

Araştırma Makalesi

Makine Öğrenmesi Algoritmaları ile Yaz Sezonu Ortalama Akım Değerlerinin Tahmini

*¹Erdem ÇOBAN

¹Haliç Üniversitesi, Mimarlık Fakültesi, Mimarlık Bölümü, İstanbul, Türkiye, erdemcoban@halic.edu.tr, ORCID ID: <http://orcid.org/0000-0002-4526-7273>

Geliş: 07.06.2024;

Kabul: 07.07.2024

Öz

Akarsu akım verilerinin tahmini su bilimi açısından kritik konuların başında gelmektedir. Özellikle yaz aylarında yağışların azalmasına ek olarak su kullanımının artması her yıl iklimsel değişikliklerin etkisinin arttığı dünyamızda tüm canlıları kuraklık riskiyle yüz yüze getirmektedir. Bu yüzden yaz aylarındaki suyun kullanımının önceden planlanması ve bu konuya daha hassas bir şekilde yaklaşılması her geçen gün daha da zorunlu hale gelmektedir. Bu planlamanın yapılmasında ise akarsu akım debilerinin tahmini, su ihtiyacının karşılanması açısından önemlidir. Bu çalışmada Beyşehir Gölünü besleyen üç akarsu üzerinde bulunan, akarsu gözlem istasyonlarından temin edilen veriler ile makine öğrenmesi modelleri kurulmuştur. Rastgele Orman (RO) ve Adaptive Yükseltme (AdaBoost) algoritmalarının kullanıldığı bu modeller ile üç girdi ve bir çıktı olacak şekilde; sonbahar, kış, ilkbahar mevsimsel ortalama akış değerlerinden yaz mevsimi ortalama akışı tahmin edilmeye çalışılmıştır. RO algoritması test ve tahmin arasındaki belirleme katsayısı (R^2) 0.9368 değerindedir. Kök ortalama kare hatası (RMSE) değeri ise 0.0275 olarak bulunmuştur. AdaBoost algoritması ise RO algoritmasına göre daha güçlü tahminde bulunarak test ve tahmin arasındaki R^2 değeri 0.981, RMSE değeri ise 0.05 olarak bulunmuştur.

Anahtar kelimeler: Makine Öğrenmesi, Tahmin, Akarsu Akım

*¹Sorumlu yazar

Bu makaleye atıf yapmak için

Çoban, E. (2024). Makine Öğrenmesi Algoritmaları ile Yaz Sezonu Ortalama Akım Değerlerinin Tahmini. *Journal of Innovations in Civil Engineering and Technology (JICIVILTECH)*, 6(2), 73-81. <https://doi.org/10.60093/jiciviltech.1497771>

Estimation of Summer Season Average Flow Values with Machine Learning Algorithms

Abstract

Estimation of stream flow data is one of the critical issues in hydrology. In addition to the decrease in precipitation, especially in the summer months, the increase in water use puts all living things in our world at risk of drought, where the impact of climatic changes increases every year. Therefore, it is becoming more and more necessary to plan the use of water in the summer months in advance and approach this issue more sensitively. In making this planning, estimation of stream flow rates is important in terms of meeting water needs. In this study, machine learning models were established with data obtained from stream observation stations on three streams feeding Beyşehir Lake. With these models using Random Forest (RF) and Adaptive Boosting (AdaBoost) algorithms, there are three inputs and one output; An attempt was made to estimate the summer average flow from the autumn, winter and spring seasonal average flow values. The coefficient of determination (R^2) between RF algorithm test and prediction is 0.9368. The root mean square error (RMSE) value was found to be 0.0275. The AdaBoost algorithm made a stronger prediction than the RF algorithm, and the R^2 value between test and prediction was found to be 0.981 and the RMSE value was 0.05.

