T.C.

YILDIZ TEKNİK ÜNİVERSİTESİ

FEN BİLİMLERİ ENSTİTÜSÜ

CRISPR-CAS13D İLE HEDEF DIŞI ETKİLERİN DERİN ÖĞRENMEYLE MODELLENMESİ: DEEPCAS13 MODELİNİN YAPISAL gRNA VERİSİYLE İYİLEŞTİRİLMESİ

Buket BARIŞ

YÜKSEK LİSANS TEZİ

Moleküler Biyoloji ve Genetik Anabilim Dalı

Moleküler Biyoloji ve Genetik Programı

Danışman

Doç. Dr. Alper YILMAZ

Kasım, 2025

T.C.

YILDIZ TEKNİK ÜNİVERSİTESİ

FEN BİLİMLERİ ENSTİTÜSÜ

CRISPR-CAS13D İLE HEDEF DIŞI ETKİLERİN DERİN ÖĞRENMEYLE MODELLENMESİ: DEEPCAS13 MODELİNİN YAPISAL gRNA VERİSİYLE İYİLEŞTİRİLMESİ

Buket BARIŞ tarafından hazırlanan tez çalışması 19.11.2025 tarihinde aşağıdaki jüri tarafından Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü Moleküler Biyoloji ve Genetik Anabilim Dalı, Moleküler Biyoloji ve Genetik Programı **YÜKSEK LİSANS TEZİ** olarak kabul edilmiştir.

Doç. Dr. Alper YILMAZ
Yıldız Teknik Üniversitesi
Danışman

Jüri Üyeleri

Doç. Dr. Alper YILMAZ, Danışman
Yıldız Teknik Üniversitesi _____

Prof. Dr. Nehir ÖZDEMİR ÖZGENTÜRK, Üye
Yıldız Teknik Üniversitesi _____

Doç. Dr. Sezen CANIM ATEŞ, Üye
İstanbul Yeni Yüzyıl Üniversitesi _____

Danışmanım Doç. Dr. Alper YILMAZ sorumluluğunda tarafımca hazırlanan "CRISPR-Cas13d İle Hedef Dışı Etkilerin Derin Öğrenmeyle Modellenmesi: DeepCas13 Modelinin Yapısal gRNA Verisiyle İyileştirilmesi" başlıklı çalışmada veri toplama ve veri kullanımında gerekli yasal izinleri aldığımı, diğer kaynaklardan aldığım bilgileri ana metin ve referanslarda eksiksiz gösterdiğimi, araştırma verilerine ve sonuçlarına ilişkin çarpıtma ve/veya sahtecilik yapmadığımı, çalışmam süresince bilimsel araştırma ve etik ilkelerine uygun davrandığımı beyan ederim. Beyanımın aksinin ispatı halinde her türlü yasal sonucu kabul ederim.

Buket BARIŞ

İmza

TEŞEKKÜR

Bu tez çalışmasının her aşamasında bilimsel rehberliği, kıymetli katkıları ve sabırlı desteğiyle bana yol gösteren değerli danışmanım Doç. Dr. Alper YILMAZ'a içten teşekkürlerimi sunuyorum.

Her koşulda yanımda olan, sevgileriyle bana güç veren ve en zor zamanlarda bile inançlarını yitirmeden beni destekleyen sevgili anneme ve babama en derin minnettarlığımı sunuyorum.

Buket BARIŞ

İÇİNDEKİLER

KISALTMA LİSTESİ ix

ŞEKİL LİSTESİ x

TABLO LİSTESİ xi

ÖZET xii

ABSTRACT xiv

1 GİRİŞ 1

1.1 Problemin Tanımı ve Araştırmanın Gerekçesi 1

1.2 Tezin Amacı 3

1.3 Hipotez 3

2 CRISPRCAS SİSTEMİ VE ETKİ TAHMİNİ 5

2.1 CRISPR Teknolojisi 5

2.2 CRISPR-Cas Sisteminin Sınıflandırılması 7

2.3 Hedef Dışı Etkilerin Önemi ve Zorlukları 12

2.3.1 HEPN Alanlarının Yanal RNaz Aktivitesi……..……...13

2.3.2 crRNA ve Hedef RNA Arasındaki Uyumsuzluklar……..13

2.3.3 Dizi Benzerliği ve Homoloji.……14

2.3.4 Hedef RNA İkincil Yapısı..……...……14

2.3.5 Cas13d ve Hedef RNA'ların Ekspresyon Seviyeleri…..…..……14

2.4 Hedef Dışı Etki Tespitinin Büyük Önemi 15

2.5 CRISPR-Cas13d Hedef Dışı Aktivitesini Etkileyen Faktörler 17

2.5.1 Rehber RNA Tasarımı ve Dizi Tamamlayıcılığı……...17

2.5.2 Hedef RNA İkincil Yapısı ve Erişilebilirliği……17

2.6 Hedef Dışı Etkilerin Tespit ve Tahmin Stratejileri 18

2.6.1 Deneysel Metodolojiler……18

2.6.2 Hesaplamalı ve Makine Öğrenimi Yaklaşımları……..……19

2.6.2.1 Cas9 Sistemleri İçin Geliştirilen Modeler..……19

2.6.2.2 Cas12 Sistemlerine Yönelik Yaklaşımlar……..……20

2.6.2.3 Cas13 Sistemleri ve Güncel Yaklaşımlar……..……20

2.7 Grna Sekansinin Temek Bileşenleri ve Fonksiyonu 21

2.8 İkincil Yapının Tanımı ve Biyolojik Önemi 22

2.9 Tersiyer Yapı (3D Konformasyon) 23

3 DERİN ÖĞRENME STRATEJİSİ **25**

3.1 Biyolojik Veri Analizinde Derin Öğrenme 25

3.1.1 Biyolojik Dizilerin Vektör Temsilleri……..……26

3.1.2 3D Yapıların Graf Gömme Yöntemleri……27

3.2 CRISPR Sistemlerinde Hedef Dışı Etki Tespiti için Derin Öğrenme..... 27

3.2.1 Cas9 Sistemleri için Derin Öğrenme Modelleri……28

3.2.2 Cas12 Sistemleri için Derin Öğrenme Modelleri..……28

3.2.3 Cas13 Sistemleri için Derin Öğrenme Modelleri……..……29

3.2.4 Cas13 Sistemleri için Derin Öğrenme Yaklaşımlarının Eksiklikleri……29

4 MATERYAL METOD **31**

4.1 Mevcut DeepCas13 Modeli 31

4.2 Girdiler ve Çıktılar 32

4.2.1 Girdiler……..32

4.2.2 Çıktılar…….……..33

4.3 Sınırlamalar 33

4.3.1 Sekans ve İkincil Yapı ile Sınırlı Girdi……33

4.3.2 Kollateral Aktivite ce Hedef Dışı Etkilerin Tam Olarak Modellenememe……34

4.4 Model Geliştirme Yaklaşımı 34

4.5 Veri Seti Toplama ve Oluşturma 35

4.5.1 Eğitim ve Test Veri Seti……35

4.5.2 In Siliko Optimizasyon Veri Seti……35

4.6 3D Yapısal Özellik Çıkarımı İş Akışı 36

4.6.1 RNA Yapılarının Graf Temsiline Dönüştürülmesi (RNAglib)……37

4.6.2 Yüksek Boyutlu Özelliklerin İndirgenmesi (UMAP)....……38

4.7 Geliştirilmiş Model Mimarisi..... 39

4.8 Model Eğitimi ve Değerlendirme 39

5 SONUÇ **41**

5.1 Çalışmanın Klinik Önemi ve Literatür Karşılaştırması 41

5.1.1 Metodolojik Yaklaşım ve Performans Değerlendirmesi…..……41

5.1.2 Geliştirilen Modelin Performans Değerlendirmesi……..……42

5.1.2.1 gRNA Verimlilik Skorlarının Tahmini (Regrasyon Başarısı)……42

5.1.2.2 Hedef Dışı Etkilerin Sınıflandırılması (Sınıflandırma)……..…43

5.2 Mevcut gRNA Dizileri için Verimlilik Skorlarının Tahmini..... 47

5.2.1 Yenilikçi Tahmin Metodolojisi ve Zenginleştirilmiş Veri Seti.....47

5.2.2 Tahmin Sonuçları ve DeepScore Analizi…….……..…48

5.2.3 En Yüksek Verimlilik Skoruna Sahip Adayın Belirlenmesi……..…49

5.3 Hedef Dizi için Yeni gRNA Tasarımı..... 50

5.3.1 Tasarım Metodolojisi: Kayan Pencere ve Yapısal Skorlama…...…50

5.3.2 Tasarım Sonuçlarının Analizi: En Verimli Adayın Tespiti.....……51

5.4 Kısıtlılıklar ve Gelecek Perspektif..... 52

5.4.1 Mevcut Kısıtlılıklar……...52

5.4.2 Gelecek Yönelimler……...53

KAYNAKÇA **54**

TEZDEN ÜRETİLMİŞ YAYINLAR **63**

KISALTMA LİSTESİ

AUC	ROC Eğrisi Altında Kalan Alan
AUPR	Hassasiyet-Duyarlılık Eğrisi Altında Kalan Alan
CNN	Evrişimsel Sinir Ağı
CRISPR	Kümelenmiş Düzenli Aralıklı Kısa Palindromik Tekrarlar
CRNN	Evrişimsel Tekrarlayan Sinir Ağı
crRNA	CRISPR RNA
DNA	Deoksiribonükleik Asit
DSB	Çift Zincir Kırığı
GNN	Graf Sinir Ağları
gRNA	Kılavuz RNA
HEPN	Yüksek Ökaryot ve Prokaryot Nükleotit Bağlama
LFC	Logaritmik Kat Değişimi
LLM	Büyük Dil Modeli
MCC	Matthews Korelasyon Katsayısı
PAM	Protospacer Bitişik Motif
PCC	Pearson Korelasyon Katsayısı
PFS	Protospacer Yan Dizisi
RNA	Ribonükleik Asit
RNN	Tekrarlayan Sinir Ağı
RNP	Ribonükleoprotein
t-SNE	t-Dağıtımlı Stokastik Komşu Yerleştirme
WYL	Triptofan-Tirozin-Lösin
YSA	Yapay sinir ağları

ŞEKİL LİSTESİ

Şekil 2.1 Gen düzenleme ve genetik mühendisliği alanındaki gelişmeler **5**
Şekil 2.2 CRISPR-Cas mekanizması **6**
Şekil 2.3 CRISPR-Cas13 mekanizması **7**
Şekil 2.4 Cas proteinlerinin sınıflandırılması **8**
Şekil 2.5 CRISPR-Cas9 sisteminin çalışma mekanizması ve DNA kesimi **8**
Şekil 2.6 CRISPR-Cas12 sisteminin çalışma mekanizması ve DNA kesimi **9**
Şekil 2.7 Cas13 sisteminin çalışma mekanizması ve RNA kesimi **10**
Şekil 2.8 Cas13 alt tiplerinin ortalama protein büyüklükleri **12**
Şekil 2.9 Hedef dışı etkilerin gösterimi **16**
Şekil 2.10 CRISPR-Cas9 hedef dışı etki tahmini için hibrit derin öğrenme mimarisi **19**
Şekil 2.11 CRISPR-Cas12 hedef dışı etki tahmini için derin öğrenme mimarisi **20**
Şekil 4.1 DeepCas13 model gösterimi **32**
Şekil 4.2 Geliştirilen model mimarisi **35**
Şekil 4.3 6e9f moleküler yapısı **36**
Şekil 4.4 6e9f içinden alınan gRNA 3 boyutlu yapısı **36**
Şekil 4.5 RNAglib görev şeması **37**
Şekil 4.6 Bir RNA yapısının rnaglib 2.5D graf temsili **38**
Şekil 4.7 UMAP algoritmasının topolojik çalışma prensibi **39**
Şekil 5.1 Geliştirilen modelin pearson korelasyonu **43**
Şekil 5.2 DeepCas13 modelinin pearson korelasyonu **43**
Şekil 5.3 Geliştirilen modelin AUC değeri **44**
Şekil 5.4 DeepCas13 modelinin AUC değeri **45**
Şekil 5.5 Geliştirilen modelin AUPR değeri **46**
Şekil 5.6 DeepCas13 modelinin AUPR değeri **46**
Şekil 5.7 UMAP verilerinin görselleştirilmesi **48**

TABLO LİSTESİ

Tablo 5.1 Verimlilik tahmini**.....49**

Tablo 5.2 Verimlilik tahmini sonuçlarının değerlendirilmesi sonucu DeepScore tespiti**.....50**

Tablo 5.3 Tasarlanan gRNA aday havuzundan elde edilen örnek sonuçlar **.....51**

Tablo 5.4 Yeni gRNA tasarımı sonucu elde edilen verimli dizilerHata! Yer işareti tanımlanmamış. **.....51**

ÖZET

CRISPR-Cas13d ile Hedef Dışı Etkilerin Derin Öğrenmeyle Modellenmesi: DeepCas13 Modelinin Yapısal gRNA Verisiyle İyileştirilmesi

Buket BARIŞ

Moleküler Biyoloji ve Genetik Anabilim Dalı
Yüksek Lisans Tezi

Danışman: Doç. Dr. Alper YILMAZ

CRISPR-Cas13d sistemi, RNA düzenleme alanında önemli bir potansiyele sahip olmakla birlikte, bu sistemin klinik uygulamalardaki güvenilirliği, hedef dışı etkilerin minimize edilmesine bağlıdır. Mevcut hesaplamalı modeller (örn. DeepCas13), gRNA tasarımını çoğunlukla 1-boyutlu (sekans) ve 2-boyutlu (ikincil yapı) bilgilere dayanarak yapmakta, hedef spesifikliği için kritik olan 3-boyutlu (3D) konformasyonel yapıyı göz ardı etmektedir.

Bu çalışma, DeepCas13 modelini, 3D yapısal bilgiyi entegre eden çoklu-girdi (multi-modal) bir yaklaşımla iyileştirmeyi amaçlamıştır. Bu doğrultuda, RNAglib (graf temsili) ve UMAP (boyut indirgeme) tekniklerine dayalı yenilikçi bir 3D özellik çıkarım işlem hattı oluşturulmuştur. Geliştirilen bu işlem hattı, literatürden elde edilen 5.700 gRNA'lık 'Eğitim ve Test Veri Seti'ndeki her bir gRNA için 32 boyutlu yapısal vektörler üretmek ve bu verileri modele entegre etmek amacıyla kullanılmıştır.

Geliştirilen model, 5-katlı çapraz doğrulama yöntemiyle eğitilmiş ve değerlendirilmiştir. Sonuçlar, 3D yapısal bilginin entegrasyonunun, temel

DeepCas13 modeline (PCC: -0.3792, AUC: 0.7238, AUPR: 0.7063) kıyasla belirgin bir performans artışı sağladığını göstermiştir. Geliştirilen modelin, test (doğrulama) verisi üzerindeki ortalama performansı; Pearson Korelasyon Katsayısı (PCC) -0.7984, ROC Eğrisi Altında Kalan Alan (AUC) 0.9549 ve Hassasiyet-Duyarlılık Eğrisi Altında Kalan Alan (AUPR) 0.8269 olarak hesaplanmıştır. Yüksek AUC/AUPR skorları, modelin 'verimli' ve 'verimsiz' gRNA'ları ayırt etmede mükemmele yakın bir başarı sergilediğini doğrulamaktadır.

Ayrıca modelin pratik uygulanabilirliği, 10.250 adet gRNA yapısından oluşan mutasyonlu validasyon seti (in silico validasyon seti) üzerinde test edilmiştir. Model, bu geniş aday havuzu içerisinden en verimli adayı (DeepScore: 0.995) başarıyla seçerek ve yeni hedefler için "de novo" (yeni) yüksek skorlu gRNA'lar (DeepScore: 0.950) tasarlayarak başarısını kanıtlamıştır. Bu sonuçlar, 3D yapısal bilginin entegrasyonunun, CRISPR-Cas13d sisteminin hedef spesifikliğini artırmak ve potansiyel hedef dışı etkileri minimize etmek için kritik bir yaklaşım sunduğunu ortaya koymaktadır.

Anahtar Kelimeler: CRISPR-Cas13d, hedef dışı etkiler, derin öğrenme, DeepCas13, gRNA'nın üç boyutlu yapısı.

ABSTRACT

Modeling Off-Target Effects of CRISPR-Cas13d via Deep Learning: Refining DeepCas13 with Structural gRNA Data

Buket Barış

Department of Molecular Biology and Genetics
Master of Science Thesis

Supervisor: Assoc.Prof. Dr. Alper YILMAZ

Although the CRISPR-Cas13d system holds significant potential in the field of RNA editing, the reliability of this system in clinical applications depends on the minimization of off-target effects. Current computational models (e.g., DeepCas13) largely rely on 1-dimensional (sequence) and 2-dimensional (secondary structure) information for gRNA design, overlooking the 3-dimensional (3D) conformational structure which is critical for target specificity.

This study aimed to enhance the DeepCas13 model through a multi-modal approach that integrates 3D structural information. To this end, an innovative 3D feature extraction pipeline based on RNAglib (graph representation) and UMAP (dimensionality reduction) techniques was developed. This pipeline was utilized to generate 32-dimensional structural vectors for each gRNA within the 'Training and Test Dataset' comprising 5,700 gRNAs obtained from the literature, and to integrate these data into the model.

The developed model was trained and evaluated using a 5-fold cross-validation method. The results demonstrated that the integration of 3D structural information

yielded a significant performance improvement compared to the baseline DeepCas13 model (PCC: -0.3792, AUC: 0.7238, AUPR: 0.7063). The average performance of the developed model on the test (validation) data was calculated as follows: Pearson Correlation Coefficient (PCC) of -0.7984, Area Under the ROC Curve (AUC) of 0.9549, and Area Under the Precision-Recall Curve (AUPR) of 0.8269. High AUC/AUPR scores confirm that the model exhibits near-perfect success in distinguishing between 'efficient' and 'inefficient' gRNAs.

Furthermore, the practical applicability of the model was tested on a mutated validation set (in silico validation set) comprising 10,250 gRNA structures. The model demonstrated its success by effectively selecting the most efficient candidate (DeepScore: 0.995) from this extensive candidate pool and by designing "de novo" high-scoring gRNAs (DeepScore: 0.950) for new targets. These findings indicate that the integration of 3D structural information offers a critical approach for enhancing the target specificity of the CRISPR-Cas13d system and minimizing potential off-target effects.

Keywords: CRISPR-Cas13d, off-target effects, deep learning, DeepCas13, three-dimensional structure of gRNA.

1
GİRİŞ

1.1 Problemin Tanımı ve Araştırmanın Gerekçeşi

Gen düzenleme teknolojileri, son yıllarda moleküler biyoloji ve genetik alanlarında çığır açan ilerlemeler kaydetmiştir. Bu teknolojiler arasında CRISPR sistemi, DNA ve RNA düzenleme çalışmalarında devrim yaratmıştır. CRISPR-Cas sistemi, bakteri hücresinin yabancı genetik materyali tanıyıp yok edebildiği veya faj enfeksiyonuna karşı direnç gösterebildiği bir adaptif bağışıklık stratejisidir. Bu sistem, bakterinin çevresindeki virüsleri tanıyıp bunlara karşı savunma geliştirebilmesini sağlar. CRISPR sistemleri, başlangıçta prokaryotlarda adaptif bağışıklık sistemi olarak keşfedilmiş olsa da günümüzde gen düzenleme çalışmalarında yaygın olarak kullanılmaktadır [1]. Bakteri hücresinin yabancı genetik materyali tanıyıp yok edebildiği veya faj enfeksiyonuna karşı direnç gösterebildiği bir adaptif bağışıklık stratejisidir. Bu sistem, bakterinin çevresindeki virüsleri tanıyıp bunlara karşı savunma geliştirebilmesini sağlar. CRISPR sistemleri içinde en çok bilineni Cas9 proteini DNA'yı hedeflerken, Cas13 ailesi RNA moleküllerini hedefleyebilme yeteneğiyle öne çıkmaktadır.

Cas13d, Cas13 ailesinin en küçük üyesi olmasına rağmen, RNA düzenleme kapasitesi bakımından oldukça etkilidir [2]. Özellikle kompakt yapısı, viral vektörlerle taşınabilmesi ve RNA'ya özgü kesim aktivitesi sayesinde, gen ekspresyonunun düzenlenmesinde ve RNA seviyesinde yapılan müdahalelerde tercih edilmektedir. Cas13d'nin bir diğer avantajı, RNA'yı hedeflemek için kullanılan gRNA (single guide RNA) tasarımında esneklik sağlamasıdır.

Ancak, CRISPR-Cas13d sisteminin sağladığı bu avantajlara rağmen, hedef dışı etkiler önemli bir sorun olarak karşımıza çıkmaktadır. Hedef dışı etkiler, gRNA'nın hedef RNA dizisi dışındaki RNA dizilerine bağlanması sonucu ortaya çıkan

istenmeyen kesimlerdir. Bu durum, özellikle klinik uygulamalarda ciddi yan etkilere neden olabilmektedir [3]. Bu nedenle, CRISPR-Cas13d sisteminin hedef dışı etkilerinin minimize edilmesi, sistemin güvenilirliğini ve etkinliğini artırmak açısından kritik öneme sahiptir.

Literatürde, CRISPR sistemlerinin hedef dışı etkilerini azaltmak için çeşitli yaklaşımlar önerilmiştir. Bunlar arasında gRNA tasarımının optimizasyonu, Cas enzimlerinin mühendisliği ve hesaplamalı yöntemlerin kullanılması yer almaktadır [4]. Son yıllarda, derin öğrenme tabanlı yaklaşımlar, CRISPR sistemlerinin hedef dışı etkilerini tahmin etmek ve azaltmak için umut verici sonuçlar sunmaktadır.

Örneğin, SpCas9 aktivitesini öngörmek için geliştirilen DeepCRISPR, geniş çaplı veri setleri üzerinde eğitilerek hem hedef içi hem de hedef dışı etkileri yüksek başarıyla tahmin etmiştir [5]. Benzer şekilde, DeepCpf1 modeli, Cas12a (Cpf1) sistemleri için kromatin erişilebilirliği bilgisini de kullanarak DNA üzerindeki kesim verimliliğini optimize etmiştir [6]. Ancak bu modeller, DNA'nın çift sarmallı ve nispeten kararlı yapısı ile kromatin organizasyonu üzerine kurgulanmıştır. RNA'nın tek sarmallı doğası, dinamik ikincil yapıları ve hücre içindeki 3 boyutlu katlanma biçimleri, DNA tabanlı bu algoritmaların RNA hedeflemesinde yetersiz kalmasına neden olmaktadır.

DeepCas13 ise, CRISPR-Cas13 sisteminin hedef spesifikliğini artırmak için geliştirilen derin öğrenme tabanlı bir modeldir [7]. Bu model, gRNA dizisi ve hedef RNA dizisi arasındaki etkileşimleri analiz ederek, hedef dışı etkilerin tahmin edilmesini sağlamaktadır. Ancak, mevcut modeller genellikle gRNA'nın birincil ve ikincil dizisine odaklanmakta olup gRNA'nın üç boyutlu yapısını yeterince dikkate almamaktadır.

