

**İSTANBUL TEKNİK ÜNİVERSİTESİ ★ LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ**

**TUZLU SU DESALİNASYONUNDA  
YAPAY ZEKA KULLANIMI**

**YÜKSEK LİSANS TEZİ**

**Rabiya KARALI TÜRKER**

**Elektronik ve Haberleşme Mühendisliği Anabilim Dalı**

**Telekomünikasyon Mühendisliği Programı**

**EYLÜL 2025**



**İSTANBUL TEKNİK ÜNİVERSİTESİ ★ LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ**

**TUZLU SU DESALİNASYONUNDA  
YAPAY ZEKA KULLANIMI**

**YÜKSEK LİSANS TEZİ**

**Rabiya KARALI TÜRKER  
(504151323)**

**Elektronik ve Haberleşme Mühendisliği Anabilim Dalı**

**Telekomünikasyon Mühendisliği Programı**

**Tez Danışmanı: Dr. Öğr. Üyesi Cemanur AYDINALP PENÇE**

**EYLÜL 2025**



**ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL**

**APPLICATION OF ARTIFICIAL INTELLIGENCE  
IN SALTWATER DESALINATION**



**M.Sc. THESIS**

**Rabiya KARALI TÜRKER  
(504151323)**

**Department of Electronics and Communication Engineering**

**Telecommunications Engineering Programme**

**Thesis Advisor: Asst. Prof. Dr. Cemanur AYDINALP PENÇE**

**SEPTEMBER 2025**



İTÜ, Lisansüstü Eğitim Enstitüsü'nün 504151323 numaralı Yüksek Lisans Öğrencisi Rabiya KARALI TÜRKER, ilgili yönetmeliklerin belirlediği gerekli tüm şartları yerine getirdikten sonra hazırladığı "TUZLU SU DESALİNASYONUNDA YAPAY ZEKA KULLANIMI" başlıklı tezini aşağıda imzaları olan jüri önünde başarı ile sunmuştur.

**Tez Danışmanı :** **Dr. Öğr. Üyesi Cemanur AYDINALP PENÇE** .....  
İstanbul Teknik Üniversitesi

**Jüri Üyeleri :** **Dr. Öğr. Üyesi Gülşah YILDIZ ALTINTAŞ** .....  
İstanbul Teknik Üniversitesi

**Dr. Öğr. Üyesi Resul TUGAY** .....  
Atatürk Üniversitesi

**Teslim Tarihi :** **1 Eylül 2025**  
**Savunma Tarihi :** **10 Eylül 2025**





*Eşime ve aileme,*



## ÖNSÖZ

Bu tez çalışması, yalnızca akademik bir sürecin tamamlanması değil, aynı zamanda içimde bir yerde eksik kalan bir hikayenin tamamlanması anlamına geliyor.

Lisans mezuniyetimden sonra VESTEL’de komponent mühendisi olarak çalışma hayatına başladım. İki yıl sonra, hem yeni bir hayata adım atmak hem de İTÜ’de yüksek lisans eğitimime başlamak üzere büyük bir umutla İstanbul’a taşındım. Yüksek lisans derslerimi iki yıl içinde tamamladım ve tez sürecindeyken yeniden iş hayatına döndüm. İş hayatına dönüş ve İstanbul’da yaşamının zorluğu hayallerimdeki gibi bir yolculuk sunmadı; ancak kariyer yolculuğuma kaldığım yerden devam etmek yeni fırsatların önünü açtı.

Yıllar içinde kariyerimin bir sonraki durağı ASELSAN oldu ve burada komponent mühendisliği biriminin kurulmasında görev alarak yeni bir oluşumun ana karakterlerinden biri olma şansı yakaladım. Tarifi kolay olmayan bambaşka bir macera ve duygu ile yabancı olduğum bir şehirde yeniden başlamak ve tüm bu yoğunluk içerisinde yeniden yüksek lisans eğitimime devam edebilmek oldukça kıymetli. Zaman zaman imkansız gibi görünen bu yolculuk; sabırla, inançla ve yakınlarımın desteğiyle yeniden umutla doldu. Bu süreci başarıyla tamamlayabilmek, hayatımdaki güzel izlerden bir tanesi ve çok büyük bir mutluluk.

En büyük teşekkürü, umudumu yitirdiğim anda yeniden yollarımızın kesişmesiyle bana projesinde çalışma fırsatı veren kıymetli danışman hocam Sayın Dr. Öğr. Üyesi Cemanur AYDINALP PENÇE’ye etmek istiyorum. Çalışmamın her aşamasında bilgi birikimiyle olduğu kadar moral ve motivasyonu da yanımda olan sevgili danışman hocama; sabırla, özveriyle bana rehberlik ettiği için minnettarım.

Bununla birlikte, yarım kalan bu yolculuğumu tamamlamam için bana motivasyon sağlayan ve akademik bilgisiyle bana mentörlük eden değerli eşime, her zaman başarabileceğimi hatırlatarak yanımda olan sevgili anneme, babama, kardeşlerime, dostlarıma ve ekibime teşekkür ediyorum. Bu tez, onların sevgisi ve desteğiyle tamamlandı.

Bu çalışma; MGA-2025-46879 numaralı İTÜ Bilimsel Araştırma Projesi ve 125E412 numaralı TÜBİTAK Projesi tarafından desteklenmiştir.

Eylül 2025

Rabiya KARALI TÜRKER  
(Takım Lideri)



## İÇİNDEKİLER

	<u>Sayfa</u>
<b>ÖNSÖZ</b> .....	<b>ix</b>
<b>İÇİNDEKİLER</b> .....	<b>xi</b>
<b>KISALTMALAR</b> .....	<b>xiii</b>
<b>SEMBOLLER</b> .....	<b>xv</b>
<b>ÇİZELGE LİSTESİ</b> .....	<b>xvii</b>
<b>ŞEKİL LİSTESİ</b> .....	<b>xix</b>
<b>ÖZET</b> .....	<b>xxi</b>
<b>SUMMARY</b> .....	<b>xxv</b>
<b>1. GİRİŞ</b> .....	<b>1</b>
1.1 Su Arıtma Teknolojileri .....	1
1.2 Desalinasyon .....	4
1.3 Ters Ozmoz (RO) Teknolojisi .....	6
1.4 Membran .....	8
1.5 Tezin Amacı .....	10
1.6 Literatürdeki Uygulamalar .....	11
<b>2. HESAPLAMALI AKIŞKANLAR DİNAMİĞİ VE KANAL TASARIMI</b> ..	<b>17</b>
2.1 Kanal Geometrisi Tasarımı ve Çözüm Ağının Oluşturulması .....	17
2.2 Hesaplamalı Akışkanlar Dinamiği (HAD) Modeli .....	19
2.3 HAD Simülasyonu .....	20
2.4 Akış Kanal Geometrisi için Performans Metrikleri .....	21
2.5 Ön Çalışma-1 .....	22
2.6 Ön Çalışma-2 .....	24
2.7 Ön Çalışma Sonuçları .....	26
<b>3. YAPAY ZEKA MODELİ VE PERFORMANS TAHMİNLEMESİ</b> .....	<b>29</b>
3.1 Makine Öğrenmesi Yöntemleri .....	29
3.2 Regresyon Modelleri .....	30
3.2.1 Klasik regresyon yöntemleri .....	31
3.2.2 Ağaç tabanlı regresyon yöntemleri .....	33
3.2.3 Komşuluk tabanlı regresyon yöntemi .....	35
3.3 Hafif Gradyan Artırmalı Regresyon Makinesi: LightGBM .....	35
3.4 Aşırı Gradyan Artırma: XGBoost .....	38
3.5 Kategorik Artırma: CatBoost .....	41
3.6 Uyarlanabilir Yükseltme: AdaBoost .....	43
3.7 Gradyan Artırmalı Regresyon: GBR .....	45
3.8 Fizikle Öğrenen Yapay Sinir Ağı Yöntemi: PINN .....	47
3.9 Yapay Zeka Temelli Performans Tahminleme Modeli .....	49
<b>4. DENEYSEL ÇALIŞMALAR</b> .....	<b>53</b>
4.1 Regresyon Deneyleri .....	53
4.1.1 Veri .....	54
4.1.2 Eğitim ve test .....	55
4.2 Modelin Uygulanması .....	59
4.3 Regresyon Performans Sonuçları .....	60
4.4 Bulgular ve Tartışma .....	62
<b>5. SONUÇLAR VE GELECEKTEKİ ÇALIŞMALAR</b> .....	<b>71</b>
<b>KAYNAKLAR</b> .....	<b>75</b>



## KISALTMALAR

<b>AdaBoost</b>	: Uyarlanabilir Yükseltme
<b>AI</b>	: Yapay Zeka
<b>ANN</b>	: Yapay Sinir Ağı
<b>BPM</b>	: Birleşik Performans Metriği
<b>CatBoost</b>	: Kategorik Artırma
<b>CDI</b>	: Kapasitif Deiyonizasyon
<b>CP</b>	: Konsantrasyon Polarizasyonu
<b>ED</b>	: Elektrodializ
<b>FO</b>	: İleri (İleriye Dönük) Ozmoz
<b>EFB</b>	: Özel Değişken Paketi
<b>GBDT</b>	: Gradyan Artırmalı Karar Ağaçları
<b>GBR</b>	: Gradyan Artırmalı Regresyon
<b>GOSS</b>	: Gradyan Tabanlı Tek Yönlü Örnekleme
<b>HAD</b>	: Hesaplamalı Akışkanlar Dinamiği
<b>HDH</b>	: Nemlendirme–Nem Alma ile Desalinasyon
<b>HY</b>	: Hidratlama Yöntemi
<b>IXR</b>	: İyon Değişim Reçinesi
<b>LightGBM</b>	: Hafif Gradyan Artırma Makinesi
<b>MED</b>	: Çok İşlemlili Distilasyon
<b>MD</b>	: Membran Damıtma
<b>ML</b>	: Makine Öğrenmesi
<b>MSF</b>	: Çok Kademeli Flaş Distilasyonu
<b>MVC</b>	: Mekanik Buhar Sıkıştırma Damıtma
<b>NF</b>	: Nanofiltrasyon
<b>NMRI</b>	: Nükleer Manyetik Rezonans Görüntüleme
<b>PINN</b>	: Fizikle Öğrenen Yapay Sinir Ağı
<b>PIV</b>	: Parçacık Görüntü Velosimetrisi
<b>RF</b>	: Rastgele Orman
<b>RO</b>	: Ters Ozmoz
<b>SC</b>	: Güneş Bacası Yöntemi
<b>SRF</b>	: İkincil Soğutucu ile Dondurarak Desalinasyon
<b>SSD</b>	: Solar Damıtma
<b>SVM</b>	: Destek Vektör Makineleri
<b>SVR</b>	: Destek Vektör Regresyonu
<b>SWM</b>	: Spiral Sarımlı Membran
<b>TFC</b>	: İnce Film Kompozit
<b>TFN</b>	: İnce Film Nanokompozit
<b>TMP</b>	: Transmembran Basınç
<b>TVC</b>	: Termal Buhar Sıkıştırma Damıtma

**VF** : Vakumla Dondurma Desalinasyonu  
**XGBoost** : Aşırı Gradyan Artırma



## SEMBOLLER

$A_p$	: Tuz konsantrasyonuna baęlı ozmotik basınç katsayısı
$C_s$	: Membran yüzeyindeki tuz konsantrasyonu
$\hat{C}_s$	: Normalize edilmiş membran yüzeyindeki tuz konsantrasyonu
$C_b$	: Çözelti içindeki tuz kütle oranı
$C_e$	: Membran yüzeyine adsorbe olabilecek azami tuz konsantrasyonu
$\eta$	: Öğrenme oranı
$f_m(x)$	: $m$ . iterasyonda öğrenilen karar ağacı
$F_m(x)$	: $m$ . iterasyonda elde edilen tahmin fonksiyonu
$F_{m-1}(x)$	: Bir önceki iterasyondaki model tahmin fonksiyonu
$\gamma$	: Yaprak oluşturma maliyeti
$g_i$	: $i$ . örnek için modelin kayıp fonksiyonuna göre hesaplanan birinci türevi (gradyan)
$h_i$	: $i$ . örnek için modelin kayıp fonksiyonuna göre hesaplanan ikinci türevi (Hessian)
$\mathcal{L}$	: Kayıp fonksiyonu
$J_w$	: Permeat akısı
$\hat{J}_w$	: Normalize edilmiş permeat akısı
$P_\Delta$	: Kanal giriş ve çıkışı arasındaki basınç farkı
$\hat{P}_\Delta$	: Normalize edilmiş kanal giriş ve çıkışı arasındaki basınç farkı
$n$	: Tahmin edilen örnek sayısı
$k$	: Membran özgül geçirgenliği
$k_w$	: Membran geçirgenliğini
$k_1$	: Adsorpsiyon katsayısı
$k_2$	: Desorpsiyon katsayısı
$\lambda$	: L2 düzenleme terimi
$\mu$	: Dinamik viskozite
$R_s$	: Kayma gerilmesine baęlı desorpsiyonu
$\mathbf{U}$	: Üç boyutlu hız vektörü
$y_i$	: $i$ . gözleme ait gerçekleşmiş performans metrięi
$\hat{y}_i$	: $i$ . gözleme ait model tarafından tahmin edilen performans metrięi
$\bar{y}$	: Gerçek performans deęerlerinin ortalaması
$\alpha$	: Kayma gerilmesine baęlı desorpsiyon katsayısı
$\rho$	: Tuzlu su yoğunluğu
$\nu$	: Kinematik viskozite



## ÇİZELGE LİSTESİ

	<u>Sayfa</u>
<b>Çizelge 2.1:</b> HAD simülasyonunda kullanılan sınır koşulları.....	<b>21</b>
<b>Çizelge 2.2:</b> Her bir kanal için membran yüzeyindeki ortalama değerler ve kanaldaki giriş çıkış arasındaki basınç farkları. ....	<b>23</b>
<b>Çizelge 2.3:</b> Her bir kanal için membran yüzeyindeki ortalama değerler ve kanaldaki giriş çıkış arasındaki basınç farkları. ....	<b>25</b>
<b>Çizelge 4.1:</b> Regresyon modellerinin eğitim veri seti üzerinde 10 katlı çapraz doğrulama sonucu performanslarının karşılaştırılması.....	<b>61</b>
<b>Çizelge 4.2:</b> Eğitim veri seti üzerinde en iyi performansı gösteren CatBoost regresyon modelinin, test verisi üzerinde hiperparametre ayarı sonrası elde edilen performans ölçütleri. ....	<b>62</b>



## ŞEKİL LİSTESİ

	<u>Sayfa</u>
<b>Şekil 1.1:</b> Desalinasyon teknolojilerinin çalışma prensiplerine göre sınıflandırılması [17].	5
<b>Şekil 1.2:</b> Ters ozmoz (RO) yöntemi [28].	7
<b>Şekil 1.3:</b> Spiral sarılmış membran modülü ve filtrasyon işlemi [39].	9
<b>Şekil 1.4:</b> Spiral sarılmış membran modülünün kesit alanı [39].	9
<b>Şekil 2.1:</b> Rastgele seçilen noktalar ve B-spline yöntemi kullanılarak elde edilen rastgele geometri tasarımı örneği.	18
<b>Şekil 2.2:</b> Rastgele hazırlanmış kanal geometrileri için oluşturulan çözüm ağı örneği.	18
<b>Şekil 2.3:</b> Membran kirlenmesi dağılımı: Rastgele 10-nokta, 15-nokta ve 20-nokta, sinüzoidal model ve referans model.	23
<b>Şekil 2.4:</b> Membran kirlenmesi dağılımı: performans metriğine göre daha yüksek performansa sahip Model 1, Model 2 ve Model 3, ve görece daha düşük performansa sahip Model 4, Model 5 ve Model 6 kanal geometrileri.	26
<b>Şekil 3.1:</b> Membran bilimine uygulanan popüler makine öğrenmesi yöntemlerinin kısa bir sınıflandırması [89].	29
<b>Şekil 3.2:</b> Regresyon modellerinin sınıflandırılması.	31
<b>Şekil 3.3:</b> Regresyon temelli performans tahminleme modelinin akış diyagramı.	52
<b>Şekil 4.1:</b> Veri hazırlığı ve birleşik performans metriğinin (BPM) elde edilmesi.	55
<b>Şekil 4.2:</b> Tüm veri setinin oluşturulması.	56
<b>Şekil 4.3:</b> 10 katlı çapraz doğrulama yöntemi.	57
<b>Şekil 4.4:</b> En iyi performans gösteren CatBoost modelinin hata dağılımı.	63
<b>Şekil 4.5:</b> En iyi performans gösteren CatBoost modelinin tahminleme hata dağılımı.	65
<b>Şekil 4.6:</b> CatBoost için çapraz doğrulamalı yinelemeli özellik eleme (RFECV) eğrisi.	66
<b>Şekil 4.7:</b> CatBoost için öğrenme eğrisi (learning curve).	68



## TUZLU SU DESALINASYONUNDA YAPAY ZEKA KULLANIMI

### ÖZET

Tuzlu suyun desalinasyonu, artan küresel su krizine karşı geliştirilen en kritik teknolojilerden biridir. Geleneksel tatlı su kaynaklarının, nüfus artışı, kentleşme ve iklim değişikliğinin getirdiği talepleri karşılamakta yetersiz kalması nedeniyle alternatif çözümler aranmaktadır. Bu çözümler arasında ters ozmoz (RO), membran temelli desalinasyon yöntemleri içinde en verimli, ölçeklenebilir ve yaygın olarak uygulanan yöntem olarak öne çıkmaktadır. Ancak RO'nun etkinliğine rağmen enerji tüketimi ve membran kirlenmesi gibi operasyonel zorlukları vardır. Bu zorluklar, sistem verimliliğini azaltmakta, işletme maliyetlerini artırmakta ve membran modüllerinin kullanım ömrünü sınırlamaktadır. Bu nedenle, RO sistemlerinde kullanılan membranların, özellikle iç akış kanalı geometrisinin optimize edilmesi, sistem performansını ve sürdürülebilirliğini artırmak açısından önemli hale gelmiştir.

Son yıllarda yapay zeka (AI), özellikle makine öğrenmesi (ML) yöntemleri, membran teknolojisindeki çok boyutlu ve karmaşık problemleri ele almak için güçlü bir araç olarak öne çıkmaktadır. Geleneksel deneysel ve fizik temelli yaklaşımlar her ne kadar temel oluştursa da; zaman, maliyet ve ölçeklenebilirlik açısından sınırlı kalmakta ve membran geometrisi, işletme parametreleri ve performans çıktıları arasındaki karmaşık ilişkileri tam olarak ortaya koyamamaktadır. Literatürde, AI'nın bu sınırlamaları aşarak yüksek boyutlu verileri işleyebilme, doğrusal olmayan desenleri keşfedebilme ve tasarım sürecini yönlendirme potansiyeli giderek daha fazla vurgulanmaktadır. Ancak çoğu mevcut çalışma, ML'yi ya tek başına ya da dar kapsamlı problemler için uygulamakta; fiziksel simülasyonlardan elde edilen derin içgörülerini yeterince entegre edememektedir.

Bu tez çalışması, yenilikçi membran kanal geometrilerinin performansını araştırmak, değerlendirmek ve öngörmek amacıyla hesaplamalı akışkanlar dinamiği (HAD) simülasyonları ve denetimli ML yöntemlerini birleştiren entegre bir çerçeve sunmaktadır. Çalışma, membran performansının akış dağılımı ve yerel türbülans gibi geometrik faktörlerden etkilendiği ve bu performansın alışılmışın dışında kanal yapıları tasarlanarak artırılabilmesi hipotezine dayanmaktadır. Bu yapıların amacı, kirlenmeyi azaltmak, basınç düşüşünü azaltmak ve permeat akısını artırmak yoluyla daha enerji verimli bir desalinasyon sağlamaktır.

Araştırmanın ilk aşamasında, yapılan ön çalışmalar ile bu hipotezin doğruluğu gösterilmiştir. Öncelikle HAD simülasyonları kullanılarak rastgele akış kanal geometrileri oluşturulmuştur. HAD simülasyonlarında; Navier-Stokes, adveksiyon difüzyon, Langmuir adsorption-desorption, permeat akısı (permeate flux) denklemleri ile Darcy yasası kullanılmış ve literatürde varolan düz ve sinüzoidal yapıdaki

geometriler ile karşılaştırılmıştır. Rastgele oluşturulan yeni kanal geometri tasarımlarının membran performansını artırmada etkili olduğu, basınç kaybını minimum seviyede tutarken enerji ve su kalitesini iyileştirdiği tespit edilmiştir.

Araştırmanın ikinci aşamasında, sayısal simülasyonlarla oluşturulan veri seti artırılmıştır. B-spline eğrileri kullanılarak üst ve alt kanal yüzeylerini tanımlayan kontrol noktaları rastgele seçilerek 1045 farklı membran akış kanalı geometrisi oluşturulmuştur. Bu geometriler, basit sinüzoidal şekillerden karmaşık, düzensiz dalga desenlerine kadar çeşitli tasarımları içermektedir. Her geometri için yapılandırılmış bir ağ oluşturularak akış dinamikleri, konsantrasyon dağılımları ve kirlenme davranışını analiz eden 3 boyutlu HAD simülasyonları gerçekleştirilmiştir. Her simülasyon, sabit çalışma koşullarında 100 saniye sürdürülmüştür.

HAD sonuçlarından üç temel performans metriği çıkarılmıştır: (1) permeat akısı (membrandan geçen su miktarı), (2) kirlenme indeksi (membran yüzeyinde çözünen partikül birikimi) ve (3) kanal giriş ve çıkışı arasındaki basınç düşüşü (akışı sürdürmek için gerekli enerji). Bu metrikler normalize edilip ağırlıklandırılarak enerji verimliliği ve su üretkenliği arasındaki dengeyi temsil eden birleşik performans metriği (BPM) olarak ifade edilmiştir.

Oluşturulan 1045 farklı membran akış kanalı geometrisinden ağ oluşturma ve sayısal kararlılık amacıyla nihai 1025 geometri seçilerek veri seti oluşturulmuştur. Kanal geometrileri, spline eğrilerini oluşturan kontrol noktalarının 42 y-koordinatından oluşan sayısal vektörler olarak ifade edilmiştir. Bu veriler, ML modeline giriş olarak kullanılırken BPM ise regresyon hedefi olarak tanımlanmıştır.

Çalışmanın devamında, geometrik özelliklerden membran performansını doğrudan tahmin etmek amacıyla denetimli regresyon yöntemi kullanılmıştır. Veri seti rastgele iki gruba ayrılmıştır: 900 geometri eğitim ve çapraz doğrulama için, 125 geometri ise test seti olarak ayrılmıştır. Çapraz doğrulama 10 katlı strateji ile gerçekleştirilmiş ve doğrusal regresyon, ridge regresyon, rastgele orman, k-en yakın komşu, kategorik artırma (CatBoost), gradyan yükseltme ve hafif gradyan artırma makinesi (LightGBM), fizikle öğrenen yapay sinir ağı (PINN) yöntemi gibi yirmi bir çeşit regresyon modeli test edilmiştir. Modeller ortalama mutlak hata (MAE), ortalama kare hata (MSE), karekök ortalama kare hata (RMSE), belirlilik katsayısı ( $R^2$ ), karekök ortalama logaritmik kare hata (RMSLE) ve ortalama mutlak yüzde hata (MAPE) gibi performans ölçütleri kullanılarak değerlendirilmiştir.

En yüksek doğruluk ve genelleme performansını CatBoost modeli sağlamıştır. Test veri setinde yaklaşık %8,8 MAPE ve 5,66 MAE değerleriyle yüksek tahmin doğruluğu elde edilmiştir. Bu sonuçlar, ML modellerinin pahalı ve zaman alan HAD simülasyonlarına güvenilir bir alternatif olduğunu doğrulamıştır. Modelin uygulanma aşamasında eğitim süreleri, tahmin gecikmeleri ve parametre optimizasyonları incelenmiştir. CatBoost, doğruluk açısından olduğu kadar hızlı çalışma süresi ve düşük hesaplama kaynağı kullanımı açısından da öne çıkmıştır. Modelin hiperparametreleri grid arama (grid search) yöntemiyle optimize edilmiştir.

Tezin önemli bir katkısı, sonlu sayıdaki HAD simülasyonlarından membran performansının öğrenilip geliştirilebileceğinin gösterilmesidir. Bu yöntem, tasarım süreçlerini hızlandırarak membran mühendisliğinde veri odaklı optimizasyonlar için

yeni olanaklar sunmuştur. Ayrıca, fizik bilgili derin öğrenme ağları (PINN) gibi daha karmaşık bir öğrenme mimarisi, regresyon yaklaşımı için kullanılmış olup diğer modeller ile karşılaştırılabilir bir performans elde edilmiştir.

Bu çalışma, akışkanlar mekaniğinde hesaplamalı tasarım alanı için de geniş etkilere sahiptir. Fiziksel simülasyon verilerinin ML modelleriyle entegrasyonu, tasarım süreçlerinin tamamen deneysel veya deneme-yanılma yaklaşımlarından, akıllı ve öngörücü modelleme yaklaşımlarına geçişi sağlamaktadır. Ayrıca, tezde önerilen yöntem geometri ile performans ilişkisinin güçlü olduğu havacılık, mikroakışkanlar, biyomedikal cihaz tasarımları gibi farklı mühendislik alanlarına da uyarlanabilir.

Sonuç olarak bu tez, HAD simülasyonları ve makine öğrenimini birleştirerek membran geometrisi optimizasyonu için sağlam, ölçeklenebilir ve etkili bir metodoloji sunmuştur. Önerilen regresyon tahminleme modeli, performans tahminini hızlandırmış, tasarım döngü sürelerini azaltmış ve membran tasarımında akıllı karar almayı desteklemiştir. Fizik tabanlı modelleme ile veri odaklı çıkarımlar arasında köprü kurarak, yüksek performanslı ve sürdürülebilir desalinasyon sistemlerinin yeni nesline katkı sağlamıştır.

İleriye dönük olarak, araştırma kapsamında daha fazla geliştirme fırsatları bulunmaktadır. Modelin genelleme yeteneğini artırmak için akış kanal geometrileri, membran malzeme çeşitliliği ve çalışma koşulları kapsamı artırılarak veri seti genişletilebilir. Fizik bilgili derin öğrenme ağlarına (PINN) daha fazla fizik bilgisi sağlanarak modelin performansının artabileceği öngörüsü yapılabilir.



# **APPLICATION OF ARTIFICIAL INTELLIGENCE IN SALTWATER DESALINATION**

## **SUMMARY**

Saline water desalination is one of the most critical technologies developed in response to the increasing global water crisis. As conventional freshwater sources become insufficient to meet the growing demands of population growth, urbanization, and climate change, alternative solutions have been investigated. Among them, reverse osmosis (RO) has emerged as the most efficient, scalable, and widely implemented membrane-based desalination technique. However, despite its effectiveness, RO faces several operational challenges, particularly those related to energy consumption and membrane fouling. These challenges not only reduce system efficiency but also increase operational costs and limit the lifespan of membrane modules. Consequently, optimizing of the membranes used in RO systems, particularly internal flow channel geometry, has become essential for enhancing system performance and sustainability.

In recent years, artificial intelligence (AI), particularly machine learning (ML) methods, has emerged as a powerful tool for addressing the high-dimensional and complex problems in membrane technology. Although traditional experimental and physics-based approaches provide a foundation, they are constrained by time, cost, and scalability, and they cannot fully elucidate the intricate relationships among membrane geometry, operating parameters, and performance outputs. The literature increasingly highlights AI's potential to overcome these limitations by processing high-dimensional data, uncovering nonlinear patterns, and guiding the design process. However, most existing studies apply ML either in isolation or to narrowly scoped problems and do not sufficiently integrate the deep insights obtained from physical simulations.

This thesis presents an integrated framework that leverages both computational fluid dynamics (CFD) simulations and supervised ML techniques to explore, evaluate, and predict the performance of innovative membrane channel geometries. The study is rooted in the hypothesis that membrane performance, which is largely affected by geometrical factors such as flow distribution and local turbulence, can be enhanced by designing non-conventional flow channel structures. These structures aim to reduce fouling, lower pressure drop, and increase permeate flux, thereby contributing to more energy-efficient desalination operations.

The first phase of the research focuses on the validation of this hypothesis through preliminary studies. Initially, randomized flow channel geometries were generated using CFD simulations. In these CFD simulations, Navier-Stokes, advection-diffusion, Langmuir adsorption-desorption, permeate flux equations, and Darcy's law were used, and the resulting geometries were compared with existing flat and sinusoidal channel designs in the literature. It was determined that randomly generated new channel

geometries effectively improved membrane performance, maintaining minimal pressure loss while enhancing energy efficiency and water quality.

The second phase of the research focuses on increasing data generation through numerical simulation. A total of 1045 unique membrane flow channel geometries were generated by randomly sampling control points to define the upper and lower surfaces of the flow channel using B-spline curves. These geometries represent a diverse design space, ranging from simple sinusoidal forms to complex, irregular wave patterns. For each geometry, a structured mesh was generated, followed by a 3D CFD simulation to evaluate flow dynamics, concentration profiles, and fouling behavior. Each simulation was run for 100 seconds under constant operating conditions.

Three key performance metrics were extracted from the CFD results: (1) permeate flux, representing the volume of water that passes through the membrane; (2) the fouling index, indicating the accumulation of solute particles on the membrane surface; and (3) the pressure drop between the inlet and outlet of the channel, which reflects the energy required to maintain flow. These metrics were normalized and weighted, and then combined into a single scalar value referred to as the composite performance metric (CPM), which captures the trade-off between energy efficiency and water productivity.

From the 1045 distinct membrane flow channel geometries generated, a final set of 1025 geometries was selected to ensure meshing and numerical stability, thereby constituting the dataset. The channel geometries were represented as numerical vectors comprising the 42 y-coordinates of the control points that define the spline curves. These data were used as inputs to the ML model, while CPM was designated as the regression target.

The second phase of the study employed a supervised regression approach to predict membrane performance directly from geometric features. The dataset was randomly partitioned into two subsets: 900 geometries for training and cross-validation, and 125 geometries reserved as a holdout test set to evaluate model generalization. Cross-validation was performed using a 10-fold strategy to ensure statistical robustness. Twenty one regression models were tested, including linear regression, ridge regression, random forest, k-nearest neighbors, categorical boosting (CatBoost), gradient boosting, and light gradient boosting machine (LightGBM), physics-informed neural networks (PINN). The models were assessed using multiple performance indicators: mean absolute error (MAE), mean squared error (MSE), root mean squared error (RMSE), R-squared ( $R^2$ ), root mean squared logarithmic error (RMSLE), and mean absolute percentage error (MAPE).

Among all tested models, CatBoost demonstrated the highest accuracy and generalization performance. It achieved a MAPE of approximately 8.8% on the test set, indicating that the model was able to predict the composite membrane performance metric with reasonable precision. Additionally, CatBoost showed a low MAE of around 5.66, suggesting minimal deviation from actual simulation results. These outcomes validated the use of machine learning as a reliable surrogate for expensive and time-consuming CFD simulations. The implementation phase included a systematic analysis of model training times, prediction latency, and parameter tuning efficiency. CatBoost not only yielded the best accuracy but also maintained a fast runtime and minimal computational resource consumption. The model's hyperparameters were optimized through grid search.

A principal contribution of this thesis is the demonstration that geometry-driven membrane performance can be learned and generalized from a finite set of CFD simulations. By accelerating design processes, the proposed approach opens new possibilities for data-driven optimization in membrane engineering. Furthermore, a more sophisticated learning architecture, physics-informed neural networks (PINNs), was employed for the regression task, achieving performance comparable to other models.

This study also has broad implications for the field of computational design in fluid mechanics. The integration of physics-based simulation data with ML models enables design processes to transition from purely experimental or trial-and-error approaches to intelligent, predictive modeling paradigms. Moreover, the method proposed in this thesis can be adapted to various engineering domains in which the geometry–performance relationship is strong, including aerospace, microfluidics, and biomedical device design.

In conclusion, this thesis presents a robust, scalable, and effective methodology for membrane geometry optimization by integrating CFD simulations with machine learning. The proposed regression-based predictive model accelerates performance estimation, shortens design cycle times, and supports intelligent decision-making in membrane design. By bridging physics-based modeling with data-driven inference, it contributes to the next generation of high-performance, sustainable desalination systems.

Looking ahead, there are further opportunities for development within the scope of this research. To enhance the model’s generalization capability, the dataset can be expanded by broadening the coverage of flow-channel geometries, membrane material diversity, and operating conditions. It is anticipated that the model’s performance could be improved by providing additional physical knowledge to physics-informed neural networks (PINNs).



