

GRAFİK SİNİR AĞLARI İLE İLAÇ KEŞFİ

VEYSEL GİDER

TEMMUZ 2024

DİYARBAKIR

DİCLE ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

GRAFİK SİNİR AĞLARI İLE İLAÇ KEŞFİ



VEYSEL GİDER

DİCLE ÜNİVERSİTESİ LİSANSÜSTÜ EĞİTİM-ÖĞRETİM VE SINAV
YÖNETMELİĞİNİN BİR PARÇASI OLARAK
ELEKTRİK-ELEKTRONİK MÜHENDİSLİĞİ ANA BİLİM DALINDA
DOKTORA TEZİ
OLARAK HAZIRLANMIŞTIR

TEMMUZ 2024

DİYARBAKIR

GRAFİK SİNİR AĞLARI İLE İLAÇ KEŞFİ

Veysel GİDER tarafından Dicle Üniversitesi Lisansüstü Eğitim-Öğretim ve Sınav Yönetmeliği'nin bir parçası olarak hazırlanan bu çalışma, aşağıda bilgileri yazılı jüri üyeleri tarafından değerlendirilerek **Elektrik-Elektronik Mühendisliği Ana Bilim Dalı**'nda **Doktora Tezi** olarak kabul edilmiştir.

Prof. Dr. Neslihan DALKILIÇ
Müdür, **Fen Bilimleri Enstitüsü**

Doç. Dr. Cafer BUDAK
Danışman, **Elektrik-Elektronik Mühendisliği Bölümü,**
Dicle Üniversitesi

Sınav Jürisi:

Doç. Dr. Serdar EKİNCİ (*)
Bilgisayar Mühendisliği Bölümü, Batman Üniversitesi

Doç. Dr. Cafer BUDAK (**)
Elektrik-Elektronik Mühendisliği Bölümü, Dicle Üniversitesi

Prof. Dr. Neval Berrin ARSERİM
Klinik Öncesi Bilimleri Bölümü, Dicle Üniversitesi

Dr. Öğr. Üyesi Mehmet Emin ASKER
Elektrik ve Enerji Bölümü, Dicle Üniversitesi

Doç. Dr. Davut İZCİ
Bilgisayar Mühendisliği Bölümü, Batman Üniversitesi



Savunma Tarihi: 30 / 07 / 2024

(*) Jüri Başkanı.

(**) Tez Danışmanı.

Eşim ve Aileme ...

Dicle Üniversitesi Fen Bilimleri Enstitüsü Tez Yazım Kurallarına uygun olarak hazırlanan bu tez çalışmasında yer alan tüm bilgilerin akademik kurallara ve etik ilkelere uygun olarak elde edildiğini ve sunulduğunu beyan ederim. Ayrıca, bahse konu bu kural ve ilkelerin gerektirdiği üzere, bu çalışmada özgün olmayan tüm bilimsel içerikleri kurallara uygun biçimde alıntılıyıp kaynak gösterdiğimi beyan ederim. Beyanımınla çelişen herhangi bir delil bulunduğu takdirde tüm sorumluluğu üstleneceğimi kabul ederim.

Ad, Soyad: Veysel GİDER

İmza:

TEŞEKKÜR

Öncelikle danışmanım Doç. Dr. Cafer BUDAK'a çalışmanın daha fikir aşamasından, verilere erişim, algoritma oluşturma, analizlerin yorumlanması ve tez düzenlenmesi konularına kadar gösterdiği destek ve sabırdan dolayı teşekkür ederim. Jüri üyelerime de bu tezin daha nitelikli bir hale gelmesi için yaptıkları eleştiri ve önerileri için teşekkür ederim.

Dicle Üniversitesi Bilimsel Araştırma Projeleri Koordinatörlüğü'ne (DÜBAP), bu tezin bir parçası olan proje çalışması için sağladıkları finansal destek için (Proje No: DÜBAP-MÜHENDİSLİK.23.016) için teşekkürü bir borç bilirim.

Hayatımın bu zorlu doktora yolculuğunda, araştırma tutkumu canlı tutmamda ve her adımda yanımda olmalarından dolayı sevgili eşime, değerli anneme, babama, ablama ve kardeşlerime sonsuz teşekkür ederim.

İÇİNDEKİLER

TEŞEKKÜR.....	v
ŞEKİLLER LİSTESİ	viii
TABLOLAR LİSTESİ.....	xi
SİMGELER VE KISALTMALAR LİSTESİ	xii
ÖZET.....	xvi
ABSTRACT.....	xvii
1. GİRİŞ	1
2. ÖNCEKİ ÇALIŞMALAR.....	11
3. MALZEME VE YÖNTEM	19
3.1 İlaç Keşfi	19
3.1.1 İlaç yeniden kullanımı.....	20
3.2 İlaç keşfi sürecinde yapay zekâ uygulamaları.....	22
3.2.1 Grafik sinir ağları (GSA)	23
4. BULGULAR VE TARTIŞMA	35
4.1 Bölüm 1: Salgın Hastalıklar	35
4.1.1 Ebola virüsünün tedavisi için ilaçların grafik ağı ve benzerlik algoritması yöntemleriyle moleküler analizi ve bir ilaç kombinasyon önerisi	35
4.1.2 COVID-19 ve akciğer kanser ilaçlarının benzerlik analizi ve grafik sinir ağları tabanlı ilaç-protein bağlanma enerjisinin hesaplanması	60
4.2 Bölüm 2: İlaç çözünürlük tahmini	79
4.2.1 İlaç çözünürlük tahmininde GSA ve diğer hesaplamalı yöntemlerin karşılaştırılması	79
4.2.2 Grafik Sinir Ağları ile İlaç Molekülü Çözünürlük Tahmini	93
4.3 Bölüm 3: Moleküler modelleme	103
4.3.1 Grafik sinir ağları tabanlı moleküler modelleme ile ilaç-protein etkileşimlerinin analizi ve ilaç keşfi: Chlorpromazine-6LU7 projesi	103
4.4 Bölüm 4: İlaç yapı analizleri	132

4.4.1	Yapay zekâ ve ilaç keşfi: ebola tedavisi için yedi farklı ilacın benzerlik ve moleküler yapı analizi	132
5.	SONUÇLAR VE ÖNERİLER	144
5.1	Sonuçlar.....	144
5.2	Öneriler	145
	KAYNAKLAR	147
	ÖZGEÇMİŞ	168



ŞEKİLLER LİSTESİ

Şekil 1.1 Yapay zekâ ve alt dallarının gelişimi.....	4
Şekil 3.1 Bir ilacı geliştirme süreci	20
Şekil 3.2 İlaç yeniden kullanımı genel bakış	22
Şekil 3.3 GSA uygulamalar alanlarının gösterimi	24
Şekil 3.4 Temel grafik yapısı	25
Şekil 3.5 Laplasyan matrisi	26
Şekil 3.6 Tek bir düğümün yerel komşuluğundan gelen mesajları nasıl topladığına genel bakış.....	27
Şekil 3.7 Bir GSA'nın genel model yapısı.....	29
Şekil 4.1 Ebola tedavisinde ilaç-ilaç benzerliğini ve moleküler iskeleleri analiz etmek için önerilen yaklaşımın akış şeması.....	42
Şekil 4.2 Moleküler parmak izi yapısı	43
Şekil 4.3 Atom çifti parmak izi yapısı	45
Şekil 4.4 Invacraftor ilacı için iskele ağı, ilacın kompozit iskelelere ayrıştırılabileceği tüm olası yolları gösterimi	46
Şekil 4.5 Benzerliği çıkarılan yedi ilacın grafik ağı algoritması tarafından oluşturulan bu ilaçların moleküler iskelelerinin temsili.....	51
Şekil 4.6 Tüm ilaçların iskelelerinin düğüm grafikleri şeklinde gösterimi.....	52
Şekil 4.7 Grafik ağı algoritması tarafından yedi (7) ilaç için oluşturulan iskelelerin moleküler yapıları da dahil olmak üzere gösterimidir. Ayrıca Şekil 4.6'daki 2. gruba ait ilaç iskelelerini ve birbirleriyle olan ilişkilerini gösterimi.....	53
Şekil 4.8 Favipiravir ve Molnupiravir'in iskele ve yan zincirleri gösteren 2B gösterimi	54
Şekil 4.9 Favipiravir ve Molnupiravir iskelelerinin düğüm grafikleri şeklinde gösterimi.....	55
Şekil 4.10 Favipiravir ve Molnupiravir ilaç moleküllerinin iskele ağacı gösterimi ..	56
Şekil 4.11 Önerilen yöntemin akış diyagramı.....	64
Şekil 4.12 Moleküler parmak izi yapısı	65
Şekil 4.13 Tanimoto benzerlik algoritmasının yapısı	66
Şekil 4.14 Tanimoto benzerlik sonuçları	69
Şekil 4.15 Tanimoto/Jaccard benzerlik analizi ile elde edilen ilaçların moleküler yapıları.....	70

Şekil 4.16 Atom çifti benzerlik sonuçları	71
Şekil 4.17 Kinaz inhibitörlerinin hedef protein bağlanma değerleri.....	74
Şekil 4.18 Önerilen metodolojinin şematik gösterimini bir akış diyagramı aracılığıyla gösterimi.....	84
Şekil 4.19 Bir MLP yapısı.....	87
Şekil 4.20 Eğitim sırasındaki kayıp değerleri	91
Şekil 4.21 Gerçek ve tahmin edilen çözünürlük değerlerinin grafiksel gösterimi.....	92
Şekil 4.22 Veri setinde yer alan ilaç moleküllerinden seçilen 16 ilaç	97
Şekil 4.23 Eğitim sırasındaki elde edilen AUC sonuçları.....	98
Şekil 4.24 Eğitim sırasındaki elde edilen kayıplar.....	99
Şekil 4.25 Çözünürlük diyagramı	99
Şekil 4.26 AUC sonuçları	100
Şekil 4.27 Confusion matrisi.....	101
Şekil 4.28 İlaç-protein etkileşimlerini analiz etmek için önerilen yaklaşımının model yapı gösterimi.....	107
Şekil 4.29 TMAP (Ağaç MAP) yöntemi kullanılarak FDA onaylı ilaçların görselleştirilmesi	114
Şekil 4.30 FDA onaylı tüm ilaçlar arasında Klorpromazin'in yapısı	115
Şekil 4.31 Klorpromazinin 2B ve 3B gösterimi.....	115
Şekil 4.32 6LU7 proteininin tüm proteinler arasında temsili.....	116
Şekil 4.33 6LU7 proteininin tüm proteinler arasından temsiline farklı gösterimi..	117
Şekil 4.34 6LU7 Proteazın yapısı	118
Şekil 4.35 Önerilen modele göre iterasyon sayısına bağlı olarak test ve eğitim aşamalarında elde edilen kayıp değerlerinin grafiksel gösterimi	122
Şekil 4.36 Kodlayıcılara göre ROC Eğrisi kullanılarak ilaç-protein hedef etkileşimi tahmin yönteminin performansları	123
Şekil 4.37 Ortalama Kare Hata (MSE) (a), Pearson Korelasyon (b) ve Uyum Endeksi (c) 100 iterasyon için elde edilen sonuçlar.....	127
Şekil 4.38 Yapay sinir ağ yapısı.....	134
Şekil 4.39 Ebola tedavisinde ilaç-ilaç benzerliğini ve moleküler iskeleleri analiz etmek için önerilen yaklaşımın akış şeması.....	135
Şekil 4.40 Invacator ilacı için iskele ağı, ilacın kompozit iskelelere ayrıştırılabileceği tüm olası yolları göstermektedir.	137

Şekil 4.41 Benzerliği çıkarılan 7 ilacın Scaffold grafik ağı algoritması tarafından oluşturulan bu ilaçların moleküler iskelelerinin temsili.....	140
Şekil 4.42 Yedi (7) ilaca ait iskelelerin grafiksel gösterimi, hangi iskelenin hangi düğümle ilişkili olduğunu ve aralarındaki bağlantılarının gösterimi	141
Şekil 4.43 Grafik ağı algoritması tarafından yedi (7) ilaç için oluşturulan iskelelerin moleküler yapıları da dahil olmak üzere gösterimi.....	142



TABLÖLAR LİSTESİ

Tablo 4.1 Çalışmada kullanılan ilaçlar ve genel özellikleri	41
Tablo 4.2 Atom-çifti analizi ile favipiravir ilacının kimyasal yapısı ve diğer ilaçların benzerlik sonuçları	48
Tablo 4.3 İlaç veri setleri	63
Tablo 4.4 Birleştirilmiş SARS ilaçları ve akciğer kanseri ilaçlarının benzerlik değerlerini gösterimi.	70
Tablo 4.5 ESOL veri seti.....	85
Tablo 4.6 Modellerin çözünürlük tahmin hata miktarları	90
Tablo 4.7 Analiz sonucu tahmin edilen çözünürlük değerleri	97
Tablo 4.8 Modelin performans metriklerinin sonuçları	102
Tablo 4.9 Diğer ilaçlar hakkında genel bilgiler	112
Tablo 4.10 6LU7 proteini ilgili genel bilgiler.....	116
Tablo 4.11 İlaç-protein bağlanma enerjilerinin değerleri	119
Tablo 4.12 İlaç-protein etkileşim sonuçları	120
Tablo 4.13 İlaç kodlamalar ve açıklamaları	121
Tablo 4.14 Hedef kodlamalar ve açıklamaları	121
Tablo 4.15 Atom çifti analizi ile favipiravir ilacının kimyasal yapısı ile diğer ilaçların benzerlik sonuçları	138

SİMGELER VE KISALTMALAR LİSTESİ

Simge	Açıklama
G	Grafik
V	Düğüm
E	Kenar
v_i	i.inci düğüm
v_j	j.inci düğüm
e_{ij}	İki düğüm arası bir kenar
N	Düğüm sayısı
D	Derece matrisi
L	Laplace
a	Hiper parametre sayısı
b	bias
σ	Sigma (Aktivasyon fonksiyonu)
h_v	Düğüm v'nin özelliği
S	İlaç ve analogu arasındaki benzerlik
A	İlaca ilişkin aktivite
T_i	Atom çiftinin yol uzunluğu
n_i	İlaçtaki i.hetero atom çiftinin sayısı
Δn_i	İlaçtaki i.hetero atom çiftinin analogundan farkı
T_x	Düğümleri ilaca ait altyapıları olan bir iskele ağacı
$w_{komşu}$	Komşu için ağırlık matrisi
w_{kendi}	Kendisi için ağırlık matrisi
c	Düğüm özellik vektör boyutu
X	Bir grafiğin düğüm özellikleri
u	Herhangi bir düğümü ifade eder
v	Herhangi bir kenarı ifade eder
h_u	U için gömme
N_u	Grafik komşuluğu

$m_{N(u)}$	Grafik komşuluktan toplanan mesaj
k	İterasyon
h_u^{k-1}	Bir önceki gömme özelliği
z_u	Gömme özellikleri
s_i^k	Durum matrisi
V_i	Temsili A molekülü ifadesi
V_j	Temsili B molekülü ifadesi
P	Protein konsantrasyonu
$k_{açık}$	Birleşme hız sabiti
$k_{kapalı}$	Ayrışma hız sabiti
m	Mesaj geçirme fonksiyonu
K_{leaf}	Yaprak düğümlerin ifadesi
$K - K_{leaf}$	Yaprak olmayan düğümlerin ifadesi

Kısaltma

Açıklama

GSA	Grafik Sinir Ağları
ML	Makine Öğrenmesi (Machine Learning)
DL	Derin Öğrenme (Deep Learning)
NLP	Doğal Dil İşleme (Natural Language Processing)
FDA	Food and Drug Administration (Gıda ve İlaç Dairesi)
QSAR	Quantitative Structure-Activity Relationships (Kantitatif Yapı-Aktivite İlişkileri)
KG	Knowledge Graph (Bilgi Grafiği)
YZ	Yapay Zekâ (Artificial Intelligence)
COVID-19	Coronavirus Disease 2019
Sars-Cov-2	Severe Acute Respiratory Syndrome Coronavirus 2
BAP	Bilimsel Araştırma Projesi

EGFR	Epidermal Growth Factor Receptor (Epidermal Büyüme Faktörü Reseptörü)
AML	Acute Myeloid Leukemia (Akut Myeloid Lösemi)
KD	Dissociation Constant (Dissosiyasyon Sabiti)
GEFA	Graph-in-Graph Neural Network (Grafik- içinde-Grafik Sinir Ağı)
AP	Atom Pair (Atom Çifti)
DTA	Drug Target Affinity (İlaç Hedef Afinitesi)
RNA	Ribonucleic Acid (Ribonükleik Asit)
P	Protein
kon	Association Rate Constant (Bağlanma Oranı Sabiti)
koff	Dissociation Rate Constant (Dissosiyasyon Oranı Sabiti)
AUC	Area Under Curve (Eğri Altında Kalan Alan)
GCN	Grafik Evrişimli Ağ
GAT	Grafik Dikkat Ağları
GIN	Grafik İzomorfizm Ağları
2B	İki boyutlu
3B	Üç boyutlu
SNG	Scaffold grafik ağı
WHO	Dünya Sağlık Örgütü
RDkit	Kimya bilgi işlem kiti
DGL	Derin grafik kütüphanesi
SMILES	Basitleştirilmiş Moleküler Girdi Hattı Giriş Sistemi
DTI	İlaç-Hedef Etkileşimi
MLP	Çok katmanlı algılayıcılar
ANN	Yapay sinir ağları
RF	Rastgele ormanlar
SVM	Destek vektör makineleri
ET	Extra ağaçlar

k-NN	En yakın komşu sinir ağıları
GB	Gradyan boosting
ESOL	Tahmini yapılmış sulu çözünürlükler
RMSE	Ortalama karekök hatası
MAE	Ortalama mutlak hata
R-kare	R kare
AUC	Eğri altında kalan alan
TP	Doğru pozitif
TN	Doğru negatif
FP	Yanlış doğru
FN	Yanlış negatif
CNN	Konvolüsyon sinir ağıları
RNN	Yinelemeli sinir ağıları

ÖZET

GRAFİK SİNİR AĞLARI İLE İLAÇ KEŞFİ

GİDER, Veysel

Doktora, Elektrik-Elektronik Mühendisliği Bölümü

Danışman: Doç. Dr. Cafer BUDAK

Temmuz 2024, 187 sayfa

İlaç keşfi, yüksek maliyetler, düşük başarı oranları ve uzun süreçler nedeniyle zorlu ve karmaşık bir alandır. COVID-19 pandemisi gibi küresel sağlık krizleri, bu sürecin daha hızlı, etkili ve yenilikçi çözümler gerektirdiğini ortaya koymuştur. Bu tez çalışması, grafik sinir ağları (GSA) kullanarak ilaç keşfi süreçlerini hızlandırmayı ve maliyetleri azaltmayı amaçlamaktadır. GSA'lar, düğümler ve kenarlarla ifade edilen grafik yapılarının analizine olanak tanıyan güçlü algoritmalar. Bu algoritmalar, moleküler yapıları ve etkileşimleri modelleyerek ilaç-protein bağlanma tahmini, ilaç benzerliği analizi, ilaç iskeleleri çıkarma ve ilaç yan etkilerinin tahmini gibi çeşitli uygulamalarda etkili bir şekilde kullanılabilir. Özellikle ilaçların yeniden kullanımı stratejisi, mevcut onaylı ilaçların yeni tedavi alanlarına adapte edilmesi veya farklı hastalıkların tedavisinde kullanılması anlamına gelir. Bu strateji, pandemi gibi acil durumlarda hızlı ve etkili çözümler sunma potansiyeline sahiptir. GSA'lar, bu yeniden kullanım sürecinde önemli bir rol oynayarak, mevcut ilaçların yeni hedefler için uygunluğunu hızlı ve doğru bir şekilde belirleyebilir. GSA'nın bilim dünyasındaki ilerleyişi, özellikle Türkiye'de öncü çalışmalar yapılmasını sağlamıştır. Öncü çalışmalardan biri olma özelliği taşıyan bu tez çalışmasında, DrugBank ve PubChem gibi kaynaklardan elde edilen moleküler yapılar, GSA modelleri ile analiz edilmiş ve ilaç-protein etkileşimleri tahmin edilmiştir. Kullanılan yöntemler arasında atom çifti benzerlik analizi, Tanimoto benzerliği ve moleküler parmak izi teknikleri bulunmaktadır. Ayrıca kinaz inhibitörleri üzerinde de yoğunlaşmıştır. Kinaz inhibitörleri, hedef proteinlerle etkileşime girerek aktivitelerini inhibe eden bileşiklerdir ve kanser tedavisinde kritik bir rol oynar. Bulgular, GSA'ların COVID-19, Ebola gibi acil durumlarda ilaç keşfi ve yeniden kullanımındaki etkinliğini göstermiştir. GSA'ların ilaç keşfi süreçlerini hızlandırma ve maliyetleri azaltma potansiyelini ortaya koymaktadır. GSA'ların ilaç keşfi ve geliştirme süreçlerinde daha geniş kapsamlı araştırmalarda kullanılmasının ve bu alandaki teknolojik gelişmelerin yakından takip edilmesinin önemi vurgulanmaktadır.

Anahtar Kelimeler: İlaç keşfi, İlaç Yeniden Kullanım, Yapay Zeka, Grafik Sinir Ağları

ABSTRACT

DRUG DISCOVERY WITH GRAPH NEURAL NETWORKS

GIDER, Veysel

PhD of Science in Department of Electric-Electronic Engineering

Supervisor: Assoc. Prof. Dr. Cafer BUDAK

July 2024, 187 pages

Drug discovery is a challenging and complex field characterized by high costs, low success rates, and lengthy timelines. Global health crises, such as the COVID-19 pandemic, have underscored the urgent need for faster, more effective, and innovative solutions in this domain. This thesis aims to accelerate drug discovery processes and reduce associated costs by utilizing Graph Neural Networks (GNNs). GNNs are powerful algorithms capable of analyzing graph structures, represented by nodes and edges, allowing for the modeling of molecular structures and interactions. These algorithms can be effectively applied in various drug discovery applications, including drug-protein binding prediction, drug similarity analysis, scaffold extraction, and the prediction of drug-induced side effects. A particularly promising strategy in this context is drug repurposing, which involves adapting existing approved drugs to new therapeutic applications or using them to treat different diseases. This strategy holds significant potential for providing rapid and effective solutions during emergencies, such as pandemics. GNNs play a crucial role in this repurposing process by quickly and accurately determining the suitability of existing drugs for new targets. The advancement of GNN technology has facilitated pioneering research, especially in Turkey. This thesis exemplifies such pioneering efforts by analyzing molecular structures obtained from resources like DrugBank and PubChem using GNN models to predict drug-protein interactions. The methods employed include atom pair similarity analysis, Tanimoto similarity, and molecular fingerprint techniques. Furthermore, the study focuses on kinase inhibitors, compounds that interact with target proteins to inhibit their activities, playing a critical role in cancer therapy. The findings demonstrate the efficacy of GNNs in drug discovery and repurposing during emergencies like COVID-19 and Ebola, highlighting their potential to expedite the drug discovery process and reduce costs. The importance of further research on the broader application of GNNs in drug discovery and development and the need to keep abreast of technological advancements in this rapidly evolving field are emphasized.

Keywords: Drug Discovery, Drug Repurposing, Artificial Intelligence, Graph neural networks

1. GİRİŞ

Tıbbi bilimlerdeki hızlı ilerlemeler, insan sağlığını iyileştirmeye yönelik kritik bir hedef olarak ilaç keşfi sürecine odaklanmayı vurgulamaktadır. Ancak bu süreç, karmaşıklığı, yüksek maliyetleri, düşük başarı oranları ve uzun zaman çizelgeleri gibi zorluklarla dolu bir süreçtir. Bu zorluklar, araştırmacıları ve ilaç endüstrisini daha etkili, hızlı ve yenilikçi çözümler aramaya yönlendirmiştir.

İlaç keşfi sürecinde karşılaşılan temel sorunlardan biri yüksek maliyettir. Özellikle yeni kimyasal yapıların geliştirilmesi için gereken fonların artması, bu maliyeti daha da belirgin hale getirmiştir. Amerika Birleşik Devletleri Gıda ve İlaç İdaresi'nin (FDA) 2003 yılında onayladığı bir kimyasal yapı için ayrılan 802 milyon dolarlık bütçe, ilaç endüstrisinin finansal zorluklarını gözler önüne sermektedir (Light ve Warburton, 2011). Aynı zamanda, ilaç şirketlerinin 2015 yılında toplamda harcadığı 26 milyar dolarlık bütçe, ilaç keşfi sürecinin ne kadar maliyetli olduğunu vurgulamaktadır (Annett, 2021).

Şubat 2022 itibarıyla dünya çapında altı milyondan fazla ölümle yaklaşık 385 milyon insanı etkileyen COVID-19 salgını, özellikle yeni antiviral ilaçlar ve aşılar için hem zaman hem de maliyet açısından büyük bir zorluk oluşturmuştur (Alexandridi vd., 2022). Bu olağanüstü durum, ilaç endüstrisinin hızlı ve etkili çözümler bulma baskısını artırmıştır. Bu bağlamda, ilaç keşfi sürecini iyileştirmek için yeni teknolojilere ve yöntemlere olan talep daha da belirgin hale gelmiştir. Ancak, yüksek maliyetin ötesinde, ilaç keşfi süreci, bir ilacın pazarlamaya sunulmasından önce yoğun bir klinik araştırma ve geliştirme aşamasını içermektedir. FDA'dan nihai onayının alınması, sadece ilacın etkili olduğunu değil, aynı zamanda güvenli olduğunu da onaylamayı gerektirir. Bu kapsamlı değerlendirme, yıllarca süren bir süreçtir ve milyonlarca dolarlık yatırımın geri dönmesini sağlamak adına kaçınılmaz bir adımdır.

Klinik deneme aşamasında ise başarı oranları düşüktür; insanlarda test edilen ilaçların yaklaşık yüzde 90'ı başarısız olmaktadır (D. Sun vd., 2022). Bu durum, sadece finansal kaynakların israf edilmesine değil, aynı zamanda umut vadeden birçok molekülün gerçek dünya uygulamalarında etkili olamamasına işaret etmektedir. Bu, birçok araştırmacı ve ilaç şirketi için büyük bir risk oluşturur ve başarısızlıklar ilaç

endüstrisindeki maliyetleri daha da artırabilir. Bu durum, farklı stratejilere yönelmeye neden olmuştur.

İlaçların yeniden kullanımı stratejisi (B. Karaman ve Sippl, 2019; Klug vd., 2016; Oprea ve Mestres, 2012; Zhan vd., 2022), mevcut ilaçların yeni tedavi alanlarına uyarlanması veya farklı hastalıkların tedavisinde kullanılması anlamına gelir. Bu strateji, özellikle pandemi gibi acil durum senaryolarında, hızlı ve etkili bir şekilde değerlendirilmesini sağlamaktadır. COVID-19 salgını, bu stratejinin önemini bir kez daha vurgulayarak, onaylanmış ilaçların listesinden güçlü ilaçların belirlenmesi konusunda önemli bir ilgiyi beraberinde getirmiştir.

Pandemi senaryosu, hızlı ve etkili bir şekilde tedavi seçeneklerinin bulunmasını gerektirir. İlaçların yeniden kullanımı stratejisi, bu ihtiyaca yanıt verme potansiyeli taşır. Onaylanmış ilaçlar, klinik denemelerden geçmiş, güvenliği ve etkinliği değerlendirilmiş ve genellikle rutin olarak belirli hastalıkların tedavisinde kullanılmaktadır. Bu ilaçlar, yeni bir hastalık veya acil bir durum ortaya çıktığında hızlı bir çözüm sunma potansiyeli taşıdıkları için büyük bir değere sahiptir. Özellikle COVID-19'a karşı ilaç geliştirme çabalarında, mevcut onaylı ilaçlardan potansiyel adayları belirleme süreci hızlandırılmıştır. Bilim insanları, virüsün yayılmasını durdurma veya hastalığın şiddetini azaltma potansiyeline sahip ilaçları hızlı bir şekilde tanımlamak için bu stratejiyi kullanmışlardır. Bu, yeni ilaç-hedef etkileşimlerinin keşfedilmesine ve bu etkileşimlerin farklı hastalıkların tedavisinde kullanılmasına olanak tanımıştır.

Ayrıca, onaylanmış ilaçların listesinden güçlü ilaçları belirleme sürecinde, bilgisayar destekli tasarım ve yapay zekâ (YZ) gibi teknolojiler de etkili bir şekilde kullanılmaktadır. Bu yöntemler, büyük veri setlerini analiz ederek potansiyel ilaç adaylarını hızlı bir şekilde değerlendirebilir ve mevcut ilaçlar arasında hedefe yönelik etkileşimleri tahmin edebilir.

Yapay zeka'nın tarihçesi

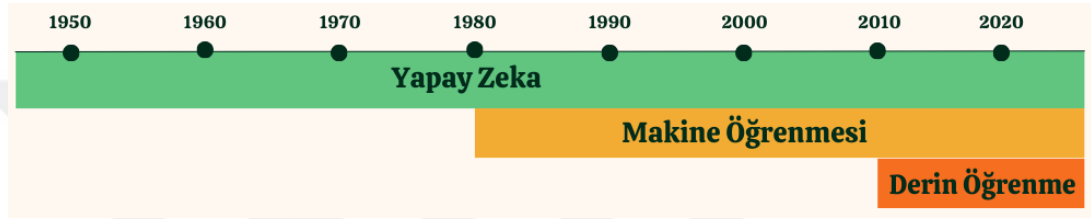
YZ, bilgisayarların insan zekâsı süreçlerini simüle etmesini ifade eder. Bu süreç, bilgiyi edinme, bilgi kullanımı için kurallar oluşturma, sonuç çıkarma ve kendi kendini düzeltme yeteneklerini kapsar. YZ, insan davranışlarını taklit eden bilgisayar tekniklerinin uygulanması olarak anlaşılmaktadır (Mak ve Pichika, 2019).

"Yapay zekâ" kavramı, 1956 yılında John McCarthy tarafından Dartmouth Konferansı'nda, "akıllı makineler yaratma bilimi ve mühendisliği" anlamında kullanılmıştır (McCarthy ve Hayes, 1981). McCarthy'nin bu tanımı günümüzde de geçerliliğini korumaktadır. YZ, bilgisayar bilimi, matematik, psikoloji, dilbilim, felsefe, nörobilim ve yapay psikoloji gibi çeşitli disiplinlerden alınan içgörülerini birleştiren, çok disiplinli bir alandır. Son dönemdeki entelektüel ve mühendislik alanındaki gelişmeler, bu alanın teorik çalışmalardan, yaşamın çeşitli alanlarındaki sorunları çözmeye yönelik akıllı sistemlerin uygulamasına geçişini hızlandırmıştır. Bu uygulamalar, doğal dil işleme, konuşma tanıma, bilgisayarlı görme, otonom robotlar ve uzman sistemler gibi çeşitli alanları kapsar (Yang vd., 2019). YZ'nin ele aldığı çeşitli problemlere rağmen, bilgi temsili, çözüm arama, mantıksal akıl yürütme ve makine öğrenmesi gibi temel yöntemler, her durumda kritik bir rol oynamaktadır.

1930'larda Alan Turing, bugün "evrensel Turing makinesi" olarak bilinen ve tüm olası bilgisayarları simüle edebilen bir kavram geliştirmiştir. YZ'nin tarihsel gelişimi, zaman içinde yükselme ve düşüşlerle karakterize edilmiştir (Yang vd., 2019). 1950'ler ve 1960'lar, bu alanın erken dönemdeki altın çağını temsil eder. Bu dönemde, sezgisel araştırmalar, mantıksal akıl yürütme kavramları ve insan-makine etkileşimine yönelik sembolik yöntemler geliştirilmiştir. 1967 yılında, matematiksel teoremlerin ve ifadelerin mantıksal çıkarımlarını gerçekleştiren ilk makine olan STUDENT, erken YZ örneği olarak tasarlanmıştır (Bobrow, 1967). Ayrıca, 1966 yılında sınırlı da olsa insan diyalogunu taklit edebilen ELIZA adlı bir makine geliştirilmiştir (Weizenbaum, 1966). 1980'lerin başında YZ, ikinci bir zirveye ulaşmıştır. Bu dönemde, çok katmanlı ileri beslemeli sinir ağları ve geri yayılım algoritması gibi YZ'ye ilişkin matematiksel modellerde önemli ilerlemeler kaydedilmiştir (Baum, 1988; Rumelhart vd., 1988). Bu araçlar, evrenin soyut bir modelini oluşturmayı ve bu modeli giriş verileriyle güncelleyerek öğrenmeyi (geri beslemeden öğrenme) mümkün kılar. Bu tür yöntemler,

kimya ve moleküler biyolojiye ilk olarak, protein dizisi bilgisinden ikincil yapının tahmin edilmesi şeklinde giriş yapmıştır (Qian ve Sejnowski, 1988).

YZ teknolojileri, verilerden kavramlar ve ilişkiler çıkararak, insanların veri modellerine bağımlı olmadan öğrenme yeteneklerini keşfetmelerine olanak tanır. Bu teknolojiler arasında bilgisayarla görme, derin öğrenme (DL), makine öğrenimi (ML), doğal dil işleme, robotik, konuşma, denetimli ve denetimsiz öğrenme gibi uygulamalar yer almaktadır.



Şekil 1.1 Yapay zekâ ve alt dallarının gelişimi

ML, bilgisayar sistemlerinin açıkça programlanmadan, deneyimle belirli bir görevdeki performanslarını otomatik olarak geliştirmelerini sağlayan algoritmalar ve istatistiksel modeller oluşturmayı içeren YZ'nin bir alt kümesidir.

Özünde ML, bir bilgisayar sisteminin verilerdeki kalıpları tanıması ve bu kalıplara dayalı tahminler veya kararlar vermesi için eğitilmesini içerir. Bu, bir modeli veya algoritmayı eğitmek için büyük veri setleri kullanarak ve ardından doğruluğunu ve etkinliğini değerlendirmek için modeli yeni veriler üzerinde test ederek elde edilir. Model daha sonra bir yinelemeli öğrenme süreci aracılığıyla rafine edilebilir ve optimize edilebilir.

Üç ana ML türü vardır: denetimli öğrenme, denetimsiz öğrenme ve pekiştirmeli öğrenme. Denetimli öğrenmede model, her girdi için doğru çıktının sağlandığı etiketli bir veri kümesi üzerinde eğitilir. Bu, modelin yeni, etiketsiz veriler için çıktıları tahmin etmeyi öğrenmesini sağlar. Denetimsiz öğrenmede, model etiketlenmemiş bir veri kümesi üzerinde eğitilir ve amaç verilerdeki kalıpları ve yapıyı tanımlamaktır. Takviyeli öğrenme, modelin geri bildirim ve ödüllere dayalı kararlar alması için eğitilmesini içerir ve deneme yanılma yoluyla öğrenmesine olanak tanır.

ML, görüntü ve konuşma tanımadan doğal dil işleme ve tahmine dayalı analitiğe kadar geniş bir uygulama yelpazesine sahiptir. Finans, sağlık ve pazarlama gibi sektörlerde tahminlerde bulunmak, süreçleri optimize etmek ve görevleri otomatikleştirmek için kullanılır. ML ayrıca bilimsel araştırmalarda büyük veri kümelerini analiz etmek ve karmaşık sistemlerdeki kalıpları ve ilişkileri tanımlamak için kullanılır. Her biri belirli bir sorunu çözmek için tasarlanmış çok sayıda makine öğrenimi algoritması vardır. Bununla birlikte, en popüler makine öğrenimi algoritmalarından bazıları şunlardır:

- Doğrusal Regresyon: Bir bağımlı değişken ile bir veya daha fazla bağımsız değişken arasındaki ilişkiyi modellemek için kullanılır.
- Lojistik Regresyon: Bir veya daha fazla yordayıcı değişkene dayalı olarak ikili veya kategorik bir sonucun olasılığını modellemek için kullanılır.
- Karar Ağaçları (DT): Kararları veya birden fazla koşula dayalı kararları modellemek için kullanılır.
- Rastgele Orman (RF): Çok sayıda karar ağacı oluşturan ve doğruluğu artırmak için tahminlerini birleştiren bir topluluk öğrenme yöntemi.
- Destek Vektör Makineleri (SVM'ler): Optimum hiper düzlemi bularak verileri farklı sınıflara ayırmak için kullanılır.
- Naive Bayes (NB): Birden fazla tahmin edici değişkene dayalı olarak kategorik bir sonucun olasılığını tahmin etmek için kullanılır.
- K-En Yakın Komşular (KNN): Verileri en yakın k komşuya göre sınıflandırmak için kullanılır.
- Sinir Ağları: Verilerdeki kalıpları ve ilişkileri tanıyabilen, genellikle görüntü ve konuşma tanıma için kullanılan bir dizi algoritma.

Bunlar, popüler ML algoritmalarının sadece birkaç örneğidir ve eldeki belirli soruna ve verilere bağlı olarak çok daha fazlası vardır. Genel olarak makine öğrenimi, büyük ve karmaşık veri kümelerini analiz etmek ve anlamlandırmak için güçlü bir araçtır ve birçok sektörü ve alanı dönüştürme potansiyeline sahiptir.

- Derin Öğrenme (DL): Derin öğrenme, büyük karmaşık veri kümelerini analiz etmek ve bunlardan bir şeyler öğrenmek için birden fazla

katmana sahip sinir ağlarının kullanılmasını içeren bir makine öğrenimi alt kümesidir. Bilginin işlenmesi ve yorumlanmasında insan beyninin işleyişini taklit etmeyi amaçlayan bir YZ biçimidir.

DL’de, bir sinir ağı birbirine bağlı düğümlerden oluşan birden fazla katmandan oluşur ve her katman giriş verilerinden belirli özellikleri çıkarmaktan sorumludur. Bir katmandan elde edilen çıktı bir sonraki katmana girdi olarak aktarılır ve sonraki her katman bir önceki katman tarafından öğrenilen özellikler üzerine inşa edilir. Bu süreç, ağ nihai bir çıktı üretene kadar devam eder.

DL’nin temel avantajlarından biri, geleneksel makine öğrenimi teknikleri kullanılarak analiz edilmesi zor olabilen görüntüler, videolar ve doğal dil gibi yapılandırılmamış verileri işleme yeteneğidir. DL, görüntü ve konuşma tanıma, doğal dil işleme ve otonom araçlar da dahil olmak üzere çok çeşitli uygulamalarda kullanılmıştır.

Eylem halindeki derin öğrenmenin bir örneği görüntü tanımadır. DL modelleri, her bir görüntünün içerdiği nesne veya özelliğe göre etiketlendiği büyük görüntü veri kümeleri üzerinde eğitilebilir. Sinir ağı daha sonra görüntülerdeki kenarlar, şekiller ve dokular gibi temel özellikleri tanımlamayı öğrenir ve bu bilgiyi yeni görüntüleri özelliklerine göre sınıflandırmak için kullanır.

Örneğin, kedi ve köpek görüntülerinden oluşan bir veri kümesi üzerinde eğitilen bir derin öğrenme modeli, kulaklarının şekli, kürklerinin rengi ve vücutlarının şekli gibi özelliklere dayanarak iki hayvan arasında ayrım yapmayı öğrenebilir. Model eğitildikten sonra, bu görüntüleri daha önce hiç görmemiş olsa bile yeni kedi ve köpek görüntülerini doğru bir şekilde sınıflandırabilir. Aşağıdakiler de dahil olmak üzere çeşitli derin öğrenme türleri vardır:

- **Evrişimli Sinir Ağlar (CNN):** Bu derin öğrenme türü öncelikle görüntü ve video tanıma görevleri için kullanılır. CNN'ler, görüntülerden özellikler çıkarabilen ve bunları bu özelliklere göre sınıflandırabilen birbirine bağlı düğümlerden oluşan birçok katmanla görsel verileri işlemek için tasarlanmıştır.

- Yinelemeli Sinir Ağlar (RNN): RNN'ler, zaman serisi verileri veya doğal dil verileri gibi sıralı verileri işlemek için tasarlanmıştır. Önceki girdilerin hafızasını tutabilir ve bu bilgileri tahminler yapmak veya yeni diziler oluşturmak için kullanabilirler.
- Üretken Çekişmeli Ağlar (GAN): GAN'lar, bir jeneratör ve bir diskriminatör olmak üzere iki sinir ağından oluşan bir tür derin öğrenme modelidir. Üretici ağ yeni veriler üretirken, ayırt edici ağ üretilen verilerin gerçek mi yoksa sahte mi olduğunu değerlendirir. Bu ağlar genellikle gerçekçi görüntüler, videolar veya sesler üretmek için kullanılır.
- Otomatik kodlayıcılar: Otomatik kodlayıcılar, gizli uzay adı verilen giriş verilerinin sıkıştırılmış bir temsilini öğrenmek için tasarlanmış sinir ağlarıdır. Boyut azaltma veya özellik çıkarma için kullanılabildikleri gibi yeni veri üretmek için de kullanılabılır.
- Derin İnanç Ağları (DBN'ler): DBN'ler, verilerin hiyerarşik temsillerini çıkarmak için denetimsiz öğrenmeyi kullanan çok katmanlı sinir ağlarıdır. Genellikle görüntü tanıma ve konuşma tanıma gibi görevler için kullanılırlar.

Bu derin öğrenme türlerinin her birinin kendine özgü güçlü ve zayıf yönleri vardır ve farklı uygulama türleri için en uygunudur. Genel olarak DL, makinelerin karmaşık verilere dayalı olarak öğrenmesini ve karar vermesini sağlayarak birçok sektör ve uygulamada devrim yaratma potansiyeline sahip güçlü bir tekniktir.

İlaç keşfi ve geliştirilmesinde yapay zeka

İlaç keşif süreci, mevcut literatürün okunması ve analiz edilmesinden potansiyel ilaçların hedeflerle etkileşim yollarının test edilmesine kadar uzanır. İlaç keşfinin klinik öncesi geliştirme aşaması, potansiyel ilaç hedeflerinin hayvan modellerinde test edilmesini içerir. Bu aşamada YZ'nin kullanılması denemelerin sorunsuz bir şekilde yürütülmesine yardımcı olabilir. Araştırmacıların bir ilacın bir hayvan modeliyle nasıl etkileşime girebileceğini daha hızlı ve başarılı bir şekilde tahmin etmelerini sağlayabilir. İlaç klinik öncesi gelişim sürecinden geçtikten ve FDA'dan onay aldıktan sonra, araştırmacılar ilacı insan katılımcılarla test etmeye başlar. Genel olarak bu dört

aşamalı bir süreçtir ve genellikle üretim yolculuğunun en uzun ve en pahalı aşaması olarak kabul edilir.

İlaç keşfinde YZ için en önemli fırsat, ilaçların yeniden tasarlanmasıdır. Donner ve arkadaşları, gen ifadesi verilerine dayalı olarak bileşik fonksiyonel benzerliğini ölçmek için yeni bir yöntem bildirmiştir. Bu yaklaşım, bütünlüğü bozan verileri en aza indirerek bir moleküldeki fonksiyonel grubun yapısını en yakın analoglarla taşıyan bileşikler bulmak için DNN'nin avantajlarını kullanmaktadır. Bu yöntemin, bileşikler yapısal olarak farklı olsa bile ortak terapötik ve biyolojik hedefleri olan ilaçları belirleyebildiği ve böylece bileşikler arasında daha önce bildirilmemiş fonksiyonel ilişkileri ortaya çıkardığı bildirilmektedir. Bu yaklaşım DeepCodex'te çevrimiçi olarak kullanıma sunulmuştur. Bu metrik, bileşikler arasında daha önce bilinmeyen işlevsel ilişkileri ortaya çıkarmış ve yapısal olarak farklı olmalarına rağmen biyolojik hedefleri paylaşan bileşiklerin tanımlanmasına olanak sağlamıştır (Donner vd., 2018).

Bu sürecin amacı, son zamanlarda ortaya çıkan YZ uygulamalarından biri olan Grafik Sinir Ağlarını (GSA, İngilizce kısaltması GNN) kullanarak ilaç keşif sürecini hızlandırmak ve ilaç-ilaç benzerliklerini, ilaç-hedef protein ilişkilerini ve ilaç-ilaç etkileşimlerini çıkarmaktır. Bu sayede öncelikli hedefi insan sağlığı olan ilaç sektörüne ve literatüre katkı sağlamayı amaçlıyoruz. Şu anda en yaygın olarak uygulananlardan biri olan GSA adlı DL algoritmasının bir alt türüne odaklanmaktadır. Yeni olmasına rağmen, GSA kullanan DL algoritmalarının kullanımı, yüksek doğruluk oranları nedeniyle bazı araştırmacılar tarafından son teknoloji olarak kabul edilmektedir (Alves vd., 2022).

YZ, ilaç keşfi sürecinde geleneksel yöntemlere alternatif bir perspektif sunarak yeni ufuklar açan önemli bir alan haline gelmiştir. YZ'nin algoritmalar ve veri analizi kullanarak deneyimden öğrenme yeteneği, ilaç keşfinin farklı aşamalarında etkili bir şekilde uygulanabilir. Bu teknoloji, büyük veri kümelerini hızlı bir şekilde işleme kapasitesi ile öne çıkarken, ilaç keşfi sürecine çeşitli avantajlar kazandırmaktadır. YZ'nin ilaç keşfi alanında sunduğu avantajlardan biri, yeni bilgilerin keşfedilmesi ve karmaşık problemlerin çözülmesidir. Algoritmalar, büyük veri setlerindeki desenleri tanımlayabilir ve bu verileri analiz ederek potansiyel ilaç adaylarını belirleyebilir.

Ayrıca, YZ, geleneksel yöntemlerle belki de fark edilemeyecek ilişkileri tespit ederek ilaç keşfinde inovasyon sağlar.

YZ'nin ilaç keşfi alanında sunduğu bir diğer önemli avantaj ise üretken modeller, de novo ilaç tasarımı, çok hedefli ilaçlar ve kişiselleştirilmiş tıp gibi alanlarda potansiyel gelişmeleri teşvik etmesidir (Banjare vd., 2023; Proschak vd., 2019). Bu yenilikçi yaklaşımlar, belirli hastalıklara yönelik özelleştirilmiş tedavilerin tasarlanmasını ve daha etkili ilaçların geliştirilmesini mümkün kılacak potansiyeller sunmaktadır.

GSA, COVID-19 pandemisi ile birlikte 2019'dan beri YZ'nin ilaç keşfi alanındaki en dikkat çekici tekniklerinden biri haline gelmiştir (Budak vd., 2023; Doshi & Chepuri, 2022). Moleküler yapıları ve etkileşimleri modelleme amacıyla kullanılan GSA, bir grafik yapısı üzerinde moleküllerin atomlarını ve bağlarını temsil eder. Bu sayede, moleküllerin kimyasal ve fiziksel özellikleri kodlanabilir. GSA, ilaç-protein bağlanma değerinin tahmininden ilaç benzerliği analizine, ilaç iskeleleri çıkarmadan ilaç yan etkilerinin tahminine kadar bir dizi uygulamada etkin bir rol oynamaktadır. GSA'nın kullanımı, ilaç keşfi sürecini hızlandırmak, maliyeti azaltmak ve yeni ilaç adaylarını keşfetmek için önemli bir araç olarak öne çıkmaktadır.

GSA'ların tarihi, araştırmacıların sinir ağlarını grafik yapıları verilerine uygulamanın yollarını keşfetmeye başladıkları 1980'lere kadar uzanabilir (Fontanesi vd., 2023; Scarselli vd., 2009). 1990'larda, ağaç yapıları verileri işlemek için RNN adı verilen bir tür sinir ağı geliştirildi (Richard Socher vd., 2011). RNN'ler daha sonra özyinelemeli bir mesaj iletim şeması eklenerek grafik yapıları verileri işleyecek şekilde genişletildi ve bu da GSA ile sonuçlandı (Gori vd., 2005). Ancak, 2010'ların başına kadar GSA'ların popülerlik kazanmaya başlaması, yeni mimarilerin ve öğrenme algoritmalarının geliştirilmesi sayesinde oldu (Bronstein vd., 2017). 2015 yılında, grafiklere evrimsel sinir ağı işlemleri uygulayan Grafik Evrimsel Ağ (GCN) tanıtıldı (Zhang vd., 2019). GCN, çeşitli kıyaslama veri kümelerinde son teknoloji sonuçlar elde ederek GSA'lara olan ilginin yenilenmesini sağladı. O zamandan beri, Grafik Dikkat Ağları (GAT'ler), GraphSAGE ve Grafik İzomorfizm Ağları (GIN'ler) dâhil olmak üzere çok çeşitli GSA mimarileri ve algoritmaları geliştirilmiştir (P. Li & Leskovec, 2022; M. Shi vd., 2021; Yao vd., 2024). Bu modeller, doğal dil işleme, bilgisayarla görme, ilaç keşfi ve öneri sistemleri gibi çeşitli alanlara uygulanmıştır.

Tez çalışması, COVID-19 gibi hızlı yayılan salgınlarla mücadelede alternatif çözümler geliştirmeyi amaçlamaktadır. Bu bağlamda, özel olarak tasarlanmış GSA yapılarını kullanarak hızlı ve etkili çözümler üretme potansiyelini araştıracaktır. Bu çerçevede elde edilen bulgular, acil durumlar ve salgın yönetimi için stratejik bir öneme sahip olabilir. Tez, ilaç keşfi ve salgın yönetimi alanlarında GSA'ların pratik uygulamalarını değerlendirmeyi amaçlamakta ve bu alandaki bilgi birikimini genişletmeyi hedeflemektedir.



2. ÖNCEKİ ÇALIŞMALAR

GSA ve ilaç keşfi hakkında, özellikle YZ tabanlı ilaç-ilaç benzerliği, ilaç-hedef protein etkileşimi ve ilaç-ilaç etkileşimi ile ilgili çeşitli çalışmalar mevcut literatürde bulunmaktadır.

Wang ve arkadaşları, RNA-hastalık ilişkisi tahmini üzerinde yaptıkları çalışmada, Dairesel RNA'ların (sirkRNA'lar) hastalıkların oluşumunda ve gelişiminde yaygın olarak rol oynadığını bu nedenle potansiyel hastalıkla ilişkili sirkRNA'ları tahmin etmek için derin öğrenme Grafik Evrişimli Ağlar (FastGCN) algoritması ile hızlı öğrenme tabanlı GCNCDA adlı bir hesaplama yöntemi önerdiler. Sonuç olarak, GCNCDA'nın potansiyel sirkRNA hastalığı ilişkilerini etkili bir şekilde tahmin edebileceğini ve biyolojik deneyler için oldukça güvenilir adaylar sağlayabileceğini gösterdiler (L. Wang vd., 2020).

J. Wang ve arkadaşları ise scGSA, tek hücreli RNA-Seq analizleri için yeni bir GSA önerdiler. Bu çalışmaya göre, tek hücreli RNA dizilimi (scRNA-Seq), dokuların, organizmaların ve karmaşık hastalıkların heterojenliğini ve dinamiklerini ortaya çıkarmak için yaygın olarak kullanılmaktadır, ancak analizleri hala gen ekspresyonundaki dizileme seyrekliği ve karmaşık diferansiyel kalıplar da dahil olmak üzere birçok büyük zorluktan muzdariptir. scRNA-Seq analizleri için hipotezsiz bir derin öğrenme çerçevesi sağlamak için scGSA'yi (tek hücreli grafik sinir ağı) tanıttılar. Bu çerçeve, GSA'lar ile hücre-hücre ilişkilerini formüle eder ve toplar ve sol kesilmiş bir karışım Gauss modeli kullanarak heterojen gen ekspresyon modellerini modeller. scGSA, üç yinelemeli çok modlu oto kodlayıcıyı entegre eder ve dört kıyaslama scRNA-Seq veri kümesinde gen atama ve hücre kümelemesi için mevcut araçlardan daha iyi performans gösterir. Postmortem beyin dokularından 13.214 tek çekirdekli bir Alzheimer hastalığı çalışmasında, scGSA hastalığa bağlı sinirsel gelişimi ve diferansiyel mekanizmayı başarıyla göstermiştir (J. Wang vd., 2021).

T. Wang ve diğerleri ise biyomedikal sınıflandırma için Multi-Omics Graph Convolutional Networks (MORONET) adlı yeni bir multi-omik bütünleştirici yöntem sundular. MORONET etkili multi-omik veri sınıflandırması için omiklere özgü öğrenmeyi ve çapraz omik korelasyon öğrenimini ortaklaşa araştırır. MORONET'in,

mRNA ekspresyon verilerini, DNA metilasyon verilerini ve miRNA ekspresyon verilerini kullanan çok çeşitli biyomedikal sınıflandırma uygulamalarından diğer son teknoloji denetimli multi-omik bütünleştirici analiz yaklaşımlarından daha iyi performans gösterdiğini sundular. Ayrıca, MORONET araştırılan hastalıklarla ilişkili farklı omik veri türlerinden önemli biyobelirteçleri tanımlayabilmektedir (T. Wang vd., 2021).

Gligoriyevic ve arkadaşları, “grafik evrişimli sinir ağları kullanarak yapı tabanlı protein fonksiyonu tahmini” adlı çalışmalarında, bir protein dili modelinden ve protein yapılarından elde edilen dizi özelliklerinden yararlanarak protein fonksiyonlarını tahmin etmeye yönelik bir Grafik Evrişimli Ağ olan DeepFRI'yi tanıttılar. Önerilen yöntem ile deneysel yapıların eğitim setini homoloji modelleriyle genişletmek, öngörülebilir fonksiyonların sayısını önemli ölçüde artırmamıza olanak tanıyabileceklerini gösterdiler (Gligoriyevic vd., 2021).

Nguyen ve arkadaşları ilaçları grafik olarak temsil eden ve ilaç-hedef afinitesini tahmin etmek için GSA'larını kullanan GraphDTA adlı yeni bir model önerdiler. GSA'ların sadece ilaç hedefi afinitesini derin olmayan öğrenme modellerinden daha iyi tahmin etmekle kalmayıp, aynı zamanda rakip DL yöntemlerinden daha iyi performans gösterdiğini ispatladılar. Sonuçlarında, DL modellerinin ilaç hedefi bağlanma afinitesi tahmini için uygun olduğunu ve ilaçların grafik olarak gösterilmesinin daha fazla iyileştirmeye yol açabileceğini doğruladılar (T. Nguyen vd., 2021).

Long ve arkadaşları, insan mikrop-ilâç ilişkilerini tahmin etmek için GCNMDA adlı yeni bir GCN tabanlı model önermiştir. Bu çerçevede, GCN'nin gizli katmanında benzer düğümlerin (mikroplar veya ilâçlar) benzer temsillere sahip olmasını sağlamak amacıyla Koşullu Rastgele Alan'dan daha etkin bir şekilde faydalanılmıştır. Ayrıca, sırasıyla ilâçlar ve mikroplar için değerli özellikleri öğrenmek için hem ilâç hem de mikrop benzerlik ağlarında yeniden başlatma tabanlı şema ile rastgele bir yürüyüş gerçekleştirdiler. Üç farklı veri kümesindeki deneysel sonuçlar, GCNMDA modellerinin sürekli olarak yedi son teknoloji ürünü yöntemden daha iyi performans elde ettiğini göstermiştir. SARS-CoV-2 ve iki antimikrobiyal ilâç (yani Siprofloksasin

ve Moksifloksasin) dahil olmak üzere üç mikrop için yapılan vaka çalışmaları, GCNMDA'nın potansiyel mikrop-ilaç ilişkilerini tanımlamadaki etkinliğini daha da doğruladı (Long vd., 2020).

Mao ve arkadaşları, çalışmalarında hastalar, karşılaşmalar, laboratuvar testleri ve ilaçlar olmak üzere dört tür tıbbi varlığı ilişkilendiren bir grafik oluşturdu. Bu grafik üzerinden, ilaç önerisi ve laboratuvar testi ataması yapmak amacıyla düğüm gömmelerini öğrenmek için bir GSA uyguladılar. MedGCN (akıllı bir tıbbi sistem) modellerini iki gerçek dünya veri kümesinde doğruladılar: NMEDW ve MIMIC-III. Her iki veri kümesindeki deneysel sonuçlar, modellerinde her iki görevde de en son teknolojiye göre daha iyi performans gösterebileceğini göstermektedir. Yenilikçi sistemlerinde, hekimlere ilaç reçeteleri hazırlamalarına ve potansiyel olarak gereksiz laboratuvar testlerinde maliyetlerden tasarruf etmelerine yardımcı olmak için umut verici ve güvenilir bir yol sağlayabileceğini gösterdiler (Mao vd., 2022).

Sacha ve arkadaşlarının çalışması, otomatik sentez planlamasında çeşitli kimyasal reaksiyonların sonuçlarını tahmin etme ve üretme zorluğunu ele almaktadır. Bu amaçla, Molekül Düzenleme Grafiği Dikkat Ağı (MEGAN) adlı bir uçtan uca kodlayıcı-kod çözücü modelini tanıtmışlardır. MEGAN, kimyasal reaksiyonları bir dizi grafik düzenlemesi olarak ifade eden modellerden ilham almıştır ve retrosentez tahminine (bir kimyasal reaksiyonun ürününden substratların tahmini) genişletilmiştir. Model, kimyasal reaksiyonların olası uzayını verimli bir şekilde keşfetmekte ve standart ölçütlerde en yüksek doğruluğu elde ederek mevcut yöntemlere göre üstünlük sağlamaktadır (Sacha vd., 2021).

Budak ve arkadaşları, çalışmalarında moleküllerin ve proteinlerin yapısal bilgilerini kullanarak, GSA tabanlı GEFA modeli ile yeniden amaçlanan ilaç adaylarının bir listesini hazırlamışlardır. Analiz için halka açık veri tabanları olan DrugBank ve PubChem veri setlerinden elde edilen veriler kullanılmıştır. Tanimoto/Jaccard benzerlik analizi kullanılarak, COVID-19 tedavisinde kullanılan ilaçlar ile diğer hastalıkların tedavisinde kullanılan ilaçlar karşılaştırılarak benzer ilaçların bir listesi oluşturulmuştur. Bu ilaçlar daha sonra akciğer kanserinde kullanılan ilaçlarla karşılaştırılmış ve bir ilaç ile bir hedef arasındaki bağlanma gücü hesaplanarak yeniden

kullanılabilecek ilaçlar belirlenmiştir. Çalışmadan elde edilen kinaz inhibitörleri (erlotinib, lapatinib, vandetanib, pazopanib, cediranib, dasatinib, linifanib ve tozasertib), ABL2, ABL1, EGFR, AAK1, FLT3 ve JAK1 gibi hedefleri bloke eden ajanların (gefitinib, osimertinib, fedratinib, baricitinib, imatinib, sunitinib ve ponatinib) veya antiviral tedavilerin (ribavirin, ritonavir-lopinavir ve remdesivir) bir kombinasyonu olarak kullanılabilir (Budak vd., 2023).

Zhao ve arkadaşları, çalışmalarına göre yeni ilaç-hedef etkileşimlerinin (DTI'lar) tanımlanması, ilaç keşfinde önemli ancak zaman alıcı ve maliyetli bir adımdır. Son yıllarda, bu dezavantajları hafifletmek için araştırmacılar, hesaplamalı yaklaşımları kullanarak DTI'ları tanımlamaya çalışmışlardır. Bununla birlikte, mevcut yöntemlerin çoğu, ilaç ağlarını ve hedef ağları ayrı ayrı inşa eder ve daha sonra ilaç-protein çiftleri (DPP'ler) arasındaki ilişkileri hesaba katmadan, ilaçlar ve hedefler arasındaki bilinen ilişkilere dayanarak yeni DTI'ları tahmin eder. DPP'ler arasındaki ilişkileri DTI modellemesine dahil etmek için, DPP'lerin düğümler olduğu ve DPP'ler arasındaki ilişkilerin ağı kenarları olduğu birden fazla ilaç ve proteine dayalı bir DPP ağı oluşturdular. Daha sonra DTI tanımlaması için yeni bir öğrenme tabanlı model olan 'GCN)-DTI' önerdiler. Model ilk olarak her DPP'nin özelliklerini öğrenmek için bir GCN kullanır. İkincisi, özellik gösterimini bir girdi olarak kullanarak, son etiketi tahmin etmek için derin bir sinir ağı kullanır. Analizlerinin sonuçları, önerilen çerçevenin bazı son teknoloji yaklaşımlardan büyük bir farkla daha iyi performans gösterdiğini göstermektedir (T. Zhao vd., 2021).

Wu ve arkadaşları, BridgeDPI adlı yeni bir yöntem önerdiler. İlaçlar ve proteinler arasındaki boşluğu kapatmak ve öğrenilebilir bir ilaç-protein ilişkilendirme ağı oluşturmak için bir sanal düğüm sınıfı tasarladılar. Ağ, aşağı akış görevinden gelen denetlenen sinyallere (DPI tahmini) göre optimize edilir. Bu ilaç-protein ilişkilendirme ağında geçen bilgiler sayesinde, bir GSA, çeşitli ilaçlar ve proteinler arasındaki ağ düzeyindeki bilgileri yakalayabilir. Ağ düzeyindeki bilgileri ve öğrenme tabanlı yöntemi birleştirerek, BridgeDPI üç gerçek dünya DPI veri kümesinde önemli bir iyileşme sağlar. Ayrıca, vaka çalışması BridgeDPI'nin etkinliğini ve güvenilirliğini daha da doğrulamaktadır (Wu vd., 2022).

