

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

**SAYISAL HARİTALAMA TEKNİKLERİ KULLANILARAK DNA
DİZİMLERİ ÜZERİNDEN LÖSEMİ HASTALIĞININ TEMEL TÜRLERİNİN
YAPAY ZEKA TABANLI ALGORİTMALAR İLE SINIFLANDIRILMASI**

DOKTORA TEZİ

Fatma AKALIN

Bilgisayar Mühendisliği Anabilim Dalı

Bilgisayar Mühendisliği Bilim Dalı

ŞUBAT 2023

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

**SAYISAL HARİTALAMA TEKNİKLERİ KULLANILARAK DNA
DİZİMLERİ ÜZERİNDEN LÖSEMİ HASTALIĞININ TEMEL TÜRLERİNİN
YAPAY ZEKA TABANLI ALGORİTMALAR İLE SINIFLANDIRILMASI**

DOKTORA TEZİ

Fatma AKALIN

Bilgisayar Mühendisliği Anabilim Dalı

Bilgisayar Mühendisliği Bilim Dalı

Tez Danışmanı: Prof. Dr. Nejat YUMUŞAK

ŞUBAT 2023

Fatma Akalın tarafından hazırlanan “Sayısal Haritalama Teknikleri Kullanılarak DNA Dizilimleri Üzerinden Lösemi Hastalığının Temel Türlerinin Yapay Zeka Tabanlı Algoritmalar ile Sınıflandırılması” adlı tez çalışması 13.02.2023 tarihinde aşağıdaki jüri tarafından oy birliği ile Sakarya Üniversitesi Fen Bilimleri Enstitüsü Bilgisayar Mühendisliği Anabilim Dalı Bilgisayar Mühendisliği Bilim Dalı’nda Doktora tezi olarak kabul edilmiştir.

Tez Jürisi

Jüri Başkanı :	Prof. Dr. Pakize ERDOĞMUŞ Düzce Üniversitesi	
Jüri Üyesi :	Prof. Dr. Nejat YUMUŞAK (Danışman) Sakarya Üniversitesi	
Jüri Üyesi :	Prof. Dr. İbrahim TÜRKOĞLU Fırat Üniversitesi	
Jüri Üyesi :	Prof. Dr. İsmail Hakkı CEDİMOĞLU Sakarya Üniversitesi	
Jüri Üyesi :	Dr. Öğr. Üyesi Ali GÜLBAĞ Sakarya Üniversitesi	



ETİK İLKE VE KURALLARA UYGUNLUK BEYANNAMESİ

Sakarya Üniversitesi Fen Bilimleri Enstitüsü Lisansüstü Eğitim-Öğretim Yönetmeliğine ve Yükseköğretim Kurumları Bilimsel Araştırma ve Yayın Etiği Yönergesine uygun olarak hazırlamış olduğum “SAYISAL HARİTALAMA TEKNİKLERİ KULLANILARAK DNA DİZİMLERİ ÜZERİNDEN LÖSEMİ HASTALIĞININ TEMEL TÜRLERİNİN YAPAY ZEKA TABANLI ALGORİTMALAR İLE SINIFLANDIRILMASI” başlıklı tezin bana ait, özgün bir çalışma olduğunu; çalışmamın tüm aşamalarında yukarıda belirtilen yönetmelik ve yönergeye uygun davrandığımı, tezin içerdiği yenilik ve sonuçları başka bir yerden almadığımı, tezde kullandığım eserleri usulüne göre kaynak olarak gösterdiğimi, bu tezi başka bir bilim kuruluna akademik amaç ve unvan almak amacıyla vermediğimi ve 20.04.2016 tarihli Resmi Gazete’de yayımlanan Lisansüstü Eğitim ve Öğretim Yönetmeliğinin 9/2 ve 22/2 maddeleri gereğince Sakarya Üniversitesi’nin abonesi olduğu intihal yazılım programı kullanılarak Enstitü tarafından belirlenmiş ölçütlere uygun rapor alındığını, çalışmamla ilgili yaptığım bu beyana aykırı bir durumun ortaya çıkması halinde doğabilecek her türlü hukuki sorumluluğu kabul ettiğimi beyan ederim.

(13/02/2023)

Fatma AKALIN



TEŞEKKÜR

Tez çalışmamın her aşamasında değerli tecrübeleri ile bana yol gösteren ve bilim dünyasına katkı sağlamanın hazzını bizzat örnek kimliği ile bana hissettiren danışman hocam Prof. Dr. Nejat Yumuşak’a sonsuz teşekkürlerimi sunarım.

Bu süreçte motivasyon ve desteğini benden asla esirgemeyen, yaşadığım tüm anlarımı güzel emekleri ile süsleyen ve her koşulda yanımda olan çok değerli aileme yürekten teşekkür ederim.

Fatma Akalın



İÇİNDEKİLER

Sayfa

ETİK İLKE VE KURALLARA UYGUNLUK BEYANNAMESİ	v
TEŞEKKÜR	vii
İÇİNDEKİLER	ix
KISALTMALAR	xiii
TABLO LİSTESİ	xv
ŞEKİL LİSTESİ	xvii
ÖZET	xix
SUMMARY	xxiii
1. GİRİŞ	1
1.1. DNA Genom Dizilimlerindeki Ekson ve İntron Bölgelerinin Sınıflandırılması	2
1.2. DNA Dizilimleri ile ALL ve KML Malignitelerinin Sınıflandırılması	4
1.3. Mikrodizi Veri Kümesi Kullanılarak ALL ve AML Malignitelerinin Sınıflandırılması	5
1.4. Mikrodizi Veri Kümesi Kullanılarak ALL, AML ve MLL Malignitelerinin Sınıflandırılması	6
1.5. Dijitalleştirilmiş Periferik Kan Yayması Görüntüleri Kullanılarak ALL Malignitesinin Tespiti	7
1.6. Mikrodizi Veri Kümesi Kullanılarak T-ALL, B-ALL ve T-LL Malignitelerinin Sınıflandırılması	8
1.7. Beyaz Kan Hücrelerinin Tespiti ve Sınıflandırılması	10
2. KAYNAK ARAŞTIRMASI	13
2.1. DNA Genom Dizilimlerindeki Ekson ve İntron Bölgelerinin Sınıflandırılmasına İlişkin Çalışmalar	13
2.2. DNA Dizilimleri Kullanılarak Anomalilerin Sınıflandırılmasına İlişkin Çalışmalar.....	15
2.3. Mikrodizi Veri Kümesi Kullanılarak Anomalilerin Sınıflandırılmasına İlişkin Çalışmalar.....	17
2.4. Mikrodizi Veri Kümesi Kullanılarak Anomalilerin Sınıflandırılmasına İlişkin Çalışmalar – 2.....	18
2.5. Dijitalleştirilmiş Periferik Kan Yayma Görüntüleri Kullanılarak Anomalilerin Tespitine İlişkin Çalışmalar	20
2.6. Mikrodizi Veri Kümesi Kullanılarak Anomalilerin Sınıflandırılmasına İlişkin Çalışmalar -3	22
2.7. Beyaz Kan Hücrelerinin Tespiti ve Sınıflandırılmasına İlişkin Çalışmalar	23
3. MATERYAL VE YÖNTEM.....	25
3.1. Materyal – Kullanılan Araçlar.....	25
3.2. Veri Kümeleri.....	25

3.2.1. DNA genom dizilimlerindeki ekson ve intron bölgelerinin sınıflandırılması için tercih edilen veri kümesi	25
3.2.2. DNA dizilimleri ile ALL ve KML malignitelerinin sınıflandırılması için tercih edilen veri kümesi	26
3.2.3. Mikrodizi veri kümesi kullanılarak ALL ve AML malignitelerinin sınıflandırılması için tercih edilen veri kümesi	28
3.2.4. Mikrodizi veri kümesi kullanılarak ALL, AML ve MLL malignitelerinin sınıflandırılması için tercih edilen veri kümesi	28
3.2.5. Dijitalleştirilmiş periferik kan yayma görüntüleri kullanılarak ALL malignitesinin tespiti için tercih edilen veri kümesi.....	29
3.2.6. Mikrodizi veri kümesi kullanılarak T-ALL, B-ALL ve T-LL malignitelerinin sınıflandırılması için tercih edilen veri kümesi	30
3.2.7. Mikrodizi veri kümesi kullanılarak T-ALL, B-ALL ve T-LL malignitelerinin sınıflandırılması için tercih edilen veri kümesi	31
3.3. Yöntemler	32
3.3.1. DNA genom dizilimlerindeki ekson ve intron bölgelerinin sınıflandırılmasına ilişkin iş akışı.....	32
3.3.1.1. DNA genom dizilimlerindeki ekson ve intron bölgelerinin sınıflandırılması için kullanılan yöntemler	35
3.3.2. DNA dizilimleri ile ALL ve KML malignitelerinin sınıflandırılmasına ilişkin iş akışı.....	46
3.3.2.1. DNA dizilimleri ile ALL ve KML malignitelerinin sınıflandırılması için kullanılan yöntemler.....	47
3.3.3. Mikrodizi veri kümesi kullanılarak ALL ve AML malignitelerinin sınıflandırılmasına ilişkin iş akışı.....	57
3.3.3.1. Mikrodizi veri kümesi kullanılarak ALL ve AML malignitelerinin sınıflandırılması için kullanılan yöntemler	59
3.3.4. Mikrodizi veri kümesi kullanılarak ALL, AML ve MLL malignitelerinin sınıflandırılmasına ilişkin iş akışı.....	68
3.3.4.1. Mikrodizi veri kümesi kullanılarak ALL, AML ve MLL malignitelerinin sınıflandırılması için kullanılan yöntemler	69
3.3.5. Periferik kan yayma görüntüleri üzerinde YOLOv4 algoritması kullanılarak ALL hastalığının erken teşhisine ilişkin iş akışı	70
3.3.5.1. Periferik kan yayma görüntüleri üzerinde YOLOv4 algoritması kullanılarak ALL hastalığının erken teşhisi için kullanılan yöntemler	71
3.3.6. Mikrodizi veri kümesi kullanılarak T-ALL, B-ALL ve T-LL malignitelerinin sınıflandırılmasına ilişkin iş akışı	73
3.3.6.1. Mikrodizi veri kümesi kullanılarak T-ALL, B-ALL ve T-LL malignitelerinin sınıflandırılması için kullanılan yöntemler	75
3.3.7. Beyaz kan hücrelerinin tespitine ve sınıflandırılmasına ilişkin iş akışı	77
3.3.7.1. Beyaz kan hücrelerinin tespiti ve sınıflandırılması için kullanılan yöntemler.....	78
4. ARAŞTIRMA BULGULARI	85
4.1. DNA Genom Dizilimlerindeki Ekson ve İntron Bölgelerinin Sınıflandırılmasının Sonucunda Elde Edilen Bulgular.....	85
4.2. DNA Genom Dizilimleri ile ALL ve KML Malignitelerinin Sınıflandırılmasının Sonucunda Elde Edilen Bulgular.....	95

4.3. Mikrodizi Veri Kümesi Kullanılarak ALL ve AML Malignitelerinin Sınıflandırılmasının Sonucunda Elde Edilen Bulgular	97
4.4. Mikrodizi Veri Kümesi Kullanılarak ALL, AML ve MLL Malignitelerinin Sınıflandırılmasının Sonucunda Elde Edilen Bulgular	100
4.5. Periferik Kan Yayma Görüntüleri Kullanılarak ALL Malignitesine İlişkin Tespitin Sonucunda Elde Edilen Bulgular	100
4.6. Mikrodizi Veri Kümesi Kullanılarak T-ALL, B-ALL ve T-LL Malignitelerinin Sınıflandırılmasının Sonucunda Elde Edilen Bulgular	102
4.7. Beyaz Kan Hücrelerinin Tespiti ve Sınıflandırılması Sonucunda Elde Edilen Bulgular	104
5. TARTIŞMA	109
5.1. DNA Genom Dizilimlerdeki Ekson ve İntron Bölgelerinin Sınıflandırılmasının Sonucunda Ulaşılan Çıktılar.....	109
5.2. DNA Dizilimleri ile ALL ve KML Malignitelerinin Sınıflandırılmasının Sonucunda Ulaşılan Çıktılar.....	111
5.3. Mikrodizi Veri Kümesi Kullanılarak ALL ve AML Malignitelerinin Sınıflandırılmasının Sonucunda Ulaşılan Çıktılar.....	112
5.4. Mikrodizi Veri Kümesi Kullanılarak ALL, AML ve MLL Malignitelerinin Sınıflandırılmasının Sonucunda Ulaşılan Çıktılar.....	114
5.5. Dijitalleştirilmiş Periferik Kan Yayma Görüntüleri Kullanılarak ALL Malignitesinin Tespitinde Ulaşılan Çıktılar	115
5.6. Mikrodizi Veri Kümesi Kullanılarak T-ALL, B-ALL ve T-LL Malignitelerinin Sınıflandırılmasının Sonucunda Ulaşılan Çıktılar.....	117
5.7. Beyaz Kan Hücrelerinin Tespitinde ve Sınıflandırılmasında Ulaşılan Çıktılar	119
6. SONUÇLAR	125
KAYNAKLAR	127
ÖZGEÇMİŞ.....	141



KISALTMALAR

ABC	: Artificial Bee Colony (Yapay Arı Kolonisi)
A	: Adenin
AHE	: Adaptive Histogram Equalization (Adaptif Histogram Eşitleme)
ALL	: Akut Lenfoblastik Lösemi
ALL-IDB	: Acute Lymphoblastic Leukemia Image Database (Akut Lenfoblastik Lösemi Görüntü Veri Kümesi)
AML	: Akut Miyeloid Lösemi
ANFIS	: Adaptive Neuro Fuzzy Inference System (Uyarlanabilir Nöro Bulanık Çıkarım Sistemi)
ACO	: Ant Colony Optimization (Karıncı Kolonisi Optimizasyonu)
B-ALL	: B-cell acute lymphoblastic leukemia (B-hücreli akut lenfoblastik lösemi)
BO	: Başarı Oranı
CNN	: Convolutional Neural Network (Evrişimsel Sinir Ağı)
CSP	: Cross-Stage-Partial-Connections (Çapraz Aşamalı Kısmi Bağlantılar)
CLAHE	: Contrast Limited Adaptive Histogram Equalization (Kontrast Sınırlı Uyarlanabilir Histogram Eşitleme)
DN	: Doğru Negatif
DNA	: Deoxyribonucleic Acid (Deoksiribo Nükleik Asit)
DGCNN	: Deep Graph Convolutional Neural Network (Derin Graf Evrişimsel Sinir Ağı)
DP	: Doğru Pozitif
DYR	: Duyarlılık
F Ö.	: F Ölçütü
FPN	: Feature Pyramid Network (Özellik Piramit Ağı)
GMDH	: Group Method of Data Handling (Grup Veri İşleme Yöntemi)
GNN	: Graph Neural Network (Graf Sinir Ağı)
G	: Guanin
GÜF	: Gauss Üyelik Fonksiyonu
HE	: Histogram equalization (Histogram Eşitleme)

IIR	: Individual Image-Level Representation (Bireysel Görüntü Düzeyinde Temsil)
İÇE	: İkili çapraz entropi
KLL	: Kronik Lenfositik Lösemi
KML	: Kronik Miyeloid Lösemi
KNN	: K-Nearest Neighbors (K En Yakın Komşular)
KSN	: Kesinlik
LSTM	: Long Short-Term Memory (Uzun Kısa Süreli Bellek)
MLL	: Mixed-Lineage Leukemia (Karışık Soy Lösemisi)
MPNN	: Message Passing Neural Networks (Mesaj Geçiren Sinir Ağları)
NCBI	: National Center for Biotechnology Information (Ulusal Biyoteknoloji Bilgi Merkezi)
ÖZG	: Özgüllük
PAN	: Path Aggregation Network (Yol Toplama Ağı)
Ph Kromozu	: Philadelphia kromozomu
PSO	: Particle Swarm Optimization (Parçacık Sürü Optimizasyonu)
RAR	: Relation-Aware Representation (İlişkiye Duyarlı Temsil)
SDD	: Sürekli Dalgacık Dönüşümü
C	: Sitozin
SPM	: Spatial Pyramid Matching (Uzaysal Piramit Eşleştirme)
SPP	: Spatial Pyramid Pooling (Uzaysal Piramit Havuzu)
T-ALL	: T-Cell Acute Lymphoblastic Leukemia (T Hücreli Akut Lenfoblastik Lösemi)
T-LL	: T-cell lymphoblastic lymphoma (T-Hücreli Lenfoblastik Lenfoma)
T	: Timin
YN	: Yanlış Negatif
YP	: Yanlış Pozitif
YOLO	: You Only Look Once
WHO	: World Health Organization (Dünya Sağlık Örgütü)

TABLO LİSTESİ

Sayfa

Tablo 3.1. BCR-ABL genlerindeki ekson ve intron bölgelerinin lokasyonları.....	25
Tablo 3.2. ALL ve KML lösemi türlerine ilişkin DNA dizilimleri.....	26
Tablo 3.3. Sayısal haritalama tekniklerine göre sayısallaştırılan örnek bir DNA dizilimi.	38
Tablo 3.4. Sayısal haritalama tekniklerinin avantajları ve dezavantajları.....	39
Tablo 3.5. Derinlik sayılarına göre bazı transfer öğrenme mimarileri.	45
Tablo 4.1. Reel haritalama tekniği için elde edilen deneysel ölçütler.....	86
Tablo 4.2. Moleküler kütle haritalama tekniği için elde edilen deneysel ölçütler. ...	87
Tablo 4.3. EIIP haritalama tekniği için elde edilen deneysel ölçütler.....	87
Tablo 4.4. Shannon entropi temelli haritalama tekniği için elde edilen deneysel ölçütler.....	87
Tablo 4.5. Eşleştirilmiş haritalama tekniğinin hidrojen bağı enerji kuralı için elde edilen deneysel ölçütleri.....	88
Tablo 4.6. Kısa zamanlı fourier dönüşümü kullanılarak elde edilen spektrogramların DGCNN yaklaşımı ile sınıflandırılmasının sonucunda elde edilen deneysel ölçütler.....	95
Tablo 4.7. Sürekli dalgacık dönüşümü kullanılarak elde edilen spektrogramların DGCNN yaklaşımı ile sınıflandırılmasının sonucunda elde edilen deneysel ölçütler.....	96
Tablo 4.8. Uyarlanabilir bulanık mantık yaklaşımı kullanılarak gerçekleştirilen sınıflandırma sonucunda elde edilen deneysel ölçütler.....	97
Tablo 4.9. Karınca optimizasyon algoritması kullanılarak seçilen potansiyel genlerin sınıflandırma performansı üzerinde etkisini gösteren deneysel kriterler. .	98
Tablo 4.10. Parçacık sürü optimizasyon algoritması kullanılarak seçilen potansiyel genlerin sınıflandırma performansı üzerinde etkisini gösteren deneysel kriterler.....	99
Tablo 4.11. Balina optimizasyon algoritması kullanılarak seçilen potansiyel genlerin sınıflandırma performansı üzerinde etkisini gösteren deneysel kriterler. .	99
Tablo 4.12. T-ALL, B-ALL ve T-LL malignitelerinin tespit performansına ilişkin değerlendirme kriterleri.....	104
Tablo 4.13. YOLOv5x ve önerilen YOLOv5x mimarilerine ilişkin performans metrikleri.	106



ŞEKİL LİSTESİ

Sayfa

Şekil 1.1. Ökaryot hücre DNA'sına ilişkin görsel tanımlamalar.....	3
Şekil 1.2. BCR-ABL genine ilişkin translokasyon sürecinin görsel tasviri.....	4
Şekil 1.3. Hematopoetik kök hücrenin farklılaşma aşamaları.	9
Şekil 3.1. Leukemia isimli veri kümesine ait gen ekspresyon verileri.	28
Şekil 3.2. MLL isimli veri kümesine ait gen ekspresyon verileri.	29
Şekil 3.3. ALL-IDB1 veri kümesindeki kan elementleri ve bu kan elementleri arasından hematologlar tarafından tanımlanan blast hücrelerinin koordinatları.	30
Şekil 3.4. ALL-IDB2 veri kümesindeki blast hücre ve sağlıklı hücre.....	30
Şekil 3.5. GSE1577 isimli veri kümesine ait gen ekspresyon verileri.....	31
Şekil 3.6. Raabin Health veri kümesinden tedarik edilen dijitalleştirilmiş beyaz kan hücreleri.....	32
Şekil 3.7. Ekson ve intron bölgelerinin sınıflandırılması amacıyla ilk adımda önerilen akış şeması.	33
Şekil 3.8. Örnek bir DNA diziliminden görüntüye dönüşüm sürecini ifade eden temsili akış.....	34
Şekil 3.9. Ekson ve intron bölgelerinin sınıflandırılması amacıyla ikinci adımda önerilen akış şeması.	35
Şekil 3.10. Durağan olan ve durağan olmayan zaman serilerine ilişkin dönüşümün tasviri.....	41
Şekil 3.11. Zaman serisine uygulanan kısa zamanlı fourier dönüşümü ve pencereleme.	43
Şekil 3.12. Ekson ve intron bölgelerinin sınıflandırılması için çalışmada kullanılan transfer öğrenme mimarilerinin iş akışı.....	46
Şekil 3.13. Löseminin temel türleri olan ALL ve KML malignitelerinin sınıflandırılmasına ilişkin ikinci bölümde önerilen akış diyagramı.....	47
Şekil 3.14. Dalgacık dönüşümünün gösterimi.	49
Şekil 3.15. DGCNN mimarisinin temsili yapısı.	50
Şekil 3.16. KNN temelli komşuluk matrisinin gösterimi.	51
Şekil 3.17. 4 katmanlı graf evrimsel sinir ağının gösterimi.	52
Şekil 3.18. Temsili ANFIS yapısı.	56
Şekil 3.19. Löseminin temel türleri olan ALL ve AML malignitelerinin sınıflandırılmasına ilişkin üçüncü bölümde önerilen akış diyagramı.	58
Şekil 3.20. Örnek bir spektrogram üzerinde CLAHE yönteminin uygulanma aşamaları.	64
Şekil 3.21. K en yakın komşu algoritmasının çalışma akışı.	65
Şekil 3.22. N adet karar ağacının bir araya gelmesi ile oluşan bir orman yapısı.....	67
Şekil 3.23. Temsili destek vektör makineleri algoritmasının çalışma hiyerarşisi.....	67

Şekil 3.24. Löseminin temel türleri olan ALL, AML ve MLL malignitelerinin sınıflandırılmasına ilişkin dördüncü bölümde önerilen akış diyagramı.	68
Şekil 3.25. LSTM sinir ağının temsili yapısı	69
Şekil 3.26. Löseminin temel türü olan ALL malignitesinin sınıflandırılmasına ilişkin beşinci bölümde önerilen akış diyagramı.	71
Şekil 3.27. YOLO nesne tanıma algoritmasının çalışma akışı.	72
Şekil 3.28. YOLOv4 yaklaşımının mimarisi.	73
Şekil 3.29. Lösemi hastalığına ilişkin T-ALL, B-ALL ve T-LL malignitelerinin sınıflandırılması için altıncı bölümde önerilen akış diyagramı.	74
Şekil 3.30. Lösemi hastalığının tanısında önemli bir biyobelirteç olan lenfosit hücre tipinin yüksek bir doğruluk oranı ile tespiti için yedinci bölümde önerilen akış şeması.	78
Şekil 3.31. Regresyon/Sınıflandırma temelli ve bölge önerisi temelli iki ayrı yönteme ilişkin değerlendirme yaklaşımlarının farkı.	79
Şekil 3.32. YOLOv5 yaklaşımının mimarisi.	80
Şekil 3.33. Detectron yaklaşımındaki temsili yol yapısı.	81
Şekil 4.1. Hidrojen bağı enerji kuralı ile sayısallaştırılan DNA dizilimlerinin EfficientNetB7 transfer öğrenme mimarisi kullanılarak sınıflandırılmasının sonucunda ekson ve intron bölgelerinin ayırt edilmesine ilişkin başarımın ROC eğrisi ile tasviri.	91
Şekil 4.2. BCR-ABL gen dizilimlerindeki ekson ve intron bölgelerine ilişkin spektrogramlar.	92
Şekil 4.3. 4 farklı test kümesi üzerinde gerçekleştirilen sınıflandırma sonucunda elde edilen doğruluk oranları.	92
Şekil 4.4. 4 farklı test kümesi üzerinde gerçekleştirilen sınıflandırma sonucunda elde edilen kayıp değerleri.	93
Şekil 4.5. Veri çoğaltma yaklaşımı ile sayısı artırılan spektrogramların EfficientNetB7 derin öğrenme mimarisi ile sınıflandırılması.	94
Şekil 4.6. ALL-IDB1+ALL-IDB2 veri kümesi üzerinde 3000 iterasyon boyunca gerçekleştirilen eğitim süreci için üretilen kayıp değerleri.	101
Şekil 4.7. ALL-IDB1 test veri kümesinden rastgele seçilen iki örnek için eğitilen model tarafından üretilen çıktı.	102
Şekil 4.8. ALL-IDB2 test veri kümesinden rastgele seçilen iki örnek için eğitilen model tarafından üretilen çıktı.	102
Şekil 4.9. Lenfosit hücrelerinin tanısında kullanılan YOLOv5s, YOLOv5x ve Detectron2 sinir ağı mimarileri ile elde edilen karışıklık matrisleri.	105
Şekil 4.10. Önerilerilen YOLOv5x mimarisinin YOLOv5x mimarisine göre lenfosit nesne tanıma oranında gösterdiği iyileşmenin kesinlik-duyarlılık eğrisi ile tasviri.	105
Şekil 4.11. Genel nesne tespit performansının artırılması için hibrit yapı kullanılarak elde edilen karışıklık matrisleri.	106

SAYISAL HARİTALAMA TEKNİKLERİ KULLANILARAK DNA DİZİLİMLERİ ÜZERİNDEN LÖSEMİ HASTALIĞININ TEMEL TÜRLERİNİN YAPAY ZEKA TABANLI ALGORİTMALAR İLE SINIFLANDIRILMASI

ÖZET

Kanser, vücudun herhangi bir yerindeki hücrelerin kontrolsüz çoğalması ile ortaya çıkan bir hastalıktır. Bu hastalık ile vücudun düzenli bir şekilde çalışan mekanizması yavaş yavaş bozulur ve erken teşhisin sağlanamaması ile düzeltilemeyecek bir duruma gelir. Bu süreç hastalığın türünün konulabilmesi ve tedavi yöntemlerinin belirlenebilmesi açısından önemlidir. Bu çalışmada kanser türleri arasında yer alan lösemi malignitesi irdelenmiştir. Çünkü lösemi, en sık rastlanan kanser türüdür. Tüm yaş gruplarında görülme ihtimali vardır. Hastalığın kişide seyretme durumuna göre akut veya kronik olarak iki grupta incelenir. Özellikle hastalığın aniden belirdiği ve vücut içerisinde yayılımın hızlı gerçekleştiği akut lösemi hastalarında uzun bir yaşam beklentisi için erken tanı şarttır. Bununla birlikte yaklaşık 40 yıl önce doktorlar tarafından löseminin temel alt türlerinden olan ALL ve AML tanısının konulduğu belli hasta tiplerinde karışık soy lösemisi (MLL) olarak tanımlanan bir lösemi türü keşfedilmiştir. MLL lösemi türünün yeni bir tür olarak keşfedilmesinden sonra lösemi hastalığının halen keşfedilebilecek yeni alt oluşumlar barındırdığı düşünülmektedir. Geniş bir çerçevede ele alınan lösemi kanserine ilişkin alt türlerin net bir şekilde ayırt edilmesi, doğru tanı sürecinin gerçekleşmesi ile birlikte doğru tedavi protokolünün uygulanmasını sağlayacaktır. Öte yandan tedaviye verilen yanıtın değerlendirilmesi ve sonraki tedaviler için uygulanacak protokolün netleştirilmesi açısından da hayatidir. Tıp dünyasında kesin sonuçlara ulaşmak için periferik kan yayması, kemik iliği aspirasyonu, kemik iliği biyopsisi, immüfenotipleme, çeşitli görüntülemeler ve testler kullanılmaktadır. Ancak bu yöntemler bazen belirsiz durumlar içermektedir. Örneğin, periferik kan yayma yönteminin manuel değerlendirme süreci doktorun bilgi, tecrübe, fiziksel yoğunluğu ve zihinsel yorgunluğu gibi parametrelere bağlıdır. Kanser hastası olan bir kişiden biyopsi yöntemi ile alınan parçada kanserli doku yer almayabilir. Löseminin tanısında ayrıca olan lenfosit hücre tipine ilişkin artış aynı zamanda hepatit virüsleri ve brusella hastalığı sonucunda da artabilir. Görüntüleme yöntemlerinin çeşitli tıbbi parametrelere bağlı olarak uygulanma başarısı ya da ulaşılan görüntülemelerin doktorlar tarafından değerlendirilmesi standart bir çıktı üretmeyebilir. Bu nedenle genetik temelli maligniteler üzerinde moleküler ve sitogenetik değerlendirme yapmak tanının doğruluğunu netleştirecektir. Ancak genom teknolojisindeki son gelişmeler DNA dizilimlerinde bir artışa neden olmuştur. Bu nedenle sitogenetik analiz doğrultusunda verilerin manuel olarak doğru ve hızlı bir şekilde yorumlanması güçleşmiştir. Öte yandan moleküler değerlendirme imkanı sunan mikrodizi teknolojisi değerlendirilmiştir. Ancak kanser araştırmalarında umut kaynağı olan mikrodizi teknolojisinin analiz işlemi, genleri temsil eden özellikler üzerinden sağlandığı için yüksek boyut sunar. Makine öğrenmesinde boyutluluğun laneti olarak bilinen bu

durum sonucunda hesaplama yükü ve yanlış kararlar üreten gürültü oluşabilir. Tüm bu yöntemler kapsamında kişi üzerinde uygulanması planlanan tedavi yaklaşımlarının sayısı, maliyet parametresi çerçevesinde hasta üzerinde anksiyeteye neden olabilir. Aynı zamanda gereksiz radyasyon maruziyeti ve doktorlar için iş yükü ve zaman kaybı diğer olumsuz çıktılardır. Bu çalışmada daha az tıbbi yöntem ile erken tanının sağlanması amacıyla 7 ayrı kısımda bilgisayar destekli bir yapı inşa edilmiştir. Doktorlara fikir vermek amacı ile tasarlanan bu yapı ile 7 farklı amaç doğrultusunda lösemi hastalığı analiz edilmiştir.

Sitogenetik değerlendirmenin sağlandığı ilk aşamada, löseminin temel türlerinden olan ALL ve KML malignitelerinin tanısında önemli bir ayıraç olan BCR-ABL genleri analiz edilmiştir. NCBI veri kümesinden tedarik edilen BCR-ABL genleri haritalama teknikleri kullanılarak sayısallaştırılmıştır. Ardından sayısallaştırılan dizilimler üzerinde canlılığa ilişkin kritik bilgiler içeren ekson bölgelerinin tespiti için sinyal işleme yaklaşımı kapsamında fourier dönüşümü ve kısa zamanlı fourier dönüşümü yöntemleri kullanılmıştır. Sinyal işleme yöntemleri ile DNA dizilimlerinden elde edilen spektral yoğunluk bilgileri EfficientNetB7 transfer öğrenme mimarisi ile sınıflandırılmıştır. Farklı DNA baz uzunluğuna ilişkin spektral yoğunluğun ifade edildiği görsel verilerin eğitim veri kümesinde yer alması sonucunda test veri kümesi üzerinde ulaşılan doğruluk oranı %100'dür. Bununla birlikte sadece belirli aralıkta yer alan DNA baz uzunluğuna ilişkin spektral yoğunluğun ifade edildiği görsel verilerin eğitim veri kümesinde yer alması sonucunda test veri kümesinde ulaşılan doğruluk oranı %50 - %70 arasındadır.

Sitogenetik değerlendirmenin sağlandığı ikinci aşamada, NCBI veri kümesi vasıtasıyla ALL ve KML hastalarından elde edilen BCR-ABL genleri kullanılarak malignitelerin ayırt edilmesi sağlanmıştır. Bu süreç iki ayrı kısımda gerçekleşmiştir. İlk kısımda sayısallaştırılan DNA dizilimlerinin spektral yoğunluk bilgisi kısa zamanlı fourier dönüşümü ve sürekli dalgacık dönüşümü yöntemleri ile spektrogramlara yansıtılmıştır. Ardından zamansal ve uzaysal bilgi çıkarımı sağlayan DGCNN yöntemi ile spektrogramlar sınıflandırılmıştır. DNA baz uzunluğuna bağlı bir sınıflandırmanın gerçekleştiği bu kısımda hem kısa zamanlı fourier dönüşümü hem de sürekli dalgacık dönüşümü yöntemleri için ulaşılan maksimum doğruluk oranı %75 olarak elde edilmiştir. Diğer kısımda DNA baz uzunluğundan bağımsız bir çalışma gerçekleştirilmiştir. Bu doğrultuda Shannon entropi temelli haritalama tekniği kullanılarak sayısallaştırılan DNA dizilimlerinden istatistiksel ve yapısal özellikler çıkarılmıştır. Ardından bu özellikler adaptif bulanık mantık algoritmasına girdi olarak verilmiştir. DNA'nın bulanık konfigürasyon yapısına uygun olduğu düşünülen bu yöntem ile %80 doğruluk oranı elde edilmiştir.

Moleküler değerlendirmenin sağlandığı üçüncü aşamada, bioinformatics laboratory vasıtasıyla temin edilen löseminin temel türleri olan ALL ve AML malignitelerine ait mikrodizi veri kümesi kullanılmıştır. İlk aşamada parçacık sürü optimizasyon algoritması, karınca optimizasyon algoritması ve balina optimizasyon algoritması kullanılarak yüksek boyut sunan mikrodizi veri kümesinden potansiyel genler seçilmiştir. Ardından her bir optimizasyon algoritması için seçilen genlere ilişkin spektral yoğunluk bilgisi sürekli dalgacık dönüşümü yöntemi ile spektrogramlara yansıtılmıştır. Son olarak ALL ve AML kategorilerinin eşleştirildiği spektrogramlar DGCNN yöntemi ile sınıflandırılmıştır. Maksimum başarı oranı karınca optimizasyon

algoritması kullanılarak seçilen potansiyel genlerin DGCNN yöntemi ile sınıflandırılmasının sonucunda %93.33 doğruluk oranı elde edilmiştir.

Moleküler değerlendirmenin sağlandığı dördüncü aşamada bioinformatics laboratory vasıtasıyla temin edilen ALL, AML ve MLL lösemi türlerine ilişkin mikrodizi veri kümesi kullanılmıştır. Yüksek boyut sunan mikrodizi veri kümesinden balina optimizasyon algoritması kullanılarak seçilen genler LSTM sinir ağı mimarisine girdi olarak verilmiştir. Seçilen genlerin birbiri ile ilişkisini hatırlama eylemleri ile dinamik tutma yetisine sahip olan LSTM mimarisinin farklı seed değerleri ile oluşturulan 100 farklı veri kümesindeki ortalama doğruluk oranı %89.883 olarak bulunmuştur.

Patolojik değerlendirmenin sağlandığı beşinci aşamada Acute Lymphoblastic Leukemia Image Database vasıtasıyla temin edilen ALL-IDB1 ve ALL-IDB2 periferik kan yayma görüntüleri kullanılmıştır. Löseminin temel türü olan ALL malignitesinin tanısında önemli bir ayıraç olan blast hücrelerinin tespiti ve sayım bilgisinin üretilmesine ilişkin gerçek zamanlı bir çıktı üreten bu çalışmanın eğitimi YOLOv4 algoritması tarafından gerçekleştirilmiştir. Modele verilen girdi için 3-4 saniye içerisinde çıktı üreten bu sistemin doğruluk oranı %98.87 olarak bulunmuştur.

Moleküler ve immünofenotipik değerlendirmenin sağlandığı altıncı aşamada bioinformatics laboratory vasıtasıyla temin edilen lenfoblast hücre türlerine ilişkin T-ALL, B-ALL ve T-LL malignitelerine ait mikrodizi veri kümeleri kullanılmıştır. İlk olarak yüksek boyut sunan mikrodizi veri kümelerinden balina optimizasyon algoritması kullanılarak seçilen potansiyel genler ANFIS yapısı ile sınıflandırılmıştır. Böylece bulanık çıkarımlar elde edilmiştir. Ardından bulanık mantığın çıkarım gücü ile yapay zekanın veriler üzerindeki öğrenme gücünü birleştiren ANFIS yapısındaki üyelik fonksiyonuna ilişkin parametre optimizasyonu sırasıyla ABC ve PSO optimizasyon algoritmaları ile iyileştirilmiştir. Son olarak ANFIS, ANFIS+ABC ve ANFIS+PSO yapılarından gelen 3 farklı bulanık çıktı, kolektif öğrenme yaklaşımı kapsamında lojistik regresyon algoritması ile sınıflandırılmıştır. Ulaşılan doğruluk oranı %86.6 olarak elde edilmiştir.

Patolojik değerlendirmenin sağlandığı yedinci aşamada Raabin Health veri kümesi vasıtasıyla temin edilen bazofil, eozinofil, lenfosit, monosit ve nötrofil beyaz kan hücrelerine ilişkin dijitalleştirilmiş periferik kan yayma görüntüleri kullanılmıştır. Lösemi hastalığının tanısında önemli bir ayıraç olan lenfosit hücre tipinin tanıma oranını geliştirmek için önerilen YOLOv5x yaklaşımı ve hibrit yapı sunulmuştur. Önerilen YOLOv5x yaklaşımı ile lenfosit tanıma oranı ve genel tespit doğruluğuna ilişkin tanıma oranı için sırasıyla 0.133 ve 0.006 doğruluk oranında bir iyileşme sağlanırken hibrit yapı için %3,44 ile %14,7 doğruluk oranında bir iyileşme sağlanmıştır.

Bu çalışmada farklı alt türlere sahip olan lösemi hastalığının bilgisayar destekli sistemler ile tanı ve tedavi sürecine ilişkin sitogenetik, moleküler, moleküler ve immünofenotipik, patolojik değerlendirmelerin yer aldığı bütünsel bir analiz yapılmıştır. Güncel bir araştırma alanı olan ve halen non-invaziv ya da alternatif yaklaşımların keşfi üzerinde çalışmalar yapılan lösemi kanserine ilişkin sunulan 7 ayrı değerlendirme ile karar destek sistemi inşa edilmiştir. Doktorlara fikir vermek amacıyla tasarlanan bu sistem vasıtasıyla doğru ve erken tanı ile başlayan başarılı tedavilerin gerçekleşmesi hedeflenmektedir.



CLASSIFICATION OF MAIN TYPES OF LEUKEMIA DISEASE WITH ARTIFICIAL INTELLIGENCE-BASED ALGORITHMS ON THE DNA SEQUENCES USING DIGITAL MAPPING TECHNIQUES

SUMMARY

A cell is a unit containing structural and functional properties related to life. It has the ability of controlled divisibility. Thus, injured tissues are repaired or dead cells are renewed. However, genetic or environmental parameters can negatively affect this division process. In such a case, malignancies occurrences can happen. Malignancies are described as malignant tumors. It is formed by the uncontrolled proliferation and spread of cells in a particular tissue or organ.

Cancer is a malignancy that occurs with the uncontrolled proliferation of cells in any part of the body. With this disease, the body's regular working mechanism gradually deteriorates and becomes irremediable if an early diagnosis is not provided. This process is important in terms of determining the type of disease and deciding the treatment methods. In this study, leukemia malignancy, which is among the cancer types, was examined. Because leukemia is the most common type of cancer. It can be seen in all age groups. According to the course of the disease in the person, it is examined in two groups as acute or chronic. Early diagnosis is essential for a long life expectancy, especially in patients with acute leukemia, where the disease appears suddenly and spreads rapidly throughout the body. However, about 40 years ago, a type of leukemia, defined by doctors as mixed lineage leukemia (MLL), was discovered in certain types of patients diagnosed with ALL and AML, which are the main subtypes of leukemia. After the discovery of the MLL leukemia type as a new species, it is thought that the leukemia disease still contains new sub-formations that can be discovered. Clearly distinguishing the subtypes of leukemia cancer, which is considered in a wide framework, will ensure the implementation of the correct treatment protocol together with the realization of the correct diagnosis process. On the other hand, it is also vital in terms of evaluating the response to treatment and clarifying the protocol to be applied for subsequent treatments. Peripheral blood smear, bone marrow aspiration, bone marrow biopsy, immunophenotyping, and various imaging and tests are used to reach definitive results in the medical world. However, these methods sometimes involve uncertain situations. For example, the manual evaluation process of the peripheral blood smear method depends on the doctor's knowledge, experience, physical intensity and mental fatigue. Cancerous tissue may not be present in the part taken by the biopsy method from a person with cancer. The increase in lymphocyte cell type, which is an indicator in the diagnosis of leukemia, may also increase as a result of hepatitis viruses and brucella disease. The success of applying imaging methods depending on various medical parameters or the evaluation of the achieved imaging by doctors may not produce standard output. Therefore,

molecular and cytogenetic evaluation of genetic-based malignancies will clarify the accuracy of the diagnosis. However, recent advances in genome technology have led to an increase in DNA sequences. For this reason, it has become difficult to manually interpret data accurately and quickly in line with cytogenetic analysis. On the other hand, microarray technology, which offers the possibility of molecular evaluation, was evaluated. However, the analysis process of microarray technology, which is a source of hope in cancer research, offers a high dimension because it is provided through features that represent genes. This situation, known as the curse of dimensionality in machine learning, can result in computational burden and noise that produces wrong decisions. In the scope of all these methods, the number of treatment approaches planned to be applied to the person may cause anxiety in the patient within the framework of the cost parameter. At the same time, unnecessary radiation exposure and workload and time loss for doctors are other negative outcomes. In this study, a computer-aided structure was built in 7 separate sections in order to provide an early diagnosis with fewer medical methods. With this structure, which was designed to give an idea to doctors, leukemia disease was analyzed for 7 different purposes.

In the first stage, where cytogenetic evaluation is achieved, BCR-ABL genes, which are an important indicator in the diagnosis of ALL and CML malignancies, which are the main types of leukemia, were analyzed. BCR-ABL genes supplied from the NCBI dataset were digitized using mapping techniques. Then, fourier transform and short-time fourier transform methods were used in the scope of the signal processing approach to detect exon regions containing critical information about life on the digitized array. Spectral density information obtained from DNA sequences with signal processing methods was classified with EfficientNetB7 transfer learning architecture. With the visual data expressing the spectral density of different DNA base lengths included in the training dataset, the accuracy rate reached on the test dataset is 100%. The accuracy rate achieved in the test dataset is between 50% and 70% for visual data expressing the spectral density of DNA base length in a certain range in the training dataset.

In the second stage, where cytogenetic evaluation is achieved, malignancies were differentiated by using BCR-ABL genes obtained from ALL and CML patients via the NCBI dataset. This process took place in two separate parts. In the first part, the spectral density information of the digitized DNA sequences was reflected to the spectrograms with short-time fourier transform and continuous wavelet transform methods. Then, the spectrograms were classified with the DGCNN method, which provides temporal and spatial information extraction. In this part, where a classification based on nucleotide base length takes place, the maximum accuracy rate achieved for both short-time fourier transform and continuous wavelet transform methods was obtained as 75%. In the other part, a study independent of nucleotide base length was performed. In this direction, statistical and structural features were extracted from digitized DNA sequences using Shannon entropy-based mapping technique. Then, these features are given as input to the adaptive fuzzy logic algorithm. With this method, which is thought to be suitable for the fuzzy configuration structure of DNA, a accuracy rate of 80% was achieved.

In the third stage, where molecular evaluation is achieved, the microarray dataset of ALL and AML malignancies, which are the main types of leukemia, obtained by the bioinformatics laboratory, was used. In the first step, potential genes were selected

from the high-dimensional microarray dataset using the particle swarm optimization algorithm, ant optimization algorithm and whale optimization algorithm. Then, the spectral density information of the selected genes for each optimization algorithm was reflected in the spectrograms with the continuous wavelet transform method. Finally, spectrograms matching ALL and AML categories were classified by the DGCNN method. The maximum accuracy rate was obtained as 93.33% as a result of the classification of the potential genes selected using the ant optimization algorithm with the DGCNN method.

In the fourth stage, where molecular evaluation is achieved, a microarray dataset of ALL, AML and MLL leukemia types supplied by the bioinformatics laboratory was used. Genes selected from the high-dimensional microarray dataset using the whale optimization algorithm were given as input to the LSTM neural network architecture. The average accuracy rate in 100 different datasets created with different seed values of the LSTM architecture, which has the ability to keep dynamic with the actions of remembering the relationship of the selected genes with each other, was found to be 89.883%.

In the fifth step, where pathological evaluation is achieved, ALL-IDB1 and ALL-IDB2 peripheral blood smear images obtained from the Acute Lymphoblastic Leukemia Image Database were used. The training of this study, which produces a real-time output for the detection of blast cells and the production of count information, which is an important indicator in the detection of ALL malignancy, which is the main type of leukemia, was carried out by the YOLOv4 algorithm. The accuracy rate of this system, which produces an output within 3-4 seconds for the input given to the model, was found to be 98.87%.

In the sixth stage, where molecular and immunophenotypic evaluation is achieved, microarray datasets of T-ALL, B-ALL and T-LL malignancies related to lymphoblast cell types obtained by the bioinformatics laboratory were used. First, potential genes selected from high-dimensional microarray datasets using the whale optimization algorithm were classified with the ANFIS structure. Thus, fuzzy inferences were obtained. Then, the parameter optimization of the ANFIS structure, which combines the inference power of fuzzy logic and the learning power of artificial intelligence on data, is provided by ABC and PSO optimization algorithms, respectively. Finally, 3 different fuzzy outputs from ANFIS, ANFIS+ABC and ANFIS+PSO structures were classified by logistic regression algorithm within the scope of the ensemble learning approach. The achieved accuracy rate is 86.6%.

In the seventh stage, where pathological evaluation is achieved, digitized peripheral blood smear images (basophil, eosinophil, lymphocyte, monocyte, and neutrophil white blood cells) were supplied through the Raabin Health dataset. The proposed YOLOv5x approach and hybrid structure were presented to improve the recognition rate of lymphocyte cell type, which is an important indicator in the diagnosis of leukemia disease. With the proposed YOLOv5x approach, an improvement of 0.133 and 0.006 was achieved for the recognition rate for lymphocyte recognition rate and overall detection accuracy, respectively, while an improvement of 3.44% and 14.7% was achieved for the hybrid structure.

In this study, a wide analysis including cytogenetic, molecular, molecular and immunophenotypic, pathological evaluations related to the diagnosis and treatment

process of leukemia, which has different subtypes, was performed with computer-aided systems. A decision support system was built with 7 separate evaluations presented on leukemia cancer, which is a current research area and is still being studied for the discovery of non-invasive or alternative approaches. It is aimed to realize successful treatments that start with accurate and early diagnosis through this system, which is designed to give an idea to doctors for leukemia, which reduces the quality of life of the person, causes loss of workforce loss and is life-threatening.



1. GİRİŞ

Hücre, canlılığın yapısal ve işlevsel özelliklerini barındıran temel birimdir. Bu birimin kontrollü bölünebilme yetisi, yaralanan dokuların onarılmasını ya da ölen hücrelerin yenilenebilmesini sağlar. Ancak genetik ya da çevresel etkenler bölünme sürecini etkileyebilir. Böyle bir ihtimalin gerçekleşmesi durumunda malignite oluşumları meydana gelir [1]. Fakat maligniteler tümör ile karıştırılmaması gereken yapılardır. Çünkü tümör, anormal büyüyen hücrelerin kötü huylu (kanseri olan) ya da iyi huylu (kanseri olmayan) olarak nitelendirilmesidir [2]. Öte yandan maligniteler kötü huylu tümörler olarak tanımlanmaktadır.

Kanser, ilgili doku ya da organdaki hücrelerin hızlı ve kontrolsüz bir şekilde çoğalıp yayılması ile ortaya çıkan bir malignitedir[2]. Lösemi, akciğer, kolon, meme ve prostat en sık görülen kanser türleri arasında yer almaktadır [3].

Lösemi kanseri, vücudun kan üretim sistemini (kemik iliği ve lenfatik sistem) etkileyen ve kan hücrelerinin miktarında değişikliğe neden olan bir hastalıktır. Bu nedenle kan kanseri olarak da ifade edilmektedir [1]. Farklı yaş gruplarında görülebilir. Lösemi kanseri hücre tipi ve kişide seyreden ilerleme durumuna göre 2 ana türe ayrılır [4]. Bu çerçevede hastalığın aniden belirdiği ve vücut içerisinde gözlemlenen yayılımın hızlı gerçekleştiği lösemi türleri akut lösemiler olarak isimlendirilir. Öte yandan teşhisin hemen konulamadığı ve vücut içerisindeki yayılımın yavaş seyrettiği lösemi türleri kronik lösemiler olarak isimlendirilir [1]. Miyeloid ve lenfoid şeklindeki nitelendirmeler ise hücre tipini ifade etmektedir. Löseminin temel türleri, Akut Miyeloid Lösemi(AML), Kronik Miyeloid Lösemi(KML), Akut Lenfoblastik Lösemi(ALL) ve Kronik Lenfositik Lösemi(KLL)'dir [5].

Löseminin vücutta oluşum süreci, olgunlaşmamış kötü huylu hücrelerin kemik iliği hücrelerinin yerine geçmesi ve ilik kısmında meydana gelen bir hasarı içerir. Bu durum savunmada önemli olan lökosit miktarı ile pıhtılaşmada önemli olan platelet miktarına zarar verir ve savunma mekanizmasının güçsüzleşmesine neden olur. Bu durumun sonucu olarak hastada enfeksiyon riski artar [1]. Ek olarak kişinin yaşam kalitesi düşer

ve iş gücü kaybı ortaya çıkar. İleriki aşamalarda ölümle sonuçlanabilme ihtimali barındıran lösemi hastalığı için erken teşhis hayati bir konudur.

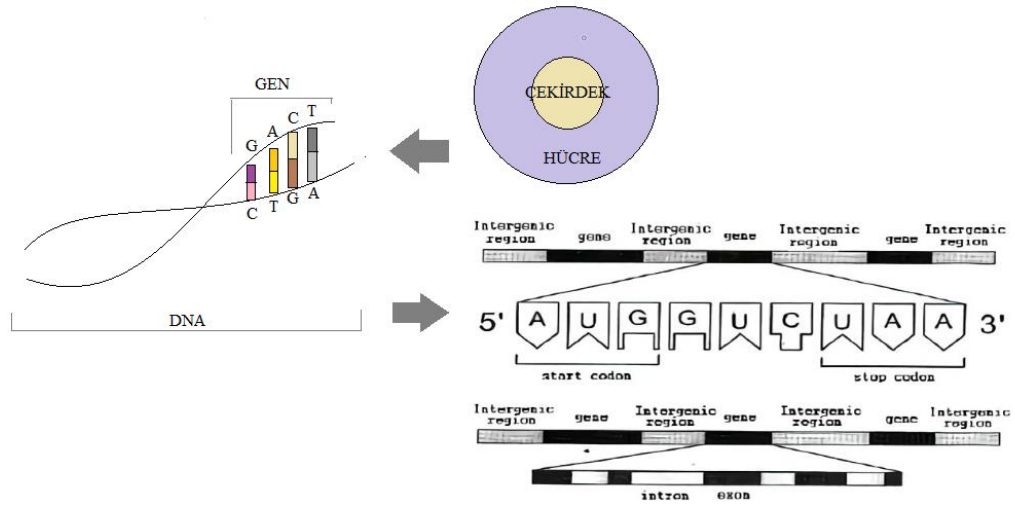
Tıp dünyasında, fiziksel muayene tanı koyma sürecinin ilk aşamasıdır. Yeterli bir çıktının elde edilemediği bu kısımda biyopsi ve kan tahlili işlemleri gerçekleştirilir. Bununla birlikte doktorlar tarafından ihtiyaç duyulan durumlarda farklı görüntülemeler (ultrason ya da CT tarama) ve testler (akış sitometrisi, karaciğer fonksiyonu testi) de tanı sürecinin bir parçasıdır. Öte yandan hücre tiplerinin ve sayısının analiz edilmek suretiyle kansere ilişkin alt türlerin teşhis edilmesinde periferik kan yayma yöntemi kullanılabilir [6]. Hastalığı genetik ya da moleküler temelli olgular üzerinden incelemek tercih edilebilen değerlendirme metodlarından diğerleridir. Bu yöntemler kapsamında kişi üzerinde uygulanması planlanan tedavi yaklaşımlarının artışı, maliyet parametresi çevresinde hasta üzerinde anksiyete oluşturabilir. Aynı zamanda doktorlar için iş yüküne ve zaman kaybına neden olur. Ek olarak yapılan analizlerin değerlendirilmesi, kişinin bilgisi ve uzmanlığı ile ilişkilidir. Bu durum manuel hataların ortaya çıkmasına neden olabilir.

Bu çalışmada, bilgisayar destekli sistemler ile doktorlara fikir vermek ve daha az tıbbi yöntem ile erken teşhisi sağlamak hedeflenmiştir. Bu kapsamda, 7 farklı açıdan lösemi hastalığı analiz edilmiş ve 7 farklı başlık altında ele alınmıştır.

1.1. DNA Genom Dizilimlerindeki Ekson ve İntron Bölgelerinin Sınıflandırılması

Kanser, yaşamı tehdit eden bir hastalıktır. Canlılığın temel birimi olan hücrelerin genetik bileşenlerinde ortaya çıkan değişimler sonucunda meydana gelir [7]. Bu durum hücre çekirdeğindeki DNA ve özel proteinlerin birleşmesinden oluşan kromozomlar ile ifade edilir [8]. Bu birleşmede hayati işlevselliği bulunan DNA genom parçası, genler ve genler arası [9] boşluklardan meydana gelir. Protein sentezinde görevli olan bu gen bölgeleri ökaryot hücreler için protein kodlayan (ekson) ve protein kodlamayan (intron) bölgelerin ayırımına sahiptir [10]. Ekson bölgeleri, protein senteziyle ilgili tüm bilgiyi barındırır ve yaşama ilişkin önemli ayrıntıların değerlendirilmesinde önemli role sahiptir [11]. Öte yandan, ökaryot bir hücre DNA'sının yaklaşık %3'lük bir kısmı ekson bölgelerinden oluşmaktadır [11].

Bahsedilen tıp terimlerinin eşleştirildiği ökaryot hücre DNA'sına ilişkin görsel tanımlamalar Şekil 1.1'de verilmiştir.



Şekil 1.1. Ökaryot hücre DNA'sına ilişkin görsel tanımlamalar [9].

