Artvin Çoruh Üniversitesi
Doğal Afetler Uygulama ve Araştırma Merkezi
Doğal Afetler ve Çevre Dergisi

Artvin Çoruh University
Natural Hazards Application and Research Center
Journal of Natural Hazards and Environment

Araştırma Makalesi / Research Article, Doğ Afet Çev Derg, 2026; 12(2): 258-272, 10.21324/dacd.1852115

Doğu Akdeniz Bölgesi İçin Makine Öğrenmesi Tabanlı Buharlaşma Modelleri

Mehmet Ali Tor[1,*], Ramazan Koçak[2], Halil Emre Kuşçu[3], Ahmet Emre Tekeli[4]

[1]Çukurova Üniversitesi, Ziraat Fakültesi, Tarımsal Yapılar ve Sulama Bölümü, 01330, Adana
[2]İstanbul Teknik Üniversitesi, Uçak ve Uzay Bilimleri Fakültesi, Meteoroloji Mühendisliği Bölümü, 34485, İstanbul
[3]Çukurova Üniversitesi, Ziraat Fakültesi, Tarımsal Yapılar ve Sulama Bölümü, 01330, Adana
[4]Iğdır Üniversitesi, Mühendislik Fakültesi, İnşaat Mühendisliği Bölümü, 76000, Iğdır

Özet

Buharlaşma; su kaynaklarının verimli kullanılabilmesi için yapılan planlama çalışmalarında, rezervuar alanlarının işletilmesinde ve tarımsal verimlilik konularında büyük öneme sahip meteorolojik bir değişkendir. Çalışmada, buharlaşmanın makine öğrenmesi algoritmaları kullanılarak hangi doğrulukta tahmin edilebileceği ve farklı algoritmaların verimlilikleri karşılaştırılarak en iyi buharlaşma tahminini veren algoritmaların tespiti amaçlanmıştır. Çalışmada 8 farklı makine öğrenmesi algoritması kullanılmıştır. Buharlaşmayı etkileyen birçok meteorolojik değişken olmasına rağmen, operasyonel çalışmalarda çoğunlukla kolay bir şekilde ulaşılabilen ve daha sık ölçüm noktası bulunan hava sıcaklığı değişkeni bağımsız değişken olarak kullanılmıştır. Doğu Akdeniz Bölgesi'nde yer alan 13 farklı meteoroloji gözlem istasyonunda (MGİ) istasyon bazında aynı zaman aralığında ölçülmüş aylık ortalama sıcaklık ve aylık toplam buharlaşma değerleri ile modeller kurulmuştur. Değişen MGİ konumlarına göre farklı makine öğrenmesi algoritmaları daha iyi başarı göstermiştir. Dörder istasyonda AdaBoost ve Lineer Regresyon, iki istasyonda Random Forest ve kalan birer istasyonlarda ise K-Nearest Neighbors, Gradient boosting ve Support Vector Machines algoritmaları en iyi tahminleri yapmışlardır. Bununla beraber, en iyi tahminleri yapan modellerin hata metrikleri arasında çok düşük farklar olduğu tespit edilmiştir. İstasyonlar bazında belirlenen en iyi modellerden en düşük kök ortalama karesel hata (RMSE) hatasına (16,4 mm) AdaBoost modeli ile Erdemli MGİ, en yüksek RMSE hatasına (39,2 mm) ise Rastgele Orman modeli ile Antakya MGİ'nun sahip olduğu sonucuna varılmıştır. Makine öğrenmesi algoritmalarının sıcaklık parametresi kullanılarak buharlaşma tahminleri yapan modeller kurulmasında kullanılabilir olduğu belirlenmiştir.

Anahtar Sözcükler
Buharlaşma, Makine Öğrenmesi, Su Kaynakları, Rezervuar İşletmesi

Machine Learning–Based Evaporation Models for the Eastern Mediterranean Region

Abstract

Evaporation is a key meteorological variable influencing water resources planning, reservoir operation, and agricultural productivity. Accurate estimation of evaporation is therefore essential for efficient water resources management. This study investigates the capability of machine learning (ML) algorithms to predict evaporation and identifies the most effective algorithms through a comparative performance assessment. Eight different ML algorithms were implemented to estimate evaporation using air temperature as the sole predictor variable. Although evaporation is controlled by several meteorological factors, air temperature was selected due to its wide availability and dense observational coverage, making it suitable for operational applications.The models were developed using monthly mean air temperature and monthly total evaporation data obtained from 13 meteorological observation stations located in the Eastern Mediterranean Region. The results indicate that the predictive performance of ML algorithms varies across station locations. AdaBoost and Linear Regression provided the best predictions at four stations each, Random Forest performed best at two stations, while K-Nearest Neighbors, Gradient Boosting, and Support Vector Machines achieved the best results at one station each. Despite these variations, the differences among the error metrics of the best-performing models were relatively small. Among the station-based optimal models, the lowest root mean square error (RMSE) (16.4 mm) was obtained using the AdaBoost algorithm at the Erdemli station, whereas the highest RMSE (39.2 mm) was produced by the Random Forest model at the Antakya station. Overall, the findings demonstrate that machine learning algorithms can effectively estimate evaporation using only air temperature as an input variable, highlighting their potential for practical applications in data-limited environments.

Keywords
Evaporation, Machine Learning, Water Resources, Reservoir Operation

1. Giriş

Hidrolojik döngünün önemli bir bileşeni olan buharlaşma, suyun sıvı halden gaz hale geçmesi olarak tanımlanmaktadır ve hidrolojik çevirimin enerji dengesiyle doğrudan ilişkilidir. Yeryüzünde su içeren her yüzeyden, her sıcaklıkta

* Sorumlu Yazar: Tel: +90 (322) 4590590/137 Faks: +90 (322) 4532774
E-posta: tor.mehmet1@gmail.com, kocakr@itu.edu.tr, holbey01@gmail.com, aemre.tekeli@igdir.edu.tr
Gönderim Tarihi / Received : 30/12/2025
Kabul Tarihi / Accepted : 30/03/2026

buharlaşma gerçekleşirken moleküllerin enerjileri artmaktadır. Su yüzeylerinden meydana gelen buharlaşmaya "evaporasyon", bitkilerden terleme ile meydana gelen su kaybına "transpirasyon", ikisinin toplamından oluşan toplam su kaybına ise "evapotranspirasyon" adı verilmektedir. Atmosferdeki su buharının yaklaşık %90'lık kısmını buharlaşma, %10'luk kısmını ise transpirasyon oluşturmaktadır (Savaş, 2023).

Su kaynaklarının verimli kullanılması için buharlaşmanın doğru hesaplanması ve modellenebilmesi oldukça önemlidir (Bozorgi vd., 2020). Buharlaşma birçok meteorolojik değişkenden etkilenmekle beraber, sıcaklık, rüzgâr, nem, radyasyon ve güneşlenme süresi buharlaşmayı doğrudan etkileyen meteorolojik faktörlerdir (Yan vd., 2019).

Buharlaşma farklı yöntemlerle ve cihazlarla ölçülmektedir. Gölgedeki buharlaşma miktarı genellikle Piş Evaporimetresi ve açık alanda buharlaşma miktarı Class A Pan tipi yuvarlak buharlaşma havuzlarıyla ölçülür. Piş Evaporimetre kullanılırken tüp içerisine su konularak seviye kaydedilir. Buharlaşmayla azalan seviye ölçülerek buharlaşma miktarı hesaplanır. Buharlaşmanın ölçümünde sıklıkla kullanılan diğer bir yöntem ise açık su yüzeyinden buharlaşmanın ölçüldüğü buharlaşma havuzlarıdır. Class A Pan tipi yuvarlak buharlaşma havuzlarında, su seviyesindeki değişim hem elle yapılan ölçümlerle hem de elektronik sistemler aracılığıyla belirlenerek, belirli zaman dilimi içinde gerçekleşen buharlaşma miktarı (mm) hesaplanır (Dünya Meteoroloji Örgütü [World Meteorological Organization (WMO)], 2008). Ülkemizdeki meteoroloji gözlem istasyonlarında (MGİ) açık su yüzü buharlaşması Class A Pan tipi buharlaşma havuzları ile ölçülmektedir. 2023 yılı itibariyle ülkemiz genelinde 96 adet buharlaşma istasyonu bulunmaktadır. Yukarıda bahsedilen yöntemler haricinde buharlaşmayı rüzgar hızı, sıcaklık, nem ve güneş radyasyonu değişkenlerini kullanarak ampirik olarak hesaplayan yöntemler de kullanılmaktadır. Ayrıca, sıvı ağırlık değişimi veya basınç ölçümleri ile de buharlaşma hesaplayan cihazlar bulunmaktadır (Verhoef & Campbell, 2005).

Hidrolojide buharlaşma hesapları temel olarak, rezervuar alanlarının işletilmesi ve su kaynaklarının verimli kullanılması için yapılmaktadır. Tarımda ise buharlaşma, bitkilerin su ihtiyaçlarının belirlenmesinde ve tarımsal verimliliğin artmasında önemli bir etkendir. Hidrolojik ve tarımsal faaliyetlerde büyük öneme sahip buharlaşmanın tahmin edilmesinde genellikle ampirik yöntemler kullanılmaktadır. Bu yöntemler meteorolojik değişkenleri kullanarak buharlaşmayı hesaplamaktadır. Penman-Monteith, Thornthwaite ve Hargreaves-Samani bu konuda en sık kullanılan ampirik yöntemlerdir (Ngongondo vd., 2013). Ampirik yöntemler, bölgesel kalibrasyon gerektirmekte ve yerel koşullara bağlı olarak farklı büyüklükte hatalı sonuçlar verebilmektedir (Cruz-Blanco vd., 2025). Bu nedenle buharlaşma tahminlerinde son zamanlarda yeni yöntemler denenmektedir. Bu kapsamda, makine öğrenmesi (MÖ) yöntemleri ve yapay sinir ağları (YSA) ile buharlaşma tahminleri yapılmaktadır. MÖ ve YSA gibi yeni istatistiksel algoritma tabanlı yöntemler ve ampirik yöntemlerin entegre olarak çalıştırılması buharlaşma tahminlerinde yenilikçi yöntemlerin önünü açacaktır (Yaseen vd., 2020).

