

Sentinel-2A MSI Verisinin Makine Öğrenmesi Tabanlı Destek Vektör Makinesi, Rastgele Orman ve En Büyük Olasılık Algoritmalarını Kullanarak Piksel Tabanlı Kontrollü Sınıflandırılmadaki Etkilerinin Araştırılması

Investigation of the Effects of Machine Learning-Based Support Vector Machine, Random Forest and Maximum Likelihood Algorithms on Pixel-Based Supervised Classification of Sentinel-2A MSI Data

Nursaç Serda Kaya^{1*} , Orhan Dengiz¹ 

¹Ondokuz Mayıs Üniversitesi, Ziraat Fakültesi, Toprak Bilimi ve Bitki Besleme Bölümü, 55139, Samsun/Türkiye.

ARAŞTIRMA MAKALESİ

*Sorumlu yazar:

Nursaç Serda Kaya
nursackaya@gmail.com

doi: 10.48123/rsgis.1410250

Yayın süreci

Geliş tarihi: 26.12.2023
Kabul tarihi: 02.06.2024
Basım tarihi: 26.09.2024

Özet

Bu araştırma makalesinde, Sinop havzasına yönelik 03.05.2023 tarihli Sentinel-2A MSI verisinin destek vektör makinesi (DVM), rastgele orman (RO) ve en büyük olasılık (EBO) algoritmalarını kullanarak piksel tabanlı kontrollü sınıflandırılması ve daha sonra her bir sınıflandırma algoritmasına ait genel doğruluk değerlerinin belirlenmesi ile her bir arazi kullanımı/arazi örtüsü sınıfı için üretici doğruluğu, kullanıcı doğruluğu, doğruluk, kesinlik, hassasiyet, F1-skoru ve ROC-AUC (İşlem Karakteristik Eğrisi-Eğri Altında Kalan Alan) metriklerine ait değerlerin kıyaslanması amaçlanmıştır. Elde edilen sonuçlar DVM ve RO algoritmalarının EBO yöntemine göre daha yüksek ve benzer genel doğruluk değerleri verdiğini göstermiştir (0.88). Her bir sınıflandırma algoritması için su kütleleri ve mera sınıflarının en yüksek doğruluk, kesinlik, hassasiyet ve F1-skoru değerlerine sahip olduğu gözlemlenmiştir. Ancak düşük AUC değerleri, eğitim setinin oluşturulduğu aşamada bazı arazi kullanımı/arazi örtüsü sınıfları için çok sayıda piksel toplanırken bazı sınıfların ise daha az piksel kullanılarak temsil edilmesi ya da yüksek doğruluk değerlerine rağmen düşük hassasiyet ve kesinlik değerlerinin varlığı gibi durumlar veri setlerindeki dengesizliği ortaya koymuştur.

Anahtar kelimeler: Uzaktan algılama, Destek vektör makinesi, Rastgele orman, En büyük olasılık, Kontrollü sınıflandırma

Abstract

In this research paper, we aimed to compare different machine learning algorithms such as Support Vector Machine (SVM), Random Forest (RF), and Maximum Likelihood for pixel-based supervised classification of Sentinel-2A MSI data from the Sinop basin on May 3, 2023. We evaluated the overall accuracy values and compared various metrics (producer accuracy, user accuracy, accuracy, precision, sensitivity, F1-score, and ROC-AUC (Receiver Operating Characteristic-Area Under Curve) for each land use/land cover class. The results showed that the SVM and RF algorithms gave higher and similar overall accuracy values than the Maximum Likelihood method (0.88). For each classification algorithm, water and pasture classes had the highest accuracy, precision, sensitivity and F1-score values. However, low AUC values, the fact that many pixels were collected for some land use/land cover classes while others were represented by fewer pixels at the stage of training set creation, or the presence of low precision and accuracy values despite high accuracy values revealed the imbalance in the datasets.

Keywords: Remote sensing, Support vector machine, Random forest, Maximum likelihood, Supervised classification

1. Giriş

Arazi kullanımı/razi örtüsünde meydana gelen değişikliklerin tespiti ile ilişkili gerçekleştirilen bilimsel çalışmalardaki genel amaçlar arasında, gözleme dayalı kentsel ve bölgesel planlamalar (Hashem & Balakrishnan, 2015), çevresel hassasiyet ve etki değerlendirmeleri (Liou vd., 2017; Nguyen vd., 2016; Nguyen & Liou, 2019; Pal & Talukdar, 2018), doğal felaketler ve zararlarının izlenmesi (Che vd., 2014; Dao & Liou, 2015; Liou vd., 2008; Liou vd., 2011; Zhang vd., 2019; Talukdar & Pal, 2019), toprak erozyonu ve toprak tuzluluğunun belirlenmesi (Talukdar vd., 2020; Chen vd., 2019; Braun & Hochschild, 2017; Nguyen vd., 2020) gibi konular yer almaktadır. Arazi kullanımı/razi örtüsü hakkında detaylı bilgi sahibi olunması, çevre koruma ve mekânsal planlamalara yönelik ticari, politik ve idari amaçların yerine getirilmesinde bir önem teşkil eder. Çevre çalışmalarında bir girdi olarak kullanılması ya da iklim değişikliğine yönelik modellerin oluşturulmasında arazi kullanımı/razi örtüsünün sınıflandırılması çok önemlidir (Disperati vd., 2015). Arazi kullanımı/razi örtüsünün birlikte kombinasyonu, jeolojik, biyofiziksel ve sosyoekonomik sistem davranışları hakkında kapsamlı bilgi sahibi olunmasında etkilidir (Moran vd., 2004).

Arazi kullanımı/razi örtüsü hakkında bilgi sahibi olunmasını sağlamak adına uzaktan algılama ve coğrafi bilgi sistemleri tekniklerinin birlikte kullanılması çok yararlıdır. Uzaktan algılama, coğrafi bilgi sistemleri analizlerinde kritik olan arazi kullanımı/razi örtüsü karakteristiklerini içeren çeşitli tematik veriler için temel bir kaynaktır. Uydu görüntüleri, arazi örtüsü dağılımlarının değerlendirilmesinde ve mevcut coğrafi özelliklerin güncellenmesinde sıklıkla kullanılan bir envanterdir. Uzaktan algılama sistemleri ve görüntü işleme yönelik yazılımların kullanılması, uzaktan algılama ve coğrafi bilgi sistemlerinin popülaritesinin artırılmasında çok önemlidir (Merchant & Narumalani, 2009). Uzaktan algılama verilerinin ve teknolojilerinin kullanımının artması, coğrafi süreçlerin daha hızlı ve güçlü ilerlemesine fayda sağlamasıyla birlikte artan hata olasılıklarını da beraberinde getirmiştir (Murty & Tiwari, 2015). Daha önceki görüntü sınıflandırma çalışmalarında doğruluğun değerlendirilmesi bir öncelik teşkil etmiyordu. Ancak, dijital sınıflandırmanın beraberinde getirdiği hata olasılıklarının artması nedeniyle doğruluk değerlerinin hesaplandığı metriklerin kullanılması hayati bir öneme sahip olmuştur (Congalton, 1991).

Uzaktan algılama teknolojileri kullanarak gerçekleştirilen çalışmaların tarım politikalarına yönelik değerli görüşlerin bildirilmesine katkı sağlamasının yanı sıra, tarım alanlarının izlenmesi ve etkin kullanımında çok önemli bir yeri vardır. Uydu verileri, tarım, doğal arazi örtüsü ve ormancılık amaçlarına yönelik tasarlanmış ve geliştirilmiştir (Üstüner vd., 2014). CBS, bilgisayar tabanlı sistemler olup, mekânsal verilerin eldesi, izlenmesi ve görüntülenmesi amacıyla tasarlanmıştır. CBS ve UA, tarım (Demir vd., 2018), ekoloji (Selim & Demir, 2018), planlama, ulaşım ve jeolojinin (Orhan vd., 2020) de dahil olduğu birçok alan ve disiplinde kullanılır. Literatürde, arazi kullanımı/razi örtüsünün belirlenmesine yönelik kullanılan sınıflandırma metotları kontrollü ve kontrolsüz sınıflandırma olmak üzere ikiye ayrılmıştır. Son 20 yıl içerisinde, kontrollü sınıflandırma yöntemlerinden olan *K*-en yakın komşu (Samaniego vd., 2008), rastgele orman (Breiman, 2001), sinir ağları (Civco, 1993), regresyon ağacı ve destek vektör makinesi (Melgani & Bruzzone, 2004) gibi ileri metotlar arazi kullanımı/razi örtüsü belirleme çalışmalarında kullanılmaya başlanmıştır (Carranza-García vd., 2019).

Görüntü işleme, genellikle arazi kullanımı/razi örtüsünü belirlemede yaygın olarak kullanılan bir süreçtir. Günümüzde gelişen bilgisayar teknolojileri ve yazılımlar ile birlikte artan işlem gücü, makine öğrenmesi metotlarını popüler hale getirmiştir. İnsansız hava araçları kullanılarak hiper spektral görüntü sınıflandırmasında ya da uydu görüntülerinin sınıflandırılmasında destek vektör makinesi (DVM), rastgele orman (RO), yapay sinir ağları ya da en büyük olasılık (EBO) gibi makine öğrenmesi yöntemlerinden yararlanılır. Yapay sinir ağlarıyla kıyaslandığında, küçük veri setlerinde bile DVM algoritmasının daha başarılı sonuçlar verdiği görülmektedir (Bromová vd., 2014; Gao vd., 2015; Su vd., 2017; Zhang vd., 2017). Bu özelliklere ek olarak DVM ile daha ucuz, iş gücü gerektirmeyen ve zamandan tasarruf edilen çalışmaların gerçekleştirilebilmesi, onu hassas tarım uygulamalarının vazgeçilmez bir parçası haline getirmiştir. Görüntü sınıflandırma çalışmalarında kullanılan ve oldukça popüler bir algoritma olan RO, esnek bir topluluk öğrenme metodudur (Guo vd., 2011; Rodriguez-Galiano vd., 2012; Tian vd., 2016). RO, aşırı uyuma mahal vermeden yüksek boyutlu ya da gürültülü veri kümelerini işlemede çok etkilidir (Dietterich, 2000; Breiman, 2001) ve DVM ve EBO gibi diğer sınıflandırıcılara göre daha başarılı doğruluk sonuçları verebilir (Nitze vd., 2012; Rodriguez-Galiano vd., 2012). Üstelik RO algoritması için karmaşık hiper parametre ayarlamalarının yapılması da gerekmez. EBO ise uzaktan algılama uygulamalarında yararlanılan en yaygın kontrollü sınıflandırma yöntemidir. Parametrik bir istatistiksel metot olan EBO yönteminde (Lillesand vd., 2004; SEOS, 2018), veriye dayalı olarak en uygun model parametreleri bulunmaya çalışılır. Bu parametreler, veri setinin gözlemlenme olasılığını maksimize eden değerlerdir. Sınıflandırma problemlerinde EBO yöntemi, hangi sınıfa ait olduğu tahmin edilmek istenen bir veri noktasının en olası sınıfını belirlemek için kullanılır. Bu yöntem, veriye en uygun modeli seçmek için kullanılan bir optimizasyon tekniğidir. Bu gibi çeşitli özellikleri bünyelerinde barındırmalarından dolayı bu araştırma makalesinde, Sentinel-2A uydu görüntüsünün piksel tabanlı olarak analiz edilmesinde kontrollü öğrenme yöntemlerinden olan EBO, RO ve DVM'den yararlanılmıştır.

