

İstanbul İçin CBS Tabanlı Makine Öğrenmesi İle Sel Duyarlılık Haritasının Oluşturulması

Zehra Koyuncu^{1,*}, Ömer Ekmekcioğlu²

¹İnşaat Mühendisliği Bölümü, Mühendislik ve Doğa Bilimleri Fakültesi, İstanbul Medipol Üniversitesi, İstanbul.

²Afet ve Acil Durum Yönetimi Ana Bilim Dalı, Afet Yönetim Enstitüsü, İstanbul Teknik Üniversitesi, İstanbul.

Özet

Bu çalışma kapsamında meydana gelebilecek olası bir sel olayının gerçekleşebileceği yerin önceden tahmini ve tespiti için makine öğrenmesi yöntemleri kullanılarak coğrafi bilgi sistemleri (CBS) tabanlı bir sel duyarlılık haritalama modeli oluşturulması amaçlanmıştır. Çalışma kapsamında incelen bölge olarak ise Türkiye'nin metropol kenti olan İstanbul ili seçilmiştir. Literatürden elde edilen sel envanteriyle oluşturulan örneklem kümesi önce sel olmayan noktaların rastgele oluşturulması ile genişletilmiş olup, ardından sınıf dengesizliği rastgele alt örnekleme (RUS) tekniği ile giderilmiştir. Bu yaklaşım Türkiye'de gerçekleştirilen sel duyarlılık haritalamaları çalışmaları için ilk kez uygulanmıştır. Rastgele orman (RF), stokastik gradyan artırma (SGB) ve XGBoost algoritmaları olmak üzere üç farklı makine öğrenmesi algoritmasının performans karşılaştırmaları gerçekleştirilmiştir. En yüksek model performansının XGBoost ile elde edildiği, bu metodu ise sırasıyla SGB ve RF'nin takip ettiği sonucuna ulaşılmıştır. Ayrıca, RF ve SGB modellerinin sel olmayan noktaların neredeyse tamamını doğru olarak bulduğu, sel olan noktalarda ise %90.67'lik bir başarı sergilediği görülmüştür. Fakat, çalışmanın esas amacını kapsayan sel gerçekleşen noktaların belirlenmesinde XGBoost modeli %92.00'lik bir başarı ile diğer iki metoda üstünlük sergilediği tespit edilmiştir. Sel olayını etkileyen parametreler incelendiğinde ise İstanbul için seli en önemli parametrenin yağış olduğu sonucuna ulaşılmış olup, yağışı sırasıyla drenaj ağına uzaklık ve eğri numarası takip etmiştir. Sonuç olarak çalışma kapsamında İstanbul'da gerçekleştirilen sel duyarlılık haritalamaları çalışmaları için ilk kez uygulanan bu çerçevenin kullanımının sayısı ve etkileri giderek artırılarak sel olaylarına karşı daha yaygın alanlara uygulanması gelecek vadeci bir yaklaşım olacaktır.

Anahtar Sözcükler

Coğrafi Bilgi Sistemleri, İstanbul, Makine Öğrenmesi, Risk Yönetimi, Sel Duyarlılık Haritalaması, Uzaktan Algılama

Generating the Flood Susceptibility Map for Istanbul with GIS-Based Machine Learning Algorithms

Abstract

The main objective of the current study is to generate a flood hazard map by using the machine learning algorithms hybridized with the geographic information systems (GIS). In this regard, the province of Istanbul, which is the metropolitan city of Turkey, was selected as the focal region within the scope of the study. The class imbalance was tackled through the commonly used random under sampling (RUS) technique in order to create a fair comparison datum line. It is worth mentioning that this is the first time this approach has been used for flood hazard mapping studies in Turkey. Random forest (RF), stochastic gradient boosting (SGB), and XGBoost algorithms were used. The best predictive performance was obtained with the XGBoost algorithm, followed by SGB and RF, respectively. The RF and SGB models showed a 90.67% success rate in determining the inundation points, while the XGBoost model outperformed its counterparts with a 92.00% success rate in determining the inundation points. In this research, the importance levels of the flood triggering variables were further investigated in order to enliven the comprehensibility of the obtained results. Thus, the most important variable was the precipitation, followed by the distance to the drainage network and the number of curves, respectively. Finally, it is suggested that flood vulnerability mapping attempts can be considered as promising approaches against increasing flood incidents over the years.

Keywords

Geographical Information Systems, Istanbul, Machine Learning, Risk Management, Flood Hazard Mapping, Remote Sensing

1. Giriş

Canlı ve cansız bütün ekosistemler için çok önemli bir madde olan suyun faydaları yanında zararlı etkileri de bulunmaktadır. Bu zararlara doğa kaynaklı afetler olarak da ifade edilen fakat insan etkisinin de ciddi derecede olduğu seller örnek olarak verilebilmektedir. Sel afeti geçmişten günümüze kadar özellikle iklim değişikliğinin de etkileri ile hem sayısal hem de şiddet açısından artış göstermekte olan, etkisi uzun süreler devam eden ve zararlı etkisi oldukça fazla olan

* Sorumlu Yazar: Tel: +90 (212) 4448544 Faks: +90 (212) 5212377

E-posta: zehra.akti@std.medipol.edu.tr (Zehra K), omer.ekmekcioglu@itu.edu.tr (Ömer E)

Gönderim Tarihi / Received : 22/02/2023

Kabul Tarihi / Accepted : 04/09/2023

hidro-meteorolojik karakterli bir doğal afettir (Chakraborty vd., 2021). Sel afeti geçmişte ve günümüzde sadece can ve mal kaybına zarar vermek ile kalmayarak insan yaşamını da olumsuz yönde etkileyerek sosyal yaşamda aksamalara sebep olmuştur. Zira, Anılan vd. (2021) Trabzon'un Beşikdüzü ilçesinde meydana gelen taşkın afetinin hidrolojik ve sosyal yönden etkileri taşkın frekans analizi, anketler ve çok kriterli karar destek sistemleri ile incelenerek, erken uyarı sistemlerinin geliştirilmesi için çeşitli önerilerde bulunulmuştur. Bu çalışma, sel afetinin zararlı etkilerini en az seviyeye indirerek, meydana gelebilecek maddi ve manevi hasarların da önüne geçilebileceğini öne sürmektedir. Bunu desteklemek adına, sel sonrası yaşanabilecek ekonomik kayıpların ön görülebilmesini sağlayan çalışmalar da etkililiğini sürdürmektedir. Örneğin, Demir ve Keskin (2022a), Orta Karadeniz Bölgesi'nde yer alan Mert Irmağı Havzası için iki boyutlu bir taşkın modellemesi gerçekleştirilerek havza geneli için olası bir sel durumunda yaşanabilecek ekonomik kayıpları öngörmüşlerdir. Günümüz dünyasının güncel problemlerinden olan yoğun nüfus artışı, aşırı kentleşme ve yanlış arazi kullanımı, atmosfere sera gazı salınımını artırmış ve iklim değişimini tetikleyecek mertebede atmosferin kimyasal yapısının bozulmasına ve kentlerde ısı adası etkisinin artmasına neden olmuştur. Teknolojik gelişmelerin ve bilgisayar kapasitelerinin artışı, sel duyarlılık haritalarının ve sel modellerinin hazırlanmasında uzaktan algılama verileri ve coğrafi bilgi sistemlerinden (CBS) daha çok yararlanılmasına katkı sağlamıştır (Güvel vd., 2023). Bu tekniklerden yararlanarak sel duyarlılık haritalama çalışmalarında başta meteorolojik ve hidrolojik gözlemlere (Demir & Keskin, 2022b), çalışma sahasına ait topoğrafik ve arazi kullanımı verilerine gereksinim vardır (Şen, 2009). Sel oluşumunu etkileyen bu faktörler ile geçmişte taşkın olaylarının gözlemlendiği noktalar bir arada ele alınarak herhangi bir çalışma bölgesinin taşkın haritasının CBS teknikleri ile üretilmesi olanaklıdır. Sel duyarlılık haritalarının üretiminde çeşitli teknikler kullanılmıştır. Bu tekniklere örnek olarak istatistiksel yöntemler (frekans oranı, kanıt ağırlıkları vb.), çok kriterli karar destek sistemleri (analitik hiyerarşi süreci vb.) ve makine öğrenmesi yöntemleri verilebilir (Özcan, 2017; Demir & Keskin, 2022c).

Sel duyarlılık haritalaması çalışmalarında son yıllarda makine öğrenmesi yöntemleri oldukça yaygın bir şekilde kullanılmaktadır. Örneğin, Lojistik Regresyon (LR), Destek Vektör Makineleri (SVM), Yapay Sinir Ağları (ANN) gibi makine öğrenmesi yaklaşımları sıklıkla kullanılmıştır (Tona vd., 2022). Tehrany vd. (2014) Malezya'nın Terengganu kenti için kanıt ağırlığı yöntemine ek olarak SVM algoritmasını da kullanarak sel duyarlılık haritası üretmişlerdir. SVM algoritmasının farklı çekirdek türleri ile eğitildiği bu çalışma ile taşkın duyarlılık haritalarında test seti için %88.64 doğruluk elde edilmiştir. Ghosh vd. (2022), Ganj Nehri Havzası için 128 sel kaydını kullanarak bir sel duyarlılık haritası hazırlamışlardır. Bu çalışmada frekans oranı yöntemi ile LR algoritmasının performansları denetlenmiş olup, sırasıyla %78 ve %80 oranında doğruluk elde edilmiştir. Bu sayede, makine öğrenmesi algoritmalarının istatistiksel yaklaşımlara göre daha yüksek doğruluk sunabildiği gösterilmiştir.

Son yıllarda, tekil makine öğrenmesi yöntemlerinin yanı sıra karar ağacı tabanlı topluluk (İng. ensemble) algoritmaları sel duyarlılık haritalarının üretilmesi için sıklıkla kullanılmaktadır. Topluluk algoritması, eğitim örnekleri üzerinde bir dizi karar ağacı üretip ve bir önceki karar ağacına göre performansını ardışık bir şekilde yükseltiyorsa, buna 'gradyan artırma' yöntemi (İng. boosting) denir. Gradyan artırma yöntemlerinin çeşitli varyasyonları bulunmaktadır. Topluluk algoritması, eğitim örnekleri üzerinde aynı anda birden fazla karar ağacı üretip, bu karar ağaçlarının bir topluluğunu (orman) oluşturarak en iyi performans gösteren karar ağacını tespit etmek üzerine de çalışabilmektedir. Bu yöntem ise 'torbalama' adı verilir ve Rastgele Orman (RF) algoritması en yaygın torbalama algoritmasıdır. Towfiqul Islam vd. (2021) Bangladeş'in kuzey bölgesindeki Teesta Nehri Havzasında sel duyarlılık haritasını modellemek için dagging topluluk modeli, rastgele alt uzay metodu, ANN, RF ve SVM algoritmalarını bir arada kullanmıştır. Elde edilen sonuçlara göre dagging topluluk modeli 0.873 AUC değeri ile doğrulama setinde en iyi sonucu verirken, onu 0.865, 0.839, 0.835 ve 0.811 ile sırasıyla SVM, rastgele alt uzay, ANN ve RF algoritmaları takip etmiştir. Deroliya vd. (2022) Hindistan'ın alt Mahanadi Nehri havzasında yer alan Jagatsinghpur'da sel duyarlılığını belirlemek için tekil karar ağacını, RF ve gradyan artırma algoritmalarını kullanmışlardır. Analizler neticesinde, RF algoritmasının diğer iki yöntemle göre daha iyi performans gösterdiğine işaret etmişlerdir.