Hargreaves-Samani, Ritchie ve Turc ampirik yöntemleri ve YSA kullanarak, Taşar vd. (2018); buharlaşma tahminleri yapmışlardır. Çalışma alanı olarak Massachusett, U.S.A (Cambridge Hazne ve havzası) seçilmiştir. Günlük ortalama hava sıcaklığı, ortalama rüzgar hızı, güneşlenme miktarı ve bağıl nem değerleri bağımsız değişkenler olarak seçilerek buharlaşma bağımlı değişkenini tahmin edecek modeller kurulmuştur. Modeller eğitim verileri ile eğitilmiş ve test verileri üzerinde tahminler test edilmiştir. YSA yöntemi geri beslemeli olarak optimize edilmiştir. Hargreaves-Samani, Ritchie ve Turc yöntemleri ile karşılaştırıldığında en düşük hatayla buharlaşma tahminlerini YSA modelinin yaptığı belirlenmiştir.

Gümüş vd. (2018), çalışmalarında Şanlıurfa ve Diyarbakır merkez MGİ verileri kullanılmıştır. Sıcaklık değerleri kullanılarak Class A Pan buharlaşma değerlerini tahmin eden modeller kurulmuştur. YSA, Uyarlamalı Ağ Tabanlı Bulanık Mantık Çıkarım Sistemi (ANFIS) ve Gen Ekspresyon Programlama (GEP) yöntemleri uygulanmış ve hata metrikleri karşılaştırılmıştır. Tüm yöntemlerin birbirine yakın şekilde başarılı sonuçlar verdiği görülmüştür. GEP yöntemi ise diğer yöntemlere göre daha düşük hata sonuçları vererek buharlaşma miktarı tahmininde daha başarılı olmuştur.

Abed vd. (2021), Malezya'da bulunan iki MGİ'ye ait günlük meteorolojik istatistikleri kullanmışlardır. Maksimum sıcaklık, ortalama sıcaklık, minimum sıcaklık, rüzgar hızı, bağıl nem ve güneş radyasyonu gibi iklimsel verilere dayanarak buharlaşma tahmin modelleri oluşturmuşlardır. Stewart ve Thornthwaite ampirik yöntemleri ile Extreme Gradient Boosting, ElasticNet Linear Regression, Long Short-Term Memory (LSTM) algoritmalarını kullanarak tava buharlaşma tahminleri yapmışlardır. Maksimum, ortalama ve minimum sıcaklık değerleri, rüzgar hızı, bağıl nem ve güneşlenme miktarı parametreleri 2000-2019 periyodu için bağımsız değişkenler olarak kullanılmıştır. MÖ algoritmalarının ampirik yöntemlere göre daha iyi tahminler yaptığı belirlenmiştir. LSTM algoritmasının ise en iyi buharlaşma tahminini yaptığı sonucuna varılmıştır. LSTM modeli ile Alor Setar ve Kota Bharu istasyonları için buharlaşma RMSE değerleri 0,166 mm/gün ve 0,074 mm/gün olarak hesaplanmıştır.

Al Sudani ve Salem (2022), Irak'ta bulunan iki MGİ noktasında aylık iklimsel değişkenleri kullanarak buharlaşma miktarını MÖ algoritmaları ile tahmin etmişlerdir. Extreme Learning Machine (ELM), Gradient Boosting Machine (GBM), Quantile Random Forest (QRF) ve Gaussian Process Regression (GPR) algoritmaları ile yapılan tahminler hata metrikleri ile karşılaştırılmıştır. GBM algoritmasının Diyala ve Erbil'de yapılan buharlaşma tahminlerinde diğer modellere kıyasla çok daha iyi performans gösterdiği tespit edilmiştir.

Tezel ve Büyükyıldız (2016), Beyşehir MGİ'den elde edilen sıcaklık, bağıl nem, rüzgâr hızı ve yağış parametrelerini model girdisi kullanarak Class A Pan buharlaşmasını tahmin etmişlerdir. Modeller, YSA ve destek vektör makinesi yöntemleri (SVR) ile kurulmuştur. YSA yöntemleri olarak çok katmanlı algılayıcı (MLP) ve radyal taban fonksiyon ağı (RBFN) yöntemleri kullanılmıştır. MLP yöntemi içerisinde geri yayılım (GDX), Levenberg–Marquardt (LVM),

ölçeklenmiş konjugat gradyan (SCG) ve dirençli geri yayılım (RBP) yöntemleri kullanılmıştır. Ayrıca, Romanenko ve Meyer ampirik yöntemleriyle de tahminler yapılmıştır. Modeller 10 katlı çapraz doğrulama ile regresyon hata metrikleri kullanılarak test edilmiştir. Tüm yöntemler Romanenko ve Meyer ampirik yöntemlerinden daha iyi performans göstermiştir. En iyi performansı ise MLP yöntemi için kullanılan ölçeklenmiş konjugat gradyan (SCG) değeri Determinasyon katsayısı (Coefficient of Determination, $R^2 = 0.905$) vermiştir. Ortalama mutlak hata (MAE) ve kök ortalama karekök hatası (RMSE) değerlerine göre eğitim verilerinde en iyi tahmin performansını LVM yöntemi (MAE = 10.721 mm/ay ve RMSE = 14.621 mm/ay), test verilerinde ise SVR yöntemi (MAE = 17.485 mm/ay ve RMSE = 22.647 mm/ay) göstermiştir.

Jasmine vd. (2022), Amerika Birleşik Devletleri'nin Arizona eyaletinde bulunan bir MGİ tarafından 2010-2017 yılları arasında ölçülen aylık bazda maksimum, minimum ve ortalama hava sıcaklığı, güneşlenme süresi, rüzgâr hızı ve bağıl nem verileri kullanarak MÖ yöntemleri ile buharlaşma tahminleri yapmışlardır. Maksimum, minimum ve ortalama hava sıcaklığı, güneşlenme süresi, rüzgâr hızı ve bağıl nem parametreleri modele girdi olarak verilmiş ve buharlaşma değeri çıktı olarak alınmıştır. Model olarak ANFIS kullanılmıştır. ANFIS; Genetik Algoritma (GA), Ateşböceği Algoritması (FFA) ve Parçacık Sürüsü Optimizasyonu (PSO) algoritması ile optimize edilen üç farklı şekilde uygulanmıştır. Modellere ait hata metrikleri R^2, RMSE, açıklanan varyans (Variance Accounted For, VAF) ve dağılım indeksi (Scatter Index, SI) ile hesaplanmıştır. ANFIS-PSO modelinin en iyi buharlaşma tahminlerini yaptığı ($R^2 = 0.99$, VAF = %98.85, RMSE = 9.73 mm/ay, SI = 0.05) sonucuna varılmıştır. Ancak, ANFIS modelinin de ($R^2 = 0.99$, VAF = %99.04, RMSE = 8.92 mm/ay, SI = 0.05) ANFIS-PSO hata metriklerine çok yakın değerler verdiği görülmüştür. ANFIS modelinin daha kolay uygulanabilir olması nedeniyle bu modelin buharlaşma tahminlerinde kullanımına öncelik verilmesi önerilmiştir.

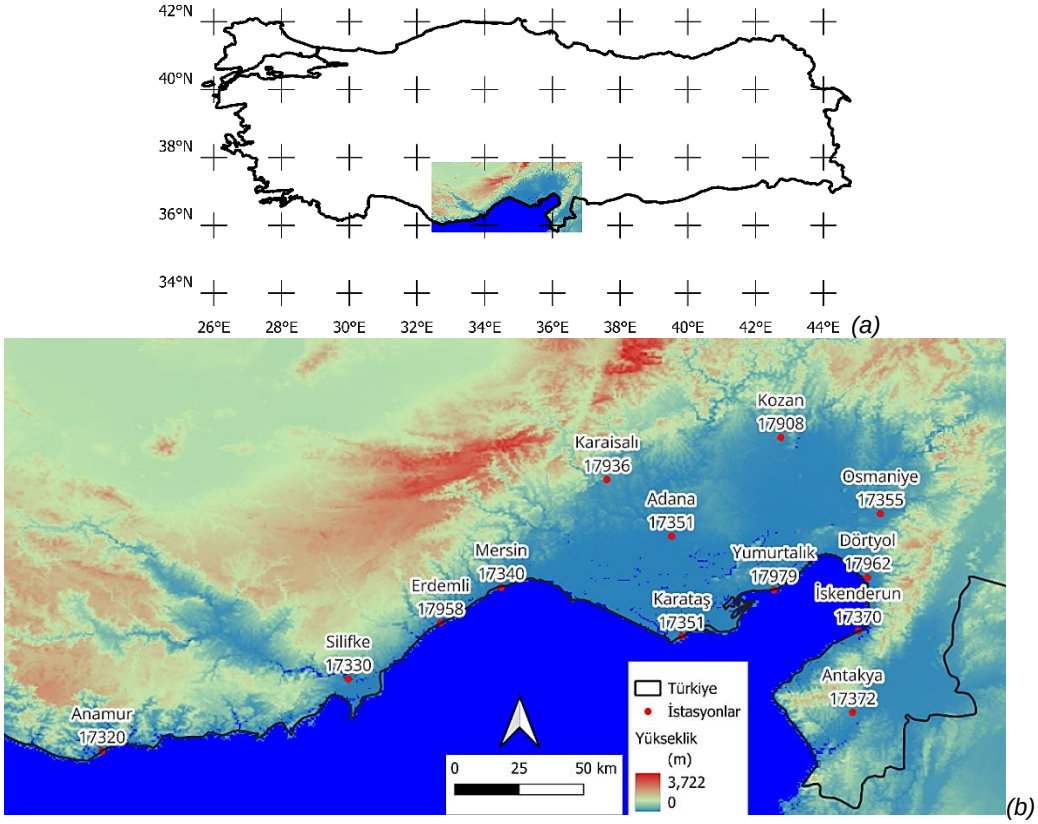
Yukarıda bahsedilen çalışmalarda kullanılan meteorolojik değişkenlerin tümünün güvenilir ve sürekli şekilde ölçülmesi operasyonel olarak çok olası değildir. Ayrıca, Devlet Su İşleri (DSİ) Mühendislik Hidrolojisi Teknik Şartnamesi (2006) ve Hidrojeolojik Etüt Şartnamesi'nde (2019) buharlaşma hesapları, buharlaşma ve sıcaklık arasındaki ilişki kullanılarak yapılmaktadır. Bu kapsamda, Class A Pan tipi buharlaşma havuzlarından ölçülen değerler ve aynı noktada ölçülen sıcaklık değerleri için kurulan regresyon denklemi kullanılarak proje yerinde ölçülen veya proje yerine en yakın MGİ'den kot ve enlem düzeltmesi yapılarak taşınan sıcaklık değerlerinden proje yerindeki buharlaşma değerleri hesap edilmektedir (Devlet Su İşleri Genel Müdürlüğü, 2006; 2019). Hidroloji uygulamalarında proje yerlerinde uzun süreli meteorolojik ölçümler çoğunlukla bulunmadığından DSİ gibi operasyonel çalışmalar yapan kurumlara, bu yaklaşım, sıcaklık verilerinin hem yaygın ölçülmesi hem de uzun dönem sürekliliği açılarından pratik bir çözüm sunmaktadır.

