramı	50
5.16. Gradyan artırma yöntemi	51
6.1. Algoritmaların doğruluk yüzdeleri.....	53
6.2. Yapay sinir ağları algoritmasına göre tahmin sayıları	54
6.3. K-NN algoritmasına göre tahmin sayıları	55
6.4. Lojistik regresyon algoritmasına göre tahmin sayıları.....	56
6.5. Destek vektör makinesi algoritmasına göre tahmin sayıları	57
6.6. CART algoritmasına göre tahmin sayıları	58
6.7. Rastgele orman algoritmasına göre tahmin sayıları	59
6.8. Gradyan artırılmış ağaçlar algoritmasına göre tahmin sayıları	60

ÇİZELGELER DİZİNİ

ÇİZELGE

	<u>Sayfa</u>
5.1. Veri kümesinin açıklaması.....	29
5.2. Akut apandisit veri kümesinde laboratuvar belirteçleri	30
5.3. İşlenmiş veri kümesinden bir kesit.....	30
5.4. Veri kümesinin dağılımı ve tanımlayıcı istatistikler	31
5.5. Özellikler ve ağırlıklandırma	32
5.6. Değerlendirme ölçütleri	37
6.1. Yapay sinir ağları algoritmasına göre karışıklık matrisi	54
6.2. K-NN algoritmasına göre karışıklık matrisi.....	55
6.3. Lojistik regresyon algoritmasına göre karışıklık matrisi	56
6.4. Destek vektör makinesi algoritmasına göre karışıklık matrisi.....	57
6.5. CART algoritmasına göre karışıklık matrisi	58
6.6. Rastgele orman algoritmasına göre karışıklık matrisi.....	59
6.7. Gradyan artırılmış ağaçlar algoritmasına göre karışıklık matrisi.....	60
6.8. Gradyan artırılmış ağaçlar algoritmasına göre değerlendirme ölçütleri	61

KISALTMALAR DİZİNİ

AA	Akut Apandisit
BT	Bilgisayarlı Tomografi
CRISP-DM	Cross Industry Standard Process for Data Mining (Çapraz-Endüstri Standardı Süreci)
CRP	C-reaktif protein
FN	False Negative (Yanlış Negatif)
FP	False Positive (Yanlış Pozitif)
HGB	Hemoglobin
HTC	Hematokrit (kan ayırıcı)
LYM	Lenfosit
KEEL	Evrimsel Öğrenmeye Dayalı Bilgi Çıkarımı
K-NN	K En Yakın Komşu
MCV	Ortalama eritrosit hacmi
ML	Machine Learning(Makine öğrenmesi)
MPV	Ortalama trombosit hacmi
MRG	Manyetik Rezonans Görüntüleme
NEU	Nötrofil
PLT	Thrombosis (Damarda kan pıhtılaşması)
TP	True Positive (Doğru Pozitif)
TN	True Negative (Doğru Negatif)
USG	Ultrasonografi
VTBK	Veri Tabanından Bilgi Keşfi
WBC	Beyaz kan hücreleri (lökosit)
YSA	Yapay Sinir Ağları

1. GİRİŞ

Bilişim sektöründeki gelişmeler, hayatımızın her alanında kendisini günden güne daha fazla hissettirmektedir. Özellikle sağlık sektöründe hasta doktor iletişimde kullanılan bilişim teknolojileri ürünleri, gerek donanım gerekse yazılım olarak çeşitlenerek artmaktadır. Günümüzde birçok hastalık için bulguların tespitinde geleneksel yöntemler kullanılmaktadır. Daha önceden yapılmış ve literatürde bulunan birçok çalışma, dijital ortamda alınan verilerin daha güvenli, erişilebilir ve işlenebilir halde olduğunu belirtmektedir [1]. Dijital ortamda elde edilen verilerin veri madenciliği yöntemleri kullanılarak analiz edilmesi, hasta verilerinin anlamlandırılması, tıbbi teşhis, uygun tedavi sürecinin belirlenmesi, teşhis ve tedavi etkinliğinin ölçülmesi ve uygun ilaç dozunun ayarlanması gibi amaçları bulunmaktadır.

Bilişim teknolojilerinin gelişimiyle sağlık sektöründe tele-sağlık hizmetleri gelişmeye ve çeşitlenmeye başlamıştır. Ülkemizde de çeşitli hastalıkların teşhisinde ve takibinde bilişim teknolojileri kullanılmaya başlanmıştır. Burada amaç insan hatalarının azaltılması, sağlık personelinin ve kaynakların daha verimli kullanılması, sağlık hizmetlerinin kalitesinin yükseltilmesi ve daha sürdürülebilir, etkin tedavi hizmeti sunmanın yanı sıra hızla artan tıbbi verilerin toplanması, saklanması ve analiz edilerek karar destek aşamasında uzman hekimin hizmetine sunulmasıdır.

Tıbbi tanı koyma, hastanın gerçekçi verileri, doğru tıbbi literatür bilgisi ve klinik tecrübeler gerektiren karmaşık ve önemli bir süreçtir. Tıbbi tanı süreci, içinde birçok beklenmedik durum bulundurduğundan dolayı diğer sektörlerdeki tanımlama süreçlerine göre çok daha karmaşık bir yapıdadır. Klinik kararlar çoğu zaman doktorların algı ve deneyim temeline dayanılarak alınır [2]. Bununla beraber hastalar her zaman şikâyetlerini doğru bir şekilde aktaramayabilir. Veri miktarının da çok hızlı bir şekilde artması karar verme açısından zorluklara sebep olmaktadır. Mevcut sağlık bilgi sistemleri birçok yönden hasta verisini anlamlandırma ve tedavi etkinliğini ölçmeyi sağlayacak analiz yöntemleri içermemektedir. Fakat bu tür analiz ve veriyi anlamlandırma yöntemleri gerek tedavi ekibinin zamanını etkin kullanması gerekse hastanın tedavi sürecinin seyrinin mercek altına alınması için yüksek önem taşımaktadır.

Apandisit komplikasyonlarını ve negatif apendektomiye önlemek için kesin bir tanı gereklidir. Apandisit tanısı öncelikle iltihaplanma için belirleyici olan biyokimyasal testlere dayanır. Bu yöntemin tek başına kullanılması, %12,3-19 oranında negatif bir apendektomi oranı ortaya çıkarmıştır. Modern görüntüleme ve laboratuvar incelemelerinde, gözle görülür iyileşmeler olmasına rağmen doğru teşhis bir problem olarak devam etmektedir [3-6]. Herhangi bir komplikasyonsuz tanı koymak için CRP (C-reaktif protein), WBC (lökosit) sayısı, NEU (nötrofil) oranı, bilirubin gibi ulaşımı kolay yöntemlerin yanı sıra, ultrason ve tomografi gibi görüntüleme teknikleriyle de çalışmalar yapılmıştır [7, 8]. Radyolojik yöntemler, özellikle ultrasonografi ve bilgisayarlı tomografi, akut apandisit ve komplikasyonlarının tanısında yaygın ve başarılı bir şekilde kullanılmaktadır [9, 10]. Bununla birlikte, bu yöntemler özel ekipman ve deneyimli radyologlar gerektirdiği için makine öğrenmesi, hastalığın teşhisinde alternatif bir yöntem olarak kullanılmıştır.

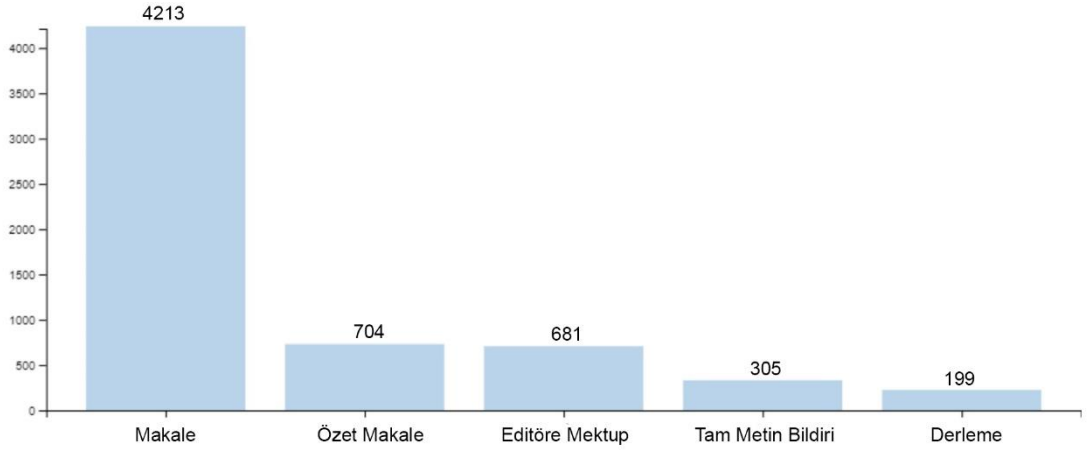
Veri madenciliği, istatistiksel yaklaşımları kullanarak büyük veri setlerinden gizli kalıpları keşfetmeye yönelik bir metodolojidir [11]. Geleneksel istatistiksel tekniklerin aksine, veri madenciliği temel olarak verileri bilgiye dönüştürme ve bu verilerden öğrenme ile ilgilidir [12]. Veri madenciliğinin çeşitli uygulamaları arasında, sağlık alanında kullanımı istisna değildir. Tıbbi veri setlerinin analizinde ve güçlü kalıpların çıkarılmasında çok faydalı olduğu kanıtlanmıştır [13, 14]. Bu çalışmada, hastaların kan örneklerinden elde edilen veriler kullanılarak cerrahinin gerekliliğini tahmin etmek üzerinde durulmuştur. Bu nedenle, hastalıkla ilgili tanılarının doğruluğunu test etmek, kaynak tüketimini en aza indirmek ve sınırlı sağlık hizmetlerinden daha doğru yararlanılmasını sağlamak amaçlanmaktadır.

Akut apandisit şüphesi olan çoğu hastada, teşhis prosedürleri kan değerleri veya elde edilen görüntüler kullanılarak gerçekleştirilir. 2008 yılında temel istatistiksel yöntemler ile akut apandisiti araştıran çalışmada, 302 apendektomi yapılan hastanın cinsiyet, yaş ve WBC değerleri incelenmiştir [15]. Ancak, bilişim teknolojileri kullanarak veri madenciliği veya makine öğrenmesi ile akut apandisiti öngörme veya tedavi etme konusunda çok az çalışma vardır. Demirhan ve arkadaşları, tıpta yapay zekâ uygulamaları araştırması yapmıştır. Yapay sinir ağlarının (YSA) akut sağ kasık ağrısı olan hastalar için tanıdaki rolünü değerlendiren bir sistem geliştirilmiştir. Geliştirilen sistemde, bir eğitim hastanesinden toplanan hasta verileri YSA'nın eğitim

ve testinde kullanılmıştır. Sistemin performansı Alvarado skoru ve deneyimli doktorların değerlendirmeleriyle karşılaştırılmıştır. Hastaların semptomları ve bulguları, hematolojik değerlendirme ve demografik bilgiler YSA'da giriş verileri olarak kullanılmıştır [16]. Bir başka apandisit çalışmasında, apandisit verilerinde bulanık çıkarım sistemlerinin ve lojistik regresyon sınıflandırıcılarının performansı karşılaştırılmıştır [17]. KEEL (Evrimsel Öğrenmeye Dayalı Bilgi Çıkarımı) veritabanından seçilen örneklem üzerinde seçilen R programı ile yapılan çalışmada, bulanık çıkarım sistemleri ve lojistik regresyonun doğruluk değerleri yüksektir. Verilerin kümeleme yöntemiyle işlendiği bir başka çalışmada cinsiyet, karın ağrısı, yaş parametreleri ve WBC, MPV (trombosit), LYM (lenfosit), NEU ve CRP değerleri kullanılmıştır. Çalışmadaki ideal küme sayısı dört olarak belirlenmiştir. Bu kümeleme çalışmasının doktor görüşlerini içeren ideal bir karar destek sistemi tasarlaması amaçlanmıştır [18].

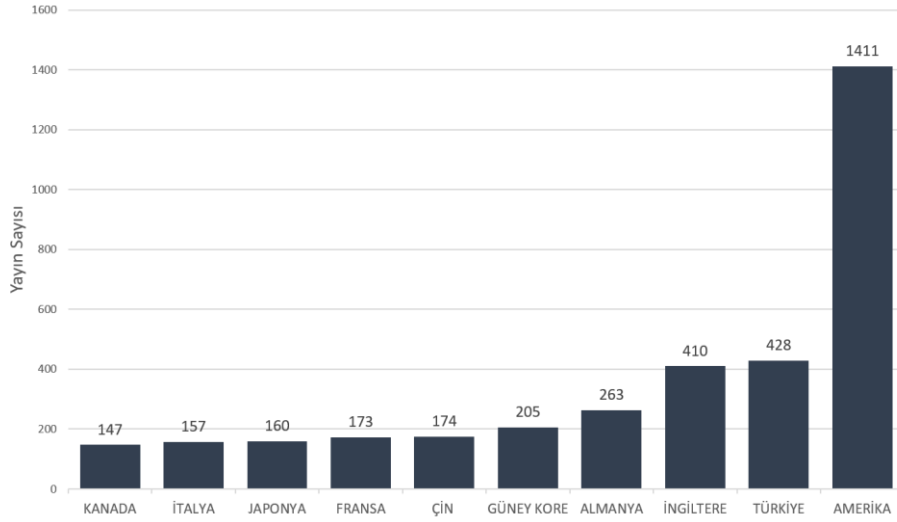
2. AKUT APANDİSİT İLE İLGİLİ YAYINLARIN BİBLİYOMETRİK ANALİZİ

Thomson Reuters Web of Science indeksine göre “akut appendicitis” ve “appendectomy” anahtar kelimesi ile çalışmaların başlığını kapsayan “Title” listesinde 1980 ve 2019 yılları arasındaki çalışmaların taranması sonucunda “5835” çalışma bulunmuştur. (Erişim tarihi: 21.01.2020; Title: ("acute appendicitis") or Title: (appendectomy); Analysis: [excluding]: Document Types: (Editorial Material or Note or Correction or Discussion or Early Access or Book Chapter or News Item or Book or Book Review or Hardware Review or Item About an Individual or Poetry or Retracted Publication); Timespan: 1980-2019). Şekil 2.1’e göre 1980 ve 2019 yılları arasındaki çalışmaların doküman tiplerine göre dağılımı; 4213’ü (%72,2) makale, 704’ü (%12,06) özet bildiri, 681’i (%11,67) editöre mektup, 305’i (%5,22) tam metin bildiri, 199’u (%3,41) derleme olarak görülmektedir.



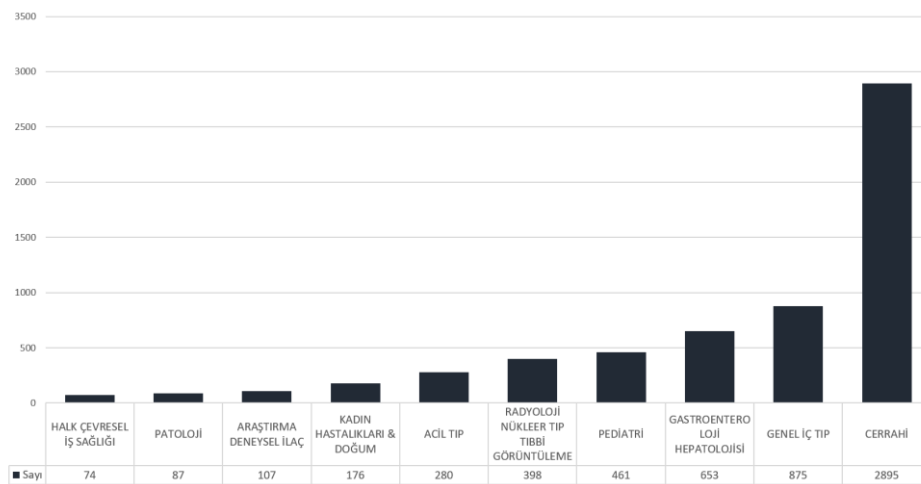
Şekil 2.1. Çalışmaların doküman tiplerine göre dağılımı

5835 çalışma incelendiğinde yayın yapan ilk on ülke sırasıyla Amerika (1411), Türkiye (428), İngiltere (410), Almanya (263), Güney Kore(205), Çin Halk Cumhuriyeti (174), Fransa (173), Japonya (160), İtalya (157) ve Kanada (147) olarak bulunmuştur. Bu konudaki çalışmaların %24’ü sadece Amerika’ya aittir. Türkiye ise 428 (%7,3) yayınlı bu konuda dünyada en çok çalışma yapan ülkeler arasında 2. sırada yer almaktadır. Şekil 2.2’de yayın sayısına göre ilk on ülkenin grafiği görülmektedir.



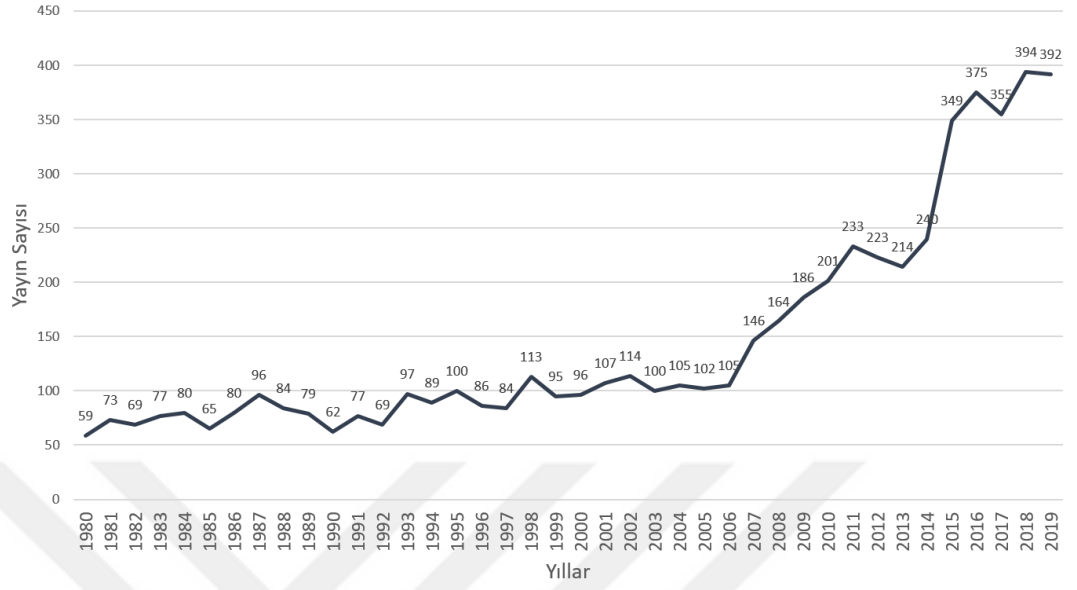
Şekil 2.2. Yayın sayısına göre ilk on ülke

5835 çalışma incelendiğinde, çalışmaların dahil oldukları ilk on alanın sırasıyla Cerrahi (Surgery) 2895 (%49,61), Genel İç Tıp (General Internal Medicine) 875 (%14,99), Gastroenteroloji Hepatolojisi (Gastroenterology Hepatology) 653 (%11,19), Pediatri (Pediatrics) 461 (%7,9), Radyoloji Nükleer Tıp ve Tıbbi Görüntüleme (Radiology Nuclear Medicine Medical Imaging) 398 (%6,82), Acil Tıp (Emergency Medicine) 280 (%4,79), Kadın Hastalıkları ve Doğum (Obstetrics Gynecology) 176 (%3,01), Araştırma Deneysel İlaç (Research Experimental Medicine) 107 (%1,83), Patoloji (Pathology) 87 (%1,49) ve Halk Çevresel İş Sağlığı (Public Environmental Occupational Health) 74 (%1,26) olduğu görülmüştür.



Şekil 2.3. Çalışmaların dahil oldukları ilk on alan

Şekil 2.4'te yıllara göre yayın sayılarının grafiği yer almaktadır. Bu grafik incelendiğinde yıllara göre yayın sayısının genellikle arttığı görülmektedir.



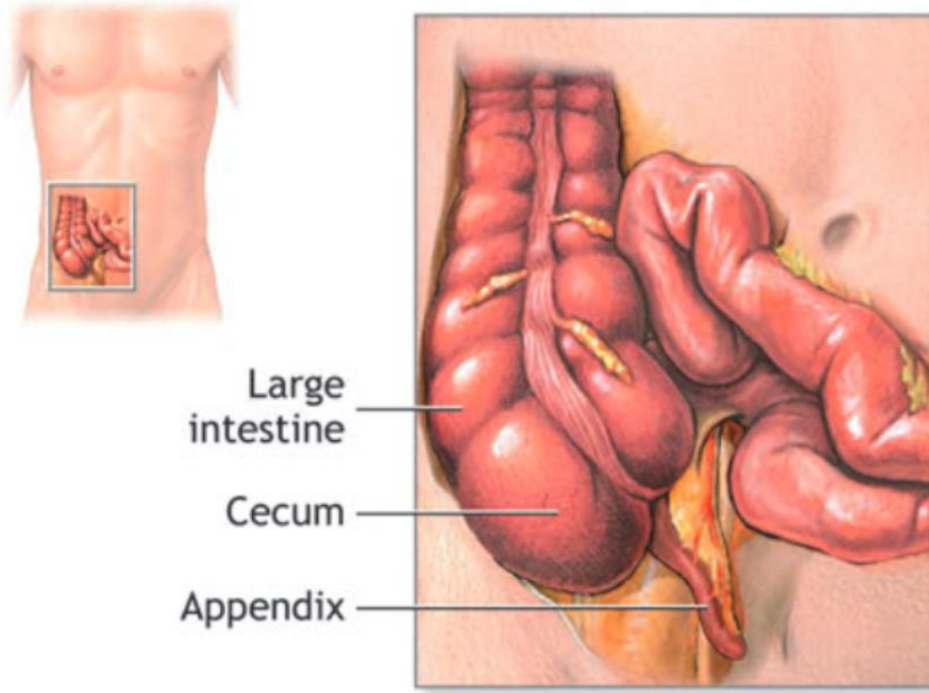
Şekil 2.4. Yayınların yıllara göre dağılımı

3. AKUT APANDİSİT

3.1. Tanım

Her yaş grubunda görülebilen akut apandisit (AA), Appendix vermiformisin iltihabi bir hastalığı olarak bilinmektedir. Ayrıca %7 hayat riskiyle beraber acil abdominal cerrahide en sık karşılaşılan endikasyondur [19].

AA için apendektomi, genel cerrahlar tarafından uygulanan en yaygın acil cerrahi işlemlerden biridir. Gecikmiş cerrahi, perforate apandisit oranını artırabileceği için apendektomi tanıdan birkaç saat sonra yapılmalıdır [20-23]. Apandisit, karnın en sık görülen cerrahi patolojisi olmasına rağmen teşhisi, takibi ve tedavisi cerrahın tecrübesine ve tercihine göre değişmektedir [24, 25]. Hastaların yaklaşık üçte biri atipik şikâyetler göstermektedir [26]. Bu durum genellikle tanıdaki gecikmeler ve oluşan komplikasyonlar ile ilişkilidir. AA şüphesi olan birçok hasta, herhangi bir komplikasyon olmadan tanı konup ameliyat edilse bile, %22 ila %62 arasında komplikasyon geliştirebilir [27]. Şekil 3.1’de Appendix vermiformis’in görüntüsü yer almaktadır.



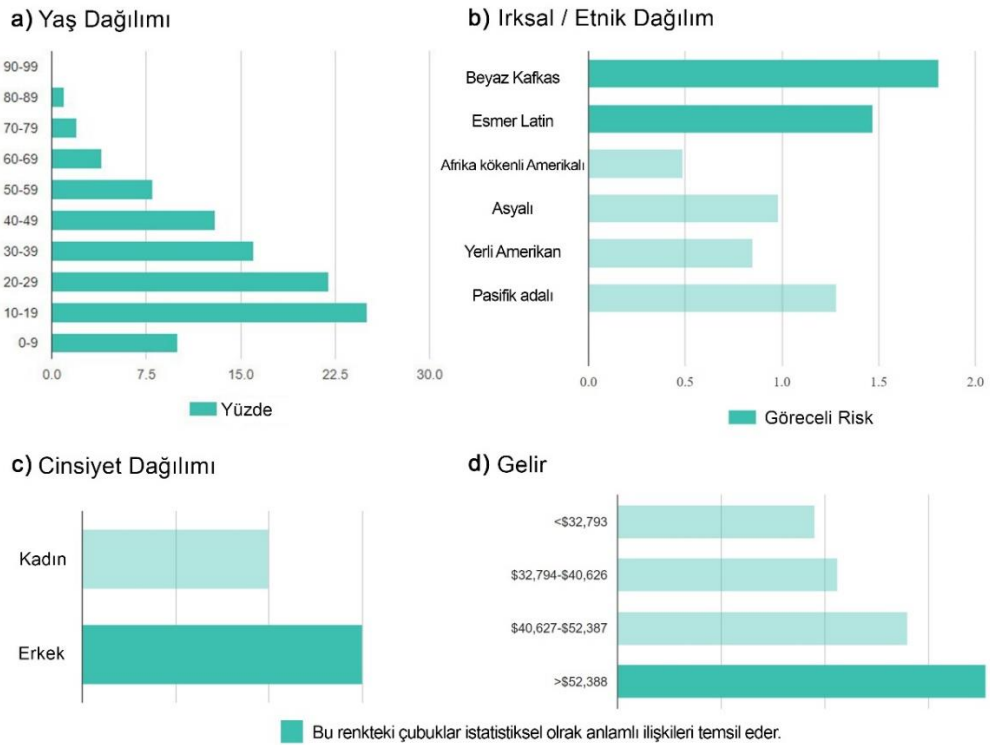
Şekil 3.1. Appendix vermiformis [28]

3.2. İnsidans (Sıklık)

AA, her yaştan geniş bir hasta grubunda görülür. Özellikle 10-30 yaş arasındaki genç erişkinlerde yaygındır. Bununla birlikte nüfusun yaklaşık %7'sinde apandisit görülmektedir [29-31].

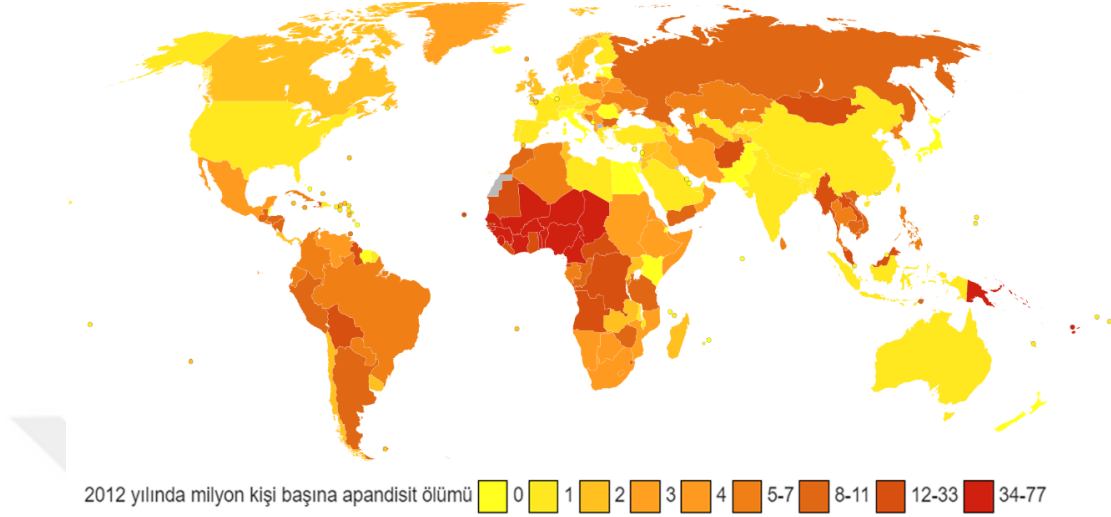
Non-travmatik akut cerrahi batının en sık nedeni akut apandisit [29]. AA'nın her yaşta ortaya çıkması mümkündür. Yetişkinlerde hastalık genellikle 20-30 yaşlarında görülmektedir [32]. 40 yaş sonrası ise görülme sıklığı giderek azalır. 10-29 yaş aralığındaki hastalarda akut apandisit belirtilerinin ortaya çıkma oranı %40'dır. Bu yüzden AA için daha çok gençlere özgü bir hastalık denilebilir. Ayrıca kadınlardaki görülme sıklığı erkeklere göre hafif daha azdır [33].

