

Atf İçin: Ekmekcioğlu, Ö. (2023). CBS Tabanlı Melez Makine Öğrenmesi Uygulamalarının Ani Sel Duyarlılık Haritalamasında Kullanımı. *İğdır Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, 13(2), 1067-1084.

To Cite: Ekmekcioğlu, Ö. (2023). Incorporating the GIS-Based Hybrid Machine Learning Applications into the Flash Flood Susceptibility Mapping. *Journal of the Institute of Science and Technology*, 13(2), 1067-1084.

CBS Tabanlı Melez Makine Öğrenmesi Uygulamalarının Ani Sel Duyarlılık Haritalamasında Kullanımı

Ömer EKMEKÇİOĞLU^{1*}

Öne Çıkanlar:

- Ani sel haritalaması hibrit makine öğrenmesi teknikleri ile gerçekleştirilmiştir.
- Önerilen model coğrafi bilgi sistemleri ile entegre edilmiştir.
- Ani taşkınları etkileyen en önemli parametreler yağış, NDVI ve eğri numarası olarak bulunmuştur

Anahtar Kelimeler:

- Taşkın haritalama
- Rastgele orman
- Parçacık sürü optimizasyonu
- Makine öğrenmesi
- Coğrafi bilgi sistemleri

ÖZET:

Bu çalışmada Kentucky Nehri havzasında son yirmi yılda meydana gelen ani sel baskınları kayıtlarına dayanarak makine öğrenmesi yöntemleri kullanılarak taşkın tehlike haritalamasının yapılması amaçlanmıştır. Tahminlerin gerçekleştirilebilmesi için yaygın olarak kullanılan ve pratik bir algoritma olan rastgele orman (RF) yöntemi kullanılmıştır. Ayrıca, bu yöntemin içsel parametreleri (ağaç sayısı ve maksimum ağaç derinliği) ise parçacık sürü optimizasyonu (PSO) algoritması ile optimize edilmiştir. Bu bağlamda 343 adet geçmiş ani sel kayıtlarına ilaveten havza sınırları içerisinde yer alacak şekilde aynı sayıda rastgele nokta atanmıştır. Tüm bu noktalara 12 adet ani sel tehlikesini tetikleyecek faktörler tanıtılmış olup, tahminler bu doğrultuda gerçekleştirilmiştir. Tahmin sonuçları birçok performans değerlendirme indikatörü göz önüne alınarak analiz edildiğinde melez PSO-RF modelinin test veri setinde oldukça başarılı sonuçlar gösterdiği görülmüştür. Öyle ki hem ani sel olan noktalar hem de ani sel gerçekleşmeyen noktalar %70 oranında doğruluk ile tahmin edilmiştir. Yapılan detaylı değerlendirmeler sonucu ise ikili sınıflandırma problemlerinde önemli bir gösterge olan AUROC değeri ise 0.79 olarak hesaplanmıştır. Ayrıca, ani selleri tetikleyen faktörlerin sonuçlar üzerindeki tekil etkileri incelendiğinde şiddetli yağış faktörü en etkili değişken olarak bulunmuş olup, onu sırasıyla topoğrafya, NDVI ve eğri numarası faktörleri izlemiştir. Öte yandan, litoloji faktörünün ani sellerin modellenmesi üzerindeki etkisi ise diğer faktörlere göre oldukça az olduğu sonucuna varılmıştır. Tüm bu bulgular ışığında elde edilen sonuçlar hem taşkın tehlike haritalaması literatürüne katkı yapacak, hem de ilgili bölgede yaşanacak gelecek ani sel olayları meydana gelmeden alınması gereken tedbirler ile ilgili yol gösterici nitelikte olacaktır.

Incorporating the GIS-Based Hybrid Machine Learning Applications into the Flash Flood Susceptibility Mapping

Highlights:

- The flash flood susceptibility mapping was conducted through a hybrid machine learning framework.
- The proposed model was incorporated into geographical information system.
- Heavy rainfall, NDVI, and curve number was found as the most determining factors in producing flash flood susceptibility maps.

Keywords:

- Flood mapping
- Random forest
- Particle swarm optimization
- Machine learning
- Geographical information system

ABSTRACT:

This study chiefly aimed to perform flash flood susceptibility mapping by means of machine learning methods based on the records attained in the Kentucky River basin over the last two decades. To carry out analysis, one of the widely adopted practical tree-based machine learning tools, i.e., the random forest (RF) method, was utilized, while the hyperparameters (i.e., number of trees and maximum tree depth) of the RF algorithm were tuned via the particle swarm optimization (PSO) strategy. In this vein, a total of 343 flash-flooded and the same number of random (non-flash flooded) points were assigned within the Kentucky River basin boundaries. In addition, a total of 12 factors triggering flash floods have been introduced to the corresponding points and the predictions were conducted in this regard. Many performance evaluation indicators considered within the scope of this study illustrated that the hybrid PSO-RF model revealed quite accurate predictive results based on the blinded testing set; such that both flash-flooded and non-flash flooded points exist in the test set were estimated with an accuracy of 70%. In addition, one of the promising performance indicators in assessing binary classification implementations, called AUROC, was calculated as 0.79. Further analysis regarding the individual impacts of the triggering factors also highlighted that the heavy rainfall probability factor was found to be the most effective variable, followed by topography, NDVI, and curve number, respectively. On the other hand, it was concluded that the effect of the lithology on the flash flood modeling is considerably lower compared to its counterparts. Overall, the results acquired in the light of all these findings have important potential in terms of both contributing to the flood susceptibility mapping literature and guiding with respect to the measures that should be taken prior to the flash flood incidents in the corresponding region.

¹ Ömer EKMEKÇİOĞLU ([Orcid ID: 0000-0002-7144-2338](https://orcid.org/0000-0002-7144-2338)), İstanbul Teknik Üniversitesi, Afet Yönetimi Enstitüsü, Afet ve Acil Durum Yönetimi Anabilim Dalı, İstanbul, Türkiye

*Sorumlu Yazar/Corresponding Author: Ömer EKMEKÇİOĞLU, e-mail: omer.ekmekcioglu@itu.edu.tr

GİRİŞ

Taşkınlar tüm dünyada büyük ekonomik ve can kayıplarına neden olmaktadır. Öyle ki, son yirmi yılda Dünya’da 2 milyar dolardan fazla ekonomik kayıp taşkınlardan dolayı meydana gelmektedir (Ekmekcioğlu ve ark., 2021). Ayrıca, bu taşkınlardan dolayı iki milyondan fazla insan doğrudan ve/veya dolaylı olarak etkilenmiştir. Özellikle kentleşmenin artması ile nüfus yoğunlukları belirli kesimlerde yoğunlaşmış ve arazi kullanım yapısı geçirimsiz yapıya bürünerek şiddetli yağışların tetiklediği ani sel baskınlarının sayısında artışı beraberinde getirmiştir (Thieken ve ark., 2006). Literatürde ve pratikte taşkınlar öncesi önleme stratejilerinin ve taşkınlar sonucu risk azaltma uygulamalarının belirlenmesinde coğrafi bilgi sistemlerine dayalı haritalama tekniklerinin kullanılması son yıllarda oldukça hız kazanmıştır (Gigović ve ark., 2017). Afet olaylarından önce yapılması gereken ilk adım ise potansiyel olarak risk altında bulunan bölgelerin belirlenmesidir. Zira, bu algoritmalar coğrafi bilgi sistemleri ile de entegre hale getirilebilmektedir. Makine öğrenmesi algoritmaları sadece tehlike altındaki bölgelerin belirlenmesinde değil aynı zamanda bu bölgelerde olası tehlikeleri tetikleyen faktörlerin sebep sonuç ilişkileri ışığında belirlenmesine olanak tanımaktadır. Ayrıca bu yöntemler, çok boyutlu problemler ile başa çıkabilme yetenekleri, gerektiğinde problemin boyutunun düşürülmesinde sağladığı avantajlar, yüksek doğruluk oranları, kullanım kolaylıkları ve diğer modellere adapte edilebilme yetenekleri ile başka sel tehlike haritalaması tekniklerinden önemli ölçüde ayrılmaktadır (Ngo ve ark., 2018; Ali ve ark., 2020; Pham ve ark., 2021; Pourghasemi ve ark., 2020). Dolayısıyla, literatürde makine öğrenmesi yöntemlerine dayalı taşkın tehlike haritalarının oluşturulmasına yönelik çalışmalar mevcuttur (Khosravi ve ark., 2018; Costache ve Tien Bui, 2019; Talukdar ve ark., 2020). Bu çalışmada ise ani sel tehlike haritalamalarının gerçekleştirildiği ilgili çalışmalardan güncel olanlarından bazıları Çizelge 1’de sunulmuştur.

Çizelge 1. Makine öğrenmesi metotları yardımıyla ani sel haritalaması gerçekleştirilen çalışmalar

| Kaynak | Bölge | Ülke | Alan (km ²) | Kullanılan Metot | Nokta Sayısı | Eğitim/ Test |
|------------------------------|----------------------|-----------|-------------------------|------------------------|--------------|--------------|
| Khosravi ve ark., (2018) | Haraz Havzası | İran | 4014 | A, B, C, D | 76 | 70/30 |
| Costache ve Tien Bui, (2019) | Putna Nehri Havzası | Romanya | 2509 | E, F, G | 132 | 70/30 |
| Tien Bui ve ark., (2019) | Bac Ha ve Bao Yen | Vietnam | 1510.4 | G, H, I, J | 654 | 70/30 |
| Costache ve ark., (2020) | Bâsca Chiojdului | Romanya | 340 | E, G, J, K, L, M | Poligon | 70/30 |
| Nhu ve ark., (2020) | Van Ban Bölgesi | Vietnam | 1435 | A, J, L, M, N | 200 | 70/30 |
| Talukdar ve ark., (2020) | Teesta Nehri Havzası | Bangladeş | 2284 | B, L, O, P | 413 | 80/20 |
| Tien Bui ve ark., (2020) | Lao Cai İli | Vietnam | 1465.07 | E, J, R | 732 | 70/30 |
| Shahabi ve ark., (2021) | Haraz Havzası | İran | 4014 | A, B, C, D, K, S, T, U | 194 | 80/20 |
| Panahi ve ark., (2021) | Golestan Eyaleti | İran | 12000 | V, Y, Z | 143 | 70/30 |

* **Kısaltmalar:** **A:** Lojistik Model Ağacı (LMT), **B:** Ağaç Budama Hatasında Azalma (REPTree), **C:** Naive Bayes Ağaçları (NBT), **D:** Değişken Karar Ağaçları (ADT), **E:** Çok Katmanlı Algılayıcı (MLP), **F:** Rotasyon Ağaçları (ROF), **G:** Sınıflandırma ve Regresyon Ağaçları (CART), **H:** Çok Değişkenli Uyarlanabilir Regresyon Eğrileri (MARS), **I:** Geri Yayılımlı Sinir Ağları (BPNN), **J:** Destek Vektör Makineleri (SVM), **K:** Lojistik Regresyon (LR), **L:** Rastgele Orman (RF), **M:** Karar Ağaçları (DT), **N:** Rastgele Alt Uzay Karar Ağaçları (RSTree), **O:** M5 Model Ağacı (M5P), **P:** Rastgele Ağaç (RT), **R:** Derin Öğrenme Sinir Ağları (DLNN), **S:** Derin İnanç Ağları (DBN), **T:** Artırılmış Lojistik Regresyon (BLR), **U:** Uyarlanabilir Nöro Bulanık Çıkarım Sistemi (ANFIS), **V:** Uzun Kısa Süreli Bellek Sinir Ağları (LSTM), **Y:** Derin Sinir Ağları (DNN), **Z:** Evrimsel Sinir Ağları (CNN).