## 1. GİRİŞ

Bu tez çalışmasının giriş bölümünde, küresel ölçekte artan su ihtiyacına çözüm olarak öne çıkan su arıtma teknolojileri genel hatlarıyla tanıtılacak; özellikle yeryüzünde yüksek oranda bulunan tuzlu su kaynaklarından tuzun arındırılması için uygulanan desalinasyon teknolojileri ve bu teknolojilerin içerisinde yaygın olarak kullanılan ters ozmoz (RO) yöntemi detaylandırılacaktır. RO sistemlerinin temel bileşenlerinden biri olan membranlar ve bu membranlardaki akış kanal geometrisinin sistem performansı üzerindeki belirleyici rolü vurgulanacaktır. Akış kanal geometrisinin kütle transferi, enerji verimliliği ve membran kirlenmesi üzerindeki etkileri teorik temellere dayandırılarak açıklanacak; bu kapsamda literatürde uygulanan geleneksel ve deneysel yöntemler ile bu alanda uygulanan yapay zeka (AI) yaklaşımları ayrıntılı biçimde ortaya konulacaktır.

### 1.1 Su Arıtma Teknolojileri

Su arıtımına yönelik ilk uygulamalar, M.Ö. 2000’li yıllarda Antik Hindistan ve Mısır uygarlıklarında başlamış; çökeltme, süzme ve kaynatma gibi basit fiziksel yöntemlerle suyun kalitesi artırılmaya çalışılmıştır [1]. Ancak modern anlamda bilimsel ve sistematik su arıtımı süreci, 19. yüzyılda su kaynaklı hastalıkların yaygınlaşmasıyla ivme kazanmış; özellikle 1854’te kolera salgınının içme suyu ile ilişkilendirilmesi, halk sağlığı açısından arıtma teknolojilerinin gelişimini hızlandırmıştır [2]. 20. yüzyılın ortalarında klorlama, hızlı kum filtrasyonu ve koagülasyon gibi kimyasal ve fiziksel prosesler yaygınlaşırken; 1970’li yıllardan itibaren membran teknolojileri, gelişmiş oksidasyon yöntemleri ve çok kademeli sistemler ön plana çıkmıştır [3,4]. Günümüzde ise nüfus artışı, sanayi gelişimi, iklim değişikliği ve kuraklık gibi faktörlerle tatlı suya olan ihtiyaç her geçen gün artmaktadır [5]. Dünya yüzeyinin yaklaşık %71’i suyla kaplı olmasına rağmen, bu suyun %97’si tuzlu sudur ve doğrudan kullanım için uygun değildir [6]. Su arıtımı yalnızca içme suyu üretimi için değil; tarımda verimliliği

artırmak, sanayi süreçlerini güvenli hale getirmek ve çevreyi korumak için de kritik bir gerekliliktir. Küresel ölçekte yaygın olan ancak çoğunlukla kullanıma uygun olmayan bu kaynakların değerlendirilebilir hale getirmek için fiziksel, kimyasal ve biyolojik prensiplere dayanan su arıtma yöntemlerine ihtiyaç duyulmakta; bu yöntemler arıtım sürecinin ön arıtma, birincil, ikincil ve ileri aşamalarında farklı hedeflere yönelik olarak uygulanmaktadır. Literatürde yaygın olarak sınıflandırılan başlıca su arıtma yöntemlerine aşağıda değinilmiştir.

Fiziksel yöntemler, arıtma prosesinin ön arıtma veya birincil aşamasında kullanılmaktadır ve temel olarak suda askıda bulunan katı maddelerin, bulanıklığın, kum, çamur gibi büyük partiküllerin giderilmesini hedeflemektedir. Yaygın fiziksel yöntemler arasında; çöktürme (sedimentation), süzme (filtrasyon), elek ve ızgara sistemleri, flotasyon ve kum filtreleri yer almaktadır. Bu yöntemler, özellikle yüzey sularının ilk arıtımında, biyolojik ve kimyasal arıtma ünitelerine gelen yükü azaltmak amacıyla uygulanmaktadır. Ayrıca ultrafiltrasyon gibi bazı membran temelli fiziksel yöntemler ileri arıtım aşamalarında da kullanılabilir [3,7].

Kimyasal yöntemler, çözülmüş kirleticilerin giderilmesini amaçlayan ve fiziksel arıtmadan sonra devreye giren proseslerdir. Bu grup içinde koagülasyon ve flokülasyon, pH ayarlaması, kimyasal oksidasyon (örneğin ozonlama ve klorlama), kimyasal çöktürme ve iyon değişimi (IX) gibi yöntemler yer almaktadır. Özellikle renk, tat, koku, fosfat ve ağır metal giderimi gibi hedefler için uygundur. Ayrıca dezenfeksiyon amacıyla da çeşitli kimyasallar (örneğin klor, hipoklorit, klor dioksit) kullanılmaktadır. Bu yöntemler, biyolojik sistemlerle kombine edilerek yüksek verimli hibrit arıtım süreçleri oluşturulabilir [8].

Biyolojik yöntemler, başta evsel atık sular olmak üzere organik madde içeriği yüksek olan suların arıtılmasında etkin olarak kullanılmaktadır. Bu yöntemler, mikroorganizmaların biyokimyasal faaliyetlerinden yararlanarak biyolojik oksijen ihtiyacı, kimyasal oksijen ihtiyacı, azot ve fosfor gibi kirleticilerin giderimini sağlamaktadır. Başlıca biyolojik yöntemler arasında aktif çamur prosesi, damlatmalı filtreler (trickling filters), anaerobik çürütücüler ve membran biyoreaktörler (MBR) yer alır. İkincil arıtım aşamasında yaygın olarak kullanılan biyolojik sistemlerin verimi;

sıcaklık, organik yüklenme oranı ve mikrobiyal aktivite gibi parametrelere bağlıdır [7,9].

Klasik fiziksel, kimyasal ve biyolojik yöntemlerle giderilemeyen mikro kirlenmelerin ve patojenlerin arıtımı için gelişmiş arıtım teknikleri kullanılmaktadır. Gelişmiş oksidasyon prosesleri, ozonlama, UV oksidasyon, fotokatalitik sistemler gibi yöntemler bu grupta yer almaktadır. Bu teknolojiler, özellikle ilaç kalıntıları, hormonlar ve pestisit gibi refrakter organiklerin gideriminde yüksek etki göstermektedir. Dezenfeksiyon işlemi son arıtım adımı olarak uygulanarak klorlama, ultraviyole (UV) ışınları veya ozon gibi yöntemlerle patojen mikroorganizmaların inaktivasyonu sağlanmaktadır. Bu işlemler, içme suyu kalitesinin güvence altına alınmasında kritik rol oynamaktadır [10,11].

Membran temelli teknolojiler, çözülmüş ve askıda maddelerin yarı geçirgen zarlar yardımıyla fiziksel olarak ayrılması prensibine dayanmaktadır. Bu yöntemler; mikrofiltrasyon (MF), ultrafiltrasyon (UF), nanofiltrasyon (NF), ters ozmoz (RO), elektrodializ (ED), ters elektrodializ (EDR), membran distilasyonu (MD) ve ileri osmoz (FO) gibi çeşitli teknikleri içermektedir. Membran yöntemleri, hem içme suyu üretimi hem de atık su geri kazanımı amacıyla yaygın olarak kullanılmaktadır. RO, çözülmüş tuzları ve iyonları uzaklaştırmada en yüksek seçiciliğe sahip membran teknolojisidir. Bu yöntemlerin performansı, membran malzemesi, gözenek çapı, uygulanan basınç ve besi suyu karakteristiği gibi faktörlere bağlıdır [5,12].

Desalinasyon, yüksek tuzluluk içeren deniz suyu veya acı su gibi kaynaklardan içilebilir nitelikte tatlı su elde edilmesini amaçlayan özel bir su arıtma teknolojisidir. Literatürde, desalinasyon yöntemleri genel olarak membran temelli ve termal temelli yöntemler olmak üzere iki ana gruba ayrılır. Günümüzde en yaygın kullanılan ve enerji verimliliği açısından öne çıkan teknoloji RO sistemleridir. Termal yöntemler ise genellikle enerji fazlası olan bölgelerde tercih edilmektedir [12,13].

Bu tez çalışması, tuzlu suların yeniden kullanılabilir hale getirilmesini hedefleyen desalinasyon yöntemine odaklanmaktadır. Bir sonraki bölümde bu teknoloji detaylı olarak açıklanmaktadır.

## 1.2 Desalinasyon

Desalinasyon, tarihsel olarak insanlığın suya erişim mücadelesinin bir parçası olarak ortaya çıkmıştır. Yüksek tuzluluk içeren deniz suyu veya acı su gibi kaynakların arıtılarak kullanılabilir tatlı suya dönüştürülmesini amaçlamaktadır. Bu alandaki ilk sistematik uygulamalar 18. yüzyılda, uzun deniz yolculukları sırasında gemi mürettebatının içme suyu ihtiyacını karşılamak amacıyla geliştirilmiştir. 1770'li yıllarda İngiliz donanması tarafından kullanılan basit distilasyon sistemleri, deniz suyunu buharlaştırarak içilebilir forma dönüştüren erken dönem desalinasyon örnekleridir [14]. 20. yüzyılın ortalarına gelindiğinde ise, özellikle su kaynakları sınırlı olan Orta Doğu ve Kuzey Afrika gibi bölgelerde büyük ölçekli termal desalinasyon tesisleri kurulmuştur. Bu dönemde çok kademeli flaş (MSF) ve çok etkili distilasyon (MED) teknolojileri ön plana çıkarken, 1970'li yıllardan itibaren yarı geçirgen membran teknolojilerinin gelişimiyle birlikte ters ozmoz (RO) sistemleri daha yaygın hale gelmiş, düşük enerji tüketimi ve ölçeklenebilir yapıları sayesinde sektörde baskın konuma ulaşmıştır [5,12].

Desalinasyon teknolojileri günümüzde yalnızca içme suyu üretimiyle sınırlı kalmamakta, aynı zamanda farklı sektörlerdeki çeşitli su ihtiyaçlarını karşılamak üzere çok yönlü bir çözüm olarak değerlendirilmektedir. Tarımsal sulamada kullanılabilir kalitede su sağlamak, endüstriyel proseslerde yüksek saflıkta su temin etmek ve enerji santrallerinde kazan besleme suyu hazırlamak gibi farklı amaçlara hizmet etmektedir [15]. Ayrıca, gemilerde, denizaltılarda ve afet bölgelerinde mobil sistemlerle su üretimi gibi askeri ve insani yardım uygulamalarında da kritik rol oynamaktadır [16]. Bu geniş uygulama yelpazesi, desalinasyonun yalnızca bir su arıtma teknolojisi değil, aynı zamanda su güvenliği stratejilerinin temel unsurlarından biri haline geldiğini göstermektedir.

Desalinasyon teknolojilerinin çalışma prensiplerine göre sınıflandırılması, teknolojik çeşitliliğin ve uygulama alanlarının daha iyi anlaşılabilmesi açısından önem taşımaktadır. Desalinasyon yöntemlerinin buharlaşma ve yoğunlaşma teknolojileri, filtrasyon teknolojileri ve kristalizasyon teknolojileri olmak üzere üç ayrı kategoriye ayrıldığı Şekil 1.1'de gösterilmektedir [17].



**Şekil 1.1:** Desalinasyon teknolojilerinin çalışma prensiplerine göre sınıflandırılması [17].

Buharlaştırma ve yoğunlaşma teknolojileri temel olarak tuzlu suya ısı enerjisi uygulayarak buhar oluşturulması ve sonrasında oluşan buharın yoğunlaştırılmasıyla tuzun sudan ayrılması prensibine dayanmaktadır [18,19]. Uygulanan ısı enerjisinin türüne göre kendi içinde de ikiye ayrılmaktadır. Çok işlemli damıtma (MED), çok kademeli flaş damıtma (MSF) ve termal buhar sıkıştırma (TVC) atık ısı veya yakıt yanması yöntemlerini kullanırken; solar damıtma (SSD), güneş bacası (SC) ve nemlendirme ve nem alma (HDH) yöntemleri güneş radyasyonu gibi yenilikçi ısı enerjisi çözümünü içermektedir. Bunların haricinde, bu kategoride mekanik süreçlerin yer aldığı mekanik buhar sıkıştırma (MVC) yöntemi yer almaktadır.

Kristalizasyon teknolojisi ise tuzdan arındırma sürecinde, suyun donma noktasına kadar soğutulması ve ardından saf buz kristallerinin oluşumunun sağlanması prensibine dayanmaktadır. Suda bulunan safsızlıklar ve tuzlar, kristal yapısına giremediğinden buz

kristalleri sadece saf su içermektedir. Bu yöntemde ikincil soğutucu dondurma (SRF), hidratlama (HY) yöntemleri yer almaktadır [20].

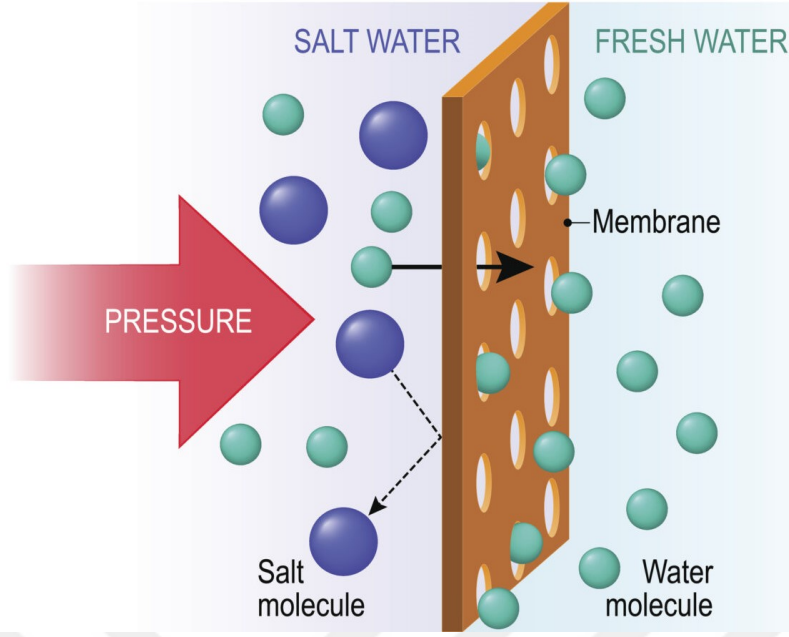
Filtrasyon teknolojilerinin temeli, yarı geçirgen bir membran kullanarak tuzlu suda bulunan farklı moleküllerin boyutlarına ve doğasına göre geçirgenlik davranışlarından faydalanarak filtreleme işleminin yapılması prensibine dayanmaktadır [21]. Filtrasyon teknolojileri kategorisinde membran tabanlı teknolojiler olarak; ters ozmoz (RO), ileri ozmoz (FO), nanofiltrasyon (NF), membrane damıtma (MD) ve kapasitif deiyonizasyon (CDI) yöntemleri yer almaktadır. Bir diğer filtrasyon yöntemi ise iyon değişim reçinesi (IXR) isimli, yarı geçirgen membran yerine çözünmüş iyonları kimyasal olarak yakalayan farklı malzemeler kullanan yöntemdir [22]. Bu yöntem çok düşük tuz konsantrasyonuna sahip su üretmek için tercih edilmektedir. Enerji maliyetlerinin artmasıyla birlikte su arıtım süreçlerinin enerji verimliliğini artırmak ve maliyetleri düşürmek amacıyla yarı geçirgen membranlara dayalı yöntemler yaygınlaşmıştır [23]. Günümüzde, RO, desalinasyon teknolojileri arasında %68.7 ile en yaygın kullanılan yöntemdir; bunu %17.6 ile MSF ve %6.9 ile MED izlemektedir [24].

Sonraki bölümlerde sırasıyla RO teknolojisi ve bu teknolojiye kullanılan membranlar açıklanmaktadır.

### **1.3 Ters Ozmoz (RO) Teknolojisi**

Ozmoz, su moleküllerinin daha yüksek yoğunluktaki ortamdan daha düşük yoğunluktaki ortama yarı geçirgen bir membran aracılığıyla ilerlemesi fenomenine dayanmaktadır. Ters ozmoz (RO) yöntemi ise ozmotik basıncı yenmek için hidrolik basınç uygulanarak tuz ve diğer safsızlıkları bir dizi yarı geçirgen membran kullanarak sudan uzaklaştırma prensibine dayanmaktadır [25]–[27]. RO yöntemi Şekil 1.2’de gösterilmektedir [28]. RO, ilk çıktığı zamanlarda yüksek elektrik tüketimi ve membranların kısa ömürleri nedeniyle sınırlı bir kullanım alanına sahipken, zamanla geliştirilen enerji geri kazanım sistemleri sayesinde daha ekonomik ve verimli bir hale gelmiştir [29].

RO, yüksek toplam çözünmüş katı madde (TDS) konsantrasyonlarını (45.000 mg/L’ye kadar) uzaklaştırabilmektedir ve hem acı hem de deniz suyunun arıtımında



**Şekil 1.2:** Ters ozmoz (RO) yöntemi [28].

kullanılmaktadır [30]. Bununla birlikte, RO'nun yüksek enerji gereksinimi vardır ve gerekli basınç, besleme suyundaki TDS konsantrasyonuna bağlıdır. Acı suyun desalinasyonu için 140–400 psi, deniz suyu için ise 1200 psi'ye kadar pompa basıncı gerekmektedir. Bu nedenle, besleme suyunun TDS konsantrasyonu, enerji tüketimi ve üretilen suyun maliyeti üzerinde büyük bir etkiye sahiptir [31]. RO teknolojisi, 1970'lerden bu yana hızla gelişmiş ve günümüzde desalinasyon pazarında baskın hale gelmiştir. Uluslararası Desalinasyon Derneği'ne (IDA) göre, RO dünya genelinde en hızlı büyüyen desalinasyon teknolojisidir. Bu büyüme, membran teknolojisindeki gelişmeler, daha düşük enerji gereksinimi, enerji geri kazanım cihazlarının (ERD'lerin) yaygınlaşması, farklı kalite ve tuzluluktaki besleme suları için uygunluğu ve modüler yapısı gibi faktörlere dayanmaktadır [32]. Deniz suyu RO (SWRO) desalinasyonu 55–70 bar basınçta %40–55 geri kazanım oranıyla çalışırken, acı su RO (BWRO) desalinasyonu 20–25 bar basınçta %70–85 geri kazanım oranı ile çalışmaktadır. Enerji tüketimi deniz suyu için 2–5 kWh/m<sup>3</sup>, acı su için ise daha düşük olmaktadır [33]. Bu yöntem, ısı veya faz dönüşümüne ihtiyaç duymadığı için termal desalinasyon tekniklerine kıyasla daha enerji verimlidir [34]. Ancak, RO teknolojisinin düşük geçirgen akış, sınırlı seçicilik, kirlenme eğilimi ve yüksek ekipman maliyetleri gibi bazı sınırlamaları bulunmaktadır. Bu zorluklar, prosesin ekonomikliğini etkileyen

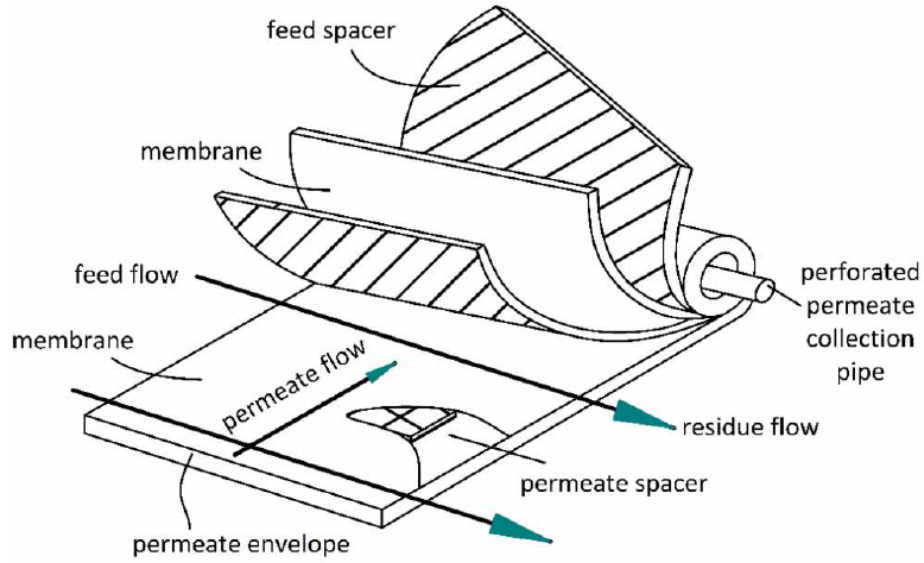
unsurlardır [35]. RO'nun gelişim süreci, membran teknolojisindeki iyileştirmelerle birlikte akış hızını artırma, uzun ömür sağlama, tuz geçişini azaltma ve sıkıştırmaya karşı dayanıklılığı artırma gibi alanlarda ilerlemiştir [36]. RO'nun en önemli sınırlamalarından biri, membranlarda kirlenme (fouling) sorunudur. Kirlenme, mikroorganizmaların kolonileşmesi ve istenmeyen maddelerin membran yüzeyinde veya gözeneklerinde birikmesiyle meydana gelmektedir. Bu durum, su akışını ve tuz tutma performansını düşürerek membranın etkinliğini azaltmaktadır [37,38].

#### **1.4 Membran**

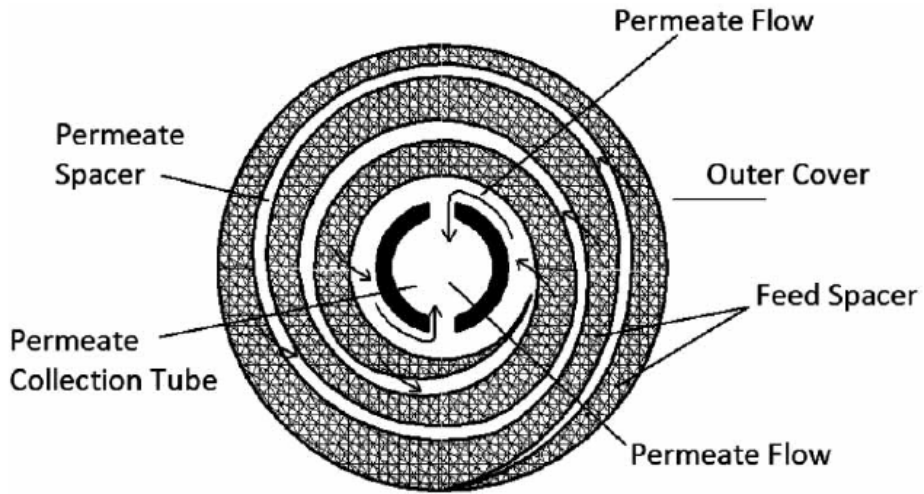
Membran modülleri, membran sistemlerinin sıvıları işleyiş şeklini ve tasarımını tanımlamaktadır. Her membran modül tipi, farklı proseslere, sıvı akışına ve uygulama gereksinimlerine göre tasarlanmaktadır. Spiral sarılmış membranlar (SWM) kompakt bir modül yapısı sunarken, tübüler membranlar daha büyük olup endüstriyel uygulamalar için uygundur. Daldırma membranları, atık su arıtma gibi düşük maliyetli çözümler sunmaktadır. Plaka ve çerçeve membranları, esnek ve bakım dostudur. Hollow fiber membranlar, kompakt yapısı ile yüksek verimlilik sunabilmektedir. Çift sıralı membranlar, birden fazla arıtma aşaması gerektiren süreçlerde kullanılmaktadır ve yüksek verimlilik sunmaktadır. Membran türlerinin verimlilik, maliyet ve uygulama açısından uygunluğu; proses koşulları ve kullanım amacına göre değişiklik göstermektedir. Aşağıda RO sistemlerinde en yaygın kullanılan membran modeli SWM detaylarıyla açıklanmaktadır.

SWM, ince membranlar ve taşıyıcı bir şemsiye plakasının bir araya geldiği bir tasarımdan oluşmaktadır [39]. Membranlar, birbirine sarılarak kompakt bir yapı oluşturmakta ve sıvı spiralin iç kısmında dolaşmaktadır. Ters ozmoz (RO), nanofiltrasyon (NF), ultrafiltrasyon (UF) ve içme suyu arıtma sistemlerinde yaygın olarak kullanılmaktadır. Kompakt yapısı nedeniyle daha az alan gerektirdiği için yüksek alan verimliliği sağlamaktadır. Yüksek verimlilik ve enerji tasarrufu sağlayan bu membran modülünde tıkanma riski yüksek olup spiral yapısı sebebiyle temizliği zor olmaktadır. SWM yapısı ve filtrasyon işlemi Şekil 1.3'te, kesit alanı ise Şekil 1.4'te gösterilmektedir [39]. SWM, temel olarak temiz suyun geçtiği permeat tüpü, permeat

ara parçası, akış kanalı aralayıcı ve zarf şeklinde sarılmış membran yaprakları gibi parçalardan oluşmaktadır.



Şekil 1.3: Spiral sarılmış membran modülü ve filtrasyon işlemi [39].



Şekil 1.4: Spiral sarılmış membran modülünün kesit alanı [39].

Bu modüllerde bulunan akış kanalı (feed spacer) yapısı, Şekil 1.3'te gösterildiği gibi iki membran zarfı arasında yer alarak besleme suyunun membran yüzeylerine paralel akışını sağlamaktadır. Bu yapı, sadece membranların birbirine yapışmasını önlemekle kalmamakta; aynı zamanda kanal içerisinde türbülans yaratarak kütle transferini artırmakta ve konsantrasyon polarizasyonunu (CP) azaltmaktadır [40]–[42]. Bu yapılar, akış koşullarını düzenleyerek sistem performansını iyileştirmektedir. Bu nedenle,

membran yüzeyinde partikül ve biyolojik birikimin (foouling) önlenmesi açısından kritik öneme sahiptir [42]–[44].

SWM modüllerinde akış kanalının hidrodinamik özellikleri, kanal içindeki türbülans seviyesini ve çapraz akışı etkileyerek sistem verimliliğini doğrudan belirlemektedir. Bu yapı sayesinde, besleme suyunun tüm membran yüzeyine eşit yayılması, basınç düşüşünün optimize edilmesi ve membran kirlenmesi (fouling) riskinin azaltılması sağlanmaktadır [45]–[47]. Ancak bu avantajların sağlanabilmesi, akış kanal geometrisinin dikkatle tasarlanmasına bağlıdır. Geometrik parametreler (örneğin; kanal geometrisinin kalınlığı, filament aralığı, dizilimi) hem enerji tüketimi hem de membran ömrü üzerinde belirleyici rol oynamaktadır [48]–[51].

### **1.5 Tezin Amacı**

Bu tez, ters ozmoz (RO) sistemlerinde kullanılan membran akış kanal geometrilerinin optimizasyonuna ve performans tahminlemesine yönelik hibrit bir çalışmayı kapsamaktadır. RO sistemlerinde karşılaşılan en önemli sorunlardan biri olan membran kirlenmesi, enerji tüketimini artırmakta ve su arıtma verimliliğini olumsuz etkilemektedir. Bu bağlamda, öncelikle, hem enerji verimliliğini yükseltmek hem de membran ömrünü uzatmak amacıyla akış kanal geometrisinin sistem performansına etkileri detaylı olarak incelenmiştir. Daha sonra, klasik deneysel ve hesaplamalı yaklaşımların ötesine geçerek yapay zeka (AI) ile sistem performansı tahminlemesi için bir model geliştirilmiştir. Hesaplamalı akışkanlar dinamiği (HAD) simülasyonları ile elde edilen veri seti kullanılarak geliştirilen AI modeli sayesinde, yüksek performans potansiyeline sahip kanal geometrilerinin daha kısa sürede ve daha düşük maliyetle öngörülmesi hedeflenmiştir. Bu doğrultuda, HAD ve AI yöntemlerinin entegrasyonu ile RO sistemlerinde enerji tüketimi, membran kirlenmesi ve su kalitesi arasında optimum bir denge sağlayacak yeni tasarım yaklaşımları ortaya konmuştur. Tez kapsamında, membran akış kanal tasarımının akış performansına etkisi, HAD destekli AI yaklaşımıyla modellenmiştir.

Bir sonraki bölümde, literatüre geçmiş olan, membran performansını artırmaya yönelik yapılan deneysel çalışmalar ve bu alanda AI kullanan uygulamalar açıklanmaktadır.

## 1.6 Literatürdeki Uygulamalar

Literatürde, membran performansını artırmak ve sistemin daha geniş bir uygulama alanına hitap etmesini sağlamak için çeşitli teknik ve yapısal iyileştirmeler önerilmiştir. Bunların başında, membran kimyasının geliştirilmesi gelmektedir. Selüloz asetat bazlı membranların yerini alan poliamid bazlı kompozit membranlar, hem su geçirgenliğini artırmış hem de tuz geri kazanımını iyileştirmiştir. Literatürde, bu tür membranların aynı çapta (8") olan eski membranlara göre iki kat daha yüksek akı ve üç kat daha düşük tuz geçişi sağladığı bildirilmiştir [52]–[54]. Bu boyuttaki spiral sarımlı RO modülleri, daha yüksek membran yüzey alanı sunarak (genellikle 37–40 m<sup>2</sup>) daha fazla su akısı sağlamıştır. Dolayısıyla, büyük ölçekli tesislerde daha az modül ile daha fazla üretim yapılmasına imkan tanıyarak endüstriyel uygulamalarda standart haline gelmiştir.

SWM modül tasarımında yapılan değişiklikler, performans üzerinde önemli bir etki yaratmıştır. Örneğin, zarfların sayısı ve genişliği optimize edilerek modül içindeki aktif membran alanı maksimize edilmiştir. Yapıştırıcı çizgisi (glue-line) genişliğinin dikkate alınmasının zarfların optimal sayısını değiştirdiği gösterilmiştir [43]. Benzer şekilde, daha dar zarf genişlikleri kullanıldığında zarfların sayısı artmıştır ve transmembran basınç (TMP) dağılımı daha homojen olmuştur [55].

Akış kanalı tasarımı özelinde, akış kanal aralayıcının kalınlığının artırılmasıyla membran kirlenmesi ve basınç kayıplarında düşüş sağlanmıştır. Örneğin, 0.7 mm kalınlığındaki akış kanal aralayıcıya kıyasla 0.86 mm kalınlığındaki bir aralayıcı, daha düşük kirlenme eğilimi ve temizlik ihtiyacı sunmasının yanında sistemdeki basınç düşüşünü azaltmıştır [56]. Ancak daha kalın akış kanal geometrileri, membran başına düşen yüzey alanını da azalttığı için bu kazanım bir üretim kapasitesi kaybıyla birlikte değerlendirilmelidir.

Akış kanal aralayıcıların geometrileri de literatürde çalışılmış bir diğer konudur ve bu tezin odağını oluşturduğu için tezin teknik kısmında detayları ile ele alınacaktır. Literatürde, elmas şekilli (rhombus), zigzag, merdiven tipi (ladder) ve çok katmanlı (multi-layer) yapılar karşılaştırmalı olarak analiz edilmiştir. Özellikle üç katmanlı akış kanal aralayıcıların daha fazla akış instabilitesi yarattığı ve dolayısıyla kütle transferini

%20'ye kadar artırabildiği görülmüştür [57,58]. Ancak bu yapılar aynı zamanda daha fazla enerji tüketimi gerektirmiştir.

Kanal aralayıcıların üretim malzemeleri, performans açısından önemli bir diğer unsurdur. Antimikrobiyal metallerle (örneğin Cu, Ag) kaplama yapılması, yüzey modifikasyonu yoluyla biyofilm oluşumunu engellemek için uygulanmıştır. Ancak bu kaplamaların etkinliği zamanla azalmış ve mikroorganizmaların oluşturduğu hücre dışı polimerik maddeleri kaplamanın etkisini sınırlamıştır [59]–[61]. Ayrıca bu tür kaplamaların membran ömrü ve su kalitesi üzerindeki uzun vadeli etkileri yeterince net değildir.

Sayısal modelleme (hesaplamalı akışkanlar dinamiği, HAD) ve deneysel görselleştirme teknikleri ile yapılan çalışmalar, akış kanal geometrisinin Reynolds sayısı (Re), Sherwood sayısı ve basınç kaybı üzerindeki etkilerini ortaya koymuştur [46,62,63]. Özellikle parçacık görüntü velosimetrisi (PIV) ve nükleer manyetik rezonans görüntüleme (NMRI) gibi yöntemlerle, aralayıcıların oluşturduğu türbülans yapıları, ölü bölgeler ve kirlenme dağılımları gözlemlenmiştir [64]–[66].

Filament kesiti (yuvarlak, üçgen, dikdörtgen), torsiyon (bükülme), filament yönelimi ( $45^\circ$  veya  $90^\circ$ ), katman sayısı, kanal içi konum (submerged spacer), gibi faktörlerin tümü sistem performansında farklı etkiler yaratmıştır [47,58,67]–[69]. Örneğin,  $45^\circ$  lik akış saldırı açısına sahip aralayıcılar daha yüksek türbülans üretse de daha fazla basınç kaybına neden olmuştur; buna karşın  $90^\circ$  lik yapılar daha dengeli performans sunabilmiştir [49,70].