Şekil 3.2'de anonimleştirilmiş tıbbi kayıtlardan hesaplanan verilere göre oluşturulmuş grafiklere yer verilmiştir. Buna göre apandisit tanısı olan kişilerin yaş ortalaması 27'dir. Apandisit, erkekler arasında daha yaygındır. Kafkasyalılar ve Latinler arasındaki yaygınlık daha yüksektir. Yüksek gelirli bireylerde ise prevalans daha yüksektir [34].



Şekil 3.2. Apandisit ile ilgili demografik bilgiler, risk faktörleri ve gelir etkisi [34]

Küresel olarak 1990 yılında 88.000 olan ölüm sayısı 2013 yılında azalarak 72.000'e düşmüştür [35]. Dünya sağlık örgütünün (WHO) istatistiklerine göre 2012 yılında milyon kişi başına apandisit ölümü Şekil 3.3'te görülmektedir.



Şekil 3.3. 2012 yılında küresel çapta milyon kişi başına apandisit ölümü [36]

3.3. Tanı

AA tanısı, hasta öyküsü, fiziki muayene, laboratuvar belirteçleri ve görüntüleme yöntemleri ile konulmaya çalışılır [37, 38]. Apandisitinin temel belirtisi karın (abdominal) ağrısıdır. Şekil 3.4'te Apandisit iltihabını gösteren 3 boyutlu bir görsel bulunmaktadır.



Şekil 3.4. Apandisit iltihabını gösteren 3 boyutlu grafik [39]

Hastaların büyük çoğunluğunda ağrıyla beraber iştahsızlık, bulantı ve kusma da görülür. Belirtilerin meydana gelme sırası ayırıcı tanı için önemlidir. Semptomların sıralaması iştahsızlık, karın ağrısı ardından kusma şeklinde görülür. Karın ağrısı olmadan kusma gerçekleşmişse, ayırıcı tanı için apandisit düşünülmemelidir [33].

Morbidite ve mortalite için en önemli faktörün tanıdaki gecikme olduğu bilinmektedir. Ayırıcı tanı hastanın çocuk, kadın veya yaşlı olması durumunda daha da zordur. Küçük çocukların ve yaşlıların belirtileri anlatamamaları ve yeterli iletişim kurulamaması tanıyı geciktirebilir. Öte yandan, üreme çağındaki kadınlarda yumurtalık ve tubal hastalığının semptomları ve bulguları ortaya çıkabilir. Hekim, akut apandisiti ve karışabilecek diğer hastalıkları da iyi bilmelidir. Bu şartlar altında ilave tanı testleri gerekli olabilir. Doğru teşhis için Alvaroda skoru gibi bir skorlama veya tomografi, ultrason gibi görüntüleme teknikleri gerekebilir [40-42].

3.3.1. Fiziki Bulgular

AA'nın klinik tanısı detaylı hasta öyküsüne ve kapsamlı fizik muayenesine dayanır. Hastada ateş, yüzde kızarıklık ve ağız kokusu görülebilir. Ağrı, öksürmekle artış gösterir. Apandisiti doğrulamada sıçrama testi çocuklar için basit bir testtir. Çocuklardan sıçraması istendiğinde ağrısı olduğu için reddedecektir. Rektal veya vajinal muayene bulguları normal olabilmektedir. Rektal muayene, akut karın ağrısı olan hastada değerlendirmenin bir parçasıdır, ancak apandisit tanısında rektal incelemenin değeri tartışmalıdır. 1204 hastada yapılan bir çalışmadaki bulgulara göre rektal incelemede sağ taraflı ağrının varlığı apandisitli hastalarda daha sık olmasına rağmen tanıdaki katkısının çok az olduğu bildirilmiştir [43-45].

3.3.2. Laboratuvar Bulguları

Karın ağrısı ile hastaneye başvuran hastalar için tam kan sayımı, üre ve elektrolit analizleri için kan tahlili yapılmaktadır. AA tanısında en sık faydalanan belirteçler WBC ve CRP'dir. Özellikle CRP, WBC sayısı ve NEU yüzdesi tanı için değerlidir. Bu belirteçler skorlama sistemlerinde parametre olarak kullanılmaktadır. Bu testler tanıda kullanılabileceği gibi komplikasyon olup olmadığı hakkında da bilgi verebilir [15, 46].

CRP'nin kandaki düzeyleri yaş, doku travması, enfeksiyon, kronik inflamasyon, malignite, otoimmün hastalıklar başta olmak üzere pek çok nedene bağlı olarak yükselebilir. Duyarlı ancak non-spesifik bir belirteçtir. İnflamasyonun başlangıcından itibaren 6-8 saat içinde yükselmeye başladığı için ilk muayenede diğer tetkiklerle birlikte istenmelidir. Kontrol amaçlı alınan kan örneğinde WBC sayısı ile beraber CRP düzeyinde de azalma saptanırsa apandisit tanısından uzaklaşılır. Yükselen değerler ise tanıyı destekler. CRP değerinde yükselme olmaması tanıyı dışlatmaz [47].

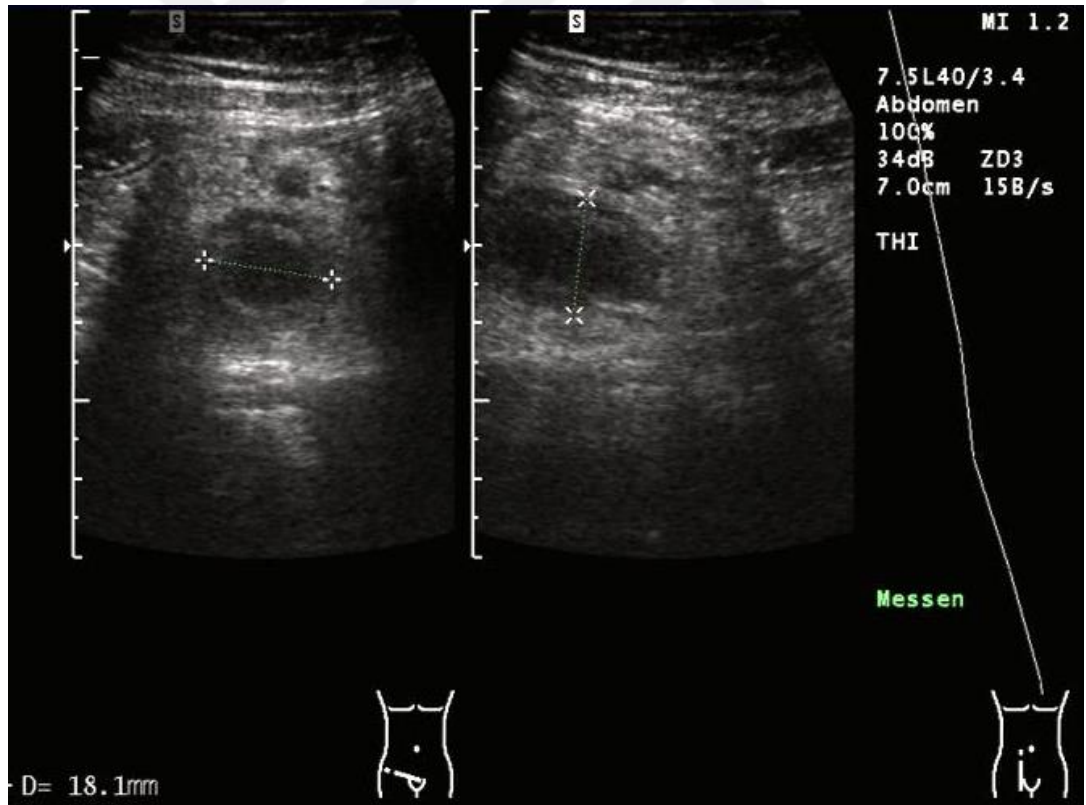
AA, çok sayıda çalışmaya konu olsa da tanısal değerleri yoktur. Klinik olarak gözlemlenen hastalar için yapılan testlerin tekrarlanması yararlı olabilir. Ancak yapılan iki testin normal olması durumunda testlerin tekrarlanması veya daha ileri tetkiklerin yapılması gerektiği belirtilmektedir. İnflamatuvar belirteçler (CRP, WBC ve NEU sayısı) normal ise AA tanısı olası değildir. Bu testlerdeki performanslar çalışılan popülasyonlarla bariz bir şekilde ilişkilidir. Laboratuvar sonuçlarını okurken dikkat edilmesi gereken nokta farklı merkezlerde az sayıda kişinin katılımı ile yapılan farklı çalışmalarda, bu değerlerin yüksekliğinin tanımlanmasında farklı kesim noktalarının kullanılmış olmasıdır. AA tanısı için laboratuvar bulgularının ayırıcı tanıda yer alan diğer patolojilerle de desteklenmesi gerektiğine vurgu yapılmaktadır. Ayrıca amilaz, lipaz, karaciğer fonksiyon testleri ve pıhtılaşma testleri de AA ayırıcı tanısı için gerekebilmektedir [48-56].

Hasta hikâyesi ve muayenenin tanıdaki doğruluk oranı %84 iken, ilave olarak WBC sayımı ve CRP konsantrasyonunu içeren bilgilerin değerlendirilmesi doğruluğu %92'ye kadar çıkarır. Kesin apandisit teşhisi olan hastalar, sıvı takviyesi ile antibiyotik almalıdır. Ameliyata alınmaları için daha fazla araştırma yapılmasına gerek yoktur. Ancak kesin olmayan bulgular ile hastaneye başvuran hastalar için görüntüleme yöntemleri değerlendirilmelidir [19].

3.3.3. Radyolojik Bulgular

Muayene ve laboratuvar bulguları ile tanı koyabilmek hastaların büyük çoğunluğu için mümkündür. AA tanısı öykü, fiziki muayene ve laboratuvar bulguları klasik olduğunda klinik olarak konulabilir ve görüntüleme yapılmayabilir. Ancak atipik vakalar için görüntüleme yöntemlerinin de kullanılması gerekebilir [57].

Ultrasonografik (USG) değerlendirme, rutin uygulamada akut apandisit şüpheli hastalar için ilk başvuru olan yöntemidir. USG incelemesinde çapının >6 cm olması, bası ile farklılık göstermemesi ve çevresinde sıvı varlığı apandisit lehine düşündürülen bulgular olarak sıralanabilir. Akut apandisit için diğer patolojilerle ayırımında USG incelemesi oldukça değerli sonuçlar vermektedir [58]. USG radyasyon içermediği için çocuk ve genç insanlarda tercih edilen bir tekniktir. Bununla birlikte tecrübeli kişiler için apandisit göstermede değerli bir yöntemidir [59]. Ucuz ve hızlı olması, radyasyon içermemesi, kontrast madde gerektirmemesi, gebelerde kullanılabilmesi ve karın ağrısının diğer nedenlerinin de belirlenebilmesi gibi özellikler ultrasonografinin avantajları arasında sayılabilir. Ancak uzmana olan bağımlılık ve bağırsaktaki gazlar nedeniyle apendiksin görülemediği durumlar ise dezavantajdır [59, 60]. Akut apandisit tanısında ultrasonografi ancak yardımcı olabilir. USG, fizik muayene ve öykünün önüne geçmemelidir [61]. Şekil 3.5'te Akut apandisit gösteren bir ultrason görüntüsü yer almaktadır.



Şekil 3.5. Akut apandisit ultrason görüntüsü [62]

Günümüzde giderek yaygınlaşan diğer bir görüntüleme yöntemi de bilgisayarlı tomografidir (BT). Özellikle USG ile tanınan dışlanamadığı durumlarda genellikle

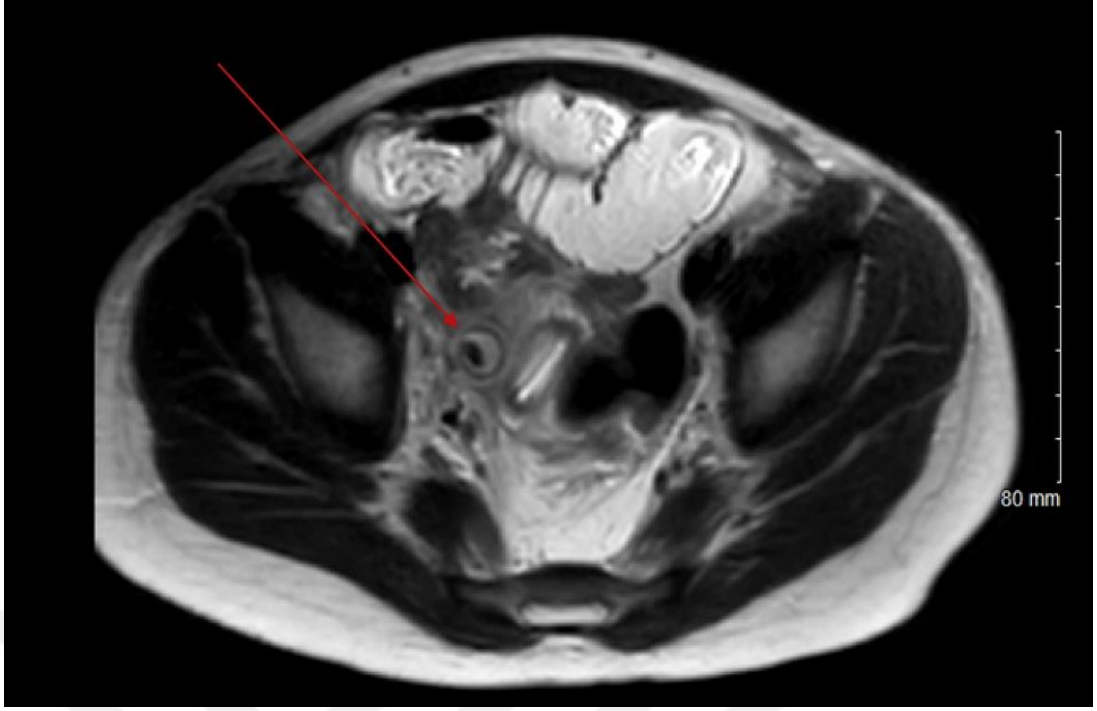
başvurulmaktadır. Apendiks çapında artış, duvarda kalınlaşma ve/veya kontrast tutulumu ve çevre yağ dokuların değerlendirilmesi BT’de akut apandisit lehine düşündüren bulgular olarak sıralanabilir [58, 63]. BT, hastanede kalış süresinde azalmaya sebep olmaktadır. Yine de yalnızca akut apandisit tanısında şüphelenilen hastalarda radyolojik bulgulara bakılmalıdır. Cerrahi müdahalenin yerine geçmesine veya cerrahi müdahalede gecikmeye sebep olmamalıdır [64]. Yüksek dozda radyasyona maruz kalma ve gebelerde kullanılamaması dezavantajlarındandır. Şekil 3.6’da akut apandisit gösteren bir BT taraması görülmektedir.



Şekil 3.6. Akut apandisit gösteren bir BT taraması [65]

AA hastalarında tanı için kullanılan bir başka görüntüleme yöntemi de manyetik rezonans görüntülemesidir (MRG). AA tanısında kullanımı son yirmi yılda artmıştır. MRG’nin kolay ulaşılabilir olmaması, fazla maliyetli olması, görüntüleme süresinin uzun olması ve hasta uygunluğunun aranması kullanımını sınırlamaktadır.

Şekil 3.7’de akut apandisiti gösteren bir MRG taraması görülmektedir. Gebeliğin dönemine göre USG’deki değişen sonuçlar, gebelerde MRG’nin AA tanısında tercih edilebilir bir yöntem olmasına neden olmuştur. MRG’nin gebelerde AA tanısında duyarlılığı %100, özgüllüğü %90 civarındadır [66-68].



Şekil 3.7. Akut apandisit gösteren bir MRG taraması [69]

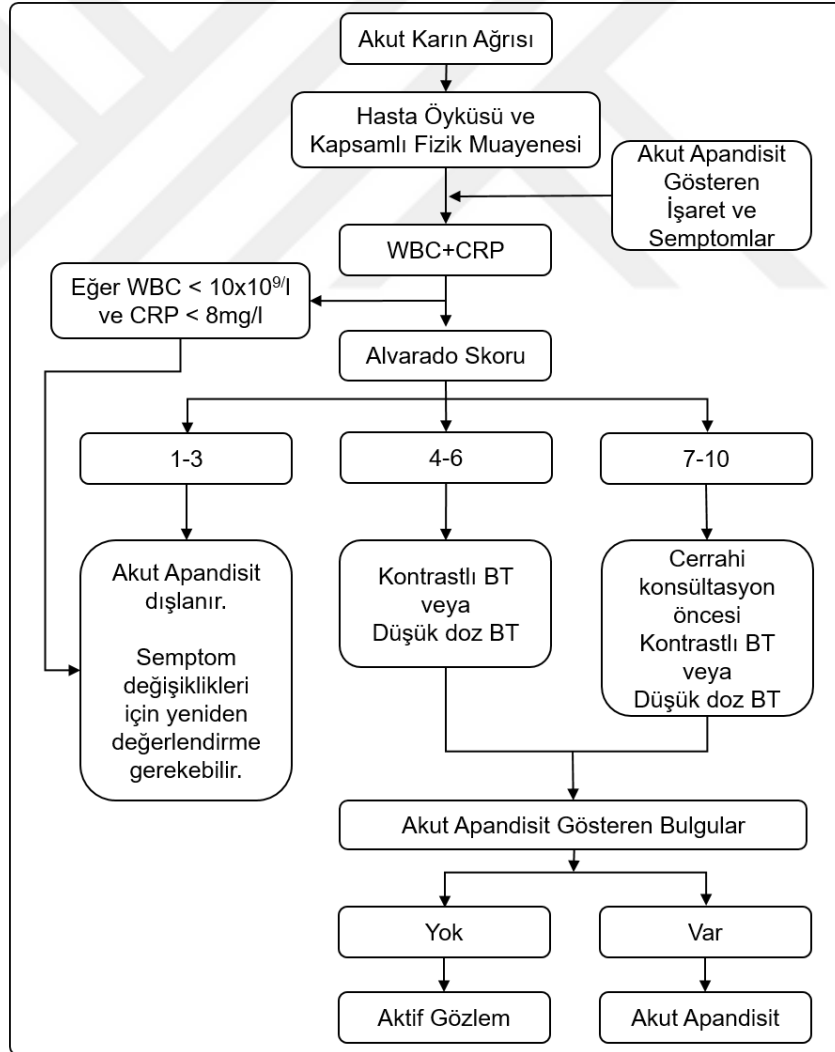
3.3.4. Skorlama Sistemleri

AA tanısında skorlama sistemleri temel olarak klinik, fizik muayene, laboratuvar incelemeleri ve görüntüleme yöntemlerinin bir arada kullanılması ile oluşturulmaktadır. Tanıda her ne kadar yeni uygulamaların da devreye girmesi ile birlikte başarı elde edilse de hala negatif apendektomi ve perforasyon hızlarının fazla olması farklı yaklaşımlara ihtiyaç olduğunu ortaya koymuştur [70]. Şekil 3.8’de Alvaroda skorlama sistemini görülmektedir. 1986 yılında Alvarado, 8 parametreden oluşan ve bugün en çok kabul gören bir skorlama sistemi geliştirmiştir. Bu skorlama toplam 10 puan üzerinden değerlendirilmektedir. Toplam puan >6 ise hastaların ameliyat edilmesi, 5-6 puan alanların izlenmesi, <5 puan alan hastalarda da AA olasılığının çok düşük olacağı için tanının dışlanabileceği önerilmiştir [71]. Alvarado skoruna göre klinik değerlendirmeyi, fiziki ve laboratuvar bulgularını ve görüntülemeyi birleştiren sistematik bir değerlendirme Şekil 3.9’da bulunmaktadır [72].

Alvarado Skorlama Sistemi		
Semptomlar	Ağrının sağ alt kadrana göçü	1
	Bulantı, Kusma	1
	İştahsızlık	1
Klinik Bulgular	Sağ alt kadranda hassasiyet	2
	Rebaund bulgusu	1
	Yüksek ateş	1
Laboratuvar Bulgular	Lökositoz	2
	Nötrofilik sola kayma	1
Toplam		10

1-4	5-7	8-10
-----	-----	------

Şekil 3.8. Alvarado skorlama sistemi



Şekil 3.9. Apandisit şüphesi olan hastalarda yaklaşım [72].

Daha sonra birçok çalışmada Alvarado skoru sistemi değiştirilerek yeni skoru sistemi sistemleri üzerinde çalışılmıştır. Ohmann, Tzanakis, Andersson skoru sistemi bunlardan bazılarıdır. Skoru sistemlerinin, özellikle görüntüleme yöntemlerinin kullanılmadığı sağlık merkezlerinde negatif apendektomi oranlarını azalttığı ve komplikasyon gelişmeden operasyon kararı alınmasına yardımcı olduğu görülmüştür [73]. Doğru tanıyı bulmak için skoru, görüntüleme yöntemleriyle desteklenmelidir [42].

3.4. Cerrahi Tedavi

AA'nın kesin tedavisi cerrahidir. Günümüzde antibiyoterapi gündeme gelse de en sık kullanılan tedavi yöntemi apendektomidir [64]. Komplikasyonları önlemek için tedavinin altın standardı, erken apendektomi olarak önerilmiştir. Bütün apendektomilerin %84'ü akut patoloji için yapılmaktadır. Bununla birlikte, yapılan bir araştırmada cerrahi girişim uygulanan hastaların %16'sında normal bir appendix bulunmuştur. Bu oran da %68'ini kadın hastalar oluşturmaktadır. [74, 75].

Cerrahinin zamanlaması oldukça önemlidir. Bulgular başladıktan sonra geçen 36 saatlik sürede perforasyon oranları %16-36 bulunmuştur. Takip eden her 12 saatte riskin %5 arttığı bildirilmiştir [20]. Perforasyonda morbidite, basit apandisit göre 100 kat daha fazladır ve mortalite %0,1 iken %10'lara ulaşmaktadır. Bu nedenle erken cerrahi girişim ön planda olmalıdır [76]. Apendektomi açık cerrahi olarak veya laparoskopik girişimle yapılabilir.

4. VERİ MADENCİLİĞİ VE MAKİNE ÖĞRENMESİ

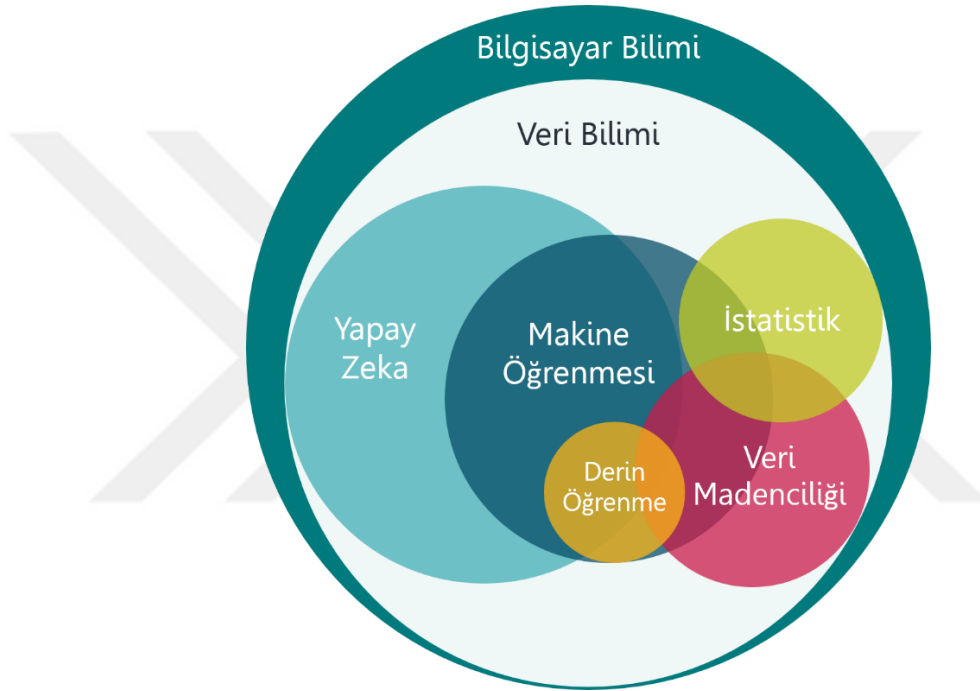
Günümüzde sürekli olarak ve hayatın her alanında çok sayıda verinin kaydı yapılmaktadır. Genel bir söylem olarak bilgi çağında yaşadığımız ifade edilir. Ancak gerçekte veri çağında yaşamaktayız. Her geçen gün hayatın neredeyse bütün alanları hakkında terabayt veya petabayt büyüklüğünde veri bilgisayar ağlarımızda birikmektedir [77]. Bu durumda hali hazırda var olan verilerden nasıl daha etkin faydalanılacağı konusu üzerinde bir gereklilik söz konusudur. Bu gereklilik, veri madenciliğinin doğmasına neden olmuştur. Buradaki en önemli konu ise geçmişin en iyi şekilde anlaşılmasını sağlamak ve daha iyi karar verebilmeyi sağlayan tahmin yapabilecek ve en etkili girdileri tespit edebilecek verilerin ortaya çıkarılabilmesi için uygun bilgiyi ayıklayabilmektir [78].

Son yıllarda veri bilimi ile ilgili gelişmeler neticesinde birçok terimin ortaya çıkması, karışıklığa neden olabilmektedir. Veri madenciliği ve makine öğrenmesi kavramları bu duruma örnek olarak gösterilebilir. Her ikisinin de analitik süreçler olması, her ikisinin de örüntü tanımada iyi olması, her ikisinin de karar vermeyi geliştirebilmek için verilerden öğrenmekle ilgilenmesi ve büyük miktarda veri gerektirmesi iki kavram arasındaki benzerlikleri gösterir. Makine öğrenmesi, modeller oluşturmak ve kalıplar bulmak için bazı veri madenciliği tekniklerini kullanabilir, böylece daha iyi tahminler yapabilir. Bununla beraber veri madenciliği de bazen daha doğru analiz üretmek için makine öğrenme tekniklerini kullanabilir. Bu anlamda bu iki kavramın iç içe olduğu söylenebilir. Ancak bir takım farklılıkları da vardır. Örneğin veri madenciliği verilerde zaten mevcut olan kalıpları ararken, makine öğrenmesi, önceden var olan verilere dayanarak gelecekteki sonuçları tahmin etmek için kullanılabilir. Veri madenciliği, insan müdahalesine ve karar almaya dayanan daha manuel bir süreçken, makine öğrenmesi, genellikle ilk kurallar uygulandıktan sonra, öğrenme ve bilgi çıkarma süreci otomatiktir. Yani insan müdahalesi olmadan gerçekleşir ve makine kendi kendine öğrenmeye başlar. Ayrıca veri madenciliği, kalıpları bulmak için mevcut bir veri kümesinde kullanılırken, makine öğrenmesi yeni veri kümeleri hakkında tahminler yapmayı öğreten bir “eğitim” veri kümesi üzerinde eğitilir [79].

Makine öğrenmesi veri madenciliği alanıyla da doğrudan ilişkilidir. Makine öğrenmesi tekniklerinin büyük veri tabanlarına uygulanması işlemine veri madenciliği adı

verilmiştir [80]. Makine öğrenmesi için veri madenciliği sürecinin uygulama aşaması da denilebilir. İki terim arasındaki en önemli fark; veri madenciliği ortaya çıkan bilgi ile ilgilenirken, makine öğrenmesi, öğrenme yöntemlerinin geliştirilmesi ve tahminlerin en iyi performansı sergilemesi ile ilgilenir. Makine öğrenmesi sadece veriler üzerinde uygulanan bir teknikten ibaret değildir. Aynı zamanda yapay zekâ konusudur [81].