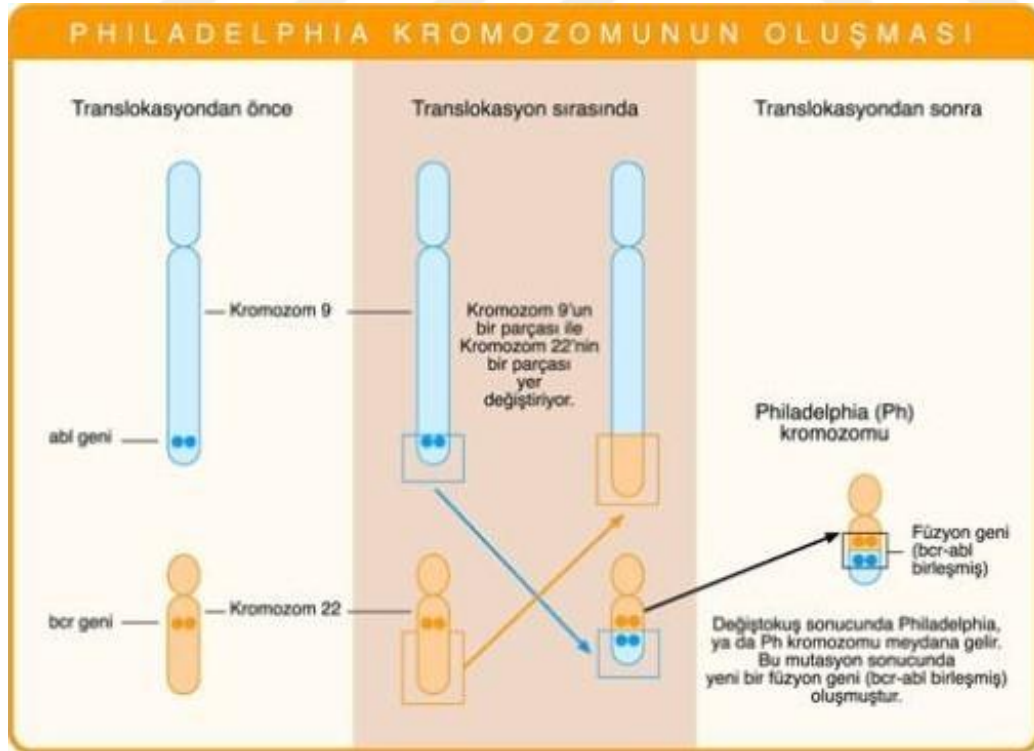
Şekil 1.1'de gösterilen DNA molekülü; şeker, fosfat ve organik bazların bir araya gelmesi ile oluşan büyük bir biyomoleküldür [8, 9]. Sembolik bir yapıya sahip olan bu DNA molekülü; Adenin, Timin, Guanin ve Sitozin organik bazları ile ifade edilmektedir [8]. Sarmal bir yapıda birbirini tamamlayan düzende iki ipliğe sahip olan DNA biriminde, adenin bazı karşı iplikteki timin bazı ile eşleşirken guanin bazı sitozin bazı ile eşleşmektedir [9]. Bu eşleşme sürecinde hidrojen bağları vasıtasıyla bazların birbirlerine tutunması sağlanır. Tutunma eylemi için Adenin ve Timin arasında ikili hidrojen bağı, Guanin ve Sitozin arasında üçlü hidrojen bağı oluşur. Ek olarak azotlu bazlar iki ayrı tip olarak değerlendirilir. Bu doğrultuda, Adenin ve Guanin, pürin bazlarıdır. Sitozin ve Timin, pirimidin bazlarıdır [9]. Aynı zamanda, DNA diziliminin mevcut düzeni içerisinde genetik bilginin saklandığına inanılmaktadır [9].

Bu bölümde genetik bilgilerin yer aldığı genom parçası olan [9] ve DNA dizilimi içerisinde yer alan gen bölgelerindeki ekson ve intron kısımlarının ayırt edilmesi hedeflenmiştir. Yaşam süreçlerinde oluşan veya oluşabilecek birçok olayın aydınlatılmasında önemli bir nokta olan ekson bölgeleri; canlılığın büyüme ve gelişme süreçlerinde yaşanan düzenlemelerin, hücrelerin çoğalma ve sonlanma durumlarının, kök hücrelerin yaşadıkları değişikliklere ilişkin bilgilerin, kanserin gelişmesi ile ilgili değerlendirmelerin ve genlerin mutasyon geçirip geçirmediğine ilişkin verilerin barındırıldığı önemli bir ayrıçtır [8, 12]. Bununla birlikte ekson bölgelerinin ayırt edilmesi araştırmacılar tarafından zorlu bir problem olarak ifade edilmiştir [9, 11, 13, 14].

Birinci bölümde bilgisayarlı tanı sistemlerinin desteği ile lösemi kanserinin temel türlerinden olan ALL ve KML'nin tanısında önemli bir ayıraç olarak nitelendirilen BCR-ABL geni üzerindeki ekson ve intron bölgelerinin sınıflandırılması için alternatif bir yaklaşım önerilmiştir. Bu kısımda sitogenetik analiz yapılmıştır.

1.2. DNA Dizilimleri ile ALL ve KML Malignitelerinin Sınıflandırılması

Vücutta genomik yapının korunması amacıyla DNA'da meydana gelme ihtimali olan problemlerin düzeltilmesinde sorumlu hücresel mekanizmalar görev alır. Ancak bazı durumlarda bu mekanizmalar gerekli onarımları sağlayamaz ve anomaliler ortaya çıkar. Örneğin genlerde meydana gelen mutasyonlardan bazıları kansere neden olmaktadır [1]. Bu doğrultuda löseminin temel türleri olan kronik miyeloid lösemi ve ph-pozitif adı verilen akut lenfoblastik lösemi hastalığının üzerinde etkili olan Philadelphia kromozomu incelenmiştir [15]. PH kromozomu ya da BCR-ABL geni olarak ifade edilen bu anomali, 22 numaralı kromozomun q11 bandındaki BCR geni ve 9 numaralı kromozomun q34 bandındaki ABL genleri arasında gerçekleşen karşılıklı parça değişimi $t(9; 22)(q34; 11)$ olarak tanımlanmaktadır [4, 16]. BCR-ABL genine ilişkin translokasyon sürecinin görsel tasviri Şekil 1.2'de sunulmuştur.



Şekil 1.2. BCR-ABL genine ilişkin translokasyon sürecinin görsel tasviri [1].

Yetişkin KML hastalarının %90'ı, çocuk KML hastalarının %5'i, yetişkin ALL hastalarının yaklaşık %30'u ve çocuk ALL hastalarının yaklaşık %5'i için PH kromozomu, hastalığın tanısının konulmasında önemli bir ayıraçtır [1, 15]. Aynı zamanda bu ayıraç tıp dünyasında da kullanılmaktadır. Ancak maliyetin minimuma indirilmesi, zamanın verimli kullanılması ve hastanın vücut bütünlüğünün bozulmaması amacıyla alternatif pcr tekniklerinin ve non-invaziv yöntemlerinin gelişimi halen devam etmektedir [17-22].

İkinci bölümde bilgisayarlı tanı sistemlerin desteği ile BCR-ABL geni üzerinde lösemi kanserinin temel türlerinden olan ALL ve KML hastalarının sınıflandırılması için alternatif bir yaklaşım önerilmiştir. Bu kısımda sitogenetik analiz yapılmıştır.

1.3. Mikrodizi Veri Kümesi Kullanılarak ALL ve AML Malignitelerinin Sınıflandırılması

Canlı vücudu kendi kendini onaran ve düzenleyen mükemmel bir mekanizmaya sahiptir. Ancak hücrelerin kontrollü bölünebilme özelliklerini yitirmeleri sonucunda ortaya çıkan kanser hastalığı ile düzenli bir şekilde işleyen bu mekanizma yavaş yavaş düzeltilemeyecek bir duruma gelir. Bu süreç hastalığın türünün konulabilmesi ve tedavi yöntemlerinin belirlenebilmesi açısından önemli bir aşamadır. Bu aşamada genlerin ekspresyon seviyeleri ile başarılı bir sınıflandırma gerçekleştirilebilir [23]. Bu nedenle mikrodizi teknolojisi, analiz sürecinde tercih edilen bir metottur. Böylece, genlerin fonksiyonel protein yapılarına dönüşüm süreci değerlendirilir [24].

Organizmanın genetik kodunun alt kümesi olarak ifade edilen gen ekspresyonu [23] farklı birçok genin nispi ekspresyon seviyesi üzerinden moleküler tanı imkanı sağlayarak belirli hastalıkların sınıflandırmasında öncü olan bir yaklaşımdır [25]. Analiz işlemi genleri temsil eden özellikler üzerinden sağlanır. Bu nedenle yüksek boyutluluk mevcuttur [26]. Gen ekspresyon seviyelerindeki bu boyutsal üstünlük, makine öğrenmesinde boyutluluğun laneti olarak tanımlanmaktadır [27]. Bu durum sınıflandırma aşamasında ekstra bir hesaplama yükü oluşturmaktadır [27] ve hastalıkla ilişkisiz genlerin hastalıkla ilişkili genler üzerinde gürültü oluşturmasıyla meydana gelebilecek yanlış yönlendirmelere neden olmaktadır [28, 29]. Bu problemin üstesinden gelebilmek ve hastalıkla ilişkili bilgili genleri veri kümesinden titizlikle

seçebilmek için özellik seçim metotları kullanılmıştır [25]. Ardından moleküler sınıflandırma sağlanmıştır

Üçüncü bölümde bilgisayarlı tanı sistemlerin desteği ile lösemi kanserinin temel türlerinden olan ALL ve AML hastalarının mikrodizi teknolojisi kullanılarak sınıflandırılması için alternatif bir yaklaşım önerilmiştir. Bu kısımda moleküler analiz yapılmıştır.

1.4. Mikrodizi Veri Kümesi Kullanılarak ALL, AML ve MLL Malignitelerinin Sınıflandırılması

Lösemi hastalığına ilişkin spesifik genetik değişiklikler, doğru bir tanının konulabilmesi için ivedilikle teşhis edilmesi gereken anormalliklerdir. Çünkü lösemi, alternatif durumların halen keşfedildiği bir kanserdir [30]. Bu doğrultuda yaklaşık 40 yıl önce doktorlar tarafından löseminin temel alt türlerinden olan ALL ve AML malignitesinin konulduğu belli hasta tiplerinde; hastalık sürecinin daha şiddetli geçtiği gözlemlenmiş ve ilerleyen zamanlarda bu hastalık karışık soy lösemisi (MLL) olarak tanımlanmıştır [31]. Baskın olarak pediatrik hastalarda görülen bu malignite, 11q23 kromozomunda mevcut olan ve MLL genini etkileyen agresif bir lösemi türü olarak tanıtılmıştır [31]. Aynı zamanda bu hastalık kötü bir prognoza sahiptir. Bu durum erken teşhisin gerekliliğini ön plana çıkarmaktadır. Tüm lösemi vakalarının yaklaşık %10'u [32], bebek ALL vakalarının yaklaşık %80'i, pediatrik AML vakalarının yaklaşık %35-50'si, yetişkin ALL vakalarının yaklaşık %5'i ve yetişkin AML vakalarının yaklaşık %5-%10'u 11q23 kromozomunda görülen anomalilerden kaynaklanmaktadır [33]. 11q23 kromozomuna ait tüm bu oranların ortak paydası ise hastaya uygulanacak yöntemlerin ve hastalığa ilişkin seyrin belirsizlik içermesidir [33]. Örneğin, bebek MLL hastası için uygulanacak tedavi ALL ya da AML şeklinde belirlenen ilk sınıflandırma sonrasında öngörülmektedir. Ek olarak, vakaların ALL ya da AML türlerine olan yatkınlığı da tedavi sürecini etkileyen farklı bir parametredir [33]. Hem ALL hem de AML alt tiplerinde gözlemlenen bu malignitenin farklı birçok belirsizlik içermesinden dolayı dünya sağlık örgütü (WHO) 11q23 sapmalarını farklı bir başlık altında değerlendirmiştir [31].

Bu malignitelerin analiz edilmesinde nispeten yeni bir teknoloji olarak kullanılan ve kanser araştırmalarında olağanüstü bir etkisi olan mikrodizi veri kümesi [34, 35] tercih

edilmiştir. Ancak mikrodizi veri kümesi karmaşık ve maliyetli bir yapıya sahiptir [35]. Öte yandan az örnek sayısı ile birlikte yüksek boyut sunması nedeni ile sınıflandırma doğruluğunu olumsuz yönde etkilemektedir [34]. Bu problemin üstesinden gelebilmek ve hastalıkla ilişkili bilgili genleri veri kümesinden titizlikle seçebilmek için özellik seçim metotları kullanılmıştır [25]. Ardından moleküler sınıflandırma sağlanmıştır.

Dördüncü bölümde bilgisayarlı tanı sistemlerinin desteği ile lösemi kanserinin türlerinden olan ALL, AML ve MLL'nin mikrodizi teknolojisi kullanılarak sınıflandırılması için alternatif bir yaklaşım önerilmiştir. Bu kısımda moleküler analiz yapılmıştır.

1.5. Dijitalleştirilmiş Periferik Kan Yayması Görüntüleri Kullanılarak ALL Malignitesinin Tespiti

Kan oksijen, karbondioksit ve minerallerin damarlar vasıtasıyla dolaşımını mümkün kılan kırmızı renkli yaşam sıvısıdır [36]. Bu kan sıvısının içerisinde yer alan çeşitli unsurların miktarındaki niteliksel ve niceliksel değişiklikler vücutta oluşum gösteren çeşitli hastalıklar sonucunda meydana gelmektedir [36, 37]. Özellikle lösemi hastalığının temel türlerinden olan ALL, kanda meydana getirdiği değişiklikler ile verilebilecek başarılı bir örnektir.

ALL olgunlaşmamış, kötü huylu beyaz kan hücrelerinin (blast hücreleri) kontrolsüz ve aşırı bir şekilde çoğalmasına neden olan bir malignitedir [38]. Kan dolaşımına ve diğer hayati organlara hızlı bir şekilde yayılır. Canlılığın sürdürülebilmesi için erken teşhis temel hedeftir [38]. Bununla birlikte hastalığın semptomları halsizlik, anemi, nefes darlığı, eklem ya da kemik ağrıları gibi genel belirtilerdir [38]. Spesifik bir belirti göstermediğinden dolayı doğru ve erken teşhisin sağlanması güçtür.

ALL malignitesinin teşhisinde önemli bir biyobelirteç olan blast hücreleri, kemik iliğinde üretilir. Ancak zaman içerisinde blast hücrelerinin miktarında artış görülür. Böylece kemik iliğine sığmayan blast hücreleri kanda da belirmeye başlar. Kan sıvısının değerlendirilip mevcut çıktılar üzerinden yorum yapılması için tıp dünyasında periferik kan yayma yöntemi tercih edilir [38]. Ancak mikroskop aracı kullanılarak bir cam üzerine yayılan kan sıvısının boyanması ve ardından farklı büyütme oranları ile elde edilen görüntüler üzerinde manuel değerlendirmelerin gerçekleştirilmesi yavaş ve zaman alıcıdır. Çünkü bu süreçte yapılan morfolojik analiz

değerlendirmesi, ilgili uzmanın ya da operatörün bilgisine ve tecrübesine göre değişir. Aynı zamanda kişinin yorgunluk ve meşguliyeti de nihai çıktıyı etkileyen faktörlerdir [38]. Ek olarak mikroskop vasıtasıyla temin edilen görüntülerdeki blast hücrelerinin arka plan kontrastına ilişkin bir eksiklik dahi hücrelerin morfolojik analizi üzerinde temellenen manuel değerlendirmeleri zorlaştırmaktadır [36]. Hücrelerin tanınması ile sağlanan ALL teşhisinin yanı sıra tanınan hücrelerin kandaki yoğunluğunun belirlenmesi de tedaviye verilen yanıtın değerlendirilebilmesi açısından hayatidir. Bu doğrultuda hücrelerin sayılmasına yönelik ürünler vardır. Ancak bu ürünlerin analiz esnasında kan örneklerini yok ettikleri belirtilmektedir [37]. Öte yandan akış sitometrisi adı verilen bir teknoloji ile blast hücrelerine ilişkin sayısal bir analiz yapmak mümkündür. Ancak bu teknoloji pahalıdır. Bu nedenle ücretsiz ve gerçek zamanlı bir yapı sunan sistem gereklidir. Böylece periferik kan yayma görüntülerindeki blast hücreleri üzerinde gerçekleştirilen tespit süreci ile ALL hastalığına ilişkin tanının konulabilmesi ve tespit edilen blast hücrelerinin toplam sayısının ifade edilebilmesi ile tedaviye verilen yanıt doktorlar tarafından değerlendirilecektir.

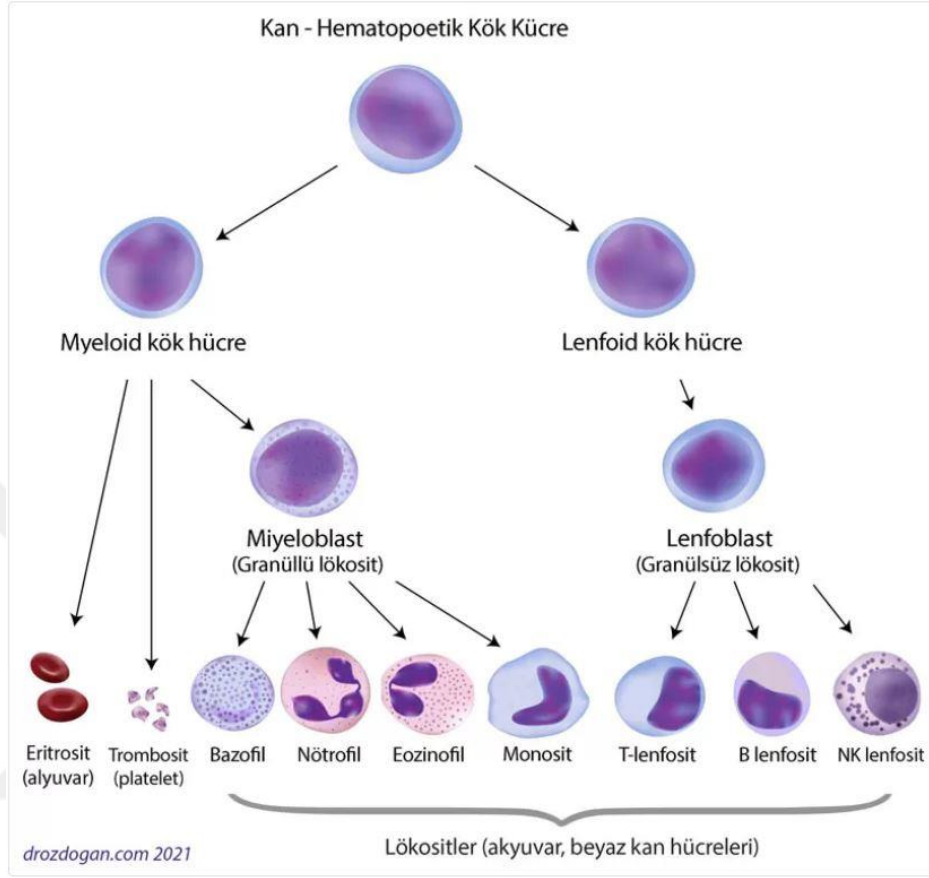
Beşinci bölümde bilgisayarlı tanı sistemlerinin desteği ile lösemi kanserinin temel türlerinden olan ALL'nin periferik kan yayma yöntemi ile sınıflandırılması için bir yaklaşım önerilmiştir. Bu kısımda patolojik analiz yapılmıştır.

1.6. Mikrodizi Veri Kümesi Kullanılarak T-ALL, B-ALL ve T-LL Malignitelerinin Sınıflandırılması

Lösemi, farklı karakteristik bulgular gösteren bir kanser türüdür [39]. Hücre kökenine göre miyeloid veya lenfoid olarak sınıflandırılmaktadır. Ayrıca bu sınıflandırma, hastalığın vücut içerisinde ilerlemesine göre akut veya kronik şeklinde farklılaşmaktadır. Bu özellikler, 4 ana lösemi türünün varlığına işaret etmektedir. Akut Myeloid Lösemi(AML), Kronik Miyeloid Lösemi(KML), Akut Lenfoblastik Lösemi(ALL) ve Kronik Lenfositik Lösemi(KLL) temel lösemi türleridir [5].

Lösemi, tüm yaş gruplarında görülebilen bir malignitedir. Bu malignite çocukluk çağında görülen kanserlerin %25-30'unu oluşturmaktadır. Bununla birlikte bu ölçekte değerlendirilen kanserlerin %97'si akut lösemilerden meydana gelmektedir [39]. Akut lösemiler, miyeloid ve lenfoid hücrelerin farklılaşma sürecinde ortaya çıkan bir

bozukluktur. Bu bozukluk, kemik iliğinde kontrolsüz çoğalan ve ardından kana ve dokulara geçen blast hücrelerin varlığı ile ifade edilmektedir [39]. Kök hücrenin farklılaşma aşamalarını içeren tasvir Şekil 1.3'te verilmiştir.



Şekil 1.3. Hematopoetik kök hücrenin farklılaşma aşamaları [40].

Lökosit olarak nitelendirilen akyuvar ve beyaz kan hücrelerinin kemik iliğinde oluşması ve farklılaşması ile lösemi hastalığı oluşmaktadır. Öte yandan kemik iliğinde oluşan lökositlerin olgunlaşmaları için lenfoid organlara taşınması ile lenfoma hastalığı meydana gelmektedir [40].

Blast hücrelerin alt tiplerinin sınıflandırılması safhasında immünofenotipik, morfolojik ve biyokimyasal özelliklerin değerlendirilmesi önemlidir. Özellikle immünofenotipik değerlendirme, ALL tanısının teyit edilmesi ve T/B ya da non T/B hücre sınıfının belirlenebilmesi açısından gerçekleştirilmektedir [39]. Prekürsör B akut lenfoblastik lösemi (B-ALL), prekürsör T akut lenfoblastik lösemi (T-ALL), T lenfoblastik lenfoma(T-LL); B ve T hücre serisinden meydana gelen ve canlılığın mekanizması dışında davranan oluşumlardır [41]. Dünya Sağlık Örgütü (WHO)'nın 2001

sınıflandırmasında, immünolojik, sitogenetik, patolojik, klinik ve moleküler özelliklerinden dolayı ALL ve LBL maligniteleri ortak terminoloji altında değerlendirilmiştir [41]. Ancak bu yapıları birbirinden ayıran farklı özellikler de bulunmaktadır. Örneğin teşhis ve tedavi sürecinde faydalanan kemik iliği tutulumu ve periferik kan incelemeleri ALL vakalarının değerlendirilmesinde kullanılırken kitlesel doku bozukluklarının analiz edilmesi LBL vakalarının değerlendirilmesinde kullanılmaktadır [41, 42]. Aynı zamanda T-ALL ve B-ALL vakalarının genomik ve moleküler anormallikleri farklıdır. Dolayısıyla tedavi protokollerine farklı yanıtlar verirler [42, 43]. Ek olarak T-ALL ve T-LL vakaları için kök hücre nakli, mediastinal ışınlama ve kemoterapi yöntemleri de farklı şekilde uygulanır [44]. Bu nedenle alt türlerin net bir şekilde ayırt edilmesi, doğru tanı sürecinin gerçekleşmesi ve uygulanacak tedavi protokolünün belirlenmesi açısından hayati bir kriterdir. Periferik yayma, kemik iliği aspirasyonu, kemik iliği biyopsisi, immünofenotipleme ve sitogenetik incelemeler; lösemi hastalığı kapsamında kesin sonuçlara ulaşabilmek amacıyla faydalanan yöntemlerdir.

Altıncı bölümde bilgisayarlı tanı sistemlerinin desteği ile ortak ve farklı özelliklere sahip olan T-ALL, B-ALL ve T-LL vakalarının mikrodizi teknolojisi kullanılarak sınıflandırılması için alternatif bir yaklaşım önerilmiştir. Bu kısımda moleküler ve immünofenotipik analiz yapılmıştır.

1.7. Beyaz Kan Hücrelerinin Tespiti ve Sınıflandırılması

Vücudun metabolizma hızının kontrol edilmesinde rol oynayan kan oksijen, karbondioksit veya minerallerin taşınmasına yardımcı olan yaşam sıvısıdır [36, 45]. Pıhtılaşma ya da bağışıklık gibi vücudun temel fonksiyonlarından sorumludur [45]. 4 temel bileşenden meydana gelir ve sağlıklı bir insana ait vücut ağırlığının yaklaşık %7'sini oluşturur[46].

Kan sıvısının 4 ana bileşeni; plazma, trombositler, beyaz kan hücreleri ve kırmızı kan hücreleridir [45, 46]. Minimum maliyet ile yaşamsal fonksiyonların değerlendirilmesi için hücre tiplerine ilişkin kan testleri, tıp dünyasında yaygın olarak kullanılmaktadır. Periferik kan yayma, mevcut kan testlerinden biridir [46]. Bu hematoloji testinde hastadan alınan kan, cam malzemesi üzerine yayılır. Boyalar vasıtasıyla kan boyanır. Ardından cam lamı, mikroskop altına konulur ve farklı büyütme oranları ile morfolojik

analizi gerçekleştirilir. Doktorlar ya da operatörler tarafından gerçekleştirilen bu değerlendirme sürecinde, çeşitli hücre tipleri boyaya farklı tepkiler verir. Bu durum hücre tiplerinin ayırt edilmesini kolaylaştırır [46]. Bununla birlikte hekimin bilgisi ve tecrübesi manuel değerlendirme aşamasının başarısını etkiler [38]. Dolayısıyla manuel değerlendirme ile periferik yaymanın okunması kişiden kişiye değişir. Ek olarak zaman alıcı bir işlemdir. Aynı zamanda kan elementlerinin incelendiği görüntülerin kontrast durumu, bu elementlerin operatörler ya da doktorlar tarafından segmente edilmesini de zorlaştırır [36, 38]. Öte yandan kan sıvısında, kırmızı kan hücreleri beyaz kan hücrelerinin 600 katıdır [45]. Bu durum yaşam sıvısının manuel analiz edilmesini olumsuz etkileyen diğer bir konudur. Fakat yüksek doğruluk oranı ile kanın değerlendirilmesi önemlidir. Özellikle hayati yapılar olan ve vücudun korunmasında rol oynayan beyaz kan hücrelerinin analizi birçok hastalığın keşfedilmesinde yardımcı olmaktadır.

Beyaz kan hücreleri 5 alt türe sahiptir. Monosit, Eozinofil, Bazofil, Lenfosit ve Nötrofil olarak isimlendirilen bu elementler; çekirdek, sitoplazma ve hücre duvarından meydana gelir [45]. Beyaz kan hücrelerinin boyama maddelerine verdikleri tepkiler farklı olduğu için periferik kan yayma yöntemi ile mikroskop altında hücre tiplerinin ayırt edilmesi mümkündür [46]. Böylece elde edilen analiz çıktılarına göre tedavi planı oluşturulacaktır.

Beyaz kan hücreleri, kemik iliği ve lenfoid dokularda üretilen yapılardır. Beyaz kan hücrelerine ilişkin alt türlerin kandaki yoğunluğu farklı hastalıklara işaret etmektedir. Örneğin; hemoliz, metabolik bozukluklar ya da kanama durumları için nötrofil yapıları kan sıvısında normal değerin üzerinde görülür. Bakteriyel ve viral enfeksiyonlar, kandaki monosit oranında bir artışa sebep olur. Alerjik ya da atopik rahatsızlıklar, kanda bulunan eozinofil oranındaki yükselişe işaret eder. Malign miyeloproliferatif ya da hemolitik anemi, bazofil hücre tipinin sayısını etkiler. Kandaki lenfosit oranı ise su çiçeği ya da tüberküloz da azalırken hepatit virüsler, brusella ya da lösemi de artar [47].

Yedinci bölümde bilgisayarlı tanı sistemlerinin desteği ile lösemi hastalığına ilişkin değerlendirmelerin yapılabilmesi amacıyla Monosit, Eozinofil, Bazofil, Lenfosit ve Nötrofil hücrelerinin periferik kan yayma yöntemi ile sınıflandırılması için bir yaklaşım önerilmiştir. Bu kısımda patolojik analiz yapılmıştır.

Sunulan yedi alt bölümde, lösemi hastalığının sitogenetik, moleküler, patolojik, moleküler ve immünojenotipik değerlendirmeleri için bilgisayar destekli analizler yapılmıştır. Bu analizler ile doktorlara yardımcı fikirler üretilerek tıp dünyasında kullanılan yöntem miktarını azaltmak temel amaçtır. Böylece maliyet parametresi çerçevesinde hem hastalar hem de doktorlar için zaman kaybı azalacak, uzun ve yorucu süreç nedeniyle hastalarda oluşan anksiyete çıktılarına bir çözüm sağlanacaktır.



2. KAYNAK ARAŞTIRMASI

Son yıllarda teknolojinin büyümesi ile hastalıklara ilişkin tanı modelleri geliştirilmeye başlanmıştır. Özellikle tıp ve bilişim dünyasının hibrit çalışmaları aktif bir süreç içerisinde. Bu süreçte bilgisayar destekli sistemler vasıtasıyla yapılan analizlerde geçmiş veya güncel verilerden faydalanılmaktadır. Böylece birçok model ve algoritma geliştirilmektedir.

Bu kısımda lösemi hastalığı kapsamında ele alınan 7 farklı konu üzerinde önerilen analizler incelenmiş ve değerlendirmeler yapılmıştır.

2.1. DNA Genom Dizilimlerindeki Ekson ve İntron Bölgelerinin Sınıflandırılmasına İlişkin Çalışmalar

[8] çalışmasında, vücutta gerçekleşen ya da gerçekleşecek hayati olayların değerlendirilmesinde önem taşıyan ekson ve intron bölgeleri analiz edilmiştir. İlk olarak MEFV genine ait sembolik DNA dizilimleri 9 farklı haritalama tekniği ile sayısallaştırılmıştır. Ardından fourier dönüşümü, destek vektör makineleri ve k en yakın komşu metotları ile sınıflandırma sağlanmıştır. Maksimum başarı integer haritalama tekniği ile sayısallaştırılan DNA dizilimlerinin Ayırık Fourier Dönüşümü yöntemi ile sınıflandırılmasının sonucunda %96.2 olarak elde edilmiştir.

[10] çalışmasında, protein kodlayan bölgelerin tanımlanması için nükleotit düzeyinde bir çalışma gerçekleştirilmiştir. İlk olarak 10 farklı haritalama tekniği ile veriler sayısallaştırılmıştır. Ardından kısa zamanlı ayırık fourier dönüşümü (ST-DFT) yönteminin rectangular, bartlett, hann, hamming ve blackman pencereleme fonksiyonları ile AUC değerleri üzerinden değerlendirilmiştir. Maksimum AUC değeri hem Voss hem de Tetrahedron haritalama tekniği ile sayısallaştırılan DNA dizilimlerinin Hamming pencereleme yaklaşımını kullanan ST-DFT yöntemiyle analiz edilmesinin sonucunda %83.21 olarak elde edilmiştir.

[48] çalışmasında, sembolik bir yapıya sahip olan DNA dizilimlerinin sayısallaştırılması için yeni bir yaklaşım önerilmiştir. Geleneksel haritalama

tekniklerinin yanı sıra önerilen bu yeni yaklaşım üzerinde tekil değer ayrışımı, ayırık fourier dönüşümü (DFT) ve kısa zamanlı fourier dönüşümü yöntemi (STFT) ile sınıflandırma sağlanmıştır. DFT ve STFT yöntemleri için ulaşılan performanslar sırasıyla %96.116 ve %92.037 olarak elde edilmiştir.

[14] çalışmasında, zorlayıcı bir görev olarak nitelendirilen ekson ve intron bölgelerinin tanımlanabilmesi için ilk aşamada nükleotit dizilimleri üzerinden istatistiksel bilginin çıkarılabilmesi hedeflenmiş ve wavelet temelli zaman serisi yaklaşımı önerilmiştir. Ardından varyans bilgisi ile özellik vektörü oluşturulmuştur ve optimize edilmiş destek vektör makineleri algoritması ile sınıflandırılmıştır. Homo sapiens türü için eğitim sırasında görülmeyen ekson ve intron veri kümeleri üzerinde %88.95 doğruluk oranına ulaşıldığı belirtilmiştir.

[49] çalışmasında, kodlama bölgelerinin tahmini üzerinde sayısal sinyal işleme tabanlı uyarlamalı bir yaklaşım önerilmiştir. Önerilen tekniğin performansının SCM, OSCM ve MGWT tekniklerine kıyasla daha başarılı olduğu belirtilmiştir. Önerilen yaklaşımın ASP67, HMR195 ve ALLSEQ veri kümeleri üzerinde ulaştığı doğruluk oranları sırasıyla %83, %79 ve %77 olarak bulunmuştur.

[50] çalışmasında, ekson ve intron bölgelerinin net bir şekilde tanımlanabilmesi amacıyla topolojik entropi hesaplaması, genomik sinyal işleme ve tekil değer ayrışımı yaklaşımlarının kullanıldığı bir analiz önerilmiştir. Ardından ekson ve intron bölgelerindeki karmaşıklığın ölçümü ile sınıflandırma sağlanmıştır. AJ229040 geninde ekson bölgelerinin tespiti için ROC eğrisi altında kalan alan 0.69 olarak bulunmuştur.

[51] çalışmasında, ekson bölgelerinin tahmini için EIIP haritalama tekniği ile sayısallaştırılan DNA dizilimleri, parametrik olmayan spektral tahmin teknikleri ile grafikler üzerinden değerlendirilmiştir.

[52] çalışmasında, DNA sinyallerinin değerlendirilmesinde önemli olan periyodiklik kavramı irdelenmiştir. Bu çerçevede STDFT ve ANF gibi sayısal sinyal işleme tabanlı yöntemler kullanılarak spektral alandaki üç temel periyodiklik özelliği üzerinde deneyler yapılmıştır ve grafikler üzerinden değerlendirilmiştir. Gerçekleştirilen analizler sonucunda periyot-3 özelliğinin uzun dizilime sahip eksonlar için daha başarılı olduğu ifade edilmiştir.

[53] çalışmasında, ekson ve intron bölgelerinin tanımlanması amacıyla güçlü ve zayıf hidrojen bağlarının varlığından faydalanarak numerik değerlerin ataması yapılmıştır. Ardından parametrik spektral tahmin yöntemleri uygulanmıştır. Ortalama standart sapma, varyans ve normalize edilmiş varyans istatistiksel değerlerinin hesaplanılmasının sonucunda ekson ve intron bölgelerinin spektral yoğunluğuna ilişkin ayırt edici bir özellik gösterdiği belirtilmiştir.

[13] çalışmasında, ekson bölgelerinin tahmini için trigonometrik haritalama ve uyarlanabilir kaiser pencere yöntemlerinin birlikte kullanıldığı hibrit bir yapı oluşturulmuştur. Önerilen yaklaşımın ulaştığı maksimum AUC değeri E3 ekson bölgesi için 0.923 olarak elde edilmiştir.

[11] çalışmasında, hem kısa hem de uzun dizilime sahip ekson bölgelerinin tanımlanabilmesi için bulanık kural üzerinde temellen gabor dalgacık dönüşümü tekniğinin ölçekleme faktörü uyarlanmıştır. Önerilen yaklaşımın *C. elegans* için ürettiği maksimum başarı oranı 0.943 olarak E5 ekson bölgesi kapsamında elde edilmiştir.

2.2. DNA Dizilimleri Kullanılarak Anomalilerin Sınıflandırılmasına İlişkin Çalışmalar

[54] çalışmasında, kanserli ve kanserli olmayan gen dizilimlerinin sınıflandırılması hedeflenmiştir. İlk olarak sayısallaştırılan gen dizilimlerine, ayırık dalgacık dönüşümü metodu uygulanmış ve ortaya çıkan dalgacık alanının ortalama, medyan, standart sapma, çeyrekler arası aralık, basıklık ve çarpıklık gibi istatistiksel değerleri elde edilmiştir. Ardından bu istatistiksel özellikler, destek vektör makineleri algoritması ile sınıflandırılmıştır ve %100 doğruluk oranı elde edilmiştir.

[55] çalışmasında, prostat, mide, kolon, meme ve sağlıklı genler EIIP haritalama tekniği ile sayısallaştırılmıştır. Ardından ayırık fourier dönüşümü ve bayesian fusion tekniği ile analiz edilmiştir ve grafikler üzerinden değerlendirilmiştir. Bu çalışmanın sağlıklı ve kanserli genlerin ayırt edilmesinde başarılı olduğu ifade edilmiştir.

[56] çalışmasında, kanser hastalıklarının sınıflandırılması için ilk olarak EIIP haritalama tekniği kullanılmış ve dizilimler sayısallaştırılmıştır. Ardından ayırık dalgacık yöntemi uygulanarak sayısallaştırılan bu dizilimlerin ekson bölgelerindeki anormallikler üzerinden kanser tanımlamaları grafikler ile değerlendirilmiştir.

[57] çalışmasında, prostat verileri üzerinden kanser ile ilişkili olabilecek genetik dizilim varyantlarının tespiti ile ilgili bir çalışma yapılmıştır. Bu doğrultuda, kanserli ve normal örneklerin dizilimlerine ilişkin varyantların istatistiksel özellikleri üzerinden karşılaştırma sağlanarak anlamlı bir istatistiksel akış elde edilmiştir.

[7] çalışmasında, kanser hastalığının oluşumunda ve yayılımında rol oynayan DNA dizilimleri ele alınmıştır. Öncelikle dizilimler üzerinden gerekli analizlerin gerçekleştirilebilmesi için z eğrisi yöntemi ile dizilimler sayısallaştırılmıştır. Ardından ayrık fourier dönüşümü ile anti-notch filtresinin birlikte kullanıldığı bir yaklaşım uygulanmıştır. Elde edilen çıktılar üzerinden yapılan sınıflandırma işlemi sonucunda ulaşılan maksimum başarı oranı 0.98 olarak SVM-RBF ile elde edilmiştir.

[58] çalışmasında, kansere neden olabilen genetik anormalliklerin belirlenmesi hedeflenmiştir. Bu nedenle sembolik bir yapıya sahip olan normal DNA dizilimleri ve mutasyona uğramış DNA dizilimleri kullanılarak mutasyon noktalarının tespiti için dalgacık dönüşümü yöntemi ile her bir DNA diziliminden elde edilen spektrogramlarda; ortalama genlik, standart sapma ve ortalama genlik/standart sapma değerleri ile mutasyon noktaları saptanmıştır.

[59] çalışmasında, normal prostat hücrelerinden kanserli prostat hücrelerinin ayırt edilmesi için EIIP haritalama tekniği ile sayısallaştırılan aminoasit dizilimleri ayrık fourier dönüşümüne girdi olarak verilmiştir. Böylece her bir DNA dizilimi için spektral tahmin elde edilmiştir ve spektrumlardaki genlik değerleri hesaplanmıştır. Ardından bu genlik değerlerinin de kullanıldığı PCA yöntemi ile korelasyon katsayıları hesaplanmıştır. 8 kanserli ve 8 prostat hücresi için %0.8 bir hata değerinin üretildiği belirtilmiştir.

[60] çalışmasında karaciğer kanseri ile ilişkilendirilen hepatit b virüsünün DNA dizilimleri amino asit bileşimine dönüştürülmüştür. Ardından KNN, DT ve SVM algoritmalarının kolektif kullanılması ile karaciğer kanserine ilişkin analiz yapılmıştır ve %88.4 doğruluk oranı elde edilmiştir.

[61] çalışmasında, NCBI veri kümesinden elde edilen genlere ilişkin tanıma süreçlerinin gerçekleştirilebilmesi için entropi tahmini, karşılıklı bilgi hesaplama ve sayısal sinyal işleme yöntemleri kullanılmıştır. Ardından minimum entropi değerleri

ile kanser gen tiplerine ilişkin grafikler üzerinden analiz (prostat, göğüs ve kolon kanseri) sağlanmıştır.

2.3. Mikrodizi Veri Kümesi Kullanılarak Anomalilerin Sınıflandırılmasına İlişkin Çalışmalar

[24] çalışmasında, mikrodizi veri kümeleri üzerinde 4 farklı kanser türü için gerçekleştirilen tanı sürecinde özellik seçim yaklaşımı ile destek vektör makineleri algoritmasının birlikte kullanıldığı aktif bir öğrenme modeli geliştirilmiştir. Önerilen yaklaşımın Prostate, DLBCL, Leukaemia, ALLGSE412 veri kümeleri için ürettiği maksimum başarı oranı sırasıyla %93.86, %93.78, %93.56 ve %85.89 olarak elde edilmiştir.

[27] çalışmasında, kanser türleri arasında doğru tanının sağlanabilmesinde umut kaynağı olan mikrodizi veri kümesi kullanılmıştır. İlk olarak parçacık sürü eniyileme yöntemi ile potansiyel genler seçilmiştir. Ardından bu genler uyarlanabilir rezonans teorisine dayanan ve sınıflandırma için uygun olduğu ifade edilen ssEAM sinir ağı mimarisi ile sınıflandırılmıştır. Acute leukemia ve ALL veri kümesi için ssEAM sinir ağı ile elde edilen maksimum başarı oranı sırasıyla %97.50 ve %98.29 olarak elde edilmiştir.

[62] çalışmasında, yüksek boyutlu veri kümelerinde kanser araştırmaları için hastalıkla ilişkili gen belirteçlerinin belirlenmesinin kritik öneme sahip olduğu belirtilmiş ve ilk olarak RFGS ve SVST yöntemleri ile veri kümelerinden ilişkili genler seçilmiştir. Ardından tanı koyma performansı SNR, t-Test, LSD, TNoM, MDMR, WEPO, RFGS, SVST metotları ile değerlendirilmiş ve grafikler üzerinden karşılaştırma sağlanmıştır.

[63] çalışmasında, mikrodizi veri kümelerinde kansere ait ayraçların doğru bir şekilde tanımlanması için çekirdekleştirilmiş bulanık kaba küme yöntemi kullanılmıştır. Ardından yarı denetimli destek vektör makineleri ile sınıflandırılmıştır. Yöntemlerin 12 ayrı kombinasyonu ile elde edilen çıktılar grafikler ile değerlendirilmiştir.

[64] çalışmasında, mikrodizi veri kümesi üzerinde MOCSVMEN isminde yeni bir destek vektör makinesi tabanlı küme topluluğu yöntemi geliştirilmiştir. Bu yaklaşımın genlerin kümelenmesi için başarılı bir performans gösterdiği belirtilmiştir.

[65] çalışmasında, yüksek boyut sunan mikrodizi veri kümeleri üzerinde yeni bir topluluk yöntemi önerilmiştir. Mikrodizi veri kümeleri üzerinde korelasyon analizi teknikleri kullanılarak seçilen gen özellikleri, sınıflandırıcı komitenin oluşturulması için önerilen topluluk yöntemine girdi olarak verilmiştir. Lösemi, kolon, akciğer kanseri ve yumurtalık veri kümelerinde test doğruluk oranları sırasıyla %98.6, %95.2, %100, %99.6 olarak elde edilmiştir.

[66] çalışmasında, mikrodizi veri kümeleri kullanılarak kanser türleri arasında belirleyici gen dizilimlerini bulmak amacıyla güçlü hiper kutu muhafazası (hyper-box enclosure/HBE) yöntemi önerilmiştir. Parametre optimizasyonuna ihtiyaç duymayan bu yöntem ile farklı veri kümeleri üzerinde de etkili ve tutarlı bir çıktı sağlandığı ifade edilmiştir. HBE yöntemi ile maksimum başarı oranı leave-one-out cross validation yaklaşımı kapsamında %98.61 oranı ile leukemia(AML-ALL) veri kümesinde elde edilmiştir.

[67] çalışmasında, mikrodizi veri kümesi üzerinde tümörlerin kümelenmesi için kendi kendini organize eden haritalar adı verilen tanımlama görevine uygun bir yaklaşım önerilmiştir. Bu teknik ile veri kümesinde her merkez kendine uygun olan bir kümeyi tanımlamaktadır. Böylece ALL ve AML lösemi türleri arasındaki farkın keşfedilmesi için bir strateji sunulmuştur.

2.4. Mikrodizi Veri Kümesi Kullanılarak Anomalilerin Sınıflandırılmasına İlişkin Çalışmalar - 2

[35] çalışmasında, mikrodizi veri kümelerinden potansiyel genlerin seçimi için fil arama ve ateş böceği arama algoritmaları kullanılmıştır. Ardından seçilen genler Derin Sinir Ağı sınıflandırıcısı ile sınıflandırılmıştır. Bu doğrultuda Leukemia veri kümesi üzerinde %100 ve MLL veri kümesi üzerinde %80.56 başarı oranı elde edilmiştir.

[68] çalışmasında, mikrodizi veri kümelerinden uygun gen alt kümesi elde edebilmek amacıyla knn tabanlı uyarlanabilir PSO algoritması önerilmiştir. Bu yaklaşımda tercih edilen k değeri ile uygun sayıda komşuluk oluşturularak veri kümesinin başarılı bir şekilde sınıflandırılması sağlanmıştır. Önerilen yöntem ALL-AML ve MLL veri kümeleri üzerinde sırasıyla %97.05 ve %100 doğruluk oranları üretmiştir.

[69] çalışmasında, mikrodizi veri kümeleri üzerinde genetik algoritma ve ileri beslemeli yapay sinir ağının hibrit kullanımını sağlayan bir model geliştirilmiştir.

Uygun parametreler ile uygulanan model sonucunda kanserin özel bir tipini ifade eden belirli genler için biyolojik olarak anlamlı çıkarımlar elde edilmiştir ve MLP, SVM, RF algoritmaları ile sınıflandırılmıştır.

[70] çalışmasında, mikrodizi veri kümesinin yüksek boyutunu azaltmak için k-ortalama kümeleme ve sinyal gürültü oranı sıralama yöntemlerinin birlikte kullanıldığı bir yaklaşım önerilmiştir. Ardından karıncı kolonisi optimizasyonu ile birleştirilmiş hücresel öğrenme otomatu kullanılarak optimize edilmiş gen alt kümelerine ulaşılmıştır ve destek vektör makinesi, en yakın komşu algoritması, naive bayes algoritması ile sınıflandırılmıştır. ALL-AML veri kümesi üzerinde destek vektör makinesi, k-en yakın komşu algoritması ve naive bayes algoritması için ulaşılan başarı oranları sırasıyla %58.82, %70.59 ve %94.13 olarak elde edilmiştir.

[71] çalışmasında, mikrodizi veri kümesi üzerinde kanser sınıflandırması için SBDNE yöntemi sunulmuştur. Sınıflar arası saçılımı maksimize ederken optimal izdüşüm matrisinin bulunması için sınıf içi dağılımı minimize eden bu yaklaşımın kanser sınıflandırmasında başarılı çıktılar ürettiği belirtilmiştir.

[34] çalışmasında, mikrodizi veri kümeleri içerisinde uygun genlerin seçilmesi için yapay arı kolonisi ile genetik algoritmanın hibrit kullanımı önerilmiştir. Genetik arı kolonisi olarak adlandırılan bu yaklaşım ile seçilen genler, destek vektör makinesi algoritması ile sınıflandırılmıştır. Colon, Leukemia1, Lung, SRBCT, Lymphoma ve Leukemia2 veri kümeleri için ulaşılan sınıflandırma doğruluğu sırasıyla %98.38, %100, %100, %100, %100 ve %100 olarak elde edilmiştir.

[72] çalışmasında, farklı kanser tiplerine ait mikrodizi veri kümeleri üzerinde gen seçimi ve transdüktif destek vektör makinelerinin birlikte kullanıldığı yeni bir yaklaşım önerilmiştir. Ulaşılan maksimum doğruluk oranı SRBCT veri kümesi üzerinde CBFS+TSVM yöntemlerinin birlikte kullanılması ile %93.71 olarak elde edilmiştir.

[73] çalışmasında, göğüs kanseri, CNS, kolon kanseri, DLBCL-MIT, DLBCL-Stanford, Leukemia ALL-AML, Leukemia-MLL, akciğer kanseri ve prostat tümörü mikrodizi veri kümelerinin yüksek boyut sunan yapısına bir çözüm üretmek için üç aşamalı gen seçimi önerilmiştir. Birçok state-of-the art Bayesian sınıflandırıcıları ile sınıflandırma sağlanmıştır. Yazarlar tarafından önerilen C-RPDAG sınıflandırıcısı ile

dokuz veri kümesi üzerinde ulaşılan ortalama başarı oranı %99.84 olarak elde edilmiştir.

[74] çalışmasında, mikrodizi veri kümesinden potansiyel genlerin seçimi için özyinelemeli memetik algoritma modeli önerilmiştir. Ardından seçilen genler MLP, KNN, SVM algoritmaları ile sınıflandırılmıştır ve grafikler üzerinden analiz edilmiştir.

2.5. Dijitalleştirilmiş Periferik Kan Yayma Görüntüleri Kullanılarak Anomalilerin Tespitine İlişkin Çalışmalar

[75] çalışmasında, beyaz kan hücreleri olan lenfosit, monosit, nötrofil, eozinofil, bazofil ve anormal hücrelerin ayırt edilmesi için geleneksel görüntü işleme ile derin öğrenme yaklaşımının kullanıldığı bir sınıflandırma süreci gerçekleştirilmiştir. Ortalama sınıflandırma sonuçları sırasıyla %99.8 ve yaklaşık %99 olarak elde edilmiştir.

[76] çalışmasında, kan elementleri içerisindeki nötrofil, eozinofil, bazofil, lenfosit, monosit, olgunlaşmamış granülosit, eritroblast ve trombosit elementlerinin sınıflandırılması amacıyla transfer öğrenme yaklaşımı kullanılmıştır. Genel sınıflandırma doğruluğu %96.2 olarak elde edilmiştir.

[46] çalışmasında, insanın sağlık durumu hakkında önemli bilgiler taşıyan beyaz kan hücrelerinin sayısı ve yapısı üzerinde değerlendirme sağlamak amacıyla bazofil, eozinofil, lenfosit, monosit ve nötrofil hücreleri yeni bir derin öğrenme yöntemi olan kapsül ağları ile sınıflandırılmıştır. Test veri kümesi üzerinde %96.8 doğruluk oranı elde edilmiştir.

[77] çalışmasında, vücutta meydana gelen birçok düzensizliğin sebebi olan ve kişinin sağlık durumu ile ilgili bilgiler barındıran kan sıvısı üzerinde gerçekleştirilen analizler ile ilgili bir tarama (review) çalışması yapılmıştır.

[78] çalışmasında, lökositlerin tanınmasına ve sınıflandırılmasına ilişkin bir çalışma yapılmıştır. Bu kapsamda DeepLabv3+ yaklaşımı ile segmente edilen bazofil, eozinofil, lenfosit, monosit ve nötrofil hücrelerine ilişkin mikroskopik görüntüler AlexNet transfer öğrenme yöntemi ile sınıflandırılmıştır. Önerilen algoritma beyaz kan hücresi lokalizasyonunda %98.42 ortalama kesinliğe ve %98.87 sınıflandırma doğruluğuna ulaşmıştır.

[79] çalışmasında, insanın bağışıklık sisteminde önemli rol oynayan beyaz kan hücreleri incelenmiştir. Bu doğrultuda BCCD, ALL-IDB2, JTSC ve CellaVision veri kümelerinden faydalanarak lokalize edilen beyaz kan hücreleri önerilen CNN yapısı ile sınıflandırılmıştır. Önerilen CNN yapısının nötrofil, eozinofil, monosit ve lenfosit hücre tipleri için doğruluk oranları sırasıyla %95.93, %96.48, %99.45 ve %99.92'dir.

[80] çalışmasında beyaz kan hücreleri (WBC) ile kırmızı kan hücrelerinin (RBC) segmentasyon ve sayım işlemleri için yinelemeli yapılandırılmış daire algılama algoritması kullanılarak hızlı ve daha doğru sonuç sunduğu belirtilen bir yaklaşım önerilmiştir. Önerilen yöntemin ortalama doğruluğu RBC'ler için %95.3 ve WBC'ler için %98.4 olarak elde edilmiştir.

[81] çalışmasında, hastalıkların tanısında ilk adım olan kan hücrelerinin tespiti ve sayımı için kırmızı ve beyaz kan hücreleri üzerinde bulanık mantık algoritması uygulanmıştır. Bu çalışma sonucunda kısa sürede çıktı sağlayan bir analiz gerçekleştirildiği ifade edilmiştir.

[82] çalışmasında, kan hastalıklarına ilişkin nitelendirmeler için kullanılan beyaz kan hücrelerindeki çekirdeğin morfolojisi üzerinde aktif kontür tekniği ile segmentasyon süreci gerçekleştirilmiştir. Ardından hücrelerin tipleri belirlenmiş ve sınıflandırma sağlanmıştır.

[83] çalışmasında lenfositlerin lenfoblastlardan ayırt edilmesi için spektral ve uzaysal özelliklerin birleştirildiği SVM-RFE yöntemi kullanılmıştır. Ardından entegre öznitelikler üzerinden tanımlamayı sağlamak amacıyla MLVQ sinir ağı yönteminden faydalanılmıştır. Doğruluk, duyarlılık ve özgüllük değerleri sırasıyla %92.9, %93.3 ve %92.5 olarak elde edilmiştir.

[84] çalışmasında, beyaz kan hücreleri üzerinde birkaç dijital görüntü işleme algoritmasının entegrasyonu ile çekirdek ve sitoplazma bölgelerinin segmentasyonu üzerinde temellenen bir yaklaşım önerilmiştir. Çekirdek segmentasyonu için %92 ve sitoplazma segmentasyonu için %78 doğruluk oranı elde edilmiştir.

2.6. Mikrodizi Veri Kümesi Kullanılarak Anomalilerin Sınıflandırılmasına İlişkin Çalışmalar - 3

[85] çalışmasında, mikrodizi veri kümesinden kanser geninin tanınması için ANFIS yaklaşımı kullanılmıştır. Ancak ANFIS yaklaşımı ile ulaşılan tanıma doğruluğunun ve genel sınıflandırma başarısının iyileştirilmesi hedeflenmiştir. Bu nedenle MaFO (Manta Işını Arama Optimizasyonu/Manta ray foraging optimization) algoritması ANFIS algoritması ile birlikte kullanılmıştır. Leukemia, Colon, Lymphoma, Prostate ve Breast Cancer veri kümeleri için sırasıyla %68, %92, %95, %99.65 ve %69 başarı oranları elde edilmiştir.

[86] çalışmasında farklı türdeki mikrodizi veri kümelerinden özelliklerin seçilmesi için t-testi ve genetik algoritmaya dayalı bir yaklaşım sunulmuştur. Ardından en uygun özellik alt kümesinin elde edilmesi amacıyla İç İç Genetik Algoritma (Nested Genetic Algorithm) isimli bir yaklaşım tercih edilmiştir. Son olarak en küçük optimal gen alt kümelerinin açığa çıkarılması için Artımlı Özellik Seçim (Incremental Feature Selection) tekniğinden faydalanılmıştır. Nested-GA iki farklı kanser alt tipi içeren akciğer kanserine ilişkin veriler üzerinde %98.4 sınıflandırma doğruluğu üretmiştir.

[87] çalışmasında, mikrodizi veri kümelerindeki sınıflandırma performansını iyileştirmek için Symmetrical Uncertainty (SU) ve Multi Layer Perceptron yöntemleri kullanılarak dağıtılmış özellik seçim yaklaşımı önerilmiştir. Bu doğrultuda önerilen yaklaşım kullanılarak çeşitli kümeler arasında tekrar olmaksızın özellikler dağıtılmıştır. Ardından MLP ile birçok kümeden sınırlı sayıda özelliğe sahip bir küme oluşturulmuş ve Ridor, Simple Cart, KNN, SVM algoritmaları ile sınıflandırılmıştır. Önerilen yöntemin lösemi veri kümesi üzerinde ulaştığı maksimum başarı oranı SVM yöntemi ile %98.61 olarak elde edilmiştir. MLL veri kümesi üzerinde ulaşılan maksimum başarı oranı SVM yöntemi ile %100 olarak elde edilmiştir.

