



Çok zamanlı Sentinel 2 uydu görüntüleri ve makine öğrenmesi tabanlı algoritmalar ile arazi örtüsü değişiminin belirlenmesi

Esmâ Efe¹ , Ugur Algancı^{1*} 

¹*İstanbul Teknik Üniversitesi, İnşaat Fakültesi, Geomatik Mühendisliği Bölümü, İstanbul, Türkiye*

Anahtar Kelimeler

Uzaktan Algılama
Sentinel 2
LandCover 2.0
Makine Öğrenmesi
Boyut İndirgeme

Araştırma Makalesi

Geliş:24.03.2022
Revize:20.05.2022
Kabul:24.05.2022
Yayınlanma:10.08.2022



Öz

Yeryüzündeki değişimin saptanması ve izlenmesi her zaman dikkate değer bir konu olmuştur. Zaman içinde insan faaliyetleri genişlemiş ve arazi örtüsünde bu faaliyetlerin etkisi açıkça görülmüştür. Sanayi faaliyetlerinin artması, yerleşimin fazlaşması gibi konular nedeniyle arazi örtüsündeki değişimi izlenmesi ve takip edilmesi karar vericiler açısından kritik bir konu haline gelmiştir. Uzaktan algılama alanında bu konu ile ilgili çokça çalışma yapılmış ve yeryüzündeki değişimin en doğru sonuçla belirlenmesi için yöntemler ve araçlar sürekli gelişim göstermiştir. Çalışma kapsamında, Kocaeli ilinde şehirleşme ve tarımsal faaliyete bağlı olarak gelişen arazi örtüsü değişimini LandCover 2.0 standartlarına göre dinamik değişim belirleme çerçevesinde saptamak için çok zamanlı Sentinel 2 uydu görüntüleri kullanılarak Yerleşim İndisi-Rastgele Orman, Temel Bileşen Analizi-Rastgele Orman, Yerleşim İndisi-Regresyon Ağacı ve Temel Bileşen Analizi-Regresyon Ağacı olmak üzere dört farklı veri indirgeme – sınıflandırma yöntem kombinasyonlarının başarımı karşılaştırmalı olarak değerlendirilmiştir. Google Earth Engine platformu üzerinde gerçekleştirilen sınıflandırma analizlerinin sonuçları tematik harita haline getirilmiş ve doğruluk değerlendirmesi gerçekleştirilmiştir. Çalışma sonucunda yüzde 83,88 doğruluk oranı ile Temel Bileşen Analizi-Regresyon Ağacı yöntem ikilisinin en yüksek doğruluk sağlayan yaklaşım olduğu ortaya konmuştur.

Determination of land cover change with multi-temporal Sentinel 2 satellite images and machine learning-based algorithms

Keywords

Remote sensing
Sentinel 2
LandCover 2.0
Machine Learning
Dimension Reduction

Research Article

Received:24.03.2022
Revised: 20.05.2022
Accepted:24.05.2022
Published:10.08.2022

Abstract

Detecting and monitoring change on Earth has always been a subject of considerable interest. Over time, human activities have expanded and the impact of these activities on the land cover has been clearly seen. Detecting and monitoring the change in land cover has become a critical issue for decision-makers due to issues such as the increase in industrial activities and the increase in settlement. Several works have been performed on this subject in the field of remote sensing, and methods and tools have continuously improved to determine the change in the earth to achieve the most accurate result. Within the scope of the study, multi-temporal Sentinel 2 satellite images were used in order to determine the land cover change due to urbanization and agricultural activity in Kocaeli province within the framework of dynamic change determination according to LandCover 2.0 standards. Four different data reduction – classification method combinations were applied, which are Built-up Index-Random Forest, Principal Component Analysis-Random Forest, Built-up Index-Regression Tree and Principal Component Analysis-Regression Tree and their performances were evaluated comparatively. The results of the classification analyses performed on the Google Earth Engine platform were turned into thematic maps and an accuracy assessment was carried out. As a result of the study, it has been revealed that Principal Component Analysis-Regression Tree method pair is the approach that provides the highest accuracy, with an accuracy rate of 83.88 percent.

*Sorumlu Yazar

(efee16@itu.edu.tr) ORCID ID 0000 – 0002 – 2500 – 365X
(alganci@itu.edu.tr) ORCID ID 0000 – 0002 – 5693 – 3614

Kaynak Göster (APA)

Efe, E., & Algancı, U. (2023). Çok zamanlı Sentinel 2 uydu görüntüleri ve makine öğrenmesi tabanlı algoritmalar ile arazi örtüsü değişiminin belirlenmesi. *Geomatik, 8(1), 27-34*

1. Giriş

Bir bölgenin arazi örtüsü ve arazi kullanımı (AÖAK) senaryosu, zaman ve uzayda insanlar tarafından kullanılan ve değiştirilen dünya yüzeyinin fiziksel ve doğal özelliklerinin sonucudur (Rawat ve Kumar 2015). Bu senaryoyu saptamak; sağlık, ekoloji, politika yönetimi, tarım ve afet yönetimi (Bégué ve ark., 2018) gibi çeşitli alanlardaki çalışmalara önemli katkılar sağlar. AÖAK sınıflandırması, insan faaliyetleri ve fiziksel çevre ile yakın bir ilişki gösteren yeryüzünü izlemek için kilit bir değer oluşturur. Toplumların hızlı gelişimi, çeşitli türlerdeki faaliyetlere bağlı olarak arazi örtüsü-arazi kullanımı üzerinde fark edilir bir etkiyle sonuçlanır (Petropoulos ve ark., 2013). Bu etkiyi ortaya koymak ve izlemek için çeşitli araçların kullanımı gelişen teknolojiye bağlı olarak değişerek süregelmiştir. AÖAK değişimini saptamak karar vericilere altlık oluşturmak açısından da önemli bir araştırma konusudur (Morsy ve Hadi 2022; Ahady ve Kaplan 2022).

AÖAK modellerini ve dinamiklerini haritalamak için, geleneksel yersel haritalama ve uydu tabanlı haritalama dahil olmak üzere çeşitli teknikler geliştirilmiştir. Saha ölçmesi olarak bilinen harita üretim yaklaşımı, haritanın farklı düzeylerde hassasiyetle bilgileri birleştiren çeşitli ölçeklerde üretilebildiği doğrudan bir harita üretme yoludur. Ancak büyük alanlar için insan gücü temelli yaklaşım, zaman ve maliyet yoğun bir seçenektir (Langat ve ark., 2019). Öte yandan uzaktan algılama uygulamalarının bu hedef için doğru araçlara sahip olduğu yıllardır yapılan araştırma ve projelerle ortaya konmuştur. Arazi örtüsü - arazi kullanımı değişimini ortaya koyarken yeterli doğruluğa sahip çalışmalar yapmak amacıyla çeşitli yaklaşımlar ve yöntemler denenmiştir. Bu kapsamda yapılan sınıflandırma işlemleri öne çıkmış, sınıflandırma için kullanılan algoritmalar süreç içinde çeşitlenmiş ve teknolojinin gelişmesi uygun yöntemlerin sayıca artmasına yardımcı olmuştur (Apaydın ve Abdikan 2021).