Bu bağlamda, akış kanal aralayıcı tasarımı konusunda hedeflenen optimal yapı; minimum basınç kaybıyla maksimum kütle transferini sağlayan, kirlenme eğilimi düşük ve üretim kapasitesi yüksek bir geometridir [48,51]. Literatürde, bu hedef doğrultusunda yapılan çok sayıda çalışma, SWM modüllerinin daha ekonomik, verimli ve sürdürülebilir hale gelmesine katkı sağlamıştır.

Son yıllarda, yapay zeka (AI) ve özellikle makine öğrenmesi (ML) tabanlı modelleme ve optimizasyon yaklaşımları, hem membran malzeme tasarımında hem de sistem performansının öngörülmesinde giderek daha fazla araştırma konusu haline gelmiştir. Bu gelişmeler, geleneksel deneysel yaklaşımların sınırlarını aşmak, yüksek boyutlu

verileri işlemek ve geniş malzeme tasarım uzayında anlamlı desenler keşfetmek için kullanılmıştır [71]. Literatürdeki güncel çalışmalar, ML yöntemlerinin membran geliştirme yaşam döngüsünün her aşamasında (malzeme seçimi, üretim koşullarının belirlenmesi, proses tasarımı ve bilgi çıkarımı gibi alanlarda) etkin biçimde uygulanabildiğini göstermiştir [72].

ML uygulamaları, genellikle denetimli öğrenme (supervised learning), denetimsiz öğrenme (unsupervised learning) ve aktif öğrenme çerçevesinde ele alınmaktadır. Denetimli öğrenme kapsamında; regresyon ve sınıflandırma problemleri için sıklıkla karar ağaçları, destek vektör makineleri (SVM), Gaussian süreçleri ve yapay sinir ağları (ANN/MLP) kullanılmıştır [73]–[75]. Denetimsiz öğrenme kapsamında, kümeleme (clustering) ve boyut indirgeme (dimension reduction) teknikleri yaygın olarak kullanılmaktadır. Aktif öğrenme (active learning), sınırlı etiketli veri ile çalışmayı mümkün kılan ve özellikle deneysel verinin az bulunduğu senaryolar için umut vadeden bir yöntemdir. Henüz membran teknolojileri alanında doğrudan uygulanmış olmasa da, literatürdeki eğilimler aktif öğrenmenin yakın gelecekte membran tasarımına entegre edileceğini göstermektedir [76].

Literatürde, ince film kompozit (TFC) RO membranları için geliştirilen ML modelleri dikkat çekicidir. Gao ve arkadaşları [77], organik çözücü nanofiltrasyon (NF) membranları dahil olmak üzere, RO sistemlerinin seçicilik ve akı değerleri üzerinde etkili olan üretim parametrelerini (örneğin monomer türü, polimerleşme süresi, çözücü tipi) analiz etmek amacıyla aşırı gradyan artırma (XGBoost) ve kategorik artırma (CatBoost) gibi gelişmiş karar ağacı algoritmalarını kullanmıştır. İlgili çalışmada, üretim verilerindeki eksiklikler klasik istatistiksel yöntemlerle doldurulmak yerine, eksik verileri doğrudan içeren bir modelleme yaklaşımı benimsenmiş ve böylece veri kaybı minimize edilmiştir. Modelin çıktıları, Bayesyen optimizasyon algoritmalarıyla birlikte kullanılarak optimum üretim koşulları belirlenmiş ve önerilen koşullar altında üretilen membranların deneysel test sonuçları, modellerin doğruluğunu teyit etmiştir.

RO sistemleriyle ilgili bir diğer dikkat çekici uygulama ise çok katmanlı (layer-by-layer, LbL) membran üretim tekniklerinin ML ile modellenmesidir. Wessling ve ekibi [78], LbL yöntemiyle üretilmiş NF ve RO membranlarının iyon tutma ve su geçirgenlik

özelliklerini tahmin etmek amacıyla hibrit bir model geliştirmiştir. Bu hibrit yapı, yapay sinir ağı (ANN) ile belirlenen membran parametrelerini (örneğin gözenek çapı ve yüzey yük yoğunluğu) klasik taşınım modelleriyle birleştirmiştir. Modelin başarımı, k-katlı çapraz doğrulama yöntemi ile değerlendirilmiş ve üç girişli (örneğin birleştirilmiş kaplama süresi ve akısı) basitleştirilmiş modelin dahi yüksek doğrulukla çalıştığı gösterilmiştir.

Ayrıca, RO sistemleri için membran tasarımı aşamasında da ML uygulamalarına yer verilmiştir. Bir araştırma grubunun RO performansı üzerinde çeşitli proses parametrelerinin (basınç, geri kazanım oranı, besleme bileşimi vb.) etkilerini değerlendirdiği çalışmada, XGBoost, rastgele orman (RF) ve ANN modelleri, çok değişkenli doğrusal regresyon ile karşılaştırılmıştır. Sonuçlar, ANN modelinin özellikle basınç tahmini konusunda en başarılı yaklaşım olduğunu, RF modelinin ise tuz geçiş oranı öngörüsünde daha iyi performans sergilediğini ortaya koymuştur [79].

Membran teknolojilerinde karşılaşılan başlıca operasyonel zorluklardan biri, membran yüzeylerinde gerçekleşen kirlenme (fouling) olgusudur. Bu durum, özellikle RO ve NF sistemlerinde, performans kaybı, enerji maliyetlerinde artış ve membran ömrünün kısılması gibi sonuçlara yol açmaktadır. Kirlenme, membran yüzeyinde veya gözeneklerinde oluşan adsorpsiyon, birikim ya da blokaj mekanizmalarıyla gerçekleşmekte ve hem fiziksel hem kimyasal etkileşimlere bağlı olarak gelişmektedir [80]. Bu süreçlerin modellenmesi, klasik yöntemlerle sınırlı kalmakta ve sistemin çok boyutlu doğası nedeniyle yeterince bütüncül öngörüler sunulamamaktadır. Bu bağlamda, ML tabanlı modeller, membran kirlenme eğilimlerinin tahmini ve kontrolü için yeni bir yol sunmaktadır.

Literatürde, membran kirlenmesi ile ilgili olarak kullanılan ML yöntemleri arasında özellikle destek vektör makineleri (SVM) ve kümeleme algoritmaları öne çıkmaktadır. Örneğin, SVM algoritması, membran sistemlerinde yağ-su ayırımına yönelik uygulamalarda permeat akısındaki düşüşlerin modellenmesi ve olası kirlenme mekanizmalarının sınıflandırılması amacıyla kullanılmıştır. Bu model, sınıflandırma görevlerinde yüksek doğruluk sağlamış ve özellikle küçük veri setleriyle çalışabilme avantajı sayesinde sahada uygulanabilir sonuçlar üretmiştir [81]. Aynı şekilde, destek

vektör regresyonu (SVR) modelleri de membran biyoreaktörlerinde performans tahmini ve kirlenme davranışlarının kantitatif analizinde kullanılmıştır [75]. Bu modeller, hem kısa vadeli performans değişimlerinin öngörülmesini hem de sistem davranışlarının daha iyi anlaşılmasını sağlamıştır.

Literatürdeki bir diğer çalışmada, AI yöntemlerinden biri olan denetimsiz öğrenme algoritmalarının membran kirlenme tahmini için kullanılmasıdır. Özellikle k-ortalama ve hiyerarşik kümeleme gibi yöntemlerle, farklı çalışma koşullarındaki membranların kirlenme eğilimleri gruplanmış, böylece belirli operasyonel pencereler içerisinde kirlenmeyi azaltacak çalışma aralıkları belirlenmiştir. Bu uygulamalar sayesinde, sistem tasarımı sırasında hangi su kaynağının daha az kirlenme eğilimi gösterdiği tespit edilebilmiş ve kirlenme riskine göre ön işlem stratejileri geliştirilebilmiştir [82].

Membran performansındaki bozulmaların (örneğin plastikleşme, yaşlanma, kimyasal bozulma gibi) önceden tahmin edilerek önlem alınması, sistem performansı ve işletme maliyetleri için önemli konulardan biridir. Bu amaçla, Gaussian süreci durum uzayı modelleri önerilmiştir. Özellikle proton değişimli membran yakıt hücrelerinde (PEMFC), membran bozulma mekanizmalarının tam olarak anlaşılmadığı durumlarda, Gaussian süreçleriyle oluşturulan zaman serisi modelleri kullanılmıştır. Bu modeller, özellikle zamanla meydana gelen performans kayıplarının nedenlerini ayırt etme ve bakım planlaması yapma açısından değerli bilgiler sunmuştur [83].

AI destekli performans iyileştirme çalışmalarının bir diğerinde, Masoud ve arkadaşlarının ANN tabanlı modellemesiyle, ince film nanokompozit (TFN) membranlarında permeat akısı ve kirlenme uzaklaştırma performansları tahmin edilmiştir. Yapılan duyarlılık analizleri, bazı üretim parametrelerinin (örneğin temas açısı ve son işlem sıcaklığı) kirlenme üzerinde doğrudan etkili olduğunu ortaya koymuştur [84]. Bu sonuçlar, yalnızca tahmin değil, aynı zamanda kirlenmeyi azaltacak stratejik üretim koşullarının da ML modelleriyle belirlenebileceğini göstermiştir.

Akış kanalı geometrisi gibi yapısal tasarım unsurlarına yönelik doğrudan bir ML uygulamasına literatürde henüz yer verilmemiştir. Ancak, genel anlatımda bu tür çok boyutlu ve fizik temelli tasarım problemlerinde AI'nın kullanılmasının önemli

potansiyeye sahip olduđu vurgulanmaktadır. Bu tez çalışmasında, hesaplamalı akışkanlar dinamiği (HAD) verisiyle desteklenen ve akış kanal tasarımı gibi karmaşık geometrilerin performansa etkisini öngörebilecek ML modelleri geliştirilmiştir.



## 2. HESAPLAMALI AKIŞKANLAR DİNAMİĞİ VE KANAL TASARIMI

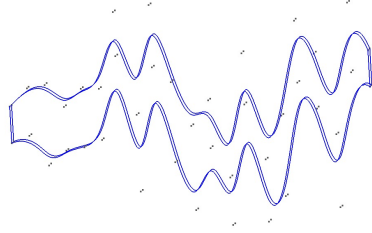
Ters ozmoz (RO) sistemlerinde kullanılan membranların akış kanalı geometrisi; akış düzeni, membran kirlenmesi ve konsantrasyon polarizasyonu üzerinde etkilidir. Bu tez kapsamında, öncelikle HAD modeli ve simülasyonları kullanılarak rastgele geometriye sahip farklı kanal geometrileri oluşturulmuş olup düz ve sinüzoidal geometrilerdeki akış kanalları ile karşılaştırılmıştır. Yapılan ön çalışmada, öncelikle HAD simülasyonlarında aşağıdaki adımlar izlenmiştir, sonra kanal geometrisi tasarımı yapılarak çözüm ağı oluşturulmuştur. Son adımda ise HAD simülasyonu gerçekleştirilerek performans metrikleri elde edilmiştir.

- CAD modelleri kullanılarak akış kanalının geometrisi oluşturulmuştur.
- Bu modeller, numerik işleme için çözüm ağı oluşturulması (meshleme işlemi) ile hücrelere ayrıştırılmıştır.
- Akışkanın akış kanalına giriş hızı, basınç, sıcaklık ve çözünür konsantrasyonu gibi sınır koşullarına ait parametreler belirlenmiştir.
- Kanal geometrileri için oluşturulan çözüm ağları OpenFOAM (v2412) simülasyon programında belirlenen sınır koşulları kullanılarak çözdürülmüştür.
- Rastgele oluşturulmuş farklı kanal geometrileri için ayrı ayrı simülasyonlar tekrarlanmıştır.
- Sonucunda, membran performansını belirleyen tortu birikimi (yüzey kirlenmesi), sıvı akısı parametreleri elde edilmiştir.

### 2.1 Kanal Geometrisi Tasarımı ve Çözüm Ağının Oluşturulması

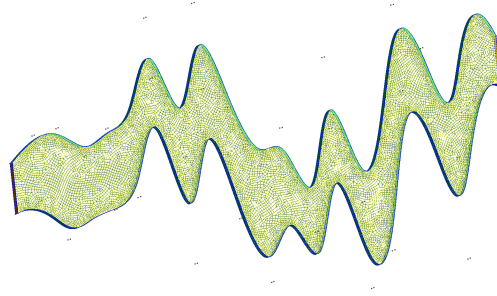
Çalışmada kullanılan kanal geometrilerinin tasarımı açık kaynaklı SALOME platform (versiyon 9.13) modelleme yazılımı kullanılarak yapılmıştır. Kanal geometrisi boyutları, kanal uzunluğu 100 mm, kanal yüksekliği 1 mm, giriş ve çıkış kanal genişliği ise 10

mm olacak şekilde belirlenmiştir. Membran kalınlığı  $150\mu\text{m}$  olarak tanımlanmıştır. Çalışmada simüle edilmek üzere oluşturulan rastgele kanal geometrilerinin tasarımı için Bspline yöntemi uygulanmıştır. Kanalın akış yönünde sağ ve sol duvarlarını oluşturacak şekilde rastgele (x,y) noktaları farklı sayılarda seçilmiştir. Seçilen nokta sayıları, Bölüm 2.5'te ön çalışma-1 kapsamında detayları ile açıklanacaktır. Bu noktalar Bspline'lara tanımlanarak rastgele nokta içeren kanal geometrileri oluşturulmuştur (bkz. Şekil 2.1). Böylece, kanal geometrisinin akış dinamiği, dolayısıyla membran kirlenmesi ve permeat akışı üzerindeki etkilerini incelemek için rastgele akış kanal tasarımları elde edilmiştir.



**Şekil 2.1:** Rastgele seçilen noktalar ve B-spline yöntemi kullanılarak elde edilen rastgele geometri tasarımı örneği.

Daha sonra, tasarlanmış kanal geometrileri için çözüm ağları, en büyük eleman uzunluğu  $25 \times 10^{-5}$  olacak şekilde oluşturulmuştur (bkz. Şekil 2.2). Çözüm ağı, hem tuzlu su akış profili hem de kanal tabanında yer alan membranın zamana bağlı kirlenmesini simüle edecek şekilde üç boyutlu olarak tasarlanmıştır. Çözüm ağının oluşturulması için aynı şekilde SALOME platform (versiyon 9.13) modelleme yazılımı kullanılmıştır.



**Şekil 2.2:** Rastgele hazırlanmış kanal geometrileri için oluşturulan çözüm ağı örneği.

## 2.2 Hesaplamalı Akışkanlar Dinamiği (HAD) Modeli

Akış kanal geometrisinin RO sistemlerindeki etkisini incelemek için yaygın olarak kullanılan HAD simülasyon yöntemi, akışkanların hareketini ve bu hareket sırasında ortaya çıkan fiziksel etkileri matematiksel modeller kullanarak incelemektedir. Bu çalışmada, su akış profili, membran kirlenmesi ve permeat akışını hesaplamaya yönelik zamana bağlı simülasyon gerçekleştirilmiştir. İlk aşamada, hız ve basınç profilleri Navier-Stokes denklemleri kullanılarak elde edilmiştir [85,86]:

$$\frac{\partial \mathbf{U}}{\partial t} + \frac{1}{\rho} \nabla P + (\mathbf{U} \cdot \nabla) \mathbf{U} = \nabla \cdot (\nu \nabla \mathbf{U}) \quad (2.1)$$

Bu denklemde,  $\mathbf{U}$  (m/s) üç boyutlu hız vektörünü,  $P$  ( $\text{kg m}^{-1} \text{s}^{-2}$ ) basıncı,  $\rho$  ( $\text{kg m}^{-3}$ ) çözelti yoğunluğunu ve  $\nu$  ( $\text{m}^2 \text{s}^{-1}$ ) kinematik viskozite değerini göstermektedir. Bu çalışmada yer çekimi ivmesi ihmal edilmiştir.

Adveksiyon difüzyon denklemi, akış sistemlerinde kütle transferini analiz etmek için kullanılmaktadır. Böylece, tuz konsantrasyonunun akış kanal geometrisi ve akış hızına bağlı olarak dağılımı incelenebilmektedir. Bu çalışmada, simüle edilen tuzlu su içerisinde tuzlu su konsantrasyon profilini elde etmek için adveksiyon difüzyon denklemi kullanılmıştır [85,86]:

$$\frac{\partial C_b}{\partial t} = \nabla \cdot [D \nabla C_b] - \nabla \cdot (\mathbf{U} C_b) \quad (2.2)$$

Bu denklem, tuz konsantrasyonunun zamana bağlı değişimini ifade etmektedir. Denklemde sağ tarafında bulunan birinci terim difüzyon etkisini, ikinci terim ise hız etkisi ile tuz taşınımını göstermektedir. Bu denklemde  $C_b$ , çözelti içerisindeki tuz kütle oranını,  $D$  [ $1.47 \times 10^{-9} \text{ m}^2 \text{ s}^{-1}$ ] ise moleküler difüzyon katsayısını ifade etmektedir.

Membran kirlenmesini hesaplamak üzere membran yüzeyinde Langmuir adsorpsiyon-desorpsiyon denklemi tanımlanmıştır [85,86]:

$$\frac{\partial C_s}{\partial t} = k_1 \left( 1 - \frac{C_s}{C_e} \right) C_b - \frac{k_2 C_s}{C_e} - R_s \quad (2.3)$$

Bu denklemde,  $C_s$  membran yüzeyindeki tuz konsantrasyonunu,  $k_1$  (1/s) adsorpsiyon katsayısını,  $C_e$  membran yüzeyine adsorbe olabilecek azami tuz konsantrasyonunu,  $k_2$  (1/s) desorpsiyon katsayısını,  $R_s$  ise kayma gerilmesine bağlı desorpsiyonu ifade etmektedir. Bu denklemde,  $k_1$  ve  $k_2$  katsayıları deneysel olarak elde edilmektedir.

Bu denklemin hem organik hem de membran yüzeyinde kristal oluşumunu inceleyen modellerde kullanılması mümkündür.  $k_1/k_2$  oranı süreç ve ortam parametrelerine bağlı olarak 0.1 ile  $1 \times 10^{-6}$  arasında değişiklik göstermektedir [85].

Kayma gerilmesine bağlı desorpsiyon ( $R_s$ ) ise laminar akış ortamında şu denklemle hesaplanmaktadır [85]:

$$R_s = \frac{\partial U}{\partial y} \cdot \nu \cdot \rho \cdot \alpha \quad (2.4)$$

Burada,  $\partial U/\partial y$  membran yüzeyine normal hız gradyanını,  $\nu$  kinematik viskoziteyi,  $\rho$  tuzlu su yoğunluğunu,  $\alpha$  ise kayma gerilmesine bağlı desorpsiyon katsayısını ifade etmektedir.

Bu çalışmada  $k_1$ ,  $k_2$  ve  $\alpha$  değerleri, [86]'da verilen deneysel çalışmalardaki kirlenme ve permeat akışı verilerine uygun olacak şekilde sırasıyla 0.1, 0.01 ve  $6.67 \times 10^{-3}$  olarak belirlenmiştir.

Test edilen farklı kanal geometrilerinde membran kirlenmesinin permeat akışı üzerine etkisini analiz etmek için aşağıda verilen denklem kullanılmıştır [87]:

$$J_w = k_w \cdot \left(1 - \frac{C_s}{C_e}\right) \cdot (P - A_p C_b) \quad (2.5)$$

$J_w$  (Lmh) permeat akışı hesaplama denkleminde  $k_w$  (Lmh/bar) membran geçirgenliğini,  $P$  (bar) membran yüzey basıncını,  $A_p$  (bar) ise tuz konsantrasyonuna bağlı ozmotik basınç katsayısını ifade etmektedir. Membran geçirgenliği  $k_w$ , Darcy yasası kullanılarak şu şekilde hesaplanmıştır [87]:

$$k_w = \frac{k}{\mu \cdot \Delta x} \quad (2.6)$$

Burada,  $k$  ( $m^2$ ) membran özgül geçirgenliğini ifade etmektedir. Bu değer, tuz ve organik arıtımında kullanılan bazı endüstriyel membranların geçirgenlik değeri olan 100 Lmh/bar değerine denk gelecek şekilde  $4.167 \times 10^{-17} m^2$  olarak alınmıştır. Membran kalınlığı ise 150  $\mu m$  olarak tanımlanmıştır. Denklem 2.6'daki  $\mu$  ise dinamik viskozite değerini ifade etmektedir.

### 2.3 HAD Simülasyonu

HAD simülasyonunu gerçekleştirmek üzere openFOAM (versiyon 2412) yazılımı kullanılmıştır. Membran akış simülasyonu; zamana bağlı, sıkıştırılmaz, türbülant

ve laminer akış koşullarını simüle edebilen pimpleFOAM çözücüsü modifiye edilerek gerçekleştirilmiştir. PimpleFOAM çözücüsü içerisine konsantrasyon  $C_b$ , yüzey konsantrasyonu  $C_s$  ve permeat akışı  $J_w$  denklemleri skaler büyüklükler olarak eklenerek hız ve basınç profiline bağlı olarak bütün denklemlerin beraber çözülmesi sağlanmıştır. Bütün profillere  $1 \times 10^{-7}$  (residual value) hassasiyet değeri atanmıştır. HAD sürecinde kullandığımız sınır koşulları, Çizelge 2.1’de verilmiştir.

**Çizelge 2.1:** HAD simülasyonunda kullanılan sınır koşulları.

	Giriş	Çıkış	Duvarlar	Membran
U	0.15 m/s	$\frac{dU}{dt} = 0$	0	kısmi kayma
P	$\frac{dP}{dt} = 0$	40 bar	$\frac{dP}{dt} = 0$	$\frac{dP}{dt} = 0$
$C_s$	0	0	0	$\frac{dC_s}{dt} = 0$
$C_b$	1	$\frac{dC_b}{dt} = 0$	$\frac{dC_b}{dt} = 0$	$\frac{dC_b}{dt} = 0$
$J_w$	0	0	0	$\frac{dJ_w}{dt} = 0$

Porlu yüzeyler üzerindeki laminar akış hızı Darcy yasasına göre denklem 2.7’de gösterildiği şekilde hesaplanmaktadır [87].

$$U = -\frac{k}{\mu} \cdot \frac{\partial p}{\partial x} \quad (2.7)$$

Burada, membran yüzeyi üzerindeki hız için  $k/\mu$  kısmi kayma değeri verilerek membran yüzeyindeki hız profili  $U$  elde edilmiştir.  $C_b$  değeri, girişte 0.035 kg/kg tuz konsantrasyonu 1 değerine denk gelecek şekilde atanmış ve  $C_e$  ile  $A_p$  değerleri buna göre hesaplanmıştır. Simülasyon esnasında membranın tuz rejeksiyonu %100 olarak varsayılmıştır.

#### 2.4 Akış Kanal Geometrisi için Performans Metrikleri

Bu çalışmada, RO membran sistemlerinde farklı kanal geometrilerinin karşılaştırmalı olarak değerlendirilmesini mümkün kılmak için performans metrikleri tanımlanmıştır. Bir sonraki bölümde anlatılacak olan ön çalışmalarda, performans metriği olarak yüzey kirlenmesi ( $C_s$ ) ile ters orantılı, permeat akışı ( $J_w$ ) ile düz orantılı bir katsayı (denklem 2.8) kullanılmıştır.

$$Performans = \frac{J_w}{C_s} \quad (2.8)$$

Çalışmanın ikinci fazında ise birden fazla fiziksel çıktıyı içeren birleşik bir performans metriği (denklem 2.9) tanımlanmıştır. Hesaplamalı akışkanlar dinamiği (HAD) simülasyonları sonucunda elde edilen üç temel çıktı aşağıda verilmiştir:

- $J_w$ : Membran yüzeyinden geçen akış miktarı (flux)
- $P_\Delta$ : Kanal giriş ve çıkışı arasındaki basınç farkı
- $C_s$ : Membran kirlenme indeksi (geçirgenlik azalmasına dayalı kirlilik göstergesi)

Farklı fiziksel boyutlara sahip bu çıktılar normalize edilerek ve ağırlıklandırılarak tek bir performans metriğinde birleştirilmiştir. Tanımlanan birleşik performans metriği (BPM), aşağıdaki şekilde ifade edilmektedir:

$$BPM = w_{J_w} \cdot \hat{J}_w - w_{P_\Delta} \cdot \hat{P}_\Delta - w_{C_s} \cdot \hat{C}_s \quad (2.9)$$

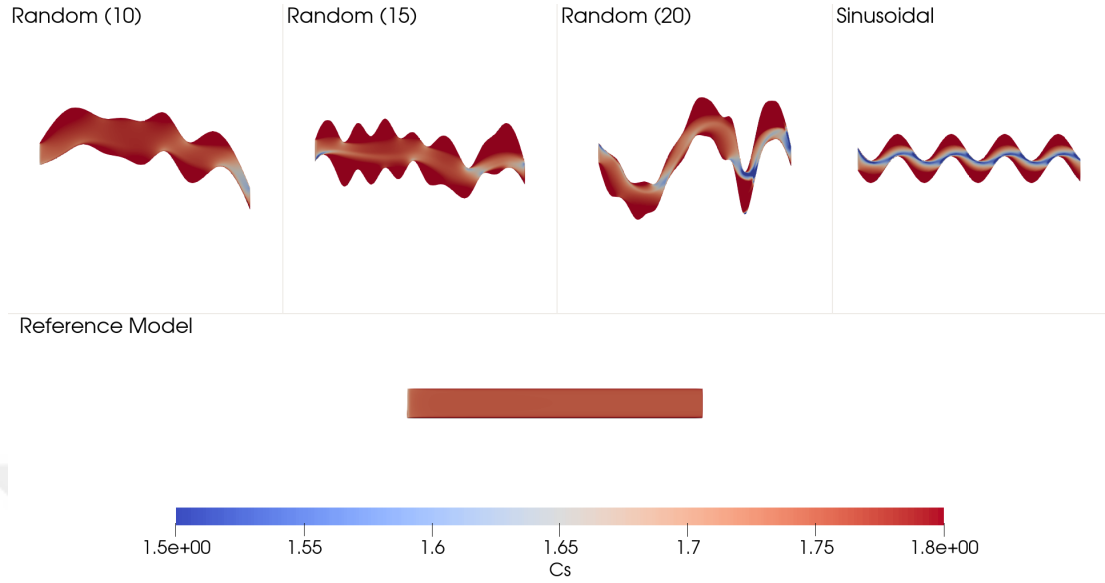
- $\hat{J}_w$ ,  $\hat{P}_\Delta$  ve  $\hat{C}_s$  sırasıyla normalize edilmiş permeat akısı, basınç farkı ve kirlenme indeks değerlerini,
- $w_{J_w} = 0.4$ ,  $w_{P_\Delta} = 0.3$ ,  $w_{C_s} = 0.3$  ise her bir çıktının önem derecesini temsil eden ağırlık katsayılarını ifade etmektedir.

Tanımlanan bu metrik, yüksek akış (flux), düşük basınç farkı ve minimum kirlenme indeksini aynı anda optimize etmeyi amaçlamaktadır. Bu nedenle,  $PM$  değeri ne kadar yüksekse, membran kanal geometrisinin performansı o kadar iyi kabul edilmektedir. Modelleme sürecinde bu metrik, makine öğrenmesi algoritmalarına hedef değişken (label) olarak verilmiştir.

## 2.5 Ön Çalışma-1

Rastgele geometriler ile akış kanalı tasarlanırken ilk önce farklı sayılarda noktalar belirlenmiştir. Kanal akış yönünde sağda ve solda ayrı ayrı 10, 15 ve 20 farklı rastgele nokta seçilmiştir ve bu noktalar Bspline'lara tanımlanarak 20, 30, 40 farklı rastgele nokta içeren kanal geometrileri oluşturulmuştur (Şekil 2.3). Daha sonra bu geometrilere

HAD simülasyonları uygulanmıştır ve elde edilen performans metrikleri sinüzoidal ve düz kanal geometrilerinin performans metrikleri ile karşılaştırılmıştır.



**Şekil 2.3:** Membran kirlenmesi dağılımı: Rastgele 10-nokta, 15-nokta ve 20-nokta, sinüzoidal model ve referans model.

**Çizelge 2.2:** Her bir kanal için membran yüzeyindeki ortalama değerler ve kanaldaki giriş çıkış arasındaki basınç farkları.

Model	Ortalama $J_w$		Yüzey Alanı	Basınç Farkı
	(Lmh) $10^{-3}$	Ortalama $C_s$		
Ref model	2.75	1.76	1.00	0.20
Random (10)	2.79	1.77	1.72	0.37
Random (15)	2.77	1.78	1.88	0.59
Random (20)	3.30	1.73	1.78	1.00
Sinüzoidal	2.83	1.72	1.00	1.20

Ön çalışma-1 kapsamında, farklı karmaşıklıklarda 10, 15 ve 20 nokta içeren üç rastgele modelin yanı sıra, geleneksel düz geometriye sahip bir referans model ve sinüzoidal model incelenmiştir. Simülasyon sonuçlarının karşılaştırmalı analizi, Çizelge 2.2'de sunulmakta olup; farklı kanal tasarımları için elde edilen ortalama permeat akısı ( $J_w$ ), ortalama yüzey kirlenme değeri ( $C_s$ ) ve giriş-çıkış basınç farkı gibi performans metriklerine dair önemli gözlemler elde edilmiştir. Bu metrikler, yüzey alanından bağımsızdır; yüzey alanı yalnızca ek bilgi olarak tabloya dahil edilmiştir. Ayrıca, tüm modellerin nihai kirlenme dağılımını gösteren görseller Şekil 2.3'de sunulmuştur.

Elde edilen bulgular, rastgele geometriye sahip tüm modellerin, referans model ile kıyaslandığında en az onun kadar veya daha yüksek ortalama  $J_w$  değerlerine ulaştığını göstermiştir. Özellikle Random (20) modeli, akı değerlerinde kayda değer bir artış göstermiştir ve bu özelliğiyle diğer modellerden ayrılmıştır. Bu durum, nokta sayısındaki artış ile akı performansı arasında potansiyel bir ilişki olabileceğini düşündürmektedir. Ancak 10 ve 15 noktalı modeller arasındaki farklar yeterince belirgin olmadığından bu konuda kesin sonuçlar çıkarılamamaktadır.

Hidrolik performans açısından önemli bir parametre olan basınç düşüşü değerlendirildiğinde, Random (20) modelinin sinüzoidal modele göre avantaj sağladığı görülmüştür. Bu durum, yüksek nokta sayılı yapıların hem kütle transfer verimliliğini hem de hidrolik direnci optimize edebileceğini göstermektedir.

Buna karşılık, kirlenme ( $C_s$ ) değerleri modeller arasında belirgin farklılık göstermemiştir. Bu da, kirlenme davranışının kanal geometrisindeki değişimlere daha az duyarlı olabileceğini ya da daha hassas bir analiz için ek deneysel verilere ihtiyaç duyulduğunu göstermektedir.

Genel olarak simülasyon sonuçları, özellikle nokta sayısının arttığı Random (20) modeli ile rastgele yapıların umut verici potansiyelini ortaya koymaktadır. Bu modeller, geleneksel dikdörtgen ve sinüzoidal yapıların verimlilik sınırlarını sorgulamakta ve artan geometrik karmaşıklığın sistem performansını artırabileceğine işaret etmektedir. Bu ön bulgular referans alınarak ikincil ön çalışma yapılmıştır, detayları bir sonraki bölümde açıklanmıştır.

## 2.6 Ön Çalışma-2

İkincil ön çalışmada, ön çalışma-1'de iyi performans gösterdiği tespit edilen 20 farklı nokta (Random 20 modeli) seçilerek 25 farklı rastgele geometri tasarımı yapılmıştır. Simüle edilmiş bu rastgele geometrilerden, belirlenen performans metriği kullanılarak üç adet en iyi ve üç adet görece en kötü performansa sahip toplam altı adet farklı geometri seçilmiştir. Daha sonra, referans amaçlı düz kanal geometrisi benzer aşamalardan geçirilmiştir ve oluşturulan bu altı membran kanalı ile karşılaştırılmak üzere simüle edilmiştir (Şekil 2.4).