Çizelgeden de anlaşılacağı üzere literatürde çok çeşitli makine öğrenmesi algoritmaları bu amaca yönelik olarak kullanılmıştır. Öte yandan, gerek makine öğrenmesi yöntemlerinin içerisinde

barındırdıkları belirsizliklerin giderilmesi gerekse de amaca yönelik uygun model konfigürasyonlarının belirlenebilmesi için farklı optimizasyon yöntemlerinin entegrasyonu gerçekleştirilerek melez modeller inşa edilmektedir.

Literatürde yer alan çalışmalar detaylı incelendiğinde ise makine öğrenmesi algoritmalarına dayalı taşkın haritalama çalışmalarında genel olarak taşkın tehlike haritalarının oluşturulduğu gözlemlenmektedir. Risk kavramının en önemli iki ayağı olan tehlike ve hasar görebilirlik kavramlarından tehlike ayağı daha çok atmosferik ve meteorolojik değişkenlere bağlı olarak belirlenebildiğinden dolayı ve topoğrafya ile topoğrafyaya bağlı olan değişkenlerin önemli oranda taşkın tehlikesinde tetikleyici rol oynadığı düşünüldüğünde veri bulunmasının kolaylığı bu sonucu açıklamaktadır. Öyle ki, hasar görebilirlik fonksiyonu genel olarak arazi kullanım yapısına dayalı detaylı değişkenlere ve insan etkisi ile ortaya çıkan faktörlere (hasar görebilir yapılar, nüfus yoğunlukları, hasar görebilir popülasyonlar, ulaşım ağları, vb.) bağlı olarak belirlenebilmektedir. Bu verilerin ise özellikle havza ölçeğinde gerçekleştirilen grid bazlı analizlerde elde edilmesinin zor olduğundan dolayı araştırmacılar tehlike fonksiyonunun belirlenmesine daha fazla odaklanmıştır. Bu doğrultuda, tehlike haritalarının elde edilmesinde web tabanlı platformlardan da kolayca erişilebilen sayısal yükseklik verileri ile birlikte bu verilerden elde edilebilen eğim, bakı, eğrilik, topoğrafik ıslaklık indeksi ve topoğrafik pürüzlülük indeksi gibi değişkenler yaygın olarak kullanılmaktadır. Örneğin Khosravi ve ark., (2018), lojistik model ağaçları (LMT), azaltılmış hata budama ağaçları (REPT), naif Bayes ağaçları (NBT) ve alternatif karar ağaçları (ADT) olmak üzere dört farklı karar ağacı tabanlı algoritmayı kullanarak İran'ın Haraz havzası sınırları içerisinde yer alan bölgelerde ani sele eğilimli yerleri belirlemiştir. Kapsamlı performans değerlendirmelerine dayanarak, bilgi kazanım oranı (information gain ratio) yaklaşımının da yardımı ile en önemli faktörleri arazi eğimi, nehirden uzaklık ve nehir yoğunluğu olarak elde ederken, ADT' nin en yüksek doğruluğa sahip makine öğrenmesi algoritması olduğu sonucuna varmıştır. Costache ve Tien Bui, (2019) ani sel tehlikesi altında olan bölgeleri uzamsal olarak tahmin etmek için frekans oranı (frequency ratio) ve kanıt ağırlıkları (weights of evidence) gibi iki değişkenli istatistiksel yaklaşımların yanı sıra yapay sinir ağları (ANN), rastgele orman (RF) ve sınıflandırma ve regresyon ağaçları (CART) gibi makine öğrenmesi tekniklerini kullanarak topluluk (ensemble) modelleri oluşturmuştur. Elde edilen sonuçlara göre, RF ve kanıt ağırlıkları entegrasyonunun diğer melez modellere göre tahmin başarısı açısından üstünlüğü ortaya çıkmıştır. Ayrıca, ani sel değerlendirmelerinde eğim açısı, nehirden uzaklık ve arazi yüksekliğinin belirleyici roller oynadığı, buna karşın hidrolojik toprak grubu, litoloji ve topoğrafik ıslaklık indeksi gibi jeolojik değişkenlerin elde edilen en başarılı modelin tahmin kabiliyetine en düşük etkiyi yaptığı tespit edilmiştir. Son yıllarda birçok araştırma alanında yaygın olarak kullanılan derin öğrenme tekniklerinin ani sel tehlike haritalanması modellerine entegre edilebildiğini göstermek için Tien Bui ve ark., (2020), derin öğrenme sinir ağını (DLNN) Vietnam'ın tropik fırtına etkisi altında olan alanlarından birine uygulamıştır. Öne sürülen derin öğrenme modeli ayrıca ANN ve destek vektör makineleri (SVM) gibi diğer makine öğrenmesi yöntemleriyle kıyaslanmıştır ve bu iki yönteme kıyasla hassas bölgeleri daha iyi tespit ettiğini göstermişlerdir. Benzer olarak, Panahi ve ark., (2021) literatürde derin öğrenme algoritmalarında yaygın olarak benimsenen konvolüsyonel sinir ağlarını (CNN) ve tekrarlayan sinir ağlarını (RNN) adlı iki farklı mimariyi kullanarak ani sel duyarlılık haritaları oluşturmuştur. Her iki derin öğrenme metodolojisini eğitmek için adım adım ağırlık değerlendirme oranı (step-wise weight assessment ratio) analizini hesaba katmışlardır ve İran'ın Golestan Eyaleti için yapılan değerlendirmelerde CNN'nin RNN' den az farkla da olsa daha yüksek tahmin performansı gösterdiği sonucuna varılmıştır.

Ek olarak, modern modelleme tekniklerindeki gelişmeler, taşkın duyarlılık haritalama alanında optimizasyon tekniklerinin kullanımını da kolaylaştırmıştır. Böylelikle, yakın literatürde hem grid arama ve Bayes optimizasyonu gibi klasik arama algoritmaları hem de genetik algoritma ve parçacık sürü optimizasyonu gibi son teknoloji meta-sezgisel algoritmaların adapte edilen makine öğrenmesi algoritmalarının içsel parametrelerinin ayarlanmasında kullanımı gerçekleştirilmiştir. Bu çalışma kapsamında da Kentucky Nehri Havzası için ilk kez ani sel tehlike haritalaması gerçekleştirilmiştir. Bu amaca yönelik olarak ise rastgele orman (RF) ve parçacık sürü optimizasyonu (PSO) algoritması entegre edilmiştir. Bu çalışma kapsamında Kentucky Nehri havzası için makine öğrenmesi yöntemlerinin uygulanması ise Python programlama dilinde gerçekleştirilmiş olup, RF ve PSO algoritmalarında sırasıyla “sklearn.ensemble” ve “NiaPy” kütüphanelerinden yararlanılmıştır.

Grid bazlı elde edilen tahminler neticesinde ani sel tehlikesi bakımından tetikleyici rol oynayan faktörler ise ifade edilerek tartışılmıştır. Böylelikle kara kutu modeller olarak ifade edilen makine öğrenmesi yöntemlerine ait sonuçların yorumlanabilirliğinde gelişmeler kaydedilmiştir. Özetle, bu çalışmada esas olarak amaçlananlar:

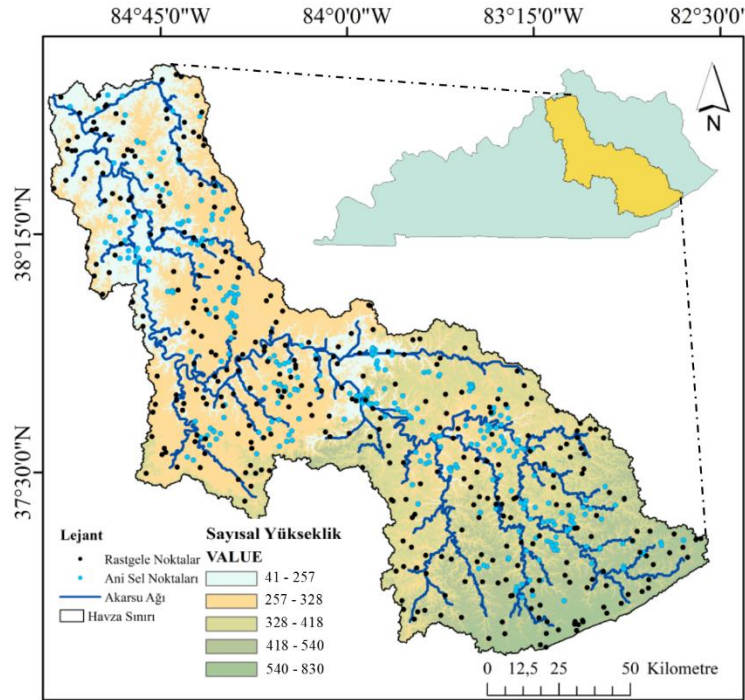
- Ağaç tabanlı algoritmalar arasında en yaygın olanlarından olan RF yönteminin ani sel duyarlılık haritalamasında kullanılabilirliğinin gösterilmesi,
- Klasik optimizasyon algoritmaları yerine meta-sezgisel PSO algoritmasının RF içsel parametrelerinin ayarlanması sürecine dahil edilmesi,
- Makine öğrenmesi teknikleri ile elde edilen sonuçların yorumlanabilirliğinin değerlendirilmesi şeklinde ifade edilebilmektedir.

Bu çalışmada geliştirilen sistemin gerek metodolojik gerekse de pratik bakış açıları ile ani sel duyarlılığı değerlendirmeleri alanına önemli katkılar sağlayacağına, araştırmacılara ve uygulayıcılara yorumlanabilir ani sel duyarlılık haritaları oluşturmalarında yardımcı olabileceğine inanılmaktadır.

MATERYAL VE METOT

Çalışma Alanı

Yaklaşık 105000 kilometrekare yüzey alanına sahip Kentucky eyaleti Amerika Birleşik Devletleri'nin güney doğusunda yer almaktadır. Arazi kullanım yapısı bakımından oldukça farklılık içeren eyalet ormanlık alanları, tarımsal arazileri ve kentsel bölgeleri barındırmaktadır. Çalışma alanı olarak seçilen Kentucky Nehri Havzası da eyaletin sınırları içerisinde yer almaktadır. Kentucky Nehri yaklaşık olarak 420 kilometrelik uzunluğa sahip olup, eyaletin doğusunda bulunan dağlık arazilerden doğup Ohio Nehri'ne dökülmektedir. Kentucky Nehri ayrıca yan kolları açısından da oldukça zengindir. Öyle ki, Red Nehri, Dix Nehri, North Fork Kentucky Nehri, South Fork Kentucky Nehri ve Middle Fork Kentucky Nehri bu yan kolların en büyük besleme noktalarını oluşturmaktadır. Drenaj alanı bakımından eyaletin önemli bir kısmını kaplayan Kentucky Nehri 18000 kilometrekarelik yüzey alanına sahiptir. Çalışma alanı Şekil 1'de gösterilmiştir. Şekilden de anlaşılacağı üzere 41 metreden başlayan havza yükseklikleri yer yer 830 metrelik seviyelere ulaşmaktadır. Görece dağlık ve yüksek rakıma sahip bölgeler havzanın güney batısında yer alırken, havzanın orta kesimleri kentsel alanları kapsamaktadır.



