

T.C.
FIRAT ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ



**COVID-19 TANISINDA BİYOKİMYA PARAMETRE
BASKINLIĞININ MAKİNE ÖĞRENİMİ YÖNTEMLERİ
KULLANILARAK BELİRLENMESİ**

Çağla DANACI

Yüksek Lisans Tezi

YAZILIM MÜHENDİSLİĞİ ANABİLİM DALI

HAZİRAN 2022

T.C.
FIRAT ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

Yazılım Mühendisliği Anabilim Dalı

Yüksek Lisans Tezi

COVID-19 TANISINDA BİYOKİMYA PARAMETRE BASKINLIĞININ
MAKİNE ÖĞRENİMİ YÖNTEMLERİ KULLANILARAK
BELİRLENMESİ

Tez Yazarı
Çağla DANACI

Danışman
Doç. Dr. Seda ARSLAN TUNCER

HAZİRAN 2022
ELAZIĞ

T.C.
FIRAT ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

Yazılım Mühendisliği Anabilim Dalı

Yüksek Lisans Tezi

Başlığı: Covid-19 Tanısında Biyokimya Parametre Baskınlığının Makine Öğrenimi Yöntemleri Kullanılarak Belirlenmesi

Yazarı: Çağla DANACI

İlk Teslim Tarihi: 25.05.2022

Savunma Tarihi: 30.06.2022

TEZ ONAYI

Fırat Üniversitesi Fen Bilimleri Enstitüsü tez yazım kurallarına göre hazırlanan bu tez aşağıda imzaları bulunan jüri üyeleri tarafından değerlendirilmiş ve akademik dinleyicilere açık yapılan savunma sonucunda OYBİRLİĞİ ile kabul edilmiştir.

İmza

Danışman: Doç. Dr. Seda ARSLAN TUNCER Onayladım
Fırat Üniversitesi, Mühendislik Fakültesi

Başkan: Doç. Dr. Muhammed TALO Onayladım
Fırat Üniversitesi, Mühendislik Fakültesi

Üye: Dr. Öğr. Üyesi Turab SELÇUK Onayladım
Kahramanmaraş Sütçü İmam Üniversitesi, Mühendislik Fakültesi

Bu tez, Enstitü Yönetim Kurulunun/...../20..... tarihli toplantısında tescillenmiştir.

İmza

Prof. Dr. Kürşat Esat ALYAMAÇ
Enstitü Müdürü

BEYAN

Fırat Üniversitesi Fen Bilimleri Enstitüsü tez yazım kurallarına uygun olarak hazırladığım “Covid-19 Tanısında Biyokimya Parametre Baskınlığının Makine Öğrenimi Yöntemleri Kullanılarak Belirlenmesi ” Başlıklı Yüksek Lisans Tezimin içindeki bütün bilgilerin doğru olduğunu, bilgilerin üretilmesi ve sunulmasında bilimsel etik kurallarına uygun davrandığımı, kullandığım bütün kaynakları atıf yaparak belirttiğimi, maddi ve manevi desteği olan tüm kurum/kuruluş ve kişileri belirttiğimi, burada sunduğum veri ve bilgileri unvan almak amacıyla daha önce hiçbir şekilde kullanmadığımı beyan ederim.

30.06.2022

Çağla DANACI



ÖNSÖZ

Covid-19 hastalığı kısa sürede tüm Dünya’da etkisini göstermeye başlayarak salgın bir hastalık haline gelmiştir. Tüm Dünya sağlık sistemi kısa süre içerisinde Covid-19 hastalığı ile mücadele içerisine girerek hastalığın hızlı tanı ve teşhisi için birçok farklı yöntem arayışına girmiştir. Bu aşamada, hastalığın erken teşhisi için laboratuvar testleri önemli bir rol oynayarak teşhise yardımcı bir araç olarak kullanılmaya başlanmıştır. Laboratuvar parametrelerindeki özellik sayısının çok olması teşhis aşamasında yorumlamayı zorlaştırdığı için daha az parametre ile daha yüksek performans gereksinimi ortaya çıkmıştır. Bu tez çalışmasında Covid-19 hastalığının teşhisine yönelik, özellik seçim yöntemlerini kullanarak makine öğrenimi tabanlı bir çalışma gerçekleştirilerek daha az özellik ile minimum maliyet ve maksimum performans sağlanacağı gösterilmek istenmiştir.

Tez çalışması sürecinde her zaman yol gösterici olan, bilgi birikimi ile çalışmama farklı açılardan bakmamı sağlayarak katkıda bulunan değerli hocam ve tez danışmanım Doç. Dr. Seda ARSLAN TUNCER’e sonsuz teşekkürü borç bilirim. Ayrıca, tüm hayatım boyunca yanımda olan, aldığım kararları her zaman destekleyen, tez çalışmam boyunca beni cesaretlendiren aileme teşekkürlerimi sunarım.

Çağla DANACI
ELAZIĞ, 2022

İÇİNDEKİLER

	Sayfa
ÖNSÖZ.....	iv
İÇİNDEKİLER	v
ÖZET	vi
ABSTRACT	vii
ŞEKİLLER LİSTESİ	viii
TABLolar LİSTESİ	ix
SİMGELER VE KISALTMALAR	x
1. GİRİŞ	1
1.1. Covid-19 Teşhisine Yönelik Literatürde Gerçekleştirilen Çalışmalar	3
2. SARS-CoV-2 PANDEMİSİ VE KLİNİK LABORATUVARLAR	7
2.1. Covid-19	7
2.2. Klinik Laboratuvarlar	10
2.2.1. Klinik Hematoloji.....	11
2.2.2. Klinik Biyokimya.....	11
2.2.3. Klinik Patoloji	12
2.2.4. Klinik Mikrobiyoloji	12
3. YAPAY ZEKÂ	13
3.1. Yapay Zekâ Alt Dalları.....	14
3.1.1. Uzman Sistemler	14
3.1.2. Genetik Algoritmalar.....	15
3.1.3. Bulanık Mantık.....	16
3.1.4. Yapay Sinir Ağları	16
3.1.5. Makine Öğrenmesi	19
4. MATERYAL VE METOT	25
4.1. Materyal.....	25
4.2. Metot	35
4.2.1. Boş Verilerin Tespiti ve Uygun Yöntemlerle Temizlenmesi	36
4.2.2. Aykırı Verilerin Analizi	36
4.2.3. Veri Normalizasyon İşlemleri	38
4.2.4. Özellik Seçimi	42
4.2.5. Sınıflandırma.....	49
4.3. Performans Değerlendirme	50
5. BULGULAR VE TARTIŞMA	52
6. SONUÇLAR.....	57
KAYNAKLAR.....	58
ÖZGEÇMİŞ	

ÖZET

Covid-19 Tanısında Biyokimya Parametre Baskınlığının Makine Öğrenimi Yöntemleri Kullanılarak Belirlenmesi

Çağla DANACI

Yüksek Lisans Tezi

FIRAT ÜNİVERSİTESİ
Fen Bilimleri Enstitüsü

Yazılım Mühendisliği Anabilim Dalı

Haziran 2022, Sayfa: x + 62

Covid-19 2020 yılında ortaya çıkan ve kısa sürede hızla yayılım göstererek pandemi haline gelen hastalıktır. Öksürük, nefes darlığı vb. belirtiler ile ortaya çıkan bu hastalık zaman içerisinde etkinliğini artırarak ölümcül bir hastalık haline gelmiştir. Bu sebeple Covid-19 hastalığının erken teşhisi önem arz etmektedir. Covid-19 hastalığının teşhisi için laboratuvar testleri, görüntüleme teknikleri vb. birçok farklı yöntem kullanılmaktadır. Gerçek zamanlı polimeraz zincir reaksiyonu (RT-PCR) testi altın standart olarak kabul edilmekte ve yaygın olarak kullanılmaktadır. Laboratuvar testleri kullanılarak gerçekleştirilen tanı işleminde parametre yoğunluğundan doğacak iş yükü artmakta, bu durumda tanıyı zorlaştırmaktadır. Bu nedenle uzmanlara tanı aşamasında yardımcı bir karar destek mekanizmasına ihtiyaç vardır.

Bu tez çalışmasında uzmanlara, Covid-19 hastalığının karar verme aşamasında destek olacak sistem makine öğrenimi algoritmaları ile tasarlanmıştır. Makine öğrenimi geniş kullanım alanı sayesinde birçok alanda uzmanlara destekleyici rol oynamaktadır. Sıklıkla kullanılan bazı makine öğrenimi algoritmaları K-en Yakın Komşu Algoritması, Destek Vektör Makineleri, Karar Ağaçları ve Yapay Sinir Ağlarıdır. Parametre fazlalığından doğacak sorunları önlemek amacıyla özellik seçim yöntemleri makine öğrenim algoritmaları ile birlikte kullanılarak parametre sayısını optimuma indirgenmiştir. Bu durumda daha az maliyetle uzmanlara teşhis aşamasında kolaylık sağlamaktadır.

Çalışmada biyokimya parametreleri üzerinde, 13 farklı özellik seçim yöntemi kullanarak her bir parametrenin Covid-19 hastalığını tanısındaki etkinliği analiz edilmiştir. Çalışma kapsamında, Elazığ Fethi Sekin Şehir Hastanesinden elde edilen 100 negatif ve 121 pozitif olmak üzere toplamda 16 özellikten oluşan 221 hasta verisi kullanılmıştır. Öznitelik seçme yöntemleri kullanılarak parametre sayısı azaltılmış ve sınıflandırma işlemi gerçekleştirilmiştir. Çalışma sonucunda elde edilen yeni özellik kümesinin tüm özellikler ile gerçekleştirilen işlemde başarılı olduğu gözlemlenmiştir. Özellik seçim yöntemleriyle elde edilen 5 özellik, %91.3 doğruluk oranında Uyarlanabilir Yapısal Öğrenme İle Özellik Seçimi yöntemiyle Yapay Sinir Ağları için en iyi performansı vermiştir. Çalışmanın klinik çalışmalarda hekimlere yardımcı olacağı ve yapılacak çalışmalara ışık tutacağı düşünülmektedir.

Anahtar Kelimeler: Covid-19, Makine Öğrenimi, Özellik Seçimi, Sınıflandırma

ABSTRACT

Determination of Biochemistry Parameter Dominance in the Diagnosis of Covid-19 Using Machine Learning Methods

Çağla DANACI

Master's Thesis

FIRAT UNIVERSITY
Graduate School of Natural and Applied Sciences
Department of Software Engineering

June 2022, Pages: x + 62

COVID-19 is a disease that appeared in 2020 and became a pandemic in a short time. Cough, shortness of breath, etc. - this disease, which appears with symptoms, has become a deadly disease as its effectiveness increases with time. For this reason, early detection of Covid-19 disease is important. Various laboratory tests, imaging techniques, etc. are used to diagnose Covid-19 disease. The real-time polymerase chain reaction test (RT-PCR) is considered the gold standard and is widely used. When diagnosing by laboratory tests, the workload increases due to the density of parameters, which makes diagnosis difficult in this case. Therefore, at the stage of diagnosis to specialists, an additional decision support mechanism is required.

In this work, the system for assisting experts in decision making in the case of Covid-19 disease was developed using machine learning algorithms. Machine learning plays a supporting role for experts in many fields thanks to its wide range of applications. Some commonly used machine learning algorithms include the K-nearest Neighbor algorithm, support vector machines, decision trees, and artificial neural networks. To avoid problems arising from the redundancy of parameters, the number of parameters was reduced to an optimum by using feature selection methods in conjunction with machine learning algorithms. In this case, the specialists are relieved in the diagnosis phase with lower costs.

The study analyzed the effectiveness of each parameter in diagnosing Covid-19 disease using 13 different feature selection methods for biochemical parameters. A total of 221 patient data consisting of 16 characteristics were used in the study, including 100 negative and 121 positive from Elazığ Fethi Sekin City Hospital. The number of parameters was reduced by using attribute selection methods and the classification process was performed. As a result of the study, it was found that the newly obtained feature set from the process performed with all features was successful. The 5 features obtained by feature selection methods gave the best performance for Artificial Neural Networks by feature selection method with Adaptive Structural Learning with an accuracy rate of 91.3%. It is expected that the study will help clinicians in clinical trials and provide information on the studies to be conducted.

Keywords: Covid-19, Machine Learning, Feature Selection, Classification

ŞEKİLLER LİSTESİ

	Sayfa
Şekil 2.1. Dünya Geneli Koronavirüs Dağılım Haritası	7
Şekil 2.2. Koronavirüs Salgını 6 Ülke İçin Vaka Dağılımları	8
Şekil 2.3. Klinik Laboratuvar Bölümleri	11
Şekil 3.1. Uzman Sistem Genel Yapısı.....	15
Şekil 3.2. Genetik Algoritma Akış Şeması.....	16
Şekil 3.3. Sinir Hücresi Yapısı	16
Şekil 3.4. Yapay Sinir Hücresi	17
Şekil 3.5. Yapay Sinir Ağı Yapısı	19
Şekil 3.6. Makine Öğrenimi Modelleri.....	20
Şekil 3.7. KNN Algoritması	22
Şekil 3.8. (a) Doğrusal Olarak Ayrılabilme Durumu (b) Doğrusal Olarak Ayrılamama Durumu.....	23
Şekil 3.9. Karar Ağacı Yapısı.....	24
Şekil 4.1. Pozitif ve Negatif Sınıflara Ait Frekanslar	25
Şekil 4.2. Covid-Pozitif ve Covid-Negatif Hastaların Özelliklere Göre Dağılımları	28
Şekil 4.3. Veri Özellik Dağılımları Histogram Grafikleri	32
Şekil 4.4. Süreç Tasarımı.....	35
Şekil 4.5. Veri Ön İşleme Adımları	36
Şekil 4.6. Aykırı Değer Analizi.....	37
Şekil 4.7. Tek Sınıflı Destek Vektör Makinesi Örnek Görsel	38
Şekil 4.8. Normalizasyon İşlemi Uygulanan Veri Dağılımları.....	41
Şekil 4.9. Özellik Seçim İşlemi	42
Şekil 4.10. Özellik Seçim Süreç Adımları.....	43
Şekil 4.11. Özellik Seçim Yöntemleri	43
Şekil 4.12. Korelasyon Temelli Özellik Seçim Süreci	45
Şekil 4.13. Karmaşıklık Matrisi.....	50
Şekil 5.1. (a) Tüm Özellikler İçin En İyi Performansa Sahip YSA Algoritmasının Karmaşıklık Matrisi (b) Relief-F Yöntemi İle Seçilen Özellikler İçin En İyi Performansa Sahip YSA Algoritmasının Karmaşıklık Matrisi (c) FSASL Yöntemi ile Seçilen Özellikler İçin En İyi Performansa Sahip YSA Algoritmasının Karmaşıklık Matrisi	56
Şekil 5.2. Yüksek Performans Sağlayan Özellik Seçim Yöntemlerine Ait Performans Grafiği	56

TABLÖLAR LİSTESİ

	Sayfa
Tablo 2.1. Covid-19 Salgını 6 Ülke İçin Vaka Verileri	8
Tablo 2.2. Covid-19 Tanısında Önemli Biyokimya Parametreleri.....	9
Tablo 3.1. Biyolojik Sinir Hücresi ve Yapay Sinir Hücresi	18
Tablo 4.1. Veriye Ait Betimleyici İstatistikler	26
Tablo 4.2. Box-cox Dönüşüm Tablosu	39
Tablo 4.3. Özellik Seçim Yöntemlerine Ait Türler.....	49
Tablo 5.1. Özellik Seçim Yöntemleri ile Elde Edilen Özellikler	52
Tablo 5.2. Her Özellik Seçim Yöntemi için Elde Edilen Sınıflandırma Performansı	53
Tablo 5.3. Yüksek Performans Sağlayan Özellik Seçim Yöntemleri.....	55

SİMGELER VE KISALTMALAR

Kısaltmalar

RT-PCR	: Gerçek Zamanlı Polimeraz Zincir Reaksiyonu
TT-PZR	: Ters Transkripsiyon-Polimeraz Zincir Reaksiyonu
DSÖ	: Dünya Sağlık Örgütü
AUC	: Roc Eğrisi Altında Kalan Alan
KNN	: K-en Yakın Komşu Algoritması
DVM	: Destek Vektör Makinesi
YSA	: Yapay Sinir Ağları
MCV	: Ortalama Korpüsküler Hacim
CBC	: Tam Kan Sayımı
LDH	: Laktat Dehidrojenaz
RBC	: Eritrosit
ALT	: Alanin Aminotransferaz
AST	: Aspartat Aminotransferaz
CRP	: C-reaktif protein
WBC	: Beyaz Kan Hücresi
CFS	: Korelasyon Temelli Özellik Seçim Yöntemi
ILFS	: Sonsuz Gizli Özellik Seçimi
UFSOL	: Sıralı Yerellik ile Denetimsiz Özellik Seçimi
DGUFS	: Bağımlılık Güdümlü Denetimsiz Özellik Seçimi
INFFS	: Sonsuz Özellik Seçimi
MCFS	: Çoklu Küme Özellik Seçimi
LLCFS	: Yerel Öğrenmeye Dayalı Kümeleme Kullanarak Özellik Seçimi
LS	: Laplacian Skoru
FSVFS	: İçbükey Minimizasyon Yoluyla Özellik Seçimi
FSASL	: Uyarlanabilir Yapısal Öğrenme İle Özellik Seçimi
EDFS	: Denetimsiz Ayrımcı Özellik Seçimi
DP	: Doğru Pozitif
YP	: Yanlış Pozitif
DN	: Doğru Negatif
YN	: Yanlış Negatif

1. GİRİŞ

Geçmişten bugüne birçok salgın hastalık ile mücadele eden insanlık, bu salgınların önlenmesi için yüksek düzeyde mücadele sergileyerek birçok yöntem arayışına girmiştir. 31 Aralık 2019 itibariyle Çin merkezli olarak ortaya çıkan korona virüs salgını kısa sürede insanlığın mücadele ettiği salgınlardan biri olmuş, tüm dünyayı etkisi altına alarak Dünya Sağlık Örgütü'nün kabulü ile bir pandemi haline gelmiş ve Covid-19 olarak adlandırılmıştır [1]. Salgın kapsamındaki yüksek vaka artışları, dünya çapında sağlık hizmetleri alanında yoğunluğa sebebiyet vermiş ve bu durum salgın ile mücadeleyi zor hale getirmiştir. Covid-19 hastalığı şiddetli akut solunum sendromu ve kısa sürede ölüme yol açabilen çoklu organ yetmezliği dâhil olmak üzere insan vücudu üzerinde çeşitli etkilere sahiptir [1]. Gün geçtikçe binlerce insanın ölümüne sebebiyet veren korona virüs salgını, etkinliğini artırarak yayılmaya, kalıcı hasarlar bırakmaya ve ölümlere sebep olmaya devam etmektedir. Esas olarak damlacık yolu ile bulaş haline gelen Covid-19 hastalığı, öksürük, solunum sıkıntısı, halsizlik, yüksek ateş, tat-koku kaybı ve son zamanlarda çoğunlukla grip gibi belirtiler ile karşımıza çıkmaktadır [2]. Çalışmalara göre hastalığın ilk aşamasında görülen semptomlardan yüksek ateş %98, öksürük %76 ve bunu takiben halsizlik ve yorgunluk %44 oranında görülmektedir [2]. Hastada belirlenen bu semptomlar doğrultusunda, hastalığın tanı ve teşhisine yönelik olarak birçok yöntem kullanılmaktadır. Temel olarak Covid-19 hastalığını teşhis etmek için kullanılan küresel yöntem Ters Transkripsiyon-polimeraz Zincir Reaksiyonudur (TT-PZR). Dezavantajı duyarlılık ve özgüllükten dolayı tam performans sağlayamamak olan bu yöntem, uzak ve kırsal bölgelerde test kitlerinin az olması sebebi ile aktif olarak kullanılamamaktadır [3]. Aktif olarak kullanılamayan bu küresel yöntemin yanı sıra aktif kullanılabilecek yöntemler mevcuttur. Bu yöntemlerden bazıları; hasta klinik belirtileri, temas öyküsü, bilgisayarlı tomografi ile akciğer bulgularının tanımlanması gibi yardımcı tetkiklere dayanmaktadır [4]. Radyolojik yöntemler, benzer akciğer hastalıkları ile aynı özellikleri göstermesi sebebiyle tanı aşamasında Covid-19 hastalığı için yetersiz kalabilmekte ve bu durumda ek laboratuvar testlerine aktif olarak ihtiyaç duyulabilmektedir.

Laboratuvarlar, hastalık tanı ve tedavi aşamasında gerek duyulan tetkik ve analizler ile tanıya ve tedaviye önemli ölçüde destek sağlayan birimlerdir. Laboratuvarların ortaya koyacağı sonuçlar tanı aşamasında önemli role sahip olmakla beraber yüksek oranda doğruluk sağlamaktadır. Gerçekleştirilen bir çalışmaya göre hasta kayıtlarındaki bulgular %94 oranında laboratuvar verilerine dayanmaktadır [5]. Laboratuvarlar sağladıkları bu önem derecesiyle güvenilir test sonuçları üretmek için tasarlanmış birimler olarak tanımlanabilir.

Klinik laboratuvarlar, hastalık tanı ve süreç izleme adımlarında vücuttan elde edilen çeşitli örneklerin analizlerinin gerçekleştirildiği laboratuvar birimleridir. Klinik Laboratuvarda yapılan analizler ile tayin etme, ölçme veya çeşitli maddelere ve mikroorganizmalara ait varlık ve yokluk

durumlarını açıklama işlemleri gerçekleştirilir. Klinik laboratuvarlar test işlemini katı ve mantık çerçevesinde gerçekleştirirler [6]. Klinik laboratuvarlarda gerçekleştirilen işlemler örneklerin alınıp, bu örneklerle göre sonuç üretmenin ötesindedir. Alınan örneklerin doğru zamanda alınmış olması, doğru kişiden alınması, en doğru ve en güvenilir metodun seçilmesi, hasta ve doktorun doğru yönlendirilmesi klinik laboratuvarların en önemli görevlerindendir. Klinik laboratuvarlar günümüzde gerçekleştirilen analizlerin en ileri teknolojiler kullanılarak yapıldığı birimler olarak karşımıza çıkmaktadır. Kaliteli laboratuvar hizmetlerinin sağlanması açısından ileri teknolojilere sahip olmak ve bu teknolojileri doğru bir şekilde kullanabilmek büyük önem taşımaktadır. Hızla gelişen teknoloji ile paralel olarak hastalıkların erken teşhis aşamasında ve doğru tedavi edilmesinde, klinik laboratuvarlar önemli bir konumda bulunmaktadır. Teknoloji ile paralel gelişim gösteren laboratuvar hizmetleri günümüzde çoğunlukla manuel olmaktan uzaklaşarak otonom hale gelmiştir.

Son yıllarda teknolojik gelişmeler hem hayatımızı kolaylaştıran hem de yaşamımıza yön veren gücüyle her alanda hızlı bir şekilde gelişim göstermeye devam etmektedir. Bu teknolojik gelişime katkı sağlayan alanlardan biri olarak yapay zekâ karşımıza çıkmaktadır. Yapay zekâ, insan zekâsını taklit ederek, topladığı bilgiler doğrultusunda sürekli olarak kendini iyileştirebilen sistemler olarak tanımlanabilir. Kendi içerisinde birçok alt dala ayrılan yapay zekânın alt dallarından bazıları genetik algoritmalar, uzman sistemler, bulanık mantık, doğal dil işleme olarak sıralanabilir. Her yapay zekâ alt dalı kendi içerisinde farklı kullanım alanlarına sahip olup, bu alanlara eğitim, eğlence, siber güvenlik ve tıp örnek olarak verilebilir [7].

1970'li yıllardan itibaren sağlık alanında faaliyet göstermeye başlayan yapay zekâ, tıp alanında önemli gelişmelerin yaşanmasını sağlamıştır [8]. Yapay zekânın tıp alanında temel olarak amacı teşhis ve tedavi tabanlı yöntemler sunabilmektir. Yapay zekâ çalışmaları, karmaşık problemleri çözmek için uzman bilgisinin kullanılmasına izin verdiği için tıbbi uygulamaya ve öğretime katkıda bulunabilir [8]. Günümüzde bir pandemi haline gelen Covid-19 ile ilişkilendirildiğinde ise yapay zekâ ölüm oranı tahmini, ilaç kullanımı, tanı ve tedavi yöntemleri geliştirme gibi farklı alanlarda yardımcı olmaktadır.

Bu çalışma Covid-19 tanısı için elde edilen biyokimya laboratuvar parametreleri üzerinde, 13 farklı özellik seçim yöntemini kullanarak optimum özellik sayısı ile sınıflandırma işlemi gerçekleştirilerek Covid-19 tanısı koymayı amaçlamaktadır. Bu amaç doğrultusunda Covid-19 tanısı için hangi parametrelerin daha belirleyici olduğu araştırılmıştır. Gerçekleştirilen çalışmada Covid-19 hastalığının teşhisine yönelik daha önce gerçekleştirilen çalışmalar incelenerek derlenmiştir. İncelenen çalışmalardan yola çıkarak tanı aşamasında kullanılan yöntemler ayrıntılı olarak ele alınmış ve ele alınan çalışmalardan çıkarılan sonuçlar doğrultusunda özellik seçim yöntemleri kullanılarak gerçekleştirilen Covid-19 teşhisine yönelik çalışmalar araştırılmıştır. Çalışma kapsamında makine öğrenimi aşamasında Destek Vektör Makinesi (DVM), K-en Yakın Komşu

Algoritması (KNN), Karar Ağaçları ve XGBoost algoritması kullanılarak elde edilen sonuçlar doğrultusunda özellik seçme algoritmalarının performansları, karşılaştırılmalı olarak değerlendirilmiştir.

1.1. Covid-19 Teşhisine Yönelik Literatürde Gerçekleştirilen Çalışmalar

Literatür incelendiğinde Covid-19 teşhisine yönelik makine öğrenimi kullanılarak gerçekleştirilen birçok çalışma bulunduğu görülmektedir. Literatürde mevcut çalışmalar Covid-19 teşhisinde makine öğrenimi ve Covid-19 teşhisinde özellik çıkarımı kullanılarak makine öğrenimi kategorilerinde analiz edilmiştir. Gerçekleştirilen analizlerin sonucu olarak derlenen literatür çalışması aşağıda yer almaktadır.

