

**T.C.
SÜLEYMAN DEMİREL ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

**AÇIKLANABİLİR YAPAY ZEKA MODELLERİ İLE İLAÇ YENİDEN
KONUMLANDIRMASI İÇİN BENZETİM ORTAMLARI
GELİŞTİRİLMESİ**

İlhan UYSAL

**Danışman
Doç. Dr. Utku KÖSE**

**DOKTORA TEZİ
BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI
ISPARTA - 2023**



© 2023 [İlhan UYSAL]

İÇİNDEKİLER

	Sayfa
İÇİNDEKİLER.....	i
ÖZET	ii
ABSTRACT	iv
TEŞEKKÜR.....	vi
ŞEKİLLER DİZİNİ	vii
ÇİZELGELER DİZİNİ	viii
SİMGELER VE KISALTMALAR DİZİNİ	ix
1. GİRİŞ	1
2. KAYNAK ÖZETLERİ	9
3. MATERYAL VE YÖNTEM	16
3.1. Çocukluk Çağı Akut Lösemisi	16
3.2. İlaç Etkileşimleri.....	17
3.3. Histon Deasetilaz (HDAC).....	17
3.4. Lipinski Kuralları (Lipinski'nin 5 Kuralı)	19
3.5. Açıklanabilir Yapay Zekâ (Explainable Artificial Intelligence – XAI)	21
3.6. ChEMBL (Chemical Biology Information Resource) Veritabanı.....	24
3.7. Python Kütüphaneleri.....	25
3.8. Yapay Zekâ Modelleri	27
4. ARAŞTIRMA BULGULARI.....	30
5. TARTIŞMA VE SONUÇLAR.....	56
KAYNAKLAR	82
EKLER.....	94
EK A. Diğer Performans Metrikleri – 1.....	95
EK A. Diğer Performans Metrikleri – 2.....	97
ÖZGEÇMİŞ.....	99

ÖZET

Doktora Tezi

AÇIKLANABİLİR YAPAY ZEKA MODELLERİ İLE İLAÇ YENİDEN KONUMLANDIRMASI İÇİN BENZETİM ORTAMLARI GELİŞTİRİLMESİ

İlhan UYSAL

Süleyman Demirel Üniversitesi
Fen Bilimleri Enstitüsü
Bilgisayar Mühendisliği Anabilim Dalı

Danışman: Doç. Dr. Utku KÖSE

Farmakoloji alanı, yapay zekânın yaygın olarak benimsenmesiyle birlikte önemli gelişmelere sahne olmuştur. Özellikle, ilaç keşfinde yapay zekâ kullanımı, geleneksel ilaç geliştirme süreçleriyle ilişkili maliyet ve zaman gibi büyük zorlukların üstesinden gelme konusunda umut vaat etmektedir. Yeni bir ilaç için geleneksel onay süreci sadece maliyetli değil, aynı zamanda zaman alıcıdır ve genellikle 12 ile 14 yıllık bir süreyi kapsamaktadır. Bu uzun keşif aşaması, ilaç geliştirmeyi hızlandırmak için alternatif yaklaşımların araştırılmasına yol açmıştır. Bu yaklaşımlardan biri, halihazırda onaylanmış ilaçların yeni endikasyonlar için kullanılmasını içeren ilaç yeniden konumlandırma. Çok sayıda ilaç şirketi, onaylanmış ve daha önce başarısız olmuş ilaç moleküllerini yeniden geliştirmek ve böylece çeşitli tıbbi durumlarda yeni tedavi seçenekleri sunmak için ilaç yeniden konumlandırmayı benimsemektedir.

Bu tez çalışmasında çocukluk çağı akut lösemi tedavisinde kullanılan Mercaptopurine, Tioguanine, Prednisolone, Dexamethasone, Vincristine, Methotrexate, Daunorubicin, Cytarabine ilaçlarının diğer ilaçlarla olan etkileşimlerinin karmaşıklığına yönelik olarak ağ grafikleri oluşturulmuş ve bu ilaçlarla etkileşime giren ilaçların sahip olduğu moleküller araştırılmıştır. Histon deasetilaz (HDAC) genlerinin çocukluk çağı akut lösemisi ile ilişkisi üzerinden molekül hesaplamaları yapılmış ve yeniden kullanım sürecinde hangi moleküller parmak izlerinin ne oranda etki ettiği açıklanabilir yapay zekâ yöntemleri ile belirlenmiştir. HDAC genlerinde en fazla frekansa sahip olan IC50 ve inhibisyon standart türleri için ayrı ayrı değerlendirmeler yapılmıştır. Regresyon ve sınıflandırma algoritmaları ile model performansları karşılaştırılmış ve en başarılı algoritma Decision Tree olmuştur. Ayrıca elde edilen tüm bulgular için bir benzetim ortamı tasarlanmıştır.

Çalışma, ilaçların yeniden konumlandırılmasında bir ortam sunduğu için literatür açısından önemli çıkarımlara sahiptir ve potansiyel olarak daha etkili terapötik stratejilerin belirlenmesine rehberlik etmektedir. Geliştirilen sistematik çerçeve, gelecekteki ilaç yeniden konumlandırma çalışmalarının tekrarlanabilirliğini ve ölçeklenebilirliğini artırmakta, sonuçta yeni terapötik seçenekler keşfetme potansiyeline sahip daha verimli ve hedefli ilaç yeniden

konumlandırma yaklaşımlarının ilerlemesine katkıda bulunmaktadır. Gelecek çalışmalarda veri setinin genişletilmesi ve iyileştirilmesi, topluluk yöntemleri ile algoritmaların birleştirilmesi yapılabilir. Ayrıca bulguların güvenilirliği ve doğruluğu açısından klinik deneyler de önem arz etmektedir.

Anahtar Kelimeler: Histon deasetilaz, çocukluk çağı akut lösemisi, açıklanabilir yapay zekâ, regresyon, sınıflandırma, ilaç-ilaç etkileşimi, ağ grafiği, benzetim.

2023, 100 sayfa



ABSTRACT

Ph.D. Thesis

DEVELOPING SIMULATION ENVIRONMENTS FOR DRUG REPURPOSING WITH EXPLAINABLE ARTIFICIAL INTELLIGENCE MODELS

Ilhan UYSAL

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

Supervisor: Assoc. Prof. Dr. Utku KOSE

The field of pharmacology has witnessed significant developments with the widespread adoption of artificial intelligence. In particular, the use of AI in drug discovery shows promise in overcoming the major challenges associated with traditional drug development processes, such as cost and time. The traditional approval process for a new drug is not only costly but also time-consuming, typically spanning 12 to 14 years. This long discovery phase has led to the exploration of alternative approaches to accelerate drug development. One such approach is drug repositioning, which involves using already approved drugs for new indications. Numerous pharmaceutical companies are adopting drug repositioning to redevelop approved and previously unsuccessful drug molecules and thus offer new treatment options for various medical conditions.

In this thesis, network graphs were created for the complexity of the interactions of Mercaptopurine, Tioguanine, Prednisolone, Dexamethasone, Vincristine, Methotrexate, Daunorubicin, Cytarabine drugs used in the treatment of childhood acute leukaemia with other drugs and the molecules possessed by the drugs interacting with these drugs were investigated. Molecular calculations were made on the relationship between histone deacetylase (HDAC) genes and childhood acute leukaemia, and the molecular fingerprints that affect the reuse process were determined by explainable artificial intelligence methods. Separate evaluations were made for IC50 and inhibition standard types, which have the highest frequency in HDAC genes. Regression and classification algorithms and model performances were compared and the most successful algorithm was DecisionTreeRegressor. In addition, a simulation environment was designed for all the findings obtained.

The study has important implications for the literature as it provides a setting for drug repositioning and potentially guides the identification of more effective therapeutic strategies. The systematic framework developed increases the reproducibility and scalability of future drug repositioning studies, ultimately contributing to the advancement of more efficient and targeted drug repositioning approaches with the potential to discover new therapeutic options. Future work could include expanding and improving the dataset and combining

ensemble methods and algorithms. Clinical trials are also important for the reliability and accuracy of the findings.

Keywords: Histone deacetylase, childhood acute leukaemia, explainable artificial intelligence, regression, classification, drug-drug interaction, network graph, simulation.

2023, 100 pages



TEŞEKKÜR

Bu araştırma için beni yönlendiren, karşılaştığım zorlukları bilgi ve tecrübesi ile aşmamda yardımcı olan değerli Danışman Hocam Doç. Dr. Utku KÖSE'ye teşekkürlerimi sunarım. Tez izleme komitesinde yer alan, yönlendirme ve destekleri ile bana yardımcı olan değerli hocalarım Prof. Dr. Tuncay AYDOĞAN ve Prof. Dr. Ecir Uğur KÜÇÜKSİLLE'ye teşekkür ederim.

Tezimin her aşamasında beni yalnız bırakmayan aileme sonsuz sevgi ve saygılarımı sunarım.

İlhan UYSAL
ISPARTA, 2023



ŞEKİLLER DİZİNİ

	Sayfa
Şekil 1.1. Geleneksel ilaç keşif süreci ile ilacın yeniden kullanım amacının karşılaştırılması.....	4
Şekil 1.2. Kemik anatomisi	7
Şekil 3.1. HDAC Genleri	19
Şekil 4.1. İlaç etkileşimleri veri seti	30
Şekil 4.2. En çok etkileşime giren ilk 8 ilaç ve karşılık gelen düğüm derecesi.....	31
Şekil 4.3. En çok etkileşime giren ilk 8 ilacın ağ görselleştirmesi.....	32
Şekil 4.4. Bazı HDAC moleküllerinin SMILES Formatları	34
Şekil 4.5 Google Drive'a bağlanma ve Glob ile dosyaları birleştirme	35
Şekil 4.6. Standart Türlerin Frekans Değerleri	36
Şekil 4.7. IC50 için Moleküllerin SMILES, pIC50 ve Lipinski Değerleri	37
Şekil 4.8. İnhibisyon için Moleküllerin SMILES, pIC50 ve Lipinski Değerleri.....	37
Şekil 4.9. HDAC genleri aday moleküllerinin ic50 ve inhibisyon standart türleri için aktiflik durumları	39
Şekil 4.10. Moleküllerin LogP ve MW değerleri	40
Şekil 4.11. Moleküller arasındaki PIC50 ve biyoaktivite ilişkisi.....	42
Şekil 4.12. Moleküllerin moleküler ağırlık ve biyoaktivite ilişkisi	43
Şekil 4.13. DecisionTreeRegressor modelinin IC50 için regresyon grafiği	48
Şekil 4.14. DecisionTreeRegressor modelinin İnhibition için regresyon grafiği.....	49
Şekil 4.15. Smiles molekül yapılarının görseli.....	50
Şekil 4.16. ConfusionMatrix (Karmaşıklık Matrisi)	51
Şekil 5.1. IC50 için SHAP değerlerine göre waterfall grafiği.....	57
Şekil 5.2. İnhibisyon için SHAP değerlerine göre waterfall grafiği.....	57
Şekil 5.3. IC50 için SHAP değerleri force grafiği	62
Şekil 5.4. İnhibisyon için SHAP değerleri force grafiği.....	62
Şekil 5.5. IC50 için decision grafiği-1	63
Şekil 5.6. IC50 için decision grafiği-2	64
Şekil 5.7. İnhibisyon için decision grafiği-1.....	65
Şekil 5.8. İnhibisyon için decision grafiği-2	66
Şekil 5.9. IC50 için BeeSwarm Grafiği.....	68
Şekil 5.10. İnhibisyon için BeeSwarm Grafiği	70
Şekil 5.11. IC50 için Violin Grafiği	72
Şekil 5.12. İnhibisyon için Violin grafiği	73
Şekil 5.13. IC50 için Bar grafiği	74
Şekil 5.14. İnhibisyon için Bar grafiği	75
Şekil 5.15. IC50 için Scatter grafikleri.....	765
Şekil 5.16. İnhibisyon için Scatter grafikleri	77
Şekil 5.17. IC50 için Heatmap grafiği	78
Şekil 5.18. İnhibisyon için Heatmap grafiği	78
Şekil 5.19. Çocukluk çağı akut lösemisi tedavisinde kullanılan ilaçların etkileşimi için oluşturulan benzetim ortamı.....	79

ÇİZELGELER DİZİNİ

	Sayfa
Çizelge 1.1. ALL Tedavisi için İlaçlar ve Faz Aşamaları	7
Çizelge 3.1. Kimyasal veritabanlarının karşılaştırılması	24
Çizelge 4.1. IC50 için Regresyon performans metriklerinin karşılaştırılması.....	45
Çizelge 4.2. İnhibisyon için Regresyon performans metriklerinin karşılaştırılması.....	47
Çizelge 4.3. Sınıflandırma modellerinin performans metrikleri	55
Çizelge 5.1. FP moleküllerinin modele etkisi ve bit altyapısı(Şekil 5.1 için)	58
Çizelge 5.2. FP moleküllerinin modele etkisi ve bit alt yapısı (Şekil 5.2 için)	61



SİMGELER VE KISALTMALAR DİZİNİ

AI	Yapay Zekâ
ALL	Akut Lenfobastik Lösemi
AML	Akut Miyeloid Lösemi
CADD	Bilgisayar Destekli İlaç Tasarımı
FDA	Amerika Birleşik Devletleri Gıda ve İlaç Dairesi
HDAC	Histon Deasetilazlar
IC50	Yarı maksimum inhibitör konsantrasyon
Log P	Lipofilite katsayısı
ML	Makine Öğrenmesi
mRNA	Messenger Ribonükleik Asit
PIC50	İlaç potens ölçütü
RMSE	Root Mean Squared Error (Kök Ortalama Kare Hatası)
QAFFP	QSAR Türevi Afinite Parmak İzleri
QSAR	Kantitatif Yapı-Aktivite İlişkisi
SHAP	SHapley Additive ExPlanations
XAI	Açıklanabilir Yapay Zekâ

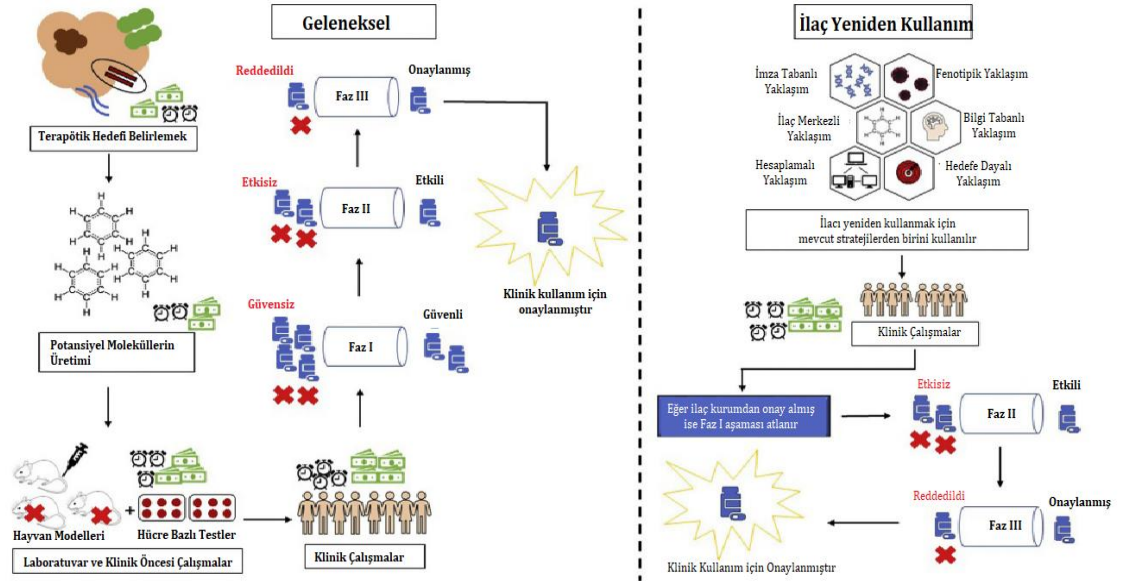
1. GİRİŞ

İlaç keşfi, değerli ve etkili ilaç moleküllerinin keşfedilmesini ve geliştirilmesini gerektiren çok yönlü ve zorlu bir yolculuktur. Terapötik faydaları olan ticari olarak uygulanabilir bir ilacın oluşturulması veya yeniden kullanımı için çeşitli disiplinler arası iş birliğine dayalı bir yaklaşım gerektirmektedir. İlaç araştırmalarının amacı, hastalıkların tedavisine yönelik yeni perspektifler sunarak hastalık süreçlerine odaklanmak ve yeni ilaçlar keşfetmektir. İlaç keşfi ve geliştirme süreci, potansiyel ilaç adaylarının belirlenmesinden başlayarak hedef molekülün sentezi, karakterizasyonu, taranması ve terapötik özelliklerinin değerlendirilmesine kadar birçok aşamayı kapsamaktadır. Bu aşamalarda, binlerce bileşik içinden sadece bir molekül seçilmektedir. Seçilen molekül, etkinlik açısından tatmin edici sonuçlar veriyorsa, klinik deneylerle ilaç geliştirme süreci başlamaktadır. İlaç keşfi ve geliştirme süreci yüksek bütçeler ve ileri düzey teknolojiler gerektiren pre-klinik, klinik ve klinik sonrası araştırmaları içermektedir. Ortalama olarak, ilk ilaç keşfi aşamasından bir farmasötik ürünün nihai olarak piyasaya sürülmesine kadar geçen süre yaklaşık 12 ila 14 yıldır. İlaç keşif ve geliştirme sürecinde, hedef molekülün belirlenmesi için literatür araştırmaları yapılır. Klinik öncesi araştırma, hayvan modelleri üzerinde yapılan deneyler yoluyla ilacın farmakolojik ve toksikolojik yönlerinin değerlendirilmesini kapsamaktadır. Bunu takiben, ilacın etkilerini değerlendirmek ve uygun dozajları belirlemek için sağlıklı bireyler üzerinde Faz 1 çalışmaları yürütülmektedir. Faz 2 çalışmalarına geçildiğinde, ilacın etkinliğini, dozaj ayarlamalarını ve güvenlik profilini daha fazla araştırmak için küçük bir hasta grubu incelenmektedir. Faz 3 çalışmaları ise ilacın daha geniş bir popülasyonda etkinliğini kanıtlamak ve yan etkilerini belirlemek için yapılmaktadır. İlaç keşif ve geliştirme süreci, hedef molekül seçimi, klinik öncesi çalışmalar ve klinik deneylerin sonraki aşamaları (Faz 1, Faz 2 ve Faz 3) dahil olmak üzere çeşitli aşamaları içermektedir. Her aşama, bir sonraki aşamaya geçmeden önce ilaç adayının güvenliğinin, etkinliğinin ve potansiyelinin değerlendirilmesinde çok önemli bir rol oynamaktadır (Kırboğa vd. 2022).

Yapay zekâ ve farmakoloji alanındaki gelişmeler son zamanlarda büyük bir ivme kazanmıştır. Yapay zekâ teknolojileri, ilaç keşfi çalışmalarında etkili bir şekilde kullanılmaktadır. İlaç keşfi sürecinde karşılaşılan maliyet ve zaman gibi sorunlar, yapay zekânın yardımıyla artık daha az sorun teşkil etmektedir. Bir ilacın onaylanması süreci oldukça pahalı ve uzun bir süre gerektirebilmektedir (Chan vd., 2019; Çelik vd., 2021). Araştırmalar, ilaç keşfi Ar-Ge harcamalarının ortalama 2,8 milyar Amerikan dolarına kadar çıktığını ve bütün geliştirme sürecinin ortalama 12-14 yıl sürdüğünü göstermektedir (DiMasi vd., 2016; Lim, 2023; Sarkar vd., 2023; Wouters vd., 2020; Chang vd., 2023). Bu uzun süre, ilaç geliştirme sürecini kısaltmak için alternatif bir yaklaşım olan ilacın yeniden amaca uygun hale getirilmesi (yeniden konumlandırılması) fırsatlarını ortaya çıkarmaktadır. İlaç yeniden kullanımı, çeşitli düzenleyici kurumlar tarafından onaylanmış bir ilacın yeni bir endikasyon için kullanılması anlamına gelmektedir (Uysal ve Köse, 2022). Birçok ilaç şirketi, daha önce başarısız olan moleküllerini yeniden tasarlayarak çeşitli hastalık koşulları için yeni tedaviler geliştirmektedir. Yeni teknolojiler ve hesaplama araçları, ilaç keşfini daha uygun maliyetli hale getirmiştir. Bu yaklaşım, son zamanlarda FDA tarafından onaylanan ilaçların yaklaşık % 30'unu oluşturmaktadır (Parvathaneni vd., 2019). İlaçların yeniden kullanımı, özellikle nadir görülen bozukluklar gibi tıbbi ihtiyaçların karşılanmadığı durumlarda önemli bir potansiyele sahiptir.

İlaç yeniden kullanım alanında yapılan çalışmalar her geçen gün daha çok önem kazanmaktadır. Ancak uygulama aşamasında ortaya çeşitli zorluklar çıkmaktadır. İlaç adaylarının yeniden kullanım amacına yönelik katı ve hızlı düzenleyici yönergeler olmadığından, düzenleyici kurumlara ilgili bilgileri sağlamak, yeni başlayanlar için aşılması zor bir engel olarak görülmektedir. Genel olarak, yeniden kullanım adayı olan ilaçlar, özellikle önceden amaçlanan bir endikasyon için başarısız olduysa, potansiyel bir zaman kaybı riski taşımaktadır. Bu durumda, aynı ilaç veya bileşiğin birden fazla endikasyon için değerlendirildiği bir geliştirme programının tasarlanması önerilmektedir. İlaçın yeniden kullanım amacı, ilacın biyolojik ve moleküler yöntemlerle etkileşimlerinin tam olarak anlaşılmasını gerektirmektedir. Kapsamlı bir bilgiyle, bu hipotez riski önemli ölçüde azaltılabilmekte ve ilacın başarılı bir şekilde yeniden kullanılmasına yol

açabilmektedir. İlaç firmaları, uygun maliyetli ve kazançlı keşif alanlarına odaklanmaktadır. Ancak, nadir görülen ve ihmal edilen hastalıklar için ilaçların yeniden kullanılması ekonomik getiri garantisi sunmamaktadır. Bu nedenle, bir firma uzmanlaşmış ve daha yerleşik bir araştırma yönergelerine odaklanmalıdır. Ayrıca, ilaç şirketlerinin karşılaştığı bir diğer engel de finansal teşvik ve araştırma fonlarının eksikliğidir. Örneğin, nadir görülen kanser vakalarında yatırım getirisi garantisi olmadan ilaçlarla ilgili araştırmalara yatırım yapılması için çok az teşvik bulunmaktadır. Tüm bu zorluklara rağmen, ilacın yeniden kullanımı yaratıcı stratejiler ve ilaç şirketlerinin isteği ve arzusuyla devam edeceği öngörülmektedir (Vaidya vd, 2019; Kulkarni vd, 2019; Pantziarka vd., 2019; Parvathaneni vd., 2019; Gard, 2019; Shineman vd., 2014; Croset, 2014; Padhy ve Gupta, 2011; Arrowsmith, 2011; Uysal ve Köse, 2022). Şekil 1.1’de geleneksel ilaç keşif yaklaşımları ve ilaç yeniden kullanım amaçlı yaklaşımların temel farklılıkları ve faydaları sunulmuştur. İlaç yeniden kullanım, eski-mevcut ilaçlara yeni tedavi edici kullanımlar bulma sürecidir. İlaç keşfi ise yeni moleküler hedefler keşfederek veya tasarlayarak yeni ilaç molekülleri geliştirme sürecidir. İlaç yeniden kullanım, ilaç keşfine göre daha etkili, zaman tasarruflu, düşük maliyetli ve başarısızlık riski daha az olan bir stratejidir. İlaç yeniden kullanım, ilacın terapötik değerini arttırmakta ve başarı oranını yükseltmektedir. Bu nedenle, ilaç yeniden kullanım, geleneksel ilaç keşfi sürecine alternatif bir yaklaşımdır. İlaç yeniden kullanım, ilacın güvenliği ve farmakokinetiği hakkında zaten bilgi sahibi olduğu için daha az test aşamasına ihtiyaç duyarken, ilaç keşfi yeni bir ilacın güvenliği ve farmakokinetiği hakkında bilgi edinmek için daha fazla test aşamasına ihtiyaç duymaktadır. İlaç yeniden kullanım, daha az zaman ve kaynak gerektirdiği için daha hızlı sonuç alabilirken, ilaç keşfi daha uzun zaman ve kaynak gerektirdiği için daha yavaş sonuç almaktadır. İlaç yeniden kullanım, daha az patent sorunu yaşadığı için daha az engelle karşılaşırken, ilaç keşfi daha fazla patent sorunu yaşadığı için daha fazla engelle karşılaşmaktadır (Rudrapal vd., 2020). Dolayısıyla ilaç yeniden kullanım ya da konumlandırma strateji belirlendikten sonra ilaç keşfine göre Faz1 aşamasını atlayarak Faz 2 aşamasına geçtiği için zaman ve maliyet yönünden daha avantajlıdır.



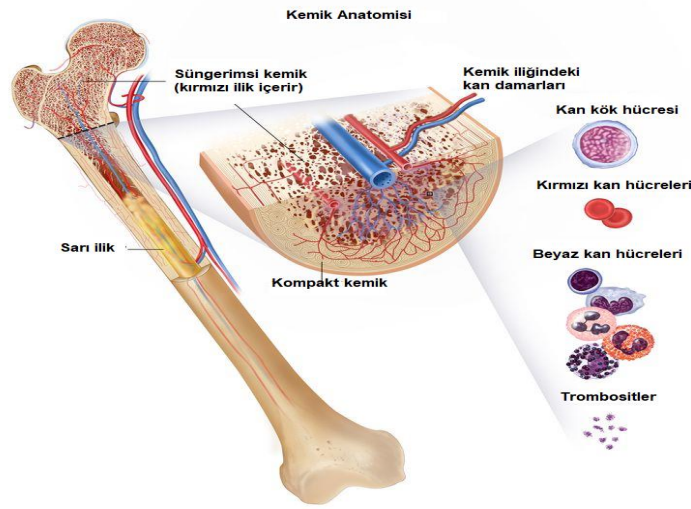
Şekil 1.1. Geleneksel ilaç keşif süreci ile ilacın yeniden kullanım amacının karşılaştırılması (Parvathaneni vd., 2019; Uysal ve Köse, 2022).

Bilgisayar destekli ilaç tasarımı (CADD), ilaçlar ve reseptörler arasındaki moleküler etkileşimleri anlamaya odaklanmaktadır. Bu etkileşimler moleküler yüzey, elektrostatik kuvvet, hidrofobik etkileşim ve hidrojen bağı oluşumu gibi çeşitli faktörlerden etkilenmektedir. Amaç, ilaç-reseptör bağlanması hakkında bilgi edinmek için bu etkileşimleri tahmin ve analiz etmektir (Dhamal vd., 2022). CADD, hastalıkla ilişkili doğal veya sentetik proteinlerin, moleküllerin veya ürünlerin bilgisayar ortamında tasarlanması ve test edilmesi yoluyla üretim sürecine katkıda bulunmaktadır. İlaç keşfinde CADD gibi hesaplamalı yöntemler, bileşik sentezini hızlandırmak için yüksek verimli tarama tekniklerinden ve otomatik sistemlerden yararlanmaktadır (Tang vd., 2006). Bu teknolojileri kullanarak CADD, daha fazla değerlendirme ve optimizasyon için bileşiklerin hızlı bir şekilde üretilmesini sağlamaktadır. Seçilen öncü bileşiğin toksisite, kanserojenlik, sentezin karmaşıklığı ve verimlilik gibi faktörler test edildikten sonra kabul edilmekte veya reddedilmektedir. CADD süreci, hedef molekülün bağlanma bölgesini belirlemekle başlar ve aday molekülün hedef yapılarla etkileşiminden elde edilen biyolojik afinite skorlarıyla devam etmektedir (Vemula vd., 2022). Lipinski kuralları, uygun ilaç özelliklerine sahip olabilecek bileşiklerin seçilmesi için kriterler sağlamaktadır (Karami vd., 2022).

CADD, çeşitli veri tabanlarından yararlanarak ve Lipinski'nin kuralına bağlı olarak ilaç keşfi ve geliştirme sürecini hızlandırmada önemli bir rol oynamaktadır (Katsila vd., 2016). CADD teknikleri, KantitatifYapı-Aktivite İlişkisi (QSAR), kombinatoriyal kimya, sanal tarama ve yerleştirme gibi yazılım araçlarını kullanarak belirli hastalıklara özel ligandların, aktivatörlerin veya inhibitörlerin tasarlanmasına olanak tanımaktadır. Bu metodolojiler, ilaç keşfinde molekül keşfinin verimliliğini ve kapsamını artırmaktadır (Duarte vd., 2019). QSAR küçük moleküllerin deneysel olarak gözlemlenen bağlanma özelliklerini, kimyasal yapı ve fizikokimyasal özellikler gibi moleküler tanımlayıcıları ile ilişkilendiren bir yöntemdir. QSAR modelleri, bu ilişkileri analiz ederek küçük molekül adaylarının aktivitelerini tahmin edebilmekte ve isabet optimizasyonuna rehberlik etmektedir. Farmakofor modelleme ve homoloji modelleme, CADD'de önemli rol oynamakta, ilaç-reseptör etkileşimlerinin üç boyutlu yönlerine ilişkin içgörüler sunmakta ve yenilikçi terapötiklerin geliştirilmesine yardımcı olmaktadır. Bu teknikler, ilaçların hedefleriyle nasıl etkileşime girdiğine dair araştırmacıların idrakını geliştirmekte ve gelişmiş özelliklere sahip yeni bileşiklerin tasarlanmasına yardımcı olmaktadır. (Kırboğa vd., 2022). CADD, moleküler kimya açısından maliyeti ve işçiliği azaltarak birçok ilacın geliştirilmesini kolaylaştırırken yapay zekâ algoritmalarını kullanmaktadır. Ayrıca, ilaç keşfi ve kemoinformatik arasındaki sinerjiyi artıran moleküler tanımlayıcıları vurgulayan makine öğrenimi yaklaşımları da bulunmaktadır.

İlaç yeniden kullanım çalışmaları lösemi gibi kanserlerde kritik bir önem arz etmektedir. Araştırmalar, yeni ilaç keşiflerindeki % 13.8'lik başarı içerisinde kanser ilaçlarının sadece % 3.4'lük bir oranda pay sahibi olduğunu göstermektedir (Lim, 2023). Ayrıca, son gelişmeler kanser ilaçlarında erişimde dünya çapında krizlerin olduğunu işaret etmektedir (Anadolu Agency, 2022; Cohen ve Musa, 2023; Das, 2022; Ellis-Petersen ve Senanayake, 2022; Nonzee ve Luu, 2019; Rees, 2023). Nature Reviews Clinical Oncology dergisi de bu durumu uzmanlar eşliğinde onaylamaktadır (Barrios vd., 2023). Çocukluk çağı akut lösemisi, çocuklarda kan ve kemik iliğini etkileyen bir kanser türünü ifade etmektedir. Bir kemiğin yapısı kompakt kemik, süngerimsi kemik ve kemik iliğinden oluşmaktadır. Kemiğin anatomisini gösteren görsel Şekil 1.2'de

verilmiştir. Kemiğin dış tabakası kompakt kemikten oluşmaktadır. Süngerimsi kemik ağırlıklı olarak kemiklerin uç kısımlarında bulunmaktadır ve kırmızı ilik içermektedir. Çoğu kemiğin merkezinde kan damarları açısından zengin olan kemik iliği bulunmaktadır. Kırmızı ve sarı olmak üzere iki tür kemik iliği vardır. Kırmızı ilik kırmızı kan hücrelerine, beyaz kan hücrelerine veya trombositlere dönüşme yeteneğine sahip kan kök hücrelerini barındırmaktadır. Öte yandan, sarı ilik esas olarak yağdan oluşmaktadır (Ulusal Kanser Enstitüsü, 2022). Bu kanser türü çocuklar arasında yaygındır ve bu yaş grubundaki tüm kanser vakalarının yaklaşık % 30'unu oluşturmaktadır. Çocukluk çağı akut lösemisinin iki ana türü vardır: Akut Lenfoblastik Lösemi (ALL) ve Akut Miyeloid Lösemi (AML). Çocukluk çağı akut lösemisinin belirtileri değişkenlik gösterebilir ancak yorgunluk, soluk cilt, sık enfeksiyonlar, kolay morarma veya kanama, kemik veya eklem ağrısı, şişmiş lenf düğümleri ve karın rahatsızlığı olabilmektedir. Çocukluk çağı akut lösemisi için tedavi seçenekleri mevcuttur. ALL ve AML için tedavinin temel dayanağı, kanser hücrelerini öldürmek için ilaçların kullanılmasını içeren kemoterapiyi içermektedir. Kemoterapi rejimlerinde kullanılan spesifik ilaçlar, lösemnin türüne, evresine ve ayrıca bireysel hasta faktörlerine bağlı olarak değişebilmektedir. Çocukluk çağı akut lösemisi için yaygın olarak kullanılan kemoterapi ilaçları arasında metotreksat, vinkristin, daunorubisin, sitarabin ve prednizon bulunmaktadır. Bazı durumlarda, lösemnin spesifik özelliklerine ve hastanın durumuna bağlı olarak radyasyon tedavisi, kök hücre nakli veya hedefe yönelik tedavi gibi ek tedaviler önerilmektedir. Çocukluk çağı akut lösemisi için tedavi planlarının bireyselleştirilmesi, çocuğun yaşı, genel sağlık durumu ve lösemnin spesifik özellikleri de dahil olmak üzere çeşitli faktörlere dayalı olarak sağlık uzmanlarından oluşan multidisipliner bir ekip tarafından belirlenmesi gerekmektedir. ALL tedavisinde ilaç firmalarının, tedavide kullanılabilecek ilaçlar için hangi aşamada olduklarını gösteren rapordan bir kesit Çizelge 1.1'de verilmiştir



Şekil 1.2. Kemik anatomisi

Çizelge 1.1. ALL Tedavisi için İlaçlar ve Faz Aşamaları (DelveInsight, 2023)

İlaçlar	Şirket	Faz
Inaticabtagene autoleucel	Juventas Cell Therapy/CASI Pharmaceuticals	Kayıtlı
Venetoclax	AbbVie	Faz III
CPX-351	Jazz Pharmaceuticals	Faz II
Obecabtagene autoleucel	Autolus	Faz I/II
Navitoclax	AbbVie	Faz I/II
UCART22	Collectis	Faz I

AML ve ALL’de tedavide Faz 3 aşamasına geçilemediği, ilaç firmalarının da ilaç üretme çabasında da Faz 3 aşamasından ileriye henüz geçemediği DelveInsight’ın Nisan 2023 araştırmasından anlaşılmaktadır. Bu durum çalışmanın motivasyonunu oluşturan en önemli unsurlardan biridir. Bu bağlamda tezin kapsamı, histon deasetilaz genleri çerçevesinde çizilmiş ve bu genlerle ilişkili olduğu bilinen çocukluk çağı akut lösemisinde ağ grafiklerinden yararlanarak moleküler hesaplamaların açıklanabilir yapay zekâ ile desteklenmesi ve tedavide kullanılan ilaçların etkileşimindeki karmaşıklığı gidermeyi içermektedir. Çalışma kapsamında, ChEMBL veritabanı üzerinden çocukluk çağı akut lösemisi ile ilişkili olduğu bilinen histon deasetilaz genleri seçilmiştir. Lipinski'nin kurallarına bağlı kalınarak moleküller seçildikten sonra en uygun adayları belirlemek için çeşitli

yapay zekâ algoritmaları kullanılmıştır. Açıklanabilir yapay zekâ (Explainable Artificial Intelligence – XAI) teknolojisinin kullanılması ve moleküler özelliklerin ilgili modeller ile değerlendirilmesi yoluyla, modellerin moleküler parmak izleri üzerindeki performans metrikleri hesaplanmıştır. Tez çalışmasını diğer araştırmalardan farklı ve özgün kılan yönü, moleküller üzerinde hem regresyon hem de sınıflandırma yapılarak XAI ile elde edilen sonuçların bir dayanağı olması ve ilaç ilaç etkileşimlerinde karmaşıklığın ağ grafikleri ile analiz edilerek kullanılmasıdır. Literatüre, moleküler özelliklerin incelenmesi, ilaç etkileşimlerinin ağ grafikleri ile analiz edilerek karmaşıklığın giderilmesi ve XAI'nın başta kanser olmak üzere ilaç yeniden kullanım çalışmalarında etkin olarak kullanılabilmesi konusunda önemli bir katkı sağlamaktadır.