Bu bağlamda, çalışmada Akdeniz Bölgesi'nin doğusunda yer alan 13 farklı MGİ'deki uzun yıllar boyunca ölçülmüş olan aylık ortalama sıcaklık ve aylık toplam buharlaşma değerleri kullanılmıştır. Aylık ortalama sıcaklık ve aylık toplam buharlaşma verileriyle her istasyon için 8 farklı MÖ algoritması kullanılarak buharlaşma modelleri kurulmuştur. Sıcaklık değişkeni bağımsız değişken buharlaşma değerleri bağımlı değişken olarak belirlenmiş ve regresyon modelleri çalıştırılmıştır. Sonuçlar hata metrikleri üzerinden değerlendirilerek her bir istasyon için en iyi buharlaşma tahmini veren modeller belirlenmiştir.

Çalışma; a) Doğu Akdeniz iklim kuşağında çoklu istasyonda uzun yıllar veri setlerini içeren bir çalışma sunması, b) modellerde hiperparametre optimizasyonu içermesi c) özellikle DSİ gibi operasyonel kurumlar için sıcaklık tabanlı basit ama güvenilir buharlaşma modeli önermesi açılarından özgündür. Çalışma, uygulamaya dönük sonuçlar sunarak, rezervuar alanlarının işletilmesi, su kaynaklarının verimli kullanılması ve tarımsal verimliliğin artması konularında veri tabanlı yaklaşımların yaygınlaştırılmasına katkı sağlamaktadır.

1.1. Çalışma alanı

Çalışma alanı, Doğu Akdeniz Bölgesi sınırları içerisindeki 13 farklı MGİ'yi içermektedir (Şekil 1). Çalışma alanı, tipik Akdeniz iklimi gösterip, yazlar kurak ve sıcak, kışlar yağışlı ve ılık geçmektedir. Çalışma alanında yüksek yıllık ortalama sıcaklıklar olması ve uzun güneşlenme sürelerinin bulunması, bölgedeki buharlaşma miktarlarının Türkiye ortalamasının üzerinde olmasına neden olmaktadır. Bu durumlardan dolayı çalışma alanında su kaynaklarının optimum kullanımı ve rezervuar işletimi açılarından buharlaşmanın doğru tahmin edilmesi büyük önem arz etmektedir. Çalışmada kullanılmış MGİ'lere ait temel bilgiler Tablo 1 de verilmiştir.



Şekil 1: Çalışma Alanının Türkiye'deki Konumu (a) Çalışma Alanındaki MGİ Konumları ve Çalışma Alanının Sayısal Yükseklik Modeli (b)

Tablo 1: MGİ kodları, adları ile Sıcaklık ve Buharlaşma veri süreleri

İstasyon		**Veri Tipi ve Süresi**	
Kodu	Adı	Sıcaklık	Buharlaşma
17351	Adana	1930-2024	1948-2024
17936	Karaisalı	1964-2024	1993-2012
17351	Karataş	1963-2024	1986-2012
17908	Kozan	1959-2024	1992-2012
17979	Yumurtalık	1964-2024	1986-2012
17320	Anamur	1948-2024	1986-2024
17958	Erdemli	1963-2024	1974-2012
17340	Mersin	1940-2024	1986-2024
17330	Silifke	1930-2024	1951-2024
17355	Osmaniye	1986-2024	2004-2024
17372	Antakya	1940-2024	1962-2024
17962	Dörtyol	1930-2024	1971-2012
17370	İskenderun	1940-2024	1986-2024

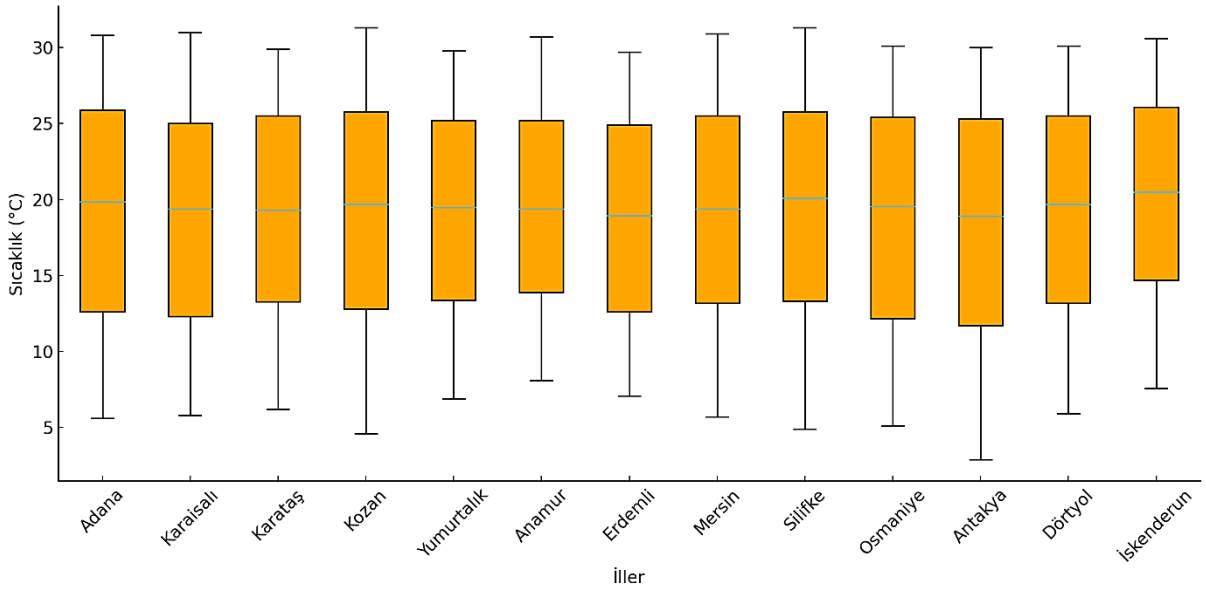
MGİ'lerin denizden yükseklikleri ve coğrafi konumları farklılık göstermektedir. Bu durum, bölgesel sıcaklık ve buharlaşma davranışlarının çeşitlenmesine neden olmaktadır. Bundan dolayı, her bir MGİ için modeller ayrı ayrı kurulmuş ve değerlendirilmiştir. Ayrıca, MGİ'nin dağılımı farklı mikroklimatik etkilerin model performanslarına nasıl yansıdığını inceleme açısından önemli bir çeşitlilik sağlamaktadır.

2. Materyal ve Metot

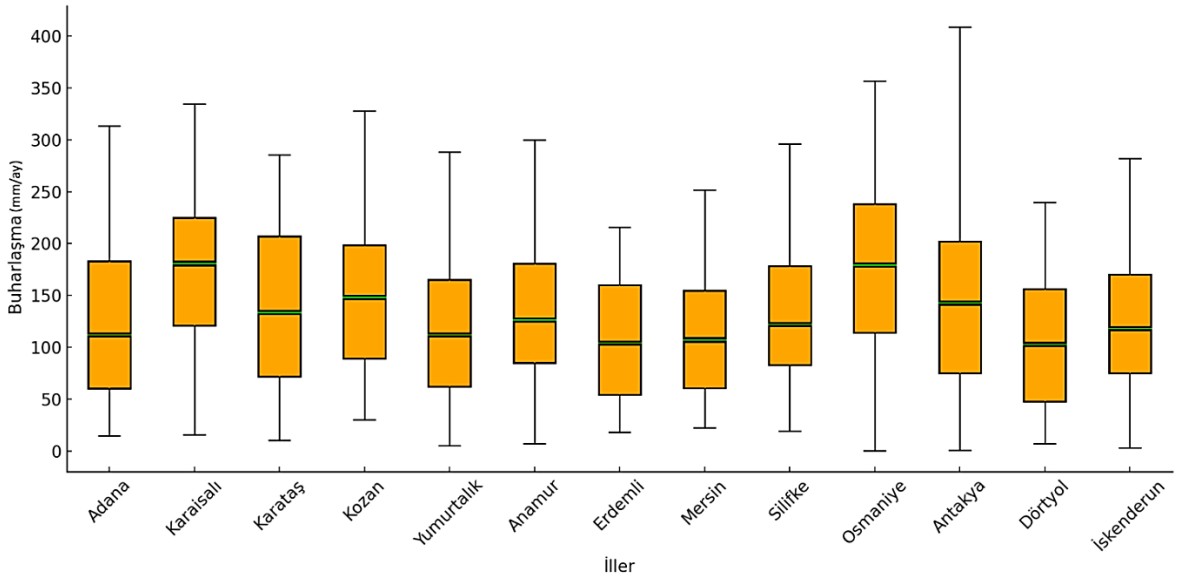
Çalışmada kullanılan veriler, MÖ algoritmaları, algoritmaların uygulanması ve hata metrikleri aşağıdaki bölümlerde detaylı olarak verilmiştir.

