T.C.
YILDIZ TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

SİLİYOPATİ ADAY GENLERİNİN ÖNCELİKLENDİRİLMESİ İÇİN AÇIKLANABİLİR MAKİNE ÖĞRENMESİ MODELİ

Rumeysa KOYUNCUOĞLU

YÜKSEK LİSANS TEZİ
Moleküler Biyoloji ve Genetik Anabilim Dalı
Moleküler Biyoloji ve Genetik Programı

Danışman
Prof. Dr. Nehir ÖZDEMİR ÖZGENTÜRK

Aralık, 2025

T.C.
YILDIZ TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

SİLİYOPATİ ADAY GENLERİNİN ÖNCELİKLENDİRİLMESİ İÇİN AÇIKLANABİLİR MAKİNE ÖĞRENMESİ MODELİ

Rumeysa KOYUNCUOĞLU tarafından hazırlanan tez çalışması 24.12.2025 tarihinde aşağıdaki jüri tarafından Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü Moleküler Biyoloji ve Genetik Anabilim Dalı Moleküler Biyoloji ve Genetik Programı **YÜKSEK LİSANS TEZİ** olarak kabul edilmiştir.

Prof. Dr. Nehir ÖZDEMİR ÖZGENTÜRK
Yıldız Teknik Üniversitesi
Danışman

Jüri Üyeleri

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

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

Prof. Dr. Günay YETİK ANACAK, Üye
Acıbadem Üniversitesi _____

Danışmanım Prof. Dr. Nehir ÖZDEMİR ÖZGENTÜRK sorumluluğunda tarafımca hazırlanan SİLİYOPATİ ADAY GENLERİNİN ÖNCELİKLENDİRİLMESİ İÇİN AÇIKLANABİLİR MAKİNE ÖĞRENMESİ MODELİ başlıklı çalışmada veri toplama ve veri kullanımında gerekli yasal izinleri aldığımı, diğer kaynaklardan aldığım bilgileri ana metin ve referanslarda eksiksiz gösterdiğimi, araştırma verilerine ve sonuçlarına ilişkin çarpıtma ve/veya sahtecilik yapmadığımı, çalışmam süresince bilimsel araştırma ve etik ilkelerine uygun davrandığımı beyan ederim. Beyanımın aksinin ispatı halinde her türlü yasal sonucu kabul ederim.

Rumeysa KOYUNCUOĞLU

İmza

Nadir hastalıklarla

mücadele eden

tüm bireylere

ve ailelerine...

TEŞEKKÜR

Yüksek lisans eğitimim boyunca yol gösteren danışman hocam Prof. Dr. Nehir ÖZDEMİR ÖZGENTÜRK'e en içten teşekkürlerimi sunarım. Akademik rehberliğinin ötesinde; nezaketi, desteği ve hoşgörüsü için kendisine minnettarım.

Varlıklarıyla bana güç veren, fedakarlıklarıyla yanımda olan canım annem ve babama sonsuz teşekkürlerimi sunarım. Bu başarının gururunu onlarla paylaşıyor olmak, benim için tüm akademik kazanımların en anlamlısıdır.

Motivasyonumu her daim yüksek tutan, dostluklarıyla bu yolu anlamlı kılan sevgili arkadaşlarım Aslı, Berkay, Bilge ve Funda'ya teşekkür ederim. Varlığınız ve desteğiniz, bu çalışmanın her aşamasında büyük bir güç kaynağı oldu.

Son olarak, varlığıyla hayatıma güzellik katan ve kalbimdeki yeri bambaşka olan biricik kardeşim Mete'ye... İyi ki varsın.

Rumeysa KOYUNCUOĞLU

İÇİNDEKİLER

SİMGE LİSTESİ viii

KISALTMA LİSTESİ ix

ŞEKİL LİSTESİ xi

TABLO LİSTESİ xii

ÖZET xiii

ABSTRACT xiv

1 GİRİŞ 1

1.1 Silyum ve Siliyopatilere Genel Bakış 1

1.2 Siliyopatilerin Genetik ve Klinik Heterojenliği 2

1.3 Gen Önceliklendirmesinde Makine Öğrenmesi 2

1.4 Tezin Amacı, Kapsamı ve Hipotezi 3

2 GENEL BİLGİLER 5

2.1 Silyumun Yapısı, Biyogenezi ve Moleküler Mimarisi 5

2.1.1 Bazal Cisim ve Vezikül Trafiği 5

2.1.2 Geçiş Bölgesi (Transition Zone) 6

2.1.3 Aksonem 7

2.1.4 İntraflagellar Taşıma (IFT) Mekanizması 8

2.2 Siliyer Sinyal İletim Mekanizmaları 9

2.2.1 Hedgehog (Hh) Sinyal Yolağı 9

2.2.2 Wnt Sinyal Yolağı 9

2.2.3 PDGF Sinyal Yolağı 10

2.3 Siliyopati Spektrumu 11

2.3.1 Motil Siliyopatiler: Primer Siliyer Diskinezi 11

2.3.2 Non-motil Siliyopatiler 13

2.3.3 Genotip-Fenotip İlişkisindeki Karmaşıklık 19

2.4 Biyoinformatik Veri Kaynakları 20

2.4.1 Genomik ve Fenotipik Veri Kaynakları 20

2.4.2 İnsan Fenotip Ontolojisi (HPO) Veritabanı 21
2.4.3 Genome Aggregation Database (gnomAD) Veritabanı . . . 22
2.4.4 STRING Veritabanı 23
2.5 Makine Öğrenmesi ve Hesaplamalı Yöntemler 23
2.5.1 Topluluk Öğrenmesi (Ensemble Learning) 23
2.5.2 XGBoost (eXtreme Gradient Boosting) Algoritması 25
2.5.3 Model Performans Değerlendirme Metrikleri 26
2.5.4 XAI ve SHAP Yöntemi 27

3 MATERYAL VE YÖNTEM 29
3.1 Çalışma Tasarımı ve Veri Kaynakları 30
3.1.1 Negatif (Kontrol) Setinin Oluşturulması 30
3.2 Öznitelik Mühendisliği (Feature Engineering) 31
3.2.1 Genomik Kısıtlılık ve Mutasyon Toleransı 31
3.2.2 Klinik ve Fenotipik Öznitelikler 32
3.2.3 İnteraktom Analizi ve Ağ Merkeziyeti 32
3.3 Veri Ön İşleme (Data Preprocessing) 33
3.3.1 Eksik Veri Yönetimi (Missing Data Handling) 33
3.3.2 Veri Setinin Ayrılması (Train-Test Split) 33
3.4 Makine Öğrenmesi Modelinin Eğitilmesi 34
3.5 Model Performansının Ölçülmesi 34
3.5.1 Sınıflandırma Metrikleri 35
3.5.2 Eşik Değerinden Bağımsız Eğri Analizleri (ROC ve AUPRC) 35
3.6 Modelin Uygulanması ve Karar Mekanizması 36
3.6.1 Aday Genlerin İn-silico Taraması 36
3.6.2 SHAP Analizi 37
3.6.3 Biyolojik Validasyon: PPI ve Fonksiyonel Zenginleştirme . 37

4 BULGULAR 39
4.1 Veri Seti İstatistikleri ve Öznitelik Dağılım Bulguları 39
4.1.1 Veri Seti İstatistikleri 39
4.2 Özniteliklerin İstatistiksel Dağılımı 40
4.3 Model Performans Bulguları 42
4.3.1 Sınıflandırma Performansı Metrikleri 42
4.3.2 Karışıklık Matrisi (Confusion Matrix) 43
4.3.3 ROC ve PR Eğrileri 44
4.4 Aday Gen Önceliklendirme Bulguları 44
4.5 Modelin Karar Mekanizması Bulguları 46
4.5.1 Açıklanabilir Yapay Zeka (XAI): Global SHAP Analizi . . . 46

4.5.2 Açıklanabilir Yapay Zeka (XAI): CRYBB3 Geni Lokal SHAP Analizi 47
4.6 Aday Genlerin Ağ ve Yolak Analizi Bulguları 48
4.6.1 Protein-Protein Etkileşim (PPI) Ağı Bulguları 48
4.6.2 Fonksiyonel Zenginleştirme ve Yolak Analizi Bulguları . . . 49

5 TARTIŞMA **51**
5.1 Model Performansı ve Metodolojinin Değerlendirilmesi 52
5.2 Model Karar Mekanizmasının Değerlendirilmesi 53
5.3 Sistem Biyolojisi Perspektifiyle Değerlendirilmesi 53
5.4 Çalışmanın Kısıtlılıkları ve Gelecek Perspektifi 55

6 SONUÇ **56**

A EK A: Aday Gen Listesi **76**

B EK B: Python Kodları **77**

C EK C: Fenotip Zenginleştirme Analizi Sonuçları **78**

SİMGE LİSTESİ

μ	Aritmetik Ortalama (Mean)
R^2	Belirlilik Katsayısı (Coefficient of Determination)
$\cup$	Birleşim Kümesi (Union)
p	İstatistiksel Anlamlılık Değeri (P-value)
$\cap$	Kesişim Kümesi (Intersection)
n	Örneklem Sayısı
ϕ	SHAP Değeri / Shapley Katkı Payı (SHapley Additive exPlanation)
σ	Standart Sapma (Standard Deviation)

KISALTMA LİSTESİ

AI	Yapay Zeka (Artificial Intelligence)
AP	Ortalama Hassasiyet (Average Precision)
AUC	Eğri Altında Kalan Alan (Area Under the Curve)
AUPRC	Hassasiyet-Duyarlılık Eğrisi Altında Kalan Alan
BBS	Bardet-Biedl Sendromu (Bardet-Biedl Syndrome)
CiliaCarta	Siliyopati İle İlişkili Genler Veri Tabanı
DisGeNET	Hastalık-Gen İlişkilendirme Veri Tabanı
EVC	Ellis-van Creveld Sendromu
FN	Yanlış Negatif (False Negative)
FP	Yanlış Pozitif (False Positive)
gnomAD	Genom Agregasyon Veri Tabanı (Genome Aggregation Database)
GO	Gen Ontolojisi (Gene Ontology)
HPO	İnsan Fenotip Ontolojisi (Human Phenotype Ontology)
IFT	İntraflagellar Taşıma (Intraflagellar Transport)
JBTS	Joubert Sendromu (Joubert Syndrome)
KEGG	Kyoto Gen ve Genom Ansiklopedisi
KS	Kartagener Sendromu (Kartagener Syndrome)
LCA	Leber Konjenital Amarozu (Leber Congenital Amaurosis)
LOEUF	Gözlenen/Beklenen Üst Sınır Oranı (LoF O/E Upper Bound Fraction)
MKS	Meckel-Gruber Sendromu (Meckel-Gruber Syndrome)
ML	Makine Öğrenmesi (Machine Learning)
NGS	Yeni Nesil Dizileme (Next Generation Sequencing)
NPHP	Nefronofitizi / Böbrek Hastalığı (Nephronophthisis)

PCD	Primer Siliyer Diskinezi (Primary Ciliary Dyskinesia)
pLI	Fonksiyon Kaybına İntolerans Olasılığı (Prob. of LoF Intolerance)
PPI	Protein-Protein Etkileşimi (Protein-Protein Interaction)
PR	Precision-Recall Eğrisi (Precision-Recall)
ROC	Alıcı İşletim Karakteristiği (Receiver Operating Characteristic)
ROC-AUC	ROC Eğrisi Altında Kalan Alan (Area Under the ROC Curve)
SCGS	Siliyer Gen Seti (Ciliary Gene Set)
SHAP	Shapley Katkı Payı Açıklamaları (SHapley Additive exPlanations)
SLS	Senior-Løken Sendromu (Senior-Løken Syndrome)
STRING	Search Tool for the Retrieval of Interacting Genes/Proteins
TN	Doğru Negatif (True Negative)
TP	Doğru Pozitif (True Positive)
WES	Tüm Ekzom Dizileme (Whole Exome Sequencing)
XAI	Açıklanabilir Yapay Zeka (Explainable Artificial Intelligence)
XGBoost	Ekstrem Gradyan Artırma (eXtreme Gradient Boosting)

ŞEKİL LİSTESİ

Şekil 2.1 Silyum Tipleri 7
Şekil 3.1 Biyoinformatik İş Akışı 29
Şekil 4.1 Pozitif Set Venn Şeması 39
Şekil 4.2 Violin Dağılım Grafikleri 41
Şekil 4.3 Karışıklık Matrisi (Confusion Matrix) 43
Şekil 4.4 ROC ve PR Grafikleri 44
Şekil 4.5 Global SHAP Beeswarm Grafiği 46
Şekil 4.6 Lokal SHAP Force Plot Grafiği 47
Şekil 4.7 PPI Ağı 48
Şekil 4.8 Fonksiyonel Zenginleştirme Grafiği 49

TABLO LİSTESİ

Tablo 3.1 XGBoost Hiperparametre Yapılandırması 34
Tablo 4.1 Analiz Süreçlerindeki Gen Dağılım İstatistikleri 40
Tablo 4.2 Model performans metrikleri özeti 42
Tablo 4.3 En yüksek skorlu ilk 20 siliyopati aday geni 45

ÖZET

SİLİYOPATİ ADAY GENLERİNİN ÖNCELİKLENDİRİLMESİ İÇİN AÇIKLANABİLİR MAKİNE ÖĞRENMESİ MODELİ

Rumeysa KOYUNCUOĞLU

Moleküler Biyoloji ve Genetik Anabilim Dalı
Yüksek Lisans Tezi

Danışman: Prof. Dr. Nehir ÖZDEMİR ÖZGENTÜRK

Silyum, ökaryotik hücrelerin yüzeyinden ekstraselüler ortama uzanan mikrotübül temelli bir organeldir. Bu yapıda meydana gelen mutasyonlar, genetik heterojenliğe sahip multisistemik bir hastalık grubu olan siliyopatilere yol açarak tanı süreçlerini zorlaştırmaktadır. Bu çalışmada, yeni aday genlerin keşfedilmesi için denetimli makine öğrenmesi tabanlı bir model geliştirilmiştir. XGBoost algoritması kullanılarak eğitilen model, bağımsız test setinde 0,84 AUC ve 0,88 AUPRC performansı sergilemiştir. Modelin karar mekanizması XAI ve SHAP analizleri ile aydınlatılmış, genomik kısıtlılık ve fenotipik benzerlik verilerinin model tahminindeki rolü kantitatif olarak gösterilmiştir. Analiz sonucunda CRYBB3 ve VPS33B en güçlü adaylar olarak saptanmıştır.

Anahtar Kelimeler: Siliyopati, Makine Öğrenmesi, XGBoost, Aday Gen Tahmini, Biyoinformatik

YILDIZ TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

ABSTRACT

Explainable Machine Learning Model for Prioritization of Ciliopathy Candidate Genes

Rumeysa KOYUNCUOĞLU

Department of Molecular Biology and Genetics
Master of Science Thesis

Supervisor: Prof. Dr. Nehir ÖZDEMİR ÖZGENTÜRK

The cilium is a microtubule-based organelle extending from the surface of eukaryotic cells into the extracellular environment. Mutations within this structure lead to "ciliopathies", a group of multisystemic disorders with high genetic heterogeneity that complicate diagnostic processes. In this study, a supervised machine learning-based model was developed for the discovery of novel candidate genes. The model, trained using the XGBoost algorithm, demonstrated a performance of 0.84 AUC and 0.88 AUPRC on an independent test set. The decision mechanism of the model was elucidated through XAI and SHAP analyses, quantitatively demonstrating the roles of genomic constraint and phenotypic similarity data in model predictions. As a result of the analysis, CRYBB3 and VPS33B were identified as the strongest candidates.

Keywords: Ciliopathy, Machine Learning, XGBoost, Candidate Gene Prediction, Bioinformatics

YILDIZ TECHNICAL UNIVERSITY
GRADUATE SCHOOL OF SCIENCE AND ENGINEERING

1
GİRİŞ

1.1 Silyum ve Siliyopatilere Genel Bakış

Silyum (Latince: *cilium*; çoğulu silya/*cilia*), ökaryotik hücrelerin yüzeyinden ekstraselüler ortama uzanan, evrimsel süreçte yüksek oranda korunmuş, mikrotübül temelli bir organeldir [19]. İlk keşfedildiği dönemlerde işlevsiz bir evrimsel kalıntı olduğu düşünülen bu yapı, ilerleyen yıllarda yapılan moleküler çalışmalarla biyolojinin dinamik konularından biri haline gelmiştir. Günümüzde primer silyumun, hücre dışı ortamdan gelen mekanik ve kimyasal sinyalleri algılayarak; Hedgehog (Hh), Wnt ve PDGF gibi önemli sinyal yolaklarını düzenleyen özelleşmiş bir organel olduğu kanıtlanmıştır [16].

Yapısal olarak silyum, hücre zarının özelleşmiş bir uzantısıdır ve temel iskeletini "aksonem" adı verilen mikrotübül demetleri oluşturmaktadır. Çoğu somatik hücrede bulunan ve genellikle hareketsiz olan primer silyum, "9+0" mikrotübül düzenine sahipken; solunum yolu epiteli, epandimal hücreler ve sperm kuyruğunda bulunan motil (hareketli) silyum ise genellikle "9+2" düzenine sahiptir. Bu yapısal ve fonksiyonel çeşitlilik, organelin sadece mukosiliyer temizlik veya hücre hareketiyle sınırlı kalmayıp; fotoreseptör dış segmentinin oluşumu, böbrek tübüllerinde idrar akışının algılanması ve embriyonik gelişimde organ asimetrisinin belirlenmesi gibi hayati süreçleri yönetmesini sağlamaktadır [79, 88, 138].

Bu organelin biyogenez (siliyogenez) sürecinde veya sinyal iletim fonksiyonlarındaki genetik defektler, "Siliyopatiler" adı verilen geniş ve heterojen bir hastalık spektrumuna yol açmaktadır. Silyumun neredeyse tüm dokularda bulunması nedeniyle, siliyopatiler multisistemik bir karakter sergilemekte olup; retinal dejenerasyon, kistik böbrek hastalığı, polidaktili, obezite, iskelet displazileri ve santral sinir sistemi malformasyonları bu spektrumun sık görülen klinik bulgularıdır [147].

Silyumun biyolojik önemi sadece hücresel sinyal iletimiyle sınırlı olmayıp; aynı

zamanda embriyonik gelişimde vücut planının (sol-sağ asimetrisi) belirlenmesinde de merkezi bir rol oynamaktadır. Bu organelin fonksiyon bozukluğu, organların ters yerleşimi (*situs inversus*) gibi sonuçlara yol açabilmektedir. Bu durum, silyumun aynı zamanda organizmanın makro-anatomik mimarı olduğunu da göstermektedir [12, 43].

1.2 Siliyopatilerin Genetik ve Klinik Heterojenliği

Siliyopatilerin tanı ve araştırmasındaki en temel zorluk, bu hastalık grubunun olağanüstü genetik ve klinik heterojenliğidir. Bu durum, siliyopatilerin neden bir "spektrum" olarak tanımlandığını açıklamaktadır.

Genetik düzeyde bu heterojenlik iki ana başlıkta incelenmektedir:

1. **Lokus Heterojenitesi:** Çok sayıda farklı genin, benzer bir klinik tabloya (fenotipe) neden olmasıdır. Örneğin, Joubert Sendromu (JS) fenotipine neden olduğu bilinen 40'tan fazla gen vardır.
2. **Alelik Heterojenite ve Pleiotropi:** Aynı gen üzerindeki farklı mutasyonların, farklı şiddette sendromlara yol açmasıdır. Örneğin *CEP290* genindeki mutasyonlar; izole Leber Konjenital Amaurosis (LCA) tablosundan, letal Meckel-Gruber Sendromuna (MKS) kadar uzanan geniş bir yelpazede hastalığa neden olmaktadır.

Genetik karmaşıklığa ek olarak, değişken ekspresivite (aynı genetik mutasyona sahip bireylerde hastalığın şiddetinin farklı olması) ve eksik penetrans gibi olgular, klinik tabloyu daha da karmaşık hale getirmektedir. Böbrek kistleri veya retinal distrofi gibi fenotiplerin birçok farklı siliyopati sendromunda ortak olarak görülmesi (fenotipik örtüşme), sadece klinik bulgulara dayalı ayırıcı tanıyı neredeyse imkansız kılmaktadır. Bu durum, klinikte hastaların kesin tanıya ulaşana kadar yıllarca süren testlere maruz kaldığı "Tanısal Odyssey" sürecine neden olmaktadır [77, 51, 85].

1.3 Gen Önceliklendirmesinde Makine Öğrenmesi

Siliyopatiler gibi genetik olarak heterojen hastalıklarda, tek gen analizleri veya dar kapsamlı gen panelleri tanısal bir darboğaz oluşturmaktadır. Yeni Nesil Dizileme (NGS) teknolojilerinin, özellikle Tüm Ekzom Dizileme (WES) yönteminin yaygınlaşması, tanı oranlarını arttırmakla birlikte veri yorumlama problemini de beraberinde getirmiştir. Bir WES analizinde tespit edilen yüzlerce varyant

arasından, patojenik olanı ayırt etmek ve "Anlamı Bilinmeyen Varyant" (VUS) yükünü azaltmak, modern tıbbi genetiğin öncelikli hedefidir.

Bu noktada biyoinformatik yaklaşımlar ve Denetimli Makine Öğrenmesi (Supervised Machine Learning) algoritmaları devreye girmektedir. Makine öğrenmesi modelleri, genlerin evrimsel korunumu (PhastCons), genomik kısıtlılık skorları (pLI / LOEUF), protein-protein etkileşim ağları (PPI) ve fenotipik benzerlikler gibi çok boyutlu verileri entegre ederek, karmaşık örüntüleri tespit edebilmektedir [56].

Eğitilen model, bilinen siliyopati genlerinin ortak özelliklerini öğrenerek, henüz hastalıkla ilişkilendirilmemiş genler için bir "aday skoru" (prediction score) üretebilmektedir. Bu tez çalışmasında kullanılan XGBoost (eXtreme Gradient Boosting) gibi topluluk öğrenme (ensemble learning) algoritmaları, özellikle dengesiz veri setlerinde yüksek performans göstermeleri nedeniyle varyant önceliklendirme çalışmalarında sıklıkla tercih edilmektedir [106, 112].

1.4 Tezin Amacı, Kapsamı ve Hipotezi

Siliyopatilerde gözlenen yüksek genetik heterojenite, halen moleküler tanısı konulamamış büyük bir hasta grubuna sebep olmaktadır. Bu durum, siliyer fonksiyonlarda rol oynayan ancak güncel literatürde henüz siliyopatiler ile ilişkisi tanımlanmamış genlerin keşfedilmeyi beklediğini göstermektedir.

Bu çalışmanın temel amacı; genomik, transkriptomik ve fenotipik verileri bütünleştiren bir biyoinformatik iş akışı ve makine öğrenmesi modeli geliştirerek, siliyopati spektrumu için yeni aday genleri yüksek doğrulukla önceliklendirmektir.

Çalışmanın Hipotezi: "Bilinen siliyopati genlerinin (Pozitif Set) ve silyum ilişkisi olmayan genlerin (Negatif Set); genomik kısıtlılık, biyolojik süreç benzerliği ve protein etkileşim ağı topolojisi gibi öznitelikleri kullanılarak eğitilen bir XGBoost modeli, insan genomunda henüz siliyopati ile ilişkilendirilmemiş genleri, siliyopatiye neden olma potansiyellerine göre başarılı bir şekilde sıralayabilir."

Tezin kapsamı şu adımlardan oluşmaktadır:

1. **Veri Seti Oluşturma:** CiliaCarta ve SYSCILIA veritabanlarından yüksek güvenilirlikli siliyer gen setinin derlenmesi.
2. **Öznitelik Mühendisliği (Feature Engineering):** Genomik (gnomAD), evrimsel (PhyloP) ve fenotipik (HPO) veritabanlarından makine öğrenmesi

için girdi vektörlerinin oluşturulması.

3. **Model Eğitimi ve Validasyon:** XGBoost algoritması ile tahmin modelinin geliştirilmesi, ROC eğrisi ve Precision-Recall metrikleri ile performansının değerlendirilmesi.

4. **Açıklanabilirlik (XAI):** SHAP (SHapley Additive exPlanations) analizleri ile modelin karar mekanizmasının ve öznitelik önem düzeylerinin biyolojik olarak doğrulanması.

5. **Aday Genlerin Keşfi:** Analiz sonucunda elde edilen yüksek skorlu adayların literatür ışığında tartışılması.

Bu tez çalışması, sadece istatistiksel bir sıralama sunmakla kalmayıp; PPI ağ analizi ve fonksiyonel zenginleştirme (Enrichment) testleri ile desteklenen, deneysel validasyona hazır güçlü aday genler sunmayı hedeflemektedir.

2
GENEL BİLGİLER

2.1 Silyumun Yapısı, Biyogenezi ve Moleküler Mimarisi

Silyum (Latince: *cilium*; çoğulu silya/*cilia*), ökaryotik hücrelerin apikal yüzeyinden ekstraselüler ortama doğru uzanan, evrimsel süreçte yüksek oranda korunmuş, mikrotübül temelli ve membranla çevrili özelleşmiş bir organeldir. Fonksiyonel ve ultrastrüktürel özelliklerine göre silyum temel olarak iki ana sınıfta incelenmektedir [151]:

1. **Primer Silyum:** Çoğu somatik hücrede (böbrek tübül epiteli, fibroblastlar, nöronlar vb.) bulunan, "9+0" mikrotübül düzenine sahip, hareketsiz (immotil) yapıdır. Temel görevi, ortamdaki mekanik ve kimyasal değişimleri algılayarak Hedgehog (Hh), Wnt ve PDGF gibi sinyal yolaklarını tetiklemektir. Siliyopatilerin büyük çoğunluğu bu yapının disfonksiyonundan kaynaklanmaktadır [79].
2. **Motil Silyum:** Solunum yolu epiteli, beyin ventriküllerini döşeyen epandimal hücreler ve sperm kuyruğunda bulunan; ATP enerjisiyle mekanik güç üreterek sıvı akışı sağlayan, "9+2" mikrotübül düzenine sahip, hareketli yapıdır [35].

Silyumun oluşumu (siliyogenez) ve yıkımı, hücre döngüsü ile sıkı bir koordinasyon içindedir. Silyum, hücre bölünmesi öncesinde sentrozomun serbest kalabilmesi için kontrollü bir şekilde yıkılır; bölünme tamamlandıktan sonra ise tekrar sentezlenmektedir. Bu dinamik süreç, hücre proliferasyonu ile diferansiyasyonu arasındaki dengenin korunması için elzemdir [6].

2.1.1 Bazal Cisim ve Vezikül Trafiği

Siliyogenez süreci, sentrozomu oluşturan iki sentriyolden daha yaşlı olanın (Anne Sentriyol / Mother Centriole) farklılaşarak plazma zarına kenetlenmesiyle

başlar. Bu olgunlaşmış yapı Bazal Cisim (Basal Body) olarak adlandırılır ve silyumun mikrotübül organizasyon merkezi (MTOC) olarak görev yapar [7]. Bazal cismin distal ucunda, mikrotübül tripletlerine bağlı "Distal Apendiks" (DA) ve "Subdistal Apendiks" adı verilen proteinöz yapılar bulunur. Özellikle CEP164 gibi distal apendiks proteinleri, siliyer veziküllerin (ciliary vesicles) yakalanması ve membrana füzyonu için bir platform oluşturur [48]. Siliyogenezin ilk aşamasında, Golgi aygıtından gelen ve siliyer membran proteinlerini taşıyan veziküller, bazal cismin ucunda birleşerek siliyer vezikül (Ciliary Vesicle) oluşturmaya başlar. Bu süreçte, Rabin8 (Rab8 GTPaz için guanin nükleotid değişim faktörü) ve TRAPPII kompleksleri tarafından yönetilen vezikül trafiği kritik öneme sahiptir. Golgi'den siliyer tabana doğru gerçekleşen bu polarize taşıma, silyumun lipid ve protein kompozisyonunu belirler.

2.1.2 Geçiş Bölgesi (Transition Zone)

Bazal cismin distal ucu ile aksonemin başlangıcı arasında yer alan "Geçiş Bölgesi" (Transition Zone (TZ)), siliyer kompartımanı sitoplazmadan ayıran, yapısal ve fonksiyonel olarak özelleşmiş bir sınır bölgesidir [156]. Geçiş bölgesi, siliyer kompartımanın özgünlüğünü iki temel mekanizma aracılığıyla korumaktadır:

- **Sitoplazmik-Siliyer Protein Ayrımı (Protein Gating):** TZ, silyum içine giren proteinler için fiziksel ve biyokimyasal bir elek görevi görür. 50 kDa'dan küçük proteinler pasif difüzyonla geçebilirken; daha büyük proteinlerin ve veziküllerin geçişi, siliyer sinyaller (CTS) ve IFT sistemi aracılığıyla, enerji bağımlı olarak gerçekleşir. Siliyer sinyal reseptörlerinin (Smoothened, Polycystin-1/2, SSTR3) silyum içinde tutulması bu bariyer sayesinde sağlanır [156].
- **Lipid Bariyeri ve Membran Kompozisyonu:** Siliyer membran, plazma zarının devamı olmasına rağmen fosfoinositid içeriği açısından farklılaşmıştır. TZ bölgesinde bulunan Septin halkası (Septin ring), sitoplazmik membrandaki PI(4,5)P2'nin silyum içine difüzyonunu engelleyerek, siliyer membranın PI(4)P açısından zengin lipid kimliğini korumasını sağlamaktadır.

