

KARADENİZ TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

HARİTA MÜHENDİSLİĞİ ANABİLİM DALI

UYDU GÖRÜNTÜLERİNDEN YERLEŞİM ALANI DEĞİŞİMİ TESPİTİNDE
YENİDEN KULLANILABİLİR KOD ÜRETİMİNDE BOYUTLARIN
BELİRLENMESİ

YÜKSEK LİSANS TEZİ

Miraç Bükre ATMACA

EKİM 2025
TRABZON



KARADENİZ TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

HARİTA MÜHENDİSLİĞİ ANABİLİM DALI

**UYDU GÖRÜNTÜLERİNDEN YERLEŞİM ALANI DEĞİŞİMİ TESPİTİNDE
YENİDEN KULLANILABİLİR KOD ÜRETİMİNDE BOYUTLARIN
BELİRLENMESİ**

Miraç Bükre ATMACA

Karadeniz Teknik Üniversitesi Fen Bilimleri Enstitüsünce

"HARİTA YÜKSEK MÜHENDİSİ"

Unvanı Verilmesi İçin Kabul Edilen Tezdir.

Tezin Enstitüye Verildiği Tarih : 18 / 09 / 2025

Tezin Savunma Tarihi : 27 / 10/ 2025

Tez Danışmanı : Prof. Dr. Çetin CÖMERT

Trabzon 2025

**KARADENİZ TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

**Harita Mühendisliği Anabilim Dalında
Miraç Bükre ATMACA Tarafından Hazırlanan**

**UYDU GÖRÜNTÜLERİNDEN YERLEŞİM ALANI DEĞİŞİMİ TESPİTİNDE YENİDEN
KULLANILABİLİR KOD ÜRETİMİNDE BOYUTLARIN BELİRLENMESİ**

**başlıklı bu çalışma, Enstitü Yönetim Kurulunun 25 / 09 / 2025 gün ve 2121 sayılı
kararıyla oluşturulan jüri tarafından yapılan sınavda
YÜKSEK LİSANS TEZİ
olarak kabul edilmiştir.**

Jüri Üyeleri

Başkan : Prof. Dr. Çetin CÖMERT

Üye : Doç.Dr. Volkan YILMAZ

Üye : Prof.Dr. Sedat DOĞAN

Prof. Dr. Asim KADIOĞLU

Enstitü Müdürü

ÖNSÖZ

“Uydu Görüntülerinden Yerleşim Alanı Değişimi Tespitinde Yeniden Kullanılabilir Kod Üretiminde Boyutların Belirlenmesi” isimli bu tez Karadeniz Teknik Üniversitesi Fen Bilimleri Enstitüsü Harita Mühendisliği Anabilim Dalı, Yüksek Lisans Programı’nda hazırlanmıştır.

Tez çalışmam boyunca desteğini esirgemeyen, bilgileri ve tecrübeleriyle yolumu aydınlatan, araştırma sürecinde bana yalnızca doğruyu göstermekle kalmayıp farklı bakış açıları kazandırarak geniş düşünmeyi öğreten danışmanım Sayın Prof. Dr. Çetin CÖMERT’e teşekkürü bir borç bilirim. Ayrıca, bu süreçte yardımlarından dolayı Sayın Dr. Öğr. Üyesi Gülten KARA’ya ve eğitim öğretim hayatımda katkısı olan tüm hocalarıma teşekkür ederim.

Tez çalışması sürecinde bana desteklerini esirgemeyen TÜRKSAT Uydu Ar-Ge ve Teknoloji Müdürlüğü’ndeki tüm değerli çalışma arkadaşlarıma teşekkürlerimi sunarım.

Son olarak, her zaman yanımda olan, kararlarımda bana rehberlik eden ve maddi-manevi desteklerini daima hissettiren anneme, babama ve tüm aileme teşekkür eder, minnettarlığımı ifade etmek isterim.

Miraç Bükre ATMACA

Trabzon 2025

TEZ ETİK BEYANNAMESİ

Yüksek Lisans Tezi olarak sunduğum “Uydu Görüntülerinden Yerleşim Alanı Değişimi Tespitinde Yeniden Kullanılabilir Kod Üretiminde Boyutların Belirlenmesi” başlıklı bu çalışmayı baştan sona kadar danışmanım Prof. Dr. Çetin CÖMERT’in sorumluluğunda tamamladığımı, verileri/örnekleri kendim topladığımı, deneyleri/analizleri ilgili laboratuvarlarda yaptığımı/yaptırdığımı, başka kaynaklardan aldığım bilgileri metinde ve kaynakçada eksiksiz olarak gösterdiğimi, çalışma sürecinde bilimsel araştırma ve etik kurallara uygun olarak davrandığımı ve aksinin ortaya çıkması durumunda her türlü yasal sonucu kabul ettiğimi beyan ederim. 27/10/2025

Miraç Bükre ATMACA

İÇİNDEKİLER

	<u>Sayfa No</u>
ÖNSÖZ.....	III
TEZ ETİK BEYANNAMESİ.....	IV
İÇİNDEKİLER.....	V
ÖZET	VII
SUMMARY.....	VIII
ŞEKİLLER DİZİNİ	IX
TABLolar DİZİNİ.....	X
SEMBOLLER VE KISALTMALAR DİZİNİ	XI
1. GENEL BİLGİLER	1
1.1. Giriş.....	1
1.2. Problem Tanımı.....	2
1.3. Çalışmanın Amacı	3
1.4. Temel Kavramlar.....	4
1.4.1. Makine Öğrenmesi	4
1.4.2. Denetimli Öğrenme.....	5
1.4.3. Karar Ağaçları.....	6
1.4.4. Random Forest	7
1.4.5. Karar Ağaçları ve Random Forest Farkı	8
2. YAPILAN ÇALIŞMALAR.....	10
2.1. Materyaller	10
2.1.1. Veri Kaynakları.....	10
2.1.2. Kullanılan Geliştirme Ortamları ve Araçları.....	11
2.2. Yöntem	12
2.2.1. İş Akışı	14
2.2.2. Çalışma Alanı.....	15
2.2.3. Veri Temini	17
2.2.4. Spektral İndekslerin Hesaplanması	19
2.2.5. Örneklem, Eğitim ve Doğrulama	22
2.2.6. Random Forest Modeli ve Otomatik Ağaç Sayısı Seçimi	23
2.2.7. Olasılık Tabanlı Sınıflandırma ve Otomatik Threshold Belirleme	24

2.2.8. Bina Maskesi Üretimi, Yama Temizleme ve Değişim Analizi.....	25
2.2.9. Değişken Seti ve Senaryo Bazlı Model Boyutu Analizi	26
3. BULGULAR VE TARTIŞMA	28
3.1. Model Performansı ve Değerlendirme Sonuçları	28
3.2. Alansal Değişim Analizi	29
3.2.1. Ömerli Barajı Havzası Değişim Sonuçları	30
3.2.2. Çubuk-1 Barajı Havzası Değişim Sonuçları	31
3.3. Otomatik Ağaç Sayısı Seçimi (OOB Analizi).....	32
3.4. Otomatik Threshold (Eşik) Seçiminin Sonuçlara Etkisi	33
3.5. Spektral Bant ve İndeks Senaryolarının Etkisi.....	35
3.6. Mekânsal Sızıntının Engellenmesi: Grid Tabanlı Eğitim-Test Ayrımının Etkisi.....	36
3.7. Bulguların Literatür ile Karşılaştırılması	37
3.8. Sınırlılıklar ve Belirsizlik Analizi	39
4. SONUÇLAR.....	41
5. KAYNAKLAR	44
ÖZGEÇMİŞ	

Yüksek Lisans Tezi

ÖZET

UYDU GÖRÜNTÜLERİNDEN YERLEŞİM ALANI DEĞİŞİMİ TESPİTİNDE YENİDEN KULLANILABİLİR KOD ÜRETİMİNDE BOYUTLARIN BELİRLENMESİ

Miraç Bükre ATMACA

Karadeniz Teknik Üniversitesi
Fen Bilimleri Enstitüsü
Harita Mühendisliği Anabilim Dalı
Danışman: Prof. Dr. Çetin CÖMERT
2025, 49 Sayfa

Bu çalışma, Sentinel-2 uydu görüntülerinden yerleşim alanı değişimlerinin yeniden kullanılabilir bir kod ile belirlenmesini amaçlamaktadır. Bu kapsamda veri ön işleme, mevsimsel bileşik üretimi, değişken seti optimizasyonu, mekansal sızıntıyı azaltan grid tabanlı eğitim-test ayrımı, Random Forest parametrelerinin otomatik seçimi ve veri temelli eşik belirleme adımlarını içeren bütünlük bir analiz altyapısı geliştirilmiştir.

Model performansı; spektral bantlar, türetilmiş indeksler ve doku özelliklerinden oluşturulan farklı değişken setleri üzerinden test edilmiş ve değişkenlerin katkısının bölgenin spektral özelliklerine bağlı olduğu görülmüştür. Bu sonuçlar, model boyutunun her bölge için veri odaklı biçimde optimize edilmesi gerektiğini göstermiştir.

Çalışmada rastgele örnekleme yerine mekânsal sızıntıyı kontrol eden grid tabanlı doğrulama kullanılmış ve bu yaklaşımın daha gerçekçi doğruluk değerleri ürettiği görülmüştür. Ayrıca Random Forest'ın ağaç sayısı ve karar eşiğinin otomatik seçimi, hem sınıflandırma performansını hem de model kararlılığını artırmıştır.

Oluşturulan bina değişim haritaları, 2018–2024 döneminde Ömerli Havzası'nda yaklaşık 3.3 km², Çubuk-1 Havzası'nda ise 0.33 km² net yerleşim artışı olduğunu göstermektedir. Bu sonuçlar, yöntemin havza yönetimi, arazi kullanım izlemesi ve kentsel büyüme analizleri için uygulanabilirliğini ortaya koymaktadır.

Genel olarak çalışma, farklı bölgelere minimum uyarlama ile uygulanabilen, parametreleri otomatik optimize eden ve mekânsal olarak güvenilir sonuçlar üreten yeniden kullanılabilir bir analiz iş akışı sunmaktadır.

Anahtar Kelimeler: Sentinel-2, Random Forest, Mekânsal Sızıntı, Bina Değişim Tespiti, Spektral İndeksler, Uzaktan Algılama, Mekânsal Planlama

Master Thesis

SUMMARY

DETERMINING DESIGN DIMENSIONS FOR REUSABLE CODE IN BUILT-UP AREA CHANGE DETECTION FROM SATELLITE IMAGERY

Miraç Bükre ATMACA

Karadeniz Teknik Üniversitesi
Fen Bilimleri Enstitüsü
Harita Mühendisliği Anabilim Dalı
Danışman: Prof. Dr. Çetin CÖMERT
2025, 49 Page

This study aims to detect settlement area changes from Sentinel-2 satellite imagery using a reusable and adaptable code framework. To this end, an integrated analysis workflow was developed, comprising data preprocessing, seasonal composite generation, variable-set optimization, grid-based training–testing separation to reduce spatial leakage, automatic selection of Random Forest parameters, and data-driven threshold determination.

The performance of the model was evaluated using different variable sets constructed from spectral bands, derived indices, and texture features. The results showed that the contribution of these variables depends on the spectral characteristics of the region, indicating that model dimensionality must be optimized in a data-driven manner for each study area.

Instead of random sampling, the study employed a grid-based validation strategy to control spatial leakage, which produced more realistic accuracy estimates. Additionally, the automatic selection of the number of trees and decision thresholds in the Random Forest classifier improved both model stability and classification performance.

The generated building-change maps indicate a net settlement increase of approximately 3.3 km² in the Ömerli Basin and 0.33 km² in the Çubuk-1 Basin between 2018 and 2024. These findings demonstrate the applicability of the proposed method for watershed management, land-use monitoring, and urban growth analyses.

Overall, the study presents a reusable analysis workflow that can be applied to different regions with minimal adjustments, automatically optimizes model parameters, and produces spatially reliable results.

Key Words: Sentinel-2, Random Forest, Spatial Leakage, Building Change Detection, Spectral Indices, Remote Sensing, Spatial Planning

ŞEKİLLER DİZİNİ

	<u>Sayfa No</u>
Şekil 1. Karar Ağacı Algoritması.....	7
Şekil 2. Random Forest Algoritmasının Şematik Gösterimi	8
Şekil 3. İş Akış Şeması	15
Şekil 4. Ömerli Barajı Çalışma Alanı	16
Şekil 5. Çubuk-1 Barajı Çalışma Alanı.....	16
Şekil 6. 2018 Yılı Ömerli Barajı Sentinel-2 Mevsimsel Bileşik Görüntüsü.....	18
Şekil 7. 2018 Yılı Ömerli Barajı Sentinel-2 Mevsimsel Bileşik Görüntüsü.....	18
Şekil 8. 2018 Yılı Çubuk-1 Barajı Sentinel-2 Mevsimsel Bileşik Görüntüsü	19
Şekil 9. 2024 Yılı Çubuk-1 Barajı Sentinel-2 Mevsimsel Bileşik Görüntüsü	19
Şekil 10. 2018 Yılı Ömerli Barajı Yerleşim Alanı (Built-up Class) Dağılımı.....	30
Şekil 11. 2024 Yılı Ömerli Barajı Yerleşim Alanı (Built-up Class) Dağılımı.....	31
Şekil 12. 2018 Yılı Çubuk-1 Barajı Yerleşim Alanı (Built-up Class) Dağılımı.....	31
Şekil 13. 2018 Yılı Çubuk-1 Barajı Yerleşim Alanı (Built-up Class) Dağılımı.....	32

TABLÖLAR DİZİNİ

	<u>Sayfa No</u>
Tablo 1. Çalışma Alanlarına Ait Koordinatlar ve Alan Büyüklükleri	17
Tablo 2. Ömerli Barajı RF Model Performans	28
Tablo 3. Çubuk-1 Barajı RF Model Performansı.....	29



SEMBOLLER VE KISALTMALAR DİZİNİ

AOI	: Çalışma Alanı
AUC	: Eğri Altındaki Alan
DW	: Dynamic World
FN	: Yanlış Negatif
FP	: Yanlış Pozitif
FPR	: Yanlış Pozitif Oranı
GEE	: Google Earth Engine
GLCM	: Gri Seviye Eş Oluşum Matrisi
KM ²	: Kilometrekare
NDBI	: Normalize Fark Yerleşim İndeksi
NDVI	: Normalize Fark Bitki İndeksi
NDWI	: Normalize Fark Su İndeksi
QGIS	: Quantum Geographic Information System
RF	: Random Forest
ROC	: Receiver Operating Characteristic
SWIR	: Kısa Dalga Kızılötesi
TIRS	: Termal Kızılötesi Sensör
TN	: Doğru Negatif
TP	: Doğru Pozitif
TPR	: Doğru Pozitif Oranı
USGS	: ABD Jeoloji Araştırmaları Kurumu
UI	: Kentsel İndeks
VS Code	: Visual Studio Code

1. GENEL BİLGİLER

1.1. Giriş

Günümüzde hızlı nüfus artışı, kentlere yönelen göç hareketleri ve ekonomik faaliyetlerin yoğunlaşması, özellikle büyükşehirlerde yerleşim alanlarının hızla genişlemesine neden olmaktadır. Bu dinamik süreç, düzensiz yapılaşma, doğal alanların daralması, tarım ve orman arazilerinin baskı altına girmesi gibi önemli çevresel ve mekânsal sorunları beraberinde getirmektedir. Özellikle su kaynakları ve içme suyu havzaları çevresinde gerçekleşen kontrolsüz şehirleşme, ekosistemler üzerinde telafisi güç etkiler oluşturmakta ve havza yönetimini daha kritik hale getirmektedir. Bu nedenle, yerleşim alanlarındaki mekânsal değişimlerin uzaktan algılama teknikleriyle düzenli bir biçimde izlenmesi, hem doğal kaynakların korunması hem de sürdürülebilir çevresel planlamaya yönelik kritik bilgiler sağlamaktadır.

Artan nüfus ve barınma ihtiyacı, yerleşim alanlarının genişlemesini kaçınılmaz hale getirmiş olsa da, bu genişlemenin planlı, dengeli ve çevresel eşiklere uygun şekilde yönetilmesi büyük önem taşımaktadır. Çarpık kentleşmenin önlenmesi, doğal alanların korunabilmesi ve altyapı hizmetlerinin etkin olarak planlanabilmesi için, yapılaşma eğilimlerinin düzenli, doğru ve analiz edilebilir biçimde izlenmesi gerekmektedir. Bu doğrultuda, yerleşim alanı değişimlerinin bilimsel yöntemlerle ortaya konulması, sağlıklı kentleşme politikalarının oluşturulmasına önemli katkılar sunmaktadır.

Bu bağlamda uzaktan algılama ve makine öğrenmesi yöntemleri, kentsel gelişim süreçlerinin tespiti ve analizi için güçlü ve etkin araçlar sağlamaktadır. Yüksek mekânsal ve zamansal çözünürlüğe sahip uydu görüntüleri, yerleşim alanlarının zaman içerisindeki değişimini hem nicel hem de nitel açıdan değerlendirme olanağı sunarken; makine öğrenmesi algoritmaları, çok boyutlu ve karmaşık verilerden anlamlı çıktılar üretme yeteneği sayesinde şehirleşme eğilimlerinin daha doğru, nesnel ve ölçeklenebilir biçimde ortaya konmasına imkân tanımaktadır.

Son yıllarda, bu tür analizlerin yalnızca belirli bir bölgeye veya projeye özgü olarak değil, yeniden kullanılabilir, genellenebilir ve farklı çalışma alanlarına kolayca uyarlanabilir biçimde tasarlanması giderek daha fazla önem kazanmaktadır. Özellikle Google Earth Engine gibi bulut tabanlı büyük veri platformlarının sağladığı olanaklar, uydu görüntülerine

dayalı makine öğrenmesi süreçlerinin modüler, tekrar edilebilir ve kod temelli bir yapıda geliştirilmesine imkân vermektedir. Böylece farklı şehirlerde, farklı yıllarda veya farklı mekânsal ölçeklerde benzer analizlerin tekrarlanması mümkün hale gelmekte; araştırmaların hem şeffaflığı hem de bilimsel geçerliliği artırılmaktadır.

Bu çalışma, sadece yerleşim alanı değişimini incelemeyi değil, aynı zamanda makine öğrenmesi tabanlı bina tespiti modellerinin yeniden kullanılabilir bir kod yapısıyla nasıl tasarlanabileceğini ortaya koymayı amaçlamaktadır. Bu kapsamda geliştirilen Python–Google Earth Engine tabanlı yaklaşım; mevsimsel bileşik üretimi, dinamik etiketleme, mekânsal kaçak (spatial leakage) engelleme, otomatik parametre seçimi ve farklı bölgelerde uygulanabilir model optimizasyonu gibi bileşenleri içermekte olup, benzer çalışmalar için tekrarlanabilir bir şablon niteliği taşımaktadır.

1.2. Problem Tanımı

Günümüzde hızlı nüfus artışı ile birlikte iç ve dış göç hareketleri, özellikle büyük şehirlerde barınma ihtiyacını artırmakta ve buna bağlı olarak yerleşim alanlarının hızla genişlemesine neden olmaktadır. Kontrolsüz biçimde gerçekleşen bu büyüme; çarpık kentleşme, doğal ve tarımsal alanların kaybı, çevresel bozulma ve altyapı yetersizlikleri gibi önemli sorunları beraberinde getirmektedir. Yerleşim alanlarının düzenli ve sürdürülebilir şekilde gelişiminin sağlanabilmesi için söz konusu büyümenin mekânsal ve zamansal olarak izlenmesi kritik bir gereklilik haline gelmiştir. Özellikle yapılaşma sınırları ve korunan bölgeler içerisinde, mevcut yerleşim alanlarının genişleme eğilimlerinin sürekli olarak takip edilmesi, planlı şehirleşmenin sağlanması, doğal kaynakların korunması ve kentsel altyapının etkin biçimde yönetilebilmesi açısından büyük önem taşımaktadır.