RNA moleküllerinin fonksiyonu, sadece dizi bilgisine değil, aynı zamanda üç boyutlu yapılarına da bağlıdır. RNA'nın üç boyutlu yapısı, RNA-protein etkileşimlerini, RNA'nın kararlılığını ve fonksiyonel aktivitesini doğrudan etkilemektedir [8]. Bu nedenle, gRNA'nın üç boyutlu yapısının CRISPR-Cas13d sisteminin hedef spesifikliğine etkisinin incelenmesi, sistemin etkinliğinin artırılması açısından önemlidir.

Bu çalışmada, CRISPR-Cas13d sisteminin hedef dışı etkilerinin azaltılması için gRNA'nın üç boyutlu yapısının derin öğrenme modellerine entegre edilmesi

amaçlanmıştır. Bu amaçla, DeepCas13 modeli temel alınarak, gRNA'nın üç boyutlu yapısal özellikleri modele dahil edilerek modelin performansı değerlendirilecektir. Çalışmanın sonuçları, gRNA'nın üç boyutlu yapısının dikkate alınmasının, CRISPR-Cas13d sisteminin hedef spesifikliğini artırmadaki etkisi gözlemlenecektir.

1.2 Tezin Amacı

Çalışmanın amacı, CRISPR-Cas13d sisteminin RNA düzenleme süreçlerinde gözlemlenen hedef dışı etkilerin sistematik olarak tespit edilmesi ve bu istenmeyen etkilerin minimize edilmesine yönelik yenilikçi metodolojik yaklaşımlar geliştirmektir.

Araştırma kapsamında, gRNA'nın üç boyutlu yapısal karakteristiklerinin derin öğrenme algoritmaları ile entegre edilmesi yoluyla mevcut DeepCas13 modelinin prediktif performansının artırılması hedeflenmektedir.

Bu bağlamda, gRNA'nın birincil, ikincil ve tersiyer yapısal özelliklerinin kapsamlı bioinformatik analizlerle sistematik olarak değerlendirilmesi ve bu moleküler özelliklerin gelişmiş derin öğrenme mimarilerine entegre edilmesi suretiyle, geleneksel sekans tabanlı tahmin metodolojilerinin sınırlılıklarını aşan innovatif bir yaklaşım geliştirmek amaçlanmıştır.

Çalışmanın metodolojik katkısı, derin öğrenme teknolojileri ile RNA yapısal biyolojisi disiplinlerini birleştirerek CRISPR sistemlerinin spesifiklik ve hassasiyetini artıran multidisipliner bir çerçeve sunmaktır. Bu yaklaşımın, gen düzenleme teknolojilerinde hedef spesifikliğinin moleküler temellerinin derinlemesine anlaşılmasına önemli katkılar sağlaması ve yapay zeka ile moleküler biyoloji alanları arasındaki sinerjik potansiyelin klinik uygulamalara yansıtılmasını mümkün kılması beklenmektedir.

1.3 Alt Başlık

Bu çalışmanın temel hipotezi, gRNA'nın üç boyutlu konformasyonel karakteristiklerinin derin öğrenme mimarilerine sistematik entegrasyonunun, konvansiyonel sekans tabanlı metodolojik yaklaşımlara kıyasla CRISPR-Cas13d

sisteminin hedef dışı etkilerinin tahmininde anlamlı bir performans optimizasyonu sağlayacağıdır.

Bu hipotez, RNA moleküllerinin fonksiyonel aktivitesinin yalnızca primer nükleotid dizilimine değil, aynı zamanda tersiyer konformasyonel yapılandırmalarına da kritik ölçüde bağımlı olduğu temel moleküler biyoloji prensibine dayandırılmıştır.

gRNA'nın moleküler kıvrım paternleri ve uzamsal konfigürasyonunun, hedef tanıma mekanizmalarında birincil dizi homolojisinden daha determinant bir faktör olacağı ve hedef dışı etkileşimlerin önemli bir kısmını açıklayabilme potansiyeline sahip olduğu varsayılmaktadır.

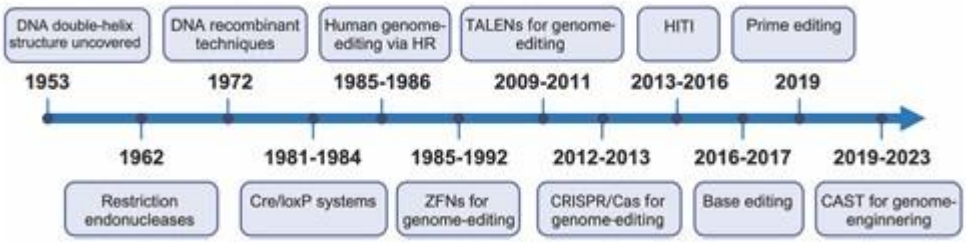
Ayrıca, RNA'nın üç boyutlu yapısal dinamiklerinin, Cas13d proteini ile gerçekleşen moleküler etkileşim geometrisini doğrudan modüle ederek yanlış hibridizasyon olasılıklarını belirleyeceği ve gRNA'nın konformasyonel esnekliğinin hedef RNA ile etkileşim kinetiklerini etkileyerek termodinamik stabilite ve katalitik verimlilik parametrelerini şekillendireceği öngörülmektedir.

Bu hipotez doğrultusunda, DeepCas13 modeline gRNA'nın üç boyutlu yapısal özelliklerinin entegrasyonu ile prediktif doğruluk, spesifiklik ve hassasiyet ölçümlerinde substansiyel iyileşmeler elde edilmesi hedeflenmektedir.

2
CRISPR-CAS SİSTEMİ VE ETKİ TAHMİNİ

2.1 CRISPR Teknolojisi

CRISPR, bilim dünyasında gen düzenleme konusunda devrim niteliğinde bir buluştur. Bu teknoloji, genom üzerinde hassas değişiklikler yapma potansiyeli sunarak birçok hastalığın tedavi edilmesi için umut vermektedir.



Şekil 2.1 Gen düzenleme ve genetik mühendisliği alanındaki gelişmeler [7]

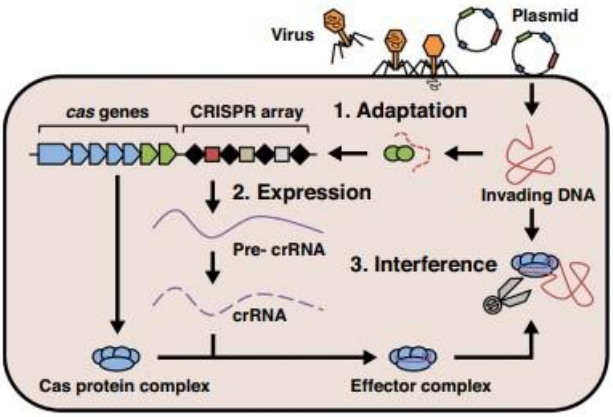
Gen düzenleme çalışmaları, restriksiyon enzimlerinin keşfiyle başladı. Bu enzimler, DNA'nın belirli bölgelerini tanımlayıp kesme yeteneğine sahipti. Bakteriyofaj lamda gibi sistemlerde, DNA'yı keserek işlevlerini gösterdiler. Ancak insan genomunda benzer bir kesme işlemi yapmak kontrol edilemeyen sonuçlar doğurabiliyordu.

Sonrasında, zinc finger proteinleri bulundu. Bu proteinler, DNA'yı tutan ve restriksiyon enzimleriyle birleştirilerek spesifik bölgeleri tanımlayabilen yapılar olarak dikkat çekti. Ancak, bu yöntemdeki kısıtlamalar nedeniyle tam olarak istenilen sonuçları elde etmek mümkün olmadı.

Talenlerin keşfi, gen düzenleme çalışmalarını daha da ileri taşıdı. Talenler, tek nükleotid üzerinde değişiklik yapabilen ve bu sayede daha hassas bir şekilde

genomu düzenleme potansiyeli sunan yapılar olarak ön plana çıktı. Ancak, 2012 yılında CRISPR'nin keşfiyle birlikte gen düzenleme çalışmalarında devrim niteliğinde bir adım atıldı (Şekil 2.1) [9, 10].

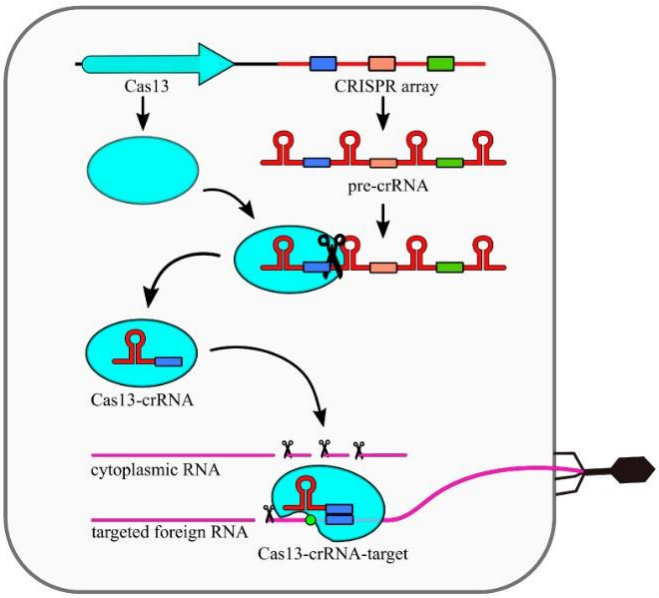
CRISPR, bakterilerde doğal olarak bulunan bir savunma mekanizmasıdır ve DNA'yı hedefleyerek kesme yeteneğine sahiptir. Bakteriler de insanlar gibi virüsler tarafından saldırıya uğrayabilirler. Bu virüsler, kendi DNA'larını bakteri hücresine enjekte ederek, bakteriyi enfekte etme ve çoğalma amacı güderler. Bakteri, virüsün saldırısına karşı kendini savunmak zorundadır çünkü virüslerin çoğalması için bakteriyi kullanır ve eğer bakteri kendini koruyamazsa, virüslerin yayılması sonucunda bakteri yok olur. Bu savunma mekanizması, bakterilerin kullandığı CRISPR-Cas sistemiyle gerçekleşir.



Şekil 2.2 CRISPR-Cas genel mekanizması [10]

CRISPR-Cas sistemi, bakteri hücresinin yabancı genetik materyali tanıyıp yok edebildiği veya faj enfeksiyonuna karşı direnç gösterebildiği bir adaptif bağışıklık stratejisidir. Bu sistem, bakterinin çevresindeki virüsleri tanıyıp bunlara karşı savunma geliştirebilmesini sağlar (Şekil 2.2).

Klasik CRISPR sistemleri (örneğin Cas9) birincil olarak DNA'yı hedefleyip çift zincirli kırıklar oluştururken, son yıllarda yapılan keşifler bu sistemlerin çeşitliliğini ortaya koymuştur. Özellikle Tip VI (Cas13) sistemleri, DNA yerine tek zincirli RNA moleküllerini hedeflemesiyle diğerlerinden ayrılır (Şekil 2.3) [11,12].



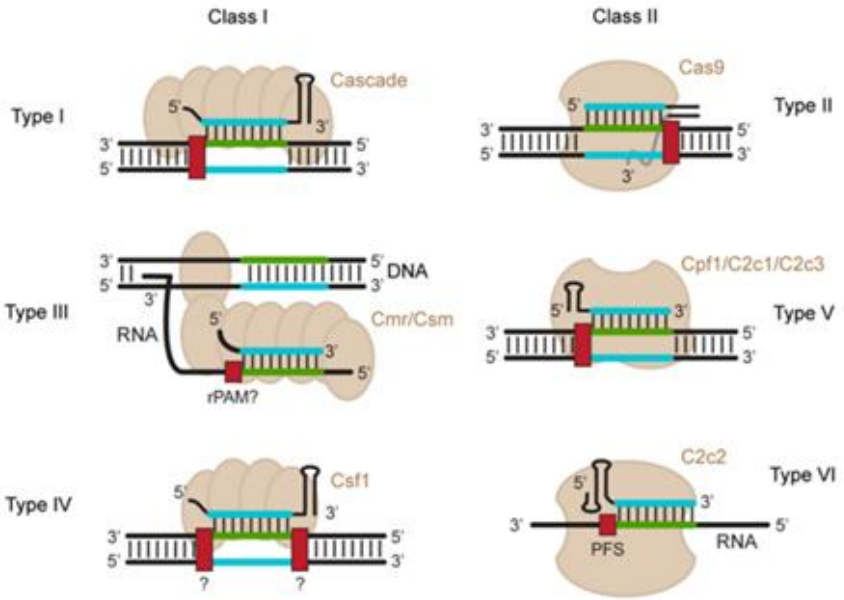
Şekil 2.3 CRISPR-Cas13 mekanizması [13]

2.2 CRISPR-Cas Sisteminin Sınıflandırılması

CRISPR-Cas sistemleri yapısal ve işlevsel özelliklerine göre iki ana sınıfa ayrılmaktadır:

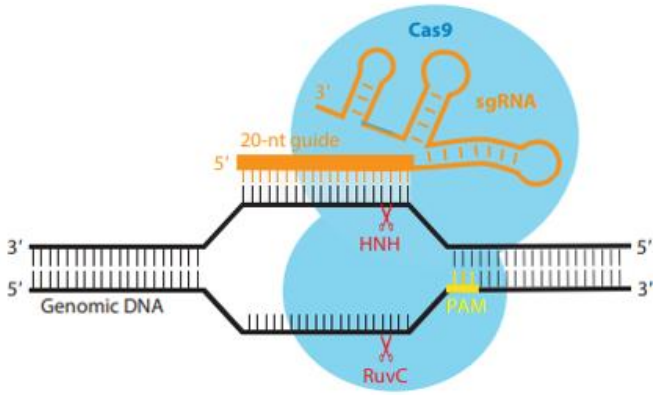
Sınıf 1 Sistemleri: Çoklu protein kompleksleri içeren ve daha karmaşık yapılar sergileyen bu sistemler (Tip I, III ve IV), doğada yaygın olarak bulunmalarına rağmen, laboratuvar uygulamaları için daha zorlayıcı karakteristikler taşımaktadır.

Sınıf 2 Sistemleri: Tek bir efektör proteini kullanan bu sistemler (Tip II Cas9, Tip V Cas12 ve Tip VI Cas13), basit yapıları ve kullanım kolaylığı nedeniyle gen düzenleme teknolojilerinde tercih edilmektedir. Sınıf 2 Cas proteinleri (Tip II Cas9, Tip V Cas12 ve Tip VI Cas13), yalnızca bir endonükleaz içerirler. Bu endonükleazlar, crRNA rehberliğinde hedef nükleik asit dizilerini tanır ve keserler. Sınıf 2 Cas proteinleri, genellikle daha basit bir yapıya ve işleyişe sahiptirler ve bu nedenle gen düzenleme teknolojilerinde daha yaygın olarak kullanılırlar (Şekil 2.4) [14,15].



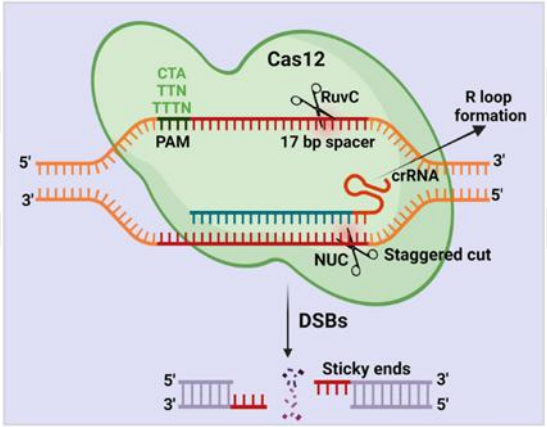
Şekil 2.4 Cas proteinlerinin sınıflandırılması [14]

Sınıf 2 CRISPR sistemleri arasında en çok çalışılan Tip II (Cas9), aktivite için iki parçalı bir RNA yapısına (crRNA ve tracrRNA) veya bunların birleştirilmesiyle oluşturulan tek rehber RNA'ya (gRNA) ihtiyaç duyar [16]. Cas9 nükleazı, hedef DNA'yı tararken "Protospacer Adjacent Motif" (PAM) olarak adlandırılan ve genellikle G-zengini (örneğin 5'-NGG-3') olan dizileri, hedef dizinin 3' ucunda tanır [16, 17]. Hedefin tanınmasıyla birlikte Cas9 enzimi, HNH ve RuvC olmak üzere iki nükleaz domainini aktive eder. HNH domaini hedef (tamamlayıcı) zinciri keserken, RuvC domaini hedef olmayan (tamamlayıcı olmayan) zinciri keser (Şekil 2.5) [16, 18]. Bu eş zamanlı kesim işlemi sonucunda DNA üzerinde "küt uçlu" (blunt end) kırıklar meydana gelir [16].



Şekil 2.5 CRISPR-Cas9 sisteminin çalışma mekanizması ve DNA kesimi [19]

Buna karşın, Sınıf 2'nin bir diğer önemli üyesi olan Tip V (Cas12, eski adıyla Cpf1) sistemleri daha farklı bir mekanizmaya sahiptir. Cas12 enzimleri, tracrRNA'ya ihtiyaç duymaz ve sadece crRNA rehberliğinde çalışabilir, bu da onları yapısal olarak daha basit kılar [20]. Cas12, hedef DNA üzerinde genellikle T-zengini (örneğin 5'-TTTV-3') olan ve hedef dizinin 5' ucunda yer alan PAM dizilerini tanır [20]. En belirgin farklardan biri kesim türüdür; Cas9'un aksine Cas12, DNA'yı küt uçlu değil, 5' çıkıntılı (overhang) "yapışkan uçlu" (staggered) şekilde keser [20]. Bu özellik, genetik materyalin belirli bir yönde ve hassasiyetle entegre edilmesini (knock-in) kolaylaştırdığı için gen düzenleme uygulamalarında önemli bir avantaj sağlar [20, 21]. Ayrıca Cas12, kendi pre-crRNA'sını işleyebilen ikili bir nükleaz (hem DNaz hem RNaz) aktivitesine sahiptir, bu da çoklu gen düzenleme (multiplexing) çalışmalarında tek bir transkript üzerinden birden fazla hedefin düzenlenmesine olanak tanır [22].



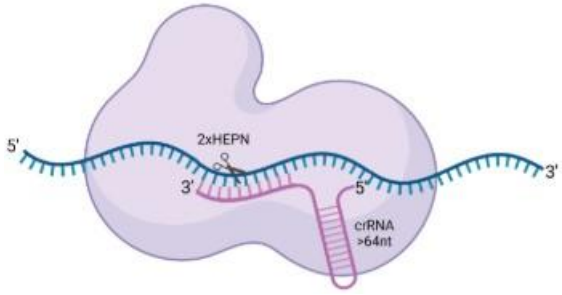
Şekil 2.6 Cas12 sisteminin çalışma mekanizması ve DNA kesimi [23]

CRISPR sistemleri içinde en çok bilineni Cas9 ve Cas12 proteinleri DNA'yı hedeflerken, Cas13 ailesi RNA moleküllerini hedefleyebilme yeteneğiyle öne çıkmaktadır.

Tip VI sistemleri, prokaryotlardaki bilinen tek RNA hedefli CRISPR-Cas sistemidir.

Cas13 sistemleri, tek bir efektör protein ve crRNA'dan oluşan nispeten basit bir yapıya sahiptir. Bu sistemlerde, Cas13 proteini üzerinde bulunan iki adet HEPN (Higher Eukaryotes and Prokaryotes Nucleotide-binding) domaini, RNA kesim aktivitesinden sorumludur. Cas13, hedef RNA dizisini tanıdığında yapısal bir değişikliğe uğrayarak aktif hale gelir. Bu aktivasyon, sadece hedef RNA'nın değil, çevredeki diğer RNA moleküllerinin de spesifik olmayan bir şekilde kesilmesiyle

(collateral cleavage) sonuçlanabilir. Cas13 ailesi kendi içinde Cas13a, Cas13b, Cas13c ve tez çalışmamızın odak noktası olan Cas13d gibi alt türlere ayrılmaktadır. Özellikle Cas13d gibi varyantlar, RNA hedeflemedeki yüksek özgüllükleri ve programlanabilir yapıları sayesinde modern biyoteknolojide gen düzenleme ve tanı kitleri gibi alanlarda güçlü adaylar haline gelmiştir [13].



Şekil 2.7 Cas13 sisteminin çalışma mekanizması ve RNA kesimi [24]

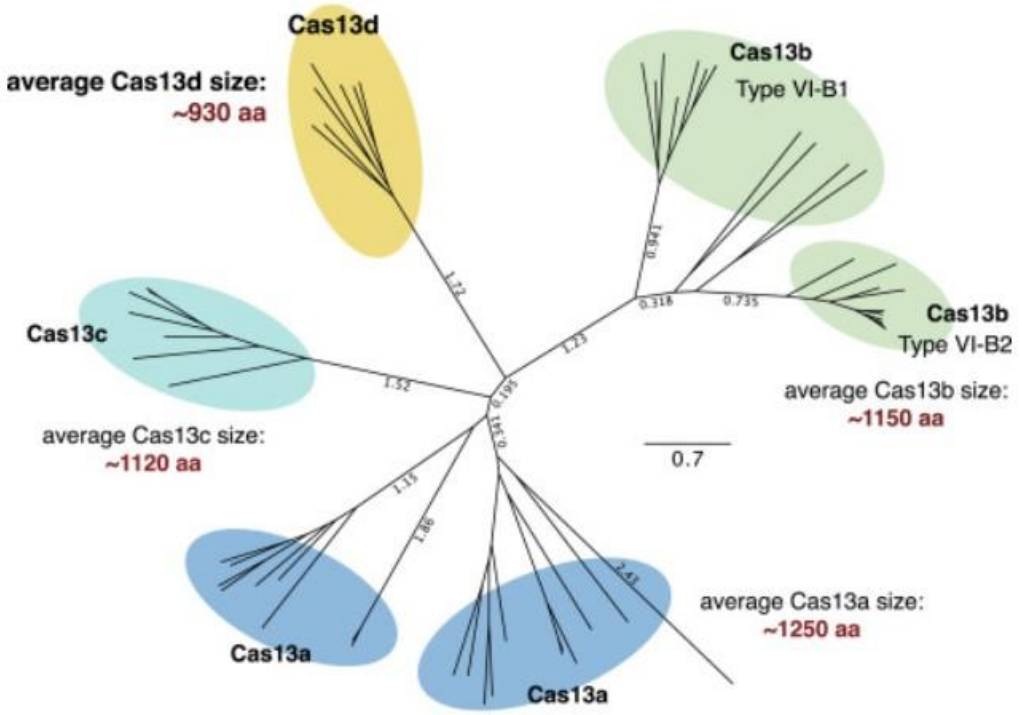
Cas13a (Tip VI-A); Cas13 ailesinin ilk keşfedilen üyesidir ve genellikle Leptotrichia shahii gibi bakterilerde bulunur. RNA hedefleme özelliğinin yanı sıra, hedef RNA'yı tanıdıktan sonra tetiklenen kolateral (yanal) RNase aktivitesi, Cas13a'yı özellikle RNA sensör teknolojileri için önemli bir aday haline getirmektedir. crRNA işleme kapasitesi, bünyesinde bulunan Helical-1 domaini sayesinde sağlanır. Ancak, hedef RNA dizisinin çevresindeki Protospacer Yan Dizisi (PFS) bölgelerine karşı hassasiyet göstermesi, hedefleme esnekliğini sınırlayabilir. Ayrıca, yapılan karşılaştırmalarda Cas13b'ye kıyasla daha düşük RNase aktivitesine sahip olduğu bildirilmiştir.

