



Yarı Nemli Bir Bölgede Günlük Referans Bitki Su Tüketiminin Tahmini İçin Üç Farklı Makine Öğrenimi Modellerinin Geliştirilmesi ve Karşılaştırılması

Development and Comparison of Three Different
Machine Learning Models for Estimating
Daily Reference Evapotranspiration in
Sub-Humid Region

Sevda TAŞAN¹, Mehmet TAŞAN²

¹Ondokuz Mayıs Üniversitesi, Ziraat Fakültesi, Tarımsal Yapılar ve Sulama Bölümü, Samsun
· sevda.safi@omu.edu.tr · ORCID > 0000-0002-4335-4074

²Karadeniz Tarımsal Araştırma Enstitüsü, Toprak ve Su Kaynakları Bölümü, Samsun
· mehmettasan17@gmail.com · ORCID > 0000-0002-5592-5022

Makale Bilgisi/Article Information

Makale Türü/Article Types: Araştırma Makalesi/Research Article

Geliş Tarihi/Received: 29 Kasım/November 2022

Kabul Tarihi/Accepted: 07 Ocak/January 2023

Yıl/Year: 2023 | **Cilt-Volume:** 38 | **Sayı-Issue:** 2 | **Sayfa/Pages:** 235-254

Atıf/Cite as: Taşan, S., Taşan, M. "Yarı Nemli Bir Bölgede Günlük Referans Bitki Su Tüketiminin Tahmini İçin Üç Farklı Makine Öğrenimi Modellerinin Geliştirilmesi ve Karşılaştırılması" Anadolu Tarım Bilimleri Dergisi, 38(2), Haziran 2023: 235-254.

Sorumlu Yazar/Corresponding Author: Sevda TAŞAN

YARI NEMLİ BİR BÖLGEDE GÜNLÜK REFERANS BİTKİ SU TÜKETİMİNİN TAHMİNİ İÇİN ÜÇ FARKLI MAKİNE ÖĞRENİMİ MODELLERİNİN GELİŞTİRİLMESİ VE KARŞILAŞTIRILMASI

ÖZ

Su kaynaklarının planlanması ve tarımsal su yönetimi için referans bitki su tüketiminin (ET_0) doğru tahmin edilmesi oldukça önemlidir. Bu çalışmada üç farklı makine öğrenimi yönteminin, çok değişkenli uyarlanabilir regresyon eğrileri (MARS), rassal ağaç (RT) ve gauss süreç regresyonu (GPR), yarı nemli iklim koşullarına sahip bir bölgeden alınan iklim verileri kullanılarak günlük ET_0 'ı tahmin etme performansları araştırılmıştır. Modelleri eğitmek ve test etmek için Bafra ilçesinde 2018–2020 döneminde toplanan günlük meteorolojik verilerin beş girdi kombinasyonu [günlük ortalama (T_{ort}), maksimum (T_{max}) ve minimum sıcaklık (T_{min}), rüzgar hızı (u_2), bağıl nem (RH_{ort}) ve güneşlenme şiddeti (R_s) hem tam hem de eksik kombinasyonları dahil] kullanılmıştır. Makine öğrenimi modellerinin performansı ise FAO-56 Penman Monteith (FAO-56 PM) standart denklemi ile karşılaştırılmıştır. Modellerin doğruluğunu değerlendirmek için dört farklı istatistiksel performans indeksi kullanılmıştır [belirleme katsayısı (R^2), ortalama mutlak hata (MAE), hata kareleri toplamının karekökü (RMSE) ve Nash–Sutcliffe etkinlik katsayısı (NSE)]. Sonuçlar, genel olarak MARS modellerinin RT ve GPR modellerinden daha iyi performans gösterdiğini ortaya koymuştur. Tüm verilerin girdi olarak kullanıldığı beşinci senaryo için MARS modeli tarafından tahmin edilen ET_0 değerlerinin FAO-56 PM değerleri ile iyi bir uyum içinde olduğu belirlenmiştir ($R^2=0.982$, MAE=0.250, RMSE=0.305, NSE=0.965). Bununla birlikte meteorolojik veriler tam olmadığında bile sadece T_{ort} , RH_{ort} ve R_s ile oldukça yüksek günlük ET_0 tahminleri elde edilmiştir. Genel olarak, radyasyona dayalı makine öğrenimi modelleri, sıcaklığa dayalı makine öğrenimi modellerinden daha iyi performans göstermiştir. Sonuçlar, yarı nemli iklim özelliklerine sahip bir bölgede MARS modelinin ET_0 'ı oldukça verimli ve doğru bir şekilde modellemek için etkili bir şekilde kullanılabileceğini göstermektedir.

Anahtar Kelimeler: Çok Değişkenli Uyarlanabilir Regresyon Eğrileri, Modelleme, Rassal Ağaç, Referans Bitki Su Tüketimi.



DEVELOPMENT AND COMPARISON OF THREE DIFFERENT MACHINE LEARNING MODELS FOR ESTIMATING DAILY REFERENCE EVAPOTRANSPIRATION IN SUB-HUMID REGION

ABSTRACT

Accurate estimation of reference evapotranspiration (ET_0) is very important for water resource planning and agricultural water management. This study investigates the performance of three different machine learning methods, multivariate adaptive regression splines (MARS), random tree (RT), and gaussian process regression (GPR), for predicting daily ET_0 using climate data from a sub-humid climate region. Five input combinations [including both complete and incomplete combinations of daily average (T_{avg}), maximum (T_{max}) and minimum (T_{min}) temperature wind speed (u_2), relative humidity (RH_{avg}), and solar radiation (R_s)] of daily meteorological data collected in Bafra district during the 2018–2020 period were used for model training and testing. The performance of machine learning models was compared with the standard FAO-56 Penman-Monteith equation (FAO-56 PM). Four different statistical performance indices were used to evaluate the accuracy of the models [coefficient of determination (R^2), mean absolute error (MAE), root mean square error (RMSE), and Nash-Sutcliffe efficiency coefficient (NSE)]. The results revealed that MARS models outperform the RT and GPR models. For scenario 5, using all data as input, the ET_0 values estimated by the MARS models agreed well with the FAO-56 PM values ($R^2=0.982$, $MAE=0.250$, $RMSE=0.305$, $NSE=0.965$). However, even if the meteorological data are incomplete, very high daily ET_0 estimates have been obtained using only T_{avg} , RH_{avg} , and R_s . Overall, the radiation-based machine learning models outperformed the temperature-based machine learning models. The results show that the MARS model can be used effectively to model ET_0 quite efficiently and accurately in a region with sub-humid climate characteristics.

Keywords: Multivariate Adaptive Regression Splines, Modeling, Random Tree, Reference Evapotranspiration.



1. GİRİŞ

Bitki su tüketimi, toprak yüzeyinden buharlaşma ve bitki yapraklarından terleme yoluyla atmosfere geçen su miktarının toplamı olup, bitkiler tarafından belirli bir süre boyunca ihtiyaç duyulan suyu temsil etmektedir (Aghelpour ve Norooz-Valashedi, 2022). Yüzey suyunun ortama göre bir ölçüsü olduğu için hidrolojik

döngüdeki en önemli bileşenlerden biri olarak kabul edilmektedir (Jing ve ark., 2019). Hidrolojik döngüdeki hayati rolü nedeniyle, bir bölgedeki sulama suyu miktarının hesaplanmasında ve programlanmasında önemli bir rol oynamaktadır. Bu nedenle bitki su tüketiminin doğru bir şekilde belirlenmesi, tarımsal amaçlı su kaynaklarının yöneticileri ve kullanıcıları için büyük önem taşımaktadır.