Sakai ve arkadaşları, araştırmalarında belirttiği gibi birçok terapötik ilaç, etki bölgesinde önemli afinite belirleyicileri içeren basit kimyasal yapılarla temsil edilebilen bileşiklerdir. Son zamanlarda, GCN modelleri, bu tür bileşiklerin aktivitesini sınıflandırmada mükemmel sonuçlar göstermiştir. Bu çalışmada, yalnızca bileşiklerin iki boyutlu yapısal bilgilerinden oluşturulan GCN modellerinin, ChEMBL veri tabanından 127 farklı hedefe karşı yüksek derecede aktivite öngörülebilirliği gösterdiğini bildirdiler. Bilgi entropisini bir metrik olarak kullanarak, yapısal çeşitliliğin tahmin performansı üzerinde daha az etkisi olduğunu da gösterdiler. Son olarak, modellerini kullanarak sanal taramanın, in vitro olarak pazarlanan bir ilacınla karşılaştırılabilir aktiviteye sahip yeni bir serotonin taşıyıcı inhibitörü tanımladığını ve davranışsal çalışmalarda antidepresan etkiler sergilediğini bildirdiler (Sakai vd., 2021).

Zhou ve arkadaşları, COVID-19 üzerinde terapötik etkisi olabilecek ilaçları taramak için LUNAR adlı yeni bir model önerdiler. Bu, manuel özellik mühendisliği gerektirmeyen yeni bir uçtan uca modeldir. Karmaşık heterojen ağları, sadece SARS-CoV-2 için ilaçları keşfetmekle kalmayıp aynı zamanda diğer virüslere de uygulanabilen ve ilaç geliştirmede iyi bir yardımcı olabilecek CNN'ler ve dikkat mekanizmaları aracılığıyla bütünleştirir. Bir GCN yapısı, yapılandırılmış bir veri modeline ilişkisel indüksiyon yanlılığının uygulanmasıyla karakterize edilir. LUNAR modelinin yüksek performans endeksi, modelin güçlü tahmin yeteneğini göstermektedir. Ayrıca, LUNAR tarafından öngörülen bazı ilaçlar, modelin etkinliğini daha da kanıtlamak için klinik çalışmalarda ortaya çıkmıştır. Bu nedenle, COVID-19 üzerinde etkisi olabilecek ilaçları eledebileceğine ve ilaç geliştirmeyi hızlandırabileceğine inandıklarını belirttiler (Zhou vd., 2021).

Wang ve arkadaşları, spesifik kanser hücrelerinin yaşayabilirliğini etkili bir şekilde inhibe edebilecek ilaç kombinasyonlarını tanımlamak için GSA ve dikkat mekanizmasına dayanan bir derin öğrenme modeli önerdiler. İlaç molekül yapısının ve gen ekspresyon profillerinin özellik gömmeleri, sinerjik ilaç kombinasyonlarını tanımlamak için çok katmanlı ileri besleme sinir ağına girdi olarak alınmıştır. DeepDDS'yi (İlaç-İlaç Sinerjisi tahmini için Derin Öğrenme) klasik makine öğrenme yöntemleri ve kıyaslama veri setindeki diğer DL tabanlı yöntemlerle karşılaştırdılar ve

bir kereye mahsus deneysel sonuçlar, DeepDDS'nin rekabetçi yöntemlerden daha iyi performans elde ettiğini gösterdi. Ayrıca, tanınmış ilaç şirketi AstraZeneca tarafından yayınlanan bağımsız bir test setinde DeepDDS, rekabetçi yöntemlerden %16'dan fazla öngörücü hassasiyetle üstün olduğunu belirttiler. Ayrıca, grafik dikkat ağıнын yorumlana bilirliliğini araştırdılar ve atomik özelliklerin korelasyon matrisinin ilaçların önemli kimyasal alt yapılarını ortaya çıkardığını buldular. DeepDDS'nin, daha fazla ıslak laboratuvar deneyi doğrulaması için sinerjik ilaç kombinasyonlarına öncelik veren etkili bir araç olduğuna inandıklarını belirttiler (J. Wang vd., 2022).

Liu ve arkadaşları, çalışmalarında, moleküler özellik tahmini için yeni bir derin öğrenme modeli olan MolGSA'yı tanıtmaktadır. MolGSA, kimyasal moleköl yerleştirmenin hesaplamalı öğrenimi için bir GSA uygulamaktadırlar. Büyük ölçüde etiketlenmiş deneysel verilere dayanan son teknoloji yaklaşımlarla karşılaştırıldığında, yöntemleri ön eğitim aşamasında manuel etiketler olmadan eşdeğer veya üstün tahmin performansı ve yalnızca birkaç etiketle veriler üzerinde mükemmel performans sağlamaktadır. Sonuçlar, MolGSA'nin kıt eğitim verilerine karşı dayanıklı olduğunu ve dolayısıyla birkaç adımlı güçlü bir öğrenme aracı olduğunu gösterdi. MolGSA, hem insan Janus kinazlarına hem de SARS-CoV-2 ana proteazına karşı, sırasıyla sitokin fırtınası COVID-19 semptomlarını hafifletmeyi ve viral replikasyonu baskılamayı amaçlayan ilaçlar için tercihli hedefler olan birkaç çok hedefli moleköl öngörmüşlerdir. Ayrıca SARS-CoV-2'nin neden olduğu hücre ölümünü potansiyel olarak inhibe eden molekülleri de öngördüler. MolGSA'nin en iyi tahminlerinden birkaçı, yöntemimizin potansiyel değerini gösteren mevcut deneysel ve klinik kanıtlarla desteklenmektedir (Liu vd., 2021).

Xu ve arkadaşları çalışmalarında, GCN ve DeepWalk'u birleştiren GDNet-DTI olan çoklu grafik gösterim yöntemlerine dayanan geliştirilmiş bir tahmin modeli önermektedir. İlk olarak, atomları düğüm olarak ve kimyasal bağları kenarlar olarak içeren bir moleküler harita, ilaçların SMILES dizisi kullanılarak üretilir ve daha sonra atomlar arasındaki karmaşık etkileşimleri daha iyi elde etmek için moleküler haritanın özelliklerini çıkarmak için GIN kullanılır. Hedef proteinler için, protein dizisi önce bir kelime vektörü ile temsil edilir ve daha sonra tek boyutlu evrişim, farklı özellik seviyelerini çıkarmak için özellikleri çıkarmak için kullanılır. Daha sonra, elde edilen

ilaç özelliklerine ve hedef özelliklerine dayanarak, ilaçların ve hedeflerin düğümler olarak temsil edildiği ve etkileşimlerin kenarlar olarak temsil edildiği bir DTI grafiği oluşturulur. Son olarak, GDNet-DTI, DTI-grafiginin düğüm komşuluk bilgilerini ve grafik topolojisi bilgilerini elde etmek için kullanılır. Diğer gelişmiş modellerle karşılaştırıldığında, sonuçlar GDNet-DTI'nın birden fazla grafik özelliğiyle birleştirildiğinde DTI'ları DrugBank ve dört kıyaslama veri kümesi ile daha doğru ve etkili bir şekilde tahmin edebileceğini göstermektedir. Ek olarak, önerilen yöntemin gerçek DTI'ları tahmin etme potansiyeline sahip olduğunu ve ilaç keşfinin geliştirilmesine katkıda bulunabileceğini gösteren COVID-19 verilerini içeren bir vaka çalışması sunulmaktadır (Xu vd., 2021).

Lai ve arkadaşları, derin üretici modellerdeki son çalışmalardan esinlenerek, ilaç moleküler yapı üretimi için kısaca MGRNN (Moleküler Grafik Tekrarlayan Sinir Ağları) olarak adlandırılan bir grafik tekrarlayan sinir ağı modeli önerdiler. MGRNN, hem yinelemeli moleküler üretim algoritmasının avantajlarını hem de eğitim stratejilerinin verimliliğini birleştirir. Ayrıca, MGRNN şunları gösterir: (i) eğitim için verimli hesaplama; (ii) veriler için yüksek model sağlamlığı ve (iii) değerlik kontrolü için kimyasal alan uzmanlığının kullanılmasına izin veren yinelemeli bir örnekleme süreci. Deneysel sonuçlar, MGRNN'nin kimyasal bilgisi olmadan bile kimyasal olarak geçerli moleküllerin %69'unu ve kimyasal kuralları olan %100 geçerli molekülleri üretebildiğini göstermektedir (Lai vd., 2021).

Nguyen ve arkadaşları çalışmalarında, problem için GCN'ye dayanan yeni bir yöntem olan GraphDRP'yi önerdiler. GraphDRP'de, ilaçlar atomlar arasındaki bağları doğrudan yakalayan moleküler grafiklerde temsil edilirken, hücre çizgileri genomik sapmaların ikili vektörleri olarak tasvir edildi. İlaçların ve hücre hatlarının temsili özellikleri, evrişim katmanları tarafından öğrenildi, daha sonra her ilaç-hücre hattı çiftini temsil etmek üzere birleştirildi. Son olarak, her bir ilaç-hücre hattı çiftinin yanıt değeri, tamamen bağlı bir sinir ağı tarafından tahmin edildi. İlaçların özelliklerini öğrenmek için GCN'lerin dört çeşidi kullanılmıştır. GraphDRP'nin tüm deneyler için tüm performans ölçümlerinde tCNNS'den daha iyi performans gösterdiğini buldular. Ayrıca, ortaya çıkan GraphDRP modellerinin belirginlik haritaları sayesinde, genomik

sapmaların yanıtla katkıını keşfettiler. Sonuç olarak ilaçların grafik olarak gösterilmesi, ilaç yanıtı tahmininin performansını artırabilir (T. Nguyen vd., 2022).



3. MATERYAL VE METOT

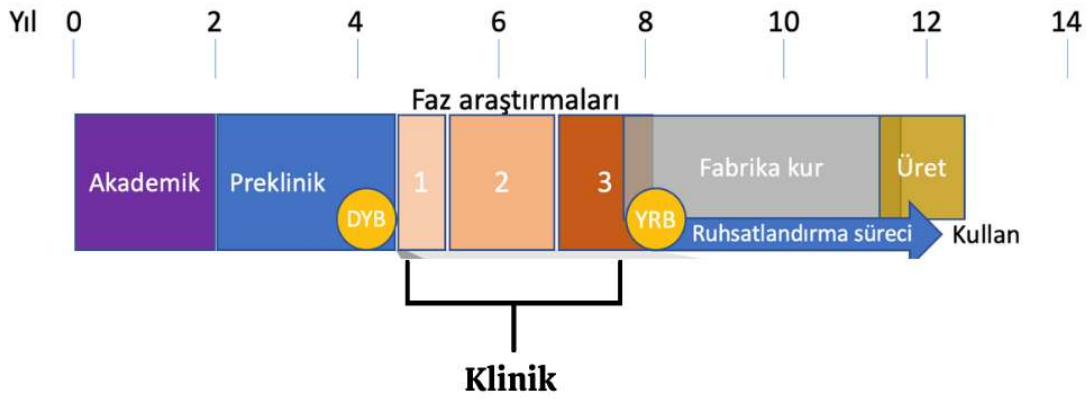
İlaç keşfi ve geliştirme aşamaları, insan sağlığı ve refahına önemli katkılarda bulunan en kritik translayonel bilim faaliyetleri arasındadır. Ancak, yeni bir ilacın geliştirilmesi süreci son derece karmaşık, maliyetli ve zaman alıcıdır. Bu bağlamda, maliyetlerin nasıl azaltılacağı ve yeni ilaçların keşfinin nasıl hızlandırılacağı soruları, endüstride acil ve çözülmesi gereken temel konular olarak öne çıkmaktadır. YZ'nin yeni deneysel teknolojilerle entegrasyonu, ilaç arayışını daha hızlı, daha düşük maliyetli ve daha etkili hale getirme potansiyeline sahiptir.

Bu tezin amacı, ilaç keşif sürecini hızlandırmak ve ilaç-ilaç benzerlikleri, ilaç-hedef protein ilişkileri ve ilaç-ilaç etkileşimlerini belirlemek için son dönemde ortaya çıkan YZ uygulamalarından biri olan GSA yöntemini kullanmaktır. Bu çalışma ile ilaç endüstrisine ve akademik literatüre, birincil hedefi insan sağlığını iyileştirmek olan önemli katkılar sağlanması hedeflenmektedir.

3.1 İlaç Keşfi

İlaç keşfi ve geliştirilmesi, son derece maliyetli, uzun süreli ve riskli bir süreç olarak bilinmektedir. Yeni bir ilacın piyasaya sürülmesine yönelik ortalama maliyetin 1,24 milyar dolar olduğu tahmin edilmektedir (DiMasi vd., 2016). Geleneksel ilaç geliştirme yöntemleri, Şekil 3.1'de de belirtildiği gibi, dört temel aşamadan oluşmaktadır: ilaç keşfi, klinik öncesi çalışmalar, klinik deneyler ve pazarlama sonrası güvenlik izleme. Bu süreç, hayvan deneyleri ve klinik çalışmalarda güvenlik, etkinlik ve kaliteye ilişkin düzenleyici gereklilikler nedeniyle yaklaşık 12-17 yıl sürmektedir (Paul vd., 2010).

Öte yandan, ilaç yeniden kullanımı stratejisi, geliştirme süresini önemli ölçüde kısaltma potansiyeline sahiptir. Araştırmacıların yeni bir tedavi amacıyla mevcut bir ilacı yeniden geliştirmesi yalnızca 3-12 yıl almaktadır (Şekil 3.1).



Şekil 3.1 Bir ilacı geliştirme süreci

Ayrıca, ilaç yeniden kullanımı maliyetleri, geleneksel ilaç geliştirme maliyetlerine kıyasla oldukça düşüktür. Bu yaklaşım, araştırma kurumlarına daha az finansal yatırımla ilaç geliştirme fırsatları sunmaktadır. Bu nedenle, ilaç şirketleri, ilaç geliştirme süreçlerini hızlandırmak, başarısızlık riskini azaltmak ve maliyetleri düşürmek amacıyla ilaç yeniden kullanımı etkili bir alternatif olarak benimsemiştir.

3.1.1 İlaç yeniden kullanımı

Son yıllarda, ilaç araştırma ve geliştirmeye yapılan yatırımın artmasına rağmen, onaylanmış ilaç sayısı hala sınırlı kalmaktadır. Bu durum, çeşitli faktörlerden dolayı ilaç yeniden kullanımına olan ilgiyi artırmıştır. İlaç yeniden kullanımının amacı, mevcut ilaçlar için yeni endikasyonları belirlemektir ve bu endikasyonlar aşağıdaki kategorilere ayrılabilir: (i) Başarılı bir şekilde pazarlanan ilaçlar, (ii) Klinik çalışmalarda spesifik endikasyonları ve güvenlik sorunları olmayan ilaçlar, (iii) Bütçe ve diğer nedenlerle kullanımı durdurulan ilaçlar. Geleneksel ilaç keşfi ile karşılaştırıldığında, ilaç yeniden kullanımı bileşiklerin Faz II aşamasına geçişine olanak tanıyarak pazara sunma süresini kısaltabilir ve yatırım maliyetlerini düşürebilir. Ayrıca, bir geliştirme projesinin başlangıcında mevcut olan klinik veriler, daha ileri geliştirme ile ilişkili riskleri büyük ölçüde azaltır (Ashburn & Thor, 2004).

İlaç yeniden kullanımının bir diğer önemli avantajı, birçok durumda klinik öncesi çalışmalarda yeterince güvenli olduğunun kanıtlanmış olmasıdır. Onaylanmış bir ilaç için sadece klinik öncesi bilgiler değil, aynı zamanda klinik profiller de mevcut olduğundan, ilaç keşfinin gelişimi hızlanır (Chong & Sullivan, 2007). İlacın yeniden

kullanımının temel amaçlarından biri, yeni ilaç-hastalık ilişkileri bulmaktır. Bu amaca ulaşmak için hesaplama yöntemleri, biyolojik deneysel yöntemler ve her ikisinin kombinasyonunu içeren çeşitli yöntemler geliştirilmiştir (J. Li vd., 2016). İlaçla ilgili birçok veri tabanının açık kaynaklı olması, hesaplamalı yaklaşımların gelişimini hızlandırmıştır. Aynı zamanda, biyolojik ve hesaplama yöntemlerini birleştiren karma yaklaşımlar, maliyetleri düşürürken ilacın yeniden kullanımının verimliliğini büyük ölçüde artırır.

İlaç yeniden kullanımı stratejisinde, aceleci kavram kanıtlarından kaçınılmalıdır. Bunun yerine, ayrıntılı ve kapsamlı bir plan önerilmeli ve önceki geliştirme döngüsünde elde edilen bilgiler, süreci basitleştirmek için kullanılmalıdır. Klinik öncesi araştırma yapılırken, ilacın farmakolojisi ve doz yoğunluğu dikkate alınmalıdır. Erken gelişim aşamasında bilimsel denetimler atlanmamalıdır (Nosengo, 2016).

Bununla birlikte, ilaç yeniden kullanımının bazı büyük zorluklarla karşı karşıyadır. Örneğin, ilk klinik çalışmalar mevcut düzenleyici gereklilikleri karşılamıyorsa, verileri tamamlamak için yeni Faz I denemeleri gerekebilir. Başarılı vakaların çoğu, ilacın orijinal endikasyonlarının retrospektif bir analizini takiben farmasötik farmakoloji veya klinik etkinlik anlayışından elde edilmektedir. Bu yeniden ilaç kullanımı, şirketlerin yatırım getirisini azaltacak daha zayıf fikri mülkiyet korumalarına sahiptir. İlacın yeniden kullanımına yönelik uygulamaların çoğu, pozitif klinik veya epidemiyolojik gözlemlere ya da in vivo etkinlik keşfine dayanmaktadır. Bilgi işlem yöntemlerinin ve veri patlamasının hızlı gelişimi ile, ilacın yeniden kullanımının sunduğu karmaşık avantajlar daha belirgin hale gelecektir. Şekil 3.2’de ilaç yeniden kullanımına ait genel yapıyı göstermektedir.



Şekil 3.2 İlaç yeniden kullanımı genel bakış

Son yıllarda hastalık ve ilaçla ilgili veri setlerinin hızlı büyümesi, hesaplamalı ilaç yeniden kullanımına, özellikle de yapıya dayalı ilaç yeniden kullanımına hızlı bir gelişme getirmiştir. ML ve ağ farmakolojisi gibi bilgiye dayalı haritalama gibi stratejiler, ilaçlar ve hedefler arasındaki iç ilişkiyi daha doğru bir şekilde analiz edebilir ve böylece daha anlamlı bilgiler elde edebilir. COVID-19 gibi ani küresel büyük halk sağlığı olayı göz önüne alındığında, hesaplamalı ilaç yeniden kullanımı en uygun stratejilerden biridir. İlacın yeniden kullanımında farklı yöntemlerin rolünü daha iyi anlamak ve gelecekteki halk sağlığı acil durumlarına yönelik metodolojik rehberlik sağlamak amacıyla, ilaç keşfi alanında ilacın yeniden kullanımının etkinliğini artırmak üzere YZ tabanlı GSA modellerinin uygulanabilirliğini ve yeni yaklaşımların geliştirilmesi için bu bölümde genel bilgiler açıklanmaktadır. Bölüm 3'te ise bilimsel çalışmalarla destekli farklı uygulama alanları sunulmaktadır.

3.2 İlaç keşfi sürecinde yapay zekâ uygulamaları

YZ teknolojileri, verilerden kavram ve ilişkiler çıkararak, veri modellerine bağımlı olmadan öğrenme kapasitesi ile insanların gerçekleştirebileceği görevleri araştırmaktadır. YZ tabanlı algoritmalar, moleküler özelliklerin tahmini gibi fiziksel yasaların tam olarak bilinmediği veya deneysel ilişkilerin çok karmaşık olduğu durumlar için özellikle uygundur. İlaç geliştirme de dahil olmak üzere birçok alanda YZ tabanlı yaklaşımlardan yararlanılmaktadır.

YZ uygulamaları, ML algoritmalarının kullanımı ile gerçekleştirilmektedir. İlaç keşfi sürecinde, Rastgele orman (H. Shi vd., 2019), Destek Vektör Makinesi (Song vd., 2019), K-En Yakın Komşu (Das & Mazumder, 2024), Naive Bayesian (Bai vd., 2018), Karar Ağaçları (Nilar, 2023) ve DL (Lavecchia, 2019) gibi çeşitli algoritmalar

uygulanmıştır. Son zamanlarda yapılan çalışmalar, ilaç keşfi alanında DL ve diğer ML algoritmalarının artan bir şekilde kullanıldığını göstermektedir.

3.2.1 Grafik sinir ağıları (GSA)

Grafikler, bir kümeyi modelleyen bir tür veri yapısıdır: nesneler (düğümler) ve bunların ilişkileri (kenarlar) (Huber vd., 2007). Son zamanlarda, makine öğrenimi ile grafiklerin analizine yönelik araştırmalar nedeniyle giderek daha fazla ilgi görmektedir (Dara et al., 2022). Grafiklerin büyük ifade gücü, çeşitli sistemlerde çok sayıda yapının temsil edilebilmesiyle ortaya çıkar. Sosyal bilimlerde (örneğin, sosyal ağlar), doğa bilimlerinde (fiziksel sistemler ve protein-protein etkileşim ağları gibi), bilgi grafikleri ve birçok diğer araştırma alanında grafikler kullanılabilir. Benzersiz bir Öklid dışı veri olarak makine öğrenimi için yapı, grafik analizi aşağıdakilere odaklanır: düğüm sınıflandırması, bağlantı tahmini ve kümeleme (Veličković, 2023).

Şu anda en çok uygulananlardan biri olan GSA adlı derin öğrenme algoritmasının bir alt türüne odaklanmaktadır. Yeni olmasına rağmen, GSA kullanan DL algoritmalarının kullanımı, bazı araştırmacılar tarafından yüksek doğruluk oranları nedeniyle son teknoloji olarak kabul edilmektedir. Geçtiğimiz birkaç yıl içinde, GSA grafiklerle modellenebilen herhangi bir problem için güçlü ve pratik araçlar haline geldi. GSA, Şekil 3.3'de görüldüğü gibi araştırma literatürü genişlemekte ve birçok uygulama sahası mevcuttur.



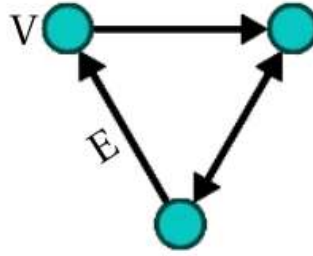
Şekil 3.3 GSA uygulamalar alanlarının gösterimi

Son yıllarda, derin öğrenme uygulamalarında önemli bir artış gözlemlenmiştir. Bu uygulamalar arasında kantitatif yapı-aktivite ilişkileri (QSAR), sanal tarama, ilacın yeniden kullanımı, in silico çalışmaları, farmakokinetik özelliklerin (emilim, dağılım, metabolizma ve atılım- ADME) ve toksisitenin tahmini yer almaktadır (Shah vd., 2024). Biyolojik ve kimyasal verilerin karmaşıklığı, belirsizliği, çeşitliliği ve yüksek boyutluluğu gibi özellikler, derin öğrenmenin sunduğu avantajlarla daha etkili bir şekilde analiz edilebilmektedir (Choudhury vd., 2022). DL, sinir ağlarının karmaşıklığı ve mimarilerinin belirli problemlere adapte edilebilme esnekliği sayesinde, büyük kimyasal kütüphanelerin işlenmesinde tahmine dayalı hesaplama modelleri sağlamada etkili bir araçtır (Budak vd., 2023). Bu sebeple, araştırmacılar yeni kemoterapötik ajanlar bulmak için bu metodolojiyi sıkça kullanmaktadır. Ayrıca, DL, protein-ligand bağlanma afinitelerini tahmin etmede de yüksek performans göstermektedir (M. Sun vd., 2020).

DL, birçok alanda olduğu gibi ilaç keşfi sürecinde de devrim niteliğinde değişiklikler yaratmaktadır. İlaç keşfi, zaman alıcı ve maliyetli bir süreç olup, DL bu süreci hızlandırmak ve maliyetleri düşürmek için kullanılabilir (Ma & Lei, 2023). İlaç keşfi süreci üç ana başlıkta incelenebilir: ilaç özellikleri tahmini, ilaç keşfi ve ilaç-hedef etkileşimi tahmini.

İlaç geliştirme, disiplinler arası iş birliği gerektiren ve birçok evrede birbirine bağımlı çalışmaların yürütüldüğü, sonuca ulaşması yıllar alan bir süreçtir. COVID-19 pandemisi, bu süreci hızlandırmış ve araştırmaları koronavirüs tedavisine yönlendirmiştir. YZ teknolojileri, ilaç keşfi süreçlerini daha hızlı ve verimli hale getirerek kritik bir rol üstlenmiştir. Bu bağlamda, YZ destekli ilaç keşfi girişimleri büyük önem kazanmıştır. GSA, ilaç keşfi için en etkili yöntemlerden biri olarak öne çıkmaktadır.

DL puanlama algoritmaları, komşu ligandları ve reseptör atomlarını mekânsal ve kimyasal olarak kodlayan GSA kullanmaktadır (Alves vd., 2022). Şekil 3.4, temel bir grafik yapısını göstermektedir.



Şekil 3.4 Temel grafik yapısı

GSA'nın temeli olan grafik teorisi, Euler'in ufuk açıcı çalışmalarıyla matematik müfredatının bir parçası olmuştur ve modern çağda bilgisayar bilimine dahil edilmiştir, ancak Euler'in çalışmasından önce grafik teorisinin temellerine dair kanıtlar vardır (Maziarz, 1979). Denklem (3.1)'de gösterildiği gibi basit bir grafik,

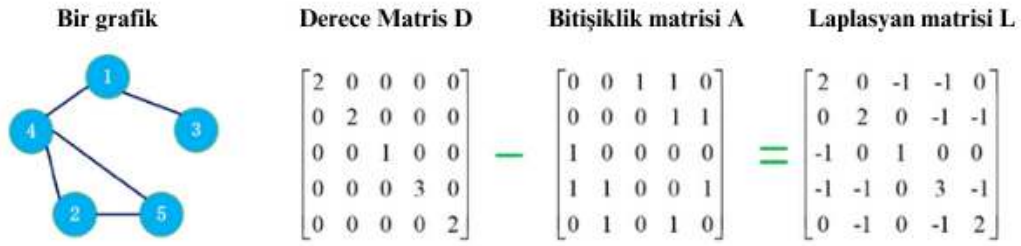
$$G = G(V, E) \quad (3.1)$$

olarak temsil edilebilir; burada V ve E , sırasıyla düğüm ve kenar kümeleridir. $v_i \in V$ bir düğüm olsun ve $e_{ij} = (v_i, v_j) \in E$, v_i ile v_j arasındaki bir kenarı gösterebilir. Daha sonra, bir v düğümünün komşuluğu $N(v) = \{u \in V \mid (v, u) \in E\}$ şeklinde tanımlanabilir. Genellikle bir grafik, N 'nin düğüm sayısı olduğu, yani $N = |V|$ olduğu bir komşuluk matrisi $A \in R^{N \times N}$ ile tanımlanabilir. Özellikle, $A_{ij} = 1$ ise $\{v_i, v_j\} \in E$ ve $i \neq j$; $A_{ij} = 0$, diğer durumlarda. Yönsüz grafikte A_{ij} , v_i ve v_j düğümleri

arasındaki bir kenar bağlantısını belirtirken, yönlü bir grafikte A_{ij} , v_i 'den v_j 'ye işaret eden bir kenarı temsil eder. Pratik uygulamalarda, bir grafiğin düğüm özellikleri (öznitelikler olarak da adlandırılır) $X \in R^{N \times c}$ olabilir, burada c bir düğüm özellik vektörünün boyutudur. Bir derece matrisi $D \in R^{N \times N}$, olabilen köşegen bir matristir $D_{ij} = \sum_{j=1}^N A_{ij}$ olarak elde edilir. Bir grafik, Denklem (3.2)'deki şekilde tanımlanan Laplasyan matrisi L ile de temsil edilebilir:

$$L = D - A \quad (3.2)$$

Denklem (3.2)'deki matematiksel ifadenin yani yukarıdaki üç matris arasındaki ilişkinin bir gösterimi Şekil 3.5'te gösterilmektedir.



Şekil 3.5 Laplasyan matrisi

Grafik yapılı veriler için karmaşık kodlayıcılar geliştirmedeki temel zorluk, alışlagelmiş derin öğrenme araç kutumuzun geçerli olmamasıdır. Örneğin, CNN'ler yalnızca ızgara yapılı girdiler (ör. görüntüler) üzerinde iyi tanımlanırken, RNN'ler yalnızca diziler (ör. metin) üzerinde iyi tanımlanır. Genel grafikler üzerinde derin bir sinir ağı tanımlamak için yeni bir tür derin öğrenme mimarisi tanımlamamız gerekir. Temel GSA'ların iki temel yaklaşım vardır: Ana fikir, yerel komşulara dayalı olarak düğüm yerleştirmeleri oluşturmak ve sezgisel, düğümler, sinir ağlarını kullanarak komşularından bilgi toplamaktır (Reiser vd., 2022). Şekil 3.6 burada bahsedilen düğüm yerleştirmelerini ve komşulardan gelen bilgi toplama fonksiyonunu göstermektedir.

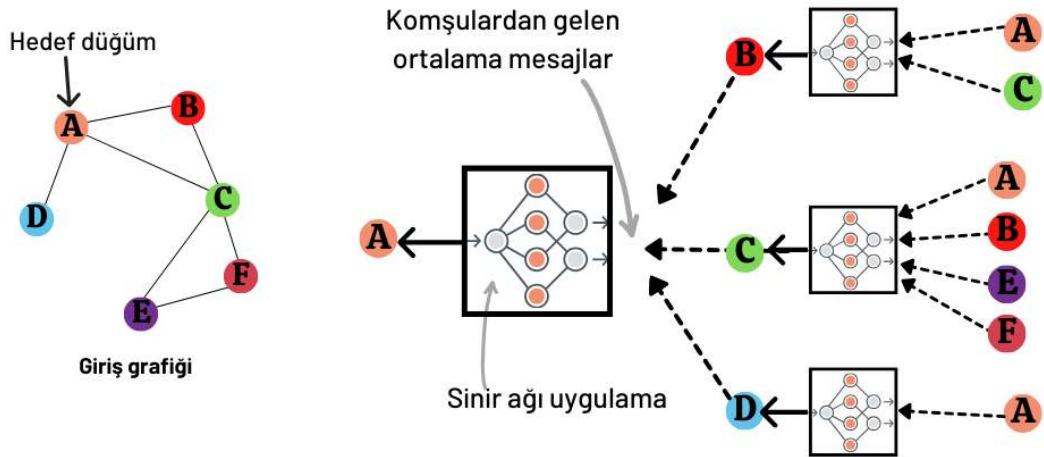
Sinirsel mesaj iletimi:

Temel GSA modeli çeşitli şekillerde motive edilebilir. Aynı temel GSA modeli, konvolüsyonların Öklid olmayan verilere genelleştirilmesi, inanç yayılımının

farklılaştırılabilir bir çeşidi ve klasik grafik izomorfizm testlerine benzetilerek türetilmiştir (Wein vd., 2022). Motivasyon ne olursa olsun, bir GSA'nın tanımlayıcı özelliği, vektör mesajlarının düğümler arasında değiş tokuş edildiği ve sinir ağıları kullanılarak güncellendiği bir tür sinirsel mesaj geçişi kullanmasıdır (Reiser vd., 2022).

Bu bölümünün geri kalanında, bu nöral mesaj geçirme çerçevesinin temellerini detaylandıracağız. Mesaj geçirme çerçevesinin kendisine odaklanacağız ve GSA modellerinin eğitimi ve optimizasyonu ile ilgili tartışmaları bölüm 4'te açıklayacağız. Bu bölümün büyük bir kısmı, bir dizi düğüm özelliği $X \in Rd \times |V|$ ile birlikte $G = (V, E)$ girdi grafiğini nasıl alabileceğimizi ve bu bilgileri $z_u, \forall u \in V$ düğüm katıştırmalarını oluşturmak için nasıl kullanabileceğimizi açıklayacaktır. Ancak, GSA çerçevesinin alt grafikler ve tüm grafikler için katıştırmalar oluşturmak için nasıl kullanılabileceğini de tartışacağız.

Bir GSA'daki her mesaj geçirme iterasyonu sırasında her $u \in v$ düğümüne yanıt veren gizli bir gömme $h_u^{(k)}$, u 'nun grafik komşuluğu $N(u)$ 'dan toplanan bilgilere göre güncellenir.



Şekil 3.6 Tek bir düğümün yerel komşuluğundan gelen mesajları nasıl topladığına genel bakış

Şekil 3.6'ya göre, model, A'nın yerel grafik komşularından (yani B, C ve D) gelen mesajları toplar ve bu komşulardan gelen mesajlar da kendi komşuluklarından toplanan bilgilere dayanır ve bu böyle devam eder. Bu görselleştirme, mesaj iletme

modelinin iki katmanlı bir versiyonunu göstermektedir. GSA'nın hesaplama grafiğinin, hedef düğümün etrafındaki komşuluğu açarak bir ağaç yapısı oluşturduğuna dikkat edin. Denklem (3.3) ve Denklem (3.4)'te mesaj iletiminin matematiksel ifadeleri yer almaktadır.

$$h_u^{(k+1)} = \text{Güncelleme}^{(k)} \left(h_u^{(k)}, \text{Toplam}^{(k)} \left(\{ h_v^{(k)}, \forall v \in N(u) \} \right) \right) \quad (3.3)$$

$$= \text{Güncelleme}^{(k)} \left(h_u^{(k)}, m_{N(u)}^{(k)} \right) \quad (3.4)$$

Burada “Güncelleme” ve “Topla” keyfi farklılaştırılabilir fonksiyonlardır (yani, nöral ağırlar) ve $m_{N(u)}$ u’nun grafik komşuluğundan toplanan “mesaj”dır. Mesaj aktarımının farklı iterasyonlarında gömülmeleri ve fonksiyonları ayırt etmek için üst simgeler kullanıyoruz.

- Genel Bakış: Bu süreç, bir düğümün (örneğin A) yerel grafik komşularından (B, C ve D) gelen mesajları toplamasını içerir. Bu mesajlar, komşularının çevrelerinden gelen bilgilere dayanmaktadır.
- Süreç: Her iterasyonda (k), “Topla” işlevi, düğüm u’nun komşularının $N(u)$ gömme özelliklerinden bir mesaj $m_{N(u)}^{(k)}$ oluşturur. Ardından “Güncelleme” işlevi bu mesajı düğüm u’nun önceki gömme özelliği $h_u^{(k-1)}$ ile birleştirerek güncellenmiş gömme özelliği $h_u^{(k)}$ üretir.
- Başlangıç Gömmeleri: Tüm düğümler için başlangıçta gömme özellikleri düğümlerin girdi özelliklerine ayarlanır: ($h_u^{(0)} = x_u, \forall u \in V$).
- Nihai Çıktı: (K) iterasyon sonunda, son katmanın çıktısı her düğüm için gömme özelliklerini tanımlar. Denklem (3.5), gömme özelliklerinin matematiksel ifadesini göstermektedir.

$$z_u = h_u^{(K)}, \forall u \in V \quad (3.5)$$

Bu formülasyon, GSA’ların nasıl çalıştığını anlamak için oldukça önemlidir.

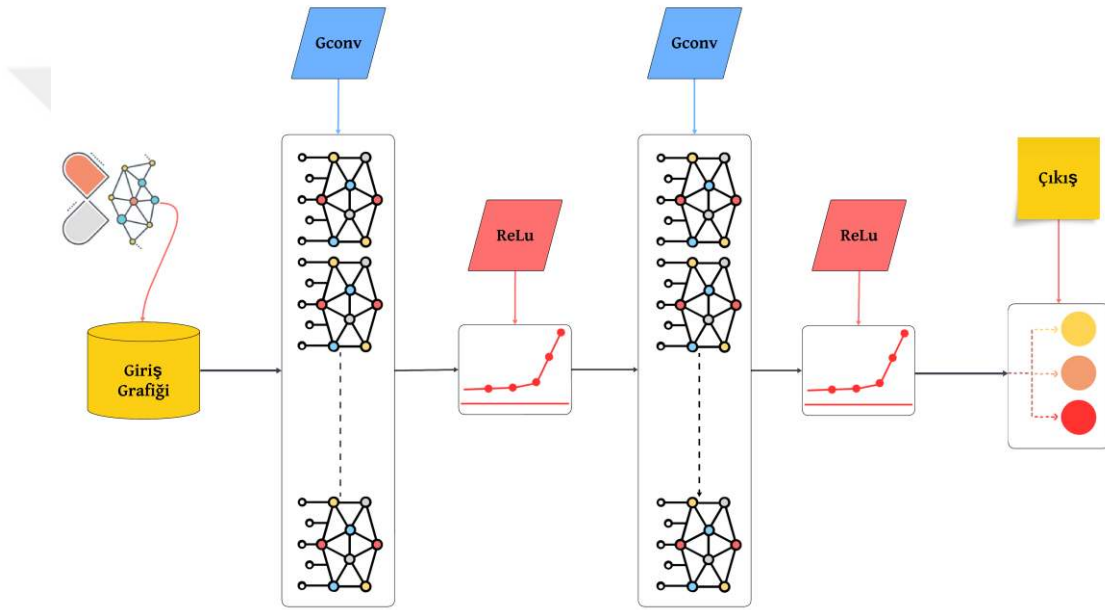
Temel GSA çerçevesine ve nasıl somut bir şekilde uygulayabileceğimize daha yakından bakalım.

1. Soyut GSA çerçevesi:

GSA'ları, "Güncelleme" ve "Topla" işlevlerini kullanarak bir dizi mesaj iletimi olarak tartıştık (Denklem (3.3))'te tanımlandığı gibi). Ancak bu soyut çerçeveyi pratik hale getirmek için bu işlemlere özgül anlamlandırmalar yapmamız gerekiyor.

2. Temel GSA çerçevesi ve model parametreleri

En temel GSA çerçevesi, (Merkwirth & Lengauer, 2005) ile (Scarselli vd., 2009) tarafından önerilen orijinal GSA modellerinin basitleştirilmesidir. Şekil 3.7'de bir GSA'nın model yapısı yer almaktadır.



Şekil 3.7 Bir GSA'nın genel model yapısı

Temel GSA mesaj iletimini denklem (3.6)'daki gibi tanımlayalım:

$$h_u^{(k)} = \sigma \left(w_{kendü}^{(k)} h_u^{(k-1)} + w_{komşu}^{(k)} \sum_{v \in N(u)} h_v^{(k-1)} + b^{(k)} \right) \quad (3.6)$$

Burada:

- $(h_u^{(k)})$, (k) iterasyon sonrası düğüm (u) 'nin gömme vektörünü temsil eder.
- $\sum_{v \in N(u)} h_v^{(k-1)}$, düğüm (u) 'nin komşuluklarını belirtir.

- $(w_{kendi}^{(k)})$ ve $(w_{komşu}^{(k)})$ sırasıyla kendi bilgisi ve komşuluk birleştirmesi için ağırlık matrisleridir.
- $(b^{(k)})$, bir bias terimidir. Genellikle basitlik için atlanır. Ancak güçlü performans elde etmek için bias terimini dahil etmek önemlidir.
- $(\sigma$ (sigma)), bir aktivasyon fonksiyonudur (örneğin sigmoid veya ReLU). GSA modeline doğrusallık katmaktadır.

2. Somut anlamlandırmalar:

GSA'ları uygulamak için gömme birleştirme ve güncelleme işlevlerini belirtmemiz gerekiyor. “Topla” için yaygın seçenekler toplama, ortalama veya dikkat tabanlı birleştirmedir. “Güncelleme” işlevleri basit besleme ileri sinir ağları veya daha karmaşık mekanizmalar olabilir. Uygulamada, GSA'lar bu temel çerçeveyi atlayan, dikkat mekanizmaları, artıklar ve diğer karmaşıklıkları içeren modellerle geliştirilir. GSA'lar ve temel işlemleri olan “Güncelleme” ve “Topla” fonksiyonları hakkında konuşuyorsunuz. Bu fonksiyonları GSA'lar bağlamında açıklayalım:

“Topla” fonksiyonu: Bu fonksiyon, komşu düğümlerin özelliklerini toplamak için kullanılır. Notasyonumuzda Denklem (3.7)'deki gibi temsil edilir:

$$m_{N(u)} = \sum_{v \in N(u)} h_v \quad (3.7)$$

Burada, $m_{N(u)}$, düğüm u 'nın komşularının toplanmış özelliğidir, $N(u)$ düğüm u 'nın komşularını temsil eder ve h_v düğüm v 'nin özelliğidir.

“Güncelleme” fonksiyonu: Bu fonksiyon, bir düğümün özelliğini, kendi özelliği ve komşularının toplanmış özelliği temelinde güncellemek için kullanılır. Notasyonumuzda Denklem (3.8)'teki gibi temsil edilir:

$$Güncelleme(h_u, m_{N(u)}) = \sigma(W_{kendi}h_u + W_{komşu}m_{N(u)}) \quad (3.8)$$

Burada, h_u düğüm u 'nun özelliğidir, $m_{N(u)}$ düğüm u 'nun komşularının toplanmış özelliğidir ve σ (*sigma*) bir aktivasyon fonksiyonudur (ReLU, sigmoid vb.).

Bu iki fonksiyon, GSA'larda mesaj geçiş tekniğinin temelini oluşturur, burada bilgi bir düğümden komşularına geçirilir ve düğüm özellikleri güncellenir. Bu işlem genellikle birkaç yineleme veya “katman” için tekrarlanır. Son düğüm temsilleri, grafiğin yapısal bilgisini yakalar.

3. Veri setleri ve metodoloji

Veri setleri:

Bu tez süreci boyunca çalışmalarda kullanılan veri setleri, DrugBank (Wishart, 2006), PubChem (Ihlenfeldt, 2018) ve diğer halka açık biyoinformatik veri tabanlarından elde edilmiştir. Bu veri setlerinin detayları, içerdikleri moleküller, proteinler ve etkileşimler hakkında bilgi bu bölümde sunulmaktadır.

o Veri ön işleme:

Veri setleri, eksik değerlerin doldurulması, normalizasyon ve veri artırma gibi adımlarla ön işleme tabi tutulmuştur. Moleküler yapılar, grafik formatında temsil edilmek üzere dönüştürülmüştür.

GSA modelleri, ilaç keşfi ve moleküler özellik tahmini gibi çeşitli kimya ve biyoinformatik uygulamalarında güçlü araçlardır. Bu uygulamalarda, moleküler yapıları temsil etmek için sıklıkla SMILES (Basitleştirilmiş Moleküler Giriş Satır Giriş Sistemi) dizileri kullanılır. SMILES veya benzeri veri girdileri kullanılarak GSA modellerinde yapılan veri ön işleme adımlarının genel bir özeti aşağıda belirtilmiştir.

1. Molekül temsiliinin dönüştürülmesi

- **SMILES'tan Grafik Temsiline Dönüşüm:** İlk adım, SMILES dizilerini grafik temsillere dönüştürmektir. Bu bağlamda, moleküldeki her atom bir düğüm, atomlar arasındaki her bağ ise grafikte bir kenar olarak kabul edilir.
- **Kanonikleşme:** Her molekülün benzersiz bir SMILES temsiliinin olması sağlanır, bu da yinelenen girişlerin ortadan kaldırılmasına yardımcı olur.

2. Dügüm ve kenar özelliklerinin çıkarılması

Dügüm özellikleri: Moleküldeki her atom (dügüm) için aşağıdaki gibi özellikler çıkarılır:

- Atom tipi (örneğin, karbon, azot, oksijen)
- Atom derecesi (bağ sayısı)
- Hibridizasyon durumu (örneğin, sp, sp², sp³)
- Formal yük
- Aromatiklik

Kenar özellikleri: Atomlar arasındaki her bağ (kenar) için aşağıdaki gibi özellikler çıkarılır:

- Bağ tipi (tek, çift, üçlü, aromatik)
- Bağ yönlülüğü (stereokimya için)

3. Grafik yapısının oluşturulması

- **Komşuluk matrisi:** Dügümler (atomlar) arasındaki bağlantıları temsil eden bir komşuluk matrisi oluşturulur. Bu matris, hangi düğümlerin kenarlarla bağlı olduğunu gösterir.
- **Özellik matrisleri:** Dügümler ve kenarlar için çıkarılan özellikleri kullanarak özellik matrisleri oluşturulur. Bu matrisler GSA'ya giriş olarak kullanılır.

4. Normalizasyon ve ölçekleme

- **Özellik Normalizasyonu:** Özelliklerin benzer bir ölçeğe getirilmesi sağlanır. Bu, GSA modelinin performansını ve stabilitesini artırabilir.
- **Veri Artırma:** Bazen, ek eğitim örnekleri oluşturmak için veri artırma teknikleri kullanılır. Moleküler veriler için bu, molekülün farklı konformasyonlarını oluşturmayı içerebilir.

5. Grafik doldurma ve toplu işleme

- **Doldurma:** Bir toplu iş içindeki farklı boyutlardaki molekülleri işlemek için doldurma gerekebilir. Bu, tüm grafiklerin aynı boyuta sahip olmasını sağlamak için sahte düğümler ve kenarlar eklemeyi içerir.

- **Toplu işleme:** Verimli işlem için birden fazla grafik temsili bir toplu işte birleştirilir. Grafik verileri için çeşitli toplu işleme teknikleri kullanılarak toplama yapılabilir.

6. Kodlama ve vektörleştirme

- **Grafik gömme:** Grafik yapısını ve özelliklerini sürekli bir vektör uzayına kodlama işlemi gerçekleştirilir. Bu adım, GSA'nın moleküler grafiklerden anlamlı desenler öğrenmesini sağlar.
- **Tek-sıcak kodlama (One-hot encoding):** Kategorik özellikler (örneğin, atom tipleri) için, bu özellikleri GSA girişi için uygun bir sayısal formata dönüştürmek amacıyla tek-sıcak kodlama kullanılır.

Araçlar ve Kütüphaneler

Moleküler verilerin GSA modelleri için ön işlenmesini kolaylaştıran birkaç araç ve kütüphane bulunmaktadır:

- **RDKit:** SMILES'tan moleküler grafiklere dönüştürme, moleküler tanımlayıcıları hesaplama ve daha fazlasını sağlayan açık kaynaklı bir kimya bilgi işlem kiti.
- **DGL (Deep Graph Library):** Grafik verileri ile çalışmak için araçlar sağlayan, GSA'ları oluşturma ve eğitme imkânı sunan bir kütüphane.
- **PyTorch Geometric:** Grafik yapıdaki verilerle çalışmak için araçlar sağlayan ve grafik verilerinin işlenmesi için kullanışlı araçlar sunan, PyTorch üzerine inşa edilmiş bir kütüphane.

SMILES kullanarak GSA modelleri için veri ön işleme, SMILES dizilerini grafik temsillere dönüştürmeyi, düğüm ve kenar özelliklerini çıkarmayı, özellikleri normalleştirmeyi, grafik yapıları oluşturmaya ve verileri model eğitimi için hazırlamayı içerir. Bu ön işleme adımları, GSA'nın moleküler verilerden etkili bir şekilde öğrenmesini ve doğru tahminler yapmasını sağlamak için kritiktir.

Metodoloji:

1. GSA Modelleri:

Bu çalışmada, GCN, GAT ve GIN gibi çeşitli GSA modelleri kullanılmıştır. Her bir modelin mimarisi, katman sayısı, düğüm ve kenar özellikleri, aktivasyon fonksiyonları ve diğer hiper parametreler detaylandırılmıştır.

2. Eğitim Süreçleri:

Modeller, eğitim ve doğrulama veri setlerine bölünerek eğitilmiştir. Eğitim sırasında kullanılan kayıp fonksiyonları, optimizasyon algoritmaları ve erken durdurma kriterleri belirtilmiştir. Ayrıca, modellerin performansı, doğruluk, hassasiyet, özgüllük ve F1 skoru gibi metriklerle değerlendirilmiştir.

4. BULGULAR VE TARTIŞMA

Bu tez çalışması Ebola virüsü, COVID-19 gibi salgın hastalıkların tedavisi için potansiyel ilaçların keşfinde moleküler yapılarının analiz edilmesi üzerine odaklanmıştır. GSA'lar, moleküler yapıları ve etkileşimleri modellemede yüksek doğruluk oranlarına ulaşmıştır. Bu, ilaç keşfi alanında yeni ve etkili bir yaklaşım sunmaktadır. Elde edilen sonuçlar, pandemi senaryolarında hızlı ve etkili çözümler sunma potansiyeline sahiptir. Özellikle acil durumlarda mevcut ilaçların yeniden kullanımı stratejisi ile hızlı bir şekilde tedavi seçenekleri belirlenebilir. Bulgular, GSA'ların ilaç keşfi süreçlerinde daha geniş kapsamlı araştırmalarda kullanılabileceğini göstermektedir.

Bu bölümde uluslararası düzeydeki yayınlarla, moleküler benzerlik ve etkileşim analizlerine odaklanan çeşitli araştırmalar yapılmıştır. Bu araştırmalarda, atom çifti ve grafik ağları gibi yöntemlerin kullanıldığı, özellikle favipiravir ve molnupiravir gibi ilaçların kombinasyonunun ebola virüsü tedavisindeki etkinliğini inceleyen çalışmalar yapılmıştır. Ayrıca, ulusal düzeydeki yayınlarda ve projelerde, ilaç çözünürlüğü tahmini ve ilaç-protein etkileşimlerinin analizi gibi konulara odaklanan araştırmalar ele alınmıştır. Bu çalışmanın amacı, mevcut literatürdeki bu bulguları derinlemesine değerlendirmek ve GSA tabanlı moleküler düzeyde ilaç tasarımı ve keşfi alanındaki gelecek çalışmalara ışık tutmaktır. Bu bağlamda, bu bölümde öncelikle uluslararası ve ulusal düzeydeki kabul alan çalışmalarımızdan elde edilen bulgular özetlenecek ve ardından bu bulguların önemi ve gelecek araştırmalar için potansiyel etkileri tartışılacaktır.

4.1 Bölüm 1: Salgın Hastalıklar

4.1.1 Ebola virüsünün tedavisi için ilaçların grafik ağı ve benzerlik algoritması yöntemleriyle moleküler analizi ve bir ilaç kombinasyon önerisi

Ebola virüsü (EBOV), ilk olarak Demokratik Kongo Cumhuriyeti'nde ortaya çıkmıştır ve Filoviridae ailesinin bir üyesidir. EBOV, 1976'dan bu yana periyodik salgınlar nedeniyle binlerce ölümden sorumludur. 2018-2020 Ebola virüsü salgını, yüksek ölüm oranları nedeniyle Dünya Sağlık Örgütü (WHO) tarafından 3. seviye acil durum olarak

sınıflandırılmıştır (WHO, 2022). Virüsle ilişkili diğer hastalıklar 1976'dan beri bilinmesine rağmen, bu virüsün 2014-2016'daki en son salgınının tehlike boyutu ve küresel etkisi fark edildiğinde, bunları tedavi etmek için etkili aşı ve ilaçlara olan ihtiyaç ortaya çıkmıştır. Bugüne kadar, Ebola virüsüne karşı ticari bir aşı (Ebanga) ile birlikte önerilen ve bazıları klinik deneylerde olan birkaç ilaç bulunmaktadır. İlaç tasarımının birincil amacı, arzu edilen farmakolojik özelliklere sahip yeni bileşikler keşfetmektir. Kimyasal bir iskele, bir grup bireysel molekülü bir alt yapı olarak karakterize eden ortak bir çekirdek yapı olarak kimyada uzun bir geçmişe sahiptir.

Ancak FDA onaylı tek bir ilacın geliştirilmesi ortalama on yıl sürmekte ve 2 ila 3 milyar dolar arasında bir maliyete mal olmaktadır (Wouters vd., 2020). Bir ilacı yeniden kullanmak, yeni ilaç bulmak için kullanılan bir başka tekniktir. Klinik olarak test edilmiş bir ilacın yeni bir durumu etkili bir şekilde tedavi edip edemeyeceğini belirlemek için basitçe test ve analiz gerektirir. Güvenlik riski zaten belirlenmiştir. İlaç keşfi, araştırma açısından çok caziptir çünkü yeni ortaya çıkan hastalıkların tedavisi için eski ilaçların yeniden kullanılması, özellikle yeni ortaya çıkan bir viral hastalık söz konusu olduğunda, de novo tasarımdan daha avantajlıdır. İlk olarak, yeniden kullanılan ilaç adayları zaten klinik gelişimin birkaç aşamasından geçmiş olacağından, geliştirme riski önemli ölçüde daha düşüktür. Ayrıca, güvenli ve köklü farmakolojik profillere sahip olacağı için daha düşük geliştirme maliyetlerine ve daha kısa geliştirme sürelerine sahip olacaktır. Bazı ilaç adayları için güvenlik değerlendirmesi ve geliştirme formülasyonu zaten tamamlanmışsa, klinik öncesi denemeleri atlamak ve doğrudan faz II klinik denemelere girmek mümkün olabilir. İkinci olarak, ilaç tarama kütüphanelerinin yanı sıra akademik ve küçük laboratuvarların mevcudiyeti ilaç keşfine yardımcı olabilir. Bu, bir biyofarmasötiğe uygun yeniden kullanım adayları bulmak için farmakopeyi gözden geçirme süresini kısaltır. Tedavi seçeneklerinin son derece sınırlı olduğu mevcut salgın ortamında, daha önce FDA onaylı ilaçların yeniden kullanımı, ilaçların hızlı bir şekilde yaygınlaştırılması için bir fırsat sunmaktadır.

Ebola ve diğer küresel salgın krizlerinin yanı sıra gelecekteki patojenik salgınlarının etkilerini incelemek ve azaltmak için güvenli ilaçlar veya ilaç adayları bulmak, zamanında ve etkili terapötik yaklaşımlar geliştirmek için kritik öneme sahiptir. Ebola

virüsü replikasyonunu ve virölansını engelleme potansiyeline sahip olanları belirlemek için bu ilaçlar arasındaki kimyasal yapı benzerliklerini veya birbirine bağlı altyapı moleküllerini değerlendirmek ve anlamak kritik önem taşımaktadır. İn vitro veya in vivo taramanın zor veya imkânsız olduğu durumlarda, hesaplamalı yaklaşımlar potansiyel ilaç yeniden kullanım fırsatlarını belirlemek için cazip bir yöntemdir.

Gelecekteki araştırmalar, bu ilaçların moleküler yapılarının ve yapısal benzerliklerinin karşılaştırılmasından fayda sağlayacaktır. Moleküler benzerlik kavramı ilaç keşfi ve geliştirilmesinde kritik öneme sahiptir. Yapısal olarak benzer moleküllerin sıklıkla benzer özelliklere sahip olduğu önermesine dayanır. Küçük moleküllerin benzerliğinin değerlendirilmesi, çeşitli ilaçların keşfi ve geliştirilmesinde önemlidir. Basitlikleri, doğrulukları ve verimlilikleri nedeniyle, iki boyutlu (2B) benzerlik yaklaşımları özellikle popülerdir. Moleküler şekli tanımlamak ve moleküller arasındaki yapısal benzerliği belirlemek için birçok yöntem geliştirilmiştir (Xue vd., 2018). Bu yöntemler arasında en yaygın olarak kullanılan atomik mesafe tabanlı yöntemlerdir. İki molekülün yapısal olarak benzer durumları genellikle benzer fiziksel ve biyolojik özellikleri paylaşabilecekleri fikrine dayanır.

Moleküler benzerlik analizi, yapısal temsillere ve iki yapısal temsil arasındaki benzerliğin nicel ölçümlerine dayanmaktadır. Bu iki yapısal temsil arasındaki benzerlik derecesini belirlemek için çok sayıda yöntem vardır. Tanimoto katsayısı, zar indeksi, kosinüs katsayısı, Öklid mesafesi vb. bazı örneklerdir. Moleküler benzerlik yaklaşımları, yapısal temsillerine göre iki türe ayrılır: 2B ve 3B benzerlik yöntemleri. 2B benzerlik yöntemleri, yalnızca 2B yapısal bilgilere dayanan en popüler, hızlı ve verimli yöntemler arasındadır. Bu yöntemler arasında parmak izi benzerlik taraması, altyapı taraması ve 2B tanımlayıcılar bulunmaktadır. Moleküler parmak izleri, moleküllerin bit dizileri olarak temsil edilmesidir. Temel prensibi, bir moleküler grafiğin yapısal özelliklerini yakalamak ve bunları daha sonra bir çift bileşikle karşılaştırmak için bir bit dizisine kodlamaktır (Budak vd., 2023). Klasik Lipinski kurallarının sınırları dahilinde, moleküler parmak izleri genellikle küçük moleküllü ilaçlar için tasarlanır, onaylanır ve kullanılır. Her parmak izi, bir moleküldeki her atomu çevreleyen belirli dairesel alt yapıların varlığını tespit ederek küçük organik moleküllerin biyolojik aktivitelerini tahmin eder. Moleküllerin boyut ve şekil gibi

küresel özelliklerini algılamakta zorlanır. Bu kısıtlamalar, moleküler formu kodlayan ve yaygın olarak kullanılan atom çifti parmak izleri kullanılarak aşılabilir.

Son zamanlarda, ilaç-ilaç ve ilaç-protein benzerlikleri, benzer ilaçların benzer hedefleri paylaşabileceği veya tam tersi varsayımıyla yürütülen çalışmalarla, ilaç kimyasal yapılarından ve protein dizilerinden yararlanılmaya çalışılmaktadır. İlaç-ilaç benzerliği çalışmalarına baktığımızda amaç, ilgilenilen ilaca benzer farmakolojik özelliklere sahip ilaçlar bulmaktır. Hipotez, benzer ilaçların etkilerinin benzer olabileceğidir. İlaç-ilaç benzerliği çeşitli kaynaklardan hesaplanabilir ve ilaç yeniden kullanımı (Bibi et al., 2022), ilaç-ilaç etkileşimi tahmini (Shtar et al., 2019), ilaç hedefleme (Karasev vd., 2022; Rifaioglu vd., 2019) ve ilaç yan etkileri tahmini (X. Zhao et al., 2019) dahil olmak üzere geniş bir uygulama alanına sahiptir. İlaç-ilaç etkileşimleri üzerine yapılan diğer çalışmalar arasında, benzer ilaçların neredeyse benzer etkileşimlere sahip olabileceğini varsayan moleküler yapı benzerlik analizine (Vilar vd., 2012) dayalı bir komşu önerme yönteminin kullanılması ve ilaç etkileşimlerinin ve ilgili önerilerin çıkarılması için bir hesaplama çerçevesinin geliştirilmesi yer almaktadır. Burada, Ebola'nın virüsünün replikasyonunu engelleme potansiyeline sahip ABD FDA onaylı ve deneysel ilaçlarla benzerliklerin çıkarılması, Ebola aktivitesini doğrulamak için bir basamak olarak kullanılabilir. Bu, Ebola için yeni terapötikler tasarlamak için yeni fırsatlar sunabilir.

DL modellerinin son zamanlarda çeşitli alanlardaki başarısı nedeniyle, YZ'nin üçüncü dalgasında olduğumuz yaygın olarak ifade edilmektedir. Son zamanlarda, GSA veya genel olarak grafik ağlarının kullanımında bir artış olmuştur. Grafik tabanlı moleküler tasarımın arkasındaki yol gösterici ilke, istenen ilaç benzeri moleküllerin tüm kriterlerini karşılayan grafikler oluşturmaktır. Birden fazla doğrusal grafiğin toplanmasıyla bir iskele ağacı oluşturulur. İskele ağacı, bir molekülü alt iskelelerin doğrusal bir düzenlemesi olarak temsil etmek için kullanılan bir yöntemdir (Mercado vd., 2021).

Çekirdek yapılar olarak belirli iskeleler içeren moleküller tasarlamak, potansiyel ilaç adayları elde etmenin etkili bir yoludur. İskele kavramı her yerde bulunmasına rağmen, molekülün hangi kısmının çekirdek olarak kabul edilmesi gerektiği konusunda

belirsizlik varsa, bir iskelenin tanımı silico'da öznel olabilir. Bemis ve Murcko (Bemis ve Murcko, 1999) tarafından yapılan popüler tanım, bir iskeleyi halkaların ve onları birbirine bağlayan atomların bir kombinasyonu olarak tanımlayarak bu belirsizlikleri göz ardı etmektedir. Belirsizliği azaltmak için, bir molekülün, Bemis ve Murcko yapısından çevresel halkaların iteratif olarak çıkarılmasıyla tanımlanan, birbiriyle ilişkili iskelelerden oluşan bir ağaç veya ağ ile temsil edilebileceği önerilmiştir.