2. KAYNAK ÖZETLERİ

Bu tez konu kapsamı dikkate alınarak gerçekleştirilen literatür taramasında; histon deasetilaz (HDAC), çocukluk çağı akut lösemisi, kimyasal yapı ve biyoaktivite ile ilaç yeniden kullanım (drug repurposing) başlıkları altındaki farklı yayınlar incelenmiştir. Literatürde yer alan çalışmalarda HDAC'lerin, kanser tedavisinde ilgi çekici bir hedef olduğu ve hematolojik kanserler dahil olmak üzere birçok kanserde farklı karakteristikte bir ekspresyon gösterdiği, bazı HDAC genlerinin, ALL gibi kanserlerde yüksek düzeyde ifade edildiği, HDAC inhibitörlerinin, kanser tedavisinde umut verici ilaçlar olarak ortaya çıktığını, ancak toksisiteleri nedeniyle kullanımlarının sınırlı olduğu, HDAC inhibitörlerinin yeniden tasarlanmasının, yeni tedavi seçenekleri keşfetme potansiyeline sahip olduğunu, moleküler yerleştirme ve QSAR gibi hesaplama yöntemleriyle keşfedilebileceği ve mevcut ilaçların yeniden tasarlanmasının yeni inhibitörlerin keşfedilmesi için etkili bir yol olabileceği ifade edilmiştir. Bu bağlamda, mevcut tez çalışmasının kuruluş ve karşılaştırmalı analiz aşamalarını önemli ölçüde etkilemiş olan dikkate değer araştırmalar önem taşımaktadır. Aşağıdaki temel çalışmalar, bu araştırmanın temel unsurlarının şekillendirilmesinde ve karşılaştırma sürecinin kolaylaştırılmasında önemli bir rol oynamıştır :

HDAC'ler, kanser tedavisinde ilgi çekici bir hedeftir ve hematolojik kanserler de dahil olmak üzere birçok kanserde değişmiş bir ekspresyon göstermektedir. Moreno vd. (2010), yaptıkları çalışmada, 94 çocukluk çağı ALL örneği üzerinde HDAC genlerinin mRNA ekspresyon profili incelenmiştir. ALL örneklerinde, normal kemik iliği örneklerine göre bazı HDAC genleri (HDAC2, HDAC3, HDAC8, HDAC6 ve HDAC7) daha yüksek düzeyde ifade edilmiştir. Bazı HDAC genleri ise T-ALL veya B-hücreli ALL gibi alt tiplerde daha yüksek ekspresyon göstermiştir. Örneğin, HDAC3'ün yüksek ekspresyonu genel hasta grubunda ve T-ALL hastalarında daha düşük sağkalım ile ilişkilendirilmiştir. Benzer şekilde, HDAC7 ve HDAC9'un yüksek ekspresyonu da düşük sağkalım ile ilişkilendirilmiştir. Bu bulgular, HDAC7 ve HDAC9'un çocukluk çağı ALL'sinde kötü prognozla ilişkili olduğunu ve yeni tedavi hedefleri olabileceğini göstermektedir.

Vega-García vd. (2018), histone deasetilaz inhibitörlerinin lösemide umut verici ilaçlar olarak ortaya çıktığını, ancak, özgünlük eksikliğinden kaynaklanan toksisiteleri nedeniyle kullanımlarının sınırlı olduğunu ve bu nedenle, HDAC'lerin belirli durumlardaki rolünün aydınlatılması gerektiğini ifade etmişlerdir. Çocukluk çağı lösemisinde HDAC ifadesinin incelenmesi, kişiselleştirilmiş bir yaklaşımda seçilmiş adaylar için daha spesifik HDAC inhibitörleri seçmeye yardımcı olduğunu belirtmişlerdir. Vega-García vd., akut lösemi tanısı konmuş 211 pediyatrik hastada HDAC inhibitörlerini analiz etmişlerdir ve sonuçta çocukluk çağı lösemisinde HDAC ifadesinin karmaşık resmi hakkında faydalı bilgiler sunulduğunu ve akut lösemi olan seçilmiş pediyatrik hastalarda spesifik HDAC inhibitörlerinin yönlendirilmiş kullanımını desteklediğini ifade etmişlerdir.

Liu vd. (2020), yeni HDAC inhibitörlerini tanımlamak için ilaç yeniden kullanım, farmakofor modelleme, 3D-QSAR ve docking çalışmalarının kullanımını araştırmışlardır. Kanseri ve diğer hastalıkları tedavi etmek için kullanılabilecek yeni inhibitörler keşfetmeyi hedeflemişler ve araştırmaları sayesinde daha önce çalışılmamış birkaç potansiyel HDAC inhibitörü tanımlamışlardır. Ayrıca, güvenlik ve etkinlik açısından zaten test edilmiş olan mevcut ilaçların kullanılmasına izin verdiği için, ilacın yeniden tasarlanmasının yeni inhibitörler keşfetmenin etkili bir yolu olabileceğini de belirtmişlerdir. Dolayısıyla yeni HDAC inhibitörlerini tanımlamak için hesaplama yöntemleri ve ilacın yeniden tasarlanmasının bir kombinasyonunu kullanma potansiyelini belirtmektedir; bu da sonuçta kanser ve diğer hastalıklar için yeni tedavilerin geliştirilmesine imkân verebilmektedir.

Zhao vd. (2015), hepatit C virüsü (HCV) enfeksiyonunun tedavisi için HDAC inhibitörlerinin yeniden tasarlanma potansiyelini araştırmışlardır. HDAC inhibitörlerinin HCV'ye karşı doğrudan etkili antiviral ilaçlar olarak etkinliğini ve anti-HCV ilaç keşfi için mevcut ilaçların yeniden tasarlanmasının avantajlarını tartışmışlardır. HDAC inhibitörlerine ve HCV'ye karşı etki mekanizmalarına genel bir bakış sunduktan sonra, HDAC inhibitörlerinin HCV replikasyonunu inhibe

etme ve konakçı immün yanıtlarını modüle etme potansiyelini tartışarak, onları anti-HCV tedavisi için umut verici adaylar olduğunu belirlemişlerdir.

Sarı vd. (2021), çalışmalarında FDA onaylı ilaçları moleküler yerleştirme kullanarak pan-HDAC ve pan-SIRT inhibitörleri olarak yeniden tasarlamak için potansiyel adayları belirlemeyi amaçlamışlardır. FRED, Glide ve AutoDock Vina kullanarak mevcut kristal yapılara sahip bir dizi klasik HDAC ve SIRT'ye karşı 1502 FDA onaylı ilacı taradıktan sonra, pan-HDAC ve pan-SIRT inhibitör aktivitesine sahip umut verici ilaçları ortaya çıktığını ve kanser de dahil olmak üzere çeşitli durumların tedavisi için yeni olanaklar sağladığını belirtmişlerdir.

Gruhn vd. (2013), çalışmalarında ALL'de HDAC4 ekspresyonunu ve bunun klinik ve biyolojik özelliklerle potansiyel ilişkisini araştırmış ve çocukluk çağı ALL'si için ilgili HDAC izoformlarını tanımlamayı ve bunların tedaviye yanıt ve prognoz üzerindeki etkilerini belirlemeyi amaçlamışlardır. Çalışmada HDAC1, HDAC2 ve HDAC8'in ALL örneklerinde önemli ölçüde daha yüksek ekspresyon gösterdiği belirtilmiştir. Ayrıca, yüksek HDAC4 seviyeleri olumsuz prognostik faktörlerle ilişkilendirilmiş ve HDAC4'ün çocukluk çağı ALL'sinde prednizona kötü yanıtta rol oynayabileceğini ifade etmişlerdir.

Pacaud vd. (2023), yaptıkları çalışmada HDAC inhibitörlerinin klinik uygulamalarından bahsetmişlerdir. Kanser hücrelerindeki anormal epigenetik değişiklikleri tersine çevirmek için yapısal ve işlevsel olarak farklı birkaç HDAC inhibitörü (HDACi) geliştirildiğini ifade etmişlerdir. Giderek artan bir literatürle birlikte, bu ilaçları kanserin çeşitli aşamalarında ve ortamlarında, tek ajan olarak veya hematolojik hem de solid tümör malignitelerinde terapötik bir strateji olarak rasyonel olarak tasarlanmış kombinasyonlarda test etmek için yeterli klinik öncesi gerekçe bulunduğunu ve ilk HDACi vorinostat (Zolinza) kutanöz T-hücreli lenfoma hastalarının tedavisi için onaylandığını ayrıca epilepsi, kistik fibroz, spinal musküler atrofi ve insan immün yetmezlik virüsü enfeksiyonu gibi çeşitli kanser dışı hastalıkların tedavisinde de HDACi'lerin büyük umut vaat ettiğini belirtmişlerdir.

Fernández-de Gortari vd. (2017), çalışmalarında moleküler veritabanlarını temsil etmek için Veritabanı Parmak İzi (DFP) adı verilen yeni bir yaklaşım önermişlerdir. DFP, tüm bir bileşik kütüphanesinin tek bir ikili parmak izi ile temsil edilmesine olanak tanıyarak depolama alanı ve gereksiz hesaplamalar sorunlarını ele almaktadır. Çalışma, çeşitli parmak izi türleri ve veri setleri kullanarak DFP'nin geliştirilmesini ve uygulanmasını göstermekte ve bileşik koleksiyonlarının temel bilgilerini yakalamadaki performansını ortaya koymaktadır. Ayrıca DFP'nin performansını diferansiyel Shannon entropi, k-ortalama kümeleme ve DFP/Tanimoto benzerlik analizlerini kullanarak değerlendirmişlerdir.

Cortés-Ciriano vd. (2020), çalışmalarında potens tahmini için QSAR türevi afinite parmak izlerinin (QAFFP) modelleme performansını araştırmışlar ve QAFFP'nin hesaplanması için, bileşiklerin biyoaktivite alanındaki benzerliklerine göre kodlanmasına ve karşılaştırılmasına olanak tanıyan bir çerçeve önermişlerdir. QAFFP'nin tahmin gücü, çeşitli kanser hücre hatları ve protein hedef veri setleri için ChEMBL veritabanından IC50 verileri kullanılarak kıyaslamışlardır. Çalışmada QAFFP'nin, Morgan2 parmak izleri ve fizikokimyasal tanımlayıcılar kullanılarak elde edilen tahmin gücünün yanı sıra ChEMBL'deki heterojen IC50 verilerinin belirsizliği ile karşılaştırılabilir olan $\sim 0.6-0.9$ pIC50 birimleri aralığında RMSE değerleri ile oldukça öngörücü modellerin oluşturulmasını sağladığını ifadet etmişlerdir. Bu performans seviyesi, benzerlik arama, bileşik sınıflandırma ve iskele atlama görevlerinde QAFFP ile elde edilen yüksek tahmin gücü ile uyumlu olduğunu belirtmişlerdir.

Scalfani vd. (2022), Python RDKit ve NetworkX iş akışı kullanılarak Kimyasal Uzay Ağlarının (CSN'ler) nasıl oluşturulacağından bahsetmişlerdir. Biri RDKit 2D parmak izi Tanimoto benzerlik değerlerine, diğeri ise maksimum ortak alt yapı benzerlik değerlerine dayanan iki farklı CSN oluşturmak için adım adım bir yaklaşım sunmuşlardır. Biyoaktivite öznitelik değerine dayalı olarak düğümleri renkle temsil etme yöntemleri, benzerlik değerine dayalı olarak farklı çizgi stillerine sahip kenarlar ve daire düğümlerini 2D yapı tasvirleriyle değiştirme gibi çeşitli farklı CSN görselleştirme özellikleri belirtilmiş ve son olarak

kümeleme katsayısı, derece çeşitliliği ve modülerlik dahil olmak üzere bazı yaygın ağ özellikleri ve analiz hesaplamaları verilmiştir.

Sada Del Real ve Rubio (2023), yaptıkları çalışmada kanser ilaçlarının etkinliğini tahmin etmek için kullanılan Derin Sinir Ağlarındaki (DNN'ler) açıklanabilirlik eksikliğini gidermeyi amaçlayan SparseGO adlı yeni bir sinir ağı mimarisi uygulayarak bu sınırlamaların üstesinden geldiğini ifade etmişlerdir. İlaçların Etki Mekanizmasını (MoA) hesaplamalı olarak keşfetmek için Destek Vektör Makineleri (SVM'ler) ile DeepLIFT adı verilen bir Açıklanabilir Yapay Zeka (XAI) tekniğini entegre etmişlerdir. Odak noktası, kanser hücre hatlarında ilaç etkinliğini tahmin etmek ve tahminlerin altında yatan nedenleri anlamak olan çalışmada, SparseGO'nun seyrek uygulaması, önceki yöntemlere kıyasla GPU bellek kullanımını ve eğitim hızını önemli ölçüde azalttığını, gen ifadesi verilerinin 200 ilaçtan oluşan bir eğitim seti kullanılarak MoA'nın tahmin edilmesine olanak sağladığını vurgulamışlardır. SparseGO, parbendazol, 6-bio ve elesclomol gibi az çalışılmış ilaçlar üzerinde doğrulanmış ve iyi çalışılmış bileşiklerin ötesinde ilaç yanıtını anlamadaki etkinliğini gösterdiğini belirtmişlerdir.

Kırboğa vd. (2022), kalıtsal Friedreich Ataksisi (FA) için potansiyel terapötik seçenekleri araştırmak üzere yapay zekâ tekniklerinin uygulanmasına odaklanmışlardır. Çalışmada özellikle demir şelasyon moleküllerini ve HDAC inhibitörlerini FA için potansiyel tedaviler olarak araştırılmıştır. ChEMBL veri tabanındaki bileşikler kullanarak bir kantitatif yapı-aktivite ilişkisi (QSAR) analizi yapılmıştır. Fe şelasyonu için 436 bileşik ve HDAC inhibisyonu için 1.163 bileşiğin biyoaktivitesi IC₅₀ birimi kullanılarak ölçülmüştür. Model oluşturmak için Rastgele Orman tekniğini kullanılmıştır. PubChem parmak izi kullanılarak oluşturulan modeller diğerlerinden daha iyi performans göstererek yorumlama için uygunluğunu göstermiştir. Çalışmada, XAI ile analiz edilen bileşiklerdeki azot içeren fonksiyonel grupların ve aromatik halkaların önemi vurgulanmıştır.

Li vd. (2023), çalışmalarında onaylı ilaçlar, klinik çalışmalarda test edilen ilaçlar, patentli/klinik öncesi ilaçlar ve araştırma ajanlarına ilişkin kapsamlı bir literatür taraması yapmışlardır. Bu ilaçlarla etkileşime giren farmakolojik öneme sahip moleküller toplanmış ve 32.487 ilaç/ilaç adayı ile etkileşime giren toplam 5067 farmakolojik öneme sahip molekül tespit etmişlerdir. 21.781 hastanın 108 hastalık bölgesi arasında 1539 molekülün, 236 sağlıklı bireyin ADME ile ilgili dört tipik organı arasında 1323 molekülün ve 32 fizyolojik/yaşam için gerekli bölge/organ arasında 1286 molekülün ifade kalıplarını toplamışlardır. Son olarak, 50.180 ilaç-molekül etkileşimi içeren kapsamlı ve hassas bir 'ağ' oluşturmuşlardır. Her ilaç için alt ağa ve tüm ilaçlar için tüm ağa DrugMAP'te ücretsiz olarak erişilebildiğini belirtmişlerdir.

Wang vd. (2019), çalışmalarında ilaçların mevcut kullanımlarının dışında farklı hastalıklar için kullanılabilme potansiyelini araştırmak amacıyla bir konsensüs ters yerleştirme stratejisi geliştirme ve değerlendirme üzerine odaklanmışlardır. Doğrulanmış ilaç hedefleri ve ilaç bileşiklerinden oluşan bir veri tabanı oluşturduktan sonra, bu hedefler üzerinde bir tarama yapmak için konsensüs ters yerleştirme protokolünü kullanmışlardır. Protokolün performansını değerlendirmek için, belirli kriterlere göre seçilen 51 ticari ilaçtan oluşan bir test seti kullanılmıştır. Tahmin performansı, ROC analizi ile incelenmiştir. Çalışmanın sonucunda, kullanılan konsensüs stratejisinin ilaç yeniden kullanım çalışmalarının başarı oranını artırma potansiyeline sahip olduğu sonucuna varmışlardır.

Literatür araştırmalarında HDAC inhibitörlerinin kanser hücrelerinin büyümesini ve yayılmasını engellemek gibi mekanizmalar üzerindeki etkileri incelenmiştir. Bu çalışmalar, çocukluk çağı akut lenfoblastik lösemi (ALL) gibi özel kanser türlerinde de HDAC genlerinin ifade değişikliklerinin hastalık prognozu ve tedavisi üzerindeki potansiyel etkilerini araştırmıştır. Dolayısıyla modellerin elde ettiği çıktıları neye göre belirlediğinin bilinmemesi, HDAC genlerinden genelde IC50 standart türünün kullanılması, ilaç etkileşimleri ile ilgili karmaşıklığın bulunması ve etkileşimli bir simülasyon ortamının olmaması literatürdeki bazı boşluklar olarak değerlendirilmektedir. Bu tez çalışması, HDAC

genlerinden IC50 ve İnhibisyon standart türlerinin aynı anda değerdendirilmesi, modellerin açıklanabilir olması, ilaçların karmaşık ilişkilerinin ağ grafiğı ile temsil edilmesi ve elde edilen çıktılarının bir benzetim ortamına taşınması ile literatürdeki bu boşluğın doldurulması amaçlanmıştır.



3. MATERYAL VE YÖNTEM

Tezin bu bölümü, çalışma içerisinde kullanılan yöntem ve metotların detaylı bir açıklamasını içermektedir. Artan kanser vakalarının olumsuz yönleri ve etkili tedavi seçeneklerinin kısıtlılığı göz önünde bulundurulduğunda, yapılan araştırma ilaç yeniden kullanım stratejilerinin kanser tedavisinde ne denli etkili olabileceğini araştırmaya odaklanmaktadır. Bölümün içerisinde sırası ile çocukluk çağı akut lösemisi, ilaç etkileşimleri, histon deasetilaz genleri, Lipinski'nin beş kuralı, açıklanabilir yapay zekâ, ChEMBL veri tabanı, Python programlama kütüphaneleri ve yapay zekâ modelleri ele alınmaktadır.

3.1. Çocukluk Çağı Akut Lösemisi

Lösemi, kanser tiplerinden biri olup, hücrelerin anormal bir şekilde çoğalması sonucunda ortaya çıkmaktadır. Genetik yatkınlıklar, kalıtsal hastalıklar ve çevresel faktörler lösemiye neden olabilen çeşitli etkenler arasındadır. Özellikle çocukluk çağında sık görülen lösemi, çocuk kanser vakalarının % 30'unu oluşturmakta ve en yaygın malignite (kötü huylu tümör) türü olarak bilinmektedir. Akut lösemiler, çocukluk çağı lösemilerinin % 98'ini oluştururken, köken aldığı hücreye bağlı olarak ALL ve AML olarak sınıflandırılmaktadır. ALL, en sık görülen lösemi şeklidir. Kemik iliği, periferik kan ve diğer organlarda olgunlaşmamış lenfoid hücrelerin aşırı çoğalmasıyla karakterize edilen heterojen bir hematolojik (kan hastalıkları) malignitedir. Kemik iliği, yeterli ve olgun alyuvar, trombosit ve nötrofil üretme kapasitesini kaybeder. Löseminin belirtileri arasında yorgunluk, ateş, gece terlemeleri, kilo kaybı, nefes darlığı, baş dönmesi, enfeksiyon riskinin artması, siyanoz olarak adlandırılan ciltte mavimsi-mor renk değişiklikleri ve kanama bulguları yer almaktadır. Çocuklarda, kol veya bacakların uç kısımlarında veya eklemlerde ağrı tek başına bazen tek belirti olarak ortaya çıkabilmektedir (Akalın ve Yumuşak, 2023; Bordbar vd., 2023; Carroll ve Bhatla, 2016; Demir, 2023).

3.2. İlaç Etkileşimleri

Çocukluk Çağı Akut Lösemisi tedavisinde kullanılan ilaçların diğer ilaçlarla olan ilişkilerini görmek amacıyla ağ analizi ve görselleştirme üzerinden ilaç-ilaç etkileşimlerinin karmaşıklığı incelenmiştir. İlaç-ilaç etkileşimleri, iki veya daha fazla ilacın birbiriyle reaksiyona girerek beklenmeyen zararlı reaksiyonlarla sonuçlanmasıyla ortaya çıkmaktadır. İlaç-ilaç etkileşimleri, çoklu varlıklar arasındaki bağlantıları içerdiği için, bu ilaç etkileşimlerini ağlarla analiz etmenin uygun olacağı değerlendirilmektedir.

3.3. Histon Deasetilaz (HDAC)

Kanser, küresel ölçekte önemli bir ölüm sebebidir ve Dünya Sağlık Örgütü'ne göre her yıl yaklaşık 8,2 milyon insan kanser nedeniyle yaşamını yitirmektedir. Kanser araştırmaları alanında, kanser ve epigenetik modifikasyonlar arasındaki korelasyon nedeniyle yeni anti-kanser ilaçları arayışında bir artış olmuştur. Epigenetik ilaçların dikkate değer bir kategorisi, gen ifadesini değiştirme yeteneğine sahip olan ve sonuçta tümörler içinde programlanmış hücre ölümünü indükleyen HDAC inhibitörleridir. Bu gelişmekte olan alan son yıllarda önemli ilerlemelere tanık olmuştur (Thotala vd., 2015). HDAC inhibitörleri, gen ekspresyonunu modüle ederek ve özellikle tümör hücrelerinde hücre ölümünü indükleyerek etkilerini gösteren önemli bir anti-kanser ilaç kategorisi olarak kabul edilmektedir (Eyal vd., 2005).

HDAC enzimleri, hücresel olayları etkileyebilmeleri nedeniyle potansiyel ilaç hedefleri haline gelmiş ve bu enzimleri inhibe edebilen bileşikler olarak tanımlanmıştır. HDAC inhibitörleri olarak adlandırılan bu bileşikler, söz konusu enzimlerin aktif bölgelerine bağlanarak histonları asetil formunda tutmakta ve gen ifadelerini değiştirmektedir. HDAC inhibitörleri, hücre bölünmesi, apoptozis ve farklılaşma gibi temel hücresel süreçler üzerinde etki göstermekte ve çeşitli hastalıkların tedavisinde kullanılmak üzere araştırılmaktadır. HDAC inhibitörleri, başta kanser olmak üzere spinal kaslar atrofi, Alzheimer hastalığı, diyabet, psikiyatrik hastalıklar ve paraziter enfeksiyonlar gibi çeşitli

hastalık gruplarının tedavilerinde önemli bir potansiyele sahiptir (Küçükoğlu, 2013).

HDAC inhibitörleri, relaps/refrakter ALL durumlarında kullanılan tedavi ajanları olarak yer almaktadır. HDAC inhibitörleri, histon deasetilaz enzimlerini hedef alarak tümör hücrelerinde gen ifadesinde değişikliklere neden olurlar. Vorinostat, relaps/refrakter ALL tedavisinde kullanılan bir HDAC inhibitörüdür (Jeha ve Pui, 2009; Pieter ve Carroll, 2008; Lee-Sherick vd., 2010). HDAC inhibitörleri, gen ekspresyonunu değiştirerek hücre bölünmesi, apoptozis (hücrenin ölmesi beklenen bir zamanda kontrollü olarak ölmesi) ve farklılaşma gibi hücresel süreçleri etkileyerek tedavi potansiyeline sahip olabilir. HDAC gen ailesi, HDAC1'den HDAC11'e kadar farklı üyelerden oluşmaktadır. Mutasyonlar ve ifade seviyeleri de dahil olmak üzere HDAC genlerindeki değişiklikler, hücre döngüsü düzenlemesi, farklılaşma, programlanmış hücre ölümü (apoptoz), hücre yapışması, göç ve yeni kan damarlarının oluşumu (anjiyogenez) gibi çeşitli hücresel süreçlerde değişikliklere yol açmaktadır. Bu değişiklikler tümörlerin gelişimi ve ilerlemesiyle yakından ilişkilidir. Araştırmalar, HDAC mutasyonlarının ve anormal gen ifadelerinin çeşitli kanser tiplerinde ve hematolojik malignitelerde sıkça gözlemlendiğini göstermektedir. Çalışmalar, sağlıklı çocuklardan alınan kemik iliği örnekleriyle karşılaştırıldığında ALL'li çocukların kemik iliği örneklerinde HDAC2, HDAC3, HDAC4, HDAC6, HDAC7 ve HDAC8 gen ekspresyonlarında kayda değer bir artış olduğunu göstermiştir (Klimek et al., 2008; Cress ve Seto, 2000; Balı et al., 2018). Bu tez çalışması kapsamında da HDAC1'den HDAC11'e kadar olan genler veritabanından alınarak listelenmiş, fakat çocukluk çağı akut lösemisi ile ilgili olduğu bilinen HDAC1, HDAC2, HDAC3, HDAC4, HDAC5, HDAC6, HDAC7 ve HDAC8 genleri kullanılmış ve Şekil 3.1'de görseli verilmiştir. Bu listede, organizma, molekül ve hedef moleküllere ilişkin detaylı bilgiler bulunmaktadır.

organism	pref_name	score	species_group_flag	target_chembl_id	target_components	target_type
Homo sapiens	Histone deacetylase 8	7.0	False	CHEMBL3192	[{'accession': 'Q9BY41', 'component_descriptio...	SINGLE PROTEIN
Homo sapiens	Histone deacetylase 10	7.0	False	CHEMBL5103	[{'accession': 'Q969S8', 'component_descriptio...	SINGLE PROTEIN
Homo sapiens	Histone deacetylase 2	7.0	False	CHEMBL1937	[{'accession': 'Q92769', 'component_descriptio...	SINGLE PROTEIN
Homo sapiens	Histone deacetylase 3	7.0	False	CHEMBL1829	[{'accession': 'O15379', 'component_descriptio...	SINGLE PROTEIN
Homo sapiens	Histone deacetylase 4	7.0	False	CHEMBL3524	[{'accession': 'P56524', 'component_descriptio...	SINGLE PROTEIN
Homo sapiens	Histone deacetylase 6	7.0	False	CHEMBL1865	[{'accession': 'Q9UBN7', 'component_descriptio...	SINGLE PROTEIN
Homo sapiens	Histone deacetylase 9	7.0	False	CHEMBL4145	[{'accession': 'Q9UKV0', 'component_descriptio...	SINGLE PROTEIN
Homo sapiens	Histone deacetylase 11	7.0	False	CHEMBL3310	[{'accession': 'Q96DB2', 'component_descriptio...	SINGLE PROTEIN
Homo sapiens	Histone deacetylase 5	7.0	False	CHEMBL2563	[{'accession': 'Q9UQL6', 'component_descriptio...	SINGLE PROTEIN
Homo sapiens	Histone deacetylase 7	7.0	False	CHEMBL2716	[{'accession': 'Q8WUI4', 'component_descriptio...	SINGLE PROTEIN
Homo sapiens	Histone deacetylase 1	7.0	False	CHEMBL325	[{'accession': 'Q13547', 'component_descriptio...	SINGLE PROTEIN

Şekil 3.1. HDAC Genleri

3.4. Lipinski Kuralları (Lipinski'nin 5 Kuralı)

Lipinski kuralı, Veber kuralı ve Ghose kuralı, biyolojik aktivite gösterme potansiyeline sahip ligandların ilaç benzerliğini değerlendirmede kullanılan kurallardan bazılarıdır. Bu kurallar, lipofiliklik ve suda çözünürlük gibi önemli fizikokimyasal özelliklere dayanarak ilaç adayı moleküllerin özelliklerini değerlendirmeye yardımcı olmaktadır. Lipinski kuralı, ilaç benzerliğini değerlendirmede sıklıkla tercih edilen bir kural olarak öne çıkmaktadır. Bunun birkaç nedeni bulunmaktadır (Lipinski, 2000):

- **İyi tanınmış bir kural:** Lipinski kuralı, ilaç tasarımı alanında yaygın olarak kabul gören ve uzun süredir kullanılan bir kuraldır. Bu nedenle, ilaç adayı moleküllerin değerlendirilmesinde güvenilir bir referans noktası olarak kabul edilmektedir.
- **İlaç benzerliği için sağlam bir temel:** Lipinski kuralı, bir molekülün ilaç benzerliği potansiyelini değerlendirmek için moleküler özelliklerin bir kombinasyonunu dikkate alır. Bu özellikler, lipofiliklik, suda çözünürlük, hidrojen bağı verme ve alım gibi önemli fizikokimyasal özelliklerdir. Bu kural, molekülün hücresel geçirgenlik, dağılım ve farmakokinetik özellikleri üzerinde etkili olabilecek temel faktörleri ele almaktadır.
- **Genel geçerlilik:** Lipinski kuralı, genellikle bir molekülün ilaç benzerliği ve oral biyoyararlanım potansiyelini tahmin etmek için kullanılmaktadır.

Bu kuralın temel prensipleri, birçok ilaç molekülünde gözlenen ortak özelliklere dayanmaktadır. Bu nedenle, geniş bir molekül yelpazesinde etkili bir şekilde uygulanabilmektedir.

Lipinski'nin beş kuralına göre, bir molekülün ilaç adayı olarak kabul edilmesi için aşağıdaki özellikleri taşıması gerekmektedir (Lipinski, 2004; Lipinski vd., 2012):

- moleküler ağırlık (MW) ≤ 500 g/mol,
- lipofilite katsayısı $\text{LogP} \leq 5$,
- hidrojen bağı vericisi sayısı ≤ 5 ,
- hidrojen bağı alıcısı sayısı ≤ 10 ve
- molar kırılma değerleri 4-130 arasında olmalıdır.

Moleküler ağırlığın, ilaç moleküllerinin bağırsak ve merkezi sinir sistemi boyunca geçirgenliğini belirlemede önemli bir rolü vardır (Pardridge, 1995; Navia ve Chaturvedi, 1996). Molekül ağırlığı ile ilaç geçirgenlik oranı arasında ters bir ilişki bulunmaktadır ve yüksek molekül ağırlıklı bileşikler daha düşük geçirgenlik oranları göstermektedir. Daha düşük moleküler ağırlığa sahip bileşiklerin oral olarak aktif olma olasılığı daha yüksektir. Bu nedenle, ilaç keşif sürecinde moleküler ağırlığın önemi belirlenmiştir.

Bir ilaç molekülünün lipofiliklik seviyesi, emilim oranını belirleyen önemli bir faktördür. Bu fizikokimyasal özellik genellikle ilacın sulu ortamdaki oranının logaritmasını temsil eden LogP kullanılarak ölçülmektedir.

Bir bileşikte çok sayıda hidrojen bağı donör grubunun bulunması, hücre zarı çift katmanını geçme kabiliyetini engelleyebilir (Paterson, 1994). Bu tür bileşikler, lipofilik ortamlardan ziyade su gibi güçlü hidrojen bağı çözücülerine daha yüksek bir afiniteye sahip olma eğilimindedir. Tarama amacıyla kullanılan ilaç kütüphanelerinin yaklaşık % 92'si beş veya daha az hidrojen bağı donörüne sahip bileşiklerden oluşmaktadır. Bu özellikler ilaç keşfi alanında büyük önem taşımaktadır (Pollastri, 2010).

Bir bileşik su gibi hidrojen bağı oluşturan bir çözücü ile temas ettiğinde, hidrojen bağı alıcılarının varlığı etkileşimlerde çok önemli bir rol oynamaktadır ve sonuçta geçirgenliği etkilemektedir. Bir bileşikteki azot ve oksijen atomlarının sayısının toplanması, hidrojen bağı alıcılarının sayısını tahmin etmek için güvenilir bir yaklaşım sunmaktadır ve bu da oral biyoyararlanımı değerlendirmek için bir gösterge olarak kullanılabilir (Pollastri, 2010).

Bu kurallar, ilaç keşfi ve geliştirme sürecinde potansiyel ilaç adaylarının özelliklerini değerlendirmek için kullanılan önemli araçlardır. Moleküler tasarım ve ilaç optimizasyonunda bu kuralların göz önünde bulundurulması, ilaç moleküllerinin biyolojik aktivitelerini artırma ve farmakokinetik özelliklerini optimize etme açısından büyük önem taşımaktadır.

3.5. Açıklanabilir Yapay Zekâ (Explainable Artificial Intelligence – XAI)

İlaç keşfi alanında, gelişmiş görüntü analizi teknikleri ve moleküler yapı ve işlevin öngörücü modelleri çok önemli araçlar olarak ortaya çıkmıştır. Bu sebeple, bilgisayar destekli ilaç keşfini kolaylaştırmak ve genel süreci geliştirmek için çeşitli yapay zekâ yöntemleri benimsenmiştir (Gawehn, 2016; Zhang, 2017; Chen, 2018; Tang, 2018; Yang, 2019; Muratov, 2020; Hummerich, 2020). İlaç keşfi alanında, sürecin verimliliğini ve etkinliğini artırmak için derin öğrenme algoritmaları, sinir ağları ve modelleme teknikleri yaygın olarak kullanılmaktadır. Bu gelişmiş teknolojiler, modern ilaç keşfi çabalarında çok önemli bir rol oynamaktadır (LeCun vd., 2015; Schmidhuber, 2015; Lenselink, 2017). Girdi verileri ile çıktı ilişkileri arasındaki korelasyonların araştırılması ve nicel yapı-etkinlik ilişkilerinin kurulması, ilaç keşfi alanında kullanılan teknikler arasındadır. Bu yöntemler, verilerin özelliklerini ve bunların potansiyel ilaç adaylarının etkinliği ile nasıl ilişkili olduğunu anlamaya yardımcı olmaktadır (Goh, 2017; Unterthiner, 2014; Wallach, 2015). Bilgisayar destekli ilaç keşfi alanı, moleküler de novo tasarım teknikleri, kimyasal sentezin stratejik planlaması, makromoleküllerin tanımlanması ve protein hedeflerinin tahmini dahil olmak üzere çeşitli alanların ilerlemesinde çok önemli bir rol oynamıştır. Bu alanlarda devrim yaratma konusunda önemli adımlar atmış ve ilaç keşfinin ilerlemesine

büyük katkı sağlamıştır (Mendez, 2020; Merk, 2018; Zhavoronkov, 2019; Schwaller, 2018; Senior, 2020; Yang, 2020; Öztürk, 2018).

Yapay zekâ yöntemlerinin daha doğru bir şekilde yorumlanması için XAI önemli bir rol oynamaktadır. XAI, modellerin açıklanabilirliğini sağlayarak temel karar verme süreçlerini şeffaflandırmak ve doğru tahminleri belirlemek için kullanılmaktadır (Lapuschkin, 2019; Murdoch, 2019; Miller, 2019; Doshi-Velez, 2017).