Bitki su tüketimi doğrudan Bowen oranı enerji dengesi yaklaşımı, girdap kovaryans teknikleri ve lizimetreler kullanılarak çeşitli deneysel yöntemlerle ölçülebilir (Roy ve ark., 2021). Alternatif olarak meteorolojik veriler kullanılarak referans bitki su tüketimi (ET_0) hesaplanarak dolaylı olarak da elde edilebilir. Bu dolaylı yöntemde bitki su tüketimi, referans bitki su tüketimine (ET_0) karşılık gelen bitki katsayısı (k_c) ile çarpılmasıyla elde edilebilir (Allen ve ark., 1998). Bu nedenle bitki su tüketiminin doğru hesaplanabilmesi için ET_0 'ın doğru bir şekilde tahmin edilmesi gerekmektedir. ET_0 'ı dolaylı olarak tahmin etmek için geliştirilen çok sayıda matematiksel model bulunmaktadır. ET_0 'ı hesaplamak için sıcaklık tabanlı Thornthwaite (1948), Blaney-Criddle (Blaney ve Criddle, 1950) ve Hargreaves-Samani (Hargreaves ve Samani, 1985), radyasyon tabanlı (Makkink 1957), Priestley-Taylor (Priestley ve Taylor, 1972) ve kombinasyon tabanlı FAO-56 Penman-Monteith (Allen ve ark., 1998) modelleri önerilmiştir. Bununla birlikte FAO-56 Penman-Monteith (FAO-56 PM) yönteminin, diğer denklemlerle karşılaştırıldığında üstün olduğu birçok çalışmada doğrulanmıştır (Djaman ve ark., 2016). Bu nedenle FAO-56 PM yöntemi, Birleşmiş Milletler Gıda ve Tarım Örgütü (FAO), Uluslararası Sulama Drenaj Komisyonu (ICID) ve ASCE-Evapotranspirasyon Sulama ve Hidroloji Komitesi (ASCE-EWRI, 2005) tarafından diğer modelleri değerlendirmek için standart model olarak önerilmektedir. Bu dolaylı yöntem, karmaşıklık veya maliyet nedeniyle doğrudan ölçümlerin mevcut olmadığı dünyanın birçok yerinde yaygın olarak kullanılmaktadır (Allen ve ark., 1998). Ancak FAO-56 PM denklemini hesaplamak için maksimum ve minimum hava sıcaklığı, rüzgar hızı, bağıl nem ve net radyasyon gibi çeşitli iklim parametrelerine ihtiyaç vardır. Bu da FAO-56 PM modelinin dünya çapında uygulanmasını önemli ölçüde sınırlandırmaktadır.

ET_0 , doğrusal olmayan bir hidrolojik süreç olduğundan, çeşitli makine öğrenmesi yöntemleri değişkenler arasındaki doğrusal olmayan ilişkileri modellemek için başarı ile kullanılmıştır (Huang ve ark., 2019; Gocić ve Arab Amiri, 2021). ET_0 'ın tahmin doğruluğu, eksik veri bulunması durumunda bağımlı ve bağımsız değişkenler arasındaki doğrusal olmayan ilişkilerin üstesinden gelme konusundaki kapasiteleri nedeniyle makine öğrenmesi modelleri tarafından önemli ölçüde iyileştirilebilir. Çeşitli makine öğrenmesi yöntemleri, dünyanın çeşitli iklim bölgelerinde ET_0 'ı tahmin etmek için önerilmiştir. Bunlar arasında en yaygın olanı nöron tabanlı yapay sinir ağlarıdır (YSA) (Makwana ve ark., 2023). Çekirdek tabanlı makine öğrenmesi algoritmaları arasında en yaygın kullanılan modeller gauss süreç regresyonu (GPR) ve destek vektör makineleridir (DVM). Bu modeller ET_0 tahmini için büyük ölçüde uygulanmıştır (Karbasi, 2018; Yin ve ark., 2021).

Çok değişkenli uyarlanabilir regresyon eğrileri (MARS) gibi eğri tabanlı makine öğrenmesi modelleri de ET_0 'ın tahmininde başarı ile kullanılmıştır (Dong ve ark., 2022). Mehdizadeh ve ark. (2017)'nin aylık ortalama ET_0 tahmini için MARS, DVM, gen ekspresyon programlama (GEP) ve ampirik denklemlerin performansını karşılaştırdıkları çalışmada en iyi tahmin performansını MARS modelinin gösterdiği bildirilmiştir. ET_0 tahmininde yaygın olarak kullanılan bir diğer algoritma ise ağaç tabanlı makine öğrenmesi algoritmalarıdır. Bunlar arasında en yaygın olanları rassal ağaç (RT) (Salam ve Islam, 2020) ve rassal orman (RF) (Wu ve ark., 2020) algoritmalarıdır. Salam ve Islam (2020) nemli bir iklim bölgesinde ET_0 'ı tahmin etmek için rassal ağaç (RT), rassal orman (RF) ve destek vektör makinesi (DVM) modellerini karşılaştırmıştır. Çalışmada RT modelinin, ET_0 'ı tahmin etmede diğer modellerden daha iyi performans gösterdiğini belirlemişlerdir. Ayrıca rüzgar hızı ve güneş radyasyonu değişkenlerinin, gözlemlenen ve tahmin edilen ET_0 'ı en çok etkileyen iki değişken olduğunu bildirmişlerdir. Chen ve ark. (2014) beş ampirik model ve üç süreç tabanlı modelin (YSA, RT ve DVM) Çin ve komşu bölgelerindeki bitki su tüketimini tahmin etme performanslarını karşılaştırmışlardır. Süreç tabanlı modellerin daha yüksek tahmin performansı gösterdiği bildirilmiştir. Yukarıdaki açıklamalar doğrultusunda farklı iklim bölgeleri ve farklı girdi değişkenlerine göre modellerin tahmin performanslarının değişkenlik gösterdiği söylenebilir. Bu sonuç muhtemelen, makine öğrenimi modellerinin farklı veri kümelerinde farklı performans göstermesinden kaynaklanmaktadır (Zhou, 2016). Dolayısıyla, ET_0 'ın tahmini için en doğru modelin seçimi konusunda ortak bir fikir birliği yoktur, bu nedenle yeni modeller gereklidir ve doğrulanmalıdır.

Bu çalışma, Türkiye'nin yarı nemli iklim özelliklerine sahip önemli tarımsal faaliyetlerin gerçekleştirildiği Bafra ovasında günlük referans bitki su tüketimini (ET_0) tahmin etmek için yapılmıştır. Bu amaçla beş farklı girdi kombinasyonu altında üç farklı makine öğrenimi (MARS, RT ve GPR) modelinin tahmin performansları karşılaştırılmıştır. FAO-56 PM denklemi modellerden elde edilen ET_0 tahminlerini karşılaştırmak için referans olarak kullanılmıştır. Çünkü bu, ölçülen bir lizimetre verisi yoksa, standart prosedür olarak kabul edilmektedir (Irmak ve ark., 2003). Çalışmanın ana hedefleri şunlardır: 1) tam ve eksik meteorolojik girdi değişkenleri ile günlük ET_0 tahmini için bir makine öğrenimi modeli geliştirmek, 2) çeşitli makine öğrenimi modellerinin performanslarını karşılaştırmak ve 3) çalışma alanı için günlük ET_0 'ı tahmin etmek için en uygun makine öğrenimi modelini önermektir.

2. MATERYAL VE YÖNTEM

2.1. Çalışma Alanı ve Verilerin Toplanması

Samsun ili, Bafra ilçesi Türkiye'nin kuzeyinde, Karadeniz bölgesinde 41° 28'-41° 45' kuzey enlemleri ve 35° 43'- 35° 58' doğu boylamları arasında yer almaktadır. Bölgede Karadeniz iklimi hakim olup, yarı nemli iklim özelliği göstermektedir. Bölgenin uzun yıllar ortalama en yüksek sıcaklık değeri Ağustos ayında 23 °C olup, ortalama en düşük sıcaklık değeri 5.7 °C ile Şubat ayında gerçekleşmiştir. Yıllık toplam yağış miktarı ise 794.2 mm'dir (Taşan, 2018). Bölgede yoğun olarak çeltik tarımı yapılmakla birlikte çeşitli yazlık (domates, biber, karpuz) ve kışlık (pırasa, kırmızı ve beyaz baş lahana) sebze ve buğday tarımı da yapılmaktadır.

Bu çalışmada referans bitki su tüketimini (ET_0) farklı makine öğrenmesi yöntemleriyle tahmin etmek için, hava sıcaklığı (T), rüzgar hızı (u_2), bağıl nem (RH) ve güneşlenme şiddetini (Rs) içeren günlük iklim verileri girdi olarak kullanılmıştır. İklim verileri, Meteoroloji Genel Müdürlüğü'ne (MGM) bağlı Bafra Toprak Su Araştırma (TAGEM) meteoroloji istasyonundan temin edilmiştir. Veriler Ocak 2018- Aralık 2020 tarihleri arasında kapsamakta olup, günlük değerlerdir. İstasyondaki eksik veriler doğrusal enterpolasyon yöntemi ile tamamlanmıştır (Zhu ve ark., 2020).