Keywords: Machine Learning, Prediction, Stream Flow

1. Giriş

Küresel iklim değişikliğinin yaşamsal alandaki etkileri her geçen gün artmaktadır (Ziervogel vd., 2014). Bu etkilerin belki de en büyük karşılığı hidrolik alanında görülmektedir. Yağışların eğilimlerinin azalması, su çevriminin bozulması, akarsu akımlarının düşmesi, içilebilir temiz su kaynaklarına ulaşımın giderek zorlaşması bu etkilere verilebilecek örneklerden sadece birkaç tanesidir. İklimsel değişikliğe ilave olarak ortalama insan yaşam süresinin ve nüfusun artması, temizlik kültürünün yaygınlaşması, sanayii kaynaklı su ihtiyacının da artması gibi sebeplerden kaynaklı su ihtiyacı da her geçen gün artmaktadır (Wang vd., 2016). Özellikle yağışların azaldığı hatta durduğu, bununla birlikte su ihtiyacının arttığı yaz döneminde su ihtiyacının karşılanabilir olup olmadığı sorusunun cevaplanması ve buna bağlı olarak su planlamasının yapılması adına tahminlerde bulunmak adına makine öğrenmesi yöntemlerinin güçlü tahmin becerilerinden yararlanılmaktadır. Buna göre yaz dönemine ait debi değerlerinin, o yıla ait diğer mevsimlerdeki debi değerleri üzerinden tahmin edilmesi, su yapılarının yapılması, taşkın koruma, hidroelektrik santral tesislerinin planlanması gibi su kaynakları yönetimi konusunda önemli bir yere sahiptir (Yang vd., 2023).

Hidroloji alanında tahmin algoritmaları bilişim teknolojisinin gelişmesiyle birlikte son yıllarda oldukça artan eğilimdedir (Güçlü vd., 2014; Şişman ve Kizilöz, 2020). Saphoğlu (2023) aylık

akım tahmininde bulunmak adına K-En Yakın Komşu (KNN), Yapay Sinir Ağı (ANN) ve Uyarlanabilir Nöro Bulanık Çıkarım Sistemi (ANFIS), olarak üç farklı yapay zekâ modeli kullanmışlardır. Türkiye'nin batısındaki Ege bölgesinde yer alan Gediz Nehri Havzasının verilerini kullanmışlardır. Özel ve Büyükyıldız (2019) yapay Sinir Ağları (YSA), Destek Vektör Regresyonu (DVR), Adaptif Ağ Tabanlı Bulanık Çıkarım Sistemi (ANFIS) metotları kullanarak Konya Kapalı Havzası'nda yer alan Karaman istasyonuna ait meteorolojik verileri kullanarak farklı giriş kombinasyonları ile aylık buharlaşma miktarı tahmin etmişlerdir. En başarılı sonucu veren algoritmanın DVR olduğu sonucuna varmışlardır. Çubukçu vd. (2022) Mersin Lamas Nehri aylık ortalama akım verileri kullanılarak Yapay Sinir Ağları (YSA), Destek Vektör Makineleri (DVM) ve Derin Öğrenme (DÖ) optimizasyonları kullanılarak akım modellemesi yapmışlar ve sonuç olarak en iyi yöntemin DÖ olduğu YSA ve DVM yöntemlerinin ise buna göre daha başarısız sonuçlar verdiği gözlemlemişlerdir. Rasouli vd. (2012) günlük akım tahmininde bulunmak adına Bayes sinir ağı (BNN), destek vektör regresyonu (SVR) ve Gauss süreci (GP) kullanmışlar ve çoklu doğrusal regresyon (MLR) ile bu yöntemleri karşılaştırmışlardır. Cheng vd. (2020) makine öğrenimi yöntemlerini kullanarak günlük ve aylık akım tahmininde bulunmuşlardır. Yapay sinir ağı (ANN) ve uzun kısa süreli hafıza (LSTM) yöntemini kullanmışlardır. Lu vd. (2023) derin öğrenme ağlarına dayalı

mevsimsel yağış tahmininde bulunmuşlardır.