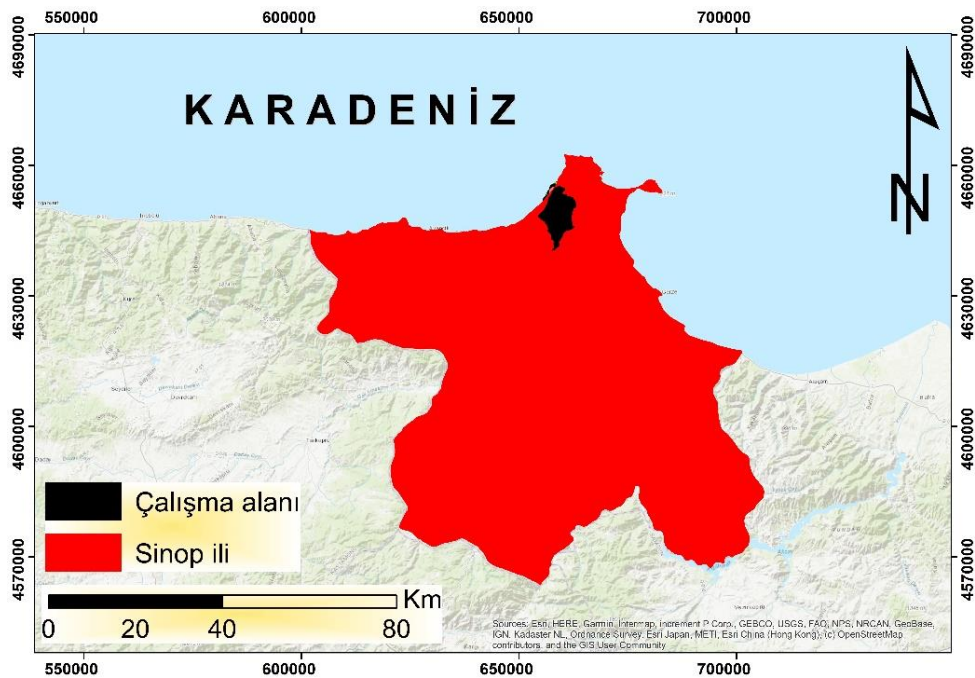
Bu çalışmada, Sinop ili Sarıkum havzası için arazi kullanımı/razi örtüsü sınıflarını oluşturmaya yönelik 03.05.2023 tarihli Sentinel-2A MSI verisinin makine öğrenmesi tabanlı destek vektör makinesi, rastgele orman ve en büyük olasılık algoritmalarını kullanarak piksel tabanlı kontrollü sınıflandırılması ve daha sonra her bir sınıflandırma algoritmasına ait

genel doğruluk değerlerinin belirlenmesi ile her bir arazi kullanımı/razi örtüsü sınıfı için doğruluk (accuracy), kesinlik (precision), hassasiyet (sensitivity), F1-skoru ve ROC-AUC metriklerine ait değerlerin kıyaslanması amaçlanmıştır. Çalışmada DVM, RO ve EBO algoritmalarının birlikte Sentinel-2A uydu görüntüsünün sınıflandırılmasında kullanılması ve elde edilen her bir arazi kullanımı/razi örtüsü sınıfının çeşitli metriklerden yararlanılarak kapsamlı bir şekilde değerlendirilmesi gibi yönlerinden dolayı bu araştırma makalesinin literatüre bir katkı sunacağı görüşündeyiz.

2. Materyal ve Metot

2.1. Araştırma Alanının Genel Özellikleri

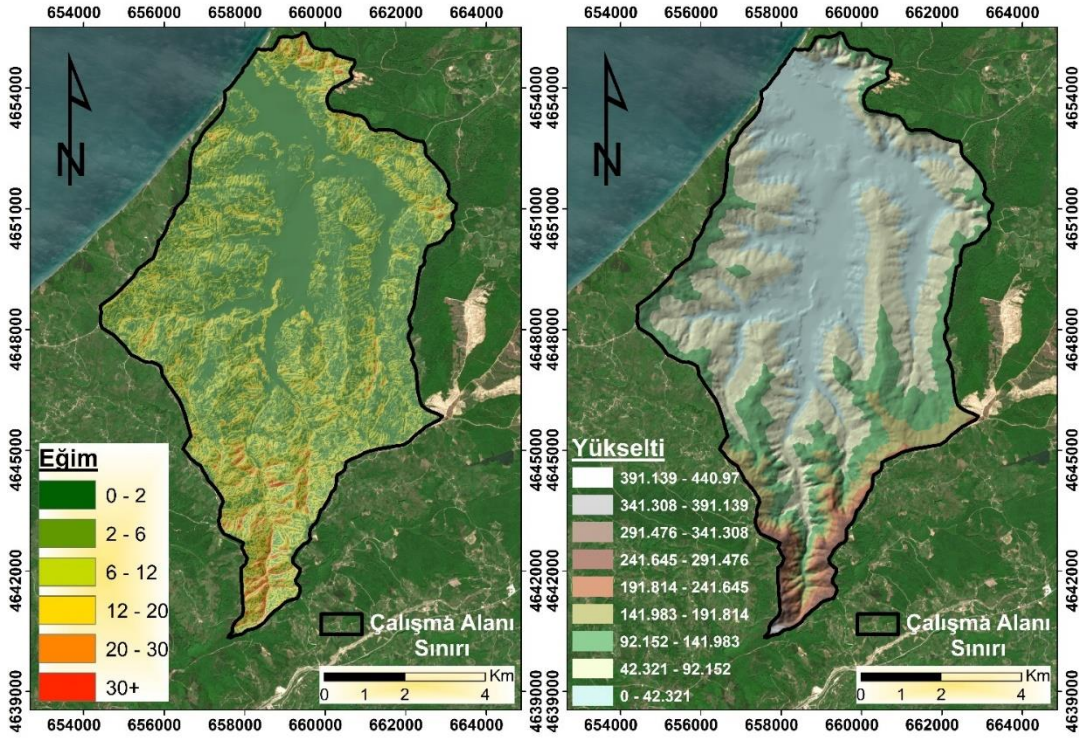
Araştırma alanı, Orta Karadeniz Bölgesi'nde yer alan Sinop ilinin kuzeyinde konumlanmış Sarıkum havzası olup, 654000-664000 D ile 4642000-4654000 K (36 Zone, UTM, WGS-84 m) koordinatları içerisinde yer almaktadır. Havza 6773.7 ha alana sahip olup, havza alanı içerisinde Sarıkum, Başaran, Gümüşsuyu, Tekke ve İncepınar köyleri yer almaktadır. Çalışma alanına ait lokasyon haritası Şekil 1'de verilmiştir.



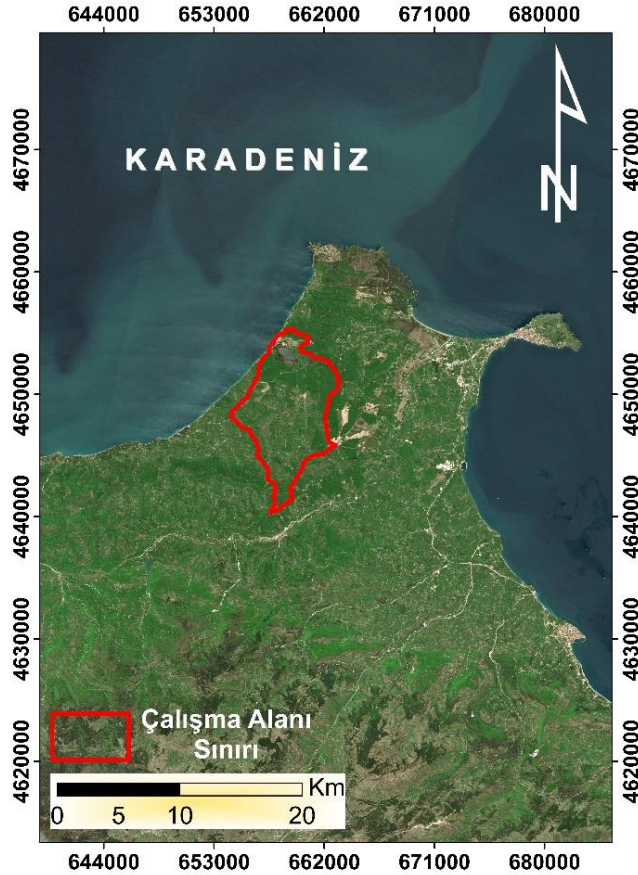
Şekil 1. Çalışma alanına ait lokasyon haritası

Yükselti haritasına göre (Şekil 2) çalışma alanı deniz seviyesinden 0 ile 440 m arasında değişen yüksekliğe sahiptir. Yükselteler kuzeyden doğuya doğru gidildikçe artmaktadır. Havzaya ait eğim haritası Şekil 2'de verilmiştir. Havzanın %25.7'lik kısmı düz ya da hafif eğimli alanlardan oluşurken %46.2'si ise dik ve sarp yamaçlar üzerinde konumlanmıştır. Çalışmada kullanılan yükselti ve eğim haritaları, arazi kullanımı/razi örtüsü sınıflarının oluşturulmasında kullanılan yardımcı CBS verilerdir. Nitekim Soyaslan ve Hepdeniz (2016), CBS ve UA teknolojilerini kullanarak Burdur ilindeki arazi kullanımının zamansal değişimini çevre açısından değerlendirmeyi amaçladıkları bir çalışmada eğim, sayısal yükselti haritası ve kabartı haritalarından gerek arazi kullanımı/razi örtüsü sınıflarının oluşturulmasında gerekse de sınıflandırma hassasiyetini artırmada yararlanmıştır.

Sınıflandırma performansını artıracığı düşünüldüğünden bu çalışmada 4 bandında 10 m, 6 bandında 20 m ve 3 bandında 60 m mekansal çözünürlüğe sahip olan 03.05.2023 tarihli Sentinel-2A MSI verisinde bulunan toplam 13 bandın tamamı kullanılarak arazi kullanımı/razi örtüsü sınıfları belirlenmeye çalışılmıştır. Nitekim Zhang vd. (2019) hassas tarımda sınıflandırma problemlerine yönelik potansiyel Sentinel-2A bantlarını inceledikleri bir çalışmada, Sentinel-2A MSI verisinde yer alan toplam 13 bandın birlikte kullanıldığında genel doğruluğu artırdığını gözlemleyerek sınıflandırma performansını iyileştirdiğini rapor etmişlerdir. Çalışmada kullanılan 03.05.2023 tarihli Sentinel-2A uydu görüntüsü Şekil 3'te verilmiştir. Sentinel-2A görüntüsü, Copernicus açık erişim merkezinden (<https://scihub.copernicus.eu/>) temin edilmiştir. Sentinel-2A uydu sensörüne ait özellikler Tablo 1'de verilmiştir. Literatürde Sentinel-2A verisinden yararlanılarak piksel tabanlı kontrollü sınıflandırmanın gerçekleştirildiği pek çok çalışma bulunmaktadır (Topaloğlu vd., 2016; ED Chaves vd., 2020; Phiri vd., 2020; Eskandari vd., 2022; Rauf vd., 2022).



Şekil 2. Çalışma alanına ait yükselti ve eğim haritaları



Şekil 3. 03.05.2023 tarihli Sentinel-2A uydu görüntüsü

Tablo 1. Sentinel-2A uydu sensörünün özellikleri

Bant	Bant Adı	Dalga Boyu (nm)	Bant Geniliği (nm)	Mekânsal Çözünürlüğü (m)
B1	Kıyı aerosol	443	20	60
B2	Mavi	490	65	10
B3	Yeşil	560	35	10
B4	Kırmızı	665	30	10
B5	Vejetasyon kırmızı-kenar	705	15	20
B6	Vejetasyon kırmızı-kenar	740	15	20
B7	Vejetasyon kırmızı-kenar	783	20	20
B8	Kızıl ötesi	842	115	10
B8A	Vejetasyon kırmızı-kenar	865	20	20
B9	Su buharı	945	20	60
B10	Sirrus	1380	30	60
B11	Kısa dalga kızıl ötesi (SWIR) 1	1610	90	20
B12	SWIR2	2190	180	20

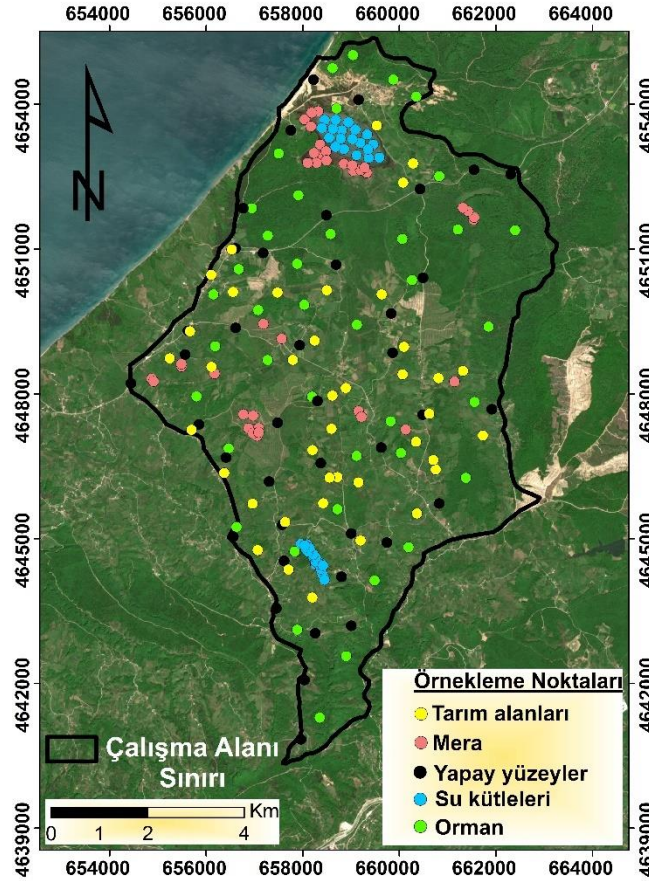
2.2. Eğitim Setinin Oluşturulması

Yüksek doğruluk değerleri elde edebilmek için geniş setlerin oluşturulmasından ziyade, eşit dağılmış örnekleme noktalarından oluşan bir eğitim setinin hazırlanması daha önemlidir (Gumma vd., 2020). Bu çalışmada her bir sınıf için 40 adet olmak üzere toplam 200 adet örnekleme noktası, Google Earth Pro kullanılarak temin edilmiştir. Bu verilerin %75'i ArcGIS 10.7v yazılımından yararlanılarak rastgele ayrılmış ve 'tarım alanları', 'su kütleleri', 'orman', 'mera' ve 'yapay yüzeyler' olmak üzere toplam 5 farklı arazi kullanımı/arazi örtüsü sınıfından oluşan bir eğitim seti oluşturulmuştur.