Tıbbi kimya, ilaç keşfi ve yeniden kullanım sürecini kapsamaktadır ve açıklanabilir yapay zekâ, ilaç etkilerinin yorumlanmasını ve güvenilirliğini artırmada çok önemli bir rol oynamaktadır. XAI sayesinde, biyolojik ve fizikokimyasal etkiler arasında bağlantı kurulmasını sağlayarak ve bu korelasyona dayalı kesin ve uygun modellerin geliştirilmesini kolaylaştırarak, kurulan modellerin güvenilirliği artırılmıştır. Dolayısıyla, XAI'nın amacı ilaç keşfi sürecinde yapılan işlemleri, nasıl yapıldıklarını ve bunlarla ilgili bilgileri açığa çıkarmaktır (Christoph, 2020). Bu tez çalışması kapsamında modellerden elde edilen sonuçların anlaşılabilmesi ve açıklanabilmesi için Python'da Shapley modülü kullanılmıştır.

Shapley değerleri, oyun teorisi ve tahminsel analizlerde önemli bir rol oynayan bir kavramdır. Bu değerler, oyuncuların birlikte çalıştığı bir oyun veya problemde her bir oyuncunun katkısını ölçmek için kullanılmaktadır (Sellereite ve Jullum, 2019). Shapley değerleri, her bir oyuncunun diğer oyuncularla olan etkileşimini ve katkısını dikkate alarak adil bir şekilde dağılım yapmayı sağlar. Shapley değerleri, Lloyd Shapley tarafından 1953 yılında tanıtılmıştır ve başlangıçta oyuncuların kazançlarının nasıl paylaşılacağına ilişkin bir çözüm kavramı olarak ortaya çıkmıştır (Aas vd., 2021). Ancak son yıllarda makine öğrenmesi ve yapay zekâ alanında model açıklanabilirliği ve özellik önem derecelendirmesi gibi birçok uygulama alanında kullanılmaktadır (Lundberg vd., 2020). Shapley değerleri, bir oyuncunun bir oyunda diğer oyuncularla iş birliği yaparak elde ettiği katkıyı ölçmektedir (Shapley, 1953). Shapley değeri, oyuncunun herhangi bir sıralama veya belirli bir pozisyonla bağlantılı olmadan, oyunun tüm olası

koalisyonları üzerinden hesaplanır ve oyuncunun her bir koalisyon içerisindeki ortalama katkısıdır. Shapley değerleri, oyuncuların katılımlarının birbirinden bağımsız olmadığı durumlarda özellikle etkilidir, yani oyuncuların katkıları arasında bağımlılıklar veya etkileşimler bulunmaktadır. Shapley değerleri, oyuncuların katkılarının belirlenmesinde aşağıdaki temel prensiplere dayanmaktadır (Aas vd., 2021; Song vd., 2016; Fryer v., 2020; Covert vd., 2020; Chalkiadakis vd., 2011; Kóczy, 2018; Frye vd., 2020; Merrick ve Taly, 2019; Grömping, 2012):

- **Katılımcı Adilliliği:** Her oyuncu, diğer oyuncularla olan etkileşimine dayalı olarak kazanca adil bir şekilde katkıda bulunmaktadır.
- **Marginallik:** Her oyuncunun katkısı, diğer oyuncuların oyuna katılmasından kaynaklanan marginallik etkilerini yansıtmaktadır.

Shapley değerleri, matematiksel olarak oyun teorisi çerçevesinde hesaplanmaktadır. Bir oyunda N oyuncu olduğunda, Shapley değeri her bir oyuncu için şu şekilde belirlenmektedir (Algaba vd., 2019):

$$\text{Shapley} = \sum [(N - 1)! / (s! (N - s - 1)!)] * [V(S \cup \{i\}) - V(S)] \quad 3.4$$

Burada, S diğer oyuncuların bir kombinasyonunu, s kombinasyondaki oyuncu sayısını, V(S) oyuncuların S kombinasyonunda elde ettiği kazancı ve V(S $\cup$ {i}) oyuncunun S kombinasyonuna kendisi eklenerek elde ettiği kazancı temsil etmektedir (Shapley, 1953).

Shapley değerleri, tahminsel modellerde özelliklerin önem sıralamasını belirlemek, özellik etkileşimlerini anlamak ve model tahminlerini açıklamak için kullanılmaktadır. Bu değerler, model çıktılarındaki değişikliklerin hangi özelliklere bağlı olduğunu belirlemek için önemli bir araçtır. Ayrıca Shapley değerleri, modelin adaletini değerlendirmek ve öznitelik tabanlı haksızlık veya ayrımcılık tespiti gibi uygulamalarda da kullanılabilir (Lundberg ve Lee, 2017; Lundberg v., 2020; Kumar vd., 2020).

3.6. ChEMBL (Chemical Biology Information Resource) Veritabanı

ChEMBL, biyoaktivite, kimyasal yapı, ilaç hedefleri, ilaç keşfi projeleri ve ilaç moleküllerinin farmakolojik profilleri gibi çeşitli kimyasal ve biyolojik verileri içeren bir veritabanıdır. Bu verilerde, biyolojik testlerdeki bileşiklerin aktiviteleri, hedef molekülleriyle olan etkileşimleri ve bu etkileşimlerin niceliksel değerlendirmeleri bulunmaktadır. ChEMBL aynı zamanda kimyasal bileşiklerin yapısal bilgilerini, moleküler hedeflerin tanımlamalarını ve ilaç keşfi projeleriyle ilgili verileri de barındırmaktadır (Gaulton vd., 2017).

Çizelge 3.1. Kimyasal veritabanlarının karşılaştırılması

Özellik	CHEMBL	PubChem	DrugBank
Bileşik sayısı	252.396 (v29)	194.333.804 (v2022.01.10)	44.224 (v5.1.8)
Protein hedeflerinin sayısı	23.547 (v29)	161.323	6.081
Etkileşim sayısı	12.265.411 (v29)	1.613.819.088	1.700.579
Yayın kapsamı	70%	40%	70%
Veri kalitesi	Yüksek	İyi	İyi
Lisans	CC BY-NC-SA 4.0	CC0	CC BY-NC-SA 4.0
Versiyon	v29 (2022-05-26)	v2022.01.10	v5.1.8 (2022-01-20)

ChEMBL, verileri farklı formatlarda sunma yeteneğine sahiptir ve hedef molekül sayısı açısından en büyük veritabanıdır. Ayrıca aşağıda verilen özelliklerden dolayı da ChEMBL veritabanı tercih edilmiştir:

- **Kapsamlı veri kaynağı** : ChEMBL, büyük bir ilaç keşfi ve kimyasal biyoloji veritabanıdır. Biyoaktivite, kimyasal yapı, ilaç hedefleri, ilaç keşfi projeleri ve ilaç moleküllerinin farmakolojik profilleri gibi geniş bir veri yelpazesini içermektedir. ChEMBL, çeşitli kaynaklardan toplanan ve birleştirilen verileri barındırmaktadır. Bu nedenle, ilaç keşfi ve kimyasal biyoloji alanında kapsamlı bir veri tabanına erişmek isteyenler için ChEMBL tercih edilen bir veritabanıdır.

- **Açık erişim ve kullanıcı dostu arayüz** : ChEMBL, açık erişimli bir veritabanıdır ve kullanıcıların kolayca erişim sağlamasını sağlayan bir web tabanlı arayüz sunmaktadır. Verileri arama, filtreleme ve görselleştirme yetenekleriyle birlikte kullanıcı dostu bir deneyim sağlamaktadır. Bu da araştırmacıların ve ilaç geliştirme profesyonellerinin hızlı bir şekilde ilgili bilgilere erişebilmelerine imkân tanımaktadır.
- **İlaç keşif projeleri için değerli kaynak** : ChEMBL, ilaç keşif sürecindeki araştırmacılar ve ilaç geliştirme şirketleri için değerli bir kaynak sağlamaktadır. Veritabanı, ilaç etkinliği, hedefler, bileşikler ve ilaç keşfi projeleriyle ilgili kapsamlı bilgiler içermektedir. Bu bilgiler, ilaç keşfinde strateji geliştirmek, ilaç tasarımı yapmak ve ilaçların farmakolojik profillerini anlamak için kullanılmaktadır.
- **Bilimsel literatüre dayalı veriler** : ChEMBL veritabanı, bilimsel literatürden elde edilen verileri bulundurmaktadır. Bu, güvenilir ve doğrulanabilir bilgilerin kaynağı olarak ChEMBL'i tercih edilmesini sağlamaktadır. Ayrıca, yayınlarla ilişkilendirilmiş ilaç etkinlik verilerini sunduğu için de verilerin güvenilirliğini arttırmaktadır.

3.7. Python Kütüphaneleri

Python, geniş bir kullanıcı topluluğu tarafından desteklenen ve popülerliği hızla artan bir programlama dilidir. Birçok amaç için kullanılan Python, zengin bir kütüphane ekosistemiyle birlikte gelmektedir. Python kütüphaneleri, açık kaynaklı olarak geliştirilmektedir ve genellikle topluluk tarafından desteklenmektedir. Bu, kullanıcıların kütüphaneleri kolayca kullanabilmesi, geliştirebilmesi ve özelleştirebilmesi anlamına gelmektedir. Python kütüphaneleri, programlamayı daha kolay ve verimli hale getiren güçlü araçlardır. Kullanıcı dostu yapıları ve zengin özellikleri sayesinde, Python ile farklı projelerde hızlı ve etkili çözümler üretilebilmektedir. Bu tez kapsamında moleküler özellikleri ve SMILES gösterimlerini analiz etmek için kullanılan bazı kütüphaneler şunlardır (Pypi, 2023):

- Pandas: Veri analizi için kullanılan güçlü bir kütüphanedir. Veri setlerini okuma, filtreleme, dönüştürme ve birleştirme gibi işlemleri kolaylıkla gerçekleştirmeyi sağlamaktadır (Pandas, 2023).
- RDKit: Kimyasal bilim ve moleküler modelleme alanında kullanılan bir kütüphanedir. Moleküler yapıları temsil etmek, kimyasal özellikleri hesaplamak, ilaç keşfi çalışmalarında kullanılan çeşitli araçlar sunmaktadır (Landrum, 2019).
- Seaborn: Görselleştirme için kullanılan bir kütüphanedir. Matplotlib üzerine inşa edilmiş ve daha estetik ve bilgilendirici grafikler oluşturmaya imkân tanımaktadır (Melitoshevich ve Alikulovich, 2023).
- NumPy: Bilimsel hesaplamalar için kullanılan temel bir kütüphanedir. N-boyutlu dizileri ve matematiksel işlemleri etkili bir şekilde yönetmek için kullanılmaktadır (Harris vd., 2020).
- Matplotlib: Veri görselleştirme için kullanılan bir kütüphanedir. Grafikler, çizimler, histogramlar ve daha fazlasını oluşturmayı sağlamaktadır (Matplotlib, 2023).
- Lazypredict: Makine öğrenmesi modellerini hızlı bir şekilde değerlendirmek için kullanılan bir kütüphanedir. Birçok farklı modeli otomatik olarak uygulayarak başarı metriklerini göstermektedir (Lazypredict Documentation, 2023).
- Glob : Python'da dosya adlarını veya dizinleri belirli bir kalıba göre eşleştirmek ve bu eşleşen dosya veya dizinler üzerinde işlemler yapmak için kullanılan bir modüldür. Özellikle dosya adlarının belirli bir örüntüye göre eşleştirilmesi için sıklıkla kullanılmaktadır. Örneğin, belirli bir dizindeki tüm .csv dosyalarını veya belirli bir örüntüye sahip dosyaları bulmak için tercih edilmektedir (Şirin, 2019).
- NetworkX: Karmaşık ağların yapısı, dinamikleri ve işlevlerinin oluşturulması, manipülasyonu ve incelenmesi için kullanılan bir Python kütüphanesidir. Ağ oluşturma, düğüm ve kenar manipülasyonu, ağ görselleştirme ve ağ analizi için algoritmalar dahil olmak üzere ağları veya grafikleri analiz etmek ve modellemek için araçlar sağlamaktadır. NetworkX, sosyal ağ analizi, ulaşım ağları, biyolojik ağlar ve daha fazlası gibi çeşitli alanlarda yaygın olarak kullanılmaktadır (Hagberg vd., 2008).

- Pyvis: Etkileşimli ağ görselleştirmelerinin oluşturulmasına ve görselleştirilmesine olanak tanıyan bir Python kütüphanesidir. NetworkX'in üzerine inşa edilmiştir ve dinamik ve etkileşimli ağ görselleştirmeleri oluşturmak için üst düzey bir arayüz sağlamaktadır. Pyvis, özelleştirilebilir düğüm ve kenar nitelikleri, farklı düzen algoritmaları, yakınlaştırma ve kaydırma gibi etkileşim seçenekleri ve HTML veya JavaScript formatlarına aktarma yetenekleri dahil olmak üzere çeşitli görselleştirme özelliklerini desteklemektedir. Karmaşık ağları görselleştirmek ve keşfetmek için kullanışlı bir araçtır (Pyvis, 2018).
- Streamlit: Veri bilimi ve makine öğrenimi görevleri için etkileşimli web uygulamaları oluşturmayı kolaylaştıran bir Python kütüphanesidir. Etkileşimli kullanıcı arayüzleri oluşturmak için üst düzey bir API sağlayarak veri odaklı uygulamalar oluşturma ve dağıtma sürecini basitleştirir. Streamlit ile Python kodunu kullanarak web tabanlı gösterge tablolarını, görselleştirmeler ve etkileşimli formlar hızlı bir şekilde oluşturabilir ve özelleştirilebilir. Pandas ve Matplotlib gibi popüler veri işleme ve analiz kütüphaneleri ile iyi entegre olmakta ve web uygulaması için de veri işleme ve görselleştirmenin sorunsuz entegrasyonuna olanak tanımaktadır (Streamlit, 2023).

3.8. Yapay Zekâ Modelleri

Yapay zekâ modelleri, açıklanabilir yapay zekâ modelleri ile ilaç yeniden kullanım çalışmalarının temelini oluşturan önemli bir bileşendir. Yapay zekâ alanında yaygın olarak kullanılan ve ilaç yeniden kullanımı çalışmalarında etkili olan çeşitli algoritmalar bulunmaktadır. Bu algoritmaları, büyük veri setleri üzerinde analiz yaparak ilaçların yeniden kullanım potansiyelini ve yan etkileri değerlendirmek, ilaç etkileşimlerini incelemek, yeni tedavi stratejileri geliştirmek ve ilaç keşfi süreçlerini optimize etmek gibi önemli görevleri yerine getirmektedir. Bu tez kapsamında, aşağıda verilen lazypredict paketindeki lazyregressor ve lazyclassifier kütüphanelerindeki temel yapay zekâ yaklaşımları kullanılmıştır:

- **Lazy Regressor** : Çoklu regresyon modellerinin performansını değerlendirmek için hızlı ve kullanışlı bir yol sağlayan bir Python kütüphanesidir. Varsayılan hiperparametrelerle çeşitli regresyon modellerini eğitme ve test etme sürecini otomatikleştirir ve her model için performans ölçümleri sağlamaktadır. Kütüphane scikit-learn üzerine inşa edilmiştir ve regresyon modellerini değerlendirmek için kullanımı kolay bir arayüz imkânı vermektedir. Doğrusal regresyon, karar ağacı regresyonu, rastgele orman regresyonu, destek vektör regresyonu ve daha fazlası dahil olmak üzere çok çeşitli regresyon algoritmalarını desteklemektedir. LazyRegressor, her model için R-kare (belirleme katsayısı), düzeltilmiş R-kare, ortalama kare hata (MSE), ortalama mutlak hata (MAE) gibi çeşitli performans ölçümlerini hesaplamaktadır. Bu ölçütler, regresyon modellerinin uyum iyiliği (modeldeki bağımsız değişkenlerin bağımlı değişkendeki varyasyonu açıklama kabiliyeti) ve tahmin doğruluğu hakkında bilgi vermektedir (Priyadharsini vd., 2022 ; Velmurugan ve Ida, 2023; Bundela ve Rahul, 2022; Pedregosa vd., 2011; Pandala, 2020; LazyPredict Documentation, 2022).
- **Lazy Classifier** : Lazy Classifier, Lazypredict paketi içinde yer alan ve belirli bir veri kümesi üzerinde birden fazla sınıflandırma modelinin performansını değerlendirmek için hızlı ve kolay bir yol sağlayan bir modüldür. Varsayılan ayarları kullanarak çeşitli sınıflandırma modellerini otomatik olarak oluşturmakta ve eğitmekte, ardından her model için performans ölçümlerini hesaplamakta ve görüntülemektedir. LazyClassifier, Logistic Regression, K-Nearest Neighbors (KNN), Support Vector Machines (SVM), Gaussian Naive Bayes, Decision Trees, Random Forest, Gradient Boosting, AdaBoost, Extreme Gradient Boosting (XGBoost), Light Gradient Boosting Machine (LGBM), CatBoost gibi popüler algoritmalar da dahil olmak üzere çok çeşitli sınıflandırma modellerini desteklemektedir. Doğruluk, F1-skoru, Kesinlik, Geri Çağırma, Alıcı İşletim Karakteristik Eğrisi Altındaki Alan (ROC-AUC), Kesinlik-Geri Çağırma Eğrisi Altındaki Alan (PR-AUC), Matthew'un Korelasyon Katsayısı (MCC), Dengeli Doğruluk, Log Kaybı gibi birçok performans metriği kullanmaktadır. LazyClassifier varsayılan olarak her model için doğruluk

metriğini hesaplamaktadır, ancak metrikleri parametre olarak belirtip özelleştirilebilmektedir (Pandala, 2020; LazyPredict Documentation, 2022).



4. ARAŞTIRMA BULGULARI

Bu tez çalışması kapsamında, çocukluk çağı akut lösemi tedavisinde kullanılan onaylı ilaçların diğer ilaçlarla olan etkileşimleri incelenmiş ve histon deasetilaz genlerindeki moleküllerin moleküler hesaplamaları açıklanabilir yapay zekâ modelleri ile değerlendirilerek ilaç yeniden kullanım çalışmaları için bir benzetim ortamı geliştirilmiştir. Bu hedef doğrultusunda öncelikle 'Mendeley Data' web sitesindeki çoklu tip ilaç-ilaç etkileşimlerinin verileri çocukluk çağı akut lösemi tedavisinde kullanılan ilaçlar için filtrelenerek kullanılmıştır (Hui, 2020). Ham veriler, her satırın belirli bir ilaç çifti arasındaki etkileşimi ifade ettiği 1609 satır ve 3 sütundan oluşmaktadır. Bu veri kümesinin bazı satırları Şekil 4.1'de verilmiştir (Uysal ve Köse, 2023).

	drug1_name	drug2_name	weight
435	Gramicidin D	Vincristine	1
628	Cyclosporine	Vincristine	1
664	Cyclosporine	Methotrexate	1
1032	Cyclosporine	Vincristine	1
1192	Cyclosporine	Cytarabine	1
...	...	...	...
190509	Dexamethasone	Nandrolone decanoate	1
190510	Dexamethasone	Phenyl aminosalicylate	1
190511	Dexamethasone	Methyl salicylate	1
190512	Dexamethasone	Trolamine salicylate	1
190513	Dexamethasone	Corticotrin ovine trifluacetate	1

1609 rows × 3 columns

Şekil 4.1. İlaç etkileşimleri veri seti

Çalışmada kullanılan veri setinde, ilaç etkileşimlerinin öneminin sağlanamaması bir dezavantaj olarak düşünülmektedir. Bu sebeple, tüm etkileşimler için "eşit" önem derecesini belirtmek üzere 1 (bir)' lerle doldurulmuş özel bir sütun (ağırlık) eklenmiştir (Uysal ve Köse, 2023). İlaçlar arasında bir dizi

yönlendirilmemiş bağlantı olduğu dikkate alındığında, ağların oluşturulması, yönlendirilmesi ve analiz edilmesi için Python paketlerinden NetworkX paketi kullanılmıştır. NetworkX ile ağlar standart ve standart olmayan veri formatlarında yüklenebilir ve depolanabilir, birçok rastgele ve klasik ağ türü oluşturulabilir, ağ yapısı analiz edilebilir, ağ modelleri oluşturulabilir, yeni ağ algoritmaları tasarlanabilir, ağlar çizilebilir ve çok daha fazlası yapılabilir (Hagberg vd., 2008).

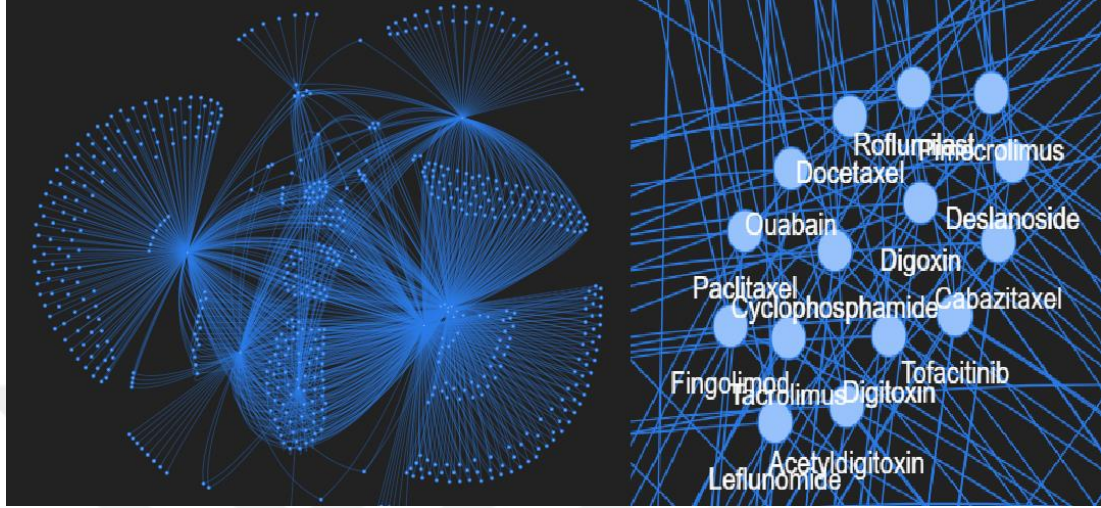
Bir düğümün sahip olduğu bağlantı sayısı düğüm derecesi ile ifade edilmektedir. Bir düğümün merkezi derecesi, sahip olduğu kenar sayısıdır. Bu, 100 etkileşime sahip bir ilaç düğümünün merkezi derecesinin 100 olacağını ifade etmektedir. Her bir düğümün derece merkeziliği hesaplanarak, ilaç etkileşimlerinde en sık yer alan ilk 5 ilaç Şekil 4.2’de verilmiştir. Buna göre çocukluk çağı akut lösemisi tedavisinde kullanılan ilaçlardan Dexamethasone, 389 düğüm derecesi ile en fazla etkileşime giren ilaç olduğu belirlenmiştir. Lenfoma ve lösemi gibi bazı kanser türlerinin tedavisinde kullanılsa da en yaygın kullanımı, inflamasyonu ve bağışıklık sistemi reaksiyonlarını kontrol etmek için çeşitli alerjik reaksiyonları, astımı, romatoid artriti ve diğer otoimmün hastalıkları tedavi etmektir.

```
Top 5 drugs by degree:  
( 'Dexamethasone', 389)  
( 'Prednisolone', 365)  
( 'Vincristine', 309)  
( 'Methotrexate', 168)  
( 'Daunorubicin', 161)
```

Şekil 4.2. En çok etkileşime giren ilk 5 ilaç ve karşılık gelen düğüm derecesi

Ağ görselleştirmesinin görünümünü yapılandırmak veya değiştirmek için ForceAtlas2Based, Hrepulsion, Repulsion, Barnes Hut gibi pek çok özelleştirme bulunmaktadır. Bu çalışmada The Popular Repulsion Solver ve Barnes Hut Physics modelleri, hiyerarşik olmayan düzenler için en hızlı, varsayılan ve önerilen çözücü oldukları için tercih edilmiştir (Uysal ve Köse, 2023). Çocukluk çağı akut lösemi tedavisinde kullanılan ve en çok etkileşime giren 8 ilacın

(mercaptopurine, tioguanine, prednisolone, dexamethasone, vincristine, methotrexate, daunorubicin, cytarabine) Pyvis Barnes Hut grafiđi ile çizdirildiđi zaman ortaya çıkan görüntü Şekil 4.3’de verilmiştir.



Şekil 4.3. En çok etkileşime giren ilk 8 ilacın ağ görselleştirilmesi (Soldaki grafik genel görünüm, sağdaki grafik yakından görünüm)

Ağ grafiđinde, ilaçlar noktalar veya düğümmler olarak temsil edilmektedir. İki ilaç arasındaki etkileşim ise çizgiler veya kenarlar olarak gösterilmektedir. Kenarların kalınlığı veya rengi, etkileşimin şiddetini veya tipini belirtebilmektedir. En çok etkileşime giren Dexamethasone, Prednisolone, Vincristine, Methotrexate ve Daunorubicin ilaçları, çoğunlukla akut lenfoblastik lösemi (ALL) gibi kanser türlerinin tedavisinde kullanılmaktadır. Bu ilaçların birlikte kullanılması, tedavinin etkinliğini artırabilir, ancak aynı zamanda yan etkileri de artırabilmektedir. İlaç etkileşimlerini ağ grafikleriyle göstermenin karmaşıklığı gidermesi son derece etkili bir yöntemdir. Ağ grafiđinde sekiz ilacın hepsiyle etkileşime giren paclitaxel, ouabain, docetaxel, roflumilast, pimecrolimus, deslanoside, digoxin, cabazitaxel, tofacitinib, digitoxin, acetyldigitoxin, leflunomide, tacrolimus, fingolimod, cyclophosphamide ilaçları keşfedilmiştir. Paclitaxel, docetaxel ve cabazitaxel, mikrotübül montaj dinamiklerini bozarak ve hücre döngüsünü G2/M fazında durdurarak apoptozu tetikleyen terpen alkaloidleri sınıfına ait antikanser ilaçlardır. Ouabain, deslanoside, digoxin, digitoxin ve acetyldigitoxin, kardiyak glikozitler sınıfına ait

ilaçlardır. Roflumilast, pimecrolimus, tofacitinib, leflunomide, tacrolimus ve fingolimod, immünomodülatör ilaçlardır. Bu ilaçlar, bağışıklık sisteminin aktivitesini değiştirerek çeşitli inflamatuvar ve otoimmün hastalıkların tedavisinde kullanılmaktadır. Cyclophosphamide, alkilleyici ajanlar sınıfına ait bir antikanser ilaçtır. Bu ilaç, DNA ile çapraz bağlar oluşturarak hücre bölünmesini engellemektedir. Bu ilaçların moleküler yapı bakımından ortak özellikleri vardır. Paclitaxel, docetaxel ve cabazitaxel, aromatik halkalar içeren karmaşık terpen yapılarına sahiptir. Ouabain, deslanoside, digoxin, digitoxin ve acetyldigitoxinde, steroid çekirdeği ile lakton halkası içeren kardiyak glikozit yapıları bulunmaktadır. Roflumilast, pimecrolimus ve tacrolimus, makrolid halkası içeren makrosiklik lakton yapılarına sahiptir. Tofacitinib ve leflunomide, aromatik halkalar ile amid bağı içeren heterosiklik bileşiklerdir. Fingolimod, siklofosfamid ve ouabain, hidrojen bağı oluşturabilecek hidroksil grupları içermektedir (Sousa-Pimenta vd., 2023; Rocha vd., 2023). Bu ilaçların molekül yapıları göz önüne alınarak veritabanından seçilen uygun hedeflerin ilaç potansiyelleri açıklanabilir yapay zekâ modelleri ile değerlendirilmiştir.

Çocukluk çağı akut lösemisi için halihazırda kimyasal yapıların ve aktivitelerin olduğu veritabanından uygun hedefler seçilerek ilaç olabilme durumu araştırılmıştır. Moleküler yapının temel alındığı değerlendirme ile hedefin ilaç potansiyeli belirlenmiştir. Bu kriterlerden biri olan Canonical Smiles (Simplified Molecular Input Line Entry System), molekül verilerinin paylaşılması, arama yapılması ve işlenmesi için yaygın olarak tercih edilen ve bir molekülün kimyasal yapısını ifade etmek için kullanılan standartlaştırılmış bir gösterimdir. Şekil 4.4'de ChEMBL veri tabanından elde edilen histon deasetilaz moleküllerinin SMILES gösterimleri verilmiştir.

canonical_smiles
<chem>CC(=O)N(O)CCCCC(=O)Nc1ccccc1</chem>
<chem>CC(=O)N(O)CCCCCCC(=O)Nc1ccccc1</chem>
<chem>O=CN(O)CCCCC(C(=O)Nc1ccc2ncccc2c1)C(=O)Nc1ccc...</chem>
<chem>O=CN(O)CCCCC(C(=O)Nc1cnc2cccc2c1)C(=O)Nc1cnc...</chem>
<chem>O=CN(O)CCCCC(=O)Nc1ccccc1</chem>
...
<chem>CN1c2ccc(F)cc2C(=O)N2CCc3c([nH])c4ccc(OCCCCC(=...</chem>
<chem>CN1c2ccc(F)cc2C(=O)N2CCc3c([nH])c4ccc(OCCCCC(...</chem>
<chem>CN1c2ccc(F)cc2C(=O)N2CCc3c([nH])c4ccc(OCc5ccc(C...</chem>
<chem>CN1c2ccc(F)cc2C(=O)N2CCc3c([nH])c4ccc(OCc5cccc(...</chem>
<chem>CN1c2ccc(F)cc2C(=O)N2CCc3c([nH])c4ccc(OCc5cccc...</chem>

Şekil 4.4. Bazı HDAC moleküllerinin SMILES Formatları

SMILES (Simplified Molecular Input Line Entry System), bir molekülün kimyasal yapısını temsil etmek için kullanılan bir metin formatıdır. SMILES formatı, molekülün topolojisini ve bağlanabilirlik bilgisini içermektedir. Atomlar, sembollerle temsil edilirken bağlar, atom sembollerinin ardından kullanılan sayılarla gösterilmektedir. Moleküldeki halka yapıları, parantez içindeki numaralarla belirtilmektedir. Bu format, bileşiklerin benimseyebileceği yapısal konfigürasyonları gözlemlemek, ilaç benzerliğini değerlendirmek için Lipinski kurallarını uygulamak ve moleküler yapıdan türetilen aktivite özelliklerini analiz etmek için değerlidir.

Hedef veriler, ChEMBL veritabanından manuel olarak csv formatında indirilebilir ya da Python ile chembl-webresource-client kütüphanesi kullanılarak otomatik olarak alınabilmektedir. ChEMBL çevrimiçi platform üzerinden “histon deasetilaz” ile arama yapılmış ve sonuçta 319 tanesi “Targets” olmak üzere 4964 adet bileşik listelenmiştir. Targets sekmesinden yalnızca insanlarda görülen bir hastalık üzerinde çalışılacağı için sadece “Homo Sapiens” filtrelemesi ve protein komplekslerinde birçok molekülün hem kendi aktiviteleri hem de kompleksin aktivitesi bulunduğundan dolayı, hedef molekülün aktivite verileri tam olarak

doğrulanamadığı için "Single Protein" filtrelemesi yapılmıştır. ChEMBL veri tabanından alınan moleküller, tekrarlayan yapıları olan ve eksik bilgileri bulunan moleküller elenerek seçilmiştir. SMILES formatındaki moleküllerin standart değerleri yine ChEMBL veri tabanından alınmıştır. Moleküllerin standart değeri, farklı koşullardaki özelliklerini hesaplamak için bir referans noktası olarak kullanılmaktadır (Kırboğa vd., 2022).

HDAC genlerinin (HDAC2 - HDAC3 - HDAC4 - HDAC5 - HDAC6 - HDAC7 - HDAC8 - HDAC9) Çocukluk Çağı Akut Lösemisi ile ilişkisi değerlendirileceği için bu veriler csv formatında indirildikten sonra Python'da Glob kütüphanesi ile tek bir dosya haline getirilmiştir. Google Colab'da Google Drive'a bağlanarak glob kütüphanesi ile bu işlem gerçekleştirilmiş ve Şekil 4.5'de kodları verilmiştir.

```
from google.colab import drive
drive.mount('/content/drive')

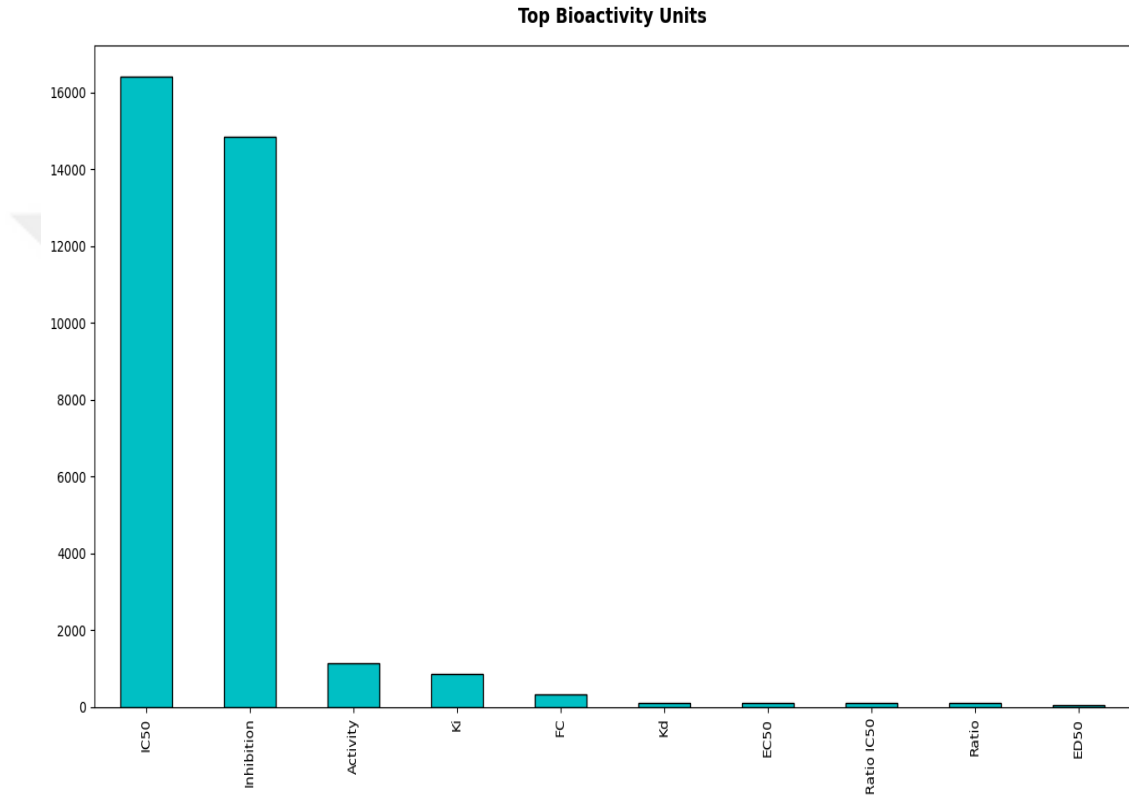
Mounted at /content/drive

x = pd.DataFrame()
for f in glob.glob("/content/drive/MyDrive/doktora_tez/hdac/*.csv"):
    df = pd.read_csv(f, sep=";")
    x = x.append(df, ignore_index=True)
```

Şekil 4.5. Google Drive'a bağlanma ve Glob ile dosyaları birleştirme

Biyoaktivite çalışmalarında, bileşiklerin aktivitesini veya gücünü ölçmek için IC50, Inhibition, Ki, Kd gibi çeşitli standart türler kullanılmaktadır. IC50 (Yarı Maksimal İnhibitör Konsantrasyon), biyolojik bir süreci veya aktiviteyi % 50 oranında engellemek için gereken bileşik konsantrasyonu olarak bilinmektedir. Genellikle enzim inhibisyon deneylerinde kullanılmaktadır ve bileşiğin hedefi inhibe etme gücünü temsil etmektedir. Inhibition (inhibisyon) ise genellikle bir bileşiğin bir enzim veya reseptör gibi belirli bir hedefin işlevini engelleme veya inhibe etme yeteneğini ölçmek için kullanılmaktadır. Biyoaktivite çalışmalarında, inhibisyon çeşitli tahliller veya testler kullanılarak kantitatif olarak ölçülebilmektedir. İnhibisyon derecesi tipik olarak bir kontrol veya referans

numuneye kıyasla inhibisyonun derecesini gösteren bir yüzde veya değer olarak ifade edilmektedir. Örneğin, bir inhibitör % 50 inhibisyon gösterebilir, bu da hedefin aktivitesini yarı yarıya azalttığını göstermektedir. Tek bir dosya haline getirilen HDAC genleri veri kümesindeki "Standard Type" sütununda bulunan biyoaktivite birimlerinin frekansını hesaplanmış ve en çok kullanılan 10 birimi bir çubuk grafik olarak görselleştirilerek Şekil 4.6'de verilmiştir.