2.1. Veriler

Çalışmada kullanılan 13 adet MGİ'de sıcaklık ve buharlaşma ölçümleri yapılmaktadır. İstasyonlara ait aylık ortalama sıcaklık ve aylık toplam buharlaşma verilerinin tanımlayıcı istatistikleri Şekil 2 ve 3'te verilmiştir.



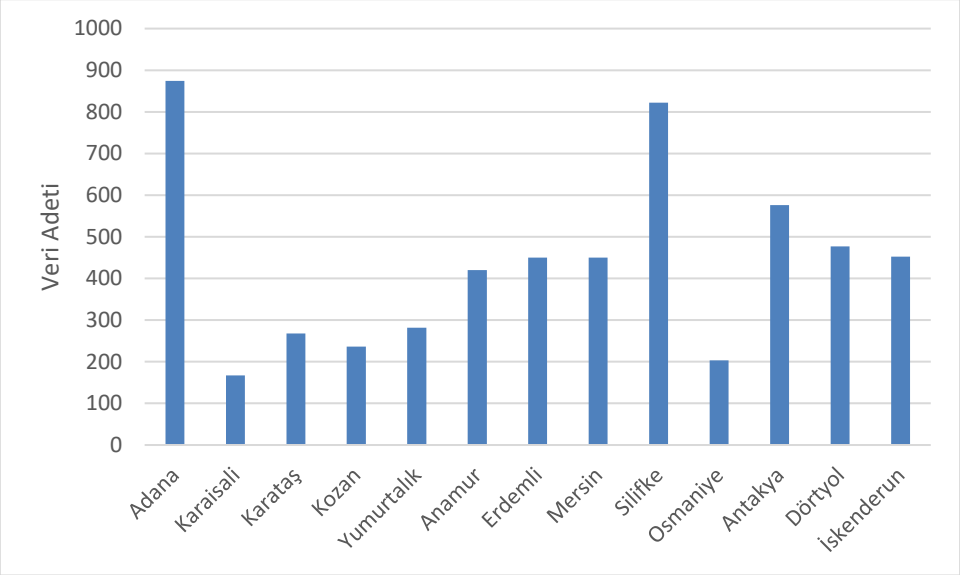
Şekil 2: MGİ Aylık Ortalama Sıcaklık Verileri Kutu Grafikleri



Şekil 3: MGİ Aylık Toplam Buharlaşma Verileri Kutu Grafikleri

İstasyon bazında zamansal olarak sıcaklık ve buharlaşma değerlerinin ortak olduğu dönemler kullanılmıştır. Cihazların çalışmamasından dolayı ölçülemeyen eksik veriler için bir tamamlama yöntemi uygulanmamış ve eksik veri bulunan satırlar veri setinden çıkarılmıştır. Çalışmada çeyrekler arası aralık ve kutu grafik yöntemiyle aykırı değer analizleri yapılmış ancak aykırı değerlerin doğal varyasyonları temsil ettiği görülmüştür. Bu durum Wada (2020)'de belirtilen, bazı veri setlerinde aykırı değerler doğal durumları temsil edebilir ve bu değerlerin olduğu gibi kullanılması daha doğru sonuçlar verir, bulgusu ile uyumludur. Bu kapsamda, veri setinin doğallığının korunması amacıyla bu değerler

olduğu gibi kullanılmıştır. Her bir istasyon için kullanılan veri sayıları Şekil 4'te gösterilmiştir. Ortak zaman periyoduna sahip veri setleri %70 eğitim ve %30 test seti olarak ayrılmıştır (Géron, 2019).



Şekil 4: Çalışmada MGİ Bazında Kullanılan Sıcaklık ve Buharlaşma Veri Seti Adetleri

2.2. Makine Öğrenmesi Yöntemleri ve Kütüphaneler

Çalışmada sekiz farklı; Linear Regression (Lineer Regresyon), K-Nearest Neighbors (K-en Yakın Komşu), Support Vector Machine (Destek Vektör Makinesi), Decision Tree (Karar Ağacı), Random Forest (Rastgele Orman), Gradient Boosting (Gradyan Artırma), AdaBoost (Adaptif Artırma), XGBoost (Ekstrem Gradyan Artırma) MÖ algoritmaları kullanılmıştır. Modeller, Python Scikit-learn ve Xgboost kütüphaneleri kullanılarak hazırlanmıştır.

MÖ yöntemleri son yıllarda hidrolojik süreçlerin modellenmesinde yaygın olarak kullanılmaktadır. Bu çalışmada kullanılan MÖ yöntemleri, hidrolojik tahmin problemlerinde farklı avantajlar sunmaktadır. Linear Regression (Lineer Regresyon) modeli hidrolojik değişkenler arasındaki temel ilişkilerin belirlenmesinde ve referans model olarak sıklıkla kullanılmaktadır. K-Nearest Neighbors (K-en Yakın Komşu) yöntemi benzer meteorolojik koşullara sahip gözlemleri dikkate alarak yağış-akış ilişkilerinin ve akım değerlerinin tahmin edilmesinde kullanılabilmektedir. Support Vector Machine (Destek Vektör Makinesi), doğrusal olmayan ilişkileri modelleyebilme yeteneği sayesinde akım tahmini, yağış tahmini ve buharlaşma modellemelerinde kullanılmaktadır. Decision Tree (Karar Ağacı) tabanlı yöntemler ise veri setini karar kurallarına göre bölerek hidrolojik değişkenlerin modellenmesine olanak sağlamaktadır. Bu yaklaşımın geliştirilmiş bir versiyonu olan Random Forest (Rastgele Orman) algoritması çok sayıda karar ağacının bir araya getirilmesiyle daha yüksek tahmin doğruluğu sağlamaktadır ve akım tahmini, buharlaşma hesapları ve su kalitesi modellemelerinde yaygın olarak kullanılmaktadır. Gradient Boosting (Gradyan Artırma) ve AdaBoost (Adaptif Artırma) gibi artırma (boosting) algoritmaları ise zayıf öğrenicilerin ardışık olarak eğitilmesi prensibine dayanmakta ve model hatalarını azaltarak tahmin performansını artırmaktadır. Gradient Boosting (Gradyan Artırma) algoritmasının gelişmiş bir versiyonu olan XGBoost (Ekstrem Gradyan Artırma) ise düzenlileştirme ve paralel hesaplama gibi özellikleri sayesinde yüksek doğruluk ve hesaplama verimliliği sağlayarak akım tahmini, yağış-akış modellemesi ve buharlaşma tahmini gibi birçok hidrolojik uygulamada başarılı sonuçlar vermektedir. Bu özellikleri nedeniyle söz konusu MÖ yöntemleri hidrolojik süreçlerin modellenmesi ve tahmini amacıyla literatürde yaygın olarak kullanılmaktadır (Mosavi vd., 2018). Her bir modelin çalışma prensibi aşağıda özetlenmiştir.

2.2.1. Linear Regression (Lineer Regresyon)

Temel olarak, bağımlı bir değişken ile bir veya daha fazla bağımsız değişken arasındaki sadece doğrusal ilişkinin modellenmesinde kullanılır. En düşük hatayı veren model, veri setini en iyi temsil eden doğrudur (Shen & Zhang, 2018).

2.2.2. K-Nearest Neighbors (K-en Yakın Komşu)

Veri kümesindeki komşuluklara bağlı olarak modelleme yapılır. Öncelikle her veri noktasının birbiri arasındaki mesafe belirlenir. Daha sonra veri noktası etrafındaki belirlenen komşu sayısı kadar en yakın komşunun ortalaması hesaplanır. Sınıflandırma uygulamalarında ise ortalama yerine en çok tekrar eden sınıf değeri belirlenir. Kullanımı kolaydır ancak büyük veri setlerinde kullanmaya uygun değildir (Kohli vd., 2020).

2.2.3. Support Vector Machine (Destek Vektör Makinesi)

Hem doğrusal hem de doğrusal olmayan veri setlerini modelleyebilir. Verileri sınırlarla böler. Bu sınıra hiper düzlem adı verilir. Verilerin bu sınıra en yakın olduğu durum modelin en düşük hataya sahip olduğu durumdur. Destek vektörleri ile kritik veri noktalarını kullanarak verilerin belirli bir aralıkta kalması hedeflenir. Büyük veri setlerinde ve karmaşık ilişkilerin modellenmesinde kullanılabilir (Zhang & O'Donnell, 2020).

2.2.4. Decision Tree (Karar Ağacı)

Verileri bir ağacın dalları ve yaprakları gibi modeller. Veri seti içerisinde kurallar belirleyerek bölümlemeler yapar. Bu şekilde ağacın dalları oluşur. Yapraklar ise sonuç değerlerdir. Aşırı öğrenmeye dirençlidir ve hiperparametre optimizasyonları ile daha doğru sonuçlar vermesi sağlanabilir. Görsel olarak tanımlanması ve anlaşılması kolay bir algoritmadır (Pekel, 2020).

2.2.5. Random Forest (Rastgele Orman)

Birçok karar ağacının bir araya gelmesiyle oluşur. Bu algoritma her ağacı eğitim verisinden aldığı rastgele değerlerle eğitir. Sonuçta tüm ağaçların yaptığı tahminlerin ortalaması alınarak sonuç değerleri hesaplanır. Her ağaç rastgele farklı özellikleri modelleyerek çeşitliliği arttırır. Aşırı öğrenmeye dirençlidir. Optimizasyon teknikleriyle genellenebilirliği arttırılır (Jain vd., 2021).

2.2.6. Gradient Boosting (Gradyan Artırma)

Gradient Boosting, modellerin art arda eğitilmesi ile oluşan bir topluluk algoritmasıdır. Modeller kendinden önceki modellerin hatalarından ders çıkarır. Bu şekilde her model daha az hataya sahip olur. Aşırı öğrenmeye karşı dirençli ve yüksek doğruluğa sahip modeller oluşturur. Hiperparametre değerleri optimize edilerek doğruluk arttırılabilir. Birçok model kurulması gerektiği için büyük veri setlerinde yavaş çalışır (Singh vd., 2021).