Arazi kullanımı/arazi örtüsü sınıfları ve sınıf değerleri Tablo 2'de verilirken, eğitim setinin oluşturulmasında kullanılan örnekleme noktaları ise Şekil 4'te verilmiştir. Eğitim seti kullanılarak makine öğrenmesi tabanlı destek vektör makinesi, rastgele orman ve en büyük olasılık algoritmalarına yönelik eğitilmiş sınıflandırmanın yapılabilmesinde ArcGIS 10.7v yazılımından yararlanılmıştır. Eğitim seti oluşturulurken her bir arazi kullanımı/arazi örtüsü sınıfı için belirli bir sayıda piksel seçilir. Bu seçilen piksellerin sayısı, her bir sınıf için temsil edilen kaç tane örnek pikselin olduğunu belirtir. Eğitim seti için kullanılan her bir arazi kullanımı/arazi örtüsü sınıfına ait piksel sayıları Tablo 3'te verilmiştir.

Tablo 2. Arazi kullanımı/arazi örtüsü sınıfları

Sınıf Değerleri	Arazi Kullanımı/Arazi Örtüsü Sınıfları
0	Su kütleleri
1	Orman
2	Yapay yüzeyler
3	Tarım alanları
4	Mera



Şekil 4. Eğitim setlerinin oluşturulmasında kullanılan örnekleme noktaları

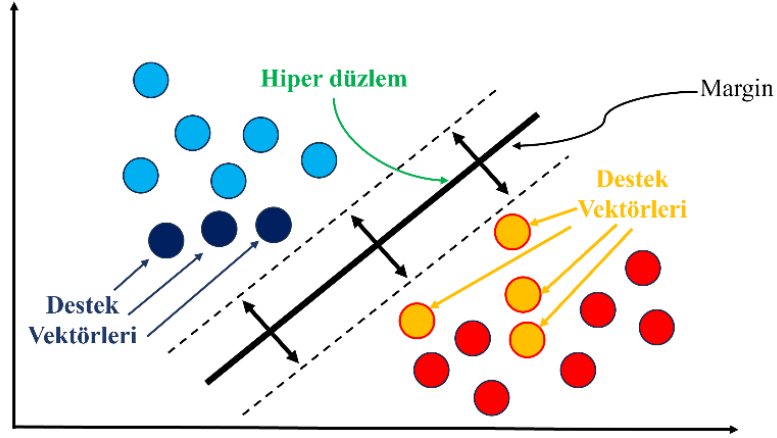
Tablo 3. Arazi kullanımı/arazi örtüsü sınıflarına ait piksel sayıları

Arazi kullanımı/arazi örtüsü sınıfı	Eğitim seti
Su kütleleri	934
Orman	4531
Yapay yüzeyler	649
Tarım alanları	1632
Mera	670
Toplam	8416

2.3. Sınıflandırma Algoritmaları

2.3.1. Destek Vektör Makinesi (DVM)

Destek vektör makinesi, istatistiksel öğrenme teorisine ve yapısal risk minimizasyonuna dayalı bir regresyon ve kontrollü öğrenme algoritmasıdır (Tehrany vd., 2015). DVM, genelleştirme hatasının üst sınırının minimize edilmesi prensibiyle çalışan bir istatistiksel öğrenme teorisidir. Kullanıcı tahmininin üst sınırını minimize etmeyi amaçlayarak eğitim ve kapasite arasında bir denge kurmaya çalışır. Destek vektör makinesindeki çekirdek yaklaşımı, iki sınıf arasındaki en iyi ayrımın sağlanmasına yönelik bir hiper düzlem tanımlar (Şekil 5) (Eşitlik 1). Bu sınıflandırma algoritması, iki sınıf kümesi arasındaki çok boyutlu boşluğun maksimize edilmesini amaçlar. Bu hiper düzlem, eğitim seti olarak bilinen bir alt kümenin kullanılmasıyla geliştirilmiştir ve genelleme yeteneği test verisi adı verilen bir bağımsız alt kümenin kullanılarak doğrulanır. Literatürde DVM algoritmasının görüntü sınıflandırma problemlerinde kullanıldığı pek çok çalışma bulunmaktadır (Kavzoğlu & Çölkesen, 2009, 2010; Okwuashi & Ndehedehe, 2020; Sheykhmousa vd., 2020; Bawa vd., 2023).



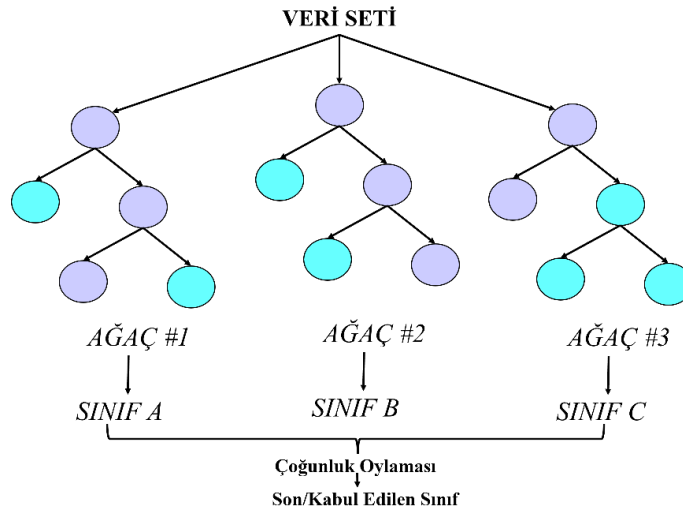
Şekil 5. Verinin destek vektör makinesi kullanılarak sınıflandırılması (Tao vd., 2006)

$$f(x) = \text{sgn}(\sum_{i=1}^n \alpha_i y_i (x_i \cdot x) + b) \quad (1)$$

Eşitlikte, $f(x)$ bir karar fonksiyonu olup, veri noktası x 'in hangi sınıfa ait olduğunu, α_i bir Lagrange çarpanı olup, her bir veri seti için hesaplanan katsayıyı, y_i veri noktasının sınıf etiketlerini, x_i eğitim veri setindeki veri noktalarını ve x sınıflandırılacak olan yeni veri noktasını belirlerken; b bias terimi olup, modelin eğitimi sırasında belirlenen bir sabiti ifade etmektedir.

2.3.1. Rastgele Orman (RO)

Kontrollü öğrenme ve regresyon görevleri açısından rastgele orman, oldukça tanınmış bir makine öğrenmesi algoritmasıdır ve popülaritesi her geçen gün daha da artmaktadır (Cunningham vd., 2008). Metot, sınıflandırma modellerinde performans artıran ve daha karmaşık problemlerin çözümünde önemli bir rol oynayan topluluk öğrenmesine dayanır. RO, özellikle spesifik bir veri setinin farklı bölümlerinde eğitilmiş karar ağaçlarından elde edilen sonuçların ortalamasını alan bir sınıflandırma yöntemidir (Rodriguez-Galiano vd., 2012). RO, her bir karar ağacına ait tahminleri toplar ve son tahmin olarak, tek bir ağaca bağlı olmak yerine, en çok oyu alan karar ağacını baz alır (Şekil 6) (Eşitlik 2).



Şekil 6. Verinin rastgele orman algoritması kullanılarak sınıflandırılması (Öztürk, 2022)

RO, üstün tahmin doğruluğu sağlar ve ezberlemeyi en aza indirir (Pouyan vd., 2021). RO algoritmalarının çok yönlülüğü, kategorizasyon problemlerinde mükemmel performans sağlar. Arazi örtüsü/arazi kullanımının belirlenmesinde rastgele orman metodunun kullanıldığı pek çok çalışma bulunmaktadır (Pal, 2005; Gislason vd., 2006; Akar & Güngör, 2012; Ghansah vd., 2021; Nakach vd., 2022; Osenberg vd., 2023).

$$\hat{y}(x) = \frac{1}{n} \sum_{i=1}^n \hat{y}_i(x) \quad (2)$$

Yukarıdaki eşitlikte, $\hat{y}(x)$ RO modelin x girdisi için yaptığı nihai tahmini, n RO içindeki toplam karar ağacı sayısını, $\hat{y}_i(x)$ i -inci bağımsız karar ağacının x girdisi için yaptığı tahmini ve i ise her bir karar ağacını temsil etmektedir.

Bu çalışmada kontrollü sınıflandırma işlemi gerçekleştirilirken RO ve DVM algoritmaları için kullanılan hiper parametreler ArcGIS 10.7v yazılımı bünyesinde bulunan ve varsayılan (default) hiper parametreler olup, açıklamalarıyla birlikte Tablo 4'te verilmiştir. Python programlama dili gibi daha güncel yazılımların farklı optimizasyon tekniklerini sunmaları, makine öğrenmesi algoritmalarına ait hiper parametre optimizasyonlarının yapılmasını daha pratik kılacaktır. Örneğin Muslim (2020), destek vektör makinesine ait hiper parametrelerin optimize edilmesinde örgü arama (grid search) yöntemini kullanıldığı bir çalışmada, Python 3.8 programlama dilinden yararlanmıştır. Benzer şekilde Tiwari vd. (2022) makine öğrenmesi algoritmalarının optimize edilmesine yönelik bir çalışmalarında, Python 3.7 yazılımını tercih etmiştir.

Tablo 4. Kontrollü sınıflandırmada kullanılan RO ve DVM algoritmaları için kullanılan varsayılan (default) hiper parametreler

Algoritma	Hiper parametre	Açıklama
RO	MaxNumTrees=50	Maksimum ağaç sayısı belirler.
	MaxTreeDepth=30	Her bir ağacın maksimum derinliğini belirler.
	MaxSampleClass=1000	Bir sınıfın maksimum örnek sayısını belirler.
DVM	Maximum Number of Samples Per Class=500	Her sınıf için maksimum örnek sayısı belirler.

2.3.1. En Büyük Olasılık (EBO)

Sınıflandırma problemlerinde sıklıkla kullanılan bir yöntem olan en büyük olasılık metodu, istatistiksel öğrenmeye dayalı bir algoritmadır. Ortalama değerleri baz almasının yanında, varyans ve kovaryans değerleri de bu sınıflandırma algoritmasında dikkate alınır. Sınıf kümesine ait verilerin normal bir dağılıma sahip olduğu varsayılan bu sınıflandırma algoritmasında, olasılık yoğunluk fonksiyonları Eşitlik (3) kullanılarak hesaplanmaktadır. Sınıflandırılacak pikseller, üyelik olasılığı en yüksek olan sınıflara atanır. Eğer pikseller için hesaplanan sınıflara ait olma olasılık değerleri, önceden belirlenen eşik değerinden düşükse, bu pikseller ilgili sınıflar açısından belirsiz olarak sınıflandırılmaktadır (Myung, 2003; Çölkesen, 2009; Günlü, 2012; Topaloğlu vd., 2016).

$$P(x) = \ln(a_c) - [0.5x \ln(|S_x|)] - [0.5x(x - M_c)^T x (S_x^{-1}) x (x - M_c)] \quad (3)$$

Eşitlikte, P olasılık değeri, C örnek bir sınıfı, x aday pikselin ölçüm vektörü, M sınıfın ortalaması, a aday pikselin sınıfa ait olma yüzdesi ve S_x varyans-kovaryans matrisidir.

2.4. Yer Doğrulama Verileri ve Doğruluk Analizi

Bu çalışmada örnekleme noktalarının %25'i ArcGIS 10.7v yazılımından yararlanılarak rastgele ayrılmış ve yer doğrulama verilerinin oluşturulmasında kullanılmıştır. Eğitim verileri kullanılarak sınıflandırılmış görüntülerin doğruluk analizlerinin gerçekleştirilmesinde yer doğrulama verilerinden yararlanılmıştır. Doğruluk analizi ya da doğrulama, sınıflandırılmış uzaktan algılama verilerinin değerlendirilmesinde çok önemli bir adımdır. Sınıflandırılmış uydu görüntüsü hakkında değerli bilgilerin sağlanması adına hata metriklerinin hesaplanması bir gerekliliktir. Uzaktan algılama verisinin verimli bir şekilde kullanılması ancak o verinin kalitesi hakkında bilgi sahibi olunmasıyla mümkün olur. Sınıflandırılmış görüntüden elde edilen genel doğruluk değeri, bir sınıflandırma modelinin ne kadar iyi performans gösterdiğini değerlendiren temel bir ölçüttür.

