

TOPLULUK TABANLI MAKİNE ÖĞRENMESİ ALGORİTMALARI İLE ORMAN YANGINI DUYARLILIK HARİTALAMASI: HATAY ÖRNEĞİ

Furkan BİLÜCAN¹, Alihan TEKE¹, Taşkın KAVZOĞLU²

¹Arş. Gör., Gebze Teknik Üniversitesi, Harita Mühendisliği Bölümü, 41400, Gebze, Kocaeli, f.bilucan2020@gtu.edu.tr, a.teke2020@gtu.edu.tr

² Prof. Dr., Gebze Teknik Üniversitesi, Harita Mühendisliği Bölümü, 41400, Gebze, Kocaeli, kavzoglu@gtu.edu.tr

ÖZET

Orman ekosistemleri, biyoçeşitliliğin korunması, oksijen üretimi ve küresel karbon döngüsünün sağlanması gibi hayati doğal aktivitelerin sürdürülebilmesinde önemli bir rol oynamaktadır. Öte yandan, özellikle son yıllarda iklim ve arazi kullanımındaki değişiklikler, bazı antropojenik faaliyetler, küresel ısınma ve iklim değişikliği gibi faktörler, Akdeniz tipi iklim bölgelerinde yangın görülme sıklığının ve şiddetinin artmasına neden olmuştur. Potansiyel orman yangını bölgelerinin doğru bir şekilde tespit edilmesinde ve etkilerinin minimize edilebilmesinde yangın duyarlılık haritaları literatürde sıklıkla tercih edilen önemli araçlardan birisidir. Bu çalışmada, CatBoost, LightGBM ve Rastgele Orman olmak üzere üç makine öğrenme algoritması ile Türkiye'nin güney kıyı bölgesinde yer alan ve stratejik bir öneme sahip olan Hatay iline ait orman yangını duyarlılık haritaları üretilmiştir. Yangın envanter haritası, NASA tarafından 2001 ve 2021 yılları arasında sağlanan açık kaynaklı MODIS verileri ve çalışma alanının ana karakteristiğini yansıtan toplam 11 faktör kullanılarak üretilmiştir. Elde edilen sonuçlara göre, LightGBM algoritması ile en yüksek doğrulukta (%88.31) yangın duyarlılık modeli üretilirken, CatBoost algoritması ile %87.21 ve Rastgele Orman algoritması ile %85.90 doğrulukta duyarlılık modelleri üretilmiştir. Üretilen duyarlılık haritalarının doğrulukları arasındaki farkın istatistiksel olarak anlamlı olup olmadıklarını belirlemek için McNemar's testi uygulanmıştır. İstatistiksel analiz sonuçları CatBoost ve LightGBM yöntemlerinin performanslarının birbirinden istatistiksel olarak anlamsız olduğunu ancak bu iki algoritmanın Rastgele Orman algoritmasından istatistiksel olarak daha iyi sonuçlar ürettiğini ortaya çıkarmıştır. Ayrıca, çalışmada kullanılan her bir faktörün duyarlılık haritası üzerindeki etkilerini araştırmak için SHapley Additive exPlanations (SHAP) yaklaşımı uygulanmıştır. Sonuç olarak, en etkili faktörlerin yükseklik, yağış, bitki örtüsü ve rüzgar hızı olduğu tespit edilirken, en az katkıya sahip faktörler ise baki, topografik ıslaklık indeksi (TWI) ve eğim olarak bulunmuştur. Haritalar tematik olarak yorumlandığında ise, potansiyel orman yangını riskinin en fazla olduğu yerlerin kıyı bölgeleri ile Amanos Dağları'nın doğu kısımları olduğu gözlemlenmiştir.

Anahtar Sözcükler: Orman Yangını, Duyarlılık Haritalama, Makine Öğrenmesi, SHAP, Hatay.

ABSTRACT

FOREST FIRE SUSCEPTIBILITY MAPPING WITH ENSEMBLE-BASED MACHINE LEARNING ALGORITHMS: A CASE STUDY OF HATAY

Forest ecosystems play an important role in sustaining important natural activities such as protecting biodiversity, producing oxygen, and ensuring the global carbon cycle. On the other hand, factors such as climate and land use changes, some anthropogenic activities, global warming, and climate change, have recently led to an increase in the incidence and severity of fires in Mediterranean-type climate regions. Fire susceptibility maps are one of the most preferred tools in the literature for accurately detecting potential forest fire areas and minimizing their effects. In this study, forest fire susceptibility maps of Hatay province, which is located in the southern coastal region of Turkey and has strategic importance, were produced using three machine learning algorithms, namely CatBoost, LightGBM and Random Forest. The fire inventory map was produced using open-source MODIS data provided by NASA (Fire Information for Resource Management System) between 2001 and 2021 and a total of 11 factors representing the main characteristic of the study area. According to the results, the highest accuracy (88.31%) wildfire susceptibility model was produced with the LightGBM algorithm, followed by CatBoost (87.21%) and Random Forest (85.90%). McNemar's test was applied to determine whether the difference between the accuracy of the produced susceptibility maps was statistically significant. Statistical analysis results revealed that the performances of CatBoost and LightGBM methods were statistically insignificant from each other, but these two algorithms produced statistically better results than the Random Forest algorithm. In addition, the SHapley Additive exPlanations (SHAP) approach was applied to investigate the effects of each factor used in the study on the susceptibility map. As a result, it was determined that the most effective factors were elevation, precipitation, vegetation, and wind speed while the aspect, topographic wetness index (TWI), and slope had the least contribution to the model. When the maps were interpreted thematically, it was observed that the areas with the highest potential forest fire risk were the coastal areas and the eastern parts of the Amanos Mountains.