2.2. Referans Bitki Su Tüketimi modeli (Tam Veri Seti ile FAO56 Penman-Monteith)

Makine öğrenmesi modellerinin performansını geliştirmek ve değerlendirmek için FAO tarafından önerilen FAO56 Penman Monteith denklemi ile hesaplanan referans bitki su tüketimi (ET_0) değerleri referans olarak alınmıştır. Bu yöntem farklı iklimlerde ET_0 'ı tahmin etmek için evrensel olarak en çok kabul gören denklemdir (Fan ve ark., 2018).

$$ET_0 = \frac{0.408\Delta(R_n - G) + \gamma \frac{900}{T + 273} u_2 (e_s - e_a)}{\Delta + \gamma (1 + 0.34 u_2)} \quad (1)$$

ET_0 , referans bitki su tüketimi (mm gün⁻¹); R_n , bitki yüzeyindeki net radyasyon (MJ m⁻² gün⁻¹); G, toprak ısı akış yoğunluğu (MJ m⁻² gün⁻¹); T, 2 m yükseklikte ortalama günlük hava sıcaklığı (°C); u_2 , 2 m yükseklikte rüzgar hızı (m s⁻¹), e_s , doymun buhar basıncı (kPa); e_a , gerçek buhar basıncı (kPa); $e_s - e_a$, doymun buhar basıncı açığı (kPa); Δ , buhar basıncı eğrisinin eğimi (kPa °C⁻¹); γ , psikrometrik sabittir (kPa °C⁻¹). FAO56 PM denkleminin ayrıntıları, Allen ve ark. (1998)'de bulunabilir.

2.3. Referans Bitki Su Tüketimini Tahmin Etmek için Kullanılan Makine Öğrenimi Algoritmaları

2.3.1. Çok Değişkenli Uyarlanabilir Regresyon Eğrileri (MARS)

Çok değişkenli uyarlanabilir regresyon eğrileri (MARS) modeli, Friedman (1991) tarafından önerilen parametrik olmayan bir regresyon yöntemidir. Bu yöntemde hedef değişkeni tahmin etmek için esnek regresyon modelleri oluşturulur. Bu algoritmada girdi değişkenleri aralıklara (alt kümelere) ayrılır ve her aralığa bir temel fonksiyon uydurulur. Bağımsız değişkenler hakkındaki bilgiler temel fonksiyon ile temsil edilir. Bir temel fonksiyonun başlangıç ve bitiş noktaları düğüm olarak adlandırılır ve belirli bir aralıkta tanımlanır. Temel fonksiyonlar, girdi ve hedef değişkenler arasındaki ilişkiyi aşağıdaki gibi gösterir.

$$y = \max(0, x - c) \quad (2)$$

$$y = \max(0, c - x) \quad (3)$$

Burada c seçilen eşik değeri, x girdi değişkeni ve y ise hedef değişkendir.

Bu çalışmada oluşturulacak maksimum temel fonksiyon sayısı deneme yanılma yoluyla belirlenmiştir. Friedman (1991), ceza için en iyi değerin 2 ile 4 arasında olduğunu bildirmiştir. Bu çalışmada ceza değeri 2 olarak alınmıştır. Son derece karmaşık modellerden kaçınmak için etkileşim derecesi (\max_degree) 2 olarak ayarlanmıştır. MARS hakkında daha ayrıntılı bilgi Cheng ve Cao (2014)'de bulunabilir. Bu çalışmada MARS modelini eğitmek ve doğrulamak için R Studio programı *earth* (enhanced adaptive regression through hinges) paketi kullanılmıştır.

2.3.2. Rassal Ağaç (RT)

Rassal ağaç (RT) modeli, çok sayıda bireysel sınıflandırıcı üreten, değiştirilmiş bir topluluk öğrenme algoritmasıdır. Rassal ağaçlar ilk olarak Leo Breiman ve Adele Cutler tarafından tanıtılmıştır. RT modeli, orman olarak da bilinen bir ağaç tahmin edicileri topluluğudur. Bu yöntem, rassal orman (RF) ve torbalama olarak bilinen iki topluluk öğrenme algoritmasının birleşimidir (Mishra ve Ratha, 2016). Bu yöntemde bir karar ağacı oluşturmak için rastgele bir veri seti oluşturmak üzere bir torbalama fikri kullanır. Rassal bir ormanda, her düğüm, o düğümde rassal olarak seçilen tahmin edicilerin alt kümesi arasındaki en iyi kullanılan bölünür. Bu algoritma hem sınıflandırma hem de regresyon problemlerinde kullanılabilir. RT modeli hakkında daha fazla ayrıntı Mishra ve Ratha (2016)'da bulunabilir. Bu çalışmada RT modelini eğitmek ve doğrulamak için Weka 3.8.5 yazılımı kullanılmıştır.

2.3.3. Gauss Süreç Regresyonu (GPR)

Gauss süreç regresyonu (GPR), birçok alanda geniş uygulama alanı olan Bayes teorisine dayalı bir çerçevedir. Bu model Rasmussen ve Williams (2006) tarafından ortak bir Gauss dağılımına sahip karmaşık bir rasgele değişkenler kümesi olarak tanımlanmıştır. GPR genellikle iki fonksiyonla açıklanır. Bunlar: ortalama ve kovaryans fonksiyonlarıdır. Ortalama fonksiyon bir vektör iken, kovaryans fonksiyonu bir matristir. Kovaryans değeri, girdiler ve çıktılar arasındaki korelasyonu belirlemektedir. GPR modeli, parametrik olmayan kernel tabanlı bir kara kutu tekniği olarak nitelendirilebilir. Bu çalışmada en iyi kernel fonksiyonu, deneme-yanılma yoluyla belirlenmiştir. GPR modeli hakkında daha fazla ayrıntılı bilgi Rasmussen ve Williams (2006)'da bulunabilir. Bu çalışmada GPR modelini eğitmek ve doğrulamak için Weka 3.8.5 yazılımı kullanılmıştır.

2.4. Girdi Kombinasyonları

Bu çalışmada makine öğrenmesi modellerinin eğitimi ve testi için Tort, Tmax, Tmin, RHort, u_2 ve Rs dahil olmak üzere günlük meteorolojik değişkenlerin 5 farklı girdi kombinasyonu dikkate alınmıştır (Çizelge 1). Bu girdi kombinasyonları oluşturulurken radyasyon tabanlı modeller, sıcaklık tabanlı modeller ve kombinasyon modellerinde kullanılan girdi değişkenleri göz önüne alınmıştır. Böylece girdiler, hem tüm meteorolojik değişkenlerin kullanıldığı tam kombinasyondan (kombinasyon 5) hem de farklı değişkenlerin bir araya getirildiği eksik kombinasyonlardan (kombinasyon 1-4) oluşmaktadır (Çizelge 1). Eğitim aşamasından önce tüm veriler iki gruba ayrılmıştır: 2018 ve 2019 yılı verileri ($n=730$) makine öğrenmesi modellerini eğitmek için kullanılırken, 2020 yılı verileri ($n=366$) modelleri doğrulamak için kullanılmıştır. Her modelin optimal mimarileri ve hiperparametre özellikleri eğitim aşamasında deneme yanılma yoluyla belirlenmiştir.

Çizelge 1. Üç farklı makine öğrenmesi modeli için oluşturulan girdi kombinasyonları

Table 1. Input combinations created for three different machine learning models

Kombinasyonlar	Modeller			Girdi Değişkenleri
	MARS	RT	GPR	
1	MARS1	RT1	GPR1	Tort, Rs
2	MARS2	RT2	GPR2	Tort, Rs, RH
3	MARS3	RT3	GPR3	Tort, Tmin, Tmax
4	MARS4	RT4	GPR4	Tort, Tmin, Tmax, Rs
5	MARS5	RT5	GPR5	Tort, Tmin, Tmax, Rs, u_2 , RH

2.5. İstatistiksel Değerlendirme

Modellerin tahmin doğruluğunu değerlendirmek için yaygın olarak kullanılan dört istatistiksel gösterge kullanılmıştır. Bunlar: belirleme katsayısı (R^2), MAE (Ortalama Mutlak Hata), RMSE (Ortalama Hata Kareleri Toplamının Karekökü) ve Nash–Sutcliffe etkinlik katsayısı (NSE) (Nash ve Sutcliffe, 1970)'dir. İstatistiksel göstergelerin matematiksel denklemleri aşağıda açıklanmıştır.

$$R^2 = \frac{\sum_{i=1}^n ((O_i - \bar{O})(P_i - \bar{P}))^2}{\sum_{i=1}^n (O_i - \bar{O})^2 \sum_{i=1}^n (P_i - \bar{P})^2} \quad (4)$$

$$MAE = \frac{\sum |O_i - P_i|}{n} \quad (5)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (O_i - P_i)^2}{n}} \quad (6)$$

$$NSE = 1.0 - \frac{\sum_{i=1}^n (O_i - P_i)^2}{\sum_{i=1}^n (O_i - \bar{O})^2} \quad (7)$$

Burada O_i , P_i , sırasıyla belirli bir değişken için ölçülen ve tahmin edilen değerleri temsil eder ve, \bar{O} , \bar{P} ve n sırasıyla ölçülen ET_0 'ın ortalaması, tahmin edilen ET_0 'ın ortalaması ve gözlem sayısıdır. Daha yüksek R^2 ve NSE değerleri, daha iyi model performansını gösterirken, tersine daha düşük MAE ve RMSE değerleri ise daha iyi tahmin doğruluğunu göstermektedir.