2.2.7. AdaBoost (Adaptif Artırma)

AdaBoost algoritması, hata değerleri yüksek modelleri belirler. Bu modellerden bir topluluk oluşturur. Topluluk içerisinde her iterasyonda zayıf tahmincilerin ağırlıklarını azaltarak eğitir. Bu şekilde daha düşük hataya sahip topluluk modeli oluşturur (Ileberi vd., 2021).

2.2.8. XGBoost (Ekstrem Gradyan Artırma)

Gradient Boosting algoritmasının optimize edilerek geliştirilmiş versiyonudur. Hesaplama hızı artmıştır. Model performansını artıran optimizasyonlar yapılmıştır (Xie & Zhang, 2021). L1 (Lasso Regularization) ve L2 (Ridge Regularization) cezalandırma yöntemleri ağaçların yaprak ağırlıklarını derecelendirerek modele etkilerini düzenler. Bu sayede modelde aşırı öğrenme durumları kontrol altına alınarak modelin genellenebilirliği artırılır. Büyük veri setleri ile verimli şekilde kullanılabilir (Chen & Guestrin, 2016).

2.3. Hata Metrikleri

Çalışmada eğitim verileri ile eğitilen veriler test verileri ile test edilmiştir. Tahmin verileri ve test verileri ortalama mutlak hata (MAE), ortalama kare hatası (MSE), kök ortalama karekök hatası (RMSE) ve determinasyon katsayısı (R^2) hata metrikleri ile test edilmiştir. Eşitlik 1-4 de verilen denklemlerde;

n: veri sayısını,
y_i: gerçek değerleri,
$\hat{y}_i$: model tahmin değerleri,
$\bar{y}$: gerçek değerler ortalamasını göstermektedir.

2.3.1. Mean Absolute Error (MAE) (ortalama mutlak hata)

Tahmin değerleri ve gerçek verilerin farklarının mutlak toplamları hesaplanır. Bu değer, veri sayısına bölünerek "ortalama mutlak hata" hesaplanır (Willmott & Matsuura, 2005).

$$MAE = (1/n) \sum |y_i - \hat{y}_i| \quad (1)$$

2.3.2. Mean Squared Error (MSE) (ortalama kare hatası)

Tahmin değerleri ve gerçek değerlerin farkları hesaplanır. Bu farkların kareleri alınarak toplamları bulunur. Bu değer, veri sayısına bölünerek "ortalama kare hatası" elde edilir (Harville & Jeske, 1992).

$$MSE = (1/n)\sum(y_i - \hat{y}_i)^2 \quad (2)$$

2.3.3. Root Mean Squared Error (RMSE) (kök ortalama karekök hatası)

Ortalama kare hatası değerinin karekköküdür. Büyük hataları verilerin mertebesine yakın şekilde öne çıkarır (Karunasingha, 2022).

$$RMSE = \sqrt{(1/n)\sum(y_i - \hat{y}_i)^2} \quad (3)$$

2.3.4. Coefficient of Determination (R^2) (determinasyon katsayısı)

Determinasyon katsayısı, modelin hedef değişkenin varyansını ne kadar iyi açıkladığını gösterir. 0 ve 1 arasında değerler alır. Determinasyon katsayısı 1'e yaklaştıkça model performansı artmaktadır (Barrett, 1974).

$$R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(yi - \bar{y})^2} \quad (4)$$

2.4. Makine Öğrenmesi Yöntemlerinin Uygulanması

MÖ algoritmaları, her MGİ için ayrı ayrı modeller kurularak kullanılmıştır. Class A Pan su yüzeyinden gerçekleşen aylık toplam buharlaşma verileri bağımlı değişken, aylık ortalama sıcaklık değerleri ise bağımsız değişken olarak belirlenmiştir. Çalışmada kullanılan Linear Regression (Lineer Regresyon, LR), K-Nearest Neighbors (K-en Yakın Komşu, KNN), Support Vector Machine (Destek Vektör Makinesi, SVM), Decision Tree (Karar Ağacı, DT), Random Forest (Rastgele Orman, RF), Gradient Boosting (Gradyan Artırma, GB), AdaBoost (Adaptif Arttorma, AB), XGBoost (Ekstrem Gradyan Artırma, XGB) algoritmaları için istasyon bazında kurulan her modelde hiperparametre optimizasyonları yapılmıştır. Bu sayede en düşük hata değerlerini veren genellebilirliği yüksek modeller elde edilmiştir. Lineer regresyon modelleri için hiperparametre optimizasyonu yapılmamıştır.

Hiperparametre optimizasyonlarında; 5 katlı çapraz doğrulama yapılmıştır. Öncelikle, Python Scikit-learn kütüphanesinde yer alan GridsearchCV modülü ile eğitim verileri için bölümlemeler yapılmıştır. Ancak çalışmamızdaki veriler zaman serisi verileri olduğu için bu bölümlemenin TimeSeriesSplit modülündeki "n_split" ile zaman serisi sıralamasına uygun olarak yapılmasına karar verilmiştir. TimeSeriesSplit modülü ile yapılan çapraz doğrulamanın daha düşük hata metrikleri verdiği görülmüştür. Bu nedenle, zaman serisi bağımlılığını bozmayan bu modülle çapraz doğrulama yapılarak en iyi hiperparametre değerleri elde edilmiştir (Tablo 2). Bu değerlerle son modeller kurularak hata metrikleri hesaplanmıştır.

Tablo 2: MÖ Modelleri İçin Belirlenen Başlangıç Hiperparametre Aralıkları

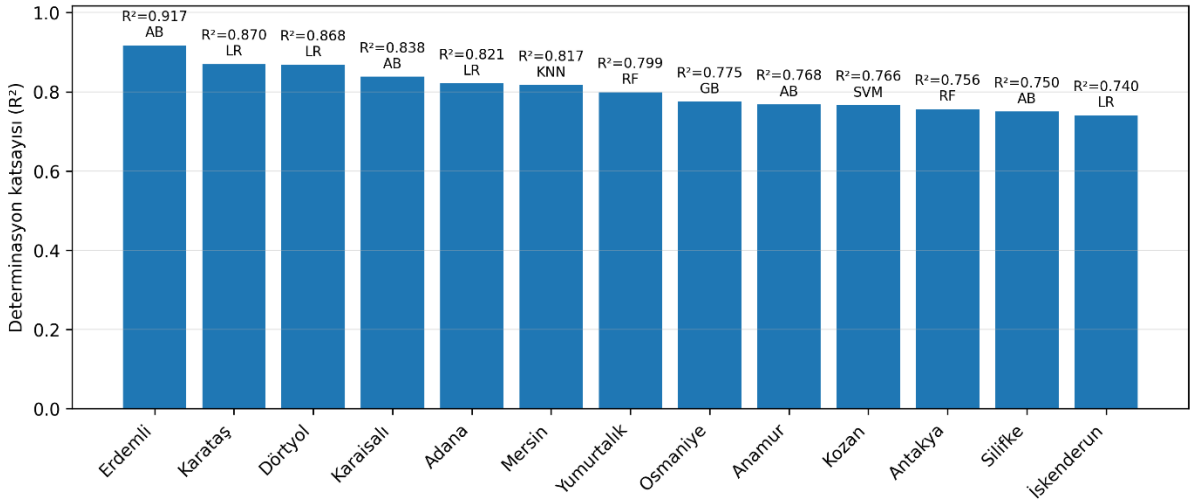
Modeller	Başlangıç Hiperparametre Aralıkları
Random Forest	{'n_estimators': [10, 50], 'max_depth': [2, 10]}
Support Vector Machine	{'C': [0.1, 1], 'gamma': [0.01, 0.1], 'kernel': ['linear', 'rbf']}
K-Nearest Neighbors	{'n_neighbors': [3, 5], 'weights': ['uniform', 'distance']}
Decision Tree	{'max_depth': [5, 10], 'min_samples_split': [2, 5]}
Gradient Boosting	{'n_estimators': [10, 50], 'learning_rate': [0.05, 0.1]}
AdaBoost	{'n_estimators': [10, 50], 'learning_rate': [0.5, 1]}
XGBoost	{'n_estimators': [10, 50], 'learning_rate': [0.05, 0.1]}

3. Bulgular

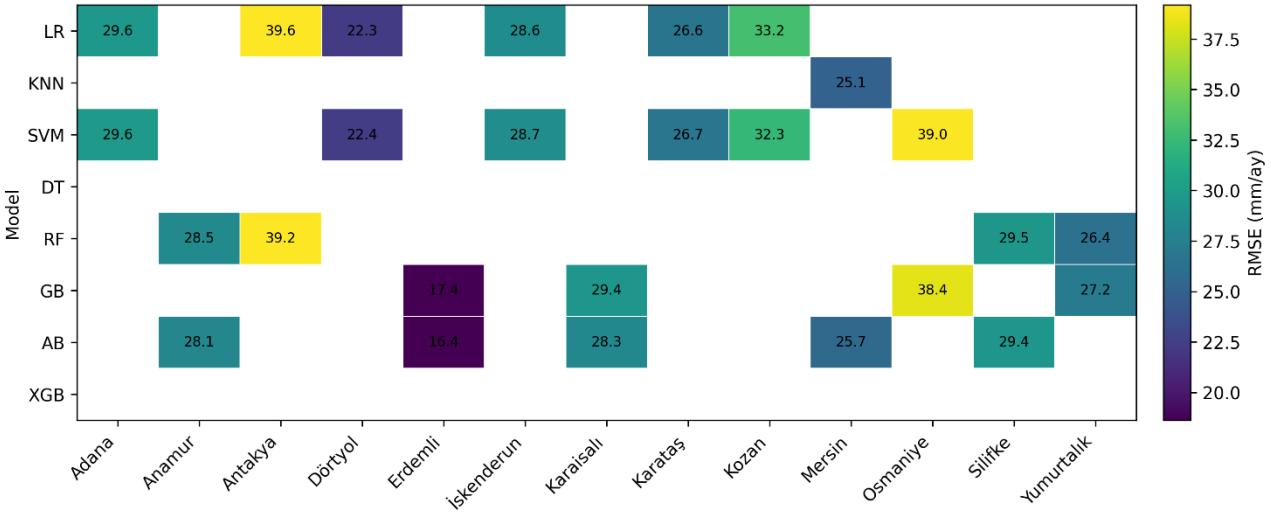
Çalışmada eğitim verileri ile eğitilen ve en iyi hiperparametre değerleriyle kurulan modeller test verileri ile regresyon hata metrikleri kullanılarak test edilmiştir. MGİ bazında en düşük hatayı veren ilk iki modele ait hata metrikleri ve en iyi hiperparametre değerleri Tablo 3'de verilmiştir. Şekil 5'de istasyon bazında en düşük hatayı veren modeller ve R^2 değerleri gösterilmiştir. Şekil 6'da ise Tablo 3'te verilen istasyon bazında diğerlerine kıyasla daha iyi tahmin yapan iki modele ait RMSE hata metrikleri ısı haritası olarak verilmiştir.