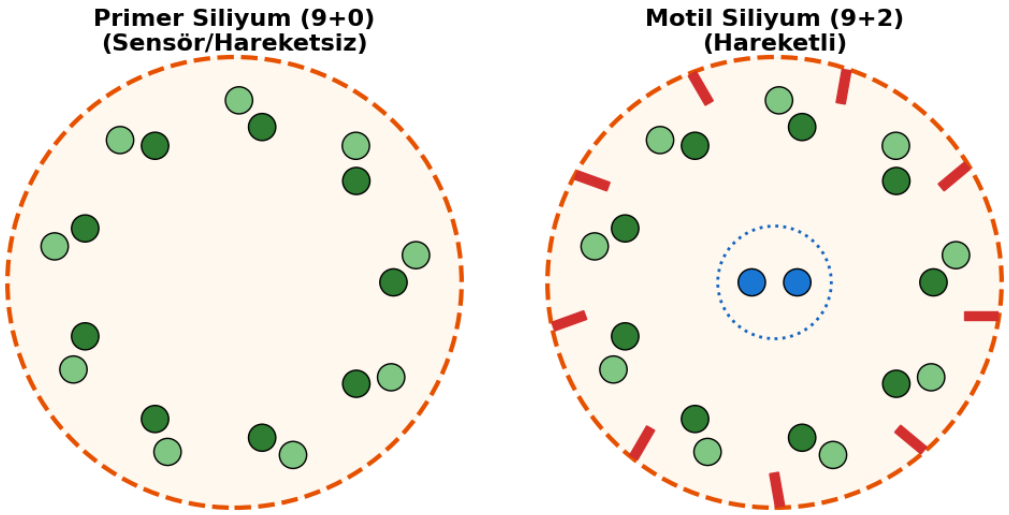
Genetik açıdan TZ, siliyopatiler için bir mutasyon "sıcak bölgesi"dir (hotspot). MKS veya JBTS modülündeki tek bir genin (örn. *TMEM67*) kaybı, bariyer fonksiyonunun bozulmasına (leaky cilium) neden olur.

2.1.3 Aksonem

Silyumun ana iskeletini, bazal cisimden uzayan mikrotübül demetleri, yani aksonem oluşturmaktadır. Aksonem, evrimsel olarak korunmuş 9 adet çevresel mikrotübül çiftinden (dublet) meydana gelmektedir. Her bir dublet, 13 protofilamentli A-tübülü ve ona yapışık 10 protofilamentli B-tübülünden oluşur. Bu iskelet, stabiliteyi artırmak ve protein trafiğini düzenlemek amacıyla yoğun post-translasyonel modifikasyonlara, özellikle tübülin asetilasyonu ve glutamilasyonuna uğrar [157].

Silyumun fonksiyonel kimliği (motil veya non-motil olması), bu temel iskelet üzerine eklenen özelleşmiş protein kompleksleri ile belirlenir. Motil silyumda, mikrotübül çiftleri üzerinde ATP enerjisini mekanik güce dönüştüren iç ve dış dinein kolları ile hareketi düzenleyen radyal uzantılar (*radial spokes*) ve neksin-dinein düzenleyici kompleks (N-DRC) bulunur. Bu kinetik birimler, mikrotübüllerin birbirleri üzerinde kaymasını (sliding) sağlayarak siliyer motiliteyi sağlar [157, 158].

Buna karşın, primer silyumda merkezi mikrotübül çifti ve dinein kolları yer almaz. Hareket kabiliyetinin kaybolması, organelin duyusal süreçler için özelleşmesini sağlamıştır. Primer ve motil silyumun aksonemal mikrotübül düzenindeki yapısal ve moleküler farklılıklar Şekil 2.1'te şematize edilmiştir.



A. Primer Silyum (9+0) B. Motil Silyum (9+2)

Şekil 2.1 Silyum tiplerinin aksonemal enine kesit şemaları. (A) Primer silyum (B) Motil silyum (Görsel: Yazar, Matplotlib)

2.1.4 İntraflagellar Taşıma (IFT) Mekanizması

Siliyer aksonem ve membran, protein sentezinin gerçekleştiği sitoplazmadan "Geçiş Bölgesi" bariyeri ile ayrılmıştır. Silyum içerisinde ribozom bulunmadığı için silyumun uzaması (elongation), yapısal idamesi ve sinyal iletimi için gerekli tüm yapıtaşlarının (tubulin, reseptörler, enzimler) siliyer tabandan uca doğru aktif olarak taşınması zorunludur. Bu çift yönlü, enerji bağımlı ve motor proteinler tarafından yürütülen moleküler kargo taşımacılığına "İntraflagellar Transport (IFT)" adı verilir [160].

IFT mekanizması, mikrotübülüzerinde hareket eden motor proteinler ve yükü taşıyan "IFT partikülleri" (IFT-A ve IFT-B kompleksleri) arasındaki dinamik etkileşime dayanmaktadır:

- **Anterograd (İleri) Taşıma:** Siliyer tabandan uca (Base-to-Tip) doğru gerçekleşen bu hareket, heterotrimerik Kinesin-2 motor kompleksi (KIF3A, KIF3B ve KAP3) tarafından yürütülür. Bu motorlar, IFT-B kompleksi (örn. IFT88, IFT52, IFT20) ile birleşerek IFT partikül dizilerini oluşturur. IFT-B kompleksinin temel görevi, sitoplazmadan gelen tubulin dimerlerini yakalayarak silyum ucundaki polimerizasyon bölgesine taşımaktır. IFT-B veya Kinesin-2 genlerindeki mutasyonlar, silyumun hiç oluşamamasına (siliyogenez yetmezliği) neden olmaktadır [159].

- **Retrograd (Geri) Taşıma:** silyum ucundan tabana gerçekleşen geri dönüş hareketi, Sitoplazmik Dinein-2 motor kompleksi (ağır zincir DYNC2H1 ve hafif zincirler) tarafından sağlanır. Bu süreci yöneten IFT-A kompleksi, silyum ucunda işlevi biten ürünlerin, yıkılacak sinyal reseptörlerinin ve IFT bileşenlerinin sitoplazmaya geri döndürülmesinden sorumludur. IFT-A defektleri, silyumun kısalmasına veya ucunda materyal birikmesine (bulbed fenotipi) yol açarak, özellikle iskelet siliyopatileri (örn. Jeune Sendromu) ile sonuçlanmaktadır [161, 119].

- **BBSome Kompleksi**: IFT kompleksleri tek başlarına membran proteinlerini tanıyamamaktadır. BBSome kompleksi, IFT kompleksleri ile siliyer membran arasında bir köprü kurarak; özellikle G-protein kenetli reseptörlerin (GPCR) (örn. SSTR3, MCHR1) silyuma girişini ve çıkışını düzenleyerek, sinyal yolağının hassasiyetini kontrol eder [164, 154].

Bu döngüdeki en ufak bir aksama, siliyer trafik sıkışıklığına ve organel kaybına neden olmaktadır.

2.2 Siliyer Sinyal İletim Mekanizmaları

Primer silyum, hücre dışı ortamdan gelen mekanik (sıvı akışı) ve biyokimyasal (ligand) uyarıların algılanmasında, hücre içi sinyal transdüksiyonunda görev alan özelleşmiş bir organeldir. Siliyer membran, spesifik reseptörlerin ve efektör moleküllerin yüksek konsantrasyonda biriktiği, sitoplazmadan izole edilmiş bir Ca^{2+} ve cAMP mikro-alanı (microdomain) oluşturur. Bu yapı, Hedgehog (Hh), Wnt, PDGF ve Notch gibi gelişimsel yolakların spatiotemporal kontrolünü sağlamaktadır [16].

2.2.1 Hedgehog (Hh) Sinyal Yolağı

Hedgehog (Hh) sinyal yolağı, siliyer yapının varlığına ve İntraflagellar Transport (IFT) mekanizmasının işleyişine mutlak bağımlılık gösteren temel sinyalizasyon mekanizmasıdır. Embriyonik gelişimde nöral tüp kapanması, ekstremite oluşumu ve doku polaritesinin belirlenmesinde merkezi rol oynar [173, 175].

Yolağın aktivasyonu, reseptör ve transkripsiyon faktörlerinin silyum içindeki dinamik translokasyonu ile gerçekleşmektedir:

- **Basal Durum (Ligand Yokluğu):** Hh ligandının yokluğunda, reseptör *Patched1 (Ptc1)* siliyer membranda lokalize olur ve sinyal transduseri olan *Smoothened (Smo)* proteininin silyum içine girişini inhibe eder. Bu süreçte, Gli transkripsiyon faktörleri (Gli2/3), silyum tabanında *SuFu* (Suppressor of Fused) ve *Kif7* kontrolünde proteozomal yıkıma uğrayarak transkripsiyonel baskılayıcı (Gli Repressor - GliR) formlarına dönüşür.
- **Aktif Durum (Ligand Varlığı):** Ligand (Sonic Hedgehog - Shh) bağlanması ile Ptc1 silyumdan uzaklaştırılır. İnhibisyonun kalkmasıyla Smo, siliyer veziküller ve IFT sistemi aracılığıyla siliyer membrana taşınırak birikir. Smo birikimi, Gli proteinlerinin tam uzunluktaki aktif forma (Gli Activator - GliA) dönüşmesini sağlar. Aktif Gli, silyum ucundan çekirdeğe taşınarak hedef gen ekspresyonunu başlatır.

Bu moleküler trafik, siliyer motor proteinlerin (Kinesin / Dinein) ve vezikül füzyon mekanizmalarının koordinasyonunu gerektirmektedir [134].

2.2.2 Wnt Sinyal Yolağı

Wnt sinyalizasyonu, embriyonik gelişimde vücut ekseninin belirlenmesi ve erişkin doku homeostazında kök hücrelerin korunması için elzemdir. Bu yolak, β-katenin

bağımlı (kanonik) ve β-katenin bağımsız (non-kanonik / Düzlemsel Hücre Polaritesi (PCP)) olmak üzere iki ana kolda incelenmektedir. Primer silyum, bu iki yolak arasındaki dengeyi düzenleyen bir "moleküler şalter" (switch) görevi görmektedir [**wnt**, 179, 12].

- **Kanonik Wnt Sinyalizasyonu:** Siliyer bazal cisim bölgesi, sitoplazmik β-katenin seviyesini kontrol eden yıkım kompleksinin (APC, Axin, GSK3β) regülasyonunda görev alır. Siliyer disfonksiyon durumlarında, β-katenin yıkılamaz ve çekirdekte birikerek *c-Myc* ve *Cyclin D1* gibi proliferatif genleri aşırı aktive eder. Bu kontrolsüz hücre bölünmesi, özellikle böbrek tübül epiteli hücrelerinde lümen genişlemesine ve polikistik böbrek hastalığı patogenezine yol açar [79, 177, 179].
- **Düzlemsel Hücre Polaritesi (PCP):** Hücrelerin doku düzlemindeki yönelimini (oryantasyonunu) düzenleyen PCP yolağı, siliyer proteinler (Inversin, Vangl2, Prickle) aracılığıyla kontrol edilir. Özellikle *Inversin (INVS)* geni, siliyer proksimal bölgede (Inversin Kompartımanı) lokalize olarak, Dishevelled (Dvl) proteinini degrade eder ve böylece kanonik Wnt yolağını baskılayıp non-kanonik PCP yolağını aktive eder. Bu mekanizma, hücrelerin doku düzlemindeki yönelimini düzenler. Bu mekanizma sayesinde, iç kulaktaki tüy hücrelerinin stereosilya demetleri uygun polaritede hizalanırken; böbrek epitelyumunda hücre bölünme ekseni kontrol edilerek tübüler genişleme ve kist oluşumu süprese edilir [79, 178].

2.2.3 PDGF Sinyal Yolağı

Trombosit Türevli Büyüme Faktörü (PDGF) sinyal yolağı, özellikle mezenkimal kökenli hücrelerin (fibroblastlar, düz kas hücreleri) migrasyonu ve proliferasyonu için gereklidir. PDGF reseptörü alfa (PDGFRα), hücre döngüsünün G0 fazında primer silyum membranına spesifik olarak lokalize olur.

Ligand (PDGF-AA) bağlanması, reseptörün dimerizasyonunu ve oto-fosforilasyonunu tetikler. Bu aktivasyon, silyum içinde PI3K/Akt ve MEK/ERK (MAPK) gibi aşağı akım (downstream) kinaz enzimini aktive eder. Reseptörün siliyer membrana hedeflenmesi, Golgi kaynaklı veziküler taşıma mekanizmalarına bağımlıdır. Bu süreçte görev alan vezikül trafiği proteinlerindeki (örn. *VPS* ailesi) veya siliyer yapıdaki (örn. *IFT* proteinleri) defektler, reseptörün silyuma ulaşamamasına ve hücrenin kemotaktik yanıtının (yara iyileşmesi vb.) kaybına yol açmaktadır [15].

2.3 Siliyopati Spektrumu

Silyumun yapısal bileşenlerini (aksonem, bazal cisim, geçiş bölgesi) veya fonksiyonel mekanizmalarını (İntraflagellar Transport, vezikül trafiği, sinyal transdüksiyonu) kodlayan genlerdeki mutasyonlar, "Siliyopatiler" adı verilen geniş ve heterojen bir hastalık grubuna neden olmaktadır. Silyumun insan vücudundaki hemen hemen tüm hücre tiplerinde bulunması (ubiquitous) ve üstlendiği kritik gelişimsel roller nedeniyle, bu hastalıklar multisistemik, kronik ve ilerleyici bir fenotip sergiler [147].

Klinik tablo; embriyonik dönemde letal olan ağır gelişimsel bozukluklardan (örneğin Meckel-Gruber Sendromu), erişkin dönemde ortaya çıkan dejeneratif bozukluklara (örn. Retinitis Pigmentosa) kadar uzanan geniş bir yelpazededir. Siliyopatiler, etkilenen organelin motilite yeteneğine (hareketli veya hareketsiz) ve klinik bulguların dağılımına (sendromik veya izole) göre sınıflandırılmaktadır.

2.3.1 Motil Siliyopatiler: Primer Siliyer Diskinezi

Motil silyumun biyogenezi, ultrastrüktürel yapısı veya aksonemal motilite mekanizmasındaki defektlerden kaynaklanan bozukluklar, “Hareketli Siliyopatiler” başlığı altında incelenmektedir [35]. Bu grubun en iyi klinik yansıması, otozomal resesif kalıtım gösteren ve canlı doğumlarda yaklaşık 1/30.000 sıklığında görülen Primer Siliyer Diskinezi’dir. [10].

Primer Siliyer Diskinezi (PCD) solunum yolu, beyin ventrikülleri ve üreme sistemi gibi motil silyum içeren dokulardaki fonksiyon kaybıyla seyreden, genetik olarak heterojen bir hastalıktır. PCD patolojisinin temelinde, siliyer hareketin koordinasyonunu sağlayan aksonemal motor proteinlerin eksikliği veya yapısal bozukluğu yatmaktadır. Sağlıklı bir motil silyum, merkezde iki tekli mikrotübül ve çevrede dokuz adet mikrotübül çiftinden oluşan karakteristik "9+2" aksonem düzenine sahiptir. PCD vakalarında, bu düzendeki sapmalar aksonemin yapısal bütünlüğünü bozarak siliyer hareketin durmasına veya bozulmasına neden olur. Siliyer hareketin (motilite) enerji kaynağı ve mekanik gücü, mikrotübül çiftleri üzerine yerleşmiş olan ve ATP hidrolizi yapan Dinein Kolları (Dynein Arms) tarafından sağlanır. Bu kolların genetik mutasyonlar sonucu gelişen yokluğu, mikrotübüller arasındaki kayma hareketini engelleyerek silyumun biyomekanik işlevini tamamen yitirmesine yol açar [10, 34].

PCD vakalarının moleküler etyolojisi incelendiğinde, ultrastrüktürel düzeyde dört temel defekt grubu gözlemlenmektedir:

1. **Dış Dinein Kolu (ODA) Defektleri:** Ana motor kompleks olan Dış Dinein Kollarının eksikliği, silyumun hareketsiz kalmasına neden olur. ODA kompleksinin silyuma taşınmasından sorumlu olan *DNAH5* ve *DNAI1* genlerindeki mutasyonlar, PCD vakalarının yaklaşık %30-40'ını oluşturur [47].

2. **İç Dinein Kolu (IDA) Defektleri:** İç Dinein Kollarının kaybı, silyumun titreşmesine rağmen efektif bir sıvı akışı oluşturamamasına (diskinezi) yol açar. *CCDC39* ve *CCDC40* genlerindeki mutasyonlar, vücutta "moleküler cetvel" görevi gören proteinlerin kaybına neden olur [41].

3. **Radyal Spoke ve Merkezi Çift Defektleri:** Çevresel mikrotübül çiftleri ile merkezi çift arasındaki sinyal iletimini sağlayan Radyal Spoke yapılarının bozulması, siliyer hareketin dairesel (rotasyonel) forma dönüşmesine neden olur (*RSPH* gen ailesi).

4. **Dinein Düzenleyici Kompleks (N-DRC) Defektleri:** Mikrotübül çiftleri arasındaki kayma hareketini sınırlayan neksin bağlarındaki defektler, mikrotübül kaymasının bükülme hareketini bozar. Bu durum, siliyer vuruşun paternini kısıtlayarak silyumun verimsiz titreşmesine neden olur.

Motil silyumun temel fizyolojik görevi, solunum yollarında mukus tabakasının orofarenkse doğru süpürülmesini sağlayan "Mukosiliyer Klirens" mekanizmasını yürütmektir. Sağlıklı silyum, "Metakronal Dalga" (sıralı ve koordineli vuruş) hareketiyle mukusu taşır. PCD hastalarında bu mekanizmanın kaybı, mukus stazına (birikimine) ve bakteriyel kolonizasyona yol açar [88, 11, 34]:

- **Neonatal Respiratuar Distres:** Yenidoğan döneminde açıklanamayan solunum sıkıntısı.
- **Kronik Enfeksiyonlar:** Sürekli mukus birikimine bağlı rekürren rinosinüzit ve otitis media.
- **Bronşektazi:** Kronik inflamasyon sonucu bronş duvarlarının kalıcı hasarı.
- **İnfertilite:** Erkeklerde sperm motilite kaybı, kadınlarda tübal disfonksiyon.

2.3.1.1 Kartagener Sendromu

PCD hastalarının yaklaşık %50'sinde, situs anomalileri ile karakterize Kartagener Sendromu (Situs Inversus Totalis, Kronik Sinüzit, Bronşektazi) görülmektedir

[181]. Siliyer disfonksiyon ile organ yerleşimi arasındaki bu ilişki, gelişimsel biyolojide "Nodal Akış Teorisi" ile açıklanmaktadır [12].

Gastrulasyon evresindeki memeli embriyosunun ventral yüzeyinde, "Hensen Düğümü" (Node) bölgesinde bulunan hücreler, "9+0" düzeninde olmalarına rağmen dinein kolları içeren özelleşmiş "Nodal silyum" taşımaktadır. Sağlıklı bir embriyo gelişiminde sol-sağ asimetrisinin (lateralite) belirlenmesi şu aşamalarla gerçekleşir [45, 12]:

1. **Mekanik Akışın Oluşturulması:** Nodal silyum, posterior eğimleri sayesinde saat yönünde dönerek, ekstraembriyonik sıvıda (perinodal sıvı) sağdan sola doğru bir laminer akış (Nodal Flow) oluşturur.
2. **Mekanik ve Kimyasal Algılama:** Bu sol yönlü akış, düğümün sol kenarında bulunan immotil mekanosensör silyum (PKD1L1 ve PKD2 proteinlerini içeren) tarafından algılanır. Akışın yarattığı mekanik stres, bu silyum üzerindeki kalsiyum kanallarını açarak hücre içi kalsiyum (Ca^{2+}) konsantrasyonunu artırır.
3. **Sinyalizasyon Yolağının Aktivasyonu:** Sol taraftaki kalsiyum artışı; *Nodal*, *Lefty2* ve *Pitx2* gibi asimetri belirleyici genlerin ekspresyonunu tetikler. Özellikle *Pitx2*, organogenez boyunca sol taraf kimliğinin korunmasını sağlayan ana transkripsiyon faktörüdür.

Kartagener Sendromunda, dinein kollarındaki (ODA / IDA) mutasyonlar nedeniyle nodal silyum hareketsizdir (immotil). Yönlendirici bir akışın yokluğunda, morfogenler rastgele dağılırak, sol-sağ ekseni şansa bağlı (stokastik) olarak belirlenir. Bu durum, hastaların yarısında organların ayna hayali şeklinde ters yerleştiği *Situs Inversus Totalis* tablosuna yol açar [181]. Bu mekanizma, siliyer proteinlerin sadece hücresel düzeyde değil, organizmanın makro-anatomik planının kurulmasında da merkezi bir rol oynadığını göstermektedir.

2.3.2 Non-motil Siliyopatiler

Primer silyumun biyogenezi, yapısal idamesi veya hücre dışı sinyalleri (Hedgehog, Wnt, PDGF, Notch, mTOR) algılama ve iletme kapasitesindeki bozukluklar, "Sensör Siliyopatiler" veya yapısal tanımıyla "Hareketsiz Siliyopatiler" grubunu oluşturur [16, 147]. Primer silyum, embriyonik gastrulasyon evresinden erişkin doku homeostazına kadar çok sayıda moleküler süreci koordine eden merkezi bir sinyalizasyon organeli olduğundan, bu organelin disfonksiyonu sadece lokal bir

doku hasarına değil; ektoderm, mezoderm ve endoderm kökenli çoklu sistemleri eş zamanlı etkileyen pleiotropik sendromlara yol açmaktadır. Bu hastalık grubu, klinik bulguların şiddeti ve tutulan organ sistemlerine göre; perinatal dönemde letal (ölümcül) seyreden ağır formlardan, yaşam boyu süren kronik metabolik ve dejeneratif bozukluklara kadar geniş bir spektrum oluşturmaktadır.

2.3.2.1 Bardet-Biedl Sendromu

Bardet-Biedl Sendromu (BBS), yüksek genetik heterojeniteye sahip, otozomal resesif kalıtım gösteren bir siliyopatidir. Siliyer membran proteinlerinin hücre içi trafiğini ve siliyer lokalizasyonunu yöneten mekanizmalardaki bozukluklardan kaynaklanır. Hastalığın prevalansı, genel popülasyonda yaklaşık 1/160.000 iken, izole toplumlarda 1/13.500'e kadar yükselebilmektedir [168].

Bardet-Biedl Sendromu'nun (BBS) moleküler patogenezi, evrimsel süreçte yüksek oranda korunmuş oktamerik bir yapı olan BBSome kompleksi etrafında şekillenmektedir. Kriyoelektron mikroskopi (Cryo-EM) ve X-ışını kristalografisi analizleri; bu kompleksin, COPI ve COPII proteinleri ile sergilediği yapısal homoloji nedeniyle siliyer membran proteinlerine özgü bir kargo adaptörü olarak işlev gördüğünü ortaya koymuştur . Hücresel düzeyde BBSome; Golgi aygıtından tomurcuklanan siliyer kargoların (ciliary cargo) (örn. *SSTR3*, *MCHR1*), Rab8 GTPaz ve HOPS kompleksi eşliğinde siliyer tabana yönlendirilmesini koordine eder [164, 163].

Klinik açıdan sendromun en karakteristik yansıması, %90'ın üzerinde penetrans ile gözlenen Retinitis Pigmentosa (RP) tipi retinal dejenerasyondur. Bu patolojinin biyolojik temeli, fotoreseptör dış segmentinin modifiye bir primer silyum olması esasına dayanmaktadır. Rodopsin gibi kritik siliyer bileşenlerin "Bağlantı Silyumu" (Connecting Cilium) üzerinden taşınması, IFT ve BBSome kompleksleri aracılığıyla gerçekleştirilir. Bu düzenekteki defektler sonucu proteinlerin yanlış lokalizasyonu; fotoreseptörlerde proteotoksik strese ve apoptoza yol açarak görme kaybı sürecini tetikler [164, 168, 169]. Benzer şekilde obezite, hipotalamik nöronlarda Leptin reseptörlerinin siliyer lokalizasyon kusurları ile ilişkilidir. Polidaktili, ekstremite morfogenezi sırasında Shh (Sonic Hedgehog) yolağının siliyer regülasyonundaki aksaklıklardan; böbrek anomalileri ise renal tübüllerdeki Wnt sinyalizasyonu ve homeostaz bozukluklarından kaynaklanmaktadır.

2.3.2.2 Joubert Sendromu

Joubert Sendromu (JBTS), beyin sapı ve serebellum (beyincik) gelişimindeki malformasyonlarla tanımlanan, genetik ve klinik olarak yüksek heterojenite gösteren bir nörogelişimsel bozukluktur. İlk kez 1969 yılında Marie Joubert tarafından tanımlanan sendromun prevalansı yaklaşık 1/80.000-1/100.000 arasındadır [170].

Hastalığın genetik altyapısı incelendiğinde, JBTS fenotipine neden olduğu bilinen 35'ten fazla genin ürünlerinin (JBTS proteinleri), primer silyumun en stratejik noktası olan Geçiş Bölgesinde (Transition Zone (TZ)) lokalize olduğu görülür. Bazal cismin hemen üzerinde yer alan ve siliyer membranı aksoneme bağlayan "Y-bağları" (Y-links) ile karakterize bu bölge, siliyer kompartımanı sitoplazmadan ayıran seçici geçirgen bir moleküler kapı işlevi görmektedir [170].

JBTS proteinleri, geçiş bölgesinde büyük makromoleküler kompleksler (MKS ve NPHP modülleri) oluşturarak siliyer membran lipit kompozisyonunu korur ve siliyer sinyal reseptörlerinin giriş çıkışını denetler [171]. Bu proteinlerdeki mutasyonlar, geçiş bölgesi fonksiyonunun çökmesine (leaky cilium) neden olur. Sonuç olarak, Sonic Hedgehog (Shh) ve Wnt gibi nöronal gelişim, akson yönelimi ve nöronal göç süreçlerini yöneten sinyal yolakları deregüle olur. Özellikle serebellar granül hücre prekürsörlerinin proliferasyonundaki bozukluk ve Aksonal deküssasyon defekti, hastalığın nöroanatomik temelini oluşturur [179].

JBTS'nin tanısı, beyin Manyetik Rezonans (MR) görüntülemesinde saptanan ve beyin sapının aksiyel kesitlerinde bir azı dişini andıran patognomonik "Molar Diş Belirtisi" (Molar Tooth Sign) ile konulur. Bu karakteristik görüntü üç ana anomaliden kaynaklanır: (1) Serebellar Vermis Hipoplazisi: Beyinciğin iki yarısını birleştiren orta hat yapısının gelişmemesi; (2) Derin İnterpedinküler Fossa: Beyin sapında anormal çukurlaşma; (3) Kalınlaşmış ve Yatay Seyirli Süperior Serebellar Pedinküller. Bu nöroanatomik hasarların klinik yansıması; bebeklik döneminde belirgin hipotoni, serebellar ataksi, düzensiz solunum paterni ve okülomotor apraksi şeklindedir. Silyumun ubiquiter doğası gereği, JBTS hastalarında nörolojik bulgulara ek olarak retinal distrofi, kistik böbrek hastalığı ve polidaktili gibi ekstranöral organ tutulumları da sıkça görülmektedir [170].

2.3.2.3 Meckel-Gruber Sendromu

Siliyopati spektrumunun fenotipik olarak en ağır ve genellikle perinatal dönemde letal (ölümcül) seyreden formu Meckel-Gruber Sendromu'dur (MKS) [183]. Otozomal resesif kalıtım gösteren bu sendrom, siliyer fonksiyon kaybının

organogenez (organ gelişimi) üzerindeki kritik etkisini en net gösteren tablodur. Genetik açıdan yüksek heterojenite gösteren hastalığa, bugüne kadar tanımlanmış 15'ten fazla genin (örn. *MKS1*, *TMEM67*, *RPGRIP1L*, *CEP290*) mutasyonları neden olmaktadır.

Hastalığın moleküler patogenezi, siliyer tabanda (geçiş bölgesi) "MKS Modülü" veya "B9 Kompleksi" olarak adlandırılan büyük protein ağının bozulmasına dayanır. Bu modül, bazal cismi plazma zarına sabitleme ve siliyer kompartıman ile sitoplazma arasındaki bariyer işlevini düzenlemeden sorumludur. MKS genlerindeki fonksiyon kaybı mutasyonları, bu bariyerin kurulamamasına ve siliyer membran proteinlerinin silyuma hedeflenememesine yol açar. Sonuç olarak, hücreler silyum oluşturma yeteneğini büyük ölçüde kaybederek siliyogenez yetmezliği gösterir veya fonksiyonel olmayan silyum üretir [184].

Silyumun yokluğu veya ağır yapısal kusurları, embriyonik gelişimin erken evrelerinde morfojenetik sinyal yolaklarının (özellikle kanonik olmayan Wnt/PCP ve Sonic Hedgehog) tamamen sessizleşmesine neden olur. Bu moleküler blokaj, MKS'nin patognomonik klinik triadını oluşturur: (1) Shh sinyalizasyonunun nöral plaka üzerindeki etkisinin kaybı sonucu gelişen "Oksipital Ensefalosel"; (2) Fetal idrar üretiminin olmaması ve buna bağlı oligohidramnios / akciğer hipoplazisi ile sonuçlanan "Kistik Displastik Böbrekler"; ve (3) Ekstremite tomurcuğundaki Shh gradyanının bozulmasıyla ortaya çıkan "Postaksiyel Polidaktili". Ayrıca, vakaların çoğunluğunda görülen karaciğerde "Duktal Plak Malformasyonu", siliyer sinyalizasyonun doku morfogenezindeki rolünü kanıtlar niteliktedir [68].

