

T.C.
AKDENİZ ÜNİVERSİTESİ
SAĞLIK BİLİMLERİ ENSTİTÜSÜ
TIBBİ BİYOTEKNOLOJİ ANABİLİM DALI

Ahmet Ozan ÖZGEN

YÜKSEK LİSANS TEZİ

2024-ANTALYA

ANTİMİKROBİYAL PEPTİTLERİN ANTİKANSER
ÖZELLİKLERİNİN IN SİLİKO YÖNTEMLER İLE
ARAŞTIRILMASI

Ahmet Ozan ÖZGEN

YÜKSEK LİSANS TEZİ

2024-ANTALYA

T.C.
AKDENİZ ÜNİVERSİTESİ
SAĞLIK BİLİMLERİ ENSTİTÜSÜ
TIBBİ BİYOTEKNOLOJİ ANABİLİM DALI

ANTİMİKROBİYAL PEPTİTLERİN ANTİKANSER
ÖZELLİKLERİNİN IN SİLİKO YÖNTEMLER İLE
ARAŞTIRILMASI

Ahmet Ozan ÖZGEN

YÜKSEK LİSANS TEZİ

DANIŞMAN
Dr. Öğr. Üyesi Devrim DEMİR DORA

“Kaynakça gösterilerek tezimden yararlanılabilir”

2024-ANTALYA

Sağlık Bilimleri Enstitüsü Müdürlüğüne;

Bu çalışma jürimiz tarafından Tıbbi Biyoteknoloji Anabilim Dalı Tıbbi Biyoteknoloji Programında yüksek lisans tezi olarak kabul edilmiştir. 19/12/2023

İmza

Tez Danışmanı : Dr. Öğr. Üyesi Devrim DEMİR DORA
Akdeniz Üniversitesi

.....

Üye : Doç. Dr. Uğur BİLGE
Akdeniz Üniversitesi

.....

Üye : Doç. Dr. Ozan Emre EYÜPOĞLU
İstanbul Medipol Üniversitesi

.....

Bu tez, Enstitü Yönetim Kurulunca belirlenen yukarıdaki jüri üyeleri tarafından uygun görülmüş ve Enstitü Yönetim Kurulu'nun/...../..... tarih ve/..... sayılı kararıyla kabul edilmiştir.

Enstitü Müdürü

ETİK BEYAN

Bu tez çalışmasının kendi çalışmam olduğunu, tezin planlanmasından yazımına kadar bütün safhalarda etik dışı davranışımın olmadığını, bu tezdeki bütün bilgileri akademik ve etik kurallar içinde elde ettiğimi, bu tez çalışmasıyla elde edilmeyen bütün bilgi ve yorumlara kaynak gösterdiğimi ve bu kaynakları da kaynaklar listesine aldığımı beyan ederim.

Ahmet Ozan ÖZGEN

Tez Danışmanı

Dr. Öğr. Üyesi Devrim DEMİR DORA

TEŞEKKÜR

Öncelikle, ülkemizde çağdaş eğitimin temellerini atan ve akademik alanda gelişmiş ülkelerin karşısında eşit paydaşlar olarak yer alabilmemizi sağlayan Mustafa Kemal ATATÜRK'e,

Yüksek lisans eğitimim boyunca bilimsel ve akademik anlamda bana rehberlik eden, öğrencisi olmanın verdiği gurur ve minneti ömür boyu taşıyacağım değerli tez danışmanım Dr. Öğr. Üyesi Devrim DEMİR DORA'ya,

Yüksek lisans çalışmalarım süresince bilgi, tecrübe ve desteğini esirgemeyen, bana yeni bakış açıları kazandıran Doç. Dr. Uğur BİLGE'ye,

Lisans eğitimim sona erdikten sonra dahi akademik çalışmalarında bana yer vererek bilgi ve deneyimlerini paylaşan değerli lisans danışmanım Doç. Dr. Ozan Emre EYÜPOĞLU'na,

Tıbbi biyoteknoloji programının kurulmasındaki çabaları ve bize bu imkanı sunduğu için Anabilim Dalı Başkanımız Prof. Dr. Ahmet Yılmaz ÇOBAN'a,

Her şart altında, bilgisi ve desteğiyle her zaman ilham veren, beni motive eden ve güçlendiren Edanur ERSÖZ'e,

Yüksek lisans eğitimime başladığım ilk günden itibaren manevi destekleri ve akademik katkılarıyla yanımda olan arkadaşım Gizem TUTKUN'a,

Ve son olarak, hayatım boyunca bana gösterdikleri sonsuz sevgi, değerli destek ve inançları için sevgi ve saygıyla bağlı olduğum aileme,

Derin saygılarımı ve teşekkürlerimi sunarım.

ÖZET

Amaç: Bu çalışma, antimikrobiyal peptitlerin (AMP) kanser tedavisindeki potansiyellerinin öngörülebilmesi amacıyla çeşitli *in-silico* yöntem ve araçları kullanarak bir süreç geliştirmeyi hedeflemektedir. Geliştirilen bu süreç sonucunda 10.000 AMP'nin antikanser etkinliklerinin ve fizikokimyasal özelliklerinin tahmini gerçekleştirilecektir. Ayrıca, sonuçların AMP'lerin olası ikili etkilerine ışık tutması beklenmektedir.

Yöntem: Ana (eğitim) veri seti kullanılarak araştırmacı tarafından özellik seçimi yapılmış olup bir makine öğrenimi modeli eğitilmiştir. Bu model antikanser etkinliğinin tespit edilmesi için çalışmanın temelini oluşturmaktadır. Model performansı temel başarı metrikleri ile ölçülmüştür. Sonuçların normal dağılım gösterme durumları Shapiro-Wilk testi ve farklı veri setlerindeki performansları da MANOVA testi ile değerlendirilmiştir. Bunun yanı sıra, peptit sekanslarının kümeleme işlemleri, fizikokimyasal özelliklerin tahmini ve üç boyutlu homoloji modellemesi için kullanılan bağımsız araçlar da bu sürecin bir parçası olarak entegre edilmiştir.

Bulgular: En başarılı model, rastgele orman algoritması ile amino asit ve dipeptit kompozisyonu özellik kombinasyonu ile elde edilmiştir. Geliştirilen makine öğrenimi modeli bağımsız (validasyon) veri setinde %79,5 doğruluk göstermiştir. Eğitilen modelin ana ve bağımsız veri setlerindeki performansları arasında istatistiksel olarak anlamlı bir farklılık bulunmamaktadır. Ayrıca, deneysel veri setinde test edilen 10.000 peptitten antikanser etkinliği tespit edilen ve optimal fizikokimyasal özellik gösteren 1648 peptit keşfedilmiştir.

Sonuç: Elde edilen bulgular, AMP'lerin antikanser özelliklerinin daha iyi anlaşılması ve bu peptitlerin kanser tedavisindeki potansiyel uygulamalarının geliştirilmesi için değerli bilgiler sunmaktadır.

Anahtar Kelimeler: Antimikrobiyal peptitler, antikanser peptitler, makine öğrenimi, özellik seçimi, in silico

ABSTRACT

Objective: This study aims to develop a process using various in-silico methods and tools to predict the potential of antimicrobial peptides (AMPs) in cancer treatment. Through the developed process, the anticancer activity, and physicochemical properties of 10,000 AMPs will be predicted. Additionally, it is anticipated that the results will shed light on the possible dual effects of AMPs.

Method: The researcher performed feature selection using the main (training) dataset and trained a machine learning model. This model forms the basis of the study for predicting anticancer activity. The success of the model was measured using fundamental performance metrics. The normal distribution of the results was assessed through the Shapiro-Wilk test, and their performances on different datasets were evaluated using the MANOVA test. Additionally, independent tools, such as clustering of peptide sequences, prediction of physicochemical properties, and three-dimensional homology modelling, were integrated as part of this process.

Results: The most successful model was achieved using the random forest algorithm with a combination of amino acid and dipeptide composition features. The developed machine learning model demonstrated an accuracy of 79.5% on the independent (validation) dataset. There was no statistically significant difference in the model's performances between the main and independent datasets. Additionally, out of the 10,000 peptides tested in the experimental dataset, 1648 peptides were discovered to exhibit anticancer activity with optimal physicochemical properties.

Conclusion: The findings provide valuable information for a better understanding of the AMPs' anticancer properties and the potential applications of these peptides in cancer treatment.

Keywords: Antimicrobial peptides, anticancer peptides, machine learning, feature selection, in silico

İÇİNDEKİLER

ÖZET	i
ABSTRACT	ii
İÇİNDEKİLER	iii
TABLolar DİZİNİ	iv
ŞEKİLLER DİZİNİ	viii
SİMGELER ve KISALTMALAR	x
1. GİRİŞ	1
2. GENEL BİLGİLER	3
2.1. Kanser	3
2.1.1. Kanser Biyolojisi ve Genetiği	3
2.1.2. Kanser Tipleri ve Sınıflandırılması	5
2.1.3. Kanser Patogenezi ve Progresyonu	6
2.1.4. Kanser Hücrelerinin Direnç Mekanizmaları	8
2.2. Kemoterapötikler	10
2.2.1. Antineoplastik Ajanlar	10
2.2.2. Antimikrobiyal Ajanlar	11
2.3. Peptitler	12
2.3.1. Peptitlerin Biyolojik Fonksiyonları	14
2.3.2. Peptitlerin Sınıflandırılması	15
2.3.3. Antimikrobiyal Peptitler	19

2.3.4. Antikanser Etki Gösteren Peptitler	23
2.3.5. Antimikrobiyal ve Antikanser Hedefler için İkili Aktivite Gösteren Peptitler	26
2.3.6. Peptitlerin Fizikokimyasal Özellikleri	26
2.3.7. Antimikrobiyal Peptitlerin Terapötik Potansiyelleri ve Kısıtları	27
2.4. Peptit Veri Tabanları ve Kullanımları	29
2.4.1. Peptit Veri Tabanlarında Arama ve Filtreleme Seçenekleri	30
2.4.2. Önemli Peptit Veri Tabanları	31
2.5. Peptit Araştırmalarında In Silico Yöntemler	37
2.5.1. Literatürdeki Biyoinformatik Araçlar	37
2.5.2. Özellik Seçimi	40
2.5.3. Makine Öğrenmesi Algoritmaları	41
3. GEREÇ ve YÖNTEM	45
3.1. İşletim Sistemi ve Yazılım Dağıtımı	45
3.1.1. İşletim Sistemi (OS)	45
3.1.2. Yazılım Dağıtımı	45
3.1.3. Entegre Geliştirme Ortamı (IDE)	46
3.2. Programlama Dilleri	46
3.2.1. Python	47
3.2.2. Visual Basic for Applications (VBA)	47

3.3. Python Kütüphaneleri ve Paketleri	47
3.4. Bağımsız Yazılım Protokoller	49
3.4.1. USEARCH (UCLUST algoritması)	50
3.4.2. PepFun	51
3.5. Literatür Taraması ve Kelime Analizi	52
3.6. Veri Seti Kaynakları ve Seçimi	53
3.6.1. Ana Veri Seti	54
3.6.2. Bağımsız Veri Seti	54
3.6.3. Deneysel Veri Seti	55
3.7. Veri Setlerinin İşlenmesi ve Hazırlanması	55
3.7.1. Kümeleme ve Dengeleme İşlemleri	56
3.7.2. Numerik Haritalama	58
3.7.3. Özellik Seçimi	59
3.8. Makine Öğrenimi Modeli Geliştirme	61
3.8.1. Değerlendirme Metodolojisi	62
3.8.2. İstatistiksel Analiz	63
3.8.3. Optimal Özellik ile Makine Öğrenimi Algoritmasının Seçimi ve Karşılaştırılması	64
3.8.4. Hiperparametre Ayarlama ve Optimizasyonu	65
3.8.5. Çalışma Modelinin Eğitilmesi, Değerlendirilmesi ve Kaydedilmesi	67

3.9. Bağımsız Veri Seti ile Validasyon	68
3.10. Antikanser Özelliklerin Tahmini	68
3.11. Fizikokimyasal Özelliklerin Tahmini	69
3.12. Peptitlerin Üç Boyutlu Modellerinin Oluşturulması	70
4. BULGULAR	71
4.1. Derlenen Literatüre Ait Kelime Analizi	71
4.2. En Başarılı Makine Öğrenimi Algoritması ile Özellik Eşleşmesi	74
4.3. RF Modeli İçin İdeal Hiperparametreler	75
4.4. Çalışma Modelinin Ana Veri Seti ile Performansı	75
4.5. Bağımsız Veri Seti ile Validasyon Sonuçları	80
4.5.1. Modelin Ana ve Bağımsız Veri Setleriyle Başarısının Karşılaştırılması	82
4.6. Deneysel Veri Setinde Antikanser Özelliği Tespit Edilen Peptitler ve Özellikleri	84
4.7. Üç Boyutlu Peptit Modelleri	87
5. TARTIŞMA	94
6. SONUÇ VE ÖNERİLER	104
KAYNAKLAR	106
EKLER	126
EK-1. 10.000 AMP içeren deneysel veri setinde antikanser etkinlik gösteren ve fizikokimyasal özellikler için belirlenen aralıkta bulunan 1648 peptit	
ÖZGEÇMİŞ	127

TABLÖLAR DİZİNİ

Tablo 2.1.	Amino asitler ve kısaltmaları.	13
Tablo 2.2.	Çeşitli terapötik peptitler ve klinik durumları.	19
Tablo 4.1.	Farklı özellik ve algoritma eşleşmelerine ait başarı metrikleri.	74
Tablo 4.2.	Çalışma modelinin ana veri seti ile başarı metrikleri.	76
Tablo 4.3.	Çalışma modelinin bağımsız veri seti ile başarı metrikleri.	81
Tablo 4.4.	Antikanser etkinliği tespit edilen rastgele seçilmiş 15 antimikrobiyal peptit ve çeşitli özellikleri.	85

ŞEKİLLER DİZİNİ

Şekil 2.1. Direkt ve dolaylı yoldan meydana gelen temel DNA hasarları.	4
Şekil 2.2. Kanser hücrelerinin kemoterapötiklere karşı potansiyel direnç mekanizmaları.	9
Şekil 2.3. Folik asit yolağı ve metotreksat.	11
Şekil 2.4. AMP'lerin olası membran etkileşim mekanizmaları.	22
Şekil 3.1. Merkez-tabanlı, orta ile yüksek özdeşlikte kümeleme.	51
Şekil 3.2. Veri setlerinde yer alan peptit sekanslarının uzunlukları ve dağılımı, A) Ana veri seti, B) Bağımsız veri seti.	57
Şekil 3.3. Veri setlerinde yer alan peptit sekanslarının uzunlukları ve antikanser özellikleri, A) Ana veri seti, B) Bağımsız veri seti.	57
Şekil 3.4. Veri setlerinde yer alan antikanser özelliğe sahip peptitlere ait amino asit sayımları, A) Ana veri seti, B) Bağımsız veri seti.	58
Şekil 4.1. En sık alıntılanan kelimelere ait kelime bulutu.	71
Şekil 4.2. Literatürde yer alan amino asitler ve alıntılanma sayıları.	72
Şekil 4.3. Bakteriler ve alıntılanma sayıları.	72
Şekil 4.4. Yapısal özellikler ve alıntılanma sayıları.	73
Şekil 4.5. Makine öğrenimi modellerinin alıntılanma sayıları.	73
Şekil 4.6. Ana veri seti hata matrisi.	77
Şekil 4.7. Ana veri seti ile eğitilen çalışma modeline ait bir karar ağacı dalı.	78

Şekil 4.8. Ana veri seti ile eğitilen çalışma modele ait en etkin 20 özelliğin önem derecesi.	79
Şekil 4.9. Eğitilen modelin ana veri seti ile öğrenme eğrisi.	80
Şekil 4.10. Modelin bağımsız veri setine ait hata matrisi.	82
Şekil 4.11. Modelin her iki veri seti ile hesaplanan ROC AUC grafiği.	84
Şekil 4.12. “seq1” numaralı peptite ait ait üç boyutlu model.	87
Şekil 4.13. “seq2” numaralı peptite ait ait üç boyutlu model.	87
Şekil 4.14. “seq3” numaralı peptite ait ait üç boyutlu model.	88
Şekil 4.15. “seq4” numaralı peptite ait ait üç boyutlu model.	88
Şekil 4.16. “seq5” numaralı peptite ait ait üç boyutlu model.	89
Şekil 4.17. “seq6” numaralı peptite ait ait üç boyutlu model.	89
Şekil 4.18. “seq7” numaralı peptite ait ait üç boyutlu model.	90
Şekil 4.19. “seq8” numaralı peptite ait ait üç boyutlu model.	90
Şekil 4.20. “seq9” numaralı peptite ait ait üç boyutlu model.	90
Şekil 4.21. “seq10” numaralı peptite ait ait üç boyutlu model.	91
Şekil 4.22. “seq11” numaralı peptite ait ait üç boyutlu model.	91
Şekil 4.23. “seq12” numaralı peptite ait ait üç boyutlu model.	92
Şekil 4.24. “seq13” numaralı peptite ait üç boyutlu model.	92
Şekil 4.25. “seq14” numaralı peptite ait üç boyutlu model.	93
Şekil 4.26. “seq15” numaralı peptite ait üç boyutlu model.	93

SİMGELER ve KISALTMALAR

5-FU: 5-Florourasil

AAC: Amino asit kompozisyonu

A-CaMP: Antikanser ve antimikrobiyal peptit tasarımı için araç

ACP: Antikanser peptit

ACT: Artemisinin bazlı kombinasyon tedaviler

ACTH: Adrenokortikotropik hormon

ADT: Androjen depresyon tedavisi

ALL: Akut lenfositik lösemi

AML: Akut miyeloid lösemi

AMP: Antimikrobiyal peptit

AMPA: Antimikrobiyal peptit tarayıcısı

AntiCP: Antikanser peptitlerin tahmini ve tasarlanması

APD: Antimikrobiyal peptit veri tabanı

ApInAPDB: Apoptoz indükleyici antikanser peptitler veri tabanı

BACTIBASE: Bakteriyosinlere Özel Veri tabanı

BAGEL: Web Tabanlı Bakteriyosin Genom Madenciliği Aracı

BER: Baz eksizyon onarımı

Bootstrap: Önyükleme

C: Sistein

CAMP: Antimikrobiyal peptit koleksiyonu

CancerPPD: Antikanser peptit ve protein veri tabanı

CAR-T Hücreleri : Kimerik antijen reseptör t hücreleri

Centroid: Merkez

CLL: Kronik lenfositik lösemi

CpACpP: Hücreye nüfuz eden antikanser peptit tahmini

CPP: Hücreye nüfuz eden peptit

CPPsite: Hücreye nüfuz eden peptitler veri tabanı

CPU: Merkezi işlemci birimi

Criterion: Kalite ölçüm fonksiyonu

CTD: Dağılım

CyBase: Döngüsel protein dizileri ve yapıları için veri tabanı

Da: Dalton

DBAASP: Antimikrobiyal aktivite ve peptit yapısı veri tabanı

DF-HAS: İnsan serum albümini içeren rekombinant özelleştirilmiş defensin

DPC: Dipeptit kompozisyonu

DRAMP: Antimikrobiyal peptitlerin veri havuzu

DSÖ: Dünya Sağlık Örgütü

EGF: Epidermal büyüme faktörü

FGF: Fibroblast büyüme faktörü

GMQE: Global model kalite tahmini

GPU: Grafik işleme birimi

hBD: β -defensinler

HIPdb: Deneysel olarak doğrulanmış HIV inhibe edici peptit veri tabanı

HMM: Gizli Markov modelleri

IDE: Entegre geliştirme ortamı

IDR peptitleri: Doğuştan savunma düzenleyici peptitleri

Ig: İmmünoglobulin

IUPAC: Uluslararası temel ve uygulamalı kimya birliği

İYE: İdrar Yolu Enfeksiyonu

K: Lizin

k-NN: K-en yakın komşular

LAMP: Antimikrobiyal peptitleri bağlayan bir veri tabanı

logP: Partisyon katsayısı

LPS: Lipopolisakkarit

LR: Lojistik regresyon

MANOVA: Çok değişkenli varyans analizi

max_depth: Maksimum derinlik

max_features: Maksimum özellik sayısı

MDR: Çoklu ilaç direnci

MIC: Minimum inhibitör konsantrasyonu

min_samples_leaf: Yaprak düğümde minimum örnek sayısı

min_samples_split: Minimum örnek bölme sayısı

miRNA: mikro RNA

MMR: Eşleşme Onarımı

MRSA: Metisiline dirençli *Staphylococcus aureus*

MSH: Melanosit stimüle edici hormonlar

NaN: Geçersiz

NER: Nükleotit eksizyon onarımı

NK: Doğal katil

NK-B: Nörokinin B

NLP: Doğal dil işleme

OHE: Tek-sıcak kodlama

Oob_score: OOB skoru

OS: İşletim sistemi

P: Prolin

PBP: Penisilin bağlayıcı proteinler

PDGF: Trombosit kaynaklı büyüme faktörü

PenBase: Prokaryotik ve ökaryotik penetratin veri tabanı

PhytAMP: Antimikrobiyal bitki peptitlerine özel veri tabanı

POMC: Proopiomelanokortin

PPY: Pankreatik Polipeptit

PS: Fosfatidilserin

PSSM: Pozisyona özel puanlama matrisi

R grubu: Yan zincir

R: Arjinin

RAM: Rasgele erişimli bellek

RBF: Radyal bazlı fonksiyon

RF: Rastgele orman

RiPP: Translasyon sonrası modifiye edilmiş peptitler

ROC AUC: Alıcı işletim karakteristik eğrisi altında kalan alan

ROS: Reaktif oksijen türleri

SCC: Kısa devre akımı

SD: Standart sapma

SE: Standart hata

SVM: Destek vektör makinesi

SVM-RFE: Yinelemeli özellik eleme

SW testi: Shapiro-Wilk testi

TB: Tüberküloz

TLR: Toll benzeri reseptörler

TME: Tümör mikroçevre

TMP-SMX: Trimetoprim-sülfametoksazol

TNM: Tümör, lenf düğümü, metastaz

TRH: Tirotropin salgılatıcı hormon

TSH: Tiroid uyarıcı hormon

VBA: *Visual Basic for Applications*

VIP: Vazoaktif bağırsak peptitleri

VRE: Vankomisine dirençli *Enterococcus*

YADAMP: Bir diğer antimikrobiyal peptit veri tabanı

1. GİRİŞ

Kanser, genellikle kontrolsüz hücre bölünmesi, replikatif ölümsüzlük ve hücre ölümü süreçlerine dirençle karakterize edilen bir hastalıktır. Günden güne gelişen teknoloji, tıp alanında, kompleks yapılar sergileyen kanserler için farklı terapötik seçenekler sunmaktadır (Leite ve ark., 2018). Kanser tedavisinde cerrahi müdahale, radyoterapi, kemoterapi gibi geleneksel tedavi yaklaşımlarının yanı sıra; immünoterapi, gen tedavisi ve bu tez çalışmasının kapsamında olan antimikrobiyal peptitler (AMP'ler) gibi terapötik potansiyeli yüksek yeni tedavi seçenekleri mevcuttur (Trapani ve Darcy, 2017; Jafari ve ark., 2022).

AMP'ler, neredeyse tüm canlı organizmalarda bulunan, genellikle 10-100 amino asitten oluşan; bakteriler, mantarlar, protozoalar ve virüslere karşı geniş spektrumda etki gösterebilen moleküllerdir. Antimikrobiyal etkinliklerinin ötesinde, immün regülasyon, yara iyileşmesi ve antikanser aktiviteleri çeşitli hastalıkların tedavisinde değerlendirilmelerini sağlamıştır. Antikanser etkilerini, kanser hücrelerinin hücre zarlarını hedef alarak veya büyümesini/yayılmasını engelleyen diğer mekanizmalar yoluyla gerçekleştirebilirler. AMP'lerin kanser üzerindeki etkilerini anlamak için yürütülen araştırmalar, farklı *in-vitro*, *in-vivo* ve *in-silico* metotlar aracılığıyla bu peptitlerin kanser hücrelerinin büyümesini inhibe etme, apoptoza neden olma ve tümör mikroçevresini etkileme gibi yeteneklerini incelemektedir. AMP'lerin kanser tedavisinde uygulamaları henüz sınırlı olmakla birlikte, bu alanda prelinik ve klinik araştırma aşamasında olan birçok terapötik peptit bulunmaktadır (Haney ve ark., 2017).

Tedavide kullanılmalarına yönelik zorluklar; istenmeyen bağışıklık yanıtlarını tetikleme potansiyeli, stabilite ve biyoyararlanım gibi problemleri içermektedir. İlaç formülasyonu geliştirilmesiyle ilişkili zorlukların yanında, bu peptitlerin terapötik kullanımı için daha etkili ve daha az toksik türevlerinin geliştirilmesi, tedavi sürecinde risk-fayda oranını iyileştirmek açısından büyük önem taşımaktadır (Haney ve ark., 2017).

Antikanser özelliklere sahip AMP'lerin keşfi için *in-silico* araçların ve yöntemlerin kullanımı son yıllarda rağbet görmektedir. Bu yaklaşım, bilgisayar tabanlı teknikler

kullanarak, büyük veri setlerinden potansiyel antikanser peptitlerin tanımlanmasını ve sınıflandırılmasını içerir. *In-silico* yöntemlerin tercih edilmesinin en önemli nedenlerinden biri, laboratuvar deneyleriyle karşılaştırıldığında, daha hızlı ve maliyet etkin alternatifler sunmalarıdır (Gačić, 2023).

Bu çalışma, AMP'lerin kanser tedavisindeki potansiyellerinin öngörülebilmesi amacıyla çeşitli *in-silico* yöntem ve araçları kullanarak bir süreç geliştirmeyi hedeflemektedir. Ayrıca, sonuçların AMP'lerin olası ikili etkilerine ışık tutması beklenmektedir. Araştırmacı tarafından geliştirilen makine öğrenimi modeli ile antikanser etkinliğinin tespit edilmesi çalışmanın temelini oluşturmaktadır. Bunun yanı sıra, peptit sekanslarının kümeleme işlemleri, fizikokimyasal özelliklerin tahmini ve üç boyutlu homoloji modellemesi için kullanılan bağımsız araçlar da bu sürecin bir parçası olarak entegre edilmiştir. Süreç ve yöntemler bu alandaki araştırmacıların kullanım kolaylığı gözetilerek planlanmıştır.

Araştırma kapsamında, antikanser peptitleri tespit etmek için geliştirilen çeşitli makine öğrenimi modelleri ve algoritmaları kapsamlı bir şekilde incelenmiştir. Özellikle, peptit dizilerinden elde edilen özellikleri tanıyan ve işleyen çeşitli modelleme teknikleri değerlendirilmiştir. Bu modellerin performansı, çeşitli veri setleri kullanılarak test edilmiş ve karşılaştırılmıştır. İncelenen modeller, destek vektör makinesi ve RF gibi farklı algoritmaları kapsamaktadır. Her modelin avantajları ve sınırlamaları, bulgular ile birlikte tartışılmıştır.

Elde edilen bulgular, AMP'lerin kanser tedavisindeki araştırmaları desteklemek için belli temel bilgileri sağlamaktadır. Ayrıca, AMP'lerin antikanser özelliklerinin daha iyi anlaşılması ve bu peptitlerin kanser tedavisindeki potansiyel uygulamalarının geliştirilmesi için ileri araştırmaların içeriklerine dair öneriler paylaşılmıştır. Bu öneriler, AMP'lerin kanserle mücadeledeki rolünü daha iyi anlamak ve bu peptitlerin terapötik olarak kullanımını değerlendirmek için temel bir çerçeve sunmaktadır.

2. GENEL BİLGİLER

2.1. Kanser

“Kanser”, farklı dokuları ve hücre tiplerini etkileyen yüzden fazla farklı hastalığı kapsayan bir olgudur (Leite ve ark., 2018). Kontrolsüz hücre bölünmesi, replikatif ölümsüzlük ve hücre ölümüne yol açan süreçlere direnç, kanser türlerinin genel özelliklerindendir. Hematolojik kanserler dışında, kanser hücrelerinin proliferasyonu sonucu oluşan ve tümör olarak bilinen anormal hücre kütleleri mevcuttur (Pérez-Herrero ve Fernández-Medarde, 2015).

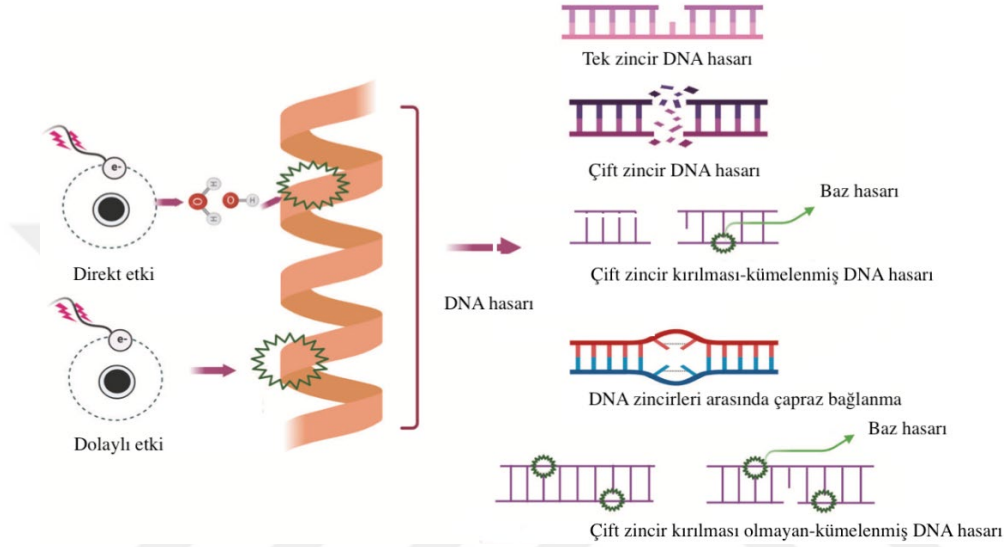
2.1.1. Kanser Biyolojisi ve Genetiği

Kanser, genetik mutasyonların birikimi ve hücre döngüsü kontrol mekanizmalarının bozulmasıyla karakterize edilir. Bu mutasyonlar, hücrelerin ölümünü engelleyip kontrolsüz çoğalmalarına yol açabilir. Onkogenlerin aşırı aktivasyonu ve tümör baskılayıcı genlerin işlevsizleşmesi, DNA hasarına yanıt veren onarım mekanizmalarının başarısızlığıyla birleştiğinde kanser gelişimine neden olur. Kanserli hücreler ayrıca epigenetik değişiklikler ve çevresel faktörlerden etkilenir; bu değişiklikler, gen ifadesini değiştirerek kanser gelişimini etkileyebilir (Torgovnick ve Schumacher, 2015).

DNA Hasarı ve Onarımı

Hücrelerin hayatta kalımı ve çoğalması için genom kararlılığının korunması şarttır. İnsan vücudundaki yaklaşık 10^{13} hücrenin her biri günde binlerce DNA hasarına maruz kalır. DNA hasarı, endojen ve eksojen faktörlerin doğrudan teması ya da bu faktörlerin radikalleri aktive etmesiyle doğrudan veya dolaylı olarak meydana gelir (Fu ve ark., 2021). Endojen materyaller, hasar sonrası veya dışarıdan gelen maddelerin metabolize edilmesiyle serbest kalır (Gupta ve ark., 2020). DNA onarım mekanizmalarının düzeltemediği hasarlar, hücre işlevlerini bozabilir, kanser oluşumuna veya hücre ölümüne yol açabilir. Genomik kararsızlık, proto-onkogenler ve anti-onkogenler içeren yanıt kaskadlarıyla tetiklenir (Huang R. ve Zhou, 2021). Organizmanın devamlılığı için etkin DNA tamir mekanizmaları gereklidir (Basu, 2018).

DNA hasarı sonucu meydana gelen tek-çift zincir kırıkları, baz hasarı ve çapraz bağlanma gibi çeşitli yollarla oluşur. Çift zincir kırıklarının tamir mekanizmalarınca onarımı daha, zor olduğu için, kanser oluşturma potansiyelleri daha yüksektir. (Frankenberg-Schwager, 1990). Bu süreçler Şekil 2.1’de açıklanmaktadır.



Şekil 2.1. Direkt ve dolaylı yoldan meydana gelen temel DNA hasarları (Huang R. ve Zhou, 2021).

Memeli hücrelerinde baz eksizyon onarımı (BER), eşleşme onarımı (MMR), nükleotit eksizyon onarımı (NER) ve çift iplik kırığı onarımı gibi mekanizmalar DNA hasarlarını onarmak için kullanır (Hoeijmakers, 2001).

Onkogenler ve Tümör Baskılayıcı Genlerin Rolü

Proto-onkogenler, sağlıklı hücrelerde temel biyolojik süreçlerin devamlılığı ve çoğalma için kritik öneme sahiptir. Genetik değişikliklerle (kromozomal translokasyon, nokta mutasyonu ve gen amplifikasyonu vb.) onkogenlere dönüşebilir ve hücre fonksiyonlarını bozarak kanser gelişimini tetikleyebilirler (Lee ve Muller, 2010).

Proto-onkogenler, kendi transkripsiyon promotörleri tarafından düzenlenir. Bu promotörler, gen ifadesinin derecesini belirleyen spesifik DNA dizileridir ve genin farklı fizyolojik sinyallere tepkilerini kontrol etmektedirler. Proto-onkogen ifadesi, hücresel ihtiyaçlara bağlı olarak değişir; farklı senaryolarda azalabilir veya artabilir (Lee ve Muller, 2010).

İnsan genomunda ayrıca neoplazinin önlenmesinde rol oynayan çeşitli tümör baskılayıcı genler bulunur. Tümör baskılayıcı genler, hücre bölünmesini düzenleyen ve anormal hücre büyümesini engelleyen genlerdir. Bu genler, hücre döngüsünü kontrol eden proteinleri kodlar ve hücrelerin kontrolsüz bir şekilde bölünmesini ve çoğalmasını engeller, bu da kanserin başlıca nedenlerinden biridir. Tümör baskılayıcı genlerin işlev bozukluğu, hücrelerin kontrolsüz bir şekilde bölünmelerine ve tümörlerin oluşmasına neden olabilir (Kontomanolis E. N. ve ark., 2020).

Epigenetik Değişiklikler ve Kanser

Memelilerde, epigenetik mekanizmalar, hücre veya dokuya özgü gen ifadesinin düzenlenmesi için kritiktir. Bu mekanizmaların bozulması, gen fonksiyonlarında değişikliklere ve kanser gibi hastalıklara yol açabilir. Günümüzde kanser oluşumunda genetik mutasyonların yanı sıra epigenetik sapmaların da etkili olduğu kabul edilmektedir. Kanser epigenetiği alanındaki yeni bulgular, DNA metilasyonu, histon modifikasyonları ve mikroRNA üretimi gibi epigenetik mekanizmalarda ciddi farklılaşmalar olduğunu göstermiştir (Sharma ve ark., 2010).

2.1.2. Kanser Tipleri ve Sınıflandırılması

Kanser, temel olarak solid tümörler ve hematolojik maligniteler olmak üzere iki sınıfta incelenebilir. Solid tümörler, belirli organlarda gelişir ve köken aldıkları dokuya göre isimlendirilirken, hematolojik maligniteler kan, kemik iliği ve lenf sistemini etkileyen lösemi, lenfoma ve miyeloma gibi türleri kapsar. Bu ayrım, kanserin tanı, tedavi ve prognozunun belirlenmesinde temeldir (Zhang H. ve ark., 2023). Kanserler histolojik özellikler, tümör, lenf düğümü, metastaz (TNM) sistemi ve moleküler profiller ile sınıflandırılır. Histolojik sınıflandırma, tümörün hücre tipi ve dokusuna dayanır; TNM, tümör boyutu, lenf düğümü yayılımı ve metastazı değerlendirir; moleküler profil ise genetik ve moleküler özellikleri temel alır ve kişiselleştirilmiş tedavilerde önem taşır (Koeppel ve ark., 2021).

Kanser hücreleri, Dünya Sağlık Örgütü (DSÖ)'nün “Tümör Sınıflandırması” ile histolojik özellikleri ve kaynak dokularına göre sınıflandırılır. Karsinomlar, epitel hücrelerinden kaynaklanır ve skuamöz hücreli karsinom, adenokarsinom gibi alt tiplere ayrılır. Sarkomlar, mezenkimal dokulardan kaynaklanır; osteosarkom ve liposarkom gibi

tipleri içerir. Lösemiler ve lenfomalar, sırasıyla lenfatik ve kan oluşturan dokulardan gelen kanserleri ifade eder. Gliomalar ve melanomalar, sırasıyla glial hücrelerden ve melanositlerden kaynaklanan diğer alt tiplerdir (Louis ve ark., 2016).

Solid Tümörler

Katı tümörler, genellikle kist veya sıvı alanları olmayan, malign veya benign anormal doku kütleleridir. Bu tümörler, oluşturan hücre türüne göre sınıflandırılır: sarkomlar destekleyici veya bağ dokusundan, karsinomlar epitelyal veya bez hücrelerinden kaynaklanır. Solid tümörler kanser ve stromal hücrelerini içerir; stromal hücreler ekstraselüler matrikse gömülü fibroblastlar ve enflamatuvar hücrelerden oluşur ve bir damar ağı tarafından beslenir. Tümör bileşenleri bölgeye göre değişkenlik gösterir. Solid tümörler ile hematolojik kanserler arasındaki fark, hedef antijen tespitindeki karmaşıklıktır. Hematolojik kanserler genellikle belirli B-hücre işaretleyicilerini, örneğin CD19'u ifade eder (Trédan ve ark., 2007).

Hematolojik Maligniteler

Kan, kemik iliği ve lenf nodlarını etkileyen kanserler, hematolojik maligniteler olarak adlandırılır. Bu kategori, lösemileri (örn. akut lenfositik lösemi (ALL), kronik lenfositik lösemi (CLL), akut miyeloid lösemi (AML)) kapsamaktadır. Ayrıca, miyelom ve lenfomaları da (Hodgkin ve Hodgkin olmayan) içermektedir (Banerjee ve ark., 2018).

2.1.3. Kanser Patogenezi ve Progresyonu

Otto Warburg tarafından tanımlanan kanser hücrelerindeki farklılaşmış metabolik enerji üretim süreçleri, kanser patogenezi ve progresyonunun anlaşılmasında temeldir. “Warburg etkisi” olarak bilinen bu değişiklikler, kanser hücrelerinin artan glikolitik aktivite göstermesi ve normal hücrelere kıyasla daha fazla glukoz tüketmesidir. Bu, glukozun enerji üretimi ve anabolik işlemler için kullanılmasını sağlar (Kroemer ve Pouyssegur, 2008).

Artan glikoliz, oksijen varken bile laktat üretimi ve tümör çevresinde asidik bir mikroçevre oluşumuna yol açar. Glikoliz, oksidatif fosforilasyona göre daha az verimli olsa da kanser hücrelerinin büyümesi ve çoğalması için gerekli metabolik ara ürünleri üretir (DeBerardinis ve ark., 2007). Ayrıca, kanser hücreleri, yeterli oksijen varken bile

hipoksiye ilişkin yollarını etkinleştiren bir adaptasyon gösterir, buna “psödo-hipoksi” denir (Iommarini ve ark., 2017).

Anjiyogenez ve Metastaz

Kanserin ölümcüllüğündeki temel etken, kanser hücrelerinin birincil bir lezyondan distal organlara yayıldığı metastaz sürecidir. Bu süreçte birçok biyolojik yolk etkin rol oynar. Bu yollar arasında stromayı geçme veya stroma ile etkileşim kurma, bağışıklık sistemini engelleyerek veya kullanarak anti-tümör mekanizmaları devre dışı bırakma, doku mikroçevresini manipüle etme ve tedaviye karşı direnç geliştirme gibi mekanizmalar bulunmaktadır (Massagué ve Obenauf, 2016; Suhail ve ark., 2019).

Hücre Proliferasyonu ve Apoptoz

Farklılaşmış hücreler çoğunlukla metabolik olarak aktif ancak bölünmeyen G0 fazındadır. Mitojenik uyarılarla, hücreler DNA'nın replike edildiği S fazına girer, bu hücre döngüsünün başlangıcını işaret eder. Hücre döngüsü, S fazı, G2 fazı, G1 fazı (G1a, G1b, G1c, G1d olarak ayrılır) ve M fazından oluşur. Mitojen etkisi devam ederse hücre döngüsü tekrarlanır, aksi takdirde hücreler G0 fazına döner (Kontomanolis E. ve ark., 2021). Hücre döngüsünün aşamaları, Trombosit kaynaklı büyüme faktörü (PDGF), epidermal büyüme faktörü (EGF), fibroblast büyüme faktörü (FGF) gibi büyüme faktörleri, hormonlar ve sitokinler gibi mitojenler tarafından tetiklenir. Bu mitojenler, proto-onkogenlerin ifade ürünleri olup, hücre döngüsüne giriş ve ilerlemeyi düzenler. Hücre dışı matriks bileşenleri ve hücre yüzeyindeki integrinler de mitojenik sinyalizasyonda rol oynar, hücrelerin çevresel sinyallere uyumunu sağlar (Gavalas ve ark., 2013).

Apoptoz, organizmanın sağlık bir şekilde büyümesi için kritik, evrimsel olarak korunmuş bir programlanmış hücre ölümü sürecidir (Du Toit, 2013). Apoptozun moleküler mekanizması, mitokondriyal yolk ve ölüm reseptörlerinin aktivasyonu ile tetiklenen kaspaz proteazlarının aktivasyonuna dayanır (Green, 2018).

Mikroçevre ve Tümör Heterojenitesi

Tümör mikroçevre (TME), kanser hücrelerinin karmaşık hücresel ortamını tanımlar ve bağışıklık hücreleri, endotel hücreleri, kan damarları, hücrelerarası matriks, fibroblastlar

ve çeşitli sinyal moleküllerini kapsar (Grivennikov ve ark., 2010). Bu bileşenlerin etkileşimleri, kanserin başlangıcını ve ilerlemesini etkiler. Endotel hücreleri, tümörlerin büyümesini ve bağışıklık sisteminden kaçışını destekleyen damarları oluşturur (Hanahan ve Weinberg, 2011). TME içindeki malign olmayan hücreler (fibroblastlar gibi), tümörün gelişimine katkıda bulunur. Malign hücreler ise, sağlıklı dokulara ve diğer vücut bölgelerine yayılabilirler (LeBleu, 2015; Arneth, 2019).