Cas13b (Tip VI-B); genellikle Prevotella ve Ruminococcus türlerinde bulunan bir Tip VI-B CRISPR-Cas sistemi efektörüdür. Cas13a'ya kıyasla rehber RNA (gRNA) ile farklı etkileşim dinamiklerine sahiptir ve daha güçlü ve kararlı bir RNase aktivitesi sergiler. Aktivitesi, bazı türlerde Csx28 gibi aksesuar proteinler aracılığıyla pozitif ya da negatif yönde düzenlenebilir. Ayrıca, bazı Cas13b varyantlarında PFS bağımlılığı gözlemlenmiştir. Cas13c ise Tip VI-C sistemine ait olup daha az karakterize edilmiştir, ancak yapısal olarak Cas13a ve Cas13b ile benzerlik gösterir. Bu proteinler, çift HEPN domaini içerir ve crRNA işleme ve RNA kesme aktivitelerine sahiptir.

Cas13c (Tip VI-C); henüz sınırlı sayıda çalışmayla karakterize edilmiş olmakla birlikte, genel RNA hedefleme mekanizmasına sahiptir. Yapısal olarak Cas13a ve Cas13b ile benzerlik gösterir. Bu protein, çift HEPN domaini içerir ve crRNA işleme ile RNA kesme aktivitelerine sahiptir.

Cas13d (Tip VI-D); kompakt boyutuyla dikkat çekmektedir; diğer Cas13 varyantlarından (Cas13a-c) 190 ila 300 amino asit daha küçüktür ve RNA parçalanması için PFS gerektirmemesi gibi minimal hedefleme kısıtlamalarına sahiptir. Bu küçük boyut ve esneklik kombinasyonu, özellikle viral vektörler aracılığıyla hücre içine taşınmasında önemli bir avantaj sağlar. Mekanizması, Cas13d'nin kendi CRISPR-RNA'larını (crRNA'lar) işlemesini içerir; bu crRNA'lar daha sonra enzimi tamamlayıcı bir hedef RNA'ya bağlanmaya ve onu parçalamaya (cis-cleavage) yönlendirir [8,15,25]. Daha da önemlisi, Cas13d, spesifik hedef RNA'sı tarafından aktive edildiğinde, "yanal parçalanma" veya "trans-parçalanma" olarak bilinen non-spesifik bir RNaz aktivitesi sergiler ve yakınındaki diğer hedef olmayan RNA moleküllerini degrade eder. Gen modülasyonunda hedef dışı etkilerin bir kaynağı olmasına rağmen, bu benzersiz yanal aktivite, SARS-CoV-2 gibi viral enfeksiyonlar için nükleik asit tespiti için hızlı ve hassas teşhis araçlarının geliştirilmesinde başarıyla kullanılmış temel bir özelliktir. Cas13d'nin geniş kullanım alanı, RNA görüntüleme, RNA-baz düzenleme, RNA epigenom düzenleme ve terapötik müdahaleler gibi çeşitli uygulamaları kapsamaktadır [15,25,26]. Cas13d'nin kompakt boyutu ve PFS gereksiniminin olmaması gibi başlangıçtaki avantajları, sistemin basitliğini düşündürebilir. Ancak, daha derinlemesine incelemeler, Cas13d'nin hedef RNA içindeki ikincil yapıya "son derece duyarlı" olduğunu ve crRNA-hedef RNA bölgelerindeki uyumsuzlukların, distal bölgelerdeki eklemelerin ve silmelerin Cas13d bağlanmasını güçlü bir şekilde engellediğini ortaya koymaktadır. Bu durum, basit diziye bitişik gereksinimleri aşsa da, özgüllüğünün daha karmaşık ve tahmin edilmesi zor RNA yapısal etkileşimleri tarafından yönetildiğini göstermektedir. Bu, mühendislik zorluğunu basit dizi kurallarından (Cas9 için PAM gibi) RNA ikincil ve üçüncül yapılarının sofistike bir şekilde anlaşılmasına ve tahmin edilmesine kaydırmaktadır. Bu karmaşıklık, etkili ve spesifik rehber RNA'ları doğru bir şekilde tasarlamak için gelişmiş hesaplama modellerini (DeepCas13, TIGER, DeepFM-Crispr gibi derin öğrenme

modelleri) gerekli kılmakta ve devam eden araştırma ve teknolojik gelişme için kritik bir alanı vurgulamaktadır [15].



Şekil 2.8 Cas13 alt tiplerinin ortalama protein büyüklükleri [2]

Özellikle Cas13d'nin kompakt yapısı ve yüksek etkinliği, çalışmamızda bu proteini tercih etmemizin temel nedenidir. Cas13d'nin yapısal özellikleri, gen ekspresyonunun hassas bir şekilde düzenlenmesine olanak tanırken, gRNA tasarımında da esneklik sağlamaktadır.

2.3 Hedef Dışı Etkilerin Önemi ve Zorlukları

CRISPR-Cas13d sisteminin sağladığı birçok avantaja rağmen, hedef dışı etkiler önemli bir sorun olarak karşımıza çıkmaktadır.

"Hedef dışı etkiler", bir deneydeki istenmeyen sonuçları kapsayan geniş bir terimdir. Bu istenmeyen etkiler, gRNA'nın hedef RNA dizisi dışındaki RNA dizilerine bağlanması sonucu ortaya çıkmakta ve potansiyel tedavi stratejilerinin güvenliğini tehlikeye atmaktadır [3]. CRISPR sistemleri bağlamında, bu genellikle nükleazın beklenmedik bölgelere lokalize olması ve programlanmış işlevini orada gerçekleştirmesi anlamına gelir. Cas13d için bu, özel olarak RNA moleküllerinin

istenmeyen parçalanması anlamına gelmektedir. Cas13d hedef dışı aktivitesinin birincil tezahürü, enzimin spesifik hedef RNA'sı tarafından aktive edildikten sonra hücredeki diğer hedef olmayan RNA moleküllerini ayrım gözetmeksizin parçaladığı "yanal parçalanma" (trans-parçalanma olarak da bilinir) yeteneğidir.

Cas13d'nin hedef dışı aktivitesine sebep olan mekanizmalar şu şekilde sıralanabilir:

2.3.1 HEPN Alanlarının Yanal RNaz Aktivitesi

Tüm Cas13 nükleazları, RNaz olarak işlev gören iki Yüksek Ökaryot ve Prokaryot Nükleotit Bağlama (HEPN) alanına sahiptir. Bir rehber RNA'nın (gRNA) hedef RNA'sına bağlanmasıyla aktive edildiklerinde, bu HEPN alanları konformasyonel değişikliklere uğrar, açığa çıkar ve sadece hedef RNA'yı değil, aynı zamanda yakındaki diğer erişilebilir RNA moleküllerini de non-spesifik bir şekilde parçalayabilir. Bu, tüm Cas13 endonükleaz ailesinin paylaştığı doğal bir rastgele parçalama yeteneğidir [27]. Bu HEPN alanlarının yanal parçalanmanın kaynağı olarak sürekli vurgulanması, bunun sadece bir "hedef dışı" yan etki değil, Cas13'ün içsel bir mekanik özelliği olduğunu göstermektedir. Bu alanların spesifik bir hedef tarafından aktivasyonu, daha sonra diğer RNA'ların non-spesifik degradasyonuna yol açar. Bu, mükemmel rehber RNA tasarımına rağmen yanal etkinin bir dereceye kadar devam edebileceği anlamına gelir. Bu temel mekanik özellik, azaltma stratejilerinin basit rehber RNA tasarımının ötesine geçmesini gerektirir. Bu stratejiler, HEPN alanlarının aktivasyon eşiğini hassas bir şekilde ayarlayan veya hedef aktivasyonu geniş yanal aktiviteden ayıran Cas13 varyantlarının mühendisliğini içerebilir. Bu, etkinlik ve özgüllük arasında bir denge sağlamak için HEPN alan regülasyonunun daha derinlemesine yapısal ve biyokimyasal anlaşılmasına duyulan kritik ihtiyacı vurgular.

2.3.2 crRNA ve Hedef RNA Arasındaki Uyumsuzluklar

Cas13d'nin bağlanma ve parçalama aktivitesi, crRNA ve hedef RNA dizisi arasındaki uyumsuzluklara karşı hassastır. crRNA-hedef RNA çiftinin distal bölgesindeki uyumsuzluklar, eklemeler ve silmeler Cas13d bağlanmasını güçlü bir şekilde engeller. Tersine, proksimal bölgedeki değişiklikler, bağlanmayı önemli ölçüde etkilemeden nükleaz aktivitesini inhibe edebilir [15]. İlginç bir şekilde, G-U wobble eşleşmeleri gibi belirli uyumsuzluk türleri, diğer tek bazlı

uyumsuzluklardan daha iyi tolere edilir, bu da Cas13d aktivitesi üzerinde karmaşık, konum ve bağlama bağlı bir etki olduğunu gösterir.

2.3.3 Dizi Benzerliği ve Homoloji

CRISPR sistemlerinin genel prensipleri, amaçlanan hedefe yeterli dizi benzerliği taşıyan beklenmedik bölgelere lokalize olabileceğini düşündürmektedir. DNA'yı hedefleyen Cas9 için açıkça detaylandırılmış olsa da, homolog dizilere rastgele bağlanmanın altında yatan mekanizma, RNA'yı hedefleyen Cas13d için de geçerlidir ve hedef dışı RNA parçalanmasına katkıda bulunur.

2.3.4 Hedef RNA İkincil Yapısı

Cas13d aktivitesi, hedef RNA içindeki ikincil yapıya "son derece duyarlıdır". Genel hedef RNA ikincil yapısını artıran nükleotit sübstitüsyonları, Cas13d'nin bağlanma afinitesini önemli ölçüde azaltabilir. Özellikle hedef RNA'nın distal segmentindeki yapısal elementler, Cas13d bağlanmasını engelleyebilir. Ancak, Cas13d'nin yüksek derecede kararlı RNA yapılarına bile bağlanma ve parçalama konusunda olağanüstü bir yetenek sergilediğini belirtmek önemlidir, bu konuda RNA interferansını (RNAi) geride bırakmaktadır. Yalnızca istisnai derecede kararlı, mükemmel eşleşmiş saç tokası yapıları (örneğin, -31.7 kcal/mol veya daha yüksek termodinamik stabiliteye sahip olanlar) Cas13d hedeflemesini önemli ölçüde engellemiştir [15].

2.3.5 Cas13d ve Hedef RNA'ların Ekspresyon Seviyeleri

Yanal aktivitenin şiddeti, hem Cas13d efektörünün hem de hedef RNA'nın ekspresyon seviyeleriyle pozitif korelasyon göstermektedir. Örneğin, yüksek RfxCas13d ekspresyonu, daha belirgin yanal RNA degradasyonuna yol açabilir [28].

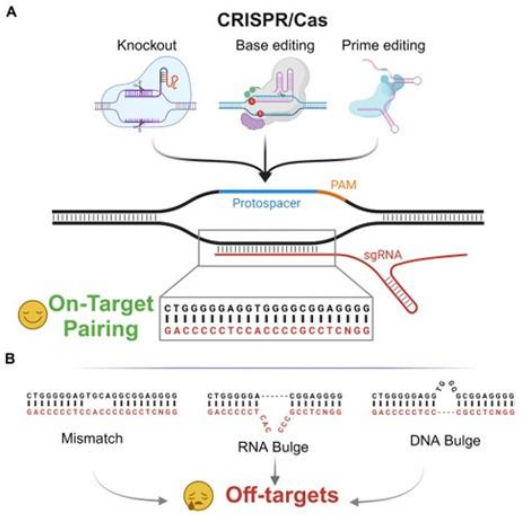
Bu faktörlerin karmaşık etkileşimi, hedef dışı etkilerin tahmin edilmesini ve azaltılmasını zorlaştırmaktadır. Uyumsuzluk tipi ve konumu, hedef RNA ikincil yapısı ve hem Cas13d enziminin hem de hedefinin ekspresyon seviyeleri gibi çeşitli faktörler hedef dışı etkilere katkıda bulunur. Ayrıca, hedef dışı etkilerin boyutu hücre tipine göre değişebilir. Bu, hedef dışı etkileri tahmin etmek veya azaltmak için basit, tek bir kural yaklaşımının yetersiz olduğunu göstermektedir. Bu karmaşık

etkileşim, hedef dışı tahmin ve azaltmanın izole kurallara dayanamayacağı anlamına gelir. Bunun yerine, bu çeşitli moleküler ve hücresel faktörleri entegre edebilen sofistike, çok parametreli hesaplama modellerine (DeepCas13, TIGER ve DeepFM-Crispr gibi derin öğrenme modelleri) ihtiyaç duyulmaktadır. Bu karmaşıklık, bir deneysel kurulumdan veya hücre hattından elde edilen hedef dışı etki bulgularının diğerine doğrudan aktarılamayacağı anlamına da gelir, bu da sonuçların genellenebilirliğini ve klinik uygulamalara çeviri yolunu zorlaştırmaktadır [15].

Yanal aktivite, yüksek düzeyde ifade edilen genleri hedeflerken en şiddetli şekilde görülür. Hedef dışı etkilerin boyutu, hedef mRNA ekspresyon seviyeleriyle pozitif korelasyon gösterebilir. Yüksek RfxCas13d ekspresyonu, yanal RNA degradasyonuna yol açabilir. Aktif Cas13 türünün konsantrasyonu üzerinde daha hassas kontrol, yanal etkileri azaltabilir ve sitotoksisiteyi düşürebilir. Bu durum, hedef dışı etkiler için doza bağımlı bir ilişki olduğunu göstermektedir. Bu, sadece Cas13d'yi iletmenin yeterli olmadığı anlamına gelir; ekspresyonunun miktarı ve süresi, hedef dışı etkileri yönetmek için kritik öneme sahiptir. Terapötik uygulamalar için, bu, yanal hasarı minimize etmek amacıyla sıkı bir şekilde düzenlenmiş, uzamsal ve zamansal olarak kontrol edilebilir dağıtım sistemlerine (örneğin, indüklenebilir promotörler veya geçici dağıtım yöntemleri kullanarak) ihtiyaç duyulduğunu ima eder. Bu, odağı sadece rehber RNA tasarımından, tüm dağıtım ve ekspresyon stratejisine kaydırmaktadır [8,15].

2.4 Hedef Dışı Etki Tespitinin Büyük Önemi

CRISPR teknolojilerinin güvenliği tartışılırken, tarihsel süreçte ilk odak noktası DNA hedefleyen Sınıf 2 sistemleri olmuştur. Cas9 ve Cas12 nükleazları için hedef dışı aktivite tespiti, oluşturdukları hasarın kalıcı ve geri döndürülemez niteliği nedeniyle hayati önem taşır. Bu sistemlerde meydana gelen istenmeyen bir kesim, hedef olmayan genomik bölgelerde kalıcı çift zincir kırıklarına (DSB) yol açabilir. Literatürdeki çalışmalar, bu kırıkların yanlış onarılması sonucunda istenmeyen insersiyon-delesyon (indel) mutasyonlarının oluşabileceğini, daha da kritik olarak büyük kromozomal translokasyonların ve genomik instabilitenin tetiklenebileceğini göstermiştir [29, 30].



Şekil 2.9 Hedef dışı etkilerin gösterimi [31]

Özellikle tümör baskılayıcı genlerde (örneğin p53) meydana gelebilecek hedef dışı mutasyonlar, terapötik uygulamalarda kanserleşme (onkogenez) riskini beraberinde getirir [32]. Cas12 sistemleri, Cas9'a kıyasla daha yüksek özgüllüğe sahip olsa da, yanlış eşleşme toleransı nedeniyle benzer genomik riskleri taşımaya devam etmektedir [20]. Dolayısıyla DNA düzenleyiciler için 'hedef dışı etki', genomun bütünlüğünü bozan kalıcı bir 'genetik yara' anlamına gelmektedir. CRISPR-Cas13d sistemlerinde hedef dışı etkilerin tespiti, hem temel bilimsel araştırmaların güvenilirliği hem de terapötik uygulamaların güvenliği ve klinik çevirisi için kritik öneme sahiptir. Tespit edilmeyen hedef dışı etkiler, deneysel sonuçları karıştırabilir ve fenotiplerin yanlış yorumlanmasına yol açarak araştırma bulgularının iç geçerliliğini ve tekrarlanabilirliğini tehlikeye atar. Büyük ölçekli CRISPR taramalarında bile, hedef dışı canlılık etkileri yanlış pozitiflere neden olabilir, bu da transkriptom mühendisliği uygulamalarının bütünlüğü için doğru tahmin ve tespiti zorunlu kılar [33]. Terapötik uygulamalarda, hedef dışı mutasyonlar veya yaygın RNA degradasyonu, hücresel toksisite, esansiyel genlerin bozulması, immünojenisite ve onkogenez gibi ciddi güvenlik endişeleri taşır. İnsan klinik deneyleri için kabul edilemez riskler oluşturduğundan, bu durum Cas13d tabanlı tedavilerin klinik çevirisini ve düzenleyici onayını önemli ölçüde yavaşlatır. Yanal parçalanma, protein kodlayan mRNA'lar ve kodlamayan RNA'lar dahil olmak üzere geniş bir RNA yelpazesinin kontrolsüz eliminasyonuna yol açar, stres yanıtlarının ve doğal bağışıklık yollarının indüksiyonuna, apoptoz hücre ölümüne ve azalmış

hücre canlılığına neden olabilir [27]. Bu yaygın hücresel bozukluk, hedef dışı etkileri kritik bir toksisite sorununa dönüştürür.

2.5 CRISPR-Cas13d Hedef Dışı Aktivitesini Etkileyen Faktörler

Cas13d'nin hedef dışı aktivitesi, çeşitli moleküler ve hücresel faktörlerin karmaşık bir etkileşiminden etkilenir. Bu faktörlerin anlaşılması, hem hedef dışı etkilerin doğru bir şekilde tespit edilmesi hem de etkili bir şekilde azaltılması için temeldir.

2.5.1 Rehber RNA Tasarımı ve Dizi Tamamlayıcılığı

Rehber RNA dizisi, Cas13d'nin özgüllüğünü belirler. Hassas Cas parçalanması, dizi, uzunluk ve GC içeriği gibi faktörleri dikkate alan etkili bir gRNA tasarımı gerektirir. crRNA ve hedef RNA dizisi arasındaki uyumsuzluklar, özellikle crRNA-hedef RNA çiftinin distal bölgelerinde, Cas13d bağlanmasını güçlü bir şekilde engeller. Bununla birlikte, proksimal bölgedeki değişiklikler, bağlanmayı etkilemeden nükleaz aktivitesini inhibe edebilir. G-U wobble eşleşmeleri gibi belirli uyumsuzluk türleri, diğer tek bazlı uyumsuzluklardan daha iyi tolere edilir. Hedef dışı etkileri azaltmak için genomun başka yerlerinde düşük dizi benzerliğine sahip bir gRNA seçmek önemlidir [15].

2.5.2 Hedef RNA İkincil Yapısı ve Erişilebilirliği

Cas13d, hedef RNA içindeki ikincil yapıya karşı son derece hassastır. Genel hedef RNA ikincil yapısını artıran nükleotit sübstitüsyonları, Cas13d bağlanma afinitesini önemli ölçüde azaltabilir. Rehber RNA'nın ikincil yapısı da nakavt verimliliği için önemli bir faktördür. RNA yapısı tek bir nükleotit sübstitüsyonuyla bile önemli ölçüde değişebilir ve hem bağlanmayı hem de parçalanmayı etkileyebilir. Ancak, Cas13d, RNA interferansını (RNAi) geride bırakarak yüksek derecede kararlı RNA yapılarına bağlanma ve onları parçalama konusunda olağanüstü bir yetenek sergiler. Yalnızca istisnai derecede kararlı, mükemmel eşleşmiş saç tokası yapıları (örneğin, -31.7 kcal/mol veya daha yüksek ΔG'ye sahip olanlar) Cas13d hedeflemesini engellemiştir.

Bu durum, hem dizi tamamlayıcılığı (uyumsuzluklar) hem de RNA ikincil yapısının kritik faktörler olduğunu göstermektedir. Ancak, Cas13d'nin yapıya duyarlı

olmasına rağmen yüksek derecede kararlı yapıları da parçalayabilmesi, belirli bir eşik veya hedeflemeden kaçan spesifik yapısal motiflerin varlığını düşündürmektedir. Bu karmaşıklık, basit dizi hizalama araçlarının Cas13d özgüllüğünü tahmin etmek için yetersiz olduğu anlamına gelir. Gelişmiş hesaplama modelleri, RNA katlanma dinamiklerini ve yapısal erişilebilirliği dahil etmelidir. Bu aynı zamanda, etkili ve spesifik rehber RNA'lar için "tasarım alanının" bu biyofiziksel özellikler tarafından oldukça kısıtlandığı anlamına gelir, bu da rasyonel tasarımı yinelemeli deneysel doğrulama ve hesaplama iyileştirmesi gerektiren önemli bir zorluk haline getirir [15].

2.6 Hedef Dışı Etkilerin Tespit ve Tahmin Stratejileri

Hedef dışı etkilerin tespiti ve tahmini, CRISPR-Cas13d sistemlerinin güvenilirliğini ve güvenliğini sağlamak için kritik öneme sahiptir. Bu alanda hem deneysel hem de hesaplamalı metodolojilerde önemli ilerlemeler kaydedilmiştir.

2.6.1 Deneysel Metodolojiler

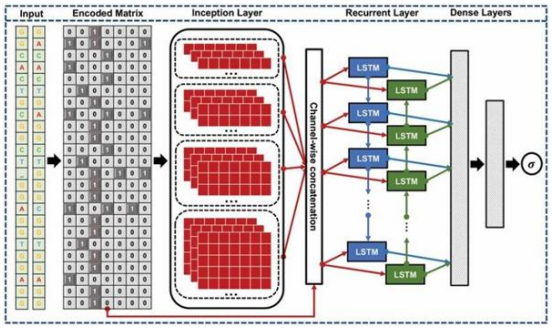
CRISPR-Cas13d sistemlerinde hedef dışı etkilerin tespiti için çeşitli deneysel yaklaşımlar kullanılmaktadır. Bunlar arasında, hücrelerin Cas13 efektörleri, rehber RNA'lar ve floresan proteinler veya lusiferazlar gibi raporlayıcı gen plazmitleri ile geçici olarak transfekte edildiği ko-transfeksiyon deneyleri yer alır; bu deneylerde hedef olmayan raporlayıcı RNA'nın tükenmesi yanal hasarı gösterir [34]. Çift floresan raporlayıcı sistemler, hedef içi ve hedef dışı parçalanmayı ayırt etmek için özel olarak tasarlanmıştır [35]. Ayrıca, hedef içi aktiviteyi ve hedef dışı canlılık etkilerini araştırmak için esansiyel/esansiyel olmayan genleri hedefleyen CRISPR-Cas13d proliferasyon taramaları da uygulanmaktadır [36]. Moleküler teknikler arasında, spesifik RNA transkriptlerini doğrudan görselleştiren ve nicelendiren Northern blotlama [34], RNA seviyelerinin son derece hassas ve kantitatif ölçümünü sağlayan RT-qPCR ve transkriptom geneli hedef dışı etkileri karakterize etmek için tüm RNA değişikliklerinin küresel bir görünümünü sunan RNA-seq bulunmaktadır. RNA-CHAMP gibi yöntemler ise RNA-protein etkileşimlerinin büyük ölçekli paralel profillemesini sağlar [27].