2.3.2.4 Leber Konjenital Amorozisi

Genetik körlüğün en erken başlangıçlı ve en ağır formu olan Leber Konjenital Amorozisi (LCA), temelinde fotoreseptör hücrelerindeki siliyer fonksiyon bozukluğunun yattığı bir hastalıktır. Bu hastalığın moleküler merkezinde; metabolik merkez olan "iç segment" ile ışığın algılandığı "dış segment" (modifiye silyum) arasında köprü görevi gören "Bağlantı Silyumu" (Connecting Cilium) yer alır [71]. Fotoreseptör dış segmenti, her gün toplam kütlesinin 10%'u kadar yenilenen oldukça dinamik bir yapıya sahiptir. Ancak rodopsin ve fototransdüksiyon için gereken diğer temel bileşenler bu segmentte sentezlenemez. Bu durum, iç segmentte üretilen proteinlerin bağlantı silyumu üzerinden IFT mekanizması aracılığıyla dış segmente kesintisiz taşınmasını zorunlu kılar. *CEP290*, *LCA5* (Lebercilin) ve *RPGRIP1* gibi genlerdeki mutasyonlar, bu lojistik trafiğin tıkanmasına ve bağlantı silyumundaki "geçiş bölgesi" (transition zone) bariyer bütünlüğünün kaybına yol açar. Opsinlerin dış segmente ulaşamayarak iç

segmentte birikmesi, proteotoksik strese ve fotoreseptörlerin hızlı dejenerasyonu ile sonuçlanan apoptoz sürecini tetikler. Bu nedenle LCA, güncel literatürde klasik bir göz hastalığından ziyade, sistemik siliyopatilerin göze özgü bir yansıması olan "siliyer transport hastalığı" olarak tanımlanmaktadır [71, 53].

2.3.2.5 Nefronofitiz

Çocukluk ve ergenlik döneminde böbrek yetmezliğinin en yaygın kalıtsal nedeni olan Nefronofitiz (NPHP), kistik böbrek hastalıkları arasında kendine özgü klinik ve histolojik özellikleriyle ayrışmaktadır [186]. Polikistik Böbrek Hastalığının (PKD) aksine NPHP vakalarında böbrek boyutları genellikle korunur veya atrofiye bağlı küçülme gözlenir. Bununla birlikte parankim dokusunda, yoğun interstisyel fibrozis, tübüler bazal membranda düzensiz kalınlaşma ve karakteristik kortikomedüller kist oluşumu ile seyreden ağır bir yapısal dejenerasyon izlenir [185].

Bu patolojik tablonun moleküler temelinde, böbrek tübül epitel hücrelerinin lümenine uzanan primer silyumun mekanosensör işlevini yitirmesi yatmaktadır. Sağlıklı bir fizyolojide primer silyum, idrar akışının yarattığı mekanik kayma kuvvetini (shear stress) algılayarak hücre içine kalsiyum (Ca^{2+}) girişini sağlar. Bu sinyal iletimi, hücre içi cAMP düzeylerini baskılamanın yanı sıra, inversin proteini aracılığıyla β-katenin yıkımını tetikleyerek kanonik Wnt yolağını kontrol altında tutar. NPHP1, NPHP4 ve Inversin (INVS) gibi proteinler, silyumun "İnversin Kompartımanı" olarak adlandırılan bölgesinde lokalize olarak bu sinyalizasyon hiyerarşisini yönetir [185, 185].

NPHP genlerindeki mutasyonlar, silyumun akış sensörü özelliğini ortadan kaldırarak Wnt yolağının kontrolsüz aktivasyonuna yol açar. Bu moleküler sapma, tübül hücrelerinin yönlendirilmiş hücre bölünmesi (planar cell polarity) yeteneğini bozmaktadır. Hücrelerin tübül eksenine paralel bölünmek yerine rastgele oryantasyonla çoğalması, tübüllerin boyuna uzamak yerine enine genişlemesine, kistik yapıların oluşmasına ve nihayetinde fibrotik sürecin tetiklenmesine neden olmaktadır [76, 76].

2.3.2.6 Senior-Løken Sendromu

Siliyopatilerin klinik ve moleküler düzeyde bir "spektrum" oluşturduğunun en çarpıcı örneği; Nefronofitiz (NPHP) ile retinal dejenerasyonun (LCA veya Retinitis Pigmentosa) eş zamanlı gözlendiği Senior-Løken Sendromu'dur (SLS). Bu sendrom, siliyer proteinlerin böbrek tübülü ve retina fotoreseptörü gibi farklı

dokularda benzer ancak özelleşmiş görevler üstlendiğini kanıtlayan model hastalık niteliğindedir [81, 186, 53].

Hastalığın genetik altyapısı incelendiğinde, siliyer geçiş bölgesi (transition zone) proteinlerinin dokuya özgü modüler bir organizasyon sergilediği görülmektedir. Özellikle *NPHP1*, *NPHP3*, *NPHP4* ve *IQCB1/NPHP5* genlerinin ürünleri; hem böbrek epitel hücrelerinin primer silyum hem de fotoreseptörlerin bağlantı silyumu (*connecting cilium*) yapısında lokalize olur. Ancak bu proteinlerin fonksiyonel eşikleri dokular arasında değişkenlik gösterir. Örneğin NPHP5 proteinini kodlayan *IQCB1* geni, *CEP290* ile doğrudan etkileşime girerek fotoreseptörlerdeki dış segmente protein trafiğini yönetmektedir. Bu genetik etkileşimdeki bozulmalar, retinada erken ve ağır bir yıkıma (LCA) yol açarken; böbrekte daha geç başlangıçlı ve sinsi ilerleyen bir fibrozis (NPHP) tablosuna neden olmaktadır [81].

SLS etyolojisinde merkezi bir rol oynayan *CEP290* mutasyonları, genetik literatüründeki "pleiotropi" kavramı için karakteristik bir örnektir. Bu gen, böbreklerde Wnt sinyal yolağının disregülasyonu sonucu kist oluşumunu tetiklerken, retinada rodopsin transportunu engelleyerek fotoreseptör kaybını başlatmaktadır. SLS vakalarında gözlenen bu pleiotropik etki, siliyopatilerin klinik değerlendirmesinde tek bir organ odaklı yaklaşım yerine, bütüncül değerlendirmenin önemini göstermektedir [77, 81, 85].

2.3.2.7 Jeune Sendromu (Asfiksiyen Torasik Distrofi)

İskelet siliyopatileri grubunun prototipi olan Jeune Sendromu veya Asfiksiyen Torasik Distrofi (ATD), otozomal resesif kalıtım gösteren nadir bir hastalıktır. Sendromun klinik karakterizasyonu; dar ve çan biçimindeki göğüs kafesi, kısa kaburgalar, kısalmış ekstremiteler (mikromeli) ve polidaktilidir. Göğüs kafesindeki darlık, akciğer gelişimini kısıtlayarak (pulmoner hipoplazi) yenidoğan döneminde solunum yetmezliğine neden olmaktadır. İskelet bulgularına ek olarak vakaların bir kısmında progresif nefronofitiz, karaciğer fibrozisi ve retinal dejenerasyon gibi diğer siliyopati semptomları da eşlik edebilmektedir [119, 187].

Hastalığın moleküler patogenezinde, siliyer retrograd taşıma mekanizmasındaki aksaklıklar merkezi rol oynar. Özellikle sitoplazmik Dynein-2 motor kompleksi ağır zincirini kodlayan *DYNC2H1* geni veya IFT-A kompleksine ait genlerdeki (*IFT140*, *IFT122* vb.) mutasyonlar, siliyer aksonem boyunca protein trafiğini bozar [117]. Retrograd trafiğin tıkanması sonucu siliyer uçta protein birikimi meydana gelerek morfogenez için kritik olan sinyal iletimi engellenir. Bu süreçte, kıkırdak büyüme plağındaki kondrositlerde Indian Hedgehog (Ihh) sinyalinin iletilememesi,

kondrosit proliferasyonunu ve farklılaşmasını durdurarak endokondral kemikleşme sürecini sekteye uğratır. Sonuç olarak, kemik dokusunun boyuna büyümesini ve ekstraselüler matriks organizasyonunu yöneten genetik yapının bozulması, iskelet displazilerine yol açmaktadır [126, 161, 124].

2.3.2.8 Ellis-van Creveld Sendromu

Ellis-van Creveld (EvC) sendromu, orantısız cücelik (kısa ekstremiteler), postaksiyel polidaktili, ektodermal displazi (tırnak ve diş anomalileri) ve konjenital kalp defektleri (özellikle atriyal septal defekt) tetradı ile karakterize bir kondro-ektodermal displazidir . Otozomal resesif kalıtılan bu hastalığın moleküler patogenezinden, siliyer bazal cisim ve proksimal aksonem bölgesinde lokalize olan *EVC* ve *EVC2* genlerindeki mutasyonlar sorumludur [118].

EVC ve *EVC2* genleri tarafından kodlanan EVC ve EVC2 proteinleri, silyum tabanında "EvC Zonu" olarak adlandırılan özelleşmiş bir protein kompleksi oluşturur. Bu transmembran kompleksi, Hedgehog (Hh) sinyal yolağının pozitif bir regülatörü olarak işlev görerek Smoothened (Smo) reseptöründen gelen sinyallerin Gli transkripsiyon faktörlerine aktarılmasını modüle eder. EVC proteinlerinin yokluğunda siliyer yapısal bütünlük korunsa dahi, Hh sinyal transdüksiyonu (iletimi) kritik düzeyde bozulmaktadır. Bu moleküler aksaklık; embriyonik gelişim sırasında uzuv tomurcuklarında (limb buds) ve kalp yastıkçıklarında (endocardial cushions) hücre proliferasyonunun hatalı yönetilmesine yol açarak, sendroma özgü iskelet ve kardiyak doku şekillenme kusurlarıyla sonuçlanır [134, 118].

2.3.3 Genotip-Fenotip İlişkisindeki Karmaşıklık

Yukarıda tanımlanan hastalıklar teorik olarak ayrı başlıklar altında sınıflandırılsa da, moleküler düzeydeki sınırları oldukça belirsizdir. Siliyopatiler, Mendelyen kalıtımın "tek gen-tek hastalık" kuralını zorlayan bir genetik karmaşıklık sergilemekte olup, bu durum tanı süreçlerini doğrudan etkilemektedir. Bu karmaşıklığın temelinde iki ana genetik unsur yer almaktadır:

İlki, allelik heterojenitedir. Aynı gen üzerindeki (örn. CEP290) farklı mutasyon tipleri, protein fonksiyonundaki bozulma düzeyine bağlı olarak tamamen farklı klinik tabloların ortaya çıkmasına yol açabilmektedir. Bu spektrum; izole retinal dejenerasyonla seyreden LCA'dan, çoklu organ tutulumu gösteren JBTS'ye veya letal karakterdeki MKS'ye kadar uzanabilmektedir. Dolayısıyla, sadece genetik veriler üzerinden kesin bir klinik sonuç öngörmek moleküler düzeyde oldukça güçtür [77, 85].

İkinci kritik unsur ise lokus heterojenitesi ve buna eşlik eden fenotipik örtüşmedir. Siliyopatilerde farklı genlerdeki mutasyonlar birbirine yakın klinik bulgular gösterebilmektedir. Örneğin, klinik belirtileriyle BBS tanısı alan bir hastada, aslında JBTS veya NPHP ile ilişkili olduğu bilinen gen mutasyonları tespit edilebilmektedir.

Geleneksel yöntemlerin (örn. Sanger dizileme, gen panelleri) bu genetik karmaşıklık karşısında yetersiz kalması, hastaların yıllarca süren bir "Tanı Arayışı" (Diagnostic Odyssey) süreciyle karşı karşıya kalmasına neden olmaktadır. Bu durum, siliyopatilerin aydınlatılmasında Tüm Ekzom Dizileme (WES) gibi kapsamlı yaklaşımlar ve biyoinformatik temelli analiz stratejilerinin kullanımını zorunluluk haline getirmiştir.

2.4 Biyoinformatik Veri Kaynakları

Biyoinformatik analizlerin başarısı, kullanılan verilerin niteliğine ve bu verilerin kapsamlı bir biyolojik çerçevede birleştirilmesine bağlıdır. Siliyopatiler gibi genetik heterojenitesi yüksek hastalıkların moleküler temellerini aydınlatmak; gen ekspresyonu, protein-protein etkileşimleri, evrimsel korunmuşluk ve klinik fenotipler gibi farklı türdeki verilerin entegrasyonunu zorunlu kılmaktadır.

Bu bölümde, aday genlerin belirlenmesi ve önceliklendirilmesi sürecinde modelin veri altyapısını oluşturan temel veritabanları, ontolojik yapılar ve bu kaynakların analitik süreçteki işlevleri ele alınacaktır.

2.4.1 Genomik ve Fenotipik Veri Kaynakları

İnsan genetiği çalışmalarında, genotip-fenotip ilişkisinin incelenmesi ve nadir hastalıkların tanısı için güncel veritabanları temel bir gerekliliktir. Bu kaynaklar, literatürdeki verileri standardize ederek hesaplamalı analizler için gerekli olan güvenilir veri altyapısını sunmaktadır.

2.4.1.1 Online Mendelian Inheritance in Man (OMIM) Veritabanı

Online Mendelian Inheritance in Man (OMIM), insan genleri ile genetik fenotipler arasındaki ilişkileri kapsayan en kapsamlı biyoinformatik veritabanı olarak kabul edilmektedir. Johns Hopkins Üniversitesi bünyesinde kürasyonu yapılan ve küresel ölçekte standart bir referans olan bu kaynak, Mendelyen kalıtım gösteren tüm genetik bozukluklar için temel bir bilgi bankasıdır. OMIM, her bir gen ve fenotip kaydı için özgün bir "MIM Numarası" atayarak siliyopatiler gibi karmaşık

sendromların genetik heterojenitesini sistematik bir yapıda sınıflandırır. Veritabanı bünyesindeki "Fenotip-Gen Haritaları" (Phenotype-Gene Map); bir hastalığın kromozomal lokalizasyonunu, ilişkili olduğu geni ve kalıtım modelini bütüncül bir şekilde sunmaktadır [93, 94].

2.4.1.2 Orphanet Veritabanı

Orphanet, nadir hastalıklar ve yetim ilaçlar (orphan drugs) konusunda uzmanlaşmış veritabanıdır. OMIM'den farklı olarak Orphanet, hastalıkların sadece genetik temelini değil; epidemiyolojik verilerini (prevalans, insidans), başlangıç yaşını ve klinik seyrini de içeren geniş bir perspektif sunmaktadır. Orphanet'in geliştirdiği "OrphaCode" sistemi, hastalıkların uluslararası terminolojide standardize edilmesini sağlar. Ayrıca, gen-hastalık ilişkilerinin validasyon durumunu puanlayarak, araştırmacılara güvenilir gen listeleri oluşturma konusunda rehberlik etmektedir. Bu özellikleri sayesinde Orphanet, nadir hastalıkların moleküler tanısında ve translasyonel araştırmalarda kaynak olarak kullanılmaktadır [96].

2.4.1.3 DisGeNET Veritabanı

DisGeNET, insan genleri ile hastalıklar arasındaki ilişkiyi, farklı veri kaynaklarını entegre ederek sunan kapsamlı bir bilgi bankasıdır. Bu platform, uzman onaylı veritabanlarından gelen kayıtları, bilimsel literatürden metin madenciliği (text mining) ile elde edilen verileri ve deneysel modelleri standardize ederek tek bir yapıda birleştirmektedir. Veritabanı bünyesindeki her bir gen-hastalık ilişkisi, kanıtların güvenilirliğine göre hesaplanan bir "İlişki Skoru" (Association Score) ile derecelendirilmiştir [97, 32].

2.4.2 İnsan Fenotip Ontolojisi (HPO) Veritabanı

Klinik gözlemlere dayalı fenotip tanımlarının (retinal dejenerasyon, polidaktili vb.) hesaplamalı analizlerde kullanılabilmesi için standart bir dile dönüştürülmesi gerekmektedir. İnsan Fenotip Ontolojisi (HPO), insan hastalıklarındaki klinik anormallikleri hiyerarşik bir yapıda tanımlayan ve bu verileri bilgisayarlı analizler için standardize eden temel bir referans kaynağıdır [99].

Matematiksel olarak "Yönlendirilmiş Asiklik Grafik" (Directed Acyclic Graph - DAG) yapısını kullanan HPO, terimleri genelden özele doğru bir hiyerarşiyle birbirine bağlar. Örneğin; üst basamakta yer alan "Göz Anomalisi" (HP:0000478) terimi, hiyerarşinin alt dallarında yer alan "Retina Anomalisi" (HP:0000479) ve daha spesifik bir tanım olan "Retinitis Pigmentosa" (HP:0000510) terimlerini

kapsar [99]. HPO'nun sunduğu bu yapı, biyoinformatik analizlerde "Semantik Benzerlik" (Semantic Similarity) hesaplamalarına olanak tanır. Bu süreçte her klinik terime, popülasyon frekansına dayalı bir "Bilgi İçeriği" (Information Content - IC) skoru atanır. Nadir görülen spesifik bulgular (yüksek IC), genel bulgulara kıyasla tahmin modellerinde daha yüksek matematiksel ağırlığa sahiptir. Bu yaklaşım; siliyopatiler gibi farklı sistemleri etkileyen karmaşık hastalıklarda, aday genlerin fenotipik profiller ile ne derece örtüştüğünü hesaplayarak gen önceliklendirme (Gene Prioritization) çalışmaları için sayısal bir temel teşkil etmektedir [30].

2.4.3 Genome Aggregation Database (gnomAD) Veritabanı

İnsan genetiği çalışmalarında bir varyantın hastalık yapıcı etkisini değerlendirmek için, o varyantın popülasyondaki yaygınlığı ve ilgili genin mutasyonlara karşı tolerans düzeyi kritik öneme sahiptir. Genome Aggregation Database (gnomAD), 140.000'den fazla bireyden elde edilen ekzom ve genom dizileme verilerini bir araya getiren, dünyanın en geniş kapsamlı popülasyon varyasyon veritabanıdır [101]. gnomAD, sadece varyant frekanslarını sunmakla kalmayıp, aynı zamanda genlerin evrimsel süreçteki hassasiyetini ölçen "Genomik Kısıtlılık" (Genomic Constraint) metriklerini de sağlamaktadır.

Genomik kısıtlılık hipotezine göre, sağlıklı bireylerden oluşan bir popülasyonda bir genin beklenen oranda "Fonksiyon Kaybı" (Loss-of-Function - LoF) mutasyonu taşımaması, o genin güçlü bir negatif seleksiyon baskısı altında olduğunu gösterir. Diğer bir deyişle, bu genlerdeki mutasyonlar hayatta kalmayı veya üremeyi olumsuz etkilediği için evrimsel süreçte popülasyondan temizlenmektedir. Bu baskıyı sayısallaştırarak genlerin hastalık yapma potansiyelini öngörmek için iki temel metrik kullanılmaktadır [29]:

İlk metrik olan pLI (Probability of Loss-of-function Intolerance), bir genin heterozigot LoF mutasyonlarını tolere edememe olasılığını 0 ile 1 arasında puanlar. pLI skorunun $0,9$ ve üzerinde olması, ilgili genin "haployetmezlik" (haploinsufficiency) gösterdiğini, yani tek bir kopya kaybının dahi ağır klinik tablolara yol açabileceğini ifade eder. Bu genler genellikle embriyonik gelişim için elzem olan temel hücresel süreçleri yönetir.

İkinci metrik ise LOEUF (Loss-of-function Observed/Expected Upper Bound Fraction), gende gözlemlenen LoF varyant sayısının, beklenen sayıya oranını temsil eder. LOEUF skoru ne kadar düşükse, genin mutasyonlara karşı toleransı o kadar az, yani kısıtlılığı o kadar yüksektir. Özellikle resesif geçişli siliyopatilerde veya

pLI skoru ile net sınıflandırılamayan genlerde LOEUF, genin fonksiyonel önemini belirleyen en güçlü göstergelerden biridir.

2.4.4 STRING Veritabanı

Sistem biyolojisi yaklaşımı, proteinlerin hücre içinde izole birimler olarak değil, dinamik ve karmaşık ağlar (interaktom) oluşturarak bulunduğunu savunmaktadır. Bir genin fonksiyonunu veya hastalıkla olan ilişkisini tam olarak kavrayabilmek için, o genin ürününün hücresel ağ içerisindeki konumunu ve etkileşim ortaklarını analiz etmek gereklidir. Bu analitik yaklaşımın temelinde yer alan "fonksiyonel yakınsama" (Guilt by Association) ilkesine göre; fonksiyonu henüz tanımlanmamış bir protein, belirli bir biyolojik süreçte görev alan proteinlerle fiziksel veya fonksiyonel etkileşim halindeyse, muhtemelen kendisi de aynı süreçte rol almaktadır. Bu tür topolojik analizler ve ağ modellemeleri için en yaygın başvurulan kaynaklardan biri STRING veritabanıdır. STRING, sadece doğrudan fiziksel etkileşimleri değil, aynı zamanda gen komşuluğu, birlikte ekspresyon (co-expression) ve metin madenciliği gibi farklı kanıt türlerini entegre ederek her etkileşim çifti için bir "Güven Skoru" (Confidence Score) üretmektedir [103].

2.5 Makine Öğrenmesi ve Hesaplamalı Yöntemler

Modern biyoinformatikte omik teknolojileriyle üretilen büyük veri setleri, geleneksel analiz yöntemlerini yetersiz bırakmaktadır. Genotip-fenotip arasındaki ilişkinin karmaşık yapısı, makine öğrenmesi (Machine Learning) algoritmalarını zorunluluk haline getirmiştir. Özellikle siliyopatiler gibi genetik çeşitliliği yüksek hastalıkların incelenmesinde, veriler içerisindeki anlamlı bağlantıları yakalayabilen denetimli öğrenme modelleri önemli bir yer tutmaktadır. Bu süreçte, topluluk öğrenme (ensemble learning) algoritmaları hem de açıklanabilir yapay zeka (Explainable AI (XAI)) yöntemleri öne çıkmaktadır [109, 116]. Takip eden bölümlerde, bu yöntemlerin temel çalışma prensipleri ele alınacaktır.

2.5.1 Topluluk Öğrenmesi (Ensemble Learning)

Makine öğrenmesi kuramında, tek bir tahmin modelinin (örneğin tek bir karar ağacının) başarısı genellikle "Yanlılık-Varyans Dengesi" (Bias-Variance Tradeoff) ile sınırlıdır. Basit yapıdaki modeller verideki karmaşık ilişkileri yakalamakta yetersiz kalırken (high bias), fazla karmaşık modeller eğitim verisine aşırı uyum sağlayarak (overfitting) yeni veriler üzerinde düşük performans sergiler (high variance). Topluluk Öğrenmesi, birden fazla temel öğrenicinin (base learner)

birleştirilerek, hem yanlılığın hem de varyansın minimize edildiği daha dayanıklı bir sistem oluşturulması prensibine dayanır [189, 110]. Bu yaklaşım, bireysel modellerin hatalarını birbirleriyle telafi etmelerini sağlayarak, tek bir modelin tek başına ulaşamayacağı daha yüksek bir tahmin başarısına ve kararlılığa ulaşılmasını hedeflemektedir. Literatürde Bagging ve Boosting olmak üzee iki temel topluluk stratejisi öne çıkmaktadır:

2.5.1.1 Bagging (Bootstrap Aggregating)

Bagging (Bootstrap Aggregating) yöntemi, eğitim veri setinden rastgele ve yerine koymalı örnekleme (bootstrap sampling) yoluyla çok sayıda alt veri kümesi oluşturulması esasına dayanır. Bu yaklaşımda, her bir alt veri kümesi üzerinde birbirinden bağımsız öğreniciler eğitilerek bir model topluluğu oluşturulur. Modellerin ürettiği bireysel sonuçlar, sınıflandırma problemlerinde "çoğunluk oylaması" (majority voting), regresyon problemlerinde ise "ortalama alma" yöntemiyle birleştirilerek nihai tahmin değerine ulaşılır. Bagging stratejisinin temel motivasyonu, modelin varyansını düşürerek aşırı uyum (overfitting) riskini minimize etmek ve daha istikrarlı tahminler üretmektir. Biyoinformatik analizlerde sıkça başvurulan Random Forest algoritması, bu metodolojinin en karakteristik ve başarılı örneği olarak kabul edilmektedir [193, 191].

2.5.1.2 Boosting

Boosting yaklaşımı, modellerin birbirinden bağımsız eğitilmesi yerine ardışık (sequential) bir hiyerarşide yapılandırıldığı daha gelişmiş bir stratejidir. Bu yöntemde prensip, her yeni modelin (M_t), kendisinden önce gelen modelin (M_{t-1}) yaptığı hataları ve tahmin kalıntılarını (residuals) düzeltmeye odaklanmasıdır. Matematiksel olarak süreç, öğrenicilerin belirli ağırlık katsayıları ile bir araya getirilerek güçlü bir tahmin yapısının oluşturulması esasına dayanır:

$$F_T(x) = \sum_{t=1}^{T} \alpha_t h_t(x) \tag{2.1}$$

Burada $h_t(x)$ her bir öğreniciyi, α_t ise ilgili modelin nihai tahmine olan katkısını ifade etmektedir. Boosting algoritmaları özellikle sınıflandırılması güç olan nadir hastalık genleri gibi zor verilerin öğrenilmesinde, Bagging yöntemlerine kıyasla genellikle daha üstün bir performans sergilemektedir [196]. Bu ardışık öğrenme süreci, modelin hem yanlılığını azaltmakta hem de tahmin gücünü maksimize etmektedir.

2.5.2 XGBoost (eXtreme Gradient Boosting) Algoritması

XGBoost, Tianqi Chen ve Carlos Guestrin tarafından geliştirilen ve klasik Gradient Boosting modelinin hız, ölçeklenebilirlik ve regülasyon kapasitesi açısından optimize edilmiş halidir. Özellikle dengesiz ve gürültülü biyolojik veri setlerinin analizinde XGBoost'un tercih edilmesinin temelinde, algoritmanın sunduğu özgün matematiksel optimizasyon teknikleri yer almaktadır [22]. Geleneksel Gradient Boosting yaklaşımları kayıp fonksiyonunun yalnızca birinci türevini (gradyan) dikkate alırken, XGBoost optimizasyon sürecine ikinci derece Taylor serisi açılımını dahil etmektedir. Böylece, modelin hatanın yönüne (gradyan) ek olarak, hatanın değişim hızını (Hessian/eğrilik) da hesaplamasını sağlayarak daha hızlı ve isabetli bir yakınsama (convergence) sağlanır [106]. Algoritmanın minimize etmeyi hedeflediği toplam objektif fonksiyon (Obj) şu şekilde ifade edilir:

$$Obj(\Theta) = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k) \tag{2.2}$$

Burada l ifadesi türevlenebilir dışbükey kayıp fonksiyonunu, Ω ise model karmaşıklığını baskılayarak, aşırı uyumu (overfitting) engelleyen regülasyon terimini temsil eder. Bu regülasyon desteği, XGBoost'un biyoinformatik veri setlerinde daha dayanıklı ve genelleştirme yeteneği yüksek modeller oluşturmasına imkan tanımaktadır [191, 106].

2.5.2.1 Aşırı Uyumun (Overfitting) Denetimi

Biyolojik veri setlerinde örneklem sayısının genellikle kısıtlı olması, modellerin eğitim verisindeki gürültüleri öğrenerek aşırı uyum (overfitting) gösterme riskini arttırmaktadır. XGBoost, bu riski yönetmek için iki temel regülasyon bileşeni ekleyerek model karmaşıklığını denetim altında tutar [23]:

$$\Omega(f) = \gamma T + \frac{1}{2}\lambda ||w||^2 \tag{2.3}$$

- γT **(Gamma):** Bir ağacın yeni bir yaprak oluşturması için gereken minimum kayıp azalmasını belirleyen kısıtlamadır. Modelin gereksiz yere karmaşıklaşmasını ve derinleşmesini engeller.
- $\lambda ||w||^2$ **(L2 Regülasyonu):** Yaprakların sahip olduğu tahmin skorlarının (w) büyüklüğünü sınırlandırır. Bu sayede, tek bir yaprağın modelin genel kararı üzerinde baskın bir etki yaratmasının önüne geçilerek daha dengeli ve genellenebilir bir tahmin yapısı oluşturulur.

2.5.2.2 Seyrek Veri (Sparsity) ve Eksik Değer Yönetimi

Biyoinformatik veritabanlarında farklı kaynaklardan gelen verilerin birleştirilmesiyle oluşan datalar, yüksek oranda eksik değer (missing data) içerebilmektedir. Geleneksel algoritmalar genellikle bu verileri ortalama veya medyan değerlerle doldurma (imputation) yoluna giderken; XGBoost verinin orijinal yapısını koruyan "Seyrek Veri Farkındalıklı Bölünme" (Sparsity-aware Split Finding) algoritmasını kullanmaktadır. Eğitim sürecinde, eksik verilerin olduğu her karar noktasında kazancı en üst düzeye çıkaran bir "varsayılan yön" (default direction) belirlenir. Bu özellik, eksik bilginin bir hata olarak görülmesi yerine, modelin öğrenme sürecinde anlamlı bir örüntü olarak değerlendirilmesine imkan sağlamaktadır. Böylece biyolojik verinin doğal varyasyonu korunarak daha gerçekçi sonuçlar elde edilebilmektedir [112, 105].

2.5.3 Model Performans Değerlendirme Metrikleri

Bir sınıflandırma modelinin başarısını ölçmek için kullanılan tüm metrikler, modelin tahminleri ile gerçek değerler arasındaki ilişkiyi özetleyen "Hata Matrisi" (Confusion Matrix) üzerinden türetilir. Bu matris; modelin performansını dört temel bileşen üzerinden analiz etmeye olanak tanımaktadır:

- **Doğru Pozitif (TP):** Modelin pozitif olarak tahmin ettiği ve gerçekten pozitif olan örnek sayısıdır.
- **Yanlış Pozitif (FP):** Modelin pozitif dediği ancak gerçekte negatif olan (hata) örnek sayısıdır.
- **Doğru Negatif (TN):** Modelin negatif olarak tahmin ettiği ve gerçekten negatif olan örnek sayısıdır.
- **Yanlış Negatif (FN):** Modelin negatif dediği ancak gerçekte pozitif olan (gözden kaçan) örnek sayısıdır.