Bağışıklık hücreleri, tümörlerin neden olduğu enflamatuvar reaksiyonlara katılır. Makrofajlar, tümör gelişimini çok yönlü etkileyen fonksiyonlara sahiptir ve metastatik yayılımı destekleyebilir (Melaiu ve ark., 2019). Doğal katil (NK) hücreleri, hasar görmüş ve enfekte hücrelere karşı sitotoksik etki gösterir. Ancak TME'nin immünsüpresif özellikleri bu hücrelerin işlevselliğini olumsuz etkiler (Barrow ve ark., 2019). TME'nin tüm bu bileşenleri, kanserin ilerlemesi üzerinde etkili olup, kanser tedavisi ve yönetiminde kritik bir rol oynamaktadır (Arneth, 2019).

2.1.4. Kanser Hücrelerinin Direnç Mekanizmaları

Kanser hücrelerinin direnç mekanizmaları, kanser tedavisinde önemli bir zorluk teşkil eder. Bu hücreler, çeşitli biyolojik yollar ve moleküler mekanizmalar aracılığıyla kemoterapi ve diğer antikanser tedavilerine karşı direnç geliştirir. Bu direnç, genetik mutasyonlar, epigenetik değişiklikler, ilaç atılımını artıran proteinlerin aşırı ifadesi ve hücre ölümü yollarının inhibisyonu gibi mekanizmalarla gerçekleşir. Kanser tedavisindeki bu zorlukları aşmak için, bu direnç mekanizmalarını anlamak ve bunları hedef alan yeni tedavi stratejileri geliştirmek hayati öneme sahiptir (Emran ve ark., 2022).

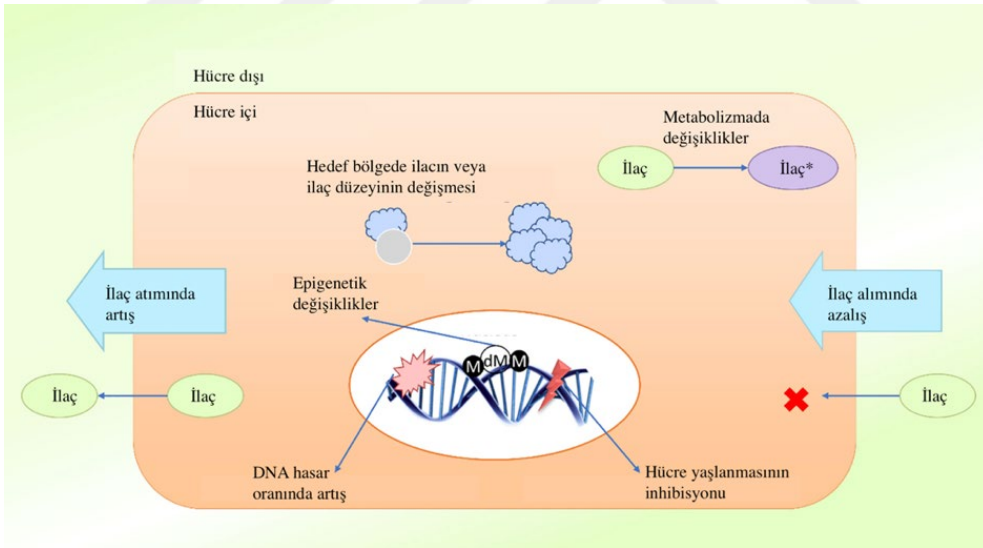
Çoklu İlaç Direnci

Çoklu ilaç direnci (MDR), kanser hücrelerinin bir kemoterapi ilacına direnç kazanırken aynı anda farklı yapılara ve etki mekanizmalarına sahip diğer kemoterapötiklere karşı da direnç kazanma yeteneğini ifade eder (Zahreddine ve Borden, 2013). Kanser hücreleri, ilaç alımının azalması, atılımın artması, DNA hasar onarımının iyileşmesi, apoptozun bastırılması, ilaç metabolizmasının değişimi, ilaç hedefinin modifikasyonu, epigenetik değişiklikler ve hedef gen amplifikasyonu yoluyla direnç geliştirir (Emran ve ark., 2022). Tedaviye direnç, ilk başlarda moleküler pompaların tümör hücrelerinden

kemoterapi ilaçlarını atmasıyla ilişkilendirilmiştir. Bu pompalar arasında P-glikoprotein ve MDR ile ilişkili protein, kanser tedavisinde önemli hedeflerdir (Persidis, 1999).

MDR genleri, ilaç direncinin gelişiminde önemli bir rol oynar. İnsanlarda *MDR1* ve *MDR2* gibi iki ana MDR geni ifade edilir. *MDR1*, P-glikoprotein kodlayarak belirli kemoterapötik ajanlara karşı direnç gelişimini sağlar. İlaç direnci, hem intrinsik (önceden var olan) hem de kazanılmış (ilaçlar tarafından indüklenen) mekanizmalar aracılığıyla oluşabilir (Wang Z. ve ark., 2015).

Genetik değişikliklere bağlı kalıtsal faktörler, tümör hücrelerinin çoğunluğunda kemoterapi ve hedef ilaçlara azalmış yanıtlara neden olabilir. Mutasyonlar, gen amplifikasyonları, transpozon elementler, gen delesyonları, gen translokasyonları ve mikro RNA (miRNA) değişiklikleri gibi genotipik farklılaşmalar, kanser hücrelerinde genomik kararsızlık ve genetik çeşitliliğe neden olmaktadır (Emran ve ark., 2022).



Şekil 2.2. Kanser hücrelerinin kemoterapötiklere karşı potansiyel direnç mekanizmaları (Emran ve ark., 2022).

Olası direnç mekanizmaları Şekil 2.2’de özetlenmiştir. Bu mekanizmalar, ilaç alımının azalması, atılımının artması, DNA hasar onarımının iyileşmesi, apoptozun baskılanması, ilaç metabolizmasında ve hedeflerinde değişiklikler, epigenetik modifikasyonlar ve hedef gen amplifikasyonu aracılığıyla kanser hücrelerinde ilaç direncine yol açar (Emran ve ark., 2022).

2.2. Kemoterapötikler

Kemoterapötik ajanlar, kanserden enfeksiyonlara kadar çeşitli hastalıkların tedavisinde kullanılır. Antineoplastik ve antimikrobiyal ajanlar olarak iki ana kategoriye ayrılırlar. Antineoplastikler, kanser hücrelerinin çoğalmasını engeller veya yok ederler. Antimikrobiyal ajanlar enfeksiyonlara neden olan patojenleri hedeflerler (Kayaalp, 2012).

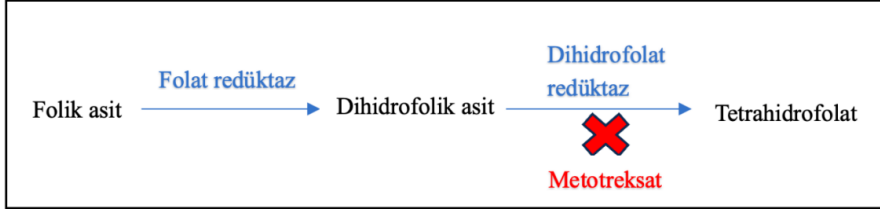
Antimikrobiyal ajanlar, bakteriyel hücre duvar sentezini inhibe eden antibakteriyeller, viral replikasyonu engelleyen antiviraller, mantar hücre zarlarını hedefleyen antifungal ve antiparazitik ajanları içerir. Bu ilaçların amacı, enfeksiyon hastalıklarının semptomlarını azaltmak ve yayılımını engellemektir (Kayaalp, 2012). Antineoplastik ajanlar ise, kanser hücrelerinin DNA replikasyonunu ve mitoz süreçlerini hedef alır. Bu kategori, topoizomeraaz inhibitörleri, antimitotik ajanlar, antimetabolitler ve alkilleyici ajanlar gibi çeşitli ilaçları içerir (Chabner ve Roberts, 2005).

2.2.1. Antineoplastik Ajanlar

Geleneksel platin bazlı kemoterapileri ve kansere karşı yeni hedefe yönelik tedavileri kapsayan birçok antineoplastik ajan bulunmaktadır. Bunların hemen hemen hepsi belirli derecede hepatotoksisiteye neden olmakta ve karaciğer hasarı genellikle doğrudan, intrinsik toksisiteden kaynaklanmaktadır. Bu tedavi çeşitlerinden örnekler aşağıda detaylandırılmıştır (Brunton ve ark., 2018).

Alkilleyici ajanlar, DNA bazlarına bağlanarak geniş spektrumda etkiler gösterirler ve bu sayede kanser tedavisinde önemli bir role sahiptirler. Pürin ve pirimidin bazları bu ajanların birincil hedefleridir ve alkilasyon süreci hem sitotoksik hem de mutajenik sonuçlar doğurabilir. Örneğin, siklofosfamid gibi azotlu hardallar grubu, hematolojik ve solid tümörlerde kullanılır (Kayaalp, 2012).

Antimetabolitler ise, nükleik asit bileşenlerini taklit ederek DNA ve RNA sentezini inhibe ederler. Metotreksat gibi ilaçlar, Şekil 2.3'te gösterildiği üzere dihidrofolat redüktazı inhibe ederek etki gösterir (Allegra ve ark., 1985).



Şekil 2.3. Folik asit yolağı ve metotreksat (Allegra ve ark., 1985).

Vinka alkaloidleri, hücre bölünmesini mikrotübül mekanizmalarını hedef alarak engellerken, sitotoksik antibiyotikler DNA işlevlerini bozarak hücre ölümüne yol açarlar. Platin bazlı ilaçlar, DNA ile platin-DNA kompleksleri oluşturarak kanser hücrelerini öldürür. L-asparaginaz, kanser hücrelerinin protein sentezi için gerekli olan asparajini hedef alır. Hormon tedavileri ve antagonistleri, hormon duyarlı tümörlerde kullanılır ve hormon reseptörlerini hedefler (Kayaalp, 2012).

İmmünoterapi, kanser ile savaşan bağışıklık sistemini düzenleyerek etki eder ve CAR-T hücre tedavisi gibi yenilikçi yaklaşımları içerir. Hassas tıp, hastanın tümörünün genetik yapısına göre uyarlanan tedaviler sunar ve spesifik mutasyonları hedef alır. Bu geniş yelpazede ilaçlar, kanser tedavisinde çeşitli mekanizmalarla etkinlik gösterir ve tedavinin özelleştirilmesine olanak tanır (Trapani ve Darcy, 2017; Szeto ve Finley, 2019).

2.2.2. Antimikrobiyal Ajanlar

Antibiyotikler ve antiviral ilaçlar, mikrobiyal hastalıkların tedavisinde kritik öneme sahiptirler. Beta-laktam antibiyotikleri, bakteriyel hücre duvar sentezini inhibe ederek etki gösterir, ancak beta-laktamaz direnci, bu sınıfın etkinliğini sınırlandıran bir sorundur. Makrolid, linkozamid ve streptogramin antibiyotikler, 50S ribozomal alt birimine bağlanarak protein sentezini inhibe ederler ve özellikle Gram-pozitif bakterilere karşı etkilidirler. Tetrasiklinler, 30S ribozomal alt birimine bağlanarak protein sentezini bozarlar. Amfenikoller, geniş spektrumlu etkileriyle dikkat çeker, ancak ciddi yan etkileri nedeniyle sınırlı kullanıma sahiptirler. Aminoglikozidler, güçlü bakterisidal aktivite gösterirler ancak nefro- ve oto-toksisiteleri nedeniyle kullanımları dikkatle yapılmalıdır (Kayaalp, 2012).

Dar spektrumlu antistafilokokal ve antianaerobik ilaçlar, hedefe yönelik tedavi sağlarlar. Polipeptit yapılı antibiyotikler, özellikle çok dirençli gram-negatif enfeksiyonlar için önemlidir. Florokinolonlar, DNA replikasyonunu inhibe ederek bakteriyel hücre ölümüne yol açarlar. Sülfonamidler ve trimetoprim, bakteriyostatik ve bakterisidal etkileriyle folik asit sentezini hedef alırlar. Antifungaller, mantar enfeksiyonlarına karşı çeşitli mekanizmalarla etkilidir ve ergosterol sentezini hedef alarak mantar hücre zarı bütünlüğünü bozarlar (Kayaalp, 2012).

Tüberküloz ve diğer mikobakteriyel enfeksiyonlar, özel tedavi rejimleri gerektirir, bu rejimler genellikle birden fazla ilacın kombinasyonunu içerir. İdrar yolu enfeksiyonlarının tedavisinde özel ilaçlar kullanılır ve bu ilaçlar genellikle yüksek idrar konsantrasyonlarına sahip olup direnç oranları düşüktür. Antimalaryaller, antihelmintikler ve antiviraller, özgül mikroorganizmalara karşı geliştirilmiş özel tedaviler sunarlar. Antiviraller, viral girişi engelleyerek veya viral replikasyonu bozarak işlev görürler ve modern terapötik stratejilerde önemli bir rol oynarlar. (Kayaalp, 2012).

2.3. Peptitler

Peptitler, genellikle 10 ila 100 amino asit uzunluğunda olan ve peptit bağları ile birbirine bağlanmış kısa amino asit zincirleridir. Bu bileşenler, biyolojik sistemler için temel öneme sahip olup, yapısal olarak proteinlere benzemelerine rağmen daha küçük boyutlarıyla ayrılırlar. Peptitlerin sentezi, ribozomların mesajcı RNA tarafından iletilen talimatları takip ederek gerçekleşen translasyon sürecinde amino asitlerin birleştirilmesiyle olmaktadır (Weder ve Belitz, 2003).

Tablo 2.1’de Uluslararası Temel ve Uygulamalı Kimya Birliği’nin (IUPAC) tanımladığı amino asitler ve kısaltmaları listelenmiştir.

Tablo 2.1. Amino asitler ve kısaltmaları.

Tek harfli kod	3 harfli kod	Amino asit
A	Ala	Alanin
B	Asx	Aspartat veya Asparajin
C	Cys	Sistein
D	Asp	Aspartat
E	Glu	Glutamat
F	Phe	Fenilalanin
G	Gly	Glisin
H	His	Histidin
I	Ile	İzolösın
J	Xle	Lösın ve İzölösın
K	Lys	Lizin
L	Leu	Lösın
M	Met	Metiyonin
N	Asn	Asparajin
O	Pyl	Pirolizin
P	Pro	Prolin
Q	Gln	Glutamin
R	Arg	Arjinin

Tablo 2.1. (Devam) Amino asitler ve kısaltmaları.

S	Ser	Serin
T	Thr	Treonin
U	Sec	Selenosistein
V	Val	Valin
W	Trp	Triptofan
X	Xaa	Spesifik olmayan veya bilinmeyen
Y	Tyr	Tirozin
Z	Glx	Glutamin veya Glutamat

(IUPAC, 1984; Kotyk, 1999) kaynak alınarak düzenlenmiştir. Renklendirilmiş satırlar, standart olmayan amino asitlere aittir.

2.3.1. Peptitlerin Biyolojik Fonksiyonları

Peptitler, sinyal iletiminden hücresel fonksiyonların düzenlenmesine kadar çeşitli biyolojik roller üstlenir. Örneğin, insülin ve glukagon gibi peptit yapıları hormonlar glikoz metabolizması gibi birçok fizyolojik süreci düzenlerler (Weder ve Belitz, 2003).

Sinir sisteminde, peptitler nörotransmitter veya nöromodülatör olarak işlev görür. P maddesi ve enkefalinler gibi opioid peptitler, ağrı algısının modülasyonuna katkıda bulunur. Bu moleküller, nöronlar arasında veya nöronlar ile kaslar arasında iletişimi kolaylaştırır. Ayrıca, AMP'ler doğuştan gelen bağışıklık sisteminin kritik bir bileşeni olup, mikrobiyal hücre zarlarını bozarak patojenlere karşı etkili bir ilk savunma hattı oluşturur. Peptitlerin terapötik potansiyeli, yüksek özgüllük ve etkinlikleri, düşük toksisite ve immünojenisite profilleri ile son yıllarda artan ilgi görmektedir. Zikonotid gibi şiddetli kronik ağrı tedavisinde ve antikanser ajan linaklotid gibi ilaçlarda klinikte kullanılan peptitler, bu alanın öne çıkan örnekleridir (Weder ve Belitz, 2003).

Peptitler, teşhis alanında da kritik bir öneme sahiptir. Biyobelirteç olarak işlev gören peptitler, hastalık durumlarını veya terapötik yanıtları belirlemede göstergeler olarak kullanılır. Özellikle tümör görüntülemeye kullanılan peptitler, hedeflenen hücreleri veya dokuları tanımlama kapasitesiyle tanı prosedürlerinin doğruluğunu ve etkinliğini artırır (Forbes ve Krishnamurthy, 2023).

2.3.2. Peptitlerin Sınıflandırılması

Birincil yapı, peptidin genel şeklini ve işlevini belirleyen temel amino asit dizisini ifade eder. İkincil yapı, proteinlerin üç boyutlu katlanmalarını ifade eder. En yaygın ikincil yapılar, alfa-helikal ve beta-yaprak oluşumlarıdır. Alfa-helikal yapısına sahip proteinlerde, amino asitler arasında oluşan hidrojen bağları, düzenli bir sarmal oluştururlar. Hidrojen bağları, bir amino asidin amonyum grubunun bir sonraki amino asidin karboksil grubu ile etkileşime girmesiyle meydana gelir. Oluşan sarmal yapının her tam turu genellikle 3,6 amino asit uzunluğundadır. Beta-sarmallar ise paralel ya da anti-paralel bir şekilde yan yana dizilmiş beta iplikçiklerinden oluşur. Bu iplikler, bitişik ipliklerin karbonil oksijen ve amid hidrojen grupları arasındaki hidrojen bağları ile bir arada tutulur. Beta-setlerin sertliği ve kıvrımlı görünümü, onları birçok proteinde güç ve stabilite sağlayan önemli bir yapısal bileşen haline getirir (Wang G., 2017).

Amino Asit Sayısına Göre Sınıflandırılması

Bir peptidin yapısal karmaşıklığı ve biyolojik aktivitesi, içerdiği amino asit miktarından etkilenir. Dipeptitler, iki amino asidin peptit bağı ile birleşmesi sonucu oluşan en basit peptit yapısını temsil eder. Bu yapılar, vücutta sinyal iletiminde rol alabilir veya protein parçalanması sonucunda yan ürün olarak meydana gelebilir. Amino asit miktarı arttıkça peptidin yapısal karmaşıklığı ve biyolojik aktivitesi de artar. (Weder ve Belitz, 2003).

Üç amino asitten oluşan tripeptitler ise daha karmaşık yapılara ve rollere sahiptir. Tripeptit yapısına sahip glutatyon, glutamat, sistein ve glisinden oluşan, hücresel antioksidatif faaliyetler için temel bir moleküldür. Bir diğer önemli tripeptit ise glutamat, histidin ve prolinden oluşan Tirotropin Salgılatıcı Hormon'dur (TRH). TRH, hipofiz bezinden tiroid uyarıcı hormon (TSH) salınımını uyarak tiroit bezinin düzenlenmesinde aktif rol alır (Nelson ve Cox, 2021).

Genellikle 4-10 amino asitten oluşan oligopeptitler daha büyük bir grubu temsil eder. Bu peptitler hormon, nörotransmitter ve antibiyotik olarak işlev görebilir. Nispeten küçük boyutlara sahip olmalarının sağladığı avantajla, hücresel membranlara kolaylıkla nüfuz edebilirler, özellikle sinyalizasyonda ve hücresel regülasyonlarda etkin rol oynarlar (Weder ve Belitz, 2003).

Bir diğer önemli oligopeptit grubu, tipik olarak altı ila sekiz amino asit içeren defensinlerdir. Bu peptitler doğuştan gelen bağışıklık sisteminin önemli bileşenleridir ve doğal antibiyotik görevi görürler. Defensinler bakteri, mantar ve virüsler de dahil olmak üzere çok çeşitli patojenlere karşı antimikrobiyal aktivite sergiler. Bu patojenlerin hücre zarlarını bozarak çalışarak tehditleri etkisiz hale getirir (Weder ve Belitz, 2003).

Ondan fazla amino asit içeren polipeptitler, dokulardaki yapısal bileşenlerde, enzimlerde ve hormonlarda yer alabilir. Polipeptit zincirinin uzunluğu oldukça değişkenlik gösterebilir, bazıları birkaç yüz amino aside varan uzunluklara sahip olabilir. Bu uzunluk ve amino asitlerin spesifik dizilimi, polipeptidin son derece spesifik ve karmaşık olabilen üç boyutlu yapısını ve işlevini belirler (Nelson ve Cox, 2021).

Şekillerine Göre Sınıflandırılması

Peptitler, şekillerine göre sınıflandırıldığında, en yaygın olanı doğrusal peptitlerdir. Bu peptitler, sıralı, dallanmamış bir zincir halinde düzenlenmiş bir dizi amino asitten oluşur. Doğrusal yapıları, çeşitli biçimler almalarına izin veren esneklik kazandırır. Bu yapısal esneklik, biyolojik reseptörler ve enzimlerle etkileşimlerinde çok önemlidir. Ancak, aynı zamanda onları enzimatik bozulmaya daha duyarlı hale getirir, bu da biyolojik sistemlerdeki stabilitelerini ve yarı ömürlerini önemli ölçüde etkileyen bir faktördür (Nelson ve Cox, 2021).

Öte yandan, siklik peptitler daha karmaşık bir konformasyonu temsil eder. Bu yapılarda, peptit zinciri kendi üzerine geri dönerek, tipik olarak peptit zincirinin N-terminal ve C-terminal uçları arasında kovalent bağlanma yoluyla döngüsel bir yapı oluşturur. Bu kapalı döngü yapısı peptite daha fazla stabilite kazandırarak onu enzimatik parçalanmaya karşı daha dirençli hale getirir. Ayrıca, döngüsel üç boyutlu yapısı

peptidin esnekliğini kısıtlar ve genellikle biyolojik hedeflerle etkileşimlerde daha yüksek özgülük ve afiniteye yol açar (Nelson ve Cox, 2021).

Dallanmış peptitler, peptit zincirinin tek bir çekirdekten birden fazla dala ayrıldığı başka bir kategoriye oluşturur. Bu dallanma, dendrimerler veya multimerik kompleksler gibi çeşitli şekillerde meydana gelebilir. Dallanmış peptitler aynı anda birden fazla bağlanma bölgesiyle etkileşime girerek çok hedefli biyolojik süreçlerdeki etkinliklerini artırabilir (Brunetti ve ark., 2018).

Biyolojik Rollerine Göre Sınıflandırılması

Kısa amino asit zincirlerinden oluşan peptitler, çok sayıda biyolojik sürecin ayrılmaz bir parçasıdır ve çeşitli biyolojik rollerine göre sınıflandırılabilirler. İnsülin ve glukagon gibi hormonal peptitler metabolik düzenlemede önemli rol oynarken, P maddesi ve opioid peptitler gibi nörotransmitter peptitler sinirsel iletişim ve ağrı modülasyonunda hayati öneme sahiptir (Rose, 2019; Forbes ve Krishnamurthy, 2023).

Bir diğer önemli sınıf ise doğuştan gelen bağışıklık sisteminin kritik bir bileşenini oluşturan antimikrobiyal peptit (AMP)'lerdir. Bu peptitler, hücre zarlarını bozarak bakteriler, mantarlar ve virüsler gibi pek çok patojenlerle mücadele edebilirler. Bu geniş spektrumlu aktivite ve terapötik uygulama potansiyeli AMP'leri kapsamlı araştırmaların odağı haline getirmektedir (Gordon ve ark., 2005).

Peptitler ayrıca çeşitli biyokimyasal reaksiyonlarda enzimatik olarak işlev görür. Bazıları bizzat enzim olarak işlev görürken, diğerleri substrat veya inhibitör olarak görev yaparak önemli biyolojik yolları düzenler. Örneğin, anjiyotensin kan basıncı ve sıvı dengesinde önemli bir rol oynar (Rose, 2019). Peptitler, vücutta çeşitli ve hayati roller oynarlar (Rose, 2019). Kanseri ve çeşitli biyolojik hedefler ile ilişkili farklı peptit gruplarına ait örnekler aşağıda incelenmektedir.

Magaininler ilk olarak Afrika pençeli kurbağası (*Xenopus laevis*) derisinden izole edilmiş ve bu peptitlerin geniş spektrumlu antibakteriyel, antifungal ve bazı durumlarda antiviral aktivitelere sahip oldukları gözlemlenmiştir. Moleküler düzeyde, genellikle hücre zarlarını hedef alırlar ve bu zarların bütünlüğünü bozarak mikroorganizmaların

veya kanser hücrelerinin ölümüne yol açarlar. Bu mekanizma, magaininlerin hücre içi direnci aşabilmesine ve antibiyotik direnci olan bakteri türlerine karşı etkili olabilmesine olanak tanır (Moore ve ark., 1991).

Cecropin ailesi, doğal olarak bulunan antimikrobiyal peptitlerden oluşur ve birçok organizmanın enfeksiyonlara karşı savunma mekanizmasının bir parçasıdır. Bu peptitler, özellikle böceklerde bulunur ve patojen mikropları hedef alarak hücre zarlarını bozarlar, böylece mikroorganizmaların ölümüne yol açarlar (Hultmark ve ark., 1982).

Katelisidin ailesi, memelilerin ve bitkilerin doğal savunma sisteminin bir parçası olarak patojenlerin hücre zarlarını hedef alarak onları etkisiz hale getirirler. Bu peptitler, bitkilerin enfeksiyonlara karşı korunmasına yardımcı olur ve aynı zamanda bitkilerin herbivorlara karşı savunma mekanizmalarında da rol oynayabilirler (Febriza ve ark., 2019).

Defensin ailesi, antimikrobiyal peptitlerin oluşturduğu bir gruptur ve patojen mikropların hücre zarlarını hedef alarak birçok organizmanın enfeksiyonlara karşı doğal savunma mekanizması içinde rol oynar (Seebah ve ark., 2007).

Taşikinin peptitleri, sinir sistemi ve bağışıklık sistemiyle ilişkili çeşitli biyolojik işlevlere sahiptir. Madde P, ağrı algısı ve enflamasyonun artırılması gibi roller üstlenirken, kassinin kurbağa kaynaklı olup sinir iletiminde etkilidir (De Felipe ve ark., 1998). Nörokinin A ve B, sinir ve bağışıklık sistemlerinde iletişimi sağlar, enflamasyon ve ağrı yanıtlarını düzenler (De Felipe ve ark., 1998; dos Santos ve ark., 2012).

Vazoaktif Bağırsak Peptitleri (VIP) ve türevleri, gastrointestinal sistemde vazodilatasyon yaparak kan basıncını düşürür ve bağışıklık sistemini modüle eder (La Rosa, 2022). Sekretin, sindirimi düzenleyen bir peptit hormondur, pankreas salgılarını ve safra akışını uyarır (Chu ve ark., 2006). Pankreatik Polipeptit (PPY), besin alımına yanıt olarak safra kesesi kontraktilesini azaltır ve sindirimi yavaşlatır (Deng ve Whitcomb, 2006).

Opioid peptitler, Proopiomelanokortin (POMC) ve türevleriyle ağrı modülasyonu ile enerji dengesinde rol oynar (Millington, 2007). Enkefalinler, opioid reseptörlerine bağlanarak ağrıyı düzenler (Cullen ve Cascella, 2023). Kalsitonin ise, kalsiyum

metabolizmasını düzenleyen ve kemik sağlığı üzerinde etkileri olan bir peptit hormondur (Sexton ve ark., 1999).

2.3.3. Antimikrobiyal Peptitler

AMP'ler, konakçının doğuştan gelen bağışıklığına katkıda bulunan, ribozomlar tarafından üretilen küçük moleküler yapılardır ve çeşitli canlı organizmalarda bulunurlar. Genellikle katyonik ve amfipatik yapıda olup, bakteri, mantar, protozoa ve virüslere karşı geniş spektrumlu antimikrobiyal etkilere sahiptir (Haney ve ark., 2017). İmmün regülasyon, anjiyogenez, yara iyileşmesi ve antitümör aktiviteleri gibi özellikleri, bu yapıların kanser ve enfeksiyon hastalıklarında terapötik potansiyeline işaret eder (Zhang Q.-Y. ve ark., 2021).

AMP'ler, çeşitli amino asit bileşimleri, kökenleri ve yapılarına göre sınıflandırılır. Sistein içermeyen α -helikal doğrusal moleküller ve sistein içeren polipeptitler iki ana grup oluşturur. Amfipatik α -sarmal peptitler içeren katelisidin LL-37, disülfür bağları ile stabilize β -yaprak içeren defensinler ve döngü peptitler gibi çeşitli alt kategoriler bulunur. Memeli defensinleri ve katelisidinler, araştırmalarda öne çıkar. Defensinler, üç disülfür köprüsü ve altı korunmuş sistein kalıntısı içerir; α -, β - ve θ -defensinler olarak kategorize edilir. İnsan α -defensinleri, doğal öldürücü hücreler, mast hücreleri, monositler ve makrofajlar tarafından salınır, β -defensinler (hBD'ler) ise mukoza ve integüment epitel hücrelerinde ifade edilir ve sitokinlerle modüle edilebilir (Mookherjee ve ark., 2020). Klinikte kullanılan birçok peptidin; araştırma başlangıç tarihleri, uygulama yolları ve menşei ülkeleri gibi bilgiler Tablo 2.2'de görülmektedir.

Tablo 2.2. Çeşitli terapötik peptitler ve klinik durumları.

Araştırma Başlangıcı	Uygulama Yolu	Etken Madde	Menşei Ülke	Durum
1945	Lokal	Basitrasin	ABD	Kullanımda
-	Lokal	Gramisidin	ABD	Kullanımda
1953 (preklinik)	Sistemik, IV/Oral	Vankomisin	ABD	Kullanımda

Tablo 2.2. (Devam) Çeşitli terapötik peptitler ve klinik durumları.

-	Lokal, İnhalasyon	Kolistin	-	Kullanımda
1986 (preklinik)	Sistemik, IV	Daptomisin	ABD	Kullanımda
1996 (preklinik)	Sistemik, SC	Enfuvirtid	ABD	Kullanımda
-	Sistemik, IV	Telavansin	ABD	Kullanımda
2005 (Faz II)	Sistemik, Oral	Boceprevir	ABD	Kesildi
2007 (Faz II)	N/A	Telaprevir	ABD	Kesildi
2001 (Faz II)	Sistemik, IV	Dalbavansin	ABD	Kullanımda
2001 (N/A)	Sistemik, IV	Oritavansin	ABD	Kullanımda
-	Sistemik, IV/IM	Teikoplanin	Fransa	Kullanımda
2010	Sistemik, IV	Tebentafusp	ABD	Kullanımda
1974	Oral ve SC	Buserelin	Avrupa	Kullanımda
2016 (Faz I)	Sistemik, IV	Plitidepsin	Avrupa	Kullanımda
1975	SC	Triptorelin	ABD	Kullanımda
-	Sistemik, IV	Daktinomisin	ABD	Kullanımda

(Wang L. ve ark., 2022) kaynak alınarak düzenlenmiştir.

Doğal sentetik veya ribozomal sentetik peptitler ve non-ribozomal peptitler antibakteriyel peptitler için diğer kategorilerdir. Ayrıca, elektrostatik yük AMP sınıflandırması için önemli bir özelliktir. Bu nedenle, bu özelliğe dayalı olarak katyonik peptitler ve katyonik olmayan peptitler olarak sınıflandırılır. AMP'lerin ve antikanser peptit (ACP)'lerin amino asitlerinin türü, sayısı ve bileşimi aktivitelerinde kritik bir rol oynar. Yapısal değerlendirmelerinde, ikincil yapılarına göre dört ana gruba ayrılırlar. Bunlar rastgele sarmallar, β -sarmal peptitler, genişletilmiş yapılar ve α -helikal çubuk

konformasyonlarını içerir. Alfa-helikal ACP'ler olarak bilinen en yaygın ACP türü basit ve kısadır, genellikle amfibilerin epidermisinde bulunur ve kapsamlı bir şekilde araştırılmıştır. β -yaprak peptitler çok sayıda disülfid bağıyla daha karmaşık yapılar oluştururlar, bitkisel veya hayvansal kaynaklı olabilirler. Rastgele sarmal ACP'ler, prolin ve glisin kalıntıları bol miktarda bulunur ve karakteristik bir ikincil yapıdan yoksundur. Son olarak, siklik ACP'ler, sistein düğümleri veya baştan kuyruğa siklizasyon oluşturan disülfid bağlantılarının omurgasına sahip kapalı peptitlerdir ve doğrusal formlara kıyasla daha stabildirler (Jafari ve ark., 2022).

Biyosentezi ve Yapısı

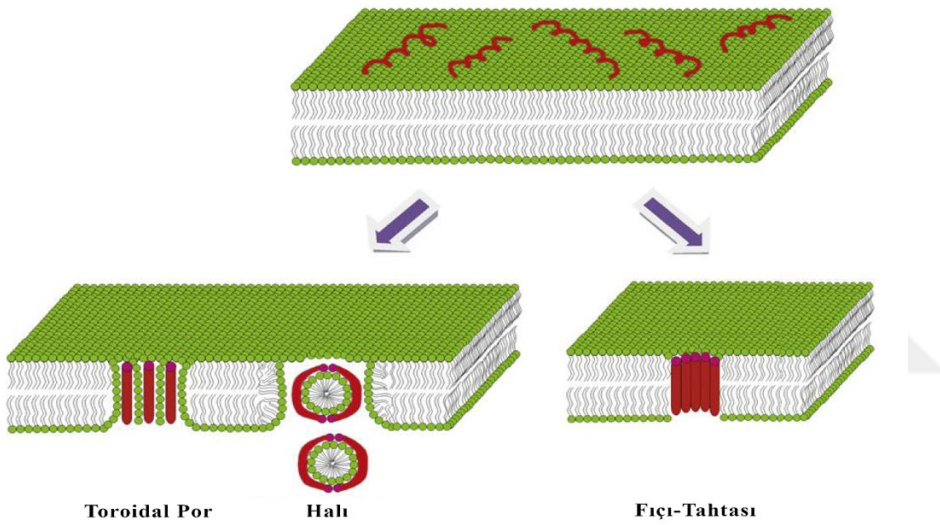
AMP'ler memelilerde genellikle epitelyal hücreler tarafından üretilir. Bu peptitler, vücudun günlük olarak birçok bakteri türüne maruz kaldığı bölgelerde bulunabilir. Bu bölgeler, dokular, sıvılar ve mukoz zarlar gibi vücut yüzeylerini içerir. Ek olarak, AMP'ler bağışıklık sistemi içindeki çeşitli hücreler tarafından üretilir, özellikle fagositler gibi hücrelerde granüller içerisinde bulunurlar (Haney ve ark., 2017).

AMP'ler üzerine yapılan erken araştırmalar, doğal kaynaklardan peptitlerin saflaştırılmasını ve elde edilen örneklerin antibakteriyel aktivitesinin değerlendirilmesini içermektedir. Ancak, çok düşük miktarlarda peptit elde etmek için bile büyük miktarda işlenmemiş biyolojik malzemenin gerekmesi, bu süreci sınırlandıran en önemli faktördür. Ayrıca, doğal kaynak temelli AMP'ler, genellikle daha büyük öncü proteinlerin proteolitik olarak kesilmesiyle üretilmektedir. AMP'lerin post-translasyonel modifikasyonları da istenen peptidin benzersiz ve/veya etkin bir formda saflaştırılmasını zorlaştırır. Sıklıkla, peptitlerin yüksek verimlilik ve saflıkla elde edilmesi için kimyasal sentez yöntemleri tercih edilmektedir (Haney ve ark., 2017).

Etki Mekanizmaları

AMP'ler, birbirlerinden farklı aktivite profillerine sahiptir ve bu nedenle ortak bir etki mekanizmasından söz etmek zordur. Bakterilerle etkileşim söz konusu olduğunda, AMP'ler başlangıçta bakteri hücresinin yüzeyindeki anyonik moleküllere elektrostatik kuvvetler aracılığıyla çekilir. Gram-negatif bakterilerde lipopolisakkaritlere (LPS) ve gram-pozitif bakterilerde lipoteik asitlere bağlanırlar. AMP'ler bakteri hücre duvarının içine geçtiklerinde, özellikle fosfatidilgliserol ve kardiyolipin gibi negatif yüklü lipitlerle

etkileşime girer. Bu, stabilizasyonu sağlayan iki değerlikli katyonları yerinden çıkararak membranı bozmaktadır (Haney ve ark., 2017). Böylece negatif yüklü zarların geçirgenliğini artırarak hücrenin lize edilmesine veya hücre içi içeriklerin salınması neticesiyle hücre ölümünün gerçekleşmesine neden olurlar. Ancak bu aşamaya ulaşmadan önce, AMP'ler kapsüler polisakkaritler ve Gram-negatif bakterilerdeki LPS gibi yapılar ve Gram-pozitif bakterilerdeki lipoteik asit ve peptidoglikan gibi diğer hücresel bileşenleri aşmak zorundadır. Bu aşamada, konformasyonel değişiklikler ve peptit-lipit oranı önem kazanmaktadır (Zhang C. ve ark., 2019).



Şekil 2.4. AMP'lerin olası membran etkileşim mekanizmaları (Schrack ve ark., 2013).

Literatürde, AMP'lerin hücre zarları üzerindeki etkilerini açıklamaya yönelik teorileri temsil eden modeller bulunmaktadır. Bu modeller, fıçı-tahtası, toroidal por, kümelenme ve halı modelleridir (Şekil 2.4).

AMP'lerin hücre membranlarıyla etkileşimleri farklı modellerle açıklanabilir. Toroidal modelde, helezonal yapıda dizilmiş peptitler, membrana paralel olarak yerleşir. Bu modelde, peptitlerin hidrofobik amino asitleri membranın polar grupları ile yer değiştirerek lipid katmanında boşluklar oluşturur. Bu boşluklar, membranın bütünlüğünü bozarak hücrenin işlevselliğini etkiler (Leontiadou ve ark., 2006).

Fıçı-tahtası modelinde ise, peptitler hücre membranını çemberleyerek kendilerini hidrofobik çekirdeğe yerleştirirler. Bu yerleşim, membran içinde silindirik gözeneklerin

oluşumuna sebep olur. Bu gözenekler, hücre içeriği akışını bozarak hücrenin zarar görmesine neden olur (Schrank ve ark., 2013).

Toroidal por modeli, AMP'lerin membrana yerleştiğini ve lipidleri bükerek peptite kaplı bir gözenek oluşturduğunu öne sürer. Bazı AMP'lerin bu modelle gözenek oluşturduğu öne sürülmüş olsa da bu yapı için yeterli bilimsel destek bulunmamaktadır (Schrank ve ark., 2013).

Kümelenme modelinde, AMP'ler membranın içine girerek kümeleşir ve büyük agregatlar oluşturur. Bu agregatlar, membran içinde iyonların ve diğer moleküllerin geçişine izin veren kanalların oluşumuna yol açar. Bu kanallar, hücre içi ve dışı arasında dengeyi bozarak hücrenin fonksiyonlarını etkiler ve potansiyel olarak hücre ölümüne neden olur (Fjell ve ark., 2012).

Halı modelinde, AMP'ler membran yüzeyinde geniş bir alanı kaplayacak şekilde düzenlenir. Bu model, peptitlerin membranın üst katmanında birikmesi ve lipid çift katmanına nüfuz ederek, yapısal bozulmalara ve gözeneklerin oluşmasına yol açar. Bu gözenekler hücre içeriği sızıntısına ve sonuçta hücre ölümüne neden olur (Schrank ve ark., 2013).

AMP'lerin biyolojik membranlarla etkileşiminin tam mekanizması ve her peptit türünün etkileşim biçimi çeşitlilik gösterir. Bu çeşitlilik, AMP'lerin membran bariyer özelliklerini nasıl değiştirdiklerini anlamak için farklı modellerin incelenmesini gerektirir. Bu etkileşimlerin anlaşılması, AMP'lerin hücre zarı üzerindeki etkilerini ve terapötik potansiyellerini belirlemek için önemlidir. AMP'ler, hücre içine alındıklarında DNA, RNA ve protein sentezini engelleyebilir ve sitoplazmik enzimatik aktiviteyi bozarak hücrenin metabolik işlevlerini etkileyebilir. (Fjell ve ark., 2012).

2.3.4. Antikanser Etki Gösteren Peptitler

Tümör hücrelerinin ACP'ler tarafından tanınması, tümör mikroçevresini değiştiren yüksek düzeyde reaktif oksijen türleri (ROS) ve hipoksi nedeniyle yüzeylerindeki negatif yüklü fosfatidilserin (PS) moleküllerinin varlığı ile kolaylaştırılır. Tümör hücreleri plazma zarının dış ve iç tabakaları arasındaki fosfolipid dağılımının simetrisini

kaybederler ve böylece tümör hücrelerinin dış yüzeyinde PS'yi açığa çıkarırlar (Tornesello ve ark., 2020).

Tümörlerin seçici olarak hedeflenmesi için yapı gerekliliklerine ilişkin sonuçlar hala belirsizliğini korumaktadır. Mevcut sonuçlar, ACP'lerin katı tümörleri çeşitli mekanizmalarla hedeflediğini göstermektedir. Meme, prostat, uterus serviksi, karaciğer ve akciğer ACP'lerin geliştirilmesi için hedeflenen tümörlerden bazılarıdır. Bu peptitlerden bazıları hücrel membranlara zarar verdikten sonra apoptotik veya nekrotik mekanizmalarla, diğerleri ise hücre içi hedeflerle malign hücreleri etkisiz hale getirir ve tek bir peptidin birden fazla antikanser aktivite modu sunması da mümkündür (Gaspar ve ark., 2013).

