



Makine öğrenmesi yöntemleri ile yağış tahmini: Orta ve Doğu Karadeniz Bölgesi örneği

İbrahim Edibali ATALAY^{1,*}, Serdar NESLİHANOĞLU²

¹Lisansüstü Eğitim Enstitüsü, Artvin Çoruh Üniversitesi, Artvin

²Fen Fakültesi İstatistik Bölümü, Eskişehir Osmangazi Üniversitesi, Eskişehir

*Corresponding author: ibrahimatalay3@gmail.com

İbrahim Edibali ATALAY: <https://orcid.org/0000-0002-3842-9086>

Serdar NESLİHANOĞLU: <https://orcid.org/0000-0001-8451-8023>

ESER BİLGİSİ / ARTICLE INFO

Araştırma Makalesi / Research Article

Geliş 29 Nisan 2026 / Received 29 April 2026

Düzeltilmelerin gelişi 2 Temmuz 2026 / Received in revised form 2 July 2026

Kabul 3 Temmuz 2026 / Accepted 3 July 2026

Yayınlanma 9 Ağustos 2026 / Published online 9 August 2026

ÖZET: Küresel ısınmanın bir sonucu olarak bölgesel yağış miktarlarındaki değişimler belirgin biçimde hissedilmektedir. Ülkemizde bu etkinin en belirgin hissedildiği yerler Orta ve Doğu Karadeniz bölgesidir. Bu bölgedeki bölgesel yağış rejimindeki zamansal ve mekânsal değişimlerin araştırılması ve modellenmesi bölgesel su yönetimi ve orman kaynaklarının sürdürülebilirliği açısından stratejik bir öneme sahiptir. Bu amaç doğrultusunda, Orta ve Doğu Karadeniz Bölgesi'ndeki illerin (Artvin, Bayburt, Giresun, Gümüşhane, Rize, Samsun, Tokat ve Trabzon) 2015-2024 yılları arasındaki günlük yağış miktarının modellenmesi ve tahmini araştırılmıştır. Bu süreçte makine öğrenmesi modellerinden CatBoost, XGBoost, LightGBM ve POA-VMD-LSTM'in yağış miktarı modelleme ve tahmin performansları model kıyaslama kriterlerinden MAE ve RMSE ile karşılaştırılmıştır. Sonuç olarak, söz konusu periyotta yağış miktarını en etkin modelleyen LightGBM modeli olurken, CatBoost modelinin gelecek tahmini konusundaki etkin performansı gözlemlenmiştir. Buna ek olarak, makine öğrenmesi modellerinin genel olarak yağış miktarı tahmini konusunda etkin performansa sahip olduğu görülmüştür.

Anahtar kelimeler: Makine öğrenmesi, Orta ve Doğu Karadeniz bölgesi, Yağış tahmini, POA-VMD-LSTM

Rainfall prediction using machine learning methods: The case of the Central and Eastern Black Sea Region

ABSTRACT: The changes in regional precipitation amounts as a result of global warming are being felt in a pronounced manner. The places in our country where this effect is most distinctly felt are the Central and Eastern Black Sea region. Investigating and modeling the temporal and spatial changes in the regional precipitation regime in this area holds strategic importance in terms of regional water management and the sustainability of forest resources. With this objective in mind, the modeling and forecasting of daily precipitation amounts in the provinces of the Central and Eastern Black Sea Region (Artvin, Bayburt, Giresun, Gümüşhane, Rize, Samsun, Tokat, and Trabzon) between the years 2015 and 2024 have been investigated. In this process, the precipitation modeling and forecasting performances of the machine learning models CatBoost, XGBoost, LightGBM, and POA-VMD-LSTM were compared using the model evaluation criteria MAE and RMSE. As a result, LightGBM emerged as the most effective model for modeling precipitation amounts during the specified period, while the effective performance of the CatBoost model in future forecasting was also observed. In addition, it was found that machine learning models generally demonstrate effective performance in precipitation amount forecasting.

Keywords: Machine Learning, Central and Eastern Black Sea Region, Rainfall Prediction, POA-VMD-LSTM

GİRİŞ

Küresel ısınmanın etkileri en belirgin olarak bölgesel yağış miktarlarındaki değişimlerde görülmektedir. Bolat ve Şensoy (2024), iklim değişikliğinin etkilerinin özellikle sıcaklık ve yağıştaki değişimler şeklinde ortaya çıktığını belirtmiştir. Yağış, atmosferdeki su buharının yoğunlaşmasıyla yeryüzüne düşen katı ve sıvı parçacıklar olarak tanımlanır ve su döngüsünün merkezinde yer alır (Temuçin Kılıçer 2025; Salaeh ve ark. 2022). Bu bağlamda, küresel ısınmaya bağlı olarak ortaya çıkabilecek doğa olaylarına karşı strateji geliştirmenin afet yönetimi, su kaynakları yönetimi ve tarım alanlarının kullanımı gibi geniş bir alanda kritik öneme sahip olduğu literatürde ortaya konulmuştur. Örneğin, Dünya Meteoroloji Örgütü (World Meteorological Organisation, WMO)'nın "Küresel İklim Durumu 2022" adlı raporuna göre Afrika'da gerçekleşen yağışların son 40 yıl ortalamasının altında kalmasının Kenya, Somali ve Etiyopya ülkelerinde yaşayan insanları etkileyen, bu ülkelerdeki hayvancılık ve tarıma zarar veren bir kriz yaşandığı belirtilmiştir. Ayrıca, Pakistan'da yaşanan aşırı yağış sebebiyle meydana gelen sel afetinin ise 1700 can kaybının yaşanmasına sebep olduğu ve 33 milyon insanın su baskınlarından etkilendiği belirtilmiştir (WMO, 2023). Avrupa Çevre Ajansı (EEA)'na göre, fırtınalar, sıcak hava dalgaları ve su baskınları gibi aşırı hava olayları son 40 yılda Avrupa'da 85.000 ila 145.000 kişinin ölümüne sebep olduğu ve bu ölümlerin %85'i sıcak hava dalgalarından kaynaklandığını belirtmiştir (EEA, 2024). Thomas (2023) çalışmasında ABD ve Avrupa'da yaşanan sıcak hava dalgaları ve kontrol edilemeyen yangınlar ABD'nin bazı bölgelerini etkileyen seller; Çin'in nehirlerini kurutan sıcak hava dalgası ve kuraklık ve yaz mevsiminden hemen önce yaşanan epik yağışlar, Güney Afrika'daki sel ve toprak kaymaları iklim değişikliğinin etkilerine örnek olarak belirtilmiştir. Batan ve Toprak (2015) ise çalışmasında iklimsel olaylardaki değişimin, ekosistemleri ve biyolojik çeşitliliği, sıcaklığı, yağış rejimini, gıda ve su güvenliğini ve insan sağlığını etkileyeceği vurgulanmıştır. Bu

küresel tablo, ulusal ve bölgesel ölçekte de yağışın yakından izlenmesini ve doğru biçimde tahmin edilmesini zorunlu kılmaktadır.

Türkiye’deki yağışın izlenmesi, eğilimlerinin belirlenmesi ve tahmin edilmesinin afet yönetimi, su kaynakları yönetimi ve tarım alanlarının kullanımı gibi geniş bir alanda strateji geliştirilmesi açısından kritik öneme sahip olduğu literatürde ortaya konulmuştur. Reis ve ark. (2008), Rize’de meydana gelen heyelanlar ile yağış arasındaki ilişkiyi araştırmak amacıyla 1975-2006 dönemine ait 32 yıllık yağış verilerini kullanmışlardır. Yağış eşik değerleri ve yağışlı gün sayılarını analiz ederek Ardeşen, Çamlıhemşin, Çayeli, Merkez ve Pazar’ın bu durumdan en çok etkilenen ilçeler olduğunu belirlemişlerdir. Kaya ve Aladağ (2009), Konya meteoroloji istasyonuna ait 1975-2006 yıllarına ait sıcaklık ve yağış verilerini değerlendirmişlerdir. Elde edilen bulgular, bölgede bitki yaşamını sınırlayan temel unsurun sıcaklık ile su kaynakları arasındaki zamansal uyumsuzluk olduğunu ortaya koymuştur. Palta ve ark. (2019), Göksu Nehri Havzası’nın 2005-2014 dönemine ait yağış ve akış verilerini regresyon analizi ve akış katsayısı yöntemleriyle incelemişlerdir. İklim değişikliğine bağlı bir yağış azalması saptamış; yaz debilerindeki artışın yapay su takviyelerinden kaynaklandığını belirterek havzaya özgü ve sürdürülebilir su yönetimi planlarının gerekliliğini vurgulamışlardır. Durmuş ve ark. (2021), 18 istasyondan elde edilen 1980-2019 verileriyle WMO ve CLIVAR tarafından belirlenen 17 sıcaklık ve 6 yağış indisini Mann-Kendall ve Sen eğilim testleriyle analiz etmişlerdir. Analiz sonuçları, iklim değişikliğinin Antalya’da uzun vadede su kaynaklarının tükenmesi ve kuraklık riskini artıracığına işaret etmiştir. Özçelik (2025), Isparta ve çevresindeki 8 meteoroloji istasyonuna ait 2005-2024 yıllarına ait sıcaklık, nem ve yağış verilerini analiz etmiştir. Buradan hareketle, tarımsal üretimde suyu verimli kullanan tekniklere geçilmesi ve kuraklığa dayanıklı türlerin tercih edilmesi gerektiğini önermiştir. Uslu ve ark. (2025), Menemen Ovası’na ait 1994-2024 dönemine ait yağış ve sıcaklık verilerini Mann-Kendall ve Sen eğilim testleriyle incelemişlerdir. Temel iklim göstergelerinin görece kararlı seyrettiğini belirlemiş sulama verimliliğinin artırılması ve kuraklığa dayanıklı türlerin seçilmesini önermişlerdir. Görüldüğü üzere Türkiye’deki çalışmalar ağırlıklı olarak geçmiş dönem verilerine dayalı eğilim analizleriyle sınırlı kalmakta, ileriye dönük yağış tahminine ise sınırlı ölçüde değinilmektedir.