Aljoue ve arkadaşları gerçekleştirdikleri çalışmada hastaya ait göğüs röntgen görüntüleri, laboratuvar verileri, altta yatan sağlık durumu, yaş ve cinsiyet verilerinin birleşiminden oluşan bir yapı kullanarak Covid-19 hastalığının erken tanı ve teşhisinde yönelik bir makine öğrenimi modeli geliştirmeyi amaçlamışlardır. Çalışma kapsamında Riyad'daki King Abdulaziz Medical City'ye kabul edilen geçmişe yönelik toplamda 5739 Covid-pozitif hasta verisi üzerinde çalışmışlardır. Çalışmada bazı sınırlandırmalar (18 yaş üstü hastaların çalışmaya dâhil edilmesi vb.) kullanarak veriyi kısıtlamışlardır. Sınırlandırmalar sonrası istenilen kriterleri karşılayan toplam hasta sayısını 1508 olarak belirlemişlerdir. Elde edilen veriler üzerinde öncelikle veri ön işleme ve özellik seçimi işlemlerini uygulamışlardır. Eksik verisi bulunan hastaları analiz çalışmasına dâhil etmemişlerdir. Ardından z-skor yöntemini kullanarak veriler üzerinde standartlaştırma işlemi uygulamışlardır. Ön işleme adımlarını tamamladıktan sonra Relief-F özellik seçim yöntemini kullanarak özellik seçim işlemi gerçekleştirmişlerdir. Eğitim verisinde dengesizlik sorununu Yapay Azınlık Aşırı Örneklemeye yöntemini kullanarak gidermişler ve veride dengesizlik sorununu çözümlemişlerdir. Veri üzerinde gerçekleştirilen tüm işlemleri tamamladıktan sonra %80 eğitim %20 test verisi olarak veriyi bölüp, Destek Vektör Makineleri, Rastgele Orman, Doğrusal Regresyon ve eXtreme Gradient Boosting algoritmalarını kullanarak çalışmayı modellemişlerdir. Çalışma sonucunda %82 oranında bir roc eğrisi altında kalan alan (AUC) puanı elde ederek çalışmayı sonlandırmışlardır [9].

Cabitza ve arkadaşları Covid-19 hastalığının erken tanısı için 19 Şubat 2020-31 Mayıs 2020 tarihleri arasında San Raffaele Hastanesi'nde bulunan rutin kan testlerinden oluşan toplamda 1925 hasta verisi üzerinde çalışmışlardır. İlk adımda %75'ten fazla kayıp içeren verileri çalışma kapsamı dışında bırakmışlardır. Bu işlemten sonra toplamda 1624 hasta verisi tam veri özelliği taşımakta olarak belirlenmiştir. Hesaplama, veri üzerinde normalleştirme çalışmaları, özellik seçim işlemi ve sınıflandırma işlemini kapsayan dört adımlı bir çalışma prensibi uygulamışlardır. Özellik seçim aşamasında öznitelik eleme algoritmasını kullanmışlar ve sonrasında veriyi Rastgele Orman, Naive Bayes, Lojistik Regresyon, DVM ve KNN olmak üzere toplamda 5 algoritma kullanarak

modellemişlerdir. Çalışma sonucunda en iyi performansı %78 kesinlik, %74 duyarlılık, %81 özgüllük, %78 AUC değeri ve %94 doğruluk oranı ile KNN algoritmasından elde etmişlerdir [10].

Formica ve arkadaşları Tor Vergata Üniversite Hastanesi'nin acil servisine şüpheli Covid-19 belirtileri (ateş, öksürük vb.) ile başvuran toplamda 171 hastaya ait tam kan sayımı verilerini kullanarak sürüntü testi olmadan Covid-19 hastalığını erken teşhis edebilmek için çalışmaya dâhil etmiştir. Çalışma kapsamında toplamda 12 değişken kullanmışlardır. Bu değişkenleri cinsiyet, yaş, hemoglobin, ortalama alyuvar hacmi, kırmızı hücre dağılım genişliği – varyasyon katsayısı ve kırmızı kan hücreleri, trombositler, nötrofiller, lenfositler, monositler, eozinofiller ve bazofiller olarak belirlemişlerdir. Veriler üzerinde öncelikle LASSO özellik seçim yöntemi sekiz önemli değişkeni (yaş, hemoglobin, ortalama korpüsküler hacim (MCV), varyasyon katsayısı (RDW-CV), trombositler, monositler, eozinofiller ve bazofiller) belirlemişlerdir. Çalışma sonucunda tam kan sayımı verilerinin Covid-19 hastalığının tahmininde belirleyici bir rol oynadığını belirlemişlerdir. Çıkarılan bu sonuca göre de kan sayımı gerçekleştirilen hastaların sürüntü testine ihtiyaç duymayacağı yönünde bir çıkarımda bulunmuşlardır [11].

Banerjee ve arkadaşları tam kan sayımı verilerinden Covid-19 hastalığını tahmin etmek için makine öğrenimi tabanlı bir çalışma gerçekleştirmişlerdir. Çalışma kapsamında kullanılan toplamda 598 tam kan sayımı içeren hasta verisini Israelita Albert Einstein Hastanesinden elde etmişlerdir. Kullanılan veri değişkeni olarak yaş (yüzdelik grup), sürüntü testi sonucu ve standart tam kan sayımı (hematokrit, hemoglobin, trombositler, ortalama trombosit hacmi, kırmızı kan hücreleri, lenfositler, ortalama korpüsküler hemoglobin yoğunlaşması, lökositler, bazofiller, nötrofiller, ortalama korpüsküler hemoglobin, eozinofiller, ortalama korpüsküler hacim, monositler ve kırmızı kan hücresi dağılım genişliği) değişkenlerini içermektedir. Çalışmanın modelleme aşamasında makine öğrenimi modeli olarak Rastgele Orman modelini tercih etmişlerdir. Makine öğrenimine ek olarak bir Yapay sinir ağı modeli kullanarak çalışmayı yapay zekâ kapsamında da değerlendirmişlerdir. Kullanılan yapay sinir ağı modelini üç ara katman içeren bir model oluşturmuşlardır. Modelin performansını 10 katlı çapraz doğrulama ile değerlendirmişlerdir. Veri kümesinde bulunan pozitif ve negatif hasta sayısı dengesizliği model performansını düşürdüğü için SMOTE tekniği kullanarak bu dengesizliği gidermeye çalışmışlardır. Uygulanan tüm işlemler sonrasında düzenli serviste kontrolü sağlanan hastalar için 10 katlı çapraz doğrulama ile değerlendirilen yapay sinir ağlarının performans ölçütlerini doğruluk, AUC, duyarlılık ve özgüllük olarak sırasıyla %87, %80, %43 ve %91 olarak elde etmişlerdir. Rastgele Orman algoritması kullanılarak ise sırasıyla %82, %86, %60 ve %88 olarak belirlemişlerdir. Aynı işlemleri serviste takibi sağlanmayan ayakta hastalar için uyguladıklarında ise YSA ile sırasıyla %85, %94, %90, %95 duyarlılık, özgüllük, kesinlik ve AUC değerlerini elde ederken rastgele orman algoritması için sırasıyla %82, %90, %91, %94 değerlerini elde etmişlerdir [12].

Kukar ve arkadaşları rutin kan testi sonuçlarını kullanarak Covid-19 tanısını belirlemeye yönelik bir makine öğrenimi çalışması gerçekleştirmişlerdir. Çalışma için kullanılan veriyi 2020 Mart-Nisan dönemleri içerisinde Slovenya Üniversitesi Tıp Merkezi Ljubljna Enfeksiyon Hastalıkları Departmanında verisi bulunan 131 pozitif ve 5221 negatif hastadan elde etmişlerdir. Elde edilen veri seti yaş, cinsiyet, rutin kan testi değişkenlerini içermektedir. Pozitif hasta grubunda bulunan 117 kan parametresi içinden ayrıştırılmalar gerçekleştirerek toplamda 35 parametre seçimi gerçekleştirmişlerdir. Her parametre için medyan değerlerini ve bağıl referans aralıklarını hesaplayarak parametre değerlerini ölçeklendirmişlerdir. Makine öğrenimi aşamasında XGBoost, Rastgele Orman, Sinir Ağları, DVM algoritmalarını karşılaştırmalı olarak kullanmışlardır. Çalışma sonucunda en yüksek performansı %81.9 duyarlılık, %97.9 özgüllük ve %91 AUC değerleri ile XGBoost algoritmasından elde etmişlerdir [13].

Çubukçu ve arkadaşları rutin laboratuvar test sonuçlarını kullanarak Covid-19 hastalığının teşhisine yardımcı olacak makine öğrenimi tabanlı bir karar destek modeli geliştirmeyi amaçlamışlardır. Çalışmada kullanılan veri seti bilirubin, Alanin Aminotransferaz (ALT), aspartat aminotransferaz (AST), Laktat Dehidrojenaz (LDH), kreatinin ve C-reaktif proteini (CRP), eozinofiller, monositler, lenfositler, kırmızı hücre dağılım genişliği, trombositler, ortalama korpüsküler hemoglobin, lökositler, ortalama eritrosit hacmi, ortalama korpüsküler hemoglobin yoğunlaşması, nötrofiller, hemoglobin, hematokrit, bazofiller ve Eritrosit (RBC) parametrelerinden oluşmaktadır. Çalışma kapsamında Rastgele Orman Sınıflandırıcısı, XGBoost, Lojistik Regresyon ve DVM kullanılmıştır. Çalışmada en iyi performansı %79.58 duyarlılık, %91.54 özgüllük, %84.64 F1 skoru ve %85.30 doğruluk ile Rastgele Orman Sınıflandırıcısı ile elde etmişlerdir [14].

Silveira gerçekleştirdiği çalışmada hastaya ait tam kan sayımı ve yaş verilerini kullanarak Covid-19 tanısına yönelik bir model geliştirmeyi hedeflemiştir. Çalışma kapsamında kullanılan verileri veri paylaşımı deposu tarafından açık kullanıma sunulan toplamda 1557 hastadan elde etmiştir. Çalışmayı birinci adım olarak veri analizi ikinci adım olarak da tahmin modelinin oluşturulması adımı ile iki grupta çözümlemiştir. Değişkenlerin istatistiksel olarak anlamlılıklarını incelediğinde hemoglobin ölçümü, ortalama alyuvar hacmi, kırmızı hücre dağılım genişliği, ortalama trombosit hacmi ve nötrofil-lenfosit oranı değişkenleri dışında kalan tüm kan verilerinin arasında anlamlı ilişkiler olduğunu gözlemlemiştir. %75.6 duyarlılık, %82 özgüllük ve %80 doğruluk oranlarına XGBoost algoritması ile ulaşmıştır. Model kararı üzerine en çok etki eden değişkenleri ise bazofil, eozinofil ve lökosit olarak belirlemiştir [15].

Alves ve arkadaşları rutin kan testlerini kullanarak Covid-19 teşhisine yönelik akıllı bir tahminleme sistemi geliştirmeyi amaçlamışlardır. Albert Einstein Hastanesi'ne ait halka açık bir veri kümesinden elde ettikleri verileri ön işleme ve veri temizleme aşamalarından geçirdikten sonra toplamda 111 özellikten oluşan 84 pozitif ve 608 negatif olmak üzere 692 hasta verisi üzerinde

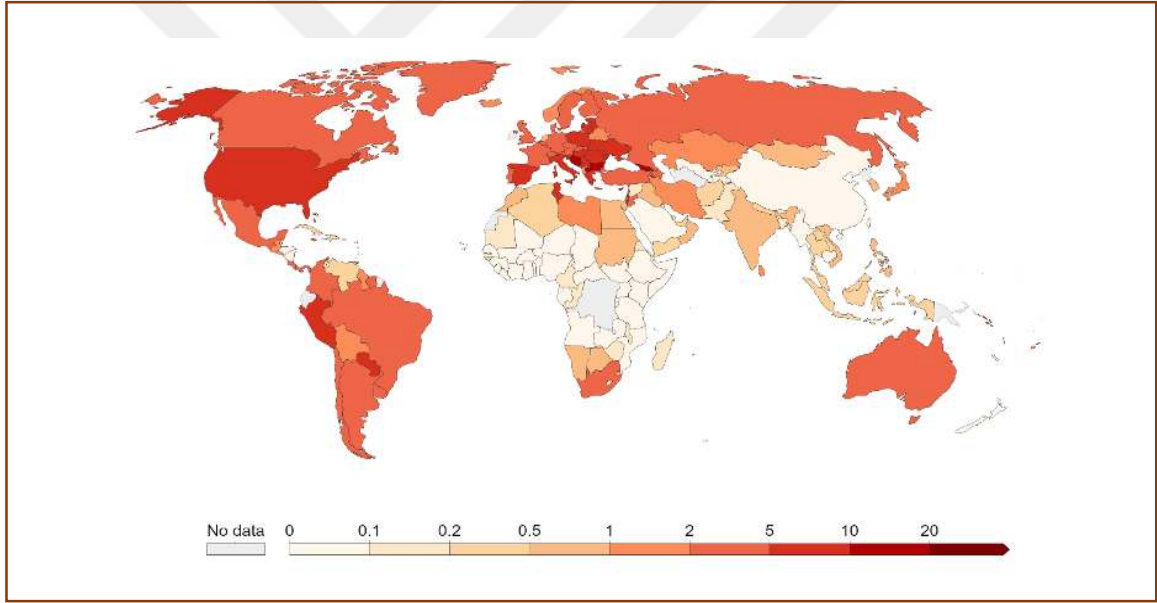
çalışmışlardır. Tahminleme aşamasında Lojistik Regresyon, XGBoost ve Rasgele Orman algoritmalarını kullanmışlardır. Kullanılan algoritmalar karşılaştırmalı olarak incelendiğinde %88 doğruluk, %91 özgüllük %86 AUC değeri ile en iyi performans Rastgele Orman ile elde edilmiştir [16].

Yao ve arkadaşları hastaya ait klinik bilgiler ve kan/idrar testi verilerini kullanarak homojen bir veri seti oluşturup Covid-19 hastalığını tahminlemeyi amaçlamışlardır. Huazhong Bilim ve Teknoloji Üniversitesi'ne Bağlı Tongji Hastanesi'nden 8 klinik, 16 idrar ve 18 kan testi olmak üzere toplamda 100 özellikli bir veri seti elde etmişlerdir. İlk adımda veri üzerinde ön işlem adımlarını uygulayarak veriyi düzenli ve modellenabilir forma dönüştürmüşlerdir. Bir sonraki adımda t-testi uygulayarak veri anlamlılıklarını inceleyip veriden anlamsız özelliklerin çıkarılmasını sağlamışlardır. Modelleme aşamasında DVM, KNN, Rastgele Orman sınıflandırıcılarını kullanmışlardır. %79 doğruluk, %76 duyarlılık, %69 özgüllük ve 0.07 standart sapma oranına sahip DVM ile en anlamlı sonucu elde etmişlerdir [17].

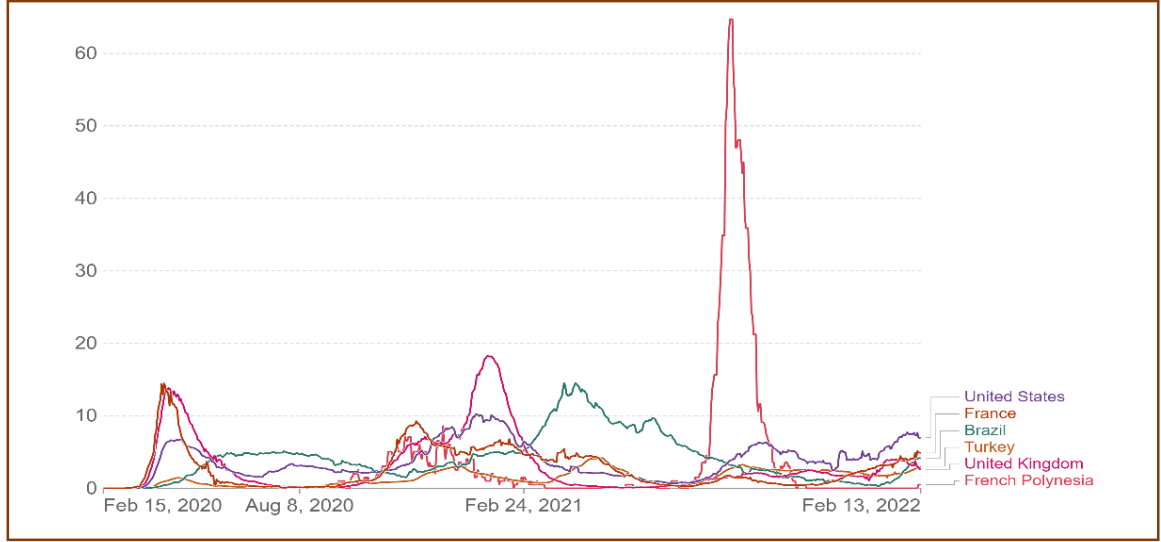
2. SARS-COV-2 PANDEMİSİ VE KLİNİK LABORATUVARLAR

2.1. Covid-19

31 Aralık 2019 tarihinde Çin merkezli olarak ortaya çıkan Covid-19 hastalığı, hızlı yayılımı ve tüm Dünya'yı etkisi altına alması sebebiyle 2020 yılında Dünya Sağlık Örgütü (DSÖ) tarafından pandemi olarak kabul edilmiştir [18]. Bu hastalığın adı ilk olarak 2019 Novel Coronavirüs olarak adlandırılırken daha sonra Covid-19 olarak adlandırılmış ve hala bu isimle literatürde yer almaya devam etmektedir [19]. Dünya genelinde halka açık olarak yayınlanan enfekte hasta sayıları incelendiğinde günden güne salgın etkisini artırmaya devam etmekte, tanımlanamayan kalıcı hasarlara ve ölümlere sebebiyet vermektedir. Gerçekleştirilen incelemelere ait veriler Şekil 2.1, Şekil 2.2 ve Tablo 2.1'de yer almaktadır.



Şekil 2.1. Dünya Geneli Koronavirüs Dağılım Haritası [19]



Şekil 2.2. Koronavirüs Salgını 6 Ülke İçin Vaka Dağılımları [19]

Tablo 2.1. Covid-19 Salgını 6 Ülke İçin Vaka Verileri[19]

ÜLKELER (2020-2022 Şubat)	TOPLAM VAKA SAYISI (2020-2022 Şubat)	VEFAT SAYISI (2020-2022 Şubat)
Amerika Birleşik Devletleri	77.6 Mn	918 B
Hindistan	42.7 Mn	509 B
Brezilya	27.5 Mn	639 B
Fransa	21.2 Mn	132 B
Birleşik Krallık	18.4 Mn	160 B
Türkiye	12.9 Mn	90 B
Tüm Dünya	412 Mn	5.82 Mn

Etiyolojik yapısı incelendiğinde korona virüsler tek zincirli yapıya sahip 4 ana cins ile karşımıza çıkmaktadır. Daha önce tanımlanmış olan farklı 6 tip korona virüse ek olarak mevcut ortaya çıkan Covid-19 bu ailenin 7. Üyesi olmuştur. Kendini genel olarak soğuk algınlığı belirtileri (ateş, öksürük, burun akıntısı vb.) ile gösteren korona virüs salgını, bağlı olduğu SARS-CoV ailesi sebebiyle daha ciddi hastalıklara sebebiyet verebilmektedir [20]. Hastalığa ait bulaş yolu incelendiğinde ilk gerçekleştirilen tanımlamalara göre, virüsün damlacık yolu ile yayılım gösterdiği görülmektedir. Damlacık yolu ile bulaşım genel itibariyle öksürük veya hapsirme gibi durumların ortaya çıkmasıyla oluşmaktadır [18]. Bulaş riskinin kolay ve hızlı olması sebebiyle korona virüs salgını Delta, Omicron vb. varyantları ile hızla yayılım göstermeye devam etmektedir. Hızla yayılım gösteren bu salgının erken teşhisi için aktif olarak birçok yöntem kullanılmaktadır. Bu yöntemlerden bazıları gerçek zamanlı polimeraz zincir reaksiyonu (PCR) testleri, görüntüleme

testleri ile bulguların incelenmesi, biyokimya parametreleri, hastaya ait klinik belirtilerin incelenmesi vb. olarak sıralanabilir [21]. Hastaya ait klinik belirtilerin incelenmesi anlam ifade etse de tek başına Covid-19 tanısı için yeterli değildir. Benzer klinik belirtilere ait farklı hastalıklarda teşhis edilebilir. Bu noktada kesin tanı işlemi için ek testlere ihtiyaç duyulmaktadır. PCR testleri aktif olarak kullanım sağlasa da yanıtın saptanabilir duruma gelerek sonuca dönüştürülmesi zaman almaktadır. Bu duruma ek olarak PCR testleri sağlık sektörüne maliyet yükü oluşturmaktadır. Görüntüleme testleri ile hastalığa ait görüntülerin incelenmesi durumunda Covid-19'un diğer bazı hastalıklar ile benzer bulgulara sahip olması sebebiyle yanlış payının yüksek olduğu görülmektedir. Diğer tüm testlerin yetersiz kaldığı durumlarda laboratuvar testleri hastalığın teşhisinde belirgin rol oynamaktadır. Laboratuvar testleri bir hastaya ait sağlık durumunu değerlendirmek için ihtiyaç duyulan testlerdir. Hastalığın tanı aşamasında rutin olarak tam kan sayımı(CBC), pıhtılaşma bulgularını araştıran testler (PT, D-Dimer) ve CRP testlerine başvurulmaktadır [22]. Klinik laboratuvar testleri hızlı ve ulaşılabilir yapısı ile Covid-19 hastalığının şiddetinin izlenmesi, uygun tedavi yöntemlerinin seçilmesi vb. birçok kritik alanda önemli bir konumda yer almaktadır [23]. Covid-19 hastalığının tanısında önem arz eden bazı biyokimya parametreleri Tablo 2.2'de verilmiştir.

Tablo 2.2. Covid-19 Tanısında Önemli Biyokimya Parametreleri[23]

Parametre	Artan Eğilim	Azalan Eğilim
Beyaz Kan Hücresi Sayısı	x	
Nötrofil Sayısı	x	
Lenfosit Sayısı		x
Albümin Sayısı		x
Laktat Dehidrojenaz(LDH)	x	
Alanin Aminotransferaz (ALT)	x	
Aspartat Aminotransferaz(AST)	x	
Toplam Bilirubin	x	
D-Dimer	x	
Kreatinin	x	
C-reaktif protein(CRP)	x	

Tablo 2.2 incelendiğinde artan ve azalan eğilime sahip biyokimya parametrelerindeki değişimlerin Covid-19 hastalığı üzerindeki etkileri görülmektedir. Literatür çalışması kapsamında incelenen çalışmalara bakıldığında Tablo 2.2’de verilen parametreler içerisinde CRP, D-Dimer, ALT, AST ve LDH parametrelerinin önemli bir konuma sahip olduğu görülmektedir. Bu parametrelere ait eğilimler daha kolay ve daha erişilebilir kaynaklar olması sebebiyle tanı, takip ve tedavi aşamasında erken tespit sürecini hızlandırmada önemli rol oynamaktadır.

2.2. Klinik Laboratuvarlar

Laboratuvarlar, yalnızca test sonucu üretmeyen klinik karar vermeyi destekleyerek hastalığın tanısı, takibi ve hastalığın önlenmesi gibi çeşitli hizmetler içeren bilim dallarıdır. Literatür genelinde inceleme gerçekleştirildiğinde tıbbi kararların büyük bölümünün laboratuvar sonuçları dikkate alınarak verildiğini görülmektedir. Bu nedenle sağlık hizmetinin yürütülmesinde laboratuvarlar büyük öneme sahiptir. Gerçekleştirilen bir çalışmaya göre hekim tarafından alınan klinik kararların %70’i hastaya ait bulgulara dayanırken, %94’ü laboratuvar verilerine dayanmaktadır [24].

Klinik laboratuvarların temel görevi hastalığın tanısı, önlenmesi ve tedavisine yönelik bilgi sağlayarak klinik karar vermeyi destekleyip hekimlere ve hastalara yardımcı olmaktır [22]. Klinik Laboratuvarlar, temel görevlerini yerine getirmek amacıyla insan vücudundan alınan materyallere ilişkin biyolojik, patolojik, kimyasal vb. analizleri yapabilen ve bu analizlere ait sonuçları yorumlayıp ileriye yönelik tavsiye dâhil bir laboratuvara ait tüm araştırma kollarını kapsayan laboratuvarlardır [25]. Klinik laboratuvarlarda gerçekleştirilen işlemler ile çeşitli maddelerin ve mikroorganizmaların varlık ve yoklukları açıklanabilmektedir. Klinik laboratuvarlar kritik önem taşıyan işlemleri içerdikleri için gerekli testleri katı ve mantıklı bir şekilde gerçekleştirirler.

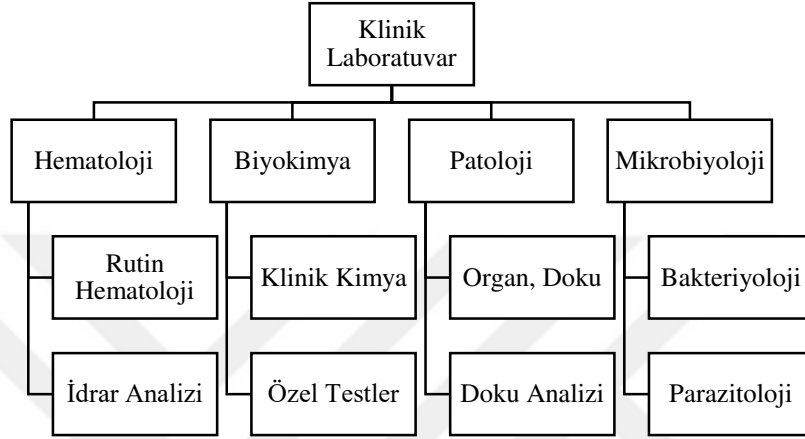
Klinik laboratuvarlarda yapılan işlemler dışarıdan görünen durumun aksine sadece örneklerle ait analizlerin gerçekleştirilip sonuç vermeden fazlasıdır. İncelenen olan numunelerin doğru zamanda, doğru ortamda ve doğru kişiden alındığına emin olmak, en uygun, en doğru, en güvenilir metodu uygulamak ve hasta-hekim yönlendirmesini doğru bir şekilde yapabilmek klinik laboratuvarların önemli görevleri arasındadır. Klinik laboratuvarlar bu görevleri yerine getirirken dışarıdan alınan biyolojik örnekler üzerinde analiz işlemlerini gerçekleştirir. Bu amaç doğrultusunda alınan biyolojik örnekler şunlardır:

- Kan Örnekleri
- İdrar Örnekleri
- Doku Parçaları

Yukarıda belirtilen genel olarak alınan biyolojik örneklerle ek olarak daha az sıklıkla alınan analiz örnekleri şunlardır [26]:

- Eklem Sıvısı
- Göğüs ve karın boşluğu sıvı örnekleri

Tüm işlevleri içerisinde barındıran bir klinik laboratuvar dört bölümden oluşur. Bu dört bölüm; biyokimya, hematoloji, mikrobiyoloji ve patolojidir. Şekil 2.3'te örnek bir klinik laboratuvar şeması verilmiştir [27].