Şekil 1. Çalışma alanı

Ani Selleri Tetikleyen Faktörler

Ani seller, şiddetli yağış, hızlı kar erimesi ve baraj duvarı çökmesi gibi nedenlerden dolayı suyun bir bölgede hızla birikmesi nedeniyle genellikle 6 saatten daha kısa sürede meydana gelen, ani ve yoğun seller olarak tanımlanmaktadır (Marchi ve ark., 2010). Ani seller her bölgede meydana gelebilir ve can ve mal kaybı için tehdit oluşturacak önemli hasara neden olabilmektedir. Ani seller, nehir selleri ve kıyı selleri gibi diğer sel türlerinden farklı olarak, aniden ve çok daha kısa bir sürede oluşur (Ekmekcioğlu ve ark., 2022). Örneğin nehir taşkınları, nehirlerin kıyılarından taşmasına neden olan uzun süreli şiddetli yağışlardan kaynaklanırken, kıyı taşkınlarına yüksek gelgitler, fırtınalar ve şiddetli yağışların birlikte neden olduğu görülmektedir (Ikeuchi ve ark., 2017). Ani seller genel olarak kentsel bölgelerde yaşansa da iklim şartlarının değişimi ile birlikte meydana gelen ani ve şiddetli yağışlar ile arazinin her bölgesinde meydana gelebilir hale bürünmüştür. Arazi şartlarının halen çok etkili olduğu ani seller toprak yapısına, arazi kullanım durumuna, yağışlara ve birtakım hidro-jeolojik değişkenlere de bağlı olarak meydana gelebilmektedir. Bu çalışma kapsamında 12 farklı ani selleri tetikleyen faktör göz önünde bulundurulmuştur (Şekil 2). Bu faktörler aşağıda kısa açıklamaları ile birlikte listelenmiştir:

1) Sayısal Yükseklik: Topoğrafya olarak da nitelendirilen bu değişken hemen her doğal afetin meydana gelmesinde veya meydana geliş biçiminde etkili rol oynamaktadır. Genellikle yüksek noktalar taşkınların daha az meydana geldiği, düşük rakımlı noktalar ise taşkınların daha sık görüldüğü noktalar olarak literatürde öne sürülmüştür (Arora ve ark., 2021). Bu gerçek arazi kullanım yapısı ve bölgelerin eğim karakteristiklerine göre de istisnalara sahiptir.

2) Eğim: Bu faktör suyun daha çok nehirden uzaklaşmasını veya nehre yakınsamasını kontrol etmektedir. Eğim değerlerinin düşük olduğu bölgelerde taşkınların daha fazla olması beklenmektedir (Fang ve ark., 2020). Yüksek noktalarda ise suyun yer çekiminin de etkisi ile düşük rakımlı bölgelere taşınması dolayısıyla ani seller daha az görülmektedir.

3) Bakı: Bu faktör özellikle toprak nem içeriği ile yüksek oranda ilişkilidir. Yoğun güneşe maruz kalan topraklarda toprak su içeriği görece daha az olacağından ani sellerin daha az olacağı, doymuş topraklarda ise taşkın olaylarının daha sık görüleceği beklenmektedir.

4) Eğrilik: Eğrilik, içbükey ve dışbükey gibi farklı şev oluşum şekillerine dayalı şev stabilitesini yansıtmaktadır. Yüzey suyun drenajında önemli olan eğrilik faktörü, plan ve profil eğriliği olarak ikiye ayrılmaktadır.

5) Akarsu yoğunluğu: Yüzey akışının yayılmasını ve bu yayılmanın büyüklüğünü kontrol ettiği için taşkınlar açısından önemlidir. Akarsu yoğunluğunun fazla olduğu bölümlerde bölgenin arazi kullanım yapısına da bağlı olmakla birlikte taşkınların sayısında genellikle artış gözlemlenmektedir. Akarsu yoğunluğu raster haritası ArcMap 10.3 bünyesinde yer alan ve her hücrenin etrafındaki bir yarıçap içinde kalan çoklu çizgi (*polygon*) özelliklerinden birim başına büyüklük hesaplayan “line density” aracı ile sayısal değer kazanmıştır. Elde edilen raster harita diğer girdilerle homojen özellik taşıması bakımından sayısal yükseklik verisi ile aynı çözünürlük olan 30m x 30m piksel büyüklüklerinde üretilmiştir.

6) Akarsuya uzaklık: Bu faktör hidro-morfolojik açıdan büyük öneme sahiptir. Öyle ki, akarsuya yakın olan bölgeler taşkınlar açısından diğer bölgelere nazaran daha riskli olarak ön plana çıkmaktadır. Akarsuya uzaklık kriterinin elde edilmesinde Öklid uzunluğundan yararlanılarak 30m x 30m çıktı raster dosyası ArcMap 10.3 ile elde edilmiştir.

7) Şiddetli yağış: Bu faktör ile bölgede meydana gelen daha yüksek tekerrür aralığına sahip yoğun yağış olaylarının olasılığı tanımlanmaktadır. Şiddetli yağış olayları genellikle toprak neminin doymuşlaşmasına yol açmaktadır. Böylelikle, şev içindeki basınca bağlı olarak şev duyarsızlığı meydana gelebilmektedir (Kim ve ark., 2012). Tabii olarak, şiddetli yağışların daha çok görüldüğü bölgelerde ani sellerin de meydana gelmesi sıklaşmaktadır.

8) Litoloji: Bir bölgenin jeolojik özelliklerinin tanımlanmasında kullanılmaktadır. Farklı litoloji tiplerinde şev stabilitesi de farklılık göstermektedir ve buna bağlı olarak, toprak kayma gerilmelerinde değişkenlikler meydana gelmektedir.

9) Eğri numarası: Eğri numarası hesaplanma stratejisi gereği arazi kullanımı, toprak tipi ve toprak nemi gibi kriterleri içerisinde barındırmaktadır (Jaafar ve ark., 2019). Özel olarak ise, bölgenin geçirimsizlik durumunu temsil etmek için kullanılmaktadır. 0 ile 100 arasında değişen eğri numarası arttıkça geçirimsizlik de artar ve taşkın olaylarının görülme olasılığı artmaktadır (Costache ve ark., 2020).

10) NDVI: Bitki yoğunluklarına bağlı olarak yüzeyin yansıtıcılık özelliklerini nicel olarak tarif etmektedir. Bir bölgede NDVI değerinin artması o bölgedeki yeşil alanın fazlalığını ve yağışa bağlı olarak oluşan yüzey suyunun daha geç akışa geçmesini temsil etmektedir. Dolayısıyla, bu gibi bölgelerde taşkın olaylarının daha az meydana gelmesini sağlamaktadır.

11) TWI: Genel olarak bir bölgedeki toprağın ıslaklığının nicel olarak ifade edilme yöntemidir. Topoğrafik özellikler ve toprak nemi ile doğru orantılı olarak nehir çıkışına doğru yer çekimi göz önüne alınarak hareket edildiğinde, TWI değerleri artma eğilimindedir (Andaryani ve ark., 2021).

12) TRI: Bu faktör bir havzanın lokal topoğrafik koşulları ile ilişkili olmakla birlikte, topoğrafyanın heterojenliğini nicel olarak açıklayan bir indekstir (Riley ve ark., 1999).

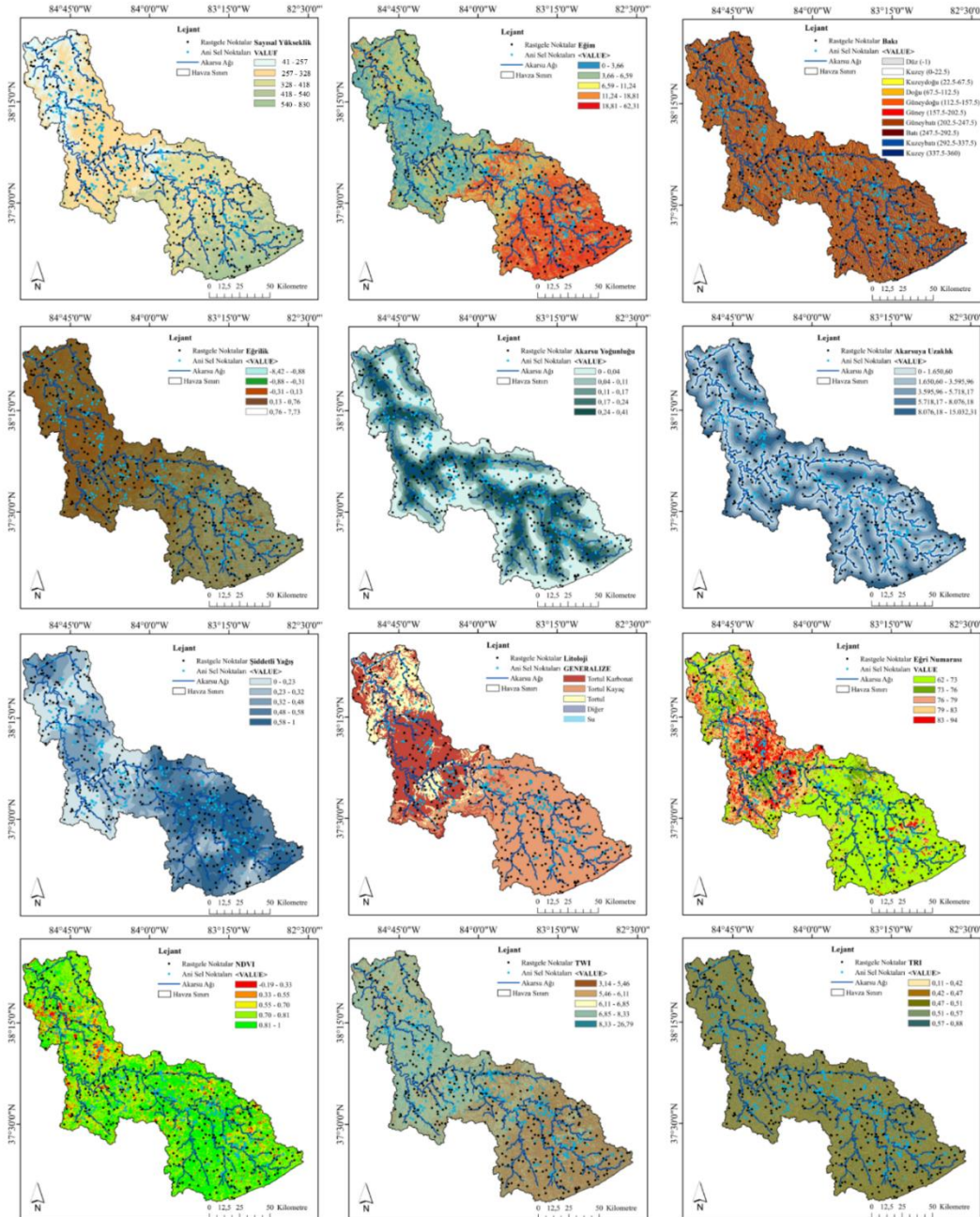
Çalışmada kullanılan ani selleri tetikleyici faktörlere karşılık gelen istatistiksel bilgiler Çizelge 2’de verilmiştir.