Uluslararası literatürde ise yağış tahmini, klasik istatistiksel zaman serisi yaklaşımlarından makine öğrenmesi ve derin öğrenme temelli yöntemlere doğru belirgin bir dönüşüm geçirmiştir. Gradyan Artırma (Gradient Boosting) ailesine ait CatBoost, XGBoost ve LightGBM algoritmaları, yağışın doğrusal ve durağan olmayan yapısını yakalamadaki başarılarıyla öne çıkmakta; örneğin Kumar ve ark. (2023), kentsel günlük ve haftalık yağış tahmininde CatBoost ve XGBoost’un üstün performans sergilediğini göstermiştir. Tek başına kullanılan modellerin sınırlılıklarını aşmak amacıyla geliştirilen hibrit yaklaşımlarda ise Değişimsel Mod Ayrıştırma (VMD), yağış serisini frekans bileşenlerine ayırarak LSTM ağlarının tahmin başarısını artırmakta; model parametrelerinin metasezgisel algoritmalarla en iyilenmesinin doğruluğu daha da yükselttiği rapor edilmektedir (Wang ve ark., 2021). Bu doğrultuda POA-VMD-LSTM gibi hibrit modeller, yağış gibi yüksek değişkenliğe sahip serilerin tahmininde umut verici sonuçlar sunmaktadır. Ancak bu modeller, Türkiye’de ve özellikle bölgesel ölçekte henüz yeterince sınanmamıştır.

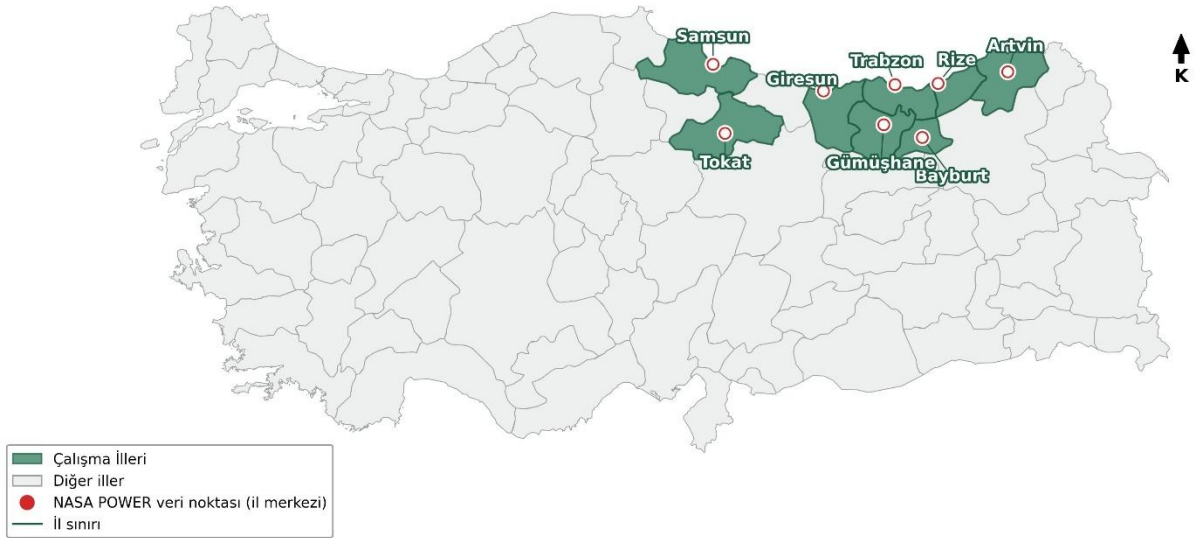
Bu çerçevede literatür iki temel boşluğa işaret etmektedir. Birincisi, Türkiye’deki yağış araştırmaları ağırlıklı olarak istatistiksel trend analizleri ve tek istasyonlu gözlem verileriyle yürütülmekte; Orta ve Doğu Karadeniz Bölgesi gibi heterojen iklim yapısına sahip alanlar il bazında ve makine öğrenmesi yöntemleriyle kapsamlı biçimde araştırılmamaktadır. İkincisi,

Gradient Boosting ailesi (CatBoost, XGBoost, LightGBM) ve hibrit derin öğrenme modellerinin (POA-VMD-LSTM) yağış tahmini konusundaki karşılaştırmalı analizlerin uluslararası literatürde yer bulmuş olsa dahi bu tür modellerin Türkiye'nin en yağışlı bölgesinde, il bazlı ve günlük veriler kullanılarak herhangi bir çalışmaya rastlanılmamıştır. Bu doğrultuda, bu araştırma literatüre üç temel katkıyı sunması hedeflemektedir: (i) Orta ve Doğu Karadeniz Bölgesi'ndeki Artvin, Bayburt, Giresun, Gümüşhane, Rize, Samsun, Tokat ve Trabzon illerindeki 2015–2024 dönemine ait günlük yağış verilerini modelleyerek bölgenin yağış rejimini yansıtmak, (ii) CatBoost, XGBoost, LightGBM ve POA-VMD-LSTM modellerinin hem modelleme hem de gelecek tahmini performanslarını MAE ve RMSE kriterleri ile sistematik olarak karşılaştırmak ve (iii) elde edilen bulgularla bölgesel su yönetimi, afet erken uyarı sistemleri ve orman kaynaklarının sürdürülebilir planlanması için karar destek süreçlerine ampirik bir temel sağlamak.

MATERYAL VE YÖNTEM

Materyal

Türkiye'nin en fazla yağış alan bölgesi olarak bilinen Orta ve Doğu Karadeniz Bölgesi'nin temel problemlerinden birisi de yağışların miktarının en az hata payı ile tahmin edilmesidir. Bölgedeki dağların kıyaya paralel uzanmasının bir sonucu olarak nemli deniz havasının tutularak konvektif ve orografik süreçlerini içinde heterojen bir iklim yapısının oluşmasını sağlamaktadır. Bu nedenle, tek bir istasyondan elde edilen verilerin bölgeyi temsil etme kapasitesi oldukça sınırlı kalmaktadır. Araştırmada NASA (2024) tarafından sağlanan Orta ve Doğu Karadeniz Bölgesi'ndeki illerin (Artvin, Bayburt, Giresun, Gümüşhane, Rize, Samsun, Tokat ve Trabzon) 1 Ocak 2015 - 31 Aralık 2024 yılları arasındaki günlük yağış miktarı incelenmiştir.



Şekil 1. Çalışma alanı içerisindeki NASA POWER veri noktaları (Kaynak: NASA 2024)

Şekil 1’de, Orta ve Doğu Karadeniz Bölgesi’nde yer alan Artvin, Bayburt, Giresun, Gümüşhane, Rize, Samsun, Tokat ve Trabzon illerinin sınırları ile bu illere ait NASA POWER veri noktalarının coğrafi dağılımı gösterilmektedir. Kıyı boyunca uzanan dağ sıralarının

bölgeyi iç ve kıyı kesimleri olarak ayırması, çalışma alanındaki iklim heterojenliğinin temel kaynağını oluşturmaktadır. Çalışmada kullanılan yağış verileri, NASA Prediction of Worldwide Energy Resources (POWER) projesi aracılığıyla, MERRA-2 yeniden çözümleme (reanalysis) ürününden elde edilmiş olup, bu veri seti yer istasyonu gözlemlerine değil, yaklaşık $0,5^\circ \times 0,625^\circ$ mekânsal çözünürlüğe sahip model tabanlı ızgara (grid) verilere dayanmaktadır. Her il için günlük yağış değerleri, ilgili il merkezinin coğrafi koordinatlarına (enlem-boylam) karşılık gelen ızgara hücresinden sorgulanarak il düzeyinde temsili bir seri oluşturulmuş; ancak ızgara yapısının doğası gereği tek bir hücrenin, bölgenin orografik kaynaklı yerel değişkenliğini sınırlı ölçüde temsil edebileceği bir kısıt olarak göz önünde bulundurulmuştur. Burada, araştırma verisinin 10 yıllık olarak tercih edilmesi hem mevsimsel döngüleri hem de kısa vadeli serilerin gözden kaçırabileceği yıllık değişkenlikleri ve ekstrem hava olaylarını da kapsayabilecek bir şekilde derinlik sağlamaktadır. Çalışmada son 3 aylık (92 günlük) periyot tahmin kümesi olarak ayrılmıştır. Zaman serisi analizlerinin doğası gereği, verinin kronolojik bütünlüğünü koruyan İlerleyen Pencere Doğrulaması (Walk-Forward Validation) stratejisinin benimsenmiş olması, modelin zamansal bağımlılıkları yakalama ve "concept drift" olgusuna karşı direncini test etme noktasında literatürle uyumlu bir geçerlilik sunmaktadır.

Hedef değişken olarak düzeltilmiş günlük yağış miktarı tercih edilmiştir. Ölçüm birimi milimetre (mm) olup $1 \text{ mm} = 1 \text{ kg/m}^2$ 'ye eşdeğerdir. Modelleme ve tahmin öncesinde veriler birim tutarlılığı ve yapısal bütünlük açısından gözden geçirilmiş; sistematik sapmalar (bias) giderildikten sonra veri seti analize hazır hale getirilmiştir.