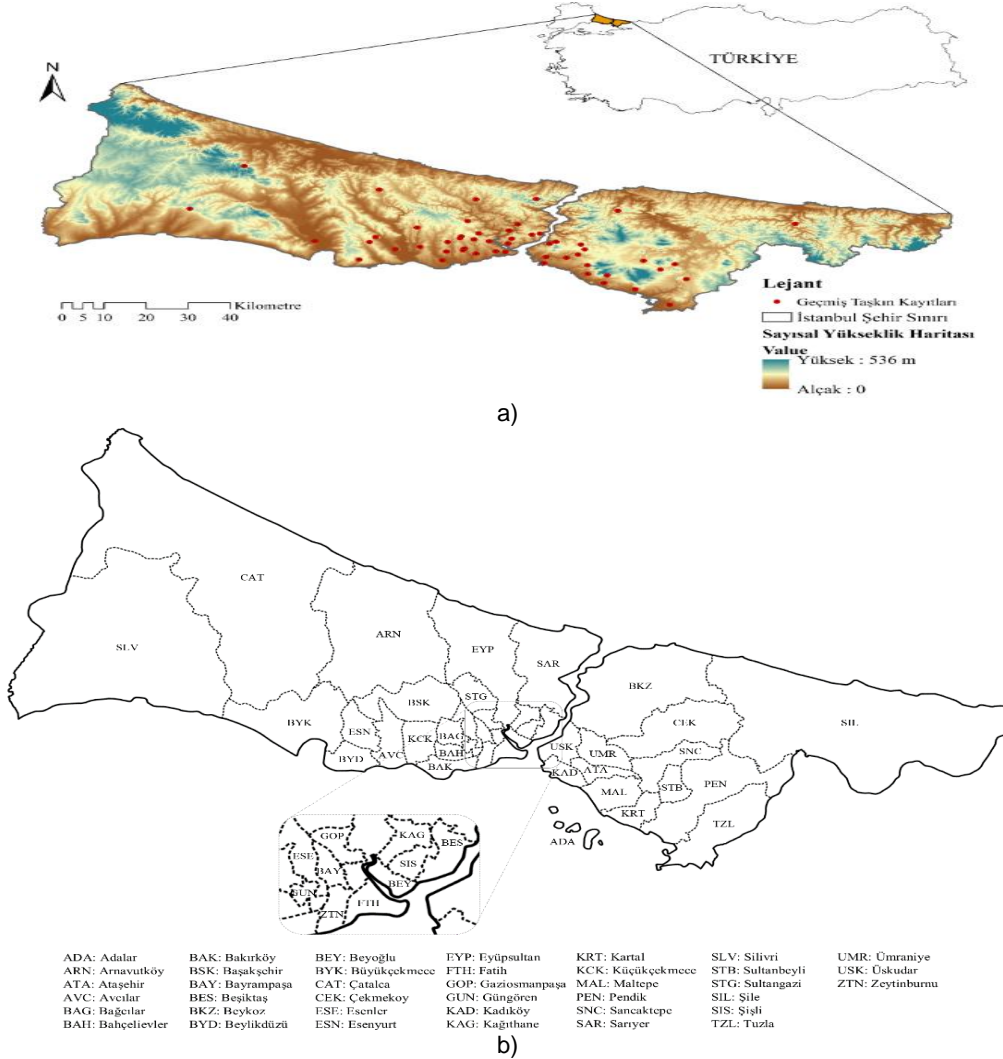
Ülkemizde ise benzer çalışmalar araştırmacılar tarafından gerçekleştirilse de sel duyarlılık haritalamalarında makine öğrenmesi algoritmalarının kullanımı oldukça sınırlıdır. Fakat, Aydın vd. (2022) bu konu üzerinde durarak Türkiye'nin Akdeniz kıyısındaki Adana ili için sel duyarlılık haritalamalarının geliştirilmesi amacıyla makine öğrenmesi algoritmalarını kullanmıştır. İncelenen bu çalışma kapsamında beş kategori altında toplam on dört taşkın koşullandırma parametresi kullanılmış, AdaBoost ve LightGBM modelleri test doğruluğu açısından en yüksek başarıyı sağlamış olup (0.8978), onu sırasıyla GB ve NGBoost (0.8832), XGBoost (0.8759), RF (0.8613) ve CatBoost (0.8102) takip etmiştir.

Öte yandan, bazı bölgelerde sel olaylarının önlenmesi için sel risk yönetimi çalışmaları yapılarak çok kriterli karar verme yöntemleri (ÇKKV) yardımıyla bu çalışmanın da odak noktasını teşkil eden İstanbul ili için sel risk haritaları oluşturulmuştur. Örneğin, Ekmekcioğlu vd. (2022a) İstanbul'un ilçelerini sel riski açısından önceliklendirmek için çok adımlı bir ÇKKV metodolojisi sunmuş ve böylece sel riskinin azaltılmasına yönelik karar verme sürecine katkıda bulunmak için kapsamlı bir sel risk haritası oluşturmuştur. Bu kapsamda, tehlike kriterlerini değerlendirmek için analitik hiyerarşi süreci (AHP) yöntemi benimsenirken, kırılganlık kümesi DEMATEL-ANP yöntemine tabi tutulmuştur. VIKOR analizini gerçekleştirmek için her bir ilçeye her bir kritere göre atfedilen veriler bu yöntemlerle hesaplanan kriter ağırlıkları ile birleştirilmiştir. Benimsenen metodolojinin sağlamlığını ve araştırma bulgularının istikrarını sağlamak için ilçe bazlı sel risk analizi yapılmıştır. İstanbul'un sel risk haritası on kritere göre oluşturulmuş ve sonuçlar İstanbul'un en yoğun nüfuslu ve kentleşmiş bölgelerinin önemli ölçüde sel riski altında olduğunu göstermiştir.

Bu çalışma kapsamında ise uzaktan algılama yöntemleri ve CBS tabanlı makine öğrenmesi yöntemi kullanılarak İstanbul bölgesi için sel duyarlılık haritalarının elde edilmesi amaçlanmıştır. Literatür incelendiğinde daha önceden İstanbul bölgesi için yapılmış makine öğrenmesine dayalı sel duyarlılığı çalışmalarının olmadığı, diğer alanlar için ise sel duyarlılığı özelinde yapılan çalışmaların havza bazlı yapıldığı için literatürde bu konu özelinde eksikliklerin olduğu, bu çalışma kapsamında ise şehir özelinde çalışılarak yenilikçi bir konu olarak literatüre katkısı olacağı ifade edilebilmektedir.

2. Çalışma Alanı

Çalışma kapsamında İstanbul ili çalışma alanı olarak seçilmiştir. İstanbul, Türkiye'nin kuzeybatısında yer alan ve 5461 km² yüzölçümüne sahip metropol bir şehirdir. Şehrin kuzeyinde Karadeniz'e ve güneyinde ise Marmara Denizi'ne kıyıları mevcuttur. Ayrıca şehir, İstanbul Boğazı tarafından ikiye ayrılan Avrupa ve Asya yakaları olmak üzere iki kıtadan oluşmaktadır (Şekil 1). İstanbul, 14'ü Asya'da, 25'i Avrupa'da olmak üzere 39 ilçeden oluşmaktadır. Ayrıca, şehir nüfusunun %65'i Avrupa yakasında bulunurken, geri kalan %35'i Asya yakasında yer almaktadır. Şehrin en yüksek noktası Asya yakasında yer alan 537 m yüksekliğindeki Aydos Tepesi'dir. İstanbul'daki orman arazilerinin yüzölçümü 535.250 hektardır. Ancak şehir metropoliten bir alan özelliği taşıdığı için ormanların şehir içindeki dağılımı düzensizdir. İstanbul'un iklim özellikleri Karadeniz ve Akdeniz iklimleri arasındaki geçiş özelliği gösterdiğinden dolayı ılıman iklim özelliği taşımaktadır. İstanbul yazları sıcak ve nemli iken, kışları genellikle soğuk, yağışlı ve bazen de karlı olmaktadır (Nefeslioglu vd., 2010). Kış aylarında aylık ortalama sıcaklıklar 6 ile 9 °C civarındayken, yaz aylarında sıcaklık değerleri yaklaşık 22 °C ve 24 °C civarını bulmaktadır. En sıcak aylar ortalama 29.5 °C ile Temmuz ve Ağustos ayları iken en soğuk aylar 4 °C ile Ocak ve Şubat aylarıdır (Meteoroloji Genel Müdürlüğü, 2023). Ayrıca İstanbul'un yıllık toplam yağış miktarı 843.9 mm'dir ve bu yağışlar yıl boyunca meydana gelmektedir. İstanbul'da yağışların %38'i kış mevsiminde, %18'i ilkbahar mevsiminde, %13'i yaz mevsiminde ve %31'i sonbahar mevsiminde gerçekleşmektedir.



Şekil 1: Çalışma alanı: a) sayısal yükseklik haritası, b) ilçeler

3. Materyal

Bu çalışma kapsamında meydana gelebilecek olası bir sel olayının gerçekleşebileceği yerin önceden tahmini ve tespiti için makine öğrenmesi yöntemleri ve sel envanterleri kullanılarak İstanbul ili için makine öğrenmesine dayalı coğrafi bilgi sistemleri (CBS) tabanlı bir sel duyarlılık haritalama modeli oluşturulması amaçlanmıştır. Bu bağlamda, üç farklı makine öğrenmesi metodu bir dizi ön işleme adımından geçirilen ham veri setine uygulanarak tahmin modellemeleri gerçekleştirilmiştir. Sellerin oluşumunu tetikleyen faktörlere ait geçmiş literatürden elde edilen veriler rastgele alt ölçeklemeye tabi tutularak gerçek dünya koşullarının analizlere yansıtılması sağlanmıştır. Verinin eğitim ve test setlerine ayrılmasında ise literatürde sıklıkla kullanılan 70/30 stratejisi benimsenmiştir. İstanbul'a ait topoğrafya verisi elde edilerek, bu topoğrafya verisine dayanan ve sel görüngüsünün oluşumuna doğrudan etki eden parametreler belirlenmek suretiyle bu parametrelere ait haritalar çıkarılmıştır.