Bu çalışmada; Beyşehir Gölünü besleyen üç akarsuya ait akım değerleri kullanılarak sonbahar, kış ve ilkbahar mevsim ortalama akım değerlerinden oluşan veri kümesi, makine öğrenmesi algoritmalarından olan Random Forest (RF) ve AdaBoost algoritmaları ile yazlık mevsimsel ortalama akım değerleri tahmin edilmeye çalışılmıştır. Yaz dönemi ortalama akışın güçlü tahmini hem su yönetimi hem de gölün beslenmesinin öngörülebilirliği açısından önem teşkil etmektedir. Ayrıca literatür incelendiğinde benzer çalışmaların yapılmadığı görülmüş ve makine öğrenmesi algoritmasının bu tarz tahmin algoritmalarındaki başarılarının hidroloji alanında daha fazla kullanılması hedeflenmiştir. Farklı parametreler ile birçok tahin algoritması olmasına rağmen mevsimsel ortalama değerlerin girdi parametresi olarak kullanıldığı alternatif bir çalışma olması hedeflenmiştir.

2. Materyal ve Yöntem

Makine öğrenmesi algoritmaları veriden öğrenip; sınıflandırma, tahmin etme gibi becerilere sahip yapay zekânın bir alt kümesidir. Bu çalışmada bu algoritmalarından iki tanesi rastgele orman (Random Forest) ve AdaBoost kullanılmıştır.

2.1 Random Forest (Rastgele Orman)

Rastgele orman algoritması Leo Breiman tarafından geliştirilen topluluk makine öğrenmesi algoritmasıdır (Breiman,

2001). Topluluk sınıflandırma algoritmaları tek bir sınıflandırma yerine, birden fazla sınıflandırma kullanılmaktadır. Rastgele orman algoritması yapısında birden fazla karar ağacı bulunmaktadır ve bu algoritma bu karar ağaçlarından aldığı sonuçların ortalamalarını alarak sonuç vermektedir (Bayrakçı, 2022). Eğitilmiş bir k-sayı ağacı denklem 1'de tanımlanan rastgele orman modelinde toplanır.

$$H(X, \theta_j) = \sum_{i=0}^k(x, \theta_j) \quad (j = 1, 2, 3, \dots, m) \quad (1)$$

Denklem (1)'de $H(X, \theta_j)$ ifadesi bir meta karar ağacı sınıflandırma işlevini görmektedir. x , eğitim veri kümesinin giriş özellik vektörünü temsil ederken, θ_j ise ağacın büyüme sürecini belirleyen bağımsız ve düzgün dağıtılmış rastgele bir vektördür.

2.2 Adaptive Yükseltme (AdaBoost)

Adaptive Boost (AdaBoost) algoritması örnek özellik uzayını yinelemeli olarak arar ve eğitim özelliği ağırlıklarını bulur. Yinelemeli süreçte eğitim örneklerinin özellik ağırlıkları sürekli olarak ayarlanır (Freund, 1997). Bu algoritmanın en temel özelliği zayıf modellerin iyi taraflarını alarak en başarılı modeli oluşturmaktır. AdaBoost'ta birden fazla model oluşturularak her modelin güçlü yönleri birleştirilir. Algoritma daha sonra bu modellerden başarılı bir model elde etmeye çalışır. AdaBoost algoritması zayıf modelleri birleştirerek en başarılı modelin çıktısını üretir.