Ortak bir iskeleyi paylaşan yapıların genellikle ortak bir sentetik yolu paylaştığı varsayılabilir ve tipik kombinatorial kütüphanelerde tüm bileşikler ortak bir iskeleye dayanır. Bu iskeleden, sadece bir halka kalana kadar halkalar birer birer çıkarılır. Bir halkanın çıkarılması, başka bir halkanın parçası olan atomlar ve bağlar olmaksızın halkanın parçası olan bağların ve atomların çıkarılması anlamına gelir. Ayrıca, çıkarılan halka atomlarına bağlı olan herhangi bir ekzosiklik çift bağ da çıkarılır. Çıkarılan halka kalan iskeleye asiklik bir bağlayıcı (açık zincirli bileşik) aracılığıyla bağlıysa, bu bağlayıcı artık bir terminal yan zincirdir ve ayrıca çıkarılır. Eğer bunu yapmak ayrık bir yapıyla sonuçlanacaksa bir halka çıkarılamaz. Birden fazla doğrusal grafiğin birleştirilmesiyle bir iskele ağacı oluşturulur. Tüm olası sapmaların kapsamlı bir şekilde sayılmasıyla, iskele ağı yaklaşımı (Varin vd., 2011) en kapsamlı halkanın belirlenmesindeki belirsizliği giderir. Çoklu molekül ağları, büyük çok parçalı asiklik grafikler oluşturmak için birleştirilebilir. İskeleler, büyük tarama veri kümelerini yapılandırmak ve keşfetmek için küçük moleküllü ilaç keşfinde yaygın olarak kullanılır. İskeleler ayrıca tıbbi kimyagerler tarafından optimizasyon projelerinde bir düzenleme prensibi olarak kullanılır. İskele ağacı yaklaşımları (Varin vd., 2011) hiyerarşik bir dizi parçalama kuralı uygular, böylece her bir iskele tam olarak iki alt parçaya bölünür ve ağaç benzeri bir grafik elde edilir.

Bunlarla ilgili çalışmalar incelendiğinde COVID-19 tedavisinde kullanılan ilaçlar ile akciğer kanserinde kullanılan ilaçlar Tanimoto/Jaccard benzerlik analizi ve GSA'ya dayalı bir GEFA modeli kullanılarak karşılaştırılmıştır. Daha sonra ilaç ve hedef protein arasındaki bağlanma gücü hesaplanarak uygun ilaçlar belirlenmiştir (Budak vd., 2023).

Başlangıçta Ebola virüsüne karşı geliştirilen bir ilaç olan Remdesivir, şu anda COVID-19'u tedavi etmek için de kullanılmaktadır (Deb ve Reeves, 2021). Remdesivir'in RNA polimeraz (RdRp) üzerindeki antiviral aktiviteleri Ebola virüsü, Orta Doğu Solunum Sendromu koronavirüsü (Middle East Respiratory Syndrome coronavirus-MERS-CoV), Şiddetli solunum sendromu koronavirüsü (acute respiratory syndrome coronavirus-SARS-CoV), insan koronavirüsleri (Human coronavirus-CoV-OC43 ve CoV-229E) ve Domuz deltakoronavirüsü (PDCoV) gibi diğer virüslere karşı bildirilmiştir (Ng vd., 2021).

Yukarıda bahsedilen ilaç-ilaç benzerliği ve kimyasal iskele yaklaşımları doğrultusunda Ebola tedavisine yönelik ilaçlar incelenmiştir. Bunlar diğer hastalıklar için FDA onaylıdır veya klinik geliştirme aşamasındadır ya da EBOV'a karşı etkili oldukları gösterilmiştir. Çalışmada kullanılan ilaçlar PubChem ve DrugBank'tan alınmıştır. Bu ilaçlardan, diğer hastalıklar için FDA onaylı Favipiravir ve kullanılan diğer onaylı veya ilaç adayı ilaçlar (örneğin, Molnupiravir COVID-19 için ilk oral ilaç adaydır) ile ilaç-ilaç benzerlik analizi yapılmıştır.

Bu bölümde, atom çifti benzerlik analizi kullanılarak bu işlem gerçekleştirilmiştir. Bu ilaçların moleküler iskeleleri bir sonraki adımda grafik ağı algoritması ile çıkarılmış ve grafiksel olarak gösterilmiştir. Tüm bu sonuçlar göz önüne alındığında, başta favipiravir ve molnupiravir olmak üzere diğer ilaçların kullanımının tavsiye edilmesinin, Ebola virüsünün tedavisinde klinik ve bilgisayar analiz süreçlerine yardımcı olabileceği öngörülmüştür. Bu, aynı zamanda, hedefe yönelik tedavi için hızlı bir yol haritası sağlayacaktır.

Yöntemler

Moleküler benzerlik, kimyasal bileşiklerin özelliklerinin tahmin edilmesinde, belirli özelliklere sahip kimyasalların tasarlanmasında ve en önemlisi ilaç keşif çalışmalarının yürütülmesinde kritik öneme sahiptir. Genellikle mevcut veya potansiyel olarak mevcut kimyasal yapıların büyük indeksleri taranarak oluşturulur. Atom çifti algoritması ile Ebola tedavisinde kullanılmak üzere araştırılmakta olan veya aday ilaçların ilaç-ilaç benzerlik analizini gerçekleştirmek amacıyla bu mantık çerçevesinde bir moleküler benzerlik listesi elde edilmiştir. Daha sonra yüksek

benzerliğe sahip ilaçların moleküler iskeletlerinin çıkarılması sonuçları değerlendirilmiştir.

Veri seti 1

SMILES, bu çalışmada veri seti olarak kullanılan ilaçların yapılandırılmış formatıdır. SMILES, bir molekülün 2B yapısının bir seri gösterimidir. Herhangi bir molekülü, daha sonra 2B bir yapıya yeniden eşlenebilen benzersiz bir dizeye eşler. Bazı durumlarda, farklı moleküller aynı SMILES dizisiyle eşleştirilebilir ve bu da modelin performansını düşürür. Bu çalışmada, Ebola tedavisi için seçilen FDA onaylı ilaçlar arasındaki benzerlikleri çıkarmak için halka açık veri tabanlarından DrugBank, PubChem ve Ebola için yayınlanan ilaç makaleleri bir veri kümesi olarak incelenmiştir. DrugBank, ilaç hedefleri, ilaç metabolize eden enzimler, ilaç etkileşimleri ve ilaç taşıyıcıları gibi ilaçla ilgili bilgileri içeren güvenilir bir veri tabanıdır. PubChem ilaç yapıları için bir veri tabanıdır. Kullanılan ilaçlara ilişkin bilgiler Tablo 4.1'de gösterilmektedir.

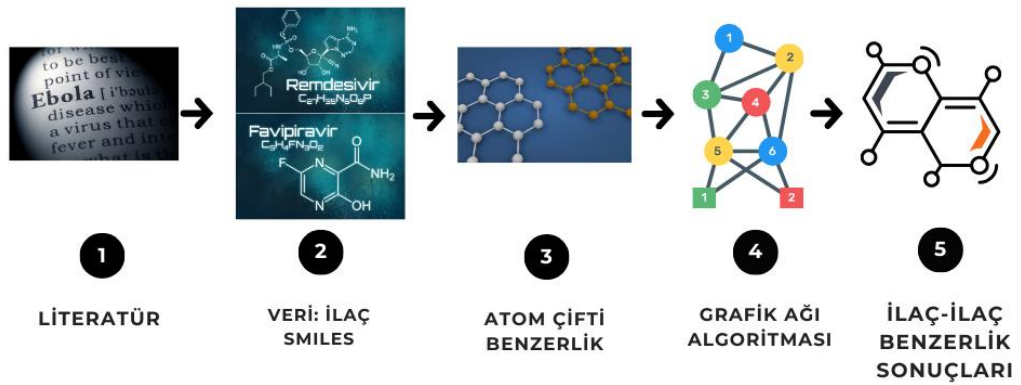
Tablo 4.1 Çalışmada kullanılan ilaçlar ve genel özellikleri

İlaç	İlaç ayrıntıları
Favipiravir	Toyama Chemical Co Ltd. tarafından geliştirilen geniş spektrumlu bir antiviraldir. Daha önce İnfluenza, Lassa virüsü, COVID-19 tedavileri için önerilen ve şu anda Ebola tedavisi için çalışılan Favipiravir'in EBOV'a karşı in vitro yüksek aktiviteye sahip olduğu gösterilmiştir (DrugBank, 2022; PubChem, 2022)
Molnupiravir	COVID-19'u tedavi etmek için kullanılan, oral yoldan biyoyararlanabilen bir izopropil ester sitidin analogudur. FDA ayrıca 23 Aralık 2021 tarihinde Molnupiravir acil kullanım yetkisi vermiştir, ancak henüz tam olarak onaylanmamıştır (DrugBank, 2022; PubChem, 2022)
Bepridil	Önemli antianjinal aktiviteye sahip, uzun etkili, seçici olmayan bir kalsiyum kanal blokeri ilaçtır (DrugBank, 2022; PubChem, 2022)
Remdesivir	COVID-19 da dahil olmak üzere RNA virüs enfeksiyonlarını tedavi etmek için kullanılan bir nükleozid analogudur. Remdesivir (GS-5734), ilk olarak 2016 yılında literatürde Ebola için potansiyel bir tedavi olarak tanımlanan bir adenosin trifosfat analogudur. Bugüne kadar Arenaviridae, Flaviviridae, Filoviridae, Paramyxoviridae, Pneumoviridae ve Coronaviridae viral ailelerine karşı in vitro aktivite göstermiştir. 19 Kasım 2020'de FDA, COVID-19 tedavisi için baricitinib ile birlikte kullanılan Remdesivir'e Acil Kullanım İzni vermiştir (DrugBank, 2022; PubChem, 2022)
Amodiaquine	Antimalaryal bir ilaçtır. Amodiaquine en az klorokin kadar etkilidir ve diğer ilaç maddeleri ile kombine edildiğinde farklı hastalıklarda olumlu sonuçlar verebilir (DrugBank, 2022; PubChem, 2022)
Azithromycin	Çeşitli bakteriyel enfeksiyonları tedavi etmek için kullanılan bir makrolid antibiyotiktir. FDA tarafından 1991 yılında onaylanmıştır (DrugBank, 2022; PubChem, 2022)
Brincidofovir	İnsan çiçek hastalığı enfeksiyonlarını tedavi etmek için kullanılan oral bir antiviral ilaçtır. Chiremax tarafından Tembexa markası altında geliştirilen Brincidofovir, çiçek hastalığı enfeksiyonunun tedavisi için Haziran 2021'de FDA tarafından onaylanmıştır (DrugBank, 2022; PubChem, 2022)

Bu çalışma için belirli ilaçlar seçilmiştir. Bu listenin bu ilaçlarla sınırlandırılmasının nedeni, grip, COVID-19, Lassa ateşi ve Ebola gibi viral hastalıklarda yaygın olarak kullanılmalarıdır. Bu ilaçların seçilmesinin bir diğer nedeni de bu ilaçların ya Ebola tedavisinde kullanılıyor olması ya da klinik kullanım için üzerinde çalışılıyor olmasıdır. Buradaki temel amaç, bu ilaçlar arasındaki moleküler yapı benzerliğine, yani bu farklı ilaçların neden aynı hastalığa etki ettiğine bakmak ve bu benzerlikten yola çıkarak kombinasyonlarının etkili olup olmayacağı konusunda bir öneride bulunmaktır. Sonuç olarak bu ilaçlar üzerinde araştırmalar yoğunlaşmıştır. İlaç-ilaç benzerlik analizi, ilacın klinikte sıklıkla kullanılmış olması koşuluyla geçerli bir alternatif olarak düşünülebilir.

Önerilen yöntem

Ebola virüsü hastalığına yönelik tedavilerin, önceki virüslerin hastalık belirtilerini tedavi etmek üzere tasarlanmış ilaçlarla yapılıyor olması, FDA onaylı ilaçların yeniden kullanımını haklı çıkarmaktadır. Bu çalışma, FDA onaylı ilaçların moleküler iskeletlerini analiz ederek hızla yayılan Ebola virüsü için yeniden tasarlanmış ilaçlar kullanarak salgınla mümkün olan en kısa sürede mücadele edilmesine yardımcı olmayı amaçlamaktadır. Şekil 4.1, bu amaçla uygun ilaç adaylarının yeniden belirlenmesi için önerilen yöntemi göstermektedir.



Şekil 4.1 Ebola tedavisinde ilaç-ilaç benzerliğini ve moleküler iskeleleri analiz etmek için önerilen yaklaşımın akış şeması

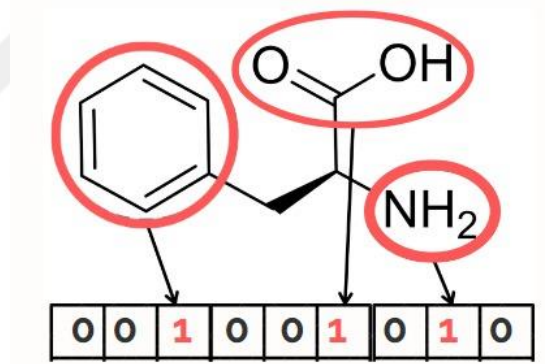
Parmak izi tabanlı benzerlik hesaplamaları için tüm ikili kimyasal benzerlikler hesaplanmıştır. Favipiravir ve Ebola tedavisi için araştırılan diğer ilaçları

karşılaştırmak ve analiz etmek için atom çifti benzerlik yöntemi kullanılmıştır. Grafik ağı algoritması, en yüksek benzerliğe sahip ilaçları seçerek moleküler iskeleleri çıkarmak için kullanılmıştır. İki ilaç benzerliğinin çekirdek yapı molekülleri böylece belirlenmiştir.

İlaç-ilaç benzerlik analizi

Kimyasal yapılar sıklıkla, yapısal özellikleri bir bit vektöründeki bitlere veya bir sayım vektöründeki sayılara dönüştüren moleküler parmak izi çıkarımı ile temsil edilir. Bu temsil, kimyasal yapıların verimli bir şekilde hesaplanmasını, analiz edilmesini ve karşılaştırılmasını sağlar.

Bu tür parmak izleri, iki molekül arasındaki benzerliği çıkarmak için kullanılır. Moleküler parmak izinin yapısı Şekil 4.2'de gösterilmektedir (Budak vd., 2023).



Şekil 4.2 Moleküler parmak izi yapısı

Belirli bir bit konumu önceden tanımlanmış bir özellik ile ilişkilendirilir. Özelliğin moleküldeki varlığı "1" ile temsil edilirken, yokluğu "0" ile temsil edilir. İlaçlar arasındaki ilaç-ilaç etkileşiminin sıfır (0) olarak değerlendirilmesi, etkileşimlerine dair hala bir kanıt olmadığını göstermesi, sağlık ve ilaç geliştirme alanındaki en kritik konulardan biridir. Sonuç olarak, birbirleriyle etkileşime girebilirler (Budak vd., 2023).

Bu çalışmada, ilaç-ilaç mesafesini hesaplamak için Jaccard benzerlik katsayısı kullanılmıştır. Bu benzerlik katsayısı ilaç-tedavi ilişkisinde iyi performans göstermektedir. Bu işlemi gerçekleştirmek için değerler ikili bitlere dönüştürülmüştür.

Anlamlı giriş değerleri 1'e, anlamsız giriş değerleri ise 0'a ayarlanmıştır. Denklem (4.1) kullanılarak, aynı tanı kodunda 1 değerine sahip her iki ilacın bulunduğu bitlerin oranı olarak hesaplanabilir. İlaç-ilaç mesafesi bu çalışmada Denklem (4.2) kullanılarak hesaplanmıştır (Budak vd., 2023).

$$J(di, dj) = \frac{|di \cap dj|}{|di \cup dj|} \quad (4.1)$$

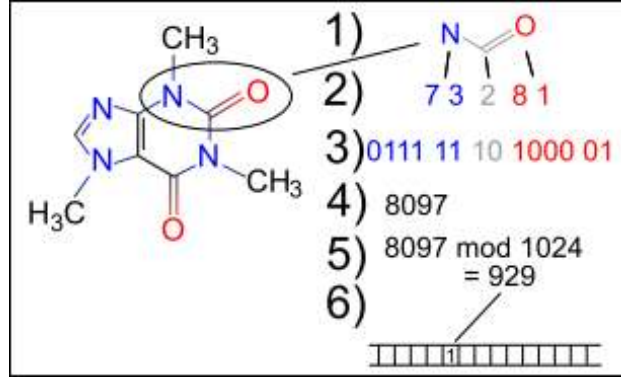
$$dj(di, dj) = 1 - J(di, dj) \quad (4.2)$$

Atom çifti benzerlik analizi

Moleküler parmak izleri, moleküler (kimyasal) yapıları bit dizeleri (1'ler ve 0'lar) olarak temsil etmek için kullanılır. Temel mantık, bir molekülün yapısal özelliklerini grafiksel olarak yakalamak ve daha sonra iki bileşik arasındaki benzerliği karşılaştırmada kullanmak üzere bunları bir bit dizisinde kodlamaktır. Bu sürecin birincil faydası, böyle bir temsili saklamak için avantajlı olması ve iki moleküler grafiği karşılaştırma hesaplama sürecini bit dizelerini karşılaştırmaktan önemli ölçüde daha hızlı hale getirerek zaman kazandırmasıdır. Atom Çifti (AP) parmak izinin uygulanması daha basittir ve Tanimoto algoritmasından daha iyi performans gösterir (Budak vd., 2023). Denklem (4.3) bir atom çifti için konfigürasyon formülünü göstermektedir:

$$A = s \left[1 - \sum_{i=1}^m \left(\frac{\tau_i}{\sum \tau_i} \cdot \frac{\Delta n_i}{n_i} \right) \right] \quad (4.3)$$

A standart ilaca ilişkin aktivite, S ilaç ve analogu arasındaki benzerlik, τ_i atom çiftinin yol uzunluğu, Δn_i standart ilacın i.heteroatom çiftinin analogundan farkı ve n_i standart ilaçtaki i.heteroatom atom çiftinin sayısıdır. Genel süreç Şekil 4.3'te özetlenmiştir (Budak vd., 2023).



Şekil 4.3 Atom çifti parmak izi yapısı

Şekil 4.3'te gösterilen adımlar aşağıdaki gibidir. Bir AP parmak izi oluşturulurken, her bir ağır atom çifti için aşağıdaki adımlar gerçekleştirilir (Budak vd., 2023):

- 1) verilen atom çiftinin ve aralarındaki en kısa yolun çıkarılması;
- 2) tanımlayıcıların kodlanması (atom tipi ve her iki atom için bağ sayısı ve topolojik uzaklıkları);
- 3) bit dizilerine dönüştürme;
- 4) bit dizelerinin bir sayıya birleştirilmesi;
- 5) sayının indeks alanına hashlenmesi;
- 6) parmak izindeki ilgili pozisyonun 1 olarak ayarlanması.

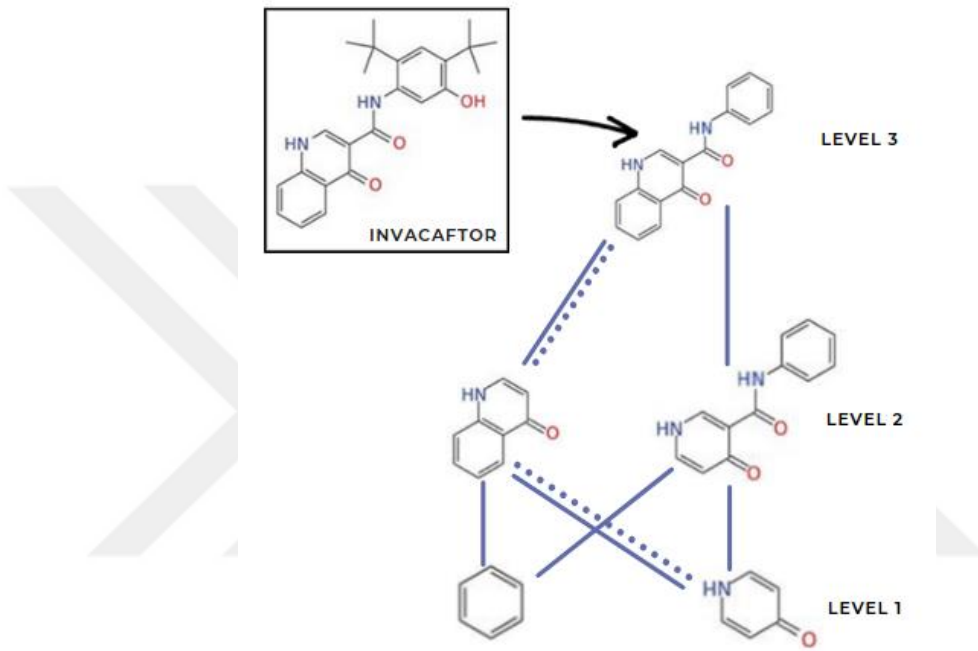
AP algoritmasının renklendirme açıklamaları aşağıdaki gibidir:

- Yeşil pozitif bir farkı gösterir (yani bitlerin çıkarılması benzerliği veya olasılığı azaltır)
- Pembe, negatif bir farkı gösterir,
- Gri, değişiklik olmadığını gösterir.

Scaffold grafik ağı algoritması

Scaffold grafik ağı (SNG), ilaç geliştirme ve keşifte kullanılan yeni bir komut satırı aracıdır. SNG, birçok özelliğe sahip büyük molekül kümelerinden moleküler iskele ağlarını ve grafiklerini çıkarmak için kullanılır. Bu ağ ve grafikler, ilaçlara dönüştürülebilecek iskeleleri belirlemek amacıyla büyük kimyasal bilgi veri tabanlarını görselleştirmek ve analiz etmek için otomatik bir hatta kullanılabilir (W.

Jin et al., 2018). İskelelerin tanımlanması, tespit edilmesi ve analiz edilmesi öznelidir çünkü bir molekülün yapısının hangi kısımlarının çekirdek yapı ve hangi kısımlarının yan zincir olarak kabul edilmesi gerektiği belirsizdir. Bu alandaki ilk çalışma, bu tür belirsizlikleri ortadan kaldırmak veya göz ardı etmek ve her molekülü ideal olmayabilecek tek bir iskeleye atamak için bir algoritma kullanmaktadır- örneğin, Şekil 4.4'te en üst seviyede Invacaftor'un Bemis-Murko iskelesidir.



Şekil 4.4 Invacaftor ilacı için iskele ağı, ilacın kompozit iskelelere ayrıştırılabileceği tüm olası yolları gösterimi

Şekil 4.4'e bakıldığında, Seviye 3'teki izole molekül ilacın Bemis-Murcko iskelesidir. Seviye 1 ve 2 alt iskeleleri, bir periferik halkanın ve ortaya çıkan yan zincirlerin bağlı oldukları ana iskelelerden çıkarılmasını temsil eder. Noktalı çizgiler İskele Ağacı algoritmasının Invacaftor ilacını nasıl ayrıştırdığını göstermektedir. SNG tüm ağı veya sadece ağacı hesaplayabilir. Bir ağ veya ağaçta birbirine bağlı aynı iskele veya iskelelere sahip moleküller benzer biyolojik aktiviteye sahip olabilir.

Bu alandaki diğer araştırmalar, molekülleri bir ağdaki düğümlere veya birbirine bağlı iskelelerden oluşan bir ağaçtaki düğümlere atayarak iskele çıkarmanın karmaşıklığını daha iyi temsil etmeye çalışmıştır. İskele Ağacı, iskele belirsizliği ile başa çıkmak için kullanılan ilk stratejidir. Her molekül tek bir Bemis-Murcko iskelesine atanır ve bu

iskele daha sonra en dıştaki halkanın iteratif olarak ayrılmasıyla ayrıştırılır. Ardışık ayrışmalar, hepsi moleküle atanan doğrusal bir iskele grafiği üretir. Şekil 4.4, Invacaftor ile oluşturulan doğrusal bir grafiği çift noktalı çizgilerle göstermektedir. İskele ağacı, birden fazla molekülden elde edilen doğrusal grafiklerin toplamıdır. İskele ağaçları faydalıdır ancak "en periferik" halkanın belirlenmesinde yeni belirsizliklere yol açabilir.

İkinci strateji olan İskele Ağları, her bir iskeleyi mümkün olan her şekilde ayrıştırır. Bu, her molekül için bir iskele ağının oluşmasıyla sonuçlanır. Şekil 4.4, Invacaftor için iskele ağını göstermektedir. Çoklu molekül ağları, büyük çok parçalı grafikler oluşturmak için üst üste bindirilebilir (Varin vd., 2011). Bu grafikteki düğümlere bağlı moleküller (her biri farklı bir iskeleye karşılık gelir) sıklıkla benzer özelliklere sahiptir. Bu iskele ağı, ilaçlara dönüştürülmek üzere en umut verici iskeleleri seçmek için tarama verilerinin analizine ve görselleştirilmesine yardımcı olur.

İskele Ağacı: Atomları düğüm ve kimyasal bağları kenar olarak içeren moleküler grafik, bir molekülün en temel matematiksel tanımıdır. Bununla birlikte, halkaların varlığı, nispeten büyük boyutları ve sınırlı kimyasal geçerlilikleri nedeniyle, moleküler grafikleri açıkça grafik olarak oluşturmak zordur. Hesaplamayı basitleştirmek için, aşağıdaki adımları izleyerek, bir moleküler grafik daha üst düzey bir temsil veya bir alt yapı ağacı olarak bir iskele ağacına dönüştürebileceğimizi görebiliriz (W. Jin vd., 2018).

Adım 1 (Altyapı): Alt yapılar bir atom veya tek halkalar olabilir. Alt yapı kümesi, ilaç benzeri moleküllerde sık görülen atomları ve tek halkaları içeren (kelime kümesi) ile gösterilir.

Adım 2 (İskele ağacı T): Bir iskele ağacı T_x , düğümleri alt yapılar olan bir yayılan ağaçtır. Moleküler grafik, X 'in üst düzey temsidir.

T_x (i) düğüm gösterge matrisi, (ii) bitişiklik matrisi ve (iii) düğüm ağırlık vektörü ile temsil edilir. T_x 'de yaprak ve yaprak olmayan düğümleri ayırt ederiz. T_x 'deki K^1 düğümleri arasında K_{leaf} düğümleri (sadece bir kenara bağlanan düğümler) ve $K - K_{leaf}$ yaprak olmayan düğümler (diğer durumlarda) vardır. Yaprak düğümler ve yaprak olmayan düğüm kümeleri sırasıyla V_{leaf} ve $V_{nonleaf}$ olarak adlandırılır.

Adım 3: N düğüm gösterge matrisi $N = \begin{pmatrix} N_{nonleaf} \\ N_{leaf} \end{pmatrix} \in \{0,1\}^{K \times |S|}$ 'e ayrıştırılır, burada $N_{nonleaf} \in \{0,1\}^{(K-K_{leaf}) \times |S|}$ yaprak olmayan düğümlere karşılık gelir ve $N_{leaf} \in \{0,1\}^{K_{leaf} \times |S|}$ yaprak düğümlere karşılık gelir. Her bir N satırı, düğümün hangi altyapıya ait olduğunu gösteren tek atışlı bir vektördür.

Adım 4: Komşuluk matrisi $A \in \{0,1\}^{K \times K}$ ile gösterilir. $A_{ij} = 1$, i_{th} düğümü ile j_{th} düğümünün bağlantılı olduğunu, 0 ise olmadığını gösterir.

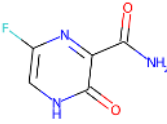
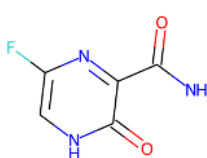
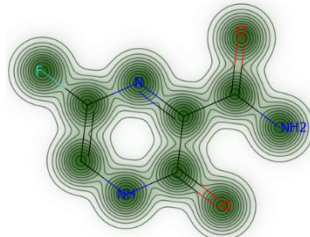
Adım 5: Düğüm ağırlık vektörü, $w = [1, \dots, 1]^T \in R^K$, iskele ağacındaki K düğümlerinin eşit ağırlıklı olduğunu gösterir (Tianfan Fu vd., 2022).

Bulgular 1:

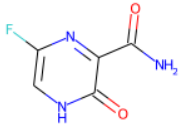
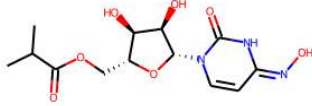
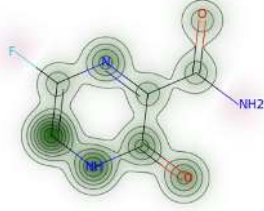
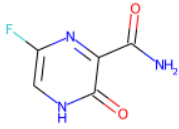
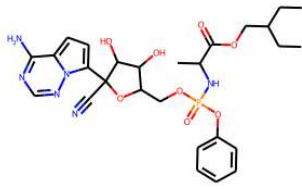
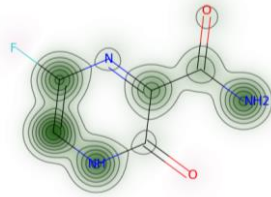
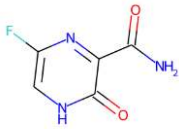
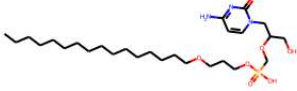
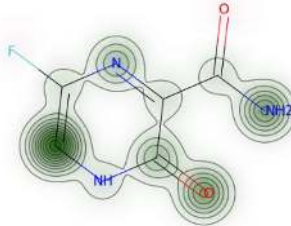
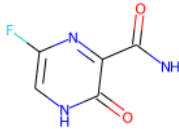
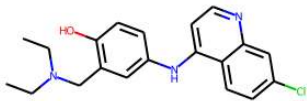
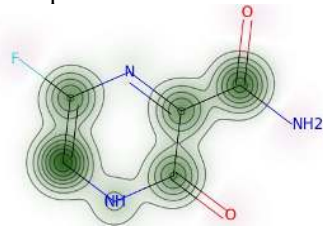
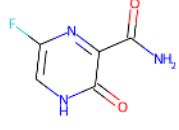
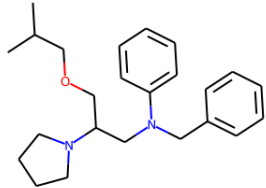
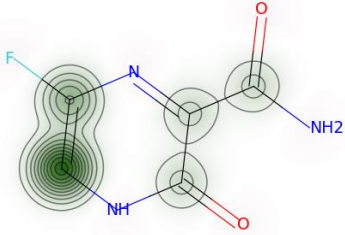
Bu çalışmada AP parmak izi benzerlik algoritması ve iskele ağ grafiği algoritması kullanılmıştır. Ebola terapötiklerine uygun mevcut ilaçları yeniden kullanmak için ilaç-ilaç benzerlik analizi yapılmıştır. Tablo 4.2, ilaç-ilaç benzerliği için önerilen yöntemin atom çifti benzerlik analizinin sonuçlarını göstermektedir.

Denklem (4.2)'de belirtildiği gibi, ilaç-ilaç benzerlikleri, Tablo 4.2'de görülebileceği gibi, ilaçların ilaç moleküllerinin ne kadar yakın olduğunu göstermektedir. Birbirine çok benzeyen ilaçlar tamamen yeşil renkte görünür.

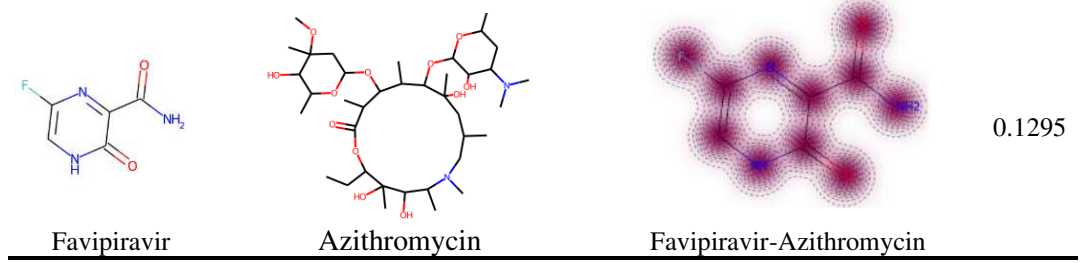
Tablo 4.2 Atom-çifti analizi ile favipiravir ilacının kimyasal yapısı ve diğer ilaçların benzerlik sonuçları

2B Referans İlaç	2B Eşleştirilen ilaç	2B Atom Çifti Benzerlik	Benzerlik Değeri
 Favipiravir	 Favipiravir	 Favipiravir-Favipiravir	1.0

Tablo 4.2 Atom-çifti analizi ile favipiravir ilacının kimyasal yapısı ve diğer ilaçların benzerlik sonuçları (Devamı)

			0.8638
Favipiravir	Molnupiravir	Favipiravir-Molnupiravir	
			0.5145
Favipiravir	Remdesivir	Favipiravir-Remdesivir	
			0.4921
Favipiravir	Brincidofovir	Favipiravir-Brincidofovir	
			0.3682
Favipiravir	Amodiaquine	Favipiravir-Amodiaquine	
			0.1995
Favipiravir	Bepridil	Favipiravir -Bepridil	

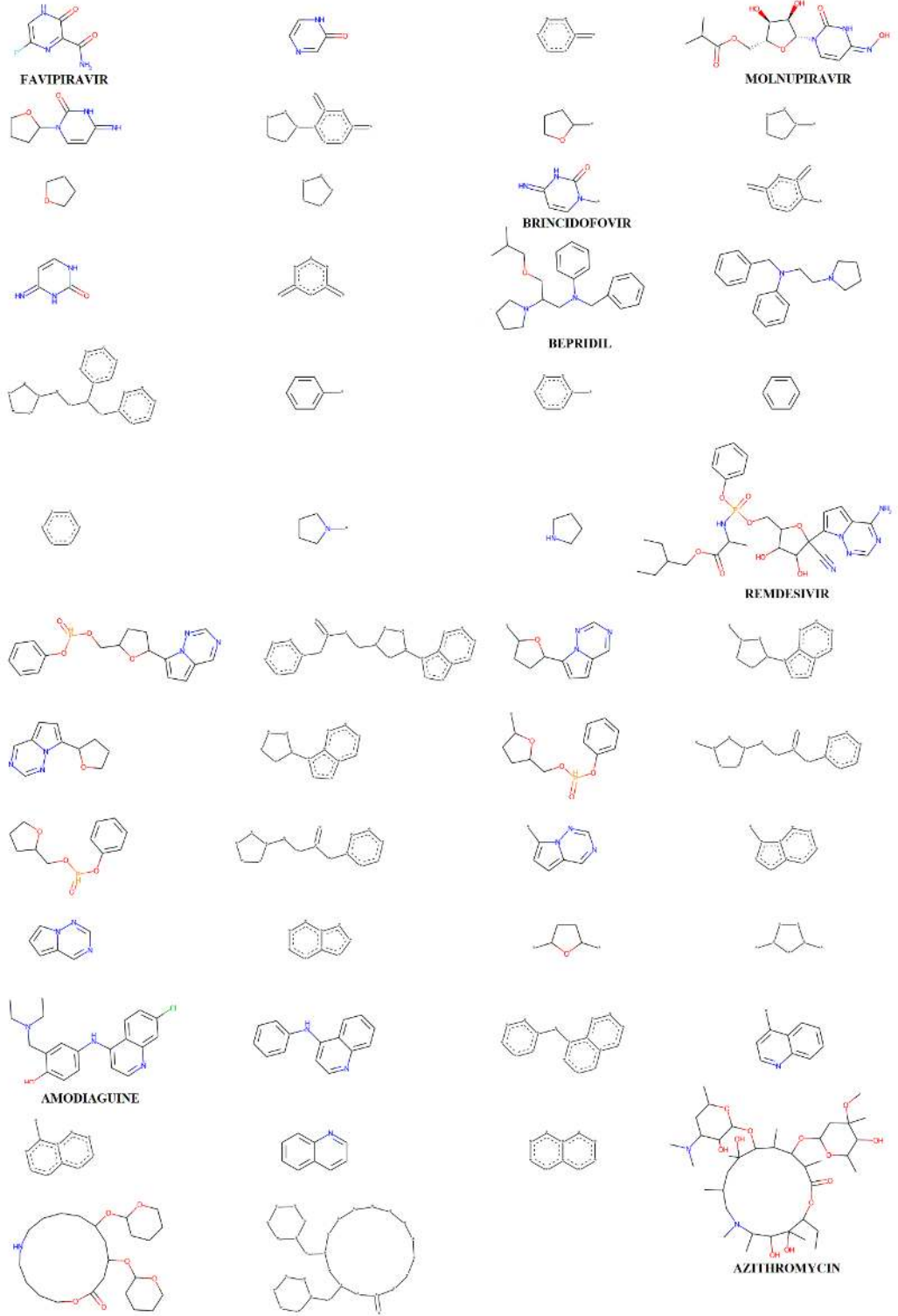
Tablo 4.2 Atom-çifti analizi ile favipiravir ilacının kimyasal yapısı ve diğer ilaçların benzerlik sonuçları (Devamı)



Tablo 4.2'deki sonuçlar, favipiravir-molnupiravir'in en yüksek benzerlik değerine sahip ilaçlar arasında olduğunu göstermektedir. Favipiravir-azitromisin ilaçları ise en düşük benzerlik değerine sahiptir.

Yeşil renk, Tablo 4.2'de gösterilen Favipiravir ilacının literatürdeki diğer ilaçlarla ne kadar örtüştüğünü gösterir. Başka bir deyişle, yeşil renk iki ilacın ne kadar benzer olduğunu temsil eder. Benzerlik haritaları, moleküler yapılar arasındaki parmak izi benzerliklerinin atomik olarak görselleştirilmesi için basit ve etkili bir yöntemdir. Atomik ağırlıklar, ilgili atomdan bitlerin çıkarılmasıyla elde edilen benzerliğin bir önceki parmak izinin (değişmemiş) benzerliği ile karşılaştırılmasıyla hesaplanır. Her parmak izi için, bitlerin karşılık gelen bir atom veya alt yapıya kadar izlenebilmesini sağlayan benzerlik haritaları oluşturulabilir.

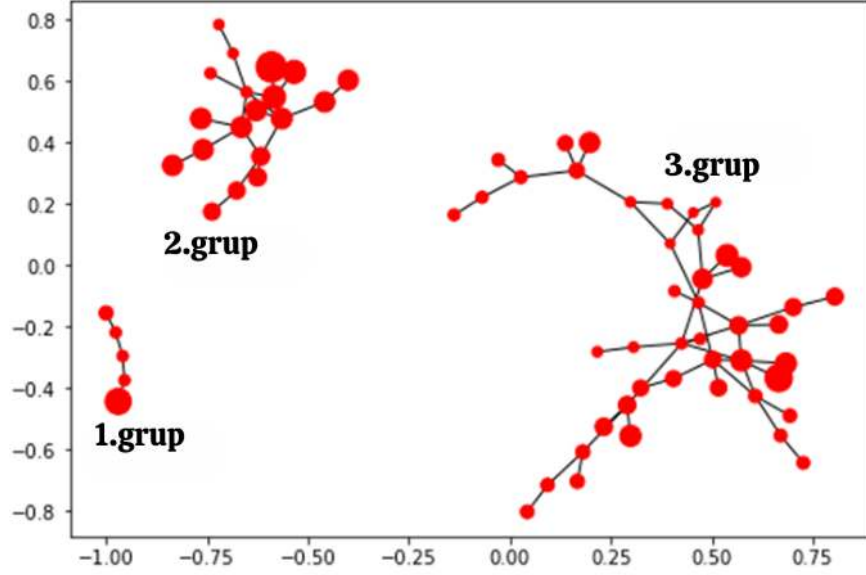
İkinci aşamada kullanılan grafik ağı, moleküler iskele ağları ve ağaçları oluşturmak ve analiz etmek için komut satırı açık kaynaklı bir Python kütüphane aracıdır. Çok sayıda girdi molekülü ile başa çıkabilir. Şekil 4.5, Tablo 4.2'de listelenen ilaçlar için grafik ağı algoritması tarafından oluşturulan iskele yapılarını göstermektedir.



Şekil 4.5 Benzerliği çıkarılan yedi ilacın grafik ağı algoritması tarafından oluşturulan bu ilaçların moleküler iskelelerinin temsili

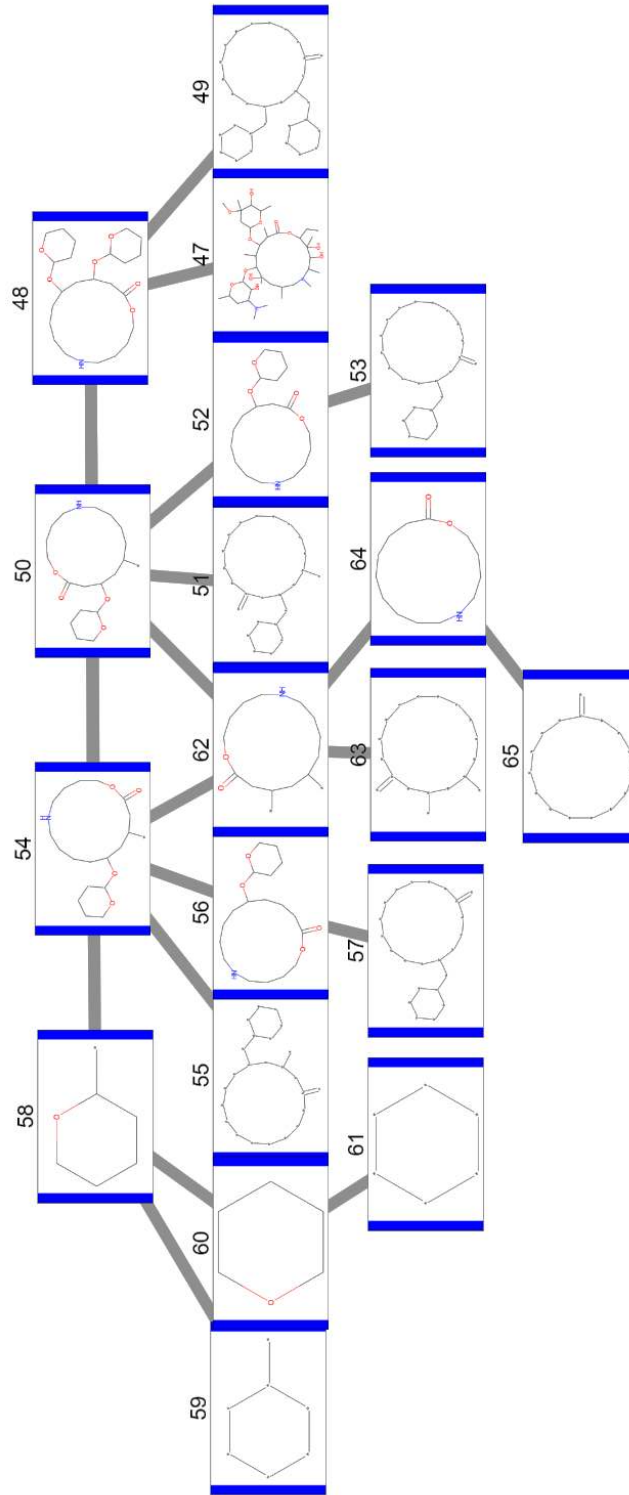
Şekil 4.5'te Tablo 4.2'deki atom çifti algoritması kullanılarak benzerlikleri de çıkarılan yedi ilacın moleküllerinin olası tüm altyapısal iskeleleri çıkarılmıştır. İskeleler, hangi

iskelenin hangisine bağı olduğunu göstermek için Şekil 4.6'da grafiksel olarak gösterilmiştir.



Şekil 4.6 Tüm ilaçların iskelelerinin düğüm grafikleri şeklinde gösterimi

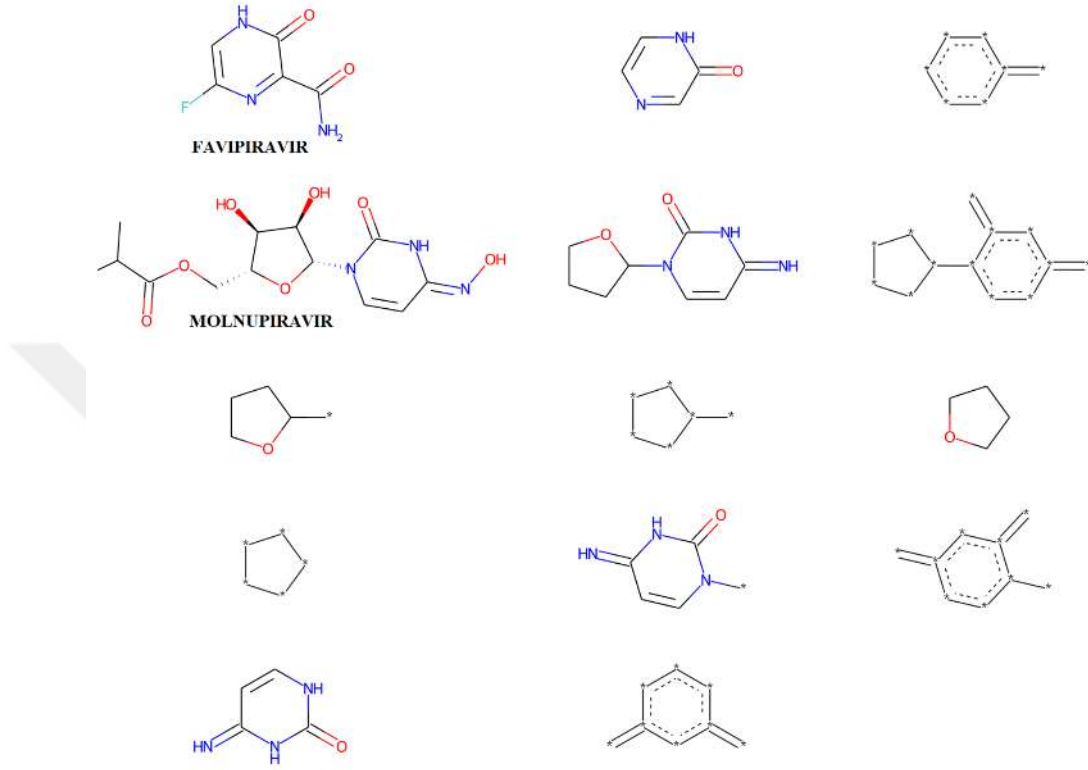
Şekil 4.6, yedi ilacın moleküllerinin iskeletlerini grafiksel olarak göstermektedir. Burada toplam 68 düğüm bulunmaktadır. Bu da yedi ilacın toplam 68 alt yapıya ve yan zincire sahip olduğunu göstermektedir. Bu düğümlerin sayısının yüksek olması, ilaçların benzerliğinin bir ölçüsü olarak yorumlanabilir. Şekil 4.6'daki ikinci grubun temel iskele ve yan zincirleri, düğümdeki moleküler yapı görüntülerini gösteren Şekil 4.7'de gösterilmiştir ve düğümler burada iskele ağacı algoritması kullanılarak yeniden çizilmiştir.



Şekil 4.7 Grafik ağı algoritması tarafından yedi (7) ilaç için oluşturulan iskelelerin moleküler yapıları da dahil olmak üzere gösterimidir. Ayrıca Şekil 4.6'daki 2. gruba ait ilaç iskelelerini ve birbirleriyle olan ilişkilerini gösterimi

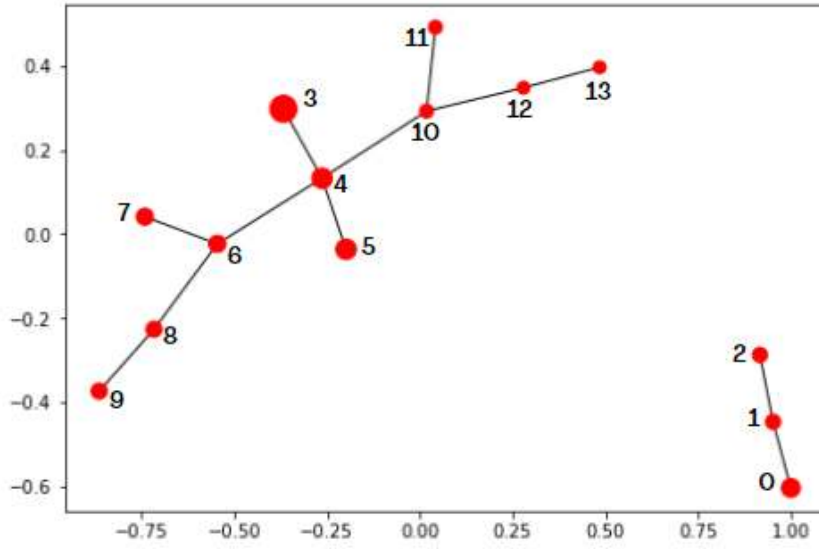
Bu analiz, ilaç benzeri moleküllere daha da optimize edilebilecek iskeleleri bulmak için kullanılabilir. Bu noktaya kadar, influenza, Ebola ve COVID-19'u tedavi etmek

için önerilen bir ilaç olan favipiravir referans ilaç olarak kullanılmıştır. Şekil 4.8 ve Şekil 4.9, Tablo 4.2'de gösterildiği gibi en yüksek benzerlik değerlerine sahip favipiravir-molnupiravir ilaçlarının iskele ağını ve ağaç gösterimlerini göstermektedir.



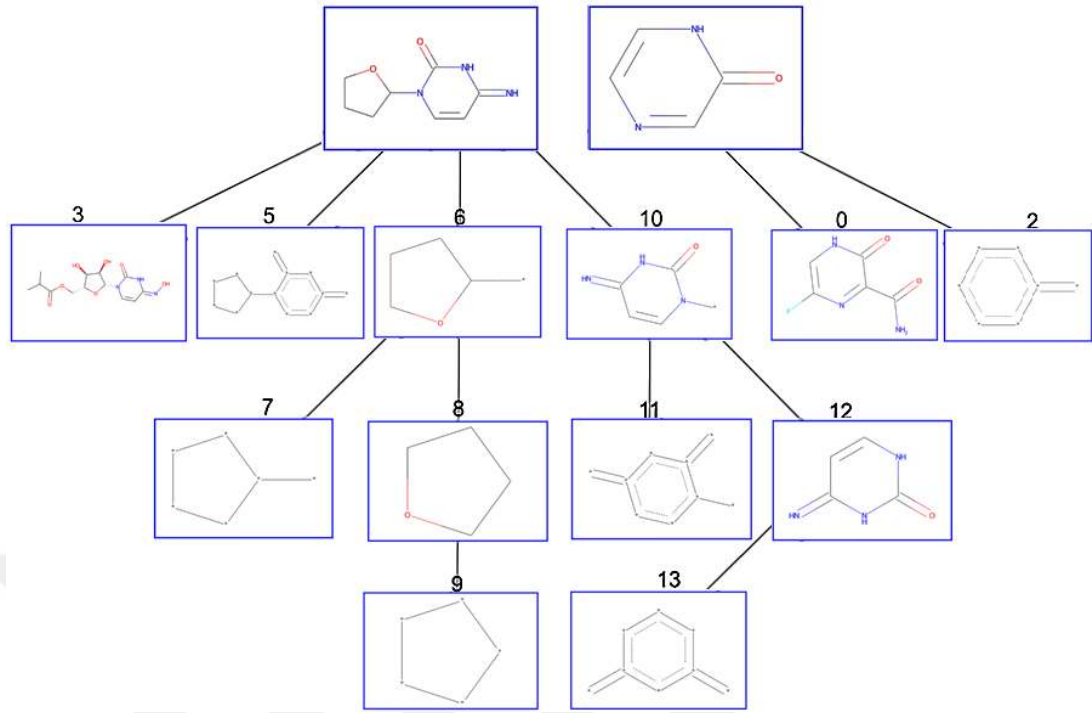
Şekil 4.8 Favipiravir ve Molnupiravir'in iskele ve yan zincirleri gösteren 2B gösterimi

Moleküllerin alt yapıları karşılaştırıldığında iki ilaç arasında 14 farklı alt yapı ve yan zincir bulunmaktadır. Şekil 4.10'da hangi yapıların hangi yapılarla ilişkili olduğu grafiksel olarak gösterilmektedir.



Şekil 4.9 Favipiravir ve Molnupiravir iskelelerinin düğüm grafikleri şeklinde gösterimi

Şekil 4.9'da toplam 14 düğüm ve bunlar arasındaki bağlantılar gösterilmektedir. Bunun nedeni, Şekil 4.8'de gösterildiği gibi, iskele algoritmasının iki ilaç molekülünün altyapı iskelelerini oluşturmak için kullanılması ve toplam 14 yapı elde edilmesidir. Şekil 4.9'da iskele ağacının çizimi gösterilmektedir. İskele ağacı bir ağ olarak çizilebilir, ancak molekül yapıları düğümde gösterilmez. Yapısal görüntü kimyager için kritik önem taşıdığından, Şekil 4.10'da yapı görüntüleri düğümde olan ağ gösterilmektedir.



Şekil 4.10 Favipiravir ve Molnupiravir ilaç moleküllerinin iskele ağacı gösterimi

Şekil 4.9 ve 4.10 karşılaştırıldığında, aynı moleküller ve alt yapılar numaralandırılmıştır. Favipiravir ve molnupiravir molekülleri ve altyapı molekülleri Şekil 4.9'da düğümler ve aralarındaki bağlantılarda kenarlar olarak gösterilmiştir. Şekil 4.10'da gösterildiği gibi düğümlerde moleküler yapı görüntüleri de verilerek ilaçlar hakkında yorum yapabilmemiz sağlanmıştır. Bu, iki ilacı birbirine bağlayan düğüm sayısının yüksek olduğu ve yapılarının benzer olduğu anlamına gelmektedir.

Bir hastalığı tedavi etmek için mevcut ilaçları yeniden kullanmak, ilaç geliştirme sürecinde yeni bir ilaç keşfetmekten çok daha hızlı olabilir. Bu mevcut ilaçlar başka bir hastalığı tedavi etmek için onaylanmışsa, daha hızlı ve daha yüksek oranda yeni tedaviler sağlayabilirler; ancak, yeni bir hastalığı tedavi etmek için onaylanmış ilaçları yeniden kullanırken dikkate alınması gereken birçok faktör vardır. İlaç geliştirmenin en önemli alanlarından biri olan ilaç-ilaç benzerliği karşılaştırmaları, alternatif ilaç hedeflerinin ve belirli bir ilacın hedef dışı etkilerinin keşfedilmesine yardımcı olur. Bu, farklı farmakolojik sınıflara ait olan veya farklı kimyasal yapıya sahip olan benzer etki şekillerine sahip ilaçları tanımlamak için yapılır. İki farklı ilacın kimyasal yapısı, kimyasal yapı ve biyolojik aktivite karşılaştırılırken yapısal benzerlikler

arasındaki potansiyel ortak biyolojik aktiviteyi arařtırmak için karřılařtırılabilir. Byle bir karřılařtırma son derece yararlıdır, ancak ilalar benzer bir farmakolojik yapıya sahip olsalar bile, etkileri bazen tamamen farklı olabilir. Birok ilacın eřitli farmakolojik etkileri vardır nk vcutta eřitli yerlerde alıřırlar. Bu durum, bu karřılařtırmaların zaman zaman ne kadar zorlayıcı olabileceğini gstermektedir. zellikle son zamanlarda kresel salgınlarla boėuřan dnyada, virsle iliřkili proteinleri hedef alan influenza, ebola ve COVID-19 gibi hastalıklar için eřitli ilalar ve/veya teraptik adaylar řu anda klinik geliřtirme ařamasındadır. COVID-19 ve ebola pandemilerini nlemek için ařılar mevcuttur. Bununla birlikte, mutasyona uėramıř viral suřların ortaya ıkması, ařı retimi ve daėıtım zorlukları ve diėer faktrler nedeniyle salgın zor olmaya devam etmektedir. Ařılanmıř olanlar bile, zellikle de mevcut ařılara daha az duyarlı olan varyantlar sz konusuysa, enfeksiyon ve hastalıėa karřı baėıřık olmayabilir. İnfluenza, ebola ve zellikle COVID-19 pandemilerinin dnya apında yayılma hızının endiře verici olması, mevcut antiviral ila seeneklerimizin ne kadar yetersiz olduėunu ortaya koymuřtur. Hızlı klinik denemelerin ardından, yeni geliřtirilen tm antiviral ilaların tedavisi hızlandırıldı. Bu durum gz nne alındığında, ebola için bir tedavinin tesadfen keřfedilmesi olası deėildir, ancak bilinen ilalar zerinde yayınlanan bazı alıřmalar bize bakmamız gereken doėru yn gsterebilir.

COVID-19 için remdesivir'e benzer bir etki mekanizmasına sahip, aėızdan uygulanan bir ila olan Favipiravir, kullanımını destekleyecek daha az saėlam veriye sahiptir. Bununla birlikte, hafif ila orta dereceli vakalarda dikkate alınması gereken bir ajandır. İnfluenza, Ebola ve COVID-19 virslerine karřı etkinliėi olan favipiravir; arenavirs, bunyavirs, filovirs, Batı Nil virs, sarı humma virs ve řap virsnn hcre kltr ve hamster modellerinde teraptik olarak etkilidir. COVID-19 tedavisi için ilk oral antiviral ila adayı olan molnupiravir, profilaksi, tedavi ve bulařmanın nlenmesi dahil olmak zere oklu SARS-CoV-2 prelinik modellerinde aktiftir. Molnupiravirin prelinik ve klinik alıřmalarda en yaygın SARS-CoV varyantlarına karřı etkili olduėu gsterilmiřtir (Agrawal vd., 2020).

Kombinatoryal terapi, yan etkilerin azalması veya etkinliėin artması nedeniyle karmařık hastalıkların tedavisinde umut verici bir strateji olabilir. rneėin, Abdelnabi

ve arkadaşları tarafından yapılan çalışmalar, influenza ve hemorajik ateş virüsleri de dahil olmak üzere bir dizi virüse karşı aktivite kanıtı ile hem Favipiravir hem de Molnupiravir'in panviral aktivitesini göstermektedir (Oliver ve Hinks, 2021). Yakın zamanda yapılan bir çalışmada, Favipiravir ve Molnupiravir antiviral kombinasyonunun COVID-19'a karşı etkili olduğu bulunmuştur (Hussien ve Abdelaziz, 2020). Bu kombinasyon çalışması, çalışılan hamster modellerinin akciğerlerinden SARS-CoV-2 virüsünün eradike edilmesinde %60'lık bir başarı oranına katkıda bulunmuştur (Hussien ve Abdelaziz, 2020). Favipiravir ve Molnupiravirin birlikte formülasyonu mümkündür çünkü her iki ilaç da oral yoldan biyoyararlanıma sahiptir (Hussien ve Abdelaziz, 2020). Bu bağlamda sonuçlarımıza baktığımızda Favipiravir ve Molnupiravir ilaçlarının benzerlik oranlarının yüksek olduğunu görebiliyoruz. Bu temelde, Molnupiravir ve Favipiravir kombinasyonları Ebola tedavisinde etkili olabilir.

Önerilen bu kombinasyon klinik çalışmalarda test edilmeden önce bazı soruların iyice araştırılması gerekmektedir. Her iki ilaç da hamster modelleri gibi deneysel modellerde antiviral dozlarda elde edilmiştir. Örneğin favipiravir, bir hamster modelinde SARS-CoV-2'ye karşı antiviral aktivite göstermiş, ancak daha büyük hayvan modellerinde veya insanlarda gösterememiştir. Bunun nedeni favipiravir farmakokinetiğinin karmaşıklığı ve insanlarda SARS-CoV-2'ye karşı böyle bir antiviral aktivite elde etmek için daha önce kullanılandan daha yüksek dozlara ihtiyaç duyulması olabilir (Abdelnabi vd., 2021).

Favipiravir, çeşitli viral salgınları tedavi etmek için kullanılan güçlü bir antiviral ilaçtır. İlaç-ilaç benzerlik analizi prosedürleri ve iskele çıkarımı sonuçları, diğer ilaçlarla kimyasal yapı benzerlikleri alınarak ve Favipiravir referans olarak kullanılarak elde edilmiştir. Bu bulgulara göre molnupiravir, favipiravir ile en fazla benzerlik gösteren ilaç olarak belirlenmiştir. Azitromisin ise moleküler yapı olarak en az benzerlik gösteren ilaçtır. Burada neredeyse hiç yapısal benzerlik yoktur. Scaffold grafik ağı algoritması ile elde edilen sonuçlar incelendiğinde favipiravir ve molnupiravir'in moleküler ve altyapı moleküler iskeleleri olmak üzere toplam 14 molekül ayrıştırılmıştır. Bunlardan 11 tanesinin benzer yapıya sahip olduğu ve birbirleriyle ilişkili olduğu tespit edildi. Bu 14 moleküler yapıyı grafiksel olarak temsil

ettiğimizde, ortak düğüm sayısının yüksek olması (11) ilaçlar arasındaki benzerliği göstermektedir. Benzer yapıya sahip ilaçlar benzer etkilere sahip olabilir. Bu da ilaç-ilac benzerliğinin nasıl son derece faydalı olabileceğini göstermektedir. Ayrıca, iskele algoritmasının grafiksel sonuçlarının yorumlana bilirliliğini artırmak için yapısal moleküllerin grafiklerdeki düğümlerde temsil edilmesi kritik önem taşımaktadır.