Yöntem

POA-VMD-LSTM

POA-VMD-LSTM modeli 3 farklı algoritmanın entegre edilmesi ile oluşturulmuştur. Bu algoritmalar:

Pelikan Optimizasyon Algoritması (POA), pelikanların avlanma sırasındaki doğal davranışlarından yola çıkılarak Trojovský ve Dehghani (2022) tarafından geliştirilen, doğadan esinli bir metasezgisel optimizasyon yöntemidir. Algoritmanın performansı, unimodal ve multimodal tiplerden yirmi üç amaç fonksiyonu üzerinde test edilmiştir (Trojovský ve Dehghani, 2022). Güçlü bir keşif kapasitesine sahip olan POA, genellikle az sayıda iterasyonla optimal çözüme ulaşmaktadır. Karmaşık optimizasyon problemlerinin çözümünde POA'nın matematiksel modellemesi aşağıdaki şekilde tanımlanmıştır.

$$x_{i,j} = l_j + rand * (u_j - l_j) \quad i = 1,2,\dots,N \quad j = 1,2,\dots,m \quad (1)$$

Burada $x_{i,j}$ i'inci aday çözüm tarafından j'inci değişkenin değeridir. N popülasyon üyelerinin sayısı, m ise problem değişkenlerinin sayısıdır. $rand, [0,1]$ aralığında rastgele bir sayı, l_j problem değişkenlerinin j'inci alt sınırını ifade ederken u_j ise üst sınır olarak tanımlanmıştır.

Varyasyonel Mod Ayırıştırma (VMD), Dragomiretskiy ve Zosso (2014) tarafından tanımlanan mod ayırıştırma temelinde karmaşık sinyal ayırıştırma yöntemidir. Merkez frekansı elde edebilmek için öncelikle sinyali ayarlanmış mod sayısına göre farklı merkezlere sahip frekanslara sahip sonlu bir bant genişliğine ayırır her mod ve merkez frekansı sürekli olarak güncellemek için alternatif yön çarpanı yöntemi ile kademeli olarak temel frekans bandına dönüştürülür. Kısaca

verinin ön işleme kısmıdır. Ham bir zaman serisini doğrudan modellemeye almak yerine VMD yönetebilmek için alt bileşenlerine ayırır. Bu durum LSTM'in öğrenmesi gereken örüntüleri belirginleştirerek karmaşık sinyalleri farklı frekans bantlarını basit alt sinyallere böler bu sayede verideki anlamsız, rastgele veya hatalı verileri azaltır. VMD aşağıdaki gibi tanımlanmıştır.

$$\min \left\{ \sum_{k=1}^K \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \right\} \quad (2)$$

$$s. t. \sum_{k=1}^K u_k = f \quad (3)$$

Burada, f analiz edilen orijinal veri sinyalini ifade etmektedir. u_k ayrılan k . Mod yani alt bileşen ve $u_1, u_2, \dots, u_k$ parçalarına bölünür. ω_k her bir modun merkez frekansı, $\delta(t)$ dirac dağılımı, $\left(\delta(t) + \frac{j}{\pi t} \right) u_k(t)$ ise u_k sinyalinin Hilbert dönüşümüdür. Burada sinyal sadece pozitif frekanslara sahip bir analitik sinyale dönüştürülmektedir. $e^{-j\omega_k t}$ üstel bir terim analitik sinyali merkez frekansına ω_k değerine göre kaydırarak sinyalin temel bant genişliğini hesaplar. $\|\dots\|$ ise yani normun karesi, her bir modun bant genişliğinin toplamını temsil etmektedir.

Uzun Kısa-Vadeli Bellek (LSTM) bir tahmin merkezi olarak kullanılmaktadır. LSTM'in buradaki işlevi, VMD tarafından ayrıştırılan her bir alt bileşeni (IMF) girdi olarak almak ve bu verilerdeki zamansal bağımlılıkları öğrenerek gelecekteki değerleri tahmin etmektir. Bu yapı, zaman serilerindeki uzun vadeli eğilimleri ve mevsimselliği yüksek başarıyla yakalayabilmesi nedeniyle tercih edilmektedir. LSTM, öz yinelemeli bir sinir ağı olarak zaman serisi verilerini işlemek için tasarlanmıştır (Zhang ve ark., 2019). Birçok farklı problemde yaygın olarak kullanılan ve uzun vadeli bağımlılık sorunlarının üstesinden gelmek için geliştirilen LSTM, muadillerinin aksine zaman serisi içerisindeki gizli durumları yakalayıp öğrenebilmekte ve önceki zaman adımlarındaki verilerle bağlantı kurabilmektedir. Zincir benzeri bir yapıya sahip olan LSTM'in en önemli unsuru hücre durumudur. Bu durum, bilginin değiştirilmeden zincir boyunca akması; yani verinin adeta bir taşıma bandı üzerinde ilerlemesi şeklinde düşünülebilir. Uzun veri dizileri için tercih edilen LSTM; unutma kapısı, giriş kapısı, hücre durumu ve çıkış kapısı olmak üzere dört ana bileşenden oluşmaktadır (Çökekoğlu, 2023). LSTM ağıının çalışma süreci aşağıda formüle edilmiştir.

$$f_t = \sigma(W_f * [h_{t-1}, x_t] + b_f) \quad (4)$$

$$i_t = \sigma(W_i * [h_{t-1}, x_t] + b_i) \quad (5)$$

$$C_t = \tanh(W_c * [h_{t-1}, x_t]) + b_c \quad (6)$$

$$c_t = f_t * c_{t-1} + i_t * C_t \quad (7)$$

$$O_t = \sigma(W_o * [h_{t-1}, x_t] + b_o) \quad (8)$$

$$h_t = O_t * \tanh(c_t) \quad (9)$$

Burada, x_t şuan ki girdiyi temsil ederken, t ve h_{t-1} o andaki biriminin çıktısı, C_t t anındaki yeni bilgiyi, $\tanh$ aksiyon fonksiyonu ile C_t hücrenin güncellenmiş durumudur. σ verileri 0-1 arasında sıkıştırarak aktivasyon işlemi için kullanmakta olup bu durum sigmoid aksiyonu olarak tanımlanmaktadır. O_t çıkış kapısının hesaplanmasını ve gizli durumdan ne kadarının çıkışa aktarılacağını belirlediği aşamadır.

Bu bütünleşik yapı çerçevesinde POA, VMD'nin ayrıştırma performansını belirleyen parametreleri adaptif biçimde optimize ederek veri gürültüsünün azaltılmasına katkı sağlamakta; VMD, ham veriyi bu optimal parametrelerle farklı frekans bileşenlerine ayrıştırmakta; LSTM ise her bir alt bileşen için ayrı ayrı tahmin üretmekte ve bu tahminler birleştirilerek birleşik tahmin sonucu elde edilmektedir. Bu doğrultuda Zhou ve ark. (2025), önerdikleri POA-VMD-LSTM modelinin karşılaştırılan diğer modellere kıyasla daha düşük hata değerleri (MAE, RMSE, MAPE) ve daha yüksek belirleme katsayısı (R^2) ile üstün tahmin performansı sağladığını tespit etmiştir.

XGBoost

XGBoost, Gradient Boosting Machine (GBM) modelinin optimize edilmiş (Yeşilyurt ve Dalkılıç, 2021) ve yüksek performanslı bir ağaç tabanlı gradyan artırma kütüphanesidir (Chen ve Guestrin, 2016). XGBoost, sık karar ağaçlarını sıralı bir şekilde eğiterek güçlü bir tahminleyici oluşturma prensibine dayanır. Burada, her yeni ağaç, bir önceki ağacın hatalarını telafi edecek şekilde eğitilir. XGBoost yüksek tahmin gücü, azaltılmış karmaşıklık, eksik verilerin yönetilmesi, aşırı öğrenimin önlenmesi ve çalışma hızı gibi önemli avantajları sağlamaktadır. XGBoost aşağıdaki gibi tanımlanmıştır.

$$\hat{y}_i = \sum_{k=1}^k f_k(x_i) \quad (10)$$

$$L(\theta) = \sum_{i=1}^n L(y_i, \hat{y}_i) + \sum_{i=1}^n \Omega(f_k) \quad (11)$$

$$\Omega(f_k) = y^T + \frac{1}{2} \lambda \|\omega\|^2 \quad (12)$$

Burada, dönemdeki tahmin edilen değer, toplamdaki veri sayısı, için k. ağaç tahminini temsil etmektedir. Modelin hedef değişkeni ne kadar iyi tahmin ettiğini ölçen kayıp fonksiyonu, modelin karmaşıklığını cezalandıran düzeltme terimidir. Ayrıca, ağaçtaki yaprak sayısını ve modelin karmaşıklığını kontrol eden düzenleme parametresidir.

CatBoost

CatBoost modeli hataları en aza indirmek ve tahminleri iyileştirmek için kategorik özelliklerle çalışmaları desteklerken bilgi kaybını en aza indirmek için tasarlanmıştır. CatBoost diğer modellere göre, daha az ön işleme ihtiyaç duymakta ve direkt veriyi kullanmaktadır Dorogush ve ark. (2018). Ayrıca, modelin aşırı öğrenmesini engellemek için veri setinin rastgele permütasyonlarını kullanarak, o andaki veriden önceki verilerin istatistikleri hesaplanır. Buna ek olarak, CatBoost simetrik ağaç yapısını kullanır. Bu yapı, modelin tahmin hızını artırır ve daha kararlı sonuçlar üretmesine olanak tanır. CatBoost modeli aşağıdaki gibi formülize edilmiştir.

$$h^t = \arg \arg L(F^{t-1} + h) \quad h \in H \quad (13)$$

$$= \arg \arg L(y, F^{t-1}(x) + h(x)) \quad h \in H$$

$$h(x) = \sum_j b_j I\{x \in R_j\} \quad (14)$$

Burada, y çıktısını temsil ederken, h ise H fonksiyon ailesinden seçilen bir gradyan adım fonksiyonudur. R_j ise karar ağacının yapraklarına karşılık gelen bölgeleri, b_i ise karşılık gelen bölgelerin tahmin değerlerini ifade ederken, I ise bir gösterge fonksiyonu olarak tanımlanabilir (Tunç ve Atasever 2025).