Şekil 2.3. Klinik Laboratuvar Bölümleri [27]

2.2.1. Klinik Hematoloji

Klinik Hematoloji;

- Kanın içyapısını,
- Kan hücrelerine ait şekil ve yapıları,
- Kanın görev ve işleyişini

İnceleyen laboratuvar dalıdır. Bu laboratuvarlarda yapılan testler hematolojik testler olarak adlandırılır ve hematologlar tarafından incelenir [28]. Odak noktası kanın bileşenlerini değerlendirmek olan bu laboratuvarlarda tam kan sayımı, B12, demir vb. parametre düzeyi kontrolleri gerçekleştirilmektedir.

2.2.2. Klinik Biyokimya

Klinik Biyokimya bölümü laboratuvarların büyük bir kısmını kapsamaktadır. Vücuda ait çeşitli sıvı, salgı ve doku örnekleri üzerinde gerçekleştirilen analiz çalışmaları ile hastalığın tanı ve süreç ilerleyişini inceleme adımlarında kullanılan laboratuvarlar dalıdır. Klinik biyokimya laboratuvarlarında bulunan ve analiz çalışmasına yardımcı olarak kullanılan araçlar birden fazla testi aynı anda gerçekleştirebilme özelliğine sahiptirler. Bu yönüyle klinik biyokimya en sık

başvurulan laboratuvar dallarından biridir. Biyokimya laboratuvarlarında analizi gerçekleştirilen başlıca testler;

- Glikoz,
- Kolesterol,
- Trigliserid,
- Hormon testleri şeklinde sıralanabilir [27].

2.2.3. Klinik Patoloji

Patoloji bilim dalı, organ ve dokularda oluşan yapısal ve fonksiyonel bozuklukları inceleyen bilim dalıdır. Klinik patoloji laboratuvarları organ ve doku yapılarını bozan insan gözü ile ayırt edilemeyen, anlaşılamayan etkenlerin araştırıldığı ortamlardır. Klinik biyokimya laboratuvarlarında kullanılan yardımcı araçların oynadığı büyük rolün aksine, patoloji laboratuvarlarında önemli rol insan etmenidir. Doku örneklemeleri vb. araç gerektiren işlemler dışında patoloji laboratuvarları alanında uzman çalışanlara ihtiyaç duyar. Klinik patoloji uzman çalışan odaklı olarak öne çıkan en belirgin laboratuvar dalıdır. Patoloji laboratuvarlarında gerçekleştirilen testlere endoskopi ve klonoskopi örnek olarak verilebilir [28].

2.2.4. Klinik Mikrobiyoloji

Klinik mikrobiyolojide canlıda hastalığa yol açan gözle görülemeyecek kadar küçük mikroorganizmalar ele alınır. Bu mikroorganizmalar izole edilerek izlenir ve gerçekleştirilen testlere paralel doğrultuda sonuçlar yorumlanır. Bu doğrultuda mikrobiyoloji laboratuvarları teşhis ve tedavi aşamalarında süreçlerin önemli bir parçasıdır. Klinik mikrobiyoloji laboratuvarlarında hastadan uygun şartlarda alınan örnekler (kan, idrar, sürüntü vb.) ile genellikle virüs, bakteri ve mantar kaynaklı hastalıkların teşhisine ve kontrolüne yönelik analiz çalışmaları gerçekleştirilir [28].

Klinik laboratuvarlar hastalıkların teşhis ve tedavisinde gerek duyulan tetkik ve örnek analizlerinin en ileri teknoloji kullanılarak yapıldığı birimlerdir. Laboratuvar hizmetlerinde kaliteli hizmet sağlamak için teknolojinin önemi büyüktür. Teknolojinin önemi ne kadar büyükse, bu teknolojinin iyi bir şekilde kullanılmasının önemi de o kadar büyüktür. Günümüzde hızla gelişen teknolojiler ve yeni testlerin geliştirilmesi sonucunda, hastalıkların erken teşhisinde, doğru ve etkili tedavi edilmesinde, laboratuvarlar önemli bir konuma gelmiştir. Klinik laboratuvarların işleyişi, modern hayatın değişimine paralel olarak hızlı gelişim kaydetmiştir. Önceleri elle gerçekleştirilen işlemler, günümüzde otonom hale gelmiştir [29]

3. YAPAY ZEKÂ

Beynin nesneler arası ilişkiyi kavrama yeteneğiyle, bilgiyi alarak bir amaca yönelik çözüm bulma, karar verebilme ve doğru analiz edebilme yeteneği zekâ olarak adlandırılmaktadır [30]. Matematiksel, pratik, evrensel gibi dallara ayrılan yapısıyla karşılaştığımız yeni bir olgu ve olaya karşı düşünebilme, öğrenme ve durumu kavrayabilme yeteneği zekâ ile gerçekleştirilmektedir. Bilgisayarların insan zekâsına benzer yapıda yeteneği kazanabilmesi için zekâ bu yönde modellenerek yapay zekâ kavramı ortaya çıkmıştır [30].

İnsan beyni biyolojik yapısı gereği hızlı anlayabilme, yorumlayabilme ve değerlendirebilme yeteneğine sahiptir. Daha önce karşılaşılmayan durumlarda geçmiş deneyimlerden elde edilen bilgiler doğrultusunda insan beyni bu durumu algılayarak hızlı dönüş sağlayabilmektedir. Yapay zekâ bu noktada insan beyninin bilgisayarlara yansması olarak tanımlanabilir. Yapay zekâ insan gibi davranış sergileyebilme, problem çözebilme, yeni fikirler üretebilme, tecrübelerden öğrenme, sayısal mantık yürütebilme, ses ve görüntü algılayabilme gibi birçok insana özgü yeteneğe sahip yapıdır [31].

Yapay zekâ birçok aşamadan geçerek günümüzdeki güncel haline ulaşmıştır. Yapay zekâ geçmişten günümüze insan gibi düşünebilen, rasyonel düşünebilen, insan gibi hareket edebilen ve rasyonel hareket edebilen yapılarıyla evrilerek gelmiştir. Güncel olarak günümüzde rasyonel hareket edebilen yapay zekâ yapısı kullanılmaktadır.

Yapay zekâ insanın düşünme yetisine bağlı olarak güçlü ve zayıf yapay zekâ olarak sınıflandırılmaktadır. Makinelerin programlanarak insana ait zekâ kavramıyla zeki davranış gösterme özelliğini kullanması zayıf yapay zekâ olarak tanımlanır. Temelde yatan prensip makinelerin zeki gibi davranma özelliğinin varoluşudur. Zayıf yapay zekâ belirli bir işi yerine getirmek amacı ile tasarlanır. Örneğin; bilgisayara karşı oynanan oyunlar göz önüne alındığında insan zekâsı düşünerek, kendi kararlarını vererek hamlelerini gerçekleştirir. Fakat bilgisayar söz konusu olduğunda bilgisayar kendine öğretileni yapar. Düşünerek ya da zekâsını kullanarak karar veremez. İnsan tarafından ne ile beslendiyse karşı hamlesini o şekilde gerçekleştirir [32]. Güçlü yapay ise zekâ insan gibi düşünme yeteneğine sahip olan ve düşünme yetisiyle görevleri yerine getirdiği düşünülen yapıdır. Hiçbir müdahale olmadan problemi çözebilmeli mantığıyla tasarlanmışlardır. Güçlü yapay zekâ sistemlerine sürücüsüz araçlar vb. akıllı yapılar örnek olarak verilebilir. Güçlü yapay zekâ sisteminin temel özellikleri planlama, akıl yürütebilme, karar verebilme ve iletişim kurabilme olarak sıralanabilir.

Yapay zekâ her ne kadar doğal zekâdan ilham alsa da yapısal farklılığı göz önüne alındığında birbirinden farklı birçok özelliğe sahiptir. Yapay zekâda bilgi kalıcıdır fakat doğal zekâ bilgiyi unutabilir. Yapay zekâ sadece öğretileni düşünebilirken doğal zekâ fikir üretme mekanizmasına sahiptir. Doğal zekâ yapısı gereği maliyet gerektirmez fakat yapay zekâ maliyetlidir. Doğal zekâ

anlık olarak sürekli deęişim gösterdiği için raporlanması zordur. Yapay zekâ ise depolandığı için raporlanabilir. Doğal zekâda dışa bağımlılık yoktur fakat yapay zekâ doğrudan insan zekâsına baęlı olarak gelişim gösterir [33].

3.1. Yapay Zekâ Alt Dalları

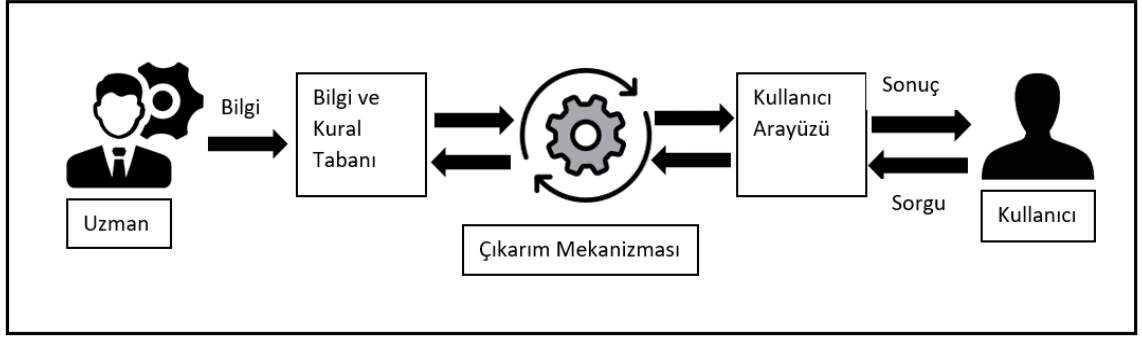
Yapay zekâ teknolojide zekâ kavramının temelini oluşturmakla beraber içerisinde birden çok alt dal barındırır. Yapay zekâ alt dalları;

- Uzman Sistemler
- Genetik Algoritmalar
- Bulanık Mantık
- Yapay Sinir Ağları
- Makine Öğrenmesi

olarak sıralanabilir.

3.1.1. Uzman Sistemler

Uzman Sistemler, belirli bir alana ait ilgili bilgilerle donatılmış ve problemlere o alanda uzman bir kişinin sunduęu şekilde çözümler sunabilen bir yapay zekâ alt dalı olarak tanımlanabilir. Uzman sistemler insan uzmanlığı gereken fonksiyonları yerine getirerek karar aşamasında kişilere destekleyici bir rol sunarken aynı zamanda uzman olmayan kişiler tarafından problemlerin çözümü için kullanılır. Mantıklı çözümler üretebilme yeteneğine sahip iyi geliştirilmiş bir uzman sistem alanında uzman kişiler gibi teşhis etme, yorumlama, kontrol sağlama, tavsiyede bulunma davranışlarını sergiler [34]. Yapay zekânın bir alt dalı olan uzman sistemler genel yapay zekâ davranışından farklı olarak sadece alanında uzman kişilerin davranışını benimser ve taklit eder. Uzman sistemler iki adımlı bir süreç sonucunda oluşur. İl adımda, uzman sisteme öğretilmek istenen bilgiler alanında uzman kişilerden toplanır ve ikinci adımda bu bilgiler uzman sistemlere öğretilerek bir sonraki problemin çözümünde kullanılmaya hazır hale getirilir. Şekil 3.1’de bir uzman sisteme ait çalışma mantığının genel yapısı verilmiştir.



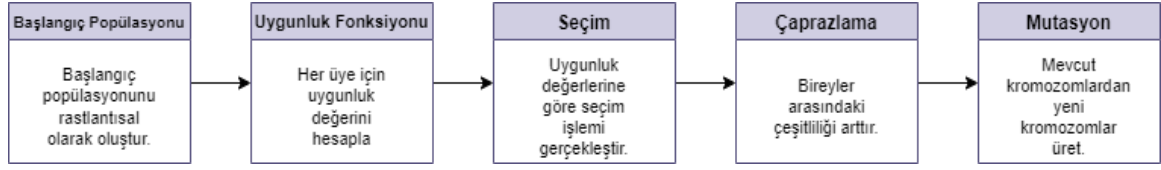
Şekil 3.1. Uzman Sistem Genel Yapısı [34]

Şekil 3.1’de genel yapısı verilen uzman sistemlerin sağladığı faydalardan bazıları şu şekildedir [35]:

- Maliyet tasarrufu sağlar.
- Kullanıcıya hazır bilgi sağlar.
- Hızlı cevap verme yetenekleri bulunduğu için verimlilik artar.
- Bilgiler gerçek dünyada unutilan, kaybolan bilgilerin aksine uzman sistemler için kalıcıdır.
- Hata oranını minimize ederek kalite artışı sağlarlar.
- Sonuca nasıl ulaşıldığına dair bilgiyi gerekirse alanında uzman biri gibi açıklayabilirler.

3.1.2. Genetik Algoritmalar

Genetik algoritmalar, canlı sistemine benzer bir şekilde öğrenme prensibini benimseyen ve doğal seçim ile hayatta kalmayı başaran bireylerin davranışını temel alan tekniktir [35]. Genellikle karmaşık optimizasyon problemlerinde kullanılan bu teknik rastgele oluşturduğu çözümler arasından probleme en uygun çözümü sunar. Genetik algoritmalar problemin hem çözüm hem de modelleme aşamasında kullanılmaktadır. Genetik algoritmalar modelleme ve çözümleme işlemlerini canlı dünyayı taklit ederek gerçekleştirirler. Gerçek dünyayı önce modeller ardından bu modellemelerde karşılaşılan sorunlara çözüm üretirler. Genetik algoritmaların yapısal özellikleri incelendiğinde kromozom, popülasyon, çaprazlama, mutasyon kavramlarını barındırdıkları görülmektedir. Şekil 3.2’de genetik algoritmaların adımlarını içeren genel yapı verilmiştir.



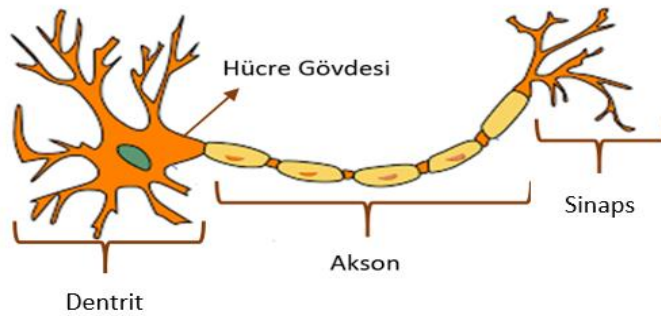
Şekil 3.2. Genetik Algoritma Akış Şeması [35]

3.1.3. Bulanık Mantık

Klasik mantıkta bir olay 0 ya da 1'dir. Yani bir olgu ya doğru ya yanlıştır. Bulanık mantıkta ise bir olgu geniş karar aralığına sahiptir. Yani bulanık mantıkta kesinlik yoktur, yaklaşım vardır. Bulanık mantığın temelini küme ve alt küme kavramları oluşturmaktadır. Bulanık varlıkların kümesinde her varlığa ait bir üyelik derecesi bulunur ve bu derece 0-1 aralığındadır [36]. Bulanık mantık model oluşturulması güç olan sistemlerde tercih edilir. Bunun sebebi kesinlikten ziyade bulanıklık kavramının getirdiği biraz doğru ya da biraz yanlış gibi bir seçilimin mevcut olmasıdır. Bulanık mantığa dayalı sistemler oluşturulurken dört temel bölüm ele alınır. Bu bölümler; veri tabanı, çıkarım motoru, bulandırma ve durulama ara yüzleridir [37].

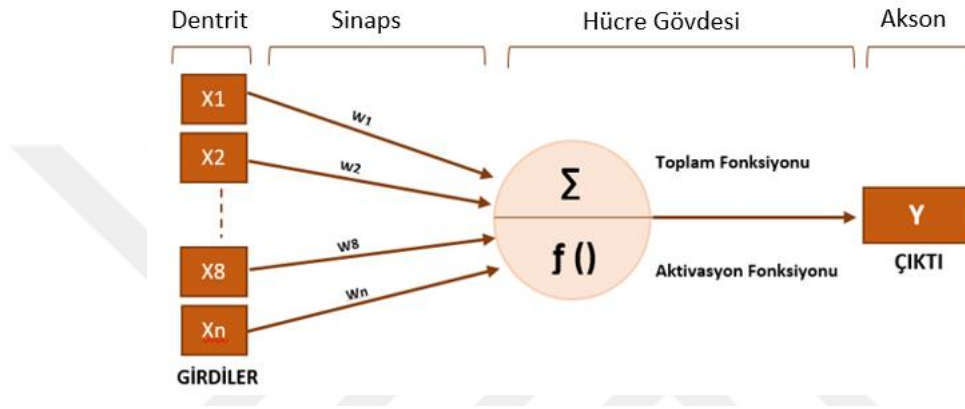
3.1.4. Yapay Sinir Ağları

Yapay Sinir Ağları (YSA), biyolojik sistemi örnek alan, nöronların bir araya gelerek oluşturduğu, insan beyninin bilgi işleme yeteneğinden esinlenerek geliştirilen teknolojidir. YSA'da öğrenme işlevi daha önce elde edilen tecrübelerle ilgili olarak gelişim gösterir. İnsanlarda öğrenme olayı nöronlar arasındaki sinaptik bağlantılar ile gerçekleşir. Dendritlerden gelen uyarıları alan nöronlara girişler yerleştirilerek bu girişlerin belirli bir eşik değerini aşması durumunda bir etki durumu oluşur [38]. Bir nöronun basit yapıda oluşumu bu şekilde gerçekleşir. Şekil 3.3'de bir nöron yapısına ait örnek görsel verilmiştir.



Şekil 3.3. Sinir Hücresi Yapısı [39]

Şekil 3.3’de verilen yapı özetlenecek olursa sinaptik sinyaller dentritler tarafından girdi olarak alınır, ardından hücre gövdesinde bu sinyaller işlenerek aksonlar tarafından denetimi gerçekleştirir ve mevcut sinyalin farklı bir sinir hücresine iletimi gerçekleşir. Verilen sistemde öğrenme, algılama vb. insana ait özellikleri barındıran birim sinapslardır [39]. Basit bir nöronun çalışma prensibinden yola çıkılarak geliştirilen YSA modelinde genel olarak nöronlar 3 katman halinde ve her katman içinde paralel olarak bir araya gelerek yapay sinir ağlarını oluştururlar. Şekil 3.4’de temel bir yapay sinir hücresine ait yapı biyolojik sinir hücre yapısı ile karşılaştırmalı olarak verilmiştir.



Şekil 3.4. Yapay Sinir Hücresi [39]

Şekil 3.4’de verilen yapay sinir hücresine ait tanımlamalar aşağıda verilen şekildedir:

Girdiler: Dış dünyadan elde edilen ağın öğrenmesi gereken bilgileri içerir ve ağa giriş değerleri olarak verilir.

Ağırlıklar: Dış dünyadan elde edilen ve ağa verilen bilgilerin önem derecelerini ve hücre üzerindeki etkilerini gösteren değerlerdir. Her girdi değeri kendine ait bir ağırlık değerine sahiptir ve bu değer 0, 0’dan küçük veya 0’dan büyük olabilir.

Toplam Fonksiyonu: Net girdiyi hesaplamaya yarayan fonksiyondur. Maksimum, minimum, toplam, çarpım vb. birçok fonksiyonu içerisinde barındırır ve problemin çözümüne uygun olarak gerekli fonksiyonun kullanılmasına olanak sağlar. Genellikle toplam fonksiyonu tercih edilir.

Aktivasyon Fonksiyonu: Toplam fonksiyonu tarafından elde edilen net girdi değerini işleyerek hücrenin üreteceği çıktıyı belirler. Yapay sinir ağlarının doğrusal olmama özelliği, doğrusal olmayan aktivasyon fonksiyonlarından gelir. Genel olarak kullanılan aktivasyon fonksiyonları sigmoid, hiperbolik tanjant ve basamak fonksiyonu olarak sıralanabilir.

Çıktı: Aktivasyon fonksiyonu tarafından elde edilen çıktı değeridir. Çıktı değeri başka bir hücrenin girdi değeri olarak kullanılabilir.

Şekil 3.3 ve Şekil 3.4’de verilen hücre yapılarına göre biyolojik hücre yapısında bulunan her bir elemana karşılık gelen yapay sinir hücre yapısı elemanı Tablo 3.1’de verilmiştir.

Tablo 3.1. Biyolojik Sinir Hücresi ve Yapay Sinir Hücresi [40]

Biyolojik Sinir Hücresi	Yapay Sinir Hücresi
Dentrit	Toplama Fonksiyonu
Hücre Gövdesi	Transfer Fonksiyonu
Aksonlar	Çıktı
Sinapslar	Ağırlıklar

Yapay Sinir Ağlarının Özellikleri

Yapay sinir ağlarının özellikleri doğrusal olmama, öğrenme yeteneğine sahip olma, eksik veriler ile çalışabilme, kolay adaptasyon becerisi, uyarlanabilir yapıda olma, hata toleransına ve genelleme yeteneğine sahip olma şeklinde sıralanabilir [40].

Yapay sinir ağları haberleşme, savunma teknolojileri, bankacılık, elektronik, eğlence, sağlık, robotik, görüntü ve ses tanıma, güvenlik vb. birçok alanda aktif olarak kullanılmaktadır. Bu alanlarda yapay sinir ağlarının kullanım amaçları genel hatları ile şu şekildedir [40]:

Tahminleme: Yapay sinir ağları tahminleme amacıyla kullanıldığında eldeki girdi değerlerinden çıktı tahminleri üretmeye çalışır.

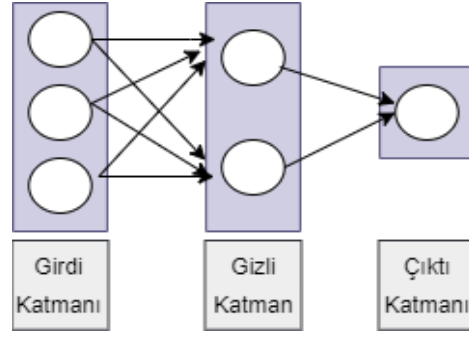
Sınıflandırma: Sınıflandırma amacıyla kullanılan yapay sinir ağları elde bulunan eğitilmiş girdileri kullanarak yeni gelecek olan girdinin hangi sınıfa ait olduğunu bulma amacıyla aktif olarak kullanılır.

Veri Yorumlama: Daha önceden eğitilen ağı ait girdileri analiz ederek olaylar karşısında bu bilgileri kullanarak yeni yorumlar üretir.

Veri İlişkilendirme: Yapay sinir ağları veri ilişkilendirme ile öğrenilen bilgiler doğrultusunda konularla ilişkilendirme gerçekleştirerek eksik bilgilerin tamamlanmasını sağlar.

Yapay Sinir Ağlarının Yapısı

Sadece girdi ve çıktı katmanlarından oluşan yapıda bulunan ağlar tek katmanlı yapay sinir ağları olarak adlandırılır. Çok katmanlı ağların yapısında ise girdi, birden fazla gizli ve çıkış katmanı bulunur. Yapay sinir ağları çalışma prensibi olarak sınıflandırma algoritmaları ile benzer yapıda davranış gösterirler. Eğitim aşamasında sistem girdilerin ağırlıklarına göre eğitilir ve çıktı hesaplanır. Test aşamasında ise sisteme eğitim dışı yeni örnekler vererek ağın doğru tahminleme gerçekleştirmesi beklenir [41]. Bir yapay sinir ağına ait temel yapı Şekil 3.5’de verilmiştir.



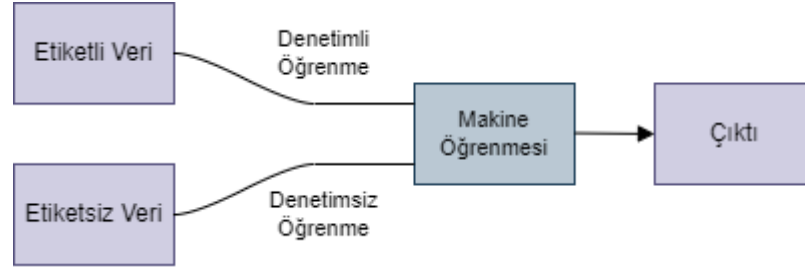
Şekil 3.5. Yapay Sinir Ağı Yapısı [41]

Şekil 3.5’de verilen yapay sinir ağı yapısı incelendiğinde bir yapay sinir ağının, girdi katmanı, ara katman ve çıktı katmanı olmak üzere toplamda 3 katmandan oluştuğu görülmektedir. Girdi katmanında dış dünyadan elde edilen bilgiler alınarak bir sonraki katman olan ara katmana iletilir. Girdi katmanından elde edilen bilgiler ara katmanda işlenerek çıktı katmanında iletilir. Bir ağda tek bir ara katman bulunabileceği gibi birden fazla ara katmanda bulunabilir. Çıktı katmanına iletilen işlenmiş bilgiler çıktı katmanında tekrardan işlenerek, girdi katmanında sunulan örnek için üretilmesi gereken çıktı üretilir. Çıktı katmanı nihai sonuçları elde edilen katman olarak tanımlanabilir. Yapay sinir ağları yönelimlerine göre ileri beslemeli ve geri beslemeli yapay sinir ağları olmak üzere ikiye ayrılırlar. İleri beslemeli sinir ağlarında doğrusal bir ilişki söz konusudur. Temel yapıda anlatılan girdi katmanı-ara katman-çıktı katmanı ilişkisi ileri beslemeli yapay sinir ağlarında doğrusal olarak ilerler ve her katmandaki yapay sinir hücresi bir önceki yapay sinir hücresinden beslenir [41]. Geri beslemeli yapay sinir ağları, ileri beslemeli yapay sinir ağlarının tersi yönde özelliklere ve davranışlara sahip sinir ağları olarak tanımlanabilir. Geri beslemeli yapay sinir ağlarında akış geriye doğru ilerler.