Bu çalışmada makine öğrenmesi modellerinin gereksinimlerini karşılamak için modelin eğitiminden önce hem girdi hem de çıktı değişkenleri 0 ile 1 aralığında aşağıdaki gibi normalize edilmiştir.

$$X_{norm} = \frac{X_i - X_{min}}{X_{max} - X_{min}} \quad (8)$$

Burada X_i ham girdi verisidir, X_{norm} X_i 'nin normalize edilmiş değeridir, X_{max} ve X_{min} sırasıyla ham verilerin maksimum ve minimum değerleridir.

3. BULGULAR

3.1. Veri Özeti İstatistikleri

Bu çalışmada referans bitki su tüketimi (ET_0) tahmininde makine öğrenmesi modellerini geliştirmek için 1 Ocak 2018 ile 31 Aralık 2020 tarihleri arasındaki günlük Tort,

Tmin, Tmax, Rs, u_2 ve RHort değerleri kullanılmıştır. Modellerde kullanılan veri setleri iki alt kümeye ayrılmıştır. Bunlar, eğitim ve test veri kümelerini içermektedir. Eğitim veri seti 1 Ocak 2018 ile 31 Aralık 2019 tarihleri arasında ölçülen iklim verilerini içerirken, test veri seti ise 1 Ocak 2020 ile 31 Aralık 2020 tarihleri arasındaki gözlemlenen verileri içermektedir. Eğitim ve test veri setleri için ortalama, minimum, maksimum, standart sapma, varyasyon katsayısı, çarpıklık ve basıklık gibi istatistiksel parametreler ve ET_0 ile iklim parametreleri arasındaki korelasyon katsayısı Çizelge 2'de özetlenmiştir. Tüm değişkenler için eğitim ve test veri setlerinin ortalama, en küçük ve en büyük değerlerin birbirine yakın olduğu görülmektedir. ET_0 ile en yüksek korelasyon Rs değişkeni arasında belirlenmiştir (Çizelge 2). Tort ve Tmax ile ET_0 arasında oldukça iyi bir korelasyon belirlenmiştir. RHort ile ET_0 arasında orta derecede negatif korelasyon belirlenmiştir. Liu ve ark. (2017) tarafından yapılan çalışmada da benzer sonuçlar bulunmuştur. En düşük korelasyon ise ET_0 ile u_2 arasında belirlenmiştir.

Çizelge 2. Makine öğrenimi modellerinin eğitim ve test aşamalarında kullanılan değişkenlerin tanımlayıcı istatistikleri

Table 2. Descriptive statistics of variables used in the training and testing phases of machine learning models

Değişken	Veri Seti	AO	Max	Min	SS	VK	Çarpıklık	Basıklık	Korelasyon
Tort	Eğitim	15.36	27.60	1.90	6.50	42.35	0.01	-1.29	0.85
	Test	15.06	26.40	0.70	6.62	43.94	0.00	-1.40	0.83
Tmin	Eğitim	11.05	25.40	-3.10	6.21	56.15	0.07	-1.14	0.79
	Test	10.55	23.20	-2.70	6.15	58.25	0.06	-1.18	0.74
Tmax	Eğitim	20.29	31.80	3.10	7.27	35.84	-0.19	-1.20	0.85
	Test	20.04	33.20	3.50	7.48	37.35	-0.15	-1.33	0.84
RHort	Eğitim	72.61	99.00	36.80	11.83	16.30	0.52	-0.55	-0.30
	Test	74.01	95.20	31.90	10.05	13.58	-0.58	0.89	-0.52
u_2	Eğitim	2.45	7.70	0.70	1.11	45.22	1.12	1.23	-0.10
	Test	2.44	8.80	1.00	1.05	43.10	1.64	4.37	0.11
Rs	Eğitim	14.03	29.58	3.13	7.50	53.49	0.29	-1.11	0.88
	Test	14.74	29.93	3.13	8.07	54.73	0.25	-1.22	0.89
ET_0	Eğitim	2.74	6.14	0.30	1.51	55.13	0.41	-1.02	1.00
	Test	2.77	6.67	0.42	1.62	58.56	0.39	-0.93	1.00

Not: Tort: ortalama hava sıcaklığı (°C), Tmin: en düşük hava sıcaklığı (°C), Tmax: en yüksek hava sıcaklığı (°C), RHort: ortalama bağıl nem (%), u_2 : 2 m yükseklikteki rüzgar hızı ($m s^{-1}$), Rs: güneşlenme şiddeti ($MJ m^{-2}gün^{-1}$), ET_0 : günlük referans bitki su tüketimi ($mm gün^{-1}$), AO: aritmetik ortalama, SS: standart sapma, VK: varyasyon katsayısı (%)

3.2. Beş Girdi Kombinasyonu Altında Üç Makine Öğrenimi Modelinin Performans Değerlendirmesi

Bu çalışmada üç farklı makine öğrenimi algoritmasının (MARS, RT ve GPR) günlük meteorolojik değişkenler olan T_{ort} , T_{min} , T_{max} , R_s , u_2 ve RH_{ort} ile referans bitki su tüketimini (ET_0) tahmin etme yetenekleri araştırılmıştır. Bu amaçla FAO-56 PM yönteminden tahmin edilen ET_0 değerleri, test aşamasında makine öğrenmesi modellerinin performansını değerlendirmek için referans olarak kabul edilmiştir. Bafra'da günlük ET_0 'ı tahmin etmeye yönelik üç farklı makine öğrenimi modelinin istatistiksel analiz sonuçları Çizelge 3, Çizelge 4 ve Çizelge 5'de verilmiştir. Bu çizelgelerde hem eğitim hem de test aşamalarında beş girdi kombinasyonu altında sırasıyla MARS, RT ve GPR modellerinin istatistiksel analiz sonuçları sunulmuştur. Çizelgelerden görüldüğü gibi, tahmin edilen ET_0 değerleri, araştırılan modele ve girdi kombinasyonlarına bağlı olarak önemli ölçüde farklılık göstermiştir. Genel olarak tüm modeller için en düşük tahmin performansını sadece sıcaklık verilerinin girdi olarak kullanıldığı üçüncü kombinasyon göstermiştir. Bununla birlikte en yüksek tahmin performansı tüm değişkenlerin girdi olarak kullanıldığı beşinci kombinasyonda gözlenmiştir. Sonuçlar, genel olarak MARS modelinin, RT ve GPR modelinden daha iyi performansa sahip olduğunu göstermiştir.

Karşılaştırılan MARS modellerinin ET_0 'ı tahmin etmek için en iyi performansı tüm girdi değişkenlerinin kullanıldığı MARS5 için gözlenmiştir. Sadece hava sıcaklığı verilerinin girdi olarak kullanıldığı MARS3 modelinin test alt kümesi için R^2 (0.703), MAE (0.844 mm gün⁻¹), RMSE (1.078 mm gün⁻¹) ve NSE (0.558) değerleri dikkate alındığında en düşük tahmin performansına sahip olduğu belirlenmiştir. Sıcaklık verilerinin yanına güneşlenme şiddetinin (R_s) eklenmesiyle test alt kümesi için MARS3 ve MARS4 modelleri karşılaştırıldığında R^2 değeri %27 artmış, MAE ve RMSE değerleri sırasıyla %36.3 ve %40.4 azalmıştır. MARS1 ve MARS2 modelleri karşılaştırıldığında ise girdi olarak ortalama bağıl nem (RH_{ort}) değişkeninin eklenmesi modelin R^2 ve NSE'sini 0.895 ve 0.818'den 0.955 ve 0.916'a yükseltmiş, MAE ve RMSE'sini ise 0.600 ve 0.691'den 0.389 ve 0.470'e düşürmüştür. Bu sonuçlara benzer olarak Adnan ve ark. (2021) çalışmasında, makwa Dhaka istasyonu için ET_0 tahmininde diğer değişkenlerin yanına RH değişkeninin eklenmesiyle modelin NSE'sinin 0.866'dan 0.922'ye yükseldiğini belirlemiştir. Son olarak MARS5'e rüzgar hızı (u_2) ve bağıl nem (RH) birlikte eklendiğinde modelin R^2 'si %10 artmıştır. Rüzgar hızı ve bağıl nem, ET_0 'ın karmaşık ve doğrusal olmayan davranışını modellemek için iyi bir girdi değişkeni olduğu bulunmuştur. Makwana ve ark. (2023), modele rüzgar hızının eklenmesiyle R^2 'nin %30 arttığını belirlemiş ve girdi kombinasyonuna rüzgar hızı değişkeninin eklenmesinin, ET_0 üzerindeki adveksiyon etkisinden dolayı tahmin doğruluğunu iyileştirdiğini bildirmiştir.