LightGBM

LightGBM modeli, Microsoft tarafından büyük veri ve çok boyutlu özellikler üzerinde tahmin yapılabilmesi için tasarlanmıştır. LightGBM modeli büyük verileri hızlı bir şekilde işleyip tahmin sunmaktadır. LightGBM, karar ağaçlarında öğreniminde seviye odaklı (level-wise or depth-wise) veya yaprak odaklı (leaf-wise) olarak iki strateji kullanılabilir. Seviye odaklı stratejide ağaç büyürken ağacın dengesi korunur, iken yaprak odaklı stratejide ise kaybı azaltan yapraklardan bölünme işlemi devam eder. LightGBM'in bu özelliği ile diğer boosting algoritmalarından ayrılmaktadır. LightGBM modeli, daha hızlı öğrenme ve daha az hata ile tahmin için, her seferinde daha az yaprak ile daha iyi bir performans göstermiştir (Eşidir, 2025). LightGBM modeli aşağıdaki gibi formülize edilmiştir.

$$L^{(t)} = \sum_{i=1}^n l(y_i, \hat{y}^{(t-1)} + f_t(x_i) + \Omega(f_t)) \quad (15)$$

Burada, $L^{(t)}$ t adımıdaki toplam kaybı, l kayıp fonksiyonunu, y_i i . gerçek değeri ve $\hat{y}^{(t-1)}$, $t - 1$ adımıdaki tahmini, $f_t(x_i)$ ise t . adımıdaki model tarafından eklenen yeni fonksiyonu ve $\Omega(f_t)$ yeni fonksiyonun cezasını temsil etmektedir.

Model Kıyaslama Kriterleri

Bu çalışmada, modelleme ve gelecek tahmini performanslarını karşılaştırmak için Mean Absolute Error (Ortalama Mutlak Hata, MAE) ve Root Mean Square Error (Hata Kareler Ortalamasının Karekökü, RMSE) kriterleri kullanılmıştır. MAE, tahmin hatalarının ortalama mutlak büyüklüğünü hatanın yönünden bağımsız biçimde verirken; RMSE, hataların karesi üzerinden hesaplandığından büyük sapmalara daha duyarlıdır. Bu iki kriter, yağış tahminine yönelik makine öğrenmesi çalışmalarında en yaygın kullanılan değerlendirme kriterleri arasında yer almaktadır (El Hafyani 2024; Neyestani vd., 2025). Bu kriterler birlikte kullanıldıklarında hatanın hem genel düzeyini hem de aykırı değerlere yönelik davranışını yansıttığından, özellikle yağış benzeri çevresel değişkenlerin karşılaştırmalı analizinde standart bir çerçeve sunmaktadır (Neslihanoglu vd., 2021).

$$MAE_i = \sum_{t=1}^N \frac{|\hat{y}_{it} - y_{it}|}{N} \quad (16)$$

$$RMSE_i = \sqrt{\sum_{t=1}^N \frac{(\hat{y}_{it} - y_{it})^2}{N}} \quad (17)$$

Burada, y_{it} ve $\hat{y}_{it}$ sırasıyla i . varlığın t . zamandaki gözlem değerini ve modelleme (gelecek tahmin) tahmini değerini göstermektedir. Burada, küçük MAE ve RMSE değerine sahip modelin kıyaslama aşamasında diğerlerine göre daha iyi modelleme (gelecek tahmini) performansına sahip olduğu söylenebilir. Bu çalışmada Python programlama dili ve PyTorch,

NumPy, pandas, scikit-learn, SciPy, XGBoost, CatBoost, LightGBM, VMDpy, Matplotlib paketleri kullanılmıştır.

Uygulama Kısıtları ve Model Parametreleri

Araştırmada yer alan tüm modellerin hiperparametreleri, kronolojik bir hold-out doğrulama şeması altında sistematik ızgara arama (grid search) yöntemiyle belirlenmiştir. Önerilen hibrit modelde (POA-VMD-LSTM), VMD sinyal ayrıştırma parametreleri K için [3, 6] ve α için [500, 2500] arama uzayında Pelikan Optimizasyon Algoritması (POA) ile dinamik olarak optimize edilirken; LSTM ağ yapısı ızgara arama yoluyla yapılandırılmıştır. En iyi konfigürasyon olarak gizli birim sayısı $h = 64$, katman sayısı $L = 2$, dropout oranı $p_{drop} = 0,2$ ve öğrenme oranı $\eta = 0.001$ değerleri belirlenmiştir.

Karşılaştırmalı analizde yansızlığı sağlamak amacıyla ağaç tabanlı baz modelleri (XGBoost, CatBoost, LightGBM) için de eşdeğer bir optimizasyon protokolü uygulanmıştır. Hold-out doğrulama kaybını en aza indiren en iyi yapılandırmalar şu şekildedir: XGBoost için $n_{est} = 400$, $\eta = 0.05$, $d_{max} = 6$; CatBoost için $T = 500$, $\eta = 0.05$, $d = 6$ ve LightGBM için $n_{est} = 500$, $\eta = 0.05$ deneylerin tekrarlanabilirliği amacıyla tüm modellerde rastgelelik tohumu $s = 42$ olarak sabitlenmiştir.

Tahmini için genişleyen pencere yaklaşımına dayalı ileri doğrulama (expanding walk-forward validation) yöntemi kullanılmıştır. Bu yaklaşımda eğitim kümesinin başlangıç tarihi (2015) sabit tutulmuş, bitiş tarihi ise 2023'ten itibaren her adımda bir gün ileriye kaydırılmıştır; kaydırmayı izleyen tek gün ise test verisi olarak değerlendirilmiştir. Söz konusu süreç, yaklaşık dört aylık bir işlem dönemini kapsayacak biçimde toplam 92 adımda yinelenmiştir. Sürecin adımları şu şekildedir:

Adım 1. Eğitim kümesi, başlangıç tarihi (2015) sabit kalacak biçimde tanımlanır; ilk adımda eğitim penceresi 2015 başından 2023'teki ilk test gününden bir önceki güne kadar uzanır.

Adım 2. Model bu eğitim kümesi üzerinde tahmin edilir ve bir sonraki tek gün ($t+1$) için öngörü üretilir; bugün test verisi olarak ayrılır.

Adım 3. Test günü gözlemlendikten sonra eğitim kümesine eklenir; böylece pencerenin bitiş tarihi bir gün ileriye kayarak pencere bir gün genişler ve süreç bir sonraki gün için yinelenir. Bu yineleme toplam k ($k = 92$) adımda gerçekleştirilmiştir. Her adımda eğitim kümesi bir önceki adıma kıyasla bir gözlem daha içermekte, test kümesi ise her zaman tek bir ileri günden oluşmaktadır.

Buradaki süreçte her adımda iki tür hata üretilmektedir: Eğitim kümesi üzerinde (Adım 1'de tanımlanan ve Adım 3'te bir gözlem genişleyen kayar pencere) hesaplanan uyum hatası ile o adımın test günü (Adım 2'deki $t+1$ öngörüsü) üzerinde hesaplanan tahmin hatası. Modelleme performansı, her adımdaki eğitim penceresi üzerinde elde edilen MAE ve RMSE değerlerinin 92 adım boyunca birleştirilmesiyle değerlendirilmiş ve Çizelge 2'de sunulmuştur. Gelecek tahmini performansı ise her adımda ayrılan tek günlük test gözlemleri üzerindeki MAE ve RMSE değerlerinin 92 adımlık öngörü dizisi boyunca toplanmasıyla hesaplanmış ve Çizelge 3'te sunulmuştur.