Çalışmada kullanılan ani sellerin tetikleyen 12 adet parametre veri erişilebilirliği ve kullanım kolaylığı gözetilerek seçilmiştir. Öyle ki, bu parametrelerden birçoğu ücretsiz olarak çevrimiçi erişilebilirliği bulunan ve 30m x 30m uzamsal çözünürlüğe sahip sayısal yükseklik haritasından elde edilmiştir. Bu parametreler eğim, bakı, eğrilik, akarsu yoğunluğu, akarsuya uzaklık, TWI ve TRI olarak sıralanabilmektedir.

Çizelge 2. Çalışma kapsamında kullanılan girdi değişkenlerine ait istatistiksel özet

| Faktör | Birim | Minimum | Ortalama | Maksimum | Standart Sapma |
|-------------------|--------|---------|-----------|------------|----------------|
| Sayısal yükseklik | m | 123 | 314.5742 | 830 | 92.7356 |
| Eğim | derece | 0 | 12.3262 | 43.8533 | 8.8513 |
| Bakı | - | 0 | 178.7057 | 359.2930 | 102.9426 |
| Eğrilik | - | -2.9439 | -0.0740 | 3.0808 | 0.6264 |
| Akarsu yoğunluğu | - | 0 | 0.1265 | 0.3867 | 0.0877 |
| Akarsuya uzaklık | m | 0 | 3021.2398 | 12497.7002 | 2595.2947 |
| Şiddetli Yağış | - | 0 | 0.4687 | 1 | 0.2357 |
| *Litoloji | - | - | - | - | - |
| Eğri numarası | - | 0 | 76.4117 | 91 | 7.4414 |
| NDVI | - | 0 | 0.7435 | 0.9918 | 0.1451 |
| TWI | - | 3.7858 | 7.2205 | 23.6912 | 2.4393 |
| TRI | - | 0.1944 | 0.4881 | 0.8333 | 0.0894 |

* Litoloji faktörü sınıfsal bir değişken olduğundan çizelgede yer verilmemiştir. Bu çalışmada kullanılan çalışma alanında gözlenen litoloji sınıfları dörde ayrılmaktadır: LT2: Tortul karbonat, LT3: Tortul kaycaç, LT4: Tortul, LT6: Diğer)



Şekil 2. Çalışmada yararlanılan ani sel tehlikesini tetikleyen faktörlerin harita gösterimi

Söz konusu parametrelerin yanı sıra, şiddetli yağış olasılığı verisi Amerika Birleşik Devletleri Ulusal Okyanus ve Atmosfer Dairesi (NOAA)'nin sitesinden (<https://www.noaa.gov/>) doğrudan ücretsiz olarak elde edilmiştir. Ayrıca, odaklanılan bölgeye ait litoloji verisi Amerika Birleşik Devletleri Jeoloji Araştırmaları Kurumu (USGS) kaynakları (<https://www.usgs.gov/>) içerisinde elde edilmiştir. Bu araştırma kapsamında sel duyarlılık analizlerinde kullanılan eğri numarası (CN) parametresi ve NDVI parametresine dair veriler ise sırasıyla Jaafar ve ark., (2019) ve Moderate Resolution Imaging Spectroradiometer (MODIS) Terra 6 kaynaklarından temin edilerek Kentucky havzası özelinde uygulanmıştır.

Ani Sel Geçmiş Kayıtları

Veriye dayalı tahmin metotları kullanımında en önemli çıkış noktası geçmişte yapılan ölçümlerdir. Bu ölçümler gerek sınıflandırma gerekse de regresyon problemlerinde ileriye dönük gerçekleştirilecek projeksiyonlara temel teşkil etmektedir. Bu bağlamda, geçmişte meydana gelen taşkın olaylarına ait kayıtların da tutulması geleceğe dönük taşkın haritalama çalışmaları için ciddi önem arz etmektedir. Bu çalışma kapsamında kullanılan makine öğrenmesi yöntemlerine dayalı ani sellerin haritalanması çalışmasında ise benzer şekilde son yirmi yılda (2000-2020) Kentucky Nehri havzasında taşkın kayıtlarından yararlanılmıştır. İlgili veriler Amerika Birleşik Devletleri'ne bağlı Ulusal Çevresel Bilgi Merkezine (NCEI 2021) kurumuna ait internet sitesinden açıkça erişilebilir konumda olduğundan doğrudan temin edilmiştir. Bu doğrultuda, Kentucky Nehri havzası sınırları içerisinde meydana gelmiş olan 343 ani sel olayına ait koordinat verilerinden yararlanılmıştır. Bu veriler Google Earth yazılımı ve ArcGis 10.3 yazılımları yardımıyla coğrafi bilgi sistemine aktarılmış ve analizlere ait geri kalan veriler bu noktalar özelinde temin edilmiştir. Tahmin işlemlerinin adil bir platformda gerçekleştirilebilmesi için ise ani sel sayıları ile aynı sayıda (343) rastgele nokta (ani sel olmayan nokta olarak) belirlenmiştir. Bu noktalar belirlenirken ArcGis 10.3 yazılımının içerisinde yer alan Rastgele Nokta Oluştur (Create Random Points) fonksiyonundan yararlanılmıştır. Toplamda elde edilen 686 noktaya tekabül eden tüm girdi değişkenlerine ait sayısal ve kategorik veriler tanıtılmış olup, tüm veri seti %70 eğitim seti ve %30 test seti olmak üzere iki ana gruba ayrılmıştır. Oluşturulan melez makine öğrenmesi metodunun performans değerlendirmesinde ise test setinden yararlanılmıştır.

Rastgele Orman Algoritması

Rastgele Orman (RF) algoritması ilk olarak Breiman, (2001) tarafından hem sınıflandırma hem de regresyon problemlerinin çözümünde kullanılmak üzere ortaya atılmıştır. RF metodu parametrik olmayan bir metottur ve özelinde torbalama (bagging) ve önyükleme (bootstrapping) yöntemlerinin kombine uygulanmasını esas almaktadır. Torbalama işleminde çoklu ağaçların kararları toplanıp, önyükleme işleminde ise en uygun dallanmanın belirlenebilmesi için rastgele bir biçimde alt-örneklemeler gerçekleştirilmektedir. En uygun dallanmanın belirlenmesinde kantitatif olarak kullanışlı bir gösterge olan Gini indisi kullanılmaktadır (Liu ve ark., 2021; Xu ve ark., 2022). Gini indisi ayrıca seçilen alt-örneklemeler arasındaki homojeniteye de odaklanmaktadır (Goswami ve ark., 2018). Öyle ki, Gini indisindeki azalma modeldeki heterojeniteyi artırmakta iken, indis değerinde artış görülmesi homojenitenin artmasına işaret etmektedir (Habba ve ark., 2018). RF metodu karar ağaçlarının oluşturulması için geleneksel olarak gösterilen sınıflandırma ve regresyon ağaçları (CART) yöntemini kullanmaktadır (Ekmekcioğlu ve ark., 2020). Fakat, CART' dan farklı olarak RF algoritmasında eğitim için ayrılan veri setine göre sınırlı hassasiyete sahip olunması ve görece daha düşük varyans değerlerini sağlaması avantaj sağlamaktadır. Bunun yanı sıra, RF metodu diğer ağaç tabanlı algoritmalarına nazaran kullanım kolaylığı, eksik verileri tahmin etme becerisi, aykırı değerleri ayıklama becerisi ve çok girdiye sahip yüksek boyutlu veri setlerindeki tahmin başarısı gibi nedenler ile daha sık olarak kullanılmaktadır.

RF algoritmasında dikkat edilmesi gereken en önemli nokta ise yöntemin aşırı öğrenmeye meyilli olmasıdır. Bu nedenle, RF yönteminin kullanılmasında eğitim setinin optimizasyon algoritmaları ile eğitilmesi ve çapraz doğrulama ile valide edilmesi önerilmektedir (Koc ve ark., 2021). Bu sayede, algoritmanın iç parametreleri optimize edilir ve veri setine uygun bir model konfigürasyonunun oluşturulması sağlanmaktadır.

Parçacık Sürü Optimizasyonu

Kuşların doğada hareket etme davranışlarını esas alan PSO yöntemi Kennedy ve Eberhart, (1995) tarafından önerilmiş ve sonrasında literatürde en yaygın olarak kullanılan stokastik meta sezgisel algoritmalar arasına girmiştir. PSO algoritmasında iteratif bir süreç izlenmek suretiyle parçacıkların arama uzayı arasındaki lokasyonları ayarlanmaktadır. En uygun sonuç ise yine bu parçacıkların uzaydaki hareket hızına ve pozisyonuna bağlı olarak bulunmaktadır. Her bir parçacık için uygunluk değeri (fitness value) hesaplanılarak optimizasyon süreçleri gerçekleştirilmektedir. Aşağıda yer alan Denklem 1 ve Denklem 2’de hesaplanma yöntemleri ifade edilen parçacık hızı ve pozisyonu en iyileme sürecinde kilit rol oynamaktadır:

$$V_i(t + 1) = wV_i(t) + r_1c_1 (x_{pbest_i} - x_i(t)) + r_2c_2 (x_{gbest_i} - x_i(t)) \quad (1)$$

$$x_i(t + 1) = x_i(t) + V_i(t + 1) \quad (2)$$

Burada, V ve x sırasıyla parçacık hızı ve pozisyonunu ifade etmektedir. Ayrıca, P_{best} ise i parçacığının optimum lokasyonunu ve G_{best} ise bu parçacığın her bir iterasyondaki en uygun değerini ifade etmektedir. r_1 ve r_2 değişkenleri 0 ve 1 arasında değişen ve üniform olarak dağılım karakteristiği gösteren rastgele değerleri temsil etmektedir. w ise eylemsizlik ağırlığını gösterirken, C_1 ve C_2 değişkenleri sırasıyla bilişsel bileşeni (cognitive component) ve sosyal bileşeni (social component) göstermektedir. Tipik bir PSO algoritmik süreci şu şekilde özetlenebilmektedir. Öyle ki, süreç konum ve hız bilgilerini içeren rastgele oluşturulmuş sürüler aracılığıyla başlatılır ve popülasyondaki parçacıklar için uygunluk değeri, amaç fonksiyonuna dayalı olarak hesaplanmaktadır. Daha sonra, yerel ve global optimumlar elde edilir ve her parçacığın hız ve konum bilgisi güncellenip gerekli yer değiştirmesi gerçekleştirilmektedir. Son olarak ise, aramanın sonlandırılması için önceden belirlenen kriter sağlandığı takdirde arama sonlandırılır veya istenilen yakınsama gerçekleştirilene kadar ikinci adıma geri dönülerek tüm işlemler tekrarlanmaktadır.