Kanser Hücrelerinin Proliferasyonu, Metastaz ve İnvazyonunun İnhibisyonu

HBD-3'ün, tümör içine sızan monositler tarafından üretildiği ve bu bileşiğin, kolon kanseri hücrelerinin invazivlik ve hareketliliğini doza bağlı olarak engellediği belirlenmiştir (Uraki ve ark., 2014). İnsan β -defensin-2 ve insan serum albümini içeren rekombinant özelleştirilmiş defensin (DF-HSA), K-Ras mutasyonuna sahip MIA PaCa-2 hücrelerinin çoğalmasını ve pankreas karsinomu xenograftının büyümesini baskılamada etkili olmuştur (Du Y. ve ark., 2016). Laktöferrisin B'nin, kanser hücrelerinin sitoplazmik ve mitokondriyal zarlarına yönelik ölümcül ve seçici destabilize edici etkileri, kolorektal, nöroblastoma ve melanom kanser hücreleri üzerindeki etkinliği ile kanıtlanmıştır. Cecropin A ve B, mesane neoplazmlarında seçici sitotoksik ve sitostatik etkiler sergilerken, insan fibroblast hücre hatlarına karşı etkisiz olduğu gözlemlenmiştir. Cecropin A'nın, HL-60 promiyelosit hücre hattında ROS homeostazı aracılığıyla apoptoza yol açtığı bulunmuştur. Ayrıca, insan hepatoselüler karsinom hücrelerinde Fas, Fas-L, kaspaz-3 ve -8 ekspresyonu aracılığıyla etkisini gösterdiği belirtilmiştir (Roudi ve ark., 2017).

Antikanser etkili peptitler, kanser hücrelerinin proliferasyonunu, metastazını ve invazyonunu çeşitli mekanizmalar aracılığıyla inhibe eder. Ayrıca, hücre hareketliliğini ve metastatik potansiyeli azaltarak tümör hücrelerinin yayılmasını engellerler. Hücre zarı destabilizasyonu ve iyon kanal fonksiyonlarının bozulması da bu peptitlerin temel inhibisyon mekanizmalarındandır (Tornesello ve ark., 2020).

İmmün Yanıtın Modülasyonu

Konak savunma peptitleri olarak da bilinen AMP'ler doğrudan antimikrobiyal etkilerinin yanı sıra, bağışıklık tepkisini modüle etmedeki rolleri nedeniyle büyük ilgi görmüştür. Çeşitli yaşam formlarında yaygın olarak bulunan bu küçük, katyonik moleküller, patojenlerle doğrudan mücadele edebilir ve aynı zamanda bağışıklık sisteminin düzenlenmesine çok yönlü bir katılım göstererek hem doğal hem de adaptif bağışıklık sistemine etkiyebilir (Guryanova ve Ovchinnikova, 2022).

AMP'ler immün modülasyonda temel olarak, immünomodülatör ajanlar olarak hareket ederler. Dentritik hücreler, monositler ve T hücreleri dahil olmak üzere çeşitli bağışıklık hücrelerinin aktive edilmesini ve yönlendirilmesini sağlar. Katelisidinler ve defensinler gibi AMP'ler bağışıklık hücrelerini enfeksiyon bölgelerine çekebilir ve böylece patojenlere karşı hızlı yanıtı kolaylaştırabilir (Duarte-Mata ve Salinas-Carmona, 2023).

AMP'ler, enflamasyonu hem indükleyerek hem de inhibe ederek bağışıklık reaksiyonlarını düzenler ve böylece enflamatuvar dengeyi korur. Aşırı veya yetersiz enflamatuvar tepkiler, sırasıyla doku hasarı veya enfeksiyon kontrolünde başarısızlığa neden olabilir. AMP'ler, sitokin ve kemokin salınımını tetikleyerek ve T hücrelerinin farklılaşmasını ve işlevini modüle ederek adaptif bağışıklık sistemini etkileyebilir. Ayrıca, doğuştan gelen bağışıklıkta, özellikle α -defensinler ve katelisidinler gibi peptitler, kemokin benzeri işlevler görerek farklı bağışıklık hücrelerini enfeksiyon bölgelerine çeker. Sentetik AMP'ler de benzer kemotaktik etkiler sergileyebilir. Örneğin, farelerde yapılan bir çalışmada HNP1-3 defensinlerinin spesifik IgG ve IgM üretimini artırdığı, ancak IgA üzerinde etkili olmadığı bulunmuştur, bu da defensinlerin sistemik ama mukozal bağışıklık yanıtını güçlendirmede farklılaştığını gösterir (Duarte-Mata ve Salinas-Carmona, 2023).

AMP'ler yara iyileşmesi ve doku onarımında da rol oynamaktadır. Epitelleşmeyi, anjiyogenezi ve doku onarımı için gerekli hücrelerin toplanmasını teşvik ederler. AMP'lerin bu yönü, bariyer bütünlüğünün ve işlevinin korunmasına katkıda bulundukları deri ve mukozal bariyerlerde spesifik olarak önem taşımaktadır. Ayrıca, AMP'lerin toll benzeri reseptörlerin (TLR) inhibisyonu ve/veya enflamatuvar sitokinlerin sitokin aracılı sentezi ve anti-endotoksin aktivitesi gibi diğer

immünomodulator işlevler ile sepsis gibi hiperbolik ve zararlı pro-enflamatuvar yanıtları önler (Duarte-Mata ve Salinas-Carmona, 2023).

2.3.5. Antimikrobiyal ve Antikanser Hedefler için İkili Aktivite Gösteren Peptitler

Tümör hücreleri genellikle sağlıklı hücrelerden daha hızlı büyüdüğünden, geleneksel kemoterapi tedavileri hızla bölünen hücrelere odaklanır. Spesifik olmayan bu yaklaşım, hedeflenen kanser hücrelerinin yanı sıra, hızlı bölünen sağlıklı vücut hücrelerinin etkilenmesi ve sonuç olarak saç dökülmesi, mukozit, trombositopeni ve miyelosüpresyon gibi yan etkilere neden olmaktadır (Gaspar ve ark., 2013).

Terapötik peptitler, kanser tedavisinde son zamanlarda oldukça kabul görmektedirler. Bir çalışmada antikanser etkili peptitleri üç genel grupta tanımlanmışlardır: (A) Tüm canlılar tarafından doğal olarak üretilen, aynı zamanda kanser karşıtı etki gösteren antimikrobiyal/gözenek oluşturan peptitler, (B) Hücre geçirgen peptitler ve (C) Tümör hedefleyen peptitler (Boohaker ve ark., 2012). Terapötik peptitlerin birçok önemli avantajı bulunmaktadır; bunlar arasında küçük boyutları, yüksek aktiviteleri, özgüllükleri ve afiniteleri, az ilaç-etkileşimi, membranlardan geçme yetenekleri ve bu peptitlerin böbrek ve karaciğer gibi hayati organlarda birikim göstermeyerek düşük toksisiteye sahip olmaları sayılabilir (Parchebafi ve ark., 2022). AMP'ler ve ACP'ler, sağlıklı hücrelere zarar vermeden kanser hücrelerini hedef alarak tedavi edebilir. Kimyasal modifikasyonlar ve nanoparçacık teknolojisi, peptitlerin etkinliğini ve hedef odaklılığını artırır. ACP'ler, kanser hücre zarlarının negatif yükünü hedef alarak daha spesifik bir tedavi sağlarken, sağlıklı hücrelerin daha rijit membranları bu peptitlerin girişini engeller. Ancak, bazı peptitler sağlıklı hücrelerde de toksik olabilir, bu yüzden tedavi geliştirilirken bu etkiler dikkate alınmalıdır (Haney ve ark., 2017).

2.3.6. Peptitlerin Fizikokimyasal Özellikleri

Biyoyararlanım ve yarı ömür, peptitlerin klinik etkinliği için önemlidir. Boyut, çözünürlük, enzimatik bozulmaya duyarlılık ve atılım yöntemleri gibi faktörler etkiler. Bu fizikokimyasal özelliklerin manipüle edilmesi, peptitlerin terapötik potansiyelini optimize etmek için gereklidir (Pauletti ve ark., 1997). Peptitler, çeşitli fizikokimyasal özellikler sergilemektedir. Molekül ağırlıkları ve boyutları, biyolojik ve terapötik

işlevlerini etkiler; daha küçük peptitler biyolojik membranlara kolayca nüfuz edebilir ve daha iyi çözünmektedir (Torrent ve ark., 2011).

Moleküler yük/ağırlık, partiyon katsayısı (logP) ve hidrofobiklik, peptitlerin biyolojik işlevlerini ve terapötik potansiyellerini belirler. Molekül ağırlığı, hücrel etkileşim, vücut içi dağılım ve atılımı etkiler. Terapötik uygulamalarda genellikle 5000 Da'dan küçük ve 500 Da'dan büyük peptitler tercih edilir. Küçük peptitlerin renal eliminasyonu daha yüksektir. Moleküler yük, çözünürlük ve diğer moleküllerle etkileşimi etkiler. LogP, bir peptidin hidrofobikliğini gösterir ve ilaç tasarımında önemlidir. Yüksek logP, hücre zarlarını geçme yeteneğini artırırken, düşük logP, suda çözünürlüğü artırır ve intravenöz uygulamalar için tercih edilir (Frøkjaer ve ark., 2004). Peptit yapıları ilaçlardan örnek olarak semaglutide (logP: -18), leuprolide (logP: -2,4), teicoplanin (logP: -2,4) ve somatostatin (logP: -8,2) gösterilebilir (Wishart ve ark., 2018).

Amino asit bileşimi ve pH, peptitlerin yükünü etkiler. İzoelektrik nokta, sadece çözünürlüğü değil, aynı zamanda biyomoleküllerle etkileşimleri de etkiler. Peptitlerin ikincil yapıları, hidrojen bağı etkileşimlerine bağlı olarak α -helikal, β -yaprak veya rastgele sarmal olabilir ve bu yapılar çevresel koşullara yanıt olarak değişebilir (Torrent ve ark., 2011).

Lipinski'nin beş kuralı, küçük molekülü ilaçların oral biyoyararlanımını değerlendirmek için kullanılır ve beşten fazla hidrojen bağı vericisi, ondan fazla hidrojen bağı alıcısı, molekül ağırlığının 500 Da altında ve logP'nin 5 altında olmasını önerir. Bu kural, küçük ve orta derecede lipofilik moleküllerin oral olarak etkili ilaçlar olduğu gözlemine dayanır (Lipinski ve ark., 2001). Peptitler, bu kuralın eşiklerine tam olarak uymasa da altta yatan ilkeler geçerlidir ve hidrofobiklik, molekül ağırlığı gibi özelliklerin değerlendirilmesi membran geçirgenliği ve potansiyel oral biyoyararlanım hakkında bilgi sağlar (Lipinski ve ark., 2001).

2.3.7. Antimikrobiyal Peptitlerin Terapötik Potansiyelleri ve Kısıtları

AMP'ler antiviral, antifungal, anti-mitojenik, antikanser ve anti-enflamatuvar özellikleri içeren geniş bir biyolojik aktivite spektrumu göstermektedir. Bu muhtemel terapötik potansiyellerin yanı sıra, AMP'ler gıda koruyucuları ve bağışıklık modülatörleri olarak

hareket edebilir. Bu nedenle, AMP'ler geleneksel ilaçların yerini alma potansiyeline sahiptir ve önemli bir küresel ilaç pazar payı kazanabilir. Birçok AMP *in-vitro* veya *in-vivo* olarak terapötik potansiyel göstermiş olsa da çoğu durumda çeşitli sorunlar nedeniyle klinik deneylerde başarısız olmuşlardır (Parchebafi ve ark., 2022).

Tedavide Kullanılmalarını Kısıtlayan Faktörler

AMP'lerin klinik çalışmalarının çoğu topikal uygulamalarla sınırlıdır, zira sistemik toksisite riski, proteazlar tarafından kolayca bozulma ve yüksek böbrek klirensi gibi faktörler bulunmaktadır (Haney ve Hancock, 2013). Doğal AMP'lerin üretimi ve saflaştırılması zorluğu, sentetik AMP'lerin terapötik uygulamalarındaki zorlukları artırmaktadır. Yüksek üretim maliyetleri, potansiyel toksisite, seçicilik eksikliği ve düşük biyoyararlanım gibi faktörlerin aşılması gerekmektedir (Tornesello ve ark., 2020).

Peptitlerin proteolitik bozunmaya duyarlılığını azaltmak için D-enantiyomerleri kullanılarak kiralite değiştirilebilir veya konformasyonel esnekliği azaltılmış döngüsel peptitler kullanılabilir. Antitümör ve antimikrobiyal özelliklere sahip α -helikal bir peptit olan Polybia-CP'nin proteazlara karşı direnci, D-lizin ile değiştirilerek artırılmıştır. Ayrıca, peptitlerin proteaz bozunmasından korunmasına yardımcı olan C-terminus amidasyonu gibi stratejiler de kullanılmaktadır (Deslouches ve Di, 2017).

Sistemik Uygulama için Formülasyon Geliştirme Stratejileri

AMP'lerin umut verici *in vitro* sonuçlardan kliniğe uyarlanabilen terapötiklere dönüştürülmesi, özellikle sistemik uygulamalar düşünüldüğünde önemli zorluklarla karşılaşmıştır. Yaşamı tehdit eden, acil ve etkili tedavi çözümleri gerektiren sistemik enfeksiyonlarla mücadelede, sistemik uygulanabilirliğin avantajları yadsınamaz. Sistemik uygulamada temel esas, hedef için spesifik etkini sağlarken, konakçı için olası toksik etki riskinin minimuma indirilmesidir (Marr ve ark., 2006). AMP'lerin proteolitik bozunma ve sistemik toksisite gibi sorunlarına çözüm olarak D-amino asit kullanımı ve peptidomimetikler gibi daha stabil yapılar araştırılmaktadır. Ayrıca, AMP'lerin sitotoksik etkilerini sınırlamak için hedeflenen dokularda birikim sağlayan lipozomlar ve biyolojik olarak parçalanabilen nanopartiküller gibi taşıyıcı sistemler kullanılarak enkapsülasyon yöntemleri geliştirilmektedir. Bu yaklaşımlar, AMP'lerin etkinliğini artırıp, yan etkilerini azaltmayı amaçlar. (Marr ve ark., 2006).

AMP'ler ve Gen Tedavisi

AMP'ler ve gen tedavisi birlikte kanser tedavisinde yeni ve umut verici bir yaklaşımı temsil etmektedir. Gen tedavisi, hastalığı tedavi etmek için bireyin hücrelerine genetik materyalin aktarılmasını, değiştirilmesini veya çıkarılmasını içerir. Kanserde gen tedavisi amacıyla, terapötik AMP'leri kodlayan genetik materyal doğrudan tümör hücrelerine aktarılabilir. Bu yaklaşım AMP'lerin lokal olarak ekspresyonunu sağlayarak tümör bölgesindeki konsantrasyonlarını artırır ve potansiyel sistemik toksisiteyi azaltır (Ramos-Espinosa ve ark., 2016).

AMP'lerin gen tedavisiyle kombinasyonu kanser tedavisinde, tümör hücrelerine özgü hedefleme ile sağlıklı dokuların korunmasını, gen tedavisi vektörlerinin kanser hücrelerinde spesifik AMP ekspresyonunu teşvik etme yeteneği ile ve MDR sorununun aşılmasını sağlayarak avantajlar sunar. Bu yaklaşım, AMP'lerin farklı bir mekanizma kullanarak dirençli kanser hücrelerini öldürme potansiyeline dayanır. (Huang G. T. ve ark., 2002).

Sonuç olarak, AMP'ler ve gen terapisinin entegrasyonu, kanser tedavisinde özgüllük, ilaç direncinin üstesinden gelme potansiyeli ve azaltılmış yan etkiler sunan umut verici yeni bir yaklaşım sunmaktadır. Bu alandaki araştırmalar ilerledikçe, daha etkili ve kişiselleştirilmiş kanser tedavi stratejileri vadetmektedir (Huang G. T. ve ark., 2002; Ramos-Espinosa ve ark., 2016).

2.4. Peptit Veri Tabanları ve Kullanımları

AMP veri tabanlarının geliştirilmesi ve kullanımı, tıbbi ve biyolojik araştırma alanında giderek daha fazla önem kazanmaktadır. Bu veri tabanları; bakterilere, virüslere, mantarlara ve hatta bazı kanser türlerine karşı geniş bir yelpazede mücadele edebilen, doğal veya sentetik AMP'ler hakkında kapsamlı bilgi sağlayan koleksiyonlardır (Hancock ve Sahl, 2006).

AMP veri tabanları, AMP'lerin birincil dizileri (amino asit sırası), üç boyutlu yapıları ve spesifik patojen türlerine karşı aktiviteleri gibi çeşitli bilgiler sunar. Bu bilgiler, bu peptitlerin biyolojik aktivitelerini ve etki mekanizmalarını anlamak için kritik öneme sahiptir. Ayrıca, bazı veri tabanları AMP'lerin terapötik ajanlar olarak potansiyelini değerlendirmek için toksisite ve etkinlik verilerini de içerir (Aruleba ve ark., 2021).

Kaynaklar, arařtırmacıların yeni terapötikler için potansiyel adayları belirlemelerine yardımcı olmak, özellikle artan antibiyotik direnci karşısında geniş bir yelpazede patojenlere etkili şekilde mücadele edebilecek ilaçlar tasarlamak için bir başlangıç noktası olarak hizmet eder (Aruleba ve ark., 2021).

2.4.1. Peptit Veri Tabanlarında Arama ve Filtreleme Seçenekleri

Antimikrobiyal veri tabanlarında arama ve filtreleme, çeşitli parametreler kullanılarak gerçekleştirilir. Çeşitli parametreler, arařtırmacılara spesifik bilgilere hızlı ve doğru bir şekilde ulaşma imkanı sunar (Pirtskhalava ve ark., 2021; Wang G., 2023):

- **Kaynak:** Antimikrobiyal peptitlerin kökenini belirtir. Örneğin, hayvansal, bitkisel, mikrobiyal veya sentetik kaynaklar seçilebilir.
- **Peptit Sınıfı:** Peptitlerin yapısal sınıflandırmasını içerir. Örneğin, α -helikal, β -yaprak, döngüsel gibi yapısal özelliklere göre sınıflandırma yapılabilir.
- **Etki Mekanizması:** Peptitlerin hücre zarına etki edip etmediği veya hücre içi hedeflere müdahale edip etmediği gibi bilgileri içerir.
- **Molekül Ağırlığı ve Amino Asit Kompozisyonu:** Peptitlerin molekül ağırlıkları ve amino asit dizilimleri gibi yapısal özellikleri.
- **Biyoaktivite Parametreleri:** Peptitlerin minimum inhibitör konsantrasyonu (MIC) gibi biyoaktivite göstergelerini içerir.
- **Kimyasal Modifikasyonlar:** Peptitlerin modifiye edilip edilmediğini ve yapılan modifikasyon türlerini gösterir.
- **İmmünojenik Potansiyel:** Peptitlerin immün sistem üzerindeki potansiyel etkileri hakkında bilgi verir.
- **Sentez Türü:** Peptitlerin üretim yöntemlerini (nonribozomal, ribozomal, sentetik) filtrelemek için kullanılır. Bu, sentez yöntemlerinin biyolojik özellikler üzerindeki etkilerini incelemek isteyen arařtırmacılar için yararlıdır.

- Kimyasal Yapı: Peptitlerin N ve C uçlarındaki modifikasyonları incelemek için kullanılır. Bu, peptitlerin işlevselliği ve etkinliği hakkında daha derin anlayış sağlar.
- Hedefler (Grup, Nesne, Tür): Bu filtre, peptitlerin hedef aldığı biyolojik yapıları veya türleri belirtir. Araştırmacılar, peptitlerin belirli hücre yapılarına, organlara veya türlere yönelik etkilerini incelemek için bu filtreyi kullanabilirler.

2.4.2. Önemli Peptit Veri Tabanları

Antimikrobiyal Peptit Veri Tabanı (APD)

APD, 2600'ü hayvan kaynaklı olarak elde edilmiş, 3569 antimikrobiyal peptit içeren, manuel olarak düzenlenmiş bir veri tabanıdır. 153 adet insan-konak savunma peptiti; 363 adet memeli kaynaklı peptit; 1117'si kurbağalardan ve 75'i semenderlerden olmak üzere 1196 adet amfibilerden elde edilen aktif peptit; 114 adet balık peptiti; 49 sürüngen peptiti; 45 kuş peptiti; toplamda 633 eklembacaklılardan elde edilen peptit; 52 adet yumuşakçadan ve 6 adet protozoa kaynaklı antimikrobiyal peptit içermektedir (Wang Guangshun ve ark., 2016).

Antimikrobiyal Peptit Koleksiyonu (CAMP)

CAMP, antimikrobiyal peptitlerle ilgili çalışmalarını hızlandırmak ve genişletmek amacıyla oluşturulan bir veri tabanıdır. Bu veri tabanı, antimikrobiyal peptit ailesine ait dizi kompozisyonlarını içerir ve yeni antimikrobiyal peptitlerin keşfi ve tasarımı için önemli bilgiler sunar. CAMP, 45 farklı aileyi temsil eden 1386 antimikrobiyal peptidin sekanslarını ve Gizli Markov Modellerini (HMM) içermektedir. Kullanıcılar için dizi, protein tanımı, erişim numaraları, aktivite, kaynak organizma, hedef organizmalar ve ilgili veri tabanlarına bağlantılar gibi çeşitli bilgilere erişim sağlar. Ayrıca, bu veri tabanı dizi hizalaması, desen oluşturma ve HMM tabanlı arama gibi analiz araçları sunarak araştırmacıların verileri daha ayrıntılı bir şekilde incelemelerine olanak tanır. CAMP, antimikrobiyal peptitlerle ilgili çalışmalarını desteklemek ve peptit tasarımını kolaylaştırmak için önemli bir kaynaktır (Waghu ve ark., 2016).

Antimikrobiyal Peptitlerin Veri Havuzu (DRAMP)

DRAMP, peptitlerle ilgili çeşitli bilgileri içeren ve manuel olarak düzenlenen bir açık erişimli veri tabanıdır. Bu veri tabanı peptitlerin dizileri, yapıları, aktiviteleri, fizikokimyasal özellikleri, patentleri, klinik denemeler ve referans bilgileri gibi çeşitli bilgileri barındırmaktadır (Kang ve ark., 2019).

DRAMP, toplamda 22499 kayıt içermektedir. Bu girişlerin 6105'i genel AMP'leri (doğal ve sentetik AMP'leri içerenler), 16110'u patentli AMP'leri ve 96'sı ilaç geliştirme aşamasında olan AMP'leri temsil etmektedir. Ayrıca AMP tasarımının kapsamını genişletme amacıyla, DRAMP, antibakteriyel aktiviteleri test edilmemiş olan 5909 aday AMP'yi içermektedir. Bu aday AMP'ler, ayrı bir bölüm olarak sunulur ve toplam 22259 girdi içerisine dahil edilmezler (Kang ve ark., 2019).

Antimikrobiyal Aktivite ve Peptit Yapısı Veri tabanı (DBAASP)

DBAASP, bilimsel topluluğa yüksek terapötik indeksli antimikrobiyal bileşikler tasarlamak için bilgi ve analitik kaynaklar sağlamak amacıyla oluşturulmuş manuel olarak düzenlenen bir veri tabanıdır. DBAASP, kullanıcılara belirli hedef türlerine karşı peptitlerin ayrıntılı yapısı (kimyasal, üç boyutlu) ve antimikrobiyal aktivitesi hakkında bilgi sunar. Veri tabanı, antimikrobiyal aktivite gösteren ribozomal, ribozomal olmayan ve sentetik peptitlerin bilgilerini içerir ve bu peptitleri monomerler, multimerler ve multi-peptitler olarak sınıflandırır. DBAASP, sinerjik aktivitelerle ilgili veriler sunar. Veri tabanındaki belirli bir peptidin belirli bir hedef suşuna karşı başka bir peptit veya antimikrobiyal ile birlikte kombinasyon aktivitesini içermektedir (Pirtskhalava ve ark., 2021).

Bir Diğer Antimikrobiyal Peptit Veri tabanı (YADAMP)

YADAMP, toplamda 2133 girişi içeren bir antimikrobiyal peptit veri tabanıdır. Bu veri tabanı, kapsamlı bir literatür araştırmasının sonucunda oluşturulmuştur. Diğer antimikrobiyal peptit veri tabanlarından en önemli farkı, en yaygın bakteri suşlarına karşı belirgin bir antimikrobiyal aktivitenin varlığının açıkça belirtilmesidir. Peptitlerin bilgilerine peptit adı, amino asit sayısı, net yük, dizi motifleri gibi kriterleri kullanarak kolayca erişilebilir. YADAMP, antimikrobiyal peptitler hakkındaki bilgilere göz atmak ve peptit yapı-fonksiyon analizleri için uygundur (Piotto ve ark., 2012).

Antimikrobiyal Peptitleri Baęlayan Bir Veri tabanı (LAMP)

LAMP, AMP'leri baęlayan bir veri tabanıdır ve yeni antimikrobiyal ajanlar olarak AMP'lerin keşfini ve tasarımını desteklemek amacıyla bir araç olarak hizmet verir. Mevcut LAMP sürümü, 3904 doğal AMP ve 1643 sentetik peptit olmak üzere toplam 5,547 girişı içerir. Bu veri tabanı, yalnızca anahtar kelimeleri kullanarak veya kombinasyonel koşullarla sorgulanabilir. Detaylı antimikrobiyal aktivite ve sitotoksosite verileri ile donatılan LAMP içinde uygulanan çapraz baęlantı ve en benzer AMP'ler işlevleri, AMP'ler hakkındaki anlayışımızı artırmaya yardımcı olacak ve yeni tıbbi uygulamalar için AMP'lerin gelişimini hızlandırabilir (Ye ve ark., 2020).

Bakteriyosinlere Özel Veri tabanı (BACTIBASE)

BACTIBASE, hem Gram-pozitif (206) hem de Gram-negatif bakteriler (19) tarafından üretilen 230 bakteriyosinin fizikokimyasal özelliklerini içeren bir veri tabanıdır. Bu veri tabanındaki bilgiler oldukça kolay çıkarılabilir ve bu peptitlerin yapı/fonksiyon ilişkileri ve hedef organizmalarının hızlı tahminini mümkün kılar, bu da tıp ve gıda sektörlerinde biyolojik aktivitelerinin daha etkin kullanımını sağlar (Hammami ve ark., 2007).

Bu tür bir veri tabanında sunulan mikrobiyal fizikokimyasal ve yapısal özellikler, antimikrobiyal peptitlerin bu özel grubunun daha iyi ve kapsamlı bir yapısal ve fonksiyonel analizine olanak tanır. Bu, gıda koruma veya gıda güvenliği uygulamalarında kesinlikle kullanışlı olacaktır, ancak aynı zamanda tıbbi kullanım için yeni ilaçların geliştirilmesi açısından da önem taşır (Hammami ve ark., 2007).

Web Tabanlı Bakteriyosin Genom Madencilięi Aracı (BAGEL)

BAGEL web sunucusu, bakteriyel DNA'da modifiye edilmemiş bakteriyosinlerin ve ribozomal olarak sentezlenmiş ve translasyon sonrası modifiye edilmiş peptitlerin (RiPP'ler) üretiminde yer alan gen kümelerinin tanımlanmasını ve görselleştirilmesini kolaylaştırır. Ayrıca, bakteriyosinler için veri tabanı sağlar. Yeni antibiyotiklere duyulan ihtiyaç ve bunların gıda koruma, mikrobiyal ekoloji ve bitki biyokontrolündeki önemi, bu bileşiklere olan ilgiyi artırmaktadır. Genetik olarak kodlanan yirmi amino asidin ötesinde, RiPP'lerin doğal yapısal çeşitlilięi post translasyonel deęişikliklerle artırılmaktadır (van Heel ve ark., 2018).

Döngüsel Protein Dizileri ve Yapıları İçin Veritabanı (CyBase)

CyBase, N ve C terminalleri geleneksel bir amid bağı ile birleştirilmiş bir döngüsel yapıya sahip yeni bir protein sınıfının incelenmesine adanmış bir veri tabanıdır (Wang C. K. ve ark., 2008).

Yapılan çalışmalar, döngüsel proteinleri daha fazla karakterize etmeye, ticari ve tıbbi kullanım için uyarlamaya yöneliktir. Özellikle, siklotidler adı verilen bir döngüsel protein sınıfı üzerinde çalışıyoruz. Bu proteinler, Rubiaceae ve Violaceae bitkilerinde bulunur ve özel hedeflerimiz arasında, siklotidlerin bitkilerde oynadığı rolü belirlemek, siklotidler tarafından sergilenen geniş biyolojik aktivitelerin mekanizmasını keşfetmek (anti-HIV, antibakteriyel ve böcek öldürücü aktivite dahil), siklotidlerin genetik özelliklerini karakterize etmek ve yenilerinin keşfini yapmak yer almaktadır (Wang C. K. ve ark., 2008).

Antimikrobiyal Bitki Peptitlerine Özel Veri Tabanı (PhytAMP)

Antimikrobiyal bitki peptitleri, patojenik saldırılara karşı biyokimyasal bir seçenek sunan, doğuştan gelen bitki bağışıklığının esansiyel bileşenlerindendir. Tipik olarak kısa ve sistein bakımından zengin olan bu peptitler, alfa-defensinler, tiyoinler, lipit transfer proteinleri, snakinler ve hevein benzeri peptitler dahil olmak üzere çeşitli ailelerdeki dizi homolojisine göre sınıflandırılır. Kapsamlı bir veri tabanı olan PhytAMP, ayrıntılı taksonomik, mikrobiyolojik ve fizikokimyasal veriler için erişilebilir bir platform sağlar. Verilerin değerlendirilmesi, AMP'lerin antibiyotik direncine karşı koyma ve genetik mühendisliği yoluyla bitki patojen direncini artırma potansiyelleri noktasında önem taşımaktadır. PhytAMP'in ücretsiz erişilebilirliği, farmasötik geliştirme ve tarımsal iyileştirme alanlarında bitki peptit biyolojik aktivitelerinden yararlanmak için bir araç olarak faydasını artırmaktadır (Hammami ve ark., 2009).

Prokaryotik ve Ökaryotik Penetratin Veri Tabanı (PenBase)

PenBase, hücre içine nüfuz edebilen peptitler olarak bilinen özel bir sınıf olan penetratinlere ilişkin olarak tasarlanmış kapsamlı bir kaynaktır. Bu peptitler, hücre zarlarını geçme yetenekleriyle tanınır ve bu özellikleriyle moleküler biyoloji alanında ve potansiyel ilaç taşıma sistemlerinde değerli araçlar haline gelmişlerdir. PenBase, hem prokaryotik hem de ökaryotik penetratinler hakkında, onların dizileri, yapıları,

kaynakları ve biyolojik aktiviteleri gibi çeşitli yönlerini kapsayan geniş bir bilgi birikimi sunmayı amaçlamaktadır (Gueguen ve ark., 2006).

Defensin Bilgi Bankası

Defensin bilgi bankası, çeşitli türlerde doğuştan gelen bağışıklık tepkisinde kritik bir bileşen olan antimikrobiyal peptitlerin defensin ailesine adanmış küratörlü bir veri tabanıdır. Antimikrobiyal özellikleriyle bilinen defensinler, bakteriler, mantarlar ve virüsler de dahil olmak üzere çok çeşitli patojenlere karşı savunmada önemli bir rol oynamaktadır (Seebah ve ark., 2007).

Defensin bilgi bankası, alfa, beta ve teta defensinler gibi çeşitli defensin türlerinin genetik, biyokimyasal ve yapısal özelliklerini içeren kapsamlı veriler sunar. Bu bilgiler, doğuştan gelen bağışıklık ve antimikrobiyal mekanizmaların yanı sıra yeni terapötik ajanların geliştirilmesi alanında çalışan araştırmacılar için değerlidir. (Seebah ve ark., 2007).

PepBank

Gelişmiş bir peptit deposu olan PepBank, mevcut halka açık peptit veri tabanlarının entegrasyonu ile en son sekans metin madenciliği metodolojilerinin birleşimi niteliğindedir. Bu birleşme, bir dizi ilişkili verinin yanı sıra sağlam ve kapsamlı bir peptit dizileri özeti sunar (Shtatland ve ark., 2007).

PepBank, önceden var olan peptit veri tabanlarından gelen verileri birleştirerek kendini farklılaştırmaktadır. Bu entegrasyon, yapısal detayları, fonksiyonel özellikleri, sentetik metodolojileri ve potansiyel terapötik uygulamaları kapsayan peptitle ilgili çeşitli bilgilere sorunsuz erişimi kolaylaştıran bütünsel ve merkezi bir platform sunmada çok önemlidir. Bu tür bir birleştirme, peptit alanındaki araştırmacılar için paha biçilmez bir kaynak görevi görmektedir (Shtatland ve ark., 2007).

Deneyisel Olarak Doğrulanmış HIV İnhibe Edici Peptit Veri Tabanı (HIPdb)

HIPdb, HIV'yi inhibe eden deneyisel olarak doğrulanmış peptitlere odaklanan bir veri tabanıdır ve HIV araştırmalarında önemli bir kaynaktır. Bu veri tabanı, çeşitli deneyisel çalışmalardan toplanan verileri bir araya getirerek, HIV'e karşı etkinlik gösteren

peptitleri içerir. Araştırmacılara güvenilir, bilimsel olarak temellendirilmiş bilgiler sunan HIPdb, peptitlerin dizileri, yapısal özellikleri, HIV'e karşı etki mekanizmaları ve deneysel koşulları hakkında kapsamlı bilgiler içerir. Bu detaylar, yeni terapötik stratejilerin geliştirilmesi için değerli bilgiler sağlayarak, HIV'le etkileşimleri ve inhibisyon mekanizmalarını anlamaya yardımcı olur (Qureshi ve ark., 2013).

Antikanser Peptit ve Protein Veri Tabanı (CancerPPD)

CancerPPD, deneysel olarak doğrulanmış ACP'ler ve proteinlerle ilgili detaylı bilgiler sunan, benzersiz bir kaynaktır. Veriler hem yayımlanan makalelerden, patentlerden hem de diğer veri depolarından manuel olarak derlenmiştir. Antikanser aktivitesinde yapıların önemli bir rol oynadığı göz önüne alındığında, antikanser peptitlerin üçüncül yapılarını PEPstr adlı son teknoloji bir yöntem kullanarak tahmin ettik ve ikincil yapı durumları DSSP kullanılarak atandı. CancerPPD'nin önemli bir özelliği, aynı zamanda dış doğal, D-amin asitler, ornitin gibi değiştirilmiş amino asitler gibi çeşitli kimyasal değişikliklerle ilgili bilgi sağlamasıdır. Bu veri tabanı, ACP'lerle ilgili kapsamlı bilgi sağlamak amacıyla çeşitli diğer ilgili kaynaklarla çapraz bağlantılıdır (Tyagi ve ark., 2015).

TumorHoPe

TumorHoPe, 744 giriş içeren ve deneysel olarak karakterize edilmiş tümör hücrelerine yönelik peptitlerin kapsamlı bir veri tabanıdır. Bu peptitler, tümör dokularını ve tümörle ilişkili mikro çevreyi, tümör metastazını da içerecek şekilde tanır. Bu sayede, ilaçlar tümörlere seçici bir şekilde iletilir (Kapoor ve ark., 2012)

Hücreye Nüfuz eden Peptitler (CPPsite)

CPPsite yaklaşık 1700 benzersiz hücreye nüfuz eden peptit (CPP) içerir. Bu veri tabanı, deneysel olarak doğrulanmış CPP'leri (lineer & döngüsel; modifiye & doğal olmayan kalıntılar) ve bu peptitler kullanılarak hücrelere teslim edilen geniş bir yelpazede yükler (nükleik asitler, proteinler, nanopartiküller) bakımından sürdürür (Gautam ve ark., 2012; Agrawal Piyush ve ark., 2015).

Apoptoz İndükleyici Antikanser Peptitler Veri Tabanı (ApInAPDB)

ApInAPDB, apoptoz indükleyen anti-kanser peptitlerin bir veri tabanıdır. Bu veri tabanı, 818 adet apoptoz indükleyen anti-kanser peptiti içerir ve bu peptitler araştırma makalelerinden manuel olarak toplanmıştır. Veri tabanı, peptitlerle ilgili işlev, bağlanma hedefi ve afinite, IC50 ve benzeri bilgileri sağlar. Ayrıca, GRAVY (hidropati büyük ortalaması), pH 7'deki net yük, hidrofobisite ve diğer fizikokimyasal özellikler hesaplanmış ve sunulmuştur (Faraji ve ark., 2022).

2.5. Peptit Araştırmalarında In Silico Yöntemler

Her alanda olduğu gibi *in-silico* yöntemler, peptitler ile ilgili araştırmalarda da sıklıkla tercih edilmektedir. *In-silico* yöntemler ile peptitlerin biyolojik aktiviteleri, etki mekanizmaları ve fizikokimyasal özellikleri hakkında değerli bilgiler elde edilebilir (Zhou P. ve Huang, 2015).

Bu yaklaşımlar, peptitlerin genomlardan tespit edilmesi, antimikrobiyal aktivitelerinin tahmin edilmesi ve hedeflenen patojenlere karşı etkinliklerinin belirlenmesi için kullanılabilir. Örneğin, ökaryotik genomlardan antimikrobiyal peptitlerin tahmin edilmesi ve yeni antibiyotiklerin geliştirilmesi mümkün hale gelmiştir. Ayrıca, bu yöntemler peptitlerin anti-biyofilm özelliklerini öngörmek için de kullanılmaktadır (Zhou P. ve Huang, 2015).

In-silico araştırmalar, peptitlerin yapısal özelliklerini ve etki mekanizmalarını anlamada da önemli bir rol oynamaktadır. Özellikle, peptitlerin hücre zarlarına etkilerini anlamak ve bunları modifiye etmek için biyofiziksel ölçümlerden yararlanılmaktadır. Bu sayede, peptitlerin hedeflenen patojenlere karşı etkinliğini artırmak ve yan etkilerini azaltmak mümkün olmaktadır (Zhou P. ve Huang, 2015).

2.5.1. Literatürdeki Biyoinformatik Araçlar

Antikanser Peptitlerin Tahmini ve Tasarlanması (AntiCP)

AntiCP 2.0, yüksek doğrulukta anti-kanser peptitleri tahmin etmek ve tasarlamak için geliştirilmiş bir makine öğrenmesi modelidir. Bu çalışma, geniş bir antikanser ve antikanser olmayan peptit veri setini kullanmaktadır. Ana veri seti, deney ile

doğrulanmış 861 anti-kanser peptiti ve 861 antikanser olmayan veya doğrulanmış antimikrobiyal peptitten oluşmaktadır. Sunucunun ana özellikleri arasında peptitlerin antikanser etkinliğe sahip olup olmadığını tahmin eden bir modül, bir peptidin tüm olası mutantlarını üreten ve her birinin anti-kanser potansiyelini tahmin eden bir tasarım modülü ve sorgu sekanslarındaki antikanser motiflerini tarayan ve eşleştiren bir motif tarama modülü bulunmaktadır. Veri setleri araştırmacılar tarafından indirilebilmektedir. Sunucu, ayrıca çeşitli fizikokimyasal özellikleri hesaplayabilmektedir (Agrawal Piyush ve ark., 2021).

Antimikrobiyal Peptit Tarayıcı (AMPA)

AMPA, yeni antimikrobiyal ilaçların tasarımına odaklanarak proteinlerin antimikrobiyal bölgelerini değerlendiren bir web uygulamasıdır. Uygulama, patojenlere karşı yeni peptit tabanlı ilaçlar geliştirmek için kullanılabilecek proteinlerdeki antimikrobiyal desenlerin hızlı keşfini sağlar. Sonuçlar, kullanıcı dostu bir grafik arayüzde gösterilir ve daha sonraki incelemeler için ham veri olarak indirilebilir (Torrent ve ark., 2012).

PeptideRanker

PeptideRanker, yapay sinir ağlarına dayalı olarak biyoaktif peptitlerin tahmin edilmesi için bir sunucudur. Geliştirilen antimikrobiyal etkinlik tahminleyicisinin, özellikle toksin ve zehir peptitleri gibi belirli sınıflara ait peptitleri tanımlama konusunda başarılı olduğu gözlemlenmiştir. Kısa peptitlere (4–20 amino asit) ve uzun peptitlere odaklanan iki model eğitilerek web sunucusuna entegre edilmiştir (Mooney ve ark., 2012).

Antikanser ve Antimikrobiyal Peptit Tasarımı için Araç (A-CaMP)

A-CaMP, veri tabanlarında yer alan mutasyona uğramış kanserli protein dizilerini kullanarak antikanser peptitleri tespit eden bir biyoinformatik araçtır. Bu araç, kullanıcı tarafından sağlanan hedef peptit veya protein sekanslarını veri tabanlarına karşı tarayarak ACP ve AMP'lerin hızlı bir şekilde belirlenmesine olanak tanır. PERL dilinde geliştirilmiş olan A-CaMP, tutarlı bir kodlama mimarisi ile donatılmıştır. Yapılan çalışmalar, A-CaMP'ın %93,4'lük bir tahmin doğruluğuna ulaştığını ve bu oranın mevcut araçlardan daha yüksek olduğunu göstermektedir. Ayrıca, dizi hizalama çalışmaları, AMP'lerin tanımlanmasında A-CaMP'den faydalanmaktadır. A-CaMP, peptitlere dair

güncel klinik verileri kullanan ve bu peptitlerle ilgili çeşitli bilgileri sağlayan açık kaynaklı bir araçtır (Kaushik ve ark., 2021).

CancerGram

CancerGram, antikanser peptitlerin tahmini için n-gram kodlama ve rastgele orman algoritması kullanan yenilikçi bir araçtır. Araştırmacılar açık kaynaklı koda erişerek kendi sistemlerinde çalıştırabilmektedirler. Aynı zamanda, kullanıcı dostu arayüze sahip web sunucusundan da erişim sağlanabilmektedir. Kullanılabilmesi için “CancerGramModel” adında, antikanser peptitlerin tahmini için gerekli olan harici bir paket yüklenmelidir. CancerGram, birden fazla peptiti aynı anda analiz edebilme yeteneğine sahiptir, ancak hesaplamalar veri seti büyüklüğüne göre uzun sürebilmektedir ve web sunucusu üzerinden aynı anda en fazla 50 sekans analiz edilebilir. Daha geniş sorgular için, CancerGram paketi R programlama dilinde çalıştırılmalıdır (Burdukiewicz ve ark., 2020).