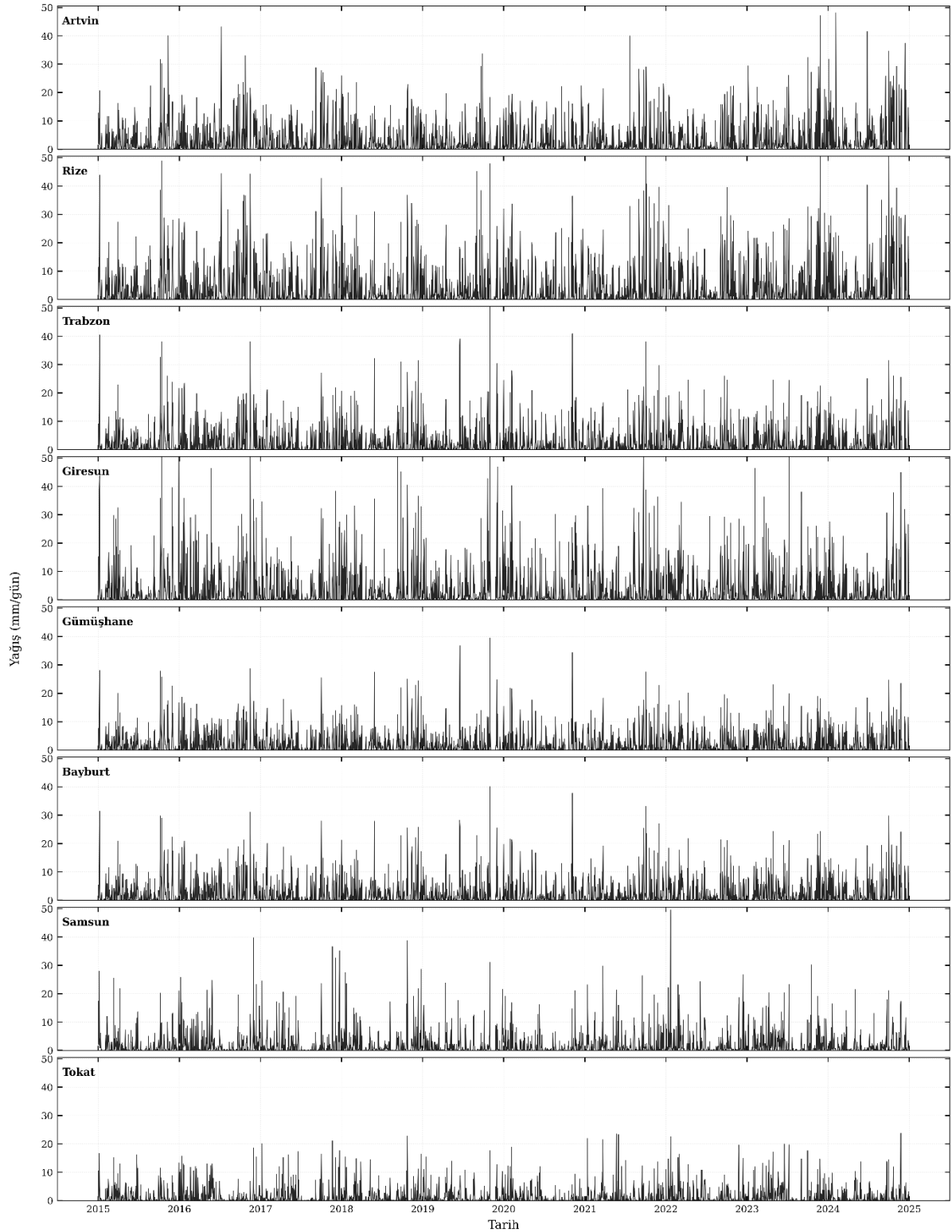
BULGULAR

Bu bölümde, Orta ve Doğu Karadeniz Bölgesi'ndeki illere ait günlük yağış verilerinin betimsel istatistikleri ile zamansal ve dağılımsal özellikleri sunulmaktadır. Buna ek olarak CatBoost, XGBoost, LightGBM ve POA-VMD-LSTM modellerinin modelleme ve gelecek tahmini performansına ilişkin bulgulara yer verilmektedir. Elde edilen sonuçlar, model kıyaslama kriterleri olan MAE ve RMSE değerleri temelinde değerlendirilmiştir. Günlük yağış miktarına ilişkin betimsel istatistikler Çizelge 1'de sunulmaktadır.

Çizelge 1. Yağış verilerinin betimleyici istatistikleri

İl	Minimum	1. Çeyrek	Medyan	Ortalama	3. Çeyrek	Maksimum	StdSapma
Artvin	0,0	0,0	0,6	3,0	3,5	48,1	5,2
Rize	0,0	0,0	0,6	3,8	4,4	55,4	6,9
Trabzon	0,0	0,0	0,3	2,5	2,9	55,8	4,8
Giresun	0,0	0,0	0,6	3,6	3,8	64,3	6,9
Gümüşhane	0,0	0,0	0,4	2,1	2,6	39,5	3,8
Bayburt	0,0	0,0	0,4	2,3	2,9	40,1	4,2
Samsun	0,0	0,0	0,2	1,9	1,8	49,5	4,0
Tokat	0,0	0,0	0,1	1,4	1,4	23,7	2,9

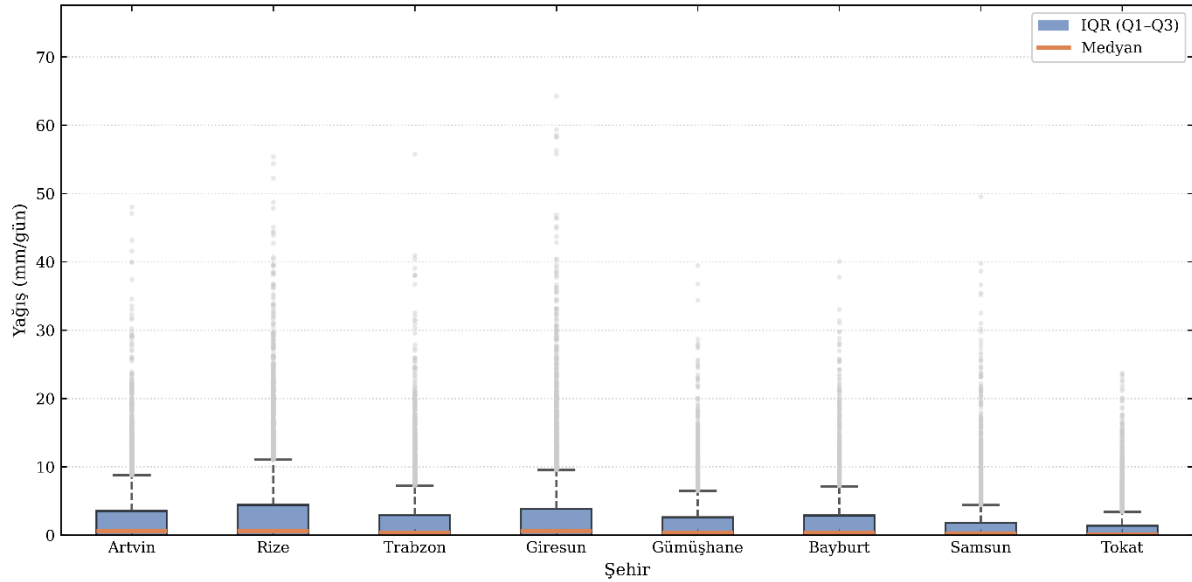
Tüm illerde yağış miktarı belirgin bir mevsimsellik olduğu Şekil 1'de görülmektedir. Kıyı illerinde (Rize, Trabzon, Artvin) sonbahar-kış aylarındaki yağış maksimum noktaları daha belirgin ve sık iken, iç kesimlerdeki illerde (Tokat, Bayburt, Gümüşhane) ise yaz aylarının belirgin biçimde kurak geçtiği görülmektedir. Rize ve Giresun'a ait serilerde 39–65 mm aralığını aşan keskin dalgalanmalar gözlemlenmektedir. Bu dalgalanmalar, ani ve şiddetli yağış olaylarının varlığına işaret etmekte; serinin genel düşük yağış arka planı üzerinde anlık sinyaller biçiminde kendini göstermektedir. Bu yapı, zaman serisinin yüksek bir aralık örüntüsüne sahip olduğunu ortaya koymaktadır. Kıyı illerinde (özellikle Rize ve Giresun) seri boyunca yoğun bir yağış dokusu gözlenmiş iken, Tokat ve Bayburt gibi iç Anadolu etkisindeki illerde zamanın önemli bir bölümünde yağış değerleri sıfıra yakın seyretmektedir. Bu görsel farklılık, araştırmada tercih edilen tek bir model yerine il bazlı modelleme yaklaşımının neden tercih edildiğini desteklemektedir.



Şekil 2. İller bazında günlük yağış miktarı zaman serisi grafikleri

Şekil 3’de, 8 ile ait günlük yağış dağılımının sağa çarpık bir yapı sergilediği açıkça görülmektedir. Tüm illerde medyan değerler 1 mm/gün’ün altında kalmakta (0.11–0.58 mm aralığında) ve ortalamanın altında seyretmektedir. Bu durum günlerin büyük çoğunluğunun az yağışlı ya da yağışsız geçtiğini göstermektedir. Ayrıca, yağışsız günlerin oranı Tokat’ta %34,7’ye, Trabzon’da %28,4’e ulaşmaktadır. IQR (Q1–Q3) aralıkları incelendiğinde ise Rize

(0.01–4.42 mm) ve Giresun (0.02–3.83 mm) en geniş kutu genişliğine sahip iken Tokat (0.00–1.36 mm) ve Samsun (0.01–1.76 mm) belirgin biçimde dar genişlikte kalmaktadır. Analizler, kıyı şeridindeki yağışın hem yoğunluk hem de değişkenlik bakımından baskın olduğunu, iç kesimlere gidildikçe ise yağış rejiminde belirgin bir daralma yaşandığını ortaya koymaktadır. Özellikle, kutu grafikleri üzerinde kümelenen aykırı değerler, tüm illerde dikkat çekici bir yoğunluğa sahiptir. Giresun’da 64.3 mm, Rize’de 55.4 mm, Trabzon’da ise 55.8 mm’ye ulaşan ekstrem günlük olaylar bu durumun somut göstergeleridir. Burada, fazla aykırı değer varlığı, bölgenin konvektif karakterli yağışlara açık olduğunun kanıtı niteliğindedir. Bu durum ise, modelleme aşamasında ekstrem olayların kestirimini en temel sorunlardan birisi haline getirmektedir.



Şekil 3. İller bazında günlük yağış miktarı kutu grafikleri

Modelleme

Bu bölümde, Orta ve Doğu Karadeniz Bölgesi’ndeki 8 ilin günlük yağış verilerinin CatBoost, XGBoost, LightGBM, ve POA-VMD-LSTM makine öğrenmesi modelleri ile modelleme performansının MAE ve RMSE kriterlerine göre karşılaştırma sonuçları Çizelge 2’de sunulmuştur.