3.1.5. Makine Öğrenmesi

Makine öğrenmesi, insana ait öğrenme yapısından esinlenerek geliştirilmiş, çıkarımlarını matematiksel ve istatistiğe dayalı yöntemler kullanarak gerçekleştiren ve yaptığı bu çıkarımlar ile tahminlemede bulunan yapay zekânın bir alt dalıdır [42]. Makine öğrenmesi uygulamaları genel başlıkları ile bilgisayarlı görme, tahminleme, sınıflandırma ve doğal dil işleme alanlarında gerçekleştirilmektedir. Tüm bu uygulama alanları için veri makine öğreniminde önemli bir role sahiptir. Veri madenciliği ve makine öğrenmesi alanları iç içe kullanım sağlanan alanlardır. Veri madenciliğinde, veriden bilgi çıkarımı gerçekleştirebilmek için makine öğrenimi yöntemleri kullanılmaktadır. Bu duruma paralel olarak makine öğrenimi verileri işlemek için veri madenciliği yöntemlerini kullanmaktadır [42]. Makine öğrenimi modelleri, Şekil 3.6’da gösterildiği gibi

verinin türüne (etiketli ya da etiketsiz) bağlı olarak denetimli ve denetimsiz öğrenme olarak ikiye ayrılmaktadır.



Şekil 3.6. Makine Öğrenimi Modelleri [42]

Denetimsiz öğrenme modellerinde girdi verileri etiketli değildir ve bilinen bir sonucu yoktur. Amaç veri kümesindeki verileri yorumlayarak ortak noktaları bulmak ve bunlardan anlamlı bir veri elde edebilmektir. Denetimsiz öğrenme problemlerinde örneklere ait çıktılar bilinmediği için problemler sınıflandırma ya da regresyon problemi değildir. Denetimsiz öğrenme problemleri “kümeleme”, “birliktelik kuralı” ve “boyut azaltma” olarak üçe ayrılır. Verilerin yakınlık, uzaklık, benzerlik gibi ölçütlere göre analiz edilerek sınıflara ayrılmasına kümeleme denir [43]. Kümeleme, gözlemleri homojen bölgelere ayırır. Kümelemenin asıl amacı, farklı kümelerdeki noktalar benzer olmamakla birlikte bir küme içindeki noktalar benzer olacak şekilde veri noktaları grupları oluşturmaktır. Birliktelik kuralı, geçmiş verilerin analiz edilerek bu veriler içindeki birliktelik davranışlarının tespiti ile geleceğe yönelik çalışmalar yapılmasını destekleyen bir yaklaşımdır. Boyut azaltma işlemi ise değişken sayısını farklı yöntemler kullanarak anlamlı olarak azaltma işlemi olarak tanımlanabilir.

Denetimli öğrenmede, veri kümemizin ne olduğunu ve bu verilerden istediğimiz çıktının ne olması gerektiğini biliriz. Denetimli öğrenme modelinde sistem etiketli veriler kullanılarak eğitilir ve öğrenme etiketli veriler ile sağlanır. Eğitilmiş sisteme test verileri verilerek sistemin bir tahmin üretmesi beklenir [44]. Denetimli öğrenme problemleri sınıflandırma ve regresyon problemleri olarak ikiye ayrılmaktadır. Regresyon problemlerinde her bir gözlem için öğrenilen veriye göre tahmin üretilir. Problemden sayısal değerler verilir ve bu değerler arasındaki ilişkiye göre matematiksel bir bağıntı oluşturulur. Ardından sisteme gelen girdilere göre sonuç tahmini gerçekleştirilir. Bir eğitim verisi üzerinden bir metod ile model elde edip, elde edilen modeli tahminde kullanma işlemine sınıflandırma denir. Başka bir deyişle sınıflandırma mevcut veriden elde ettiği tecrübe ile yeni gelen verileri tahmin etme işlemidir. Bir sınıflandırma işleminde kullanılan veri temel olarak eğitim ve test verisi olmak üzere iki kategoriye ayrılır. Sınıflandırma modeli eğitim verisi ile eğitilir ve test verisi ile değerlendirilir. Asıl amaç her örnek için test verisi sisteme verildiğinde hedef sınıfı doğru tahmin edebilmektir.

Makine Öğrenimi Algoritmaları

Makine öğrenimi denetimli ve denetimsiz olmak üzere birçok algoritma barındırmaktadır. Bu algoritmaların genel amacı veriden bilgi elde etmektir. Sıklıkla kullanılan makine öğrenimi algoritmalarından bazıları KNN, DVM ve Karar Ağacı algoritmalarıdır.

K-En Yakın Komşu Algoritması

KNN algoritması, T. M. Cover ve P. E. Hart tarafından önerilen, örnek veri noktasının bulunduğu sınıfın ve en yakın komşunun, k değerine göre belirlendiği bir sınıflandırma yöntemidir [45]. KNN gürültülü eğitim verilerine karşı dirençli olması sebebiyle en çok tercih edilen sınıflandırma algoritmalarından biridir. Fakat çok sayıda bellek alanına ihtiyaç duyması vb. gibi sebeple KNN algoritmasının dezavantajlarını beraberinde getirmektedir. KNN algoritması denetimli bir öğrenme algoritmasıdır ve diğer denetimli öğrenme algoritmalarının aksine, bu algortmada eğitim aşaması mevcut değildir. KNN algoritması eğitim verilerini öğrenmek yerine bu verileri ezberler. KNN, sınıfları belli olan bir veri kümesinden yararlanarak işlem yapmaktadır. Veri setine katılacak olan yeni örneğin, mevcut verilere göre uzaklığı hesaplanıp, k sayıda yakın komşuluğuna bakılır ve hangi sınıfa ait olduğu en yakın k komşuya göre belirlenir [45]. Uzaklık hesaplama işleminde genellikle Öklid, Manhattan, Minkowski ve Hamming uzaklık fonksiyonları kullanılır.

$$\sqrt{\sum_{i=1}^k (x_i - y_i)^2} \quad (3.1)$$

$$\sum_{i=1}^k |x_i - y_i| \quad (3.2)$$

$$(\sum_{i=1}^k (|x_i - y_i|)^q)^{1/q} \quad (3.3)$$

x_i = Uzaklık hesaplaması gerçekleştirilecek noktalardan birincisi

y_i = Uzaklık hesaplaması gerçekleştirilecek noktalardan ikincisi

q = Norm

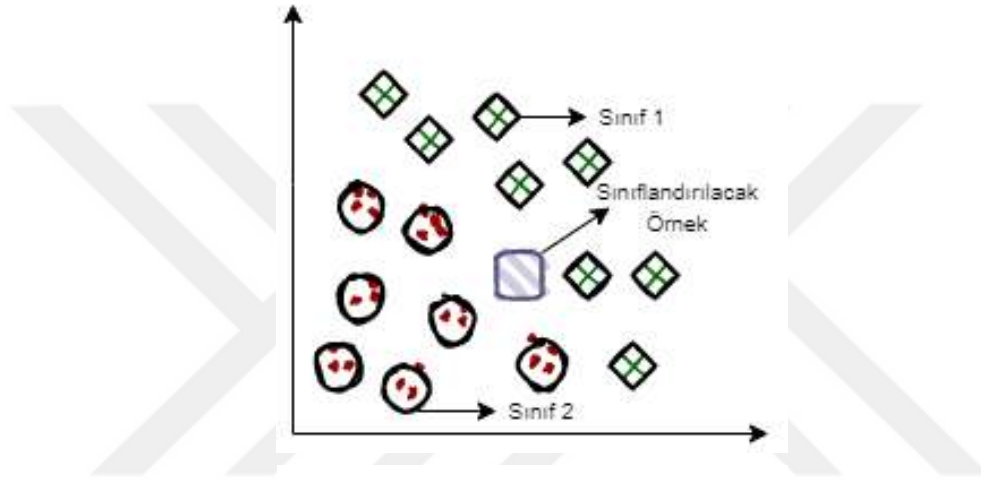
olarak ifade edilmektedir.

Denklem 3.1 Öklid, denklem 3.2 Manhattan ve denklem 3.3 Minkowski uzaklıklarının matematiksel formüllerini temsil etmektedir.

KNN algoritması adımları şu şekildedir [45]:

1. İlk olarak k parametresi belirlenir. k parametresi verilen noktaya en yakın komşuların sayısıdır.
2. Yeni verilen girdi örneğinin mevcut verilere olan uzaklıkları tek tek hesaplanır.
3. Uzaklıklar sıralanır ve en minimum uzaklıklar ele alınarak yakın komşular bulunur.
4. En yakın komşuların hangi sınıfa ait oldukları belirlenir, en sık tekrar eden sınıf seçilir.
5. Seçilen sınıf yeni verilen örneğin sınıfı olarak kabul edilir.

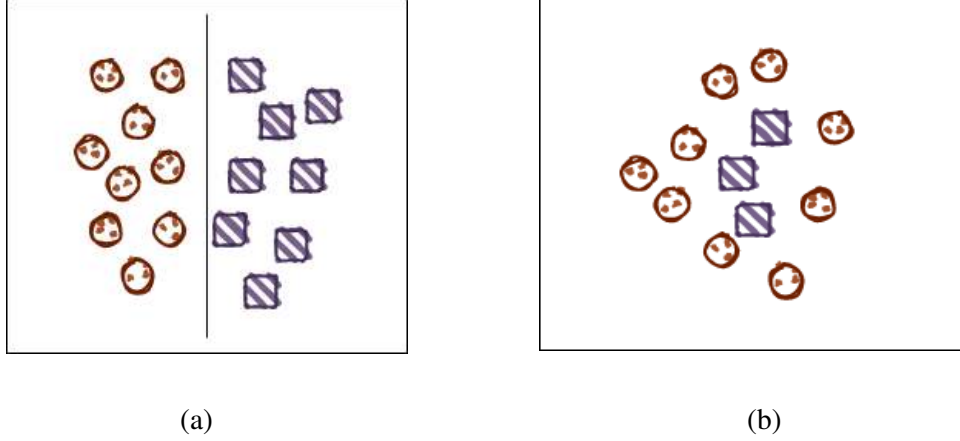
Yukarıda adımları verilen KNN algoritmasına ait örnek bir yapı Şekil 3.7’de verilmiştir.



Şekil 3.7. KNN Algoritması

Destek Vektör Makineleri

Bir destek vektör makinesi, verileri optimal olarak iki kategoriye ayıran n -boyutlu bir hiperdüzlem oluşturmaktadır [46]. DVM algoritması denetimli bir öğrenme algoritmasıdır. Değişkenler arasındaki örüntülerin bilinmediği veri setlerindeki sınıflandırma problemleri için önerilmiş bir yöntemdir. İstatistiksel öğrenme teorisine ve yapısal risk minimizasyonuna dayanmaktadır. DVM hiçbir parametre almayan bir sınıflandırıcıdır. DVM’de amaç, farklı sınıflara ait destek vektörleri arasındaki uzaklığı maksimize etmektir. DVM’de karşılaşılabilecek iki durum, verilerin doğrusal olma ve doğrusal olmama durumlarıdır. Doğrusal olmayan bir veri için DVM doğrusal bir hiperdüzlem oluşturamaz. DVM doğrusal olmayan durumlar için Doğrusal Fonksiyon, Polinomal Fonksiyon, Sigmoid Fonksiyon gibi çeşitli çekirdek fonksiyonları kullanır. Şekil 3.8’ de doğrusal olarak ayrılabilen ve doğrusal olarak ayrılamayan veriler için örnek bir gösterim verilmiştir.



Şekil 3.8. (a) Doğrusal Olarak Ayrılabilme Durumu (b) Doğrusal Olarak Ayrılamama Durumu [46]

Karar Ağaçları

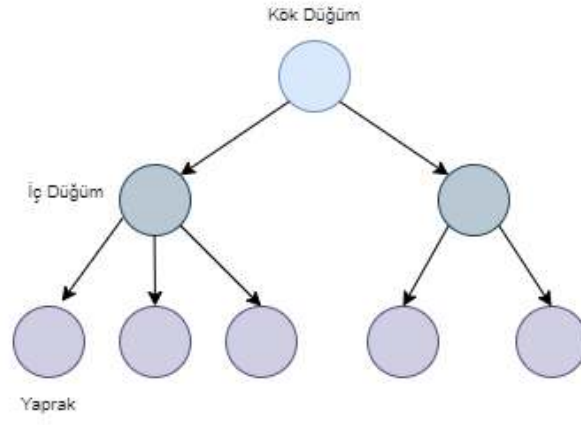
Karar ağaçları yaygın kullanıma sahip olan bir makine öğrenimi algoritmasıdır. Bu yöntemin yaygın kullanım sebebi, ağaç yapısı oluşturulurken kuralların anlaşılabilir olmasıdır [47]. Karar ağacı düğüm, dallar ve yapraklardan oluşur. Ağaç yapısında özellikler düğümler tarafından temsil edilir. Ağaç yapısına ait diğer elemanlar ise dallar, yapraklar ve köktür. En üst kısım kök olarak adlandırılırken en altta kalan son kısım ise yaprak olarak adlandırılır. Kök ve yapraklar arasında kalan kısımlar ise dal olarak adlandırılmaktadır [48]. Düğümler giriş verilerinin test edildiği, soruların sorulduğu ve hangi yöne yönlendirilmesi gerektiğini belirleyen karar düğümleridir. Dallar sorulan soruların cevaplarını temsil ederler. Yapraklar ise kategorinin bulunduğu sınıf etiketleridir.

Karar ağacı algoritmasının adımları;

1. Karar ağacının hangi kararı alacağı belirlenir.
2. Sistemin entropi değeri hesaplanır.
3. Kökte konumlanacak özellik belirlenir. Özellik belirleme işlemi için her özellik için kazanç değeri hesaplanır ve kazancı en yüksek olan değer kök düğüm seçilir.
4. İşlemler;
 - Örneklerin tamamı aynı sınıfta bulunuyorsa
 - Örnekleri ayıracak herhangi bir özellik bulunmuyor,
 - Kalan özelliklerin değerini taşıyan örnek yok,

Durumlarından biri oluşana kadar devam eder.

Şekil 3.9’da örnek bir karar ağacı yapısı verilmiştir.

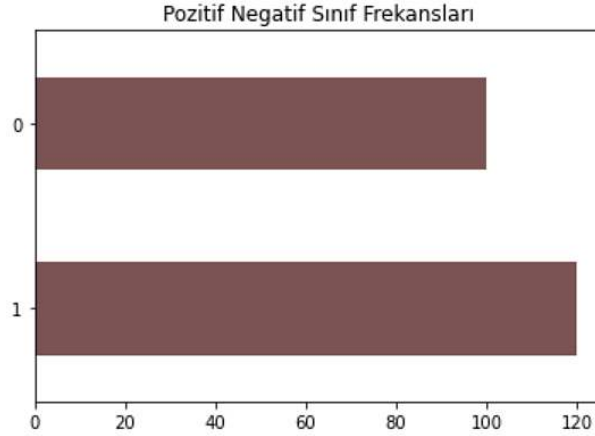


 ekil 3.9. Karar Aėacı Yapısı

4. MATERYAL VE METOT

4.1. Materyal

Tez çalışmasında kullanılan veri seti Elazığ Fethi Sekin Şehir Hastanesine başvuran ve uzman hekimler tarafından tanısı koyulmuş olan hastalardan elde edilmiştir. Veriler 100 Covid-negatif ve 121 Covid-pozitif hasta olmak üzere toplamda 221 hastaya ait biyokimya verisinden oluşmaktadır. Veri seti toplamda Covid-19 hastalığının teşhisi için rutin olarak kullanılan 16 biyokimya parametresini içermektedir. Bu parametreler sırasıyla (C-reaktif protein, Dimer, Üre, Kreatinin, Protein, Albümin, Aspartat Aminotransferaz, Alanin Aminotransferaz, Laktat dehidrogenaz, Bilirubin, Beyaz Kan Hücresi (WBC) , Trombosit, Nötrofil, Lenfosit, Monosit, Eozonofil) şeklindedir. Mevcut biyokimya verilerinde karşılık gelen Covid-pozitif ve Covid-negatif sınıfları veride 0 ve 1 şeklinde belirtilmiştir. 0 sınıfı mevcut veride Covid-negatif, 1 sınıfı veride Covid-pozitif hastaları temsil etmektedir. Mevcut sınıflara ait sınıf frekansları Şekil 4.1’de verilmiştir.



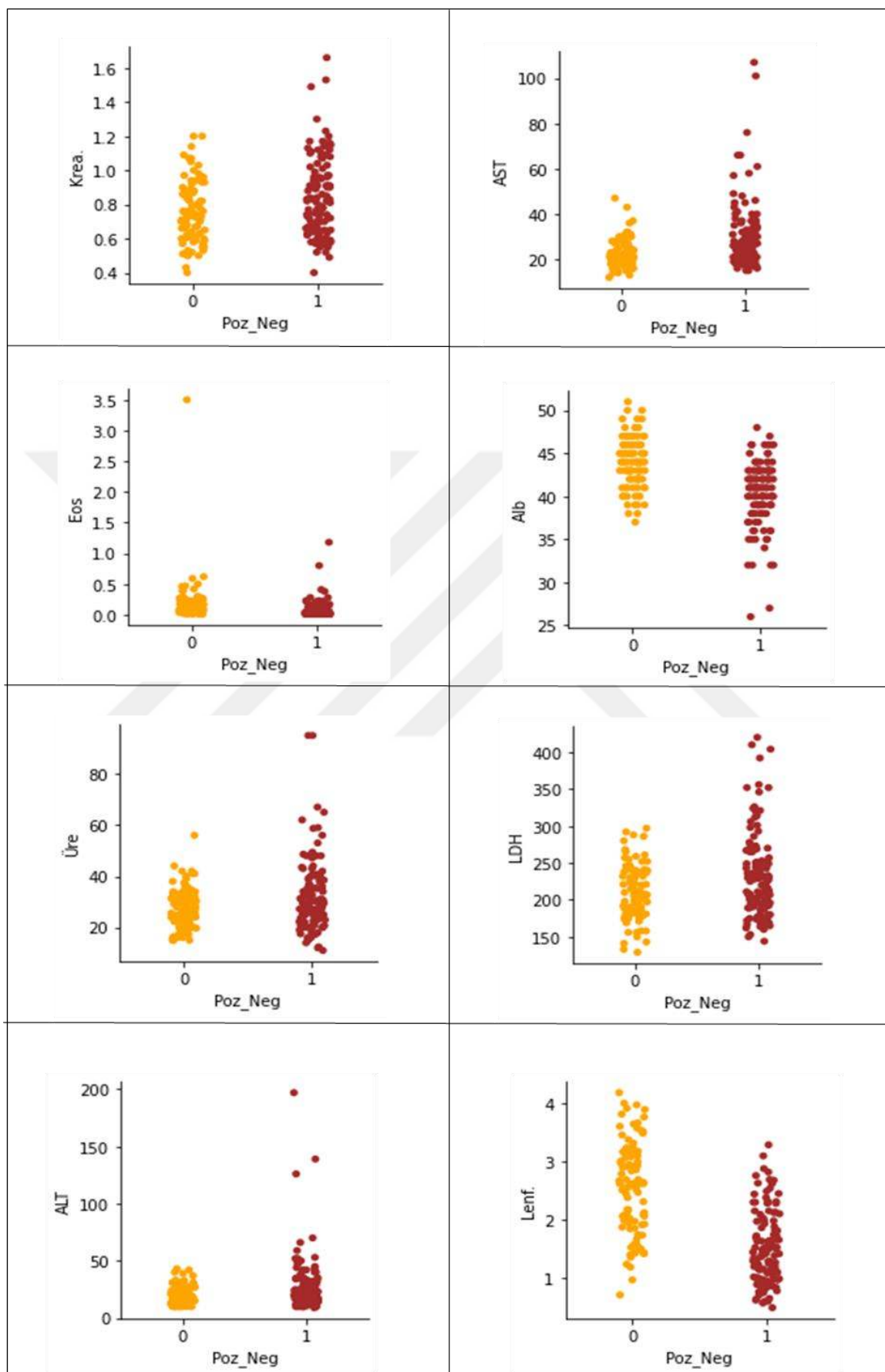
Şekil 4.1. Pozitif ve Negatif Sınıflara Ait Frekanslar

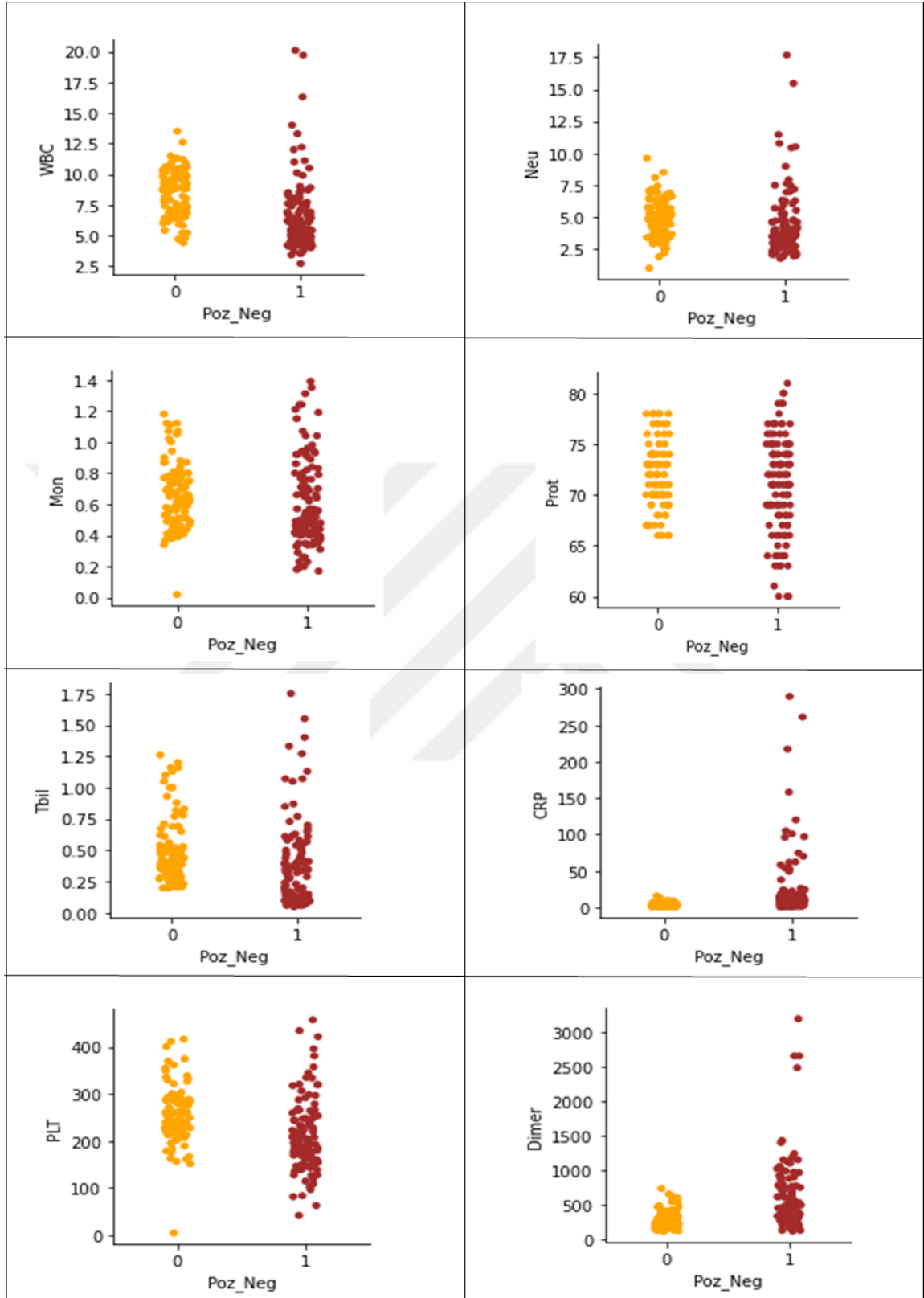
Veri setinde bulunan her özellik Covid-pozitif ve Covid-negatif olma durumlarına göre birbirinden farklı değer aralıklarına ve farklı istatistiklere sahiptir. Veriye ait betimleyici istatistikler Tablo 4.1’de verilmiştir.

Tablo 4.1. Veriye Ait Betimleyici İstatistikler

	Ortalama	S. Sapma	Minimum	25%	50%	75%	Maksimum
CRP	14.490685	35.64223	1.000000	2.015000	3.800000	10.750000	289.00000
Dimer	453.69117	428.0068	111.0000	217.50000	323.00000	498.00000	3200.0000
Üre	30.236.364	11.707.95	11.000.00	23.000.000	28.000.000	34.850.000	95.000.000
Kreatinin	0.794136	0.210576	0.400000	0.630000	0.750000	0.922500	1.660.000
Protein	71.412.844	4.081.454	60.000.00	69.000.000	71.500.000	74.000.000	81.000.000
Albümin	42.009.174	3.840.112	26.000.00	40.000.000	42.000.000	44.000.000	51.000.000
AST	26.600.000	12.329.71	12.000.00	20.000.000	23.000.000	30.000.000	107.000.00
ALT	23.900.000	19.270.46	9.000.000	14.000.000	20.500.000	27.250.000	197.000.00
LDH	221.229.35	50.351.41	129.000.0	188.000.00	211.000.00	242.000.00	420.000.00
Blirubin	0.422603	0.315068	0.050000	0.200000	0.360000	0.525000	1.750.000
WBC	7.394.818	2.658.356	2.700.000	5.400.000	6.900.000	9.000.000	20.100.000
Trombosit	230.703.5	71.750.81	4.780.000	185.000.00	228.000.0	266.000.00	458.000.00
Nötrofil	4.573.864	2.135.464	0.990000	3.065.000	4.115.000	5.545.000	17.680.000
Lenfosit	1.988.636	0.866485	0.480000	1.305.000	1.895.000	2.622.500	4.190.000
Monosit	0.629545	0.245823	0.020000	0.457500	0.590000	0.770000	1.390000
Eozonofil	0.127000	0.268507	0.010000	0.010000	0.070000	0.160000	3.500000

Tablo 4.1’de verilen minimum parametresi özelliğe ait en küçük değeri, 25% parametresi veride özelliklerin %25 'ine ait ortalama değeri, 50% parametresi veride bulunan ortanca değeri yani verinin medyan değerini, 75% parametresi veride özelliklerin %75'ine ait ortalama değeri, maksimum değer ise özelliğe ait en büyük değeri temsil etmektedir. Tablo 4.1’de verilen betimleyici istatistikler incelendiğinde verilerin gerçek hayat verisi olması sebebiyle aykırı değerler içerebileceği buna bağlı olarak bazı verilerin standart sapma değerlerinin yüksek olduğu gözlemlenmektedir. Aykırı veriler sonraki aşamalarda temizlenerek, gerekli veriler üzerinde normalizasyon işlemi uygulanacaktır. Şekil 4.2’de betimleyici istatistik tablosunda verilen her özelliğe ait Covid-pozitif ve Covid-negatif sınıf dağılımlarını gösteren grafikler verilmiştir.