Hücreye Nüfuz Eden Antikanser Peptit Tahmini (CpACpP)

CpACpP, hücreye nüfuz eden antikanser peptitlerin tahmin edilmesi için kullanılan yenilikçi bir biyoinformatik araçtır. Bu araç, rastgele orman (RF), destek vektör makinesi (SVM) ve XGBoost algoritmalarını kullanarak tasarlanmıştır. CpACpP, çeşitli yapısal ve fizikokimyasal temelli özellikleri, çok katmanlı yinelemeli özellik eleme yöntemi ile birleştiren ve seçen gelişmiş bir özellik mühendisliği içermektedir. Bir web sunucusu üzerinden geliştirilen araca erişilebilmektedir (Nasiri ve ark., 2021).

PepVAE

PepVAE, değişken otokodlayıcı ve antimikrobiyal aktivite tahmin modelleri üzerinde geliştirilmiş bir araçtır. Bu araç, sadece sekans ve MIC verilerini kullanarak antimikrobiyal peptitlerin tespit edilmesi ve tasarlanması için kullanılmaktadır. Bu alanda etkili bir çözüm sunan PepVAE’ye bir web sunucusu üzerinden erişim sağlanabilmektedir (Dean ve ark., 2021).

AntiMPmod

AntiMPmod, kimyasal olarak modifiye edilmiş peptitlerin antimikrobiyal özelliklerinin tahmin edilmesi için geliştirilmiş modelleri içeren bir web sunucusudur. Bu modeller,

948 pozitif ve 931 negatif örneği içeren, kimyasal olarak modifiye edilmiş ve doğal rezidüler barındıran bir veri seti üzerinde eğitilmiş, test edilmiş ve değerlendirilmiştir. Peptitlerin tersiyer özellikleri kullanılarak destek vektör makinesi algoritması ile geliştirilmiştir (Agrawal P. ve Raghava, 2018).

RandSeq, Deep-AmPEP30, iACP, DRACP

RandSeq, rastgele dizilmiş protein veya peptit sekansları oluşturmak için kullanılan bir araçtır. Bu rastgele diziler, protein veya peptit yapısının ve fonksiyonunun anlaşılmasında, biyoinformatik çalışmalarında kullanılabilir (Gasteiger ve ark., 2003).

Deep-AmPEP30, derin öğrenme yöntemlerini kullanarak kısa uzunluğa sahip AMP'lerin tahmin edilmesine odaklanan bir araştırma projesidir. Bu, potansiyel antimikrobiyal özelliklere sahip yeni peptitlerin keşfedilmesine yardımcı olabilir (Yan ve ark., 2020).

iACP, ilaç tasarımında kullanılmak üzere, peptitlerin antikanser aktivite tahminini gerçekleştiren bir araçtır. Bu proje, potansiyel ilaç adaylarını hızlı ve etkili bir şekilde belirlemeye yardımcı olabilir (Chen W. ve ark., 2016).

DRACP, benzer bir şekilde antikanser peptitlerin tespit edilmesi için çeşitli makine öğrenmesi algoritmaları üzerinde geliştirilmiş bir modeldir. Açık kaynaklı kod ile araştırmacılar yöntemi kendi sistemlerinde deneyebilmektedirler (Zhao ve ark., 2020).

2.5.2. Özellik Seçimi

Özellik seçimi, makine öğrenmesi ve in silico araçlarında kritik bir rol oynar. Bu süreç, veri setlerindeki özneliliklerden en bilgilendirici ve ilgili olanlarını belirlemeyi amaçlar. Özellik seçiminin temel amacı, modelin genel performansını iyileştirmek, aşırı uyumu önlemek ve hesaplama maliyetini azaltmaktır (Agrawal S. ve ark., 2017). Özellikle peptitlerin antikanser aktivitelerini tahmin etmek için kullanılan in silico modellerde, doğru özelliklerin seçimi, modelin doğruluğu ve güvenilirliği açısından hayati öneme sahiptir (Tang ve ark., 2016).

Özellik seçimi genellikle üç ana strateji altında incelenir: filtre, gömülü ve sarmalayıcı yöntemler. Filtre yöntemleri, özelliklerin istatistiksel ölçümlerine dayanarak bir ön

seim yapar. Bu yntemler, modelin kendisinden bağımsız olarak alışır ve genellikle hızlı ve etkili bir n seim saėlar. Gml yntemler, ėrenme algoritmasının bir parası olarak zellik seimini gerekleřtirir. Bu yaklařımlar, model eėitimi sırasında zelliklerin nemini deėerlendirir ve genellikle daha doėru sonular retir. Sarmalayıcı yntemler ise, farklı zellik kombinasyonlarını deneyerek en iyi model performansını saėlayan zellik setini seer. Bu hesaplama gereksinimleri aısından maliyetlidir (Agrawal S. ve ark., 2017).

Peptit dizilerinden elde edilen znitelikler, modelin eėitimi iin girdi olarak kullanılır. Bu znitelikler, amino asit kompozisyonu, dipeptit kompozisyonu, kimyasal ve fiziksel zellikler gibi eřitli biyolojik ve yapısal bilgileri ierebilir. Ancak, tm zniteliklerin modelin doėruluėuna katkısı eřit deėildir. Bu nedenle, zellik seimi sreci, modelin karmařıklıėını azaltmak ve tahmin bařarısını artırmak iin kritik neme sahiptir. Her bir yntemin avantajları ve dezavantajları vardır ve spesifik uygulama gereksinimlerine gre seim yapılmalıdır (Tang ve ark., 2016).

2.5.3. Makine ėrenmesi Algoritmaları

Makine ėrenmesi algoritmaları, saėlık bilimlerinde de byk veri setlerinden karmařık ve deėerli bilgileri ıkarmak iin kullanılmaktadır. Bunlar, veri madenciliėi ve istatistiksel analiz teknikleri kullanarak, biyolojik ve tıbbi verilerden anlamlı sonular elde etmeyi amalar. Hastalıkların teřhisinden tedavi yntemlerinin geliřtirilmesine kadar geniř bir yelpazede kullanılan bu algoritmalar; hastalık tahmini, risk deėerlendirmeleri ve tedavi sonularının iyileřtirilmesinde nemli roller oynar (Ghosh ve Dasgupta, 2022).

Bu algoritmalar ile geliřtirilen modeller, mevcut veriye dair iėr sunmak, yeni veri oluřturmak veya gelecekteki olayları tahmin etmek iin eřitli yntemler kullanır. ėrenme sreleri, genellikle denetimli, denetimsiz veya pekiřtirmeli olmak zere  ana kategoriye ayrılır. Denetimli ėrenmede, model, etiketlenmiř eėitim verilerinden ėrenir ve bu bilgileri yeni, grlmemiř verilere uygular. Denetimsiz ėrenme ise, etiketlenmemiř verileri kullanarak veri setlerindeki gizli yapıları ve kalıpları keřfeder. Pekiřtirmeli ėrenme ise, algoritmaların dl ve ceza mekanizmaları aracılıėıyla en iyi

kararları almayı öğrenmelerine dayanır. Bu algoritmalar, veri analizi, tahmin yapma ve karmaşık problem çözme konularında önemli araçlardır (Ghosh ve Dasgupta, 2022).

Aşağıda AMP'lerin etkinlik tahmini ile ilgili araştırmalarda sıklıkla tercih edilen dört makine öğrenmesi algoritması açıklanmaktadır.

Rastgele Orman (RF)

RF algoritması, makine öğrenmesinde popüler bir toplu öğrenme tekniğidir hem sınıflandırma hem de regresyon görevlerinde kullanılır. Birden fazla karar ağacının birleştirilmesiyle oluşturulan bu model, Breiman ve Cutler tarafından geliştirilmiştir. Bir tek karar ağacına göre daha tutarlı ve doğru tahminler sağlar (Breiman, 2001; Speiser ve ark., 2019).

Her bir karar ağacı, veri ve özelliklerin rastgele alt kümelerinden oluşturulur, bu yaklaşım modelin varyansını azaltır ve aşırı uyma karşı direnç sağlar. RF, her ağacı tam olarak büyütür ve tahminlerde çoğunluk oyu veya ortalama değeri kullanır, bu sayede birden fazla modelin birikimini etkili bir şekilde yakalar (Breiman, 2001).

RF modelleri çok yönlülükleri ve kullanım kolaylıkları ile öne çıkar. Sayısal ve kategorik verileri işleyebilir, özellik ölçeğine duyarlıdır ve özelliklerin önemini anlamada değerli içgörüler sunar. Ancak, özellikle büyük veri kümeleri ve yüksek sayıda ağaç ile karmaşık hesaplamalar yüksek kapasitede donanım gerektirebilir. Çoklu ağaçların karar verme süreci tek bir karar ağacına göre daha zor takip edilebilmektedir (Speiser ve ark., 2019).

Sonuç olarak, RF, makine öğrenmesindeki etkili ve çok yönlü tekniklerden biridir. Aşırı uyma karşı sağlamlığı ve yüksek doğrulukta tahminler üretme yeteneği, geniş bir uygulama alanı sağlamaktadır (Breiman, 2001; Speiser ve ark., 2019).

Destek Vektör Makineleri (SVM)

SVM, Vladimir Vapnik ve Alexey Chervonenkis tarafından geliştirilen, sınıflandırma ve regresyon görevleri için kullanılan popüler bir denetimli öğrenme modelidir. İkili sınıflandırmada, en uygun hiperdüzlemi bularak sınıflar arasındaki ayrım marjını en üst düzeye çıkarır. Bu hiperdüzleme en yakın olan destek vektörleri, modelin karar

vermesinde kritik rol oynar. Doğrusal olmayan durumlarda, SVM çekirdek işlevleri kullanarak verileri daha yüksek boyutlu bir özellik uzayı yaratır. Polinom, radyal-bazlı fonksiyon (RBF) ve sigmoid gibi yaygın çekirdekler, karmaşık ilişkileri yakalamak için tercih edilir. SVM'in güçlü yönleri arasında, yüksek boyutlu alanlarda tutarlılığı ve aşırı uyma karşı direnci yer alır (Cortes ve Vapnik, 1995).

SVM, görüntü tanıma ve metin sınıflandırmaya kadar çeşitli alanlarda uygulanmıştır. Özellikle örnek sayısının boyut sayısını aştığı durumlarda etkindir. Ancak, çekirdek seçimi ile parametre ayarları modelin performansını önemli ölçüde etkileyebilir ve bu yapılandırmaları belirlemek zor olabilir. Büyük veri setlerinde veya yüksek gürültü seviyelerinde performans sorunlarına neden olabilir. Özetle, SVM, sınıflar arasındaki ayrımı en üst düzeye çıkarmak ve çeşitli veri türlerinde başarı sağlamak için çok yönlü bir makine öğrenmesi algoritmasıdır (Cortes ve Vapnik, 1995).

Lojistik Regresyon (LR)

LR, ikili sınıflandırma problemlerinde yaygın olarak kullanılan bir denetimli öğrenme modelidir. David Cox tarafından 1958'de geliştirilen bu yöntem, bir sonucun olasılığını tahmin etmek için bağımsız değişkenlerin lineer kombinasyonunu kullanır. Bağımlı değişkenlerin iki olası değeri arasında seçim yapar, bu bir e-postanın spam olup olmadığı veya bir hastalığın varlığı olabilir (Stoltzfus, 2011; Msaouel ve ark., 2022).

Model, olasılıkları 0 ile 1 arasında bir değer alan logit fonksiyonu aracılığıyla hesaplar. Lojistik regresyon, bağımsız değişkenlerin etkisini ve bu değişkenlerin bağımlı değişken üzerindeki ilişkilerini anlamada değerli bilgiler sunar (Sperandei, 2014).

Tıbbi risk değerlendirmelerinden finansal modellere kadar geniş bir uygulama yelpazesine sahiptir. Basit yapısı ve yüksek yorumlanabilirliği ile popülerdir, ancak karmaşık ilişkileri modellemekte yetersiz kalabilir ve aşırı uyma karşı hassastır. (Sperandei, 2014).

K-En Yakın Komşular (k-NN)

K-NN, araştırmacılar tarafından sıkça kullanılan bir denetimli öğrenme yöntemidir. Basit, etkili bir sınıflandırma ve regresyon algoritması olan k-NN, bir veri noktasının

sınıfını veya değerini, en yakın komşularının sınıflarını veya değerlerini temel alarak tahmin eder. 1951'de Evelyn Fix ve Joseph Hodges tarafından geliştirilen bu yöntem, özellikle düşük boyutlu veri kümeleri üzerinde etkilidir (Uddin ve ark., 2022). K-NN algoritmasının temel prensibi, bir veri noktasının komşularına dayanarak bu noktanın sınıfını belirlemektir. "K" değeri, algoritmanın dikkate alacağı en yakın komşu sayısını belirtir. Regresyon görevlerinde ise, bu komşuların çıktı değerlerinin ortalaması alınır (Zhang Z., 2016).

K-NN'nin başarısı büyük ölçüde, veri noktaları arasındaki mesafeyi ölçmek için kullanılan metriğe ve K'nin seçimine bağlıdır. Mesafe ölçümü için genellikle Öklid, Manhattan veya Minkowski gibi metrikler kullanılır. K'nin doğru bir şekilde seçilmesi, algoritmanın performansı üzerinde önemli bir etkiye sahiptir; çok küçük bir K değeri modelin aşırı uymasına, çok büyük bir K değeri ise modelin aşırı genelleştirmesine yol açabilir (Feng ve ark., 2023).

Eğitim aşamasında, k-NN algoritması tüm eğitim veri setini referans olarak kaydeder. Ancak, bu modelin hesaplama maliyeti, veri seti büyüklüğü ile doğru orantılı olarak artar. Ayrıca, k-NN, veri özelliklerinin ölçeklenmesine duyarlıdır; verilerde normalleştirme veya standardizasyon gerektirebilir (Zhang Z., 2016).

Sonuç olarak, k-NN, özellikle küçük ve orta ölçekli veri setlerinde etkili olan, kolay anlaşılır bir makine öğrenmesi algoritmasıdır. Sınıflandırma ve regresyon görevlerinde geniş bir uygulama yelpazesi sunar, ancak büyük veri setlerindeki hesaplama maliyeti ve ölçeklendirme sorunları göz önünde bulundurulmalıdır (Zhang Z., 2016).

3. GEREÇ ve YÖNTEM

Çalışma ortamı, araştırmanın gereksinimleri doğrultusunda planlanmıştır. Araştırma yöntemleri ve araçları, sistemin temel donanımı ve kısıtları gözetilerek seçilmiştir. Araştırma kapsamında veri işleme ve makine öğrenimi modellemesi için donanım olarak yalnızca merkezi işlemci birimi (CPU) ve rasgele erişimli bellek (RAM) kullanılmıştır. Grafik işleme birimi (GPU) gibi özel donanım birimleri kullanılmamıştır.

CPU olarak, altı çekirdekli yapıya ve 3.00 GHz işlem hızına sahip bir birim (AMD Ryzen 5 4600HS) kullanılmıştır. Bu çalışmada, veri işleme görevleri için yeterli düzeyde performans sunmuştur. RAM olarak ise 16 GB kapasiteli bir birim (Micron Technology) kullanılmıştır. 3200 MHz hızına sahip bu bellek, işlemcinin veri işleme kapasitesini desteklemekte ve çalışma için yeterli bir bellek alanı sağlamaktadır.

Mevcut bellek ve işlemci kapasitesi, proteinlere ait büyük veri setlerinin işlenmesi sırasında bazı kısıtlamalara yol açabilir, ancak peptitlere ait veri setlerinin temel analizi ve işleme görevleri için uygun düzeydedir.

Bu çalışmada Python programlama dili, veri analizi ve işleme görevlerinde ana araç olarak kullanılmaktadır. Ayrıca, *Visual Basic for Applications* (VBA) da belirli görevler için kullanılmıştır.

3.1. İşletim Sistemi ve Yazılım Dağıtımı

3.1.1. İşletim Sistemi (OS)

Bu çalışma, 64-bit mimariye sahip Windows 11 Pro (OS Build 22621.2715) işletim sisteminde gerçekleştirilmiştir. Windows, günümüzde yaygın olarak kullanılan birçok uygulama ve araştırma aracı ile uyumludur. Bu işletim sistemi sürümü, araştırma ve veri analizi için gerekli olan temel yazılım ve uygulamaların sorunsuz bir şekilde çalıştırabilmiştir (Panay, 2021).

3.1.2. Yazılım Dağıtımı

Bu araştırma kapsamında, veri bilimi ve makine öğrenimi uygulamaları için temel bir yazılım dağıtım aracı olan Anaconda Navigator 2.5.0 kullanılmaktadır. Bu platform,

açık kaynak kodlu olmakla birlikte, Python ve R programlama dilleri için kapsamlı bir paket ve ortam yönetimi sağlar. Bilimsel hesaplama, veri analizi ve makine öğrenimi projeleri için gerekli olan birçok kütüphane ve aracı içermektedir ("Anaconda software distribution," 2023).

Anaconda'nın ortam kanalları arasında "conda-forge", "bioconda" ve "defaults" bulunmaktadır. "Conda-forge", geniş bir kullanıcı ve geliştirici topluluğu tarafından desteklenen standart bir yazılım paketi koleksiyonu sunar. "Bioconda" ise, biyolojik bilimler alanında kullanılan özelleşmiş araçlar ve kütüphaneler için bir kanaldır. "Defaults" kanalı ise Anaconda tarafından sağlanan temel paketleri ve yazılımları içerir. Bu kanallar, araştırma sürecinde ihtiyaç duyulan çeşitli araçların ve kütüphanelere etkili bir şekilde erişilmesini sağlamaktadır ("Anaconda software distribution," 2023).

3.1.3. Entegre Geliştirme Ortamı (IDE)

Bu çalışmada, Entegre Geliştirme Ortamı (IDE) olarak Jupyter Notebook versiyon 6.5.4 tercih edilmiştir. Jupyter Notebook, veri bilimi ve makine öğrenimi projeleri için popüler bir araçtır. Bu platform, Python başta olmak üzere birçok programlama dilini desteklemekte ve biyoinformatik alanındaki araştırmacılara uygun bir ortam sağlamaktadır ("Jupyter notebook," 2023).

Jupyter Notebook, kod bloklarını ve açıklamalarını bir arada sunabilen bir yapıya sahiptir. Bu özellik, hataları kolay tespit etmek ve sonuçları kolayca gözden geçirmek için idealdir. Ayrıca, araştırma sürecinde yapılan analizler, grafikler ve sonuçlar, tek bir dokümanda interaktif bir şekilde sunulabilir. Bu, veri analiz sürecinin şeffaflığını artırarak sonuçların sunumu ve paylaşımını kolaylaştırır ("Jupyter notebook," 2023).

Notebook ortamının esnek yapısı, farklı veri setleri ve kütüphanelerle de çalışmaya imkan sunmaktadır ("Jupyter notebook," 2023).

3.2. Programlama Dilleri

Sağlık bilimlerinde *in-silico* çalışmaların yaygınlaşmasıyla araştırmacılar birçok programlama dilini kullanmaya başlamışlardır. Bu alanda çözüm sunan lisanslı yazılımlara erişim, yüksek maliyetler nedeniyle kısıtlanmaktadır. Açık kaynaklı araçlar

da genellikle özgül sorulara yanıt verebilmekte ve farklı sorulara uyarlamaya imkân tanımamaktadır. Sonuç olarak, bu durum araştırmacıları sıklıkla kendi algoritmalarını ve protokollerini geliştirmeye yöneltmektedir (Terstappen ve Reggiani, 2001).

Bu araştırmada, Python programlama dili kullanılmıştır. Sağlık ve biyoloji bilimlerindeki araştırmacılar tarafından geliştirilen açık kaynaklı paket ve kütüphanelerin çeşitliliği bu tercihin temel sebebidir. Python, veri analizi, biyoistatistik ve makine öğrenimi uygulamaları için güçlü kütüphanelere sahip olmasıyla öne çıkar. Bu dilin kullanımı, biyolojik bilimlerdeki araştırmaların kapsamını genişletmekte ve veri tabanlı keşifleri kolaylaştırmaktadır (Bassi, 2017).

3.2.1. Python

Bu çalışma için Python (sürüm 3.11.5) programlama dili kullanılmıştır. Python, nesne yönelimli, yorumlanabilir ve yüksek seviyeli bir programlama dilidir. Dinamik tiplmesi ve otomatik bellek yönetimi özellikleriyle dikkat çeker. Python'un, biyoinformatik alanında veri manipülasyonu ve analiz süreçlerini kolayca yönetmeye olanak tanıyan yapısı, bu zenginlikte bir ekosistem oluşmasını sağlamıştır. Son olarak, Python'un Jupyter Notebook ile entegrasyonu, araştırma yöntemlerinin ve sonuçlarının daha anlaşılır ve erişilebilir olmasını sağlamaktadır (Bassi, 2017; Python, 2023).

3.2.2. Visual Basic for Applications (VBA)

Bu araştırmada, Excel'in 16.0.16924 sürümüyle entegre olarak VBA da kullanılmıştır. VBA, özellikle veri temizliği, basit analizler ve veri setlerinin ön hazırlığı gibi görevler için tercih edilmiştir. VBA'nın sunduğu makro yazma ve çalıştırma yeteneği, tekrar eden görevleri otomatikleştirmek ve veri setlerini hızlı bir şekilde düzenlemek için kullanılmıştır. VBA'nın Excel ile entegrasyonu, bu araştırmada veri setlerinin etkili bir şekilde işlenmesi ve analiz edilmesi için temel bir rol oynamıştır (Windows, 2023).

3.3. Python Kütüphaneleri ve Paketleri

Bu araştırma, çeşitli Python modülleri, kütüphaneleri ve paketleri kullanarak gerçekleştirilmiştir. Bunlar, veri işleme, analiz, görselleştirme ve makine öğrenimi gibi farklı ihtiyaçları karşılamak üzere seçilmiştir. Her biri, özgün işlevleri ile araştırmanın

verimliliğini ve etkinliğini artırmaktadır. Aşağıda, kullanılan modül, kütüphane ve paketlerin listesi ve her birinin kısa açıklamaları yer almaktadır:

- ast: soyut sözdizim ağaçları modülü, Python kodunu öğelerine ayırmak ve işlemek için kullanılır (Python, 2023).
- biopython (1.81): Biyoinformatik için tasarlanmış bir kütüphanedir, DNA, RNA ve protein sekans analizlerinde geniş kullanım alanları sunar (Cock ve ark., 2009).
- igraph (0.11.2): Ağ analizi ve görselleştirme için kullanılan bir kütüphanedir, karmaşık ağ yapılarını işlemek ve analiz etmek için uygundur (Csárdi ve ark., 2023).
- joblib (1.2.0): Büyük veri setlerinin etkili bir şekilde serileştirilmesi ve paralel işlem görmesi için kullanılır. Python'da makine öğrenimi modellerini kaydetmek ve yüklemek için kullanılabilir (Joblib, 2023).
- matplotlib (3.8.0): Veri görselleştirme için kullanılan bir kütüphanedir, çizimler ve grafikler oluşturmak için geniş imkanlar sunar (Hunter, 2007).
- more-itertools (8.12.0): İterasyonlar ve döngüler için ekstra işlevsellikler sağlayan bir kütüphane, veri işleme süreçlerinde kolaylık sağlar (Bayles ve Rose, 2021).
- nltk (3.8.1): Doğal Dil İşleme (NLP) görevleri için kullanılan kapsamlı bir kütüphane, metin analizi ve dil işleme işlevleri sunar (Loper ve Bird, 2002).
- numpy (1.24.3): Sayısal hesaplamalar için kullanılan temel bir kütüphanedir, dizi ve matris işlemlerinde yüksek performans sağlar (Harris ve ark., 2020).
- os: İşletim sistemi ile etkileşim ve dosya yönetimi işlemleri için kullanılan standart bir Python modülüdür (Python, 2023).
- pandas (2.1.1): Veri analizi ve manipülasyonu için kullanılan, DataFrame yapısı ile bilinen bir kütüphanedir (Team, 2023).

- pycairo (1.25.1): 2D grafikler ve görseller oluşturmak için kullanılan, Cairo grafik kütüphanesi için Python bağlantısı sunar (Reiter ve stuaxo, 2023).
- rdkit (2023.9.1): Kimyasal bilişim ve moleküler modelleme işlemleri için kullanılan bir kütüphanedir (Landrum ve ark., 2023).
- re (Regüler İfadeler): Python'un standart kütüphanesidir. Metin işleme ve arama işlemleri için kullanılır. Metin içinde belirli örüntülere dair aramalar yapmayı ve metni dönüştürmeyi sağlar (Python, 2023).
- scikit-learn (1.3.0): Makine öğrenimi algoritmaları için kullanılan geniş kapsamlı bir kütüphanedir (Pedregosa ve ark., 2011).
- scipy (1.11.3): Bilimsel hesaplamalar için kullanılan kapsamlı bir kütüphane, matematiksel işlemler ve istatistiksel analizler için geniş araçlar sunar (Virtanen ve ark., 2020).
- seaborn (0.12.2): Matplotlib temelli gelişmiş veri görselleştirme kütüphanesi, estetik grafikler ve çizimler için kullanılır (Waskom, 2021).
- sys: Python yorumlayıcısı ve çalışma ortamı ile ilgili bilgileri elde etmek ve yürütme ortamını yönetmek için kullanılan standart bir modüldür (Python, 2023).
- textract (1.6.1): Belgelerden metin çıkarmak için kullanılan bu kütüphane, çeşitli belge türlerini işleyebilme kapasitesine sahiptir (Amazon, 2023).
- wordcloud (1.9.3): Kelime bulutu oluşturmak için kullanılmaktadır (Python, 2023)
- WordNetLemmatizer: nltk kütüphanesine ait olan bu araç, lemmatizasyon işlemi için kullanılır. Lemmatizasyon, bir kelimenin kök halini bulmayı içerir, farklı eklere sahip aynı köke ait kelimelerin analizini kolaylaştırır (Loper ve Bird, 2002).

3.4. Bağımsız Yazılım Protokoller

Bu araştırmada, iki önemli dış kaynaklı yazılım protokolünden faydalanılmıştır. İlk protokol (USEARCH), benzer sekanstaki peptitleri elemek ve araştırmada yanlılığı

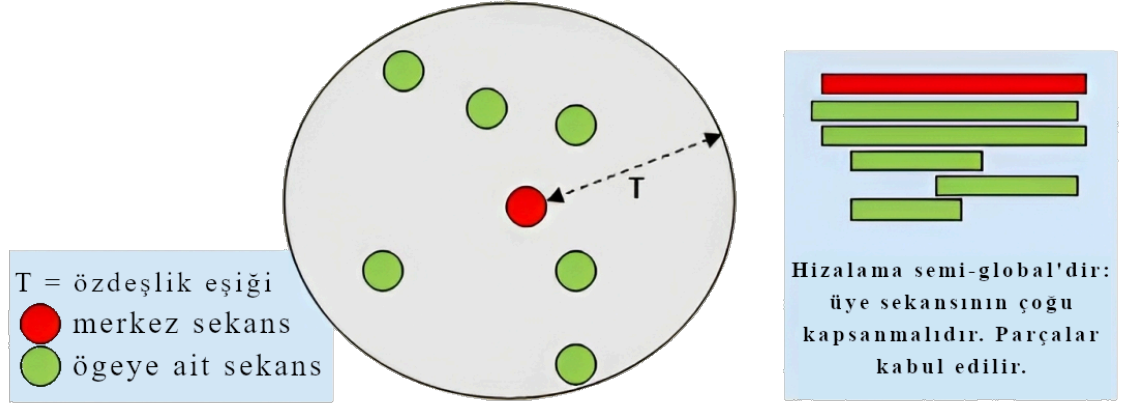
önlemek için seçilmiştir. Bu protokol, veri setlerinin temizlenmesi ve daha doğru sonuçlar elde edilmesi için kullanılmıştır (Edgar, 2010). İkinci protokol (PepFun) ise, peptitlerin fizikokimyasal özelliklerinin hesaplanmasında kullanılmıştır. Bu protokol, peptitlerin moleküler yapısını ve etkileşim özelliklerini analiz etmek için gerekli hesaplamaları gerçekleştirmektedir (Ochoa ve Cossio, 2021).

3.4.1. USEARCH (UCLUST algoritması)

Bu çalışmada, USEARCH'ün 11.0.667 Windows 32-bit sürümü kullanılmıştır. USEARCH, özellikle protein ve peptit, amino asit sekanslarının kümeleme ve karşılaştırılma işlemleri için tasarlanmış bir araçtır. Bu araç, UCLUST algoritmasını temel alarak, benzer sekansları hızlı ve etkili bir şekilde tanımlamak ve gruplamak için kullanılır. USEARCH'ün seçilmesindeki temel neden, Windows işletim sistemi ile uyumluluğudur. Alternatif programlar olan MeShClust, CD-HIT ve DNAClust ile karşılaştırıldığında USEARCH bu yüzden tercih edilmiştir (Edgar, 2010).

UCLUST, her kümenin bir merkez (centroid) dizisi tarafından tanımlandığı ve her sekansın, belirli bir benzerlik eşiğinin üzerinde olan bir merkez dizisi ile eşleşmesi gerektiği bir yapı kullanır (Şekil 3.1). UCLUST, proteinler ve peptitler için yaklaşık %50 ve üzeri özdeşlik eşiklerinde etkilidir. Daha düşük eşiklerde, sonuçların kalitesi sorgulanmaktadır çünkü hizalama kalitesi düşer ve homoloji güvenilir bir şekilde belirlenemez. (Edgar, 2010). Algoritmanın temel prensipleri şunlardır:

- Tüm merkezlerin birbirleriyle olan benzerlikleri, belirlenen eşik değeri T'den düşük olmalıdır (Edgar, 2010).
- Tüm üye diziler, bir merkez ile T veya üzeri benzerlik göstermelidir (Edgar, 2010).



Şekil 3.1. Merkez-tabanlı, orta ile yüksek özdeşlikte kümeleme (Edgar, 2010).

USEARCH, benzer peptit dizilerini etkin bir şekilde tespit ederek, araştırmada kullanılan veri setlerinin kalitesini ve tutarlılığını artırmıştır. Bu, özellikle büyük veri setlerinin işlenmesi ve analiz edilmesi sırasında, veri yoğunluğunu azaltmak, yanlışlığı önlemek ve hesaplama verimliliğini artırmak için önemlidir. Bu sayede, araştırma sonuçlarının güvenilirliği artırılır ve biyolojik veriler üzerinden yapılacak çıkarımlar daha sağlam bir temele oturtulur (Edgar, 2010).

3.4.2. PepFun

Bu araştırmada, PepFun sürüm 1.0 kullanılmıştır. PepFun, peptitlerin çeşitli özelliklerini biyoinformatik yöntemlerle incelemek için tasarlanmış bir araçtır. Bu araç, peptitlerin logP, hidrofobisite, karasızlık indeksi, molekül ağırlığı ve moleküler yük gibi önemli fizikokimyasal özelliklerinin hesaplanmasında kullanılmıştır. PepFun, peptitlerin sekans ve yapılarını içeren geniş veri setleri üzerinden çeşitli analizlere imkan tanıyan fonksiyonlara sahiptir (Ochoa ve Cossio, 2021).

PepFun, Python programlama dili kullanılarak geliştirilmiş olup BioPython ile RDKit gibi açık kaynak projelerin fonksiyon ve metotlardan yararlanmaktadır. Bu araç, MHC sınıf II ve Granzyme B proteaz gibi sistemlerdeki peptit bağlayıcıların analiz edilmesinde kullanılmıştır. PepFun'un sağladığı bu çeşitlilik, peptitlerin çok yönlü analizini sağlayarak, biyolojik ve kimyasal verilerin daha kapsamlı bir şekilde değerlendirilmesine imkan tanır. Araştırmada bu yazılımın kullanılması, peptitlerin detaylı fizikokimyasal profillemesine olanak sağlamakta ve böylece peptitlerin işlevsel

özellikleri üzerine daha derinlemesine analizler yapılmasını mümkün kılmaktadır (Ochoa ve Cossio, 2021).

3.5. Literatür Taraması ve Kelime Analizi

Literatür taraması ve analizi için özel bir metot, antimikrobiyal peptitlerin geniş etki yelpazesi ve son yıllarda artan popülerliği göz önünde bulundurularak geliştirilmiştir. Özellikle *in-silico* yöntemlerin standartlaştırılamaması nedeniyle, bu alandaki çeşitli araştırma yöntemleri, sonlanım noktaları ve başarı ölçütleri üzerine detaylı bir araştırma yapılmıştır.

Araştırmanın hazırlık aşamasında toplam 154 makale dikkatle incelenmiştir. Bu makalelerin 59'u derleme niteliğindedir ve konu hakkında geniş bir perspektif sunmaktadır. Literatür taraması sürecinde, “Google Scholar” ve “EBSCO Discovery Service” gibi önde gelen akademik arama motorları kullanılmıştır.

Araştırma sürecindeki literatür taraması, antimikrobiyal peptitlerin kansere karşı etkilerinin anlaşılması ve *in-silico* yöntemlerle analiz edilmesi üzerine odaklanmıştır. İlk arama terimi olarak belirlenen (“*anticancer properties*” AND “*antimicrobial peptides*” AND “*in silico*”) kombinasyonu, çalışmanın temel konusunu yansıtmaktadır.

İkinci arama terimi olarak seçilen (“*host defence peptides*” AND “*in silico*”) kombinasyonu, antimikrobiyal peptitlerin konak savunma mekanizmalarındaki rollerini ve bu konudaki bilgisayar tabanlı analizleri kapsamaktadır.

Yeni teknolojiler ve metodolojilerin sürekli evrim geçirmesi, özellikle 2010 yılından itibaren yapılan çalışmaların incelenmesini gerektirmiştir. Bu çalışma kapsamında gerçekleştirilen literatür analizi, araştırmacı tarafından özenle seçilen sonuçlar üzerinden yürütülmüştür. İlgili literatür, ileri analizler için PDF formatında kaydedilmiştir. Bu PDF dosyaları, araştırmacı tarafından geliştirilen özel bir algoritma kullanılarak kelime analizine tabi tutulmuştur. Belirli bir klasördeki tüm dosyalar taranarak, bu dosyalardaki kelimelerin sayısı hesaplanmıştır.

Bu tarama işlemi sırasında alttaki kurallı ifade kullanılmıştır.

$$r''[\alpha - \omega A - \Omega a - zA - Z0 - 9/''-] + '' \quad (3.1)$$

Bu örüntü, Latin alfabesini kapsayacak şekilde özelleştirilmiştir. Böylece “ω-3” ve “α-helikal” gibi terimler, metin içerisinde doğru bir şekilde yakalanabilmektedir. Bu yaklaşım, belgelerdeki biyolojik ve kimyasal terimlerin doğru bir şekilde işlenmesini ve analiz edilmesini sağlamıştır.

Araştırmada kullanılan bir diğer önemli algoritma ise kelimelerin kök halini bulmayı amaçlamaktadır. Bunun için “nltk” kütüphanesinin “WordNetLemmatizer” aracı kullanılmıştır. Bu sayede, farklı eklere sahip aynı kelimeler birleştirilerek daha sağlıklı bir analiz gerçekleştirilebilmiştir.

Bu tarama, literatürün geniş bir bölümünü kapsayarak çalışmanın güvenilirliğini ve kapsamını artırmıştır. Ayrıca, analiz sonuçları, daha odaklı bir literatür taraması gerçekleştirilmesine yönelik olarak araştırmacı tarafından değerlendirilmiştir.

Elde edilen sonuç bir Microsoft Excel sayfası (“xlsx”) formatında kaydedilerek, çeşitli analiz ve görselleştirme süreçlerine tabi tutulmuştur.

3.6. Veri Seti Kaynakları ve Seçimi

Bu araştırmada kullanılan ana ve bağımsız veri setleri; Tyagi ve ark., 2013 ile Boopathi ve ark., 2019 çalışmalarından derlenmiştir. Deneysel veri seti ise “Antimikrobiyal Aktivite ve Peptit Yapısı Veri Tabanı” (DBAASP) veri tabanından elde edilmiştir (Pirtskhalava ve ark., 2021). Ana ve bağımsız veri setlerinin seçiminde negatif ve pozitif örneklerin dengeli bir dağılıma sahip olması, pozitif örneklerin *in-vitro* ve/veya *in-vivo* çalışmalar ile doğrulanmış kaynaklardan alınmış olması ve negatif örneklerin rastgele oluşturulmamış olması gibi kriterlere özellikle dikkat edilmiştir.

Veri seti seçimi, bu alandaki araştırmalar için kritik öneme sahiptir. Dengeli veri setleri, makine öğrenimi modellerinin eğitiminde yanlılığın oluşmasını önlemektedir. Pozitif örneklerin doğrulanmış kaynaklardan elde edilmesi, araştırmanın bilimsel geçerliliğini artırırken, negatif örneklerin dikkatli bir şekilde seçilmesi, yanıltıcı sonuçların önlenmesine yardımcı olmuştur. Bu yaklaşım, modelin genel performansını ve tahmin

doğruluğunu iyileştirir ve sonuçların yorumlanmasında sağlam bir temel oluşturur (Tyagi ve ark., 2013; Boopathi ve ark., 2019).

3.6.1. Ana Veri Seti

Ana veri seti; Tyagi et al., 2013 ve Boopathi ve ark., 2019 çalışmalarının ana veri setlerinden derlenmiştir. Bu çalışmada kullanılan ana veri seti, virgülle ayrılan değerler (“csv”) formatında kaydedilmiştir. Dosyada, peptitler ile ilgili iki kayıt mevcuttur.

İlk olarak, incelenen peptitlerin amino asit sekanslarını içermektedir. Her sekans, belirli bir peptidin yapısal özelliklerini temsil eden bir dizi amino asit olarak karakter tipinde kodlanmıştır. İkinci olarak ise, her bir peptidin antikanser etkisini gösteren ikili (binary) verileri barındırır. Burada “1”, peptidin antikanser etkinliğe sahip olduğunu; “0” ise antikanser etkisi olmadığını göstermektedir.

Bu çalışmada ele alınan ana veri seti, modelin eğitimi ve test edilmesi amacıyla iki ayrı bölüme ayrılmıştır. Eğitim ve test veri setleri olarak %80 - %20 oranında bölünmektedir. Eğitim veri seti, modelin peptitlerin amino asit sekansları ve antikanser etkileri üzerinden öğrenmesini sağlamak için kullanılmıştır. Test veri seti ise, eğitilen modelin öğrendiklerini doğru bir şekilde genelleyip genellemeyeceğini test etmek amacıyla kullanılmıştır.

704 adet peptit sekansı içermektedir ve dengeli bir veri setidir. Veri setinin dengeli olması negatif ve pozitif örneklerin eşit olduğunu belirtmektedir. En uzun peptit 113 amino asit ve en kısa peptit 5 amino asit içermektedir. Ortalama sekans uzunluğu ise 25,83’tür.

3.6.2. Bağımsız Veri Seti

Aynı şekilde bağımsız veri seti de Tyagi et al., 2013 ve Boopathi ve ark., 2019 çalışmalarının bağımsız veri setlerinden derlenmiştir ve “csv” formatında kaydedilmiştir. Ayrıca, ana veri seti ile aynı yapıdadır.

Derlenen bağımsız veri seti validasyon veri seti olarak kullanılmaktadır. Validasyon süreci, modelin gerçek dünya verileri üzerindeki performansını test etmek ve bu şekilde elde edilen sonuçların genel geçerliğini sağlamak amacıyla gerçekleştirilmiştir. Bu

kapsamda, bağımsız veri seti, modelin doğruluk ve güvenilirliğinin belirlenmesinde esas alınan bir araç olarak ön plana çıkmaktadır.

Dengeli olan bu veri seti 396 kayıt içermektedir. En uzun peptit 54 amino asit ve en kısa peptit 6 amino asit içermektedir. Ortalama sekans uzunluğu ise 26,53'tür.

3.6.3. Deneysel Veri Seti

Bu çalışmada kullanılan deneysel veri seti, DBAASP veri tabanından elde edilmiştir (Pirtskhalava ve ark., 2021). Veri seti, 10 ile 40 amino asit uzunluğunda olan ve çeşitli hedef gruplara (mantar, gram-pozitif ve gram-negatif bakteriler, molliküt, parazit, virüs) karşı potansiyel etkisi bulunan toplam 15.264 peptiti kapsamaktadır.

Sekans uzunluğuna ait aralık, optimal fizikokimyasal özelliklerin elde edilebilmesi amacıyla seçilmiştir. Fizikokimyasal özellikler, peptitlerin biyolojik aktivitesinin yanı sıra, ilaç geliştirme süreçlerinde karşılaşılabilecek potansiyel zorluklardan olan suda çözünürlük, logP ve stabilite vb. özellikleri belirlemektedir.

Mikrobiyal hedef gruplarına yönelik peptitler, kanser üzerine etkinlikleri bilinmeyen bir veri seti oluşturmak için tercih edilmiştir. Böylece peptitlerin, ikili etkilerinin araştırılması ve hem mikrobiyal hem de kanser hedefleri üzerindeki etkilerinin daha geniş bir perspektifle değerlendirmesi amaçlamaktadır.

Bu veri seti, “csv” formatında kaydedilmiştir ve yalnızca peptit sekanslarına ait bilgileri içermektedir.

3.7. Veri Setlerinin İşlenmesi ve Hazırlanması

Bu çalışmada, ana ve bağımsız veri setlerinin işlenmesi ve hazırlanması, makine öğrenimi modelinin doğruluğunu ve güvenilirliğini artırmak amacıyla titizlikle gerçekleştirilmiştir. İlk olarak, her iki veri setinde bulunan ve aynı olan örnekler elenmiştir. Ardından, ayrı olarak her bir veri setindeki, belirli oranda benzerlik gösteren sekanslara sahip örnekler de model eğitiminde yanlılık oluşmasını önlemek amacıyla çıkarılmıştır.

Bir sonraki adımda veri setleri rastgele karıştırılarak yeniden dengelemeye tabi tutulmuş, negatif ve pozitif örneklerin miktarları eşitlenmiştir. Bu işlem, modelin her iki sınıfı adil bir şekilde temsil etmesini sağlamak için önemlidir.

Veri setlerinin mevcut durumunu değerlendirmek için ön istatistiksel analizler yapılmıştır. Bu analizler, veri setlerinin genel özelliklerini ve sekansların amino asit dağılımlarını öğrenmek için gerçekleştirilmiştir.