Çizelge 2. Makine öğrenmesi yöntemlerinin yağış verilerinin modelleme karşılaştırmaları

MAE				
İl	POA-VMD-LSTM	XGBoost	CatBoost	LightGBM
Artvin	1,826	0,471	1,753	0,336
Rize	2,047	0,589	2,315	0,392
Trabzon	1,526	0,429	1,690	0,318
Giresun	2,048	0,553	2,278	0,400
Gümüşhane	1,270	0,341	1,307	0,248
Bayburt	1,385	0,403	1,489	0,274
Samsun	1,150	0,310	1,207	0,266

Tokat	0,896	0,213	0,912	0,160
<i>Ortalama</i>	<i>1,519</i>	<i>0,414</i>	<i>1,619</i>	<i>0,300</i>
RMSE				
İl	POA-VMD-LSTM	XGBoost	CatBoost	LightGBM
Artvin	2,605	0,659	2,614	0,522
Rize	3,405	0,830	3,497	0,591
Trabzon	2,468	0,612	2,542	0,554
Giresun	3,570	0,782	3,512	0,620
Gümüşhane	1,966	0,483	1,952	0,382
Bayburt	2,161	0,568	2,243	0,414
Samsun	2,148	0,446	1,971	0,555
Tokat	1,510	0,309	1,458	0,251
<i>Ortalama</i>	<i>2,479</i>	<i>0,586</i>	<i>2,473</i>	<i>0,486</i>

Çizelge 2’deki sonuçlara göre, LightGBM’in illerin tercih edilen periyottaki yağış miktarlarını modellemede ortalama olarak en düşük MAE (0,300) ve RMSE (0,486) değerlerine sahip olduğu, CatBoost’un ise en yüksek MAE (1,619) ve RMSE (2,473) değerlerine sahip olduğu görülmüştür. Sonuç olarak LightGBM modeli, Orta ve Doğu Karadeniz Bölgesi’ndeki 8 ilin günlük yağış verilerinin modellenmesinde diğer modellere kıyasla daha etkin bir performans göstermiştir.

Gelecek Tahmini

Bu bölümde, araştırma verilerinin CatBoost, XGBoost, LightGBM, ve POA-VMD-LSTM makine öğrenmesi modelleri ile gelecek 3 aylık (92 günlük) tahmin performansı MAE ve RMSE kriterleri sonuçları Çizelge 3’de gösterilmiştir.

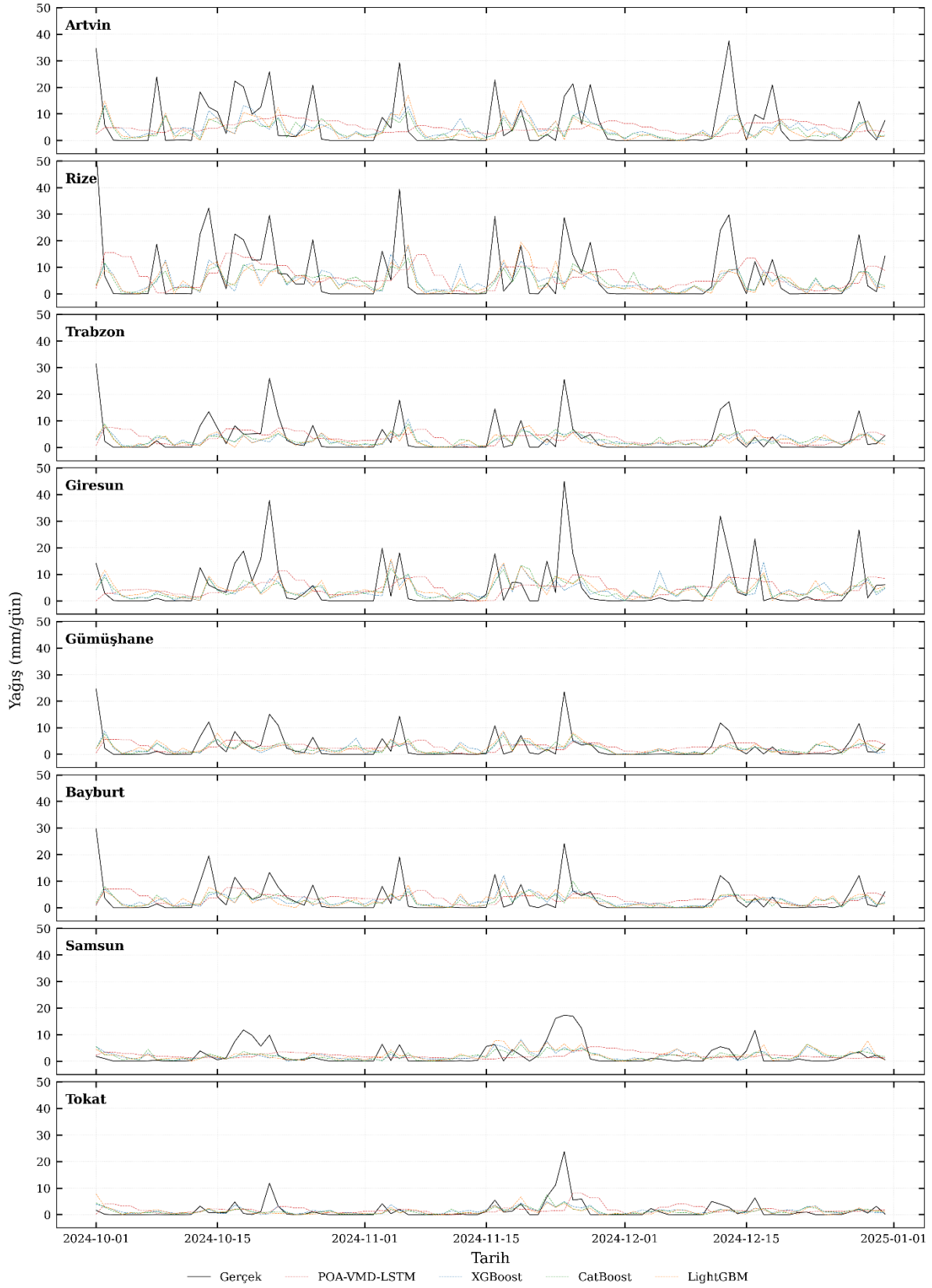
Çizelge 3. Makine Öğrenmesi Yöntemlerinin Yağış verilerinin Gelecek Tahmini Karşılaştırmaları

MAE				
İl	POA-VMD-LSTM	XGBoost	CatBoost	LightGBM
Artvin	6,674	5,832	5,626	5,615
Rize	7,865	6,774	6,473	6,410
Trabzon	4,431	3,557	3,578	3,410
Giresun	5,948	5,518	5,261	5,626
Gümüşhane	3,524	2,845	2,807	2,899
Bayburt	4,099	3,472	3,196	3,362
Samsun	2,571	2,459	2,347	2,516
Tokat	2,254	1,570	1,487	1,566
<i>Ortalama</i>	<i>4,670</i>	<i>4,003</i>	<i>3,847</i>	<i>3,926</i>
RMSE				
İl	POA-VMD-LSTM	XGBoost	CatBoost	LightGBM
Artvin	9,671	8,679	8,743	8,885
Rize	12,186	10,615	10,390	10,540

Trabzon	7,334	6,017	5,943	5,925
Giresun	9,757	8,856	8,392	8,647
Gümüşhane	5,806	4,821	4,732	4,752
Bayburt	6,539	5,706	5,441	5,511
Samsun	3,684	3,770	3,606	3,744
Tokat	3,570	2,883	2,876	2,992
<i>Ortalama</i>	<i>6,982</i>	<i>6,418</i>	<i>6,265</i>	<i>6,375</i>

Çizelge 3’te, Orta ve Doğu Karadeniz Bölgesi’ndeki 8 ilin günlük yağış miktarı verilerinin gelecek tahmininde ortalama olarak en düşük MAE ve RMSE değerlerine CatBoost modelinin, en yüksek MAE ve RMSE değerlerine ise POA-VMD-LSTM modelinin sahip olduğu görülmüştür. Ortalama MAE (RMSE) açısından, CatBoost’un gelecek tahmininde POA-VMD-LSTM’ye göre %21,39 (%11,44) daha iyi performans gösterdiği belirlenmiştir. Özet olarak CatBoost, diğer modellere kıyasla daha iyi bir gelecek tahmini performansı sergilemiştir.

Şekil 4’te söz konusu makine öğrenmesi modelleri genel olarak yağış miktarı verilerinin gelecek tahmini konusunda yakın performanslı olduğu ve sonuçların Çizelge 3 ile eşleştiği görülmüştür. Buna ek olarak, söz konusu modellerin iller bazındaki yağış miktarı verilerinin gelecek tahmini konusundaki genel trendlerini yakalamada başarılı, fakat genel olarak iller için ani dalgalanmaların olduğu gözlemlerde ise yetersiz kaldığı tespit edilmiştir.



Şekil 4. İllere göre yağış miktarı verilerinin 3 aylık gelecek tahminleri

TARTIŞMA

Bu araştırmanın bulguları, modelleme başarısı ile gelecek tahmin performansının her zaman aynı modelde birleşmediğini ortaya koymaktadır. Modelleme aşamasında en yüksek başarıyı gösteren LightGBM'in performansı, yaprak odaklı (leaf-wise) büyüme stratejisi ve histogram tabanlı bölme yaklaşımıyla ilişkilendirilebilir; bu yapı, modelin verideki keskin geçişleri ve aykırı değerleri daha etkin biçimde temsil etmesine olanak tanımaktadır (Ke vd., 2017). Kayıp fonksiyonunu en fazla azaltan yapraklardan bölünerek ilerleyen bu yaklaşım, verideki ani dalgalanmaların yüksek çözünürlükle öğrenilmesini sağlamakta ve bu yönüyle Orta ve Doğu Karadeniz yağış serilerinin yüksek frekanslı değişken yapısıyla uyum göstermektedir. Ayrıca, söz konusu yüksek frekanslı değişimleri temsil edebilme kapasitesi, modelin düşük yanlılık (bias) üretmesine katkı sağlamış ve modelleme aşamasındaki üstün performansını desteklemiştir. Elde edilen bulgu, güncel yağış modelleme çalışmalarıyla da uyum göstermektedir: Tyralis vd. (2023), uydu ve istasyon yağış verilerinin birleştirilmesinde LightGBM'in özellikle dağılımın uç kuantillerini tahmin etmede rastgele orman temelli yöntemlere kıyasla üstün olduğunu bildirmiş; Papacharalampous vd. (2024), mekânsal yağış tahminlerinde belirsizlik kestiriminde LightGBM'in karşılaştırılan diğer yöntemler (rastgele orman ve sinir ağı varyantları dâhil) arasında en yüksek performansı sergilediğini göstermiş; Islam vd. (2025) ise güncel karşılaştırmalı çalışmalarda LightGBM'in de içinde bulunduğu gradyan artırma temelli yöntemlerin, yüksek değişkenliğe sahip günlük yağış serilerinin modellenmesinde MAE ve RMSE açısından daha başarılı sonuçlar ürettiğini ortaya koymuştur. Bu bulgular, LightGBM'in az sayıda yaprakla dahi düşük hata değerleri üretebildiğini gösteren Eşidir (2025) ile Gradient Boosting ailesinin yağış tahminindeki başarısını ortaya koyan Kumar ve ark. (2023) tarafından elde edilen sonuçları bölgesel ve günlük ölçekte desteklemektedir.