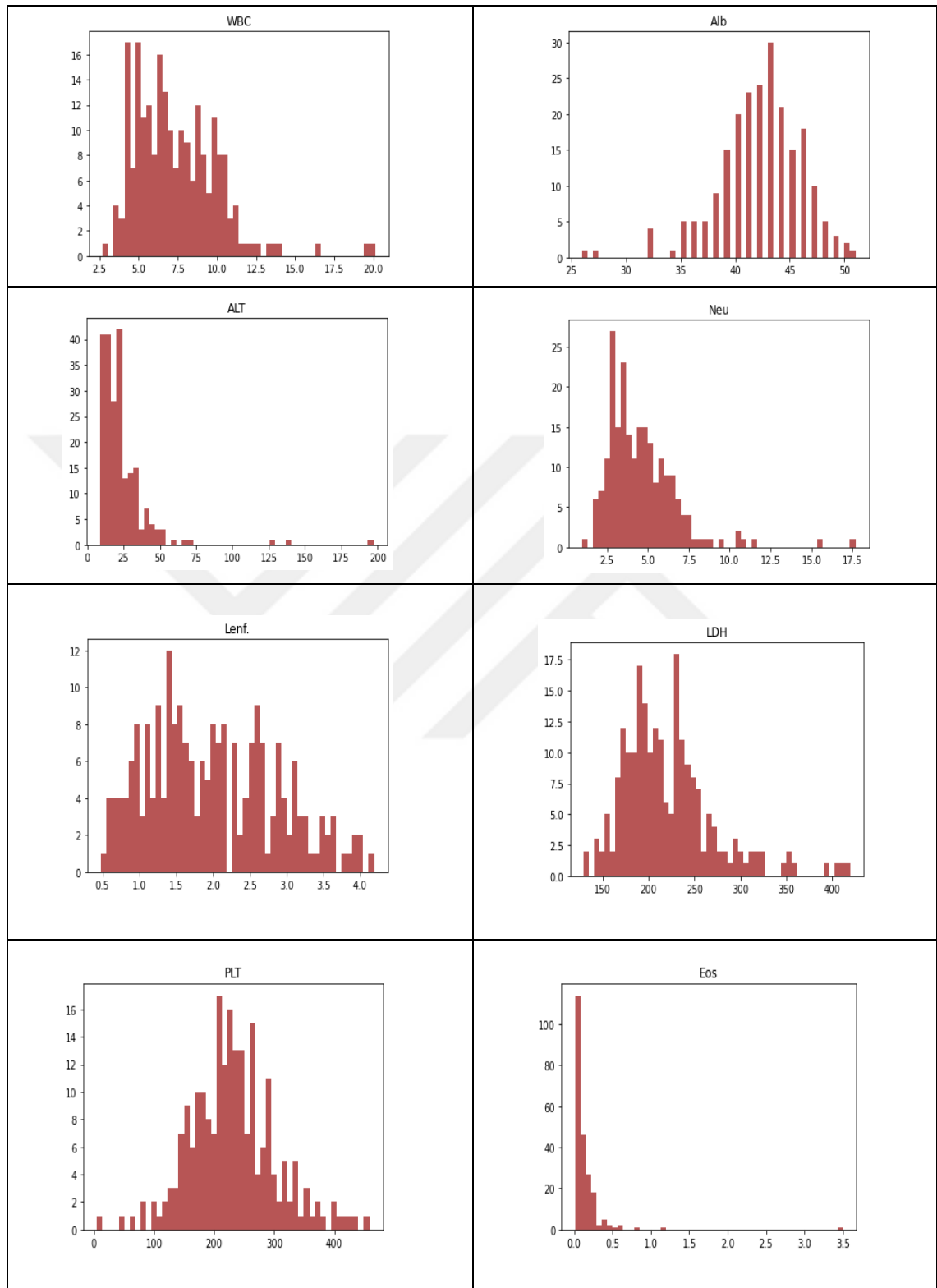
Şekil 4.2. Covid-Pozitif ve Covid-Negatif Hastaların Özelliklere Göre Dağılımları

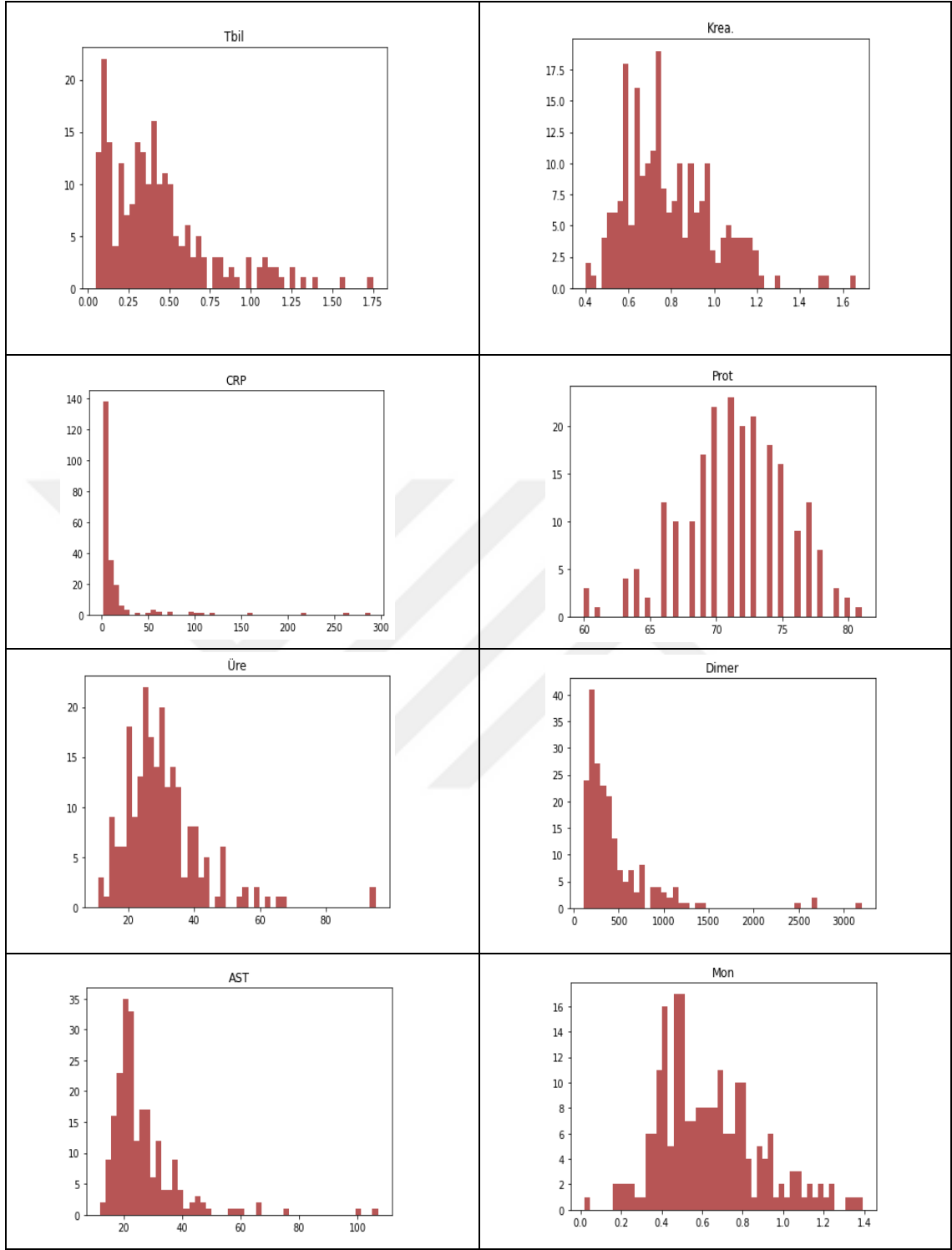
Şekil 4.2’de verilen dağılım grafikleri incelendiğinde her bir özellik aşağıdaki şekilde yorumlanabilir:

- CRP: Covid-negatif hastalarda CRP değer aralığının genel itibariyle 0-25 değerleri arasında yoğunlaştığı görülmektedir. Covid-pozitif hastalar incelendiğinde ise bu değer aralığının 0-300 aralığına genişlediğini ve yoğunluğun 50-300 değer aralığında oluştuğunu gözlemlemekteyiz. CRP değeri için belirleyici aralığın 50-300 aralığı olduğu grafik verilerine göre söylenebilir.
- Dimer: Dimer parametresi dağılımı incelendiğinde Covid-negatif hastalarda değer aralığının 11-800 değerleri arasında olduğu ve yoğunluğun genel olarak 11-500 değer aralığında olduğunu gözlemlemekteyiz. Covid-pozitif hastalar incelendiğinde ise bu değer aralığının 11-3500 değer aralığında genişlediği gözlemlenmektedir. Yoğunluk dağılımının ise 11-1500 aralığında oluştuğu görülmektedir.
- Üre: Üre parametresinin Covid-negatif hastalar için 11-60 aralığında dağılım gösterdiği, yoğunluk olarak 11-40 değer aralığında bir birikim olduğu gözlemlenmektedir. Covid-pozitif hastalar için ise 11-95 aralığında dağılımın olduğu ve yoğunluğun 11-75 değer aralığında oluştuğu görülmektedir.
- Kreatinin: Kreatinin parametresi incelendiğinde Covid-negatif hastalar için yoğunluğun 0.4-1.2 aralığında yoğunlaştığı, Covid-pozitif hastalar için ise bu yoğunluğun 1.8 değerlerine ulaşabildiği görülmektedir.
- Protein: Covid-pozitif hastalarda protein değerleri belirli bir aralıkta düzenli bir dağılım göstermektedir. Grafik incelendiğinde bu aralığın yaklaşık olarak 66 ile 78 değerleri arasında olduğu gözlemlenmektedir. Covid-negatif hastalar için ise bu aralık yaklaşık olarak 60 ile 85 arasında değişiklik göstermektedir.
- Albümin: Albümin parametresi incelendiğinde Covid-negatif hastaların albümin değer aralığının şimdiye kadar incelenen parametrelerin aksine Covid-pozitif hastalara kıyasla daha yüksek değer aralığında yer aldığı görülmektedir.
- AST: Covid-pozitif hastalarda gözlemlenen AST değeri 10-45 aralığında yoğunluk gösterse de 100 ve üstü değerler grafikte Covid-pozitif hastalar için yer almaktadır. Covid-negatif hastalar için bu değer aralığı 11-50 aralığında olup yoğunluğun yaklaşık olarak 11-35 aralığında oluştuğu gözlemlenmektedir.
- ALT: ALT parametresi incelendiğinde Covid-pozitif hastalarda benzer dağılımın dışına çıkarak aralığın yaklaşık 5-200 değerlerine kadar genişlediği gözlemlenmektedir. Covid-negatif hastalarda bu aralığın yaklaşık 5-45 arasında olduğu görülmektedir.

- LDH: Covid-negatif hastalarda LDH değeri sınırının yaklaşık olarak 300 olduğu görülmektedir. Covid-pozitif hastalar incelendiğinde ise bu değeri sınırının 400 ve üstü bir aralığa kadar genişlediğini ve yoğunluğun yaklaşık 170-350 değeri aralığında olduğunu gözlemlemekteyiz.
- Bilirubin: Bilirubin parametresi incelendiğinde Covid-negatif hastaların bilirubin değeri aralığının Covid-pozitif hastalara kıyasla daha yüksek değeri aralığında yoğunlaştığı görülmektedir.
- WBC: WBC parametresinin Covid-negatif hastalar için 4-13 aralığında dağılım gösterdiği, Covid-pozitif hastalar için ise yaklaşık olarak 2.5-21 aralığında bir dağılım olduğu görülmektedir.
- Trombosit: Covid-pozitif hastaların trombosit değeri aralığının yaklaşık 50-450 değeri aralığında değişkenlik gösterdiği ve 150-280 değeri arasında yoğunlaştığı görülmektedir. Covid-negatif hastalarda ise bu yoğunluğu 200-300 değeri aralığında olduğu gözlemlenmektedir.
- Nötrofil: Grafik incelendiğinde Covid-pozitif hastalarda değeri aralığının yaklaşık olarak 2 ile 17.5 değeri arasında olduğu gözlemlenmektedir. Covid-negatif hastalar için ise bu aralık yaklaşık olarak 1 ile 10 arasında değişiklik göstermektedir.
- Lenfosit: Covid-negatif hastalar lenfosit parametresinde yaklaşık olarak 0.5 ile 4.5 aralığında yayılım göstermektedir. Covid-pozitif hastalar ise yaklaşık olarak 0.1 ile 3.5 aralığında yayılım göstermektedir. Grafığe göre Covid-pozitif hastalarda lenfosit değeri aralığının azalan eğilimde olduğunu söyleyebiliriz.
- Monosit: Monosit parametresi incelendiğinde, Covid-pozitif hastalarda daha yüksek değeri aralığında vaka gözlemlenmesine rağmen Covid-negatif hastaların daha yüksek değeri aralığında yoğunlaştığı görülmektedir.
- Eozonofil: Grafik incelendiğinde Covid-pozitif ve Covid-negatif hastalarda eozonofil değeri aralığının yoğunluk aralıklarının benzerlik gösterdiği görülmektedir. Bu yüzden eozonofil parametresi için artan ya da azalan bir eğilimden söz edilemez.

Verilerin sınıflara göre dağılımları ve incelemelere ait bulgular yukarıda verilmiştir. Verilerin kendi içerisinde dağılımlarını gösteren histogram grafikleri Şekil 4.3’de verilmiştir. Histogramlar incelenerek verilerin ihtiyaç duyduğu ön işleme adımlarına betimleyici istatistiklerde göz önünde bulundurularak karar verilecektir.





Şekil 4.3. Veri Özellik Dağılımları Histogram Grafikleri

Şekil 4.3’de verilen dağılım grafikleri incelendiğinde normal dağılıma yakın ve normal dağılımdan uzak parametrelerin olduğu gözlemlenmektedir. Sırasıyla dağılımlarla ilgili elde edilen gözlemler şu şekildedir:

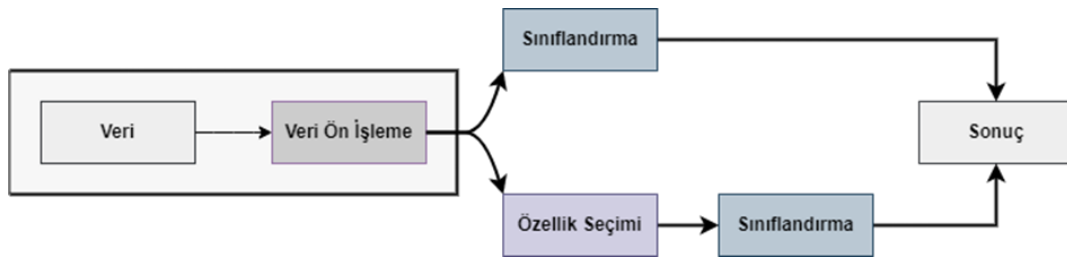
- CRP: CRP parametresine ait dağılım grafiği incelendiğinde eksponansiyel bir dağılım olduğu gözlemlenmektedir. Bu durum CRP parametresinin düzensiz dağılım davranışı gösterdiğini ifade etmektedir. Gözlemlenen dağılıma göre CRP parametresi için uygun normalizasyon işlemi olan box-cox dönüşümü uygulanmıştır. Metot başlığında CRP parametresinde gözlemlenen eksponansiyel dağılımı normal dağılıma yaklaştırmak amacıyla gerçekleştirilen normalizasyon işlemi ayrıntılı bir şekilde ele alınacaktır.
- Dimer: Dimer parametresine ait dağılım grafiği incelendiğinde CRP parametresine benzer eksponansiyel bir dağılımın olduğu gözlemlenmektedir. Dimer parametresini normal dağılıma yakın hale getirmek için kullanılacak olan uygun yöntem box-cox dönüşümüdür.
- Üre: Üre parametresinin dağılımı incelendiğinde CRP ve Dimer parametrelerinden farklı bir dağılım olduğu görülmektedir. Mevcut dağılım grafiğine göre üre parametresi için kesin bir yorum yapmak doğru olmayacaktır. Bu yüzden üre parametresine ait dağılımın, grafikte beraber tanımlayıcı istatistiklerde incelenerek yorumlanması doğru olacaktır. Grafik ve tanımlayıcı istatistikler beraber incelendiğinde normale yakın bir dağılımın varlığından söz edilebilir. Fakat elimizde bulunan veri gerçek hayat verisi olduğu için tamamen normal bir dağılımdan söz edilemez. Üre parametresi normale yakın dağılımlı davranış göstermektedir.
- Kreatinin: Kreatinin parametresinin dağılımı incelendiğinde Üre parametresi ile benzer bir dağılımın olduğu görülmektedir. Bu sebeple dağılım grafiği tanımlayıcı istatistiklerle beraber incelenmiştir. Tanımlayıcı istatistikler incelendiğinde Kreatinin parametresine ait minimum-maksimum değerleri arasında anlam ifade edecek bir farklılık olmadığı, standart sapma değerinin minimum düzeyde olduğu gözlemlenmektedir. Kreatinin parametresi için bir dönüşüm işlemi uygulanmayacaktır. Kreatinin parametresi normalde yakın dağılımlı davranış göstermektedir.
- Protein: Protein parametresine ait dağılım grafiği incelendiğinde normale yakın bir dağılımın olduğu görülmektedir. Mevcut dağılım ve tanımlayıcı istatistikler göz önüne alındığında bir normalizasyon işlemine ihtiyaç duyulmamaktadır.
- Albümin: Albümin parametresi incelendiğinde protein parametresine benzer dağılım gösterdiği görülmektedir. Protein parametresinde olduğu gibi albümin parametresinde de bir normalizasyon işlemine ihtiyaç duyulmamaktadır.
- AST: AST parametresinin dağılım grafiği incelendiğinde normalden uzak eksponansiyel bir dağılımın olduğu görülmektedir. AST parametresini normal dağılıma yakın hale getirmek için kullanılacak olan uygun yöntem box-cox dönüşümüdür.

- ALT: ALT parametresi incelendiğinde eksponansiyel bir dağılımın olduğu ve minimum-maksimum değerler arasında görülebilir bir uzaklık olduğu gözlemlenmektedir. Tanımlayıcı istatistiklerde göz önüne alındığında minimum ve maksimum değerler arasında dağılım grafiğinde gözlemlenen uzaklığın tanımlayıcı istatistiklerde de mevcut olduğu görülmektedir. Bu durumda ALT parametresini normale yaklaştırmak için uygulanacak olan uygun yöntem min-max dönüşümüdür.
- LDH: LDH parametresi için tanımlayıcı istatistikler ile birlikte inceleme yapıldığında standart sapmanın elimizde bulunan veriye göre yüksek olduğu gözlemlenmektedir. Dağılım grafiği incelendiğinde minimum ve maksimum değerler arasında belli bir uzaklık söz konusu olduğu görülmektedir. Bu durumda LDH parametresi için uygulanacak olan uygun yöntem min-max dönüşümüdür.
- Blirubin: Blirubin parametresine ait dağılım grafiği ve tanımlayıcı istatistikler birlikte ele alındığında eksponansiyel bir dağılımın söz konusu olduğu gözlemlenmektedir. Bu durumda Blirubin parametresi için uygulanacak olan uygun yöntem box-cox dönüşümüdür.
- WBC: WBC parametresinin dağılımı incelendiğinde normale yakın bir dağılımın mevcut olduğu görülmektedir. WBC parametresi için bir normalizasyon işlemine ihtiyaç duyulmamaktadır.
- Trombosit: Trombosit parametresi tanımlayıcı istatistiklerle beraber incelendiğinde minimum ve maksimum değer arasında uzaklık olduğu, buna bağlı olarak standart sapma değerinin yüksek olduğu görülmektedir. Bu durumda trombosit parametresini normale yakın bir dağılıma getirmek için min-max dönüşümü uygulanmalıdır.
- Nötrofil: Nötrofil parametresinin dağılımı incelendiğinde normale yakın bir dağılımın mevcut olduğu görülmektedir. Nötrofil parametresi için bir normalizasyon işlemine ihtiyaç duyulmamaktadır.
- Lenfosit: Lenfosit parametresinin dağılımı incelendiğinde normale yakın bir dağılımın mevcut olduğu görülmektedir. Lenfosit parametresi için bir normalizasyon işlemine ihtiyaç duyulmamaktadır.
- Monosit: Monosit parametresinin dağılımı incelendiğinde normale yakın bir dağılımın mevcut olduğu görülmektedir. Monosit parametresi için bir normalizasyon işlemine ihtiyaç duyulmamaktadır.
- Eozonofil: Eozonofil parametresine ait dağılım grafiği ve tanımlayıcı istatistikler birlikte ele alındığında eksponansiyel bir dağılımın söz konusu olduğu gözlemlenmektedir. Bu durumda Eozonofil parametresi için uygulanacak olan uygun yöntem box-cox dönüşümüdür.

Verilerin kendi içerisinde dağılımları ve incelemelere ait bulgular yukarıda verilmiştir. Histogramlar incelenerek verilerin ihtiyaç duyduğu ön işleme adımlarına betimleyici istatistiklerde göz önünde bulundurularak karar verilmiştir. Metot alt başlığı altında belirlenen ön işlem adımlarına ayrıntılı yer verilecek ve uygulanarak verilere ait dağılımlar yeniden incelenecektir.

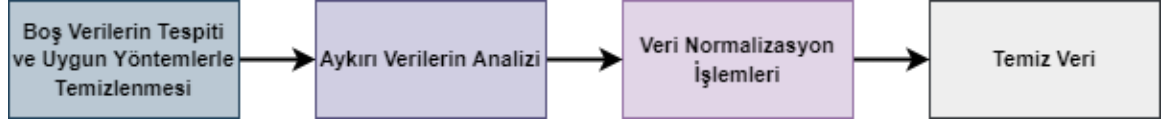
4.2. Metot

Tez çalışması temelde iki ana yaklaşıma dayanarak gerçekleştirilmiştir. Bu yaklaşımlardan ilki elimizde bulunan biyokimya parametrelerinin tamamını kullanarak makine öğrenmesi yöntemleriyle Covid-19 hastalığının tanısını sağlamaktır. Çalışmaya ait diğer yaklaşım ise özellik seçim yöntemlerini makine öğrenimine dâhil ederek, daha az özellik ile daha anlamlı sonuçlar elde edip Covid-19 hastalığının teşhisini gerçekleştirmektir. Bu durum sağlık sektöründe zamandan ve maliyetten tasarruf sağlayacaktır. Bu tez çalışması daha az parametre ile daha doğru sonuç yaklaşımı benimsenerek gerçekleştirilmiştir. Birinci yaklaşımda toplamda 221 hastaya ait 16 parametreden oluşan veri seti üzerinde gerekli veri ön işleme adımları uygulandıktan sonra veriler sınıflandırma işlemine tabi tutulmuştur. İkinci yaklaşımda ise bu 16 parametreden elde edilen sonuçları daha az parametre ile elde edebilmek amacıyla 13 farklı özellik seçim yöntemi kullanılıp, her yöntem için 5 özellik seçilerek özellikler sınıflandırmaya dâhil edilmiş ve seçilen 5 özelliğin sınıflandırma performansı üzerindeki etkileri karşılaştırmalı olarak incelenmiştir. Çalışmaya özellik sayısını maksimum tutarak başlanmış yüksek performans sonuçlarını minimum özellik sayısı ile elde edene kadar deneysel işlemler gerçekleştirilmiştir. Gerçekleştirilen işlemler sonucunda minimum parametre sayısı 5 olarak belirlenmiştir. Gerçekleştirilen çalışmaya ait süreç tasarımı Şekil 4.4’de verilmiştir.



Şekil 4.4. Süreç Tasarımı

Veriye ait incelemeler ve açıklamalar 4.1 Materyal başlığı altında ayrıntılı olarak verilmiştir. Veriye ait ön işleme adımları Şekil 4.5’de verilmiştir.



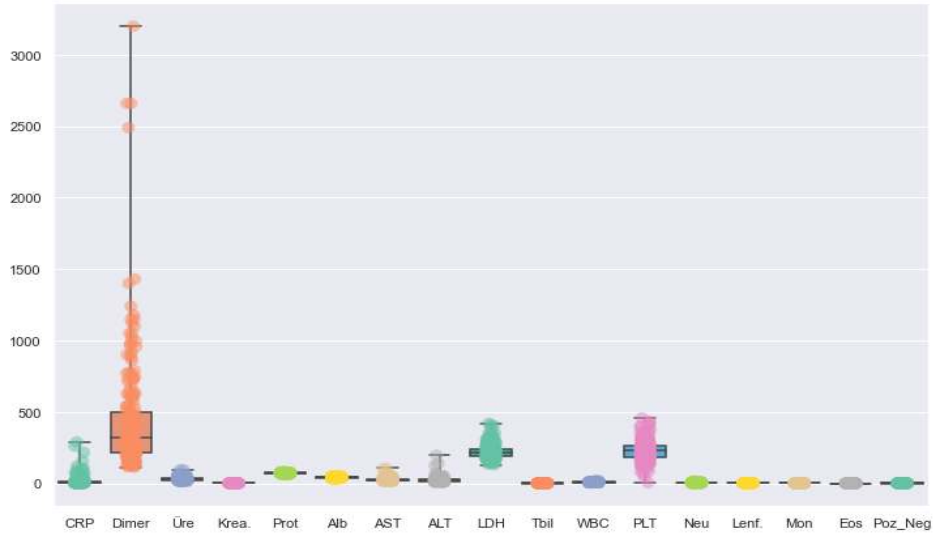
Şekil 4.5. Veri Ön İşleme Adımları

4.2.1. Boş Verilerin Tespiti ve Uygun Yöntemlerle Temizlenmesi

Veri ön işleme aşamasının ilk adımında sınıflandırma performansını etkileyeceği için veride bulunan boş verilerin tespiti ve boş veri sorununun giderilmesi gerçekleştirilmiştir. Veride bulunan parametreler incelendiğinde sırasıyla CRP, Dimer, Protein, Albümin, LDH ve Bilirubin parametrelerinde 1, 16, 2, 2, 2, 1 boş veri olduğu gözlemlenmiştir. CRP, Protein, Albümin, LDH ve Bilirubin parametreleri az sayıda boş veri içerdiği için satır bazlı olarak boş veriler silinerek boş veri sorunu giderilmiştir. Dimer parametresi çok sayıda boş veri içerdiği için öncelikle veri dağılımı incelenmiştir. İnceleme sonucunda veri aralığının birbirine yakın olduğu gözlemlenmiştir. Genellikle küçük veri setlerinde ve veri aralığının birbirinden çok ayrık olmadığı durumlarda kullanılan ortalama ile doldurma yöntemi Dimer parametresi için kullanılmıştır.