Amino asit sekansları sayısal formata dönüştürülmüştür. Bu dönüşüm, makine öğrenimi algoritmalarının verileri işleyebilmesi için gereklidir ve sekans bilgisinin kantitatif analizine imkan tanır. Son olarak, özellik çıkarımı yapılmıştır. Bu süreç, peptitlerin biyolojik ve kimyasal özelliklerini yansıtan önemli bilgilerle, makine öğrenimi modellerinin etkili bir şekilde eğitilmesini sağlamaktadır.

3.7.1. Kümeleme ve Dengeleme İşlemleri

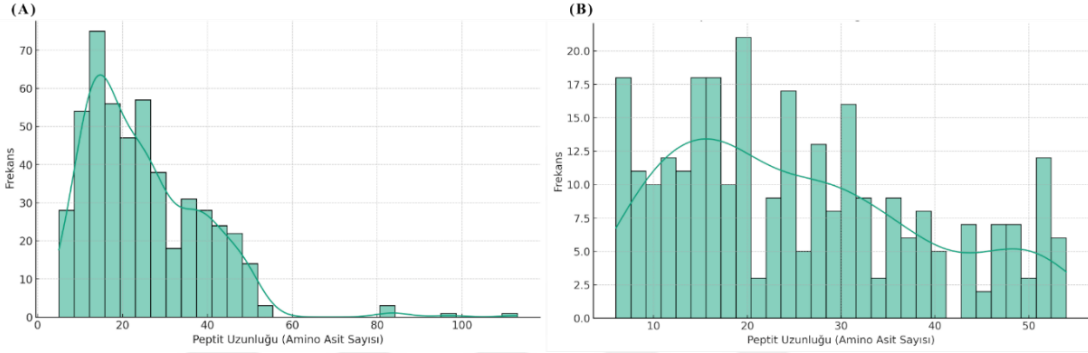
Ana ve bağımsız veri setinin kümeleme işlemi için USEARCH kullanılmıştır. İlk olarak veri setindeki peptit sekansları “FASTA” formatına çevrilerek USEARCH programı için hazırlanmıştır. Bu işlem sırasında antikanser özelliği, kaybolmaması amacıyla, 0 veya 1 olarak sekans başlığı içerisine, sekans numarası ile birlikte, yazdırılmıştır. Ardından, USEARCH komutu ile %50 benzerlik eşiği kullanılarak eşik üstünde kalan peptitler elenmiştir.

Sonuç FASTA formatında elde edilmiştir. Bu yüzden, veri setinin tekrardan bir “csv” dosyasına yazılması gerekmektedir. FASTA dosyasındaki her bir kayıt ayrıştırılarak; sekans numarası, antikanser özelliği ve amino asit sekansı alınmıştır.

Her iki veri setinden de kayıtlar, negatif ve pozitif örnekler için farklı oranlarda elenmiştir. Bu yüzden veri setinin tekrardan dengelenmesi gerekmektedir. İlk olarak, “antikanser” sütununa göre veri seti negatif (0) ve pozitif (1) sınıflara ayrılmıştır. Her sınıfın örnek sayısı hesaplanarak daha az sayıda örneğe sahip sınıf belirlenmiştir. Sonrasında, yanlılığı önlemek için veri seti rastgele karıştırılmıştır. Bu işlemin ardından, her iki sınıfa ait kayıtlardan, azınlık sınıfın sayısı kadar örnek seçilerek birleştirilmiştir. Sonrasında, analiz için kullanılacak nihai veri setini oluşturan yeni bir “csv” dosyasına

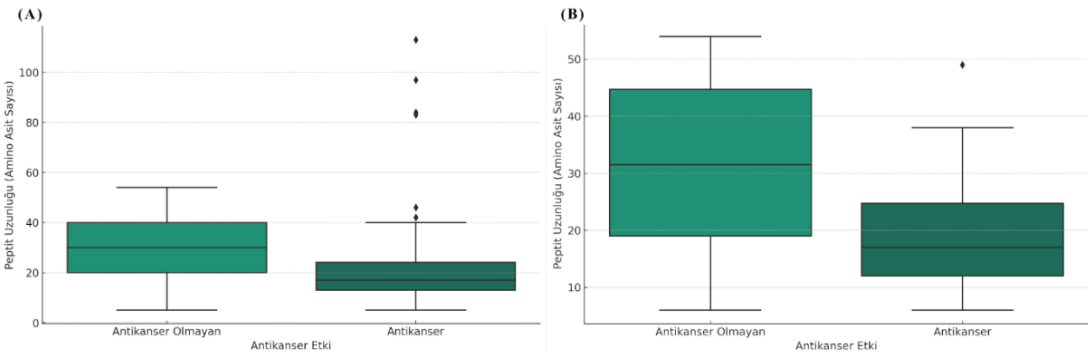
kaydedilmiştir. Bu adım, her iki sınıfın da eşit sayıda örneğe sahip olmasını sağlayarak dengeli bir veri seti oluşturmuştur.

Elde edilen ve çalışmada kullanılacak ana veri setinde 500 peptit (250 negatif ve 250 pozitif) sekansı kalmıştır. En uzun peptit 113, en kısa peptit ise hala 5 amino asit uzunluğundadır. Ortalama sekans uzunluğu ise 24,81'e düşmüştür.



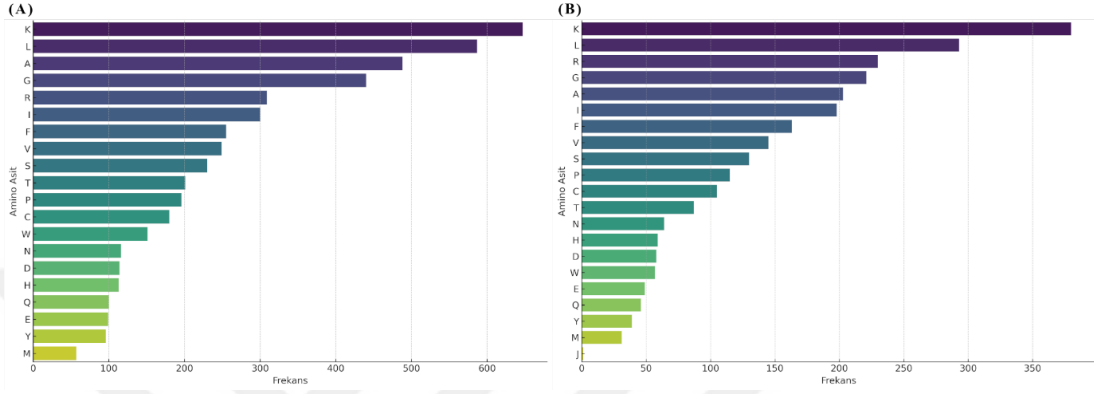
Şekil 3.2. Veri setlerinde yer alan peptit sekanslarının uzunlukları ve dağılımı, A) Ana veri seti, B) Bağımsız veri seti.

Bağımsız veri setinde ise peptit sayısı 284'e (142 pozitif ve 142 negatif) düşmüştür. Aynı şekilde en uzun ve en kısa peptitlere ait sekans uzunlukları değişmemiştir. Ortalama peptit uzunluğu ise 25,45'tir. Her iki veri setinde yer alan peptitlerin uzunlukları ve dağılımı bir histogram ile (Şekil 3.2) gösterilmiştir.



Şekil 3.3. Veri setlerinde yer alan peptit sekanslarının uzunlukları ve antikanser özellikleri, A) Ana veri seti, B) Bağımsız veri seti.

Antikanser etkinliği ile peptit uzunluklarını karşılaştıran kutu diyagramı ise Şekil 3.3'te görülmektedir. Son olarak, her iki veri setinde antikanser özelliğe sahip peptitlerin amino asit frekansları görselleştirilmiştir (Şekil 3.4).



Şekil 3.4. Veri setlerinde yer alan antikanser özelliğe sahip peptitlere ait amino asit sayımları, A) Ana veri seti, B) Bağımsız veri seti

3.7.2. Numerik Haritalama

Peptit sekansları, makine öğrenimi algoritmaları tarafından işlenebilir hale getirilmek için sayısal değerlere dönüştürülmüştür. Bu süreç, her amino asidi temsil eden bir sayı ile oluşturulmuş bir sözlük kullanılarak gerçekleştirilmiştir. Standart ve non-standart amino asitler dikkate alınarak, özgül sayısal değerler atanmıştır (örn. “A” için 1, “R” için 2). “X” (25) ise veri setinde bulunan ve tanımlı olmayan (kukla) amino asitleri temsil etmek için kullanılmıştır.

Veri seti, bu sayısal dönüşüm işleminden önce temizlenmiştir. Geçersiz (NaN) değerler çıkarılmıştır. Daha sonra, tüm sekanslar büyük harfe dönüştürülerek karakter dizisi tipinde oldukları doğrulanmıştır. Bu adım, olası hataların önüne geçmek için önemlidir.

Her bir peptit sekansı, tanımlanan sözlük kullanılarak liste formatında sayısal dizilere dönüştürülmüştür. Son olarak, veri seti sonraki analiz aşamalarında kullanılmak üzere; “sekans”, “numerik_sekans” ve “antikanser” kolonları ile yeni bir “csv” dosyasına kaydedilmiştir.

3.7.3. Özellik Seçimi

Makine öğrenimi modelleri için özellik seçimi, bu modelleri eğitmede kullanılacak en ilgili ve etkili özellikleri belirleme sürecidir. Bu, veri setini temsil etmekte zayıf olan özellikleri çıkararak ve en etkili olanları seçerek modelin performansını arttırmayı ve modelin genelleştirme yeteneğini artırmayı amaçlamaktadır.

Bu çalışma kapsamında üç özellik seçilmiştir; amino asit kompozisyonu (AAC), dipeptit kompozisyonu (DPC) ve ikili ayırlama veya diğer adıyla tek-sıcak kodlama (OHE). AAC ve DPC, sözlük formatında oldukları ve birlikte daha yüksek başarı gösterdikleri için kombine olarak kullanılmışlardır. Öte yandan, OHE ise tek-boyutlu liste dizileri (vektör) olarak işlenmektedir ve tek başına kullanılmıştır.

Bu özelliklerin tercih edilmesinin temel nedenleri, aynı alandaki araştırmalarda yaygın olarak faydalanılmaları, analizleri için ileri düzey donanım gerektirmemeleri ve çıkarım yöntemlerinin standardize edilmiş olmasıdır. AAC ve DPC kombinasyonu, her peptit için hem genel bir perspektif hem de amino asitler arasındaki ilişkilere dair özgün bilgiler sağlamaktadır. OHE ise, peptitlerin yapısal özelliklerinin bir tek özellik içerisinde daha detaylı incelenbilmesine olanak sağlamaktadır. Ayrıca, vektörize yapısı analiz verimliliğini arttırmaktadır.

Özellik çıkarımı işlemi için ilk olarak veri seti yüklenip “numerik_sekans” kolonu ayrılarak bir veri çerçevesine kaydedilmiştir.

AAC, bir peptidin bileşenleri olan amino asitlerin bileşimini ifade etmektedir. AAC, her bir amino asidin peptit içerisindeki oranını verir. AAC, her amino asidin peptit içerisindeki sayısını toplam amino asit sayısına bölerek yüzdesel olarak hesaplanır. Matematiksel formülü aşağıdaki şekildedir:

$$AAC(i) = Ri \div N \times 100 \quad (3.2)$$

Burada, $AAC(i)$ amino asidin yüzde kompozisyonunu, Ri bir amino asidin peptitteki sayısını ve N ise peptidin toplam amino asit sayısını temsil eder. Bu hesaplama, her bir amino asidin peptit içindeki göreceli ağırlığını belirlemek için kullanılır. Sonuç olarak,

her peptit için 26 (non-standart amino asitler de dahil) boyutlu bir sözlük elde edilerek oluşturulmuş olan veri çerçevesine eklenir.

DPC özelliği, peptitlerin yapısal ve fonksiyonel özelliklerini daha ayrıntılı bir şekilde anlamak için kullanılan önemli bir özellik çıkarımı yöntemidir. DPC, tüm amino asit çiftlerinin (dipeptitlerin) kompozisyonunu içerir. Bu yöntem, peptitlerin lokal amino asit sıralamasını temsil eder ve 676 boyutlu bir vektör (26×26) ile ifade edilebilir.

DPC hesaplamak için, ilk olarak 26 amino asidin sayısal değerlerinden oluşan tüm olası dipeptit çiftleri hesaplanır (Örn. 1-1, 1-2, 1-3). Daha sonra, bu dipeptitlerin peptit içindeki kompozisyonu belirlenir. Matematiksel olarak DPC'nin hesaplanması şu formülle ifade edilir:

$$DPC(i) = \frac{\text{Toplam Dipeptit Sayısı}(i)}{(N-1)} \times 100 \quad (3.3)$$

Burada, $DPC(i)$ belirli bir dipeptit türünün kompozisyonunu, N ise peptit uzunluğunu temsil eder. DPC hesaplaması; peptit içindeki her bir dipeptit çiftinin sayısını hesaplamak ve bu sayıları peptit uzunluğunun bir eksiği ile bölerek yüzde olarak elde edilir. Aynı şekilde, sözlük formatında elde edilen özellik oluşturulmuş olan veri çerçevesine eklenir.

OHE yöntemi, peptitlerin amino asit kompozisyonlarını ikili kodlama tekniği ile ifade etmekte kullanılan popüler bir yöntemdir. Bu metot, her bir amino asidi, belirli bir pozisyonda 1 ve geri kalan 25 pozisyonda 0 değerlerini alacak şekilde 26 boyutlu bir vektör ile temsil eder. Bu sayede, her amino asit benzersiz bir ikili vektör ile karakterize edilmiş olur.

OHE'de tüm peptit sekansları, C-terminus sonrasına bir kukla amino asit eklenerek veri setindeki en uzun peptidin uzunluğuna tamamlanır. Bu kukla amino asit, tüm değerleri sıfır olan bir vektör ile temsil edilir. Bu yöntem, peptitlerin farklı uzunluklarının standartlaştırılmasına ve makine öğrenimi modellerinin daha etkili bir şekilde eğitilmesine yardımcı olur.

OHE işlemi, her bir amino asidin sayısal değerini alır ve bu değere karşılık gelen pozisyonda 1, diğer tüm pozisyonlarda 0 olan bir vektör oluşturur. Bu işlem, peptidin tüm amino asitleri için tekrarlanır. Eğer peptit maksimum uzunluğa ulaşmazsa, kalan kısımlar tamamen sıfırlardan oluşan vektörler ile doldurulur. Yani, OHE, peptitlerin amino asit kompozisyonunu basit ve etkili bir şekilde ifade eden bir yöntemdir. Sonuç olarak, veri setindeki en uzun peptidin amino asit sayısında ve her amino asit için 26 boyutlu bir vektör elde edilir.

Elde edilen bu özellik de diğer iki özelliğin sonuna oluşturulmuş olan veri çerçevesinde eklenir. AAC ve DPC özellikleri sözlük formatında, OHE ise vektör formatındadır. Bu üç özelliği de içeren veri çerçevesi bir sonraki analiz için bir “csv” dosyasına aktarılmıştır.

3.8. Makine Öğrenimi Modeli Geliştirme

Bu bölümde, peptitlerin antikanser etkinliklerinin tanımlanmasında kullanılacak makine öğrenim modelinin oluşturulması için kapsamlı bir metodoloji izlenmiştir. İlk olarak OHE özelliği ve AAC ile DPC özellik kombinasyonu, ana veri seti kullanılarak, çeşitli makine öğrenimi modellerinde denenmiştir. Bu denemelerin amacı, peptitlerin sınıflandırılmasında en yüksek başarıyı gösteren model ve özellik eşleşmesini belirlemektir.

İlk olarak RF, lojistik regresyon gibi çeşitli makine öğrenimi algoritmaları karşılaştırılmıştır. Bu karşılaştırma, her bir algoritmanın AAC ve DPC kombinasyonu ile OHE özellikleri üzerindeki etkinliğini değerlendirmek amacıyla yapılmıştır. En başarılı model ve özellik kombinasyonu belirlendikten sonra, ilgili modelin performansını arttırmak için hiperparametre optimizasyonu gerçekleştirilmiştir. Bu süreç, modelin antikanser peptit tanımlamada güvenilir ve etkili bir şekilde görev almasını sağlamak için kritik öneme sahiptir.

En iyi model/özellik seçimi için başarı, temel değerlendirme kriterleri (doğruluk, hassasiyet vb.) ile ölçülmüştür. Hiperparametre optimizasyonu gerçekleştirilmiş modelin genelleştirme yeteneği ve performansı ise daha detaylı bir ölçüm yöntemi olan 10-katlı çapraz doğrulama kullanılarak test edilmiştir.

3.8.1. Değerlendirme Metodolojisi

Makine öğrenimi modellerinin performansının kapsamlı bir şekilde değerlendirilmesi için çeşitli istatistiksel ölçütler ve metodolojiler kullanılmıştır. Modelin başarısını değerlendirmek için altta açıklanan temel kriterler kullanılmıştır.

Doğruluk, modelin hem pozitif hem de negatif sınıfları ne kadar doğru tahmin ettiğinin genel bir ölçüsüdür. Doğru tahmin edilen örneklerin toplam örnek sayısına oranı olarak hesaplanır ve modelin genel performansını yansıtır.

$$\text{Doğruluk} = \frac{\text{Gerçek Pozitif (GP)} + \text{Gerçek Negatif (GN)}}{\text{GP} + \text{Yanlış Pozitif (YP)} + \text{GN} + \text{Yanlış Negatif (YN)}} \quad (3.4)$$

Duyarlılık, modelin pozitif sınıfı ne kadar doğru tahmin ettiğini gösterir. Gerçek pozitiflerin, tüm gerçek pozitiflerin toplamına oranı olarak hesaplanır. Diğer bir deyişle, gerçekten pozitif olan durumların kaçının model tarafından pozitif olarak doğru saptandığını gösterir.

$$\text{Duyarlılık} = \frac{GP}{GP + YN} \quad (3.5)$$

Özgüllük, modelin negatif sınıfı ne kadar doğru tahmin ettiğinin bir ölçüsüdür. Gerçek negatiflerin, tüm gerçek negatiflerin toplamına oranı olarak hesaplanır. Gerçekten negatif olan durumların kaçının model tarafından negatif olarak doğru saptandığını gösterir.

$$\text{Özgüllük} = \frac{GN}{GN + YP} \quad (3.6)$$

Kesinlik, modelin pozitif olarak tahmin ettiği durumların gerçekten pozitif olma oranını gösterir. Gerçek pozitiflerin, toplam tahmin edilen pozitiflere oranı olarak hesaplanır. Yani, modelin yaptığı pozitif tahminlerin doğruluk oranını gösterir.

$$\text{Kesinlik} = \frac{GP}{GP + YP} \quad (3.7)$$

F1 skoru, modelin duyarlılık ve kesinlik değerlerinin ortalamasını olarak hesaplanır. Bu metrik, modelin hem pozitifleri hem de negatifleri dengeli bir şekilde tahmin etme

kabiliyetini değerlendirir ve özellikle dengesiz veri setleri ile geliştirilen modellerin değerlendirilmesinde önemlidir.

$$F1 Skoru = 2 \times \frac{Kesinlik \times Duyarluluk}{Kesinlik + Duyarluluk} \quad (3.8)$$

Alıcı işletim karakteristik eğrisi altında kalan alan (ROC AUC) skoru, modelin sınıflandırma başarısını farklı eşik değerleri için değerlendiren bir ölçüttür. ROC AUC, modelin rastgele seçilen bir pozitif örneği rastgele seçilen bir negatif örneğe tercih etme olasılığını gösterir ve modelin ayırma kabiliyetini ifade eder. Gerçek pozitif oranı ve yanlış pozitif oranı arasındaki ilişkiyi temsil eden bir eğrinin altında kalan alanı ifade eder. Özellikle dengesiz veri setlerinin değerlendirilmesi için tercih edilmektedir.

Bu temel ölçütler, modellerin genel performansını, yanlılık eğilimini ve sınıflandırma yeteneğini objektif bir şekilde değerlendirmek için kritik öneme sahiptir.

Hiperparametre optimizasyonu gerçekleştirildikten sonra ise, model performansının daha güvenilir bir şekilde, ana ve bağımsız veri setleri kullanılarak, test edilebilmesi için 10-katlı çapraz validasyon tekniği kullanılmıştır. Bu yöntemde, ana veri seti rastgele 10 eşit boyutlu alt kümeye ayrılır. Bu alt kümelerden dokuzu modelin eğitimi için kullanılırken, kalan küme ise iç test için doğrulama verisi olarak kullanılır. Bu işlem her küme için tekrarlanır, yani her bir alt küme bir kez doğrulama seti olarak kullanılır. Son olarak, 10 denemenin sonuçlarının ortalaması, standart sapması ve standart hatası hesaplanır.

10-katlı çapraz validasyon, modelin genelleştirme kabiliyetini ve tutarlılığını değerlendirmek için literatürde sıkça tercih edilmiştir. Bu yöntem, modelin farklı örneklerle karşı performansını objektif bir şekilde test etmeyi sağlar. Ayrıca, bu yöntem modelin aşırı uyma eğilimini azaltmaya yardımcı olur ve daha güvenilir sonuçlar elde edilmesini sağlar.

3.8.2. İstatistiksel Analiz

Ana veri seti ile eğitilen makine öğrenmesi modelinin ve validasyon sonuçlarının arasındaki anlamlılık t-testi ve çok değişkenli varyans analizi (MANOVA)

kullanılmıştır. Bu sonuçlar bir önceki bölümde açıklandığı üzere 10-katlı çapraz validasyon metodu ile oluşturulmuştur. Ayrıca, tüm temel metrikler için değerler içermektedir.

Student'ın t-testi, istatistiksel bir hipotez testi olan t-testi, iki grup arasındaki ortalama farkları değerlendirir. Bağımsız veya eşleştirilmiş örnekler için uygundur ve normal dağılıma dayanır. Gözlemlenen ortalama farkın, şans eseri meydana gelme olasılığını belirler (Pituch ve Stevens, 2015).

MANOVA, birden fazla bağımlı değişken üzerindeki grup etkilerini aynı anda değerlendiren bir istatistiksel yöntemdir. Korelasyonları dikkate alır ve gruplar arasındaki genel farklılıkları test eder. Çok değişkenli normallik ve eş varyans varsayımlarına dayanır. Bağımlı değişkenlerin birbiriyle ilişkili olması ve verilerin normal dağılım göstermesi beklenir (Pituch ve Stevens, 2015).

Shapiro-Wilk (SW) testi ise, bir veri kümesinin normal dağılıma uyup uymadığını değerlendirmek için kullanılan bir istatistiksel testtir. Örneklem dağılımının normal dağılıma ne kadar yakın olduğunu belirleyen bu test, özellikle küçük örneklem boyutları için etkilidir. MANOVA ve t-testi kullanılmadan önce veri setinin normal dağılıma uyup uymadığı bu test aracılığıyla kontrol edilmiştir (Pituch ve Stevens, 2015).

3.8.3. Optimal Özellik ile Makine Öğrenimi Algoritmasının Seçimi ve Karşılaştırılması

Bu aşamada antimikrobiyal peptitlerin antikanser etkinliklerinin değerlendirilmesi için dört farklı makine öğrenimi algoritması incelenmiştir; RF, destek vektör makinesi (SVM), lojistik regresyon (LR) ve k-NN. Benzer araştırmalarda sık sık kullanılan bu algoritmalar, veri seti için en uygun olanın seçilmesi amacıyla karşılaştırılmıştır.

En başarılı özellik ve model eşleşmesinin bulunabilmesi için ana veri seti kullanılmıştır. OHE ve AAC/DPC özellikleri sırayla tüm modellerde denenmiştir. Her bir model için, antikanser etki hedef değişken olarak belirlenerek, veri setinin yüzde yirmisi test verisi olacak şekilde bölünmüştür. Performans ölçümünde, bir önceki bölümde bahsedilen temel başarı kriterleri göz önünde bulundurulmuştur. Ancak, araştırmada kullanılan veri

setinin dengeli olması ve sınıflandırmanın genel doğruluğunun araştırmanın temel amacı olması göz önünde bulundurularak, en uygun özellik/model eşleşmesi doğruluk değerine göre belirlenmiştir. Bu durum, ileriki aşamalarda deneysel veri setindeki gerçek pozitif ve negatifleri tespit etmenin öneminden kaynaklanmaktadır. Bu yaklaşım, modelin sınıflandırma performansını ve tahmin yeteneğini daha kapsamlı bir şekilde değerlendirmeyi amaçlamaktadır.

Her bir modelin eğitimi ve test edilmesi sırasında, algoritmaların farklı özellik kombinasyonlarına olan tepkileri incelenmiştir. Bu süreç, modelin veri setindeki özelliklerden en etkili şekilde yararlanma kapasitesini ortaya koymaktadır. Yapılan denemeler, modellerin farklı özellik yapılarına olan uyumlarını ve bu özelliklerin model performansına etkilerini göstermektedir.

İlk olarak, veri seti yüklenmiş ve gerekli özellikler ile hedef değişken ayrıştırılmıştır. Özellikler modelin anlayabileceği düzleştirilmiş bir format haline getirilmiştir. Ayrıca, sözlük formatındaki AAC/DPC özellikleri farklı uzunluklarda oldukları için özellik ölçeklendirme gerçekleştirilmiştir. Sonuçlar bir veri çerçevesine sırayla eklenerek “csv” formatında kaydedilmiştir.

3.8.4. Hiperparametre Ayarlama ve Optimizasyonu

Bu kısımda, RF modeli için hiperparametre ayarlaması ve optimizasyonu ele alınmıştır. Hiperparametre optimizasyonu, algoritmanın performansını artırmak ve en iyi sonuçları elde etmek için kullanılan bir süreçtir. En iyi parametreler, çapraz doğrulama (n=5) yöntemiyle saptanmıştır. Birçok farklı hiperparametre kombinasyonunu denenerek, en iyi performansı veren değerler seçilir ve kaydedilir.

İlk olarak, peptit dizilerinin AAC ve DPC özelliklerini içeren veri seti yüklenmiştir. Özellikler ve hedef değişken ayrıştırılarak, veri seti eğitim ve test setlerine bölünmüş, özellikler ölçeklendirilmiştir. Modelin hiperparametre optimizasyonu için “Grid Search” yöntemi kullanılmıştır. Bu yöntem, belirlenen parametre aralıkları içerisinde en uygun kombinasyonları belirleyerek modelin performansını olabildiğince yükseltmeyi amaçlamaktadır. Kullanılan hiperparametreler altta açıklanmışlardır (Probst ve ark., 2019; scikit-learn developers, 2023):

- Karar ağacı sayısı (n_estimators), modelde kullanılan karar ağaçlarının toplam sayısıdır. Genellikle daha fazla ağaç, daha iyi performans ve daha düşük aşırı uyma riski anlamına gelir, ancak donanım gereksinimini ve işlem süresini artırır.
- Maksimum özellik sayısı (max_features), her karar ağacı bölünürken göz önünde bulundurulacak maksimum özellik sayısını belirtir. Bu parametre, modelin genelleme yeteneğini belirler.
- Maksimum derinlik (max_depth), karar ağaçlarının maksimum derinliğini ifade eder. İdeal değer üstü, aşırı uyum riski oluşturur.
- Minimum örnek bölme sayısı (min_samples_split), bir düğümün bölünmesi için gerekli olan minimum örnek sayısını belirtir. Bu değer arttıkça modelin karmaşıklığı azalır, ancak çok yüksek değerler eksik uyma neden olabilir.
- Yaprak düğümde minimum örnek sayısı (min_samples_leaf), bir yaprak düğümünde olması gereken minimum örnek sayısıdır. Bu, modelde yumuşatma etkisi yaratabilir. Ayrıca, yaprakların her biri belirli limite ulaştığında ağacın bölünmesini durdurarak aşırı uyumu önlemektedir.
- Kalite Ölçüm Fonksiyonu (criterion), karar ağacının bölünmesi sırasında kullanılan fonksiyondur. Bu ağacın hangi özellikleri bölünme için seçeceğini etkilemektedir.
- Önyükleme (bootstrap); her ağaç, önyükleme kullanıldığında gözlem sayısının bir alt kümesi üzerinde eğitilir. Bunlar orijinal veri kümesinden örnekleme yapılarak seçilir. Bu, veri noktalarının sırasının önemli olduğu, amino asit dizileri ile çalışırken olduğu gibi, durumlarda faydalıdır.
- OOB skoru (oob_score); “Out-of-Bag” skoru, eğitim sırasında kullanılmayan örnekler üzerinden modelin performansını değerlendirir. Bu, model eğitilirken uyumu ve genelleme yeteneğini iyileştirir.
- Sıcak başlatma (warm_start); bu parametre ise aktifken, model, yeni ağaçlar eklenerek bir önceki modelin üzerinde eğitilir. Bu, özellikle model aşamalı olarak

eğitilmek istendiğinde kullanışlıdır. 10-katlı çapraz doğrulama yapılırken, her çapraz doğrulama adımının bağımsız olması için, sıcak başlatma parametresi genellikle “False” olarak ayarlanır. Çünkü, her adımda eğitilen model bir sonraki modeli etkilemektedir.

Parametre kombinasyonlarının performansı doğruluk değeri ile ölçülmüştür. En iyi parametreler bir değişken içerisine kaydedilerek sonraki adımda kullanılmıştır.

3.8.5. Çalışma Modelinin Eğitilmesi, Değerlendirilmesi ve Kaydedilmesi

Araştırmanın bu aşaması, modelin eğitilerek performansının daha detaylı değerlendirilmesini ve ileri analizler için “joblib” kütüphanesi yardımıyla kaydedilmesini içermektedir.

İlk olarak veri seti yüklenir ve önceki aşamalardaki gibi ön işlemlere tabi tutulur. Modelin eğitimi için, bir önceki adımda belirlenen en iyi hiperparametreler kullanılmıştır. Bu parametreler, RF sınıflandırıcısının oluşturulması ve eğitilmesi için kullanılmıştır. Model, eğitim seti üzerinde eğitilmiş ve test seti üzerinde tahminler yapmıştır. Ayrıca, modelin performansının daha kolay anlaşılabilmesi için, görselleştirilmek üzere, tahminlerinin gerçek etiketlerle karşılaştırıldığı bir hata matrisi oluşturulmuştur.

Modelin performansını değerlendirmek için, 10-katlı çapraz validasyon yöntemi kullanılmıştır. Bu işlem, modelin genel performansını daha doğru bir şekilde değerlendirmek amacıyla yapılmıştır. Her çapraz doğrulama yinelemesinde, modelin temel performans metrikleri hesaplanmıştır.

Son olarak, elde edilen skorlar ve tahminler, analiz ve raporlama amaçları için veri çerçevelerine dönüştürülmüş ve Excel dosyalarına kaydedilmiştir. Bu süreç, araştırma bulgularının daha sonra detaylı analizi ve yorumlanması için önemli bir adımdır. Ek olarak, optimize edilmiş model ile özellik ölçeklendirici kaydedilmiştir. Özellik ölçeklendirici, veri ön işleme sürecinin bir parçası olarak kullanılan bir araçtır ve modelin eğitimi sırasında kullanılan verilerin ölçeğini ayarlamak için kullanılır. Bu

işlem, modelin ileride yapılacak analizlerde kolayca yeniden kullanılabilmesini sağlamaktadır.

3.9. Bağımsız Veri Seti ile Validasyon

Bu bölümde, daha önce optimize edilmiş modelin bağımsız bir veri seti üzerinde validasyon süreci detaylandırılmaktadır. Modelin, araştırma kapsamında hazırlanan bağımsız bir veri seti üzerindeki performansının değerlendirilmesi amaçlanmaktadır.

Validasyon süreci, bir modelin genelleme yeteneğini ve gerçek dünya verileri üzerindeki performansını değerlendirmek için kritik öneme sahiptir. Bir modelin eğitildiği veri seti üzerinde yüksek doğruluk oranları elde etmiş olması, modelin henüz karşılaşmadığı yeni veriler üzerinde de aynı başarıyı göstereceği anlamına gelmez. Modelin eğitim sırasında ilk gördüğü verilere aşırı uyum sağlaması riski mevcuttur. Bu durum, modelin yalnızca eğitim verilerindeki örüntüleri ezberleyerek yeni verilere karşı esnek ve etkili olmadığını gösterir.

İlk olarak veri seti yüklenip önceki aşamalardaki gibi ön işlemlere tabi tutulmuştur. Önceden eğitilmiş ve “joblib” kütüphanesi kullanılarak saklanmış model, yüklenmiştir. Modelin performansını değerlendirmek için, 10-katlı çapraz validasyon yöntemi kullanılmıştır. Her yinelemenin sonuçları hesaplanmıştır. Aynı şekilde, elde edilen skorlar ve tahminler (hata matrisi), analiz ve raporlama amaçları için veri çerçevelerine dönüştürülmüş ve Excel dosyalarına kaydedilmiştir.

Bilimsel araştırmalarda kullanımından önce modelin güvenilirliğini ve uygulanabilirliğini doğrulamak, önem taşımaktadır. Bu süreç, modelin farklı veri setlerine ne kadar iyi genelleme yaptığını anlamak için yapılan önemli bir testtir.

3.10. Antikanser Özelliklerin Tahmini

Bu bölümde, antikanser özelliklerinin tahmin edilmesi süreci, daha önceden eğitilmiş ve kaydedilmiş olan RF modeli kullanılarak gerçekleştirilmiştir. Öncelikle, peptit dizilerinin özelliklerini içeren deneysel veri seti yüklenmiştir. Daha sonra, mükerrer peptit sekanslarının önüne geçmek için veri seti “sekans” kolonuna göre gruplandırılmış ve her grup içerisinde ilk örnek seçilmiştir.

Modelin performansını deęerlendirebilmek iin, veri setinden rastgele seilmiř 10.000 rnek zerinde alıřılmıřtır. Bu seimden sonra, AAC ve DPC zellikleri, uygun řekilde iřlenmiř ve modelin girdisi olarak kullanılmak zere hazırlanmıřtır.

zellikler leklendirilerek model bunların zerinde tahmin yapmak iin kullanılmıřtır. Elde edilen tahminler, dizilim bilgileri ile birleřtirilerek kapsamlı bir veri erevesi oluřturulmuřtur.

Elde edilen veri erevesinde peptit dizilimlerini ieren “sekans” stununda 'X' karakteri bulunduran satırlar filtrelenmiřtir. AMP veri tabanlarında incelenen peptit sekansları ierisinde, sıklıkla tanımlanamayan (kukla) amino asitlere rastlanmaktadır. *İn-vitro* ve *in-vivo* yntemler kullanılarak kanser aktivitesinin arařtırılabileceęi gelecek alıřmalarda, bu tr belirsizlik ieren amino asitlerin varlıęı, peptitlerin elde edilmesi veya sentezlenmesinde zorluklara yol aabileceęinden, bu peptitlerin veri setinden ıkarılmaları tercih edilmiřtir. Bu yaklařım, gelecekteki arařtırmaların ve alıřmaların uygulanabilirlięini saęlamak iin nem tařımaktadır.

Antikanser etkileri tespit edilen peptitlerin amino asit sekansları fizikokimyasal zelliklerinin tahminlenebilmesi iin “csv” formatında kaydedilmektedir.

3.11. Fizikokimyasal zelliklerin Tahmini

Bu kısımda, peptitlerin fizikokimyasal zelliklerinin tahmin edilmesi iin PepFun aracından yararlanılmıřtır. Bu metot, peptitlerden geliřtirilebilecek potansiyel ila formlasyonları iin ngrlerde bulunmaya olanak saęlamaktadır.

Metodun uygulanmasında, nce antikanser zellięe sahip olduęu tespit edilen peptitlere ait veri setinden amino asit sekansları alınarak PepFun ile eřitli fizikokimyasal zellikleri hesaplanmıřtır. Her bir peptit iin net yk, kararsızlık indeksi, molekl aęırlıęı, logP deęeri ve ortalama hidrofobisite zellikleri tespit edilmiř ve bu deęerler, daha sonra analiz edilmek zere bir “csv” dosyasına kaydedilmiřtir. Ayrıca, PepFun ile alınan sonularda sayısal deęerler ieren ancak metin formatında kaydedilmiř veriler (zellikle negatif sayılar iin) mevcuttur. Bu deęerlerin dzgn bir řekilde iřlenebilmesi iin stunlardaki sayısal olmayan deęerler gerek sayılara dnřtrlmřtir.

3.12. Peptitlerin Üç Boyutlu Modellerinin Oluşturulması

Bu çalışmada, antikanser özelliği gösterdiği tespit edilen ve uygun fizikokimyasal özelliklere sahip örnek peptitlerin üç boyutlu yapısal modelleri oluşturulmuştur. Bu süreçte, SwissModel web sunucusu kullanılarak peptitlerin yapıları modellenmiştir. SwissModel, protein ve peptit yapılarını otomatik olarak modellemek için kullanılan, yaygın bir biyoinformatik araçtır. Bu araç, proteinlerin ve peptitlerin amino asit dizilimleri temelinde yapısal modeller oluşturmak için tasarlanmıştır (Waterhouse ve ark., 2018).

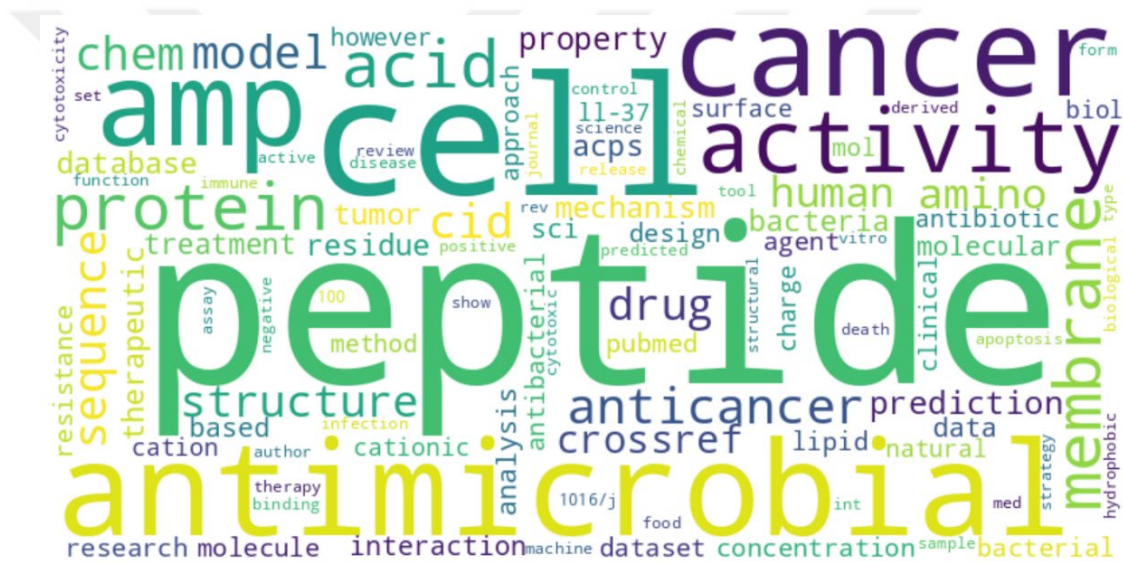
SwissModel'in veri tabanında, incelenen peptitlerin yüzde yüz sekans benzerliğine sahip yapıları mevcut ise, bu yapılar doğrudan kullanılarak zaman ve kaynak tasarrufu sağlanmıştır. Tam benzerlik gösteren yapıların bulunmadığı durumlarda, SwissModel'in homoloji modelleme tekniğini kullanarak sekanslar modellenmiştir. Homoloji modelleme, bilinen protein yapılarından yararlanarak benzer sekanslara sahip proteinlerin üç boyutlu yapılarını tahmin etmekte kullanılan bir yöntemdir (Waterhouse ve ark., 2018).

SwissModel veri tabanında yüzde yüz yapısal benzerliğe sahip peptitler bulunmadığında, Global Model Kalite Tahmini (GMQE) kullanılarak en uygun model seçimi yapılmıştır. Bu tahmin, hedef ve şablon arasındaki hizalamanın özelliklerini ve şablon yapısının niteliklerini birleştirir. Çok katmanlı bir algılayıcı, oluşturulacak modelin yerel mesafe fark testi puanını tahmin etmek üzere eğitilmiştir. GMQE, model oluşturulmadan önce mevcut olup, modelleme sorunları için en uygun şablonların seçiminde yardımcı olur. Model oluşturulduktan sonra, elde edilen modelin global puanını da dikkate alarak GMQE güncellenir, böylece kalite tahmininin güvenilirliği artırılır. 0 ile 1 arasında bir değer almaktadır; 0,7'nin üzerindeki değerler sonucun güvenilir olabileceğini göstermektedir (Waterhouse ve ark., 2018).

4. BULGULAR

4.1. Derlenen Literatüre Ait Kelime Analizi

Bu çalışmada, literatür taraması sonuçlarının analizi, öncelikle tarih bilgileri içeren kelimelerin çıkarılması ile başlamıştır. Ardından, metinlerde en sık geçen kelimeler temel alınarak bir kelime bulutu oluşturulmuştur (Şekil 4.1). Bu kelime bulutu, çalışmanın kapsadığı konuların ve terimlerin görsel bir özetini sunmakta ve araştırmanın anahtar kavramlarını vurgulamaktadır.

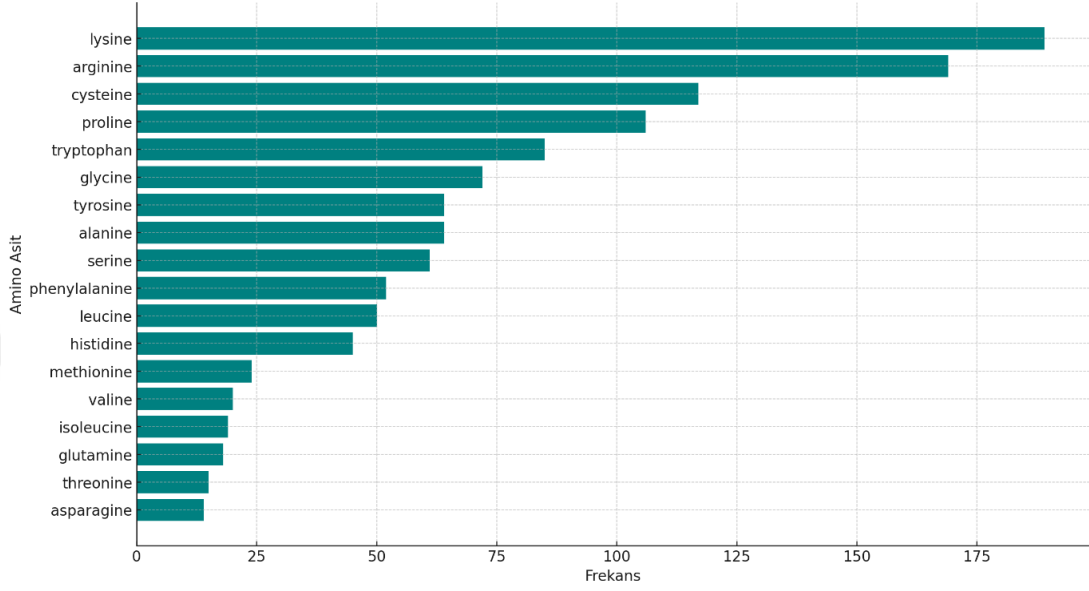


Şekil 4.1. En sık alıntılanan kelimelere ait kelime bulutu.