Özellikle yüksek mekansal çözünürlüklü sensörlerde görülen son gelişmelerle birlikte, yer gözlem teknolojisi, şehirlerin kırsal kesimlerinde AÖAK tanımlama ve yönetimi için uygun bir çözüm sunmaktadır. Bununla birlikte, yüksek doğrulukta AÖAK sınıflandırması ortaya koymak zor bir süreçtir. Özellikle kırsal ve kenti bir arada barındıran heterojen alanların sınıflandırılmasında halen zorluklar ve sorunlar yaşanmaktadır (Ruiz-Luna ve Berlanga-Robles, 2003).

Uzaktan algılanan görüntülerden doğru AÖAK bilgisinin üretilmesi, optimum görüntü sınıflandırma teknikleri gerektirir. Genel olarak bu sınıflandırıcılar, kontrollü ve kontrolsüz veya parametrik ve parametrik olmayan veya katı ve bulanık sınıflandırma veya piksel tabanlı ve alt piksel tabanlı sınıflandırıcılar olarak gruplanabilir. Sınıflandırıcıların performansı eğitim örneklerinin seçimi, çalışma alanının heterojenliği, sensörler, tanımlanacak sınıf sayısı vb. gibi çeşitli faktörlerden etkilenmektedir (Lu ve Weng, 2007). Yeni nesil sınıflandırıcılardaki gelişmeyi tanımlamak için farklı algoritmaların sistematik bir karşılaştırmalı analizi önemlidir.

Landsat-8 ve Sentinel-2 gibi ücretsiz erişimli uydu verileri, görüntü sınıflandırma algoritmalarının uzaktan

algılama alanına yönelik kullanımını artırmıştır (Belward ve Skøien 2015; Harris ve Baumann 2015). Öte yandan, kişisel bilgisayarların bilgi işlem gücü artarken, maliyeti hızla düşmektedir (Waldrop, 2016). Bu kapsamda, sınıflandırma çalışmalarında makine öğrenmesi algoritmalarının kullanılması zamanla gündeme gelmiştir. Gelişmiş makine öğrenme algoritmalarının kullanımı, AÖAK sınıflandırması ve haritalandırılması çalışmalarında daha yaygın hale gelmektedir (Jamali, 2019). Yu ve ark. (2014) karşılaştırmalı literatür değerlendirmeleri sonucunda son yıllarda makine öğrenmesi tabanlı sınıflandırıcıların klasik sınıflandırıcılara göre daha iyi performans gösterdiğini bildirmiştir. Makine öğrenmesi kökenli sınıflandırıcılar arasında karar ağacı tabanlı yaklaşımlar hızlı hesaplama kabiliyetleri, parametrik olmayan yapıları, sınıf tanımlarının gerçekleştirilmesinde basit ve anlaşılır bir yaklaşım sunmaları nedeniyle ön plana çıkmaktadır (Kavzoğlu ve Çölkesen, 2010).

Ulusal ölçekte makine tabanlı yaklaşımlar ile uydu görüntülerinden AÖAK haritası üretimi kapsamında birçok çalışma olup bunların bir kısmı karar ağacı tabanlı algoritmaları da çalışmalarına konu etmişlerdir. Genel hatları ile bu çalışmalarda sınıf tanımlarının statik olarak yapıldığı, tek ya da iki tarihli görüntü üzerinde çalışıldığı, değerlendirmelerin farklı sensör verisi ya da farklı sınıflandırıcıların başarımı üzerine yoğunlaştıkları söylenebilir (Akar ve Güngör, 2012; Akar ve Tunç Görmüş, 2019; Kavzoğlu ve Çölkesen, 2010)

Değişim belirlemeye yönelik geçmiş çalışmalar incelendiğinde ise uydu görüntüleri ile AÖAK değişim belirleme süreçlerinin farklı tarihte algılanmış olan görüntülerin statik sınıf tanımlarına göre sınıflandırılması ve sonuç haritalar üzerinde değişimin belirlenmesi şeklinde sürdürüldüğü, son yıllarda ise tüm görüntülerin bir arada değerlendirildiği ve değişimlerin de sınıf tanımlamalarına dahil edildiği dinamik yaklaşımların ön plana çıktığı görülmektedir (Alganci, 2019; Wulder ve ark., 2018).

Bu çalışma kapsamında özellikle şehirleşme ve tarımsal faaliyete bağlı olarak arazi örtüsünde meydana gelen değişimlerin belirlenmesi amacı ile LandCover 2.0 standartlarına göre Google Earth Engine tabanlı dinamik değişim belirleme çerçevesi tasarlanmıştır. Bu kapsamda ücretsiz elde edilebilen Sentinel 2 uydu görüntüleri kullanılmış, açık kaynak kodlu görüntü işleme yazılımlarından faydalanılmıştır. Çok zamanlı uydu görüntülerinden boyut indirgeme yaklaşımı ile zaman serisi oluşturma, rastgele orman (RO), ve regresyon ağaçları (RA) gibi makine öğrenmesi tabanlı algoritmalar ile görüntü sınıflandırma, sonuçların haritalandırılması ve değişime ilişkin sınıf bazında istatistiksel bilgilerin üretilmesi gibi temel analiz süreçleri gerçekleştirilmiştir. Çalışmanın temel hedefi kapsamında, bu farklı boyut indirgeme ve sınıflandırma yöntemi kombinasyonlarının başarımları karşılaştırmalı olarak değerlendirilmiştir.

2. Yöntem

2.1. Çalışma Alanı ve Veriler

Türkiye'nin en kalabalık onuncu şehri olan Kocaeli, ülkenin en büyük sanayi kentlerinden biridir. 2021 yılı

itibariyle nüfusu 2.033.411 kişi olarak belirlenmiştir. Kocaeli’de yıllık nüfus artış oranı %18’dir, tüm ilçelerinde nüfus artışı olmuştur. 1965 yılından itibaren nüfusun arttığı bölgede, şehirdeki nüfus sürekli artarken kırsaldaki nüfus azalmıştır (URL-1). İstanbul, Bursa, Sakarya ve Yalova illeriyle komşu olan Kocaeli, metropol kentine yakınlığıyla komşu illerin etkilerini taşımaktadır.