Bu temel değerler kullanılarak literatürde yaygın olarak başvurulan şu metrikler hesaplanmaktadır:

1. **Hassasiyet (Precision):** $\frac{TP}{TP+FP}$ formülü ile hesaplanır. Modelin pozitif olarak tahmin ettiği örneklerin ne kadarının gerçekten pozitif olduğunu gösterir. Sistemin hata payını (yalancı pozitiflik) ölçmek için kullanılır.
2. **Duyarlılık (Recall):** $\frac{TP}{TP+FN}$ formülü ile hesaplanır. Gerçek pozitif örneklerin ne kadarının başarıyla yakalanabildiğini gösterir.

Biyoinformatik ve veri madenciliği çalışmalarında, hedeflenen sınıfın tüm veri setine göre çok nadir olduğu durumlarda (imbalanced data), sadece bu metriklerin sayısal değerlerine değil, bu değerlerin farklı eşiklerdeki değişimini gösteren eğrilere bakılması standart bir yaklaşımdır. Bu kapsamda ROC (Duyarlılık vs. Yanlış Pozitif Oranı) ve AUPRC (Hassasiyet / Duyarlılık) eğrileri, modelin ayırt edici gücünü tek bir performans skoru yerine bütüncül bir grafiksel ölçekte değerlendirmeyi mümkün kılmaktadır [114, 148, 115].

2.5.4 XAI ve SHAP Yöntemi

Makine öğrenmesi modellerinin, özellikle topluluk öğrenmesi (Ensemble Learning) gibi karmaşık yapıların klinik karar destek sistemlerinde uygulanabilirliği, "Kara Kutu" (Black Box) problemi nedeniyle sıklıkla kısıtlanmaktadır. Bu yaklaşımlar yüksek tahmin başarısı sergilese de, modelin hangi biyolojik değişkenlere dayanarak sonuca ulaştığını bilinememektedir. Modern biyoinformatik araştırmalarda, tahminlerin biyolojik mekanizmalarla ilişkilendirilmesi gerekliliği, Açıklanabilir Yapay Zeka (Explainable AI (XAI)) yaklaşımlarının önemini artırmıştır. Modelden bağımsız (model-agnostic) yapısı ve matematiksel tutarlılığıyla öne çıkan SHAP (SHapley Additive exPlanations) yöntemi, literatürde güvenilir bir araç olarak kabul edilmektedir [116, 109, 139].

SHAP yöntemi, kökenini Lloyd Shapley'in kooperatif oyun teorisinden almaktadır. Bu teorik yapıda, tahmin modeli bir "oyun", modeli oluşturan her bir öznitelik (feature) ise bu oyuna katılan birer "oyuncu" olarak tanımlanır. Yöntemin temel amacı, modelin ürettiği toplam çıktı değerini (Prediction), girdilerin marjinal katkı paylarına göre adil bir şekilde dağıtmaktır. Klasik öznitelik önemi yöntemlerinden (Gini Impurity, Permutation Importance vb.) farklı olarak SHAP, değişkenler arasındaki karmaşık etkileşimleri ve korelasyonları da hesaplamaya dahil etmektedir. Bir i özelliğinin Shapley değeri (ϕ_i), tüm olası özellik kombinasyonları üzerinden ağırlıklı marjinal katkıların ortalaması olarak hesaplanır [25, 26]:

$$\phi_i = \sum_{S \subseteq F \setminus \{i\}} \frac{|S|!(|F| - |S| - 1)!}{|F|!} [f_{S \cup \{i\}}(x_{S \cup \{i\}}) - f_S(x_S)] \tag{2.4}$$

SHAP analizi, biyolojik verilerin çözümlenmesinde iki farklı boyutta yorumlanabilirlik sunmaktadır:

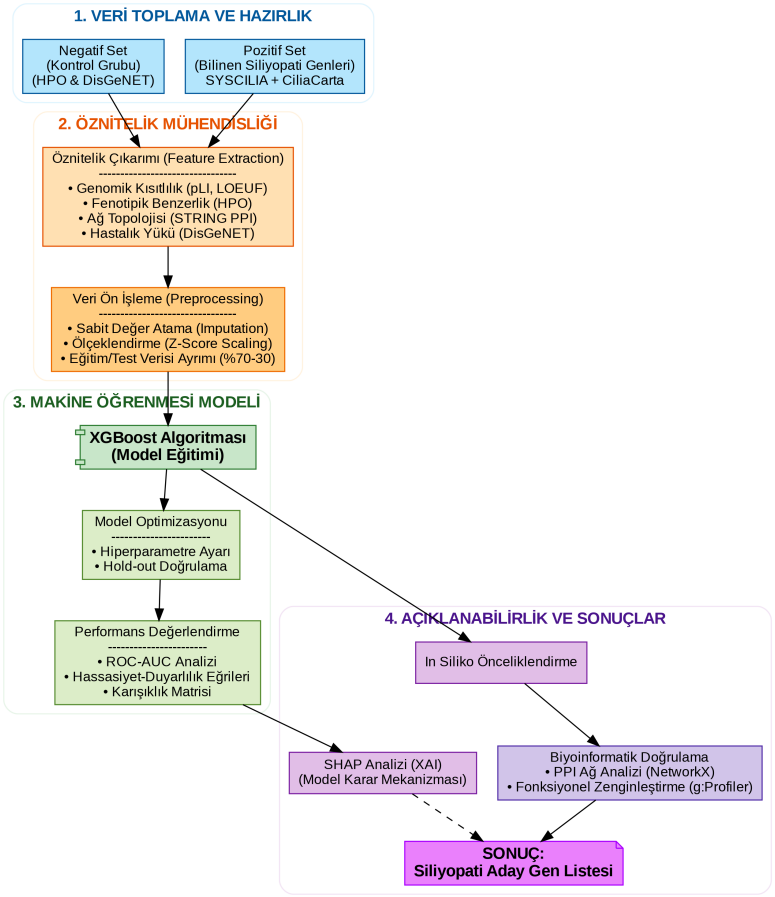
1. **Global Yorumlanabilirlik:** Tüm veri seti genelinde tahmin modelini en

çok etkileyen faktörlerin hiyerarşisini ortaya koyar. Özniteliklerin yüksek veya düşük değerlerinin, tahmin sonucunu pozitif veya negatif yönde nasıl etkilediği "Beeswarm" grafikleri ile görselleştirilir.

2. **Lokal Yorumlanabilirlik:** Tek bir örnek özelinde kişiselleştirilmiş bir açıklama sunar. "Force Plot" analizi ile baz değer (base value) üzerine eklenen risk artırıcı veya azaltıcı vektörler hesaplanır. Bu yaklaşım, in-silico tahminlerin biyolojik mantıkla (biological plausibility) uyumunun denetlenmesinde kullanılır.

3
MATERYAL VE YÖNTEM

Çalışmada, siliyopati aday genlerini önceliklendirmek amacıyla makine öğrenmesi modeli geliştirilmiştir. Yüksek boyutlu omik veri setlerinin işlenmesi ve model eğitimi süreçlerindeki yüksek performanslı hesaplama gereksinimi nedeniyle, Google Colaboratory platformu tercih edilmiştir. Analizler, Python (v3.10) ortamında *Scikit-learn*, *XGBoost*, *SHAP* ve *NetworkX* kütüphaneleriyle yürütülmüştür. Genel çalışma tasarımı Şekil 3.1'da özetlenmiştir.



Şekil 3.1 Makine öğrenmesi tabanlı biyoinformatik iş akış şeması.
(Görsel: Yazar, Graphviz)

3.1 Çalışma Tasarımı ve Veri Kaynakları

Denetimli makine öğrenmesi algoritmalarının başarısı; eğitim verisinin niteliğine, çeşitliliğine ve biyolojik sistemleri temsil etme kapasitesine doğrudan bağlıdır. Bu çalışmada, siliyopati aday genlerinin belirlenmesi süreci; genomik kısıtlılık, fenotipik benzerlik, gen-hastalık ilişki skorları ve protein ağlarındaki topolojik konumlara dayanan çok katmanlı bir sınıflandırma yaklaşımı olarak modellenmiştir. Bu doğrultuda, literatürde kabul görmüş ve uzman denetiminden geçmiş yüksek güvenilirlikli veritabanları entegre edilerek kapsamlı bir veri seti oluşturulmuştur.

3.1.0.1 Pozitif Eğitim Setinin Oluşturulması

Modelin siliyer biyoloji ile ilişkili genlerin karakteristik özelliklerini kavrayabilmesi için, bu sistemle ilişkisi deneysel olarak kesinleşmiş genlerden oluşan bir referans setine ihtiyaç duyulmuştur. Bu amaçla, "Altın Standart" (Gold Standard) niteliği taşıyan iki temel kaynak entegre edilmiştir:

1. **SYSCILIA Gold Standard (SCGSv1/2):** Avrupa Birliği SYSCILIA konsorsiyumu tarafından hazırlanan bu liste, siliyer lokalizasyonu veya fonksiyonu deneysel yöntemlerle kanıtlanmış genleri içermektedir. Veritabanındaki yüksek güvenilirlikli kayıtlar, modelin öğrenme sürecinde temel dayanak olarak pozitif sete dahil edilmiştir [91, 197].
2. **CiliaCarta:** Genlerin siliyer olma olasılığını Bayesyen istatistiksel modellerle hesaplayan güncel ve kapsamlı bir veri listesidir. Modelin pozitif eğitim setini zenginleştirmek amacıyla, CiliaCarta veritabanında yüksek olasılık skoruna sahip genler analize eklenmiştir [33].

Veri entegrasyonu aşamasında kaynaklardan elde edilen gen listeleri birleştirilmiş ve yinelenen kayıtlar temizlenmiştir. Tüm gen sembolleri HGNC (*HUGO Gene Nomenclature Committee*) standartlarına göre güncellenmiş ve sonuç olarak 1183 adet siliyer gen, modelin "Pozitif Eğitim Seti"ni oluşturmuştur.

3.1.1 Negatif (Kontrol) Setinin Oluşturulması

Makine öğrenmesi modellinin pozitif sınıfın ayırt edici özelliklerini öğrenebilmesi için, bu özellikleri taşımayan bir negatif veri setine ihtiyaç duyulmuştur. Literatürde siliyer fonksiyonla ilişkisi kesin olarak dışlanmış (non-ciliary) standardize bir gen seti bulunmadığı için, çalışmada siliyer olmayan genleri temsil etmek üzere literatürle uyumlu bir negatif örnekleme stratejisi benimsenmiştir.

Metodolojik süreç üç temel aşamada yapılandırılmıştır:

1. **Analiz Evreninin Sınırlandırılması (Universe Proxy):** Çalışma evreni; HPO, DisGeNET ve pozitif eğitim setinde yer alan genlerin birleşimiyle (union) oluşturulmuştur. Bu yaklaşım ile analizlerin yalnızca literatürde klinik veya biyolojik karşılığı bulunan genlere (Clinically Relevant Genome) odaklanması sağlanmış, böylece fonksiyonel verisi bulunmayan genlerden kaynaklanabilecek gürültünün azaltılması hedeflenmiştir.

2. **Negatif Set Seçimi ve Sınıf Dengesi:** Pozitif referans setinde yer alan genler, analiz evreninden çıkartılmıştır. Sınıf dengesizliğini (class imbalance) önlemek amacıyla, eğitim setindeki pozitif gen sayısına eşit olacak şekilde "Rastgele Alt-Örnekleme" (Random Undersampling) stratejisi uygulanmıştır. Bu yöntemle, 1183 pozitif gene karşılık aynı sayıda (1:1) negatif gen seçilerek dengeli bir "Negatif (Kontrol) Seti" elde edilmiştir.

3. **Aday Gen Havuzunun Belirlenmesi:** Negatif Set tamamlandıktan sonra, analiz evreninde kalan gelen, modelin siliyopati skoru atayacağı "Aday Gen Havuzu" olarak tanımlanmıştır. Bu havuz, eğitilen modelin aday genleri önceliklendirmesi için kullanılmıştır.

3.2 Öznitelik Mühendisliği (Feature Engineering)

Bu çalışmada, siliyopati aday genlerini önceliklendirmek amacıyla kullanılan sayısal öznitelikler; genlerin evrimsel kısıtlılık düzeyleri, fenotipik spektrumları ve protein etkileşim ağlarındaki topolojik rolleri gibi kritik biyolojik özelliklerden türetilmiştir. Veri setinin oluşturulması sürecinde, literatürde kabul görmüş ve uzman denetiminden geçmiş yüksek güvenilirlikli veri kaynakları entegre edilerek aşağıdaki öznitelikler (features) üretilmiştir.

3.2.1 Genomik Kısıtlılık ve Mutasyon Toleransı

Genlerin doğal seçilim baskısı altındaki durumunu ve varyasyon toleransını analiz etmek amacıyla *Genome Aggregation Database* (gnomAD v2.1.1) referans alınmıştır. Bu veri tabanından elde edilen ve genlerin fonksiyon kaybı (LoF) mutasyonlarına karşı duyarlılığını nicelleştiren iki temel metrik modele dahil edilmiştir:

- **pLI (Probability of Loss-of-function Intolerance):** Genin LoF mutasyonlarına karşı toleranssızlık olasılığıdır. Yüksek skor (≥ 0.9),

genin mutasyonlara karşı aşırı duyarlı olduğunu ve muhtemel bir hastalık geni olabileceğini gösterir [150].

- **LOEUF (LoF Observed / Expected Upper Bound Fraction):** Gende gözlenen LoF varyant sayısının, beklenen sayıya oranının istatistiksel üst sınırıdır. Değerin 0'a yaklaşması, genin mutasyonlara karşı kısıtlı olduğunu ve biyolojik olarak kritik bir öneme sahip olduğunu ifade eder [29].

3.2.2 Klinik ve Fenotipik Öznitelikler

Siliyopatilerin çoklu sistem tutulumu gösteren (pleiotropik) doğasını modele yansıtmak amacıyla klinik bağlantı ve fenotipik çeşitlilik skorları üretilmiştir:

- **Klinik İlişki Derinliği:** İnsan gen-hastalık ilişkilerini entegre eden DisGeNET (v7.0) platformu üzerinden, her bir genin literatürde ilişkili olduğu hastalık sayısı (*nunique*) hesaplanmıştır. Bu skor ile genin klinik assosiasyon gücü sayısal bir değere dönüştürülmüştür.
- **Fenotipik Çeşitlilik:** İnsan Fenotip Ontolojisi (HPO) `genes_to_phenotype.txt` veri seti kullanılarak, her genin ilişkili olduğu fenotipik terim sayısı hesaplanmıştır.

Bu iki metrik, genlerin klinik yansımasını ve biyolojik sistemler üzerindeki pleiotropik etkisini temsil eden sayısal öznitelikler olarak modelin eğitim setine aktarılmıştır.

3.2.3 İnteraktom Analizi ve Ağ Merkeziyeti

Aday proteinlerin fonksiyonel protein etkileşim (interaktom) ağlarındaki konumlarını belirlemek amacıyla STRING (v11.5) veri tabanı kullanılmıştır. Analizde, güvenilirlik düzeyi 400 (orta güvenilirlik) ve üzerinde olan etkileşimler baz alınmıştır. Her gen için, tanımlanan eşik değerin üzerindeki tüm siliyer protein etkileşimlerinin kümülatif toplamı hesaplanmıştır. Elde edilen ham skorlar, model girişinde diğer özniteliklerle ölçek uyumu sağlamak amacıyla *MinMaxScaler* yöntemiyle 0-1 aralığına normalize edilmiştir. Bu süreçte, STRING veritabanındaki 400 skorluk temel güven eşiği, normalizasyon sonrası 0.4 değerine tekabül edecek şekilde ölçeklendirilmiş ve proteinin ağ üzerindeki etkileşim yoğunluğunu temsil eden bir öznitelik olarak modele eklenmiştir.

3.3 Veri Ön İşleme (Data Preprocessing)

Biyolojik veri tabanlarından elde edilen ham verilerin analiz süreçlerine hazırlanması amacıyla hazırlık ve standardizasyon süreci izlenmiştir. Bu aşamada; *scikit-learn* ve *pandas* kütüphaneleri kullanılarak aşağıdaki metodolojik adımlar uygulanmıştır.

3.3.1 Eksik Veri Yönetimi (Missing Data Handling)

Biyoinformatik veri setlerindeki seyreklik (sparsity) problemini yönetmek için "Sabit Değer Atama" (Constant Imputation) stratejisi benimsenmiştir. Fenotipik kayıtların (HPO) veya protein etkileşim verilerinin (STRING) bulunmaması, biyolojik açıdan "bilgi yokluğu" (absence of evidence) olarak değerlendirilmiştir. Bu doğrultuda, eksik verilere popülasyon ortalaması atayarak yapay bir bilgi üretmek yerine, varlık/yokluk bilgisini korumak adına "0" değeri atanmıştır. Ancak gnomAD veri tabanından elde edilen LOEUF gibi genetik kısıtlılık metriklerinde, eksik değerler genin kısıtlanmadığı varsayımıyla "1" olarak tamamlanmıştır. Bu yaklaşım ile modelin bilgi bulunmayan durumlarda tarafsız (neutral) kalması hedeflenmiştir [113].

3.3.2 Veri Setinin Ayrılması (Train-Test Split)

Modelin genelleme yeteneğini ölçmek ve aşırı uyum (overfitting) riskini denetlemek amacıyla, dengelenmiş veri seti (D) iki alt kümeye ayrılmıştır. Bu işlem matematiksel olarak aşağıda gösterilmiştir:

$$D = D_{egitim} \cup D_{test}, \quad D_{egitim} \cap D_{test} = \emptyset \tag{3.1}$$

Bu çalışmada, veri setinin %70'i modelin eğitimi, %30'u ise modelin daha önce hiç görmediği veriler üzerindeki performansını test etmek amacıyla ayrılmıştır.

Bölme işlemi sırasında, pozitif (siliyopati) ve negatif (kontrol) sınıflarının her iki kümede de orijinal oranını (1:1) korumak için Tabakalı Örnekleme (Stratified Sampling) yöntemi uygulanmıştır [198]. Bu yöntem ile eğitim ve test süreçlerinde sınıf dağılımının homojen kalması sağlanarak istatistiksel yanlılığın önlenmesi amaçlanmıştır. Tekrarlanabilirlik için bölme işlemlerinde sabit bir rastgelelik değeri (*Random State: 42*) kullanılmıştır.

3.4 Makine Öğrenmesi Modelinin Eğitilmesi

Literatür verileri, siliyopati çalışmalarındaki gibi heterojen öznitelik setine ve seyrek (sparse) veri yapılarına sahip senaryolarda, gradyan artırma (*gradient boosting*) mekanizmasının klasik yöntemlere oranla daha optimize ve ölçeklenebilir sonuçlar verdiğini göstermektedir [113, 112]. Bu nedenle yüksek tahmin başarısı ve sınıf dengesizliğine karşı dirençli yapısı nedeniyle XGBoost (*eXtreme Gradient Boosting*) algoritması tercih edilmiştir.

Modelin eğitim sürecinde biyolojik verilerin karmaşıklığını yönetebilmek amacıyla `XGBClassifier` kullanılmıştır. Algoritmanın aşırı uyum (*overfitting*) riskini minimize etmek ve genelleme yeteneğini artırmak amacıyla hiperparametre optimizasyonu gerçekleştirilmiştir. Bu süreçte, biyoinformatik literatüründe benzer veri setleri için önerilen standart değerler ve deneysel ön denemeler baz alınarak parametre seçimi yapılmıştır [23, 22, 113, 106].

Analizler sonucunda saptanan ve modelin en istikrarlı performansı sergilediği nihai konfigürasyon, Tablo 3.1'de sunulmuştur. Ağaç sayısının (*n_estimators*) 100 ve maksimum derinliğin (*max_depth*) 5 olarak belirlenerek, modelin eğitim verisindeki gürültüleri ezberlemek yerine temel biyolojik örüntüleri öğrenmesi amaçlanmıştır.

Tablo 3.1 XGBoost hiperparametre konfigürasyonu.

Hiperparametre	**Değer**	**Açıklama**
`n_estimators`	100	Ardışık oluşturulan karar ağacı sayısı
`learning_rate`	0.1	Her adımda hata düzeltme hızı (Shrinkage)
`max_depth`	5	Bir ağacın ulaşabileceği maksimum derinlik
`objective`	binary:logistic	İkili sınıflandırma için lojistik hedef fonksiyonu
`random_state`	42	Analitik tekrarlanabilirlik için sabit değer

3.5 Model Performansının Ölçülmesi

Eğitilen makine öğrenmesi modelinin genelleme kapasitesi (generalization ability), eğitim sürecinde kullanılmamış bağımsız test seti (*hold-out test set*) üzerinde analiz edilmiştir. Siliyopati gibi nadir hastalıklarla ilişkili genlerin genom içindeki prevalansının düşük olması, veri setinde belirgin bir sınıf dengesizliğine (class imbalance) neden olmaktadır. Bu tür durumlarda geleneksel "Doğruluk" (Accuracy) metriği, azınlık sınıfındaki (pozitif genler) hataları maskeleyerek yanıltıcı sonuçlar verebildiği için çalışmada hassas metrikler tercih edilmiştir.

3.5.1 Sınıflandırma Metrikleri

Modelin başarısı, Karmaşıklık Matrisi (Confusion Matrix) bileşenleri olan Doğru Pozitif (TP), Yanlış Pozitif (FP), Doğru Negatif (TN) ve Yanlış Negatif (FN) değerleri üzerinden türetilen şu metriklerle ölçülmüştür:

1. **Hassasiyet (Precision):** Modelin pozitif olarak tahmin ettiği genlerin gerçekte siliyopati ile ilişkili olma oranını temsil eder.

$$\text{Hassasiyet} = \frac{TP}{TP + FP} \tag{3.2}$$

2. **Duyarlılık (Recall / Sensitivity):** Mevcut siliyopati genlerinin ne kadarının model tarafından yakalanabildiğini gösterir.

$$\text{Duyarlılık} = \frac{TP}{TP + FN} \tag{3.3}$$

3. **F1-Skoru:** Hassasiyet ve duyarlılık metriklerinin harmonik ortalamasıdır. Sınıf dengesizliğinin bulunduğu durumlarda modelin genel başarısını yansıtan dengeli bir başarı göstergesidir.

$$\text{F1-Skoru} = 2 \times \frac{\text{Hassasiyet} \times \text{Duyarlılık}}{\text{Hassasiyet} + \text{Duyarlılık}} \tag{3.4}$$

3.5.2 Eşik Değerinden Bağımsız Eğri Analizleri (ROC ve AUPRC)

Modelin ayırt edici gücünü farklı karar eşiklerinde bütüncül olarak değerlendirmek için iki temel eğri analizi uygulanmıştır:

3.5.2.1 ROC-AUC Analizi

ROC eğrisi, Yanlış Pozitif Oranı (FPR) ile Doğru Pozitif Oranı (TPR) arasındaki ilişkiyi görselleştirir. Eğri altında kalan alan (AUC), modelin sınıfları birbirinden ayırma kabiliyetini 0 ile 1 arasında puanlar. AUC değerinin 0.50 olması, modelin sınıfları birbirinden ayıramadığını ve rastgele tahmin (*random guess*) yaptığını ifade eder. Bu değer arttıkça ve 1.0 düzeyine yaklaştıkça modelin tahmin performansı ve ayırt edicilik gücü de doğrusal olarak artmaktadır [148].

3.5.2.2 AUPRC (Average Precision) Analizi

AUPRC, modelin azınlık sınıfını tespit etme performansını doğrudan Kesinlik ve Duyarlılık değerleri üzerinden ölçtüğü için nadir hastalıklar gen önceliklendirme

çalışmalarında birincil metrik olarak kabul edilmiştir. Literatürde Ortalama Kesinlik (Average Precision - AP) olarak da tanımlanan bu değer, duyarlılık seviyelerindeki kesinlik değişimlerinin ağırlıklı ortalamasını şu formülle hesaplamaktadır [115]:

$$AP = \sum_{n} (R_n - R_{n-1}) P_n \tag{3.5}$$

Burada P_n ve R_n, ilgili eşikteki kesinlik ve duyarlılık değerlerini ifade eder. AUPRC, özellikle pozitif sınıfın nadir olduğu dengesiz veri setlerinde modelin duyarlılık (*recall*) ve kesinlik (*precision*) dengesini ne kadar başarılı kurduğunu gösterir. Gerçek pozitifleri, aday listesinin en üst sıralarına taşıma yeteneğini ölçmektedir.

3.6 Modelin Uygulanması ve Karar Mekanizması

Geliştirilen XGBoost modelinin başarısı ölçüldükten sonra, modelin siliyer biyolojiye dair keşif potansiyelini ölçmek ve tahminlerinin biyolojik temellerini analiz etmek amacıyla üç aşamalı bir değerlendirme stratejisi izlenmiştir.

3.6.1 Aday Genlerin İn-silico Taraması

Çalışmanın başlangıç noktasını, HGNC veri tabanında tanımlı olan ve insan genomunun protein kodlayan kısmını temsil eden yaklaşık 19.000 genlik ana küme oluşturmuştur. Ancak, makine öğrenmesi modelinin tahmin performansını düşüren veri seyrekliğini (*data sparsity*) ve biyolojik gürültüyü minimize etmek amacıyla, bu geniş küme üzerinde daraltma işlemi uygulanmıştır. Bu kapsamda; HPO, DisGeNET ve Pozitif Setin) birleşiminden oluşan, en az bir fenotipik veya patolojik kaydı literatürde mevcut olan 6.102 özgün gen ($n = 6.102$), çalışmanın "analiz evreni" (*analysis universe*) olarak tanımlanmıştır.

Modelin daha önce eğitim ve test süreçlerinde hiç karşılaşmadığı genlerin nesnel bir şekilde skorlanabilmesi için bu evren üzerinden iki aşamalı bir dışlama prosedürü izlenmiştir. İlk aşamada, 6.102 gen içerisinden 1.183 pozitif siliyopati geni dışlanarak siliyopati durumu belirsiz gen havuzu oluşturulmuştur. İkinci aşamada ise, modelin eğitimi sırasında Negatif Set (kontrol grubu) olarak rastgele seçilen 1.183 gen de bu havuzdan çıkarılmıştır. Sonuç olarak, siliyopati ile ilişkisi literatürde henüz tanımlanmamış olan ve modelin eğitim safhasında hiç görmediği 3.736 özgün genden oluşan "keşif kümesi" ($n = 3.736$) elde edilmiştir. Bu

kümedeki her bir gen, XGBoost modelinin `predict_proba` fonksiyonu ile analiz edilerek 0 ile 1 arasında değişen bir "Siliyopati Olasılık Skoru" (*Prediction Probability Score*) ile derecelendirilmiştir. Elde edilen skorların büyükten küçüğe sıralanması neticesinde, siliyer genlerle en yüksek öznitelik benzerliği gösteren genler, ileri araştırmalar için "öncelikli aday genler" olarak elde edilmiştir.

3.6.2 SHAP Analizi

Makine öğrenmesi modellerinin, özellikle biyoinformatik karar destek süreçlerindeki en büyük kısıtı "kara kutu" (black-box) yapılarıdır. Bu çalışmada, modelin tahminlerini şeffaf ve biyolojik olarak yorumlanabilir hale getirmek amacıyla Açıklanabilir Yapay Zeka (Explainable AI (XAI)) yaklaşımı benimsenmiştir. Bu kapsamda, özniteliklerin model çıktısına katkısını matematiksel olarak hesaplayan SHAP (SHapley Additive exPlanations) yöntemi kullanılmıştır. Analizler, yöntemsel şeffaflığı sağlamak adına iki şekilde yürütülmüştür:

1. **Küresel Yorumlanabilirlik (Global Interpretability):** Modelin siliyopati kavramını hangi biyolojik değişkenler üzerinden öğrendiğini ve genel karar verme eğilimlerini belirlemek amacıyla, SHAP değerleri eğitim seti (X_{egitim_s}) üzerinden hesaplanmıştır. *Beeswarm* grafiği ile pLI, LOEUF, HPO skoru ve PPI merkeziyeti gibi özniteliklerin modelin genel tahmin başarısı üzerindeki kümülatif etkisi analiz edilmiştir.

2. **Lokal Yorumlanabilirlik (Local Interpretability):** Modelin bilinmeyen gen havuzunda yaptığı keşfin güvenilirliğini test etmek amacıyla, analizler önceliklendirilen en yüksek skorlu aday gen üzerinde yoğunlaşılmıştır. *Force Plot* grafiği ile bu aday genin siliyopati olasılık skorlarını artıran ve azaltan öznitelikler vektörel olarak görselleştirilmiştir. Bu analiz, modelin biyolojik tutarlılığını (biological plausibility) bireysel gen düzeyinde göstermek için kullanılmıştır.

3.6.3 Biyolojik Validasyon: PPI ve Fonksiyonel Zenginleştirme

Önceliklendirilen aday genlerin siliyer biyoloji ile olan potansiyel ilişkilerini doğrulamak amacıyla sistem biyolojisi yaklaşımları entegre edilmiştir.

3.6.3.1 Protein-Protein Etkileşim (PPI) Ağ Analizi

İlk 100 aday genin, siliyer interaktom içindeki fiziksel ve fonksiyonel komşuluklarını haritalandırmak ve bilinen siliyer ağa olan yapısal yakınlıklarını test etmek amacıyla PPI ağ analizi gerçekleştirilmiştir. Analizlerde, STRING veritabanında "orta güvenilirlik" eşiği olarak tanımlanan 400 ham skoru, *MinMaxScaler* standardizasyonu sonrası 0,4 normalize değerine karşılık gelen eşik ve üzeri etkileşimler baz alınarak *NetworkX* kütüphanesi ile ağ topolojisi oluşturulmuştur. Genlerin ağ içindeki konumlarını ve etkileşim yoğunluklarını belirlemek için "Derece Merkeziliği" (*Degree Centrality*) metriği hesaplanmıştır. Bu analiz ile önceliklendirilen adayların siliyer protein komplekslerine entegrasyon potansiyellerinin nicelleştirilmesi ve ağ üzerindeki topolojik rollerinin belirlenmesi amaçlanmıştır.