2.6.2 Hesaplamalı ve Makina Öğrenimi Yaklaşımları

Rehber RNA (gRNA) tasarımı ve hedef dışı etkilerin tahmini, CRISPR sistemlerinin başarısı için kritik öneme sahiptir. Bu alandaki hesaplamalı yaklaşımlar, basit kural tabanlı hizalama yöntemlerinden, büyük veri setleri üzerinde eğitilen gelişmiş derin öğrenme modellerine doğru evrilmiştir. Farklı CRISPR sınıfları için geliştirilen özelleşmiş modeller aşağıda detaylandırılmıştır.

2.6.2.1 Cas9 Sistemleri İçin Geliştirilen Modeller

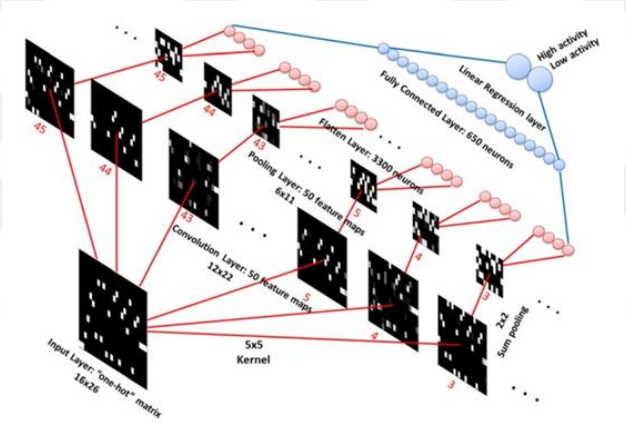
CRISPR-Cas9 sistemlerinde hedef dışı etkilerin ve kesim verimliliğinin tahmini, veri bolluğu nedeniyle hesaplamalı biyolojinin en gelişmiş olduğu alandır. İlk dönem yaklaşımları basit hizalama kurallarına dayanırken, son yıllarda derin öğrenme tabanlı modeller standart haline gelmiştir. Bu alandaki öncü çalışmalardan biri olan DeepCRISPR, hem gRNA'nın hedef içi kesim etkinliğini hem de hedef dışı profilini aynı anda tahmin edebilen birleşik bir derin öğrenme çerçevesi sunmuştur. Bu model, gRNA dizilerinin "one-hot" kodlamasını kullanarak, el yordamıyla özellik çıkarımı yapmadan ham dizi verisinden öğrenme yeteneğine sahiptir [5]. Buna ek olarak, DeepHF gibi modeller, sadece dizi bilgisini değil, aynı zamanda RNA ikincil yapılarını da modele dahil ederek tahmin hassasiyetini artırmayı hedeflemiştir. DeepHF, uzun süreli kısa bellek (LSTM) ağlarını kullanarak gRNA'nın kararlılığını etkileyen yapısal özellikleri analiz etmiştir [37]. Ayrıca CRISPRon gibi daha yeni araçlar, dizi bilgisini epigenetik faktörler (örneğin kromatin erişilebilirliği) ile birleştirerek, özellikle ökaryotik hücrelerdeki karmaşık genomik ortamda daha doğru tahminler yapabilmektedir [38].



Şekil 2.10 CRISPR-Cas9 hedef dışı etki tahmini için hibrit derin öğrenme mimarisi [39]

2.6.2.2 Cas12 Sistemlerine Yönelik Yaklaşımlar

Cas12 sistemleri (özellikle Cas12a/Cpf1), Cas9'dan farklı bir PAM yapısı (T-zengini) ve kesim mekanizması kullandığı için kendine özgü tahmin modellerine ihtiyaç duyar. Bu alandaki en etkili çalışmalardan biri olan DeepCpf1, yüksek verimli tarama verileri kullanılarak eğitilmiş evrişimli sinir ağı (CNN) tabanlı bir modeldir. DeepCpf1, Cas12a aktivitesini tahmin ederken sadece nükleotit dizisini değil, aynı zamanda hedef bölgenin kromatin erişilebilirliğini de dikkate alarak, önceki puanlama algoritmalarından çok daha yüksek bir doğruluk oranına ulaşmıştır [6]. Benzer şekilde Seq-DeepCpf1, sadece dizi bilgisine odaklanarak kromatin bilgisinin bulunmadığı durumlarda bile yüksek performans gösterecek şekilde optimize edilmiştir [6].



Şekil 2.11 CRISPR-Cas12 hedef dışı etki tahmini için derin öğrenme mimarisi [40]

2.6.2.3 Cas13 Sistemleri ve Güncel Yaklaşımlar

Rehber bağımlı hedef içi ve hedef dışı etkileri doğru bir şekilde tahmin etmek, hesaplamalı ve makine öğrenimi yaklaşımlarının geliştirilmesiyle önemli ilerlemeler kaydetmiştir [26]. Bu alanda öne çıkan derin öğrenme modelleri arasında, rehber dizilerinden ve ikincil yapılardan hedef içi aktiviteyi tahmin etmek için tasarlanmış DeepCas13 [27]; büyük veri kümeleri üzerinde eğitilmiş ve hem hedef içi hem de hedef dışı aktivite tahmininde mevcut modellerden daha iyi performans gösteren TIGER ve RNA ikincil yapılarının ve genel gRNA etkinliğinin tahminlerini geliştirmek için büyük bir dil modelinden yararlanan DeepFM-Crispr

yer almaktadır [26]. Yapay zekanın entegrasyonu, büyük ölçekli Cas13d tarama veri kümeleriyle birlikte, CRISPR araçlarının hassasiyetini ve verimliliğini önemli ölçüde artırmaktadır [27]. Bu yaklaşımlar, hedef dışı etkileri deneysel olarak tespit etmekten, rehber RNA tasarımı sırasında tahmin etmeye ve önlemeye yönelik önemli bir paradigma kaymasını temsil etmektedir.

2.7 gRNA Sekansının Temel Bileşenleri ve Fonksiyonu

Bir gRNA'nın birincil yapısı, nükleotidlerinin doğrusal dizisini ifade eder. Bu dizi, gRNA işlevinin temel belirleyicisidir ve belirli bir hedef nükleik aside tanıma ve bağlanma yeteneğini doğrudan belirler. Bir gRNA, tipik olarak kısa bir sentetik RNA molekülüdür ve iki ana işlevsel bileşenden oluşur: kullanıcı tanımlı bir ayırıcı dizi ve bir iskele dizisi. Ayırıcı dizi, genellikle yaklaşık 20 nükleotid uzunluğunda olup, hedeflenen nükleik asit dizisine tamamlayıcıdır ve Cas efektör proteinini istenen genomik veya transkriptomik lokusa yönlendirmekten sorumludur Bu tamamlayıcılık, gRNA'nın hedefine özgü bağlanmasını sağlar. İskele dizisi ise Cas proteinini bağlamak için gereklidir ve ribonükleoprotein (RNP) kompleksinin oluşumunu kolaylaştırırgRNA'nın kesin dizisi, Cas enzimini doğru hedefe yönlendirmek için kritik öneme sahiptir. Özellikle, gRNA hedefleme dizisinin 3' ucundaki 8-10 bazlık "tohum bölgesi", hedef RNA'ya ilk bağlanma ve tavlanmada belirleyici bir rol oynar [36,41]. Bu bölgedeki herhangi bir uyumsuzluk, bağlanma afinitesini ve dolayısıyla düzenleme verimliliğini önemli ölçüde etkileyebilir.

Kompakt bir RNA hedefleyen Tip VI CRISPR efektörü olan Cas13d, RNA kesim aktivitesi için bir Protospacer Yanı Dizi (PFS) gerektirmemesiyle dikkat çekicidir. Bu PFS gereksiniminin olmaması, Cas13d'yi diğer Cas13 ailesi enzimlerinden ayırır; zira bazı Cas13 varyantları (örneğin, Cas13a 5' non-G; Cas13b 5' non-C veya 3' NAN/NNA gerektirir) belirli PFS'lere ihtiyaç duyar [8]. Cas13d proteinlerinin küçük boyutu (Cas13a-c'den ortalama 190-300 amino asit daha küçük, yaklaşık 930 amino asit) onları terapötik uygulamalarda dağıtım için avantajlı kılar. Küçük boyutlarına rağmen, Cas13d ortologları hem CRISPR RNA işlenmesinde hem de hedef/yan RNA kesiminde aktiftir. Cas13d'nin bir protospacer yan dizisi (PFS) gerektirmemesi, hedef bölge seçiminde daha fazla esneklik anlamına gelir ve potansiyel olarak RNA manipülasyonu ve tespiti için uygulanabilirliğini genişletir [8]. Ancak, bu durum aynı zamanda DNA hedefleyen sistemlerde PAM'ın sağladığı

özgüllük katmanını ortadan kaldırır. Bu, gRNA tasarımının titizlikle optimize edilmemesi durumunda hedef dışı etkilerin riskini artırabilir. Bu esneklik ve doğal özgüllük arasındaki denge, Cas13d'nin biyoteknolojik ve terapötik uygulamaları için kritik bir husustur [33].

2.8 İkincil Yapının Tanımı ve Biyolojik Önemi

RNA'nın ikincil yapısı, tek zincirli bir RNA molekülü içindeki baz eşleşmeleri yoluyla oluşan lokal katlanmaları ifade eder. Bu yapılar, genellikle gövdeler (stem), ilmekler (loop), saç tokaları (hairpin) ve bulgeler gibi motifleri içerir. RNA ikincil yapısı, nükleotidler arasındaki Watson-Crick (A-U, G-C) ve sallanma (G-U) baz eşleşmeleriyle oluşan lokal etkileşimleri tanımlar [41]. Bu eşleşmeler, RNA molekülünün üç boyutlu konformasyonunu belirleyen gövdeler (çift sarmal bölgeler) ve ilmekler (eşleşmemiş bölgeler) gibi karakteristik motifleri oluşturur. Bu ikincil yapılar, RNA'nın hücresel işlevlerinde kritik rol oynar. Ribozomlarda kataliz, RNA-indüklenmiş susturma kompleksinin oluşumu ve viral replikasyon mekanizmalarında substrat görevi görme bu işlevlere örnektir [42]. Cas13d, RNA hedefleyen bir sistem olarak, hedef RNA'nın ikincil yapısına karşı benzersiz bir hassasiyet/tolerans sergiler. Cas13d'nin hedef RNA'daki ikincil yapıya karşı hassas olduğu, ancak aynı zamanda oldukça kararlı RNA yapılarını etkili bir şekilde bağlama ve kesme yeteneğine sahip olduğu gösterilmiştir [15]. RNA girişiminden (RNAi) farklı olarak, Cas13d mekanizması daha kararlı RNA ikincil yapılarını tolere edebilir [43]. Örneğin, RNAi -19.0 kcal/mol'lük termodinamik stabiliteye sahip yapıları bozamazken, Cas13d -31.7 kcal/mol veya daha yüksek stabiliteye sahip yapılarla karşılaşana kadar aktif kalmıştır [44]. Bu, Cas13d'nin kompakt ikili Cas13 crRNA kompleksinin RNA ayırıcısını çözme içsel kapasitesine bağlanabilir [43]. Cas13d'nin, diğer bazı Cas sistemlerinin aksine, oldukça kararlı RNA yapılarını kesebilmesi, viral genomları hedeflemede önemli bir avantaj sağlamaktadır. Bununla birlikte, Cas13d bağlanması, distal crRNA-hedef RNA bölgelerindeki uyumsuzluklar, eklemeler ve silmeler tarafından güçlü bir şekilde cezalandırılırken, proksimal bölgedeki değişiklikler bağlanmayı etkilemeden nükleaz aktivitesini inhibe eder. Bu durum, gRNA tasarımının hedef RNA'nın ikincil yapısını dikkate almasını zorunlu kılmaktadır. Hedef RNA'nın genel ikincil yapısını artıran nükleotid ikameleri, bağlanma afinitesini önemli ölçüde azaltır. Bu,

Cas13d'nin aktivitesinin hedef RNA ikincil yapısına karşı yüksek oranda hassas olduğu anlamına gelir ve tasarım sürecinde bu yapısal özelliklerin titizlikle değerlendirilmesi gerektiğini gösterir [44].

2.9 Tersiyer Yapı (3D Konformasyon)

RNA'nın tersiyer yapısı, ikincil yapı elemanlarının (gövdeler ve ilmekler) üç boyutlu uzayda belirli bir şekil oluşturmak üzere daha fazla katlanmasıyla ortaya çıkan karmaşık üç boyutlu konformasyonunu ifade eder. Bu, bir polipeptit zincirinin tek bir "omurgaya" sahip olduğu protein tersiyer yapısına benzer, ancak RNA için nükleotidler ve fosfodiester omurgası arasındaki etkileşimler tarafından belirlenir. Cas13d gibi RNA hedefleyen sistemlerde, gRNA'nın üç boyutlu konformasyonu, Cas proteinini doğru hedefe yönlendiren ve nükleaz aktivitesini tetikleyen ribonükleoprotein (RNP) kompleksinin oluşumu ve işlevi için hayati öneme sahiptir [45]. Cas13d, kompakt bir RNA nükleazı olarak bilinir ve kendi CRISPR-RNA'larını (crRNA'lar) işler, ardından crRNA'ya tamamlayıcı bir hedef RNA'ya bağlanır ve onu parçalar (cis-cleavage). Bu aktivasyon üzerine, Cas13d ayrıca "yanal parçalanma" veya "trans-parçalanma" olarak bilinen non-spesifik bir RNaz aktivitesi sergileyerek yakındaki diğer hedef olmayan RNA moleküllerini de degrade eder. Bu RNaz aktiviteleri, Cas13 tarafından kodlanan iki HEPN alanı tarafından katalize edilir [27]. Hedef RNA bağlandığında, Cas13d'nin HEPN alanları konformasyonel değişikliklere uğrayarak açığa çıkar ve diğer erişilebilir RNA moleküllerini de kesebilir. Cas13d'nin bağlanması, crRNA-hedef RNA bölgelerindeki uyumsuzluklar, eklemeler ve silmeler tarafından güçlü bir şekilde engellenirken, proksimal bölgedeki değişiklikler bağlanmayı etkilemeden nükleaz aktivitesini inhibe edebilir. Biyofiziksel modeller, Cas13d'nin hedef RNA'yı başlangıçta çözücüye maruz kalan distal ayırıcı bölgede tanıdığını, ardından proksimal bölgedeki hedef RNA'ya doğru RNA dubleks oluşumunun gerçekleştiğini öne sürmektedir. Cas13d, hedef RNA içindeki ikincil yapıya karşı "son derece hassas" olmasına rağmen, RNA interferansını (RNAi) geride bırakarak oldukça kararlı RNA yapılarını bile etkili bir şekilde bağlama ve kesme yeteneği sergiler [43]. Yalnızca mükemmel eşleşmiş, istisnai derecede kararlı saç tokası yapıları Cas13d hedeflemesini engellemiştir. Ayrıca, WYL alanı içeren aksesuar proteinler, Cas13d ortologlarının aktivitesini pozitif yönde modüle edebilir. Bu

dinamik ve yapısal etkileşimler, Cas13d'nin özgüllüğünü ve işlevselliğini belirlemede tersiyer yapının kritik rolünü vurgulamaktadır [46].

3
DERİN ÖĞRENME STRATEJİSİ

3.1 Biyolojik Veri Analizinde Derin Öğrenme

Biyoinformatik alanında, karmaşık biyolojik verilerin analizi ve yorumlanması, derin öğrenme tekniklerinin sunduğu yenilikçi yaklaşımlarla önemli ölçüde dönüşmüştür [47].

Yapay sinir ağları (YSA), insan beynindeki bilgi iletimini taklit eden, birbirine bağlı "nöronlardan" oluşan algoritmalardır. Girdi verilerini alırlar, otomatik olarak işlerler ve bir çıktı üretirler. Bir YSA'nın bağımsız çalışabilmesi için eğitim sürecinden geçmesi gerekir; bu süreçte verideki özellikleri tanımlamayı öğrenir [48]. YSA mimarisi tipik olarak üç ana bileşenden oluşur:

Girdi Katmanı; ağın veriyi aldığı ve bir sonraki aşamaya geçmeden önce kategorilere ayırdığı yerdir. Bu katmandaki nöronlar, girdi verisindeki özelliklere karşılık gelir (örneğin, görüntü işlemede her nöron bir piksel değerini temsil edebilir).

Gizli Katmanlar; girdi ve çıktı katmanları arasında yer alan ara katmanlardır. Ağın gerektirdiği hesaplamaların çoğunu gerçekleştirirler ve sayıları ile boyutları görevin karmaşıklığına göre değişebilir. Her gizli katman, girdi verisine bir dizi ağırlık ve sapma uygular, ardından doğrusal olmama özelliği kazandırmak için bir aktivasyon fonksiyonu kullanır.

Çıktı Katmanı; ağın işleme sonuçlarını sunan son katmandır. Çıktı katmanındaki nöron sayısı, bir sınıflandırma problemindeki sınıf sayısına veya bir regresyon problemindeki çıktı sayısına karşılık gelir. Çıktı katmanı için seçilen aktivasyon fonksiyonu, problemin türüne bağlıdır (örneğin, çok sınıflı sınıflandırma için Softmax, ikili sınıflandırma için Sigmoid, regresyon için Doğrusal) [48,49].

Nöronlar, bir YSA içindeki bilgiyi ileten, birbirine bağlı bileşenlerdir. Her nöronun bir sapma değeri ve her bağlantının (kenar) bir ağırlığı vardır. Aktivasyon fonksiyonları, bir YSA içinde bilginin ağın katmanları boyunca akarken hangi nöronların etkinleştirilmesi gerektiğini belirlemede kritik bir rol oynar. Bu fonksiyon, nöronlar arasında bilginin doğrusal olmayan hareketini sağlar, bu da YSA'nın aldığı verilerdeki karmaşık ilişkileri ve örüntüleri öğrenmesi için esastır. Aktivasyon fonksiyonları olmasaydı, YSA'lar yalnızca doğrusal olarak çalışabilir ve işleyebilecekleri veri miktarını ve karmaşıklığını önemli ölçüde sınırlardı. Aktivasyon fonksiyonlarının "bilginin doğrusal olmayan hareketini" sağladığı ve bunlar olmadan ağların "doğrusal olarak çalışabileceği, dolayısıyla işleyebilecekleri veri miktarını sınırlayacağı" açıkça belirtilmektedir. Biyolojik sistemler doğası gereği karmaşık ve doğrusal değildir. Derin öğrenmenin doğrusal olmayan ilişkileri modelleme yeteneği, protein yapısı tahmini, gen ifadesi analizi ve çoklu-omik entegrasyonu gibi biyoinformatikteki başarısının doğrudan nedensel bir faktörüdür. Bu, daha basit doğrusal modellere göre önemli bir avantajdır. Bu doğrusal olmama özelliği, derin öğrenmenin basit modellerin başaramadığı karmaşık biyolojik örüntüleri yakalamasına olanak tanır, bu da onu yüksek boyutlu ve karmaşık biyolojik veriler için uygun hale getirir [50,51].

En belirgin avantaj, derin öğrenmenin "özellikleri otomatik olarak çıkarma" yeteneğidir. Bu yetenek, geniş ve çoğu zaman keşfedilmemiş biyolojik veri ortamını keşfetmek için kritik olan "hipotezden bağımsız" keşfi kolaylaştırır ve hangi özelliklerin önemli olduğuna dair önceden tanımlanmış varsayımların ötesine geçer [47,52,53].

Derin öğrenme modellerinin biyolojik verileri işleyebilmesi için, bu verilerin uygun sayısal formatlara dönüştürülmesi gerekir. Bu bölümde, biyolojik dizilerin ve yapısal verilerin vektör uzaylarında nasıl temsil edildiği ve gömüldüğü incelenmektedir [54].

3.1.1 Biyolojik Dizilerin Vektör Temsilleri

Temsil öğrenimi, biyolojik dizileri (DNA, RNA, protein) gerçek değerli vektörlere (dağıtılmış temsiller) dönüştürür. Bu yaklaşım, doğal dil işlemeden ödünç alınmıştır; diziler "cümle" olarak, k-merler (k uzunluğundaki alt diziler) ise "kelime" olarak ele alınır. Amaç, anlamsal benzerliği korumaktır (işlevsel veya

yapısal olarak benzer diziler benzer vektörlere dönüştürülür). Avantajları arasında boyut indirgeme, hesaplama verimliliği (bir kez hesaplama yeterlidir) ve bağlamsal bilginin yakalanması yer alır [54,55].

3.1.2 3D Yapıların Graf Gömme Yöntemleri

Biyolojik moleküllerin karmaşık üç boyutlu (3D) yapılarını derin öğrenme modelleri için uygun formatlara dönüştürmek, biyoinformatikte önemli bir adımdır. Bu alanda, Graf Sinir Ağları (GNN'ler) ve 3D PointNet modelleri gibi yaklaşımlar öne çıkmaktadır. GNN'ler, protein yapılarını grafikler olarak soyutlayarak temsil etmek ve işlemek için özellikle uygundur; burada amino asitler düğüm, aralarındaki etkileşimler ise kenar olarak kabul edilir. Bu ağlar, düğümlerin ve kenarların özelliklerini çıkararak kalıntıların fizikoşimik özelliklerini ve uzaysal ilişkilerini etkili bir şekilde yakalar [56,57]. GNN'ler, protein yapılarını, amino asitlerin düğümler ve etkileşimlerinin kenarlar oluşturduğu grafikler olarak soyutlayarak temsil etmek ve işlemek için özellikle uygundur. Bu ağlar, öz vektörleme teknikleri aracılığıyla hem düğümlerin hem de kenarların özelliklerini çıkarır, kalıntıların fizikoşimik özelliklerini ve uzaysal ilişkilerini etkili bir şekilde yakalar [58]. GNN'ler, 3D yapısal veriyi işlemede mevcut en güçlü yaklaşım olarak kabul edilse de, bu tekniklerin doğrudan entegrasyonu yüksek hesaplamalı maliyetler ve karmaşık mimari değişiklikleri gerektirmektedir. Bu nedenle, bu çalışmanın metodolojisinde, 3D yapısal bilginin özünü yakalamak için hesaplamalı olarak daha verimli ve mevcut DeepCas13 mimarisine entegrasyonu daha yalın olan UMAP tekniği tercih edilmiştir.