Performans Değerlendirme Metrikleri

Sınıflandırma problemlerinde makine öğrenmesi algoritmalarının performans değerlendirmelerinin gerçekleştirilebilmesi için birçok kriter mevcuttur. Bu kriterler arasında en yaygın olarak kullanılanları precision, recall, F-1 skoru, ve ROC eğrisinin altında kalan alan (AUROC)’dır. İkili sınıflandırma problemlerinde bu kriterlerin hesaplanabilmesi için ise detayları Çizelge 3’de sunulan dört farklı alt-metrik yer almaktadır: Gerçek Pozitif (TP), Yanlış Negatif (FN), Yanlış Pozitif (FP), ve Gerçek Negatif (TN).

Çizelge 3. Sınıflandırma problemlerinde performans değerlendirmesindeki olası senaryolar

| | | Tahmin Edilen Durum | |
|----------------|-------|---------------------|---------------------|
| | | Evet | Hayır |
| Gözlenen Durum | Evet | Gerçek Pozitif (TP) | Yanlış Negatif (FN) |
| | Hayır | Yanlış Pozitif (FP) | Gerçek Negatif (TN) |

Burada Gerçek Pozitif (TP) pozitif örneklere göre doğru tahminlerin sayısını gösterirken, Yanlış Negatif (FN) pozitif örneklere göre yanlış tahminlerin sayısını temsil etmektedir. Ayrıca, Yanlış Pozitif

(FP) ve Gerçek Negatif (TN) ise sırasıyla negatif örneklere göre yanlış tahminlerin sayısını ve negatif örneklere göre doğru tahminlerin sayısını temsil etmektedir. Tüm bu hesaplamalara göre belirlenen Precision ve Recall değerleri sırasıyla Denklem 3 ve Denklem 4 ile belirlenmektedir:

$$Precision = \frac{TP}{TP+FP} \quad (3)$$

$$Recall = \frac{TP}{TP+FN} \quad (4)$$

Bu çalışma kapsamında ayrıca PSO-RF melez modeli F-1 skoru (F) metriğine göre değerlendirilmiştir. Öyle ki, F değerinin hesaplanmasında precision ve recall metriklerinin harmonik ortalamalarından yararlanılmaktadır:

$$F = \frac{2 \times precision \times recall}{precision + recall} \quad (5)$$

AUROC metriği ise 0.5 ve 1 arasında değişkenlik göstermektedir. AUROC ayrıca alıcı işletim karakteristik eğrisi (receiver operating characteristic curve) ile görsel hale getirilmiştir. Bu grafiğin x eksenini ve y eksenleri ise sırasıyla yanlış pozitif oranı ($FPR=FP/(FP+TN)$) ve gerçek pozitif oranı (TPR, yani, recall) değerlerinin işlenmesi ile elde edilmektedir (Ekmekcioğlu ve ark., 2022). AUROC, ikili sınıflandırma problemlerinde makine öğrenmesi algoritmalarının performans ölçümlerinde en yaygın olarak kullanılan ve en güvenilir yöntemlerden biri olarak literatürde yerini almıştır (Ekmekcioğlu ve Koc, 2022).

BULGULAR VE TARTIŞMA

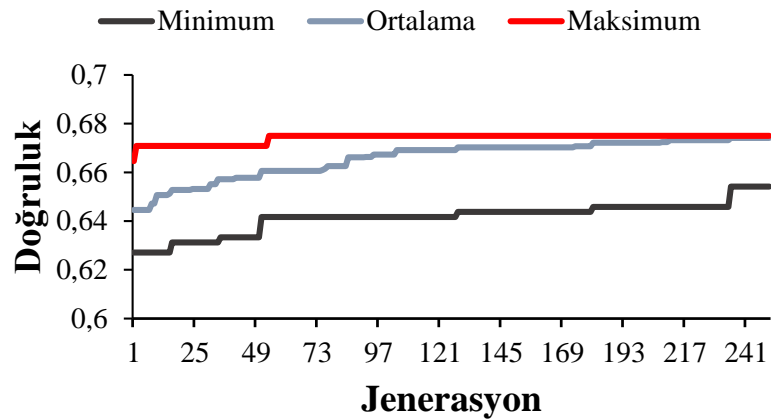
Bu çalışma kapsamında yapılan ani sel analizlerinde kullanılan makine öğrenmesi yöntemi olarak rastgele orman algoritması en uygun model konfigürasyonunun sağlanması adına parçacık sürü algoritmasına tabi tutulmuştur. Literatürde sıklıkla kullanılan RF algoritması ile farklı problemler için farklı içsel parametreler kullanılmakla birlikte, içsel parametrelerin analizi yapılacak probleme uygun olarak ayarlanması oldukça önem arz etmektedir. Bu noktada eğitim setinden yararlanılmaktadır. Yani, tüm veri seti öncelikle %70'i eğitim seti ve %30'u test seti olmak üzere iki ana gruba ayrılmış ve sonrasında eğitim seti kendi içerisinde 5-fold çapraz doğrulama uygulanmak suretiyle modellenmiştir. Bu işlem ile hem aşırı öğrenmenin önüne geçilmiş olup hem de adil bir eğitim modeli kurulmuştur. RF algoritmasında optimum hale getirilmesi için ise iki adet içsel parametre seçilmiştir. Bunlar ağaç sayısı ve ağaç derinliğidir. Öte yandan minimum ayrılan örneklem sayısı (minimum number of samples split) parametresi ise optimizasyon sürecinin daha kısa ve efektif sürdürülmesinin yanı sıra model performansında diğer iki parametreye göre görece daha az etkisi olduğu bilindiğinden dolayı sabit tutulmuştur. Bu doğrultuda, ağaç sayısı ve ağaç derinliği parametreleri sırasıyla 10 ile 200 ve 1 ile 50 aralıklarında 10'ar ve 1'er adımlar ile aranmıştır. Minimum ayrılan örneklem sayısı parametresi ise 9 olarak belirlenmiştir. Algoritma sürecinin adil olarak gerçekleştirilmesi ve her bir iteratif süreçte aynı eğitim ve test setlerinin kullanılması için ise random state parametresi 0 olarak belirlenmiştir. Eğitim seti içerisinde kullanılan PSO algoritmasında ise RF yöntemi gibi birçok içsel parametre mevcuttur. Literatürde birtakım çalışmalar da aynı RF' de gerçekleştirildiği gibi PSO algoritmasının da genetik algoritma gibi geleneksel meta sezgisel algoritmalar gibi yöntemler ile optimize edilmesini gerçekleştirirse de yaygın olarak kullanılan yöntem PSO parametrelerinin deneme yanılma ile belirlenmesi yönündedir. Bu çalışmada PSO algoritmasının içsel parametrelerinin belirlenmesinde detaylı bir literatür taraması gerçekleştirilmiş ve parametreler bu yol ile belirlenmiştir. RF ve PSO algoritmalarına ait optimize edilen ve kabulü gerçekleştirilen içsel parametreler Çizelge 4' de özetlenmiştir.

Çizelge 4. Sınıflandırma problemlerinde performans değerlendirmesindeki olası senaryolar

| Metot | İçsel parametre | Değer aralığı | Arama adımı | Optimum değerler | Toplam Kombinasyon | Optimizasyon Süresi (s) |
|-------|---------------------------------|---------------|-------------|------------------|--------------------|-------------------------|
| RF | Ağaç sayısı | 10-200 | 10 | 100 | 1000 | 97.2 |
| | Maksimum ağaç derinliği | 1-50 | 1 | 7 | | |
| | Minimum ayrılan örneklem sayısı | Sabit | - | 9 | | |
| | Random state | Sabit | - | 0 | | |
| PSO | Popülasyon sayısı | - | - | 100 | - | - |
| | Bilişsel bileşen | - | - | 2.8 | | |
| | Sosyal bileşen | - | - | 1.45 | | |
| | Eylemsizlik ağırlığı | - | - | 0.3 | | |
| | Minimum parçacık hızı | - | - | -1.5 | | |
| | Maksimum parçacık hızı | - | - | 12 | | |

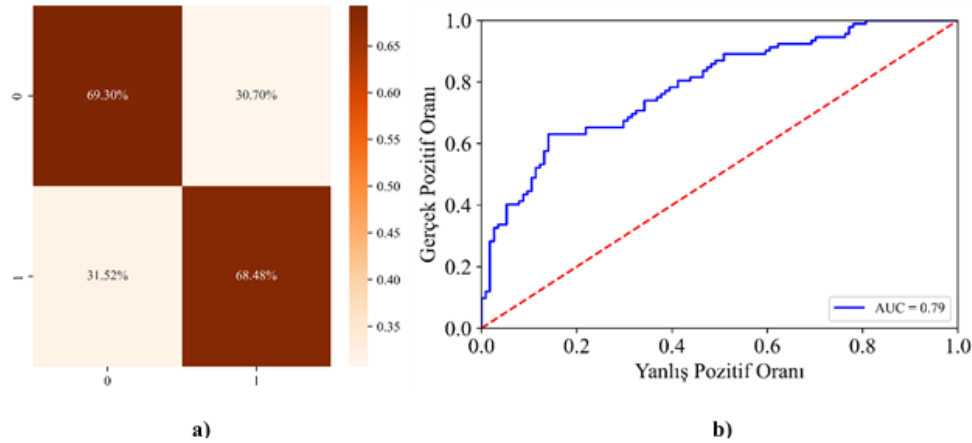
Not: Yapılan optimizasyon işlemleri AMD Ryzen 5900X 12-Çekirdekli işlemci (Windows 10, 3.70 GHz ve 32.9 GB RAM) özelliklerine sahip iş istasyonu ile gerçekleştirilmiştir.

Ek olarak, RF algoritmasının eğitim setinin tabii tutulduğu PSO algoritması ile gerçekleştirilen öğrenme sürecinde amaç fonksiyonu olarak Accuracy indikatörü belirlenmiştir. Sınıf dengesizliğinin olmadığı ikili sınıflandırma problemlerinde en yaygın olarak Accuracy kullanıldığı literatürde sıklıkla ifade edilmiştir. Bu çalışmada gerçekleştirilen optimizasyon sürecinde elde edilen doğruluk değerleri de her bir jenerasyonda (iterasyon) hesaplanmıştır. Her bir optimizasyon adımında elde edilen minimum, ortalama ve maksimum doğruluk değerleri Şekil 3’de grafiksel gösterim ile ifade edilmiştir. Optimizasyon süreçleri çalışma prensipleri gereği en iyileme işlemine odaklandığından dolayı maksimum doğruluk ile ilgilenmektedir ve sonuçları bu yaklaşım ile sunmaktadır. Bu çalışmada gerçekleştirilen optimizasyon kapsamında ise 53. iterasyonda elde edilen RF parametreleri ile en iyileme gerçekleştirilmiştir. Yani Çizelge 4’ de elde edilen ağaç sayısı ve maksimum ağaç derinlikleri bu iterasyonda elde edilmiştir. Arama uzayında belirlenen toplam jenerasyon sayısı da 250 olarak seçilmiştir. Yakınsama grafiklerine göre maksimum doğruluk değerlerinde 53. iterasyondan sonra herhangi bir değişim gözlenmediğinden dolayı maksimum iterasyon değerinin 250’den fazla olmasına gerek olmadığı gözlemlenmektedir.