Yakın zamanda COVID -19 hastalarında ön antiviral aktivite gösteren yüksek direnç bariyerine sahip iki oral ilaç olan molnupiravir ve favipiravir, hamster modellerinde SARS-CoV2 tedavisinde özellikle etkilidir (Abdelnabi vd., 2021). Bu araştırmanın birincil amacı, Ebola tedavisi için klinik ortamda çalışılan ilaçlara bakmak ve bu ilaçların etkilerinin kombinasyonlarının Ebola tedavisinde etkili olabileceğini önermektir. Bu amaçla, ilaç-ilac benzerliğini belirlemek için atom çifti algoritması ve ardından bu ilaçların moleküler iskelelerini belirlemek için iskele ağı grafiği algoritması kullanılmıştır.

Bu çalışmada, Ebola tedavisi için halihazırda araştırılmakta olan veya klinik deneme aşamasında olan yeni potansiyel ilaç adaylarını bulmak için atom çifti algoritması ile kimyasal yapı benzerlikleri için literatürdeki ilaçların benzerliği analiz edilmiş ve incelenmiştir. Daha sonra bu benzer ilaçların çekirdek molekülü ve benzer altyapı molekülleri için iskeleleri çıkarmak üzere grafik ağı algoritması kullanılmıştır. Kimyagerlerin bunları önceden bilmesi gerekir.

Sonuç olarak, şu anda COVID-19 tedavisi için klinik çalışmalarda incelenmekte olan iki ilaç veya ilaç adayı olan molnupiravir ve favipiravir kombinasyonu, ebola virüsüne karşı etkinliği test etmek için klinik çalışmalar geliştirmek için temel oluşturabilir. Ebola tedavisi için böyle bir kombinasyon, uygun dozajların ayarlanmasına ve uygulanan dozların azaltılmasına da yol açabilir. Listedeki diğer ilaçlar da aynı kombinasyonlar kullanılarak değerlendirilebilir. Bu bakımdan çalışma, ebola tedavisine yeni bir yaklaşımı temsil etmektedir.

4.1.2 COVID-19 ve akciğer kanser ilaçlarının benzerlik analizi ve grafik sinir ağları tabanlı ilaç-protein bağlanma enerjisinin hesaplanması

Kasım 2019'da Çin'in Wuhan kentinde ilk kez bildirilen yeni koronavirüs enfeksiyonu, birkaç ay içinde hızla dünyaya yayılmış ve birçok ülkeyi büyük ölçüde etkilemiştir. Bu virüs, Virüs Adlandırma Komitesi tarafından COVID-19 olarak adlandırılmıştır ve Sars-Cov-2 olarak adlandırılmıştır (Wu vd., 2020). Bu virüsün, yarasadan pangoline ve oradan insanlara geçtiği düşünülmektedir ve beta koronavirüsler sınıfına aittir, farklı ortamlarda insanlara bulaşabilir. Sars-Cov ve Mers-Cov (The WHO MERS-CoV Research Group, 2013) gibi Sars-Cov-2'ye benzer virüslerin benzeri görülmemiş sağlık, ekonomik ve sosyal etkiler yaratabileceği düşünülmektedir. WHO, 11 Mart 2020'de bu durumu pandemi olarak ilan etmiştir (Alhudhaif vd., 2021). Pandemi, dünya çapında 180'den fazla ülkede laboratuvarca doğrulanmış Sars-Cov-2 vakalarının bildirilmesiyle dramatik bir şekilde gelişmiştir. Koronavirüsler, alfa, beta, gamma ve delta olmak üzere dört alt gruba ayrılır ve genom boyutları 26 KB ile 32 KB arasında değişir. COVID-19 salgınının hızla yayılması, vaka sayısının ve ölüm oranlarının artması, yetersiz tedavi yöntemleri ve aşı seçenekleri birçok hükümeti pandemi ile mücadele etmek için karantina ve seyahat kısıtlamaları gibi sıkı önlemler almaya yöneltmiştir. Tüm bu sonuçlar, Sars-Cov-2 virüsünü önlemek için etkili tedavi seçeneklerinin belirlenmesinde önemlidir.

İlaç tasarımı ve geliştirilmesi, ilaç firmaları ve kimya bilim insanları için önemli bir araştırma alanıdır. Ancak düşük etkinlik, hedef dışı dağıtım, zaman tüketimi ve yüksek maliyet, ilaç tasarımı ve keşfini etkileyen engeller ve zorluklar sunmaktadır. Ayrıca genomik, proteomik, mikroarray verileri ve klinik denemelerden elde edilen karmaşık ve büyük veriler de ilaç keşfi sürecinde bir engel teşkil etmektedir. YZ ve ML teknolojisi, ilaç keşfi ve geliştirilmesinde önemli bir rol oynamaktadır. Deneysel Yüksek Verimlilik teknolojilerindeki son ilerlemeler, biyolojide moleküler veri miktarını ve kullanılabilirliğini artırmıştır. Biyolojik süreçlerdeki etkileşimlerin önemi göz önünde bulundurulduğunda, bu veriler genellikle biyolojik ağlar şeklinde temsil edilir. Bu verilerin artması, ağları analiz etmek için yeni hesaplama araçlarına olan ihtiyacı doğurmuştur. Bu alandaki en büyük eğilimlerden biri, derin öğrenmeyi bu amaçla kullanmak ve özellikle GSA olarak adlandırılan ağlarla çalışan yöntemleri kullanmaktır. COVID-19, akut solunum yolu hastalığı, ateş, öksürük, boğaz ağrısı, kas

ağrısı ve nefes darlığı gibi semptomlara neden olur (Karaman, 2021). Önceki çalışmalar, Sars-Cov ve Mers-Cov gibi ciddi solunum semptomlarına neden olan tek sarmallı RNA virüsleri tarafından tetiklenen enfeksiyon hastalıklarını etkili bir şekilde tedavi etmek için birçok mevcut ilacın yeniden konumlandırılmasını önermektedir. İlaç-Hedef Etkileşimleri (DTI) tahmini, ilaç keşfi için kritik bir parça olup, ilaç hedef etkileşimlerini ligandların hedef proteinler arasındaki benzerliğine dayanarak tahmin eder.

COVID-19'un yaşamı tehdit eden semptomlarını hafifletmekte faydalı olabilecek farmakolojik etkileri olan birçok onaylanmış kinaz inhibitörü, özellikle Sars-Cov ve Mers-Cov'un önemli arabulucuları olarak önerilmiştir. İdeal olarak, optimal farmakokinetik özelliklere sahip bir kinaz inhibitörü, viral hedefleme yoluyla enfeksiyonu doğrudan azaltabilir. Kinaz inhibitörleri, hastalık semptomlarını baskılayarak klinik fayda sağlayabilecek iki işlevli terapötik olarak yeniden kullanılabilir. Kinaz inhibitörleri, akciğer hasarı nedeniyle yaşamı tehdit eden durumların olasılığını azaltabilecek anti-enflamatuvar ve sitokin inhibitör aktivitesi gibi özelliklere sahiptir. Örneğin, Osimertinib, güçlü bir Epidermal Büyüme Faktör Reseptörü (EGFR) inhibitörüdür. Osimertinib, Sars-Cov-2'ye karşı in vitro aktivite gösteren FDA onaylı 24 ilaçtan biridir.

Daha önce FDA tarafından onaylanmış ilaçların Sars-Cov-2 ve ilgili koronavirüsler için tedavi olarak yeniden kullanılması, mevcut pandemi ortamında sınırlı olan tedavi seçeneklerinin hızlı dağıtımı için bir fırsat sunar. Bu amaçla, Mers-Cov, Sars-Cov ve Sars-Cov-2'den oluşan birleşik bir veri seti bu çalışmada kullanılmıştır. Çalışmada kullanılan ilaç veri setleri, PubChem (*PubChem*, 2021) ve DrugBank'ten (*DrugBank*, 2021) elde edilmiştir. Bu birleşik veri seti, FDA onaylı veri seti ile karşılaştırılmış ve ilaç-ilaç benzerlik analizi yapılmıştır. Bu işlem, Tanimoto/Jaccard benzerlik analizi kullanılarak gerçekleştirilmiştir. Elde edilen ilaçlar, akciğer kanserinde kullanılan ilaçlarla karşılaştırılmış ve ilaçlar atom çifti benzerlik yöntemi kullanılarak analiz edilmiştir. Belirlenen bu ilaçlar, daha önce Sars-Cov virüsünde kullanılmış olan ve hala etkili olabilecek kinaz inhibitörleri ve adaylarıdır. Çeşitli kinaz inhibitörleri seçilerek elde edilen ilaçlar, Davis veri setini kullanarak dört farklı deneysel ayardan oluşan GSA tabanlı Grafik Erken Füzyon Afinitesi (T. M. Nguyen vd., 2022) modelini

kullanarak İlaç-protein afinitesi analizi açısından incelenmiştir. Sars-Cov-2 virüsü tarafından neden olunan COVID-19 için en etkili antikor yanıtına katkıda bulunan moleküler özellikler hala belirsizdir. Bu çalışmada elde edilen sonuçlar, yeniden kullanılmış ilaçların COVID-19 virüsüne karşı kullanılması için ihtiyaç ve zamanın aciliyetinin önemli olduğu pandemi krizinde mümkün olan en kısa sürede mücadeleye yardımcı olmayı kolaylaştırabilir.

Veri Seti 2

Çalışmada kullanılan veri seti ve ilaç yapısı, SMILES olarak kullanılmıştır. SMILES, molekülün 2B yapısının bir sıra temsidir. Herhangi bir molekülü (genellikle) geri dönen benzersiz bir diziye eşleştirir. Bazen, farklı moleküller aynı SMILES dizisine eşlenebilir, bu da modelin performansını düşürebilir. Bu çalışmada, seçilmiş FDA onaylı ilaçların COVID-19 ve akciğer kanseri ilaçlarıyla benzerliklerini belirlemek ve bu ilaçların afinitesi bağlanmasını analiz etmek için kamu veritabanlarından (DrugBank ve PubChem) elde edilen veri seti kullanılmıştır. DrugBank veritabanı, ayrıntılı ilaç verilerini kapsamlı ilaç hedef bilgileriyle birleştirir. Bu veritabanı, biyoinformatik ve kimyasal informatik için bir kaynaktır. DrugBank Online, 14.556 ilaç girişine sahiptir. Bunların 131'i nutrasötikler, 2.698'i onaylanmış küçük molekülü ilaçlar, 1.473'ü onaylanmış biyolojik ilaçlar (proteinler, peptitler, aşılar ve alerjenler) ve 6.653'ten fazlası deneysel ilaçlardır. PubChem, kimyasal moleküllerin bilgi tabanıdır. Burada milyonlarca bileşiğin yapı ve tanımlayıcı bilgilerine erişilebilir. Sistem, Ulusal Biyoteknoloji Merkezi tarafından işletilmektedir. Bu merkez, ABD Sağlık Bakanlığına bağlı Ulusal Tıp Kütüphanesi'nin bir ajansıdır. PubChem, 110.025.926 kimyasal yapı ve 96.561 protein hedefi içerir. Ayrıca PubChem'e bağlı 32.816.125 bilimsel yayın, 29.940.379 patent ve veri sağlayan 803 kuruluş bulunmaktadır. DrugBank, ilaç hedefleri, ilaç enzimleri, ilaç etkileşimleri ve ilaç taşıyıcıları gibi ilaç bilgilerini içeren güvenilir bir veritabanıdır. PubChem, ilaç yapıları için bir veritabanıdır. Benzerlik analizi için, akciğer kanseri ilaçları, FDA onaylı ilaçlar ve daha önce Sars-Cov ve Mers-Cov hastalıklarının tedavisinde kullanılan ve şu anda Sars-Cov-2 için klinik denemelerde kullanılan ilaçlar kullanılmıştır. Tablo 4.3, kullanılan ilaçların bileşik numaralarına dair bilgileri göstermektedir.

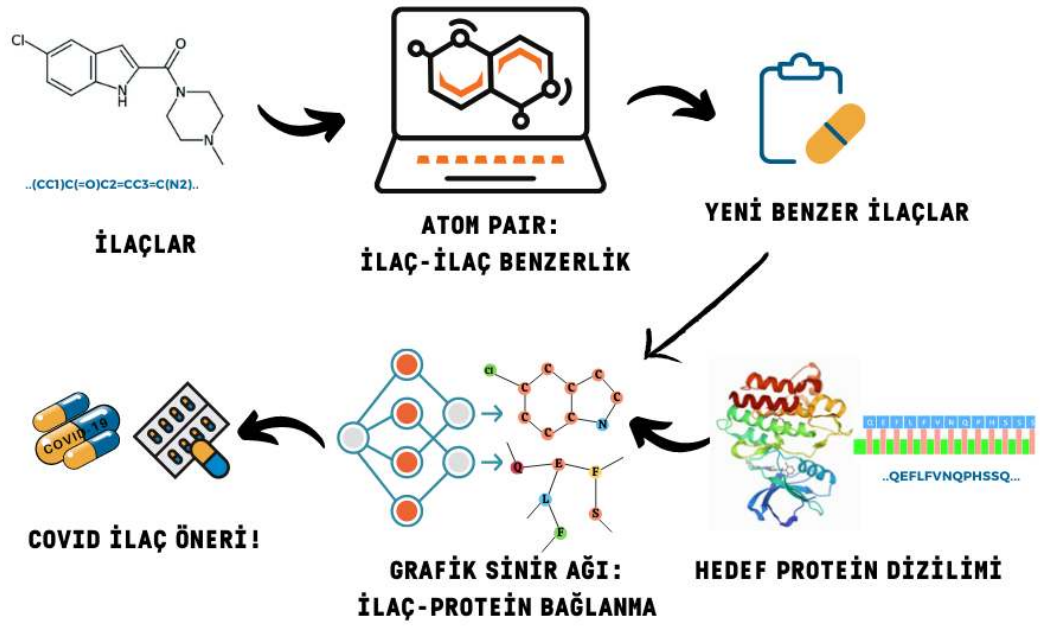
Tablo 4.3 İlaç veri setleri

Veri setleri	İlaç özellikleri	Bağlantı	Bileşenlerin sayısı
Akciğer kanser ilaçları	Moleküler yapı	https://go.drugbank.com/ , https://pubchem.ncbi.nlm.nih.gov/	401
FDA onaylı ilaçlar	Moleküler yapı	https://go.drugbank.com/ , https://pubchem.ncbi.nlm.nih.gov/	956
Sars-Cov-2 Mpro	Moleküler yapı	https://go.drugbank.com/	101
Davis veriseti	Moleküler yapı	https://go.drugbank.com/	20 000

İlaç yeniden kullanımı, bu çalışmanın stratejisi olarak, ilaç klinik olarak sıkça kullanılmışsa geçerli bir alternatif olarak değerlendirilebilir. Özellikle, COVID-19 tedavisi için yeniden değerlendirilen birçok ilaç, kanser tedavisinde kullanılmaktadır. Bunun nedeni, enfekte hücrelerin 'viral programa' uyum sağlamak için nükleik asit, protein ve lipid sentezini artırmaya ve enerji metabolizmasını artırmaya zorlanmasıdır. Aynı özellikler kanser hücrelerinde de görülmektedir. Bu, belirli kanser hücresi yollarına müdahale eden ilaçların viral replikasyonu yenmede de etkili olabileceği anlamına gelir. COVID-19, beyin, böbrek, karaciğer ve özellikle akciğerler gibi birçok organı etkileyebilir. Mortalite üzerinde etkisi olan en çok etkilenen organ, akciğerdir. Bu nedenle, solunum yolu sendromuna neden olan COVID-19'un tedavisi, akciğer hastalığının semptomlarını tedavi etmek için kullanılan ilaçlarla sağlandığından, bu çalışmada akciğer kanseri tedavisinde kullanılan ilaçlar, yeniden konumlandırılmış ilaçların listesini hazırlamak amacıyla kullanılmıştır.

Önerilen yöntem

COVID-19 hastalığının tedavisinde kullanılan ilaçların, akciğer hastalığı semptomlarını tedavi etmek için tasarlanmış ilaçlar olması, FDA onaylı ilaçların yeniden kullanımını haklı çıkarmaktadır. Bu çalışma, FDA onaylı ilaçlar ile COVID-19 ve akciğer kanseri ilaçları arasındaki benzerlikleri belirlemeyi ve bu ilaçların moleküler afinitesi bağlanma modu analizi ile hızlıca COVID-19 virüsüne karşı kullanılarak pandemiye karşı mücadeleyi desteklemeyi amaçlamaktadır. Bu amaçla, yeniden kullanım ilaç adaylarını belirlemek için önerilen yöntem Şekil 4.11'de gösterilmektedir.



Şekil 4.11 Önerilen yöntemin akış diyagramı

Önerilen yöntem iki aşamadan oluşmaktadır:

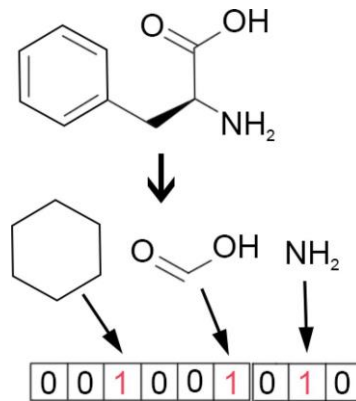
İlk aşamada, ilaç-ilaca benzerlik analizi yapılmıştır. Bu süreçte, COVID-19 tedavisinde kullanılan ilaçlar, FDA onaylı ilaçlar ve akciğer kanseri tedavisinde kullanılan ilaçlar kullanılmıştır. Yeniden kullanımı için ilaç adaylarının bir listesini oluşturmak için biyoenformatik analizler gerçekleştirilmiştir. COVID-19 hakkında sınırlı bilgi olduğu için, çalışmalarımızı benzer patojenlere odakladık ve FDA onaylı ilaçları COVID-19 tedavisinde kullanılan ilaçlarla karşılaştırarak benzer yeniden kullanılabilir ilaçların bir listesini oluşturduk. Her bir ilaç bileşiğinin genişletilmiş bağlantı parmak izlerini Jaccard benzerlik katsayısını kullanarak hesapladık. Parmak izi tabanlı benzerlik hesaplamaları için uygun bir seçenek olduğu kanıtlanmış olan Tanimoto benzerliği kullanarak tüm ikili kimyasal benzerlikleri hesapladık. Elde edilen ilaçlar, akciğer kanseri tedavisinde kullanılan ilaçlarla karşılaştırıldı ve AP benzerlik yöntemi ile analiz edildi. Böylece hem akciğer kanseri tedavisinde hem de COVID-19 tedavisinde kullanılan ve FDA onaylı benzer ilaçların bir listesi elde edildi. Bu benzer ilaçlar listesi bir sonraki adımda kullanılmıştır.

İkinci aşamada, İlaç-protein Afinitesi analizi yapılmıştır. Bu analiz, ilaç-ilac benzerlik analizinden elde edilen ilaçların bağlanma modu ve afinitesini hesaplamak için gerçekleştirilmiştir. Bu aşamada, ilk aşamada elde edilen ilaç listesi ve Tablo 1'de

belirtilen Davis veri seti kullanılmıştır. Davis veri seti, ilaçların hedef-protein dizisini içermektedir. Benzer matris tamamlama problemleri için kullanılan matris faktörizasyonu gibi diğer yöntemler, her yeni ilaç veya hedef veri setine eklendiğinde yeniden eğitilmelidir. Bu sorunu aşmak için GSA tabanlı model kullanılmıştır. Kullanılan GSA türü, doğası gereği indüktif bir modeldir. Bu, eğitim sürecini tekrarlamadan eğitim sırasında görülmeyen hedefler ve bileşikler hakkında tahminler yapmak için kullanılabileceği anlamına gelir. Özellikle, bir ilaç, daha büyük bir kalıntı-ilaç kompleksi grafında bir düğüm olarak hareket eden atomların bir grafiği olarak modellenir. Ortaya çıkan model, anlamlı derinlemesine iç içe geçmiş bir GSA'dır. Eğitimler, yeni ilaçlar veya hedefler gibi senaryoları değerlendirmek için farklı ayarlar altında gerçekleştirilir. Afinitesi değeri hesaplamak için Dissosiasyon sabiti (KD) parametresi kullanılmıştır. Önerilen yöntemin akış şemasında belirtildiği gibi, ilaç-ilaç afinitesi (AP, Tanimoto) ve DTI için birleşik yöntem, belirli bir moleküler hedefle numunelerin (ligandlar) bağlanma afinitesini tahmin eder.

İlaç-ilaç benzerlik analizi

Kimyasal benzerlik ve literatüre dayalı ilaç benzerliği ile ilişkili olan klinik ilaç-ilaca benzerlik, ilaç tedavisi benzerliğini ve hasta benzerliğini değerlendirmede birçok potansiyel uygulamaya sahiptir. Kimyasal yapılar, moleküler parmak izleri çıkararak özellikleri temsil eder, burada yapısal özellikler bir bit vektöründeki bitlere veya bir sayım vektöründeki sayılara dönüştürülür. Bu temsil, kimyasal yapıların hesaplamasını verimli bir şekilde incelemeyi ve karşılaştırmayı sağlar. Bu tür parmak izleri kullanılarak, iki molekül arasındaki benzerlik çıkarılır. Moleküler parmak izi yapısı Şekil 4.12'de gösterilmiştir.



Şekil 4.12 Moleküler parmak izi yapısı

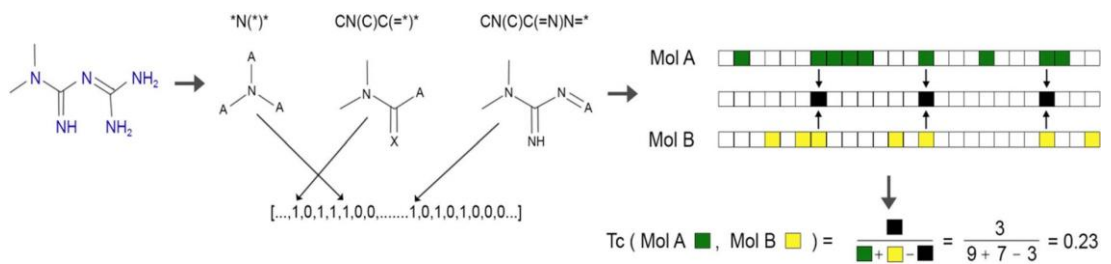
Bir bit pozisyonu, tam olarak önceden tanımlanmış bir özellik ile ilişkilendirilir. '1', molekülde bir özelliğin varlığını, '0' ise molekülde bir özelliğin yokluğunu ifade eder. İlaç geliştirme ve sağlıkta en kritik konulardan biri olan ilaç-ilaç etkileşimi sıfıra ayarlandığında, bu, henüz etkileşimlerine dair bir kanıt olmadığını gösterir. Dolayısıyla, birbirleriyle etkileşime girebilirler. Çalışmada, ilaç-ilaç mesafe, Jaccard benzerlik katsayısı kullanılarak ölçülmüştür. Bu benzerlik katsayısı, önemli bir ilaç-tanı ilişkisi gerçekleştirmiştir. Bunu yapmak için, değerler ikili bitlere dönüştürülmüş ve anlamlı girdiler 1'e, anlamsız girdiler ise 0'a ayarlanmıştır. Jaccard benzerlik katsayısı, her iki ilacın da aynı tanı kodunda 1 değerine sahip olduğu bit oranı olarak hesaplanabilir, bu da her iki ilaçtan en az birinin 1 olduğu durumlar arasındaki bit oranı olarak ifade edilir (Denklem 4.1'de belirtildiği gibi). Bu çalışmada, ilaç-ilaca mesafe Denklem 4.2'de belirtildiği gibi hesaplanmıştır.

a) Tanimoto Benzerlik Analizi

Parmak izi kavramı, moleküllerin yapılarını ve özelliklerini ikili (0 ve 1) sayı sistemiyle temsil etmenin bir yoludur. Bu temsil yöntemi, veri tarama işlemleri için uygulanmıştır. Bu algoritmaya dayalı birçok parmak izi algoritması ve benzerlik analizi mevcuttur. En yaygın kullanılan benzerlik algoritması olan Tanimoto'nun (Rogers ve Tanimoto, 1960) matematiksel formülü Denklem (4.4)'te gösterilmektedir.

$$Tan(V_i, V_j) = \frac{V_i V_j}{\sum_b V_{ib} + \sum_b V_{jb} - V_i V_j} \quad (4.4)$$

Burada V_i molekül A'yı ve V_j molekül B'yi temsil etmektedir. Bu fonksiyon, molekülün mevcut bağ ve moleküler yapı türleri gibi her özelliği için bir rastgele sayı üretici olarak çalışır. Şekil 4.13, Tanimoto algoritmasının yapısını göstermektedir.



Şekil 4.13 Tanimoto benzerlik algoritmasının yapısı

Şekil 4.13, iki molekülün benzerliklerinin, moleküler parmak izleri oluşturularak nasıl değerlendirilebileceğini göstermektedir. Her molekülün bir karma fonksiyonu vardır ve ardından özelliklere göre bir parmak izi oluşturulur. Şekil 4.13'teki parmak izi oluşturucu, belirli bir bağ mesafesi ve bu mesafe içindeki özelliklere bakar.

b) Atom Çifti Benzerlik Analizi

Moleküler (kimyasal) yapıların bit dizileri (1'ler ve 0'lar) olarak temsili moleküler parmak izleri ile yapılır. Temel mantık bir grafik molekülün yapı bilgisini yakalamak ve daha sonra bir bileşik çifti ile benzerliğini değerlendirme aşamasında kullanılmak üzere bit dizisine kodlamaktır. Bu işlemin kritik avantajı böyle bir temsili depolamanın oldukça farklı olmasıdır. İki moleküler grafiğin karşılaştırılması bit karşılaştırmasına göre daha hızlı ilerler bu da zaman tasarrufu sağlar. Bu benzerliğin temel mantığı atom tarafından ayarlanan bitlerin kaç tanesinin parmak izinde mevcut olduğuna bağlı olarak renklenmesidir. Detaylı bilgiler ve açıklamalar Bölüm 4.1.1'de yer almaktadır.

İlaç-Protein Afinitesi Analizi

İlaçların kan, serum veya plazmadaki proteinlere bağlanması, vücuttaki aktivite durumlarını, dağılımını, atılım hızını ve toksisitesini belirlemede önemli bir rol oynar. GSA'nın gelişmesiyle birlikte, son yıllarda ilaç keşfi üzerine yapılan çalışmalar hem özellik tahmini hem de inovatif tasarım için doğrudan moleküler grafik temsili kullanmaya odaklanmıştır. Bu çalışmada, hızlı ilaç yeniden kullanımı için çok önemli olan ilaç-protein afinitesi analizi için GEFA modeli kullanılmıştır. Burada, ilaç atomik bir grafik olarak modellenmiş ve daha sonra artık-ilac grafikinde bir düğüm olarak hareket etmiştir. Kullanılan GEFA modeli, dört farklı eğitim ayarından oluşmaktadır. Modelin hem proteini hem de ilacı bildiği, sadece proteinleri bildiği, sadece ilacı bildiği ve son olarak hem proteini hem de ilacı bilmediği durumlar. İlaç-hedef afinitesi değerleri, GEFA modelinde proteinlerin bilindiği deneysel ayar ve ilaç-ilaca benzerlik analizinin sonuçlarından elde edilen ilaç veri setine göre belirlenmiştir ve Sars-Cov-2 tedavisi için kullanılabilirliğini tahmin etmektedir. Bağlanma afinitesini elde etmek için gereken moleküler özellikleri ölçen rasyonel uygulama, fragmanların ve vuruşların seçimini hızlandırır. Ağların çıktıları, onları oluşturan birçok molekül ve etkileşimin niceliksel etkileşimi ile belirlendiğinden, ağ bileşenleri arasındaki ilişki

için denge sabitlerine ihtiyaç vardır. Bağlanma dengesi için, bir bağlanma partnerinin (burada protein, P) diğerinden (RNA) çok fazla olduğu koşullar altında, dengeye yaklaşma oran denklemi şu şekildedir:

$$k_{equil} = k_{açık}[P] + k_{kapalı} \quad (4.5)$$

Burada $k_{açık}$ birleşme hız sabiti, $[P]$ protein konsantrasyonu veya aşırı bağlanmanın ifadesi, $k_{kapalı}$ ise ayrışma hız sabitidir. Denklem (4.5)'te dengeye ulaşma en düşük protein konsantrasyonlarında en yavaş olduğundan, denge sürelerinin konsantrasyon aralığının alt ucundan belirlenmesi gerekir. Pratikte, protein konsantrasyonunun sıfıra yaklaştığı ($[P] \approx 0$) sınırlayıcı durumu düşünmek yararlıdır, bu şekilde Denklem (4.2), Denklem (4.6)'yı basitleştirir.

$$k_{equil,limit} = k_{kapalı} \quad (4.6)$$

İlaç hedeflerinin bağlanma gücünü ifade eden Dissosiasyon sabiti (KD) ne kadar düşükse, dengeye ulaşmak için gereken inkübasyon süresi o kadar uzun olur (Jarmoskaite vd., 2020).

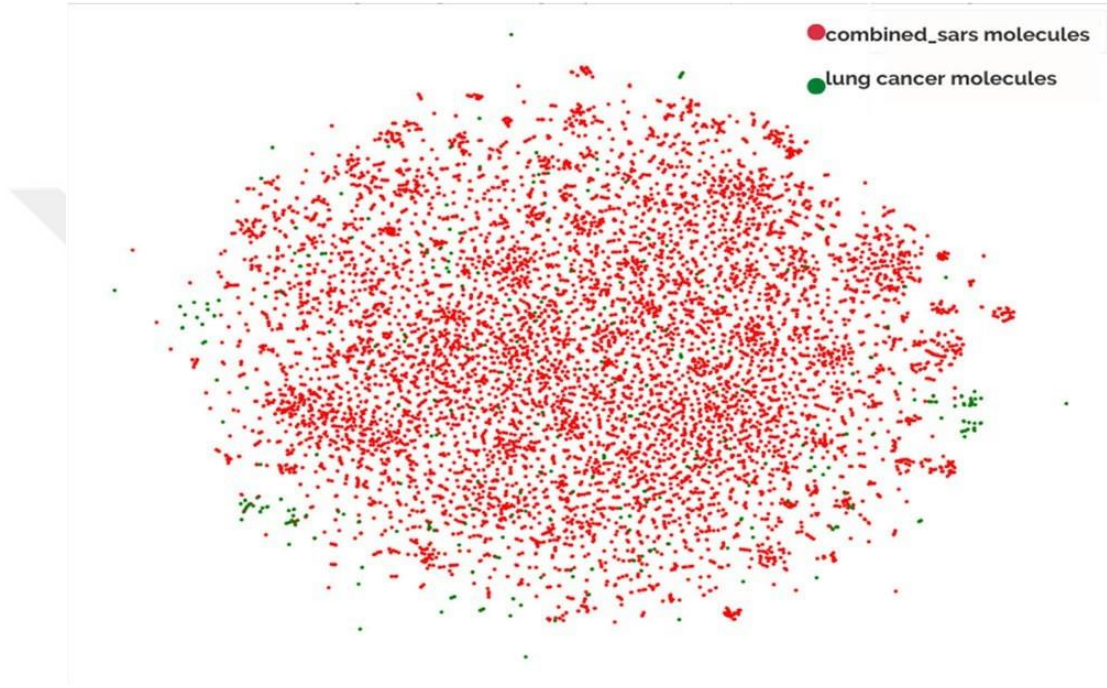
$$KD = \frac{k_{kapalı}}{k_{açık}} \quad (4.7)$$

Bir bileşiğin bir hedefe ne kadar iyi bağlandığının nihai ölçüsü, KD olarak adlandırılan ayrışma sabitidir ve bir ilaç ile hedef arasındaki bağlanma afinitesini temsil etmek için kullanılan gerçek sayısal bir ölçümdür. Denklem (4.7)'de ifade edilen KD değeri ne kadar düşükse, bağlanma o kadar güçlüdür. Bu nedenle, bu metrik bu aşamada kullanılmıştır.

Bu bölümde, GSA ile ilgili olarak daha önce Bölüm 3.2.1'de detaylı bir şekilde açıklanan bilgilere başvurulmuştur. GSA'ların temel işleyişi, düğüm ve kenar özelliklerinin güncellenmesi, komşuluk bilgisi ve ağırlıkların kullanımı konularındaki detaylı açıklamalar için Bölüm 3.2.1'e bakınız.

Bulgu 2:

Bu bölümde iki farklı yöntem kullanılmıştır: Tanimoto algoritması ve Atom Çifti (AP) algoritması. Mevcut kullanılan ilaçların COVID-19 terapötikleri olarak yeniden kullanımı için afinitesi bağlanma modu analizi gerçekleştirilmiştir. Önerilen yöntemin ilk adımında, ilaç-ilaç benzerlik için Tanimoto benzerlik sonuçları Şekil 4.14'te gösterilmiştir.



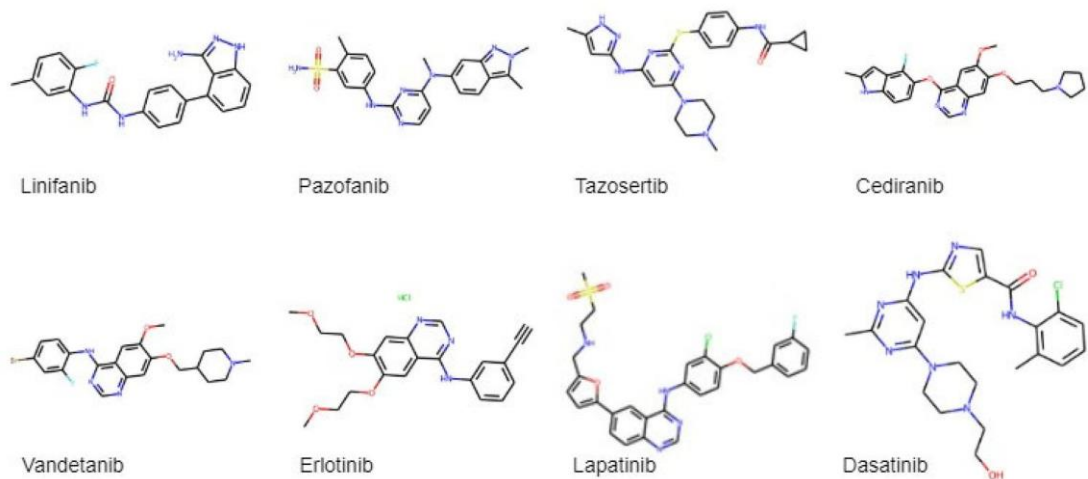
Şekil 4.14 Tanimoto benzerlik sonuçları

Şekil 4.14, Tanimoto/Jaccard algoritması mantığına göre afinitesi mesafesine göre birbirine en yakın ilaç moleküllerinin dağılımını göstermektedir. Kırmızı renkle gösterilen Mers-Cov, Sars-Cov tedavisi için kullanılan ilaçları ve Sars-Cov-2 tedavisi için klinik denemelerde onaylanmış ilaçlardan ve şimdiye kadar kullanılan ilaçların kombinasyonunu göstermektedir. Yeşil renkle gösterilenler, akciğer kanseri tedavisinde kullanılan ilaçlardır. Bu ilaçların benzerlik değerleri Tablo 4.4'de gösterilmiştir.

Tablo 4.4 Birleştirilmiş SARS ilaçları ve akciğer kanseri ilaçlarının benzerlik değerlerini gösterimi.

Akciğer Molekülü	Kanser	Birleştirilmiş SARS İlaçları	Benzerlik Oranı
Erlotinib		N-(3-Chlorophenyl)-6,7-dimethoxyquinazolin-4-amine hydrochloride	0.8013
Dasatinib		PubChem CID: 24816490	0.5520
Vandetanib		N-(3-Chlorophenyl)-6,7-dimethoxyquinazolin-4-amine hydrochloride	0.7342
Lapatinib		PubChem CID: 666049	0.5468
Gemcitabine		20-Deoxycytidine hydrochloride	0.7420
Napabucasin		3-Acetyl-2-methylnaphtho[2,3-b]furan-4,9-dione	0.7822
Diflomotecan		Camptothecin	0.8784
Bortezomib		2-Acetylamino-3-phenyl-N-[1-(pyrazine-2-carbonyl)-piperidin-4-yl]-propionamide	0.6132
Acetaminophen		Phenacetin	0.7671
P-Toluenesulfonamide		Mafenide acetate	0.8193
Pomalidomide		Thalidomide	0.8180

Denklem (4.2)'de belirtildiği gibi hesaplanan ilaç-ilaca mesafe, ilaç moleküllerinin birbirine ne kadar yakın olduğunu göstermektedir. Şekil 4.16'da görüldüğü gibi çok yakın olan, neredeyse benzer ilaçlar üst üste binmektedir. Bu benzerlik, akciğer kanseri tedavisinde kullanılan ilaçların enfeksiyon hastalığı ve akut solunum yolu sendromuna neden olan COVID-19 hastalığının tedavisinde kullanılmasına ışık tutmaktadır. Bu, çalışmanın ilk aşamasıdır ve ikinci aşamadaki ana sonuç için kullanılmıştır. Şekil 4.15, seçilen bazı SARS ilaçları ve Tanimoto/Jaccard benzerlik yöntemiyle kombine edilmiş akciğer kanseri tedavisinde kullanılan ilaçların AP 2B SMILES yapılarını göstermektedir.

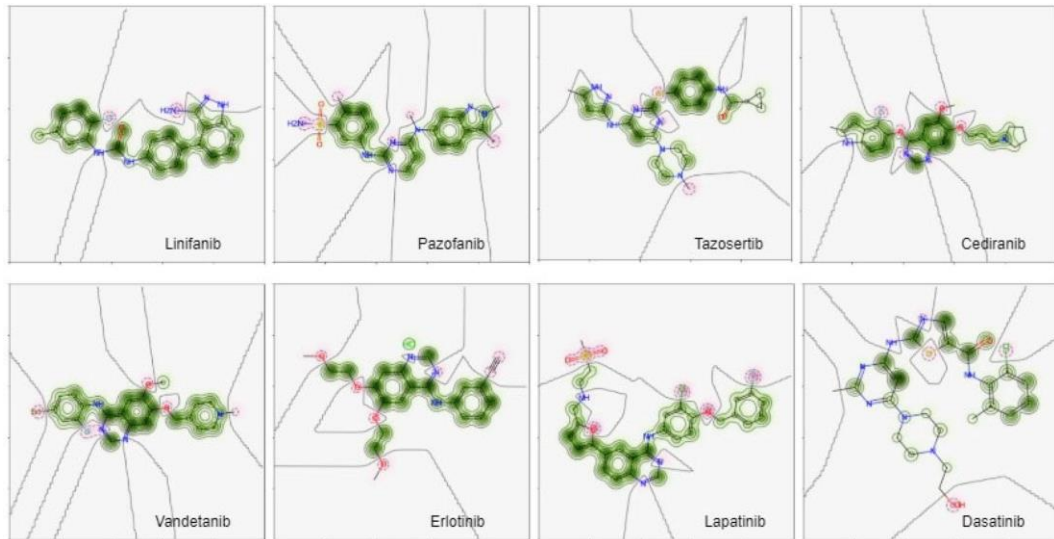


Şekil 4.15 Tanimoto/Jaccard benzerlik analizi ile elde edilen ilaçların moleküler yapıları

Şekil 4.15'te gösterildiği gibi, Tanimoto/Jaccard benzerlik analizi ile elde edilen ilaçların moleküler yapıları, parçaların dizilere nasıl kodlanması gerektiğini gösteren AP parmak izi analizi kullanılarak 2B SMILES olarak gösterilmiştir. AP parmak izleri algoritması, parçaların dizilere nasıl kodlanması gerektiğinin bir göstergesidir. Bu aşamadaki temel fikir, moleküler parmak izlerinde yapısal özellikleri (bağ yapısı, atom türü vb.) dikkate almak ve belirli bir parçanın her atomunun değerlerini almaktır. Daha sonra bunlar sınırlı sayıda bite (örneğin, bağlantı sayısı için üç bit yeterlidir) kodlanır ve yapının parmak izi dizini bit temsilini oluşturmak için birleştirilir. AP parmak izi yapısının temel mantığı şu şekildedir:

1. Tüm atom çifti parçalarını kaldırın.
2. Parçaları tamsayılara (indeksler) kodlayın.
3. n uzunluğunda bir bit dizisi oluşturun.
4. Bit dizesinin alanına indekslerin hash'lerini ekleyin.
5. Her bir hash indeksi için karşılık gelen biti açın.

Başka bir deyişle, moleküldeki atom çiftlerine karşılık gelen bitler açılır ve kalan bitler kapatılır. Şekil 4.16, elde edilen ilaçların AP benzerlik sonuçlarını göstermektedir.



Şekil 4.16 Atom çifti benzerlik sonuçları

Şekil 4.16'daki AP benzerlik sonuçlarında;

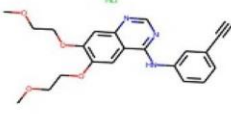
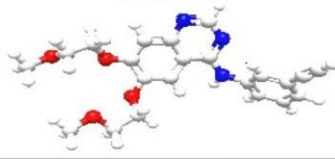
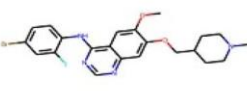
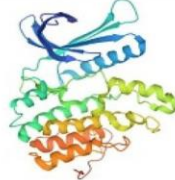
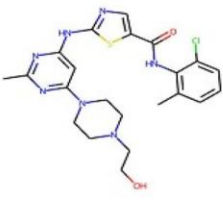
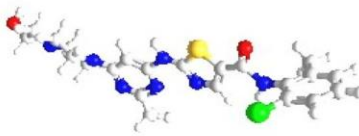
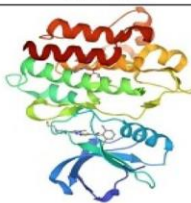
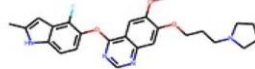
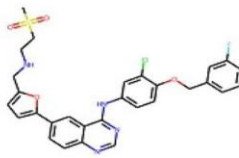
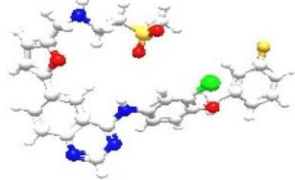
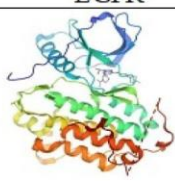
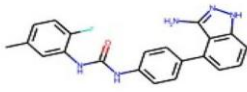
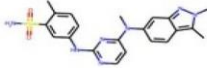
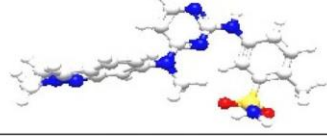
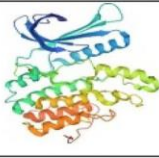
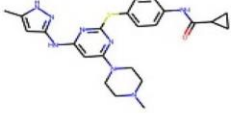
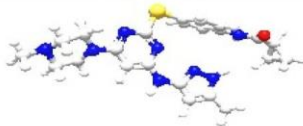
- Yeşil, pozitif bir farkı (yani, bitler çıkarıldığında benzerlik veya olasılık azalır) gösterir.
- Pembe, negatif bir farkı gösterir.
- Gri, hiçbir değişiklik olmadığını gösterir.

Yeşil renk, Şekil 4.16'da gösterilen birleştirilmiş SARS ilaçlarının akciğer kanseri tedavisinde kullanılan ilaçlarla ne kadar iyi örtüştüğünü gösterir. Başka bir deyişle, burada görülen yeşil renk, iki ilacın ne kadar benzer olduğunu gösterir. Benzerlik haritaları, moleküler yapılar arasındaki parmak izi benzerliğini atomik olarak görselleştirmek için yararlı ve anlaşılması kolay bir stratejidir. Atomik ağırlıklar, ilgili atoma ait bitlerin çıkarılması sonucu elde edilen benzerliğin önceki parmak izinin (değiştirilmemiş) benzerliği ile karşılaştırılmasıyla oluşturulur. Benzerlik haritaları, her parmak izi için oluşturulabilir ve bitlerin ilgili bir atom veya alt yapıya izlenmesine olanak tanır.

COVID-19 salgınında viral yaşam döngüsünü hedeflemek ve yaşamı tehdit eden akciğer hasarına bağlı enfeksiyon belirtilerini hafifletmek için FDA onaylı ilaçlar tarafından engellenebilecek birçok protein kinazı vardır. Bu inhibitörler, güvenlik açısından kapsamlı bir şekilde incelendikleri ve hastaların tedavisi ve klinik denemelerde test edilmesi için daha kolay erişilebilir oldukları için yeniden kullanım için çekici bir seçenek sunar. Mers-Cov, Sars-Cov ve çeşitli viral enfeksiyonların önemli arabulucuları olarak birçok kinaz önerilmiştir. Aynı proteinlerin Sars-Cov-2 enfeksiyonunu aracılık etmekte rol oynadığı tahmin edilmektedir. Büyük ilaç firmalarının ilaç keşfi planlarının %20-%30'unu oluşturan protein kinazları, bu aşamada çok uygun bir hedef haline gelmiştir. Halihazırda onaylanmış ve farmakolojik etkileri olan birçok kinaz inhibitörü, anti-enflamatuar aktivite, sitokin baskılama ve antifibrotik aktivite gibi COVID-19'un ciddi ve potansiyel olarak yaşamı tehdit eden semptomlarını iyileştirmede faydalı olabilir. İdeal olarak, bir kinaz inhibitörü, hastalık semptomlarını baskılayarak klinik fayda sağlayabilir. Kinaz inhibitörleri, ilaç hedefleri olarak çok yaygın olarak kullanılan protein türleridir. Kinaz inhibitörleri, bir kinazın aktivitesine bağlanan ve onu inhibe eden ilaçlar oldukları için ilaç-protein afinitesi

analizinde kullanılmıştır. Şekil 4.17, kinaz inhibitörlerinin hedef protein bağlanma değerlerini göstermektedir.



2D Smiles	3D Conformer	Target Protein	Affinity
Erlotinib	Erlotinib	AAK1	5,0
			
Vandetanib	Vandetanib	AAK1	5.92081
			
Dasatinib	Dasatinib	ABL2	9.76955
			
Cediranib	Cediranib	JAK1	5,0
			
Lapatinib	Lapatinib	EGFR	8.92081
			
Linifanib	Linifanib	FLT3	7.95860
			
Pazopanib	Pazopanib	AAK1	5.53760
			
Tozasertib	Tozasertib	ABL1	8.10790
			

Şekil 4.17 Kinaz inhibitörlerinin hedef protein bağlanma değerleri

İdeal bir ilaç adayı, istenen hedef ile yüksek bağlanma afinitesine (düşük KD) sahip, ancak diğer tüm bilinen biyolojik hedeflerle düşük bağlanma afinitesine (yüksek KD) sahip bir bileşik olacaktır. Bu, ilacın diğer hedeflerle etkileşime girme riskini ve istenmeyen yan etkiler oluşturma riskini en aza indirir. Bazı ilaçların birden fazla kez ortaya çıkabileceğini görebiliriz, bu ilaçlar potansiyel olarak en umut verici tedaviler olabilir çünkü çok hedefli inhibitörler olarak hizmet edebilirler. Umud verici olan şey, birçok bulgunun zaten literatürde önerilmiş veya doğrulanmış olmasıdır. Bu teşvik edicidir çünkü sonuçlarımızı doğrular.

Önerilen yöntemde, aşama 1'de elde edilen listedeki ilaçların SMILES yapıları, Davis veri setindeki kinaz inhibitörleri ile çeşitli eğitim ayarları altında eğitilmiştir. Burada, her SMILES Davis veri setindeki kinaz inhibitörleri ile hedef-protein analizi yapılmış ve KD değeri ile tüm kinaz inhibitörleri ile bağlanma oranı her ilaç için ayrı ayrı elde edilmiştir. Diğer inhibitörlere göre daha iyi bağlanma oranına sahip kinaz inhibitörü seçilmiştir. Örneğin, Cediranib için, JAK1 inhibitörü diğerlerine göre daha yüksek bir bağlanma oranı ile 5.0 KD değeri ile sonuçlanmıştır. Benzer şekilde, Dasatinib, diğer inhibitörlere göre ABL2 kinaz inhibitörüne 9.7 KD değeri ile daha iyi bağlanmıştır. Çeşitli malignitelerin tedavisi için onaylanmış kinaz inhibitörleri, solunum virüsü enfeksiyonlarının neden olduğu akciğer hasarına bağlı yaşamı tehdit eden durumların olasılığını azaltabilecek anti-enflamatuar ve sitokin inhibitör aktivitesi gibi özelliklere sahiptir. Örneğin, Osimertinib, güçlü bir EGFR inhibitörüdür. Osimertinib'in, Sars-Cov-2'ye karşı in vitro aktivite gösteren 24 FDA onaylı ilaçtan biri olduğu bildirilmiştir. Lapatinib de önerilebilir. Bu tabloya göre, akciğer hastalığı ve Akut Miyeloid Lösemi (AML) hastalarından muzdarip olanların tedavisinde kullanılan umut verici ilaçlardır. Tüm sonuçlar arasında, en umut verici olanlar EGFR (Lapatinib), AAK1 (Erlotinib, Vandetanib, Pazopanib), ABL2 ve ABL1 (Dasatinib, Tozasertib), FLT3 (Linifanib) ve JAK1 (Cediranib) olarak öne çıkmaktadır ve COVID-19 tedavisi için alternatif olarak kullanılabilirler. Bu kinaz inhibitörlerinin, solunum virüsü enfeksiyonlarının neden olduğu akciğer hasarı ile ilişkili yaşamı tehdit eden durumların olasılığını azaltma potansiyeli vardır. Kinaz inhibitörleri, anti-enflamatuar aktivite, sitokin baskılama ve antifibrotik aktivite gibi farmakolojik özelliklere sahip olup, COVID-19 tedavisinde kullanılabilecek umut verici ilaç adaylarıdır.

Sonuç olarak, önerilen yöntem, mevcut FDA onaylı ilaçların COVID-19 tedavisinde yeniden kullanılabilirliğini belirlemek için Tanimoto ve AP benzerlik analizlerini kullanmış ve bu ilaçların proteinlerle afinitesi bağlanma modu analizini gerçekleştirmiştir. Bu çalışmada elde edilen sonuçlar, akciğer kanseri tedavisinde kullanılan bazı ilaçların COVID-19 tedavisinde de etkili olabileceğini göstermektedir. Bu yöntem, yeni terapötik stratejilerin geliştirilmesine katkıda bulunabilir ve hızlı bir şekilde uygulanabilir tedavi seçeneklerinin belirlenmesine yardımcı olabilir.

Elde edilen bulgular, özellikle EGFR, AAK1, ABL2, ABL1, FLT3 ve JAK1 gibi hedefleri inhibe eden kinaz inhibitörlerinin COVID-19 tedavisinde umut verici adaylar olduğunu göstermektedir. Bu kinaz inhibitörlerinin yeniden konumlandırılması, COVID-19 pandemisine karşı etkili bir çözüm sunabilir ve bu ilaçların hızlı bir şekilde klinik denemelere girmesine olanak tanıyabilir. Bu çalışmanın sonuçları, pandemiye karşı mücadelede stratejik bir öneme sahip olabilir ve gelecekteki araştırmalara temel oluşturabilir.

İlaç geliştirme bağlamında, mevcut ilaçların başka bir hastalık için yeniden kullanımı, yeni ilaç keşfinden daha hızlı olabilir. Mevcut ilaçlar diğer hastalıklar için onaylandığında, yeni tedaviler daha hızlı sağlanabilir, ancak onaylanmış ilaçların yeni bir endikasyon için yeniden kullanımı sırasında dikkate alınması gereken birçok faktör vardır. Potansiyel olarak inhibe edilen kilit protein hedeflerini belirlemek, yeni bir terapötik uygulama için iyi bir seçenek olabilir. Ancak, pandemi krizi sırasında ihtiyaç ve zamanın aciliyeti, bir ilaca etkinliği atfeden verilerle iyi kontrollü çalışmalar yapmayı zorlaştırabilir. Virüsle ilişkili proteinleri hedefleyen çeşitli kinaz inhibitörleri, COVID-19 için klinik araştırmalar altındadır. COVID-19 pandemisi ile mücadelede birkaç aşı mevcuttur, ancak virüsün mutant türlerinin ortaya çıkması, aşılardan üretilmesi ve dağıtılması zorlukları nedeniyle pandemi ile başa çıkmak hala zorludur. İlgili hastalıkla ilişkili ilaç hedefleri ve protein yapıları bilindiğinde, yapısal biyoenformatik kullanarak mevcut ilaçları bu bilinen hedeflere karşı taramak mümkündür. Bu bağlamda, bilinen prelinik, farmakokinetik, farmakodinamik ve yeniden konumlandırılmış ilaçların toksisite profilleri, COVID-19 hastalığına karşı hızlı tedavi bulmak için önemli bir çözüm olabilir.

Bu çalışmada belirlenen Erlotinib ilacının afinitesi AKT1 için analiz edilmiş ve COVID-19 tedavisinde antiviral ilaç olarak kullanılabileceği görülmüştür. Şiddetli akut solunum yolu sendromu Sars-Cov-2 ile enfekte olan yeni teşhis edilmiş AML hastaları için en uygun tedavi bilinmemektedir. Gilteritinib'in tek ajan olarak güvenle uygulanabileceği ve de novo FLT3-ITD pozitif AML nedeniyle başvuran hastalarda remisyonda neden olabileceği sonucuna varılmıştır. Gilteritinib, FLT3 mutasyonlu AML ve şiddetli COVID-19 hastaları için bir tedavi seçeneği olarak değerlendirilmiştir, burada uzun süreli kemoterapi kaynaklı pansitopeninin sonuçları olumsuz etkileyebileceği belirtilmiştir (Wilson et al., 2020). İlaç-hedef benzerliklerinden elde edilen Linifanib ilacının, FLT3 inhibitörü olan Gilteritinib gibi, COVID-19 ile enfekte AML hastalarında etkili olabilir.

Sars-Cov ve Mers-Cov enfeksiyonlarından iyileşen hastalarda epidermal büyüme faktörü (EGFR) ile aracılık edilen bir süreç olan pulmoner fibrozis şeklinde uzun süreli veya kalıcı akciğer hasarı gözlenmiştir. EGFR'yi hedefleyen inhibitörler, birçok solunum yolu virüsü de dahil olmak üzere farklı virüsler tarafından kullanılmış ve gefitinib ve erlotinib ile gözlenmiştir (Lupberger vd., 2011). Örneğin, Osimertinib, güçlü bir EGFR inhibitörüdür ve Sars-Cov-2'ye karşı in vitro aktivite gösteren 24 FDA onaylı ilaçtan biri olduğu bildirilmiştir. Başka bir inhibitör olan lapatinib, COVID-19 tedavisinde alternatif olarak kullanılabilir. Bu çalışmada elde edilen kinaz inhibitörleri (erlotinib, lapatinib, vandetanib, pazopanib, cediranib, dasatinib, linifanib ve tozasertib), ABL2, ABL1, EGFR, AAK1, FLT3 ve JAK1 gibi engelleyici ajanların kombinasyonu olarak veya antiviral tedaviler (ribavirin, ritonavir-lopinavir ve remdesivir) olarak COVID-19 tedavisi için kullanılabilir.

Bu bölümde, dünya çapında milyonlarca ölüme neden olan COVID-19 virüsü hakkında bilinen sınırlı bilgileri, ilişkili virüslerle ilgili verilerle tamamlayarak yeniden kullanılabilir ilaç adaylarının bir listesini oluşturmak için biyoenformatik analizler gerçekleştirilmiştir. COVID-19 hakkında sınırlı bilgi olduğu için, benzer patojenlere odaklandık ve akciğer kanseri tedavisinde kullanılan ilaçları FDA onaylı ilaçlarla karşılaştırarak benzer yeniden kullanılabilir ilaçların bir listesini hazırladık. Bu ilaçların afinitesini hesaplamak için grafik sınır ağı tabanlı GEFA modeli

kullanıldı. Bir bileşik ile bir hedef arasındaki etkileşimi tahmin etmek, hızlı ilaç yeniden kullanımı için çok önemlidir. DL, DTA problemine başarıyla uygulanmıştır. Ancak, önceki derin öğrenme tabanlı yöntemler, ilaç ve protein kalıntıları arasındaki doğrudan etkileşimlerin modellenmesini göz ardı etmiştir. Bu, ilaç bağlanma etkileri nedeniyle hedef temsiliinin yanlış öğrenilmesine yol açabilir. Ayrıca, önceki DTA yöntemleri, DTA veri kümelerinin dışındaki proteinlerin kullanımını ihmal eder. Bu yöntemler, yalnızca DTA veri kümelerinde yer alan az sayıda protein dizisine dayalı olarak protein temsiliini öğrenir. GEFA, bağlanma etkileri nedeniyle hedef temsiliindeki değişiklikleri incelemek için dikkat mekanizmasıyla grafik-içinde-grafik sinir ağı olduğundan bu çalışmada seçilmiştir.

Dolayısıyla, mevcut onaylı ilaçların kullanımı, pandemiye karşı mücadelede yeni bir ilacın keşfi için gereken uzun zamana kıyasla daha hızlı bir tedavi sağlayabilir. Hızla yayılan COVID-19 virüsü için yeniden kullanılan ilaçların kullanımı, zamanın önemi ve aciliyeti dikkate alındığında, pandemi krizinde mümkün olan en kısa sürede mücadeleye yardımcı olmayı kolaylaştırabilir.

Sonuç olarak, bu bölümde, mevcut FDA onaylı ilaçların COVID-19 tedavisinde yeniden kullanılabilirliğini değerlendirmek için Tanimoto ve AP benzerlik analizlerini kullanmış ve bu ilaçların proteinlerle afinitesini hesaplamak için GEFA modelini kullanmıştır. Bu çalışmanın sonuçları, COVID-19 pandemisine karşı hızlı ve etkili tedavi seçeneklerinin belirlenmesine katkıda bulunabilir. Ayrıca, akciğer kanseri tedavisinde kullanılan bazı ilaçların COVID-19 tedavisinde de etkili olabileceğini göstermektedir. Bu çalışma, pandemi krizinde stratejik bir öneme sahip olabilir ve gelecekteki araştırmalara temel oluşturabilir.

Tek bir FDA onaylı ilacın geliştirilmesi ortalama on yıl ve iki milyar dolar almaktadır. Ancak, COVID-19 pandemisi gibi ani küresel tehditlere karşı bu süreç çok yavaş kalmaktadır. COVID-19 aşısı zaten dağıtılmaktadır, ancak küresel aşı dağıtımı bir gün içinde gerçekleştirilemez. Aşı dağıtımının tüm zorluklarına rağmen, dünya genelinde yüz binlerce insan hala Sars-Cov-2 virüsünden mustarip olmuştur. Yeni bir ilaç keşfetmeye alternatif olarak mevcut ilaçların yeniden kullanımı düşünülebilir. Bu, diğer hastalıkların tedavisi için klinik denemelerden geçmiş ilaçları bulup, yeni

hastalığın tedavisinde etkili olabileceklerini değerlendirmek anlamına gelir. Bu şekilde, ilacın yeni bir hastalığı tedavi etme yeteneğini test etmek dışında, güvenlik riski zaten bilindiği için başka bir test gerekmez.

Bu bölümde, GSA tabanlı GEFA modeli, Sars-Cov-2 tedavisi için zaten klinik denemelerden geçmiş yeni potansiyel ilaç adaylarını bulmak için kullanılmıştır. Bu bağlamda, COVID-19 hakkında sınırlı bilgi olduğu için çalışmalarımızı benzer patojenlere odakladık ve Tanimoto ve AP gibi benzerlik algoritmalarını kullanarak akciğer kanseri tedavisinde kullanılan ilaçlarla COVID-19 tedavisinde kullanılan FDA onaylı ilaçları karşılaştırarak benzer yeniden kullanılabilir ilaçların bir listesini hazırladık. Bu ilaçların bağlanma oranını hesaplamak için grafik sinir ağı tabanlı GEFA modeli uygulanmış ve sonuçlar değerlendirilmiştir. Çalışmada kullanılan ilaç veri setleri PubChem ve DrugBank'ten alınmıştır. Sars-Cov-2 ile ilişkili semptomların (pnömoni, fibrozis ve enflamasyon dahil) gelişiminde rol oynayan ana virüsle ilişkili proteinleri hedefleyen çeşitli kinaz inhibitörleri için klinik denemeler devam etmektedir.