3.2 CRISPR Sistemlerinde Hedef Dışı Etki Tespiti için Derin Öğrenme

Hedef dışı etkiler, istenmeyen genomik bölgelerin değiştirilmesiyle ortaya çıkar ve CRISPR-Cas sistemlerinin güvenliği ve etkinliği, özellikle klinik uygulamalar ve gen terapileri için önemli zorluklar yaratır. Bu etkiler, istenmeyen mutasyonlara yol açabilir, gen fonksiyonunu bozabilir, genomik kararsızlığa neden olabilir, bağışıklık tepkilerini tetikleyebilir ve potansiyel olarak onkogenleri aktive edebilir. Doğru tahmin, optimal rehber RNA'ların tasarlanması ve araştırma ve klinik çalışmalarda güvenlik ve güvenilirliğin sağlanması için kritik öneme sahiptir [59].

3.2.1 Cas9 Sistemleri için Derin Öğrenme Modelleri

Derin öğrenme, özellikle en yaygın kullanılan Cas9 sistemi için hedef dışı etki tahmininde devrim yaratmıştır. Bu alandaki modeller, gRNA ve hedef DNA dizileri arasındaki karmaşık kalıpları öğrenerek, geleneksel skorlama tabanlı yöntemlerden daha yüksek doğruluk sağlamaktadır. Literatürde, bu amaçla geliştirilmiş çok çeşitli derin öğrenme mimarileri bulunmaktadır [56,57,60].

Evrişimsel Sinir Ağları (CNN), dizi verilerindeki uzamsal hiyerarşileri ve motifleri yakalama yetenekleri nedeniyle popüler bir seçimdir. Örneğin, DeepCRISPR platformu, gRNA on-target ve off-target tahminini tek bir derin öğrenme çerçevesinde birleştirir ve bu amaçla evrişimsel otokodlayıcılar kullanır. Benzer şekilde, CNN_std gibi modeller, gRNA-DNA çiftini bir görüntü gibi ele alarak standart CNN mimarilerini bu probleme başarıyla uygular [61,62,63].

Daha karmaşık ilişkileri modellemek için CNN ve Tekrarlayan Sinir Ağları (RNN) gibi yapıları birleştiren hibrit modeller de geliştirilmiştir. Bu modeller, hem yerel dizi motiflerini (CNN ile) hem de dizinin ardışık doğasını (RNN ile) yakalayabilir. Bu kategoriye giren önemli modeller arasında Inception ve BiLSTM mimarilerini birleştiren CRISPR-Net, CNN ve BiLSTM'ye ek olarak dikkat mekanizmalarını da içeren CRISPR-IP ve CRISPR-M ve Inception ile yığınlanmış BiGRU kullanan Crispr-SGRU bulunmaktadır [64].

3.2.2 Cas12 Sistemleri için Derin Öğrenme Modelleri

Cas9'a bir alternatif olarak ortaya çıkan Cas12, özellikle AT'ce zengin genomları düzenlemede avantajlar sunar. Bununla birlikte, Cas12a için hedef dışı etki tahmini yapan derin öğrenme modelleri, Cas9'a kıyasla daha az gelişmiştir. Yapılan çalışmalar, Cas12'nın hedef dışı kesim davranışının, nükleotid yanlılığı ve protospacer bitişik motifi (PAM) bölgesine göre oluşan uyumsuzluklarla ilişkili olduğunu göstermiştir. Bu alandaki mevcut hesaplamalı modeller, genellikle Random Forest gibi geleneksel makine öğrenmesi algoritmalarına dayanmakta ve on-target verimliliğini tahmin etmeye odaklanmaktadır. Bu durum, Cas12a sistemleri için özelleşmiş ve yüksek doğruluklu derin öğrenme tabanlı hedef dışı etki tahmin araçlarına olan ihtiyacı ortaya koymaktadır [65].

3.2.3 Cas13 Sistemleri için Derin Öğrenme Modelleri

Derin öğrenme, CRISPR-Cas13d sistemlerinde hem on-target etkinliğin tahmini hem de rehber RNA (gRNA) tasarımının optimizasyonu açısından kritik bir rol oynamaktadır. Cas13d, DNA yerine RNA'yı hedefleyen ve hedef RNA'yı tanıdıktan sonra hem hedef hem de hedef dışı RNA'ları parçalayabilen benzersiz bir kolateral aktiviteye sahip bir enzimdir. Bu özgün mekanizma ve özellikle hedef RNA'nın ikincil yapısına karşı gösterdiği duyarlılık, sistemin işlevselliğini doğru şekilde modellemek için derin öğrenme yaklaşımlarını zorunlu kılmaktadır.

Bu alanda geliştirilen çeşitli modeller, Cas13d'nin hedefleme etkinliğini anlamak ve optimize etmek amacıyla öne çıkmaktadır. DeepCas13, gRNA dizilerini ve tahmin edilen RNA ikincil yapılarını girdi olarak kullanan bir evrişimsel tekrarlayan sinir ağı (CRNN) mimarisiyle, geleneksel yöntemlere kıyasla üstün performans sergiler. DeepFM-Crispr, büyük dil modelleri (LLM) ve transformatör tabanlı yapılar kullanarak evrimsel ve yapısal verileri birleştirir; böylece RNA'nın ikincil yapısını ve genel gRNA etkinliğini daha isabetli şekilde tahmin eder. TIGER modeli ise, on-target ve off-target aktivite tahmini için evrişimsel sinir ağı (CNN) mimarisi kullanır ve transkript düzeyinde ifade modülasyonu sağlamada hassasiyet sunar [66].

Bu modeller, Cas13d'nin RNA hedefleme kapasitesi ile kolateral aktivitesinin karmaşıklığını dikkate alarak, yüksek verimlilikte ve yüksek özgüllükte rehber RNA tasarımını mümkün kılar. Derin öğrenme yaklaşımları arasında özellikle CNN'ler, RNN'ler (özellikle BiLSTM), dikkat mekanizmaları ve hibrit modeller (örneğin CNN-SVM veya CNN-XGBoost kombinasyonları) yaygın olarak kullanılmaktadır. Dikkat mekanizmaları, modelin girdi dizisi içerisindeki anlamlı bölgelere odaklanarak hem performansı artırmakta hem de modelin yorumlanabilirliğini güçlendirmektedir [67,68].

3.2.4 Cas13 Sistemleri için Derin Öğrenme Yaklaşımlarının Eksiklikleri

Derin öğrenme, DNA yerine RNA'yı hedefleyen CRISPR-Cas13 sistemlerinde hem on-target etkinliğin tahmini hem de rehber RNA (gRNA) tasarımının optimizasyonu açısından kritik bir rol oynamaktadır [27]. Ancak, genom düzenleme alanında en yaygın olarak kullanılan Cas9 ve Cas12 sistemleriyle karşılaştırıldığında, Cas13 üzerine yapılan derin öğrenme çalışmaları henüz daha

sınırlı sayıdadır [69]. Cas13d, hedef RNA'yı tanıdıktan sonra hem hedef hem de hedef dışı RNA'ları parçalayabilen benzersiz bir kolateral aktiviteye sahip bir enzimdir. Bu özgün mekanizma ve özellikle hedef RNA'nın ikincil yapısına karşı gösterdiği duyarlılık, sistemin işlevselliğini doğru şekilde modellemek için derin öğrenme yaklaşımlarını zorunlu kılmaktadır [27].

Bu alanda geliştirilen çeşitli modeller, Cas13d'nin hedefleme etkinliğini anlamak ve optimize etmek amacıyla öne çıkmaktadır. DeepCas13, gRNA dizilerini ve tahmin edilen RNA ikincil yapılarını girdi olarak kullanan bir evrişimsel tekrarlayan sinir ağı (CRNN) mimarisiyle, geleneksel yöntemlere kıyasla üstün performans sergiler [27]. DeepFM-Crispr, büyük dil modelleri (LLM) ve transformatör tabanlı yapılar kullanarak evrimsel ve yapısal verileri birleştirir; böylece RNA'nın ikincil yapısını ve genel gRNA etkinliğini daha isabetli şekilde tahmin eder. TIGER modeli ise, on-target ve off-target aktivite tahmini için evrişimsel sinir ağı (CNN) mimarisi kullanır ve transkript düzeyinde ifade modülasyonu sağlamada hassasiyet sunar [26].

Bu modellerin ortak noktası, gRNA dizisi ve tahmin edilen ikincil yapı (2D) gibi özellikleri girdi olarak kullanmalarıdır. Ancak, bu yaklaşımların önemli bir sınırlılığı, RNA'nın fonksiyonelliği için kritik olan üç boyutlu (3D) yapısal bilgiyi doğrudan modellerine dahil etmemeleridir. RNA'nın katlanarak oluşturduğu karmaşık 3D mimari, moleküler etkileşimleri ve dolayısıyla enzimin hedefi tanıma ve kesme verimliliğini doğrudan etkilemektedir. Mevcut modellerin bu bilgiyi göz ardı etmesi, tahmin yeteneklerinde önemli bir eksiklik yaratmakta ve bu alanda yeni yaklaşımlara olan ihtiyacı ortaya koymaktadır.

Mevcut modeller, Cas13d'nin RNA hedefleme kapasitesi ile kolateral aktivitesinin karmaşıklığını dikkate alarak, yüksek verimlilikte ve yüksek özgüllükte rehber RNA tasarımını mümkün kılmayı hedefler. Bu amaçla kullanılan derin öğrenme yaklaşımları arasında özellikle CNN'ler, RNN'ler (özellikle BiLSTM), dikkat mekanizmaları ve hibrit modeller (örneğin CNN-SVM veya CNN-XGBoost kombinasyonları) yaygın olarak bulunmaktadır. Dikkat mekanizmaları, modelin girdi dizisi içerisindeki anlamlı bölgelere odaklanarak hem performansı artırmakta hem de modelin yorumlanabilirliğini güçlendirmektedir [27, 69].

4
MATERYAL METOD

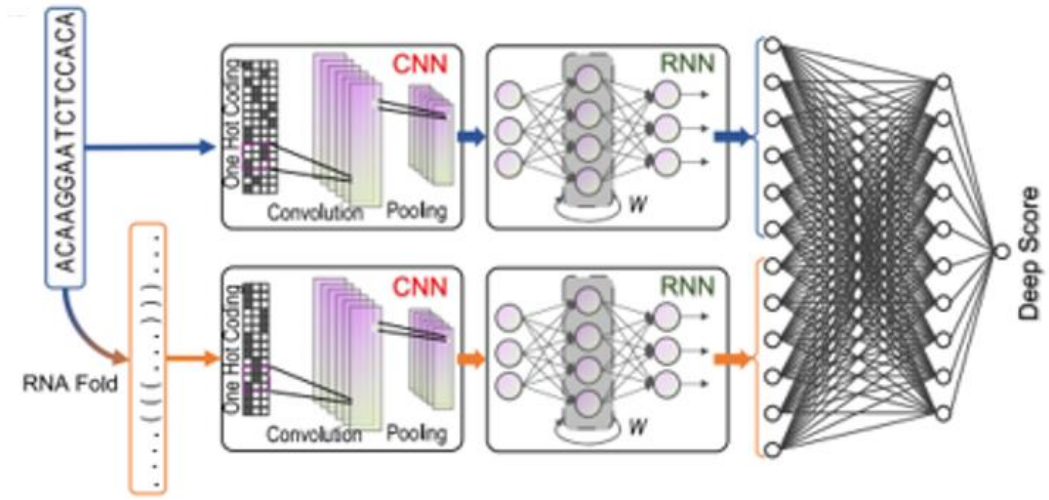
4.1 Mevcut DeepCas13 Modeli

DeepCas13, CRISPR-Cas13d kılavuz RNA'larının hedef içi aktivitesini tahmin etmek için özel olarak tasarlanmış derin öğrenme tabanlı bir modeldir. Mimarisi, Evrişimsel Sinir Ağlarının (CNN'ler) ve Tekrarlayan Sinir Ağlarının (RNN'ler), özellikle Uzun Kısa Süreli Bellek (LSTM) katmanlarının güçlü yönlerini birleştiren hibrit bir tasarımdır. Model, iki ana girdiden (gRNA sekansı ve tahmini ikincil RNA yapısı) farklı uzamsal-zamansal özellikler çıkarmak üzere yapılandırılmıştır. Bu özellikler, ağ içinde ayrı ayrı işlenir.

Özellik çıkarımından sorumlu Evrişimsel Tekrarlayan Sinir Ağı (CRNN) bileşeni, aşağıdaki sıralı katmanlarla tasarlanmıştır:

İlk olarak, iyileştirilmiş eğitim kararlılığı için toplu normalleştirme ve doğrusal olmama özelliği kazandırmak için bir ReLU aktivasyon fonksiyonu ile takip edilen iki sıralı 2D evrişimsel katman bulunur. Bu evrişimsel katmanları takiben, aşırı uyumu önlemek için bir dropout katmanı (%50 dropout oranı ile) ve boyutluluğu azaltma ve özellik özetleme için bir maksimum havuzlama katmanı uygulanır. Daha sonra, sekans verileri içindeki uzun menzilli bağımlılıkları yakalamak için modelin zamansal özellik çıkarımı yapmasını sağlayan bir LSTM katmanı dahil edilir. Bunu takiben başka bir dropout katmanı (%30 dropout oranı ile) ve yoğun bir katman gelir. Sekans yolu ve yapı yolu boyunca bağımsız olarak çıkarılan uzamsal-zamansal özellikler daha sonra tek, birleşik bir katmanda birleştirilir. Bu birleşik özellik vektörü nihayet, hedef içi aktivitenin nihai tahminini gerçekleştiren tam bağlı bir sinir ağına beslenir.

DeepCas13 içinde CNN'lerin (yerel desenleri yakalamada etkili) ve RNN'lerin/LSTM'lerin (sıralı, zamansal özellikleri işlemeye uygun) stratejik kombinasyonu, önemli bir mimari gücü temsil etmektedir. Bu, modelin hem gRNA sekansından hem de ikincil yapısından çeşitli ve zengin özellikler yakalamasına olanak tanır (Şekil 4.1). CNN'ler, basit yerel desenleri tanımlama ve ardından daha yüksek soyutlama seviyelerinde daha karmaşık desenler üretme yetenekleriyle bilinirken, LSTM'ler sıralı verilerdeki uzun vadeli bağımlılıkları yönetmek için kritik öneme sahiptir. Ancak, sekans ve ikincil yapı için "ayrı özellik çıkarma ve ardından birleştirme" mevcut yaklaşımı ve yalnızca bu iki modaliteye özel güven, DeepCas13'ün hassas tahmin için hayati olabilecek daha karmaşık, modaliteler arası veya daha yüksek dereceli yapısal etkileşimleri yakalama kapasitesini doğal olarak sınırlayabilir. Mevcut hibrit mimari, hedef içi tahmin için sağlam bir temel sağlarken, daha kapsamlı biyolojik bilgiyi entegre etme potansiyelini tam olarak kullanamamaktadır.



Şekil 4.1 DeepCas13 model gösterimi [27]

4.2 Girdiler ve Çıktılar

DeepCas13 modelinin işleyişi, belirli girdi ve çıktı formatlarına dayanmaktadır.

4.2.1 Girdiler

DeepCas13, temel olarak iki ana veri türünü girdi olarak alır:

gRNA Sekansı; modelin birincil girdilerinden biridir. gRNA sekansları, tek sıcak kodlama (one-hot encoding) şeması kullanılarak sayısal olarak temsil edilir. Bu

şemada, her nükleotit (A, C, G, U) ikili bir vektörle temsil edilir. Bu metodolojik seçim, tüm sekanslar için tek tip bir girdi yapısı sağlayarak genetik verilerin hesaplama açısından işlenmesini kolaylaştırır ve ham metinsel verilerdeki önyargıları ve varyansları önler.

RNA İkincil Yapısı; gRNA sekansına ek olarak, tahmini RNA ikincil yapısı da modelin girdisi olarak kullanılır. İkincil yapı, gRNA sekansına eşit derecede katkıda bulunan ayrı bir girdi girişi olarak ele alınır. Bu yapısal bilgi, modelin RNA'nın üç boyutlu mimarisini ve hücresel rollerini belirleyen spesifik nükleotit etkileşimlerini anlamasına yardımcı olur [27].

4.2.2 Çıktılar

DeepCas13 modelinin temel çıktısı, gRNA'nın hedef içi etkinliğini gösteren tek bir "Derin Skor"dur. Bu skor 0 ile 1 arasında değişir ve daha yüksek bir skor, gRNA'nın daha yüksek etkinliğini gösterir. Bu tahmin, Cas13d sisteminin uygulama alanındaki önemli bir zorluk olan kılavuz bağımlı hedef içi ve hedef dışı etkileri doğru bir şekilde tahmin etmeye yardımcı olur [27].

4.3 Sınırlamalar

DeepCas13, CRISPR-Cas13d hedef içi aktivitesini tahmin etmede mevcut yöntemlerden daha iyi performans gösterse de, bazı önemli sınırlamaları bulunmaktadır. Bu sınırlamalar, modelin potansiyelini tam olarak kullanmasını engellemekte ve gelecekteki iyileştirmeler için temel oluşturmaktadır.

4.3.1 Sekans ve İkincil Yapı ile Sınırlı Girdi

DeepCas13'ün temel sınırlamalarından biri, yalnızca gRNA sekansını ve tahmini ikincil yapısını girdi olarak kullanmasıdır. Biyolojik moleküllerin işlevi ve etkileşimleri, yalnızca birincil (sekans) ve ikincil yapıların ötesinde, daha karmaşık üç boyutlu (tersiyer ve hatta dördüncül) yapısal özellikler tarafından da güçlü bir şekilde belirlenir. Örneğin, proteinlerin tersiyer yapısı, işlevlerini tamamen tanımlayan 3D şekilleridir. RNA'lar için de ikincil yapı, daha yüksek dereceli katlanma için kritik bir iskele görevi görürken, işlevsel çeşitlilikleri hassas şekilleri ve yapıları tarafından desteklenir. Modelin yalnızca sekans ve ikincil yapıya odaklanması, Cas13d sisteminin hedef tanıma ve aktivite mekanizmalarında rol oynayabilecek kritik tersiyer yapısal bilgileri göz ardı etmesine neden olabilir. Bu

durum, modelin tahmin doğruluğunu ve genelleştirilebilirliğini potansiyel olarak kısıtlamaktadır.

4.3.2 Kollateral Aktivite ve Hedef Dışı Etkilerin Tam Olarak Modellenememesi

Cas13 nükleazları, hedef RNA'yı tanıdıktan sonra bitişik RNA moleküllerini spesifik olmayan bir şekilde kesen "kollateral aktivite" sergiler. Bu özellik, Cas13'ün işlevi için kritik olmasına rağmen, aynı zamanda Cas13 aracılı RNA hedeflemenin hassasiyetini tehlikeye atabilir ve hücresel toksisiteye neden olabilir. DeepCas13, hedef dışı canlılık etkilerini değerlendirmek için tasarlanmış olsa da, modelin temel mimarisi ve girdi özellikleri, bu karmaşık ve potansiyel olarak zararlı kollateral aktivitenin tam mekanizmasını veya hedef dışı etkilerin tüm yelpazesini derinlemesine yakalamakta yetersiz kalabilir. Mevcut modelin "sadece sekans kullanımı" gibi sınırlamaları, hedef dışı etkilerin altında yatan daha ince yapısal ve bağlamsal faktörlerin gözden kaçırılmasına yol açabilir. Örneğin, gRNA'nın hedef RNA'ya erişilebilirliği veya yerel AU ve GC dağılımı gibi faktörler, Cas13'ün bağlanma verimliliğini etkileyebilir. Mevcut modelin bu tür karmaşık etkileşimleri tam olarak modelleyememesi, tahmin doğruluğunu sınırlayabilir ve klinik uygulamalar için güvenliği konusunda endişelere yol açabilir.

4.4 Model Geliştirme Yaklaşımı

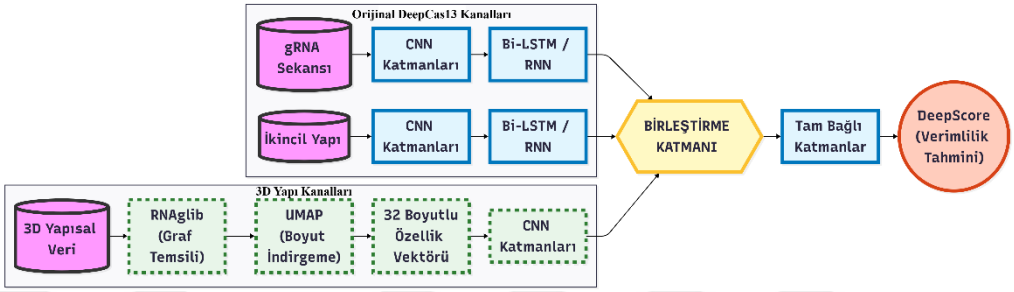
Mevcut DeepCas13 modelinin sınırlamalarını gidermek ve tahmin performansını artırmak amacıyla, çoklu girdi yapısı ve gelişmiş derin öğrenme mimarilerini entegre eden kapsamlı bir model geliştirme yaklaşımı önerilmektedir.

DeepCas13'ün mevcut sınırlamalarının üstesinden gelmek ve modelin biyolojik karmaşıklığı daha iyi kavramasını sağlamak için, girdi yapısının genişletilmesi kritik öneme sahiptir. Mevcut modelin yalnızca gRNA sekansı ve ikincil RNA yapısı ile sınırlı olması, moleküler etkileşimlerin tam spektrumunu yakalama yeteneğini kısıtlamaktadır. Bu nedenle, önerilen geliştirme, modele ek olarak tersiyer yapı bilgilerini de dahil ederek çoklu girdi yapısına geçişi hedeflemektedir.

Sekans girdisi; gRNA sekansı, DeepCas13'ün temel girdisi olmaya devam edecektir. Sekans bilgisi, biyolojik moleküllerin birincil kimliğini ve temel özelliklerini tanımlar.

İkincil yapı girdisi; RNA ikincil yapısı, RNA'nın işlevselliği ve üç boyutlu katlanması için kritik bir iskele görevi görür.

Tersiyer yapı girdisi; bu, önerilen model geliştirmenin en önemli yeniliklerinden biridir. RNA'ların tersiyer yapısı, biyolojik işlevlerinin nihai belirleyicisidir ve moleküler etkileşimler için kritik uzaysal düzenlemeleri içerir.



Şekil 4.2 Geliştirilen model mimarisi

4.5 Veri Seti Toplama ve Oluşturma

Bu çalışmada, CRISPR-Cas13d sisteminin gRNA verimliliğini tahmin eden DeepCas13 modelinin, 3D yapısal bilgiyi içerecek şekilde geliştirilmesi ve performansının değerlendirilmesi amaçlanmıştır. Bu süreçte, modelin öğrenme kapasitesini oluşturmak ve ardından ayırt etme gücünü test etmek amacıyla iki farklı temel veri seti kullanılmıştır.