3.1. Sellerin Oluşumunu Tetikleyen Faktörler

Bu çalışma kapsamında İstanbul İli sel duyarlılık haritalamasının makine öğrenmesi yöntemleri ile gerçekleştirilmesi adına sel fenomenini oluşmasında etkili olan 11 farklı parametre kullanılmıştır. Bu parametreler (1) yükseklik, (2) bakı, (3) plan eğriliği, (4) profil eğriliği, (5) drenaj yoğunluğu, (6) drenaj ağına uzaklık, (7) eğri numarası, (8) yağış, (9) topoğrafik ıslaklık indeksi (TWI), (10) topoğrafik pürüzlülük indeksi (TRI) ve (11) akış gücü indeksidir (SPI). *Sayısal yükseklik* haritaları, çalışılan bölgede yağış sonrasında ortaya çıkan akış ve sellerin belirlenmesinde önemli bir role sahiptir. Literatür incelendiğinde akarsu akımını etkileyen fizyografik bir değişken olan yüksekliğin fazla olduğu bölgelerde sel doğal afetinin meydana gelme sıklığının düşük rakımlı bölgelere göre daha fazla olduğu söylenebilmektedir (*Ekmekcioğlu vd., 2022b*). *Bakı* faktörü, güneşten alınan ısı enerjisi miktarına etki ettiği için terleme ve buharlaşma ile su kaybını etkilemektedir. Bölgesel olarak farklılık gösterebilmekle birlikte, bakı genel olarak sel riski ile ters orantılı bir özelliğe sahip parametredir. Zira, eğim yönü güneş radyasyonunun dağılımını ve böylece buharlaşma oranını etkileyebilmektedir. Dolayısıyla, taşkın duyarlılığı üzerinde önemli etkisi olan sızmayı da etkileyebilmektedir. Güneye bakan yamaçların daha fazla güneş ışınına maruz kalmasından dolayı buharlaşmalar artacak ve bu gibi bölgelerde yağış miktarı fazla olsa bile sel riskleri azalabilecektir (*Abedi vd., 2021*). Suyun hangi yöne doğru akış göstereceğini belirleyen eğrilik parametresi de seli tetikleyen diğer önemli parametrelerden bir tanesidir. Bu çalışma kapsamında hem akışın hızlanmasını ve yavaşlamasını kontrol etmek suretiyle erozyon ve çökelmeyi etkileyen *profil eğriliği* hem de akışın yakınsamasını ve sapmasını kontrol eden *plan eğriliği* girdiler arasında değerlendirilmiştir. *Drenaj yoğunluğu*, birim alana düşen akarsu uzunluğu olarak ifade edilmektedir. Drenaj yoğunluğu, havzadaki bütün derelerin uzunlukları toplamının havza alanına oranıyla hesaplanmaktadır. Havzanın drenaj yoğunluğu ne kadar fazla ise, o havzadaki drenaj ağının yağış sularının çevreye zarar vermeden akıtma yeteneğinin de o ölçüde fazla olduğu söylenebilmektedir (*Hitouri vd., 2022*). Selleri tetikleyen parametrelerden bir tanesi olan *drenaj ağına uzaklık* faktörü incelendiğinde akarsuya uzaklığın fazla olduğu bölgelerde sel tehlikesinin akarsuya yakın olan bölgelere göre daha az olduğu öngörülebilmektedir. Sellerin oluşmasını etkileyen diğer bir parametre de eğri numarasıdır. 0 ile 100 arasında değişen *eğri numarası* arttıkça geçirimsizlik de doğru orantılı bir şekilde artar ve sel olaylarının görülme olasılığı artış göstermektedir (*Costache vd., 2020*). *Yağış* parametresi de sel duyarlılığı açısından büyük önem teşkil etmektedir. Özellikle son yıllarda ülkemizde yağış miktarının fazla olduğu bölgelerde sel doğal afetinin meydana gelme sıklığında önemli ölçüde artış yaşanmıştır. Bu çalışma kapsamında İstanbul İline ait yağış değerleri (günlük maksimum) MGM'nden elde edilmiş olup, her bir istasyona ait verilerde frekans analizi gerçekleştirilmiştir. Frekans analizleri neticesinde 100 yıllık tekrerr periyoduna sahip değerler hesaplanarak Kriging interpolasyon yöntemi esas alınarak nihai yağış haritası elde edilmiştir. *Topoğrafik ıslaklık indeksi* ve *topoğrafik pürüzlülük indeksi* parametreleri ise topoğrafik özelliklere bağlı olarak bölgelere göre farklılık göstermektedir. ıslaklık indeksi büyük olan bir topoğrafyada zemin neme doygun bir durumda olduğu için sızma miktarı azalış göstermekte ve sel riski artmaktadır. Topoğrafik pürüzlülük indeksi düşük olan bir bölge için ise su akışı hızlanacağı için sel tehlikesinin azalacağı öngörülmektedir. Çalışma kapsamında incelenen *akış gücü indeksi*ne bakıldığında zaman sel tehlikesinin akış gücü indeksi büyük olan bölgelerde fazla olacağı söylenebilmektedir. Bu çalışmada genellikle elde edilmesi kolay olan verilere odaklanılmıştır. Bunun nedeni ise benzer analizlerin veri ihtiyacının karşılanmadığı bölgelerde de uygulanabilmesinin mümkün kılınmaya çalışılmasıdır. Böylelikle, hemen her bölgede ölçümü gerçekleştirilebilen yağış parametresine, uzaktan algılama teknikleri ile elde edilebilen eğri numarası parametresine ve topoğrafya ve topoğrafyaya bağlı olarak CBS teknikleri ile elde edilebilen diğer parametrelere yer verilmiştir. Bu durum her ne kadar analizlerin kapsamını kısıtlasa da benzer tahmin çerçevesinin veri kısıtı ile karşılaşan bölgelerde kullanılarak başarılı modellemeler gerçekleştirilebilmesinin önü açılması amaçlanmıştır. Çalışmada kullanılan sellerin oluşumunda etkili faktörlere karşılık gelen istatistiksel bilgiler Tablo 1'de verilirken, Şekil 2 ilgili faktörlere ait haritaları sunmaktadır. Ayrıca belirtilmesi gereken önemli husus ise makine öğrenmesi uygulamalarında sıklıkla karşılaşılan çoklu bağlantım (*İng. multicollinearity*) problemlidir. Bu problem çoğunlukla bir regresyon modelindeki iki veya daha fazla bağımsız değişken yüksek oranda ilişkili olduğunda ortaya çıkmaktadır. Her ne kadar bu çalışma kapsamında regresyon modeli yerine bir sınıflandırma analizi yapılmış olsa ve söz konusu problem birincil olarak sürekli bir değişkeni tahmin etmeye yarayan lineer regresyon modelinde aransa da, bu çalışmada da çoklu bağlantımı kontrol etme noktasında en yaygın olarak kullanılan yöntem varyans büyütme faktörü (VIF) değerleri hesaplanmıştır. Bu değerler de ayrıca Tablo 1'de sunulmuştur.

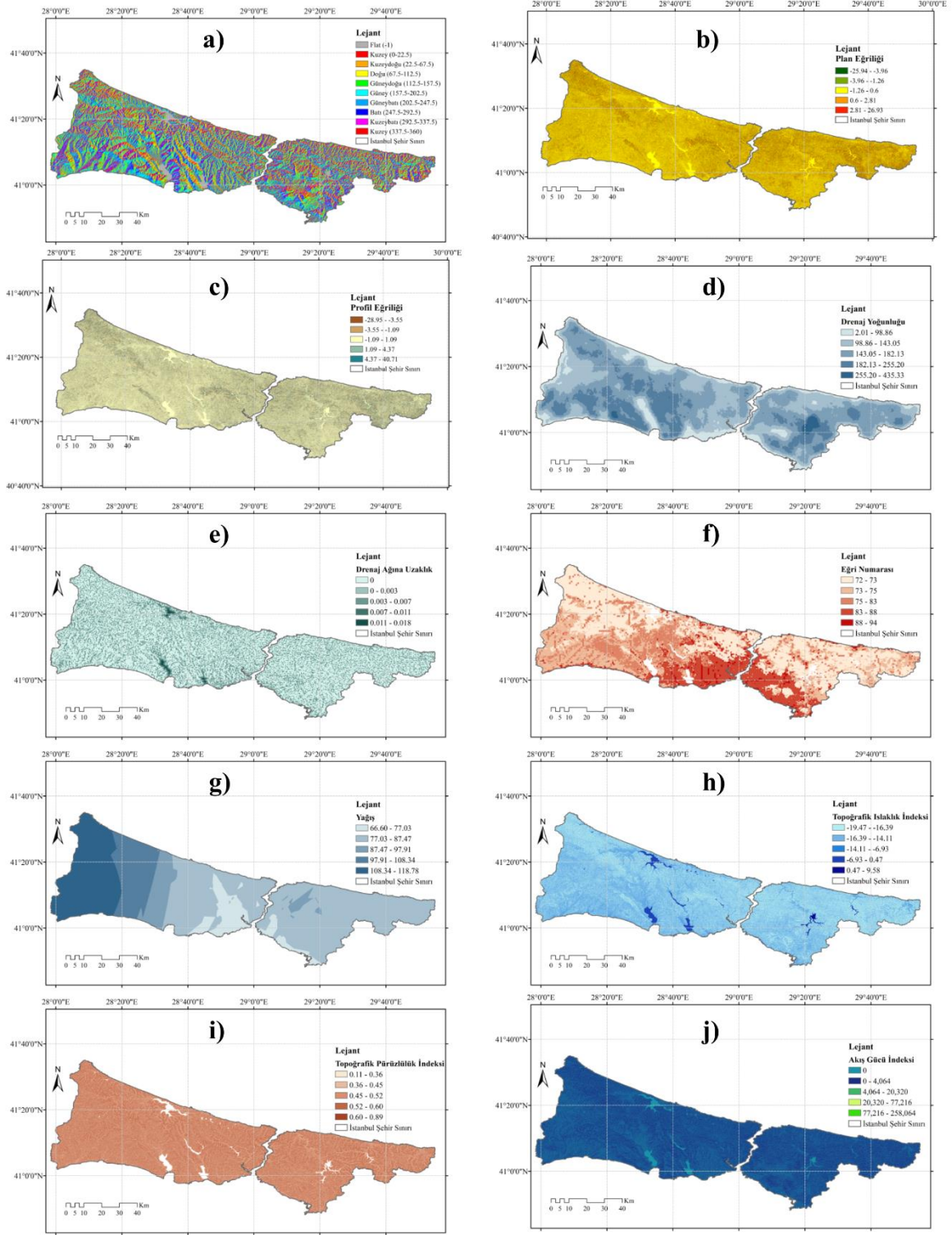
Tablo 1: Sellerin oluşumunu tetikleyici faktörlere ait verilerin istatistiksel özeti

Değişken/İstatistik	Minimum	Ortalama	Maksimum	Standart Sapma	Çarpıklık	Kurtosis	VIF
Yükseklik	0.000	118.157	480.000	71.071	0.766	0.769	1.243
Bakı	-1.000	176.993	359.284	102.963	0.012	-1.115	1.001
Plan Eğriliği	-1.27E+10	4.35E+07	1.24E+10	2.35E+09	-0.088	1.741	1.688
Profil Eğriliği	-2.10E+10	6.18E+07	1.53E+10	2.60E+09	0.078	2.503	1.958
Drenaj Yoğunluğu	2.006	149.315	410.735	39.978	0.044	1.307	1.065
Drenaj Ağına Uzaklık	0.000	0.001	0.007	0.001	0.828	-1.221	1.052
Eğri Numarası	0.000	77.066	94.000	6.223	0.761	4.083	1.252
Yağış	66.740	91.199	118.692	14.343	0.576	-1.138	1.250
TWI	-18.822	-15.965	3.575	1.643	2.503	15.359	1.645
TRI	0.000	0.497	0.889	0.089	-0.028	0.507	1.540
SPI	0.000	25.873	3342.784	100.549	12.877	279.396	1.304

Tablo 1'den de görüleceği üzere tetikleyici parametreler arasında kurulan modelin stabilitesine veya doğruluk oranını tehlikeye atacak herhangi ciddi bir ilişki yoktur. Bu çıkarım VIF değerlerinin 5' in altında olduğundan ve 1' e doğru gidildikçe çoklu bağlanım problemi riskinin azaldığı gerçeğinden hareketle yapılabilmektedir. Zira, VIF değerlerinin çok düşük olması (1'e yakın veya 2'nin altında), girdi özelliklerinin nispeten birbirinden bağımsız olduğunu ve aralarında önemli bir korelasyon olmadığını göstermektedir. Bu, modelin her bir özelliğın bireysel katkılarına dayalı olarak güvenilir ve istikrarlı tahminler yapmasına izin verdiği için arzu edilen bir sonuçtur.

3.2. Sel Envanteri

Sel envanterleri ile bir bölgenin tarihsel süreç içerisinde maruz kaldığı sel afeti eğilimi tespit edilerek gelecek yıllar için güvenli sel analizleri ve/veya tahminleri yapılabilmektedir. Türkiye'de sellerin kayıtları Devlet Su İşleri (DSİ) tarafından tutulmaktadır. Ayrıca, ilgili literatür incelendiğinde ise Türkiye özelinde gerçekleştirilmiş birçok farklı havza, bölge ve şehre odaklanmış çalışmalar olduğu görülmektedir. Bu çalışma kapsamında ise [Haltas vd. \(2021\)](#)'ın gerçekleştirdiği çalışmadan faydalanılarak İstanbul bölgesi için sel envanter verileri temin edilmiştir. Bu çalışmada araştırmacılar mevcut küresel ve ulusal veri havuzlarını kapsamlı bir şekilde gözden geçirerek Türkiye geneli bir sel envanteri oluşturmuşlardır. Mevcut ulusal veriler kullanılarak 1930 ve 2020 yılları arasında Türkiye genelinde meydana gelen sel olayı envanterinin derlemesi yapılmış olup, sayısallaştırılmış medya arşivleri de dahil olmak üzere birçok veri kaynağı incelemesi neticesinde toplamda 2101 sel olayına ait veriler elde edilmiştir. Bu çalışma kapsamında ise söz konusu çalışmada sunulan veri tabanında İstanbul sınırları içerisinde kayıt altına alınan veriler kullanılmış ve çeşitli veri-tabanlı metotlar kullanılarak sel duyarlılık analizleri gerçekleştirilmiştir. Tarihsel veri kayıtlarına ait sel noktaları, Google Earth ve ArcGIS 10.3 ([ESRI, 2016](#)) yazılımları aracılığı ile incelenmiştir. Sel olmayan noktaların belirlenmesinde ise literatürdeki genel yaklaşım olan eşit sayıda rastgele sel olmayan nokta oluşturulması mantığı yerine, gerçek dünya koşullarını daha doğru bir şekilde temsil edecek şekilde sel olaylarından çok daha fazla sel olmayan nokta oluşturulmuştur ([Ekmekcioğlu vd., 2022b](#); [Pham vd., 2023](#)). Dolayısıyla, sel gerçekleşen (1) ve gerçekleşmeyen (0) noktaların sayılarında bir sınıf dengesizliği oluşturulmuştur. Bu sınıf dengesizliği de ayrıca makine öğrenmesi yöntemlerinin güvenilirliğinin artırılması bakımından ön işleme adımları ile eğitim setinde ortada kaldırılmıştır. Bu işlemlere ait detaylar aşağıdaki alt başlıklarda detaylı olarak ifade edilmiştir.