[28] çalışmasında mikrodizi veri kümesi kullanılmıştır. Çalışmanın ilk aşamasında hastalıkla ilişkili genlerin seçilebilmesi için temel bileşenler analizi ve lojistik regresyon üzerinde temellenen bir yaklaşım önerilmiştir. Seçilen genler LNNDP, SVM RBF, SVM LIN, ELM, MLP, Bayes Net, Naive Bayes ve RBF 12 yöntemleri ile sınıflandırılmıştır. Lösemi veri kümesinde seçilen genler için ulaşılan maksimum kesinlik %88.24 olarak elde edilmiştir.

[88] çalışmasında, yüksek boyut sunan ALL-AML lösemi, prostat tümörü, MLL lösemi ve ALL-AML-4 mikrodizi veri kümeleri üzerinde fisher kriteri ile genlerin sayısı indirgenmiştir. Ardından karınca kolonisi ve hücresel öğrenme otomatu üzerinde kurulan bir yaklaşım ile özelliklerin alt kümesi seçilmiştir. Seçilen alt küme, destek vektör makineleri, en yakın komşu, naive bayes yöntemleri ile sınıflandırılmıştır ve grafikler üzerinden analizler yapılmıştır.

[89] çalışmasında, yüksek bir boyuta sahip olan mikrodizi veri kümesinde hastalıkları temsil eden genler, metasezgisel özellik tabanlı özellik seçim yöntemi ile seçilmiştir. Bu doğrultuda fedakar balina optimizasyon algoritması ile yinelemeler boyunca küresel optimuma ulaşabilecek aday çözümlerin verimli bir şekilde yayılması sağlanmıştır. Önerilen yöntem lösemi veri kümesi üzerinde %100 doğruluk oranı göstermiştir.

2.7. Beyaz Kan Hücrelerinin Tespiti ve Sınıflandırılmasına İlişkin Çalışmalar

[90] çalışmasında beyaz kan hücrelerinin 5 temel türünün ayırt edilmesi için yeni bir segmentasyon yaklaşımı önerilmiştir. Bu yaklaşım ile segmente edilen bölgelerden çıkarılan özellikler çok katmanlı algılayıcı, destek vektör makinesi ve hiperdikdörtgen kompozit sinir ağı yaklaşımları ile sınıflandırılmıştır. Maksimum doğruluk oranı %99.11 olarak elde edilmiştir.

[91] çalışmasında beyaz kan hücrelerinin sınıflandırılması için derin öğrenme temelli transfer öğrenme yaklaşımları kullanılmıştır. Xception, VGG19, EfficientNetB1 ve DenseNet121 tercih edilen transfer öğrenme mimarileridir. DenseNet-121 %94'lük genel doğruluk oranı ile maksimum başarı göstermiştir.

[92] çalışmasında hematolojik hastalıkların teşhisinde önemli olan beyaz kan hücrelerinin sınıflandırılması amacıyla mikroskobik kan görüntülerinden istatistiksel ve geometrik özellikler çıkarılmıştır. Ardından DT, k-NN, MLR, NB, RF ve SVM algoritmaları ile sınıflandırılmıştır. Ortalama %95 test başarısı ile Multinomial Logistic Regression (MLR) algoritması en iyi sonucu üretmiştir.

[93] çalışmasında Fine-Grained Interactive Attention Learning (FIAL) olarak isimlendirilen yarı denetimli kan hücresi sınıflandırma yöntemi önerilmiştir. Bu yöntemde etiketli veriler ile eğitilen model, etiketsiz beyaz kan hücreleri

görüntülerinden tahmini olasılık vektörleri üretmiş ve olasılık sıralamasına göre yüksek oranda ilişkili görüntülerin seçilmesi sağlanmıştır. Çok az miktarda açıklamalı WBC görüntüsü ile FIAL, BCCD veri kümesindeki verilerden çıkarılan özniteliklerin Resnet mimarisi ile sınıflandırılmasının sonucunda maksimum %93.59 başarı oranı üretmiştir.

[45] çalışmasında önerilen evrişimsel sinir ağı tabanlı bir mimari kullanılarak beyaz kan hücrelerine ait alt türlerin tespiti yapılmıştır. Bu doğrultuda eozinofil, lenfosit, monosit ve nötrofil hücreleri için elde edilen doğruluk oranları sırasıyla %97.35, %99.75, %99.92, %97.19 olarak elde edilmiştir.

[46] çalışmasında, beyaz kan hücrelerinin sınıflandırılması amacıyla kapsül ağları kullanılmıştır. Test verileri üzerinde %96.86 doğruluk üretmiştir.

[47] çalışmasında, kan yayma görüntüleri üzerinde beyaz kan hücrelerinin tespiti için bölgesel evrişimsel sinir ağı yöntemi kullanılmıştır. Böylece girdi olarak alınan görüntüdeki nesnelerin sahip olduğu koordinatlar ve bu nesnelerin veri etiketleri üretilmiştir. Ardından maksimum başarı oranı Resnet50 transfer öğrenme mimarisi ile gerçekleştirilen test sonucunda monosit hücre tipi için %98.40, bazofil hücre tipi için %98.48, eozinofil hücre tipi için %96.16, nötrofil hücre tipi için %95.04 ve lenfosit hücre tipi için %99.52 olarak elde edilmiştir.

[94] çalışmasında beyaz kan hücrelerinin sınıflandırılması için derin öğrenme mimarileri üzerinde temellenen kanonik korelasyon analizi gerçekleştirilmiştir. Kanonik korelasyon analizi ile tek bir görüntüden çıkarılan iki farklı görüntü özelliği CNN ve RNN'e ayrı ayrı uygulanmıştır. Ardından bireysel RNN ve CNN çıktısı birleştirilmiştir. Yoğun bir softmax katmanı kullanılarak tahmin sağlanmıştır ve grafikler üzerinden analiz edilmiştir.

[95] çalışmasında beyaz kan hücrelerinin sınıflandırılması için sıfırdan oluşturulan WBCsNet isimli yeni bir evrişimsel derin mimari önerilmiştir. WBCsNet tarafından ulaşılan genel sistem doğruluğu %96.1 olarak elde edilmiştir.

3. MATERYAL VE YÖNTEM

3.1. Materyal – Kullanılan Araç-Gereçler

Bu çalışmada, Intel i7-9750H CPU @ 2.60 GHz ve 32 GB ram özelliklerine sahip bilgisayar üzerinde R2019b matlab ve anaconda navigator 2.0.3 kullanılmıştır. Ek olarak Google Colaboratory'den faydalanılmıştır.

3.2. Veri Kümeleri

Bu çalışmada ücretsiz ve halka açık olan veri kümeleri kullanılmıştır. Her bir bölüm için faydalanılan veri kümeleri sırasıyla açıklanmıştır.

3.2.1. DNA genom dizilimlerindeki ekson ve intron bölgelerinin sınıflandırılması için tercih edilen veri kümesi

Bu bölümde, NCBI (Ulusal Biyoteknoloji Bilgi Bankası) [96] gen bankasından elde edilen AM400881.1, AM600680.1, AM886138.1, ve EU447303.1 numaralarına sahip 4 farklı BCR-ABL füzyon geni üzerindeki ekson ve intron bölgelerinin bilgisayar destekli sistemler ile tanımlanması için çalışmalar gerçekleştirilmiştir. BCR-ABL genlerindeki ekson ve intron bölgelerinin lokasyonları Tablo 3.1'de sunulmuştur.

Tablo 3.1. BCR-ABL genlerindeki ekson ve intron bölgelerinin lokasyonları.

AM400881.1	
1-14	Ekson Bölgesi
15-71	İntron Bölgesi
72-527	İntron Bölgesi
AM600680.1	
1-29	Ekson Bölgesi
30-578	İntron Bölgesi
579-1114	İntron Bölgesi
1115-1180	Ekson Bölgesi

Tablo 3.1. (Devamı) BCR-ABL genlerindeki ekson ve intron bölgelerinin lokasyonları.

AM886138.1	
1-31	Ekson Bölgesi
32-280	İntron Bölgesi
281-790	İntron Bölgesi
791-853	Ekson Bölgesi
EU447303.1	
1-145	Ekson Bölgesi
274-488	Ekson Bölgesi

NCBI veri kümesinden tedarik edilen DNA dizilimleri, ekson ve intron bölgelerinin sınıflandırılması için kullanılmıştır.

3.2.2. DNA dizilimleri ile ALL ve KML malignitelerinin sınıflandırılması için tercih edilen veri kümesi

Bu bölümde, NCBI (Ulusal Biyoteknoloji Bilgi Bankası) [96] gen bankasından elde edilen insan türüne ait ALL ve KML lösemi türlerine ilişkin örneklerin BCR-ABL genleri kullanılarak bilgisayar destekli sistemler ile ayırt edilebilmesi için bir çalışma gerçekleştirilmiştir. ALL ve KML lösemi türlerine ilişkin DNA dizilimleri Tablo 3.2’de sunulmuştur.

Tablo 3.2. ALL ve KML lösemi türlerine ilişkin DNA dizilimleri

ALL		KML	
Nükleotit Numarası	DNA Uzunluğu	Nükleotit Numarası	DNA Uzunluğu
FN820215.1	718	FN820207.1	377
FN820216.1	356	FN820208.1	205
FN820219.1	445	FN820209.1	516
FN820222.1	776	FN820210.1	767
FN820223.1	667	FN820211.1	698
FN820224.1	484	FN820212.1	545
FN820225.1	683	FN820213.1	195
FN820226.1	676	FN820214.1	723

Tablo 3.2. (Devamı) ALL ve KML lösemi türlerine ilişkin DNA dizilimleri.

FN820227.1	291	FN820217.1	561
FN820228.1	548	FN820218.1	272
FN820229.1	609	FN820220.1	262
FN820231.1	585	FN820237.1	708
FN820232.1	660	FN820238.1	835
FN820233.1	282	FN820239.1	621
FN820234.1	611	FN820243.1	761
FN820235.1	395	FN820248.1	708
FN820236.1	804	FN820250.1	218
FN820240.1	670	FN820252.1	564
FN820241.1	783	FN820255.1	357
FN820242.1	586	FN820259.1	212
FN820244.1	647	FN820260.1	422
FN820245.1	643	FN820261.1	489
FN820246.1	582	FN820262.1	373
FN820247.1	695	FN820265.1	994
FN820249.1	508	FN820266.1	707
FN820251.1	555	FN820267.1	809
FN820253.1	895		
FN820254.1	912		
FN820256.1	388		
FN820257.1	547		
FN820258.1	629		
FN820263.1	691		
FN820264.1	429		
FN820268.2	604		

NCBI veri kümesinden elde edilen DNA dizilimleri, löseminin temel türü olan ALL ve KML malignitelerinin sınıflandırılması için kullanılmıştır.

3.2.3. Mikrodizi veri kümesi kullanılarak ALL ve AML malignitelerinin sınıflandırılması için tercih edilen veri kümesi

Bu bölümde, bioinformatics laboratory [97] vasıtasıyla temin edilen gen ekspresyon verileri kullanılmıştır. Bu doğrultuda 5147 gen ve 72 örnekten (47 ALL ve 25 AML) oluşan lösemi hastalığına ait gen ekspresyon verileri excel dosyasına aktarılmış ve csv formatına dönüştürülmüştür. Leukemia isimli veri kümesine ait gen ekspresyon verilerinin bir kısmı Şekil 3.1’de sunulmuştur.

ALL	218	6479	171	-30	-97	448	2822	70
ALL	132	22143	211	101	-92	73	404	-10
ALL	-24	24327	253	23	-95	-104	153	16
ALL	-36	36417	229	98	-60	30	594	17
ALL	255	14989	67	-32	-168	147	659	-14
ALL	316	18343	32	23	-113	191	753	49
ALL	54	12801	161	4	-113	35	495	29
ALL	81	25673	366	67	-179	-62	503	79
AML	101	16716	172	18	-66	182	378	50
AML	132	45492	67	7	-79	45	372	21
AML	318	23865	-14	7	216	227	722	58
AML	325	17359	286	-11	1287	810	1378	21
AML	392	19406	227	-13	207	260	720	81
AML	241	17878	218	52	51	132	647	10
AML	193	26052	121	6	-95	-22	202	44
AML	312	11983	114	32	-435	135	423	45
AML	230	23508	61	45	-125	85	295	90
AML	330	17996	371	6	-36	123	532	51
AML	337	19679	238	-16	2003	1735	3995	12

Şekil 3.1. Leukemia isimli veri kümesine ait gen ekspresyon verileri

Bioinformatics laboratory vasıtasıyla temin edilen Leukemia isimli mikrodizi veri kümesi, lösemnin temel türlerinden olan ALL ve AML malignitelerinin sınıflandırılması için kullanılmıştır.

3.2.4. Mikrodizi veri kümesi kullanılarak ALL, AML ve MLL malignitelerinin sınıflandırılması için tercih edilen veri kümesi

Bu bölümde, bioinformatics laboratory [97] vasıtasıyla temin edilen gen ekspresyon verileri kullanılmıştır. Bu doğrultuda 12533 gen ve 72 örnekten (24 ALL, 20 MLL ve 28 AML) oluşan lösemi hastalığına ait gen ekspresyon verileri excel dosyasına aktarılmış ve csv formatına dönüştürülmüştür. MLL isimli veri kümesine ait gen ekspresyon verilerinin bir kısmı Şekil 3.2’de sunulmuştur.

ALL	-52	-62	-50	-91	326	140	-33
ALL	-102	98	21	-1	158	158	-21
MLL	-66	-57	2	-106	237	-100	-10
MLL	-79	-28	-65	-125	54	87	-43
MLL	-185	-62	-41	-45	116	259	3
MLL	-359	-86	-75	-121	325	-317	-40
MLL	-139	8	-72	-128	140	-47	-40
MLL	-164	16	-34	-29	214	-205	-52
MLL	-84	-42	26	-138	428	-300	-14
MLL	-79	53	-75	-75	246	32	142
MLL	-135	100	-10	-77	164	55	151
MLL	-232	118	-9	-178	805	-149	-59
MLL	-74	-9	-33	-60	308	37	-14
MLL	-231	132	-15	96	537	-34	-19
MLL	-363	-180	-84	-216	376	-124	-46
MLL	-177	73	-24	-112	97	-103	-21
MLL	-261	14	-46	-50	573	148	-53
MLL	-252	-8	-103	272	451	117	-39
MLL	-278	-33	-93	-135	469	68	-49
MLL	-235	-148	-105	-146	311	-179	-59
MLL	-348	88	41	-140	460	126	-49
MLL	-295	72	-24	-190	379	115	-70
AML	-217	-119	19	215	-173	184	-57
AML	-217	114	67	338	17	-26	-46
AML	-160	-114	19	-129	753	-30	-80
AML	-213	-190	-81	213	337	-258	-52
AML	-155	-117	-140	-210	-410	-201	-67

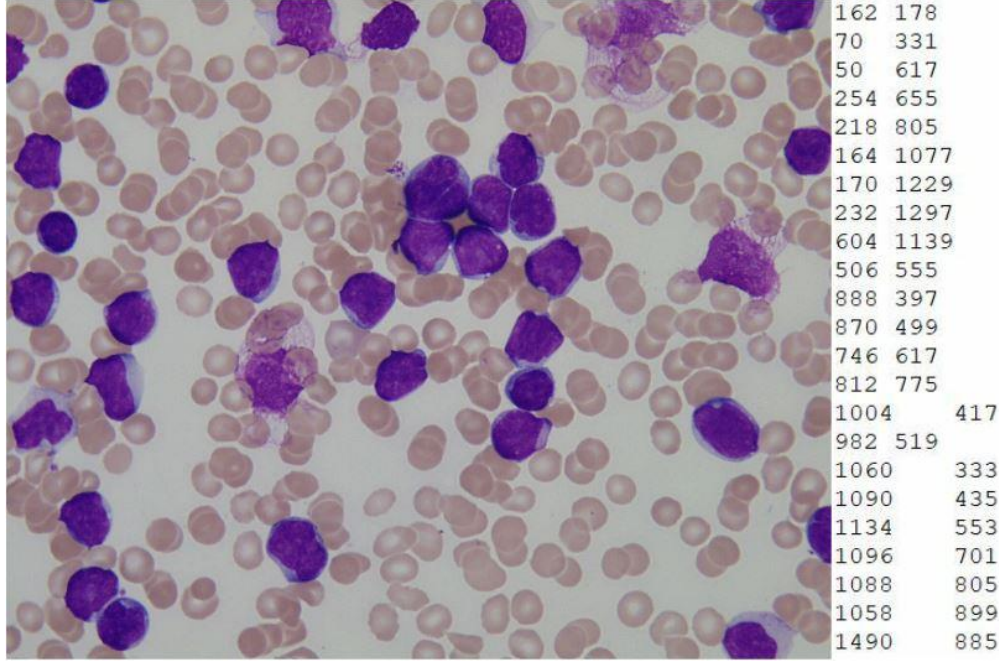
Şekil 3.2. MLL isimli veri kümesine ait gen ekspresyon verileri

Bioinformatics laboratory vasıtasıyla temin edilen MLL isimli mikrodizi veri kümesi, lösemiye ilişkin ALL, AML ve MLL malignitelerinin sınıflandırılması için kullanılmıştır.

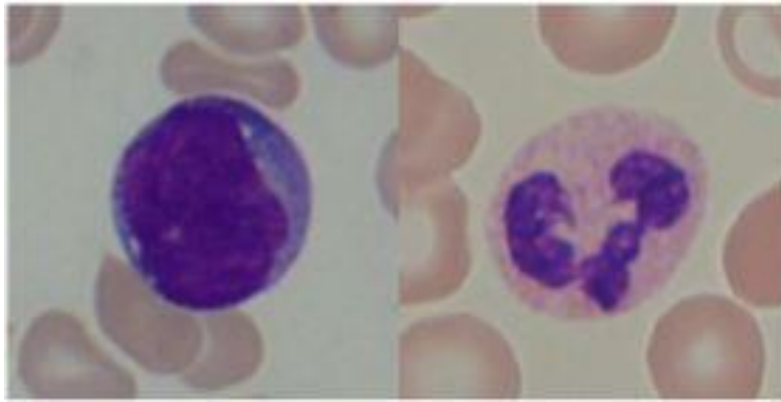
3.2.5. Dijitalleştirilmiş periferik kan yayma görüntüleri kullanılarak ALL malignitesinin tespiti için tercih edilen veri kümesi

Bu bölümde, Acute Lymphoblastic Leukemia Image Database [98] vasıtasıyla temin edilen normal ve ALL hastalıklarına sahip bireylerin dijitalleştirilmiş periferik kan yayma görüntüleri kullanılarak kan elementleri analiz edilmiştir. Canon PowerShot G5 kamera kullanılarak yakalanan jpg formatındaki tüm görüntüler, hematologlar tarafından blast ve normal hücreler olarak koordinatları ile ifade edilmiştir. ALL-IDB1 veri kümesindeki kan elementleri ve bu kan elementleri arasından hematologlar tarafından tanımlanan blast hücrelerinin koordinatları ile ALL-IDB2 veri kümesindeki

blast hücrenin ve sağlıklı hücrenin görsel temsili sırasıyla Şekil 3.3. ve Şekil 3.4'te verilmiştir.



Şekil 3.3. ALL-IDB1 veri kümesindeki kan elementleri ve bu kan elementleri arasından hematologlar tarafından tanımlanan blast hücrelerinin koordinatları



Şekil 3.4. ALL-IDB2 veri kümesindeki blast hücre ve sağlıklı hücre

Periferik kan yayma görüntülerinde hasta (blast) ve sağlıklı hücrelerin bilgisayar destekli sistemler ile ayırt edilebilmesi için modelin eğitiminde gerekli bir aşama olan segmentasyon işlemi ALL-IDB1 ve ALL-IDB2 veri kümelerinde yapılmıştır.

3.2.6. Mikrodizi veri kümesi kullanılarak T-ALL, B-ALL ve T-LL malignitelerinin sınıflandırılması için tercih edilen veri kümesi

Bu bölümde, bioinformatics laboratory [97] vasıtasıyla temin edilen gen ekspresyon verileri kullanılmıştır. Bu doğrultuda 15434 gen ve 48 örnekten (20 T-ALL, 10 B-ALL

ve 18 T-LL) oluşan lösemi hastalığına ait gen ekspresyon verileri excel dosyasına aktarılmış ve csv formatına dönüştürülmüştür. GSE1577 isimli veri kümesine ait gen ekspresyon verilerinin bir kısmı Şekil 3.5’te sunulmuştur.

3	202.5	138.6	45.9	85.2	1135.8	27.2	T-LL
5	125.5	26.5	72.4	36.2	2092	49.3	T-LL
2	315.4	105.8	48.5	19.1	2195.1	49.5	T-LL
7	127.1	82.3	54.7	3.3	611.2	66.3	T-LL
3	269.7	45.1	84.4	97	1017.1	42	T-LL
	197	190.8	37.6	49	1841.5	25.1	T-LL
5	340.6	88.2	40.1	75.9	1292.6	105.7	T-LL
3	265.4	65.1	12.2	20	1340.3	43	T-LL
5	579.1	98	31.7	20.9	3399	40.5	T-LL
5	65	53.6	37.7	25.5	242.2	53	T-ALL
3	99.3	70.4	51.3	5	632.3	39.4	T-ALL
9	183.6	66.1	41.1	1.8	391	70.1	T-ALL
3	13.7	211.6	32.2	4.4	79.5	18.6	T-ALL
9	54.5	41.4	37.6	13	179.3	62	T-ALL
9	54.5	41.4	37.6	13	179.3	62	T-ALL
5	34.1	40.9	32.8	3	93.4	34.6	T-ALL
1	63.1	23.3	19.6	3.1	309.6	72.3	T-ALL
5	75.4	116.1	34.2	5.5	810.6	22	T-ALL
5	195.2	87.1	43	4.6	778.1	23	T-ALL
.7	97.4	59.9	45.8	17.5	529.3	39.9	B-ALL
7	198.7	60	45.5	3.6	359.5	33.5	B-ALL
.8	58.5	54.8	52.9	8.4	278.6	13.2	B-ALL
9	255.3	38.3	68.1	29.9	1102.8	21.6	B-ALL
4	128.2	38.3	40.3	2	440.2	20.9	B-ALL
1	176	69.2	44.9	15.5	514.1	40.4	B-ALL
5	175.2	54.1	29.3	6.3	497.4	57.4	B-ALL

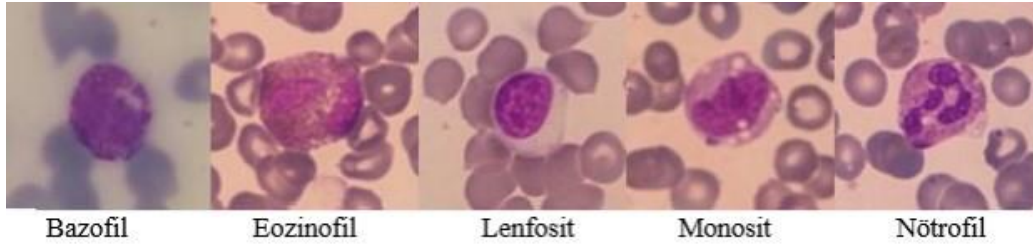
Şekil 3.5. GSE1577 isimli veri kümesine ait gen ekspresyon verileri

Bioinformatics laboratory vasıtasıyla temin edilen GSE1577 isimli mikrodizi veri kümesi, lösemiye ilişkin T-ALL, B-ALL ve T-LL malignitelerinin sınıflandırılması amacıyla kullanılmıştır

3.2.7. Beyaz kan hücrelerinin tespiti ve sınıflandırılması için tercih edilen veri kümesi

Bu bölümde hastalık oluşturan maddelere ve ölümcül enfeksiyonlara karşı vücut savunmasında rol oynayan beyaz kan hücreleri için bir analiz yapılmıştır. Bu doğrultuda ücretsiz ve halka açık olan Raabin Health veri kümesi kullanılmıştır. Bu veri kümesinde glemsa boyama tipi ile boyanan eozinofil, bazofil, monosit, nötrofil ve lenfosit kan hücreleri, Samsung Galaxy S5 ve LG G3 telefon kameraları ile 100x büyütme oranında Olympus CX18 ve Zetss mikroskopları kullanılarak elde edilmiştir.

Raabin Health veri kümesinden tedarik edilen dijitalleştirilmiş beyaz kan hücrelerinin [99] görsel tasviri Şekil 3.6’da verilmiştir.



Şekil 3.6. Raabin Health veri kümesinden tedarik edilen dijitalleştirilmiş beyaz kan hücreleri [99]

Eozinofil, bazofil, monosit, nötrofil ve lenfosit kan hücrelerinin bilgisayar destekli sistemler ile ayırt edilebilmesi için modelin eğitiminde gerekli bir aşama olan segmentasyon işlemi Raabin Health veri kümesi için gerçekleştirilmiştir.

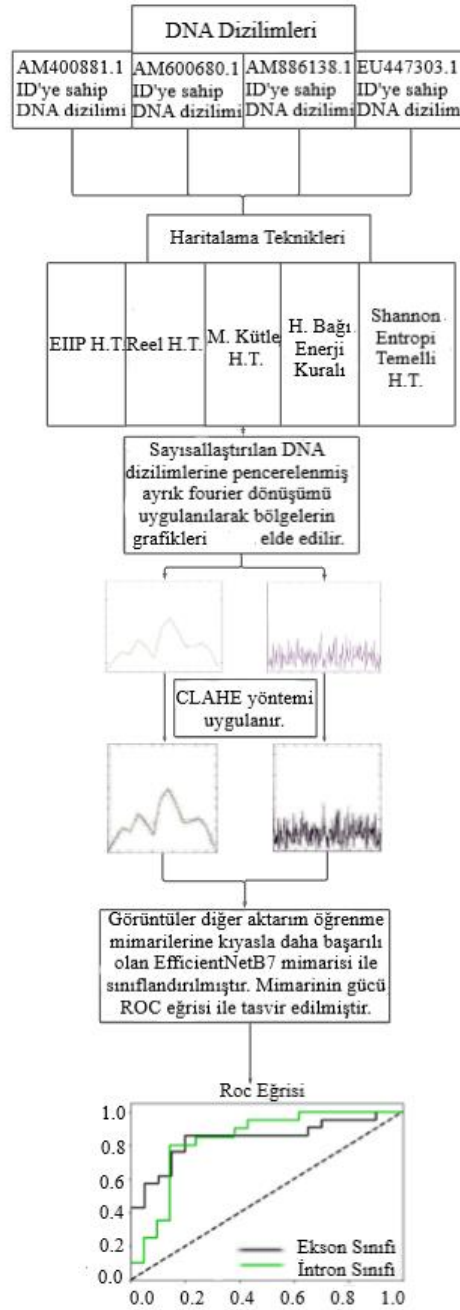
3.3. Yöntemler

3.3.1. DNA genom dizilimlerindeki ekson ve intron bölgelerinin sınıflandırılmasına ilişkin iş akışı

DNA A, T, G ve C harflerinden oluşan sembolik bir yapıya sahiptir. Genler ve genler arası boşluklardan meydana gelen bu yapı, ökaryot hücreler için ekson ve intron bölgelerinin ayırımına sahiptir [9]. Ekson bölgeleri canlılıkla ilgili bilgiler barındıran ayırt edici bir kısımdır. Bu bölgenin doğru bir şekilde tanımlanabilmesi araştırmacılar tarafından zorlu bir problem olarak ifade edilmiştir [9, 11, 13, 14].

Bu bölümde NCBI veri kümesinden elde edilen BCR-ABL DNA dizilimleri üzerinde ekson ve intron bölgelerinin tespiti ile ilgili bir çalışma yapılmıştır. Çalışmanın ilk kısmında, DNA dizilimlerinin 5 farklı sayısal haritalama yöntemi (EIIP haritalama tekniği, Reel Haritalama Tekniği, Moleküler Kütle Haritalama Tekniği, Eşleştirilmiş Haritalama Tekniğinin Hidrojen Bağlı Enerji Kuralı, Shannon Entropi Temelli Haritalama Tekniği) ile sayısallaştırılması sağlanmış ve ardından sayısal sinyal işleme yöntemi olan fourier dönüşümü ile dizilimlere ait spektral yoğunluk bilgisi grafikler ile ifade edilmiştir. Ardından ekson ve intron bölgelerinin sınıflandırılması amacıyla kullanılan 10 ayrı transfer öğrenme mimarileri ile (VGG16, VGG19, Xception, MobilenetV2, Inception, ResnetV2, DenseNet169, DenseNet201, Resnet101, Resnet150, EfficientNetB7) görüntü üzerinde yapılacak çıkarımların nihai sonucu

etkilemesinden dolayı kenar, köşe gibi önemli noktaların CLAHE yöntemi ile belirginleştirilmesi sağlanmıştır. Son olarak görüntüler, bu transfer öğrenme mimarileri ile sınıflandırılmıştır. Ekson ve intron bölgelerinin sınıflandırılması amacıyla ilk adımda önerilen akış şeması Şekil 3.7’de verilmiştir.

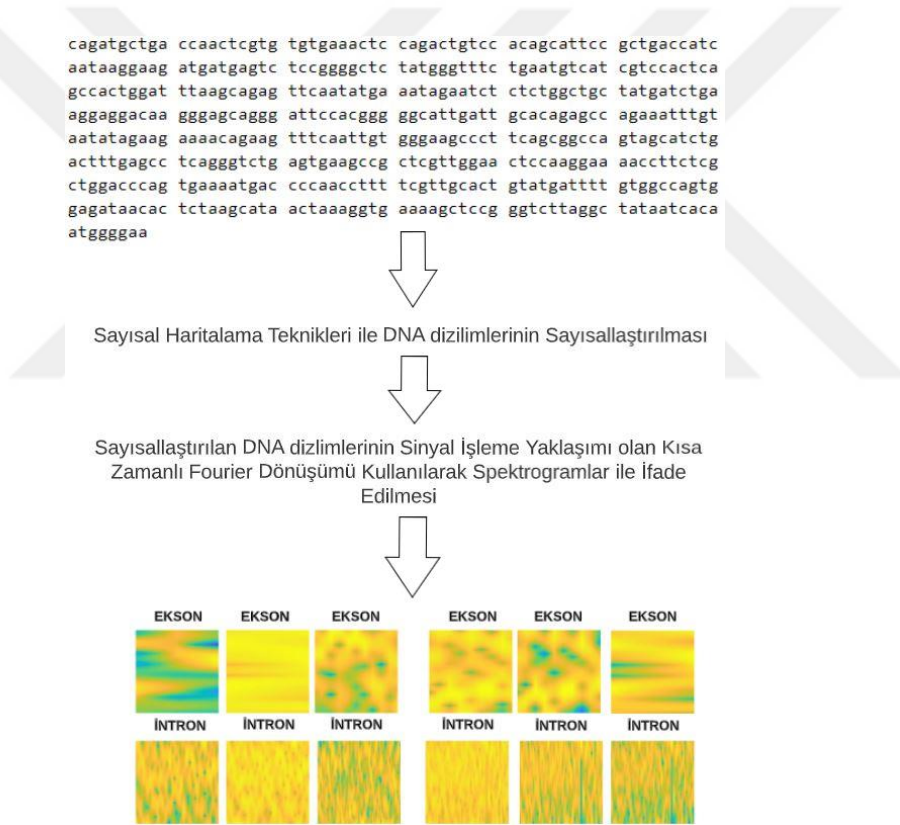


Şekil 3.7. Ekson ve intron bölgelerinin sınıflandırılması amacıyla ilk adımda önerilen akış şeması

İlk aşama için ulaşılan en başarılı performansın eşleştirilmiş haritalama tekniğinin hidrojen bağı enerji kuralı ile sayısallaştırılan dizilimlere sinyal işleme yaklaşımı

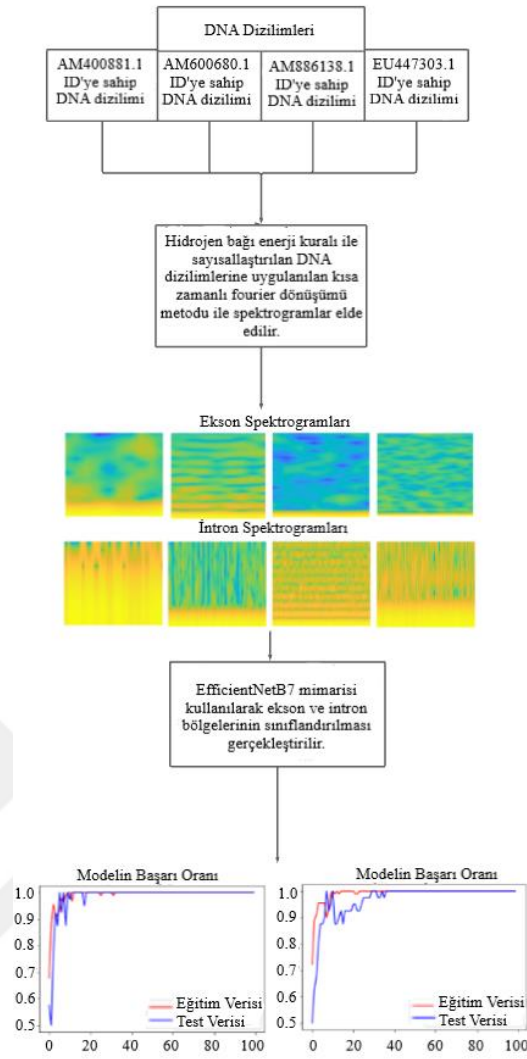
uygulanarak elde edilen grafiklerin transfer öğrenme mimarisi yöntemi olan EfficientNetB7 ile sınıflandırılması sonucunda üretildiği Tablo 4.1- Tablo 4.5’te ifade edilmiştir.

Çalışmanın ikinci kısmında, ilk aşamadaki en başarılı sayısal haritalama tekniği olan eşleştirilmiş haritalama tekniğinin hidrojen bağı enerji kuralı ile sayısallaştırılan DNA dizilimlerine kısa zamanlı fourier dönüşümü uygulanmıştır. Böylece durağan olmayan serilerin analizi için düşük performans gösteren [100] ayrık zamanlı fourier dönüşümünün bu dezavantajına problem kapsamında bir çözüm üretilmiştir. Kısa zamanlı fourier dönüşümü yöntemi ile elde edilen spektrogramların örnek bir DNA diziliminden görüntüye dönüşüm sürecini ifade eden temsili akış Şekil 3.8’de verilmiştir.



Şekil 3.8. Örnek bir DNA diziliminden görüntüye dönüşüm sürecini ifade eden temsili akış

Kısa zamanlı Fourier dönüşümü ile elde edilen spektrogram görüntüleri, ilk aşamanın en başarılı transfer öğrenme yöntemi olan EfficientNetB7 mimarisi ile sınıflandırılmıştır. Ekson ve intron bölgelerinin sınıflandırılması amacıyla ikinci adımda önerilen akış şeması Şekil 3.9’da verilmiştir.



Şekil 3.9. Ekson ve intron bölgelerinin sınıflandırılması amacıyla ikinci adımda önerilen akış şeması

Bu bölümde, DNA dizimleri üzerinde ekson ve intron bölgelerinin ayırt edilmesi amacıyla iki farklı kısımda gerçekleştirilen bir çalışma yapılmıştır.

3.3.1.1. DNA genom dizimlerindeki ekson ve intron bölgelerinin sınıflandırılması için kullanılan yöntemler

DNA, organizmanın inşa edilmesi, canlılık işlevlerinin gerçekleştirilmesi ve biyolojik gelişmelerin sürdürülebilmesi için kritik bilgilerin yer aldığı önemli bir genom parçasıdır [9]. A, T, G ve C organik bazlarından meydana gelen bu genom parçasında bazların oluşturduğu kombinasyonlar, DNA'nın gen dizimlerini ifade etmektedir [8]. Biyolojik süreçlerde görev alan genler, ökaryot hücreler için protein kodlayan (ekson) ve protein kodlamayan (intron) bölgelerin ayrımına sahiptir. Bu bölgelerin ayırt

edilmesi arařtırmacılar tarafından zorlu bir problem olarak ifade edilmiřtir [9, 11, 13-14].

Bu bölümde, yařam süreçlerinin analiz edilmesinde önemli bir ayıraç olan ekson ve intron bölgeleri arařtırılmıřtır. Ancak DNA dizilimleri sembolik bir yapıya sahiptir. Bu durum analizleri zorlařtırmaktadır. Bu nedenle sayısal haritalama teknikleri kullanılarak DNA dizilimleri sayısallařtırılmıřtır.

Sayısal haritalama teknikleri

Sembolik bir yapıya sahip olan DNA dizilimlerinin sayısallařtırılmasında kullanılan sayısal haritalama teknikleri sabit haritalama teknikleri ve fiziko-kimyasal özellik tabanlı haritalama teknikleridir [12]. Ancak yeni ve farklı haritalama teknikleri halen literatüre kazandırılmaya devam etmektedir. Literatürde yer alan bazı sayısal haritalama yöntemleri ařağıda sunulmuřtur.

Reel haritalama tekniğı

Sabit haritalama teknikleri ierisinde yer alan reel haritalama tekniğinde A-T ve G-C organik bazları, tamamlayıcı olma özelliğine sahiptir [101]. Bu durum DNA dizisinin tamamlayıcı sarmalını bulmada etkilidir [101].

Reel haritalama tekniğı ile sayısallařtırılması hedeflenen DNA dizilimindeki A, T, C ve G organik bazlarına sırasıyla -1.5, 1.5, 0.5 ve -0.5 atamaları yapılır [12]. Reel haritalama tekniğine göre sayısallařtırılan örnek bir DNA dizilimi iin yapılan sayısal temsil Tablo 3.3'te gösterilmiřtir.

Moleküler kütle haritalama tekniğı

Sabit haritalama teknikleri ierisinde yer alan haritalama tekniğidir. Moleküler kütle haritalama tekniğı ile sayısallařtırılması hedeflenen DNA dizilimindeki A, T, C ve G organik bazlarına sırasıyla 134, 125, 110 ve 150 atamaları yapılır [12]. Moleküler kütle haritalama tekniğine göre sayısallařtırılan örnek bir DNA dizilimi iin yapılan sayısal temsil Tablo 3.3'te gösterilmiřtir.

EIIP haritalama tekniğı

Fiziko-kimyasal özellik tabanlı teknikler ierisinde yer alan haritalama tekniğidir. Bu teknik kullanılarak DNA diziliminde yer alan her bir nükleotidin EIIP temsilindeki yarı deęerlik sayısı ile eřleřtirilmesi saęlanır. EIIP haritalama tekniğı ile sayısallařtırılması hedeflenen DNA dizilimindeki A, T, C ve G organik bazlarına

sırasıyla 0.1260, 0.1335, 0.1340 ve 0.0806 atamaları yapılır [12]. EIIP haritalama tekniğine göre sayısallaştırılan örnek bir DNA dizilimi için yapılan sayısal temsil Tablo 3.3'te gösterilmiştir.

Eşleştirilmiş Haritalama Tekniği

Fiziko-kimyasal özellik tabanlı teknikler içerisinde yer alan haritalama tekniğidir. Birbirini tamamlayan sarmal bir düzende iki ipliğe sahip olan DNA yapısının karşılıklı eşleşen adenin-timin ve guanin-sitozin bazlarına 1 ve -1 değerleri atanarak dizilimin sayısallaştırılması sağlanır [9, 12]. Karmaşıklığın indirgendiği bir DNA yapısı sunan bu teknik kapsamında 7 farklı kural mevcuttur [12, 101].

Pirimidin-pürin kuralı

Pirimidin-pürin kuralı ile sayısallaştırılması hedeflenen DNA dizilimindeki adenin ve guanin bazlarına 1, sitozin ve timin bazlarına -1 değerlerinin atanması ile dizilim sayısallaştırılır [12].

AA' kuralı

AA' kuralı ile sayısallaştırılması hedeflenen DNA dizilimindeki adenin bazına 1, timin, guanin ve sitozin bazlarına -1 değerlerinin atanması ile dizilim sayısallaştırılır [12].

TT' kuralı

TT' kuralı ile sayısallaştırılması hedeflenen DNA dizilimindeki timin bazına 1, adenin, guanin ve sitozin bazlarına -1 değerlerinin atanması ile dizilim sayısallaştırılır [12].

GG' kuralı

GG' kuralı ile sayısallaştırılması hedeflenen DNA dizilimindeki guanin bazına 1, adenin, timin ve sitozin bazlarına -1 değerlerinin atanması ile dizilim sayısallaştırılır [12].

CC' kuralı

CC' kuralı ile sayısallaştırılması hedeflenen DNA dizilimindeki sitozin bazına 1, adenin, timin ve guanin bazlarına -1 değerlerinin atanması ile dizilim sayısallaştırılır [12].

Hibrit kuralı

Hibrit kural ile sayısallaştırılması hedeflenen DNA dizilimindeki adenin veya sitozin bazlarına 1, timin veya guanin bazlarına -1 değerlerinin atanması ile dizilim sayısallaştırılır [12].

Hidrojen bağı enerji kuralı

Hidrojen bağı enerji kuralı ile sayısallaştırılması hedeflenen DNA diziliminde 3'lü hidrojen bağı ile birbirine bağlanan guanin ve sitozin bazlarına 1, 2'li hidrojen bağı ile birbirine bağlanan adenin ve timin bazlarına -1 değerinin atanması ile dizilim sayısallaştırılır [12].

Bu çalışmada eşleştirilmiş haritalama tekniği kapsamında incelenen 7 kural arasından hidrojen bağı enerji kuralı ile DNA dizilimleri sayısallaştırılmıştır. Hidrojen bağı enerji kuralına göre sayısallaştırılan örnek bir DNA dizilimi için yapılan sayısal temsil Tablo 3.3'te gösterilmiştir.

Tablo 3.3. Sayısal haritalama tekniklerine göre sayısallaştırılan örnek bir DNA dizilimi [12].

Haritalama Teknikleri (HT)		Örnek Dizilim için Sayısal Temsiller ... ATGCATGCAG ...
Sabit HT	Reel HT	... -1.5, 1.5, -0.5, 0.5, -1.5, 1.5, -0.5, 0.5, -1.5, -0.5 ...
	MolekülerKütle HT	... 134, 125, 150, 110, 134, 125, 150, 110, 134, 150 ...
	Tam Sayı HT	... 2, 0, 3, 1, 2, 0, 3, 1, 2, 3 ...
	Karmaşık HT	... 1+j, 1-j, -1+j, -1-j, 1+j, 1-j, -1+j, -1-j, 1+j, -1+j ...
Fiziko-Kimyasal Özellik Tabanlı HT	EHP HT	... 0.1260, 0.1335, 0.0806, 0.1340, 0.1260, 0.1335, 0.0806, 0.1340, 0.1260, 0.0806 ...
	Atomik Sayı HT	... 70, 66, 78, 58, 70, 66, 78, 58, 70, 78 ...
	DNA-Yürüyüş HT	... -1, 0, 0, 1, -1, 0, 0, 1, -1, 0 ...
	Eşleştirilmiş Sayısal HT (Hidrojen Bağı Enerji Kuralı)	... -1, -1, 1, 1, -1, -1, 1, 1, -1, 1 ...

Literatürde yer alan bazı sayısal haritalama tekniklerinin avantaj ve dezavantajları Tablo 3.4'te verilmiştir.

Tablo 3.4. Sayısal haritalama tekniklerinin avantajları ve dezavantajları [12, 101,102].

Haritalama Teknikleri	Avantajları	Dezavantajları
Reel HT	A-T ve G-C organik bazları tamamlayıcı olma özelliğine sahiptir.	DNA diziliminde bulunmayan matematiksel özelliklerin tanıtılması gerekmektedir.
MolekülerKütle HT	DNA dizilimlerinin çok boyutlu bir uzaya haritalanmasını sağlamaktadır.	Daha fazla araştırılması gerekmektedir.
Tam Sayı HT	Basit bir gösterime sahiptir. Verimli bir haritalama süreci sağlamaktadır.	DNA diziliminde bulunmayan matematiksel özelliklerin tanıtılması gerekmektedir.
Karmaşık HT	Nükleotitlerin eşlenik olma özelliğini yansıtarak daha doğru bir gen tahmini sağlamaktadır.	Zaman alanındaki baz yanılmasının (base bias) tanıtılması gerekmektedir
EIIP HT	DNA'nın fiziko-kimyasal özelliğini yansıtarak hesaplama yükünü azaltır ve gen ayırım yeteneğini geliştirir.	Bazı genomlar için kodlama bölgesinin tespiti hususunda başarısızlıkları mevcuttur.
Atomik Sayı HT	DNA'nın fiziko-kimyasal özelliğini yansıtır.	Çeşitli kodlama şemalarında farklı sonuçlar vermesinden dolayı araştırma kapsamında tekdüzeliğe neden olabilmektedir.
DNA-Yürüyüş HT	Uzun menzilli korelasyon bilgisini kullanarak nükleotit bileşimindeki değişiklikleri ortaya çıkarır.	1000 bazdan daha uzun olan dizilimler için uygun değildir.
Eşleştirilmiş Sayısal HT (Hidrojen Bağ Enerji Kuralı)	Karmaşıklığı azaltılmış bir DNA yapısı sunmaktadır.	Karmaşıklığın azaltılması amacıyla sunduğu 7 kural içerisinden çalışma kapsamında başarılı olacak kuralın seçilmesi önemlidir.

Farklı avantajlar ya da dezavantajlar içeren sayısal haritalama tekniklerine göre sayısallaştırılan DNA dizilimlerine ek olarak son zamanlarda literatürde önerilen ve haritalama teknikleri arasında yer alan Shannon Entropi Temelli Haritalama Tekniği de bu çalışmada kullanılmıştır.

Shannon entropi temelli haritalama tekniği

DNA, canlılıkla ilgili devasa bilgi barındıran genom birimidir. Sembolik tanımlamalardan oluştuğu için matematiksel olarak analiz edilebilmesi zordur. DNA dizilimlerinin sayısallaştırılması için farklı birçok çalışma gerçekleştiren bilim dünyası halen alternatif yöntemler üretmeye devam etmektedir. [48] makalesinde önerilen shannon entropi temelli haritalama tekniği, DNA dizilimlerinin sayısallaştırılması amacıyla geliştirilen yeni bir yaklaşımdır.

Bu yaklaşımın temelinde olan shannon entropisi, karmaşıklığın bir ölçümü olarak nitelendirilmektedir ve 1948 yılında Shannon tarafından tanıtılmıştır. Zaman içerisinde topolojik, metrik, kolmogorov gibi farklı yapılarda ifade edilen shannon entropisi, DNA dizilimlerinin sayısal temsili için yetersiz olarak kabul edilmiştir. Bu nedenle DNA dizilimlerinin sayısallaştırılmasında Shannon entropi temelli haritalama tekniği önerilmiştir [48]. Bu yaklaşımda, Ali Karcı tarafından geliştirilen shannon denkleminin kesirli bir türevi olan fraksiyonel shannon entropi temelli yaklaşım üzerinden DNA dizilimlerinin sayısal temsili sağlanmıştır. Kodon dağılımlarına ilişkin entropi değerlerinin hesaplandığı bu yöntemin matematiksel ifadesi denklem 3.1’de verilmiştir [12].

$$Sf = -\sum_i [(p(x_i))^{\alpha_i} p(x_i) \log p(x_i)] \quad (3.1)$$

Denklem 3.1’de ifade edilen $p(x_i)$, her bir kodonun tekrarlanma sıklığı ve α_i , dizilim üzerinden uyarlamalı olarak hesaplanılan değişkendir. α parametresinin matematiksel tanımı denklem 3.2’de verilmiştir [12].

$$\alpha_i = \frac{1}{\log(p(x_i))} \quad (3.2)$$

Bu yaklaşım ile kodon olasılıklarına ait korelasyonların net bir şekilde ifade edildiği ve DNA dizilimine ilişkin sayısal temsilin daha geniş bir ölçekte sağlandığı belirtilmiştir [12].

Çalışmada kullanılan sayısal sinyal işleme yöntemleri

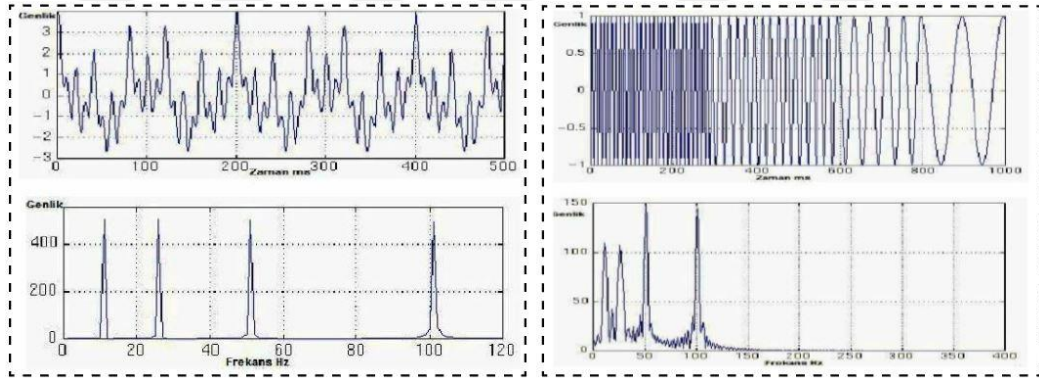
Biyoenformatik, biyolojik verilerin analizinde ve değerlendirilmesinde tercih edilen disiplinler arası bir alandır [56]. Bu alan çerçevesinde gelişmekte olan genomik sinyal işleme verilerin içerisindeki ilişkilerin, desenlerin ve periyodiklik durumlarının analiz edilmesini sağlar [56]. Aynı zamanda genomik verilere ilişkin karanlık noktaların

aydınlatılmasında umut kaynağıdır ve genom teknolojisindeki son gelişmelerle birlikte manuel değerlendirmelerin yetersiz kaldığı noktada sunduğu hızlı çıktılar ile tercih edilen bir teknolojidir [9, 56]. 2000’li yılların başından beri genomik veriler için sayısal sinyal işleme yaklaşımlarının kullanıldığı genomik sinyal işleme alanında gerçekleştirilen çalışmalar, büyüyen bir gelişim eğrisi oluşturmuştur [7, 56].

Bu bölümde, sayısallaştırılan DNA dizilimleri sayısal sinyal işleme yaklaşımları kapsamında değerlendirilen pencerelemiş fourier dönüşümü ve pencerelemiş kısa zamanlı fourier dönüşümü ile analiz edilmiştir.

Pencerelemiş fourier dönüşümü

Fourier dönüşümü, belirli bir zaman periyodunda sıralanmış verilerin zaman ortamından frekans ortamına aktarılmasını sağlayan bir yöntemdir. Aktarım süreci sonunda zamana ait frekans bilgisi korunurken zaman bilgisi yitirilir. Bu durum yöntemin dezavantajıdır. Öte yandan durağan serilerin analizinde anlamlı sonuçlar üretir. Bu nedenle frekans bilgileri üzerinden çıkarım yapılması gereken durumlar için tercih edilmektedir [100]. Durağan ve durağan olmayan zaman serilerine ilişkin dönüşümün tasviri Şekil 3.10’da verilmiştir.



Şekil 3.10. Durağan olan ve durağan olmayan zaman serilerine ilişkin dönüşümün tasviri [100]

Şekil 3.10. incelendiğinde frekanslarda görülen ani dalgalanmalar, eğimler ya da kırılmalar durağan olmayan serilerde görülmektedir. Bununla birlikte durağan seriler, fourier dönüşümünün analizi için uygun çıktılar sunmaktadır [100].

Fourier dönüşümünün temel matematiksel ifadesi denklem 3.3’te verilmiştir.

$$x(f)=\int_{-\infty}^{+\infty} x(t) e^{-2\pi ift} dt \quad (3.3)$$

Verilen eşitlikte f , frekansı; t , zamanı; $x(f)$, fourier dönüşümünü ve $x(t)$, zaman serisini tanımlamaktadır [100]. $x(t)$, sinyalin periyodik olmadığı durumlarda sinyali periyodik olarak gösterebilmek amacı ile yapay olarak uzatılan bir değişkendir. Bu durum sonlanma kısımlarında ek süreklilikler oluşturduğundan dolayı pencerelemiş fourier dönüşümü çözüm olarak kullanılmıştır [103].

Pencerelemiş Fourier dönüşümünde giriş sinyalinin pencere fonksiyonu ile ayrılan her bir bölgesi, mevcut frekans içerikleri için tek tek incelenir [103]. Ani geçişlere sahip olan sinyal için giriş verileri sonlanma noktalarında sıfır olacak şekilde pencerelemiş [103]. Periyodik olmayan sinyaller için iyileştirme imkanı sunan pencerelemiş fourier dönüşümünde pencerenin etkisi, sinyali zamanda yerelleştirmesi olarak ifade edilir [103]. Öte yandan pencerenin büyüklüğü nihai başarı oranını etkileyebilen bir parametredir [100]. Bu nedenle farklı pencereleme fonksiyonları kullanılarak elde edilen çıktılar üzerinde çeşitlilik arttırılmıştır. Bu doğrultuda çalışmada kullanılan pencereleme fonksiyonları, daha iyi spektral sonuç sağladığı ifade edilen [104] hibrit pencereleme fonksiyonları içerisinde seçilmiştir. Çalışmanın ilk bölümünde tercih edilen hibrit pencereleme fonksiyonlarının matematiksel ifadeleri denklem 3.4 - 3.12 verilmiştir [104].

$$w1 = 0.5(\cosh) + 0.5(\text{bohman}) \quad (3.4)$$

$$w2 = 0.5(\cosh) + 0.5(\cos^3(x)) \quad (3.5)$$

$$w3 = 0.5(\cosh) + 0.5(\cos^4(x)) \quad (3.6)$$

$$w4 = 0.5(\cosh) + 0.5(\text{hamming}) \quad (3.7)$$

$$w5 = 0.5(\cosh) + 0.5(\text{blackman}) \quad (3.8)$$

$$w6 = 0.5(\cosh) + 0.5(\text{optimized blackman}) \quad (3.9)$$

$$w7 = 0.5(\cosh) + 0.5(\text{triangular}) \quad (3.10)$$

$$w8 = 0.5(\cosh) + 0.5(\text{von-hann}) \quad (3.11)$$

$$w9 = 0.5(\cosh) + 0.5(\text{welch}) \quad (3.12)$$

Bu yöntem kapsamında kullanılan 9 farklı hibrit pencereleme yaklaşımı ile DNA dizilimlerinin ekson ve intron bölgelerine ait spektral yoğunluk bilgilerini ifade eden grafikler üretilmiştir.

Pencerelenmiş kısa zamanlı fourier dönüşümü

Kısa zamanlı fourier dönüşümü, fourier dönüşümünün eş zamanlı olarak üretilemeyen zaman ve frekans bilgisine ilişkin eksikliğin tolere edilebilmesi amacıyla geliştirilen bir yöntemdir. Gabor dönüşümü olarak tanımlanan kısa zamanlı fourier dönüşümünde analiz edilecek zaman serisi, zaman ve frekans fonksiyonu olarak 2 farklı boyutta ifade edilir. Ancak aynı anda iyi bir zaman ve frekans çözünürlüğü sunmaz [100].