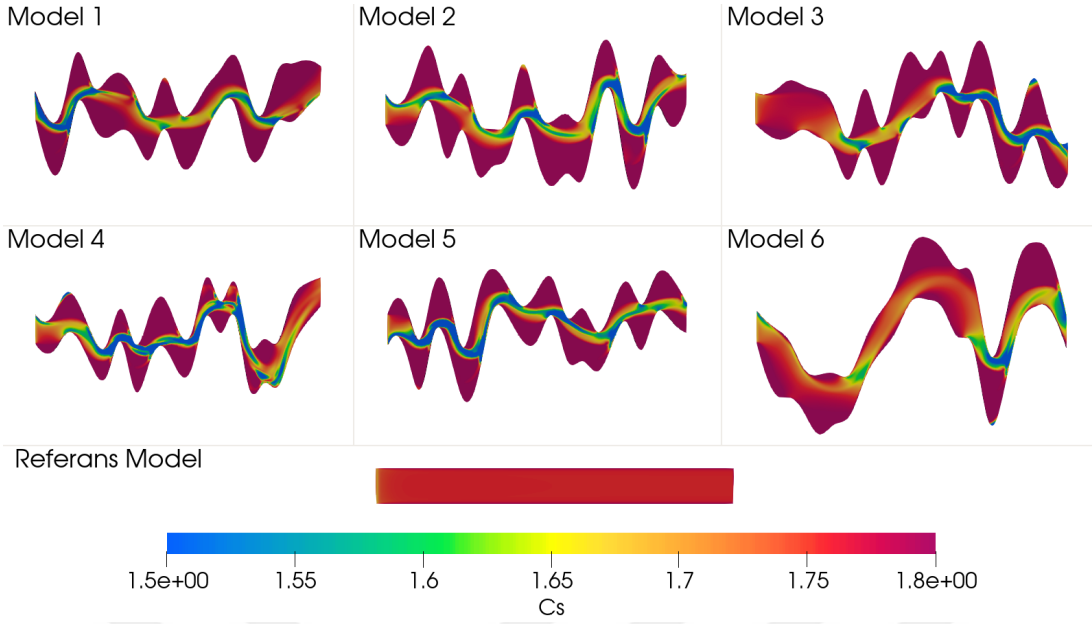
Bu modellerden 3 tanesi görece iyi performansa sahip; diğer 3 tanesi de görece kötü performansa sahip modellerden seçilmiştir. Model 1, 2, 3 yüksek performans gösteren modelleri; Model 4, 5, 6 ise düşük performans gösteren modelleri göstermektedir (Şekil 2.4). Referans model düz (dikdörtgen şeklinde) bir kanal geometrisine sahiptir. Her bir modelin simülasyon iterasyonları sonucunda ortaya çıkan performans metrikleri Çizelge 2.3’de verilmiştir. Bunlar, ortalama flux ( $J_w$ ), ortalama yüzey kirlenmesi ( $C_s$ ) ve uçtan uca basınç düşümüdür ( $\Delta P$ ). Ele alınan bu üç performans metriği yüzey alanından bağımsızdır ve her bir modelin yüzey alanı tabloda ek bilgi olarak verilmiştir. Tüm modellerin son durumdaki kirlenme karakterini gösteren görseller Şekil 2.4’de verilmiştir.

**Çizelge 2.3:** Her bir kanal için membran yüzeyindeki ortalama değerler ve kanaldaki giriş çıkış arasındaki basınç farkları.

Model	Ortalama $J_w$ (Lmh) $10^{-3}$	Ortalama $C_s$	Yüzey Alanı ( $m^2$ ) $10^{-3}$	Basınç Farkı (bar)
Ref model	2.75	1.76	1.00	0.20
Model 1	2.96	1.75	1.72	2.00
Model 2	3.29	1.73	1.76	1.60
Model 3	3.65	1.74	1.65	1.20
Model 4	3.01	1.71	1.37	2.10
Model 5	3.10	1.72	1.49	2.10
Model 6	3.30	1.73	1.78	1.00

Basınç düşüşünün ve  $C_s$  değerinin yüksek olması, sistem performansının iyi olmadığını ifade etmektedir. Buna karşılık,  $J_w$  değeri ne kadar yüksek olursa, sistem performansı açısından o kadar iyi olduğu anlamına gelmektedir. Fazla  $\Delta P$ , daha yüksek pompa gücü gerektirirken, yüksek  $C_s$  değeri membran kirlenmesinin arttığını ve dolayısıyla  $J_w$  değerinin düştüğünü göstermektedir.

Simülasyon sonuçlarının verildiği Çizelge 2.2 incelendiğinde tüm modellerin ortalama  $J_w$  değerlerinin referans modele göre daha iyi olduğu görülmektedir. Özellikle, iyi modeller arasında olan Model 3’e ait ortalama  $J_w$  değerinin diğer modellere oranla önemli derecede yüksek olduğu dikkat çekmektedir. Görece iyi ve kötü performanslı modeller karşılaştırıldığında ise ortalama  $J_w$  değerlerinin iyiler tarafında daha yüksek olduğu görülmektedir.



**Şekil 2.4:** Membran kirlenmesi dağılımı: performans metriğine göre daha yüksek performansa sahip Model 1, Model 2 ve Model 3, ve görece daha düşük performansa sahip Model 4, Model 5 ve Model 6 kanal geometrileri.

Öte yandan, basınç düşüşü açısından değerlendirildiğinde referans modelin tasarlanan rastgele kanal geometrilerine göre üstünlüğü olduğu gözlemlenmektedir.

Bununla birlikte, önemli sayılan ortalama  $C_s$  ise modeller arası bir miktar farklılık gösterse de bu sonuçlar için bilgi verici bir karakter sergilememiştir.

Sonuç olarak, simülasyon sonuçları değerlendirildiğinde rastgele modellerin referans modele göre daha umut verici sonuçlar ortaya koyduğu değerlendirilmiştir. Ayrıca, rastgele modellerin performans metriklerine göre sınıflandırılması bu modeller arasında yüksek performanslıların seçilmesi yönünde fayda sağlayabileceği gözlemlenmiştir. Bu konuda daha net yargılara erişebilmek için daha fazla modelin ele alınması ve karşılaştırmalı olarak incelenmesi gerekmektedir.

## 2.7 Ön Çalışma Sonuçları

Bu çalışma, RO sistemlerinde kullanılan akış kanalı geometrilerinin membran performansı üzerindeki etkilerini incelemek amacıyla gerçekleştirilmiştir. OpenFOAM yazılımı kullanılarak yapılan HAD simülasyonları sonucunda, yenilikçi rastgele oluşturulmuş kanal geometrilerinin geleneksel düz yapıya sahip geometrilere kıyasla daha iyi performans gösterdiği tespit edilmiştir. Özellikle, akış hızının artırılmasıyla

konsantrasyon polarizasyonunun azaltıldığı, bu sayede membran yüzeyinde oluşan kirlenmenin minimize edildiği ve permeat akışının iyileştirildiği belirlenmiştir.

Simülasyon sonuçları, belirli rastgele geometrilerin basınç kaybını artırmadan daha yüksek su akışına ve daha düşük membran kirlenmesine katkı sağladığını göstermiştir. Performans değerlendirmeleri, optimize edilmiş akış kanal tasarımlarının, RO sistemlerinde enerji tüketimi ile kütle transfer verimliliği arasındaki dengeyi optimize etmek için önemli bir araç sunduğunu ortaya koymuştur.

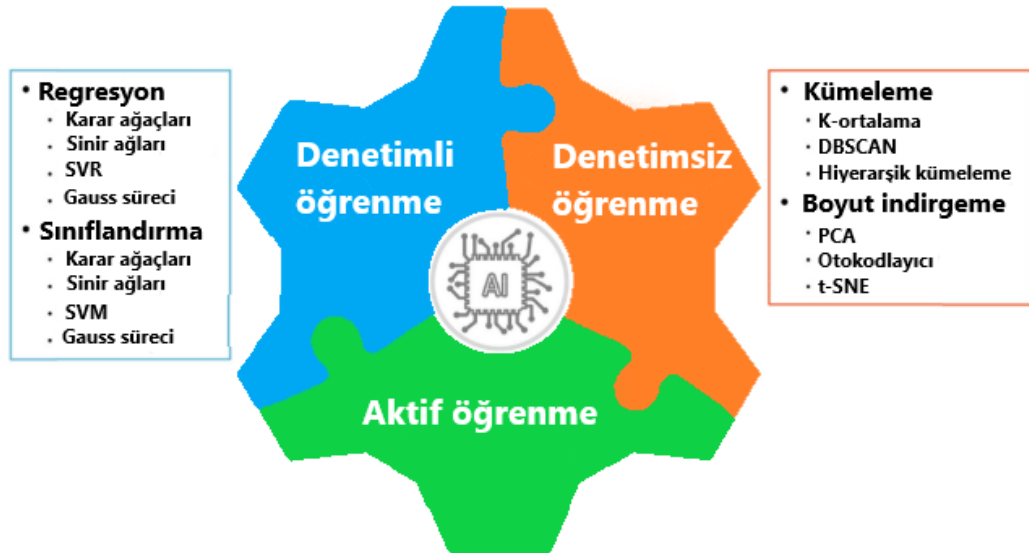
Elde edilen bulgular, RO sistemlerinde membran performansını artırmak amacıyla yeni nesil akış kanalı tasarımlarının geliştirilmesi için önemli bir rehber niteliği taşımaktadır. Bu çalışmanın devamında, daha fazla akış kanal geometrisi kullanarak eğitilen ML tabanlı yöntem ile membran sistemlerinin daha yüksek verimlilikle çalışmasını sağlayacak yüksek performanslı akış kanalı geometrilerinin seçilmesini sağlayacaktır.



### 3. YAPAY ZEKA MODELİ VE PERFORMANS TAHMİNLEMESİ

#### 3.1 Makine Öğrenmesi Yöntemleri

Makine öğrenmesi (ML), verilerden öğrenen ve veriye dayalı kararlar veren algoritmaların geliştirilmesi üzerine odaklanan bir yapay zeka (AI) dalıdır. Bu algoritmalar, açıkça programlanmadan belirli görevleri gerçekleştirebilir hale gelmektedir. ML temel olarak denetimli, denetimsiz ve aktif öğrenme olmak üzere üç ana kategoriye ayrılmaktadır ve Şekil 3.1’de gösterilmektedir [88].



**Şekil 3.1:** Membran bilimine uygulanan popüler makine öğrenmesi yöntemlerinin kısa bir sınıflandırması [89].

Denetimli öğrenme (supervised learning), etiketlenmiş veri kümeleri kullanılarak yapılan bir öğrenme yöntemidir. Bu yöntemde, algoritmaya hem girdi verisi hem de bu girdilere karşılık gelen doğru çıktılar (etiketler) sunulmaktadır. Algoritmanın amacı, girdi ve çıktı arasındaki haritalama fonksiyonunu öğrenmektir. Öğrenme tamamlandığında, model daha önce görmediği yeni girdiler için doğru çıktıyı tahmin etmeye çalışmaktadır. Bu yöntem, sınıflandırma (örneğin, bir e-postanın spam olup olmadığını belirleme) ve regresyon (örneğin, bir evin fiyatını tahmin etme)

problemleri için yaygın olarak kullanılmaktadır. Bu çalışmamızda, membranın akış kanal geometrisinin performansı tahmin edilmeye çalışılmakta olduğundan regresyon yöntemi ile performans tahminleme modeli geliştirilmiştir. Geliştirilen modelin detayları bölüm 4'te açıklanacaktır.

Denetimsiz öğrenme (unsupervised learning), etiketlenmemiş veri kümeleriyle çalışmaktadır. Bu yöntemde, algoritmaya yalnızca girdi verisi sunulmaktadır ve algoritmanın görevi, verideki gizli yapıları, örüntüleri veya ilişkileri kendi başına keşfetmektir. Denetimsiz öğrenme, kümeleme (örneğin, müşteri segmentasyonu), boyut indirgeme (örneğin, ana bileşen analizi) ve birliktelik kuralı madenciliği gibi alanlarda kullanılmaktadır.

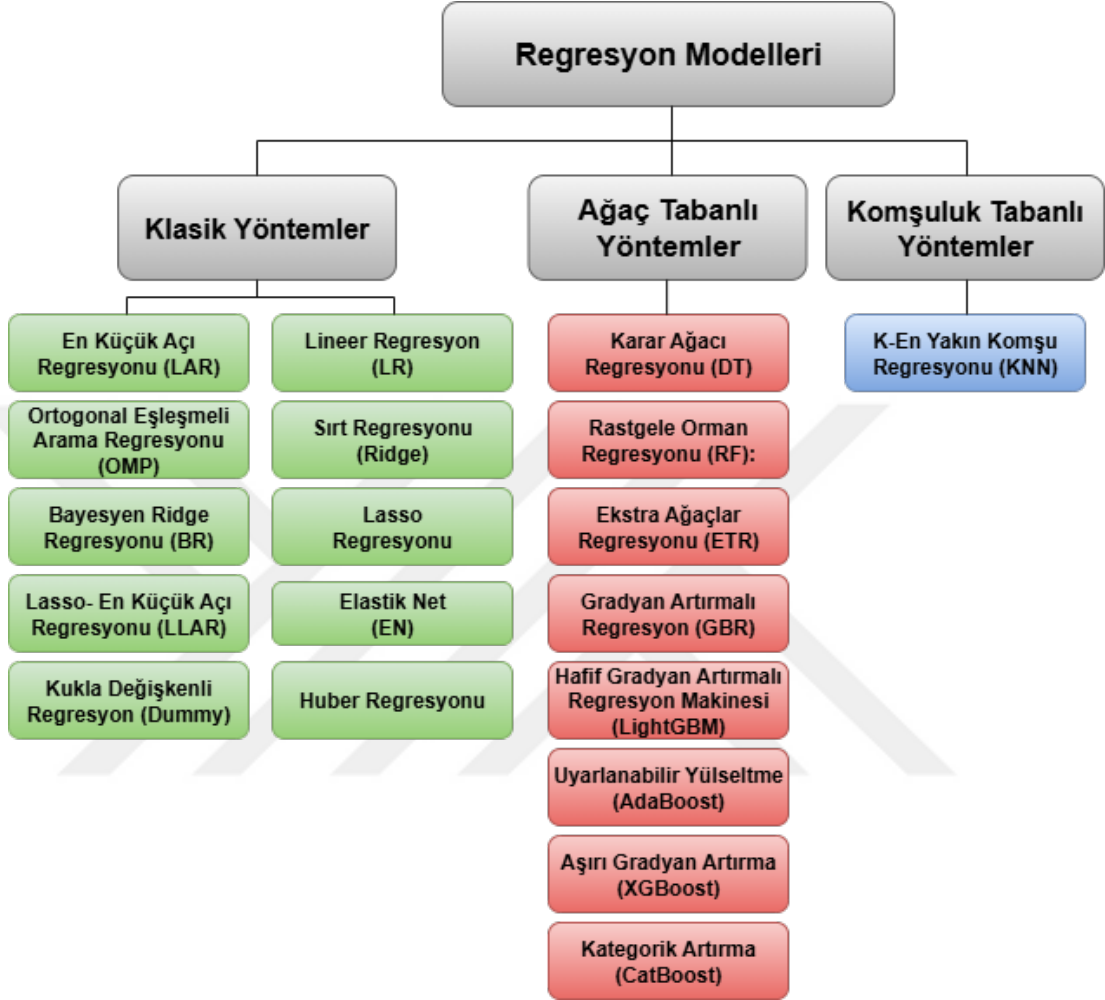
Aktif öğrenme (active learning), öğrenme algoritmasının kendi ihtiyaçlarına göre veri örneklerini seçtiği özel bir makine öğrenmesi alt dalıdır. Geleneksel denetimli öğrenmenin aksine, aktif öğrenme sistemi tüm veri kümesinin etiketlenmesini beklememektedir. Bunun yerine, modelin en çok bilgi kazanacağı ve belirsizliğini en aza indireceği örnekleri belirleyerek bir sorgulayıcının (genellikle bir insan uzmanının) bu örnekleri etiketlemesi istenmektedir. Bu yöntem, etiketleme maliyetinin yüksek olduğu durumlarda verimli ve etkili model eğitimi sağlamaktadır.

### **3.2 Regresyon Modelleri**

AI ve ML alanlarında, regresyon yöntemleri özellikle sürekli değişkenlerin tahmininde önemli bir yer tutmaktadır. Regresyon teknikleri, girdi özellikleri (bağımsız değişkenler) ile hedef değişken (bağımlı değişken) arasındaki ilişkileri modelleyerek geleceğe yönelik tahminler yapılmasına olanak sağlamaktadır. Bu çalışma, klasik regresyon yaklaşımlarından başlayarak günümüzde sıklıkla kullanılan gelişmiş ML regresyon algoritmalarına kadar kapsamlı bir açıklama sunmaktadır.

Regresyon, istatistiksel modellemenin temel yapı taşlarından biridir. Klasik, ağaç tabanlı ve komşuluk tabanlı olmak üzere üçe ayrılmaktadır ve Şekil 3.2'de gösterilmiştir. En yaygın kullanılan klasik yöntemler arasında lineer regresyon (LR), ridge regresyon, lasso, elastik net (EN) ve Huber gibi teknikler yer almaktadır. Bu yöntemler özellikle küçük ve orta ölçekli veriler üzerinde iyi sonuçlar vermektedir ve yorumlanabilirlik

açısından avantajlıdır [89]. Her bir regresyon yöntemi, farklı veri yapılarına ve problemlere daha uygun hale gelebilmektedir. Bir sonraki bölümde, her bir regresyon yöntemi kısaca tanıtılacaktır.



Şekil 3.2: Regresyon modellerinin sınıflandırılması.

### 3.2.1 Klasik regresyon yöntemleri

Lineer (doğrusal) regresyon (LR), bağımlı değişken ile bağımsız değişkenler arasındaki doğrusal ilişkiyi modelleyen en temel regresyon yöntemidir [90]. Modelin temel denklemi, denklem 3.1'de verilmiştir. Burada  $y$  tahmin edilen çıktı,  $x_i$  girdi değişkenleri,  $\beta_i$  katsayılar ve  $\epsilon$  hata terimidir. Model, en küçük kareler yöntemi (ordinary least squares - OLS) kullanarak, gözlemlenen ve tahmin edilen değerler arasındaki farkların karelerinin toplamını minimize eden katsayıları bulmaktadır.

$$y = \beta_0 + \beta_1 x_1 + \dots + \beta_n x_n + \epsilon \quad (3.1)$$

Sırt (ridge) regresyon, çoklu doğrusal bağlantı (multicollinearity) problemini ele almak için lineer regresyona bir L2 normalizasyon terimi eklemektedir [91]. Bu terim, katsayıların karelerinin toplamını minimize ederek onları sifıra yaklaştırır ancak sifır yapmaz. Bu, modelin aşırı uydurma (overfitting) yapmasını engellemektedir ve modelin genellenebilirliğini artırmaktadır.

Lasso (en küçük mutlak daralma ve seçim operatörü) regresyonu, ridge regresyona benzer şekilde bir ceza terimi kullanmaktadır; ancak bu bir L1 normalizasyon terimidir [92]. L1 ceza, katsayıların mutlak değerlerinin toplamını minimize etmektedir. Bu ceza, bazı katsayıları tam olarak sifıra indirerek değişken seçimi (feature selection) yapma özelliği sunmaktadır. Bu, modelin daha basit ve yorumlanabilir olmasını sağlamaktadır.

Elastik net (EN), lasso ve ridge yöntemlerini birleştiren hibrit bir regresyon yöntemidir. L1 ve L2 normalizasyon terimlerini birleştirerek [93], hem değişken seçimi yapma (lasso'nun avantajı) hem de çoklu doğrusal bağlantı durumunda daha kararlı sonuçlar üretme (ridge'in avantajı) yeteneğine sahiptir. Özellikle yüksek boyutlu veri setlerinde L1 ve L2 cezalarının avantajlarından yararlanmak için idealdir.

Huber regresyonu, aykırı değerlere (outliers) karşı daha dayanıklı bir regresyon modelidir [94]. Geleneksel en küçük kareler yöntemi, aykırı değerlerden büyük ölçüde etkilenirken, Huber kaybı (Huber loss) küçük hatalar için karesel, büyük hatalar için ise doğrusal bir ceza uygulamaktadır. Bu, modelin aşırı aykırı değerlerin etkisi altında bozulmasını engellemektedir.

En küçük açılı regresyonu (LAR), lasso regresyonuna benzer şekilde bir değişken seçim algoritmasıdır [95]. LAR, en yüksek korelasyona sahip değişkeni bulmakta ve katsayısını adım adım artırmaktadır. Herhangi bir noktada yeni bir değişkenin korelasyonu mevcut katsayılarla aynı seviyeye geldiğinde, yeni değişken de modele dahil edilmektedir. Bu, adım adım ve verimli bir şekilde en önemli değişkenleri bulmayı sağlamaktadır.

Ortogonal eşleşmeli arama regresyonu (OMP), seyrek sinyal işleme alanında ortaya çıkmış bir algoritma olup, regresyon bağlamında en uygun değişkenleri yinelemeli olarak seçmektedir [96]. Her adımda, artık hatalarla en yüksek korelasyona sahip

değişkeni seçmekte ve bu değişkeni modele dahil etmektedir. Bu süreç, belirlenen bir iterasyon sayısına veya hata eşiğine ulaşılan kadar devam etmektedir.

Bayesyen ridge regresyonu (BR), katsayılar için bir olasılıksal yaklaşım sunmaktadır [97]. Klasik ridge regresyonun aksine, katsayıların tek bir noktayla tahmini yerine, bu katsayılar için bir olasılık dağılımı (Gaussian dağılımı) tahmin etmektedir. Bu, modelin belirsizliğini de hesaba katarak daha sağlam tahminler yapmasını sağlamaktadır.

Lasso LAR (LLAR), LAR algoritmasını kullanarak lasso modelinin çözümlerini bulmaktadır [95]. LAR'ın verimli yapısı, lasso'nun değişken seçimi özelliğini hızlı ve stabil bir şekilde gerçekleştirmesini sağlamaktadır. Algoritma, lasso'nun karmaşık optimizasyon problemini daha basit ve sezgisel bir dizi adımda çözmektedir.

Kukla (dummy) değişkenli regresyon, regresyon analizi için bir baz çizgisi (baseline) modelidir. Amacı, ML modellerinin performansını değerlendirmek için bir referans noktası sağlamaktır. Model, tüm tahminler için sabit bir değer (örneğin, bağımlı değişkenin ortalaması) döndürmektedir veya basit bir strateji izlemektedir. Bu, karmaşık bir modelin basit bir referanstan daha iyi performans gösterip göstermediğini anlamak için kullanılmaktadır.

### **3.2.2 Ağaç tabanlı regresyon yöntemleri**

Ağaç tabanlı yöntemler, karar ağaçlarının yapısal özelliklerini kullanarak regresyon problemlerini çözmektedir. Bu algoritmalar, veriyi ardışık olarak bölerek tahminler yapmaktadır.

Karar ağacı regresyonu, bir veri kümesini daha küçük ve daha homojen alt gruplara ayırarak tahminler yapan bir modeldir [98]. Ağacın her bir düğümü, veriyi en iyi şekilde bölen bir özelliğe ve eşiğe dayanarak bir karar vermektedir. Yapraklara ulaşıldığında, bu yaprakta bulunan tüm örneklerin ortalama değeri tahmin olarak kullanılmaktadır. Karar ağaçları, yorumlanabilirlikleri ve doğrusal olmayan ilişkileri yakalayabilme yetenekleri nedeniyle popülerdir.

Ağaç tabanlı yöntemler genellikle topluluk öğrenme (ensemble learning) algoritmaları içinde kullanılmaktadır. Bu algoritmalar, tek bir modelin zayıflıklarını gidermek için birden fazla modeli bir araya getirmektedir. Bu yöntemler, aşağıda kısaca

açıklanmış olup, bu çalışmada iyi performans gösteren modeller bir sonraki bölümde detaylandırılmıştır.

Rastgele orman regresyonu (RF), bir dizi karar ağacından oluşan bir topluluk modelidir [99]. Her bir ağaç, veri kümesinden ve özelliklerden rastgele seçilen örneklerle eğitilmektedir. Tahmin, tüm ağaçların ürettiği bireysel tahminlerin ortalaması alınarak yapılmaktadır. Bu çoklu oy mekanizması, tek bir karar ağacının aşırı uydurma eğilimini azaltmakta ve modelin genelleme performansını artırmaktadır.

Ekstra ağaçlar regresyonu (ETR), rastgele orman regresyon yöntemine benzemektedir. Farkı ise, her bir ağaç için alt küme seçimi yapılırken rastgele bir eşik değeri belirlenmesidir [100]. Bu aşırı rastgelelik, modelin varyansını daha da düşürmekte ve hesaplama hızını artırabilmektedir.

Gradyan artırmalı regresyon (GBR), hata üzerine odaklanan bir topluluk öğrenme yöntemidir [101]. Algoritma, ardışık olarak yeni ağaçlar inşa etmektedir. Her yeni ağaç, bir önceki ağacın hatalarını minimize etmek için eğitilmektedir. Bu ardışık düzeltme süreci, modelin tahmin performansını kademeli olarak artırmaktadır.

Uyarlanabilir yükseltme regresyonu (AdaBoost), ağırlıklı bir topluluk öğrenme yöntemidir [102]. Modelin yanlış tahmin ettiği örneklere daha yüksek ağırlıklar vererek, bu örnekler üzerinde yeni zayıf öğreniciler (genellikle küçük karar ağaçları) eğitilmektedir. Bu süreç, modelin en zor örnekleri öğrenmesini sağlamaktadır.

Aşırı gradyan artırma regresyonu (XGBoost), gradyan artırmalı regresyonun optimize edilmiş ve ölçeklenebilir bir uygulamasıdır [103]. Geliştirilmiş düzenleme teknikleri, paralel hesaplama desteği ve eksik verilerle başa çıkma yeteneği gibi özellikleriyle yüksek performans sunmaktadır. Yarışmalarda ve endüstride yaygın olarak kullanılan güçlü bir algoritmadır.

Hafif gradyan artırmalı regresyon makinesi (LightGBM), Microsoft tarafından geliştirilen bir gradyan artırma çerçevesidir [104]. Büyük veri setleri için özellikle verimli ve hızlı olmasıyla bilinmektedir. Geleneksel GBR'nin aksine, ağaçları katman bazlı (level-wise) yerine yaprak bazlı (leaf-wise) büyütürken daha hızlı eğitim süreleri sunmaktadır.

Kategorik artırma (CatBoost), Yandex tarafından geliştirilen CatBoost, kategorik özelliklerle daha etkin bir şekilde başa çıkabilen bir GBR kütüphanesidir [105]. Geliştirilmiş sıralı artırma (ordered boosting) tekniği sayesinde, veri sızıntısını (leakage) önleyebilmekte ve kategorik değişkenleri önceden işleme gereksinimi olmadan kullanabilmektedir.

### 3.2.3 Komşuluk tabanlı regresyon yöntemi

Komşuluk tabanlı regresyon yöntemleri, bir veri noktasının tahminini en yakın komşularının özelliklerine göre yapan ve verinin yerel yapısını dikkate alan modellerdir. Bu kategoride en yaygın ve temel kullanılan yöntem K-En Yakın Komşu (KNN) Regresyonudur.

K-En Yakın Komşu regresyonu, en basit ve en sezgisel regresyon algoritmalarından biridir [106]. Yeni bir veri noktasının değeri, eğitim veri kümesindeki en yakın K komşusunun ortalama değeri olarak tahmin edilmektedir. Burada K, algoritmanın en önemli hiper-parametresidir ve komşu sayısını belirlemektedir. En yakın komşular genellikle Öklid mesafesi gibi bir uzaklık metriği kullanılarak bulunmaktadır.

### 3.3 Hafif Gradyan Artırmalı Regresyon Makinesi: LightGBM

LightGBM, gradyan artırmalı karar ağaçları (GBDT) temelli bir makine öğrenimi algoritması olup, büyük veri setleri üzerinde hızlı ve bellek açısından verimli biçimde çalışmak üzere tasarlanmıştır. Microsoft tarafından geliştirilen bu yöntem, geleneksel GBDT algoritmasına kıyasla bazı temel yapısal ve algoritmik yenilikler içermektedir. LightGBM, karar ağaçlarını iteratif olarak inşa ederek her bir iterasyonda hata terimini azaltan bir modelleme stratejisi izlemektedir. Her yeni ağaç, önceki tahminlerin hatalarını modellemeyi hedefleyerek modelin genel doğruluğunu artırmaktadır [104].

Modelin temel matematiksel yapısı, artan modeller dizisi olarak tanımlanmaktadır. Her bir iterasyonda öğrenilen model  $F_t(x)$ , bir önceki modele eklenen yeni bir karar ağacıyla güncellenmektedir. Bu yapı denklem 3.2'deki gibi ifade edilmektedir:

$$F_t(x) = F_{t-1}(x) + \eta f_t(x) \quad (3.2)$$

Burada  $F_t(x)$ ,  $t$ . iterasyonda elde edilen tahmin fonksiyonunu,  $F_{t-1}(x)$ , önceki iterasyonlara kadar yapılan toplam tahmini,  $f_t(x)$ ,  $t$ . iterasyonda öğrenilen zayıf tahminleyici olan karar ağacını,  $\eta$  ise öğrenme oranını (learning rate) temsil etmektedir. Amaç, denklem 3.3'te gösterilen toplam kayıp fonksiyonunu (loss function) minimize etmektir:

$$\mathcal{L} = \sum_{i=1}^n l(y_i, F_{t-1}(x_i) + \eta f_t(x_i)) \quad (3.3)$$

LightGBM, bu genel çerçeveyi önemli ölçüde optimize eden birkaç temel yenilik sunmaktadır. Bunların başında histogram tabanlı karar ağacı bölme yöntemi gelmektedir. Bu yöntemde, sürekli değişkenler, belirli sayıda aralığa (örneğin 255) bölünerek her özelliğe ait histogramlar oluşturulmaktadır. Böylece karar ağaçlarındaki en iyi bölme noktalarının aranması süreci, özellik değerlerinin doğrudan karşılaştırılması yerine histogramlar üzerinden gerçekleştirilmektedir. Bu yaklaşım, bellek kullanımını azaltmakla kalmamakta; aynı zamanda işlem hızını da önemli ölçüde artırmaktadır [104].

Diğer önemli farklılık, ağaç büyütme stratejisinde görülmektedir. Geleneksel GBDT algoritmalarında kullanılan katman bazlı büyüme yöntemi (level-wise tree growth), her iterasyonda tüm yaprakları eş zamanlı olarak genişletir. LightGBM ise bunun yerine yaprak bazlı büyüme stratejisi (leaf-wise growth) uygulamaktadır. Bu strateji, her iterasyonda en yüksek kayıp azaltımını sağlayan yaprağı tespit etmekte ve sadece bu yaprağı genişletmektedir. Böylece model, daha hızlı şekilde hatayı minimize edebilmektedir. Ancak, yaprak bazlı büyüme stratejisi kontrolsüz derinlik artışına yol açabileceği için genellikle maksimum yaprak sayısı veya maksimum derinlik gibi hiperparametrelerle sınırlandırılmaktadır.

Bir yaprağın bölünmesi sonucunda elde edilecek kazanç, gradyan (birinci türev) ve Hessian (ikinci türev) değerleri kullanılarak hesaplanmaktadır. Bu kazanç (split gain), denklem 3.4'teki gibi ifade edilmektedir.

$$\text{Gain} = \frac{1}{2} \left[ \frac{G_L^2}{H_L + \lambda} + \frac{G_R^2}{H_R + \lambda} - \frac{(G_L + G_R)^2}{H_L + H_R + \lambda} \right] - \gamma \quad (3.4)$$

Burada  $G_L$  ve  $G_R$ , sol ve sağ düğümdeki gradyan toplamlarını,  $H_L$  ve  $H_R$  ise Hessian toplamlarını temsil etmektedir.  $\lambda$ , L2 düzenleme terimi,  $\gamma$  ise yeni yaprak

oluşturmanın maliyetini temsil etmektedir. Bu formül, en iyi bölünmeyi belirlemek için kullanılmaktadır. L2 düzenleme, modelin aşırı öğrenmesini (overfitting) önlemek amacıyla, modelin karmaşıklığını cezalandıran bir terimdir. LightGBM’de bu düzenleme, yaprak skorlarının büyüklüğünü sınırlandırarak modelin genelleme kabiliyetini artırmaktadır. Bu bağlamda, yaprak skoru (leaf score), her bir yapraktaki toplam gradyanın Hessian toplamına oranını ifade etmektedir. Bu skorları sınırlamak için L2 cezası eklenmektedir.

LightGBM, eğitim sürecini hızlandırmak için gradyan tabanlı tek yönlü örnekleme (GOSS) adlı örnekleme yöntemini de kullanmaktadır. Bu yöntemde, gradyan değeri yüksek olan örnekler veri setinden tamamen alınırken, düşük gradyanlı örneklerden yalnızca rastgele bir kısmı örneklendirilmektedir. Sonrasında düşük gradyanlı örneklerin katkısı, örnekleme oranına göre yeniden ölçeklendirilmektedir. Bu sayede, bilgi kaybı en aza indirilirken işlem yükü azaltılmaktadır. Bir başka önemli özellik de özel değişken paketi (EFB) yöntemidir. Bu yaklaşım, birbirini dışlayan seyrek (sparse) öznitelikleri aynı paket içerisinde birleştirerek modelin öznitelik boyutunu azaltmaktadır. Özellikle yüksek boyutlu, sparse veri setlerinde bu özellik işlem verimliliğini ciddi ölçüde artırmaktadır. EFB, birlikte aktif olmayan (low conflict) öznitelikleri grafik tabanlı bir algoritma ile aynı demet içinde gruplayarak modelin bellekteki iz düşümünü daraltmaktadır.