Şekil 2: Tetikleyici faktörlere ait haritalar: a) bakı, b) plan eğriliği, c) profil eğriliği, d) drenaj yoğunluğu, e) drenaj ağına uzaklık, f) eğri numarası, g) yağış, h) TWI, i) TRI, j) SPI

4. Metodoloji

4.1. Ön İşleme Uygulamaları

4.1.1. Ölçekleme

Çoğu makine öğrenmesi yöntemi, tahminlerin gerçekleştirilmesinde kullanılan özniteliklerin sayısal olarak farklı aralıklarda olmasından ötürü yüksek performans gösteremeyebilmektedir (Koc vd., 2021). Bu sorunu çözmek için yaygın olarak kullanılan iki çözüm vardır: 1) Min-Maks ölçekleyici ve 2) Standart ölçekleyici (Bhattacharya vd., 2020). Min-Maks ölçekleyici 0 ile 1 arasında bir aralık oluşturarak sayısal öznitelik değerlerini belirli bir aralığa indirmektedir. Öte yandan, Standart ölçekleyici ise, sayısal özellikleri, sırasıyla 0 ve 1 ortalama değerine ve standart sapmasına sahip bir dağılımla sonuçlanacak şekilde dönüştürmektedir. İlgili literatüre göre karar ağacı temelli makine öğrenmesi algoritmalarının ölçekleme ve normalizasyona ihtiyaç duymadığına dair birçok çalışma bulunsa da (Kumar vd., 2022), bazı çalışmalar ise her ne kadar farklı sayı uzaylarına duyarlılık göstermeyen ağaç tabanlı algoritmalarının ölçeklemeye tabi tutulmasının performans iyileştirmeleri yapacağına işaret etmiştir (Izonin vd., 2022). Bu çalışmada, literatürde sıklıkla kullanılması (Sarker vd., 2020) ve model performanslarını artırmadaki etkinliği (Thippa Reddy vd., 2020) nedeniyle Standart ölçekleyici tekniği uygulanmıştır. Bu işlem, Denklem 1'de gösterildiği üzere örneklem x 'in normal değeri, her bir örneklem değerinden ortalama değer çıkarılarak ve ardından standart sapmaya bölünerek gerçekleştirilmektedir.

$$z = \frac{x-u}{s} \quad (1)$$

Burada u örnekleme yer alan değişkenlerin ortalamasını ve s ise standart sapmalarını ifade etmektedir.

4.1.2. Sınıf Dengesizliği

Hedef değişkenler arasında yer alan sınıf dengesizliği gerçek dünya problemlerinin belirsiz doğası göz önüne alındığında, sınıflandırma problemlerinde yaygın olarak karşılaşılan zorluklardan biridir. Bu nedenle, çözümler çoğu durumda çoğunluk sınıfları üzerinde yoğunlaşmaktadır. Örneğin, bir yağış olayı sonucu oluşacak sel olayının geniş bir alanda önemli etkileri olabilirken, seller odaklanılan bölgenin sadece bazı belirli yerlerinde meydana gelmektedir. Öncelikli amacın sel olayı meydana gelen yerleri tespit etmek olduğu düşünülürse, yağış olayından etkilenen ancak sel olayının ortaya çıkmadığı diğer noktalar da dikkate alınmalıdır. Benzer problemlerle başa çıkmak için mevcut literatür, tahminlere yönelik önyargıyı kesin olarak ortadan kaldıran sınıf dengesizliği işleme tekniklerinin uygulanabilirliğine işaret etmektedir. Bu nedenle, bu çalışmada gerçekleşen sellerin lokasyonlarının belirlenmesinde yeniden örnekleme yöntemlerinin etkinliğini ve uygunluğunu göze almıştır. Bu doğrultuda, literatürde sıklıkla kullanılan rastgele alt örnekleme (random under-sampling-RUS) algoritmasından yararlanılmıştır. RUS algoritması ile çoğunluk sınıfındaki ögeler rastgele silinerek sınıf dağılımı dengelenmektedir (Kang & Ryu, 2019). Bu algoritmanın ana dezavantajı, eğitim verilerinde bilgi kaybı riski olsa da, örneklem boyutunun küçültülmesi ile hesaplama süresinde kazanç sağlanmasıdır (Liang vd., 2020). İlgili literatür incelendiğinde, özellikle büyük veri kümesine odaklanıldığı durumlarda RUS'un uygun bir alternatif olabileceğine işaret edilmektedir (Provost, 2000).

4.2. Makine Öğrenmesi Algoritmaları

Bu çalışma kapsamında İstanbul ili sel duyarlılık haritalamasının yapılabilmesi için CBS-tabanlı makine öğrenmesi algoritmalarından yararlanılmıştır. Bu bağlamda, Aşırı gradyan artırma (XGBoost) algoritması sel duyarlılık sınıflandırılmasında kullanılmıştır. Ayrıca, bu yöntemin karşılaştırılmalı değerlendirilmesi için (İng. benchmarking) iki ayrı ağaç tabanlı algoritması (rastgele orman ve stokastik gradyan artırma) da CBS tabanlı tahmin çerçevesine entegre edilmiştir. Aşağıda yer alan alt başlıklarda kullanılan üç algoritmanın çalışma prensipleri özetle ifade edilmiştir.

4.2.1. Rastgele Orman (RF)

Rastgele orman (RF), bir topluluk öğrenme algoritması olarak tanımlanabilmektedir (Breiman, 2001). Bu algortmada, karar ağacının her düğümündeki en iyi dallanmış eleman, tüm eleman kümeleri arasında gelişigüzel olarak belirlenen elemanlar arasından seçilir. Oluşan düğümler boyunca en iyi dallanmanın sağlanabilmesi için ise Gini indeksi kullanılır. N sınıftan örneklem içeren bir veri seti için Gini indeksi aşağıdaki şekilde hesaplanmaktadır (Denklem 2):

$$Gini(t) = 1 - \sum_j [p(j|t)]^2 \quad (2)$$

Formülasyondaki j ve t sırasıyla sınıfı ve düğümü temsil etmektedir. Ayrıca $p(j|t)$, t düğümündeki j sınıfının göreceli olasılığını ifade etmektedir.

Gini indeksi ayrıca değişkenlerin homojenliği hususunda da önemli bilgilere erişilmesine imkân tanımaktadır. Sıfıra yakın Gini indeksi, artan homojenliği gösterirken, daha yüksek değer ise heterojene doğru eğilimi ifade etmektedir (Ekmekcioğlu vd., 2020).

RF algoritması, kullanılacak örneklem sayısı ve ağaç sayısı gibi hiperparametreleri içermektedir. Kullanıcı, karar ormanını N sayıda ağaçla oluşturur. Böylece her bir sınıflandırma işlemi için veri seti N sayıda karar ağacından geçer. Ayrıca eğitim setinin ve test setinin belirlenmesi RF yöntemi için kritik öneme sahiptir (Yu & Abdel-Aty, 2014). Bir RF modelinin ortalama hata oranı elde edilir ve hata oranı en düşük olan ağaç en yüksek ağırlıkla değerlendirilir. En düşük ağırlıklar ise, hata oranı en yüksek olan ağaca atanır. Genel olarak, her ağacın bir oyu olduğu göz önünde bulundurularak, nihai tahmini elde etmek için yani en uygun sınıflandırıcıyı bulmak için çoğunluğun oylaması (İng. majority voting) gerçekleştirilir.

4.2.2. Stokastik gradyan artırma

Friedman (2002), stokastik gradyan artırma (SGB) algoritmasını hem hesaplama verimliliğini hem de tahmin başarılarını artırmak için geleneksel gradyan artırmadan (İng. gradient boosting) türetmiştir. Birkaç zayıf sınıflandırıcıdan topluluk modelleri oluşturmaya dayanan gradyan artırma hem regresyon hem de sınıflandırma amaçları için kullanılabilir. Gradyan artırma temel olarak diğer artırma tekniklerine benzer şekilde adım adım oluşturulsa da, gradyan yükseltmede keyfi türevlenebilir kayıp fonksiyonunun optimizasyonuna içerdiği daha genellenebilir modellerin oluşturulması sağlanmaktadır (Devi ve Kumar, 2022). Yine de yinelemeli süreç düzgün bir şekilde düzenlenmezse, gradyan artırma aşırı uyumdan zarar görme eğilimindedir (Bentjac vd., 2021). SGB’de sınıflandırma ağaçları, önceki ağacın kayıp fonksiyonunun gradyanı olarak adlandırılan "sözde" hatalardan temel öğrencileri sırayla uydurarak oluşturulur (Moisen vd., 2006). SGB, model performansında bir artış sağlamak için eğitim örneklerine rastgeleleştirme olgusunu dahil eder. Model oluşturmak için tüm örnekleri dikkate alınmanın hem hesaplama verimliliği hem de fazla uydurma riski açısından zor olabileceği göz önüne alındığında, eğitim verilerinin yalnızca rastgele seçilmiş bir kısmını kullanmak, algoritmanın bu sorunları aşmasına ve algoritmayı daha az hassas hale getirmesine yardımcı olur. SGB’nin matematiksel hesaplama prosedürleriyle ilgili daha fazla ayrıntı bilgi Friedman (2002) ve Devi ve Kumar (2022)’in çalışmalarında yer almaktadır.

4.2.3. Aşırı gradyan artırma

Aşırı gradyan artırma (XGBoost), gradyan artırma karar ağacının (GBDT) performansını iyileştirmek için önerilen bir algoritmadır. XGBoost algoritması algoritmanın belirli kısımlarını birden fazla işlemci üzerinde aynı anda yürütme tekniği olarak adlandırılan paralelleştirme özelliğine de sahiptir (Chen & Guestrin, 2016). Bu özellik ile hesaplama verimliliği önemli ölçüde artırılır ve eğitim süresini azaltılarak yüksek boyutlu verilerde kısa sürede doğruluk oranı yüksek sonuçlar elde edilebilmektedir. Dış döngüler ağaç yapraklarını oluştururken, iç döngüler özneliklerin elde edilmesinde rol oynamaktadır. İç döngüler sonlandırılmadan dış döngüler sonlandırılmadığından, yani öznelikler hesaplanmadan ağacın yaprakları oluşturulmadığından dolayı paralelleştirme de sınırlıdır. XGBoost algoritmasında işlem süresini azaltmak için, tüm örnekler toplu olarak dikkate alınır ve algoritmaya giren döngü adımlarının sayısı belirlenir. Algoritmanın diğer bir avantajı ise XGBoost’un ağaç yapılarında kullanılan arama algoritması olan derinlemesine aramayı (DSF) gerçekleştirebilmesidir (Başakın vd., 2022). Bu algoritma, rastgele başlayan düğümden ulaşılabilecek en derin düğüme kadar arama yapılmasına izin verir. Hesaba katılacak daha derin düğüm kalmadığında, algoritma geri sarar ve derin düğümlere öncelik vererek aramaya devam eder ve en uygun sınıflandırıcının elde edilmesini sağlamaktadır.