Hem veri madenciliği hem de makine öğrenimi veri bilimine dayanır ve genellikle bu çatı altında kalır. Bu kavramların birbirleriyle ilişkisi Şekil 4.1’de görülmektedir.



Şekil 4.1 Veri bilimi ile ilgili kavramlar

4.1. Veri Madenciliği Tanımı

Disiplinler arası bir çalışma alanı olan veri madenciliğinin literatürde birçok tanımı yer almaktadır. Genel bir tanımla mevcut verilerin analiz edilmesi sonucunda verilerden anlamlı bilgilerin çıkarılması işlemidir.

Veri madenciliği (aynı zamanda bilgi keşfi süreci), istatistiksel yaklaşımlar kullanarak muazzam veri kümelerinden gizli kalıpları keşfetmeye yarayan bir metodolojidir [11]. Veri madenciliği genel olarak büyük veri kümelerinden bilgiye ulaşmak ya da bilgiyi madenleme işlemi olarak tanımlanabilir [82].

Veri madenciliği, teknolojinin sağladığı veri depolama ve veri işleme hızındaki artışla, farklı disiplinlerin sunduğu araçlarla, çok büyük miktardaki verilerle, karar vericinin etkin bir şekilde bilgiye dayalı karar vermesini sağlamak amacıyla, önceden bilinmeyen ve klasik metotlarla ortaya çıkarılması güç olan, faydalı ve ilginç olan ilişki, örüntü veya eğilimlerin otomatik ya da yarı otomatik bir şekilde ortaya çıkarılmasıdır [83]. Veri madenciliği, geniş kapsamlı bir mevcut veri setinin önceden bilinmeyen özelliklerinin araştırılmasına odaklanmaktadır [84]. Veri madenciliği yöntemleri birçok araştırmacı tarafından tahmin amacıyla da kullanılmaktadır. Veri madenciliği teknikleri altındaki sınıflandırma ve değerlendirme, eğitim verilerinin oluşturulmasında, tahmin modelinin sınıflandırılmasında ve sınıflandırma verimliliğinin test edilmesinde yardımcı olmaktadır [85].

Veri madenciliği, veri tabanında biriken veri setleri üzerinde yapay zekâ algoritmaları ile istatistiksel analiz yöntemlerinin bir kombinasyonun kullanılması sonucu, önceden öngörülemeyen karar verme süreçlerinde kullanışlı olabilecek faydalı bilgilerin çıkarımını yapabilen bir yöntemdir. Bu yöntem, test sonuçlarını tahmin etmek, ürün geliştirmek, tıbbi teşhis ve tedavi sürecini belirlemek gibi amaçlara yönelik olarak tıp, biyoloji, genetik, kimya, uzay, eğitim, yüzey analizi ve coğrafi bilgi alanları ile özellikle rekabet avantajı kazandırması nedeniyle pazarlama, işletme, bankacılık gibi daha birçok sektörde kendisine kullanım alanı bulmuştur [86]. Bugüne dek veri madenciliğinin pek çok tanımı yapılmıştır. Veri madenciliği tanımları incelendiğinde, ortak nokta, çok büyük miktarlarda veri içermesi, bu verilerden anlamlı bilgiler elde edilmesi ve çeşitli verilerden daha önce keşfedilememiş bilgilerin ortaya çıkarılmasıdır [87].

Sayısal teknolojilerin gelişmesiyle birlikte kurumlarda ve işletmelerde çok fazla veri birikimi oluşmaktadır. Verinin büyüyen hızına yetişmek her geçen gün daha da zorlaşmaktadır. Depolanan bu veri yığınları içinden anlamlı ilişkilerin, örüntülerin ve eğilimlerin ortaya çıkarılması ihtiyacı doğmuş, doğru tahminlerin yapılması önem kazanmıştır. Veri tabanı, veri ambarı, internet gibi birçok veri depolanan ortamda veriler gün geçtikçe daha fazla birikmekte ve buna karşılık veriyi inceleyecek bilim adamlarının, mühendislerin ve analistlerin sayısı bu kadar hızlı artmamaktadır [88]. Bu verilerin sunduğu bilgileri kavramak, güçlü veri analizi araçları olmadan insani

yetenekleri aşmaktadır. Kısacası bu durum karar vericiyi veri zengini, aynı zamanda bilgi fakiri yapmaktadır [82, 89].

4.2. Veri Tabanından Bilgi Keşfi

Verinin bilgiye dönüşme süreci bilgi keşfi olarak adlandırılmaktadır. Veri kalitesinin veri madenciliği, makine öğrenmesi ve bilgi keşfi için büyük önemi bulunmaktadır [90]. Veri tabanından bilgi keşfi (VTBK) olarak da bilinen bilgi keşfi süreci bazı veri alanlarında yeni bir bilginin aranması işlemleri olarak tanımlanmaktadır [91]. Bu ve benzer tanımlara bakıldığında veri madenciliğinin bir süreç olduğu görülmektedir.



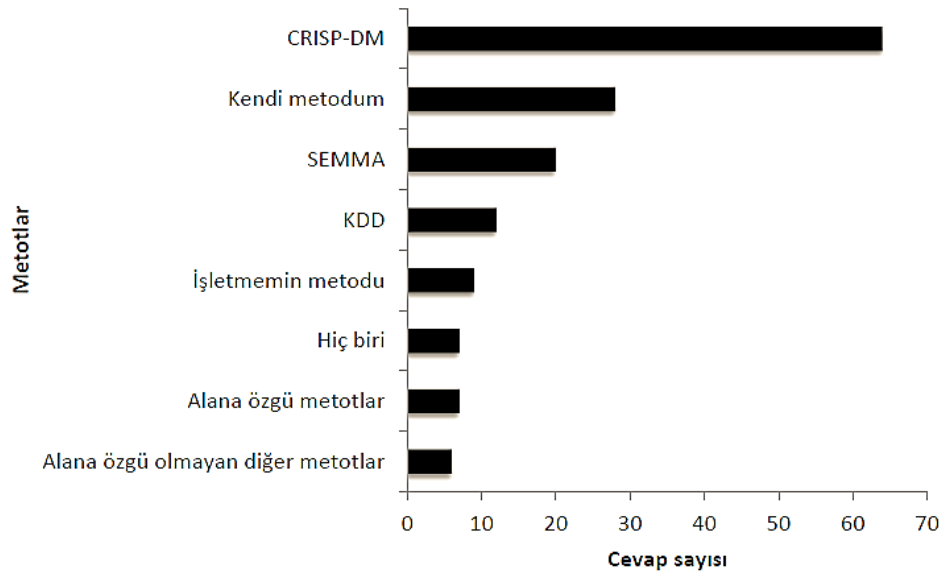
Şekil 4.2 Veri tabanından bilgi keşfi süreci [92].

Şekil 4.2’de veri tabanından bilgi keşfi sürecinin ana safhaları görülmektedir. Bilgi keşfi süreci, veri üzerinde makine öğrenmesi metotlarının uygulanmasıdır. Diğer bir ifadeyle veri madenciliği olarak adlandırılır [93, 94].

4.3. Çapraz-Endüstri Standardı Süreci

Veri madenciliği süreciyle ilgili en çok tercih edilen modellerden biri de Çapraz-Endüstri Standardı Süreci (Cross Industry Standard Process for Data Mining) modelidir. CRISP-DM şeklinde kısaltılan model, 1996 yılının sonlarına doğru Daimler Chrysler, SPSS ve NCR markalarının temsilcileri olan veri madenciliği pazarının üç deneyimli analisti tarafından tasarlanmıştır [95].

Şekil 4.3’te veri madenciliğinde temel olarak hangi yöntemi kullanıyorsunuz sorusuna verilen cevapların oranından da en yaygın olarak CRISP-DM yönteminin kullanıldığı anlaşılabilmektedir.



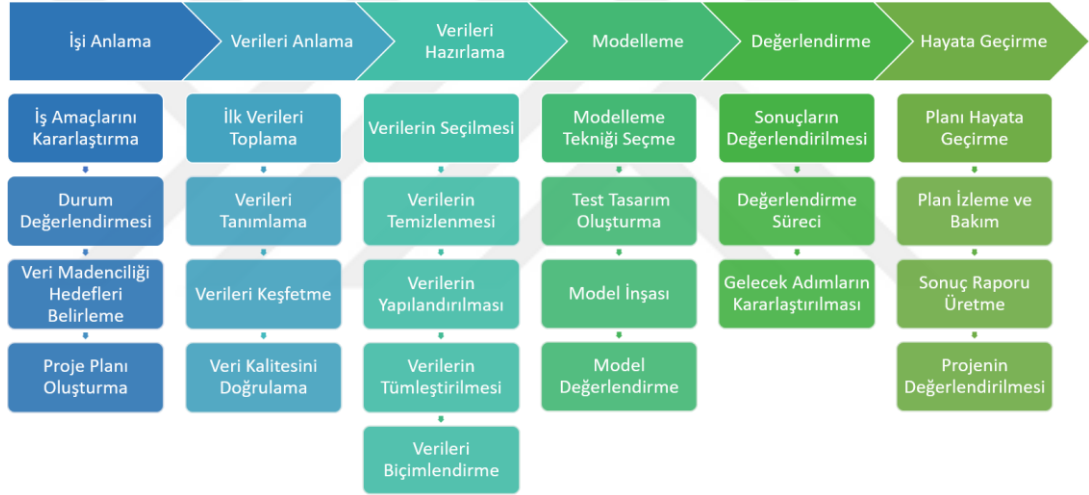
Şekil 4.3. Veri madenciliği yöntemlerinin kullanım dağılımları [96]



Şekil 4.4. CRISP-DM sürecinin aşamaları [97]

Veri madenciliğinin CRISP-DM metodolojisi, prosesin güvenilir ve standart olabilmesi için adım adım prosedürler önermektedir. Şekil 4.4'te bu adımlar görülmektedir. Bu metodoloji; iş anlayışı, veri anlayışı, veri hazırlama, modelleme,

değerlendirme ve dağıtımı içeren döngüsel bir süreci takip eder [95]. Veri madenciliği süreci, CRISP-DM'ye göre altı aşamadan meydana gelen bir döngü içerir. Her bir aşama sıra ile ilerlemektedir. CRISP-DM, veri madenciliği projesi için yol haritası görevi görmektedir. Bu döngü işin anlaşılmasıyla başlar. Daha sonra analiz için gerekli olan verilerin toplanıp anlaşılmasıyla süreç devam eder. Bu iki aşama birbirleriyle geri bildirimli şekilde çalışmaktadır. Bir sonraki aşama verilerin hazırlanmasıdır. Hazırlanan veri modellenir ve bu iki aşama da birbirlerini geri bildirimli olarak çalışmaktadır. Değerlendirme aşamasına geçildiğinde istatistiksel olarak modelin geçerliliği ve sonuçların doğruluk derecesi ölçülür. Ölçümü yapılan model, analizler için kullanılabilir hale getirilir. Kurulu modellerin gelecekteki veri setlerine uyarlanabilmesi sağlanabilir. Bu çalışmada da veri madenciliği için CRISP-DM metodolojisinden faydalanılmıştır.



Şekil 4.5. Genel veri madenciliği süreci

Şekil 4.5'te CRISP-DM'nin adımları ve bu adımların içerdiği görevler yer almaktadır.

İş anlama: CRISP-DM'nin bu ilk aşaması veri madenciliği çalışmasının hedeflerini ve ihtiyaçlarını anlamaya yöneliktir. Analistlerin hedefi, veri madenciliği uygulamalarının sonuçlarına etki edebilecek önemli etkenleri ortaya çıkarmaktır. Veri madenciliği projelerindeki başarı, projenin dikkatli bir şekilde planlanması, çok özel, gerçekleştirilebilir ve ölçülebilir bir hedefin olmasına bağlıdır [98].

Verileri anlama: İlk verilerin toplandığı ve bu toplanan verileri anlamaya ve tanımlamaya yönelik işlemlerin yapıldığı, verilerle ilgili ilk görüşlerin ortaya çıktığı

ve veri kalitesinin değerlendirildiği aşamadır. Bu aşamanın sonucunda analizi yapan kişi, çalışmada kullanılacak verilerle ilgili ön bilgi edinmiş olur.

Verileri hazırlama: Ham veriden başlayarak verinin son haline geldiği sürece kadar yapılması gereken tüm işlemlerin ele alındığı aşamadır. Verilerin hazırlanması ve analize uygun hale getirilmesi için diğer aşamalara göre genellikle daha fazla zaman ve emek sarf edilir. Analiz için uygun olan değişkenlerin belirlenmesi, verilerin dönüşümü, temizlenmesi, yapılandırılması ve biçimlendirilmesi gibi süreçleri içermektedir.

Verilerin ön işleminden geçirilmesi genellikle veri hacminin çok fazla olduğu durumlarda en çok süre isteyen aşamadır. CRISP-DM sürecinde en çok çaba ve zamanın (%80'inin) verilerin hazırlanması aşamasında harcandığı belirtilmektedir [96].

Modelleme: İşin anlama aşamasında belirlenen hedefler ışığında veri madenciliği tekniklerinin uygulandığı ve modelin oluşturulduğu aşamadır. Bu aşama kendi içinde modelleme tekniği seçimi, test verisi hazırlanması, modelin oluşturulması ve modelin değerlendirilmesi olarak dört kısımdan oluşur [95].

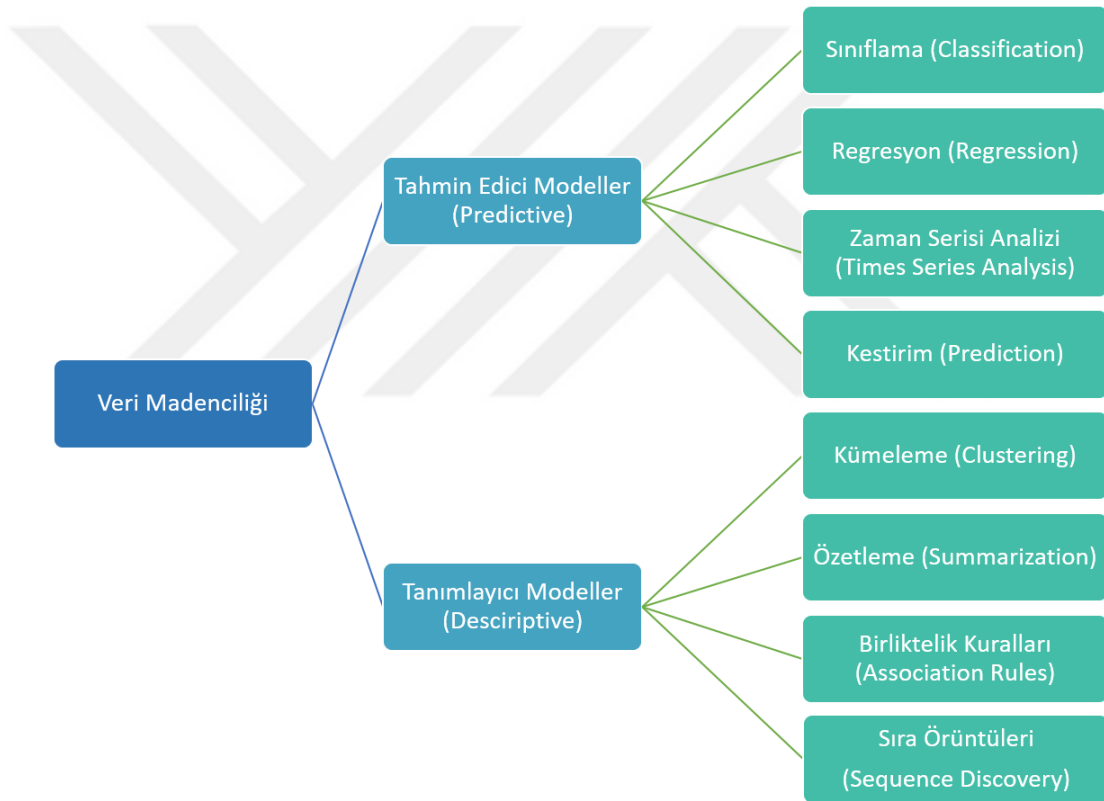
Kullanım amaçlarına göre algoritmalar tanımlayıcı ve tahmin edici yaklaşımlar olmak üzere ikiye ayrılır. Bu aşamada modelleme teknikleri içerisinde kullanıldıkları amaçlara göre en uygun olan teknik seçilir.

Değerlendirme: Modelin işi anlama aşamasında belirlenen amaçlar açısından değerlendirilmesi işlemi bu aşamada yapılır. Süreç gözden geçirilir ve bir sonraki aşamada veri madenciliği süreçlerinden hangisinin uygulanacağına karar verilir [92].

Hayata Geçirme: Sürecin son kısmı olan hayata geçirme aşamasıdır. İzleme ve bakım işlemleri için planlama yapılır ve proje sonuçları raporlanır. Sonuçlar değerlendirilerek yorumlanır ve ortaya çıkarılan bilgiler neticesinde alınan kararlar gerçek hayatta uygulanır. Nihai amaç veriden kazanılan bilgiyi arttırmak olsa da açığa çıkarılan bilginin düzenlenmesi ve karar vericinin kullanabileceği ve anlayabileceği bir şekilde sunulması önemlidir.

4.4. Veri Madenciliği Modelleri

Veri madenciliği modelleri, tahmin edici modeller ve tanımlayıcı modeller olmak üzere iki ana başlık altında toplanabilir. Tahmin edici modellerde, sonuçları bilinen verilerden yola çıkılarak bir model geliştirilmektedir. Geliştirilen bu model ile sonuçları bilinmeyen verilerin sonuç değerlerinin öngörülmesi amaçlanmaktadır [99]. Tahmin edici modellerde amaç geleceği tahmin eden çıkarımların yapılmasıdır. Tanımlayıcı modellerde ise verilerdeki örüntülerin tanımlanması ile karar vermede kılavuz olabilecek çıkarımların yapılması sağlanmaktadır. [100]. Şekil 4.6'da veri madenciliği modelleri görülmektedir.



Şekil 4.6. Veri madenciliği modelleri

Akpınar (2000) ise veri madenciliği modellerini gördükleri işlev bakımından aşağıdaki gibi sınıflandırmıştır [99]:

- Sınıflama (Classification) ve Regresyon (Regression),
- Kümeleme (Clustering),

- Birliktelik Kuralları (Association Rules) ve Ardışık Zamanlı Örüntüler (Sequential Patterns)

Sınıflama ve Regresyon, verinin bir kısmı eğitim verisi olarak kullanılır ve kural çıkarımı yapılır. Elde edilen kurallardan yola çıkarak sınıfı belli olmayan verilerin sınıfı tahmin etmek için kullanılır. Sınıflama kategorik değişkenleri öngörürken, regresyon süreklilik gösteren değişkenlerin öngörülmesinde kullanılır [101].

Kümeleme, verinin içindeki benzerlikler veya uzaklıklar dikkate alınarak gruplandırılmasıdır. Kümeleme, genellikle bölümlendirme problemlerini çözmekte kullanılır. Kümeleme analizinde, kümeler içindeki gözlem birimlerine göre önceden belirlenmiş bir özellik dikkate alınarak birbirine benzeyen kümeler bulunur. Elde edilen kümelerdeki gözlem birimleri homojendir [102].

Birliktelik Kuralları, bir veri kümesindeki sıklıkla birlikte görülen özelliklere ait ilişki kuralların ortaya çıkarılmasıdır. Pazar sepet analizi adıyla da bilinen bu model, veri tabanında yer alan bir kaydın olduğu durumda diğer kaydın olma olasılığını değerlendirerek sonuca ulaşır. Ardışık zamanlı örüntüler ise birbiriyle ilişkili ve birbirini takip eden zamanlarda meydana gelen olaylar arasındaki ilişkiyi tanımlar. Örneğin X ameliyatı gerçekleştirildikten 15 gün sonra %20 olasılıkla Y enfeksiyonu oluşur şeklinde bir çıkarım ardışık zamanlı örüntüdür [99].

4.5. Makine Öğrenmesi

Bilgisayarların öğrenme işlevini yerine getirebilmesi için çeşitli algoritmaların kullanılması ile ilgili çalışma alanına makine öğrenmesi denir [103]. Belli bir problemin çözülmesi için insanın problem çözme yeteneklerinin taklit edilmesi ve bu problemin çözülmesi için sistemin daha önceden bilgi ile donatılması makine öğrenmesi alanına girmektedir [104]. Makine öğrenmesi, analitik ve istatistiksel örüntü tanıma ve modelleme için otomatik veri analizi terimidir. Yeni verilerden tekrar tekrar öğrenmeye, oluşturulan modelleri güncellemeye ve böylece büyük veri setlerinde gizli bilgileri veya kalıpları bulmaya olanak tanır [95]. Makine öğrenme algoritmaları mevcut verileri kullanarak kendisini eğitir. Daha sonra olası yeni durumlar için tahminde bulunur. Literatüre bakıldığında çok sayıda makine öğrenmesi

yöntemi bulunmaktadır. Bu yöntemlerin başarısı, veriye göre farklılıklar göstermektedir. Bu yüzden belli bir yöntemin her veri için uygun olduğu söylemek doğru değildir.

Medikal verilerin analizinde de makine öğrenmesi algoritmaları başlangıçtan beri kullanılmaktadır. Günümüzde verilerin toplama ve depolama adına daha ucuz ve uygulanabilir yöntemlerin gelişmesi, modern hastanelerin görüntüleme ve veri toplama cihazları ile donatılması, verilerin geniş paylaşım alanlarında toplanmasını ve paylaşılmasını sağlamıştır [104].

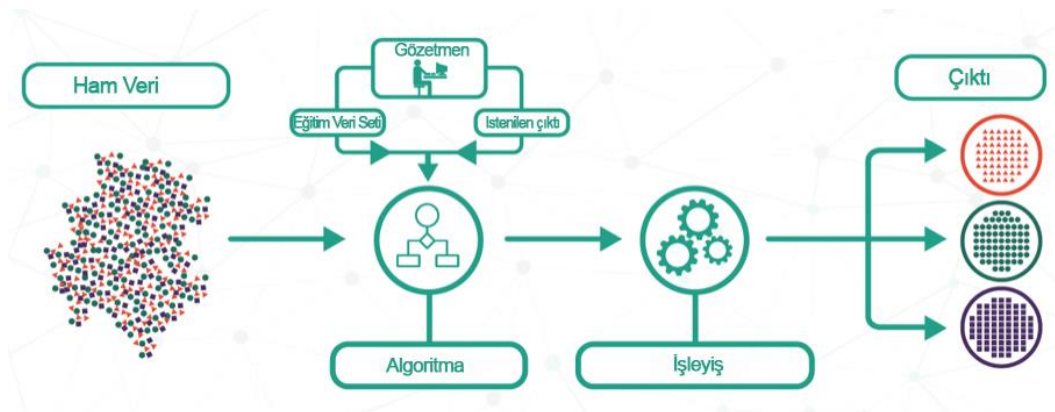
4.6. Makine Öğrenmesi Türleri

Makine öğrenmesi temelde üç başlık altında incelenir;

1. Denetimli Öğrenme (Supervised Learning)
2. Denetimsiz Öğrenme (Unsupervised Learning)
3. Takviyeli Öğrenme (Reinforcement Learning)

4.6.2. Denetimli Öğrenme

Bu öğrenme modelinde eğitim verileri kullanılarak girdiler ve çıktılar arasındaki ilişkiyi belirleyen bir fonksiyon oluşturulur.

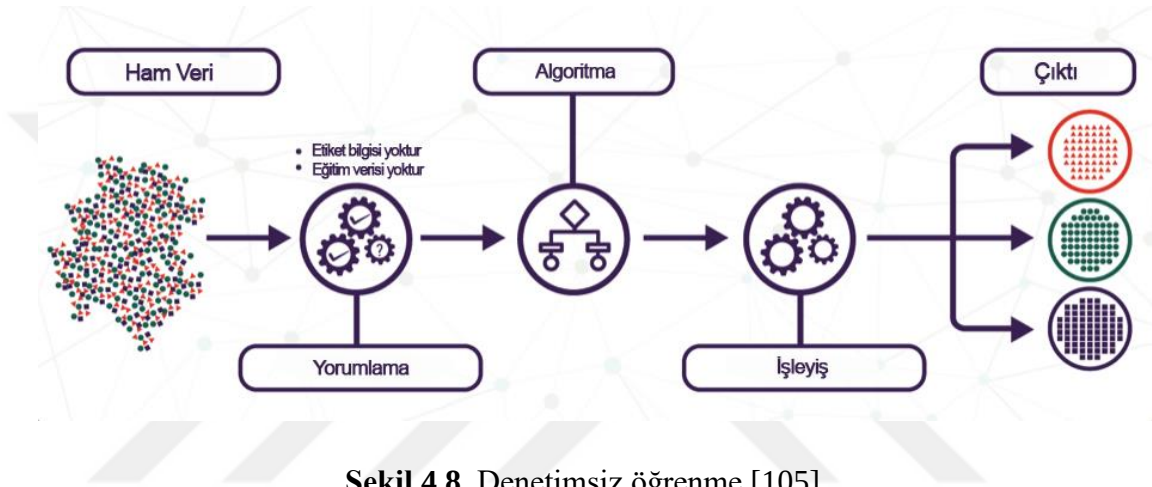


Şekil 4.7. Denetimli öğrenme [105].

Şekil 4.7’de denetimli öğrenme modelinin yapısı yer almaktadır. Denetimli öğrenme, etiketlenmiş verilerden oluşan öğrenme sürecidir. Veride, kategorik çıkışlar varsa sınıflandırma, nümerik çıkışlar varsa regresyon algoritmaları kullanılır. Geleceğe yönelik bir tahmin söz konusuysa bu model tercih edilir. Bu öğrenme modelinde öğrenmek için eğitim veri setine ihtiyaç duyulur.

4.6.3. Denetimsiz Öğrenme

Denetimsiz öğrenmede veri örneklerinin etiket bilgisi yoktur. Şekil 4.8’de denetimsiz öğrenme modelinin yapısı yer almaktadır.



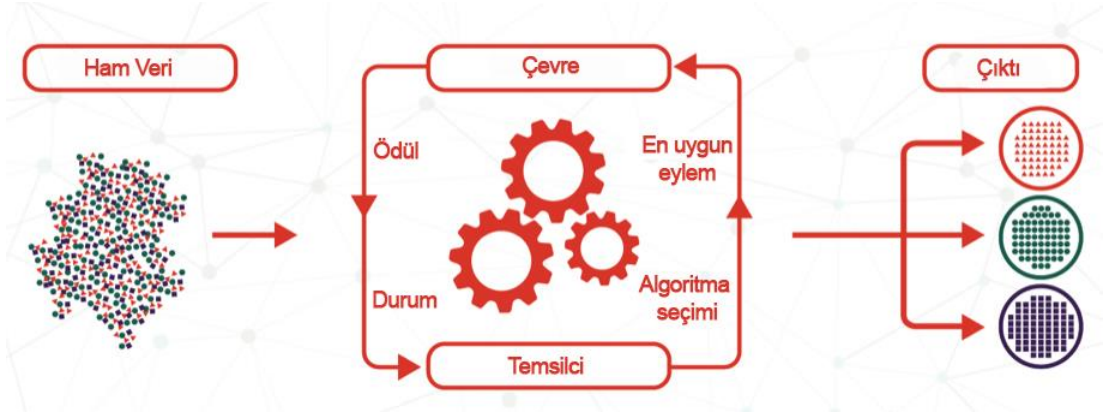
Şekil 4.8. Denetimsiz öğrenme [105].

Bu yöntemde giriş verisinin hangi sınıfa ait olduğu belirsizdir. Girdilerin uzaklık, benzerlik gibi metriklerle göre verilerin birbirlerine yakın benzerliklerinin yer aldığı kümeler öğrenilir ve bu şekilde kümeleme yapılır [80]. Bu öğrenme modelinde herhangi bir eğitim verisine ihtiyaç duyulmamaktadır.