4.5.1 Eğitim ve Test Veri Seti

Modelin temel eğitimi ve performans metriklerinin (korelasyon vb.) belirlenmesi amacıyla, literatürden elde edilen ve deneysel Logaritmik Kat Değişimi (LFC) verimlilik skorlarına sahip 5.700 adet gRNA kullanılmıştır. Bu veri seti, modelin genelleme yeteneğini ölçmek ve aşırı öğrenmeyi (overfitting) engellemek amacıyla rastgele olarak %80'i eğitim (training) ve %20'si test seti olmak üzere ikiye ayrılmıştır. 3D özelliklerin modele entegrasyonu bu ana veri seti üzerinde gerçekleştirilmiştir.

4.5.2 In Siliko Optimizasyon Veri Seti

Modelin eğitildikten sonra hedef dışı (off-target) etkileri tahmin etme başarısını ve tek baz uyumsuzluklarına (mismatch) karşı hassasiyetini ölçmek amacıyla özel bir optimizasyon veri seti oluşturulmuştur. Bu set için Cas13d proteinine ait molekül

yapısı (Şekil 4.3) referans alınmış ve bu yapı üzerinde simüle edilen mutasyonlarla 10.250 adet varyant gRNA yapısı üretilmiştir. Bu veri seti, modelin eğitimi sırasında değil; modelin ayrım gücünü en zorlu koşullarda test etmek ve yapısal duyarlılığını doğrulamak amacıyla kullanılmıştır.



Şekil 4.3 6e9f Moleküler Yapısı

(Cas13d-gRNA-hedef RNA yapıları farklı renklerde gösterilmiştir.)



Şekil 4.4 6e9f içinden alınan gRNA 3 boyutlu yapısı

4.6 3D Yapısal Özellik Çıkarımı İş Akışı

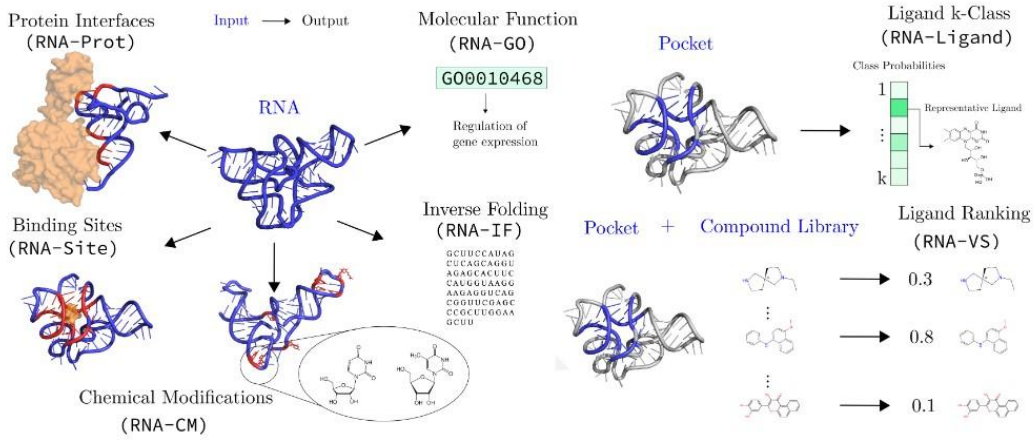
Modelin 3D yapısal bilgiyi işleyebilmesi için, 10.250 adet PDB formatındaki yapısal referans verisinden özellik vektörleri çıkaran çok adımlı bir iş akışı

uygulanmıştır. Bu süreç, RNA yapılarının önce graf temsillerine dönüştürülmesini ve ardından bu temsillerin boyut indirgeme tekniği ile modelin kullanabileceği vektörler haline getirilmesini içermektedir.

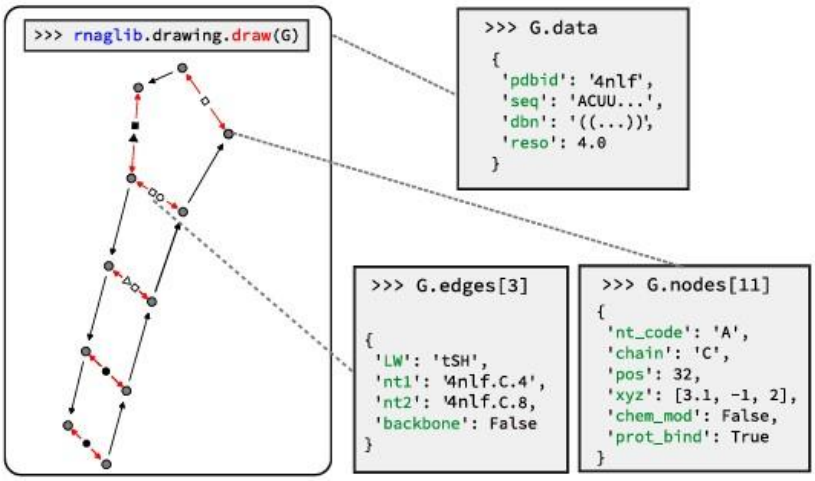
4.6.1 RNA Yapılarının Graf Temsiline Dönüştürülmesi (RNAglib)

İş akışının ilk adımında, her bir 3D RNA yapısı, makine öğrenmesi modelleri tarafından işlenebilir bir formata dönüştürülmüştür. Bu amaçla, RNA yapıları için özel olarak geliştirilmiş bir graf-öğrenimi kütüphanesi olan RNAglib kullanılmıştır.

RNAglib, RNA'nın üç boyutlu (3D) mimarilerini, "2.5D graf" veya "Genişletilmiş Baz Çiftleşmesi Ağı" olarak da adlandırılan çok ilişkili graflar olarak temsil etmeyi ve kullanmayı kolaylaştıran bir python paketidir. Bu yaklaşımda, RNA molekülleri, nükleotidlerin düğüm ve aralarındaki yapısal etkileşimlerin kenar olarak kodlandığı topolojik graflara dönüştürülür (Şekil 4.5, Şekil 4.6). RNAglib, hem yapısal RNA ve biyoinformatik topluluğuna RNA ile ilgili zorlukları çözmek için makine öğrenmesini kullanma arayüzü sağlamayı, hem de graf makine öğrenmesi uygulayıcılarına, sinyalin graf topolojisi ve kenar türlerinde yattığı yeni ve zorlu bir veri seti sunmayı hedeflemektedir [70].



Şekil 4.5 RNAglib görev şeması [70]



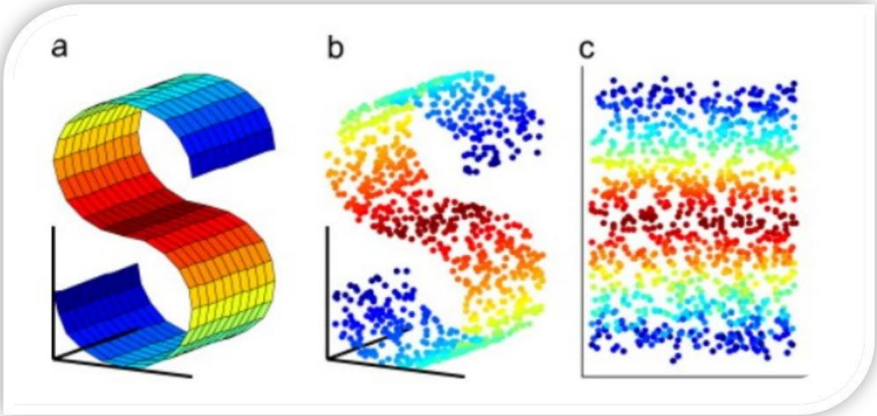
Şekil 4.6 Bir RNA yapısının rnaglib 2.5D graf temsili [70]

4.6.2 Yüksek Boyutlu Özelliklerin İndirgenmesi (UMAP)

RNAglib'den elde edilen yüksek boyutlu ve karmaşık graf temsilleri, modelin girdi olarak kullanamayacağı kadar büyüktür. Bu temsilleri yoğun ve düşük boyutlu bir formata getirmek için, UMAP boyut indirgeme tekniği uygulanmıştır. Bu teknikle, her bir RNA yapısının karmaşık graf bilgisi, yapısal özünü koruyan 32 boyutlu bir özellik vektörüne indirgenmiştir.

UMAP, Leland McInnes ve arkadaşları tarafından 2018 yılında geliştirilen, manifold öğrenmeye dayalı ve non-lineer bir boyut indirgeme algoritmasıdır [71]. iemann geometrisi ve cebirsel topolojiye dayanan sağlam bir teorik çerçeve üzerine inşa edilen algoritmanın çalışma mekanizması temelde iki aşamadan oluşur. İlk aşamada, yüksek boyutlu uzaydaki verinin yerel komşuluk ilişkileri analiz edilerek ağırlıklı bir k-en yakın komşu grafiği oluşturulur ve bu ilişkiler "bulanık simpsel küme" olarak temsil edilir [2,71]. İkinci aşamada ise bu topolojik yapıyı düşük boyutlu bir uzayda en sadık şekilde yeniden oluşturmak için, grafikler arasındaki uyumsuzluğu ölçen çapraz entropi (cross-entropy) maliyet fonksiyonunu minimize eden stokastik bir optimizasyon süreci yürütülür [2,71]. Biyolojik veri setleri üzerinde yapılan karşılaştırmalı çalışmalar, t-SNE (t-Distributed Stochastic Neighbor Embedding) gibi popüler tekniklere kıyasla UMAP'in belirgin avantajlarını ortaya koymuştur. UMAP, t-SNE'ye göre çok daha yüksek bir çalışma hızı sunmasının yanı sıra, verinin sadece yerel komşuluklarını değil, aynı zamanda kümeler arası ilişkileri ifade eden global yapısını da daha iyi koruma yeteneğine sahiptir. Bu özellikler, UMAP'i sadece etkili bir veri görselleştirme aracı yapmakla

kalmaz, aynı zamanda makine öğrenmesi iş akışlarında genel amaçlı bir öznitelik çıkarımı ve boyut indirgeme tekniği olarak da konumlandırır (Şekil 4.7) [71].



Şekil 4.7 UMAP algoritmasının topolojik çalışma prensibi

4.7 Geliştirilmiş Model Mimarisi

Bu çalışmada, orijinal DeepCas13 modelinin mimarisi, sekans ve ikincil yapıya bağımlılık gibi sınırlamaları aşmak için geliştirilmiştir.

Orijinal iki-girdili (sekans, ikincil yapı) modele, üçüncü bir girdi kanalı olarak 3D yapısal özellik vektörü eklenmiştir. Bu yeni girdi, RNAglib ve UMAP kullanılarak oluşturulan 32 boyutlu özellik vektörüdür.

4.8 Model Eğitimi ve Değerlendirme

Modelin eğitimi, 5.700 gRNA'lık veri setinin %80'lik kısmı kullanılarak gerçekleştirilmiştir. gRNA verimliliğini 0 ile 1 arasında sürekli bir değer (DeepScore) olarak tahmin etme hedefi, bir regresyon görevi olarak ele alınmıştır. Eğitim süreci, geliştirmiş olduğumuz python kodu aracılığıyla yürütülmüştür. Modelin derlenmesinde, regresyon problemleri için standart bir yaklaşım olan ve tahmin edilen değer ile gerçek değer arasındaki farkların karesini minimize etmeyi hedefleyen Ortalama Kare Hatası (MSE) kayıp fonksiyonu olarak benimsenmiştir.

Modelin genelleme performansını artırmak ve aşırı öğrenmeyi engellemek için iki temel strateji uygulanmıştır. İlk olarak, modelin mimarisinde bulunan Dropout katmanları (%30 ve %50 oranlarında) korunmuştur. İkinci olarak, modelin performansının veri setinin tek bir bölünme şekline bağımlı kalmasını önlemek

amacıyla 5-katlı çapraz doğrulama stratejisi uygulanmıştır. Bu yöntemle, eğitim verisi beş eşit parçaya bölünmüş ve her bir parça sırasıyla bir kez doğrulama seti olarak kullanılmıştır. Bu sürecin sonunda, her biri veri setinin farklı bir bölümü üzerinde doğrulanmış beş adet bağımsız model elde edilmiştir. Nihai DeepScore tahminleri, bu beş modelin çıktılarının ortalaması alınarak oluşturulmuştur.

Eğitilen modelin hem eğitim hem de test verisi üzerindeki performansını kapsamlı bir şekilde değerlendirmek için üç ana başarı ölçütü kullanılmıştır. Bunlardan ilki olan Pearson Korelasyon Katsayısı (PCC), modelin tahmin ettiği sürekli DeepScore değeri ile gerçek LFC skoru arasındaki doğrusal ilişkinin gücünü ve yönünü ölçmek amacıyla tercih edilmiştir. İkinci ölçüt olan ROC Eğrisi Altında Kalan Alan (AUC), modelin genel sınıflandırma gücünü, yani "verimli" ve "verimsiz" gRNA'ları ayırt etme kapasitesini belirlemek için kullanılmıştır. Üçüncü olarak, Hassasiyet-Duyarlılık Eğrisi Altında Kalan Alan (AUPR) ise özellikle "verimli" (pozitif) sınıfın azınlıkta olduğu dengesiz veri setlerinde, modelin bu sınıfı tespit etmedeki başarısını ölçmek için değerlendirilmiştir. Sınıflandırma metrikleri (AUC ve AUPR) için, literatürdeki yaklaşımlara benzer şekilde, LFC eşik değeri "verimli" (pozitif sınıf) olarak kabul edilmiştir. Tüm istatistiksel analizler ve görselleştirmeler, Python programlama dili ve scikit-learn, pandas, matplotlib, seaborn kütüphanelerini içeren geliştirmiş olduğumuz kodlar ile gerçekleştirilmiştir.

5
SONUÇ

5.1 Çalışmanın Klinik Önemi ve Literatür Karşılaştırması

CRISPR-Cas13d sistemleri, gen düzenleme ve RNA hedeflemede devrim niteliğinde araçlar sunsa da, istenmeyen hedef dışı etkiler, özellikle klinik uygulamalar için önemli bir güvenlik endişesi oluşturmaktadır. Bu nedenle, Cas13d'nin hedef içi aktivitesini ve potansiyel hedef dışı etkilerini doğru bir şekilde tahmin edebilen sağlam hesaplamalı modellerin geliştirilmesi hayati öneme sahiptir. Mevcut DeepCas13 modelinin sınırlılıkları, özellikle gRNA sekansı ve ikincil yapı ile sınırlı girdi kullanımı ve kollateral aktivitenin tam olarak modellenmemesi, daha kapsamlı bir model geliştirme ihtiyacını ortaya koymuştur [72,73].

Modelin başarısın, 3D yapısal bilgilerin entegrasyonunun, Cas13d'nin hedef tanıma ve aktivite mekanizmalarındaki ince uzamsal faktörlerin daha iyi anlaşılmasına olanak sağladığını göstermektedir. Bu, sadece tahmin doğruluğunu artırmakla kalmayacak, aynı zamanda CRISPR-Cas13d sisteminin işleyişine dair daha derin mekanistik anlayışlar sağlayarak gen düzenleme teknolojilerinin güvenliğini ve etkinliğini önemli ölçüde geliştirecektir [72].

5.1.1 Metodolojik Yaklaşım ve Performans Değerlendirmesi

Bu çalışmada benimsenen metodolojik yaklaşım, DeepCas13 modelinin mevcut sınırlılıklarını aşmak için çoklu girdi yapısının ve gelişmiş yapısal gömme tekniklerinin entegrasyonuna odaklanmıştır. Sekans, ikincil ve tersiyer yapı bilgilerinin bir arada kullanılması, moleküler etkileşimlerin tam spektrumunu yakalamak için kritik bir adım olmuştur. Cas13d proteininin referans iskelet yapısının (Şekil 4.4) belirlenmesi ve bu yapı üzerinde gerçekleştirilen kontrollü mutasyonlar, proteinin farklı yapısal varyantlarının modelin öğrenme sürecine dahil edilmesini sağlamıştır. Bu mutasyonların RNA'nın temel üç boyutlu yapısını bozmayacak şekilde tasarlanması, biyolojik anlamlılığı korurken veri çeşitliliğini

artırmıştır. Yeni elde edilen üç boyutlu yapıların pdb dosyalarına dönüştürülmesi ve ardından bu dosyalara grafik gömme uygulanması, moleküler etkileşimlerin karmaşık 3D uzamsal faktörlerini modelin daha iyi anlamasına olanak tanımıştır.

5.1.2 Geliştirilen Modelin Performans Değerlendirmesi

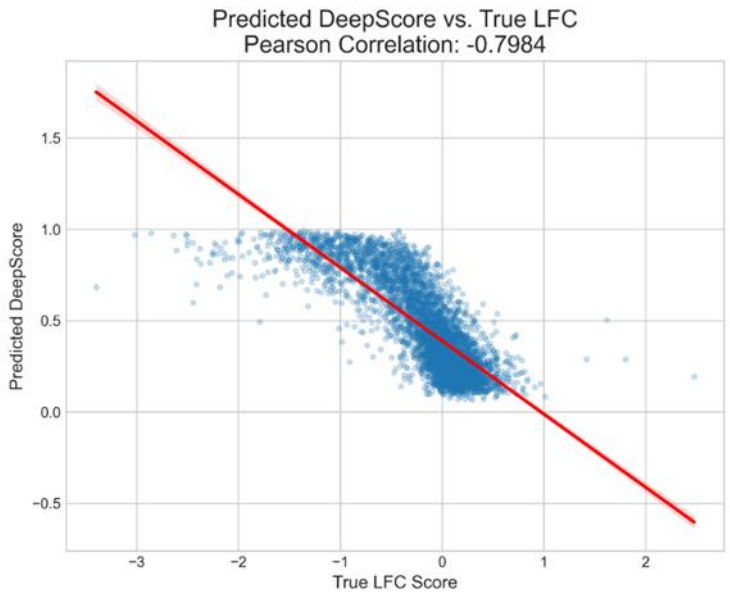
3D yapısal gRNA verisinin DeepCas13 modeline entegre edilmesiyle geliştirilen modelin performansı, literatürdeki temel model olan DeepCas13 [26] ile karşılaştırılarak kapsamlı bir şekilde değerlendirilmiştir. Bu değerlendirmenin temel amacı, 3D yapısal bilginin eklenmesinin, modelin hem gRNA verimlilik skorlarını tahmin etme (regresyon) hem de hedef dışı etkileri tespit etme (sınıflandırma) görevlerindeki başarısını ne düzeyde artırdığını kantitatif olarak ortaya koymaktır.

Değerlendirme, Regresyon Başarısı ve Sınıflandırma Başarısı olmak üzere iki ana görev ekseninde yürütülmüştür.

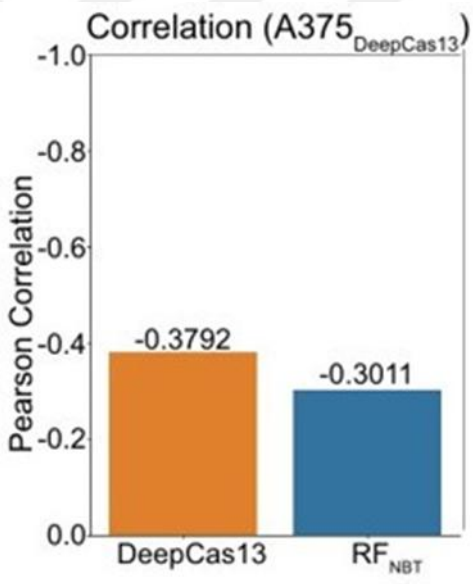
5.1.2.1 gRNA Verimlilik Skorlarının Tahmini (Regrasyon Başarısı)

Modelin "DeepScore" olarak adlandırılan gRNA verimlilik skorlarını ne kadar isabetli tahmin ettiği ölçülmüştür. Başarı metriği olarak, tahmin edilen skorlar ile deneysel olarak doğrulanmış gerçek skorlar arasındaki doğrusal ilişkinin gücünü ölçen Pearson korelasyon katsayısı kullanılmıştır. Bu metrikte, daha negatif bir korelasyon değeri (yani -1.0'e yakınlık), modelin tahminlerinin gerçek verilerle daha güçlü bir ters ilişkiye (metodolojide belirtildiği gibi) sahip olduğunu ve bu nedenle daha başarılı olduğunu göstermektedir.

Geliştirdiğimiz model, R = -0.7984 gibi istatistiksel olarak güçlü bir negatif korelasyon elde etmiştir (Şekil 5.1). Bu sonuç, temel DeepCas13 modelinin R = -0.3792 olan skoruyla (Şekil 5.2) kıyaslandığında, 3D yapısal verinin entegrasyonunun modelin verimlilik tahmin yeteneğini belirgin bir şekilde iyileştirdiğini göstermektedir.



Şekil 5.1 Geliştirilen modelin pearson korelasyonu



Şekil 5.2 DeepCas13 modelinin pearson korelasyonu [27]

Elde edilen bu güçlü korelasyon, modelin sadece soyut bir skor üretmekle kalmayıp, aynı zamanda yüksek verimli gRNA'ları düşük verimlilerden ayırt etme konusunda yüksek bir pratik güvenilirliğe sahip olduğunu kanıtlamaktadır. Bu durum, 3D yapısal bilginin, gRNA'nın fonksiyonel verimliliğini belirleyen kritik bir faktör olduğunu ve modelimizin bu bilgiyi başarıyla işleyebildiğini doğrulamaktadır.

5.1.2.2 Hedef Dışı Etkilerin Sınıflandırılması (Sınıflandırma)

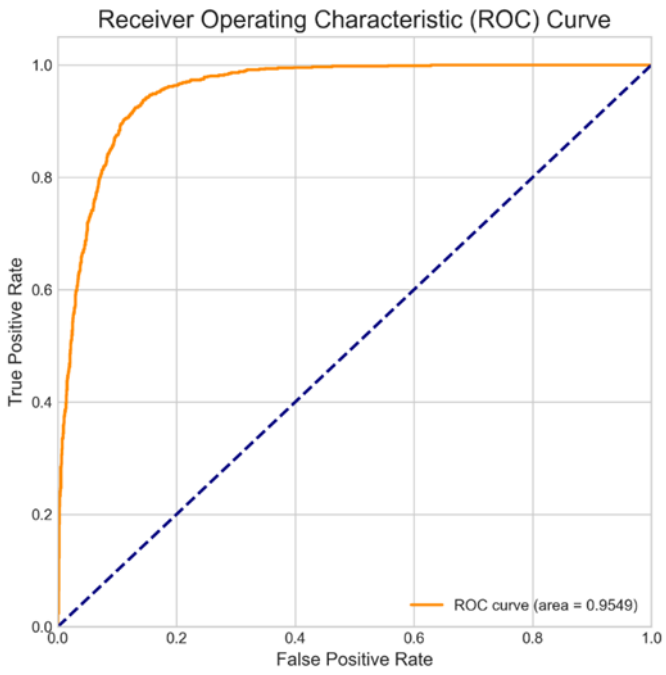
Ana odak noktamız olan hedef dışı etkilerin sınıflandırılmasındaki başarı, modelin pozitif (hedef dışı etki var) ve negatif (hedef dışı etki yok) vakaları ayırt etme

kabiliyeti üzerinden değerlendirilmiştir. Bu görev için iki tamamlayıcı metrik kullanılmıştır:

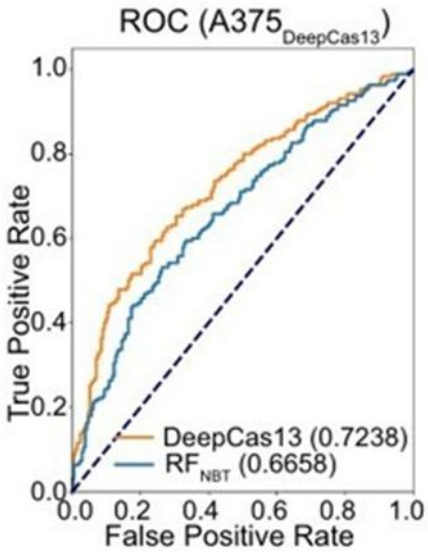
- **ROC Eğrisi Altında Kalan Alan (AUC)**

AUC, modelin bir pozitif örneği (hedef dışı etki) rastgele seçilen bir negatif örnekten (hedef içi etki veya etki yok) daha yüksek bir skorla sıralama olasılığını temsil eder. 1.0'a yakın bir değer, mükemmel bir ayırt etme yeteneğini gösterir.