Çalışmada elde edilen kinaz inhibitörleri (erlotinib, lapatinib, vandetanib, pazopanib, cediranib, dasatinib, linifanib ve tozasertib), ABL2, ABL1, EGFR, AAK1, FLT3 ve JAK1 gibi engelleyici ajanların kombinasyonu olarak veya antiviral tedaviler (ribavirin, ritonavir-lopinavir ve remdesivir) olarak COVID-19 tedavisi için alternatif olarak kullanılabilir. Sonuç olarak, gelecekteki Sars-Cov-2 tedavisi için potansiyel bir hedefi temsil edebilecek ve büyük viral faktörleri hedefleyen küçük moleküller, Sars-Cov-2 enfeksiyonuna karşı yeni tedavi sınıflarının tasarımı için bir temel ve araştırma sağlayabilir. Bu sonuçlar aracılığıyla, gelecekteki tedavilere ışık tutmayı amaçlamıştır.

4.2 Bölüm 2: İlaç çözünürlük tahmini

4.2.1 İlaç çözünürlük tahmininde GSA ve diğer hesaplamalı yöntemlerin karşılaştırılması

Günümüz dünyasında ilaç keşfi ve geliştirilmesi, ilaç endüstrisinde ve tıbbi araştırmalarda en önemli alanları temsil etmektedir. Bu karmaşık süreçte, ilaç molekülünün çözünürlüğünün doğru bir şekilde tahmin edilmesi çok büyük bir önem taşımaktadır. Bir ilaç molekülünün çözünürlüğü, ilaç tasarımı ve formülasyonunda

temel bir parametre olarak kabul edilir. Çözünürlük, bir bileşiğin suda veya başka bir çözücünde ne kadar iyi çözünebileceğini belirler ve bu da ilacın farmakokinetiğini, biyoyararlanımını ve toksisitesini etkileyebilir. Bu nedenle, bir ilaç adayının geliştirme aşamasında çözünürlüğünün doğru bir şekilde tahmin edilmesi, potansiyel sorunların erken tespiti ve çözümü için gereklidir.

Yıllar içinde, ilaç çözünürlüğü tahmininde geleneksel metodolojilerden YZ ve ML temelli gelişmiş tekniklere geçilmiştir. Bu teknolojik gelişmeler, büyük miktarlarda moleküler veriye erişim sağlayarak çözünürlük tahmininde yeni yaklaşımların geliştirilmesini mümkün kılmıştır. Geleneksel yöntemler, bileşiğin kimyasal ve fiziksel özelliklerine dayalı matematiksel denklemler kullanarak ilaç çözünürlüğünü tahmin etmeye çalışırken, YZ ve ML, karmaşık moleküler etkileşimleri ve kalıpları tanıyabilen daha esnek ve veri odaklı bir yaklaşım sunmaktadır.

Bu bağlamda, GSA'lar büyük ilgi görmüştür. GSA'lar, çeşitli alanlarda ve denetimli, yarı denetimli, kendi kendine denetimli ve denetimsiz ortamlar da dahil olmak üzere çeşitli öğrenme paradigmaları altında grafik yapıları veri kümelerini işlemede başarı göstermektedir. Grafik tabanlı metodolojilerin çoğunluğu denetimsiz öğrenme alanına girmekte ve sıklıkla Otomatik kodlayıcılara, kontrastlı öğrenmeye veya rastgele yürüyüşlerle ilgili kavramlara dayanmaktadır.

GSA'lar, diziler yerine grafik yapıları üzerinde çalışmak üzere özel olarak tasarlandıkları için geleneksel sinir ağlarından farklıdır. Grafiklerin kullanımı, öncelikle ara bağlantılarla karakterize edilen karmaşık gerçek dünya problemlerini etkili bir şekilde temsil etme konusundaki olağanüstü yetenekleri nedeniyle son yıllarda önemli bir büyüme ve tanınma yaşamıştır. Bu uygulamalar, bilgilerin test senaryoları gibi yapılandırılmamış formatlarda kullanıldığı yapılandırılmış verileri içerir. Ayrıca grafikler, sosyal ağlar, moleküler yapılar, web bağlantı verileri ve daha fazlası dahil olmak üzere çeşitli alanların modellenmesinde değerli olduğunu kanıtlamış ve kapsamlı analiz ve yorumlamaya olanak sağlamıştır. GSA'lar, grafikleri görüntüler için temsiller olarak kullanarak görüntü segmentasyonu ve nesne tespiti gibi görüntü analizi görevlerinde etkinliklerini kanıtlamıştır. Sonuç olarak, GSA'ların kullanımı, grafik yapıları verilerin işlenmesinde özel bir yaklaşıma izin vererek,

birbirine bağılılığın ve yapılandırılmış ilişkilerin en önemli hususlar olduğu çeşitli alanlarda gelişmiş performans ve sonuçları kolaylaştırır.

GSA'lar, ağ gömme, grafik sınıflandırma, düğüm sınıflandırma, uzamsal-zamansal grafik tahmini ve grafik oluşturma dahil ancak bunlarla sınırlı olmamak üzere çeşitli görevler ve alanlar yelpazesinde uygulama alanı bulur. Kullanımları geniş bir faaliyet ve alan yelpazesini kapsar. Uyarlanabilirlikleri, GSA'ların ilişkisel yapılar ve bağımlılıklarla karakterize edilen karmaşık sorunların üstesinden gelmek için çok önemli araçlar olarak konumlandırır. GSA'lar moleküler yapıları temsil etmek ve analiz etmek için grafik tabanlı bir yaklaşım kullanır. Bu yaklaşım moleküler yapıları grafikler olarak temsil eder ve bu grafikleri analiz ederek çözünürlük tahminleri yapabilir. GSA'lar moleküler etkileşimlerin daha iyi anlaşılmasına katkıda bulunabilir ve ilaç tasarımını hızlandırabilir (Budak vd., 2023; Gider & Budak, 2022).

Bununla birlikte, geleneksel regresyon modelleri ve geleneksel makine öğrenimi algoritmaları bu alanda hala etkili araçlar olarak kabul edilmektedir. Özellikle Çok Katmanlı Algılayıcı (MLP) ile örneklendirilen geleneksel makine öğrenimi metodolojileri, çözünürlük tahminleri alanında etkinliklerini göstermiş ve ilaç endüstrisinde yaygın olarak benimsenmiştir (Gardner & Dorling, 1998). Ayrıca, XGBoost, Gradient Boosting (GB) ve RF gibi önde gelen makine öğrenimi algoritmaları, özellikle kapsamlı moleküler veri kümeleri üzerinde sağlam tahmin yetenekleri göstermiştir (Hu vd., 2021).

Bu bölümde, ilaç çözünürlüğü tahmininde yapay zekâ tabanlı modelleri geleneksel ML yöntemleriyle karşılaştırmayı amaçlamaktadır. Özellikle, ilaç çözünürlüğünü tahmin etmede GSA'ların ve MLP'nin performansını değerlendireceğiz. Ayrıca RF, GB ve XGBoost gibi popüler makine öğrenimi algoritmalarının bu bağlamdaki rolünü de inceleyeceğiz. Bu araştırma, ilaç tasarım ve geliştirme süreçlerinde hangi modelin en iyi performansı göstereceğine rehberlik edebilir ve gelecekteki ilaç keşif çalışmalarına katkıda bulunabilir.

İlaç çözünürlüğünün tahmini, ilaç endüstrisinde uzun süredir devam eden bir zorluk olarak kabul edilmekte ve bu alanda kapsamlı araştırmalara yol açmaktadır.

Geleneksel yöntemler, moleküler yapıdaki kimyasal gruplar, bağ uzunlukları ve atomik yükler gibi faktörleri dikkate alarak ilaç çözünürlüğünü tahmin etmek için çeşitli matematiksel denklemler ve kurallar kullanır. Ancak bu yöntemler, bireysel ilaç moleküllerinin çözünürlüğünü tahmin etmede genellikle sınırlı başarı elde etmektedir.

Son dönemlerde, çeşitli maddelerin sulu çözünürlüğünü tahmin etme gibi karmaşık bir göreve özel bir vurgu yaparak, hesaplamalı modelleme alanında önemli ilerlemeler kaydedilmiştir. Geleneksel yöntemlerin sınırlamaları nedeniyle, YZ ve ML teknikleri ilaç çözünürlüğü tahmini için daha cazip hale gelmiştir.

Bu teknikler, büyük veri kümelerini analiz etmede ve moleküler etkileşimlerdeki karmaşık kalıpları belirlemede daha fazla esneklik sunar. Özellikle, moleküler yapıları daha iyi temsil etmek ve çözünürlük tahminlerini iyileştirmek için derin öğrenme ve GSA'lar gibi YZ yöntemleri kullanılmıştır (Selvaraj vd., 2022). Moleküllerin özelliklerini tahmin etmek için Çoklu Doğrusal Regresyon, Temel Bileşen Analizi, Kısmi En Küçük Kareler, Yapay Sinir Ağı (ANN), k-NN, SVM ve RF dahil olmak üzere çeşitli tahmin teknikleri kullanılmaktadır. Bu tahminler, kimyasal yapıların özelliklerini yakalayan tanımlayıcılara dayanmaktadır (Eros vd., 2004; Kherouf vd., 2019; Sinha vd., 2023).

GSA'lar, atomları düğümler ve kimyasal bağları kenarlar olarak modelleyen grafik tabanlı bir yaklaşım kullanarak moleküler yapıları temsil eder. GSA'lar, ilaç-protein bağlanma değerlerini tahmin etmekten ilaç benzerliklerini analiz etmeye, ilaç iskelelerini çıkarmadan ilaç yan etkilerini tahmin etmeye ve daha fazlasına kadar bir dizi uygulamada etkili bir rol oynamaktadır. GSA'ların kullanımı, ilaç keşif sürecini hızlandırmak, maliyetleri azaltmak ve moleküler etkileşimlerin daha ayrıntılı bir analizine izin vererek yeni ilaç adayları keşfetmek için önemli bir araç olarak ortaya çıkmıştır. Buna karşılık, MLP gibi geleneksel regresyon modelleri moleküler özellikleri vektörler olarak temsil eder ve bu özellik vektörlerini kullanarak tahminler yapar (Qin vd., 2022).

GSA'ların ilaç çözünürlüğü tahminindeki potansiyeli son zamanlarda yapılan birkaç çalışmada araştırılmıştır. Bir araştırmada, çözünürlük tahmini için çeşitli derin

öğrenme modelleri geliştirilmiştir. Moleküler temsili kapsüllemek için dört ayrı GSA modeli varsayılmış ve bunlar arasında AttentionFP modeli performansta kayda değer bir üstünlük sergilemiştir (Ahmad vd., 2023). Eş zamanlı olarak, yenilikçi bir Çok Düzenli Grafiksel Dikkat Ağı (MoGAT), çözünürlüğün tahmini için gelişmiş bir çerçeve olarak tanıtılmıştır. Bu önerinin temel amaçları tahmin performansını artırmak ve tahmin edilen sonuçların açıklanmasını kolaylaştırmaktır. Bulgular, MoGAT'ın performans açısından çağdaş yöntemlerden daha iyi performans gösterdiğini ve tahmin edilen sonuçların yerleşik kimyasal bilgi ile uyumluluğunu ortaya koyduğunu göstermiştir (Lee vd., 2023). Farklı bir araştırmada, farmasötik bileşiklerin suda çözünürlüğünü tahmin etmek için yeni bir grafik çerçevesi önerilmiştir. Bu kavramsal çerçeve, QM9 veri kümesi için açıkça uyarlanmış ALIGSA adlı farklı bir GSA modelini tanıtmıştır (Hamdi vd., 2022).

Bu çalışmalardan elde edilen genel bulgular, GSA'ların moleküler etkileşimlerin daha incelikli bir şekilde anlaşılmasını sağlayarak çözünürlük tahminlerini geliştirme potansiyeline sahip olduğunu göstermektedir. Bununla birlikte, MLP gibi geleneksel regresyon modelleri bu alanda etkili araçlar olarak kabul edilmeye devam etmekte ve çeşitli ilaç şirketleri tarafından yaygın olarak kullanılmaktadır. Ayrıca, RF, GB ve XGBoost gibi popüler makine öğrenimi algoritmaları ilaç çözünürlüğü tahmini için kullanılmış ve başarılı sonuçlar elde edilmiştir. Bu algoritmalar, büyük moleküler veri kümeleri üzerinde tahminler yapmak için değerli araçlar olarak kabul edilir ve ilaç çözünürlüğü tahmini için gerekli görülür (M. Li vd., 2022).

Bu bölümde, ilaç çözünürlüğünün tahminine ilişkin çeşitli ML ve YZ metodolojilerinin karşılaştırmalı bir analizini yapmaya çalışmaktadır. Bu amaç doğrultusunda, ilk olarak bir ilaç verisi derlenmiş ve çeşitli modelleme stratejileri bu veri kümesinin merceğinden incelenmiştir.

İlaç veri kümesi, çeşitli ilaç moleküllerinin kimyasal özelliklerini, yapılarını ve çözünürlük değerlerini içerir. Bu veri kümesi, ilaç tasarımı ve geliştirme süreçlerinde yaygın olarak kullanılan bir veri tabanından elde edilir ve çeşitli ilaç moleküllerini kapsar. İlaç molekülü çözünürlüğünün tahmininde GSA'lar, MLP, RF, GB ve XGBoost gibi modelleri kapsayan çeşitli metodolojiler kullanılmaktadır. Bu modeller,

cinsinden çözünürlük tahminlerini temsil etmektedir. Ek olarak, her bir bileşiğin kimyasal yapısını ve moleküler özelliklerini temsil eden bağımsız değişkenler içerir. Bu moleküler özellikler çözünürlük tahminleri için kullanılan temel verileri oluşturmaktadır (KAGGLE, 2023; Sadybekov ve Katritch, 2023).

Bu veri seti, ilaçların biyoyararlanımı, farmakokinetiği ve toksisitesi gibi kritik farmasötik parametrelerin anlaşılmasına katkıda bulunmak için kullanılmaktadır. Ayrıca, YZ tabanlı regresyon modellerinin ve DL tekniklerinin ilaç tasarım süreçlerini hızlandırmak için nasıl kullanılabileceğini anlamak için kapsamlı bir şekilde araştırılmaktadır.

İlaç endüstrisi, kimyasal araştırma ve moleküler modelleme alanlarında çok önemli bir varlık olarak kabul edilen bu veri kümesi büyük önem taşımaktadır. Bu çalışmanın amacı, ESOL veri setini kullanarak çözünürlüğü tahmin etmek ve bu tahminlerin farmasötik bileşiklerin tasarımı ve geliştirilmesi için paha biçilmez bilgiler sağlayacağı beklentisidir.

ESOL veri seti 10 sütun ve 1128 satırdan oluşmaktadır. Aşağıdaki Tablo 4.5'te bu veriye ait 19 ilaç için ilgili öznitelik sütunları gösterilmektedir.

Tablo 4.5 ESOL veri seti

Bileşen	Minimu	Mol ağırlık	H-Bandı Donörler	Halkalar	Döndürülebilir	Polar yüzey	Litre başına mol	SMILES
Amigdalin	1	457,432	7	3	7	202,32	-0,77	OCC3OC(OCC2OC(OC(C#N)c1cccc1)C(O)C(O)C2O)C(O)C(O)C3O
Fenfuram	1	201,225	1	2	2	42,24	-3,30	Cc1occc1C(=O)Nc2ccccc2
citral	1	152,237	0	0	4	17,07	-2,06	CC(C)=CCCC(C)=CC(=O)
Picene	2	278,354	0	5	0	0	-7,87	c1ccc2c(c1)ccc3c2ccc4c5ccccc5cc43
Thiophene	2	84,143	0	1	0	0	-1,33	c1ccsc1
benzothiazole	2	135,191	0	2	0	12,89	-1,50	c2ccc1scnc1c2
2,2,4,6,6'-PCB	1	326,437	0	2	1	0	-7,32	Clc1cc(Cl)c(c(Cl)c1)c2c(Cl)cccc2Cl

Tablo 4.5 ESOL veri seti (Devamı)									
Estradiol	1	272,388	2	4	0	40,46	-5,03	CC12CCC3C(CCc4cc(O)ccc34)C2CCC1O	
Dieldrin	1	380,913	0	5	0	12,53	-6,29	ClC4=C(Cl)C5(Cl)C3C1CC(C2OC12)C3C4(Cl)C5(Cl)Cl	
Rotenone	1	394,423	0	5	3	63,22	-4,42	COc5cc4OCC3Oc2c1CC(Oc1ccc2C(=O)C3c4cc5OC)C(C)=CO=C1CCCN1	
2-pyrrolidone	1	85,106	1	1	0	29,1	1,07	O=C1CCCN1	
2-Chloronaphthalene	1	162,619	0	2	0	0	-4,14	Clc1ccc2ccccc2c1	
1-Pentene	1	70,135	0	0	2	0	-2,68	CCCC=C	
Primidon	1	218,256	2	2	2	58,2	-2,64	CCC1(C(=O)NCNC1=O)c2ccccc2	
Tetradecane	1	198,394	0	0	11	0	-7,96	CCCCCCCCCCCCCCC	
2-Chloropropane	1	78,542	0	0	0	0	-1,41	CC(C)Cl	
2-Methylbutanol	1	88,150	1	0	2	20,23	-0,47	CCC(C)CO	
Benzonitrile	1	103,124	0	1	0	23,79	-1,00	N#Cc1ccccc1	
Diazinon	1	304,352	0	1	7	53,47	-3,64	CCOP(=S)(OCC)Oc1cc(C)nc(n1)C(C)C	

Bu bölümde GSA'ların entegrasyonu, farmasötik ve kimya alanlarında öncü çözünürlük tahmin metodolojilerinin gelişimini katalize etme amacına hizmet etmektedir. Bu stratejik dahil etme, ilaç tasarımıda yer alan karmaşık süreçleri hızlandırmayı ve geliştirmeyi amaçlamaktadır. Sonuç olarak, GSA tabanlı bir çözünürlük tahmin modeli, ilaç geliştirme süreçlerinde değerli bir araç olarak hizmet edebilir ve ilaçların daha etkili ve güvenli bir şekilde tasarlanmasına yardımcı olabilir. GSA'lar ile ilgili olarak detaylı açıklamalar için Bölüm 3.2.1'e bakınız.

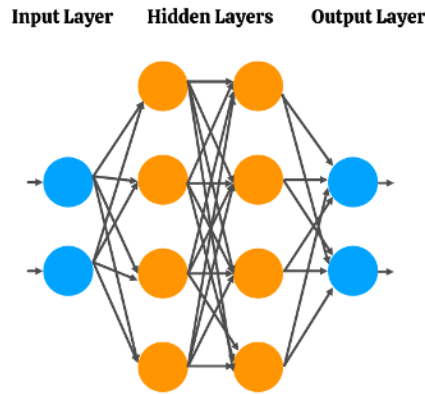
Çok Katmanlı Algılayıcı (MLP)

Bu bölüm kapsamında yapay sinir ağlarından türetilmiş bir DL modeli olarak MLP kullanılmıştır. MLP, veri madenciliği ve örüntü tanıma alanlarında yaygın olarak kullanılan bir regresyon ve sınıflandırma modelidir. Karmaşık fonksiyonları öğrenme kabiliyeti ile bilinir ve özellikle yapılandırılmış ve yapılandırılmamış veri türlerini içeren tahmin problemlerinde başarılıdır.

Çok katmanlı mimari MLP'nin ayırt edici bir özelliğidir. MLP mimarisi en az üç temel katmandan oluşur: ilk giriş katmanı, ardından bir veya daha fazla gizli katman ve son olarak çıkış katmanı. Bu katmanlardaki nöronlar tamamen birbirine bağlıdır ve hem önceki hem de sonraki katmanlardaki her nöronla bağlantı kurar. Bu yapısal özellik, modelin karmaşık işlevleri ustalıkla temsil etme ve öğrenme kapasitesini artırır (Gardner ve Dorling, 1998).

Her nöron, bir aktivasyon fonksiyonunun uygulanması yoluyla gelen verilerin işlenmesini üstlenir. Sıklıkla kullanılan aktivasyon fonksiyonları sigmoid, Rectified Linear Unit (ReLU) ve tanh'ı kapsamaktadır. Aktivasyon fonksiyonlarının seçimi, nöronların çıktılarının belirlenmesinde çok önemli bir rol oynar ve modelin doğrusal olmayan ilişkileri edinme ve kavrama kapasitesini kolaylaştırır.

MLP'nin eğitimi tipik olarak geriye yayılma adı verilen bir süreç kullanılarak gerçekleştirilir. Bu süreç, modelin tahminlerinin gerçek değerlerle karşılaştırılmasını ve bu karşılaştırmadan kaynaklanan hataların katmanlar arasında geriye doğru yayılmasını içerir. Bu, ağırlıkların güncellenmesine ve modelin eğitilmesine olanak tanır. Şekil 4.19'da bir MLP'nin yapısı gösterilmektedir.



Şekil 4.19 Bir MLP yapısı

Kullanılan bazı ML algoritmaları:

Bu bölüm, ilaç molekülü çözünürlüğünün tahmininde kullanılan çok çeşitli ML algoritmalarını kapsamaktadır. Bu algoritmalar, ilaç geliştirme süreçlerinde ve ilaç tasarımında önemli bir rol oynayan çözünürlük tahmini gibi kritik bir görevi yerine getirmek amacıyla veri madenciliği ve örüntü tanıma alanlarında kullanılmaktadır. Bu

çalışmada kullanılan bazı temel makine öğrenimi algoritmalarının kısa bir açıklaması aşağıda verilmiştir:

- Rastgele Orman (RF): RF, birden fazla karar ağacının oluşturulduğu ve birleştirildiği bir topluluk öğrenme algoritmasıdır. Her karar ağacı veri örneklerini bağımsız olarak değerlendirir ve sonuçlarını toplar. Bu yaklaşım, aşırı uyum riskini azaltırken modelin genel performansını artırır (Liu et al., 2012).
- Gradient Boosting (GB): GB, zayıf öğreniciler olarak adlandırılan basit modellerin güçlü bir model oluşturmak için birleştirildiği bir topluluk öğrenme tekniğidir. Bu yöntem, her yeni modelin öncekilerin hatalarını düzeltmeye çalıştığı yinelemeli bir süreç kullanır. GB, yüksek doğruluk elde edebilen sağlam bir regresyon modeli oluşturabilir (Friedman, 2002).
- XGBoost (Extreme Gradient Boosting): XGBoost, GB'nin bir varyasyonudur ve genellikle yüksek performansı ve ölçeklenebilirliği ile bilinir. XGBoost, özellikle yapılandırılmış veri kümelerinde ve tablo verilerinde çözünürlük tahmini gibi regresyon görevlerinde sıklıkla tercih edilmektedir (Bentéjac et al., 2021).

Bu ML algoritmaları, bu çalışmanın ilaç çözünürlüğünü tahmin etme, ilaç tasarım süreçlerinin ilerlemesine yardımcı olma ve farmasötik araştırmalara katkıda bulunma hedefinin ayrılmaz bir parçasıdır.

Bulgular 3:

Performans ölçütleri, bir modelin tahminlerinin gerçek değerlere ne kadar yakın olduğunu değerlendirmek için kullanılan kriterlerdir. Bu çalışmada kullanılan ölçütler şunlardır:

1. **Ortalama Karekök Hatası (RMSE):** RMSE, bir modelin tahminlerinin gerçek değerlerden ne kadar saptığını gösteren bir ölçüdür. RMSE hesaplanırken her bir tahmin ile gerçek değer arasındaki fark hesaplanır, bu farkların karelerinin ortalaması alınır ve son olarak bu değer karekökü alınır. RMSE, tahmin edilen değişkenle aynı birimleri paylaşır, bu da onu doğrudan yorumlanabilir hale getirir. Matematiksel formül Denklem (4.8)'de gösterilmektedir.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i^{pre} - y_i^{exp})^2}{n}} \quad (4.8)$$

2. **Ortalama Mutlak Hata (MAE):** MAE, bu farkların mutlak değerlerini kullanarak bir modelin tahminlerinin gerçek değerlerden ne ölçüde saptığını gösteren bir metrik olarak hizmet eder. Her bir tahmin ile ona karşılık gelen gerçek değer arasındaki mutlak farkların ortalamasının hesaplanmasını içerir. MAE, tahminler ve gerçek değerler arasındaki sapmaların ortalama büyüklüğü hakkında fikir verir. MAE için matematiksel ifade Denklem (4.9)'da sunulmuştur.

$$MAE = \frac{\sum_{i=1}^n |y_i^{pre} - y_i^{exp}|}{n} \quad (4.9)$$

3. **R-kare (R²):** R² bir modelin veri setindeki varyansı ne kadar iyi açıkladığını ölçer. R², 0 ile 1 arasında değişir ve modelin verilere uyum iyiliğini gösterir; yüksek değerler daha üstün bir uyumu ifade eder. R², 1'e yaklaştıkça model verileri iyi açıklamaktadır. Bununla birlikte, R² negatif de olabilir, bu da modelin verileri tahmin etmede başarısız olduğunu gösterir. Matematiksel formül Denklem (4.10)'da gösterilmektedir.

$$R^2 = \frac{\sum_{i=1}^n (y_i^{pre} - y_i^{exp})^2}{\sum_{i=1}^n (y_i^{exp} - \bar{y})^2} \quad (4.10)$$

Bu ölçütler regresyon modellerinin performansının değerlendirilmesinde yaygın olarak kullanılır. RMSE ve MAE tahminlerin doğruluğunu belirlerken R² modelin veri setine ne kadar iyi uyduğunu gösterir. Bu metriklerin kullanılması, bir modelin etkinliğini ve doğruluğunu değerlendirmek için çok önemlidir ve sonuçları anlamak ve karşılaştırmak için temel araçlar sağlar.

Analizler PYTHON ortamında 32 GB RAM ve GPU RTX3060 grafik işlemcisine sahip bir bilgisayarda gerçekleştirilmiştir. Modellerin parametreleri iterasyon değeri 50, optimizör olarak “Adam”, öğrenme oranı 0,01 olarak seçilmiştir. Veri kümesi

farklı eğitim ve test alt kümelerine bölünmüş, eğitim kümesi verilerin %80'ini, test kümesi ise kalan %20'sini oluşturmuştur. Sonraki değerlendirme ve sonuçların ölçümü yalnızca belirlenen test seti üzerinde gerçekleştirilmiştir. Çalışmanın bulguları, farklı modelleme yaklaşımlarının ilaç çözünürlüğü tahmini üzerindeki etkisini göstermektedir. Sonuçlar, GSA ve MLP'nin diğer geleneksel makine öğrenimi yöntemlerine kıyasla düşük performans gösterdiğini ortaya koymaktadır. Özellikle RMSE, MAE ve R^2 gibi performans ölçütleri, GSA ve MLP tarafından yapılan tahminlerin diğer modellere kıyasla daha fazla hata içerdiğini ortaya koymaktadır.

RF, GB ve XGBoost gibi makine öğrenmesi algoritmaları bu regresyon görevinde daha iyi sonuçlar elde etmiştir. Bu sonuçlar, geleneksel makine öğrenimi yöntemlerinin GSA ve MLP gibi yapay zekâ tabanlı yaklaşımlara kıyasla ilaç çözünürlüğü tahmini için daha etkili olduğunu göstermektedir.

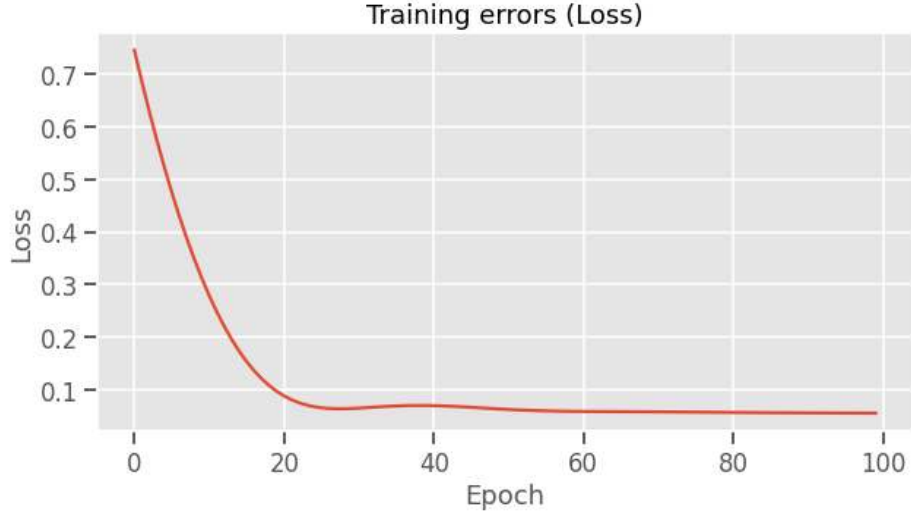
Bununla birlikte, sonuçların veri kümesi kalitesi, özellik mühendisliği, hiper parametre ayarı ve diğer hususlar dahil olmak üzere çok sayıda faktöre bağlı olarak değişebileceğini belirtmek önemlidir. Bu nedenle, ilaç çözünürlüğü tahminini iyileştirmek için farklı model yapılarının ve veri özelliklerinin daha fazla araştırılması ve dikkate alınması gerekebilir.

Tablo 4.6 Modellerin çözünürlük tahmin hata miktarları

Modeller	RMSE	MAE	R^2
Random Forest	1.2145	0.9221	0.6575
Gradient Boosting	1.2955	1.0002	0.6102
MLP	1.1473	0.8586	0.6943
GSA	1.8389	1.4684	0.2147
Random Forest	1.2145	0.9221	0.6575

Tablo 4.6'daki sonuçlar, RF ve MLP'nin ilaç çözünürlüğü tahmininde diğer modellere kıyasla üstün performans sergilediğini göstermektedir. Daha düşük RMSE ve MAE değerlerinin yanı sıra daha yüksek R^2 puanları, bu modellerin ilaç çözünürlüğünü tahmin etmede daha iyi performans gösterdiğine işaret etmektedir. Özellikle, MLP en düşük RMSE ve en yüksek R^2 puanını elde etmiştir.

Buna karşılık, GSA modeli diğer modellere kıyasla daha düşük performans göstermiştir. Yüksek RMSE ve MAE değerlerinin yanı sıra düşük R^2 puanı, GSA modelinin ilaç çözünürlüğü tahmininde daha fazla hata içerdiğini ve diğer modellere kıyasla daha az başarılı olduğunu göstermektedir.

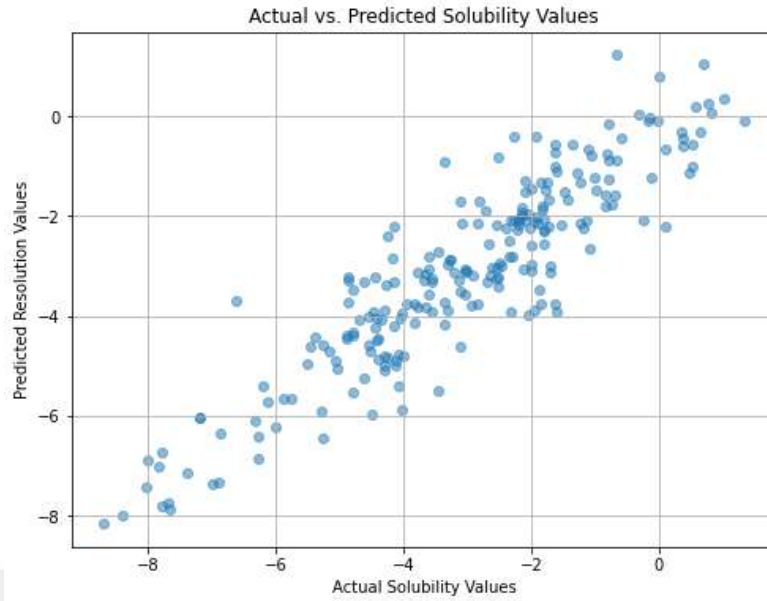


Şekil 4.20 Eğitim sırasındaki kayıp değerleri

Şekil 4.20 eğitim sırasındaki kayıp değerlerini göstermektedir. 100 iterasyondan sonra eğitim kaybı değeri 0,082'de tamamlanmıştır. Eğitilmiş bir GSA modeli yeni moleküllerin çözünürlüğünü tahmin etmek için kullanılabilir.

Tablo 4.7, RF modelinin diğer modellere kıyasla en düşük hata oranlarına sahip olduğunu göstermektedir. Bu da RF modelinin çözünürlüğü daha doğru tahmin ettiğini ve gerçek değerlere daha yakın sonuçlar verdiğini göstermektedir. Öte yandan, GSA modeli diğer modellere kıyasla daha yüksek hata oranları sergilemiştir, bu da bu özel veri kümesi için daha az doğru tahminler yaptığını ve iyileştirilmesi gerektiğini göstermektedir.

Bu grafik, her bir modelin performansının net bir şekilde karşılaştırılmasına yardımcı olur ve ilaç çözünürlüğü tahmininde hangi modelin mükemmel olduğunu ve hangisinin iyileştirme için daha fazla alana sahip olduğunu anlamamıza yardımcı olur. Bu bilgi, ilaç tasarımı ve geliştirme süreçlerinde uygun modelin seçilmesine rehberlik eder.



Şekil 4.21 Gerçek ve tahmin edilen çözünürlük değerlerinin grafiksel gösterimi

Gerçek ve tahmin edilen çözünürlük grafikleri, ilaç çözünürlüğü tahmininde kullanılan farklı modellerin performansını değerlendirmek için temel araçlardır. Şekil 4.21, her bir modelin gerçek çözünürlüğü ne kadar iyi tahmin ettiğini ve tahminleri ile gerçek değerler arasındaki sapmanın boyutunu görsel olarak temsil etmektedir.

Bu bölümde, ilaç çözünürlüğünü tahmin etmede farklı ML algoritmalarının performansını değerlendirmek için yapılmıştır. Çözünürlük tahmini, ilaç geliştirme süreçlerinde ve farmasötik tasarımda kritik öneme sahiptir. Bu nedenle, doğru ve güvenilir çözünürlük tahmin modelleri ilaç endüstrisinin başarısında önemli bir rol oynamaktadır.

Çalışmanın başlangıcında ESOL veri kümesi kullanılmıştır. Bu veri kümesi, çeşitli ilaç moleküllerinin deneysel olarak ölçülen çözünürlük değerlerini kapsamaktadır. Veri kümesinin ön işleme ve hazırlık aşamasında, moleküler grafik özellikleri çıkarılmış ve veriler eğitim ve test kümelerine ayrılmıştır.

Çalışmanın ilk aşamasında GSA modeli kullanılmıştır. GSA, ilaç çözünürlüğü tahmini için kullanılan YZ tabanlı bir yaklaşımdır. Ancak, bu çalışmada GSA modelinin performansı diğer ML modelleri ile karşılaştırıldığında daha düşük olduğu görülmüştür. RMSE, MAE ve R^2 gibi performans ölçütleri incelendiğinde, GSA

modelinin tahminlerinin diğer modellere kıyasla daha fazla hata içerdiği görülmüştür. Bu sonuçlar, GSA modelinin bu özel veri kümesinde çözünürlük tahmini için etkili olmadığını göstermektedir.

İkinci aşamada MLP modeli ve geleneksel ML algoritmaları (RF, GB, XGBoost) kullanılmıştır. Bu modeller regresyon görevleri için yaygın olarak kullanılan makine öğrenimi yaklaşımlarıdır. Sonuçlar, bu geleneksel ML algoritmalarının çözünürlük tahmininde daha yüksek başarı elde ettiğini göstermektedir. Özellikle, RF, GB ve XGBoost gibi algoritmalar daha düşük RMSE ve MAE değerleri ve daha yüksek R^2 puanları ile daha iyi performans göstererek daha doğru tahminlerde bulunmuştur.

Bu bulgular, ilaç çözünürlüğü tahmini gibi regresyon görevlerinde geleneksel makine öğrenimi algoritmalarının YZ tabanlı yaklaşımlardan (GSA gibi) daha etkili olabileceğini göstermektedir. Ancak bu sonuçlar veri kümesi, özellik mühendisliği, hiper parametre ayarı ve model seçimi gibi faktörlere bağlı olarak değişebilir. Bu nedenle, ilaç çözünürlüğü tahmini için daha fazla araştırma yapılması ve farklı model yapılarının ve veri özelliklerinin dikkate alınması gerekebilir.

Sonuç olarak, bu çalışma ilaç çözünürlüğü tahmininde kullanılan makine öğrenimi ve YZ modellerinin karşılaştırmalı bir analizini sunmaktadır. Geleneksel ML algoritmalarının belirli bir veri kümesi için daha iyi performans gösterebileceğini ve model seçiminde dikkate alınması gerektiğini vurgulamaktadır. Doğru çözünürlük tahminleri, farmasötik geliştirme süreçlerinde yeni ilaçların tasarımı ve keşfinde önemli bir avantaj sağlayabilir.

4.2.2 Grafik Sinir Ağları ile İlaç Molekülü Çözünürlük Tahmini

Farmasötik endüstrisi, ilaç keşfi ve geliştirme süreçlerinin karmaşıklığı, yüksek maliyetleri ve uzun zaman çerçeveleri nedeniyle önemli bir dönüşüm süreci geçirmektedir. Bu dönüşümün merkezinde, YZ ve ML gibi yenilikçi teknolojilerin kullanımı yer almaktadır. Bu teknolojiler, büyük miktarda veriyi hızlı bir şekilde analiz etme, kalıpları tanımlama ve tahmin yapma yetenekleri ile öne çıkar. Özellikle ilaç keşfi ve geliştirme süreçlerinde, moleküler tasarımın optimize edilmesi, yeni ilaç

adaylarının belirlenmesi ve biyolojik etkinliğin tahmin edilmesi gibi kritik aşamalarda YZ ve ML'nin kullanımı giderek artmaktadır.

GSA'lar, moleküler veri analizinde son yıllarda önemli bir ilgi odağı haline gelmiştir. GSA'lar, moleküler yapıları grafikler olarak temsil eden ve bu yapıları analiz etmek için özel olarak tasarlanmış yapay sinir ağlarıdır. Bu ağlar, moleküler özellikleri ve kimyasal bağları daha etkili bir şekilde modellemek ve anlamak için geliştirilmiştir. Moleküler yapıların bu şekilde temsil edilmesi, moleküler etkileşimleri ve yapı-aktivite ilişkilerini daha iyi anlama potansiyeli sunar. Dolayısıyla, GSA'ların ilaç moleküllerinin özelliklerini ve davranışını anlamak için etkili bir araç olduğu düşünülmektedir.

Bu çalışmanın amacı, grafik sinir ağlarının ilaç moleküllerinin çözünürlüğünün tahmin edilmesindeki etkinliğini incelemektir. İlaç moleküllerinin çözünürlüğü, ilaç keşfi ve geliştirme süreçlerindeki biyoyararlanım, toksisite ve farmakokinetik özelliklerin belirlenmesinde kritik bir faktördür (Lipinski vd., 1997). GSA'ların ilaç moleküllerinin çözünürlüğünü tahmin etme yeteneği, geleneksel yöntemlere kıyasla daha hassas sonuçlar elde etme potansiyeli taşımaktadır.

Bu çalışma, öncelikle GSA'ların ilaç moleküllerinin çözünürlüğünün tahminindeki başarısını inceleyecek ve bu yaklaşımın ilaç keşfi ve moleküler tasarım alanlarına olan katkılarını değerlendirecektir. Ayrıca, GSA'ların ilaç çözünürlüğü tahminindeki potansiyelini daha geniş bir moleküler özellik analizi bağlamında ele alacaktır. Bu çalışmanın sonuçları, YZ ve ML'nin ilaç keşfi alanındaki kullanımını genişletecek ve gelecekteki farmasötik araştırmalar için rehberlik edecektir.

Veri seti 4:

Bu bölümde, ilaç moleküllerinin çözünürlüğünün tahmin edilmesi amacıyla kullanılan veri kaynağı "Aqueous Solubility Dataset" olarak adlandırılan bir veri kümesidir (KAGGLE, 2023; Rosace vd., 2023). Bu veri kümesi, yaklaşık 9000 farklı molekülün çözünürlük verilerini içermektedir ve ilaç tasarımı alanında önemli bir referans noktası sunmaktadır.

Veri kümesi, çeşitli kimyasal bileşiklerin sulu çözünürlük değerlerini içermektedir. Her bir molekülün çözünürlüğü, çeşitli deneyler ve testler sonucunda elde edilen verilere dayanmaktadır. Bu çözünürlük değerleri, ilgili moleküllerin kimyasal yapısına, fiziksel özelliklerine ve çeşitli etkenlere bağlı olarak elde edilmiştir.

Veri kümesinin kullanımı, grafik sınır ağlarının ilaç moleküllerinin çözünürlüğünü tahmin etme yeteneğini değerlendirmek ve bu tahminlerin gerçek dünya verileri ile ne kadar uyumlu olduğunu anlamak amacını taşımaktadır. Ayrıca, bu veri kümesi sayesinde geliştirilen model ve yöntemlerin genel ilaç keşfi ve tasarım alanındaki uygulanabilirliği de test edilebilir.

Veri kümesinin paylaşılması, ilaç moleküllerinin çözünürlüğünün tahmin edilmesi alanında çalışan araştırmacılar ve ilgili endüstri uzmanları için değerli bir kaynak oluşturmaktadır. Bu veri kümesi sayesinde, ilaç tasarımının optimize edilmesi ve ilaç moleküllerinin biyolojik etkinliğinin tahmin edilmesi gibi önemli alanlarda daha hassas ve güvenilir sonuçlar elde edilmesine olanak sağlanmaktadır.

Önerilen yöntem

GSA modeli, moleküler grafiklerin analizi ve özellik çıkarımı için özel olarak tasarlanmıştır. GSA modeli, düğümlerin ve kenarların özelliklerini ve etkileşimlerini göz önünde bulundurarak, ilaçların proteinlere bağlanma afinitesini tahmin etme yeteneğine sahiptir. Bu çalışmada kullanılan GSA ile ilgili detaylı açıklamalar ve matematiksel ifadeler Bölüm 3.2.1'de yer almaktadır.

Bulgular 4:

GSA tabanlı çözünürlük tahmin modelleri, geleneksel yöntemlerle kıyaslandığında genellikle daha yüksek doğruluk ve tahmin yeteneği sunar. Bu modeller, moleküler yapının kompleks etkileşimlerini yakalamada daha etkilidir ve bu da çözünürlük tahminlerini iyileştirir. Başarım değerleri genellikle Doğruluk, precision, recall, AUC (Area Under the Curve) gibi metriklerle ölçülür. Bu metrikler, modelin gerçek verilere ne kadar yakın tahminler yaptığını gösterir. ROC-AUC eğrisi ise sol üst köşeye ne kadar yakınsa, testin genel doğruluğu da o kadar yüksek olur (Zweig ve Campbell, 1993).

Başarım performansını değerlendirmek amacıyla doğruluk, kesinlik ve duyarlılık gibi istatistiksel parametreler kullanılmıştır. Bu parametrelere ilişkin tanımlar aşağıdaki denklem (4.11-4.15)'te verilmiştir (Brown, 2018; Hand, 2009).

$$\text{Doğruluk} = \frac{TP + TN}{TP + FP + TN + FN} \quad (4.11)$$

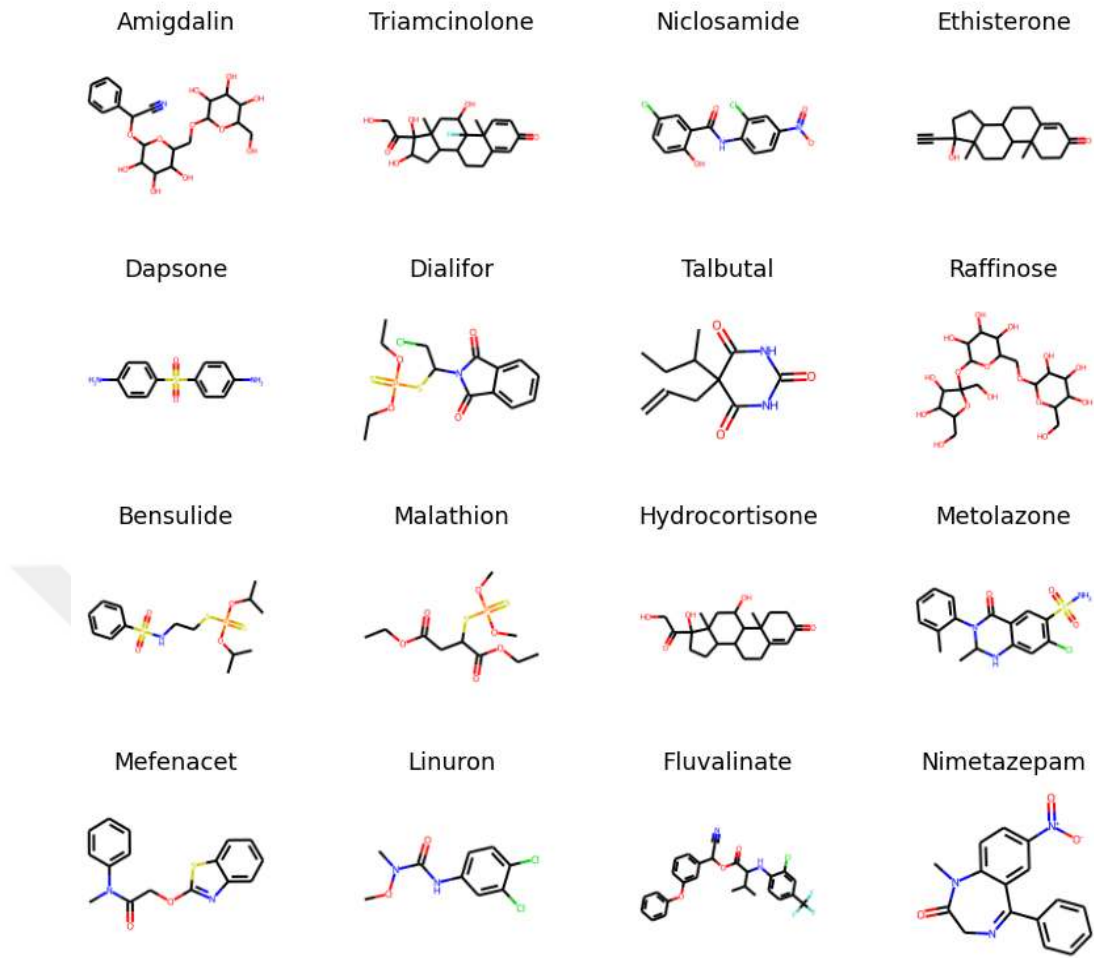
$$\text{Kesinlik} = \frac{TP}{TP + FP} \quad (4.12)$$

$$\text{Hatırlama} = \frac{TP}{TP + FN} \quad (4.13)$$

$$\text{TPR} = \frac{TP}{TP + NP} \quad (4.14)$$

$$\text{FPR} = \frac{FP}{FP + TN} \quad (4.15)$$

- Gerçek Pozitifler (TP): Bunlar modelin pozitif sınıfı doğru tahmin ettiği ve gerçek sınıfın da pozitif olduğu durumlardır.
- Yanlış Pozitifler (FP): Modelin pozitif sınıfı yanlış tahmin ettiği, ancak gerçek sınıfın aslında negatif olduğu durumlardır.
- Gerçek Negatifler (TN): Bunlar modelin negatif sınıfı doğru tahmin ettiği ve gerçek sınıfın da negatif olduğu durumlardır.
- Yanlış Negatifler (FN): Bunlar modelin negatif sınıfı yanlış tahmin ettiği ancak gerçek sınıfın aslında pozitif olduğu durumlardır.



Şekil 4.22 Veri setinde yer alan ilaç moleküllerinden seçilen 16 ilaç

Şekil 4.22, veri seti içerisinde rastgele seçilen 16 farklı ilaç molekülünün 2B yapılarını göstermektedir.

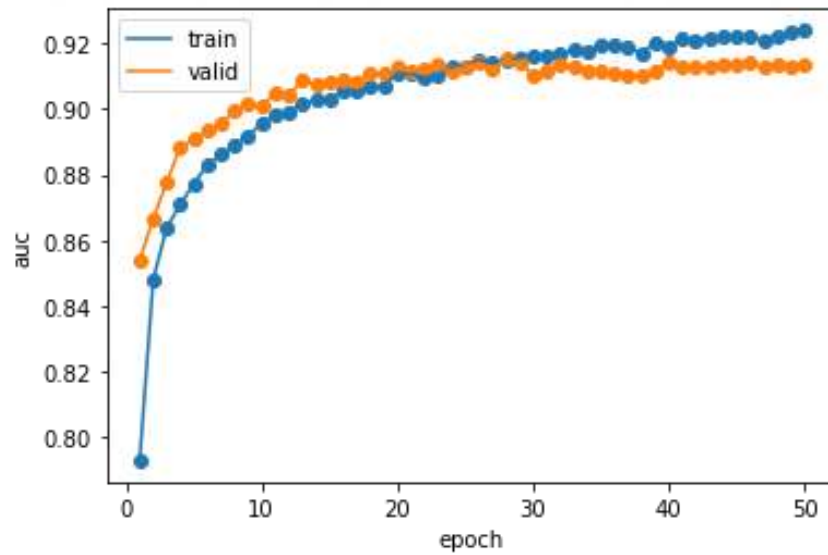
Tablo 4.7 Analiz sonucu tahmin edilen çözünürlük değerleri

SMILES	Tahmin edilen çözünürlük	Ölçülen çözünürlük	Mol Ağırlığı	isSouble
<chem>OCC3OC(OCC2OC(OC(C#N)c1ccccc1)C(O)C(O)C2O)C(O)C(O)C3O</chem>	-0,974	-0,77	457,432	1
<chem>Cc1occc1C(=O)Nc2ccccc2</chem>	-2,885	-3,3	201,225	1
<chem>CC(C)=CCCC(C)=CC(=O)</chem>	-2,579	-2,06	152,237	1
<chem>c1ccc2c(c1)ccc3c2ccc4c5ccccc5ccc43</chem>	-6,618	-7,87	278,354	1
<chem>c1ccsc1</chem>	-2,232	-1,33	84,143	1
<chem>c2ccc1scnc1c2</chem>	-2,733	-1,5	135,191	0

Tablo 4.7 Analiz sonucu tahmin edilen çözünürlük değerleri (Devamı)

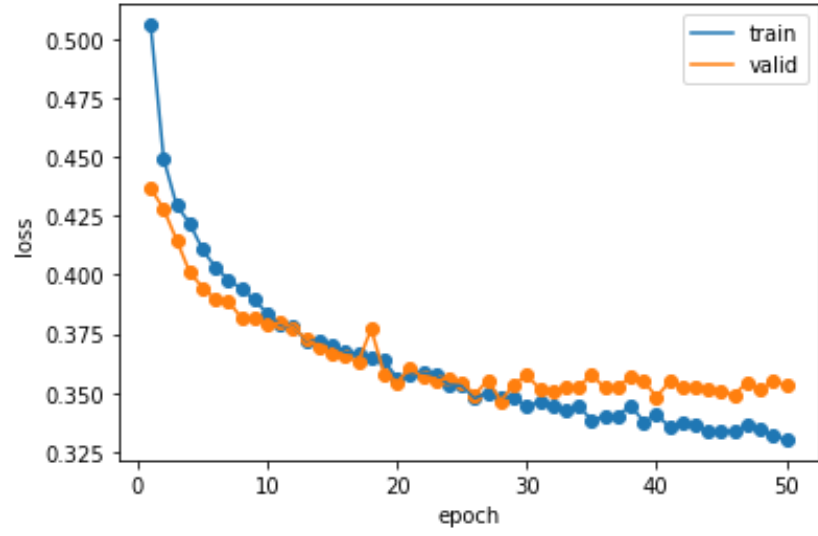
<chem>Clc1cc(Cl)c(c(Cl)c1)c2c(Cl)cccc2Cl</chem>	-6,545	-7,32	326,437	0
<chem>CC12CCC3C(CCC4cc(O)ccc34)C2CCC1O</chem>	-4,138	-5,03	272,388	0
<chem>ClC4=C(Cl)C5(Cl)C3C1CC(C2OC12)C3C4(Cl)C5(Cl)Cl</chem>	-4,533	-6,29	380,913	0
<chem>COc5cc4OCC3Oc2c1CC(Oc1ccc2C(=O)C3c4cc5OC)C(C)=C</chem>	-5,246	-4,42	394,423	1
<chem>O=C1CCCN1</chem>	0,243	1,07	85,106	1
<chem>Clc1ccc2ccccc2c1</chem>	-4,063	-4,14	162,619	0
<chem>CCCC=C</chem>	-2,01	-2,68	70,135	0
<chem>CCC1(C(=O)NCNC1=O)c2ccccc2</chem>	-1,897	-2,64	218,256	1
<chem>CCCCCCCCCCCCC</chem>	-5,45	-7,96	198,394	1
<chem>CC(C)Cl</chem>	-1,585	-1,41	78,542	1

Tablo 4.7, veri setinden yer alan 9000 ilaç arasından rastgele seçilmiş 16 farklı ilaç molekülünün SMILES, tahmin edilen çözünürlük değerleri, ölçülen çözünürlük değerleri, moleküler ağırlık ve isSouble (çözünme durumu) değerlerini göstermektedir.



Şekil 4.23 Eğitim sırasındaki elde edilen AUC sonuçları

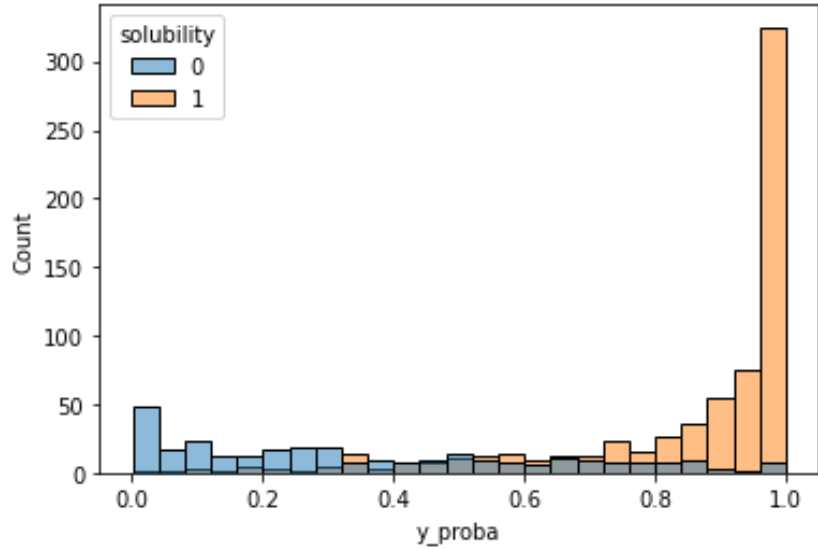
Çalışmada kullanılan veri setinin %70'i eğitim, %30'u ise test verisi olarak ayrılmıştır. 50 iterasyon ile model analiz edilmiştir. Şekil 4.23'te 50 iterasyon sonucunda eğitim verileri ile tahmin verilerinin AUC grafiği gösterilmektedir.



Şekil 4.24 Eğitim sırasındaki elde edilen kayıplar

Şekil 4.24'te eğitim sırasındaki kayıpları gösteren grafik yer almaktadır. 50 iterasyon sonucunda eğitim kayıp değeri 0.335 olarak tamamlanmaktadır.

Eğitilmiş bir GSA modeli, yeni moleküllerin çözünürlüğünü tahmin etmek için kullanılabilir. Bu, ilaç keşfi süreçlerini hızlandırmayı ve ilaç tasarımını optimize etmeyi destekler.

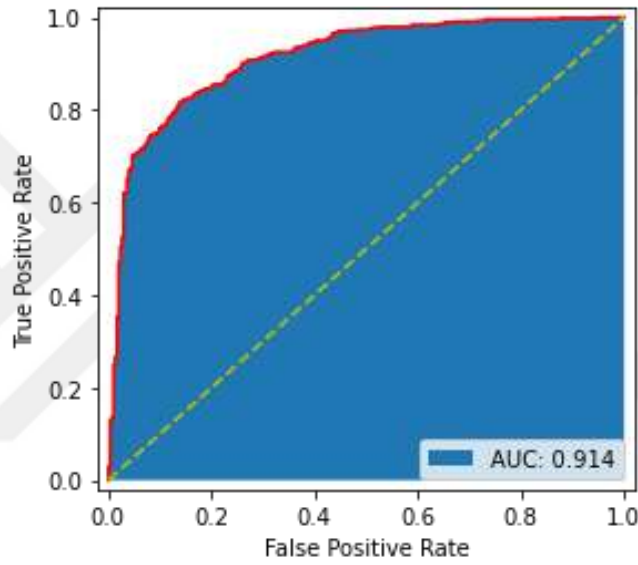


Şekil 4.25 Çözünürlük diyagramı

Şekil 4.25, ilaç moleküllerine ait çözünürlük diyagram grafiğini göstermektedir. Moleküllerin çözünürlükleri, yapısal özelliklerine, kimyasal bağlantılara ve moleküler etkileşimlere bağlı olarak değişebilir. Diyagramdaki 1 değerleri, çözünürlüğü yüksek

olan moleküllerin ve potansiyel ilaç adaylarının hangi yapısal özelliklere sahip olduğunu anlamamızı sağlar.

Çözünürlük diyagramında sıklıkla 1 ve 0 değerleri kullanılır. Bu değerler, bir molekülün çözünürlüğünün varlığını veya yokluğunu ifade eder. Eğer bir molekül çözücüde çözünebiliyorsa (örneğin, suda çözünüyorsa), o zaman genellikle 1 ile temsil edilir. Eğer çözilemiyorsa (örneğin, suda çözünmüyorsa), o zaman 0 ile temsil edilir. Bu ikili yapının kullanılması, genellikle çözünürlük verilerinin nitel bir değerlendirmesini yapmayı kolaylaştırır.



Şekil 4.26 AUC sonuçları

AUC, özellikle ikili sınıflandırma modellerinin performansını değerlendirmek için yaygın olarak kullanılan bir performans metriğidir. AUC'un kullanılması sulu çözünürlük tahmini gibi regresyon problemleriyle biraz farklıdır, çünkü AUC genellikle sınıflandırma problemlerinde kullanılan bir metriktir. Ancak, eğer sulu çözünürlük tahmini problemi ikili bir problem gibi ele alınıyorsa (örneğin, çözünürlüğü belirli bir eşiğin üstünde veya altında mı?) ve 1 ve 0 değerleri ile çalışılıyorsa, AUC değeri kullanılabilir. Sınıflandırma problemlerinde olduğu gibi, AUC değeri yine modelin yanlış sınıflandırma yapma olasılığını doğru sınıflandırma yapma olasılığına kıyasla özetler. Eğer sınıflandırma eşiği uygun bir şekilde seçilirse, AUC değeri 0 ile 1 arasında bir değer alır. 1'e yaklaştıkça modelin daha iyi performans

sergilediği anlamına gelirken, 0.5'e yaklaştıkça rastgele tahmin seviyesine yaklaştığını gösterir. Şekil 4.26'da 1 seviyesine ne kadar yaklaştığı görülmektedir.

		True label	
		0	1
Predicted label	0	208	53
	1	97	628

Şekil 4.27 Confusion matrisi

Confusion matrisi, modelin hangi tür hatalar yaptığını anlamak ve performansını değerlendirmek için kullanılmaktadır. Özellikle dengesiz sınıf dağılımına sahip veri setlerinde, modelin yanlış sınıflandırmalarının nedenlerini belirlemek ve geliştirmek için önemli bir araçtır.

Temelde dört farklı terimi içerir: true positive (TP), true negative (TN), false positive (FP) ve false negative (FN). Bu terimler, farklı yönleriyle modelin performansını açıklar. Şekil 4.27'de tahmin edilen ve gerçek değerlerin başarımlarını göstermektedir.

Tablo 4.8, çalışmada elde edilen model performans metriklerini detaylı bir şekilde göstermektedir. Bu tabloda, doğruluk, duyarlılık, özgüllük, pozitif öngörü değeri, negatif öngörü değeri, kesinlik ve hatırlama gibi önemli performans göstergeleri yer almaktadır.

- **Doğruluk (Accuracy):** Modelin tüm sınıflandırma görevlerindeki doğruluğunu temsil eder ve 0,87 olarak hesaplanmıştır.
- **Duyarlılık (Sensitivity):** Modelin pozitif örnekleri doğru bir şekilde tanımlama yeteneğini gösterir ve 0,93 olarak hesaplanmıştır.
- **Özgüllük (Specificity):** Modelin negatif örnekleri doğru bir şekilde tanımlama yeteneğini gösterir ve 0,72 olarak hesaplanmıştır.
- **Pozitif Öngörü Değeri (Positive Predictive Value, PPV):** Pozitif olarak tahmin edilen örneklerin gerçekten pozitif olma olasılığını gösterir ve 0,88 olarak hesaplanmıştır.
- **Negatif Öngörü Değeri (Negative Predictive Value, NPV):** Negatif olarak tahmin edilen örneklerin gerçekten negatif olma olasılığını gösterir ve 0,83 olarak hesaplanmıştır.
- **Kesinlik (Precision):** Modelin pozitif tahminlerinin doğruluğunu gösterir ve 0,88 olarak hesaplanmıştır.
- **Hatırlama (Recall):** Modelin tüm pozitif örnekleri tanımlama oranını gösterir ve 0,91 olarak hesaplanmıştır.