Keywords: Forest Fire, Susceptibility Mapping, Machine Learning, SHAP, Hatay.

1. GİRİŞ

Orman yangınları ekosistemlerin yapısına zarar veren ve bitki örtüsünün doğal döngüsünü etkileyen başlıca doğal afetlerden birisidir. Bunun yanı sıra, toprak özelliklerinin bozulması, su havzalarının hidrolojik davranışındaki değişiklikler, can ve mal kayıpları gibi ciddi sonuçlara neden olmaktadır (Soulis et al., 2021). Küresel ölçekte orman yangınları salınan büyük miktarlardaki karbondioksit ve iklim değişikliğinin önemli faktörlerinden birisi olan sera gazlarının artışında da önemli bir rol oynamaktadır. Türkiye’de ise ormanların çoğu yangın tehdidi altındadır ve toplam orman arazilerinin %60’ı birinci ve ikinci derece yangına duyarlı alanlar olarak sınıflandırılmaktadır (Tonbul vd., 2022). Son yıllarda iklim değişikliğinin de etkisiyle özellikle Akdeniz bölgesinde artan orman yangınları insan yaşamını ve ekosistemleri tehdit eden ciddi bir sosyal ve çevresel sorun haline gelmiştir. Bu nedenlerden dolayı, orman yangınlarının oluşumunu mümkün olduğunca en aza indirmek için doğru stratejilerin geliştirilmesi gerekmektedir. Bu stratejilerin en önemlilerinden biri ise potansiyel yangına duyarlı bölgeleri konumsal olarak gösteren orman yangını duyarlılık haritalarının üretilmesidir. Bu haritalar aynı zamanda hem afet öncesi ve sonrası bir altlık harita olarak hem de erken uyarı sistemleri için de yardımcı bir araç olarak kullanılabilir.

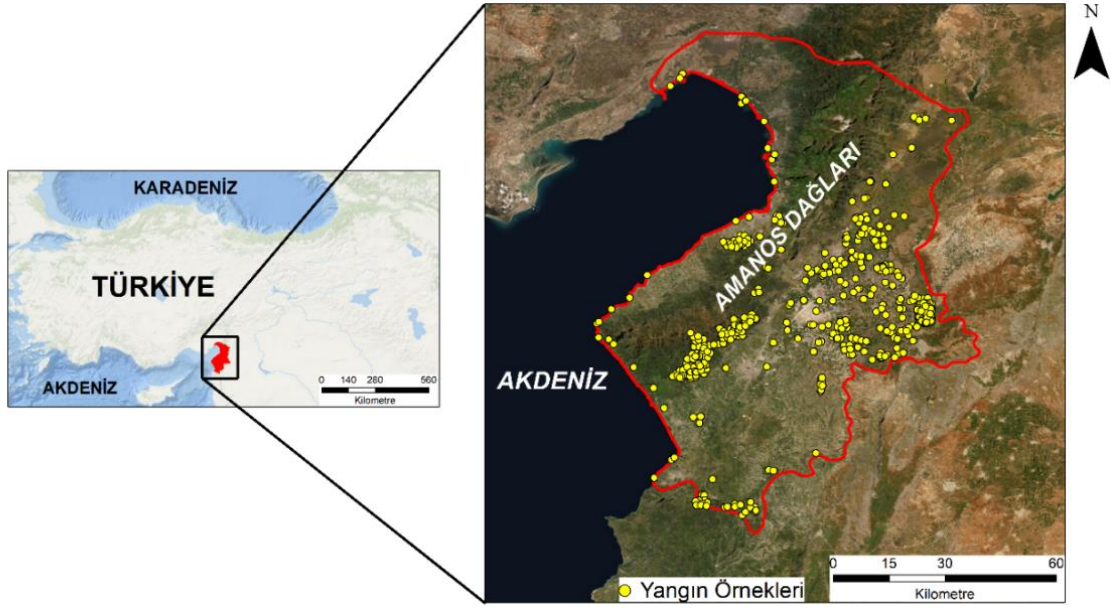
Orman yangını duyarlılık haritalarının kalitesini etkileyen önemli unsurlardan birisi de bu aktivitelerin potansiyel mekânsal dağılımlarını yansıtabilen uygun tekniklerin seçimidir. Özellikle son yıllarda uzaktan algılama, coğrafi bilgi sistemleri ve yapay zeka alanlarındaki gelişmeler genellikle modellenmesi ve tahminlenmesi nispeten zor olan doğal afetlerin potansiyel duyarlılıklarının belirlenmesi çalışmalarında sıklıkla tercih edilmektedir (Iban ve Sekertekin, 2022; Hong vd., 2018). Bu kapsamda, güvenilir ve doğru yangın duyarlılık haritaları üretmek için lojistik regresyon (Pourghasemi, 2016), destek vektör makineleri (Gigović vd., 2019) ve yapay sinir ağları (Bui, 2019) gibi geleneksel makine öğrenme algoritmaları literatürde yaygın olarak kullanılmaktadır. Ancak, bu tür yaklaşımların performansları söz konusu problemin ve çalışma alanının karakteristiğine göre değişebilmektedir. Örneğin, bir makine öğrenmesi algoritması bir duyarlılık tahminleme probleminde yüksek bir doğruluk üretebilirken, farklı bir test sahasında düşük bir performansa sahip olabilir (Kavzoglu ve Teke, 2022a). Bu nedenle, bu tür tekil tahminleme mekanizmasına sahip tekniklerden ziyade son zamanlarda topluluk tabanlı algoritmalar birçok alanda kullanılmaktadır. Bu algoritmaların temel çalışma mekanizmalarında daha iyi bir performans elde edebilmek için farklı tekil algoritmaların tahminlerini farklı oylama stratejileri kullanarak birleştirme prensibi bulunmaktadır. Böylece, herhangi bir tekil algoritma tarafından yapılan yanlış ya da hatalı tahminleme bir diğeri tarafından telafi edilerek daha kararlı modeller oluşturulabilir.