Bu bağlamda çalışma alanlarının belirlenmesinde, 2017 yılında yürürlüğe giren “İçme-Kullanma Suyu Havzalarının Korunmasına Dair Yönetmelik” (Resmî Gazete, 2017) temel alınmıştır. Bu yönetmelik, daha önce Su Kirliliği Kontrolü Yönetmeliği (1988) kapsamında tanımlanan mutlak ve kısa mesafeli koruma alanı hükümlerini yürürlükten kaldırmış; içme suyu havzalarında yapılaşma koşullarını yeniden tanımlayarak birçok yerleşim alanı üzerinde doğrudan etkide bulunmuştur. Bu nedenle araştırmada örnek alan olarak İstanbul Ömerli Barajı ve Ankara Çubuk-1 Barajı çevreleri seçilmiş; her iki bölgenin hem içme suyu teminindeki kritik rolü hem de 2017 yılı yönetmelik değişikliklerinden etkilenmiş olmaları dikkate alınmıştır.

Bu alanlarda yerleşim baskısının zaman içinde ne ölçüde değiştiğinin ortaya konulması, yalnızca yönetmelik sonrası gelişmelerin değerlendirilmesi açısından değil, aynı zamanda havza koruma politikalarının sürdürülebilirliğinin analiz edilmesi açısından da büyük önem taşımaktadır. Bununla birlikte, uydu görüntülerinin çeşitliliği, mekânsal ölçek farklılıkları, bulutluluk problemleri ve çalışma alanlarının coğrafi özellikleri nedeniyle yerleşim alanı değişiminin güvenilir bir şekilde modellenmesi teknik açıdan çeşitli zorluklar içermektedir.

Bu noktada bir diğer önemli problem, literatürde yer alan birçok çalışmanın belirli bir bölgeye özgü, tekrar kullanılması güç, statik ve karmaşık kod yapılarıyla gerçekleştirilmesidir. Bu durum, analizlerin farklı bölgelerde yeniden uygulanmasını zorlaştırmakta; parametrik esnekliği düşük, sürdürülebilirliği sınırlı modeller ortaya çıkarmaktadır. Bu çalışma ise, yalnızca yerleşim alanı değişimini tespit etmeyi değil aynı zamanda farklı havzalar, farklı yıllar ve farklı coğrafi koşullar için kolayca uyarlanabilir, modüler ve yeniden kullanılabilir bir makine öğrenmesi tabanlı analiz yapısının geliştirilmesini problem olarak ele almaktadır. Google Earth Engine ve Python tabanlı bu yaklaşım; mevsimsel bileşik üretimi, otomatik etiketleme, mekânsal kaçak kontrolü, veri dengesi, otomatik ağaç sayısı optimizasyonu ve eşige dayalı bina değişimi tespiti gibi süreçleri kapsayarak, genişletilebilir ve tekrarlanabilir bir analiz altyapısı sunmayı amaçlamaktadır.

1.3. Çalışmanın Amacı

Bu çalışmanın temel amacı, uydu görüntülerinden yerleşim alanı tespitini mümkün kılan makine öğrenmesi tabanlı yaklaşımların Yeniden kullanılabilir, modüler ve farklı çalışma alanlarına kolaylıkla uyarlanabilir bir kod yapısı ile geliştirilmesidir. Literatürde sıklıkla karşılaşılan bölgeye özgü, tekrarlanabilirliği sınırlı ve parametrik esnekliği düşük algoritmaların aksine, bu çalışmada hedef, farklı yıllar, farklı coğrafi koşullar ve farklı havza alanları üzerinde aynı metodolojik yapının minimum müdahale ile çalıştırılabilmesini sağlayan bir analiz altyapısı oluşturmaktır.

Bu doğrultuda, örnek uygulama alanları olarak İstanbul'un önemli içme suyu kaynaklarından biri olan Ömerli Barajı ile Ankara'nın içme suyu temininde kritik role sahip Çubuk-1 Barajı çevreleri seçilmiştir. Bu bölgeler hem hızlı kentleşme baskısına maruz

kalmaları hem de 2017 yılında yenilenen havza koruma yönetmeliklerinden doğrudan etkilenmeleri nedeniyle çalışmanın değerlendirme alanları olarak uygun görülmüştür.

Çalışmada, Sentinel-2 uydu görüntüleri kullanılarak mevsimsel bileşikler üretilmiş, Dynamic World veri kümesinden otomatik etiketleme sağlanmış, mekânsal kaçak (spatial leakage) kontrol edilerek eğitim–test ayrımı gerçekleştirilmiş ve Random Forest algoritması Python ortamında uygulanmıştır. Oluşturulan bu modüler yapı; otomatik ağaç sayısı seçimi, otomatik eşik belirleme, grid tabanlı örnekleme, bulutluluk toleransı, bant seçimi ve ölçek parametrelerinin yeniden ayarlanabilir olması sayesinde geniş bir kullanım alanı sunmaktadır.

Bu kapsamda çalışma, yalnızca Ömerli ve Çubuk-1 baraj havzalarındaki yerleşim alanı değişimlerinin belirlenmesini değil; aynı zamanda bu tür analizlerin farklı bölgelerde tekrar edilebilmesini sağlayan, esnek, parametrik ve yeniden kullanılabilir bir kod altyapısının geliştirilmesini amaçlamaktadır. Geliştirilen yaklaşımın, sürdürülebilir yerleşim politikalarının desteklenmesi, havza koruma çalışmalarının güçlendirilmesi ve uzaktan algılama tabanlı kentleşme analizlerinde standartlaştırılabilir bir model yapısının oluşturulması açısından önemli katkılar sunması hedeflenmektedir.

1.4. Temel Kavramlar

1.4.1. Makine Öğrenmesi

Makine öğrenmesi, bilgisayar sistemlerinin açıkça programlanmaksızın veriden öğrenerek belirli bir görevi yerine getirme yeteneği kazanmasını konu alan bir yapay zekâ alt alanıdır. Mitchell (1997), makine öğrenmesini “bir bilgisayar programının, belirli bir görevdeki performansının, deneyim ile birlikte iyileşmesi” olarak tanımlayarak, alanın merkezinde yer alan deneyimden öğrenme kavramını vurgulamaktadır.

Jordan ve Mitchell (2015), makine öğrenmesini büyük veri kümeleri üzerinden kalıpların çıkarılması ve bu kalıplardan hareketle tahmin veya karar üretimine yönelik algoritmalar bütünü olarak ele almıştır. Günümüzde makine öğrenmesi yalnızca kestirimsel modeller oluşturmak için değil, aynı zamanda karmaşık sistemleri çözümlemek, örüntüleri keşfetmek ve çok boyutlu verilerden anlam çıkarmak için de kullanılmaktadır. Bunun en önemli avantajlarından biri, verideki doğrusal olmayan ilişkileri modelleyebilme ve farklı veri tiplerini (sayısal, kategorik, mekânsal vb.) bir arada işleyebilme kapasitesidir.

Bishop (2006) ise makine öğrenmesini, istatistiksel çıkarım ile hesaplamalı modellerin birleştiği bir alan olarak tanımlamakta; öğrenmenin temel amacının, gözlemlenen veriden olasılıksal bir model kurarak genelleme yapabilmek olduğunu belirtmektedir. Bu çerçevede makine öğrenmesi, hem kuramsal hem de uygulamalı açıdan güçlü bir metodolojik temel sunmaktadır.

Uzaktan algılama alanında makine öğrenmesi yöntemleri, özellikle sınıflandırma, nesne tanıma, arazi örtüsü tespiti ve değişim analizi gibi süreçlerde yaygın olarak kullanılmaktadır. Belgiu ve Drăguț (2016), makine öğrenmesi teknikleri içinde Random Forest algoritmasının yüksek doğruluk oranı, gürültüye dayanıklılığı ve karmaşık veriler için uygunluğu sayesinde uzaktan algılama çalışmalarında öne çıktığını belirtmektedir. Bu nedenle makine öğrenmesi, uydu tabanlı yerleşim alanı tespitinde güçlü ve esnek bir yaklaşım sunmaktadır.

1.4.2. Denetimli Öğrenme

Denetimli öğrenme (supervised learning), makine öğrenmesinin en yaygın kullanılan alt alanlarından biridir ve temel olarak giriş verileri ile bu verilere karşılık gelen doğru sınıfların (etiketlerin) birlikte kullanıldığı bir öğrenme sürecini ifade eder. Bu yaklaşımda amaç, etiketli örneklerden yola çıkarak giriş değişkenleri ile çıktı değişkeni arasındaki ilişkiyi öğrenen bir model geliştirmek ve bu modeli yeni verilere genelleştirebilmektir (Mohri, Rostamizadeh & Talwalkar, 2018).

Goodfellow, Bengio ve Courville (2016), denetimli öğrenmeyi “girdi-çıktı çiftlerinden oluşan bir veri kümesi üzerinden, bu ilişkileri yakalayacak bir fonksiyonun öğrenilmesi” olarak tanımlamakta ve bu sürecin istatistiksel tahmin teorisinin temel ilkelerine dayandığını belirtmektedir. Bu bağlamda öğrenme algoritmasının amacı, eğitim verisindeki örüntüleri en iyi şekilde yakalayıp, daha önce görülmemiş veriler üzerinde yüksek doğrulukla tahmin yapabilmektir. Bunun için modelin hem eğitim hatasını düşük tutması hem de aşırı öğrenmeden kaçınarak genelleme performansını koruması gerekmektedir.

Shalev-Shwartz ve Ben-David (2014), denetimli öğrenmenin başarısının üç temel unsura bağlı olduğunu vurgular: (i) yeterli sayıda ve dağılımı temsil eden etiketli veri, (ii) uygun bir hipotez sınıfı (model), ve (iii) aşırı öğrenmeyi önleyecek bir düzenleme (regularization) stratejisi. Bu unsurlar arasındaki dengenin sağlanması, özellikle karmaşık,

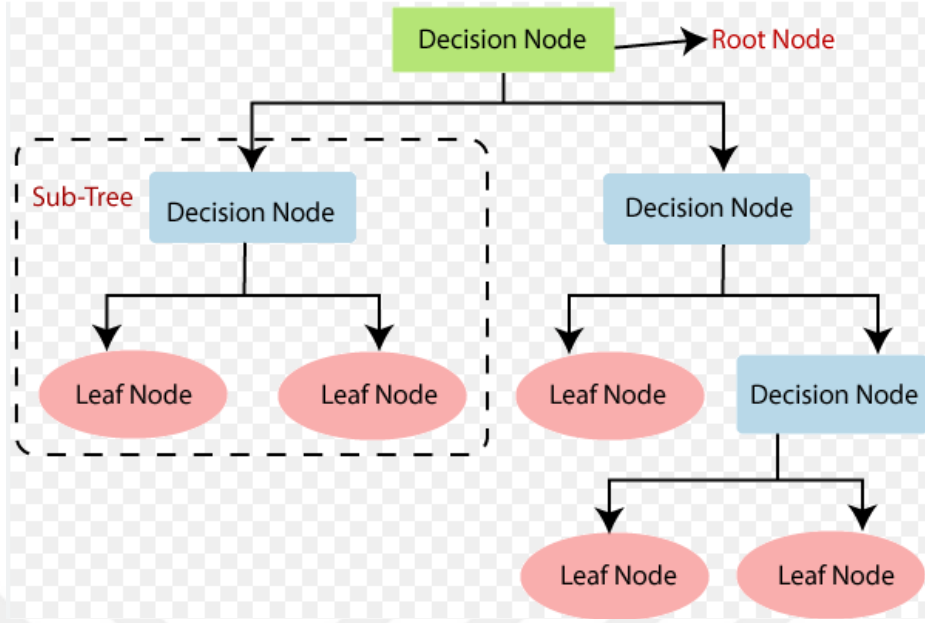
çok boyutlu ve gürültülü veriler içeren uzaktan algılama uygulamalarında kritik öneme sahiptir.

Uzaktan algılama literatüründe denetimli öğrenme, özellikle arazi örtüsü sınıflandırması, bina tespiti, nesne tanıma ve değişim izleme gibi uygulamalarda yaygın olarak kullanılmaktadır. Piksel düzeyinde sınıflandırmalarda, spektral bantlar ve türetilmiş indeksler giriş özelliklerini (features), referans veriler ise etiketleri oluşturur. Böylece algoritma, tanımlanan örneklerden öğrenerek geniş alanlara yayılmış uydu görüntülerini otomatik olarak sınıflandırabilir (Lary et al., 2016). Bu yönüyle denetimli öğrenme, büyük ölçekli çevresel izleme çalışmalarında yüksek doğruluk ve esneklik sunan bir yaklaşım olarak öne çıkmaktadır.

1.4.3. Karar Ağaçları

Karar ağaçları, sınıflandırma ve regresyon problemlerinde yaygın olarak kullanılan, basit ama güçlü makine öğrenmesi yöntemlerinden biridir. İlk olarak Morgan & Sonquist (1963) tarafından önerilen bu yöntem, veriyi özniteliklere dayalı olarak dallara ayırarak, kökten yapraklara doğru ilerleyen bir karar yapısı oluşturur. Breiman vd., (1984) karar ağaçlarını sistematik bir çerçevede ele alarak “Classification and Regression Trees (CART)” yaklaşımını geliştirmiştir.

Karar ağaçlarının en önemli avantajları, kolay yorumlanabilir olmaları ve görselleştirilebilmeleridir. Bu yöntem, veriyi öznitelik değerlerine göre yinelemeli olarak böler ve her bir düğümde en iyi ayrımı sağlayan öznitelik seçilir (Quinlan, 1986). Böylece karmaşık karar süreçleri basit kurallar biçiminde ifade edilebilir. Ancak karar ağaçları tek başına kullanıldığında, özellikle karmaşık ve gürültülü veri setlerinde aşırı öğrenme (overfitting) eğilimi gösterebilir. Bu durum modelin eğitim verisine fazla uyum sağlamasına ve yeni veriler üzerindeki tahmin performansının düşmesine neden olabilir (Breiman vd., 1984).



Şekil 1. Karar Ağacı Algoritması (Komarasamy & Ravishankar, 2022)

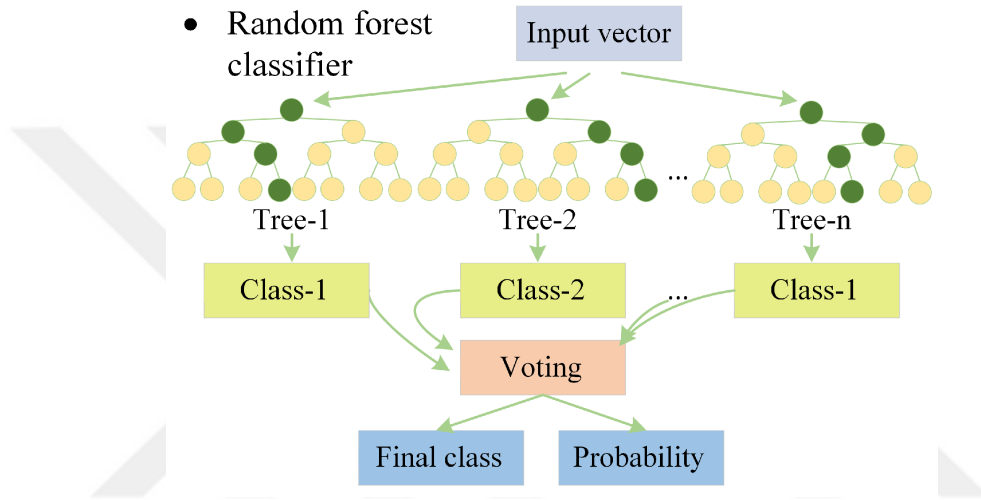
Şekil 1, bir karar ağacının temel yapısını göstermektedir. En üstte yer alan kök düğüm (root node), tüm veriyi temsil eder ve ilk bölünmenin yapıldığı noktadır. Daha sonra gelen karar düğümleri (decision nodes), belirli özelliklere göre veriyi alt gruplara ayırır. En uçlarda bulunan yaprak düğümler (leaf nodes) ise sınıflandırma veya regresyon sonucunu gösterir. Ayrıca, şeklin sol tarafında görüldüğü gibi birden fazla düğüm ve yapraktan oluşan kısım alt ağaç (sub-tree) olarak adlandırılır. Bu yapı sayesinde karmaşık karar süreçleri, adım adım basit alt kararlar hâline getirilir.

Uzaktan algılama ve coğrafi bilgi sistemleri uygulamalarında karar ağaçları, özellikle arazi örtüsü sınıflandırmalarında sıkça tercih edilmektedir. Karar ağaçları, Random Forest gibi topluluk yöntemlerinin de temel yapı taşını oluşturmaktadır (Belgiu & Drăguţ, 2016).

1.4.4. Random Forest

Random Forest, Breiman (2001) tarafından önerilen ve karar ağaçlarına dayalı topluluk (ensemble) öğrenme yöntemlerinden biridir. Bu yöntem, çok sayıda karar ağacının rastgele seçilmiş veri alt kümeleri ve öznitelik alt kümeleri üzerinde eğitilmesi ve sonuçların oylanması prensibine dayanır. Böylece tek bir karar ağacına kıyasla daha yüksek doğruluk elde edilmekte ve aşırı öğrenme (overfitting) riski azaltılmaktadır.

Random Forest algoritmasının en önemli avantajları, büyük boyutlu verilerle çalışabilmesi, yüksek doğruluk sağlaması ve değişken önem derecelerini (feature importance) hesaplayabilmesidir (Breiman, 2001). Uzaktan algılama alanında da yaygın bir kullanım alanı bulan bu yöntem, özellikle sınıflandırma ve değişim analizi çalışmalarında tercih edilmektedir. Belgiu & Drăguț (2016), Random Forest'ın uzaktan algılama uygulamalarında en sık kullanılan sınıflandırma algoritmalarından biri haline geldiğini ve yüksek performansı nedeniyle literatürde geniş yer bulduğunu belirtmiştir.



Şekil 2. Random Forest Algoritmasının Şematik Gösterimi (Zeng vd., 2022)

Şekil 2'deki görselde, eğitim verisi bootstrap yöntemiyle farklı alt kümelerle ayrılır, her alt küme ile bağımsız karar ağaçları eğitilir ve test verisi bu ağaçlardan geçirilir. Her ağaç kendi sınıflandırma sonucunu üretir ve en sonunda çoğunluk oyu (majority vote) yöntemiyle nihai sınıf belirlenir. Bu yaklaşım, modelin genelleme performansını artırırken aşırı öğrenme riskini azaltır (Zeng vd., 2022).

1.4.5. Karar Ağaçları ve Random Forest Farkı

Karar ağaçları, sınıflandırma ve regresyon problemlerinde kolay anlaşılır ve yorumlanabilir yapıları nedeniyle yaygın şekilde kullanılan yöntemlerdir (Quinlan, 1986). Ancak tek bir karar ağacı, özellikle karmaşık ve büyük veri kümelerinde aşırı öğrenmeye (overfitting) yatkın olabilir ve genelleme kabiliyeti düşük kalabilir (Breiman vd., 1984).

Bu sınırlamaları gidermek amacıyla Breiman (2001) tarafından geliştirilen Random Forest algoritması, birden fazla karar ağacını rastgele seçilmiş veri alt kümeleri ve öznitelik alt kümeleri üzerinde bağımsız biçimde eğiterek sonuçları birleştiren topluluk (ensemble) tabanlı bir öğrenme yöntemidir. Bu yaklaşım, karar ağaçlarının çeşitliliğinden yararlanarak modelin varyansını azaltmakta, genelleme gücünü artırmakta ve aşırı öğrenme riskini önemli ölçüde düşürmektedir. Rodriguez-Galiano vd. (2012), Random Forest algoritmasının uzaktan algılama çalışmalarında yüksek doğruluk ve kararlılık sağladığını belirtmiştir. Cutler vd. (2007) ise, Random Forest modelinde kullanılan değişken önem dereceleri (feature importance) metriklerinin sınıflandırma sürecinde hangi özniteliklerin belirleyici olduğunu ortaya koyarak modelin yorumlana bilirliliğini artırdığını vurgulamıştır.