Kelime bulutu incelendiğinde peptitlerin elde edildikleri kaynakların, yapısal ve terapötik özelliklerinin ön plana çıktığı görülmektedir. Veri tabanlarının birçoğu bu özelliklere göre yapılandırılmaktadır. Ayrıca, insan antimikrobiyal peptidi olan “LL-37” sıklıkla alıntılanmıştır. Geniş spektrumlu antibakteriyel ve anti-biyofilm aktivitelerinin yanı sıra immünomodülatör işlevler sergilediği için konvansiyonel antibiyotiklere alternatif bir seçenek olarak araştırılmaktadır (Ridyard ve Overhage, 2021).

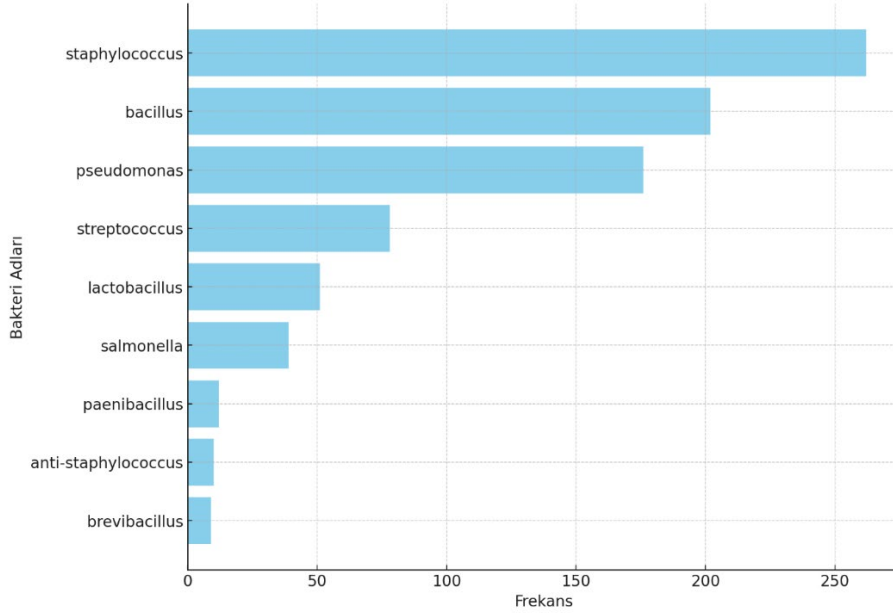
İncelenen literatürde yer alan peptitlerin amino asit kompozisyonları da görselleştirilerek, bu peptitlerin en çok hangi amino asitlerden oluştukları değerlendirilmiştir. Bu analizde, üçten az sayıda geçen peptitler göz ardı edilmiştir

(Şekil 4.2). Lizin (K), Arjinin (R), Sistein (C) ve Prolin (P) amino asitlerinin, antimikrobiyal peptitler ile ilgili literatürde sık sık alıntıldıkları görülmektedir.



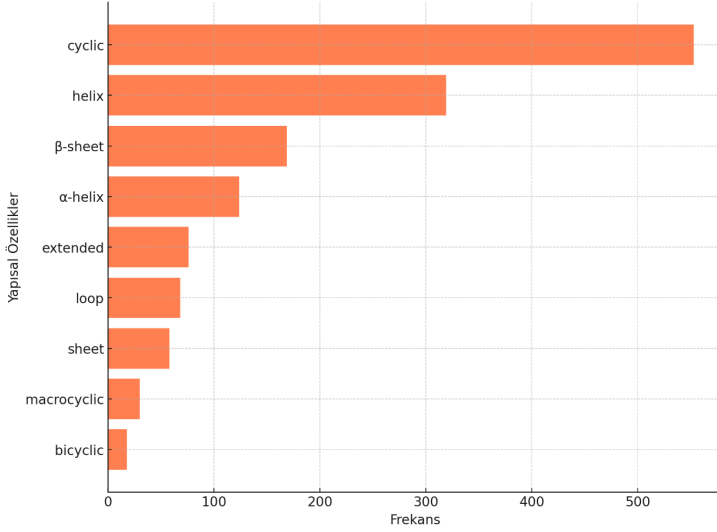
Şekil 4.2. Literatürde yer alan amino asitler ve alıntılanma sayıları.

Literatürde sıkça alıntılanan bakteri adları da benzer bir görselleştirme sürecinden geçirilmiş, bu sayede çalışmalarda ön plana çıkan bakteri türleri ortaya konulmuştur (Şekil 4.3).



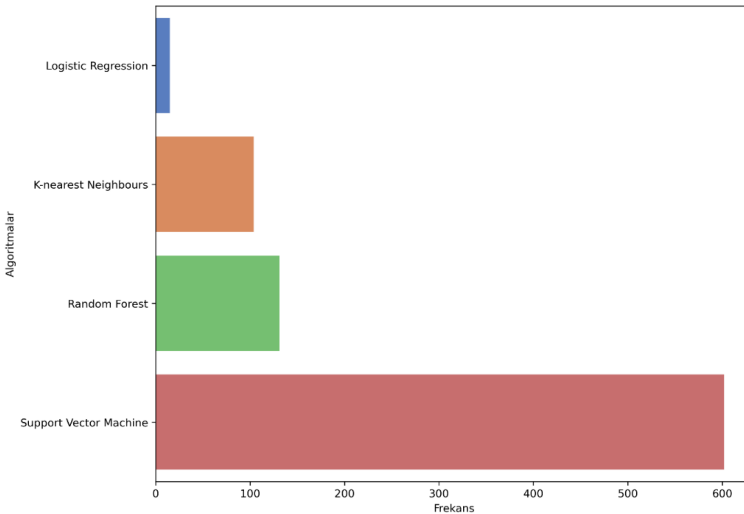
Şekil 4.3. Bakteriler ve alıntılanma sayıları.

Peptitlerin yapısal özellikleriyle ilgili bilgiler de incelenmiş ve bu yapısal özelliklerin alıntılanma sayılarına dayalı bir görselleştirme yapılmıştır (Şekil 4.4). Benzer çalışmalarda araştırılan peptitlerin sıklıkla siklik, sarmal ve beta-yaprak yapıda oldukları görülmektedir.



Şekil 4.4. Yapısal özellikler ve alıntılanma sayıları.

Makine öğrenmesi metodolojilerini kullanan çalışmalarda tercih edilen modeller de incelenmiş ve en sık kullanılan ilk dört model, görsel olarak ifade edilmiştir (Şekil 4.5). Peptitlerin etkinlik ve yapısal özelliklerini inceleyen birçok çalışmada SVM ve RF makine öğrenmesi algoritmaları kullanılmıştır.



Şekil 4.5. Makine öğrenimi modellerinin alıntılanma sayıları.

Bu sistematik ve stratejik literatür taraması yaklaşımı, antimikrobiyal peptitlerin kanserle mücadeledeki rolü ve *in-silico* yöntemlerin bu alanda sağladığı katkılar hakkında kapsamlı ve güncel bir bilgi tabanı oluşturmuştur.

4.2. En Başarılı Makine Öğrenimi Algoritması ile Özellik Eşleşmesi

AAC/DPC ve OHE özellik setlerinin farklı makine öğrenmesi algoritmalarıyla performansları, başarı metrikleri ile birlikte Tablo 4.1’te özetlenmiştir.

Tablo 4.1. Farklı özellik ve algoritma eşleşmelerine ait başarı metrikleri.

Algoritma	Özellik	Doğruluk	Duyarlılık	Özgüllük	Kesinlik	F1 Skoru	ROC AUC
RF	OHE	%76,0	%77,8	%73,9	%77,8	%77,8	%88,1
	AAC & DPC	%81,0	%75,9	%87,0	%87,2	%81,2	%92,4
SVM	OHE	%79,0	%85,2	%71,7	%78,0	%81,4	%86,4
	AAC & DPC	%79,0	%79,6	%78,3	%81,1	%80,4	%91,0
LR	OHE	%78,0	%81,5	%73,9	%78,6	%80,0	%86,7
	AAC & DPC	%80,0	%77,8	%82,6	%84,0	%80,8	%85,8
k-NN	OHE	%63,0	%75,9	%47,8	%63,1	%68,9	%72,0
	AAC & DPC	%65,0	%83,3	%43,5	%63,4	%72,0	%73,2

Analiz sonuçlarına göre, RF algoritması AAC/DPC özellik setiyle kullanıldığında en yüksek doğruluğa (%81,0), özgüllüğe (%87,0) ve kesinliğe (%87,2) ulaşmıştır. Ayrıca, ROC AUC metriğinde %92,4 ile en yüksek performansı göstermiştir. OHE özelliği ise,

SVM’de %79,0 doğruluk ve %86,4 ROC AUC değeriyle yakın bir performans sergilemiştir.

Tüm makine öğrenmesi algoritmalarında doğruluk metriği incelendiğinde; OHE’nin ortalaması %74,0 ve AAC/DPC ortalamasının %76,3 çıktığı sonuçlar elde edilmiştir. Diğer taraftan, k-NN algoritması, AAC/DPC ile en düşük performansa sahiptir; %65,0 gibi görece düşük bir doğruluk oranı sergilemiştir. Aksine, duyarlılık değeri %75,9 ile diğer modellere yakındır. Bu, modelin pozitif sınıfları doğru bir şekilde tanıma konusunda yüksek bir yeteneğe sahip olduğunu göstermektedir. Ancak, modelin genel başarısı, düşük özgüllük (%47,8) ve kesinlik (%63,1) nedeniyle düşük çıkmıştır.

Sonuç olarak, bu çalışmada elde edilen sonuçlar, AAC/DPC özellik setinin RF algoritmasıyla peptitlerin antikanser etkinliklerinin tahmininde en yüksek performansı sağladığını göstermektedir.

4.3. RF Modeli için İdeal Hiperparametreler

Uygulanan hiperparametre optimizasyonu sürecinde, modelin en üstün performansı gösterdiği parametreler karakterize edilmiştir. Önyükleme, OOB skoru ve sıcak başlatma için “True” değeri belirlenmiştir. Karar ağacı bölünme kriteri olarak “entropy” kullanılmıştır. Maksimum derinlik “44”, maksimum özellik sayısı “log2”, minimum yaprak örneği “1”, minimum örnek bölme sayısı için “14” ve karar ağacı sayısı için “115” olarak belirlenmiştir.

Bu parametre kombinasyonu ile elde edilen doğruluk değeri %87,5 (öncesi %81,0) olarak kaydedilmiştir. Bu ayarlamalar, modelin başarısını artırmak için kritik önem taşımaktadır.

4.4. Çalışma Modelinin Ana Veri Seti ile Performansı

Çalışma modeli, RF sınıflandırıcısı ile ana veri seti kullanılarak ve optimize edilmiş hiperparametrelerle eğitilmiştir. Tablo 4.2’de sunulan başarı metrikleri, modelin eğitim ve test veri setleri üzerindeki performansını göstermektedir. Bu metrikler, modelin tahmin doğruluğunu ve güvenilirliğini değerlendirmek için 10-katlı çapraz doğrulama yöntemiyle elde edilmiştir.

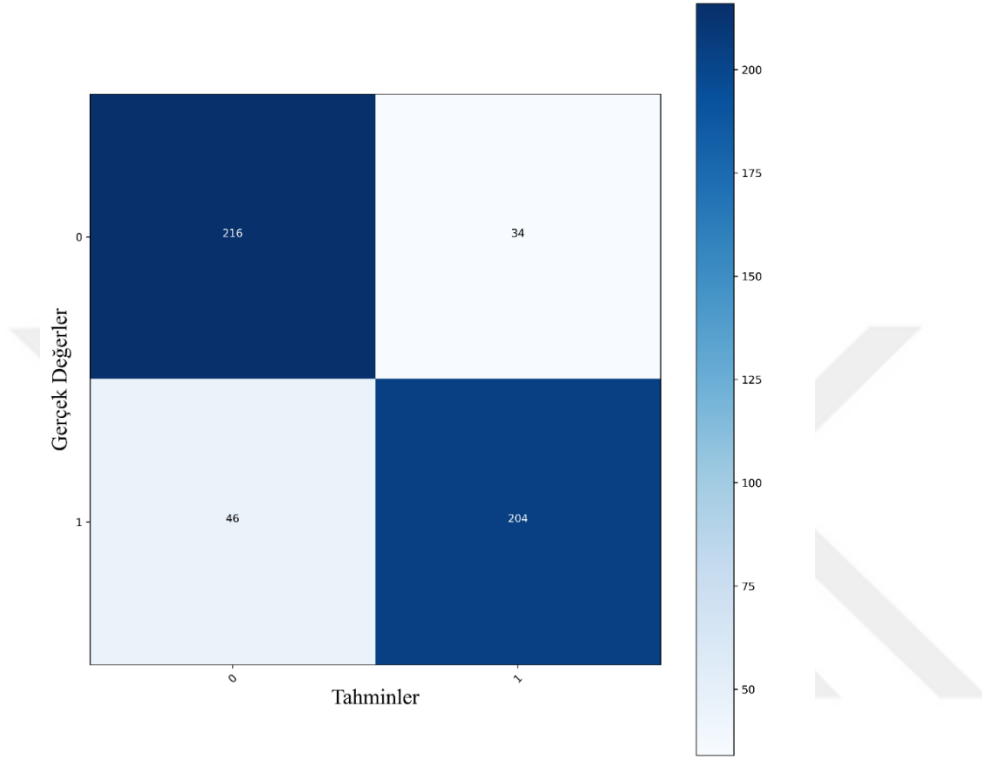
Tablo 4.2. Çalışma modelinin ana veri seti ile başarı metrikleri.

Yineleme Sırası	Doğruluk	Kesinlik	Duyarlılık	Özgüllük	F1 Skoru	ROC AUC
1	%78,0	%81,8	%72,0	%84,0	%76,6	%86,4
2	%82,0	%86,4	%76,0	%88,0	%80,9	%93,4
3	%84,0	%84,0	%84,0	%84,0	%84,0	%93,0
4	%82,0	%83,3	%80,0	%84,0	%81,6	%88,2
5	%84,0	%84,0	%84,0	%84,0	%84,0	%89,3
6	%88,0	%82,8	%96,0	%80,0	%88,9	%92,8
7	%84,0	%90,5	%76,0	%92,0	%82,6	%90,2
8	%92,0	%95,7	%88,0	%96,0	%91,7	%95,5
9	%84,0	%84,0	%84,0	%84,0	%84,0	%91,0
10	%82,0	%86,4	%76,0	%88,0	%80,9	%93,9
Ortalama	%84,0	%85,9	%81,6	%86,4	%83,5	%91,4
SD	0,038	0,042	0,071	0,047	0,042	0,029
SE	0,012	0,013	0,022	0,015	0,013	0,009
p-değeri	0,175	0,021	0,463	0,108	0,478	0,866

* Shapiro-Wilk testine ait değerdir.

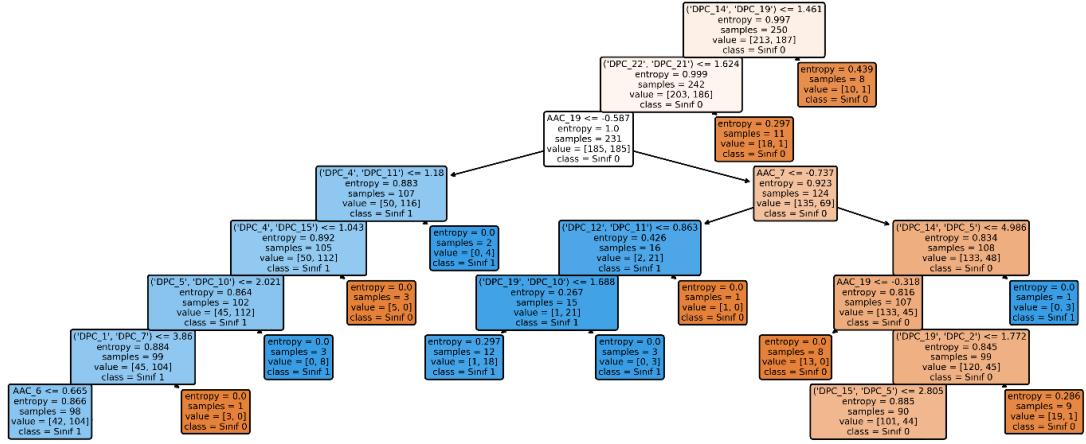
Model, doğruluk, kesinlik, duyarlılık, özgüllük, F1 skoru ve ROC AUC değerleri açısından tutarlı ve yüksek performans göstermiştir. Standart sapma (SD) ve standart hata (SE) değerlerinin düşük çıkması, bu da modelin çeşitli yinelemeler arasında tutarlı sonuçlar verdiğini gösterir niteliktedir. Özellikle ROC AUC değerinin yüksek olması modelin sınıflandırma performansının güçlü olduğunu belirtir. Bu sonuçlar, modelin

antikanser etkinliđi tanıma yeteneđinin genel olarak güvenilir ve tekrarlanabilir olduđunu göstermektedir.



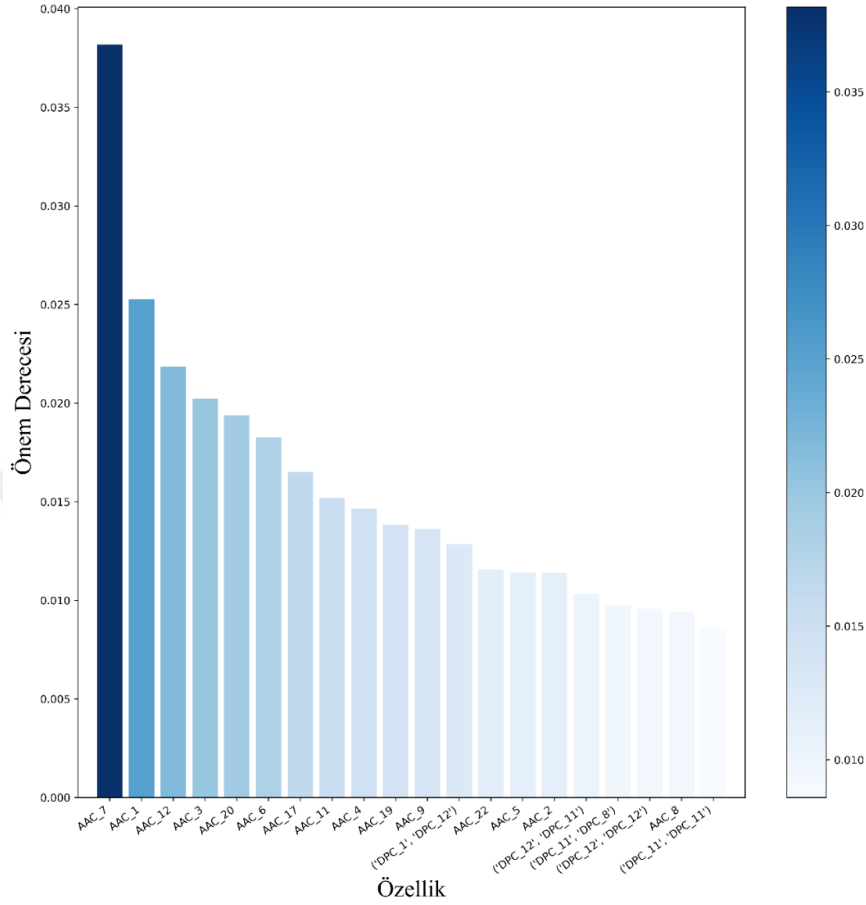
Şekil 4.6. Ana veri seti hata matrisi.

Hata matrisinin görselleştirilmesi için bir ısı haritası grafiđi kullanılmıştır (Şekil 4.6). Hata matrisinde, açık mavi tonları daha düşük deđerleri, koyu tonlar ise yüksek deđerleri temsil etmektedir. Bu görselleştirme ile dođru ve yanlış tahminlerin dađılımını rahatlıkla incelenebilmektedir. Modelin GN'leri, GP'lere göre daha dođru tespit ettiđi görölmektedir. Ayrıca, bu durum YP'lere göre gerçekte pozitif olan örneklerin daha çok negatif olarak sınıflandırıldıklarını göstermektedir.



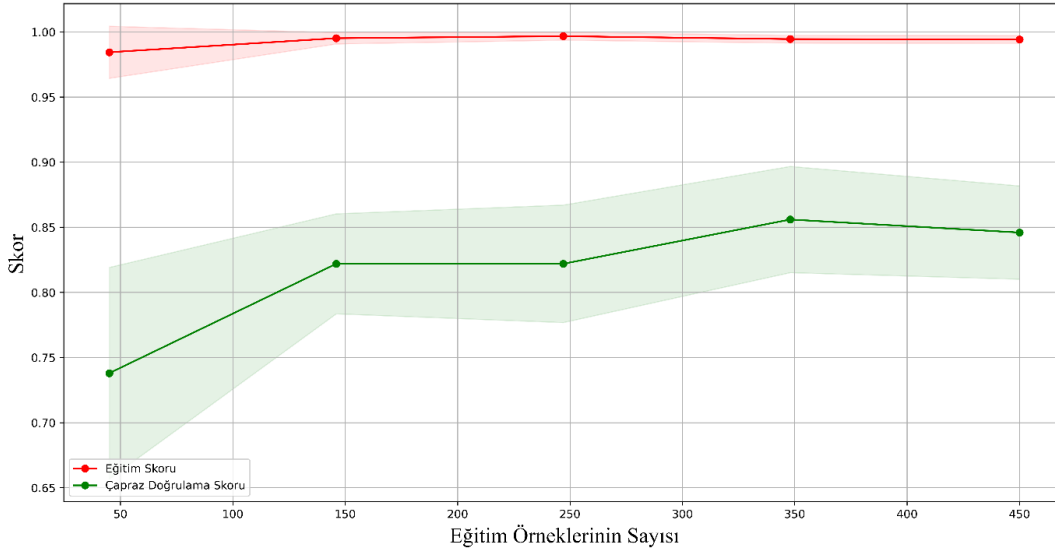
Şekil 4.7. Ana veri seti ile eğitilen çalışma modeline ait bir karar ağacı dalı.

Şekil 4.7’de karar ağacı modelinin dallarından birisinin belli bir bölümü incelenebilmektedir. Düğümler, belirli öz niteliklere göre veri setinin nasıl bölündüğünü temsil etmektedir. Kök düğüm en üsttedir ve bütün kararların başlangıcı olarak kabul edilir. Her bir iç düğüm, verinin daha homojen alt kümelerine ayrılmasını ve yaprak düğümlerine varılmasını sağlar. Yaprak düğümler ise modelin tahminini verir ve bu düğümlerdeki sınıf dağılımı, renk tonları ile ifade edilir. Doygun mavi tonlar, birinci sınıfın hakim olduğu, turuncu tonlar ise sıfıncı sınıfın baskın olduğu düğümleri gösterir. “entropy” katsayısı, düğümün saflık seviyesini gösterir; düşük “entropy” değeri, yüksek saflığı ifade eder. Görseldeki her bir yol, modelin karar verme sürecinde izlediği kuralları yansıtır.



Şekil 4.8. Ana veri seti ile eğitilen çalışma modele ait en etkin 20 özelliğin önem derecesi.

Modele ait özelliklerin önem dereceleri incelendiğinde toplam 702 özellik içerisinde, 286'sının hiçbir etki derecesine sahip olmadığı görülmektedir. Bu özelliklerin, modelin tahminleri üzerinde fark edilir bir etkisinin olmadığı anlaşılmaktadır. Ayrıca, en etkili özellikler incelendiğinde ilk sırada yüzde 3,82 etki derecesiyle "AAC_7" özelliği yer almaktadır. Bu, "AAC_7" özelliğinin, modelin karar verme süreçlerinde diğer özelliklere kıyasla daha baskın bir etkiye sahip olduğunu işaret etmektedir. Şekil 4.8'de, azalan sıralama ile devam eden diğer özelliklerin önem dereceleri de gösterilmekte ve bu dağılım, modelin hangi özelliklerin sonucu daha fazla etkilediğini anlamada kullanışlı bir referans noktası sunmaktadır.



Şekil 4.9. Eğitilen modelin ana veri seti ile öğrenme eğrisi.

Son olarak, öğrenme eğrisi grafiği oluşturulmuştur. Eğitim skorları, modelin eğitim verileri üzerinde ne kadar iyi performans gösterdiğini, çapraz doğrulama skorları ise genelleme yeteneğini göstermektedir.

Eğitim skoru, eğitim örneklerinin sayısının artmasıyla sabit kalmakta ve 1 değerine oldukça yakın bir düzeyde seyretmektedir (Şekil 4.9). Bu, modelin eğitim verilerini mükemmel bir şekilde öğrendiğini ve ezberlediğini gösterir. Ancak, modelin eğitim veri setine aşırı uyma yapabildiğini de işaret eder. Çapraz doğrulama skoru ise, eğitim örneklerinin sayısı arttıkça iyileşme olduğunu göstermektedir. Bu durum, modelin genelleştirme kabiliyetinin, daha fazla veri ile eğitildikçe arttığını ve aşırı uyma riskinin azaltıldığını işaret eder.

4.5. Bağımsız Veri Seti ile Validasyon Sonuçları

Bu bölümde, çalışma modelinin bağımsız veri seti üzerindeki performansı ele alınmaktadır. Tablo 4.3, modelin bağımsız veri seti ile gerçekleştirdiği 10-katlı çapraz validasyon sonuçlarını göstermektedir. Bu analiz, modelin genelleme kabiliyetini ve gerçek dünya verileri üzerindeki etkinliğini değerlendirmektedir. Modelin sadece eğitim verilerindeki örüntüleri ezberlemekten öteye geçip geçmediğini ve yeni, karşılaşmadığı verilere karşı nasıl bir performans gösterdiğini anlamak için bu analizler kritik rol oynamaktadır.

Tablo 4.3. Çalışma modelinin bağımsız veri seti ile başarı metrikleri.

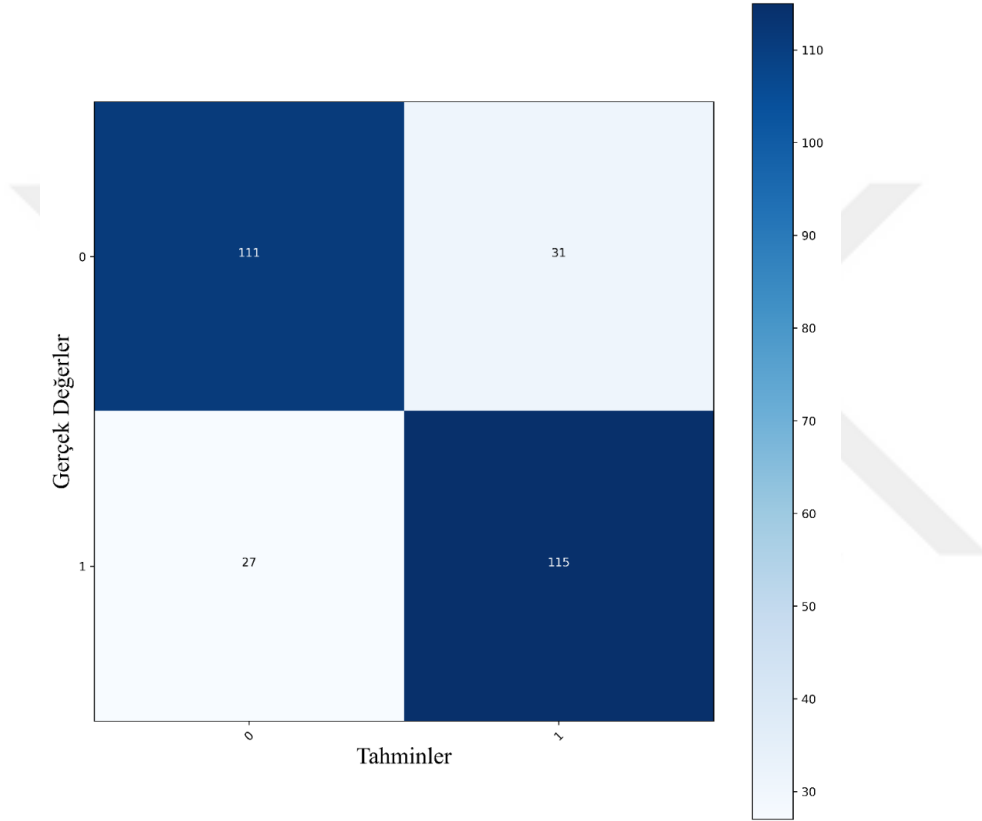
Yineleme Sırası	Doğruluk	Kesinlik	Duyarlılık	Özgüllük	F1 Skoru	ROC AUC
1	%86,2	%85,7	%85,7	%86,7	%85,7	%91,9
2	%86,2	%81,3	%92,9	%80,0	%86,7	%93,8
3	%86,2	%82,4	%93,3	%78,6	%87,5	%91,4
4	%79,3	%80,0	%80,0	%78,6	%80,0	%92,9
5	%82,1	%80,0	%85,7	%78,6	%82,8	%96,4
6	%64,3	%64,3	%64,3	%64,3	%64,3	%77,6
7	%78,6	%75,0	%85,7	%71,4	%80,0	%85,2
8	%89,3	%92,3	%85,7	%92,9	%88,9	%96,4
9	%67,9	%69,2	%64,3	%71,4	%66,7	%86,2
10	%75,0	%76,9	%71,4	%78,6	%74,1	%89,8
Ortalama	%79,5	%78,7	%80,9	%78,1	%79,7	%90,2
SD	0,083	0,080	0,107	0,080	0,087	0,058
SE	0,026	0,025	0,034	0,025	0,027	0,018
p-değeri*	0,264	0,929	0,079	0,538	0,142	0,239

* Shapiro-Wilk testine ait değerdir.

Modelin bağımsız veri seti üzerindeki performansı, ortalama değerler üzerinden incelendiğinde %79,5 doğruluk, %78,7 kesinlik, %80,9 duyarlılık, %78,1 özgüllük, %79,7 F1 skoru ve %90,2 ROC AUC olarak gözlemlenmiştir. Bu ortalama değerler, modelin ana veri setiyle gösterdiği performansa yakındır. Ancak, yinelemelerin bir kısmında modelin performansının dalgalandığı gözlemlenmiştir. Özellikle, 6. ve 9.

yinelemelerde gözlenen düşük değerler, modelin her senaryoya aynı tepkiyi vermediğini göstermektedir.

Ana veri setiyle alınan sonuçlarda olduğu gibi SD ve SE'nin düşük olması, modelin genel olarak tutarlı bir performans sergilediğini göstermektedir.



Şekil 4.10. Modelin bağımsız veri setine ait hata matrisi.

Modelin performansının daha kolay anlaşılabilmesi için tahminlerinin gerçek etiketlerle karşılaştırıldığı hata matrisi görselleştirilmiştir (Şekil 4.10). Bu sefer ana veri setinin aksine modelin, bağımsız veri setinde GP'leri, GN'lere göre daha doğru tespit ettiği görülmektedir. Ayrıca, bu durum YP'lerin YN'lerden fazla olduğunu göstermektedir.

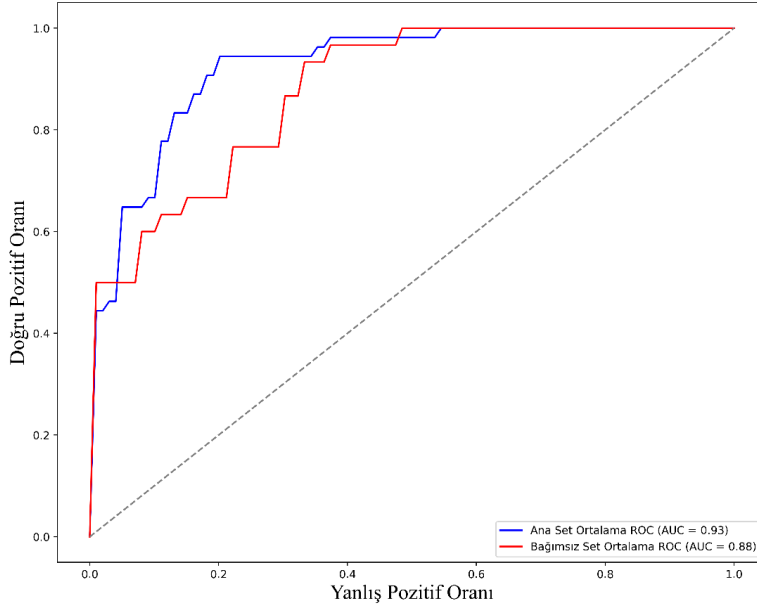
4.5.1. Modelin Ana ve Bağımsız Veri Setleriyle Başarısının Karşılaştırılması

Ana veri setinin sonuçları için gerçekleştirilen Shapiro-Wilk testleri, çoğu metrik için normal dağılıma işaret etmektedir. Doğruluk, Duyarlılık, Özgüllük, F1 Skoru ve ROC AUC metriklerinin p-değerleri, 0,05'ten büyük olduğundan, bu metriklerin normal

dağılıma uygun olduğu kabul edilmektedir. Bununla birlikte, Kesinlik metriği için elde edilen p-değeri (0,021), bu metriğin normal dağılımdan farklı olduğunu göstermektedir. Validasyon sonuçları için yapılan Shapiro-Wilk testleri ise, tüm metriklerin normal dağılıma uygun olduğunu göstermektedir, çünkü tüm p-değerleri 0,05'ten büyük çıkmıştır.

Her iki veri setinin sonuçlarından, t-testi ile doğruluk ve ROC AUC başarı metriklerini tek başlarına analiz ettiğimizde ise p-değerleri doğruluk için 0,138 ve ROC AUC için 0,561 çıkmıştır. İki veri seti arasında, bu metrikler açısından istatistiksel olarak anlamlı bir fark olmadığı görülmektedir. Yani, modelin her iki veri seti üzerindeki performansı arasında, bu metrikler için, önemli bir farklılık bulunmamaktadır.

MANOVA analizi ise, veri seti türünün (ana ve bağımsız veri seti olarak ifade edilen), makine öğrenimi modelinin çeşitli başarı metrikleri (Doğruluk, Kesinlik, Duyarlılık, Özgüllük, F1 Skoru, ROC AUC) üzerindeki etkisini değerlendirmiştir. Bu analizde, veri seti türünün başarı metrikleri üzerindeki etkisi için hesaplanan p-değeri 0,1882 olarak bulunmuştur. Bu değer, genel olarak kabul edilen 0,05 anlamlılık eşiği üzerindedir. Bu sonuç, iki veri arasında başarı metrikleri açısından istatistiksel olarak anlamlı bir fark olmadığını göstermektedir. Dolayısıyla, modelin her iki veri seti üzerindeki performansının benzer olduğu sonucuna varılabilir, bu da modelin genelleştirme kabiliyetinin iyi olduğuna işaret etmektedir.



Şekil 4.11. Modelin her iki veri seti ile hesaplanan ROC AUC grafiği.

Şekil 4.11’de, iki veri seti için ROC eğrilerini ve eğri altındaki alanları karşılaştırmaktadır. Ana veri seti için AUC değeri 0,93 iken, bağımsız veri seti için bu değer 0,88’dir. Her iki eğri de ideal noktaya (1) yakınlık sergilemekte, ancak ana veri setinin eğrisi, bağımsız veri setinin eğrisine göre daha yüksek bir AUC değeri ile daha iyi bir sınıflandırma performansı göstermektedir. Bu, modelin ana veri setinde öğrendiklerini bağımsız veri setinde de görece iyi genelleştirdiğini, fakat ana veri setindeki performansın biraz daha üstün olduğunu göstermektedir.

4.6. Deneysel Veri Setinde Antikanser Özelliği Tespit Edilen Peptitler ve Özellikleri

Geliştirilen RF modeli kullanılarak, çalışma kapsamında incelenen deneysel veri setindeki 10.000 peptitten 6117’sinin antikanser özellik gösterme potansiyeline sahip olduğu belirlenmiştir. Bu peptitlerin detaylı fizikokimyasal özellikleri PepFun aracı kullanılarak hesaplanmıştır. İlk aşamada, kararsızlık indeksi 40’ın üzerinde olan peptitlerin potansiyel kararsızlık gösterme ihtimali göz önünde bulundurularak bu peptitler analiz dışı bırakılmış ve sonuçta 4024 peptit elde edilmiştir. Bu peptitler arasından, molekül ağırlıkları 500 ile 5000 Da arasında değişen ve logP değerleri -18 ile 5 arasında olanlar seçilmiştir. Son filtreleme aşamasında, hidrofobisite değeri 0’ın altında olan peptitler de dikkate alınarak 1648 peptit elde edilmiştir. Antikanser etkinliği

tespit edilen rastgele 15 peptit kararsızlık indeksine göre sıralanmış olup çeşitli fizikokimyasal özellikleri ile birlikte Tablo 4.4'te görülmektedir.

Tablo 4.4. Antikanser etkinliği tespit edilen rastgele seçilmiş 15 antimikrobiyal peptit ve çeşitli özellikleri.

No	Peptit Sekansı	Net Yük (pH 7)	Molekül Ağırlığı (Da)	logP	Hidrofobisite	Kararsızlık İndeksi
seq1	KKHRKHRKHRK HWKKWSKKWK HWIPQCKKFGK K	17,44	4444,46	-10,15	-24,92	8,06
seq2	GWGCNIFGGND YRCHRHCKSIKG YKGGYCKLGGIC KCY	5,92	4247,00	-17,14	-1,70	22,88
seq3	KWKVFKKIEKK WKVFKKIEKAGP KWKVFKKIEK	11,99	4190,28	-1,39	-10,12	20,54
seq4	KFFKKLKKSVKK HVKKFFKKPKVI GVSIPF	12,09	3590,60	-3,42	-3,95	-7,82
seq5	GRFRRLRKKTRK RLKKIGKVLKWI PPIVGSIPLGCG	12,95	4157,25	-11,53	-8,94	27,28
seq6	NGAICWGPCPTA FRQIGNCGHFKV RCCKIR	4,86	3337,03	-13,90	-0,19	-1,33
seq7	RLGNFFRKAKK KIGRGLKKIGQKI KDFLGNLVPRTE S	9,99	4243,14	-15,95	-10,23	-1,15

Tablo 4.4. (Devam) Antikanser etkinliđi tespit edilen rastgele seçilmiş 15 antimikrobiyal peptit ve çeşitli özellikleri*.

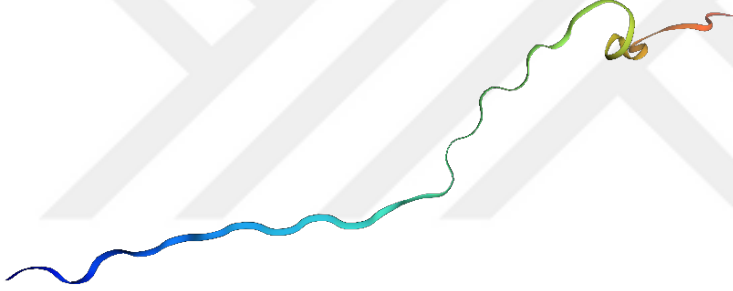
seq8	KWKVFKKIEKM GRNIRNGIVKAG PKWKVFKKIEK	10,99	4114,13	-7,92	-9,02	-36,23
seq9	RWKVFKKIEKV GRNVRDGIKAG PAIGVLGQAKAL G	6,99	3874,70	-12,35	-0,42	28,29
seq10	RICRCRIGRCLGL EYVFGVCFLHGR LARRCCR	6,82	3786,70	-13,49	-5,30	28,01
seq11	GRADYNFGYGL GRGTRKFFNGIG RWVRKTF	6,00	3497,99	-14,20	-4,62	3,04
seq12	GGLKKLGKKLE GAGKRVFNAAE KALPVVAGAKA LRK	8,00	3675,48	-13,60	-2,25	21,75
seq13	GGGGRFKRFRK KFKKLFKKLSPV IPLLHLG	10,10	3454,32	-6,82	-3,51	32,29
seq14	GWKDWAKKAG GWLKKKGPGMA KAALKAAMQ	7,00	3213,93	-8,20	-1,06	14,22
seq15	RCRKGLFCRPKC GQKEHVIGTCPK GLICCR	6,82	3391,23	-14,30	-7,11	11,50

* Daha tutarlı homoloji modelleme yapılabilmesi için sekans uzunluđu 29 üzeri olan peptitler arasından seçilmiştir.

4.7. Üç Boyutlu Peptit Modelleri

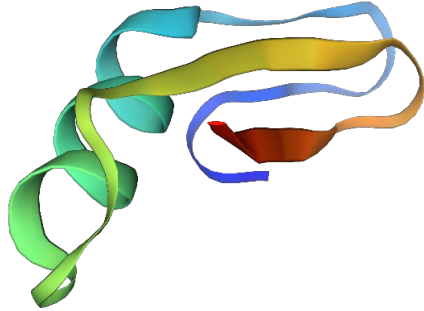
Antikanser etkinliğine sahip olduğu tespit edilen ve fizikokimyasal özellikleri uygun bulunan peptitlerden beşinin üç boyutlu yapısal modelleri oluşturulmuştur. SwissModel web sunucusu kullanılarak gerçekleştirilen bu modelleme işlemi, peptitlerin yapısal özelliklerinin daha detaylı bir şekilde incelenmesini sağlamıştır. Görselleştirilen peptitler GMQE skoru ve sekans benzerliği gözetilerek seçilmiştir.