4.6.4. Takviyeli Öğrenme

Diğer iki öğrenme yönteminden daha farklıdır. Şekil 4.9’da takviyeli öğrenme modelinin yapısı yer almaktadır. Model, adım adım öğrenilir. Üretilen sonuçlar için doğru veya yanlış şeklinde geri besleme veren bir eğitmen vardır. Makine bir karar verir ve kararın doğru olduğu durumlar için sistem ödüllendirilir. Kararın yanlış olduğu durumlarda ise sistem cezalandırılır. Asıl amaç makinenin çevreyle etkileşimi ve bu etkileşim sonucunda aldığı geri beslemeyle en uygun hareket tarzını

geliştirmesidir. Takviyeli öğrenmede büyüme, makinelerin yaptıkları şeyin sonucunu öğrenmelerine yardımcı olan çok çeşitli algoritmaların üretilmesine yol açmıştır.



Şekil 4.9. Takviyeli öğrenme [105].

5. MATERYAL VE METOT

Önceki bölümlerde akut apandisit ile ilgili yayınların bibliyometrik analizi, akut apandisit tanımı, tanısı ve tedavisi, veri madenciliği ve makine öğrenmesi literatür taramaları ile gözden geçirilmiştir. Çalışmada, makine öğrenmesi algoritmaları kullanarak erken akut apandisit tanısı için kolay, hızlı ve doğru bir tahmin yöntemi geliştirmek amaçlanmıştır ve yapay sinir ağları, K-NN, lojistik regresyon, destek vektör makinesi, sınıflandırma ve regresyon ağaçları (CART), rastgele orman ve gradyan artırılmış ağaçlar algoritmaları olmak üzere 7 adet makine öğrenme algoritması denenmiştir. Verilerin analizi için Python (3.7), RStudio ve Rapid Miner Studio (9.5) programları kullanılmıştır.

5.1. Veri Toplama

Veriler, Hitit Üniversitesi Eğitim ve Araştırma Hastanesi'nde toplanan klinik kayıtlardan elde edilmiştir. Bu veriler, apandisit şüphesi ile acil servise gelen 18 yaşından küçük hastaların cinsiyet, yaş, laboratuvar göstergeleri ve ameliyat geçirip geçirmediğini göstermektedir. Çalışmada, 348 erkek (%58,49) ve 247 kadın (%41,51) olmak üzere toplam 595 klinik kayıt kullanılmıştır. Ön işlemeden sonra klinik kayıt sayısı 595'den 428'e düşmüştür. Veri hakkındaki açıklamalar, laboratuvar belirteçleri, veri kümesinden bir kesit, dağılım ve tanımlayıcı istatistikler sırasıyla Çizelge 5.1, Çizelge 5.2, Çizelge 5.3, Çizelge 5.4'te verilmiştir.

Çizelge 5.1. Veri kümesinin açıklaması

İsim	Tip	Tanım	Rol
Grup	Kategorik	Ameliyat olmayan/Ameliyat olan	Hedef
Cinsiyet	Kategorik	Kadın/Erkek	Giriş
HGB	Nümerik	Hemoglobin	Giriş
NEU	Nümerik	Nötrofil	Giriş
LYM	Nümerik	Lenfosit	Giriş
MCV	Nümerik	Ortalama eritrosit hacmi	Giriş
MPV	Nümerik	Ortalama trombosit hacmi	Giriş
HTC	Nümerik	Hematokrit (kan ayırıcı)	Giriş
PLT	Nümerik	Trombosit (Damarda kan pıhtılaşması)	Giriş
CRP	Nümerik	C-reaktif protein	Giriş
WBC	Nümerik	Beyaz kan hücreleri (lökosit)	Giriş

Veri kümesinin açıklamasını gösteren her kayıt için veri adı, veri tipi ve tanımı ve işlevi Çizelge 5.1’de gösterilmektedir.

Çizelge 5.2. Akut apandisit veri kümesinde laboratuvar belirteçleri

İsim	Değer aralığı	Ameliyat olan	Ameliyat olmayan	P değeri
Grup	0 veya 1			
Cinsiyet	1 veya 2	77/137	104/110	< 0,008
HGB	1,4-130	13,63±8,18	13,02±1,43	0,348
NEU	0,8-29	11,82±5,31	9,11±5,77	< 0,001
LYM	0,2-94	3,09±7,09	2,44±1,17	< 0,016
MCV	4-97,3	80,98±8,76	81,2±5,79	0,434
MPV	6-99	14,03±20,02	9,31±7,93	0,708
HTC	3-99	40,32±6,24	39,69±3,49	0,143
PLT	111-593	272,49±73,58	274,37±82,73	0,954
CRP	0-302	39,61±51,62	22,86±40,64	< 0,001
WBC	6-31590	15280±5302	12541±6259	< 0,001

Çizelge 5.2, veri setindeki laboratuvar belirteçlerini göstermektedir. Grup değişkeninde 0, ameliyat olmayanları ve 1 ameliyat olanları ifade etmektedir. Cinsiyet değişkeni için 1 kadın ve 2 erkek anlamına gelir. Diğer veriler kan numunelerindeki değerlerdir. Değerler ortalama ± standart sapma olarak sunulmuştur. Kalın olanlar, anlamlı P değerini 0,05'ten az gösterir.

Çizelge 5.3. İşlenmiş veri kümesinden bir kesit

Grup	Cinsiyet	HGB	NEU	LYM	MCV	MPV	HTC	PLT	CRP	WBC
0	1	13,3	7,28	1,58	85,6	7,5	39,4	246	3,02	9460
0	1	10,6	10,33	1,57	76,1	6,8	32,9	466	14,7	13200
0	2	12,1	2,27	2,07	84,3	8,3	36,6	207	63,7	5210
0	1	14,3	14,94	2,13	84,2	8,4	44,1	324	4,75	18030
1	2	13,9	5,62	3,26	90,1	7,2	43,7	265	95,8	10160
1	2	14,9	10,48	1,5	86,3	8,6	45,7	167	3,14	12670
1	2	13,1	8,95	1,83	75,9	7,9	40,1	193	7,43	11600
1	1	13,9	7,43	1,8	84,5	7,8	41,5	355	27,3	18720
1	1	14,8	8,99	0,94	83,1	8,1	43	241	52,5	10390
1	1	13,3	9,32	1,42	88,2	7,3	41,3	299	3,14	11530

Çizelge 5.3'te veri kümesinden bir kesit yer almaktadır.

Çizelge 5.4. Veri kümesinin dağılımı ve tanımlayıcı istatistikler

	Count	Mean	Std	Min	%25	%50	%75	Max
Cinsiyet	428			1				2
Grup	428			0				1
HB	428	13,32	5,87	1,4	12,2	13	14	130
NEU	428	10,46	5,7	0,8	5,6	10	14,02	29
LYM	428	2,77	5,08	0,2	1,5	2,1	3	94
MCV	428	81,09	7,42	4	78,3	81,6	84,82	97,3
MPV	428	11,67	15,39	6	7,8	8,4	9,2	99
HTC	428	40,01	5,06	3	37,2	40	42,4	99
PLT	428	273,42	78,2	113	219	257	321,25	593
CRP	428	31,23	47,15	0	3,3	9,3	36,25	302
WBC	428	13911,3	5954,01	6	8987,5	13205	18140	31590

Çizelge 5.4'teki veri kümesinin dağılımı ve tanımlayıcı istatistikler şunları içermektedir;

- Sayı (Count): giriş sayısı
- Mean (Ortalama): girişlerin ortalaması
- Std: Standart sapma
- Min: minimum giriş
- %25: İlk kartil
- %50: Medyan veya ikinci kartil
- %75: Üçüncü kartil
- Max: Maksimum giriş

5.2. Özellik Seçimi ve Ağırlıklandırma

Karın ağrısı şikâyeti ile hastaneye başvuran hastalarda apandisit tanısı koymak için önemli olan özellikleri tanımlamada çeşitli çalışmalar yapılmıştır. Literatürde tanı için önemli olarak belirtilen nitelikler ve doktor tavsiyelerinin yanı sıra niteliklerin ağırlıkları da dikkate alınmıştır. Korelasyon ve ki-kare istatistiği ile etiket niteliğine göre özelliklerin ağırlıkları hesaplanmıştır. Bir niteliğin ağırlığı ne kadar yüksek olursa, o kadar ilgili sayılır. Korelasyon, iki özellik arasındaki ilişki derecesini ölçen -1 ile +1

arasında bir sayıdır. Korelasyon için pozitif bir değer, pozitif bir ilişki anlamına gelir. Korelasyon için negatif bir değer, negatif veya ters bir ilişkiyi ifade eder. Ki-kare istatistiği, gözlemlenen frekansların dağılımının teorik olarak beklenen frekanslardan farklı olup olmadığını belirlemek için kullanılan parametrik olmayan bir istatistiksel tekniktir. Ki-kare istatistikleri nominal veriler kullanır, dolayısıyla araçlar ve varyanslar yerine bu test frekansları kullanır. Hesaplanan ağırlıklar 0 ile 1 arasında normalleştirilmiştir ve Çizelge 5.5’de görülmektedir.

Çizelge 5.5. Özellikler ve ağırlıklandırma

No	Özellikler	Ağırlıklandırma (Korelasyon)	Özellikler	Ağırlıklandırma (Ki-kare)
1	NEU	1	NEU	1
2	WBC	0,948	WBC	0,994
3	CRP	0,769	CRP	0,436
4	MPV	0,676	MPV	0,201
5	Cinsiyet	0,535	LYM	0,120
6	LYM	0,270	HTC	0,119
7	HTC	0,255	Cinsiyet	0,079
8	HB	0,211	MCV	0,075
9	MCV	0,021	MPV	0,049
10	MPV	0	HB	0

5.3. Veri Ön İşleme

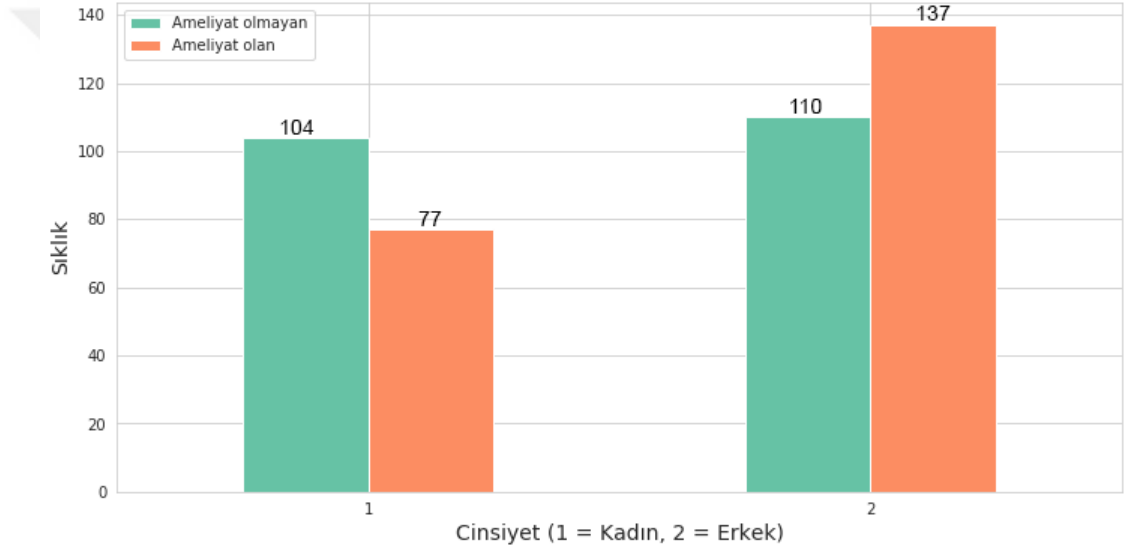
Verilerin kalitesi, tahminin sonucunu büyük ölçüde etkiler. Veri hazırlama, veri temizleme, veri dönüştürme ve veri indirgeme gibi süreçler tahmin doğruluğu üzerinde etkisi büyüktür. Biyolojik, kimyasal, meteorolojik veya anket çalışmalarındaki verileri içeren veri kümeleri eksik değerler içerebilmektedir. Eksik değerlerin meydana gelme sebeplerinin çok sayıda sebebi bulunabilir. Bunlardan bazıları;

- Elektronik sensor bozuklukları,
- Ölçüm hatalarından kaynaklanan nedenler,
- Bilimsel bir deney sonucunun ortaya çıkmaması (örneğin kimyasal bir tepkimenin gerçekleşmemesi),
- Görüntü dosyalarının yeterli çözünürlüğe sahip olmaması,

- Dijital sistemlerde veri transferinde meydana gelen problemler,
- Bilgisayar operatörünün veri giriş hataları,
- Anket katılımcılarının bazı sorulara cevap vermekten kaçınması

gibi nedenler gösterilmektedir [106-109]. Ön işlemenin gerektiği gibi yapılmadığı durumlarda çalışmanın başarısında ciddi düşüşler olabilir. Bu, ön işlemenin modelde önemli bir rol oynadığı anlamına gelir [110].

Kayıtlardaki eksik değerler niteliklerin doğruluğunu korumak için silinmiştir. Ön işleme aşamasında, eksik veriler temizlenmiştir ve sonuç olarak, klinik kayıt sayısı 595 kayıttan 428'e düşmüştür.



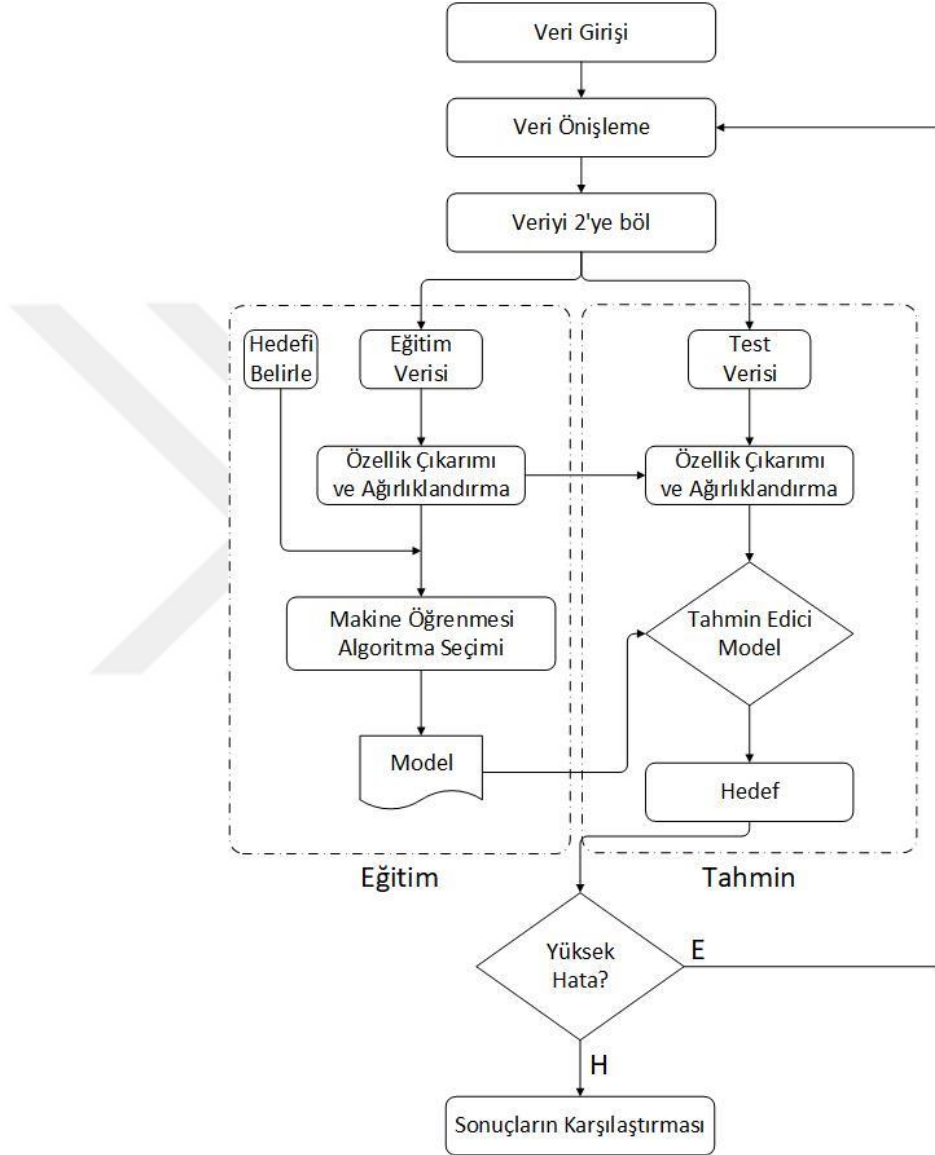
Şekil 5.1. Ön işlemeden sonra cinsiyete göre ameliyat geçirenlerin grafiği

Şekil 5.1'e göre ön işleme sonrasında ameliyat olanlar grubunda 77 (%36) kadın ve 137 (%64) erkek, ameliyat olmayanların grubunda ise 104 (%48,6) kadın ve 110 (%51,4) erkek vardı. Ayrıca son durumda 428 kaydın yarısı kadın, diğer yarısı erkektir.

5.4. Makine Öğrenmesi Modelinin Oluşturulması

Tahmin, makine öğrenme algoritmaları ile yapılmış ve bu algoritmaların doğrulukları karşılaştırılmıştır. Modellerde verilerin bir kısmı eğitim verisi olarak bir kısmı da modelin doğruluğunu test etmek için kullanılmıştır. Önerilen mimari, Şekil 5.2'de

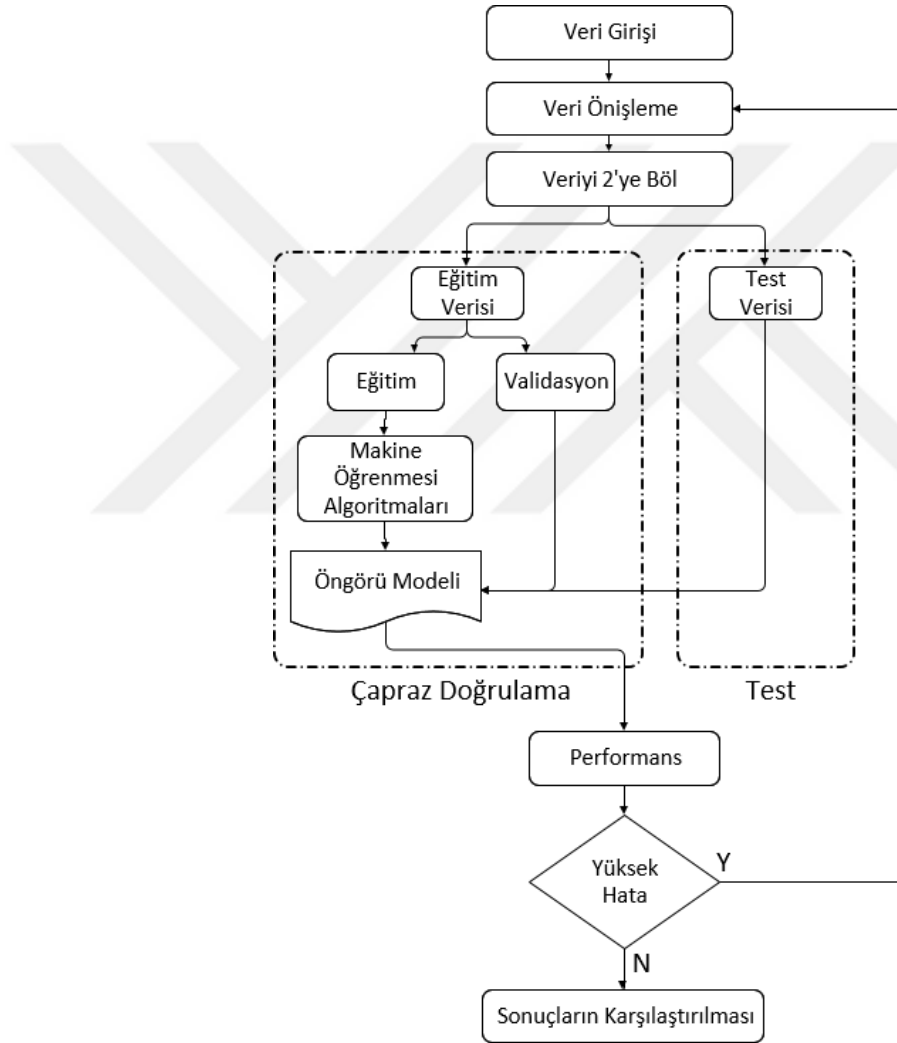
gösterilmiştir. Şekil 5.2'de görülen mimariye göre toplanan veriler ön işlemeye tabi tutulmuştur. Bağımlı değişkeni belirledikten sonra veriler, eğitim ve test verisi olarak iki gruba ayrılmıştır. Verilerin yüzde yetmişi eğitim verisi olarak yüzde otuzu ise modeli test etmek için kullanılmıştır. Son olarak modelin performansı ve apandisit ameliyatı olması gerekenler değerlendirilmiştir.



Şekil 5.2. Önerilen sistemin mimarisi

Çalışmada uygulanan diğer bir mimari, Şekil 5.3'te gösterilmiştir. Şekil 5.3'te görülen mimariye göre toplanan veriler ön işlemeye tabi tutulmuştur. Bağımlı değişkeni belirledikten sonra veriler eğitim ve test verisi olarak iki gruba ayrılmıştır. Verilerin yüzde yetmişi eğitim verisi olarak yüzde otuzu ise modeli test etmek için

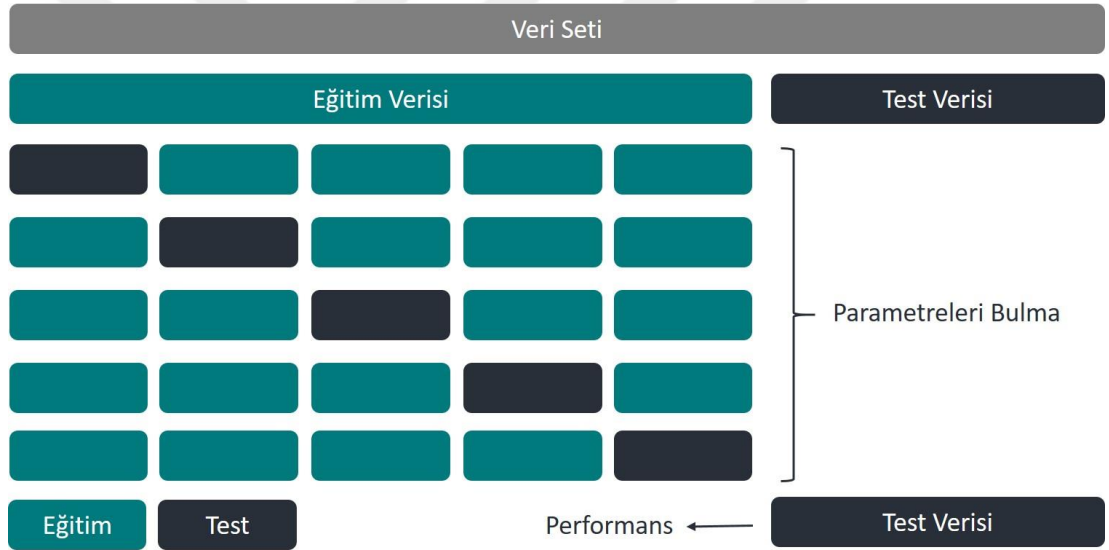
kullanılmıştır. Modelin birden fazla eğitim ve test grubu ile eğitilmesine izin verdiği için genellikle çapraz doğrulama tercih edilir. Gruplardan biri test seti, geri kalanı eğitim seti olarak kullanılır. Her grup, modeli eğitmek için bu şekilde tekrarlanır. Bu, modelin yeni verilerle ne kadar iyi performans göstereceği hakkında daha iyi bir fikir verir ve modelin doğruluğu için gereklidir. Genel olarak, veri kümesi 10 eşit parçaya bölünür. Bunlardan 1 tanesi test için, 9 tanesi eğitim için kullanılmaktadır. Her yinelemede, test sonuçları hesaplanır, ortalaması alınır ve performans elde edilir. Bu çalışmada, veri kümesi 10 eşit parçaya bölünmüştür.



Şekil 5.3. Önerilen sistemin mimarisi (çapraz doğrulama)

Uygulanan her iki mimaride sonuçların birbirine çok yakın olduğu görülmüştür. Bu durum, ilk mimaride bu veri seti için aşırı uyum (over-fitting) probleminin ortaya çıkmadığı şeklinde yorumlanabilir.

Makine öğrenmesinde, modelin gerçek veriler için doğru çalışıp çalışmayacağı konusunda tam olarak bilgi sahibi olunamadığı için çapraz doğrulama tekniği kullanılmaktadır. Çapraz doğrulama bir modelin gerçekte ne kadar doğru performans göstereceğini tahmin etmek için kullanılır. Bir modelin bağımsız test setleri üzerindeki performansının değerlendirilmesi, sonraki veriler üzerindeki performansının ne kadar iyi olacağı hakkında iyi bir fikir verir. Bu yöntemde veri kümesi rastgele k adet gruba ayrılır. Gruplardan biri test seti olarak kullanılırken geri kalan $k-1$ grup eğitim seti olarak kullanılır. Her bir grup sırasıyla test seti olarak tekrarlanır ve model eğitilir. Çapraz doğrulama işlemi k kez tekrarlanır. İterasyonlardan elde edilen k sonuçlarının ortalaması ile tek bir tahmin oluşturulur. K değeri veri setine göre ayarlanabilir. Bu modelde verilerin tamamı eğitim ve test olarak kullanılır ve modelin doğruluğu için oldukça fayda sağlar.



Şekil 5.4. Çapraz doğrulama

5.5. Model Performans Değerlendirmesi

Kurulan modelin analizi yapıldığında en doğru sonucu bulmak için elde edilen ölçümlerin performansları değerlendirilir. Bu değerlerin hesaplanmasında doğru pozitif, yanlış pozitif, doğru negatif ve yanlış negatif değerleri kullanılmaktadır. Şekil 5.5'te karışıklık matrisi yer almaktadır. Buna göre;

Doğru Pozitif (TP): Gerçekte pozitif sınıfına ait olan ve modelin de pozitif tahmini yaptığı durum. Örneğin hasta insanlar hasta olarak doğru tanımlanır (Doğru tahmin).

Yanlış Pozitif (FP): Gerçekte negatif sınıfına ait olan ve modelin pozitif tahmini yaptığı durum. Sağlıklı insanlar hasta olarak yanlış tanımlanır (Tip I hatası).

Doğru Negatif (TN): Gerçekte negatif sınıfına ait olan ve modelin de negatif tahmini yaptığı ve durum. Sağlıklı insanlar sağlıklı olarak doğru tanımlanır (Doğru tahmin).

Yanlış Negatif (FN): Gerçekte pozitif sınıfına ait olan ve modelin negatif tahmin yaptığı durum. Hasta insanlar olarak sağlıklı olarak yanlış tanımlanır (Tip II hatası).