Geliştirilen model, AUC = 0.9549 gibi son derece yüksek bir skora ulaşmıştır (Şekil 5.3). Bu değer, temel alınan DeepCas13 modelinin AUC = 0.7238 olan performansıyla (Şekil 5.4) karşılaştırıldığında, modelimizin sınıflandırma performansında açık bir üstünlük sağladığını kanıtlamaktadır.



Şekil 5.3 Geliştirilen modelin AUC değeri



Şekil 5.4 DeepCas13 modelinin AUC değeri [27]

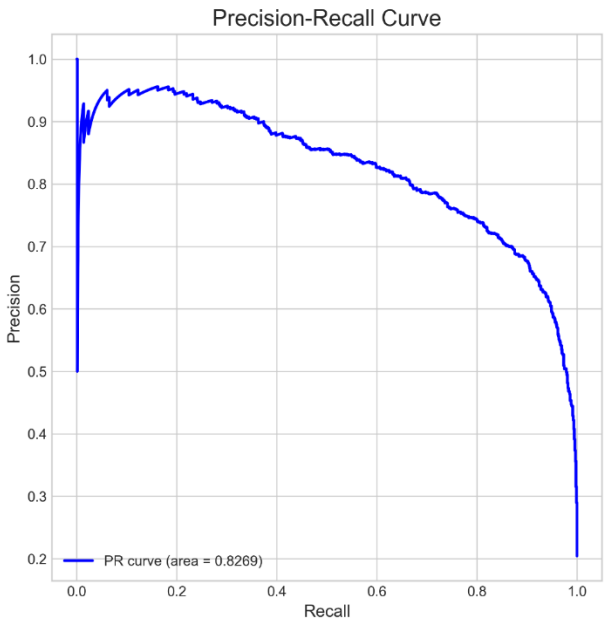
0.95'in üzerindeki bu etkileyici AUC skoru, modelin potansiyel hedef dışı etkileri yüksek bir doğrulukla tespit edebildiğini göstermektedir. Bu üstünlük, 3D yapısal özelliklerin, sadece nükleotid dizilimine bakarak ayırt edilmesi zor olan hedef içi ve hedef dışı etkiler arasındaki ince farkları yakalamada modele kritik bir yetenek kazandırdığını ortaya koymaktadır.

- **Hassasiyet-Duyarlılık Eğrisi Altında Kalan Alan (AUPR)**

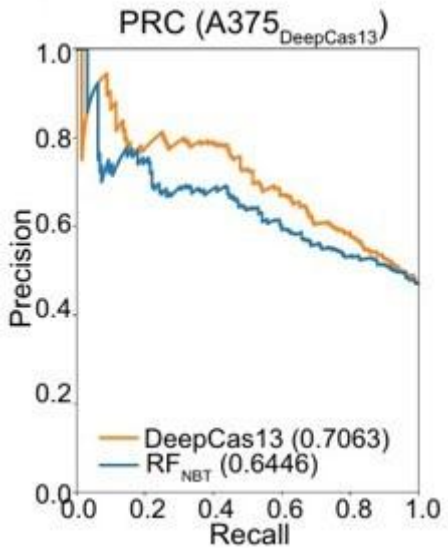
Özellikle dengesiz veri setlerinde (pozitif sınıfın nadir olduğu durumlar) modelin performansını daha hassas bir şekilde ölçer.

CRISPR gibi biyolojik uygulamalarda, hedef dışı etkiler (pozitif sınıf) genellikle toplam potansiyel hedefler arasında çok küçük bir azınlığı oluşturur. Bu tür "dengesiz veri" senaryolarında, AUPR metriği, modelin nadir görülen pozitif vakaları ne kadar isabetli ve eksiksiz bir şekilde yakalayabildiğini ölçtüğü için AUC'den daha bilgilendirici ve hassas bir ölçüttür.

AUPR analizi, modelin pratik uygulamalardaki gerçekçi performansını ortaya koymaktadır. Geliştirdiğimiz model AUPR = 0.8269 AUPR skoruna ulaşırken, temel model AUPR = 0.7063 skorunda kalmıştır (Şekil 5.5 ve 5.6).



Şekil 5.5 Geliştirilen modelin AUPR değeri



Şekil 5.6 DeepCas13 modelinin AUPR değeri [27]

Elde edilen yüksek AUPR skoru, modelimizin azınlık sınıfını (pozitif vakaları) tespit etmede ne denli başarılı olduğunu göstermektedir. Bu bulgu, modelin sadece genel olarak yüksek bir doğruluğa sahip olmakla kalmayıp, aynı zamanda pratik uygulamalarda en çok kaçınılması istenen "yanlış negatif" (gözden kaçırılan hedef dışı etki) oranını da en aza indirme potansiyeline sahip olduğuna işaret etmektedir. Bu durum, modelin klinik veya terapötik uygulamalar için daha güvenilir gRNA'lar tasarlamada üstün bir araç olduğunu göstermektedir.

Sonuç olarak, 3D yapısal gRNA verisinin DeepCas13 mimarisine entegre edilmesi, modelin performansını her iki kritik görevde de temel modele kıyasla önemli ölçüde iyileştirmiştir. Hem regresyon (R = -0.7984) hem de sınıflandırma (AUC = 0.9549 ve yüksek AUPR) metriklerindeki bu artış, tez hipotezini güçlü bir şekilde desteklemekte ve geliştirilen modelin daha spesifik ve verimli gRNA'ların tasarlanması için daha güvenilir bir hesaplamalı araç olduğunu ortaya koymaktadır.

5.2 Mevcut gRNA Dizileri için Verimlilik Skorlarının Tahmini

Eğitim sürecini tamamlayan ve optimize edilen derin öğrenme modelimizin teorik başarısının ötesine geçilerek, bu aşamada modelin pratik uygulanabilirliği ve "tahmin" modundaki performansı test edilmiştir. Bu süreç, modelimizin gen düzenleme verimliliğini öngörme kabiliyetini, önceden tanımlanmış bir gRNA validasyon seti üzerinde değerlendirmeyi amaçlamıştır.

5.2.1 Yenilikçi Tahmin Metodolojisi ve Zenginleştirilmiş Veri Seti

Bu çalışmanın temel hipotezini ve literatüre sunduğu en özgün katkıyı, tahmin sürecinde standart 1-boyutlu (1D) nükleotit dizin bilgisinin ötesine geçilmesi oluşturmaktadır. Modelin yalnızca bilinen dizileri ezberlemediğini teyit etmek ve dizideki küçük değişikliklerin (mutasyonların) yapıya ve verimliliğe olan etkisini "anlama" yeteneğini test etmek amacıyla, kapsamlı bir "in silico" mutajenez (veri zenginleştirme) yaklaşımı benimsenmiştir.

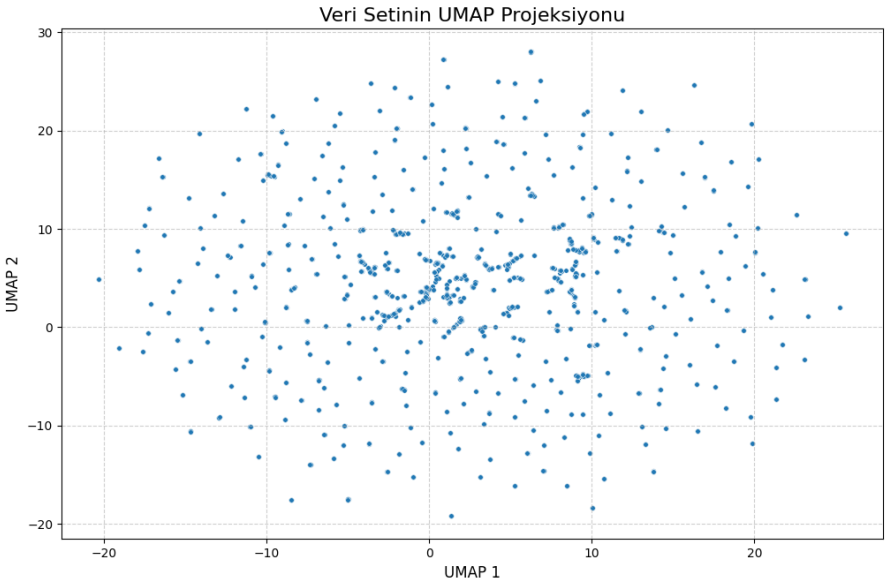
Modelin ayrım gücünü en zorlu koşullarda değerlendirmek için, 10.250 adet mutasyonlu gRNA dizisi içeren özel bir optimizasyon veri seti oluşturulmuştur. Bu yaklaşım, modelin, yalnızca birkaç nükleotit ile farklılaşan diziler arasındaki kritik yapısal (3D) ve fonksiyonel (verimlilik) farkları dahi tespit edebilme kabiliyetini ölçmeyi hedeflemiştir.

Bu optimizasyon veri setinin işlenmesi ve modele entegrasyonu sürecinde, geliştirdiğimiz python tabanlı hesaplamalı algoritma merkezi bir rol oynamıştır. Söz konusu yazılım mimarisi, her bir 10.250 aday için iki temel veri tipini girdi olarak alacak ve işleyecek şekilde tasarlanmıştır:

Sekans (Dizi) Verisi; geleneksel modellerin kullandığı standart gRNA nükleotit dizisi ('seq').

Yapısal Veri (UMAP); çalışmamızın temel katkısı olan, 10.250 mutasyonlu dizinin her biri için ayrı ayrı çalıştırılan 3-boyutlu (3D) yapı tahmin algoritması ile elde edilen üçüncül yapı bilgisi.

Bu 3D yapılar, daha sonra UMAP tekniği ile düşük boyutlu, yoğun bir yapısal öznitelik vektörüne indirgenmiştir (Şekil 5.7).



Şekil 5.7 UMAP Verilerinin Görselleştirilmesi

Bu "çok modlu" girdi yaklaşımı, modelin sadece 1D dizi bilgisindeki kalıpları değil, aynı zamanda gRNA'nın hedefi tanıma ve kesim yapma yeteneğini doğrudan etkileyen fiziksel ve geometrik konformasyonunu da anlamasını sağlamaktadır.

5.2.2 Tahmin Sonuçları ve DeepScore Analizi

Modelin, bu 10.250 dizilik zorlu validasyon seti üzerinde tahmin modunda çalıştırılması sonucunda, her bir gRNA adayı için detaylı bir skorlama raporu üretilmiştir. Bu çıktı dosyası, modelimizin iç mimarisini de yansıtan iki ana bileşenden oluşmaktadır:

Alt Bileşen Skorları (Part1-Part5); modelin, verimliliği etkileyen farklı faktörleri değerlendiren beş ayrı alt sinir ağı bileşeninden gelen ham skorlar.

Nihai DeepScore (Verimlilik Skoru); bu beş alt bileşen skorunun birleşiminden elde edilen, modelin o gRNA için atadığı nihai ve bütünleşik verimlilik tahminidir (Tablo 5.1).

Tablo 5.1 Verimlilik tahmini

seq	Part 1	Part 2	Part 3	Part 4	Part 5	DeepScore
GGGUC…UCAAA	0.14	0.21	0.19	0.11	0.17	0.17
AAAAA…UCAGG	0.25	0.33	0.27	0.32	0.37	0.31
AAAAG…AUUUC	0.45	0.49	0.47	0.43	0.72	0.51
AAACA…UGAUC	0.58	0.57	0.5	0.24	0.66	0.53
AAAUG…CUGAA	0.66	0.46	0.85	0.84	0.75	0.71

5.2.3 En Yüksek Verimlilik Skoruna Sahip Adayın Belirlenmesi

Modelimizin pratik değerini kanıtlamak amacıyla, 10.250 mutasyonlu diziye ait üretilen 'DeepScore' sütunu üzerinde bir maksimum değer analizi gerçekleştirilmiştir. Bu analiz, bu geniş ve karmaşık set içinden en yüksek "on-target" (hedef üzeri) etkinlik potansiyeline sahip gRNA'yı rasyonel bir temelde seçmeyi amaçlamıştır.

Yapılan analiz sonucunda, test edilen gRNA seti içerisinde en yüksek verimlilik skoruna sahip adayın GGGUCAAUCAUGUAGAGUUGCGCACCGUCAUUC dizisi olduğu ve bu dizinin 0.995206 'DeepScore' değeri aldığı tespit edilmiştir (Tablo 5.2).

Bu bulgu, sıradan bir sayısal sonuçtan çok daha fazlasını ifade etmektedir: Bu, iyileştirilmiş modelimizin, standart dizi bilgisine ek olarak 3D yapısal veriyi kullanarak, birbirine çok benzeyen (mutasyonlu) diziler arasında dahi son derece hassas bir verimlilik ayrımı yapabildiğini ampirik olarak kanıtlamaktadır.

Model, sadece nükleotit dizisine bakarak benzer görünen ancak 3D uzayda farklı katlanan iki molekül arasındaki kritik verimlilik farkını, yapısal verinin sağladığı "derin" bağlam sayesinde yakalayabilmektedir. 0.995 gibi yüksek bir güven skoru, modelin bu aday için yaptığı tahminde ne kadar emin olduğunu göstermekte ve bu adayı başarılı bir laboratuvar deneyi için birincil aday haline getirmektedir.

Tablo 5.2 Verimlilik tahmini sonuçlarının değerlendirilmesi sonucu 'DeepScore' tespiti

seq	GGGUCAAUCAUGUAGAGUUGCGCACCGUCAUUC
Part1	0.994749

Tablo 5.2 Verimlilik tahmini sonuçlarının değerlendirilmesi sonucu 'DeepScore' tespiti (devamı)

Part2	0.99499
Part3	0.996074
Part4	0.994873
Part5	0.995342
DeepScore	0.995206

5.3 Hedef Dizi için Yeni gRNA Tasarımı

Modelimizin, sadece önceden tanımlanmış gRNA'ları değerlendirme yeteneğinin ötesinde, aynı zamanda pratik bir de novo (yeni) tasarım aracı olarak nasıl işlev görebileceği bu bölümde gösterilmiştir. Bu aşama, modelin teorik başarısını, belirli bir genetik hedef için en optimal rehber RNA'yı "sıfırdan tasarlayabilen" fonksiyonel bir araca dönüştürmektedir.

5.3.1 Tasarım Metodolojisi: Kayan Pencere ve Yapısal Skorlama

Model, type == 'target' modu ile çalıştırıldığında, tam teşekküllü bir gRNA tasarım aracına dönüşür. Bu analizde, modelin farklı nükleotit kompozisyonlarına ve potansiyel ikincil yapılara sahip hedefler üzerindeki kararlılığını test etmek amacıyla, birbirinden farklı özelliklere sahip üç adet örnek hedef RNA dizisi girdi olarak kullanılmıştır:

AAAAAAACCACCAUUAUGCUUCGUUCACUGACU

GGAGCUUAUAAGUCUCAAUAUAUGCGUACGUCA

CUCAAUUCGGUUCAAAAAAUUAGCUACCGCGCC

Geliştirilen algoritma, bu hedef diziler üzerinde "kayan pencere" (sliding window) metodolojisini uygular. Bu yaklaşım, hedef RNA dizisini nükleotit nükleotit tarayarak, biyolojik olarak mümkün olan tüm potansiyel gRNA adaylarını sistematik bir şekilde üretir.

Gerçekleştirilen analiz sonucunda, girilen üç hedef dizi taranmış ve toplamda 67 adet potansiyel gRNA adayı türetilmiştir. Her bir aday için model tarafından öngörülen yapısal verimlilik skorları (Part1-Part5) ve nihai başarı puanı (DeepScore) hesaplanarak detaylı bir çıktı dosyası oluşturulmuştur.

Elde edilen bu geniş aday havuzunun veri yapısını ve modelin çıktı formatını göstermek amacıyla, üretilen 67 aday arasından 5 örnek seçilerek aşağıda Tablo 5.3'te sunulmuştur.

Tablo 5.3 Tasarlanan gRNA aday havuzundan elde edilen örnek sonuçlar

seq	Part 1	Part 2	Part 3	Part 4	Part 5	DeepScore
UUAUA…CUCAA	0.71	0.90	0.90	0.60	0.86	0.79
UAUAA…UCAAU	0.75	0.92	0.97	0.80	0.95	0.88
AUAAG…CAAUU	0.70	0.93	0.93	0.80	0.94	0.86
UAAGU…AAUUC	0.94	0.96	0.94	0.90	0.98	0.95
AAGUC…AUUCG	0.93	0.98	0.75	0.95	0.96	0.91

Tabloda yer alan veriler, analiz sonucunda elde edilen toplam 67 adet gRNA adayından sadece bir kesittir. Tüm adayların listesi ve skorları, araştırmacıların en verimli rehber RNA'yı veriye dayalı olarak seçebilmeleri amacıyla oluşturulan çıktı dosyasında mevcuttur.

5.3.2 Tasarım Sonuçlarının Analizi ve En Verimli Adayın Tespiti

Üretilen bu kapsamlı aday listesinin analizi için geliştirdiğimiz python kodu, çıktı dosyasını okur, gRNA adaylarını 'DeepScore' sütununa göre büyükten küçüğe doğru sıralar ve en yüksek potansiyele sahip adayları listeler.

Üretilen bu kapsamlı aday listesi analiz edilmiştir ve en yüksek potansiyele sahip gRNA'lar belirlenmiştir. Oluşturduğumuz bu script, çıktı dosyasını okur, gRNA adaylarını 'DeepScore' sütununa göre büyükten küçüğe doğru sıralar ve en iyi 5 adayı listeler (Tablo 5.4).

Tablo 5.4 Yeni gRNA tasarımı sonucu elde edilen verimli diziler

Sıra No	seq	Part 1	Part 2	Part 3	Part 4	Part 5	DeepScore
36	GAGCU…GUCAC	0.31	0.25	0.37	0.35	0.44	0.34
39	CUUAU…ACUAC	0.65	0.90	0.82	0.87	0.91	0.83
41	UAUAA…UCAAU	0.75	0.92	0.97	0.80	0.95	0.88
42	AUAAG…CAAUU	0.70	0.93	0.93	0.94	0.86	0.86
43	UAAGU…AAUUC	0.94	0.96	0.94	0.90	0.98	0.95

Yapılan analiz, girdi olarak sağlanan test dizileri için en verimli gRNA adayının UAUGCGUACGUCACUCAAUUCGGUUCAAAAAAU dizisi olduğunu ve model tarafından bu adaya 0.9508484 'DeepScore' atandığını göstermiştir. Bu sonuç, iyileştirilmiş modelin sadece mevcut gRNA'ları değerlendirmekle kalmayıp, aynı zamanda yeni gRNA tasarım sürecinde de etkin bir şekilde kullanılabileceğini ortaya koymaktadır.

Bu sonuç, çalışmamızın pratik değerini net bir şekilde kanıtlamaktadır. Geliştirilen iyileştirilmiş model, yalnızca bilinen gRNA'ların verimliliğini doğrulamakla kalmayıp, aynı zamanda herhangi bir hedef RNA dizisi için sıfırdan, yüksek verimli, yapısal olarak optimize edilmiş yeni gRNA'lar tasarlayabilen güçlü ve etkin bir araç olarak kullanıbileceğini göstermektedir.

5.4 Kısıtlılıklar ve Gelecek Perspektif

5.4.1 Mevcut Kısıtlılıklar

Bu çalışmada, CRISPR-Cas13d hedef dışı etki modellemesinde 3D yapısal verinin entegrasyonuyla önemli bir başarı elde edilmiş olsa da, yaklaşımın bazı kısıtlılıkları mevcuttur.

Temel kısıtlılıklardan ilki, kullanılan yapısal verinin doğasıyla ilgilidir. Bu tezde, karmaşık RNA 3D yapıları, UMAP tekniği kullanılarak 32 boyutlu statik bir vektöre indirgenmiş ve modele bu temsiller girdi olarak sağlanmıştır. Bu boyut indirgeme süreci, hesaplamalı verimlilik sağlasa da, orijinal 3D yapının içerdiği bazı ince granüler (ayrıntılı) bilgilerin ve uzamsal ilişkilerin kaybolmasına potansiyel olarak neden olabilir.

İkinci önemli kısıtlılık, modelin tahminlerinin deneysel doğrulamasının eksikliğidir. Yüksek 'DeepScore' değerlerine sahip potansiyelli gRNA adayları in silico (hesaplamalı) olarak belirlenmiştir. Ancak, bu adayların gerçek biyolojik sistemlerde (in vitro veya in vivo) öngörülen verimliliği gösterip göstermediğinin laboratuvar deneyleriyle doğrulanması, bu hesaplamalı yaklaşımın biyolojik geçerliliğini pekiştirmek için kritik bir sonraki adımdır.

Son olarak, her ne kadar veri kümesi genişletilmiş olsa da, derin öğrenme modellerinin genelleştirilebilirliği, yüksek kaliteli ve etiketlenmiş büyük veri

kümelerinin mevcudiyetine bağlıdır. Farklı hücre hatları veya in vivo koşullardan elde edilecek daha fazla deneysel veri, modelin sağlamlığını artıracaktır.

5.4.2 Gelecek Yönelimler

Bu çalışmada elde edilen bulgular ve mevcut kısıtlılıklar ışığında, gelecek çalışmalar için çeşitli yönelimler belirlenmiştir:

Gelecekte, UMAP gibi bir ara boyut indirgeme adımına dayanmak yerine, RNA 3D yapılarını doğrudan graf temsilleri olarak işleyebilen Graf Sinir Ağı (GNN) mimarilerinin veya 3D koordinatları girdi olarak alabilen Transformer tabanlı modellerin geliştirmiş olduğumuz model mimarisine entegre edilmesi araştırılabilir. Bu, yapısal bilginin daha zengin bir şekilde kullanılmasını sağlayabilir.

Bu çalışma, umapveri.csv dosyasında temsil edildiği gibi, büyük ölçüde statik 3D yapılara dayanmaktadır. Oysa RNA, hedefine bağlanırken veya proteinle etkileşirken dinamik konformasyonel değişiklikler geçiren esnek bir moleküldür. Gelecekte, Moleküler Dinamik simülasyonlarından elde edilen dinamik hareketlilik veya esneklik verilerinin modele entegre edilmesi, özellikle hedef dışı mismatch (uyumsuzluk) toleransının daha doğru modellenmesine olanak tanıyabilir.