Dönüşüm sırasında zaman serisi küçük parçalara bölünür ve bu parçaların durağan olduğu kabul edilir. Ardından bir zaman penceresi içerisinde fourier dönüşümünü uygulamak amacıyla zaman serisine ait küçük parçalar pencere fonksiyonu ile çarpılır. Böylece kısa zamanlı fourier dönüşümü yapılır [100].

Örnek bir zaman ekseninde K noktasına yerleştirilecek $w(t)$ pencere fonksiyonunun f frekansındaki fourier dönüşümünün matematiksel ifadesi denklem 3.13'te verilmiştir [100].

$$g_{f,t}(t) = e^{2\pi ift} w(t - K) \quad (3.13)$$

Pencerenin kaydırılması ve fourier dönüşümünün tekrar alınması ile işlem devam eder. Bu işlemin matematiksel ifadesi denklem 3.14'te verilmiştir [100].

$$KZFD(t,f) = \int_{-\infty}^{+\infty} x(t) w(t - K) e^{-2\pi ift} dt = \langle g_{f,t}(t), x(t) \rangle \quad (3.14)$$

Kısa zamanlı fourier dönüşümünde tercih edilecek pencere fonksiyonlarına ait mevcut enerjinin sonsuz olmamasından dolayı integrali alınır. Öte yandan pencere fonksiyonu ($w(t)$) dönüşüme ilişkin performansı belirler [100].

Şekil 3.11'de zaman serisine uygulanan kısa zamanlı fourier dönüşümü ve pencerelenmenin tasviri verilmiştir.



Şekil 3.11. Zaman serisine uygulanan kısa zamanlı fourier dönüşümü ve pencerelenme [100]

Zaman ve frekans bileşenleri arasında zaman serisine ilişkin bir uyumun yakalandığı kısa zamanlı fourier dönüşümü, frekans ve olay arasındaki bağlantının bilgisini sağlar. Bu durum fourier dönüşümüne kıyasla öne çıkan bir özelliktir. Ancak elde edilen bilgi pencerenin büyüklüğü ile de ilişkili olduğu için farklı pencere fonksiyonları kullanılarak elde edilen çıktıların çeşitliliği arttırılmıştır [100]. Bu doğrultuda bartlett, blackman, blackman-harris, bohman, chebwin, gausswin, hamming, rectangular ve triang pencereleme fonksiyonları kullanılmıştır.

Görüntülerin İyileştirilmesi

Devasa bilgi barındıran DNA genom parçası, protein üreten ve protein üretmeyen bölgelere sahiptir. Bu bölgelerin analiz edilmesi canlılıkla ilgili farklı birçok konunun aydınlatılmasını sağlamaktadır. Bu amaçla çalışmanın ilk aşamasında sayısallaştırılan DNA dizilimlerine sinyal işleme yöntemi olan fourier dönüşümü uygulanmıştır. Böylece ekson ve intron bölgelerinin spektral yoğunluk bilgisi grafikler ile ifade edilmiştir. Fakat görüntülerdeki kenar, köşe gibi kritik bölgelerin konumu önemli olduğu için bu kritik bölgelerin CLAHE (Contrast Limited Adaptive Histogram Equalization) yöntemi ile netleştirilmesi ve belirginleştirilmesi sağlanmıştır. Sonuç olarak güçlü çıkarımların üretilmesi hedeflenmiştir.

CLAHE yöntemi

Dijital görüntülerdeki kontrastı iyileştirmek için histogram tabanlı bir yaklaşım olan histogram eşitleme (HE), görüntünün tamamını içeren histogram dağılımı üzerinde işlem yaparak görüntünün genel kontrastını iyileştiren bir yöntemdir. Fakat homojen olmayan alanlarda görüntünün kontrastının düzeltilmesinde başarılı çıktılar verirken homojen bölgelerde gürültüyü arttırmaktadır. Bu nedenle uyarlanabilir histogram eşitleme (AHE) yöntemi, bahsedilen probleme bir çözüm olarak sunulmuştur. Bu yöntem ile görüntü çakışmayacak şekilde alt bölümlere ayrılmakta ve her bir parçanın histogram eşitleme yöntemi ile kontrastı iyileştirilmektedir. Ardından iyileştirilmiş her bir alt bloktan elde edilen sonuçlar birleştirilmektedir. Ancak AHE yöntemi için de yerel bölgelerdeki gürültü ve özelliklerin ayırt edilmesi zordur ve homojen bölgelerdeki gürültüler büyür. Homojen alanlardaki tüm bu olumsuz durumların önüne geçmek amacıyla kontrast sınırlı uyarlamalı histogram eşitleme yöntemi geliştirilmiştir. Kontrast iyileştirmenin sınırlandırıldığı CLAHE yönteminde görüntü parçalara ayrılır ve her bir parçanın histogramı elde edilir. Ardından histogram kırpma

limiti ile kırılan her bir parçanın kontrastı histogram eşitleme yöntemi ile geliştirilir ve tüm parçalar birleştirilir. Böylece görüntünün yerel kontrastı gürültünün etkisi arttırılmadan iyileştirilir [105].

Sınıflandırma süreci

Sinyal işleme yöntemleri kullanılarak frekans alanında ifade edilen görüntülerin analiz edilmesi ile potansiyel çözümler sunan çıktılar ilişkilendirilir ve elde edilen bilgiler anlamlandırılır. Bu çalışmada ilişkilendirme ve anlamlandırma süreci, yapay zeka tabanlı transfer öğrenme yöntemi ile gerçekleştirilmiştir. Böylece ekson ve intron bölgelerine ilişkin bir sınıflandırma sağlanmıştır.

Transfer Öğrenme

Bir görevin eğitiminde sıfırdan kazanılan bilgilerin farklı bir görev için yeniden kullanıldığı makine öğrenme tekniğidir [106]. Bu yöntemde yeni görevin hedef veri kümesine uyarlanması sürecinde eğitilen ağıın son birkaç katmanı kaldırılır ve model, yeni görev için uygun hale getirilir. Hedef veri kümesine uyarlanacak transfer öğrenme mimarileri sıfırdan üretilmediği için zaman ve donanım kısıtından dolayı modelin yeterince eğitilememesi problemine karşı bir çözüm üretilir [106, 107]. Ek olarak her bir transfer öğrenme mimarisi farklı bir yapıya sahiptir. Bu nedenle modelin eğitim aşamasında evrişim katmanları vasıtasıyla çıkarılan özelliklerin aynı olmamasından dolayı kazanılan bilgilerin hedef göreve uyarlandıktan sonra üreteceği başarı oranları da farklılık oluşturacaktır. Bu çalışmada 10 farklı transfer öğrenme mimarisi kullanılmış ve karşılaştırmalar yapılmıştır. Derinlik sayılarına göre çalışmada kullanılan bazı transfer öğrenme mimarileri Tablo 3.5'te verilmiştir.

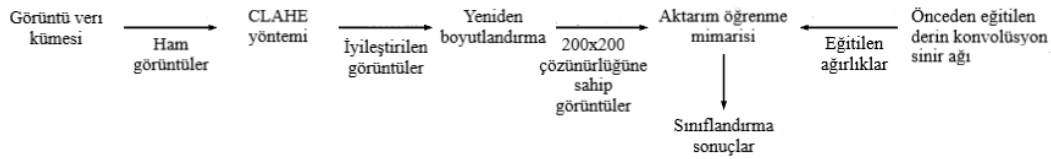
Tablo 3.5. Derinlik sayılarına göre bazı transfer öğrenme mimarileri

Transfer öğrenme Mimarileri	Derinlik Sayısı
VGG16	16
VGG19	19
Xception	71
MobileNetV2	53
InceptionResnetV2	164
DenseNet169	169
DenseNet201	201
Resnet101	101
Resnet150	150
EfficientNetB7	813

Tablo 3.5'te yer alan her bir transfer öğrenme mimarisinin sınıflandırma aşamasında 200x200 olarak yeniden boyutlandırılan görüntüler için kullandığı kayıp fonksiyonu, ikili çapraz entropidir. Matematiksel ifadesi denklem 3.15'te verilmektedir.

$$İÇE(h,t)=-(h*\log(t)+(1-h)*\log(1-t)) \quad (3.15)$$

Yukarıda verilen denklemde h, hedef vektörünü ve t, tahmin edilen sınıf vektörünü tanımlamaktadır [108]. Ekson ve intron bölgelerinin sınıflandırılması için çalışmada kullanılan transfer öğrenme mimarilerinin iş akışı Şekil 3.12'de sunulmuştur.



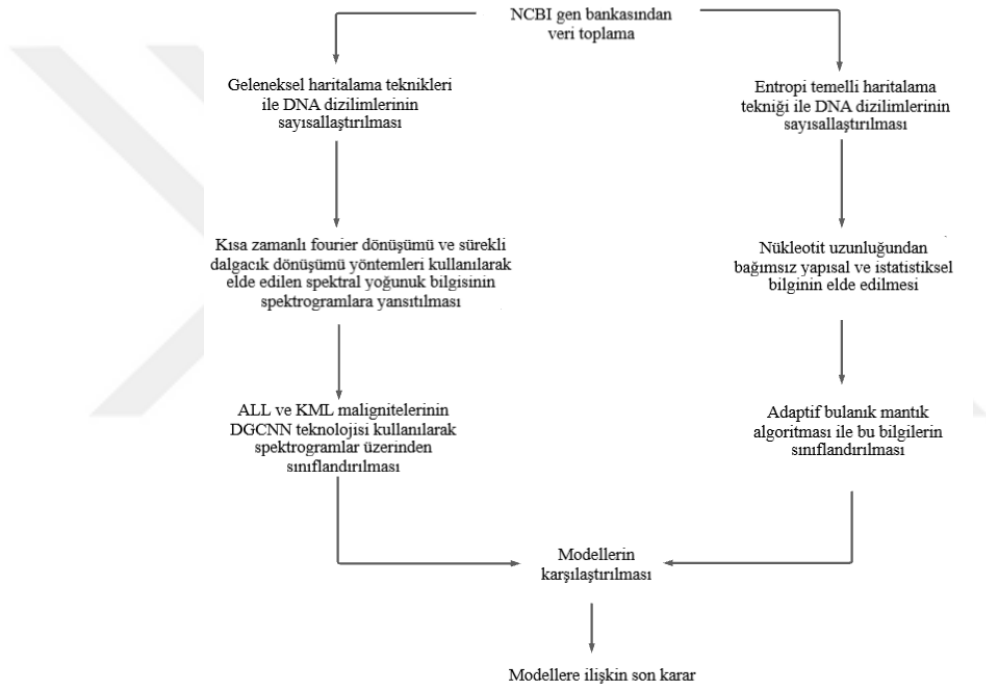
Şekil 3.12. Ekson ve intron bölgelerinin sınıflandırılması için çalışmada kullanılan transfer öğrenme mimarilerinin iş akışı [108]

Transfer öğrenme yönteminde kullanılan her bir mimarinin eğitimi, RMSprop optimize edicisinin 0.00007 öğrenme oranı, 0.7 momentum ve 1e-07 epsilon değeri ile sağlanır. İnce ayarı yapılan mimarilerin her bir yineleme sırasında kayıp değeri iyileştirilir. Böylece minimum kayıp değerine ve maksimum başarı oranına ulaşılır.

3.3.2. DNA Dizilimleri ile ALL ve KML Malignitelerinin Sınıflandırılmasına İlişkin İş Akışı

DNA canlılığın inşa edilmesine, geliştirilmesine ve sürdürülmesine ilişkin devasa bilgi barındıran hücre parçasıdır [9]. Sembolik bir yapıya sahip olan bu hücre parçası, DNA molekülünün analiz edilmesini zorlaştırmaktadır. Dolayısıyla çalışmanın ilk aşamasında DNA dizilimlerinin sayısallaştırılması sağlanmış ve ardından sayısal sinyal işleme yöntemi olan kısa zamanlı fourier dönüşümü ile dizilimlerin içerisindeki gizli bilgiler spektrogramlara yansıtılmıştır. Ancak kısa zamanlı fourier dönüşümü, geniş frekans aralıklarındaki sinyallerin analizi için uygun değildir [109]. Bu nedenle, daha başarılı bir analiz için sayısal sinyal işleme yöntemi olan sürekli dalgacık dönüşümü yöntemi de kullanılmıştır. Hem Kısa Zamanlı Fourier Dönüşümü hem de Sürekli Dalgacık Dönüşümü yöntemi vasıtasıyla spektrogramlara aktarılan spektral yoğunluk bilgisi, yapay zeka temelli sinir ağı mimarisi olan DGCNN yöntemi ile sınıflandırılmıştır.

Bu bölümün ikinci aşamasında, farklı nükleotit uzunluklarına sahip olan ALL ve KML malignitelerinin DNA baz uzunluğundan bağımsız sınıflandırma performansı araştırılmıştır. Bunun için sembolik DNA dizilimlerinin sayısallaştırılmasında DNA baz uzunluğundan bağımsız bir işlem uygulanmıştır. Bu doğrultuda, [46] çalışmasındaki kodon olasılıkları arasındaki korelasyonu ifade eden ve DNA dizilimlerinin sayısallaştırılmasında daha geniş bir aralık sunan [12] entropi temelli haritalama tekniği kullanılmıştır. Ardından sayısallaştırılan dizilimlerin istatistiksel, yapısal bilgileri çıkarılmış ve adaptif bulanık mantık algoritması ile sınıflandırılmıştır. Löseminin temel türleri olan ALL ve KML malignitelerinin sınıflandırılmasına ilişkin ikinci bölümde önerilen akış diyagramı Şekil 3.13'te sunulmuştur.



Şekil 3.13. Löseminin temel türleri olan ALL ve KML malignitelerinin sınıflandırılmasına ilişkin ikinci bölümde önerilen akış diyagramı

Bu bölümde, löseminin temel türlerinden olan ALL ve KML malignitelerinin sınıflandırılması için iki farklı aşamada gerçekleştirilen bir çalışma yapılmıştır.

3.3.2.1. DNA Dizilimleri ile ALL ve KML Malignitelerinin Sınıflandırılması için Kullanılan Yöntemler

Lösemi farklı karakteristik bulgular gösteren bir kanser oluşumudur. Tüm yaş gruplarında görülebilir ve hastalığın vücut içerisinde ilerleme durumuna göre kronik

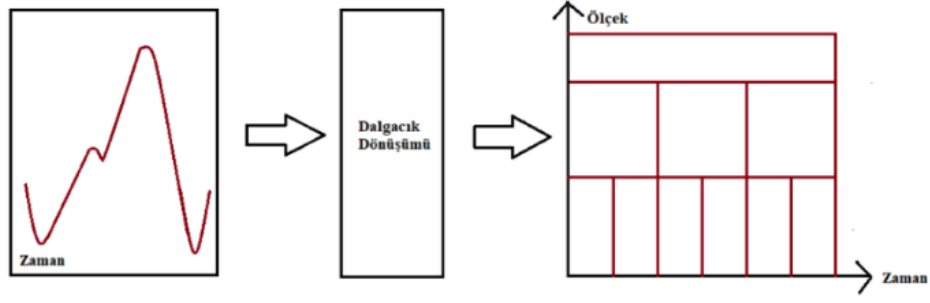
veya akut olarak farklılaşır. Aynı zamanda hücre kökenine göre miyeloid veya lenfoid şeklinde sınıflandırılır [4].

Bu bölümde lösemnin temel türlerinden olan akut lenfoblastik lösemi (ALL) ve kronik miyeloid lösemi (KML) malignitelerinin DNA dizilimleri üzerinden sınıflandırılması sağlanmıştır. Bu sınıflandırmada ALL ve KML lösemi türlerinin ayırt edilmesinde hayati bir ayıraç olan Philadelphia(Ph) kromozomu analiz edilmiştir [15].

Bu çalışma iki aşamadan oluşmuştur. İlk aşamada sayısal haritalama teknikleri (Reel Haritalama Tekniğı, Moleküler Kütle Haritalama Tekniğı, Tamsayı Haritalama Tekniğı, Karmaşık Haritalama Tekniğı, EIIP Haritalama Tekniğı, Atomik Sayı Haritalama Tekniğı, DNA-Yürüyüş Haritalama Tekniğı, Eşleştirilmiş Haritalama Tekniğinin Hidrojen Bağı Enerji Kuralı) ile dizilimler sayısallaştırılmıştır. Ardından sayısallaştırılan biyolojik verilerin analizi için genomik sinyal işleme yaklaşımı tercih edilmiştir. Bu yaklaşım kapsamında, Pencerelemiş Kısa Zamanlı Fourier Dönüşümü ve Sürekli Dalgacık Dönüşümü yöntemleri kullanılmıştır. Pencerelemiş Kısa Zamanlı Fourier Dönüşümü yönteminde tercih edilen pencereler hamming, gausswin, chebwin, blackman ve bartlett olarak belirlenmiştir. Pencerelemiş Kısa Zamanlı Fourier Dönüşümü için ayrıntılı açıklama 3.3.1.1. bölümünde verilmiştir.

Dalgacık Analizi

Morlet ve Grossmann tarafından tanıtılan dalgacıklar, sinyalin içerdiği bilgiyi farklı frekans bileşenlerine ayıran ve sonrasında belirli bir ölçeğe sahip bileşenler üzerinde çalışan matematiksel fonksiyonlardır. Gerçekleştirilen analizlerde dalganın ötelenmesi (zaman ekseninde dalganın kaydırılması) ve ölçeklenmesi (dalganın genişletip daraltılması) ile oluşturulan baz fonksiyonlarına işaretin izdüşümü alınır. Böylece farklı frekanslara sahip bileşenlerin zaman bilgisi elde edilir. Bu analizde tüm sinyalin bir pencere vasıtasıyla incelenmesinin ardından pencerenin kaydırılması ile sinyalin içerisindeki en küçük anların dahi bilgisi değerlendirilir. Dinamik sinyallerin incelenmesinde kullanılan dalgacık dönüşümünün tasviri Şekil 3.14'te verilmiştir [110-113].



Şekil 3.14. Dalgacık dönüşümünün gösterimi [110]

Şekil 3.14'te verilen zaman-frekans düzlemi incelendiğinde yüksek frekansların kısa süreli, düşük frekansların uzun süreli dalgacık ile belirtilmesi; güçlü bir zaman ve frekans çözünürlüğünü ifade etmektedir [110].

Öte yandan dalgacık dönüşümleri, biyosinyaller gibi zaman ve frekans bilgisinin önemli olduğu dinamik sinyallerin analizinde çoklu çözünürlük yaklaşımı ile başarılı bir değerlendirme sağlamaktadır. Dalgacık dönüşümleri sürekli ve ayrık alanda ifade edilmektedir. Bu çalışmada sürekli dalgacık dönüşümü yöntemi kullanılmıştır [56].

Sürekli dalgacık dönüşümü

Sürekli Dalgacık Dönüşümü, kaydırılan dalgacık fonksiyonunun belirli bir ölçek ile çarpılmasından sonra zaman alanı boyunca toplanması ile elde edilmektedir. Matematiksel ifadesi denklem 3.16'da verilmiştir [112].

$$SDD_{(s, \tau)} = \int_{-\infty}^{+\infty} g(t) \cdot \psi_{s, \tau}^*(t) dt \quad (3.16)$$

Denklem 3.16'da verilen $\psi_{s, \tau}(t)$, dalgacık fonksiyonunu; $g(t)$, dönüşümü yapılacak fonksiyonu; τ , kaydırma parametresini; s , belirlenen ölçeğin parametresini ve $*$, kompleks eşleniği ifade etmektedir. Sürekli dalgacık dönüşümü fonksiyonu, ana dalgacık fonksiyonu vasıtasıyla ölçek ve kaydırma faktörlerinin kullanılması ile hesaplanır. Matematiksel ifadesi denklem 3.17'de verilmiştir [112].

$$\psi_{s, \tau}^*(t) = \frac{1}{\sqrt{s}} \psi\left(\frac{t - \tau}{s}\right) \quad (3.17)$$

Denklem 3.17 de verilen $1/\sqrt{s}$ ifadesi, farklı ölçeklere sahip normalizasyon faktörü olarak ifade edilmektedir. Bu eşitliğin 3.16 denkleminde yerine konulması ile denklem 3.18 elde edilmektedir [112].

$$SDD_{(s, \tau)} = \frac{1}{\sqrt{s}} \int_{-\infty}^{+\infty} g(t) \cdot \psi_{s, \tau}^*\left(\frac{t - \tau}{s}\right) dt \quad (3.18)$$

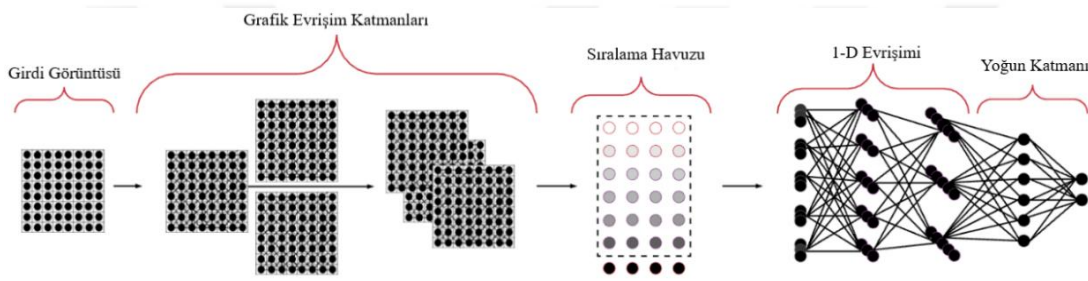
Seilen potansiyel genlere uygulanan srekli dalgacık dnřm sonucunda iliřkili genlerin spektral yoęunluk bilgisinin ifade edildięi spektrogramlar elde edilmiřtir.

Kısa zamanlı Fourier dnřm ve Srekli Dalgacık Dnřm yntemleri ile elde edilen spektrogramlar DGCNN yaklařımı ile sınıflandırılmıřtır.

Derin graf evriřimsel sinir aęları

DGCNN (deep graph convolutional neural network), spatial-temporal graf temelli bir yapıdır. Graf evriřim katmanları ile 1 boyutlu CNN yapılarının ard arda modellenmesi ile oluřur. Bu yapıda graf evriřim katmanları ile uzaysal baęımlılık ortaya ıkarılır. Ardından 1 boyutlu CNN katmanları ile zamansal baęımlılık yakalanır [114].

DGCNN mimarinin temsili yapısı řekil 3.15’de tasvir edilmektedir [115].



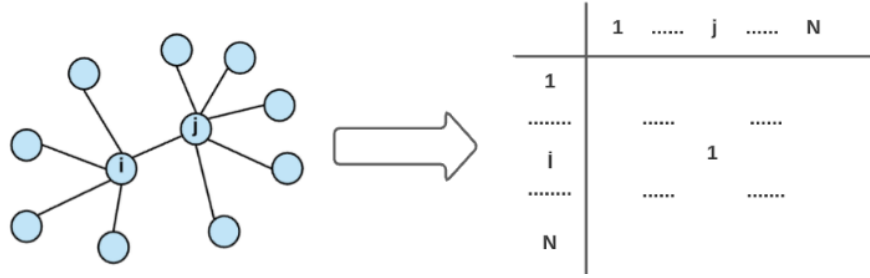
řekil 3.15. DGCNN mimarisinin temsili yapısı [115]

Bu blmde akut lsemilerin temel trlerinden olan ALL ve KML malignitelere ait spektrogramlar, DGCNN yntemi ile sınıflandırılmıřtır. İlk ařamada ALL ve KML malignitelere zg spektrogramlara SIFT (Scale Invariant Feature Transform) algoritması uygulanmıř ve grntlerdeki bireysel grnt dzeyi temsilleri (individual image-level representationlar(D)) aıęa ıkarılmıřtır. Ardından elde edilen bu temsillere k ortalama kmeleme yntemi uygulanmıřtır ve N tane kme merkezi ($X \in \mathbb{R}^{N \times D}$) belirlenmiřtir. Kme merkezleri korelasyonu grntlerin iliřkilerini tasvir eden bir baęıntıdır. Bu doęrultuda her bir grnt iin tasarlanan komřuluk matrisi $A \in \mathbb{R}^{N \times N}$, denklem 3.19’da verilmiřtir [116].

$$A_{mn} = \begin{cases} 1 & \text{if } X_m \in \text{KNN}(X_n) \vee X_n \in \text{KNN}(X_m) \\ 0 & \text{aksi durumda} \end{cases} \quad (3.19)$$

Burada V operandı “or” ve KNN () kosins benzerlięi zerinde temellenen en yakın k tane komřu anlamına gelir. A komřuluk matrisi ve grntnn sahip olduęu dęm

özellikleri tasarlanan 4 katmanlı GNN'e girdi olarak verilir. Şekil 3.16'da KNN temelli A komşuluk matrisinin temsili bir görseli bulunmaktadır [116].



Şekil 3.16. KNN temelli komşuluk matrisinin gösterimi [116]

Spektrogramların ifade edilmesinde kullanılan her bir piksel, düğüm olarak ifade edilir. Bu nedenle piksellerin çevresindeki pikseller ile bağlantılı olması, bilginin yayılmasını mümkün kılar. Bu durum mesaj geçiren sinir ağının (MPNN) varlığı ile açıklanmaktadır. Graf evrişim işleminde mesaj geçişi bir düğümden diğer bir düğüme iletilerek sağlanır ve bilgilerin daha fazla yayılması için mesaj geçişlerinin iterasyon sayısı bilgisi kullanılır. Graf evrişim işleminde gerçekleşen mesaj geçiş fonksiyonu denklem 3.20'de verilmiştir [114]

$$h_v^{(k)} = U_k(h_v^{(k-1)}, \sum_{u \in N(v)} M_k(h_v^{(k-1)}, h_u^{(k-1)}, x_{vu}^e)) \quad (3.20)$$

Denklem 3.20'de, $h_v^{(0)} = x_v$ ve $U_k(.)$, $M_k(.)$ öğrenilebilir parametreleri içeren fonksiyonları; k , iterasyon sayısını; v her bir pikseli ifade eden düğümü; u , komşuyu temsil etmektedir [114-116].

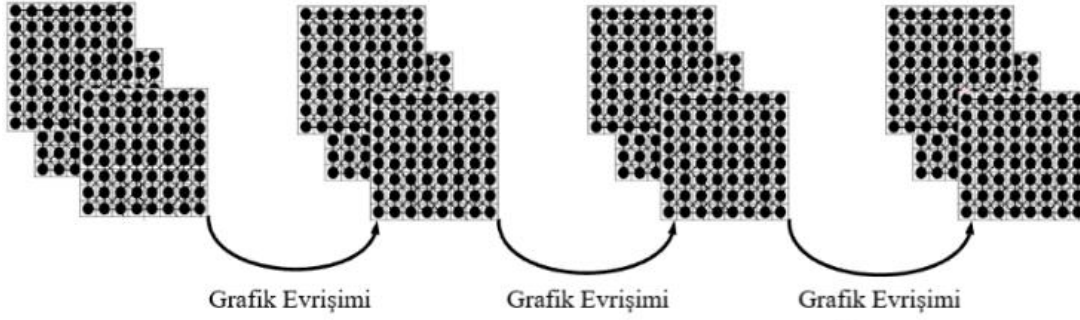
Bu kapsamda düğümlerin V olarak ifade edildiği bir görüntü $G=(V,E)$ için E , her bir düğüme karşılık gelen bağlantılara (V_i, V_j) sahiptir. Tüm düğümlerin ilişkisini içeren komşuluk matrisi $A \in V \times V$ olarak tanımlanır. Her bir düğümün özellik boyutu (D) vasıtasıyla görüntülerin kodlanmasını gerçekleştiren graf sinir ağı yaklaşımı $f(X,A)$, $X \in V \times D$ ile ifade edilmektedir. Denklem 3.21'de düğüm özelliklerinin güncellenmesini sağlayan kural verilmiştir.

$$H^{(l+1)} = \text{frelu}(\hat{A}H^{(l)}W^{(l)}) \quad (3.21)$$

Denklem 3.21'de verilen $\hat{A}$, komşuluk matrisinin normalize edilmiş temsidir ve komşuluk matrisi ile derece matrisi vasıtasıyla hesaplanır. Derece matrisinin matematiksel ifadesi denklem 3.22'de verilmiştir [116].

$$\delta_{ij} = \begin{cases} \deg(v_i) & \text{if } i = j \\ 0 & \text{otherwise} \end{cases} \quad (3.22)$$

F_{relu} çıkışa relu fonksiyonunun uygulandığını ifade eder. H^l , l'inci katmanın özelliklerini belirtir. $H^{(l)} \in |V| * D^l$, l'inci katmanın özellik temsilidir [116]. Şekil 3.17'de gösterilen 4 katmanlı GNN yapısında düğüm özelliklerinin güncellenmesini sağlamak amacıyla denklem 3.23-3.26'da gösterilen işlemler gerçekleştirilmektedir.



Şekil 3.17. 4 katmanlı graf evrişimsel sinir ağının gösterimi [116]

$$H^{(1)} = f_{\text{relu}}(\hat{A}XW^{(0)}) \quad (3.23)$$

$$H^{(2)} = f_{\text{relu}}(\hat{A}H^{(1)}W^{(1)}) \quad (3.24)$$

$$H^{(3)} = f_{\text{relu}}(\hat{A}H^{(2)}W^{(2)}) \quad (3.25)$$

$$H^{(4)} = f_{\text{relu}}(\hat{A}H^{(3)}W^{(3)}) \quad (3.26)$$

Bu yaklaşımda GNN yapısı ile girdinin tüm düğüm bilgileri komşu düğümler vasıtasıyla yayılır. Ardından köşe özelliklerinin sıralandığı sıralama havuzu (sort pooling) katmanına iletilir. Bu katmanda $n * \sum_1^h c_t (Z^{1:h})$ tensor yapısında olan girdi için her bir satır, köşenin özellik tanımlayıcısı ve her bir sütun, özellik kanalıdır. Sıralama havuzu katmanında girdi $Z^{1:h}$ 'e göre satır bazında sıralama gerçekleştirir. Aynı zamanda graf evrişim katmanlarındaki h değerlerinin yüksek bir sayı kullanılarak ifade edilmesiyle köşelere ince bir ayar verilir. DGCNN mimarisinde graf evrişim katmanının çıktıları sürekli WL renkleri (Z^t , $t=1,2,...,h$) olduğu için sıralama havuzu katmanı köşelerin sıralanmasında bu durumu kullanır. Böylece grafların yapısal rolüne göre köşelerin sıralanması gerçekleştirilir. Katmanın çıkışında elde edilen çıktı değeri, $k * \sum_1^h c_t$ boyutunda Z^{sp} tensörüdür. CNN yaklaşımını çıktı üzerinde gerçekleştirebilmek amacıyla ilk olarak Z^{sp} satır bazında bir $k(\sum_1^h c_t) \times 1$ vektörüne yeniden şekillendirilir ve havuzlama katmanından sonra CNN mimarisi eklenir. 2 katmanlı ve 1 boyutlu

geleneksel sinir ağı mimarisi ile tam bağlantılı katmanlar kullanılarak sıralanmış graf temsilleri üzerinden tahminler sağlanır. Burada belirtilen k değeri kullanıcı tanımlı (user-defined integer) bir değerdir [115].

RAR ve IIR bilgilerinin açığa çıkartılmasını sağlayan yapay zeka temelli DGCNN yaklaşımı, hem zamansal hem de uzaysal bilgi çıkarımı gerçekleştiren bir eğitim süreci sunar [115].

Bu bölümün ilk aşamasında geleneksel haritalama teknikleri ile sayısallaştırılan DNA dizilimleri, Kısa Zamanlı Fourier Dönüşümü ve Sürekli Dalgacık Dönüşümü yöntemleri ile spektrogram olarak ifade edilmiştir. Ardından DGCNN yöntemi ile sınıflandırılmıştır. Böylece DNA baz uzunluğuna bağlı olarak bir sonuç elde edilmiştir.

Bu bölümün ikinci aşamasında, DNA baz uzunluğundan bağımsız bir yapı inşa edilmiştir. Bu yapıda ilk olarak Shannon Entropi Temelli Haritalama Tekniği ile sayısallaştırılan DNA dizilimlerinden istatistiksel ve yapısal özellikler çıkarılmıştır. Ardından çıkarılan bu özellikler adaptif bulanık mantık algoritması ile sınıflandırılmıştır. Sayısallaştırılan DNA dizilimlerinden çıkarılan özellikler aşağıda açıklanmıştır.

Ortalama mutlak değer

Verilerin işaret genliğinin mutlak değer ortalamasıdır. Matematiksel ifadesi denklem 3.27’de verilmiştir [12]

$$\frac{1}{N} \sum_{i=1}^N |x_i| \quad (3.27)$$

Standart sapma

Verilerin nasıl dağıldığını ve ortalamadan ne kadar uzakta olduğunu açıklayan bir ölçüttür. Matematiksel ifadesi denklem 3.28’de verilmiştir [117].

$$S = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (x_i - \mu)^2} \quad (3.28)$$

Denklem 3.28’de μ , verinin ortalamasını ifade etmektedir.

Varyans

Verilerin aritmetik ortalamadan sapmalarının kareler toplamıdır. Matematiksel ifadesi denklem 3.29’da verilmiştir [117].

$$V = \frac{1}{N-1} \sum_{i=1}^N |x_i - \mu|^2 \quad (3.29)$$

Denklem 3.29’da μ , verinin ortalamasını ifade etmektedir.

Basit kare integral

Verinin enerjisini ifade etmek amacı ile kullanılan bir ölçüttür. Matematiksel ifadesi denklem 3.30’da verilmiştir [12].

$$\alpha = \sum_{i=1}^N |x_i|^2 \quad (3.30)$$

Dalga boyu

Zaman dilimi boyunca uzanan verinin kümülatif uzunluğunun ölçüsüdür. Genlik, frekans ve zaman terimleri ile ilişkilidir. Matematiksel ifadesi denklem 3.31’de verilmiştir [12].

$$\tau = \sum_{i=1}^{N-1} |x_{i+1} - x_i| \quad (3.31)$$

Kovaryans

İki rastgele değişkene ilişkin değişimlerin incelendiği bir ölçüttür. Matematiksel ifadesi denklem 3.32’de verilmiştir [117] [118].

$$\text{cov}(A,B) = \frac{1}{N-1} \sum_{i=1}^N |A_i - \mu_A| * |B_i - \mu_B| \quad (3.32)$$

Denklem 3.32’de μ , verinin ortalamasını ifade etmektedir.

Çarpıklık

Ortalama çevresindeki verilerin simetriden sapma oranını ifade eden asimetrisinin ölçüsüdür. Matematiksel ifadesi denklem 3.33’te verilmiştir [117].

$$S = E(x - \mu)^3 / \sigma^3 \quad (3.33)$$

Denklem 3.33’te μ , verinin ortalamasını; σ , verinin standart sapmasını ve E , beklenen değeri temsil etmektedir.

Basıklık

Dağılımın aykırı değere olan eğilimine ilişkin bir ölçüttür. Matematiksel ifadesi denklem 3.34’te verilmiştir [117].

$$k = E(x - \mu)^4 / \sigma^4 \quad (3.34)$$

Denklem 3.34’te μ , verinin ortalamasını; σ , verinin standart sapmasını ve E , beklenen değeri temsil etmektedir

Entropi

Düzensizlik derecesinin ölçülmesinde kullanılan bir ölçüttür. Matematiksel ifadesi denklem 3.35’te verilmiştir [12].

$$H(p) = -\sum_{i=1}^M p_i \log p_i \quad (3.35)$$

GMDH (Group method of data handling) ağları

Bir dizi çoklu giriş ve tek çıkış veri çiftlerinin performansını değerlendirmek amacıyla oluşturulan ileri beslemeli bir yöntemdir. Bu yöntemde girdi ve çıktı arasındaki ilişki Polinom Sinir Ağları şeklinde tanımlanmaktadır. Bu yapının matematiksel ifadesi denklem 3.36’da verilmiştir [119].

$$Y(x_1, \dots, x_n) = a_0 + \sum_{i=1}^n a_i x_i + \sum_{i=1}^n \sum_{j=i}^n a_{ij} x_i x_j + \sum_{i=1}^n \sum_{j=i}^n \sum_{k=j}^n a_{ijk} x_i x_j x_k + \dots \quad (3.36)$$

GMDH ağları, eğitim sürecinde kendi kendine organize olur ve yüksek öngörme özelliklerine sahiptir. Bu nedenle veri madenciliği, optimizasyon ve görüntü tanıma gibi birçok yapay zeka uygulamasında tercih edilen regresyon tabanlı bir yaklaşımdır [120].

Sayısallaştırılan DNA dizilimlerinden çıkarılan bu özellikler, adaptif bulanık mantık algoritması ile sınıflandırılmıştır.

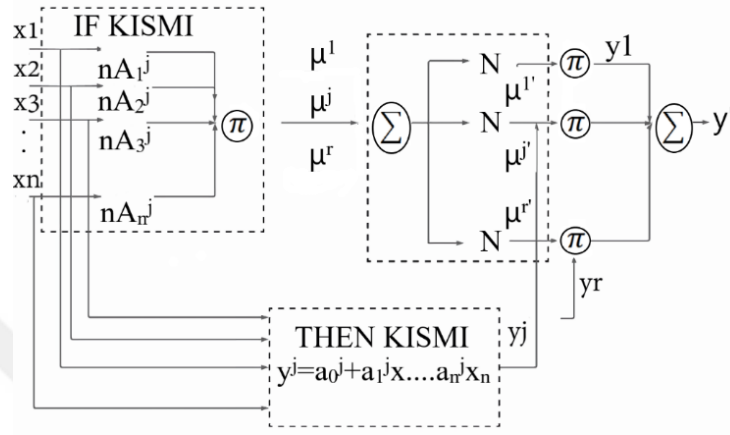
ANFIS (Adaptive neuro fuzzy inference system)

ANFIS bir yapay sinir ağı tekniğidir. Yapay sinir ağlarının öğrenme performansı ile bulanık mantık algoritmasının veriler üzerindeki madencilik yeteneğinin birlikte kullanılması ile verilerden çıkarım sağlar [121, 122]. IF THEN kuralı çerçevesinde çalışır ve bulanıklaştırıcı modülü ile seçilen üyelik fonksiyonuna göre (üçgen, yamuk, gauss, sigmoidal vb.) karakterize edilen bulanık kümeler net girdi modellerini eşler. Bununla birlikte nöro adaptif teknikler bulanık kural tabanlı sistemler için bir öğrenme stratejisi sağlayabilir. Bu durum girdi çıktı çiftlerinden uygun kuralların öğrenmesini ifade eder. Aynı zamanda ANFIS, girdi çıktı veri kümesinden üyelik fonksiyonları ile sonuç parametrelerini ayarlayabilme yetisine sahiptir. Matematiksel olarak ANFIS, denklem 3.37’de ifade edildiği gibi rastgele doğrusal fonksiyon içeren bir sonuç kısmı sunmaktadır [121].

$$R_i = \text{IF } x_1 \text{ is } A_1^j \text{ AND } x_2 \text{ is } A_2^j \text{ AND } \dots \text{AND } x_n \text{ is } A_n^j$$

$$\text{THEN } y^j = a_0^j + a_1^j x_1 + \dots + a_n^j x_n \quad (3.37)$$

Denklem 3.37’de verilen R^j , j’inci bulanık kuraldır. N-boyutlu girdi vektörünün k’ıncı girdi değişkeni x_k ’dır ve j’inci bulanık kural için x_k ile ilişkilendirilen bulanık üyelik fonksiyonu, A_k^j ’dir. Denklem 3.37’deki $a_0^j, \dots, a_n^j$, doğrusal katsayıları tanımlamaktadır. Şekil 3.18’de temsili bir ANFIS mimarisi verilmiştir [121].



Şekil 3.18. Temsili ANFIS yapısı [121]

Bu çalışmada girdi örüntülerini bulanık değerlere dönüştürmek için gauss üyelik fonksiyonu (GÜF) seçilmiştir. Bu yapının ilk katmanında GÜF, denklem 3.38’deki gibi bulanıklaştırılmaktadır [121].

$$\eta_{A_k^j} = \exp \left[-0.5 \left(\frac{x_k - c_k^j}{\sigma_k^j} \right)^2 \right] \quad (3.38)$$

Denklem 3.38’de verilen c_k^j ve σ_k^j , k’ıncı girdi değişkeni için j’inci GMF’nin merkezini ve genişliğini temsil eder. İkinci katman, ateşleme gücünü ve üçüncü katman, normalleştirilmiş çıktısını hesaplar. Öncül kısmında VE işlemi ile her bir kural için elde edilen çıktı denklem 3.39’da gösterilmektedir [121].

$$\mu^j = \prod_{k=1}^n \eta_{A_k^j} \quad (3.39)$$

j. kuralın normalleştirilmiş sonucu denklem 3.40’da verilmiştir.

$$\mu^j = \frac{\mu^j}{\sum_{j=1}^R \mu^j} \quad (3.40)$$

Denklem 3.40’da verilen μ^j , ateşleme gücüdür ve son ağ çıkışındaki her bir kural için katkı miktarına denklem 3.41’de gösterildiği gibi karar vermektedir [121].

$$y' = \sum_{j=1}^R \mu^i y_j \quad (3.41)$$

Sonuçların matematiksel hesabı dördüncü katmanda yapılır ve kesin çıktı fonksiyonu vasıtasıyla sonuçlanan kuralların sonuç bölümlerine doğrusal bir kombinasyon ile durulaştırma işlemi gerçekleştirilir [121].

Adaptif bulanık mantık algoritması kullanılarak istatistiksel ve yapısal bilgiler üzerinden gerçekleştirilen sınıflandırma eylemi ile DNA baz uzunluğundan bağımsız bir çıktı elde edilmiştir.

3.3.3. Mikrodizi Veri Kümesi Kullanılarak ALL ve AML Malignitelerinin Sınıflandırılmasına İlişkin İş Akışı

Kanser, kontrolsüz bölünebilme yetisi kazanmış hücrelerin vücutta oluşturdukları eylemler sonucunda ortaya çıkan bir malignitedir. Hastalığın tanı sürecinde kullanılan geleneksel yöntemler ile elde edilen parametreler nihai kararın verilmesi yönünde belirsizlikler barındırabilmektedir. Ancak genlerle ilişkili kanser hastalıkları için moleküler tanı imkanı sunan mikrodizi teknolojisi, kanser araştırmalarında bir umut kaynağı olmuştur [27]. Bu çalışmada, ALL ve AML maligniteleri mikrodizi teknolojisi ile sınıflandırılmıştır.

Bu çalışmada ilk olarak yüksek boyuta sahip mikrodizi veri kümelerindeki hastalıkla ilişkili potansiyel genler özellik çıkarım yöntemleri ile seçilmiştir. Ardından uygun genlere ilişkin spektral yoğunluk bilgisi, sinyal işleme yöntemi olan sürekli dalgacık dönüşümü yöntemi ile spektrogramlara yansıtılmıştır. Spektral yoğunluk bilgilerin netleştirilmesi amacıyla CLAHE yönteminin uygulandığı spektrogramlar, yapay zeka temelli DGCNN yaklaşımı ile sınıflandırılmıştır. Löseminin temel türleri olan ALL ve AML malignitelerinin sınıflandırılmasına ilişkin üçüncü bölümde önerilen akış diyagramı Şekil 3.19'da verilmiştir.

Bu bölümde löseminin temel türleri olan ALL ve AML malignitelerinin sınıflandırılması için bir çalışma yapılmıştır.

3.3.3.1. Mikrodizi Veri Kümesi Kullanılarak ALL ve AML Malignitelerinin Sınıflandırılması için kullanılan yöntemler

Canlı vücudu kendini onarabilen bir yapıya sahiptir. Ancak bazı durumlarda hücreler kontrollü bölünebilme yetilerini yitirir ve düzenli çalışan bu mekanizmanın yavaş yavaş bozulmasına neden olur. Bu aşamada doğru tanıya ve doğru tedavi protokolüne karar vermek önemlidir. Bu nedenle kanser hastalıklarının tanısında umut kaynağı olan ve moleküler tanı imkanı sunan mikrodizi teknolojisi ile akut lösemiler sınıflandırılmıştır [27]. Bununla birlikte mikrodizi veri kümelerinde genleri temsil eden özellikler yüksek boyutluluğa sahiptir [26]. Bu durum makine öğrenmesinde boyutluluğun laneti olarak ifade edilmektedir [27].

Bu bölümde ilk olarak yüksek boyutlu mikrodizi veri kümelerinin boyutu Parçacık Sürü, Karınca ve Balina özellik seçim algoritmaları ile indirgenmiştir. Ardından her bir özellik seçim algoritması için indirgenen mikrodizi veri kümesi ile akut lenfoblastik lösemi ve akut miyeloid lösemi maligniteleri sınıflandırılmıştır.

Özellik seçim yöntemleri

Parçacık sürü optimizasyon algoritması

PSO, çok boyutlu bir uzayda parçacık sürüsü ile optimum durumu bulmayı amaçlayan bir algoritmadır. Bu algortmada her bir parçacık, problem için olası bir çözümü temsil eder ve optimum çıktı üretme sürecine katkı sağlar. Algoritmanın ilk aşamasında parçacıklar çok boyutlu bir uzayda rastgele bir konuma atanır. Her bir parçacığın pozisyonu, kendi tecrübeleri ve komşularının tecrübelerine göre belirlenir. Her bir parçacık için konumu tasvir eden pozisyon vektörü tarafından optimal bir çözüm üretilir. İterasyon sayısı boyunca parçacıkların kendi deneyimleri ile diğer parçacıkların deneyimleri üzerinden bir arama davranışı oluşturulur. Bu yinelemeler sonucunda parçacıkların sunduğu optimum durum ile tarihsel bir yol inşa edilir [65, 123, 124].

Parçacıkların konum ve hız durumlarının güncellenmesine ilişkin matematiksel ifadeler, denklem 3.42 ve denklem 3.43’de verilmiştir.

$$V_i[t+1]=wv_i[t]+c_1r_1(X_{i,best}[t]-x_i[t])+c_2r_2(X_{g,best}[t]-x_i[t]) \quad (3.42)$$

$$X_i[t+1] = x_i[t] + v_i[t+1] \quad (3.43)$$

Denklem 3.42'deki $v_i[t]$ değişkeni, parçacığın mevcut durumunu belirten hız vektörüdür; w sabiti, parçacığın kendi yönündeki hareketinin katsayısıdır; $x_i[t]$ değişkeni, parçacığın mevcut durumunu belirten konum vektörüdür; $x_{i,best}[t]$ değişkeni, parçacığın geçmiş yinelemeler çerçevesinde elde ettiği en iyi konumdur; c_1 sabiti, parçacığın kendi geçmişinden gelen öğrenme katsayısıdır; $x_{g,best}[t]$ değişkeni, tüm parçacıklar kapsamında elde edilen en iyi konumdur; c_2 sabiti, parçacığın toplumun geçmişinden gelen öğrenme katsayısıdır; r_1 ve r_2 ise (0,1) aralığında atanan rastgele değerleri ifade eden değişkenlerdir. Çalışmada kullanılan algoritmanın sözde kodu aşağıda verilmiştir.

Algoritma 1: PSO Algoritması [123]

N boyutlu bir sürü oluştur ve başlat;

DO

FOR i FROM 1 TO n DO

if $f(x_{(i)}) < p_{best(i)}$ **then**

$(p_{best(i)} = x_{(i)});$

end

if $p_{best(i)} < g_{best(i)}$ **then**

$(g_{best} = p_{best(i)});$

end

if $g_{best(i)} > threshold_value$ **then**

$(selection_features_{(i)} = g_{best(i)});$

end

END FOR

FOR i FROM 1 TO n DO

Denklem 1 ile parçacığın hızını güncelle

Denklem 1 ile parçacığın pozisyonunu güncelle

END FOR

WHILE (maksimum iterasyona ulaşılan kadar)

Mikrodizi veri kümeleri üzerinde parçacık sürü optimizasyon yöntemi ile gerçekleştirilen özellik çıkarımı sonucunda potansiyel genler elde edilmiştir.

Balina optimizasyon algoritması

2016 yılında Seydali ve Andrew Lewis tarafından önerilen balina optimizasyon algoritması, kambur balinalara özgü olan kabarcık-ağ avlanma metodundan esinlenerek oluşturulmuştur. Bu yaklaşım avın kuşatılması, ava doğru ilerleme ve av arama şeklinde balinalara özgü olan 3 temel avlanma stratejisinden oluşmaktadır [125].

İlk aşamada algoritma tarafından en iyi çözüm av ya da ava en yakın çözüm olarak kabul edilir Ardından en iyi arama ajanına göre diğer arama ajanlarının konumu güncellenir. Bu süreçle ilişkin matematiksel ifadeler denklem 3.44 ve 3.45'te verilmiştir [125].

$$\vec{D} = |\vec{C}\vec{X}^*(t) - \vec{X}(t)| \quad (3.44)$$

$$\vec{X}(t+1) = |\vec{X}^*(t) - \vec{A}\vec{D}| \quad (3.45)$$

$\vec{X}$, arama ajanına ilişkin konum vektörüdür; X^* , optimum konum vektörüdür. Optimum konum vektörü her iterasyon sonunda incelenir ve daha iyi bir çözümün elde edilmesi için güncellenir [125]. $\vec{A}$ ve $\vec{C}$, katsayı vektörleridir. Bu vektörlerin hesaplanabilmesi için kullanılan matematiksel ifadeler denklem 3.46 ve 3.47 'de verilmiştir [125].

$$\vec{A} = 2\vec{a}\vec{r} - \vec{a} \quad (3.46)$$

$$\vec{C} = 2\vec{r} \quad (3.47)$$

Denklem 3.46 ve 3.47'de verilen r vektörü, $[0,1]$ aralığında atanan rastgele bir değerdir. Denklem 3.46'da verilen a vektörü, iterasyon boyunca 2'den 0'a doğru azalan bir değerdir [125].

Ava doğru hareket eylemi, kabarcık ağ yaklaşımı için 2 farklı matematiksel model sunmaktadır. İlk davranış modeli, daralan kuşatma mekanizmasıdır. Bu eylem, denklem 3.46’da yer alan $\vec{a}$ vektörünün azalması ile gerçekleşir. Diğer davranış modeli, spiral harekettir. Spiral hareket modelinde balinanın (X,Y) şeklindeki konumu ile avın (X^*,Y^*) konumu arasındaki mesafe hesaplanır. Ardından kambur balinaların sarmal hareketi, denklem 3.48’de gösterildiği gibi modellenir [125].

$$\vec{X}(t+1) = \vec{D}' \cdot e^{bl} \cdot \cos(2\pi l) + \vec{X}^*(t) \quad (3.48)$$

Denklem 3.48’de verilen l, [-1,1] aralığında rastgele atanan bir sayıyı; b, spiral modelin belirlenebilmesi için kullanılan sabiti ifade etmektedir. $\vec{D}$ vektörü, balina ile av arasında elde edilen optimum çözümdür. $\vec{D}$ vektörünün matematiksel ifadesi denklem 3.49’da verilmiştir [125].

$$\vec{D}' = \vec{X}^*(t) - \vec{X}(t) \quad (3.49)$$

Kabarcık ağ metoduna göre kambur balinaların hem daralan kuşatma mekanizması hem de spiral hareket modeline göre eş zamanlı olarak yüzme davranışı 2 farklı sürecin seçilmesine ilişkin $\frac{1}{2}$ ‘lik olasılık meydana getirir. Bu harekete ilişkin matematiksel model denklem 3.50’de verilmiştir [125].

$$\begin{aligned} \vec{X}(t+1) &= \{ \vec{X}^*(t) - \vec{A}\vec{D}, & p < 0,5 \\ \vec{D}' \cdot e^{bl} \cdot \cos(2\pi l) + \vec{X}^*(t) \} & & p \geq 0,5 \end{aligned} \quad (3.50)$$

Kambur balinaların kabarcık ağ metodu dışında faydalandıkları av arama davranışı, rastgele avlanma biçimidir. Kabarcık ağ modelinden farklı olarak bu davranış modelinde arama ajanı rastgele seçilen bir arama ajanıdır. $\vec{A}$ vektörünün 1’den büyük olması durumunda balina optimizasyon algoritması global arama yapar. Bu mekanizmaya ilişkin matematiksel model denklem 3.51 ve 3.52’de verilmiştir [125].

$$\vec{D} = |\vec{C} \cdot \vec{X}_{rand} - \vec{X}| \quad (3.51)$$

$$\vec{X}(t+1) = \vec{X}_{rand} - \vec{A} \cdot \vec{D} \quad (3.52)$$

Denklemlerde verilen X_{rand} değişkeni, rastgele atanan bir çözüm vektörüdür. $\vec{A}$ vektörü aldığı değere göre global aramanın ya da yerel aramanın yapılacağına karar verir [126-127].

Mikrodizi veri kümeleri üzerinde balina optimizasyon yöntemi ile gerçekleştirilen özellik çıkarımı sonucunda potansiyel genler elde edilmiştir.

Karınca kolonisi optimizasyon algoritması

Karıncaların koloni şeklinde yiyecek arama davranışından esinlenerek geliştirilen metasezgisel optimizasyon algoritmasıdır. Temel amaç karıncanın yuvası ile yemek kaynağı arasındaki yolun optimum olarak belirlenmesidir. Algoritma 3 temel aşamaya sahiptir [128].

1-) Karıncaların çözüm yollarını oluşturmak [128].

2-) Karıncaların belirlenen yollardan geçerken en uygun mesafeyi bulmak amacı ile salgıladığı feromon maddesini buharlaştırarak algoritmanın tek bir çözüme yakınsamasını sağlamak [128].

3-) Adım 2'deki yerel bilgiler ile adım 3'teki global bilgiler arasındaki farkın dikkate alınması koşuluyla global bilgiye dayalı olarak karar vermek [128].

Bu 3 temel aşama yakınsamanın gerçekleşmesi ya da belirlenen iterasyon sayısına ulaşılmasına kadar devam edecektir [128].

Karınca algoritmasında k'ncı karıncanın kaynak noktası olan i düğümünden hedef noktası olan j düğümüne hareket ederken gideceği yolu seçmek amacı ile kullandığı kuralın matematiksel modeli denklem 3.53'de verilmiştir [129].

$$P_{i,j} = \{ \operatorname{argmax} l \in N_i \{t_{il} [n_{il}]^b\}, \text{ if } q \leq q_0; \\ J, \text{ otherwise } \} \quad (3.53)$$

Burada q , $[0,1]$ aralığında tekdüze (uniform) dağılan rastgele bir değişkendir. q_0 , $(0 \leq q_0 \leq 1)$ aralığındaki bir değişkendir. J , olasılık dağılımına göre rastgele seçilen bir değerdir. Olasılık tabanlı matematiksel model, denklem 3.54'te verilmiştir [129].

$$J = [t_{i,j}]^a [n_{i,j}]^b / \sum [t_{i,l}]^a [n_{i,l}]^b, \text{ if } j \in N_i \text{ and } l \in N_i \quad (3.54)$$

Denklem 3.54'te n , iki nokta arasındaki maliyeti; t , iki nokta arasındaki feromon seviyesini; a , feromon seviyesinin formüle etkisini ve b , maliyetin formüle etkisini ölçer [129].