Topluluk tabanlı makine öğrenme algoritmalarının performanslarının orman yangının duyarlılık haritalarının üretimindeki performanslarını karşılaştırmak için bu çalışma kapsamında, CatBoost, LightGBM ve Rastgele Orman olmak üzere toplamda üç teknik kullanılmıştır. Mevcut literatür ve çalışma alanının karakteristiği göz önünde bulundurularak orman yangınlarının oluşumunda önemli role sahip 11 parametre kullanılarak Hatay iline ait yangın duyarlılık haritası oluşturulmuştur. İlk olarak, faktörler arasında potansiyel bir korelasyon olup olmadığı çoklu bağlantı (multicollinearity) testi ile analiz edilmiştir. Daha sonra, bu faktörlerin model üzerindeki katkılarını ölçebilmek için SHAP analizi uygulanmış ve üretilen duyarlılık haritalarının doğrulukları arasındaki fark McNemar’s istatistiksel anlamlılık testi ile incelenmiştir.

2. ÇALIŞMA ALANI VE VERİ SETİ

Hatay 35° 49’ – 37° 00’ doğu boylamları ve 35° 44’ – 36° 43’ Kuzey enlemleri arasında Türkiye’nin güneyinde yer alan bir kıyı kentidir (**Şekil 1**). Şehir kuzeyde Türkiye’deki bitki türlerinin yaklaşık yarısını barındıran Amanos Dağları, batıda Akdeniz ve doğuda ise Suriye sınırı ile çevrilidir. Akdeniz iklim tipinin görüldüğü şehirde yazlar sıcak ve kurak, kışlar ise serin ve yağışlıdır. 2021’de yıllık ortalama sıcaklık 18.3 iken yaz aylarında 42 C°’ nin üstünde sıcaklıklar görülmektedir (<https://www.mgm.gov.tr/>). Ülkemizde 2021 yılında meydana gelen 2793 adet yangınının 64 tanesi Hatay da gerçekleşmiştir ve bu yangınlarda toplam 334.63 hektar alan zarar görmüştür (<https://www.ogm.gov.tr/>).

Çalışma kapsamında NASA tarafından sağlanan açık kaynaklı 2001-2021 yılları arasına ait tarihsel yangın verileri orman yangını envanter haritasının oluşturulması için kullanılmıştır. Buna ek olarak, çalışma alanının temel özellikleri ve geçmiş çalışmalar göz önünde bulundurularak toplamda 11 parametre (yükseklik, yağış, normalize edilmiş fark bitki örtüsü indeksi (NDVI), rüzgar hızı, sıcaklık, vadi derinliği, yollara olan uzaklık, arazi kullanımı/arazi örtüsü (LULC), bakı, TWI ve eğim) seçilmiştir. Çalışma alanındaki yerel istasyonlara ait yağış, rüzgar hızı, sıcaklık verileri Meteoroloji Genel Müdürlüğü’nden temin edilmiştir. Yükseklik, vadi derinliği, bakı, TWI ve eğim parametreleri ise SRTM sayısal yükseklik modeli ile üretilmiştir. NDVI tematik katmanı Landsat 8 OLI uydu görüntülerinin yakın kızılötesi ve kırmızı bantları kullanılarak Google Earth Engine platformunda üretilirken, LULC ise açık kaynak veri sağlayan GlobeLand30 platformundan temin edilmiştir. Yollara olan uzaklık haritası mevcut yol ağı ve Öklid uzaklığı fonksiyonu kullanılarak ArcGIS ortamında tematik olarak elde edilmiştir.



Şekil 1. Çalışma alanının coğrafi konumu.

3. YÖNTEM

3.1. CatBoost

CatBoost algoritması sınıflandırma problemlerini çözmek için Yandex tarafından 2018 yılında geliştirilmiş topluluk tabanlı bir algoritmadır. CatBoost algoritması, klasik gradyan artırma iki şekilde farklılık gösterir. Bunlardan ilki, standart gradyan artırma algoritmasına bir alternatif olan permütasyon yönteminin kullanılmasıyla sıralı artırma uygulamasıdır (Prokhorenkova vd., 2018). Diğeri kategorik özellikleri kodlamak için sıralı hedef istatistiklerini kullanmasıdır. Algoritma iki temel prensibe göre çalışmaktadır. İlk olarak veri seti rastgele alt kümelerle bölünerek etiketler tam sayılara dönüştürülür, ikinci adımda ise kategorik özellikler sayısal hale dönüştürülür. Algoritmanın en önemli avantajı ise aşırı öğrenme sorununu azaltmasıdır. Ayrıca CatBoost makine öğrenme algoritması regresyon ve sınıflandırma problemlerinde kullanılabilir.