LightGBM, herhangi bir diferansiyellenebilir kayıp fonksiyonu ile çalışabilmektedir. Regresyon için sıklıkla kullanılan kare hata fonksiyonu ya da sınıflandırma için çapraz entropi (log loss) gibi fonksiyonlar tercih edilmektedir. Kayıp fonksiyonu, genellikle ikinci dereceden Taylor açılımı ile yaklaşık olarak ifade edilmektedir. Bu açılım, eğitim sürecinde her iterasyonda kullanılacak karar ağacını optimize etmek için kullanılmaktadır ve denklem 3.5’te gösterilmektedir.

$$\mathcal{L} \approx \sum_{i=1}^n \left[ g_i f(x_i) + \frac{1}{2} h_i f(x_i)^2 \right] \quad (3.5)$$

Bu formülde  $\mathcal{L}$  toplam kaybı (loss) temsil ederken,  $f(x_i)$  öğrenilmekte olan yeni karar ağacının (veya zayıf öğrenicinin)  $i$ . örnek için yaptığı tahmini ifade etmektedir. Bu tahmin, önceki iterasyonlarda elde edilen tahminlere eklenecek düzeltme fonksiyonu olarak düşünülebilir.  $g_i$ ,  $i$ . örnek için modelin kayıp fonksiyonuna göre hesaplanan

birinci türevini, yani gradyanı ifade etmektedir. Gradyan, modelin hatasının hangi yönde arttığını ya da azaldığını belirlemektedir ve optimizasyonun yönünü tayin etmektedir. Matematiksel olarak denklem 3.6'daki gibi ifade edilmektedir.

$$g_i = \frac{\partial l(y_i, F_{t-1}(x_i))}{\partial F_{t-1}(x_i)} \quad (3.6)$$

Burada  $l$ , kayıp fonksiyonudur;  $y_i$  gerçek değer,  $F_{t-1}(x_i)$  ise önceki iterasyona ait tahmin değerini ifade etmektedir. Diğer taraftan,  $h_i$ ,  $i$ . örnek için kayıp fonksiyonunun ikinci türevidir ve Hessian değeri olarak adlandırılmaktadır. Hessian, gradyanın eğimini ifade etmektedir ve optimizasyonun kararlılığını artırmak için kullanılmaktadır. İlgili formül, denklem 3.7'deki gibi gösterilmektedir.

$$h_i = \frac{\partial^2 l(y_i, F_{t-1}(x_i))}{\partial F_{t-1}(x_i)^2} \quad (3.7)$$

Bu değerler, optimum bölünmelerin belirlenmesinde kullanılmaktadır. LightGBM, bu gradyan ve Hessian değerlerini kullanarak her iterasyonda yeni bir karar ağacı oluşturmakta ve böylece modelini güncellemektedir.

Modelin başarımını etkileyen birçok hiperparametre mevcuttur. Örneğin, yaprak sayısını kontrol eden `num_leaves`, öğrenme oranını belirleyen `learning_rate`, örnekleme oranları olan `feature_fraction` ve `bagging_fraction`, düzenleme terimleri olan `lambda_l1` ve `lambda_l2` gibi parametreler, modelin genelleme başarımı ve hesaplama süresi üzerinde doğrudan etkilidir. Bu parametreler deneysel olarak veya otomatik optimizasyon yöntemleriyle (örneğin `grid search`, `bayesian optimization`) belirlenebilmektedir.

LightGBM, klasik GBDT algoritmasının sınırlılıklarını aşmak üzere geliştirilmiştir. Özellikle büyük ölçekli ve yüksek boyutlu veri kümelerinde üstün performans sunmaktadır. Yaprak bazlı büyüme stratejisi, histogram temelli bölme, GOSS ve EFB gibi yenilikçi yaklaşımları sayesinde bellek verimliliğini artırmakta, eğitim süresini azaltmakta ve doğruluk oranlarını artırmaktadır. Günümüzde Kaggle gibi veri bilimi yarışmalarında en çok tercih edilen modellerden biridir.

### 3.4 Aşırı Gradyan Artırma: XGBoost

XGBoost, gradyan artırmalı karar ağaçları (GBDT) algoritmasının optimize edilmiş bir versiyonudur ve hem hesaplama verimliliği hem de model başarımını açısından birçok

yenilikçi bileşeni bünyesinde barındırmaktadır. XGBoost, 2016 yılında Tianqi Chen ve Carlos Guestrin tarafından geliştirilmiş ve özellikle büyük ölçekli veri setlerinde ve yarışma ortamlarında sunduğu yüksek doğruluk ve hız nedeniyle hızla yaygınlaşmıştır [103].

Modelin temel çalışma prensibi, hataları azaltmak amacıyla ardışık olarak karar ağaçları oluşturmaya dayanmaktadır. Her bir iterasyonda, mevcut modelin yaptığı hataları telafi edecek şekilde yeni bir zayıf öğrenici oluşturulmakta ve bu yeni model, önceki tahminlere eklenerek genel model güncellenmektedir. Bu yapı, LightGBM’de olduğu gibi aynı matematiksel denklem ile gösterilmektedir (bkz. denklem 3.2).

Toplam kayıp fonksiyonu, tahmin hatalarını ve modelin karmaşıklığını içeren iki terimin toplamı olarak tanımlanmaktadır ve denklem 3.8’deki şekilde gösterilmektedir.

$$\mathcal{L} = \sum_{i=1}^n l(y_i, F_t(x_i)) + \sum_{k=1}^t \Omega(f_k) \quad (3.8)$$

Burada  $l$ , gerçek değer  $y_i$  ile tahmin değeri  $F_t(x_i)$  arasındaki kaybı ölçen fonksiyonu (örneğin, kare hata veya log-kayıp fonksiyonu) ifade etmektedir.  $\Omega(f_k)$ ,  $k$ . ağacın karmaşıklığını kontrol eden düzenleme terimini ifade etmektedir. XGBoost’un diğer GBDT uygulamalarından en önemli farkı, bu düzenleme terimini açık biçimde kayıp fonksiyonuna dahil etmesidir. Bu durum, modelin aşırı öğrenmeye (overfitting) karşı dayanıklılığını artırmaktadır.

XGBoost’ta  $\Omega(f)$  denklem 3.9’deki şekilde tanımlanmaktadır ve aşağıda gösterilmiştir.

$$\Omega(f) = \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T w_j^2 \quad (3.9)$$

Bu formülde  $T$ , karar ağacının toplam yaprak (leaf) sayısını ifade ederken,  $w_j$  ifadesi  $j$ . yaprağın skorunu (ağırlığını) temsil etmektedir. Modelin karmaşıklığını kontrol altına almak için formüle dahil edilen  $\gamma$ , yeni bir yaprak eklemenin sabit maliyeti olarak tanımlanmaktadır ve gereksiz bölünmeleri engellemeye yardımcı olmaktadır. Aynı zamanda,  $\lambda$  sembolü ile gösterilen L2 düzenleme (regularizasyon) parametresi, her bir yaprak skorunun büyüklüğüne ceza uygulayarak modelin öğrenme kapasitesini sınırlamaktadır. Bu bileşenlerin birlikte çalışması, modelin aşırı karmaşık yapılar öğrenmesini önleyerek daha genelleştirilebilir (overfitting’e dirençli) bir yapı elde edilmesini sağlamaktadır.

Kayıp fonksiyonunun optimizasyonu için ikinci dereceden Taylor açılımı kullanılmaktadır. Bu sayede, doğrusal olmayan kayıp fonksiyonları da yaklaşık olarak optimize edilebilmektedir. İkinci dereceden açılım denklem 3.10'da gösterilmektedir.

$$\mathcal{L} \approx \sum_{i=1}^n \left[ g_i f_t(x_i) + \frac{1}{2} h_i f_t(x_i)^2 \right] + \Omega(f_t) \quad (3.10)$$

Burada da gradyan  $g_i$ , bu kayıp fonksiyonunun tahmin değerine göre birinci türevidir (denklem 3.6) ve modelin yönünü belirlerken kullanılmaktadır. Hessian  $h_i$  ise ikinci türevidir (denklem 3.7) ve adım büyüklüğünün ayarlanmasında rol oynamaktadır.

Karar ağacı üzerinde hangi bölünmenin daha faydalı olacağına karar vermek için kazanç fonksiyonu kullanılmaktadır. Bir yaprağın iki alt yaprağa bölünmesiyle elde edilen kazanç, denklem 3.4'teki gibi ifade edilmektedir. LightGBM ve XGBoost modelleri, düğüm bölme kazancını benzer matematiksel formüllerle ifade etmektedir; ancak kazanç değerleri aynı değildir. Aralarındaki temel fark, hesaplama yönteminden kaynaklanmaktadır. LightGBM, histogram tabanlı yaklaşık algoritmalar kullanarak gradyan ( $G$ ) ve Hessian ( $H$ ) değerlerini belirli aralıklara göre toplulaştırmaktadır, bu da kazanç değerini yaklaşık hale getirmektedir. Buna karşın XGBoost, gerçek değerlerle doğrudan çalışmaktadır ve daha hassas kazanç hesaplamaları yapmaktadır [103,104]. XGBoost kazanç değeri pozitifse bölme yapılmakta; negatifse yaprak bölünmemektedir. Böylece, sadece anlamlı bilgi taşıyan bölünmeler gerçekleştirilmektedir.

XGBoost, ağaç yapısını büyütürken, her düğümde en yüksek kazancı sağlayacak özelliği seçmekte ve bu işlem, modelin karmaşık yapılar öğrenebilmesini sağlamaktadır. Bununla birlikte, ağaç büyüklüğü, maksimum derinlik, minimum örnek sayısı, öğrenme oranı gibi hiperparametrelerle sınırlandırılmaktadır. Bu parametrelerin uygun şekilde ayarlanması, modelin performansı açısından kritik önem taşımaktadır.

XGBoost'un bir diğer önemli özelliği, eksik verilerle etkili biçimde çalışabilmesidir. Model, eğitim sırasında her eksik veri noktası için otomatik olarak en uygun yönü (sol/sağ dal) belirlemekte ve bu yolu takip etmektedir. Bu özellik, özellikle gerçek dünya veri setlerinde eksik verilerle karşılaşıldığında modelin doğruluğunu korumasını sağlamaktadır.

XGBoost, aynı zamanda paralel işlemeye olanak tanıyan yapısı sayesinde büyük veri setlerinde oldukça verimli çalışmaktadır. Split noktalarının paralel olarak hesaplanabilmesi, modelin hem CPU hem de GPU üzerinde yüksek hızla eğitilebilmesini sağlamaktadır. Bu da onu büyük ölçekli ticari ve akademik projelerde tercih edilen bir algoritma haline getirmiştir.

XGBoost, gradyan artırmalı karar ağaçları çerçevesini hem teorik hem de pratik düzeyde geliştiren bir yöntem olarak kabul görmüştür. Modelin başarımı, düzenleme terimleri, ikinci dereceden optimizasyon yaklaşımı ve paralel işleme yetenekleri sayesinde yüksek genelleme gücü, hesaplama verimliliği ve doğruluk sağlamaktadır. Günümüzde sınıflandırma, regresyon, sıralama ve anomali tespiti gibi pek çok farklı problemde yaygın olarak kullanılmakta ve hala veri bilimi yarışmalarında en yüksek performanslı modeller arasında yer almaktadır [103,107].

### **3.5 Kategorik Artırma: CatBoost**

CatBoost, Yandex tarafından geliştirilmiş bir gradyan artırmalı karar ağaçları (GBDT) algoritmasıdır ve özellikle kategorik (nitel) değişkenlerin etkili şekilde işlenmesi için optimize edilmiştir [108]. Model, gradyan artırma yaklaşımının temel ilkelerini takip ederken, overfitting (aşırı öğrenme) riskini azaltmaya, veri sızıntısını (target leakage) önlemeye ve öğrenme sürecini istatistiksel olarak kararlı hale getirmeye yönelik özel yapılar içermektedir.

CatBoost, LightGBM ve XGBoost gibi gradyan artırma algoritmalarında olduğu gibi, model çıktısını iteratif olarak oluşturmaktadır. Her bir adımda, önceki modelin hatalarını telafi edecek şekilde yeni bir karar ağacı inşa edilir. Model çıktısı denklem 3.2'deki gibi diğer GBDT tabanlı modellerle (XGBoost, LightGBM vb.) aynı şekilde ifade edilmektedir.

CatBoost'un diğer modellerden ayrıldığı en önemli nokta, tahmin hatalarının (residual) nasıl hesaplandığı ve bu hataların modele nasıl yansıtıldığıdır. Geleneksel GBDT algoritmalarında, modelin çıktısı üzerinden doğrudan gradyan (birinci türev) ve Hessian (ikinci türev) değerleri hesaplanarak kayıp fonksiyonu yaklaşıkleştirilmektedir. Bu yaklaşımda kullanılan genel ikinci dereceden Taylor açılımı aşağıdaki gibidir:

$$L(y, F_{t-1}(x) + f_t(x)) \approx L(y, F_{t-1}(x)) + g_i \cdot f_t(x_i) + \frac{1}{2} h_i \cdot f_t(x_i)^2 \quad (3.11)$$

Burada  $g_i$  ve  $h_i$ , sırasıyla, gradyan ve Hessian değerleridir. CatBoost da bu ikinci dereceden açılım yaklaşımını kullanmaktadır; ancak bu türevleri hesaplama ve güncelleme sürecini diğer modellerden farklı olarak sıralı artırma (ordered boosting) adı verilen bir teknikle uygulamaktadır. Sıralı artırma, eğitim verisini rastgele sıralayarak her örnek için yalnızca ondan önce gelen örneklerin bilgilerini kullanarak gradyan ve hedef temelli istatistikleri hesaplayan bir yaklaşımdır. Bu yöntem, eğitim sürecinde bilgi sızıntısını (data leakage) engelleyerek modelin genellebilirliğini artırmaktadır. Bu bağlamda, CatBoost'un teorik farklılıkları, türevlerin yapısından değil, bu türevlerin ne zaman ve nasıl hesaplandığından kaynaklanmaktadır.

Kayıp fonksiyonu olarak CatBoost, regresyon problemleri için genellikle ortalama kare hata (MSE) kullanmaktadır ve denklem 3.12'deki gibi ifade edilmektedir.

$$L = \frac{1}{n} \sum_{i=1}^n (y_i - F_{t-1}(x_i))^2 \quad (3.12)$$

Burada  $y_i$ , gerçek hedef değeri;  $F_{t-1}(x_i)$ , önceki iterasyondaki tahmini;  $n$ , toplam örnek sayısını ifade eder. Sınıflandırma problemlerinde ise logaritmik kayıp fonksiyonu (log-loss) tercih edilmektedir.

CatBoost'un en önemli yeniliklerinden biri, kategorik verilerin işlenmesindeki yaklaşımıdır. Geleneksel yöntemlerde kategorik değişkenler, tek seçimli kodlama (one-hot encoding) veya etiket kodlama (label encoding) gibi yöntemlerle sayısallaştırılmaktadır. Bu tür dönüşümler, yüksek boyutluluk (curse of dimensionality) ve bilgi kaybı gibi sorunlara neden olabilmektedir. CatBoost, bunun yerine hedef temelli kodlama (target-based encoding) yönteminin sıralı bir versiyonunu kullanmaktadır. Bu yöntemde, her kategorik değer karşılığı denklem 3.13'teki şekilde hesaplanmaktadır.

$$\text{Kodlanmış Değer}_i = \frac{\sum_{j < i} y_j \cdot \mathbb{I}(x_j = x_i) + a \cdot P}{\sum_{j < i} \mathbb{I}(x_j = x_i) + a} \quad (3.13)$$

Bu denklemde  $\mathbb{I}(x_j = x_i)$  gösterim fonksiyonu,  $a$  smoothing parametresi,  $P$  ise hedef değişkenin genel ortalamasıdır. Bu yapı, küçük örnek sayılarına sahip kategorilerde ezberleme (overfitting) riskini azaltmak için önemlidir.

CatBoost, XGBoost ve LightGBM gibi diğer gradyan artırmalı karar ağaçları algoritmalarıyla benzer matematiksel temellere dayansa da, özellikle kategorik veri işleme ve modelin genellenebilirliği açısından önemli farklılıklar sunmaktadır. CatBoost, kategorik değişkenleri işlemek için sıralı hedef temelli kodlama (ordered target-based encoding) yöntemini kullanarak bilgi sızıntısını önlemeyi başarmaktayken [105]; XGBoost ve LightGBM modelleri bu tür verilerin önceden kullanıcı tarafından sayısallaştırılmasını gerektirmektedir. Ayrıca CatBoost, modelin her iterasyonunda yalnızca önceki örneklerin istatistiksel bilgilerini kullanarak model güncellemesi gerçekleştiren sıralı artırma (ordered boosting) yaklaşımını benimsemektedir. Bu yapı, veri sızıntısını önlemekte ve aynı zamanda aşırı öğrenme riskini önemli ölçüde azaltmaktadır. Diğer modeller ise tüm eğitim verisini aynı anda kullanarak gradyanları hesaplamakta ve bu durum, özellikle küçük veri setlerinde modelin genelleme kapasitesini olumsuz etkileyebilmektedir.

CatBoost, aynı zamanda geleneksel karar ağaçları yerine simetrik ağaçlar (oblivious trees) kullanmaktadır. Bu yapı, her seviyede aynı özellik ve eşik değeri üzerinden karar verilmesini sağlayarak, modelin daha kararlı ve düzenli bir yapıya sahip olmasını mümkün kılmaktadır. Aynı zamanda, GPU üzerinde paralel hesaplama avantajı sunduğu için eğitim sürecinde ciddi performans kazanımları sağlamaktadır. XGBoost ve LightGBM ise genellikle asimetrik ağaç yapısı ya da histogram tabanlı ayırıştırma tekniklerini tercih etmektedir. Dolayısıyla, CatBoost, gradyan artırmalı karar ağaçlarının genel yapısını korurken; sıralı artırma, hedef temelli kodlama ve simetrik ağaç yapısı gibi özgün yaklaşımlarıyla, özellikle kategorik veri içeren ve aşırı öğrenme riski taşıyan problemlerde yüksek doğruluk ve güçlü genellenebilirlik sunan etkili bir yöntem olarak öne çıkmaktadır. Bu tez çalışmasında da CatBoost'un diğer regresyon modellerine göre daha üstün performans gösterdiği ilerideki bölümlerde açıklanmıştır.

### **3.6 Uyarlanabilir Yükseltme: AdaBoost**

AdaBoost, sınıflandırma problemlerinde zayıf sınıflayıcıları ardışık şekilde birleştirerek güçlü bir topluluk (ensemble) model oluşturan önyargı azaltma yöntemlerinden biridir. Freund ve Schapire tarafından 1997 yılında geliştirilen bu yöntem, özellikle karar

ağaçları gibi basit sınıflayıcılarla yüksek doğruluklu modeller üretme kapasitesi sayesinde istatistiksel öğrenme kuramında önemli bir yer edinmiştir [102].

AdaBoost'un temel mantığı, her iterasyonda yanlış sınıflandırılan örneklerin ağırlıklarını artırmak ve doğru sınıflandırılanların ağırlıklarını azaltarak, bir sonraki sınıflayıcının bu hatalı örneklere odaklanmasını sağlamaktır. Böylece, ardışık olarak öğrenen zayıf sınıflayıcılar, ağırlıklı oylama ile birleşerek güçlü bir sınıflayıcı elde etmektedir.

AdaBoost'un matematiksel algoritması, verilen eğitim veri seti  $\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$  ile  $x_i \in \mathbb{R}^d$  ve  $y_i \in \{-1, +1\}$  olmak üzere, aşağıdaki adımlarla işlemektedir.

İlk olarak tüm örneklere denklem 3.14'te gösterildiği gibi eşit ağırlık atanmaktadır.

$$w_1(i) = \frac{1}{n}, \quad \text{tüm } i \text{ değerleri için} \quad (3.14)$$

Her iterasyonda, zayıf öğrenici  $f_t(x)$ , mevcut ağırlıklı verilerle eğitilmektedir ve hata oranı denklem 3.15'te gösterildiği gibi hesaplanmaktadır.

$$\epsilon_t = \sum_{i=1}^n w_t(i) \cdot \mathbb{I}(f_t(x_i) \neq y_i) \quad (3.15)$$

Zayıf sınıflayıcının güven katsayısı (alpha değeri) denklem 3.16'daki gibi hesaplanmaktadır.

$$\alpha_t = \frac{1}{2} \ln \left( \frac{1 - \epsilon_t}{\epsilon_t} \right) \quad (3.16)$$

Ağırlıklar bir sonraki iterasyon için güncellenmekte (denklem 3.17) ve güncellenen ağırlıklar normalize edilerek toplamları 1 yapılmaktadır. Tüm zayıf sınıflayıcılar toplandıktan sonra nihai tahmin denklem 3.18'deki gibi hesaplanmaktadır.

$$w_{t+1}(i) = w_t(i) \cdot e^{-\alpha_t y_i h_t(x_i)} \quad (3.17)$$

$$H(x) = \text{sign} \left( \sum_{t=1}^T \alpha_t h_t(x) \right) \quad (3.18)$$

AdaBoost algoritması, zayıf sınıflayıcıların ardışık şekilde bağlanmasıyla adaptif öğrenme gerçekleştirilmesi açısından artırma (boosting) algoritma ailesinin öncüsü olarak kabul edilmektedir. Ancak, AdaBoost'un bazı sınırlamaları mevcuttur. Özellikle, aykırı değerler ve gürültülü veriler üzerinde aşırı duyarlıdır; çünkü yanlış

sınıflandırılan örneklerin ağırlıkları her iterasyonda artmaktadır ve model bu örneklere fazlaca odaklanabilmektedir. Bu da aşırı öğrenme riskini artırmaktadır.

LightGBM, XGBoost ve CatBoost gibi modern artırma algoritmaları ise daha gelişmiş gradyan artırma çerçevesine dayanmaktadır. Bu algoritmalar, kayıp fonksiyonunun gradyanını minimize etmeye çalışmaktayken; AdaBoost, sınıflandırma hatasına doğrudan odaklanmaktadır. Özellikle XGBoost ve LightGBM, ikinci dereceden türevleri de kullanarak daha hassas optimizasyon yapmaktadır. Ayrıca bu modeller, eksik veri, regülasyon ve paralel hesaplama gibi alanlarda AdaBoost'tan çok daha üstündür.

CatBoost ise AdaBoost'un eksik kaldığı kategorik veri işleme konusuna çözüm getiren sıralı kodlama ve aşırı öğrenme önleyici sıralı artırma gibi yeniliklerle donatılmıştır. Buna karşın AdaBoost'un teorik olarak daha sade ve yorumu kolay bir yapısı vardır; bu nedenle özellikle giriş seviyesi sınıflandırma problemlerinde tercih edilebilmektedir.

### 3.7 Gradyan Artırmalı Regresyon: GBR

Gradyan artırmalı regresyon (GBR) yöntemi, zayıf öğrenicilerin (weak learners) ardışık olarak eğitilip birleştirilmesiyle oluşturulan güçlü bir tahminleme yöntemidir. Temel amaç, her iterasyonda mevcut modelin ürettiği hataları (residuals) her adımda azaltacak şekilde yeni bir zayıf öğrenici eklemektir. Bu bağlamda,  $t$ . iterasyondaki toplam tahmin fonksiyonu  $F_t(x)$ , bir önceki tahmin fonksiyonu  $F_{t-1}(x)$  ile o iterasyonda öğrenilen zayıf öğrenicinin  $\eta$  çarpanı ile ağırlıklı katkısının toplamı olarak ifade edilmektedir. Bu güncelleme kuralı, tüm modern tüm artırma (boosting) yöntemlerinde ortak bir yapı olup denklem 3.2'de gösterilmektedir.

GBR'nin optimizasyon süreci, toplam kayıp fonksiyonunun iteratif olarak en aza indirilmesine dayanmaktadır. Özellikle ortalama kare hata (MSE) gibi bir kayıp fonksiyonları için Friedman'ın [101] tanımladığı gradyan tabanlı yaklaşım, her adımda mevcut tahmin ile gerçek hedef arasındaki farkın yönünü ve büyüklüğünü temsil eden negatif gradyanı (pseudo-residual) hesaplanmaktadır. Bu değer,  $i$ . örnek için denklem 3.19'da gösterildiği gibi tanımlanmaktadır.

$$r_i^{(t)} = - \left. \frac{\partial l(y_i, F(x_i))}{\partial F(x_i)} \right|_{F(x)=F_{t-1}(x)} \quad (3.19)$$

Burada  $l$ , kayıp fonksiyonunu;  $F_{t-1}(x_i)$ , bir önceki iterasyondaki tahmin değerini;  $r_i^{(t)}$  ise o iterasyonda  $i$ . örnek için hesaplanan negatif gradyanı ifade etmektedir. Yeni öğrenici  $f_t(x)$ , bu negatif gradyan değerlerini mümkün olduğunca doğru tahmin edecek şekilde eğitilmekte ve model, her iterasyonda hatayı en hızlı azaltma yönünde güncellenmiş olmaktadır. Negatif gradyan kavramı, LightGBM, XGBoost ve CatBoost gibi diğer gradyan artırma yöntemlerinde de mevcut olup,  $F_t$  güncelleme kuralı bu modellerle ortaktır.

Bu bağlamda, artırma (boosting) algoritmalarının temelinde, her iterasyonda mevcut modelin hatasını yeni bir zayıf öğrenici ekleyerek azaltmak vardır. Hata, kayıp fonksiyonunun gradyanı ile sayısal olarak ifade edilmekte, negatif gradyan optimizasyon sürecinde en hızlı hata azaltma yönünü vermekte ve her iterasyonda hedef değer olarak kullanılmaktadır. Model, bu negatif gradyanı doğru şekilde tahmin edecek zayıf öğreniciyi ekleyerek performansını kademeli olarak artırmaktadır.

Klasik GBR, birinci-ordu (first-order) yaklaşım kullanmaktadır ve yalnızca gradyan bilgisinden yararlanmaktadır. Buna karşılık, XGBoost ve LightGBM gibi daha gelişmiş sürümler, ikinci-ordu (second-order) bilgiyi, yani Hessian'ı da kullanarak optimizasyon sürecini iyileştirmektedir ve daha hızlı yakınsama sağlamaktadır. Taylor açılımı ile yapılan ikinci dereceden yaklaşıklaştırma, gradyan  $g_i$  (denklem 3.6) ve Hessian  $h_i$  (denklem 3.7) tanımları tüm bu yöntemlerde ortaktır. Ancak, uygulamada bu değerlerin hesaplanma biçimleri farklılık göstermektedir.

Ağaç tabanlı modellerde, düğüm bölünme kararlarını değerlendirmek için kullanılan kazanç fonksiyonu, ortak kullanılmaktadır ve denklem 3.4'teki gibi ifade edilmektedir. Kazanç, bir bölünmenin toplam kayıpta sağladığı azalmayı nicel olarak ölçmektedir. XGBoost, LightGBM ve CatBoost uygulamalarında kazanç formülünün matematiksel yapısı benzerdir; fakat kullanılan  $G$  ve  $H$  değerlerinin tam mı yoksa yaklaşık mi hesaplandığına bağlı olarak pratikte farklı sonuçlar üretilmektedir. Klasik GBR, özünde birinci-ordu yöntemi olduğundan Hessian terimini açıkça kullanmamaktadır. XGBoost, LightGBM ise ikinci-ordu terimlerini aktif biçimde kullanarak bölünme seçimlerini ve yaprak ağırlık optimizasyonunu iyileştirmektedir.

### 3.8 Fizikle Öğrenen Yapay Sinir Ağı Yöntemi: PINN

Fizikle öğrenen yapay sinir ağı (PINN) [109], modelleri, fiziksel süreçlerin yapay sinir ağları (ANN) aracılığıyla temsil edildiği bir çerçeve sunmaktadır. Bu modellerin temel farkı, yalnızca veriye dayalı öğrenim yapmaları değil, aynı zamanda fiziksel sistemlerin karakteristiklerini temsil eden diferansiyel denklemleri veya başka fiziksel önbilgileri doğrudan öğrenme sürecine entegre etmeleridir. Böylece, model yalnızca bir gözlem fonksiyonuna (örneğin hata fonksiyonu) değil, aynı zamanda fiziksel süreçlerin kısıtlarını da minimize ederek eğitilebilmektedir. Bu yaklaşım, Raissi ve arkadaşları tarafından sistematikleştirilmiş ve geniş bir uygulama alanı bulmuştur [110].

Bu çalışmada kullanılan PINN modeli, fiziksel bir sürecin belirli tasarım parametrelerine (membranın akış kanalının geometri noktalarına) bağlı olarak elde edilen performans çıktısını öğrenmeyi amaçlamıştır. Model, çok katmanlı tam bağlantılı bir yapay sinir ağı (fully-connected feedforward neural network) olarak tasarlanmıştır. Girdi uzayı, fiziksel sistemin yapısal özelliklerini tanımlayan sayısal vektörlerden oluşmaktadır; çıktı ise bu yapıya karşılık gelen tekil bir performans değeri olarak ifade edilmiştir. Model matematiksel olarak denklem 3.20'deki şekilde tanımlanmıştır.

$$\hat{y} = f_{\theta}(\mathbf{x}), \quad \mathbf{x} \in \mathbb{R}^d, \quad \hat{y} \in \mathbb{R} \quad (3.20)$$

Burada  $\mathbf{x}$ ,  $d$ -boyutlu bir girdi vektörünü;  $\hat{y}$ , modelin tahmin ettiği çıktıyı; ve  $f_{\theta}$ , parametreleri  $\theta$  ile gösterilen sinir ağı fonksiyonunu ifade etmektedir. Bu sinir ağı, her katmanda doğrusal dönüşüm (linear transformation) ve doğrusal olmayan aktivasyon fonksiyonlarından oluşmaktadır. Modelin mimarisi aşağıdaki şekilde temsil edilmiştir:

$$f_{\theta}(\mathbf{x}) = L_n \circ \sigma \circ L_{n-1} \circ \dots \circ \sigma \circ L_1(\mathbf{x}) \quad (3.21)$$

Burada  $L_i$  terimleri tam bağlantılı katmanları (dense layers),  $\sigma$  ise Tanh (hiperbolik tanjant) gibi doğrusal olmayan aktivasyon fonksiyonlarını temsil etmektedir. Modelin çıktısı tek bir gerçek sayı olup, tahmin edilmek istenen fiziksel performans metriğine karşılık gelmektedir.

Modelin eğitimi sırasında kullanılan toplam kayıp fonksiyonu (loss function) iki bileşenden oluşmaktadır: veri kaybı (data loss) ve fizik bilgilendirmeli kayıp (physics-informed loss). Toplam kayıp aşağıdaki şekilde ifade edilmiştir:

$$\mathcal{L}_{\text{toplam}} = \mathcal{L}_{\text{veri}} + \lambda_{\text{fizik}} \cdot \mathcal{L}_{\text{fizik}} \quad (3.22)$$

Bu formülde  $\mathcal{L}_{\text{veri}}$ , gözlenen veriler ile modelin tahminleri arasındaki klasik ortalama karesel hata (MSE) terimini ifade etmektedir. Bu terim, aşağıdaki şekilde hesaplanmıştır:

$$\mathcal{L}_{\text{veri}} = \frac{1}{N} \sum_{i=1}^N \left( f_{\theta}(\mathbf{x}^{(i)}) - y^{(i)} \right)^2 \quad (3.23)$$

Burada  $N$  adet gözlem çifti  $(\mathbf{x}^{(i)}, y^{(i)})$  bulunmaktadır. Modelin klasik regresyon yapısından ayrıldığı temel nokta, fiziksel bilgilendirme terimidir. Bu çalışmada, fizik bilgilendirme doğrudan fiziksel sistemin açık bir diferansiyel denklemine dayanmasa da, sistem çıktılarının türevsel süreklilik (derivative smoothness) özelliği taşıması beklendiğinden, fizik bilgilendirmeli kayıp terimi bu türevsel düzgünlüğü sağlayacak şekilde tasarlanmıştır.

Fizik bilgilendirmeli kayıp, modelin çıktısının girdilere göre ikinci türevinin kare ortalamasını minimize etmektedir. Böylelikle, modelin öğrenmiş olduğu fonksiyonun hem sürekli hem de türevlerinin de sürekli olması sağlanmaktadır. Bu tür bir yaklaşım, fiziksel sistemlerin genellikle düzgün değişim göstermesi ve ani sıçramalardan kaçınması gerekliliğinden kaynaklanmaktadır. Kullanılan fiziksel kayıp (physics loss) terimi aşağıdaki şekilde ifade edilmektedir:

$$\mathcal{L}_{\text{fizik}} = \frac{1}{N} \sum_{i=1}^N \left\| \nabla^2 f_{\theta}(\mathbf{x}^{(i)}) \right\|^2 \quad (3.24)$$

Eğer giriş uzayı çok boyutluysa, bu ikinci türevler Hessian matrisi ile ifade edilmektedir ve ilgili kare norm aşağıdaki şekilde açılmaktadır:

$$\left\| \nabla^2 f_{\theta}(\mathbf{x}) \right\|^2 = \sum_{j=1}^d \sum_{k=1}^d \left( \frac{\partial^2 f_{\theta}}{\partial x_j \partial x_k}(\mathbf{x}) \right)^2 \quad (3.25)$$

Fakat bu çalışmada yalnızca yönsel ikinci türevler (diagonal elemanlar) göz önüne alınarak hesaplama yapılmıştır. Pratikte bu terim, ardışık türev hesaplamaları yoluyla PyTorch gibi otomatik türevleme destekleyen derin öğrenme kütüphaneleriyle uygulanabilmektedir. Uygulamada kullanılan yaklaşımda, önce model çıktısının birinci türevi alınmakta, ardından bu türev tekrar türevlenerek ikinci türev elde edilmekte ve sonuçta bu türevin karesi alınarak ortalama hesaplanmaktadır.