Random Forest, karar ağaçlarının basitliğini ve diğer avantajlarını korurken, topluluk yapısı sayesinde daha yüksek doğruluk, istatistiksel sağlamlık ve gürültüye karşı dayanıklılık sunmaktadır. Pal (2005), Random Forest algoritmasının uzaktan algılama uygulamalarında karmaşık arazi örtüsü yapılarında yüksek doğruluk sağladığını ve diğer sınıflandırma yöntemlerine kıyasla daha tutarlı sonuçlar ürettiğini belirtmiştir. Belgiu & Drăguț (2016) ise Random Forest'ın arazi örtüsü ve kullanım sınıflandırmalarında en güvenilir ve etkili yöntemlerden biri olarak öne çıktığını, özellikle heterojen bölgelerde üstün performans sergilediğini vurgulamıştır.

Random Forest algoritmasının tercih edilmesinin bir diğer nedeni, veri dağılımı, parametrik varsayımlar veya doğrusal ilişkiler gerektirmemesidir. Bu yönüyle RF, SVM (Support Vector Machine) veya Yapay Sinir Ağları (ANN) gibi karmaşık parametre ayarları gerektiren yöntemlere kıyasla daha az kullanıcı müdahalesine ihtiyaç duymaktadır. Foody & Mathur (2006), uzaktan algılama uygulamalarında parametrik varsayımlar gerektirmeyen algoritmaların, özellikle sınıf dağılımlarının dengesiz olduğu durumlarda daha güvenilir sonuçlar verdiğini vurgulamıştır. Rodriguez-Galiano vd. (2012) ise Random Forest'ın yüksek boyutlu ve dengesiz veri kümelerinde K-Nearest Neighbors (KNN) ve Naive Bayes gibi yöntemlere kıyasla daha istikrarlı sonuçlar ürettiğini belirtmiştir.

Sonuç olarak, karar ağaçları yorumlana bilirlilik açısından avantaj sağlarken, Random Forest algoritması çok zamanlı ve yüksek boyutlu uydu verilerinde yüksek doğruluk, tutarlı performans ve değişken önem analizi gibi ek yetenekleri sayesinde bu çalışmada tercih edilmiştir.

2. YAPILAN ÇALIŞMALAR

2.1. Materyaller

2.1.1. Veri Kaynakları

Bu çalışmada temel uydu verisi olarak Sentinel-2 Multispektral Enstrüman (MSI) görüntülerinin Level-2A yüzey yansımaya (Surface Reflectance) ürünleri kullanılmıştır. Sentinel-2, Avrupa Uzay Ajansı (ESA) ve Avrupa Birliği'nin Kopernik Programı kapsamında işletilen, yüksek mekânsal ve zamansal çözünürlüğe sahip bir uydu sistemidir. 10–20 m çözünürlükte çok bantlı optik görüntüler sunması nedeniyle özellikle kentsel alan tespiti, arazi örtüsü sınıflandırması ve değişim izleme çalışmalarında yaygın olarak tercih edilmektedir. Drusch vd. (2012), Sentinel-2'yi, geniş bant aralığı, yüksek yenileme süresi ve atmosferik düzeltme kapasitesi ile çevresel izleme çalışmalarında önemli bir boşluğu dolduran yeni nesil bir misyon olarak tanımlamaktadır.

Çalışmada 2018 ve 2024 yıllarına ait Sentinel-2 L2A görüntüleri, Google Earth Engine (GEE) üzerinden temin edilmiştir. GEE, kullanıcıların küresel uydu veri arşivine doğrudan erişmesine, bu verileri bulut ortamında hızlı biçimde işlemesine ve büyük veri setlerini ölçeklenebilir şekilde analiz etmesine olanak tanıyan bir platformdur. Gorelick vd. (2017) GEE'yi, araştırmacılara ve uygulayıcılara gezegen ölçeğinde analiz imkânı sağlayan yenilikçi bir sistem olarak tanımlamaktadır.

Sınıflandırma sürecinde eğitim verisi olarak Dynamic World (DW) küresel arazi örtüsü veri seti kullanılmıştır. Dynamic World, Sentinel-2 görüntülerinden Google tarafından derin öğrenme tabanlı modellerle üretilmiş, 10 m mekânsal çözünürlükte, piksel bazlı sınıflandırma ve olasılık temelli çıktı sağlayan bir arazi örtüsü ürünüdür. Brown vd. (2022), Dynamic World'ü gerçek zamanlıya yakın güncellenebilen ve yüksek mekânsal çözünürlüklü arazi örtüsü verileri sunan ilk küresel ölçekli yapay zekâ ürünü olarak tanımlamaktadır.

Dynamic World veri setinde dokuz farklı arazi örtüsü sınıfı bulunmaktadır; bu çalışmada özellikle yerleşim alanı (built-up) olasılık bantları kullanılmıştır. DW'nin 10 m çözünürlüğü, Sentinel-2 verileriyle uyumlu olduğu için Landsat tabanlı çalışmalarla karşılaştırıldığında daha ayrıntılı ve güncel bir eğitim verisi sağlamaktadır. Bu durum,

Sentinel-2 görüntülerinin Random Forest algoritması ile sınıflandırılmasında elde edilen doğruluğun artmasına katkı sağlamaktadır (Brown vd., 2022).

Ayrıca Dynamic World'ün açık erişimli yapısı, geniş coğrafi alanlarda, tekrarlanabilir ve ölçeklenebilir analizlerin Python ve GEE ile entegrasyonunu kolaylaştırmış; böylece çalışmanın yeniden kullanılabilirliği ve geliştirilebilirliği daha da artmıştır.

2.1.2. Kullanılan Geliştirme Ortamları ve Araçları

Bu çalışmada Sentinel-2 uydu görüntülerinin işlenmesi, sınıflandırılması ve mekânsal analizlerin yürütülmesi için birbirini tamamlayan birden fazla yazılım aracı ve geliştirme ortamı birlikte kullanılmıştır. Verilerin temini, büyük ölçekli görüntü koleksiyonlarının yönetilmesi ve ön işleme adımlarının önemli bir kısmı Google Earth Engine (GEE) platformu üzerinden gerçekleştirilmiştir. GEE, geniş uydu arşivine erişim sunan bulut tabanlı altyapısı sayesinde kullanıcılara yüksek hacimli verilerle çalışmayı, zaman serisi analizleri yürütmeyi ve coğrafi ölçekli görüntü işleme süreçlerini yerel donanımdan bağımsız biçimde gerçekleştirme olanağı tanımaktadır. Gorelick ve arkadaşları (2017), GEE'yi gezegen ölçeğinde mekânsal analiz yapılmasına imkân veren bir sistem olarak tanımlamakta ve özellikle Sentinel-2 gibi yüksek çözünürlüklü verilerle yapılan çalışmalarda hız ve kolaylık sağladığını vurgulamaktadır. Bu çalışma kapsamında GEE, mevsimsel bileşik görüntülerin oluşturulması, bulut maskesi uygulamaları, Dynamic World arazi örtüsü verilerinin elde edilmesi ve eğitim-etiket veri setinin hazırlanması amacıyla etkin biçimde kullanılmıştır.

Verilerin sınıflandırılması, model eğitimi, model kalibrasyonu ve sonuçların türetilmesi Python programlama dili ile yürütülmüştür. Python, geniş kütüphane desteği ve esnek yapısı sayesinde bilimsel hesaplamalar ve uzaktan algılama çalışmalarında yaygın kullanılan bir araç haline gelmiştir. Sayısal işlemler için NumPy (Harris vd., 2020), tablo yapısındaki verilerin yönetimi için Pandas, görselleştirme süreçleri için Matplotlib (Hunter, 2007) ve makine öğrenmesi algoritmalarının uygulanması için scikit-learn (Pedregosa vd., 2011) kütüphaneleri kullanılmıştır. Özellikle Random Forest modellerinin eğitilmesi, OOB hata hesaplamaları, model kalibrasyonu (isotonic ve sigmoid yöntemleri) ve metrik değerlendirmeleri bu kütüphaneler aracılığıyla gerçekleştirilmiştir. Ayrıca, Python ile GEE arasında etkileşim kurulmasını sağlayan geemap kütüphanesi, hem veri çekme hem de harita tabanlı çıktıları görselleştirme açısından önemli bir rol üstlenmiştir.

Çalışmanın sürdürülebilir ve yeniden üretilebilir bir yapıda olması amacıyla Python ortamı Anaconda dağıtımı üzerinden yönetilmiştir. Anaconda, bağımlılıkların ve paket sürümlerinin kontrol altında tutulmasına olanak sağlayarak kodun farklı cihazlarda aynı şekilde çalışmasını kolaylaştırmıştır. Tüm analiz süreci için özel bir sanal ortam oluşturulmuş ve bu ortamda gerekli paketlerin kararlı sürümleri kullanılarak çalışma tekrarlanabilir hâle getirilmiştir (Anaconda Software Distribution, 2020).

Kod geliştirme aşamasında Visual Studio Code (VS Code) tercih edilmiştir. VS Code'un Python uzantıları, hata ayıklama araçları, entegre terminali ve kullanıcı dostu arayüzü, kodların modüler biçimde yazılması ve yönetilmesini kolaylaştırmıştır. Ayrıca Jupyter Notebook ortamı, özellikle ara testlerin yapılması, sınıflandırma sonuçlarının anlık olarak incelenmesi ve görselleştirilmesi açısından önemli bir katkı sağlamıştır. Kluyver ve arkadaşları (2016), Notebook yapısının bilimsel analizlerde şeffaflık ve yeniden üretilebilirlik sağladığını vurgulamaktadır. Bu çalışma kapsamındaki tüm deneysel süreçler, Jupyter Notebook'un hücresel yapısı sayesinde adım adım kaydedilerek açıklanabilir ve tekrar uygulanabilir bir hale getirilmiştir.

Bu bütünleşik yapı sayesinde, Sentinel-2 verilerinin ön işlenmesinden Random Forest sınıflandırmasına, bina değişimi hesaplamalarından mekânsal çıkarımların elde edilmesine kadar tüm süreç tutarlı, modüler ve yeniden kullanılabilir bir analiz akışı içerisinde gerçekleştirilmiştir.

2.2. Yöntem

Bu çalışmada, İstanbul'daki Ömerli Barajı ve Ankara'daki Çubuk-1 Barajı çevresinde 2018–2024 dönemi boyunca meydana gelen yerleşim alanı değişimlerinin, Sentinel-2 uydu görüntüleri ve makine öğrenmesi tabanlı bir yaklaşım ile belirlenmesi amaçlanmakta; aynı zamanda bu sürecin yeniden kullanılabilir, modüler ve farklı çalışma alanlarına uyarlanabilir bir kod yapısı ile nasıl gerçekleştirilebileceği incelenmektedir. Yöntem, hem uzaktan algılama verilerinin işlenmesine dayalı klasik bir yerleşim alanı değişim analizi, hem de kullanılan kodun bileşenlerinin (bant/indeks seti, ağaç sayısı, eşik seçimi, mekânsal grid vb.) model başarımı üzerindeki etkilerini sistematik olarak değerlendiren deneysel bir çerçeve sunmaktadır.

Çalışmada temel veri kaynağı olarak Sentinel-2 Level-2A yüzey yansıma ürünleri kullanılmış, her iki çalışma alanı için 2018 ve 2024 yıllarına ait görüntüler Google Earth

Engine (GEE) platformu üzerinden temin edilmiştir. Yöntemin önemli bir bileşeni, tek bir tarih yerine yıl içinde farklı mevsimleri temsil eden görüntülerin bileşik (composite) olarak kullanılmasıdır. Bu amaçla her yıl için dört mevsim (DJF, MAM, JJA, SON) ayrı ayrı ele alınmış; bulutluluk, kar ve gölge gibi istenmeyen etkiler, Sentinel-2 SCL bandına dayalı maskeleme ile giderilmiş ve her mevsim için median bileşik görüntüler üretilmiştir. Daha sonra bu mevsimsel bileşikler bir araya getirilerek, karar verme sürecinde hem zamansal hem de spektral çeşitlilik içeren çok bantlı bir özellik kümesi oluşturulmuştur. Bu kümede Sentinel-2'nin temel spektral bantlarına ek olarak, klasik spektral indeksler (örneğin NDVI, NDWI, NDBI, kentsel indeks türevleri) de hesaplanarak girdi değişkeni (özellik uzayı) genişletilmiştir.

Modelin etiket bilgisi, Dynamic World veri kümesinin yerleşim (built-up) olasılık bandından türetilmiştir. Dynamic World verileri, ilgili yıl boyunca çalışma alanını kapsayan tüm sahalardan ortalanarak her piksel için 0 ile 1 arasında değişen bir yerleşim olasılığı elde edilmiştir. Bu olasılık haritası, belirli bir eşik değeri ile ikili sınıfa (yerleşim / yerleşim dışı) dönüştürülmüş ve böylece Random Forest modeli için eğitim ve test etiketleri oluşturulmuştur. Eşik değeri, çalışma kapsamında sabit alınmak yerine, model performansını maksimize edecek biçimde otomatik olarak belirlenmiştir. Bu amaçla, olasılık çıktılarına karşılık gelen kesinlik-duyarlılık (precision-recall) eğrileri üzerinden F1 skorunun en yüksek olduğu eşik değeri seçilmiş ve her yıl için optimum karar sınırı veri odaklı olarak tanımlanmıştır.

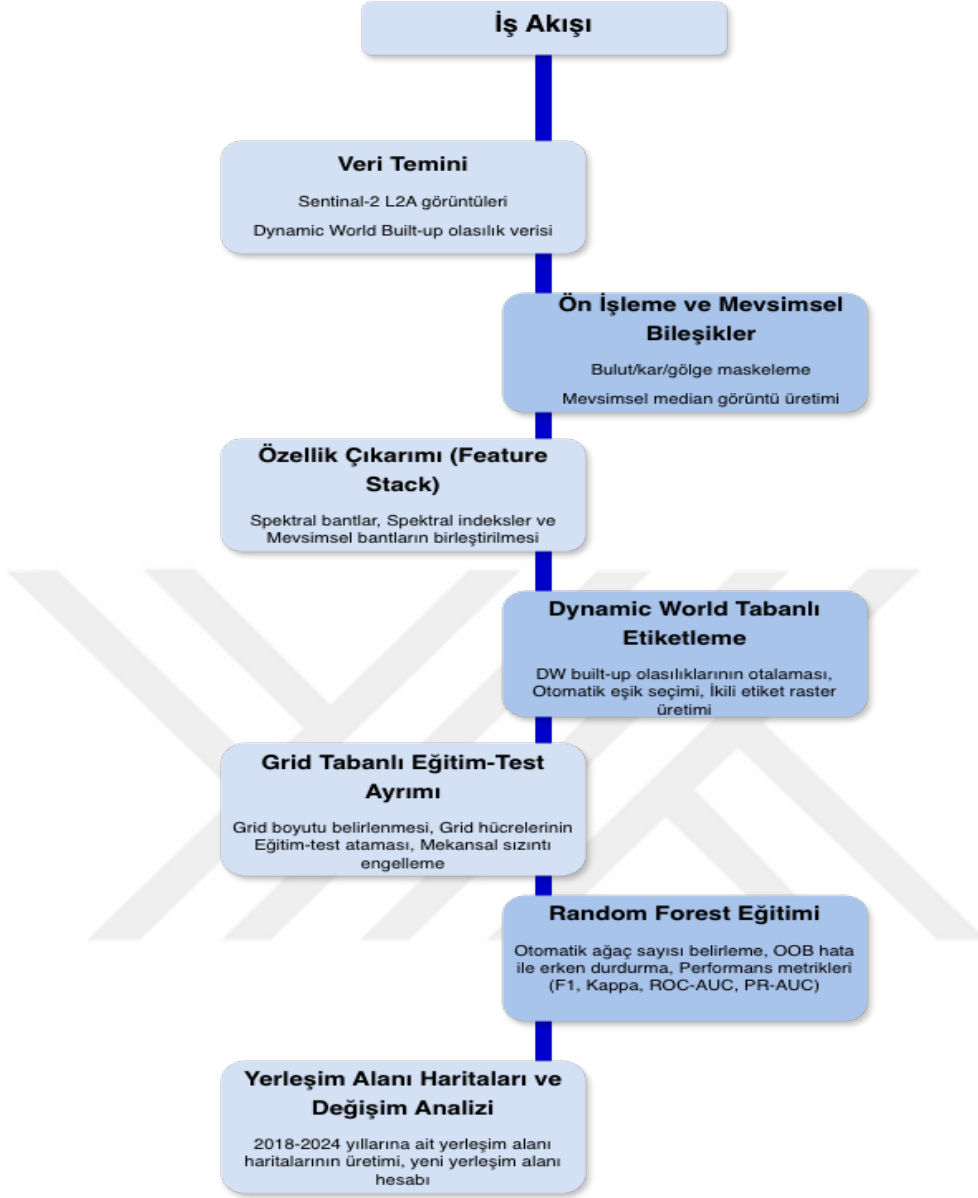
Sınıflandırma süreci, Random Forest algoritması kullanılarak yürütülmüştür. Model, yalnızca sabit bir ağaç sayısı ile değil, çalışma alanının büyüklüğü ve kullanılan örnek sayısına bağlı olarak ağaç sayısını uyarlayan bir kural ve Out-of-Bag (OOB) hata temelli erken durdurma mantığıyla yapılandırılmıştır. Böylece, ağaç sayısı otomatik olarak belirlenmiş ve farklı bölgeler ya da farklı veri hacimleri için yeniden kullanılabilir bir yapı elde edilmiştir. Eğitim ve test veri setleri, mekânsal sızıntıyı (spatial leakage) engellemek amacıyla rastgele piksel seçiminden ziyade, çalışma alanının otomatik üretilen bir grid yapısına bölünmesi ile oluşturulmuştur. AOI (Area of Interest) alanına bağlı olarak grid boyutu (örneğin 3×3, 5×5, 6×6, 8×8) otomatik seçilmiş; belirli hücreler eğitim, diğerleri test için kullanılarak, aynı mekânsal bloktan hem eğitim hem test örneği alınmasının önüne geçilmiştir. Bu yaklaşımla modelin uzamsal genelleme kabiliyetinin daha gerçekçi biçimde değerlendirilmesi hedeflenmiştir.

Yöntemin özgün yönlerinden biri, bu modüler yapının yalnızca “nihai model” üretmek için değil, aynı zamanda kod bileşenlerinin gerekliliğini ve etkisini sorgulamak için de kullanılmasıdır. Bu kapsamda, temel iş akışı referans senaryo olarak belirlenmiş ve ardından aşamalı olarak sadeleştirilmiş veya değiştirilmiş çeşitli senaryolar denenmiştir. Örneğin, spektral indekslerin girdi değişkenlerinden çıkarıldığı, ağaç sayısının otomatik belirlenmek yerine sabit tutulduğu, karar eşiğinin otomatik seçilmeyip 0.50 gibi varsayılan bir değere sabitlendiği ve mekânsal grid ayrımının kaldırılarak klasik rastgele örneklemeye dönüldüğü alternatif modeller oluşturulmuştur. Her bir senaryoda model performansı F1 skoru, Kappa istatistiği, ROC-AUC ve PR-AUC gibi metrikler üzerinden değerlendirilmiş; ayrıca elde edilen bina alanı ve bina değişimi (yeni bina, bina kaybı, net değişim) sonuçları referans senaryo ile karşılaştırılmıştır. Böylece, yeniden kullanılabilir bir kod yapısında hangi bileşenlerin model başarımına ne ölçüde katkı sağladığı, hangi durumlarda sadeleştirmenin kabul edilebilir olduğu ve hangi bileşenlerin vazgeçilmez olduğu sistematik olarak ortaya konmuştur.

Son aşamada, seçilen referans senaryo kullanılarak 2018 ve 2024 yılları için her iki çalışma alanında bina alanı dağılımı haritaları üretilmiş ve bu haritalar üzerinden bina değişim analizi gerçekleştirilmiştir. Bina değişimi, iki yıl arasındaki olasılık farkına dayalı olarak tanımlanmış; belirli bir eşik değeri üzerinde artış gösteren pikseller “yeni bina”, azalış gösterenler “bina kaybı” ve değişmeyen alanlar “stabil” sınıfında değerlendirilmiştir. Min-yama (speckle) filtreleme adımı ile çok küçük ve izole lekeler temizlenmiş, sonuçta hem Ömerli Barajı hem de Çubuk-1 Barajı çevresinde bina değişim miktarları ve mekânsal desenleri nicel ve görsel olarak değerlendirilmiştir. Bu sonuçlar yalnızca kentsel büyüme dinamiklerini ortaya koymakla kalmamış, aynı zamanda yeniden kullanılabilir kod yapısının farklı senaryolar altında ürettiği çıktıların kıyaslanmasına olanak tanımıştır.