Bu performans metrikleri, modelin genel başarısını değerlendirmek için kritik öneme sahiptir ve modelin farklı koşullardaki etkinliğini anlamak için kullanılmaktadır.

Tablo 4.8 Modelin performans metriklerinin sonuçları

Performans metriği	Sonuç
Doğruluk (Accuracy)	0.87
Duyarlılık (Sensitivity)	0.93
Özgüllük (Specificity)	0.72
Pozitif Öngörü Değeri (PPV)	0.88
Negatif Öngörü Değeri (NPV)	0.83
Kesinlik (Precision)	0.88
Hatırlama (Recall)	0.91

Bu bölümde, GSA'ların ilaç moleküllerinin çözünürlüğünün tahminindeki potansiyelinin incelenmesi amaçlanmıştır. Elde edilen sonuçlar ve yapılan analizler,

GSA'ların ilaç moleküllerinin çözünürlüğü tahmininde etkili bir araç olabileceğini göstermektedir.

GSA'lar, moleküler yapının grafiksel temsili ve komşuluk ilişkilerinin dikkate alınması sayesinde ilaç moleküllerinin çözünürlüğünü daha doğru bir şekilde tahmin etme yeteneği sunar. Bu özelliği, moleküler yapının kompleks etkileşimlerini ve yapısını anlamamıza yardımcı olabilir. Elde edilen sonuçlar, geleneksel yöntemlere kıyasla daha hassas ve güvenilir çözünürlük tahminleri sağlayabileceğini göstermektedir.

Ayrıca, GSA'ların ilaç tasarımı ve moleküler keşif alanlarında da büyük potansiyel taşıdığı görülmüştür. Bu yaklaşım, ilaç adaylarının özelliklerini daha iyi anlama, biyoyararlanım ve toksisite profillerini daha iyi öngörme imkânı sunabilir. GSA'ların kullanımıyla, ilaç keşfi süreçlerinin hızlandırılması ve daha etkili ilaç adayları belirlenmesi sağlanabilir.

Ancak, bu çalışmanın sınırlamaları da göz önünde bulundurulmalıdır. GSA'ların başarıları, veri kalitesi, modelin karmaşıklığı, hiper parametre ayarları gibi faktörlere bağlıdır. Ayrıca, GSA'ların moleküler çözünürlük tahmininde tam anlamıyla geleneksel yöntemleri geçip geçmediğini değerlendirmek için daha fazla karşılaştırmalı çalışmalara ihtiyaç vardır.

Sonuç olarak, bu çalışma, GSA'ların ilaç moleküllerinin çözünürlüğü tahmininde potansiyelini açığa çıkarmış ve ilaç tasarımı alanındaki ilerlemeler için yeni bir yaklaşım sunmuştur. Bu yaklaşımın ilaç keşfi ve geliştirme süreçlerine katkı sağlama potansiyeli büyüktür ve gelecekteki araştırmalarda daha da derinlemesine incelenmelidir.

4.3 Bölüm 3: Moleküler modelleme

4.3.1 Grafik sinir ağıları tabanlı moleküler modelleme ile ilaç-protein etkileşimlerinin analizi ve ilaç keşfi: Chlorpromazine-6LU7 projesi

SARS-CoV-2 koronavirüsün neden olduğu COVID-19, küresel sağlık tehdidi oluşturan ciddi bir solunum yolu hastalığıdır. Pandeminin aciliyeti, hızlı yayılımı ve

ciddi sađlık etkisi nedeniyle tedavi bulma abalarının artmasına yol amıřtır. İyileřtirici tedavi arayıřını hızlandırmak iin, birok ila keřif grubu, COVID-19'a karřı potansiyel etkinlik iin mevcut ilaları arařtırarak ila yeniden kullanım stratejilerini benimsemiřtir. Bu yaklařım, tedavi geliřimini hızlandırmayı ve zaman alan yeni ila keřfetme srecini atlatmayı amalamaktadır. (Budak vd., 2023).

Mevcut antiviral ilalar COVID-19'u ynetmede bazı faydalar gstermektedir, ancak sınırlamaları vardır ve uzun ařı geliřtirme srecinde zorluklarla karřılařırlar. Bu nedenle, alternatif tedavi yaklařımlarını keřfetmek ve yeni ila adayları bulmak ok nemlidir. Yeniden kullanılabilir ilalar, bilinen farmakokinetik standartlara ve ilgili hastalıklarda insan kullanımı iin nceden FDA onayına sahip ilalardır (FDA, 2023). Bu ilaları yeni endikasyonlar iin klinik alıřmalara dođrudan dahil ederek, SARS-CoV-2 enfeksiyonu iin teraptiklerin geliřtirilmesini hızlandırma ve COVID-19 tedavileri iin FDA onayını kolaylařtırma potansiyeli vardır.

Son zamanlarda COVID-19 iin ila keřfi ve tedavi seenekleri kapsamlı bir řekilde arařtırılmıřtır. alıřmalar, remdesivir ve favipiravir gibi mevcut antiviral ilaların COVID-19 tedavisindeki potansiyelini vurgulamıřtır (Baby vd., 2021). Ayrıca, immnosupresan deksametazon yeniden tasarlanarak sitokin fırtınasını etkili bir řekilde azaltmıř ve COVID-19 hastalarında lm oranlarını dřrmřtr. Ritonavir ve nelfinavir gibi HIV ilalarının, Mpro proteazı hedefleyerek COVID-19'da yeniden tasarlanması nerilmiřtir (Sang vd., 2020). İla yeniden kullanımı umut vaat etse de hedef seimi ve dođrulama sreteki nemli adımlar olmaya devam etmektedir. Bazı yeniden tasarlanmış ilaların, dikkatle deđerlendirilmesi gereken yan etkiler ve potansiyel ila etkileřimleri ile ilgili sınırlamaları olabilir.

Yeni ila adaylarının belirlenmesi ve karakterizasyonu, molekler dzeyde analizler ve hesaplamalı modelleme teknikleri ile desteklenen hedef seimi ve dođrulama gibi nemli sreleri ierir. Bu yaklařımlar, ilacın hedef proteinlere bađlanma afinitesi, molekler etkileřimler ve inhibitr etkiler hakkında deđerli bilgiler sunar. Sonu olarak, ila moleklleri ve hedef proteinler arasındaki etkileřim mekanizmalarını anlamak ve analiz etmek, ila keřif srecinde ok nemli adımlardır. Geleneksel olarak, molekler etkileřim analizleri molekler modelleme ve simlasyon

yöntemlerine dayanmaktadır. Ancak, moleküler yapıların ve etkileşim ağlarının karmaşıklığı nedeniyle, daha sağlam ve etkili bir yaklaşım gereklidir. YZ ve ML yaklaşımları, modern ilaç keşfi araştırmalarında giderek daha önemli hale gelmektedir. Bu teknolojiler, geniş veri kümelerini analiz edebildikleri ve karmaşık moleküler etkileşimleri kavrayabildikleri için ilaç geliştirmede yeni fırsatlar sunmaktadır.

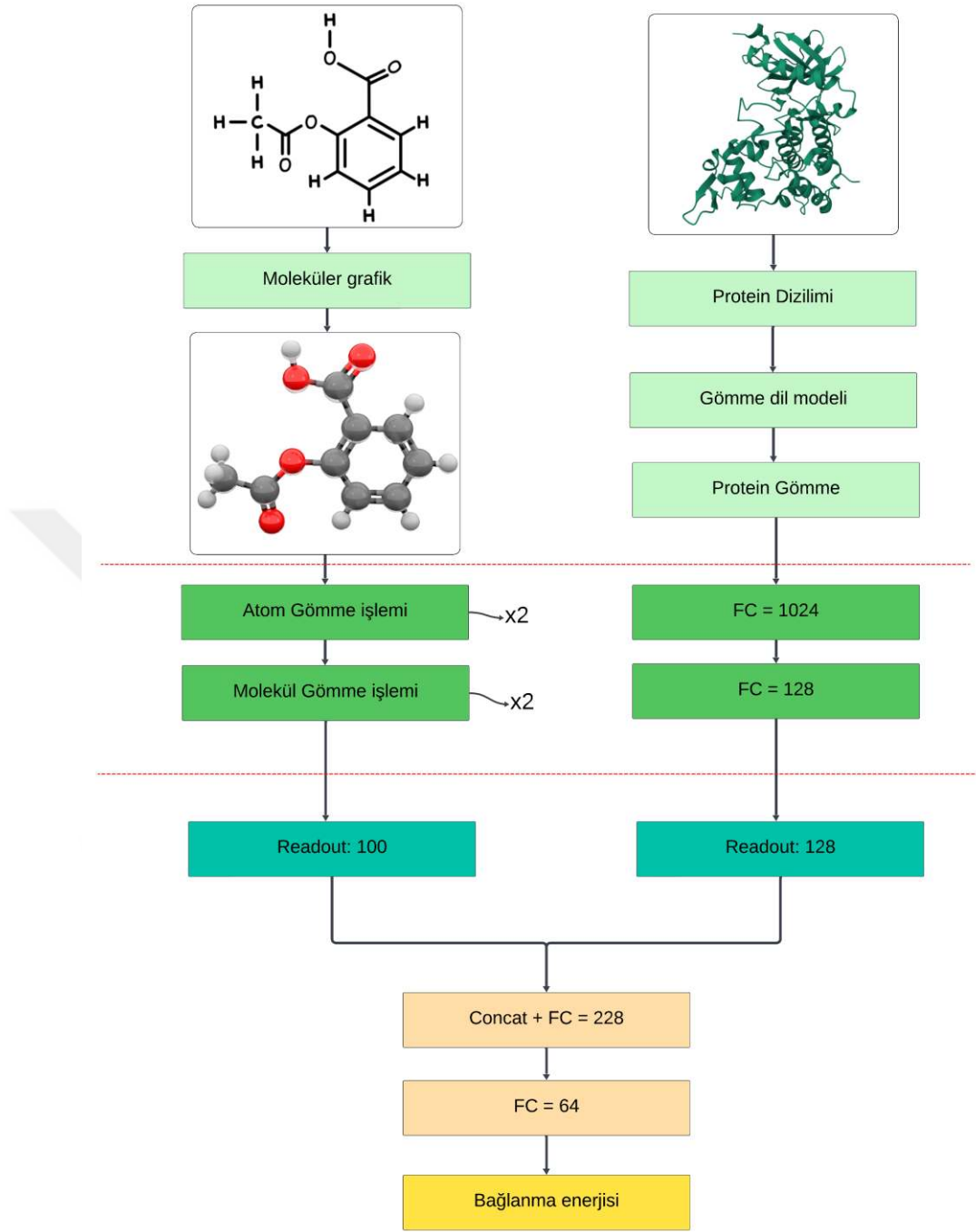
SARS-CoV-2'nin 3-kimotripsin benzeri proteazının (3CLpro) yapısal analizini geliştirmek ve potansiyel ilaç adaylarının tasarlanmasına yardımcı olmak için GSA'lar gibi gelişmiş analitik yöntemler kullanılmaktadır. GSA'ların protein-ligand etkileşimlerini tahmin etmede ve bağlanma yakınlıklarını hesaplamada etkili olduğu, 6LU7 proteazın yapısal özelliklerinin daha iyi anlaşılmasını sağladığı ve daha etkili ilaç tasarımı kolaylaştırdığı kanıtlanmıştır. Bu bölümde, 6LU7 proteazın yapısal analizine odaklandık ve antipsikotik bir ilaç olan klorpromazin ile 6LU7 proteaz arasındaki etkileşimi araştırılmıştır. GSA tabanlı analitik yöntemler kullanarak, klorpromazinin 6LU7 proteaz ile bağlanma afinitesini ve etkileşim mekanizmasını hesapladık. Sonuçlar, klorpromazinin 6LU7 proteaza güçlü bir bağlanma kapasitesini ortaya koydu ve COVID-19 tedavisi için umut verici bir ilaç adayı olarak potansiyelini gösterdi. 6LU7 proteazın hedeflenmesi ve COVID-19 tedavisi bağlamında ilaç tasarımından yararlanılması yeni bir bakış açısı sunmaktadır. GSA'lar gibi gelişmiş analitik yöntemlerin kullanılması, ilaç keşif sürecinde daha hassas ve kişiye özel analizler yapılmasına olanak tanır. Bu da potansiyel ilaç adaylarının daha etkili bir şekilde değerlendirilmesini sağlar ve COVID-19 gibi kritik hastalıklarla mücadele etmek için yeni terapötik yaklaşımların geliştirilmesinin önünü açar.

Sonuç olarak, bu bölümde anlatılan proje çalışması, 6LU7 proteazın yapısı ve etkileşim mekanizması hakkındaki anlayışımızı geliştirmeyi ve COVID-19 tedavisinde Klorpromazin'in potansiyelini değerlendirmeyi amaçlamaktadır. Bu araştırma, GSA'lar ve hesaplamalı modelleme teknikleri gibi gelişmiş analitik yöntemlerden yararlanarak, ilaç keşfi ve geliştirmedeki önemlerinin altını çizecek ve yeni tedavi yaklaşımlarının keşfedilmesini sağlayacaktır. Bu çalışmanın bulguları gelecekteki araştırma çabalarına ilham verebilir ve COVID-19 tedavisi için yeni ilaç adaylarının belirlenmesine katkıda bulunabilir. Ayrıca, ilacın yeniden konumlandırılmasındaki fırsatları ortaya koyarak, tedavi seçeneklerinin hızlı ve etkili

bir şekilde geliştirilmesini kolaylaştırabilir. Genel olarak, bu çalışma COVID-19 için potansiyel terapötik müdahaleler hakkında değerli bilgiler sağlama vaadini taşımaktadır ve daha hedefli ve verimli ilaç keşif çabalarının önünü açabilir.

Bu bölüm, FDA onaylı ilaçları GSA ile analiz ederek ve pandemiyle hızlı bir şekilde mücadele etmek için yeniden kullanıma uygun adayları belirleyerek COVID-19 pandemisine acil müdahaleye katkıda bulunmayı amaçlamaktadır. Şekil 4.28, bu amaçla potansiyel ilaç adaylarının yeniden tanımlanması için önerilen yaklaşımı göstermektedir.





Şekil 4.28 İlaç-protein etkileşimlerini analiz etmek için önerilen yaklaşımının model yapı gösterimi

İlaç-hedef etkileşimi

İlaç-hedef etkileşimi (DTI), kimyasal bileşiklerin protein hedeflerine bağlanmasını araştırarak ilaç keşif sürecinde çok önemli bir rol oynar. İlaç taraması ve ilacın yeniden konumlandırılması gibi kritik görevleri içerir. İlaç taraması, hedef proteinlere spesifik

olarak bağlanabilen ligand moleküllerini tanımlamayı amaçlarken, ilacın yeniden konumlandırılması mevcut ilaçlar için yeni terapötik kullanımları ortaya çıkarmayı amaçlamaktadır. DL teknikleri, geleneksel hesaplama yöntemlerine kıyasla bu görevlerde üstün performans göstermiştir. Klasik kimyasal parmak izi ilaç kodlayıcılarına ek olarak çeşitli derin öğrenme tabanlı ilaç kodlayıcıları kullanılmaktadır. Bunlar arasında oto kodlayıcılar, CNN'ler, RNN'ler, dönüştürücüler ve GSA'lar bulunmaktadır (Cheung & Moura, 2020).

GSA'lar, COVID-19 ilaç keşfi de dahil olmak üzere çeşitli alanlarda uygulanan ve başarılı sonuçlara yol açan analitik bir araç olarak potansiyellerini göstermiştir (Cheung & Moura, 2020). Grafik yapıları etkili bir şekilde işleme yetenekleri, onları ilaç-hedef etkileşimlerini tahmin etmede ve COVID-19 dahil olmak üzere çeşitli hastalıklar için potansiyel ilaç adaylarını belirlemede değerli kılmaktadır. Bu bölümde, GSA'ların moleküler modelleme ve ilaç-protein etkileşimlerinin analizi konusundaki uygulamaları tartışılmıştır. Daha önce Bölüm 3.2.1'de detaylı bir şekilde açıklanan GSA işleyişi ve bileşenleri hakkında bilgi edinmek için Bölüm 3.2.1'e bakınız.

Diğer makine öğrenme teknikleri

Bu bölümde, GSA modelinin yanı sıra CNN ve MLP gibi diğer makine öğrenimi tekniklerinden de yararlanılmıştır. Bu teknikler moleküler yapıların analizinde ve ilaç etkileşim mekanizmalarının anlaşılmasında farklı bakış açıları sunmaktadır.

MLP, sınıflandırma ve regresyon problemlerindeki etkinliği ile bilinen ve yaygın olarak kullanılan bir sinir ağı mimarisidir. Tahmin edilen ve gerçek girdi değerleri arasındaki hatayı en aza indirmek için en uygun parametreleri öğrenir, bu da onu çok yönlü hale getirir ve derin öğrenme çerçevelerinden destek alarak istikrarlı performans elde edebilir (Qin vd., 2022).

MLP'nin geniş uygulanabilirliği, parmak izleri (FP), transkriptomik veriler, biyoinformatik veriler ve moleküler özellikler gibi çeşitli veri türleriyle kombinasyonuna izin verir. Öte yandan CNN, bitişik özellikleri aynı hesaplama filtresi aracılığıyla işleyerek yerel özellik çıkarımı içeren görevler için yaygın olarak tercih

edilmektedir. Yapısal verilerle uyumluluğu nedeniyle moleküler tasarım ve ilaç keşfi alanlarında sıklıkla kullanılmaktadır. Ayrıca CNN, artan veri hacimlerinde bile verimli eğitim süreçleri ve giriş verilerinden kaynaklanan gürültüye karşı göreceli esneklik sergiler.

Atomik geometri gibi yapısal verilerle uyumluluğu göz önüne alındığında CNN, DEEPScreen'in molekülün 2B görüntüsünü kullandığı moleküler tasarım ve RoseNet, AK-score ve DeepDrug3D'nin protein-ligand çiftlerini voksellere dönüştürdüğü DTI tahmini gibi çeşitli alanlarda sıkça uygulama alanı bulmaktadır. Ayrıca, RNN'ler için uygun olmayan SMILES veya amino asit dizileri gibi sıralı ifadeler içeren veriler için CNN'ler tercih edilmektedir. Sıralı girdi verilerine sahip QSAR modellerinin oluşturulmasında hem DeepConv-DTI hem de transformer-CNN kullanılmıştır (Rifaioğlu vd., 2020).

Araştırmacılar, bu çeşitli ML tekniklerini kullanarak, karmaşık moleküler verilerden değerli içgörüler elde etmek ve ilaç keşif ve geliştirme süreçlerini optimize etmek için kendi güçlerinden yararlanabilirler.

Önerilen yöntem

Buradaki ana amaç, hazır paket programlarına alternatif olabilecek ücretsiz bir GSA tabanlı moleküler modelleme ile ilaç-protein etkileşimlerinin analizi ve ilaç keşfi için klorpromazinin COVID-19 tedavisindeki potansiyelini değerlendirmek ve ilaç geliştirmede YZ'ye dayalı yenilikçi bir strateji sunmaktır. Ayrıca, ilaçlar ve hedef proteinler arasındaki bağlanma afinitesinin tahmini, moleküler etkileşimlerin araştırılması ve inhibitör etkilerinin değerlendirilmesi konusunda değerli bilgiler sunmaktadır. Bunu başarmak için, bu yaklaşımların ilaç keşfi ve geliştirmedeki öneminin altını çizmek için GSA'lar ve çeşitli hesaplamalı modelleme teknikleri kullanılmıştır. Bu katkı, yeni terapötik metodolojilerin araştırılmasını daha da ileriye götürmektedir.

Bölüm 3.2.1'de detaylı şekilde belirtildiği her grafik $G = (V, E)$ bir dizi köşe V ve bir dizi kenardan $E \subseteq V \times V$ oluşur. Her köşe $v_i \in V$ özniteliklerin l_i bir vektörü ile etiketlenir. Her kenar $(v_i, v_j) \in E$ özniteliklerin $e_{i,j}$ bir vektörü ile etiketlenir. E'ye dayalı olarak, her bir $v_i \in V$ tepe noktasına bir dizi $Ne_{(v_i)}$ komşusu atayan bir

komşuluk fonksiyonu $Ne_{(vi)} = \{v_j: (v_i, v_j) \in E\}$ olarak tanımlanır (Bongini vd., 2021).

Her tepe noktası $v_i \in V$ 'ye bir durum s_i atanır. Şu şekilde başlatılır $s_i^0 = l_i$. Daha sonra tüm köşelerin durumu, yakınsama ve/veya maksimum iterasyon sayısına ulaşılan kadar iteratif ve paralel olarak güncellenir. Bu işlem sırasında, her bir köşe mesaj gönderip alarak komşularıyla iletişim kurar. Adım k 'da köşe v_i 'nin durumu s_i^k , önceki adım s_i^{k-1} 'daki durumuna ve komşularından gelen mesajlara dayalı olarak Denklem (4.16)'daki gibi hesaplanır $v_j \in Ne_{(vi)}$ (Bongini vd., 2021).

$$s_i^k = F\left(s_i^{k-1}, A(\{M(v_i, v_j, k): v_j \in Ne(v_i)\})\right) \quad (4.16)$$

Denklem (4.18)'de, bir mesaj geçirme fonksiyonu M , k adımında v_j 'ten v_i 'ye gönderilen mesajı tanımlar. Bir komşuluk toplama fonksiyonu A , v_i 'nin tüm komşularından gelen mesajların nasıl toplandığını tanımlar. Son olarak, bir durum güncelleme fonksiyonu F yeni durumun nasıl hesaplanacağını tanımlar. Prensip olarak, tüm bu fonksiyonlar sinir ağları tarafından öğrenilebilir. Daha önceki modellerde (Beigel et. Al., 2020), M bir MLP tarafından uygulanırken, diğer iki işlev sabit tutulmuştur. Bu çalışmada M ve A 'yı sabit tutarken F 'i bir MLP ile öğrenme tercih edilmiştir. Bu seçim aynı zamanda (Abdelrheem vd., 2020; Z. Jin vd., 2020) genel olarak özetlenen diğer son gelişmelerle de uyumludur. Özellikle, M 'yi kaynak düğüm s_j^{k-1} 'nin durumu ve $e_{i,j}$ kenarının $(v_i$ ve $v_j)$ etiketinin birleşimi olarak Denklem (4.17)'de tanımladık.

$$M(v_i, v_j, k) = (s_j^{k-1}, e_{i,j}) \quad (4.17)$$

Denklem 4.18'de gösterildiği gibi, gelen mesajların aritmetik ortalaması ya da eleman bazında toplamı olarak tanımlanan iki farklı komşuluk toplama fonksiyonu kullanılmıştır.

$$A_{avg} = (\{M(v_i, v_j, k): v_j \in Ne(v_i)\}) = \frac{1}{|Ne(v_i)|} \sum_{j: v_j \in Ne(v_i)} M(v_i, v_j, k) \quad (4.18)$$

$$A_{sum} = (\{M(v_i, v_j, k): v_j \in Ne(v_i)\}) = \sum_{j: v_j \in Ne(v_i)} M(v_i, v_j, k) \quad (4.18)$$

Bu nedenle A_{avg} ve A_{sum} toplama fonksiyonları arasında seçim yapmamızı sağlayan bir hiperparametre katsayısı a 'yı, sırasıyla $\frac{1}{|Ne(v_i)|}$ veya 1 değerlerini atayarak tanıtabiliriz. Sonuç olarak, Denklem (4.16)'nın son hali Denklem (4.19)'daki gibi yazılabilir.

$$s_i^k = F(s_i^{k-1}, a \sum_{j: v_j \in Ne(v_i)} (s_j^{k-1}, e_{i,j})) \quad (4.19)$$

Durum güncelleme süreci, durum yakınsaması veya maksimum iterasyon sayısı $k = k_{max}$. Durum yakınsamasının, mevcut durum ile bir önceki durum arasındaki mesafe ihmal edilebilir hale geldiğinde gerçekleştiği varsayılır: $\|s_i^k - s_i^{k-1}\| < \epsilon$. ϵ ve k_{max} hiperparametreler olarak ayarlanır, $\| \cdot \|$ Öklid normudur.

Veri seti 5:

Model girdi verileri, ilaçların kimyasal bileşimlerini farklı ve geri dönüşümlü bir 2B konfigürasyonda göstermek için SMILES kullanılmıştır. SMILES, bir bileşiğin moleküler bileşimini özetleyen doğrusal bir dizi sunar. Ancak, bazı durumlarda farklı moleküllerin aynı SMILES dizisine bağlanabileceğini ve bunun da modelin etkinliğini potansiyel olarak etkileyebileceğini kabul etmek çok önemlidir.

Bu araştırma için veri seti, COVID-19 tedavisi için seçilen FDA onaylı ilaçlara odaklanmıştır. Veriler, halka açık veritabanları DrugBank ve PubChem'den elde edilmiştir. DrugBank, ilaç hedefleri, enzimler, etkileşimler ve ilaç taşıyıcıları dahil olmak üzere ilaçlar hakkında temel bilgiler sağlayarak ilaçla ilgili güvenilir bir veri kaynağı haline gelmiştir. Öte yandan PubChem, ilaç yapıları hakkında kapsamlı ayrıntılar sunarak araştırmacıların ilaçların moleküler bilgilerine erişmesini sağlamıştır.

İlgilenilen protein yapısı, viral yaşam döngüsünde önemli bir enzim olan ve ilaç geliştirme için cazip bir hedef olarak kabul edilen SARS-CoV-2 virüsünden 3CLpro

idi. Bu bölümde kullanılan spesifik protein yapısı Protein Veri Bankası'ndan (PDB ID: 6LU7) elde edilmiştir.

Çalışmada ayrıca, araştırmada kullanılan ilaçlarla ilgili spesifik bilgileri özetlemek için Tablo 4.19'dan yararlanılmıştır. Bu tablo muhtemelen ilaç isimleri, SMILES veya SDF formatındaki kimyasal yapılar ve ilaçlarla ilgili diğer bilgiler gibi ayrıntıları içermektedir. Genel olarak, DrugBank, PubChem ve Protein Data Bank'tan alınan verilerle birlikte SMILES gösteriminin kullanılması, çalışmamızın COVID-19 tedavisi bağlamında moleküler yapı analizi ve ilaç keşfi için uygun bir veri kümesi oluşturmaya olanak sağlamıştır.

Tablo 4.9 Diğer ilaçlar hakkında genel bilgiler

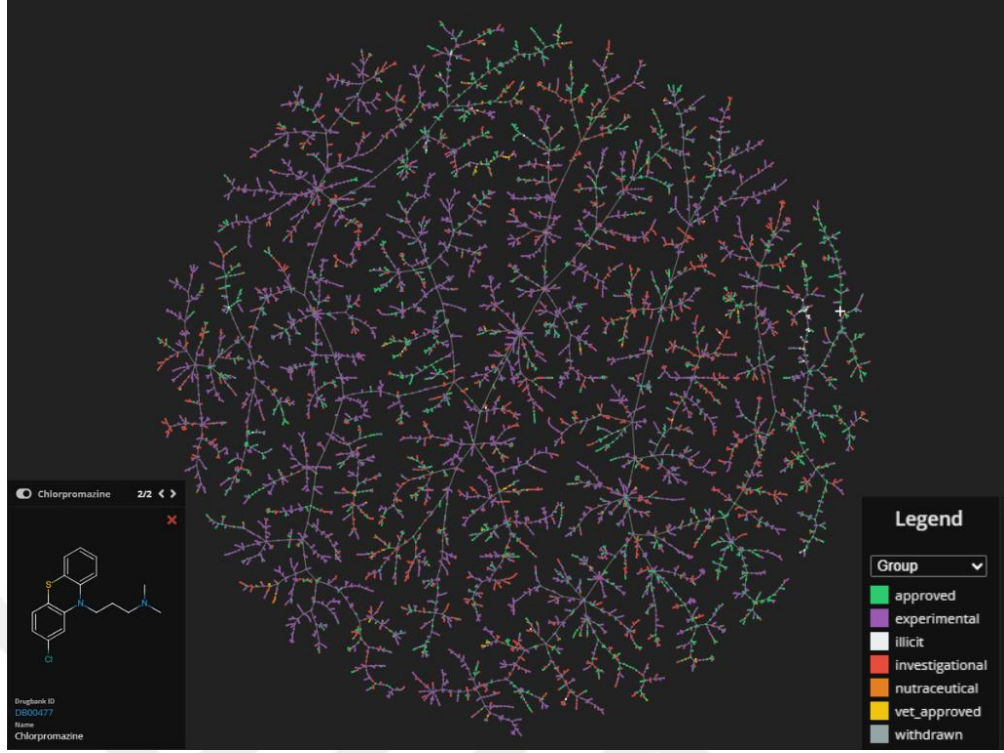
İlaç	İlaç özellikleri
Chlorpromazine	Bir fenotiyazin antipsikotik olan klorpromazin, bulantı, kusma, ameliyat öncesi anksiyete, şizofreni, bipolar bozukluk ve çocuklarda ciddi davranış sorunları gibi çeşitli durumların tedavisinde uygulama alanı bulur). (DrugBank, 2023; PubChem, 2023)
Letermovir	Letermovir, allojenik hematopoetik kök hücre nakli hastalarında sitomegalovirüs (CMV) enfeksiyonunun profilaksisinde kullanımı için 8 Kasım 2017 tarihinde FDA onayı aldı. DNA terminaz kompleks inhibitörleri olarak bilinen yeni bir CMV anti-enfektif sınıfının ilk üyesi olan Letermovir, bu alanda önemli bir ilerlemeyi temsil etmektedir. Ayrıca, FDA tarafından hem öncelikli hem de yetim ilaç statüsü verilmiştir. Letermovir şu anda Prevymis markası altında ticari olarak temin edilebilmektedir (DrugBank, 2023; PubChem, 2023).
Ribavirin	Ribavirin, Hepatit C'nin belirli formlarının tedavisinde kullanılan bir guanozin nükleozid ilacıdır (DrugBank, 2023; PubChem, 2023).
Remdesivir	Bir nükleozid analogu olan Remdesivir, COVID-19 da dahil olmak üzere RNA virüs enfeksiyonlarının tedavisi için umut verici bir antiviral ilaç olarak ortaya çıkmıştır. İlk olarak 2016 yılında Ebola için potansiyel bir tedavi olarak tanımlanması, çeşitli viral aileleri hedef alan geniş spektrumlu antiviral aktivitesini vurgulamaktadır. FDA tarafından 19 Kasım 2020 tarihinde verilen Acil Kullanım İzni, COVID-19 pandemisi sırasında etkili tedavilere duyulan acil ihtiyacı daha da vurgulamaktadır. Remdesivir'in onayı ve ardından baricitinib ile birlikte kullanımı, COVID-19 hastaları için değerli bir tedavi seçeneği olma potansiyelinin altını çizmektedir (Beigel vd., 2020; DrugBank, 2023; PubChem, 2023).
Lopinavir	Lopinavir, insan immün yetmezlik virüsü (HIV) enfeksiyonunu etkili bir şekilde tedavi etmek için ritonavir ile kombinasyon halinde kullanılan bir HIV-1 proteaz inhibitörüdür (DrugBank, 2023; PubChem, 2023).
Hydroxychloroquine	Hidroksiklorokin, öncelikle komplike olmayan sıtma vakalarının tedavisi ve belirli bölgelerde kemoprofilaksi için kullanılan bir antimalaryal ilaç olarak hizmet eder. Ayrıca, hastalık modifiye edici bir anti-romatizmal ilaç (DMARD) olarak işlev görür ve romatoid artrit ve lupus eritematozus tedavisinde endikedir (DrugBank, 2023; PubChem, 2023).

Tablo 4.9 Diğer ilaçlar hakkında genel bilgiler (Devamı)

Maraviroc	Maraviroc, CCR5 ko-reseptör antagonisti olarak sınıflandırılır ve CCR5-tropik HIV-1 enfeksiyonunu etkili bir şekilde tedavi etmek için diğer antiretroviral ilaçlarla birlikte uygulanır (DrugBank, 2023; PubChem, 2023).
Favipiravir	Favipiravir, öncelikle influenza tedavisinde kullanılan antiviral bir ilaçtır ve diğer viral enfeksiyonların hedeflenmesinde de potansiyel uygulamaları vardır (DrugBank, 2023; PubChem, 2023).
Loviride	Lovirid, efavirenz ve nevirapini de içeren nükleozid olmayan ters transkriptaz inhibitörleri ailesine aittir. Bu ilaçlar anti-HIV tedavi rejimlerinin bir parçası olarak kullanılır (DrugBank, 2023; PubChem, 2023).

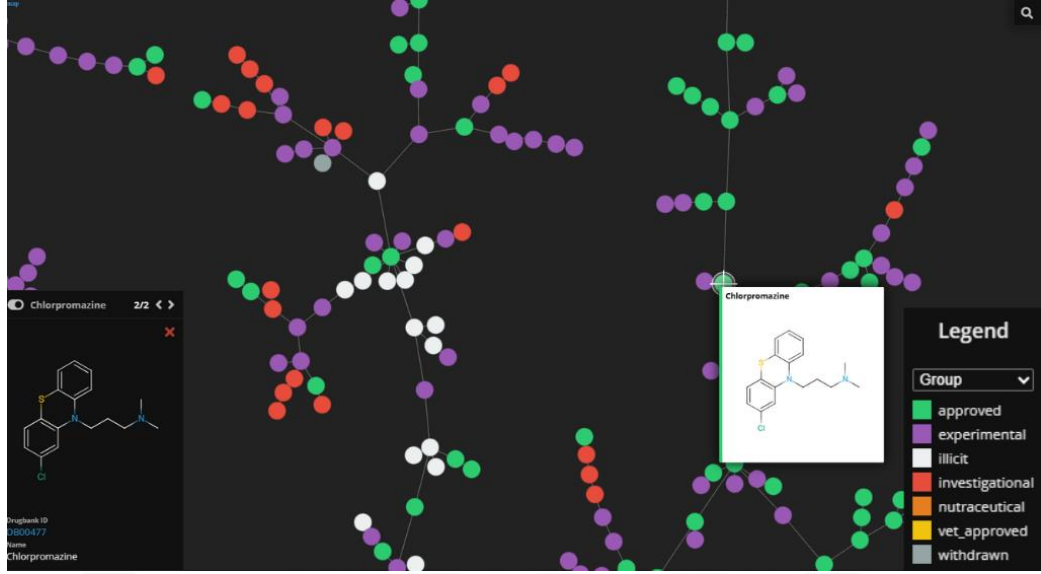
İlaç-protein etkileşimleri, bu ilaçların hedef proteinlere bağlanma enerjileri hesaplanarak analiz edilmiştir. Bu analizler, bu ilaçların etkileşim mekanizmalarını anlamak ve potansiyel olarak yeni terapötik stratejiler geliştirmek için çok önemlidir. Seçilen bu ilaçlara odaklanan kapsamlı araştırmalar, klinik ortamlarda yaygın olarak kullanıldıklarını ve mevcut hastalıkların tedavisinde önemli rol oynadıklarını göstermektedir. İlaç-protein analizinin sonuçları, bu ilaçların mevcut tedavi seçeneklerine alternatifler sunabileceğini göstermektedir. Bu çalışma, bu ilaçların hedef proteinlerle etkileşimini değerlendirerek yeni uygulama alanlarını ve tedavi seçeneklerini keşfetme potansiyeline sahiptir.

Böylece, spesifik ilaçlara odaklanan bu çalışma, mevcut tedavi protokollerinin iyileştirilmesi ve viral hastalıklarla mücadele için daha etkili ilaçların geliştirilmesinde kritik bir adımı temsil etmektedir. Bu ilaçların etkileşim mekanizmalarının anlaşılması, hedeflenen dağıtımlarını ve terapötik etkinliklerini artırabilir.



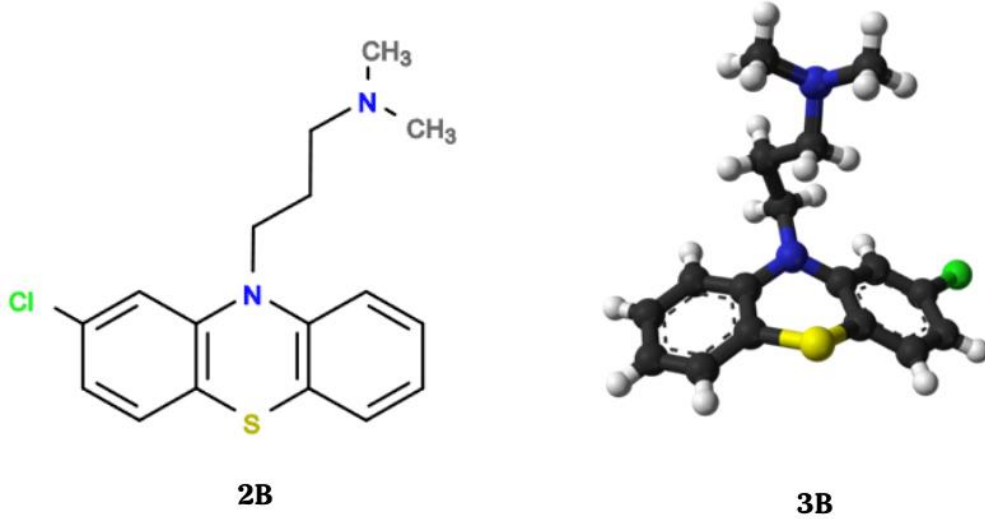
Şekil 4.29 TMAP (Ağaç MAP) yöntemi kullanılarak FDA onaylı ilaçların görselleştirilmesi

Moleküler veri tabanlarındaki kimyasal benzerlikleri temsil etmek için kullanılan TMAP (Tree MAP) yöntemi, bu çalışmada ilaç çiftlerinin yapısal benzerliklerini anlamak ve ilaç geliştirme için kritik içgörüler sağlamak için kullanılmıştır (Probst & Reymond, 2020). Bu yaklaşım, ilgili moleküler özellikleri görselleştirerek ve moleküller arasındaki ilişkileri ortaya çıkararak veri tabanlarının analizini kolaylaştırmaktadır. Şekil 4.29, bu çalışmada kullanılan ilaçları görsel olarak göstermekte ve böylece moleküller arasındaki ilişkileri ortaya çıkararak veri tabanlarının analizini kolaylaştırmaktadır.



Şekil 4.30 FDA onaylı tüm ilaçlar arasında Klorpromazin'in yapısı

Yakın zamanda geliştirilen TMAP algoritmasını kullanarak kimyasal uzayın görsel bir temsilini oluşturduk ve ilaç çiftlerini analiz ettik. TMAP, veri kümelerindeki ilaç çiftlerine genel bir bakış elde etmemizi ve ilgili ilaçları görselleştirmemizi sağladı. Şekil 4.30, Klorpromazin'in TMAP haritasındaki konumunu göstermekte ve ilacın kimyasal benzerliklerini ve ilişkilerini anlamayı kolaylaştırmaktadır.



Şekil 4.31 Klorpromazinin 2B ve 3B gösterimi

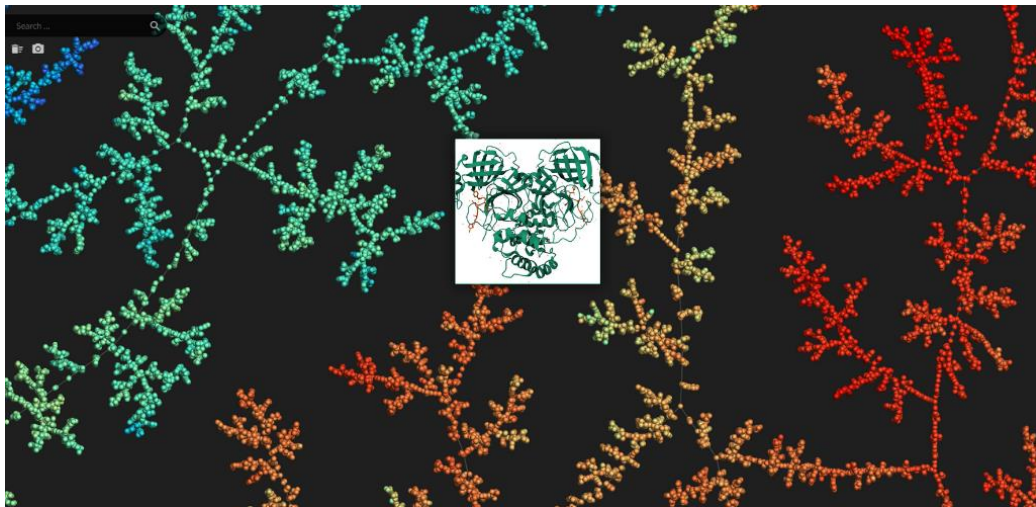
Şekil 4.31'de Klorpromazinin 2B ve 3B kimyasal yapısı gösterilmektedir. Bu görselleştirme, ilacın atomlarını, bağlarını ve kimyasal gruplarını gösteren moleküler yapısını temsil etmektedir. Klorpromazinin 2B yapısı, ilacın kimyasal bileşimini ve

yapısal özelliklerini anlamak için kullanılır. Bu tür görselleştirmeler, ilaç tasarımı ve ilaç etkileşimi çalışmalarında önemli araçlar olarak hizmet eder.

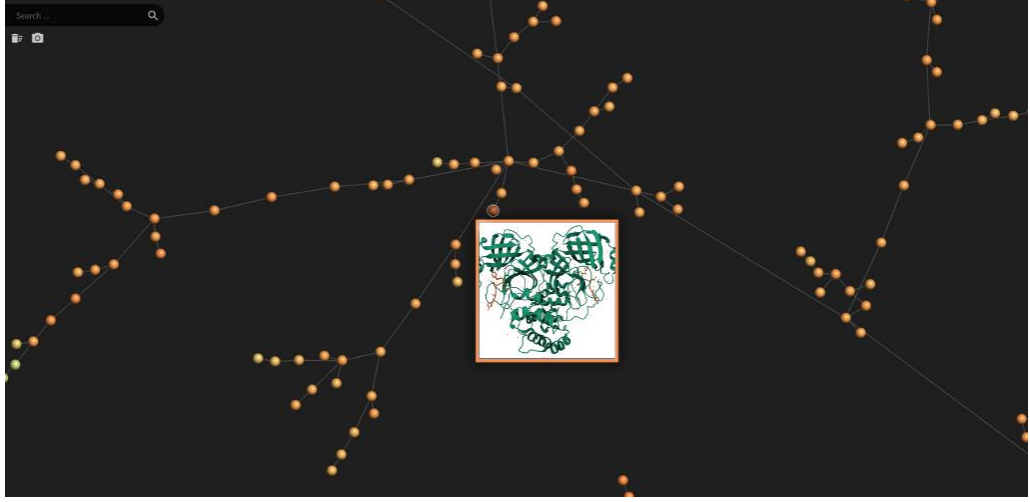
Tablo 4.10 6LU7 proteini ilgili genel bilgiler

Protein	Protein ayrıntıları
6LU7 proteaz	SGFRKMAFPSGKVEGCMVQVTCGTTTLNGLWLDDVVYCPRHVICTSEDM LNPNYEDLLIRKSNHNFLVQAGNVQLRVIGHSMQNCVLKLKVDLTANPKTP KYKFVRIQPGQTFSVLACYNGSPSGVYQCAMRPNFTIKGSFLNGSCGSGVGF NIDYDCVSFCYMHMELPTGVHAGTDLEGNFYGPFVDRQTAQAAGTDTTI TVNVLAWLAAVINGDRWFLNRFTTLNDFNLVAMKYNIEPLTQDHVDI LGPLSAQTGIAVLDMCASLKELLQNGMNGRTILGSALLEDEFTPFVVRQC SGVTFO

Tablo 4.10, SARS-CoV-2 virüsünün bir proteaz enzimi olan ve viral replikasyon ve olgunlaşma süreçlerinde çok önemli bir rol oynayan 6LU7 proteazı sunmaktadır. Bu enzim, virüsün genetik materyalinin işlenmesi ve yeni viral bileşenlerin sentezlenmesi için gereklidir. Sonuç olarak, 6LU7 proteaz SARS-CoV-2'nin hayati bir bileşeni olarak kabul edilir ve ilaç geliştirme sürecinde önemli bir hedef haline gelir. 6LU7 proteaz, virüsün poliprotein zincirini parçalama ve parçalama yeteneğine sahiptir, bu da onu virüsün replikasyon sürecinde kritik hale getirir. 6LU7'nin inhibe edilmesi virüsün replikasyonunu durdurabilir ve enfeksiyonun kontrol altına alınmasına yardımcı olabilir. Bu nedenle, 6LU7 proteaz COVID-19 tedavisi için potansiyel bir terapötik hedef olarak değerlendirilmektedir. Şekil 4.32-4.33'te, 6LU7 proteaz ile etkileşimini anlamak için önemli bir araştırma alanı oluşturan tüm proteinler arasından hedef olarak seçilen 6LU7 proteinine ilişkin veriler gösterilmektedir.



Şekil 4.32 6LU7 proteininin tüm proteinler arasında temsili



Şekil 4.33 6LU7 proteininin tüm proteinler arasından temsili farklı gösterimi

Şekil 4.33'te, 6LU7 proteini önemli yapısal detaylarıyla birlikte gösterilmektedir. Protein, monomerik durumunda 306 amino asit içeren tek bir zincir olarak temsil edilmektedir. Ancak, biyolojik olarak aktif formunda dimer olarak bulunur, yani başka bir 6LU7 molekülü ile bir kompleks oluşturur.

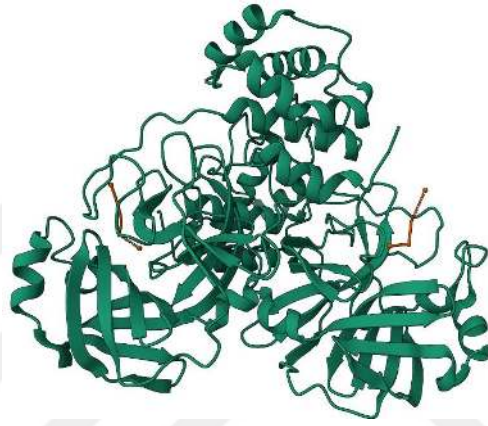
6LU7 proteininin yapısı, I, II ve III olarak etiketlenen üç farklı yapısal alan halinde düzenlenmiştir. Etki alanı I ve II, proteinin aktif bölgesinin oluşturulmasında çok önemli bir rol oynar. Aktif bölge, protein içinde enzimatik işlevini yerine getirmesini sağlayan belirli bir bölgedir. Bu işlev virüsün replikasyon ve olgunlaşma süreci için gereklidir.

Öte yandan, alan III dimer oluşumundan sorumludur. Dimerizasyon, iki özdeş protein molekülünün işlevsel bir birim oluşturmak üzere birbirine bağlandığı süreçtir. 6LU7 durumunda, alan III iki ayrı 6LU7 monomeri arasındaki etkileşimi kolaylaştırarak aktif dimerik durumun oluşmasını sağlar.

Şekil 4.33'te gösterilen yapıda, 6LU7 proteini, N3 (N-[(5-METHYLISOXAZOL-3-YL)CARBONYL]ALANYL-L-VALYL-N1-((1R, 2Z)-4-(BENZYLOXY)-4-OXO-1-[[3R)-2-OXOPYROLIDIN-3-YL]METHYL}BUT-2-ENYL)-L-LEUCINAMIDE). N3 inhibitörü, 6LU7'nin aktif bölgesiyle etkileşime girecek ve böylece enzimatik aktivitesini bloke edecek şekilde tasarlanmış küçük bir moleküldür. N3 tarafından 6LU7'nin bu şekilde inhibe edilmesi, viral replikasyon sürecine

müdahale edebileceği için onu ilaç geliştirme için potansiyel bir aday haline getirmektedir.

Şekil 4.33'te verilen yapısal bilgiler, 6LU7'nin etki mekanizmasını ve bir ilaç hedefi olarak potansiyelini anlamak için gereklidir. Araştırmacıların ve ilaç geliştiricilerin, SARS-CoV-2'ye karşı etkili terapötiklerin geliştirilmesine yol açabilecek ve COVID-19'un yönetimine katkıda bulunabilecek 6LU7 ile etkileşime girebilecek ve onu inhibe edebilecek potansiyel ilaç adayları tasarımlarına yardımcı olur.



Şekil 4.34 6LU7 Proteazın yapısı

Bulgular 5:

DTI, ilaç bileşiklerinin belirli protein hedeflerine bağlanmasını değerlendirdiği için ilaç geliştirmenin önemli bir parçasıdır. DTI tahmin doğruluğu, çeşitli ilaç geliştirme görevleri için kritik öneme sahiptir; ilaç taraması ve yeniden kullanım iki önemli uygulamadır. İlaç taraması, spesifik hedeflere bağlanabilen muhtemel ligandları tanımlarken, ilaç yeniden kullanımı mevcut ilaçlar için yeni terapötik kullanımları araştırır. Bu uygulamalar, COVID-19'da olduğu gibi etkili ve güvenli tedavilerin belirlenmesinde özellikle önemli olan ilaç araştırmalarıyla ilgili zaman, maliyet ve çabayı büyük ölçüde en aza indirir.

Bunun için, DTI modelleri bir kodlayıcı-kod çözücü çerçevesi kullanılarak oluşturulmuş ve ilaç yeniden tasarlama ve tarama uygulamalarına olanak sağlamıştır. Model, ilacın SMILES dizesi ve hedefin amino asit dizisinden oluşan ilaç-hedef

çiftlerini girdi olarak alır ve ilaç ile hedef arasındaki bağlanma aktivitesini temsil eden bir puan oluşturur. Bu yaklaşım, ilaç keşfini hızlandırmada ve COVID-19 dahil olmak üzere çeşitli hastalıklar için tedavilerin geliştirilmesini kolaylaştırmada muazzam bir potansiyele sahiptir.

Analiz için GSA, CNN ve MLP dahil olmak üzere çeşitli makine öğrenimi teknikleri kullanılmıştır. Bu tekniklerden elde edilen sonuçlar, klorpromazinin SARS-CoV-2 virüsünün hedef proteinlerinden biri olan 6LU7 proteaz ile etkileşime girme kabiliyetine sahip olduğunu ortaya koymuştur. Bu etkileşim, klorpromazinin virüse karşı potansiyel bir inhibitör olarak hizmet edebileceğini göstermektedir.

Çalışmada yer alan Tablo 4.11, ilaç-protein bağlanma enerjisi değerlerini sunmakta ve klorpromazin ile hedef protein arasındaki etkileşimlerin gücü hakkında önemli bilgiler sağlamaktadır. Bu bulgular, COVID-19 ile mücadelede klorpromazinin potansiyel terapötik rolünü anlamak için önemli etkilere sahiptir ve bu küresel sağlık sorunu için yeni tedavi stratejilerinin geliştirilmesine katkıda bulunabilir.

Tablo 4.11 İlaç-protein bağlanma enerjilerinin değerleri

Sıra	İlaç-adı	Hedef protein-adı	Bağlanma Skoru (K _d) (-) kcal/mol
1.	Chlorpromazine	SARS-CoV2 3CL-Protease (6LU7)	1.03
2.	Loviride	SARS-CoV2 3CL-Protease (6LU7)	3.50
3.	Favipiravir	SARS-CoV2 3CL-Protease (6LU7)	8.75
4.	Letermovir	SARS-CoV2 3CL-Protease (6LU7)	22.03
5.	Ribavirin	SARS-CoV2 3CL-Protease (6LU7)	26.68
6.	Remdesivir	SARS-CoV2 3CL-Protease (6LU7)	45.98
7.	Lopinavir	SARS-CoV2 3CL-Protease (6LU7)	52.04
8.	Maraviroc	SARS-CoV2 3CL-Protease (6LU7)	61.25
9.	Hydroxychloroquine	SARS-CoV2 3CL-Protease (6LU7)	71.08

Tablo 4.11, bir bileşiğin belirli bir hedefe ne kadar iyi bağlandığını belirlemede çok önemli olan ilaç-protein bağlanma enerjisi değerlerini sunmaktadır. Bir ilaç ve hedefi arasındaki bağlanma afinitesini ölçmek için, ayrışma sabiti (K_d) gerçek değerli bir metrik olarak kullanılır. K_d değeri, bağlanma etkileşiminin gücünün nihai ölçüsü olarak hizmet eder ve potansiyel ilaç adaylarının etkinliğinin değerlendirilmesinde çok önemli bir rol oynar. Daha düşük bir K_d değeri, ilaç ile hedef arasında daha güçlü bir bağlanma afinitesine işaret eder ve daha fazla geliştirme için daha umut verici bir aday

anlamına gelir. Bu bağlamda, Denklem (4.20), ilaçlar ve ilgili protein hedefleri arasındaki bağlanma afinitesini değerlendirmek için bir kriter olarak hizmet eder ve ilaç keşfi ve geliştirme bağlamında ilaç adaylarının potansiyel etkinliği hakkında değerli bilgiler sağlar (Budak vd., 2023).

$$Kd = \frac{k_{kapalı}}{k_{açık}} \quad (4.20)$$

Burada $K_{a\check{c}ık}$ K_{on} birleşme hız sabitini, $K_{kapalı}$ ise ayrışma hız sabitini temsil eder.

Bu çalışmanın bulguları, ilaçların 6LU7 proteaza -1.03 ila -71.08 kcal/mol arasında değişen farklı bağlanma serbest enerjileri ile bağlandığını göstermiştir. Tablo 4.11'e göre, en güçlü bağlanma enerjisi Klorpromazin ilacı durumunda gözlenmektedir. Öte yandan, Tablo 4.12 ilaç-hedef etkileşimlerini, özellikle de bu ilaçların 6LU7 proteaz ile DTI oranlarını sunmaktadır.

Tablo 4.12'ye göre, en yüksek etkileşim oranına (%97) sahip ilaç Klorpromazin'dir. Bu sonuçlar, klorpromazinin 6LU7 proteaz ile güçlü bir etkileşim sergilediğini ve potansiyel bir anti-COVID-19 ajanı olarak düşünülebileceğini göstermektedir.

Tablo 4.12 İlaç-protein etkileşim sonuçları

Sıra	İlaç-Adı	Hedef protein- adı	Etkileşim	Olabilirlik
1	Chlorpromazine	SARS-CoV2 3CL Protease	EVET	0.97
2	Loviride	SARS-CoV2 3CL Protease	EVET	0.74
3	Favipiravir	SARS-CoV2 3CL Protease	EVET	0.70
4	Letermovir	SARS-CoV2 3CL Protease	EVET	0.64
5	Ribavirin	SARS-CoV2 3CL Protease	EVET	0.64
6	Remdesivir	SARS-CoV2 3CL Protease	HAYIR	0.43
7	Lopinavir	SARS-CoV2 3CL Protease	HAYIR	0.38
8	Maraviroc	SARS-CoV2 3CL Protease	HAYIR	0.29
9	Hydroxychloroquine	SARS-CoV2 3CL Protease	HAYIR	0.22

Tablo 4.13 ve Tablo 4.14, algoritma sonuçlarının elde edilmesinde kullanılan çeşitli moleküler ilaç ve hedef protein kodlama türlerini içermektedir. Bu tablolarda farklı moleküler yapıların ve veri gösterimlerinin kullanıldığı yöntemler sunulmaktadır. Bu

kodlama türleri, algoritmanın ilaç ve protein verilerini eğitim ve tahmin aşamaları için uygun bir formatta temsil etmesini sağlar.

Tablo 4.13 İlaç kodlamalar ve açıklamaları

Sıra	İlaç Kodlama	Açıklama
1	Daylight	Daylight-tipi parmakizi
2	CNN	SMILES üzerinde Evrişimsel Sinir Ağı
3	MPNN	Mesaj geçiren sinir ağı
4	DGL_GCN	Grafik Evrişimli Ağlar

Tablo 4.13, ilaç kodlaması için kullanılan çeşitli yöntemleri ve bunların algoritmanın performansı üzerindeki etkilerini göstermektedir. İlaçların farklı özellikleri, yapıları ve bileşenleri farklı kodlama türlerinin kullanılmasını gerektirebilir. Bu tablo FP, SMILES ve diğer moleküler özellik tabanlı kodlama yöntemlerini içerebilir.

Tablo 4.14 Hedef kodlamalar ve açıklamaları

Sıra	Hedef Kodlama	Açıklama
1	AAC	3-mers'e kadar amino asit bileşimi
2	CNN	Hedef dizilim üzerinde Evrişimsel Sinir Ağı
3	MPNN	Mesaj geçiren sinir ağı
4	DGL_GCN	Grafik Evrişimli Ağlar

Tablo 4.14'te hedef proteinleri kodlamak için kullanılan farklı teknikler ve bunların algoritmanın performansı üzerindeki etkileri sunulmaktadır. Proteinlerin amino asit dizilimleri veya yapısal özellikleri çeşitli kodlama türlerinin kullanılmasını gerektirebilir.

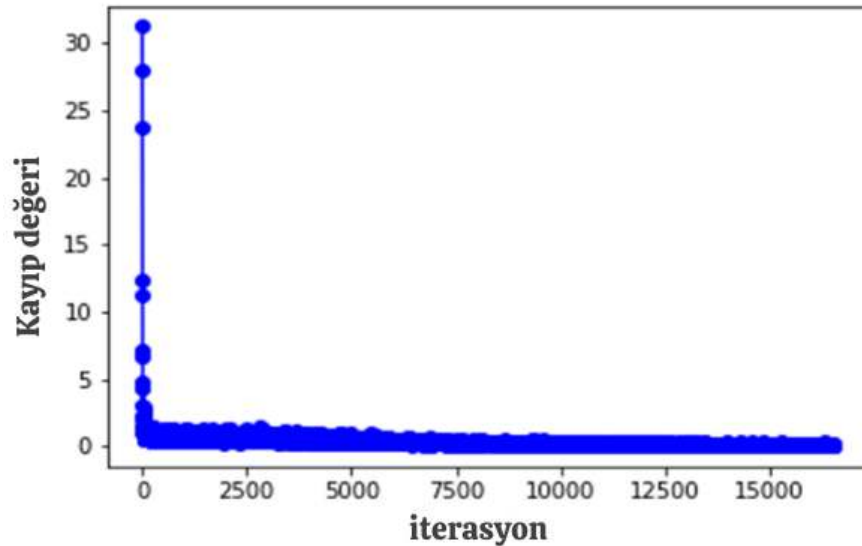
Tablo 4.13 ve 4.14, hangi kodlama yöntemlerinin daha iyi sonuçlar verdiğini ve algoritmanın daha etkili çalışmasını sağladığını belirlemek için algoritmanın performansını farklı kodlama türleriyle karşılaştırmaktadır. Bu tablolar, ilaç-hedef etkileşim analizlerinin ve ilaç geliştirme süreçlerinin optimize edilmesine katkıda bulunabilir.

Bu çalışmada, ilaç bileşikleri ve protein dizileri için iki farklı moleküler gösterim türü kullanılmıştır: İlaç bileşikleri için gün ışığı parmak izleri ve protein dizileri için AAC (Amino Asit Kompozisyonu). Veri seti üç alt kümeye ayrılmıştır: verilerin %70'ini içeren bir eğitim seti, verilerin %10'unu içeren bir doğrulama seti ve verilerin %20'sini içeren bir test seti. Önceden eğitilmiş modeller için eğitim süreci, sinir ağının eğitim

katmanlarındaki birkaç kritik parametrenin yapılandırılmasını, başlatılmasını ve optimize edilmesini içeriyordu. Bu parametreler şunları içerir:

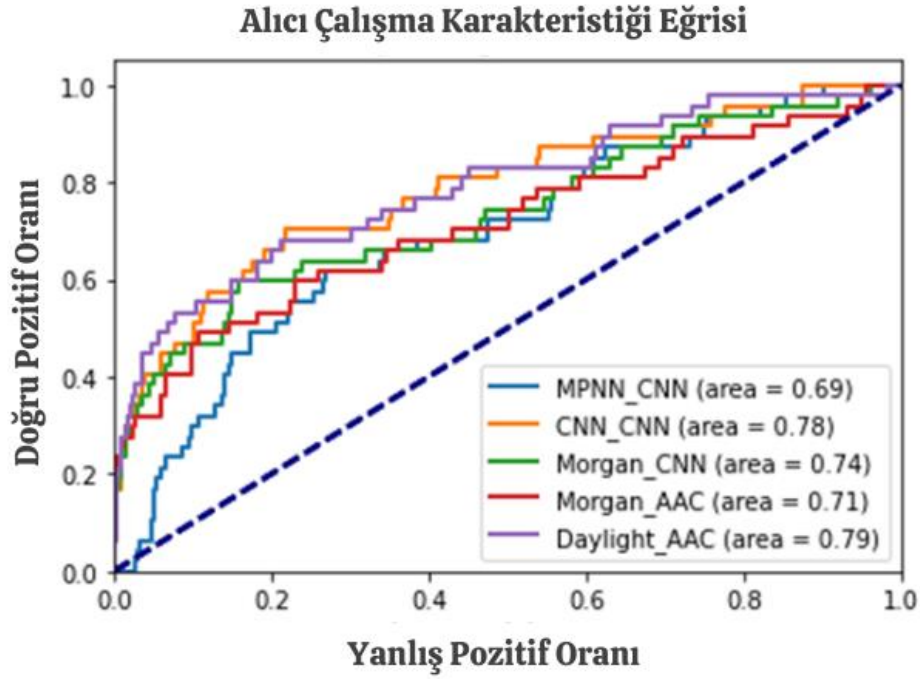
- İterasyon (Epoch): Eğitim sırasında tüm eğitim veri setlerinin sinir ağında yinelenme sayısı. Daha yüksek sayıda epok daha kapsamlı bir eğitim sağlayabilir, ancak aynı zamanda modelin eğitim verilerine aşırı uyum sağlama riskini de artırır.
- Batch boyutu (Batch size): Her iterasyonda ağ üzerinden yayılan veri örneklerinin sayısı. Daha büyük yığın boyutları eğitimi hızlandırabilir, ancak daha fazla bellek gerektirebilir ve daha az genelleştirilmiş bir modelle sonuçlanabilir.
- Öğrenme oranı (Learning rate): Eğitim sırasında modelin ağırlıklarının ne kadar ayarlanacağını belirleyen bir hiperparametredir.