Karınca kolonisi optimizasyon algoritmasında tüm karıncaların kullandığı buharlaşma stratejisi ile yerel feromon izi güncellenir. Ardından en iyi performans gösteren karınca

tarafından feromon eklenmesi ile global feromon izi güncellenir. Yerel ve global feromon maddesinin güncellenmesine ilişkin matematiksel modeller sırasıyla denklem 3.55 ve 3.56’da verilmiştir [129].

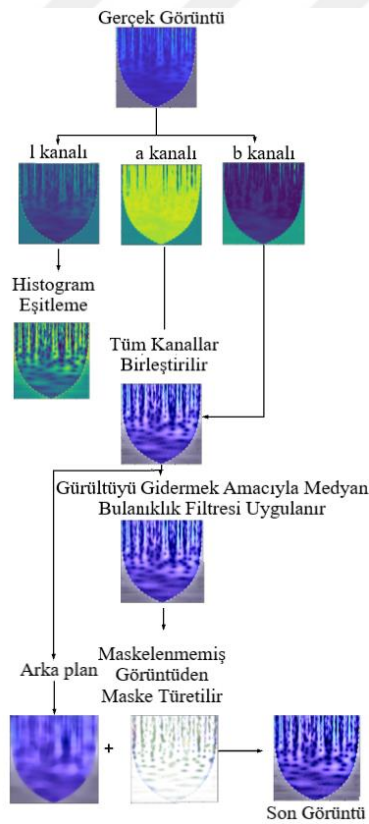
$$t_{ij} = (1-e) t_{ij} + e t_0 \quad (3.55)$$

$$(t_{i,j}) = (1-r) (t_{i,j}) + r \Delta t_{i,j}^{bs}, \quad \forall (i,j) \in T^{bs} \quad (3.56)$$

Verilen denklemlerde e , $0 < e < 1$ aralığındaki değişkeni; t_0 , feromon izinin başlangıç değerini ve r , buharlaşma oranını gösteren bir sabiti ifade etmektedir [129].

Mikrodizi veri kümeleri üzerinde karınca optimizasyon algoritması ile gerçekleştirilen özellik çıkarımı sonucunda potansiyel genler elde edilmiştir.

3 farklı özellik seçim algoritması ile elde edilen hastalıkla ilişkili potansiyel genlerin spektral yoğunluk bilgisi sürekli dalgacık dönüşümü yöntemi ile spektrogramlara yansıtılmıştır. Ardından görüntünün yerel kontrastı gürültünün etkisi arttırılmadan CLAHE yöntemi ile iyileştirilmiştir. Örnek bir spektrogram üzerinde CLAHE yönteminin uygulanma aşamaları Şekil 3.20’de verilmiştir.



Şekil 3.20. Örnek bir spektrogram üzerinde CLAHE yönteminin uygulanma aşamaları

CLAHE yöntemi ile iyileştirilen spektrogramlar, transfer öğrenme mimarileri ve DGCNN yaklaşımı ile sınıflandırılmıştır.

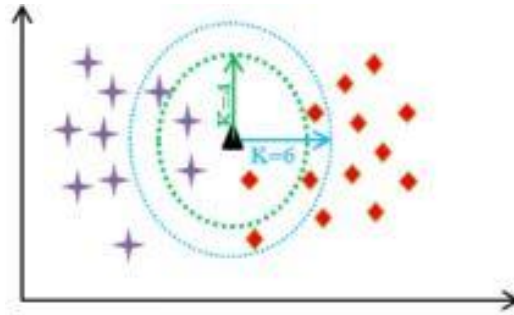
Ek olarak özellik çıkarım aşamasından sonra ulaşılan potansiyel genlere uygulanan klasik makine öğrenme teknikleri ile bir kıyaslama süreci daha gerçekleştirilmiştir.

Klasik makine öğrenmesi algoritmaları

Makine Öğrenmesi, hedef görev doğrultusunda mevcut veriler için bir matematiksel model üzerinde temellenen çerçevedir. Makine öğrenmesi denetimli, denetimsiz ve pekiştirmeli öğrenme yaklaşımlarına sahiptir. Bu yaklaşımlar ile veriler içerisindeki örüntülerin ortaya çıkarılmasını hedefler [130]. Bu çalışmada, denetimli öğrenme yaklaşımı ile ALL ve AML maligniteleri sınıflandırılmıştır. Bu çalışma için kullanılan sınıflandırma algoritmaları K En Yakın Komşu Algoritması (KNN), Karar Ağacı Algoritması, Rastgele Orman Algoritması ve Destek Vektör Makineleri Algoritmasıdır.

K en yakın komşu (KNN) algoritması

Bu algoritma, en yakın k tane komşu örneği ile parametrik olmayan bir sınıflandırma sistemi inşa eder. Mevcut tüm durumlar arasından benzerlik ölçüsüne göre yeni durumların sınıflandırılmasını sağlar. Basit ve en eski sınıflandırma algoritmaları arasında değerlendirilir. K en yakın komşu algoritmasının çalışma akışı Şekil 3.21’de verilmiştir [130].



Şekil 3.21. K en yakın komşu algoritmasının çalışma akışı [130].

Şekil 3.21’de verilen örnekte, hedef nesnenin sınıflandırılması için en yakın komşu sayısı 4 ve 6 olarak belirlenmiştir. Sınıflandırma işlemi sonucunda komşu sayısı 4 olan bir hesap için yeni nesnenin çıktısı yıldız; komşu sayısı 6 olan bir hesap için yeni nesnenin çıktısı kare olarak tasvir edilmiştir [130].

Karar ağacı algoritması

Karar Ağacı algoritması tümevarım mantığını kullanan basit ve anlaşılır bir yöntemdir. Ayrık değerler ile çalışabilir. Bu durum algoritmaya gürültüye karşı dayanıklı olma yetisi kazandırır. Karar ağacı algoritması altında benzer yaklaşımlar üzerinde temellenen ID3, C4.5 ve CART yöntemleri bulunmaktadır. Bu çalışmada GINI(CART) algoritması kullanılmıştır [131].

Gini (CART) algoritması

Gini, hedef ağacın karar düğümünden itibaren iki ayrı dala ayrılmasını temsil eden bir algoritmadır. Ağacın bölünme değeri ve bölüneceği nitelik seçimi gini indeks değeri sonucunda belirlenmektedir. Gini indeks değeri sol ve sağ gini değerlerinin hesaplanması ile bulunur. $Gini_{sol}$ ve $Gini_{sağ}$ ölçütlerinin matematiksel ifadeleri denklem 3.57 ve 3.58’te verilmiştir [132].

$$Gini_{sol} = 1 - \sum_{i=1}^k \left[\frac{L_i}{|T_{sol}|} \right]^2 \quad (3.57)$$

$$Gini_{sağ} = 1 - \sum_{i=1}^k \left[\frac{R_i}{|T_{sağ}|} \right]^2 \quad (3.58)$$

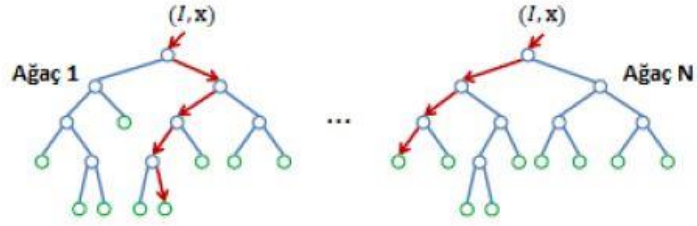
Denklem 3.57 ve denklem 3.58’de k, sınıfların sayısı; T_{sol} , ağacın sol kolundaki örneklerin sayısı; $T_{sağ}$, ağacın sağ kolundaki örneklerin sayısı; L_i , ağacın sol kolunun i kategorisinde yer alan örneklerin sayısı; R_i , ağacın sağ kolunun i kategorisinde yer alan örneklerin sayısı olarak karakterize edilmektedir. Sağ ve sol gini çıktılarından yararlanarak bir nitelik için hesaplanan gini değeri, denklem 3.59’da verilmiştir [132].

$$Gini_j = \frac{1}{n} (|T_{sol}| Gini_{sol} + |T_{sağ}| Gini_{sağ}) \quad (3.59)$$

Her bir nitelik için hesaplanan gini değerleri içerisinde en küçük değere sahip olan nitelik üzerinden bölünme gerçekleşir. Kalan veri kümesi için gerçekleştirilecek diğer bölünmede ilgili adımlar yeniden uygulanır ve sınıflandırma işlemi yapılır [132].

Rastgele orman algoritması

Karar Ağaçları topluluğundan meydana gelen sınıflandırma algoritmasıdır. Model oluşturma sürecinin kısa bir zaman diliminde gerçekleşmesi, basit, hızlı ve esnek olması algoritmanın birçok alanda kullanılmasını sağlar. Bu çalışmada ağaçların üretilmesi için CART algoritması tercih edilmiştir. Şekil 3.22’de N adet karar ağacının bir araya gelmesi ile oluşan bir orman yapısı yer almaktadır [131].

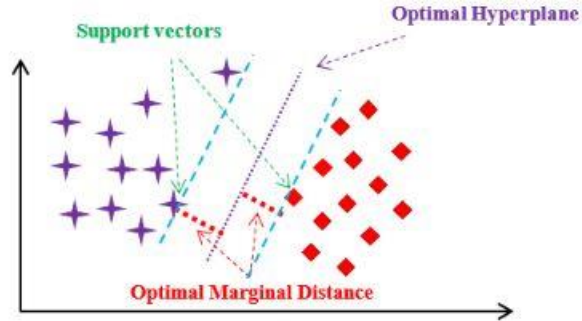


Şekil 3.22. N adet karar ağacının bir araya gelmesi ile oluşan bir orman yapısı [131]

Rastgele orman algoritması ile girdi veri kümesindeki bir nesnenin sınıflandırılması hedeflendiğinde ilgili girdi, ormanda inşa edilen tüm ağaçlara uygulanır. Ardından mevcut tüm ağaçlar ile bir sınıflandırma sonucu elde edilir ve ağaçlar arasında bir oylama işlemi gerçekleştirilir. Son aşamada en çok oyu alan sınıfa atama yapılır [131].

Destek vektör makineleri algoritması

İstatistiksel öğrenme teorisine dayalı, hassas sonuçlar üreten etkili bir algoritmadır. Bir hiper düzlem vasıtasıyla verilerin sınıflandırılmasını sağlar. SVM algoritması, minimum sınıflandırma hatasına ve maksimum marjinal mesafeye sahip bir karar düzlemi inşa eder. Mesafelerine göre doğrusal ayrılabilen veriler için Şekil 3.23'te temsili destek vektör makineleri algoritmasının çalışma hiyerarşisi verilmiştir [131].



Şekil 3.23. Temsili destek vektör makineleri algoritmasının çalışma hiyerarşisi [130].

Şekil 3.23'te gösterilen destek vektörleri, sınıflandırmaya ilişkin bilgi veren ölçütlerdir. Hiper düzlem ve destek vektörlerine ilişkin matematiksel ifadeler sırasıyla denklem 3.60 ve denklem 3.61'de verilmiştir [131].

$$W.X+b=0 \quad (3.60)$$

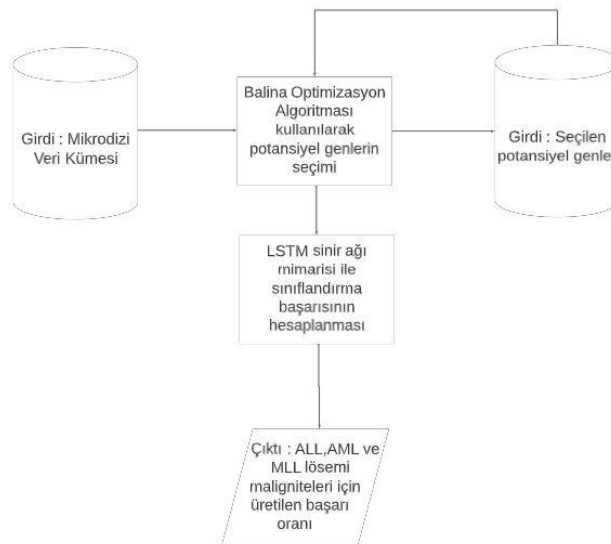
$$W.X+b=\pm 1 \quad (3.61)$$

Denklem 3.60 ve 3.61 'de ifade edilen W, ağırlık vektörünü; $\{-1,+1\}$, sınıf etiketlerini; b, eğilim değerini ve X, N boyutlu bir uzayı ifade etmektedir [131].

Bu bölümde mikrodizi veri kümelerinin boyutu Parçacık Sürü, Karınca ve Balina özellik seçim algoritmaları ile indirgenmiştir. Ardından indirgenen mikrodizi veri kümeleri ile akut lenfoblastik lösemi ve akut miyeloid lösemi maligniteleri sınıflandırılmıştır.

3.3.4. Mikrodizi Veri Kümesi Kullanılarak ALL, AML ve MLL Malignitelerinin Sınıflandırılmasına İlişkin İş Akışı

Genetik ya da çevresel etkenler sonucunda meydana gelen lösemi hastalığı, tedavi edilmediği takdirde ölümle sonuçlanabilen bir malignitedir. Hastalık sürecinde genetik anormallikleri tanımlamak, lösemi türlerinin ayırt edilmesi ve tedavi protokolünün belirlenmesi için önemli bir faktördür. Hastalığın tanısı için moleküler keşif imkanı tanıyan mikrodizi teknolojisi, genler arası ilişkilerin ortaya çıkarılmasında ve anormal durumların ifade edilmesinde tercih edilmektedir [27]. Bu çalışmada, ALL, AML ve MLL malignitelerinin sınıflandırılmasında yüksek boyut sunan mikrodizi veri kümeleri kullanılmıştır. Ancak mikrodizi veri kümeleri üzerinde hesaplama karmaşıklığının indirgenmesi ve seçilen potansiyel genler ile başarılı çıkarımların elde edilmesi için balina optimizasyon algoritmasından faydalanılmıştır. Ardından seçilen uygun genlerin birbiri ile etkileşiminin değerlendirilmesi için LSTM sinir ağı mimarisi kullanılmıştır. Löseminin temel türleri olan ALL, AML ve MLL malignitelerinin sınıflandırılmasına ilişkin dördüncü bölümde önerilen akış şeması Şekil 3.24'te verilmiştir.



Şekil 3.24. Löseminin temel türleri olan ALL, AML ve MLL malignitelerinin sınıflandırılmasına ilişkin dördüncü bölümde önerilen akış şeması.

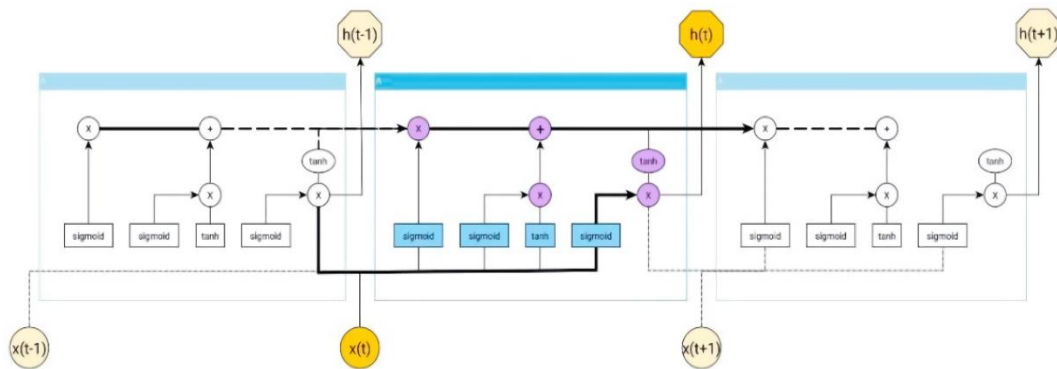
Bu bölümde löseminin temel türleri olan ALL, AML ve MLL malignitelerinin sınıflandırılması için bir çalışma yapılmıştır.

3.3.4.1. Mikrodizi Veri Kümesi Kullanılarak ALL, AML ve MLL Malignitelerinin Sınıflandırılması için kullanılan yöntemler

Tüm lösemi türlerinin yaklaşık %10'luk bir kısmını oluşturan MLL translokasyonları hem ALL hem de AML hücrelerinde bulunma ihtimali olan yapılardır. Bu yapıların doğru bir şekilde sınıflandırılması tedavi protokolünün belirlenmesinde hayati bir konudur. Bu çalışmada, ALL, AML ve MLL mikrodizi veri kümelerinden Balina Optimizasyon Algoritması (3.3.3.1. kısmında Balina Optimizasyon Algoritması açıklanmıştır) ile potansiyel genler seçilmiştir. Ardından seçilen bu genler LSTM sinir ağı ile sınıflandırılmıştır.

LSTM (Long-Short Term Memory) sinir ağı

LSTM, uzun bir menzil boyunca bilgileri depolayan sinir ağıdır. Giriş kapısı, giriş modülasyon kapısı, unutma kapısı ve çıkış kapısı bu sinir ağındaki 4 önemli kısımdır. Giriş kapısı, dışarıdan gelen yeni verileri alır ve işler. Bellek hücresi giriş kapısı, son iterasyonda LSTM sinir ağı hücresinin çıkışından girdiyi alır. Unutma kapısı, çıkış sonuçlarının ne zaman unutulacağına karar verir. Aynı zamanda hücrenin önceki durumunun hatırlanması veya unutulması da bu kapı ile sağlanır. Çıkış kapısı, tüm sonuçları alarak LSTM sinir ağı için çıkış üretme eylemini gerçekleştirir ve bellek hücresinin çıkışlar üzerindeki etkisini belirleyebilir. LSTM sinir ağının temsili yapısı Şekil 3.25'de verilmiştir [133].



Şekil 3.25. LSTM sinir ağının temsili yapısı [133].

Girdi zaman serisi, gizli durum hücreleri ve çıkış dizisi sırasıyla $X=(xx, xx2, \dots, xxTT)$, $HH=(h1, h2, \dots, hTT)$, ve $YY=(yy1, yy2, \dots, yyTT)$ olan LSTM sinir ağıları için hesaplama işlemleri denklem 3.62 ve 3.63'te verilmiştir.

$$h_t = H(W_{hy}x_t + W_{hh}h_{t-1} + b_h) \quad (3.62)$$

$$y_t = W_{hy}h_t + b_y \quad (3.63)$$

Denklem 3.62. ve denklem 3.63'te tasvir edilen LSTM yapısı, denklem 3.64 - 3.68'de verilen fonksiyonlar ile uygulanır ($t=1, \dots, T$ için).

$$i_t = \varphi(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i) \quad (3.64)$$

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f) \quad (3.65)$$

$$c_t = f_t * c_{t-1} + i_t * \tan h(W_{xc}x_t + W_{hc}h_{t-1} + b_c) \quad (3.66)$$

$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_t + b_o) \quad (3.67)$$

$$h_t = o_t \tan h(c_t) \quad (3.68)$$

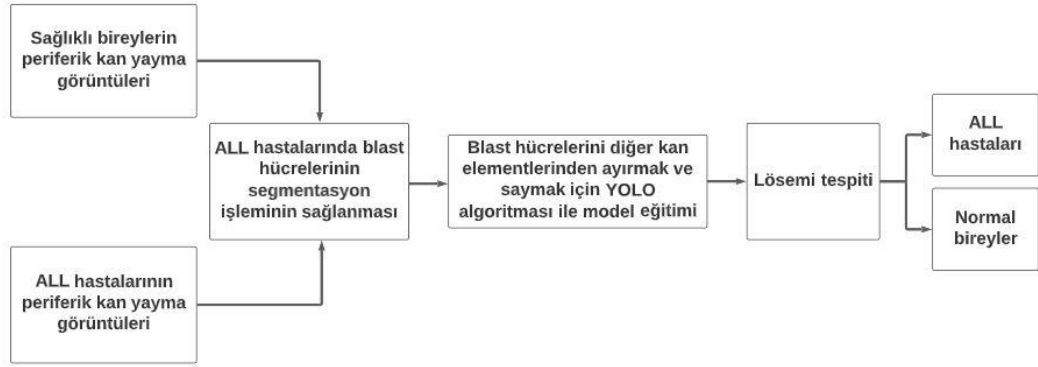
Denklemlerdeki σ , standart sigmoid fonksiyonunu temsil eder ve σ ile $\tanh$, eleman bazında uygulanan aktivasyon fonksiyonlarıdır. Giriş kapısı, i ; unutma kapısı, f ; çıkış kapısı, o ve hücre aktivasyon vektörleri, c ile karakterize edilmektedir. W , ağırlık matrisini temsil etmektedir [133].

LSTM, genellikle metin analizinde ve zaman serilerinin tahmininde kullanılır. Ancak tıp alanında üretilecek karar destek sistemleri için de sahip olduğu fiziksel üniteler yardımıyla başarılı çıkarımlar sağlamaktadır.

3.3.5. Periferik kan yayma görüntüleri üzerinde YOLOv4 algoritması kullanılarak ALL hastalığının erken teşhisine ilişkin iş akışı

Lösemnin temel türlerinden biri olan ALL, kan dolaşımına ve hayati organlara hızlı bir şekilde yayıldığı için erken teşhisi önemli olan bir hastalıktır. Bu hastalığın tanı ve tedavi sürecinin öngörülmesinde kullanılan periferik kan yayma yöntemi ile olgunlaşmamış beyaz kan hücreleri tespit edilir ve kandaki yoğunluğu incelenir. Ardından elde edilen sonuçlar ile bir değerlendirme yapılır. Bu çalışmada ALL hastalığının tanısında ya da tedaviye verilen yanıtın değerlendirilmesinde doktorlara yardımcı bir fikir sunmak amacıyla YOLOv4 algoritması kullanılarak beyaz kan hücrelerinin değerlendirildiği bilgisayar destekli bir sistem inşa edilmiştir. Lösemnin

temel türü olan ALL malignitesinin sınıflandırılmasına ilişkin beşinci bölümde önerilen akış diyagramı şekil 3.26’da verilmiştir.



Şekil 3.26. Löseminin temel türü olan ALL malignitesinin sınıflandırılmasına ilişkin beşinci bölümde önerilen akış diyagramı

Bu bölümde löseminin temel türü olan ALL malignitesinin değerlendirilmesine ilişkin bir çalışma yapılmıştır.

3.3.5.1. Periferik kan yayma görüntüleri üzerinde YOLOv4 algoritması kullanılarak ALL hastalığının erken teşhisi için kullanılan yöntemler

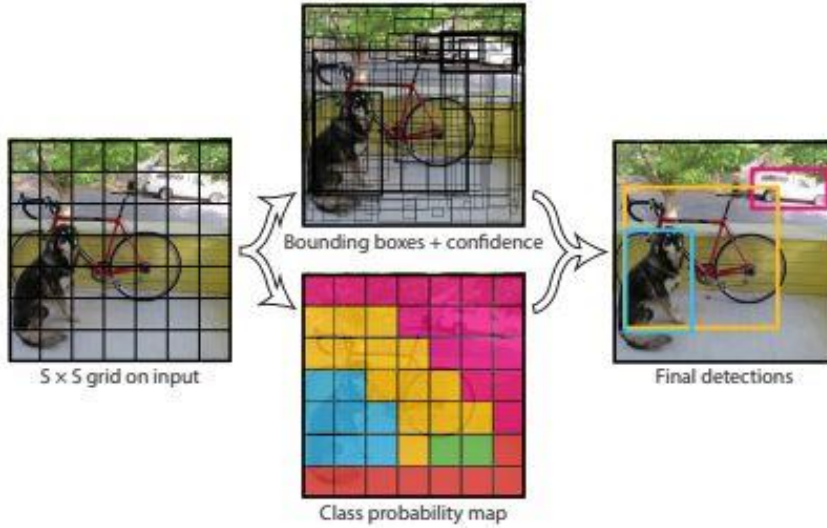
Akut lenfoblastik lösemi (ALL), beyaz kan hücresi kanseri olarak bilinen hematik bir rahatsızlıktır. Bu hastalığın ilk safhasında blast ya da lenfoblast olarak bilinen kötü huylu kan hücrelerinin kemik iliğinde aşırı üretimi görülür. Bu üretim sonucunda kemik iliğine sığmayan kötü huylu blast hücreleri kan sıvısında da görülmeye başlar. Periferik kan yayma, kan sıvısının incelenmesi için kullanılan bir tekniktir.

Bu çalışmada dijitalleştirilmiş periferik kan yayma görüntüleri üzerinde YOLOv4 nesne tanıma yaklaşımı kullanılarak blast hücrelerinin otomatik tespiti ve sayımı yapılmıştır.

YOLO nesne tanıma yaklaşımı

YOLO, [134] çalışmasında önerilen yeni bir nesne tanıma yaklaşımıdır. Bu yaklaşım, sınırlayıcı kutular üzerinde temellenen ve ilişkili sınıf olasılıklarına göre hesaplanan bir süreci içerir. Gerçek zamanlı bir çıktı sunan bu yapının ilk aşamasında giriş görüntüsü SXS’lik ızgara hücrelerine bölünür. Izgara hücrelerinin her biri nesnenin algılanmasından sorumlu olduğundan dolayı her bir hücre için sınırlayıcı kutular ve

kutuların güven puanları tahmin edilir. Güven puanları, modelin kutu içerisinde nesne olma durumunun doğruluğunu ve tahmin ettiği nesnenin ne kadar doğru olduğunu belirler. YOLO nesne tanıma algoritmasının temel mantığı Şekil 3.27’de tasvir edilmiştir.

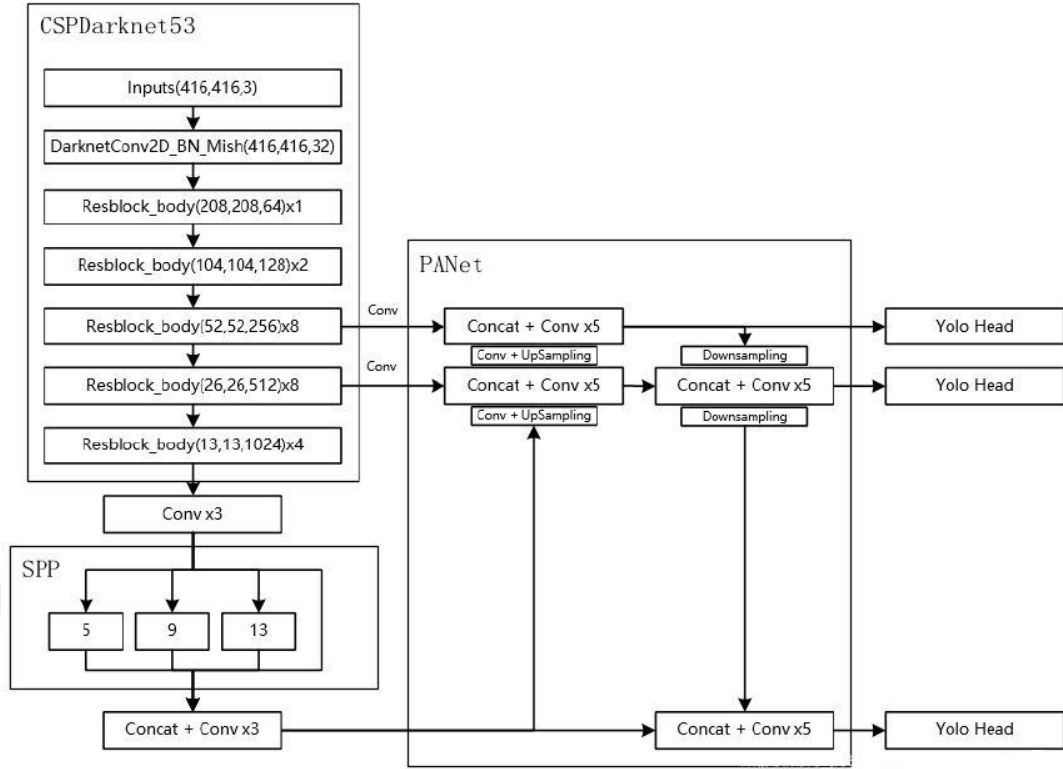


Şekil 3.27. YOLO nesne tanıma algoritmasının çalışma akışı [134]

YOLO zaman içerisinde geliştirilmiştir. Bu çalışmada, YOLOv3 mimarisinin geliştirilmiş versiyonu olan YOLOv4 mimarisi kullanılmıştır. YOLOv4 omurga (backbone), boyun (neck) ve baş (head) kısımlarından oluşmaktadır [135]. Omurga kısmı CSPNet ve Darknet53 mimarilerinin birlikte kullanıldığı CSPDarknet53 yapısıdır. Özniteliklerin çıkarılmasını sağlayan bu kısım, hesaplama ve bellek maliyetlerini azaltır ve CNN’in öğrenme kabiliyetini geliştirir [136]. Bu durum temel (base) katmana ilişkin özellik haritasının 2 parçaya bölünmesi ve ardından çapraz aşamalı hiyerarşik yapı ile birleştirilmesinden kaynaklanır.

Boyun kısmı SPP ve PAN yaklaşımlarından oluşur. SPP, alıcı alanı iyileştiren modüldür. Mekansal Piramit Eşleştirmeden (SPM) meydana gelir. SPM, giriş görüntüsünü SXS’lik ızgaraya böler ve SPP modülünün yardımı ile CNN’e entegre eder. PAN modülü, farklı omurga seviyelerinden parametreleri toplar [135-137]. Baş kısmı, farklı boyutlardaki nesnelerin algılanmasını sağlar [138].

YOLOv4 ağının mimarisi Şekil 3.28’de verilmiştir.



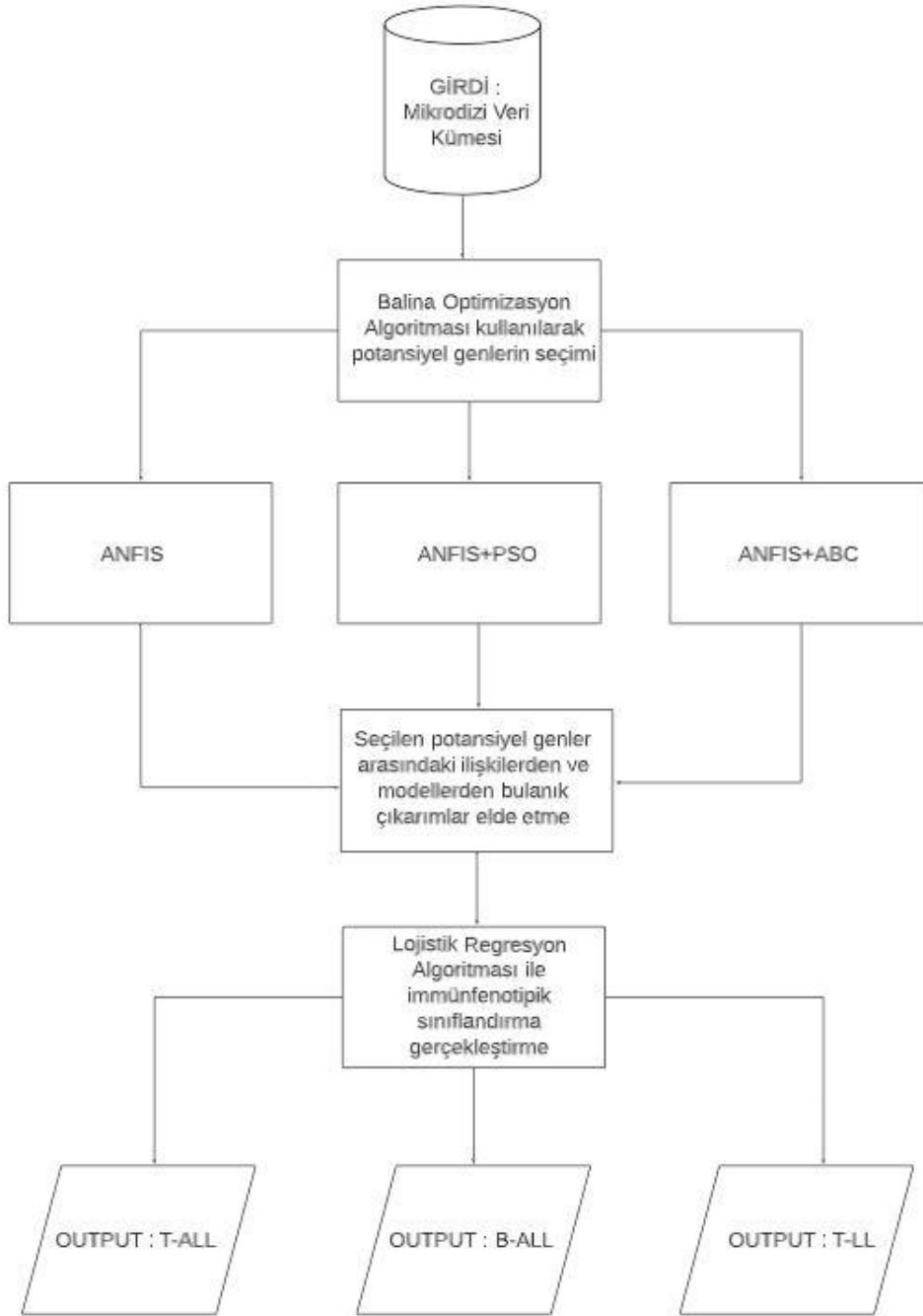
Şekil 3.28. YOLOv4 yaklaşımının mimarisi [139]

Bu bölümde lösemnin temel türü olan ALL malignitesinin periferik kan yayma tekniği ile tespiti için YOLOv4 nesne tanıma yöntemi kullanılmıştır. Aynı zamanda tedavi protokolünün belirlenmesi ya da tedaviye verilen yanıtın değerlendirilmesinde önemli bir kriter olan blast hücre sayımı da bu bölüm kapsamında gerçekleştirilmiştir.

3.3.6. Mikrodizi veri kümesi kullanılarak T-ALL, B-ALL ve T-LL malignitelerinin sınıflandırılmasına ilişkin iş akışı

Lösemi ve lenfoma; klinik, patolojik, sitogenetik, moleküler ve immünolojik özellikler olarak birbiri ile örtüşen malignitelerdir [41]. Ancak tanı ve tedavi sürecinde birbirinden ayrılan kısımların mevcut olması malignitelerin sınıflandırma gerekliliğini ön plana çıkarmıştır. Örneğin, lösemi hastalığının tanı sürecinde periferik kan yayma ve kemik iliği tutulumu üzerinden incelemeler yapılır. Ancak lenfoma hastalığı için bu incelemeler çok fazla bir etkiye sahip değildir [41, 42]. Öte yandan T-ALL ve T-LL maligniteleri için kullanılan mediastinal ışınlama, kök hücre nakli ve kemoterapi yöntemlerinin hastaya uygulanma biçimlerinde farklılıklar vardır [44]. Ek olarak T-ALL ve B-ALL malignitelerinin sahip olduğu farklı moleküler ve genomik yapılar, tedavi yöntemlerinde farklı çıktılar sunmaktadır. Bu nedenle T-ALL, B-ALL ve T-LL

ayrı ayrı incelenmesi gereken malignitelerdir [42-43]. Lösemi hastalığına ilişkin T-ALL, B-ALL ve T-LL malignitelerinin sınıflandırılması için altıncı bölümde önerilen akış şeması Şekil 3.29’da verilmiştir.



Şekil 3.29. Lösemi hastalığına ilişkin T-ALL, B-ALL ve T-LL malignitelerinin sınıflandırılması için altıncı bölümde önerilen akış şeması

Bu çalışmada lösemi kanseri ile ilişkili olan T-ALL, B-ALL ve T-LL malignitelerinin sınıflandırılmasına ilişkin bir çalışma yapılmıştır.

3.3.6.1. Mikrodizi veri kümesi kullanılarak T-ALL, B-ALL ve T-LL malignitelerinin sınıflandırılması için kullanılan yöntemler

Akut lösemiler, miyeloid ve lenfoid hücrelerin farklılaşma sürecinde meydana gelen malign hastalıklardır ve blast hücrelerin varlığı ile karakterize edilir [39]. Blast hücrelerine ilişkin alt türlerin sınıflandırılmasında farklı değerlendirmeler mevcuttur. Bu bölümde immünofenotipik değerlendirme irdelenecektir. Böylece ALL türü tespit edilecek ve T/B ya da non T/B şeklinde hücre sınıfı belirlenecektir.

Bu çalışmada sitogenetik, patolojik, immünofenotipik, klinik ve moleküler özellikleri örtüşen ve bununla birlikte tanı ve tedavi süreçleri farklılık gösteren T-ALL, B-ALL ve T-LL vakaları sınıflandırılmıştır. İlk olarak balina optimizasyon algoritması ile mikrodizi veri kümesinden hedef hastalıklarla ilişkili genler seçilmiştir. Ardından seçilen potansiyel genler arasındaki desenlerden ANFIS algoritması ile bulanık çıkarımlar elde edilmiştir. Yapay sinir ağları ile bulanık mantık yapısını birlikte kullanan ANFIS, birçok alanda ve farklı sistemlerin modellenmesinde tercih edilmektedir. Ancak nitelikli çıktılar üretmek amacıyla ANFIS yapısının eğitim süreci önemlidir [122]. Bu nedenle PSO ve ABC optimizasyon algoritmaları ile ANFIS yapısındaki üyelik fonksiyonunun parametre optimizasyonu sağlanmıştır. Ardından ANFIS, ANFIS+ABC, ANFIS+PSO yapıları doğrultusunda ulaşılan çıkarımlar kullanılarak daha güçlü, daha istikrarlı ve daha başarılı sonuçlar üretmek amacıyla topluluk makine öğrenmesi yaklaşımı çerçevesinde yer alan lojistik regresyon algoritması ile maligniteler sınıflandırılmıştır.

ANFIS yapısının optimizasyonunda yapay arı kolonisi algoritması

Yapay arı kolonisi (ABC), bal arılarının yiyecek arama davranışından esinlenen bir optimizasyon algoritmasıdır. Kolonide tüm işler için özelleşmiş arılar mevcuttur. Bu arılar, iş bölümü yapabilme ve kendi kendine organize olma özelliklerine sahiptir [124].

Yapay arı kolonisi optimizasyon algoritmasında 3 farklı grup vardır. Bu gruplar işçi, gözcü ve izci arılarıdır. Her bir kaynağa ilişkin nektar bir görevli arı tarafından alındığı için işçi arıların sayısı toplam yiyecek kaynağına eşittir. Öte yandan işçi ve gözcü arıların da eşit sayıda olduğu kabul edilir. Nektarı tükenen besinin görevli arısı son durumda kaşif arı olarak nitelendirilir [122-124].

ABC algoritmasında optimizasyon işlemi 3 aşamadan meydana gelmektedir. İlk olarak, denklem 3.69’da belirlenen matematiksel ifade ile arama uzayında çözümlere karşılık gelen rastgele yiyecek kaynağı yerleri üretilir [122-124].

$$X_{ij} = X_j^{min} + rand(0,1)(X_j^{max} - X_j^{min}) \quad (3.69)$$

Bu optimizasyon algoritmasında SN, popülasyon boyutudur. [1, SN] arasındaki herhangi bir değer, i tarafından kullanılır. $X_{i,i}$ ’inci çözümü ifade eder. Her çözüm optimize edilecek parametre sayısını belirten D elemanlı bir vektörden oluşur. X_j^{min} ve X_j^{max} sırasıyla en düşük ve en yüksek değeri tanımlar [122-124].

Her kaynak, besin kaynağının bulunduğu çevrede yeni kaynaklar belirleyen işçi arılardan oluşmaktadır. Bu arılar, tespit ettikleri kaynağın kalitesini bir önceki kaynağın kalitesine göre değerlendirir ve yeni kaynağın daha kaliteli olması durumunda hafızasında bu bilgiyi saklar. Denklem 3.70’de yeni bir besin kaynağının kaynak komşuluğunda belirlenmesine ilişkin benzetimin matematiksel ifadesi yer almaktadır [122-124].

$$V_{ij} = X_{ij} + \Phi_{ij}(X_{ij} - X_{kj}) \quad (3.70)$$

Algoritmada, BN işçi arı sayısını gösteren bir parametredir. [1, BN] arasındaki herhangi bir değer, k tarafından kullanılır. Φ_{ij} , [-1,1] arasında rastgele atanan bir sayıdır ve V_i besin kaynağı, X_i ’nin bir parametresi değiştirilerek bulunur.

Arama işlemi tüm görevli arılar tarafından tamamlandıktan sonra dans alanında kaynaklara ilişkin bilgiler gözcü arılar ile paylaşılır. Gözcü arılar paylaşılan bilgileri yorumlar ve olasılık değerine göre yeni kaynak seçimi gerçekleştirir. Yeni kaynağın miktarına göre mevcut pozisyon, önceki pozisyon ile karşılaştırılır ve değerlendirilir. Yeni kaynağın gözcü arı tarafından seçilmesi denklem 3.71’de verilen P olasılık değerine göre belirlenir [122].

$$P_i = \frac{fitness_i}{\sum_{j=1}^{SN} fitness_j} \quad (3.71)$$

Denklem 3.71’de yer alan $fitness_i$ parametresi, i’inci çözümün görevli arısı tarafından elde edilen uygunluk değeri ya da besinin kalitesidir. SN, besin kaynaklarının sayısıdır [124].

Bu bölümde ANFIS yapısı, parametre optimizasyonu için ABC algoritması ile birlikte kullanılmıştır. Bu yapının öncül ve sonuç kısmındaki parametreler, besin kaynağını ve besin kaynaklarının konumu problem ile ilgili bir çözümü temsil etmektedir [122]. Eğitim sonucunda en iyi çözümü bulmak için kaşif arıların başlangıç sayısı 10, değişken sayılarının minimum değeri -10 ve maksimum değeri +10 olarak seçilmiştir. Böylece ABC algoritması ile ANFIS yapısında kullanılan gauss üyelik fonksiyonuna ilişkin parametrelerin iyileştirilmesi sağlanmıştır.

ANFIS yapısıyla birlikte kullanılan ABC algoritması ile parametre optimizasyonu sağlanmıştır. Böylece daha güçlü çıkarımlar yapılmıştır.

ANFIS yapısının optimizasyonunda PSO algoritması

PSO, kuşların yemek arama davranışından esinlenerek oluşturulan metasezgisel optimizasyon algoritmasıdır [121] ve 3.3.3.1.'de detaylıca anlatılmıştır.

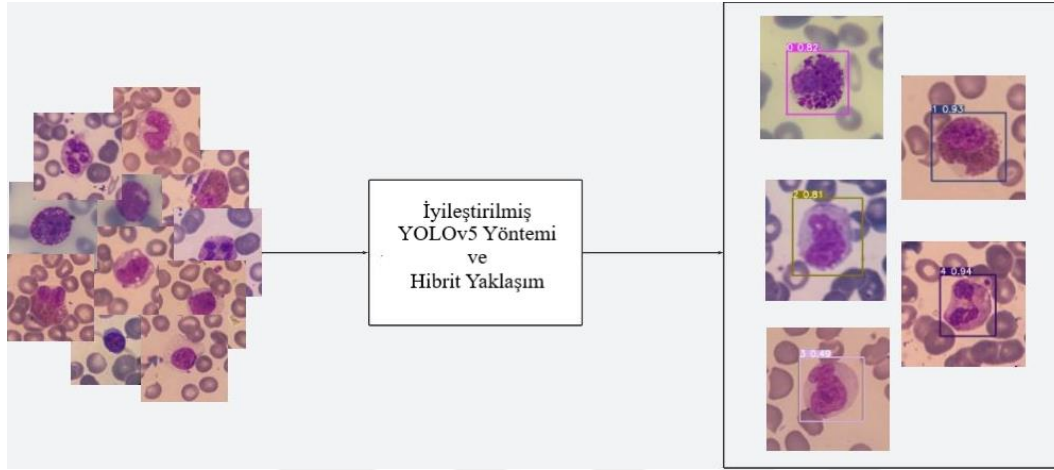
Bu kısımda verilere ilişkin daha güçlü çıkarımlar yapılabilmesi için ANFIS yapısı PSO algoritması ile birlikte kullanılmıştır. Bu yapının öncül ve sonuç kısmındaki parametreler besin kaynağını temsil etmektedir. Ek olarak gıda kaynaklarının konumu, problem kapsamında bir çözümü temsil etmektedir. Eğitim sırasında en iyi çözümü bulabilmek için öğrenme faktörleri 1.5, kendi yolunun hareket katsayısı 0.8, atalet ağırlığının azalma oranı 0,99, popülasyon büyüklüğü 10, değişken sayılarının minimum değeri -10 ve maksimum değer +10'dur. Böylece PSO algoritması ile ANFIS yapısında kullanılan gauss üyelik fonksiyonuna ilişkin parametrelerin iyileştirilmesi sağlanmıştır.

ANFIS yapısıyla birlikte kullanılan PSO algoritması ile parametre optimizasyonu sağlanmıştır. Böylece daha güçlü çıkarımlar yapılmıştır.

3.3.7. Beyaz kan hücrelerinin tespitine ve sınıflandırılmasına ilişkin iş akışı

Hayati fonksiyonların gerçekleşmesini mümkün kılan kan, hastanın mevcut durumunun değerlendirilebilmesini sağlayan yaşamsal bir sıvıdır [45]. Tıp dünyasında minimum maliyet ile kişinin sağlık durumunun değerlendirilebilmesi için kan üzerinde testler yapılmaktadır [46]. Periferik kan yayma yöntemi, mevcut testlerden biridir. Özellikle hastalık oluşturan maddelere ve ölümcül enfeksiyonlara karşı savunma gösteren beyaz kan hücrelerinin değerlendirilmesi için periferik kan yayma yöntemi tercih edilen hemotoloji testidir [47]. Ancak periferik kan yayma yönteminin manuel

olarak değerlendirilmesi çevresel ve insani olarak olumsuz birçok parametre içermektedir [36-38]. Bu çalışmada bilgisayar destekli sistemlerin yardımıyla periferik kan yayma yöntemi kullanılarak beyaz kan hücreleri, gerçek zamanlı olarak tespit edilmiş ve sınıflandırılmıştır. Lösemi hastalığının tanısında önemli bir biyobelirteç olan lenfosit hücre tipinin yüksek bir doğruluk oranı ile tespiti için yedinci bölümde önerilen akış şeması Şekil 3.30’da verilmiştir.



Şekil 3.30. Lösemi hastalığının tanısında önemli bir biyobelirteç olan lenfosit hücre tipinin yüksek bir doğruluk oranı ile tespiti için yedinci bölümde önerilen akış şeması

Bu çalışmada beyaz kan hücrelerinin alt türleri olan Monosit, Eozinofil, Bazofil, Lenfosit ve Nötrofil yapılarının değerlendirilmesine ilişkin bir analiz yapılmıştır.

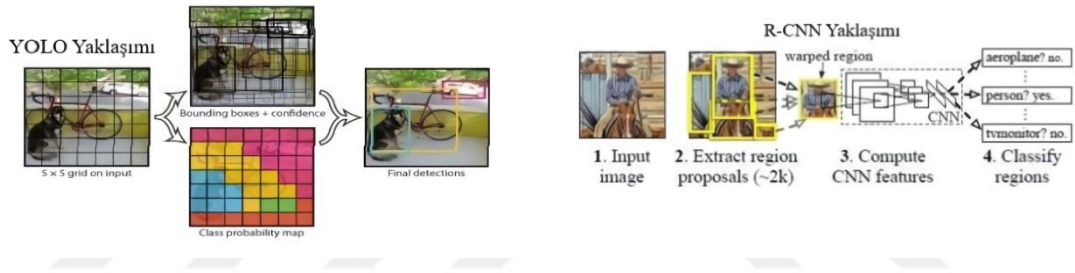
3.3.7.1. Beyaz kan hücrelerinin tespiti ve sınıflandırılması için kullanılan yöntemler

Beyaz kan hücreleri, hastalığa neden olan maddeler için savaşan yapılardır. Aynı zamanda ölümcül enfeksiyonlara karşı savunma gösterirler [3]. Bu hücrelerin kandaki yoğunlukları, hastalıklara ilişkin doğru tanının konulmasında önemli bir ayrıçtır. Bu nedenle sunulan çalışmada Monosit, Eozinofil, Bazofil, Lenfosit ve Nötrofil isimli dijitalleştirilmiş beyaz kan hücreleri YOLOv5 ve Detectron2 nesne tanıma mimarileri ile sınıflandırılmıştır. Ardından bu mimarilerden elde edilen başarı performanslarının geliştirilmesi için iyileştirilmiş YOLOv5 ve hibrit yapının kullanıldığı iki yaklaşım önerilmiştir.

Nesne tanıma yaklaşımı

Nesne tanıma, veri kümesinde yer alan görüntüler üzerinde hedef nesneleri bulmak, dikkörtgen sınırlayıcı kutular ile göstermek, güven değerini belirtmek ve sınıflandırmak için tercih edilmektedir. Nesne tanıma kapsamında iki temel yaklaşım vardır. Birinci yaklaşımda bölge önerileri yapılır ve ardından yapılan öneriler, nesne kategorisine göre sınıflandırılır. Detectron2 mimarisi birinci yaklaşım doğrultusunda değerlendirilir. İkinci yaklaşımda, nesne algılama süreci bir regresyon ya da sınıflandırma problemi olarak ele alınır ve birleşik bir çerçeve mantığı ile nihai çıktılar elde edilir. YOLOv5 mimarisi, ikinci yaklaşım doğrultusunda değerlendirilir [140].

Nesne algılama çerçevesinde yer alan ve iki ayrı değerlendirme yaklaşımı kullanan Regresyon/Sınıflandırma temelli ve bölge önerisi temelli iki ayrı yönteme ilişkin değerlendirme yaklaşımlarının farkı Şekil 3.31’de tasvir edilmiştir.

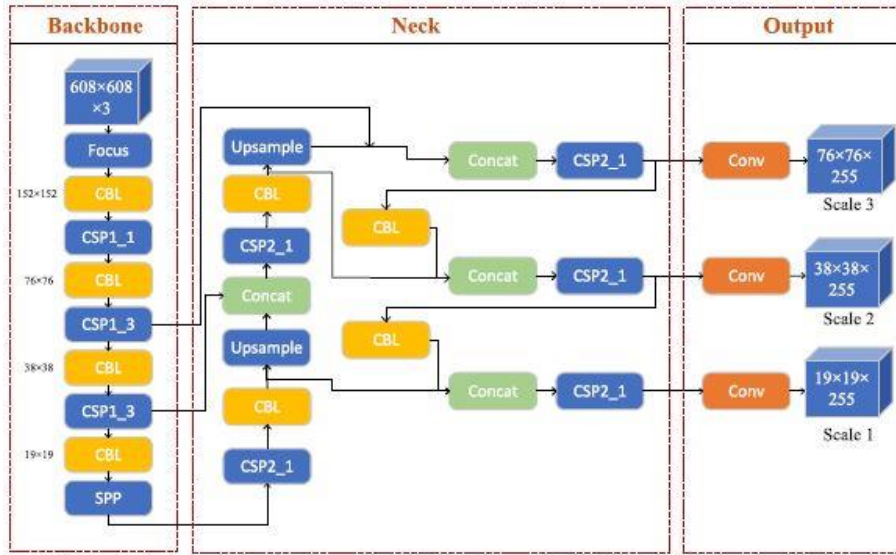


Şekil 3.31. Regresyon/Sınıflandırma temelli ve bölge önerisi temelli iki ayrı yönteme ilişkin değerlendirme yaklaşımlarının farkı [140]

Çalışma kapsamında nesne tespit süreci için 0.3 eşik değeri kullanan YOLOv5 ile Detectron2 mimarilerine ilişkin detaylar ve önerilen yeni yaklaşımlar aşağıda açıklanmıştır.

YOLOv5 yaklaşımı

YOLOv5 yüksek tespit doğruluğu, hafif sıklet özelliği ve ivedi algılama hızı sunan bir ağıdır. Giriş (Input), Omurga (Backbone), Boyun (Neck) ve Baş (Head) kısımlarından oluşur. YOLOv5 ağı, dört temel ön eğitilmiş mimariye sahiptir [141]. Bu mimariler YOLOv5s, YOLOv5m, YOLOv5l ve YOLOv5x'tir. Mimariler arasındaki temel fark, özellik çıkarma modüllerinin ve evrişim çekirdeklerinin sayısıdır. Bu doğrultuda YOLOv5s ve YOLOv5m mimarisi, basitleştirilmiş modeller; YOLOv5l karşılaştırmalı model ve YOLOv5x genişletilmiş model olarak tanımlanmaktadır [141]. YOLOv5 yaklaşımının mimarisi Şekil 3.32’de gösterilmiştir.



Şekil 3.32. YOLOv5 yaklaşımının mimarisi [142]

Mimarinin omurga kısmı öz niteliklerin çıkarıldığı evrimsel sinir ağıdır. Burada 152x152, 76x76, 38x38 ve 19x19 boyutlarında özellik haritaları mevcuttur. Boyun ağı bu özellik haritalarını birleştirerek bilgi kaybının azalmasına yardımcı olur. Füzyon işleminde ise güçlü semantik özellikleri, üst özellik haritalarından alt özellik haritalarına aktaran FPN yapısı ile düşük özellikli haritalardan daha yüksek özellikli haritalara güçlü yerleştirme özelliklerini aktaran PAN yapısı kullanılır. Bu iki yapı sayesinde boyun ağının özellik birleştirme kabiliyeti artar. Böylece $76 \times 76 \times 255$, $38 \times 38 \times 255$ ve $19 \times 19 \times 255$ boyutlarında üç özellik füzyon katmanı oluşur. Büyük nesnelerin algılanması için 19x19x255 özellik haritaları ve küçük nesnelerin algılanması için 76x76x255 özellik haritaları uygundur. Bu yeni özellik haritaları vasıtasıyla nesnelerin tespiti ve sınıflandırılması çıkış ağında gerçekleşir.

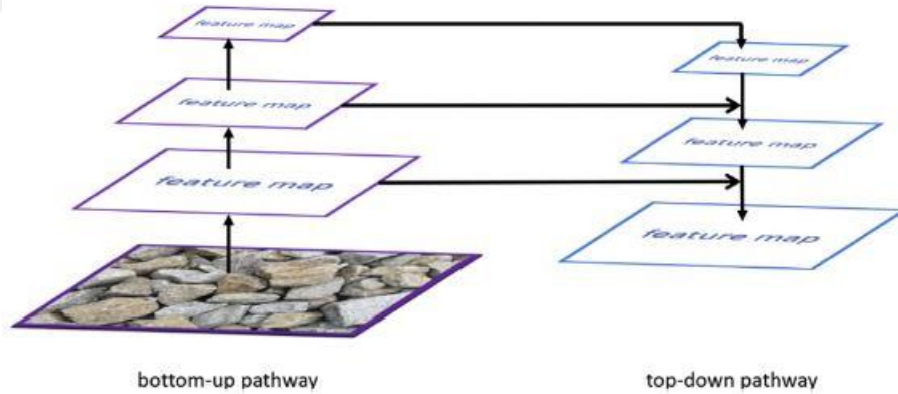
Mimarideki odak (Focus), alt örnekleme sırasında görüntülerin ayrılmasını ve ardından birleştirilmesini gerçekleştirerek özelliklerin daha nitelikli çıkarılmasını sağlayan modüldür. SPP (spatial pyramid pooling), görüntü özelliklerini yüksek bir soyutlama düzeyinde temsil eden uzaysal piramit havuzlama modülüdür. Bu modül, özellik haritasını küçülterek ağıın hesaplama karmaşıklığını basitleştirir ve ana özellikleri çıkarır. CSP (cross-stage partial network), ağıın çıkarım hızını arttıran ve ön ile arka katmanları birbirine bağlayan kısımdır. Yapısal farklılıkları olan iki ayrı CSP ağı mevcuttur. Omurgada yer alan CSP ağı bir ya da daha fazla residual ünitelerden meydana gelir. Bununla birlikte boyun kısmında yer alan CSP ağı, CBL modülleri ile

residual ünitelerin değiştirilmesi ile oluşur. Concat modülü tensörlerin birleştirilmesinden sorumludur [142].