3.6.3.2 Fonksiyonel Zenginleştirme Analizi

Model tarafından en yüksek olasılıkla önceliklendirilen ilk 100 aday genin biyolojik tutarlılığını ve fonksiyonel benzerliklerini saptamak amacıyla zenginleştirme analizi uygulanmıştır. Bu amaçla *g:Profiler* aracı üzerinden Gene Ontology (GO:BP), KEGG ve HPO veritabanları kullanılmıştır. İstatistiksel anlamlılığın korunması için Benjamini-Hochberg FDR düzeltmesi uygulanmış ve düzeltilmiş p-değeri ($< 0,05$) olan terimler anlamlı kabul edilmiştir. Bu analizin temel amacı, aday gen kümesinin siliyer biyogenez, organel fonksiyonu veya siliyopati ilişkili klinik fenotipler (örneğin kistik böbrek, retinitis pigmentosa) ile istatistiksel olarak anlamlı bir örtüşme sergileyip sergilemediğini metodolojik olarak test etmektir.

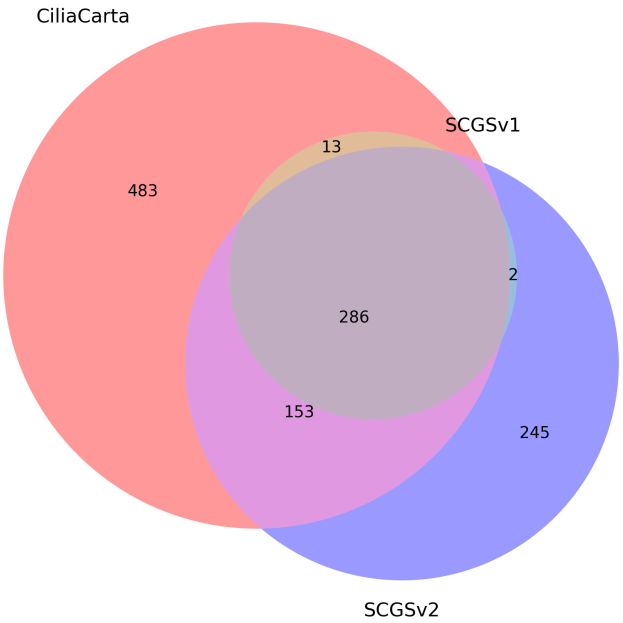
4
BULGULAR

Bu bölümde, Bölüm 3'te detaylandırılan makine öğrenmesi metodolojisi kullanılarak elde edilen analiz sonuçları sunulmaktadır. Bulgular; (1) veri setinin istatistikleri, (2) öznitelik önem düzeyleri, (3) modelin performans metrikleri, (4) tahmin edilen aday genler ve (5) aday genlerin PPI ve fonksiyonel zenginleştirme analizlerini kapsamaktadır.

4.1 Veri Seti İstatistikleri ve Öznitelik Dağılım Bulguları

4.1.1 Veri Seti İstatistikleri

Pozitif veri seti, literatürdeki üç siliyer kaynağın (CiliaCarta, SYSCILIA v1 ve SYSCILIA v2) birleştirilmesi ile elde edilmiştir. Bu kaynakların kesişim kümesi Şekil 4.1'de sunulmuştur.



Şekil 4.1 Pozitif Seti oluşturan veri kaynaklarının Venn şeması.

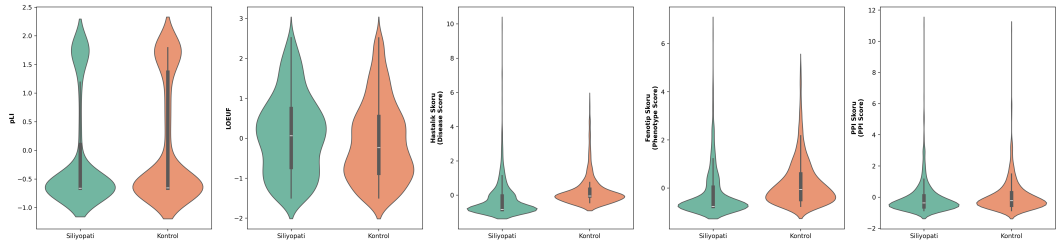
Analiz sürecinde, HGNC veritabanında yer alan yaklaşık 19.000 protein kodlayan gen başlangıç noktası olarak alınmıştır. Ancak modelin tahmin başarısını optimize etmek ve veri seyrekleşmesinden (data sparsity) kaynaklanabilecek hesaplamalı hataları önlemek amacıyla uygulanan filtreleme (HPO, DisGeNET ve Pozitif Set) neticesinde, 6.102 genlik analiz evreni oluşturulmuştur. Veritabanlarının entegrasyonu ve tekrar eden kayıtların temizlenmesi neticesinde, Pozitif Eğitim Seti için toplam 1.183 adet siliyer gen tanımlanmıştır; şemada (Şekil 4.1) her bir veritabanından elde edilen gen sayıları ve kaynaklar arasındaki ortak genlerin dağılımı gösterilmiştir. Evren içerisinden belirlenen 1.183 pozitif gen ile rastgele seçilen aynı sayıda (1.183) negatif genin sonucunda toplam 2.366 genden oluşan dengeli bir eğitim ve test seti oluşturulmuştur. Modelin eğitim aşamasında kullanılmayan ve 3.736 gen ise nihai aday gen havuzunu oluşturmuştur. Veri setlerinin sayısal dağılımı Tablo 4.1'te sunulmuştur.

Tablo 4.1 Çalışmada elde edilen gen setlerinin dağılımı.

Veri Kümesi	**Sayı (n)**	**Tanım / Kapsam**
Protein Kodlayan Genom (HGNC)	∼19.000	Teorik Başlangıç Havuzu
Analiz Evreni (Filtrelenmiş)	**6.102**	**Klinik ve Fonksiyonel İlişkili Genom**
• Pozitif Eğitim Seti	1.183	Tanımlanmış Siliyer Genler
• Negatif Eğitim Seti	1.183	Kontrol Grubu (Siliyer Olmayan)
• **Toplam Eğitim ve Test Seti**	**2.366**	**Model Eğitimi ve Doğrulama**
Aday Gen Havuzu (Tahmin)	3.736	Önceliklendirilen Bilinmeyen Genler

4.2 Özniteliklerin İstatistiksel Dağılımı

Egitim setini oluşturan pozitif ve negatif setlerin, model girdisi olarak kullanılan beş biyolojik öznitelik üzerindeki dağılım profilleri Şekil 4.2'de sunulmuştur. Grafiklerde yeşil renk siliyopati genlerini, turuncu renk ise kontrol grubunu temsil etmektedir. Grafiklerdeki keman şekillerinin genişligi ilgili değerdeki veri yoğunluğunu; ortadaki beyaz noktalar medyan değeri; siyah kalın çubuklar çeyrekler arası açıklığı (IQR); ince siyah çizgiler ise uç deger sınırlarını temsil etmektedir. Grafikler soldan başlayarak sırasıyla; pLI, LOEUF, Hastalık Skoru, Fenotip Skoru ve PPI Skoru olmak üzere sıralanmıştır. Analiz edilen özniteliklerden pLI ve LOEUF skorları genlerin fonksiyon kaybına (LoF) karşı toleransını ve evrimsel kısıtlılığını; hastalık ve fenotip skorları ise genlerin siliyopatilerle olan klinik ve fenotipik benzerliğini ölçmektedir. PPI skorları ise protein ürünlerinin interaktom üzerindeki fonksiyonel etkileşimlerini yansıtmaktadır. Bu öznitelik seti, aday genlerin siliyer sistem içindeki yerini hem biyolojik hem de klinik açıdan bütüncül bir şekilde analiz etmektedir.



Şekil 4.2 Öznitelik dağılım profillerinin Violin analizi ile karşılaştırılması.

Elde edilen bulgular şu şekildedir:

- **pLI:** Her iki grupta da bimodal (iki tepeli) bir dağılım gözlenmiştir. pLI değerinin siliyopati genleri için tek başına tam bir ayırt edici olmadığı, ancak dağılımın sağ ucundaki fonksiyon kaybına (LoF) toleranssız bölgede pozitif sınıfın daha belirgin bir kümelenme eğilimi oluşturduğu saptanmıştır.
- **LOEUF:** Siliyopati genlerinin, kontrol grubuna kıyasla daha dar bir aralıkta ve düşük değerlere (yüksek evrimsel kısıtlılık) doğru daha tutarlı bir eğilim sergilediği gözlenmiştir.
- **Fenotip Skoru:** Pozitif sınıfın, yüksek skorlara doğru uzanan karakteristik "uzun kuyruk" (*long-tail*) yapısı ile pleiotropik etki açısından kontrol grubundan belirgin şekilde ayrıştığı gözlenmiştir.
- **Hastalık Skoru:** Kontrol grubu ile siliyopati grubu arasında bu öznitelik bazında belirgin bir ayrışma saptanmıştır. Siliyopati grubunda çeyrekler arası açıklığı (IQR) temsil eden kalın siyah çubuğun, kontrol grubuna kıyasla hem daha uzun olması hem de yüksek skorlara doğru genişlemesi, siliyer genlerin daha geniş bir pleiotropik spektrum sergilediğini göstermektedir.
- **PPI Skoru:** Siliyopati ve kontrol gruplarının bu öznitelik özelinde oldukça benzer (izomorfik) bir dağılım sergilediği ve her iki grupta da yoğunluğun düşük skorlarda kümelendiği görülmektedir. Çeyrekler arası açıklığı (IQR) temsil eden siyah çubuğun siliyopati grubunda milimetrik düzeyde daha uzun ve yukarı meyilli olması, siliyer proteinler arasında çok kısıtlı ancak biyolojik açıdan bir alt kümenin "Hub" (merkezi etkileşim noktası) özelliği taşıdığına işaret etmektedir. Bu görsel benzerlik, STRING verisinin tek başına belirleyici bir kriter olmadığını, ancak diğer özniteliklerle birlikte ele alındığında modelin tahmin gücüne katkı sağladığını göstermektedir.

4.3 Model Performans Bulguları

Geliştirilen XGBoost modelinin performansı; toplam 2366 genin rastgele %70 eğitim ve %30 test şeklinde bölümlendirilmesiyle elde edilen bağımsız test seti üzerinde değerlendirilmiştir. Test seti üzerindeki analizler; (1) temel başarı metrikleri, (2) karışıklık matrisi üzerinden hata dağılımları ve (3) modelin ayırt etme gücünü gösteren eğri analizleri olmak üzere üç başlık altında sunulmuştur.

4.3.1 Sınıflandırma Performansı Metrikleri

Modelin bağımsız test seti üzerindeki genel başarısını gösteren metrikler Tablo 4.2'te sunulmuştur. Geliştirilen XGBoost modelinin performansı; toplam 2366 genden oluşan veri setinin rastgele %70 eğitim (≈1653 gen) ve %30 test (≈713 gen) şeklinde bölümlendirilmesiyle elde edilen bağımsız veriler üzerinde ölçülmüştür.

Tablo 4.2 Modelin bağımsız test seti üzerindeki performans istatistikleri.

Sınıf / Metrik	**Hassasiyet**	**Duyarlılık**	**F1-Skoru**	**Örnek Sayısı**
Kontrol (0)	0,73	0,87	0,80	357
Siliyopati (1)	0,84	0,68	0,75	356
Ortalama / Toplam	**0,79**	**0,78**	**0,77**	**713**

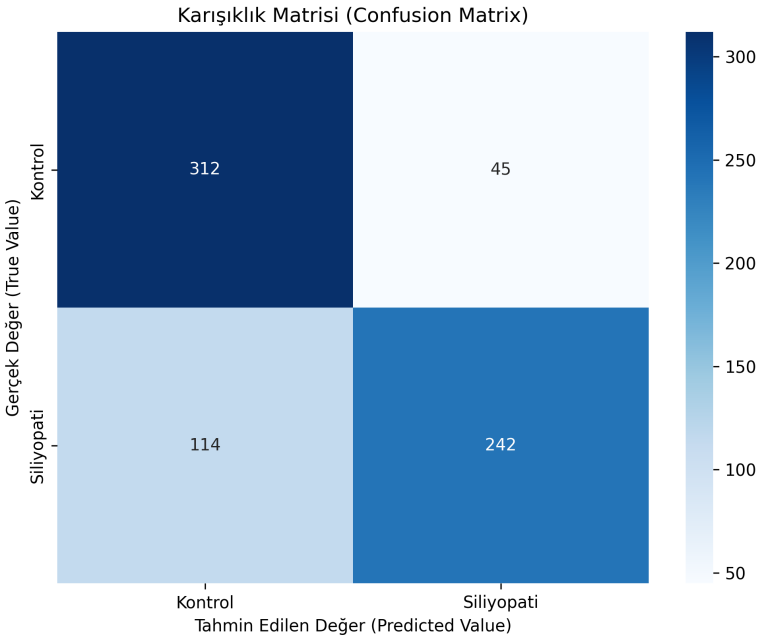
Genel Başarım Metrikleri	**Değer**
Doğruluk (Accuracy)	0,78
Eğri Altı Alan (ROC-AUC)	0,837
Ortalama Hassasiyet (AUPRC/AP)	0,880

Tablo 4.2 verileri incelendiğinde; siliyopati sınıfı için hassasiyet (*precision*) değerinin 0,84, duyarlılık (*recall*) değerinin ise 0,68 olduğu görülmektedir. Bu durum, modelin siliyopati olarak tahmin ettiği genlerin büyük bir bölümünün gerçek pozitiflerden oluştuğunu, ancak pozitif genlerin bir kısmının negatif olarak etiketlendiğine (yanlış negatif) işaret etmektedir. Kontrol grubu üzerinde elde edilen 0,87'lik duyarlılık oranı ise siliyopati ile ilişkisi bulunmayan genlerin model tarafından yüksek başarıyla elendiğini göstermektedir.

Toplam 713 bağımsız test seti üzerinden yapılan analizde model; 0,78 genel doğruluk (*accuracy*), 0,837 ROC-AUC ve 0,880 ortalama hassasiyet (AP) skorlarına ulaşmıştır. Siliyopati sınıfındaki hassasiyet oranı, aday genlerin güvenilirliğini yansıtırken; 0,77'lik ortalama F1-skoru modelin her iki sınıfı ayırt etme kapasitesinin dengeli olduğunu göstermektedir.

4.3.2 Karışıklık Matrisi (Confusion Matrix)

Modelin test setindeki karar dağılımı, Şekil 4.3'de sunulan Karışıklık Matrisi (*Confusion Matrix*) ile görselleştirilmiştir.



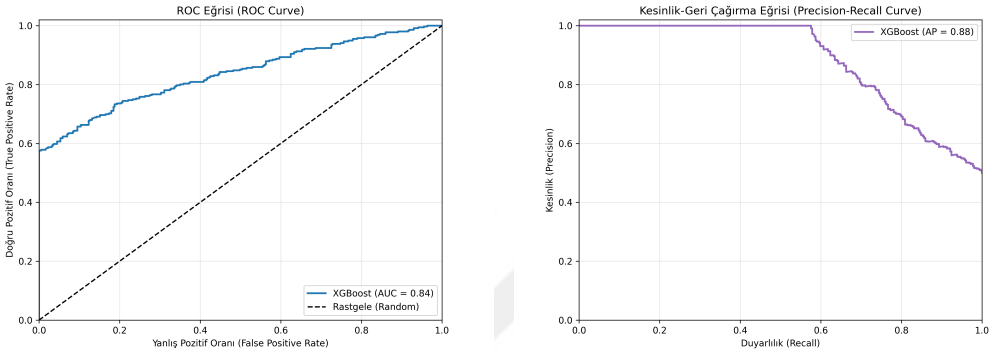
Şekil 4.3 Test seti ($n = 713$) üzerinden elde edilen karışıklık matrisi.

Şekil 4.3'de sunulan karışıklık matrisinde dikey eksen (Y-ekseni) genlerin gerçek biyolojik durumunu (*ground truth*), yatay eksen (X-ekseni) ise modelin bu genler için yaptığı sınıflama tahminlerini göstermektedir. Matris verileri incelendiğinde, modelin kontrol grubunda yer alan 357 genin 312'sini doğru tahmin ederek negatif olarak sınıflandırdığı görülmektedir. Bu gerçek negatif oranı, modelin siliyopatiler ile ilişkisi olmayan genleri ayırt etme başarısını yansıtmaktadır.

Siliyopati grubunda yer alan 356 genin ise 242'si (gerçek pozitif) model tarafından doğru tespit edilmiştir. Buna karşın, 114 gen model tarafından yanlış negatif olarak sınıflandırılmıştır. Bu durum, bazı siliyopati genlerinin sahip olduğu biyolojik sinyallerin kontrol grubuyla benzerlik göstermesinden kaynaklanabilmektedir. Modelin siliyopati olmadığı halde pozitif olarak etiketlediği gen sayısı (yanlış pozitif) ise 45'tir. Yanlış pozitif sayısının düşük olması, modelin tahmin yaparken seçici bir yaklaşım sergilediğini ve aday genlerin deneysel doğrulama süreçlerinde yüksek başarı potansiyeline sahip olduğunu göstermektedir. Sonuçlar, modelin performans istatistiklerini desteklerken; aynı zamanda yanlış tahmin riskinin de minimize edildiğini göstermektedir.

4.3.3 ROC ve PR Eğrileri

Modelin siliyopati ve kontrol gruplarını ayırt etme performansı Şekil 4.4'deki ROC (*Receiver Operating Characteristic*) ve PR (*Precision-Recall*) eğrileriyle sunulmuştur. ROC eğrisi, modelin pozitif ve negatif genleri birbirinden ayırma yeteneğini ölçmektedir. Precision-Recall eğrisi ise siliyopati tahminlerinin ne kadarının doğru olduğunu (*precision*) ve gerçek siliyopati genlerinin ne kadarının başarıyla yakalanabildiğini (*recall*) göstermektedir.



A. ROC Eğrisi B. Precision-Recall Eğrisi

Şekil 4.4 Modelin sınıflandırma performansını gösteren eğri analizleri. (A) ROC (Receiver Operating Characteristic) Eğrisi (B) PR (Precision-Recall) Eğrisi

Şekil 4.4A'da sunulan ve 0,84 olarak hesaplanan ROC-AUC (*Eğri Altı Alan*) skoru, modelin genel sınıflandırma kapasitesini yansıtmaktadır. Bu skor, siliyopati genlerinin kontrol grubuna kıyasla daha yüksek olasılık puanlarıyla önceliklendirildiğini göstermektedir. Şekil 4.4B'de sunulan ve özellikle pozitif sınıfa odaklanan Ortalama Hassasiyet (*Average Precision* (AP)) skoru ise 0,88 olarak belirlenmiştir. PR eğrisinin altında kalan alanı temsil eden bu metrik, modelin hedef genleri tespit etme yeteneğini ölçmektedir. 0,88 düzeyindeki AP skoru, siliyopati genlerinin yüksek doğrulukla ayırt edilerek aday gen listesinin üst sıralarına yerleştirildiğini göstermektedir.

4.4 Aday Gen Önceliklendirme Bulguları

Geliştirilen XGBoost modeli, HGNC veri tabanında tanımlı yaklaşık 19.000 protein kodlayan geni kapsayan başlangıç havuzu üzerinden, *in siliko* önceliklendirme amacıyla eğitilmiştir. Modelin eğitim ve doğrulama aşamalarında kullanılmayan beş temel biyolojik özniteliğin entegre edilmesiyle, 3.736 özgün gen 'keşif kümesi' olarak analize alınmıştır. Olasılık tahminleri sonucunda, $0,50$ eşik değerinin üzerinde siliyopati skoru alan toplam 415 potansiyel aday gen saptanmıştır.

En yüksek skorla önceliklendirilen ilk 20 aday gen ve bu genlerin literatürdeki fonksiyonel özellikleri Tablo 4.3'de sunulmuştur.

Tablo 4.3 XGBoost modeli tarafından tahmin edilen en yüksek skorlu ilk 20 siliyopati aday geni ve fonksiyonel özellikleri.

Sıra	Gen Sembolü	Skor	Bilinen Temel Fonksiyon / Hastalık İlişkisi
1	*CRYBB3*	0,900	Lens gelişimi, Katarakt (*LCA/Alström benzeri*)
2	*VPS33B*	0,876	Vezikül trafiği, ARC Sendromu
3	*PIK3CA*	0,859	Sinyal iletimi (PI3K/AKT), CLOVES Sendromu
4	*LMNA*	0,852	Nükleer zar yapısı, Laminopati
5	*PAX6*	0,845	Göz ve nöral gelişim
6	*COL2A1*	0,842	Kollajen sentezi, İskelet displazileri
7	*MED12*	0,841	Transkripsiyonel regülasyon
8	*SCN1A*	0,841	İyon kanalı, Epileptik ensefalopati
9	*SLCO2A1*	0,838	Prostaglandin taşıyıcısı
10	*COL1A1*	0,838	Ekstraselüler matriks, Osteogenesis Imperfecta
11	*BRAF*	0,837	MAPK sinyal yolağı, RASopatiler
12	*FGFR1*	0,837	Büyüme faktörü reseptörü
13	*BSCL2*	0,831	Lipid metabolizması, Seipinopati
14	*VPS13C*	0,831	Vezikül trafiği, Mitokondriyal homeostaz
15	*PDGFRL*	0,826	PDGF sinyalizasyonu
16	*GLB1*	0,821	Lizozomal enzim, GM1 Gangliosidoz
17	*BSND*	0,817	İyon kanalı alt birimi, Bartter Sendromu
18	*HBB*	0,817	Hemoglobin alt birimi
19	*CACNA1A*	0,816	Kalsiyum kanalı, Ataksi
20	*IL12RB1*	0,816	İmmün sinyalizasyon

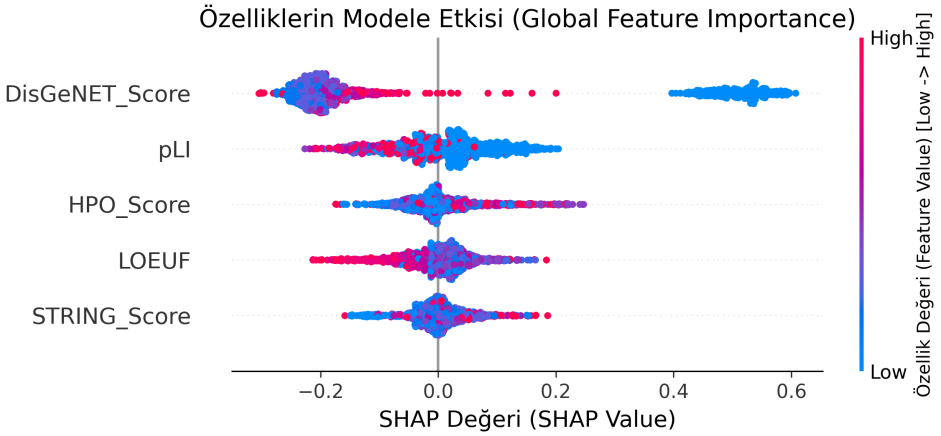
Aday gen listesi incelendiğinde, modelin siliyer fonksiyonlarla, fonksiyonel yakınsama gösteren genleri önceliklendirdiği saptanmıştır. Listenin üst sıralarında yer alan *CRYBB3* (0,90) ve *PAX6* (0,84); oküler gelişim ve retinal homeostaz süreçlerindeki rolleri nedeniyle, siliyopatilerin karakteristik bir bulgusu olan görme kayıpları ve oküler anomalilerle ilişkilendirilmiştir. Modelin bu genleri yüksek olasılık skorlarıyla tahmin etmesinde, fenotipik benzerlik verileri ile oküler morfogenez süreçlerindeki genetik etkileşimlerin belirleyici olduğu görülmektedir. Benzer şekilde, siliyer aksonem ve membran organizasyonu için elzem olan vezikül trafiği ve endolizozomal taşıma proteinlerinden *VPS33B* (0,87) ve *VPS13C* (0,83) genlerinin önceliklendirilmesi; modelin, siliyer biyojenez için gerekli hücre içi taşıma mekanizmalarını ayırt etme kapasitesine işaret etmektedir. Ayrıca, 'İskelet Siliyopatileri' (*Skeletal Ciliopathies*) grubunda izlenen osteokondrodisplazi ve ekstraselüler matriks anomalileriyle uyumlu moleküler karakteristikler sergileyen *COL2A1*, *COL1A1* ve *FGFR1* gibi genlerin varlığı, tahmin sonuçlarının siliyopatilerin klinik tablosuyla gösterdiği paralellik ile uyumludur. Listenin ilk sırasında yer alan *CRYBB3* geninin yüksek olasılıkla tahmin edilme nedenleri, Bölüm 4.5.2'de detaylandırılan lokal SHAP analizi ile açıklanmıştır.

4.5 Modelin Karar Mekanizması Bulguları

Geliştirilen XGBoost modelinin, siliyopati adaylarını belirlerken hangi biyolojik parametreleri dikkate aldığını çözümlemek ve modelin "Kara Kutu" (*Black Box*) niteliğini şeffaflaştırmak amacıyla SHAP (*SHapley Additive exPlanations*) analizi yapılmıştır. Bu analiz, modelin sadece istatistiksel bir korelasyonu mu öğrendiğini, yoksa biyolojik nedenselliğe dayalı tutarlı bir karar mekanizması mı geliştirdiğini doğrulamak için kritik öneme sahiptir. Analiz süreci; (1) özniteliklerin model üzerindeki genel etkilerini kapsayan global SHAP analizi ve (2) seçili aday gen üzerindeki tahmin mekanizmasını açıklayan lokal SHAP analizi olmak üzere iki aşamada gerçekleştirilmiştir.

4.5.1 Açıklanabilir Yapay Zeka (XAI): Global SHAP Analizi

Özniteliklerin model kararları üzerindeki etkileri ve önem sıralaması, Şekil 4.5'deki SHAP özet (beeswarm) grafiği ile analiz edilmiştir.



Şekil 4.5 Özniteliklerin modele etkilerini gösteren SHAP (Beeswarm) grafiği.

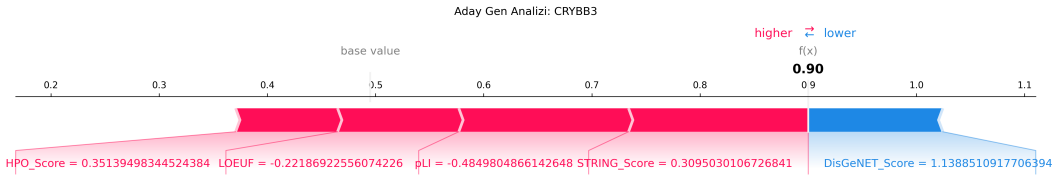
Grafikte öznitelikler önem sırasıyla yukarıdan aşağıya doğru sıralanmış olup; her nokta bir geni temsil etmektedir. Renkler öznitelik değerini (kırmızı: yüksek, mavi: düşük), yatay konum ise model tahminine etki yönünü göstermektedir. Sağ taraftaki noktalar siliyopati olasılığını artıran, sol taraftakiler ise azaltan yöndeki katkıyı ifade etmektedir.

Analiz sonuçlarına göre, siliyopati genlerini ayırt etmede en yüksek etkiye sahip özniteliğin DisGeNET olduğu görülmektedir. DisGeNET değerlerinin pozitif SHAP bölgesinde geniş yayılım sergilemesi, hastalık ilişkileri güçlü olan genlerin model tarafından siliyopati sınıfı lehine değerlendirildiğini göstermektedir. Genomik kısıtlılık metrikleri incelendiğinde, pLI özniteliğinin

model kararlarında ikinci en önemli değişken olduğu görülmektedir. Fonksiyon kaybına toleransı düşük olan genlerin pozitif SHAP değerleriyle ilişkilenmesi, mutasyon intoleransının siliyopati tahmininde belirleyici bir parametre olarak kullanıldığını göstermektedir. Üçüncü değişken olan HPO ise çoklu klinik fenotiplerle ilişkili genlerin pozitif tahmin olasılığını artırması, fenotipik çeşitliliğin model tarafından ayırt edici bir özellik olarak işlendiğine işaret etmektedir. Dördüncü sırada yer alan LOEUF özniteliğinde ise düşük değerlerin (yüksek genomik kısıtlılık) pozitif SHAP değerleriyle örtüşmesi, modelin mutasyonel baskı bilgisini biyolojik açıdan tutarlı bir şekilde öğrendiğini göstermektedir. Son olarak, STRING skoru, genlerin ağ içerisindeki konumlarının modelin karar almasında en az katkıyı sağladığını göstermektedir.

4.5.2 Açıklanabilir Yapay Zeka (XAI): CRYBB3 Geni Lokal SHAP Analizi

Modelin spesifik bir aday geni neden yüksek siliyopati olasılığı ile sınıflandırdığını anlamak amacıyla lokal SHAP analizi (*Force Plot*) uygulanmıştır. Bu doğrultuda, analiz sonucunda 0,90 tahmin skoruyla ilk sıradaki aday gen olarak belirlenen *CRYBB3* (*Crystallin Beta B3*) geni incelenerek, sonuçlar Şekil 4.6'de sunulmuştur.



Şekil 4.6 CRYBB3 geni SHAP (Force Plot) grafiği.