4.3. Performans Değerlendirme Kriterleri

Sel duyarlılığı tahmininin önemli bir parçası doğrulama sürecidir. Elde edilen sonuçların bilimsel değeri sınırlı olduğu için çalışmalar hızlı bir şekilde kontrol edilerek doğrulama yapılmalıdır. Analiz sonuçlarını doğrulamak için birçok performans değerlendirme göstergesi vardır. Doğrulama sürecinde dört temel istatistiksel önlem kullanılır. Bu önlemler sırasıyla gerçek pozitif (TP), gerçek negatif (TN), yanlış pozitif (FP) ve yanlış negatif (FN) önlemlerdir. Burada TP, doğru sınıflandırılan sel olayı sayısını, TN doğru sınıflandırılan sel olmayan olayların sayısını, FP yanlış sınıflandırılan sel olayları sayısını, FN ise yanlış sınıflandırılmış sel olmayan olayların sayısını temsil etmektedir. Bu dört ölçüte göre, hassasiyet (İng. precision) ve duyarlılık (İng. recall) değerleri hesaplanmıştır. Fakat özellikle sınıf dengesizliğinin olduğu durumlarda bu iki performans göstergesinin model başarı değerlendirmesi için kullanılmasının sonuçların yorumlanmasında hatalara sebep olacağından, bunların harmonik ortalaması ile elde edilen F-1 skoru değerleri hesaplanmıştır (Denklem 3). Ayrıca, sınıf dengesizliğinin olduğu durumlarda oldukça güvenilir bir değerlendirme kriteri olan alıcı işletim karakteristiği altındaki alan (AUROC) da kullanılmıştır.

$$F - 1 \text{ skoru} = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \quad (3)$$

AUC’un 0.5 değeri yetersiz performanstaki tahminleri temsil ederken, 1 mükemmel uyumlu modeli ifade etmektedir. AUC aşağıdaki formül kullanılarak hesaplanmaktadır (Feng vd., 2020):

$$AUC = \int_0^1 TPR(FPR)dFPR \quad (4)$$

Formülde yer alan TRP gerçek pozitif oranı (TP/(FN+TP)) temsil ederken, FPR ise yanlış pozitif oranı (FP/(TN+FP)) temsil etmektedir. Dolayısıyla, TPR (FPR) terimi gerçek pozitif oranının yanlış pozitif oranının fonksiyonu ile bulunduğunu göstermektedir.

Tablo 2: 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)

Tablo 3: Performans metriklerinin sınır değerleri

Metrik	En Başarısız Durum	En Başarılı Durum	Eşik Değeri
Hassasiyet	0	1	N/A
Duyarlılık	0	1	N/A
F1-skoru	0	1	0.8*
AUROC	0.5	1	0.7*

N/A: Sınıf dengesizliği durumunda kullanılması önerilmemektedir; * Başarılı sonuçları ifade eden eşik değerini temsil etmektedir.

5. Bulgular ve Tartışma

Bu çalışmanın amacı makine öğrenmesi ve farklı yöntemleri kullanılarak İstanbul iline ait sel duyarlılık haritasının elde edilmesidir. Ayrıca, sel fenomeninin doğası gereği birçok yağış olayı meydana geldiğinde aynı bölgede bir kısım yerler sel ile karşılaşırken bir kısım yerlerde ise seller meydana gelmediğinden dolayı sınıf dengesizliği oluşmaktadır (Ekmekcioğlu vd., 2022b). Bu sınıf dengesizliği ise makine öğrenmesi yöntemlerine tanıtılmak suretiyle analizler gerçekleştirilmiştir. Sınıf dengesizliği problemindeki dengesizlik oranının belirlenmesinde ise Ekmekcioğlu vd. (2022) çalışması esas alınmıştır. İlgili çalışmada Alabama Nehri Havzası için yapılan analizlerde gerçeği en doğru şekilde yansıtan dengesizliğinin, sel olaylarının meydana gelme sayısının 50 ile çarpılarak sel meydana gelmeyen noktaların belirlenmesine göre ele alınmasına işaret edilmiştir. Bu doğrultuda, bu çalışma kapsamında da İstanbul ili için elde edilen sel envanterinde yer alan sel gerçekleşen noktaların sayısı 50 ile çarpılarak sel meydana gelmeyen nokta sayısı belirlenmiştir. Bu noktaların belirlenmesi ise ArcGIS 10.3 yazılımı yardımıyla rastgele noktaların oluşturulması suretiyle gerçekleştirilmiştir. Tüm bu veri seti elde edildikten sonra ise, eğitim ve test setleri olmak üzere iki eşit parçaya ayrılmıştır. Eğitim seti kullanılarak model kalibrasyonu gerçekleştirilmiş ve test setinden yararlanılarak ise modelin performans değerlendirmesi yapılmıştır. Fakat, model tahminlerinde yanlışlığın asgariye indirilmesi adına eğitim setinde meydana gelen sınıf dengesizliği RUS yöntemi kullanılarak çıktı değişkenleri arasında sayısal dengenin oluşturulması suretiyle elimine edilmiştir. Aksi takdirde, makine öğrenmesi modeli sayıca üstün olan sel meydana gelmez sonucuna yakınsama potansiyeli göstermektedir (Koc vd., 2022). Böylelikle tüm eğitimin sel meydana gelmez yaklaşımı ile yapılması ve test setine göre değerlendirmelerde ise modelin hemen hemen tüm örneklerde sel meydana gelmez sonucuna eğilim göstereceği öngörüldüğünden bu ön-işleme adımı sınıf dengesizliği yer alan makine öğrenmesi problemlerinde oldukça önem arz etmektedir.

Modelin eğitim setine göre kalibrasyonunun yapılmasında ise uygulanan üç makine öğrenmesi yöntemi (RF, SGB ve XGBoost) için de deneme yanılma tekniği kullanılmıştır. Bu teknik esnek, pratik ve çoğu zaman içsel parametrelerin düşük hesaplama maliyetlerine imkân sağladığından tercih edilmektedir (Guo vd., 2019). Her bir makine öğrenmesi yöntemine ait içsel parametrelerin elde edilen optimum değerlerine ise Tablo 4'te yer verilmiştir. Bu aşamadan sonra ise ilgili parametreler kullanılarak kurulan modeller bağımsız ve daha önce modele tanıtılmayan test setine tabi tutulmuştur. Böylelikle, daha güvenilir ve adil bir model performans karşılaştırılması gerçekleştirilmiştir.

Test setine uygulanan modeller ise birçok farklı performans değerlendirme kriteri baz alınarak kapsamlı bir biçimde incelenmiştir. Bu performans metrikleri literatürde ikili sınıflandırma problemlerinde yaygın olarak kullanılan Hassasiyet, Duyarlılık, F1-skoru ve AUROC olarak belirlenmiştir (Ekmekcioğlu & Koc, 2022). Her ne kadar Hassasiyet ve Duyarlılık özellikle sınıf dengesizliğinin olduğu problemlerde dikkatli yaklaşılması gereken metrikler olsa da (Kamalov, 2020; Kim & Byun, 2022) bu çalışmada literatürde yaygın olarak kullanıldığından dolayı yer verilmiştir.

Tablo 4: Çalışma kapsamında kullanılan yöntemlere ait içsel parametreler ve karşılık gelen optimize edilmiş değerler

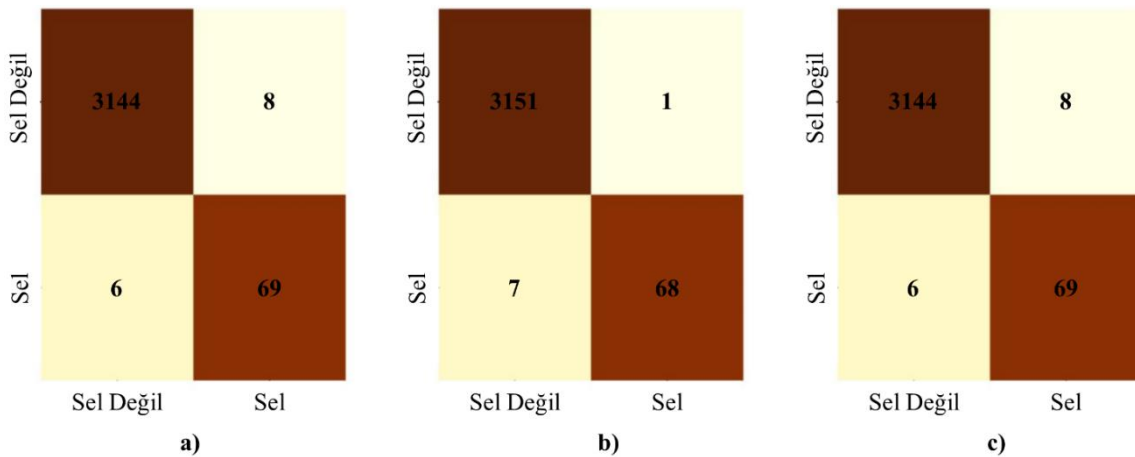
Hiperparametre /Yöntem	Maksimum ağaç derinliği	Ağaç sayısı	Minimum örneklem ayırması	Öğrenme oranı	Minimum yaprak sayısı	Alt örneklem oranı
RF	150	50	5	-	-	-
SGB	50	250	5	0.05	4	0.4
XGBoost	250	50	-	0.1	-	-

Fakat esas performans değerlendirmeleri sınıf dengesizliğinden görece en az etkilendiği literatürde vurgulanan F-1 skoru ve AUROC’ a göre gerçekleştirilmiştir. İlgili sonuçlar ise Tablo 5’de sunulmuştur. Tüm sonuçlar incelendiğinde hemen her makine öğrenmesi modelinin yüksek performans gösterdikleri görülmektedir. Bu durum her modelin hem F-1 skoru için hem de AUROC için 0.9’ dan yüksek sonuçlar elde edilmesi ile açıklanabilmektedir (Choi vd., 2020). Modellerin ayrı ayrı karşılaştırılmasında ise, en yüksek model performansının XGBoost ile elde edildiği, bu metodu ise sırasıyla SGB ve RF’ nin takip ettiği görülmektedir. Zira XGBoost metodu ile F-1 skoru 0.9957 ve AUROC değeri 0.9840 olarak bulunmuştur. Öte yandan, SGB yöntemi ile ise F-1 skoru 0.9975 ve AUROC değeri 0.9803 olarak elde edilmiştir. Bu noktada, SGB ile elde edilen F-1 skoru değerinin XGBoost ile elde edilen değerden yüksek olduğu dikkat çekmektedir. Fakat aradaki farkın oldukça az olması ve XGBoost için elde edilen AUROC değerinin SGB için elde edilen değerden önemli ölçüde fazla olması genel değerlendirmelerde XGBoost modelini ön plana çıkarmaktadır. Ayrıca, benzer durum RF modeli için de geçerlidir. RF modeli kullanılarak elde edilen test setine göre sonuçlarda da SGB modeli ile aynı olacak şekilde F-1 skoru 0.9978 olarak elde edilmiştir. Fakat AUROC performans kriterine göre yapılan hesaplamalarda ise RF modeli 0.9742 sonucunu vermiştir. Bu değer de benzer şekilde XGBoost ile elde edilen değerden önemli ölçüde düşük olarak göze çarpmaktadır.