Çizelge 3. Farklı girdi değişkenlerine sahip MARS modelinin eğitim ve test aşamasındaki istatistiksel değerleri

Table 3. Statistical values of the MARS model with different input variables during the training and testing phase

Model	Eğitim Veri Seti				Test Veri Seti			
	R ²	MAE	RMSE	NSE	R ²	MAE	RMSE	NSE
		(mm gün ⁻¹)	(mm gün ⁻¹)			(mm gün ⁻¹)	(mm gün ⁻¹)	
MARS1	0.901	0.358	0.475	0.901	0.895	0.600	0.691	0.818
MARS2	0.963	0.221	0.288	0.963	0.955	0.389	0.470	0.916
MARS3	0.833	0.496	0.615	0.833	0.703	0.844	1.078	0.558
MARS4	0.909	0.354	0.455	0.909	0.893	0.538	0.642	0.843
MARS5	0.987	0.138	0.175	0.987	0.982	0.250	0.305	0.965

Not: Tüm modeller arasındaki en iyi istatistiksel göstergeler koyu renkle işaretlenmiştir.

RT modelleri, MARS ve GPR modelleriyle karşılaştırıldığında tüm girdi kombinasyonları için eğitim veri setinde en yüksek R²'yi ve en düşük MAE ve RMSE'yi veren model olmuştur. Bunu sırasıyla MARS ve GPR modelleri izlemiştir. Test aşamasında ise RT modelleri daha düşük performans göstermiş ve üçüncü girdi kombinasyonu (GPR3) hariç diğer kombinasyonlar için MARS modeli, diğer iki makine öğrenmesi modelinden daha yüksek tahmin performansı göstermiştir. Bu durum RT modellerinde belirli ölçüde aşırı uyum sorunu olduğunu göstermiştir. Yukarıdaki durumun nedeni, ağaç algoritmalarının yapısı ile ilgili olabilir. En yüksek performans test alt kümesi için tüm girdi değişkenlerinin kullanıldığı RT5 modelinde elde edilmiş ve bu modelin R², MAE, RMSE ve NSE değerleri sırasıyla 0.921, 0.388 mm gün⁻¹, 0.499 mm gün⁻¹ ve 0.905 olarak belirlenmiştir (Tablo 4). Sadece Tort, Tmin ve Tmax değişkenlerinin girdi olarak kullanıldığı RT3 modeli ise test aşamasında en düşük tahmin performansını vermiştir (R²=0.504, MAE=0.960 mm gün⁻¹, RMSE=1.213 mm gün⁻¹, NSE=0.440). Bununla birlikte RT modelleri, 5. kombinasyon haricinde ilk dört kombinasyon için RMSE'de %10.85 ile %22.27'lik bir artışla MARS modeliyle karşılaştırılabilir günlük ET0 tahminleri üretmiştir. Tüm girdi değişkenlerinin kullanıldığı kombinasyon 5'de ise MARS5 modeli, RT5 modellerinden oldukça yüksek tahmin performansı göstermiştir. YSA modeli ile karşılaştırıldığında RT modeli için RMSE'deki artış sırasıyla %63.61 olarak meydana gelmiştir.

Çizelge 4. Farklı girdi değişkenlerine sahip RT modelinin eğitim ve test aşamındaki istatistiksel değerleri

Table 4. Statistical values of the RT model with different input variables during the training and testing phase

Model	Eğitim Veri Seti				Test Veri Seti			
	R ²	MAE	RMSE	NSE	R ²	MAE	RMSE	NSE
		(mm gün ⁻¹)	(mm gün ⁻¹)			(mm gün ⁻¹)	(mm gün ⁻¹)	
RT1	1.000	0.010	0.019	1.000	0.821	0.630	0.808	0.752
RT2	1.000	0.013	0.022	1.000	0.926	0.419	0.521	0.897
RT3	1.000	0.007	0.016	1.000	0.504	0.960	1.213	0.440
RT4	1.000	0.011	0.020	1.000	0.805	0.611	0.785	0.765
RT5	1.000	0.016	0.025	1.000	0.921	0.388	0.499	0.905

Not: Tüm modeller arasındaki en iyi istatistiksel göstergeler koyu renkle işaretlenmiştir.

Eğitim ve test alt kümeleri için günlük ET_0 'ı tahmin etmek amacıyla beş girdi kombinasyonu altında GPR'nin en iyi performansı, kernel fonksiyonu olarak radial temelli fonksiyon (rbf) kullanıldığında ve beşinci kombinasyonda gözlenmiştir ($R^2=0.911$, $MAE=0.532$ mm gün⁻¹, $RMSE=0.638$ mm gün⁻¹, $NSE=0.845$). GPR modelleri arasında en düşük tahmin performansını veren model ise Tort, Tmin ve Tmax değişkenlerinin kullanıldığı üçüncü kombinasyondan (GPR3) elde edilmiştir (test veri seti için $R^2=0.687$, $MAE=0.855$ mm gün⁻¹, $RMSE=1.005$ mm gün⁻¹, $NSE=0.616$). Bununla birlikte üçüncü girdi kombinasyonu için karşılaştırılan üç makine öğrenmesi modeli arasında ise en yüksek tahmin performansını GPR3 modeli göstermiştir. Bunun yanı sıra GPR modeli, 2. ve 5. kombinasyon haricinde diğer kombinasyonlar için RMSE'de %18.2 ile 21.41'lik bir artışla MARS modeliyle karşılaştırılabilir günlük ET_0 tahminleri üretmiştir. Tüm girdi değişkenlerinin kullanıldığı kombinasyon 5'de ise MARS5 modeli, GPR5 modelinden oldukça yüksek tahmin performansı göstermiştir. MARS5 modeli ile karşılaştırıldığında GRP5 modeli için RMSE'deki artış %109.18 olarak meydana gelmiştir.

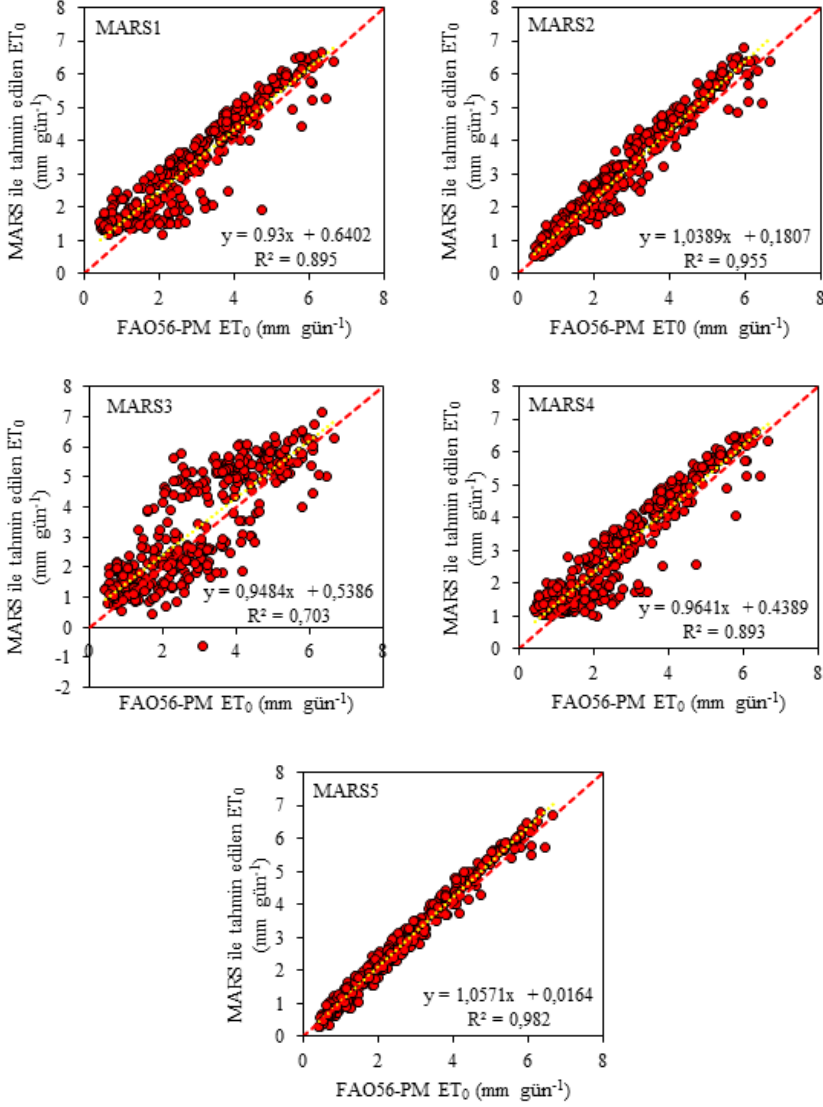
Çizelge 5. Farklı girdi parametrelerine sahip GPR modelinin eğitim ve test aşamasındaki istatistiksel değerleri