İlde Kocaeli Sanayi Odası’na bağlı yaklaşık 3500 sanayi kuruluşu vardır ve bu sanayi kuruluşları genelde Gebze, İzmit ve Körfez ilçelerinde toplanmıştır. Önemli markaların fabrikalarının yer aldığı il, önemli bir sanayi ve ticaret merkezidir (URL-2). Kocaeli’nin göç alan bir bölge olması, kurulan fabrikalar, yapılması tasarlanan/yapılan kentleşme çalışmaları vb. gibi etkilerle arazi örtüsü/arazi kullanımı açısından incelenmeye uygun olduğunu ortaya koymaktadır. Çalışma alanına ait görsel Şekil 1’de verilmektedir.

Çalışma kapsamında; Kocaeli ilinin idari sınırlarını kapsayan ve 2017-2020 yılları arasında elde edilen Sentinel 2A uydu görüntüleri kullanılmıştır. Kullanılan görüntüler, zamansal aralığın düzenliliğini sağlamak ve mevsimsel etkileri en aza indirmek için aynı yaz mevsiminde olacak şekilde seçilmiştir. Bu kapsamda, 29 Haziran 2017, 4 Temmuz 2018, 29 Temmuz 2019 ve 23 Temmuz 2020 tarihli görüntüler kullanılmıştır.

Sentinel-2 ikiz uyduları, Avrupa Komisyonu ve Avrupa Uzay Ajansı tarafından ortaklaşa uygulanan Çevre ve Güvenlik için Küresel İzleme programına veri sağlamak amacı ile fırlatılmış olup algılanan görüntüler son kullanıcılar tarafından ücretsiz elde edilebilir. Uyduların yörünge yüksekliği ortalama 785 km olup, iki eşlenik uydu takımı sayesinde, ekvator da 5 günde bir ve orta enlemlerde 2-3 günde bir tekrarlanan çekimler mümkün olmaktadır.



Şekil 1. Çalışma alanının Türkiye üzerinde coğrafi konumu (Google Earth©) ve 2020 tarihli Sentinel 2A uydu görüntüsü ile yakın görünümü

2.2. Ön işleme ve boyut indirgeme

Çalışma kapsamında elde edilen Sentinel 2A uydu görüntülerinin ön işleme adımları kapsamında 20 m mekânsal çözünürlüklü bantların 10 m mekânsal çözünürlüklü bantlarla eşlenmesi amacı ile yeniden örneklenmesi ilk aşamayı oluşturmaktadır. Bu işlem kapsamında ilgili bantlara kübik enterpolasyon işlemi uygulanmıştır. Sonraki aşamada uydu görüntüleri il sınıırı vektör verisine göre kesilmiştir. Temin edilen görüntüler

2A işleme seviyesinde olduklarından geometrik ve atmosferik düzeltmeleri hâlihazırda gerçekleştirilmiş olup bantlar yüzey yansıtım değerlerini içermektedir.

Çalışmada kullanılan boyut indirgeme metodlarından biri spektral indeks tabanlı yöntemdir. Araştırmada şehirleşmeye bağlı olarak arazi örtüsünde meydana gelen değişimler saptanacağından, yerleşim indeksine dayalı yaklaşım benimsenmiştir. Bu kapsamda Normalleştirilmiş Yerleşim İndeksi (NDBI) ve Normalize Edilmiş Fark Bitki Örtüsü İndeksi (NDVI)’nin entegrasyonu ile Yerleşim İndeksi (Yİ) (Builtup İndice-BUI) Erdas Imagine yazılımı kullanılarak türetilmiştir. Denklem (1) yerleşim indeksinin hesaplanmasın ilişkin formülü içermektedir.

$$Yİ = \frac{SWIR - NIR}{SWIR + NIR} - \frac{NIR - RED}{NIR + RED} \quad (1)$$

SWIR, kısa dalga kızılötesi bandına karşılık gelirken, NIR yakın kızılötesi bandına ve RED kırmızı görüntü bandına karşılık gelir (Alganci, 2019). Yİ verileri uydu görüntülerinden hesaplanmış ve tek bir bant olarak saklanmıştır (Şekil 2a).

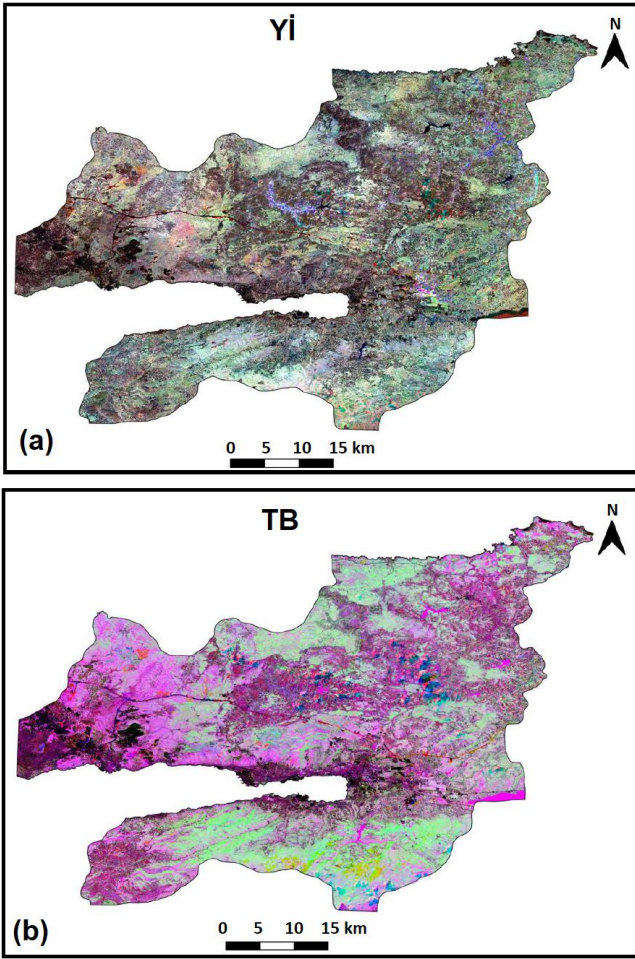
Çalışmada kullanılan diğer boyut indirgeme metodu Temel Bileşen (TB) analizidir. TB, orijinal görüntü bantlarının oluşturduğu n boyutlu öznelik uzayından varyansı maksimize edecek şekilde birbirine dik yeni eksenler tanımlanması ile yeni bir öznelik uzayı oluşturmayı hedefleyen istatistiksel bir prosedürdür. Analiz sonucunda verilerin dönüşümü ve veri boyutunun azaltılması gürültüyü, fazlalığı ve ilgisiz bilgileri ortadan kaldırarak verileri sıkıştırır. Yeni çok değişkenli öznelik uzayındaki korelasyonu minimuma indirgenmiş değişkenlere temel bileşenler denir. İlk temel bileşen (ilk öz vektörden türetilen TB1), projeksiyonların en büyük varyansa sahip olduğu uzaydaki yöndür. Sonraki ana bileşen (TB2), önceki ana bileşene ortogonal olan tüm yönler arasındaki varyansı en üst düzeye çıkararak yöndür. Geri kalan ana bileşen görüntülerinin varyansları, karşılık gelen öz değerlerin büyüklükleriyle orantılı olarak azalır (Deng ve ark., 2008).