		Gerçek Durum		Toplam
		Durum Pozitif	Durum Negatif	
Tahmin edilen durum	Tahmin edilen durum pozitif	True Positive – TP Doğru Pozitif	False Positive - FP Yanlış Pozitif Tip I hatası	P'
	Tahmin edilen durum negatif	False Negative - FN Yanlış Negatif Tip II hatası	True Negative - TN Doğru Negatif	N'
Toplam		P	N	P+N

Şekil 5.5. Karışıklık matrisi

Çizelge 5.6. Değerlendirme ölçütleri [111, 112]

Ölçüt	Açıklama	Formül
Doğruluk (Accuracy, Recognition rate)	Doğru tahmin edilen örneklerin, tüm örneklerin sayısına oranıdır.	$\frac{TP + TN}{TP + FP + TN + FN} = \frac{TP + TN}{P + N}$
Hata oranı (Error rate, Misclassification rate)	Yanlış tahmin edilen örneklerin, tüm örneklerin sayısına oranıdır. Doğruluk değeri ile toplamı bire eşittir.	$\frac{FP + FN}{TP + FP + TN + FN} = \frac{FP + FN}{P + N}$ veya $1 - \text{doğruluk}$
Duyarlılık (Sensitivity, Recall, True positive rate-TPR)	Doğru tahmin edilen pozitif sınıftaki örneklerin, gerçekte pozitif sınıftaki örneklerin sayısına oranıdır.	$\frac{TP}{TP + FN} = \frac{TP}{P}$

Çizelge 5.6. Değerlendirme ölçütleri (devam)

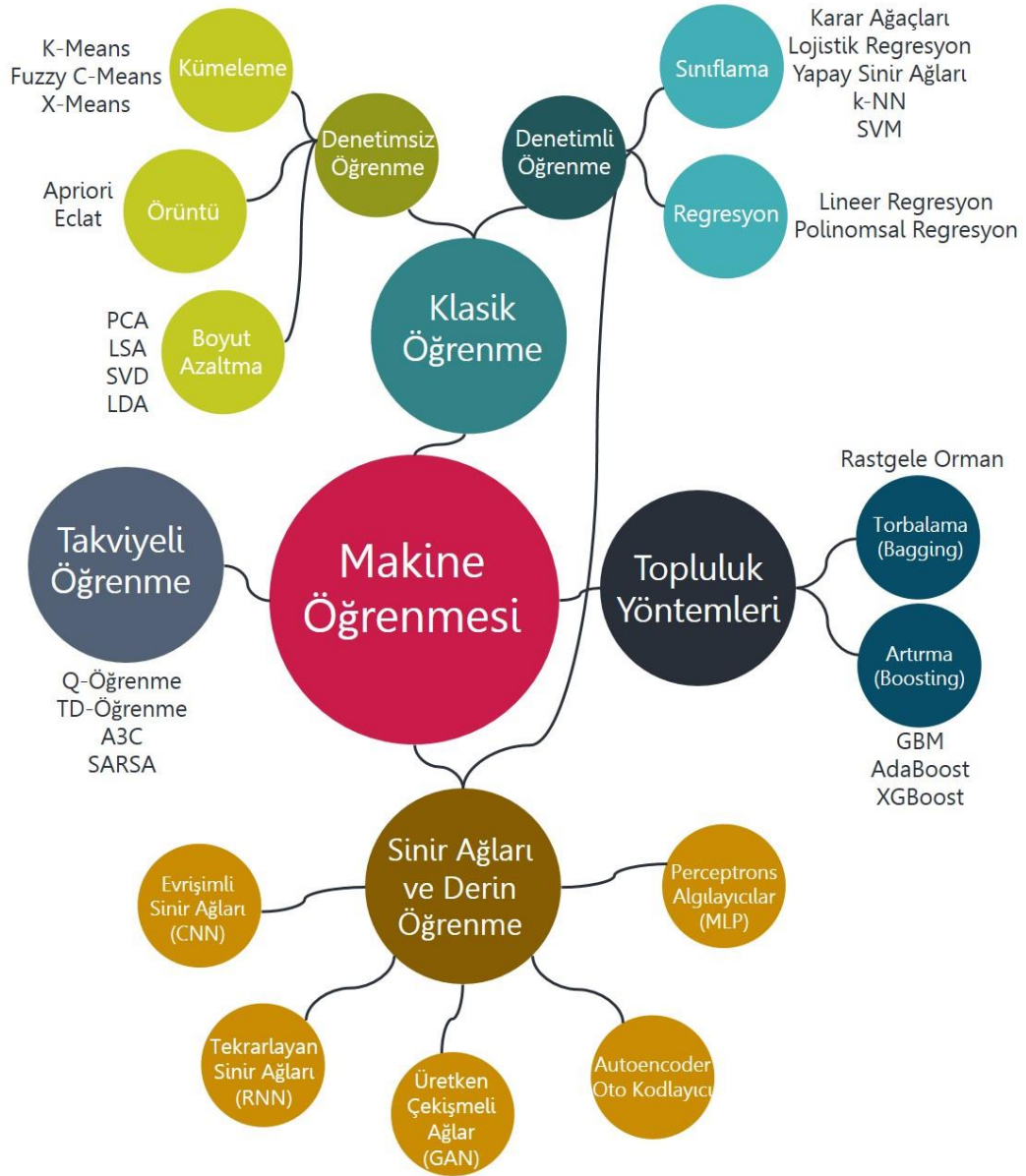
Belirleyicilik (Specificity, True negative rate- TNR)	Doğru tahmin edilen negatif sınıftaki örneklerin, gerçekte negatif sınıftaki örneklerin sayısına oranıdır.	$\frac{TN}{TN + FP} = \frac{TN}{N}$
Kesinlik (Pozitif öngörü değeri, Positive predictive value, precision)	Doğru tahmin edilen pozitif sınıftaki örneklerin, pozitif sınıfta tahmin edilen örneklerin sayısına oranıdır.	$\frac{TP}{TP + FP} = \frac{TP}{P'}$
Kesinlik (Negatif öngörü değeri Negative predictive value)	Doğru tahmin edilen negatif sınıftaki örneklerin, negatif sınıfta tahmin edilen örneklerin sayısına oranıdır.	$\frac{TN}{TN + FN} = \frac{TN}{N'}$
Yanlış pozitif oranı (False positive rate- FPR)	Yanlış tahmin edilen negatif sınıftaki örneklerin, negatif sınıftaki örneklerin sayısına oranıdır.	$\frac{FP}{FP + FN} = \frac{FP}{N}$ veya $1 - \text{belirleyicilik}$
Yanlış negatif oranı (False negative rate- FNR)	Yanlış tahmin edilen pozitif sınıftaki örneklerin, pozitif sınıftaki örneklerin sayısına oranıdır.	$\frac{FN}{FN + TP} = \frac{FN}{P}$ veya $1 - \text{duyarlılık}$
F-ölçütü (F-score)	Duyarlılık ve kesinliğin birlikte değerlendirildiği ölçüttür.	$\frac{2 \times \text{Kesinlik} \times \text{Duyarlılık}}{\text{Kesinlik} + \text{Duyarlılık}}$

Çizelge 5.6'da değerlendirme ölçütlerini içeren tablo, ölçütlerin açıklaması ve formülleri görülmektedir.

5.6. Makine Öğrenmesi Algoritmaları

Makine öğrenmesi söz konusu olduğunda akla birçok algoritma gelmektedir. Bu başlık altında çalışmada kullanılan algoritmaların yapısı gösterilmiştir. Bir algoritmanın diğerinden üstün ya da zayıf olduğunu söylemek mümkün değildir. Sonuçlar, veri setinin boyutu, yapısı gibi durumlara göre değişiklik gösterebilir. Bu nedenle verinin ve problemin durumuna göre farklı algoritmalar denenmeli ve en doğru makine öğrenmesi algoritmasına başvurulmalıdır. Bu çalışmada verinin ve problemin yapısına uygun düşen 7 algoritma kullanılmıştır. Bunlar, lojistik regresyon, yapay sinir ağları, k en yakın komşu, destek vektör makineleri, sınıflandırma ve regresyon ağaçları, rastgele orman ve gradyan artırılmış ağaçlar algoritmalarıdır. En başarılı algoritma

gradyan artırılmış ağaçlarıdır. Şekil 5.6’da makine öğrenmesi haritasına görülmektedir.



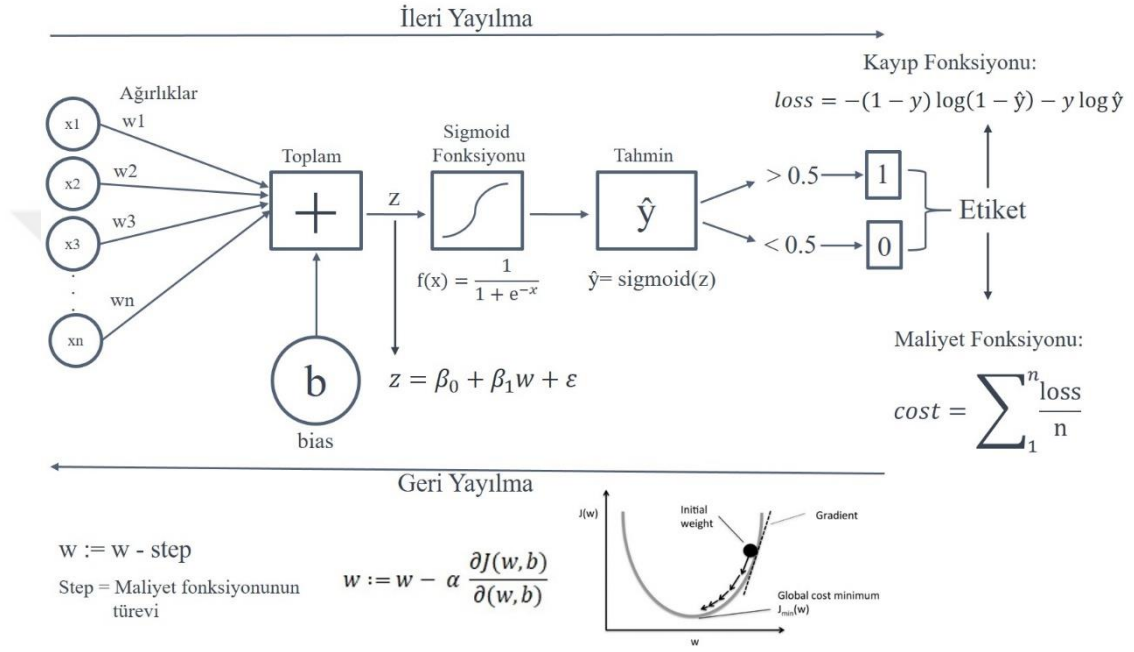
Şekil 5.6. Makine öğrenmesi haritası

5.6.1. Lojistik Regresyon

Regresyon analizi, tahmin amaçlı kullanılan ve bağımlı değişken ile bağımsız değişkenler arasındaki ilişkinin fonksiyonel şeklini belirleyen bir yöntemdir [113]. Lojistik Regresyon, bir sonucu belirleyen ve bağımsız değişkenleri bulunan bir veri

setini analiz etmek amacıyla kullanılan istatistiksel bir yöntemdir. Regresyon olarak adlandırılmasına rağmen burada bir sınıflandırma söz konusudur.

Son zamanlarda lojistik regresyon, kullanım kolaylığı ile beraber sayısal verilerin rahat yorumlanabilir olması nedeniyle ön plana çıkmış ve yoğun bir şekilde kullanılan bir yöntem haline gelmiştir. Genellikle tıp, biyoloji, ekonomi gibi alanlarda kullanılmaktadır [114].



Şekil 5.7. Lojistik regresyon

Lineer regresyonun modifikasyonu ile ortaya çıkan lojistik regresyonda, sigmoid fonksiyonu 0 ile 1 arasında bir olasılık çıkarır.

Sigmoid Fonksiyonu:

$$f(x) = \frac{1}{1 + e^{-x}} \quad (5.1)$$

Doğrusal regresyon modelinin formülü:

$$z = \beta_0 + \beta_1 w + \varepsilon \quad (5.2)$$

Lojistik regresyonun yapısı Şekil 5.7’de görülmektedir. Veri setindeki her bir değişkene ait değerler başlangıç ağırlıklarıyla çarpılmaktadır ve bias ile toplanır. Böylece z değeri elde edilir. Literatürde ağırlıkların başlangıç değerleri genellikle 0,01 olarak alınır. Bias için bu değer 0’dır. Elde edilen z değeri sigmoid fonksiyonuyla 0 ile 1 arasındaki sonucu vermektedir. Bu aşamaya kadar olan kısım ileri yayılmayı ifade eder. Başlangıcı rastgele belirlenen ağırlık ve bias değerleri neticesinde sonuçlar yanlış üretilirse yani maliyet (cost) değeri büyükse geriye doğru gidilir. Bu kısım geri yayılma olarak adlandırılır. Maliyet fonksiyonunun türevi alınır ve ağırlıklar ile bias güncellenir. Bu işlem cost değeri minimum olana kadar tekrarlanır.

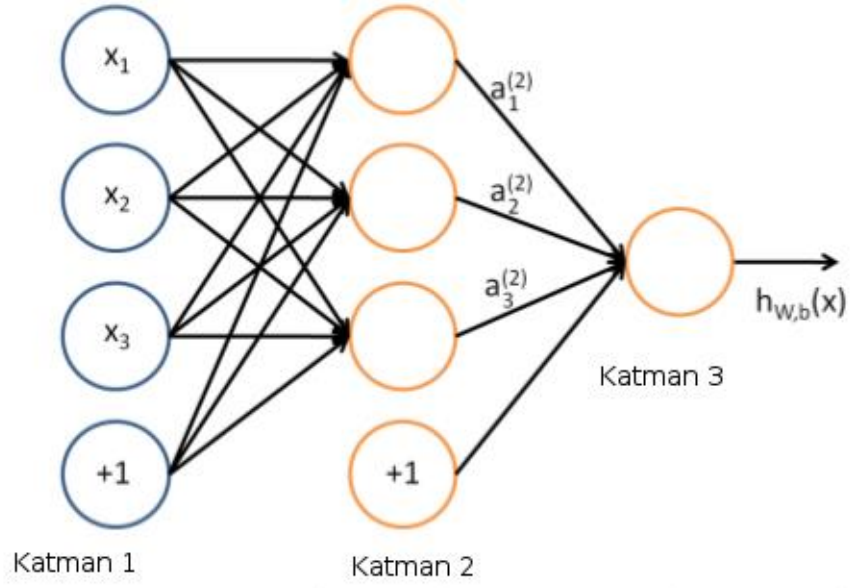
5.6.2. Yapay Sinir Ağları

Yapay sinir ağları (YSA), insan beyninin çalışma prensibini benzetim yoluyla taklit ederek analiz yapan, bu verilerden yeni bilgiler oluşturan bir yöntemdir. YSA’nın genel avantajları aşağıdaki gibi sıralanabilir [115].

- Lineer olmayan yapıda olması
- Girdi ve çıktı eşleştirmeleri ile tasarlanabilmesi
- Adapte olabilmesi
- Hataya karşı toleranslı olması

YSA, bilgisayarların öğrenmesine yönelik çalışmaları kapsayan, yapay zekâ biliminin altında olan ve araştırmacıların da yoğun ilgilerinin olduğu bir alanıdır [116]. Basit bir tanımla adından da anlaşılacağı gibi yapay sinir ağlarının biyolojik sinir ağını taklit eden bir program olduğu söylenebilir. YSA, ağırlıklandırılmış şekilde birbirlerine bağlanmış olan birçok yapay nöronun oluşturduğu matematiksel bir modeldir. Bir sinir ağı birbirine bağlı bir yapay nöron grubundan oluşur.

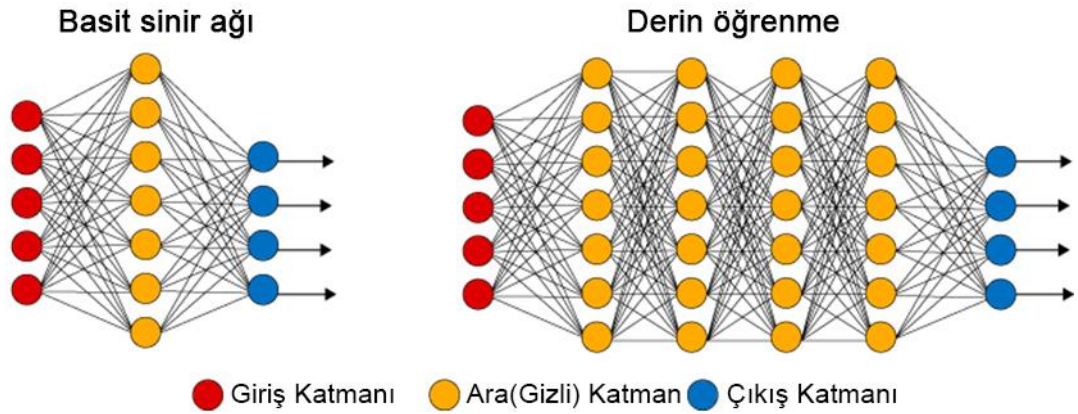
Şekil 5.8’de basit bir yapay sinir ağı yapısı görülmektedir. Görüldüğü gibi çok katmanlı durumlarda bir katmanın çıktısı bir diğer katmanın girdisi haline gelmektedir. YSA’da öğrenme, ağa gösterilen veriler için doğru çıktıları üretebilen optimum ağırlık değerlerinin bulunmasıyla gerçekleşir. En doğru ağırlık değerlerine ulaşıldığında verilerin temsil ettiği olay hakkında genellemeler yapılabilme yeteneğine kavuşulur. Verinin az olması, YSA’nın başarısında olumsuz etkiye neden olabilir.



Şekil 5.8. Basit yapay sinir ağı yapısı

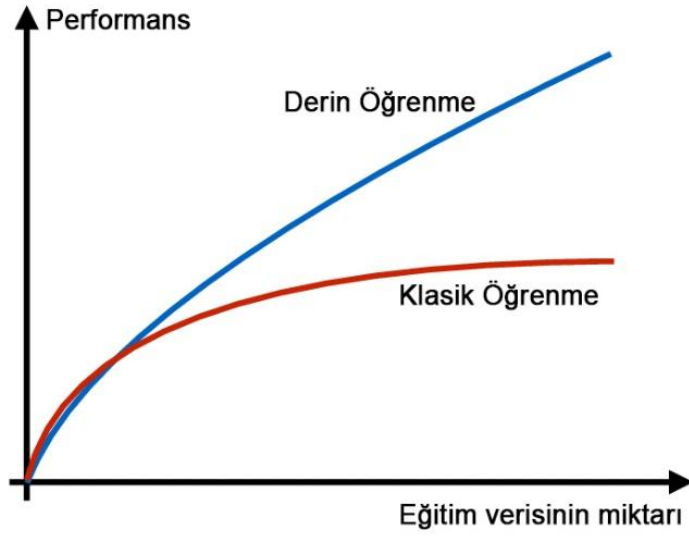
5.6.2.1. Derin Öğrenme

Bir makine öğrenmesi sınıfı olan derin öğrenme, özellik çıkarma ve dönüştürme için birçok doğrusal olmayan işlem birimi katmanını kullanmaktadır. Şekil 5.9’da yapay sinir ağı ve derin öğrenme diyagramı görülmektedir. Her ardışık katman, bir önceki katmandaki çıktıyı girdi olarak almaktadır [117].



Şekil 5.9. Yapay sinir ağı ve derin öğrenme

Derin öğrenmede yapay sinir ağları kullanılır ve görüntü kümelemede doğru sonuçlar elde etmek açısından mükemmel sonuçlar vermektedir.



Şekil 5.10. Veri miktarına göre derin öğrenme ve klasik öğrenme

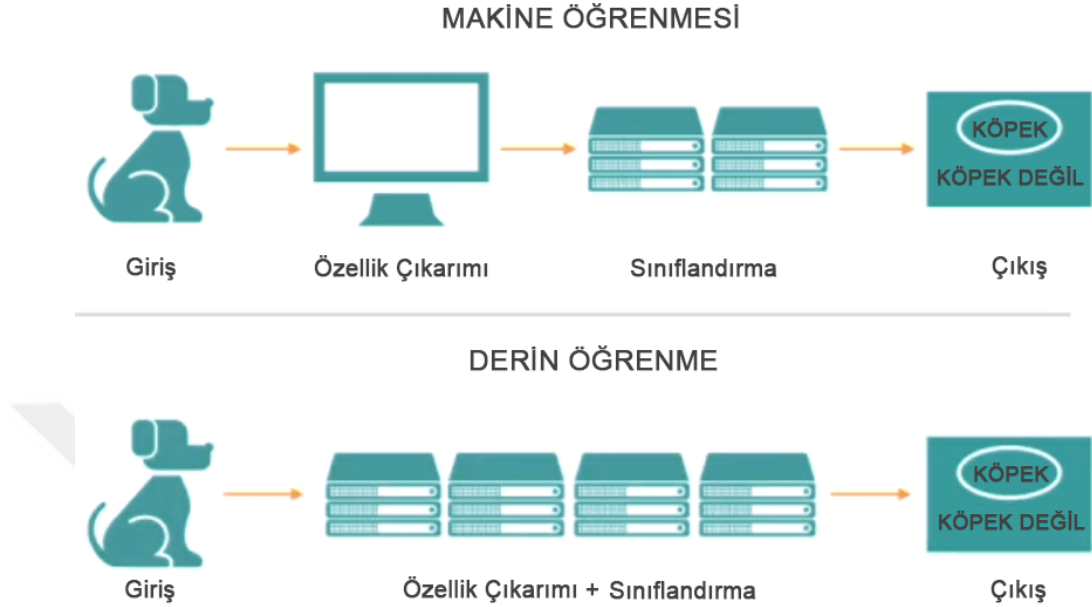
Son yıllardaki veri miktarının hızla artması ve bu artış karşısında makine öğrenme yöntemlerinin performans açısından yetersiz kalması derin öğrenmeye olan ihtiyacı artırmıştır. Şekil 5.10'daki grafiğe göre belli bir yere kadar her iki yöntemde performans açısından çok fark olmasa da veri miktarının çok fazla artması durumunda derin öğrenme öne geçmektedir.

Günümüzde kamera sayılarının artmasıyla beraber dijital veride çok büyük miktarda bir birikim olduğu söylenebilir. Veri miktarının artması ve donanımdaki gelişmeler neticesinde derin öğrenme giderek yaygınlaşmaktadır. Derin öğrenme, yüz tanıma, ses tanıma, parmak izi okuyucular, plaka okuyucular, sürücüsüz arabalar ve güvenlik kameraları gibi alanlarda da kullanılmaktadır.

İnsanlar bir hayvan görüntüsü tanımada, işaretlerin okumasında veya bir insanın yüzünün tanınmasında bir çaba göstermezler. Beynimiz görüntüleri kolay bir şekilde anlayabilirken bilgisayarlar için çözülmesi zor problemlerdir. Son birkaç yılda, bu zor problemleri ele alma konusunda muazzam bir ilerleme kaydedilmiştir. Özellikle, derin bir konvolüsyonel nöral ağ olarak adlandırılan bir tür model ile görsel tanıma görevleri için makul seviyede bir performans sağlanmıştır. Nesnelerin görüntülerinin bilgisayar tarafından anlamlı hale getirilmesi büyük kolaylıklar sağlamaktadır.

Makine öğrenmesinde özellikler makineye manuel olarak verilir. Ancak derin öğrenmede özellikler doğrudan veriden çıkarılır. Bu durum Şekil 5.11'de

gösterilmiştir. Şekle göre özellik çıkarımı ve sınıflandırma, derin öğrenme yönteminde model tarafından yapılır. Her derin öğrenme bir makine öğrenmesi iken her makine öğrenmesi bir derin öğrenme değildir.



Şekil 5.11. Makine öğrenmesi ve derin öğrenmede özellik çıkarımı

Derin öğrenme modelinin ideal hale gelmesi ve özelliklerin model tarafından çıkarılması için çok fazla veriye ihtiyaç duyulmaktadır. Makine öğrenmesinde ise değerlendirilecek özellikler bir kişi tarafından algoritmaya verildiği için çok daha az veri ile problemin çözümü mümkündür. Burada veri miktarının artması donanım yetersizliği gibi bir sonucu ortaya çıkarsa da bu problem içinde çeşitli çözüm yolları ortaya konmaktadır. Sonuç olarak, makine öğrenimi ve derin öğrenmenin birbirine göre üstün ve zayıf yönleri olduğu söylenebilir. Kısaca problemin durumu ve veri miktarı hangi yöntemin daha uygun olduğu konusunda bize fikir vermelidir.

5.6.3. K-NN (K En Yakın Komşu)

K En Yakın Komşu (K-NN) sınıflama yapmak amacıyla kullanılan makine öğrenmesi yöntemlerinden birisidir. K-NN, denetimli sınıflandırma yöntemleri arasındadır.

K değeri sınıflandırmadaki eleman sayısını belirler. Tahmin yapılırken veri setindeki en yakın komşuları arar. Uzaklık hesaplamalarında genellikle Öklid olmak üzere

Manhattan, Minkowski ve Hamming fonksiyonları da kullanılabilir. Örnek örüntü vektörü, kendisine en çok benzeyen örüntünün sınıfından kabul edilir.

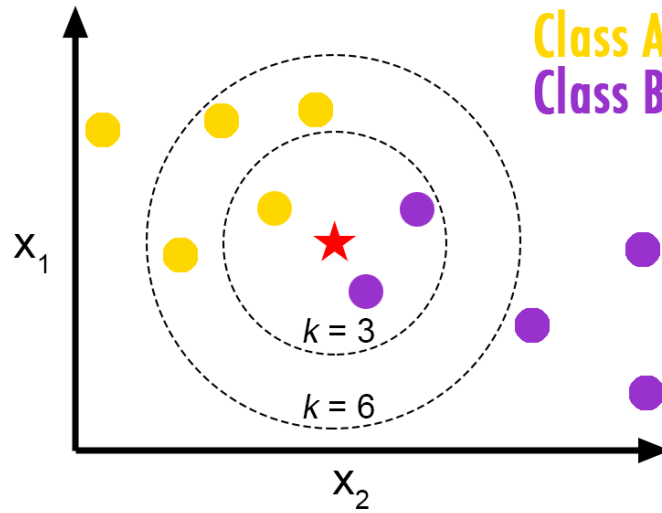
Öklid uzaklığı fonksiyonu;

$$d(p, q) = d(q, p) = \sqrt{(q_1 - p_1)^2 + (q_2 - p_2)^2 + \dots + (q_n - p_n)^2} \quad (5.3)$$

$$d(p, q) = d(q, p) = \sqrt{\sum_{i=1}^n (q_i - p_i)^2} \quad (5.4)$$

KNN algoritmasının adımları şu şekildedir [118];

- İlk olarak K değeri belirlenir.
- Kullanılan uzaklık fonksiyonuna göre diğer nesnelerden hedef nesneye olan mesafeler hesaplanır.
- Uzaklıklar sıralanır ve en küçük uzaklığa göre en yakın komşuluklar bulunur.
- En yakın komşuların y kategorisi değerleri toplanır.
- En uygun komşu kategorisi seçilir.



Şekil 5.12. K En Yakın Komşu sınıflandırması

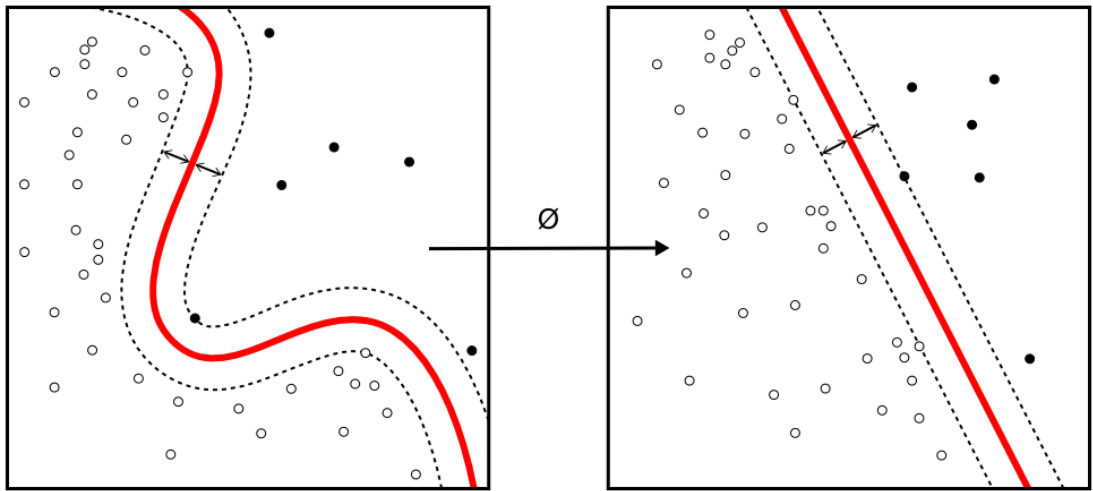
Şekil 5.12’de k=3 için ve k=6 için sınıflandırma durumu görülmektedir. En yakın komşu sayısının seçimi K-NN sınıflandırmasında çok önemlidir.

5.6.4. Destek Vektör Makineleri

Destek Vektör Makineleri (DVM) sınıflama yapmak amacıyla kullanılan makine öğrenmesi yöntemlerinden birisidir. DVM, denetimli sınıflandırma yöntemleri arasındadır.