Bu parametrelerin optimize edilmesi, eğitim sırasında optimum model performansının elde edilmesinde çok önemlidir. Bu parametrelerin dikkatli bir şekilde ayarlanması ve seçilmesi, önceden eğitilmiş modellerin daha iyi genelleştirilmesine ve tahmin doğruluğuna yol açabilir.



Şekil 4.35 Önerilen modele göre iterasyon sayısına bağlı olarak test ve eğitim aşamalarında elde edilen kayıp değerlerinin grafiksel gösterimi

Şekil 4.35'te test ve eğitim aşamalarında elde edilen kayıp değerlerinin iterasyon sayısına göre grafikleri gösterilmektedir. Görselden de anlaşılacağı üzere, hem test hem de eğitim aşamaları için genel kayıp değerleri iterasyon sayısı arttıkça azalmaktadır. Ayrıca, 100 iterasyonun ötesinde, kayıp değerleri sabit kalmaktadır, bu da modelin eğitiminin o noktada optimizasyona ulaştığını ve daha fazla iterasyonun kayıp değeri üzerinde fark edilebilir bir etkisi olmadığını göstermektedir.



Şekil 4.36 Kodlayıcılara göre ROC Eğrisi kullanılarak ilaç-protein hedef etkileşimi tahmin yönteminin performansları

Değerlendirmemizde, özellikle tamamlanmamış proteinler için ilaç-protein etkileşimlerini tahmin etmede yöntemimizin doğruluğunu değerlendirmek için bir araç olarak alıcı çalışma karakteristiği (ROC) eğrisini kullanıyoruz. ROC eğrisi, optimum kesme noktasını belirlemek ve test doğruluğunu değerlendirmek için tanı testlerinde yaygın olarak kullanılan bir araçtır. Şekil 4.36'da önerilen model algoritmanın ROC eğrisi sunulmakta ve hesaplanan tahminlere dayalı olarak ROC Eğrisi Altındaki Alan (AUC ROC) değeri hesaplanmaktadır. ROC eğrisi, ilaç-protein etkileşimlerinin varlığını veya yokluğunu belirlememizi sağlayan 1-spesifikliğe (yanlış pozitif oran-FPR) karşı duyarlılığı (gerçek pozitif oran-TPR) çizer. Daha yüksek bir AUC ROC değeri, modelimizin daha iyi tahmin performansını gösterir (Hand, 2009; Brown,

2018). ROC eğrisi analizinin formülasyonu aşağıdaki Denklem (4.21) ve (4.22)'de verilmiştir.

$$TPR = \frac{TP}{TP+NP} \quad (4.21)$$

$$FPR = \frac{FP}{FP+TN} \quad (4.22)$$

- Gerçek Pozitifler (TP): Bunlar modelin pozitif sınıfı doğru tahmin ettiği ve gerçek sınıfın da pozitif olduğu durumlardır.
- Yanlış Pozitifler (FP): Modelin pozitif sınıfı yanlış tahmin ettiği, ancak gerçek sınıfın aslında negatif olduğu durumlardır.
- Gerçek Negatifler (TN): Bunlar modelin negatif sınıfı doğru tahmin ettiği ve gerçek sınıfın da negatif olduğu durumlardır.
- Yanlış Negatifler (FN): Bunlar modelin negatif sınıfı yanlış tahmin ettiği ancak gerçek sınıfın aslında pozitif olduğu durumlardır.

Çalışma, DTI araştırmaları için ilaçlar için dört kodlayıcı ve proteinler için dört kodlayıcı kullanan basit ve etkili bir çerçeve sunmaktadır. Metodoloji aşağıdaki adımları içermekte olup, her adım karşılık gelen bir kod satırı ile temsil edilmektedir:

- Kodlayıcı spesifikasyonu: Hem ilaçlar hem de proteinler için kodlayıcıların seçilmesi ve tanımlanması.
- Veri kodlama ve bölümlendirme: İlaç ve protein verilerinin uygun formatlara dönüştürülmesi ve veri setinin eğitim, doğrulama ve test setlerine bölünmesi.
- Model yapılandırması oluşturma: DTI modelinin mimarisini ve yapılandırmasını ayarlama.
- Model başlatma: DTI modelinin seçilen kodlayıcılar ve konfigürasyonlarla başlatılması.
- Model eğitimi: Modelin parametrelerini optimize etmek için eğitim veri kümesini kullanarak modeli eğitme.

- Model tahmini ve yeniden yapılandırma/eleme: İlaç-protein etkileşimlerini tahmin etmek için eğitilmiş modeli kullanma ve gerekirse sonuçları daha da iyileştirme.
- Model kaydetme ve yükleme: Eğitilmiş modelin ileride kullanılmak üzere saklanması ve gerektiğinde yüklenmesi.

Bu adımlar sayesinde çalışmamız, eksik protein verileri için bile ilaç-protein etkileşimlerinin doğru bir şekilde tahmin edilmesine olanak tanıyan DTI araştırmalarına kapsamlı ve pratik bir yaklaşım sağlamayı amaçlamaktadır. Bu yöntem, analiz süreci için açık ve basit bir çerçeve sağlayarak ilaç-protein etkileşimlerini keşfetmek ve değerlendirmek için kullanılabilir.

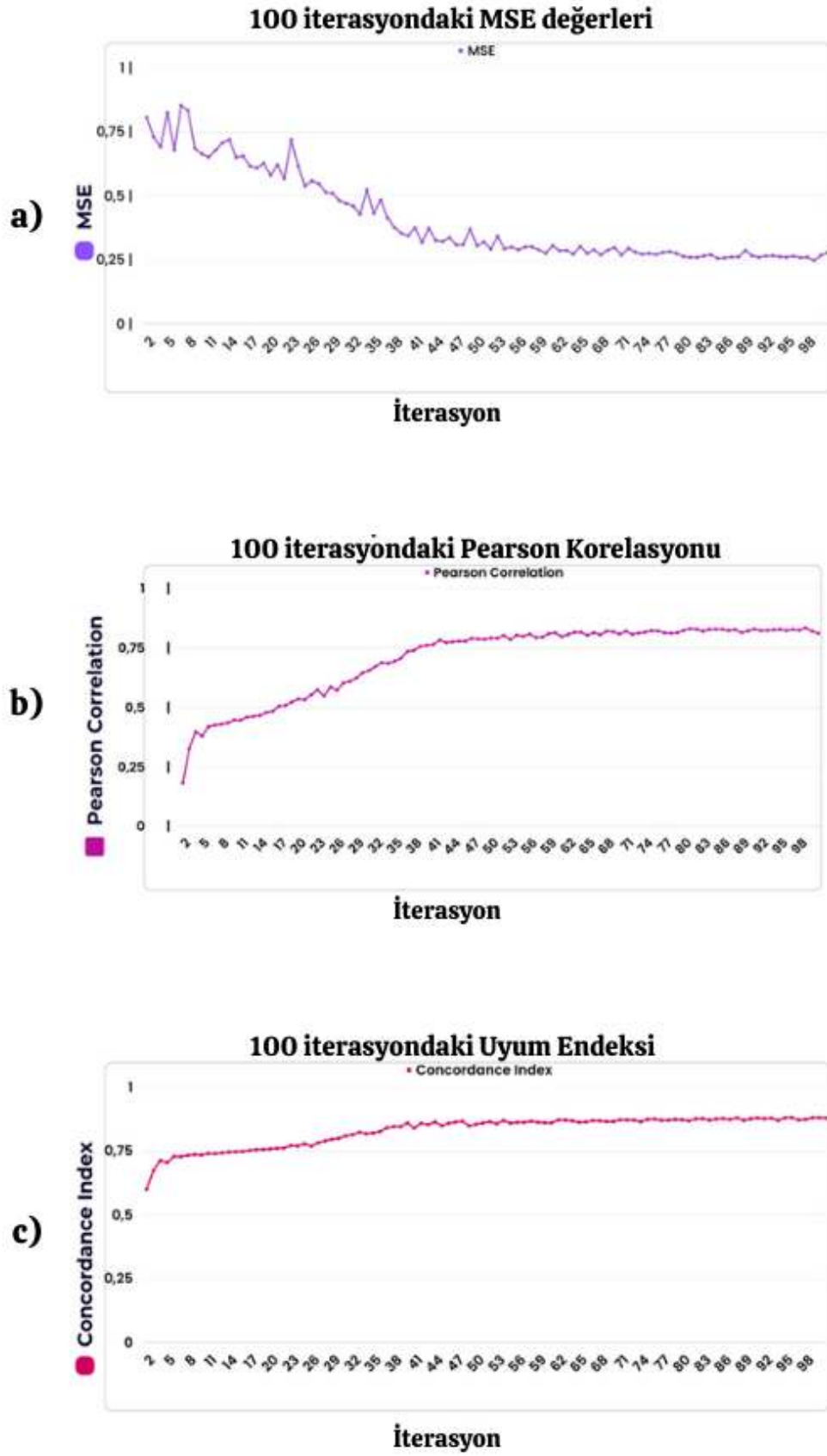
(Kodlayıcı spesifikasyonu): Gerekli veri formatını elde ettikten sonra, bunları kodlayıcılar için hazırlamamız gerekir. Bu nedenle ilaçlar ve proteinler için kullanılacak kodlayıcılar belirlenmiştir. Kullanılan kodlayıcı çiftleri GPU'lara göre seçilmiştir.

Çalışmamızda en dikkat çekici fark, ilaç SMILES ve protein dizilerinin kodlanması için kullanılan kodlayıcıların seçiminde yatmaktadır. İlaç SMILES'leri için, ilaç bileşiklerinin farklı temsillerini sağlayan (1) CNN, (2) Morgan, (3) MPNN ve (4) Daylight olmak üzere dört farklı kodlayıcı kullandık. Öte yandan, protein dizilerini kodlamak için, protein yapılarının benzersiz özelliklerini yakalayan (1) CNN ve (2) AAC tekniklerini kullanılmıştır.

Bu kodlayıcıları birleştirerek, ilaçlar ve proteinler arasındaki etkileşimleri etkili bir şekilde yakalayabilen bir dizi tahmin modeli oluşturduk. Bu kombinasyon, farklı kodlayıcıların güçlü yönlerinden yararlanmamızı ve ilaç-protein etkileşimleri hakkında daha kapsamlı bilgiler elde etmemizi sağlıyor. Ayrıca, model yapılandırması için çeşitli hiperparametreleri (örn. öğrenme hızı, epoklar, yığın boyutu) ve model parametrelerini (örn. gizli boyutlar, filtre boyutu) kolayca özelleştirmemizi sağlayan bir işlev geliştirdik. Bu esneklik, sınırlı yapılandırılabilirliğe sahip hazır yazılım paketlerinin kullanılmasına kıyasla model eğitim sürecini daha hızlı ve daha kullanışlı hale getirmektedir. Genel olarak, ilaçlar ve proteinler için çoklu kodlayıcıların

kullanımı ve model yapılandırmasındaki esneklik, ilaç-protein etkileşimlerini tahmin etmede yaklaşımımızın etkinliğine ve verimliliğine katkıda bulunan temel unsurlardır.

Şekil 4.37, modelin 100 iterasyon boyunca eğitilmesinden sonra elde edilen Ortalama Kare Hata (MSE) (a), Pearson Korelasyon (b) ve Uyum İndeksi (c) değerlerini göstermektedir. Bu performans ölçümleri ve değerlendirme puanları çeşitli ilaç-protein etkileşimleri için verilmiştir. MSE, tahmin edilen ve gerçek değerler arasındaki ortalama kare farkını ölçer ve tahmin doğruluğunun bir ölçüsü olarak hizmet eder. Pearson Korelasyonu iki değişken arasındaki doğrusal ilişkiyi değerlendirirken, Uyum Endeksi sıralama düzenindeki uyumu değerlendirir. Bu değerler, modelin performansını değerlendirmek için kullanılan temel ölçütleri temsil eder.



Şekil 4.37 Ortalama Kare Hata (MSE) (a), Pearson Korelasyonu (b) ve Uyum Endeksi (c) 100 iterasyon için elde edilen sonuçlar

Şekil 4.37 incelendiğinde, 100 iterasyon sonrasında değerlerin aynı kaldığı görülmektedir. Bu bulgu, daha fazla iterasyon gerçekleştirmenin modelin sonuçlarında önemli bir değişikliğe yol açmadığını göstermektedir.

Bu bölümde, klorpromazinin COVID-19 patogenezi üzerindeki etkisini anlamak için moleküler düzeyde bir analiz gerçekleştirdi. Hesaplamalı modelleme ve simülasyonlar GSA, CNN ve MLP gibi çeşitli ML teknikleri kullanılarak gerçekleştirilmiştir. Sonuçlar, klorpromazinin SARS-CoV-2'nin hedef proteinlerinden biri olan 6LU7 ile etkileşime girebileceğini ve potansiyel olarak inhibitör etkiler sergileyebileceğini göstermektedir.

6LU7 proteazı, COVID-19 ile mücadeleye yönelik antiviral ilaç geliştirmede çekici bir potansiyel hedef olarak araştırmanın ana odak noktası olmuştur. Bu hedef proteazı etkili bir şekilde inhibe edebilen ve potansiyel olarak hastalık için tedavi görevi görebilen ilaç adaylarını belirlemek için çeşitli çalışmalar yapılmıştır. Abdelrheem ve arkadaşları, caulerpin adı verilen doğal bir bileşiğin SARS-CoV-2'nin ana proteazına karşı inhibitör potansiyelini araştırdı. Bulguları, karnabaharın bir inhibitör olarak ümit verici olduğunu ve COVID-19 tedavisi için potansiyel bir aday olabileceğini gösterdi (Abdelrheem vd., 2020). Peele ve arkadaşları, SARS-CoV-2'ye karşı antiviral ilaç adaylarını belirlemek için silico tarama yöntemlerini kullandı. Hesaplamalı analizleri aracılığıyla, ana proteaz (6LU7) ile umut verici etkileşimler gösteren lopinavir, amodiakin ve theaflavin digallat (TFDG) olmak üzere üç bileşik tanımladılar. Bu bileşiklerin virüse karşı etkili antiviral ajanlar olarak potansiyele sahip olduğu öne sürüldü (Peele vd., 2020). Prateeksha ve arkadaşları, kriptogamik ikincil metabolitlerin (CSM'ler) SARS-CoV-2'nin 3CLpro ve spike glikoprotein reseptör bağlama alanı (RBD) hedeflerine karşı inhibitör potansiyelini araştırdı. Çalışmaları, CSM'lerden biri olan Marchantin E'nin proteaza karşı etkili bir inhibitör aday olarak potansiyel sergilediğini ortaya çıkardı. Bu çalışmalardaki moleküler bağlanma analizi, yerleştirme simülasyonları ve moleküler etkileşimleri tahmin etmek için yaygın olarak kullanılan bir paket program olan Autodock 4.2 yazılımı kullanılarak gerçekleştirilmiştir (Prateeksha vd., 2021).

Bu bilgi, COVID-19'un ortaya çıkmasından bu yana SARS-CoV-2'ye yönelik birçok küçük moleköl inhibitörün belirlendiğini ve incelendiğini belirtiyor. Bunlar arasında remdesivir, molnupiravir ve nirmatrelvir (PF-07321332) gibi sınırlı sayıda inhibitör, bu çalışmanın tarihinden itibaren COVID-19 tedavisi için onay almıştır (Hou vd., 2023). Bununla birlikte, bu onaylanmış seçeneklere rağmen, SARS-CoV-2 için daha büyük potansiyele, düşük toksisiteye ve daha iyi biyoyararlanıma sahip inhibitörlere yönelik kritik bir ihtiyaç bulunmaktadır. Yeni ve daha etkili inhibitörlerin sürekli aranması, COVID-19 için daha güvenli ve erişilebilir tedaviler geliştirme isteğiyle motive edilmektedir. İdeal inhibitörler, virüsü hedeflemede etkili olmalı, insan vücudu üzerindeki olumsuz etkileri minimuma indirmeli ve iyi emilim ve dağılım özelliklerine sahip olmalıdır, böylece geniş kapsamda uygulanabilirlik ve bulunabilirliklerini sağlamalıdır.

Bu çalışmanın bulguları, klorpromazin gibi mevcut ilaçların yeniden değerlendirilmesi ve COVID-19 gibi salgınlarda potansiyel tedavilerin hızla belirlenmesi için çok önemli bir strateji olabilir. Ancak bu aşamada dikkatli olunmalıdır. Klorpromazin gibi antipsikotik ilaçlar yan etkileri ve güvenlik profilleri nedeniyle dikkatle değerlendirilmelidir. Klorpromazinin COVID-19 tedavisindeki potansiyelini açıklığa kavuşturmak için daha fazla deneysel çalışmaya ve klinik çalışmalara ihtiyaç vardır. GSA'ların ilaç-protein etkileşimlerini tahmin etme ve bağlanma afinitelerini hesaplama yeteneği, ilaç keşfinde yeni yollar açmıştır. Büyük veri kümelerini analiz ederek ve karmaşık moleküler etkileşimleri anlayarak, GSA'lar daha kesin ve özel analizlere olanak tanır, potansiyel ilaç adaylarının değerlendirilmesine yardımcı olur ve yeni terapötik yaklaşımların geliştirilmesini kolaylaştırır. Genel olarak, GSA'lar ve diğer ML teknikleri, araştırmacılara geniş kimyasal alanı keşfetmeleri ve COVID-19 da dahil olmak üzere çeşitli hastalıklar için umut verici ilaç adaylarını belirlemeleri için güçlü araçlar sunarak ilaç keşif sürecinde devrim yaratmıştır.

Elde edilen sonuçlar, klorpromazinin COVID-19 tedavisindeki potansiyel rolünü göstermektedir. Ancak klorpromazin kullanımına ilişkin bazı önemli noktaların dikkate alınması gerekir. Öncelikle klorpromazin antipsikotik bir ilaç olarak biliniyor. COVID-19 tedavisinde kullanımıyla ilgili olarak sadece antiviral etkisinin değil aynı zamanda antiinflatuar ve immünomodölatör etkilerinin de dikkate alınması

önemlidir. Klorpromazin, proinflatuar sitokinlerin salınımını azaltabilir ve bağışıklık tepkisini düzenleyebilir. Bu özellikler, COVID-19 hastalarında bağışıklık tepkisinin kontrol edilmesinde ve sitokin fırtınaları gibi zararlı yan etkilerin azaltılmasında potansiyel bir avantaj sunabilir.

Ancak klorpromazinin COVID-19 tedavisinde kullanımına ilişkin bazı önemli endişeler bulunmaktadır. İlk olarak, klorpromazin gibi antipsikotik ilaçların sıklıkla ciddi yan etkileri olabilir. Bu yan etkiler arasında sedasyon, kardiyak aritmiler, ekstrapiramidal etkiler ve nörolojik sorunlar yer alır. Bu nedenle, COVID-19 hastalarında klorpromazin kullanımını düşünmeden önce yan etkilerin dikkatli bir şekilde değerlendirilmesi gerekmektedir.

İkincisi, klorpromazinin COVID-19 tedavisindeki etkinliği ve güvenliği konusunda yeterli klinik veri eksikliği bulunmaktadır. Bu çalışma, klorpromazinin potansiyel etkisini moleküler düzeyde analiz yoluyla göstermiştir. Ancak ilacın etkinliğinin ve güvenliğinin gerçek hastalarda değerlendirilmesi için bu bulguların klinik çalışmalarla desteklenmesi gerekiyor.

Ayrıca klorpromazin gibi mevcut ilaçların yeniden değerlendirilmesi, yeni tedavi stratejilerinin hızla bulunması açısından önemli bir yaklaşım olabilir. Ancak klorpromazin gibi mevcut ilaçların COVID-19 tedavisinde kullanımı yasal ve düzenleyici gerekliliklere uygun olmalıdır. İlacın yeniden kullanımı için gerekli izinlerin alınması, dozajın ve uygulamanın belirlenmesi ve yan etkilerin yönetilmesi dikkatle ele alınmalıdır.

Sonuç olarak bu çalışma, klorpromazinin SARS-CoV-2'nin 3CL proteazı ile bağlanma etkileşimlerini analiz ederek, klorpromazinin COVID-19 tedavisine aday olma potansiyeline ışık tutuyor. Ancak klorpromazin gibi antipsikotik ilaçların COVID-19 tedavisinde kullanımını düşünmeden önce yan etkileri ve güvenlik profilini dikkate almak önemlidir. COVID-19 hastalarında etkinliğini ve güvenilirliğini değerlendirmek için daha fazla klinik veriye ve titiz çalışmalara ihtiyaç vardır. Çalışma, COVID-19 için etkili tedavi arayışında klorpromazin gibi mevcut ilaçların geçerli bir strateji olarak yeniden değerlendirilmesinin önemini altını çiziyor.

GSA'lar gibi gelişmiş ML tekniklerinin kullanımının, potansiyel ilaç adaylarının belirlenmesinde ve ilaç keşif sürecinde karmaşık moleküler etkileşimlerin anlaşılmasında etkili olduğu kanıtlanmıştır. Gelecekteki araştırmalar, klorpromazinin COVID-19 tedavisindeki tam potansiyelini keşfetmek için bu çalışmanın bulgularına dayanabilir. Ek olarak, etki mekanizmasını ve güvenlik profilini COVID-19 hastaları bağlamında araştırmak, terapötik etkinliği hakkında değerli bilgiler sağlayabilir. ML tekniklerinin ilaç keşfine entegrasyonu, yeni ilaç adaylarının tanımlanmasını hızlandırabilir ve hastalıklara dair anlayışımızı geliştirerek daha etkili ve hedefe yönelik terapötik müdahalelere yol açabilir.

Bu bölümde, GSA tabanlı moleküler modelleme ve ML teknikleri kullanılarak Klorpromazin ilacının 6LU7 proteine bağlanma değeri hesaplanmıştır ve ilaç-etkileşim analizi için önemli bilgiler sağlanmaktadır. Çalışmanın hedefleri arasında ilaç etkileşim afinitesinin belirlenmesi, moleküler etkileşim modelinin oluşturulması, ML algoritmalarının performansının değerlendirilmesi, ilaç tasarımına yönelik bilgi sağlanması ve hazır modelleme paketlerine ücretsiz bir alternatif olması yer almaktaydı. Bu çalışmanın sonuçları, ilaç-etkileşim analizi ve ilaç tasarımı alanlarında değerli bir bilimsel birikim sunacak ve ilgili sektörlere önemli katkılar sağlayabilecektir.

Sonuç olarak çalışmadan erişilen çıktılar şunlardır:

- Klorpromazin ilacının 6LU7 proteine bağlanma afinitesi değerinin belirlenmesi: Proje, ilacın 6LU7 proteine bağlanma derecesini nicel olarak hesaplanmıştır. Bu bilgi, ilacın etkisini ve etkileşim mekanizmasını daha iyi anlamamıza yardımcı olacaktır.
- İlaç-etkileşim analizi için moleküler etkileşim modelinin oluşturulması: Projede kullanılan algoritmalar, ilacın 6LU7 proteiniyle olan etkileşimini modellemek için kullanılmıştır. Bu, ilacın protein üzerindeki etkileşim mekanizmasını ve bağlanma bölgesini belirlememize yardımcı olacaktır.
- İlaç tasarımı için hedefe yönelik bilgi sağlanması: Projenin sonuçları, ilacın hedef proteine olan bağlanma özelliklerini analiz ederek, ilaç

tasarımı sürecinde kullanılabilir bilgiler sunmaktadır. Bu, daha etkili ve spesifik ilaçların geliştirilmesine katkıda bulunabilir.

- ML algoritmalarının performansının değerlendirilmesi: Projede kullanılan farklı algoritmaların performansı karşılaştırılmıştır ve hangi algoritmanın en iyi sonuçları verdiği değerlendirilmiştir. Bu, ilaç-etkileşim analizi alanında gelecekteki çalışmalara rehberlik edebilecek ve daha iyi yöntemlerin geliştirilmesine katkı sağlayabilecektir.
- İlaç-etkileşim analizinde kullanılan modelleme paketlerine göre avantajların belirlenmesi: Projenin sonuçları, hazır modelleme paketlerine kıyasla GSA, CNN ve MLP algoritmalarının kullanımının avantajlarını ve farklarını değerlendirmektedir. Bu, ilgili araştırmacılar ve endüstri uzmanları için ilaç tasarım süreçlerinde daha etkili ve verimli bir yaklaşım sunabilir.

4.4 Bölüm 4: İlaç yapı analizleri

4.4.1 Yapay zekâ ve ilaç keşfi: ebola tedavisi için yedi farklı ilacın benzerlik ve moleküler yapı analizi

İlaç keşfi, çeşitli hastalıkların tedavisi için yeni ilaçların tanımlanmasını ve geliştirilmesini içeren karmaşık ve zaman alıcı bir süreçtir. Belirli bir hastalıkla ilişkili hedef proteinlerin tanımlanmasını, bu hedeflerle etkileşime girebilecek bileşiklerin geliştirilmesini ve daha sonra bu bileşiklerin insan kullanımı için güvenli ve etkili olduklarından emin olmak için kapsamlı testler yapılmasını içerir. Çeşitli hastalıkları tedavi etmek için yeni ilaçlar keşfetme süreci, tipik olarak 15 yıl kadar süren ve milyarlarca dolara mal olan uzun ve zorlu bir yolculuktur. Maliyetlerin nasıl azaltılacağı ve yeni ilaç keşfinin nasıl hızlandırılacağı, sektör için zorlu ve acil bir soru haline gelmiştir.

Şu anda ilaç endüstrisi, artan Ar-Ge maliyetleri ve azalan verimlilik nedeniyle ilaç geliştirme programlarını sürdürmede zorluklarla karşı karşıya. Bununla birlikte, YZ ve ML'deki ilerlemelerle, ilaç keşfi alanı, sürecin önemli ölçüde hızlandırılacağı ve başarı oranının artırılacağı yeni bir döneme giriyor.

Son yıllarda, YZ ilaç keşfi alanında yaygın olarak benimsenmiş ve yeni terapötik bileşiklerin tanımlanması ve etkinliklerinin optimize edilmesi için muazzam bir potansiyel göstermiştir. İlaç keşfinde YZ kullanımı, süreci kolaylaştırarak yeni ilaçlar keşfetmek için gereken zaman ve kaynak miktarını azaltabilmiş ve daha önce tedavi edilemez olduğu düşünülen hastalıklar için potansiyel tedavilerin belirlenmesinde büyük umut vaat etmiştir. YZ, karmaşık verileri analiz etmek, yeni terapötik bileşikleri tanımlamak ve etkinliklerini optimize etmek için yeni yollar sunarak ilaç keşfinde güçlü bir araç haline gelmiştir. YZ algoritmaları, büyük miktarda biyolojik ve kimyasal veriyi hızlı bir şekilde analiz ederek araştırmacıların yeni ilaç hedeflerini belirlemelerine ve yeni bileşikleri her zamankinden daha hızlı geliştirmelerine olanak tanır. Ayrıca yapay zekâ, ilaç tasarım sürecini optimize etmeye yardımcı olarak klinik deneylerde daha yüksek bir başarı oranı sağlayabilir.

Bu bölümde, daha önce giriş bölümünde anlatılan, YZ ve ML tekniklerini kullanarak ilaç keşfinin mevcut durumuna genel bir bakış sağlayacaktır. İlaç keşfinde kullanılan en son teknolojiler ve algoritmaların yanı sıra bu alanda YZ kullanımının zorlukları ve sınırlamaları da ele alınacaktır. Ayrıca bu kitap, YZ ve ML kullanan başarılı ilaç keşif projelerinin vaka çalışmalarını ve bunların tıp alanında yarattığı etkiyi keşfedecektir. Ayrıca, bu bölümde Ebola tedavisi için yedi farklı ilacın benzerlik ve moleküler yapı analizi çalışmaları sunulacaktır. YZ'nin yeni ilaç hedeflerini belirlemek, yeni bileşikler tasarlamak ve bunların klinik deneylerdeki güvenliğini ve etkinliğini optimize etmek için nasıl kullanıldığını inceleyeceğiz.

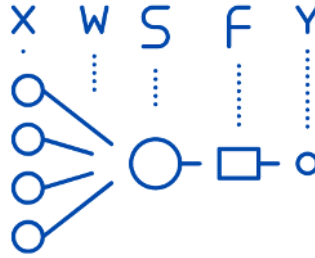
YZ yeni deneysel teknolojilerle birleşmesinin yeni ilaç arayışını daha hızlı, daha ucuz ve daha etkili hale getirmesi beklenmektedir. Bu durum, gelecekte ilaç endüstrisi devlerinin YZ destekli ilaç keşif şirketleri ile iş birliğini artıracaktır.

İster bir bilim insanı, araştırmacı olun, ister teknoloji ve tıbbın kesiştiği noktalara ilgi duyun, bu bölüm çalışmasında YZ kullanarak ilaç keşfi alanına dair değerli bilgiler sağlayacaktır. Bu bölümün ilaç keşfi alanında yeni fikirlere ve yeniliklere ilham vereceğini ve nihayetinde tüm insanlığın yararına daha etkili ve verimli tedavilerin geliştirilmesine yol açacağını umuyoruz.

Önerilen yöntem

Son birkaç yılda, GSA'lar, grafik ile modellenen her türlü problem için güçlü ve pratik araçlar haline gelmiştir. GSA temeli olan grafik teorisi, galaksilerden atom altı parçacıklara kadar neredeyse her şeyin ilişkilerini modelleyebilen son derece genel bir veri temsidir. Sosyal ağlar, şehirleri birbirine bağlayan yollar, beyin bağlantıları, protein ve DNA etkileşimleri, diğerlerinin yanı sıra, grafik yapılar olarak modellenen ilişki örnekleridir.

Genel olarak bir yapay sinir ağına bakıldığında ise biyolojik sinir sisteminden tasarlanan yapay sinir sistemi de birbirine bağlı bir yapıya sahiptir. Yapay sinir hücreleri (yapay nöronlar) oluşturdukları bağlantılar ile YSA'nın yapısını oluştururlar. Yapay sinir hücrelerinin işlem elemanları olarak adlandırılan beş temel elemanı vardır. Her yapay sinir hücresinin gelen bilgileri alan girdileri (x), gelen bilgileri işleyen ve bağlantıları oluşturan ağırlıkları (w), toplama fonksiyonu (S), aktivasyon fonksiyonu (f) ve işlenen bilgileri dış dünyaya sunan çıktıları (y) veya çıkış elemanları vardır. Şekil 4.38'de bir yapay sinir ağının genel yapısı gösterilmektedir.



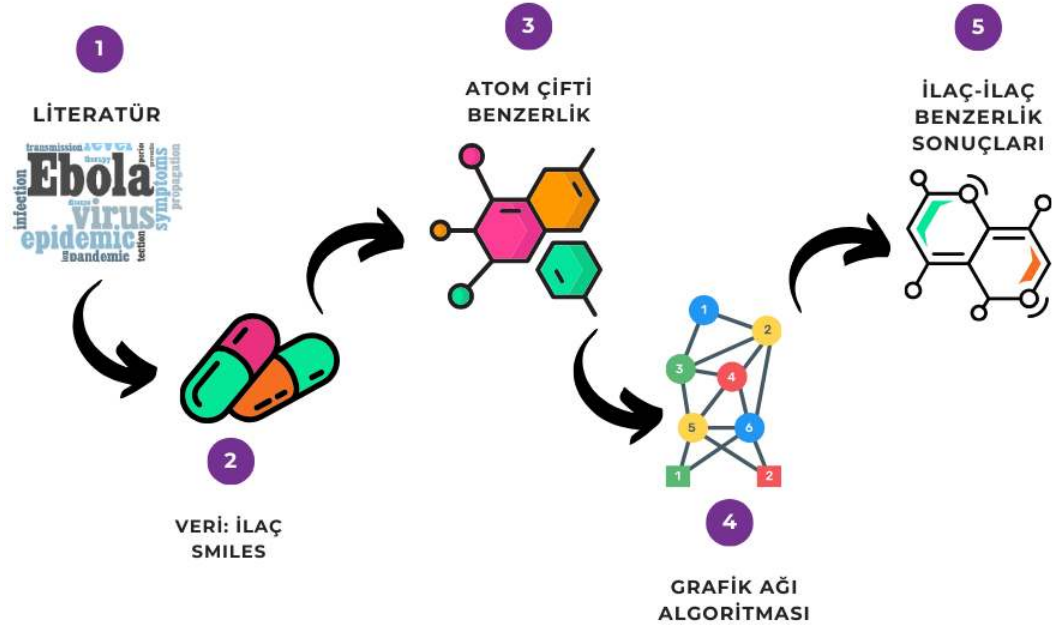
Şekil 4.38 Yapay sinir ağ yapısı

Bu yapay sinir ağının daha gelişmiş versiyonları olan grafik sinir ağlarının çeşitli alt türleri geliştirilmiştir. Genel olarak, bir molekül grafiksel (3B yapılar gibi), metinsel (yani moleküler formül) ve diğerleri dahil olmak üzere birçok temsile sahip olabilir. Bir molekülü temsil etmenin bir yolu, moleküler modelleme kitinin bir tür soyut versiyonu olarak çalışan grafikleri kullanmaktır.

Buradaki amaç, son zamanlarda YZ alanında giderek daha başarılı hale gelen grafiksel bilgiye dayalı grafiksel sinir ağlarının ilaç keşfi, geliştirilmesi ve analizi üzerine araştırmalar yapmaktır. Daha sonra bu araştırmalar sonucunda bilgisayar destekli ilaç

analizi sektörüne, ilaç-ilaç benzerliği, ilaç-hedef protein etkileşimi, ilaç-ilaç etkileşimi, ilaç tasarımı, ilaç analizi gibi konularda farklı yaklaşım önerileri ile katkı sağlamaktır. Bu nedenle grafik bilgisine dayanan Scaffold grafik ağı algoritması ve ilaç-ilaç benzerlik algoritmalarından AP algoritması kullanılarak Ebola tedavisi için ilaç araştırması yapılmıştır. Bu bölümde, Ebola tedavisi için araştırılmakta olan veya klinik test aşamasında olan ilaçlar analiz edilmiştir.

Ebola virüsü hastalığına yönelik tedavilerin daha önceki virüslerin hastalık belirtilerini tedavi etmek üzere tasarlanmış ilaçlarla yapılıyor olması, FDA onaylı ilaçların yeniden kullanılmasını haklı çıkarmaktadır. Bu çalışma, FDA onaylı ilaçların benzerliklerini çıkararak ve moleküler iskeletlerini analiz ederek hızla yayılan Ebola virüsü için yeniden tasarlanan ilaçları kullanarak salgınla mücadeleye mümkün olan en kısa sürede yardımcı olmayı amaçlamaktadır. Bu amaçla uygun ilaç adaylarının yeniden belirlenmesi için önerilen metodoloji Şekil 4.39'da gösterilmektedir.



Şekil 4.39 Ebola tedavisinde ilaç-ilaç benzerliğini ve moleküler iskeleleri analiz etmek için önerilen yaklaşımın akış şeması

DL modellerinin çok çeşitli alanlardaki son başarısı nedeniyle, genellikle YZ'nin üçüncü dalgasında olduğumuz söylenir. Son zamanlarda, GSA veya daha genel olarak Grafik Ağlarının (GN) kullanımında bir artış olmuştur. Grafik tabanlı moleküler

tasarımın arkasındaki yol gösterici ilke, istenen ilaç benzeri moleküllerin tüm kriterlerini karşılayan grafikler oluşturmak olarak düşünülebilir. Çoklu doğrusal grafiklerin toplanması bir iskele ağacı oluşturur. Bir iskele ağacı, bir molekülün alt iskelelerin doğrusal bir düzenlemesi ile temsil edildiği yöntemlerden biridir (Mercado vd., 2021). Çekirdek yapılar olarak belirli iskelelere sahip moleküller tasarlamak, potansiyel ilaç adayları elde etmenin etkili bir yoludur (Y. Li vd., 2020). İskele kavramı her yerde bulunmasına rağmen, molekülün hangi kısmının çekirdek olarak kabul edilmesi gerektiği konusunda belirsizlik varsa, bir iskelenin tanımı silico'da öznel olabilir. Bemis ve Murcko tarafından yapılan popüler tanım, bir iskeleyi halkaların ve onları birbirine bağlayan atomların bir kombinasyonu olarak tanımlayarak bu belirsizlikleri göz ardı etmektedir. Belirsizliği azaltmak için, bir molekülün, Bemis ve Murcko yapısından çevresel halkaların iteratif olarak çıkarılmasıyla tanımlanan, birbiriyle ilişkili iskelelerden oluşan bir ağaç veya ağ ile temsil edilebileceği önerilmiştir (Bemis & Murcko, 1999).

Bu çalışmada, ilk aşaması AP benzerlik analizi kullanılarak gerçekleştirilmiştir. Bir sonraki adımda, bu ilaçların moleküler iskeleleri çıkarılmış ve İskele Ağ Grafiği algoritması ile grafiksel olarak temsil edilmiştir. Tüm bu sonuçlar göz önünde bulundurulduğunda, başta favipiravir ve molnupiravir olmak üzere diğer ilaçların kullanılmasının önerilmesinin, Ebola virüsünün tedavisinde klinik ve bilgisayar analiz süreçlerine yardımcı olabileceğini öngörüyoruz. Bu aynı zamanda hedefe yönelik tedavi için hızlı bir yol haritası sağlayacaktır.

İlaç analizi uygulamaları için algoritmalar

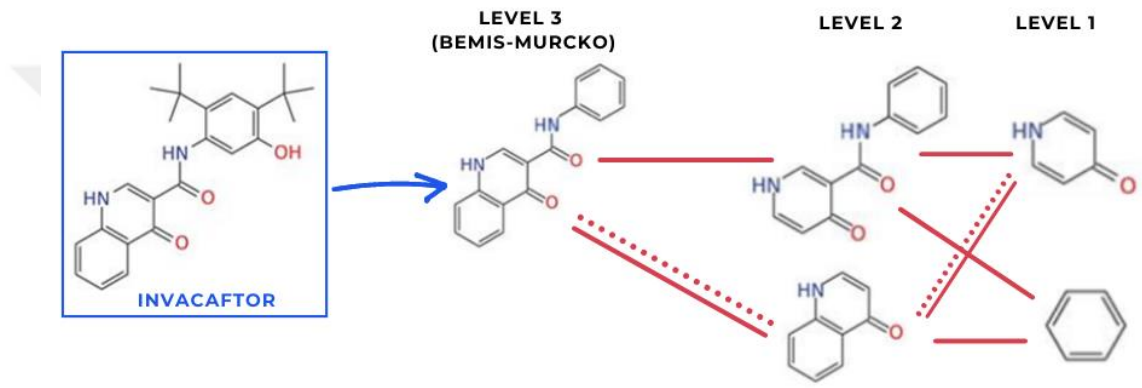
a) Atom çifti benzerlik algoritması

Bu bölümde, AP Benzerlik analizi ile ilgili olarak daha önce Bölüm 4.1.1'de detaylı bir şekilde açıklanan bilgilere başvurulmuştur. AP benzerlik analizinin temel mantığı ve uygulamaları hakkında detaylı bilgi için Bölüm 4.1.1'e bakınız.

b) Scaffold grafik ağı algoritması

İlaç keşfi ve geliştirilmesinde kullanılan yeni bir komut satırı aracı olan SNG özellikle büyük molekül kümelerinden moleküler yapı ağlarını ve grafiklerini çıkarmak için kullanılır. Bu ağlar ve grafikler, ilaçlara dönüştürülebilecek iskeleleri tanımlamak için

kullanılabilir. Ayrıca, büyük kimyasal bilgi verilerini görselleştirmek ve analiz etmek için otomatik bir boru hattına dahil edilebilir (Matlock vd., 2013). İskelelerin belirlenmesi, tespit edilmesi ve analiz edilmesi öznedir çünkü bir molekülün yapısının hangi kısmının çekirdek yapı ve hangi kısımlarının yan zincir olarak kabul edilmesi gerektiği konusunda belirsizlik vardır. Bu alandaki ilk çalışma, bu tür belirsizlikleri ortadan kaldırmak veya göz ardı etmek ve her molekülü ideal olmayabilecek tek bir iskeleye atamak için bir algoritma kullanmaktadır - örneğin, Invacator'un Şekil 4.40'taki en üst seviyedeki Bemis-Murcko iskelesi.



Şekil 4.40 Invacator ilacı için iskele ağı, ilacın kompozit iskelelere ayrıştırılabileceği tüm olası yolları göstermektedir.

İskele ağı ile ilgili matematiksel ifadeler ve detaylı açıklamalar Bölüm 4.1.1'deki "Scaffold grafik ağı algoritması" başlığı altında yer almaktadır.

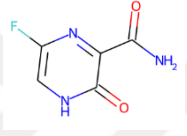
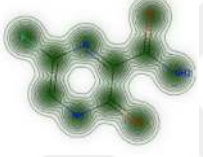
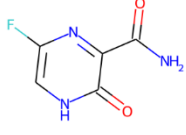
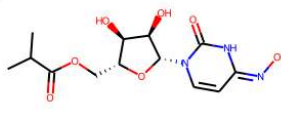
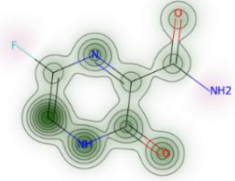
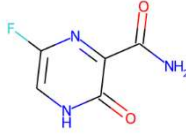
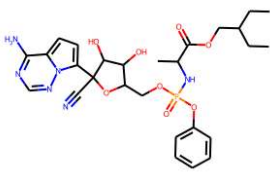
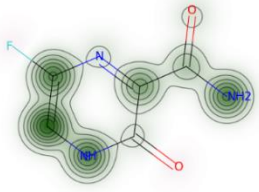
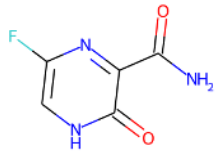
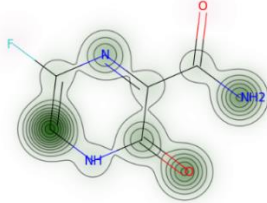
Bulgular 6:

Bu bölümde, veri olarak belirli ilaçlar seçilmiştir. Bu ilaçlar İnfluenza, COVID-19, Lassa ve Ebola gibi viral hastalıkların tedavisinde yaygın olarak kullanılan veya tavsiye edilen ve FDA onaylı ilaçlar oldukları için seçilmiştir. Bir diğer neden ise bu ilaçların Ebola tedavisi için araştırılıyor olması veya klinik ortamlarda kullanımının araştırılıyor olmasıdır. Bu nedenlerle çalışma bu ilaçlara odaklanmıştır. Çalışmanın bir stratejisi olan ilaç-ilac benzerlik analizi, ilacın klinikte sıklıkla kullanılması koşuluyla geçerli bir alternatif olarak düşünülebilir. Bunun için atom çifti parmak izi benzerlik algoritması ve iskele ağ grafiği algoritması kullanılmıştır. Ebola tedavisine uygun mevcut ilaçların yeniden kullanımı için ilaç-ilac benzerlik analizi yapılmıştır. İlaç-ilac

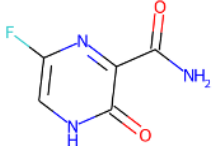
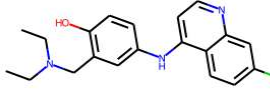
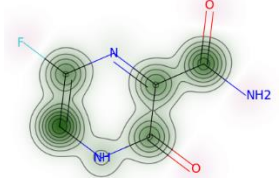
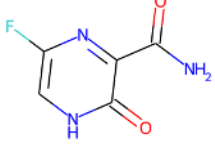
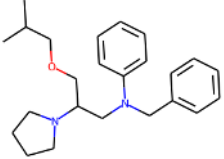
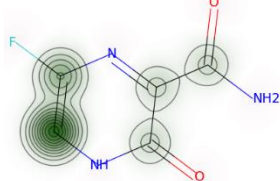
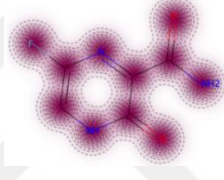
benzerliği için önerilen yöntemin AP benzerlik analizi sonuçları Tablo 4.15'te gösterilmiştir.

Denklem (4.2)'de belirtildiği gibi ilaç-ilaç benzerlikleri, Tablo 4.15'te görüldüğü gibi ilaçların ilaç moleküllerinin ne kadar yakın olduğunu göstermektedir. Birbirine çok benzeyen ilaçlar tamamen yeşil renkte görünmektedir.

Tablo 4.15 Atom çifti analizi ile favipiravir ilacının kimyasal yapısı ile diğer ilaçların benzerlik sonuçları

2B Referans İlaç	2B Eşleştirilen İlaç	2B Atom çifti benzerlik sonuçları	Benzerlik Değeri
			1.0
Favipiravir	Favipiravir	Favipiravir-Favipiravir	
			0.8638
Favipiravir	Molnupiravir	Favipiravir-Molnupiravir	
			0.5145
Favipiravir	Remdesivir	Favipiravir-Remdesivir	
			0.4921
Favipiravir	Brincidofovir	Favipiravir-Brincidofovir	

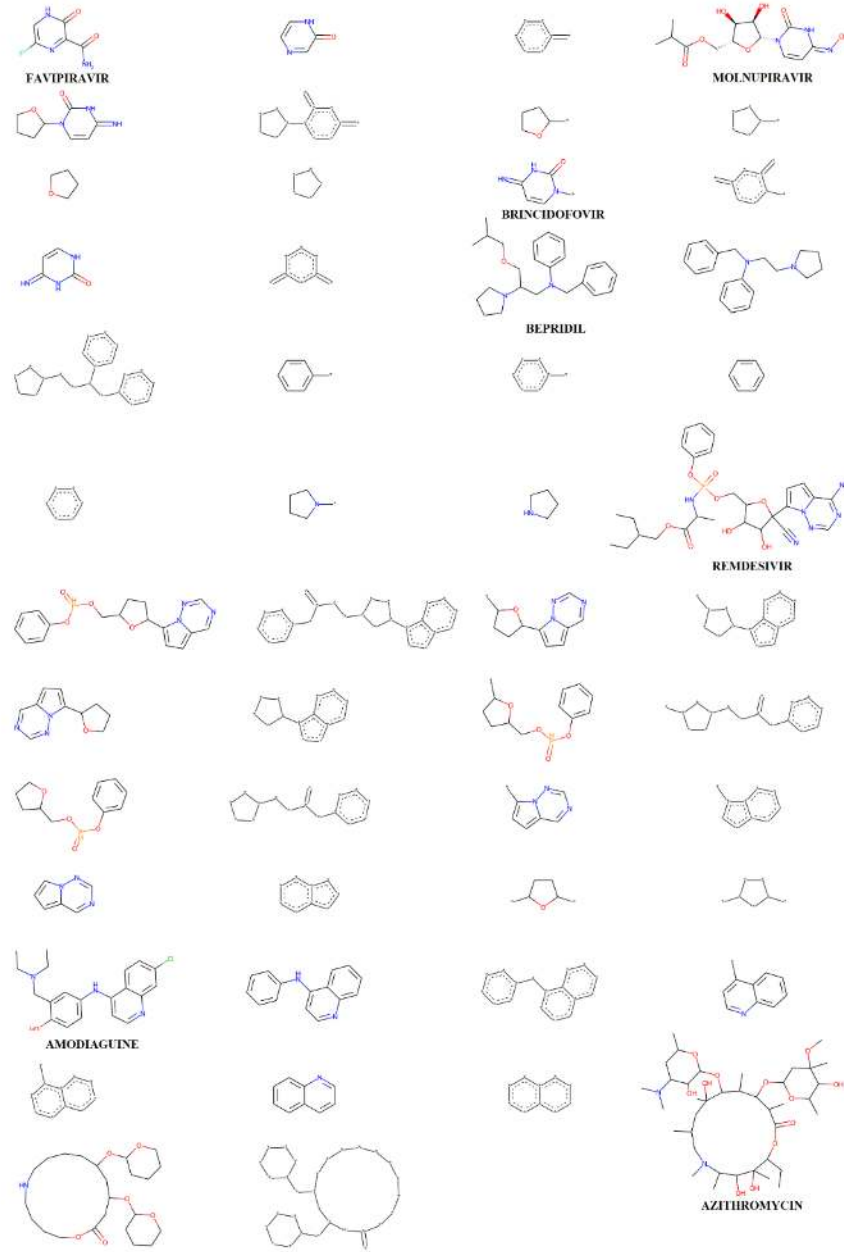
Tablo 4.15 Atom çifti analizi ile favipiravir ilacının kimyasal yapısı ile diğer ilaçların benzerlik sonuçları (Devamı)

			0.3682
Favipiravir	Amodiaquine	Favipiravir-Amodiaquine	
			0.1995
Favipiravir	Bepridil	Favipiravir -Bepridil	
			0.1295
Favipiravir	Azithromycin	Favipiravir-Azithromycin	

Tablo 4.15'te gösterilen yeşil renk, Favipiravir ilacının literatürden alınan Ebola'ya karşı diğer ilaçlarla ne kadar örtüştüğünü gösterir. Başka bir deyişle, yeşil renk iki ilacın ne kadar benzer olduğunu gösterir. Benzerlik haritaları, moleküler yapılar arasındaki parmak izi benzerliklerinin atomik olarak görselleştirilmesi için çok kullanışlı ve anlaşılması kolay bir stratejidir. Atomik ağırlıklar, ilgili atomun bitlerinin çıkarılmasıyla elde edilen benzerliğin bir önceki parmak izinin (değişmemiş) benzerliği ile karşılaştırılmasıyla elde edilir. Benzerlik haritaları her parmak izi için oluşturulabilir ve bitlerin karşılık gelen bir atoma veya alt yapıya kadar izlenmesine olanak tanır.

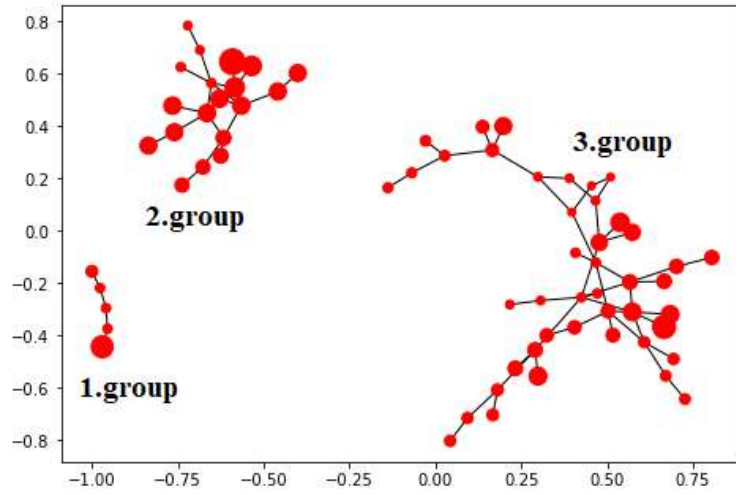
Tablo 4.15'teki sonuçlara göre, Favipiravir-Molnupiravir en yüksek benzerlik değerine sahiptir. Favipiravir-Azitromisin ise en düşük benzerlik değerine sahiptir. İkinci aşamada kullanılan iskele ağ grafiği, moleküler iskele ağları ve ağaçları oluşturmak ve analiz etmek için açık kaynaklı bir Python kütüphanesi komut satırı aracıdır. Çok sayıda girdi molekülünü işleme kapasitesine sahiptir. Şekil 4.41, Scaffold grafik ağı

algoritması kullanılarak oluşturulan Tablo 4.15'te listelenen ilaçların iskele yapılarını göstermektedir.



Şekil 4.41 Benzerliği çıkarılan 7 ilacın Scaffold grafik ağı algoritması tarafından oluşturulan bu ilaçların moleküler iskelelerinin temsili

Şekil 4.41'de, Tablo 4.15'teki atom çifti algoritması kullanılarak benzerlikleri de elde edilen yedi (7) ilacın moleküllerinin olası tüm iskeleleri çıkarılmıştır.



Şekil 4.42 Yedi (7) ilaca ait iskelelerin grafiksel gösterimi, hangi iskelenin hangi düğümle ilişkili olduğunu ve aralarındaki bağlantılarının gösterimi

Şekil 4.42, yedi ilacın moleküllerinin iskeletlerini grafiksel olarak göstermektedir. Burada toplam 68 düğüm bulunmaktadır. Bu da yedi ilaç için toplam 68 alt yapı ve yan zincir olduğunu göstermektedir. Bu düğümlerin sayısının yüksek olması, ilaçların benzerliğinin bir ölçüsü olarak alınabilir. İskele ağacı algoritması kullanılarak Şekil 4.42'deki ikinci gruba ait temel iskele ve yan zincirler Şekil 4.43'te gösterilmiş olup, burada düğümdeki moleküler yapı görüntüleri ve düğümler yeniden çizilmiştir.

Bilim ve tıp çevreleri, COVID-19 gibi salgınlarla mücadele etmek, enfekte hastaları tedavi etmek ve nihayetinde gelecekte yeni salgınları önlemek için YZ kullanıyor. YZ teknolojisi ilaç keşfinde kritik bir noktaya evrilerek tüm bu süreçleri daha hızlı ve verimli hale getirdi. Bu nedenle YZ destekli ilaç keşfi yapan girişimler önem kazanmıştır. İlaç keşfinde, ilaç analiz süreçlerinin gerçekleştirilmesi için benzerlik algoritmaları etkili bir süreçtir. Bunlardan biri de bir ilaç-ilaç benzerlik algoritması olan AP parmak izidir. Bu algoritma ile ilaç-ilaç benzerlikleri analiz edilerek Ebola tedavisi için şu anda araştırılmakta olan veya klinik çalışmalarda olan yeni potansiyel ilaç adayları bulunmuştur. Bu benzer ilaçların benzer alt yapı molekülleri ve çekirdek moleküler iskeleleri daha sonra SGN algoritması kullanılarak çıkarılmıştır. Kimyagerlerin bunları önceden bilmesi gerekir.

Şu anda klinik ortamda COVID-19 tedavisi için araştırılmakta olan ilaçlar veya ilaç adayları, Ebola virüsünün etkinliğini test etmek için klinik deneyler tasarlamının temelini oluşturabilir. Ebola tedavisi için bir kombinasyon oluşturmak, uygun dozajların ayarlanmasına ve uygulanan dozların azaltılmasına da yol açabilir. Listedeki tüm ilaçlar için farklı kombinasyonlar düşünülebilir.

5. SONUÇLAR VE ÖNERİLER

5.1 Sonuçlar

GSA ile ilaç keşfi üzerine yapılan bu çalışma, ilaç keşfi sürecinde önemli katkılar sağlayarak dört ana alanda sonuçlar sunmaktadır: salgın hastalıklar, ilaç çözünürlük tahmini, moleküler modelleme ve ilaç yapı analizleri.

Salgın hastalıklar kapsamında yapılan analizlerde, Ebola virüsünün tedavisi için grafik ağları ve benzerlik algoritmaları kullanılarak potansiyel ilaçlar belirlenmiştir. Bu yöntem, moleküler iskele analizi ile Ebola gibi acil durumlarda etkili tedavi stratejileri geliştirilmesine olanak tanımıştır. COVID-19 ve akciğer kanseri tedavisinde kullanılan ilaçların grafik sinir ağları ile yapılan benzerlik analizi, ilaç-protein bağlanma enerjisi hesaplamaları ile ilaçların potansiyel etkileşimlerini ve benzerliklerini ortaya koymuştur. GSA, özellikle COVID-19 gibi hızla yayılan salgınlarda, mevcut ilaçların yeniden kullanımı için kritik bir araç olarak önemli bir rol oynamaktadır.

İlaç çözünürlük tahmini konusunda, GSA, diğer hesaplamalı yöntemlere kıyasla daha düşük performans göstermiştir. GSA'nın moleküler yapıların detaylı tahminini sağlaması, ilaç geliştirme süreçlerinde daha etkili kararların alınmasına olanak tanımıştır.

Moleküler modelleme alanında, GSA'lar ile ilaç-protein etkileşimlerinin analizinde başarılı sonuçlar elde edilmiştir. Özellikle Chlorpromazine-6LU7 projesi kapsamında yapılan analizler, GSA tabanlı moleküler modellemenin ilaç-protein etkileşimlerini doğru bir şekilde tahmin edebildiğini göstermiştir. Bu bulgu, ilaç-protein etkileşimleri üzerine yapılan çalışmaların doğruluğunu artırarak yeni ilaç keşfi ve geliştirme süreçlerine katkıda bulunmaktadır.

İlaç yapı analizleri kapsamında, YZ ve GSA teknikleri kullanılarak yapılan moleküler yapı analizleri, ebola tedavisi için yedi farklı ilacın moleküler benzerliklerini detaylı bir şekilde ortaya koymuştur. Bu analizler, ilaç keşif süreçlerinde kullanılabilecek yeni yöntemlerin geliştirilmesine olanak tanımaktadır.

Genel olarak, bu çalışmanın sonuçları GSA'ların ilaç keşfi süreçlerinde geniş bir uygulama alanı bulduğunu ve önemli avantajlar sağladığını göstermektedir. GSA, ilaç-protein bağlanma tahmini, ilaç benzerliği analizi, moleküler yapı analizi ve ilaç çözünürlüğü gibi alanlarda etkin bir şekilde kullanılabilir. Bu yöntemler, ilaç keşfi sürecinin hızlandırılması ve maliyetlerin azaltılması açısından büyük potansiyel taşımaktadır. GSA'ların, ilaç keşfi ve geliştirme süreçlerinde daha kapsamlı ve etkili bir şekilde kullanılması önerilmektedir.

Bu bağlamda, GSA tekniklerinin geliştirilmesi ve mevcut yöntemlerle entegrasyonunun artırılması gerekmektedir. Ayrıca, GSA ile yapılan tahminlerin klinik deneylerle doğrulanması, bu yöntemlerin geçerliliğini artıracaktır. İlaç keşfi süreçlerinde akademik ve endüstriyel iş birliklerinin artırılması, GSA gibi ileri teknolojilerin yaygın kullanımını sağlayacaktır. GSA ve benzeri YZ tekniklerinin ilaç keşfi alanında kullanılmasının yaygınlaştırılması için akademik ve endüstriyel iş birlikleri ve eğitim programları düzenlenmelidir. Son olarak, GSA ve diğer YZ tabanlı yöntemlerdeki teknolojik gelişmelerin yakından takip edilmesi ve bu yeniliklerin ilaç keşfi süreçlerine entegrasyonu sağlanmalıdır. Bu çalışma, GSA'ların ilaç keşfi alanında sunduğu potansiyeli ortaya koymakta ve gelecekteki araştırmalar için yeni perspektifler sunmaktadır.

Bu bağlamda elde edilen bulgular, ilaç keşfi ve salgın yönetimi alanında stratejik bir değere sahiptir. Tez, uluslararası düzeyde yayımlanmış makaleler, sunulan bildiri, kabul edilen kitap bölümü ve Bilimsel Araştırma Projesi (BAP) ile akademik katkısını göstermektedir. Bu tez çalışması, ilaç endüstrisi ve tıbbi araştırmalar için yeni stratejiler geliştirmeye ve acil tıbbi ihtiyaçlara cevap vermeye yönelik bir adım olarak değerlendirilebilir.