Şekil 4.6. Standart Türlerin Frekans Değerleri

Literatürdeki çalışmalar incelendiğinde genellikle standart türlerden hedefe uygun olarak bir tanesi seçilip kimyasal yapı biyoaktivite değerlerinin hesaplandığı görülmektedir. Bu tez çalışmasında frekans değerleri dikkate alınarak hem ic50 hem de inhibisyon standart türleri için moleküler veri analizi yapılmıştır. İlaç tasarımı ve yeniden kullanımı sürecinde, moleküllerin aktiflik durumu değerlendirilerek, etkili ve güvenli ilaç adaylarının seçilmesi hedeflenmektedir. Moleküllerin ilaç olarak kullanılabilirlik potansiyellerini anlamak için aktiflik durumları (aktif, inaktif ve orta değer) hesaplanmış ve

SMILES, pIC50 (IC50 verilerinin daha düzgün bir şekilde dağılmasını sağlamak için negatif logaritmik ölçeğe dönüştürülmüş hali) ve Lipinski özelliklerini (moleküler ağırlık – MW, oktanol ile su arasındaki değişim katsayısı olan lipofilik – LogP, hidrojen bağı donörlerin sayısı – NumHDonors, hidrojen bağı alıcılarının sayısı – NumHAcceptors) gösteren görseller IC50 için Şekil 4.7’ de, inhibisyon için ise Şekil 4.8’de verilmiştir.

	molecule_chembl_id	canonical_smiles	class	MW	LogP	NumHDonors	NumHAcceptors	pIC50
0	CHEMBL138293	CC(=O)N(O)CCCCC(=O)Nc1cccc1	inactive	264.325	2.4232	2.0	3.0	4.000000
1	CHEMBL138626	CC(=O)N(O)CCCCC(=O)Nc1cccc1	inactive	292.379	3.2034	2.0	3.0	4.000000
2	CHEMBL336935	O=CN(O)CCCCC(=O)Nc1ccc2ncccc2c1)C(=O)Nc1ccc...	active	485.544	4.3844	3.0	6.0	6.096910
3	CHEMBL140525	O=CN(O)CCCCC(=O)Nc1cnc2cccc2c1)C(=O)Nc1cnc...	active	485.544	4.3844	3.0	6.0	6.161151
4	CHEMBL141082	O=CN(O)CCCCC(=O)Nc1cccc1	intermediate	250.298	2.0331	2.0	3.0	5.167491
...	...	...	...	...	...	...	...	...
7898	CHEMBL5090035	CN1c2ccc(F)cc2C(=O)N2CCc3c([nH]c4ccc(OCCCCC(=O)Nc1ccc...	active	541.627	5.6140	3.0	5.0	6.494850
7899	CHEMBL5075736	CN1c2ccc(F)cc2C(=O)N2CCc3c([nH]c4ccc(OCCCCC(=O)Nc1ccc...	active	555.654	6.0041	3.0	5.0	7.602060
7900	CHEMBL5088654	CN1c2ccc(F)cc2C(=O)N2CCc3c([nH]c4ccc(Oc5ccc(C(=O)Nc1ccc...	active	561.617	5.8674	3.0	5.0	7.508638
7901	CHEMBL5093940	CN1c2ccc(F)cc2C(=O)N2CCc3c([nH]c4ccc(Oc5ccc(C(=O)Nc1ccc...	intermediate	561.617	5.8674	3.0	5.0	5.443697
7902	CHEMBL5092068	CN1c2ccc(F)cc2C(=O)N2CCc3c([nH]c4ccc(Oc5ccc(C(=O)Nc1ccc...	intermediate	561.617	5.8674	3.0	5.0	5.721246

7903 rows x 8 columns

Şekil 4.7. IC50 için Moleküllerin SMILES, pIC50 ve Lipinski Değerleri

	molecule_chembl_id	canonical_smiles	class	MW	LogP	NumHDonors	NumHAcceptors	pIC50
0	CHEMBL235210	COc1ccc(C(=O)Nc2cc(-c3cccs3)ccc2N)cc1	active	324.405	4.25820	2.0	4.0	7.301030
1	CHEMBL236061	COc1ccc(C(=O)Nc2ccccc2N)cc1	active	242.278	2.52970	2.0	3.0	7.301030
2	CHEMBL235842	CC(=O)Nc1ccc(C(=O)Nc2cc(-c3cccs3)ccc2N)cc1	active	351.431	4.20800	3.0	4.0	7.301030
3	CHEMBL235191	CC(=O)Nc1ccc(C(=O)Nc2ccccc2N)cc1	active	269.304	2.47950	3.0	3.0	7.301030
4	CHEMBL235413	COc1ccc(Nc2ccc(C(=O)Nc3cc(-c4cccs4)ccc3O)cc2)cc1	active	460.555	6.00230	3.0	6.0	7.301030
...	...	...	...	...	...	...	...	...
1078	CHEMBL4851419	COc1cccc(Nc2nc(Nc3ccc(NC(=O)CCCCC(=O)NO)cc3)cc2)cc1	active	546.550	5.77570	5.0	8.0	7.251812
1079	CHEMBL4879096	CC(C)S(=O)(=O)c1cccc1Nc1nc(Nc2ccc(NC(=O)CCCCC(=O)NO)cc2)nc1Nc1cc...	active	572.663	5.06950	5.0	9.0	7.154902
1080	CHEMBL4867091	Cc1cnc(Nc2ccc(NC(=O)CCCCC(=O)NO)cc2)nc1Nc1cc...	active	568.700	5.23882	5.0	9.0	7.200659
1081	CHEMBL4866452	COc1cnc(Nc2ccc(NC(=O)CCCCC(=O)NO)cc2)nc1Nc1cc...	active	584.699	4.93900	5.0	10.0	7.167491
1083	CHEMBL4849600	O=C(NO)c1ccc(Cl)c(NC(=O)c2ccc(-c3ccccc3)cc2Cl)c1	active	401.249	5.03170	3.0	3.0	8.154902

1031 rows x 8 columns

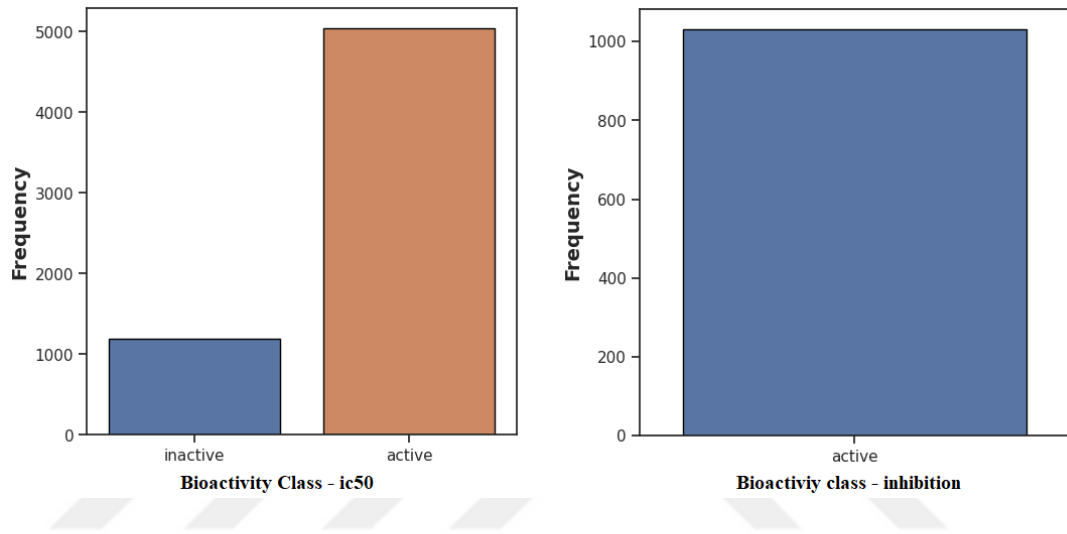
Şekil 4.8. Inhibisyon için Moleküllerin SMILES, pIC50 ve Lipinski Değerleri

HDAC genlerinin PIC50 değerlerinin standart sapması, veri noktalarının ortalamadan ne kadar uzaklaştığını göstermek için kullanılmıştır. Standart sapma büyük olduğunda, veri noktalarının daha geniş bir aralığa yayıldığını göstermektedir. Tersine, küçük bir standart sapma, veri noktalarının birbirine sıkıca kümelendiğini ve aralarında daha az değişkenlik olduğunu ifade

etmektedir. Standart sapmanın büyük olması, testin daha ayırt edici olduğunu ve değerler arasındaki farkın yüksek olduğunu ifade etmektedir. Bu durum, ilaç komplekslerinde inaktif bileşenlerin de bulunduğunu belirtmektedir. Yardımcı maddeler olarak da bilinen inaktif bileşenler, tipik olarak aktif bileşenin terapötik etkinliği üzerinde doğrudan bir etkisi olmayan maddelerdir. Standart sapmanın küçük olması ise, testin daha az ayırt edici olduğunu ve değerler arasındaki farkın düşük olduğunu göstermektedir. Bu durum, aktif bileşenlerin olduğunu ve grupların homojen olduğunu belirtmektedir. Ayrıca, reseptör ve ligand arasındaki afinite (bağlanma gücü) hakkında bilgi edinmek için bu istatistiksel yöntemlerin kullanılabileceği ve afinite ne kadar yüksekse, IC50 değerinin o kadar düşük olduğu da değerlendirilebilir. Hedef molekülün farmakolojik kullanımına geçişi değerlendirmek için kemoinformatik yöntemler kullanılmıştır. Python da Rdkit kütüphanesi ile kullanılan bileşiklerin yapı ve özellikleri analiz edilmiştir. İlacın vücuda alınmasından başlayarak etkisini göstermesine kadar olan süreçler ADME olarak ifade edilen Absorpsiyon (emilim), Distribution (dağılım), Metabolizma (metabolizasyon) ve Eliminasyon (atılım)'dur. İlaçların vücutta hareketi ve etkileşimi açısından önemlidir ve ilacın etkili ve güvenli bir şekilde kullanılabilmesi için dikkate alınması gereken faktörlerdir. ADME özellikleri, vücudun ilahtan ne kadar faydalandığını ölçen bir parametre olan biyoyararlanım değerlerini ölçmek için hesaplanmıştır. Bu özellikler, moleküler esneklik ve hidrojen bağı sayısı gibi faktörlerin biyoyararlanımın temel belirleyicileri olduğunu göstermektedir (Veber, 2002).

Kemoinformatik, kimyasal bileşiklerin moleküler yapılarından yola çıkarak ilaç potansiyelleri ve metabolizma üzerindeki etkileri hakkında önemli bilgiler sağlayabilmektedir (Ferreira, 2019; Özkan, 2019). Bu yöntemler, binlerce molekül arasından potansiyel ilaç adaylarını belirlemek için kullanılmaktadır. ADME özellikleri, ilaç olabilme potansiyeli açısından değerlendirilirken Lipinski'nin 5 kuralı bu değerlendirmede önemli bir rol oynamaktadır (Lipinski, 2004; Oprea, 2001; Leeson, 2007). Bu kurallara uygun moleküller, klinik denemeler sırasında daha iyi performans gösterme eğilimindedir ve klinik denemede düşük yıpranma ile pazarlama şansları artmaktadır (Kırboğa vd., 2022). Moleküllerin biyolojik etkinliklerinin daha net bir şekilde belirlenebilmesi

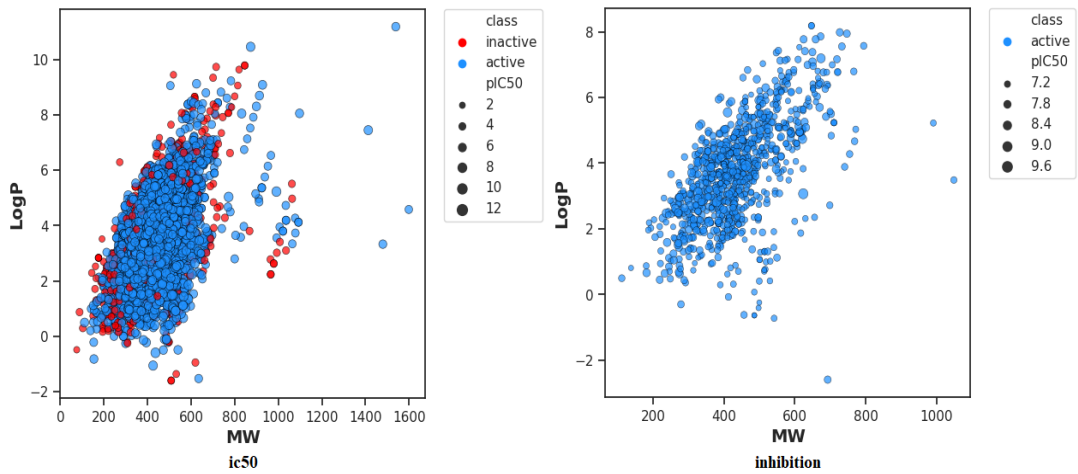
ve bu moleküllerin ilaç potansiyellerinin daha doğru bir şekilde yorumlanabilmesi için orta seviyedeki aktiflik gösteren moleküller çıkarılmıştır. Sonuçta ortaya çıkan aktiflik durumu IC50 ve İnhibisyon standart türleri için görselleştirilerek Şekil 4.9'da verilmiştir. Buna göre IC50 standart türü için aktiflik gösteren moleküllerin sayısı 5042, inaktiflik gösteren moleküllerin sayısı 1190 iken inhibisyon standart türü için inaktiflik gösteren molekül bulunmazken aktiflik gösteren molekül sayısı 1031 olarak tespit edilmiştir.



Şekil 4.9. HDAC genleri aday moleküllerinin ic50 ve inhibisyon standart türleri için aktiflik durumları

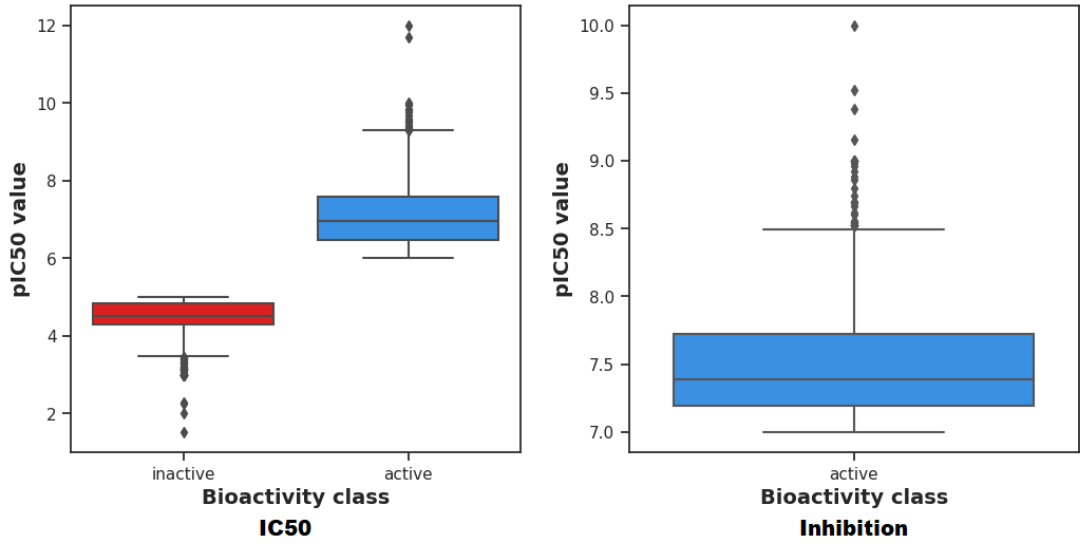
MW değeri ile LogP değeri arasındaki ilişkileri incelemek amacıyla hem IC50 hem de inhibisyon standart türleri için Şekil 4.10'da gösterilen şekilde bir görsel oluşturulmuştur. Grafiğin amacı, ilaç keşfi çalışmasında hangi moleküllerin daha iyi performans gösterdiğini ve hangilerinin elenmesi gerektiğini belirlemeye yardımcı olmaktır. Grafiğin x-ekseni molekül ağırlığını (MW), y-ekseni ise lipofilikliği (LogP) göstermektedir. Noktaların rengi moleküllerin aktivitesini, büyüklüğü ise pIC50 değerini belirtmektedir. pIC50 değeri, bir molekülün hedef enzimi inhibe etmek için gerekli konsantrasyonun negatif logaritmasıdır. Yani pIC50 değeri ne kadar yüksekse, molekül o kadar güçlü bir inhibitördür. Görselde soldaki grafik için (ic50), aktif moleküllerin çoğu düşük MW ve LogP değerlerine sahip olduğu tespit edilmiştir. Bu, onların daha küçük ve daha hidrofilik (bir molekülün hidrojen bağları kurarak suya bağlanabilme özelliği) olduğunu

göstermektedir. Bu özellikler genellikle ilaç olarak kullanılmaya uygun moleküller için arzu edilmektedir. Aktif moleküller arasında pIC50 değeri yüksek olanlar da bulunmaktadır. Bu moleküller hedef enzimi çok iyi inhibe etmektedirler ve potansiyel ilaç adayları olarak görülmektedir. İnaktif moleküllerin çoğunun yüksek MW ve LogP değerlerine sahip olduğu belirlenmiştir. Dolayısıyla onların daha büyük ve daha hidrofobik (bir molekülün sudan kaçınma özelliği) olduğunu göstermektedir. Bu özellikler genellikle ilaç olarak kullanılmaya uygun olmayan moleküllerde görülmektedir. İnaktif moleküller arasında pIC50 değeri yüksek olanlar da vardır. Bu moleküllerin hedef enzimi inhibe etmelerine rağmen, başka faktörlerden dolayı ilaç olarak kullanılamadıklarını göstermektedir. Bu faktörler toksisite, metabolizma veya farmakokinetik olabilmektedir. Görseldeki sağdaki inhibisyon grafiğinde ise noktaların çoğu sol alttan sağ üste doğru bir çizgi oluşturmuştur. Bu durum, MW ve LogP arasında pozitif bir korelasyon olduğunu göstermektedir. Yani molekül ağırlığı arttıkça, lipofiliklik de artmaktadır. Noktaların rengi ise aktif molekülleri belirtmektedir. Grafiğin sol alt köşesindeki noktaların en koyu renge sahip olduğu görülmektedir. Bu durum, bu moleküllerin en yüksek aktiviteye sahip olduğunu ifade etmektedir. Bu moleküller potansiyel ilaç adayları olarak görülmektedir. Grafiğin sağ üst köşesindeki noktalar en açık renk olduğu için, bu moleküllerin en düşük aktiviteye sahip olduğunu göstermektedir. Dolayısıyla bu moleküller ilaç olarak kullanılmaya uygun değildir.



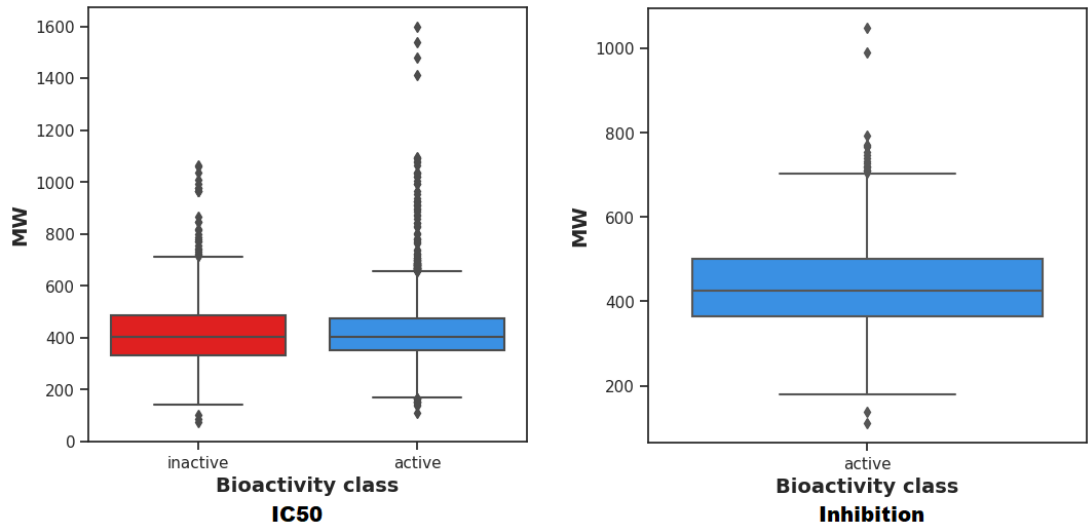
Şekil 4.10. Moleküllerin LogP ve MW değerleri

Biyoaktivite özellikleri ile pIC50 değerleri arasındaki ilişkilerin değerlendirilmesi için Şekil 4.11'deki gibi görselleştirilmiştir. Görseldeki ic50 (soldaki) grafiğinde iki kutu grafiği bulunmakta ve biri inaktif diğeri aktif biyoaktivite sınıfıdır. Kutu grafiği, verilerin dağılımını gösteren bir grafik türüdür. Kutu grafiğinin alt ve üst kenarları, verilerin % 25 ve % 75'lik yüzdelerlerini göstermektedir. Kutu grafiğinin içindeki çizgi ise verilerin ortanca değerini belirtmektedir. Kutu grafiğinin dışında kalan noktalar ise aykırı değerlerdir. Aktif biyoaktivite sınıfının kutu grafiği mavi renkte ve inaktif biyoaktivite sınıfının kutu grafiği kırmızı renkte gösterilmiştir. Aktif biyoaktivite sınıfının ortanca pIC50 değeri yaklaşık 7 iken, inaktif biyoaktivite sınıfının ortanca pIC50 değeri yaklaşık 4.5'tir. Bu durum, aktif moleküllerin inaktif moleküllere göre çok daha güçlü inhibitörler olduğunu ifade etmektedir. Her iki kutu grafiğinde de aykırı değerler bulunmaktadır. Yani bazı moleküllerin biyoaktivite sınıflarına göre beklenenden çok farklı pIC50 değerlerine sahip olduğunu belirtmektedir. Sağdaki inhibisyon görseli için ise grafiğin sadece aktif biyoaktivite sınıfı için bir kutu grafiği vardır. Aktif biyoaktivite sınıfının kutu grafiği mavi renkte ve ortanca pIC50 değeri yaklaşık 7.5'tir. Kutu grafiğinin bıyıkları yaklaşık 7 ile 8.5 arasında uzanmaktadır. Bu, aktif moleküllerin çoğunun bu aralıkta pIC50 değerine sahip olduğunu göstermektedir. Kutu grafiğinin üzerinde birkaç aykırı değer bulunmaktadır. En yüksek değer yaklaşık 10'dur. Bu durum, bu moleküllerin çok daha güçlü inhibitörler olduğunu ifade etmektedir.



Şekil 4.11. Moleküller arasındaki PIC50 ve biyoaktivite ilişkisi

Moleküler ağırlık ile biyoaktivite arasındaki ilişkiyi görselleştirmek için Şekil 4.12'de bir grafik oluşturulmuştur. Grafiğin amacı, ilaç keşfi çalışmasında hangi moleküllerin daha küçük veya daha büyük olduğunu ve bunun biyoaktivite üzerindeki etkisini göstermektir. Grafiğin x-ekseni biyoaktivite sınıfını, y-ekseni ise molekül ağırlığını (MW) ifade etmektedir. Molekül ağırlığı, bir moleküldeki atomların toplam kütesidir ve bir molekülün fiziksel ve kimyasal özelliklerini etkileyebilmektedir. Görselde ic50 (soldaki) için grafiğin iki kutu grafiği bulunmakta ve biri inaktif diğeri aktif biyoaktivite sınıfına aittir. İnaktif biyoaktivite sınıfının kutu grafiği kırmızı renkte ve aktif biyoaktivite sınıfının kutu grafiği mavi renktedir. İnaktif biyoaktivite sınıfının ortanca molekül ağırlığı aktif biyoaktivite sınıfından daha yüksektir dolayısıyla inaktif moleküller daha büyüktür. Her iki kutu grafiğinde de siyah elmaslarla gösterilen aykırı değerler vardır. Bu durum, bazı moleküllerin biyoaktivite sınıflarına göre beklenenden çok farklı molekül ağırlıklarına sahip olduğunu göstermektedir. Sağdaki inhibisyon biyoaktivite sınıfının kutu grafiği mavi renkte ve ortanca molekül ağırlığı yaklaşık 400'dür. Kutu grafiğinin bıyıkları yaklaşık 200 ile 700 arasındadır ve inhibisyon moleküllerinin çoğunun bu aralıkta molekül ağırlığına sahiptir. Kutu grafiğinin üzerinde ve altında ikişer tane aykırı değer bulunmaktadır. Dolayısıyla bazı moleküllerin çok daha küçük veya çok daha büyük olduğunu ifade etmektedir.



Şekil 4.12. Moleküllerin moleküler ağırlık ve biyoaktivite ilişkisi

Özellikle ilaç tasarımı, moleküler tanımlayıcılar, ilaç moleküllerinin hedef proteine olan etkileşimlerini ve biyoyararlanımlarını tahmin etmek için kullanılabilen Padel-Descriptor yazılımı aynı zamanda moleküler özelliklerin hesaplanması ve karşılaştırılması, bileşiklerin benzerlik ve farklılık derecelerinin belirlenmesinde yardımcı olmaktadır. Moleküllerin inhibisyonu, bir kimyasal bileşiğin belirli bir biyolojik hedef üzerindeki etkisini engelleme veya azaltma sürecidir ve özellikle ilaç geliştirme sürecinde, hastalıkların tedavisinde hedef molekülleri hedefleyen ilaçların etkinliğini artırmak veya istenmeyen biyolojik süreçleri durdurmak için önemlidir. İnhibisyon sürecinde yer alan moleküllerin moleküler parmak izleri, PubChem veri tabanından elde edilen 881 farklı moleküler parmak izi sağlayan PaDeL-Descriptor yazılımı kullanılarak hesaplanmıştır. Bu parmak izleri, bileşiklerin moleküler özelliklerini analiz etmek ve değerlendirmek için kullanılmıştır. ChEMBL veri tabanından histon deasetilaz için 4695 adet IC50 molekülü ve 1012 adet İnhibisyon molekülü getirilmiştir. İlaç olabilecek moleküller içerisinde bu moleküler parmak izlerinin varlığını tespit edebilmek amacıyla binary matriks oluşturulmuştur. Modellerin değerlendirilmesi için oluşturulan matriks üzerinden R^2 , Düzeltilmiş R^2 (Adjusted R^2) ve Kök Ortalama Kare Hata (Root Mean Squared Error-RMSE) metrikleri hesaplanmıştır. R^2 , bağımsız değişkenlerin bağımlı değişkendeki

değişikliği açıklama gücünü temsil etmektedir. Ancak, R^2 tek başına kullanıldığında modelin karmaşıklığından kaynaklanan yanıltıcı sonuçlar verebilmektedir. Düzeltilmiş R^2 , R^2 değerini düzeltmek için kullanılmaktadır. Bu değer, modelin açıklama gücünü, bağımsız değişken sayısı ve örneklem büyüklüğü gibi faktörlerin etkisinden arındırmaktadır. Düzeltilmiş R^2 , daha fazla bağımsız değişkenin modele eklenmesi durumunda artış veya azalışın ne kadarının modelin performansını gerçekten iyileştirdiğini göstermektedir. RMSE, tahminlerin gerçek değerlerden ne kadar uzak olduğunu ölçmektedir ve bu hataları kare olarak hesaplamaktadır. Aynı zamanda hataların büyüklüğünü ve yayılmasını yansıtmaktadır. Aykırı değerlerin büyük bir etkisi vardır ve daha büyük hataların ağırlıklı olarak dikkate alınmasını sağlamaktadır. Python ile hesaplanan modellerin performans metriklerinin karşılaştırması IC50 için Çizelge 4.1 'de verilmiştir. Buna göre en yüksek Düzeltilmiş R-Kare ve R-Kare değerine sahip olan modeller DecisionTreeRegressor, ExtraTreeRegressor ve GaussianProcessRegressor'dır. Bu modeller % 85 oranında bağımlı değişkenin varyansını açıklamışlardır. En düşük Düzeltilmiş R-Kare ve R-Kare değerine sahip olan modeller ise PassiveAggressiveRegressor'dır. Bu model %-80 oranında bağımlı değişkenin varyansını açıklamıştır. En düşük RMSE değerine sahip olan modeller de DecisionTreeRegressor, ExtraTreeRegressor ve GaussianProcessRegressor'dır. Bu modeller 0.42 oranında ortalama hata yapmışlardır. En yüksek RMSE değerine sahip olan model ise PassiveAggressiveRegressor'dır. Bu model 1.47 oranında ortalama hata yapmıştır. En kısa sürede eğitilen model Ridge'dir. Bu model 0.06 saniyede eğitilmiştir. En uzun sürede eğitilen ve test edilen model ise QuantileRegressor'dır. Bu model 2205.80 saniyede eğitilmiştir.

Çizelge 4.1. IC50 için Regresyon performans metriklerinin karşılaştırılması

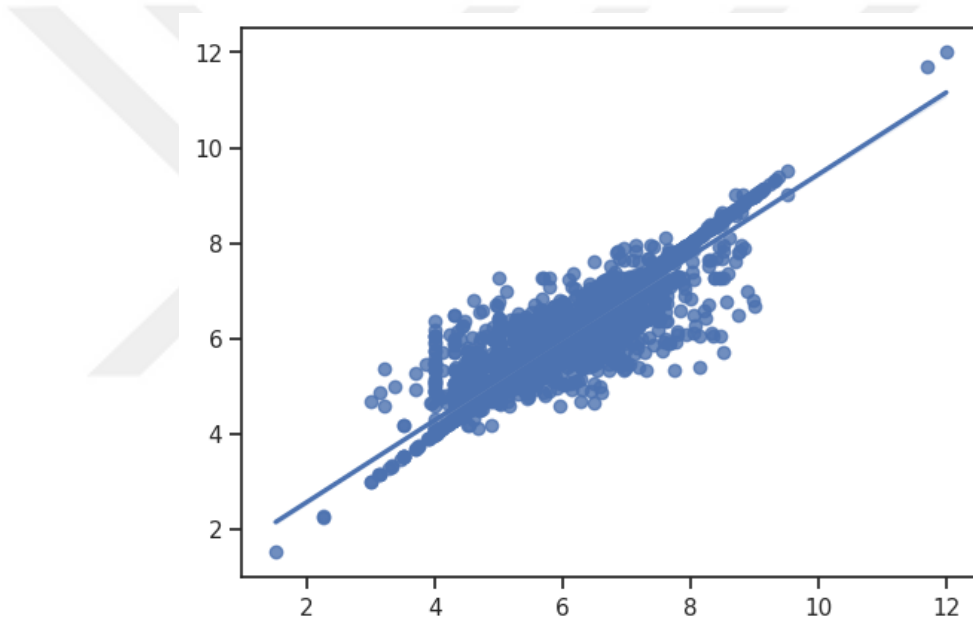
Model	Adjusted R-Squared	R-Squared	RMSE	Time Taken
DecisionTreeRegressor	0.85	0.85	0.42	0.20
ExtraTreeRegressor	0.85	0.85	0.42	0.18
ExtraTreesRegressor	0.85	0.85	0.42	12.69
GaussianProcessRegressor	0.85	0.85	0.42	18.48
RandomForestRegressor	0.78	0.79	0.51	4.13
MLPRegressor	0.75	0.76	0.54	9.33
BaggingRegressor	0.75	0.76	0.54	1.02
XGBRegressor	0.70	0.71	0.59	2.99
HistGradientBoostingRegressor	0.48	0.50	0.78	7.65
LGBMRegressor	0.48	0.49	0.78	0.26
SVR	0.47	0.49	0.79	4.76
KNeighborsRegressor	0.46	0.48	0.79	0.39
NuSVR	0.46	0.48	0.79	2.78
GradientBoostingRegressor	0.22	0.24	0.96	2.23
TransformedTargetRegressor	0.11	0.13	1.02	0.13
LinearRegression	0.11	0.13	1.02	0.13
Ridge	0.11	0.13	1.02	0.06
RidgeCV	0.11	0.13	1.02	0.14
Lars	0.10	0.13	1.03	0.08
LassoLarsIC	0.10	0.12	1.03	0.13
PoissonRegressor	0.10	0.12	1.03	0.34
BayesianRidge	0.10	0.12	1.03	0.30
HuberRegressor	0.10	0.12	1.03	0.54
SGDRegressor	0.08	0.11	1.04	0.13
LinearSVR	0.08	0.10	1.04	0.79
TweedieRegressor	0.06	0.09	1.05	0.27
GammaRegressor	0.06	0.09	1.05	0.31
AdaBoostRegressor	0.04	0.06	1.06	0.99
OrthogonalMatchingPursuit	0.04	0.06	1.06	0.04
OrthogonalMatchingPursuitCV	-0.01	0.01	1.09	0.08
LarsCV	-0.03	0.00	1.10	0.25
LassoLarsCV	-0.03	0.00	1.10	0.27
LassoLars	-0.03	0.00	1.10	0.04
LassoCV	-0.03	0.00	1.10	1.87
Lasso	-0.03	0.00	1.10	0.07
ElasticNet	-0.03	0.00	1.10	0.06
QuantileRegressor	-0.03	-0.00	1.10	2205.80
PassiveAggressiveRegressor	-0.85	-0.80	1.47	0.07

Inhibition için oluşturulan regresyon performans metriklerinin karşılaştırması Çizelge 4.2’de verilmiştir. Buna göre en yüksek Düzeltilmiş R-Kare ve R-Kare değerine sahip olan modeller GaussianProcessRegressor, DecisionTreeRegressor ve ExtraTreeRegressor’dır. Bu modeller % 79 oranında bağımlı değişkenin varyansını açıklamışlardır. En düşük Düzeltilmiş R-Kare ve R-Kare değerine sahip olan modeller ise PassiveAggressiveRegressor’dır. Bu model % -14 oranında bağımlı değişkenin varyansını açıklamıştır. En düşük RMSE değerine sahip olan modeller de DecisionTreeRegressor ve ExtraTreeRegressor’dır. Bu modeller 0.17 oranında ortalama hata yapmışlardır. En yüksek RMSE değerine sahip olan model ise PassiveAggressiveRegressor’dır. Bu model 0.43 oranında ortalama hata yapmıştır. En kısa sürede eğitilen model Ridge’dır. Bu model 0.03 saniyede eğitilmiştir. En uzun sürede eğitilen ve test edilen model ise QuantileRegressor’dır. Bu model 26.54 saniyede eğitilmiştir.