Tablo 3: Hata Metrikleri ve Hiperparametre Değerleri ({}: hiperparametre optimizasyonu yapılmamıştır.)

MGİ	Model	MAE (mm/ay)	MSE (mm/ay)	RMSE (mm/ay)	R^2	En İyi Hiperparametre
Adana	LR	22,81	874,49	29,57	0,821	{}
	SVM	22,84	879,30	29,65	0,820	{'C': 1, 'gamma': 0.01, 'kernel': 'linear'}
Anamur	AB	22,03	791,74	28,14	0,768	{'learning_rate': 0.5, 'n_estimators': 10}
	RF	22,40	810,67	28,47	0,762	{'max_depth': 2, 'n_estimators': 50}
Antakya	RF	28,34	1.538,60	39,23	0,756	{'max_depth': 2, 'n_estimators': 50}
	LR	27,88	1.571,54	39,64	0,750	{}
Dörtyol	LR	17,81	495,31	22,26	0,868	{}
	SVM	17,82	501,29	22,39	0,867	{'C': 0.1, 'gamma': 0.01, 'kernel': 'linear'}
Erdemli	AB	13,05	268,91	16,40	0,917	{'learning_rate': 0.5, 'n_estimators': 10}
	GB	14,29	303,15	17,41	0,906	{'learning_rate': 0.05, 'n_estimators': 50}
İskenderun	LR	21,65	821,02	28,65	0,740	{}
	SVM	21,70	823,45	28,70	0,740	{'C': 1, 'gamma': 0.01, 'kernel': 'linear'}
Karaisali	AB	22,98	798,90	28,26	0,838	{'learning_rate': 0.5, 'n_estimators': 10}
	GB	23,26	867,54	29,45	0,824	{'learning_rate': 0.05, 'n_estimators': 50}
Karataş	LR	21,45	706,52	26,58	0,870	{}
	SVM	21,64	712,00	26,68	0,869	{'C': 1, 'gamma': 0.01, 'kernel': 'linear'}
Kozan	SVM	24,81	1.044,17	32,31	0,766	{'C': 0.1, 'gamma': 0.01, 'kernel': 'linear'}
	LR	26,33	1.099,46	33,16	0,754	{}
Mersin	KNN	18,48	629,52	25,09	0,817	{'n_neighbors': 5, 'weights': 'uniform'}
	AB	19,57	662,15	25,73	0,807	{'learning_rate': 0.5, 'n_estimators': 10}
Osmaniye	GB	30,08	1.471,49	38,36	0,775	{'learning_rate': 0.05, 'n_estimators': 50}
	SVM	30,67	1.524,66	39,05	0,767	{'C': 1, 'gamma': 0.01, 'kernel': 'linear'}
Silifke	AB	23,23	865,11	29,41	0,750	{'learning_rate': 1, 'n_estimators': 10}
	RF	22,91	870,18	29,50	0,748	{'max_depth': 2, 'n_estimators': 10}
Yumurtalık	RF	20,03	696,52	26,39	0,799	{'max_depth': 2, 'n_estimators': 50}
	GB	21,36	737,51	27,16	0,787	{'learning_rate': 0.05, 'n_estimators': 50}