Table 5. Statistical values of the GPR model with different input variables during the training and testing phase

Model	Kernel	Eğitim Veri Seti				Test Veri Seti			
		R ²	MAE (mm gün ⁻¹)	RMSE (mm gün ⁻¹)	NSE	R ²	MAE (mm gün ⁻¹)	RMSE (mm gün ⁻¹)	NSE
GPR1	rbf	0.869	0.606	0.737	0.761	0.890	0.718	0.839	0.732
GPR2	rbf	0.926	0.560	0.669	0.803	0.931	0.635	0.734	0.795
GPR3	rbf	0.665	0.964	1.123	0.444	0.687	0.855	1.005	0.616
GPR4	rbf	0.854	0.546	0.670	0.802	0.863	0.637	0.759	0.781
GPR5	rbf	0.910	0.482	0.590	0.847	0.911	0.532	0.638	0.845

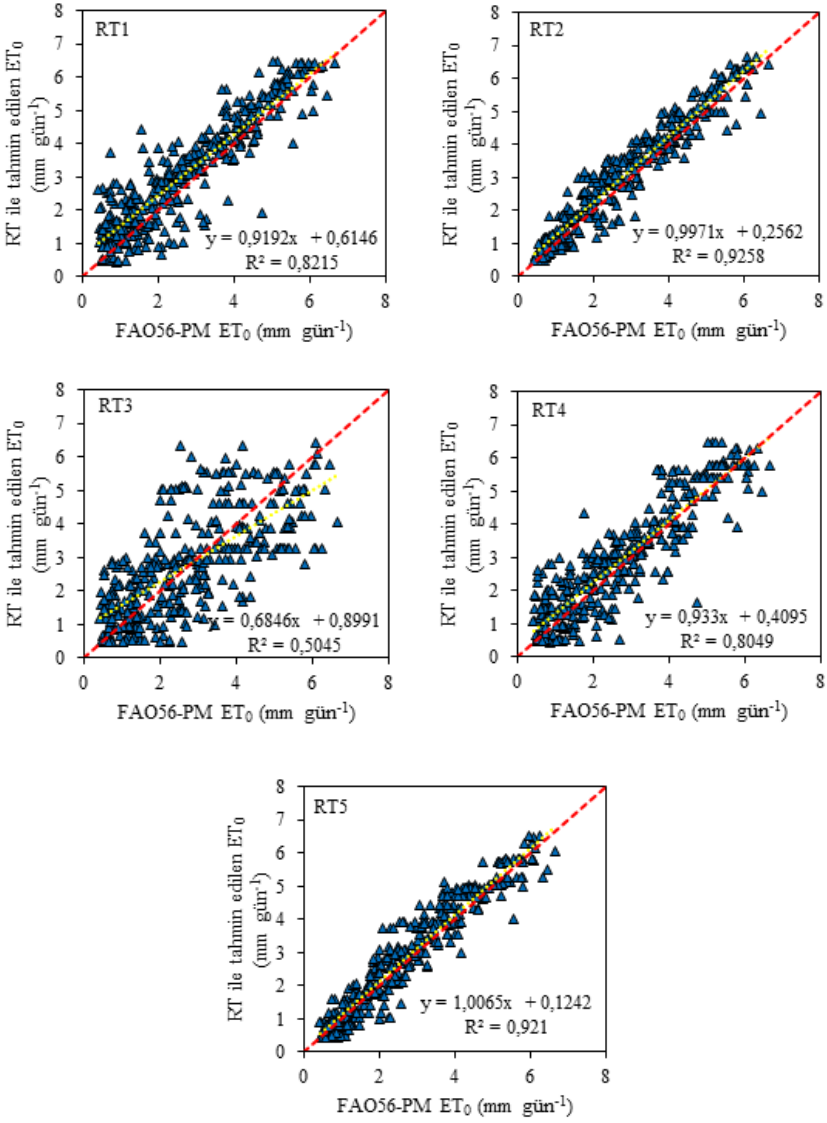
Not: Tüm modeller arasındaki en iyi istatistiksel göstergeler koyu renkle işaretlenmiştir.

Test aşamasında MARS, RT ve GPR modelleri tarafından tahmin edilen ET₀ değerleri ile bunlara karşılık gelen FAO56-PM değerleri arasındaki saçılım grafikleri sırasıyla Şekil 1, Şekil 2 ve Şekil 3'de gösterilmiştir. MARS modeli kullanılarak tahmin edilen ET₀ değerlerinin, FAO56-PM yöntemi ile elde edilen ET₀ değerleriyle iyi bir uyum içinde olduğu saçılım grafiklerinden görülmektedir. Sonuçlar beş farklı girdi kombinasyonu karşılaştırıldığında MARS'ın diğer kombinasyonlardan daha iyi performans gösterdiğini açıkça göstermektedir. Genel olarak tüm modellerin performansları karşılaştırıldığında MARS modellerinin en iyi performansı gösterdiği, RT ve GPR modellerinin ise MARS modeli ile yakın tahmin performansı ürettiği görülmektedir.



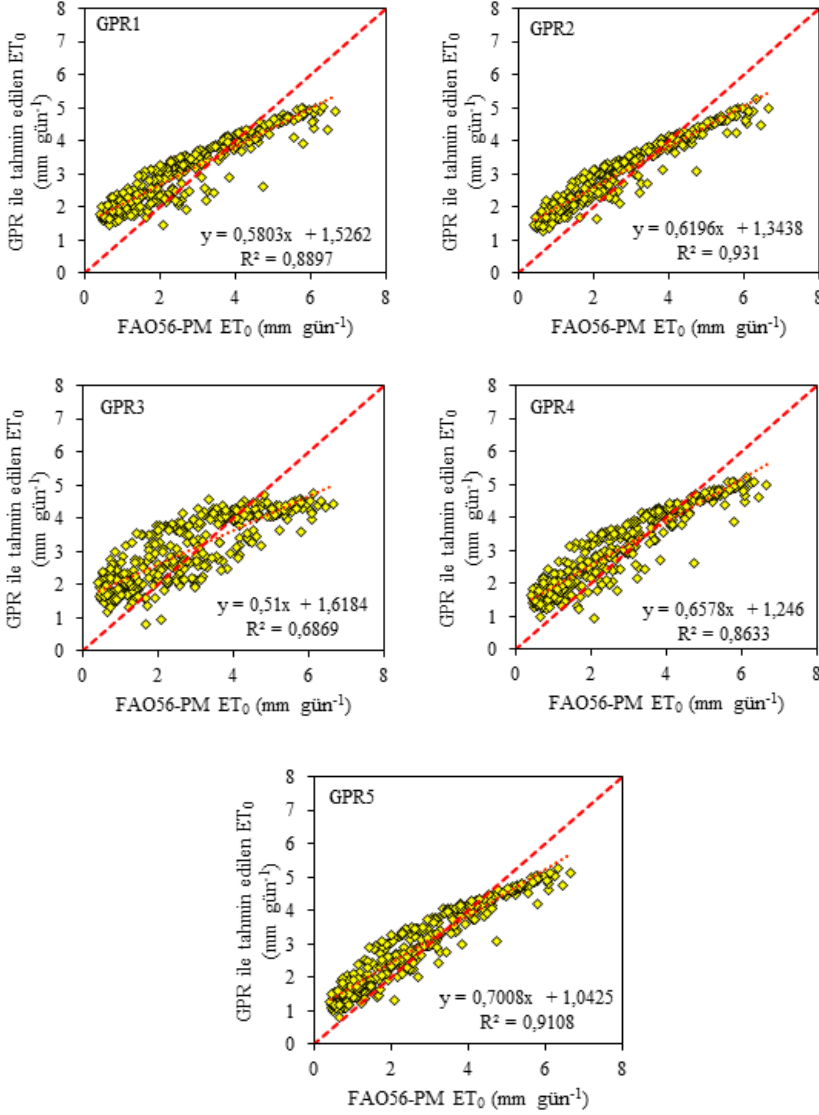
Şekil 1. Test aşamasında beş girdi kombinasyonu için FAO56-PM değerlerine karşı MARS yöntemi ile tahmin edilen ET₀ değerlerinin saçılım grafikleri