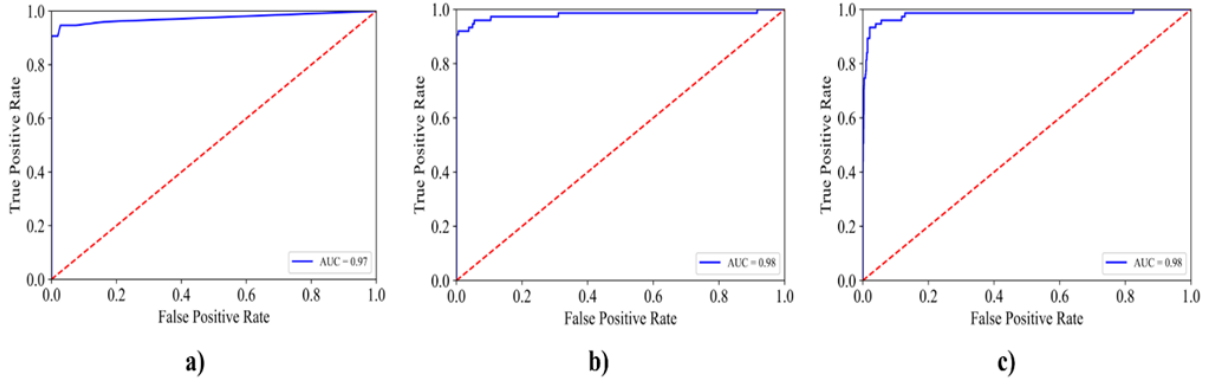
Tablo 5: Test setine göre elde edilen performanslar

Yöntem	Hassasiyet	Duyarlılık	F1-skoru	AUROC
RF	0.9978	0.9978	0.9978	0.9742
SGB	0.9975	0.9975	0.9975	0.9803
XGBoost	0.9957	0.9957	0.9957	0.9840

Performans kıyaslarının daha kapsamlı değerlendirilmesi adına hata matrisleri Şekil 3’de sunulmuştur. Hata matrisleri incelendiğinde ise RF modelinin sel olmayan noktaların tamamını doğru olarak bulduğu (%100), sel olan noktalarda ise %90.67’lik bir başarı sergilediği görülmektedir. Benzer şekilde, SGB modeli ile de sel olmayan noktaların belirlenmesinde %99.97’lik bir başarı sergilenirken, sel olan noktalarda ise %90.67’lik bir tahmin başarısı elde edilmiştir. Fakat, çalışmanın esas amacını kapsayan sel gerçekleşen noktaların belirlenmesinde XGBoost modeli %92.00’ lik bir başarı ile diğer iki metoda üstünlük sergilemiştir. Ayrıca, XGBoost modeli sel olmayan yerlerin tahmininde de oldukça yüksek bir performans olan %99.75 başarı etkisi göstermiştir. Tüm bu sonuçlar da değerlendirildiğinde kullanılan üç modelin karşılaştırılmasında XGBoost modeli bir adım öne çıkmaktadır. Bu çalışmada ayrıca, tahmin modellerinin gösterdiği performanslar AUC grafikleri de sunulmuştur (Şekil 4).

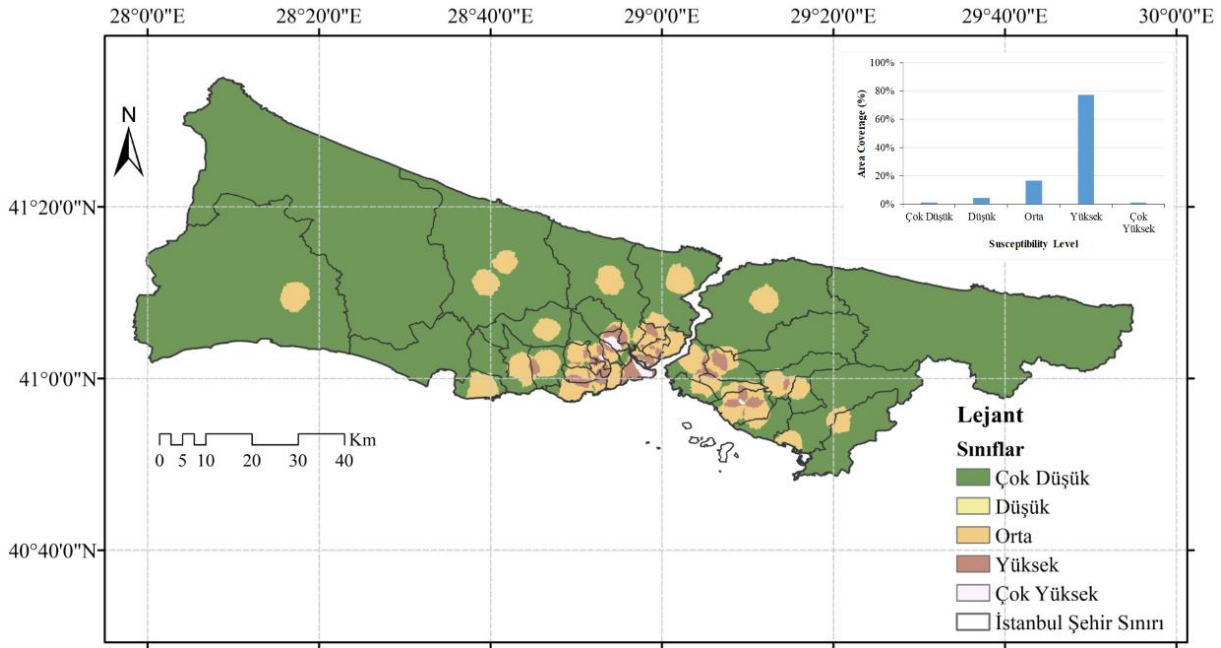


Şekil 3: Hata matrisleri: a) RF, b) SGB, c) XGBoost



Şekil 4: AUC-ROC grafikleri: a) RF, b) SGB, c) XGBoost

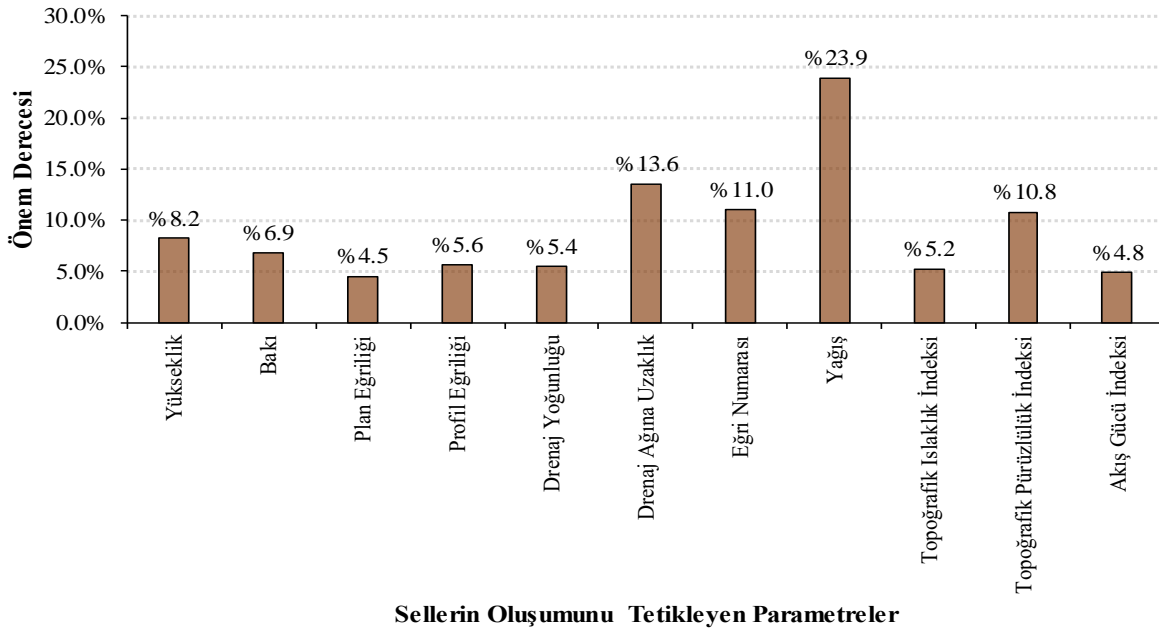
Bu çalışma kapsamında, İstanbul iline ait sel duyarlılık haritasının elde edilmesinde XGBoost yöntemine ait sonuçlar kullanılmıştır. Dolayısıyla, XGBoost metodu ile ulaşılan test seti sonuçlarına göre sel meydana gelen ve sel meydana gelmeyen bölgeler ikili sınıflandırma mantığı gereği sırasıyla 1 ve 0 olarak noktasal bazda (koordinatları ile) elde edilmiştir. Daha sonra ise, literatürde yaygın olarak kullanılan metotlardan biri olan Kriging yöntemi ile 1 ve 0 noktaları interpolate edilmek suretiyle İstanbul sınırları esas alınarak sel duyarlılık haritası üretilmiştir (Şekil 5). Elde edilen haritanın farklı sınıflara bölünmesinde ise daha önce birçok çalışmada seli tetikleyen parametrelerin haritalanmasında kullanılan ve farklı veri tiplerine uygunluk, sınıf heterojenliğine daha düşük hassasiyet gösterme, anlaşılması kolay sonuçlar verme gibi avantajları nedeniyle Jenks Natural Break metodundan (Ekmekcioğlu vd., 2022b; Pham vd., 2023) yararlanılmıştır (Şekil 5). Harita incelendiğinde Avrupa yakasının tarihi yarımada kısımlarında sel duyarlılığının oldukça yüksek olduğu görülmektedir. Asya yakasında ise Marmara kıyısına yakın olan Maltepe, Kadıköy ve Ümraniye bölgelerinde sel duyarlılığının diğer bölgelere göre daha fazla olduğu görülmüştür.



Şekil 5: İstanbul ili sel duyarlılık haritası

Makine öğrenmesi yöntemlerinin kullanılmasında literatür tarafından en çok eleştirilen hususlardan biri de bu modellerin kara kutu modeller olarak çalıştığı, yani model yorumlanabilirliğinin oldukça kısıtlı olmasıdır (Zhang vd., 2018). Bu nedenle, bu çalışma kapsamında tahmin sonuçlarını etkileyen faktörlerin önem dereceleri Gini impurity kriteri yardımıyla elde edilerek yüzdelik cinsten sıralanmıştır (Şekil 6). Şekil incelendiğinde sel duyarlılık haritalamasına en fazla etkisi olan değişkenin %23,9 önem derecesi ile yağış parametresinin olduğu görülmektedir. Literatür incelendiğinde yağışın selleri etkilediği ve birçok çalışmada en önemli değişkenlerden biri olduğu yükselti farkının fazla olduğu dağlık bölgelerde yağışın daha fazla olduğu (Norollahi & Seyed Kaboli, 2021; Tien Bui vd., 2019), yağış ile doğru orantılı olarak artış gösteren sel afetinin de yağış miktarına bağlı olarak artmakta olduğu tespit edilmiştir.

Ek olarak, bu çalışma neticesinde elde edilen ikinci önemli değişken ise %13,6 önem derecesi ile drenaj ağına olan uzaklık parametresidir. Drenaj ağına olan uzaklığın az olduğu bölgelerde geçmişten günümüze kadar sel olaylarının etkisinin daha fazla meydana geldiği görülmektedir (Avand vd., 2021; Fiori & Volpi, 2020). Hatta, ekstrem yağış olaylarının neden olduğu sel olaylarında akarsuların kapasitesinin yetmediği durumlarda kesitten taşan suyun civar bölgelerde sellere sebebiyet verdiği de bilinmektedir. Üçüncü olarak ise %11 önem derecesi ile eğri numarası en önemli değişkenlerden biri olarak elde edilmiştir. Eğri numarası bir bölgedeki geçirimsizlik durumunu temsil etmektedir. Eğri numarasının yüksek olduğu bölgeler daha geçirimsiz bölgelere işaret ederken, düşük olduğu bölgeler geçirimsiz alanların görece daha yaygın olduğunu ifade etmektedir (Jaafar vd., 2019). Bu nedenle, ekstrem ve ani yağışlarda özellikle geçirimsizlikten kaynaklanan yağışın doğrudan veya kısa bir sürede akışa geçmesi sel olaylarının sayısında da artışa neden olmaktadır. Diğer değişkenler ile kıyaslandığında plan eğriliği, akış gücü indeksi ve topoğrafik ıslaklık indeksi parametreleri de tahmin performanslarına katkı bakımından en düşük değişkenler arasında bulunduğu görülmektedir. Literatür ile karşılaştırıldığında elde edilen sonuçların oldukça tutarlı olduğu gözlemlenmiştir. Halihazırda, yüksek tahmin başarısına sahip olan modelin model performans değerlendirme kriterlerinin yüksek olmasının yanı sıra, sonuçların literatür ile de doğrulandığı ifade edilebilmektedir. Bu sonuçlar ile de yapılan analizlerin doğruluk oranına ilaveten fiziksel olarak da anlamlı bir biçimde açıklanabildiği gösterilmektedir.