Bu fizik bilgilendirme terimi, modelin yalnızca verilere uymasını değil, aynı zamanda fiziksel olarak anlamlı, sürekli ve yumuşak (smooth) bir çıktı üretmesini de garanti altına almaktadır. Bu yönüyle fizik bilgilendirmeli kayıp, diferansiyel denklem bazlı açık kısıtlar bulunmasa dahi fiziksel yapının doğasına dair örtük ön kabulleri (örneğin süreklilik, düzgünlük, düşük frekanslı davranış) modelin öğrenme sürecine entegre etmektedir. Bu yaklaşım, özellikle az veriyle eğitilen modellerde genelleme başarımını anlamlı ölçüde artırabilmektedir [111].

Bu çalışmada kullanılan PINN yapısı, klasik yapay sinir ağı modellerinin esnekliğini fiziksel sistemlerin türevsel düzgünlük varsayımlarıyla birleştirerek hem veriye uygun, hem de fiziksel anlamlı tahminler elde etmeyi mümkün kılmıştır. Bu yapı, fiziksel modelleme ve makine öğrenmesi arasında köprü kuran çağdaş bir yöntem olarak öne çıkmaktadır.

### **3.9 Yapay Zeka Temelli Performans Tahminleme Modeli**

Ters ozmoz (RO) sistemlerinin performansı üzerinde belirleyici etkiye sahip en önemli bileşenlerden biri, Bölüm 2’de detaylarıyla anlatıldığı gibi membran modülü içerisindeki akış kanalının geometrisidir. Kanal boyunca akan akışkanın yönlendirilme şekli, oluşan türbülans düzeyi, sınır tabaka kalınlığı ve konsantrasyon kutuplanması gibi hidrodinamik etkiler doğrudan membran verimliliğini etkilemektedir. Düz geometrilerde bu etkiler daha öngörülebilir olsa da akışın istenilen düzeyde karışması ve kirlenmenin azaltılması gibi hedeflere ulaşmakta yetersiz kalabilmektedir. Bu nedenle, literatürde son yıllarda dalgalı ve kompleks yapılarla desteklenmiş kanal tasarımları ön plana çıkmış; bu yapıların kirlenmeyi azaltarak enerji verimliliğini artırabildiği gösterilmiştir. Geometrinin değişimiyle birlikte yalnızca akı değil, aynı zamanda membran ömrü ve işlem maliyeti gibi önemli parametrelerde de iyileşme sağlanabilmiştir. Bu gelişmeler, bu tez çalışmasının temelini oluşturmuştur.

Bununla birlikte, söz konusu geometrik tasarımların her biri için fiziksel prototip oluşturmak ya da hesaplamalı akışkanlar dinamiği (HAD) simülasyonlarıyla analiz gerçekleştirmek oldukça zaman alıcı ve maliyetlidir. Özellikle çok sayıda geometrik varyantın incelendiği geniş tasarım aralıklarında, klasik yöntemlerle ilerlemek pratik değildir. HAD simülasyonlarının doğruluğu yüksek olsa da binlerce tasarımı

değerlendirmek için gereken süre ve işlem gücü sınırlayıcıdır. Bu bağlamda, veri odaklı yaklaşımlar, özellikle de makine öğrenmesi (ML) tabanlı modeller, sistematik bir öngörü süreci sağlamak açısından büyük potansiyel sunmaktadır. Bu modeller, geçmişte simülasyon ya da deneylerle analiz edilmiş tasarımların verilerinden öğrenerek, yeni bir geometri için olası performansı tahmin edebilecek şekilde eğitilebilmektedir.

Regresyon analizleri, bu tür tahmin modellerinde sıklıkla kullanılan yöntemlerden biridir. Özellikle yapısal tasarıma ilişkin problemler söz konusu olduğunda, regresyon modelleri giriş verileriyle (örneğin, geometri noktaları) çıktı verileri (örneğin, performans metriği) arasındaki sürekli ilişkiyi öğrenme açısından uygundur. Bu tez çalışmasında uygulanan yöntem, kanal geometrisini tanımlayan spline kontrol noktalarının y-koordinatları ile elde edilen fiziksel çıktılar arasında bir matematiksel model kurmaya yöneliktir. Bu modelin öğrenme süreci, yalnızca geçmiş verilerdeki örüntüleri keşfetmekle kalmamakta; aynı zamanda yeni ve görülmemiş geometri tasarımları için öngöründe bulunarak tasarım sürecini yönlendirebilmektedir. Böylece, fiziksel doğrulama gereksinimi olmadan binlerce yeni kanal geometrisi hızlıca taranabilmektedir.

Veri temelli performans modeli geliştirmek için öncelikle kullanılan özneliklerin sistemin davranışını yeterli düzeyde temsil edebilmesi gerekmektedir. Bu çalışmada, her bir geometri üst ve alt olmak üzere toplam 42 spline noktasından oluşmuştur ve bu noktalar sayesinde kanalın yapısı belirlenmiştir. Simülasyon çıktıları ise permeat akısı, kirlenme oranı ve kanal içindeki basınç farkı gibi temel parametreleri içermiştir. Bu üç parametre birleştirilerek, 0 ile 100 aralığında ölçeklenmiş birleşik bir performans metriği elde edilmiştir. Modelin eğitilmesinde yalnızca geometrik veriler kullanılmış, performans çıktıları ise bu verilerin etiketleri olarak tanımlanmıştır. Böylelikle modelin tahmin gücü, yalnızca geometriye dayalı olarak değerlendirilmiştir.

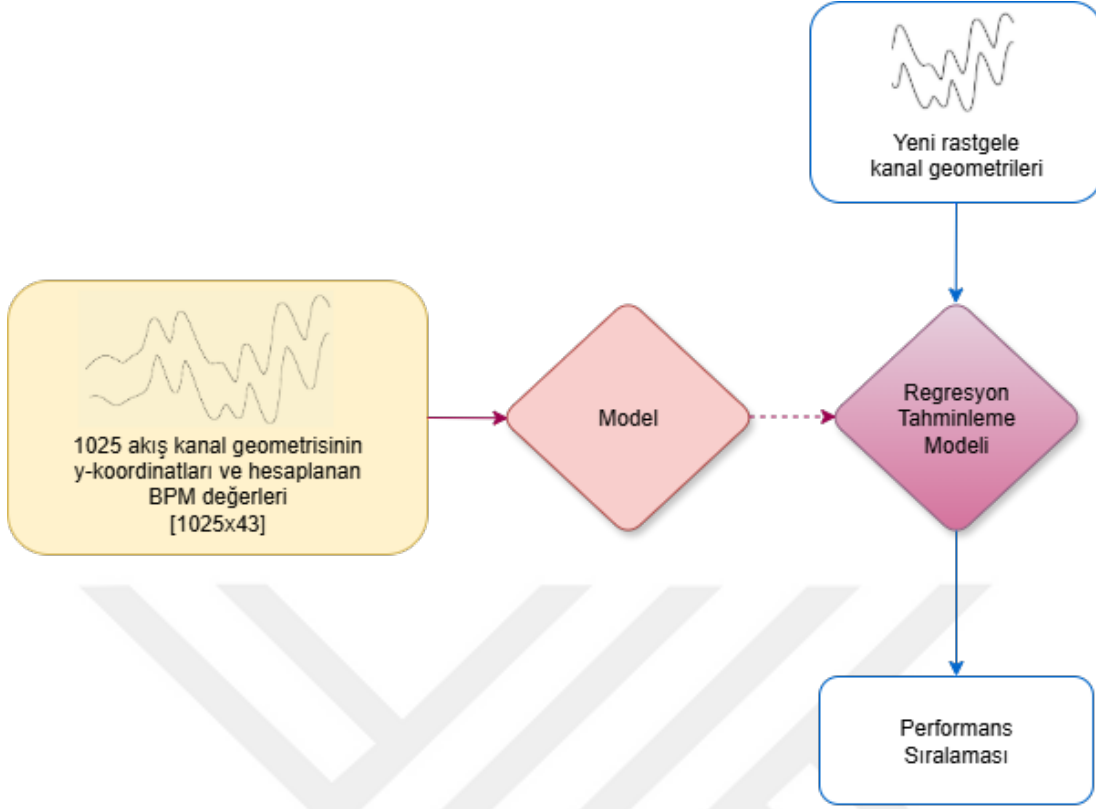
Bu yaklaşımla oluşturulan regresyon modeli sayesinde, yeni bir kanal tasarımı önerildiğinde, herhangi bir fiziksel deney ya da HAD simülasyonu yapılmadan bu tasarımın performans seviyesi tahmin edilebilmektedir. Bu, özellikle optimizasyon ve tasarım taraması senaryoları için son derece avantajlı bir durumdur. Örneğin, tasarım

uzayından rastgele oluşturulan binlerce geometri arasından yalnızca yüksek performans potansiyeli taşıyanlar seçilebilecektir. Böylece, simülasyon veya üretim için kaynaklar daha verimli kullanılabilir. Ayrıca, regresyon modeli optimizasyon algoritmaları (örneğin genetik algoritmalar) ile entegre edilerek çok kriterli tasarım süreçlerinde de etkin bir şekilde kullanılabilir hale getirilebilir.

Bu çalışmada önerilen yaklaşım, yalnızca bugünkü ihtiyaçlara değil, aynı zamanda gelecekteki sistem gereksinimlerine de uyarlanabilir yapıdadır. Bu bölümde anlatılan teorik çerçeve, RO membran modüllerinde kanal geometrisi tasarımının veri odaklı yöntemlerle nasıl optimize edilebileceğine ışık tutmaktadır. Regresyon temelli tahmin modelleri, klasik simülasyon süreçlerine güçlü bir alternatif sunarak tasarım döngüsünü hızlandırmakta ve yeni nesil membran sistemlerinin geliştirilmesine katkı sağlamaktadır. Bu metodoloji, yalnızca RO sistemleri için değil, geniş bir uygulama alanı bulunan tüm membran teknolojileri için referans alınabilecek nitelikte bir yöntemsel altyapı sunmaktadır.

Bu çalışmanın ikinci fazında, hesaplamalı akışkanlar dinamiği (HAD) analizleri sonucunda elde edilen performans verileri kullanılarak membran akış kanal geometrilerinin regresyon temelli ML yöntemleriyle modellenmesi üzerine çalışılmıştır. Regresyon tahminleme modelinin akış diyagramı Şekil 3.3'de gösterilmiştir. Bu çalışmada, 1025 akış kanal geometrisinin y-koordinatları regresyon tahminleme modeline girdi olarak verilip model eğitilmiştir. Nihai elde edilen regresyon tahminleme modelinin, yeni oluşturulacak rastgele akış kanal geometrilerinin potansiyel performanslarını tahmin etmesi amaçlanmaktadır.

Oluşturulan bu model ile tasarlanacak olan akış kanal geometrilerinin potansiyel performansı, HAD simülasyonuna ihtiyaç duyulmadan hesaplanabilmektedir. Böylelikle, HAD analizlerinin zaman ve işlem maliyeti azaltılırken, tasarım süreci önemli ölçüde hızlandırılmaktadır. Bu AI destekli yaklaşım, gelecekteki membran tasarımlarında etkili kanal geometrilerini hızlıca taramak ve önermek için bir temel oluşturmaktadır.



**Şekil 3.3:** Regresyon temelli performans tahminleme modelinin akış diyagramı.

## 4. DENEYSEL ÇALIŞMALAR

### 4.1 Regresyon Deneyleri

Bu bölümde, rastgele oluşturulan akış kanalı geometrilerine ait performans çıktılarının tahmini amacıyla gerçekleştirilen regresyon modelleme süreci detaylı olarak ele alınmıştır. Sırasıyla, veri setinin nasıl oluşturulduğu, oluşturulan veri seti ile modelin nasıl eğitildiği ve test edildiği, modelin nasıl uygulandığı ve sonuç bölümlerine yer verilmiştir.

Kanal geometrileri için öncelikle 1045 adet örnek rastgele oluşturulmuş olup, ağ oluşturma ve sayısal kararlılık amacıyla içlerinden 1025 adet örnek seçilerek veri seti elde edilmiştir. Seçilen bu geometriler, hesaplamalı akışkanlar dinamiği (HAD) simülasyonlarına uygun hale getirilmiştir. Her bir geometri için permeat akısı, basınç farkı ve kirlenme miktarı gibi sistem performansını etkileyen kritik çıktılar elde edilmiş ve bu değerler kullanılarak denklem 2.9'daki formül ile birleşik performans metriği (BPM) elde edilmiştir ve regresyon çalışmaları gerçekleştirilmiştir.

Modelleme sürecinde, yalnızca kanal geometrisini tanımlayan spline noktalarının y-koordinatları girdi olarak kullanılmış, hedef değişken ise normalize edilmiş BPM olmuştur. Eğitim ve test ayrımı yapılarak, farklı regresyon modellerinin karşılaştırılması sağlanmış, en iyi performansı gösteren model hiperparametre optimizasyonu ile güçlendirilmiştir. Elde edilen modellerin başarımı; ortalama mutlak hata (MAE), ortalama kare hata (MSE), karekök ortalama kare hata (RMSE), belirlilik katsayısı ( $R^2$ ), karekök ortalama logaritmik kare hata (RMSLE) ve ortalama mutlak yüzde hata (MAPE) gibi istatistiksel metrikler aracılığıyla değerlendirilmiş, sonuçlar grafik ve tablolarla desteklenerek sunulmuştur. Bu kapsamda yapılan çalışmalar, yapay zeka destekli geometri-performans tahminleme sürecinin etkinliğini ortaya koymakta ve gelecekteki tasarımlar için hızlı tahminleme altyapısı oluşturmaktadır.

#### 4.1.1 Veri

Bu çalışmada regresyon modellerinin eğitimi amacıyla kullanılan veri seti, toplam 1045 adet rastgele oluşturulmuş akış kanalı geometrisine dayanmaktadır. Kanal geometrileri, Python ortamında geliştirilen özel bir script aracılığıyla, her biri 42 kontrol noktası içeren (üst ve alt duvarlar için 21'er adet) spline noktaları kullanılarak tanımlanmıştır. Bu noktalar B-spline interpolasyon yöntemi ile birleştirilerek her bir örneğe özgü akış kanal geometrisi oluşturulmuştur.

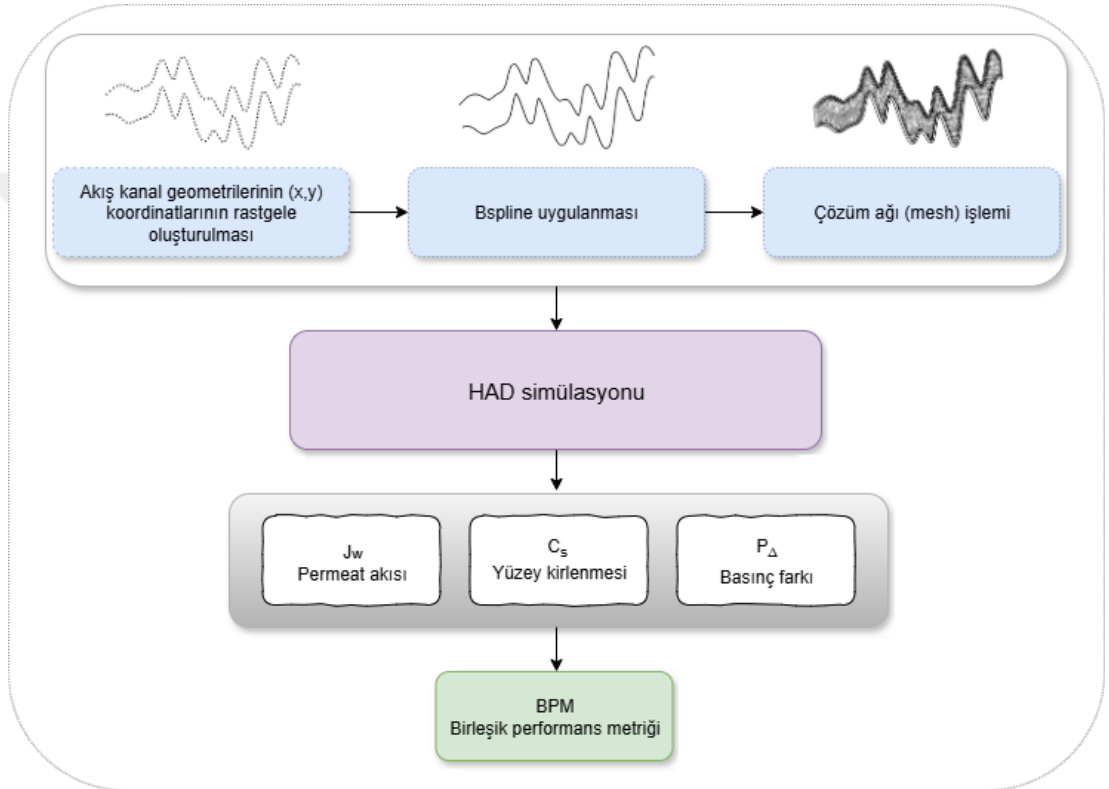
Elde edilen bu parametrik geometriler, sayısal çözümleme için uygun ağ (mesh) yapısına dönüştürülmüş ve sonlu hacim yöntemine dayalı olarak yapılandırılan hesaplamalı akışkanlar dinamiği (HAD) simülasyonlarına hazır hale getirilmiştir. Her bir kanal geometrisi için 100 saniyelik simülasyon süresi boyunca analiz gerçekleştirilmiştir. Bu sürecin sonunda, kanal performansını değerlendirebilmek amacıyla üç temel çıktı kaydedilmiştir: (i) Permeat akısı (flux), (ii) kirlenme miktarı (fouling index) ve (iii) kanal giriş-çıkış noktaları arasındaki basınç farkı (pressure drop). Bu parametreler, her geometriye özgü performans karakterizasyonu sağlamak üzere sistematik biçimde çıkarılmıştır.

Simülasyon çıktıları, regresyon modellerinin hedef değişkenini oluşturmak üzere programatik olarak ayrıştırılmış ve okunmuştur. Elde edilen bu veriler, her bir geometrik örneğe ait kontrol noktaları ile birlikte tek bir veri seti içerisinde yapılandırılmıştır. Her satır, bir kanal geometrisine karşılık gelecek şekilde, geometriyi tanımlayan spline noktalarını ve ona karşılık gelen üç fiziksel çıktıyı içerecek biçimde birleştirilmiştir.

Simülasyon aşaması ve çıktıların otomatik olarak işlenmesi sırasında çeşitli sayısal ya da fiziksel nedenlerle (örneğin mesh başarısızlıkları, sınır koşulu sorunları veya hesaplamada kararsızlık) doğru biçimde yapılandırılmayan örnekler elenmiştir. Bu ön işleme sonucunda toplamda 20 adet geçersiz kayıt veri setinden çıkarılmış ve nihai olarak 1025 geçerli örnek ile regresyon deneyleri gerçekleştirilmiştir.

Veri setinde her bir geometri için, kanalın sistem üzerindeki genel etkisini temsil edebilecek şekilde oluşturulan bir BPM tanımlanmıştır. Bu metrik, denklem 2.9'da sunulan formülasyona göre hesaplanmış olup; normalize edilmiş permeat akısı, basınç

farkı ve kirlenme indeks değerlerinin ağırlıklı kombinasyonu ile elde edilmiştir. Her bir parametre için 100. saniyede, simülasyon çıktılarının ortalama değeri alınarak sistemin kararlı hal performansı esas alınmıştır. Bu birleşik metrik BPM, regresyon modellerinin tahmin etmeyi amaçladığı etiket (target) olarak kullanılmıştır. Bu etiket, bir sonraki bölüm olan eğitim ve test sürecinde,  $y_i$  sembolü ile gösterilmiş olup  $i$ . gözleme ait gerçekleşmiş performans metriği olarak ifade edilmiştir. Veri setinin hazırlanması ve BPM'nin elde edilmesi Şekil 4.1'de görselleştirilmiştir.



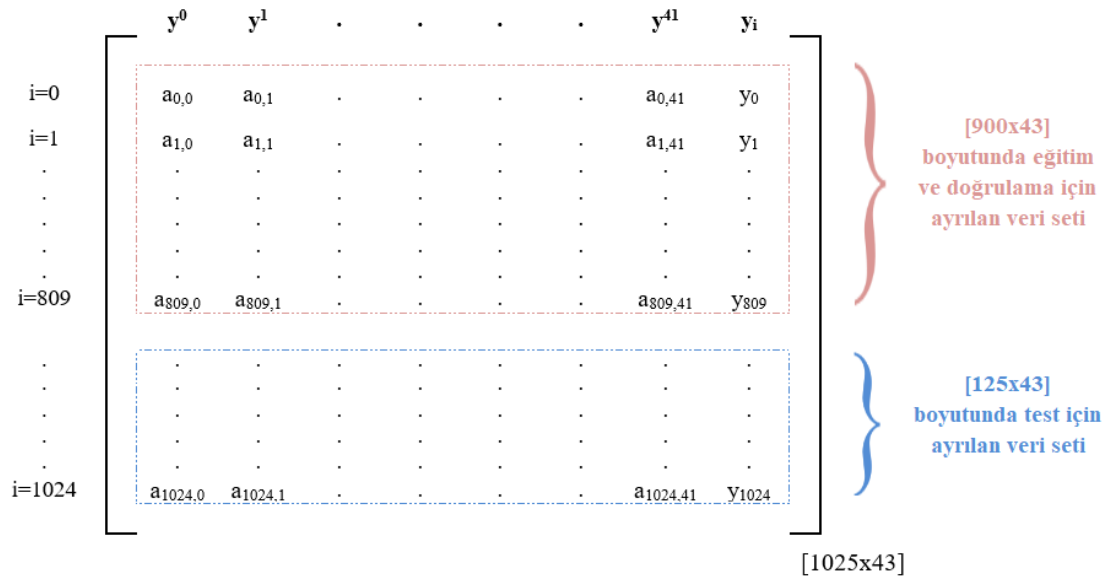
**Şekil 4.1:** Veri hazırlığı ve birleşik performans metriğinin (BPM) elde edilmesi.

#### 4.1.2 Eğitim ve test

Bu çalışmada, regresyon modellemesinin girdisi olarak yalnızca kanal geometrisine ait B-spline kontrol noktalarının y-eksenine ait koordinatları kullanılmıştır. Her bir geometri 42 noktadan oluşmakta olup, bu noktaların y-koordinatları doğrudan modelin açıklayıcı değişkenlerini (features) oluşturmaktadır. Modelin tahmin etmeye çalıştığı hedef değişken (etiket) ise daha önce tanımlanmış olan ve kanal geometrisinin fiziksel performansını temsil eden birleşik performans metriğidir.

Söz konusu performans metriği, karşılaştırılabilirlik ve ölçek bağımsızlık sağlamak amacıyla 0–100 aralığında normalize edilmiştir. Böylece, hem modelin öğrenme sürecindeki stabilitesi artırılmış hem de çıktılarının anlamlılık düzeyi güçlendirilmiştir.

Toplamda 1025 adet örnekten oluşan veri seti, eğitim ve test aşamaları için ayrıştırılmıştır. Bu kapsamda, 900 örnek eğitim ve doğrulama (validation) işlemleri için, geri kalan 125 örnek ise test kümesi olarak kullanılmıştır. Ayrılan test kümesi, modeli gerçek dünya koşullarında değerlendirmek amacıyla daha önce modelin hiçbir aşamasında kullanılmamış olup sınav seti yöntemine (holdout method) göre yapılandırılmıştır. İlgili veri seti, Şekil 4.2’de gösterilmiştir. Veri setinin son sütunu, HAD simülasyonları sonucu denklem 2.9 kullanılarak hesaplanan gerçek BPM değerlerini; diğer sütunlar ise  $y^r$ ,  $r = \{0, \dots, 41\}$  olmak üzere, akış kanal geometrisinin alt ve üst kenarlarına ait y-eksen koordinatlarını göstermektedir.



Şekil 4.2: Tüm veri setinin oluşturulması.

Yapay zeka modelinin eğitim aşamasında 10 katlı çapraz doğrulama (10-fold cross-validation) yöntemi uygulanarak her bir model farklı veri bölütleri üzerinde eğitilmiş ve doğrulanmıştır. Bu yöntemin uygulanması için öncelikle eğitim ve doğrulama için kullanılmak üzere ayrılan 900 örnekle veri seti, 10 eşit alt parçaya ayrılmıştır. Her alt parça, 90 farklı örnek içermekte olup, her adımda, bu alt kümelerden bir tanesi test için, geri kalan dokuz tanesi (toplamda 810 örnek) eğitim için kullanılmıştır. Her adımda kullanılan örnekler, Şekil 4.3’te gösterildiği gibi sıralı

şekilde kullanılmış ve ilk oluşturulan alt parçalar korunmuştur. Böylece, model 10 kez kurularak sınanmış olup modelin eğitimi tamamlanmıştır. Bu strateji, modelin genelleme kabiliyetini daha güvenilir şekilde değerlendirmeyi mümkün kılmıştır. Çapraz doğrulama sürecinde yirmi bir farklı regresyon algoritması denenmiş, her bir modelin doğrulama metrikleri kaydedilmiştir. Test sonuçlarının ortalaması olan çapraz doğrulama başarımı Çizelge 4.1’de raporlanmıştır. En iyi performansı gösteren model belirlenerek (CatBoost), hiperparametre optimizasyonu (tuning) gerçekleştirilmiştir.



**Şekil 4.3:** 10 katlı çapraz doğrulama yöntemi.

Optimizasyonu tamamlanan model, 900 örnekten oluşan tüm eğitim verisi üzerinde yeniden eğitilmiş ve böylece nihai model elde edilmiştir. Bu model, yalnızca test kümesinde bulunan 125 örnek üzerinde sınanmış ve modelin gerçek veri üzerindeki tahmin başarımı değerlendirilmiştir.

Henüz gerçekleştirilmemiş olmakla birlikte, ilerleyen süreçte elde edilecek yeni geometri verilerinde kullanılmak üzere, tüm 1025 örnek üzerinden optimum parametrelerle eğitilmiş final modelin oluşturulması planlanmaktadır. Bu model, genişletilmiş veri seti ile daha sağlam tahmin yeteneğine sahip bir yapay zeka temelli analiz aracı işlevi görecektir.

Regresyon modelinin tahmin performansını değerlendirmek için aşağıdaki hata metrikleri kullanılmıştır.

- *Ortalama mutlak hata (MAE):* Tahmin hatalarının mutlak değerlerinin ortalamasıdır. Hedef değişkenle aynı birime sahiptir ve yorumlaması kolaydır. Büyük hataları MSE

ve RMSE'ye göre daha az ağır cezalandırdığı için uç değerlere karşı daha dayanıklıdır.

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (4.1)$$

- *Ortalama kare hata (MSE)*: Hataların karesinin ortalamasıdır. Büyük hataları karesel olarak daha fazla cezalandırmaktadır ve bu yüzden aykırı değerlere duyarlıdır. Türevlenebilirliği nedeniyle eğitimde yaygın kullanılmaktadır; fakat birimi hedefin karesi olduğu için doğrudan yorumlamak güçtür.

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (4.2)$$

- *Karekök ortalama kare hata (RMSE)*: MSE'nin karekökü olduğundan hedefle aynı birimde ifade edilebilmektedir ve MSE'ye göre yorumlanması kolaylaşmaktadır. Büyük hataları güçlü biçimde vurgulamaktadır. Büyük sapmaların özellikle pahalı olduğu uygulamalarda tercih edilmektedir.

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (4.3)$$

- *Belirlilik (determinasyon) katsayısı ( $R^2$ )*: Modelin, hedef değişkendeki toplam varyansın ne kadarını açıkladığını göstermektedir.  $R^2$  değeri, 1'e yaklaştıkça uyum iyidir; 0, modelin sadece ortalama tahmini kadar iyi olduğunu; negatif değerler ise ortalamadan daha kötü performansı olduğunu ifade etmektedir. Varyansı çok düşük veri kümelerinde yanıltıcı olabilmektedir; çoklu açıklayıcı modellerde ayarlı  $R^2$  kullanımı daha uygundur.

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (4.4)$$

- *Karekök ortalama logaritmik kare hata (RMSLE)*: Hataları logaritmik ölçekte değerlendirmektedir; bu nedenle farkların mutlak büyüklüğünden çok oransal (nispi) farklılıklara duyarlıdır. Denklem 4.5'teki şekilde ifade edilmektedir. Hedef değerini geniş aralıkta değiştiği veya ağır kuyruklu (heavy tailed) dağılımlarda, büyük mutlak hataların etkisini yumuşatmaktadır. ( $y_i \geq 0$ ) ve ( $\hat{y}_i \geq 0$ ) koşulu gerekmektedir; ( $\log(1+x)$ ) kullanımı sayesinde sıfırlar tolere edilmektedir. Küçük ve orta düzey göreceli hatalara odaklanmak istendiğinde uygundur.

$$\text{RMSLE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (\log(1 + \hat{y}_i) - \log(1 + y_i))^2} \quad (4.5)$$

- *Ortalama mutlak yüzde hata (MAPE)*: Hataları yüzde olarak ifade etmektedir, ölçekten bağımsız ve anlatması kolaydır. Ancak  $y_i = 0$  için tanımsızdır ve küçük gerçek değerlerde hata payını orantısız büyütürük yanıltıcı olabilmektedir.

$$\text{MAPE} = \frac{100}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (4.6)$$

*Semboller:*

- $n$  : Tahmin edilen örnek sayısı (veri setindeki toplam gözlem sayısı)
- $y_i$  :  $i$ . gözleme ait gerçek (gerçekleşmiş) performans metriği değeri
- $\hat{y}_i$  :  $i$ . gözleme ait model tarafından tahmin edilen performans metriği değeri
- $\bar{y}$  : Tüm gerçek değerlerin ortalaması

Bu metriklerden özellikle MAPE değeri, modelin genel başarımını yüzde cinsinden ifade etmesi açısından önemli bulunmuş ve en iyi modeli belirlemede temel kriter olarak kullanılmıştır. Model hiperparametreleri, MAPE minimizasyonu hedeflenerek optimize edilmiş ve böylece tahmin doğruluğu en üst seviyeye çıkarılmıştır.

## 4.2 Modelin Uygulanması

Bu çalışmanın yapay zeka bileşeni kapsamında regresyon modelleme süreci, Python programlama dili kullanılarak gerçekleştirilmiştir. Öncelikle, kanal geometrilerini tanımlayan rastgele nokta setlerinin üretimi yine Python ortamında geliştirilmiş özel bir algoritma ile sağlanmıştır. Bu noktalar, üst ve alt kanal duvarlarını tanımlayan spline eğrileri aracılığıyla birleştirilmiş ve her bir geometri fiziksel olarak anlamlı bir akış kanalı yapısına dönüştürülmüştür.

Akışkanlar dinamiği yönünden değerlendirmek amacıyla, bu geometrilere Bölüm 2'de detayları ile açıklanan hesaplamalı akışkanlar dinamiği (HAD) simülasyon adımları uygulanmış ve her bir örnek için sistem performansını temsil eden üç temel çıktı (akış hızı, basınç farkı ve kirlenme oranı) elde edilmiştir. Bu çıktıların normalize

edilerek ağırlıklandırılmasıyla birleşik performans metriği hesaplanmış ve regresyon modellerine hedef değişken olarak atanmıştır.

Veri seti oluşturulduktan sonra, Python ortamında, aynı eğitim ve test verilerini kullanarak farklı regresyon algoritmalarının karşılaştırmalı analizi yapılmıştır. Bu kapsamda, modelleme sürecini otomatikleştirmek ve değerlendirme metriklerini karşılaştırmak amacıyla yirmi bir çeşit regresyon modeli denenmiş, eğitim ve doğrulama aşamaları sistematik şekilde yürütülmüştür. Nihai olarak, en iyi genel başarıma sahip model belirlenmiş ve hiperparametre optimizasyonuna tabi tutularak tahmin gücü artırılmıştır. Denenen modeller ve sonuçları sonraki bölümlerde açıklanmıştır.

### **4.3 Regresyon Performans Sonuçları**

Bu çalışmada uygulanan regresyon modelleme süreci sonucunda, kullanılan 21 çeşit farklı algoritma arasında yapılan karşılaştırmalarda tüm değerlendirme metriklerine göre en iyi performansı kategorik artırma (CatBoost) regresyon modeli göstermiştir (bkz. Çizelge 4.1). Eğitim verisi olarak seçilen 900 örnek üzerinde gerçekleştirilen 10 katlı çapraz doğrulama (10-fold cross-validation) sonucu, en başarılı model için ortalama mutlak yüzde hata (MAPE) değeri %12,28 olarak elde edilmiştir. Bu oran, normalize edilerek 0–100 aralığına ölçeklenmiş olan birleşik akış kanalı performans metriği açısından değerlendirildiğinde, modelin tahminleme başarımının oldukça tatmin edici düzeyde olduğu gösterilmiştir. Tahmin edilen ve gerçek performans değerleri arasındaki farklara dayalı olarak hesaplanan ortalama mutlak hata (MAE) değeri 6,85 olarak ölçülmüş ve modelin ortalama mutlak tahminleme hatasının sınırlı düzeyde kaldığı gözlemlenmiştir.