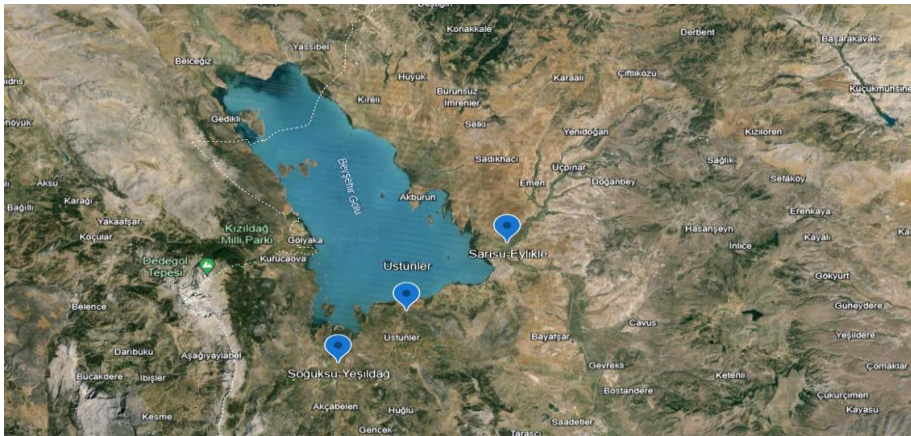
2.1 Çalışma Alanı

Bu çalışmada, Konya Havzasındaki Beyşehir İlçe sınırlarında yer alan ve Beyşehir Gölünü besleyen akarsular üzerinde bulunan D16A104 Sarısu – Eylikle, D16A111 Soğuksu - Yesildağ Köprüsü, D16A015 Üstünler akım gözlem istasyonlarından alınan, 1974 ile 2010 yılları arasındaki günlük akım verileri kullanılmıştır. İstasyonlara ait enlem, boylam ve rakım bilgileri Tablo 1’de gösterilmiştir. Ayrıca istasyonların harita üzerindeki konumları da Şekil 1’de gösterilmiştir. Eksik verinin çok olmasından kaynaklı bu üç istasyondaki veriler öncelikle 12 aylık veriler olarak düzeltilmiştir. Ancak makine öğrenmesi algoritmalarının daha verimli ve hızlı çalışması adına bu aylık veriler mevsimsel ortalama olarak 4 parametreye indirgenmiştir. Bu şekilde algoritma mimarisi 3 girdi ve bir

çıktıdan oluşan bir hale getirilmiştir. Girdi parametreleri olarak sonbahar, kış ve ilkbahar ortalama akım değerleri, tahmin edilecek olan çıktı parametresi olarak da yaz mevsimi ortalama akım değerleri kullanılmıştır. Veriler, her mevsimdeki akım verilerinin yelpazesinin geniş olması dezavantajını azaltmak adına normalize edilerek 0 ile 1 aralığına getirilmiş ve veriler boyutsuzlaştırılmıştır. Her bir parametrede 70, toplamda 280 adet verinin kullanıldığı bu çalışmada, eğitim için verilerin yüzde 80’i, test için ise yüzde 20’si kullanılmıştır. Elde edilen sonuçlar hem eğitimin ölçülmüş çıktılarıyla hem de test için ayrılmış verilerin çıktı değerleri ile kıyaslanmıştır. Bu kıyaslama da saçılma diyagramlarıyla görselleştirilerek grafik üzerinde korelasyon katsayı değerleriyle birlikte gösterilmiştir.

Tablo 1. İstasyonlara ait coğrafi bilgiler.

İstasyon	Enlem	Boylam	Rakım (m)
Sarısu – Eylikle	37°42'43"N	31°44'13"E	1140
Soğuksu-Yesildağ	37°32'31"N	31°29'32"E	1124
Üstünler	37°36'57"N	31°35'29"E	1126



Şekil 1. İstasyonların harita üzerindeki konumları.

2.3. Doğrulama Modelleri

Regresyon modelinin performansının değerlendirilmesi için iki adet ölçüt kullanılmıştır.

2.3.1. Belirleme katsayısı (R^2)

Belirleme katsayısı, istatistikte, R^2 bir modelin doğrusal regresyon ortamında bir sonucu tahmin etme veya açıklama yeteneğini değerlendiren bir ölçüdür. Daha spesifik olarak R^2 , bağımlı değişkendeki (Y) doğrusal regresyon ve yordayıcı değişken (X, bağımsız değişken olarak da bilinir) tarafından tahmin edilen veya açıklanan varyansın oranını gösterir. R^2 ne kadar yüksek bir değerde olursa model verileri o kadar iyi tanımlar (Di Bucchianico, 2008). Formül aşağıdaki gibi verilmiştir

$$R^2 = 1 - \frac{\sum_{j=i}^N (y_j - \hat{y}_j)^2}{\sum_{j=i}^N (y_j - \bar{y}_j)^2} \quad (2)$$

Denklem (2)'de; y_j gerçeğe dayalı değerleri, \hat{y}_j tahmine dayalı değerleri ve \bar{y}_j veri kümesinin ortalama değerini temsil etmektedir. N ise örnek boyutu (değer sayısı) temsil etmektedir (Güler ve Kandemir, 2022)

2.3.2. Kök Ortalama Kare Hatası

Kök ortalama kare hatası (RMSE), bir regresyon modelinin iki ana performans göstergesinden biridir. Bir model tarafından tahmin edilen değerler ile gerçek değerler arasındaki ortalama farkı ölçer. Modelin hedef değeri