Şekil 6: Sellerin oluşumunu tetikleyen faktörlerin önem dereceleri ve karşılık gelen sıralamaları

6. Sonuçlar

Günümüzde küresel ısınma ve iklim değişikliği ile birlikte formasyonlarında da değişimler gerçekleşen sel olayları can kayıpları başta olmak üzere birçok kayıplara neden olmaktadır. Özellikle atmosferik ve meteorolojik faktörlere insan kaynaklı etkiler de eklendiğinde, doğa kaynaklı tehlike olarak ifade edilebilen seller afete dönüşmektedir. Dünyanın her bölgesi olduğu gibi İstanbul da bu afetlerden etkilenmeye devam etmektedir. Sellerin önlenmesi veya etkilerinin azaltılabilmesi için yapılabilecek ilk adım bu olayların gerçekleştiği bölgelerin tespit edilebilmesidir. Bu bağlamda, bu çalışmada İstanbul iline ait geçmişte yaşanan sel olaylarının lokasyonları literatürde yer alan bir çalışmadan elde edilmiş olup, makine öğrenmesi yöntemleri ile sel duyarlılık haritası oluşturulması amaçlanmıştır. Ayrıca, bölgelerin karakteristiklerine göre aynı yağış olayı neticesinde birtakım bölgelerde sel olurken diğer bölgelerde sel olayı meydana gelmeyebilmektedir. Dolayısıyla, bu durumun da hesaba katılabilmesi adına bu çalışmada literatüre de dayandırılmak suretiyle x kadar sel (ikili sınıflandırma terminolojisinde “1”) olmuşsa $50x$ kadar sel gerçekleşmeyen nokta (ikili sınıflandırma terminolojisinde “0”) rastgele olarak belirlenmiştir. Fakat bu işlemin gerçekleştirilmesinden sonra “1” ve “0” sayıları arasında model performanslarını yanlı bir hale getirecek sınıf dengesizliği ortaya çıkmıştır. Bunun üstesinden gelmek adına ise literatürde en yaygın olarak kullanılan yöntemlerden rastgele alt ölçekleme algoritması sadece eğitim setine uygulanmıştır. Çalışma kapsamında öne sürülen XGBoost algoritması ayrıca aynı ağaç tabanlı makine öğrenmesi ailesinden gelen RF ve SGB algoritmaları ile karşılaştırılmıştır. Elde edilen sonuçlara göre, makine öğrenmesi algoritmaları farklı performans metrikleri ile hem “1” hem de “0” olaylarını tahmin etme başarısı bakımından ayrı ayrı değerlendirilmiştir. Buna göre, uygulanan tüm makine öğrenmesi yöntemleri sel olmayan bölgelerin neredeyse tamamını doğru tahmin etmiştir.

Fakat her ne kadar üç yöntem sel gerçekleşen noktaları da birbirine yakın performans ile belirlese de, çalışmanın ana amacı olan sel noktalarını belirleme hususunda XGBoost algoritması % 92'lik bir başarı ile ön plana çıkmıştır. Farklı performans indikatörlerine göre gerçekleştirilen genel değerlendirmelerde de XGBoost, özellikle sınıf dengesizliğinin olduğu problemlerde karşılaştırmanın yapılmasının en doğru olduğu AUROC metriğine göre 0.9840 değeri ile en iyi sonucu veren algoritma olmuştur.

Bu çalışma ayrıca, makine öğrenmesi algoritmalarının sonuçlarını CBS ortamında inceleyerek sel duyarlılık haritasını da İstanbul için oluşturmuştur. Elde edilen verilerin kısıtlı olması ve bu haritaların kısıtlı miktarının sadece test kısmına ayrılanları ile oluşturulduğu düşünüldüğünde, il sınırları içerisinde sel duyarlılığının olduğu bölgelerde belirli noktalarda yoğunlaşmıştır. Bu bağlamda, Avrupa yakasının tarihi yarımada kısımlarında sel duyarlılığının oldukça yüksek olduğu görülmektedir. Asya yakasında ise Marmara kıyısına yakın olan Maltepe, Kadıköy ve Ümraniye bölgelerinde sel duyarlılığının diğer bölgelere göre daha fazla olduğu görülmüştür.

Makine öğrenmesi algoritmaları kısıtlı yorumlanabilir özellikleri ile literatür tarafından fiziksel dinamikleri yansıtmamalarından dolayı eleştirilmektedir. Bu çalışmada ayrıca, sel olayını etkileyen faktörlerin önem derecelerinin ortaya koyulması bu nedenle gerçekleştirilmiştir. Sonuçlara göre ise, beklendiği gibi yağış faktörü sel duyarlılık haritalamasında en önemli değişken olarak bulunurken, bu faktörü sırasıyla drenaj ağına uzaklık, eğri numarası ve topoğrafik pürüzlülük katsayısı takip etmiştir. Öyle ki, şehrin özellikle kentsel bir karakteristikte olması ve arazi kullanım yapısı ile geçirimsizliği birlikte temsil eden eğri numarası ve topoğrafik pürüzlülük katsayısı faktörlerinin sel gerçekleşen veya gerçekleşmeyen noktalardaki özellikleri oldukça başarılı bir biçimde temsil edilmiştir.

Literatürde yer alan hemen her çalışmada olduğu gibi bu çalışmada da her ne kadar yenilikçi bir yaklaşım kapsamlı analizler ile adapte edilse de birtakım kısıtlar hem kavramsal hem de metodolojik açıdan mevcuttur. Örneğin, sel olaylarının gerçekleştiği noktalara dair bilginin kısıtlı sayıda olması bu eksikliklerden en önemlisidir. Ayrıca kavramsal açıdan, daha fazla girdi değişkeninin model tahminlerine dahil edilmesi (model karmaşıklığını yönetilebilir seviyede tutmak suretiyle) sonraki çalışmalarda farklı perspektifler ortaya koyacaktır. Bu bağlamda sadece yağış, eğri numarası, topoğrafya ve topoğrafyaya bağlı diğer parametrelere ilaveten sonuçlar üzerinde etkisi olabilecek farklı parametrelerin de sel duyarlılık haritalaması işlemlerine dahil edilmesi faydalı olacaktır. Öte yandan, sel olaylarının sadece duyarlılık veya meydana gelebilirlik boyutlarından değil, aynı zamanda maruziyet seviyelerinin de ortaya koyularak risk boyutunda incelenmesi ilerleyen çalışmalar için önemli gelişmeler kaydedilmesini sağlayabilecektir. Böylelikle, ani sel olaylarının sonuçlarının sosyal ve ekonomik boyutlardan da incelenmesine ve alınabilecek tedbirlerin bu çıktılara göre belirlenebilmesi de sağlanabilecektir. Metodolojik açıdan değerlendirildiğinde ise yine bu çalışma birtakım sınırlamalara sahiptir. Öyle ki, çalışma kapsamında sadece rastgele alt örnekleme sınıf dengesizliği ile baş etme adına uygulanmıştır. İleride gerçekleştirilecek çalışmalarda ise, farklı sınıf dengesizliği etkisini minimize etmek adına başka örnekleme teknikleri de adapte edilebilecektir. Bu noktada, sadece bu tekniklerin performanslara etkisinin değerlendirilmesi dahi önemli bir araştırma konusudur. Öte yandan, çalışma kapsamında sadece ağaç tabanlı makine öğrenmesi algoritmalarından yararlanılmıştır. Kapsamlı bir analizde bu yöntemlere ilaveten başka makine öğrenmesi ailelerinden de algoritmaların (olasılık tabanlı, uzaklık tabanlı, sinir ağları tabanlı vb.) entegre edilmesi oldukça önemli bulgular elde edilmesini potansiyeline sahiptir. Farklı makine öğrenmesi tekniklerine ilaveten, bünyesinde birçok içsel parametre barındıran bu tekniklerin güncel koşulları daha iyi modellenmesine olanak sağlayabilecek optimizasyon tekniklerinin de ani sel haritalamalarına katkı sağlayabileceğine inanılmaktadır. Ek olarak, değişken sayılarının hassasiyet analizleri ile düşürülerek daha az girdi değişkeni ile etkin sonuçlar alınıp alınamayacağı araştırılması da incelenebilecek önemli hususlardandır. Gelecek çalışmalarda ayrıca, sel olaylarının verisel bazda sayılarının artırılması ile etkin rol oynayabilecek derin öğrenme algoritmalarının uygulamada performans değerlendirmesine tabi tutulması da ilgili literatüre ışık tutucu olacaktır. Sonuçların yorumlanabilirliği adına her ne kadar önemli bir adım atılsa da değişken önemlerini daha detaylı sunan ve değişkenlerin kendi değerlerinde meydana gelen değişimlerin sel olaylarında ne gibi varyasyonlara neden olduğunu daha derinden incelemek için SHAP algoritması gibi adapte edilen makine öğrenmesi algoritmasından bağımsız çalışan ve gelişmiş yöntemlerin uygulanması da önemlidir. Fakat, çalışmanın ifade edilen tüm bu kısıtlarına rağmen, bu çalışmanın özellikle karar vericilere ve politika yapıcılara İstanbul ili özelinde sel tehlikesinin etkilerinin azaltılabilmesi adına ilk olarak odaklanması gereken bölgeleri sunması adına yol gösterici nitelikte olduğuna inanılmaktadır. Ayrıca, sel olaylara dair kayıtların titizlikle tutularak koordinat bazında tüm bu verilerin envanterlere dahil edilmesi bu ve benzeri çalışmaların doğruluk oranının artırılmasına ve daha etkin sel duyarlılık, tehlike, hasar görülebilirlik ve risk yönetimine katkı sağlayacaktır.

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*, 37(19), 5479–5496. <https://doi.org/10.1080/10106049.2021.1920636>
- Anılan, T., Durmuş, H., Akçalı, E., & Yüksek, M. (2021). Taşkın farkındalık ve erken uyarı sistemleri değerlendirilmesi: Trabzon Beşikdüzü örneği. *Doğal Afetler ve Çevre Dergisi*, 7(1), 110–123. <https://doi.org/10.21324/dacd.722798>
- Avand, M., Khiavi, A. N., Khazaei, M., & Tiefenbacher, J. P. (2021). Determination of flood probability and prioritization of sub-watersheds: A comparison of game theory to machine learning. *Journal of Environmental Management*, 295, Article 113040. <https://doi.org/10.1016/j.jenvman.2021.113040>