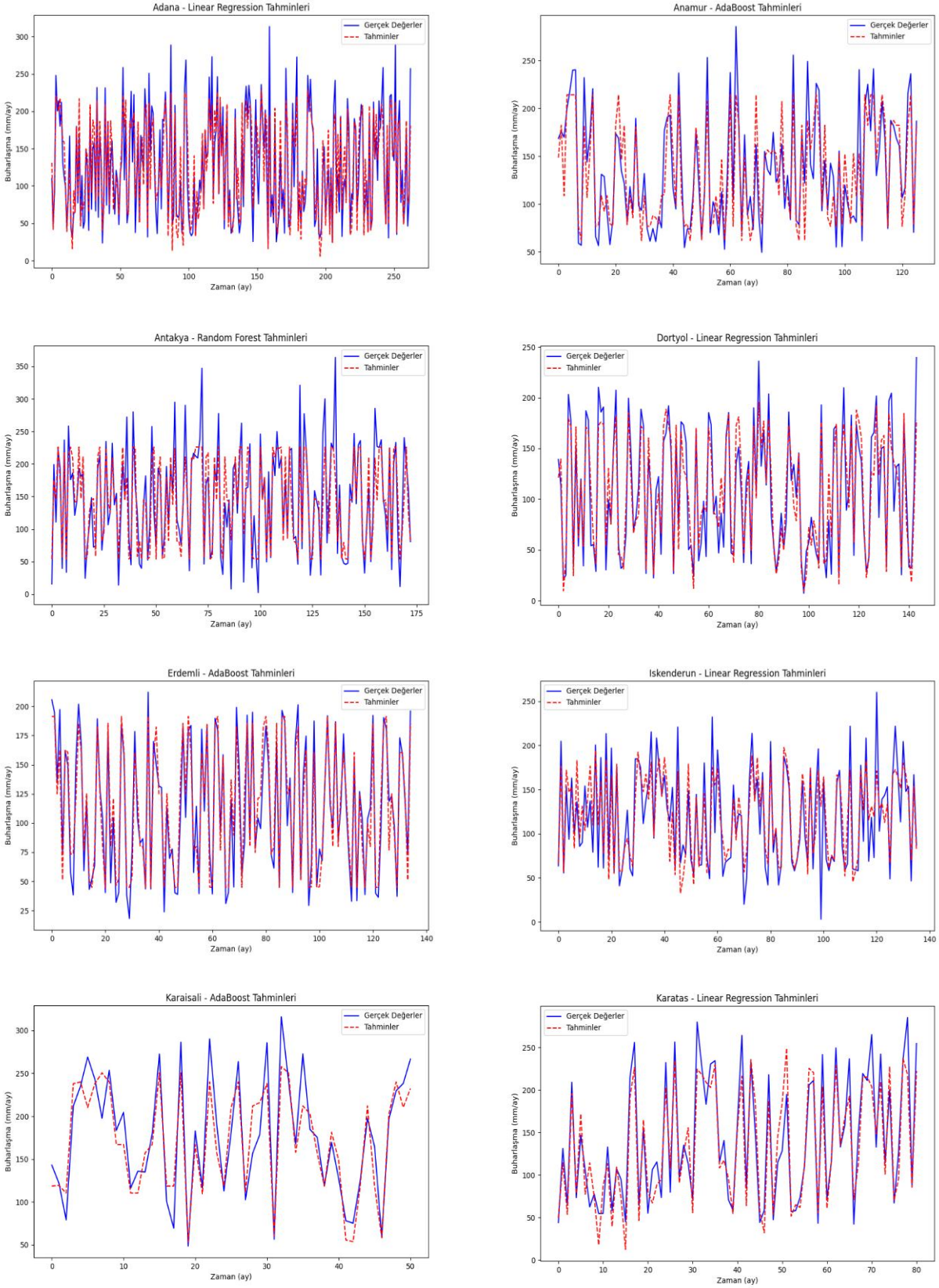


Şekil 5: İstasyon Bazında En Düşük Hataya Sahip Modellerin R² Değerleri

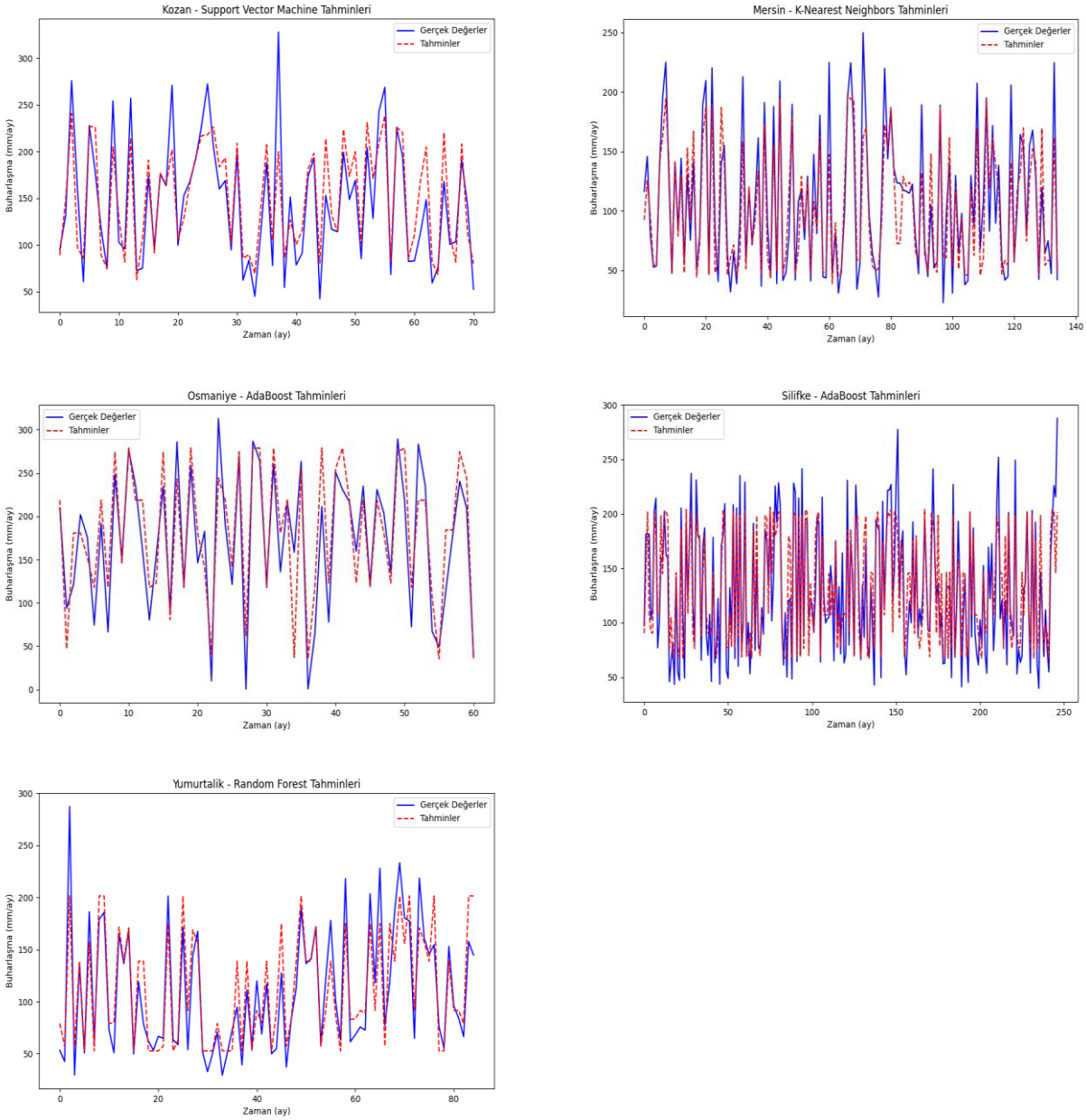


Şekil 6: İstasyon Bazında En İyi Tahminci İki Modelin RMSE (mm/ay) Değerleri

İstasyon bazında en düşük hata metriklerine sahip modellere ait buharlaşma tahmin değerleri ve ölçülen buharlaşma verilerini gösteren grafikler Şekil 7'de verilmiştir.



Şekil 7: İstasyon Bazında En Düşük Hata Metriğine Sahip Model Buharlaşma Tahminleri (mm/ay) ve Ölçülen Buharlaşma Değerleri (mm/ay)



Şekil 7: Devamı

4. Tartışma

Bu çalışmada; 4 istasyonda Lineer Regresyon algoritması, 4 istasyonda AdaBoost algoritması, 2 istasyonda Rastgele Orman algoritması, 1 istasyonda Destek Vektör Regresyonu algoritması, 1 istasyonda En Yakın Komşu algoritması ve 1 istasyonda ise Gradyan Azaltma algoritması en düşük hata metrikleri ile diğer modellere göre daha iyi sonuçlar vermiştir. En iyi ikinci modellere bakıldığında ise Lineer Regresyon algoritması ve Destek Vektör Regresyonu algoritmalarının birbiri ardına geldiği; AdaBoost, Gradyan Azaltma ve Rastgele Orman algoritmalarının da birbirlerini takip ettiği görülmüştür.

Hata metriklerine göre AdaBoost algoritmasıyla en düşük RMSE hatasına 16,4 mm/ay ile Erdemli istasyonu, en yüksek RMSE hatasına ise En Yakın Komşu algoritmasıyla 43,72 mm/ay ile Antakya istasyonunun sahip olduğu sonucuna varılmıştır. İstasyonlar bazında en iyi modeller arasındaki en yüksek RMSE hatasına ise Rastgele Orman algoritmasıyla 39,23 mm/ay ile yine Antakya istasyonu sahiptir. En düşük hataya sahip ilk 2 modelin her istasyonda birbiriyle çok yakın hata metriklerine sahip olduğu tespit edilmiştir. Tüm istasyonlarda en düşük hataya sahip ilk üç model incelendiğinde; %61 oran ile en sık Lineer Regresyon algoritması, %54 oran ile ise Destek Vektör Regresyonu, AdaBoost ve Gradyan Artırma algoritmalarının yer aldığı belirlenmiştir. Lineer Regresyon ve Destek Vektör Regresyonu algoritmalarının düşük hata değerleri sıcaklık ve buharlaşma ilişkisinin çoğunlukla doğrusal olduğunu göstermektedir. AdaBoost ve Gradyan Artırma algoritmalarının düşük hata değerleri ise zayıf tahmincilerden güçlü topluluk oluşturma

özelliklerinden kaynaklanmaktadır. Bazı istasyonlarda daha yüksek hata değerleri elde edilmesinin temel nedeni, istasyonların iklimsel ve topoğrafik açıdan birbirinden farklı olmasıdır. Örneğin Antakya istasyonunda; Amanos Dağları'nın orografik etkileri, nemli bir bölge olması gibi doğrusal ilişkiyi bozan ek faktörlerden dolayı RMSE hata değerleri diğer istasyonlara kıyasla daha yüksektir.

Diğer modellere göre daha iyi sonuç veren Lineer Regresyon ve AdaBoost algoritmalarının model sayısı fazla olsa da tüm buharlaşma tahminlerinde kullanılması için bir algoritma önerilememektedir. Ayaz vd. (2021) Hindistan ve Yeni Zellanda'daki farklı iklim bölgeleri için MÖ yöntemleriyle buharlaşma tahmin modelleri oluşturarak elde ettikleri sonuçlarda olduğu gibi çalışmamızda kullanılan algoritmaların veri setlerine ve bölgesel etkenlere göre farklı performanslar gösterdikleri belirlenmiştir. MÖ algoritmaları farklı MGİ'lerde farklı başarılar gösterdiğinden buharlaşma tahminlerinde tüm algoritmaların denenerek en iyi sonucu veren modelin seçilmesi önerilmektedir. Ancak, modellerin hata metrikleri arasındaki farkların az olması ve kolay uygulanabilir olması nedeniyle Lineer Regresyon modelleri daha sık tercih edilebilir. Lineer regresyon modellerinin zaman maliyetini azaltması ve anlaşılabilir olması dolayısıyla da genellikle lineer ilişkiye sahip sıcaklık ve buharlaşma modellerinde kullanılması avantajlı olabilmektedir.

Çalışmanın sonuçları literatürdeki araştırmalarla uyum göstermektedir. Rai vd. (2022) tarafından 3 MÖ algoritması ve 3 ampirik yöntem karşılaştırması sonucunda MÖ algoritmalarının daha iyi tahminler yapması, buharlaşma tahminlerinde MÖ modellerinin kullanılabilir olduğunu göstermektedir. Yaseen vd. (2020), MÖ algoritmalarının kurak ve yarı kurak bölgelerde ampirik formüllerden daha başarılı sonuçlar ürettiğini göstermiştir. Zouzou ve Çıtakoğlu (2021) tarafından yapılan çalışmada sıcaklık ve güneş radyasyonunun evapotranspirasyon tahminlerinde en etkili değişkenler olduğu bildirilmiştir. Bu çalışmayla MÖ modellerinin buharlaşma tahminlerinde kullanılabilir oldukları ve iyi tahminler yaptıkları sonucuna varılmıştır.

5. Sonuçlar

Çalışmada Doğu Akdeniz Bölgesi'nde yer alan ve buharlaşma rasadı bulunan 13 adet MGİ'nin uzun yıllar boyunca istasyon bazında aynı zaman periyodunda ölçülmüş aylık ortalama sıcaklık ve aylık toplam buharlaşma değerleri kullanılarak buharlaşma tahmininde bulunan MÖ modelleri oluşturulmuştur. Hata metrikleri üzerinden farklı MÖ modellerinin buharlaşma tahmin performansları değerlendirilmiştir. Sadece sıcaklık değişkenini kullanarak buharlaşma tahmini yapan MÖ modellerinin ampirik yöntemlere iyi bir alternatif olabileceği sonucuna ulaşılmıştır.

İstasyonların büyük çoğunluğunda AdaBoost ve Lineer Regresyon (her biri 4 er istasyonda) diğer modellere göre daha düşük hata metriklerini vermiştir. Bu durum, sıcaklık ile buharlaşma arasındaki ilişkinin çoğunlukla doğrusal olduğu bilgisi ile uyumludur. En düşük ve en yüksek RMSE değerleri16,4 mm/ay ve 39,23mm/ay değerleri ile Erdemli MGİ'de AdaBoost ve Antakya MGİ'de Rastgele Orman modelleriyle elde edilmiştir. MÖ modellerinin hata metriklerinin küçük olması çalışma bölgesinde sadece sıcaklık değişkeni yapılacak buharlaşma tahminlerinin güvenilir bir şekilde uygulanabileceğini göstermektedir. Farklı MÖ algoritmalarının hata değerlerinin birbirine yakın olması, buharlaşma tahminlerinde uygulaması daha az maliyetli modellerin tercih edilmesine neden olmaktadır.

Her istasyon için farklı modeller daha iyi sonuç vermiştir. Bu nedenle sıcaklıktan buharlaşma tahminlerinde bir modelin diğer modellerden daha üstün performans sergilediği genellemesi yapılamamaktadır. Modeller arası temel farklılığı MGİ'lerin bulunduğu noktalardaki iklimsel ve topografik farklılıkların oluşturduğu düşünülmektedir. Sıcaklık ve buharlaşma arasındaki lineer ilişki, nem ve rüzgar etkisine açık kıyı bölgelerindeki MGİ'lerde azalmıştır. Nispeten kıyıdan uzak iç kesimlerde sıcaklık artışı doğrudan buharlaşmayı artırmıştır.

Çalışma, Doğu Akdeniz Bölgesi için ilk kapsamlı MÖ tabanlı buharlaşma karşılaştırmasını sunmuş ve MÖ yöntemlerinin mühendislik uygulamalarında operasyonel olarak kullanılabilir olduğunu ortaya koymuştur.