Figure 1. Scatter plots of ET₀ values estimated with the MARS method versus FAO56-PM values for five input combinations in the test phase



Şekil 2. Test aşamasında beş girdi kombinasyonu için FAO56-PM değerlerine karşı RT yöntemi ile tahmin edilen ET_0 değerlerinin saçılım grafikleri

Figure 2. Scatter plots of ET_0 values estimated with the RT method versus FAO56-PM values for five input combinations in the test phase



Şekil 3. Test aşamasında beş girdi kombinasyonu için FAO56-PM değerlerine karşı GPR yöntemi ile tahmin edilen ET₀ değerlerinin saçılım grafikleri

Figure 3. Scatter plots of ET₀ values estimated with the GPR method versus FAO56-PM values for five input combinations in the test phase

4. TARTIŞMA

Bu çalışmada genel olarak, sıcaklık ve güneşlenme şiddetinin kullanıldığı radyasyona dayalı makine öğrenimi modelleri, sadece sıcaklık verilerine dayalı makine öğrenimi modellerinden daha iyi performans göstermiştir. Bunun nedeni güneşlenme şiddetinin (R_s), bitki su tüketimi için birincil enerji kaynağı olmasından kaynaklanabilir. Bu sonuç, Fan ve ark. (2018) ve Zhu ve ark. (2020)'nin bulgularıyla uyumludur. Minimum ve maksimum hava sıcaklıklarının girdi olarak kullanıldığı kombinasyon 4 ile bağıl nemin girdi olarak kullanıldığı kombinasyon 2 karşılaştırıldığında bağıl nemin model performansına katkısı daha fazla olmuştur. Bununla birlikte, sıcaklığa dayalı modellerle karşılaştırıldığında, bağıl nem ve rüzgar hızı gibi daha fazla iklimsel verinin girdi olarak kullanılmasıyla modellerin doğruluğu önemli oranda artmıştır. Bu sonuç, gerçek zamanlı sulama tahminleri için halen kullanılmakta olan sadece sıcaklığa dayalı referans bitki su tüketimi tahmin modellerinin dikkatli kullanılması gerektiğini doğrulamaktadır. Yapılan önceki çalışmalarda, meteorolojik girdi değişkenlerinin hepsinin kullanıldığı modellerin daha doğru tahmin sonuçları verdiği belirlenmiştir. Fan ve ark. (2018) girdi olarak kullanılan farklı meteorolojik değişkenlerin kombinasyonlarının, makine öğrenimi modelleriyle ET_0 'ın tahmin doğruluğunu belirlemede birincil rol oynadığını bildirmiştir. Shan et al. (2020) Güneybatı Çin'in yarı kurak bölgelerinde günlük ET 'yi tahmin etmek için MARS, YSA ve ampirik modelleri karşılaştırmıştır. Genel olarak, MARS modellerinin tüm büyüme aşamalarında diğer modellerden daha iyi performans gösterdiği bildirilmiştir.

Referans bitki su tüketimi, çok sayıda meteorolojik faktöre bağlı karmaşık, dinamik ve doğrusal olmayan bir hidrolojik olgudur. Bu çalışmadan elde edilen bulgular, MARS modellerinin ET_0 ile meteorolojik değişkenler arasındaki karmaşık ve doğrusal olmayan ilişkileri başarılı bir şekilde modelleyebildiğini göstermiştir. Bununla birlikte RT ve GPR modelleri, ET_0 'ı tahmin etmede MARS ile karşılaştırılabilir iyi bir performansa sahip olmuştur. Bu çalışmada oluşturulan MARS modelleri, eksik meteorolojik veri olması durumunda bile yarı nemli iklim özelliklerine sahip bölgelerde referans bitki su tüketimini tahmin etmek için önerilebilir. Tüm bunların sonucunda Türkiye'nin yarı nemli iklim özelliğine sahip Bafra Ovası'nda su kaynaklarının planlanmasında ve sulama suyu miktarının belirlenmesinde MARS modeli kullanılarak ET_0 'ın tahmin edilmesinin daha doğru olacağı sonucuna varılabilir.

5. SONUÇ

Referans bitki su tüketiminin doğru, hızlı ve güvenilir tahmini, bölgesel su yönetimi ve tarımsal planlama başta olmak üzere sulama uygulamalarının başarılı olabilmesi için önemli bir rol oynamaktadır. Bu çalışmada Bafra Ovası için günlük ET_0 'ı tahmin etmek amacıyla farklı girdi kombinasyonlarında MARS, RT ve GPR makine öğrenmesi modelleri değerlendirilmiştir. FAO-56 PM denklemi, modellerden elde edilen ET_0 tahminlerini karşılaştırmak için referans olarak kullanılmıştır. Sonuçlar, tam veya eksik girdi değişkenlerinin modellerin performansını önemli ölçüde etkilediğini göstermiştir. Genel olarak, radyasyona dayalı makine öğrenimi modelleri, sıcaklığa dayalı makine öğrenimi modellerinden daha iyi performans göstermiştir. Güneşlenme şiddeti (R_s), rüzgar hızı (u_2) ve bağıl nem (RH), ET_0 tahmininde sıcaklık değişkenlerine oranla tahmin sonucunu daha çok etkilemiştir. Farklı performans indekslerine göre her üç modelin de tahmin performansı oldukça benzer ve yakın olmasına rağmen MARS modelinin RT ve GPR modellerinden daha iyi olduğu belirlenmiştir. En doğru sonuçlar, diğer iki modele (RT ve GPR) kıyasla daha iyi tahmin performansı gösteren MARS ile ve altı girdi değişkeninin hepsinin kullanıldığı beşinci girdi kombinasyonu ile elde edilmiştir (Test veri seti için $R^2=0.982$, $MAE=0.250$ mm gün⁻¹, $RMSE=0.305$ mm gün⁻¹, $NSE=0.965$). Bu nedenle çalışma alanı gibi yarı nemli iklim özelliğine sahip alanlarda günlük ET_0 'ı tahmin etmek için MARS modeli tavsiye edilmektedir. Bununla birlikte yapılan çalışmaların çoğu sınırlı iklim verisiyle ET_0 'ı tahmin etmeye odaklanmıştır. Oysa ki ET_0 'ın tahmin doğruluğunun geliştirilmesi de önemlidir. Bu amaçla gelecekteki çalışmalarda modellerin optimizasyon parametrelerinin belirlenmesinde biyolojik ilhamlı sezgisel algoritmalar araştırılabilir.

Çıkar Çatışması

Yazarlar herhangi bir çıkar çatışması olmadığını beyan eder.

Etik

Bu çalışma etik kurul onayı gerektirmez.