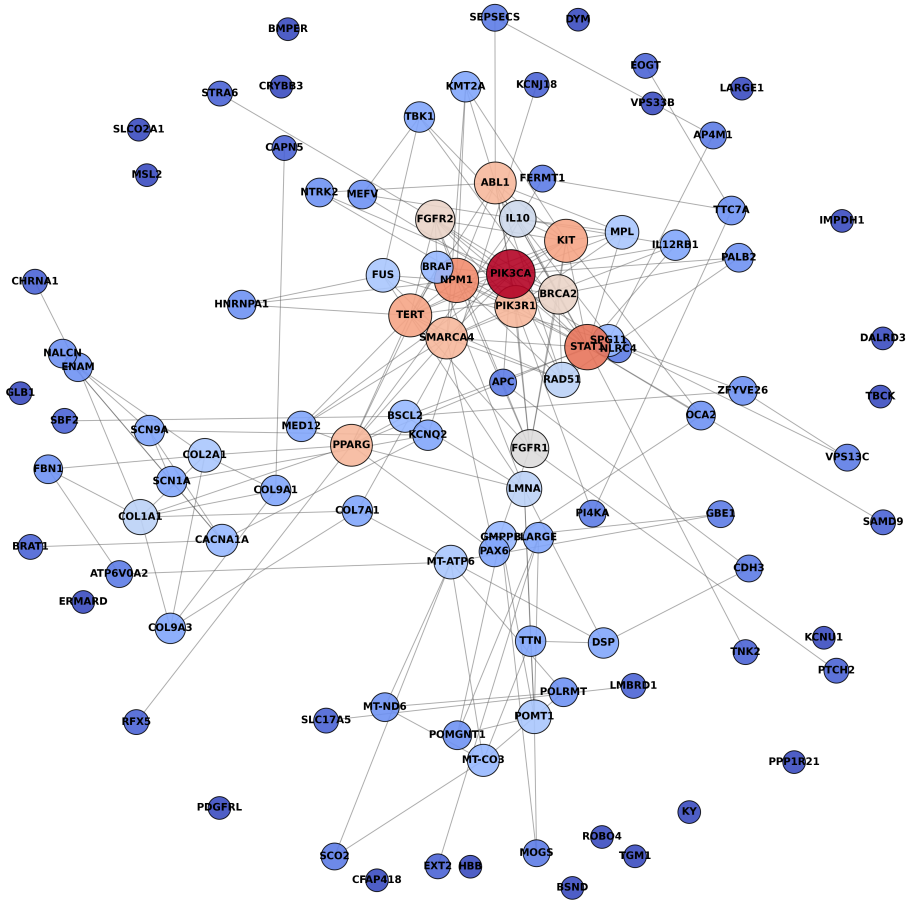
Şekil 4.6'de sunulan lokal SHAP analizi (*force plot*), *CRYBB3* geninin 0,90 düzeyindeki yüksek olasılık skoruyla siliyopati sınıfına atanmasını sağlayan öznitelik katkılarını göstermektedir. Modelin baz değerinden bu tahmin skoruna yükselmesini sağlayan pozitif katkılar (kırmızı oklar); genin siliyer etkileşim ağlarındaki stratejik rolüne ve mutasyonlara karşı evrimsel intoleransına işaret etmektedir. DisGeNET verisinden gelen sınırlayıcı etkiye rağmen (mavi oklar), genomik kısıtlılık ve protein ağ verilerinin birleşik etkisi bu durumu dengeleyerek genin siliyopati sınıfı lehine güçlü bir şekilde önceliklendirilmesini sağlamıştır. Bu analiz, modelin literatürde katarakt ile tanımlanan *CRYBB3* genini, siliyopati spektrumundaki genlerle benzer biyolojik özelliklere sahip olması nedeniyle ileri düzey araştırmalar için potansiyel bir aday olarak önceliklendirebildiğini göstermektedir.

4.6 Aday Genlerin Ağ ve Yolak Analizi Bulguları

XGBoost modeli tarafından önceliklendirilen aday genlerin oluşturduğu protein-protein etkileşim ağları ve ve bu genlerin istatistiksel olarak zenginleştiği sinyal yolakları ve biyolojik süreçler analiz edilmiştir.

4.6.1 Protein-Protein Etkileşim (PPI) Ağı Bulguları

Önceliklendirilen aday genlerin biyolojik bir bütünlük sergileyip sergilemediğini ve siliyopati patogeneziyle ilişkili proteinlerle etkileşimlerini değerlendirmek amacıyla Protein-Protein Etkileşimi (PPI) analizi gerçekleştirilmiştir. Bu kapsamda, modelin en yüksek olasılık skoruna sahip ilk 100 aday geni arasındaki interaktom yapısı STRING veritabanı üzerinden haritalandırılmıştır. Analizde hem fiziksel etkileşimler hem de fonksiyonel ortaklıklar dikkate alınarak, orta-yüksek güvenilirlik düzeyi (skor $> 0,400$) eşik değer olarak belirlenmiştir. Analiz sonucunda elde edilen interaktom haritası Şekil 4.7'de sunulmuştur.

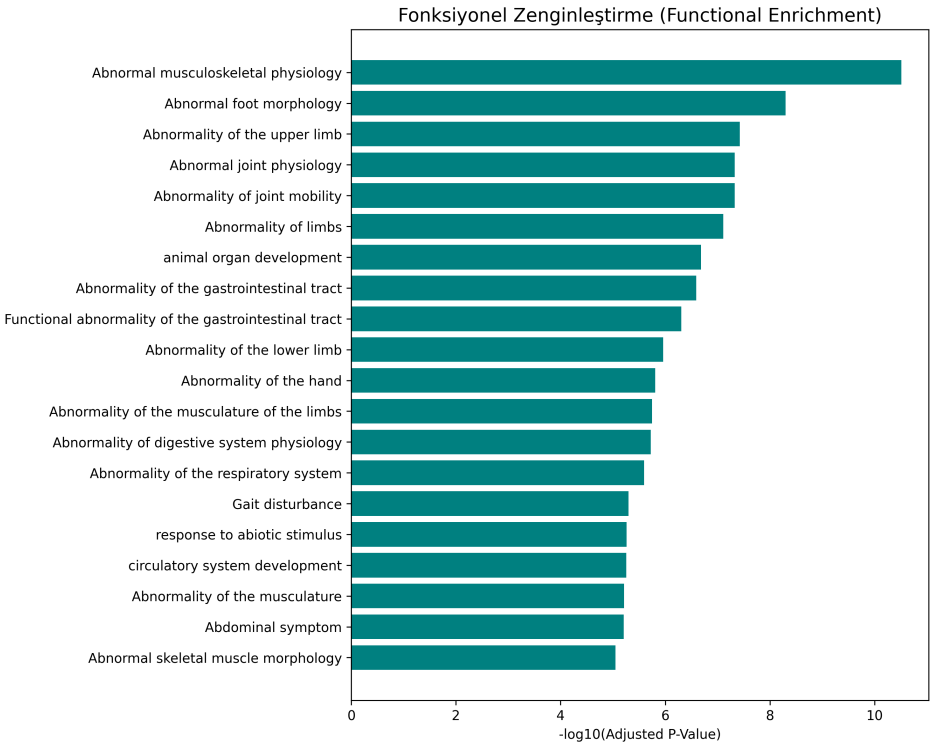


Şekil 4.7 İlk 100 aday genin STRING tabanlı PPI ağı.

Şekil 4.7'de sunulan interaktom haritasında düğümler genleri, çizgiler ise belirlenen eşik değer üzerindeki etkileşimleri simgelemektedir. Analiz sonucunda ağdaki etkileşim sayısının, rastgele beklenen değerden istatistiksel olarak anlamlı derecede yüksek olduğu (PPI zenginleşme p-değeri: $< 1,0 \times 10^{-16}$) saptanmıştır. Özellikle *PIK3CA* gibi yüksek bağlantı derecesine (*degree connectivity*) sahip genlerin merkezi konumda yer alması, modelin bilinen biyolojik ağlarla uyumlu sonuçlar ürettiğini göstermektedir. Buna karşın, olasılık sıralamasında ilk sırada yer alan *CRYBB3* geninin ağ yapısında bir merkezi düğüm (hub) özelliği göstermemesi ve ana gövdeden izole kalması, modelin karar mekanizmasının yalnızca protein etkileşim verisine dayanmadığını yansıtmaktadır. Bu durum, modelin fenotipik benzerlik ve genomik kısıtlılık gibi diğer öznitelikleri belirleyici olarak kullanarak, interaktom haritaları henüz tam aydınlatılmamış olan potansiyel adayları da "fonksiyonel yakınsama" (*guilt by association*) prensibi çerçevesinde önceliklendirebildiğini doğrulamaktadır.

4.6.2 Fonksiyonel Zenginleştirme ve Yolak Analizi Bulguları

Önceliklendirilen ilk 100 aday genin biyolojik fonksiyonlarını değerlendirmek amacıyla g:Profiler üzerinden zenginleştirme analizi yapılmıştır. Elde edilen biyolojik süreç (GO:BP) ve fenotipik küme (HPO) sonuçları Şekil 4.8'de sunulmuştur:



Şekil 4.8 İlk 100 aday genin fonksiyonel zenginleştirme analizi.

Zenginleştirme analizi sonuçları, aday genlerin siliyopatilerin karakteristik klinik spektrumuyla yüksek derecede örtüştüğünü göstermektedir. En yüksek anlamlılık düzeyine sahip HPO terimleri arasında iskelet sistemi fizyolojisindeki bozukluklar ($p = 3,10 \times 10^{-11}$) ve ekstremite anomalilerinin ($p = 5,01 \times 10^{-9}$) yer alması, modelin *COL2A1*, *COL1A1* ve *FGFR1* gibi iskelet displazileriyle ilişkili genleri ayırt etme yeteneğini göstermektedir. Bunun yanı sıra solunum, sindirim ve dolaşım sistemlerini kapsayan viseral organ gelişim terimlerinin zenginleşmesi, siliyer fonksiyonun organogenez ve sinyal iletimindeki rolüyle uyumludur. Özellikle aday listesinin zirvesinde yer alan *CRYBB3* ve *PAX6* gibi genlerin ilişkili olduğu göz ve retina gelişimine yönelik terimlerin Şekil 4.8'de en üst sıralarda yer almaması, bu fenotiplerin anlamsız olduğu anlamına gelmemektedir. Kapsamlı analiz dosyası (Ek C) incelendiğinde, oküler sistemle ilgili terimlerin istatistiksel olarak anlamlı ($p < 0,05$) olduğu, ancak iskelet sistemi fenotiplerinin genom genelinde daha yaygın ve baskın bir zenginleşme paterni sergilemesi nedeniyle grafiksel sıralamada alt basamaklarda kaldığı saptanmıştır. Bu durum, modelin siliyopati spektrumunu sadece tek bir organ sistemiyle sınırlı kalmadan, çoklu sistemik tutulum sergileyen geniş bir perspektifte değerlendirebildiğini göstermektedir.

5
TARTIŞMA

Bu tez çalışması kapsamında, klinik ve genetik heterojenliği nedeniyle moleküler tanısı ve fonksiyonel karakterizasyonu zorlu hastalık gruplarından biri olan siliyopatiler için, yeni aday genleri önceliklendiren bir makine öğrenmesi modeli geliştirilmiştir. Özellikle primer silyum, sinyal iletimi, hücre döngüsü kontrolü ve embriyonik gelişim gibi temel süreçlerin koordinasyon merkezi olarak işlev gördüğü için bu organeldeki disfonksiyonlar geniş bir multisistemik klinik spektruma yol açmaktadır. Siliyopatilerin bu karmaşık doğası, sadece tekil veri setlerine dayalı analizlerin sınırlı kalmasına neden olmakta ve genomik, proteomik ve fenotipik verilerin bütünsel bir yaklaşımla entegre edilmesini zorunlu kılmaktadır.

Çalışmanın bulguları, heterojen biyolojik veri kaynaklarının sistem biyolojisi prensipleri çerçevesinde birleştirilmesinin, siliyer biyolojinin moleküler altyapısını çözümlemede geleneksel yöntemlerden daha etkin bir kapasiteye sahip olduğunu göstermektedir. XGBoost algoritması kullanılarak geliştirilen model, genlerin mutasyon toleransını, protein etkileşim ağlarındaki konumlarını ve fenotipik benzerlik skorlarını aynı analitik düzlemde analiz ederek, siliyopati gen havuzuna dahil edilme potansiyeli yüksek yeni adaylar saptamıştır.

Bu tartışma bölümünde, ilk olarak geliştirilen modelin performans çıktıları ve tahmin yürütürken benimsenen istatistiksel stratejiler ele alınacaktır. Ardından, aday gen sıralamalarında kullanılan özniteliklerin modelin karar mekanizmasına olan katkısı irdelenecektir. Bölümün son kısmında ise, önceliklendirilen aday genlerin siliyer mekanizmalardaki rolleri, protein etkileşim ağlarındaki entegrasyon düzeyleri ve siliyopati patogenezi üzerindeki biyolojik geçerlilikleri değerlendirilecek; çalışmanın metodolojik kısıtlılıkları ve gelecekteki fonksiyonel doğrulama çalışmaları için sunduğu perspektifler tartışılacaktır.

5.1 Model Performansı ve Metodolojinin Değerlendirilmesi

Geliştirilen XGBoost modelinin başarısı, kod kurgusunda benimsenen veri entegrasyon stratejileri ve istatistiksel seçimler ekseninde analiz edilmelidir. Modelin bağımsız test setinde sergilediği 0,84 AUC değeri, kullanılan özniteliklerin siliyer biyolojiyle ilişkili genetik sinyalleri ayırt etmede yüksek kapasiteye sahip olduğunu göstermektedir. Ancak, bu performansın arka planındaki metodolojik kararlar, sonuçların yorumlanması açısından kritik öneme sahiptir.

Çalışmanın veri evreni kurgulanırken, genomdaki tüm genleri kapsamak yerine, "*all_genes_proxy = hpo_genes.union(dis_genes).union(positive_set)*" stratejisi benimsenmiştir. Bu yaklaşım, modelin analiz kapsamını sadece HPO veya DisGeNET veritabanlarında klinik ve fenotipik karşılığı bulunan genlerle sınırlandırmıştır. Yaklaşık 13.000 genin analiz dışı kalması, bu genlerin fenotipik veya hastalık ilişkisi bilgisinin bulunmamasından kaynaklanan teknik bir zorunluluktur. Bu genlerin modele dahil edilmesi, öznitelik matrisinde aşırı seyreklik (*sparsity*) yaratarak modelin öğrenme kapasitesini düşüreceği için, analiz süreci bilgi temelli bir evrende yürütülmüştür [113].

Veri önişleme aşamasında eksik değerlerin (missing values) 0 veya 1 gibi baz değerlerle doldurulması (*imputation*), veri setlerini birleştirmeyi amaçlayan bir tercihtir. Bu sayede, örneğin güçlü bir fenotipik verisi (HPO) olan ancak protein etkileşim (PPI) veya gnomAD kısıtlılık verisi henüz literatüre girmemiş olan genlerin analizden elenmesi engellenmiştir. Modelin bu eksiklikleri kanıt yokluğu olarak değerlendirip analize devam etmesi, aday gen havuzunun biyolojik çeşitliliğini korumuştur [25, 26].

Performans metriklerinde gözlenen Hassasiyet (Precision: 0,84) değerinin Duyarlılık (Recall: 0,68) değerinden yüksek olması, modelin yüksek özgüllük odaklı bir sınıflandırma yürüttüğünü doğrulamaktadır. Eğitim aşamasında uygulanan dengeli örnekleme (*undersampling*) stratejisi, modelin bir geni siliyopati adayı olarak tanımlaması için yüksek düzeyde kanıt eşiği aramasını sağlamıştır. Bu durum, yanlış pozitif (*false positive*) oranını minimize ederken, zayıf sinyalli bazı gerçek pozitif genlerin gözden kaçmasına (düşük recall) yol açmıştır.

Aday gen sıralamalarında gözlenebilen mikro düzeydeki değişimler, rastgele negatif set seçimine (*np.random.choice*) dayanmaktadır. Sabit bir rastlantısallık kilidi (*SEED=42*) kullanılmasına rağmen, kontrol grubunu oluşturan negatif genlerin havuz içinden her seferinde yeniden seçilmesi, XGBoost'un siliyopati olmayan genleri tanımladığı referans noktasında kaymalar yaratabilmektedir. Bu durum, sınırda kalan genlerin olasılık skorlarını etkileyebilmektedir.

5.2 Model Karar Mekanizmasının Değerlendirilmesi

Makine öğrenmesi modellerinin, özellikle de XGBoost gibi kompleks algoritmaların "kara kutu" (black box) olarak nitelendirilen kapalı yapısı, biyoinformatik araştırmalarda sonuçların yorumlanabilirliğini kısıtlayabilmektedir. Bu çalışmada, modelin tahmin stratejisini şeffaf hale getirmek ve özniteliklerin karar sürecine katkısını nicelleştirmek amacıyla Açıklanabilir Yapay Zeka (XAI) yaklaşımlarından biri olan SHAP (SHapley Additive exPlanations) analizi uygulanmıştır. SHAP analizi sonuçları, modelin sadece istatistiksel korelasyonları değil, siliyer biyolojinin temel evrimsel ve klinik prensiplerini de birer karar kriteri olarak öğrendiğini doğrulamaktadır.

SHAP değerlerinin dağılımı incelendiğinde, yüksek pLI skorunun modelin olasılık çıktısını pozitif yönde etkilediği görülmektedir. Silyum, embriyonik gelişim ve organogenez süreçlerinde sinyal iletim merkezi olarak kritik roller üstlendiği için, siliyer proteinleri kodlayan genlerde meydana gelen fonksiyon kaybı mutasyonları genellikle ağır fenotipik sonuçlar doğurmaktadır. Bu durum, siliyer genlerin evrimsel süreçte güçlü bir negatif seleksiyon baskısı altında kalmasına neden olmuştur. Modelin, mutasyon toleransı düşük olan genleri siliyopati adayı olarak önceliklendirmesi, evrimsel kısıtlılık ile hastalık patogenezi arasındaki ilişkiyi doğru bir şekilde modellediğinin göstergesidir. Modelin karar hiyerarşisinde etkili olan bir diğer temel bileşen ise HPO tabanlı fenotipik benzerlik skorudur. Bu durum, modelin genetik veriyi klinik yansımalarla entegre edebilme yeteneğini göstermektedir. Özellikle *CRYBB3* ve *PAX6* gibi genlerin aday listesinin zirvesinde yer alması, modelin yüksek fenotipik benzerliği başarılı bir şekilde saptayabildiğini ve genotip-fenotip korelasyonunu analitik bir düzleme taşıdığını desteklemektedir [77].

5.3 Sistem Biyolojisi Perspektifiyle Değerlendirilmesi

Önceliklendirilen aday genlerin biyolojik rollerini anlamlandırmak amacıyla yürütülen protein etkileşim ağı (PPI) ve fonksiyonel zenginleştirme analizleri, modelin siliyopati patogenezini sistemik bir bütünlük içerisinde kavradığını göstermektedir. Elde edilen bulgular, Sistem Biyolojisi'nin temel prensiplerinden biri olan "Fonksiyonel Yakınsama" (*guilt by association*) ilkesini destekler niteliktedir. PPI ağı incelendiğinde (Şekil 4.7), aday genlerin büyük bir kısmının izole düğümler yerine, çoklu bağlantılar kuran yüksek entegrasyonlu yapılar sergilediği görülmüştür.

Ağın merkezinde yer alan ve yüksek bağlantı derecesine (*degree connectivity*) sahip olan *BRCA2* geninin hub gen olması dikkat çekicidir. Her ne kadar *BRCA2* literatürde DNA tamir mekanizmalarıyla özdeşleşmiş olsa da, ağın merkezindeki konumu, bu genin bazal cisimcik (*basal body*) lokalizasyonu ve siliyer aksonem stabilitesi üzerindeki fonksiyonel rolleriyle uyumludur. Olasılık sıralamasında birinci sırada yer alan *CRYBB3* geninin PPI ağında bir hub özelliği sergilememesi ve ağ yapısında izole bir düğüm olarak kalması, modelin karar mekanizmasının analitik gücünü yansıtmaktadır. Bu durum, modelin tahmini yürütürken protein etkileşim verilerine bağımlı kalmayarak, fenotipik benzerlik ve evrimsel kısıtlılık gibi öznitelikleri daha efektif kullandığını göstermektedir. Bir genin ağda az bağlantılı olmasına rağmen yüksek skor alması, literatürde henüz PPI haritaları tam çıkarılmamış gizli adaylar olduğuna işaret etmektedir.

Diğer yandan, fonksiyonel zenginleştirme analizi sonuçları, aday genlerin siliyopatilerin karakteristik klinik spektrumuyla yüksek derecede örtüştüğünü göstermektedir. Şekil 4.8'de gözlenen istatistiksel dağılımda, iskelet sistemi fizyolojisindeki anomalilerin ezici bir baskınlık kurduğu görülmüştür. Bu durum, *COL2A1* ve *FGFR1* gibi adayların siliyer sinyalizasyonun (özellikle Hedgehog yolağı) iskelet displazileri üzerindeki kritik kontrolünü temsil etmesinden kaynaklanmaktadır. Silyumun kondrosit proliferasyonu ve matriks üretimi üzerindeki doğrudan etkisi, iskelet sistemine yönelik terimlerin analizde en düşük p-değerleriyle (en yüksek anlamlılık) ön plana çıkmasını sağlamıştır [119, 180, 173].

Analiz listesinin zirvesinde oküler fenotiplerle ilişkili *CRYBB3* ve *PAX6* gibi genlerin yer almasına rağmen, Şekil 4.8'deki ilk basamaklarda gözle ilgili terimlerin doğrudan bulunmaması istatistiksel sıralamadan kaynaklanmaktadır [142, 143]. Detaylı analiz verileri (Ek Tablo C), oküler sistemle ilgili terimlerin de istatistiksel olarak anlamlı ($p < 0,05$) düzeyde zenginleştiğini; ancak iskelet ve çoklu sistemik gelişim süreçlerinin aday havuzu içerisindeki sayısal yoğunluğunun daha yüksek olması nedeniyle görsel sıralamada bu terimlerin arkasında kaldığını göstermektedir. Bu bulgular, modelin siliyopati spektrumunu tek bir organa odaklanmadan, geniş bir multisistemik perspektifte değerlendirebildiğini ve siliyopati-göz eksenini (retinal proteostaz) ana klinik tabloların bir parçası olarak başarıyla analiz ettiğini ortaya koymaktadır.

5.4 Çalışmanın Kısıtlılıkları ve Gelecek Perspektifi

Bu tez çalışmasında geliştirilen XGBoost modeli, siliyopati aday genlerinin tahmin edilmesinde istatistiksel olarak anlamlı bir performans ve biyolojik tutarlılık sergilemiş olsa da, kullanılan verilerin yapısından ve metodolojik tercihlerden kaynaklanan bazı kısıtlılıklar bulunmaktadır. Bu kısıtlılıkların değerlendirilmesi, elde edilen bulguların doğru yorumlanması ve gelecekteki araştırma stratejilerinin belirlenmesi açısından önem taşımaktadır. Çalışmanın temel kısıtlılıklarından biri, modelin eğitim setinin genel bir siliyopati havuzu üzerinden kurgulanmış olmasıdır. Siliyer proteinler, farklı dokularda özgün izoformlar halinde bulunabilmekte veya dokuya özel protein kompleksleriyle etkileşime girebilmektedir [199]. Geliştirilen model, tüm dokulardaki ortak siliyer özellikleri öğrenmeye odaklandığı için, sadece belirli bir dokuda veya spesifik bir gelişimsel evrede eksprese edilen genleri gözden kaçırma riski taşımaktadır. Bu durum, modelin duyarlılık (Recall) değerinin %68 seviyesinde kalmasının temel nedenlerinden biri olarak değerlendirilmektedir. İkinci bir kısıtlılık ise, protein-protein etkileşim (PPI) verilerinin statik doğasıdır. STRING veritabanından elde edilen etkileşim skorlarını sunsa da, bu etkileşimlerin hangi hücresel koşullarda veya hücre döngüsünün hangi aşamasında gerçekleştiği bilgisini içermemektedir. Siliyogenez süreçlerinin dinamik yapısı göz önüne alındığında, gelecek çalışmalarda statik ağ verileri yerine dokuya özgü ko-ekspresyon ağlarının modele entegre edilmesi, tahmin çözünürlüğünü artıracak bir geliştirme alanı olarak öne çıkmaktadır.

Bu tez çalışması, kapsamlı bir *in silico* önceliklendirme altyapısı sunmakla birlikte, elde edilen aday gen listesinin klinik ve biyolojik geçerliliğinin kesinleşmesi için laboratuvar ortamında doğrulanması gerekmektedir. Bu bağlamda önerilen gelecek yol haritası, çok aşamalı bir validasyon sürecini kapsamaktadır. İlk aşamada, listenin zirvesinde yer alan *CRYBB3* ve *VPS33B* gibi yüksek skorlu adayların siliyer fonksiyon üzerindeki etkileri model organizmalarda test edilmelidir. Siliyer genetiği yüksek oranda korunan *Danio rerio* (Zebrafish), model organizma olarak önerilmektedir. CRISPR/Cas9 teknolojisiyle oluşturulan fonksiyon kaybı (loss-of-function) *Danio rerio* (Zebrafish) modellerinde retina dejenerasyonu gibi siliyopati bulgularının izlenmesi, bu genlerin hastalıktaki rolünü netleştireceği düşünülmektedir. Eş zamanlı olarak, tanısı konulamamış (*undiagnosed*) siliyopati şüpheli hasta kohortlarına ait Tüm Ekzom Dizileme (WES) verilerinin, bu çalışmada oluşturulan önceliklendirme listesi referans alınarak yeniden filtrelenmesi, klinik tanı oranlarını artırma potansiyeline sahiptir.

6
SONUÇ

Bu tez çalışmasında, XGBoost algoritmasına dayalı, yüksek performanslı bir aday gen önceliklendirme modeli geliştirilmiştir. Yaklaşık 19.000 genlik insan protein kodlayan genomu içerisinden klinik ve fenotipik verilerle süzülen 6.102 genlik analiz evreni üzerinde yürütülen bu çalışma; 0,77 ortalama F1-skoru ve 0,84 AUC değeri ile istatistiksel açıdan güvenilir bir performans sergilemiştir.

Modelin eğitim safhasında hiç görmediği 3.736 genlik keşif kümesi üzerinde yapılan tahminler sonucunda, 0,50 eşik değerinin üzerinde skor alan 415 potansiyel aday gen saptanmıştır. SHAP analizleri ve Protein-Protein Etkileşim (PPI) ağı incelemeleri, modelin tahminlerini sadece teorik verilerle değil; mutasyonel tolerans, fenotipik genişlik ve fonksiyonel yakınsama gibi biyolojik açıdan tutarlı parametrelerle dayandığını doğrulamıştır. Özellikle listede ilk sıralarda öne çıkan adayların, siliyopati patogenezinde daha önce tanımlanmamış kritik rollere sahip olabileceği düşünülmektedir.

Çalışmanın sunduğu *in silico* altyapı, nadir hastalıkların genetik yapısını anlamada tek boyutlu analizler yerine çok boyutlu veri entegrasyonunun önemini vurgulamaktadır. Elde edilen 415 genlik aday listesi; klinik tanısı konulamamış hasta kohortlarına ait ekzom verilerinin yeniden yorumlanması, *Danio rerio* modellerinde fonksiyon kaybı analizleri ve hücresel lokalizasyon çalışmaları için güçlü bir rehber niteliğindedir.

Sonuç olarak bu çalışma, sistem biyolojisi ve makine öğrenmesi yaklaşımlarının siliyopati gibi kompleks genetik hastalıkların aydınlatılmasında ne kadar etkin kullanılabileceğini gösteren bütüncül bir çerçeve sunmaktadır.

KAYNAKÇA

[1] Jose L. Badano et al. “The ciliopathies: an emerging class of human genetic disorders”. *Annual Review of Genomics and Human Genetics* 7 (2006), pp. 125–148. doi: 10.1146/annurev.genom.7.080505.115610.

[2] Jeremy F. Reiter, Michel R. Leroux. “Genes and molecular pathways underpinning ciliopathies”. *Nature Reviews Molecular Cell Biology* 18.9 (2017), pp. 533–547. doi: 10.1038/nrm.2017.60.

[3] Amrita Dasgupta, Jeffrey D. Amack. “Cilia in vertebrate left-right patterning”. *Philosophical Transactions of the Royal Society B: Biological Sciences* 371.1710 (2016), p. 20150410. doi: 10.1098/rstb.2015.0410.

[4] Sarah C. Goetz, Kathryn V. Anderson. “The primary cilium: a signalling centre during vertebrate development”. *Nature Reviews Genetics* 11.5 (2010), pp. 331–344. doi: 10.1038/nrg2774.

[5] Joel L. Rosenbaum, George B. Witman. “Intraflagellar transport”. *Nature Reviews Molecular Cell Biology* 3.11 (2002), pp. 813–825. doi: 10.1038/nrm952.

[6] Hiroaki Ishikawa, Wallace F. Marshall. “Ciliogenesis: building the cell’s antenna”. *Nature Reviews Molecular Cell Biology* 12.4 (2011), pp. 222–234. doi: 10.1038/nrm3085.

[7] Takanori Kobayashi, Brian D. Dynlacht. “Regulating the transition from centriole to basal body”. *The Journal of Cell Biology* 193.3 (2011), pp. 435–444. doi: 10.1083/jcb.201101005.

[8] Francesc R. Garcia-Gonzalo, Jeremy F. Reiter. “Open Sesame: How Transition Fibers and the Transition Zone Control Ciliary Composition”. *Cold Spring Harbor Perspectives in Biology* 9.2 (2017), a028134. doi: 10.1101/cshperspect.a028134.

[9] Maxence V. Nachury et al. “A core complex of BBS proteins cooperates with the GTPase Rab8 to promote ciliary membrane biogenesis”. *Cell* 129.6 (2007), pp. 1201–1213. doi: 10.1016/j.cell.2007.03.053.