Genel doğruluk, her bir arazi kullanımı/razi örtüsü sınıfı için doğru olarak sınıflandırılmış piksellerin, bu piksellere karşılık gelen yer doğrulama verileri ile karşılaştırılmasıyla elde edilir. Genel doğruluğun hesaplanmasına yönelik formül Eşitlik (4)'te verilmiştir.

$$\text{Genel Doğruluk} = \frac{\sum_{k=1}^q n_{kk}}{n} \times 100 \quad (4)$$

Üretici doğruluğu, her sınıf için doğru sınıflandırılmış piksellerin sayısının bu sınıf için kullanılan referans piksellerinin sayısına bölünmesiyle bulunur. Üretici doğruluğu, referans piksellerinin ne kadar doğru sınıflandırılabilirdiğini gösterir. Üretici doğruluğunun hesaplanmasına yönelik formül Eşitlik (5)'te verilmiştir.

$$\text{Üretici Doğruluğu} = \frac{n_{ii}}{n_{ki}} \quad (5)$$

Burada n_{ii} , doğru olarak sınıflandırılmış piksel ve n_{ki} , o sınıf için seçilen gerçek örtü tipi örnekleme sayısı toplamı anlamına gelmektedir.

Kullanıcı doğruluğu, her sınıf için doğru sınıflandırılmış piksel sayısının, o sınıfa atanan piksellerin toplamına bölünmesiyle bulunur (Yan, 2006). Kullanıcı doğruluğu, her sınıf için atanan piksellerin o sınıfı gerçekte temsil etme olasılığını gösterir (Congalton, 1991; Campbell and Wynne, 2002; Jensen, 2005). Kullanıcı doğruluğunun hesaplanmasında kullanılan formül Eşitlik (6)'da verilmiştir.

$$\text{Kullanıcı Doğruluğu} = \frac{n_{ii}}{n_{ik}} \quad (6)$$

Burada n_{ii} , sınıflandırılmış piksel ve n_{ik} , sınıflandırılan toplam piksel anlamına gelmektedir. Sınıflandırmanın başarısının ölçülmesinde genel doğruluğun hesaplanması her zaman tek başına yeterli olmayabilir. Özellikle dengesiz dağılan sınıf koşullarında sınıflandırma performansını değerlendirmek için, doğruluk (accuracy), kesinlik (precision), hassasiyet (sensitivity/recall) ve F1-skoru gibi metrikler kullanılır. Bu hata metrikleri, sınıflar arasındaki dengesizlikleri hesaba katarak daha kapsamlı bir doğruluk analizi sağlar. Doğruluk, kesinlik, hassasiyet ve F1-skorunun nasıl hesaplandığına dair formüller, sırasıyla Eşitlik (7), Eşitlik (8), Eşitlik (9) ve Eşitlik (10)'da verilmiştir.

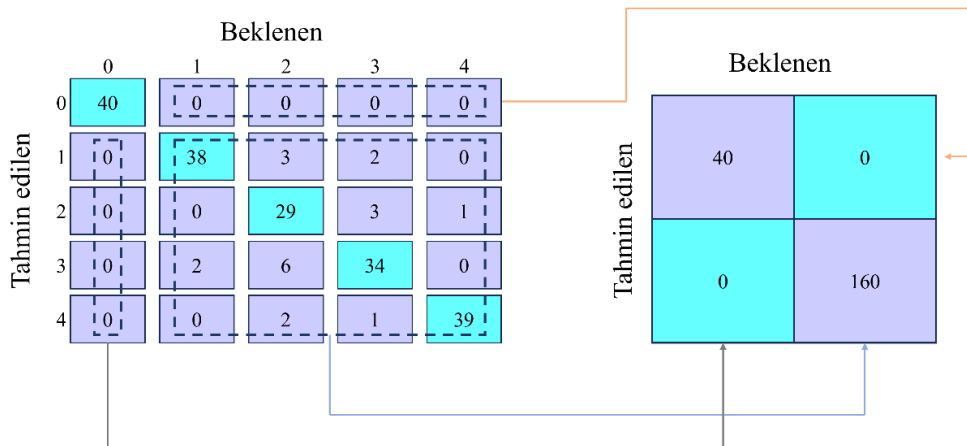
$$\text{Doğruluk} = \frac{DP+DN}{DP+YP+DN+YN} \quad (7)$$

$$\text{Kesinlik} = \frac{(DP)}{(DP+FP)} \quad (8)$$

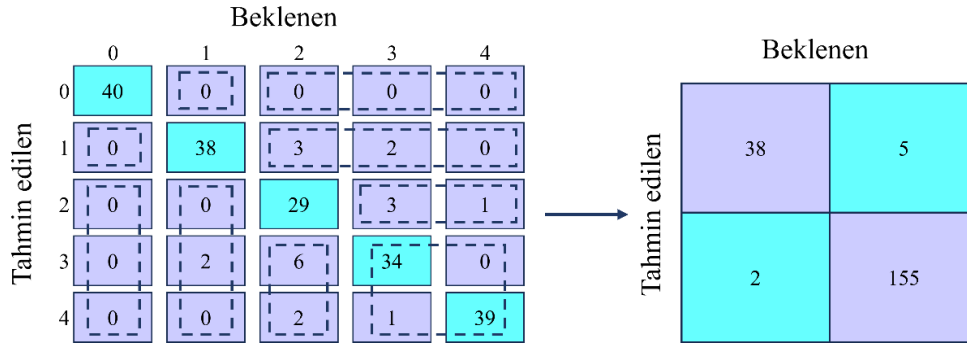
$$\text{Hassasiyet} = \frac{(DP)}{(DP+YN)} \quad (9)$$

$$\text{F1 - skoru} = 2 * \frac{(\text{Kesinlik} \times \text{Hassasiyet})}{(\text{Kesinlik} + \text{Hassasiyet})} \quad (10)$$

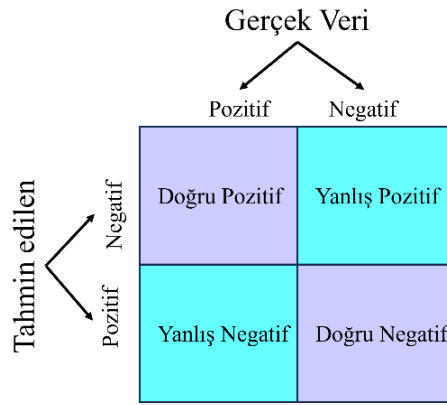
Şekil 7 ve Şekil 8'de her bir sınıflandırma algoritmasına yönelik doğruluk (accuracy), kesinlik (precision), hassasiyet (sensitivity/recall) ve F1-skorunun hesaplanması amacıyla 5x5'lik karışıklık matrisinden bire-karşı-hepsi yöntemi kullanılarak Şekil 9'daki 2x2'lik karışıklık matrisi elde edilmiştir. Her bir sınıflandırma algoritmasına yönelik oluşturulan arazi kullanımı/arazi örtüsü sınıfları için bire-karşı-hepsi yöntemi kullanılarak elde edilmiş 2x2'lik karışıklık matrisleri ise Şekil 10'da verilmiştir.



Şekil 7. Destek vektör makinesi, rastgele orman ve en büyük olasılık algoritmaları için çok sınıflı bir karışıklık matrisinin bire-karşı-hepsi (sınıf 0 için) matrisine dönüştürülmesi (örnek gösterim) (Kundu, 2022)

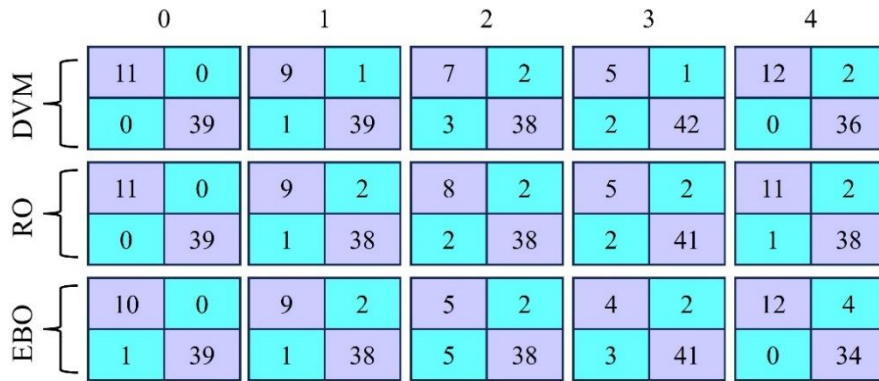


Şekil 8. Destek vektör makinesi, rastgele orman ve en büyük olasılık algoritmaları için çok sınıflı bir karışıklık matrisinin bire-karşı-hepsi (sınıf 1 için) matrisine dönüştürülmesi (örnek gösterim) (Kundu, 2022)



Şekil 9. 2x2'lik karışıklık matrisi (Kundu, 2022)

Doğru pozitif: Model pozitif olarak sınıflandırır ve gerçek etiket de pozitiftir. Yanlış pozitif: Model pozitif olarak sınıflandırır, ancak gerçek etiket negatiftir. Doğru negatif: Model negatif olarak sınıflandırır ve gerçek etiket de negatiftir. Yanlış negatif: Model negatif olarak sınıflandırır ancak etiket aslında pozitiftir.



Şekil 10. Her bir sınıflandırma algoritmasına yönelik oluşturulan arazi kullanımı/arazi örtüsü sınıfları için bire-karşı-hepsi yöntemi kullanılarak elde edilmiş nihai 2x2'lik karışıklık matrisleri

Burada; DVM: Destek vektör makinesi, RO: Rastgele orman; EBO: En büyük olasılık, 0: Su kütleleri, 1: Orman, 2: Yapay yüzeyler, 3: Tarım alanları ve 4: Mera sınıfını ifade etmektedir. Şekil 10'da verilen 2x2'lik karışıklık matrisleri, doğruluk analizlerinin gerçekleştirilmesinde kullanılan yer doğrulama verilerinden elde edilmiştir.

İşlem Karakteristik Eğrisi (Receiver Operating Characteristic) (ROC), hassasiyet ve özgülük ile birlikte kullanılarak modelin başarı oranını hesaplamak için kullanılan bir başka ölçüt olup, eğri altındaki alan (Area Under Curve) (AUC) modelin gelecekteki olayları ne kadar iyi tahmin ettiğini gösterir. 0 ile 1 arasında değişen AUC değeri, modele dayalı performans değerlendirmesini Hanley ve McNail (1982)'e göre zayıf (<0.6), orta (<0.7), iyi (0.7-0.8), çok iyi (0.8-0.9) ve mükemmel (> 0.9) olmak üzere kategorilere ayırır.

Daha yüksek bir AUC değeri, sınıflar arasında daha güçlü bir ayırım yapma yeteneğini ifade eder (Miao & Zhu, 2022). Bu çalışmada, her bir sınıflandırma algoritmasına yönelik ROC eğrileri, Python 3.8 programlama dili, geliştirme ortamı olarak Anaconda3 ve kod düzenleyici olarak Spyder 5.2.2 kullanılarak elde edilmiştir.

3. Bulgular ve Tartışma

Çalışmada 2023 tarihli Sentinel-2A uydu görüntüsü çeşitli makine öğrenmesi tabanlı algoritmalar kullanılarak sınıflandırılmış ve her bir sınıflandırma algoritması için genel doğruluk değerleri belirlenmiştir. Buna ek olarak sınıflandırma sonuçlarının daha kapsamlı yorumlanması adına, arazi kullanımı/arazi örtüsü sınıflarına ait doğruluk, kesinlik, hassasiyet ve F1-skoru değerleri tüm sınıflandırma yöntemleri için ayrı ayrı hesaplanmıştır. Bu doğrultuda, destek vektör makinesi ve rastgele orman algoritmalarının, en büyük olasılık yöntemine göre daha yüksek ve benzer genel doğruluk değerleri verdiği gözlemlenmiştir (0.88) (Tablo 5).