2.2.1. İş Akışı

Bu çalışmada bina alanlarının belirlenmesi ve 2018–2024 dönemindeki değişimin hesaplanması için geliştirilen yöntem, hem Sentinel-2 görüntülerine dayalı spektral bilgi çıkarımını hem de Random Forest sınıflandırma sürecini içeren çok aşamalı bir iş akışına sahiptir. İş akışı, yeniden kullanılabilir bir kod yapısı oluşturmak amacıyla modüler biçimde tasarlanmış olup, her bir bileşenin bağımsız olarak değiştirilebilmesine ve farklı senaryolar altında sınanabilmesine olanak tanımaktadır.

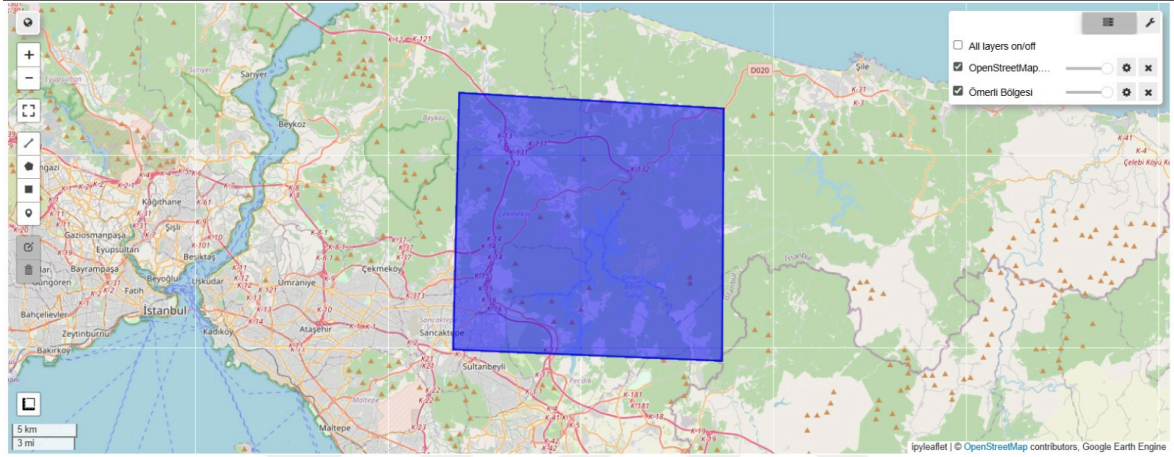


Şekil 3. İş Akış Şeması

2.2.2. Çalışma Alanı

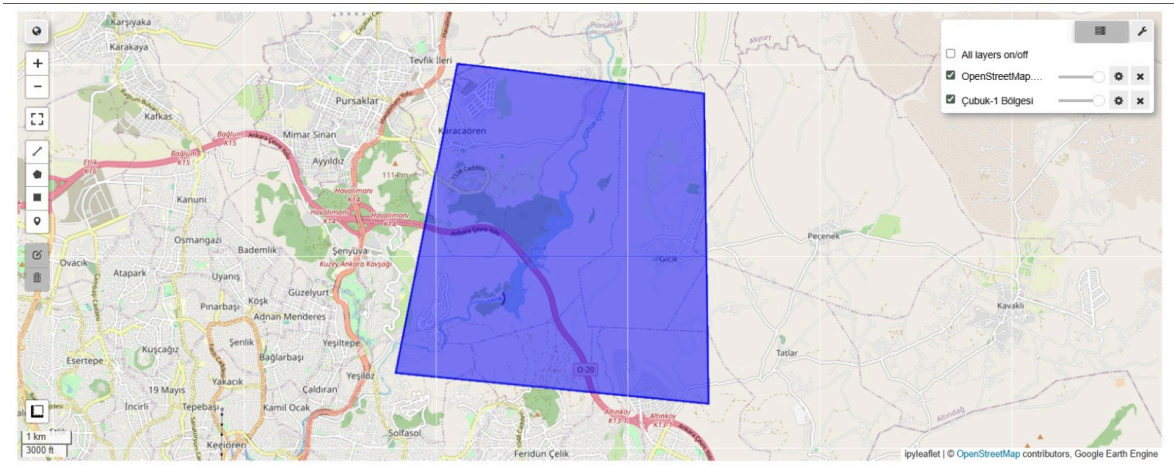
Bu çalışma, Türkiye'nin iki büyük metropolünde içme suyu arzı açısından kritik öneme sahip olan İstanbul Ömerli Barajı ve Ankara Çubuk-1 Barajı havzalarında yürütülmüştür. Her iki havza, bulundukları kentlerin su güvenliğinde merkezi bir rol üstlenmekte olup, çevrelerinde gerçekleşen yerleşim alanı baskıları hem ekosistem sağlığı hem de sürdürülebilir havza yönetimi açısından önemli bir risk oluşturmaktadır. Bu nedenle çalışma alanları, yerleşim alanı değişimlerinin zamansal olarak izlenmesi için uygun örnek bölgeler olarak seçilmiştir.

Ömerli Barajı, İstanbul'un Anadolu Yakası'nda konumlanan ve kente içme suyunun bir bölümünü sağlayan kaynağıdır. Barajın çevresi, son yıllarda artan nüfus, ulaşım yatırımları, konut projeleri ve kontrolsüz yapılaşma baskısı nedeniyle İstanbul'un en hızlı dönüşen bölgelerinden biridir. Bu durum, özellikle havza koruma sınırları içinde gerçekleşen yapılaşmaların düzenli olarak izlenmesini zorunlu kılmaktadır.



Şekil 4. Ömerli Barajı Çalışma Alanı

Ankara'da yer alan Çubuk-1 Barajı, Türkiye Cumhuriyeti'nin planlı şehircilik uygulamalarının ilk örneklerinden biri olarak inşa edilmiş olup, kentin su temininde uzun yıllar boyunca önemli bir rol oynamıştır. Çubuk-1 Barajı çevresindeki yerleşim alanı gelişmeleri havza bütünlüğü ve çevresel sürdürülebilirlik açısından önem taşımaktadır.



Şekil 5. Çubuk-1 Barajı Çalışma Alanı

Her iki çalışma alanının sınırları, kesin koordinat değerlerine dayalı olarak çokgen (polygon) geometri biçiminde oluşturulmuş ve tüm mekânsal işlemler bu ortak geometri üzerinde yürütülmüştür. Sentinel-2 uydu görüntülerinden oluşturulan mevsimsel bileşikler (DJF, MAM, JJA, SON), bu sınırlar dahilinde kesilmiş (clip edilmiş) ve tüm sınıflandırma, eğitim-test ayırımı, bina maskesi üretimi ve değişim analizi süreçlerinde aynı çalışma geometrisi kullanılmıştır.

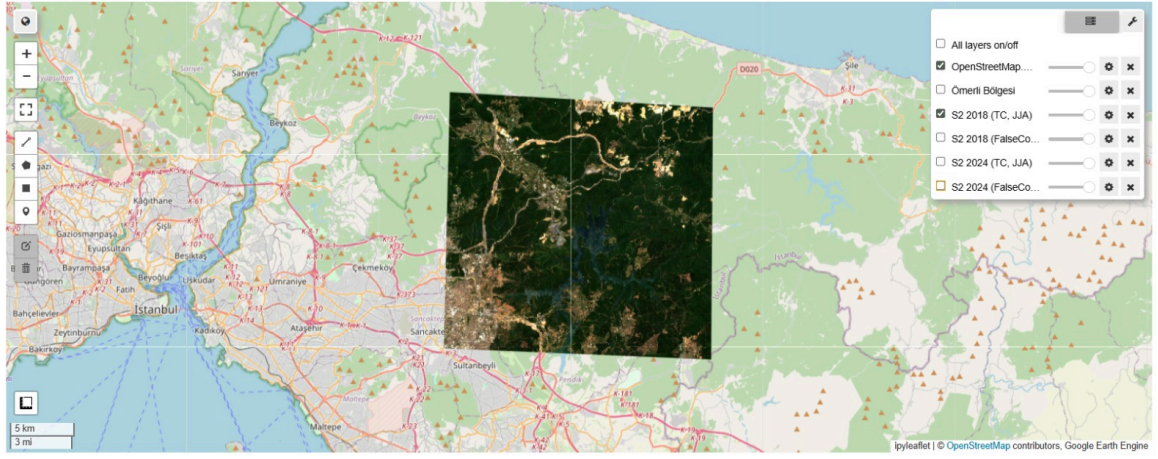
Bu sayede, Ömerli ve Çubuk-1 baraj havzaları için 2018–2024 yılları arasında yerleşim alanı değişimleri karşılaştırılabilir, sistematik ve yeniden üretilebilir bir yöntemle analiz edilmiştir.

Tablo 1. Çalışma Alanlarına Ait Koordinatlar ve Alan Büyüklükleri

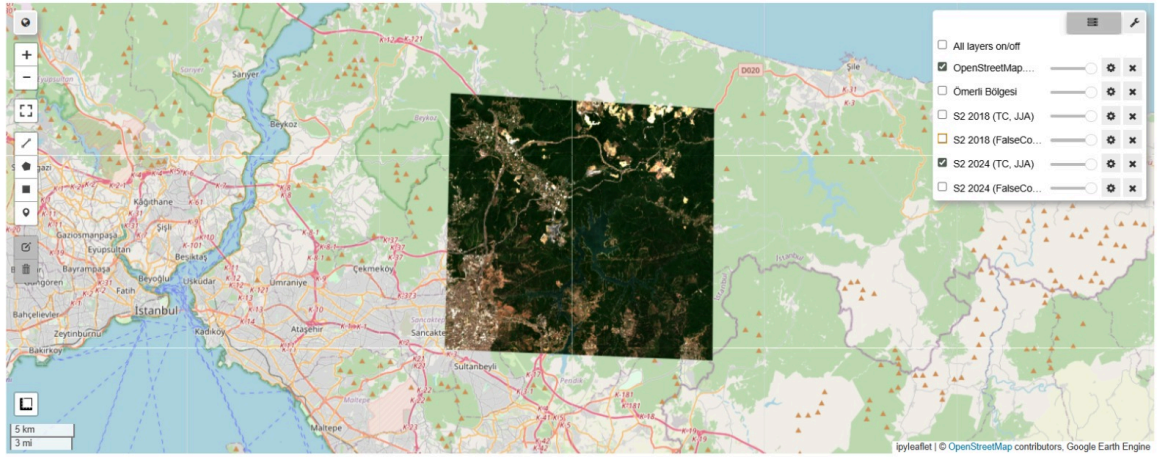
Çalışma Alanı	Çalışma Alanı Köşe Koordinatları		Çalışma Alanı(km ²)
	Longitude	Latitude	
Ömerli Barajı	29.221388	41.174444	398 km ²
	29.509444	41.161666	
	29.507500	40.952222	
	29.214722	40.961388	
Çubuk-1 Barajı	33.007025	40.032889	63 km ²
	33.077085	40.033573	
	33.076171	39.983533	
	33.007702	39.983085	

2.2.3. Veri Temini

Bu çalışmada kullanılan uydu verileri, Avrupa Uzay Ajansı (ESA) tarafından işletilen Sentinel-2A ve Sentinel-2B uydularına ait Level-2A yüzey yansıma (Surface Reflectance) ürünleridir. Sentinel-2 verileri, 10–20 m mekânsal çözünürlüğü, yüksek spektral çeşitliliği ve 5 günlük tekrar geçiş süresi sayesinde yerleşim alanlarının tespiti ve değişim analizlerinde yaygın olarak tercih edilmektedir.



Şekil 6. 2018 Yılı Ömerli Barajı Sentinel-2 Mevsimsel Bileşik Görüntüsü

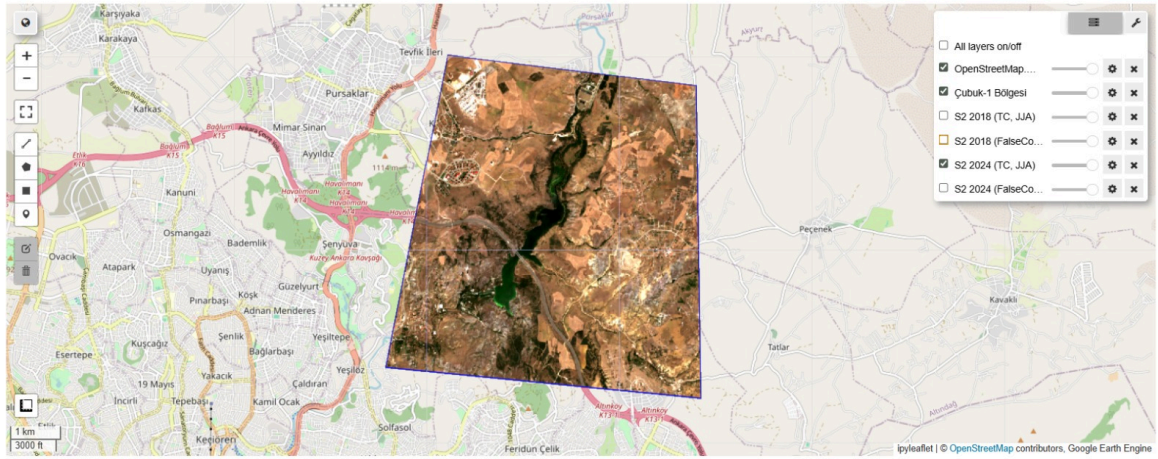


Şekil 7. 2018 Yılı Ömerli Barajı Sentinel-2 Mevsimsel Bileşik Görüntüsü

Çalışmada hem İstanbul Ömerli Barajı hem de Ankara Çubuk-1 Barajı için 2018–2024 yılları arasındaki veriler kullanılmıştır. Tek bir görüntü yerine, yıl boyunca elde edilen tüm görüntüler içinden seçilen dört farklı mevsime (DJF, MAM, JJA, SON) ait sahneler birleştirilerek mevsimsel median bileşikler oluşturulmuştur. Bu yaklaşım, tek bir tarihe bağımlı olmanın doğurabileceği bulut, mevsimsel değişkenlik veya atmosferik etkiler gibi sorunları azaltarak daha temsil gücü yüksek veri üretmiştir.



Şekil 8. 2018 Yılı Çubuk-1 Barajı Sentinel-2 Mevsimsel Bileşik Görüntüsü



Şekil 9. 2024 Yılı Çubuk-1 Barajı Sentinel-2 Mevsimsel Bileşik Görüntüsü

Uydu görüntülerinin temin edilmesi ve ön filtreleme işlemleri Google Earth Engine (GEE) platformu üzerinden gerçekleştirilmiştir. GEE, küresel Sentinel-2 arşivine doğrudan erişim sağlamakta; tarih, mevsim, bulutluluk oranı ve çalışma alanı ile kesişme gibi kriterlere göre veri seçimini otomatik olarak gerçekleştirmektedir. Bu çalışmada, bulutluluk oranı sahne bazında %60–85 aralığında sınırlandırılmış, bulut ve gölge maskesi uygulanarak temiz veri seti elde edilmiştir.

2.2.4. Spektral İndekslerin Hesaplanması

Bu çalışmada Sentinel-2 uydu verileri üzerinde yerleşim, bitki örtüsü ve su yüzeylerinin ayırt edilebilmesi için çeşitli spektral indeksler hesaplanmış ve mevsimsel bileşik görüntülere ek öznitelikler olarak dahil edilmiştir. Spektral indeksler, Random Forest

sınıflandırma modelinin ayırt etme gücünü artırmış; yerleşim–bitki, yerleşim–su ve toprak–yapay yüzey ayrımlarını belirginleştirmiştir.

NDVI (Normalized Difference Vegetation Index), uzaktan algılama çalışmalarında canlı yeşil bitki örtüsünün durumu ve yayılımını değerlendirebilmek için yaygın olarak kullanılan bir bitki indeksi çeşididir. İlk kez 1970’lerde Rouse ve arkadaşları tarafından geliştirilmiş olup, bitki yapraklarının kırmızı ışığı büyük ölçüde absorbe etmesi, yakın kızılötesi ışığı ise güçlü şekilde yansıtması prensibine dayanmaktadır (Rouse vd., 1974). NDVI şu formül ile hesaplanır:

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

İndeksin değerleri -1 ile +1 arasında değişmekte olup, yüksek pozitif değerler sağlıklı ve yoğun bitki örtüsüne; düşük pozitif veya sıfıra yakın değerler seyrek bitki örtüsü ya da çıplak toprağa; negatif değerler ise su yüzeyi, kar veya bulut gibi bitki olmayan yüzeylere işaret etmektedir (Tucker, 1979). NDVI’nin hesaplanması görece basit olmakla birlikte, farklı uydu platformlarından elde edilen kırmızı ve NIR bantları kullanılarak uygulanabilmektedir. Bu özelliği sayesinde zamansal değişimlerin incelenmesine olanak tanımakta ve geniş bir kullanım alanı sunmaktadır. Pettorelli vd. (2005), NDVI’nin ekosistem izleme ve çevresel değişimlerin değerlendirilmesinde güçlü bir gösterge olduğunu belirtmiştir. Ji & Peters (2003) ise NDVI’nin bitki örtüsündeki mevsimsel ve yıllık değişimlerin izlenmesinde etkin biçimde kullanılabileceğini ortaya koymuştur. Ancak, literatürde de vurgulandığı üzere NDVI’nin yüksek bitki yoğunluklarında doygunluk problemi yaşaması, atmosferik etkilerden ve toprak yansımalarından etkilenmesi gibi bazı sınırlamaları bulunmaktadır (Huete vd., 2002).