Detectron 2 yaklaşımı

Detectron2, görüntülerdeki nesnelerin tespiti ve segmentasyonu amacıyla kullanılan Facebook yapay zeka araştırması kütüphanesidir [143]. Detectron ve MaskRCNN Benchmark'ın halefi olan bu yaklaşımın Detectron ve MaskRCNN'e göre en gelişmiş sonuçlar ürettiği kabul edilir [143].

Detectron2 ağı, R50-C4, R50-DC5, R50-FPN, R101-C4, R101-DC5, R101-FPN ve X101-FPN şeklinde ön eğitilmiş modellere sahiptir. Farklı omurgalar yardımıyla elde edilen hazır ön eğitilmiş modeller https://github.com/facebookresearch/detectron2/blob/main/MODEL_ZOO.md linkinde mevcuttur [27]. Bu ön eğitilmiş modeller COCO veri kümesi ile eğitilmiştir. Modellerdeki R50 ifadesi ResNet50 yaklaşımının kullanıldığını ve FPN son eki omurgada FPN yapısının yer aldığını gösterir. FPN, yukarıdan aşağıya ve aşağıdan yukarıya bir yol ile özelliklerin giriş görüntüsünden çıkarılmasını sağlayan kısımdır [144]. Detectron yaklaşımındaki temsili yol yapısı Şekil 3.33'de gösterilmektedir.



Şekil 3.33. Detectron yaklaşımındaki temsili yol yapısı [144]

Ön eğitilmiş modellerin hedef veri kümesine uygulanması sürecinde mevcut sınıf sayısına göre ön eğitilmiş model ayarlanır ve ardından ince ayar yapılarak eğitime uygun hale getirilir [27].

İyileştirilmiş YOLOv5x mimarisi

Bu bölümün orijinal ilk kısmı olarak nitelendirilen iyileştirilmiş YOLOv5 mimarisinde iki farklı katkı mevcuttur. İlk katkı, mimaride yer alan Convolution modülünün DeepWise Convolution modülü ile değiştirilmesidir. Bu durum parametre ve işlem boyutunun küçültülmesini sağlar. Aynı zamanda model gücünün artırılmasına imkan tanıyan bir yapı oluşturulur. İkinci katkı, sinir ağının eğitiminde önemli bir rol oynayan aktivasyon fonksiyonunun optimize edilmesidir. Aktivasyon fonksiyonu, eğitim performansının artırılmasını sağlayarak gerçek dünya problemlerinin öğrenilmesinde titizlikle seçilmesi gereken bir kriterdir [30]. Bu nedenle, en uygun çözümü bulmak ve gerçekçi bir çağrışım sağlamak amacı ile bu çalışmada optimize edilmiştir. Optimizasyon, sigmoid fonksiyonundan esinlenerek oluşturulan SiLU aktivasyon fonksiyonu üzerinde gerçekleştirilmiştir [32]. SiLU aktivasyon fonksiyonunun matematiksel ifadesi denklem 3.72’de verilmiştir.

$$\text{SiLU}(x) = x * \text{sigmoid}(x) \quad (3.72)$$

SiLU aktivasyon fonksiyonunda yer alan sigmoid fonksiyonu Altın Oran Arama algoritması ile optimize edilmiştir. Ardından optimum çıktılar elde etmek ve doğrusal olmayan durumları sinir ağına öğretmek amacıyla optimize edilen sigmoid fonksiyonun yer aldığı SiLU aktivasyon fonksiyonu kullanılmıştır. Bu yöntemin detaylı açıklaması aşağıda verilmiştir.

Altın oran arama algoritması

Optimizasyon ya da hesaplama gibi farklı uygulamalarda tercih edilen altın oran arama algoritması, türevlenemeyen ya da zor problemlerin çözümünde kullanılan bir yaklaşımdır. Örneğin $[a,b]$ aralığında yer alan fonksiyonun $f(x)$ minimum noktasını bulmak amacıyla x_1 ve x_2 olarak tanımlanan iki nokta üzerinde hesaplama sağlanır [145]. Bu noktaların matematiksel eşitliği denklem 3.73 ve 3.74 ‘te verilmiştir [145].

$$c = \frac{-1+\sqrt{5}}{2}$$
$$x_1 = ca + (1-c)b \quad (3.73)$$

$$x_2 = (1-c)a + cb \quad (3.74)$$

Hesaplama sonrasında x_1 ile x_2 olarak tanımlanan noktalar arasında bir karşılaştırma yapılır. Fonksiyonun f olarak karakterize edildiği $f(x_1)$ ve $f(x_2)$ çıktıları için $f(x_1)$ 'in $f(x_2)$ 'den küçük olması durumunda $[a, x_2]$ noktası, $f(x_1)$ 'in $f(x_2)$ 'den büyük olması durumunda $[x_1, b]$ noktası optimal nokta olarak belirlenir. Arama işlemi durdurma kriteri sağlanana kadar devam eder [145].

Doğrusal olmayan durumların sinir ağına öğretilmesi için kullanılan bu mimaride durdurma kriteri $1e-5$ ve aralık $[-1, +1]$ olarak belirlenmiştir.

Altın oran arama algoritması vasıtasıyla optimize edilen aktivasyon fonksiyonu ve DeepWise Convolution modülüyle değiştirilen Convolution modülü ile yeniden oluşturulan ağ yapısı kullanılarak nesne tespit oranında iyileşme sağlanmıştır.

Hibrit nesne tanıma yaklaşımı

Beyaz kan hücrelerinin tespiti ve sınıflandırılması için üretilecek sistemin güvenilirliğini arttırmak amacıyla YOLOv5 ile Detectron2 yöntemlerinin hibrit kullanıldığı bir yapı önerilmiştir. Bu hibrit yapı farklı değerlendirme yaklaşımlarına sahip olan YOLOv5 ve Detectron2 yöntemleri ile üretilen nesne tespit oranlarının karşılaştırılmasını içerir. Bu doğrultuda Detectron2 mimarisinin ürettiği nesne tespit başarı oranı yanlış ise YOLOv5 yaklaşımı ile elde edilen başarı oranına bakılır. YOLOv5 yaklaşımının başarı oranının Detectron2 algoritmasından büyük olması durumunda YOLOv5 yaklaşımının ürettiği çıktı doğru cevap olarak kabul edilir. Aksi bir durumda ise karşılaştırma yapılmaz.

Bu bölümde monosit, eozinofil, lenfosit, bazofil ve nötrofil isimli dijitalleştirilmiş beyaz kan hücrelerin tespiti ve sınıflandırılması için 2 orijinal yaklaşım önerilmiştir. Ardından bu yaklaşımlar kullanılarak ulaşılan sonuçlar ile YOLO ve Detectron yöntemlerinden elde edilen sonuçlar BÖLÜM 4'te karşılaştırılmıştır.



4. ARAŞTIRMA BULGULARI

Kanser, hücrelerin kontrolsüz ve hızlı bir şekilde yayılması ile oluşum gösteren bir malignitedir [2]. Lösemi, akciğer, kolon, meme ve prostat en sık görülen kanser türleridir [3].

Bu tezde lösemi hastalığı ele alınmıştır ve ardından sitogenetik, moleküler, patolojik, moleküler ve immünofenotipik değerlendirme yöntemleri ile analiz edilmiştir. Çünkü lösemi alt türlere sahip bir malignitedir ve alt türlere ilişkin doğru bir tanı sağlamak hayatidir. Doğru bir tanı sonucunda hastaya uygun tedavi protokolü belirlenecektir. Öte yandan tedaviye verilen yanıtın değerlendirilmesi ve bu değerlendirme sonucuna göre yeni ve başarılı bir sürecin başlaması da doğru tanı kararı ile sağlanacaktır. Ek olarak MLL lösemi türünün yeni bir tür olarak keşfedilmesinden sonra lösemi hastalığının halen keşfedilebilecek yeni alt oluşumlar barındırdığı düşünülmektedir. Bu nedenle lösemi hastalığının farklı değerlendirme yöntemleri ile analiz edilmesi ve üretilen çıktılar ile yorumlanması önemlidir.

Bu çalışmada lösemi hastalığı 7 farklı hedef için analiz edilmiştir. Elde edilen sonuçlar değerlendirilmiştir ve literatürdeki benzer çalışmalar ile karşılaştırılmıştır.

4.1. DNA Genom Dizilimlerindeki Ekson ve İntron Bölgelerinin Sınıflandırılmasının Sonucunda Elde Edilen Bulgular

Bu bölümde ekson ve intron bölgelerinin sınıflandırılması amacıyla 2 aşamada gerçekleştirilen bir çalışma yapılmıştır. Çalışma kapsamında NCBI gen bankasından insana ait 4 ayrı BCR-ABL geni tedarik edilmiştir. Bu genler AM600680.1, AM400881.1, AM886138.1 ve EU447303.1 numaralarına sahiptir ve farklı sayıda organik bazdan oluşmaktadır.

Bu çalışmanın ilk aşamasında AM600680.1, AM400881.1, AM886138.1 ve EU447303.1 numaralarına sahip sembolik DNA dizilimleri sayısallaştırılmıştır. Sayısallaştırma işlemi için EIIP Haritalama Tekniği, Reel Haritalama Tekniği, Moleküler Kütle Haritalama Tekniği, Eşleştirilmiş Haritalama Tekniğinin Hidrojen

Bağı Enerji Kuralı ve Shannon Entropi Temelli Haritalama Tekniği kullanılmıştır. Sayısallaştırılan DNA yapılarına pencerelemiş fourier dönüşümü uygulanmıştır. Böylece zaman alanı sinyali eşdeğer frekans alanı sinyaline dönüştürülmüştür ve ilgili dizilim bölgelerinin sinyal çıktıları elde edilmiştir. Bu çıktılar, görüntü kontrastının iyileştirilmesi amacıyla kullanılan CLAHE yöntemi ile belirginleştirilmiştir. Sınıflandırmada farklı pencereleme yaklaşımları kullanılarak ekson ve intron bölgelerine ait grafikler arttırılmıştır ve 63 adet ekson bölgesi ile 54 adet intron bölgesini tasvir eden grafikler vasıtasıyla oluşturulan veri kümesi kullanılmıştır. Ardından VGG16, VGG19, Xception, MobilenetV2, InceptionResnetV2, DenseNet169, DenseNet201, Resnet101, Resnet150 ve EfficientNetB7 transfer öğrenme mimarileri ile sınıflandırma sağlanmıştır. Her bir transfer öğrenme için RMSprop optimize edicisi, 0.00007 öğrenme oranı, 0.7 momentum, 1e-07 epsilon değeri ve ikili çapraz entropi kayıp fonksiyonu kullanılarak eğitim gerçekleştirilmiştir. Eğitim aşamasında son 3 katman (tamamen bağlı katman/fully connected layer) kaldırılmıştır ve modelimize uygun bir şekilde yeni katmanlar eklenmiştir. Eklenen ilk katman transfer öğrenme yapmak istediğimiz modeldir. Modelimiz içerisinde kullanılan ağırlıklar imagenet veri kümesi üzerinde gerçekleştirilen eğitimin ağırlıklarıdır. Ardından, aşırı öğrenme problemini azaltılması amacıyla GlobalAveragePooling2D ve Dropout (0.1 unutmama oranı) katmanları sırasıyla eklenmiştir. Son olarak sigmoid aktivasyon fonksiyonu ile modele 2 ayrı çıkış sunulmuştur. Her bir haritalama tekniği için tercih edilen transfer öğrenme mimarileri ve bu mimariler tarafından elde edilen sonuçlar Tablo 4.1 - 4.5'te verilmiştir.

Tablo 4.1. Reel haritalama tekniği için elde edilen deneysel ölçütler.

Öğr. Akt. M.	DN	YP	YN	DP	BO	DYR	ÖZG	KSN	F Ö.
VGG16	21	0	20	0	%51.2	0	%100	0	0
VGG19	21	0	20	0	%51.2	0	%100	0	0
Xception	3	18	1	19	%53.6	%95	%14.2	%51.3	%66.6
MobilenetV2	4	17	1	19	%56	%95	%19	%52.7	%67.7
Inception ResnetV2	15	6	3	17	%78.04	%85	%71.4	%73.9	%79
DenseNet169	0	21	0	20	%48.7	%100	0	%48.7	%65.5
DenseNet201	0	21	0	20	%48.7	%100	0	%48.7	%65.5
Resnet101	15	6	5	15	%73.1	%75	%71.4	%71.4	%73.1

Tablo 4.1. (Devamı) Reel haritalama tekniği için elde edilen deneysel ölçütler.

Resnet150	16	5	4	16	%78.04	%80	%76.1	%76.1	%78
EfficientNetB7	18	3	6	14	%78.04	%70	%85.7	%82.3	%75.6

Tablo 4.2. Moleküler kütle haritalama tekniği için elde edilen deneysel ölçütler.

Öğr. Akt. M.	DN	YP	YN	DP	BO	DYR	ÖZG	KSN	F Ö.
VGG16	21	0	20	0	%51.2	0	%100	0	0
VGG19	1	20	0	20	%51.2	%100	%0.047	%50	%66.6
Xception	8	13	4	16	%58.5	%80	%38	%55.1	%65.2
MobilenetV2	18	3	13	7	%60.9	%35	%85.7	%70	%46.6
Inception ResnetV2	1	20	2	18	%46.3	%90	%0.047	%47.3	%62
DenseNet169	4	17	2	18	%53	%90	%19	%51.4	%65.4
DenseNet201	6	15	2	18	%58.5	%90	%28.5	%54.5	%67.8
Resnet101	13	8	6	14	%65.8	%70	%61.9	%63.6	%66.6
Resnet150	6	15	6	14	%48.7	%70	%28.5	%48.2	%57
EfficientNetB7	11	10	5	15	%63.4	%75	%52.3	%60	%66

Tablo 4.3. EIIP haritalama tekniği için elde edilen deneysel ölçütler.

Öğr. Akt. M.	DN	YP	YN	DP	BO	DYR	ÖZG	KSN	F Ö.
VGG16	14	7	4	16	%73.1	%80	%66.6	%69.5	%74.3
VGG19	12	9	6	14	%63.4	%63.6	%57.1	%60.8	%62.1
Xception	8	13	3	17	%60.9	%85	%38.0	%56.6	%67.9
Mobilenet V2	21	0	18	2	%56.0	%10	%100	%100	%18.1
Inception ResnetV2	0	21	1	19	%46.3	%95	0	%47.5	%63.3
DenseNet 169	17	4	6	14	%75.6	%70	%80.9	%77.7	%73.6
DenseNet 201	11	10	4	16	%65.8	%80	%52.3	%61.5	%69.5
Resnet101	19	2	13	7	%63.4	%35	%90.4	%77.7	%48.2
Resnet150	12	9	6	14	%63.4	%70	%57.1	%60.8	%65.0
EfficientNet B7	14	7	5	15	%70.7	%75	%66.6	%68.1	%71.3

Tablo 4.4. Shannon entropi temelli haritalama tekniği için elde edilen deneysel ölçütler.

Öğr. Akt. M.	DN	YP	YN	DP	BO	DYR	ÖZG	KSN	F Ö.
VGG16	21	0	20	0	%51.2	0	%100	0	0
VGG19	21	0	20	0	%51.2	0	%100	0	0
Xception	12	9	1	19	%75.6	%95	%57.1	%67.8	%79.1

Tablo 4.4. (Devamı) Shannon entropi temelli haritalama tekniği için elde edilen deneysel ölçütler.

MobileNet V2	5	16	0	20	%60.9	%100	%23.8	%55.5	%70.9
Inception ResnetV2	16	5	4	16	%78.04	%80	%76.1	%76.1	%78
DenseNet169	13	8	5	15	%68.2	%75	%61.9	%65.2	%69.7
DenseNet 201	2	19	1	19	%51.2	%95	%0.095	%50	%65.5
Resnet101	17	4	5	15	%78.04	%75	%80.9	%78.9	%76.9
Resnet150	14	7	8	12	%63.4	%60	%63.6	%63.1	%61.5
EfficientNetB7	15	6	5	15	%73.1	%75	%71.4	%71.4	%73.1

Tablo 4.5. Eşleştirilmiş haritalama tekniğinin hidrojen bağı enerji kuralı için elde edilen deneysel ölçütler.

Öğr. Akt. M.	DN	YP	YN	DP	BO	DYR	ÖZG	KSN	F Ö.
VGG16	21	0	20	0	%51.2	0	%100	0	0
VGG19	21	0	20	0	%51.2	0	%100	0	0
Xception	14	7	4	16	%73.1	%80	%66.6	%69.5	%74.3
MobileNet V2	19	2	12	8	%65.8	%40	%90.4	%80	%53.3
Inception ResnetV2	15	6	5	15	%73.1	%75	%71.4	%71.4	%73.1
DenseNet 169	10	11	4	16	%63.4	%80	%47.6	%59.2	%68.0
DenseNet 201	7	14	2	18	%60.9	%90	%33.3	%56.2	%69.1
Resnet101	13	8	3	17	%73.1	%85	%61.9	%68	%75.5
Resnet150	15	6	6	14	%70.7	%70	%71.4	%70	%70
EfficientNetB7	16	5	3	17	%80.4	%85	%76.1	%77.2	%80.9

Tablo 4.1 - 4.5 incelendiğinde performans değerlendirme kriterlerine göre eşleştirilmiş haritalama tekniğinin hidrojen bağı enerji kuralı kullanılarak sayısallaştırılan DNA dizilimleri üzerinde EfficientNetB7 mimarisinin ulaştığı başarımlar diğer haritalama teknikleri ve transfer öğrenme mimarilerine kıyasla daha başarılı sonuçlar üretmiştir. Aynı zamanda EfficientNetB7 mimarisi için önemli bir kısım olan mobil ters çevrilmiş dar boğaz konvolüsyonu (Mobile Inverted Bottleneck Layer/MBConv) ile konvolüsyon katmanının başı ve sonu arasında oluşturulan bir geçiş ile hesaplama oranında iyileşme sağlanmıştır. Sonuçların yorumlanmasında kullanılan değerlendirme kriterleri aşağıda açıklanmıştır.

Doğru Negatif (DN)

Uzmanlar tarafından ekson bölgesi olarak etiketlenen DNA dizilimleri için bilgisayar destekli sistemler tarafından ekson bölgesi olarak yapılan doğru tahminlerin sayısıdır.

Yanlış Pozitif (YP)

Uzmanlar tarafından ekson bölgesi olarak etiketlenen DNA dizilimleri için bilgisayar destekli sistemler tarafından intron bölgesi olarak yapılan yanlış tahminlerin sayısıdır.

Yanlış Negatif (YN)

Uzmanlar tarafından intron bölgesi olarak etiketlenen DNA dizilimleri için bilgisayar destekli sistemler tarafından ekson bölgesi olarak yapılan yanlış tahminlerin sayısıdır.

Doğru Pozitif (DP)

Uzmanlar tarafından intron bölgesi olarak etiketlenen DNA dizilimleri için bilgisayar destekli sistemler tarafından intron bölgesi olarak yapılan doğru tahminlerin sayısıdır.

Başarı Oranı (BO)

Ekson ve intron bölgeleri için yapılan doğru tahmin sayısının tüm tahminlere oranını ifade eden değerlendirme ölçütüdür.

Matematiksel ifadesi denklem 3.75'te verilmiştir.

$$DO=(DP+DN)/(DP+YP+DN+YN) \quad (3.75)$$

Elde edilecek başarı oranı kriterinin %100 ya da %100'e yakın bir sonuç üretmesi hedeflenmektedir.

Duyarlılık (DYR)

Uzmanlar ve bilgisayar destekli sistemler tarafından intron bölgeleri için yapılan ortak tahminlerin intron bölgeleri kapsamında yapılan tüm tahminlere oranını ifade eden değerlendirme ölçütüdür. Matematiksel ifadesi denklem 3.76'da verilmiştir.

$$DYR=DP/(DP+YN) \quad (3.76)$$

Elde edilecek duyarlılık kriterinin %100 ya da %100'e yakın bir sonuç üretmesi hedeflenmektedir.

Özgüllük (ÖZG)

Uzmanlar ve bilgisayar destekli sistemler tarafından ekson bölgeleri için yapılan ortak tahminlerin ekson bölgeleri kapsamında yapılan tüm tahminlere oranını ifade eden değerlendirme ölçütüdür. Matematiksel ifadesi denklem 3.77'de verilmiştir.

$$\text{ÖZG} = \text{DN} / (\text{DN} + \text{YP}) \quad (3.77)$$

Elde edilecek özgüllük kriterinin %100 ya da %100'e yakın bir sonuç üretmesi hedeflenmektedir.

Kesinlik (KSN)

Ekson ve intron bölgeleri kendilerine özgü kritik bilgiler taşıyan bölgelerdir ve bu bölgelerin doğru bir şekilde birbirinden ayırt edilmesi önemlidir. Kesinlik değerlendirme ölçütü, bilgisayar destekli sistemler için bölgelerin ayırt edilme doğruluğu ifade eder. Matematiksel ifadesi denklem 3.78'de verilmiştir.

$$\text{KSN} = \text{DP} / (\text{DP} + \text{YP}) \quad (3.78)$$

Elde edilecek kesinlik kriterinin %100 ya da %100'e yakın bir sonuç üretmesi hedeflenmektedir.

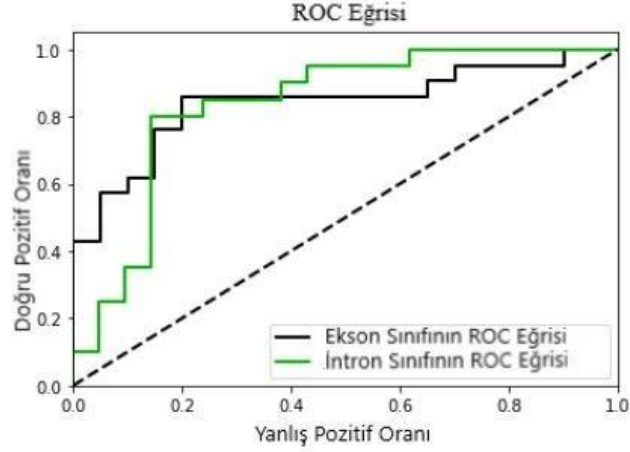
F Ölçütü (F Ö.)

Kesinlik ve duyarlılık kriterlerinin harmonik ortalamasının ifade edildiği değerlendirme ölçütüdür. Matematiksel ifadesi denklem 3.79'da verilmiştir.

$$\text{F Ö.} = (2 * \text{kesinlik} * \text{duyarlılık}) / (\text{kesinlik} + \text{duyarlılık}) \quad (3.79)$$

Elde edilecek F Ölçütü'nün %100 ya da %100'e yakın bir sonuç üretmesi hedeflenmektedir.

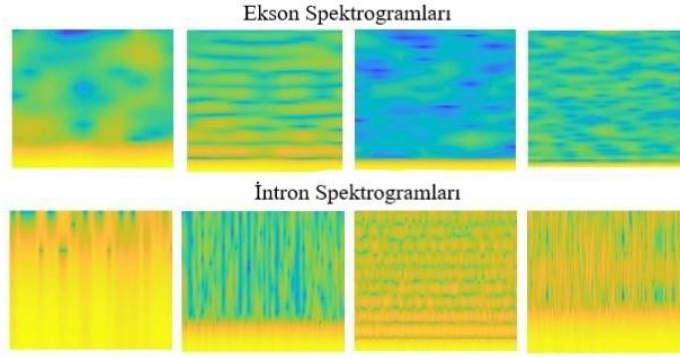
Bu bilgiler ışığında Tablo 4.6 - 4.10'da yer alan performans ölçütleri incelendiğinde çalışmanın birinci kısmı için en başarılı haritalama tekniği, eşleştirilmiş haritalama tekniğinin hidrojen bağı enerji kuralı ve en başarılı transfer öğrenme mimarisi, EfficientNetB7'dir. Aynı zamanda hidrojen bağı enerji kuralı ile sayısallaştırılan DNA dizilimlerinin EfficientNetB7 transfer öğrenme mimarisi kullanılarak sınıflandırılmasının sonucunda ekson ve intron bölgelerinin ayırt edilmesine ilişkin başarımın ROC eğrisi üzerinden tasviri Şekil 4.1'de gösterilmiştir.



Şekil 4.1. Hidrojen bağı enerji kuralı ile sayısallaştırılan DNA dizilimlerinin EfficientNetB7 transfer öğrenme mimarisi kullanılarak sınıflandırılmasının sonucunda ekson ve intron bölgelerinin ayırt edilmesine ilişkin başarımın ROC eğrisi ile tasviri

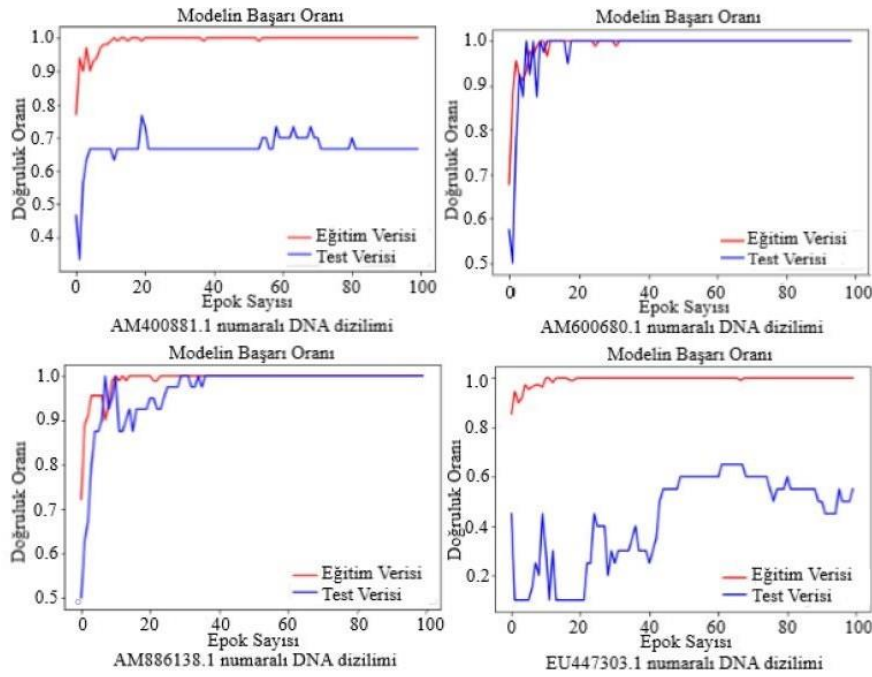
ROC eğrisi grafiğinde her bir kategori için çizilen eğrilerin altında kalan alanın %100 ya da %100'e yakın olması hedeflenir. Bu durum mevcut kategorilere ilişkin tanıma oranlarında başarılı tahminler üretildiğini göstermektedir. Bu doğrultuda Şekil 4.1'de verilen ROC eğrisi grafiği incelendiğinde ekson ve intron bölgelerinin tanıma performansında başarılı çıktılar üretildiği görülmüştür. Ancak ulaşılan tanıma performansının iyileştirilmesi planlanmış ve ikinci aşama gerçekleştirilmiştir.

Çalışmanın ikinci aşamasında, AM600680.1, AM400881.1, AM886138.1 ve EU447303.1 numaralarına sahip olan sembolik DNA dizilimleri ilk aşamadaki en başarılı haritalama tekniği olan Eşleştirilmiş Haritalama Tekniğinin Hidrojen Bağı Enerji Kuralı ile sayısallaştırılmıştır. Ardından sayısallaştırılan DNA dizilimlerine Pencerelemiş Kısa Zamanlı Fourier Dönüşümü uygulanmıştır. Böylece kısa zamanlı fourier dönüşümü yöntemi ile fourier dönüşümü yönteminin eş zamanlı olarak üretilmeyen zaman ve frekans bilgisine ilişkin eksiklik kısmen tolere edilmiştir. Pencerelemiş kısa zamanlı Fourier dönüşümü yöntemi ile elde edilen spektrogramlar Şekil 4.2'de verilmiştir.

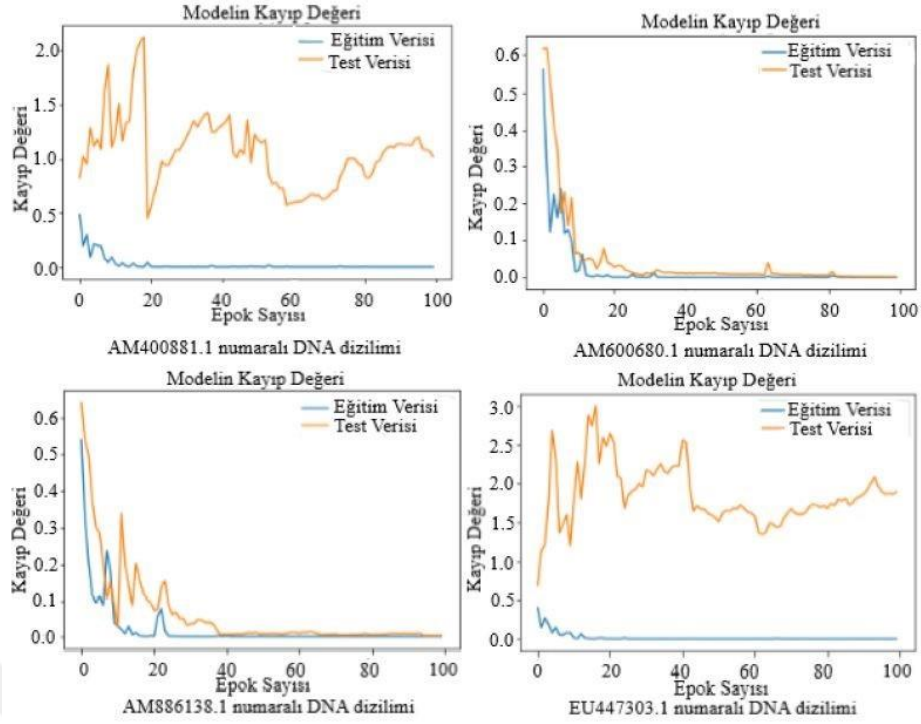


Şekil 4.2. BCR-ABL gen dizilimindeki ekson ve intron bölgelerine ilişkin spektrogramlar

Pencerelenmiş kısa zamanlı fourier dönüşümü ile elde edilen spektrogramlar, birinci aşamadaki en başarılı transfer öğrenme mimarisi olan EfficientNetB7 ile sınıflandırılmıştır. Sınıflandırmada farklı pencereleme yaklaşımları kullanılarak ekson ve intron bölgelerine ait spektrogramlar arttırılmıştır ve 63 adet ekson bölgesi ile 54 adet intron bölgesini tasvir eden spektrogramlar vasıtasıyla oluşturulan veri kümesi kullanılmıştır. AM600680.1, AM886138.1, AM400881.1, EU447303.1 numaralarına sahip her bir DNA diziliminde yer alan bölgelerin test veri kümesi ve kalan verilerin eğitim veri kümesi olarak sırasıyla kullanılmasının sonucunda ulaşılan doğruluk oranlarına ve kayıp değerlerine ilişkin grafikler Şekil 4.3. ve Şekil 4.4'te verilmiştir.



Şekil 4.3. 4 farklı test kümesi üzerinde gerçekleştirilen sınıflandırma sonucunda elde edilen doğruluk oranları



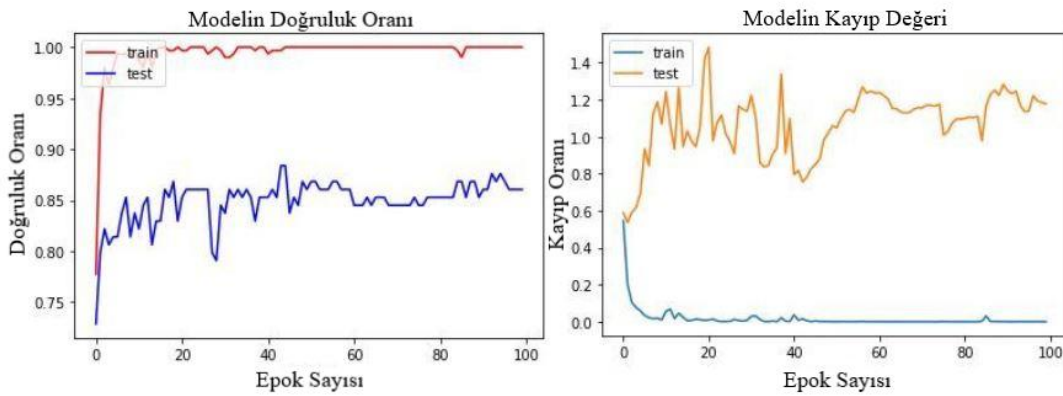
Şekil 4.4. 4 farklı test kümesi üzerinde gerçekleştirilen sınıflandırma sonucunda elde edilen kayıp değerleri

Doğruluk oranına ilişkin grafikler incelendiğinde AM600680.1 ve AM886138.1 numaralı DNA dizilimleri için başarı oranları %100'e ulaşmaktadır ve kayıp değeri 0'a yaklaşmaktadır. Bununla birlikte AM400881.1 ve EU447303.1 numaralı DNA dizilimlerine ilişkin başarı oranları %50-%70 seviyelerindedir ve kayıp değerleri 1.0-2.0 arasında seyretmektedir. Aynı zamanda AM400881.1 ve EU447303.1 numaralı DNA dizilimleri için başarı oranları ve kayıp değerlerine ait çıktılar istikrarsızdır.

Bu çalışmada üretilen AM600680.1 ile AM886138.1 ve AM400881.1 ile EU447303.1 DNA dizilimlerine ilişkin grafikler iki farklı şekilde yorumlanmaktadır. İlk olarak AM600680.1 ile AM886138.1 DNA dizilimleri için yüksek performans gösteren bir sınıflandırma işlemi gerçekleşmiştir. Bunun sebebi eğitim veri kümesinde farklı DNA baz uzunluğuna sahip olan DNA dizilimlerinin spektral yoğunluk bilgisini içeren spektrogramların yeterli sayıda bulunmasıdır. Farklı uzunluktaki DNA dizilimlerinin spektral yoğunluk bilgisini içeren zengin bir veri kümesi sayesinde eğitilen model çeşitli DNA uzunluklarına ait spektral yoğunluk bilgisini öğrenecektir. Böylece test veri kümesi öğrendiği bilgiler üzerinden test edilecektir. Öte yandan düşük performans gösteren AM400881.1 ve EU447303.1 DNA dizilimleri için modelin eğitildiği eğitim

veri kümesi belirli uzunluktaki nükleotit dizilimlerinden oluşmaktadır. Böylece yetersiz veri barındıran eğitim veri kümesi vasıtasıyla spektral yoğunluk bilgisi ile eşleşen kategori bilgisi yapay zeka tabanlı transfer öğrenme mimarileri tarafından eksiksiz bir şekilde öğrenilemeyecektir. Bu durumun sonucunda kararsız tahminler üretilecek ve başarı oranı düşecektir.

Bu kısımda AM600680.1, AM886138.1, AM400881.1, EU447303.1 DNA dizilimlerinin her biri, test kümesi olarak kalan dizilimlerin de eğitim kümesi olarak 63+54 görüntü üzerinden analiz edilmiştir. Bununla birlikte yeni bir değerlendirme süreci daha sunulmuştur. Bu doğrultuda eşleştirilmiş haritalama tekniğinin hidrojen bağı ile sayısallaştırılan dört ayrı DNA diziliminin ekson ve intron bölgelerine ait toplam 117 adet spektrograma veri çoğaltma yaklaşımı uygulanmıştır. Böylece ekson ve intron bölgelerine ilişkin spektrogramlar; yatay çevirme, dikey çevrime, döndürme, kırpma, gürültü ekleme, kontrast değiştirme teknikleri ile arttırılmıştır. Ekson bölgesine ait 220 adet spektrogram ve intron bölgesine ait 210 adet spektrogram ile yeni veri kümesi oluşturulmuştur. Bu veri kümesindeki spektrogramlar, eğitim ve test kümesi için %70 ve %30 oranlarında 42 olarak atanan seed parametresi ile rastgele dağıtılmıştır. Ardından EfficientNetB7 derin öğrenme mimarisi ile sınıflandırma aşaması gerçekleştirilmiştir. Veri çoğaltma yaklaşımı ile sayısı arttırılan spektrogramların EfficientNetB7 derin öğrenme mimarisi ile sınıflandırılmasının sonucunda elde edilen başarı oranına ve kayıp değerine ilişkin grafikler Şekil 4.5'te verilmiştir.



Şekil 4.5. Veri çoğaltma yaklaşımı ile sayısı arttırılan spektrogramların EfficientNetB7 derin öğrenme mimarisi ile sınıflandırılması

Eğitim veri kümesindeki 301 adet görüntü ile eğitilen model, 129 adet görüntünün yer aldığı test veri kümesinde test edilmiştir. Eğitim ve test veri kümesinde ulaşılan başarı oranları sırasıyla %100 ve %86.05 olarak elde edilmiştir. Eğitim ve test veri kümesine rastgele atanan spektrogramlar ile gerçekleştirilen eğitim boyunca istikrarlı sonuçlar üretilmesi başarılı bir eğitim sürecini ve ulaşılan çıktılarının kararlı olduğunu göstermektedir.

4.2. DNA Dizilimleri ile ALL ve KML Malignitelerinin Sınıflandırılmasının Sonucunda Elde Edilen Bulgular

Bu bölümde lösemnin temel türlerinden olan ALL ve KML malignitelerinin sınıflandırılması amacıyla 2 farklı aşamada gerçekleştirilen bir çalışma sunulmuştur. Bu çalışmada genetik temelli ALL ve KML malignitelerinin tanısında önemli bir ayıracı olan BCR-ABL geni kullanılmıştır. NCBI gen bankasından tedarik edilen ve farklı DNA baz uzunluklarına sahip olan ALL tanısı almış 35 ayrı BCR-ABL geni ile KML tanısı almış 27 ayrı BCR-ABL geni analiz edilmiştir.

Bu çalışmanın ilk aşamasında 35+27 farklı BCR-ABL gen dizilimi sabit haritalama teknikleri ve fiziko-kimyasal özellik tabanlı haritalama teknikleri ile sayısallaştırılmıştır. Sayısallaştırılan DNA dizilimlerinin spektral yoğunluk bilgisi, Kısa Zamanlı Fourier Dönüşümü ve Sürekli Dalgacık Dönüşümü yöntemleri ile spektrogramlara yansıtılmıştır. Ardından ALL ve KML malignitelerine ilişkin kategoriler ile eşleştirilen spektrogramlar DGCNN yaklaşımı ile sınıflandırılmıştır. Her bir haritalama tekniği için tercih edilen sinyal işleme yöntemi ile ulaşılan sonuçlar Tablo 4.6. ve 4.7’de verilmiştir.

Tablo 4.6. Kısa zamanlı fourier dönüşümü kullanılarak elde edilen spektrogramların DGCNN yaklaşımı ile sınıflandırılmasının sonucunda elde edilen deneysel ölçütler.

Haritalama Teknikleri	DP	YP	YN	DN	BO	DYR	ÖZG	KSN	F Ö.
Reel HT	9	2	6	3	0,55	0,60	0,50	0,54	0,56
Moleküler Kütle HT	7	4	4	5	0,55	0,57	0,50	0,72	0,63
Tamsayı HT	11	0	5	4	0,65	0,62	0,75	0,90	0,73
Karmaşık HT	6	5	6	3	0,55	0,57	0,50	0,72	0,63
EIIP HT	9	2	3	6	0,45	0,50	0,44	0,09	0,15

Tablo 4.6. (Devamı) Kısa zamanlı fourier dönüşümü kullanılarak elde edilen spektrogramların DGCNN yaklaşımı ile sınıflandırılmasının sonucunda elde edilen deneysel ölçütler.

Atomik Sayı HT	10	1	5	4	0,75	0,75	0,75	0,81	0,77
DNA-Yürüyüş HT	8	3	7	2	0,55	0,57	0,50	0,72	0,63
Eşleştirilmiş Sayısal HT	9	2	5	4	0,65	0,64	0,66	0,81	0,71

Tablo 4.7. Sürekli dalgacık dönüşümü kullanılarak elde edilen spektrogramların DGCNN yaklaşımı ile sınıflandırılmasının sonucunda elde edilen deneysel ölçütler.

Haritalama Teknikleri	DP	YP	YN	DN	BO	DYR	ÖZG	KSN	F Ö.
Reel HT	6	5	4	5	0,60	0,60	0,60	0,81	0,68
Moleküler Kütle HT	8	3	6	3	0,60	0,63	0,55	0,63	0,63
Tamsayı HT	10	1	6	3	0,75	0,68	1	1	0,81
Karmaşık HT	8	3	6	3	0,55	0,50	0,37	0,54	0,51
EIIP HT	1	10	1	8	0,75	0,75	0,75	0,81	0,77
Atomik Sayı HT	9	2	3	6	0,70	0,66	0,80	0,90	0,76
DNA-Yürüyüş HT	8	3	6	3	0,50	0,53	0,40	0,72	0,61
Eşleştirilmiş Sayısal HT	9	2	5	4	0,65	0,64	0,66	0,81	0,71

Tablo 4.6 ve 4.7’de verilen değerlendirme kriterleri incelendiğinde sürekli dalgacık dönüşümü yöntemi ile elde edilen sonuçların kısa zamanlı fourier dönüşümü yöntemine kıyasla daha başarılı çıktılar ürettiği görülmektedir. Bu durum dalgacık dönüşümü yönteminin güçlü bir zaman ve frekans çözünürlüğü sunması ile açıklanmaktadır. Ek olarak dalgacık dönüşümü yöntemi, yüksek frekansların kısa süreli ve düşük frekansların uzun süreli dalgacık ile belirtilmesi nedeniyle dinamik sinyallerin analizinde tercih edilmektedir. Bununla birlikte çalışmanın bu kısmında yapılan analizler DNA uzunluğu ile ilişkili çıktılar sunmaktadır. Bu nedenle ALL ve KML malignitelerinin sınıflandırma performansını iyileştirmek için ikinci aşama gerçekleştirilmiştir.

Çalışmanın ikinci aşamasında 35+27 farklı BCR-ABL gen dizilimi için sayısallaştırma işlemi [48] makalesinde önerilen Shannon Entropi Temelli Haritalama Tekniği ile

yapılmıştır. Bu teknik ile kodon dağılımlarına ilişkin entropi değerleri hesaplanmıştır. Ardından DNA uzunluğundan bağımsız bir değerlendirme sağlamak için sayısallaştırılan DNA dizilimlerinden istatistiksel ve yapısal özellikler çıkarılmıştır. Bu özellikler, uyarlanabilir bulanık mantık algoritması (ANFIS) ile sınıflandırılmıştır. Uyarlanabilir bulanık mantık algoritması ile gerçekleştirilen sınıflandırma işlemi sonunda ulaşılan değerlendirme kriterleri Tablo 4.8’de verilmiştir.

Tablo 4.8. Uyarlanabilir bulanık mantık yaklaşımı kullanılarak gerçekleştirilen sınıflandırma sonucunda elde edilen deneysel ölçütler

Haritalama Teknikleri	DP	YP	YN	DN	BO	DYR	ÖZG	KSN	F Ö.
Bulanık Mantık Yöntemi*	8	4	2	6	0,70	0,80	0,60	0,66	0,72
Bulanık Mantık Yöntemi**	10	2	2	6	0,80	0,83	0,75	0,83	0,83

DGCNN Approach1*=GMDH ağırları ile elde edilen sayısal çıkarımın özellikler içerisinde yer almaması sonucunda elde edilen deneysel ölçütler
DGCNN Approach2**=GMDH ağırları ile elde edilen sayısal çıkarımın özellikler içerisinde yer alması sonucunda elde edilen deneysel ölçütler

Tablo 4.8’de verilen değerlendirme kriterlerine göre ANFIS algoritması tarafından ALL ve KML malignitelerinin ayırt edilme gücü ve tespit doğruluğu iyileştirilmiştir. Bu durum DNA uzunluğundan bağımsız bir çıkarım sağlanması ve DNA’nın bulanık konfigürasyon yapısına uygun olan ANFIS algoritmasının kullanılması ile açıklanmaktadır.

4.3. Mikrodizi Veri Kümesi Kullanılarak ALL ve AML Malignitelerinin Sınıflandırılmasının Sonucunda Elde Edilen Bulgular

Bu bölümde löseminin temel türlerinden olan ALL ve AML malignitelerinin sınıflandırılması amacıyla 2 farklı aşamada gerçekleştirilen bir çalışma sunulmuştur. Bu çalışmada Bioinformatics laboratory [97] vasıtasıyla temin edilen gen ekspresyon verileri analiz edilmiştir. Organizmanın genetik kodunun alt kümesi olarak ifade edilen gen ekspresyonu [23] farklı birçok genin nispi ekspresyon seviyesi üzerinden moleküler tanı imkanı sağlamaktadır [25]. Bu çalışmada kullanılan gen ekspresyon

verileri 5147 gen ve 72 örnekten (47 ALL ve 25 AML) oluşmaktadır. Ancak 5147 gen verisinin hepsi araştırılan malignitelere ilişkin bilgi taşımaz. Bu durum hesaplama yükü ve zaman kaybı oluşturur. Aynı zamanda hastalıkla ilişkisiz genlerin hastalıkla ilişkili genler üzerinde gürültü oluşturmasıyla yanlış çıkarımlar oluşturma ihtimali barındırır. Bu problemlerin üstesinden gelmek ve başarılı bir çıkarım yapmak amacıyla yüksek boyut sunan mikrodizi veri kümesine sırasıyla parçacık sürü optimizasyon algoritması, karınca optimizasyon algoritması ve balina optimizasyon algoritması uygulanmıştır. Her bir optimizasyon algoritması için elde edilen potansiyel genlerin spektral yoğunluk bilgisi sürekli dalgacık dönüşümü yöntemi ile spektrogramlara yansıtılmıştır. Ardından spektrogramların yerel kontrastı CLAHE yöntemi ile iyileştirilmiştir. Son olarak ALL ve AML maligniteleri için iyileştirilen bu spektrogramlar DGCNN yaklaşımı ile sınıflandırılmıştır. Her bir optimizasyon tekniğine göre tercih edilen sınıflandırma algoritmaları için ulaşılan sonuçlar Tablo 4.9, 4.10. ve 4.11’de verilmiştir.

Tablo 4.9. Karınca optimizasyon algoritması kullanılarak seçilen potansiyel genlerin sınıflandırma performansı üzerindeki etkisini gösteren deneysel kriterler

Yöntemler	BO	DYR	ÖZG	KSN	F Ö.
VGG19 Mimarisi	%73.3	0	%100	0	0
Inceptionv3	%73.3	0	%100	0	0
Mimarisi					
MobilenetV2	%66.6	%75	%63.6	%42.8	%54.5
Mimarisi					
DenseNet201	%73.3	%100	%63.6	%50	%66.6
Mimarisi					
Resnet150	%80	%75	%81.8	%60	%66.6
Mimarisi					
DGCNN Yaklaşımı	%93.3	%80	%100	%100	%88.8
KNN Algoritması	%80	%100	%72.72	%57.14	%72.72
Karar Ağacı Alg.	%80	%50	%90.90	%66.6	%57.11
Rastgele Orman	%73.33	%25	%90.90	%50	%33.33
Alg.					
SVM Alg.	%86.6	%75	%90.90	%75	%75

Tablo 4.10. Parçacık sürü optimizasyon algoritması kullanılarak seçilen potansiyel genlerin sınıflandırma performansı üzerindeki etkisini gösteren deneysel kriterler

Yöntemler	BO	DYR	ÖZG	KSN	F Ö.
VGG19 Mimarisi	%73.3	0	%100	0	0
Inceptionv3 Mimarisi	%33.3	%100	%09.09	%28.5	%44.4
MobilenetV2 Mimarisi	%66.6	%100	%54.5	%44.4	%61.5
DenseNet201 Mimarisi	%33.3	%100	%09.09	%28.5	%44.4
Resnet150 Mimarisi	%73.3	%50	%81.8	%50	%50
DGCNN Yaklaşımı	%86.6	%80	%90	%80	%80
KNN Algoritması	%80	%75	%81.81	%60	%66.6
Karar Ağacı Alg.	%80	%50	%90.90	%66.6	%57.11
Rastgele Orman Alg.	%73.33	%25	%90.90	%50	%33.33
SVM Alg.	%86.6	%75	%90.90	%75	%75

Tablo 4.11. Balina optimizasyon algoritması kullanılarak seçilen potansiyel genlerin sınıflandırma performansı üzerindeki etkisini gösteren deneysel kriterler

Yöntemler	BO	DYR	ÖZG	KSN	F Ö.
VGG19 Mimarisi	%73.3	0	%100	0	0
Inceptionv3 Mimarisi	%26.6	%100	0	%26.6	%42.1
MobilenetV2 Mimarisi	%80	%100	%72.7	%57.1	%72.7
DenseNet201 Mimarisi	%46.6	%100	%27.27	%33.3	%50.01
Resnet150 Mimarisi	%53.3	%100	%36.3	%36.3	%53.3
DGCNN Yaklaşımı	%86.6	%60	%100	%100	%75
KNN Algoritması	%73.3	%50	%81.81	%50	%50
Karar Ağacı Alg.	%80	%50	%90.90	%66.6	%57.11
Rastgele Orman Alg.	%73.33	%25	%90.90	%50	%33.33
SVM Alg.	%80	%75	%90.90	%60	%66.6

Tablo 4.9, 4.10. ve 4.11. incelendiğinde performans değerlendirme kriterlerine göre transfer öğrenme mimarileri ve geleneksel sınıflandırma algoritmalarına kıyasla üretilen maksimum başarı oranı graf evrişim katmanları ile CNN katmanlarının birlikte kullanılmasına imkan tanıyan DGCNN yöntemi ile ulaşılmıştır. Ek olarak sürü

zekası temelli karınca optimizasyon algoritması kullanılarak seçilen potansiyel genlerin DGCNN yöntemi ile sınıflandırılmasının sonucunda en iyi performans elde edilmiştir.

4.4. Mikrodizi Veri Kümesi Kullanılarak ALL, AML ve MLL Malignitelerinin Sınıflandırılmasının Sonucunda Elde Edilen Bulgular

Bu bölümde lösemi hastalığına ilişkin ALL, AML ve MLL malignitelerinin sınıflandırılması amacıyla Bioinformatics laboratory [97] vasıtasıyla temin edilen gen ekspresyon verileri analiz edilmiştir. Bu çalışmada kullanılan gen ekspresyon verileri 12533 gen ve 72 örnekten (24 ALL, 28 AML ve 20 MLL) oluşmaktadır. Ancak 12533 gen verisinin hepsi araştırılan malignitelere ilişkin bilgi taşımamaktadır. Bu nedenle 12533 genden oluşan veri kümesine balina optimizasyon algoritması uygulanmış ve hastalıkla ilişkili potansiyel genler seçilmiştir. Ardından seçilen potansiyel genler LSTM sinir ağına girdi olarak verilmiştir. Test veri kümesinin 100 farklı seed değeri ile karıştırılmasının sonucunda elde edilen 100 ayrı test kümesi için ulaşılan ortalama başarı oranı %89.883 olarak bulunmuştur. Bu durum LSTM modelinin üç farklı tür için ürettiği dizi tahmin işlevinin modele başarı bir şekilde uygulandığını göstermektedir.

Bu bölümde tercih edilen LSTM mimarisi genellikle metin analizi ya da zaman serilerinin tahmin süreçlerinde kullanılmaktadır. Ancak farklı birçok amaç doğrultusunda üretilecek karar destek sistemleri için de başarılı çıkarımlar yapabilecek fiziksel ünitelere sahiptir. Böylece birbiri ile etkileşim içerisinde olarak nihai sonucu üreten döngüsel bağlantılara ve ağın geçici durumunu rastgele periyotlar ile hatırlayabilme eylemine sahip olan LSTM sinir ağı mimarisi [146], gen ekspresyon verilerinin birbiri ile ilişkisini hatırlama işlevleri ile dinamik tutarak güçlü bir sınıflandırma süreci inşa etmiştir.

4.5. Dijitalleştirilmiş Periferik Kan Yayma Görüntüleri Kullanılarak ALL Malignitesine İlişkin Tespitin Sonucunda Elde Edilen Bulgular

Bu bölümde lösemi hastalığına ilişkin ALL malignitesinin tespiti amacıyla Acute Lymphoblastic Leukemia Image Database [98] vasıtasıyla temin edilen normal ve ALL hastalıklarına sahip bireylerin dijitalleştirilmiş periferik kan yayma görüntüleri analiz edilmiştir. Periferik kan yayma görüntülerindeki blast hücrelerin tespiti ve

kandaki yoğunluğu löseminin temel türü olan ALL malignitesinin tanısında ve tedaviye verilen yanıtın değerlendirilmesinde önemlidir. Bu nedenle sunulan çalışmada periferik kan yayma görüntüleri üzerinde YOLOv4 algoritması ile blast hücreleri tespit edilmiş ve ardından blast hücre sayısına ilişkin bilgi verilmiştir.

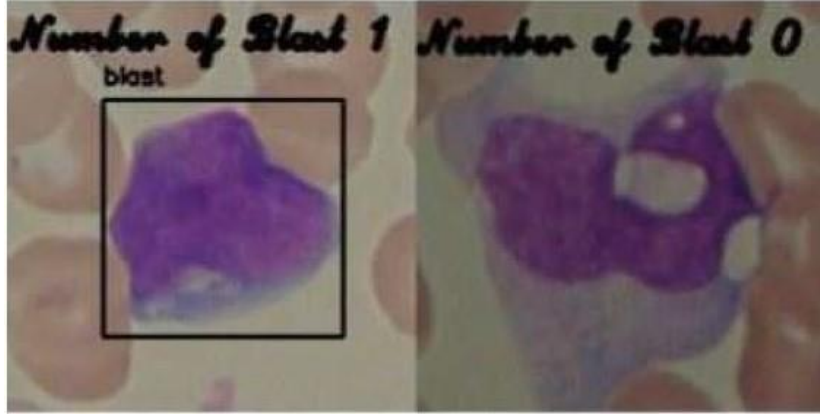
Bu bölümde tercih edilen YOLOv4 nesne tespit algoritması, CSPDarknet53 omurga yapısına sahiptir. Bu kısım CNN'in öğrenme kabiliyetini geliştirir. Aynı zamanda bellek ve hesaplama maliyetlerini azaltır [136]. Böylece tıbbi karar destek sistemlerinde harcanılan zamanın indirgenmesini sağlar. Ek olarak YOLO, iki aşamalı nesne tanıma çerçevelerine kıyasla sahip olduğu tek aşamalı nesne tanıma çerçevesi ile daha hızlı nesne tespit süreci inşa eder [141]. YOLOv4 algoritması ile ALL-IDB1 ve ALL-IDB2 veri kümeleri üzerinde 3000 iterasyon boyunca gerçekleştirilen eğitim sonucunda elde edilen kayıp değerleri Şekil 4.6'da verilmiştir.