3.2. LightGBM

Yeni nesil topluluk öğrenme algoritmalarından biri olan hafif gradyan artırma makineleri (LightGBM) 2017 yılında Microsoft tarafından üretilmiştir. Yüksek işlem hızına sahip olan algoritma, ağaç dallanma yapısının farklı olması sebebiyle diğer gradyan artırma yöntemlerine göre farklılık göstermektedir (Ke et al., 2017). Diğer gradyan artırma yaklaşımlarında karar ağaçları seviye bazlı (level-wise) büyüme stratejisi gösterirken, LightGBM algoritmasında karar ağaçlarının yaprak bazlı (leaf-wise) büyümesi söz konusudur. Ayrıca LightGBM topluluk öğrenme yaklaşımında işlem hızının artmasını sağlayan, gradyan alt yapı tek yönlü örnekleme (Gradient-Based One-Side Sampling, GOSS) ve ayrıcalıklı öznel destekleme (Exclusive Feature Bundling, EFB) olmak üzere iki yeni strateji mevcuttur. GOSS ile doğruluk oranı korunarak, veri sayısında azaltma veri dağılımı çok fazla etkilenmeden gerçekleştirilir. EFB algoritmasında ise seyrek halde bulunan özellikler daha yoğun bir hale getirilir ve işlem karmaşıklığı giderilir.

3.3. Rastgele Orman

Rastgele Orman, karar ağaçlarını birleştirerek karar ormanı oluşturan bir topluluk öğrenme yöntemidir (Breiman, 2001). Söz konusu ormandaki her ağaç, eğitim örneklerinin rastgele seçilmiş bir alt kümesiyle eğitilir ve ağaçların oy çokluğuna bağlı olarak sınıflandırma işlemi gerçekleştirilir. Alt kümelerin üçte ikisi karar ağaç yapısını oluşturmak için kullanılırken, diğer üçte birlik kısım ise ağaç yapısının geçerliliğini test etmek için çapraz doğrulama sürecinde kullanılır (Kavzoglu vd., 2020). Karar ağaçlarının her biri sınıflandırma işlemi sonucunda bir oy alır ve en çok oyu alan ağaçların tespit edilmesine bağlı olarak ağaç yapısı belirlenir. Rastgele Orman yönteminde dallanma şeklinin belirlenmesi, bir özelliğin safsızlığının ölçüsü olan Gini indeksi metriği ile gerçekleşir (Kavzoglu ve Colkesen, 2013). Ayrıca algoritmanın, her düğümü bölmek için kullanılan örnek sayısı ile ağaç sayısı olmak üzere iki önemli parametresi bulunmaktadır.

3.4. McNemar's Testi

Sınıflandırma işlemlerinde kullanılan algoritmaların çalışma prensiplerinin ve performanslarının aynı olmamasının doğal bir sonucu olarak hesaplanan genel doğruluklar birbirinden farklılık göstermektedir. Elde edilen farklı doğruluk oranlarının değerlendirilmesi için Ki-kare dağılım yöntemine dayanan McNemar's (χ^2) testi kullanılmaktadır. Özellikle, uzaktan algılama uygulamalarında metod performanslarının analizinde sıklıkla başvurulanan bir istatistiksel testtir (Ouyang ve Ma 2006; Manandhar vd., 2009; Pal vd., 2013). Parametrik olmayan McNemar's testinin hesaplanması için 2x2 boyutlu bir matris kullanılmaktadır (Foody, 2004). Değerlendirme yapılırken belirlenen güven aralığına göre Ki-kare tablo dağılımı esas alınır ve hesaplanan değer tabloda bulunan değerden büyük ise iki doğruluk arasındaki farkın istatistiksel olarak anlamlı olduğu sonucuna ulaşılır. Ayrıca güven aralığının %95 olarak belirlendiği durumlarda, söz konusu tablo değeri 3.84'tür. McNemar's test hesabının gösterildiği eşitlik 1'de, i sınıflandırma yöntemi ile hatalı ve j sınıflandırma yöntemi ile doğru tahmin edilen piksel sayısı n_{ij} ile temsil edilmiştir. Ek olarak, j sınıflandırma algoritması tarafından hatalı ve i sınıflandırma algoritması tarafından doğru tahmin edilen piksel sayısı ise n_{ji} ile ifade edilmiştir.

$$\text{McNemar's } (\chi^2) = \frac{(|n_{ij} - n_{ji}| - 1)^2}{n_{ij} + n_{ji}} \quad (1)$$

3.5. SHAP Analizi

SHAP (SHapley Additive exPlanations) algoritması (Lundberg ve Lee, 2017), oyun teorisi yaklaşımına bağlı olarak, modeldeki her bir değişkenin modelin sonucu üzerindeki etkisinin, boyutunun ve türünün değerlendirilmesini sağlar (Kopeć et al. 2020). Eklenebilir özellik atama yönteminin kullanıldığı algoritmada, Shapley değerleri özelliklerin tüm olası permütasyonları boyunca artan katkının ortalamasına bağlı olarak hesaplanır. SHAP değerleri ana iki kısımdan oluşmaktadır: (i) modelin sonucunu ifade eden oyun, (ii) modelin özelliklerinin temsil edildiği oyuncular. Bu yüzden, modelde bulunan her özelliğin modele olan katkısının tahmin edilmesiyle genel yorumlanabilirliğin yanı sıra özellikler için yerel yorumlanabilirlikte sağlamaktadır (Kavzoglu ve Teke, 2022b). Her gözlemdaki modelin sonucu Eşitlik 2'deki gibi toplam SHAP değeri ile tahmin edilmektedir. Eşitlikte ϕ_i , i özellik sayısının katkısı, $g(z')$ açıklayıcı modeli, M ise girdi özellik sayısını temsil etmektedir.