4.2.2. Aykırı Verilerin Analizi

Aykırı değerler bir veri kümesinde bulunan verilerin çoğundan büyük ölçüde sapma gösteren değerlerdir. Aykırı değerlerin tespiti tıbbi tanı ve teşhis, hatalı tespit gibi alanlarda sorun yaratmakta ve bu sorunun çözümü önem arz etmektedir. Bu çalışmada veriye ait aykırı değerlerin tespiti için öncelikle boxplot grafiği incelenerek aykırı değerlerin varlığı gözlemlenmiştir. Boxplot grafiği ile gözlemlenen aykırı değerlerin varlığının ispatı, var olan aykırı değerlerin sayısı ve temizlenmesi için Tek Sınıflı Destek Vektör Makinesi kullanılmıştır. Gözlemlenen aykırı değerlere ait boxplot Şekil 4.6'da verilmiştir.

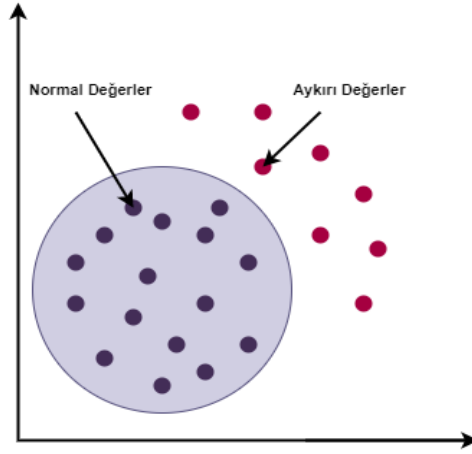


Şekil 4.6. Aykırı Değer Analizi

Kullanılan Tek Sınıflı Destek Vektör Makinesi yöntemi ile $\nu=0.1$ değeri için veride toplamda 21 aykırı değer olduğu belirlenmiştir. Aykırı değerleri temizlemek özellikle küçük boyutlu verilerde veri kaybına yol açarak sınıflandırma performansına etki edebilir. Bu sorunun elimizde bulunan veri için mevcut olup olmadığını belirlemek için elimizdeki veriye aykırı değerler temizlenerek ve temizlemeden iki farklı sınıflandırma işlemi uygulanmıştır. Aykırı değerler temizlenerek gerçekleştirilen sınıflandırma işleminin daha yüksek performans ile sonuç verdiği gözlemlenmiştir. Mevcut veriler için tez çalışmasına aykırı değerler temizlenerek devam edilmiştir.

Tek Sınıflı Destek Vektör Makinesi

Tek sınıflı destek vektör makinesi yöntemi temelde tüm veri noktalarının yakın olduğu varsayımı fikri üzerine inşa edilmiştir. Tek sınıflı destek vektör makinesi algoritmasında algoritma neyin normal olduğunu öğrenmek için eğitilir. Model, eğitilen veriler etrafında bir hiperdüzlem oluşturarak belirlenen bu hiperdüzlem dışında kalan veri noktalarını aykırı değer olarak işaretler. Yeni bir veri algoritmaya giriş olarak verildiğinde algoritma yeni verinin normal gruba ait olup olmadığını tespit eder. Normal gruba ait olmayan girişler aykırı değer olarak kabul edilir [49]. Şekil 4.7’de tek sınıflı destek vektör makinesine ait örnek bir görsel verilmiştir.



Şekil 4.7. Tek Sınıflı Destek Vektör Makinesi Örnek Görsel

4.2.3. Veri Normalizasyon İşlemleri

Gerçek hayat verileri değişkenlik gösteren yapıları sebebiyle genelde normal dağılımdan uzak bir dağılım davranışı gösterirler. Bu çalışmada kullanılan veriler gerçek hayat verileri olduğu için bazı özellikler normalden uzak dağılım göstermektedir. Normal dağılım göstermeyen veriler mevcut olduğunda, verilerin normal dağılım göstermeme nedenleri belirlenerek normale yaklaştırmak için uygun yöntem arayışına girilmelidir. Bir değişkenin normal dağılım göstermemesi durumunda çeşitli veri dönüştürme yöntemleri uygulanmaktadır. Kullanılan bu veri dönüşüm yöntemleri ile veri dağılımlarının normale yaklaşımı sağlanmaktadır. Şekil 10’da verilen özellik dağılımları incelenerek normal dağılımdan uzak özellikleri normal dağılıma yaklaştırmak için gerekli olan normalizasyon yöntemleri belirtilmiştir. Mevcut dağılımlara göre CRP, Dimer, AST, Bilirubin ve Eozonofil parametreleri için gösterdikleri exponansiyel dağılım sebebiyle box-cox dönüşümü uygulanmıştır. ALT, LDH ve Trombosit parametreleri için ise min-max dönüşümü gerçekleştirilmiştir. Diğer özellikler normale yakın dağılım gösterdikleri için veriler üzerinde bir normalizasyon işlemi gerçekleştirilmemiştir.

Box-Cox Dönüşümü

Box-cox dönüşümü, normal dağılmayan çarpık verileri normal dağılıma yaklaştırmak için kullanılan istatistiksel bir yöntemdir. Box-cox yönteminde normal olmayan dağılımı normal dağılıma benzetmek için lambda (λ) parametresini kullanılır. Mevcut yöntem, en iyi değeri bulana kadar $-5 < \lambda < 5$ aralığında arama gerçekleştirir. Lambda parametresinin aldığı değere göre farklı box-cox dönüşümleri uygulanmaktadır [50]. Elimizde bir Y verisi olduğunu varsayarsak lambda değerlerine karşı Y' e uygulanacak dönüşümler ve formülleri Tablo 4.2’de verilmiştir.

Tablo 4.2. Box-cox Dönüşüm Tablosu

Lamda Değeri	Dönüşüm Yöntemi	Formül
-1	Karşıt Dönüşüm	$Y^{-1} = 1/Y^1$
-0.5	Karşıt Karekök Dönüşüm	$Y^{-0.5} = 1/(\text{Sqr}(Y))$
0.0	Logaritmik Dönüşüm	$\log(Y)$
0.5	Karekök Dönüşüm	$Y^{0.5} = \text{Sqrt}(Y)$
1.0	Dönüşüm Yok	$Y^1 = Y$

Min-Max Dönüşümü

Min-max dönüşümü kullanılarak [0-1] aralığında normal dağılım sergilemeyen veriler üzerinde doğrusal bir dönüşüm işlemi gerçekleştirilir. Minimum; bir verinin alabileceği en küçük değeri temsil ederken maksimum ise bir verinin alabileceği en büyük değeri temsil eder. Min-max dönüşümünde amaç minimum değeri 0, maksimum değeri 1 yaparak diğer verileri [0-1] aralığına yayıp verileri ortak bir sayı sistemi içerisinde ölçeklendirmektir [51]. 4.1 numaralı denklemde min-max dönüşümüne ait matematiksel gösterim verilmiştir.

$$x' = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \quad (4.1)$$

Verilen 1 numaralı denklemde;

X' = Normalize edilmiş olan veriyi,

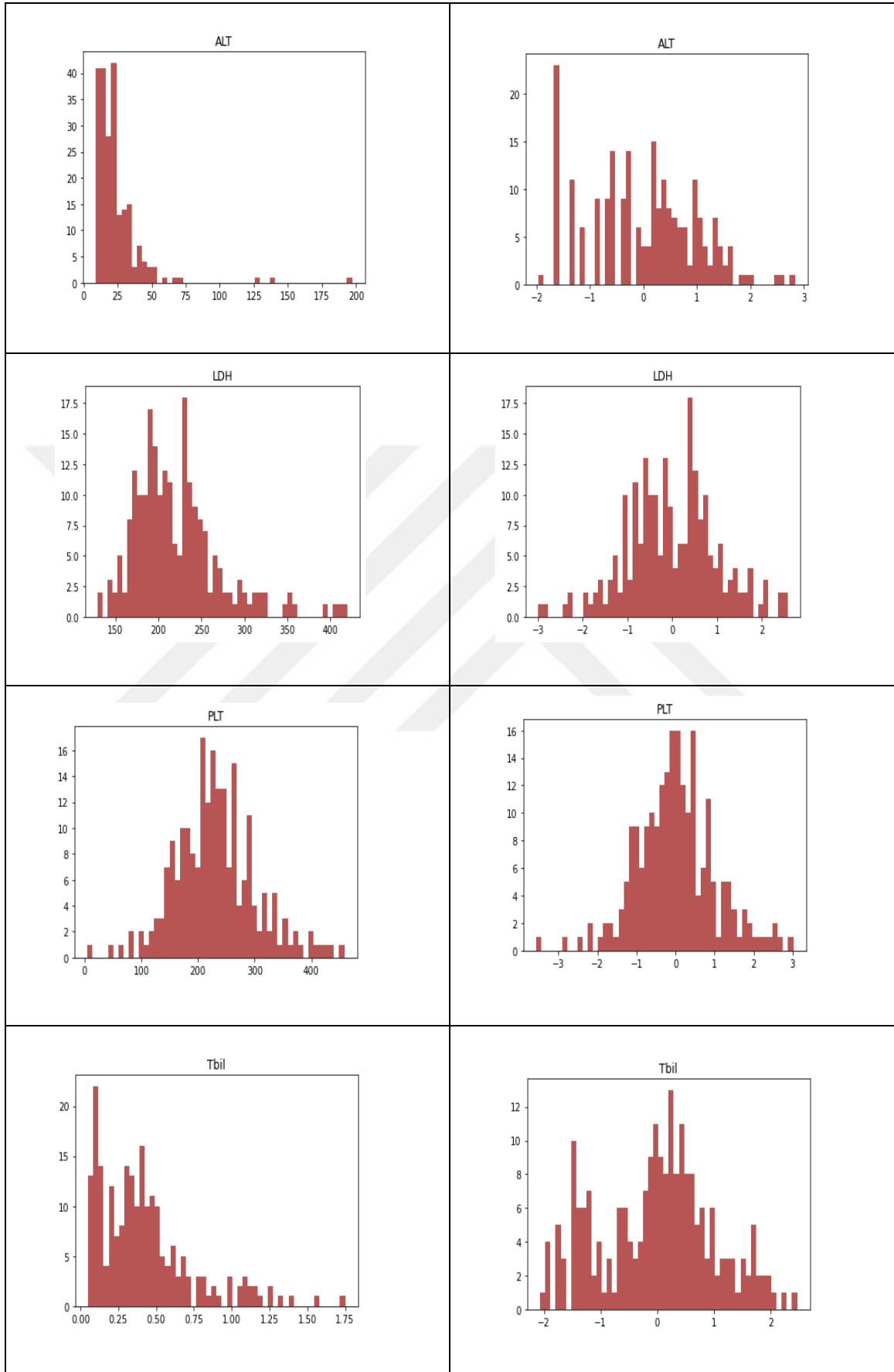
x_i = Girdi değerini,

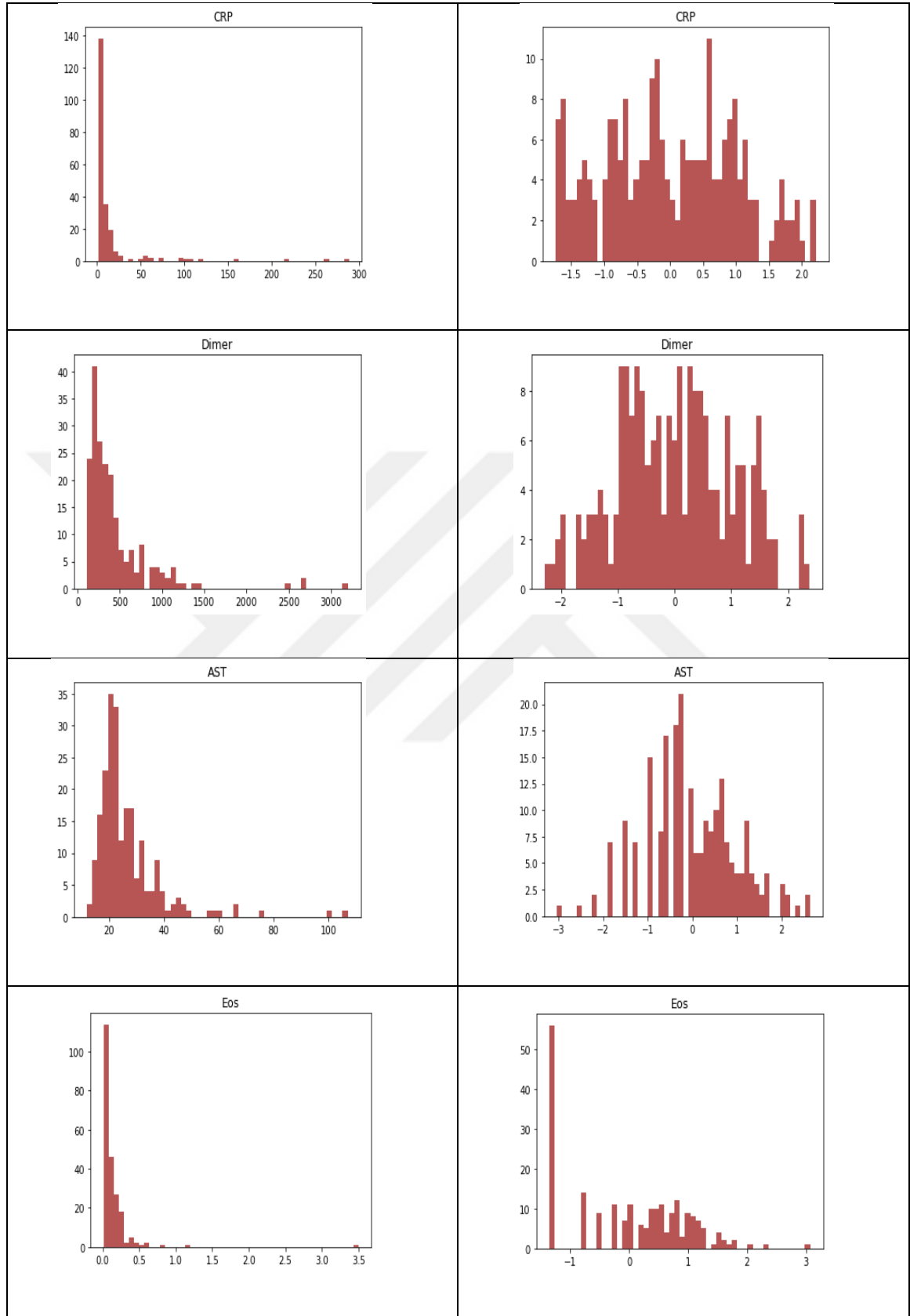
$x_{\min}$ = Girdi verileri içerisinde yer alan en küçük sayıyı,

$x_{\max}$ = Girdi verileri içerisinde yer alan en büyük sayıyı,

temsil etmektedir.

Şekil 4.8’de normalizasyon işlemi uygulanmadan önce ve uygulandıktan sonra verimizde bulunan parametrelere ait dağılım grafikleri karşılaştırmalı olarak verilmiştir.





Şekil 4.8. Normalizasyon İşlemi Uygulanan Veri Dağılımları

4.2.4. Özellik Seçimi

Gerçek dünyayı temsil edebilmek için genelde çok özellikli veriler ile karşılaşırız. Fakat bu özelliklerden sadece birkaçı gerçek dünya temsiline yeterli olabilmektedir. Hangi özelliklerin bu yeterliliği sağladığını öğrenebilmek için özellik seçim işlemine başvurulmaktadır. Özellik seçimi, çok özellikli veri setini betimleyecek en iyi özellik alt kümesinin seçilmesi olarak tanımlanabilir. Özellik seçim işlemi ile problemin çözümüne en uygun alt küme araştırılarak bir özellik alt kümesi oluşturulması amaçlanır [52]. Özellik seçim işlemine ait temsili gösterim Şekil 4.9’da verilmiştir.

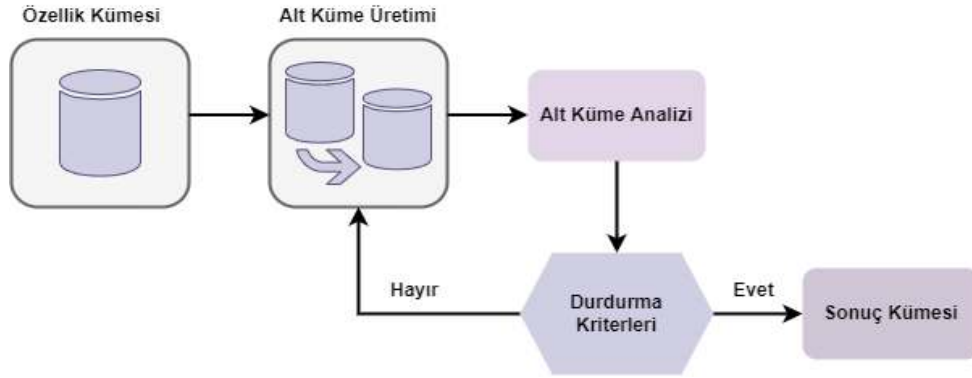


Şekil 4.9. Özellik Seçim İşlemi

Özellik seçim işlemi birçok avantaj sağlamaktadır. Bu avantajlardan bazıları şu şekildedir [52]:

- Özellik seçim işlemi ile oluşturulan özellik alt kümeleri ile boyut azalmasına bağlı olarak model hızı artmaktadır.
- Veri setinde etkisi olmayan sadece boyut fazlalığı yaratan gürültülü veri ortadan kalkar.
- Veri için ayrılan hafıza alanı azalır bu sayede daha çok veri depolama alanında saklanabilir.
- Elde edilen alt küme en iyi özellikleri barındırdığı için model başarımı bu duruma paralel olarak artış gösterir.
- Veri kümesindeki karmaşıklık sorunu giderilir. Bu sayede verinin daha anlaşılabilir olması sağlanır.
- Veri kaynağı sağlama aşamasında maliyetten tasarruf sağlanır.

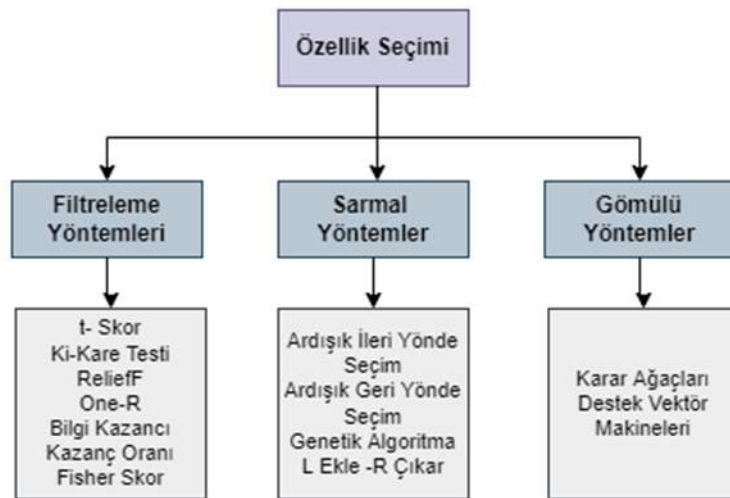
Özellik seçim işlemi metin madenciliği, makine öğrenmesi vb. birçok farklı çalışma alanında kullanılmaktadır. Şekil 4.10’da birçok farklı alanda aynı mantık ile kullanımı sağlanan özellik seçim işlemine ait süreç adımları verilmiştir.



Şekil 4.10. Özellik Seçim Süreç Adımları

Şekil 4.10’da verilen adımlar incelendiğinde ilk adımda özellik seçim işlemine tabi tutulacak özellik kümesi belirlenerek sürece giriş olarak verilir. Ardından verilen özellik kümesinden Komple Arama, Ardışık Arama ya da Rastgele Arama yöntemlerinden biri seçilip uygulanarak bir alt küme üretimi gerçekleştirilir. Bağımlı ya da bağımsız değerlendirme ölçütleri ile analiz edilen alt kümelerden en iyi özellik alt kümesi seçilerek sürece devam edilir. Arama işleminin tamamlanması, minimum özellik sayısı vb. kriterlere ulaşılması, model performansını en iyiye ulaştıran alt kümeye ulaşılması gibi durdurma kriterlerinin sağlanması durumunda en iyi özellik kümesi belirlenmiş olarak sonuç elde edilir [52].

Özellik seçim işlemi bünyesinde birçok farklı yöntem barındırmaktadır. Şekil 4.11’de özellik seçim yöntemlerini içeren şema verilmiştir.



Şekil 4.11. Özellik Seçim Yöntemleri

Filtreleme Yöntemleri

Özellikler ile hedef değişken arasındaki ilişkiyi esas alarak özellik seçimi gerçekleştiren yöntemlerdir. Filtreleme yöntemleri istatistiğe dayalı kriterler aracılığıyla herhangi bir algoritmaya ihtiyaç duymadan mesafe, bilgi vb. kesin değerlendirme ölçütleri kullanarak özellik seçim işlemi gerçekleştirirler [53]. Filtreleme yönteminde her özellik için değerlendirme fonksiyonu kullanılarak bir değer elde edilir. Elde edilen değerler büyükten küçüğe sıralanır ve önem dereceleri belirlenmiş olur. En yüksek değere sahip olan en değerli özelliklerden bir özellik alt kümesi oluşturulur. Filtreleme yöntemlerinin kullanılma sebeplerinden ve avantajlarından bazıları şu şekilde sıralanabilir:

- Temel kullanım sebebi basitliktir. Bu yapısı ile alakalı özellik alt kümesi üretimi gerçekleştirir.
- Hesaplama maliyetleri azdır.
- Aşırı öğrenme sorununu ortadan kaldırır.
- Diğer özellik seçim yöntemlerine kıyasla daha hızlı çalışır.
- Özellik seçimini bir kez gerçekleştirir ve farklı sınıflandırıcılarda kullanılır.

Tüm avantajları göz önüne alındığında filtreleme yöntemleri ön plana çıksa da sınıflandırma performansının diğer yöntemlere kıyasla daha düşük olması sebebiyle temelde bir dezavantaj barındırır [53].

Sarmal Yöntemler

Sarmal yöntemler algoritmalara ihtiyaç duyarak özellik seçim işlemi gerçekleştiren yöntemlerdir. Bu yöntemde algoritmaya uygun özellik araştırılarak, model performansının seçilen özellikler ile iyileştirilmesi amaçlanır. Sarmal yöntemler kullanılarak oluşturulan her özellik alt kümesi için makine öğrenimi algoritmaları kullanılarak performans değerlendirmesi gerçekleştirilir. En iyi model performansını veren özellikler seçilerek nihai özellik alt kümesi belirlenir. Sarmal yöntemler alt küme seçiminde model ile etkileşimli olması sebebiyle daha çok tercih edilen ve daha yüksek performans sağlayan özellik seçim yöntemleridir. Bu yöntemlerin filtreleme yöntemlerine göre dezavantajı ise hesaplama açısından maliyetli olmasıdır [53].

Gömülü Yöntemler

Gömülü yöntemler, hem sınıflandırma algoritmalarını hem de özellik seçim algoritmalarını bünyelerinde barındırırlar. Gömülü yöntemlerdeki temel yaklaşım özellik seçim işlemi eğitim sürecine dâhil etmektir. Bu yaklaşım sebebiyle gömülü yöntemler sınıflandırma ve özellik seçim işlemlerini eş zamanlı olarak birlikte gerçekleştirirler. Gömülü yöntemler performansı açısından tercih edilen özellik seçim yöntemlerinden biri olsa da filtreleme yöntemlerine göre sarmal yöntemlerde olduğu gibi hesaplama açısından daha maliyetlidir [54].

Çalışma Kapsamında Kullanılan Özellik Seçim Yöntemleri

Tez çalışması kapsamında filtreleme, sarmal ve gömülü yöntemlere ait toplamda 13 farklı özellik seçim yöntemi kullanılmıştır. Özellik seçim yöntemleri, özellik sayısını optimum düzeye getirerek minimum özellik sayısı ile yüksek performans elde etmek için kullanılmıştır. Özellik seçim yöntemleri yaygın kullanıma sahip olan kısaltmaları ile birlikte verilmiştir.

Korelasyon Temelli Özellik Seçim Yöntemi (CFS)

Korelasyon tabanlı özellik seçim yöntemi, her özelliğin kendine özgü tahmin performansını dikkate alıp korelasyon temelli bir değerlendirme gerçekleştirerek özellik alt kümelerini bu yapıya uygun olarak sıralayan bir filtreleme yöntemidir [55,56]. KTÖS özellik alt kümeleri ile model arasındaki korelasyonun yanı sıra özelliklerin birbirleri ile olan korelasyon ilişkilerini de göz önünde bulundurur. Birbiri ile ilgisiz özellikler düşük korelasyonlu, birbiri ile alaka düzeyi yüksek olan özellikler ise yüksek korelasyonlu özellikler olarak belirlenir ve özellik kümesinden elenir. Korelasyon temelli özellik seçim yönteminde temel amaç, en iyi özellik alt kümesini, birbiri ile ilgisiz ancak sınıf ile alaka düzeyi yüksek olan özellikleri seçerek oluşturmaktır. Korelasyon temelli özellik seçim yöntemine ait matematiksel gösterim 4.2 numaralı denklemde verilmiştir.

$$S_A = \frac{nr_{kf}}{\sqrt{n+n(n-1)r_{ff}}} \quad (4.2)$$

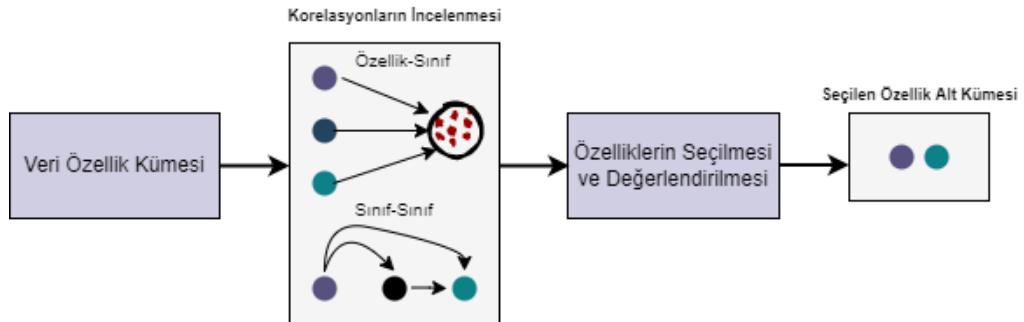
S_A = n adet özellik içeren S altkümesi

n = Özellik sayısı

nr_{kf} = Sınıf ile özellikler arasındaki korelasyon

r_{ff} = Özelliklerin birbirleri ile olan korelasyonları ifade etmektedir.