**Şekil 3.** Melez PSO-RF algoritmasına ait eğitim veri setine göre yakınsama grafikleri

Optimizasyon süreci neticesinde elde edilen optimum model konfigürasyonuna göre model inşası gerçekleştirilmiş ve test veri setine bu model uygulanmıştır. Test verisine uygulanarak elde edilen model sonuçları ise gerçek gözlem değerleri ile karşılaştırılarak modelin test performansı değerlendirilmiştir. Bu performans değerleri melez PSO-RF modeline daha önce hiç tanıtılmayan bir veri seti olduğundan

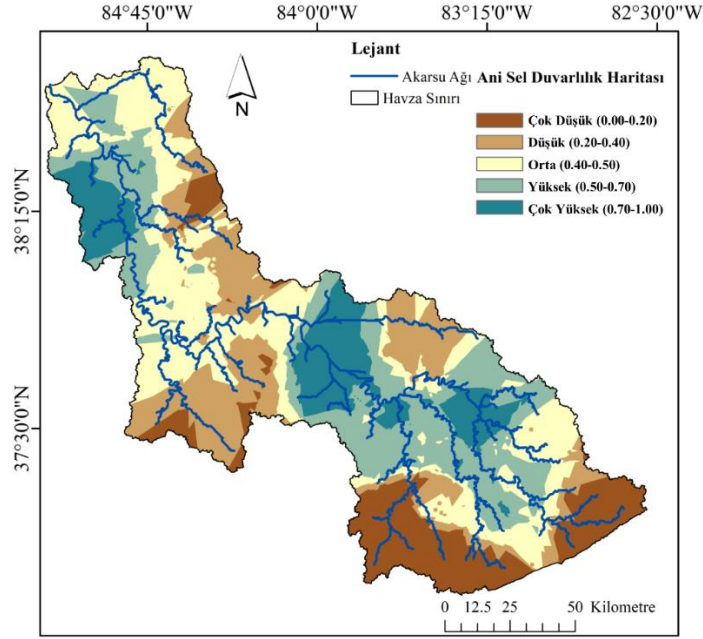
dolayı modelin gerçek performansını yansıtmaktadır. Test veri seti kullanılarak elde edilen model başarısı hata matrisi (confusion matrix) (Şekil 4a) ve AUROC grafiği (Şekil 4b) ile görsel olarak ifade edilmiştir. Şekil 4a incelendiğinde melez PSO-RF modelinin ani sel olmayan durumların %69.30' unu doğru tahmin ettiği ve ani sel olan durumların %68.48' ini doğru tahmin ettiği gözlemlenmektedir. Öte yandan, her iki senaryo için de yaklaşık %30' luk bir hata payı mevcuttur. Bu hata payının sebebi görece büyük bir havzada meydana gelen ani sel sayılarının az olmasından doğmaktadır. İkili sınıflandırma problemlerinde de diğer sınıflandırma ve regresyon problemlerinde olduğu gibi genel olarak örneklem sayısı arttıkça modelin öğrenme becerisi artacak ve benzer doğrultuda modelin test verisi ışığındaki performansı da yükselme göstermektedir. Ayrıca, havzanın sınırları içerisinde yer alan arazi kullanım tiplerinin lokal olarak büyük farklılıklar içermesi de makine öğrenmesi modellerinin başarı performansında düşüklüklere sebebiyet verebilmektedir. Bu bağlamda, hata matrisi sadece baz alınması gereken bir performans başarı göstergesi değildir. Melez PSO-RF modelinin performansı ayrıca hesaplanan AUROC değeri göz önünde bulundurularak değerlendirildiğinde oldukça başarılı olarak göze çarpmaktadır. Zira, literatürde sınıflandırma problemlerinin performans değerlendirmesi ile ilgili yapılan tartışmalarda özellikle ikili sınıflandırma problemlerinde AUROC değerlerinin 0.8 ve daha üzeri değerleri elde edildiğinde modelin çok yüksek bir performans başarısına sahip olduğu kabul edilmektedir. Bu çalışmada kullanılan modele ait AUROC değeri de 0.79 olarak elde edilmiştir (Şekil 4b). Test veri seti için oldukça yüksek olan bu değer literatürce yüksek performansı gösteren 0.8 değerine oldukça yakındır.



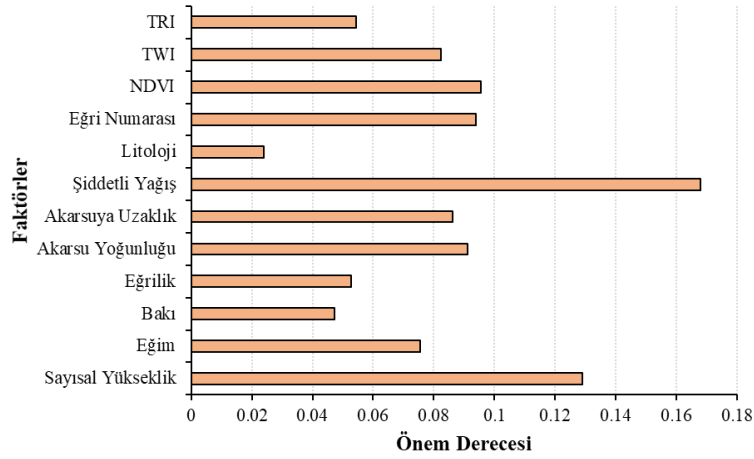
Şekil 4. Melez PSO-RF modeline ait test verisi göz önüne alınarak elde edilen hata matrisi ve AUROC grafiği

Makine öğrenmesi yöntemlerine dayalı taşkın haritalaması çalışmalarında elde edilen sonuçlar harita üzerinde interpolasyon yöntemleri ile yayıldığında bölgesel bağlamda taşkın tehlike haritaları elde edilebilmektedir. Bu çalışma kapsamında da test için ayrılan veri seti için yapılan tahmin sonuçları her bir nokta için elde edilmiş olup, harita üzerinde 0 (ani sel olmaz) ve 1 (ani sel olur) olacak şekilde noktasal bazda tanıtılmıştır. Daha sonra ise ArcGis 10.3 yazılımı kullanılarak Kriging yöntemi ile interpolate edilmiştir. İnterpolasyon neticesinde elde edilen haritada ise grid bazında elde edilen değerler Jenk's Natural Break Metodu kullanılarak beş farklı kategoriye (Çok düşük, düşük, ortalama, yüksek ve çok yüksek tehlikeli) ayrılmıştır. Natural Breaks (Jenks) yöntemi, CBS ve mekânsal istatistiklerde kullanılan bir veri sınıflandırma tekniğidir. Bir veri kümesindeki benzer değerleri istatistiksel özelliklerine göre bir dizi sınıf veya küme halinde gruplandırmak için kullanılır (Lu ve ark., 2021). Ayrıca, bu çalışmada adapte edilmesinin sebepleri ise birçok çalışmada yaygın olarak kullanılması (Tang ve ark., 2020; Saha ve ark., 2021), optimal sınıflandırma sağlama yeteneği (Hou ve ark., 2022), parametrik olmayan yapısı ve kullanım kolaylığıdır (Long ve ark., 2022).

Söz konusu beş kategoriye ve karşılık gelen alanları içeren Kentucky Nehri havzasına ait ani sel duyarlılık haritası Şekil 5' de sunulmuştur. Bu harita incelendiğinde farklı kategorilere bölünmüş duyarlılık alanlarının havza geneline homojen dağıldığı gözlemlenmektedir. Ayrıca, havzanın özellikle kuzeybatı ve güneye bakan orta kesimlerinde yüksek ani sel riski görülmektedir. Öte yandan, havzanın dik eğimlere sahip yüksek kesimlerinin yer aldığı güney kısmında ise ani sel riski oldukça düşük olarak gözlemlenmektedir. Ayrıca, havzanın kuzey doğuya bakan orta kesimlerinde ise yüksek eğri numarasına sahip kentsel bölgeler yer almasına rağmen, bu kesimlerin görece düşük ani sel tehlikesi ile karşı karşıya olduğu ifade edilebilmektedir. Bunun sebebi ise söz konusu bölgelerdeki yağış rejimleri ile açıklanabilmektedir. Öyle ki, bu bölgelerde şiddetli yağış dağılımlarına göre yüksek tekrür aralığına sahip yağışların seyrek olduğu gözlenmektedir. Özellikle şiddetli yağış faktörünün Şekil 6' da ifade edilen faktörlerin önem derecelerinin sıralanmasında en yüksek önem derecesine sahip olduğu düşünüldüğünde bu sonuç beklenebilmektedir. Bu çalışmada değişkenlerin önem sıralamasının belirlenmesi için ise RF algoritmasındaki her ağaçta bölme işlemlerinin verimliliğine dayan önem puanları hesaplanmaktadır. Bu işlem, tahmin edilen hedef değişkenin varyansındaki azalmanın ölçüsüne dayanmaktadır. Yani, bir değişkenin eklenmesiyle elde edilen azalmayı, o değişkenin eklenmediği durumda elde edilen azalmaya oranlayarak değişkenlerin önem puanları hesaplanıp, sonrasında bu önem puanları toplanarak değişkenlerin tahmin başarısındaki göreceli önemi belirlenmektedir (Breiman, 2001). Şekil 6' da sunulan önem dereceleri incelendiğinde en yüksek ikinci sırada yer alan faktör ise sayısal yükseklik verileri ile açıklanan topoğrafya kriteri ön plana çıkmaktadır. Yüksek kotlu bölgelerde ani sel tehlikesinin istisnai durumlar dışında diğer bölgelere göre daha düşük olduğu literatürde yer alan birçok çalışmada ifade edilmiştir (Abedi ve ark., 2021). Şekil 6 incelendiğinde şiddetli yağış ve sayısal yükseklik faktörlerini sırasıyla NDVI ve eğri numarası takip etmektedir. Bu iki faktörün önem dereceleri arasında oldukça yakın bir fark çıkması anlaşılabilir haldedir. Çünkü NDVI ve eğri numarası faktörleri havza karakteristikleri açısından benzer göstergelere işaret etmektedir (Darabi ve ark., 2021). Aralarında ters korelasyon olan bu iki faktöre göre yüksek NDVI değeri olan bölgelerde yeşil alan miktarının fazla olduğu, dolayısıyla karşılık gelen bölgelerin düşük eğri numarasına sahip olacağı ifade edilebilmektedir. Düşük eğri numarasına sahip bölgeler arazi kullanım yapısı bakımından daha çok dağlık/ormanlık bölgeleri göstermektedir (Norallahi ve Seyed Kaboli, 2021). Bu bölgelerde yağın öncelikle toprağın doymasını sağlayacak ve yer altı suyunun beslenmesine katkı sağlayacak, sonrasında ise toprak doygunluk derecesine eriştiği durumda yüzey akışına dönüşerek sel tehlikesi oluşturacaktır. Havzanın kuzey batısına bakan kesimlerde düşük NDVI değerleri gözlemlendiğinde bu bölgelerde ani sel tehlikesinin diğer faktörlerin (yüksek geçirimsizlik ve düşük eğim) de katkısıyla yüksek olması anlaşılabilir hale gelmektedir. Faktörlerin önem dereceleri sıralamasında yukarıda ifade edilen dört faktörü birbirleriyle ilişkili olan ve önem dereceleri açısından oldukça az bir fark gösteren akarsu yoğunluğu, akarsuya uzaklık ve TWI faktörleri takip etmektedir. Bu noktada, havzanın güneye bakan orta kesimlerinde yüksek akarsu yoğunluğu ve düşük akarsuya uzaklık değerleri yukarıda bahsedilen faktörlere sel tehlikesinin artması bakımından önemli katkı sağlarken (Arabameri ve ark., 2020), havzanın güney batısına bakan orta kesimlerinde havzanın çıkış noktasına da yakın olması sebebiyle meydana gelen yüksek TWI değerleri sel tehlikesinin yüksek çıkmasına önemli katkıyı gerçekleştirilmiştir. Faktörlerin önem dereceleri detaylı incelendiğinde ise litoloji faktörünün ani sel riski açısından diğer faktörlere göre çok daha az etkiye sahip olduğu ifade edilebilmektedir. Bu bağlamda, ilgili literatür incelendiğinde her ne kadar litoloji kriteri taşkın analizlerinde kullanılsa da bu kriter ana kayanın jeomekanik ve hidrolik özelliklerini doğrudan yansıttığı ve toprak örtüsünün özelliklerini etkilediği için (Catani ve ark., 2013) daha çok heyelan/toprak kayması afeti bakımından tetikleyici rol oynamaktadır (Rahmati ve ark., 2019; Tien Bui ve ark., 2019; Chen ve Li, 2020).