Yazar Katkı Oranları

Çalışmanın Tasarlanması (Design of Study): ST(%50), MT(%50)

Veri Toplanması (Data Acquisition): ST(%20), MT(%80)

Veri Analizi (Data Analysis): ST(%80), MT(%20)

Makalenin Yazımı (Writing Up): ST(%70), MT(%30)

Makalenin Gönderimi ve Revizyonu (Submission and Revision): ST(%70), MT(%30)

KAYNAKLAR

- Adnan, R.M., Mostafa, R.R., Islam, A.R.M.T., Kisi, O., Kuriqi, A., Heddam, S., 2021. Estimating reference evapotranspiration using hybrid adaptive fuzzy inferencing coupled with heuristic algorithms. *Computers and Electronics in Agriculture*, 191, 106541. doi:10.1016/j.compag.2021.106541
- Aghelpour, P., Norooz-Valashedi, R., 2022. Predicting daily reference evapotranspiration rates in a humid region, comparison of seven various data-based predictor models. *Stochastic Environmental Research and Risk Assessment*, 1-23. doi:10.1007/s00477-022-02249-4
- Allen, R.G., Pereira, L.S., Raes, D., Smith, M., 1998. *Crop evapotranspiration-Guidelines for computing crop water requirements-FAO Irrigation and drainage paper 56*, FAO - Food and Agriculture Organisation of the United Nations, Rome, 300(9), D05109.
- ASCE-EWRI, 2005. The ASCE standardized reference evapotranspiration equation. Standardization of reference evapotranspiration task committee final report, R. G. Allen, I. A. Walter, R. L. Elliott, T. A. Howell, D. Itenfisu, M. E. Jensen, and R. L. Snyder, eds., ASCE, Reston, VA, (<http://www.asce.org/Product.aspx?id=2147485918>), 216.
- Blaney, H.F., Criddle, W.D., 1950. Determining Water Requirements in Irrigated Areas from Climatological Irrigation Data. Technical Paper No. 96, US Department of Agriculture, Soil Conservation Service, Washington, D.C., 48 pp.
- Cheng, M.Y., Cao, M.T., 2014. Accurately predicting building energy performance using evolutionary multivariate adaptive regression splines. *Applied Soft Computing*, 22, 178-188. doi: 10.1016/j.asoc.2014.05.015
- Djaman, K., Tabari, H., Balde, A. B., Diop, L., Futakuchi, K., Irmak, S., 2016. Analyses, calibration and validation of evapotranspiration models to predict grass-reference evapotranspiration in the Senegal river delta. *Journal of Hydrology: Regional Studies*, 8, 82-94. doi:10.1016/j.ejrh.2016.06.003
- Dong, J., Zhu, Y., Jia, X., Han, X., Qiao, J., Bai, C., Tang, X., 2022. Nation-scale reference evapotranspiration estimation by using deep learning and classical machine learning models in China. *Journal of Hydrology*, 604, 127207. doi: 10.1016/j.jhydrol.2021.127207
- Fan, J., Yue, W., Wu, L., Zhang, F., Cai, H., Wang, X., Lu, X., Xiang, Y., 2018. Evaluation of SVM, ELM and four tree-based ensemble models for predicting daily reference evapotranspiration using limited meteorological data in different climates of China. *Agricultural and Forest Meteorology*, 263, 225-241. doi:10.1016/j.agrformet.2018.08.019
- Gocić, M., Arab Amiri, M., 2021. Reference evapotranspiration prediction using neural networks and optimum time lags. *Water Resources Management*, 35(6), 1913-1926. doi:10.1007/s11269-021-02820-8
- Hargreaves, G.H., Samani, Z.A., 1985. Reference crop evapotranspiration from temperature. *Applied engineering in agriculture*, 1(2), 96-99. doi:10.13031/2013.26773
- Huang, G., Wu, L., Ma, X., Zhang, W., Fan, J., Yu, X., Zeng, W., Zhou, H., 2019. Evaluation of CatBoost method for prediction of reference evapotranspiration in humid regions. *Journal of Hydrology*, 574, 1029-1041. doi:10.1016/j.jhydrol.2019.04.085
- Irmak, S., Irmak, A., Allen, R. G., Jones, J. W., 2003. Solar and net radiation-based equations to estimate reference evapotranspiration in humid climates. *Journal of irrigation and drainage engineering*, 129(5), 336-347. doi:10.1061/(ASCE)10733-9437
- Jing, W., Yaseen, Z.M., Shahid, S., Saggi, M.K., Tao, H., Kisi, O., Salih, S.Q., Al-Ansari, N., Chau, K.W., 2019. Implementation of evolutionary computing models for reference evapotranspiration modeling: short review, assessment and possible future research directions. *Engineering applications of computational fluid mechanics*, 13(1), 811-823. doi:10.1080/19942060.2019.1645045
- Karbasi, M., 2018. Forecasting of multi-step ahead reference evapotranspiration using wavelet-Gaussian process regression model. *Water resources management*, 32(3), 1035-1052.
- Liu, X., Xu, C., Zhong, X., Li, Y., Yuan, X., Cao, J., 2017. Comparison of 16 models for reference crop evapotranspiration against weighing lysimeter measurement. *Agricultural Water Management*, 184, 145-155. <https://doi.org/10.1016/j.agwat.2017.01.017>
- Makkink, G.F., 1957. Testing the Penman formula by means of lysimeters. *J. Inst. Water Eng.*, 11(3): 277-288
- Makwana, J.J., Tiwari, M.K., Deora, B.S., 2023. Development and comparison of artificial intelligence models for estimating daily reference evapotranspiration from limited input variables. *Smart Agricultural Technology*, 3, 100115. doi:10.1016/j.atech.2022.100115
- Mehdizadeh, S., Behmanesh, J., Khalili, K., 2017. Using MARS, SVM, GEP and empirical equations for estimation of monthly mean reference evapotranspiration. *Computers and electronics in agriculture*, 139, 103-114. doi:10.1016/j.compag.2017.05.002
- Mishra, A.K., Ratha, B.K., 2016. Study of random tree and random forest data mining algorithms for microarray data analysis. *International Journal on Advanced Electrical and Computer Engineering*, 3(4), 5-7.

- Nash, J.E., Sutcliffe, J.V., 1970. "River Flow Forecasting through Conceptual Models: Part I. A Discussion of Principles." *Journal of Hydrology* 10 (3): 282-90. doi:10.1016/0022-1694(70)90255-6
- Priestley, C.H.B., Taylor, R.J., 1972. On the assessment of surface heat flux and evaporation using large-scale parameters. *Monthly Weather Review*, 100(2): 81-92. doi:10.1175/1520-0493
- Rasmussen, C.E., Williams, C.K.I., 2006. *Gaussian Processes for Machine Learning*, in: Adaptive Computation and Machine Learning series, MIT Press, Cambridge, Massachusetts, ISBN 0-262-18253-X, 2005.
- Roy, D.K., Saha, K.K., Kamruzzaman, M., Biswas, S.K., Hossain, M.A., 2021. Hierarchical fuzzy systems integrated with particle swarm optimization for daily reference evapotranspiration prediction: A novel approach. *Water Resources Management*, 35(15), 5383-5407. doi:10.1007/s11269-021-03009-9
- Salam, R., Islam, A.R.M.T., 2020. Potential of RT, Bagging and RS ensemble learning algorithms for reference evapotranspiration prediction using climatic data-limited humid region in Bangladesh. *Journal of Hydrology*, 590, 125241. doi:10.1016/j.jhydrol.2020.125241
- Shan, X., Cui, N., Cai, H., Hu, X., Zhao, L., 2020. Estimation of summer maize evapotranspiration using MARS model in the semi-arid region of northwest China. *Computers and Electronics in Agriculture*, 174, 105495. doi:10.1016/j.compag.2020.105495
- Taşan, S., 2018. Bafra ovası sağ sahil topraklarının sulama açısından bazı fiziksel ve kimyasal özelliklerindeki değişimin modeller ile tahmini. Doktora Tezi. Ondokuz Mayıs Üniversitesi Fen Bilimleri Enstitüsü, 346s, Samsun.
- Thorntwaite, C.W., 1948. An approach toward a rational classification of climate. *Geographical review*, 38(1), 55-94.
- Wu, M., Feng, Q., Wen, X., Deo, R. C., Yin, Z., Yang, L., Sheng, D., 2020. Random forest predictive model development with uncertainty analysis capability for the estimation of evapotranspiration in an arid oasis region. *Hydrology Research*, 51(4), 648-665. <https://doi.org/10.2166/nh.2020.012>
- Yin, L., Tao, F., Chen, Y., Liu, F., Hu, J., 2021. Improving terrestrial evapotranspiration estimation across China during 2000–2018 with machine learning methods. *Journal of Hydrology*, 600, 126538.
- Zhou, Z. H., 2016. Learnware: on the future of machine learning. *Frontiers Comput. Sci.*, 10(4), 589-590.
- Zhu, B., Feng, Y., Gong, D., Jiang, S., Zhao, L., Cui, N., 2020. Hybrid particle swarm optimization with extreme learning machine for daily reference evapotranspiration prediction from limited climatic data. *Computers and Electronics in Agriculture*, 173, 105430. doi:10.1016/j.compag.2020.105430