Temel Bileşen Analizi, multispektral görüntüler için anlamlı parametre çıkarımı, verilerdeki gürültünün azaltılması ve AÖAK değişikliklerinin sınıflandırmasında doğruluğun artırılması gibi çeşitli konularda kullanılmıştır (Rana ve Suryanarayana, 2019). Bu çalışmada TB verileri Erdas Imagine yazılımı ile uydu görüntülerinden hesaplanmış ve birinci bileşen tek bir bant olarak saklanmıştır (Şekil 2b).

2.3. Arazi örtüsü sınıf tanımları

Çalışmada kullanılan arazi örtüsü sınıfları Anderson (1976) Arazi Örtüsü lejantı ve Sentinel görüntülerinin çözümlenme kapasitesi göz önüne alınarak tanımlanmıştır (Anderson ve ark, 1976). Ayrıca Wulder ve diğerleri tarafından tanımlanan “Land Cover 2.0” kavramından esinlenilerek dinamik değişim sınıfı olarak bir adet değişiklik sınıfı dâhil edilmiştir (Wulder ve ark., 2018)

Bu kapsamda belirlenen sınıflar ve tanımları Tablo 1’de verilmiştir.



Şekil 2. Boyut indirgenmiş veri setinin RGB görünümü, a) YI, b) TB (2017:R 2018:G 2020:B)

Tablo 1. Çalışmada haritaları oluşturulan arazi örtüsü sınıflarının tanımları.

AÖ Sınıfı	Tanım
Yapay Alanlar	Yerleşim alanları, sanayi ve ticaret alanları, ulaşım ağları, limanlar, havalimanları, şantiyeler, harfiyat sahaları
Ormanlar	İğne yapraklı ormanlar, geniş yapraklı ormanlar, karışık ormanlar
Su Kütleleri	Nehirler, göller, rezervuarlar
Çıplak ve Yarıdoğal Alanlar	Çıplak kaya, çıplak toprak, seyrek bitkili alanlar
Tarım Alanları	Ekilebilir araziler, yıllık bitkiler, daimi bitkiler, meralar
Kentsel Genişleme	2017-2020 yılları arasında yapılaşmaya dönüşen alanlar

2.4. Sınıflandırma

Bu çalışmada mevcut AÖAK durumunu ve üç yıllık bir süre içindeki değişiklikleri ortaya koymak amacıyla iki makine öğrenmesi tabanlı algoritma test edilmiştir.

Bunlardan ilki olan RA algoritması (Breiman ve diğ., 1984) uzaktan algılama uygulamalarında yaygın olarak kullanılan karar ağacı tabanlı bir yaklaşım sunar (Friedl ve ark., 2002). Bir karar ağacının yapısını belirlemenin

anahtarı her bölmede bir tanımlayıcı öznelik ve bu özneliğe bağlı olarak alt küme bölünmesini sağlayacak eşik değeri seçmektir. Bu yaklaşımda ağacın her bir düğümünde ilgili katmandaki verinin alt kümelere bölünmesi için tanımlayıcı özneliğin seçilmesi önem taşır. Bu kapsamda öznelik kümesi kendi içinde karşılaştırılarak en yüksek normalleştirilmiş bilgi kazancı değerine sahip öznelik seçimi gerçekleştirilir (Bishop, 2006). RA algoritması, eğitim verilerine uyacak şekilde büyümesine izin verilirse, aşırı uyumla sonuçlanabilir (Shao ve Lunetta, 2011). Bahsi geçen aşırı uyum riskine rağmen, sınıflandırma doğruluğu ve hız performansı sebebiyle Regresyon Ağacı algoritması yaygın olarak kullanılan bir AÖAK sınıflandırma algoritması olarak karşımıza çıkmaktadır (Lawrence ve Wright, 2001).

Bahsi geçen dezavantaj RO sınıflandırıcı algoritmalar ile başarıyla ele alınabilir (Shetty ve ark., 2021). RO sınıflandırıcı, rastgele seçilen örneklem (eğitim) kümesi ve bu örneklemelerden elde edilen değişkenler alt kümesini kullanarak birden çok karar ağacı üreten bir sınıflandırıcıdır. Bu sınıflandırıcı, sınıflandırmalarının doğruluğu nedeniyle uzaktan algılama topluluğu içinde popüler hale gelmiştir (Belgiu ve Drăgu, 2016). RO, farklı karar ağaçları ile elde edilen sınıflandırıcıların birleşiminin tek bir değişken kümesi ile çalıştırılan sınıflandırıcıdan daha iyi performans gösterebileceği ön görüşüne dayanan bir toplu öğrenme algoritmasıdır (Breiman, 2001). Bu kapsamda her bir karar ağacı için örneklem verilerinin arasından rastgele seçilerek oluşturulan alt eğitim seti kullanılarak ağaçlar arasındaki parametrik bağımlılık (korelasyon) en alt seviyeye indirgenir. Bu kapsamda eğitim seti olarak kullanılmayan örneklemelerin oranı değerlendirmeye dahil edilir ve "torba dışı" örnekler olarak adlandırılır. Farklı değişken grupları ile oluşturulan çok sayıda karar ağacı ile elde edilen sonuçların aritmetik ortalaması, her sınıfın çoğunluk oyu olarak tanımlanmaktadır.

RO algoritmasının çalışma prensibine göre kullanıcının tanımladığı iki parametre sınıflandırmayı gerçekleştirmek için gereklidir. Bu parametreler her bir düğümde alt bölünmeyi hesaplamak için kullanılan değişkenlerin sayısı (m) ve toplamda üretilecek olan ağaçların sayısı (N)'dir (Breiman ve Cutler, 2005).