Şekil 5. Kentucky Nehri havzasına ait ani sel tehlikesi duyarlılık haritası



Şekil 6. Ani sel tehlikesinde etkili olan değişkenlerin önem dereceleri

Bu çalışmada litoloji kriterinin dahil edilmesinin nedeni ise ani yağışlara bağlı olarak meydana gelen toprak kopmaları beraberinde heyelan olaylarını ve bu olaylar neticesinde suyun çok hızlı bir şekilde yüksek kotlardan düşük kotlu bölgelere intikal etmesine sebebiyet verme ihtimali olarak gösterilebilmektedir. Tüm bu sonuçlar göz önünde bulundurulduğunda özellikle kara kutu modeller olarak ifade edilen makine öğrenmesi yöntemleri ile sebep sonuç ilişkisi açıklanabilir bir ağaç tabanlı model (RF) oluşturulmuş ve RF modelinin içerisinde barındırdığı belirsizlikler PSO algoritması ile minimize edilmiştir. Çalışma sonucunda ise hem Kentucky Nehri havzasının sınırları içerisinde yer alan ani sel tehlikesine sahip bölgeler belirlenmiş hem de bu bölgelerde yüksek sel tehlikesini tetikleyen faktörler ifade edilmiştir.

SONUÇ

Taşkınlar sadece gerçekleştiği bölgelerde önemli yıkımlara sebebiyet verip mal ve can kaybına neden olmamakta, aynı zamanda bu bölgelerde altyapıların hasar görmesine, halka yapılacak hizmetlerin aksamasına ve dahi su kaynaklı kirlenmenin neden olduğu hastalıklara sebebiyet vermektedir. Özellikle iklim değişikliği ile birlikte yağış rejimlerinde meydana gelen değişimlerin ve dünya genelinde plansız yapılaşmaya bağlı kentleşmenin de etkileri ile taşkın afetlerin gerek sayısı gerekse de tahrip boyutları

artmaktadır. Bu çalışma kapsamında da Amerika Birleşik Devletleri'nin Kentucky eyaleti sınırları içerisinde yer alan ve eyaletin önemli bir kısmını kapsayan Kentucky Nehri havzası için ani sellere ait duyarlılık haritalarının oluşturulması amaçlanmıştır. Gerçekleştirilecek kapsamlı analizler için ise veri tabanlı makine öğrenmesi algoritmalarından yorumlanabilirliği artırma yeteneğine de sahip ağaç tabanlı rastgele orman algoritması kullanılmıştır. Öte yandan, RF metodunun eğitilmesi esnasında parçacık sürü optimizasyonu da entegre edilerek öğrenme sistemi kuvvetlendirilmiş ve 5-fold çapraz doğrulama sayesinde aşırı öğrenmeye dayalı problemlerin önüne geçilmesi sağlanmıştır. Bağımsız olarak ayrılan test veri seti ile öncesinde eğitimi gerçekleştirilen melez PSO-RF tahmin çerçevesi test edilerek model performansları elde edilmiştir. Elde edilen sonuçlara göre önerilen melez model geçmişte meydana gelen ani sel kayıtları ışığında yüksek tahmin performans başarısı göstermiştir. Ek olarak ise, bölgede alınabilecek aksiyonların ve risk azaltma stratejilerinin belirlenmesi adına uygulanan önem dereceleri sıralamalarına göre ise, Kentucky Nehri havzası için ani selleri öncelikli olarak tetikleyen faktörler şiddetli yağışlar, topoğrafya, NDVI ve eğri numarası olarak belirlenmiştir. Kentucky Nehri havzasında son yirmi yılda gerçekleşen ve kayıt altına alınan 343 adet ani sel noktalarına ait bilgiler ışığında yapılan analizler ile ilgili literatüre önemli bir katkı gerçekleştirilmiştir. Ayrıca, analiz sonuçlarının bölgede görev yapan profesyonellere ve bu alanda benzer incelemeler yapacak araştırmacılara yol gösterici nitelikte olduğuna inanılmaktadır. Tüm bu yapılan analizlere rağmen, söz konusu çalışma halen birtakım sınırlamalar çerçevesinde gerçekleştirilmiştir. Örneğin, RF algoritması literatürde yer alan Adaboost, XGBoost gibi farklı ağaç tabanlı topluluk modelleri ile performans karşılaştırmasına tabi tutulup, sonuçların karşılaştırmalı gösterimi sağlanabilecektir. Ayrıca, bu çalışmada kullanılan PSO arama algoritmasına ilaveten genetik algoritma, gri kurt algoritması gibi farklı meta-sezgisel optimizasyon araçlarının da kullanılacak makine öğrenmesi yöntemlerine entegre edilmesi ve bunların performans karşılaştırılmasının yapılması gelecekteki çalışmalar için faydalı olabilecektir. Bu çalışmada kullanılan faktör önem ağırlıklarının incelenmesi metoduna ilaveten, faktörlerin ekstrem değerlerine dair davranışlarının sonuçlar üzerindeki etkilerinin incelenmesini sağlayan ve sonuçların yorumlanabilirliğine dair daha detaylı çıktılar sunan Shapley Additive Explanations gibi modelden bağımsız olarak çalışan farklı yaklaşımlar da sonraki çalışmalarda ilgili araştırmacılar tarafından entegre edilebilecektir. Kullanılan faktörler açısından ise örneğin, havzanın bağıl yükseklik farkını ortaya koyabilecek havzanın jeomorfolojik yaşına göre veya havzanın hipsometrik yükseklik eğimine göre sayısal/nicel analiz yapılması da sonraki çalışmalara ışık tutacaktır. Son olarak ise, bu çalışmada önerilen melez yaklaşımın hem Dünyada hem de Türkiye'de yer alan diğer havzalar özelinde de test edilebileceği ifade edilebilmektedir.

KAYNAKLAR

- Abedi, R., Costache, R., Shafizadeh-Moghadam, H., & Pham, Q. B. (2021). Flash-flood susceptibility mapping based on XGBoost, random forest and boosted regression trees. *Geocarto International*, 0(0), 1–18. <https://doi.org/10.1080/10106049.2021.1920636>
- Ali, S. A., Parvin, F., Pham, Q. B., Vojtek, M., Vojteková, J., Costache, R., ... Ghorbani, M. A. (2020). GIS-based comparative assessment of flood susceptibility mapping using hybrid multi-criteria decision-making approach, naïve Bayes tree, bivariate statistics and logistic regression: A case of Topľa basin, Slovakia. *Ecological Indicators*, 117(June), 106620. <https://doi.org/10.1016/j.ecolind.2020.106620>
- Andaryani, S., Nourani, V., Haghighi, A. T., & Keesstra, S. (2021). Integration of hard and soft supervised machine learning for flood susceptibility mapping. *Journal of Environmental Management*, 291(April), 112731. <https://doi.org/10.1016/j.jenvman.2021.112731>

- Arabameri, A., Saha, S., Mukherjee, K., Blaschke, T., Chen, W., Ngo, P. T. T., & Band, S. S. (2020). Modeling spatial flood using novel ensemble artificial intelligence approaches in northern Iran. *Remote Sensing*, 12(20), 1–30. <https://doi.org/10.3390/rs12203423>
- Arora, A., Arabameri, A., Pandey, M., Siddiqui, M. A., Shukla, U. K., Bui, D. T., ... Bhardwaj, A. (2021). Optimization of state-of-the-art fuzzy-metaheuristic ANFIS-based machine learning models for flood susceptibility prediction mapping in the Middle Ganga Plain, India. *Science of the Total Environment*, 750, 141565. <https://doi.org/10.1016/j.scitotenv.2020.141565>
- Breiman, L. (2001). Random forests. *Machine Learning*, 45, 5–32. <https://doi.org/10.1023/A:1010933404324>
- Catani, F., Lagomarsino, D., Segoni, S., & Tofani, V. (2013). Landslide susceptibility estimation by random forests technique: Sensitivity and scaling issues. *Natural Hazards and Earth System Sciences*, 13(11), 2815–2831. <https://doi.org/10.5194/nhess-13-2815-2013>
- Chen, W., & Li, Y. (2020). GIS-based evaluation of landslide susceptibility using hybrid computational intelligence models. *Catena*, 195(December 2019), 104777. <https://doi.org/10.1016/j.catena.2020.104777>
- Costache, R., Hong, H., & Pham, Q. B. (2020). Comparative assessment of the flash-flood potential within small mountain catchments using bivariate statistics and their novel hybrid integration with machine learning models. *Science of the Total Environment*, 711, 134514. <https://doi.org/10.1016/j.scitotenv.2019.134514>
- Costache, R., Pham, Q. B., Sharifi, E., Linh, N. T. T., Abba, S. I., Vojtek, M., ... Khoi, D. N. (2020). Flash-flood susceptibility assessment using multi-criteria decision making and machine learning supported by remote sensing and GIS techniques. *Remote Sensing*, 12(1). <https://doi.org/10.3390/RS12010106>
- Costache, R., & Tien Bui, D. (2019). Spatial prediction of flood potential using new ensembles of bivariate statistics and artificial intelligence: A case study at the Putna river catchment of Romania. *Science of the Total Environment*, 691, 1098–1118. <https://doi.org/10.1016/j.scitotenv.2019.07.197>
- Darabi, H., Torabi Haghighi, A., Rahmati, O., Jalali Shahrood, A., Rouzbeh, S., Pradhan, B., & Tien Bui, D. (2021). A hybridized model based on neural network and swarm intelligence-grey wolf algorithm for spatial prediction of urban flood-inundation. *Journal of Hydrology*, 603(PA), 126854. <https://doi.org/10.1016/j.jhydrol.2021.126854>
- Ekmekcioğlu, Ö., Başakın, E. E., & Özger, M. (2020). Tree-based nonlinear ensemble technique to predict energy dissipation in stepped spillways. *European Journal of Environmental and Civil Engineering*, 0(0), 1–19. <https://doi.org/10.1080/19648189.2020.1805024>
- Ekmekcioğlu, Ö., Koc, K., & Özger, M. (2021). Stakeholder perceptions in flood risk assessment: A hybrid fuzzy AHP-TOPSIS approach for Istanbul, Turkey. *International Journal of Disaster Risk Reduction*, 60(May). <https://doi.org/10.1016/j.ijdrr.2021.102327>
- Ekmekcioğlu, Ö., & Koc, K. (2022). Explainable step-wise binary classification for the susceptibility assessment of geo-hydrological hazards. *CATENA*, 216, 106379. <https://doi.org/10.1016/j.catena.2022.106379>
- Ekmekcioğlu, Ö., Koc, K., Özger, M., & Işık, Z. (2022). Exploring the additional value of class imbalance distributions on interpretable flash flood susceptibility prediction in the Black Warrior River basin, Alabama, United States. *Journal of Hydrology*, 610, 127877. <https://doi.org/10.1016/j.jhydrol.2022.127877>
- Fang, Z., Wang, Y., Peng, L., & Hong, H. (2020). Integration of convolutional neural network and conventional machine learning classifiers for landslide susceptibility mapping. *Computers and Geosciences*, 139(February), 104470. <https://doi.org/10.1016/j.cageo.2020.104470>
- Gigović, L., Pamučar, D., Bajić, Z., & Drobnjak, S. (2017). Application of GIS-interval rough AHP methodology for flood hazard mapping in Urban areas. *Water (Switzerland)*, 9(6), 1–26. <https://doi.org/10.3390/w9060360>
- Goswami, S., Murthy, C. A., & Das, A. K. (2018). Sparsity measure of a network graph: Gini index. *Information Sciences*, 462, 16–39. <https://doi.org/10.1016/j.ins.2018.05.044>
- Habba, M., Ameer, M., & Jabrane, Y. (2018). A novel Gini index based evaluation criterion for image segmentation. *Optik*, 168, 446–457. <https://doi.org/10.1016/j.ijleo.2018.04.045>