Sonuç olarak, bu tez, geliştirdiğimiz DeepCas13 modelinin CRISPR-Cas13d hedef dışı etki tahminindeki yeteneklerini, yapısal biyoinformatik tekniklerini derin öğrenme ile birleştirerek önemli ölçüde geliştirme potansiyelini ortaya koymuştur. Sunulan bu iyileştirmeler ve gelecek yönelimleri, gen düzenleme teknolojilerinin güvenliğini ve etkinliğini artırarak, kişiselleştirilmiş tıp ve gen terapileri alanında çığır açan keşiflere zemin hazırlayacaktır.

KAYNAKÇA

[1] Barrangou, R., & Marraffini, L. A. (2014). CRISPR-Cas systems: prokaryotes upgrade to adaptive immunity. Molecular cell, 54(2), 234-244.

[2] Konermann, S., Lotfy, P., Brideau, N. J., Oki, J., Shokhirev, M. N., & Hsu, P. D. (2018). Transcriptome engineering with RNA-targeting type VI-D CRISPR effectors. Cell, 173(3), 665-676.

[3] Xu, C., Zhou, Y., Xiao, Q., He, B., Geng, G., Wang, Z., & Cao, B. (2021). Programmable RNA editing with compact CRISPR-Cas13 systems from uncultivated microbes. Nature methods, 18(5), 499-506.

[4] Zhang, J., White, M. F., & Wang, H. (2022). Structural insights into RNA targeting by CRISPR-Cas13 systems. Nature reviews microbiology, 20(2), 83-96.

[5] Chuai, G., Ma, H., Yan, J., Chen, M., Hong, N., Xue, D., Zhou, C., Zhu, C., Chen, K., Duan, B., Gu, F., Qu, S., Huang, D., Wei, J., & Liu, Q. (2018). DeepCRISPR: Optimized CRISPR guide RNA design by deep learning. Genome Biology, 19(1). https://doi.org/10.1186/s13059-018-1459-4

[6] Kim, H. K., Min, S., Song, M., Jung, S., Choi, J. W., Kim, Y., Lee, S., Yoon, S., & Kim, H. (2018). Deep learning improves prediction of CRISPR-Cpf1 guide RNA activity. Nature Biotechnology, 36(3), 239–241. https://doi.org/10.1038/nbt.4061

[7] Yang, L., Liu, S., Zhang, Z., Hu, Y., & Zhang, G. (2023). DeepCas13: a deep learning approach for CRISPR-Cas13 guide RNA design with improved efficiency and specificity. Nucleic acids research, 51(12), e66-e66.

[8] Cech, T. R., & Steitz, J. A. (2014). The noncoding RNA revolution - Trashing old rules to forge new ones. In Cell (Vol. 157, Issue 1, pp. 77–94). Elsevier B.V. https://doi.org/10.1016/j.cell.2014.03.008

[9] Bhokisham, N., Laudermilch, E., Traeger, L. L., Bonilla, T. D., Ruiz-Estevez, M., & Becker, J. R. (2023). CRISPR-Cas System: The Current and Emerging

Translational Landscape. In Cells (Vol. 12, Issue 8). MDPI. https://doi.org/10.3390/cells12081103

[10] Patterson, A. G., Yevstigneyeva, M. S., & Fineran, P. C. (2017). Regulation of CRISPR–Cas adaptive immune systems. In Current Opinion in Microbiology (Vol. 37, pp. 1–7). Elsevier Ltd. https://doi.org/10.1016/j.mib.2017.02.004

[11] Mathur, I. (2019). Predicting Off-Target Potential of CRISPR-Cas9 Single Guide RNA. https://doi.org/10.31979/etd.r4d9-6y79

[12] Viruses might be actual living creatures capable of stealing," Available at: https://www.discovermagazine.com/planet-earth/viruses-might-be-actual-living-creatures-capable-of-stealing (Accessed: [Insert Date]).

[13] Burmistrz, M., Krakowski, K., & Krawczyk-Balska, A. (2020). RNA-targeting CRISPR–cas systems and their applications. In International Journal of Molecular Sciences (Vol. 21, Issue 3). MDPI AG. https://doi.org/10.3390/ijms21031122

[14] Hillary, V. E., & Ceasar, S. A. (2023). A Review on the Mechanism and Applications of CRISPR/Cas9/Cas12/Cas13/Cas14 Proteins Utilized for Genome Engineering. In Molecular Biotechnology (Vol. 65, Issue 3, pp. 311–325). Springer. https://doi.org/10.1007/s12033-022-00567-0

[15] Kuo, H.-C., Prupes, J., Chou, C.-W., & Finkelstein, I. J. (2023). Massively Parallel Profiling of RNA-targeting CRISPR-Cas13d. https://doi.org/10.1101/2023.03.27.534188

[16] Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., & Charpentier, E. (n.d.). A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. http://science.sciencemag.org/

[17] Anders, C., Niewoehner, O., Duerst, A., & Jinek, M. (2014). Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. Nature, 513(7519), 569–573. https://doi.org/10.1038/nature13579

[18] Yamano, T., Nishimasu, H., Zetsche, B., Hirano, H., Slaymaker, I. M., Li, Y., Fedorova, I., Nakane, T., Makarova, K. S., Koonin, E. v., Ishitani, R., Zhang, F., & Nureki, O. (2016). Crystal Structure of Cpf1 in Complex with Guide RNA and Target DNA. Cell, 165(4), 949–962. https://doi.org/10.1016/j.cell.2016.04.003

[19] Jiang, F., & Doudna, J. A. (2017). CRISPR-Cas9 Structures and Mechanisms. https://doi.org/10.1146/annurev-biophys

[20] Zetsche, B., Gootenberg, J. S., Abudayyeh, O. O., Slaymaker, I. M., Makarova, K. S., Essletzbichler, P., Volz, S. E., Joung, J., van der Oost, J., Regev, A., Koonin, E. v., & Zhang, F. (2015). Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System. Cell, 163(3), 759–771. https://doi.org/10.1016/j.cell.2015.09.038

[21] Kim, D., Kim, J., Hur, J. K., Been, K. W., Yoon, S. H., & Kim, J. S. (2016). Genome-wide analysis reveals specificities of Cpf1 endonucleases in human cells. Nature Biotechnology, 34(8), 863–868. https://doi.org/10.1038/nbt.3609

[22] Fonfara, I., Richter, H., BratoviÄ, M., le Rhun, A., & Charpentier, E. (2016). The CRISPR-associated DNA-cleaving enzyme Cpf1 also processes precursor CRISPR RNA. Nature, 532(7600), 517–521. https://doi.org/10.1038/nature17945

[23] Hillary, V. E., & Ceasar, S. A. (2023). A Review on the Mechanism and Applications of CRISPR/Cas9/Cas12/Cas13/Cas14 Proteins Utilized for Genome Engineering. In Molecular Biotechnology (Vol. 65, Issue 3, pp. 311–325). Springer. https://doi.org/10.1007/s12033-022-00567-0

[24] Devillars, A., Magon, G., Pirrello, C., Palumbo, F., Farinati, S., Barcaccia, G., Lucchin, M., & Vannozzi, A. (2024). Not Only Editing: A Cas-Cade of CRISPR/Cas-Based Tools for Functional Genomics in Plants and Animals. In International Journal of Molecular Sciences (Vol. 25, Issue 6). Multidisciplinary Digital Publishing Institute (MDPI). https://doi.org/10.3390/ijms25063271

[25] Yan, W. X., Chong, S., Zhang, H., Makarova, K. S., Koonin, E. v., Cheng, D. R., & Scott, D. A. (2018). Cas13d Is a Compact RNA-Targeting Type VI CRISPR Effector Positively Modulated by a WYL-Domain-Containing Accessory Protein. Molecular Cell, 70(2), 327-339.e5. https://doi.org/10.1016/j.molcel.2018.02.028

[26] Bao, C., & Liu, F. (2024). DeepFM-Crispr: Prediction of CRISPR On-Target Effects via Deep Learning. http://arxiv.org/abs/2409.05938

[27] Cheng, X., Li, Z., Shan, R., Li, Z., Wang, S., Zhao, W., Zhang, H., Chao, L., Peng, J., Fei, T., & Li, W. (2023). Modeling CRISPR-Cas13d on-target and off-target effects using machine learning approaches. Nature Communications, 14(1). https://doi.org/10.1038/s41467-023-36316-3

[28] Ai, Y., Liang, D., & Wilusz, J. E. (2022). CRISPR/Cas13 effectors have differing extents of off-target effects that limit their utility in eukaryotic cells. Nucleic Acids Research, 50(11), E65–E65. https://doi.org/10.1093/nar/gkac159

[29] Kosicki, M., Tomberg, K., & Bradley, A. (2018). Repair of double-strand breaks induced by CRISPR–Cas9 leads to large deletions and complex rearrangements. Nature Biotechnology, 36(8), 765–771. https://doi.org/10.1038/nbt.4192

[30] Tsai, S. Q., Zheng, Z., Nguyen, N. T., Liebers, M., Topkar, V. v., Thapar, V., Wyvekens, N., Khayter, C., Iafrate, A. J., Le, L. P., Aryee, M. J., & Joung, J. K. (2015). GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. Nature Biotechnology, 33(2), 187–198. https://doi.org/10.1038/nbt.3117

[31] Wang, G., Wang, A., Wen, J., Kim, P., Song, Q., Liu, X., & Zhou, X. (2025). CRISPRoffT: comprehensive database of CRISPR/Cas off-targets. Nucleic Acids Research, 53(D1), D914–D924. https://doi.org/10.1093/nar/gkae1025

[32] Haapaniemi, E., Botla, S., Persson, J., Schmierer, B., & Taipale, J. (2018). CRISPR–Cas9 genome editing induces a p53-mediated DNA damage response. Nature Medicine, 24(7), 927-930.

[33] Liu, W., Li, L., Jiang, J., Wu, M., & Lin, P. (2021). Applications and challenges of CRISPR-Cas gene-editing to disease treatment in clinics. In Precision Clinical Medicine (Vol. 4, Issue 3, pp. 179–191). Oxford University Press. https://doi.org/10.1093/pcmedi/pbab014

[34] Kontarakis, Z., Revah, O., & Rossi, A. (2024). CRISPR-Cas13d: RNA's own Jedi Master in the fight against viral darkness. In Molecular Therapy Nucleic Acids (Vol. 35, Issue 4). Cell Press. https://doi.org/10.1016/j.omtn.2024.102321

[35] Moreno-Sánchez, I., Hernández-Huertas, L., Nahón-Cano, D., Martínez-García, P. M., Treichel, A. J., Gómez-Marin, C., Tomás-Gallardo, L., da Silva Pescador, G., Kushawah, G., Egidy, R., Perera, A., Díaz-Moscoso, A., Cano-Ruiz, A., Walker, J. A., Muñoz, M. J., Holden, K., Galcerán, J., Nieto, M. Á., Bazzini, A. A., & Moreno-Mateos, M. A. (2025). Enhanced RNA-targeting CRISPR-Cas technology in zebrafish. Nature Communications , 16(1). https://doi.org/10.1038/s41467-025-57792-9

[36] Ai, Y., Liang, D., & Wilusz, J. E. (2022). CRISPR/Cas13 effectors have differing extents of off-target effects that limit their utility in eukaryotic cells. Nucleic Acids Research, 50(11), E65–E65. https://doi.org/10.1093/nar/gkac159

[37] Wang, D., Zhang, C., Wang, B., Li, B., Wang, Q., Liu, D., Wang, H., Zhou, Y., Shi, L., Lan, F., & Wang, Y. (2019). Optimized CRISPR guide RNA design for two high-fidelity Cas9 variants by deep learning. Nature Communications, 10(1). https://doi.org/10.1038/s41467-019-12281-8

[38] Xiang, X., Corsi, G. I., Anthon, C., Qu, K., Pan, X., Liang, X., Han, P., Dong, Z., Liu, L., Zhong, J., Ma, T., Wang, J., Zhang, X., Jiang, H., Xu, F., Liu, X., Xu, X., Wang, J., Yang, H., … Luo, Y. (2021). Enhancing CRISPR-Cas9 gRNA efficiency prediction by data integration and deep learning. Nature Communications, 12(1). https://doi.org/10.1038/s41467-021-23576-0

[39] Lin, J., Zhang, Z., Zhang, S., Chen, J., & Wong, K. C. (2020). CRISPR-Net: A Recurrent Convolutional Network Quantifies CRISPR Off-Target Activities with Mismatches and Indels. Advanced Science, 7(13). https://doi.org/10.1002/advs.201903562

[40] Luo, J., Chen, W., Xue, L., & Tang, B. (2019). Prediction of activity and specificity of CRISPR-Cpf1 using convolutional deep learning neural networks. BMC Bioinformatics, 20(1). https://doi.org/10.1186/s12859-019-2939-6

[41] Korkin, D., ammartayakun, A., Busaranuvong, P., & Khosravi-Far, R. (2024). Graph Convolutional Network for predicting secondary structure of RNA. https://doi.org/10.21203/rs.3.rs-3798842/v1

[42] Wang, Y., Shen, Y., Li, J., Wang, T., Peng, J., & Shang, X. (2025). Enhanced RNA secondary structure prediction through int egr ativ e deep learning and structural context analysis. Nucleic Acids Research, 53, 533. https://doi.org/10.1093/nar/gkaf533

[43] Kontarakis, Z., Revah, O., & Rossi, A. (2024). CRISPR-Cas13d: RNA's own Jedi Master in the fight against viral darkness. In Molecular Therapy Nucleic Acids (Vol. 35, Issue 4). Cell Press. https://doi.org/10.1016/j.omtn.2024.102321

[44] Hussein, M., Liu, Y., Vink, M., Kroon, P. Z., Das, A. T., Berkhout, B., & Herrera-Carrillo, E. (2024). Evaluation of the effect of RNA secondary

structure on Cas13d-mediated target RNA cleavage. Molecular Therapy Nucleic Acids, 35(3). https://doi.org/10.1016/j.omtn.2024.102278

[45] Wei, J., Chen, S., Zong, L., Gao, X., & Li, Y. (2022). Protein–RNA interaction prediction with deep learning: structure matters. In Briefings in Bioinformatics (Vol. 23, Issue 1). Oxford University Press. https://doi.org/10.1093/bib/bbab540

[46] Yan, W. X., Chong, S., Zhang, H., Makarova, K. S., Koonin, E. v., Cheng, D. R., & Scott, D. A. (2018). Cas13d Is a Compact RNA-Targeting Type VI CRISPR Effector Positively Modulated by a WYL-Domain-Containing Accessory Protein. Molecular Cell, 70(2), 327-339.e5. https://doi.org/10.1016/j.molcel.2018.02.028

[47] Yousef, M., & Allmer, J. (2023). Deep learning in bioinformatics. Turkish Journal of Biology, 47(6), 366–382. https://doi.org/10.55730/1300-0152.2671

[48] Zupan, J. (1994). Introduction to Artificial Neural Network (ANN) Methods: What They Are and How to Use Them*. In Acta Chimica Slovenica (Vol. 41).

[49] Sharifani, K., & Amini, M. (n.d.). Machine Learning and Deep Learning: A Review of Methods and Applications. In World Information Technology and Engineering Journal (Vol. 10). https://ssrn.com/abstract=4458723

[50] Parhi, R., & Nowak, R. D. (2020). The Role of Neural Network Activation Functions. IEEE Signal Processing Letters, 27, 1779–1783. https://doi.org/10.1109/LSP.2020.3027517

[51] Proceedings of the 30th Chinese Control and Decision Conference (2018 CCDC) : 09-11 June 2018, Shenyang, China. (2018). IEEE Industrial Electronics (IE) Chapter.

[52] Kamath, C. N., Bukhari, S. S., & Dengel, A. (2018, August 28). Comparative study between traditional machine learning and deep learning approaches for text classification. Proceedings of the ACM Symposium on Document Engineering 2018, DocEng 2018. https://doi.org/10.1145/3209280.3209526

[53] Ahmed, S. F., Alam, M. S. bin, Hassan, M., Rozbu, M. R., Ishtiak, T., Rafa, N., Mofijur, M., Shawkat Ali, A. B. M., & Gandomi, A. H. (2023). Deep learning modelling techniques: current progress, applications, advantages, and challenges. Artificial Intelligence Review, 56(11), 13521–13617. https://doi.org/10.1007/s10462-023-10466-8

[54] Iuchi, H., Matsutani, T., Yamada, K., Iwano, N., Sumi, S., Hosoda, S., Zhao, S., Fukunaga, T., & Hamada, M. (2021). Representation learning applications in biological sequence analysis. In Computational and Structural Biotechnology Journal (Vol. 19, pp. 3198–3208). Elsevier B.V. https://doi.org/10.1016/j.csbj.2021.05.039

[55] Lewis, B., Smith, I., Fowler, M., & Licato, J. (2017). The robot mafia: A test environment for deceptive robots. 28th Modern Artificial Intelligence and Cognitive Science Conference, MAICS 2017, 189–190. https://doi.org/10.1145/1235

[56] Charlier, J., Nadon, R., & Makarenkov, V. (2021). Accurate deep learning off-target prediction with novel sgRNA-DNA sequence encoding in CRISPR-Cas9 gene editing. Bioinformatics, 37(16), 2299–2307. https://doi.org/10.1093/bioinformatics/btab112

[57] Toufikuzzaman, M., Samee, M. A. H., & Sohel Rahman, M. (2024). CRISPR-DIPOFF: an interpretable deep learning approach for CRISPR Cas-9 off-target prediction. In Briefings in Bioinformatics (Vol. 25, Issue 2). Oxford University Press. https://doi.org/10.1093/bib/bbad530

[58] Meng, Y., Zhang, Z., Zhou, C., Tang, X., Hu, X., Tian, G., Yang, J., & Yao, Y. (2025). Protein structure prediction via deep learning: an in-depth review. In Frontiers in Pharmacology (Vol. 16). Frontiers Media SA. https://doi.org/10.3389/fphar.2025.1498662

[59] Emtiaj, A. H., Rafi, R. H., Nayeem, M. A., & Rahman, M. S. (2025). CRISMER: A transformer-based Interpretable Deep Learning Approach for Genome-wide CRISPR Cas-9 Off-Target Prediction and Optimization. https://doi.org/10.1101/2025.05.03.652008

[60] Sherkatghanad, Z., Abdar, M., Charlier, J., & Makarenkov, V. (2023). Using traditional machine learning and deep learning methods for on- and off-target prediction in CRISPR/Cas9: a review. In Briefings in Bioinformatics (Vol. 24, Issue 3). Oxford University Press. https://doi.org/10.1093/bib/bbad131

[61] Vibhuti Choubisa. (2024). CRISPR-Cas9 off-target predictions using CNN and double CNN: Comparative analysis. International Journal of Science and Research Archive, 12(1), 1074–1080. https://doi.org/10.30574/ijsra.2024.12.1.0927

[62] Chuai, G., Ma, H., Yan, J., Chen, M., Hong, N., Xue, D., Zhou, C., Zhu, C., Chen, K., Duan, B., Gu, F., Qu, S., Huang, D., Wei, J., & Liu, Q. (2018). DeepCRISPR: Optimized CRISPR guide RNA design by deep learning. Genome Biology, 19(1). https://doi.org/10.1186/s13059-018-1459-4

[63] Lin, J., & Wong, K. C. (2018). Off-target predictions in CRISPR-Cas9 gene editing using deep learning. Bioinformatics, 34(17), i656–i663. https://doi.org/10.1093/bioinformatics/bty554

[64] Zhang, G., Luo, Y., Xie, H., & Dai, Z. (2024). Crispr-SGRU: Prediction of CRISPR/Cas9 Off-Target Activities with Mismatches and Indels Using Stacked BiGRU. International Journal of Molecular Sciences, 25(20). https://doi.org/10.3390/ijms252010945

[65] O'Brien, A., Bauer, D. C., & Burgio, G. (2023). Predicting CRISPR-Cas12a guide efficiency for targeting using machine learning. PLoS ONE, 18(10 October). https://doi.org/10.1371/journal.pone.0292924

[66] Zhang, C., Konermann, S., Brideau, N. J., Lotfy, P., Wu, X., Novick, S. J., Strutzenberg, T., Griffin, P. R., Hsu, P. D., & Lyumkis, D. (2018). Structural Basis for the RNA-Guided Ribonuclease Activity of CRISPR-Cas13d. Cell, 175(1), 212-223.e17. https://doi.org/10.1016/j.cell.2018.09.001

[67] Zhu, W., Xie, H., Chen, Y., & Zhang, G. (2024). CrnnCrispr: An Interpretable Deep Learning Method for CRISPR/Cas9 gRNA On-Target Activity Prediction. International Journal of Molecular Sciences, 25(8). https://doi.org/10.3390/ijms25084429

[68] Lee, M. (2023). Deep learning in CRISPR-Cas systems: a review of recent studies. In Frontiers in Bioengineering and Biotechnology (Vol. 11). Frontiers Media SA. https://doi.org/10.3389/fbioe.2023.1226182

[69] Liu, J., & Ma, J. (n.d.). Comparative Analysis of Machine Learning Algorithms for Predicting On-Target and Off-Target Effects of CRISPR-Cas13d for gene editing.

[70] Mallet, V., Oliver, C., Broadbent, J., Hamilton, W. L., & Waldispühl, J. (2022). RNAglib: A python package for RNA 2.5 D graphs. Bioinformatics, 38(5), 1458–1459. https://doi.org/10.1093/bioinformatics/btab844

[71] McInnes, L., Healy, J., & Melville, J. (2020). UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. http://arxiv.org/abs/1802.03426

[72] Zhu, G., Zhou, X., Wen, M., Qiao, J., Li, G., & Yao, Y. (2024). CRISPR–Cas13: Pioneering RNA Editing for Nucleic Acid Therapeutics. In BioDesign Research (Vol. 6). American Association for the Advancement of Science. https://doi.org/10.34133/bdr.0041

[73] Moreno-Sánchez, I., Hernández-Huertas, L., Nahón-Cano, D., Martínez-García, P. M., Treichel, A. J., Gómez-Marin, C., Tomás-Gallardo, L., da Silva Pescador, G., Kushawah, G., Egidy, R., Perera, A., Díaz-Moscoso, A., Cano-Ruiz, A., Walker, J. A., Muñoz, M. J., Holden, K., Galcerán, J., Nieto, M. Á., Bazzini, A. A., & Moreno-Mateos, M. A. (2025). Enhanced RNA-targeting CRISPR-Cas technology in zebrafish. Nature Communications, 16(1). https://doi.org/10.1038/s41467-025-57792-9

TEZDEN ÜRETİLMİŞ YAYINLAR

Konferans Bildirileri

1. Barış, B., & Yılmaz, A. (2025). CRISPR-Cas13d'nin Hedef Dışı Etkilerinin gRNA'nın 3 Boyutlu Yapısı ile Beraber Derin Öğrenme ile İyileştirilmesi. In 11. Uluslararası Ege Doğa ve Tıp Bilimleri Kongresi (AEGEAN-2025). Türkiye.