NDBI (Normalized Difference Built-up Index), kentsel alanların tespiti amacıyla Zha vd. (2003) tarafından geliştirilmiştir. İndeks, kısa dalga kızılötesi (SWIR) ile yakın kızılötesi (NIR) bantların farkının toplamına oranlanması ile hesaplanır. Bu hesaplama, yapay yüzeylerin SWIR bandında yüksek, NIR bandında ise düşük yansıtma özelliğine sahip olmasına dayanır (Zha vd., 2003). NDBI’nin yüksek değerleri genellikle yapay yüzeylere, düşük değerleri ise bitki örtüsüne karşılık gelir. Literatürde, NDBI’nin özellikle hızlı kentleşmenin izlenmesinde ve kentsel yayılımın nicel olarak değerlendirilmesinde etkili bir araç olduğu ifade edilmektedir (Liu vd., 2010). Bununla birlikte, çıplak toprak ile yapay yüzeyler arasında benzer spektral özellikler nedeniyle karışıklık yaşanabilmesi bu indisin

başlıca sınırlamaları arasında gösterilmektedir (He vd., 2010). NDBI şu formül ile hesaplanır:

$$NDBI = \frac{SWIR - NIR}{SWIR + NIR} \quad (2)$$

NDWI (Normalized Difference Water Index): özellikle su yüzeylerinin tespiti için geliştirilmiş bir uzaktan algılama indisi. İlk olarak McFeeters (1996) tarafından önerilen bu indis, yeşil bant ile yakın kızılötesi (NIR) bant arasındaki farkın normalleştirilmesine dayanır. NDWI, açık su yüzeylerinin yansıtma özelliklerini güçlendirmekte ve bu sayede su kütlelerinin diğer arazi örtüsü sınıflarından ayrımını kolaylaştırmaktadır (McFeeters, 1996). İndeksin değerleri -1 ile +1 arasında değişmekte olup, pozitif değerler genellikle su yüzeylerini temsil ederken negatif değerler bitki örtüsü veya yapay yüzeyleri işaret etmektedir (Xu, 2006). NDWI'nın en önemli avantajı, göl, nehir, baraj gibi yüzeysel su varlıklarının zamansal değişimlerinin kolaylıkla izlenmesine olanak tanımasıdır. Bununla birlikte, bitki örtüsü veya yapay yapılarla kaplı bölgelerde yanlış sınıflamalara yol açabilmesi literatürde vurgulanan kısıtlamaları arasındadır (Xu, 2006). NDWI şu formül ile hesaplanır:

$$NDWI = \frac{GREEN - NIR}{GREEN + NIR} \quad (3)$$

UI (Urban Index)i yerleşim alanlarının tespiti için literatürde önerilen ve NDBI ile NDVI arasındaki farkın alınmasına dayalı bir göstergedir. Bu yaklaşım, yapay yüzeylerin NDBI değerlerinin yüksek, bitki örtüsünün ise NDVI değerlerinin yüksek olması prensibine dayanmaktadır (Kawamura vd., 1996). Dolayısıyla, iki indeks arasındaki farkın kullanılması, bitki örtüsü ile yapay yüzeylerin daha belirgin bir şekilde ayrılmasını sağlamaktadır. UI, özellikle kentsel alanların hızlı ve doğru biçimde haritalanması amacıyla kullanılmakta ve diğer tekil indekslere kıyasla arazi örtüsü sınıfları arasındaki karışıklığı azaltma potansiyeli sunmaktadır (Xu, 2008). Bununla birlikte, çıplak toprak alanları ile düşük yoğunluklu yerleşim bölgelerinin benzer yansıma özellikleri göstermesi, UI'nın sınırlamalarından biri olarak değerlendirilmektedir (Li vd., 2017). UI şu formül ile hesaplanır:

$$UI = NDBI - NDVI \quad (4)$$

2.2.5. Örnekleme, Eğitim ve Doğrulama

Bu çalışmada yerleşim alanı sınıflandırması, Sentinel-2 mevsimsel bileşiklerinden (DJF, MAM, JJA, SON) oluşturulan çok bantlı özellik kümesi ile Dynamic World yerleşim (built-up) olasılık haritalarından türetilen etiket rasterlarının birleştirilmesiyle gerçekleştirilmiştir. Dynamic World'ün built-up olasılık değerleri, çalışmada geliştirilen yöntem kapsamında her bölge ve yıl için otomatik olarak belirlenen optimum eşik değeri kullanılarak ikili sınıf etiketine dönüştürülmüştür. Böylece sabit bir eşik değeri yerine, çalışma alanındaki spektral ayrışmaya en uygun eşik seçilmiş ve etiket kalitesi artırılmıştır.

Model eğitiminde mekânsal sızıntıyı önlemek temel bir kriter olduğundan, çalışma alanı önce otomatik olarak belirlenen boyuta sahip bir mekânsal ızgaraya (grid) bölünmüş, belirli hücreler “test bölgesi”, geri kalanları “eğitim bölgesi” olarak tanımlanmıştır. Böylece aynı bölgedeki komşu piksellerin hem eğitim hem test aşamasında kullanılmasının önüne geçilmiş, modelin gerçek genelleme performansını ölçmek mümkün olmuştur.

Örnekleme işlemi, bu mekânsal bölünmüş geometriler üzerinde tabakalı örnekleme (stratified sampling) yöntemi ile yapılmıştır. Dynamic World etiketlerinin pozitif (built-up) ve negatif (non-built-up) sınıfları çalışma alanında dengesiz olduğundan, her iki sınıftan da eşit sayıda piksel seçilerek sınıf dengesinin korunması sağlanmıştır. Toplam örnek sayısı çalışma alanı büyüklüğüne göre otomatik ölçeklendirilmiş; her iki sınıf için sınıf başına yaklaşık 4.000–7.000 piksel örnek alınmıştır. Bu yaklaşım, hem veri miktarını kontrol altında tutmakta hem de eğitim/test örneklerinin çalışma alanına orantılı biçimde belirlenmesine olanak vermektedir.

Eğitim verisi üzerinde Random Forest modeli uygulanmış ve ağaç sayısı ($n_{estimators}$) çalışma alanı büyüklüğü ve örnek sayısına göre otomatik olarak belirlenmiştir. İlk olarak ampirik bir başlangıç değeri hesaplanmış, ardından artan ağaç sayıları için OOB (Out-Of-Bag) hatası izlenerek erken durdurma (early stopping) uygulanmış ve en uygun değer seçilmiştir. Bu sayede gereksiz yere büyük modeller oluşturulmamış, optimize edilmiş model karmaşıklığı sağlanmıştır.

Bu yaklaşım, klasik “sabit örnekleme–sabit ağaç sayısı–sabit eşik” yapılarından farklı olarak tamamen veri ve bölge odaklı, yeniden kullanılabilir, parametrik olmayan bir

sınıflandırma çerçevesi sunmakta ve farklı coğrafyalarda minimum ayarla kullanılabilir hale gelmektedir.

2.2.6. Random Forest Modeli ve Otomatik Ağaç Sayısı Seçimi

Bu çalışmada yerleşim (built-up) alanlarının sınıflandırılmasında Rastgele Orman (Random Forest, RF) algoritması tercih edilmiştir. RF, çok sayıda karar ağacının rastgele seçilmiş örnekler ve öznitelikler üzerinde bağımsız olarak eğitilmesi ve bu ağaçlardan elde edilen tahminlerin çoğunluk oyu ile birleştirilmesi prensibine dayanır (Breiman, 2001).

Klasik RF uygulamalarında ağaç sayısı genellikle sabit bir değer olarak belirlenirken (ör. 100, 300, 500 ağaç), bu çalışma farklı olarak otomatik ağaç sayısı seçimi kullanmaktadır. Bu kapsamda, bootstrap örnekleme yapısı gereği her ağacın eğitiminde seçilmeyen örnekler “Out-of-Bag (OOB)” verisi olarak adlandırılır ve model performansı bu veri üzerinde hesaplanır (Breiman, 2001).

Bu çalışmada Random Forest modelinin ağaç sayısı, sabit bir değer olarak belirlenmek yerine OOB (Out-of-Bag) hata oranına dayalı otomatik bir seçim süreciyle optimize edilmiştir. Bu amaçla önce minimum ve maksimum ağaç sayılarından oluşan bir aralık (örneğin 100–400) tanımlanmış, ardından bu aralıktaki her bir ağaç sayısı için ayrı bir RF modeli eğitilmiştir. Her model için OOB hata oranı hesaplanarak hata eğrisinin değişimi izlenmiştir. OOB hata değerinin belirgin şekilde azalmadığı, diğer bir ifadeyle eğrinin plato yaptığı nokta “erken durdurma” yaklaşımıyla belirlenmiş ve OOB’un en düşük olduğu ya da azalışın anlamlı düzeyin altına düştüğü ağaç sayısı optimum değer olarak seçilmiştir. Böylece gereksiz derecede büyük modellerin oluşturulmasının önüne geçilmiş, ağaç sayısının çalışma alanı ve yıl özelindeki spektral karmaşıklığı yansıtması sağlanmış ve modelin hem açıklanabilirliği hem de yeniden üretilebilirliği artırılmıştır.

Uygulamada yapılan denemeler, Ömerli ve Çubuk-1 havzaları için optimum ağaç sayısının yıllara göre değişiklik gösterdiğini ortaya koymuştur. Örneğin 2018 yılı için optimum değer 150 ağaç olarak belirlenirken, 2024 yılı için bu sayı 200 ağaç seviyesine yükselmiştir. Bu farklılık, her yılın mevsimsel bileşiklerinin spektral çeşitliliği, Dynamic World etiketlerinin mekânsal dağılımı, kullanılan bant ve spektral indeks çeşitliliği ile sınıf ayrışmasının zorluk derecesi gibi veri yapısına özgü etmenlerden kaynaklanmaktadır.

Random Forest modellerinin optimum ağaç sayısının veri karmaşıklığına göre değişebildiği yönündeki bulgular literatürle de uyumludur. Rodriguez-Galiano vd. (2012),

optimum ağaç sayısının veri setinin yapısına bağlı olarak önemli ölçüde farklılaşabileceğini göstermiştir. Benzer şekilde Millard ve Richardson (2015) da farklı uzaktan algılama senaryolarında optimum ağaç sayısının sabit olmadığını ve veri özelliklerine göre yeniden belirlenmesi gerektiğini vurgulamıştır.

2.2.7. Olasılık Tabanlı Sınıflandırma ve Otomatik Threshold Belirleme

Bu çalışmada yerleşim (built-up) sınıflandırması, ikili (0/1) karar üretmek yerine, her piksel için yerleşim olma olasılığını tahmin eden olasılık tabanlı bir yaklaşım ile gerçekleştirilmiştir. Random Forest algoritması, her bir piksel için 0 ile 1 arasında bir olasılık değeri üretmekte ve bu değer, pikselin yerleşim sınıfına ait olma ihtimalini ifade etmektedir. Olasılık temelli çıktılar, özellikle sınıf sınırlarının belirsiz olduğu bölgelerde daha doğru bir yorum yapmaya imkân tanımaktadır (Foody & Mathur, 2006). Benzer şekilde, düşük yoğunluklu yerleşim alanları veya geçiş zonları gibi karmaşık örneklerde, tahmin olasılıklarının değerlendirilmesi sınıflandırma güvenilirliğini artırmakta ve yanlış sınıflandırmaların azaltılmasına katkı sağlamaktadır (Han vd., 2011).

Sınıflandırma sonrası elde edilen olasılık değerlerinin ikili sınıfa dönüştürülmesi için sabit bir eşik (örneğin 0.50) kullanmak yerine, otomatik threshold belirleme yöntemi uygulanmıştır. ROC (Receiver Operating Characteristic) eğrisi ve AUC (Area Under Curve) değerlerinin hesaplanmasıyla, duyarlılık (true positive rate) ve özgüllük (true negative rate) arasındaki farkın maksimum olduğu nokta Youden J istatistiği ile tanımlanmaktadır (Youden, 1950). ROC eğrisi ve AUC metrinin genel prensipleri ise Fawcett (2006) tarafından ayrıntılı biçimde açıklanmıştır.

Bu yaklaşım, özellikle sınıf dağılımlarının dengesiz olabildiği uzaktan algılama uygulamalarında önerilmektedir. Pontius & Parmentier (2014) çalışmasında, dengesiz sınıf problemlerinde ROC temelli eşik seçiminin doğruluk metriklerini önemli ölçüde iyileştirdiği gösterilmiştir. Benzer biçimde Foody (2002) de uzaktan algılamada olasılık temelli sınıflandırmalar için ROC ve eşik optimizasyonunun kritik olduğunu vurgulamaktadır.

Bu çalışmada otomatik threshold yöntemi, 2018 ve 2024 yılları için yaklaşık 0.48–0.50 aralığında eşik seçilmesine yol açmıştır. Bu değerler, uluslararası literatürde Random Forest + kalibrasyon yöntemleriyle kentsel alan sınıflandırması yapan çalışmalarda rapor edilen eşik aralıklarıyla uyumludur. Bu durum, kullanılan otomatik seçim sürecinin

istatistiksel olarak tutarlı ve bölgeye uyarlanabilir bir karar mekanizması sunduğunu göstermektedir.

Sonuç olarak, otomatik threshold belirleme yaklaşımı sabit eşik kullanımının neden olabileceği yanlış pozitif ve yanlış negatif sınıflandırmaları önemli ölçüde azaltmış; hem yıllar arasında hem de farklı çalışma alanları arasında daha karşılaştırılabilir ve güvenilir yerleşim maskeleri elde edilmesini sağlamıştır.

2.2.8. Bina Maskesi Üretimi, Yama Temizleme ve Değişim Analizi

Sınıflandırma modeli tarafından üretilen olasılık haritaları, ROC/AUC analizi ile belirlenen optimum eşik değerinin uygulanmasıyla ikili bina maskelerine (0 = bina değil, 1 = bina) dönüştürülmüştür. Bu yaklaşım, geleneksel sabit eşik yöntemlerine göre daha esnek olup, her yıl ve çalışma bölgesi için veri dağılımına uyarlanabilen daha duyarlı bina-sınıf ayrımı sağlamaktadır (Foody, 2002). ROC/AUC temelli eşikleme yöntemlerinin uzaktan algılama çalışmalarında performansı artırdığı çeşitli araştırmalarda da vurgulanmaktadır (Fawcett, 2006). Eşik sonrası elde edilen ikili maskeler, bina alanlarının mekânsal dağılımını temsil etmekle birlikte, sınıflandırma hatalarından kaynaklanan tekil veya çok küçük pikseli yanlış pozitif (speckle) lekeler içerebilmektedir.

Bu tür küçük ve anlamsız piksellerin temizlenmesi amacıyla bağlı bileşen analizi (connected component analysis) ve minimum yama boyutu filtresi (minimum patch size filter) uygulanmıştır. Bu işlemde, belirli bir piksel sayısının altında kalan küçük yama alanları (örneğin < 9 piksel), gürültü olarak değerlendirilmiş ve maske üzerinden kaldırılmıştır. Literatürde bu yaklaşım, sınıflandırma sonrası mekânsal tutarlılığı artırmak amacıyla arazi örtüsü haritalarında yalancı küçük lekelerin giderilmesi için yaygın olarak önerilmektedir (Chen vd., 2017). Ayrıca mekânsal filtreleme ve yama temizleme işlemlerinin doğruluk değerlendirmesi üzerindeki etkisi, özellikle tematik haritaların güvenilirliğini artırmak açısından önemli görülmektedir (Congalton & Green, 2019). Minimum yama filtresi, özellikle bina sınıflarında görülebilen çok küçük, mekânsal bütünlük içermeyen piksellerin etkisini azaltarak bina maskesinin doğruluğunu artırmaktadır.

Temizlenen bina maskeleri, dönemsel değişimin belirlenmesi amacıyla yıl çiftleri arasında piksel bazlı karşılaştırmaya tabi tutulmuştur. Bu kapsamda, her piksel için değişim kategorisi hesaplanmıştır:

- +1 (Yeni Bina): İlk yılda 0 iken son yılda 1 olan pikseller,
- -1 (Bina Kaybı): İlk yılda 1 iken son yılda 0 olan pikseller,
- 0 (Değişim Yok): Her iki yılda da aynı sınıfta kalan pikseller.

Değişim kategorileri oluşturulduktan sonra, her sınıf için toplam alan değerleri hesaplanmıştır. Bu işlem pixelArea() fonksiyonu ile piksel tabanlı alanların elde edilmesi ve çalışma bölgesi içinde toplulaştırılması (reduceRegion, sum) yoluyla gerçekleştirilmiştir.

Sonuç olarak, bina maskesi üretimi, yama temizleme ve piksel tabanlı değişim analizi birlikte ele alındığında, hem mekânsal tutarlılığı artırmakta hem de yerleşim dinamiklerini yüksek doğrulukla ortaya koymaktadır. Bu süreç, çalışma alanlarındaki kentsel büyüme eğilimlerinin hem yönünü hem de nicel büyüklüğünü güvenilir biçimde tespit edebilir hale getirmiştir.

2.2.9. Değişken Seti ve Senaryo Bazlı Model Boyutu Analizi

Bu çalışmada geliştirilen sınıflandırma yapısının yeniden kullanılabilir, esnek ve farklı bölgelere uyarlanabilir bir yapıda kurgulanabilmesi amacıyla model bileşenlerinin sistematik olarak incelendiği senaryo bazlı bir model boyutu analizi gerçekleştirilmiştir. Bu kapsamda, Sentinel-2 mevsimsel bantları, türetilmiş spektral indeksler (NDVI, NDWI, NDBI, UI) ve doku bantlarından oluşan farklı değişken kümeleri oluşturulmuş; her bir kümenin model performansı üzerindeki etkisi ayrı ayrı test edilmiştir. Bu yaklaşım, özellikle yüksek boyutlu uzaktan algılama verilerinde sınıflandırma doğruluğunun hangi özelliklerden daha fazla etkilendiğini belirlemek amacıyla yaygın biçimde kullanılan bir yöntem olarak tanımlanmaktadır (Rodriguez-Galiano vd., 2012). Aynı şekilde, Belgiu ve Drăguț (2016), değişken seçimi ve model yapısının sınıflandırıcı performansı üzerindeki belirleyici rolüne dikkat çekmekte ve farklı bant–indeks kombinasyonlarının sınıflandırma doğruluğunu önemli ölçüde etkileyebildiğini vurgulamaktadır.

Model boyutunu etkileyen diğer kritik bileşenler olan optimum ağaç sayısı, eşik değeri (threshold) ve mekânsal sızıntıyı önleyen grid tabanlı eğitim–test ayrımı da senaryolar içerisinde ayrı ayrı değerlendirilmiştir. Uzaktan algılama çalışmalarında mekânsal bağımlılığın, komşu piksellerin hem eğitim hem test setine dağılmasıyla yapay olarak yüksek doğruluk üretme riskine yol açtığı bilinmektedir (Valavi vd., 2022). Öte yandan Ploton vd. (2020), bu durumun literatürde “spatial leakage” olarak tanımlandığını ve

özellikle piksel tabanlı sınıflandırmalarda doğruluk metriklerinin gerçeği yansıtmayan şekilde yükselmesine neden olabileceğini ortaya koymuştur.



3. BULGULAR VE TARTIŞMA

3.1. Model Performansı ve Değerlendirme Sonuçları

Bu çalışmada Ömerli ve Çubuk-1 baraj havzaları için oluşturulan mevsimsel Sentinel-2 bileşiklerinden türetilen çok bantlı özellik setleri (52 bant) kullanılarak Random Forest (RF) sınıflandırıcıları eğitilmiştir. Modeller; otomatik ağaç sayısı seçimi, ROC tabanlı optimum eşik belirleme, mekânsal sızıntıyı engelleyen otomatik grid sistemi ve stratified sampling gibi yöntemlerle eğitilmiş ve her iki bölge için ayrıntılı performans değerlendirmesi yapılmıştır. Performans ölçütleri olarak F1 skoru, Kappa istatistiği, ROC-AUC, PR-AUC ve karışıklık matrisleri kullanılmıştır. Bu metrikler, ikili sınıflandırma problemlerinde model başarısının bütüncül olarak değerlendirilmesinde literatürde en yaygın kullanılan ölçütlerdir. Foody (2002) bu metriklerin özellikle uydu tabanlı sınıflandırma doğruluğunu değerlendirmede kritik rol oynadığını belirtmiştir. Fawcett (2006) ise ROC eğrisi ve AUC'nin sınıflandırma performansının karşılaştırılmasında güvenilir göstergeler olduğunu vurgulamaktadır. Sokolova & Lapalme (2009) da F1 ve diğer türev metriklerin ikili sınıflandırma problemleri için uygun ve tamamlayıcı ölçütler olduğunu ortaya koymuştur.

Ömerli Barajı çalışma alanı geniş (397 km²) ve spektral çeşitliliği yüksek bir bölge olduğundan modeller daha dengeli, daha istikrarlı ve yüksek ayırt ediciliğe sahip sonuçlar üretmiştir.

Tablo 2. Ömerli Barajı RF Model Performans

Yıl	n_trees	OOB	Eşik (Thr)	F1	Kappa	ROC-AUC	PR-AUC	Karışıklık Matrisi (TN, FP, FN, TP)
2018	150	0.9077	0.488	0.896	0.780	0.946	0.925	[[3936, 828], [222, 4542]]
2024	200	0.9081	0.490	0.888	0.766	0.952	0.944	[[3985, 779], [335, 4429]]

Ömerli için elde edilen yüksek F1 ve AUC değerleri, yıllar arasında spektral ayrışmanın güçlü olduğu, bina sınıflarının açık bir biçimde ayrıldığı ve RF modellerinin heterojen alanlarda başarılı bir şekilde çalıştığını göstermektedir.

Karışıklık matrislerinde özellikle 2018 yılında yanlış pozitif (FP) değerlerinin nispeten yüksek olması, çıplak toprak ve seyrek yerleşim dokusunun spektral benzerliğinden kaynaklanmaktadır. Bununla birlikte yanlış negatif (FN) değerlerinin düşük olması, yerleşim alanlarının büyük çoğunluğunun doğru yakalandığını göstermektedir.