Çizelge 4.2. İnhibisyon için Regresyon performans metriklerinin karşılaştırılması

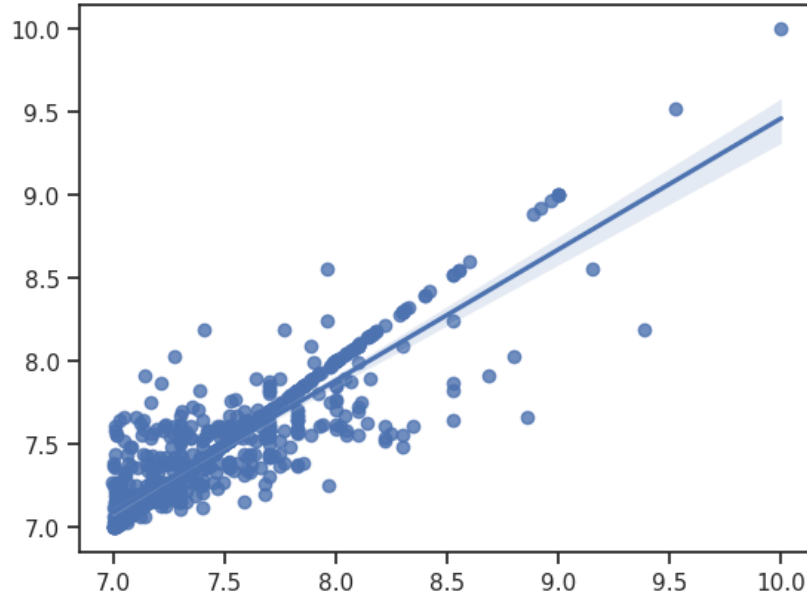
Model	Adjusted R-Squared	R-Squared	RMSE	Time Taken
DecisionTreeRegressor	0.79	0.82	0.17	0.09
ExtraTreeRegressor	0.79	0.82	0.17	0.11
ExtraTreesRegressor	0.79	0.82	0.17	3.68
XGBRegressor	0.74	0.77	0.19	1.40
RandomForestRegressor	0.69	0.72	0.21	1.23
BaggingRegressor	0.65	0.69	0.22	0.14
MLPRegressor	0.56	0.61	0.25	4.30
HistGradientBoostingRegressor	0.49	0.55	0.27	3.20
LGBMRegressor	0.48	0.54	0.27	0.08
SVR	0.45	0.51	0.28	0.28
NuSVR	0.44	0.50	0.28	0.32
GradientBoostingRegressor	0.32	0.39	0.31	0.55
KNeighborsRegressor	0.27	0.35	0.32	0.22
TransformedTargetRegressor	0.18	0.26	0.34	0.05
LinearRegression	0.18	0.26	0.34	0.09
Ridge	0.18	0.26	0.34	0.03
RidgeCV	0.17	0.26	0.34	0.08
LassoLarsIC	0.15	0.24	0.35	0.21
SGDRegressor	0.14	0.23	0.35	0.07
PoissonRegressor	0.13	0.22	0.35	0.54
HuberRegressor	0.13	0.22	0.35	0.16
BayesianRidge	0.09	0.19	0.36	0.05
Lars	0.05	0.15	0.37	0.10
TweedieRegressor	0.05	0.15	0.37	0.29
LinearSVR	0.05	0.15	0.37	0.48
GammaRegressor	0.05	0.15	0.37	0.31
OrthogonalMatchingPursuit	-0.01	0.10	0.38	0.03
OrthogonalMatchingPursuitCV	-0.10	0.02	0.40	0.06
LassoLarsCV	-0.12	0.00	0.40	0.37
LassoLars	-0.12	0.00	0.40	0.06
LassoCV	-0.12	0.00	0.40	1.33
Lasso	-0.12	0.00	0.40	0.03
LarsCV	-0.12	0.00	0.40	0.24
ElasticNetCV	-0.12	0.00	0.40	1.63
ElasticNet	-0.12	0.00	0.40	0.04
DummyRegressor	-0.12	0.00	0.40	0.04
AdaBoostRegressor	-0.15	-0.02	0.41	0.13
QuantileRegressor	-0.23	-0.10	0.42	26.54
PassiveAggressiveRegressor	-0.27	-0.14	0.43	0.05

DecisionTreeRegressor modelinin seçilmesiyle IC50 molekülü için Şekil 4.13'deki gibi, Inhibition molekülü için Şekil 4.14'daki gibi bir eğitim seti regresyon görseli oluşturulmuştur. Şekil 4.13'deki grafiğin x-ekseni ve y-ekseni 2 ile 12 arasında değişmektedir. Veri noktaları mavi daireler olarak gösterilmiştir. Lineer regresyon çizgisi, veri noktalarının arasından geçen en iyi uyumlu çizgiyi ifade etmektedir. Veri noktaları lineer regresyon çizgisi ile pozitif bir ilişki göstermiştir. Yani, x-eksenindeki değer arttıkça, y-eksenindeki değer de artmıştır. Ayrıca veri noktaları lineer regresyon çizgisine oldukça yakındır ve bu da lineer regresyon modelinin veriyi iyi açıkladığını ifade etmektedir.



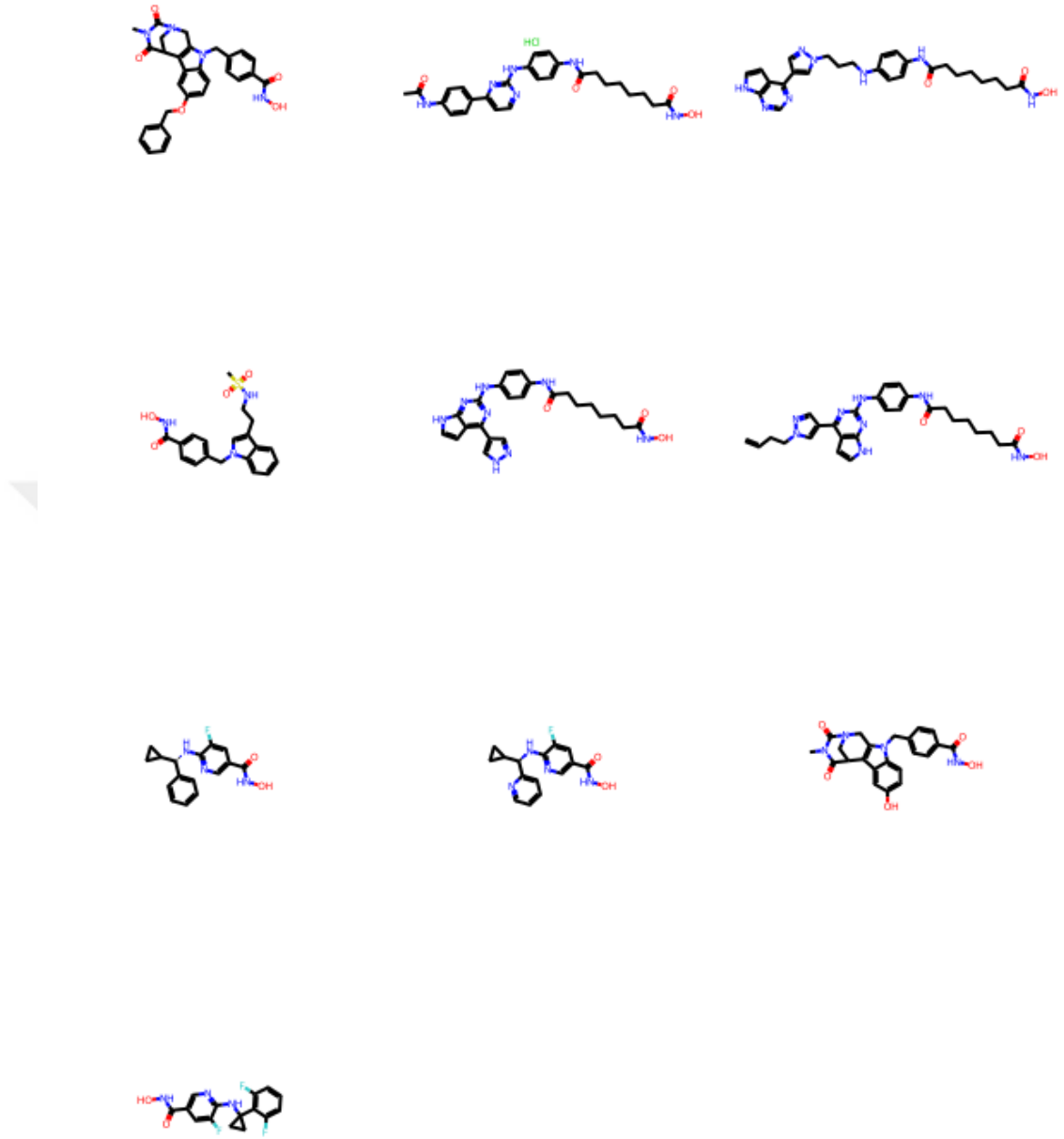
Şekil 4.13. DecisionTreeRegressor modelinin IC50 için regresyon grafiği

Şekil 4.14'deki grafiğin x-ekseni ve y-ekseni 7.0 ile 10.0 arasında değişmektedir. Veri noktaları regresyon çizgisi ile pozitif bir ilişki içerisindedir. Yani, x-eksenindeki değer arttıkça, y-eksenindeki değer de artmıştır. Veri noktaları regresyon çizgisine yakındır, bu da regresyon modelinin veriyi iyi açıkladığını ifade etmektedir.



Şekil 4.14. DecisionTreeRegressor modelinin Inhibition için regresyon grafiği

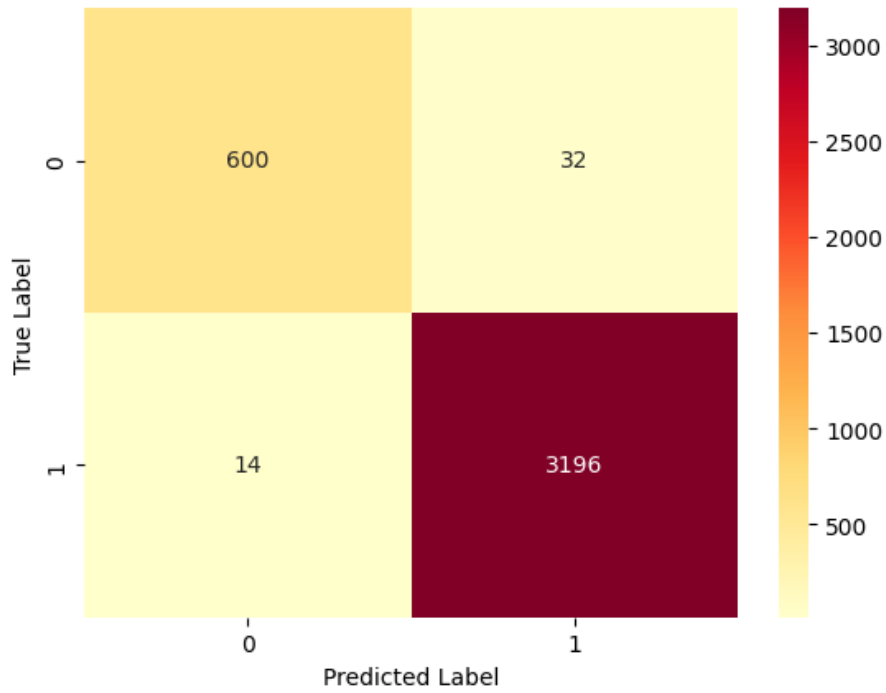
Smiles sütunundaki değerler, `rdkit.Chem` ve `rdkit.Chem.Draw` modülleri ve `Chem.MolFromSmiles` fonksiyonu kullanılarak molekül objelerine dönüştürülmüştür. `MolsToGridImage` fonksiyonu ile ilk 10 molekül yapısı görselleştirilmiş ve Şekil 4.15’de verilmiştir. Daha sonra moleküllerin hesaplanan morgan fingerprints değerleri ile label sütunundaki aktiflik durumuna göre 1 ve 0 şeklinde etiketlenen değerler birleştirilerek elde edilen veri seti, eğitim ve test kümelerine bölünmüştür. `RandomForestClassifier` ile eğitim seti üzerinde bir model oluşturularak test verileri ile modelin tahminleri hesaplanmıştır. Sınıflandırma modelinin performansı ROC eğrisi ile hesaplanmış ve 0.9725 değeri ile yüksek bir başarı elde edilmiştir.



Şekil 4.15. Smiles molekül yapılarının görseli

Veri集中的 bileşikler "active" ve "inactive" olarak etiketlenmiştir. Daha sonra, bu bileşikler etiket değerlerine göre sınıflandırılmıştır. Smiles formatındaki kimyasal yapılar, kimyasal parmak izi olarak bilinen Morgan Fingerprints

kullanılarak temsil edilmiştir. Kimyasal parmak izi, kimyasal bileşiklerin özelliklerini temsil etmek için kullanılan sayısal bir vektördür. Bu sayede, makine öğrenimi algoritmaları tarafından işlenebilecek bir forma dönüştürülmüşlerdir. Modelin gerçek ve tahmin edilen sınıflarını değerlendirmek için pycm kütüphanesinden ConfusionMatrix sınıfı kullanılarak gerçek etiketlerle tahmin edilen etiketler arasındaki karmaşıklık matrisi hesaplanmış ve ısı haritası (heatmap) ile görselleştirilerek Şekil 4.16’de verilmiştir. Buna göre gerçek pozitif sınıfı doğru bir şekilde pozitif olarak tahmin edilen örnek sayısı 3196, gerçek negatif sınıfı yanlış bir şekilde pozitif olarak tahmin edilen örnek sayısı 32, gerçek pozitif sınıfı yanlış bir şekilde negatif olarak tahmin edilen örnek sayısı 14 ve gerçek negatif sınıfı doğru bir şekilde negatif olarak tahmin edilen örnek sayısı 600 olarak elde edilmiştir.



Şekil 4.16. ConfusionMatrix (Karmaşıklık Matrisi)

Yapılan sınıflandırma sonucunda modelin performansını daha net görebilmek için bazı metrikler ile değerlendirme yapılmıştır. Ek-1 ve Ek-2’de bu metriklerin değerleri verilmiştir. Buna göre ACC Macro, tüm sınıfların doğru sınıflandırma oranlarının ortalamasıdır ve 0.98803 olarak hesaplanmıştır. ARI - Ayarlanmış

Rand Endeksi, kümeleme algoritmasının gerçek sınıflandırmaya ne kadar yakın olduğunu gösteren bir ölçüttür. Değer ne kadar yüksekse, modelin daha iyi bir kümeleme performansı olduğunu göstermektedir ve 0.94022 olarak hesaplanmıştır. AUNP, pozitif sınıfın altında kalan alanın normalleştirilmiş versiyonudur. Değer ne kadar yüksekse, modelin pozitif sınıfı iyi bir şekilde sınıflandırdığını göstermektedir ve bu değer 0.9725 olarak hesaplanmıştır. AUNU, negatif sınıfın altında kalan alanın normalleştirilmiş versiyonudur. Değer ne kadar yüksekse, modelin negatif sınıfı iyi bir şekilde sınıflandırdığını göstermektedir ve 0.9725 olarak hesaplanmıştır. F1 Macro - Makro F1 skoru, tüm sınıfların F1 skorlarının ortalamasıdır. F1 skoru, hassasiyet (precision) ve duyarlılık (recall) metriklerinin harmonik ortalaması olarak ifade edilmektedir ve 0.97797 olarak hesaplanmıştır. Mikro F1 skoru, tüm sınıfların TP, FP ve FN değerlerinin toplamı üzerinden hesaplanan bir F1 skorudur ve 0.98803 olarak hesaplanmıştır. Hamming Loss - Hamming kaybı, doğru sınıflandırılmayan etiketlerin toplam sayısının, toplam etiket sayısına oranıdır ve 0.01197 olarak hesaplanmıştır. Kappa, modelin gerçek sınıfları rastgele tahmin etme olasılığına karşı düzeltilmiş bir sınıflandırma doğruluğu ölçüsüdür. Değer 1'e ne kadar yakınsa, modelin daha iyi bir sınıflandırma performansı olduğunu göstermektedir ve 0.95594 olarak bulunmuştur. Matthews korelasyonu, doğru ve yanlış pozitif/negatif sınıflandırmalar arasındaki dengeli bir ölçüdür. Değer ne kadar yüksekse, modelin daha iyi bir sınıflandırma performansı olduğunu göstermektedir ve 0.95608 olarak elde edilmiştir. A1 (Landis & Koch), Landis ve Koch tarafından önerilen ölçüte göre, modelin sınıfları neredeyse mükemmel bir şekilde sınıflandırdığını göstermektedir.

Ek 2'de elde edilen sonuçlara göre, her iki sınıf (0 ve 1) için de yüksek performans elde edildiği görülmektedir. ACC (Accuracy) değeri hem sınıf 0 hem de sınıf 1 için yaklaşık olarak % 98.8 olarak hesaplanmıştır, yani doğru sınıflandırılan örneklerin oranı oldukça yüksektir. AUC (Area under the ROC curve) ve AUPR (Area under the PR curve) değerleri de her iki sınıf için yaklaşık olarak 0.97 olarak hesaplanmıştır. Bu da modelin sınıflandırma yeteneğinin yüksek olduğunu göstermektedir. Precision (kesinlik) ve recall (hatırlama) ölçümlerinin harmonik ortalaması olan F1 skoru, her iki sınıf için yüksek değerlerde bulunmaktadır. Sınıf

0 için F1 skoru 0.96, sınıf 1 için ise 0.99 olarak hesaplanmıştır. Yanlış pozitif oranı (FPR) ve yanlış negatif oranı (FNR) da her iki sınıf için düşük değerlerde bulunmaktadır. Bu da modelin hem yanlış pozitif sınıflandırmaları hem de yanlış negatif sınıflandırmaları minimize ettiğini göstermektedir. Genel olarak, verilen metrikler sonucunda modelin yüksek doğruluk, hassasiyet ve özgünlük değerlerine sahip olduğu görülmektedir.

Diğer modellerin sınıflandırma performanslarını görmek için hazırlanan Çizelge 4.3'de accuracy, balanced accuracy, ROC-AUC, F1 Score ve Time Taken değerleri karşılaştırılmıştır. Bu metrikler, bir sınıflandırma modelinin ne kadar iyi çalıştığını değerlendirmek için kullanılan ölçülerdir. Doğruluk (Accuracy), modelin doğru tahmin ettiği örneklerin toplam örnek sayısına oranıdır. Doğruluk değeri ne kadar yüksekse, model o kadar iyidir. Dengeli Doğruluk (Balanced Accuracy), modelin her bir sınıf için doğru tahmin ettiği örneklerin oranının ortalamasıdır. Dengeli doğruluk değeri ne kadar yüksekse, model o kadar dengelidir. ROC AUC, modelin gerçek pozitif oranı ile yanlış pozitif oranı arasındaki ilişkiyi gösteren bir eğrinin altında kalan alanın büyüklüğüdür. ROC AUC değeri ne kadar yüksekse, model o kadar iyidir. F1 Skoru, modelin hassasiyeti ve duyarlılığı arasındaki dengeyi gösteren bir skordur. Hassasiyet, modelin pozitif olarak tahmin ettiği örneklerden kaç tanesinin gerçekten pozitif olduğunu gösterir. Duyarlılık, modelin gerçekten pozitif olan örneklerden kaç tanesini doğru tahmin ettiğini gösterir. F1 skoru ne kadar yüksekse, model o kadar dengelidir. Harcanan Süre (Time Taken), modelin eğitilmesi ve test edilmesi için harcadığı süredir. Süre ne kadar düşükse, model o kadar hızlıdır. Performans metriklerinin karşılaştırılmasına göre, en yüksek doğruluk değerine sahip olan modeller LabelPropagation, ExtraTreeClassifier ve RandomForestClassifier'dır. Bu modeller % 100 oranında doğru tahmin yapmışlardır. En düşük doğruluk değerine sahip olan model ise GaussianNB'dir. Bu model % 71 oranında doğru tahmin yapmıştır. En yüksek dengeli doğruluk değerine sahip olan modeller de LabelPropagation, ExtraTreeClassifier ve RandomForestClassifier'dır. Bu modeller % 100 oranında dengeli doğruluk değerine sahiptirler. En düşük dengeli doğruluk değerine sahip olan model ise GaussianNB'dir. Bu model % 70 oranında dengeli doğruluk değerine sahiptir. En

yüksek ROC AUC değerine sahip olan modeller de LabelPropagation, ExtraTreeClassifier ve RandomForestClassifier'dır. Bu modeller % 100 oranında ROC AUC değerine sahiptirler. En düşük ROC AUC değerine sahip olan model ise GaussianNB'dir. Bu model % 77 oranında ROC AUC değerine sahiptir. En yüksek F1 skoruna sahip olan modeller de LabelPropagation, ExtraTreeClassifier ve RandomForestClassifier'dır. Bu modeller % 100 oranında F1 skoruna sahiptirler. En düşük F1 skoruna sahip olan model ise GaussianNB'dir. Bu model % 71 oranında F1 skoruna sahiptir. En kısa sürede eğitilen ve test edilen model LogisticRegression'dur. Bu model 0.02 saniyede eğitilmiş ve test edilmiştir. En uzun sürede eğitilen ve test edilen model ise MLPClassifier'dır. Bu model 2.29 saniyede eğitilmiş ve test edilmiştir.

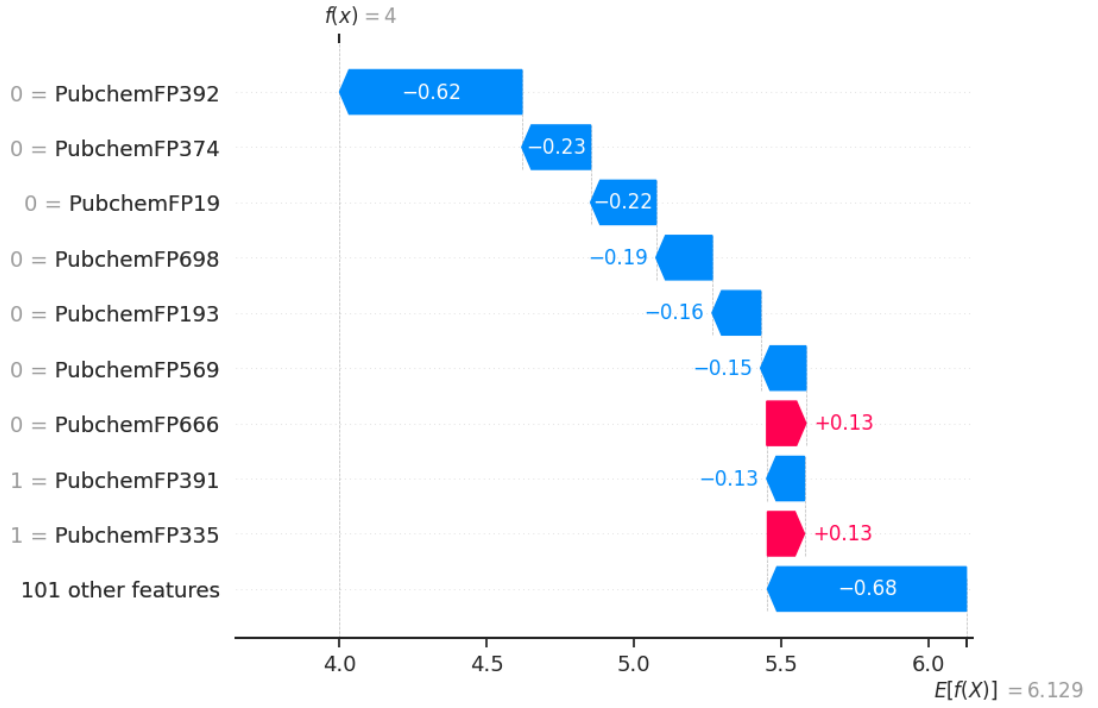
Çizelge 4.3. Sınıflandırma modellerinin performans metrikleri

Model	Accuracy	Balanced Accuracy	ROC AUC	F1 Score	Time Taken
LabelPropagation	1.00	1.00	1.00	1.00	30.86
DecisionTreeClassifier	1.00	1.00	1.00	1.00	9.87
ExtraTreeClassifier	1.00	1.00	1.00	1.00	1.53
ExtraTreesClassifier	1.00	1.00	1.00	1.00	16.50
LabelSpreading	1.00	1.00	1.00	1.00	35.40
RandomForestClassifier	1.00	1.00	1.00	1.00	10.25
BaggingClassifier	1.00	1.00	1.00	1.00	50.34
LogisticRegression	0.99	0.99	0.99	0.99	4.33
LGBMClassifier	0.99	0.98	0.98	0.99	4.30
LinearSVC	0.99	0.98	0.98	0.99	26.22
Perceptron	0.98	0.98	0.98	0.98	2.26
PassiveAggressiveClassifier	0.99	0.98	0.98	0.99	3.89
XGBClassifier	0.99	0.97	0.97	0.99	82.86
LinearDiscriminantAnalysis	0.98	0.97	0.97	0.98	22.70
RidgeClassifier	0.98	0.97	0.97	0.98	3.92
RidgeClassifierCV	0.98	0.97	0.97	0.98	17.91
SGDClassifier	0.98	0.96	0.96	0.98	9.55
SVC	0.98	0.96	0.96	0.98	122.85
KNeighborsClassifier	0.97	0.95	0.95	0.97	18.96
CalibratedClassifierCV	0.98	0.94	0.94	0.98	107.79
QuadraticDiscriminantAnalysis	0.95	0.88	0.88	0.95	29.52
BernoulliNB	0.91	0.86	0.86	0.91	2.19
NearestCentroid	0.91	0.84	0.84	0.91	1.51
AdaBoostClassifier	0.92	0.84	0.84	0.92	37.52
GaussianNB	0.71	0.82	0.82	0.74	2.27
DummyClassifier	0.82	0.50	0.50	0.74	1.34

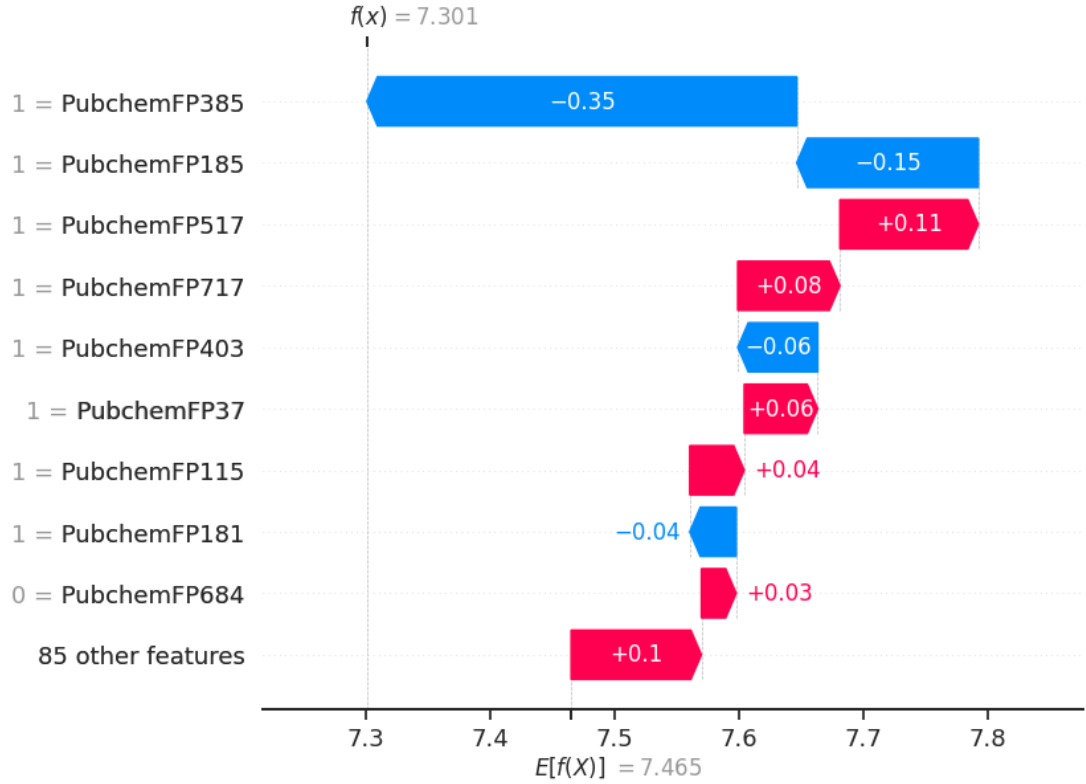
5. TARTIŞMA VE SONUÇLAR

Bu araştırma, çocukluk çağı akut lösemisi için potansiyel adayları belirlemek üzere histon deasetilaz genlerinin sanal taramasına odaklanmıştır. Çalışmada moleküler parmak izleri kullanılmış ve bu parmak izleri ile ilişkili histon deasetilaz genlerinin keşfedilmesinde IC50 ve İnhibisyon standart tiplerinin katkısı değerlendirilmiştir. Ayrıca, çocukluk çağı akut lösemisinin tedavisinde kullanılan ilaçların ilaç etkileşimlerini görselleştirmek için ağ grafikleri kullanılmıştır. Bu yaklaşımlar kullanılarak çocukluk çağı akut lösemisi ile ilgili potansiyel hedefler ve ilaç etkileşimleri hakkında bilgi edinme amaçlanmıştır. Moleküler parmak izleri belirlendikten sonra, moleküler özelliklerin önemini değerlendirmek için açıklanabilir yapay zekâ yaklaşımı kullanılmıştır. Aday bileşiklerin moleküler özellikleri üzerinde sanal tarama gerçekleştirilerek, Decision Tree Regressor ve Gradient Boosted Trees modellerinin en uygun seçenekler olduğu belirlenmiştir. Moleküler parmak izi verilerini içeren ikili matris, bu modelleri uygulamak için kullanılmış ve moleküler özelliklerin kapsamlı bir şekilde incelenmesini kolaylaştırmıştır. Açıklanabilir yapay zekâ tekniklerinin uygulanabilmesi için her bir moleküler parmak izinin modeller üzerindeki etkisi Shap değerleri üzerinden hesaplanmış ve gözlemlenmiştir. Moleküler özelliklerin modeller üzerindeki etkilerini görselleştirmek için çeşitli grafik türleri kullanılmış ve kapsamlı bir analize olanak sağlanmıştır.

Waterfall grafikleri, bir başlangıç değeri (genellikle bir temel tahmin değeri veya ortalama değer) üzerindeki değişimleri adım adım göstermektedir. Her adımda, bir özellik katkısı toplam değere eklenmekte veya çıkarılmaktadır. Bu şekilde, özelliklerin toplam katkısı belirlenmektedir. Şekil 5.1, IC50 için tasarlanan grafiğin görsel temsilini sunarak her bir moleküler parmak izinin tahminler üzerindeki etkisinin analizini sağlamaktadır. Benzer şekilde, Şekil 5.2 inhibisyon için görselleştirmeyi sunarak her bir moleküler parmak izinin tahminler üzerindeki etkisinin değerlendirilmesine yardımcı olmaktadır.



Şekil 5.1. IC50 için SHAP değerlerine göre waterfall grafiği



Şekil 5.2. İnhibisyon için SHAP değerlerine göre waterfall grafiği

Waterfall grafiğinin alt kısmı, modelin beklenen çıktı değeri olan $E[f(X)]$ ile başlamaktadır. Her satırda, her bir özelliğin pozitif (kırmızı) veya negatif (mavi) katkısının, beklenen çıktı değerinden modelin çıktı değerine nasıl taşıdığı gösterilmektedir. Sonuçlara göre, IC50 için beklenen model çıktı değeri 6.129 iken mevcut modelin çıktı değeri 4, inhibisyon için beklenen model çıktı değeri 7.465 iken mevcut modelin çıktı değeri 7.301 olarak bulunmuştur. Özellik adlarının önündeki gri sayılar, bu örnekte her bir özelliğin değerini temsil etmektedir. Özelliklerin 1 değerine sahip olduğu durumlar var olarak, 0 değerine sahip olduğu durumlar yok olarak ifade edilmektedir. Şekil 5.1’de verilen ic50 için waterfall grafiğinde modele etki eden moleküllerin etkisi, modele katkısı ve bit alt yapısı Çizelge 5.1’de verilmiştir.

Çizelge 5.1. FP moleküllerinin modele etkisi ve bit altyapısı(Şekil 5.1 için)

FP	Modele Katkısı	Bit Altyapısı
PubChemFP392	-0.62	$N(\sim C)(\sim C)(\sim H)$
PubChemFP374	-0.23	$C(\sim H)(\sim H)(\sim H)$
PubChemFP19	-0.22	$\geq 2 O$
PubChemFP698	-0.19	$O-C-C-C-C-C-C$
PubChemFP193	-0.16	≥ 3 saturated or aromatic carbon-only ring size 6
PubChemFP569	-0.15	$N-C-C-N$
PubChemFP666	+0.13	$C=C-C-O-C$
PubChemFP391	-0,13	$N(\sim C)(\sim C)(\sim C)$
PubChemFP335	+0,13	$C(\sim C)(\sim C)(\sim C)(\sim H)$

Şekil 5.1’de PubchemFP335 ve PubchemFP666’nın modele pozitif yönde etki ettiği görülmektedir. Her iki molekül de 5.5’den 5.63’e kadar 0.13 oranında model çıktısına pozitif katkı sağlamıştır. PubChem substructure fingerprint(FP) listesine göre, PubchemFP335, " $C(\sim C)(\sim C)(\sim C)(\sim H)$ " alt yapısını temsil etmektedir. Bu alt yapı, karbon atomunun üç farklı karbon atomu ve bir hidrojen

atomu ile bağlandığını ifade etmektedir. Bu alt yapı, karbon atomunun çevresinde üç adet karbon atomu ve bir hidrojen atomu bulunduğunu göstermektedir. IC50 değerleri için taranan histon deasetilaz gen inhibitörlerinin analizi, aromatik bağları çevreleyen atomların nispeten daha düşük bir öneme sahip olduğunu ortaya koymuştur. Bu aromatik bağlardaki komşu atomların varlığının, +0,13'lük bir etki oranı ile pozitif bir etkiye sahip olduğu bulunmuştur. Bu durum, aromatik bağların kendisinden ziyade komşu atomların spesifik özelliklerinin, moleküllerin histon deasetilaz genlerine karşı aktivitesinde önemli bir rol oynadığını göstermektedir. FP'ye göre, aromatik bağların sayısı dikkate alınmaksızın, diğer moleküllerle olan tekli ve çiftli bağlarına odaklanılmalıdır. Aromatik halkaların tekli ve çiftli bağları, histon deasetilaz genlerinin potansiyel ilaç özellikleri açısından önemli bir rol oynamaktadır. PubChemFP392, $(N(\sim C)(\sim C)(\sim H))$ 'yi temsil ederken, modelin tahminine -0.62 katkıda bulunmuştur. Bu durum çocukluk çağı akut lösemisi için bir ilaç adayı olarak pek uygun olmadığını göstermektedir. Diğer moleküller de farklı derecelerde katkıda bulunmuştur. Bazıları modelin tahminine pozitif katkıda bulunurken diğerleri negatif katkı vermiştir. PubChemFP19 ($\geq 2 O$), modelin tahminine -0.22 katkı verirken, PubChemFP666 ($C=C-C-O-C$) ise +0.13 katkı vermiştir. Bu durum, oksijen atomlarının varlığının modelin tahminini olumsuz etkilediğini, ancak belirli bir karbon-oksijen bağının varlığının ise tahmini olumlu yönde etkilediğini ifade etmektedir.