Çalışmada, eğitim örnekleri veri setleri üzerinden manuel olarak seçilmiştir. Seçim yapılırken Google Earth üzerinden kontrol yapılarak ilerlenilmiştir. Eğitim örneklerinin sınıf dağılımı şu şekilde özetlenebilir: Yapay alanlar (153 adet poligon), ormanlar (160 adet poligon), su kütleleri (134 adet poligon), çıplak ve yarıdoğal alanlar (106 adet poligon), tarım alanları (159 adet poligon), kentsel genişlemenin görüldüğü alanlar (32 adet poligon). Eğitim örneklerinin farklı sınıflar arasında spektral ayrılabilirliği, seçilen örneklerin istatistiksel olarak ne kadar iyi eşleştiğini gösteren önemli bir ölçüdür (Alganci, 2019).

Google Earth Engine ile sınıflandırma algoritmalarının veri setleri üzerinde uygulanması esnasında Tablo 2'de verilen parametre seti uygulanmıştır.

Tablo 2. Farklı sınıflandırıcılar için Google Earth Enginede kullanılan girdi parametre değerleri

Sınıflandırıcı	Parametre	Değer	Referans
Regresyon Ağacı	Budama için çapraz doğrulama faktörü	5 ve 10	Kohavi, 1995
	Ağaç sayısı	50, 100, 150, 200	Belgiu ve Dragut 2016
Rastgele Orman	Bölme başına değişken sayısı	Girdi değişkenlerin karekökü	

2.5. Doğruluk analizi

Doğruluk değerlendirmesi, çeşitli sınıflandırıcıların performansının ve eğitim örnekleme tasarımlarının etkisinin değerlendirilmesine yardımcı olur. Sınıflandırıcıların performansını değerlendirmek için genel doğruluk, kullanıcı doğruluğu ve üretici doğruluğu metrikleri kullanılır. Genel doğruluk, doğru sınıflandırılan birimlerin sayısı ile bunların toplam sayısı arasındaki oran tarafından verilen toplam sınıflandırma yüzdesidir (Congalton ve Green, 2019). Kullanıcı ve üretici doğruluğu ise sınıf bazında sınıflandırma doğruluğuna atıfta bulunur (Story ve Congalton, 1986). Kullanıcı doğruluğu belirli bir sınıftaki doğru sınıflandırılmış ve tüm sınıflandırılmış birimler arasındaki oran iken, üretici doğruluğu doğru sınıflandırılmış birimler sayısı ile belirli bir sınıftaki doğrulama birimlerinin sayısı arasındaki orandır (Pratico ve ark., 2021).

Çalışmada her sınıf için 30 adet olacak şekilde rastgele kontrol noktası ile bahsedilen metrikler hesaplanmıştır.

Çalışma kapsamında bu metriklere ek olarak her sınıflandırıcı-boyut indirgeme kombinasyonu için kappa metriği hesaplanmıştır (Congalton, 1991). Ayrıca her sınıflandırma metodu sonucu ortaya çıkan alanlar, sınıf bazında performans değerlendirilmesi ve karşılaştırması yapabilmek adına ortaya konmuştur.

Bu analizlere ek olarak sınıflandırma doğruluk analizi kapsamında elde edilen dört adet hata matrisi kullanılarak bu matrisler arasında anlamlı bir fark olup olmadığını test etmek için McNemar testi gerçekleştirilmiştir.

McNemar testi, hata matrislerinden hesaplanan standartlaştırılmış normal test istatistiğine dayanan parametrik olmayan bir testtir (Eşitlik 2).

$$Z = \frac{f_{12} - f_{21}}{\sqrt{f_{12} + f_{21}}} \quad (2)$$

Burada f_{12} , birinci hata matrisinde yanlış sınıflandırılan ancak ikinci hata matrisinde doğru şekilde sınıflandırılan örneklerin sayısını gösterir. f_{21} , ikinci hata matrisinde yanlış sınıflandırılan ancak birinci hata matrisinde doğru sınıflandırılan örneklerin sayısını gösterir. Z değeri 1,96'dan büyükse, hata matrisleri arasındaki doğruluk farkı istatistiksel olarak anlamlıdır ($p \leq 0.05$) (Foody, 2004; Leeuw ve ark. 2006).

3. Bulgular

Yİ ve TB veri setlerinden RO ve RA sınıflandırıcıları kullanılarak üretilen arazi örtüsü haritaları Şekil 3'te sunulmuştur. Ayrıca nokta bazlı doğruluk sonuçları da Tablo 3'te verilmiştir. Sonuçlar değerlendirildiğinde RO sınıflandırmasının RA yöntemine göre daha homojen dağılımlı haritalar ürettiği sonucu çıkarılmaktadır. Çıplak ve yarıdoğal alanların sınıflandırılmasında yaşanan zorluk tüm sınıflandırıcılar için geçerli olmuştur.

Boyut indirgeme yöntemleri değerlendirildiğinde ise TB veri setlerinden elde edilen sınıflandırma sonuçlarının Yİ veri setlerine kıyasla daha yüksek doğruluk sağladığı görülmüştür. Bu kısımda da çıplak ve yarıdoğal alanların sınıflandırılmasının zorluğu ve kentsel genişlemenin saptanmasının getirdiği sorunlar sonuçları etkilemiş ve her yöntem için bu doğruluk değerleri diğer sınıflara kıyasla daha düşük çıkmıştır. Yİ tabanlı yöntemde farklı arazi örtüsü sınıflarını ayırt etme verimliliğinin, TB tabanlı yöntemde göre daha az olduğu sonucunun ortaya konduğu bu çalışmada TB tabanlı sistemin verimli, hızlı ve karmaşıklığı az bir yöntem olması deneyimlenmiştir.

Sonuçlar genel olarak değerlendirildiğinde aynı eğitim veri seti ile uygulanan analizlerde her iki boyut indirgeme yaklaşımı ayrı ayrı ele alındığında tek bir veri seti için sınıflandırıcı başarımlarında ortaya çıkan doğruluk farkı oldukça düşük olup, doğruluğa etki eden temel faktörün boyut indirgeme yaklaşımı olduğu ortaya çıkmaktadır. TB veri seti üzerinde her iki sınıflandırıcı ile %83 civarı doğruluğa ulaşılırken, aynı örneklem seti ve aynı sınıflandırıcılar ile Yİ veri seti üzerinde yaklaşık %10'luk bir düşüş yaşanmıştır. Her iki sınıflandırıcı kapsamında kappa istatistiği karşılaştırması sonucunda TB veri seti için 0,80 civarında gözlemlenen değerlerde Yİ veri seti için yaklaşık 0,11 ile 0,16 arasında düşüş yaşanmıştır. Özellikle TB - RA veri seti - sınıflandırıcı ikilisi ile özellikle diğer kombinasyonlarda başarımın düşük olduğu yapay alanlar, çıplak ve yarı doğal alanlar be kentsel genişleme sınıfları için başarım daha yüksek olup, üretici doğruluğu ve kullanıcı doğruluğu arasındaki yüzdesel fark daha düşük çıkmıştır. Yİ veri seti için yapay alanlarda kullanıcı doğruluğu, çıplak ve yarı doğal alanlarda ise üretici doğruluğu oldukça düşük çıkmaktadır.