Korelasyon temelli özellik seçim süreci birinci adım olarak sınıf ile özellikler arasındaki korelasyonun belirlenmesi, özelliklerin kendi aralarındaki korelasyonların incelenmesi, ikinci adım kapsamında özellik alt kümelerinin değerlendirilmesi olarak iki adımdan oluşmaktadır. Korelasyon temelli özellik seçim süreci Şekil 4.12’de verilmiştir.



Şekil 4.12. Korelasyon Temelli Özellik Seçim Süreci

Sonsuz Gizli Özellik Seçimi (ILFS)

Sonsuz Gizli Özellik Seçimi yöntemi çapraz doğrulama gerçekleştirerek özellikleri içsel özelliklerine göre sıralayan bir filtreleme yöntemidir. Sonsuz Gizli Özellik Seçimi yöntemi özellik seçim işlemini, veri ön işleme, grafik ağırlıklandırma ve sıralama olarak 3 adımda gerçekleştirmektedir. Ön işleme adımında, özelliklerden her biri önemini öne çıkaracak olan bir tanımlayıcı ile temsil edilir. Grafik ağırlıklandırma adımında ise bütün özelliklere ait ağırlıkların bulunduğu gizli bir ağırlıklandırma grafiği oluşturulur. Bu adımdan sonra alt kümelerdeki her özelliğin problem ile olan ilişkisi ağırlıkları incelenerek hesaplanır. Oluşturulan grafikte bulunan her yol bir özellik seçimini temsil eder. Özellik seçim işlemi ile ortaya çıkan özellik sıralaması tüm olası alt kümelerdeki fazlalığa bakarak belirlenir [57,58].

Sıralı Yerellik ile Denetimsiz Özellik Seçimi (UFSOL)

Sıralı yerellik ile denetimsiz özellik seçimi yöntemi, örneklerin komşu yakınlıklarını koruyup mesafeye dayalı kümeleme tabanlı bir yaklaşımı benimseyerek özellik seçim işlemini gerçekleştirir. Sıralı yerellik ile denetimsiz özellik seçimi yöntemi özellikleri seçerken verilerin sıralı yapılarını koruyarak daha yüksek performans elde edilmesini sağlar. Bu yapısı ile birçok denetimsiz özellik seçim yönteminden performans olarak üstündür [59].

Bağımlılık Güdümlü Denetimsiz Özellik Seçimi (DGUFS)

Bağımlılık güdümlü denetimsiz öznetelik seçimi yöntemi, aynı anda kümeleme ve özellik seçim işlemini gerçekleştiren sarmalayıcı bir yöntemdir. Bu yöntemde özellik seçimi iki adımlı grafik tabanlı kümeleme ile gerçekleştirilir. İlk adım kümeleme ile özellikler arasındaki bağımlılığı ortadan kaldırırken, ikinci adım bağımlılığı en aza indirgeyen özelliklerin seçilmesini sağlar [60].

Relief-F Özellik Seçim Yöntemi

Relief-F özellik seçim yöntemi, komşuluk ilişkilerini temel alarak özellik seçim işlemini gerçekleştiren denetimli bir filtreleme yöntemidir. Bu yöntemde temel mantık, başlangıçta rastgele örnekler seçerek özelliğin ait olduğu örneğin bulunduğu ve bulunmadığı sınıflarda yer alan en yakın komşulara ağırlıklar atayarak özellik seçimini gerçekleştirmektir. Relief-F yöntemine göre iyi bir özellik, aynı sınıftaki örnekler için benzer, farklı sınıfta bulunan örnekler için farklı değerlere sahip olmalıdır [52,61]. Relief-F yönteminin çalışma mekanizması temelde dört adım ile açıklanabilir.

Adım 1: Başlangıçta rastgele örnekler seç.

Adım 2: Özelliğin ait olduğu örneğin bulunduğu ve bulunmadığı sınıflarda yer alan en yakın komşularla ilgili olan özellikleri belirle.

Adım 3: İlgili özelliklerin ağırlıklarını hesapla.

Adım 4: El yordamı ile ayarlanan bir eşik değerinin üstünde bulunan k adet özelliği seç ve özellik seçim işlemini sonlandır.

Sonsuz Özellik Seçimi (INFFS)

Sonsuz özellik seçimi yöntemi matrislere ait güç serilerinin yakınsaklık özelliklerinden yararlanan grafik tabanlı bir özellik seçim yöntemidir. Sonsuz özellik seçimi yönteminde her özellik alt kümesi yol olarak belirlenerek her özellik için bir önem değeri atanır. Sonsuz özellik seçimi yöntemine göre her özellik birer düğüm ve bu özelliklerin oluşturduğu özellik alt kümeleri ise yol olarak tanımlanır. Yöntem kapsamında bu yollar değişikliğe uğrayabilir ve sayısı sonsuza uzayabilir. Sonsuz özellik seçim yöntemi her yola bir son puan atayarak, yolun performans açısından önemini ortaya koyar [62].

Çoklu Küme Özellik Seçimi (MCFS)

Çoklu küme özellik seçim yöntemi en eski kullanıma sahip kümeleme tabanlı denetimsiz bir özellik seçim yöntemidir. Özelliklerin birbirleri ile olan korelasyonlarını belirlemek için bir spektral analiz gerçekleştirir. Ardından Laplacian grafiğini kullanarak kümeleme gerçekleştirilir ve her özellik için bir özellik puanı hesaplanır. Hesaplanan puan matrisine göre en önemli k adet özelliğin seçimi gerçekleştirilir [62,63].

Yerel Öğrenmeye Dayalı Kümeleme Kullanarak Özellik Seçimi (LLCFS)

Yerel öğrenmeye dayalı kümeleme yönteminde her özellik bir ağırlık ile ilişkilendirilir. Ardından ilişkilendirilen bu ağırlıklar her özelliğin uygunluğu göz önünde bulundurularak yerel öğrenmeye dayalı kümeleme yönteminin barındırdığı yerleşik düzenlemesine dâhil edilir. Böylece ağırlıklar iteratif bir şekilde güncellenmeye ve tahmin edilmeye devam edilir. İlgisiz özelliklerin mevcut ağırlıkları sıfıra indirgenebilir ve bu sayede önemsiz özellikler özellik kümesinden çıkarılır. Yerel öğrenmeye dayalı kümeleme yöntemi ile gerçekleştirilen işlem sonucunda ayrılmış bir küme matrisi ile seçilen özellikler elde edilir [64].

LASSO

Lasso yöntemi 1996 yılında Robert Tibshirani tarafından gömülü bir özellik seçim yöntemi olarak geliştirilmiştir [65]. Bu yöntem özellik seçim ve düzenleme işlemlerini aynı anda gerçekleştirerek tahminleme doğruluğuna paralel olarak model performansını dengede tutabilir. Lasso yöntemi öncelikle modelde kullanılan özelliklerinin mutlak değer toplamlarına bir sınırlandırma getirir. Her değişken için bir küçültme işlemi uygulanarak ilgisiz özelliklerin katsayıları sıfıra indirgenir. Katsayısı sıfıra indirgenemeyen özellikler modele dâhil edilir. Bu yöntem beraberinde birçok avantaj sağlamaktadır. Bu avantajlardan bazıları; veride sapmayı artırmadan yüksek tahmin performansı sağlamak, yorumlanabilirliği artırarak yüksek korelasyonu azaltmak olarak sıralanabilir [65].

Laplacian Skoru (LS)

Laplacian Skoru yönteminde her özelliğin Laplace değeri, varyansına ve yerellik koruma gücüne bakılarak değerlendirilir. Bu yöntemde, Manifold yapısı adı verilen ve Laplacian matrisi tarafından temsil edilen yapıyı en çok koruyan özellikler için yüksek ağırlıklı değerler atanır. Yüksek ağırlıklı değerler atanan özellikler içerisinde k adet özellik seçilerek süreç tamamlanır. Bu yapı birbirine yakın iki nesnenin aynı küme ile ilişkili olduğu fikrinden esinlenerek, yakın özellikler için benzer değerler, birbirinden uzak özellikler için ise farklı değer atanmalı prensibinden esinlenmiştir. Laplacian Skoru yöntemi için ana fikir en basit haliyle, yerellik koruma gücüne göre özellik seçim işleminin gerçekleştirilmesi olarak tanımlanabilir [66].

İçbükey Minimizasyon Yoluyla Özellik Seçimi (FSVFS)

İçbükey küçültme yoluyla özellik seçimi yöntemi, ikili sınıflandırma problemlerinin çözümünü yinelemeli olarak gerçekleştirdiği için sarmalayıcı özellik seçim yöntemi olarak kabul edilir. İçbükey küçültme yoluyla özellik seçimi yönteminde, destek vektör makinelerinde benzer mantıkla çalışan iki sınıf arasındaki ayrımı maksimize etmeye yarayan bir hiperdüzlem mevcuttur. Problem hiper düzlemde bulunan boyutları minimize edecek şekilde sınırlandırılarak özellik seçim işlemi gerçekleştirir [67].

Uyarlanabilir Yapısal Öğrenme İle Özellik Seçimi (FSASL)

Uyarlanabilir Yapısal Öğrenme ile Özellik Seçimi yöntemi yapı öğrenmeyi ve özellik seçimini bir arada kullanan denetimsiz bir özellik seçim yöntemidir. Bu yöntemde yapılar gerçekleştirilen özellik seçim işleminin çıktılarını kullanarak eşzamanlı ve uyarlamalı olarak güncellenir. Özetlenecek olursa FSASL yöntemi ile verinin altında bulunan anlamlı veri yapılarını koruyacak özelliklerin seçimi gerçekleştirilir [68].

Denetimsiz Ayrımcı Özellik Seçimi (EDFS)

Denetimsiz Ayrımcı Özellik Seçimi yöntemi gömülü bir özellik seçim yöntemidir. Denetimsiz Ayrımcı Özellik Seçimi yönteminde, dağılım matrislerinde ve özellikler arası ilişkilerde bulunan ayırt edici bilgiler kullanılarak özellik alt kümesinin seçimi gerçekleştirilir. Bu yöntemi diğer denetimsiz özellik seçim yöntemlerinden ayıran özellik ise özellik alt kümesini toplu olarak tüm özellik kümesinden elde etmesidir. UDFS probleme bazı ek kısıtlamalar koyar ve problemi optimize edebilmek için verimli bir model önerir. UDFS yöntemi özellik sıralamasını özellik ağırlıklarında göre büyükten küçüğe sıralar ve ağırlığı büyük olan k sayıda özelliği seçer [66].

Çalışma kapsamında kullanılan özellik seçim yöntemlerinin ait olduğu özellik seçim kolları Tablo 4.3’de verilmiştir.

Tablo 4.3. Özellik Seçim Yöntemlerine Ait Türler

Özellik Seçim Kolu	Özellik Seçim Yöntem Adı
Filtreleme Yöntemleri	ILFS, LS, Relief-F, MCFS
Sarmal Yöntemler	DGUFS, FSVFS
Gömülü Yöntemler	EDFS, LASSO, UFSOL

Çalışma kapsamında her özellik seçim türünden farklı yöntemler kullanılarak özellik seçim türlerinin ve yöntemlerinin çalışma üzerindeki etkisi incelenmiştir.

4.2.5. Sınıflandırma

Sınıflandırma işlemi makine öğreniminin temelini oluşturmaktadır. Sınıflandırma yöntemi veri içerisindeki bilgiyi ortaya çıkarmak için kullanılır. Sınıflandırma işlemini gerçekleştirmek için birçok makine öğrenimi algoritması mevcuttur. Sınıflandırma algoritmaları hedef bilgi doğrultusunda veriyi sınıflara ayırarak işlem yapar. Bu sınıflar $[0,1]$ formunda ikili sınıflar olabileceği gibi çoklu sınıflarda olabilir.

Çalışma kapsamında sınıflandırma işlemini gerçekleştirebilmek için DVM, KNN, Karar Ağaçları ve YSA kullanılmıştır. Çalışma iki aşamada gerçekleştirilmiştir. İlk adımda eğitim verisi ile eğitilen modele ikinci adımda test verileri girdi olarak verilerek modelin performansı değerlendirilmiştir. Veri seti sınıflandırma işleminde kullanılmak üzere k-katlı çapraz doğrulama ile eğitim ve test verisi olarak ayrılmıştır. Çalışma kapsamında k sayısı 5 olarak belirlenmiştir. K-katlı çapraz doğrulama yöntemi model geliştirilirken aşırı öğrenme ya da eksik öğrenme gibi sorunların önlenmesinde rol oynamaktadır. Test aşamasında ise en iyi model performans çıktısını elde etmeyi hedeflemektedir. Bu yöntemde ilk adım olarak eğitim verisi karıştırılıp bölünerek k adet eşit büyüklükte küme elde edilir. Aynı işlem k kez tekrarlanır ve her k adımda alt küme eğitim kümesinden çıkarılarak test verisi olarak kullanılır. Her k kümesi için bu işlemler gerçekleştirildiğinde, çapraz doğrulama süreci tamamlanmış olur [69].

Eğitim ve test süreçleri tamamlandıktan sonra model performansını değerlendirebilmek için model değerlendirme metriklerine ihtiyaç duyulmaktadır. Bu çalışmada sınıflandırma işleminin model performansını değerlendirebilmek için doğruluk, duyarlılık, özgüllük, hassasiyet ve F_1 -skor metrikleri kullanılmıştır.

4.3. Performans Değerlendirme

Sınıflandırma işleminin performansı doğru ve yanlış sınıflandırılan örneklerin karşılaştırılması ile elde edilmektedir. Çalışma kapsamında model değerlendirmek için kullanılan doğruluk, duyarlılık, özgüllük, hassasiyet ve F₁-skor metrikleri karmaşıklık matrisi kullanılarak elde edilmektedir. Sınıflandırma modelinin performans değerlendirme aşamasında kullanılan karmaşıklık matrisi Şekil 4.13’de verilmiştir.

		Tahmin Edilen Değer	
		Yanlış	Doğru
Gerçek Değer	Yanlış	Doğru Negatif (DN)	Yanlış Negatif (YN)
	Doğru	Yanlış Pozitif (YP)	Doğru Pozitif (DP)

Şekil 4.13. Karmaşıklık Matrisi

Şekil 4.13’de verilen karmaşıklık matrisinde satır değerleri bize gerçek sınıfları verirken, sütun değerleri tahmin edilen sınıfları göstermektedir. Yukarıda verilen karmaşıklık matrisinde kullanılan parametreler için temel tanımlar aşağıdaki gibidir [69]:

Doğru Pozitif (DP): Covid-pozitif olup Covid-pozitif olarak tanımlanmış hasta sayısı.

Yanlış Pozitif (YP): Covid-pozitif olup Covid-negatif olarak tanımlanmış hasta sayısı.

Doğru Negatif (DN): Covid-negatif olup Covid-negatif olarak tanımlanmış hasta sayısı.

Yanlış Negatif (YN): Covid-negatif olup Covid-pozitif olarak tanımlanmış hasta sayısı.

Yukarıda verilen temel tanımlardan faydalanarak elde edilen model değerlendirme metriklerine ait bilgiler aşağıda verilmiştir.

Doğruluk: Sınıflandırma sonucunda doğru olarak sınıflandırılmış örnek sayısının toplam girdi sayısına oranı kullanılarak hesaplanır. Hasta ve sağlıklı olan vakaların hangi doğrulukla tahmin edebildiğine dair sonuç veren değerdir. Denklem 4.3’de doğruluk değerinin nasıl hesaplandığı formüle edilmiştir.

$$Doğruluk = \frac{DP+DN}{DP+YP+DN+YN} \quad (4.3)$$

Duyarlılık: Duyarlılık değeri, doğru olarak sınıflandırılan pozitif örnek sayısının, toplam pozitif örnek sayısına oranı kullanılarak hesaplanır. Hasta olan vakaların doğru bir şekilde belirlenmesi için kullanılır. Denklem 4.4’de duyarlılık değerinin nasıl hesaplandığı formüle edilmiştir.

$$Duyarlılık = \frac{DP}{DP+YP} \quad (4.4)$$

Özgüllük: Özgüllük değeri, doğru olarak sınıflandırılan negatif örnek sayısının, toplam negatif örnek sayısına oranı kullanılarak hesaplanır. Sağlıklı olan vakaları doğru bir şekilde belirleme işlemidir. Özgüllük değeri matematiksel olarak Denklem 4.5’te verilmiştir.

$$Özgüllük = \frac{DN}{DN+YP} \quad (4.5)$$

Hassasiyet: Pozitif olan vakaları tahmin etme işlemidir. Hassasiyet değeri matematiksel olarak Denklem 4.6’da verilmiştir.

$$Hassasiyet = \frac{DP}{DP+YN} \quad (4.6)$$

F₁ Ölçütü: Hassasiyet ve duyarlılık arasındaki harmonik ortalama ile hesaplanıp, sınıflandırma işleminin genel doğruluğunu bize verir. F₁-skor değerinin nasıl hesaplandığı Denklem 4.7’de verilmiştir.

$$F1\ Score = \frac{2DP}{2DP+YP+YN} \quad (4.7)$$

5. BULGULAR VE TARTIŞMA

Tez çalışması Covid-19 hastalığının doğru bir şekilde tahmin edilebilmesi için iki yaklaşıma dayanarak gerçekleştirilmiştir. Birinci yaklaşımda veri setinde bulunan 221 hastaya ait 16 özellik sınıflandırma işlemine dâhil edilirken, ikinci yaklaşımda özellik seçim yöntemleri kullanılarak optimum özellik sayısı arayışı gerçekleştirilmiştir. Bu aramada veri kümesi sırası ile minimuma indirgenerek test edilmiş ve en iyi sonuçlar 5 özellik ile elde edilmiştir. Buna bağlı olarak özellik sayısı 5'e indirgenmiş ve model her yöntem için 5 özellik ile test edilmiştir. Özellik seçim işlemi ile daha az özellik kullanıp maliyeti düşürmek ve daha yüksek performans elde etmek hedeflenmiştir. Özellik seçim yöntemlerinin performansları karşılaştırmalı olarak değerlendirilmiştir. Tablo 5.1'de her özellik seçim yöntemi için seçilen özelliklerin sıralaması önem seviyelerine göre sıralı olarak verilmiştir.

Tablo 5.1. Özellik Seçim Yöntemleri ile Elde Edilen Özellikler

Özellik Seçim Yöntemi	Seçilen Özellikler
CFS	ALT, Alb., LDH, Prot, Krea
ILFS	Lenf., WBC, Tbil, Neu, Eos
UFSOL	Lenf., Neu, WBC, CRP, Alb
RELIEFF	Eos, Lenf., Alb, Tbil, WBC
INFFS	CRP, Dimer, Üre, Krea., Prot
MCFS	Eos, Mon, Neu, WBC, Lenf.
LLCFS	Dimer, LDH, PLT, Üre, CRP
LASSO	Mon, Krea., Prot., Eos, ALT
LAPLACIAN	Dimer, PLT, LDH, CRP, Üre
FSVFS	Dimer, Lenf, WBC, Prot, AST
FSALS	Alb, Lenf., WBC, Prot, Dimer
EDFS	Krea., Mon, Eos, Neu, Lenf
DGUFS	Tbil, Lenf, CRP, Dimer, Üre

Her bir özellik seçim yöntemi için seçilen özellikler ile 4 farklı sınıflandırıcı kullanılarak sınıflandırma işlemi gerçekleştirilmiştir. Tablo 5.2'de tüm özellikler kullanılarak gerçekleştirilen sınıflandırma işlemine ait performans değerleri ve her yöntem için seçilen 5 özellik ile elde edilen sınıflandırma işlemine ait başarımlar metrikleri verilmiştir.

Tablo 5.2. Her Özellik Seçim Yöntemi için Elde Edilen Sınıflandırma Performansı

		Doğruluk	Duyarlılık	Özgüllük	Hassasiyet	F1 Değeri
Tüm Özellikler	DVM	%86.4	%85.9	%87	%88.8	%87.3
	KNN	%81.4	%77.6	%86	%87	%82.1
	Karar A.	%85.9	%83.4	%89	%90	%86.7
	YSA	%90.6	%89.2	%92	%93.1	%91.1
NCA	DVM	%82.7	%79.3	%86	%87.2	%83.1
	KNN	%82.3	%79.8	%84.3	%87.1	%83.3
	Karar A.	%80.5	%77.5	%81.4	%85.6	%80.2
	YSA	%84.1	%82.6	%86	%87.7	%85.1
LASSO	DVM	%75	%70	%80	%80	%75.2
	KNN	%70.5	%66.9	%75	%76.4	%71.3
	Karar A.	%70	%66.1	%75	%76.1	%70.8
	YSA	%80.9	%74.3	%87	%87.3	%80.3
RELIEFF	DVM	%85	%84.5	%83.1	%87.4	%86.2
	KNN	%83.2	%80.1	%81	%86.1	%83
	Karar A.	%85.5	%85.9	%85.2	%88	%86.6
	YSA	%90.6	%90	%91	%92.3	%91.2
ILFS	DVM	%79.5	%71.7	%89	%87.7	%78.6
	KNN	%77.7	%69.3	%87	%85.5	%76.6
	Karar A.	%80.5	%72.9	%89	%88	%79.8
	YSA	%80.9	%72	%90	%89.8	%80
FSASL	DVM	%86	%82.6	%90	%90	%86.5
	KNN	%83.2	%82.6	%85	%86.9	%84.7
	Karar A.	%85.5	%82	%89	%90	%86.2
	YSA	%91.3	%91	%92	%93.2	%92
UFSOL	DVM	%75	%74.3	%76	%78.9	%76.6
	KNN	%80	%72	%90	%88.7	%79
	Karar A.	%79.1	%71.1	%88	%86.8	%78.2
	YSA	%82.3	%75.6	%89	%88.4	%81.5
DGUFS	DVM	%80.5	%76.8	%85	%86.1	%81.2
	KNN	%81.8	%78.5	%86	%87.1	%82.6
	Karar A.	%82.7	%78.5	%85	%86.3	%82.2
	YSA	%86.4	%83	%92.1	%91	%87
INFFS	DVM	%76.4	%78.5	%75	%79.1	%78.4
	KNN	%69.5	%66.1	%74	%75.4	%70.4
	Karar A.	%74.1	%74.3	%74	%77.5	%75.9
	YSA	%78.6	%80.1	%77	%82.3	%80.5
FSVFS	DVM	%85	%84.6	%82.8	%87.2	%85.9
	KNN	%86.4	%83	%92.1	%91	%87
	Karar A.	%83.2	%82.2	%84	%85.3	%82.6
	YSA	%85.5	%81.8	%88	%89.6	%86

LAPLACIAN	DVM	%76.9	%74.3	%80	%81.8	%77.9
	KNN	%75	%76.8	%73	%77.5	%77.1
	Karar A.	%75.5	%78.5	%72	%77.2	%77.8
	YSA	%77.7	%80.9	%74	%79	%80
MCFS	DVM	%79.1	%76.8	%82	%83.7	%80
	KNN	%78.2	%76	%81	%82.8	%79.3
	Karar A.	%82.3	%80.1	%85	%86.6	%83.2
	YSA	%76.4	%74.3	%79	%81	%77.5
UDFS	DVM	%78.6	%80.9	%76	%80.3	%80.6
	KNN	%75.5	%77.6	%73	%77	%77.6
	Karar A.	%72.7	%71	%75	%77.4	%74.1
	YSA	%81.8	%80.9	%83	%85.2	%83
CFS	DVM	%81.4	%78	%84	%86.5	%82
	KNN	%75.5	%72.7	%79	%80.7	%76.5
	Karar A.	%83.2	%80.1	%87	%88.1	%83.9
	YSA	%82.3	%79	%87	%88	%83.4
LLCFS	DVM	%79.1	%78.5	%80	%82.6	%80.5
	KNN	%73.2	%71	%75.4	%76.1	%73.5
	Karar A.	%76.8	%75.4	%78.1	%79.3	%77.3
	YSA	%77.7	%76.8	%79	%80	%78.4

Tablo 5.2 incelendiğinde özellik seçim yöntemleri ile elde edilen özelliklerin sınıflandırma performanslarının bazı yöntemler için tüm özelliklere göre daha başarılı olduğu görülmektedir. Relief-F yöntemi ile seçilen Eos, Lenf., Alb, Tbil, WBC parametrelerinin tüm özellikler ile elde edilen doğruluk oranı ile aynı oranda başarıyı yakaladığı görülmektedir. Bu durum incelendiğinde daha az parametre ile aynı performans çıkarımı elde edilmektedir. FSASL özellik seçim yöntemi ile elde edilen Alb, Lenf., WBC, Prot, Dimer parametrelerinin %91.3 doğruluk oranı ile tüm özellikler ile gerçekleştirilen sınıflandırma işlemine üstünlük sağladığı görülmektedir. Bu durum ulaşılmak istenen hedefe ulaşıldığını göstermektedir. Genel olarak yöntemlerin performansları incelendiğinde yüksek performans sağlayan algoritmanın Yapay Sinir Ağları olduğu gözlemlenmiştir. Tablo 5.3’de yüksek performans sağlayan özellik seçim yöntemlerine ait doğruluk oranları karşılaştırmalı olarak verilmiştir.

Tablo 5.3. Yüksek Performans Sağlayan Özellik Seçim Yöntemleri

	DVM	KNN	Karar Ağaçları	YSA
Tüm Özellikler	%86.4	%81.4	%85.9	%90.6
Relief-F	%85	%83.2	%85.5	%90.6
FSASL	%86	%83.2	%85.5	%91.3

Tablo 5.3 incelendiğinde tüm özellikler için en yüksek doğruluk oranının %90.6 performans sağlayan YSA ile elde edildiği görülmektedir. Doğruluk performansları sırasına göre DVM algoritması için bu oranın %86.4, Karar Ağaçları için %85.9 ve KNN algoritması için %81.4 olduğu görülmektedir.