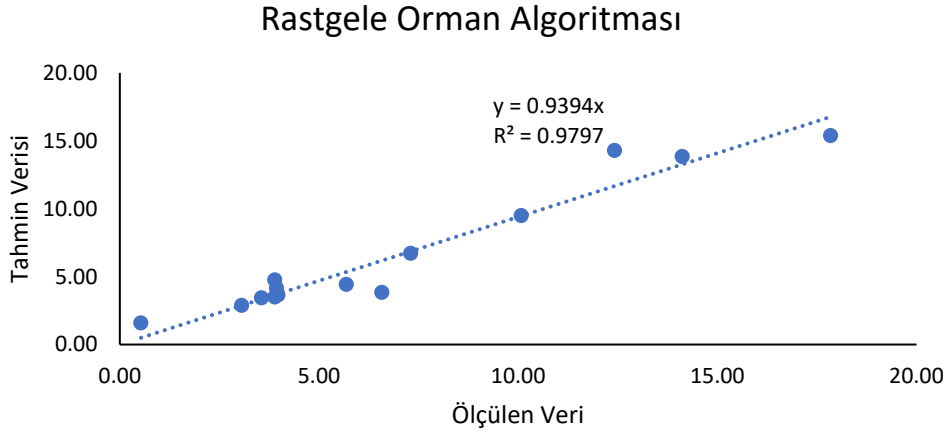
(doğruluk) ne kadar iyi tahmin edebildiğine dair bir tahmin sağlar. Denklemi aşağıda verilmiştir.

$$RMSE = \sqrt{\frac{1}{W} \sum_{i=1}^N w_i u_i^2} \quad (3)$$

Denklem (3)'de; W popülasyonun toplam ağırlığını, N gözlem sayısını, w_i i. gözlemin ağırlığını, u_i ise i. gözlemle ilişkili hatayı temsil etmektedir.

3. Bulgular

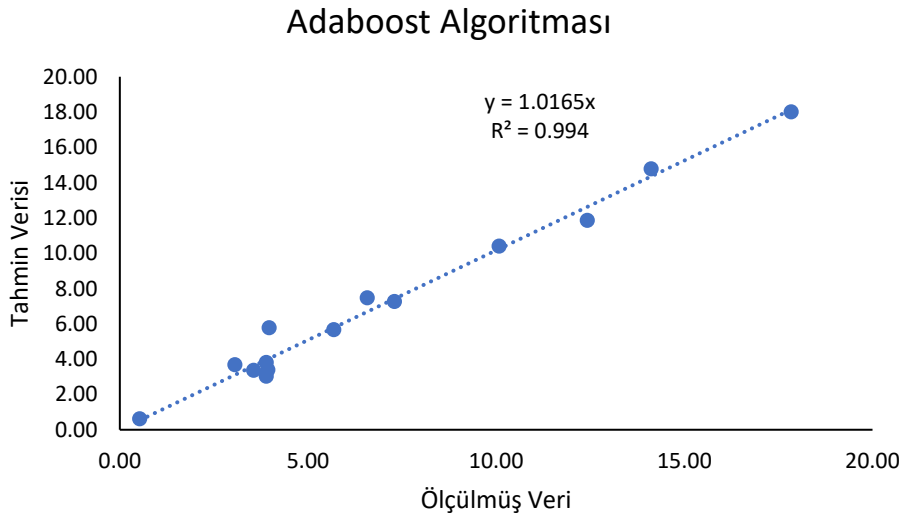
Çalışmada Beyşehir gölünü besleyen üç adet akarsuda bulunan istasyonlardan alınan veriler kullanılmıştır. Tahmin gücü yüksek Random Forest (Rastgele Orman) ve AdaBoost olmak üzere iki güçlü makine öğrenmesi algoritması yardımıyla ilkbahar ortalama akım, kış ortalama akım, sonbahar ortalama akım girdi parametreleri kullanılarak yaz ortalama akım tahmin edilmeye çalışılmıştır. Çalışmada mevsimsel ortalama seçilmesinin sebebi, modelin karmaşıklığını azaltmak ve yaz sezonu için ortalama bir tahminde bulunma çabasıdır. Sonsuz parametreye bağlı olan akım parametresinin sadece diğer mevsimlerdeki ortalama akım değerleri ile tahmini modeli basitleştirmek adına alınmış bir karar olup anlamlı sonuçları da beraberinde getirmiştir. Test değerleri ile tahmin değerleri saçılma diyagramında görselleştirilmiştir. Belirleme sayısı ve kök ortalama kare hatası değerleri ile de tahminin başarısı değerlendirilmiştir.