Sınıflandırma sonuçlarından elde edilen farkların istatistiksel anlamlılığının değerlendirilmesi amacı ile iki farklı sınıflandırıcı ve iki farklı veri seti için toplamda üretilmiş olan dört adet hata matrisi üzerinde tüm olası ikili kombinasyonlar için McNemar karşılaştırmalı istatistik testi gerçekleştirilmiştir. Burada veri setleri sabit seçilerek sınıflandırıcı kıyaslamaları (Yİ için RO - RA, TB için RO - RA) ve ek olarak sınıflandırıcı sabit seçilerek veri seti kıyaslamaları (RO için Yİ - TB ve RA için Yİ - TB) söz konusudur. İlgili karşılaştırma sonucunda elde edilen değerler Tablo 4'te verilmektedir. Bu tablo değerlendirildiğinde tüm ikili kıyaslamalar kritik Z değeri olan 1,96'dan büyük çıkmakla birlikte her iki sınıflandırıcı için veri setleri arasında hesaplanan Z değerleri her iki veri seti için sınıflandırıcılar arasında hesaplanan Z değerlerinden yüksek çıkmış olup, veri seti nedeniyle ortaya çıkan farkın anlamlılık seviyesinin daha yüksek olduğu söylenebilir. Bu sonuç sınıflandırma

doğruluğu ve kappa istatistik değerlerini destekler niteliktedir.

Sınıflandırma sonucu elde edilen alansal bilgilerin karşılaştırılması adına **Tablo 5**'te sınıflandırma sonucu oluşan alanlar verilmiştir. Bu kapsamda tanımlanan ilk beş sınıf, kendi nesne tanımlamaları kapsamında ilgili dönem aralığında değişim göstermeyen statik sınıflar olup, kentsel genişleme sınıfı ilgili dönemde yapay alan dışı bir sınıfa ait iken yapay alana dönüşen konumları temsil etmektedir. Alanlar incelendiğinde kentsel genişleme ve çıplak-yarıdoğal alanların her kombinasyon için diğer sınıflara nispeten daha farklı sonuçlar verdiği saptanmıştır. Bu iki sınıfın manuel olarak oluşturulan eğitim alanlarındaki hata payı, zamana göre değişen yansıtım farklılıkları, diğer sınıflarla benzerlik (yapay alanlar ve tarım alanları) bu sonucun ortaya çıkmasına yol açmıştır. Su kütlesi sınıfının alan bazında gösterdiği farklılık, sınıf tanımına homojenlik parametresi eklenmemesine bağlı olarak ortaya çıkmaktadır. Bu noktada yine TB tabanlı boyut indirgeme verisi ile elde edilen sınıflandırma sonuçları kentsel genişleme sınıfını daha doğru bir şekilde ortaya koyarak yüksek başarımlar göstermektedir.

4. Sonuçlar

Bu araştırmanın sonuçları son zamanlarda uzaktan algılama uygulamalarında popüler hale gelen makine öğrenmesi algoritmalarının boyut indirgeme metotlarıyla kombinasyonunun kullanılabilirliğini ortaya koymaktadır. Literatürde rastgele orman

sınıflandırması daha sık kullanılan algoritma olmakla beraber bu çalışma özelinde regresyon ağacı algoritmasının da benzer doğrulukta sonuç ürettiği görülmektedir. Çalışmanın sonuçlarına göre doğruluğu asıl etkileyen faktörün boyut indirgeme metodu ile ilişkili olduğu ve burada temel bileşen dönüşümü temelli yaklaşımın özellikle kentsel genişlemeyi daha doğru tespit ettiği ve sınıflandırma doğruluğunda yüksek başarımlar verdiği görülmüştür. Bu kapsamda temel bileşen dönüşümü ile elde edilen veri seti üzerinde her iki sınıflandırıcı da %83 sınıflandırma doğruluğu ve 0,8 üzerinde kappa değerlerine ulaşmıştır. Bu iki sınıflandırma sonucu için tüm sınıflarda %70 ve üzeri kullanıcı ve üretici doğruluğu elde edilebilmiştir.

Çalışma sonucunda Kocaeli ili arazi örtüsünde dört yıllık zaman süresince kentsel alanlarda genişleme olduğu görülmüştür. Özellikle ilin kuzeybatı ve güneyinde gözlemlenen genişleme dikkat çekicidir. Zaman içerisinde çıplak alanlarda bina yapımının olduğu ve Kuzey Marmara otoyolunun yapımına bağlı olarak yapay alanların arttığı da çalışma sonucuyla bağdaşmaktadır.

Çalışmanın yukardaki sonuçların yanı sıra, bu tip makine öğrenmesi algoritmaları ile farklı veri indirgeme yöntemlerin kombinasyonlarının kullanımında olumlu sonuçlar ortaya koyduğunu göstermiştir. Farklı uydulardan elde edilen görüntüler kullanılarak veya farklı sınıflandırıcılar değerlendirilerek çalışma kapsamı genişletilebilir, daha kapsamlı ve etkin uygulamalarla karar vericilere doğru ve güvenilir sonuçlar sunulabilir.