[10] Michael R. Knowles, Maimoona Zariwala, Margaret Leigh. “Primary Ciliary Dyskinesia”. *Clinics in Chest Medicine* 37.3 (2016), pp. 449–461. doi: 10.1016/j.ccm.2016.04.008.

[11] Michael R. Knowles et al. “Primary ciliary dyskinesia. Recent advances in diagnostics, genetics, and characterization of clinical disease”. *American Journal of Respiratory and Critical Care Medicine* 188.8 (2013), pp. 913–922. doi: 10.1164/rccm.201301-0059CI.

[12] Nobutaka Hirokawa et al. “Nodal flow and the generation of left-right asymmetry”. *Cell* 125.1 (2006), pp. 33–45. doi: 10.1016/j.cell.2006.03.002.

[13] Danwei Huangfu et al. “Hedgehog signalling in the mouse requires intraflagellar transport proteins”. *Nature* 426.6962 (2003), pp. 83–87. doi: 10.1038/nature02061.

[14] Kevin C. Corbit et al. “Kif3a constrains beta-catenin-dependent Wnt signalling through dual ciliary and non-ciliary mechanisms”. *Nature Cell Biology* 10.1 (2008), pp. 70–76. doi: 10.1038/ncb1670.

[15] Lotte Schneider et al. “PDGFRalphaalpha signaling is regulated through the primary cilium in fibroblasts”. *Current Biology* 15.20 (2005), pp. 1861–1866. doi: 10.1016/j.cub.2005.09.012.

[16] Vinay Singla, Jeremy F. Reiter. “The primary cilium as the cell’s antenna: signaling at a sensory organelle”. *Science* 313.5787 (2006), pp. 629–633. doi: 10.1126/science.1124534.

[17] David G. Birch. “Retinal degeneration in retinitis pigmentosa and neuronal ceroid lipofuscinosis: An overview”. *Molecular Genetics and Metabolism* 66.4 (1999), pp. 356–366. doi: 10.1006/mgme.1999.2829.

[18] Jan D. Marshall et al. “Alström syndrome: genetics and clinical overview”. *Current Genomics* 12.3 (2011), pp. 225–235. doi: 10.2174/138920211795677912.

[19] J. M. Brown, George B. Witman. “Cilia and Diseases”. *Bioscience* 64.12 (2014), pp. 1126–1137. doi: 10.1093/biosci/biu174.

[20] A. Moore, M. Bell. “XGBoost, A Novel Explainable AI Technique, in the Prediction of Myocardial Infarction: A UK Biobank Cohort Study”. *Clinical Medicine Insights: Cardiology* 16 (2022), p. 11795468221133611. doi: 10.1177/11795468221133611.

[21] Matthew Wiens et al. “A Tutorial and Use Case Example of the eXtreme Gradient Boosting (XGBoost) Artificial Intelligence Algorithm for Drug Development Applications”. *Clinical and Translational Science* 18.3 (2025), e70172. doi: 10.1111/cts.70172.

[22] W. Li et al. “XGBoost-based machine learning model combining clinical and ultrasound data for personalized prediction of thyroid nodule malignancy”. *Frontiers in Endocrinology* 16 (2025), p. 1639639. doi: 10.3389/fendo.2025.1639639.

[23] T. Inoue et al. “XGBoost, a Machine Learning Method, Predicts Neurological Recovery in Patients with Cervical Spinal Cord Injury”. *Neurotrauma Reports* 1.1 (2020), pp. 8–16. doi: 10.1089/neur.2020.0009.

[24] V. Zelli et al. “Classification of tumor types using XGBoost machine learning model: a vector space transformation of genomic alterations”. *Journal of Translational Medicine* 21.1 (2023), p. 836. doi: 10.1186/s12967-023-04720-4.

[25] Y. Nohara et al. “Explanation of machine learning models using shapley additive explanation and application for real data in hospital”. *Computer Methods and Programs in Biomedicine* 214 (2022), p. 106584. doi: 10.1016/j.cmpb.2021.106584.

[26] A. V. Ponce-Bobadilla et al. “Practical guide to SHAP analysis: Explaining supervised machine learning model predictions in drug development”. *Clinical and Translational Science* 17.11 (2024), e70056. doi: 10.1111/cts.70056.

[27] J. Wang et al. “Machine Learning and Shapley Additive Explanations Value Integration for Predicting the Prognostic of Anti-N-Methyl-D-Aspartate Receptor Encephalitis: Model Development and Evaluation Study”. *JMIR Medical Informatics* 13 (2025), e75020. doi: 10.2196/75020.

[28] A. Michalski, K. Duraj, B. Kupcewicz. “Leukocyte deep learning classification assessment using Shapley additive explanations algorithm”. *International Journal of Laboratory Hematology* 45.3 (2023), pp. 297–302. doi: 10.1111/ijlh.14031.

[29] Konrad J. Karczewski et al. “The mutational constraint spectrum quantified from variation in 141,456 humans”. *Nature* 581.7809 (2020), pp. 434–443. doi: 10.1038/s41586-020-2308-7.

[30] Peter N. Robinson et al. “The Human Phenotype Ontology: a tool for annotating and analyzing human hereditary disease”. *American Journal of Human Genetics* 83.5 (2008), pp. 610–615. doi: 10.1016/j.ajhg.2008.09.017.

[31] D. Szklarczyk et al. “STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets”. *Nucleic Acids Research* 47.D1 (2019), pp. D607–D613. doi: 10.1093/nar/gky1131.

[32] Janet Piñero et al. “DisGeNET: a comprehensive platform integrating information on human disease-associated genes and variants”. *Nucleic Acids Research* 45.D1 (2017), pp. D833–D839. doi: 10.1093/nar/gkw943.

[33] Teun J. P. van Dam et al. “CiliaCarta: An integrated and validated compendium of ciliary genes”. *PLoS ONE* 14.5 (2019), e0216705. doi: 10.1371/journal.pone.0216705.

[34] Katherine A. Despotes et al. “Primary Ciliary Dyskinesia: A Clinical Review”. *Cells* 13.11 (2024), p. 974. doi: 10.3390/cells13110974.

[35] Julia Wallmeier et al. “Motile ciliopathies”. *Nature Reviews Disease Primers* 6.1 (2020), p. 77. doi: 10.1038/s41572-020-0209-6.

[36] Myrofora Goutaki, Amelia Shoemark. “Diagnosis of Primary Ciliary Dyskinesia”. *Clinics in Chest Medicine* 43.1 (2022), pp. 127–140. doi: 10.1016/j.ccm.2021.11.008.

[37] C. de Ceuninck van Capelle et al. “Proteomic and structural comparison between cilia from primary ciliary dyskinesia patients with a DNAH5 defect”. *Frontiers in Molecular Biosciences* 12 (2025), p. 1593810. doi: 10.3389/fmolb.2025.1593810.

[38] A. Wilken et al. “Primary Ciliary Dyskinesia Associated Disease-Causing Variants in CCDC39 and CCDC40 Cause Axonemal Absence of Inner Dynein Arm Heavy Chains DNAH1, DNAH6, and DNAH7”. *Cells* 13.14 (2024), p. 1200. doi: 10.3390/cells13141200.

[39] J. Y. Hwang et al. “LRRC23 truncation impairs radial spoke 3 head assembly and sperm motility underlying male infertility”. *eLife* 12 (2023), RP90095. doi: 10.7554/eLife.90095.

[40] K. Cai et al. “Structural organization of the C1b projection within the ciliary central apparatus”. *Journal of Cell Science* 134.21 (2021), jcs254227. doi: 10.1242/jcs.254227.

[41] I. Aprea et al. “Pathogenic gene variants in CCDC39, CCDC40, RSPH1, RSPH9, HYDIN, and SPEF2 cause defects of sperm flagella composition and male infertility”. *Frontiers in Genetics* 14 (2023), p. 1117821. doi: 10.3389/fgene.2023.1117821.

[42] G. Burns et al. “Feasibility of Machine Learning Analysis for the Identification of Patients with Possible Primary Ciliary Dyskinesia”. *medRxiv* (2025). Preprint. doi: 10.1101/2025.04.18.25326065.

[43] T. A. Katoh et al. “Immotile cilia mechanically sense the direction of fluid flow for left-right determination”. *Science* 379.6627 (2023), pp. 66–71. doi: 10.1126/science.abq8148.

[44] D. L. Shi. “Breaking Left-Right Symmetry by the Interplay of Planar Cell Polarity, Calcium Signaling and Cilia”. *Cells* 13.24 (2024), p. 2116. doi: 10.3390/cells13242116.

[45] Y. Tanaka, A. Morozumi, N. Hirokawa. “Nodal flow transfers polycystin to determine mouse left-right asymmetry”. *Developmental Cell* 58.16 (2023), 1447–1461.e6. doi: 10.1016/j.devcel.2023.06.002.

[46] M. Messmore et al. “Cilia and Nodal Flow in Asymmetry: An Engineering Perspective”. *Critical Reviews in Biomedical Engineering* 52.3 (2024), pp. 63–82. doi: 10.1615/CritRevBiomedEng.2024051678.

[47] L. Wang et al. “Novel compound heterozygous mutations of DNAH5 identified in a pediatric patient with Kartagener syndrome: case report and literature review”. *BMC Pulmonary Medicine* 21.1 (2021), p. 263. doi: 10.1186/s12890-021-01586-4.

[48] L. Čajánek, E. A. Nigg. “Cep164 triggers ciliogenesis by recruiting Tau tubulin kinase 2 to the mother centriole”. *Proceedings of the National Academy of Sciences* 111.28 (2014), E2841–E2850. doi: 10.1073/pnas.1401777111.

[49] M. Eschment et al. “CEP290 deficiency disrupts ciliary axonemal architecture in human iPSC-derived brain organoids”. *Journal of Cell Science* 138.20 (2025), jcs264092. doi: 10.1242/jcs.264092.

[50] A. Esteve-Garcia et al. “Expanding the Clinical Spectrum of CEP290 Variants: A Case Report on Non-Syndromic Retinal Dystrophy with a Mild Phenotype”. *Genes* 15.12 (2024), p. 1584. doi: 10.3390/genes15121584.

[51] T. G. Drivas, J. Bennett. “CEP290 and the primary cilium”. In: *Advances in Experimental Medicine and Biology*. Vol. 801. Springer, 2014, pp. 519–525. doi: 10.1007/978-1-4614-3209-8_66.

[52] M. Cardenas-Rodriguez et al. “Genetic compensation for cilia defects in cep290 mutants by upregulation of cilia-associated small GTPases”. *Journal of Cell Science* 134.14 (2021), jcs258568. doi: 10.1242/jcs.258568.

[53] E. R. Burnight et al. “CEP290 gene transfer rescues Leber congenital amaurosis cellular phenotype”. *Gene Therapy* 21.7 (2014), pp. 662–672. doi: 10.1038/gt.2014.39.

[54] S. Sun et al. “Intraflagellar transport trains can switch rails and move along multiple microtubules in intact primary cilia”. *Proceedings of the National Academy of Sciences* 122.16 (2025), e2413968122. doi: 10.1073/pnas.2413968122.

[55] D. Portran et al. “Tubulin acetylation protects long-lived microtubules against mechanical ageing”. *Nature Cell Biology* 19.4 (2017), pp. 391–398. doi: 10.1038/ncb3481.

[56] N. Hadas et al. “Machine Learning Analysis of Cilia-Driven Particle Transport Distinguishes Primary Ciliary Dyskinesia Cilia from Normal Cilia”. *bioRxiv* (2025). Preprint. doi: 10.1101/2025.11.02.686130.

[57] S. R. Shinde, A. R. Nager, M. V. Nachury. “Ubiquitin chains earmark GPCRs for BBSome-mediated removal from cilia”. *The Journal of Cell Biology* 219.12 (2020), e202003020. doi: 10.1083/jcb.202003020.

[58] S. K. Singh et al. “Structure and activation mechanism of the BBSome membrane protein trafficking complex”. *eLife* 9 (2020), e53322. doi: 10.7554/eLife.53322.

[59] J. L. Wingfield, K. F. Lechtreck, E. Lorentzen. “Trafficking of ciliary membrane proteins by the intraflagellar transport/BBSome machinery”. *Essays in Biochemistry* 62.6 (2018), pp. 753–763. doi: 10.1042/EBC20180030.

[60] H. K. Hoffman, R. Prekeris. “HOPS-dependent lysosomal fusion controls Rab19 availability for ciliogenesis in polarized epithelial cells”. *Journal of Cell Science* 137.5 (2024), jcs261047. doi: 10.1242/jcs.261047.

[61] Z. Zhou et al. “Impaired cooperation between IFT74/BBS22-IFT81 and IFT25-IFT27/BBS19 causes Bardet-Biedl syndrome”. *Human Molecular Genetics* 31.10 (2022), pp. 1681–1693. doi: 10.1093/hmg/ddab354.

[62] R. Bachmann-Gagescu, J. C. Dempsey, I. G. Phelps, et al. “Joubert syndrome: a model for untangling recessive disorders with extreme genetic heterogeneity”. *Journal of Medical Genetics* 52.8 (2015), pp. 514–522. doi: 10.1136/jmedgenet-2015-103087.

[63] J. Gonçalves, L. Pelletier. “The Ciliary Transition Zone: Finding the Pieces and Assembling the Gate”. *Molecules and Cells* 40.4 (2017), pp. 243–253. doi: 10.14348/molcells.2017.0054.

[64] D. Wang et al. “FUT8-mediated core fucosylation stabilizes TMEM67 to promote ciliogenesis”. *The Journal of Cell Biology* 224.10 (2025), e202412224. doi: 10.1083/jcb.202412224.

[65] X. Wang et al. “Novel biallelic loss-of-function variants in CEP290 cause Joubert syndrome in two siblings”. *Human Genomics* 14.1 (2020), p. 26. doi: 10.1186/s40246-020-00274-4.

[66] M. Ahmed et al. “Two functional forms of the Meckel-Gruber syndrome protein TMEM67 generated by proteolytic cleavage by ADAMTS9 mediate Wnt signaling and ciliogenesis”. *bioRxiv* (2024). Preprint. doi: 10.1101/2024.09.04.611229.

[67] G. Campobasso et al. “Investigating the Role of B9D1 in Meckel-Gruber Syndrome: A Case Report and Comprehensive Literature Review”. *Genes* 16.6 (2025), p. 643. doi: 10.3390/genes16060643.

[68] V. Hartill et al. “Meckel-Gruber Syndrome: An Update on Diagnosis, Clinical Management, and Research Advances”. *Frontiers in Pediatrics* 5 (2017), p. 244. doi: 10.3389/fped.2017.00244.

[69] C. Li et al. “MKS5 and CEP290 Dependent Assembly Pathway of the Ciliary Transition Zone”. *PLoS Biology* 14.3 (2016), e1002416. doi: 10.1371/journal.pbio.1002416.

[70] A. R. Moye et al. “Ciliopathy-associated protein, CEP290, is required for ciliary necklace and outer segment membrane formation in retinal photoreceptors”. *bioRxiv* (2025). Preprint. doi: 10.1101/2025.01.20.633784.

[71] S. Faber et al. “Gene augmentation of LCA5-associated Leber congenital amaurosis ameliorates bulge region defects of the photoreceptor ciliary axoneme”. *JCI Insight* 8.10 (2023), e169162. doi: 10.1172/jci.insight.169162.

[72] Y. J. Kim et al. “Eupatilin rescues ciliary transition zone defects to ameliorate ciliopathy-related phenotypes”. *The Journal of Clinical Investigation* 128.8 (2018), pp. 3642–3648. doi: 10.1172/JCI99232.

[73] D. G. Taub, Q. Liu. “The Role of Intraflagellar Transport in the Photoreceptor Sensory Cilium”. In: *Advances in Experimental Medicine and Biology*. Vol. 854. Springer, 2016, pp. 627–633. doi: 10.1007/978-3-319-17121-0_83.

[74] S. T. Christensen, C. M. Ott. “Cell signaling. A ciliary signaling switch”. *Science* 317.5836 (2007), pp. 330–331. doi: 10.1126/science.1146180.

[75] H. W. Bennett et al. “Novel fibrillar structure in the inversin compartment of primary cilia revealed by 3D single-molecule superresolution microscopy”. *Molecular Biology of the Cell* 31.7 (2020), pp. 619–639. doi: 10.1091/mbc.E19-09-0499.

[76] S. Srivastava, J. A. Sayer. “Nephronophthisis”. *Journal of Pediatric Genetics* 3.2 (2014), pp. 103–114. doi: 10.3233/PGE-14086.

[77] S. Srivastava, S. Molinari E. Raman, J. A. Sayer. “Many Genes-One Disease? Genetics of Nephronophthisis (NPHP) and NPHP-Associated Disorders”. *Frontiers in Pediatrics* 5 (2018), p. 287. doi: 10.3389/fped.2017.00287.

[78] Q. Chen et al. “Phenotypic and genotypic analysis of pediatric nephronophthisis patients with different levels of proteinuria”. *Renal Failure* 47.1 (2025), p. 2598179. doi: 10.1080/0886022X.2025.2598179.

[79] Y. Bai et al. “Primary cilium in kidney development, function and disease”. *Frontiers in Endocrinology* 13 (2022), p. 952055. doi: 10.3389/fendo.2022.952055.

[80] B. K. Yoder. “NPHP and MKS proteins and the ciliary transition zone”. *Cilia* 1.Suppl 1 (2012), O13. doi: 10.1186/2046-2530-1-S1-O13.

[81] D. Zhou et al. “Senior-Loken Syndrome: Ocular Perspectives on Genetics, Pathogenesis, and Management”. *Biomolecules* 15.5 (2025), p. 667. doi: 10.3390/biom15050667.

[82] J. Wang et al. “Pathogenic Variants in CEP290 or IQCB1 Cause Earlier-Onset Retinopathy in Senior-Loken Syndrome Compared to Those in INVS, NPHP3, or NPHP4”. *American Journal of Ophthalmology* 252 (2023), pp. 188–204. doi: 10.1016/j.ajo.2023.03.025.

[83] M. Anand, H. Khanna. “Ciliary transition zone (TZ) proteins RPGR and CEP290: role in photoreceptor cilia and degenerative diseases”. *Expert Opinion on Therapeutic Targets* 16.6 (2012), pp. 541–551. doi: 10.1517/14728222.2012.680956.

[84] P. Mill, S. T. Christensen, L. B. Pedersen. “Primary cilia as dynamic and diverse signalling hubs in development and disease”. *Nature Reviews Genetics* 24.7 (2023), pp. 421–441. doi: 10.1038/s41576-023-00587-9.

[85] N. Vrabič, A. Fakin, M. Tekavčič Pompe. “Spectrum and frequencies of extraocular features reported in CEP290-associated ciliopathy - A systematic review”. *Journal Français d’Ophtalmologie* 47.8 (2024), p. 104232. doi: 10.1016/j.jfo.2024.104232.

[86] S. F. Wang et al. “Review of Ocular Manifestations of Joubert Syndrome”. *Genes* 9.12 (2018), p. 605. doi: 10.3390/genes9120605.

[87] I. Al Alawi et al. “The diagnostic yield of whole exome sequencing as a first approach in consanguineous Omani renal ciliopathy syndrome patients”. *F1000Research* 10 (2021), p. 207. doi: 10.12688/f1000research.40338.2.

[88] A. Horani et al. “Primary ciliary dyskinesia phenotypes and correlation with genotype”. *Current Opinion in Pulmonary Medicine* 31.6 (2025), pp. 628–634. doi: 10.1097/MCP.0000000000001212.

[89] M. G. Turan et al. “CiliaMiner: an integrated database for ciliopathy genes and ciliopathies”. *Database* 2023 (2023), baad047. doi: 10.1093/database/baad047.

[90] M. S. Pir et al. “CilioGenics: an integrated method and database for predicting novel ciliary genes”. *Nucleic Acids Research* 52.14 (2024), pp. 8127–8145. doi: 10.1093/nar/gkae554.

[91] S. S. V. Vasquez, J. van Dam, G. Wheway. “An updated SYSCILIA gold standard (SCGSv2) of known ciliary genes, revealing the vast progress that has been made in the cilia research field”. *Molecular Biology of the Cell* 32.22 (2021), br13. doi: 10.1091/mbc.E21-05-0226.

[92] R. Raycheva et al. “Challenges in mapping European rare disease databases, relevant for ML-based screening technologies in terms of organizational, FAIR and legal principles: scoping review”. *Frontiers in Public Health* 11 (2023), p. 1214766. doi: 10.3389/fpubh.2023.1214766.

[93] A. Hamosh et al. “Online Mendelian Inheritance in Man (OMIM)”. *Human Mutation* 15.1 (2000), pp. 57–61. doi: 10.1002/(SICI)1098-1004(200001)15:1<57::AID-HUMU12>3.0.CO;2-G.

[94] J. S. Amberger et al. “OMIM.org: Online Mendelian Inheritance in Man (OMIM®), an online catalog of human genes and genetic disorders”. *Nucleic Acids Research* 43.Database issue (2015), pp. D789–D798. doi: 10.1093/nar/gku1205.

[95] J. S. Amberger et al. “OMIM.org: leveraging knowledge across phenotype-gene relationships”. *Nucleic Acids Research* 47.D1 (2019), pp. D1038–D1043. doi: 10.1093/nar/gky1151.

[96] S. S. Weinreich et al. “Orphanet: een Europese database over zeldzame ziekten [Orphanet: a European database for rare diseases]”. *Nederlands tijdschrift voor geneeskunde* 152.9 (2008), pp. 518–519.

[97] Janet Piñero et al. “The DisGeNET knowledge platform for disease genomics: 2019 update”. *Nucleic Acids Research* 48.D1 (2020), pp. D845–D855. doi: 10.1093/nar/gkz1021.

[98] S. Köhler et al. “The Human Phenotype Ontology in 2021”. *Nucleic Acids Research* 49.D1 (2021), pp. D1207–D1217. doi: 10.1093/nar/gkaa1043.

[99] M. A. Gargano et al. “The Human Phenotype Ontology in 2024: phenotypes around the world”. *Nucleic Acids Research* 52.D1 (2024), pp. D1333–D1346. doi: 10.1093/nar/gkad1005.

[100] S. Best et al. “Uncovering the burden of hidden ciliopathies in the 100 000 Genomes Project: a reverse phenotyping approach”. *Journal of Medical Genetics* 59.12 (2022), pp. 1151–1164. doi: 10.1136/jmedgenet-2022-108476.

[101] S. Gudmundsson et al. “Variant interpretation using population databases: Lessons from gnomAD”. *Human Mutation* 43.8 (2022), pp. 1012–1030. doi: 10.1002/humu.24309.

[102] R. Amato et al. “A network-based approach to dissect the cilia/centrosome complex interactome”. *BMC Genomics* 15.1 (2014), p. 658. doi: 10.1186/1471-2164-15-658.

[103] D. Szklarczyk et al. “The STRING database in 2025: protein networks with directionality of regulation”. *Nucleic Acids Research* 53.D1 (2025), pp. D730–D737. doi: 10.1093/nar/gkae1113.

[104] Janet Piñero et al. “The DisGeNET cytoscape app: Exploring and visualizing disease genomics data”. *Computational and Structural Biotechnology Journal* 19 (2021), pp. 2960–2967. doi: 10.1016/j.csbj.2021.05.015.

[105] M. Yousef, J. Allmer. “Deep learning in bioinformatics”. *Turkish Journal of Biology* 47.6 (2023), pp. 366–382. doi: 10.55730/1300-0152.2671.

[106] W. Li et al. “Gene Expression Value Prediction Based on XGBoost Algorithm”. *Frontiers in Genetics* 10 (2019), p. 1077. doi: 10.3389/fgene.2019.01077.

[107] Y. Guo et al. “Gene-Based Testing of Interactions Using XGBoost in Genome-Wide Association Studies”. *Frontiers in Cell and Developmental Biology* 9 (2021), p. 801113. doi: 10.3389/fcell.2021.801113.

[108] C. Miller et al. “A review of model evaluation metrics for machine learning in genetics and genomics”. *Frontiers in Bioinformatics* 4 (2024), p. 1457619. doi: 10.3389/fbinf.2024.1457619.

[109] P. Linardatos, V. Papastefanopoulos, S. Kotsiantis. “Explainable AI: A Review of Machine Learning Interpretability Methods”. *Entropy* 23.1 (2020), p. 18. doi: 10.3390/e23010018.

[110] Pengyi Yang et al. “A Review of Ensemble Methods in Bioinformatics”. *Current Bioinformatics* 5.4 (2010). doi: 10.2174/157489310794072508.

[111] C. Shen et al. “DLRAPom: a hybrid pipeline of Optimized XGBoost-guided integrative multiomics analysis for identifying targetable disease-related lncRNA-miRNA-mRNA regulatory axes”. *Briefings in Bioinformatics* 23.2 (2022), bbac046. doi: 10.1093/bib/bbac046.

[112] X. Zhang et al. “Predicting Missing Values in Medical Data via XGBoost Regression”. *Journal of Healthcare Informatics Research* 4.4 (2020), pp. 383–394. doi: 10.1007/s41666-020-00077-1.

[113] Jian Song, Changbin Yu. “Missing Value Imputation using XGboost for Label-Free Mass Spectrometry-Based Proteomics Data”. *bioRxiv* (2021). Preprint. doi: 10.1101/2021.04.08.438945.

[114] E. Richardson et al. “The receiver operating characteristic curve accurately assesses imbalanced datasets”. *Patterns* 5.6 (2024), p. 100994. doi: 10.1016/j.patter.2024.100994.

[115] H. Zhu et al. “AUPRC: a metric for evaluating the performance of in-silico perturbation methods in identifying differentially expressed genes”. *Briefings in Bioinformatics* 26.5 (2025), bbaf426. doi: 10.1093/bib/bbaf426.

[116] P. A. Toussaint et al. “Explainable artificial intelligence for omics data: a systematic mapping study”. *Briefings in Bioinformatics* 25.1 (2023), bbad453. doi: 10.1093/bib/bbad453.

[117] M. Schmidts et al. “Exome sequencing identifies DYNC2H1 mutations as a common cause of asphyxiating thoracic dystrophy (Jeune syndrome)”. *Journal of Medical Genetics* 50.5 (2013), pp. 309–323. doi: 10.1136/jmedgenet-2012-101284.

[118] V. L. Ruiz-Perez et al. “Mutations in a new gene in Ellis-van Creveld syndrome and Weyers acrodental dysostosis”. *Nature Genetics* 24.3 (2000), pp. 283–286. doi: 10.1038/73508.

[119] N. Quadri, P. Upadhyai. “Primary cilia in skeletal development and disease”. *Experimental Cell Research* 431.1 (2023), p. 113751. doi: 10.1016/j.yexcr.2023.113751.

[120] M. Moraes de Lima Perini et al. “Primary cilia in osteoblasts and osteocytes are required for skeletal development and mechanotransduction”. *bioRxiv* (2023). Preprint. doi: 10.1101/2023.12.15.570609.

[121] Z. Chinipardaz et al. “Role of Primary Cilia in Bone and Cartilage”. *Journal of Dental Research* 101.3 (2022), pp. 253–260. doi: 10.1177/00220345211046606.

[122] N. X. Zheng et al. “Primary cilia-associated protein IFT172 in ciliopathies”. *Frontiers in Cell and Developmental Biology* 11 (2023), p. 1074880. doi: 10.3389/fcell.2023.1074880.

[123] W. Chen et al. “Genetic variations in the DYNC2H1 gene causing SRTD3 (short-rib thoracic dysplasia 3 with or without polydactyly)”. *Frontiers in Genetics* 14 (2023), p. 1125473. doi: 10.3389/fgene.2023.1125473.

[124] A. A. Asseri, A. A. Alzoani, B. M. Al Mufarrih. “Whole-Exome Sequencing Identifies DYNC2H1 Mutations as a Cause of Jeune Asphyxiating Thoracic Dystrophy”. *International Medical Case Reports Journal* 17 (2024), pp. 209–214. doi: 10.2147/IMCRJ.S447466.

[125] C. Shak et al. “Disease-associated mutations in WDR34 lead to diverse impacts on the assembly and function of dynein-2”. *Journal of Cell Science* 136.5 (2023), jcs260073. doi: 10.1242/jcs.260073.

[126] J. Walczak-Sztulpa, A. Wawrocka, M. M. Oud. “Identical IFT140 Variants Cause Variable Skeletal Ciliopathy Phenotypes-Challenges for the Accurate Diagnosis”. *Frontiers in Genetics* 13 (2022), p. 931822. doi: 10.3389/fgene.2022.931822.

[127] F. Lian et al. “Recent advances in primary cilia in bone metabolism”. *Frontiers in Endocrinology* 14 (2023), p. 1259650. doi: 10.3389/fendo.2023.1259650.

[128] S. R. Senum, Y. S. M. Li, P. C. Harris. “Monoallelic IFT140 pathogenic variants are an important cause of the autosomal dominant polycystic kidney-spectrum phenotype”. *American Journal of Human Genetics* 109.1 (2022), pp. 136–156. doi: 10.1016/j.ajhg.2021.11.016.

[129] E. Olinger, P. Phakdeekitcharoen, J. A. Sayer. “Biallelic variants in TTC21B as a rare cause of early-onset arterial hypertension and tubuloglomerular kidney disease”. *American Journal of Medical Genetics Part C* 190.1 (2022), pp. 109–120. doi: 10.1002/ajmg.c.31964.

[130] C. Thiel et al. “NEK1 mutations cause short-rib polydactyly syndrome type majewski”. *American Journal of Human Genetics* 88.1 (2011), pp. 106–114. doi: 10.1016/j.ajhg.2010.12.004.

[131] A. Costantini, H. Valta, F. Taylan. “Oligogenic Inheritance of Monoallelic TRIP11, FKBP10, NEK1, TBX5, and NBAS Variants”. *Frontiers in Genetics* 12 (2021), p. 680838. doi: 10.3389/fgene.2021.680838.