Modelin eğitim sürecinin başında ayrılan 125 örnekten oluşan test kümesi, nihai modelin gerçek veriler üzerindeki başarımını değerlendirmek için kullanılmıştır. Çapraz doğrulama (900 örnek) sonunda en iyi aday olarak belirlenen CatBoost modeli, bağımsız test kümesinde (125 örnek) hiperparametre ayarı (tuning) yapılarak test edilmiştir ve Çizelge 4.2’de gösterildiği gibi performans değerlerinin iyileştiği görülmüştür. Bu parametre ayarlamaları, modelin hem öğrenme hem de genelleme kapasitesini artırmaya yönelik olarak sistematik biçimde gerçekleştirilmiştir. İlgili

çizelge incelendiğinde, MAPE değerinin %8,8'e ve MAE değerinin 5,66 seviyesine iyileştiği gözlemlenmiştir. Doğrulama evresindeki  $R^2$  düzeyiyle ( $R^2 = 0.602$ ) yakın bir performansın ( $R^2 = 0.609$ ) testte de korunması, modelin kalibrasyon ve genelleme açısından tutarlı olduğunu göstermiştir. Elde edilen sonuçların, çapraz doğrulama sırasında elde edilen değerlerle yüksek oranda uyumlu olması, modelin genelleme kabiliyetinin tatmin edici düzeyde olduğunu ortaya koymuştur. Bu hata ölçekleri, rastgele oluşturulmuş akış kanalı geometrilerinde performansın mühendislik kararlarını yönlendirecek çözünürlükte ve güvenilirlikte kestirilebildiğine işaret etmektedir.

**Çizelge 4.1:** Regresyon modellerinin eğitim veri seti üzerinde 10 katlı çapraz doğrulama sonucu performanslarının karşılaştırılması.

Kod	Regresyon Modelleri	MAE	MSE	RMSE	$R^2$	RMSLE	MAPE
<i>CatBoost</i>	Kategorik Artırma	6.8451	78.5152	8.7920	0.6018	0.1861	0.1228
<i>LightGBM</i>	Hafif Gradyan Artırmalı Regresyon Makinesi	7.1805	87.1916	9.2619	0.5579	0.1897	0.1263
<i>PINN</i>	Fizikle Öğrenen Yapay Sinir Ağı	7.3839	93.4555	9.5927	0.5201	0.1913	0.1274
<i>GBR</i>	Gradyan Artırmalı Regresyon	7.8517	101.1262	10.0063	0.4809	0.1994	0.1367
<i>XGBoost</i>	Aşırı Gradyan Artırma	8.0360	112.6940	10.5537	0.4236	0.2100	0.1421
<i>RF</i>	Rastgele Orman Regresyonu	8.5442	122.2292	10.9830	0.3778	0.2171	0.1506
<i>ETR</i>	Ekstra Ağaçlar Regresyonu	8.6110	123.3706	11.0507	0.3693	0.2186	0.1518
<i>AdaBoost</i>	Uyarlanabilir Yükseltme Regresyonu	9.3669	144.2620	11.9287	0.2679	0.2292	0.1638
<i>LR</i>	Lineer Regresyon	10.4662	179.7948	13.3259	0.0854	0.2489	0.1831
<i>Ridge</i>	Ridge Regresyonu	10.4662	179.7927	13.3258	0.0854	0.2489	0.1831
<i>LAR</i>	En Küçük Açık Regresyonu	10.4662	179.7948	13.3259	0.0854	0.2489	0.1831
<i>Huber</i>	Huber Regresyonu	10.4106	180.5628	13.3532	0.0818	0.2501	0.1846
<i>EN</i>	Elastik Net	10.5648	181.8028	13.4022	0.0765	0.2503	0.1853
<i>Lasso</i>	Lasso Regresyonu	10.6721	184.8486	13.5137	0.0616	0.2518	0.1872
<i>LLAR</i>	Lasso–En Küçük Açık Regresyonu	10.6721	184.8474	13.5137	0.0616	0.2518	0.1872
<i>BR</i>	Bayesyen Ridge Regresyonu	10.8865	189.9970	13.7078	0.0342	0.2549	0.1909
<i>OMP</i>	Ortogonal Eşleşmeli Arama Regresyonu	10.9525	193.9900	13.8294	0.0194	0.2563	0.1926
<i>Dummy</i>	Kukla Değişkenli Regresyon	11.2027	198.1141	14.0007	-0.0079	0.2592	0.1969
<i>KNN</i>	K-En Yakın Komşu Regresyonu	10.9888	215.6505	14.5917	-0.0998	0.2714	0.2087
<i>DT</i>	Karar Ağacı Regresyonu	12.6821	266.1588	16.2882	-0.3942	0.3035	0.2123
<i>PAR</i>	Pasif Agresif Regresyon	14.0386	322.8880	17.7160	-0.6364	0.3112	0.2313

**Çizelge 4.2:** Eğitim veri seti üzerinde en iyi performansı gösteren CatBoost regresyon modelinin, test verisi üzerinde hiperparametre ayarı sonrası elde edilen performans ölçütleri.

Kod	Regresyon Modeli	MAE	MSE	RMSE	R <sup>2</sup>	RMSLE	MAPE
CatBoost	Kategorik Artırma	5.6551	51.6362	7.1858	0.6094	0.1135	0.0879

#### 4.4 Bulgular ve Tartışma

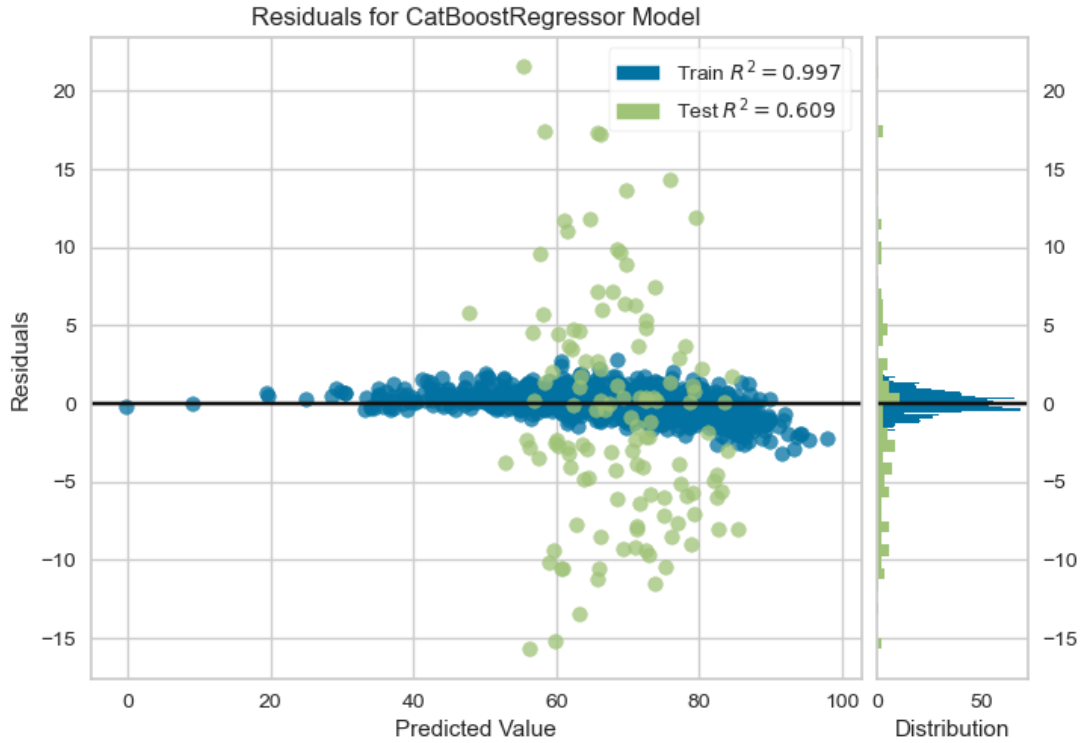
Modelin tahminleme hatalarının dağılımı Şekil 4.4'te gösterilmiştir. Bu görsel, modelin önyargısız ve dengeli tahmin yapma kabiliyetini göstermesi açısından anlamlıdır. Bu grafikte, yatay eksen model tahminlerini  $\hat{y}$  (predicted), dikey eksen ise artık değerini  $e = y - \hat{y}$  (residual) göstermektedir. Her nokta, test kümesindeki tek bir örneğe karşılık gelmektedir. Diyagramda  $e = 0$  doğrusu (yatay eksen) hatasız çizgidir; noktaların bu çizgi etrafında simetrik ve yoğun biçimde toplanması, modelin sistematik yanlılığının (bias) sınırlı olduğunu göstermektedir.

Nokta bulutu incelendiğinde, artıkların ortalama etrafında genel olarak homojen bir bantta dağıldığı, ancak öngörülen değer  $\hat{y}$  yükseldikçe yayılımın ılımlı biçimde arttığı gözlenmiştir. Bu desen, hata varyansının hedef düzeyiyle birlikte artması işaretidir ve akış kanalı geometrilerinde beklenen fiziksel çeşitlilikle uyumludur. Yüksek performans bölgelerinde (örneğin, daha karmaşık akış desenleri veya türbülans eşiği gibi) belirsizlik artmaktadır. CatBoost'un parçalı sabit ağaç tabanlı karar bölgeleri ortalama hatayı düşürmekte etkilidir; kalan varyans artışı ise veri üretim sürecinin doğal oynaklığına karşılık gelmektedir.

Grafiğin yanında ve üstünde verilen dağılım özeti (histogram ve yoğunluk) artıkların merkezinin yaklaşık sıfır olduğunu ve dağılımın ılımlı simetri sergilediğini teyit etmiştir. Ağır kuyruklar veya belirgin sağ ve sol çarpıklık görünmemesi, modelin belirli bir aralıkta sistematik olarak aşırı ve eksik tahmin yapmadığını göstermiştir. Kimlik doğrusu etrafından çok uzak tekil noktalar varsa, bunlar aykırı geometrilere veya nadir akış motiflerine işaret etmektedir ve mühendislik açıdan bu tasarımların ayrıca incelenmesi yararlı olmaktadır.

Yöntemsel olarak değerlendirildiğinde, modelin kalibrasyonu olumlu değerlendirilmiştir: artıkların ortalaması yaklaşık sıfırdır ve kalanlara ilişkin belirgin, siste-

matik bir desen gözlenmemiştir. Hata varyansı, özellikle yüksek tahmin değerleri ( $\hat{y}$ ) bölgesinde artış eğilimi göstermiştir. Bu durum, veri dağılımının uçlarında örneklem büyüklüğünün artırılması (ek HAD senaryolarıyla) ya da hedef dönüşümü ve kuantil kaybı gibi varyans-duyarlı yaklaşımların kullanılmasıyla iyileştirilebileceği sonucunu doğurmuştur. Öte yandan, model aleyhine anlamlı bir eğrilik (nonlineer kalıntı deseni) saptanmadığından, CatBoost'un yerel doğrusal olmayanlıkları yakalama kapasitesi teyit edilmiştir. Sonuç olarak, artıklar grafiği test verisi üzerindeki genellebilirlik bulgularıyla tutarlı bir hata yapısı sunmuş; kalan iyileştirme alanı esasen uç değer kalibrasyonu ve varyansın dengelenmesi üzerinde yoğunlaşmıştır.



**Şekil 4.4:** En iyi performans gösteren CatBoost modelinin hata dağılımı.

Modelin tahminleme hata (prediction error) değerlendirmesi Şekil 4.5'te verilmiştir. Bu görsel, CatBoost modelinin kalibrasyonunu ve genelleme doğruluğunu ortaya koymuştur. Bu grafikte, yatay eksen hedef değişkenin gerçek değerlerini  $y$  (actual), dikey eksen ise modelin tahminlerini  $\hat{y}$  (predicted) göstermektedir; her nokta test kümesindeki tek bir gözleme karşılık gelmektedir. Diyagonal kimlik doğrusu  $\hat{y} = y$  kusursuz kalibrasyonu temsil ederken; grafikte ayrıca çizilen en iyi uyum doğrusu (best-fit)  $\hat{y}$  ile  $y$  arasındaki doğrusal ilişkiyi özetlemektedir.

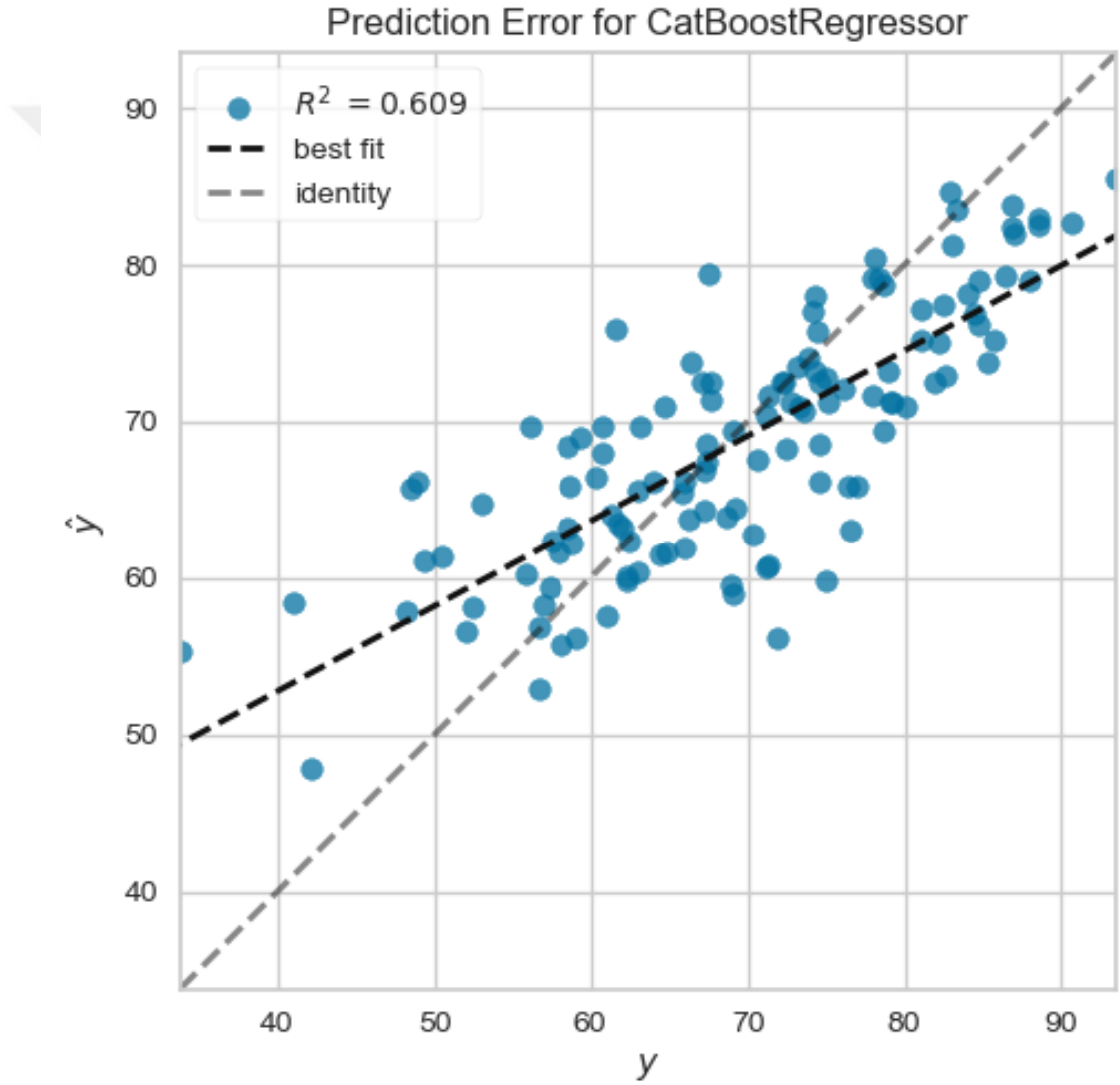
Nokta bulutu, kimlik doğrusu etrafında sıkı bir kümelenme sergileyerek modelin genel eğilimi doğru yakaladığını göstermiştir. Uyum doğrusunun kimlik doğrusuna yakın ama bir miktar daha düşük eğimli görünmesi, gradyan artırmalı ağaçlarda sık gözlenen uç değerleri sıkıştırma davranışına işaret etmektedir. Yüksek  $y$  değerleri kısmen eksik tahmin edilmektedir ( $\hat{y} < y$ ), çok düşük  $y$  değerleri ise kısmen aşırı tahmin edilmektedir ( $\hat{y} > y$ ). Kimlik doğrusuna dikey uzaklık artık (residual) büyüklüğünü vermektedir:  $e_i = y_i - \hat{y}_i$ . Noktaların bu doğru etrafında geniş bir huni yerine görece homojen bir bantta kalması, hataların yayılımının büyük ölçüde istikrarlı olduğunu ve sistematik bir sapmanın sınırlı kaldığını düşündürmektedir.

Bu sonuçların, bağımsız test kümesi (125 gözlem) üstünde elde edilen nicel sonuçlarla (bkz. Çizelge 4.2) tutarlı olduğu gözlemlenmiştir. Çizelge 4.2’de verilen bu değerler, rastgele tasarlanan akış kanallarında performansın mühendislik kararlarını yönlendirecek çözünürlükte kestirilebildiğini göstermiştir. Model, genel seviyeyi güvenilir biçimde yakalarken, yalnızca uç bölgelerde ılımlı bir sıkıştırma eğilimi sergilemektedir. Tezin hedefi olan rastgele oluşturulan geometri tasarımlarının performansını AI ile öngörmek açısından bu, iki önemli sonucu beraberinde getirmiştir: (i) model kalibre ve genellenebilir bir çizgidedir; (ii) daha iyi uç bölge kalibrasyonu için, veri dağılımının uçlarında örnek zenginleştirme (ek HAD senaryoları) veya uygun bir hedef dönüşümü ve kuantil kayıp seçimi değerlendirilirse, grafikteki uyum doğrusu kimlik doğrusuna daha da yaklaşacaktır.

Bu tahminleme hata grafiği, CatBoost’un test verisi üzerinde doğru kalibrasyon ve makul hata düzeyi sunduğunu, hatanın yapısının ise tasarım uzayındaki doğal değişkenlikle uyumlu olduğunu göstermiştir. Bu nedenle modelin, yeni kanal geometrilerinin hızlı sıralanması ve ön seçimi için pratik bir öngörü motoru olarak kullanılabilmesi; kalan iyileştirmelerin, özellikle uç değer kalibrasyonuna odaklanacak şekilde planlanması gerektiği yorumu yapılmıştır.

Çapraz doğrulamalı yinelemeli özellik eleme (RFECV), bir sarmalayıcı (wrapper) yöntem olarak, seçili tahmincinin (CatBoost) özellik çıkarımına dayalı skorunu yinelemeli eleme adımlarıyla çapraz doğrulama (CV) üzerinden optimize etmektedir. Şekil 4.6’da gözlenen plato ( $n \approx 38$  özelliğin ötesinde sınırlı kazanç), modelin bilgi

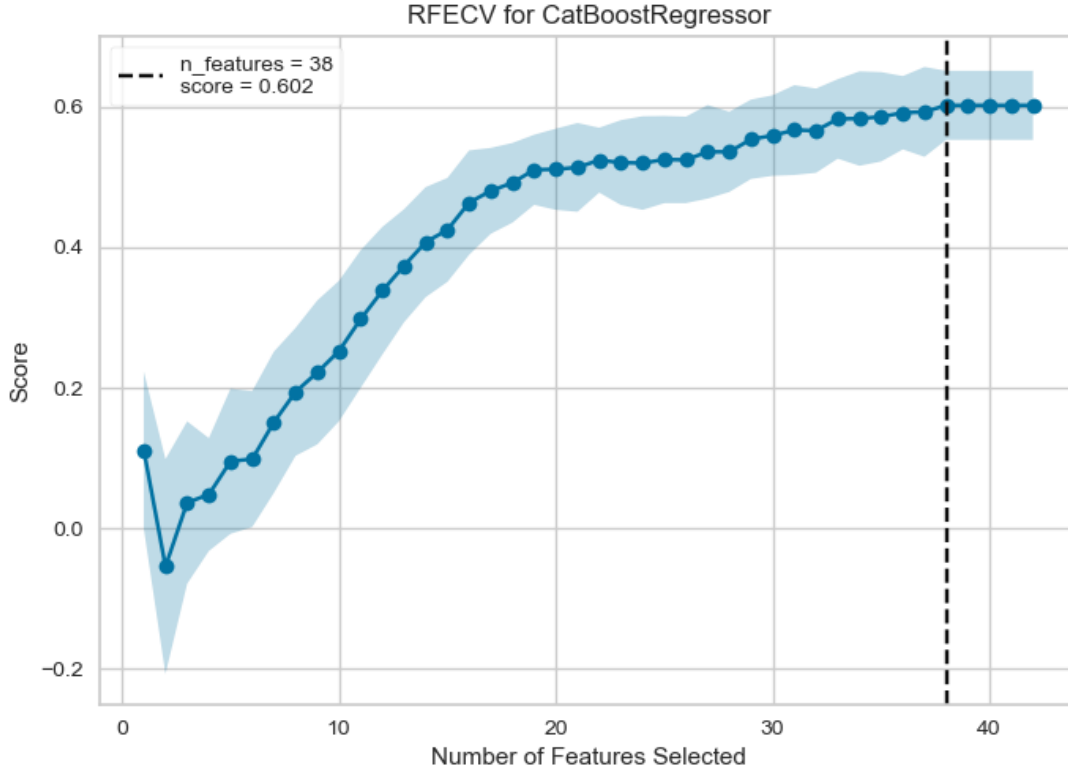
içeriğinin belirli bir altkümeye yoğunlaştığını, geri kalan değişkenlerin ise çoğunlukla artıksı ve kolinear katkılar sunduğunu göstermiştir. Simetrik (oblivious) ağaçların aynı derinlikte ortak bölme kuralı uygulaması, yüksek etkili birkaç değişkenin bileşik etkisini etkin biçimde toplarken, zayıf sinyalli özelliklerin tamamını elde tutmanın sağlayacağı marjinal kazancı sınırlar. Bu, RFECV eğrisindeki doyum eşiği ile tutarlıdır. Bu çalışmada, 42 özellik yerine yaklaşık 38 özelliğin korunması, genelleme performansını muhafaza edip gereksiz varyans artışını önleyerek daha sade ve kararlı bir model elde edilmesini sağlamıştır.



**Şekil 4.5:** En iyi performans gösteren CatBoost modelinin tahminleme hata dağılımı.

Seçim kararını istatistiksel açıdan sağlamlaştırmak için bir standart hata kuralı uygulanmaktadır: en iyi CV skorunu veren altküme yerine, en iyi skorun standart

hatası içinde kalan en küçük özellik sayısı seçilmektedir. Bu kural, aynı performans bandında daha sade bir model tercih ederek varyans maliyetini düşürmektedir. Ek olarak, RFECV'nin kararlılığını değerlendirmek için farklı tohumlarla ve tekrarlı k-katlı (k-fold) ile çalışıp, seçilen altkümeler arası Jaccard benzerliği gibi bir ölçüt raporlanabilmekte olup yüksek kararlılık, yoruma dayalı mühendislik çıkarımları için önemli görülmektedir.



**Şekil 4.6:** CatBoost için çapraz doğrulamalı yinelemeli özellik eleme (RFECV) eğrisi.

Sızıntıyı (leakage) önlemek amacıyla, tüm ön işleme adımlarının (ölçkleme ve kodlama varsa) RFECV içinde bir "pipeline" ile kapsüllenmesi gerekmektedir. Aksi halde çapraz doğrulama skorları iyimser yanlı olabilmektedir. Kolineer değişkenler için grup temelli yinelemeli özellik eleme (RFE) (örneğin, yüksek korelasyonlu kümelerden yalnızca bir temsilci bırakma) veya kümelenmiş özellik seçimi kullanılmaktadır. Bu, özellikle geometrik özniteliklerin türev veya tekrar üretimleri mevcutsa pratik fayda sağlamaktadır. RFECV bir model seçimi adımı olduğu için performansın tarafsız tahmini adına; iç döngüde özellik seçimi, dış döngüde performans tahmini olmak üzere iç içe çapraz doğrulama (nested CV) kullanılmaktadır. Bu prosedür, seçime bağlı

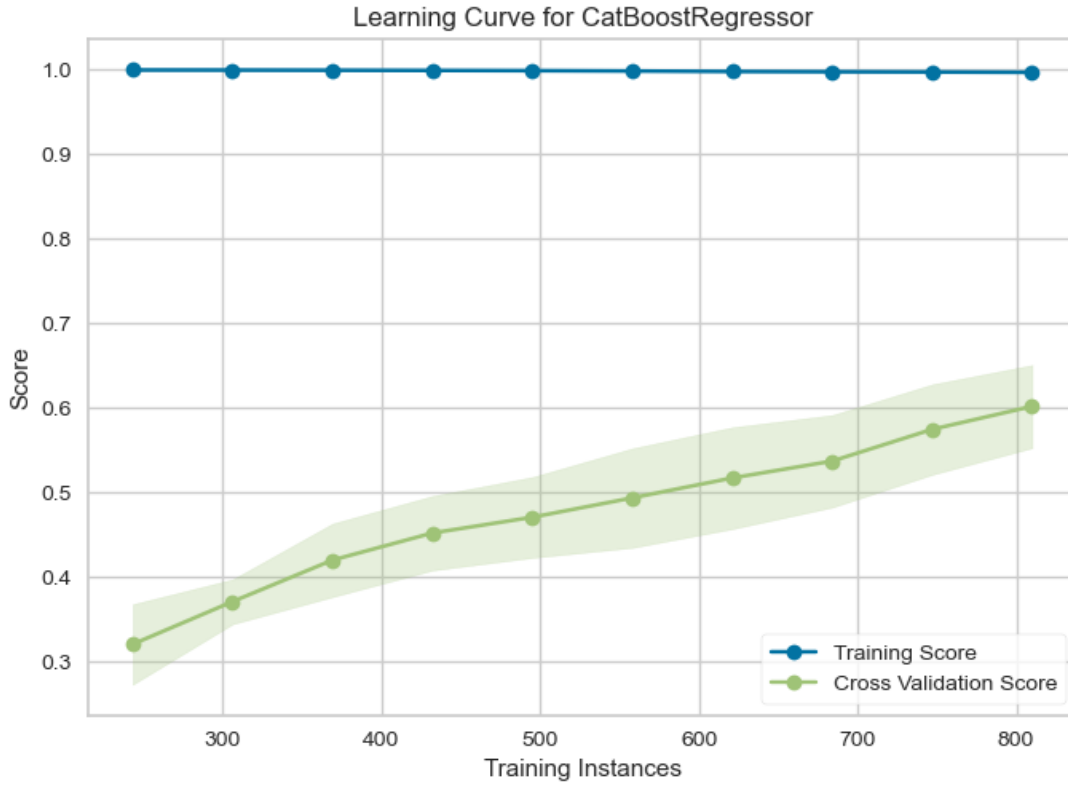
iyimserlik yanlılığını (selection bias) azaltmaktadır ve raporlanan tahmin doğruluğu oranlarının güvenilirliğini artırmaktadır.

Hesaplama maliyeti bakımından, RFECV'nin adım boyu (step) ve başlangıç özellik sayısı ile doğrusal ölçeklendiği unutulmamalıdır; ağaç tabanlı modellerde tek tek silme yerine  $n > 1$  özellik birden eleme (örneğin; step=2-3) pratikte ciddi hız kazandırmaktadır. Ancak, step büyüdükçe optimumu atlama riski arttığından, platonun çevresinde daha ince tarama yapılması (örneğin; 32-42 aralığında step=1) önerilmektedir. CatBoost'un tekrarlı özelliklere dayanıklılığı göz önüne alındığında, nihai seçimde yorumlanabilirlik ve gecikme ve karmaşıklık kısıtları da (çevrim içi kullanımda gecikme bütçesi vb.) karar kriterlerine eklenmelidir.

Şekil 4.7'de gösterilen öğrenme eğrisi (learning curve), eğitim puanı ile çapraz doğrulama (CV) puanının örneklem büyüklüğüne  $n$  bağlı seyrini birlikte göstererek modelin yanlılık-varyans (bias-variance) profilini nicel olarak teşhis etmeyi sağlamıştır. Bu çalışmada, eğitim puanının yüksek seviyede kalması (1.0 civarında sabit)) ve çapraz doğrulama puanının  $n$  arttıkça düzenli biçimde yükselmesi (0.6'ya düzenli yükseliyor), kısıtlayıcı etmenin yanlılıktan ziyade varyans olduğunu ve dolayısıyla daha fazla veriyle doğrulama başarımının artmaya devam ettiğini göstermiştir. Bu eğilim, orta ölçekli veri rejimlerinde CatBoost'un simetrik (oblivious) ağaçlarının sağladığı düzenlileştirmenin aşırı uyumu frenlediğini, ancak özellikle uç bölgelerde (çok yüksek ve çok düşük performanslı geometriler) örnek yetersizliği nedeniyle henüz tam kalibrasyon sağlayamadığını göstermiştir.

Eğrinin eğimi, azalan getiri (diminishing returns) noktasını tayin etmede kullanılmaktadır. CV hatası ve başarımı için basit bir ölçeklenme modeli,  $Err_{val}(n) \approx \alpha n^{-\beta} + c$  (veya karşılık gelen başarıma göre) şeklinde yazılmaktadır. Burada  $\beta > 0$  veri artışıyla beklenen iyileşme hızını,  $c$  ise yaklaşılabılır taban hata (irreducible error) seviyesini temsil etmektedir. Pratikte, mevcut eğriye bu tip bir güç-yasa (power-law) uyumu alınarak  $n$  iki katına çıkarıldığında beklenen kazanımın sayısal bir projeksiyonu verilmektedir. Eğimin belirgin biçimde yavaşlamadığı gözlenen bölgede, ek HAD senaryolarıyla veri genişletmeye devam etmek rasyonel olarak değerlendirilmektedir.

Ölçüm ve raporlama açısından, öğrenme eğrisinin yalnızca ortalama değil, %95 güven bantları ile birlikte verilmesi; tek bir k-katlı yerine tekrarlı k-katlı (repeated CV) kullanılması önerilmektedir. Böylece, eğrideki dalgalanmaların rastlantısal tohumlara duyarlılığı azaltılmakta ve gerçek eğilim daha güvenilir gösterilmektedir. Ayrıca, genelleme açıklığı  $\Delta(n) = \text{Score}_{\text{train}}(n) - \text{Score}_{\text{CV}}(n)$  izlenerek varyansın nicel seyri tartışılmaktadır. Burada,  $\Delta(n)$  küçülürken CV skorunun artmaya devam etmesi, veri artırmanın doğru yönde etkili olduğunu göstermektedir. Öğrenme eğrisi bulgularını, erken durdurma (early stopping), öğrenme oranı ve ağaç derinliği ve yaprak sayısı gibi CatBoost hiperparametreleriyle ilişkilendirerek (kapasite artışı  $\leftrightarrow$  varyans artışı dengesi) metodolojik gerekçe güçlendirilmektedir.



**Şekil 4.7:** CatBoost için öğrenme eğrisi (learning curve).

Regresyon modellerinin eğitim verisi üzerinde gerçekleştirilen 10 katlı çapraz doğrulama sonuçları karşılaştırıldığında, CatBoost, LightGBM, PINN ve GBR yöntemlerinin ilk sıralarda yer aldığı görülmüştür. Aralarında en yüksek doğruluk MAPE değerleri sıralandığında CatBoost'tadır. CatBoost'un simetrik (oblivious) ağaçları her derinlikte aynı bölme kuralını uygulayarak hipotez uzayını denetimli

biçimde sınırlamış; sıralı artırma (ordered boosting) gradyan yanlılığını azaltmış ve yaprak değerlerinin Newton güncellemeleri ise yakınsamayı kararlı kılmıştır. Bu birleşim, orta ölçekli örnekleme ( $n=900$  çapraz doğrulama) hem varyansı kontrol etmiş, hem de doğrusal olmayan etkileşimleri yeterince yakalamıştır. LightGBM ikinci yüksek performansı sergilemiştir. Yaprak seçimli (leaf-wise) büyüme ve agresif örnekleme politikaları onu çok güçlü kılsa da bu veri rejiminde CatBoost'a göre biraz daha değişken bir genelleme profili üretmiştir. GBR, klasik gradyan artırma yapısıyla rekabetçi olmakla birlikte, düzenleme ve erken durdurma ayarlarına daha duyarlıdır. XGBoost, güçlü bir taban çizgisi verse de asimetrik ağaçlarının küçük alt-bölgelerde aşırı uyum eğilimi, bu veri büyüklüğünde CatBoost'un gerisinde kalmasına yol açmıştır.