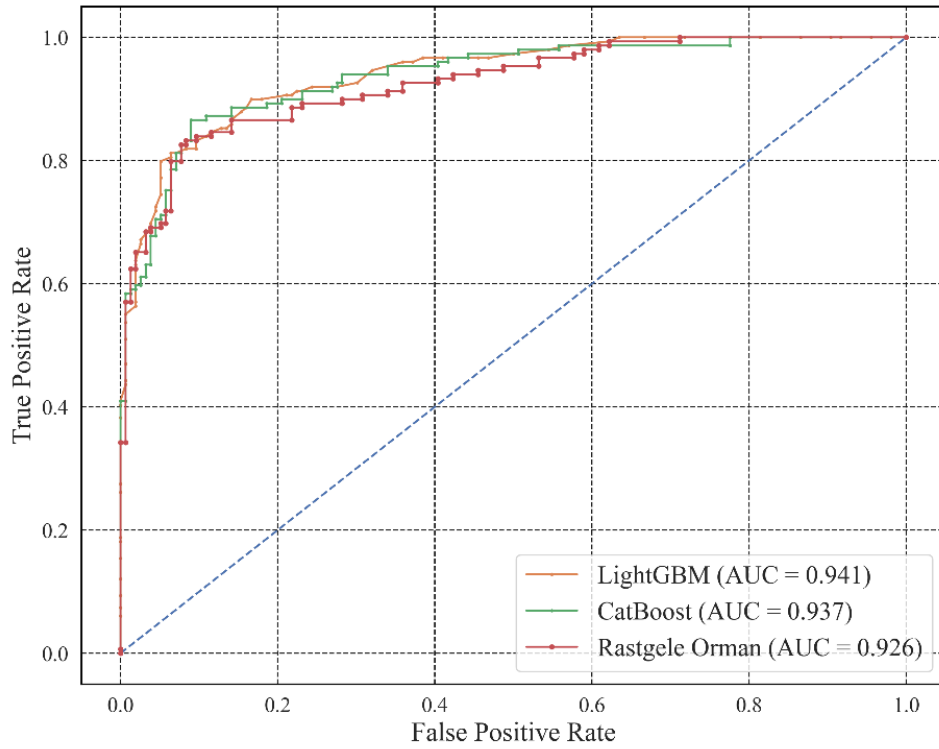
$$g(z') = \phi_0 + \sum_{i=1}^M \phi_i z'_i \quad (1)$$

4. UYGULAMA

Bu çalışmada, üç topluluk tabanlı makine öğrenme algoritması (CatBoost, LightGBM, ve Rastgele Orman) kullanılarak Türkiye'nin güneyinde bulunan Hatay iline ait yangın duyarlılık haritaları üretilmiştir. Bu amaç doğrultusunda ilk olarak çoklu doğrusallık testi yapılarak varyans artırıcı faktör (VIF) ve tolerans (TOL) metrikleri incelenmiştir. Genellikle, hesaplanan VIF değeri 10'dan büyükse ya da TOL değeri 0.1'den küçükse, çoklu doğrusallık için önemli bir gösterge olarak kabul edilir. Bu çalışmada ise maksimum VIF değeri 2.656 olarak bulunurken, en düşük TOL değeri 0.376 olarak hesaplanmış, yani faktörler arasında istatistiksel çoklu doğrusallık problemine rastlanmamıştır (Çizelge 1).

Çizelge 1. Çoklu doğrusallık testi analizi.

Faktör	TOL	VIF
Yükseklik	0.376	2.656
Eğim	0.507	1.974
Yollara olan uzaklık	0.567	1.762
Yağış	0.593	1.686
Sıcaklık	0.610	1.640
TWI	0.668	1.498
NDVI	0.780	1.282
Valley Depth	0.882	1.134
Rüzgar Hızı	0.911	1.097
LULC	0.958	1.044
Baki	0.975	1.025