[132] J. D. Da Silva et al. “Establishing an objective clinical spectrum in the Ellis-van Creveld syndrome”. *Genetics in Medicine Open* 1.1 (2023), p. 100781. doi: 10.1016/j.gimo.2023.100781.

[133] D. C. Thomas et al. “Role of primary cilia and Hedgehog signaling in craniofacial features of Ellis-van Creveld syndrome”. *American Journal of Medical Genetics Part C: Seminars in Medical Genetics* 190.1 (2022), pp. 36–46. doi: 10.1002/ajmg.c.31969.

[134] K. V. Dorn, C. E. Hughes, R. Rohatgi. “A Smoothened-Evc2 complex transduces the Hedgehog signal at primary cilia”. *Developmental Cell* 23.4 (2012), pp. 823–835. doi: 10.1016/j.devcel.2012.07.004.

[135] S. Unger et al. “Nosology of genetic skeletal disorders: 2023 revision”. *American Journal of Medical Genetics Part A* 191.5 (2023), pp. 1164–1209. doi: 10.1002/ajmg.a.63132.

[136] M. D. Briggs et al. “New therapeutic targets in rare genetic skeletal diseases”. *Expert Opinion on Orphan Drugs* 3.10 (2015), pp. 1137–1154. doi: 10.1517/21678707.2015.1083853.

[137] C. L. Bonatto Paese et al. “Pharmacological intervention of the FGF-PTH axis as a potential therapeutic for craniofacial ciliopathies”. *Disease Models & Mechanisms* 15.8 (2022), p. dmm049611. doi: 10.1242/dmm.049611.

[138] R. Kalot et al. “Primary cilia and actin regulatory pathways in renal ciliopathies”. *Frontiers in Nephrology* 3 (2024), p. 1331847. doi: 10.3389/fneph.2023.1331847.

[139] B. H. M. van der Velden et al. “Explainable artificial intelligence (XAI) in deep learning-based medical image analysis”. *Medical Image Analysis* 79 (2022), p. 102470. doi: 10.1016/j.media.2022.102470.

[140] D. P. Upadhyaya et al. “Explainable Artificial Intelligence (XAI) in the Era of Large Language Models: Applying an XAI Framework in Pediatric Ophthalmology Diagnosis using the Gemini Model”. *AMIA Joint Summits on Translational Science Proceedings*. 2025, pp. 566–575.

[141] M. A. Pramudito et al. “Explainable artificial intelligence (XAI) to find optimal in-silico biomarkers for cardiac drug toxicity evaluation”. *Scientific Reports* 14.1 (2024), p. 24045. doi: 10.1038/s41598-024-71169-w.

[142] O. A. Zin, L. M. Neves, F. L. Motta, et al. “Novel Mutation in CRYBB3 Causing Pediatric Cataract and Microphthalmia”. *Genes* 12.7 (2021), p. 1069. doi: 10.3390/genes12071069.

[143] X. Jiao, F. Kabir, B. Irum, et al. “A Common Ancestral Mutation in CRYBB3 Identified in Multiple Consanguineous Families with Congenital Cataracts”. *PLoS ONE* 11.6 (2016), e0157005. doi: 10.1371/journal.pone.0157005.

[144] M. Karahan et al. “Crystalline gene mutations in Turkish children with congenital cataracts”. *International Ophthalmology* 41.8 (2021), pp. 2847–2852. doi: 10.1007/s10792-021-01843-9.

[145] D. Rayêe, P. A. Wilmarth, A. Cvekl. “Analysis of mouse lens morphological and proteomic abnormalities following deletion of the βB3-crystallin promoter”. *Experimental Eye Research* 260 (2025), p. 110587. doi: 10.1016/j.exer.2025.110587.

[146] A. Barabino et al. “Deregulation of Neuro-Developmental Genes and Primary Cilium Cytoskeleton Anomalies in iPSC Retinal Sheets from Human Syndromic Ciliopathies”. *Stem Cell Reports* 14.3 (2020), pp. 357–373. doi: 10.1016/j.stemcr.2020.02.005.

[147] D. A. Braun, F. Hildebrandt. “Ciliopathies”. *Cold Spring Harbor Perspectives in Biology* 9.3 (2017), a028191. doi: 10.1101/cshperspect.a028191.

[148] K. Hajian-Tilaki. “Receiver Operating Characteristic (ROC) Curve Analysis for Medical Diagnostic Test Evaluation”. *Caspian Journal of Internal Medicine* 4.2 (2013), pp. 627–635.

[149] Ş. K. Çorbacıoğlu, G. Aksel. “Receiver operating characteristic curve analysis in diagnostic accuracy studies: A guide to interpreting the area under the curve value”. *Turkish Journal of Emergency Medicine* 23.4 (2023), pp. 195–198. doi: 10.4103/tjem.tjem_182_23.

[150] A. Ziegler et al. “A snapshot of some pLI score pitfalls”. *Human Mutation* 40.7 (2019), pp. 839–841. doi: 10.1002/humu.23763.

[151] V. Sreekumar, D. P. Norris. “Cilia and development”. *Current Opinion in Genetics & Development* 56 (2019), pp. 15–21. doi: 10.1016/j.gde.2019.05.002.

[152] O. V. Plotnikova, E. N. Pugacheva, E. A. Golemis. “Primary cilia and the cell cycle”. *Methods in Cell Biology* 94 (2009), pp. 137–160. doi: 10.1016/S0091-679X(08)94007-3.

[153] J. T. Eggenschwiler, K. V. Anderson. “Cilia and developmental signaling”. *Annual Review of Cell and Developmental Biology* 23 (2007), pp. 345–373. doi: 10.1146/annurev.cellbio.23.090506.123249.

[154] L. Wang et al. “Ciliary transition zone proteins coordinate ciliary protein composition and ectosome shedding”. *Nature Communications* 13.1 (2022), p. 3997. doi: 10.1038/s41467-022-31751-0.

[155] T. Fujii et al. “Defects in diffusion barrier function of ciliary transition zone caused by ciliopathy variations of TMEM218”. *Human Molecular Genetics* 33.16 (2024), pp. 1442–1453. doi: 10.1093/hmg/ddae083.

[156] P. G. Czarnecki, J. V. Shah. “The ciliary transition zone: from morphology and molecules to medicine”. *Trends in Cell Biology* 22.4 (2012), pp. 201–210. doi: 10.1016/j.tcb.2012.02.001.

[157] Takashi Ishikawa. “Axoneme Structure from Motile Cilia”. *Cold Spring Harbor Perspectives in Biology* 9.1 (2017), a028076. doi: 10.1101/cshperspect.a028076.

[158] C. R. Wood, J. L. Rosenbaum. “Proteins of the ciliary axoneme are found on cytoplasmic membrane vesicles during growth of cilia”. *Current Biology* 24.10 (2014), pp. 1114–1120. doi: 10.1016/j.cub.2014.03.047.

[159] R. Ng, C. Y. Su. “Making sense of cilia: The role of intraflagellar transport”. *PLoS Biology* 22.11 (2024), e3002924. doi: 10.1371/journal.pbio.3002924.

[160] L. B. Pedersen, J. L. Rosenbaum. “Intraflagellar transport (IFT) role in ciliary assembly, resorption and signalling”. *Current Topics in Developmental Biology* 85 (2008), pp. 23–61. doi: 10.1016/S0070-2153(08)00802-8.

[161] S. Meleppattu et al. “Mechanism of IFT-A polymerization into trains for ciliary transport”. *Cell* 185.26 (2022), 4986–4998.e12. doi: 10.1016/j.cell.2022.11.033.

[162] A. Philbrook et al. “Cilia structure and intraflagellar transport differentially regulate sensory response dynamics within and between C. elegans chemosensory neurons”. *PLoS Biology* 22.11 (2024), e3002892. doi: 10.1371/journal.pbio.3002892.

[163] F. Chiuso, R. Delle Donne, A. Feliciello. “Ubiquitylation of BBSome is required for ciliary assembly and signaling”. *EMBO Reports* 24.4 (2023), e55571. doi: 10.15252/embr.202255571.

[164] X. Tian, H. Zhao, J. Zhou. “Organization, functions, and mechanisms of the BBSome in development, ciliopathies, and beyond”. *eLife* 12 (2023), e87623. doi: 10.7554/eLife.87623.

[165] X. Zhang et al. “A cilia-independent function of BBSome mediated by DLK-MAPK signaling in C. elegans photosensation”. *Developmental Cell* 57.12 (2022), 1545–1557.e4. doi: 10.1016/j.devcel.2022.05.005.

[166] E. Forsythe, P. L. Beales. “Bardet-Biedl syndrome”. *European Journal of Human Genetics* 21.1 (2013), pp. 8–13. doi: 10.1038/ejhg.2012.115.

[167] A. Melluso, F. Secondulfo, M. Zacchia. “Bardet-Biedl Syndrome: Current Perspectives and Clinical Outlook”. *Therapeutics and Clinical Risk Management* 19 (2023), pp. 115–132. doi: 10.2147/TCRM.S338653.

[168] E. N. Suspitsin, E. N. Imyanitov. “Bardet-Biedl Syndrome”. *Molecular Syndromology* 7.2 (2016), pp. 62–71. doi: 10.1159/000445491.

[169] J. W. Tomlinson. “Bardet-Biedl syndrome: A focus on genetics, mechanisms and metabolic dysfunction”. *Diabetes, Obesity & Metabolism* 26.Suppl 2 (2024), pp. 13–24. doi: 10.1111/dom.15480.

[170] L. Spahiu, E. Behluli, G. Temaj. “Joubert syndrome: Molecular basis and treatment”. *Journal of Mother and Child* 26.1 (2023), pp. 118–123. doi: 10.34763/jmotherandchild.20222601.d-22-00034.

[171] F. Brancati, B. Dallapiccola, E. M. Valente. “Joubert Syndrome and related disorders”. *Orphanet Journal of Rare Diseases* 5 (2010), p. 20. doi: 10.1186/1750-1172-5-20.

[172] B. L. Maria, E. Boltshauser, T. X. Tran. “Clinical features and revised diagnostic criteria in Joubert syndrome”. *Journal of Child Neurology* 14.9 (1999), pp. 583–591. doi: 10.1177/088307389901400906.

[173] F. Bangs, K. V. Anderson. “Primary Cilia and Mammalian Hedgehog Signaling”. *Cold Spring Harbor Perspectives in Biology* 9.5 (2017), a028175. doi: 10.1101/cshperspect.a028175.

[174] J. Xu, Y. Zhang, J. Pan. “Primary Cilia-Mediated Hedgehog Signaling Regulates Cell Fate During ER Stress-Induced Life or Death Decisions”. *FASEB Journal* 39.18 (2025), e70982. doi: 10.1096/fj.202500391RR.

[175] A. E. Gómez et al. “Systematic analysis of cilia characteristics and Hedgehog signaling in five immortal cell lines”. *PLoS ONE* 17.12 (2022), e0266433. doi: 10.1371/journal.pone.0266433.

[176] Y. G. Han et al. “Hedgehog signaling and primary cilia are required for the formation of adult neural stem cells”. *Nature Neuroscience* 11.3 (2008), pp. 277–284. doi: 10.1038/nn2059.

[177] C. Niehrs, F. Da Silva, C. Seidl. “Cilia as Wnt signaling organelles”. *Trends in Cell Biology* 35.1 (2025), pp. 24–32. doi: 10.1016/j.tcb.2024.04.001.

[178] K. Zhang et al. “Primary cilia are WNT-transducing organelles whose biogenesis is controlled by a WNT-PP1 axis”. *Developmental Cell* 58.2 (2023), 139–154.e8. doi: 10.1016/j.devcel.2022.12.006.

[179] L. T. Vuong, M. Mlodzik. “The complex relationship of Wnt-signaling pathways and cilia”. *Current Topics in Developmental Biology* 155 (2023), pp. 95–125. doi: 10.1016/bs.ctdb.2023.09.002.

[180] H. L. May-Simera, M. W. Kelley. “Cilia, Wnt signaling, and the cytoskeleton”. *Cilia* 1.1 (2012), p. 7. doi: 10.1186/2046-2530-1-7.

[181] N. Skeik, F. I. Jabr. “Kartagener syndrome”. *International Journal of General Medicine* 4 (2011), pp. 41–43. doi: 10.2147/IJGM.S16181.

[182] J. D. Imbrie. “Kartagener’s syndrome: a genetic defect affecting the function of cilia”. *American Journal of Otolaryngology* 2.3 (1981), pp. 215–222. doi: 10.1016/s0196-0709(81)80018-x.

[183] A. E. Kheir et al. “Meckel-Gruber syndrome: A rare and lethal anomaly”. *Sudanese Journal of Paediatrics* 12.1 (2012), pp. 93–96.

[184] A. Turkyilmaz et al. “Meckel-Gruber Syndrome: Clinical and Molecular Genetic Profiles in Two Fetuses and Review of the Current Literature”. *Genetic Testing and Molecular Biomarkers* 25.6 (2021), pp. 445–451. doi: 10.1089/gtmb.2020.0311.

[185] Matthias T. F. Wolf et al. “Nephronophthisis: a pathological and genetic perspective”. *Pediatric Nephrology* 39.7 (2024), pp. 1977–2000. doi: 10.1007/s00467-023-06174-8.

[186] Friedhelm Hildebrandt, Weibi Zhou. “Nephronophthisis-associated ciliopathies”. *Journal of the American Society of Nephrology* 18.6 (2007), pp. 1855–1871. doi: 10.1681/ASN.2006121344.

[187] J. D. Phillips, J. A. van Aalst. “Jeune’s syndrome (asphyxiating thoracic dystrophy): congenital and acquired”. *Seminars in Pediatric Surgery* 17.3 (2008), pp. 167–172. doi: 10.1053/j.sempedsurg.2008.03.006.

[188] B. Aksoy, G. Tigrel. “Report of a Rare Syndromic Retinal Dystrophy: Asphyxiating Thoracic Dystrophy (Jeune Syndrome)”. *Turkish Journal of Ophthalmology* 55.1 (2025), pp. 53–56. doi: 10.4274/tjo.galenos.2024.76574.

[189] P. Mahajan et al. “Ensemble Learning for Disease Prediction: A Review”. *Healthcare (Basel, Switzerland)* 11.12 (2023), p. 1808. doi: 10.3390/healthcare11121808.

[190] B. Naderalvojoud, T. Hernandez-Boussard. “Improving machine learning with ensemble learning on observational healthcare data”. *AMIA Annual Symposium Proceedings. AMIA Symposium*. 2024, pp. 521–529.

[191] D. Che et al. “Decision tree and ensemble learning algorithms with their applications in bioinformatics”. In: *Advances in Experimental Medicine and Biology*. Vol. 696. Springer, 2011, pp. 191–199. doi: 10.1007/978-1-4419-7046-6_19.

[192] D. Nguyen et al. “Ensemble learning using traditional machine learning and deep neural network for diagnosis of Alzheimer’s disease”. *IBRO Neuroscience Reports* 13 (2022), pp. 255–263. doi: 10.1016/j.ibneur.2022.08.010.

[193] H. Kim, Y. Lim. “Bootstrap aggregated classification for sparse functional data”. *Journal of Applied Statistics* 49.8 (2021), pp. 2052–2063. doi: 10.1080/02664763.2021.1889997.

[194] T. Crawley, 3rd Palmer A. G. “Bootstrap Aggregation for Model Selection in the Model-free Formalism”. *Magnetic Resonance (Gottingen, Germany)* 2.1 (2021), pp. 251–264. doi: 10.5194/mr-2-251-2021.

[195] H. Kim, Y. Lim. “Bootstrap aggregated classification for sparse functional data”. *Journal of Applied Statistics* 49.8 (2021), pp. 2052–2063. doi: 10.1080/02664763.2021.1889997.

[196] C. Zhao et al. “BoostTree and BoostForest for Ensemble Learning”. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 45.7 (2023), pp. 8110–8126. doi: 10.1109/TPAMI.2022.3227370.

[197] T. J. P. van Dam et al. “The SYSCILIA gold standard (SCGSv1) of known ciliary components and its applications within a systems biology consortium”. *Cilia* 2.1 (2013), p. 7. doi: 10.1186/2046-2530-2-7.

[198] M. Elfil, A. Negida. “Sampling methods in Clinical Research; an Educational Review”. *Emergency (Tehran, Iran)* 5.1 (2017), e52.

[199] K. Hua, R. J. Ferland. “Primary cilia proteins: ciliary and extraciliary sites and functions”. *Cellular and Molecular Life Sciences* 75.9 (2018), pp. 1521–1540. doi: 10.1007/s00018-017-2740-5.

[200] M. Mc Fie, L. Koneva, A. K. T. Wann. “Ciliary proteins specify the cell inflammatory response by tuning NFκB signalling, independently of primary cilia”. *Journal of Cell Science* 133.13 (2020), jcs239871. doi: 10.1242/jcs.239871.

[201] P. Kohli, M. Höhne, B. Schermer. “The ciliary membrane-associated proteome reveals actin-binding proteins as key components of cilia”. *EMBO Reports* 18.9 (2017), pp. 1521–1535. doi: 10.15252/embr.201643846.

[202] F. Ehrhart, E. L. Willighagen, C. T. Evelo. “A resource to explore the discovery of rare diseases and their causative genes”. *Scientific Data* 8.1 (2021), p. 124. doi: 10.1038/s41597-021-00905-y.

A
EK A: Aday Gen Listesi

Bu ekte, makine öğrenmesi modeli tarafından, "Siliyopati Aday Skoru" ile önceliklendirilen ilk 100 aday gen sunulmuştur.

Sıra	Gen	Skor	Sıra	Gen	Skor	Sıra	Gen	Skor
1	*CRYBB3*	0.8999	35	*RFX5*	0.7806	68	*BMPER*	0.6943
2	*VPS33B*	0.8761	36	*ZFYVE26*	0.7750	69	*ENAM*	0.6938
3	*PIK3CA*	0.8589	37	*MT-ND6*	0.7722	70	*DSP*	0.6914
4	*LMNA*	0.8520	38	*MOGS*	0.7680	71	*PPARG*	0.6889
5	*PAX6*	0.8446	39	*APC*	0.7668	72	*NPM1*	0.6873
6	*COL2A1*	0.8420	40	*SBF2*	0.7659	73	*TNK2*	0.6866
7	*MED12*	0.8414	41	*LMBRD1*	0.7614	74	*AP4M1*	0.6861
8	*SCN1A*	0.8414	42	*MEFV*	0.7569	75	*DALRD3*	0.6860
9	*SLCO2A1*	0.8379	43	*KCNU1*	0.7540	76	*KMT2A*	0.6856
10	*COL1A1*	0.8378	44	*GMPPB*	0.7520	77	*TBK1*	0.6850
11	*BRAF*	0.8374	45	*KCNJ18*	0.7403	78	*ABL1*	0.6848
12	*FGFR1*	0.8371	46	*RAD51*	0.7398	79	*ERMARD*	0.6838
13	*BSCL2*	0.8311	47	*POMT1*	0.7389	80	*SLC17A5*	0.6838
14	*VPS13C*	0.8308	48	*SCO2*	0.7353	81	*BRAT1*	0.6830
15	*PDGFRL*	0.8264	49	*TTC7A*	0.7338	82	*NTRK2*	0.6826
16	*GLB1*	0.8214	50	*FERMT1*	0.7333	83	*FUS*	0.6809
17	*BSND*	0.8173	51	*FBN1*	0.7292	84	*SAMD9*	0.6803
18	*HBB*	0.8167	52	*PI4KA*	0.7268	85	*CHRNA1*	0.6797
19	*CACNA1A*	0.8163	53	*MT-CO3*	0.7219	86	*NALCN*	0.6785
20	*IL12RB1*	0.8158	54	*CAPN5*	0.7198	87	*PIK3R1*	0.6783
21	*KIT*	0.8124	55	*PPP1R21*	0.7165	88	*DYM*	0.6764
22	*COL7A1*	0.8077	56	*CFAP418*	0.7164	89	*KCNQ2*	0.6753
23	*POLRMT*	0.8073	57	*ATP6V0A2*	0.7148	90	*HNRNPA1*	0.6706
24	*FGFR2*	0.8027	58	*PALB2*	0.7139	91	*IMPDH1*	0.6695
25	*KY*	0.8020	59	*LARGE1*	0.7139	92	*IL10*	0.6694
26	*EOGT*	0.8006	60	*MSL2*	0.7128	93	*SCN9A*	0.6692
27	*EXT2*	0.7987	61	*PTCH2*	0.7086	94	*MT-ATP6*	0.6682
28	*POMGNT1*	0.7978	62	*SMARCA4*	0.7069	95	*BRCA2*	0.6667
29	*STAT1*	0.7964	63	*TBCK*	0.7069	96	*TGM1*	0.6655
30	*STRA6*	0.7946	64	*ROBO4*	0.7068	97	*CDH3*	0.6638
31	*TERT*	0.7940	65	*SPG11*	0.6995	98	*TTN*	0.6625
32	*SEPSECS*	0.7930	66	*COL9A3*	0.6962	99	*MPL*	0.6621
33	*GBE1*	0.7885	67	*COL9A1*	0.6952	100	*NLRC4*	0.6620

B
EK B: Python Kodları

Bu ekte, geliştirilen modele ait temel Python kodları sunulmuştur.

```
import pandas as pd
import numpy as np
import io, gzip, requests
from sklearn.model_selection import
    train_test_split
from sklearn.preprocessing import StandardScaler
    , MinMaxScaler
from sklearn.metrics import roc_auc_score,
    classification_report,
    average_precision_score
import xgboost as xgb
import shap
from gprofiler import GProfiler

SEED = 42
np.random.seed(SEED)

ciliacarta = set(pd.read_csv('CiliaCarta.csv')['
    Gene Name'].str.upper())
scgsv2 = set(pd.read_csv('SCGSv2.csv')['Gene
    Name'].str.upper())
positive_set = ciliacarta.union(scgsv2)

hpo_genes = set(pd.read_csv('genes_to_phenotype.
    txt', sep='\t')['gene_symbol'].dropna().str
    .upper())
dis_genes = set(pd.read_csv('genes_to_disease.
    txt', sep='\t')['gene_symbol'].dropna().str
    .upper())
all_genes_proxy = hpo_genes.union(dis_genes).
    union(positive_set)

hpo_score = pd.read_csv('genes_to_phenotype.txt'
    , sep='\t').groupby('gene_symbol')['hpo_id'
    ].nunique().to_frame('HPO_Score')
dis_score = pd.read_csv('genes_to_disease.txt',
    sep='\t').groupby('gene_symbol')['
    disease_id'].nunique().to_frame('Dis_Score'
    )

with gzip.open('gnomad_metrics.txt.bgz', 'rt')
    as f:
   gnomad = pd.read_csv(f, sep='\t')[['gene', '
       pLI', 'oe_lof_upper']].set_index('gene')
gnomad.rename(columns={'oe_lof_upper': 'LOEUF'},
     inplace=True)

links = pd.read_csv("protein.links.txt.gz", sep=
    ' ', compression='gzip')
links = links[links['combined_score'] > 400]
string_score = links.groupby('protein1')['
    combined_score'].sum().to_frame('
    STRING_Score')

master = pd.DataFrame(index=list(all_genes_proxy
    ))
master = master.join(gnomad).join(dis_score).
    join(hpo_score).join(string_score).fillna
    (0)

if master['STRING_Score'].max() > 0:
   master['STRING_Score'] = MinMaxScaler().
       fit_transform(master[['STRING_Score']])

master['label'] = master.index.map(lambda x: 1
    if x in positive_set else 0)

pos_idx = master[master['label'] == 1].index
neg_pool = master[master['label'] == 0].index
neg_sel_idx = np.random.choice(neg_pool, len(
    pos_idx), replace=False)

train_df = master.loc[pos_idx.union(neg_sel_idx)
    ]
X = train_df.drop('label', axis=1)
y = train_df['label']

X_train, X_test, y_train, y_test =
    train_test_split(X, y, test_size=0.3,
    random_state=SEED, stratify=y)

scaler = StandardScaler()
X_train_s = pd.DataFrame(scaler.fit_transform(
    X_train), columns=X.columns, index=X_train.
    index)
X_test_s = pd.DataFrame(scaler.transform(X_test)
    , columns=X.columns, index=X_test.index)

model = xgb.XGBClassifier(n_estimators=100,
    max_depth=5, learning_rate=0.1,
    random_state=SEED)
model.fit(X_train_s, y_train)

probs = model.predict_proba(X_test_s)[:, 1]
auc_score = roc_auc_score(y_test, probs)

aday_genler = list(set(neg_pool) - set(
    neg_sel_idx))
X_aday_s = pd.DataFrame(scaler.transform(master.
    loc[aday_genler].drop('label', axis=1)),
    index=aday_genler)
aday_probs = model.predict_proba(X_aday_s)[:, 1]

results = pd.DataFrame({'Gene': aday_genler, '
    Score': aday_probs}).sort_values('Score',
    ascending=False)

explainer = shap.Explainer(model.predict_proba,
    X_train_s)
shap_values = explainer(X_train_s)

gp = GProfiler(return_dataframe=True)
enr = gp.profile(organism='hsapiens', query=
    results.head(100)['Gene'].tolist())
```

C
EK C: Fenotip Zenginleştirme Analizi Sonuçları

Bu ekte, fenotip zenginleştirme analizinin ilk 50 sonucu sunulmuştur.

Sıra	ID	Fenotip / Terim Adı	*p*-değeri
1	HP:0011843	Abnormal musculoskeletal physiology	3.10×10^{-11}
2	HP:0001760	Abnormal foot morphology	5.01×10^{-09}
3	HP:0002817	Abnormality of the upper limb	3.79×10^{-08}
4	HP:0011729	Abnormality of joint mobility	4.72×10^{-08}
5	HP:0034430	Abnormal joint physiology	4.72×10^{-08}
6	HP:0040064	Abnormality of limbs	7.80×10^{-08}
7	GO:0048513	animal organ development	2.09×10^{-07}
8	HP:0011024	Abnormality of the gastrointestinal tract	2.56×10^{-07}
9	HP:0012719	Functional abnormality of the gastrointestinal tract	4.92×10^{-07}
10	HP:0002814	Abnormality of the lower limb	1.10×10^{-06}
11	HP:0001155	Abnormality of the hand	1.55×10^{-06}
12	HP:0009127	Abnormality of the musculature of the limbs	1.79×10^{-06}
13	HP:0001762	Talipes equinovarus	2.31×10^{-06}
14	HP:0000152	Abnormality of the head	2.70×10^{-06}
15	HP:0011842	Abnormality of skeletal morphology	2.84×10^{-06}
16	HP:0000924	Abnormality of the skeletal system	3.59×10^{-06}
17	HP:0011844	Abnormal connective tissue morphology	3.59×10^{-06}
18	HP:0000707	Abnormality of the nervous system	5.00×10^{-06}
19	HP:0004322	Short stature	6.20×10^{-06}
20	HP:0001871	Abnormality of blood and blood-forming tissues	8.07×10^{-06}
21	HP:0001507	Growth abnormality	9.08×10^{-06}
22	HP:0001392	Abnormality of the liver	9.43×10^{-06}
23	HP:0000478	Abnormality of the eye	1.16×10^{-05}
24	HP:0012373	Abnormal eye physiology	1.20×10^{-05}
25	HP:0002011	Morphological abnormality of the gastrointestinal tract	1.29×10^{-05}
26	HP:0002086	Abnormality of the respiratory system	2.03×10^{-05}
27	HP:0001438	Abnormality of abdomen morphology	2.61×10^{-05}
28	HP:0011442	Abnormality of central motor function	2.91×10^{-05}
29	HP:0000118	Abnormality of the genitourinary system	3.71×10^{-05}
30	HP:0003011	Abnormality of the musculature	3.93×10^{-05}
31	HP:0002664	Neoplasm	4.18×10^{-05}
32	HP:0002715	Abnormality of the immune system	4.41×10^{-05}
33	HP:0000153	Abnormality of the mouth	4.87×10^{-05}
34	HP:0001574	Abnormality of the integument	5.26×10^{-05}
35	HP:0011804	Abnormality of muscle physiology	5.83×10^{-05}
36	HP:0001197	Abnormalities of the prenatal developmental process	6.27×10^{-05}
37	HP:0000598	Abnormality of the ear	7.02×10^{-05}
38	HP:0001939	Abnormality of metabolism/homeostasis	1.09×10^{-04}
39	HP:0000119	Abnormality of the genitourinary tract morphology	1.20×10^{-04}
40	HP:0001626	Abnormality of the cardiovascular system	1.39×10^{-04}
41	HP:0012638	Abnormality of nervous system physiology	1.41×10^{-04}
42	HP:0012639	Abnormality of nervous system morphology	1.41×10^{-04}
43	HP:0000479	Abnormalities of the external eye	1.76×10^{-04}
44	HP:0001608	Abnormality of the voice	1.86×10^{-04}
45	HP:0000818	Abnormality of the endocrine system	1.87×10^{-04}
46	HP:0002087	Abnormality of the upper respiratory tract	1.98×10^{-04}
47	HP:0010923	Abnormality of the adrenal glands	2.26×10^{-04}
48	HP:0000504	Abnormality of the visual system	2.31×10^{-04}
49	HP:0012372	Abnormal visual morphology	2.31×10^{-04}
50	HP:0002088	Abnormal lung morphology	2.44×10^{-04}

TEZDEN ÜRETİLMİŞ YAYINLAR

Konferans Bildirisi

1. R. Koyuncuoğlu ve N. Ö. Özgentürk, "Siliyopati spektrumunda makine öğrenmesi modeliyle aday genlerin önceliklendirilmesi," *14. Ulusal Moleküler Biyoloji ve Biyoteknoloji Kongresi Bildiri Kitabı*, M. Karataş, Ed. Ankara, Türkiye, 2025, s. 25. E-ISBN: 978-625-92960-1-2.