En yüksek üçüncü performansı PINN modeli sağlamıştır. Bu çalışmada, modele sağlanan fiziksel bilgi, kanal geometrilerinin  $y$ -koordinatları olmuştur. PINN yönteminin literatürde yaygın kullanılan CatBoost, LightGBM gibi başarılı modellerle karşılaştırılabilir bir performans sunduğu gözlenmiştir.

Takip eden modeller RF ve ETR'dir. Torbalama (bagging) stratejisi varyansı düşürmüştür; ancak çok sayıda parçalı sabit tahminin ortalaması, bu problemdeki eşik etkilerini ve keskin yerel ilişkileri yeterince ince taneli yakalayamamıştır. Bu nedenle, boosting ailesinin gerisinde kalmıştır. AdaBoost'ta hatalı örnekleri agresif ağırlıklandırma gürültüye duyarlılığı artırmış, genel doğruluk daha düşük kalmıştır.

Diğer yandan, LR ve Ridge doğrusal varsayımları nedeniyle belirgin biçimde sınırlı performans vermiştir. Geometri kaynaklı kuvvetli doğrusal olmayan etkileşimler ve özellikler arası eşik davranışları, tek katsayılı doğrusal modellerle temsil edilemediği için hata seviyeleri yüksektir.

Sonuç olarak, aynı veri bölüşümü ve değerlendirme altında CatBoost; simetrik ağaçlarla kontrollü kapasite, sıralı artırma ile daha düşük gradyan yanlılığı ve ikinci mertebeye güncellemelerle kararlı yakınsama sayesinde ilk 10 yöntem arasında en düşük hata ve en yüksek açıklama gücünü sağlamıştır. LightGBM, GBR ve XGBoost yakın takipçi olsa da, bu tez bağlamında genellenebilirlik ve kalibrasyon bakımından CatBoost belirgin üstünlük göstermiştir.



## 5. SONUÇLAR VE GELECEKTEKİ ÇALIŞMALAR

Bu tez çalışmasında, ters ozmoz (RO) sistemlerinde kullanılan membranların akış kanal geometrilerinin performansını iyileştirmek amacıyla hem hesaplamalı akışkanlar dinamiği (HAD) hem de yapay zeka (AI) yaklaşımları entegre bir şekilde kullanılmıştır. Çalışmanın ilk aşamasında, çeşitli spline noktalardan türetilmiş 1000'in üzerinde rastgele kanal geometrisi oluşturulmuş, her biri için HAD simülasyonları yürütülerek permeat akışı, kirlenme miktarı ve basınç farkı gibi temel performans parametreleri elde edilmiştir. Bu parametreler kullanılarak oluşturulan birleşik performans metriği (BPM) aracılığıyla, sistem çıktıları tek bir sayısal göstergeye indirgenmiş ve modellemeye uygun hale getirilmiştir.

Elde edilen veri seti ile regresyon temelli bir tahminleme modeli eğitilmiş, farklı algoritmalar karşılaştırılmış ve en düşük hata oranını sağlayan model seçilmiştir. Geliştirilen regresyon modeli sayesinde yeni oluşturulacak rastgele geometriler, HAD simülasyonuna gerek kalmadan hızlı bir şekilde performans bakımından sıralanabilir hale gelmiştir. Bu yaklaşım, deneysel ve hesaplamalı yükleri azaltarak tasarım sürecini hızlandırmakta ve daha geniş tasarım alanlarının keşfedilmesine olanak tanımaktadır. Özellikle kategorik artırma (CatBoost) algoritması, test verisinde yaklaşık %8.8 MAPE hatası ile öne çıkmış, modelin genelleme başarısı yüksek bulunmuştur. Bu doğrultuda, AI destekli ön değerlendirme sürecinin, akış kanal tasarımında yüksek doğrulukla çalışabileceği kanıtlanmıştır.

Bu tezin temel katkılarından biri, geometri odaklı membran performansının sonlu bir HAD simülasyonları kümesinden öğrenilebileceğini ve genelleştirilebileceğini göstermiş olmasıdır. Bu modelleme yaklaşımı, tasarım sürecini önemli ölçüde hızlandırmış, membran mühendisliğinde veri odaklı optimizasyon için yeni yollar açmıştır. Örneğin, regresyon modelini evrimsel algoritmalar veya Bayes optimizasyonu ile birleştirerek çok amaçlı kriterlere dayalı optimum geometriler için otomatik aramalar yapmak mümkün hale gelmiştir. Bu tür hibrit yaklaşımların, akıyı maksimize ederken

kirlenmeyi ve basınç kaybını en aza indirerek denge sağlayabileceği öngörülmektedir. Ayrıca, fizik bilgili derin öğrenme ağları (PINN) gibi daha karmaşık bir öğrenme mimarisi, regresyon yaklaşımı için kullanılmış olup diğer modeller ile karşılaştırılabilir bir performans elde edilmiştir.

Bu çalışmanın sunduğu yöntemsel çerçevenin; veri seti, performans metriği, uygulanan regresyon modellerinde iyileştirme yapılması gibi çeşitli açılardan geliştirilebileceği öngörülmektedir. Öncelikle, bu çalışmada kullanılan veri seti, 1025 örnekle sınırlı uygulanabilmiştir. HAD simülasyonları ile daha fazla geometri örneği elde edilerek modelin genel geçerliliğini ve tahmin başarısını artırılabilir. Veri seti, ayrıca, farklı kanal uzunlukları, genişlikleri veya membran özellikleri gibi ek parametrik varyasyonlar dahil edilerek genişletilebilir. Diğer yandan, bu çalışmada kullanılan performans göstergesi BPM olarak tanımlanmıştır. Her bir performans ölçütünün (permeat akısı, basınç kaybı ve kirlenme miktarı) ayrı ayrı hedefler olarak modellendiği çok çıkışlı (multi-output) regresyon yaklaşımları uygulanarak sapma oranları ayrı ayrı incelenebilir. Bir diğer yandan, regresyon yaklaşımıyla uygulanan fizik bilgili derin öğrenme ağı (PINN) yönteminde daha fazla fizik bilgisi kullanılarak modelin performansının iyileşebileceği öngörüsü yapılabilir.

Bu çalışma, aynı zamanda, akışkanlar mekaniğindeki daha geniş hesaplamalı tasarım alanı için de çıkarımlara sahiptir. Fiziksel simülasyon verilerinin ML modelleriyle bütünleştirilmesi, tamamen deneysel veya deneme yanılma tasarım süreçlerinden akıllı ve öngörücü modelleme çerçevelerine geçişi mümkün kılmıştır. Bu tezde önerilen metodolojinin, geometrinin performansı güçlü bir şekilde etkilediği havacılık, mikroakışkanlar, biyomedikal cihaz tasarımı gibi diğer mühendislik alanlarına uyarlanabilir olduğu değerlendirilebilir.

İleriye dönük olarak, bu tezde ortaya konulan çıktıların (yüksek performansa sahip olarak tahminlenen geometrilerin) sadece simülasyon ortamında değil aynı zamanda üretim açısından da uygulanabilir olup olmadıkları araştırılmalıdır. Bu kapsamda, mikro akışkan sistem üretim teknikleri ile prototip üretimi yapılarak deneysel doğrulama gerçekleştirilebilir. Öte yandan, eğitilmiş modellerin gelecekte dijital ikiz (digital twin) sistemlerde, gerçek zamanlı sistem izleme ve tasarım güncellemesi

amacıyla kullanılması mümkündür. Bu tür uygulamalarda düşük gecikmeli tahminleme sistemleri kritik rol oynamaktadır.

Sonuç olarak, bu tez çalışması, veri odaklı modelleme ve AI tabanlı yaklaşımın, RO sistemleri için akış kanal tasarımında etkin bir mühendislik çözümü sunduğunu göstermiştir. Sadece membran performansını öngörmekle kalmayıp, aynı zamanda fiziksel deneysel süreci daha verimli kılacak karar destek sistemlerinin temelleri atılmıştır. Geliştirilecek ileri seviye tahminleme modelleri, çevresel sürdürülebilirlik ve su arıtma sistemlerinin optimizasyonu gibi alanlarda geniş uygulama alanları bulabilecektir.





## KAYNAKLAR

- [1] **Baker, M.N. ve Taras, M.J.** (1948). The quest for pure water: the history of water purification from the earliest records to the twentieth century, (*No Title*).
- [2] **Pontius, F.** (2003). *Drinking water regulation and health*, John Wiley & Sons.
- [3] **Spellman, F.R.** (2008). *Handbook of water and wastewater treatment plant operations*, CRC press.
- [4] **Organization, W.H.** (2004). *Guidelines for drinking-water quality*, cilt 1, World health organization.
- [5] **Shannon, M.A., Bohn, P.W., Elimelech, M., Georgiadis, J.G., Mariñas, B.J. ve Mayes, A.M.** (2008). Science and technology for water purification in the coming decades, *Nature*, 452(7185), 301–310.
- [6] **Ahuja, S.** (2021). *Handbook of water purity and quality*, Academic press.
- [7] **Tchobanoglous, G., Eddy, M., Stensel, H., Tsuchihashi, R., Burton, F., Abu-Orf, M., Bowden, G. ve Pfrang, W.** (2014). *Wastewater Engineering: Treatment and Resource Recovery*, sayiv. 1 in McGraw-Hill series in civil and environmental engineering, McGraw-Hill Education, <https://books.google.com.tr/books?id=ZHbFsgEACAAJ>.
- [8] **Association, A.W.W. ve Edzwald, J.** (2010). *Water Quality & Treatment: A Handbook on Drinking Water*, Water Resources and Environmental Engineering Series, McGraw Hill LLC, <https://books.google.com.tr/books?id=Lr2ossXJ6NwC>.
- [9] **Judd, S.** (2010). *The MBR book: principles and applications of membrane bioreactors for water and wastewater treatment*, Elsevier.
- [10] **Glaze, W.H., Kang, J.W. ve Chapin, D.H.** (1987). The chemistry of water treatment processes involving ozone, hydrogen peroxide and ultraviolet radiation.
- [11] **Bolton, J.R. ve Cotton, C.A.** (2008). *The ultraviolet disinfection handbook*, American Water Works Association.
- [12] **Elimelech, M. ve Phillip, W.A.** (2011). The future of seawater desalination: energy, technology, and the environment, *science*, 333(6043), 712–717.
- [13] **Gude, V.G.** (2017). Desalination and water reuse to address global water scarcity, *Reviews in Environmental Science and Bio/Technology*, 16(4), 591–609.

- [14] **Council, N.R., on Earth, D., Studies, L., Science, W., Board, T. ve on Advancing Desalination Technology, C.** (2008). *Desalination: A national perspective*, National Academies Press.
- [15] **Ghaffour, N., Missimer, T.M. ve Amy, G.L.** (2013). Technical review and evaluation of the economics of water desalination: Current and future challenges for better water supply sustainability, *Desalination*, 309, 197–207.
- [16] **Kesieme, U.K., Milne, N., Aral, H., Cheng, C.Y. ve Duke, M.** (2013). Economic analysis of desalination technologies in the context of carbon pricing, and opportunities for membrane distillation, *Desalination*, 323, 66–74.
- [17] **Alkaisi, A., Mossad, R. ve Sharifian-Barforoush, A.** (2017). A review of the water desalination systems integrated with renewable energy, *Energy Procedia*, 110, 268–274.
- [18] **Cipollina, A., Micale, G. ve Rizzuti, L.** (2005). A critical assessment of desalination operations in Sicily, *Desalination*, 182(1-3), 1–12.
- [19] **El-Ghonemy, A.** (2018). Performance test of a sea water multi-stage flash distillation plant: Case study, *Alexandria engineering journal*, 57(4), 2401–2413.
- [20] **Sangwai, J.S., Patel, R.S., Mekala, P., Mech, D. ve Busch, M.** (2013). Desalination of seawater using gas hydrate technology-current status and future direction, *Proceedings of the 18th International Conference on Hydraulics, Water Resources, Coastal and Environmental Engineering, HYDRO*, s.434–440.
- [21] **Talaeipour, M., Nouri, J., Hassani, A. ve Mahvi, A.** (2017). An investigation of desalination by nanofiltration, reverse osmosis and integrated (hybrid NF/RO) membranes employed in brackish water treatment, *Journal of Environmental Health Science and Engineering*, 15, 1–9.
- [22] **Xu, T.** (2005). Ion exchange membranes: State of their development and perspective, *Journal of membrane science*, 263(1-2), 1–29.
- [23] **Lee, K.P., Arnot, T.C. ve Mattia, D.** (2011). A review of reverse osmosis membrane materials for desalination—Development to date and future potential, *Journal of membrane science*, 370(1-2), 1–22.
- [24] **Jones, E., Qadir, M., Van Vliet, M.T., Smakhtin, V. ve Kang, S.m.** (2019). The state of desalination and brine production: A global outlook, *Science of the Total Environment*, 657, 1343–1356.
- [25] **Nair, M. ve Kumar, D.** (2013). Water desalination and challenges: The Middle East perspective: a review, *Desalination and Water Treatment*, 51(10-12), 2030–2040.

- [26] **Goh, P., Matsuura, T., Ismail, A. ve Hilal, N.** (2016). Recent trends in membranes and membrane processes for desalination, *Desalination*, 391, 43–60.
- [27] **Kaplan, R., Mamrosh, D., Salih, H.H. ve Dastgheib, S.A.** (2017). Assessment of desalination technologies for treatment of a highly saline brine from a potential CO<sub>2</sub> storage site, *Desalination*, 404, 87–101.
- [28] **Samco Tech** (2025). *What is Reverse Osmosis and How Is It Used for Industrial Applications?*, <https://samcotech.com/what-is-reverse-osmosis>, accessed: 2025-07-13.
- [29] **Rognoni, M.** (2010). La dissalazione dell'acqua di mare, *Descrizione, analisi e valutazione delle principali tecnologie*, Dario Flaccovio, Palermo.
- [30] **Mito, M.T., Ma, X., Albuflasa, H. ve Davies, P.A.** (2019). Reverse osmosis (RO) membrane desalination driven by wind and solar photovoltaic (PV) energy: State of the art and challenges for large-scale implementation, *Renewable and Sustainable Energy Reviews*, 112, 669–685.
- [31] **Greenlee, L.F., Lawler, D.F., Freeman, B.D., Marrot, B. ve Moulin, P.** (2009). Reverse osmosis desalination: water sources, technology, and today's challenges, *Water research*, 43(9), 2317–2348.
- [32] **Liu, S., Wang, Z., Han, M. ve Zhang, J.** (2021). Embodied water consumption between typical desalination projects: reverse osmosis versus low-temperature multi-effect distillation, *Journal of Cleaner Production*, 295, 126340.
- [33] **Alsarayreh, A.A., Al-Obaidi, M.A., Patel, R. ve Mujtaba, I.M.** (2020). Scope and limitations of modelling, simulation, and optimisation of a spiral wound reverse osmosis process-based water desalination, *Processes*, 8(5), 573.
- [34] **Akther, N., Sodiq, A., Giwa, A., Daer, S., Arafat, H. ve Hasan, S.** (2015). Recent advancements in forward osmosis desalination: A review, *Chemical Engineering Journal*, 281, 502–522.
- [35] **Wenten, I.G. ve diğerleri** (2016). Reverse osmosis applications: Prospect and challenges, *Desalination*, 391, 112–125.
- [36] **Subramani, A. ve Jacangelo, J.G.** (2015). Emerging desalination technologies for water treatment: a critical review, *Water research*, 75, 164–187.
- [37] **Majeed, T., Phuntsho, S., Jeong, S., Zhao, Y., Gao, B. ve Shon, H.K.** (2016). Understanding the risk of scaling and fouling in hollow fiber forward osmosis membrane application, *Process Safety and Environmental Protection*, 104, 452–464.
- [38] **Tijing, L.D., Woo, Y.C., Choi, J.S., Lee, S., Kim, S.H. ve Shon, H.K.** (2015). Fouling and its control in membrane distillation—A review, *Journal of Membrane Science*, 475, 215–244.

- [39] **Aliyu, U.M., Rathilal, S. ve Isa, Y.M.** (2018). Membrane desalination technologies in water treatment: A review, *Water Practice & Technology*, 13(4), 738–752.
- [40] **Crittenden, J.C., Trussell, R.R., Hand, D.W., Howe, K.J. ve Tchobanoglous, G.** (2012). *MWH's water treatment: principles and design*, John Wiley & Sons.
- [41] **van der Meer, W.G.J.** (2003). Mathematical modelling of NF and RO membrane filtration plants and modules.
- [42] **Zimmerer, C. ve Kottke, V.** (1996). Effects of spacer geometry on pressure drop, mass transfer, mixing behavior, and residence time distribution, *Desalination*, 104(1-2), 129–134.
- [43] **Schock, G. ve Miquel, A.** (1987). Mass transfer and pressure loss in spiral wound modules, *Desalination*, 64, 339–352.
- [44] **Shrivastava, A., Kumar, S. ve Cussler, E.** (2008). Predicting the effect of membrane spacers on mass transfer, *Journal of Membrane Science*, 323(2), 247–256.
- [45] **Da Costa, A.R. ve Fane, A.G.** (1994). Net-type spacers: effect of configuration on fluid flow path and ultrafiltration flux, *Industrial & engineering chemistry research*, 33(7), 1845–1851.
- [46] **Da Costa, A., Fane, A. ve Wiley, D.** (1994). Spacer characterization and pressure drop modelling in spacer-filled channels for ultrafiltration, *Journal of membrane science*, 87(1-2), 79–98.
- [47] **Neal, P., Li, H., Fane, A. ve Wiley, D.** (2003). The effect of filament orientation on critical flux and particle deposition in spacer-filled channels, *Journal of Membrane Science*, 214(2), 165–178.
- [48] **Belfort, G. ve Guter, G.A.** (1972). An experimental study of electrodialysis hydrodynamics, *Desalination*, 10(3), 221–262.
- [49] **Da Costa, A., Fane, A., Fell, C. ve Franken, A.** (1991). Optimal channel spacer design for ultrafiltration, *Journal of Membrane Science*, 62(3), 275–291.
- [50] **Light, W.G. ve Tran, T.V.** (1981). Improvement of thin-channel design for pressure-driven membrane systems, *Industrial & Engineering Chemistry Process Design and Development*, 20(1), 33–40.
- [51] **Thomas, D.G., Hayes, P.H., Mixon, W.R., Sheppard, J.D., Griffith, W.L. ve Keller, R.M.** (1970). Turbulence promoters for hyperfiltration with dynamic membranes, *Environmental Science & Technology*, 4(12), 1129–1136.
- [52] **Johnson, J. ve Busch, M.** (2010). Engineering aspects of reverse osmosis module design, *Desalination and Water Treatment*, 15(1-3), 236–248.

- [53] **Busch, M. ve Mickols, W.** (2004). Reducing energy consumption in seawater desalination, *Desalination*, 165, 299–312.
- [54] **Molina, V.G., Busch, M. ve Sehn, P.** (2009). Cost savings by novel seawater reverse osmosis elements and design concepts, *Desalination and Water Treatment*, 7(1-3), 160–177.
- [55] **Koutsou, C., Karabelas, A. ve Kostoglou, M.** (2015). Membrane desalination under constant water recovery–The effect of module design parameters on system performance, *Separation and Purification Technology*, 147, 90–113.
- [56] **Franks, R., Bartels, C. ve Anit, A.** (2010). Demonstrating improved RO system performance with new Low Differential (LD) technology, *Hydranautics, Hydranautics, Oceanside, CA*.
- [57] **Balster, J., Pünt, I., Stamatialis, D. ve Wessling, M.** (2006). Multi-layer spacer geometries with improved mass transport, *Journal of Membrane Science*, 282(1-2), 351–361.
- [58] **Li, F., Meindersma, W., De Haan, A. ve Reith, T.** (2005). Novel spacers for mass transfer enhancement in membrane separations, *Journal of Membrane Science*, 253(1-2), 1–12.
- [59] **Tsuneda, S., Aikawa, H., Hayashi, H., Yuasa, A. ve Hirata, A.** (2003). Extracellular polymeric substances responsible for bacterial adhesion onto solid surface, *FEMS microbiology letters*, 223(2), 287–292.
- [60] **Araújo, P.A., Miller, D.J., Correia, P.B., van Loosdrecht, M.C., Kruithof, J.C., Freeman, B.D., Paul, D. ve Vrouwenvelder, J.S.** (2012). Impact of feed spacer and membrane modification by hydrophilic, bactericidal and biocidal coating on biofouling control, *Desalination*, 295, 1–10.
- [61] **Hausman, R., Gullinkala, T. ve Escobar, I.C.** (2010). Development of copper-charged polypropylene feedspacers for biofouling control, *Journal of Membrane Science*, 358(1-2), 114–121.
- [62] **Schwinge, J., Neal, P., Wiley, D., Fletcher, D. ve Fane, A.** (2004). Spiral wound modules and spacers: Review and analysis, *Journal of membrane science*, 242(1-2), 129–153.
- [63] **Fimbres-Weihs, G. ve Wiley, D.** (2010). Review of 3D CFD modeling of flow and mass transfer in narrow spacer-filled channels in membrane modules, *Chemical Engineering and Processing: Process Intensification*, 49(7), 759–781.
- [64] **Von Der Schulenburg, D.G., Vrouwenvelder, J., Creber, S., Van Loosdrecht, M. ve Johns, M.** (2008). Nuclear magnetic resonance microscopy studies of membrane biofouling, *Journal of Membrane Science*, 323(1), 37–44.

- [65] **Willems, P., Deen, N., Kemperman, A.J., Lammertink, R.G., Wessling, M., van Sint Annaland, M., Kuipers, J. ve van der Meer, W.G.J.** (2010). Use of Particle Imaging Velocimetry to measure liquid velocity profiles in liquid and liquid/gas flows through spacer filled channels, *Journal of Membrane Science*, 362(1-2), 143–153.
- [66] **Haidari, A., Heijman, S. ve Van der Meer, W.** (2016). Visualization of hydraulic conditions inside the feed channel of Reverse Osmosis: A practical comparison of velocity between empty and spacer-filled channel, *Water research*, 106, 232–241.
- [67] **Ahmad, A.L., Lau, K.K., Bakar, M.A. ve Shukor, S.A.** (2005). Integrated CFD simulation of concentration polarization in narrow membrane channel, *Computers & chemical engineering*, 29(10), 2087–2095.
- [68] **Icoz, T. ve Jaluria, Y.** (2006). Design optimization of size and geometry of vortex promoter in a two-dimensional channel.
- [69] **Amokrane, M., Sadaoui, D., Dudeck, M. ve Koutsou, C.P.** (2016). New spacer designs for the performance improvement of the zigzag spacer configuration in spiral-wound membrane modules, *Desalination and Water Treatment*, 57(12), 5266–5274.
- [70] **Fiebig, M.** (1995). Vortex generators for compact heat exchangers, *Journal of Enhanced Heat Transfer*, 2(1-2).
- [71] **Mitchell, T.M.** (2006). *The discipline of machine learning*, cilt 9, Carnegie Mellon University, School of Computer Science, Machine Learning . . . .
- [72] **Günay, M.E., Tapan, N.A. ve Akkoç, G.** (2022). Analysis and modeling of high-performance polymer electrolyte membrane electrolyzers by machine learning, *International Journal of Hydrogen Energy*, 47(4), 2134–2151.
- [73] **Zhang, Z., Luo, Y., Peng, H., Chen, Y., Liao, R.Z. ve Zhao, Q.** (2021). Deep spatial representation learning of polyamide nanofiltration membranes, *Journal of Membrane Science*, 620, 118910.
- [74] **Adib, H., Raisi, A. ve Salari, B.** (2019). Support vector machine-based modeling of grafting hyperbranched polyethylene glycol on polyethersulfone ultrafiltration membrane for separation of oil–water emulsion, *Research on Chemical Intermediates*, 45(11), 5725–5743.
- [75] **Liu, J., Kang, X., Luan, X., Gao, L., Tian, H. ve Liu, X.** (2020). Performance and membrane fouling behaviors analysis with SVR-LibSVM model in a submerged anaerobic membrane bioreactor treating low-strength domestic sewage, *Environmental Technology & Innovation*, 19, 100844.
- [76] **Wilson, N., Wilhelm, D., Qian, X., Arróyave, R. ve Qian, X.** (2022). Batch active learning for accelerating the development of interatomic potentials, *Computational Materials Science*, 208, 111330.

- [77] **Gao, H., Zhong, S., Zhang, W., Igou, T., Berger, E., Reid, E., Zhao, Y., Lambeth, D., Gan, L., Afolabi, M.A. ve diğeri** (2021). Revolutionizing membrane design using machine learning-bayesian optimization, *Environmental science & technology*, 56(4), 2572–2581.
- [78] **Rall, D., Menne, D., Schweidtmann, A.M., Kamp, J., von Kolzenberg, L., Mitsos, A. ve Wessling, M.** (2019). Rational design of ion separation membranes, *Journal of membrane science*, 569, 209–219.
- [79] **Odabaşı, Ç., Dologlu, P., Gülmez, F., Kuşoğlu, G. ve Çağlar, Ö.** (2022). Investigation of the factors affecting reverse osmosis membrane performance using machine-learning techniques, *Computers & Chemical Engineering*, 159, 107669.
- [80] **Mulder, M.** (2012). *Basic principles of membrane technology*, Springer science & business media.
- [81] **Meng, X., Wang, F., Meng, S., Wang, R., Mao, Z., Li, Y., Yu, M., Wang, X., Zhao, Q. ve Yang, L.** (2021). Novel surrogates for membrane fouling and the application of support vector machine in analyzing fouling mechanism, *Membranes*, 11(12), 990.
- [82] **Teychene, B., Touffet, A., Baron, J., Welte, B., Joyeux, M. ve Gallard, H.** (2018). Predicting of ultrafiltration performances by advanced data analysis, *Water Research*, 129, 365–374.
- [83] **Zhu, L. ve Chen, J.** (2018). Prognostics of PEM fuel cells based on Gaussian process state space models, *Energy*, 149, 63–73.
- [84] **Fetanat, M., Keshtiara, M., Keyikoglu, R., Khataee, A., Daiyan, R. ve Razmjou, A.** (2021). Machine learning for design of thin-film nanocomposite membranes, *Separation and Purification Technology*, 270, 118383.
- [85] **Xie, P., Murdoch, L.C. ve Ladner, D.A.** (2019). Mitigating membrane fouling with sinusoidal spacers, *Desalination and Water Treatment*, 168, 56–64.
- [86] **Ling, B., Xie, P., Ladner, D. ve Battiato, I.** (2021). Dynamic modeling of fouling in reverse osmosis membranes, *Membranes*, 11(5), 349.
- [87] **Saffman, P.G.** (1971). On the boundary condition at the surface of a porous medium, *Studies in applied mathematics*, 50(2), 93–101.
- [88] **Yin, H., Xu, M., Luo, Z., Bi, X., Li, J., Zhang, S. ve Wang, X.** (2024). Machine learning for membrane design and discovery, *Green Energy & Environment*, 9(1), 54–70.
- [89] **James, G., Witten, D., Hastie, T., Tibshirani, R. ve diğeri** (2013). *An introduction to statistical learning*, cilt112, Springer.
- [90] **Draper, N.R. ve Smith, H.** (1998). *Applied regression analysis*, cilt326, John Wiley & Sons.

- [91] **Hoerl, A.E. ve Kennard, R.W.** (1970). Ridge regression: Biased estimation for nonorthogonal problems, *Technometrics*, 12(1), 55–67.
- [92] **Tibshirani, R.** (1996). Regression shrinkage and selection via the lasso, *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 58(1), 267–288.
- [93] **Zou, H. ve Hastie, T.** (2005). Regularization and variable selection via the elastic net, *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 67(2), 301–320.
- [94] **Huber, P.J.,** (1992). Robust estimation of a location parameter, *Breakthroughs in statistics: Methodology and distribution*, Springer, s.492–518.
- [95] **Efron, B., Hastie, T., Johnstone, I. ve Tibshirani, R.** (2004). Least angle regression.
- [96] **Pati, Y.C., Rezaiifar, R. ve Krishnaprasad, P.S.** (1993). Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition, *Proceedings of 27th Asilomar conference on signals, systems and computers*, IEEE, s.40–44.
- [97] **MacKay, D.J.** (1992). Bayesian interpolation, *Neural computation*, 4(3), 415–447.
- [98] **Breiman, L., Friedman, J., Olshen, R.A. ve Stone, C.J.** (2017). *Classification and regression trees*, Chapman and Hall/CRC.
- [99] **Breiman, L.** (2001). Random forests, *Machine learning*, 45, 5–32.
- [100] **Geurts, P., Ernst, D. ve Wehenkel, L.** (2006). Extremely randomized trees, *Machine learning*, 63(1), 3–42.
- [101] **Friedman, J.H.** (2001). Greedy function approximation: a gradient boosting machine, *Annals of statistics*, 1189–1232.
- [102] **Freund, Y. ve Schapire, R.E.** (1997). A decision-theoretic generalization of on-line learning and an application to boosting, *Journal of computer and system sciences*, 55(1), 119–139.
- [103] **Chen, T. ve Guestrin, C.** (2016). Xgboost: A scalable tree boosting system, *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, s.785–794.
- [104] **Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q. ve Liu, T.Y.** (2017). Lightgbm: A highly efficient gradient boosting decision tree, *Advances in neural information processing systems*, 30.
- [105] **Dorogush, A.V., Ershov, V. ve Gulin, A.** (2018). CatBoost: gradient boosting with categorical features support, *arXiv preprint arXiv:1810.11363*.
- [106] **Cover, T. ve Hart, P.** (1967). Nearest neighbor pattern classification, *IEEE Transactions on Information Theory*, 13(1), 21–27.

- [107] **Zhang, D. ve Gong, Y.** (2020). The comparison of LightGBM and XGBoost coupling factor analysis and prediagnosis of acute liver failure, *Ieee Access*, 8, 220990–221003.
- [108] **Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A.V. ve Gulin, A.** (2018). CatBoost: unbiased boosting with categorical features, *Advances in neural information processing systems*, 31.
- [109] **Aygun, A. ve Karakus, A.** (2022). Physics informed neural networks for two dimensional incompressible thermal convection problems, *Isı Bilimi ve Tekniği Dergisi*, 42(2), 221–232.
- [110] **Raissi, M., Perdikaris, P. ve Karniadakis, G.E.** (2019). Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations, *Journal of Computational physics*, 378, 686–707.
- [111] **Karniadakis, G.E., Kevrekidis, I.G., Lu, L., Perdikaris, P., Wang, S. ve Yang, L.** (2021). Physics-informed machine learning, *Nature Reviews Physics*, 3(6), 422–440.



## ÖZGEÇMİŞ

**Adı SOYADI: Rabiya KARALI TÜRKER**

### ÖĞRENİM DURUMU:

- **Lisans:** 2013, İzmir Yüksek Teknoloji Enstitüsü, Mühendislik Fakültesi, Elektronik ve Haberleşme Mühendisliği Bölümü.
- **Y. Lisans:** 2025, İstanbul Teknik Üniversitesi, Elektronik ve Haberleşme Mühendisliği Anabilim Dalı, Telekomünikasyon Mühendisliği Programı.

### EĞİTİM DERECEŚİ ve MESLEKİ DENEYİMLER:

- 2013 yılında İYTE Elektronik ve Haberleşme Mühendisliği Bölümünü 5. olarak tamamladı.
- 2013-2015 yılları arasında VESTEL ELEKTRONİK A.Ş'de Komponent Mühendisi olarak çalıştı.
- 2015-2017 yılları arasında İstanbul Teknik Üniversitesinde yüksek lisans derslerini tamamladı.
- 2017-2021 yılları arasında Elektro Elektronik A.Ş'de Saha Uygulama Mühendisi (FAE) olarak çalıştı.
- 2018-2020 yılları arasında Anadolu Üniversitesi'nde İşletme Yönetimi bölümünden mezun oldu.
- 2021-2024 yılları arasında ASELSAN ELEKTRONİK A.Ş'de Komponent Mühendisi olarak çalıştı.
- 2024 Kasım ayından bu yana ASELSAN ELEKTRONİK A.Ş'de Takım Lideri olarak çalışıyor.

### YÜKSEK LİSANS TEZİNDEN TÜRETİLEN YAYINLAR:

- **Türker, R.K.,** Ekici, H., Aydinalp, C. (2025). *the 4th International Graduate Research Symposium: IGRS'25 - Enhancing Reverse Osmosis Performance: A CFD-Based Study of Randomized Feed Spacer Geometries*, May 12-14, 2025 İstanbul, Türkiye. (Online Presentation)

- **Türker, R.K.**, Ekici, H., Aydıralp, C. (2025). *7th International Congress on Human-Computer Interaction, Optimization and Robotic Applications: Ters Ozmoz Sistemlerinde Akış Kanalı Tasarımı ve Analizi: Ön Çalışma*, May 23-24, 2025, Ankara, Türkiye. (In-person)