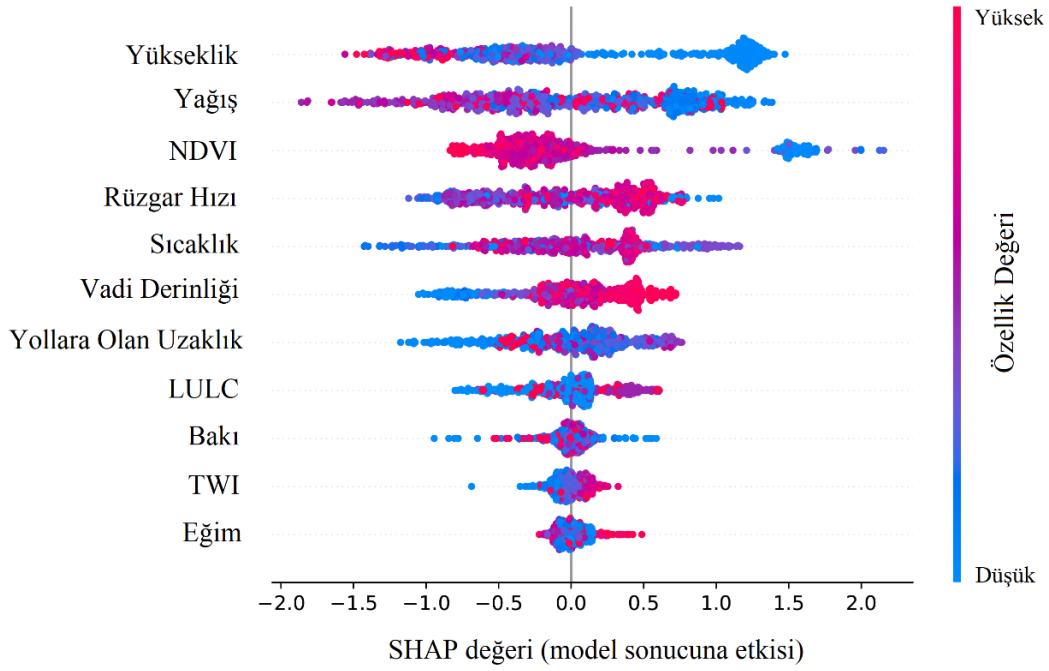
Şekil 2. ROC Eğrisi.

Farklı makine öğrenmesi algoritmaları ile üretilen haritaların kalitelerini ve doğruluklarını değerlendirmek için genel doğruluk ve alıcı işlem karakteristiği eğrisi olmak üzere iki metrik kullanılmıştır. Elde edilen sonuçlara göre, LightGBM algoritması ile en yüksek doğrulukta (%88.31) yangın duyarlılık modeli üretilirken, CatBoost algoritması ile %87.21 ve Rastgele Orman algoritması ile %85.90 doğrulukta duyarlılık modelleri üretilmiştir. Genel doğruluk skorlarına ek olarak, genellikle iki sınıfa sahip bağımlı değişkenlerin söz konusu olduğu problemlerde sıklıkla kullanılan alıcı işlem karakteristiği eğrisi (ROC) altında kalan alan metriği makine öğrenmesi algoritmalarının performanslarını ölçmek için kullanılmıştır. Temel olarak bu metrik bir algoritmanın veri kümesi içindeki etiketli negatif ve pozitif örnekleri ne kadar başarılı bir şekilde ayırt edebildiğinin bir göstergesidir. Eğri altında kalan alan 1 değerine yaklaştıkça algoritmanın sınıflar arası yüksek bir ayırım kabiliyetine sahip ideal bir model olduğu çıkarımı yapılır. Genel doğruluk skorlarına benzer şekilde, LightGBM 0.941 ile en yüksek skora sahip iken, CatBoost (0.937) ve Rastgele Orman (0.926) algoritmalarının daha düşük bir performansa sahip olduğu gözlemlenmiştir (Şekil 2).

Çizelge 2. McNemar's istatistiksel anlamlılık testi.

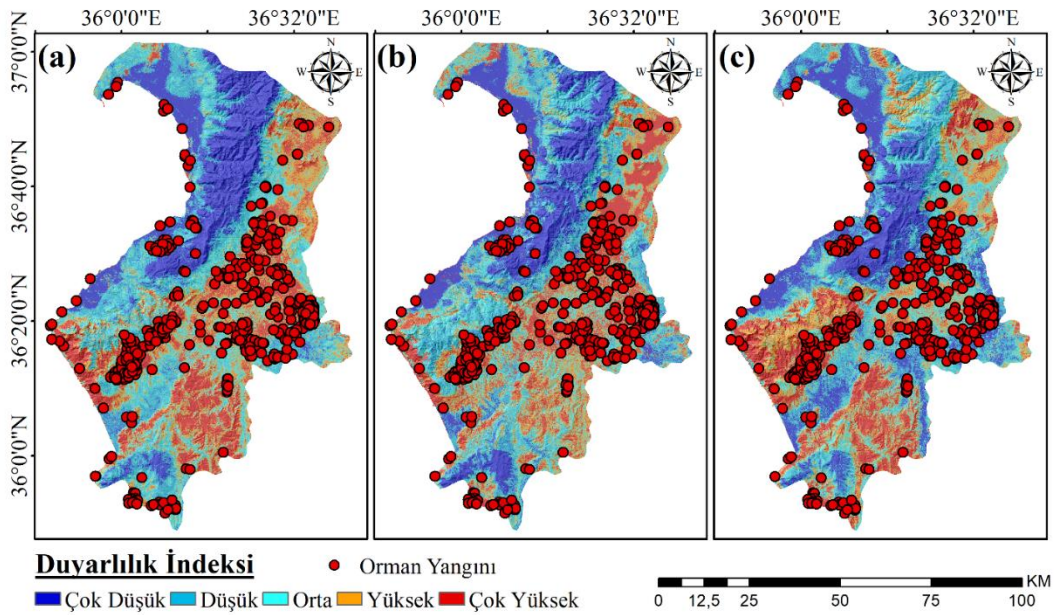
	LightGBM	CatBoost	Rastgele Orman
LightGBM	-	0.025	8.323
CatBoost		-	7.993
Rastgele Orman			-