Sınıflandırma problemlerinde yaygın olarak kullanılan DVM'nin doğruluk oranının yüksek olması, bu yöntemin kullanımını yaygınlaştırmıştır. Bu yöntemin diğerlerinden farkı, işlem sayısı ve algoritma karışıklığının az olmasıdır [119]. DVM, doğrusal ya da doğrusal olmayan şekilde olmak üzere ikiye ayrılır [120]. 1963'te Vapnik tarafından önerilen algoritma doğrusal bir sınıflandırıcı modeliydi [121]. Bununla birlikte, 1992'de Bernhard E. Boser, Isabelle M. Guyon ve Vladimir N. Vapnik doğrusal olmayan sınıflandırıcıları oluşturmanın bir yolunu önerdiler [122].

Doğrusal durumlarda, eğitim verisi ile elde edilen bir karar fonksiyonu yardımıyla sınıfların birbirinden ayrılması söz konusudur. Veri setini ikiye bölen doğruya karar doğrusu denir. Bu modeldeki temel amaç, sınıfları en uygun şekilde birbirinden ayıracak hiper düzlemi tespit etmektir. Doğrusal olmayan bir veri kümesi söz konusu olduğunda DVM'ler doğrusal bir hiper düzlem çizemediği için çekirdek (kernel) yöntemi kullanılır. Çekirdek yöntemi, doğrusal olmayan veriler için sınıflandırma doğruluğunu yüksek oranda arttırmaktadır. Şekil 5.13'te doğrusal ve doğrusal olmayan iki ayrı veri seti görülmektedir.



Şekil 5.13. Doğrusal ve doğrusal olmayan iki ayrı veri seti [123]

5.6.5. Sınıflandırma ve Regresyon Ağaçları (CART)

CART (Classification and Regression Trees) Sınıflandırma ve Regresyon Ağaçları olarak bilinir ve temelleri 1984 yılında Breiman, Freidman, Olshen ve Stone tarafından yazılan kitaba dayanmaktadır. CART algoritması sürekli ve kategorik verilerle ağaç modelinde ayırma kriterini hesaplama sırasında eksik gözlemleri önemsemeyen bir algoritmadır. Bağımlı değişken kategorik ise “sınıflandırma ağaçları”, sürekli yapıda ise “regresyon ağaçları” adını almaktadır [124]. Sınıflama kategorik değerlerin tahmininde, regresyon ise süreklilik gösteren değerlerin tahmininde kullanılır. Tahmin edilen bağımlı değişken kategorik veya süreklilik gösteren bir değere sahip olabilir.

Bu yöntemde kullanılan algoritma bağımsız değişkenlerin bağımlı değişkene olan etkisini incelemenin yanı sıra model içerisindeki etkileşimi de incelemektedir. Böylece doğrusal olmayan ilişkilerin açıklanması da mümkün hale gelmektedir.

Sınıflandırma ve regresyon ağaç yapılarında her düğümden meydana gelen iki dal yapısı mevcuttur. Her dal için yeni bölünmelerin meydana gelmesinde her düğüme belirlenmiş bir kriter uygulanır. Bu kriterlerin belirlenmesinde değişkenlerin sahip oldukları nitelikler etkilidir [125].

CART, bölme noktaları oluşturmak için Gini ya da Twoing algoritmasını kullanır. Gini yönteminde amaç her adımda en büyük veri kümesini elde etmek ve en iyi bölmeyi sağlamaktır. Gini yönteminde her bir nitelik ile ilgili olarak sol ve sağ taraftaki bölünmeler için G_{inisol} ve G_{inisag} değerleri olmak üzere;

- k : Sınıf sayısı,
- T : Bir düğümdeki örnekler,
- T_{sol} : Sol düğümdeki örnek sayısı,
- $T_{sağ}$: Sağ düğümdeki örnek sayısı,
- L_i : Sol düğümde i kategorisindeki örnek sayısı,
- R_i : Sağ düğümde i kategorisindeki örnek sayısı olmak üzere;

$$Gini_{sol} = 1 - \sum_{i=1}^k \left(\frac{L_i}{T_{sol}} \right)^2 \quad (5.5)$$

$$Gini_{sağ} = 1 - \sum_{i=1}^k \left(\frac{R_i}{T_{sağ}} \right)^2, \quad (5.6)$$

şeklinde hesaplanır. Her j niteliği için eğitim verisindeki satır sayısı n olmak üzere Gini indeksi ise;

$$Gini_j = \frac{1}{n} (|T_{sol}| Gini_{sol} + |T_{sağ}| Gini_{sağ}) \quad (5.7)$$

formülüne göre hesaplanır [126].

Her defasında ana ve yavru düğümlerin %50' sini içermeye çalıştığı için Twoing algoritması Gini' ye göre daha dengeli bir yapı sunar. Bu yüzden yavaşlama olmaktadır.

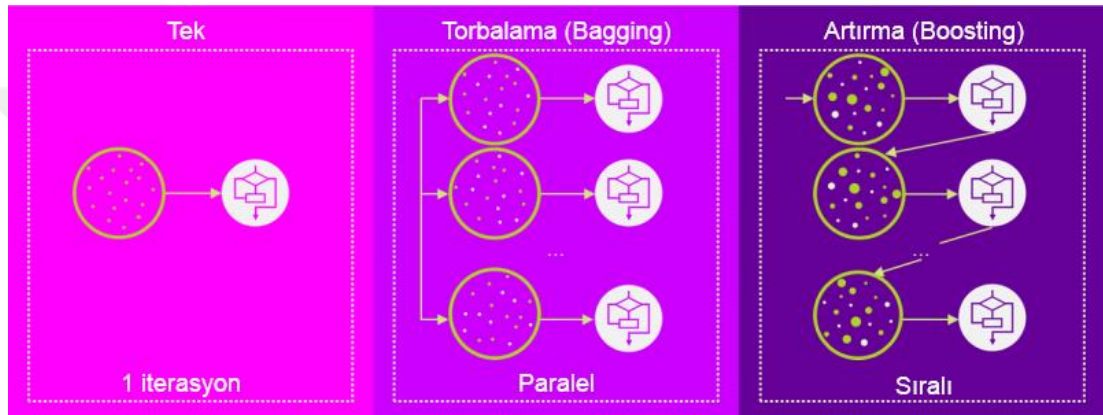
CART algoritmasında bağımlı değişken nominal ordinal veya sürekli olabilir. Belli değişkenlerin değerlerine göre oluşturulan kuralların toplanması ile model oluşur.

- Seçilen bir kuralın düğümleri ikiye bölmesindeki mantık tekrarlanarak alt düğümlere uygulanır.
- Eğer daha fazla kazanç elde edilecekse durma kurallarına göre dallanma durur.
- Her ağaç terminal bir düğümde biter.
- Her gözlem yalnızca bir terminal düğüme düşer.
- Her terminal düğümü bir kural seti ile diğerlerinden ayrı olarak tanımlanmıştır [127].

5.6.6. Rastgele Orman

Rastgele orman algoritması Karar ağaçları ve Torbalama yöntemlerinin bir arada kullanılması ilkesine dayanır ve Topluluk (Ensemble) yöntemlerine girer [128]. Topluluk yöntemleri, bir öğrenme algoritmasını kullanarak birden fazla modeli eğiten bir makine öğrenmesi kavramıdır. Bu yöntemlerden olan Torbalama (Bagging) veya Artırmayı (Boosting) kullanmak için bir temel öğrenici algoritma seçilmelidir.

Torbalama yönteminde özellikler farklı şekillerde birleştirilerek yeni ağaçlar oluşturulur ve oluşturulan ağaçlardan en popüler olan sınıf seçilir. Bu yöntemde rastgele özellik kullanımı doğruluğun artmasına katkı sağlar [128]. Torbalama yönteminde herhangi bir elemanın yeni bir veri kümesinde görünme olasılığı aynıdır. Ancak, Artırmada gözlemler ağırlıklandırılır ve bu nedenle bazıları yeni setlerde daha sık yer almaktadır. Eğitim aşaması torbalama için paraleldir. Yani her model bağımsız olarak inşa edilir. Artırma, tahminlerin bağımsız olarak değil, sırayla yapıldığı bir topluluk yöntemidir. Bir başka deyişle, sonraki tahminler öncekilerin hatalarından öğrenir [129].



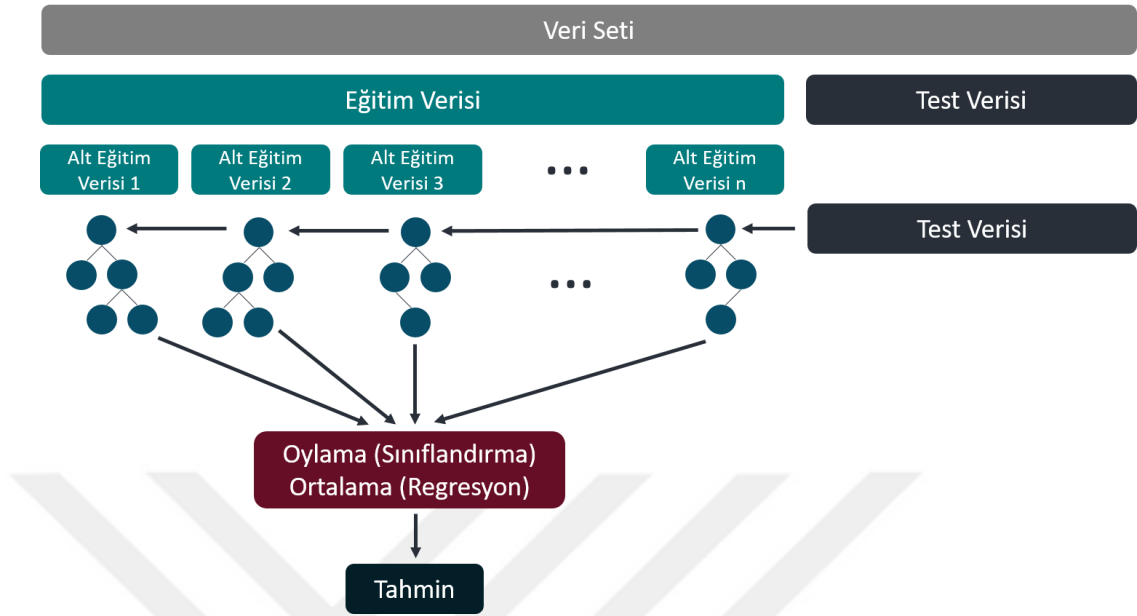
Şekil 5.14. Torbalama ve Artırma yöntemleri arasındaki fark [129]

Şekil 5.14'te Torbalama ve Artırma yöntemleri arasındaki fark görülmektedir. Genel bir prensip olarak, Artırma algoritmaları, her bir yinelemede elde edilen zayıf öğreniciyi belirli kurallar altında birleştirerek güçlü bir öğrenici elde etmeye çalışır.

Rastgele orman, regresyon veya sınıflandırma problemleri için kullanılan esnek bir makine öğrenmesi yöntemidir. En basit şekliyle tanımlamak gerekirse rastgele orman, oluşturulan çok sayıda karar ağacının daha doğru bir tahmin elde etmek için birleştirilmesidir.

Rastgele orman, karar ağaçlarında çok sık meydana gelen aşırı uyum (over-fitting) problemini çözmek için hem veri setinden hem de öznitelik setinden rastgele çok sayıda farklı alt setler seçerek eğitim yapar. Böylece çok sayıda oluşan karar ağacından her biri bir tahminde bulunur. Regresyon probleminde karar ağaçlarının tahminlerinin ortalaması alınırken sınıflandırma probleminde en çok oy alan tahmin seçilir. Bu

modelde farklı veri setleri üzerinde eğitim gerçekleştirildiği için aşırı uyum azalır. Ayrıca, veri setindeki özellikler arasından en önemli olanları tanımlamak için kullanılabilir.



Şekil 5.15. Rastgele orman diyagramı

5.6.7. Gradyan Artırılmış Ağaçlar

Gradyan artırılmış ağaçlar, regresyon veya sınıflandırma problemleri için kullanılan ve topluluk yöntemlerinden biri olan denetimli bir makine öğrenmesi tekniğidir. Her ağaç daha önce yetiştirilmiş ağaçlardan alınan bilgiler kullanılarak yetiştirilir. Regresyon problemi için gradyan artırmanın temel algoritması, x 'in özellikleri temsil ettiği ve y 'nin yanıtı temsil ettiği üzere aşağıdaki gibi genelleştirilebilir:

1. Gradyan artırma, ilk yinelemede tahminler üreten bir f_1 fonksiyonu oluşturur.

$$f_1(x) = y, \quad (5.8)$$

2. Tahminler ve hedef değer arasındaki farkı hesaplar ve bu farklılıklar için h_1 fonksiyonunu oluşturur.

$$h_1(x) = y - f_1(x), \quad (5.9)$$

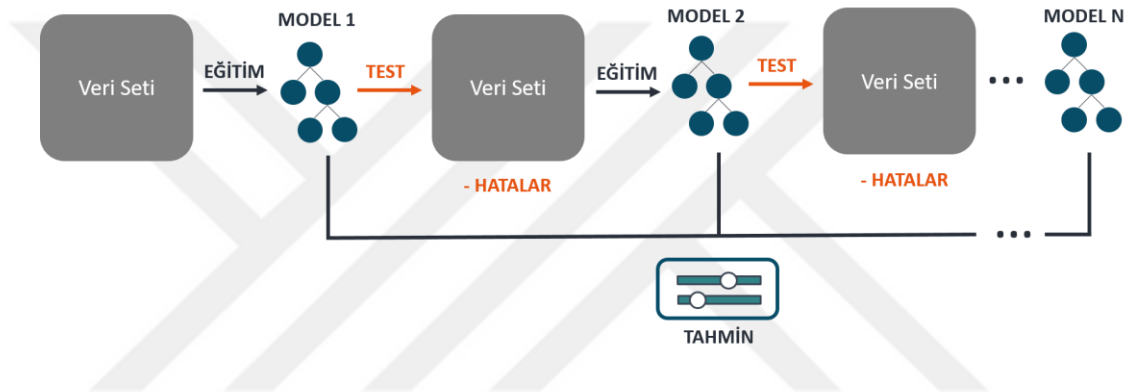
3. Yeni bir ağaç oluşturur.

$$f_2(x) = f_1(x) + h_1(x), \quad (5.10)$$

4. Tahminler ve hedefler arasındaki farkı yeniden hesaplar.

$$h_2(x) = y - f_2(x), \quad (5.11)$$

Bu şekilde, sürekli olarak “f” fonksiyonunun başarısını arttırmaya ve tahminler ile hedefler arasındaki farkı sıfıra indirmeye çalışır.



Şekil 5.16. Gradyan artırma yöntemi

Şekil 5.16’da Gradyan artırma yönteminin çalışma prensibi görülmektedir. Temel fikir, hatayı minimize etmek ve bir sonraki model için hedef çıktıları belirlemektir. Bu teknik, önceki tahmin hatalarından ders alarak sonraki tahminlerin ilerlemesine dayanır.

Gradyan iniş, çok çeşitli sorunlara en uygun çözümleri bulabilen çok genel bir optimizasyon algoritmasıdır. Tahminler, kayıp fonksiyonu (MSE) minimum olacak ve tahmin edilen değerler gerçek değerlere yakın olacak şekilde güncellenir. Gradyan inişinin genel fikri, bir maliyet fonksiyonunu en aza indirmek için parametreleri tekrar tekrar ayarlamaktır.

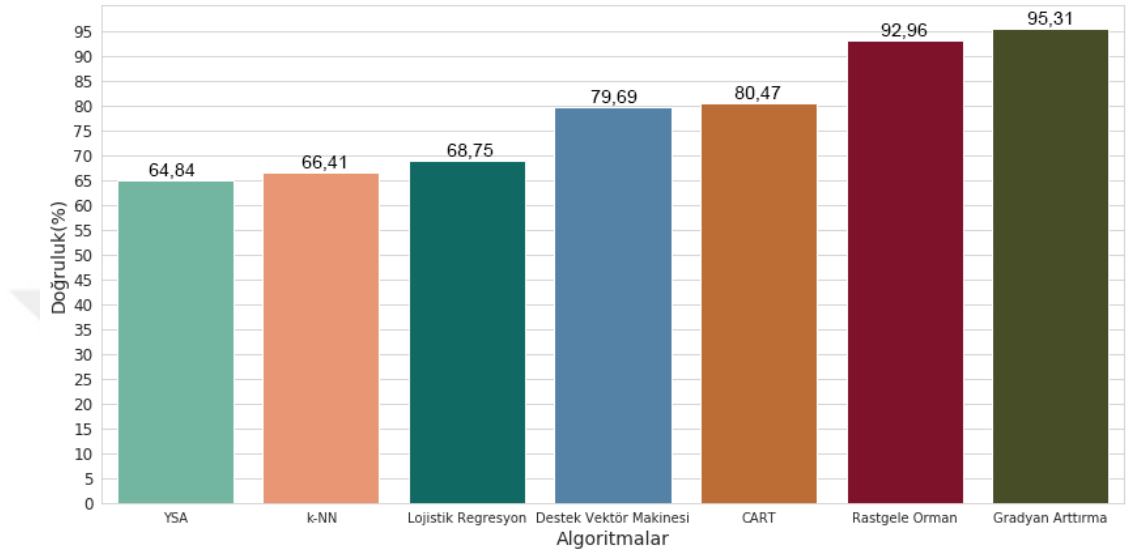
$$J(\theta) = \frac{1}{n} \sum_{i=1}^n (h_{\theta}(x^{(i)}) - y^{(i)})^2 \quad (5.12)$$

n	Eğitim örneklerinin sayısı
$x^{(i)}$	Eğitim örnekleri için girdi vektörü
$y^{(i)}$	i. eğitim örnekleri için sınıf etiketi
θ	Seçilen parametre değerleri veya ağırlıklar
$h_{\theta}(x^{(i)})$	Algoritmanın, θ parametrelerini kullanan i. eğitim örnekleri için öngörüsü.

Her yinelemede, kayıp fonksiyonunun kalıntısı gradyan iniş yöntemi kullanılarak hesaplanır ve bir sonraki yinelemenin hedef değeri haline gelir [130]. Özetle, gradyan artırılmış ağaçlar, her zayıf sınıflandırıcının türetilabilir kayıp fonksiyon değerini en aza indirmek için gradyan iniş yöntemini kullanır.

6. ARAŞTIRMA BULGULARI

Bu bölümde araştırma sürecinde uygulanan tüm algoritmaların doğruluk oranları yer almaktadır. Şekil 6.1’de araştırmada kullanılan 7 makine öğrenme algoritmasının doğruluk yüzdeleri görülmektedir.



Şekil 6.1. Algoritmaların doğruluk yüzdeleri

Şekil 6.1’e göre en başarılı algoritma gradyan artırılmış ağaçları olarak bulunmuştur. 128 klinik kayıttan 122’si doğru olarak tahmin edilmiştir. Böylece, modelin öngörme doğruluğu %95,31 olmuştur. Onu %92,96 ile rastgele orman algoritması izlemektedir. Çizelge 6.1’den 6.8’e kadar kullanılan algoritmaların karışıklık matrisi verilmiştir.

Bu tablolarda bulunan TP, FP, TN, FN kısaltmalarının anlamları şöyledir;

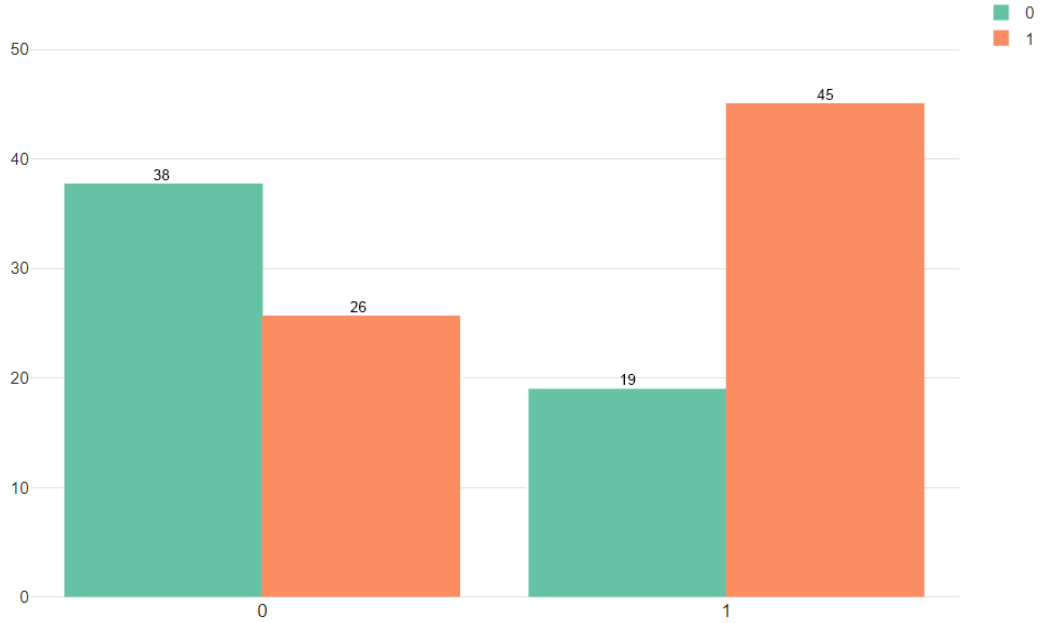
- TP (Doğru Karar): Gerçek durumda, ameliyat edilen hastaların doğru tahmin edildiği anlamına gelir.
- FP (Tip I Hatası): Gerçek durumda ameliyat edilmeyen hastalar ameliyat edilmiş olarak bulunmuştur.
- TN (Doğru Karar): Gerçek durumda ameliyat geçirmeyen hastaların doğru tahmini anlamına gelir.
- FN (Tip II Hatası): Gerçek durumda ameliyat edilen hastalar ameliyat edilmedi olarak tahmin edilmiştir.

Çizelge 6.1. Yapay sinir ağları algoritmasına göre karışıklık matrisi

Doğruluk: %64,84	Gerçek 1	Gerçek 0	Toplam	Precision (Kesinlik)
Tahmin 1	45 (TP) Doğru Karar	26 (FP) Tip I Hatası	71 (P')	%63,38
Tahmin 0	19 (FN) Tip II Hatası	38 (TN) Doğru Karar	57 (N')	%66,67
Toplam	64 (P)	64 (N)	128 (P+N)	
Recall (Duyarlılık)	%70,31	%59,38		

Çizelge 6.1'e göre doğru tahmin edilen örneklerin tüm örnek sayısına oranı %64,84'dur. Bu oran toplam doğruluğu temsil eder. TP doğru pozitifler, TN doğru negatifler, FN yanlış negatifler ve FP yanlış pozitifler olarak temsil edilir.

Şekil 6.2, yapay sinir ağları algoritmasına göre tahmin sonuçlarını göstermektedir. Ameliyat olmayanlar "0", ameliyat olanlar ise "1" olarak ifade edilmiştir. Buna göre, ameliyatsız 64 hastanın 38'i doğru olarak tahmin edilmiştir. Ayrıca ameliyat edilen 64 hastanın 45'i doğru olarak tahmin edilmiştir.



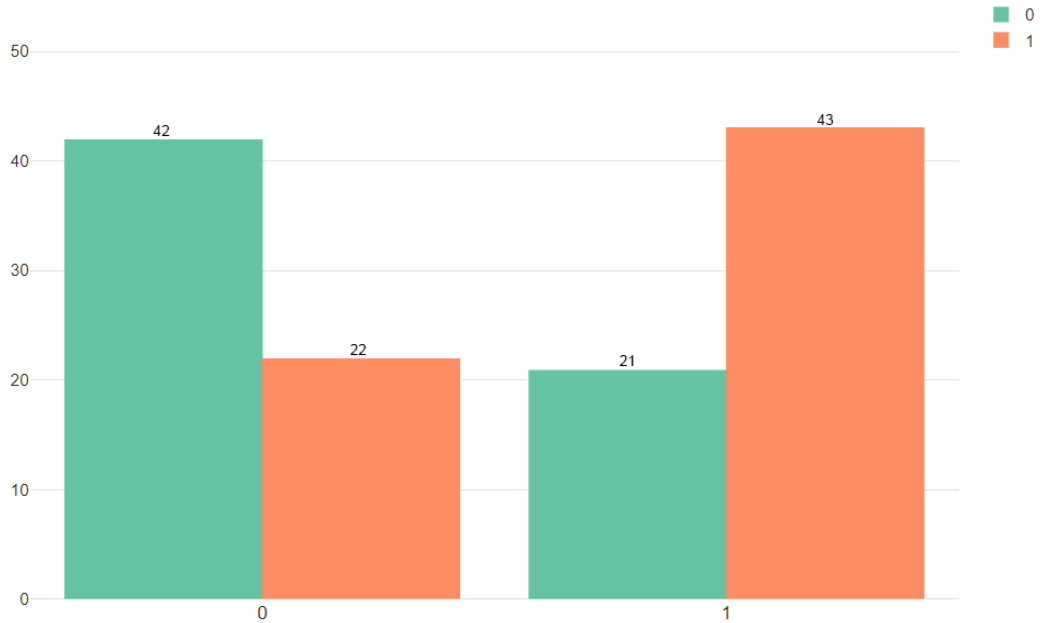
Şekil 6.2. Yapay sinir ağları algoritmasına göre tahmin sayıları

Çizelge 6.2. K-NN algoritmasına göre karışıklık matrisi

Doğruluk: %66,41	Gerçek 1	Gerçek 0	Toplam	Precision (Kesinlik)
Tahmin 1	43 (TP) Doğru Karar	22 (FP) Tip I Hatası	65 (P')	%66,15
Tahmin 0	21 (FN) Tip II Hatası	42 (TN) Doğru Karar	63 (N')	%66,67
Toplam	64 (P)	64 (N)	128 (P+N)	
Recall (Duyarlılık)	%67,19	%65,62		

Çizelge 6.2'ye göre doğru tahmin edilen örneklerin tüm örnek sayısına oranı %66,41'dir. Bu oran toplam doğruluğu temsil eder. TP doğru pozitifler, TN doğru negatifler, FN yanlış negatifler ve FP yanlış pozitifler olarak temsil edilir.

Şekil 6.3, K-NN algoritmasına göre tahmin sonuçlarını göstermektedir. Ameliyat olmayanlar "0", ameliyat olanlar ise "1" olarak ifade edilmiştir. Buna göre, ameliyatsız 64 hastanın 42'si doğru olarak tahmin edilmiştir. Ayrıca ameliyat edilen 64 hastanın 43'ü doğru olarak tahmin edilmiştir.