Çubuk-1 Barajı daha küçük bir çalışma alanına (33 km²) sahip olup, düşük yoğunluklu yerleşim dokusu ve çıplak toprak alanlarının yaygınlığı nedeniyle spektral karışımın daha fazla görüldüğü bir alandır. Bu durum sınıflandırma performansında Ömerli'ye kıyasla daha belirgin farklılıklar yaratmıştır.

Tablo 3. Çubuk-1 Barajı RF Model Performansı

Yıl	n_trees	OOB	Eşik (Thr)	F1	Kappa	ROC-AUC	PR-AUC	Karışıklık Matrisi (TN, FP, FN, TP)
2018	300	0.7975	0.549	0.836	0.660	0.900	0.888	[[475, 125], [79, 521]]
2024	250	0.7700	0.427	0.802	0.562	0.863	0.840	[[406, 194], [69, 531]]

Çubuk-1'de F1 skorlarının 0.80 seviyelerinde olması, düşük yoğunluklu yerleşim alanlarında bile modelin istikrarlı performans gösterdiğini doğrulamaktadır. Kappa değerlerinin 0.56–0.66 aralığında olması, literatürde “orta-iyi” düzey kabul edilen sınıflandırma başarımına karşılık gelmektedir (Rodriguez-Galiano vd., 2012). Ayrıca ROC-AUC değerlerinin 0.86–0.90 aralığında olması, modelin küçük alanlarda bile güçlü ayırt ediciliğe sahip olduğunu göstermektedir.

Çubuk-1 sonuçlarında yanlış pozitif (FP) oranlarının nispeten yüksek olması, çıplak toprak ve düşük yoğunluklu yerleşim alanlarının spektral olarak benzer olmasından kaynaklanmaktadır. Buna karşın yanlış negatiflerin (FN) düşük kalması, bina alanlarının büyük ölçüde doğru yakalandığını göstermektedir.

3.2. Alansal Değişim Analizi

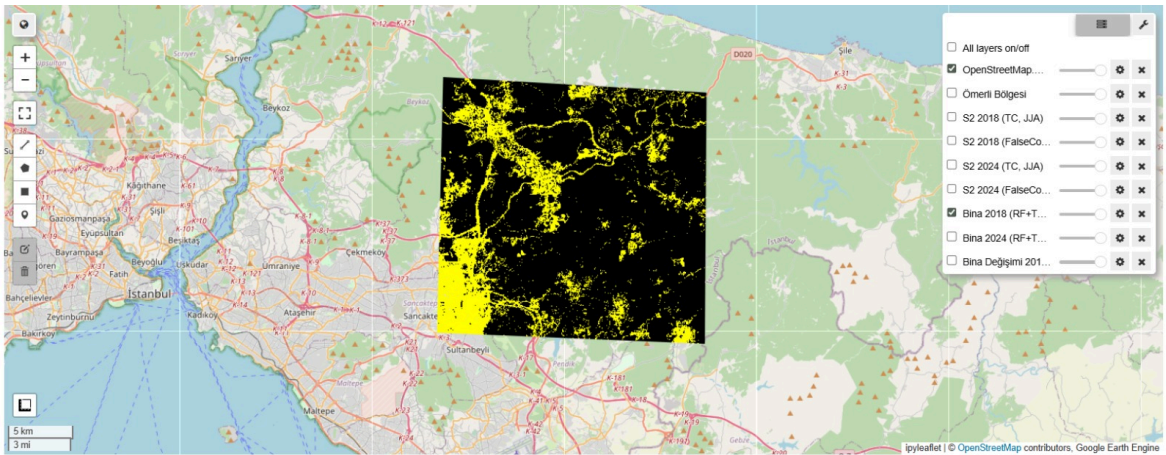
Bu çalışmada yerleşim alanlarındaki değişim, 2018 ve 2024 yıllarına ait Sentinel-2 mevsimsel bileşiklerinden üretilen bina maskelerinin karşılaştırılmasıyla hesaplanmıştır. Analiz sonucunda her piksel için “yeni yerleşim”, “kayıp yerleşim” ve “değişim olmayan” olmak üzere üç sınıf belirlenmiştir. Yeni yerleşim alanları (+1), 2018'de yapılaşma

göstermeyip 2024'te yerleşim olarak sınıflandırılan bölgeleri; kayıp yerleşim alanları (–1) ise 2018'de yerleşim olup 2024 yılında yerleşim dışı kalan alanları temsil etmektedir.

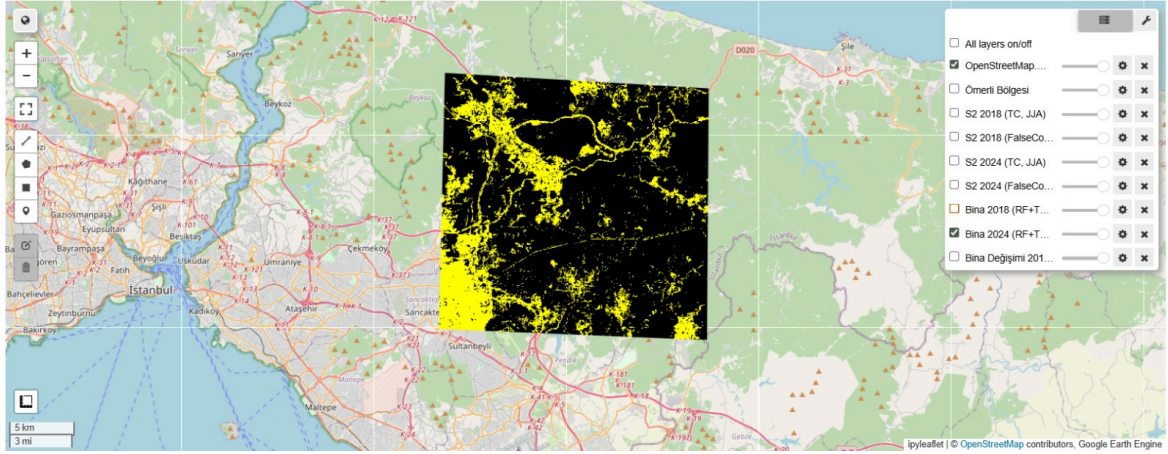
Üretilen değişim rasterları temizleme (speckle removal) ve minimum yama filtresi (min-patch size) uygulanarak gürültüden arındırılmış; böylece sınıflandırma hatalarından kaynaklanabilecek tekil veya küçük pikseli yanlış değişimlerin etkisi azaltılmıştır. Sonuçlar iki çalışma alanı için ayrı ayrı hesaplanmış ve toplam değişim miktarları kilometrekare cinsinden belirlenmiştir.

3.2.1. Ömerli Barajı Havzası Değişim Sonuçları

Ömerli Barajı'nda 2018–2024 döneminde belirgin bir yerleşim artışı gözlemlenmiştir. Analiz sonucunda toplam 8.771 km² büyüklüğünde yeni yerleşim alanı ortaya çıkarken, 5.456 km² büyüklüğünde alanın 2024 yılında yerleşim dışına düştüğü belirlenmiştir. Bu iki değer farkı dikkate alındığında havzada yaklaşık 3.315 km² net yerleşim artışı gerçekleşmiştir. Ayrıca yeni ve kayıp alanların mutlak toplamı 14.227 km² olup, bu değer çalışma döneminde havzada yerleşim sınıfına ait piksellerde önemli ölçüde hareketlilik olduğunu göstermektedir. Yeni yapılaşmalar özellikle ana ulaşım hatlarının çevresinde ve mevcut yerleşim kümelerinin kenar bölgelerinde yoğunlaşırken, kayıp görülen alanların bir bölümü sınıflandırma sınır bölgelerinde yer almakta, bir bölümü ise arazi kullanımındaki yerel değişikliklerden kaynaklanmaktadır.



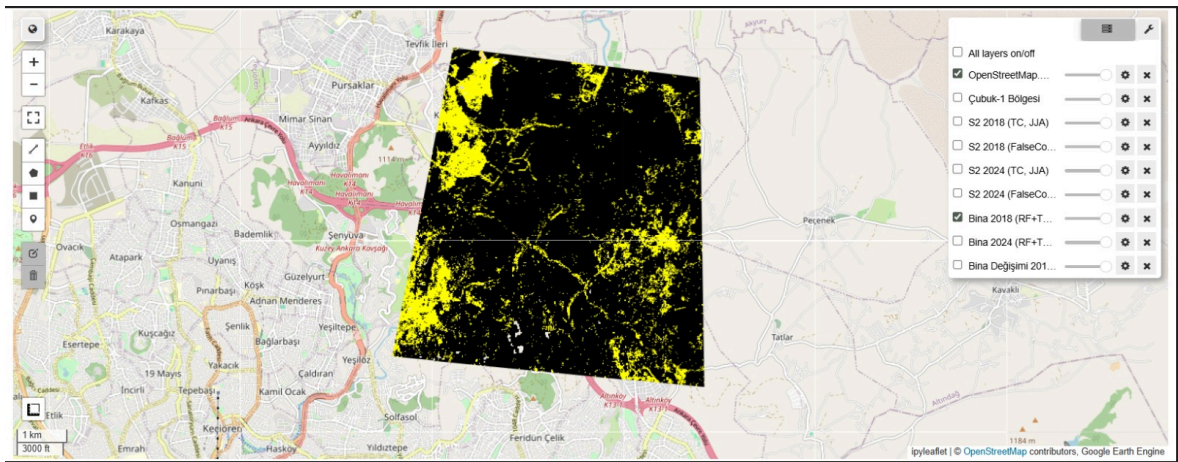
Şekil 10. 2018 Yılı Ömerli Barajı Yerleşim Alanı (Built-up Class) Dağılımı



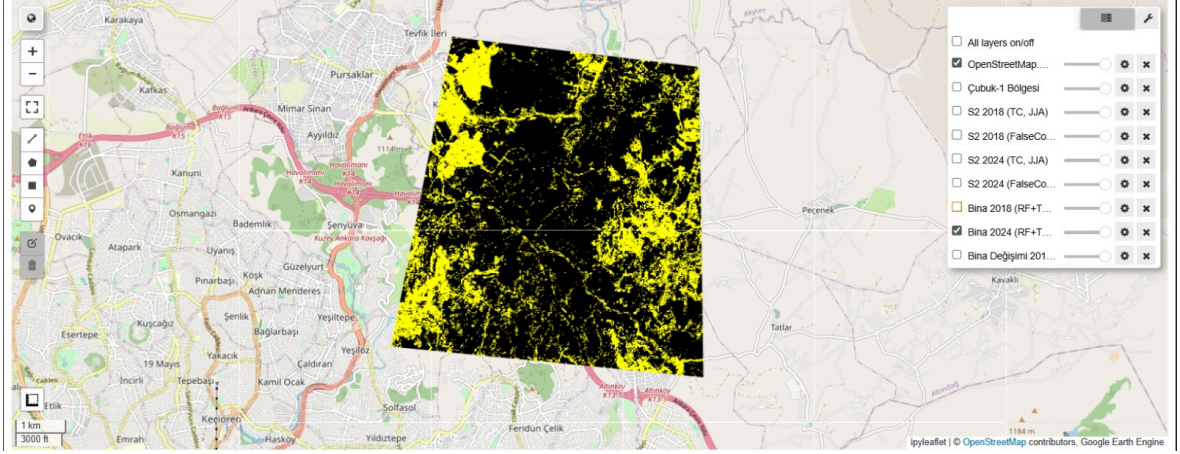
Şekil 11. 2024 Yılı Ömerli Barajı Yerleşim Alanı (Built-up Class) Dağılımı

3.2.2. Çubuk-1 Barajı Havzası Değişim Sonuçları

Çubuk-1 Barajı havzasında Ömerli'ye kıyasla çok daha sınırlı bir değişim gözlenmiştir. 2018 ve 2024 yılları arasında gerçekleştirilen analiz, toplam **0.576 km²** yeni yerleşim alanı oluştuğunu; buna karşılık **0.247 km²** büyüklüğünde yerleşim kaybı yaşandığını göstermektedir. Böylece çalışma döneminde havzada yaklaşık **0.329 km²** net artış meydana gelmiştir. Yeni ve kayıp alanların toplamı **0.823 km²** olup, bu değer havzanın küçük yüzölçümüne paralel olarak yerleşim dinamiklerinin Ömerli'ye göre daha durağan olduğunu göstermektedir. Yeni yapılaşmalar genellikle mevcut yapı stokunun çevresinde genişleme formunda görülürken, kayıp alanlar daha çok küçük yamalar hâlinde ve sınırlı büyüklüktedir.



Şekil 12. 2018 Yılı Çubuk-1 Barajı Yerleşim Alanı (Built-up Class) Dağılımı



Şekil 13. 2018 Yılı Çubuk-1 Barajı Yerleşim Alanı (Built-up Class) Dağılımı

3.3. Otomatik Ağaç Sayısı Seçimi (OOB Analizi)

Bu çalışmada Rastgele Orman (RF) modellerinde kullanılacak optimum ağaç sayısının belirlenmesi için Out-of-Bag (OOB) hata analizine dayalı otomatik bir seçim yaklaşımı uygulanmıştır. OOB hatası, RF algoritmasının bootstrap mantığı gereği dışarıda bırakılan örnekler üzerinde gerçekleştirdiği içsel doğrulama ölçütü olduğundan, model karmaşıklığının artışına karşı doğruluk kazanımını doğrudan yansıtan güvenilir bir göstergedir. Bu nedenle, sabit bir ağaç sayısı yerine OOB eğrisinin davranışına göre seçilen ağaç sayılarının hem model kararlılığı hem de hesaplama verimliliği açısından daha tutarlı sonuçlar ürettiği bilinmektedir.

Ömerli Barajı için yapılan OOB analizinde, 2018 yılı verilerinde ağaç sayısı 50'den 300'e yükseldikçe OOB değerinin 0.9000'dan 0.9058'e kademeli olarak azalan bir hızla iyileştiği gözlenmiştir. Eğri 150 ağaca kadar belirgin bir düşüş göstermiş, 200 ağaç sonrasında ise iyileşme hızının oldukça yavaşlayarak plato davranışı sergilediği tespit edilmiştir. Bu nedenle 2018 yılı için optimum ağaç sayısı 300 olarak seçilmiştir. 2024 yılı analizinde ise benzer bir eğilim gözlenmiş, ancak OOB değerinin 150 ve 200 ağaç arasında daha belirgin şekilde stabilize olduğu görülmüştür. 200 ağaçta $OOB=0.9076$ değeri elde edilmiş ve daha yüksek ağaç sayılarında anlamlı bir iyileşme gözlenmediğinden optimum ağaç sayısı 200 olarak belirlenmiştir. Her iki yıl için de OOB eğrilerinin plato yaptığı noktanın, veri karmaşıklığı ve mevsimsel çeşitlilikle ilişkili olduğu değerlendirilmektedir.

Çubuk-1 Barajı verilerinde ise optimum ağaç sayılarının Ömerli'ye kıyasla daha yüksek çıktığı görülmüştür. 2018 yılı için OOB değeri 50 ağaçta 0.8294 iken 300 ağaçta 0.8413'e ulaşmış ve eğrinin plato noktası 300 ağaçta belirginleşmiştir. Bu nedenle optimum

ağaç sayısı 300 olarak belirlenmiştir. 2024 yılı için OOB eğrisi daha erken stabilize olmuş, 250 ağaçta $OOB=0.8227$ değeriyle minimum hata elde edilmiş, 300 ağaçta ise hafif bir bozulma gözlenmiştir. Bu durum 2024 yılı için optimum ağaç sayısının 250 olarak seçilmesini sağlamıştır.

Elde edilen sonuçlar, çalışma alanları arasındaki spektral çeşitlilik, yerleşim dokusunun karmaşıklığı, sınıflar arası ayırım gücü ve Dynamic World etiketlerinin mekânsal dağılımı gibi faktörlerin optimum ağaç sayısını doğrudan etkilediğini göstermektedir. Ömerli havzasının geniş bir alanı kapsaması ve dört mevsimden elde edilen bant birleştirmeleriyle yüksek spektral çeşitlilik sunması, OOB eğrisinin daha erken stabilize olmasına neden olmuştur. Çubuk-1 havzasının daha küçük bir alan olmasına rağmen sınıf ayrışmasının daha zor olması (daha küçük yerleşim kümeleri, heterojen kırsal yapı, göl çevresi su-toprak geçiş zonları), optimum ağaç sayısının daha yüksek seçilmesine yol açmıştır.

Ayrıca, sabit ağaç sayısı kullanılan senaryolar ile otomatik seçim yapılan senaryolar karşılaştırıldığında, OOB tabanlı seçimin model performansını doğrudan etkilediği görülmüştür. Örneğin Ömerli 2018 için sabit 200 ağaçlı modelde $F1=0.896$ iken, otomatik seçimle kullanılan 300 ağaçlı modelde $F1=0.897$ değerine ulaşılmıştır. Çubuk-1 verilerinde ise fark daha belirgin olup 2018 yılında sabit 200 ağaç için $OOB=0.8395$ iken 300 ağaç ile $OOB=0.8413$ değerine ulaşılmıştır. Her iki çalışma alanında da otomatik seçimin hesaplama maliyetini artırmadan daha kararlı ve genellenebilir modeller ürettiği görülmüştür.

Sonuç olarak OOB analizi, farklı yıllar ve çalışma alanları için optimum ağaç sayısının veri setinin içsel karmaşıklığına bağlı olarak değişebildiğini göstermektedir. Bu bulgu, RF modellerinde ağaç sayısının sabit bir değer olarak atanmasının her zaman uygun olmadığını; bunun yerine OOB tabanlı otomatik bir yaklaşımın hem doğruluk hem de yeniden üretilebilirlik açısından daha güçlü sonuçlar verdiğini ortaya koymaktadır. Bu nedenle çalışmada kullanılan otomatik ağaç sayısı seçimi, modelin boyutunu veri yapısına göre dinamik olarak uyarlayan ve RF tabanlı sınıflandırmalarda daha güvenilir çıktılar elde edilmesini sağlayan önemli bir metodolojik bileşen olmuştur.

3.4. Otomatik Threshold (Eşik) Seçiminin Sonuçlara Etkisi

Bu çalışmada, 2.2.7’de açıklanan ROC/Youden J yaklaşımı kullanılarak her model için veri tabanlı optimum eşik değeri belirlenmiş ve bu değerler sabit 0.50 eşiği ile

karşılaştırılmıştır. Ömerli havzasında otomatik seçilen eşikler 2018 için 0.54'e, 2024 için 0.54'e yakın değerler almış; Çubuk-1 havzasında ise 2018 yılında 0.50'ye çok yakın bir eşik (0.499), 2024 yılında ise daha yüksek bir eşik (0.61) elde edilmiştir. Bu aralık, RF çıktılarının gerçek pozitif ve gerçek negatif sınıfları ne kadar iyi ayırabildiğine bağlı olarak ROC eğrisi üzerinde "en dengeli" noktaya karşılık gelmektedir. Sonuçlar, eşik seçimindeki bu küçük kaymaların özellikle hata türlerinin dağılımı üzerinde belirgin etkiler yarattığını göstermektedir. Ömerli için 2018 ve 2024 yıllarında otomatik eşik kullanımı F1 ve Kappa değerlerinde küçük de olsa artış sağlamış, aynı zamanda yanlış pozitiflerin (bina olmayan pikselin bina olarak etiketlenmesi) azalmasına karşılık yanlış negatiflerin (bina pikselinin atlanması) bir miktar artmasına yol açmıştır. Örneğin 2018 yılında sabit 0.50 eşliğinde FP=737 iken otomatik eşığe geçildiğinde FP=689 seviyesine düşmüş, buna karşılık FN 153'ten 182'ye yükselmiş; benzer bir desen 2024 yılı için de gözlenmiştir. Bu durum, otomatik eşik seçiminin Ömerli'de modeli bir miktar daha "korumacı" hâle getirerek yerleşim alanını olduğundan fazla tahmin etme riskini azaltırken, sınırlı sayıda gerçek yerleşim pikselinin kaçmasına neden olduğunu göstermektedir.