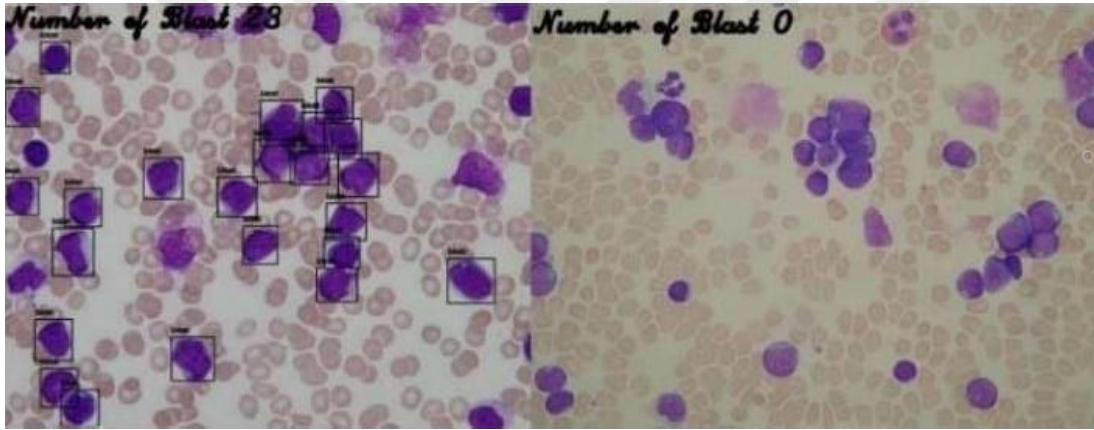
Şekil 4.6. ALL-IDB1 +ALL-IDB2 veri kümesi üzerinde 3000 iterasyon boyunca gerçekleştirilen eğitim süreci için üretilen kayıp değerleri

Şekil 4.6'da verilen grafikteki kayıp kriterine ilişkin oran, 1800 iterasyon boyunca aşamalı bir şekilde azalmıştır ve son 1200 iterasyonda ise istikrarlı bir şekilde devam ederek sonlanmıştır. Bu durum güçlü ve kararlı bir modelin inşa edildiğini göstermektedir. Aynı zamanda eğitilen bu modelin görüntüler üzerinde ürettiği blast hücre tespitinin başarısı da incelenmiştir. Bu amaçla periferik kan mikroskobunun farklı büyütme oranları ile elde edilen dijitalleştirilmiş kan yayma görüntüleri, eğitilen

modele girdi olarak verilmiştir. ALL-IDB1 ve ALL-IDB2 veri kümelerinden tedarik edilen ikişer örnek için elde edilen sonuçlar Şekil 4.7. ve Şekil 4.8’de sunulmuştur.



Şekil 4.7. ALL-IDB1 test veri kümesinden rastgele seçilen iki örnek için eğitilen model tarafından üretilen çıktı



Şekil 4.8. ALL-IDB2 test veri kümesinden rastgele seçilen iki örnek için eğitilen model tarafından üretilen çıktı

Şekil 4.7. ve Şekil 4.8’de verilen çıktılarda blast hücrelerin mevcut olduğu görüntüler üzerinde sınırlayıcı kutular ile blast hücreleri belirtilmiştir. Ek olarak tespit edilen blast hücre sayısına ilişkin bilgi de çıktı görüntüsü ile birlikte verilmiştir. Bununla birlikte blast hücrenin bulunmadığı görüntüler üzerinde farklı bir kan hücresi için blast hücre tespiti yapılmamıştır.

4.6. Mikrodizi Veri Kümesi Kullanılarak T-ALL, B-ALL ve T-LL Malignitelerinin Sınıflandırılmasının Sonucunda Elde Edilen Bulgular

Bu bölümde lösemi hastalığına ilişkin T-ALL, B-ALL ve T-LL malignitelerinin sınıflandırılması amacıyla Bioinformatics laboratory [97] vasıtasıyla temin edilen gen

ekspresyon verileri analiz edilmiştir. Bu çalışmada kullanılan gen ekspresyon verileri 15434 gen ve 48 örnekten (20 T-ALL, 18 T-LL ve 10 B-ALL) oluşmaktadır. Ancak 15434 gen verisinin hepsi araştırılan malignitelere ilişkin bilgi taşımamaktadır. Bu nedenle hastalıkla ilişkili potansiyel genler balina optimizasyon algoritması ile veri kümesinden seçilmiştir. Ardından T-ALL, B-ALL ve T-LL kategorileri için seçilen uygun genler adaptif bulanık mantık algoritması ile sınıflandırılmıştır.

T-ALL, B-ALL ve T-LL immünofenotipik, sitogenetik, patolojik, klinik ve moleküler özelliklerinden dolayı ortak terminoloji kapsamında incelenen malignitelerdir. Bununla birlikte malignitelerin ayırt edilmesini sağlayan farklı özellikler de mevcuttur. Bu çalışmada bahsedilen bulanık durumlar için malignitelerin belirsiz yapılarına uygulanan ANFIS algoritması ile genler arası ilişkilerden çıkarılan bulanık çıkarımların 3 ayrı kategori ile eşleştirilmesi sağlanmıştır.

ANFIS yapısı, yapay zekanın veriler üzerindeki öğrenme kabiliyeti ile bulanık mantığın veriler üzerindeki çıkarım gücünü birleştiren bir sinir ağıdır. Belirsizlikler içeren durumlara karşı bulanık setler için net girdi desenleri eşleyen bulanıklaştırıcı modüle sahiptir. Bu çalışma ile ANFIS'in çıkarım gücünden faydalanarak bulanık çıkarımlar elde edilmiştir. Ancak üretilen bu çıkarımları iyileştirmek amacıyla ANFIS algoritması hem parçacık sürü optimizasyon algoritması hem de yapay arı kolonisi optimizasyon algoritması ile hibrit kullanılmıştır. Böylece ANFIS algoritması için tercih edilen üyelik fonksiyonuna ait parametreler sırasıyla parçacık sürü optimizasyon ve yapay arı kolonisi optimizasyon algoritmaları ile optimize edilmiştir. Son olarak ANFIS, ANFIS+ABC ve ANFIS+PSO yapılarından elde edilen bulanık çıktılar topluluk öğrenme yaklaşımı kapsamında değerlendirilmiştir. Böylece tek bir örneği nitelendiren üç farklı çıkarım sonucunda oluşturulan veriler, lojistik regresyon algoritması ile sınıflandırılmıştır. Elde edilen başarı oranı, geleneksel sınıflandırma algoritmaları (karar ağacı, rastgele orman, destek vektör makineleri, naive bayes, k ortalama) ve k ortalama kümeleme yaklaşımı ile karşılaştırılmıştır. T-ALL, B-ALL ve T-LL malignitelerinin kesinlik, duyarlılık ve f kriterine göre elde edilen değerlendirme ölçütleri Tablo 4.12'de verilmiştir.

Tablo 4.12. T-ALL, B-ALL ve T-LL malignitelerinin tespit performanslarına ilişkin değerlendirme kriterleri

	T-LL			T-ALL			B-ALL		
Kriter	KSN	DYR	F Ö.	KSN	DYR	F Ö.	KSN	DYR	F Ö.
KA	0	0	0	%20	%100	%33,3	0	0	0
RO	%100	%77,7	%87,5	%40	%66,6	%50	%66,6	%66,6	%66,6
DVM	%100	%22,2	%36,3	%23,0	%100	%42,8	0	0	0
NB	%100	%55,5	%71,4	%37,5	%100	%54,5	%100	%66,6	%80
K-Or	0	0	0	%100	%70	%82,3	0	0	0
Öneri	%100	%100	%100	%100	%77,7	%87,5	%50	%100	%66,6

Tablo 4.12. incelendiğinde inşa edilen bu hiyerarşinin geleneksel makine öğrenmesi algoritmalarına kıyasla başarı oranı ve her bir malignite için tespit doğruluğu daha güvenilir sonuçlar sunmaktadır. Bu başarının ilk sebebi metasezgisel optimizasyon algoritmalarının türevsiz olması nedeniyle yerel optimumlardan kaçınmasıdır [147]. Böylece veri kümesinden uygun genler seçilmiştir. İkinci sebebi net göstergeleri olmayan malignitelerin sınıflandırılması için bulanık çıktılar üreten ANFIS yapısının kullanılmasıdır. Üçüncü sebebi ANFIS yapısı ile üretilen çıkarımların parçacık sürü optimizasyon algoritması ve yapay arı kolonisi algoritması ile iyileştirilmesidir. Dördüncü sebebi kolektif öğrenme yaklaşımı kullanarak birden fazla algoritmaya ilişkin gücün birleştirilmesi ile daha istikrarlı ve kararlı çıktılar elde edilmesidir. Beşinci sebebi hata değerini azaltmak için yumuşak geçişler sağlayan lojistik regresyon algoritması ile sınıflandırılmasıdır.

4.7. Beyaz Kan Hücrelerinin Tespiti ve Sınıflandırılması Sonucunda Elde Edilen Bulgular

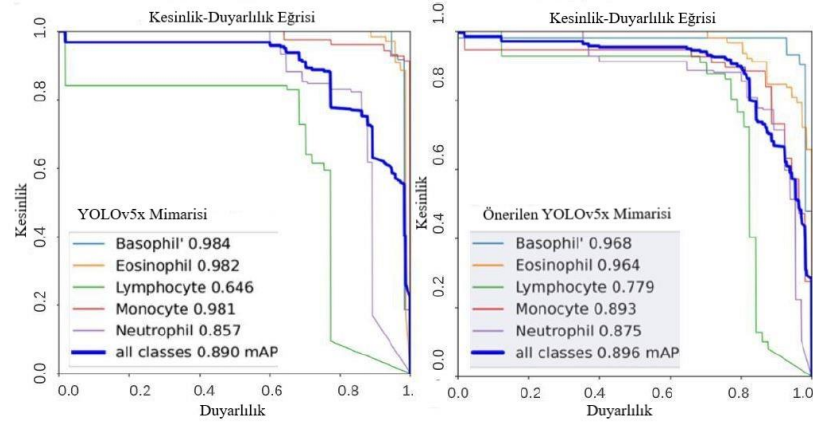
Bu bölümde bazofil, eozinofil, lenfosit, monosit ve nötrofil kan hücrelerinin sınıflandırılması amacıyla 2 aşamada gerçekleştirilen bir çalışma yapılmıştır. Çalışma kapsamında Raabin Health Database veri kümesinden [99] tedarik edilen dijitalleştirilmiş periferik kan yayma görüntüleri kullanılmıştır.

Bu çalışmanın ilk aşamasında lösemi hastalığının tanısında önemli bir ayıraç olan lenfosit hücresinin tespiti için nesne tanıma yaklaşımları olan YOLOv5s, YOLOv5x ve Detectron2 sinir ağı mimarileri kullanılmıştır. Bu mimarilerin ürettiği başarı oranı Şekil 4.9’da yer alan karışıklık matrisi ile ifade edilmiştir.

		YOLOv5s Modeli							YOLOv5x Modeli							Detectron2 Modeli				
		Tahmin							Tahmin							Tahmin				
		0	1	2	3	4			0	1	2	3	4			0	1	2	3	4
Gerçek	0	53	0	2	0	0	Gerçek	0	55	0	0	0	0	Gerçek	0	54	0	0	1	0
	1	0	40	2	0	29		1	0	69	0	1	1		1	0	71	0	0	0
	2	0	0	52	0	3		2	0	2	53	0	2		2	0	0	53	3	1
	3	0	1	10	41	1		3	0	0	9	43	1		3	0	0	3	48	1
	4	0	0	0	0	64		4	0	0	0	0	64		4	2	4	1	2	55

Şekil 4.9. Lenfosit hücrelerinin tanısında kullanılan YOLOv5s, YOLOv5x ve Detectron2 sinir ağı mimarileri ile elde edilen karışıklık matrisleri

Karışıklık matrisi ile hesaplanan başarı oranı YOLOv5s, YOLOv5x ve Detectron2 mimarileri için sırasıyla %83.3, %94.6 ve %93.6 olarak elde edilmiştir. Nesne tanıma mimarileri arasında lenfosit tespitine ilişkin tanıma doğruluğu, YOLOv5x ve Detectron2 mimarileri için aynı oranda çıktı üretmiştir. Ancak genel doğruluk oranı YOLOv5x mimarisi için maksimum başarı gösterdiğinden dolayı YOLOv5x mimarisi üzerinde iyileştirme yapılmıştır. Mimari ve aktivasyon fonksiyonu kapsamında gerçekleştirilen değişiklikleri içeren iyileştirilmiş YOLOv5x mimarisi ile YOLOv5x mimarisinin ürettiği Kesinlik-Duyarlılık eğrisi Şekil 4.10'da verilmiştir.



Şekil 4.10. Önerilen YOLOv5x mimarisinin YOLOv5x mimarisine göre lenfosit nesne tanıma oranında gösterdiği iyileşmenin kesinlik-duyarlılık eğrisi ile tasviri

Şekil 4.10'da verilen grafiklere ek olarak YOLOv5x ve önerilen YOLOv5x mimarilerine ilişkin performans metrikleri Tablo 4.13'te verilmiştir.

Tablo 4.13. YOLOv5x ve önerilen YOLOv5x mimarilerine ilişkin performans metrikleri

YOLOv5x Modeli	Kesinlik	Duyarlılık	mAP@.5	mAP@.5:.95:
Bazofil	0.974	0.947	0.984	0.539
Eozinofil	0.946	0.958	0.982	0.466
Lenfosit	0.709	0.702	0.646	0.18
Monosit	0.96	0.917	0.981	0.321
Nötrofil	0.819	0.862	0.857	0.515
Önerilen YOLOv5x Modeli	Kesinlik	Duyarlılık	mAP@.5	mAP@.5:.95:
Bazofil	0.931	0.952	0.968	0.665
Eozinofil	0.823	0.944	0.964	0.5
Lenfosit	0.875	0.735	0.779	0.232
Monosit	0.838	0.878	0.893	0.297
Nötrofil	0.886	0.718	0.875	0.478

Şekil 4.10. ve Tablo 4.13'te ulaşılan çıktılar analiz edildiğinde Kesinlik-Duyarlılık eğrisinde löseminin tespiti için önemli bir ayıraç olan lenfosit tanıma doğruluğunun 0.133 oranında iyileştirildiği görülmektedir. Aynı zamanda tüm sınıflar için tanıma doğruluğu 0.006 oranında da arttırılmıştır.

Çalışmanın ikinci aşamasında ilk aşamada elde edilen genel tespit doğruluğunun iyileştirilmesi sağlanmıştır. Bu doğrultuda regresyon/sınıflandırma tabanlı nesne değerlendirme yaklaşımını kullanan YOLO algoritması ile bölge öneri temelli nesne değerlendirme yaklaşımını kullanan Detectron2 mimarisi hibrit kullanılmıştır. Bu hibrit çalışma sonucunda YOLOv5s+Detectron2 ve YOLOv5x+Detectron2 yapıları için üretilen karışıklık matrisleri Şekil 4.11'de verilmiştir.

		YOLOv5s + Detectron2 Modeli Tahmin							YOLOv5x + Detectron2 Modeli Tahmin				
		0	1	2	3	4			0	1	2	3	4
Gerçek	0	54	0	0	1	0	Gerçek	0	54	0	0	1	0
	1	0	71	0	0	0		1	0	71	0	0	0
	2	0	0	57	0	0		2	0	0	56	1	0
	3	0	0	3	48	1		3	0	0	2	50	1
	4	0	0	0	0	64		4	0	0	0	1	63

Şekil 4.11. Genel nesne tespit performansının arttırılması için hibrit yapı kullanılarak elde edilen karışıklık matrisleri

Karışıklık matrisleri ile hesaplanan doğruluk oranı hem YOLOv5s+Detectron2 hem de YOLOv5x+Detectron2 yapıları için %98 olarak elde edilmiştir. Bu yapıların sunduğu oranlara ilişkin istikrarın sebebi regresyon/sınıflandırma temelli bir değerlendirmeye sahip olan YOLOv5s ve YOLOv5x mimarilerinin bölge önerisi tabanlı değerlendirme yaklaşımını kullanan Detectron2 R50-FPN mimarisi ile hibrit kullanılmasıdır. Böylece farklı değerlendirme yaklaşımlarını kullanan yöntemlerin gücü birleştirilmiştir.





5. TARTIŞMA

Dünyadaki başlıca ölüm nedenlerinden biri olan kanser, farklı yaş gruplarında görülebilen bir malignitedir [2, 4]. Erken teşhis, canlılığın sürdürülebilmesi açısından hayati bir eylemdir [2]. Ancak tıp dünyasında hastalığın tanısı için kullanılan geleneksel yöntemlerin bazıları net çıktılar sunmaz. Ek olarak uzmanın bilgisi, tecrübesi, fiziksel yoğunluğu ya da zihinsel yorgunluğu gibi parametreler de tanı doğruluğunu etkiler.

Bu bölümde en çok görülen kanser tiplerinden biri olan lösemi hastalığının [3] analiz edilen 7 farklı kısmı için elde edilen sonuçları tartışılmıştır.

5.1. DNA Genom Dizilimlerindeki Ekson ve İntron Bölgelerinin Sınıflandırılmasının Sonucunda Ulaşılan Çıktılar

Bu bölümde lösemi hastalığının temel türlerinden olan ALL ve KML malignitelerin tanısında önemli bir biyobelirteç olarak nitelendirilen BCR-ABL geni üzerinde ekson ve intron bölgelerinin sınıflandırılması için alternatif bir yaklaşım önerilmiştir. Çünkü ekson bölgeleri, canlılığın büyüme ve gelişme süreçlerinde yaşanan düzenlemelerin, hücrelerin çoğalma ve sonlanma durumlarının, kök hücrelerin yaşadıkları değişikliklere ilişkin bilgilerin, kanserin gelişmesi ile ilgili değerlendirmelerin ve genlerin mutasyon geçirip geçirmediğine ilişkin hayati çıkarımların sağlandığı kısımdır [8, 12]. Bununla birlikte ekson bölgelerinin ayırt edilmesi araştırmacılar tarafından zorlu bir problem olarak ifade edilmektedir [9, 11, 13, 14].

Literatürde ekson ve intron bölgelerinin ayırt edilmesi amacıyla farklı birçok çalışma gerçekleştirilmiştir. [13, 49, 148-150] çalışmalarında NCBI veri kümesinden tedarik edilen gen dizilimleri üzerinde analizler yapılmıştır. Bu doğrultuda [148] çalışmasında walsh kodları üzerinde temellenen sayısal haritalama tekniği sunulmuştur. [13] çalışmasında uyarlamalı ayarlı Kaiser pencereleme ile trigonometrik haritalamanın hibrit kullanıldığı bir yaklaşım önerilmiştir [149] çalışmasında zaman ve frekans filtreleme tekniği olan S dönüşümüne dayalı bir yaklaşım inşa edilmiştir. [49]

çalışmasında uyarlanabilir temsil şeması olarak kullanılan tahmin tekniği önerilmiştir. Dizilimlerin sayısallaştırılması sürecinde DNA yapısının tam olarak yansıtılması için [150] çalışmasında entropi temelli haritalama tekniği tasarlanmıştır. Ensemble veri kümesinin tercih edildiği [8] çalışmasında ENSG00000103313 numaralı DNA dizilimi kullanılmıştır. 9 farklı haritalama tekniği ile sayısallaştırılan DNA dizilimi, Fourier dönüşümü, destek vektör makineleri ve k-en yakın komşu algoritması ile sınıflandırılmıştır. USCS Assembly gen bankası vasıtasıyla elde edilen insan genomunun kısa dizileri (GRCh37/hg19) kullanılarak ekson ve intron bölgelerinin sınıflandırıldığı [101] çalışmasında sabit haritalama teknikleri ile fiziko kimyasal özellik tabanlı haritalama teknikleri için sınıflandırma kesinliği karşılaştırılmıştır.

Literatürdeki çalışmalara kıyasla bu bölümün özgün yanı, belirli bir kanser türü olan lösemi hastalığı ile ilişkili DNA dizilimlerinin analizidir. Bildiğimiz kadarı ile literatürde bu amaç doğrultusunda gerçekleştirilen bir çalışma yoktur. Bununla birlikte sınıflandırma sürecinde başarı çıktıları elde edebilmek için transfer öğrenme mimarileri kullanılmıştır. Böylece güçlü bilgisayarlarda geliştirilen ve inşa edilen modellere yapılan ince ayar sonucunda keşfedilen başarılı ağırlık değerlerinden faydalanılmıştır. Aynı zamanda verilerin içerisindeki ilişkilerin, desenlerin ve periyodiklik durumlarının analiz edilmesini sağlayan sinyal işleme yaklaşımları ile genomik verilere ilişkin karanlık noktaların aydınlatılması sağlanmıştır.

Ekson ve intron bölgelerinin tespiti için farklı spektral yoğunluğa sahip spektrogramlar ile zenginleştirilmiş veri kümesi üzerinde %100 başarı oranı gösteren bu çalışmanın literatüre katkı sağlayacağı düşünülmektedir. Gelecekte DNA konfigürasyonunun belirsiz yapısı için uyarlanabilir bulanık mantık algoritması ile ekson ve intron bölgelerinin sınıflandırılması planlanmaktadır. Böylece farklı DNA baz uzunluğundaki dizilimlerinden oluşmayan spektrogramların yer aldığı eğitim veri kümesi için nihai performansın artırılması hedeflenmektedir.

5.2. DNA Dizilimleri ile ALL ve KML Malignitelerinin Sınıflandırılmasının Sonucunda Ulaşılan Çıktılar

Bu bölümde lösemi hastalığının temel türlerinden olan ALL ve KML malignitelerinin DNA dizilimleri kullanılarak sınıflandırılması için alternatif bir yaklaşım önerilmiştir. Genetik temelli malignitelerden olan ALL ve KML türlerinin analizi için kullanılan

dizilimler DNA’da bir mutasyon sonucunda oluşan BCR-ABL genidir. Bu gen kullanılarak tıp dünyasında PCR tekniği ile lösemi kanserine ait bazı türlerin ayırt edilmesine ilişkin değerlendirmeler yapılmaktadır. Ancak non-invaziv yöntemlerin ve alternatif yaklaşımların bulunmasına yönelik halen çalışmalar devam etmektedir [18-22]. Bu bölüm ile mevcut çalışmalara ek olarak bilişim destekli bir değerlendirme süreci inşa edilmiştir.

Literatürde malignitelerin sınıflandırılması amacıyla farklı birçok çalışma gerçekleştirilmiştir. Bu kapsamda [151] çalışmasında 3C(chromosome conformation capture) tabanlı dizilimlerin kromatin etkileşimlerinin tahmin edildiği derin öğrenme modelleri incelenmiştir. Karaciğer kanseri ve normal karaciğer genlerine ilişkin sınıflandırmanın yapıldığı [152] çalışmasında dizilimler sayısallaştırılmıştır ve sayısallaştırılan DNA sinyali spektrogram olarak ifade edilmiştir. Ardından yapay zeka tabanlı mimariler kullanarak spektrogramlar sınıflandırılmıştır. [153] çalışmasında transkripsiyon sürecini başlatan ve organizmadaki genlerin ne zaman ve nerede ifade edileceğini düzenleyen bir DNA dizisini belirlemek için NLP temelli bir çalışma gerçekleştirilmiştir. DNA diziliminin tek bir geninde meydana gelen mutasyonların saptanması amacıyla [154] çalışmasında CNN-LSTM temelli MWHHO algoritması geliştirilmiş ve NCBI SARS-COV-2, ENCPP ve UBE3A veri kümeleri için değerlendirmeler sağlanmıştır. [155] çalışmasında genetik hastalıklar ve kanser ilerlemesi ile ilişkilendirilen G4’lü yapıların tahminine ilişkin evrişimli sinir ağı temelli G4detector yapısı sunulmuştur.

Literatürdeki çalışmalara kıyasla bu bölümün özgün yanı, belirli bir kanser türü olan lösemi hastalığının temel türlerinden olan ALL ve KML malignitelerinin sınıflandırılması için DNA dizilimlerinin analiz edilmesidir. Bildiğimiz kadarı ile literatürde bu amaç doğrultusunda gerçekleştirilen bir çalışma yoktur. Ek olarak DNA baz uzunluğundan bağımsız olarak gerçekleştirilen sınıflandırma sürecinde kullanılan ANFIS yapısı ile yapay zekanın veriler üzerindeki öğrenme kabiliyeti ve bulanık mantığın veriler üzerindeki çıkarım gücü birleştirilmiştir. Aynı zamanda DNA yapısının bulanık konfigürasyonuna karşın bulanık setler için net girdi desenleri eşleyen bulanıklaştırıcı modül ile ANFIS yapısının malignitelere ilişkin tespit gücünden faydalanılmıştır.

Net çıkarımlar sağlamak için genetik temelli analizler doğrultusunda DNA dizilimleri üzerinde gerçekleştirilen çalışmalara ek olarak mikrodizi teknolojisinden de yararlanılmaktadır. Ancak mikrodizi teknolojisi ile elde edilen mikrodizi veri kümesi, hastalıkla ilişkili olmayan gen ekspresyon verilerini de barındırmaktadır. Bu durum ekstra zaman ve maliyet çıktısı oluşturmaktadır. Aynı zamanda ilişkisiz genler ilişkili genler üzerinde gürültü oluşturacağı için yanlış çıkarımlar üretme ihtimaline sahiptir. Genetik çıkarım sağlayan DNA dizilimlerinin doğrudan analiz edilmesi ile bu problemler çözülmüş ve kişiye ilişkin hayati bilgiler barındıran yapılar kullanılarak net çıktılar üretilmiştir.

ALL ve KML malignitelerinin tespiti için farklı istatistiksel ve yapısal bilgiler kullanılarak oluşturulan veri kümesi üzerinde %80 başarı oranı elde edilen bu çalışmanın literatüre katkı sağlayacağı düşünülmektedir. Gelecekte DNA dizilimlerinin sayısallaştırılması için bulanık mantığın kullanıldığı yeni bir haritalama tekniğinin keşfedilmesi planlanmaktadır.

5.3. Mikrodizi Veri Kümesi Kullanılarak ALL ve AML Malignitelerinin Sınıflandırılmasının Sonucunda Ulaşılan Çıktılar

Bu bölümde lösemi hastalığının temel türlerinden olan ALL ve AML malignitelerinin tanısında moleküler değerlendirme sağlamak için alternatif bir yaklaşım önerilmiştir. Genlerle ilişkili kanser hastalıkları için umut kaynağı olduğu ifade edilen mikrodizi teknolojisinin [27] kullanıldığı bu yaklaşım ile mikrodizi veri kümesindeki genler arasındaki ilişkiler özellik seçim algoritmaları kullanılarak açığa çıkartılmış ve potansiyel genler seçilmiştir. Ardından seçilen genler üzerinde zamansal ve uzaysal bilgi çıkarımı sunan DGCNN mimarisi ile ilişkiye duyarlı ve görüntü seviyesinde temsil bilgileri elde edilmiştir. Tasarlanan bilgisayar destekli bu yapı ile mikrodizi veri kümesinin sahip olduğu dezavantajlar tolere edilmiştir ve avantajlar ortaya çıkarılmıştır. Böylece doğru tanı ile hastaya uygun bir tedavi protokolünün seçilmesi için doktorlara fikir verilmesi hedeflenmiştir.

Literatürde ALL ve AML malignitelerinin sınıflandırılması amacı ile farklı birçok çalışma gerçekleştirilmiştir. [24, 28, 65, 156-158] çalışmalarında ALL-AML veri kümesinden tedarik edilen gen ekspresyon verileri kullanılmıştır. Bu doğrultuda [65] çalışmasında kanser sınıflandırması için korelasyon analizine dayalı yeni bir topluluk

yöntemi önerilmiştir. Kaba küme temelli özellik seçimi ve sınıflandırma yaklaşımlarının uygulandığı [156] çalışmada gen seçiminde uygun alt kümelerin belirlenmesi ile yüksek sınıflandırma performansının elde edileceği ifade edilmiştir. Mikrodizi veri kümeleri ile kanser sınıflandırmasının gerçekleştirildiği [157] çalışmada vector-valued regularized kernel function approximation(VVRKFA) yöntemi sunulmuştur. [158] çalışmada bulanık tercih tabanlı kaba küme ile yarı denetimli SVM yaklaşımları birlikte kullanılmıştır. Gen seçimi ve sınıflandırma adımlarının sırasıyla kullanıldığı bu çalışma ile kanser teşhisi ve ilaç tanısı için bir tahmin süreci inşa edilmiştir. Kanser tanısı için [24] çalışmada Simetrik Belirsizlik ve SVM yaklaşımlarının birlikte kullanıldığı Aktif Öğrenme modeli geliştirilmiştir. Mikrodizi veri kümelerinde hastalıklarla ilişkili genlerin seçilmesi amacıyla [28] çalışmada Temel Bileşenler Analizi ve Lojistik Regresyon yaklaşımları üzerinde temellenen bir metodoloji önerilmiştir. Ardından 8 farklı sınıflandırıcı ile önerilen yöntemin etkinliği değerlendirilmiştir.

Literatürdeki çalışmalara kıyasla bu bölümün özgün yanı lösemi hastalığının temel türlerinden olan ALL ve AML malignitelerinin lösemi hastalığı ile eşleştirilerek analiz edilmesidir. Genel olarak literatürdeki çalışmalar, inşa edilen hibrit yapının ya da keşfedilen bir algoritmanın farklı mikrodizi veri kümeleri üzerinde karşılaştırma yapılarak değerlendirilmesini içerir. Bununla birlikte bu çalışmada elde edilen deneysel sonuçlar lösemi hastalığının bir bütün olarak analiz edilmesine yardımcı olan çıktılar sunmaktadır. Ek olarak bildiğimiz kadarı ile mikrodizi veri kümeleri üzerinde DGCNN yöntemi ilk defa bu çalışmada kullanılmıştır. Böylece hem zamansal hem de uzaysal bağımlılığın kullanılması ile ALL ve AML malignitelerine ilişkin güçlü çıkarımlar sağlanmıştır.

ALL ve AML malignitelerinin tespiti için gen ekspresyon veri kümesi üzerinde %93.3 başarı oranı gösteren bu çalışmanın literatüre katkı sağlayacağı düşünülmektedir. Gelecekte özgün bir optimizasyon algoritmasının keşfedilmesi planlanmaktadır. Böylece potansiyel genlerin seçildiği veri kümesi için nihai performans iyileştirilecektir.

5.4. Mikrodizi Veri Kümesi Kullanılarak ALL, AML ve MLL Sınıflandırılmasının Sonucunda Ulaşılan Çıktılar

Bu bölümde lösemi hastalığının temel türlerinden olan ALL, AML ve MLL malignitelerinin tanısında moleküler değerlendirme sağlamak için alternatif bir yaklaşım önerilmiştir. Genlerle ilişkili kanser hastalıkları için umut kaynağı olduğu ifade edilen mikrodizi teknolojisinin [27] kullanıldığı bu yaklaşımda ilk olarak mikrodizi veri kümesinde yer alan genler arasındaki ilişkiler özellik seçim algoritması kullanılarak açığa çıkartılmış ve potansiyel genler seçilmiştir. Ardından seçilen genler LSTM mimarisine girdi olarak verilmiştir. LSTM mimarisi genellikle metin analizi ve zaman serilerinin analizinde kullanılmaktadır. Ancak seçilen genlerin birbiri ile ilişkisini hatırlama eylemleri ile dinamik tutma yetisine sahip olduğundan dolayı bu çalışmada tercih edilmiştir. Tasarlanan bilgisayar destekli bu yapı ile mikrodizi veri kümesinin sahip olduğu dezavantajlar tolere edilmiştir ve avantajlar ortaya çıkarılmıştır. Böylece doğru tanı ile hastaya uygun bir tedavi protokolünün seçilmesi için doktorlara fikir verilmesi hedeflenmiştir.

Literatürde lösemiye ilişkin malignitelerinin sınıflandırılması amacıyla farklı çalışmalar gerçekleştirilmiştir. Bu doğrultuda [35] çalışmasında yüksek boyut sunan mikrodizi veri kümelerinin boyutunu indirgemek amacıyla fil arama tabanlı optimizasyon algoritması önerilmiştir. Ardından özellik seçim algoritması ile seçilen genler derin sinir ağı mimarisi ile sınıflandırılmıştır. Gen ekspresyon verileri içerisinden hastalıkla ilişkili uygun bir alt kümenin belirlenmesi amacıyla [68] çalışmasında uyarlanabilir KNN tabanlı gen seçim tekniği ile birlikte parçacık sürü optimizasyon yönteminin kullanıldığı özellik seçim yaklaşımı sunulmuştur. Ardından seçilen genlerin uygunluğu SVM algoritması ile test edilmiştir. Genellikle yüksek boyutlu, gereksiz ve gürültülü olarak nitelendirilen mikrodizi veri kümeleri üzerinde sınıflandırma doğruluğunu arttırmak için [70] çalışmasında kümeleme tabanlı hibrit bir gen seçim yaklaşımı önerilmiştir. Önerilen yöntemin başarısı destek vektör makinesi, K en yakın komşu algoritması ve Naive Bayes algoritması ile sınıflandırılmıştır. Kanser sınıflandırması ve kanser teşhisi için potansiyel genlerin seçildiği [88] çalışmasında karınca kolonisi algoritması ile optimize edilen hücresel öğrenme otomatına dayalı sarmalayıcı bir yaklaşım sunulmuştur. Bu yaklaşımın başarısı KNN, SVM ve Naive Bayes algoritmaları ile değerlendirilmiştir. [74]

çalışmasında özyinelemeli memetik algoritma (RMA) önerilmiştir. Geliştirilen bu yaklaşımın memetik algoritma ve genetik algoritmaya kıyasla daha başarılı olduğu belirtilmiştir. Yüksek boyut sunan veri kümelerinde sınıflandırma performansının iyileştirilmesi amacıyla [87] çalışmasında dağıtılmış öznitelik yaklaşımı önerilmiştir. Ardından seçilen özelliklerin alt kümeleri ile sınıflandırma sağlanmıştır.

Literatürdeki çalışmalara kıyasla bu bölümün özgün yanı lösemi hastalığına ilişkin ALL, AML ve MLL malignitelerinin lösemi hastalığı ile eşleştirilerek analiz edilmesidir. Çünkü literatürdeki çalışmalar genel olarak yüksek boyutlu mikrodizi veri kümeleri için var olan, hibrit kullanılan ya da yeni geliştirilen özellik seçim algoritmaları ile uygun genlerin elde edilmesini ve ardından farklı sınıflandırıcılar ile spesifik olmayan hastalıklara ilişkin kategorilerin ayırt edilmesini içerir.

Bu çalışmanın güçlü kısmı LSTM mimarisinin mikrodizi veri kümeleri için uygun olan fiziksel ünitelerinin ve işlevselliklerinin başarılı bir şekilde modele entegre edilmesidir. Kanseri hastalıklarının teşhisinde umut kaynağı olan mikrodizi teknolojisi vasıtasıyla elde edilen ilişkili genlerin uzun zaman periyodunda hatırlama işlevselliğine sahip olan LSTM mimarisi ile sınıflandırılması tıp alanında geliştirilen güçlü bir karar destek sisteminin varlığına işaret etmektedir.

ALL, AML ve MLL malignitelerinin sınıflandırılması için seed parametresi kullanılarak değiştirilen 100 farklı test kümesinde ortalama %89.883 başarı oranı üretilmiştir. Bu çalışmanın literatüre katkı sağlayacağı düşünülmektedir. Gelecekte bulanık mantık algoritmasıyla LSTM mimarisinin birlikte kullanılması ile nihai performansın iyileştirilmesi planlanmaktadır.

5.5. Dijitalleştirilmiş Periferik Kan Yayma Görüntüleri Kullanılarak ALL Malignitesinin Tespitinde Ulaşılan Çıktılar

Bu bölümde lösemi hastalığının temel türlerinden olan ALL malignitesinin tanısında önemli bir ayıraç olarak nitelendirilen blast hücrelerinin tespiti için alternatif bir yaklaşım önerilmiştir. Dijitalleştirilmiş periferik kan yayma görüntülerinin kullanıldığı bu yaklaşım ile manuel değerlendirmelere ilişkin zorluklar tolere edilmiştir. Aynı zamanda tespiti yapılan blast hücrelerin sayısına ilişkin verilen bilgi ile tedavi sonrasındaki süreç için doğru kararların alınmasında doktorlara fikir verilmeside planlanmıştır. Çünkü, ALL türü kan dolaşımına ve diğer hayati organlara

hızlı bir şekilde yayılır. Ek olarak halsizlik, anemi, nefes darlığı, eklem ya da kemik ağrıları gibi genel belirtiler gösterir [38]. Bu nedenle canlılığın sürekliliği için erken tanı önemlidir.

Literatürde lösemnin temel türü olan ALL malignitesinin teşhisi için ALL-IDB veri kümesinin kullanıldığı birçok çalışma [5, 36, 159-163] gerçekleştirilmiştir. Bu doğrultuda [159] çalışmasında MI+LDP ve SCA temelli derin CNN sınıflandırıcı ile lösemi tespit modülü geliştirilmiştir. GA ve ANN yöntemlerinin kullanıldığı [36] çalışmasında lökosit hücre segmentasyonu için yeni bir teknik sunulmuştur. [160] çalışmasında ALL malignitesinin alt tiplerini (L1, L2 ve L3) tanımlamak amacıyla KNN/SVM ve ANN yöntemleri kullanılmıştır. Lenfoblast ve sağlıklı hücrelerin ayırt edilmesi ve ALL hastalığının değerlendirilmesi amacıyla [161] çalışmasında VGG/GoogleNet ve Alexnet mimarileri kullanılmıştır. [162] çalışmasında SESSA ve VGGNet yaklaşımları ile lösemi hastalığına ilişkin görüntü sınıflandırması yapılmıştır. Transfer öğrenme mimarilerinin kullanıldığı [5] çalışmasında görüntülerden lösemi tespiti yapılmıştır. [163] çalışmasında YOLOv4 yöntemi kullanılarak blast hücre tespiti gerçekleştirilmiştir.

Literatürdeki çalışmalara kıyasla bu bölümün özgün yanı lösemi hastalığının temel türü olan ALL malignitesinin tespitinde önemli bir biyobelirteç olan blast hücrelerinin maksimum 3-4 saniye içerisinde gerçek zamanlı olarak tespit edilmesi ve blast hücre sayısına ilişkin bilginin üretilmesidir. Bununla birlikte son 3 yılda yayınlanan çalışmalar [5, 36, 159-163], ya ALL-IDB1 ya da ALL-IDB2 veri kümesi ile eğitilmiştir. Fakat bu çalışmada modelin eğitimi bu iki veri kümesinin birleştirilmesi ile oluşturulan yeni veri kümesinde gerçekleştirilmiştir. Böylece eğitilen bu model ile farklı büyütme oranlarına sahip periferik kan yayma görüntülerindeki blast hücre tespiti mümkün olmuştur. Ek olarak bu çalışmada blast hücrelere ilişkin verilen sayısal bir bilgi ile ALL tanısının sağlanmasına yardımcı olmanın yanı sıra tedaviye verilen tepki de değerlendirilecektir.

ALL malignitesinin tanısı için periferik kan örneklerine ilişkin mikroskopik görüntülerin yer aldığı ALL-IDB1+ALL-IDB2 veri kümesi üzerinde %98.87 başarı oranı gösteren bu çalışmanın literatüre katkı sağlayacağı düşünülmektedir. Çünkü manuel değerlendirmeler; zahmetli, zaman alıcı, kişinin bilgi ve tecrübesine göre değişebilen parametrelere sahiptir. Öte yandan bu sürecin otomatik sağlandığı akış

sitometrisi teknolojisi de pahalıdır, deneyimli personel gerektirir ve gerçek zamanlı çıktı üretmez. Bu nedenle bilgisayar destekli ve gerçek zamanlı çıktı üreten bu sistemin hem tanı tespiti hem de tedaviye verilen yanıtın değerlendirilmesinde doktorlar tarafından tercih edilmesi hedeflenmektedir. Gelecekte basit bir hiyerarşiye sahip bu sistem için blast sayısına ilişkin bir çıktı üretmek yerine ilgili görüntüdeki blast hücre yoğunluğuna ait bir bilgi verilmesi planlanmaktadır.

5.6. Mikrodizi Veri Kümesi Kullanılarak T-ALL, B-ALL ve T-LL Malignitelerinin Sınıflandırılmasının Sonucunda Ulaşılan Çıktılar

Bu bölümde lösemi hastalığına ilişkin T-ALL, B-ALL ve T-LL malignitelerinin tanısında moleküler değerlendirme sağlamak için alternatif bir yaklaşım önerilmiştir. Genlerle ilişkili kanser hastalıkları için umut kaynağı olduğu ifade edilen mikrodizi teknolojisinin [27] kullanıldığı bu yaklaşımda mikrodizi veri kümesinde yer alan genler arasındaki ilişkiler özellik seçim algoritması kullanılarak açığa çıkartılmış ve potansiyel genler seçilmiştir. Ardından seçilen genler ANFIS yapısına girdi olarak verilmiştir. ANFIS, yapay zekanın veriler üzerindeki öğrenme kabiliyeti ile bulanık mantığın veriler üzerindeki çıkarım yetisini birleştiren bir sinir ağıdır. Optimizasyon algoritmaları ile güçlendirilen bu sinir ağı vasıtasıyla üretilen çıktıların gücü topluluk öğrenme yaklaşımı ile daha istikrarlı ve kararlı hale getirilmiştir. Ek olarak tasarlanan bilgisayar destekli bu yapı ile mikrodizi veri kümesinin sahip olduğu dezavantajlar tolere edilmiştir ve avantajlar ortaya çıkarılmıştır. Böylece doğru tanı ile hastaya uygun bir tedavi protokolünün seçilmesi için doktorlara fikir verilmesi planlanmıştır.

Literatürde mikrodizi veri kümeleri ile kanser sınıflandırması yapmak amacıyla farklı çalışmalar mevcuttur. Bu doğrultuda [24] çalışmada kanser araştırmaları için öznitelik seçim yaklaşımı ve destek vektör makinesi algoritmasının birlikte kullanıldığı bir model geliştirilmiştir. [164] çalışmada genetik algoritma ve destek vektör makineleri birlikte kullanılmış ve kanser araştırmalarında başarılı bir çıktı sağlandığı belirtilmiştir. [27] çalışmada parçacık sürü optimizasyon yöntemi ile veri setinden seçilen uygun genler ssEAM sinir ağı mimarisi ile sınıflandırılmıştır. [63] çalışmada kanserle ilgili faktörleri belirlemek için bulanık bir kaba kümeleme yöntemi ve yarı denetimli destek vektör makineleri kullanılmıştır. [66] çalışmada, mikrodizi veri kümesinden ilişkili gen dizilerini belirlemek için hiper kutu kapatma

yöntemi önerilmiştir. Mikrodizi veri kümesinden uygun genleri seçmek için [35] çalışmasında, fil arama ve ateş böceği arama optimizasyon algoritmaları kullanılmıştır. Daha sonra seçilen genler derin sinir ağları tarafından sınıflandırılmıştır. [68] çalışmasında uyarlanabilir k en yakın komşu ve parçacık sürüsü optimizasyon yöntemi ile mikrodizi veri kümesinden seçilen potansiyel genler destek vektör makineleri algoritması ile sınıflandırılmıştır. Yapay arı kolonisi optimizasyon algoritmasının ve genetik algoritmanın birlikte kullanıldığı [34] çalışmasında önerilen genetik arı kolonisi algoritması ile seçilen uygun genler destek vektör makinesi algoritması ile sınıflandırılmıştır.

Literatürdeki çalışmalara kıyasla bu bölümün özgün yanı, belirli bir kanser türü olan lösemi hastalığı ile ilişkili T-ALL, B-ALL ve T-LL malignitelerinin sınıflandırılmasıdır. Bildiğimiz kadarı ile literatürde bu amaç doğrultusunda gerçekleştirilen bir çalışma yoktur. Bununla birlikte kullanılan her bir yöntemin uygunluğu çalışmanın gidişatına uygun bir şekilde seçilmiştir. Böylece basit bir hiyerarşi ve güçlü bir çıkarım süreci oluşturulmuştur. Bu durum 5 neden ile açıklanmaktadır.

Bu çalışmadaki performansın ilk sebebi kullanılan metasezgisel optimizasyon algoritmalarının türeysiz olması nedeniyle yerel optimumlardan kaçınmasıdır [147]. Böylece veri kümesinden uygun genler seçilmiştir. İkinci sebebi net göstergeleri olmayan malignitelerin sınıflandırılması için bulanık çıktılar üreten ANFIS yapısının kullanılmasıdır. Üçüncü sebebi ANFIS yapısı ile üretilen çıkarımların parçacık sürü optimizasyon algoritması ve yapay arı kolonisi algoritması ile iyileştirilmesidir. Dördüncü sebebi kolektif öğrenme yaklaşımı kullanarak birden fazla algoritmaya ilişkin gücün birleştirilmesi ile daha güçlü ve kararlı çıktılar elde edilmesidir. Beşinci sebebi hata değerini azaltmak için yumuşak geçişler sağlayan lojistik regresyon algoritmasının kullanılmasıdır.

Uygun genlerin seçildiği mikrodizi veri kümesi üzerinde T-ALL, B-ALL ve T-LL malignitelerinin sınıflandırılmasında %86.6 başarı oranı gösteren bu çalışmanın literatüre katkı sağlayacağı düşünülmektedir. Gelecekte bulanık mantık yapısı ile yinelemeli sinir ağı yaklaşımının birlikte kullanılması hedeflenmektedir.

5.7. Beyaz Kan Hücrelerinin Tespitinde ve Sınıflandırılmasında Ulaşılan Çıktılar

Bu bölümde lösemi hastalığının tanısında önemli bir ayıraç olarak nitelendirilen lenfosit hücrelerinin tespiti için alternatif bir yaklaşım önerilmiştir. Dijitalleştirilmiş periferik kan yayma görüntülerinin kullanıldığı bu yaklaşımda manuel değerlendirmelere ilişkin zorluklar tolere edilmiştir.

Bu çalışmada lenfosit tanıma oranının güçlendirilmesi için önerilen YOLOv5x yaklaşımı tasarlanmıştır. Ancak önerilen YOLOv5x yaklaşımı kapsamında ulaşılan performansın ve genel doğruluk oranının iyileştirilmesi amacıyla hibrit bir yapı daha sunulmuştur. Böylece hücrelerin tespitinde bir optimizasyon sağlanarak daha başarılı çıkarımlar ile doktorlara fikir verilmesi hedeflenmiştir.

Literatürde kan elementlerinin sınıflandırılmasına yönelik farklı birçok çalışma gerçekleştirilmiştir. Bu doğrultuda kan yayma görüntüleri üzerinde WBC'lerin otomatik tespiti için [165] çalışmasında AEFA-C(Artificial electric field algorithm with novel velocity and position bound) yaklaşımı kullanmıştır. [166] çalışmasında beyaz kan hücre çekirdeğinin tespiti için temel bileşenler analizine dayalı uyarlamalı füzyon yapısı sunulmuştur. Bölgesel evrişimli sinir ağı yönteminin kullanıldığı [47] çalışmasında kan görüntüleri üzerinde beyaz kan hücre tiplerinin tespiti sağlanmıştır. Lightweight evrişimli sinir ağının kullanıldığı [167] çalışmasında beyaz kan hücreleri tespit edilmiş ve sınıflandırılmıştır. [168] çalışmasında ReLU ve LReLU aktivasyon fonksiyonlarının kullanıldığı GoogleNet mimarisi ile beyaz kan hücre çekirdekleri sınıflandırılmıştır. CNN ve RNN modellerini birleştiren kanonik korelasyon analizinin sunulduğu [94] çalışmasında beyaz kan hücrelerine ilişkin bir sınıflandırma yapılmıştır. Çok sınıflı destek vektör makinesi ile transfer öğrenme mimarilerinin hibrit kullanıldığı [169] çalışmasında beyaz kan hücrelerinin otomatik tespiti ve sınıflandırılması sağlanmıştır. Hızlı geleneksel sinir ağı modelinin kullanıldığı [170] çalışmasında beyaz kan hücrelerinin alt tiplerine ilişkin bir sınıflandırma sağlanmıştır. YOLO_v3, SSD, Faster R-CNN ve R-FCN yapılarının kullanıldığı [171] çalışmasında kan hücrelerinin farklı türleri tespit edilmiştir. FIAL (Fine-grained interactive attention learning) yarı denetimli yaklaşımın önerildiği [93] çalışmasında beyaz kan hücreleri sınıflandırılmıştır. Tiny ve Efficient YOLOF (TE-YOLOF) modellerinin kullanıldığı [172] çalışmasında beyaz kan hücreleri, kırmızı kan hücreleri ve trombositler tespit edilmiştir.

Literatürdeki çalışmalara kıyasla bu bölümün özgün yanı lösemi hastalığının tanısında önemli bir ayıraç olan lenfosit hücre tipinin yüksek bir doğruluk oranı ile tespiti için geliştirilen 2 farklı yapıdır. Bu doğrultuda iyileştirilmiş YOLOv5x yaklaşımının önerildiği ilk yapı ile lenfosit hücre tipinin tespitinde 0.133 oranında bir iyileştirme sağlanmıştır. Hibrit yaklaşımın önerildiği ikinci yapı ile YOLO ve Detectron2 önceden eğitilmiş modellerinden elde edilen genel doğruluk oranları %3.44 ve %14.7 arasında iyileştirilmiştir. Bazofil, eozinofil, lenfosit, monosit ve nötrofil kan hücrelerinin tespiti için Raabin Health Database veri kümesi üzerinde %98 başarı oranı gösteren bu çalışmanın literatüre katkı sağlayacağı düşünülmektedir. Gelecekte YOLOv5x mimarisinin eğitiminde kullanılan aktivasyon fonksiyonunun farklı metasezgisel algoritmalar ile optimize edilmesi planlanmaktadır. Böylece doğrusal olmayan durumların modele öğretilmesinde başarılı bir performans üretecek ve nihai başarı oranı iyileştirilecektir.

Lösemi hastalığının değerlendirilmesinde kullanılan tüm bu yaklaşımlar lösemi hastalığının bir bütün olarak analiz edilmesine yardımcı çıktılar sunmaktadır. Çünkü lösemi kanseri, tek bir hastalık değildir. Kendi içerisinde farklı alt malignitelere sahip olan birçok hastalıktır. Her bir malignite için hastalığın seyri, hastalığın süreci, hastalığın yönetimi, iyileşme ihtimali ve uygulanacak tedavi protokolü farklıdır. Aynı zamanda hastalığın tanısı için kullanılan yöntemler dahi değişiklik gösterebilmektedir. Bununla birlikte hastaya doğru tanı etiketinin konulması sürecinde tıp dünyasında kullanılan yöntemlerin belirsiz çıktılar üretme ihtimali de vardır. Örneğin, fiziksel muayene ve hastanın öyküsü tanı sürecinin ilk aşamasıdır. Ancak kesin tanı için yeterli değildir. Biyopsi yöntemi tanı sürecinde kullanılan diğer bir metottur. Fakat biyopsi yöntemi ile hastadan alınan örnek dokuda kanserli hücre yer almayabilir. Görüntüleme yöntemleri ve patolojik analizler de tanı sürecinde faydalanılan yaklaşımlardır. Fakat bu yaklaşımların manuel değerlendirilmesi; doktorların bilgisi, tecrübesi, fiziksel yoğunluğu ve zihinsel yorgunluğu ile değişmektedir. Standart bir değerlendirme imkanı sunmayan bu yöntemler kapsamında doktorlara yardımcı bir fikir vermek temel hedefimizdir. Böylece maliyet parametresi çerçevesinde hasta üzerinde uygulanacak tanı yöntemlerinin azaltılması planlanmaktadır. Bu doğrultuda lösemi hastalığı sitogenetik değerlendirme, moleküler değerlendirme, moleküler ve

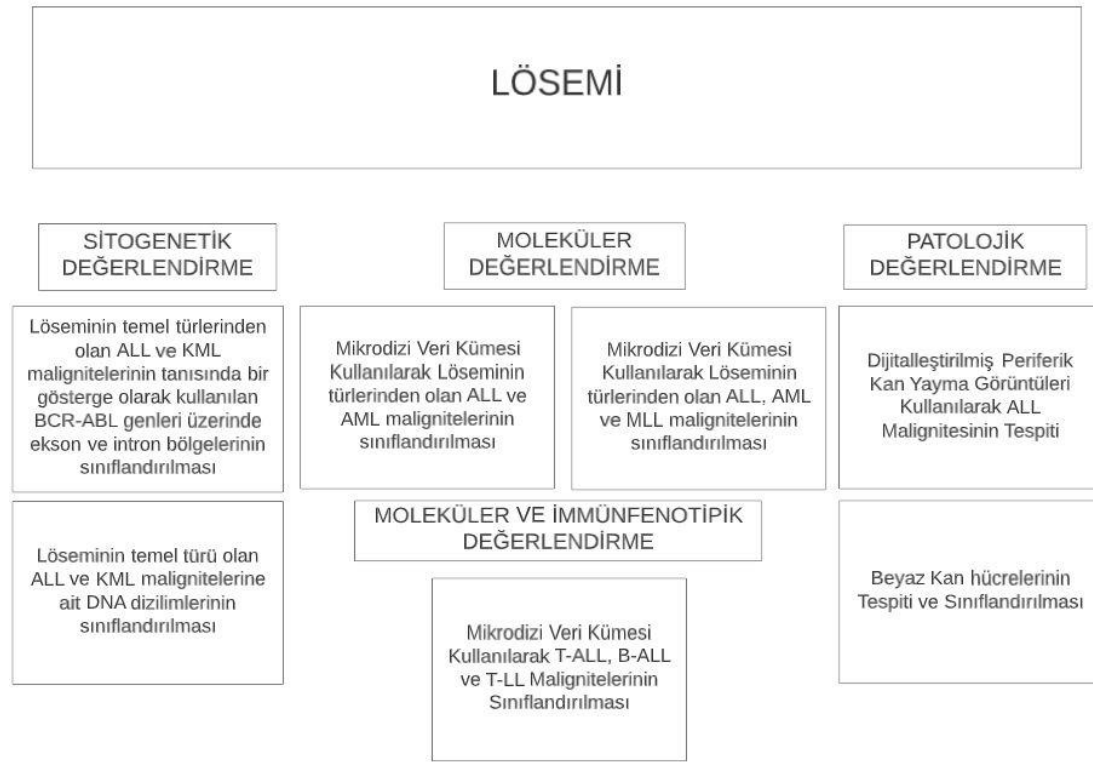
immüfenotipik deęerlendirme, patolojik deęerlendirme olarak 4 ayrı kısımda analiz edilmiştir.

İlk adımda sitogenetik deęerlendirme yapılmıştır. Böylece yaşama ilişkin devasa bilgi barındıran DNA dizilimleri üzerinden analizler sağlanılarak elde edilen çıkarımlar lösemi hastalığı ile eşleştirilmiştir. Bu kısımda üretilen sonuçlar lösemi hastalığının bir bütün olarak deęerlendirilmesine yardımcı olan çıktılar sunmaktadır.