Bundan sonraki çalışmalara ışık tutması için aşağıdaki öneriler verilmiştir:

- Rüzgar yönü ve hızı, bağıl nem ve güneşlenme süresi gibi buharlaşmayı etkileyen diğer meteorolojik değişkenlerin modele girdiler olarak sağlanması MÖ modellerinin doğruluk değerlerini artıracaktır.
- Uzun kısa süreli bellek (LSTM) gibi derin öğrenme algoritmaları ile buharlaşmayı etkileyen meteorolojik değişkenler arasındaki ilişkiler daha iyi çözümlenebilir.
- Günlük ve saatlik ölçümler ile buharlaşma modellerinin zamansal çözünürlüğü artırılabilir.
- Trend analizleriyle iklim değişikliğinin buharlaşma üzerindeki etkileri ortaya konabilir.

Son olarak, çalışmanın farklı iklim özelliklerine sahip bölgelerde de uygulanması ve daha farklı algoritmalar ile denenmesi buharlaşma tahminlerinde yeni yöntemler geliştirilmesine ışık tutacaktır.

Kaynaklar

Abed, M., Imteaz, M. A., Ahmed, A. N., & Huang, Y. F. (2021). Application of long short-term memory neural network technique for predicting monthly pan evaporation. *Scientific Reports, 11*(1), Article 20742. https://doi.org/10.1038/s41598-021-99999-y

Al Sudani, Z. A., & Salem, G. S. A. (2022). Evaporation rate prediction using advanced machine learning models: A comparative study. *Advances in Meteorology, 2022*(1), Article 1433835. https://doi.org/10.1155/2022/1433835

Ayaz, A., Rajesh, M., Singh, S. K., & Rehana, S. (2021). Estimation of reference evapotranspiration using machine learning models with limited data. *AIMS Geosciences, 7*(3), 268-290. https://doi.org/10.3934/geosci.2021016

Barrett, J. P. (1974). The coefficient of determination—some limitations. *The American Statistician, 28*(1), 19-20. https://doi.org/10.1080/00031305.1974.10479056

Bozorgi, A., Bozorg-Haddad, O., Sima, S., & Loáiciga, H. A. (2020). Comparison of methods to calculate evaporation from reservoirs. *International Journal of River Basin Management, 18*(1), 1-12. https://doi.org/10.1080/15715124.2018.1546729

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 785-794). ACM. https://doi.org/10.1145/2939672.2939785

Cruz-Blanco, M., Ortega-Farías, S., & Fuentes-Pérez, D. (2025). Regionalization of the Hargreaves–Samani coefficients to estimate reference evapotranspiration in high-altitude areas. *Atmosphere, 16*(4), Article 408. https://doi.org/10.3390/atmos16040408

Devlet Su İşleri Genel Müdürlüğü. (2006). *Mühendislik hidrolojisi teknik şartnamesi*. https://cdniys.tarimorman.gov.tr/api/File/GetFile/425/KonuIcerik/740/1088/DosyaGaleri/m%C3%BChendislik-hidrolojisi-hizmetleri-teknik-%C5%9Fartnamesi_r00_20061110.pdf

Devlet Su İşleri Genel Müdürlüğü. (2019). *Hidrojeolojik etüt şartnamesi*. https://cdniys.tarimorman.gov.tr/api/File/GetFile/425/KonuIcerik/740/1088/DosyaGaleri/hidrojeolojik-et%C3%BCt-teknik-%C5%9Fartnamesi_r01_20190527-.pdf

Géron, A. (2019). *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems* (2nd ed.). O'Reilly Media.

Gümüş, V., Yeniğün, K., Toprak, Z. F., & Baçi, N. O. (2018). Şanlıurfa ve Diyarbakır istasyonlarında sıcaklık tabanlı buharlaşma tahmininde YSA, ANFIS ve GEP yöntemlerinin karşılaştırılması. *Dicle Üniversitesi Mühendislik Fakültesi Mühendislik Dergisi, 9*(1), 553-562.

Harville, D. A., & Jeske, D. R. (1992). Mean squared error of estimation or prediction under a general linear model. *Journal of the American Statistical Association, 87*(419), 724-731. https://doi.org/10.1080/01621459.1992.10475274

Ileberi, E., Sun, Y., & Wang, Z. (2021). Performance evaluation of machine learning methods for credit card fraud detection using SMOTE and AdaBoost. *IEEE Access, 9*, 165286-165294. https://doi.org/10.1109/ACCESS.2021.3134330

Jain, P., Choudhury, A., Dutta, P., Kalita, K., & Barsocchi, P. (2021). Random forest regression-based machine learning model for accurate estimation of fluid flow in curved pipes. *Processes, 9*(11), Article 2095. https://doi.org/10.3390/pr9112095

Jasmine, M., Mohammadian, A., & Bonakdari, H. (2022). On the prediction of evaporation in arid climate using machine learning model. *Mathematical and Computational Applications, 27*(2), Article 32. https://doi.org/10.3390/mca27020032

Karunasingha, D. S. K. (2022). Root mean square error or mean absolute error? Use their ratio as well. *Information Sciences, 585*, 609-629. https://doi.org/10.1016/j.ins.2021.11.036

Kohli, S., Godwin, G. T., & Urolagin, S. (2020). Sales prediction using linear and KNN regression. In S. Patnaik, X.-S. Yang, & I. K. Sethi (Eds.), *Advances in machine learning and computational intelligence: Proceedings of ICMLCI 2019* (pp. 321-329). Springer Singapore. https://doi.org/10.1007/978-981-15-5243-4_29

Mosavi, A., Ozturk, P., & Chau, K. W. (2018). Flood prediction using machine learning models: Literature review. *Water, 10*(11), Article 1536. https://doi.org/10.3390/w10111536

Ngongondo, C., Xu, C. Y., Tallaksen, L. M., & Alemaw, B. (2013). Evaluation of the FAO Penman–Monteith, Priestley–Taylor and Hargreaves models for estimating reference evapotranspiration in southern Malawi. *Hydrology Research, 44*(4), 706-722. https://doi.org/10.2166/nh.2012.224

Pekel, E. (2020). Estimation of soil moisture using decision tree regression. *Theoretical and Applied Climatology, 139*(3), 1111-1119. https://doi.org/10.1007/s00704-019-03048-8

Rai, P., Kumar, P., Al-Ansari, N., & Malik, A. (2022). Evaluation of machine learning versus empirical models for monthly reference evapotranspiration estimation in Uttar Pradesh and Uttarakhand States, India. *Sustainability, 14*(10), Article 5771. https://doi.org/10.3390/su14105771

Savaş, K. (2023). *Artvin ormanlarında yangınların dağılımı ve büyük yangınların değerlendirilmesi* [Yüksek lisans tezi, Bezmialem Vakıf Üniversitesi]. YÖK Ulusal Tez Merkezi. https://tez.yok.gov.tr/UlusalTezMerkezi

Shen, R., & Zhang, B.-W. (2018). The research of regression model in machine learning field. *MATEC Web of Conferences, 176*, Article 01033. https://doi.org/10.1051/matecconf/201817601033

Singh, U., Rizwan, M., Alaraj, M., & Alsaidan, I. (2021). A machine learning-based gradient boosting regression approach for wind power production forecasting: A step towards smart grid environments. *Energies, 14*(16), Article 5196. https://doi.org/10.3390/en14165196

Taşar, B., Üneş, F., Demirci, M., & Kaya, Y. Z. (2018). Yapay sinir ağları yöntemi kullanılarak buharlaşma miktarı tahmini. *Dicle Üniversitesi Mühendislik Fakültesi Mühendislik Dergisi, 9*(1), 543-551.

Tezel, G., & Büyükyıldız, M. (2016). Monthly evaporation forecasting using artificial neural networks and support vector machines. *Theoretical and Applied Climatology, 124*, 69-80. https://doi.org/10.1007/s00704-015-1392-3

Verhoef, A., & Campbell, C. L. (2005). Evaporation measurement. In M. G. Anderson & J. J. McDonnell (Eds.), *Encyclopedia of hydrological sciences* (pp. 589-601). John Wiley & Sons. https://doi.org/10.1002/0470848944.hsa043

Wada, K. (2020). Outliers in official statistics. Japanese Journal of Statistics and Data Science, 3, 669–691. https://doi.org/10.1007/s42081-020-00091-y

Willmott, C. J., & Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Research, 30*(1), 79-82. https://doi.org/10.3354/cr030079

World Meteorological Organization. (2008). *Guide to meteorological instruments and methods of observation* (WMO-No. 8, 7th ed.).

Xie, D., & Zhang, S. (2021). Machine learning model for sales forecasting by using XGBoost. In *2021 IEEE international conference on consumer electronics and computer engineering (ICCECE)* (pp. 480-483). IEEE. https://doi.org/10.1109/ICCECE51280.2021.9342304

Yan, Z., Wang, S., Ma, D., Liu, B., Lin, H., & Li, S. (2019). Meteorological factors affecting pan evaporation in the Haihe River Basin, China. *Water*, *11*(2), Article 317. https://doi.org/10.3390/w11020317

Yaseen, Z. M., Al-Juboori, A. M., Beyaztas, U., Al-Ansari, N., Chau, K. W., Qi, C., Ali, M., Salih, S. Q., & Shahid, S. (2020). Prediction of evaporation in arid and semi-arid regions: A comparative study using different machine learning models. *Engineering Applications of Computational Fluid Mechanics*, *14*(1), 70-89. https://doi.org/10.1080/19942060.2019.1680576

Zhang, F., & O'Donnell, L. J. (2020). Support vector regression. In A. Mechelli & S. Vieira (Eds.), *Machine learning* (pp. 123-140). Academic Press. https://doi.org/10.1016/B978-0-12-815739-8.00007-9

Zouzou, Y., & Çıtakoğlu, H. (2021). Reference evapotranspiration prediction from limited climatic variables using support vector machines and Gaussian processes. *Avrupa Bilim ve Teknoloji Dergisi*, (28), 346-351. https://doi.org/10.31590/ejosat.999319