İlk olarak, “seq1” numaralı peptit SwissModel veri tabanında yapısal analiz için ele alınmıştır (Şekil 4.12). Monomer yapıdaki bu peptit *Naegleria fowleri*'ye ait olduğu varsayılan karakterize edilmemiş bir protein ile %73,08 sekans benzerliği göstermektedir (Varadi ve ark., 2022).



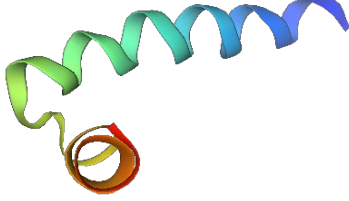
Şekil 4.12. “seq1” numaralı peptite ait üç boyutlu model (GMQE: 0,66).

Şekil 4.13'te incelenebilen monomer yapıdaki “seq2”, *Helicocarpus griseus* kökenli defensin alt birimi içeren bir protein sekansı ile %81,58 benzerliğe sahiptir (Varadi ve ark., 2022).



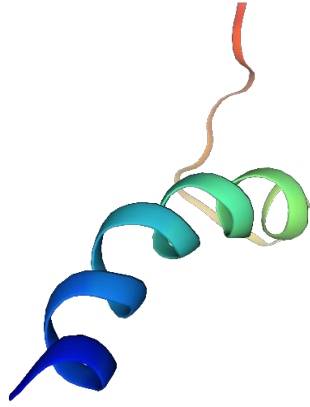
Şekil 4.13. “seq2” numaralı peptite ait üç boyutlu model (GMQE: 0,90).

Hyalophora cecropia'dan elde edilen bir antikanser peptit olan CB1a ile %100 sekans benzerliği tespit edilen “seq3” numaralı peptit monomer yapıdadır ve Şekil 4.14'te görülmektedir (Wu ve ark., 2009; Varadi ve ark., 2022).



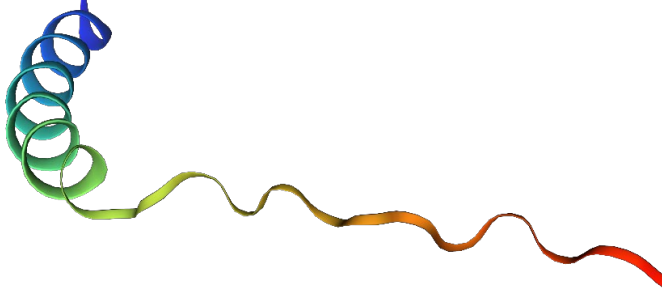
Şekil 4.14. “seq3” numaralı peptite ait üç boyutlu model (GMQE: 0,63).

Diğer bir örnekte, “seq4” numaralı peptidin, Güney Amerika çukur engereği zehrinden elde edilen katelisidin benzeri bir peptit olan “2mwt.1” ile %80 sekans benzerliği gösterdiği tespit edilmiştir. Monomer yapıdaki bu peptidin, hem gram-pozitif hem de gram-negatif bakterilere karşı, ayrıca tümör hücrelerine karşı toksisite gösterdiği belirlenmiştir (Şekil 4.15). Ancak, sağlıklı ökaryotik hücrelere karşı da toksik etkileri gözlemlenmiştir (Falcao ve ark., 2015; Varadi ve ark., 2022).



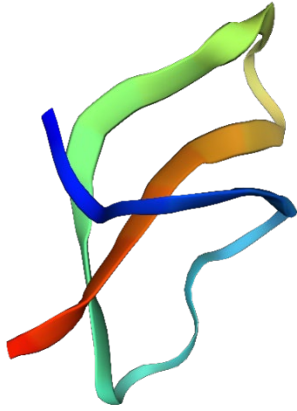
Şekil 4.15. “seq4” numaralı peptite ait üç boyutlu model (GMQE: 0,59).

Yaban domuzu (*Sus scrofa*) kaynaklı bir AMP olan PMAP-36, monomer yapıdaki “seq5” numaralı peptit ile %100 sekans benzerliği göstermektedir. PMAP-36 ve çeşitli analoglarının, in-vitro ve in-vivo modellerde geniş spektrumlu antimikrobiyal aktiviteleri tespit edilmiştir (Zhou J. ve ark., 2019; Varadi ve ark., 2022).



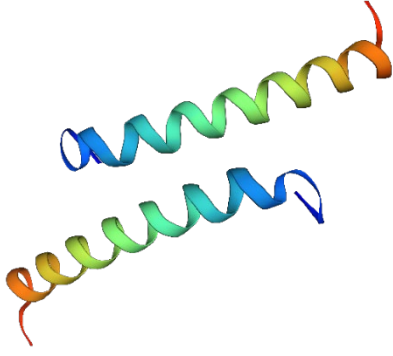
Şekil 4.16. “seq5” numaralı peptite ait ait üç boyutlu model (GMQE: 0,57).

Bir diğer örnek, “seq6” numaralı peptittir (Şekil 4.17). Bu peptit, *Mus musculus* kaynaklı beta-defensin 4 ile yüzde yüz sekans uyumu göstermektedir ve bu yapı monomer formdadır. Peptidin, gram-negatif ve gram-pozitif bakterilere karşı antimikrobiyal aktiviteye sahip olduğu belirlenmiştir. Ayrıca, C-C kemokin reseptörü CCR6 için bir ligand olarak işlev görebileceği ve fare CCR6'ya bağlanarak CCR6 ifade eden hücrelerde kemotaksisi indükleyebileceği tespit edilmiştir, ancak insan CCR6 ile böyle bir etkileşim göstermemiştir. CCR6, gastrointestinal malignitelerin gelişiminde ve metastatik yayılımında görev almaktadır (Rohrl ve ark., 2010; Varadi ve ark., 2022).



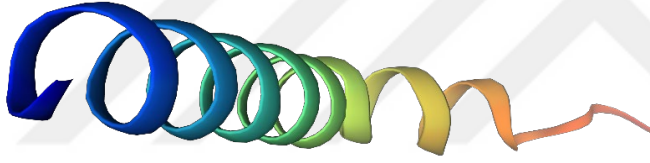
Şekil 4.17. “seq6” numaralı peptite ait ait üç boyutlu model (GMQE: 0,94).

Homo-dimer yapıda bulunan “seq7” numaralı peptit, insan katelisinini olan “LL-37” ile %72,22 sekans benzerliğine sahiptir. Model şekil 4.18’te incelenebilmektedir. Bu katelisininin antimikrobiyal ve antikanser aktivitesini gösteren çalışmalar mevcuttur (Kuroda ve ark., 2015; Niemirowicz ve ark., 2015; Ridyard ve Overhage, 2021; Varadi ve ark., 2022).



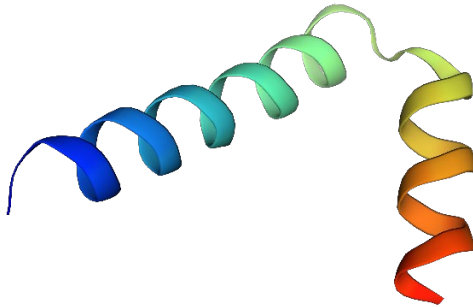
Şekil 4.18. “seq7” numaralı peptite ait üç boyutlu model (GMQE: 0,66).

34 amino asit içeren “seq8” incelendiğinde *Bombyx mori* kökenli bir cecropin B ile %73,53 sekans benzerliğine sahip olduğu görülmektedir. Monomer yapıdaki bu peptit şekil 4.19’da incelenebilmektedir (Varadi ve ark., 2022).



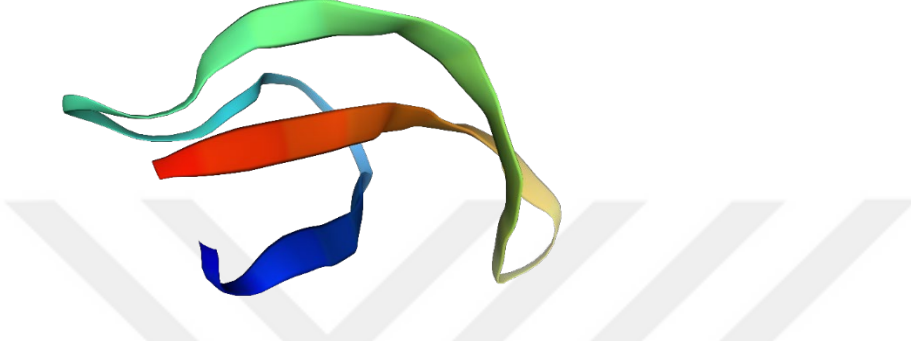
Şekil 4.19. “seq8” numaralı peptite ait üç boyutlu model (GMQE: 0,75).

Papilio xuthus’tan elde edilen cecropin-benzeri bir peptit ile “seq9” numaralı peptit %77,78 sekans benzerliğine sahiptir (Şekil 4.20). Monomer yapıdaki bu peptidin *in-vitro* testler aracılığıyla antimikrobiyal ve antienflamatuvar etkinlikleri gösterilmiştir (Kim ve ark., 2011; Varadi ve ark., 2022).



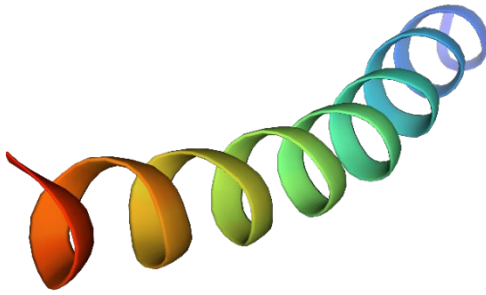
Şekil 4.20. “seq9” numaralı peptite ait üç boyutlu model (GMQE: 0,75).

Monomer yapıdaki “seq10”, *Papio hamadryas* kökenli bir defensin-1 (PHD1-3) ile yüzde 100 sekans benzerliği göstermektedir (Şekil 4.21). PHD1-3 çeşitli gram-negatif ve gram-pozitif bakteriler ile bir mantar türü olan *Candida albicans* karşı antimikrobiyal etkiye sahiptir (Tsvetkova ve ark., 2006; Varadi ve ark., 2022).



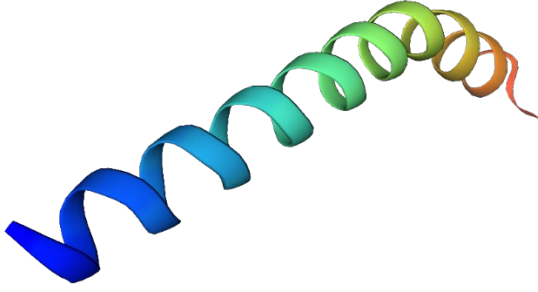
Şekil 4.21. “seq10” numaralı peptite ait üç boyutlu model (GMQE: 0,93).

Monomer yapıdaki “seq11” numaralı peptidin, *Lactobacillus plantarum* kaynaklı bir bakteriosin olan “KL-1Y” ile yüzde yüz sekans benzerliği gösterdiği belirlenmiştir (Şekil 4.22). Bu, peptidin işlevsel formunun tek bir polipeptit zincirinden oluştuğunu göstermektedir. Bir araştırmada, bu peptidin hem gram-pozitif hem de gram-negatif bakterilere karşı geniş spektrumlu inhibitör aktiviteye sahip olduğu keşfedilmiştir (Rumjuankiat ve ark., 2015; Varadi ve ark., 2022).



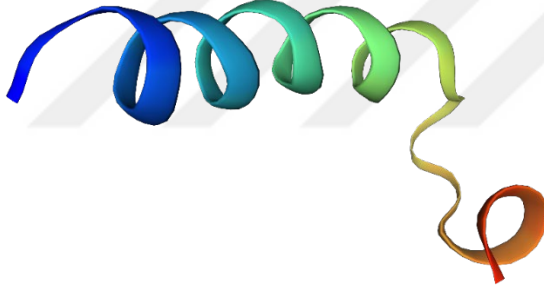
Şekil 4.22. “seq11” numaralı peptite ait üç boyutlu model (GMQE: 0,76).

“seq12” numaralı peptit, *Anopheles darlingi* kökenli bir cecropin A ile yüzde 83,33 benzerlik göstermektedir (Şekil 4.23). Monomer yapıdaki bu peptidin biyolojik fonksiyonlarına dair bir çalışma mevcut değildir (Varadi ve ark., 2022).



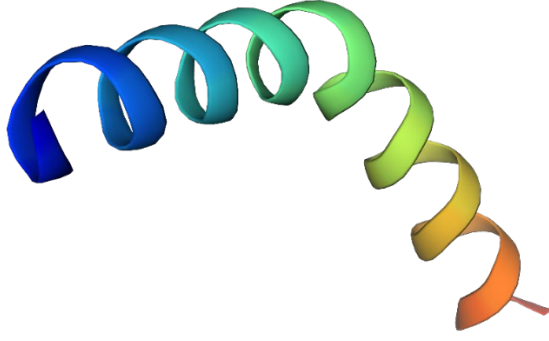
Şekil 4.23. “seq12” numaralı peptite ait ait üç boyutlu model (GMQE: 0,86).

“seq13” numaralı peptit, bir katelisinin türevi olan BMAP-27 ile %96,30 sekans benzerliğine sahiptir (Şekil 4.24). Çeşitli membran etkileşim mekanizmaları ile antimikrobiyal aktivite gösterdiği düşünülmektedir (Yang ve ark., 2019; Varadi ve ark., 2022).



Şekil 4.24. “seq13” numaralı peptite ait ait üç boyutlu model (GMQE: 0,64).

Bir diğer örnekte, “seq14” numaralı peptit incelenmiştir. *Neoponera goeldii* kaynaklı bir peptit toksini (Penoricin) ile %73,33 sekans benzerliği taşımaktadır. Penoricinler geniş spektrumlu antimikrobiyal etki ile birlikte insektisit etki göstermektedirler (Orivel ve ark., 2001; Varadi ve ark., 2022).



Şekil 4.25. “seq14” numaralı peptite ait üç boyutlu model (GMQE: 0,84).

Son olarak, “seq15” numaralı peptit sekansı modellenmiştir. *Alligator mississippiensis* (Amerikan timsahı) kaynaklı beta-defensin 11 ile yüzde yüz sekans benzerliği göstermektedir (Şekil 4.28). Etkinliğine dair bir çalışma literatürde bulunamamıştır (Varadi ve ark., 2022).



Şekil 4.26. “seq15” numaralı peptite ait üç boyutlu model (GMQE: 0,83).

5. TARTIŞMA

Bu çalışma, peptitlerin antikanser etkinliklerinin tahmini için çeşitli makine öğrenimi algoritmalarının performansını değerlendirmiştir. Araştırma konusu olan veri setinin, AAC/DPC ve OHE özellik setleriyle farklı algoritmalarda etkileşimi etkili alternatifler sunmuştur. Bulgular, AAC/DPC özellik setinin farklı algoritmalar ile kombinasyonunun, özellikle doğruluk ve ROC AUC metriklerinde yüksek performans sergilediğini göstermektedir. Yani, AAC/DPC özellik setinin, peptitlerin yapısal ve fonksiyonel özelliklerini iyi yansıttığı söylenebilmektedir. Ayrıca, bu özellik seti en iyi performans RF algoritmasında elde edilmiştir. Bu durum da RF algoritmasının bu özellikleri etkili bir şekilde işleyebildiğini göstermektedir.

Öte yandan, OHE özelliğiyle oluşturulan SVM modeli de dikkate değer bir performans göstermiştir, ancak bu kombinasyonun AAC/DPC ve RF kombinasyonuna göre daha düşük bir doğruluk oranı sunduğu gözlemlenmiştir. Bu durum, OHE'nin peptitlerin antikanser etkinlikleriyle ilgili özellikleri temsil etme yeteneğinin kıyasla daha sınırlı olabileceğine işaret edebilir. Ancak, OHE'nin daha geniş bir özellik düzlemine sahip olduğu ve literatürde sık tercih edilen SVM gibi algoritmalarla iyi sonuçlar verebileceği de göz önünde bulundurulmalıdır (Jung ve Kim, 2023). Ayrıca, OHE'nin vektörize yapısı analiz süresini kısaltmaktadır.

Dipeptit kompozisyonu özelliği, aynı peptit sekansı içerisinde yer alan amino asitler arasındaki etkileşimlerin yakalamasına imkan tanımaktadır. Bu durum, AAC/DPC özellik setinin OHE'ye karşı antikanser etkinliği sınıflandırmada daha başarılı olmasını açıklayabilir.

Her iki özellik setiyle kullanıldığında k-NN algoritmasının daha düşük bir doğruluk oranı sergilemesi, bu özelliklerin k-NN algoritmasının metodolojisine uygun olmayabileceğini gösterir. Düşük özgüllük değeri, modelin negatif sınıfları doğru bir şekilde ayırt etmede zorluklar yaşadığına işaret ederken, görece yüksek duyarlılık değeri pozitif sınıfların doğru bir şekilde tanınmasında başarılı olduğunu göstermektedir. Bu, k-NN algoritmasının peptitlerin antikanser etkinliği üzerindeki bazı özel özelliklerini

yakalayabildiğini, ancak genel bir sınıflandırma başarısı elde etmekte diğer algoritmalara göre zayıf kaldığını göstermektedir.

Veri setinin hazırlık sürecinde benzer homolojiye sahip peptitlerin kümeleme işlemi ile elenmesi, makine öğrenimi modellerinin genelleştirme yeteneğini artırmak ve aşırı uyma karşı önlem almak için kritik bir adımdır. Literatürde genellikle %70 ile %90 arasında eşik değerleri tercih edilmiştir (Tyagi ve ark., 2013; Manavalan ve ark., 2017; Schaduangrat ve ark., 2019). Bunun en büyük nedeni çalışılan veri setlerinde örnek sayısının görece az olmasıdır. Daha düşük eşik değerleri, veri setinin boyutunu önemli ölçüde azaltmakta, bu da daha az örnek ile eğitim yapılmasına ve performansın düşmesine neden olmaktadır.

Bu araştırmada, literatürdeki diğer çalışmalara kıyasla daha büyük bir veri setiyle çalışılmıştır. Bu yüzden, %50 gibi daha düşük bir eşik değerinin kullanılması, aşırı uyma riski ve genelleştirme yeteneği gözetilerek tercih edilmiştir. Bu durum, modelin daha farklı peptit sınıflarını öğrenmesine ve daha karmaşık ilişkileri keşfetmesine imkan tanımış olabilir. RF algoritması, karar ağaçlarındaki çeşitliliği kullanarak veri setindeki geniş varyasyonları yakalayabilmekte ve bu sayede her bir ağaç, veri setinin farklı bir bölümünü öğrenmektedir (Speiser ve ark., 2019). Öte yandan, SVM algoritması, marjinal sınırın etrafında konumlandırılan destek vektörlerine dayalı olarak çalışmakta ve eğitim setindeki benzer peptitlerin sayısının azaltılması, modelin bu kritik destek vektörlerini daha etkili bir şekilde belirlemesine yardımcı olabilir (Cortes ve Vapnik, 1995).

Literatürde yer alan benzer çalışmaların sıklıkla SVM algoritmasını tercih etmesi, bu yöntemin özellikle biyolojik veri setleri üzerindeki yüksek performansına ve genelleştirme yeteneğine atfedilebilir. SVM, karmaşık veri setlerinde bile doğrusal olmayan sınırların belirlenmesi için çekirdek hileleri kullanarak bilgiyi daha yüksek boyutlara taşıma ve optimal bir marj ile sınıflandırma yapma yeteneğine sahiptir. Bu, SVM'in biyolojik verilerdeki gürültülü veya eksik verileri işlerken bile güvenilir sınıflandırma yapabilmesini sağlar (Cortes ve Vapnik, 1995).

Küçük veya eksikli veri setlerinde SVM modellerinin ağaç tabanlı algoritmalara kıyasla daha üstün performans sergilediği farklı araştırmacılar tarafından ifade edilmiştir (Pragati, 2020; Cui ve ark., 2021). Mevcut çalışmanın veri setinin büyüklüğü, antikanser etkinlik tahmininde diğer araştırmalara göre daha fazla veri noktası içermesi ve bu yüzden, RF modelinin daha geniş ve çeşitlendirilmiş veri setinden öğrenerek daha yüksek başarı elde etmiş olması mümkündür. Bu durum, büyük veri setlerinin modelin genelleştirme yeteneğini ve tahmin doğruluğunu artırabileceğini göstermektedir. Yani, düşük eşik değeriyle oluşturulan veri seti, RF ve SVM gibi farklı algoritmaların başarısını farklı şekillerde etkileyebilir. RF algoritması için, daha çeşitli peptitlerle eğitilmiş bir modelin genelleştirme performansı artabilirken, SVM için daha az sayıda ancak daha temsil edici örneklerle daha net bir sınıflandırma sınırı belirlenebilir. Bu durum, her iki algoritmanın da kendi avantajlarına ve veri setinin özelliklerine bağlı olarak optimize edilmesi gerektiğini göstermektedir.

Antikanser peptitlerin sınıflandırılmasında kullanılan makine öğrenimi modelleri çeşitli özellik setlerinden yararlanmaktadır. Literatürde karşımıza çıkan bu modeller, AAC, DPC ve daha gelişmiş özellikler olan K-mer seyrek matris ve pozisyona özel puanlama matrisi (PSSM) gibi özellik setlerini kullanmaktadır (Tyagi ve ark., 2013; Chen X. G. ve ark., 2021). Özellikle, Yinelemeli Özellik Eleme (SVM-RFE) kullanımı, özelliklerin önemini belirleyerek aralarından en iyilerini seçen verimli bir yöntem olarak öne çıkmaktadır (Akbar S. ve ark., 2022; Arif ve ark., 2022).

Modeli geliştirirken makine öğrenmesi algoritması seçimi kadar özellik seçimi de önemlidir. Literatürdeki çalışmalarda yer alan özellik setlerinin çeşitliliği, antikanser peptitleri sınıflandırmada hangi özelliklerin en bilgilendirici olduğu sorusuna cevap arayışını yansıtmaktadır. Örneğin, “AntiCP 2.0” modelinde AAC, DPC ve Kompozisyon, Transisyon, Dağılım (CTD) gibi klasik özelliklerin yanı sıra, model karmaşıklığını artıran OHE gibi özellikler kullanılmıştır. Bu, genellikle peptitlerin fizikokimyasal ve yapısal özelliklerinin geniş bir yelpazesini kapsayarak daha kapsamlı bir model oluşturma çabasını gösterir (Agrawal Piyush ve ark., 2021).

“cACP” çalışmasında Geary otokorelasyon tanımlayıcısı, konjoint triad özellikleri ve quasi-sekans düzeni gibi özelliklerin kullanılması, peptitlerin daha derin yapısal ve

sekans özelliklerini dikkate alan ve dolayısıyla sınıflandırmada potansiyel olarak daha yüksek ayrımcı bilgi sağlayan özelliklere yönelik bir eğilimi işaret edebilir (Akbar Shahid ve ark., 2020; Akbar S. ve ark., 2022).

Benzer çalışmalarda kullanılan farklı özellik setlerinin incelenmesi, peptitlerin antikanser etkinliklerinin anlaşılması ve sınıflandırılması için hangi özelliklerin daha fazla bilgi içerdiğini anlamada önemli bir rol oynar. SVM ve RF gibi güçlü sınıflandırıcılar, bu özellikleri işleyebilecek ve modelin performansını en üst düzeye çıkarabilecek kapasitededir. Bununla birlikte, çalışmada kullanılan özellik setinin de modelin performansını önemli ölçüde etkileyebileceği, ancak algoritma ve veri setinin doğasının da dikkate alınması gerektiği unutulmamalıdır.

RF modelinin antikanser peptit tahminindeki performansını optimize etmek adına gerçekleştirilen hiperparametre ayarlama süreci, modelin başarısını önemli ölçüde artırmıştır. Elde edilen bulgular, önyükleme, OOB skoru ve sıcak başlatma için “True” değerinin seçilmesinin, modelin tutarlılığını ve doğruluğunu iyileştirdiğine işaret etmektedir. Bu seçeneklerin etkinleştirilmesi, her bir ağacın oluşturulması sırasında rastgele örneklem çekilmesini ve model değerlendirmesinin bir önceki iterasyonlarda kullanılmayan örneklerle yapılmasını sağlayarak, modelin aşırı uyum riskini azaltmış ve genel performansını iyileştirmiştir.

Karar ağacı bölünme kriteri olarak “entropy” seçimi, modelin bilgi kazancını maksimize etme yönünde bir yaklaşımı temsil eder. Entropi, sınıflandırma problemlerinde kullanılan bilgi teorisine dayalı bir ölçüttür ve her bölünmede elde edilen kazancı en üst düzeye çıkarmak için kullanılır. Bu, özellikle sınıflandırılacak sınıflar arasındaki belirsizliği azaltmada etkili olmuştur.

Maksimum derinlik değeri olarak “44”ün belirlenmesi, modelin veri setindeki karmaşık yapıları öğrenebilmesi için yeterli esnekliği sağlarken, aşırı derinliğin önlenmesine yardımcı olmuştur. Maksimum özellik sayısı olarak “log2” kullanılması, her bölünmede göz önünde bulundurulacak özelliklerin sayısını optimize ederek modelin hesaplama yükünü dengelerken performansını da arttırmış olabilir. Minimum yaprak örneği olarak “1” ve minimum örnek bölme sayısı için “14” değerlerinin seçilmesi, modelin her bir

yaprağı ve bölünmesini dikkatli bir şekilde düzenlemiş, böylece her karar düğümünün veri setindeki spesifik özelliklere duyarlı olmasını sağlamıştır. Karar ağacı sayısı için belirlenen “115” değeri, modelin karar gücünü artırarak farklı veri alt kümelerinden desteklenmesini sağlamıştır.

Sonuç olarak, bu hiperparametre ayarlamalarının, modelin doğruluğunu %81'den %87,5'e yükselttiği gözlemlenmiştir. Bu gelişme, RF modelinin antikanser peptit tahmininde başarısını artırmanın yanı sıra, hiperparametre ayarlamalarının model performansı üzerindeki belirgin etkisini de teyit etmiştir. Bu optimizasyon süreci, benzer çalışmalarda model ayarlarının ne derece kritik olduğunu ve doğru parametrelerin seçilmesinin, modelin genel performansını nasıl iyileştirebileceğini gösteren bir örnek teşkil eder.

Çalışmada RF modeli, antikanser peptit tahmini için ana veri seti üzerinde, dikkatlice seçilmiş ve optimize edilmiş hiperparametreler kullanılarak eğitilmiştir. Elde edilen başarı metrikleri, modelin yüksek doğruluğunu, kesinliğini, duyarlılığını ve özgüllüğünü ortaya koymuştur. 10-katlı çapraz doğrulama yöntemiyle elde edilen sonuçlar, modelin güvenilirliğini ve tahmin yeteneğini doğrulamakta ve modelin sınıflandırma performansının güçlü olduğunu öne sürmektedir. Özellikle ROC AUC değerinin yüksekliği, modelin sınıflar arası ayrımı başarıyla gerçekleştirdiğini göstermektedir.

Modelin performansını değerlendirirken SD ve SE değerlerinin düşük olması, modelin farklı iterasyonlar arasında kararlı sonuçlar verdiğini gösterir. Hata matrisinin ısı haritası kullanılarak görselleştirilmesi, modelin yanlış pozitifler ve yanlış negatifler arasındaki dengesini gözler önüne sermektedir. Modelin GN'leri GP'lere göre daha doğru tespit ettiği görülmüştür. Bu, modelin negatif sınıfın tahmininde daha hassas olduğunu ve veri setinde belirli bir sınıfa ait örneklerin diğerine göre daha güvenilir olabileceğini düşündürmektedir.

En etkin özelliklerin incelenmesi ise, modelin karar verme sürecindeki özellik dağılımını ve bunların etkinliklerini ortaya koyar. Belirli özelliklerin modelin tahminleri üzerinde önemli bir etkiye sahip olduğunu gösterirken, bazı özelliklerin ise hiçbir etki göstermediği belirlenmiştir. Bu durum, ileriki çalışmalarda modeli geliştirirken hangi

özelliklere odaklanılması gerektiği ve hangilerinin dışlanabileceği konusunda yol gösterici olabilir.

Son olarak, modelin öğrenme eğrisi, eğitim ve çapraz doğrulama skorları arasındaki ilişkiyi aydınlatmaktadır. Modelin eğitim skorlarının yüksek ve sabit kalması, modelin eğitim verilerini iyi öğrendiğini gösterirken, çapraz doğrulama skorlarının artışı, genelleştirme yeteneğinin arttığını göstermektedir. Bu eğilim, modelin genelleştirme kabiliyetinin ve tutarlılığının, eğitim veri setinin büyüklüğüne bağlı olarak nasıl değişebileceğini göstermektedir.

Modelin bağımsız veri seti üzerindeki performansının analizi, genelleştirme yeteneği ve yeni verilere adaptasyon kapasitesi açısından önemli içgörüler sunmaktadır. Ortaya çıkan metrikler, özellikle çapraz validasyon sonuçları, modelin eğitim setindeki örüntüleri öğrenme yeteneğinin ötesine geçtiğini ve bilinmeyen verilere karşı iyi bir performans sergilediğini göstermektedir.

Ancak, belli yinelemelerde performans dalgalanmaları gözlenmiş, bu da modelin her türlü veri dağılımına karşı sabit bir tepki vermediğini göstermiştir. Özellikle 6. ve 9. yinelemelerde düşük performans değerleri, modelin bazı durumlarda beklenen doğruluk seviyesini tutturamadığını ve bu senaryolarda modelin duyarlılığının ve özgüllüğünün azaldığını ortaya koymaktadır. Bu durum, modelin belirli senaryolarda, özellikle veri setinin heterojenliği veya sınıflandırılacak özelliklerin karmaşıklığı arttığında, performansının etkilenebileceğini işaret edebilir. SD ve SE değerlerinin düşük olması ise, modelin genel olarak tutarlı sonuçlar verdiğini gösterir. Ayrıca, hata matrisi incelendiğinde GP ve GN tahmin doğruluklarının birbirlerine yakın oldukları gözükmemektedir.

Çalışma modelinin bağımsız veri seti üzerindeki performansı, modelin gerçek dünya verileriyle karşılaştığında da genel olarak güvenilir ve tutarlı sonuçlar verdiğini göstermektedir. Bu, modelin sadece bilinen verilere değil, aynı zamanda yeni ve önceden karşılaşmadığı verilere karşı da etkili bir şekilde genelleme yapabildiğinin bir işaretidir. Bu bulgular, modelin pratik uygulamalarda kullanılabilirliğinin ve antikanser peptit tahmini alanında katkısının altını çizmektedir.

Çalışmada yapılan SW testleri, ana veri seti için belirtilen çoğu başarı metriğinin normal dağılıma uygun olduğunu ortaya koymuştur, bu da modelin tahminlerinin istatistiksel olarak tutarlı olduğunu göstermektedir. Ana veri setinde, kesinlik metriği için elde edilen düşük p-değeri ise bu metriğin normal dağılıma uymadığını göstermektedir. Bağımsız veri seti sonuçları için yapılan SW testleri tüm metriklerin normal dağılıma uygun olduğunu, dolayısıyla modelin yeni verilere karşı da istatistiksel olarak tutarlı bir performans sergilediğini göstermiştir.

T-testi sonuçları, iki veri seti arasında doğruluk ve ROC AUC değerleri bakımından istatistiksel olarak anlamlı bir fark olmadığını belirtmektedir. Bu, modelin eğitim sürecindeki verilerle sınırlı kalmadığını ve genelleştirme yaparken benzer bir başarı seviyesi gösterdiğini kanıtlamaktadır. Bu, modelin farklı veri setlerine uyum sağlama kapasitesinin ve karşılaştığı yeni veri noktalarını işleme yeteneğinin yüksek olduğunu işaret eden bir göstergedir.

MANOVA analizinin sonuçları, veri seti türünün modelin çeşitli başarı metrikleri üzerinde anlamlı bir etkisi olmadığını göstermiştir. Bu, modelin hem ana hem de bağımsız veri setleri üzerinde benzer bir performans sergilediğini, yani genelleştirme yeteneğinin güçlü olduğunu gösteren bir başka kanıttır. Bu bulgular, modelin antikanser peptit tahmini konusunda güvenilir ve tutarlı sonuçlar üretebildiğini desteklemektedir.

ROC AUC grafiğinin incelenmesi, ana veri setindeki eğitimle elde edilen AUC değerinin, bağımsız veri setiyle karşılaştırıldığında daha yüksek olduğunu ortaya koymuştur. Bu, her iki veri setinde de ideal bir sınıflandırma performansına yakınlık sergilendiğini, ancak ana veri setinin modelin öğrenme kapasitesini daha iyi şekilde genelleştirmesini sağladığını gösterir. Ana veri setinin bağımsız veri setine kıyasla yaklaşık iki kat daha fazla örnek içermesi ve her iki veri setinin de dengeli olması, modelin genelleştirme yeteneğinin ve performansının bu ek örneklerden olumlu yönde etkilenebileceğini düşündürmektedir.

Bu gözlem, geniş ve dengeli bir eğitim setinin, modelin karşılaşacağı yeni verilere karşı daha iyi bir genelleştirme yapmasını sağlayabileceğini gösterir. Modelin eğitim sürecinde daha fazla veriye maruz kalması, algoritmanın daha karmaşık örüntüleri

öğrenmesine ve böylece bağımsız veri setindeki çeşitli durumlarla başa çıkabilmesine imkan tanımış olabilir. Bu durum, geniş bir eğitim setinin, modelin bağımsız veri setindeki performansını güçlendirmede kritik bir rol oynayabileceğini işaret eder.

AMP'lerin antikanser etkinliklerinin daha kapsamlı ve güvenilir bir şekilde değerlendirilmesi için, gelecekteki çalışmalarda, *in-vitro* ve *in-vivo* çalışmalardan elle toplanarak bu tür veri setlerinin hazırlanması daha güvenilir olabilir. Ayrıca, elle toplanan veriler, deneysel koşulların ve peptitlerin biyolojik sistemlerdeki davranışlarının daha gerçekçi bir resmini sağlayarak, AMP'lerin potansiyel terapötik uygulamalarının geliştirilmesine katkıda bulunabilir. Bu nedenle, gelecekteki çalışmalarda, laboratuvar ve klinik çalışmalardan elde edilen verilerin entegrasyonu, AMP araştırmalarında daha güvenilir ve kapsamlı sonuçlar elde etmek için önem taşımaktadır.

Geliştirilen RF modelinin tahminleri, AMP'lerin aynı zamanda antikanser etki gösterebilecekleri terapötik potansiyellerini de vurgulamaktadır. Bu tespit, AMP'lerin ikili mekanizmalarının varlığını güçlendirmekte ve bu peptitlerin kanser tedavisindeki rolünü daha da önemli hale getirmektedir. Antikanser aktivitesi tespit edilen bu peptitlerden rastgele seçilen örneklerin bir kısmının; antikanser etkinlikleri kanıtlanmış peptitlerle benzer homolojiye sahip olması da önemli bir öngörü sunmaktadır. Yalnız, aradaki ilişkinin kanıtlanması için daha büyük örneklem büyüklükleri ve *in-vitro* farmakodinamik çalışmalar ile desteklenmesi gerekmektedir.

Fizikokimyasal özelliklerin dikkatli analizi, peptitlerin ilaç formülasyonları için uygunluğunun değerlendirilmesinde kritik bir öneme sahiptir. Bu çalışmada, PepFun aracılığıyla hesaplanan özelliklerin filtrelenmesi, potansiyel farmasötik uygulamalar için gerekli olan stabilite ve çözünürlük gibi kriterleri karşılayan peptitlerin belirlenmesine olanak tanımıştır. Kararsızlık indeksi, molekül ağırlığı ve logP değerleri gibi parametrelerin belirlenen aralıklarda sınırlandırılması, bu peptitlerin ilaç formülasyonlarında kullanılabilirliğinin bir ön değerlendirmesini sağlar. Bu seçimler, geliştirilen yöntemin başarısına ek olarak *in-vitro/in-vivo* çalışmalar için sonuçların uygulanabilirliğini arttırmaktadır.

Hidrofobisite değeri 0'ın altında olan ve kararsızlık indeksi 40'ın üzerinde olan peptitlerin dışarıda bırakılması, bu peptitlerin formülasyonlarda istenmeyen kararsızlıklar gösterme ihtimalini düşünerek yapılmış bir seçimdir ve bu, daha stabil ve formülasyona uygun peptitlerin seçimine imkan vermektedir. Molekül ağırlıkları ve logP değerleri belirli bir aralıkta olan peptitlerin seçilmesi, ilaç olarak kullanımda karşılaşılabilecek biyolojik bariyerlerle etkileşimi ve farmakokinetik profilini optimize etme ihtiyacını yansıtmaktadır. Özellikle, belirli bir molekül ağırlık aralığında olması, peptitlerin biyolojik etkinliğinin ve hedefe ulaşma kapasitesinin korunmasına katkıda bulunabilirken, logP değerlerinin belirli bir aralıkta olması, hücre içine giriş kabiliyetini ve genel hidrofobik/hidrofilik dengesini etkiler.

Antikanser etkinliğe sahip peptitlerin 3 boyutlu yapısal modellerinin oluşturulması, bu peptitlerin yapısal özellikleri hakkında önemli bilgiler sunmakta ve bu yapıların potansiyel etki mekanizmaları üzerine spekülasyon yapılmasına olanak tanımaktadır. İncelenen peptitlerin SwissModel web sunucusu kullanılarak modellemelerinin yapılması ve ardından elde edilen yapısal verilerin, peptitlerin fonksiyonel özellikleri ile ilişkilendirilmesi, bu moleküllerin kanserle mücadeledeki rolünün daha iyi anlaşılmasına yardımcı olmaktadır.

Örneğin, “KL-1Y” bakteriosini ile yüksek benzerlik gösteren peptidin tek bir polipeptit zincirinden oluşan monomer yapıda olması, bu tip yapıların belirli hücre hedeflerine etkili bir şekilde bağlanabileceğini göstermektedir. Bu durum, hem gram-pozitif hem de gram-negatif bakterilere karşı inhibitör aktivitesinin yanı sıra, kanser hücrelerindeki belirli hedeflere karşı potansiyel etkileri olduğuna işaret edebilmektedir (Rumjuankiat ve ark., 2015).

Mus musculus kaynaklı beta-defensin 4 ile tam uyum göstermesi ve monomer yapıda olması, bu peptitlerin evrimsel olarak korunmuş fonksiyonel motiflere sahip olduğunu ve bu yapıların bakteriyel patojenlere karşı doğal savunma mekanizmalarında rol aldığını göstermektedir. Ayrıca, bu peptitlerin CCR6 gibi spesifik reseptörlerle etkileşime girebileceği ve bu etkileşimin immüno-modülatör etkiler ve kemotaksis yaratabileceği belirtilmiştir (Rohrl ve ark., 2010).

Güney Amerika çukur engereği zehrinden elde edilen katelisidin benzeri bir peptit ile sekans benzerliği gösteren peptidin incelenmesi, bu peptitlerin zehirli organizmalarda evrimsel olarak gelişmiş ve optimize edilmiş etki mekanizmalarına sahip olduğunu göstermektedir. Bu peptitlerin tümör hücrelerine karşı toksik etkileri, kanser hücre hedefleme ve bunların eradikasyonu için yeni stratejiler geliştirmede önemli fırsatlar sunabilir (Du Q. ve ark., 2015).

3-boyutlu yapısal analizler ve homoloji benzerliğinin araştırılması, antikanser peptitlerin molekül yapıları ve bunların biyolojik etkileşimlerindeki rolü hakkında önemli bilgiler sağlamakla kalmayıp, aynı zamanda yeni antikanser ajanların keşfi ve geliştirilmesi için yol gösterici olabilir. Bu bulgular, ileri araştırmalar için bir temel oluştururken, AMP'lerin terapötik potansiyelinin daha geniş bir yelpazede değerlendirilmesine olanak tanımaktadır.

6. SONUÇ ve ÖNERİLER

Sonuç olarak, bu çalışmada elde edilen ana bulgular aşağıdaki şekildedir:

- Bu çalışma, antikanser peptitlerin (AMP'ler) keşfi ve karakterizasyonu alanında önemli katkılarda bulunmuştur. Geliştirilen RF modeli, AMP'ler içerisinde geniş bir antikanser etkinlik potansiyeline sahip peptit grubunu başarıyla tespit etmiştir, bu da AMP'lerin kanserle mücadeledeki çift yönlü potansiyelini destekler niteliktedir.
- Veri setinde benzer homolojiye sahip peptitlerin kümeleme işlemleriyle elenmesi, eğitilen modelin gerçek dünya verilerine karşı daha iyi sonuç almasını; eğitildiği ana veri setine aşırı uyumun önüne geçmesini sağlamaktadır.
- Fizikokimyasal özelliklerin detaylı analizi, peptitlerin ilaç formülasyonları için uygunluğunu değerlendirmede kritik bir rol oynamıştır. Kararsızlık indeksi, molekül ağırlığı ve logP gibi özelliklerin belirli aralıklarda filtrelenmesi, ilaç formülasyonları için potansiyel olarak uygun peptitlerin seçimine imkan tanımıştır.
- Projede, peptitlerin antikanser özelliklerinin yanı sıra antimikrobiyal aktiviteleri de dikkate alınmıştır. Bu yaklaşım, AMP'lerin kanser ve mikrobiyal patojenlere karşı ikili terapötik potansiyellerinin altını çizmiş ve bu alanlarda yeni terapötik stratejilerin geliştirilmesine zemin hazırlamıştır.

Öneriler:

- Veri seti hazırlanırken ana ve bağımsız veri setlerinin ayrı araştırmacılar tarafından hazırlanması yanlışlığı azaltabilir. Ayrıca, veri setlerinin *in-vitro* ve *in-vivo* farmakodinamik araştırmalardan elle seçilerek derlenmesi daha güvenilir ve tutarlı sonuçlar sağlayabilir.
- Literatürde yer alan bazı çalışmalarda olduğu gibi belli sekans uzunluklarına ait peptitler için ayrı modeller eğitilebilir. Bu modelin tanımlı veri setinde daha yüksek başarı elde etmesini sağlayacaktır.