5.2 Öneriler

Bu tez çalışması, ilaç keşfi sürecinde yenilikçi yaklaşımların ve teknolojilerin önemini vurgulamaktadır. Bu bağlamda, aşağıdaki öneriler ilaç keşfi alanında daha fazla ilerleme sağlamak için değerlendirilebilir:

- GSA'ların Kullanımının Artırılması: GSA'lar, ilaç molekülleri arasındaki etkileşimleri modelleme ve ilaç keşfi sürecindeki potansiyel adayları belirleme konusunda önemli bir araçtır. Bu teknolojinin daha geniş bir şekilde kullanılması, ilaç keşfi sürecini hızlandırabilir ve maliyetleri azaltabilir.
- Veri Paylaşımı ve İşbirliğinin Teşvik Edilmesi: İlaç keşfi alanında veri paylaşımı ve iş birliği, potansiyel ilaç adaylarını belirleme ve değerlendirme sürecini iyileştirebilir. Akademik kurumlar, endüstriyel firmalar ve hükümet kurumları arasında daha fazla iş birliği ve veri paylaşımı teşvik edilmelidir.
- Yeniden Kullanılabilir İlaçların Daha Fazla İncelenmesi: Mevcut onaylı ilaçların yeniden kullanım stratejisi, yeni tedavi alanlarının keşfedilmesi ve mevcut hastalıkların tedavisinde kullanılması için önemli bir potansiyele sahiptir. Bu nedenle, daha fazla yeniden kullanılabilir ilaç adayının incelenmesi ve değerlendirilmesi gerekmektedir.
 - YZ ve ML'nin Daha Geniş Kapsamlı Kullanımı: YZ ve ML teknikleri, ilaç keşfi sürecindeki veri analizi ve modelleme işlemlerinde daha geniş kapsamlı bir şekilde kullanılmalıdır. Bu teknikler, büyük veri setlerinden anlamlı bilgiler çıkarmak ve ilaç adaylarını belirleme sürecini iyileştirmek için değerlidir.
 - Akademik ve Endüstriyel İşbirliğinin Güçlendirilmesi: Akademik kurumlar ile endüstriyel firmalar arasındaki iş birliği, ilaç keşfi sürecinde daha etkili ve verimli sonuçlar elde etmek için önemlidir. Bu iş birlikleri, temel bilimsel araştırmalardan klinik deneylere kadar çeşitli aşamalarda gerçekleştirilebilir.

Bu önerilerin dikkate alınması, ilaç keşfi alanında daha fazla ilerleme sağlayabilir ve gelecekteki pandemi ve salgınlar gibi acil tıbbi ihtiyaçlara daha iyi hazırlanmamıza yardımcı olabilir.

KAYNAKLAR

- Abdelnabi, R., Foo, C. S., Kaptein, S. J. F., Zhang, X., Do, T. N. D., Langendries, L., Vangeel, L., Breuer, J., Pang, J., Williams, R., Vergote, V., Heylen, E., Leyssen, P., Dallmeier, K., Coelmont, L., Chatterjee, A. K., Mols, R., Augustijns, P., De Jonghe, S., ... Neyts, J. (2021). The combined treatment of Molnupiravir and Favipiravir results in a potentiation of antiviral efficacy in a SARS-CoV-2 hamster infection model. *EBioMedicine*, 72, 103595. <https://doi.org/10.1016/j.ebiom.2021.103595>
- Abdelrheem, D. A., Ahmed, S. A., Abd El-Mageed, H. R., Mohamed, H. S., Rahman, A. A., Elsayed, K. N. M., & Ahmed, S. A. (2020). The inhibitory effect of some natural bioactive compounds against SARS-CoV-2 main protease: insights from molecular docking analysis and molecular dynamic simulation. *Journal of Environmental Science and Health, Part A*, 55(11), 1373–1386. <https://doi.org/10.1080/10934529.2020.1826192>
- Agrawal, U., Raju, R., & Udawadia, Z. F. (2020). Favipiravir: A new and emerging antiviral option in COVID-19. *Medical Journal Armed Forces India*, 76(4), 370–376. <https://doi.org/10.1016/j.mjafi.2020.08.004>
- Ahmad, W., Tayara, H., & Chong, K. T. (2023). Attention-Based Graph Neural Network for Molecular Solubility Prediction. *ACS Omega*, 8(3), 3236–3244. <https://doi.org/10.1021/acsomega.2c06702>
- Alexandridi, M., Mazej, J., Palermo, E., & Hiscott, J. (2022). The Coronavirus pandemic – 2022: Viruses, variants & vaccines. *Cytokine & Growth Factor Reviews*, 63, 1–9. <https://doi.org/10.1016/j.cytogfr.2022.02.002>
- Alhudhaif, A., Polat, K., & Karaman, O. (2021). Determination of COVID-19 pneumonia based on generalized convolutional neural network model from chest X-ray images. *Expert Systems with Applications*, 180, 115141. <https://doi.org/10.1016/j.eswa.2021.115141>
- Alves, L. A., Ferreira, N. C. da S., Maricato, V., Alberto, A. V. P., Dias, E. A., & Jose Aguiar Coelho, N. (2022a). Graph Neural Networks as a Potential Tool in Improving Virtual Screening Programs. *Frontiers in Chemistry*, 9. <https://doi.org/10.3389/fchem.2021.787194>
- Alves, L. A., Ferreira, N. C. da S., Maricato, V., Alberto, A. V. P., Dias, E. A., & Jose Aguiar Coelho, N. (2022b). Graph Neural Networks as a Potential Tool in

- Improving Virtual Screening Programs. *Frontiers in Chemistry*, 9. <https://doi.org/10.3389/fchem.2021.787194>
- Angelo, J. S., Guedes, I. A., Barbosa, H. J. C., & Dardenne, L. E. (2023). Multi-and many-objective optimization: present and future in de novo drug design. *Frontiers in Chemistry*, 11. <https://doi.org/10.3389/fchem.2023.1288626>
- Annett, S. (2021). Pharmaceutical drug development: high drug prices and the hidden role of public funding. *Biologia Futura*, 72(2), 129–138. <https://doi.org/10.1007/s42977-020-00025-5>
- Ashburn, T. T., & Thor, K. B. (2004). Drug repositioning: identifying and developing new uses for existing drugs. *Nature Reviews Drug Discovery*, 3(8), 673–683. <https://doi.org/10.1038/nrd1468>
- Baby, K., Maity, S., Mehta, C. H., Suresh, A., Nayak, U. Y., & Nayak, Y. (2021). Targeting SARS-CoV-2 Main Protease: A Computational Drug Repurposing Study. *Archives of Medical Research*, 52(1), 38–47. <https://doi.org/10.1016/j.arcmed.2020.09.013>
- Bai, L.-Y., Dai, H., Xu, Q., Junaid, M., Peng, S.-L., Zhu, X., Xiong, Y., & Wei, D.-Q. (2018). Prediction of Effective Drug Combinations by an Improved Naïve Bayesian Algorithm. *International Journal of Molecular Sciences*, 19(2), 467. <https://doi.org/10.3390/ijms19020467>
- Banjare, P., Wamanrao Matore, B., Murmu, A., Kumar, V., Singh, J., & Roy, P. P. (2023). In silico Strategy: A Promising Implement in the Development of Multitarget Drugs against Neurodegenerative Diseases. *Current Topics in Medicinal Chemistry*, 23(29), 2765–2791. <https://doi.org/10.2174/1568026623666230811113231>
- Baum, E. B. (1988). On the capabilities of multilayer perceptrons. *Journal of Complexity*, 4(3), 193–215. [https://doi.org/10.1016/0885-064X\(88\)90020-9](https://doi.org/10.1016/0885-064X(88)90020-9)
- Beigel, J. H., Tomashek, K. M., Dodd, L. E., Mehta, A. K., Zingman, B. S., Kalil, A. C., Hohmann, E., Chu, H. Y., Luetkemeyer, A., Kline, S., Lopez de Castilla, D., Finberg, R. W., Dierberg, K., Tapson, V., Hsieh, L., Patterson, T. F., Paredes, R., Sweeney, D. A., Short, W. R., ... Lane, H. C. (2020). Remdesivir for the Treatment of Covid-19 — Final Report. *New England Journal of Medicine*, 383(19), 1813–1826. <https://doi.org/10.1056/NEJMoa2007764>

- Bemis, G. W., & Murcko, M. A. (1999). Properties of Known Drugs. 2. Side Chains. *Journal of Medicinal Chemistry*, 42(25), 5095–5099. <https://doi.org/10.1021/jm9903996>
- Bentéjac, C., Csörgő, A., & Martínez-Muñoz, G. (2021). A comparative analysis of gradient boosting algorithms. *Artificial Intelligence Review*, 54(3), 1937–1967. <https://doi.org/10.1007/s10462-020-09896-5>
- Bibi, N., Farid, A., Gul, S., Ali, J., Amin, F., Kalathiya, U., & Hupp, T. (2022). Drug repositioning against COVID-19: a first line treatment. *Journal of Biomolecular Structure and Dynamics*, 40(23), 12812–12826. <https://doi.org/10.1080/07391102.2021.1977698>
- Bobrow, D. G. (1967). Problems in Natural Language Communication with Computers. *IEEE Transactions on Human Factors in Electronics*, HFE-8(1), 52–55. <https://doi.org/10.1109/THFE.1967.233305>
- Bobrowski, T., Chen, L., Eastman, R. T., Itkin, Z., Shinn, P., Chen, C. Z., Guo, H., Zheng, W., Michael, S., Simeonov, A., Hall, M. D., Zakharov, A. V., & Muratov, E. N. (2021). Synergistic and Antagonistic Drug Combinations against SARS-CoV-2. *Molecular Therapy*, 29(2), 873–885. <https://doi.org/10.1016/j.ymthe.2020.12.016>
- Bongini, P., Bianchini, M., & Scarselli, F. (2021). Molecular generative Graph Neural Networks for Drug Discovery. *Neurocomputing*, 450, 242–252. <https://doi.org/10.1016/j.neucom.2021.04.039>
- Bronstein, M. M., Bruna, J., LeCun, Y., Szlam, A., & Vandergheynst, P. (2017). Geometric Deep Learning: Going beyond Euclidean data. *IEEE Signal Processing Magazine*, 34(4), 18–42. <https://doi.org/10.1109/MSP.2017.2693418>
- Brown, J. B. (2018). Classifiers and their Metrics Quantified. *Molecular Informatics*, 37(1–2). <https://doi.org/10.1002/minf.201700127>
- Budak, C., Mençik, V., & Gider, V. (2023a). Determining similarities of COVID-19 – lung cancer drugs and affinity binding mode analysis by graph neural network-based GEFA method. *Journal of Biomolecular Structure and Dynamics*, 41(2), 659–671. <https://doi.org/10.1080/07391102.2021.2010601>
- Budak, C., Mençik, V., & Gider, V. (2023b). Determining similarities of COVID-19 – lung cancer drugs and affinity binding mode analysis by graph neural network-

- based GEFA method. *Journal of Biomolecular Structure and Dynamics*, 41(2), 659–671. <https://doi.org/10.1080/07391102.2021.2010601>
- Cheung, M., & Moura, J. M. F. (2020). Graph Neural Networks for COVID-19 Drug Discovery. *2020 IEEE International Conference on Big Data (Big Data)*, 5646–5648. <https://doi.org/10.1109/BigData50022.2020.9378164>
- Chong, C. R., & Sullivan, D. J. (2007). New uses for old drugs. *Nature*, 448(7154), 645–646. <https://doi.org/10.1038/448645a>
- Choudhury, C., Arul Murugan, N., & Priyakumar, U. D. (2022). Structure-based drug repurposing: Traditional and advanced AI/ML-aided methods. *Drug Discovery Today*, 27(7), 1847–1861. <https://doi.org/10.1016/j.drudis.2022.03.006>
- Dara, S., Dhamercherla, S., Jadav, S. S., Babu, C. M., & Ahsan, M. J. (2022). Machine Learning in Drug Discovery: A Review. *Artificial Intelligence Review*, 55(3), 1947–1999. <https://doi.org/10.1007/s10462-021-10058-4>
- Das, P., & Mazumder, D. H. (2024). K1K2NN: A novel multi-label classification approach based on neighbors for predicting COVID-19 drug side effects. *Computational Biology and Chemistry*, 110, 108066. <https://doi.org/10.1016/j.compbiolchem.2024.108066>
- Deb, S., & Reeves, A. A. (2021). Simulation of Remdesivir Pharmacokinetics and Its Drug Interactions. *Journal of Pharmacy & Pharmaceutical Sciences*, 24, 277–291. <https://doi.org/10.18433/jpps32011>
- DiMasi, J. A., Grabowski, H. G., & Hansen, R. W. (2016). Innovation in the pharmaceutical industry: New estimates of R&D costs. *Journal of Health Economics*, 47, 20–33. <https://doi.org/10.1016/j.jhealeco.2016.01.012>
- Donner, Y., Kazmierczak, S., & Fortney, K. (2018). Drug Repurposing Using Deep Embeddings of Gene Expression Profiles. *Molecular Pharmaceutics*, 15(10), 4314–4325. <https://doi.org/10.1021/acs.molpharmaceut.8b00284>
- Doshi, S., & Chepuri, S. P. (2022). A computational approach to drug repurposing using graph neural networks. *Computers in Biology and Medicine*, 150, 105992. <https://doi.org/10.1016/j.combiomed.2022.105992>
- DrugBank*. (2021). <https://go.drugbank.com/drugs>.
- DrugBank*. (2022, February). <https://go.drugbank.com/drugs>.
- DrugBank*. (2023, February). <https://go.drugbank.com/drugs>.

- Dunning, J., Kennedy, S. B., Antierens, A., Whitehead, J., Ciglenecki, I., Carson, G., Kanapathipillai, R., Castle, L., Howell-Jones, R., Pardinaz-Solis, R., Grove, J., Scott, J., Lang, T., Oliaro, P., & Horby, P. W. (2016). Experimental Treatment of Ebola Virus Disease with Brincidofovir. *PLOS ONE*, *11*(9), e0162199. <https://doi.org/10.1371/journal.pone.0162199>
- Dyall, J., Nelson, E. A., DeWald, L. E., Guha, R., Hart, B. J., Zhou, H., Postnikova, E., Logue, J., Vargas, W. M., Gross, R., Michelotti, J., Deuliis, N., Bennett, R. S., Crozier, I., Holbrook, M. R., Morris, P. J., Klumpp-Thomas, C., McKnight, C., Mierzwa, T., ... White, J. M. (2018). Identification of Combinations of Approved Drugs With Synergistic Activity Against Ebola Virus in Cell Cultures. *The Journal of Infectious Diseases*, *218*(suppl_5), S672–S678. <https://doi.org/10.1093/infdis/jiy304>
- Eroglu, E., & Toprak, C. (2021). OVERVIEW OF FAVIPIRAVIR AND REMDESIVIR TREATMENT FOR COVID-19. *INTERNATIONAL JOURNAL OF PHARMACEUTICAL SCIENCES AND RESEARCH*, *12*(4), 1950–1957.
- Eros, D., Keri, G., Kovesdi, I., Szantai-Kis, C., Meszaros, G., & Orfi, L. (2004). Comparison of Predictive Ability of Water Solubility QSPR Models Generated by MLR, PLS and ANN Methods. *Mini-Reviews in Medicinal Chemistry*, *4*(2), 167–177. <https://doi.org/10.2174/1389557043487466>
- FDA. (2023, July). <https://www.fda.gov/drugs/development-approval-process-drugs/drug-approvals-and-databases>.
- Fiolet, T., Guihur, A., Rebeaud, M. E., Mulot, M., Peiffer-Smadja, N., & Mahamat-Saleh, Y. (2021). Effect of hydroxychloroquine with or without azithromycin on the mortality of coronavirus disease 2019 (COVID-19) patients: a systematic review and meta-analysis. *Clinical Microbiology and Infection*, *27*(1), 19–27. <https://doi.org/10.1016/j.cmi.2020.08.022>
- Florescu, D. F., Kalil, A. C., Hewlett, A. L., Schuh, A. J., Stroher, U., Uyeki, T. M., & Smith, P. W. (2015). Administration of Brincidofovir and Convalescent Plasma in a Patient With Ebola Virus Disease. *Clinical Infectious Diseases*, *61*(6), 969–973. <https://doi.org/10.1093/cid/civ395>
- Fontanesi, M., Micheli, A., Milazzo, P., & Podda, M. (2023). Exploiting the structure of biochemical pathways to investigate dynamical properties with neural

- networks for graphs. *Bioinformatics*, 39(11).
<https://doi.org/10.1093/bioinformatics/btad678>
- Friedman, J. H. (2002). Stochastic gradient boosting. *Computational Statistics & Data Analysis*, 38(4), 367–378. [https://doi.org/10.1016/S0167-9473\(01\)00065-2](https://doi.org/10.1016/S0167-9473(01)00065-2)
- Gardner, M. W., & Dorling, S. R. (1998). Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences. *Atmospheric Environment*, 32(14–15), 2627–2636. [https://doi.org/10.1016/S1352-2310\(97\)00447-0](https://doi.org/10.1016/S1352-2310(97)00447-0)
- Gider, V., & Budak, C. (2022). Instruction of molecular structure similarity and scaffolds of drugs under investigation in ebola virus treatment by atom-pair and graph network: A combination of favipiravir and molnupiravir. *Computational Biology and Chemistry*, 101, 107778. <https://doi.org/10.1016/j.compbiolchem.2022.107778>
- Gignoux, E., Azman, A. S., de Smet, M., Azuma, P., Massaquoi, M., Job, D., Tiffany, A., Petrucci, R., Sterk, E., Potet, J., Suzuki, M., Kurth, A., Cannas, A., Bocquin, A., Strecker, T., Logue, C., Pottage, T., Yue, C., Cabrol, J.-C., ... Ciglenecki, I. (2016). Effect of Artesunate–Amodiaquine on Mortality Related to Ebola Virus Disease. *New England Journal of Medicine*, 374(1), 23–32. <https://doi.org/10.1056/NEJMoa1504605>
- Gligorijević, V., Renfrew, P. D., Kosciolk, T., Leman, J. K., Berenberg, D., Vatanen, T., Chandler, C., Taylor, B. C., Fisk, I. M., Vlamakis, H., Xavier, R. J., Knight, R., Cho, K., & Bonneau, R. (2021). Structure-based protein function prediction using graph convolutional networks. *Nature Communications*, 12(1), 3168. <https://doi.org/10.1038/s41467-021-23303-9>
- Gori, M., Monfardini, G., & Scarselli, F. (n.d.). A new model for learning in graph domains. *Proceedings. 2005 IEEE International Joint Conference on Neural Networks, 2005.*, 729–734. <https://doi.org/10.1109/IJCNN.2005.1555942>
- Hagar, M., Ahmed, H. A., Aljohani, G., & Alhaddad, O. A. (2020). Investigation of Some Antiviral N-Heterocycles as COVID 19 Drug: Molecular Docking and DFT Calculations. *International Journal of Molecular Sciences*, 21(11), 3922. <https://doi.org/10.3390/ijms21113922>
- Hamdi, M. E., Dezhkam, R., Hajizade, A., & Shamloo, A. (2022). Prediction of Aqueous Solubility of Drug Molecules by Embedding Spatial Conformers Using

- Graph Neural Networks. *2022 29th National and 7th International Iranian Conference on Biomedical Engineering (ICBME)*, 46–50. <https://doi.org/10.1109/ICBME57741.2022.10052964>
- Hand, D. J. (2009). Measuring classifier performance: a coherent alternative to the area under the ROC curve. *Machine Learning*, 77(1), 103–123. <https://doi.org/10.1007/s10994-009-5119-5>
- Hoertel, N., Sánchez-Rico, M., Vernet, R., Jannot, A.-S., Neuraz, A., Blanco, C., Lemogne, C., Airagnes, G., Paris, N., Daniel, C., Gramfort, A., Lemaitre, G., Bernaux, M., Bellamine, A., Beeker, N., & Limosin, F. (2021). Observational Study of Chlorpromazine in Hospitalized Patients with COVID-19. *Clinical Drug Investigation*, 41(3), 221–233. <https://doi.org/10.1007/s40261-021-01001-0>
- Hou, N., Shuai, L., Zhang, L., Xie, X., Tang, K., Zhu, Y., Yu, Y., Zhang, W., Tan, Q., Zhong, G., Wen, Z., Wang, C., He, X., Huo, H., Gao, H., Xu, Y., Xue, J., Peng, C., Zou, J., ... Hu, Q. (2023). Development of Highly Potent Noncovalent Inhibitors of SARS-CoV-2 3CLpro. *ACS Central Science*, 9(2), 217–227. <https://doi.org/10.1021/acscentsci.2c01359>
- Hu, P., Jiao, Z., Zhang, Z., & Wang, Q. (2021). Development of Solubility Prediction Models with Ensemble Learning. *Industrial & Engineering Chemistry Research*, 60(30), 11627–11635. <https://doi.org/10.1021/acs.iecr.1c02142>
- Huber, W., Carey, V. J., Long, L., Falcon, S., & Gentleman, R. (2007). Graphs in molecular biology. *BMC Bioinformatics*, 8(S6), S8. <https://doi.org/10.1186/1471-2105-8-S6-S8>
- Hussien, M. A., & Abdelaziz, A. E. M. (2020). Molecular docking suggests repurposing of brincidofovir as a potential drug targeting SARS-CoV-2 ACE2 receptor and main protease. *Network Modeling Analysis in Health Informatics and Bioinformatics*, 9(1), 56. <https://doi.org/10.1007/s13721-020-00263-6>
- Ihlenfeldt, W. (2018). PubChem. In *Applied Chemoinformatics* (pp. 245–258). Wiley. <https://doi.org/10.1002/9783527806539.ch6e>
- Indari, O., Jakhmola, S., Manivannan, E., & Jha, H. C. (2021). An Update on Antiviral Therapy Against SARS-CoV-2: How Far Have We Come? *Frontiers in Pharmacology*, 12. <https://doi.org/10.3389/fphar.2021.632677>

- Jin, W., Barzilay, R., & Jaakkola, T. (2018). Junction Tree Variational Autoencoder for Molecular Graph Generation. *Proceedings of the 35th International Conference on Machine Learning*, 80, 2323–2332.
- Jin, Z., Du, X., Xu, Y., Deng, Y., Liu, M., Zhao, Y., Zhang, B., Li, X., Zhang, L., Peng, C., Duan, Y., Yu, J., Wang, L., Yang, K., Liu, F., Jiang, R., Yang, X., You, T., Liu, X., ... Yang, H. (2020). Structure of Mpro from SARS-CoV-2 and discovery of its inhibitors. *Nature*, 582(7811), 289–293. <https://doi.org/10.1038/s41586-020-2223-y>
- Johansen, L. M., DeWald, L. E., Shoemaker, C. J., Hoffstrom, B. G., Lear-Rooney, C. M., Stossel, A., Nelson, E., Delos, S. E., Simmons, J. A., Grenier, J. M., Pierce, L. T., Pajouhesh, H., Lehár, J., Hensley, L. E., Glass, P. J., White, J. M., & Olinger, G. G. (2015). A screen of approved drugs and molecular probes identifies therapeutics with anti-Ebola virus activity. *Science Translational Medicine*, 7(290). <https://doi.org/10.1126/scitranslmed.aaa5597>
- KAGGLE. (2023, August). <https://www.kaggle.com/datasets/sorkun/aqsoldb-a-curved-aqueous-solubility-dataset?resource=download>.
- Kaore, S., & Atal, S. (2022). Molnupiravir in COVID-19: A Scoping Review. *Current Drug Research Reviews*, 14(3), 203–214. <https://doi.org/10.2174/2589977514666220527105158>
- Karaman, B., & Sippl, W. (2019). Computational Drug Repurposing: Current Trends. *Current Medicinal Chemistry*, 26(28), 5389–5409. <https://doi.org/10.2174/0929867325666180530100332>
- KARAMAN, O. (2021). Boosting Performance of Transfer Learning Model for Diagnosis of COVID-19 from Computer Tomography Scans. *Süleyman Demirel Üniversitesi Fen Edebiyat Fakültesi Fen Dergisi*, 16(1), 35–45. <https://doi.org/10.29233/sdufeffd.830351>
- Karasev, D. A., Sobolev, B. N., Lagunin, A. A., Filimonov, D. A., & Poroikov, V. V. (2022). The method predicting interaction between protein targets and small-molecular ligands with the wide applicability domain. *Computational Biology and Chemistry*, 98, 107674. <https://doi.org/10.1016/j.compbiolchem.2022.107674>

- Kharwade, R., & Mahajan, N. M. (2022). Role of nanocarriers for the effective delivery of anti-HIV drugs. In *Photophysics and Nanophysics in Therapeutics* (pp. 291–310). Elsevier. <https://doi.org/10.1016/B978-0-323-89839-3.00011-7>
- Kherouf, S., Bouarra, N., Bouakkadia, A., & Messadi, D. (2019). Modeling of linear and nonlinear quantitative structure property relationships of the aqueous solubility of phenol derivatives. *Journal of the Serbian Chemical Society*, 84(6), 575–590. <https://doi.org/10.2298/JSC180820016K>
- Kirtonia, A., Gala, K., Fernandes, S. G., Pandya, G., Pandey, A. K., Sethi, G., Khattar, E., & Garg, M. (2021). Repurposing of drugs: An attractive pharmacological strategy for cancer therapeutics. *Seminars in Cancer Biology*, 68, 258–278. <https://doi.org/10.1016/j.semcancer.2020.04.006>
- Klug, D. M., Gelb, M. H., & Pollastri, M. P. (2016). Repurposing strategies for tropical disease drug discovery. *Bioorganic & Medicinal Chemistry Letters*, 26(11), 2569–2576. <https://doi.org/10.1016/j.bmcl.2016.03.103>
- Lai, X., Yang, P., Wang, K., Yang, Q., & Yu, D. (2021). MGRNN: Structure Generation of Molecules Based on Graph Recurrent Neural Networks. *Molecular Informatics*, 40(10). <https://doi.org/10.1002/minf.202100091>
- Lavecchia, A. (2019). Deep learning in drug discovery: opportunities, challenges and future prospects. *Drug Discovery Today*, 24(10), 2017–2032. <https://doi.org/10.1016/j.drudis.2019.07.006>
- Lee, S., Park, H., Choi, C., Kim, W., Kim, K. K., Han, Y.-K., Kang, J., Kang, C.-J., & Son, Y. (2023). Multi-order graph attention network for water solubility prediction and interpretation. *Scientific Reports*, 13(1), 957. <https://doi.org/10.1038/s41598-022-25701-5>
- Li, J., Zheng, S., Chen, B., Butte, A. J., Swamidass, S. J., & Lu, Z. (2016). A survey of current trends in computational drug repositioning. *Briefings in Bioinformatics*, 17(1), 2–12. <https://doi.org/10.1093/bib/bbv020>
- Li, M., Chen, H., Zhang, H., Zeng, M., Chen, B., & Guan, L. (2022). Prediction of the Aqueous Solubility of Compounds Based on Light Gradient Boosting Machines with Molecular Fingerprints and the Cuckoo Search Algorithm. *ACS Omega*, 7(46), 42027–42035. <https://doi.org/10.1021/acsomega.2c03885>

- Li, P., & Leskovec, J. (2022). The Expressive Power of Graph Neural Networks. In *Graph Neural Networks: Foundations, Frontiers, and Applications* (pp. 63–98). Springer Nature Singapore. https://doi.org/10.1007/978-981-16-6054-2_5
- Li, X., & Peng, T. (2021). Strategy, Progress, and Challenges of Drug Repurposing for Efficient Antiviral Discovery. *Frontiers in Pharmacology*, 12. <https://doi.org/10.3389/fphar.2021.660710>
- Li, Y., Hu, J., Wang, Y., Zhou, J., Zhang, L., & Liu, Z. (2020). DeepScaffold: A Comprehensive Tool for Scaffold-Based De Novo Drug Discovery Using Deep Learning. *Journal of Chemical Information and Modeling*, 60(1), 77–91. <https://doi.org/10.1021/acs.jcim.9b00727>
- Light, D. W., & Warburton, R. (2011). Demythologizing the high costs of pharmaceutical research. *BioSocieties*, 6(1), 34–50. <https://doi.org/10.1057/biosoc.2010.40>
- Lipinski, C. A., Lombardo, F., Dominy, B. W., & Feeney, P. J. (1997). Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Advanced Drug Delivery Reviews*, 23(1–3), 3–25. [https://doi.org/10.1016/S0169-409X\(96\)00423-1](https://doi.org/10.1016/S0169-409X(96)00423-1)
- Liu, Y., Wang, Y., & Zhang, J. (2012). *New Machine Learning Algorithm: Random Forest* (pp. 246–252). https://doi.org/10.1007/978-3-642-34062-8_32
- Liu, Y., Wu, Y., Shen, X., & Xie, L. (2021). COVID-19 Multi-Targeted Drug Repurposing Using Few-Shot Learning. *Frontiers in Bioinformatics*, 1. <https://doi.org/10.3389/fbinf.2021.693177>
- Long, Y., Wu, M., Kwoh, C. K., Luo, J., & Li, X. (2020). Predicting human microbe–drug associations via graph convolutional network with conditional random field. *Bioinformatics*, 36(19), 4918–4927. <https://doi.org/10.1093/bioinformatics/btaa598>
- Luo, H., Li, M., Yang, M., Wu, F.-X., Li, Y., & Wang, J. (2021). Biomedical data and computational models for drug repositioning: a comprehensive review. *Briefings in Bioinformatics*, 22(2), 1604–1619. <https://doi.org/10.1093/bib/bbz176>
- Lupberger, J., Zeisel, M. B., Xiao, F., Thumann, C., Fofana, I., Zona, L., Davis, C., Mee, C. J., Turek, M., Gorke, S., Royer, C., Fischer, B., Zahid, M. N., Lavillette, D., Fresquet, J., Cosset, F.-L., Rothenberg, S. M., Pietschmann, T., Patel, A. H., ... Baumert, T. F. (2011). EGFR and EphA2 are host factors for hepatitis C virus

- entry and possible targets for antiviral therapy. *Nature Medicine*, 17(5), 589–595. <https://doi.org/10.1038/nm.2341>
- Ma, M., & Lei, X. (2023). A dual graph neural network for drug–drug interactions prediction based on molecular structure and interactions. *PLOS Computational Biology*, 19(1), e1010812. <https://doi.org/10.1371/journal.pcbi.1010812>
- Mak, K.-K., & Pichika, M. R. (2019). Artificial intelligence in drug development: present status and future prospects. *Drug Discovery Today*, 24(3), 773–780. <https://doi.org/10.1016/j.drudis.2018.11.014>
- Mao, C., Yao, L., & Luo, Y. (2022). MedGCN: Medication recommendation and lab test imputation via graph convolutional networks. *Journal of Biomedical Informatics*, 127, 104000. <https://doi.org/10.1016/j.jbi.2022.104000>
- Matlock, M. K., Zaretski, J. M., & Swamidass, S. J. (2013). Scaffold network generator: a tool for mining molecular structures. *Bioinformatics*, 29(20), 2655–2656. <https://doi.org/10.1093/bioinformatics/btt448>
- Maziarz, E. A. (1979). *Graph Theory: 1736-1936*. N. L. Biggs, E. K. Lloyd, R. J. Wilson. *Isis*, 70(1), 164–165. <https://doi.org/10.1086/352170>
- McCarthy, J., & Hayes, P. J. (1981). Some Philosophical Problems from the Standpoint of Artificial Intelligence. In *Readings in Artificial Intelligence* (pp. 431–450). Elsevier. <https://doi.org/10.1016/B978-0-934613-03-3.50033-7>
- Mentré, F., Taburet, A.-M., Guedj, J., Anglaret, X., Keïta, S., de Lamballerie, X., & Malvy, D. (2015). Dose regimen of favipiravir for Ebola virus disease. *The Lancet Infectious Diseases*, 15(2), 150–151. [https://doi.org/10.1016/S1473-3099\(14\)71047-3](https://doi.org/10.1016/S1473-3099(14)71047-3)
- Mercado, R., Rastemo, T., Lindelöf, E., Klambauer, G., Engkvist, O., Chen, H., & Jannik Bjerrum, E. (2021a). Graph networks for molecular design. *Machine Learning: Science and Technology*, 2(2), 025023. <https://doi.org/10.1088/2632-2153/abcf91>
- Mercado, R., Rastemo, T., Lindelöf, E., Klambauer, G., Engkvist, O., Chen, H., & Jannik Bjerrum, E. (2021b). Graph networks for molecular design. *Machine Learning: Science and Technology*, 2(2), 025023. <https://doi.org/10.1088/2632-2153/abcf91>

- Merkwirth, C., & Lengauer, T. (2005). Automatic Generation of Complementary Descriptors with Molecular Graph Networks. *Journal of Chemical Information and Modeling*, 45(5), 1159–1168. <https://doi.org/10.1021/ci049613b>
- Ng, Y. L., Salim, C. K., & Chu, J. J. H. (2021). Drug repurposing for COVID-19: Approaches, challenges and promising candidates. *Pharmacology & Therapeutics*, 228, 107930. <https://doi.org/10.1016/j.pharmthera.2021.107930>
- Nguyen, T., Le, H., Quinn, T. P., Nguyen, T., Le, T. D., & Venkatesh, S. (2021). GraphDTA: predicting drug–target binding affinity with graph neural networks. *Bioinformatics*, 37(8), 1140–1147. <https://doi.org/10.1093/bioinformatics/btaa921>
- Nguyen, T. M., Nguyen, T., Le, T. M., & Tran, T. (2022). GEFA: Early Fusion Approach in Drug-Target Affinity Prediction. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 19(2), 718–728. <https://doi.org/10.1109/TCBB.2021.3094217>
- Nguyen, T., Nguyen, G. T. T., Nguyen, T., & Le, D.-H. (2022). Graph Convolutional Networks for Drug Response Prediction. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 19(1), 146–154. <https://doi.org/10.1109/TCBB.2021.3060430>
- Nilar, S. H. (2023). Mining big data in drug discovery—triaging and decision trees. In *Big Data Analytics in Chemoinformatics and Bioinformatics* (pp. 265–281). Elsevier. <https://doi.org/10.1016/B978-0-323-85713-0.00019-0>
- Nosengo, N. (2016). Can you teach old drugs new tricks? *Nature*, 534(7607), 314–316. <https://doi.org/10.1038/534314a>
- Oliver, M. E., & Hinks, T. S. C. (2021). Azithromycin in viral infections. *Reviews in Medical Virology*, 31(2). <https://doi.org/10.1002/rmv.2163>
- Oprea, T. I., & Mestres, J. (2012). Drug Repurposing: Far Beyond New Targets for Old Drugs. *The AAPS Journal*, 14(4), 759–763. <https://doi.org/10.1208/s12248-012-9390-1>
- Park, K. (2019). A review of computational drug repurposing. *Translational and Clinical Pharmacology*, 27(2), 59. <https://doi.org/10.12793/tcp.2019.27.2.59>
- Parvathaneni, V., Kulkarni, N. S., Muth, A., & Gupta, V. (2019). Drug repurposing: a promising tool to accelerate the drug discovery process. *Drug Discovery Today*, 24(10), 2076–2085. <https://doi.org/10.1016/j.drudis.2019.06.014>

- Patel, T. K., Patel, P. B., Barvaliya, M., Saurabh, M. K., Bhalla, H. L., & Khosla, P. P. (2021). Efficacy and safety of lopinavir-ritonavir in COVID-19: A systematic review of randomized controlled trials. *Journal of Infection and Public Health*, 14(6), 740–748. <https://doi.org/10.1016/j.jiph.2021.03.015>
- Pathak, Y., Mishra, A., Choudhir, G., Kumar, A., & Tripathi, V. (2021). Rifampicin and Letermovir as potential repurposed drug candidate for COVID-19 treatment: insights from an in-silico study. *Pharmacological Reports*, 73(3), 926–938. <https://doi.org/10.1007/s43440-021-00228-0>
- Paul, S. M., Mytelka, D. S., Dunwiddie, C. T., Persinger, C. C., Munos, B. H., Lindborg, S. R., & Schacht, A. L. (2010). How to improve R&D productivity: the pharmaceutical industry's grand challenge. *Nature Reviews Drug Discovery*, 9(3), 203–214. <https://doi.org/10.1038/nrd3078>
- Peele, K. A., Potla Durthi, C., Srihansa, T., Krupanidhi, S., Ayyagari, V. S., Babu, D. J., Indira, M., Reddy, A. R., & Venkateswarulu, T. C. (2020). Molecular docking and dynamic simulations for antiviral compounds against SARS-CoV-2: A computational study. *Informatics in Medicine Unlocked*, 19, 100345. <https://doi.org/10.1016/j.imu.2020.100345>
- Peng, Y., Wang, M., Xu, Y., Wu, Z., Wang, J., Zhang, C., Liu, G., Li, W., Li, J., & Tang, Y. (2021). Drug repositioning by prediction of drug's anatomical therapeutic chemical code via network-based inference approaches. *Briefings in Bioinformatics*, 22(2), 2058–2072. <https://doi.org/10.1093/bib/bbaa027>
- Polat, Ç., Karaman, O., Karaman, C., Korkmaz, G., Balçı, M. C., & Kelek, S. E. (2021). COVID-19 diagnosis from chest X-ray images using transfer learning: Enhanced performance by debiasing dataloader. *Journal of X-Ray Science and Technology*, 29(1), 19–36. <https://doi.org/10.3233/XST-200757>
- Prateeksha, G., Rana, T. S., Asthana, A. K., Singh, B. N., & Barik, S. K. (2021). Screening of cryptogamic secondary metabolites as putative inhibitors of SARS-CoV-2 main protease and ribosomal binding domain of spike glycoprotein by molecular docking and molecular dynamics approaches. *Journal of Molecular Structure*, 1240, 130506. <https://doi.org/10.1016/j.molstruc.2021.130506>
- Probst, D., & Reymond, J.-L. (2020). Visualization of very large high-dimensional data sets as minimum spanning trees. *Journal of Cheminformatics*, 12(1), 12. <https://doi.org/10.1186/s13321-020-0416-x>

- Proschak, E., Stark, H., & Merk, D. (2019). Polypharmacology by Design: A Medicinal Chemist's Perspective on Multitargeting Compounds. *Journal of Medicinal Chemistry*, 62(2), 420–444. <https://doi.org/10.1021/acs.jmedchem.8b00760>
- PubChem. (2021). <https://pubchem.ncbi.nlm.nih.gov/>.
- PubChem. (2022, February). <https://pubchem.ncbi.nlm.nih.gov/>.
- PubChem. (2023, February). <https://pubchem.ncbi.nlm.nih.gov/>.
- Qian, N., & Sejnowski, T. J. (1988). Predicting the secondary structure of globular proteins using neural network models. *Journal of Molecular Biology*, 202(4), 865–884. [https://doi.org/10.1016/0022-2836\(88\)90564-5](https://doi.org/10.1016/0022-2836(88)90564-5)
- Qin, Y., Li, C., Shi, X., & Wang, W. (2022a). MLP-Based Regression Prediction Model For Compound Bioactivity. *Frontiers in Bioengineering and Biotechnology*, 10. <https://doi.org/10.3389/fbioe.2022.946329>
- Qin, Y., Li, C., Shi, X., & Wang, W. (2022b). MLP-Based Regression Prediction Model For Compound Bioactivity. *Frontiers in Bioengineering and Biotechnology*, 10. <https://doi.org/10.3389/fbioe.2022.946329>
- Reiser, P., Neubert, M., Eberhard, A., Torresi, L., Zhou, C., Shao, C., Metni, H., van Hoesel, C., Schopmans, H., Sommer, T., & Friederich, P. (2022). Graph neural networks for materials science and chemistry. *Communications Materials*, 3(1), 93. <https://doi.org/10.1038/s43246-022-00315-6>
- Richard Socher, Cliff Chiung-Yu Lin, Andrew Y. Ng, & Christopher D. Manning. (2011). *Parsing natural scenes and natural language with recursive neural networks*. 129–136.
- Rifaioğlu, A. S., Atas, H., Martin, M. J., Cetin-Atalay, R., Atalay, V., & Doğan, T. (2019). Recent applications of deep learning and machine intelligence on in silico drug discovery: methods, tools and databases. *Briefings in Bioinformatics*, 20(5), 1878–1912. <https://doi.org/10.1093/bib/bby061>
- Rifaioğlu, A. S., Nalbat, E., Atalay, V., Martin, M. J., Cetin-Atalay, R., & Doğan, T. (2020). DEEPScreen: high performance drug–target interaction prediction with convolutional neural networks using 2-D structural compound representations. *Chemical Science*, 11(9), 2531–2557. <https://doi.org/10.1039/C9SC03414E>
- Rogers, D. J., & Tanimoto, T. T. (1960). A Computer Program for Classifying Plants. *Science*, 132(3434), 1115–1118. <https://doi.org/10.1126/science.132.3434.1115>

- Romani, D., & Brandán, S. A. (2019). Effect of the side chain on the properties from cidofovir to brincidofovir, an experimental antiviral drug against to Ebola virus disease. *Arabian Journal of Chemistry*, 12(8), 2959–2972. <https://doi.org/10.1016/j.arabjc.2015.06.030>
- Rosace, A., Bennett, A., Oeller, M., Mortensen, M. M., Sakhnini, L., Lorenzen, N., Poulsen, C., & Sormanni, P. (2023). Automated optimisation of solubility and conformational stability of antibodies and proteins. *Nature Communications*, 14(1), 1937. <https://doi.org/10.1038/s41467-023-37668-6>
- Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1988). (1986) D. E. Rumelhart, G. E. Hinton, and R. J. Williams, “Learning internal representations by error propagation,” *Parallel Distributed Processing: Explorations in the Microstructures of Cognition*, Vol. I, D. E. Rumelhart and J. L. McClelland (Eds.) Cambridge, MA: MIT Press, pp. 318-362. In *Neurocomputing, Volume 1* (pp. 675–695). The MIT Press. <https://doi.org/10.7551/mitpress/4943.003.0128>
- Sacha, M., Błaż, M., Byrski, P., Dąbrowski-Tumański, P., Chromiński, M., Loska, R., Włodarczyk-Pruszyński, P., & Jastrzębski, S. (2021). Molecule Edit Graph Attention Network: Modeling Chemical Reactions as Sequences of Graph Edits. *Journal of Chemical Information and Modeling*, 61(7), 3273–3284. <https://doi.org/10.1021/acs.jcim.1c00537>
- Sadybekov, A. V., & Katritch, V. (2023). Computational approaches streamlining drug discovery. *Nature*, 616(7958), 673–685. <https://doi.org/10.1038/s41586-023-05905-z>
- Sakai, M., Nagayasu, K., Shibui, N., Andoh, C., Takayama, K., Shirakawa, H., & Kaneko, S. (2021). Prediction of pharmacological activities from chemical structures with graph convolutional neural networks. *Scientific Reports*, 11(1), 525. <https://doi.org/10.1038/s41598-020-80113-7>
- Sang, P., Tian, S.-H., Meng, Z.-H., & Yang, L.-Q. (2020). Anti-HIV drug repurposing against SARS-CoV-2. *RSC Advances*, 10(27), 15775–15783. <https://doi.org/10.1039/D0RA01899F>
- Scarselli, F., Gori, M., Ah Chung Tsoi, Hagenbuchner, M., & Monfardini, G. (2009). The Graph Neural Network Model. *IEEE Transactions on Neural Networks*, 20(1), 61–80. <https://doi.org/10.1109/TNN.2008.2005605>

- Selvaraj, C., Chandra, I., & Singh, S. K. (2022). Artificial intelligence and machine learning approaches for drug design: challenges and opportunities for the pharmaceutical industries. *Molecular Diversity*, 26(3), 1893–1913. <https://doi.org/10.1007/s11030-021-10326-z>
- Shah, M., Patel, M., Shah, M., Patel, M., & Prajapati, M. (2024). Computational transformation in drug discovery: A comprehensive study on molecular docking and quantitative structure activity relationship (QSAR). *Intelligent Pharmacy*. <https://doi.org/10.1016/j.ipha.2024.03.001>
- Shamsi, A., Mohammad, T., Anwar, S., AlAjmi, M. F., Hussain, A., Rehman, Md. T., Islam, A., & Hassan, Md. I. (2020). Glecaprevir and Maraviroc are high-affinity inhibitors of SARS-CoV-2 main protease: possible implication in COVID-19 therapy. *Bioscience Reports*, 40(6). <https://doi.org/10.1042/BSR20201256>
- Shi, H., Liu, S., Chen, J., Li, X., Ma, Q., & Yu, B. (2019). Predicting drug-target interactions using Lasso with random forest based on evolutionary information and chemical structure. *Genomics*, 111(6), 1839–1852. <https://doi.org/10.1016/j.ygeno.2018.12.007>
- Shi, M., Huang, Y., Zhu, X., Tang, Y., Zhuang, Y., & Liu, J. (2021). GAEN: Graph Attention Evolving Networks. *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence*, 1541–1547. <https://doi.org/10.24963/ijcai.2021/213>
- Shiraki, K., Sato, N., Sakai, K., Matsumoto, S., Kaszynski, R. H., & Takemoto, M. (2022). Antiviral therapy for COVID-19: Derivation of optimal strategy based on past antiviral and favipiravir experiences. *Pharmacology & Therapeutics*, 235, 108121. <https://doi.org/10.1016/j.pharmthera.2022.108121>
- Shtar, G., Rokach, L., & Shapira, B. (2019). Detecting drug-drug interactions using artificial neural networks and classic graph similarity measures. *PLOS ONE*, 14(8), e0219796. <https://doi.org/10.1371/journal.pone.0219796>
- Sinha, P., Roshini, D., Daoo, V., Abraham, B. M., & Singh, J. K. (2023). Integrating Machine Learning and Molecular Simulation for Material Design and Discovery. *Transactions of the Indian National Academy of Engineering*, 8(3), 325–340. <https://doi.org/10.1007/s41403-023-00412-z>
- Song, D., Chen, Y., Min, Q., Sun, Q., Ye, K., Zhou, C., Yuan, S., Sun, Z., & Liao, J. (2019). Similarity-based machine learning support vector machine predictor of

- drug-drug interactions with improved accuracies. *Journal of Clinical Pharmacy and Therapeutics*, 44(2), 268–275. <https://doi.org/10.1111/jcpt.12786>
- Suay-García, B., Bueso-Bordils, J. I., Falcó, A., Antón-Fos, G. M., & Alemán-López, P. A. (2022). Virtual Combinatorial Chemistry and Pharmacological Screening: A Short Guide to Drug Design. *International Journal of Molecular Sciences*, 23(3), 1620. <https://doi.org/10.3390/ijms23031620>
- Sun, D., Gao, W., Hu, H., & Zhou, S. (2022). Why 90% of clinical drug development fails and how to improve it? *Acta Pharmaceutica Sinica B*, 12(7), 3049–3062. <https://doi.org/10.1016/j.apsb.2022.02.002>
- Sun, M., Zhao, S., Gilvary, C., Elemento, O., Zhou, J., & Wang, F. (2020). Graph convolutional networks for computational drug development and discovery. *Briefings in Bioinformatics*, 21(3), 919–935. <https://doi.org/10.1093/bib/bbz042>
- Sweiti, H., Ekwunife, O., Jaschinski, T., & Lhachimi, S. K. (2017). Repurposed Therapeutic Agents Targeting the Ebola Virus: A Systematic Review. *Current Therapeutic Research*, 84, 10–21. <https://doi.org/10.1016/j.curtheres.2017.01.007>
- The WHO MERS-CoV Research Group, -. (2013). State of Knowledge and Data Gaps of Middle East Respiratory Syndrome Coronavirus (MERS-CoV) in Humans. *PLoS Currents*. <https://doi.org/10.1371/currents.outbreaks.0bf719e352e7478f8ad85fa30127ddb8>
- Tianfan Fu, Wenhao Gao, Cao Xiao, Jacob Yasonik, Connor W. Coley, & Jimeng Sun. (2022). Differentiable Scaffolding Tree for Molecular Optimization. *The Tenth International Conference on Learning Representations (ICLR 2022)*.
- Tong, S., Su, Y., Yu, Y., Wu, C., Chen, J., Wang, S., & Jiang, J. (2020). Ribavirin therapy for severe COVID-19: a retrospective cohort study. *International Journal of Antimicrobial Agents*, 56(3), 106114. <https://doi.org/10.1016/j.ijantimicag.2020.106114>
- Ullah, I., Sharma, R., Mete, A., Biagini, G. A., Wetzel, D. M., & Horrocks, P. D. (2020). The relative rate of kill of the MMV Malaria Box compounds provides links to the mode of antimalarial action and highlights scaffolds of medicinal chemistry interest. *Journal of Antimicrobial Chemotherapy*, 75(2), 362–370. <https://doi.org/10.1093/jac/dkz443>

- Varin, T., Schuffenhauer, A., Ertl, P., & Renner, S. (2011). Mining for Bioactive Scaffolds with Scaffold Networks: Improved Compound Set Enrichment from Primary Screening Data. *Journal of Chemical Information and Modeling*, 51(7), 1528–1538. <https://doi.org/10.1021/ci2000924>
- Vatansever, E. C., Yang, K. S., Drelich, A. K., Kratch, K. C., Cho, C.-C., Kempaiah, K. R., Hsu, J. C., Mellott, D. M., Xu, S., Tseng, C.-T. K., & Liu, W. R. (2021). Bepridil is potent against SARS-CoV-2 in vitro. *Proceedings of the National Academy of Sciences*, 118(10). <https://doi.org/10.1073/pnas.2012201118>
- Veličković, P. (2023). Everything is connected: Graph neural networks. *Current Opinion in Structural Biology*, 79, 102538. <https://doi.org/10.1016/j.sbi.2023.102538>
- Vilar, S., Harpaz, R., Uriarte, E., Santana, L., Rabadan, R., & Friedman, C. (2012). Drug—drug interaction through molecular structure similarity analysis. *Journal of the American Medical Informatics Association*, 19(6), 1066–1074. <https://doi.org/10.1136/amiajnl-2012-000935>
- Wang, J., Liu, X., Shen, S., Deng, L., & Liu, H. (2022). DeepDDS: deep graph neural network with attention mechanism to predict synergistic drug combinations. *Briefings in Bioinformatics*, 23(1). <https://doi.org/10.1093/bib/bbab390>
- Wang, J., Ma, A., Chang, Y., Gong, J., Jiang, Y., Qi, R., Wang, C., Fu, H., Ma, Q., & Xu, D. (2021). scGNN is a novel graph neural network framework for single-cell RNA-Seq analyses. *Nature Communications*, 12(1), 1882. <https://doi.org/10.1038/s41467-021-22197-x>
- Wang, L., You, Z.-H., Li, Y.-M., Zheng, K., & Huang, Y.-A. (2020). GCNCDA: A new method for predicting circRNA-disease associations based on Graph Convolutional Network Algorithm. *PLOS Computational Biology*, 16(5), e1007568. <https://doi.org/10.1371/journal.pcbi.1007568>
- Wang, T., Shao, W., Huang, Z., Tang, H., Zhang, J., Ding, Z., & Huang, K. (2021). MOGONET integrates multi-omics data using graph convolutional networks allowing patient classification and biomarker identification. *Nature Communications*, 12(1), 3445. <https://doi.org/10.1038/s41467-021-23774-w>
- Wein, S., Schüller, A., Tomé, A. M., Malloni, W. M., Greenlee, M. W., & Lang, E. W. (2022). Forecasting brain activity based on models of spatiotemporal brain

- dynamics: A comparison of graph neural network architectures. *Network Neuroscience*, 6(3), 665–701. https://doi.org/10.1162/netn_a_00252
- Weizenbaum, J. (1966). ELIZA—a computer program for the study of natural language communication between man and machine. *Communications of the ACM*, 9(1), 36–45. <https://doi.org/10.1145/365153.365168>
- Wengong Jin, Kevin Yang, Regina Barzilay, & Tommi Jaakkola. (2019). Learning Multimodal Graph-to-Graph Translation for Molecular Optimization. *The Seventh International Conference on Learning Representations (ICLR 2019)*.
- WHO. (2022). <https://www.who.int/news/item/17-07-2019-ebola-outbreak-in-the-democratic-republic-of-the-congo-declared-a-public-health-emergency-of-international-concern>.
- Wilson, A. J., Troy-Barnes, E., Subhan, M., Clark, F., Gupta, R., Fielding, A. K., Kottaridis, P., Mansour, M. R., O’Nions, J., Payne, E., Chavda, N., Baker, R., Thomson, K., & Khwaja, A. (2020). Successful remission induction therapy with gilteritinib in a patient with *de novo FLT3* -mutated acute myeloid leukaemia and severe COVID-19. *British Journal of Haematology*, 190(4). <https://doi.org/10.1111/bjh.16962>
- Wishart, D. S. (2006). DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Research*, 34(90001), D668–D672. <https://doi.org/10.1093/nar/gkj067>
- Wishart, D. S., Feunang, Y. D., Guo, A. C., Lo, E. J., Marcu, A., Grant, J. R., Sajed, T., Johnson, D., Li, C., Sayeeda, Z., Assempour, N., Iynkkaran, I., Liu, Y., Maciejewski, A., Gale, N., Wilson, A., Chin, L., Cummings, R., Le, D., ... Wilson, M. (2018). DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Research*, 46(D1), D1074–D1082. <https://doi.org/10.1093/nar/gkx1037>
- Wouters, O. J., McKee, M., & Luyten, J. (2020). Estimated Research and Development Investment Needed to Bring a New Medicine to Market, 2009-2018. *JAMA*, 323(9), 844. <https://doi.org/10.1001/jama.2020.1166>
- Wu, Y., Gao, M., Zeng, M., Zhang, J., & Li, M. (2022). BridgeDPI: a novel Graph Neural Network for predicting drug–protein interactions. *Bioinformatics*, 38(9), 2571–2578. <https://doi.org/10.1093/bioinformatics/btac155>

- Wu, Y., Ho, W., Huang, Y., Jin, D.-Y., Li, S., Liu, S.-L., Liu, X., Qiu, J., Sang, Y., Wang, Q., Yuen, K.-Y., & Zheng, Z.-M. (2020). SARS-CoV-2 is an appropriate name for the new coronavirus. *The Lancet*, 395(10228), 949–950. [https://doi.org/10.1016/S0140-6736\(20\)30557-2](https://doi.org/10.1016/S0140-6736(20)30557-2)
- Xu, S., Lin, X., & Yu, H. (2021). Prediction of Drug-Target Interactions Using Molecular Graph and GDNet-DTI Model. *2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 3747–3754. <https://doi.org/10.1109/BIBM52615.2021.9669652>
- Xue, H., Li, J., Xie, H., & Wang, Y. (2018). Review of Drug Repositioning Approaches and Resources. *International Journal of Biological Sciences*, 14(10), 1232–1244. <https://doi.org/10.7150/ijbs.24612>
- Yang, X., Wang, Y., Byrne, R., Schneider, G., & Yang, S. (2019). Concepts of Artificial Intelligence for Computer-Assisted Drug Discovery. *Chemical Reviews*, 119(18), 10520–10594. <https://doi.org/10.1021/acs.chemrev.8b00728>
- Yao, H.-Y., Zhang, C.-Y., Yao, Z.-L., Chen, C. L. P., & Hu, J. (2024). A recurrent graph neural network for inductive representation learning on dynamic graphs. *Pattern Recognition*, 154, 110577. <https://doi.org/10.1016/j.patcog.2024.110577>
- Zhan, P., Yu, B., & Ouyang, L. (2022). Drug repurposing: An effective strategy to accelerate contemporary drug discovery. *Drug Discovery Today*, 27(7), 1785–1788. <https://doi.org/10.1016/j.drudis.2022.05.026>
- Zhang, S., Tong, H., Xu, J., & Maciejewski, R. (2019). Graph convolutional networks: a comprehensive review. *Computational Social Networks*, 6(1), 11. <https://doi.org/10.1186/s40649-019-0069-y>
- Zhao, T., Hu, Y., Valsdottir, L. R., Zang, T., & Peng, J. (2021). Identifying drug–target interactions based on graph convolutional network and deep neural network. *Briefings in Bioinformatics*, 22(2), 2141–2150. <https://doi.org/10.1093/bib/bbaa044>
- Zhao, X., Chen, L., Guo, Z.-H., & Liu, T. (2019). Predicting Drug Side Effects with Compact Integration of Heterogeneous Networks. *Current Bioinformatics*, 14(8), 709–720. <https://doi.org/10.2174/1574893614666190220114644>
- Zhou, D., Peng, S., Wei, D.-Q., Zhong, W., Dou, Y., & Xie, X. (2021). LUNAR :Drug Screening for Novel Coronavirus Based on Representation Learning Graph

Convolutional Network. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 18(4), 1290–1298. <https://doi.org/10.1109/TCBB.2021.3085972>

Zweig, M. H., & Campbell, G. (1993). Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. *Clinical Chemistry*, 39(4), 561–577. <https://doi.org/10.1093/clinchem/39.4.561>



ÖZGEÇMİŞ

Kişisel Bilgiler

Soyad, Ad

GİDER, Veysel

Web sayfası

(Research Gate, Academia, vs.)

www.starnnt.com

Eğitim Bilgileri

Derece	Kurum	Mezuniyet Yılı
Yüksek Lisans	Batman Üniversitesi	2020
Lisans	Batman Üniversitesi	2017
Ön Lisans	Muğla Üniversitesi	2011
Lise	Batman Gazi Lisesi	2007

İş Deneyimi

Dönem (Yıl)	Şirket, Kurum	Görev
2022-....	Batman Üniversitesi	Öğretim Görevlisi

Yabancı Dil

İngilizce

Yayınlar

1. Gider, V., & Budak, C. (2022). Instruction of molecular structure similarity and scaffolds of drugs under investigation in ebola virus treatment by atom-pair and graph network: A combination of favipiravir and molnupiravir. *Computational biology and chemistry*, 101, 107778. <https://doi.org/10.1016/j.compbiolchem.2022.107778>
2. Budak, C., Mençik, V., & Gider, V. (2023). Determining similarities of COVID-19–lung cancer drugs and affinity binding mode analysis by graph neural network-based GEFA method. *Journal of Biomolecular Structure and Dynamics*, 41(2), 659-671. <https://doi.org/10.1080/07391102.2021.2010601>
3. Gider, V., & Budak, C. (2024). Drug Solubility Prediction: A Comparative Analysis of GNN, MLP, and Traditional Machine Learning Algorithms. *Gazi University Journal of Science Part C: Design and Technology*, 1-1. <https://doi.org/10.29109/gujsc.1371519>
4. Budak, C. & Gider, V. (2023). *Artificial Intelligence and Drug Discovery: Similarity and Molecular Structure Analysis of Seven Different Drugs for Ebola Treatment*. B. Turgut (Ed.), *Reflections of Change in Engineering, Science, Maths and Education* (s. 103- 136). Klaipeda: SRA Academic Publishing.

<https://www.researchgate.net/publication/370124494> Reflections of Change in Engineering Science chapter 6

5. Gider, V., & Budak, C. (2023, Ekim 19-20). *Grafik sinir ağırları ile ilaç molekülü çözünürlük tahmini*. Dicle Üniversitesi 1. Uluslararası Fen Bilimleri Lisansüstü Araştırmalar Sempozyumu (252-257), Diyarbakır, Türkiye.
https://www.dicle.edu.tr/Contents/Announcements/Files/c2971b88-59b0-4baf-83b9-03cc97b54e6c/e530725ac17d42b2abcdd524e1ca296f_DUFEBLAS%202023%20-%20Bildiri%20Kitab%c4%b1_Proceeding%20Book.pdf

Özel İlgiler

Yüzme, Müzik

DİCLE ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
TEZ BENZERLİK BİLDİRİMİ FORMU

Öğrencinin Adı, Soyadı	Veysel GİDER
Öğrenci No	20805502
Ana Bilim Dalı	Elektrik-Elektronik Mühendisliği
Program Türü	Proje ☐ Yüksek Lisans ☐ Doktora ☒
Tez Danışmanı (Ünvanı, Adı, Soyadı)	Doç. Dr. Cafer BUDAK
(Varsa) II. Tez Danışmanı (Ünvanı, Adı, Soyadı)	
Tez Başlığı	Grafik Sinir Ağları ile İlaç Keşfi
RAPOR BİLGİLERİ	
Raporlama Aşaması	Tez Savunma Sınavı Sonrası
Sayfa Sayısı	187
Raporlama Tarihi	12.06.2024
Benzerlik Oranı (%)	6

Yukarıda bilgileri verilen tez çalışmamın toplam 187 sayfalık kısmına ilişkin, 12/06/2024 tarihinde şahsım/tez danışmanım tarafından TURNITIN isimli intihal tespit programından aşağıda belirtilen filtrelemeler uygulanarak alınmış olan intihal raporuna göre, tezimin benzerlik oranı % 6 olarak tespit edilmiştir.

Uygulanan filtrelemeler:

- ☐Başlangıç Bölümleri (Kabul ve Onay sayfası, Teşekkür sayfası, Özet/Abstract) hariç
☒ Kaynaklar hariç
☐ Alıntılar hariç/dâhil
☐ Diğer (Açıklayınız)

Tezimin benzerlik oranı, Dicle Üniversitesi Fen Bilimleri Enstitüsü İntihal Raporu Uygulama Esaslarında belirtilen üst sınır benzerlik oranını aşmamaktadır. Tez benzerlik oranı üst sınır benzerlik oranının altında olsa dahi aksinin tespit edilmesi durumunda her türlü yasal sorumluluğu kabul ettiğimi ve hukuki sonuçlarına razı olduğumu bildirir, gereğini arz ederim.

Öğrencinin Adı, Soyadı: Veysel GİDER

Tarih: 12.06.2024

İmza:

Danışman Adı, Soyadı: Doç. Dr. Cafer

İmza:

BUDAK

Tarih:12.06.2024

Ana Bilim Dalı Başkanı Adı, Soyadı: Doç.

İmza:

Dr. Cafer BUDAK

Tarih:12.06.2024