Tablo 5. Yer doğrulama verilerine dayalı karışıklık matrisleri ve sınıflandırma metodlarının doğruluğu

Destek Vektör Makinesi							
Sınıflar	Su kütleleri	Orman	Yapay yüzeyler	Tarım alanları	Mera	Genel toplam	KD, %
Su kütleleri	11	0	0	0	0	11	100.00
Orman	0	9	1	0	0	10	90.00
Yapay yüzeyler	0	0	7	2	0	9	77.78
Tarım alanları	0	1	0	5	0	6	83.33
Mera	0	0	2	0	12	14	85.71
Genel toplam	11	10	10	7	12	50	
ÜD, %	100.00	90.00	70.00	71.43	100.00		
Genel doğruluk (%)	88						
Rastgele Orman							
Sınıflar	Su kütleleri	Orman	Yapay yüzeyler	Tarım alanları	Mera	Genel toplam	KD, %
Su kütleleri	11	0	0	0	0	11	100.00
Orman	0	9	1	0	1	11	81.82
Yapay yüzeyler	0	0	8	2	0	10	80.00
Tarım alanları	0	1	1	5	0	7	71.43
Mera	0	0	0	0	11	11	100.00
Genel toplam	11	10	10	7	12	50	
ÜD, %	100.00	90.00	80.00	71.43	91.67		
Genel doğruluk (%)	88						
En Büyük Olasılık							
Sınıflar	Su kütleleri	Orman	Yapay yüzeyler	Tarım alanları	Mera	Genel toplam	KD, %
Su kütleleri	10	0	0	0	0	10	100.00
Orman	0	9	1	1	0	11	81.82
Yapay yüzeyler	1	0	5	1	0	7	71.43
Tarım alanları	0	1	1	4	0	6	66.67
Mera	0	0	3	1	12	16	75.00
Genel toplam	11	10	10	7	12	50	
ÜD, %	90.91	90.00	50.00	57.14	100.00		
Genel doğruluk (%)	80						

Burada, KD: Kullanıcı doğruluğunu ve ÜD: Üretici doğruluğunu ifade etmektedir.

DVM ve RO algoritmaları, günümüzde yaygın olarak kullanılan bilgisayar destekli makine öğrenmesi yöntemlerindedir (Rana & Suryanarayana, 2020). Adam vd. (2014), arazi kullanımı/arazi örtüsü sınıflarını destek vektör makinesi ve rastgele orman algoritmalarını kullanarak belirledikleri bir çalışmada, her iki sınıf için de benzer genel doğruluk değerleri elde ettiklerini rapor etmiştir. Rimal vd. (2020), kentsel alanları destek vektör makinesi ve en büyük olasılık yöntemlerinden faydalanarak haritaladıkları bir araştırmasında, destek vektör makinesinin en büyük olasılık yöntemine göre daha iyi bir sınıflandırma performansı gösterdiğini gözlemiştir. Benzer şekilde Halder vd. (2023), en büyük olasılık yönteminin popüler bir yöntem olduğunu, ancak günümüzde destek vektör makinesinin arazi

kullanımı/arazi örtüsü sınıflarının belirlendiği çalışmalarda daha başarılı doğruluk değerleri verdiğini belirtmiştir. Sentinel-1 ve Sentinel-2 uydularından yararlanılarak rastgele orman ve en büyük olasılık algoritmalarının sel hasarı tespitinde kullanıldığı bir çalışmada Billah vd. (2023), rastgele orman algoritmasının (%90) en büyük olasılık yöntemine göre (%74) daha başarılı bir sonuç verdiğini rapor etmiştir. Sınıflandırma problemlerinde hataları en aza indirme yetenekleri, DVM ve RO algoritmalarını en büyük olasılık gibi parametrik yöntemlerden daha üstün kılmaktadır (Huang vd., 2002; Foody & Mathur, 2004; Pal & Mather, 2005; Schneider vd., 2012; Rimal vd., 2020). Bir çalışmada Mather & Tso, (2016), RO ve DVM modellerinin benzer ve yüksek sonuçlar vermesinin, onları en popüler makine öğrenmesi sınıflandırıcıları arasına soktuğunu belirtmiştir.

Tablo 5'te yer alan destek vektör makinesine ait kullanıcı doğruluğu değerleri incelendiğinde, su kütleleri, orman, yapay yüzeyler, tarım alanları ve meranın sırasıyla %100, %90.00, %77.78, %83.33 ve %85.71'lik oranlarla sınıflandırıldığını görmekteyiz. Diğer yandan, rastgele orman algoritmasına ait kullanıcı değerleri sırasıyla su kütleleri, orman, yapay yüzeyler, tarım alanları ve mera sınıfları için %100.00, %81.82, %80.00, %71.43 ve %100.00 olarak sınıflandırılmıştır. En büyük olasılık algoritması için su kütleleri, orman, yapay yüzeyler, tarım alanları ve mera arazi kullanım/arazi örtüsüne ait sınıflar ise sırasıyla %100.00, %81.82, %71.43, %66.67 ve %75.00'lik oranlarla sınıflandırılmıştır.

Üretici doğruluğu değerlerine baktığımızda, su kütleleri, orman, yapay yüzeyler, tarım alanları ve mera sınıfları destek vektör makinesi için sırasıyla %100.00, %90.00, %70.00, %71.43 ve %100.00 doğrulukla sınıflandırılmıştır. Rastgele orman algoritmasına ait üretici doğruluğu değerleri, su kütleleri, orman, yapay yüzeyler, tarım alanları ve mera için sırasıyla %100.00, %90.00, %80.00, %71.43 ve %91.67 şeklinde sınıflandırılmıştır. Son olarak, en büyük olasılık algoritmasıyla sınıflandırılan arazi kullanım/arazi örtüsü sınıflarından su kütleleri, orman, yapay yüzeyler, tarım alanları ve mera ise sırasıyla %90.91, %90.00, %50.00, %57.14 ve %100.00 doğrulukla sınıflandırılmıştır.

Tablo 6'ya göre, DVM, RO ve EBO algoritmaları için orman sınıfı, sırasıyla %50.56, %51.08 ve %51.62'lik oranlarla çalışma alanında en fazla dağılım gösteren arazi kullanım/arazi örtüsü sınıfı olmuştur. Su kütleleri ise, %1.84, %1.85 ve %1.64'lük oranlarla sırasıyla DVM, RO ve EBO algoritmaları için çalışma alanında en az dağılım gösteren arazi kullanım/arazi örtüsü sınıfı olmuştur.

Tablo 6. Arazi kullanım/arazi örtüsü sınıflarının alansal dağılımları

Sınıflar	DVM		RO		EBO	
	Alan (ha)	Yüzde (%)	Alan (ha)	Yüzde (%)	Alan (ha)	Yüzde (%)
Su Kütleleri	124.55	1.84	125.08	1.85	110.77	1.64
Orman	3417.17	50.56	3452.5	51.08	3488.81	51.62
Yapay Yüzeyler	349.93	5.18	359.95	5.33	316.01	4.68
Tarım Alanları	2646.91	39.16	2615.26	38.70	2598.21	38.44
Mera	219.84	3.25	205.61	3.04	244.6	3.62
Toplam	6758.4	100	6758.4	100	6758.4	100

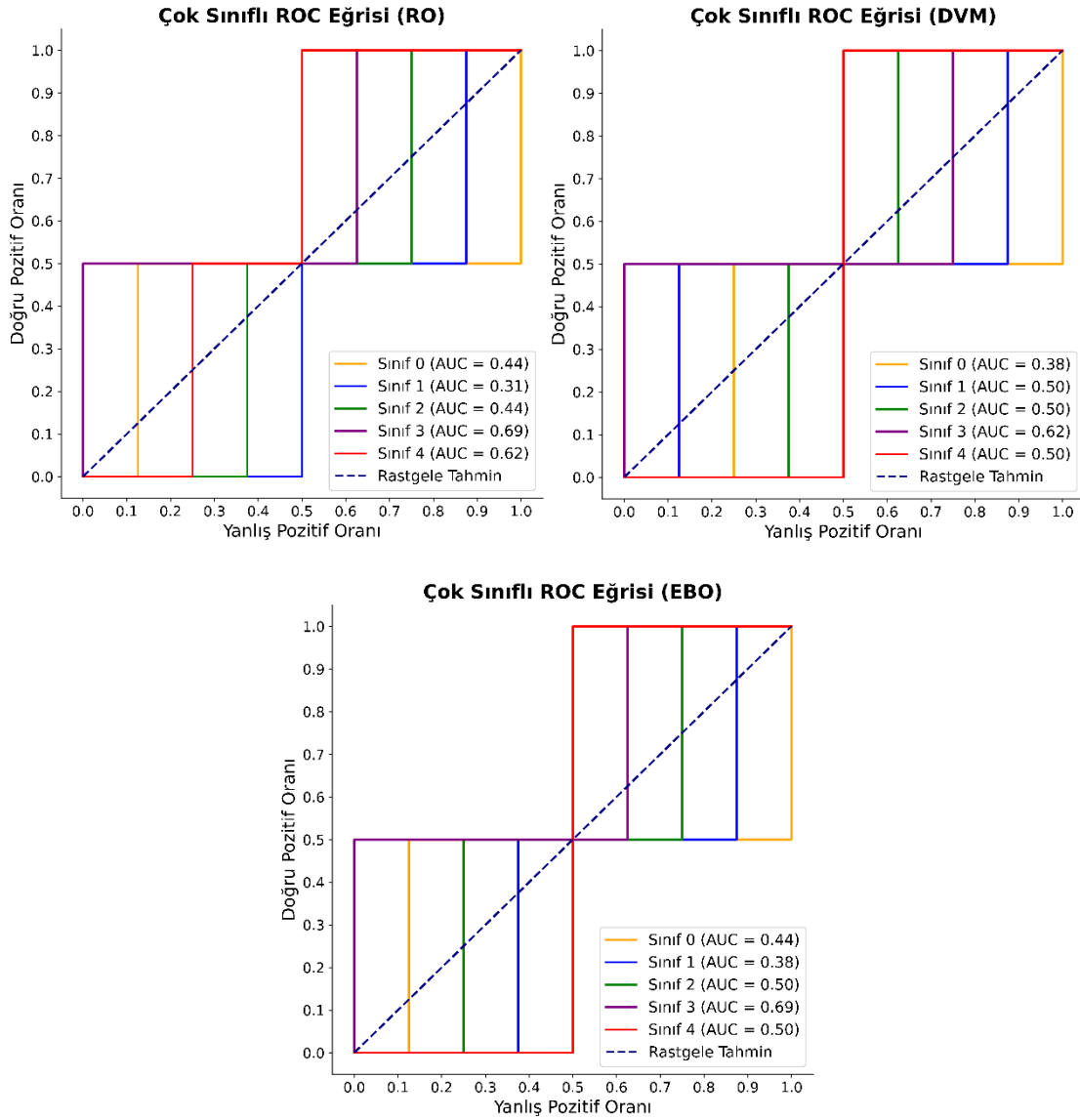
Tablo 7 incelendiğinde, su kütleleri sınıfına ait F1-skoru değerleri, DVM, RO ve EBO algoritmaları için en yüksek sonuçlara ulaşmıştır. En düşük değer, 0.50 ile yapay yüzeyler sınıfına ait hassasiyet metriğinde tespit edilmiştir.

F1-skoru, kesinlik ve hassasiyet metrislerinin harmonik bir ortalamasıdır ve 0 ve 1 arasında değişkenlik göstermektedir. F1-skoru için 0.7 değeri, modelin iyi bir performans gösterdiğini işaret eder (Goutte & Gaussier, 2005). F1-skorunun 1'e yaklaşması, bu hata matrisinin daha başarılı sonuçlar verdiği anlamına gelir. Bununla birlikte iyi bir F1-skoru değeri, sınıflandırıcının daha az oranda yanlış tahmin edilen pozitif örneklere ve yanlış tahmin edilen negatif örneklere sahip olduğu anlamına gelir (Alakus & Turkoglu, 2020).

Tablo 7. Arazi kullanım/arazi örtüsü sınıfları için DVM, RO ve EBO sınıflandırma algoritmalarının karşılaştırılması

Sınıflar	Doğruluk			Kesinlik			Hassasiyet			F1-skoru		
	DVM	RO	EBO	DVM	RO	EBO	DVM	RO	EBO	DVM	RO	EBO
Su Kütleleri	1.00	1.00	0.98	1.00	1.00	1.00	1.00	1.00	0.91	1.00	1.00	0.95
Orman	0.96	0.94	0.94	0.90	0.82	0.82	0.90	0.90	0.90	0.90	0.86	0.86
Yapay Yüzeyler	0.90	0.92	0.86	0.78	0.80	0.71	0.70	0.80	0.50	0.74	0.80	0.59
Tarım Alanları	0.94	0.92	0.90	0.83	0.71	0.67	0.71	0.71	0.57	0.77	0.71	0.62
Mera	0.96	0.98	0.92	0.86	0.85	0.75	1.00	0.92	1.00	0.92	0.88	0.86

Bu çalışmada, DVM, RO ve EBO algoritmaları kullanılarak sınıflandırılmış arazi kullanımı/arazi örtüsü sınıflarına yönelik ait ROC eğrileri oluşturulmuş ve ROC eğrisi altında kalan en yüksek AUC değerinin 0.69 ile tarım alanları sınıfı için RO ve EBO algoritmalarında elde edildiği gözlemlenmiştir. En düşük AUC değerleri ise 0.31 ile orman sınıfı için RO algoritmasında tespit edilmiştir. En büyük olasılık ve rastgele orman algoritması kullanılarak sınıflandırılmış uydu görüntüsünde tarım alanları sınıfına ait AUC değerleri hariç, diğer sınıflandırma algoritmaları ile sınıflandırılmış arazi kullanımı/arazi örtüsü sınıflarına yönelik AUC değerlerinin 0.6'dan düşük olması nedeniyle modele dayalı performansları 'zayıf' kategorisinde değerlendirilmiştir (Hanley & McNail, 1982). Bu AUC değerlerine göre, sınıflandırma algoritmalarına ait arazi kullanımı/arazi örtüsü sınıfları arasında çok güçlü bir ayırım yapılamadığı sonucuna varılmıştır (Miao & Zhu, 2022). DVM, RO ve EBO algoritmalarına yönelik oluşturulan ROC eğrileri Şekil 11'de verilmiştir.