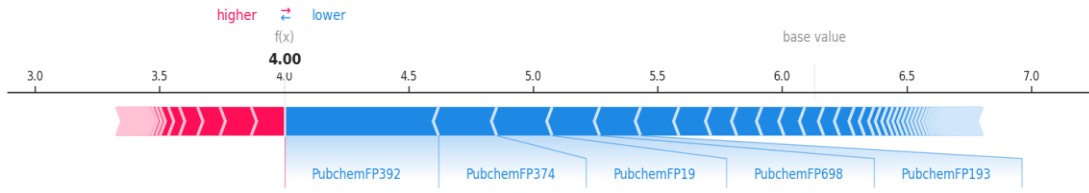
Şekil 5.2'de verilen inhibisyon için waterfall grafiğinde modele etki eden moleküllerin etkisi, modele katkısı ve bit alt yapısı Çizelge 5.2'de verilmiştir. PubchemFP385, 7.651'den 7.301'e kadar model çıktısına 0.35 oranında negatif katkı sağlamıştır. PubChemFP385 $(C(:C)(:C)(:C))$ ile bir karbon atomunun üç bağlantısı olan bir yapıyı temsil etmektedir. Modele -0.35 katkı verirken negatif bir katkı değeri olduğu için, bu molekülün varlığı ilaç potansiyelini azaltabilir veya hedef molekül üzerinde olumsuz bir etkiye sahip olabilir. PubChemFP185, (≥ 2 any ring size 6), en az iki adet 6 halkalı bir yapıyı temsil ederken modele -0.15 katkıda bulunmuştur. Yine negatif bir katkı değeri olduğu için, bu molekülün varlığı ilaç potansiyelini azaltabilir veya hedef molekül üzerinde olumsuz bir etkiye sahip olabilir. PubChemFP517, $(N-N-C-N)$ ile azot atomlarıyla bağlı bir

karbon atomunu ifade etmektedir ve modele +0.11 katkı vermiştir. Pozitif bir katkı değeri olduğu için, bu altyapının varlığı ilaç potansiyelini artırabilir veya hedef molekül üzerinde olumlu bir etki yaratabilmektedir. PubChemFP717, (Cc1ccc(Cl)cc1) ile bir klor atomu içeren bir benzen halkasını temsil etmekte ve modele +0.08 katkıda bulunmaktadır. Pozitif bir katkı değeri olduğu için, bu altyapının varlığı ilaç potansiyelini artırabilir veya hedef molekül üzerinde olumlu bir etki verebilmektedir. PubChemFP403, (N(:C)(:C)) ile bir azot atomunun iki bağlantısı olan bir yapıyı ifade ederken modele -0.06 katkıda bulunmuştur. Negatif bir katkı değeri olduğu için, bu bit altyapısının varlığı ilaç potansiyelini azaltabilir veya hedef molekül üzerinde olumsuz bir etkiye sahip olabilir. PubChemFP37, (>= 1 Cl) ile en az bir klor atomunu temsil etmekte ve modele +0.06 katkıda bulunmaktadır. Pozitif bir katkı değeri olduğu için, bu molekülün varlığı ilaç potansiyelini artırabilir veya hedef molekül üzerinde olumlu bir etki verebilir. PubChemFP115, (>= 1 any ring size 3) ile en az bir adet 3 halkalı bir yapıyı temsil etmekte ve modele +0.04 katkı vermektedir. Pozitif bir katkı değeri olduğu için, bu molekülün varlığı ilaç potansiyelini artırabilir veya hedef molekül üzerinde olumlu bir etkiye sahip olabilir. PubChemFP181, (>= 1 saturated or aromatic heteroatom-containing ring size 6) ile en az bir doymuş veya aromatik bir heteroatom içeren 6 halkalı bir yapıyı temsil etmekte ve modele -0.04 katkıda bulunmaktadır. Negatif bir katkı değeri olduğu için, bu molekülün varlığı ilaç potansiyelini azaltabilir veya hedef molekül üzerinde olumsuz bir etkiye yaratabilir. PubChemFP684, (O=C-C-C-C-C) ile bir karbonil grubu (O=C) ile bağlı bir seri karbon atomunu temsil ederken modele +0.03 katkı vermiştir. Pozitif bir katkı değeri olduğu için, bu altyapının varlığı ilaç potansiyelini artırabilir veya hedef moleküle olumlu bir etki verebilir.

Çizelge 5.2. FP moleküllerinin modele etkisi ve bit alt yapısı (Şekil 5.2 için)

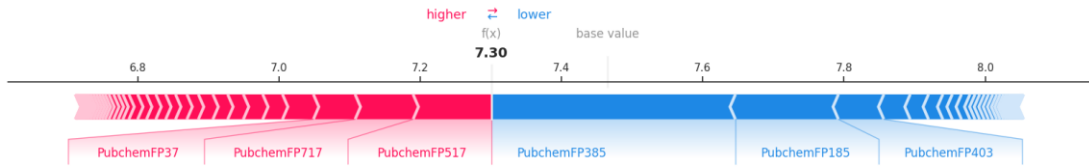
FP	Modele Katkısı	Bit Altyapısı
PubChemFP385	-0.35	C(:C)(:C)(:C)
PubChemFP185	-0.15	>= 2 any ring size 6
PubChemFP517	+0.11	N-N-C-N
PubChemFP717	+0.08	Cc1ccc(Cl)cc1
PubChemFP403	-0.06	N(:C)(:C)
PubChemFP37	+0.06	>= 1 Cl
PubChemFP115	+0.04	>= 1 any ring size 3
PubChemFP181	-0.04	>= 1 saturated or aromatic heteroatom- containing ring size 6
PubChemFP684	+0.03	O=C-C-C-C-C

Force grafiği, waterfall grafiğiyle birlikte kullanılan bir görselleştirme yöntemidir. Force grafiğinde, kırmızı oklar pozitif etkileri temsil ederken mavi oklar negatif etkileri göstermektedir. Okların boyutu, özelliğin etkisinin büyüklüğünü ifade etmektedir. Gri renkli "temel değer", modelin eğitim seti üzerindeki ortalama tahmin değerini belirtmektedir. "Çıkış değeri" ise modelin tahmini sonucudur. Force grafiği, tahmin için etkili bir özet sunarak, özelliklerin etkilerini anlamayı kolaylaştırmaktadır. Özellik değerleri grafiğin alt kısmında görüntülenmiştir. IC50 için tasarlanan force grafiği Şekil 5.3'de verilmiştir. Buna göre PubchemFP392, PubchemFP374, PubchemFP19, PubchemFP698 ve PubchemFP193'ün model çıktısına negatif yönde katkı verdiği görülmektedir.



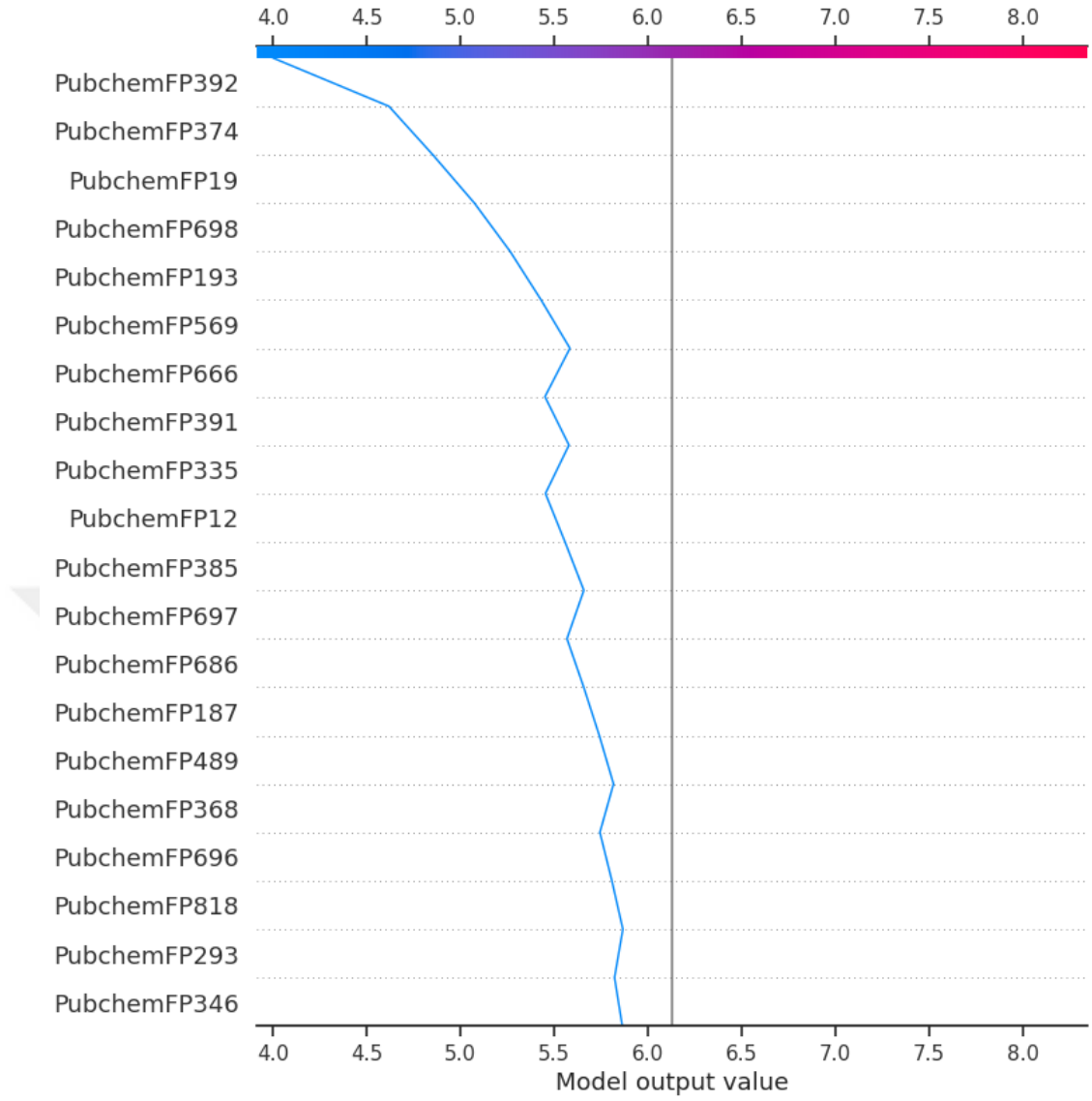
Şekil 5.3. IC50 için SHAP değerleri force grafiği

İnhibisyon molekülleri için oluşturulan force grafiği Şekil 5.4’de verilmiştir. PubchemFP37, PubchemFP517 ve PubchemFP717’nin model çıktısına pozitif yönde katkı sağlarken tahmin değerini yükselttiği, PubchemFP185, PubchemFP385 ve PubchemFP403’ün ise negatif yönde katkı sağlarken tahmin değerini düşürdüğü görülmektedir.



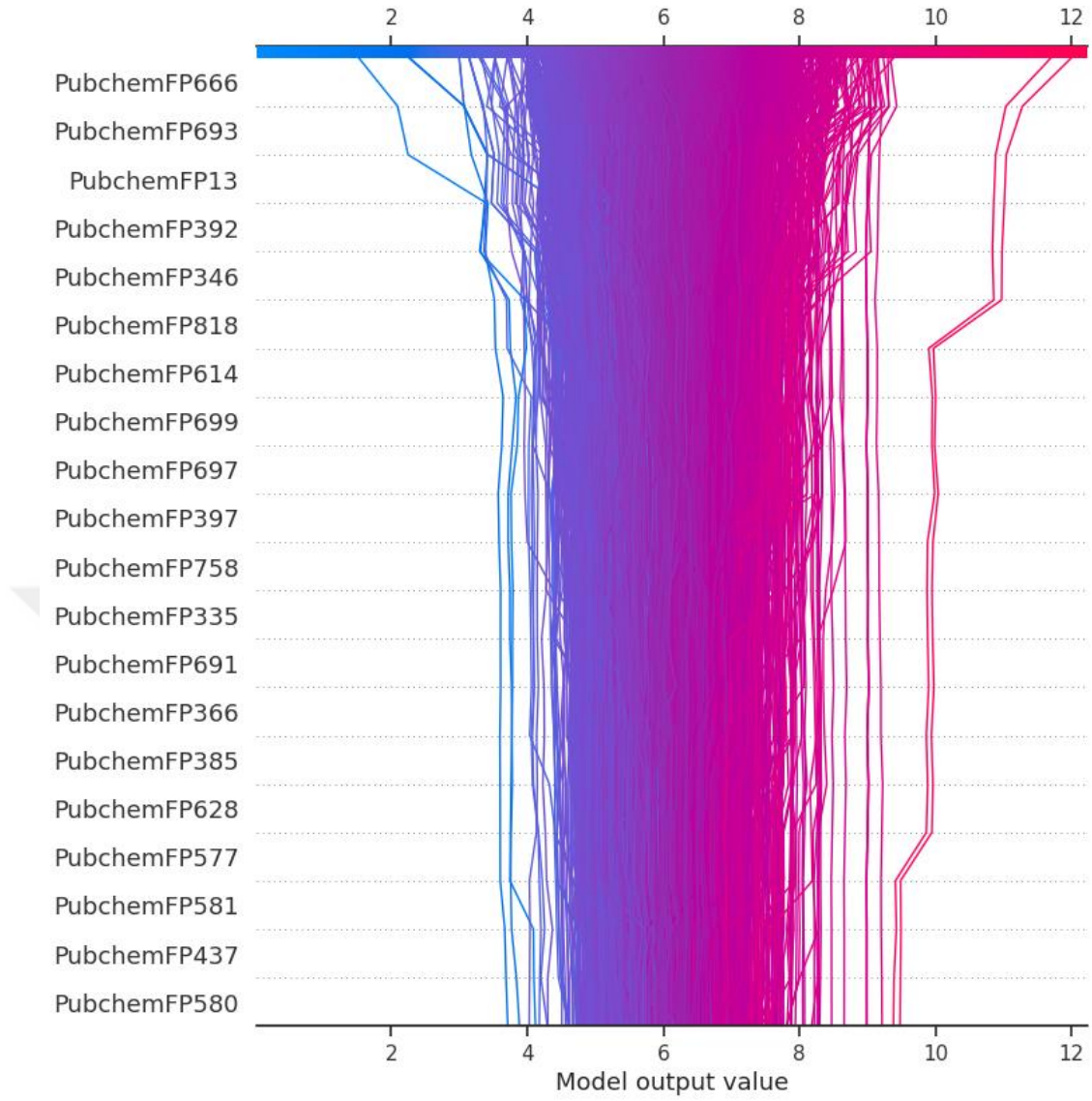
Şekil 5.4. İnhibisyon için SHPA değerleri force grafiği

IC50 için oluşturulan decision grafiklerinden ilki Şekil 5.5’de verilmiştir. Decision grafiği, force grafiğinin aksine, düz bir dikey çizgi aracılığıyla modelin temel değeri hakkında fikir vermektedir. Çizimdeki renkli çizgiler, özelliklerle ilişkili tahmin edilen değerleri temsil etmektedir. SHAP değerlerini (özellik etkileri) görselleştiren decision grafiği, temel değerden modelin nihai puanına giden yolu göstermektedir. Her bir özelliğin genel tahmine nasıl katkıda bulunduğuna dair kapsamlı bir anlayış sunmaktadır.



Şekil 5.5. IC50 için decision grafiği-1

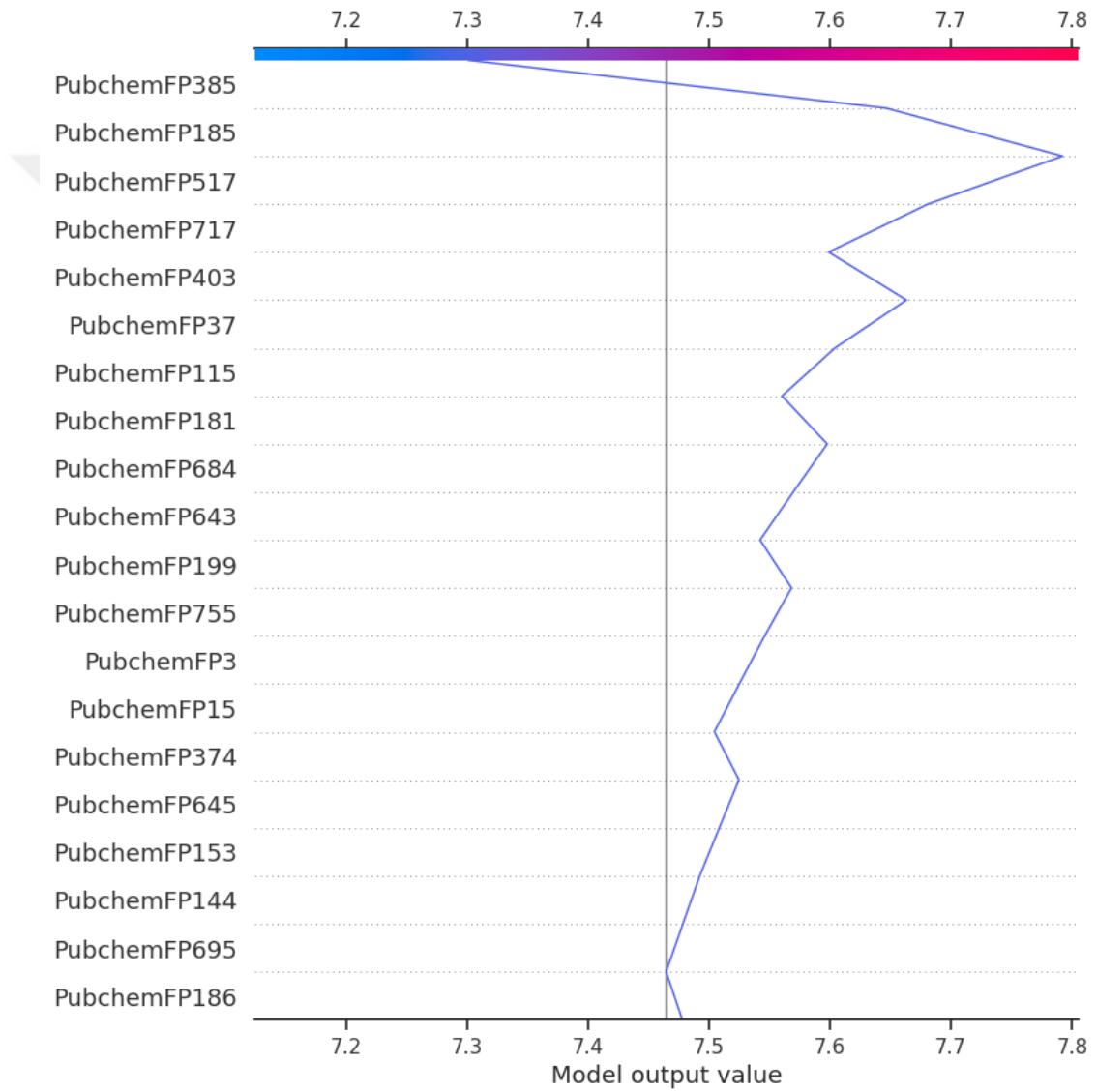
Decision ve force grafikleri birbirini desteklemektedir ve benzerlikler ifade etmektedir. Özelliklerin SHAP değerlerine göre her iki grafikte de pozitif veya negatif etkiler görülebilmektedir. Örneğin, PubchemFP392, PubchemFP374, PubchemFP19, PubchemFP698, PubchemFP193 ve PubchemFP568 molekülleri, model çıktısını negatif yönde etkilemiştir. Decision ve force grafiklerinde gözlemlenen değerler, diğer grafiksel gösterimlerle pozitif bir korelasyon göstermektedir. Bu grafikler modelin tahmin sürecine ilişkin kayda değer veriler sunarak ana etkilerin büyüklüğü ve yönünün daha iyi anlaşılmasını sağlamaktadır. IC50 için bir diğer decision görseli Şekil 5.6'da verilmiştir.



Şekil 5.6. IC50 için decision grafiği-2

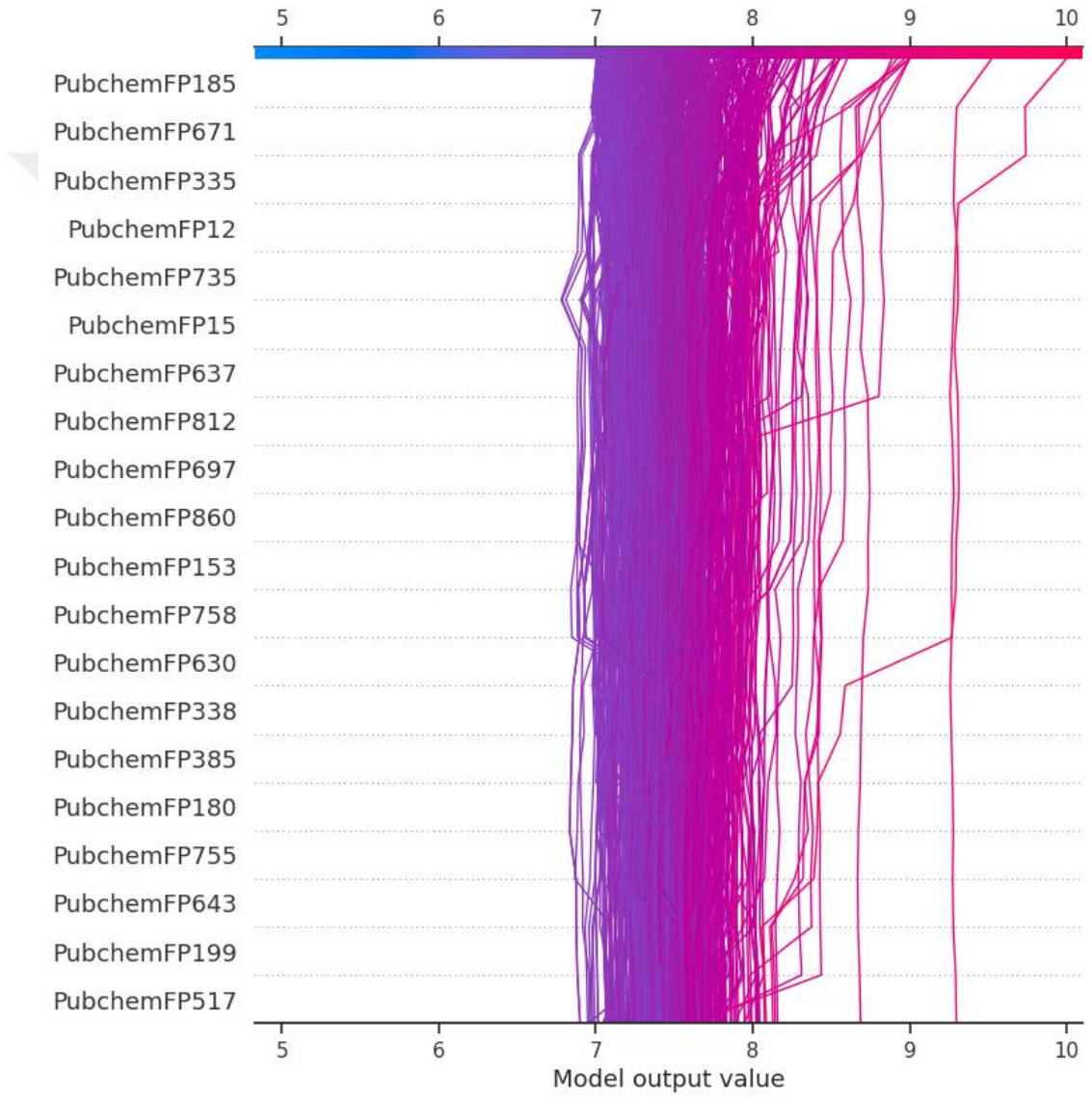
Şekil 5.6'da verilen farklı bir decision grafiği türü, aykırı özelliklerin belirlenmesine olanak sağlamakla beraber birçok örnekte hangi özelliklerin ne kadar etkili olduğunu açıklamaktadır. Bu grafik türü, veri setindeki aykırı değerleri tespit etmek için sıkça kullanılmaktadır ve özellikle veri boyutu büyük olduğunda daha da etkilidir. Görselde PubchemFP818 özelliğinin aykırı değerleri olduğu anlaşılmaktadır. Tutarlı özellik sırası ve x eksenini ölçeği ile birden fazla karar grafiği oluşturmak, doğrudan karşılaştırmaları basitleştirmekte ve yorumlama sürecini geliştirmektedir. Bu yaklaşım, çizilen verilerin daha kolay analiz edilmesini ve anlaşılmasını sağlamaktadır.

İnhibisyon için tasarlanan ilk decision görseli Şekil 5.7’de verilmiştir. Buna göre grafiğin düz dikey çizgisi, modelin temel değerini (7.446) göstermektedir. PubchemFP385 yani C(:C)(:C)(:C) özelliği model çıktısını yaklaşık 7.651’den 7.301’e kadar model çıktısına 0.35 oranında negatif yönde katkı sağlamıştır. PubchemFP517 ve PubchemFP717 ise model çıktısına pozitif yönde katkı verdiği görülmektedir.



Şekil 5.7. İnhibisyon için decision grafiği-1

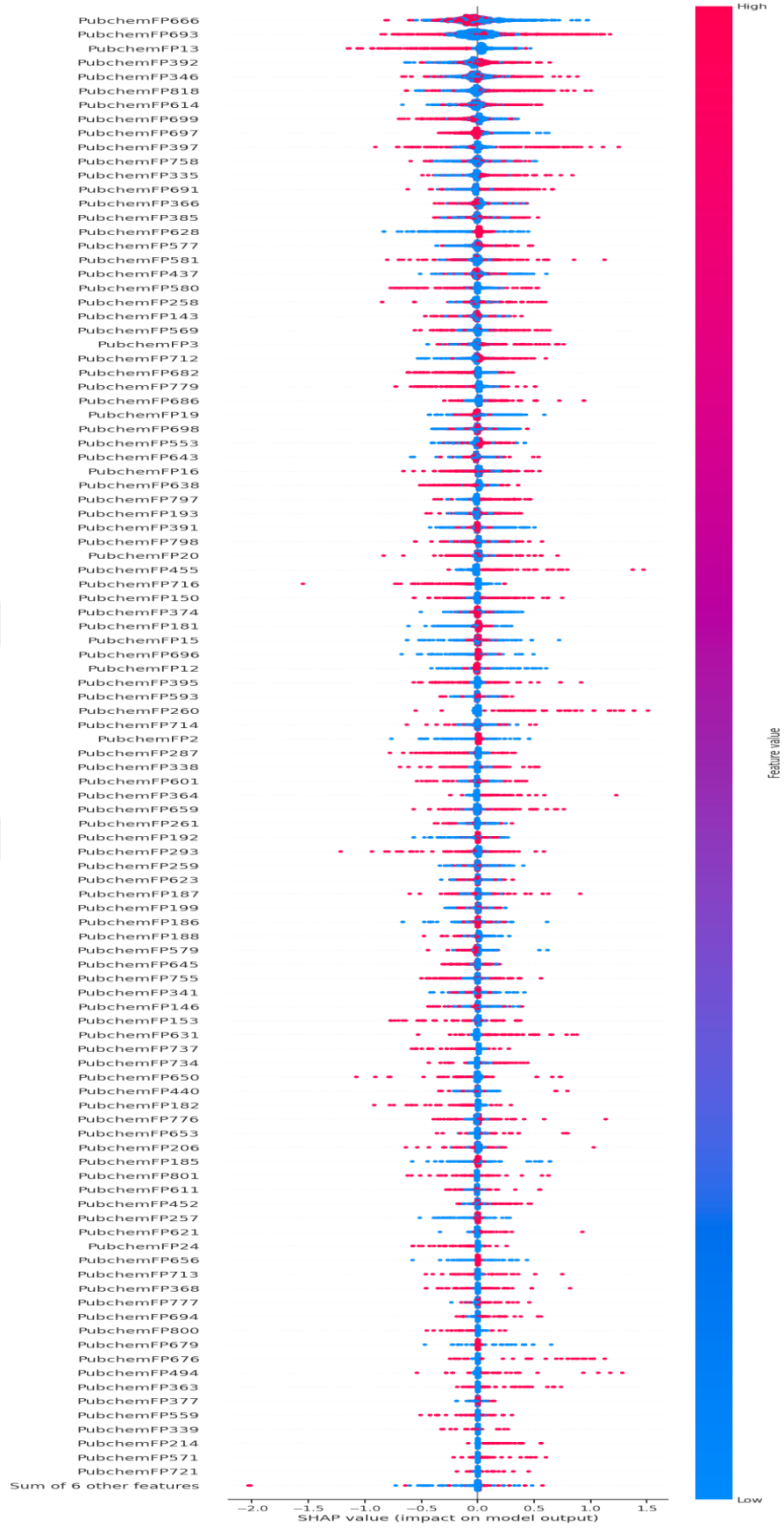
İnhibisyon için tasarlanan bir diğer decision görseli Şekil 5.8’de verilmiştir. Bu görsel, aykırı değerleri belirlemek için kullanılabilecek bir araçtır. Aykırı değerler, veri setinde genel eğilimden önemli ölçüde farklı olan veya diğer örneklerden büyük ölçüde sapmış olan veri noktaları olarak ifade edilmektedir. Bu tür değerler, modelin tahminini ve tahminin doğruluğunu etkilemektedir. İnhibisyon için oluşturulan görselde PubchemFP630 ve PubchemFP812’nin aykırı değerler bulundurduğu görülmektedir.



Şekil 5.8. İnhibisyon için decision grafiği-2

Bir veri kümesi içindeki kayda değer özelliklerin modelin çıktısı üzerindeki etkisinin kısa bir özetini sağlayan görselleştirme Beeswarm grafiği ile elde edilmektedir. Grafiğin amacı, her bir özellik için SHAP değerlerini görselleştirmek ve bu değerlerin dağılımını göstermektir. Grafiğin her bir satırı, bir özelliği temsil etmektedir. Her satırda, özelliğin farklı değerlerine sahip olan noktalar yer almaktadır. Noktaların yatay konumu, özelliğin SHAP değeriyle belirlenmektedir. Yani, bir özelliğin çıktıdaki etkisini göstermektedir. Noktaların yoğunluğu, her bir özellik satırında artmakta ve bu, özelliğin çıktı üzerindeki etkisinin gücünü göstermektedir. Ayrıca, her noktanın renk kullanılarak orijinal değeri de gösterilmektedir. Renk, özelliğin orijinal değerini görselleştirmek için kullanılmaktadır. Bu, orijinal veri setindeki farklı özellik değerlerini görmeye yardımcı olmaktadır.

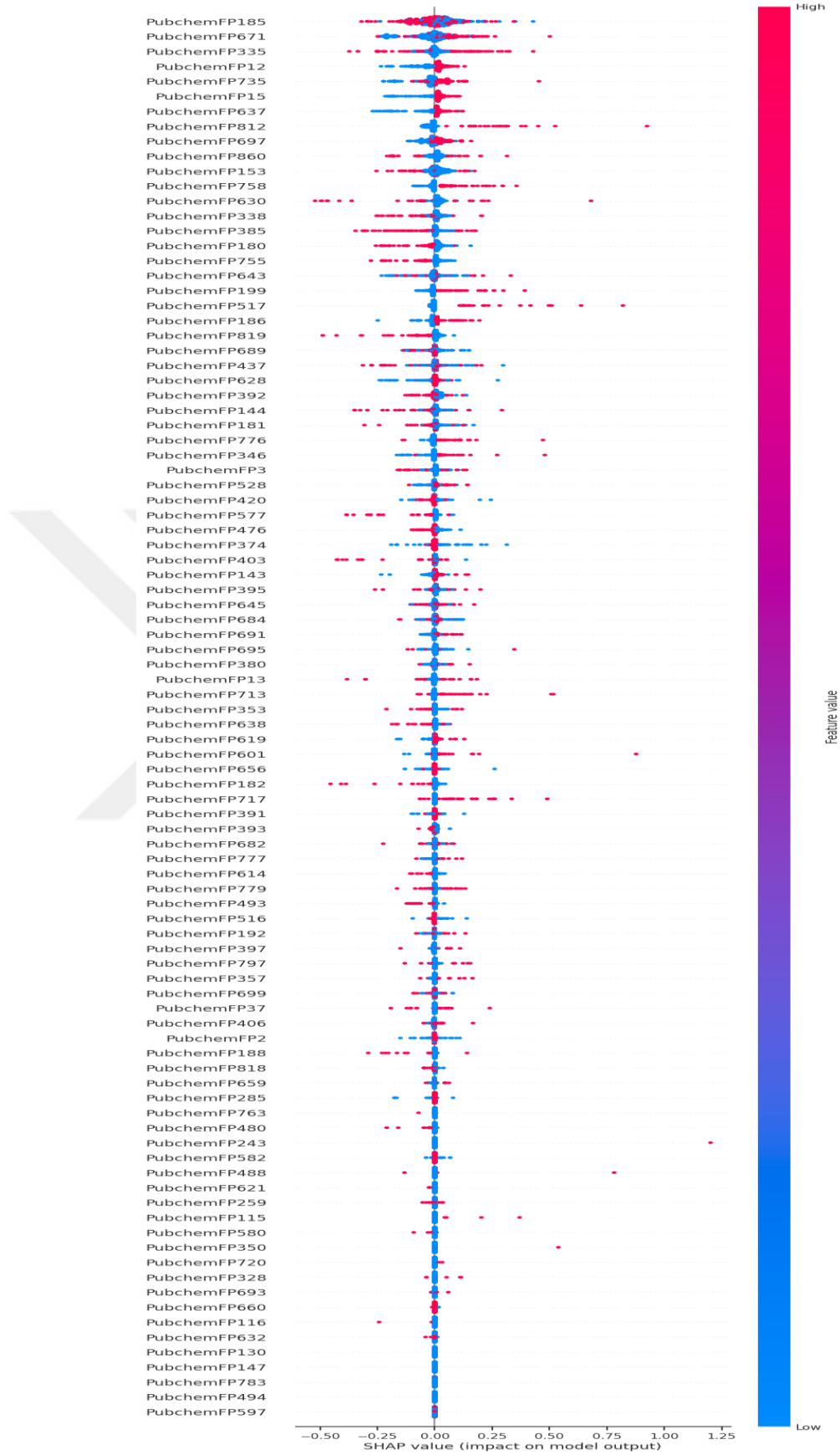
Beeswarm grafiği, özelliklerin çıktı üzerindeki etkisini anlamak ve farklı değerler arasındaki ilişkiyi görselleştirmek için kullanışlı bir araçtır. Yoğunluğu, dağılımı ve orijinal değerleri bir araya getirerek, bir veri kümesinin önemli özelliklerini anlamak için kapsamlı bir bilgi sunmaktadır. IC50 için oluşturulan beeswarm grafiği Şekil 5.9'da verilmiştir.