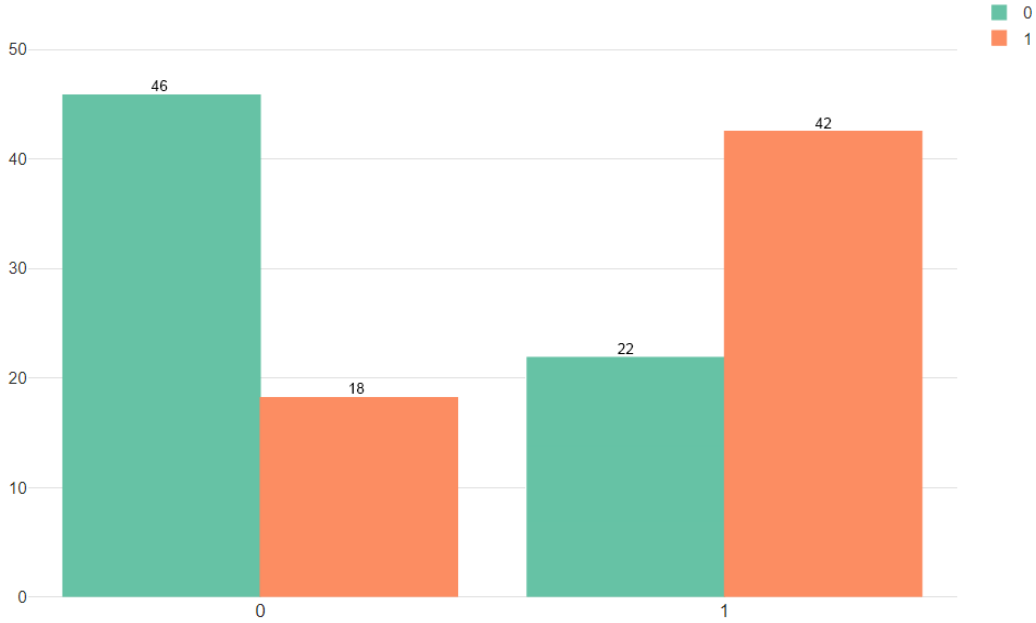
Şekil 6.3. K-NN algoritmasına göre tahmin sayıları

Çizelge 6.3. Lojistik regresyon algoritmasına göre karışıklık matrisi

Doğruluk: %68,75	Gerçek 1	Gerçek 0	Toplam	Precision (Kesinlik)
Tahmin 1	42 (TP) Doğru Karar	18 (FP) Tip I Hatası	60 (P')	%70
Tahmin 0	22 (FN) Tip II Hatası	46 (TN) Doğru Karar	68 (N')	%67,65
Toplam	64 (P)	64 (N)	128 (P+N)	
Recall (Duyarlılık)	%65,62	%71,88		

Çizelge 6.3'e göre doğru tahmin edilen örneklerin tüm örnek sayısına oranı %68,75'dir. Bu oran toplam doğruluğu temsil eder. TP doğru pozitifler, TN doğru negatifler, FN yanlış negatifler ve FP yanlış pozitifler olarak temsil edilir.

Şekil 6.4, lojistik regresyon algoritmasına göre tahmin sonuçlarını göstermektedir. Ameliyat olmayanlar "0", ameliyat olanlar ise "1" olarak ifade edilmiştir. Buna göre, ameliyatsız 64 hastanın 46'sı doğru olarak tahmin edilmiştir. Ayrıca ameliyat edilen 64 hastanın 42'si doğru olarak tahmin edilmiştir.



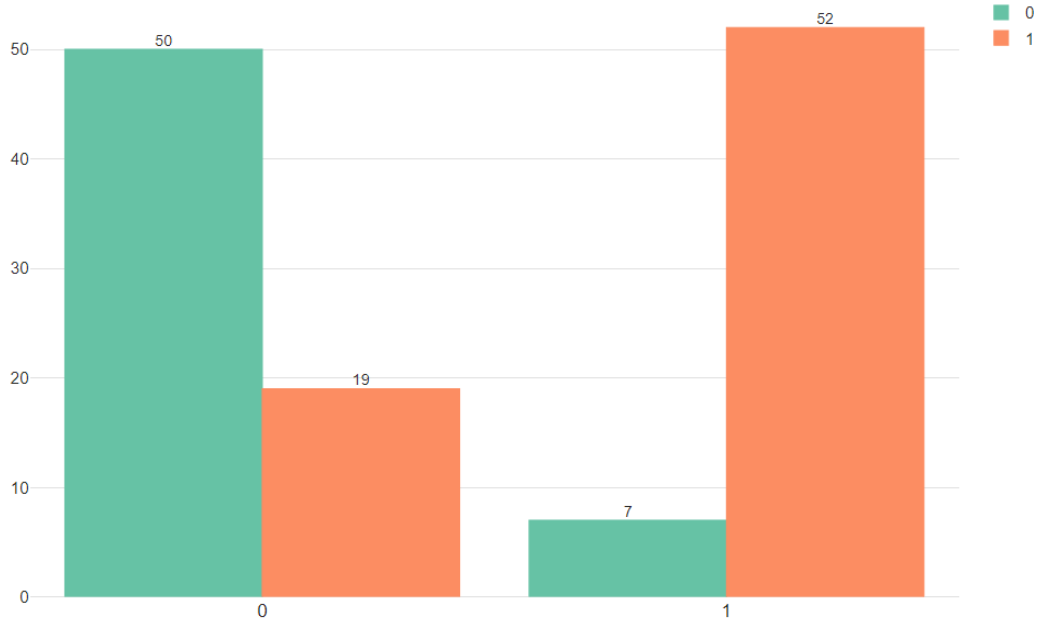
Şekil 6.4. Lojistik regresyon algoritmasına göre tahmin sayıları

Çizelge 6.4. Destek vektör makinesi algoritmasına göre karışıklık matrisi

Doğruluk: %79,69	Gerçek 1	Gerçek 0	Toplam	Precision (Kesinlik)
Tahmin 1	52 (TP) Doğru Karar	19 (FP) Tip I Hatası	71 (P')	%73,24
Tahmin 0	7 (FN) Tip II Hatası	50 (TN) Doğru Karar	57 (N')	%87,72
Toplam	59 (P)	69 (N)	128 (P+N)	
Recall (Duyarlılık)	%88,14	%72,46		

Çizelge 6.4'e göre doğru tahmin edilen örneklerin tüm örnek sayısına oranı %79,69'dur. Bu oran toplam doğruluğu temsil eder. TP doğru pozitifler, TN doğru negatifler, FN yanlış negatifler ve FP yanlış pozitifler olarak temsil edilir.

Şekil 6.5, destek vektör makinesi algoritmasına göre tahmin sonuçlarını göstermektedir. Ameliyat olmayanlar "0", ameliyat olanlar ise "1" olarak ifade edilmiştir. Buna göre, ameliyatsız 69 hastanın 50'si doğru olarak tahmin edilmiştir. Ayrıca ameliyat edilen 59 hastanın 52'si doğru olarak tahmin edilmiştir.



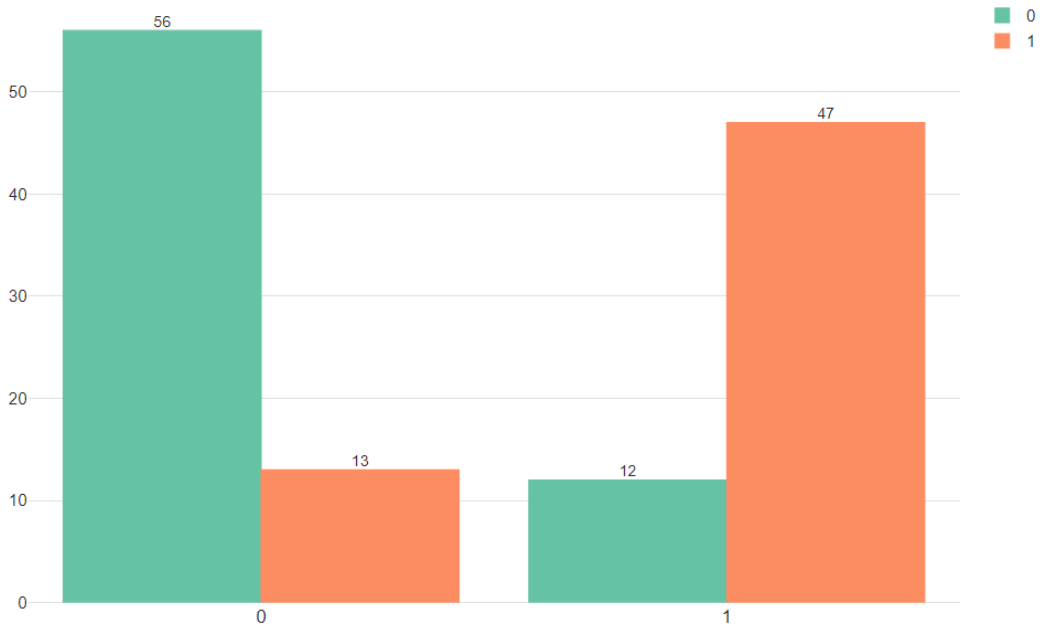
Şekil 6.5. Destek vektör makinesi algoritmasına göre tahmin sayıları

Çizelge 6.5. CART algoritmasına göre karışıklık matrisi

Doğruluk: %80,47	Gerçek 1	Gerçek 0	Toplam	Precision (Kesinlik)
Tahmin 1	47 (TP) Doğru Karar	13 (FP) Tip I Hatası	60 (P')	%78,33
Tahmin 0	12 (FN) Tip II Hatası	56 (TN) Doğru Karar	68 (N')	%82,35
Toplam	59 (P)	69 (N)	128 (P+N)	
Recall (Duyarlılık)	%79,66	%81,16		

Çizelge 6.5'e göre doğru tahmin edilen örneklerin tüm örnek sayısına oranı %80,47'dir. Bu oran toplam doğruluğu temsil eder. TP doğru pozitifler, TN doğru negatifler, FN yanlış negatifler ve FP yanlış pozitifler olarak temsil edilir.

Şekil 6.6, CART algoritmasına göre tahmin sonuçlarını göstermektedir. Ameliyat olmayanlar "0", ameliyat olanlar ise "1" olarak ifade edilmiştir. Buna göre, ameliyatsız 69 hastanın 56'sı doğru olarak tahmin edilmiştir. Ayrıca ameliyat edilen 59 hastanın 47'si doğru olarak tahmin edilmiştir.



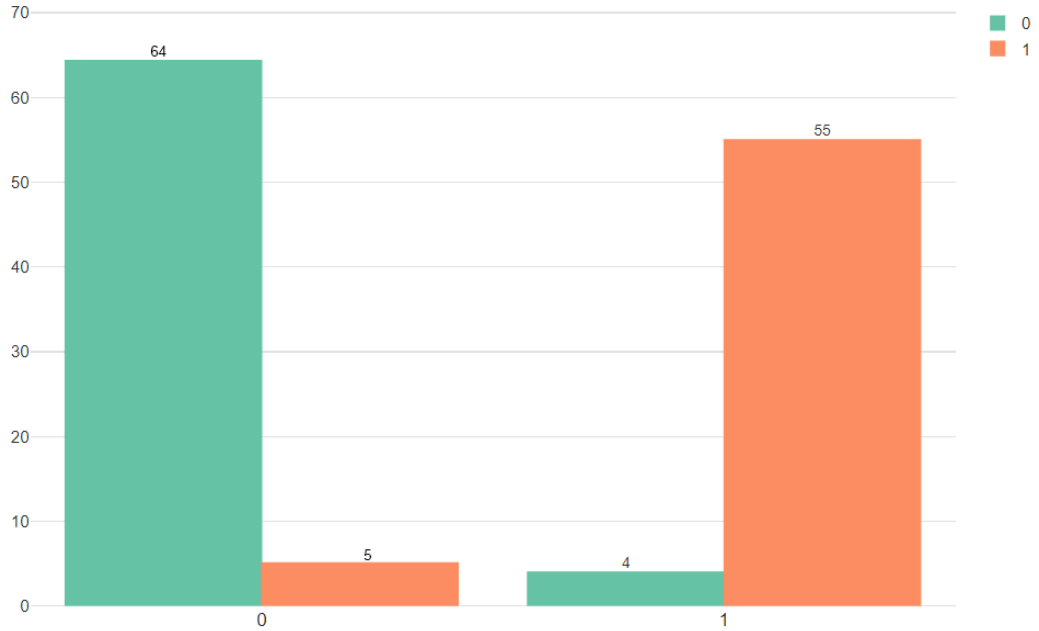
Şekil 6.6. CART algoritmasına göre tahmin sayıları

Çizelge 6.6. Rastgele orman algoritmasına göre karışıklık matrisi

Doğruluk: %92,96	Gerçek 1	Gerçek 0	Toplam	Precision (Kesinlik)
Tahmin 1	55 (TP) Doğru Karar	5 (FP) Tip I Hatası	60 (P')	%91,67
Tahmin 0	4 (FN) Tip II Hatası	64 (TN) Doğru Karar	68 (N')	%94,12
Toplam	59 (P)	69 (N)	128 (P+N)	
Recall (Duyarlılık)	%93,22	%92,75		

Çizelge 6.6'ya göre doğru tahmin edilen örneklerin tüm örnek sayısına oranı %92,96'dır. Bu oran toplam doğruluğu temsil eder. TP doğru pozitifler, TN doğru negatifler, FN yanlış negatifler ve FP yanlış pozitifler olarak temsil edilir.

Şekil 6.7, rastgele orman algoritmasına göre tahmin sonuçlarını göstermektedir. Ameliyat olmayanlar "0", ameliyat olanlar ise "1" olarak ifade edilmiştir. Buna göre, ameliyatsız 69 hastanın 64'ü doğru olarak tahmin edilmiştir. Ayrıca ameliyat edilen 59 hastanın 55'i doğru olarak tahmin edilmiştir.



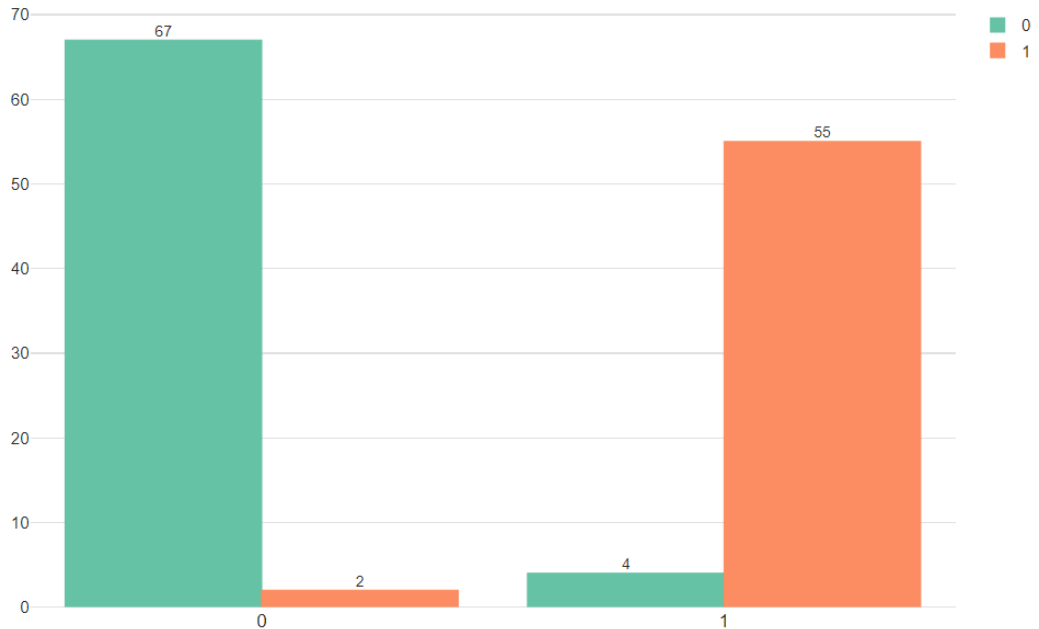
Şekil 6.7. Rastgele orman algoritmasına göre tahmin sayıları

Çizelge 6.7. Gradyan artırılmış ağaçlar algoritmasına göre karışıklık matrisi

Doğruluk: %95,31	Gerçek 1	Gerçek 0	Toplam	Precision (Kesinlik)
Tahmin 1	55 (TP) Doğru Karar	2 (FP) Tip I Hatası	57 (P')	%96,49
Tahmin 0	4 (FN) Tip II Hatası	67 (TN) Doğru Karar	71 (N')	%94,36
Toplam	59 (P)	69 (N)	128 (P+N)	
Recall (Duyarlılık)	%93,22	%97,1		

Çizelge 6.7'ye göre doğru tahmin edilen örneklerin tüm örnek sayısına oranı %95,31'dir. Bu oran toplam doğruluğu temsil eder. TP doğru pozitifler, TN doğru negatifler, FN yanlış negatifler ve FP yanlış pozitifler olarak temsil edilir.

Şekil 6.8, Gradyan artırılmış ağaçlar algoritmasına göre tahmin sonuçlarını göstermektedir. Ameliyat olmayanlar “0”, ameliyat olanlar ise “1” olarak ifade edilmiştir. Buna göre, ameliyatsız 69 hastanın 67'si doğru olarak tahmin edilmiştir. Ayrıca ameliyat edilen 59 hastanın 55'i doğru olarak tahmin edilmiştir.



Şekil 6.8. Gradyan artırılmış ağaçlar algoritmasına göre tahmin sayıları

Çizelge 6.8. Gradyan artırılmış ağaçlar algoritmasına göre değerlendirme ölçütleri

Ölçüt	Açıklama	Sonuç
Doğruluk (Accuracy, Recognition rate)	Doğru tahmin edilen örneklerin, tüm örneklerin sayısına oranıdır.	%95,31
Hata oranı (Error rate, Misclassification rate)	Yanlış tahmin edilen örneklerin, tüm örneklerin sayısına oranıdır. Doğruluk değeri ile toplamı bire eşittir.	%4,69
Duyarlılık (Sensitivity, Recall, True positive rate-TPR)	Doğru tahmin edilen pozitif sınıftaki örneklerin, gerçekte pozitif sınıftaki örneklerin sayısına oranıdır.	%93,22
Belirleyicilik (Specificity, True negative rate-TNR)	Doğru tahmin edilen negatif sınıftaki örneklerin, gerçekte negatif sınıftaki örneklerin sayısına oranıdır.	%97,1
Kesinlik (Pozitif öngörü değeri, Positive predictive value, precision)	Doğru tahmin edilen pozitif sınıftaki örneklerin, pozitif sınıfta tahmin edilen örneklerin sayısına oranıdır.	%96,49
Kesinlik (Negatif öngörü değeri Negative predictive value)	Doğru tahmin edilen negatif sınıftaki örneklerin, negatif sınıfta tahmin edilen örneklerin sayısına oranıdır.	%94,36
Yanlış pozitif oranı (False positive rate-FPR)	Yanlış tahmin edilen negatif sınıftaki örneklerin, negatif sınıftaki örneklerin sayısına oranıdır.	%2,89
Yanlış negatif oranı (False negative rate-FNR)	Yanlış tahmin edilen pozitif sınıftaki örneklerin, pozitif sınıftaki örneklerin sayısına oranıdır.	%6,78
F-ölçütü (F-score)	Duyarlılık ve kesinliğin birlikte değerlendirildiği ölçüttür.	%94,83

Çizelge 6.8’de araştırmada kullanılan gradyan artırılmış ağaçlar algoritmasına göre değerlendirme ölçütleri, ölçütlerin açıklamaları ve sonuçlar bulunmaktadır.

7. SONUÇLAR VE TARTIŞMA

Bu retrospektif çalışmada 0-18 yaş arası hastaların laboratuvar verilerini kullanarak apandisit cerrahisinin gerekliliğini öngören bir model önerilmiştir. Önerilen model, akut apandisit şüphesi olan hastalarda cerrahi karar vermeyi iyileştirmek için bir tanı aracı olarak geliştirilmiştir. Bu yöntemle AA tanısı en fazla %95,31 doğrulukla ayırt edilebilmiştir.

Muayene ve basit laboratuvar testleri sıklıkla akut apandisiti teşhis edebilir, ancak belirtiler atipik olduğunda tanı zordur. Tam kan sayımı kullanarak çalışmalarını yürüten araştırmacıların akut apandisit tanısında nötrofil/lenfosit oranları üzerinde daha fazla durdukları gözlenmiştir. Tanısal çalışmalarda, çoğunlukla bilgisayarlı görüntüleme teknikleri ve kan sayımlarından elde edilen değerlerin istatistiksel analizi ön plana çıkmaktadır. Çeşitli algoritmaların kullanımı ile bilgisayar kestirim becerilerinin katkısı az sayıda çalışmada görülmektedir.

AA tanısı, hasta öyküsü, fiziki muayene, laboratuvar belirteçleri ve görüntüleme yöntemleri ile konulmaya çalışılır [37, 38]. Ancak özellikle atipik belirtiler gösteren vakalarda olmak üzere bu yöntemlerden ilk üçüyle teşhis belirlenemediyse görüntüleme yöntemlerine başvurulur. Görüntüleme yöntemleri her ne kadar tanıyı belirlemede önemli faydalar sağlasa da bazı sınırlılıkları bulunmaktadır. Örneğin ultrasonografide deneyimli bir uzmana bağımlı olmak ve bağırsak gazlarından dolayı apendiksin görülemezliği ihtimali vardır [59, 60]. Akut apandisit tanısı için preoperatif bilgisayarlı tomografinin artan kullanımı negatif apendektomi oranlarının azaltılmasına yardımcı olmuştur [131-134]. Bununla birlikte, bilgisayarlı tomografi kullanımı radyasyona gereksiz yere maruz kalmayla ilgili endişeleri artırmıştır [135]. Radyasyona maruz kalmaya ek olarak, bu yöntemler özel ekipman ve deneyimli radyologlar gerektirir. Ayrıca bu yöntem gebelerde de kullanılamaz. Manyetik rezonans görüntüleme ise kolay ulaşılabilir olmaması, fazla maliyetli olması, görüntüleme süresinin uzun olması ve hasta uygunluğunun aranması gibi kısıtlılıklara sahiptir.

Bu çalışmanın diğerlerine göre farklı bir yönü, apendektomi tanısı için laboratuvar verilerine dayalı makine öğrenme yöntemlerinin kullanılmasıdır. Gelecekteki

alıřmalarda, modelin srekli eęitimine yeni hasta verileri eklenmelidir. Veri sayısı arttıķça tahminin doęruluęunun artabileceęi dřnlmektedir. Global olarak arařtırmaya deęer katan en nemli zellikler ařaęıdaki gibi sıralanabilir.

- Cerrahi operasyon iin karar vermede, zel ekipman ve personelin bulunmaması durumunda tanıya yardımcı olmak,
- Gereksiz apendektomilerin ve olası postoperatif komplikasyonların nlenmesine yardımcı olmak,
- Hastanede kalıř sresini azaltmak ve hasta bařına dřen tedavi maliyetlerinden tasarruf saęlamak,
- Hasta ile ilgilenen saęlık personelinin zamanlarını daha verimli kullanabilmelerini saęlamak,
- Birimlerdeki hasta yoęunluęunun dřrlmesine katkı saęlamak ve bylece kaynak tketimi azaltılmasına ve sınırlı saęlık bakımının daha doęru bir řekilde kullanılmasına yardım etmek,
- Olası hayati tehlike durumunu nceden tahmin etmek ve erken mdahale iin zaman sunabilmek,
- Hastanın ocuk veya yařlı olması durumunda yeterli iletiřim kurulamaması veya řikyetlerin doęru bir řekilde aktarılamaması nedeniyle tanının gecikme ihtimaline karřı teřhisin belirlenmesi iin hekimlere yardımcı olmak,
- Hastalıkla ilgili tanıların doęruluęunu test etmek,
- Makalenin ana problemi olan apendektominin tanısı iin kolay, hızlı ve doęru bir tahmin yntemi geliřtirmek.

EKLER

EK 1: Hitit Üniversitesi'nden Alman Etik Kurul Raporu



T.C.
HİTİT ÜNİVERSİTESİ
GİRİŞİMSSEL OLMAYAN ARAŞTIRMALAR ETİK KURULU

Sayı : 2019-113

29/03/2019

Konu: Başvuru Değerlendirme Sonucu

Sayın Öğr. Gör. Ömer Faruk AKMEŞE

Etik Kurulumuza yapmış olduğunuz başvurunuzla ilgili kurul kararımız ve ilgili bilgiler aşağıda yer almaktadır.

Bilgilerinize rica ederim.

Prof. Dr. Mehmet Ömer BOSTANCI
Başkan

Başvuru Numarası	2019-08
Sorumlu Araştırmacı	Öğr. Gör. Ömer Faruk AKMEŞE
Araştırma Başlığı	Makine Öğrenme Teknikleri ile Apandisit Cerrahisi Tahmini (Appendicitis Surgery Prediction Using Machine Learning Techniques)
Toplantı Tarihi	28/03/2019
Karar Numarası	2019-109

- ☒ Araştırma başvurunuz etik açıdan uygun bulunmuştur.
- ☒ Araştırmaya Kurum İzni/İzinleri alındıktan sonra başlanması uygun bulunmuştur.
- ☐ Başvurunun, ekte belirtilen düzeltmelerin yapılması halinde tekrar değerlendirilmesine karar verilmiştir.
- ☐ Araştırma projesi etik açıdan uygun olmadığından başvurunun reddine karar verilmiştir.

KAYNAKLAR

- [1] Salaffi, F., Gasparini, S., Ciapetti, A., Gutierrez, M., Grassi, W., Usability of an innovative and interactive electronic system for collection of patient-reported data in axial spondyloarthritis: comparison with the traditional paper-administered format. *Rheumatology*, 52 (11): 2062-2070, 2013.
- [2] Parthiban, L., Subramanian, R., Intelligent heart disease prediction system using CANFIS and genetic algorithm. *International Journal of Biological, Biomedical and Medical Sciences*, 3 (3): 157-160, 2008.
- [3] Boonstra, P., van Veen, R., Stockmann, H., Less negative appendectomies due to imaging in patients with suspected appendicitis. *Surgical Endoscopy*, 29 (8): 2365-2370, 2015.
- [4] Kansakar, N., Agarwal, P.N., Singh, R., Mishra, A., Kumar, J., Hadke, N.S., Sheth, S., Evaluation of combined use of modified Alvarado score and Ultrasound in predicting acute appendicitis: a prospective study. *International Surgery Journal*, 5 (11): 3594-3597, 2018.
- [5] van Amstel, P., et al., Gorter, R.R., van der Lee, J.H., Cense, H.A., Bakx, R., Heij, H.A., Ruling out Appendicitis in Children: Can We Use Clinical Prediction Rules? *Journal of Gastrointestinal Surgery*, 23 (10): 2027-2048, 2019.
- [6] Warner, B.W., Kulick, R.M., Stoops, M.M., Mehta, S., Stephan, M., Kotagal, U.R., An evidenced-based clinical pathway for acute appendicitis decreases hospital duration and cost. *Journal of Pediatric Surgery*, 33 (9): 1371-1375, 1998.
- [7] Behzatoğlu, B., Hatipoğlu, E., Bayramoğlu, S., Yılmaz, G., Yirik, G., Cimilli, T., Akut apandisit tanısında ultrasonografi ve bilgisayarlı tomografi bulgularının karşılaştırılması. *Bakırköy Tıp Dergisi*, 2: 22-24, 2006.

- [8] McGowan, D.R., Sims, H.M., Zia, K., Uheba, M., Shaikh, I.A., The value of biochemical markers in predicting a perforation in acute appendicitis. ANZ Journal of Surgery, 83 (1-2): 79-83, 2013.
- [9] Atema, J.J., Atema J.J., van Rossem, C.C., Leeuwenburgh, M.M., Stoker, J., Boermeester, M.A., Scoring system to distinguish uncomplicated from complicated acute appendicitis. British Journal of Surgery, 102 (8): 979-990, 2015.
- [10] Sevinç, M.M., Sevinç, M.M., Kınacı, E., Çakar, E., Bayrak, S., Özakay, A., Aren, A., Sarı S., Diagnostic value of basic laboratory parameters for simple and perforated acute appendicitis: an analysis of 3392 cases. Ulus Travma Acil Cerrahi Dergisi, 22 (2):155-162, 2016.
- [11] Colak, C., Karaman, E., Turtay, M.G., Application of knowledge discovery process on the prediction of stroke. Computer Methods and Programs in Biomedicine, 119 (3): 181-185, 2015.
- [12] Mdzingwa, N., Data Mining with Oracle 10g using Clustering and Classification Algorithms, 2005.
- [13] Brossette, S.E., Hymel, P.A. Jr, Data mining and infection control. Clinics in Laboratory Medicine, 28 (1): 119-126, 2008.
- [14] Dao, T.K., Zabaneh, F., Holmes, J., Disrude, L., Price, M., Gentry, L., A practical data mining method to link hospital microbiology and an infection control database. American Journal of Infection Control, 36 (3): S18-S20, 2008.
- [15] Aren, A., Gökçe, A.H., Gökçe, F.S., Özakay, K., Aksoy, Ş., Karagöz, B., Çelik, G., Aydın, İ., Akut apandisitın yaş, cinsiyet, lökosit değerleri ile ilişkisi. İstanbul Tıp Dergisi, 10 (3): 126-129, 2009.