İkinci adımda moleküler deęerlendirme yapılmıştır. Genler arası ilişkilerin ortaya çıkarılması ve anormal durumların keşfedilmesinde tercih edilen mikrodizi teknolojisi ile genlerin fonksiyonel protein yapılarına dönüşüm süreci deęerlendirilmiştir. Bu kısımda üretilen sonuçlar lösemi hastalığının bir bütün olarak deęerlendirilmesine yardımcı olan çıktılar sunmuştur.

Üçüncü adımda moleküler ve immüfenotipik deęerlendirme yapılmıştır. Genomik verilere ilişkin karanlık noktaların aydınlatılmasında umut kaynağı olan mikrodizi teknolojisi kullanılarak hücre tiplerinin tespiti sağlanmıştır. Bu kısımda gen ekspresyon seviyeleri üzerinden elde edilen çıkarımlar lösemi hastalığının bir bütün olarak deęerlendirilmesine yardımcı olan çıktılar sunmuştur.

Dördüncü adımda sitogenetik ve moleküler deęerlendirme yaklaşımlarına destek çıktılar üretmek ve lösemi hastalığı ile eşleştirerek bütüncül bir deęerlendirme yapmak amacıyla patolojik analiz yapılmıştır. İlk üç adımda gerçekleştirilen genetik temelli tanı sürecinin yanı sıra dördüncü adımda morfolojik temele dayanan bir tanı süreci ile doktorların hizmetine sunmak amacıyla gerçek zamanlı çıktılar üretilmiştir. Ek olarak tedavi sonrası deęerlendirmelerin yapılması için sunduğu yardımcı sonuçlar ile lösemi hastalığının deęerlendirilmesine yeni bir bakış açısı kazandırmıştır. Lösemi malignitesinin bir bütün olarak deęerlendirilmesi amacıyla bu tezde kullanılan tüm yaklaşımlar Şekil 5.1’de gösterilmiştir.



Şekil 5.1. Lösemi kanserinin bir bütün olarak değerlendirilmesi amacıyla bu tezde kullanılan tüm yaklaşımlar

Bu çalışmada, hayata ilişkin devasa bilgi barındıran DNA dizilimleri üzerinden analizlerin sağlandığı sitogenetik değerlendirme ile yaşamsal olayların sembolik DNA yapıları kullanılarak değerlendirilmesi sağlanmıştır. Bu aşama için tercih edilen her bir sayısal haritalama tekniği, kullanılan sayısal sinyal işleme yaklaşımları ve dizilimlerden çıkarılan özellikler; genomik noktaların aydınlatılmasında süreci doğrudan etkileyen önemli parametrelerdir. Öte yandan değerlendirilmesi hedeflenen DNA baz uzunluklarının sabit olmaması değerlendirme sürecini olumsuz etkilemektedir.

Genler arası ilişkilerin ortaya çıkarılmasında ve anormal durumların keşfedilmesinde kullanılan mikrodizi teknolojisi ile genlerin fonksiyonel protein yapılarına dönüşüm süreci değerlendirilmiştir. Fakat bu aşamada faydalanan mikrodizi veri kümeleri, hastalıkla ilişkili ve hastalıkla ilişkisiz genler barındırmaktadır. Bu durum gen ekspresyon verilerinin değerlendirilme aşamasında hesaplama yükü meydana getirmekte ve hastalıkla ilişkisiz genlerin hastalıkla ilişkili genler üzerinde iş yükü oluşturmaya neden olmaktadır. Bu dezavantajı yok etmek amacıyla özellik seçim

algoritmaları kullanılarak potansiyel genler seçilmiştir. Ancak özellik seçim algoritmaları ile seçilen genlerin uygunluğu malignitelerin sınıflandırma başarısını etkilemektedir.

Avantajlara ve dezavantajlara sahip olan sitogenetik ve moleküler değerlendirme yaklaşımları, genetik temelli olgular üzerinden analiz imkanı sağlamaktadır. Yaklaşımlara ilişkin tüm olumsuz durumların bilgisayar destekli sistemler ile tolere edilmesi çalışmanın ana omurgasını oluşturmaktadır. Bununla birlikte çalışma kapsamında kullanılan patolojik değerlendirme yaklaşımı ile genetik temelli çıktıların morfolojik değerlendirme ile desteklenmesi hedeflenmiştir. Tanı koyma sürecinin yanı sıra ALL malignitelerine sahip hastalar için tedaviye verilen yanıtın gerçek zamanlı olarak değerlendirilmesini de sağlayan patolojik değerlendirme ile lösemi hastalığı geniş bir ölçekte değerlendirilmiştir.

Geniş bir ölçekte değerlendirilen lösemi hastalığı için inşa edilen bilgisayar destekli bu sistemlerin doktorlara yardımcı bir fikir sunması beklenmektedir. Elde edilen tüm sonuçlar, lösemi hastalığının bir bütün olarak analiz edilmesini sağlayacak ve güçlü bir karar destek sistemi inşa edecektir.



6. SONUÇLAR

Tıp ve bilişim dünyası hastanın vücut bütünlüğüne zarar vermeden uygulanması kolay tedavi yöntemleri geliştirmek ve alternatif tanı yaklaşımları keşfetmek amacıyla çok disiplinli bir çalışma alanı oluşturmuştur. Özellikle güncel ve belirli bir uygulama alanı olan sağlıkta yapay zeka çalışmaları ile çeşitli hastalıklara ilişkin tanı sürecinde başarılı karar destek sistemleri inşa edilmiştir.

Bu çalışmada sağlıkta yapay zeka çalışmaları çerçevesinde kanser türleri arasında sıklıkla görülen lösemi kanseri incelenmiştir. Güçlü bir karar destek sistemi oluşturmak için 7 farklı hedef çerçevesinde lösemi hastalığı analiz edilmiştir. Genetik temelli tanı imkanı sağlayan sitogenetik ve moleküler analizin yanı sıra patolojik analizde yapılmıştır. Gerçek zamanlı çıktı üreten ve standart ölçütler doğrultusunda çıkarım gerçekleştiren patolojik analiz sonuçları, genetik tanı imkanı sağlayan sitogenetik ve moleküler analiz sonuçları ile birlikte değerlendirilmiştir. Değerlendirme sonucunda patolojik analiz çıktılarının lösemi hastalığının tespit ve tedaviye verilen yanıt sürecinde genetik temelli değerlendirme imkanı sağlayan sitogenetik ve moleküler analiz sonuçlarını desteklediği görülmüştür.

Lösemi kanserinin gizemli yapısının bir bütün olarak değerlendirildiği bu çalışmada gerçekleştirilen 7 farklı analiz yaklaşımı ile bilgisayar destekli bir sistem oluşturulmuştur. Doktorlara fikir vermek amacıyla inşa edilen bu sistemin gelecekteki tanı süreçlerinde yer alması hedeflenmektedir.



KAYNAKLAR

- [1] Kocabıyık, V .B. (2011). *ALL ve KML 'li hastalarda BCR ve ABL genlerindeki mutasyonların incelenmesi* [Yüksek Lisans Tezi]. Selçuk Üniversitesi.
- [2] Jauhari, S., & Rizvi, S. A. M. (2014). Mining gene expression data focusing cancer therapeutics: A digest, *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 11(3), 533–547, doi: 10.1109/TCBB.2014.2312002.
- [3] Hsu, C. H., Chen, X., Lin, W., Jiang, C., Zhang, Y., Hao, Z., & Chung Y. C., (2021). Effective multiple cancer disease diagnosis frameworks for improved healthcare using machine learning, *Measurement*, 175, 2021, doi: 10.1016/j.measurement.2021.109145.
- [4] Aydın, G. (2017). *Quercetin'in KML kök hücreleri üzerine sitotoksik etkilerinin moleküler düzeyde incelenmesi* [Yüksek Lisans Tezi]. Erciyes Üniversitesi.
- [5] Anilkumar, K. K., Manoj, V. J., & Sagi, T. M. (2021). Automated detection of leukemia by pretrained deep neural networks and transfer learning: A comparison, *Medical Engineering and Physics*, 98, 8–19, doi: 10.1016/j.medengphy.2021.10.006.
- [6] A Guide to Leukemia (2022). What is Leukemia? <https://www.healthline.com/health/leukemia#risk-factors%0A>.
- [7] Khodaei, A. Feizi-Derakhshi M. R., & Mozaffari-Tazehkand, B., (2020). A pattern recognition model to distinguish cancerous DNA sequences via signal processing methods, *Soft Computing*, 24(21), 16315–16334, doi: 10.1007/s00500-020-04942-4.
- [8] Das, B., & Türkoglu, I., (2016). Sayısal haritalama teknikleri ve Fourier dönüşümü kullanılarak DNA dizilimlerinin sınıflandırılması, *Journal of the Faculty of Engineering and Architecture of Gazi University*, 31(4), 921–932, 2016, doi: 10.17341/gazimmfd.278447.
- [9] Barman, S., Saha, S., Mandal, A., & Roy, M. (2012). Prediction of protein coding regions of a DNA sequence through spectral analysis, *IEEE/OSA/IAPR International Conference on Informatics, Electronics & Vision*, doi: 10.1109/ICIEV.2012.6317389.
- [10] Hota, M. K., & Srivastava, V. K. (2010). Performance analysis of different DNA to numerical mapping techniques for identification of protein coding regions using tapered window based short-time discrete Fourier transform, *ICPCES 2010 - International Conference on Power, Control and Embedded Systems*, doi: 10.1109/ICPCES.2010.5698675.
- [11] Das, L., Das, J. K., & Nanda, S. (2020). Detection of exon location in eukaryotic DNA using a fuzzy adaptive Gabor wavelet transform, *Genomics*, 112(6), 4406–4416, doi: 10.1016/j.ygeno.2020.07.020.

- [12] Daş, B. (2018). *DNA dizilimlerinden hastalık tanılanması için işaret işleme temelli yeni yaklaşımların geliştirilmesi* [Doktora Tezi]. Fırat Üniversitesi.
- [13] Das, L., Nanda, S., & Das, J. K., An integrated approach for identification of exon locations using recursive Gauss Newton tuned adaptive Kaiser window, *Genomics*, 111, 284–296, 2019, doi: 10.1016/j.ygeno.2018.10.008.
- [14] Gupta, R., Mittal, A., Singh, K., Bajpai, P., & Prakash, S., (2008). A Time Series Approach for Identification of Exons and Introns, 91–93, 2008, doi: 10.1109/icit.2007.54.
- [15] Arslan, S. (2014). *KML ve ALL tanılı hastalarda BCR/ABL füzyon geni mutasyonlarının taranması* [Yüksek Lisans Tezi]. Eskişehir Osmangazi Üniversitesi.
- [16] Paiva, A. S., Paiva, H. D. D. O., Cavalcanti, G. B., Leão, G. D., Leão, M. D., Vasconcelos, R. C., Fernandes, A. L., Bahia, F., Soares V. L., Gil E. A., Januario, G., Santos, K. C., Lemos, T. A., & Jardim, A. S., (2018). Detection of the BCR-ABL Gene By the Real-Time PCR Method in Patients with Chronic Myeloid Leukemia in Rio Grande Do Norte, Brazil, *blood*, 132.
- [17] Kitamura, H., Tabe, Y., Ai, T., Tsuchiya, K., Yuri, M., Misawa, S., Horii, T., Kawaguchi, A., Ohsaka, A., & Kimura, S., (2019). A new highly sensitive real-time quantitative-PCR method for detection of BCR-ABL1 to monitor minimal residual disease in chronic myeloid leukemia after discontinuation of imatinib, *PLoS One*, 14(3), 1–13, doi: 10.1371/journal.pone.0207170.
- [18] Smitalova, D., Dvorakova, D., Racil, Z., & Romzova, M. (2021). Digital PCR can provide improved BCR-ABL1 detection in chronic myeloid leukemia patients in deep molecular response and sensitivity of standard quantitative methods using EAC assays, *Practical Laboratory Medicine*, 25, doi: 10.1016/j.plabm.2021.e00210.
- [19] Uzoma, I., & NNA, E., (2017). Detection and quantitation of BCR-ABL1 fusion gene in saliva of chronic myeloid leukaemic patients in NIGERIA', *2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 2325.
- [20] Yang, R., Paparini, A., Monis, P., & Ryan, U. (2014). Comparison of next-generation droplet digital PCR (ddPCR) with quantitative PCR (qPCR) for enumeration of *Cryptosporidium* oocysts in faecal samples, *International Journal for Parasitology*, 44(14), 1105–1113, doi: 10.1016/j.ijpara.2014.08.004.
- [21] Maier, J., Lange, T., Cross, M., Wildenberger, K., Niederwieser, D., & Franke, G. N., (2019). Optimized Digital Droplet PCR for BCR-ABL, *Journal of Molecular Diagnostics*, 21(1), pp. 27–37, doi: 10.1016/j.jmoldx.2018.08.012.
- [22] Jennings, L. J., George, D., Czech, J., Yu, M., & Joseph, L., (2014). Detection and quantification of BCR-ABL1 fusion transcripts by droplet digital PCR, *Journal of Molecular Diagnostics*, 16(2), 174–179, doi: 10.1016/j.jmoldx.2013.10.007.
- [23] Jauhari S. & Rizvi, S. A. M., (2014). Mining gene expression data focusing cancer therapeutics: A digest, *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 11(3), 533–547.

- [24] Begum, S., Sarkar, R., Chakraborty, D., Sen, S., & Maulik, U., (2021). Application of active learning in DNA microarray data for cancerous gene identification, *Expert Systems with Applications*, 177, doi: 10.1016/j.eswa.2021.114914.
- [25] Wang, X., & Simon, R., (2011). Microarray-based cancer prediction using single genes', *BMC Bioinformatics*, 12, 2011, doi: 10.1186/1471-2105-12-391.
- [26] Khorshed, T., Moustafa, M. N., & Rafea, A., (2020). Learning & Visualizing Genomic Signatures of Cancer Tumors using Deep Neural Networks, *Proceedings of the International Joint Conference on Neural Networks*, doi: 10.1109/IJCNN48605.2020.9207368.
- [27] Xu, R., Anagnostopoulos, G. C., & Wunsch, D. C., (2007). Multiclass cancer classification using semisupervised ellipsoid ARTMAP and particle swarm optimization with gene expression data, *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 4(1), 65–77, doi: 10.1109/TCBB.2007.1009.
- [28] Ocampo-Vega, R., Sanchez-Ante, G., De Luna, M.A., Vega, R., Falcón-Morales L.E., & Sossa H., (2016). Improving pattern classification of DNA microarray data by using PCA and Logistic Regression, *Intelligent Data Analysis*, 20, S53–S67, doi: 10.3233/IDA-160845.
- [29] Li, J., Liang, K., & Song, X., (2022). Logistic regression with adaptive sparse group lasso penalty and its application in acute leukemia diagnosis, *Computers in Biology and Medicine*, 141, doi: 10.1016/j.combiomed.2021.105154.
- [30] Yakut T., & Gülten, T., (2005). Çocuklukçağı Lösemilerindeki Genetik Değişiklikler ve Klinik Önemi, *Uludağ Üniversitesi Tıp Fakültesi Dergisi*, 31(1), 57–62.
- [31] Slany, R. K., (2009). The molecular biology of mixed lineage leukemia, *Haematologica*, 94(7), 984–993, doi: 10.3324/haematol.2008.002436.
- [32] Winters A. C., & Bernt, K. M. MLL-rearranged leukemias- An update on science and clinical approaches, *Frontiers in Pediatrics*, 5(4), 2017, doi: 10.3389/fped.2017.00004.
- [33] Slany, R. K., (2016). The molecular mechanics of mixed lineage leukemia, *Oncogene*, 35, 5215–5223, doi: 10.1038/onc.2016.30.
- [34] Alshamlan, H. M., Badr, G. H., & Alohalı, Y. A., (2015). Genetic Bee Colony (GBC) algorithm: A new gene selection method for microarray cancer classification, *Computational Biology and Chemistry*, 56, 49–60.
- [35] Panda, M., (2020). Elephant search optimization combined with deep neural network for microarray data analysis, *Journal of King Saud University - Computer and Information Sciences*, 32(8), 940–948, doi: 10.1016/j.jksuci.2017.12.002.
- [36] Al-jaboriy, S. S., Sjarif, N. N. A., Chuprat, S., & Abdullāh, W. M., (2019). Acute lymphoblastic leukemia segmentation using local pixel information, *Pattern Recognition Letters*, 125, 85–90, doi: 10.1016/j.patrec.2019.03.024.

- [37] Piuri, V., & Scotti, F., (2004). Morphological classification of blood leucocytes by microscope images, *2004 IEEE International Conference on Computational Intelligence for Measurements Systems and Applications, CIMSAs*, 103–108, doi: 10.1109/cimsa.2004.1397242.
- [38] Scotti, F., (2005). Automatic morphological analysis for acute leukemia identification in peripheral blood microscope images, *CIMSAs 2005 – IEEE International Conference on Computational Intelligence for Measurement Systems and Applications*, 96–101, 2005, doi: 10.1109/CIMSAs.2005.1522835.
- [39] Yöntem, A. & Bayram I., (2018). Çocukluk Çağında Akut Lenfoblastik Lösemi, *Arşiv Kaynak Tarama Dergisi*, 27(4), 483–499, 2018, doi: 10.17827/aktd.397762.
- [40] Lösemi nedir, türleri nelerdir? Yaşa göre genel sağkalım oranları nasıldır? (2022). <https://www.drozdogan.com/>.
- [41] Tecimer, T., (2001). Prekürsör B ve T Lenfoblastik Lösemi / Lenfoblastik Lenfoma Patolojisi, *Türk Hematoloji Derneği Klinisyen-Patolog Ortak Lenfoma Kursu*, 24–27.
- [42] Shiraz, P., Jehangir, W., & Agrawal, V. (2021). T-cell acute lymphoblastic leukemia—Current concepts in molecular biology and management, *Biomedicines*, 9, 1–19, 2021, doi: 10.3390/biomedicines9111621.
- [43] Raetz, E. A., & Teachey, D. T., (2016). T-cell acute lymphoblastic leukemia, *Blood | American Society of Hematology*, 580–588.
- [44] Hoelzer, D., & Gökbuğet, N., (2009). T-cell lymphoblastic lymphoma and T-cell acute lymphoblastic leukemia: a separate entity?, *Clinical Lymphoma, Myeloma & Leukemia Supplement*, 9, S214–S221, doi: 10.3816/clm.2009.s.015.
- [45] Girdhar, A., Kapur, H., & Kumar, V., (2022). Classification of White blood cell using Convolution Neural Network, *Biomedical Signal Processing and Control*, 71, doi: 10.1016/j.bspc.2021.103156.
- [46] Baydilli, Y. Y., & Atila, Ü., (2020). Classification of white blood cells using capsule networks, *Computerized Medical Imaging and Graphics*, 80, doi: 10.1016/j.compmedimag.2020.101699.
- [47] Kutlu, H., Avcı, E., & Özyurt, F., (2020). White blood cells detection and classification based on regional convolutional neural networks, *Medical Hypotheses*, 135, doi: 10.1016/j.mehy.2019.109472.
- [48] Das, B. & Turkoglu, I., (2018). A novel numerical mapping method based on entropy for digitizing DNA sequences, *Neural Computing and Applications*, 29(8), 207–215, doi: 10.1007/s00521-017-2871-5.
- [49] Marhon, S. A., & Kremer, S. C., (2011). Protein coding region prediction based on the adaptive representation method, *Canadian Conference on Electrical and Computer Engineering*, 000415–000418, doi: 10.1109/CCECE.2011.6030484.
- [50] Li, J., Zhang, L., Li, H., Ping, Y., Xu, Q., Wang, R., Tan, R., Wang, Z., Liu, B., & Wang, Y., (2019). Integrated entropy-based approach for analyzing exons and introns in DNA sequences, *BMC Bioinformatics*, 20, doi: 10.1186/s12859-019-2772-y.

- [51] Dessouky, A. M., Taha, T. E., Dessouky, M. M., Eltholth, A. A., Hassan, E., & Abd El-Samie, F. E., (2019). Non-parametric spectral estimation techniques for DNA sequence analysis and exon region prediction, *Computers and Electrical Engineering*, 73, 334–348, doi: 10.1016/j.compeleceng.2018.12.001.
- [52] Singh, A. K., & Srivastava, V. K., (2020). The three base periodicity of protein coding sequences and its application in exon prediction, *2020 7th International Conference on Signal Processing and Integrated Networks, SPIN 2020*, 1089–1094, 2020, doi: 10.1109/SPIN48934.2020.9071068.
- [53] Roy, M., & Barman, S. (2010). Spectral analysis of coding and non-coding regions of a DNA sequence by Parametric method, *Proceedings of the 2010 Annual IEEE India Conference: Green Energy, Computing and Communication, INDICON 2010*, doi: 10.1109/INDCON.2010.5712676.
- [54] Liu D. W., Jia, R. P., Wang, C. F., Arunkumar, N., Narasimhan, K., Udayakumar, M., & Elamaram V., (2019). Automated detection of cancerous genomic sequences using genomic signal processing and machine learning, *Future Generation Computer Systems*, 98, 233–237, doi: 10.1016/j.future.2018.12.041.
- [55] Das J., & Barman, S., (2014). Bayesian Fusion in Cancer Gene Prediction, *International Journal of Computer Application*, 5–10.
- [56] Chakraborty S., & Gupta, V. (2016). DWT based cancer identification using EIIP, *Proceedings - 2016 2nd International Conference on Computational Intelligence and Communication Technology, CICT 2016*, 718–723, doi: 10.1109/CICT.2016.148.
- [57] Wang, B., Mohl, J., & Leung, M. Y., (2020). Computational Prediction of Functional Effects for Cancer Related Genetic Sequence Variants, *Proceedings - 2020 IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2020*, 2999–3001, doi: 10.1109/BIBM49941.2020.9313326.
- [58] Gayathri, T. T. (2017). Analysis of Genomic sequences for prediction of Cancerous cells using Wavelet technique, *International Research Journal of Engineering and Technology(IRJET)*, 4(4), 1071–1077.
- [59] Ghosh, A., & Barman, S., (2013). Prediction of Prostate Cancer Cells based on Principal Component Analysis Technique, *Procedia Technology*, 10, 37–44, doi: 10.1016/j.protcy.2013.12.334.
- [60] Muflikhah, L., Widodo, N., Mahmudy, W. F., & Solimun, (2020). Prediction of Liver Cancer Based on DNA Sequence Using Ensemble Method, *2020 3rd International Seminar on Research of Information Technology and Intelligent Systems, ISRITI 2020*, 37–41, doi: 10.1109/ISRITI51436.2020.9315341.
- [61] Das, J., & Barman, S., (2017). DSP based entropy estimation for identification and classification of Homo sapiens cancer genes, *Microsystem Technologies*, 23, 4145–4154, doi: 10.1007/s00542-016-3056-3.
- [62] Chen, A. H., Tsau, Y. W., & Lin, C. H., (2010). Novel methods to identify biologically relevant genes for leukemia and prostate cancer from gene expression profiles, *BMC Genomics*, 11(1), doi: 10.1186/1471-2164-11-274.

- [63] Chakraborty, D., & Maulik, U., (2014). Identifying Cancer Biomarkers from Microarray Data Using Feature Selection and Semisupervised Learning, *IEEE Journal of Translational Engineering in Health and Medicine*, 2, 1–11, doi: 10.1109/JTEHM.2014.2375820.
- [64] Mukhopadhyay, A., Maulik, U., & Bandyopadhyay, S., (2011). Gene expression data analysis using multiobjective clustering improved with SVM based ensemble, *In Silico Biology*, 11, 19–27, doi: 10.3233/ISB-2012-0441.
- [65] Chen, Y., & Zhao, Y., (2008). A novel ensemble of classifiers for microarray data classification, *Applied Soft Computing Journal*, 8, 1664–1669, doi: 10.1016/j.asoc.2008.01.006.
- [66] Dagliyan, O., Uney-Yuksektepe, F., Kavakli, I. H., & Turkay, M., (2011). Optimization based tumor classification from microarray gene expression data, *PLoS One*, 6(2), doi: 10.1371/journal.pone.0014579.
- [67] Golub, T. R., Slonim, D. K., Tamayo, P., Huard, C., Gassenbeck, M., Mesirov, J. P., Coller, H., Loh, M. L., Downing, J. R., Caligiuri, M. A., Bloomfield, C. D., & Lander, E.S., (1999). Molecular classification of cancer: Class discovery and class prediction by gene expression monitoring, *Science*, 286, 531–537, doi: 10.1126/science.286.5439.531.
- [68] Kar, S., Das, Sharma K., & Maitra, M., (2015). Gene selection from microarray gene expression data for classification of cancer subgroups employing PSO and adaptive K-nearest neighborhood technique, *Expert Systems with Applications*, 42(1), 612–627, doi: 10.1016/j.eswa.2014.08.014.
- [69] Tong, D. L., & Schierz, A. C., (2011). Hybrid genetic algorithm-neural network: Feature extraction for unprocessed microarray data, *Artificial Intelligence in Medicine*, 53, 47–56, 2011, doi: 10.1016/j.artmed.2011.06.008.
- [70] Babu P. S. A., Annavarapu, C. S. R., & Dara, S., (2021). Clustering-based hybrid feature selection approach for high dimensional microarray data', *Chemometrics and Intelligent Laboratory Systems*, 213, 2021, doi: 10.1016/j.chemolab.2021.104305.
- [71] Zhang, L., Qian, L., Ding, C., Zhou, W., & Li, F., (2015). Similarity-balanced discriminant neighbor embedding and its application to cancer classification based on gene expression data, *Computers in Biology and Medicine*, 64, 236–245, doi: 10.1016/j.combiomed.2015.07.008.
- [72] Maulik, U., Mukhopadhyay, A., & Chakraborty, D., (2013). Gene-expression-based cancer subtypes prediction through feature selection and transductive SVM, *IEEE Transactions on Biomedical Engineering*, 60(4), 1111–1117, doi: 10.1109/TBME.2012.2225622.
- [73] De Campos, L. M., Cano, A., Castellano, J. G., & Moral, S., (2011). Bayesian networks classifiers for gene-expression data, *International Conference on Intelligent Systems Design and Applications, ISDA*, 1200–1206, , doi: 10.1109/ISDA.2011.6121822.
- [74] Ghosh, M., Begum, S., Sarkar, R., Chakraborty, D., & Maulik, U., (2019). Recursive Memetic Algorithm for gene selection in microarray data, *Expert Systems with Applications*, 116, 172–185, doi: 10.1016/j.eswa.2018.06.057.

- [75] Hegde, R. B., Prasad, K., Hebbar, H., & Singh, B. M. K., (2019). Comparison of traditional image processing and deep learning approaches for classification of white blood cells in peripheral blood smear images, *Biocybernetics and Biomedical Engineering*, 39(2), 382–392, doi: 10.1016/j.bbe.2019.01.005.
- [76] Acevedo, A., Alf  rez, S., Merino, A., Puigv  , L., & Rodellar, J., (2019). Recognition of peripheral blood cell images using convolutional neural networks', *Computer Methods and Programs in Biomedicine*, 180, doi: 10.1016/j.cmpb.2019.105020.
- [77] Hegde, R. B., Prasad, K., Hebbar, H., & Sandhya, I., (2018). Peripheral blood smear analysis using image processing approach for diagnostic purposes: A review, *Biocybernetics and Biomedical Engineering*, 38(3), 467–480, doi: 10.1016/j.bbe.2018.03.002.
- [78] Reena M. R., & Ameer, P. M., (2020). Localization and recognition of leukocytes in peripheral blood: A deep learning approach, *Computers in Biology and Medicine*, 126, doi: 10.1016/j.combiomed.2020.104034.
- [79] Banik, P. P., Saha, R., & Kim, K. D., (2020). An Automatic Nucleus Segmentation and CNN Model based Classification Method of White Blood Cell', *Expert Systems with Applications*, 149, doi: 10.1016/j.eswa.2020.113211.
- [80] Alomari, Y. M., Sheikh Abdullah, S. N. H., Zaharatul Azma, R., & Omar, K., (2014). Automatic detection and quantification of WBCs and RBCs using iterative structured circle detection algorithm, *Computational and Mathematical Methods in Medicine*, doi: 10.1155/2014/979302.
- [81] Bhagavathi S. L., & Thomas Niba, S., (2016). An automatic system for detecting and counting rbc and wbc using fuzzy logic, *ARPJ Journal of Engineering and Applied Sciences*, 11(11), 6891–6894.
- [82] Marzuki, N. I. C., Mahmood, N. H., & Razak, M. A. A., (2015). Segmentation of white blood cell nucleus using active contour, *Jurnal Teknologi*, 74(6), 115–118, 2015, doi: 10.11113/jt.v74.4675.
- [83] Wang, Q., Wang, J., Zhou, M., Li, Q., & Wang, Y., (2017). Spectral-spatial feature-based neural network method for acute lymphoblastic leukemia cell identification via microscopic hyperspectral imaging technology, *Biomedical Optics Express*, 8(6), doi: 10.1364/boe.8.003017.
- [84] Sadeghian, F., Seman, Z., Ramli, A. R., Abdul Kahar, B. H., & Saripan, M. I., (2009). A framework for white blood cell segmentation in microscopic blood images using digital image processing, *Biological Procedures Online*, 11(1), 196–206, doi: 10.1007/s12575-009-9011-2.
- [85] Mishra P., & Bhoi, N., (2021). Cancer gene recognition from microarray data with manta ray based enhanced ANFIS technique, *Biocybernetics and Biomedical Engineering*, 41(3), 916–932, doi: 10.1016/j.bbe.2021.06.004.
- [86] Sayed, S., Nassef, M., Badr, A., & Farag, I., (2019). A Nested Genetic Algorithm for feature selection in high-dimensional cancer Microarray datasets, *Expert Systems with Applications*, 121, 233–243, doi: 10.1016/j.eswa.2018.12.022.

- [87] Potharaju S. P., & Sreedevi, M., (2019). Distributed feature selection (DFS) strategy for microarray gene expression data to improve the classification performance, *Clinical Epidemiology and Global Health*, 7(2), 171–176, doi: 10.1016/j.cegh.2018.04.001.
- [88] Vafae Sharbaf, F., Mosafer, S., & Moattar, M. H., (2016). A hybrid gene selection approach for microarray data classification using cellular learning automata and ant colony optimization, *Genomics*, 107(6), 231–238, doi: 10.1016/j.ygeno.2016.05.001.
- [89] Kundu, R., Chattopadhyay, S., Cuevas, E., & Sarkar, R., (2022). AltWOA: Altruistic Whale Optimization Algorithm for feature selection on microarray datasets, *Computers in Biology and Medicine*, 144, doi: 10.1016/j.combiomed.2022.105349.
- [90] Su, M.-C., Cheng, C.-Y., & Wang, P.-C., (2014). A neural-network-based approach to white blood cell classification, *The Scientific World Journal*, doi: 10.1155/2014/796371.
- [91] Bozkurt, F., (2021). Classification of Blood Cells from Blood Cell Images Using Dense Convolutional Network, *Journal of Scientific Technology and Engineering Research*, 2(2), 81–88, doi: 10.53525/jster.1014186.
- [92] Elen, A. & Turan, M. K., (2019). Classifying White Blood Cells Using Machine Learning Algorithms, *Uluslararası Mühendislik. Arastırma ve Geliştirme Dergisi*, 11(1), 141–152, doi: 10.29137/umagd.498372.
- [93] Ha, Y., Du, Z., & Tian, J., (2022). Fine-grained interactive attention learning for semi-supervised white blood cell classification, *Biomedical Signal Processing and Control*, 75, doi: 10.1016/j.bspc.2022.103611.
- [94] Patil, A. M., Patil, M. D., & Birajdar, G. K., (2021). White Blood Cells Image Classification Using Deep Learning with Canonical Correlation Analysis, *IRBM*, 42, 378–389, doi: 10.1016/j.irbm.2020.08.005.
- [95] Shahin, A. I., Guo, Y., Amin, K. M., & Sharawi, A. A., (2019). White blood cells identification system based on convolutional deep neural learning networks, *Computer Methods and Programs in Biomedicine*, 168, 69–80, doi: 10.1016/j.cmpb.2017.11.015.
- [96] National Library of Medicine. National Center for Biotechnology Information. (2022). BCR-ABL1 gene. <https://www.ncbi.nlm.nih.gov/>.
- [97] Cancer gene expression data sets and their visualizations, (2022). Leukemia malignancies. <https://file.biomedcentral.com/suppl/10.1186/s13059-022-02800-0/1>.
- [98] Acute Lymphoblastic Leukemia Image Database for Image Processing/ALL-IDB (2022). Microscopic images of blood samples. <https://homes.di.unimi.it/scotti/all/>.
- [99] RaabinData/Raabin Health Database (2022). Microscopic images of blood samples. <https://raabindata.com/>.
- [100] Duran, K. (2013). *Rüzgar karakteristiklerinin dalgacık dönüşümü ile ortaya konması*, [Yüksek Lisans Tezi]. İstanbul Teknik Üniversitesi.

- [101] Abo-Zahhad, M., Ahmed, S. M., & Abd-Elrahman, S. A., (2012). Genomic Analysis and Classification of Exon and Intron Sequences Using DNA Numerical Mapping Techniques, *International Journal of Information Technology and Computer Science*, 4(8), 22–36, doi: 10.5815/ijitcs.2012.08.03.
- [102] Yu, N., Li, Z., & Yu, Z., (2018). Survey on encoding schemes for genomic data representation and feature learning-from signal processing to machine learning, *Big Data Mining and Analytics*, 1(3), 191–210, doi: 10.26599/BDMA.2018.9020018.
- [103] Aygün, O. (2006). *Konuşmacı tanıma sistemlerinde dalgacık dönüşümü*, [Yüksek Lisans Tezi]. Selçuk Üniversitesi.
- [104] Avcı K., & Coskun, O., (2018). Spectral performance analysis of cosh window based new two parameter hybrid windows, *26th IEEE Signal Processing and Communications Applications Conference, SIU 2018*, 1–4.
- [105] Aktürk, S. M. (2018). *Grabcut etkileşimli bölütleme yöntemi üzerinde iyileştirme çalışmaları*, [Yüksek Lisans Tezi], Karadeniz Teknik Üniversitesi.
- [106] Atila, Ü., Uçar, M., Akyol, K., & Uçar, E., (2021). Plant leaf disease classification using EfficientNet deep learning model, *Ecological Informatics*, 61, doi: 10.1016/j.ecoinf.2020.101182.
- [107] Elmas, B., (2021). Identifying species of trees through bark images by convolutional neural networks with transfer learning method, *Journal of the Faculty of Engineering and Architecture of Gazi University*, 36(3), 1253–1269, 2021, doi: 10.17341/gazimmfd.689038.
- [108] Sreng, S., Maneerat, N., Hamamoto, K., & Win, K. Y., (2020). Deep learning for optic disc segmentation and glaucoma diagnosis on retinal images, *Applied Sciences*, 10, doi: 10.3390/app10144916.
- [109] Xia, Y., Johnson, B. K., Jiang, Y., Fischer, N., & Xia, H., (2018). A new method based on artificial neural network, Wavelet Transform and Short Time Fourier Transform for Subsynchronous Resonance detection, *International Journal of Electrical Power and Energy Systems*, 103, 377–383.
- [110] Fidan, H. (2006). *Dalgacık dönüşümü tekniği ile motor arıza tespiti*, [Yüksek Lisans Tezi], Süleyma Demirel Üniversitesi.
- [111] Zhao, Y., Zhou, Y., Liu, Y., Hao, Y., Li, M., Pu, X., Li, C., & Wen, Z., (2020). Uncovering the prognostic gene signatures for the improvement of risk stratification in cancers by using deep learning algorithm coupled with wavelet transform, *BMC Bioinformatics*, 21, 1–24, doi: 10.1186/s12859-020-03544-z.
- [112] Öner, İ. V., Yeşilyurt, M. K., & Yılmaz, E. Ç., (2017). Wavelet Analiz Tekniği Ve Uygulama Alanları, *Ordu Üniv. Bil. Tek. Derg.*, 7(1), 42–56.
- [113] Saini S., & Dewan, L., (2016). Application of discrete wavelet transform for analysis of genomic sequences of Mycobacterium tuberculosis', *Springerplus*, 5, 1–15, doi: 10.1186/s40064-016-1668-9.
- [114] Wu, Z., Pan, S., Chen, F., Long, G., Zhang, C., & Yu, P. S., (2021). A Comprehensive Survey on Graph Neural Networks, *IEEE Transactions on Neural Networks and Learning Systems*, 32(1), 4–24.

- [115] Zhang, M., Cui, Z., Neumann, M., & Chen, Y., (2018). An end-to-end deep learning architecture for graph classification, *32nd AAAI Conference on Artificial Intelligence, AAAI 2018*.
- [116] Wang, S. H., Govindaraj, V. V., Górriz, J. M., Zhang, X., & Zhang, Y. D., (2021). Covid-19 classification by FGCNet with deep feature fusion from graph convolutional network and convolutional neural network, *Information Fusion*, 67, 208–229, doi: 10.1016/j.inffus.2020.10.004.
- [117] Matlab. (2022). İstatistik komutları. <https://www.mathworks.com/help/>.
- [118] Wikipedia (2022). Kovaryans. <https://tr.wikipedia.org/wiki/Kovaryans>.
- [119] Wikipedia (2022). Group method of data handling. https://en.wikipedia.org/wiki/Group_method_of_data_handling.
- [120] Kaya, Y., Makaracı, M., Bayraklılar, M. S., & Kuncan, M., (2021). Using group method of data handling (GMDH) neural network to predict the maximum stress on elastomeric layers in spherical elastomeric bearings, *Journal of the Faculty of Engineering and Architecture of Gazi University*, 36(3), 1331–1345, doi: 10.17341/gazimmfd.722514.
- [121] Mahdevari S., & Khodabakhshi, M. B., (2021). A hybrid PSO-ANFIS model for predicting unstable zones in underground roadways, *Tunnelling and Underground Space Technology*, 117, doi: 10.1016/j.tust.2021.104167.
- [122] Karaboga D., & Kaya, E., (2020). Estimation of number of foreign visitors with ANFIS by using ABC algorithm, *Soft Computing*, 24, 7579–7591, doi: 10.1007/s00500-019-04386-5.
- [123] Houssein, E. H., Gad, A. G., Hussain, K., & Suganthan, P. N., (2021). Major Advances in Particle Swarm Optimization: Theory, Analysis, and Application, *Swarm and Evolutionary Computation*, 63, doi: 10.1016/j.swevo.2021.100868.
- [124] Karaboğa, D. (2018) *Yapay Zeka Optimizasyon Algoritmaları (6. Basım)*. Nobel.
- [125] Doğan, C., (2019). *Balina Optimizasyon Algoritması ve Gri Kurt Optimizasyonu Algoritmaları Kullanılarak Yeni Hibrit Optimizasyon Algoritmalarının Geliştirilmesi*, [Yüksek Lisans Tezi], Erciyes Üniversitesi.
- [126] Canayaz M., & Demir, M., (2017). Balina Optimizasyon Algoritması ve Yapay Sinir Ağı ile Öznitelik Seçimi, *2017 International Artificial Intelligence and Data Processing Symposium (IDAP)*. doi: 10.1109/IDAP.2017.8090247
- [127] Mirjalili S., & Lewis, A., (2016). The Whale Optimization Algorithm, *Advances in Engineering Software*, 95, 51–67, doi: 10.1016/j.advengsoft.2016.01.008.
- [128] Christensen J., & Bastien, C., (2016). Chapter | seven - Heuristic and Meta-Heuristic Optimization Algorithms, *Nonlinear Optimization of Vehicle Safety Structures*, pp. 277–314.
- [129] Alobaedy, M. M., Khalaf, A. A., & Muraina, I. D., (2017). Analysis of the number of ants in ant colony system algorithm, *2017 5th International Conference on Information and Communication Technology, ICoIC7 2017*.

- [130] El Mrabet, M. A., El Makkaoui, K., & Faize, A., (2021). Supervised Machine Learning: A Survey, *Proceedings - 4th International Conference on Advanced Communication Technologies and Networking, CommNet 2021*, doi: 10.1109/CommNet52204.2021.9641998.
- [131] Akalın, F. (2020). *İnce bağırsak görüntüleri üzerinde sezgisel algoritma teknikleri ile polip teşhisi*, [Yüksek Lisans Tezi]. Sakarya Üniversitesi.
- [132] Adak, M. F., & Yurtay, N., (2013). Gini Algoritmasını Kullanarak Karar Ağacı Oluşturmayı Sağlayan Bir Yazılımın Geliştirilmesi, *Bilişim Teknolojileri Dergisi*, 6(3), 1–6.
- [133] Li, Y., & Cao, H., (2018). Prediction for Tourism Flow based on LSTM Neural Network, *Procedia Comput. Science*, 129, 277–283, doi: 10.1016/j.procs.2018.03.076.
- [134] Redmon, J., Divvala, S., Girshick, R., & Farhadi, A., (2016). You only look once: Unified, real-time object detection’, *Arxiv*, 2016, doi: 10.1109/CVPR.2016.91.
- [135] Bochkovskiy, A., Wang, C.-Y., & Liao, H.-Y. M., (2020). YOLOv4: Optimal Speed and Accuracy of Object Detection, *arXiv*, [Online]. Available: <http://arxiv.org/abs/2004.10934>.
- [136] Pacal, I., Karaman, A., Karaboğa, D., Akay, B., Basturk, A., Nalbantoğlu, U., & Coskun, S., (2021). An efficient real-time colonic polyp detection with YOLO algorithms trained by using negative samples and large datasets, *Computers in Biology and Medicine*, doi: 10.1016/j.combiomed.2021.105031.
- [137] Chethan, Kumar, B., Punitha, R. & Mohana, (2020). ‘YOLOv3 and YOLOv4: Multiple object detection for surveillance applications’, *Proceedings of the 3rd International Conference on Smart Systems and Inventive Technology, ICSSIT 2020*, 1316–1321, doi: 10.1109/ICSSIT48917.2020.9214094.
- [138] Li, Y., Wang, H., Dang, L. M., Nguyen, T. N., Han, D. Lee, A., Jang, I., & Moon, H., (2020). A deep learning-based hybrid framework for object detection and recognition in autonomous driving, *IEEE Access*, 8, 194228–194239, 2020, doi: 10.1109/ACCESS.2020.3033289.
- [139] Li, X., Yin, P., Duan, C., & Zhang, S., (2021). Research on the application of YOLOV4 target detection network in traffic scenarios by machine vision technology, *Journal of Physics: Conference Series*, doi: 10.1088/1742-6596/2033/1/012151.
- [140] Zhao, Z. Q., Zheng, P., Xu, S. T. & Wu, X., (2019). Object Detection with Deep Learning: A Review, *IEEE Transactions on Neural Networks and Learning Systems*, doi: 10.1109/TNNLS.2018.2876865.
- [141] Song, Q., Li, S., Bai, Q., Yang, J., Zhang X., Li, Z., & Duan, Z., (2021). Object detection method for grasping robot based on improved yolov5, *Micromachines*, 12, doi: 10.3390/mi12111273.
- [142] Zhu, L., Geng, X., Li, Z., & Liu, C., Improving yolov5 with attention mechanism for detecting boulders from planetary images, *Remote Sensing*, 13, 1–19, 2021, doi: 10.3390/rs13183776.

- [143] Javier, Yague, F., Diez-Pastor, J. F., Latorre-Carmona, P., & Osorio, C. I. G., (2022). Defect detection and segmentation in X-Ray images of magnesium alloy castings using the Detectron2 framework', *arXiv*, 1–13.
- [144] Yang, D., Wang, X., Zhang, H., Yin, Z. Y., Su, D., & Xu, J., (2021) A Mask R-CNN based particle identification for quantitative shape evaluation of granular materials, *Powder Technology*, 392, 296–305, doi: 10.1016/j.powtec.2021.07.005.
- [145] Koupaei, J. A., Hosseini, S. M. M., & Ghaini, F. M. M., (2016). A new optimization algorithm based on chaotic maps and golden section search method, *Engineering Applications of Artificial Intelligence*, 50, 201–214, doi: 10.1016/j.engappai.2016.01.034.
- [146] Sak, H., Senior, A. & Beaufays, F., (2014). Long Short-Term Memory Based Recurrent Neural Network Architectures for Large Vocabulary Speech Recognition. [Online]. Available: <http://arxiv.org/abs/1402.1128>.
- [147] Zamfirache, I. A., Precup, R. E., Roman, R. C., & Petriu, E. M., (2022). Policy Iteration Reinforcement Learning-based control using a Grey Wolf Optimizer algorithm, *Information Sciences*, 585, 162–175, doi: 10.1016/j.ins.2021.11.051.
- [148] M., R. K., & Vaegae, N. K., (2020). Walsh code based numerical mapping method for the identification of protein coding regions in eukaryotes, *Biomedical Signal Processing and Control*, 58, doi: 10.1016/j.bspc.2020.101859.
- [149] Sahu, S. S., & Panda, G., (2011). Identification of protein-coding regions in DNA sequences using a time-frequency filtering approach, *Genomics, Proteomics Bioinformatics*, 9(1–2), 45–55, doi: 10.1016/S1672-0229(11)60007-7.
- [150] Daş, B., Toraman, S., & Türkoğlu, I., (2020). A novel genome analysis method with the entropy-based numerical technique using pretrained convolutional neural networks, *Turkish Journal of Electrical Engineering and Computer Sciences*, 28(4), 1932–1948, doi: 10.3906/ELK-1909-119.
- [151] Piecyk, R. S., Schlegel, L., & Johannes, F., (2022). Predicting 3D chromatin interactions from DNA sequence using Deep Learning, *Computational and Structural Biotechnology Journal*, 20, 3439–3448, doi: 10.1016/j.csbj.2022.06.047.
- [152] Das B., & Toraman, S., (2022). Deep transfer learning for automated liver cancer gene recognition using spectrogram images of digitized DNA sequences, *Biomedical Signal Processing and Control*, 72, doi: 10.1016/j.bspc.2021.103317.
- [153] Le, N. Q. K., Ho, Q. T., Nguyen, V. N., & Chang, J. S., (2022). BERT-Promoter: An improved sequence-based predictor of DNA promoter using BERT pre-trained model and SHAP feature selection, *Computational Biology and Chemistry*, 99, doi: 10.1016/j.compbiolchem.2022.107732.
- [154] Nandhini, K. & Tamilpavai, G., (2022). Hybrid CNN-LSTM and modified wild horse herd Model-based prediction of genome sequences for genetic disorders, *Biomedical Signal Processing and Control*, 78.

- [155] Barshai, M., Aubert, A., & Orenstein, Y., (2022). G4detector: Convolutional Neural Network to Predict DNA G-Quadruplexes, *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 19(4), 1946–1955, doi: 10.1109/TCBB.2021.3073595.
- [156] Wang, X. & Gotoh, O., (2009). Cancer classification using single genes, *Genome informatics*, 179–188, 2009, doi: 10.1142/9781848165632_0017.
- [157] Ghorai, S., Mukherjee, A. & Dutta, P. K., (2012). Gene Expression Data Classification by VVRKFA, *Procedia Technology*, 4, 330–335, doi: 10.1016/j.protcy.2012.05.050.
- [158] Maulik, U. & Chakraborty, D., (2014). Fuzzy preference based feature selection and semisupervised SVM for cancer classification, *IEEE Trans. Nanobioscience*, 13(2), 152–160, 2014, doi: 10.1109/TNB.2014.2312132.
- [159] Jha K. K., & Dutta, H. S., (2019). Mutual Information based hybrid model and deep learning for Acute Lymphocytic Leukemia detection in single cell blood smear images, *Computer Methods and Programs in Biomedicine*, 179, doi: 10.1016/j.cmpb.2019.104987.
- [160] Al-Tahhan, F. E., Fares, M. E., Sakr, A. A., & Aladle, D. A., (2020). Accurate automatic detection of acute lymphatic leukemia using a refined simple classification, *Microscopy Research and Technique*, doi: 10.1002/jemt.23509.
- [161] Safuan, S. N. M., Tomari, M. R. M., Zakaria, W. N. W., Mohd, M. N. H., & Suriani, N. S., (2020). Investigation of white blood cell biomaker model for acute lymphoblastic leukemia detection based on convolutional neural network, *Bulletin of Electrical Engineering and Informatics*, 9(2), 611–618, doi: 10.11591/eei.v9i2.1857.
- [162] Sahlol, A. T., Kollmannsberger, P., & Ewees, A. A., (2020). Efficient Classification of White Blood Cell Leukemia with Improved Swarm Optimization of Deep Features, *Scientific Reports*, doi: 10.1038/s41598-020-59215-9.
- [163] Khandekar, R., Shastry, P., Jaishankar, S., Faust, O., & Sampathila, N., (2021). Automated blast cell detection for Acute Lymphoblastic Leukemia diagnosis, *Biomedical Signal Processing and Control*, 68, doi: 10.1016/j.bspc.2021.102690.
- [164] Peng, S., Xu, Q., Ling, X. B., Peng, X., Du, W., & Chen, L., (2003). Molecular classification of cancer types from microarray data using the combination of genetic algorithms and support vector machines, *FEBS Letters*, 358–362, doi: 10.1016/S0014-5793(03)01275-4.
- [165] Anita, & Yadav, A., (2021). An Intelligent Model for the Detection of White Blood Cells using Artificial Intelligence, *Computer Methods and Programs in Biomedicine*, 199, doi: 10.1016/j.cmpb.2020.105893.
- [166] Makem, M. & Tiedeu, A., (2020). An efficient algorithm for detection of white blood cell nuclei using adaptive three stage PCA-based fusion, *Informatics in Medicine Unlocked*, 20, doi: 10.1016/j.imu.2020.100416.

- [167] Ridoy, M. A. R., & Islam, M. R., (2020). A Lightweight Convolutional Neural Network for White Blood Cells Classification, *ICCIT 2020 - 23rd International Conference on Computer and Information Technology, Proceedings*, 19–21.
- [168] Kousalya, K., Krishnakumar, B., Mohana, R. S., & Karthikeyan, N., (2021). Comparative analysis of White Blood Cells Classification using Deep Learning Architectures, *Proceedings - 2nd International Conference on Smart Electronics and Communication, ICOSEC 2021*, 1220–1225, doi: 10.1109/ICOSEC51865.2021.9591771.
- [169] Baby, D., Devaraj, S. J., Hemanth, J., & Anishin Raj, M. M., (2021). Leukocyte classification based on feature selection using extra trees classifier: A transfer learning approach, *Turkish Journal of Electrical Engineering and Computer Sciences*, 29(8), 2742–2757, doi: 10.3906/elk-2104-183.
- [170] Sharma, A., Thomas, S. C., Sah, A., Abhyankar, V. V., Singh, V. K., & Prakash, S., (2021). White Blood Cells Subtypes Classification Using Fast Traditional Convolutional Neural Network, *Proceedings of the 2021 International Conference on Emerging Techniques in Computational Intelligence, ICETCI 2021*, 113–117, doi: 10.1109/ICETCI51973.2021.9574045.
- [171] Shah, R., Shastri, J., Bohara, M. H., Panchal, B. Y., & Goel, P., (2022). Detection of Different Types of Blood Cells: A Comparative Analysis, 1–5, doi: 10.1109/icdcece53908.2022.9793132.
- [172] Xu, F., Li, X., Yang, H., Wang, Y., & Xiang, W., (2022). TE-YOLOF: Tiny and efficient YOLOF for blood cell detection, *Biomedical Signal Processing and Control*, 73, doi: 10.1016/j.bspc.2021.103416.

ÖZGEÇMİŞ

Ad-Soyad : Fatma Akalın

ÖĞRENİM DURUMU:

- **Lisans** : 2018, Sakarya Üniversitesi, Bilgisayar ve Bilişim Bilimleri Fakültesi, Bilgisayar Mühendisliği Bölümü
- **Yükseklisans** : 2020, Sakarya Üniversitesi, Bilgisayar ve Bilişim Bilimleri Fakültesi, Bilgisayar Mühendisliği Bölümü

MESLEKİ DENEYİM VE ÖDÜLLER:

- 2020 yılından beri Sakarya Üniversitesi'nde Araştırma Görevlisi olarak çalıştı.
- 2019 yılında 10. İksad Bilim Ödülü'nü kazandı.

TEZDEN TÜRETİLEN ESERLER:

- AKALIN FATMA, YUMUŞAK NEJAT (2022). Detection and classification of white blood cells with an improved deep learning-based approach. Turkish Journal of Electrical Engineering and Computer Sciences, 30(7), 2725-2739., Doi: 10.55730/1300-0632.3965
- AKALIN FATMA, YUMUŞAK NEJAT (2023). Lösemi hastalığının temel türlerinden ALL ve KML malignitelerinin grafik sinir ağları ve bulanık mantık algoritması ile sınıflandırılması. Gazi Üniversitesi Mühendislik-Mimarlık Fakültesi Dergisi, 38(2), 707-720., Doi: 10.17341/gazimmfd.1022624
- AKALIN FATMA, YUMUŞAK NEJAT (2022). DNA genom dizilimi üzerinde dijital sinyal işleme teknikleri kullanılarak elde edilen ekson ve intron bölgelerinin EfficientNetB7 mimarisi ile sınıflandırılması. Gazi Üniversitesi Mühendislik-Mimarlık Fakültesi Dergisi, 37(3), 1355-1371., Doi: 10.17341/gazimmfd.900987

- AKALIN FATMA, YUMUŞAK NEJAT (2023). Özellik seçim algoritmaları ve derin öğrenme tabanlı mimarilerin hibrit kullanımıyla akut lösemilerin sınıflandırılması. Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi, Doi: 10.5505/pajes.2022.62282
- AKALIN FATMA, YUMUŞAK NEJAT (2022). Early Detection of ALL Disease Using YOLOv4 Algorithm on Peripheral Blood Smear Images. Akıllı Sistemlerde Yenilikler ve Uygulamaları Konferansı, Doi: 10.1109/ASYU56188.2022.9925427 (Tam Metin Bildiri/Sözlü Sunum)
- AKALIN FATMA, YUMUŞAK NEJAT (2023). Mikrodizi veri kümesindeki ALL, AML ve MLL lösemi türlerine ilişkin gen anomalilerinin LSTM sinir ağı ile sınıflandırılması. Journal of the Faculty of Engineering and Architecture of Gazi University, 38(3), 1299-1306., Doi: 10.17341/gazimmfd.1064693 (Yayın No: 8128817)

DİĞER ESERLER:

- AKALIN FATMA, ORHAN MEHMET FATİH, BÜYÜKAVCI MUSTAFA (2022). A Decision Support System For Detecting Stage In Hodgkin Lymphoma Patients Using Artificial Neural Network and Optimization Algorithms. Sakarya University Journal of Computer and Information Sciences, 5(3), 448-461., Doi: 10.35377/saucis.05.03.1210786
- AKALIN FATMA, YUMUŞAK NEJAT (2019). İnce Bağırsak Fotoğrafları Üzerinde Sezgisel Algoritma Teknikleri ile Polip Teşhisi. International Ankara Conference on Scientific Researches, 670-679. (Tam Metin Bildiri/Sözlü Sunum)(Yayın No:7935060)
- AKALIN FATMA, SAKEN Manarbek, YUMUŞAK NEJAT (2022). Derin Öğrenme Algoritmaları ile Glukom hastalığının önceden tespiti. IV. INTERNATIONAL ANKARA MULTIDISCIPLINARY STUDIES CONGRESS (Tam Metin Bildiri/Sözlü Sunum)(Yayın No:7780862)