Bununla birlikte, modelleme aşamasında gözlenen bu üstün performans, gelecek tahmin aşamasına doğrudan yansımamıştır; bu aşamada en başarılı sonuçları CatBoost üretmiştir. LightGBM'in test verisindeki performans kaybı, yaprak odaklı yapısının eğitim verisine aşırı uyum (overfitting) gösterme eğilimiyle açıklanabilir. Buna karşılık, CatBoost'un simetrik ağaç yapısı ve sıralı artırma (ordered boosting) mekanizması, varyansı sınırlandırarak daha kararlı ve genellenebilir tahminler üretilmesini sağlamakta; bu yapısal özellik, aynı zamanda modelin aykırı değerleri yakalama kapasitesini de artırmaktadır. Elde edilen bulgu, CatBoost'un aşırı öğrenmeye karşı gösterdiği dayanıklılığı vurgulayan Dorogush ve ark. (2018) ile Tunç ve Atasever (2025) tarafından raporlanan sonuçların yanı sıra, Fas'ın yarı kurak bölgelerinde günlük yağış tahmininde CatBoost'un XGBoost, LSTM ve TCN modellerine kıyasla en yüksek başarıyı gösterdiğini ve bu üstünlüğü modelin doğrusal olmayan yağış dinamiklerini yakalama performansına bağlayan Elmotawakkil vd. (2025) bulgularıyla da uyumludur. Ayrıca, CatBoost'un POA-VMD-LSTM modeline kıyasla belirgin biçimde daha düşük hata üretmesi, sağa çarpık ve aykırı değer yoğunluğu yüksek yağış dağılımlarını temsil etmede daha başarılı olduğuna işaret etmektedir.

Ağaç tabanlı bu iki yöntemin aksine, hibrit POA-VMD-LSTM modelinin hem modelleme hem de gelecek tahmini aşamalarında en yüksek hata değerlerini üretmesi dikkat çekicidir. Bu sonuç, literatürde bildirilen olumlu bulgularla (Zhou ve ark., 2025) ilk bakışta çelişiyor gibi görünse de söz konusu çalışmaların daha durağan zaman serileri üzerinde yürütülmüş olması bu farklılığı açıklayabilir. Daha da önemlisi, bu bulgu derin öğrenme modellerinin üstün performans göstereceğine ilişkin yaygın varsayımın her veri yapısı için geçerli olmadığını ortaya koymaktadır. LSTM tabanlı mimariler, yüksek parametre sayıları nedeniyle genellikle büyük hacimli ve görece durağan veri setlerinde daha başarılı sonuçlar vermektedir. Bu

araştırmada kullanılan günlük yağış serileri, gözlemlerin önemli bir bölümünün sıfıra yakın değerlerden oluştuğu (Tokat'ta yağışsız gün oranı %34,7), buna karşılık seyrek fakat şiddetli ekstrem olayların görüldüğü kesintili (intermittent) bir yapı sergilemektedir. VMD ön işlemine rağmen bu süresiz ve çarpık dağılım, LSTM mimarisinin söz konusu veri yapısındaki zamansal bağımlılıkları yeterince temsil edemediğine işaret etmektedir. Ayrıca modelin yüksek karmaşıklığı, optimizasyon maliyeti ve sınırlı yorumlanabilirliği uygulama açısından önemli dezavantajlar oluşturmaktadır. Bu nedenle hibrit derin öğrenme modellerinin üstünlüğünün veri yapısına bağlı olduğu; yağış gibi kesintili ve yüksek değişkenlik gösteren serilerde ağaç tabanlı yöntemlerin daha sağlam, yorumlanabilir ve güvenilir bir alternatif sunduğu söylenebilir.

Bu çalışma, Türkiye'de günlük yağış tahmini alanında ağaç tabanlı artırma algoritmaları ile hibrit derin öğrenme modellerinin karşılaştırmalı olarak değerlendirildiği araştırmalara katkı sunmaktadır. Reis ve ark. (2008), Durmuş ve ark. (2021) ve Uslu ve ark. (2025) gibi çalışmalar büyük ölçüde geçmiş dönem eğilim analizleriyle (Mann–Kendall ve Sen yöntemleri) sınırlı kalırken, bu araştırmada il bazında ve günlük çözünürlükte ileriye dönük yağış tahminleri gerçekleştirilmiştir. Elde edilen bulgular, tüm modellerin mevsimsel eğilimleri başarıyla yakalayabildiğini; ancak ani ve yüksek frekanslı yağış değişimlerini öngörmeye yetersiz kaldığını göstermektedir. Bu durum, bölgedeki konvektif ve orografik kaynaklı ekstrem yağışların tahmin edilmesinin temel bir zorluk oluşturmaya devam ettiğini ortaya koymaktadır. Sonuçlar, bölgesel su yönetimi ve afet erken uyarı sistemlerinde özellikle ekstrem olaylara odaklanan olasılıksal tahmin yaklaşımlarının önemini vurgulamaktadır.

Bununla birlikte, araştırmanın bazı sınırlılıkları bulunmaktadır. Çalışma yaklaşık üç aylık (92 günlük) bir test dönemiyle sınırlandırılmış ve yalnızca yağış değişkeni kullanılmıştır. Sıcaklık, bağıl nem, atmosfer basıncı, rüzgâr hızı ve atmosferik dolaşım göstergeleri gibi açıklayıcı değişkenlerin modele dahil edilmemesi, özellikle ekstrem yağış olaylarının kestirim performansını sınırlandırmış olabilir. Ayrıca kullanılan yöntemler yalnızca noktasal tahmin üretmiş olup tahmin belirsizliği değerlendirilmemiştir. Bu nedenle gelecekte yapılacak çalışmalarda daha uzun test dönemleri, çok değişkenli girdi setleri ve ekstrem olaylara duyarlı olasılıksal tahmin yaklaşımlarının birlikte değerlendirilmesi önerilmektedir.

SONUÇ

Bu araştırmada, Orta ve Doğu Karadeniz Bölgesi'nde yer alan Artvin, Bayburt, Giresun, Gümüşhane, Rize, Samsun, Tokat ve Trabzon illerine ait 2015–2024 dönemi günlük yağış miktarı verileri modellenmiş ve üç aylık ileriye dönük tahmin performansı değerlendirilmiştir. Modelleme sürecinde CatBoost, XGBoost, LightGBM ve POA-VMD-LSTM yöntemleri kullanılmış; modelleme ve tahmin başarımları ortalama mutlak hata (MAE) ve kök ortalama kare hata (RMSE) ölçütleri üzerinden karşılaştırmalı olarak incelenmiştir.

Bulgular, gelecek tahmin aşamasında en düşük hata değerlerinin CatBoost modeli tarafından, modelleme aşamasında ise en başarılı sonuçların LightGBM modeli tarafından üretildiğini göstermektedir. Buna karşılık, POA-VMD-LSTM modeli hem modelleme hem de tahmin performansı açısından diğer yöntemlerin gerisinde kalmıştır. Genel olarak sonuçlar, günlük yağış serileri gibi kesintili, sağa çarpık ve yüksek değişkenlik gösteren veri yapılarında ağaç tabanlı artırma algoritmalarının hibrit derin öğrenme yaklaşımlarına kıyasla daha başarılı

sonular verebildiđini ortaya koymaktadır. Bu bulgu, model seiminde yalnızca yntemlerin karmaşıklılıđının deđil, veri yapısının zelliklerinin de belirleyici olduđunu gstermektedir.

Araştırma sonuları, kısa dnem yađıř tahmin dođruluđundaki iyileřmenin sulama planlaması, tařkın risk ynetimi ve gnlk su kaynakları ynetimi kararlarına operasyonel dzeyde katkı sađlayabileceđini gstermektedir. zellikle Orta ve Dođu Karadeniz Blgesi gibi yađıř deđiřkenliđinin yksek olduđu alanlarda daha gvenilir kısa dnem tahminlerin elde edilmesi, su kaynaklarının etkin ynetimi ve afet risklerinin azaltılması aısından nem tařımaktadır. Bununla birlikte, sonuların gnlk znrlkte gerekleřtirilen kısa dnem tahminler ve yalnızca yađıř deđiřkeninin kullanıldıđı tek deđiřkenli model yapısı ile sınırlı olduđu dikkate alınmalıdır.