Şekil 2. Rastgele orman algoritmasına ait saçılma diyagramı.

Şekil 2’de rastgele Orman algoritması ile tahmin edilen ve test amacıyla kullanılan değerler bir grafik üzerinde saçılma diyagramı ile gösterilmeye

çalışılmıştır. Test ve tahmin arasındaki determinasyon katsayısı (R^2) 0.9368 değerindedir. Kök ortalama kare hatası (RMSE) değeri ise 0.0275 olarak hesaplanmıştır.



Şekil 3. Adaboost algoritmasına ait saçılma diyagramı.

Şekil 3’de adaboost algoritması ile tahmin edilen ve test amacıyla kullanılan değerler bir grafik üzerinde

saçılma diyagramı ile gösterilmeye çalışılmıştır. Test ve tahmin arasındaki determinasyon katsayısı (R^2) 0.981 değerindedir. Kök ortalama kare hatası

(RMSE) değeri ise 0.05 olarak bulunmuştur.

4. Sonuç ve Tartışma

Yaz dönemi akarsu akım debisinin tahmini su kaynakları yönetimi açısından önem teşkil etmektedir. Bu çalışmada üç mevsime ait ortalama akarsu akım verileri girdi olarak kullanılarak yaz dönemine ait akım verileri tahmin edilmeye çalışılmıştır. Bu bağlamda makine öğrenmesi algoritmalarından iki tanesi bu çalışmada kullanılmıştır. Bunlardan Rastgele Orman algoritması determinasyon katsayısı (R^2) değeri 0.9368 gelirken, kök ortalama kare hatası (RMSE) değeri 0.0275 olarak hesaplanmıştır. Adaboost algoritmasına bakıldığında determinasyon katsayısı (R^2) değeri 0.981, kök ortalama kare hatası (RMSE) değeri 0.0275 olarak hesaplanmıştır.

Akarsu akım tahmini her geçen gün daha da kritik hale gelmektedir. Burada yapılan çalışmaya benzer farklı bölgelerde ve farklı verilerle alternatif yöntemler kullanılarak tahminlerde bulunmak literatüre katkı sağlayacak, su kaynakları yönetiminde ilerleme sağlanacaktır. Makine öğrenmesi yöntemleri güçlü tahmin yetenekleri bakımından birçok benzer çalışma açısından önemlidir. Özellikle su bilimleri alanında birçok tahmin modelinde diğer algoritmalarla kullanılmalıdır.

Etik Standartlar Bildirgesi

Yazarlar tüm etik standartlara uyduklarını beyan ederler.

Yazarlık Katkı Beyanı

Yazar 1: Kavramsallaştırma, Metodoloji, Doğrulama, Analiz ve yorumlama, Araştırma, Yazma/orijinal taslak, Yazma/inceleme ve düzenleme, Görselleştirme

Çıkar Çatışması Beyanı

Yazarın bu makalenin içeriğiyle ilgili olarak beyan edeceği hiçbir çıkar çatışması yoktur.

Verilerin Kullanılabilirliği

Bu çalışma sırasında oluşturulan veya analiz edilen tüm veriler, yayınlanan bu makaleye dâhil edilmiştir.

Teşekkür

Veri paylaşımı ile çalışmaya katkısından dolayı DSI'ye teşekkür ederim.