- [16] Demirhan, A., Kılıç, Y.A., Güler, İ., Tıpta yapay zeka uygulamaları. Yoğun Bakım Dergisi, 9 (1): 31-41, 2010.
- [17] Kar, İ., Bakırarar, B., Köse, S.K., Bulanık Çıkarsama Sistemleri ile Veri Madenciliği Yöntemlerinin Sınıflama Performansının Benzetim Çalışması ile Karşılaştırılması. XIX. Ulusal ve II. Uluslararası Biyoistatistik Kongresi, Ekim 2017, Belek-Antalya, s 33, 2017.
- [18] Turhan, K., Kurt, T., Kurt, B., Buçan Kirkbir, İ., Akut Apandisit Nedeniyle Opere Edilen Hastaların Kümeleme Yöntemi ile Profillerinin Belirlenmesi. Tıp Bilişiminde Yenilikler Sempozyumu, Mart 2017, Ankara, 2017.
- [19] Schein, M., Wise, L., Controversies in surgery. Springer Science & Business Media, 2001.
- [20] Bickell, N.A., Aufses Jr, A.H., Rojas, M., Bodian, C., How time affects the risk of rupture in appendicitis. Journal of the American College of Surgeons, 202 (3): 401-406, 2006.
- [21] Busch, M., Gutzwiller, F.S., Aellig, S., Kuettel, R., Metzger, U., Zingg, U., In-hospital delay increases the risk of perforation in adults with appendicitis. World Journal of Surgery, 35 (7): 1626-1633, 2011.
- [22] Jeon, B.G., Kim, H.J., Heo, S.C., CT scan findings can predict the safety of delayed appendectomy for acute appendicitis. Journal of Gastrointestinal Surgery, 23 (9): 1856-1866, 2019.
- [23] Temple, C.L., Huchcroft, S.A., Temple, W.J., The natural history of appendicitis in adults. A prospective study. Annals of Surgery, 221 (3): 278-281, 1995.
- [24] Abeş, M., Apaydın, H.Ö., Türk çocuk cerrahlarının akut apandisit yaklaşımı. Çocuk Cerrahisi Dergisi, 29 (3): 88-92, 2015.

- [25] Tsao, K., St Peter, S.D., Valusek, P.A., Spilde, T.L., Keckler, S.J., Nair, A., Ostlie, D.J., Holcomb, G.W., Management of pediatric acute appendicitis in the computed tomographic era. *Journal of Surgical Research*, 147 (2): 221-224, 2008.
- [26] Lewis, F.R., Holcroft J.W., Boey, J., Dunphy, E., Appendicitis: a critical review of diagnosis and treatment in 1,000 cases. *Archives of Surgery*, 110 (5): 677-684, 1975.
- [27] Xharra, S., Gashi-Luci, L., Xharra, K., Veselaj, F., Bicaj, B., Sada, F., Krasniqi, A., Correlation of serum C-reactive protein, white blood count and neutrophil percentage with histopathology findings in acute appendicitis. *World Journal of Emergency Surgery*, 7 (1): 27, 2012.
- [28] Rubèr, M., Immunopathogenic aspects of resolving and progressing appendicitis. Linköping University Electronic Press, Sweden, 2012.
- [29] Addiss, D.G., Shaffer, N., Fowler, B.S., Tauxe, R.V., The epidemiology of appendicitis and appendectomy in the United States. *American Journal of Epidemiology*, 132 (5): 910-925, 1990.
- [30] Feldman, M., Friedman, L.S., Brandt, L.J., Sleisenger and Fordtran's *Gastrointestinal and Liver Disease E-Book*. Elsevier Health Sciences, Philadelphia, 2015.
- [31] Köksal, H., Uysal, B., Sarıbaşı, R., Bir Devlet Hastanesinin Akut Apandisit Tecrübesi/Acute Appendicitis Experience of a State Hospital. *Journal of Academic Emergency Medicine*, 9 (1): 41-44, 2010.
- [32] Aygün, A., Akut apandisit erken tanısında fizik muayene, laboratuvar, görüntüleme yöntemlerinin tanısal değeri ve skrolama sistemlerinin karşılaştırılması. Uzmanlık Tezi. Trakya Üniversitesi, Edirne, 2015.

- [33] Kılınçer, A., Akut Apandisit Ön Tanısı Olan Hastalarda Sağ Alt Kadrana Kompresyon Uygulanarak Elde Olunan Abdomen Bt Tetkikinin Tanısal Etkinliğinin Değerlendirilmesi. Uzmanlık Tezi, Hacettepe Üniversitesi, Ankara, 2013.
- [34] Anonim, Appendicitis. <https://web.archive.org/web/20160629211607/http://voxhealth.co/conditions/Appendicitis> (Erişim tarihi: 01.01.2020)
- [35] Abubakar, I., Tillmann, T., Banerjee, A., Global, regional, and national age-sex specific all-cause and cause-specific mortality for 240 causes of death, 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013. Lancet, 385 (9963): 117-171, 2015.
36. Anonim, Appendicitis deaths per million persons in 2012. https://en.wikipedia.org/wiki/Appendicitis#/media/File:Appendicitis_world_map-Deaths_per_million_persons-WHO2012.svg (Erişim tarihi: 05.11.2019)
- [37] Doğan, S., Bulut, A.Ş., Karamürsel, B.S., Akut apandisit nadir ayırıcı tanısı: Appendiks Endometriozisi. Turkish Journal of Surgery/Ulusal Cerrahi Dergisi, 28 (3), 159-161, 2012.
- [38] Fischer, J.E., Bland, K.I., Callery, M.P., Mastery of surgery. Lippincott Williams & Wilkins, USA, 2006.
- [39] Anonim, 3D medical animation still showing an inflammation of the appendix. <https://www.scientificanimations.com/wiki-images/> (Erişim tarihi: 08.11.2019)
- [40] Horata, E., Akut Apandisit Tanısı Doğrulan Hastaların, Preoperatif Periappendiküler Bt Morfometrilerinin İncelenmesi ve Periappendiküler Bt Değerleri ile Apandisit Perforasyonu Arasındaki İlişkinin İncelenmesi. Yüksek Lisans Tezi. Afyon Kocatepe Üniversitesi, Afyon, 2017.

- [41] Sözüer, E.M., Kayhan, E., Kerek, M., Keçeli, M., Yılmaz, Z., Akut Apandisitte Tanıda Gecikme Sebepleri, Teşhis Hataları Ve Perforasyon Sorunu. Türkiye Klinikleri Journal of Medical Sciences, 19 (3): 141-144, 1999.
- [42] Yildirim, E., Karagülle, E., Kirbaş, I., Türk, E., Hasdoğan, B., Tekşam, M., Coşkun, M., Alvarado scores and pain onset in relation to multislice CT findings in acute appendicitis. Diagnostic and Interventional Radiology, 14 (1): 14-18, 2008.
- [43] Kremer, K., Kraemer, M., Fuchs, K.H., Ohmann, C., The diagnostic value of rectal examination of patients with acute appendicitis. in Langenbecks Archiv für Chirurgie. Supplement. Kongressband. Deutsche Gesellschaft für Chirurgie. Kongress, 115: 1120-1122, 1998.
- [44] Paulson, E.K., Kalady, M.F., Pappas, T.N., Suspected appendicitis. New England Journal of Medicine, 348 (3): 236-242, 2003.
- [45] Silen, W., Cope, Z., Cope's early diagnosis of the acute abdomen. Oxford University Press, USA, 2005.
- [46] Yang, H.R., Wang, Y.C., Chung, P.K., Chen, W.K., Jeng, L.B., Chen, R.J., Laboratory tests in patients with acute appendicitis. ANZ Journal of Surgery, 76 (1-2): 71-74, 2006.
- [47] Mengüçük, M.E., Ayten, R., Bülbüller, N., Gödekmerdan, A., Başbuğ, Murat, Mungan, İ., Akut apandisit tanısında serum C-reaktif protein, prokalsitonin ve neopterinin yeri. Fırat Tıp Dergisi, 15 (1): 40-43, 2010.
- [48] Beltrán, M.A., Almonacid, J., Vicencio, A., Gutiérrez, J., Cruces, K.S., Cumsille, M.A., Predictive value of white blood cell count and C-reactive protein in children with appendicitis. Journal of Pediatric Surgery, 42 (7): 1208-1214, 2007.

- [49] Çağlayan, F., Çakmak, M., Çağlayan, O., Çavuşoğlu, T., Plasma D-lactate levels in diagnosis of appendicitis. *Journal of Investigative Surgery*, 16 (4): 233-237, 2003.
- [50] Coleman, C., Thompson, J.E. Jr, Bennion, R.S., Schmit, P.J., White blood cell count is a poor predictor of severity of disease in the diagnosis of appendicitis. *The American Surgeon*, 64 (10): 983-985, 1998.
- [51] Demircan, M., Cetin, S., Uguralp, S., Sezgin, N., Karaman, A., Gozukara, E.M., Plasma D-lactic acid level: a useful marker to distinguish perforated from acute simple appendicitis. *Asian Journal of Surgery*, 27 (4): 303-305, 2004.
- [52] Eriksson, S., Granström, L., Carlström, A., The diagnostic value of repetitive preoperative analyses of C-reactive protein and total leucocyte count in patients with suspected acute appendicitis. *Scandinavian Journal of Gastroenterology*, 29 (12): 1145-1149, 1994.
- [53] Estrada, J.J., Petrosyan, M., Barnhart, J., Tao, M., Sohn, H., Towfigh, S., Mason, R.J., Hyperbilirubinemia in appendicitis: a new predictor of perforation. *Journal of Gastrointestinal Surgery*, 11 (6): 714-718, 2007.
- [54] Goodwin, A.T., Swift, R.I., Bartlett, M.J., Fernando, B.S., Chadwick, S.J., Can serum interleukin-6 levels predict the outcome of patients with right iliac fossa pain? *Annals of the Royal College of Surgeons of England*, 79 (2): 130-133, 1997.
- [55] Käser, S.A., Fankhauser, G., Willi, N., Maurer, C.A., C-reactive protein is superior to bilirubin for anticipation of perforation in acute appendicitis. *Scandinavian Journal of Gastroenterology*, 45 (7-8): 885-892, 2010.
- [56] Thuijls, G., Derikx, J.P., Prakken, F.J., Huisman, B., van Bijnen Ing, A.A., van Heurn, E.L., Buurman, W.A., Heineman, E., A pilot study on potential new

plasma markers for diagnosis of acute appendicitis. The American Journal of Emergency Medicine, 29 (3): 256-260, 2011.

- [57] Pinto Leite, N., Pereira, J.M., Cunha, R., Pinto, P., Sirlin, C., CT evaluation of appendicitis and its complications: imaging techniques and key diagnostic findings. American Journal of Roentgenology, 185 (2): 406-417, 2005.
- [58] Choi, D., Park, H., Lee, Y.R., Kook, S.H., Kim, S.K., Kwag, H.J., Chung, E.C., The most useful findings for diagnosing acute appendicitis on contrast-enhanced helical CT. Acta Radiologica, 44 (6): 574-582, 2003.
- [59] Hızal, M., Akpınar, E., Akut Abdomen. Türk Radyoloji Seminerleri, 4: 286-298, 2016.
- [60] Brown, J., Acute appendicitis: the radiologist's role. Radiology, 180 (1): 13-14, 1991.
- [61] Gökçe, A.H., Aren, A., Gökçe, F.S, Dursun, N., Barut, A.Y., Reliability of ultrasonography for diagnosing acute appendicitis. Ulus Travma Acil Cerrahi Dergisi, 17 (1): 19-22, 2011.
- [62] Anonim, Ultrasound image of acute appendicitis. <https://commons.wikimedia.org/wiki/File:SonoAppendizitis.JPG> (Erişim tarihi: 02.11.2019)
- [63] Pedrosa, I., Zeikus, E.A., Levine, D., Rofsky, N.M., MR imaging of acute right lower quadrant pain in pregnant and nonpregnant patients. Radiographics, 27 (3): 721-743, 2007.
- [64] Arer, İ.M., Laparoskopik ve tek insizyon laparoskopik apendektominin prospektif karşılaştırma sonuçları. Uzmanlık Tezi. Başkent Üniversitesi, Ankara, 2012.

- [65] Heilman, J., A cat scan demonstrating acute appendicitis, https://en.wikipedia.org/wiki/Appendicitis#/media/File:CAT_scan_demonstrating_acute_appendicitis.jpg. (Eriřim tarihi: 07.11.2019)
- [66] Byott, S., Harris, I., Rapid acquisition axial and coronal T2 HASTE MR in the evaluation of acute abdominal pain. *European Journal of Radiology*, 85 (1): 286-290, 2016.
- [67] Incesu, L., Coskun, A., Selcuk, M.B., Akan, H., Sozubir, S., Bernay, F., Acute appendicitis: MR imaging and sonographic correlation. *AJR. American Journal of Roentgenology*, 168 (3): 669-674, 1997.
- [68] Pedrosa, I., Levine, D., Eyvazzadeh, A.D., Siewert, B., Ngo, L., Rofsky, N.M., MR imaging evaluation of acute appendicitis in pregnancy. *Radiology*, 238 (3): 891-899, 2006.
- [69] Riley, R., Murphy, J., Higgins, T., MRI imaging in pediatric appendicitis. *Journal of Pediatric Surgery Case Reports*, 31: 88-89, 2018.
- [70] Meystre, S., Clinical decision-Support tool for acute appendicitis. *Swiss Medical Informatics*, 19 (51): 31-36, 2003.
- [71] Alvarado, A., A practical score for the early diagnosis of acute appendicitis. *Annals of Emergency Medicine*, 15 (5): 557-564, 1986.
- [72] De la Torre Quiroga, A., Bosques Padilla, F.J., Morales Garza, L., How do we approach an adult patient with suspected acute appendicitis? *Medicina Universitaria*, 18 (71): 125-129, 2016.
- [73] Feussner, H., Becker, V., Bauer, M., Kranzfelder, M., Schirren, R., Lüth, T., Meining, A., Wilhelm, D., Developments in flexible endoscopic surgery: a review. *Clinical and Experimental Gastroenterology*, 8: 31-42, 2015.

- [74] Lee, S.C., Park, G., Choi, B.J., Kim, S.J., Determination of surgical priorities in appendicitis based on the probability of undetected appendiceal perforation. *World Journal of Gastroenterology: WJG*, 21 (7): 2131-2139, 2015.
- [75] Schwartz, S.I., *Principles of Surgery: PreTest Self-assessment and Review*. McGraw-Hill, Health Professions Division, USA, 1999.
- [76] Atahan, K., Üreyen, O., Aslan, E., Deniz, M., Çökmez, A., Gür, S., Avci, A., Tarcan, E., Preoperative diagnostic role of hyperbilirubinaemia as a marker of appendix perforation. *Journal of International Medical Research*, 39 (2): 609-618, 2011.
- [77] Witten, I.H., Frank, E., Hall, M., Pal, C., *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann, 2016.
- [78] Duda, R.O., Hart, P.E., *Pattern Classification*. 200 (1) Wiley-Interscience, USA, 2001.
- [79] Marr, B., Strategy, D., *Data Strategy: How to Profit from a World of Big Data, Analytics and the Internet of Things*. Kogan Page Ltd., New York, 2017.
- [80] Alpaydin, E., *Introduction to machine learning*. MIT press, 2014.
- [81] Dalyan, T., *Makine öğrenmesinde 1R algoritması ve ikinci kuralın (2R) oluşturulması*. Yüksek Lisans Tezi. Kocaeli Üniversitesi, Kocaeli, 2006.
- [82] Han, J., Pei, J., Kamber, M., *Data mining: concepts and techniques*. Elsevier, 2011.
- [83] Şentürk, A., *Veri madenciliği: kavram ve teknikler*. Ekin Yayınevi, 2006.

- [84] Petit, C., Bezemer, R., Atallah, L., A review of recent advances in data analytics for post-operative patient deterioration detection. *Journal of Clinical Monitoring and Computing*, 32 (3): 391-402, 2018.
- [85] Sinha, P., Sinha, P., Comparative study of chronic kidney disease prediction using KNN and SVM. *International Journal of Engineering Research and Technology*, 4 (12): 608-612, 2015.
- [86] Altukardeş, Z.A., Tip 2 diabetes mellituslu hastalarda 24 saatlik kan basıncı değişiminin öngörülmesine yönelik uzman sistem tasarımı. Doktora Tezi. Marmara Üniversitesi, İstanbul, 2012.
- [87] Tantıuş, A.C., Veri Madenciliğın Ve Demetleme. Yüksek Lisans Tezi. İstanbul Teknik Üniversitesi, İstanbul, 2002.
- [88] Vahaplar, A.M. İnceoğlu, Veri Madenciliği ve Elektronik Ticaret, Türkiye’de İnternet Konferansları-VII. Elektronik Poster Bildiri, Kasım 2001, İstanbul, 2001.
- [89] Çakır, Ö., Veri madenciliğinde sınıflandırma yöntemlerinin karşılaştırılması: bankacılık müşteri veri tabanı üzerinde bir uygulama. Basılmamış Doktora Tezi. Marmara Üniversitesi, İstanbul, 2008.
- [90] Patil, D.V., Bichkar, R., Multiple imputation of missing data with genetic algorithm based techniques. *IJCA Special Issue on" Evolutionary Computation for Optimization Techniques*, 2: 74-78, 2010.
- [91] Cios, K., Pedrycz, W., Swiniarski, R.W., Kurgan, L.A., *Data Mining: A Knowledge Discovery Approach*. Springer, New York, 2007.
- [92] Inthasone, S., Pasquier, N., Tettamanzi, A.G.B., da Costa Pereira, C. The BioKET Biodiversity Data Warehouse: Data and Knowledge Integration and Extraction. In: Blockeel H., van Leeuwen M., Vinciotti V. (eds) *Advances in*

Intelligent Data Analysis XIII. IDA 2014. Lecture Notes in Computer Science, vol 8819. Springer, Cham, 2014.

- [93] Alpaydın, E., Zeki veri madenciliği: Ham veriden altın bilgiye ulaşma yöntemleri. Bilişim 2000 eğitim semineri, 2000.
- [94] Kohavi, R., Provost, F., Glossary of terms. Machine learning, 30: 271-274, 1998.
- [95] Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., Wirth, R., CRISP-DM 1.0: Step-by-step data mining guide. SPSS inc, 16, 2000.
- [96] Turban E., Sharda, R., De, D., Decision support and business intelligence systems. Pearson Education India, 2011.
- [97] Larose, D.T., K-nearest neighbor algorithm. Discovering Knowledge in Data: An Introduction to Data Mining, 90, 2005.
- [98] Gürsoy, U.T.Ş., Veri madenciliği ve bilgi keşfi. Pegem Akademi, Ankara, 2009.
- [99] Akpınar, H., Veri tabanlarında bilgi keşfi ve veri madenciliği. İÜ İşletme Fakültesi Dergisi, 29 (1): 1-22, 2000.
- [100] Zahn, E., Informationstechnologie und Informationsmanagement. Allgemeine Betriebswirtschaftslehre, 2: 376-428, 2001.
- [101] Han, J., Pei, J., Yin, Y., Mining frequent patterns without candidate generation. ACM Sigmod Record. 29 (2): 1-12, 2000.
- [102] Hair, J.F., Black, W.C., Babin, B.J., Anderson, R.E., Tatham, R.L., Multivariate Data Analysis, 6 th Edition, Upper Saddle River, NJ: Pearson Prentice Hall, 2005.

- [103] Uzun, E., İnternet tabanlı bilgi erişimi destekli bir otomatik öğrenme sistemi. Doktora Tezi. Trakya Üniversitesi, Edirne, 2007.
- [104] Kocamaz, A.F., Makine Öğrenmesi Tabanlı Bir Uzman Tasarımı. Doktora Tezi. Trakya Üniversitesi, Edirne, 2012.
- [105] Loon, R. van, Machine Learning Explained: Understanding Supervised, Unsupervised, and Reinforcement Learning, 2014.
- [106] Abdella, M., Marwala, T., The use of genetic algorithms and neural networks to approximate missing data in database. IEEE 3rd International Conference on Computational Cybernetics, 2005, ICC 2005.
- [107] Mohamed, S., Marwala, T., Neural network based techniques for estimating missing data in databases. 16th Annual Symposium of the Pattern Recognition Association of South Africa, Langebaan, 2005.
- [108] Nelwamondo, F.V., Mohamed, S., Marwala, T., Missing data: A comparison of neural network and expectation maximization techniques. Current Science, 93 (11): 1514-1521, 2007.
- [109] Qiao, W., Gao, Z., Harley, R.G., Continuous on-line identification of nonlinear plants in power systems with missing sensor measurements. in Proceedings. 2005 IEEE International Joint Conference on Neural Networks, August 2005, Montreal, Que., Canada, 2005.
- [110] Orabi, K.M., Kamal, Y.M., Rabah, T.M., Early Predictive System for Diabetes Mellitus Disease. In: Perner P. (eds) Advances in Data Mining. Applications and Theoretical Aspects. ICDM 2016. Lecture Notes in Computer Science, vol 9728. Springer, Cham, 2016.

- [111] Gorunescu, F., Data Mining: Concepts, models and techniques, 12th Edition. Springer Science & Business Media, Berlin Heidelberg, 2011.
- [112] Han, J. and M. Kamber, Data Mining: Concepts and Techniques, 3th Edition. USA, 2012.
- [113] Gujarati, D.N., Temel Ekonometri (Çev. Ümit Şenesen, Gülay Günlük Şenesen). Birinci Basım, İstanbul: Literatür Yayınları, 1999.
- [114] Ürük, E., İstatistiksel uygulamalarda lojistik regresyon analizi. Yüksek Lisans Tezi. Marmara Üniversitesi, İstanbul, 2007.
- [115] Haykin, S., Neural Networks. M. Horton ed., Prentice Hall: Upper Saddle River, New Jersey, 1999.
- [116] Oztemel, E., Yapay sinir aglari. Papatya Yayıncılık, İstanbul, 2003.
- [117] Deng, L., Yu, D., Deep learning: methods and applications. Foundations and Trends® in Signal Processing, 7 (3–4): 197-387, 2014.
- [118] Upretee, P., Yüksel, M.E., Accurate Classification of Heart Sounds for Disease Diagnosis by A Single Time-Varying Spectral Feature: Preliminary Results. 2019 Scientific Meeting on Electrical-Electronics & Biomedical Engineering and Computer Science (EBBT). June 2019, Istanbul, Turkey, 2019.
- [119] Osowski, S., Siwek, K., Markiewicz, T., MLP and SVM networks-a comparative study. in Proceedings of the 6th Nordic Signal Processing Symposium, 2004. NORSIG 2004, June 2004, Espoo, Finland, 2004.
- [120] Vapnik, V., The nature of statistical learning theory. Springer science & business media, 2013.

- [121] Aizerman, M.A., Theoretical foundations of the potential function method in pattern recognition learning. Automation and Remote Control, 25: 821-837, 1964.
- [122] Boser, B.E., Guyon, I.M., Vapnik, V.N., A training algorithm for optimal margin classifiers. Proceedings of the fifth annual workshop on Computational learning theory, 1992.
- [123] Sadreddini, A., Kernel machines are used to compute a non-linearly separable functions into a higher dimension linearly separable function. https://commons.wikimedia.org/wiki/File:Kernel_Machine.png (Eriřim tarihi: 22.112.2019)
- [124] Friedman, J.H., Olshen, R.A., Stone, C.J., Classification and regression trees. University of California, Berkeley, 1984.
- [125] Özkan, Y., Veri madencilięi yöntemleri. Papatya, İstanbul, 2013.
- [126] Güner, Z., Veri Madencilięinde Cart ve Lojistik Regresyon Analizinin Yeri: İlaç Provizyon Sistemi Verileri Üzerinde Örnek Bir Uygulama. Sosyal Güvence, 6: 53-99, 2014.
- [127] Gülpınar, V., Avrupa Birlięi Ülkeleri ile Türkiye'nin Ekonomik Göstergelerinin Karar Ağacı Yöntemi ile Karşılaştırılması. Yüksek Lisans Tezi. Marmara Üniversitesi, İstanbul, 2008.
- [128] Breiman, L., Random forests. Machine learning, 45 (1): 5-32, 2001.
- [129] Garrido, A.P., What is the difference between Bagging and Boosting? <https://quantdare.com/what-is-the-difference-between-bagging-and-boosting/> (Eriřim tarihi: 15.01.2020)

- [130] Bengio, Y., Simard, P., Frasconi, P., Learning long-term dependencies with gradient descent is difficult. *IEEE Transactions on Neural Networks*, 5 (2): 157-166, 1994.
- [131] Anderson, B.A., Salem, L., Flum, D.R., A systematic review of whether oral contrast is necessary for the computed tomography diagnosis of appendicitis in adults. *The American Journal of Surgery*, 190 (3): 474-478, 2005.
- [132] Krajewski, S., Brown, J., Phang, P.T., Raval, M. MD, Brown, C.J., Impact of computed tomography of the abdomen on clinical outcomes in patients with acute right lower quadrant pain: a meta-analysis. *Canadian Journal of Surgery*, 54 (1): 43-53, 2011.
- [133] Rao, P.M., Rhea, J.T., Novelline, R.A., Mostafavi, A.A., McCabe, C.J., Effect of computed tomography of the appendix on treatment of patients and use of hospital resources. *New England Journal of Medicine*, 338 (3): 141-146, 1998.
- [134] Sporn, E., Petroski, G.F., Mancini, G.J., Astudillo, J.A., Miedema, B.W., Thaler, K., Laparoscopic appendectomy—is it worth the cost? Trend analysis in the US from 2000 to 2005. *Journal of the American College of Surgeons*, 208 (2): 179-185, 2009.
- [135] Shah, D., Sachs, R., Wilson, D., Radiation-induced cancer: a modern view. *The British Journal of Radiology*, 85 (1020): e1166-e1173, 2012.

ÖZGEÇMİŞ

Adı Soyadı : Ömer Faruk AKMEŞE

Doğum Tarihi : 07/03/1986

Yabancı Dil : İngilizce

Eğitim Durumu :

Lisans : Anadolu Üniversitesi(2004-2008)

Yüksek Lisans : Kırıkkale Üniversitesi(2012-2015)

Çalıştığı Kurum/Kurumlar ve Yıl/Yıllar:

Milli Eğitim Bakanlığı 2008 - 2012

Hitit Üniversitesi 2012 - ...

Yayınları (SCI) :

Yayınları (Diğer) :

Boyacı, Ö., Karagöl, S., Akmeşe, Ö. F., Kinect Algılayıcı ile Gömülü Sistem Tabanlı Robot Kolunun Matlab Simulink Destekli Gerçek Zamanlı Kontrolü. Anka E-Dergi, 2 (2): 10-18, 2018.

Akmeşe, Ö. F., Erbay, H., Boyacı, Ö., Akça, B. A., Yönetim Bilişim Sistemlerinde Gündem. 139-143. Robotikte Kinect Algılayıcı ile Gerçek Zamanlı Kontrol. Ed: by S. Gülseçen, Ç. S. Erol, Z. A. Reis, M. Gezer. Çağlayan, İstanbul, 2018.

Kör, H., Erbay, H., Demir, E., Akmeşe, Ö. F., A study on the factors affecting the academic performance of distance education students and formal students, Hitit Üniversitesi Sosyal Bilimler Enstitüsü Dergisi, 9 (2): 1029-1050, 2016.

Akmeşe, Ö. F., Demir, E., Dünder, E. Student perceptions for distance education and efficiency analysis of the system. Hitit Üniversitesi Sosyal Bilimler Enstitüsü Dergisi, 9 (2): 981-998, 2016.

Akmeşe, Ö. F., Erbay, H., Emiroğlu, B. G., Kör, H., Uzaktan Eğitimde Ders İçeriği Oluşturma Platformu. Hitit Üniversitesi Sosyal Bilimler Enstitüsü Dergisi, 10 (2): 1521-1546, 2015.

Aksoy, H., Kör, H., Akmeşe, Ö. F., & Erbay, H., Study on Professional Guidance on Teaching Settings. Procedia-Social and Behavioral Sciences, 141: 1215-1219, 2014.

Araştırma Alanları : Uzaktan Eğitim, Veri Madenciliği, Biyoenformatik, Makine Öğrenmesi