Gelecek alıřmalarda, sıcaklık, bađıl nem, rzgr hızı, atmosfer basıncı ve byk lekli atmosferik dolařım gstergelerini ieren ok deđiřkenli modellerin geliřtirilmesi nerilmektedir. Ayrıca analizlerin farklı zamansal leklerde ve farklı iklim rejimlerine sahip blgelerde tekrarlanması, bulguların genellenebilirliđinin deđerlendirilmesine katkı sađlayacaktır. Son olarak, Transformer tabanlı gncel derin đrenme mimarilerinin ve aıklanabilir yapay zek (XAI) tekniklerinin modele entegre edilmesi, gelecekteki arařtırmalar iin nemli bir yntemsel geliřim alanı olarak deđerlendirilmektedir.

YAZAR KATKILARI

İbrahim Edibali Atalay: Araştırma literatr taraması, verilerin grselleřtirilmesi, makale taslađının hazırlanması. **Serdar Neslihanoglu:** Verilerin organizasyonu, analiz ynlendirmesi, makale yazımının yapılması.

FİNANSAL DESTEK BEYANI

alıřma iin herhangi bir maddi destek alınmamıřtır.

IKAR ATIřMASI

Yazarlar arasında ıkar atıřması bulunmamaktadır.

ETİK KURUL ONAYI

Bu alıřma iin etik kurul onayı gerekmemektedir.

KAYNAKLAR

Batan, M., & Toprak, Z. F. (2015). Kresel iklim deđiřikliđinin olumlu etkileri ve bu etkilerin iklim deđiřikliđine uyum kapsamında deđerlendirilmesi. *Dicle niversitesi Mhendislik Fakltesi Mhendislik Dergisi*, 6(2), 93–102.

- Bolat, İ., & Şensoy, H. (2024). 2012–2021 yılları arasındaki bazı meteorolojik verilerin analizi ve değişim eğiliminin irdelenmesi: Sinop ve Kastamonu illeri örneği. *Anadolu Orman Araştırmaları Dergisi*, 9(2), 1–9.
- 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).
- Çökekoğlu, S. (2023). *Uzaktan algılama sistemleri ile yangına hassas bölgelerin tahmin edilmesi* (Yüksek lisans tezi). Süleyman Demirel Üniversitesi Fen Bilimleri Enstitüsü.
- Dorogush, A. V., Ershov, V., & Gulin, A. (2018). CatBoost: Gradient boosting with categorical features support. *arXiv*, 1–7. Retrieved from <http://arxiv.org/abs/1810.11363>
- Dragomiretskiy, K., & Zosso, D. (2014). Variational mode decomposition. *IEEE Transactions on Signal Processing*, 62(3), 531–544.
- Durmuş, B., Bulut, İ., & Gönençgil, B. (2021). Antalya bölümünde sıcaklık ve yağış indislerinin değişim analizleri. *Türk Coğrafya Dergisi*, 78, 91–108.
- El Hafyani, M., El Himdi, K., & El Adlouni, S.-E. (2024). Improving monthly precipitation prediction accuracy using machine learning models: A multi-view stacking learning technique. *Frontiers in Water*, 6, 1378598.
- Elmotawakkil, A., Moumane, A., Sadiki, A., Zahi, A., Al Karkouri, J., Batchi, M., Bhagat, S. K., & Enneya, N. (2025). Forecasting short-term rainfall patterns in arid and semi-arid regions using machine learning and deep learning models: A case study from Morocco. *Theoretical and Applied Climatology*, 156, 436.
- Eşidir, K. A. (2025). Makine öğrenimi modelleri ile yetişkin eğitimi analizi: Modellerin karşılaştırmalı performansı. *Elektronik Sosyal Bilimler Dergisi*, 24(2), 946–964. doi:10.17755/esosder.1589887
- European Environment Agency. (2024). *European climate risk assessment: Executive summary*. Retrieved from <https://www.eea.europa.eu/publications/european-climate-risk-assessment>
- Islam, M. S., Shafiuzzaman, M., Mahmud, G., Nowshin, N., Reza, P., Hasan, J., Ahamed, M. F., Nahiduzzaman, M., Ayari, M. A., & Khandakar, A. (2025). Explainable deep learning for rainfall prediction: A CNN-XGBoost hybrid approach in the northern region of Bangladesh. *Neural Computing and Applications*, 37(33), 28125–28160. doi:10.1007/s00521-025-11646-z
- Kaya, B., & Aladağ, C. (2009). Konya şartlarında yağış, sıcaklık ve bitki örtüsü ilişkisi. *Selçuk Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 22, 265–278.
- Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T.-Y. (2017). LightGBM: A highly efficient gradient boosting decision tree. *Advances in Neural Information Processing Systems*, 30, 3146–3154.
- Kumar, V., Kedam, N., Sharma, K. V., Khedher, K. M., & Alluqmani, A. E. (2023). Comparison of machine learning models for rainfall forecasting in urban metropolitan cities. *Sustainability*, 15(18), 13724. doi:10.3390/su151813724
- NASA Langley Research Center. (2024). *POWER: Prediction of Worldwide Energy Resources*. Retrieved from <https://power.larc.nasa.gov>
- Neslihanoglu, S., Ünal, E., & Yozgatligil, C. (2021). Performance comparison of filtering methods on modelling and forecasting the total precipitation amount: A case study for Muğla in Turkey. *Journal of Water and Climate Change*, 12(4), 1071–1085.
- Neyestani, A., Asgari, F., & Asgari, V. (2025). Application of machine and deep learning models to forecast daily precipitation over the western part of Iran. *Meteorological Applications*, 32, e70143.

- Özçelik, M. S. (2025). İklim değişikliğinin buhar basıncı açığı ve yağışın günlük zamanlaması üzerindeki etkisi: Isparta örneği (2005–2024). *Turkish Journal of Forestry*, 26(3), 262–268.
- Palta, Ş., Yurtseven, İ., & Aksay, H. (2019). Göksu Nehri Havzası'nın yağış-akış ilişkileri. *Bartın Orman Fakültesi Dergisi*, 21(3), 860–872.
- Papacharalampous, G. A., Tyralis, H., Doulamis, N., & Doulamis, A. (2024). Uncertainty estimation of machine learning spatial precipitation predictions from satellite data. *Machine Learning: Science and Technology*, 5(3), 035044. doi:10.1088/2632-2153/ad63f3
- Reis, S., Bayrak, T., Yalçın, A., Atasoy, M., Nişancı, R., & Ekercin, S. (2008). Rize bölgesinde yağış-heyelan ilişkisi. *Jeodezi ve Jeoinformasyon Dergisi*, 99, 5–9.
- Salaeh, N., Dittthakit, P., Pinthong, S., Hasan, M. A., Islam, S., Mohammadi, B., & Linh, N. T. T. (2022). Long-short term memory technique for monthly rainfall prediction in Thale Sap Songkhla River Basin, Thailand. *Symmetry*, 14(8), 1599.
- Temuçin Kılıçer, S. (2025). *Uydu verilerine dayalı zaman serileri ile makine öğrenmesi ve derin öğrenme algoritmaları kullanılarak Samsun ilinde yağış tahmini* (Doktora tezi). Ondokuz Mayıs Üniversitesi Lisansüstü Eğitim Enstitüsü, Samsun.
- Thomas, V. (2023). *Risk and resilience in the era of climate change*. Singapore: Palgrave Macmillan.
- Trojovský, P., & Dehghani, M. (2022). Pelican optimization algorithm: A novel nature-inspired algorithm for engineering applications. *Sensors*, 22(3), 855.
- Tunç, B. N., & Atasever, Ü. H. (2025). CatBoost algoritmasının taşınmaz değerlemede kullanımı: Bayesian hiperparametre optimizasyonu ile karşılaştırmalı analiz. *Niğde Ömer Halisdemir Üniversitesi Mühendislik Bilimleri Dergisi*, 14(2), 668–679. doi:10.28948/ngumuh.1597288
- Tyralis, H., Papacharalampous, G. A., Doulamis, N., & Doulamis, A. (2023). Merging satellite and gauge-measured precipitation using LightGBM with an emphasis on extreme quantiles. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 16, 6969–6979. doi:10.1109/JSTARS.2023.3297013
- Uslu, İ., Çelik, Z., Yüceerim, G., Dirama, A., Ozdarici Ok, A., & Yürekli, K. (2025). İklim değişikliğinin izinde: Uzun dönem sıcaklık ve yağış verilerinin analizi: Menemen Ovası örneği. *ANADOLU Journal of Aegean Agricultural Research Institute*, 35(2), 156–171.
- Wang, X., Wang, Y., Yuan, P., Wang, L., & Cheng, D. (2021). An adaptive daily runoff forecast model using VMD-LSTM-PSO hybrid approach. *Hydrological Sciences Journal*, 66(9), 1488–1502. doi:10.1080/02626667.2021.1937631
- World Meteorological Organization. (2023). *State of the global climate 2022*. Geneva, Switzerland: Author. Retrieved from <https://wmo.int/publication-series/state-of-global-climate-2022>
- Yeşilyurt, S. E., & Dalkılıç, H. Y. (2021). XGBoost ve gradient boost machine ile günlük nehir akımı tahmini. In *3rd International Symposium of Engineering Applications on Civil Engineering and Earth Sciences*.
- Zhang, T., Song, S., Li, S., Ma, L., Pan, S., & Han, L. (2019). Research on gas concentration prediction models based on LSTM multidimensional time series. *Energies*, 12(1), 161. doi:10.3390/en12010161
- Zhou, H., Schertzer, D., & Tchiguirinskaia, I. (2025). Combining recurrent neural networks with variational mode decomposition and multifractals to predict rainfall time series. *Hydrology and Earth System Sciences*, 29, 4437–4455.
- Zhou, X., Ma, X., & Wang, H. (2025). Urban air quality management: PM2.5 hourly forecasting with POA–VMD and LSTM. *Processes*, 13(8), 2482.