- Hou, C., Xie, Y., & Zhang, Z. (2022). An improved convolutional neural network based indoor localization by using Jenks natural breaks algorithm. *China Communications*, 19(4), 291–301. <https://doi.org/10.23919/JCC.2022.04.021>
- Ikeuchi, H., Hirabayashi, Y., Yamazaki, D., Muis, S., Ward, P. J., Winsemius, H. C., ... Kanae, S. (2017). Compound simulation of fluvial floods and storm surges in a global coupled river-coast flood model: Model development and its application to 2007 Cyclone Sidr in Bangladesh. *Journal of Advances in Modeling Earth Systems*, 9(4), 1847–1862. <https://doi.org/10.1002/2017MS000943>
- Jaafar, H. H., Ahmad, F. A., & El Beyrouthy, N. (2019). GCN250, new global gridded curve numbers for hydrologic modeling and design. *Scientific Data*, 6(1), 1–9. <https://doi.org/10.1038/s41597-019-0155-x>
- Liu, X., Zhang, Z., Jiang, T., Li, X., & Li, Y. (2021). Evaluation of the Effectiveness of Multiple Machine Learning Methods in Remote Sensing Quantitative Retrieval of Suspended Matter Concentrations: A Case Study of Nansi Lake in North China. *Journal of Spectroscopy*, 2021, 1–17. <https://doi.org/10.1155/2021/5957376>
- Long, Y., Song, Y., & Chen, L. (2022). Identifying subcenters with a nonparametric method and ubiquitous point-of-interest data: A case study of 284 Chinese cities. *Environment and Planning B: Urban Analytics and City Science*, 49(1), 58–75. <https://doi.org/10.1177/2399808321996705>
- Lu, Y., He, T., Xu, X., & Qiao, Z. (2021). Investigation the Robustness of Standard Classification Methods for Defining Urban Heat Islands. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14, 11386–11394. <https://doi.org/10.1109/JSTARS.2021.3124558>
- Kennedy, J., & Eberhart, R. (1995). Particle swarm optimization. *Proceedings of ICNN'95 - International Conference on Neural Networks*, 4, 1942–1948. IEEE. <https://doi.org/10.1109/ICNN.1995.488968>
- Khosravi, K., Pham, B. T., Chapi, K., Shirzadi, A., Shahabi, H., Revhaug, I., ... Tien Bui, D. (2018). A comparative assessment of decision trees algorithms for flash flood susceptibility modeling at Haraz watershed, northern Iran. *Science of the Total Environment*, 627, 744–755. <https://doi.org/10.1016/j.scitotenv.2018.01.266>
- Kim, J., Jeong, S., & Regueiro, R. A. (2012). Instability of partially saturated soil slopes due to alteration of rainfall pattern. *Engineering Geology*, 147–148, 28–36. <https://doi.org/10.1016/j.enggeo.2012.07.005>
- Koc, K., Ekmeçcioğlu, Ö., & Özger, M. (2021). An integrated framework for the comprehensive evaluation of low impact development strategies. *Journal of Environmental Management*, 294, 113023. <https://doi.org/10.1016/j.jenvman.2021.113023>
- Marchi, L., Borga, M., Preciso, E., & Gaume, E. (2010). Characterisation of selected extreme flash floods in Europe and implications for flood risk management. *Journal of Hydrology*, 394(1–2), 118–133. <https://doi.org/10.1016/j.jhydrol.2010.07.017>
- NCEI. (2021). NCEI. Erişim adresi: <https://www.ncei.noaa.gov/> (Erişim tarihi: 10 Kasım, 2022)
- Ngo, P. T. T., Hoang, N. D., Pradhan, B., Nguyen, Q. K., Tran, X. T., Nguyen, Q. M., ... Bui, D. T. (2018). A novel hybrid swarm optimized multilayer neural network for spatial prediction of flash floods in tropical areas using sentinel-1 SAR imagery and geospatial data. *Sensors (Switzerland)*, 18(11). <https://doi.org/10.3390/s18113704>
- Nhu, V. H., Ngo, P. T. T., Pham, T. D., Dou, J., Song, X., Hoang, N. D., ... Bui, D. T. (2020). A new hybrid firefly-pso optimized random subspace tree intelligence for torrential rainfall-induced flash flood susceptible mapping. *Remote Sensing*, 12(17), 1–19. <https://doi.org/10.3390/RS12172688>
- Norollahi, M., & Seyed Kaboli, H. (2021). Urban flood hazard mapping using machine learning models: GARP, RF, MaxEnt and NB. *Natural Hazards*, 106(1), 119–137. <https://doi.org/10.1007/s11069-020-04453-3>
- Pham, B. T., Luu, C., Phong, T. Van, Trinh, P. T., Shirzadi, A., Renoud, S., ... Clague, J. J. (2021). Can deep learning algorithms outperform benchmark machine learning algorithms in flood susceptibility modeling? *Journal of Hydrology*, 592(July 2020), 125615. <https://doi.org/10.1016/j.jhydrol.2020.125615>
- Panahi, M., Jaafari, A., Shirzadi, A., Shahabi, H., Rahmati, O., Omidvar, E., ... Bui, D. T. (2021). Deep learning neural networks for spatially explicit prediction of flash flood probability. *Geoscience Frontiers*, 12(3), 101076. <https://doi.org/10.1016/j.gsf.2020.09.007>

- Pourghasemi, H. R., Gayen, A., Edalat, M., Zarafshar, M., & Tiefenbacher, J. P. (2020). Is multi-hazard mapping effective in assessing natural hazards and integrated watershed management? *Geoscience Frontiers*, 11(4), 1203–1217. <https://doi.org/10.1016/j.gsf.2019.10.008>
- Rahmati, O., Falah, F., Naghibi, S. A., Biggs, T., Soltani, M., Deo, R. C., ... Tien Bui, D. (2019). Land subsidence modelling using tree-based machine learning algorithms. *Science of the Total Environment*, 672, 239–252. <https://doi.org/10.1016/j.scitotenv.2019.03.496>
- Riley, S. J., De Gloria, S. D., & Elliot, R. (1999). A Terrain Ruggedness that Quantifies Topographic Heterogeneity. *Intermountain Journal of Sciences*, 5(1–4), 23–27.
- Saha, S., Sarkar, R., Thapa, G., & Roy, J. (2021). Modeling gully erosion susceptibility in Phuentsholing, Bhutan using deep learning and basic machine learning algorithms. *Environmental Earth Sciences*, 80(8), 1–21. <https://doi.org/10.1007/s12665-021-09599-2>
- Shahabi, H., Shirzadi, A., Ronoud, S., Asadi, S., Pham, B. T., Mansouripour, F., ... Bui, D. T. (2021). Flash flood susceptibility mapping using a novel deep learning model based on deep belief network, back propagation and genetic algorithm. *Geoscience Frontiers*, 12(3), 101100. <https://doi.org/10.1016/j.gsf.2020.10.007>
- Talukdar, S., Ghose, B., Shahfahad, Salam, R., Mahato, S., Pham, Q. B., ... Avand, M. (2020). Flood susceptibility modeling in Teesta River basin, Bangladesh using novel ensembles of bagging algorithms. *Stochastic Environmental Research and Risk Assessment*, 34(12), 2277–2300. <https://doi.org/10.1007/s00477-020-01862-5>
- Tang, X., Li, J., Liu, M., Liu, W., & Hong, H. (2020). Flood susceptibility assessment based on a novel random Naïve Bayes method: A comparison between different factor discretization methods. *Catena*, 190(March), 104536. <https://doi.org/10.1016/j.catena.2020.104536>
- Thieken, A. H., Petrow, T., Kreibich, H., & Merz, B. (2006). Insurability and Mitigation of Flood Losses in Private Households in Germany. *Risk Analysis*, 26(2), 383–395. <https://doi.org/10.1111/j.1539-6924.2006.00741.x>
- Tien Bui, D., Hoang, N. D., Pham, T. D., Ngo, P. T. T., Hoa, P. V., Minh, N. Q., ... Samui, P. (2019). A new intelligence approach based on GIS-based Multivariate Adaptive Regression Splines and metaheuristic optimization for predicting flash flood susceptible areas at high-frequency tropical typhoon area. *Journal of Hydrology*, 575(April), 314–326. <https://doi.org/10.1016/j.jhydrol.2019.05.046>
- Tien Bui, D., Hoang, N. D., Martínez-Álvarez, F., Ngo, P. T. T., Hoa, P. V., Pham, T. D., ... Costache, R. (2020). A novel deep learning neural network approach for predicting flash flood susceptibility: A case study at a high frequency tropical storm area. *Science of the Total Environment*, 701, 134413. <https://doi.org/10.1016/j.scitotenv.2019.134413>
- Xu, H., Fan, G., & Song, Y. (2022). Novel Key Indicators Selection Method of Financial Fraud Prediction Model Based on Machine Learning Hybrid Mode. *Mobile Information Systems*, 2022, 1–12. <https://doi.org/10.1155/2022/6542652>