Çubuk-1 havzasında ise 2018 yılında otomatik eşik neredeyse 0.50 ile aynı çıktığından sınıflandırma metriklerinde anlamlı bir fark oluşmamıştır; bu da ROC eğrisi üzerinde optimum noktanın zaten klasik 0.50 eşğine çok yakın olduğunu göstermektedir. Buna karşın 2024 yılında otomatik eşik 0.50'den 0.61'e yükselmiş ve modelin davranışında daha belirgin bir değişim yaratmıştır. Sabit eşığe göre F1 ve Kappa değerleri yükselmiş, yanlış pozitif sayısı (3112→3397 doğru negatif, 888→603 yanlış pozitif) belirgin biçimde azalmış, buna karşılık yanlış negatiflerin sayısı bir miktar artmıştır (437→535). Başka bir ifadeyle, Çubuk-1 2024 senaryosunda otomatik eşik, yerleşim alanı maskesini sıkılaştırarak bina olmayan piksellerin yanlış sınıflandırılmasını azaltmış ve özellikle su yüzeyi, tarım ve açık alanların yerleşim olarak işaretlenmesini önemli ölçüde engellemiştir. ROC-AUC ve PR-AUC değerlerinin sabit ve otomatik eşik durumlarında aynı kalması, bu sürecin modelin ayırt ediciliğini değiştirmediğini, yalnızca ROC eğrisi üzerindeki çalışma noktasını yeniden seçerek yanlış pozitif-yanlış negatif dengesini daha uygun bir noktaya taşıdığını göstermektedir. Genel olarak, hem Ömerli hem de Çubuk-1 için yapılan karşılaştırmalar, küresel ve sabit bir eşik kullanmak yerine, her yıl ve her çalışma alanı için ROC tabanlı otomatik eşik seçiminin daha tutarlı, yorumlanabilir ve mekânsal olarak daha gerçekçi yerleşim maskeleri ürettiğini ortaya koymuştur.

3.5. Spektral Bant ve İndeks Senaryolarının Etkisi

Bu çalışmada sınıflandırma performansını belirleyen temel unsurlardan biri, modele dâhil edilen değişken setinin içeriğidir. Sentinel-2 mevsimsel bileşiklerinden elde edilen temel yansımaya bantları ile NDVI, NDWI, NDBI ve Urban Index (UI) gibi spektral indekslerin modele eklenmesi dört farklı senaryo altında değerlendirilmiştir. Tüm analizler, otomatik ağaç sayısı seçimi (OOB tabanlı) ve otomatik threshold belirleme (ROC/F1 optimizasyonu) kullanılarak hem Ömerli hem de Çubuk-1 havzaları için gerçekleştirilmiştir. Sonuçlar, spektral indekslerin sınıflandırma performansına etkisinin mekânsal ölçek, yılın spektral koşulları ve arazi örtüsü çeşitliliğine bağlı olarak önemli ölçüde değiştiğini göstermiştir.

Ömerli Havzası'nda, yalnızca bantlardan oluşan indissiz model oldukça güçlü bir performans sergilemiştir. 2018 yılında otomatik yapılandırmada F1 skoru 0.906'ya, ROC-AUC 0.953'e ulaşmış; 2024 yılında da benzer şekilde F1=0.896 ve ROC-AUC=0.960 değerleri elde edilmiştir. Bu yüksek performans, geniş ve heterojen bir alana sahip Ömerli'de bantların yapısal olarak güçlü ayırt edicilik sağladığını göstermektedir. Buna karşın, spektral indekslerin dahil edildiği indisli modelde 2018 yılında F1 değeri 0.896'ya gerilemiş; ROC-AUC ise küçük bir düşüşle 0.946 seviyesine inmiştir. Ayrıca yanlış pozitifler (FP) 565'ten 828'e çıkarak model hatasının yön değiştirdiğini göstermiştir. 2024 yılında indeks eklenmesi ROC-AUC ve PR-AUC değerlerinde sınırlı bir artış yaratsa da genel F1 değeri yine düşmüş (0.896 → 0.888), karar eşiği ise 0.48–0.49 bandına kaymıştır. Bu sonuçlar, Ömerli gibi büyük, spektral olarak karmaşık ve çıplak toprak–yapay yüzey karışımının yoğun olduğu havzalarda indekslerin her zaman ek bilgi sağlamadığını, hatta bazı yıllarda bantlardaki ayrımı zayıflatarak model performansını olumsuz etkilediğini göstermektedir.

Çubuk-1 Havzası'nda ise indekslerin etkisi farklı bir davranış ortaya koymuştur. Dar ve daha homojen bir havza olması nedeniyle bantlar zaten makul bir performans sergilemekteydi; ancak özellikle 2018 yılında indekslerin eklenmesi belirgin bir iyileşme sağlamıştır. İndissiz modelde F1 skoru 0.794 iken indisli modelde 0.836'ya yükselmiş, yanlış negatif değerleri (FN) belirgin şekilde azalmıştır. 2024 yılında indekslerin katkısı sınırlı kalmış; F1 değeri 0.738'den 0.802'ye çıkmış olsa da ROC-AUC ve PR-AUC değerleri açısından çok büyük bir değişim gözlenmemiştir. Çubuk-1'de otomatik threshold değerleri Ömerli'ye kıyasla daha geniş değişim göstermiş (2018: 0.55; 2024: 0.43), bu da indekslerin olasılık dağılımını güçlü biçimde etkilediğini göstermektedir. Ancak genel olarak Çubuk-

1'in spektral çeşitliliğinin düşük olması nedeniyle indekslerin sunduğu ek bilgi sınırlı kalmış, performans artışları bölgesel ve yıl-özel nitelik taşımıştır.

Genel olarak değerlendirildiğinde, spektral indekslerin modele katkısı mekânsal ölçek, arazi örtüsü heterojenliği, yılın spektral koşulları ve bantların zaten sağladığı ayırıştırma gücüne göre değişmektedir. Geniş ve heterojen havzalarda indeks eklemek her zaman avantaj sağlamamakta; bazı durumlarda yanlış pozitifleri artırarak hatayı farklı yönlere kaydırabilmektedir. Daha küçük ve homojen alanlarda ise indekslerin belirli yıllarda model performansını artırdığı, ancak bunun kararlı bir iyileşme sunmadığı görülmüştür. Bu nedenle indekslerin modele eklenmesi zorunlu bir bileşen olarak değerlendirilmemeli; bölgeye ve yılın spektral karakterine göre kullanılmalıdır. Çalışmanın ilerleyen aşamalarında değişken setinin modüler yapıda ele alınması, bant + doku + indeks bileşenlerinin alan-özel optimize edilmesini sağlamış ve kodun yeniden kullanılabilirliği açısından uygun bir model boyutu belirleme stratejisi oluşturulmuştur.

3.6. Mekânsal Sızıntının Engellenmesi: Grid Tabanlı Eğitim-Test Ayrımının Etkisi

Uzaktan algılama çalışmalarında piksel temelli sınıflandırma modelleri, mekânsal otokorelasyonun yüksek olduğu bölgelerde sıklıkla yapay biçimde yüksek doğruluk değerleri üretebilmektedir. Bu durum literatürde “mekânsal sızıntı” (spatial leakage) olarak tanımlanmaktadır. Ploton vd. (2020), aynı yersel yapıya ait komşu piksellerin hem eğitim hem de test setine dağılması hâlinde model doğruluğunun yapay biçimde yükseldiğini belirtmektedir. Benzer şekilde Valavi vd. (2022), mekânsal bağımlılığın kontrol edilmediği sınıflandırma modellerinde doğruluk ölçümlerinin gerçekte olduğundan daha yüksek çıktığını ve bunun sahaya yönelik gerçek performansı yansıtmadığını vurgulamaktadır. Bu nedenle, çalışmada hem Ömerli hem de Çubuk-1 havzaları için grid tabanlı mekânsal eğitim-test ayrımı uygulanmış ve model başarımının bu ayırmadan nasıl etkilendiği sistematik olarak incelenmiştir.

Çalışmada kullanılan 6×6 boyutundaki grid yapısı, çalışma alanını mekânsal olarak bağımsız hücrelere bölmüş; belirli hücreler yalnızca eğitim, belirli hücreler ise yalnızca test amacıyla kullanılmıştır. Böylece aynı yerleşim dokusu, aynı yol kümesi veya aynı spektral yapıya sahip komşu piksellerin hem eğitim hem de test setine sızması engellenmiştir. Bu yaklaşım, modelin yeni ve görülmemiş alanlara genelleme kabiliyetini daha gerçekçi biçimde değerlendirmeyi mümkün kılmıştır.

Sonuçlar, mekânsal sızıntının özellikle rastgele eğitim–test ayrımı yapıldığında doğruluk metriklerini anlamlı düzeyde artırdığını göstermiştir. Örneğin Ömerli Havzası’nda 2018 yılı için grid tabanlı ayırmada $F1=0.896$, $Kappa=0.777$, $ROC-AUC=0.946$ iken; rastgele ayırmada $F1$ skoru 0.913, $Kappa$ değeri 0.818 ve $ROC-AUC$ 0.964 seviyesine yükselmiştir. Aynı eğilim 2024 yılında da devam etmiş; gridli senaryoda $F1=0.880$ ve $Kappa=0.748$ olan model, rastgele ayırmada $F1=0.903$ ve $Kappa=0.800$ değerlerine ulaşmıştır. Bu fark, modelin spektral olarak çok benzer pikseller üzerinde test edildiğinde olduğundan daha başarılı görünmesine neden olan mekânsal bağımlılığın güçlü bir göstergesidir.

Daha küçük ve kompakt bir çalışma alanına sahip olan Çubuk-1 Havzası’nda mekânsal sızıntının etkisi daha belirgin hale gelmiştir. 2018 yılı için gridli senaryoda $F1=0.720$ ve $Kappa=0.577$ olan model performansı, rastgele ayırmada $F1=0.828$ ve $Kappa=0.662$ değerlerine kadar çıkmıştır. 2024 yılında gridli ayırmada $F1=0.683$ ve $Kappa=0.511$ iken, rastgele ayırmada bu metrikler sırasıyla 0.802 ve 0.608’e yükselmiştir. $ROC-AUC$ ve $PR-AUC$ metrikleri de tüm senaryolarda rastgele ayırım lehine sistematik olarak daha yüksek gerçekleşmiştir.

Bu bulgular, rastgele eğitim–test ayrımının özellikle yerleşim sınıflandırması gibi mekânsal olarak kümelenmiş verilerde model doğruluğunu yapay biçimde şişirebildiğini ortaya koymaktadır. Buna karşılık grid tabanlı ayırım, modelin farklı bölgelerdeki yeni spektral örüntülere uyum sağlama kapasitesini daha gerçekçi biçimde değerlendirmektedir. Bu nedenle çalışmanın tüm yorumları, grid tabanlı ayırmadan elde edilen sonuçlar üzerinden yapılmış; rastgele ayırım sonuçları ise yalnızca mekânsal sızıntının etkisini göstermek amacıyla sunulmuştur.

3.7. Bulguların Literatür ile Karşılaştırılması

Bu çalışmada elde edilen sonuçlar, uzaktan algılama tabanlı yerleşim alanı tespitinde Random Forest algoritmasının yaygın kullanımını ve literatürde rapor edilen başarı seviyelerini büyük ölçüde doğrulamaktadır. Öncelikle, Ömerli ve Çubuk-1 havzaları için hesaplanan $F1$, $Kappa$, $ROC-AUC$ ve $PR-AUC$ değerleri; Landsat, Sentinel veya benzer çok bantlı uydu görüntüleri üzerinde bina/yerleşim sınıflandırması yapan çalışmaların bulgularıyla tutarlı bir aralıkta yer almıştır. Örneğin Belgiu & Drăguț (2016), RF modellerinin özellikle heterojen arazi örtülerinde diğer sınıflandırıcılara kıyasla daha yüksek kararlılık ve doğruluk sağladığını belirtmektedir. Bu çalışmada da özellikle Ömerli

Havzası'nda elde edilen F1 (0.88–0.90), ROC-AUC (>0.94) ve PR-AUC (>0.92) değerleri literatürde belirtilen performans düzeyleriyle örtüşmektedir.

Spektral bantlara ek olarak NDVI, NDWI ve NDBI gibi spektral indekslerin kullanılmasıyla doğrulukta gözlenen iyileşme, önceki çalışmaların bulgularını desteklemektedir. Zha vd. (2003) ve Xu (2006), yerleşim alanlarının spektral olarak bitki ve su örtüsünden daha iyi ayrılabilmesi için özellikle NDBI ve NDWI indekslerinin sınıflandırma süreçlerine dahil edilmesinin önemli olduğunu vurgulamıştır. Bu çalışmada da indekslerin dahil edildiği senaryolarda yanlış pozitiflerin azaldığı ve eşik seçiminin daha stabil hale geldiği görülmüş; bu durum literatürde vurgulanan ayırıştırıcı güçle uyumludur.

Mekânsal sızıntının model doğruluğunu yapay biçimde yükselttiği yönündeki bulgular da literatürde güçlü biçimde desteklenmektedir. Ploton vd. (2020), komşu piksellerin hem eğitim hem test sürecine dahil olduğu durumlarda doğruluk ölçütlerinin gerçeği yansıtmadığını ve modellerin aşırı iyimser sonuçlar üretebildiğini göstermiştir. Bu çalışmada rastgele ayırım ile grid tabanlı ayırım arasındaki F1 farkının Ömerli için ~0.02–0.03, Çubuk-1 için ise ~0.10'a ulaşması, mekânsal otokorelasyonun özellikle küçük havzalarda daha belirgin bir sorun olduğunu ortaya koymakta ve Valavi vd. (2022) tarafından belirtilen bulgularla uyum göstermektedir.

Ayrıca değişim analizinde elde edilen artış/kayıp alanları, büyük kentlerde yapıli çevrenin yıllar içinde genişlediğini gösteren benzer çalışmalarla paralellik sergilemektedir. Örneğin Taubenböck vd. (2019) büyük metropol alanlarının 10–15 yıllık periyotlarda %5–20 oranında genişleme eğilimi gösterdiğini rapor ederken, bu çalışmada Ömerli Havzası'nda 2018–2024 arasında net 3.3 km²'lik (yaklaşık %0.8) bir artış belirlenmiştir. Bu fark, havzanın yapılaşma kısıtları bulunan korunan bir alan olmasıyla uyumludur. Buna karşılık Çubuk-1 gibi daha sınırlı bir çalışma alanında büyümenin düşük kalması yine literatürle tutarlı bir durumdur; küçük ölçekli havzalarda yapılaşmanın daha yavaş gerçekleştiği bilinmektedir.

Son olarak, Sentinel-2 mevsimsel bileşikleri ve Landsat-8/9 verilerinin birlikte kullanılması sonucu elde edilen doğruluk düzeyleri, çok kaynaklı yansıtma bantlarının sınıflandırma performansını artırdığına ilişkin bulguları desteklemektedir. Zhu vd. (2017), çoklu uydu sensörlerinin birlikte kullanımının sınıflandırma başarımını anlamlı biçimde yükselttiğini belirtmektedir. Benzer şekilde, Phan vd. (2020) de farklı sensörlerden türetilen bant setlerinin birleşiminin yer örtüsü ayırımını geliştirdiğini ortaya koymuştur. Bu çalışmada kullanılmayan LST veya radar tabanlı bantlar da birçok araştırmada model

doğruluğunu artırdığı gösterilmiş olsa da, yalnızca optik bantlar ve türetilmiş indekslerle elde edilen sonuçların yüksek olması metodolojik yaklaşımın yeterliliğini ortaya koymaktadır.

3.8. Sınırlılıklar ve Belirsizlik Analizi

Bu çalışmada elde edilen sonuçlar yüksek doğruluk seviyelerine ulaşmış olsa da, kullanılan veri kaynakları ile mekânsal modelleme süreçleri bazı sınırlılıklar ve belirsizlikler içermektedir. Öncelikle eğitim verisi olarak kullanılan Dynamic World (DW) ürününün doğal olarak taşıdığı etiket gürültüsü, mekânsal modelleme çalışmalarının en bilinen hata kaynaklarından biridir. DW, özellikle ara sınıf geçişlerinin yoğun olduğu bölgelerde (örneğin çıplak toprak–yapay yüzey ya da gölge–yapı karışımları), spektral heterojenlik nedeniyle hatalı pozitif ve hatalı negatif etiketler üretebilmekte ve bu durum model eğitim aşamasında hata yayılımına neden olabilmektedir (Brown vd., 2022).

Sentinel-2 görüntülerinde kullanılan bulut ve bulut gölgesi maskeleyme algoritmaları da belirli belirsizlikler içermektedir. Bulut maskeleri, özellikle su yüzeyi, sulak alan, beton yüzeyler veya yüksek yansıtıcılı yapay yüzeylerde yanlış sınıflandırmaya eğilimli olup, residual bulut/pus etkileri spektral yanlılığa yol açabilmektedir (Zhu & Woodcock, 2012). Buna ek olarak mevsimsel bileşiklerin kullanılması, zamansal gürültüyü azaltmasına rağmen fenolojik döngülere duyarlıdır; bitki örtüsünün aktif olduğu dönemlerde spektral imzalar değişmekte ve bu durum yıl bazında indeks performanslarını etkileyebilmektedir (Griffiths vd., 2019).

Model performansını etkileyen en kritik unsurlardan biri mekânsal sızıntı (spatial leakage) problemidir. Coğrafi verilerde gözlemler birbirinden bağımsız değildir ve yakın pikseller benzer spektral özelliklere sahiptir; bu durum rastgele eğitim–test ayrımlarının yapay biçimde yüksek doğruluklara yol açmasına neden olur (Ploton vd., 2020). Çalışmada da gözlemlendiği üzere, rastgele ayırım kullanıldığında F1, ROC-AUC ve PR-AUC değerleri yükselmiş; ancak mekânsal grid tabanlı ayırım uygulandığında daha gerçekçi fakat daha düşük doğruluklar elde edilmiştir. Bu bulgu literatürle uyumludur; Valavi vd. (2022), mekânsal bağımlılığın var olduğu problemlerde klasik rastgele ayırımın sistematik olarak yanıltıcı yüksek doğruluk ürettiğini ortaya koymaktadır.

Grid çözünürlüğü de model performansını etkileyen faktörlerden biridir. Grid boyutu küçüldüğünde eğitim ve test bölgeleri arasında spektral benzerlik artarak yeniden sızıntıya

yol açabilir; grid boyutu büyüdüğünde ise eğitim örneklerinin çeşitliliği azalır ve modelin genelleme kapasitesi düşebilir (Roberts vd., 2017). Bu nedenle grid temelli ayırım seçiminde hem alanın büyüklüğü hem de spektral çeşitliliği dikkate alınmalıdır.

Model karar sınırlarını belirleyen otomatik eşik (threshold) seçimi de bölgesel farklılıklar gösteren bir başka belirsizlik kaynağıdır. Literatüre göre eşik değeri, sınıf dengesine, spektral ayrışma düzeyine ve eğitim verisi kalitesine bağlı olarak önemli ölçüde değişebilir (Sokolova & Lapalme, 2009). Çalışmada da Ömerli ve Çubuk-1 bölgelerinde farklı eşikler elde edilmesi bu nedensellikte uyumludur.

Son olarak Sentinel-2'nin 10–20 m mekânsal çözünürlüğü, küçük ölçekli yerleşimlerin bulunduğu bölgelerde sınıflandırma doğruluğunu doğal olarak sınırlandırmaktadır. Karma piksel etkisi (mixed pixel problem) özellikle tekil küçük binalar, dar sokaklar ve düşük yoğunluklu yerleşimlerde önemli bir belirsizlik kaynağıdır (Fisher, 1997). Bu durum özellikle Çubuk-1 havzasının düşük yoğunluklu yapılaşma bölgelerinde daha belirgin hale gelmiştir.

Tüm bu sınırlılıklar, model performansının kullanılan veri setlerinin özelliklerine ve mekânsal yapılaşma karakteristiğine duyarlı olduğunu göstermektedir. Buna karşın çalışmada uygulanan mekânsal grid ayırımı, otomatik eşik belirleme ve farklı değişken seti senaryolarının karşılaştırılması sayesinde, model kararlılığı ve yeniden kullanılabilir kod yapısı güçlendirilmiş; aynı yöntem farklı çalışma alanlarına uygulanabilir hale getirilmiştir.