Tablo 3. Nokta tabanlı doğruluk analizi sonuçları

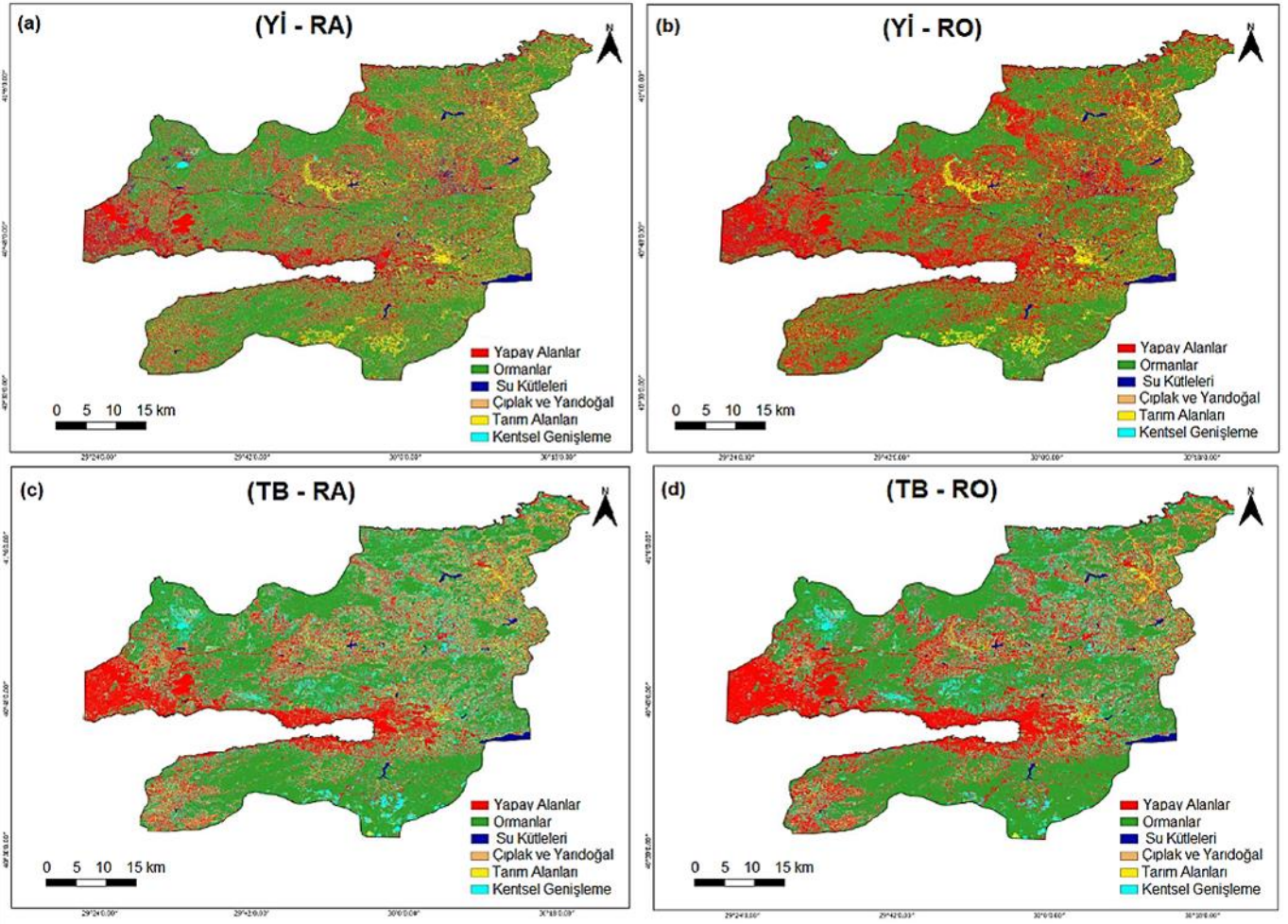
Veri/Yöntem	Yİ-RO	TB-RO	Yİ-RA	TB-RA
Sınıf	ÜD KD	ÜD KD	ÜD KD	ÜD KD
Yapay Alanlar	86,66 60,01	96,66 72,50	80,00 55,81	100 78,94
Ormanlar	99,05 76,92	96,65 82,85	86,60 81,25	96,60 82,85
Su Kütleleri	99,06 85,71	100 100	96,66 96,66	100 100
Çıplak ve Yarıdoğal	50,27 82,35	56,66 85,00	50,01 86,66	70,00 84,00
Tarım Alanları	76,66 74,19	80,00 92,30	73,33 61,10	70,00 87,50
Kentsel Genişleme	50,21 69,23	70,00 72,41	50,00 60,00	66,66 71,42
Kappa	0,69	0,80	0,65	0,81
Sınıflandırma Doğruluğu	73,74	83,33	71,27	83,88

Tablo 4. Hata matrisleri üzerinden hesaplanan McNemar istatistikleri

Veri ve Yöntem	Algoritma ve Veri Çiftleri	
	RO - RA	Yİ - TB
RO	\	10,24
RA	\	11,78
Yİ	3.64	\
TB	2.48	\

Tablo 5. Sınıflandırma sonucu oluşan alanlar (km²)

Veri/Yöntem	Yİ-RO	TB-RO	Yİ-RA	TB-RA
Sınıf	Alan	Alan	Alan	Alan
Yapay Alanlar	1136,65	1032,72	919,48	977,16
Ormanlar	1609,28	1457,26	1406,17	1392,01
Su Kütleleri	72,60	21,81	119,18	21,70
Çıplak ve Yarıdoğal	58,27	100,83	137,70	131,96
Tarım Alanları	318,02	228,76	419,62	302,98
Kentsel Genişleme	184,06	537,50	376,72	553,07



Şekil 3. Sınıflandırma sonucunda elde edilen tematik haritalar, a) Yİ-RA, b) Yİ-RO, c) TB-RA, d) TB-RO

Bilgilendirme/Teşekkür

Bu çalışma kapsamında kullanılan Sentinel 2 (ESA) uydu görüntüleri U.S. Geological Survey tarafından sağlanmıştır.

Araştırmacıların Katkı Oranı

Esmâ Efe: Literatür taraması, Analiz, Modelleme, Makale yazımı; **Uğur Algancı:** Doğruluk analizi, Düzenleme, Makale yazımı

Çatışma Beyanı

Herhangi bir çıkar çatışması bulunmamaktadır.

Kaynakça

- Ahady, A. B., & Kaplan, G. (2022). Classification comparison of Landsat-8 and Sentinel-2 data in Google Earth Engine, study case of the city of Kabul. *International Journal of Engineering and Geosciences*, 7(1), 24-31.
- Akar Ö & Güngör O (2012). Rastgele orman algoritması kullanılarak çok bantlı görüntülerin sınıflandırılması. *Jeodezi ve Jeoinformasyon Dergisi*, 1(2), 139-146.
- Akar, Ö., & Tunç Görmüş, E. (2019). Göktürk-2 ve Hyperion EO-1 Uydu Görüntülerinden Rastgele Orman Sınıflandırıcısı ve Destek Vektör Makineleri ile Arazi Kullanım Haritalarının Üretilmesi. *Geomatik*, 4(1), 68-81.