Şekil 11. RO, DVM ve EBO algoritmaları için ROC eğrileri

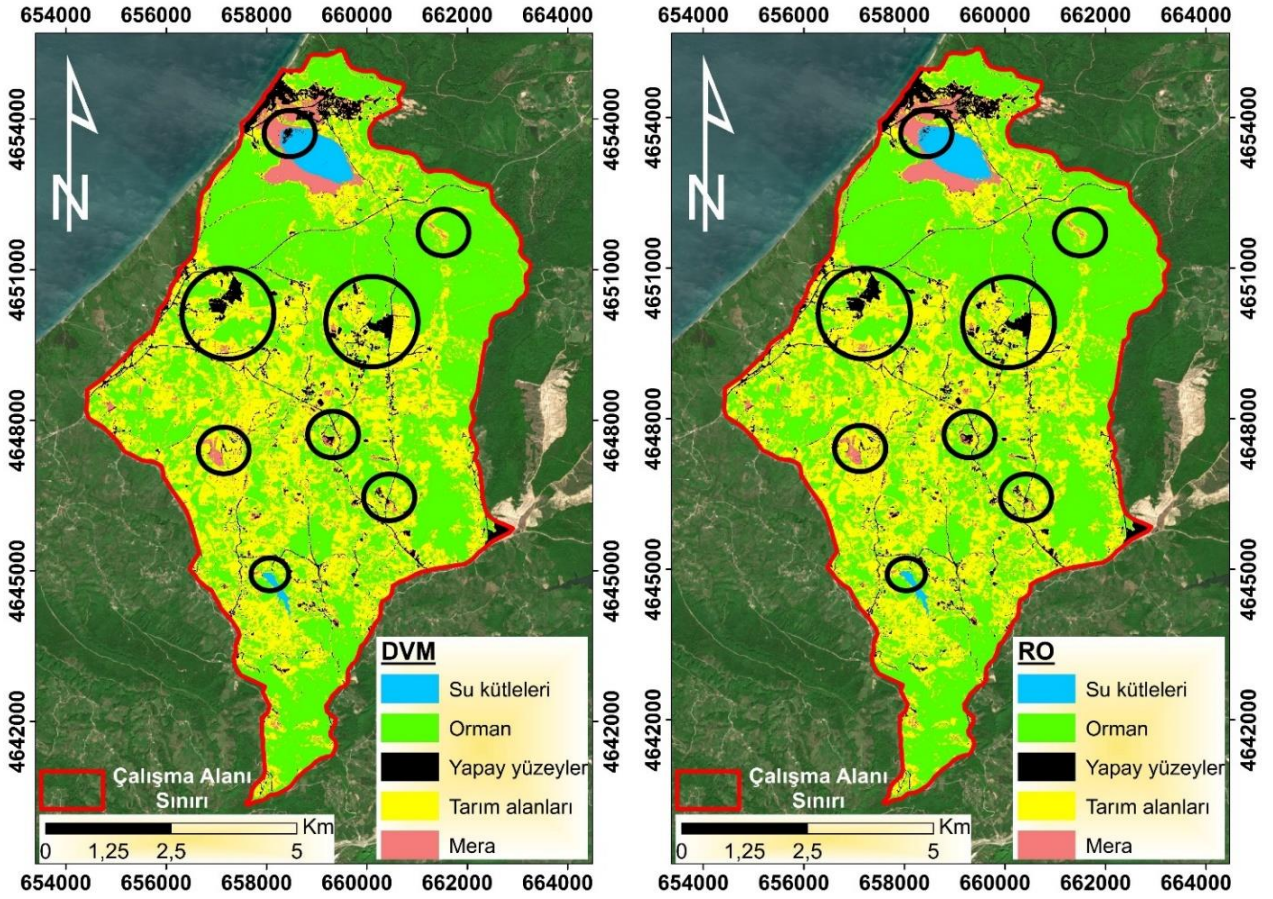
Şekil 11'e baktığımızda AUC değerlerinin, her bir sınıflandırma sistemine yönelik nerdeyse bütün arazi kullanımı/arazi örtüsü sınıfları için 'zayıf' olarak kategorize edildiğini görmekteyiz (<0.6) (Hanley & McNail, 1982). Şekildeki grafiklerde yer alan çapraz şerit çizgi 'rastgele bir tahmin' anlamına gelir ve model tarafından anlamsız kabul edilir. Eğri sol üst köşeye yaklaştıkça modelin performansı önemli ölçüde artar. Çapraz çizginin altındaki herhangi bir eğri, performansın rastgele bir tahminden daha kötü olduğunu gösterir.

Eğitim setlerinin oluşturulmasında kullanılan örnekleme noktalarının her ne kadar çalışma alanında eşit dağılmış olmasına dikkat edilse de elde edilen AUC değerlerinin nerdeyse bütün arazi kullanımı/arazi örtüsü sınıfları için düşük olması, veri setlerinde bir dengesizlik olduğunu göstermektedir.

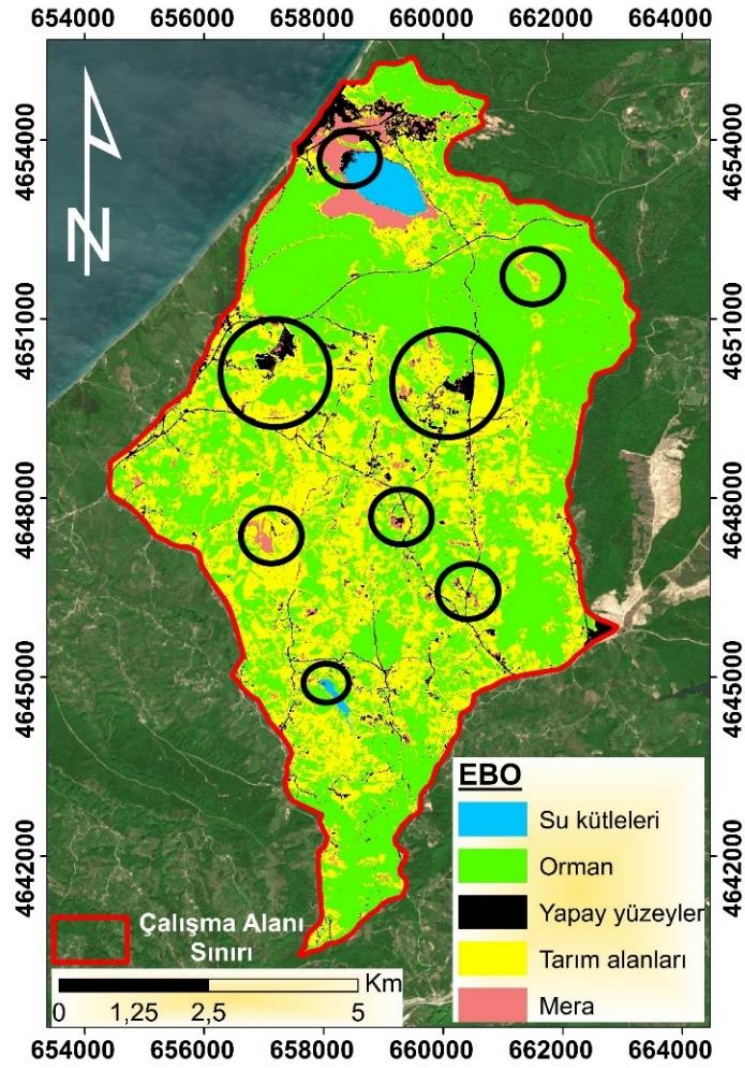
Eğitim setinin oluşturulduğu aşamada bazı arazi kullanımı/arazi örtüsü sınıfları için çok sayıda piksel toplanırken bazı sınıfların ise daha az piksel kullanılarak temsil edildiği durumlar oluşabilir. Eşit olmayan piksel dağılımı, sınıflandırma modelinin bazı sınıfları diğerlerine göre daha az temsil etmesine yol açabilir (Üstüner vd., 2018). Bu durumda, modelin daha çok temsil edilen sınıflara karşı daha iyi performans göstermesi muhtemeldir, ancak daha az temsil edilen sınıflar için performans düşebilir.

Düşük AUC değerleriyle birlikte, karışıklık matrislerindeki pozitif ve negatif olarak tahmin edilen örneklerdeki farklılıkların fazla olması, yani pozitif olarak tahmin edilen örneklerin negatif olarak tahmin edilenlerden yüksek olması ya da tam tersi durumun olması gibi nedenlerden dolayı veri setlerinin dengesiz dağıldığı durumlar veri setindeki dengesizliğe işaret etmektedir (Saito & Rehmsmeier, 2015). Aynı zamanda bazı sınıflar için doğruluk (accuracy) değeri yüksekken, hassasiyet (sensitivity) ya da kesinlik (precision) değerlerinin düşük olması da dengesiz bir verinin varlığına diğer bir kanıttır. Eşit olmayan sınıf dağılımı durumunun normal bir durum olduğu ve gerçek olan veri setlerinde karşılaşılabilecek bir durum olduğu birçok bilimsel çalışmada belirtilmiştir (Chawla vd., 2002; Chawla vd., 2004; Rao vd., 2006; Kubat vd., 1998; Khan vd., 2017; Dong vd., 2018; Wang vd., 2021).

Destek vektör makinesi, rastgele orman ve en büyük olasılık algoritmaları kullanılarak kontrollü sınıflandırılmış Sentinel-2A MSI uydu görüntüleri Şekil 12 ve Şekil 13'de verilmiştir. Şekil 12 ve Şekil 13'de verilen haritalar incelendiğinde, farklı algoritmalar kullanılarak sınıflandırılan uydu görüntüleri arasında çok küçük farklılıkların bulunduğu görülmektedir ve bu farklılıklar sınıflandırılmış uydu görüntülerinde siyah daire içerisine alınmıştır. Özellikle yapay yüzeyler sınıfının her bir sınıflandırılmış uydu görüntüsü üzerindeki farklı dağılımı dikkat çekicidir.



Şekil 12. DVM ve RO algoritmaları kullanılarak kontrollü sınıflandırılmış Sentinel-2A MSI uydu görüntüleri



Şekil 13. En büyük olasılık algoritması kullanılarak kontrollü sınıflandırılmış Sentinel-2A MSI uydu görüntüsü

4. Sonuç ve Öneriler

Bu çalışmada, Sinop havzasına yönelik arazi kullanımı/arazi örtüsü sınıflarının makine öğrenmesi tabanlı destek vektör makinesi, rastgele orman ve en büyük olasılık algoritmaları kullanılarak piksel tabanlı kontrollü sınıflandırılması ve tüm sınıflandırma algoritmalarına yönelik genel doğruluk (overall accuracy) değerleri ile, her bir arazi kullanımı/arazi sınıfına ait üretici doğruluğu (producer accuracy), kullanıcı doğruluğu (user accuracy), kesinlik (precision), doğruluk (accuracy), hassasiyet (sensitivity) ve F1-skoru metriklerine ait değerlerin kıyaslanması amaçlanmıştır.