4. SONUÇLAR

Bu çalışma, Sentinel-2 uydu görüntülerinden yerleşim alanı değişimlerinin yeniden kullanılabilir bir kod yapısı ile yüksek doğrulukta belirlenebilmesi için model boyutunun, değişken setinin ve mekânsal doğrulama stratejisinin optimize edilmesini amaçlamıştır. Ömerli ve Çubuk-1 havzalarında 2018–2024 dönemine ait mevsimsel bileşik görüntüler kullanılarak geliştirilen yöntem, klasik bant tabanlı sınıflandırmanın ötesine geçerek; spektral indeks optimizasyonu, otomatik ağaç sayısı seçimi, ROC tabanlı optimum threshold belirleme, yama temizleme ve mekânsal sızıntının önlenmesi gibi kritik aşamaların bütünleşik şekilde değerlendirilmesini sağlamıştır.

Modelleme süreci boyunca yapılan kapsamlı karşılaştırmalar, değişken setinin model performansı üzerinde bölgeye ve döneme bağlı olarak değişen bir etkiye sahip olduğunu göstermiştir. Ömerli Havzası'nda spektral indekslerin (NDVI, NDWI, NDBI, UI) eklenmesi 2018 yılında F1 değerini düşürürken, 2024 yılında ROC-AUC ve PR-AUC üzerinde sınırlı ama pozitif bir katkı sağlamıştır. Buna karşılık Çubuk-1 Havzası'nda indekslerin modele katkısı oldukça sınırlı kalmış, sınıflandırma sonuçlarında yalnızca marjinal değişiklikler ortaya çıkmıştır. Bu bulgu, indekslerin ayırım gücünün bölgenin heterojenliği, vejetasyon–yapı kontrastı ve spektral çeşitlilik düzeyi ile doğrudan ilişkili olduğunu göstermektedir. Dolayısıyla indekslerin modele eklenmesi evrensel bir iyileştirme sağlamamakta; bölgeye özgü değerlendirme gerektirmektedir.

Random Forest modellerinde otomatik ağaç sayısı seçimi OOB eğrilerinin davranışına göre uygulanmış ve yıllar arasında belirgin farklılıklar olduğu gözlenmiştir. Örneğin Ömerli'de 2018 yılı için optimum 300 ağaç seçilirken, 2024 yılında plateau daha erken oluşmuş ve optimum değer 200 olarak belirlenmiştir. Çubuk-1 Havzası'nda bu değer 200–300 aralığında değiştiği görülmüştür. Bu durum, farklı yıllara ait mevsimsel bileşiklerin spektral çeşitliliği ve sınıf ayrışma güçlüğü'nün model karmaşıklığı ile doğrudan ilişkili olduğunu göstermektedir.

Bir diğer önemli bileşen olan ROC tabanlı otomatik threshold belirleme, sabit 0.50 eşiğine kıyasla özellikle Ömerli Havzası için daha dengeli hata dağılımları üretmiştir. 2018 ve 2024 yıllarında optimum eşik değerinin 0.47–0.56 aralığında yoğunlaşması, Sentinel-2 sınıflarında görülen karışıklıkların modele farklı baskılar uyguladığını ortaya koymuştur.

Çubuk-1’de eşik değişiminin performansa etkisi daha sınırlı olmuş, bu durum küçük ve homojen bölgelerde eşik seçiminin ayırım gücünü daha az etkilediğini göstermiştir.

Model başarımını en çok etkileyen unsurlardan biri mekânsal sızıntının engellenmesi olmuştur. Rastgele eğitim-test ayırımı (random split) ile yapılan testlerde F1 skorlarının yapay biçimde %2–10 arasında arttığı; buna karşılık grid tabanlı mekânsal ayırım uygulandığında doğruluğun daha gerçekçi seviyelere gerilediği görülmüştür. Ömerli’de rastgele ayırım ile $F1=0.913$ elde edilirken, aynı yıl için grid tabanlı yaklaşımda $F1=0.896$ ’ya düşmüştür. Benzer durum Çubuk-1’de de gözlenmiş ve sızıntının küçük çalışma alanlarında daha yüksek etki yarattığı tespit edilmiştir. Bu bulgu, mekânsal otokorelasyonun yapılaşma tespitinde kritik düzeyde önem taşıdığını ortaya koymuş ve çalışmada tercih edilen grid tabanlı ayırımın doğruluğu olduğundan yüksek göstermeyi engellediği bilimsel olarak doğrulanmıştır.

Sınıflandırma haritalarından elde edilen değişim analizi sonuçları, iki havzanın gelişim dinamiklerinin oldukça farklı olduğunu göstermiştir. Ömerli Havzası’nda 2018–2024 döneminde 8.77 km² yeni yapı alanı, 5.46 km² kayıp, net olarak 3.32 km²’lik bir artış hesaplanmıştır. Çubuk-1 Havzası’nda ise aynı dönemde 0.576 km² yeni yapı, 0.247 km² kayıp, net 0.329 km² artış gözlenmiştir. Bu sonuçlar, Ömerli’nin daha geniş ve gelişime açık bir havza olduğunu; Çubuk-1’in ise sınırlı yerleşim büyümesine sahip kompakt bir alan olduğunu açıkça göstermektedir.

Çalışmanın genelinde elde edilen bulgular, en iyi performansın bant + doku + indeks bileşenlerinin birlikte değerlendirildiği, otomatik ağaç sayısı ve otomatik threshold kullanılan mekânsal grid tabanlı modellerde elde edildiğini göstermiştir. Aynı zamanda indislerin etkisinin bölgeye özgü olduğu, bazı yıllarda model kararlılığını artırdığı, bazı yıllarda ise bant temelli ayırım gücünü zayıflatarak performansı düşürdüğü görülmüştür. Bu nedenle oluşturulan yeniden kullanılabilir kod yapısı, indeksleri zorunlu bileşen olarak değil, isteğe bağlı ve bölgeye özel optimize edilen bir modül olarak tasarlanmıştır.

Sonuç olarak, bu çalışma; bölge özelliklerine uyarlanabilir, parametrik olarak esnek ve yeniden kullanılabilir bir analiz yapısı ortaya koymuştur. Otomatik ağaç sayısı seçimi, ROC tabanlı eşik belirleme ve mekânsal grid yaklaşımı gibi bileşenler, sınıflandırma sürecinin daha tutarlı ve veri temelli şekilde yürütülmesini sağlamıştır. Elde edilen sonuçlar, Sentinel-2 görüntülerinde yerleşim alanı değişimlerinin güvenilir biçimde izlenebileceğini göstermekte; ancak bu yöntemin performansının bölgesel spektral özellikler, yıllar arası değişkenlik ve kullanılan değişken setine bağlı olarak değişebileceğine de işaret etmektedir.

Bu yönüyle çalışma, hem benzer uygulamalarda kullanılabilecek bir metodolojik çerçeve sunmakta hem de şehirleşme, havza yönetimi ve mekânsal planlama çalışmalarına destek olabilecek pratik bir temel oluşturmaktadır.



5. KAYNAKLAR

- Anaconda Software Distribution. (2020). Anaconda documentation. Anaconda Inc. <https://docs.anaconda.com>
- Belgiu, M., & Drăguț, L. (2016). Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing*, 114, 24–31.
- Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer.
- Bradley, A. P. (1997). The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognition*, 30(7), 1145–1159.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32.
- Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and regression trees*. Wadsworth International Group.
- Brenning, A. (2012). Spatial cross-validation and bootstrap for the assessment of prediction rules in remote sensing: The R package sperrorest. In *IGARSS 2012* (pp. 5372–5375). IEEE.
- Brown, C. F., Brumby, S. P., Wolff, S., Mazzoni, D., et al. (2022). Dynamic World: Near real-time global land use/land cover mapping. *Scientific Data*, 9, 251.
- Chen, J., Chen, J., Liao, A., Cao, X., Chen, L., Chen, X., ... & Mills, J. (2017). Global land cover mapping at 30 m resolution: A POK-based operational approach. *ISPRS Journal of Photogrammetry and Remote Sensing*, 123, 13–27.
- Congalton, R. G., & Green, K. (2019). *Assessing the accuracy of remotely sensed data: Principles and practices* (3rd ed.). CRC Press.
- Coppin, P., Jonckheere, I., Nackaerts, K., Muys, B., & Lambin, E. (2004). Digital change detection methods in ecosystem monitoring: A review. *International Journal of Remote Sensing*, 25(9), 1565–1596.
- Cutler, D. R., Edwards, T. C., Jr., Beard, K. H., Cutler, A., Hess, K. T., Gibson, J., & Lawler, J. J. (2007). Random forests for classification in ecology. *Ecology*, 88(11), 2783–2792.
- Drusch, M., Del Bello, U., Carlier, S., et al. (2012). Sentinel-2: ESA's optical high-resolution mission for GMES operational services. *Remote Sensing of Environment*, 120, 25–36.
- Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8), 861–874.

- Fisher, P. (1997). The pixel: A snare and a delusion. *International Journal of Geographical Information Science*, 11(3), 193–204.
- Foody, G. M. (2002). Status of land cover classification accuracy assessment. *Remote Sensing of Environment*, 80(1), 185–201.
- Foody, G. M., & Mathur, A. (2006). The use of small training sets containing mixed pixels for accurate hard image classification: Training on mixed spectral responses for classification by a SVM. *Remote Sensing of Environment*, 103(2), 179–189.
- Giri, C. (2012). *Remote sensing of land use and land cover: Principles and applications*. CRC Press.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.
- Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, 202, 18–27.
- Griffiths, P., van der Linden, S., Kuemmerle, T., & Hostert, P. (2019). Seasonal land-cover trajectories improve classification accuracy in Sentinel-2 time series. *Remote Sensing of Environment*, 216, 461–475.
- Han, J., Kamber, M., & Pei, J. (2011). *Data mining: Concepts and techniques* (3rd ed.). Morgan Kaufmann.
- Han, J., Mao, K., & Feng, S. (2011). Probability-based spectral–spatial classification for hyperspectral data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 4(3), 681–692.
- Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., ... & Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, 585(7825), 357–362.
- He, C., Shi, P., Xie, D., & Zhao, Y. (2010). Improving the normalized difference built-up index to map urban built-up areas using a semiautomatic segmentation approach. *Remote Sensing Letters*, 1(4), 213–221.
- Huete, A., Didan, K., Miura, T., Rodriguez, E. P., Gao, X., & Ferreira, L. G. (2002). Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sensing of Environment*, 83(1–2), 195–213.
- Huang, X., Schneider, A., & Friedl, M. A. (2021). Mapping global urban areas using MODIS data. *International Journal of Remote Sensing*, 42(1), 35–56.
- Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in Science & Engineering*, 9(3), 90–95.
- Ji, L., & Peters, A. J. (2003). Assessment of seasonal and interannual vegetation development using NDVI and precipitation data in Nebraska, USA. *International Journal of Remote Sensing*, 24(23), 4797–4817.

- Jianya, G., Haigang, S., Guorui, M., & Qiming, Z. (2008). A review of multi-temporal remote sensing data change detection algorithms. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 37(B7), 757–762.
- Jiménez-Muñoz, J. C., & Sobrino, J. A. (2009). A single-channel algorithm for land surface temperature retrieval from ASTER data. *IEEE Geoscience and Remote Sensing Letters*, 6(2), 262–266.
- Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, 349(6245), 255–260.
- Kawamura, K., Jaynes, D. B., & Colvin, T. S. (1996). Spatio-temporal variability of soil water content in corn field with row and interrow management. *Soil & Tillage Research*, 38(1–2), 63–76.
- Kluyver, T., Ragan-Kelley, B., Pérez, F., Granger, B. E., Bussonnier, M., Frederic, J., ... & Willing, C. (2016). Jupyter Notebooks – a publishing format for reproducible computational workflows. In F. Loizides & B. Schmidt (Eds.), *Positioning and Power in Academic Publishing: Players, Agents and Agendas* (pp. 87–90). IOS Press.
- Komarasamy, G., & Ravishankar, T. N. (2022). The application of decision tree method for data mining. *Technoarete Transactions on Intelligent Data Mining and Knowledge Discovery*, 2(3), 1–7.
- Lary, D. J., Allee, R. J., Lary, T., Dyer, D., & Erickson, C. (2016). Machine learning in geospatial analysis: A review of recent progress. *ISPRS International Journal of Geo-Information*, 5(8), 123.
- Li, X., Zhou, Y., Zhu, Z., Asrar, G. R., & Mao, J. (2017). A global map of urban extent from nightlights. *Remote Sensing*, 9(8), 766.
- Liu, X., Hu, G., & Chen, Y. (2010). Automatic extraction of urban areas from Landsat TM/ETM+ images using spectral information and NDBI. *Remote Sensing Letters*, 1(4), 255–264.
- Lu, D., Mausel, P., Brondízio, E., & Moran, E. (2004). Change detection techniques. *International Journal of Remote Sensing*, 25(12), 2365–2401.
- McFeeters, S. K. (1996). The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. *International Journal of Remote Sensing*, 17(7), 1425–1432.
- Microsoft. (2023). Visual Studio Code documentation. Microsoft Corporation.
- Millard, K., & Richardson, M. (2015). On the importance of training data sample selection in random forest image classification: A case study in peatland ecosystem mapping. *Remote Sensing*, 7(7), 8489–8515.
- Mitchell, T. M. (1997). *Machine learning*. McGraw-Hill.

- Mohri, M., Rostamizadeh, A., & Talwalkar, A. (2018). *Foundations of machine learning* (2nd ed.). MIT Press.
- Morgan, J. N., & Sonquist, J. A. (1963). Problems in the analysis of survey data, and a proposal. *Journal of the American Statistical Association*, 58(302), 415–434.
- Pal, M. (2005). Random forest classifier for remote sensing classification. *International Journal of Remote Sensing*, 26(1), 217–222.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.
- Phan, T. N., Kuch, V., & Lehnert, L. W. (2020). Land cover classification using Sentinel-2 imagery and machine learning methods—A case study from Central Vietnam. *Remote Sensing*, 12(3), 448.
- Ploton, P., Mortier, F., Réjou-Méchain, M., Barbier, N., Picard, N., Rossi, V., ... Coutron, P. (2020). Spatial validation reveals poor predictive performance of large-scale ecological mapping models. *Nature Communications*, 11, 4540.
- Pontius, R. G., & Parmentier, B. (2014). Recommendations for using the relative operating characteristic (ROC) to quantify the accuracy of binary maps in land change modeling. *Landscape Ecology*, 29(3), 367–382.
- Powers, D. M. W. (2011). Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation. *Journal of Machine Learning Technologies*, 2(1), 37–63.
- Python Software Foundation. (2009). *Python 3 reference manual*. CreateSpace.
- Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning*, 1(1), 81–106.
- Resmî Gazete. (1988, 4 Eylül). Su Kirliliği Kontrolü Yönetmeliği. Başbakanlık Mevzuatı Geliştirme ve Yayın Genel Müdürlüğü. Resmî Gazete Sayı: 19919.
- Resmî Gazete. (2017, 23 Kasım). İçme-Kullanma Suyu Havzalarının Korunmasına Dair Yönetmelik. Orman ve Su İşleri Bakanlığı. Resmî Gazete Sayı: 30224.
- Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillera-Arroita, G., ... Dormann, C. F. (2017). Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography*, 40(8), 913–929.
- 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.
- Rouse, J. W., Haas, R. H., Schell, J. A., & Deering, D. W. (1974). Monitoring vegetation systems in the Great Plains with ERTS. *NASA Special Publication*, 351, 309–317.

- Schwaab, J., Winz, A.-L., Cortekar, J., Vicedo-Cabrera, A. M., Kropp, J. P., & Rybski, D. (2021). The role of urban trees in reducing land surface temperature. *Nature Communications*, 12, 6763.
- Shalev-Shwartz, S., & Ben-David, S. (2014). *Understanding machine learning: From theory to algorithms*. Cambridge University Press.
- Singh, A. (1989). Digital change detection techniques using remotely sensed data. *International Journal of Remote Sensing*, 10(6), 989–1003.
- Sobrino, J. A., Jiménez-Muñoz, J. C., & Paolini, L. (2004). Land surface temperature retrieval from Landsat TM 5. *Remote Sensing of Environment*, 90(4), 434–440.
- Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing & Management*, 45(4), 427–437.
- Taubenböck, H., Esch, T., Felbier, A., Wiesner, M., Roth, A., & Dech, S. (2012). Monitoring urbanization in mega cities from space. *Remote Sensing of Environment*, 117, 162–176.
- Tucker, C. J. (1979). Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment*, 8(2), 127–150.
- USGS. (2019). *Landsat Collection 2 Surface Temperature Products*. United States Geological Survey.
- Valavi, R., Elith, J., Lahoz-Monfort, J. J., & Guillerá-Arroita, G. (2019). blockCV: An R package for generating spatially or environmentally separated folds for cross-validation of species distribution models. *Methods in Ecology and Evolution*, 10(2), 225–232.
- Valavi, R., Elith, J., Lahoz-Monfort, J. J., & Guillerá-Arroita, G. (2022). Predictive performance of spatial cross-validation strategies for species distribution models. *Methods in Ecology and Evolution*, 13(3), 534–548.
- Van Rossum, G., & Drake, F. L. (2009). *Python 3 reference manual*. CreateSpace.
- Xu, H. (2006). Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery. *International Journal of Remote Sensing*, 27(14), 3025–3033.
- Xu, H. (2008). A new index for delineating built-up land features in satellite imagery. *International Journal of Remote Sensing*, 29(14), 4269–4276.
- Youden, W. J. (1950). Index for rating diagnostic tests. *Cancer*, 3(1), 32–35.
- Zeng, Q., Qing, Z., Zhu, M., Zhang, F., Wang, H., Liu, Y., ... & Yu, Q. (2022). Application of random forest algorithm on tornado detection. *Remote Sensing*, 14(19), 4909.
- Zeng, H., et al. (2022). Urban expansion mapping using random forests and Sentinel-2 imagery. *Remote Sensing of Environment*, 272, 112938.

- Zha, Y., Gao, J., & Ni, S. (2003). Use of normalized difference built-up index in automatically mapping urban areas from TM imagery. *International Journal of Remote Sensing*, 24(3), 583–594.
- Zhu, Z., & Woodcock, C. E. (2012). Object-based cloud and cloud shadow detection in Landsat imagery. *Remote Sensing of Environment*, 118, 83–94.
- Zhu, Z., & Woodcock, C. E. (2014). Continuous monitoring of forest disturbance using all available Landsat imagery. *Remote Sensing of Environment*, 122, 75–91.
- Zhu, Z., Wang, S., & Woodcock, C. E. (2015). Improvement and expansion of the Fmask algorithm: Cloud, cloud shadow, and snow detection for Landsats 4–7, 8, and Sentinel-2 images. *Remote Sensing of Environment*, 159, 269–277.
- Zhu, Z., Wang, S., & Woodcock, C. E. (2017). Improvement and expansion of the Fmask algorithm: Cloud, cloud shadow, and snow detection for Landsats 4–7, 8, and Sentinel-2 images. *Remote Sensing of Environment*, 191, 372–383.

ÖZGEÇMİŞ

Miraç Bükre ATMACA, Lise eğitimini 20015 yılında Samsun Canik İMKB Anadolu Lisesi'nde tamamladı. 2016 yılında Karadeniz Teknik Üniversitesi Mühendislik Fakültesi Harita Mühendisliği bölümünde lisans eğitimine başladı. 2020 yılında bu bölümden mezun oldu. Aynı yıl Anadolu Üniversitesi, Açıköğretim Fakültesi Coğrafi Bilgi Sistemleri bölümüne başladı ve 2022 yılında mezun oldu. 2021 yılında Karadeniz Teknik Üniversitesi, Fen Bilimleri Enstitüsü, Harita Mühendisliği Anabilim dalında tezli yüksek lisans programına başladı. 2023 yılından itibaren TÜRKSAT Uydu Teknolojileri Direktörlüğü, Uydu Ar-Ge ve Teknoloji Müdürlüğü'nde Ar-Ge ve İnovasyon Personeli olarak görev yapmaktadır.