- Alganci, U. (2019). Dynamic Land Cover Mapping of Urbanized Cities with Landsat 8 Multi-temporal Images: Comparative Evaluation of Classification Algorithms and Dimension Reduction Methods. *ISPRS International Journal of Geo-Information*, 8(3), 139.
- Anderson, J. R., Hardy, E. E., Roach, J. T., & Witmer, R. E. (1976). *A Land Use and Land Cover Classification System for Use with Remote Sensor Data*; Professional Paper 964; US Geological Survey Government Printing Office: Washington, DC, USA.
- Apaydın, C., & Abdikan, S. (2021). Fındık bahçelerinin Sentinel-2 verileri kullanılarak piksel tabanlı sınıflandırma yöntemleriyle belirlenmesi. *Geomatik*, 6(2), 107-114.
- Bégué A, Arvor D, Bellon B, Betbeder J, De Abelleira D P D, Ferraz R, Lebourgeois V, Lelong C, Simões M R & Verón S (2018). Remote Sensing and Cropping Practices: A Review. *Remote Sensing*, 10(1), 99.
- Belgiu, M., & Dragut, L. (2016). Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing*, 114, 24-31
- Belward AS & Skoien JO (2015). Who launched what, when and why; trends in global land-cover observation capacity from civilian earth observation satellites. *ISPRS Journal of Photogrammetry and Remote Sensing*, 103, 115-128.
- Bishop, C. M. (2006). *Pattern recognition and machine learning*, Springer.

- Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and regression trees*. Routledge.
- Breiman, L. (2001). Random Forests. *Machine Learning*, 45, 5–32.
- Breiman, L., & Cutler, A. (2005). Random forest. http://www.stat.berkeley.edu/~breiman/RandomForests/cc_home.htm, erişim tarihi:16.01.2021.
- Congalton, R. G. (1991). A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment*, 37(1), 35-46.
- Congalton, R. G., & Green, K. (2019). *Assessing the accuracy of remotely sensed data: principles and practices*. CRC press.
- Deng, J. S., Wang, K., Deng, Y. H., & Qi, G. J. (2008). PCA-based land-use change detection and analysis using multitemporal and multisensor satellite data. *International Journal of Remote Sensing*, 29(16), 4823- 4838.
- Foody, G. M. (2004). Thematic Map Comparison. *Photogrammetric Engineering & Remote Sensing*, 70(5), 627-633.
- Friedl, A. F., McIver, D. K., Hodges, J. C. F., Zhang, X.Y., Muchoney, D., Strahler, A. H., Woodcock, C. E., Gopal, S., Schneider, A., Cooper, A., Baccini, A., Gao, F., & Schaaf, C. (2002). Global land cover mapping from MODIS: algorithms and early results. *Remote Sensing of Environment*, 83(1–2), 287–302.
- Harris, R. & Baumann, I. (2015). Open data policies and satellite Earth Observations. *Space Policy*, 32, 44-53.
- Jamali, A. (2019). A Fit-for-Purpose Algorithm for environmental monitoring based on maximum likelihood, support vector machine and random forest. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-3/W7, 25–32.
- Kavzoğlu, T., & Çölkesen, İ. (2010). Karar Ağaçları ile Uydu Görüntülerinin Sınıflandırılması: Kocaeli Örneği, *Harita Teknolojileri Elektronik Dergisi*, 2(1), 36-45.
- Langat, P. K., Kumar, L., & Koech, R. (2019). Monitoring river channel dynamics using remote sensing and GIS techniques. *Geomorphology*, 325, 92-102.
- Lawrence, R. L., & Wright, A. (2001). Rule-based classification systems using classification and regression tree (CART) analysis. *Photogrammetric Engineering and Remote Sensing*, 67(10), 1137–1142.
- Leeuw, J. D., Jia, H., Yang, L., Liu, X., Schmidt, K., & Skidmore, A. (2006). Comparing accuracy assessments to infer superiority of image classification methods. *International Journal of Remote Sensing*, 27(1), 223-232.
- Lu, D., & Weng, Q. (2007). A survey of image classification methods and techniques for improving classification performance. *International Journal of Remote Sensing*, 28(5), 823-870.
- Morsy, S., & Hadi, M. (2022). Impact of land use/land cover on land surface temperature and its relationship with spectral indices in Dakahlia Governorate, Egypt. *International Journal of Engineering and Geosciences*, 7(3), 272-282.
- Petropoulos, G. P., Vadrevu, K. P., & Kalaitzidis, C. (2013). Spectral angle mapper and object-based classification combined with hyperspectral remote sensing imagery for obtaining land use/cover mapping in a Mediterranean region. *Geocarto international*, 28(2), 114-129.
- Praticò, S., Solano, F., Di Fazio, S., & Modica, G. (2021). Machine Learning Classification of Mediterranean Forest Habitats in Google Earth Engine Based on Seasonal Sentinel-2 Time- Series and Input Image Composition Optimisation. *Remote Sensing*, 13(4), 586.
- 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, 100351.
- Rawat, J. S., & Kumar, M. (2015). Monitoring land use/cover change using remote sensing and GIS techniques: A case study of Hawalbagh block, district Almora, Uttarakhand, India, *The Egyptian Journal of Remote Sensing and Space Science*, 18(1), 77-84.
- Ruiz-Luna, A., & Berlanga-Robles, C. A. (2003). Land use, land cover changes and coastal lagoon surface reduction associated with urban growth in northwest Mexico. *Landscape Ecology*, 18, 159–171.
- Shao, Y., & Lunetta, R. S. (2012). Comparison of support vector machine, neural network, and CART algorithms for the land-cover classification using limited training data points. *ISPRS Journal of Photogrammetry and Remote Sensing*, 70, 78-87.
- Shetty, S., Gupta, P. K., Belgium, M., & Srivastava, S. K. (2021). Assessing the Effect of Training Sampling Design on the Performance of Machine Learning Classifiers for Land Cover Mapping Using Multi-Temporal Remote Sensing Data and Google Earth Engine. *Remote Sensing*, 13(8), 1433.
- Story, M., & Congalton, R. G. (1986). Accuracy assessment: a user's perspective. *Photogrammetric Engineering and Remote Sensing*, 52(3), 397-399.
- URL -1: www.turkstat.gov.tr, erişim tarihi: 15.01.2021
- URL -2: <https://kosano.org.tr/rakamlarla-kocaeli/>, erişim tarihi 27.04.2021
- Yu, L., Liang, L., Wang, J., Zhao, Y., Cheng, Q., Hu, L., Liu, S., Yu, L., Wang, X., Zhu, P., Li, X., Xu, Y., Li, C., Fu, W., Li, X., Li, W., Liu, C., Cong, N., Zhang, H. ... Gong, P. (2014). Meta-Discoveries from a Synthesis of Satellite-Based Land Cover Mapping Research. *International Journal of Remote Sensing* 35 (13), 4573–4588.
- Waldrop, M. M. (2016). The chips are down for Moore's law. *Nature*, 530(7589), 144-147.
- Wulder, M. A., Coops, N. C., Roy, D. P., White, J. C., & Hermosilla, T. (2018). Land cover 2.0. *International Journal of Remote Sensing*, 39(12), 4254-4284.