- Aydin, H. E., & Iban, M. C. (2022). Predicting and analyzing flood susceptibility using boosting-based ensemble machine learning algorithms with SHapley Additive exPlanations. *Natural Hazards*, 116(3), 2957–2991. <https://doi.org/10.1007/s11069-022-05793-y>
- Başakın, E. E., Ekmekcioğlu, M., Çıtakoğlu, H., & Özger, M. (2021). A new insight to the wind speed forecasting: robust multi-stage ensemble soft computing approach based on pre-processing uncertainty assessment. *Neural Computing and Applications*, 34(1), 783–812. <https://doi.org/10.1007/s00521-021-06424-6>
- Bentéjac, C., Csörgő, A., & Martínez-Muñoz, G. (2020). A comparative analysis of gradient boosting algorithms. *Artificial Intelligence Review*, 54(3), 1937–1967. <https://doi.org/10.1007/s10462-020-09896-5>
- Bhattacharya, S., S, S. R. K., Maddikunta, P. K. R., Kaluri, R., Singh, S., Gadekallu, T. R., Alazab, M., & Tariq, U. (2020). A Novel PCA-Firefly Based XGBoost Classification Model for Intrusion Detection in Networks Using GPU. *Electronics*, 9(2), Article 219. <https://doi.org/10.3390/electronics9020219>
- Breiman, L. (2001). Random forests. *Machine Learning*, 45, 5–32.
- Chakraborty, R., Chandra Pal, S., Rezaie, F., Arabameri, A., Lee, S., Roy, P., Saha, A., Chowdhuri, I., & Moayedi, H. (2021). Flash-flood hazard susceptibility mapping in Kangsabati River Basin, India. *Geocarto International*, 37(23), 6713–6735. <https://doi.org/10.1080/10106049.2021.1953618>
- Chen, T., & Guestrin, C. (2016). XGBoost: a scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. <https://doi.org/10.1145/2939672.2939785>
- Choi, J., Gu, B., Chin, S., & Lee, J. S. (2020). Machine learning predictive model based on national data for fatal accidents of construction workers. *Automation in Construction*, 110, Article 102974. <https://doi.org/10.1016/j.autcon.2019.102974>
- Costache, R., Pham, Q. B., Avand, M., Thuy Linh, N. T., Vojtek, M., Vojteková, J., Lee, S., Khoi, D. N., Thao Nhi, P. T., & Dung, T. D. (2020). Novel hybrid models between bivariate statistics, artificial neural networks and boosting algorithms for flood susceptibility assessment. *Journal of Environmental Management*, 265, Article 110485. <https://doi.org/10.1016/j.jenvman.2020.110485>
- Demir, V., & Ülke Keskin, A. (2022a). Taşkınların ekonomik zararlarının değerlendirilmesi (Samsun-Mert Irmağı havzası). *Uluslararası Mühendislik Araştırma ve Geliştirme Dergisi*, 14(2), 663–678. <https://doi.org/10.29137/umagd.1090447>
- Demir, V., & Ülke Keskin, A. (2022b). Taşkın tehlike haritalarının oluşturulması (Samsun, Mert Irmağı örneği). *Türkiye Coğrafi Bilgi Sistemleri Dergisi*, 4(1), 47–54. <https://doi.org/10.56130/tucbis.1120501>
- Demir, V., & Ülke Keskin, A. (2022c). Yeterince akım ölçümü olmayan nehirlerde taşkın debisinin hesaplanması ve taşkın modellemesi (Samsun, Mert Irmağı örneği). *Geomatik*, 7(2), 149–162. <https://doi.org/10.29128/geomatik.918502>
- Deroliya, P., Ghosh, M., Mohanty, M. P., Ghosh, S., Rao, K. D., & Karmakar, S. (2022). A novel flood risk mapping approach with machine learning considering geomorphic and socio-economic vulnerability dimensions. *Science of the Total Environment*, 851, Article 158002. <https://doi.org/10.1016/j.scitotenv.2022.158002>
- Devi, K. K., & Kumar, G. A. S. (2022). Stochastic gradient boosting model for twitter spam detection. *Computer Systems Science and Engineering*, 41 (2), 849–859. <https://doi.org/10.32604/csse.2022.020836>
- Ekmekcioğlu, M., 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*, 26(8), 3547–3565. <https://doi.org/10.1080/19648189.2020.1805024>
- Ekmekcioğlu, M., & Koc, K. (2022a). Explainable step-wise binary classification for the susceptibility assessment of geo-hydrological hazards. *Catena*, 216, Article 106379. <https://doi.org/10.1016/j.catena.2022.106379>
- Ekmekcioğlu, M., Koc, K., & Özger, M. (2022b). Towards flood risk mapping based on multi-tiered decision making in a densely urbanized metropolitan city of Istanbul. *Sustainable Cities and Society*, 80, Article 103759. <https://doi.org/10.1016/j.scs.2022.103759>
- Ekmekcioğlu, M., Koc, K., Özger, M., & Işık, Z. (2022c). 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, Article 127877. <https://doi.org/10.1016/j.jhydrol.2022.127877>
- ESRI. (2016, Haziran 30). *What's new in ArcGIS 10.3.1 for Desktop*. 30 Haziran 2016'da <https://desktop.arcgis.com/en/arcmap/10.3/main/get-started/whats-new-in-arcgis-1031.htm> adresinden alındı.
- Feng, Y., Wang, D., Yin, Y., Li, Z., & Hu, Z. (2020). An XGBoost-based casualty prediction method for terrorist attacks. *Complex & Intelligent Systems*, 6(3), 721–740. <https://doi.org/10.1007/s40747-020-00173-0>
- Fiori, A., & Volpi, E. (2020). On the Effectiveness of LID Infrastructures for the attenuation of urban flooding at the catchment scale. *Water Resources Research*, 56(5), Article e2020WR027121. <https://doi.org/10.1029/2020wr027121>
- Friedman, J. H. (2002). Stochastic gradient boosting. *Computational Statistics & Data Analysis*, 38(4), 367–378. [https://doi.org/10.1016/s0167-9473\(01\)00065-2](https://doi.org/10.1016/s0167-9473(01)00065-2)
- Thippa Reddy G., Swarna Priya R.M., Parimala M., Chowdhary, C. L., M., P. K. R., Hakak, S., & Khan, W. Z. (2020). A deep neural networks based model for uninterrupted marine environment monitoring. *Computer Communications*, 157, 64–75. <https://doi.org/10.1016/j.comcom.2020.04.004>
- Ghosh, A., Dey, P., & Ghosh, T. (2022). Integration of RS-GIS with Frequency Ratio, Fuzzy Logic, Logistic Regression and Decision Tree Models for Flood Susceptibility Prediction in Lower Gangetic Plain: A Study on Malda District of West Bengal, India. *Journal of the Indian Society of Remote Sensing*, 50(9), 1725–1745. <https://doi.org/10.1007/s12524-022-01560-5>
- Guo, F., Zhou, X., Liu, J., Zhang, Y., Li, D., & Zhou, H. (2019). A reinforcement learning decision model for online process parameters optimization from offline data in injection molding. *Applied Soft Computing*, 85, Article 105828. <https://doi.org/10.1016/j.asoc.2019.105828>
- Güvel, Ş. P., Akgül, M. A., & Akkoyunlu, M. F. (2023). Taşkın Riski Değerlendirmesi Kapsamında Landsat-8 Uydu Verileri ile 2015 Yılı Devrek Zonguldak Heyelanının İzlenmesi ve Değerlendirilmesi. *Doğal Afetler ve Çevre Dergisi*, 9(1), 81–89. <https://doi.org/10.21324/dacd.1152670>
- Haltas, I., Yildirim, E., Oztas, F., & Demir, I. (2021). A comprehensive flood event specification and inventory: 1930–2020 Turkey case study. *International Journal of Disaster Risk Reduction*, 56, Article 102086. <https://doi.org/10.1016/j.ijdr.2021.102086>

- Hitouri, S., Varasano, A., Mohajane, M., Ijlil, S., Essahlaoui, N., Ali, S. A., Essahlaoui, A., Pham, Q. B., Waleed, M., Palateerdham, S. K., & Teodoro, A. C. (2022). Hybrid machine learning approach for gully erosion mapping susceptibility at a watershed scale. *ISPRS International Journal of Geo-Information*, 11(7), Article 401. <https://doi.org/10.3390/ijgi11070401>
- Izonin, I., Tkachenko, R., Shakhovska, N., Ilchyshyn, B., & Singh, K. K. (2022). A two-step data normalization approach for improving classification accuracy in the medical diagnosis domain. *Mathematics*, 10(11), Article 1942. <https://doi.org/10.3390/math10111942>
- 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), Article 145. <https://doi.org/10.1038/s41597-019-0155-x>
- Kamalov, F. (2020). Kernel density estimation based sampling for imbalanced class distribution. *Information Sciences*, 512, 1192–1201. <https://doi.org/10.1016/j.ins.2019.10.017>
- Kang, K., & Ryu, H. (2019). Predicting types of occupational accidents at construction sites in Korea using random forest model. *Safety Science*, 120, 226–236. <https://doi.org/10.1016/j.ssci.2019.06.034>
- Kim, D., & Byun, J. (2022). Selection of augmented data for overcoming the imbalance problem in facies classification. *IEEE Geoscience and Remote Sensing Letters*, 19, 1–5. <https://doi.org/10.1109/lgrs.2021.3103997>
- Koc, K., Ekmekcioğlu, M., & Gurgun, A. P. (2021). Integrating feature engineering, genetic algorithm and tree-based machine learning methods to predict the post-accident disability status of construction workers. *Automation in Construction*, 131, Article 103896. <https://doi.org/10.1016/j.autcon.2021.103896>
- Koc, K., Ekmekcioğlu, M., & Gurgun, A. P. (2022). Prediction of construction accident outcomes based on an imbalanced dataset through integrated resampling techniques and machine learning methods. *Engineering, Construction and Architectural Management*, 30(9), 4486–4517. <https://doi.org/10.1108/ecam-04-2022-0305>
- Kumar, S., Gupta, S., & Arora, S. (2022). A comparative simulation of normalization methods for machine learning-based intrusion detection systems using KDD Cup'99 dataset. *Journal of Intelligent & Fuzzy Systems*, 42(3), 1749–1766. <https://doi.org/10.3233/jifs-211191>
- Liang, X., Jiang, A., Li, T., Xue, Y., & Wang, G. (2020). LR-SMOTE — An improved unbalanced data set oversampling based on K-means and SVM. *Knowledge-Based Systems*, 196, Article 105845. <https://doi.org/10.1016/j.knosys.2020.105845>
- Meteoroloji Genel Müdürlüğü. (2023, Ocak 2023) de <https://www.mgm.gov.tr/veridegerlendirme/il-ve-ilceler-istatistik.aspx?k=H> adresinden alındı.
- Moisen, G. G., Freeman, E. A., Blackard, J. A., Frescino, T. S., Zimmermann, N. E., & Edwards, T. C. (2006). Predicting tree species presence and basal area in Utah: A comparison of stochastic gradient boosting, generalized additive models, and tree-based methods. *Ecological Modelling*, 199(2), 176–187. <https://doi.org/10.1016/j.ecolmodel.2006.05.021>
- Nefeslioglu, H. A., Sezer, E., Gokceoglu, C., Bozkir, A. S., & Duman, T. Y. (2010). Assessment of landslide susceptibility by decision trees in the metropolitan area of Istanbul, Turkey. *Mathematical Problems in Engineering*, 2010, Article 901095. <https://doi.org/10.1155/2010/901095>
- 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>
- Özcan, O. (2017). Taşkın tespitinin farklı yöntemlerle değerlendirilmesi: Ayamama deresi örneği. *Doğal Afetler ve Çevre Dergisi*, 3(1), 9–27. <https://doi.org/10.21324/dacd.267200>
- Pham, Q. B., Ekmekcioğlu, M., Ali, S. A., Koc, K., & Parvin, F. (2023). Examining the role of class imbalance handling strategies in predicting earthquake-induced landslide-prone regions. *Applied Soft Computing*, 143, Article 110429. <https://doi.org/10.1016/j.asoc.2023.110429>
- Provost, F. (2000). *Machine learning from imbalanced data sets 101* [Invited paper]. AAAI'2000 Workshop on Imbalanced Data Sets, Menlo Park, CA.
- Sarker, I. H., Alqahtani, H., Alsolami, F., Khan, A. I., Abushark, Y. B., & Siddiqui, M. K. (2020). Context pre-modeling: an empirical analysis for classification based user-centric context-aware predictive modeling. *Journal of Big Data*, 7(1), Article 51. <https://doi.org/10.1186/s40537-020-00328-3>
- Şen, Z. (2009). *İklim değişikliği içerikli taşkın afet ve modern hesaplama yöntemleri*. Su Vakfı Yayınları.
- Tehrany, M. S., Lee, M. J., Pradhan, B., Jebur, M. N., & Lee, S. (2014). Flood susceptibility mapping using integrated bivariate and multivariate statistical models. *Environmental Earth Sciences*, 72(10), 4001–4015. <https://doi.org/10.1007/s12665-014-3289-3>
- Tien Bui, D., Hoang, N. D., Pham, T. D., Ngo, P. T. T., Hoa, P. V., Minh, N. Q., Tran, X. T., & 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, 314–326. <https://doi.org/10.1016/j.jhydrol.2019.05.046>
- Tona, A. U., Demir, V., Kuşak, L., & Yakar, M. (2022). Su kaynakları mühendisliğinde CBS'nin kullanımı. *Türkiye Coğrafi Bilgi Sistemleri Dergisi*, 4(1), 23–33. <https://doi.org/10.56130/tucbis.993807>
- Towfiqul Islam, A. R. M., Talukdar, S., Mahato, S., Kundu, S., Eibek, K. U., Pham, Q. B., Kuriqi, A., & Linh, N. T. T. (2021). Flood susceptibility modelling using advanced ensemble machine learning models. *Geoscience Frontiers*, 12(3), Article 101075. <https://doi.org/10.1016/j.gsf.2020.09.006>
- Yu, R., & Abdel-Aty, M. (2014). Analyzing crash injury severity for a mountainous freeway incorporating real-time traffic and weather data. *Safety Science*, 63, 50–56. <https://doi.org/10.1016/j.ssci.2013.10.012>
- Zhang, Z., Beck, M. W., Winkler, D. A., Huang, B., Sibanda, W., & Goyal, H. (2018). Opening the black box of neural networks: methods for interpreting neural network models in clinical applications. *Annals of Translational Medicine*, 6(11), 216–216. <https://doi.org/10.21037/atm.2018.05.32>