Relief-F özellik seçim yöntemi ile elde edilen sonuçların %90.6 doğruluk veren YSA ile tüm özelliklerle aynı performansı sağladığı görülmektedir. Performans sırasına göre diğer algoritmalar incelendiğinde Karar Ağaçları ile %85.5, DVM ile %85 ve KNN algoritması ile %83.2 oranında doğruluk elde edildiği görülmektedir.

FSASL özellik seçim yöntemi YSA ile %91.3 oranında doğruluk performansı sağlayarak daha az özellik ile daha yüksek performans hedefine ulaşıldığını göstermektedir. YSA algoritmasını FSASL yöntemi için %86 doğruluk ile DVM takip etmektedir. %85.5 doğruluk oranı ile Karar Ağaçlarının ve %83.2 oran ile KNN algoritmasının anlamlı sonuçlar verdiği gözlemlenmektedir. Tablo 5.3’de verilen en yüksek doğruluk performansını sağlayan algoritmalara ait karmaşıklık matrisleri Şekil 5.1’de, çizgi grafiği üzerinde gösterimleri ise Şekil 5.2’de verilmiştir.

		Tahmin Edilen Değer	
		Yanlış	Doğru
Gerçek Değer	Yanlış	89	11
	Doğru	11	110

(a)

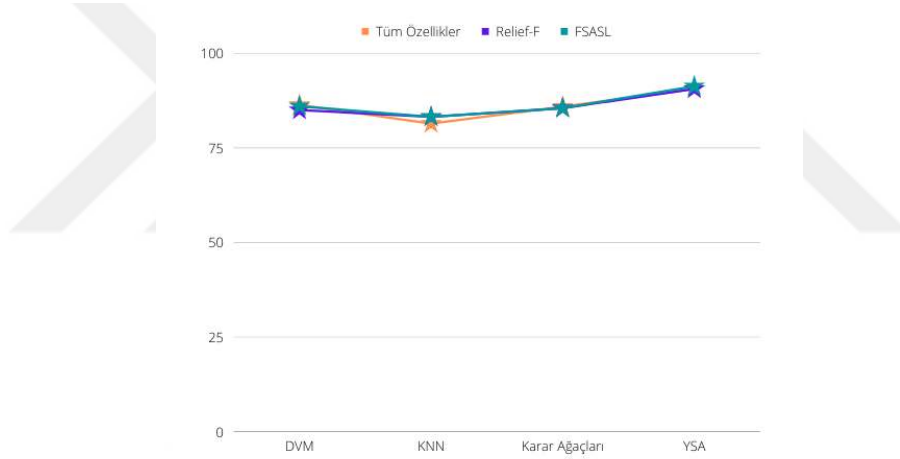
		Tahmin Edilen Değer	
		Yanlış	Doğru
Gerçek Değer	Yanlış	88	12
	Doğru	9	112

(b)

		Tahmin Edilen Değer	
		Yanlış	Doğru
Gerçek Değer	Yanlış	91	9
	Doğru	10	111

(c)

Şekil 5.1. (a) Tüm Özellikler İçin En İyi Performansa Sahip YSA Algoritmasının Karmaşıklık Matrisi
(b) Relief-F Yöntemi İle Seçilen Özellikler İçin En İyi Performansa Sahip YSA Algoritmasının Karmaşıklık Matrisi (c) FSASL Yöntemi ile Seçilen Özellikler İçin En İyi Performansa Sahip YSA Algoritmasının Karmaşıklık Matrisi



Şekil 5.2. Yüksek Performans Sağlayan Özellik Seçim Yöntemlerine Ait Performans Grafiği

Şekil 5.1 incelendiğinde elde edilen doğruluk performanslarına paralel bir dağılım gerçekleştiği grafik üzerinde de görülmektedir. Gerçekleştirilen çalışmada her özellik seçim yöntemi için tüm sınıflandırma algoritmaları karşılaştırmalı olarak değerlendirilmiş ve probleme olumlu anlamda katkısı olan yöntemlerin seçilimi gerçekleştirilmiştir. Elde edilen sonuçlar daha az parametre kullanarak model performansının iyileştirildiğini göstermektedir. Başlangıçta bulunan 16 özellik, özellik seçim yöntemleri ile 5 özelliğe indirilerek FSASL yönteminden elde edilen ve model performansına olumlu yönde etkisi bulunan özellikler sırasıyla Alb, Lenf., WBC, Prot, Dimer olarak belirlenmiştir. Elde edilen sonuçlar belirlenen bu özellik kümesinin problem üzerinde maliyet azaltıcı ve yüksek performans sağlayıcı etkisi olduğunu göstermektedir.

6. SONUÇLAR

Tez çalışmasında, salgın bir hastalık haline gelerek kalıcı hasarlara ve ölümlere sebebiyet verebilen Covid-19 hastalığını biyokimya parametreleri kullanarak teşhis etmeye yönelik, özellik seçme tabanlı olan ve olmayan iki model geliştirilmiştir. Toplamda 121 Covid-pozitif ve 100 Covid-negatif olmak üzere 221 hastaya ait biyokimya verisi çalışmaya dâhil edilmiştir.

Birinci modelde çalışmaya tüm biyokimya parametreleri dâhil edilerek sınıflandırma işlemi gerçekleştirilmiş ve %90.6 doğruluk, %89.2 duyarlılık, %92 özgüllük, %93.1 hassasiyet, %91.1 F1-skor oranlarıyla en yüksek başarımlı Yapay Sinir Ağları kullanılarak elde edilmiştir. İkinci model daha az özellik ile daha yüksek performans ve daha az maliyet ilkesi üzerinde inşa edilmiştir. İkinci modelde sınıflandırma işlemine 13 farklı özellik seçim yöntemi dâhil edilmiştir. Her yöntem için 5 özellik seçilmiş ve seçilen özellikler ile sınıflandırma işlemi gerçekleştirilmiştir. Relief-F algoritması ile seçilen Eos, Lenf., Alb, Tbil, WBC parametreleri kullanılarak gerçekleştirilen sınıflandırma işleminde %90.6 doğruluk oranı yakalanarak tüm özelliklerle kurulan modeldeki ile aynı doğruluk değeri elde edilmiştir. FSASL özellik seçim yöntemi incelendiğinde ise bu doğruluk oranı Alb, Lenf., WBC, Prot, Dimer parametreleri ile duyarlılık %91, özgüllük %92, hassasiyet %93.2, F₁-skor %92 ve doğruluk değerleri %91.3'e kadar yükselmiştir. İkinci model sonuçlarına bağımlı olarak, özellik seçim yöntemi uygulandıktan sonra gerçekleştirilen çalışmanın tüm özellikler kullanılarak gerçekleştirilen çalışmaya kıyasla daha başarılı olduğu gözlemlenmiştir.

Gerçekleştirilen çalışmanın Covid-19 tanısında sağladığı özellik eleme mekanizması sayesinde daha az maliyetle daha doğru sonuç üretmesi ve hekimlere yardımcı bir karar destek sistemi olması beklenmektedir.

KAYNAKLAR

- [1] Rustam, F., Reshi, A. A., Mehmood, A., Ullah, S., On, B.-W., Aslam, W., & Choi, G. S. (2020). COVID-19 Future Forecasting Using Supervised Machine Learning Models. *IEEE Access*, 8, 101489-101499. <https://doi.org/10.1109/ACCESS.2020.2997311>
- [2] *Makine Öğrenmesi Yöntemleri İle Covid-19 Verilerinin İncelenmesi: Türkiye Örneği görünümü.* (t.y.). Geliş tarihi 20 Mayıs 2022, gönderen https://jaihs.com/index.php/jaihs/article/view/i2y1or1/research_ozel
- [3] Özbay, E., & Altunbey Özbay, F. (2020). Derin Öğrenme ve Sınıflandırma Yaklaşımları ile BT görüntülerinden Covid-19 Tespiti. *DÜMF Mühendislik Dergisi*. <https://doi.org/10.24012/dumf.812810>
- [4] Bilge U. (2007). Tıpta Yapay Zekâ ve Uzman Sistemler, Biyoistatistik ve Tıp Bilimi AD, Akdeniz Üniversitesi Tıp Fakültesi, pp.113-118 Antalya
- [5] Bal, C., Serdar, M., Torun Güngör, O., Celik, H., Abusoglu, S., Uğuz, N., Erden, G., & Yıldırımka, M. (2014). Biyokimya parametrelerinin ölçüm belirsizliğinin hesaplanması 39, 538-543. <https://doi.org/10.5505/tjb.2014.04127>
- [6] Sayers, E. W., Barrett, T., Benson, D. A., Bolton, E., Bryant, S. H., Canese, K., Chetvernin, V., Church, D. M., DiCuccio, M., Federhen, S., Feolo, M., Fingerman, I. M., Geer, L. Y., Helmberg, W., Kapustin, Y., Krasnov, S., Landsman, D., Lipman, D. J., Lu, Z., ... Ye, J. (2012). Database resources of the National Center for Biotechnology Information. *Nucleic Acids Research*, 40(Database issue), D13-D25. <https://doi.org/10.1093/nar/gkr1184>
- [7] Bayot, L.M., Naidoo, P., (2019). Clinical Laboratory, *National Center for Biotechnology Information Search database*
- [8] Şahin A.R., Ateş S., Günay M., (2019). Klinik Mikrobiyoloji Laboratuvarlarında Yapay Zekânın Temel İşleyiş Modelleri, *J Biotechnol and Strategic Health Res*, ;3(2):66-71 DOİ: bshr.602790
- [9] Aljouie, A. F., Almazroa, A., Bokhari, Y., Alawad, M., Mahmoud, E., Alawad, E., Alsehawi, A., Rashid, M., Alomair, L., Almozaai, S., Albeshir, B., Alomaish, H., Daghistani, R., Alharbi, N. K., Alaamery, M., Bosaeed, M., & Alshaalan, H. (2021). Early Prediction of COVID-19 Ventilation Requirement and Mortality from Routinely Collected Baseline Chest Radiographs, Laboratory, and Clinical Data with Machine Learning. *Journal of Multidisciplinary Healthcare*, 14, 2017-2033. <https://doi.org/10.2147/JMDH.S322431>
- [10] *Development, evaluation, and validation of machine learning models for COVID-19 detection based on routine blood tests.* (t.y.). Geliş tarihi 20 Mayıs 2022, gönderen <https://www.degruyter.com/document/doi/10.1515/ccim-2020-1294/html>
- [11] Formica, V., Minieri, M., Bernardini, S., Ciotti, M., D'Agostini, C., Roselli, M., Andreoni, M., Morelli, C., Parisi, G., Federici, M., Paganelli, C., & Legramante, J. M. (2020). Complete blood count might help to identify subjects with high probability of testing positive to SARS-CoV-2. *Clinical Medicine*, 20(4), e114-e119. <https://doi.org/10.7861/clinmed.2020-0373>
- [12] Banerjee, A., Ray, S., Vorselaars, B., Kitson, J., Mamalakis, M., Weeks, S., Baker, M., & Mackenzie, L. (2020). *Use of Machine Learning and Artificial Intelligence to predict SARS-CoV-2 infection from Full Blood Counts in a population.*

- [13] Kukar, M., Gunčar, G., Vovko, T., Podnar, S., Černelč, P., Brvar, M., Zalaznik, M., Notar, M., Moškon, S., & Notar, M. (2021). COVID-19 diagnosis by routine blood tests using machine learning. *Scientific Reports*, 11(1), 10738. <https://doi.org/10.1038/s41598-021-90265-9>
- [14] Çubukçu, H. C., Topcu, D. İ., Bayraktar, N., Gülşen, M., Sarı, N., & Arslan, A. H. (2022). Detection of COVID-19 by Machine Learning Using Routine Laboratory Tests. *American Journal of Clinical Pathology*, 157(5), 758-766. <https://doi.org/10.1093/ajcp/aqab187>
- [15] Silveira, E. C. (2020). Prediction of COVID-19 From Hemogram Results and Age Using Machine Learning. *Frontiers in Health Informatics*, 9(1), 39. <https://doi.org/10.30699/fhi.v9i1.234>
- [16] Alves, M. A., Castro, G. Z., Oliveira, B. A. S., Ferreira, L. A., Ramírez, J. A., Silva, R., & Guimarães, F. G. (2021). Explaining machine learning based diagnosis of COVID-19 from routine blood tests with decision trees and criteria graphs. *Computers in Biology and Medicine*, 132, 104335. <https://doi.org/10.1016/j.compbiomed.2021.104335>
- [17] Yao, H., Zhang, N., Zhang, R., Duan, M., Xie, T., Pan, J., Peng, E., Huang, J., Zhang, Y., Xu, X., Xu, H., Zhou, F., & Wang, G. (2020). Severity Detection for the Coronavirus Disease 2019 (COVID-19) Patients Using a Machine Learning Model Based on the Blood and Urine Tests. *Frontiers in Cell and Developmental Biology*, 8. <https://www.frontiersin.org/article/10.3389/fcell.2020.00683>
- [18] Uğraş DiKmen, A., Kina, M., Özkan, S., & İlhan, M. N. (2020). COVID-19 Epidemiyolojisi: Pandemiden Ne Öğrendik. *JOURNAL OF BIOTECHNOLOGY AND STRATEGIC HEALTH RESEARCH*. <https://doi.org/10.34084/bshr.715153>
- [19] COVID-19 pandemisi. (2022). İçinde *Vikipedi*. https://tr.wikipedia.org/w/index.php?title=COVID-19_pandemisi&oldid=27750871
- [20] Budak, F., & Korkmaz, Ş. (2020). Covid-19 Pandemi Sürecine Yönelik Genel Bir Değerlendirme: Türkiye Örneği. *Sosyal Araştırmalar ve Yönetim Dergisi*, 1, 62-79. <https://doi.org/10.35375/sayod.738657>
- [21] Ali, D. M., Zake, L. G., & El Kady, N. K. (2021). Role of Chest Computed Tomography versus Real Time Reverse Transcription Polymerase Chain Reaction for Diagnosis of COVID-19: A Systematic Review and Meta-Analysis. *Interdisciplinary Perspectives on Infectious Diseases*, 2021, 8798575.
- [22] Akçay, E. (2021). *Pandemi Öncesi Dönem Ve Pandemi Sürecinde Covid 19 Tanısı Alan İle Covid 19 Tanısı Almayan Yoğun Bakım Hastalarının Bakteriyolojik Kültür Ve Duyarlılık Testi Sonuçlarının Karşılaştırılması*, Tıpta Uzmanlık Tezi, T.C. Sağlık Bilimleri Üniversitesi Ankara Sağlık Uygulama Ve Araştırma Merkezi
- [23] Bozkurt, İ., & Tezcan Keleş, G. (2021). COVID-19 Hastalığı Tanı ve Tedavisinde Kullanılan Laboratuvar Testleri. *Celal Bayar Üniversitesi Sağlık Bilimleri Enstitüsü Dergisi*. <https://doi.org/10.34087/cbusbed.865763>
- [24] Özmen, G.G. (2011). *Klinik Biyokimya Ölçüm Belirsizliği*, Uzmanlık Tezi, Yeditepe Üniversitesi, Tıp Fakültesi Biyokimya Anabilim Dalı
- [25] Bayot, L.M., Naidoo, P., (2019). Clinical Laboratory, *National Center for Biotechnology Information Search database*
- [26] Biyolojik Örnek Alma ve Nakli, (2015). *T.C Milli Eğitim Bakanlığı*

- [27] Öztaş, D. Y., & Bozdayi, D. G. (t.y.). *Klinik Laboratuvarlarda Temel Kavramlar*. 176.
- [28] Çon, N., (2020), Tıbbi Laboratuvar Teknikleri Programı, *Sağlık Hizmetleri Meslek Yüksekokulu*
- [29] *Klinik Laboratuvarlar Tedavinin Rotasına Yön Veriyor...* Geliş tarihi 20 Mayıs 2022, gönderen <https://www.ogunhaber.com/saglik/klinik-laboratuvarlar-tedavinin-rotasina-yon-veriyor-338691h.html>
- [30] Yılmaz, D. Ö. Ü. A., & Yayın, K. (2021). *Yapay Zekâ*. Kodlab Yayın Dağıtım Yazılım Ltd.Şti.
- [31] *Yapay zekâ nedir? Yapay Zekâ Uygulama Alanları Nelerdir?* (t.y.). Geliş tarihi 20 Mayıs 2022, gönderen <https://www.mediatick.com.tr/tr/blog/yapay-zeka-nedir>
- [32] Altunok, C., & Ünver, M. (2020). Medikal Endüstride Yapay Zeka ve Uzman Sistemlerin Sürekli İyileştirmeye Etkisi. *Academic Perspective Procedia*, 3(1), 482-490. <https://doi.org/10.33793/acperpro.03.01.95>
- [33] *Birol Akan Esen: Yapay Zekâ ile Doğal Zekanın Karşılaştırılması*. (t.y.). Geliş tarihi 20 Mayıs 2022, gönderen <https://www.akanesen.com/2017/08/yapay-zeka-ile-dogal-zekann.html>
- [34] Kılağız, Y., (1996). *Yapay Zekâ Bilgi İşlem Teknolojisi Ve Tek Aşamalı Karar Verme Problemleri İçin Bir Uzman Sistem Denemesi*, Doktora Tezi, Atatürk Üniversitesi, Fen Bilimleri Enstitüsü
- [35] Place, J.F., Truhead, A., Ozawa, K., Pardue, H., Schnipelsky, P., (1995). Use of artificial intelligence in analytical systems for the clinical laboratory, *Automatic Chemistry*, Cilt:17, Sayı: 1, pp.1-15
- [36] Atalay, M., & Çelik, E. (2017). Büyük Veri Analizinde Yapay Zekâ Ve Makine Öğrenmesi Uygulamaları - Artificial Intelligence and Machine Learning Applications in Big Data Analysis. *Mehmet Akif Ersoy Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 155-172. <https://doi.org/10.20875/makusobed.309727>
- [37] Demirhan, A., Kiliç, Y. A., & Güler, İ. (t.y.). *Tıpta Yapay Zeka Uygulamaları*. 11.
- [38] *Yapay Sinir Ağı(Artificial Neural Network) Nedir? - Veri Bilimi Okulu*. (2020, Mayıs 2). <https://www.veribilimiokulu.com/yapay-sinir-agiartificial-neural-network-nedir/>
- [39] Ersoy, E., & Karal, Ö. (2012). *Yapay Sinir Ağları Ve İnsan Beyni*. 2, 18.
- [40] *Yapay Sinir Ağları – ENNUR ÇITIR*. (t.y.). Geliş tarihi 20 Mayıs 2022, gönderen <https://ennurcitir.wordpress.com/2020/02/19/yapay-sinir-aglari/>
- [41] Yavuz, S., & Deveci, M. (2012). İstatiksel Normalizasyon Tekniklerinin Yapay Sinir Ağı Performansına Etkisi. *Erciyes Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 22.
- [42] El Naqa, I., & Murphy, M. (2015). *What Is Machine Learning?* (ss. 3-11). https://doi.org/10.1007/978-3-319-18305-3_1
- [43] *Denetimsiz öğrenme (Unsupervised learning) | yavuz blog*. (t.y.). Geliş tarihi 20 Mayıs 2022, gönderen <https://yavuz.github.io/denetimsiz-ogrenme/>
- [44] Bilgin, M. (t.y.). *Gerçek Veri Setlerinde Klasik Makine Öğrenmesi Yöntemlerinin Performans Analizi*. 6.
- [45] Ulgen, K., (2017). *Makine Öğrenimi Bölüm-2*, Medium. Alıntı: <https://medium.com/@k.ulgen90/makine%C3%B6%C4%9Frenimib%C3%B6l%C3%BCm-2-6d6d120a18e1>

- [46] Yakut, E., (2014). Yapay Sinir Ağları Ve Destek Vektör Makineleri Yöntemleriyle Borsa Endeksi Tahmini, *İktisadi ve İdari Bilimler Fakültesi Dergisi*, Cilt: 19, No: 1, pp.139-157.
- [47] Kavzoğlu, T., Colkesen, İ.,(2010). Karar Ağaçları İle Uydu Görüntülerinin Sınıflandırılması: Kocaeli Örneği, *Harita Teknolojileri Elektronik Dergisi*, Cilt: 2, No: 1, pp.36-45.
- [48] Bavaş, E., Alıntı: <http://erdoganb.com/2017/07/karar-agaclari-decision-trees-ile-veri-siniflandirma/>
- [49] *Introduction to one-class Support Vector Machines—Roemer's blog.* (t.y.). Geliş tarihi 20 Mayıs 2022, gönderen <http://rvlasveld.github.io/blog/2013/07/12/introduction-to-one-class-support-vector-machines/>
- [50] Making Data Normal Using Box-Cox Power Transformation. (2010, Şubat 26). *ISixSigma*. <https://www.isixsigma.com/tools-templates/normality/making-data-normal-using-box-cox-power-transformation/>
- [51] Yavuz, S., & Deveci, M. (2012). İstatiksel Normalizasyon Tekniklerinin Yapay Sinir Ağına Performansına Etkisi. *Erciyes Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 22.
- [52] Budak, H. (2018). Özellik Seçim Yöntemleri ve Yeni Bir Yaklaşım. *Süleyman Demirel Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, 22(Özel), 21. <https://doi.org/10.19113/sdufbed.01653>
- [53] Visalakshi, S., & Radha, V. (2015). A literature review of feature selection techniques and applications: Review of feature selection in data mining. <https://doi.org/10.1109/ICCIC.2014.7238499>
- [54] Karakaş, M., & Danışmanı, T. (t.y.). *Sınıflandırma Problemlerinde Özellik Seçimi İçin Karşılıklı Tabanlı Gri Kurt Optimizasyon Algoritması*. 80.
- [55] Karegowda, A. G., Manjunath, A. S., & Jayaram, M. A. (t.y.). *Comparative Study Of Attribute Selection Using Gain Ratio And Correlation Based Feature Selection*.
- [56] Emhan, Ö., & Akın, M. (2019). Filtreleme Tabanlı Öznelik Seçme Yöntemlerinin Anomali Tabanlı Ağ Saldırısı Tespit Sistemlerine Etkisi. *DÜMF Mühendislik Dergisi*, 10(2), 549-559. <https://doi.org/10.24012/dumf.565842>
- [57] Cai, L., Wu, H., & Zhou, K. (2021). Improved cancer biomarkers identification using network-constrained infinite latent feature selection. *PloS One*, 16(2), e0246668. <https://doi.org/10.1371/journal.pone.0246668>
- [58] Haider, F., Pollak, S., Albert, P., & Luz, S. (2021). Emotion recognition in low-resource settings: An evaluation of automatic feature selection methods. *Computer Speech & Language*, 65, 101119. <https://doi.org/10.1016/j.csl.2020.101119>
- [59] Tian, C., Lv, J., & Xu, X. (2021). Evaluation of Feature Selection Methods for Mammographic Breast Cancer Diagnosis in a Unified Framework. *BioMed Research International*, 2021, e6079163. <https://doi.org/10.1155/2021/6079163>
- [60] Roffo, G., Melzi, S., Castellani, U., Vinciarelli, A., & Cristani, M. (2021). Infinite Feature Selection: A Graph-based Feature Filtering Approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(12), 4396-4410. <https://doi.org/10.1109/TPAMI.2020.3002843>
- [61] Zhang, Y., Ren, X., Zhang, J. *Intrusion detection method based on information gain and ReliefF feature selection | IEEE Conference Publication | IEEE Xplore.* (t.y.). Geliş tarihi 20 Mayıs 2022, gönderen https://ieeexplore.ieee.org/abstract/document/8851756?casa_token=KqPWnRm_0bkAAAAA:CSdsXOFaxK0rfBGdCDVG02pEJghTj5mg1ND9nGtPLwSX33hSKjei79p7rJcFshu8yLB0Q
- [62] Bulut, C. (2018). Beyin Bilgisayar Arayüzü Uygulamaları İçin Eeg Sinyal Analizi, Yüksek Lisans Tezi, Altınbaş Üniversitesi, Fen Bilimleri Enstitüsü

- [63] Duangsoithong, R., & Windeatt, T. (2010). Correlation-Based and Causal Feature Selection Analysis for Ensemble Classifiers. İçinde F. Schwenker & N. El Gayar (Ed.), *Artificial Neural Networks in Pattern Recognition* (C. 5998, ss. 25-36). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-12159-3_3
- [64] Engin, M. A., Akçay, L., & Aras, S. (2022). Tek Kanallı Akciğer Seslerinde Süzgeç Tipi Özellik Seçim Yöntemlerini Kullanarak Solunum Patolojisinin Teşhisi. *European Journal of Science and Technology*. <https://doi.org/10.31590/ejosat.1082560>
- [65] Fonti, V., (2017). Feature Selection Using Lasso, *VU Amsterdam*
- [66] Solorio-Fernández, S., Carrasco-Ochoa, J. A., & Martínez-Trinidad, J. (2019). A review of unsupervised feature selection methods. *Artificial Intelligence Review*. <https://doi.org/10.1007/s10462-019-09682-y>
- [67] Adams, S., Meekins, R., & Beling, P. A. (2017). An Empirical Evaluation of Techniques for Feature Selection with Cost. *2017 IEEE International Conference on Data Mining Workshops (ICDMW)*, 834-841. <https://doi.org/10.1109/ICDMW.2017.153>
- [68] Du, L., & Shen, Y.-D. (2015). Unsupervised Feature Selection with Adaptive Structure Learning. *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 209-218. <https://doi.org/10.1145/2783258.2783345>
- [69] Özlüer Başer, B., Yangin, M., & Sarıdaş, E. S. (2021). Makine Öğrenmesi Teknikleriyle Diyabet Hastalığının Sınıflandırılması. *Süleyman Demirel Üniversitesi Fen Bilimleri Enstitüsü Dergisi*. <https://doi.org/10.19113/sdufenbed.842460>

ÖZGEÇMİŞ

Çağla DANACI

Category	Value
Category 1	Value 1
Category 2	Value 2
Category 3	Value 3
Category 4	Value 4
Category 5	Value 5
Category 6	Value 6
Category 7	Value 7

AKADEMİK FAALİYETLER

Bildiriler: Danacı, Ç., Arslan Tuncer, S., Ayyıldız, H., & Kalaycı, M. (2022). Determining Covid-19 with Relieff-Based Machine Learning Algorithms Using Biochemistry Parameters. *1st International Conference on Computing and Machine Intelligence (ICMI-2021)*