Şekil 5.9. IC50 için BeeSwarm Grafiği

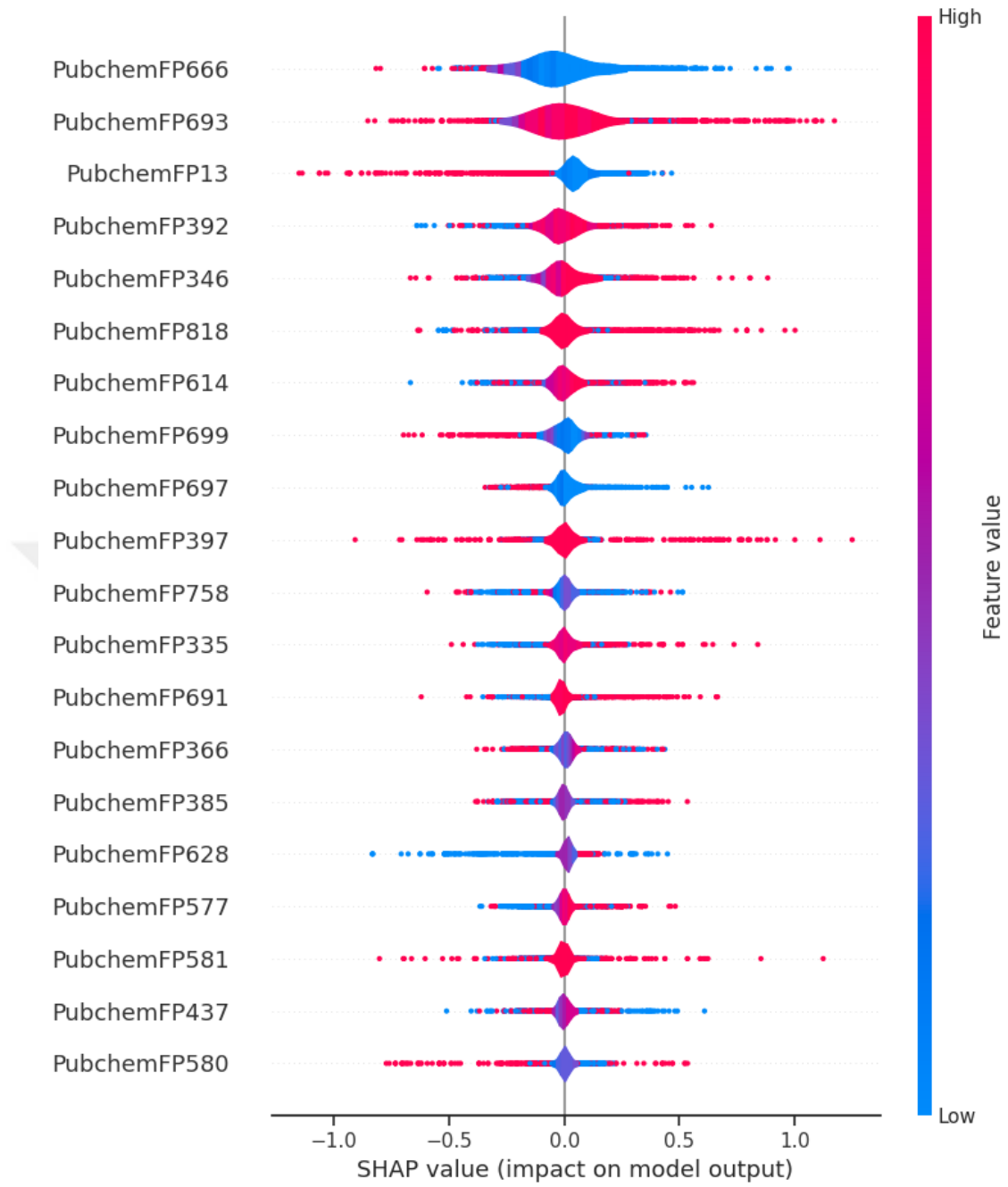
PubchemFP666'nın model çıktısı üzerindeki etkisi, olumsuz bir etkiye işaret eden mavi noktaların varlığıyla görsel olarak temsil edilmektedir. Tersine, Pubchem392'nin yüksek değeri görsel olarak kırmızı noktaların birikmesiyle temsil edilir ve model çıktısı üzerinde önemli bir etkiye işaret etmektedir. PubchemFP494 yüksek SHAP değerleri sergilemesine rağmen, noktaların dağınık dağılımı, etkisinin birkaç örnekle sınırlı olduğunu ve genel model çıktısına nispeten düşük bir katkı sağladığını göstermektedir.

İnhibisyon için oluşturulan BeeSwarm görseli Şekil 5.10'da verilmiştir. Buna göre PubchemFP755'in negatif yönde, PubchemFP12'nin pozitif yönde model çıktısına katkıda bulunduğu görselden tespit edilebilmektedir. Dolayısıyla beeswarm grafiği ile her bir özelliğin modelin tahminlerini etkilemede sahip olduğu değişen önem derecelerini görsel olarak temsil edilmiştir. Grafikteki veri noktalarının dağılımını ve konumu gözlemlenerek, hangi özelliklerin modelin çıktısı üzerinde daha önemli bir etkiye sahip olduğu ve genel performansa önemli ölçüde katkıda bulunduğunu belirlenmiştir.



Şekil 5.10. İnhibisyon için BeeSwarm Grafiği

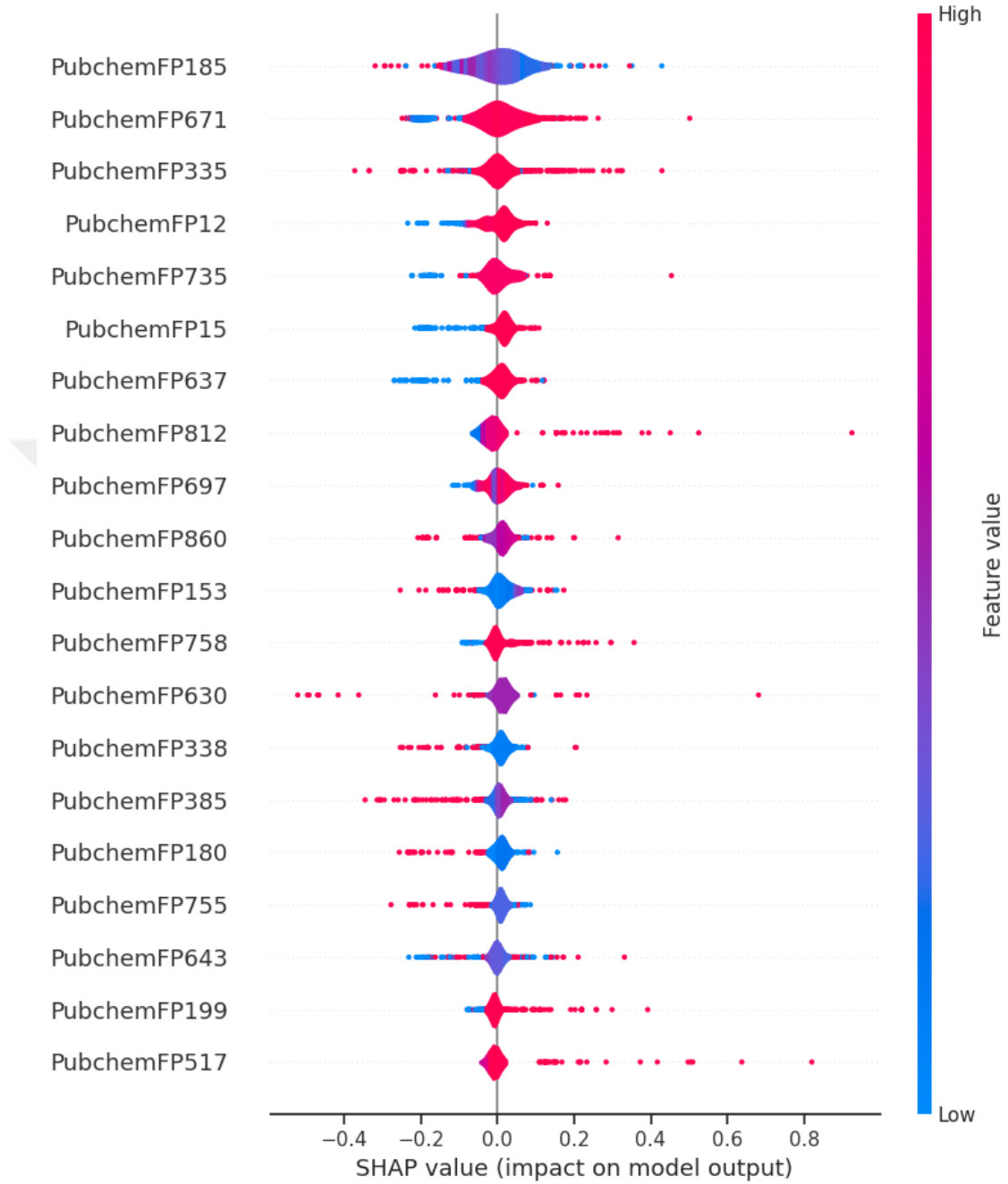
Violin grafiđi, özelliklerin çıktı üzerindeki etkisini karşılaştırmak ve farklı özellik değerleri arasındaki varyasyonu görmek için kullanışlı bir araçtır. Her bir özelliğın SHAP değerlerini göstermek için bir veya daha fazla yan yana duran keman şekline benzeyen bölgelerden oluşmaktadır. Her bir bölgenin genişliđi, SHAP değerlerinin yoğunluđunu temsil etmektedir. Yani, daha geniş bir bölge, daha yoğun bir SHAP değeri dağılımını ifade etmektedir. Grafiğın orta çizgisi, özelliğın ortalama SHAP değerini göstermektedir. Sol ve sađ taraftaki genişliđe yayılan şekiller, SHAP değerlerinin dağılımını belirtmektedir. Daha geniş bir şekil, SHAP değerlerinin daha yaygın bir dağılıma sahip olduđunu gösterirken, daha dar bir şekil daha odaklanmış bir dağılımı temsil etmektedir. Şekil 5.11’de IC50 için oluşturulan violin görseli verilmiştir. Buna göre beeswarm grafiğini destekleyecek şekilde model çıktısını, PubchemFP693’ün azaltan, PubchmeFP666’nın ise arttıran bir özellik olduđu görölmektedir.



Şekil 5.11. IC50 için Violin Grafiği

İnhibisyon için oluşturulan violin görseli Şekil 5.12’de verilmiştir. Buna göre beeswarm grafiğini destekleyen bir şekilde PubchemFP755’nin model çıktısına negatif katkı, PubchemFP12’nin pozitif katkı verdiği görülmektedir. PubchemFP185 özelliği, model çıktısını arttıran bir özellik olarak görünmektedir. Yani iki veya daha fazla altı üyeli halka içeren moleküllerin aktivitesi daha yüksek olma eğilimindedir. PubchemFP671 özelliği ise model çıktısını azaltan bir özellik

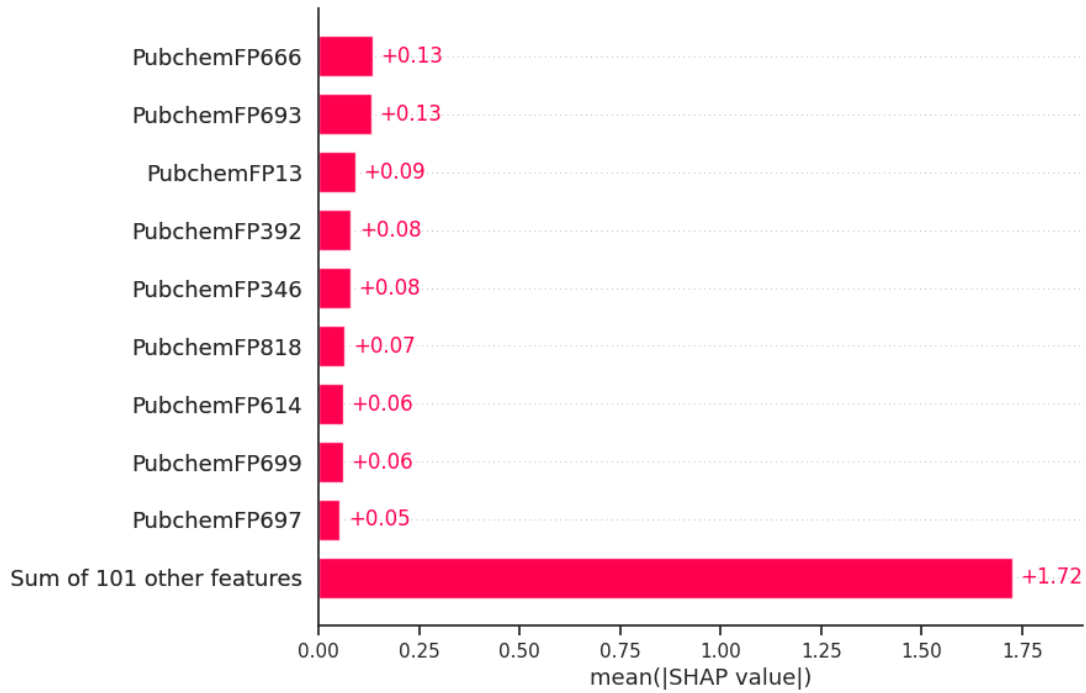
olarak görünmektedir. Yani bir aromatik halkalı nitro grubu içeren moleküllerin aktivitesi daha düşük olma eğilimindedir.



Şekil 5.12. İnhibisyon için Violin grafiği

Bar grafiği, bir özelliğin çıktı üzerindeki etkisini görselleştirmek için kullanılan bir grafik türüdür. Bu grafik, her bir özelliğin SHAP değerini bir çubuk şeklinde temsil etmektedir. IC50 için oluşturulan görsel Şekil 5.13’de verilmiştir. Görselden anlaşılacağı üzere diğer grafik türlerini destekleyen sonuçlar ortaya

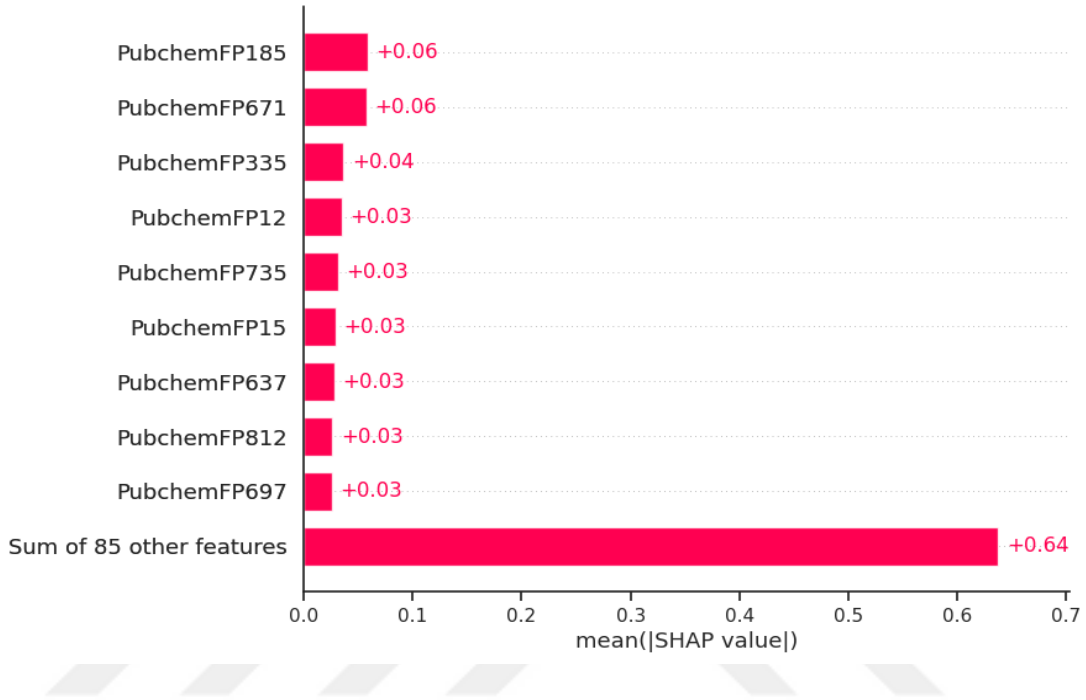
çıkmiştir. PubchemFP666, molekülün aromatik halka sayısı ile ilgilidir ve aromatik halkalar, karbon atomlarından oluşan ve elektronların paylaşıldığı dairesel yapılar olan kimyasal bileşenlerdir. Aromatik halkalar, ilaçların etkinliği ve toksisitesi üzerinde önemli bir rol oynamaktadır. PubchemFP693, molekülün hidrojen bağı vericisi sayısı ile ilgilidir. Hidrojen bağı vericileri, hidrojen atomlarına bağlı olan elektronegatif atomlar (genellikle oksijen veya azot) olan kimyasal gruplardır. Hidrojen bağı vericileri, ilaçların biyolojik aktivitesi ve çözünürlüğü üzerinde etkilidir. Literatürde, aromatik halkaların ve hidrojen bağı vericilerinin kanser hücrelerinin büyümesini ve çoğalmasını engelleyebileceği ile ilgili çalışmalar bulunmaktadır (Yalçın vd., 2017; Liu vd., 2013; Norman vd., 2012). Bu nedenle, PubchemFP666 ve PubchemFP693'ün lösemi ilaçlarının etkinliği ile ilişkili olabileceği öngörülmektedir.



Şekil 5.13. IC50 için Bar grafiği

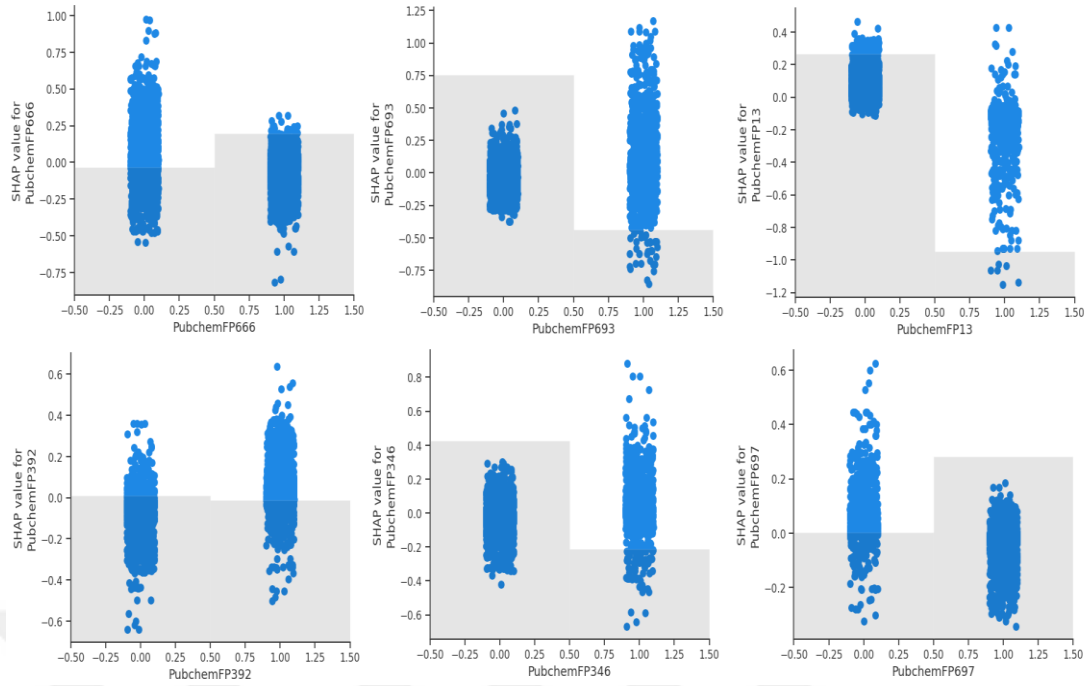
İnhibisyon için oluşturulan bar grafiği Şekil 5.14'de verilmiştir. Yine IC50'de olduğu gibi diğer grafik türlerini destekler şekilde sonuçlar ortaya çıkmıştır. PubchemFP185, molekülün içerisinde en az iki tane altı atomlu veya daha büyük halka yapısı olduğunu göstermektedir ve modelin çıktısını pozitif yönde 0.06 birim arttırmıştır. PubchemFP671 ise molekülde bir oksijen (O) atomunun bir

karbon (C) atomuna bağlı olduğu ve ardışık olarak üç karbon atomunun (C=C=C) bulunduğu bir yapıyı ifade etmektedir ve modelin sonucunu 0.06 birim pozitif yönde arttırmıştır. Grafikte diğer 85 özellik değerinin toplamının model çıktısına en büyük katkıyı sağladığı görülmektedir.



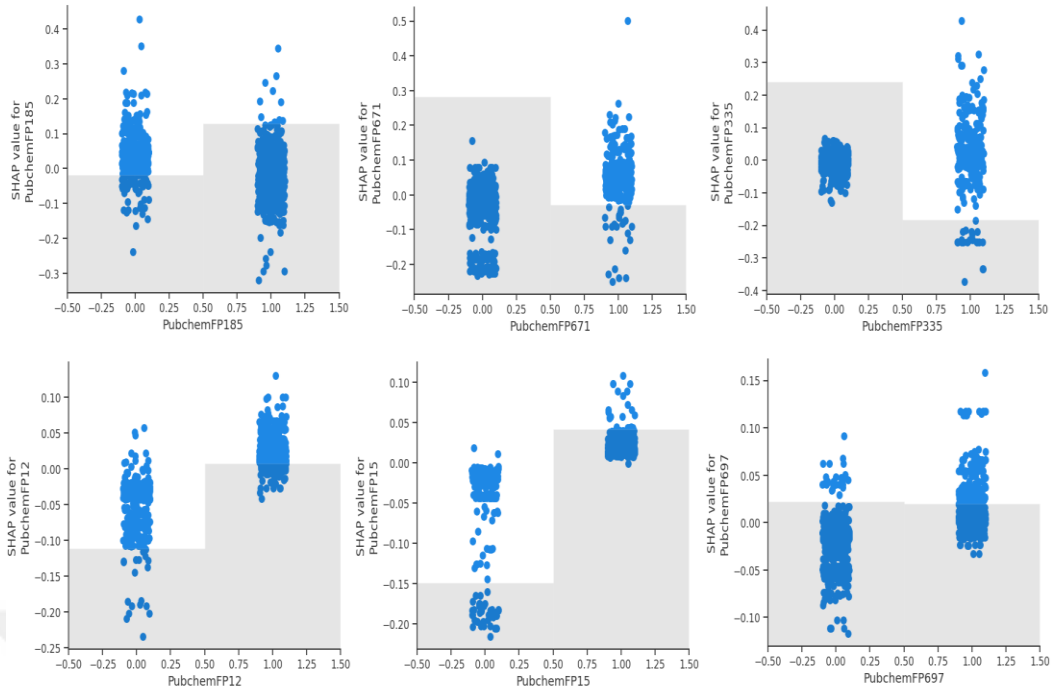
Şekil 5.14. İnhibisyon için Bar grafiği

Scatter grafikleri, özelliklerin çıktı üzerindeki etkisini analiz etmek ve özellik değerlerini karşılaştırmak için kullanılmaktadır. Ayrıca, aykırı değerleri veya belirli bir deseni gözlemlemek için de uygundur. Modelin davranışını anlamak, önemli özellikleri belirlemek ve veri setindeki ilişkileri görselleştirmek için kullanılan etkili bir araçtır. Grafikte her bir nokta, bir veri örneğini temsil etmekte ve x-ekseni üzerinde özelliğin değerini, y-ekseni üzerinde ise çıktı değerini göstermektedir. Renk veya boyut gibi ekstra görsel özellikler, başka bir özelliğin değerini veya etkisini gösterebilmektedir. IC50 için oluşturulan Scatter grafiklerinden bazıları Şekil 5.15’de verilmiştir.



Şekil 5.15. IC50 için Scatter grafikleri

Şekil 5.15'deki grafikler incelendiğinde, moleküllerin örnekler üzerindeki SHAP değerleri Y ekseninde, model tahminine olan etkileri X eksenini üzerinde görselleştirilmiştir. X eksenindeki değerler 0 ve 1 etrafında toplanmıştır. PubChem693, model çıktısına pozitif oranda etki etmiştir. 1 konumunda bulunan ve model çıktısına etki eden özelliklerin değerleri artmıştır. 693 numaralı SMARTS deseni "O=C-C-C-C-O" olarak ifade edilmektedir. Bu desen, karbon atomları arasında sırasıyla çift bağ, tek bağ, tek bağ, tek bağ, tek bağ ve son olarak tekrar çift bağ bulunan bir yapıyı temsil etmektedir. Ayrıca bu yapıda başlangıçta bir oksijen atomu (O) ve sonunda da bir oksijen atomu (O) bulunmaktadır. Bu desen, moleküllerin bu yapıya sahip olup olmadığını test etmek için kullanılmaktadır. PubChemFP666 ise model çıktısına negatif oranda etki etmiştir ve 1 konumundaki değerler 0 konumuna göre daha düşüktür. 666 numaralı SMARTS deseni "C=C-C-O-C" olarak ifade edilmektedir. Bu desen, bir çift bağ içeren karbon atomu (C), ardından tek bağlı bir karbon atomu (C), oksijen atomu (O) ve tekrar tek bağlı bir karbon atomu (C) içeren bir yapıyı temsil etmektedir.

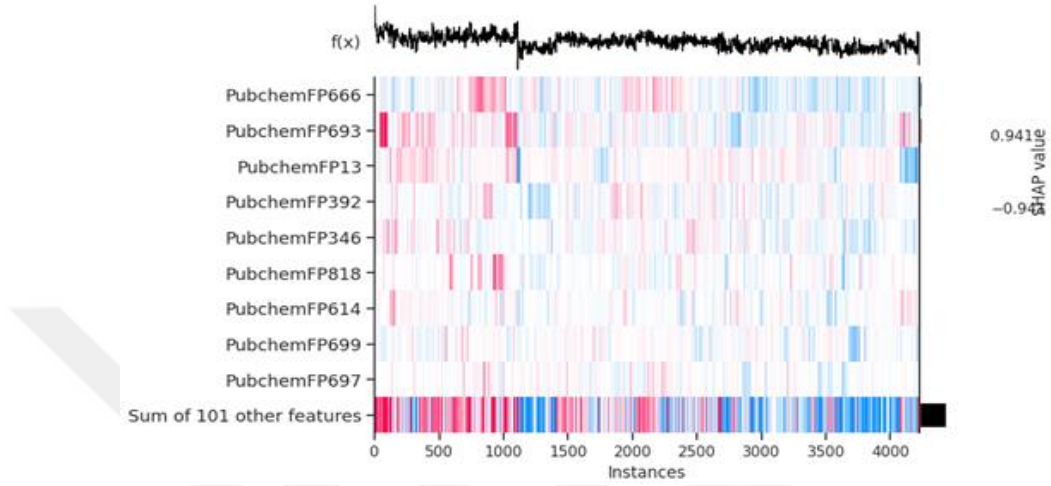


Şekil 5.16. İnhibisyon için Scatter grafikleri

Şekil 5.16'da verilen inhibisyon için scatter grafikleri incelendiğinde PubchemFP671 ve PubchemFP335'in model çıktısına pozitif oranda etki etmişlerdir ve 1 konumunda bulunan, model çıktısına etki eden özelliklerin değerleri artmış durumdadır. PubchemFP671, bir karbonil grubunu ($O=C$), ardından çift bağlı bir karbon atomunu ($C=C$), tekrar tek bağlı bir karbon atomunu (C) ve son olarak yine çift bağlı bir karbon atomunu (C) içeren bir yapı iken PubchemFP335, dört adet hidrojen bağlı karbon atomunu (C) ve bir adet tek bağlı karbon atomunu ($\sim C$) içeren bir yapıdır. PubchemFP12 ve PubchemFP697 ise model çıktısına negatif etki ettiği belirlenmiştir. PubchemFP12, en az 16 adet karbon atomu içeren bir yapı ile temsil edilirken, PubchemFP697 ise bir karbon zincirini temsil etmektedir. Bu desende C sembolü, karbon atomlarını temsil ederken, - sembolü ise karbon atomları arasındaki bağları ifade etmektedir.

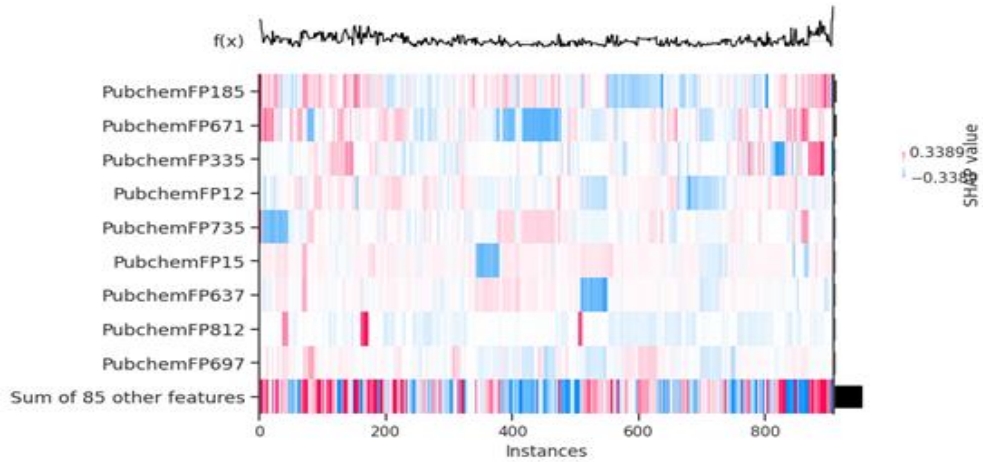
Heatmap grafiği, her bir özelliğin hedef değişken üzerindeki etkisini net bir şekilde gösteren bir grafik türüdür. Özelliklerin yüksek etkiye sahip olduğu bölgeler daha koyu renkte ve yoğunlukta gösterilirken, düşük etkiye sahip olan bölgeler daha açık renkte ve yoğunlukta gösterilmektedir. Bu grafik, veri

kümesinin özelliklerini daha iyi anlamak, önemli etkilere sahip olan özellikleri belirlemek ve modelin yorumlanabilirliğini artırmak için kullanışlı bir araç olarak ifade edilmektedir. IC50 için oluşturulan heatmap grafiği Şekil 5.17’de verilmiştir. Diğer SHAP grafiklerini destekleyecek şekilde sonuçlar ortaya çıkmıştır.



Şekil 5.17. IC50 için Heatmap grafiği

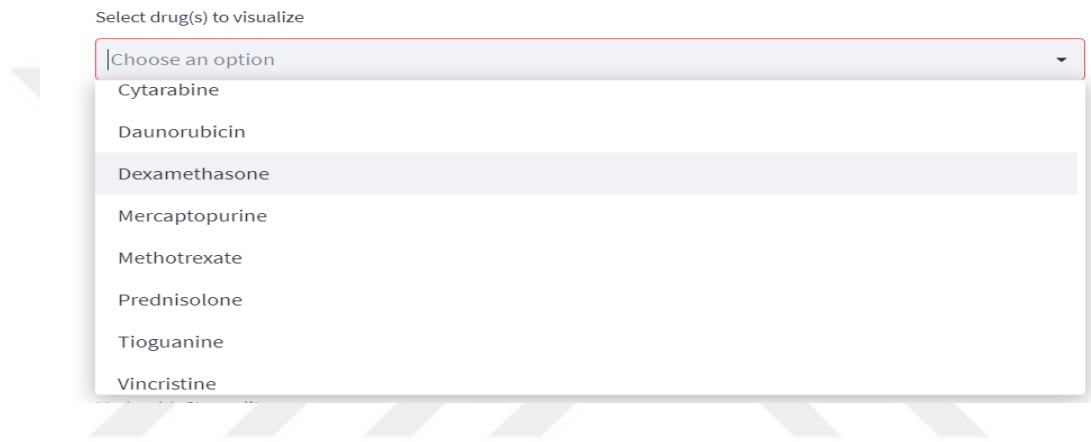
İnhibisyon için oluşturulan heatmap görseli Şekil 5.18’de verilmiştir. $f(x)$ ’deki yüksek değerler, moleküller ile ilgili yüksek tahmin değerlerini göstermektedir. Örnek sayısı arttıkça tahmindeki değerlerin de arttığı görülmektedir. Yine diğer SHAP görselleri ile benzer sonuçlar verdiği tespit edilmiştir.



Şekil 5.18. İnhibisyon için Heatmap grafiği

Çocukluk Çağı Akut Lösemisi tedavisinde kullanılan Mercaptopurine, Tioguanine, Prednisolone, Dexamethasone, Vincristine, Methotrexate, Daunorubicin, Cytarabine ilaçlarının diğer ilaçlarla olan ilişkilerine yönelik oluşturulan ağ grafikleri için bir benzetim ortamı tasarlanmış ve görseli Şekil 5.19’da verilmiştir. Streamlit kütüphanesi ile etkileşimli bir arayüz tasarlanarak listeden seçilen ilaçların ağ grafikleri görselleştirilebilmektedir.

Network Graph Visualization of Drug-Drug Interactions



Şekil 5.19. Çocukluk çağı akut lösemisi tedavisinde kullanılan ilaçların etkileşimi için oluşturulan benzetim ortamı

Etkileşimli veri görselleştirilmesi ile en az bir ilaç seçilerek de ağ grafiği oluşturulabilmektedir. Bu ilaçların kimyasal yapılarında bazı ortak veya benzer özellikler bulunmaktadır (PubMed; MedlinePlus; RxList):

- Halkalı yapılar: Bu ilaçların çoğunda aromatik halkalı yapılar bulunmaktadır. Bu halkalar, molekülün stabilitesine katkıda bulunmaktadır ve ilacın etkisini artırmaktadır.
- Hidrokarbon zincirleri: Her bir ilaç, farklı uzunluklarda ve farklı yapısal özelliklere sahip hidrokarbon zincirlerine sahiptir. Bu zincirler, ilacın lipofilik özelliklerini ve biyolojik etkileşimlerini etkilemektedir.

- Fonksiyonel gruplar: Bu ilaçların çoğunda farklı fonksiyonel gruplar bulunur. Örneğin, Mercaptopurine ve Tioguanine'de tiyol (-SH) grubu bulunurken, Prednisolone ve Dexamethasone'da keton (C=O) ve hidroksil (-OH) grupları bulunmaktadır.
- Heterosiklik halkalar: Birkaç ilaçta heterosiklik halkalar bulunmaktadır. Örneğin, Mercaptopurine ve Tioguanine'de purin halkası bulunurken, Methotrexate'te folat antagonist olarak etki eden pteridin halkası vardır.
- İyonizasyon: Bu ilaçların birçoğu asidik veya bazik özelliklere sahiptir ve çözeltide iyonize olabilmektedir. Bu, ilacın çözünürlüğünü, stabilitesini ve etki mekanizmasını etkileyebilmektedir.

Açıklanabilir yapay zekâ modelleri ile elde edilen bulgular ile etkileşimli ağ grafikleri sonucunda elde edilen sonuçların birbirini destekler nitelikte olması bu tez çalışması kapsamında önemli bir sonuçtur. Yapay zekâ modellerinin SHAP grafiklerinden elde edilen moleküler parmak izlerinin sahip olduğu molekül yapısının, ilaç etkileşimlerinin karmaşıklığının giderilmesi ile ortaya çıkan ilaçlardaki molekül yapıları ile benzerlik göstermesi yapılan analizin doğruluğunu göstermektedir. Dolayısıyla ilaç yeniden kullanım çalışmalarında, ilaç keşfindeki Faz 1 (güvenlik) aşamasının atlanarak Faz 2 (etkililik) aşamasına geçilmesinin zaman ve maliyet yönünden fayda sağlamasında yapay zekanın rolü kritik bir önem arz etmektedir.