5. Kaynaklar

- Bayrakçı, H., Keşkekçi, A. B., & Arslan, R. (2022). Classification of iris flower by random forest algorithm. *Advances in Artificial Intelligence Research*, 2(1), 7-14. <https://doi.org/10.54569/aaair.1018444>
- Breiman, L. (2001). Random forests. *Machine learning*, 45, 5-32. <http://dx.doi.org/10.1023/A:1010933404324>
- Cheng, M., Fang, F., Kinouchi, T., Navon, I. M., & Pain, C. C. (2020). Long lead-time daily and monthly streamflow forecasting using machine learning methods. *Journal of Hydrology*, 590, 125376. <https://doi.org/10.1016/j.jhydrol.2020.125376>
- Çubukçu, E. A., Demir, V., & Sevimli, M. F. (2022). Akım Verilerinin Makine Öğrenmesi Teknikleriyle Tahmin Edilmesi. *Gazi Mühendislik Bilimleri Dergisi*, 8(2), 257-272.
- Di Buccianico, A. (2008). Coefficient of determination (R²). *Encyclopedia of*

- statistics in quality and reliability. https://doi.org/10.1002/9780470061572.eq_r173
- Dirlik, C., Kandemir, H., Çetin, N., Şen, S., Güler, B., & Gürel, A. (2022). Effects of different culture media compositions on in vitro micropropagation from paradox walnut rootstock nodes. *Gazi University Journal of Science Part A: Engineering and Innovation*, 9(4), 500-515. <https://doi.org/10.54287/gujsa.1194822>
- Freund, Y., & Schapire, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of computer and system sciences*, 55(1), 119-139. <https://doi.org/10.1006/jcss.1997.1504>
- Güçlü, Y. S., Yeleşen, M. Ö., Dabanlı, İ., & Şişman, E. (2014). Solar irradiation estimations and comparisons by ANFIS, Angström–Prescott and dependency models. *Solar Energy*, 109, 118-124. <https://doi.org/10.1016/j.solener.2014.08.027>
- Lu, P., Deng, Q., Zhao, S., Wang, Y., & Wang, W. (2023). Deep learning for seasonal prediction of summer precipitation levels in eastern China. *Earth and Space Science*, 10(11), e2023EA003129. <https://doi.org/10.1029/2023EA003129>
- Özel, A., & Büyükyıldız, M. (2019). Usability of artificial intelligence methods for estimation of monthly evaporation. *Niğde Ömer Halisdemir University Journal of Engineering Sciences*, 8(1), 244-254. <https://doi.org/10.28948/ngumuh.516891>
- Rasouli, K., Hsieh, W. W., & Cannon, A. J. (2012). Daily streamflow forecasting by machine learning methods with weather and climate inputs. *Journal of Hydrology*, 414, 284-293. <https://doi.org/10.1016/j.jhydrol.2011.10.039>
- Saplıoğlu, K. (2023). Monthly streamflow prediction using ANN, KNN and ANFIS models: Example of Gediz River Basin. *Teknik Bilimler Dergisi*, 13(2), 42-49. <https://doi.org/10.35354/tbed.1298296>
- Şişman, E., & Kizilöz, B. (2020). Artificial neural network system analysis and Kriging methodology for estimation of non-revenue water ratio. *Water Supply*, 20(5), 1871-1883. <https://doi.org/10.2166/ws.2020.095>
- Wang, X. J., Zhang, J. Y., Shahid, S., Guan, E. H., Wu, Y. X., Gao, J., & He, R. M. (2016). Adaptation to climate change impacts on water demand. *Mitigation and Adaptation Strategies for Global Change*, 21, 81-99. <https://doi.org/10.1007/s11027-014-9571-6>
- Yang, S., Ling, F., Li, Y., & Luo, J. J. (2023). Improving Seasonal Prediction of Summer Precipitation in the Middle–Lower Reaches of the Yangtze River Using a TU-Net Deep Learning Approach. *Artificial Intelligence for the Earth Systems*, 2(2), 220078. <https://doi.org/10.1175/AIES-D-22-0078.1>
- Ziervogel, G., New, M., Archer van Garderen, E., Midgley, G., Taylor, A., Hamann, R., ... & Warburton, M. (2014). Climate change impacts and adaptation in South Africa. *Wiley Interdisciplinary Reviews: Climate Change*, 5(5), 605-620. <https://doi.org/10.1002/wcc.295>