Çalışma kapsamında sayısal olarak algoritmalarının performansları arasındaki farkın incelenmesine ek olarak bu farkın istatistiksel olarak anlamlı bir değer olup olmadığı McNemar's testi ile analiz edilmiştir. Sonuçlar LightGBM'in diğer algoritmalarla göre daha üstün bir performans göstermiş olsa da istatistiksel anlamlılık testi sonuçlarına göre LightGBM ve CatBoost algoritmaları için hesaplanan χ^2 değeri 0.025 olarak bulunmuştur (Çizelge 2). Yani, bu iki algoritma ile üretilen haritaların doğrulukları arasındaki farkın (yaklaşık %1) istatistiksel olarak anlamsız olduğu gözlemlenmiştir. Buna karşın, Rastgele Orman algoritmasının LightGBM ve CatBoost algoritmaları için hesaplanan χ^2 değerlerinin anlamlı olduğu sonucuna ulaşılmıştır. Diğer bir ifade ile LightGBM ve CatBoost algoritmaları ile Rastgele Orman arasındaki performans farkının istatistiksel olarak anlamlı olduğu bulunmuştur. Diğer taraftan, LightGBM ile Rastgele Orman algoritması için hesaplanan χ^2 değeri 8.322 ve CatBoost ile Rastgele Orman arasında hesaplanan değerin ise 7.993 olduğu Çizelge 2'den görülmektedir. Bu değerler %95 güven aralığında tablo değeri olan 3.84' ten büyük oldukları için aralarındaki farkın istatistiksel olarak anlamlı oldukları sonucuna ulaşılmaktadır.



Şekil 3. SHAP özet grafiği.

Bir makine öğrenmesi modelinin eğitiminden sonra yapay zekanın karar verme mekanizmasında neler olduğunu anlamak, model çıktılarının doğru yorumlanabilmesi için son derece önemli adımlardan birisidir. Çalışma kapsamında, faktörler ve oluşturulan model üzerinde anlamlı bilgilerin çıkarılabilmesi amacıyla kökeni oyun teorisine dayanan SHAP yaklaşımı uygulanmıştır. SHAP özet grafiğinde, yatay eksen, dikey ekseninde var olan bağımsız değişkenlerin kurulan model üzerindeki etkisinin büyüklüğünü temsil eder. Öznitelik seçimi tekniklerinden farklı olarak, bu analiz ile değişkenlerin bireysel olarak önem skorlarının hesaplanmasının haricinde, düşük ve yüksek değerlerinin de model üzerindeki etkileri değerlendirilebilir. SHAP analizi sonucunda model doğruluğu üzerinde en büyük etkiye sahip faktör yükseklik olurken, en az etkili olanın eğim faktörü olduğu belirlenmiştir (Şekil 3). Buna ek olarak, iklimsel faktörlerin (yağış, rüzgar hızı ve sıcaklık) en önemli ilk beş faktör içinde yer alması çalışma alanında yangın oluşumunu önemli ölçüde etkilediklerini gösteren önemli bulgulardan biridir.



Şekil 4. Orman yangını duyarlılık haritaları: (a) LightGBM, (b) CatBoost, (c) Rastgele Orman.

Orman yangını duyarlılık haritaları çok düşük, düşük, orta, yüksek ve çok yüksek olmak üzere dağılım (quantile) tabanlı sınıflandırıcı kullanılarak toplamda 5 duyarlılık dilimine ayrılmıştır. Üretilen haritalar tematik olarak incelendiğinde, Hatay'ın kuzeyi ve Amanos Dağları yangına duyarlılık bakımından daha düşük seviyede

gruplandırılırken, genel olarak kıyı ve Amanos Dağları'nın doğu kesimleri yüksek risk grubunda yer almıştır (Şekil 4). Ancak yangın duyarlılık bölgelerinin dağılımında makine öğrenmesi algoritmalarına bağlı olarak birtakım farklılıklar görmekte mümkündür. Örneğin, test sahasının güney batı bölgeleri LightGBM ve CatBoost algoritmaları tarafından çoğunlukla çok yüksek ve orta duyarlılık sınıflarına atanmışken, bu bölgeler Rastgele Orman algoritması tarafından genellikle çok yüksek ve yüksek duyarlılık sınıfları olarak gruplandırılmıştır. Duyarlılık sınıfları arasındaki bir diğer farklılığa ise çalışma alanının kuzey bölümünde rastlanmıştır.

5. SONUÇLAR

Bu çalışmada Türkiye'nin Hatay ili için yangın duyarlılık haritası üç topluluk öğrenme algoritması (CatBoost, LightGBM ve Rastgele Orman) ile üretilerek analiz edilmiştir. Çalışmada Hatay ilinin topografik, çevresel ve meteorolojik özelliklerini temsil eden toplamda 11 faktör (yükseklik, yağış, NDVI, rüzgar hızı, sıcaklık, vadi derinliği, yollara olan uzaklık, LULC, bakı, TWI ve eğim) bağımsız değişken olarak seçilmiştir. Uygulanan topluluk öğrenme algoritmalarının performansları karşılaştırıldığında en yüksek doğruluğa (%88.31) LightGBM yönteminin sahip olduğu sonucuna ulaşılmıştır. Elde edilen sonuçların anlamlı olup olmadıklarının analizi McNemar's testi ile yapılmıştır. Ayrıca, her bir parametrenin kurulan model üzerinde nasıl ve ne kadar etki ettiğinin değerlendirilmesini yapmak için kökeni oyun teorisine dayanan SHAP analizi uygulanmıştır. Analiz sonucunda, duyarlılık haritası üretiminde en etkili faktörlerin yükseklik, yağış, bitki örtüsü ve rüzgar hızı olduğu belirlenirken, en az etkili olanların ise bakı, TWI ve eğim olduğu tespit edilmiştir. Çalışmanın genel sonucu olarak, potansiyel orman yangını riskinin daha fazla olduğu bölgelerin kıyı kesimleri ile Amanos Dağları'nın doğu kısımları olduğu bulgusuna haritaların tematik olarak yorumlanmasıyla ulaşılmıştır.