Gelecek çalışmalarda yapılabilecek birkaç potansiyel bulunmaktadır. Analiz için kullanılan veri setinin genişletilmesi ve iyileştirilmesi sonuçların doğruluğunu ve sağlamlığını artırabilir. Genomik, proteomik ve klinik veriler gibi çeşitli veri kaynaklarının birleştirilmesi, ilaç etkileşimlerinin ve yeniden konumlandırma fırsatlarının daha kapsamlı bir şekilde anlaşılmasını sağlayabilir. Ek moleküler özelliklerin ve tanımlayıcıların keşfedilmesi ve dahil edilmesi daha iyi tahminlere ve daha bilgilendirici farkındalıklar ortaya çıkarabilir. Yapısal bilgilerin kodlanması veya 3D moleküler temsillerin dahil edilmesi gibi özellik mühendisliği teknikleri, makine öğrenimi modelleri için daha zengin temsiller

sağlayabilir. Sonuçların harici veri setleri veya gerçek dünya deneyleri kullanılarak doğrulanması, bulguların genellenebilirliğini değerlendirmek için gereklidir. Tahmin edilen etkileşimleri ve yeniden konumlandırılan ilaçları doğrulamak için in vitro veya in vivo deneyler yapmak, hesaplamalı tahminlerin güvenilirliğini artırabilir. Başarılı bir ilaç yeniden konumlandırması için, belirlenen ilaç adaylarının güvenliğini ve etkinliğini doğrulamak üzere klinik deneyler yapmak gerekmektedir. Tıp uzmanlarıyla iş birliği yapmak ve iyi tasarlanmış klinik çalışmalar yürütmek, yeniden konumlandırılan ilaçların pratikte uygulanabilirliğini sağlamalıdır. Araştırmanın potansiyel yan etkilerin ve ilaç-ilaç etkileşimlerinin tahminini içerecek şekilde genişletilmesi, çocukluk çağı akut lösemi tedavisi için daha güvenli ve daha etkili ilaç kombinasyonlarının belirlenmesine yardımcı olabileceği öngörülmektedir.

KAYNAKLAR

- Aas, K., Jullum, M., Løland, A. 2021. Explaining individual predictions when features are dependent: More accurate approximations to Shapley values. *Artificial Intelligence*, 298, 103502.
- Akalın, F., Yumuşak, N. 2023. Mikrodizi veri kümesindeki ALL, AML ve MLL lösemi türlerine ilişkin gen anomalilerinin LSTM sinir ağı ile sınıflandırılması. *Gazi Üniversitesi Mühendislik Mimarlık Fakültesi Dergisi*, 38(3), 1299-1306.
- Anadolu Agency, 2022. German pharmacies warn of shortage of cancer drugs. Erişim tarihi: 13 Mart 2023. <https://www.aa.com.tr>
- Algaba, E., Fragnelli, V., Sánchez-Soriano, J. (Eds.). 2019. *Handbook of the Shapley value*. CRC Press.
- Altmann, A., Toloşi, L., Sander, O., Lengauer, T. 2010. Permutation importance: a corrected feature importance measure. *Bioinformatics*, 26(10), 1340-1347.
- Arrowsmith, J. 2011. Phase III and submission failures: 2007–2010. *Nature Reviews Drug Discovery*, 10(2), 87-87.
- Balı, D. A., Kürekçi, A. E., Akar, M. N. 2018. HDAC2, HDAC4, HDAC5, HDAC7, HDAC8, HDAC9 Gen İfade Seviyelerinin Çocukluk Çağı Akut Lösemilerinde Prognoz İle İlişkisi. *SDÜ Tıp Fakültesi Dergisi*, 25(4), 400-406.
- Barrios, C., de Lima Lopes, G., Yusof, M. M., Rubagumya, F., Rutkowski, P., Sengar, M. 2023. Barriers in access to oncology drugs—a global crisis. *Nature Reviews Clinical Oncology*, 20(1), 7-15.
- Bordbar, M., Jam, N., Karimi, M., Shahriari, M., Zareifar, S., Zekavat, O. R., Mottaghipisheh, H. 2023. The survival of childhood leukemia: An 8-year single-center experience. *Cancer Reports*, 6(4), e1784.
- Bundela, A. S., Rahul, M. R. 2022. Machine learning-enabled framework for the prediction of mechanical properties in new high entropy alloys. *Journal of Alloys and Compounds*, 908, 164578.
- Carroll, W.L., Bhatla, T. 2016. Acute lymphoblastic leukemia. In: Lankowsky P, Lipton JM & Fish JD. (eds.) *Lankowsky's manual of pediatric hematology and oncology* (sixth edition). San Diego: Academic Press.
- Chalkiadakis, G., Elkind, E., Wooldridge, M. 2011. Computational aspects of cooperative game theory. *Synthesis Lectures on Artificial Intelligence and Machine Learning*, 5(6), 1-168.

- Chan, H. S., Shan, H., Dahoun, T., Vogel, H., Yuan, S. 2019. Advancing drug discovery via artificial intelligence. *Trends in Pharmacological Sciences*, 40(8), 592-604
- Chang, Y., Hawkins, B. A., Du, J. J., Groundwater, P. W., Hibbs, D. E., Lai, F. 2023. A Guide to In Silico Drug Design. *Pharmaceutics*, 15(1), 49.
- Chen, H., Engkvist, O., Wang, Y., Olivecrona, M., Blaschke, T. 2018. The rise of deep learning in drug discovery. *Drug discovery today*, 23(6), 1241–1250. <https://doi.org/10.1016/j.drudis.2018.01.039>.
- Christoph, M. 2020. Interpretable machine learning. A Guide for Making Black Box Models Explainable.
- Cohen, E., Musa, A. 2023. Thousands of people can't get full treatments of a lifesaving cancer drug. *CNN Health*. Erişim tarihi: 13 Mart 2023. <https://edition.cnn.com/2023/02/15/health/cancer-drug-shortage-bcg/index.html> Son erişim tarihi: 13 Mart 2023.
- Cortés-Ciriano, I., Škuta, C., Bender, A., Svozil, D. 2020. QSAR-derived affinity fingerprints (part 2): modeling performance for potency prediction. *Journal of Cheminformatics*, 12(1), 41.
- Covert, I. C., Lundberg, S., Lee, S. I. 2021. Explaining by removing: A unified framework for model explanation. *The Journal of Machine Learning Research*, 22(1), 9477-9566.
- Cress, W. D., Seto, E. 2000. Histone deacetylases, transcriptional control, and cancer. *Journal of cellular physiology*, 184(1), 1-16.
- Croset, S. 2014. Drug repositioning and indication discovery using description logics (Doctoral dissertation, University of Cambridge).
- Çelik, İ. N., Arslan, F. K., Tunç, R., Yıldız, İ. 2021. İlaç keşfi ve geliştirilmesinde yapay zekâ. *Journal of Faculty of Pharmacy of Ankara University*, 45(2), 17-17.
- DelveInsight, 2023. Acute Lymphocytic Lösemi Pipeline Insight 2023. Erişim Tarihi : 29.06.2023. <https://www.delveinsight.com>
- Demir, D. 2023. Çocukluk çağı lösemileri ve hemşirelik bakımı. *Sağlık & Bilim 2023: Çocuk hemşireliğinde kanıta dayalı uygulamalar*, Ed. A.Sarman. Efe Akademi Yayınları, 95.
- Dhamal K. S., Waghmare S.A., Kamble H., Shelar V.S., Bhalekar, O. S. (2022). Role Of Computer Aided Drug Design In Drug Discovery and Development. *International Journal of Research in Engineering and Science*, 10(3), 01-05.

- 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.
- Doshi-Velez, F., Kim, B. 2017. Towards a rigorous science of interpretable machine learning. arXiv preprint arXiv:1702.08608.
- Duarte, Y., Márquez-Miranda, V., Miossec, M. J., González-Nilo, F. (2019). Integration of target discovery, drug discovery and drug delivery: a review on computational strategies. *Wiley Interdisciplinary Reviews: Nanomedicine and Nanobiotechnology*, 11(4), e1554.
- Ellis-Petersen, H., Senanayake, D. 2022. People are going to die': crisis-hit Sri Lanka runs out of medicine. *The Guardian World News*. Erişim Tarihi : 13 Mart 2023. <https://www.theguardian.com>
- Eyal, S., Yagen, B., Shimshoni, J. Bialer, M., 2005. Histone Deacetylases Inhibition and Tumor Cells Cytotoxicity by CNS-Active VPA Constitutional Isomers and Derivatives. *Biochemical Pharmacology*, 69(10), 1501–1508.
- Fernández-de Gortari, E., García-Jacas, C. R., Martinez-Mayorga, K., Medina-Franco, J. L. 2017. Database fingerprint (DFP): an approach to represent molecular databases. *Journal of cheminformatics*, 9(1), 1-9.
- Ferreira, L., Andricopulo, A. D. 2019. ADMET modeling approaches in drug discovery. *Drug discovery today*, 24(5), 1157–1165. <https://doi.org/10.1016/j.drudis.2019.03.015>
- Frye, C., Rowat, C., Feige, I. 2020. Asymmetric shapley values: incorporating causal knowledge into model-agnostic explainability. *Advances in Neural Information Processing Systems*, 33, 1229-1239.
- Fryer, D., Strümke, I., Nguyen, H. 2020. Shapley value confidence intervals for attributing variance explained. *Frontiers in Applied Mathematics and Statistics*, 6, 587199.
- Fryer, D., Strümke, I., Nguyen, H. 2021. Shapley values for feature selection: The good, the bad, and the axioms. *IEEE Access*, 9, 144352-144360.
- Gaulton, A., Hersey, A., Nowotka, M., Bento, A. P., Chambers, J., Mendez, D., Leach, A. R. 2017. The ChEMBL database in 2017. *Nucleic acids research*, 45(D1), D945-D954.
- Gawehn, E., Hiss, J. A. Schneider, G. 2016. Deep Learning in Drug Discovery. *Molecular informatics*, 35 (1), 3–14. doi:10.1002/minf.201501008
- Goh, G. B., Siegel, C., Vishnu, A., Hodas, N. O. Baker, N. 2017. Chemception: a deep neural network with minimal chemistry knowledge matches the

performance of expert-developed QSAR/QSPR models. arXiv preprint arXiv:1706.06689

- Göcen, T. (2022). HDAC inhibitörü 2-(N-(2-hidroksifenil)-2-propilpentanamid)' in Kuantum Kimyasal Hesaplamalar ile Moleküler Yapı, Spektroskopik (IR, NMR, UV-VIS) İncelemeleri ve Lineer Olmayan Optik (NLO) Analizi . Afyon Kocatepe Üniversitesi Fen Ve Mühendislik Bilimleri Dergisi , 22 (1) , 1-15 . DOI: 10.35414/akufemubid.1036164.
- Grömping, U. 2009. Variable importance assessment in regression: linear regression versus random forest. *The American Statistician*, 63(4), 308-319.
- Gruhn, B., Naumann, T., Gruner, D., Walther, M., Wittig, S., Becker, S., Sonnemann, J. 2013. The expression of histone deacetylase 4 is associated with prednisone poor-response in childhood acute lymphoblastic leukemia. *Leukemia research*, 37(10), 1200-1207.
- Hagberg, A., Swart, P., S Chult, D. 2008. Exploring network structure, dynamics, and function using NetworkX (No. LA-UR-08-05495; LA-UR-08-5495). Los Alamos National Lab.(LANL), Los Alamos, NM (United States).
- Harris, C. R., Millman, K. J., Van Der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Oliphant, T. E. 2020. Array programming with NumPy. *Nature*, 585(7825), 357-362.
- Hemmerich, J. Ecker, G. F. 2020. In silico toxicology: From structure–activity relationships towards deep learning and adverse outcome pathways. *Wires computational molecular science*, 10 (4), doi:10.1002/wcms.1475
- Hui, Y. 2020. Data of multiple-type drug-drug interactions, Mendeley Data, V1, doi:10.17632/md5czfsfnd.1
- Jeha, S., Pui, C. H. 2009. Risk-adapted treatment of pediatric acute lymphoblastic leukemia. *Hematology/Oncology Clinics*, 23(5), 973-990.
- Katsila, T., Spyroulias, G. A., Patrinos, G. P., Matsoukas, M. T. (2016). Computational approaches in target identification and drug discovery. *Computational and structural biotechnology journal*, 14, 177-184.
- Karami, T. K., Hailu, S., Feng, S., Graham, R., Gukasyan, H. J. (2022). Eyes on Lipinski's Rule of Five: A New “Rule of Thumb” for Physicochemical Design Space of Ophthalmic Drugs. *Journal of Ocular Pharmacology and Therapeutics*, 38(1), 43-55.
- Kırboğa, K. K., Küçüksille, E. U., Köse, U. 2022. Ignition of Small Molecule Inhibitors in Friedreich's Ataxia with Explainable Artificial Intelligence. 10.21203/rs.3.rs-1408745/v1

- Klimek, V. M., Fircanis, S., Maslak, P., Guernah, I., Baum, M., Wu, N., Nimer, S. D. 2008. Tolerability, pharmacodynamics, and pharmacokinetics studies of depsipeptide (romidepsin) in patients with acute myelogenous leukemia or advanced myelodysplastic syndromes. *Clinical Cancer Research*, 14(3), 826-832.
- Kulkarni, N. S., Guerro, Y., Gupta, N., Muth, A., Gupta, V. 2019. Exploring potential of quantum dots as dual modality for cancer therapy and diagnosis. *Journal of Drug Delivery Science and Technology*, 49, 352-364.
- Kumar, I. E., Venkatasubramanian, S., Scheidegger, C., Friedler, S. 2020. Problems with Shapley-value-based explanations as feature importance measures. In *International Conference on Machine Learning*, 5491-5500, PMLR.
- Küçükoğlu, K. 2013. Histonların Asetilasyonu ve Histon Deasetilaz İnhibitörleri. *Turkiye Klinikleri J Pharm Sci*. 2013;2(2):55-73.
- Landrum, G. (2019). Rdkit documentation. Release, 1(1-79), 4.
- Lapuschkin, S., Wäldchen, S., Binder, A., Montavon, G., Samek, W., Müller, K. R. 2019. Unmasking Clever Hans predictors and assessing what machines really learn. *Nature communications*, 10(1), 1096.
- Laszlo, A. K. 2019. *Partition function form games: Coalitional games with externalities*. Springer.
- LazyPredict Documentation, 2022. Erişim Tarihi : 29.06.2023 <https://lazypredict.readthedocs.io/en/latest/>
- Lecun, Y., Bengio, Y. Hinton, G. 2015. Deep learning. *Nature*, 521 (7553), 436–444. doi:10.1038/nature14539
- Lee-Sherick, A. B., Linger, R. M., Gore, L., Keating, A. K., Graham, D. K. 2010. Targeting paediatric acute lymphoblastic leukaemia: novel therapies currently in development. *British journal of haematology*, 151(4), 295-311.
- Lenselink, E. B., Ten Dijke, N., Bongers, B., Papadatos, G., van Vlijmen, H. W. T., Kowalczyk, W., van Westen, G. J. P. 2017. Beyond the hype: deep neural networks outperform established methods using a ChEMBL bioactivity benchmark set. *J Cheminform* 9 (1): 45.
- Leeson, P. D. Springthorpe, B. 2007. The influence of drug-like concepts on decision-making in medicinal chemistry. *Nature reviews drug discovery*, 6 (11), s. 881–890. <https://doi:10.1038/nrd2>
- Li, F., Yin, J., Lu, M., Mou, M., Li, Z., Zeng, Z., Zhu, F. 2023. DrugMAP: molecular atlas and pharma-information of all drugs. *Nucleic acids research*, 51(D1), D1288-D1299.

- Lim, S. 2023. The Process and Costs of Drug Development 2022. FTLO Science. Erişim tarihi: 13 Mart 2023. <https://ftloscience.com/process-costs-drug-development>
- Lipinski, C. A. 2000. Drug-like properties and the causes of poor solubility and poor permeability. *Journal of pharmacological and toxicological methods*, 44(1), 235-249.
- Lipinski, C. A. 2004. Lead-and drug-like compounds: the rule-of-five revolution. *Drug discovery today: Technologies*, 1(4), 337-341.
- Lipinski, C. A., Lombardo, F., Dominy, B. W., Feeney, P. J. 2012. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Advanced drug delivery reviews*, 64, 4-17.
- Liu, J., Zhu, Y., He, Y., Zhu, H., Gao, Y., Li, Z., Li, W. 2020. Combined pharmacophore modeling, 3D-QSAR and docking studies to identify novel HDAC inhibitors using drug repurposing. *Journal of Biomolecular Structure and Dynamics*, 38(2), 533-547.
- Liu, K.J., He, J.H., Su, X.D., Sim, H.M., Xie, J.D., Chen, X.G., ...Fu, L.W. 2013. Saracatinib (AZD0530). Is a potent modulator of Abl-mediated multidrug resistance in vitro and in vivo. *International Journal of Cancer*, 132(1), 224-235. doi:10.1002/ijc.27649.
- Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., Lee, S. I. 2020. From local explanations to global understanding with explainable AI for trees. *Nature machine intelligence*, 2(1), 56-67.
- Lundberg, S. M., Lee, S. I. 2017. A unified approach to interpreting model predictions. *Advances in neural information processing systems*, 30, 1-10.
- Matplotlib, 2023. Erişim Tarihi : 28 Haziran 2023. <https://matplotlib.org>
- MedlinePlus Amerikan Ulusal Tıp Kütüphanesi, 2023. Erişim Tarihi : 18 Haziran 2023. medlineplus.gov
- Melitoshevich, V. A., Alikulovich, V. D. 2023. Development by a Graphic User Interface-Programs in the Tkinter Package Using Modern Pedagogical Technologies in the Field of Medicine. *Miasto Przyszłości*, 32, 13-17.
- Méndez-Lucio, O., Baillif, B., Clevert, D.-A., Rouquié, D. Wichard, J. 2020. De novo generation of hit-like molecules from gene expression signatures using artificial intelligence. *Nature Communications* 11, 1-10.

- Merk, D., Friedrich, L., Grisoni, F. Schneider, G. 2018. De novo design of bioactive small molecules by artificial intelligence. *Molecular Informatics* 37, 1700153.
- Merrick, L., Taly, A. 2020. The explanation game: Explaining machine learning models using shapley values. In *Machine Learning and Knowledge Extraction: International Cross-Domain Conference, Proceedings 4*, 17-38. Springer International Publishing.
- Miller, T. 2019. Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence* 267, 1–38.
- Moreno, D. A., Scrideli, C. A., Cortez, M. A. A., De Paula Queiroz, R., Valera, E. T., Da Silva Silveira, V., Tone, L. G. 2010. Differential expression of HDAC3, HDAC7 and HDAC9 is associated with prognosis and survival in childhood acute lymphoblastic leukaemia. *British journal of haematology*, 150(6), 665-673.
- Muratov, E. N., Bajorath, J., Sheridan, R. P., Tetko, I. V., Filimonov, D., Poroikov, V., Tropsha, A. 2020. QSAR without borders. *Chemical Society Reviews*, 49(11), 3525-3564.
- Murdoch, W. J., Singh, C., Kumbier, K., Abbasi-Asl, R. Yu, B. 2019. Interpretable machine learning: definitions, methods, and applications. *arXiv preprint arXiv:1901.04592*.
- Navia, M.A. and Chaturvedi, P.R. 1996. Design principles for orally bioavailable drugs. *Drug Development Today* 1, 179-189.
- Norman, R. A., Toader, D., Ferguson, A. D. 2012. Structural approaches to obtain kinase selectivity. *Trends in Pharmacological Sciences*, 33(5), 273-278. doi: <https://doi.org/10.1016/j.tips.2012.03.005>.
- Nonzee, N. J., Luu, T. H. 2019. The drug shortage crisis in the United States: impact on cancer pharmaceutical safety. *Cancer Policy: Pharmaceutical Safety*, 75-92.
- Oprea, T. I., Davis, A. M., Teague, S. J., Leeson, P. D. 2001. Is there a difference between leads and drugs? A historical perspective. *Journal of chemical information and computer sciences*, 41(5), 1308–1315. <https://doi.org/10.1021/ci010366a>
- Özkan, H. 2019. N-Alkillenmiş Tetrazol Türevi Bileşiklerin Absorpsiyon, Dağılım, Metabolizma ve Atılım (ADME) Özelliklerinin Araştırılması. *Süleyman Demirel Üniversitesi Fen Edebiyat Fakültesi Fen Dergisi*, 14(2), 384-394. <https://doi.org/10.29233/sdufeffd.631009>

- Öztürk, H., Özgür, A. Ozkirimli, E. 2018. DeepDTA: Deep drug- target binding affinity prediction. *Bioinformatics* 34, 821– 829.
- Pacaud, R., Garcia, J., Thomas, S., Munster, P. N. 2023. Clinical Applications of Histone Deacetylase Inhibitors. In *Handbook of Epigenetics*, 793-819. Academic Press.
- Padhy, B. M., Gupta, Y. K. 2011. Drug repositioning: re-investigating existing drugs for new therapeutic indications. *Journal of postgraduate medicine*, 57(2), 153.
- Pandas, 2023. Erişim Tarihi : 29.06.2023. <https://pandas.pydata.org/>
- Pandala, S. R. 2020. Lazy Predict. Erişim Tarihi : 29.06.2023 <https://github.com/shankarpandala/lazypredict>.
- Pantziarka, P., Pirmohamed, M., Mirza, N. 2018. New uses for old drugs. *Bmj*, 361.
- Pardridge, W.M. 1995. Transport of small molecules through the blood-brain barrier: Biology and methodology. *Advanced Drug Delivery Reviews*. 15, 5-36.
- 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.
- Paterson, D.A., Conradi, R.A., Hilgers, A.R., Vidmar, T.J., and Burton, P.S. 1994. A nonaqueous partitioning system for predicting the oral absorption potential of peptides. *Quantitative Structure-Activity Relationship*, 13, 4-10.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Duchesnay, É. 2011. Scikit-learn: Machine learning in Python. *the Journal of machine Learning research*, 12, 2825-2830.
- Pieters, R., Carroll, W. L. 2008. Biology and treatment of acute lymphoblastic leukemia. *Pediatric Clinics of North America*, 55(1), 1-20.
- Pollastri, M. P. 2010. Overview on the Rule of Five. *Current protocols in pharmacology*, 49(1), 9-12.
- Priyadharsini, M., Mallika, S. S., Karuppasamy, 2022. Drug Discovery for Alzheimer's Disease Using Machine Learning. In *2022 8th International Conference on Advanced Computing and Communication Systems (ICACCS)* (Vol. 1, pp. 787-792). IEEE.
- PubMed Veritabanı, 2023. Erişim Tarihi : 18 Haziran 2023. <http://www.ncbi.nlm.nih.gov/pubmed>
- Pypi, 2023. Erişim Tarihi : 10.06.2023. <https://pypi.org>

- Pyvis, 2018. Erişim Tarihi : 15 Haziran 2023. <https://pyvis.readthedocs.io/en/latest/>
- Rees, J. 2023. Cancer in crisis warning over long treatment waits in Wales. BBC News. Erişim Tarihi : 13 Mart 2023. <https://www.bbc.com/news/uk-wales-64888722>
- Rocha, S. M., Nascimento, D., Coelho, R. S., Cardoso, A. M., Passarinha, L. A., Socorro, S., Maia, C. J. 2023. STEAP1 Knockdown Decreases the Sensitivity of Prostate Cancer Cells to Paclitaxel, Docetaxel and Cabazitaxel. *International Journal of Molecular Sciences*, 24(7), 6643.
- Rudrapal, M., Khairnar, S. J., Jadhav, A. G. 2020. Drug repurposing (DR): an emerging approach in drug discovery. *Drug repurposing-hypothesis, molecular aspects and therapeutic applications*, 10.
- RxList, 2023. Erişim Tarihi : 18 Haziran 2023. <http://www.rxlist.com>
- Sada Del Real K., Rubio, A. 2023. Discovering the Mechanism of Action of Drugs with a Novel Sparse Explainable Network. <https://dx.doi.org/10.2139/ssrn.4364890>.
- Sarı, S., Avcı, A., Aslan, E. K. 2021. In silico Repurposing of Drugs for pan-HDAC and pan-SIRT Inhibitors: Consensus Structure-based Virtual Screening and Pharmacophore Modeling Investigations. *Turkish Journal of Pharmaceutical Sciences*, 18(6), 730.
- Sarkar, C., Das, B., Rawat, V. S., Wahlang, J. B., Nongpiur, A., Tiewsoh, I., Sony, H. T. 2023. Artificial Intelligence and Machine Learning Technology Driven Modern Drug Discovery and Development. *International Journal of Molecular Sciences*, 24(3), 2026.
- Scalfani, V. F., Patel, V. D., Fernandez, A. M. 2022. Visualizing chemical space networks with RDKit and NetworkX. *Journal of Cheminformatics*, 14(1), 87.
- Shineman, D. W., Alam, J., Anderson, M., Black, S. E., Carman, A. J., Cummings, J. L., Fillit, H. M. 2014. Overcoming obstacles to repurposing for neurodegenerative disease. *Annals of clinical and translational neurology*, 1(7), 512-518.
- Schmidhuber, J. 2015. Deep learning in neural networks: An overview. *Neural networks*, 61, 85–117. doi:10.1016/j.neunet.2014.09.003
- Schwaller, P., Gaudin, T., Lanyi, D., Bekas, C. Laino, T. 2018. 'Found in translation': Predicting outcomes of complex organic chemistry reactions using neural sequence-to-sequence models. *Chemical Science* 9, 6091–6098.

- Sellereite, N., Jullum, M. 2020. shapr: An R-package for explaining machine learning models with dependence-aware Shapley values, J. Open Source Softw., 5, 2027.
- Senior, A. W., Evans, R., Jumper, J., Kirkpatrick, J., Sifre, L., Green, T., Hassabis, D. 2020. Improved protein structure prediction using potentials from deep learning. Nature, 577(7792), 706-710.
- Shapley, L. S. 1953. A value for n-person games. Contributions to the Theory of Games, 307-317.
- Song, E., Nelson, B. L., Staum, J. 2016. Shapley effects for global sensitivity analysis: Theory and computation. SIAM/ASA Journal on Uncertainty Quantification, 4(1), 1060-1083.
- Sousa-Pimenta, M., Estevinho, L. M., Szopa, A., Basit, M., Khan, K., Armaghan, M., Sharifi-Rad, J. 2023. Chemotherapeutic properties and side-effects associated with the clinical practice of terpene alkaloids: paclitaxel, docetaxel, and cabazitaxel. Frontiers in Pharmacology, 14, 1157306.
- Streamlit, 2023. Erişim Tarihi : 20.05.2023. <https://docs.streamlit.io/>
- Şirin, E. (2019). Python Pandas ile Aynı Anda Birden Fazla Excel Dosyasını Okumak. Erişim Tarihi : 20 Mayıs 2023. <https://www.veribilimiokulu.com/python-pandas-ile-ayni-anda-birden-fazla-excel-dosyasini-okumak/>
- Tang, Y., Zhu, W., Chen, K., Jiang, H. (2006). New technologies in computer-aided drug design: Toward target identification and new chemical entity discovery. Drug discovery today: technologies, 3(3), 307-313.
- Tang, W., Chen, J., Wang, Z., Xie, H., Hong, H. 2018. Deep learning for predicting toxicity of chemicals: a mini review. Journal of environmental science and health. Part C, Environmental carcinogenesis & ecotoxicology reviews, 36(4), 252-271. <https://doi.org/10.1080/10590501.2018.1537563>
- Thotala, D., Karwas, R.M., Engelbach, J.A., Garbow, J.R., Hallahan, A.N., DeWees, T.A., Laszlo, A. and Hallahan, D.E., 2015. Valproic Acid Enhances the Efficacy of Radiation Therapy by Protecting Normal Hippocampal Neurons and Sensitizing Malignant Glioblastoma Cells. Oncotarget, 6(33), 35004-35022.
- Ulusal Kanser Enstitüsü, 2022. Childhood Acute Lymphoblastic Leukemia Treatment. Erişim Tarihi : 29 Haziran 2023. <https://www.cancer.gov/types/leukemia/patient/child-all-treatment-pdq>

- Unterthiner, T., Mayr, A., Klambauer, G., Steijaert, M., Wegner, J. K., Ceulemans, H., Hochreiter, S. 2014. Deep learning as an opportunity in virtual screening. In Proceedings of the deep learning workshop at NIPS, 27, 1-9.
- Uysal, İ., Köse, U. 2022. "İlaç Keşfi ve Yapay Zekâ". Yapay Zekanın Değiştirdiği Dinamikler. Ed. Bilen, M. Eğitim Yayınevi, 1(1), 19-34.
- Uysal, İ., Köse, U. 2023. İlaç-İlaç Etkileşimlerini Keşfetmek: Bir Ağ Analizi ve Görselleştirme Yaklaşımı. Journal of Materials and Mechatronics: A (JournalMM), 4(1), 257-270
- Vaidya, B., Parvathaneni, V., Kulkarni, N. S., Shukla, S. K., Damon, J. K., Sarode, A., Gupta, V. 2019. Cyclodextrin modified erlotinib loaded PLGA nanoparticles for improved therapeutic efficacy against non-small cell lung cancer. International journal of biological macromolecules, 122, 338-347.
- Veber, D. F., Johnson, S. R., Cheng, H. Y., Smith, B. R., Ward, K. W., Kopple, K. D. 2002. Molecular properties that influence the oral bioavailability of drug candidates. Journal of medicinal chemistry, 45(12), 2615-2623. <https://doi.org/10.1021>
- Vega-García, N., Malatesta, R., Estella, C., Pérez-Jaume, S., Esperanza-Cebollada, E., Torreadell, M., Camós, M. 2018. Paediatric patients with acute leukaemia and KMT 2A (MLL) rearrangement show a distinctive expression pattern of histone deacetylases. British Journal of Haematology, 182(4), 542-553.
- Velmurugan, J., Ida, M. 2023. Computational Drug Discovery Using QSAR Modeling. In Recent Trends in Computational Intelligence and Its Application: Proceedings of the 1st International Conference on Recent Trends in Information Technology and its Application (ICRTITA, 22) (p. 438). CRC Press.
- Vemula, D., Jayasurya, P., Sushmitha, V., Kumar, Y. N., Bhandari, V. (2022). CADD, AI and ML in drug discovery: A comprehensive review. European Journal of Pharmaceutical Sciences, 106324.
- Wallach, I., Dzamba, M. Heifets, A. 2015. Atomnet: a deep convolutional neural network for bioactivity prediction in structurebased drug discovery. arXiv preprint arXiv:1510.02855.
- Wang, F., Wu, F. X., Li, C. Z., Jia, C. Y., Su, S. W., Hao, G. F., Yang, G. F. 2019. ACID: a free tool for drug repurposing using consensus inverse docking strategy. Journal of cheminformatics, 11, 1-11.
- 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), 844-853.

- Yalçın, A. S., Yılmaz, A. M., Altundağ, E. M., KoçtürK, S. 2017. Kurkumin, kuersetin ve çay kateşinlerinin anti-kanser etkileri. Marmara Pharmaceutical Journal, 21(1), 19-29.
- 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. doi:10.1021/acs.chemrev.8b00728
- Yang, J., Anishchenko, I., Park, H., Peng, Z., Ovchinnikov, S., Baker, D. 2020. Improved protein structure prediction using predicted interresidue orientations. Proceedings of the National Academy of Sciences, 117(3), 1496-1503.
- Zhang, L., Tan, J., Han, D., & Zhu, H. 2017. From machine learning to deep learning: progress in machine intelligence for rational drug discovery. Drug discovery today, 22(11), 1680–1685. <https://doi.org/10.1016/j.drudis.2017.08.010>
- Zhao, F., Liu, N., Zhan, P., Liu, X. 2015. Repurposing of HDAC inhibitors toward anti-hepatitis C virus drug discovery: teaching an old dog new tricks. Future Medicinal Chemistry, 7(11), 1367-1371.
- Zhavoronkov, A., Ivanenkov, Y. A., Aliper, A., Veselov, M. S., Aladinskiy, V. A., Aladinskaya, A. V., Aspuru-Guzik, A. 2019. Deep learning enables rapid identification of potent DDR1 kinase inhibitors. Nature biotechnology, 37(9), 1038-1040.

EKLER

EK A. Diğer Performans Metrikleri - 1

EK B. Diğer Performans Metrikleri - 2



EK A. Diğer Performans Metrikleri – 1

Metric	Value
95% CI	(0.98459,0.99147)
ACC Macro	98.803
ARI	94.022
AUNP	9.725
AUNU	9.725
Bangdiwala B	98.367
Bennett S	97.605
CBA	96.973
CSI	95.615
Chi-Squared	351.193.482
Chi-Squared DF	1
Conditional Entropy	8.137
Cramer V	95.608
Cross Entropy	64.508
F1 Macro	97.797
F1 Micro	98.803
FNR Macro	275
FNR Micro	1.197
FPR Macro	275
FPR Micro	1.197
Gwet AC1	98.356
Hamming Loss	1.197
Joint Entropy	72.633
KL Divergence	12
Kappa	95.594
Kappa 95% CI	(0.94328,0.9686)
Kappa No Prevalence	97.605
Kappa Standard Error	646
Kappa Unbiased	95.594
Krippendorff Alpha	95.594
Lambda A	92.722
Lambda B	92.508
Mutual Information	55.249
NIR	8.355
NPV Macro	98.364
NPV Micro	98.803
Overall ACC	98.803
Overall CEN	8.178
Overall J	(1.9146,0.9573)

Overall MCC	95.608
Overall MCEN	7.019
Overall RACC	72.827
Overall RACCU	72.828
P-Value	None
PPV Macro	98.364
PPV Micro	98.803
Pearson C	69.106
Phi-Squared	91.409
RCI	85.663
RR	1921.0
Reference Entropy	64.496
Response Entropy	63.386
SOA1(Landis & Koch)	Almost Perfect
SOA2(Fleiss)	Excellent
SOA3(Altman)	Very Good
SOA4(Cicchetti)	Excellent
SOA5(Cramer)	Very Strong
SOA6(Matthews)	Very Strong
SOA7(Lambda A)	Very Strong
SOA8(Lambda B)	Very Strong
SOA9(Krippendorff Alpha)	High
SOA10(Pearson C)	Strong
Scott PI	95.594
Standard Error	175
TNR Macro	9.725
TNR Micro	98.803
TPR Macro	9.725
TPR Micro	98.803
Zero-one Loss	46

EK A. Diğer Performans Metrikleri - 2

Metric	Class 0	Class 1
ACC(Accuracy)	98.803	98.803
AGF(Adjusted F-score)	97.283	98.295
AGM(Adjusted geometric mean)	98.288	969
AM(Difference between automatic and manual classification)	-18	18
AUC(Area under the ROC curve)	9.725	9.725
AUCI(AUC value interpretation)	Excellent	Excellent
AUPR(Area under the PR curve)	96.328	99.286
BB(Braun-Blanquet similarity)	94.937	99.009
BCD(Bray-Curtis dissimilarity)	234	234
BM(Informedness or bookmaker informedness)	94.501	94.501
CEN(Confusion entropy)	20.844	5.727
DOR(Diagnostic odds ratio)	428.035.714	428.035.714
DP(Discriminant power)	200.214	200.214
DPI(Discriminant power interpretation)	Fair	Fair
ERR(Error rate)	1.197	1.197
F0.5(F0.5 score)	9.715	99.119
F1(F1 score - harmonic mean of precision and sensitivity)	96.308	99.285
F2(F2 score)	95.481	99.452
FDR(False discovery rate)	228	991
FN(False negative/miss/type 2 error)	32	14
FNR(Miss rate or false negative rate)	5.063	436
FOR(False omission rate)	991	228
FP(False positive/type 1 error/false alarm)	14	32
FPR(Fall-out or false positive rate)	436	5.063
G(G-measure geometric mean of precision and sensitivity)	96.318	99.286
GI(Gini index)	94.501	94.501
GM(G-mean geometric mean of specificity and sensitivity)	97.223	97.223
HD(Hamming distance)	46	46
IBA(Index of balanced accuracy)	90.149	98.896
ICSI(Individual classification success index)	92.657	98.573
IS(Information score)	257.058	24.491
J(Jaccard index)	92.879	98.581
LS(Lift score)	5.9405	118.502
MCC(Matthews correlation coefficient)	95.608	95.608
MCCI(Matthews correlation coefficient interpretation)	Very Strong	Very Strong

MCEN(Modified confusion entropy)	33.456	9.969
MK(Markedness)	96.729	0.