- Özellik seçimi sürecine etkisiz özelliklerin elenmesi de eklenmelidir. Bu modellerin aşırı uyma riskini azaltabilir ve hesaplama maliyetini düşürebilir.
- Yerel canlı florasına ait bir cinse ait peptitler izole edilerek ve antikanser/antimikrobiyal etkileri tespit edilerek bölgesel (Akdeniz bölgesi gibi) bir veri tabanı oluşturulabilir. Ayrıca, eğitilen bu makine öğrenimi modeli, araştırmacıların erişimine bir grafiksel kullanıcı arayüzü aracılığıyla bir web sitesi üzerinden sunulabilir.
- AMP'lerin antikanser özelliklerinin daha fazla anlaşılması için, bu çalışmada tespit edilen peptitler üzerinde ek deneysel (*in-vitro* gibi) çalışmalar yapılması önerilir. Bu, AMP'lerin kanser hücrelerine karşı etki mekanizmalarının daha iyi anlaşılmasına yardımcı olabilir.
- AMP'lerin ilaç formülasyonlarındaki potansiyellerinin daha iyi anlaşılması için, fizikokimyasal özelliklerin yanı sıra, canlı ortamlarda biyolojik aktiviteleri, stabilite ve geçimlilikleri üzerine detaylı çalışmalar yapılmalıdır.

Bu tez çalışması, AMP'lerin antikanser etkinliklerinin ve yapısal özelliklerinin keşfi konusunda katkı sunmuştur. Ayrıca, geliştirilen süreç bu alanda çalışan araştırmacılara yeni perspektifler sunabilir. Bu bulgular, AMP'lerin kanser tedavisindeki rolünün daha iyi anlaşılmasına ve bu peptitlerin terapötik kullanımlarının yaygınlaştırılmasına yardımcı olacaktır.

KAYNAKLAR

- Agrawal P, Bhagat D, Mahalwal M, Sharma N, Raghava GPS. Anticp 2.0: An updated model for predicting anticancer peptides. *Briefings in Bioinformatics*. 2021; 22: bbaa153.
- Agrawal P, Bhalla S, Usmani SS, Singh S, Chaudhary K, Raghava Gajendra PS, Gautam A. Cppsite 2.0: A repository of experimentally validated cell-penetrating peptides. *Nucleic Acids Research*. 2015; 44: D1098-D1103.
- Agrawal P, Raghava GPS. Prediction of antimicrobial potential of a chemically modified peptide from its tertiary structure. *Front Microbiol*. 2018; 9: 2551.
- Agrawal S, Acharya D, Adholeya A, Barrow CJ, Deshmukh SK. Nonribosomal peptides from marine microbes and their antimicrobial and anticancer potential. *Frontiers in Pharmacology*. 2017; 8.
- Akbar S, Hayat M, Tahir M, Khan S, Alarfaj FK. Caccp-deepgram: Classification of anticancer peptides via deep neural network and skip-gram-based word embedding model. *Artif Intell Med*. 2022; 131: 102349.
- Akbar S, Rahman AU, Hayat M, Sohail M. Caccp: Classifying anticancer peptides using discriminative intelligent model via chou's 5-step rules and general pseudo components. *Chemometrics and Intelligent Laboratory Systems*. 2020; 196.
- Allegra CJ, Drake JC, Jolivet J, Chabner BA. Inhibition of phosphoribosylaminoimidazolecarboxamide transformylase by methotrexate and dihydrofolic acid polyglutamates. *Proc Natl Acad Sci U S A*. 1985; 82: 4881-4885.
- Amazon. (2023). Textract (Version 1.6.1): AWS. Retrieved from <https://boto3.amazonaws.com/v1/documentation/api/latest/reference/services/textract.html>

- . Anaconda software distribution (Version 2.5.0). (2023): Anaconda. Retrieved from <https://anaconda.com>
- Arif M, Ahmed S, Ge F, Kabir M, Khan YD, Yu D-J, Thafar M. Stackacpred: Prediction of anticancer peptides by integrating optimized multiple feature descriptors with stacked ensemble approach. *Chemometrics and Intelligent Laboratory Systems*. 2022; 220.
- Arneth B. Tumor microenvironment. *Medicina (Kaunas)*. 2019; 56.
- Aruleba RT, Tincho MB, Pretorius A, Kappo AP. In silico prediction of new antimicrobial peptides and proteins as druggable targets towards alternative anti-schistosomal therapy. *Scientific African*. 2021; 12: e00804.
- Banerjee S, Parasramka MA, Paruthy SB. Chapter 15 - garcinol: Preclinical perspective underpinning chemo- and radiosensitization of cancer. In AC Bharti BB Aggarwal (Eds.), *Role of nutraceuticals in cancer chemosensitization*. Academic Press; 2018, 2, p. 297-324.
- Barrow AD, Martin CJ, Colonna M. The natural cytotoxicity receptors in health and disease. *Front Immunol*. 2019; 10: 909.
- Bassi S. Python for bioinformatics. Chapman and Hall/CRC; 2017.
- Basu AK. DNA damage, mutagenesis and cancer. *Int J Mol Sci*. 2018; 19.
- Bayles B, Rose E. (2021). More-itertools (Version 8.12.0). Retrieved from <https://pypi.org/project/more-itertools/#description>
- Boohaker RJ, Lee MW, Vishnubhotla P, Perez JM, Khaled AR. The use of therapeutic peptides to target and to kill cancer cells. *Curr Med Chem*. 2012; 19: 3794-3804.
- Boopathi V, Subramaniam S, Malik A, Lee G, Manavalan B, Yang D-C. Macppred: A support vector machine-based meta-predictor for identification of anticancer peptides. *International Journal of Molecular Sciences*. 2019; 20: 1964.

- Breiman L. Random forests. *Machine Learning*. 2001; 45: 5-32.
- Brunetti J, Falciani C, Bracci L, Pini A. Branched peptides as bioactive molecules for drug design. *Peptide Science*. 2018; 110: e24089.
- Brunton LL, Knollmann BC, Hilal-Dandan R. Goodman & Gilman's the pharmacological basis of therapeutics. McGraw Hill Medical; 2018.
- Burdukiewicz M, Sidorczuk K, Rafacz D, Pietluch F, Bakała M, Słowik J, Gagat P. Cancergram: An effective classifier for differentiating anticancer from antimicrobial peptides. *Pharmaceutics*. 2020; 12: 1045.
- Chabner BA, Roberts TG. Chemotherapy and the war on cancer. *Nature Reviews Cancer*. 2005; 5: 65-72.
- Chen W, Ding H, Feng P, Lin H, Chou K-C. Iacp: A sequence-based tool for identifying anticancer peptides. *Oncotarget*. 2016; 7: 16895-16909.
- Chen XG, Zhang W, Yang X, Li C, Chen H. Acp-da: Improving the prediction of anticancer peptides using data augmentation. *Front Genet*. 2021; 12: 698477.
- Chu JY, Yung WH, Chow BK. Secretin: A pleiotrophic hormone. *Ann N Y Acad Sci*. 2006; 1070: 27-50.
- Cock PJ, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, et al. Biopython: Freely available python tools for computational molecular biology and bioinformatics. *Bioinformatics*. 2009; 25: 1422-1423.
- Cortes C, Vapnik V. Support-vector networks. *Machine Learning*. 1995; 20: 273-297.
- Csárdi G, Nepusz T, Horvát S, Traag V, Zanini F, Noom D. Igraph. Zenodo; 2023.
- Cui L, Shen J, Yao S. The sparse learning of the support vector machine. *Journal of Physics: Conference Series*. 2021; 2078.

Cullen JM, Cascella M. Physiology, enkephalin. In Statpearls. Treasure Island (FL): StatPearls Publishing

Copyright © 2023, StatPearls Publishing LLC.; 2023.

De Felipe C, Herrero JF, O'Brien JA, Palmer JA, Doyle CA, Smith AJ , ve ark. Altered nociception, analgesia and aggression in mice lacking the receptor for substance p. *Nature*. 1998; 392: 394-397.

Dean SN, Alvarez JAE, Zabetakis D, Walper SA, Malanoski AP. Pepvae: Variational autoencoder framework for antimicrobial peptide generation and activity prediction. *Front Microbiol*. 2021; 12: 725727.

DeBerardinis RJ, Mancuso A, Daikhin E, Nissim I, Yudkoff M, Wehrli S, Thompson CB. Beyond aerobic glycolysis: Transformed cells can engage in glutamine metabolism that exceeds the requirement for protein and nucleotide synthesis. *Proc Natl Acad Sci U S A*. 2007; 104: 19345-19350.

Deng X, Whitcomb DC. Chapter 151 - pancreatic polypeptide. In AJ Kastin (Ed.), *Handbook of biologically active peptides*. Burlington: Academic Press; 2006, p. 1097-1107.

Deslouches B, Di YP. Antimicrobial peptides with selective antitumor mechanisms: Prospect for anticancer applications. *Oncotarget*. 2017; 8: 46635-46651.

dos Santos LV, Souza FH, Brunetto AT, Sasse AD, da Silveira Nogueira Lima JP. Neurokinin-1 receptor antagonists for chemotherapy-induced nausea and vomiting: A systematic review. *JNCI: Journal of the National Cancer Institute*. 2012; 104: 1280-1292.

Du Q, Hou X, Wang L, Zhang Y, Xi X, Wang H , ve ark. Aaeap1 and aaeap2: Novel antimicrobial peptides from the venom of the scorpion, androctonus aeneas: Structural characterisation, molecular cloning of biosynthetic precursor-encoding cdnas and engineering of analogues with enhanced antimicrobial and anticancer activities. *Toxins*. 2015; 7: 219-237.

- Du Toit A. Cell death: Balance through a bivalent regulator. *Nat Rev Mol Cell Biol.* 2013; 14: 546.
- Du Y, Shang BY, Sheng WJ, Zhang SH, Li Y, Miao QF, Zhen YS. A recombinantly tailored beta-defensin that displays intensive macropinocytosis-mediated uptake exerting potent efficacy against k-ras mutant pancreatic cancer. *Oncotarget.* 2016; 7: 58418-58434.
- Duarte-Mata DI, Salinas-Carmona MC. Antimicrobial peptides' immune modulation role in intracellular bacterial infection. *Front Immunol.* 2023; 14: 1119574.
- Edgar RC. Search and clustering orders of magnitude faster than blast. *Bioinformatics.* 2010; 26: 2460-2461.
- Emran TB, Shahriar A, Mahmud AR, Rahman T, Abir MH, Siddiquee MF , ve ark. Multidrug resistance in cancer: Understanding molecular mechanisms, immunoprevention and therapeutic approaches. *Front Oncol.* 2022; 12: 891652.
- Falcao CB, Perez-Peinado C, de la Torre BG, Mayol X, Zamora-Carreras H, Jimenez MA , ve ark. Structural dissection of crotalicidin, a rattlesnake venom cathelicidin, retrieves a fragment with antimicrobial and antitumor activity. *J Med Chem.* 2015; 58: 8553-8563.
- Faraji N, Arab SS, Doustmohammadi A, Daly NL, Khosroushahi AY. Apinapdb: A database of apoptosis-inducing anticancer peptides. *Scientific Reports.* 2022; 12: 21341.
- Febriza A, Hatta M, Natzir R, Kasim VN, Idrus HH. Activity of antimicrobial peptide; cathelicidin, on bacterial infection. *The Open Biochemistry Journal.* 2019; 13.
- Feng C, Zhao B, Zhou X, Ding X, Shan Z. An enhanced quantum k-nearest neighbor classification algorithm based on polar distance. *Entropy (Basel).* 2023; 25.
- Fjell CD, Hiss JA, Hancock REW, Schneider G. Designing antimicrobial peptides: Form follows function. *Nature Reviews Drug Discovery.* 2012; 11: 37-51.

Forbes J, Krishnamurthy K. Biochemistry, peptide. In Statpearls. Treasure Island (FL): StatPearls Publishing

Copyright © 2023, StatPearls Publishing LLC.; 2023.

Frankenberg-Schwager M. Induction, repair and biological relevance of radiation-induced DNA lesions in eukaryotic cells. *Radiat Environ Biophys.* 1990; 29: 273-292.

Frøkjær S, van de Weert M, Hovgaard L. Formulation, stability, and characterization of protein and peptide drugs. In.; 2004.

Fu J, Liao L, Balaji KS, Wei C, Kim J, Peng J. Epigenetic modification and a role for the e3 ligase rnf40 in cancer development and metastasis. *Oncogene.* 2021; 40: 465-474.

Gačić M. (2023, 2023//). *Cost effectiveness analysis of real and in silico clinical trials for stent deployment based on decision tree.* Paper presented at the Applied Artificial Intelligence: Medicine, Biology, Chemistry, Financial, Games, Engineering, Cham.

Gaspar D, Veiga AS, Castanho MA. From antimicrobial to anticancer peptides. A review. *Front Microbiol.* 2013; 4: 294.

Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A. ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res.* 2003; 31: 3784-3788.

Gautam A, Singh H, Tyagi A, Chaudhary K, Kumar R, Kapoor P, Raghava GP. Cppsite: A curated database of cell penetrating peptides. *Database (Oxford).* 2012; 2012: bas015.

Gavalas NG, Lontos M, Trachana SP, Bagratuni T, Arapinis C, Liacos C, et al. Angiogenesis-related pathways in the pathogenesis of ovarian cancer. *Int J Mol Sci.* 2013; 14: 15885-15909.

- Ghosh S, Dasgupta R. Machine learning in biological sciences: Updates and future prospects. Springer Nature Singapore; 2022.
- Gordon YJ, Romanowski EG, McDermott AM. A review of antimicrobial peptides and their therapeutic potential as anti-infective drugs. *Curr Eye Res.* 2005; 30: 505-515.
- Green DR. Cell death: Apoptosis and other means to an end. Cold Spring Harbor Laboratory Press; 2018.
- Grivennikov SI, Greten FR, Karin M. Immunity, inflammation, and cancer. *Cell.* 2010; 140: 883-899.
- Gueguen Y, Garnier J, Robert L, Lefranc MP, Mougenot I, de Lorgeril J, et al. Penbase, the shrimp antimicrobial peptide penaeidin database: Sequence-based classification and recommended nomenclature. *Dev Comp Immunol.* 2006; 30: 283-288.
- Gupta N, Khetan D, Chaudhary R, Shukla JS. Prospective cohort study to assess the effect of storage duration, leuko-filtration, and gamma irradiation on cell-free DNA in red cell components. *Transfus Med Hemother.* 2020; 47: 409-419.
- Guryanova SV, Ovchinnikova TV. Immunomodulatory and allergenic properties of antimicrobial peptides. *Int J Mol Sci.* 2022; 23.
- Hammami R, Ben Hamida J, Vergoten G, Fliss I. Phytamp: A database dedicated to antimicrobial plant peptides. *Nucleic Acids Res.* 2009; 37: D963-968.
- Hammami R, Zouhir A, Ben Hamida J, Fliss I. Bactibase: A new web-accessible database for bacteriocin characterization. *BMC Microbiol.* 2007; 7: 89.
- Hanahan D, Weinberg RA. Hallmarks of cancer: The next generation. *Cell.* 2011; 144: 646-674.

- Hancock REW, Sahl H-G. Antimicrobial and host-defense peptides as new anti-infective therapeutic strategies. *Nature Biotechnology*. 2006; 24: 1551-1557.
- Haney EF, Hancock RBEW. Peptide design for antimicrobial and immunomodulatory applications. *Biopolymers*. 2013; 100: 572-583.
- Haney EF, Mansour SC, Hancock REW. Antimicrobial peptides: An introduction. In PR Hansen (Ed.), *Antimicrobial peptides: Methods and protocols*. New York, NY: Springer; 2017, p. 3-22.
- Harris CR, Millman KJ, van der Walt SJ, Gommers R, Virtanen P, Cournapeau D , ve ark. Array programming with numpy. *Nature*. 2020; 585: 357-362.
- Hoeijmakers JHJ. DNA repair mechanisms. *Maturitas*. 2001; 38: 17-22.
- Huang GT, Zhang HB, Kim D, Liu L, Ganz T. A model for antimicrobial gene therapy: Demonstration of human beta-defensin 2 antimicrobial activities in vivo. *Hum Gene Ther*. 2002; 13: 2017-2025.
- Huang R, Zhou P-K. DNA damage repair: Historical perspectives, mechanistic pathways and clinical translation for targeted cancer therapy. *Signal Transduction and Targeted Therapy*. 2021; 6: 254.
- Hultmark D, ENGSTRÖM Å, BENNICH H, KAPUR R, BOMAN HG. Insect immunity: Isolation and structure of cecropin d and four minor antibacterial components from cecropia pupae. *European journal of biochemistry*. 1982; 127: 207-217.
- Hunter JD. Matplotlib: A 2d graphics environment. *Computing in Science & Engineering*. 2007; 9: 90-95.
- Iommarini L, Porcelli AM, Gasparre G, Kurelac I. Non-canonical mechanisms regulating hypoxia-inducible factor 1 alpha in cancer. *Front Oncol*. 2017; 7: 286.

- IUPAC. Iupac-iub joint commission on biochemical nomenclature (jcbn). Nomenclature and symbolism for amino acids and peptides. Recommendations 1983. Eur J Biochem. 1984; 138: 9-37.
- Jafari A, Babajani A, Sarraimi Forooshani R, Yazdani M, Rezaei-Tavirani M. Clinical applications and anticancer effects of antimicrobial peptides: From bench to bedside. *Frontiers in Oncology*. 2022; 12.
- Joblib. (2023). Joblib: Running python functions as pipeline jobs (Version 1.2.0): Joblib Development Team. Retrieved from <https://joblib.readthedocs.io/>
- Jung T, Kim J. A new support vector machine for categorical features. *Expert Systems with Applications*. 2023; 229.
- . Jupyter notebook (Version 6.5.4). (2023): Jupyter Team. Retrieved from <https://docs.jupyter.org/>
- Kang X, Dong F, Shi C, Liu S, Sun J, Chen J , ve ark. Dramp 2.0, an updated data repository of antimicrobial peptides. *Scientific Data*. 2019; 6: 148.
- Kapoor P, Singh H, Gautam A, Chaudhary K, Kumar R, Raghava GP. Tumorhope: A database of tumor homing peptides. *PLoS One*. 2012; 7: e35187.
- Kaushik AC, Mehmood A, Peng S, Zhang YJ, Dai X, Wei DQ. A-camp: A tool for anti-cancer and antimicrobial peptide generation. *J Biomol Struct Dyn*. 2021; 39: 285-293.
- Kayaalp SO. Tıbbi farmakoloji. In.: Pelikan Yayıncılık; 2012, p. 157-373.
- Kim JK, Lee E, Shin S, Jeong KW, Lee JY, Bae SY , ve ark. Structure and function of papiliocin with antimicrobial and anti-inflammatory activities isolated from the swallowtail butterfly, *papilio xuthus*. *J Biol Chem*. 2011; 286: 41296-41311.

- Koeppel F, Muller E, Harlé A, Guien C, Sujobert P, Trabelsi Grati O , ve ark. Standardisation of pathogenicity classification for somatic alterations in solid tumours and haematologic malignancies. *Eur J Cancer*. 2021; 159: 1-15.
- Kontomanolis E, Koutras A, Syllaios A, Schizas D, Kalagasidou S, Pagkalos A , ve ark. Basic principles of molecular biology of cancer cell-molecular cancer indicators. *Journal of B.U.ON. : official journal of the Balkan Union of Oncology*. 2021; 26: 1723-1734.
- Kontomanolis EN, Koutras A, Syllaios A, Schizas D, Mastoraki A, Garmpis N , ve ark. Role of oncogenes and tumor-suppressor genes in carcinogenesis: A review. *Anticancer Res*. 2020; 40: 6009-6015.
- Kotyk A. Iupac-iubmb joint commission on biochemical nomenclature (jcbn) and nomenclature committee of iubmb (nc-iubmb). *Newsletter 1999. Folia Microbiol (Praha)*. 1999; 44: 243-246.
- Kroemer G, Pouyssegur J. Tumor cell metabolism: Cancer's achilles' heel. *Cancer Cell*. 2008; 13: 472-482.
- Kuroda K, Okumura K, Isogai H, Isogai E. The human cathelicidin antimicrobial peptide ll-37 and mimics are potential anticancer drugs. *Frontiers in Oncology*. 2015; 5.
- La Rosa S. Vasoactive intestinal peptide (vip). In S La Rosa S Uccella (Eds.), *Endocrine pathology*. Cham: Springer International Publishing; 2022, p. 831-833.
- Landrum G, Tosco P, Kelley B, Ric, Cosgrove D, Sriniker , ve ark. (2023). Rdkit/rdkit: 2023_09_1 (q3 2023) release (Version Release_2023_09_1). Retrieved from <https://doi.org/10.5281/zenodo.10099869>
- LeBleu VS. Imaging the tumor microenvironment. *Cancer J*. 2015; 21: 174-178.
- Lee EY, Muller WJ. Oncogenes and tumor suppressor genes. *Cold Spring Harb Perspect Biol*. 2010; 2: a003236.

- Leite ML, da Cunha NB, Costa FF. Antimicrobial peptides, nanotechnology, and natural metabolites as novel approaches for cancer treatment. *Pharmacology & Therapeutics*. 2018; 183: 160-176.
- Leontiadou H, Mark AE, Marrink SJ. Antimicrobial peptides in action. *Journal of the American Chemical Society*. 2006; 128: 12156-12161.
- Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev*. 2001; 46: 3-26.
- Loper E, Bird S. (2002). Nltk: The natural language toolkit. In: arXiv.
- Louis DN, Perry A, Reifenberger G, von Deimling A, Figarella-Branger D, Cavenee WK, et al. The 2016 world health organization classification of tumors of the central nervous system: A summary. *Acta Neuropathol*. 2016; 131: 803-820.
- Manavalan B, Basith S, Shin TH, Choi S, Kim MO, Lee G. Mlaccp: Machine-learning-based prediction of anticancer peptides. *Oncotarget*. 2017; 8: 77121-77136.
- Marr AK, Gooderham WJ, Hancock RE. Antibacterial peptides for therapeutic use: Obstacles and realistic outlook. *Curr Opin Pharmacol*. 2006; 6: 468-472.
- Massagué J, Obenauf AC. Metastatic colonization by circulating tumour cells. *Nature*. 2016; 529: 298-306.
- Melaiu O, Lucarini V, Cifaldi L, Fruci D. Influence of the tumor microenvironment on nk cell function in solid tumors. *Front Immunol*. 2019; 10: 3038.
- Millington GW. The role of proopiomelanocortin (pomc) neurones in feeding behaviour. *Nutr Metab (Lond)*. 2007; 4: 18.
- Mookherjee N, Anderson MA, Haagsman HP, Davidson DJ. Antimicrobial host defence peptides: Functions and clinical potential. *Nature Reviews Drug Discovery*. 2020; 19: 311-332.

- Mooney C, Haslam NJ, Pollastri G, Shields DC. Towards the improved discovery and design of functional peptides: Common features of diverse classes permit generalized prediction of bioactivity. *PLoS One*. 2012; 7: e45012.
- Moore KS, Bevins C, Brasseur M, Tomassini N, Turner K, Eck H, Zasloff M. Antimicrobial peptides in the stomach of *xenopus laevis*. *Journal of Biological Chemistry*. 1991; 266: 19851-19857.
- Msaouel P, Jimenez-Fonseca P, Lim B, Carmona-Bayonas A, Agnelli G. Medicine before and after david cox. *European Journal of Internal Medicine*. 2022; 98: 1-3.
- Nasiri F, Atanaki FF, Behrouzi S, Kavousi K, Bagheri M. Cpacpp: In silico cell-penetrating anticancer peptide prediction using a novel bioinformatics framework. *ACS Omega*. 2021; 6: 19846-19859.
- Nelson DL, Cox M. *Lehninger principles of biochemistry: International edition*. Macmillan Learning; 2021.
- Niemirowicz K, Prokop I, Wilczewska AZ, Wnorowska U, Piktel E, Wątek M, et al. Magnetic nanoparticles enhance the anticancer activity of cathelicidin 11-37 peptide against colon cancer cells. *International Journal of Nanomedicine*. 2015; 10: 3843-3853.
- Ochoa R, Cossio P. Pepfun: Open source protocols for peptide-related computational analysis. *Molecules*. 2021; 26.
- Orivel J, Redeker V, Le Caer JP, Krier F, Revol-Junelles AM, Longeon A, et al. Ponericins, new antibacterial and insecticidal peptides from the venom of the ant *pachycondyla goeldii*. *J Biol Chem*. 2001; 276: 17823-17829.
- Panay P. (2021). Introducing windows 11. Retrieved from <https://blogs.windows.com/windowsexperience/2021/06/24/introducing-windows-11/>

- Parchebafi A, Tamanaee F, Ehteram H, Ahmad E, Nikzad H, Haddad Kashani H. The dual interaction of antimicrobial peptides on bacteria and cancer cells; mechanism of action and therapeutic strategies of nanostructures. *Microb Cell Fact.* 2022; 21: 118.
- Pauletti GM, Gangwar S, Siahaan TJ, Jeffrey A, T. Borchardt R. Improvement of oral peptide bioavailability: Peptidomimetics and prodrug strategies. *Advanced Drug Delivery Reviews.* 1997; 27: 235-256.
- Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O , ve ark. Scikit-learn: Machine learning in python. *Journal of Machine Learning Research.* 2011; 12: 2825-2830.
- Pérez-Herrero E, Fernández-Medarde A. Advanced targeted therapies in cancer: Drug nanocarriers, the future of chemotherapy. *European Journal of Pharmaceutics and Biopharmaceutics.* 2015; 93: 52-79.
- Persidis A. Cancer multidrug resistance. *Nature Biotechnology.* 1999; 17: 94-95.
- Piotto SP, Sessa L, Concilio S, Iannelli P. Yadamp: Yet another database of antimicrobial peptides. *Int J Antimicrob Agents.* 2012; 39: 346-351.
- Pirtskhalava M, Amstrong AA, Grigolava M, Chubinidze M, Alimbarashvili E, Vishnepolsky B , ve ark. Dbaasp v3: Database of antimicrobial/cytotoxic activity and structure of peptides as a resource for development of new therapeutics. *Nucleic Acids Research.* 2021; 49: D288-D297.
- Pituch KA, Stevens JP. Assumptions in manova. In *Applied multivariate statistics for the social sciences.* Routledge; 2015, p. 219-264.
- Pragati. (2020). Svm and random forest: A case study. Retrieved from <https://medium.com/@pandeypragati2112/svm-and-random-forest-a-case-study-6213da5be02f>

- Probst P, Wright MN, Boulesteix AL. Hyperparameters and tuning strategies for random forest. *WIREs Data Mining and Knowledge Discovery*. 2019; 9.
- Python. (2023). Python (Version 3.11.5): Python Software Foundation. Retrieved from <http://www.python.org>
- Qureshi A, Thakur N, Kumar M. Hipdb: A database of experimentally validated hiv inhibiting peptides. *PLoS One*. 2013; 8: e54908.
- Ramos-Espinosa O, Hernandez-Bazan S, Francisco-Cruz A, Mata-Espinosa D, Barrios-Payan J, Marquina-Castillo B, et al. Gene therapy based in antimicrobial peptides and proinflammatory cytokine prevents reactivation of experimental latent tuberculosis. *Pathogens and Disease*. 2016; 74.
- Reiter C, stuaxo. (2023). Pycairo (Version 1.25.1). Retrieved from <https://pycairo.readthedocs.io/>
- Ridyard KE, Overhage J. The potential of human peptide ll-37 as an antimicrobial and anti-biofilm agent. *Antibiotics (Basel)*. 2021; 10.
- Rohrl J, Yang D, Oppenheim JJ, Hehlhans T. Specific binding and chemotactic activity of mbd4 and its functional orthologue hbd2 to ccr6-expressing cells. *J Biol Chem*. 2010; 285: 7028-7034.
- Rose AJ. Role of peptide hormones in the adaptation to altered dietary protein intake. *Nutrients*. 2019; 11.
- Roudi R, Syn NL, Roudbary M. Antimicrobial peptides as biologic and immunotherapeutic agents against cancer: A comprehensive overview. *Frontiers in Immunology*. 2017; 8.
- Rumjuankiat K, Perez RH, Pilasombut K, Keawsompong S, Zendo T, Sonomoto K, Nitisinprasert S. Purification and characterization of a novel plantaricin, kl-1y, from *Lactobacillus plantarum* kl-1. *World J Microbiol Biotechnol*. 2015; 31: 983-994.

- Schaduangrat N, Nantasenamat C, Prachayasittikul V, Shoombuatong W. Acpred: A computational tool for the prediction and analysis of anticancer peptides. *Molecules*. 2019; 24: 1973.
- Schrank E, Wagner GE, Zangger K. Solution nmr studies on the orientation of membrane-bound peptides and proteins by paramagnetic probes. *Molecules*. 2013; 18: 7407-7435.
- scikit-learn developers. (2023). Sklearn.Ensemble.Randomforestclassifier. *scikit-learn documentation*. Retrieved from <https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html>
- Seebah S, Suresh A, Zhuo S, Choong YH, Chua H, Chuon D, Ve ark. Defensins knowledgebase: A manually curated database and information source focused on the defensins family of antimicrobial peptides. *Nucleic Acids Res*. 2007; 35: D265-268.
- Sexton PM, Findlay DM, Martin TJ. Calcitonin. *Current medicinal chemistry*. 1999; 6: 1067-1093.
- Sharma S, Kelly TK, Jones PA. Epigenetics in cancer. *Carcinogenesis*. 2010; 31: 27-36.
- Shtatland T, Guettler D, Kossodo M, Pivovarov M, Weissleder R. Pepbank - a database of peptides based on sequence text mining and public peptide data sources. *BMC Bioinformatics*. 2007; 8: 280.
- Speiser JL, Miller ME, Tooze J, Ip E. A comparison of random forest variable selection methods for classification prediction modeling. *Expert Syst Appl*. 2019; 134: 93-101.
- Sperandei S. Understanding logistic regression analysis. *Biochem Med (Zagreb)*. 2014; 24: 12-18.

- Stoltzfus JC. Logistic regression: A brief primer. *Acad Emerg Med*. 2011; 18: 1099-1104.
- Suhail Y, Cain MP, Vanaja K, Kurywachak PA, Levchenko A, Kalluri R, Kshitiz. Systems biology of cancer metastasis. *Cell Syst*. 2019; 9: 109-127.
- Szeto GL, Finley SD. Integrative approaches to cancer immunotherapy. *Trends Cancer*. 2019; 5: 400-410.
- Tang H, Su ZD, Wei HH, Chen W, Lin H. Prediction of cell-penetrating peptides with feature selection techniques. *Biochem Biophys Res Commun*. 2016; 477: 150-154.
- Team TPD. (2023). Pandas-dev/pandas: Pandas (Version v2.1.1): Zenodo. Retrieved from <https://doi.org/10.5281/zenodo.10107975>
<http://dx.doi.org/10.5281/zenodo.10107975>
- Terstappen GC, Reggiani A. In silico research in drug discovery. *Trends Pharmacol Sci*. 2001; 22: 23-26.
- Torgovnick A, Schumacher B. DNA repair mechanisms in cancer development and therapy. *Front Genet*. 2015; 6: 157.
- Tornesello AL, Borrelli A, Buonaguro L, Buonaguro FM, Tornesello ML. Antimicrobial peptides as anticancer agents: Functional properties and biological activities. *Molecules*. 2020; 25: 2850.
- Torrent M, Andreu D, Nogués VM, Boix E. Connecting peptide physicochemical and antimicrobial properties by a rational prediction model. *PLoS One*. 2011; 6: e16968.
- Torrent M, Di Tommaso P, Pulido D, Nogues MV, Notredame C, Boix E, Andreu D. Ampa: An automated web server for prediction of protein antimicrobial regions. *Bioinformatics*. 2012; 28: 130-131.

- Trapani JA, Darcy PK. Immunotherapy of cancer. *Aust Fam Physician*. 2017; 46: 194-199.
- Trédan O, Galmarini CM, Patel K, Tannock IF. Drug resistance and the solid tumor microenvironment. *JNCI: Journal of the National Cancer Institute*. 2007; 99: 1441-1454.
- Tsvetkova EV, Aleshina GM, Shamova OV, Leonova LE, Lehrer RI, Kokryakov VN. Alpha-defensins from blood leukocytes of the monkey *papio hamadryas*. *Biochemistry (Mosc)*. 2006; 71: 879-883.
- Tyagi A, Kapoor P, Kumar R, Chaudhary K, Gautam A, Raghava GPS. In silico models for designing and discovering novel anticancer peptides. *Scientific Reports*. 2013; 3: 2984.
- Tyagi A, Tuknait A, Anand P, Gupta S, Sharma M, Mathur D , ve ark. Cancerppd: A database of anticancer peptides and proteins. *Nucleic Acids Research*. 2015; 43: D837-D843.
- Uddin S, Haque I, Lu H, Moni MA, Gide E. Comparative performance analysis of k-nearest neighbour (knn) algorithm and its different variants for disease prediction. *Scientific Reports*. 2022; 12: 6256.
- Uraki S, Sugimoto K, Shiraki K, Tameda M, Inagaki Y, Ogura S , ve ark. Human beta-defensin-3 inhibits migration of colon cancer cells via downregulation of metastasis-associated 1 family, member 2 expression. *Int J Oncol*. 2014; 45: 1059-1064.
- van Heel AJ, de Jong A, Song C, Viel JH, Kok J, Kuipers OP. Bagel4: A user-friendly web server to thoroughly mine ripps and bacteriocins. *Nucleic Acids Research*. 2018; 46: W278-W281.
- Varadi M, Anyango S, Deshpande M, Nair S, Natassia C, Yordanova G , ve ark. Alphafold protein structure database: Massively expanding the structural

- coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res.* 2022; 50: D439-D444.
- Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D , ve ark. Scipy 1.0: Fundamental algorithms for scientific computing in python. *Nature Methods.* 2020; 17: 261-272.
- Waghu FH, Barai RS, Gurung P, Idicula-Thomas S. Camp₁₃ : A database on sequences, structures and signatures of antimicrobial peptides: Table 1. *Nucleic Acids Research.* 2016; 44: D1094-D1097.
- Wang CK, Kaas Q, Chiche L, Craik DJ. Cybase: A database of cyclic protein sequences and structures, with applications in protein discovery and engineering. *Nucleic Acids Res.* 2008; 36: D206-210.
- Wang G. Antimicrobial peptides: Discovery, design and novel therapeutic strategies, 2nd edition. CABI; 2017.
- Wang G. The antimicrobial peptide database is 20 years old: Recent developments and future directions. *Protein Sci.* 2023; 32: e4778.
- Wang G, Li X, Wang Z. Apd3: The antimicrobial peptide database as a tool for research and education. *Nucleic Acids Research.* 2016; 44: D1087-D1093.
- Wang L, Wang N, Zhang W, Cheng X, Yan Z, Shao G , ve ark. Therapeutic peptides: Current applications and future directions. *Signal Transduct Target Ther.* 2022; 7: 48.
- Wang Z, Shen Z, Li Z, Duan J, Fu S, Liu Z , ve ark. Activation of the bmp-bmpr pathway conferred resistance to egfr-tkis in lung squamous cell carcinoma patients with egfr mutations. *Proceedings of the National Academy of Sciences.* 2015; 112: 9990-9995.
- Waskom M. Seaborn: Statistical data visualization. *Journal of Open Source Software.* 2021; 6.

- Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R , ve ark. Swiss-model: Homology modelling of protein structures and complexes. *Nucleic Acids Res.* 2018; 46: W296-W303.
- Weder JKP, Belitz HD. *Protein | chemistry*. Oxford: Academic Press; 2003.
- Windows. (2023). Office visual basic for applications (vba) (Version 16.0.16924). Retrieved from <https://learn.microsoft.com/en-us/office/vba/api/overview/>
- Wishart DS, Feunang YD, Guo AC, Lo EJ, Marcu A, Grant JR , ve ark. Drugbank 5.0: A major update to the drugbank database for 2018. *Nucleic Acids Res.* 2018; 46: D1074-D1082.
- Wu JM, Jan PS, Yu HC, Haung HY, Fang HJ, Chang YI , ve ark. Structure and function of a custom anticancer peptide, cb1a. *Peptides.* 2009; 30: 839-848.
- Yan J, Bhadra P, Li A, Sethiya P, Qin L, Tai HK , ve ark. Deep-ampep30: Improve short antimicrobial peptides prediction with deep learning. *Molecular Therapy - Nucleic Acids.* 2020; 20: 882-894.
- Yang S, Lee CW, Kim HJ, Jung HH, Kim JI, Shin SY, Shin SH. Structural analysis and mode of action of bmap-27, a cathelicidin-derived antimicrobial peptide. *Peptides.* 2019; 118: 170106.
- Ye G, Wu H, Huang J, Wang W, Ge K, Li G , ve ark. Lamp2: A major update of the database linking antimicrobial peptides. *Database.* 2020; 2020: baaa061.
- Zahreddine H, Borden KL. Mechanisms and insights into drug resistance in cancer. *Front Pharmacol.* 2013; 4: 28.
- Zhang C, Yang M, Ericsson AC. Antimicrobial peptides: Potential application in liver cancer. *Frontiers in Microbiology.* 2019; 10.

- Zhang H, Qureshi MA, Wahid M, Charifa A, Ehsan A, Ip A , ve ark. Differential diagnosis of hematologic and solid tumors using targeted transcriptome and artificial intelligence. *Am J Pathol.* 2023; 193: 51-59.
- Zhang Q-Y, Yan Z-B, Meng Y-M, Hong X-Y, Shao G, Ma J-J , ve ark. Antimicrobial peptides: Mechanism of action, activity and clinical potential. *Military Medical Research.* 2021; 8: 48.
- Zhang Z. Introduction to machine learning: K-nearest neighbors. *Ann Transl Med.* 2016; 4: 218.
- Zhao T, Hu Y, Zang T. Dracp: A novel method for identification of anticancer peptides. *BMC Bioinformatics.* 2020; 21: 559.
- Zhou J, Liu Y, Shen T, Chen L, Zhang C, Cai K , ve ark. Antimicrobial activity of the antibacterial peptide pmap-36 and its analogues. *Microb Pathog.* 2019; 136: 103712.
- Zhou P, Huang J. Computational peptidology. Springer New York; 2015.

EKLER

Ek 1. 10.000 AMP içeren deneysel veri setinde antikanser etkinlik gösteren ve fizikokimyasal özellikler için belirlenen aralıkta bulunan 1648 peptit (molekül ağırlığına göre sıralanmıştır)

Google Drive indirme bağlantısı (https://docs.google.com/spreadsheets/d/1SMP350f7ec-pfNiRcwiHTJwuRsoUlXxi/edit?usp=drive_link&ouid=104854238955850120804&rtpof=true&sd=true)



ÖZGEÇMİŞ

Kişisel Bilgiler

Adı	Ahmet Ozan	Uyruğu	
Soyadı	Özgen	Tel no	
Doğum tarihi		e-posta	

Eğitim Bilgileri

Mezun olduğu kurum		Mezuniyet yılı
Lise	Dr. İlhami Tankut Anadolu Lisesi	2013
Lisans	İstanbul Medipol Üniversitesi – Eczacılık Fakültesi	2019
Yüksek Lisans	Akdeniz Üniversitesi Sağlık Bilimleri Enstitüsü – Tıbbi Biyoteknoloji	2024

İş Deneyimi

Görevi	Kurum	Süre (yıl-yıl)
Yardımcı Eczacılık	Avaroğlu Eczanesi	2019-2020
İş Geliştirme Uzmanı	Biocalis	2020-2023
Klinik Araştırmalar Uzmanı	Nobel İlaç	2023-

Yabancı Dilleri	Sınav türü	Puanı
İngilizce	e-YDS	92.5
İngilizce	IELTS Academic	7.5

Burslar-Ödüller:

Huawei, Seeds for the Future (ICT Yetenek Programı), İstanbul / Türkiye – Atina / Yunanistan, 07/2022 – 12/2022

Yayınlar ve Bildiriler:

Hakemli dergilerle yayımlanmış yayınlar (WoS)

Uslu, A., Güçarslan, İ., Sağıroğlu, B., Sağlam, O., Nacak, M., Ocak, Ö. H., Öztürk, Y., & Özgen, A. O. (2023). Pharmacokinetic Profile and a Bioequivalence Study of Apixaban Tablets in Healthy Volunteers. Nobel Medicus, 19(3), 200–208.

Hakemli dergilerle yayımlanmış yayınlar (Diğer)

Özgen, A. O., & Eyüpoğlu, O. E. (2020). Antibiotic Susceptibility Of Klebsiella pneumoniae Strains Isolated From Clinical Samples. Journal of the Turkish Chemical Society Section A: Chemistry, 7(1), 319–326. <https://doi.org/10.18596/jotcsa.635088>

Tutkun, G., Özgen, A. O., & Bilge, U. (2022). Models of spike, envelope and membrane proteins of SARS-CoV-2 variants for possible therapeutics. Journal of Medical Topics and Updates, 1(3), 133-141. <https://doi.org/10.5281/zenodo.7476170>

Sözlü Sunumlar ve Poster Bildirileri

Özgen, A. O., & Eyüpoğlu, O. E. (09/2019). Antibiotic Susceptibility Of Klebsiella pneumoniae Strains Isolated From Clinical Samples (Poster Sunumu), 31. Ulusal Kimya Kongresi, İstanbul / Türkiye.

Tutkun, G., Özgen, A. O., & Bilge, U. (09/2021). Models of spike, envelope and membrane proteins of SARS-CoV-2 variants for possible therapeutics (Sözlü Sunum), BIO Türkiye – Uluslararası Biyoteknoloji Kongresi 2021, İstanbul / Türkiye.

Özgen, A. O., & Demir-Dora, D. (11/2021). Pharmacoeconomic Evaluation of Biosimilars in Developing Countries (Sözlü Sunum), 26. Ulusal ve 1. Uluslararası Farmakoloji Kongresi, Çevrimiçi.

Özgen, A. O., & Demir-Dora, D. (02/2022). In Silico Stability of Matrixyl Peptide Solution in Thermal Stress, Uluslararası Farmasötik Teknoloji Sempozyumu (IPTS 2022), Ankara / Türkiye.

Özgen, A. O., & Demir-Dora, D. (09/2023). Pharmacoeconomic Evaluation of Nusinersen for Spinal Muscular Atrophy (SMA) in Türkiye, BIO Türkiye – Uluslararası Biyoteknoloji Kongresi 2023, İstanbul / Türkiye.

Özgen, A. O., & Demir-Dora, D. (11/2023). Pharmacoeconomic Evaluation of Ulcerative Colitis and Crohn's Disease Treatments in Türkiye in Terms of Biosimilars, 27. Ulusal ve 2. Uluslararası Farmakoloji Kongresi, Antalya / Türkiye.