Arazi kullanımı/arazi örtüsü hakkında detaylı bilgi sahibi olunması, çevre koruma, mekânsal planlama, ticari, politik ve idari amaçların yerine getirilmesinde önemli bir rol oynamaktadır. Bu bağlamda, uzaktan algılama ve coğrafi bilgi sistemleri tekniklerinin birlikte kullanılması, tematik verilerin elde edilmesinde kritik bir kaynaktır. Elde edilen sonuçlar DVM ve RO algoritmalarının EBO yöntemine göre daha yüksek ve benzer genel doğruluk değerleri verdiğini göstermiştir (0.88). Her bir sınıflandırma algoritması için su kütleleri ve mera sınıflarının en yüksek doğruluk, kesinlik, hassasiyet ve F1-skoru değerlerine sahip olduğu gözlemlenmiştir. Python programlama dilinden faydalanılarak oluşturulan ROC eğrileri ve bu eğrilerin altında kalan AUC değerleri incelendiğinde rastgele orman algoritmasıyla sınıflandırılan tarım alanları sınıfının, diğer arazi kullanımı/arazi örtüsü sınıflarına göre daha yüksek bir performans gösterdiği sonucuna varılmıştır. Ancak, dengesiz veri seti dağılımının çoğu arazi kullanımı/arazi örtüsü sınıfı için düşük AUC değerlerine yol açtığı ve bu durumun model performanslarını olumsuz yönde etkilediği görülmüştür. Bu sorunun, eğitim seti oluşturulurken bazı sınıflar için çok sayıda piksel toplanırken diğerlerinin daha az piksel ile temsil edilmesinden kaynaklanabileceği ve modelin bazı sınıfları diğerlerinden daha iyi tanımasına yol açabileceğinden ileri geldiği araştırmacılar tarafından ortaya konmuştur.

Bu çalışmada kullanılan her bir hata metriğinin kendine özgü avantajları ve dezavantajları bulunmaktadır. Veri setindeki dengesizliğin giderilmesi adına, daha fazla verinin toplanması, az örnekleme (undersampling) ya da aşırı örnekleme (oversampling) gibi çözüm yollarından faydalanılabilir.

Her bir metrik, verilerin dengeli ya da dengesiz dağılımına bağlı olarak farklılık gösterebileceğinden, gelecekte yapılacak çalışmalarda verilere ait sınıf dağılımlarının dikkate alınması ve en uygun hata metriğinin sınıflandırma algoritmasının doğruluğunun değerlendirilmesinde kullanılması araştırmacılar tarafından önerilmektedir. Bu çalışmada, ArcGIS 10.7v yazılımının bünyesinde bulundurduğu destek vektör makinesi, rastgele orman ve en büyük olasılık algoritmalarının sahip oldukları varsayılan (default) parametreler ile sınıflandırma işlemleri gerçekleştirilmiştir. Python programlama dili ya da Python ile entegre çalışabilen QGIS gibi açık kaynaklı bir yazılım, sınıflandırmada kullanılacak destek vektör makinesi ya da rastgele orman algoritmalarına ait parametrelerin optimizasyonunda kullanılarak daha sağlıklı doğruluk (accuracy) ya da AUC değerleri elde edilebilir.

Sonuç itibarıyla, uzaktan algılama ve coğrafi bilgi sistemleri teknolojilerinden faydalanılarak elde edilen bu çalışma ile farklı sınıflandırma algoritmalarının karşılaştırılması ve performans analizlerinin belirlenmesinin, bu alanda ileriye dönük yapılacak araştırmalara rehberlik edebilecek bir kaynak olacağı düşünülmektedir.

Kaynaklar

- Adam, E., Mutanga, O., Odindi, J., & Abdel-Rahman, E. M. (2014). Land-use/cover classification in a heterogeneous coastal landscape using RapidEye imagery: evaluating the performance of random forest and support vector machines classifiers. *International Journal of Remote Sensing*, 35(10), 3440-3458.
- Akar, Ö., & Güngör, O. (2012). Classification of multispectral images using Random Forest algorithm. *Journal of Geodesy and Geoinformation*, 1(2), 105-112.
- Alakus, T. B., & Turkoglu, I. (2020). Comparison of deep learning approaches to predict COVID-19 infection. *Chaos, Solitons & Fractals*, 140, Article 110120. <https://doi.org/10.1016/j.chaos.2020.110120>
- Bawa, A., Samanta, S., Himanshu, S. K., Singh, J., Kim, J., Zhang, T., ... & Ale, S. (2023). A support vector machine and image processing based approach for counting open cotton bolls and estimating lint yield from UAV imagery. *Smart Agricultural Technology*, 3, Article 100140. <https://doi.org/10.1016/j.atech.2022.100140>
- Billah, M., Islam, A. S., Mamoon, W. B., & Rahman, M. R. (2023). Random forest classifications for landuse mapping to assess rapid flood damage using Sentinel-1 and Sentinel-2 data. *Remote Sensing Applications: Society and Environment*, 30, Article 100947. <https://doi.org/10.1016/j.rsase.2023.100947>
- Braun, A., & Hochschild, V. (2017). A SAR-based index for landscape changes in African savannas. *Remote Sensing*, 9(4), Article 359. <https://doi.org/10.3390/rs9040359>
- Breiman, L. (2001). Random forests. *Machine learning*, 45, 5-32.
- Bromová, P., Škoda, P., & Vážný, J. (2014). Classification of spectra of emission line stars using machine learning techniques. *International Journal of Automation and Computing*, 11(3), 265-273.
- Campbell, J. B., & Wynne, R. H. (2002). *Introduction to remote sensing*. Guilford Press.
- Campbell, J. B., & Wynne, R. H. (2011). *Introduction to remote sensing*. Guilford Press.
- Carranza-García, M., García-Gutiérrez, J., & Riquelme, J. C. (2019). A framework for evaluating land use and land cover classification using convolutional neural networks. *Remote Sensing*, 11(3), Article 274. <https://doi.org/10.3390/rs11030274>
- Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 16, 321-357.
- Chawla, N. V., Japkowicz, N., & Kotcz, A. (2004). Special issue on learning from imbalanced data sets. *ACM SIGKDD explorations newsletter*, 6(1), 1-6. <https://doi.org/10.1145/1007730.1007733>
- Che, T., Xiao, L., & Liou, Y. A. (2014). Changes in glaciers and glacial lakes and the identification of dangerous glacial lakes in the Pumqu River Basin, Xizang (Tibet). *Advances in meteorology*, 2014, Article ID 903709. <http://dx.doi.org/10.1155/2014/903709>
- Chen, Z., Wang, L., Wei, A., Gao, J., Lu, Y., & Zhou, J. (2019). Land-use change from arable lands to orchards reduced soil erosion and increased nutrient loss in a small catchment. *Science of the Total Environment*, 648, 1097-1104.
- Civco, D. L. (1993). Artificial neural networks for land-cover classification and mapping. *International journal of geographical information science*, 7(2), 173-186.
- Congalton, R. G. (1991). A review of assessing the accuracy of classifications of remotely sensed data. *Remote sensing of environment*, 37(1), 35-46.
- Cunningham, P., Cord, M., & Delany, S. J. (2008). Supervised learning, machine learning techniques for multimedia. *Springer-Verlag Berlin, Heidelberg*.
- Çölkesen, İ. (2009). *Uzaktan algılamada ileri sınıflandırma tekniklerinin karşılaştırılması ve analizi* [Yüksek lisans tezi, Gebze Yüksek Teknoloji Enstitüsü]. YÖK Ulusal Tez Merkezi. <https://tez.yok.gov.tr/UlusalTezMerkezi>

- Dao, P. D., & Liou, Y. A. (2015). Object-based flood mapping and affected rice field estimation with Landsat 8 OLI and MODIS data. *Remote Sensing*, 7(5), 5077-5097.
- Demir, N., Sonmez, N.K., Akar, T., Ünal, S., (2018). Automated Measurement of Plant Height of Wheat Genotypes Using a DSM Derived from UAV Imagery. *Proceedings of the 2nd International Electronic Conference on Remote Sensing* 2(7), Article 350. <https://doi.org/10.3390/ecrs-2-05163>.
- Dietterich, T. G. (2020). An experimental comparison of three methods for constructing ensembles of decision trees: Bagging, boosting, and randomization. *Machine Learning*, 40, 139–157.
- Disperati, L., & Virdis, S. G. P. (2015). Assessment of land-use and land-cover changes from 1965 to 2014 in Tam Giang-Cau Hai Lagoon, central Vietnam. *Applied Geography*, 58, 48-64.
- Dong, R., Zhang, Y., & Zhao, J. (2018). How green are the streets within the sixth ring road of Beijing? An analysis based on tencent street view pictures and the green view index. *International journal of environmental research and public health*, 15(7), Article 1367. <https://doi.org/10.3390/ijerph15071367>
- ED Chaves, M., CA Picoli, M., & D. Sanches, I. (2020). Recent applications of Landsat 8/OLI and Sentinel-2/MSI for land use and land cover mapping: A systematic review. *Remote Sensing*, 12(18), Article 3062. <https://doi.org/10.3390/rs12183062>
- Eskandari, S., Sarab, S. A. M., Pourhashemi, M., & Ahmadloo, F. (2022). Selection of the best pixel-based algorithm for land cover mapping in Zagros forests of Iran using Sentinel-2A data: A case study in Khuzestan province. In H. R. Pourghasemi (Ed.), *Computers in Earth and Environmental Sciences* (pp. 181-190). Elsevier.
- Foody, G. M., & Mathur, A. (2004). Toward intelligent training of supervised image classifications: directing training data acquisition for SVM classification. *Remote Sensing of Environment*, 93(1-2), 107-117.
- Gao, Z., Cecati, C., & Ding, S. X. (2015). A survey of fault diagnosis and fault-tolerant techniques—Part I: Fault diagnosis with model-based and signal-based approaches. *IEEE transactions on industrial electronics*, 62(6), 3757-3767.
- Ghansah, B., Nyamekye, C., Owusu, S., & Agyapong, E. (2021). Mapping flood prone and Hazards Areas in rural landscape using landsat images and random forest classification: Case study of Nasia watershed in Ghana. *Cogent Engineering*, 8(1), Article 1923384. <https://doi.org/10.1080/23311916.2021.1923384>
- Gislason, P. O., Benediktsson, J. A., & Sveinsson, J. R. (2006). Random forests for land cover classification. *Pattern recognition letters*, 27(4), 294-300.
- Goutte, C., & Gaussier, E. (2005). A probabilistic interpretation of precision, recall and F-score, with implication for evaluation. In D. E. Losada & J. M. Fernández-Luna (Eds.), *Advances in Information Retrieval* (pp. 345-359). Springer Berlin Heidelberg.
- Gumma, M. K., Thenkabail, P. S., Teluguntla, P. G., Oliphant, A., Xiong, J., Giri, C., ... & Whitbread, A. M. (2020). Agricultural cropland extent and areas of South Asia derived using Landsat satellite 30-m time-series big-data using random forest machine learning algorithms on the Google Earth Engine cloud. *GIScience & Remote Sensing*, 57(3), 302-322.
- Guo, L., Chehata, N., Mallet, C., & Boukir, S. (2011). Relevance of airborne lidar and multispectral image data for urban scene classification using Random Forests. *ISPRS Journal of Photogrammetry and Remote Sensing*, 66(1), 56-66.
- Günlü, A. (2012). Landsat TM uydu görüntüsü yardımıyla bazı meşçere parametreleri (gelişim çağı ve kapalılık) ve arazi kullanım sınıflarının belirlenmesi. *Kastamonu University Journal of Forestry Faculty*, 12(1), 71-79.
- Halder, S., Das, S., & Basu, S. (2023). Use of support vector machine and cellular automata methods to evaluate impact of irrigation project on LULC. *Environmental Monitoring and Assessment*, 195(1), Article 3. <https://doi.org/10.1007/s10661-022-10588-6>
- Hanley, J. A., & McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, 143(1), 29-36.
- Hashem, N., & Balakrishnan, P. (2015). Change analysis of land use/land cover and modelling urban growth in Greater Doha, Qatar. *Annals of GIS*, 21(3), 233-247.
- Huang, C., Davis, L. S., & Townshend, J. R. G. (2002). An assessment of support vector machines for land cover classification. *International Journal of remote sensing*, 23(4), 725-749.
- Jensen, J. R. (2005). *Digital image processing: a remote sensing perspective* (3rd ed.). Prentice Hall.
- Kavzoğlu, T., & Çölkesen, İ. (2009). A kernel functions analysis for support vector machines for land cover classification. *International Journal of Applied Earth Observation and Geoinformation*, 11(5), 352-359.
- Kavzoğlu, T., & Çölkesen, İ. (2010). Destek vektör makineleri ile uydu görüntülerinin sınıflandırılmasında kernel fonksiyonlarının etkilerinin incelenmesi. *Harita Dergisi*, 144(7), 73-82.
- Khan, S. H., Hayat, M., Bennamoun, M., Soheli, F. A., & Togneri, R. (2017). Cost-sensitive learning of deep feature representations from imbalanced data. *IEEE transactions on neural networks and learning systems*, 29(8), 3573-3587.
- Kubat, M., Holte, R. C., & Matwin, S. (1998). Machine learning for the detection of oil spills in satellite radar images. *Machine learning*, 30, 195-215.