KAYNAKLAR

- Breiman, L.**, 2001, Random forests. *Machine learning* 45:5–32.
- Bui, Q-T.**, 2019, Metaheuristic algorithms in optimizing neural network: a comparative study for forest fire susceptibility mapping in Dak Nong, Vietnam. *Geomatics, Natural Hazards and Risk*, 10:136–150.
- Foody, G.M.**, 2004, Thematic map comparison: Evaluating the statistical significance of differences in classification accuracy. *Photogrammetry Engineering and Remote Sensing*, 70(5): 627-633.
- Gigović, L., Pourghasemi, H.R., Drobnjak, S., Bai, S.**, 2019, Testing a New Ensemble Model Based on SVM and Random Forest in Forest Fire Susceptibility Assessment and Its Mapping in Serbia's Tara National Park. *Forests* 10(5):408.
- Hong, H., Tsangaratos, P., Ilia, I., Liu, J., Zhu, A., Xu, C.**, 2018, Applying Genetic Algorithms to Set the Optimal Combination of Forest Fire Related Variables and Model Forest Fire Susceptibility Based on Data Mining Models. The Case of Dayu County, China. *Science of The Total Environment*, 630:1044–1056.
- Iban, M.C., Sekertekin, A.**, 2022, Machine learning based wildfire susceptibility mapping using remotely sensed fire data and GIS: A case study of Adana and Mersin provinces, Turkey. *Ecological Informatics*, 69:101647.
- Kavzoglu, T., Bilucan, F.**, 2022, Effects of auxiliary and ancillary data on LULC classification in a heterogeneous environment using optimized random forest algorithm. *Earth Science Informatics*, Doi:10.1007/s12145-022-00874-9.
- Kavzoglu, T., Bilucan, F., Teke A.**, 2020, Comparison of support vector machines, random forest and decision tree methods for classification of Sentinel - 2A image using different band combinations. *41st Asian Conference on Remote Sensing (ACRS 2020)*, Deqing, China.
- Kavzoglu, T., Colkesen, I.**, 2013, An assessment of the effectiveness of a rotation forest ensemble for land-use and land-cover mapping. *International Journal of Remote Sensing*, 34:4224–4241.
- Kavzoglu, T., Teke, A.**, 2022a, Predictive Performances of Ensemble Machine Learning Algorithms in Landslide Susceptibility Mapping Using Random Forest, Extreme Gradient Boosting (XGBoost) and Natural Gradient Boosting (NGBoost). *Arabian Journal for Science and Engineering*, 47:7367–7385.
- Kavzoglu, T., Teke, A.**, 2022b, Advanced hyperparameter optimization for improved spatial prediction of shallow landslides using extreme gradient boosting (XGBoost). *Bulletin of Engineering Geology and the Environment*, 81:201.
- Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., Liu, T. Y.**, 2017, Lightgbm: A highly efficient gradient boosting decision tree. *Advances in Neural Information Processing Systems*, 3146-3154.

- Kopeć, A., Trybala, P., Głabicki, D., Buczyńska, A., Owczarz, K., Bugajska, N., Kosińska, P., Chojwa, M., Gattner A.,** 2020, Application of remote sensing, gis and machine learning with geographically weighted regression in assessing the impact of hard coal mining on the natural environment. *Sustainability*, 12(22):1–26.
- Lundberg, S.M., Lee, S.I.,** 2017, A unified approach to interpreting model predictions. *Advance in Neural Information Processing System*, 30:4766–4775.
- Manandhar, R., Odeh, I.O.A., Ancev, T.,** 2009, Improving the Accuracy of Land Use and Land Cover Classification of Landsat Data Using Post-Classification Enhancement. *Remote Sensing*, 1(3):330–344.
- Ouyang, Y., Ma, J.,** 2006, Classification of multi-spectral remote sensing data using a local transfer function classifier. *International Journal of Remote Sensing*, 27:5401–5408.
- Pal, M., Maxwell, A.E., Warner, T.A.,** 2013, Kernel-based extreme learning machine for remote-sensing image classification. *Remote Sensing Letters*, 4:853–862.
- Pourghasemi, H.R.,** 2016, GIS-based forest fire susceptibility mapping in Iran: a comparison between evidential belief function and binary logistic regression models. *Scandinavian Journal of Forest Research*, 31:80–98.
- Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A.V., Gulin, A.,** 2018, CatBoost: unbiased boosting with categorical features. *Advances in Neural Information Processing Systems*, 31.
- Soulis, K.X., Generali, K.A., Papadaki, C., Theodoropoulos, C. ve Psomiadis, E.,** 2021, Hydrological response of natural mediterranean watersheds to forest fires. *Hydrology*, 8(1):15.
- Tonbul., H, Colkesen., I, Kavzoglu, T.,** 2022, Pixel- and Object-Based ensemble learning for forest burn severity using USGS FIREMON and Mediterranean condition dNBRs in Aegean ecosystem (Turkey). *Advances in Space Research*, 69:3609–3632.