- Kundu, R. (2022, September 13). *Confusion Matrix: How To Use It & Interpret Results*. V7 Labs. <https://www.v7labs.com/blog/confusion-matrix-guide>
- Lillesand T., Kiefer R., & Chipman, J. (2004). *Remote Sensing and Image Interpretation*. John Wiley and Sons.
- Liou, Y. A., Nguyen, A. K., & Li, M. H. (2017). Assessing spatiotemporal eco-environmental vulnerability by Landsat data. *Ecological Indicators*, 80, 52-65.
- Mather, P., & Tso, B. (2016). *Classification methods for remotely sensed data*. CRC press.
- Melgani, F., & Bruzzone, L. (2004). Classification of hyperspectral remote sensing images with support vector machines. *IEEE Transactions on geoscience and remote sensing*, 42(8), 1778-1790.
- Merchant, J. W., & Narumalani, S. (2009). Integrating remote sensing and geographic information systems. In T.A. Warner, M. D. Nellis & G. M. Foody (Eds.), *The SAGE handbook of remote sensing* (pp. 257-268). SAGE Publications.
- Miao, J., & Zhu, W. (2022). Precision–recall curve (PRC) classification trees. *Evolutionary intelligence*, 15(3), 1545-1569. <https://doi.org/10.1007/s12065-021-00565-2>
- Moran, E. F., Skole, D. L., & Turner, B. L. (2004). The development of the international land-use and land-cover change (LUCC) research program and its links to NASA's Land-Cover and Land-Use Change (LCLUC) Initiative. In G. Gutman, A. C. Janetos, C. O. Justice, E. F. Moran, J. F. Mustard, R. R. Rindfuss D. Skole, B. L. Turner & M. A. Cochrane (Eds.), *Land Change Science: Observing, Monitoring and Understanding Trajectories of Change on the Earth's Surface* (Vol. 6, pp. 1-15). Springer. https://doi.org/10.1007/978-1-4020-2562-4_1
- Murty, P. S., & Tiwari, H. (2015). Accuracy assessment of land use classification—a case study of Ken Basin. *Journal of Civil Engineering and Architecture Research*, 2(12), 1199-1206.
- Muslim, M. A. (2020). Support vector machine (svm) optimization using grid search and unigram to improve e-commerce review accuracy. *Journal of Soft Computing Exploration*, 1(1), 8-15.
- Myung, I. J. (2003). Tutorial on maximum likelihood estimation. *Journal of mathematical Psychology*, 47(1), 90-100.
- Nakach, F. Z., Zerouaoui, H., & Idri, A. (2022). Random forest based deep hybrid architecture for histopathological breast cancer images classification. In O. Gervasi, B. Burgante, E. M. T. Hendrix, D. Taniar & B. O. Apduhan (Eds.), *Computational Science and Its Applications* (Vol. 13376, pp. 3-18). Springer.
- Nguyen, A. K., Liou, Y. A., Li, M. H., & Tran, T. A. (2016). Zoning eco-environmental vulnerability for environmental management and protection. *Ecological Indicators*, 69, 100-117.
- Nguyen, K. A., & Liou, Y. A. (2019). Global mapping of eco-environmental vulnerability from human and nature disturbances. *Science of The Total Environment*, 664, 995-1004.
- Nguyen, K. A., Liou, Y. A., Tran, H. P., Hoang, P. P., & Nguyen, T. H. (2020). Soil salinity assessment by using near-infrared channel and Vegetation Soil Salinity Index derived from Landsat 8 OLI data: a case study in the Tra Vinh Province, Mekong Delta, Vietnam. *Progress in Earth and Planetary Science*, 7, 1-16.
- Nitze, I., Schulthess, U., & Asche, H. (2012, May 7-9). *Comparison of machine learning algorithms random forest, artificial neural network and support vector machine to maximum likelihood for supervised crop type classification* [Conference paper]. 4th International conference on Geographic Object-Based Image Analysis (GEOBIA), Rio de Janeiro, Brazil.
- Okwuashi, O., & Ndehedehe, C. E. (2020). Deep support vector machine for hyperspectral image classification. *Pattern Recognition*, 103, Article 107298. <https://doi.org/10.1016/j.patcog.2020.107298>
- Orhan, O., Kirtiloğlu, O. S., & Yakar, M. (2020). Konya kapalı havzası obruk envanter bilgi sisteminin oluşturulması. *Geomatik*, 5(2), 81-90.
- Osenberg, M., Hilger, A., Neumann, M., Wagner, A., Bohn, N., Binder, J. R., ... & Manke, I. (2023). Classification of FIB/SEM-tomography images for highly porous multiphase materials using random forest classifiers. *Journal of Power Sources*, 570, Article 233030. <https://doi.org/10.1016/j.jpowsour.2023.233030>
- Öztürk, M. (2022, April 13). *Python ile Sınıflandırma Analizleri – Rastgele Orman (Random Forest) Algoritması*. <https://miracozturk.com/python-ile-siniflandirma-analizleri-rastgele-orman-random-forest-algoritmasi>
- Pal, M. (2005). Random forest classifier for remote sensing classification. *International journal of remote sensing*, 26(1), 217-222.
- Pal, M., & Mather, P. M. (2005). Support vector machines for classification in remote sensing. *International journal of remote sensing*, 26(5), 1007-1011.
- Pal, S., & Talukdar, S. (2018). Drivers of vulnerability to wetlands in Punarbhaba river basin of India-Bangladesh. *Ecological Indicators*, 93, 612-626.
- Phiri, D., Simwanda, M., Salekin, S., Nyirenda, V. R., Murayama, Y., & Ranagalage, M. (2020). Sentinel-2 data for land cover/use mapping: A review. *Remote Sensing*, 12(14), Article 2291. <https://doi.org/10.3390/rs12142291>
- Pouyan, S., Pourghasemi, H. R., Bordbar, M., Rahmanian, S., & Clague, J. J. (2021). A multi-hazard map-based flooding, gully erosion, forest fires, and earthquakes in Iran. *Scientific Reports*, 11(1), Article 14889. <https://doi.org/10.1038/s41598-021-94266-6>

- Rana, V. K., & Suryanarayana, T. M. V. (2020). Performance evaluation of MLE, RF and SVM classification algorithms for watershed scale land use/land cover mapping using sentinel 2 bands. *Remote Sensing Applications: Society and Environment*, 19, Article 100351. <https://doi.org/10.1016/j.rsase.2020.100351>
- Rao, R. B., Krishnan, S., & Niculescu, R. S. (2006). Data mining for improved cardiac care. *Acm Sigkdd Explorations Newsletter*, 8(1), 3-10.
- Rauf, U., Qureshi, W. S., Jabbar, H., Zeb, A., Mirza, A., Alanazi, E., ... & Rashid, N. (2022). A new method for pixel classification for rice variety identification using spectral and time series data from Sentinel-2 satellite imagery. *Computers and Electronics in Agriculture*, 193, Article 106731. <https://doi.org/10.1016/j.compag.2022.106731>
- Rimal, B., Rijal, S., & Kunwar, R. (2020). Comparing support vector machines and maximum likelihood classifiers for mapping of urbanization. *Journal of the Indian Society of Remote Sensing*, 48(1), 71-79.
- Rodriguez-Galiano, V. F., Ghimire, B., Rogan, J., Chica-Olmo, M., & Rigol-Sanchez, J. P. (2012). An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS journal of photogrammetry and remote sensing*, 67, 93-104.
- Saito, T., & Rehmsmeier, M. (2015). The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLoS one*, 10(3), Article e0118432. <https://doi.org/10.1371/journal.pone.0118432>
- Samaniego, L., Bárdossy, A., & Schulz, K. (2008). Supervised classification of remotely sensed imagery using a modified k -NN technique. *IEEE Transactions on Geoscience and Remote Sensing*, 46(7), 2112-2125.
- Schneider, C. A., Rasband, W. S., & Eliceiri, K. W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nature methods*, 9(7), 671-675.
- Selim, S., & Demir, N. (2018). Analysis of landscape patterns and connectivity between tree clusters derived from LIDAR data. *Fresenius Environmental Bulletin*, 27(5A), 3512-3520.
- SEOS, (2018). *Introduction to Remote Sensing*. <https://seos-project.eu/remotesensing/remotesensing-c00-p02.html>
- Sheykhou, M., Mahdianpari, M., Ghanbari, H., Mohammadimanesh, F., Ghamisi, P., & Homayouni, S. (2020). Support vector machine versus random forest for remote sensing image classification: A meta-analysis and systematic review. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 13, 6308-6325.
- Soyaslan, İ., & Hepdeniz, K. (2016). Coğrafi bilgi sistemleri ve uzaktan algılama kullanılarak Burdur ili arazi kullanımının zamansal değişiminin belirlenmesi. *Mehmet Akif Ersoy Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, 7(2), 94-101.
- Su, J., Yi, D., Liu, C., Guo, L., & Chen, W. H. (2017). Dimension reduction aided hyperspectral image classification with a small-sized training dataset: experimental comparisons. *Sensors*, 17(12), Article 2726. <https://doi.org/10.3390/s17122726>
- Talukdar, S., & Pal, S. (2019). Effects of damming on the hydrological regime of Punarbhaba river basin wetlands. *Ecological Engineering*, 135, 61-74.
- Talukdar, S., Singha, P., Mahato, S., Praveen, B., & Rahman, A. (2020). Dynamics of ecosystem services (ESs) in response to land use land cover (LU/LC) changes in the lower Gangetic plain of India. *Ecological Indicators*, 112, Article 106121. <https://doi.org/10.1016/j.ecolind.2020.106121>
- Tao, D., Tang, X., Li, X., & Wu, X. (2006). Asymmetric bagging and random subspace for support vector machines-based relevance feedback in image retrieval. *IEEE transactions on pattern analysis and machine intelligence*, 28(7), 1088-1099.
- Tehrany, M. S., Pradhan, B., Mansor, S., & Ahmad, N. (2015). Flood susceptibility assessment using GIS-based support vector machine model with different kernel types. *Catena*, 125, 91-101.
- Tian, S., Zhang, X., Tian, J., & Sun, Q. (2016). Random forest classification of wetland landcovers from multi-sensor data in the arid region of Xinjiang, China. *Remote Sensing*, 8(11), Article 954. <https://doi.org/10.3390/rs8110954>
- Tiwari, R., Sharma, R., & Dubey, R. (2022, September). Microstrip Patch Antenna Parameter Optimization Prediction Model using Machine Learning Techniques. *International Journal on Recent and Innovation Trends in Computing and Communication*, 10(9), 53-59.
- Topaloğlu, R. H., Sertel, E., & Musaoğlu, N. (2016, July 12-19). *Assessment of classification accuracies of Sentinel-2 and Landsat-8 data for land cover/use mapping* [Conference paper]. The International archives of the photogrammetry, remote sensing and spatial information sciences, Prague, Czech Republic.
- Üstüner, M., Balık Şanlı, F., & Abdikan, S. (2014, 14-17 Ekim). *Kırmızı-Kenar Ve Yakın Kızılötesi Bantlarının Ürün Deseni Sınıflandırma Doğruluğuna Olan Etkisinin Araştırılması* [Tam metin bildirisi]. V. Uzaktan Algılama ve Coğrafi Bilgi Sistemleri Sempozyumu, Türkiye.
- Üstüner, M., Gökdağ, Ü., Bilgin, G., & Şanlı, F. B. (2018, May 2-5). *Comparing the classification performances of supervised classifiers with balanced and imbalanced SAR data sets* [Conference paper]. 26th Signal Processing and Communications Applications Conference, İzmir, Türkiye.

- Wang, L., Zhang, L., Qi, X., & Yi, Z. (2021). Deep attention-based imbalanced image classification. *IEEE transactions on neural networks and learning systems*, 33(8), 3320-3330.
- Yan, G., Mas, J. F., Maathuis, B. H. P., Xiangmin, Z., & Van Dijk, P. M. (2006). Comparison of pixel-based and object-oriented image classification approaches—a case study in a coal fire area, Wuda, Inner Mongolia, China. *International Journal of Remote Sensing*, 27(18), 4039-4055.
- Zhang, T. X., Su, J. Y., Liu, C. J., & Chen, W. H. (2019). Potential bands of sentinel-2A satellite for classification problems in precision agriculture. *International Journal of Automation and Computing*, 16, 16-26.
- Zhang, T., Su, J., Liu, C., Chen, W. H., Liu, H., & Liu, G. (2017, September 7-8). *Band selection in sentinel-2 satellite for agriculture applications* [Conference paper]. 23rd International Conference on Automation and Computing, Huddersfield, UK.
- Zhang, Y., Ge, T., Tian, W., & Liou, Y. A. (2019). Debris flow susceptibility mapping using machine-learning techniques in Shigatse area, China. *Remote Sensing*, 11(23), Article 2801. <https://doi.org/10.3390/rs11232801>